



Fabián López
Teresa Pérez
Manuel Prieto
Felipe Ramírez
Arturo Reyes
Francisco Treviño

Introducción a los Modelos de Análisis Multivariante

Casos de uso Práctico en los Negocios y la
Investigación Científica



**Aprenda
Ediciones**

ISBN 978-968-5980-06-7



9 789685 980067 >

Introducción a los Modelos de Análisis Multivariante.

Casos de uso práctico en los negocios y en la investigación científica.

Fabián López

FACPYA, UANL, México

Teresa Pérez

Universidad Regiomontana, México

Manuel Prieto

Universidad Regiomontana, México

Felipe Ramírez

FACPYA / FIME, México

Arturo Reyes

FACPYA, UANL, México

Francisco Treviño

FIME, UANL, México



MÉXICO • ARGENTINA • BRASIL • COLOMBIA • COSTA RICA • CHILE
ESPAÑA • GUATEMALA • PERÚ • PUERTO RICO • VENEZUELA

**López, Fabián; Pérez, Teresa; Prieto, Manuel;
Ramírez, Felipe; Reyes, Arturo; Treviño, Francisco.
(2010).**

*Introducción a los modelos de análisis multivariante.
Casos de uso práctico en los negocios y en la
investigación científica.*

Primera Edición,

Aprenda Ediciones,

México. ISBN: 978-968-5980-06-7

Formato: 21.59 x 27.94 cm.

Págs: 431

PRIMERA EDICIÓN, 2010.

DERECHOS RESERVADOS

© Aprenda Ediciones

Ventas: ventas@aprenda.mx

Información de productos,

conferencias, cursos: info@aprenda.mx

Web site: www.aprenda.mx

Reservados todos los derechos. Ni la totalidad ni parte de esta publicación, así como los materiales complementarios que acompañan la obra, pueden reproducirse, registrarse o transmitirse, por un sistema de recuperación de información, en ninguna forma ni por ningún medio, sea electrónico, mecánico, fotoquímico, magnético o electroóptico, por fotocopia, grabación o cualquier otro conocido o por conocer, sin permiso previo y por escrito del titular de los derechos.

El software, productos y marcas utilizadas en este libro son propiedad intelectual de sus autores, y su uso queda sujeto a los términos del contrato licencia que aparece al instalar los mismos, así como de la legislación vigente.

El préstamo, alquiler o cualquier otra forma de cesión de uso de este ejemplar requerirá también la autorización del titular de los derechos o de sus representantes.



Índice

Índice	iii
--------------	-----

Capítulo 1: Introducción al análisis multivariable

Importancia del análisis multivariante.....	2
Etapas para la aplicación del análisis multivariable	6
Instrumento de medición	8
Tipos de escalas de medida.....	8
Construcción del instrumento.....	12
Escalas de respuesta.....	14
Pruebas de validez y fiabilidad.....	15
Prueba de consistencia interna (alfa Cronbach).....	18
Selección de los ítems	20
Tamaño de la Muestra y Ponderación.....	22
Inferencia estadística en el análisis multivariable	24
Hipótesis	27
Análisis inicial y pruebas en los datos	28
Supuesto de normalidad	30
Supuesto de linealidad y errores no correlacionados	31
Supuesto de homocedasticidad	33
Índice del capítulo	34

Capítulo 2: Técnicas de análisis multivariable

Clasificación de técnicas de análisis multivariable	38
Análisis de la varianza	39
Análisis de varianza con un factor.....	41
Análisis de varianza de dos factores con una sola muestra por grupo.....	43
Análisis de varianza de dos factores con varias muestras por grupo.....	44
Análisis de varianza de tres factores con varias muestras por grupo.....	45
Modelo de regresión lineal múltiple	47
Introducción al análisis de regresión múltiple	48
Supuestos de la regresión para mínimos cuadrados ordinarios	50
Medición del ajuste del modelo de regresión	51
Fundamentos estadísticos del modelo de regresión	53
Significación del modelo de regresión.....	59
Validación del modelo	61
Identificación de puntos atípicos e influyentes	62
La multicolinealidad	65
Índice del capítulo	67

Capítulo 3: Las herramientas estadísticas para la investigación

Herramientas estadísticas para la investigación	70
Introducción	70
Principales errores en la utilización de las herramientas estadísticas.	71
Solución Propuesta: Estrategia general de aplicación de las herramientas estadísticas.	73
Modelo Propuesto: Metodología de aplicación de las herramientas.	76
Estudio de los conjuntos de observaciones.	77
La distribución Normal de un conjunto de observaciones.	80
Otras distribuciones estadísticas de datos.	82
Tamaño de las muestras.	83
Diseño de instrumentos de recolección de datos.	85
Confiabilidad del Instrumento por medio del Coeficiente Alfa de Cronbach.	92
Relaciones entre variables: Regresión Lineal Simple y Múltiple.	98
Análisis para identificación de clusters o subconjuntos de elementos relacionados dentro de una población.	107
Análisis de datos para comprobación de comportamiento normal prueba de Kolmogorov-Smirnov.	109
Comparación de dos conjuntos de observaciones: pruebas t-Student y Mann-Whitney.	111
Comparación de más de dos conjuntos de observaciones. ANOVA.	120
Conclusiones.	125
Referencias	126
Índice del capítulo	128

Capítulo 4: Factores críticos de éxito de e-Gobierno en México

¿Qué se encontrará en este caso?	130
Temática de la investigación.	130
Técnicas utilizadas en el caso	130
Marco teórico	131
Definición de gobierno electrónico	131
Eficacia de los portales ¿Cómo medirla?	132
Estudios antecedentes de la investigación	135
Análisis y declaración del problema	143
Experimento	145

Preguntas de investigación	145
Hipótesis	146
Tamaño de la muestra	149
Diseño del instrumento	150
Recopilación de datos.....	153
Operacionalización de las variables	159
Diseño experimental	162
Planteamiento estadístico de la hipótesis	162
Aplicación de pruebas estadísticas	164
Comprobación de hipótesis.....	185
Conclusiones y recomendaciones.....	186
Respuestas a las preguntas de investigación	186
Conclusiones	187
Recomendaciones	188
Bibliografía	188
Índice del capítulo	195

Capítulo 5: Análisis multivariante en la administración de inventarios

Introducción al caso de estudio	198
Planteamiento del problema de investigación	198
Antecedentes.....	198
Declaración del problema y propósito del estudio	199
Objetivos Específicos de la Investigación.....	200
Justificación y Limitaciones	201
Antecedentes e Investigación bibliográfica.....	202
Teoría enfocada a la estratificación de inventarios.....	202
Sistema de punto de re-orden de inventarios	202
Control con corrección anticipante	203
Técnicas de control dirigidas al futuro	203
La teoría de Gantt.....	204
La teoría de PERT	204
Obsolescencia, su naturaleza e importancia.....	205
La teoría del ciclo de vida del producto	209
Servicio de Inventario	212
Prácticas administrativas	213
Solución propuesta e hipótesis	218
Definición teórica de las Variables	218
Comprobación de la hipótesis	219
Instrumento de Medición, recolección y diseño de la base de datos	221

Diseño de la Investigación	221
Población, marco muestral y muestra	222
Selección empírica de los ítems	225
Instrumento de Investigación	226
Evaluación de las propiedades Psicométricas de la escala	226
Método de Análisis	227
Desarrollo del Diseño experimental	229
Construcción de Variables para el sector de la in- dustria de la Transformación.....	231
Modelo de Regresión Lineal Múltiple	235
Resultados y discusión.	243
Modelo con Variables ponderadas para el sector Transformación.....	244
Coefficientes de las variables independientes y su prueba T	245
Modelo con Variables escaladas para el sector Transformación.....	247
Aceptación de las variables	247
Coefficiente de determinación del modelo	247
Valor y significancia de la prueba F.	248
Coefficientes de las variables independientes y su prueba T	249
Conclusiones y líneas de investigación.....	251
Bibliografía.....	253
Índice del capítulo.....	255

Capítulo 6: Análisis y comparación de grupos de trabajo

Análisis y Comparación de Grupos de Trabajo.	258
Formación y Análisis de los equipos de trabajo utilizando méto- dos cuantitativos.	259
Introducción al caso de estudio. El desarrollo de una herramienta generadora de aplicaciones de software y la formación de equipos de trabajo para probarla.	259
Antecedentes e Investigación Bibliográfica.	262
Propuesta e Hipótesis.....	266
Modelo Propuesto para la Hipótesis.	267
Recolección de datos	274
Desarrollo y diseño del experimento.....	274
Procedimientos y mediciones utilizadas en el estu- dio.....	279
Limitaciones y alcance del estudio.....	280
Resultados Obtenidos	281
El Modelo Utilizado	287
Análisis de la sensibilidad de los equipos del expe- rimento aplicación del modelo.....	287

Comentarios sobre los resultados.....	295
Conclusiones y líneas de investigación futura.....	295
Referencias.....	296
Índice del capítulo	297

Capítulo 7: Administración del conocimiento y capital intelectual

Introducción	300
Desarrollo del caso	301
Generalidades del caso	301
Planteamiento del problema.....	302
Objetivos específicos y generales	304
Justificación de la investigación	304
Antecedentes e investigación bibliográfica.....	305
Solución propuesta e hipótesis	320
Metodología y diseño de la investigación.....	328
Análisis y Discusión de los Resultados	335
Conclusiones y líneas de investigación futura.....	369
Bibliografía.....	371
Índice del capítulo	374

Capítulo 8: Disponibilidad de los sistemas computacionales

Introducción	378
Planteamiento del problema.....	379
Justificación de la investigación	380
Objetivos.....	380
Alcance y limitaciones.....	380
Antecedentes e investigación bibliográfica.....	381
Principales escuelas del pensamiento	381
Solución propuesta e hipótesis	395
Variable dependiente	396
Variables independientes.....	397
Preguntas moderadoras.....	402
Hipótesis	403
Hipótesis nula	403
Hipótesis de investigación.....	403
Modelo propuesto para la prueba de la hipótesis	404
Elaboración de la encuesta.....	405
Recolección de datos.....	406

Determinación de la población.....	406
Descripción de la muestra	407
Selección de la muestra representativa	408
Trabajo de campo.....	408
Ficha técnica de la investigación	409
Análisis de resultados.....	409
Análisis descriptivo	410
Análisis estadístico	413
Aceptación o rechazo de la hipótesis nula	422
Comprobación de las hipótesis de investigación.....	423
Contribuciones.....	424
Comentarios generales.....	426
Líneas futuras de investigación.....	426
Referencias	428
Índice del capítulo.....	431
Índice de tablas.....	ix
Índice de figuras.....	xv
Los autores	xix
Prólogo	xxiii

Índice de tablas

Tabla 3.1	Problemas más frecuentes en el uso de las herramientas estadísticas	72
Tabla 3.2	Conjunto de Observaciones	73
Tabla 3.3	Clasificación de las hipótesis de los distintos tipos de estudio	76
Tabla 3.4	Valores de z para diferentes porcentajes de certidumbre	79
Tabla 3.5	Resumen de herramientas a utilizar para validar el instrumento	90
Tabla 3.6	Paquetes de Software para el pre-proceso de los datos estadísticos	92
Tabla 3.7	Resultado del SPSS al calcular el alfa de Cronbach de los datos	94
Tabla 3.8	Resultado del SPSS al calcular el alfa de Cronbach de los datos	95
Tabla 3.9	Resultado del SPSS al calcular el alfa de Cronbach de los datos	95
Tabla 3.10	Matriz de correlación de Pearson para los datos del ejemplo	96
Tabla 3.11	Tabla de Observación del ejemplo para Demostrar Hipótesis	99
Tabla 3.12	Resultados para los datos del Ejemplo	100
Tabla 3.13	Aproximación cuadrática de los datos del ejemplo	102
Tabla 3.14	Resultados para los modelos lineal, cuadrático y cúbico del ejemplo	103
Tabla 3.15	Ejemplo ampliado a dos variables independientes	104
Tabla 3.16	Resultados para la regresión lineal múltiple de dos valores	105
Tabla 3.17	Matriz de correlación de Pearson para el ejemplo	106
Tabla 3.18	Área de construcción y tipos de edificios en una ciudad , ejemplo de datos	107
Tabla 3.19	Área de construcción y datos alternos para calcular el tipo de edificio	109
Tabla 3.20	Resultados de la prueba de normalidad	111
Tabla 3.21	Salarios netos de los empleados de cuatro diferentes empresas	114
Tabla 3.22	Equivalente a la tabla 3.21 pero con menos columnas y mas observaciones	116
Tabla 3.23	Resultados de la prueba de normalidad Kolmogorov-Smirnov para los salarios	117
Tabla 3.24	Resultados de comprar las empresas 1 y 2	117

Tabla 3.25	Resultados de comprar las empresas 2 y 3	118
Tabla 3.26	Resultados de comparar las empresas 1 y 2 con la prueba U de Mann - Whitney	119
Tabla 3.27	Resultados de comparar las empresas 2 y 3 con la prueba U de Mann - Whitney	119
Tabla 3.28A	Datos para comparar métodos de enseñanza en universidades	122
Tabla 3.28B	Datos para comparar métodos de enseñanza codificando el método	123
Tabla 3.29	Resultados de la prueba de normalidad Kolmogorov-Smirnov para los grupos de Aprovechamiento	124
Tabla 3.30	Descriptivas Estadísticas por grupo.	125
Tabla 3.31	Prueba de homogeneidad de varianzas de Levene significancia mayos .05	125
Tabla 3.32	El análisis de Varianzas ANOVA	125
Tabla 4.1	Matriz de concurrencia entre preguntas de investigación e hipótesis	148
Tabla 4.2	Cobertura de categorías de información de información con variables seleccionadas	151
Tabla 4.3	Fundamentos teórico de las variables seleccionadas en la investigación	152
Tabla 4.4	Selección de variables	157
Tabla 4.5	Estadísticas Descriptivas	172
Tabla 4.6	Prueba Kolmogorow - Smirnov para normalidad	173
Tabla 4.7	Coefficientes de correlación, paramétricos y no paramétricos y su significancia	174
Tabla 4.8	Interpretación de los coeficientes de correlación, paramétricos y no paramétricos, entre las variables	175
Tabla 4.9	Análisis de Varianza para el modelo completo	176
Tabla 4.10	coeficientes de Determinación, para los modelos ajustados en el procedimiento "Backward" o hacia atrás	178
Tabla 4.11	Tablas Anova para los modelos ajustados	182
Tabla 4.12	Tabla de comparación de modelos ajustado propuestos por el procedimiento Backward considerando su significancia	183
Tabla 4.13	Estimaciones mínimo cuadráticas y su significado	185
Tabla 5.1	Mapa del Marco Teórico	205
Tabla 5.2	Clasificación de la obsolescencia de acuerdo a su naturaleza	207
Tabla 5.3	Marco Teórico sobre las prácticas Administrativas de Inventarios	213
Tabla 5.4	Definición teórica de las Variables	218

Tabla 5.5	Prueba de Alpha de Cronbach para el índice CVP de Transformación	231
Tabla 5.6	Prueba de KMO para índice de CVP de Transformación	232
Tabla 5.7	Prueba de Alpha de Cronbach para el índice SI, sector de Transformación	233
Tabla 5.8	Prueba de KM Opara índice SI, Sector de Transformación	233
Tabla 5.9	Prueba de alpha de Cronbach para variable PA en Transformación	234
Tabla 5.10	Prueba de KMO para variables PA en Transformación	234
Tabla 5.11	Prueba de Normalidad para la variable OI14	237
Tabla 5.12	Normalidad para las variables ponderadas del sector Transformación	240
Tabla 5.13	Normalidad para las variables escaladas del Sector Transformación	240
Tabla 5.14	No auto correlación en las variables ponderadas del Sector Transformación	241
Tabla 5.15	No auto correlación en las variables ponderadas del Sector Transformación	241
Tabla 5.16	Multicolinealidad en las variables ponderadas del Sector Transformación	242
Tabla 5.17	Multicolinealidad con Stepwise en las variables ponderadas del Sector Transformación	243
Tabla 5.18	Multicolinealidad en las variables escaladas del Sector Transformación	243
Tabla 5.19	Aceptación de variables modelo	244
Tabla 5.20	Coefficiente de determinación del modelo	244
Tabla 5.21	Valor y significancia de la prueba F en el modelo	245
Tabla 5.22	Coefficientes de las variables y su prueba T	246
Tabla 5.23	Aceptación de variables al modelo	247
Tabla 5.24	Coefficiente de determinación del modelo	247
Tabla 5.25	Valor y significancia de la prueba F en el modelo	248
Tabla 5.26	Coefficientes de las variables y su prueba T	249
Tabla 6.1	Tabla de perfiles de Belbin	265
Tabla 6.2	Matriz "E" de descripción de un equipo mediante	268
Tabla 6.3	Diseños no-experimentales tipo encuestas	277
Tabla 6.4	Diseños Pre-experimentales	277
Tabla 6.5	Diseños Experimentales	278
Tabla 6.6	Diseños Experimentales	278
Tabla 6.7	Modelo Cuasi Experimental del presente estudio	279
Tabla 6.8	Resultados de la prueba clasificando el desarrollo por el tipo de programador	281

Tabla 6.9	Equipos de Trabajo Participantes	282
Tabla 6.10	Estadísticas descriptivas de la composición de los equipo de trabajo	283
Tabla 6.11	Prueba de Normalidad Kolmogorov-Smirnov para el grupo de control	283
Tabla 6.12	Prueba de Normalidad Kolmogorov-Smirnov para el grupo experimental	284
Tabla 6.13	Prueba de Normalidad Kolmogorov-Smirnov para ambos grupos	285
Tabla 6.14	Comparación de los dos equipos de trabajo	285
Tabla 6.15	Comparación de los dos equipos de trabajo utilizando la prueba "t"	286
Tabla 7.1	Niveles de aprendizaje	306
Tabla 7.2	Definiciones de los elementos de un Sistema de Aprendizaje Organizacional	307
Tabla 7.3	Distribución por sector económico de las 200 empresas más grandes de Mty	329
Tabla 7.4	Distribución por nivel organizacional de los encuestados	332
Tabla 7.5	Distribución por sector económico	333
Tabla 7.6	Valores de la medida muestral, variable muestral, mínimos y máximos	333
Tabla 7.7	Valor de Alfa de Cronbach del instrumento de medición	337
Tabla 7.8	Valores de la Media por constructo, Desviación Estándar, Varianza, asimetría y Curtosis	337
Tabla 7.9	Valores de la Media muestral, Varianza muestral, Mínimos y Máximos	338
Tabla 7.10	Sumario de modelos de regresión lineal múltiple para el primer grupo de hipótesis	346
Tabla 7.11	Coefficientes de impacto del modelo de regresión lineal múltiple	347
Tabla 7.12	Análisis de varianza de ANOVA del modelo de regresión lineal múltiple para el primer grupo	347
Tabla 7.13	Modelo de regresión lineal múltiple para el primer grupo de hipótesis	348
Tabla 7.14	Sumario de modelos de regresión lineal múltiple para el segundo grupo de hipótesis	349
Tabla 7.15	coeficientes de impacto del modelos de regresión lineal múltiple para el segundo grupo de hipótesis	350
Tabla 7.16	Análisis de Varianza ANOVA del modelo de regresión lineal múltiple para el segundo grupo de hipótesis	350
Tabla 7.17	Modelo de regresión lineal múltiple para el segundo grupo de hipótesis	351
Tabla 7.18	Sumario de modelos de regresión lineal múltiple para	352

	el tercer grupo de hipótesis	
Tabla 7.19	Coefficientes de impacto del modelos de regresión lineal múltiple para el tercer grupo de hipótesis	352
Tabla 7.20	Análisis de Varianza ANOVA del modelo de regresión lineal múltiple para el tercer grupo de hipótesis	353
Tabla 7.21	Modelo de regresión lineal múltiple para el tercer grupo de hipótesis	354
Tabla 7.22	Sumario de modelos de regresión lineal múltiple para el cuarto grupo de hipótesis	355
Tabla 7.23	Coefficientes de impacto del modelo de regresión lineal múltiple para el cuarto grupo de hipótesis	355
Tabla 7.24	Análisis de Varianza ANOVA del modelo de regresión lineal múltiple para el cuarto grupo de hipótesis	356
Tabla 7.25	Modelo de regresión lineal múltiple para el cuarto grupo de hipótesis	358
Tabla 7.26	Sumario de los modelos de regresión lineal múltiple para el quinto grupo de hipótesis	358
Tabla 7.27	Coefficientes de impacto del modelos de regresión lineal múltiple para el quinto grupo de hipótesis	359
Tabla 7.28	Análisis de varianza ANOVA del modelo de regresión lineal múltiple para el quinto grupo de hipótesis.	360
Tabla 7.29	modelo de regresión lineal múltiple para el quinto grupo de hipótesis	361
Tabla 7.30	Sumario de modelos de regresión lineal múltiple para el sexto grupo de hipótesis	362
Tabla 7.31	Coefficientes de impacto del modelo de regresión lineal múltiple para el sexto grupo de hipótesis	363
Tabla 7.32	Análisis de varianza ANOVA del modelo de regresión lineal múltiple para el sexto grupo de hipótesis	364
Tabla 7.33	Modelo de regresión lineal múltiple para el sexto grupo de hipótesis	365
Tabla 7.34	Sumario de modelos de regresión lineal múltiple para el séptimo grupo de hipótesis	366
Tabla 7.35	Coefficientes de impacto del modelo de regresión lineal múltiple para el séptimo grupo de hipótesis	366
Tabla 7.36	Análisis de varianza ANOVA del modelo de regresión lineal múltiple para el séptimo grupo de hipótesis	367
Tabla 7.37	Modelo de regresión lineal múltiple para el séptimo grupo de hipótesis	368
Tabla 8.1	Variable Dependiente	396
Tabla 8.2	Variable Estructura Organizacional en TI y la Función de la SI	398
Tabla 8.3	Variable Herramientas de Protección	399
Tabla 8.4	Variable Seguridad Lógica, Física y ambiental	400
Tabla 8.5	Variable Cultura en SI	401

Tabla 8.6	Variable Clima Organizacional	402
Tabla 8.7	Preguntas Moderadas	403
Tabla 8.8	Ficha técnica de la investigación	409
Tabla 8.9	Resultados del Coeficiente alfa de Cronbach por variable	411
Tabla 8.10	Resultados de la R y R2 -Primer Modelo	414
Tabla 8.11	Resultados ANOVA- Primer Modelo	414
Tabla 8.12	Significancia a Detalle - Primer Modelo	415
Tabla 8.13	Resultados de la R y R2 -Segundo Modelo	417
Tabla 8.14	Resultados de la ANOVA- Segundo Modelo	417
Tabla 8.15	Resultados de la F	418
Tabla 8.16	Significancia a detalle del segundo modelo	418
Tabla 8.17	Resultados de la R y R2 - Tercer Modelo	419
Tabla 8.18	Resultados de la ANOVA- Tercer Modelo	419
Tabla 8.19	Resultados de la F	420
Tabla 8.20	Resultados de la Significancia- Tercer Modelo	420
Tabla 8.21	Resultados de ANOVA - Modelo seleccionado	421
Tabla 8.22	Del análisis de las Variables independientes del segundo modelo	422
Tabla 8.23	Resultados de la R y R2	423
Tabla 8.24	Significancia a Detalle	423

Índice de figuras

Figura 2.11	Gráfica del coeficiente R^2_p en función del número de regresores p .	54
Figura 2.2	Zona de aceptación y rechazo del estadístico F .	58
Figura 3.1	Universo y muestra de un conjunto de observaciones	77
Figura 3.2	Densidad de distribución normal estándar	81
Figura 3.3	Ejemplo de Datos para Calcular coeficiente de confiabilidad Alfa Cronbach	93
Figura 3.4	Calculo del coeficiente aplicando la ecuación 3.10 en una hoja de calculo	94
Figura 3.5	Cálculo del Alfa de Cronbach mediante el procedimiento de correlaciones	97
Figura 3.6	Los datos del Ejemplo Graficados	100
Figura 3.7	Línea recta que modela la relación entre las variables	101
Figura 3.8	Aproximaciones Lineal y aproximación cuadrática	102
Figura 3.9	Aceptación y Rechazo de la Hipótesis nula para Anova	121
Figura 4.1	Divergencia entre visitas a los portales y su posición en el ranking de portales gubernamentales estatales	134
Figura 4.2	Origen de los factores críticos de éxito de los portales gubernamentales	137
Figura 4.3	Línea del Tiempo de los tipos de estudio relativo al gobierno electrónico	140
Figura 4.4	Proceso de identificación de factores que explican el número de visitas los portales gubernamentales en México	141
Figura 4.5	Línea de tiempo. Estudios pertinentes relativos al gobierno electrónico	142
Figura 4.6	Secuencia lógica de análisis y pruebas estadísticas 1ra parte	167
Figura 4.7	Secuencia lógica de análisis y pruebas estadísticas 2da parte	170
Figura 4.8	Nivel de significancia y zona de rechazo de la investigación	171
Figura 5.1	Curva del ciclo de vida del producto Levitt (1965)	210
Figura 5.2	Gráfica 5.2	217
Figura 5.3	Modelo de Obsolescencia en los inventarios	220
Figura 5.4	Prueba de Homocedasticidad	236

Figura 5.5	P-P de Normalidad en la variable OI14	238
Figura 5.6	Distribución Normal de la Variable Dependiente OI14	239
Figura 5.7	Distribución Normal corregida de la nueva variable Log OI14	239
Figura 6.1	Productividad y tamaño del equipo	263
Figura 6.2	Resumen del análisis del Equipo1	271
Figura 6.3	Modelo utilizado. Área de alimentación de datos.	287
Figura 6.4	Modelo Aplicado al Equipo de Control Primera Parte	288
Figura 6.5	Modelo Aplicado al Equipo de Control Segunda Parte	288
Figura 6.6	Modelo Aplicado al Equipo de Control Resumen de Resultados	289
Figura 6.7	Modelo Aplicado al Equipo Experimental Primera Parte	290
Figura 6.8	Modelo Aplicado al Equipo Experimental Segunda Parte Criterios 4	290
Figura 6.9	Modelo Aplicado al Equipo Experimental Resumen de Resultados	291
Figura 6.10	Formulas del Criterio 1 tamaño del Equipo	292
Figura 6.11	Formulas para Cálculos del Criterio 2 Diversidad	293
Figura 6.12	Formulas para el Criterio 2 Diversidad	293
Figura 6.13	Formulas para el Criterio 3 Balanceo	294
Figura 6.14	Formulas para el criterio 4 personalidad de Líder	294
Figura 7.1	Aprendizaje Organizacional como un proceso Dinámico	308
Figura 7.2	Teoría de la Creación del Conocimiento Organizacional	316
Figura 7.3	Modelo de la relación del compromiso Organizacional con un Sistema de aprendizaje	321
Figura 7.4	Modelo Hipotético del Impacto del compromiso organizacional	326
Figura 7.5	Modelo Hipotético del Impacto del compromiso Organizacional en la administración de conocimiento	328
Figura 7.6	Correlación de las Variables en el modelo del Impacto del compromiso Organizacional	342
Figura 7.7	Correlación de las variables en el modelo del impacto del compromiso organizacional.	343
Figura 7.8	Impacto y su significancia estadística del compromiso organizacional en los componentes de la administración	345

Figura 8.1	Estructura de la encuesta	406
Figura 8.2	Resultados del Coeficiente alfa de Cronbach por variable	411

Los autores

Dr. Fabián López

El Dr. Fabián López es profesor e investigador en el Centro de Desarrollo Empresarial y Postgrado de la Facultad de Contaduría Pública y Administración, en la Universidad Autónoma de Nuevo León (CEDEEM/FACPyA/UANL, México), desde hace 15 años. Miembro del Sistema Nacional de Investigadores (SIN) Nivel 1. En 2004 fue el 1er graduado del programa de Doctorado en Filosofía con especialidad en Administración (FACPyA), obteniendo una mención honorífica *Summa Cum Laude*. Tiene una Maestría en Administración y otra en Ingeniería Industrial por la Escuela de Graduados en Administración y Dirección de Empresas del Instituto Tecnológico de Estudios Superiores de Monterrey (EGADE/ITESM, México). Tiene el grado CPIM (*Certified in Production & Inventory Management*) por la *American Production & Inventory Control Society* (APICS). Es Ingeniero en Sistemas, Ingeniero en Control e Ingeniero en Electrónica y Comunicaciones por la Facultad de Ingeniería Mecánica y Eléctrica (FIME/UANL). Tiene más de 15 años de experiencia gerencial en áreas de manufactura, logística, planeación y sistemas, en 8 diferentes ramas industriales. Autor de varios artículos en revistas indexadas. Sus líneas de investigación están enfocadas en el desarrollo de modelos para optimización de procesos basados en técnicas de programación mixta entera y meta-heurísticas para problemas relacionados a manufactura, logística, localización de instalaciones, programación de la producción, diseño territorial, ruteo de vehículos de transporte y distribución.

Correo electrónico: fabian.lopez@e-arca.com.mx

Dra. Teresa Pérez

La Dra. Teresa Pérez nació en Real del Monte, Estado de Hidalgo, México. Es egresada tanto de licenciatura como maestría del Instituto Politécnico Nacional, en la Unidad Profesional Interdisciplinaria de Ciencias Sociales y Administrativas (IPN/UPIICSA, México). Cursó sus estudios de doctorado en la Facultad de Contaduría y Administración de la Universidad Autónoma de Nuevo León (FACPyA/UANL, México), con proyecto de Investigación "Evaluación de los factores que determinan el grado de disponibilidad de los sistemas computacionales relacionados a la seguridad de información". Ha impartido clases tanto de licenciatura como de nivel maestría en Tecnologías de Información en diferentes Universidades, tales como la Universidad Autónoma de México (UNAM, México) y en el IPN. Actualmente cumple su labor docente en la Universidad Regiomontana (UR, México). Es conferencista en temas relacionados con Tecnologías de la Información.

Correo electrónico: teresa.perez@bancoahorrofamsa.com

Dr. Manuel Prieto

El Dr. Manuel Prieto es Ingeniero Químico por la Universidad Autónoma de Coahuila (UADEC, México), es Maestro en Ciencias con especialidad en Ciencias Computacionales por la Universidad de Nebraska en Lincoln (UNL, Estados Unidos). Es Doctor en Filosofía con especialidad en Administración por la Facultad de Contaduría y Administración de la Universidad Autónoma de Nuevo León (FACPyA/UANL, México), siendo sus líneas de investigación: Productividad en la producción de software, Métodos cuantitativos para la formación de equipos de trabajo de desarrollo e instalación de software Aplicaciones de tecnologías de información en la educación.

Correo electrónico: mprieto@interwebinternational.com

Dr. Felipe Ramírez

El Dr. Felipe Ramírez el Licenciado en Informática Administrativa por la Facultad de Contaduría Pública y Administración, en la Universidad Autónoma de Nuevo León (FACPyA/UANL, México), especializándose en bases de datos y lenguajes de programación; también es Licenciado en Derecho por la Facultad de Derecho y Criminología (FACDyC/UANL), especializándose en contratos de tercerización de servicios de TI y delitos informáticos. Cuenta con una Maestría en Informática y además obtuvo el grado de Doctor en Filosofía con especialidad en Administración (FACPyA/UANL), siendo distinguido con mención honorífica *Magna Cum Laude*. Es instructor certificado de ITIL, es director de contenidos de Aprenda.mx y catedrático a nivel licenciatura y postgrado en la Facultad de Contaduría y Administración, así como catedrático a nivel postgrado en la Facultad de Ingeniería Mecánica y Eléctrica (FIME/UANL), donde forma parte del núcleo académico de la Maestría en Ingeniería de la Información. Es director de contenidos de Aprenda.mx, en donde desarrolla conocimiento para su consumo en instituciones educativas. Conferencista internacional, tiene publicados 7 libros relacionados con bases de datos y programación de aplicaciones en plataformas de Microsoft, y además desarrolló el modelo de Alto Rendimiento Positivo (ARP), utilizado en las organizaciones para fomentar la productividad.

Correo electrónico: felipe.ramirez@aprenda.mx

Dr. Arturo Reyes

El Dr. Arturo Reyes es egresado del Doctorado en Administración de la UANL, con la especialidad en Intangibles de las Organizaciones; obtuvo su maestría en Administración con acentuación en Recursos Humanos en FACPYA, UANL, y es Lic. En Ciencias Computacionales por la Facultad de Ciencias Físico-Matemáticas de la misma Universidad. También es egresado de *Leadership College*. Ha trabajado como Consultor desde hace 15 años en el área de desarrollo organizacional en diferentes empresas en México, USA, Guatemala, Chile, España y Brasil, desarrollando modelos de evaluación de ambiente laboral. Los últimos 6 años se ha desempeñado como Director General en algunas empresas. Su experiencia impartiendo conferencias es de 20 años en temas diversos, enfocados a la rentabilidad de empresas, productividad y ambiente laboral.

Correo electrónico: arturo.reyes@microsyredes.com.mx

Dr. Francisco Treviño

El Dr. Francisco Treviño es profesor e investigador dentro del programa de posgrado de la Facultad de Ingeniería Mecánica y Eléctrica en la Universidad Autónoma de Nuevo León (UANL, México). Ha participado en los congresos 2008 y 2010 de ACACIA (Academia de Ciencias en Administración). Tiene un doctorado en filosofía con especialidad en administración por UANL. Tiene una maestría en administración con especialidad en finanzas, así como varios diplomados en ventas, temas gerenciales, y estudios de negocios basados en internet por el Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM, México). Es Licenciado en Economía por el ITESM. Tiene más de 15 años de experiencia en el área de Comercio Internacional, 10 años en el área de logística y abastecimientos en diferentes compañías. Es consultor de empresas dentro de Soluciones Integrales Treviño S.A. de C.V. Sus líneas de investigación están enfocadas en los factores que generan productos obsoletos en los inventarios de las empresas, y en las prácticas administrativas para mejorar el desempeño de las áreas comerciales.

Correo electrónico: francisco.trevino@sitconsultores.mx

Prólogo

Sentencias clásicas como “*lo que no se puede medir, no se puede controlar*” se han arraigado tanto en la memoria de todos los profesionistas de nuestra época, que se desea medirlo todo.

La pregunta fundamental es ¿para qué medimos las cosas? La respuesta se antoja simple: para disponer de datos que nos lleven a la generación de información, que a su vez nos permita tomar decisiones acertadas.

Existe en la actualidad una obsesiva necesidad de recopilación de datos. Todas las organizaciones lo hacen de alguna manera, y sin embargo, en ocasiones eso no tiene sentido, pues no saben qué hacer con los datos que recopilaron. Peor aún, hacen cosas sin saber, y toman decisiones sobre interpretaciones erróneas respecto a las cosas que los datos tienen que decir.

Los datos esconden verdades ocultas, soluciones a problemas, e incluso un poco de futuro. Claro que no basta con observarlos simplemente, sino que es necesario darles un tratamiento estadístico para que tengan sentido de existir.

Aquí es donde inicia el problema. Existen muchas herramientas estadísticas que nos permiten darle un tratamiento a los datos, y derivar conclusiones con cierto grado de certeza. Pero importa qué certeza es la que queremos obtener, importa los datos de los que disponemos para procesar, importa escoger la herramienta estadística correcta que nos permita obtener el resultado correcto.

Este libro es fundamental para aquellos profesionistas e investigadores que tienen la necesidad de respuestas y quieren saber cómo el análisis multivariante puede ayudarles a obtener dichas respuestas.

El libro explica los fundamentos teóricos de los métodos de análisis multivariante; posteriormente, explica la mecánica que nos lleva a decidir qué técnica es la adecuada para enfrentar el problema estadístico que deseamos resolver, y finalmente nos muestra casos prácticos, de investigaciones reales vistas desde diferentes ópticas, que pueden servir de referencia para otros nuevos estudios.

Esperamos que el libro sea de su agrado, y que el lector encuentre las respuestas que busca.

Los Autores.

Capítulo 1:

Introducción al análisis multivariable

Contenido

Importancia del análisis multivariante	2
Etapas para la aplicación del análisis multivariable	6
Instrumento de medición	8
Tipos de escalas de medida	8
Construcción del instrumento	12
Escala de respuesta	14
Pruebas de validez y fiabilidad	15
Prueba de consistencia interna (alfa Cronbach)	18
Selección de los ítems	20
Tamaño de la Muestra y Ponderación.....	22
Inferencia estadística en el análisis multivariable	24
Hipótesis	27
Análisis inicial y pruebas en los datos.....	28
Supuesto de normalidad.....	30
Supuesto de linealidad y errores no correlacionados	31
Supuesto de homocedasticidad	33
Índice del capítulo	34

El mundo como lo conocemos es un sistema altamente complejo. Si tuviéramos que explicar las bases sobre las cuales funciona la sociedad hoy en día, tendríamos que decir que los recursos son cada vez más limitados, y el tiempo para hacer las cosas, insuficiente. Vivimos escenarios saturados de información, pero al mismo tiempo llenos de incertidumbre; no es de extrañar que las personas y las organizaciones estén con una profunda necesidad de certezas, ávidas de una luz al final del camino que les permita el uso más eficiente de los recursos.

Este libro trata a fondo el uso de herramientas matemáticas que nos permitan ofrecer cierto grado de certezas en escenarios complejos. Los modelos de análisis multivariante ofrecen muchas verdades ocultas en los datos; cada fenómeno está determinado por un conjunto de elementos que lo determinan en mayor o menor grado, y los modelos de análisis multivariante permiten estimar resultados que de otra manera no es posible conocer.

Importancia del análisis multivariante

El *análisis multivariante* (de igual manera conocido como *análisis multivariable*), tiene su origen a principios del siglo XIX, con las teorías desarrolladas para resolver problemas del área de psicología. Estudios como los de Karl Pearson y Charles Spearman fueron claves para el desarrollo de los fundamentos del análisis multivariante. Spearman, por ejemplo, al tratar de evaluar la inteligencia de los seres humanos seleccionó un conjunto de características que se podían medir (variables), y se dio cuenta de lo siguiente: **a)** era imposible determinar la inteligencia observando una sola característica, pues se requería observar varias características al mismo tiempo; **b)** no todas las características que podían ser medidas eran relevantes, por lo cual se necesitaba saber qué características sí eran relevantes para medir la inteligencia, y cuáles no; **c)** algunas características indicaban los mismos fenómenos que otras, por lo que era necesario seleccionar el número de características mínimas suficientes y necesarias para explicar los fenómenos de la mejor manera; **d)** las características no eran independientes entre sí, pues la variación en una característica afectaba a las otras; y **e)** no existían modelos matemáticos que permitieran el análisis con un escenario tan complejo como el que estaban enfrentando.

Al enfrentar ese tipo de situaciones complejas, los científicos se dieron a la tarea de crear modelos que permitieran el análisis conjunto de variables. Fue así como nació el cálculo del coeficiente de correlación, y posteriormente, una serie de modelos relacionados con el análisis de escenarios que involucraban múltiples variables. No obstante que el marco teórico de dichos modelos tenía bastante sustento, las limitantes de procesamiento de cálculo de la época hacían que no tuvieran un uso extendido en las organizaciones.

Afortunadamente, el progreso tecnológico ha incrementado la capacidad para procesar grandes volúmenes de datos. Esto ha permitido a los investigadores abordar análisis más específicos, considerando un mayor número de variables, demostrando sus estimaciones empíricas con marcos teóricos bien fundamentados, y sin tener que esperar mucho tiempo de procesamiento de datos para ello.

Las técnicas del análisis multivariable en la actualidad están siendo aplicadas con mucho éxito en la industria, la administración de las empresas así como en los centros de investigación. También hay que decirlo, muchos de los trabajos evolucionan a prueba y error, dado el desconocimiento de las técnicas de análisis multivariante. Es común que se apliquen pruebas para situaciones inadecuadas, que no se verifiquen supuestos requeridos antes de la realización de pruebas, o que los resultados no sean adecuadamente interpretados. No obstante que existen libros que explican los fundamentos de las técnicas de análisis multivariante, el investigador con cierto conocimiento de estadística carece de herramientas para *decidir* qué técnica es la más adecuada para el trabajo de investigación que realiza, la preparación que requiere para aplicar una determinada prueba, y cómo interpretar los resultados.

Este libro se ha escrito para cubrir este vacío. Ayuda a determinar qué prueba es la más adecuada para una investigación, ayuda a realizar la revisión de supuestos requeridos para su aplicación, y la forma en que se debe realizar la interpretación de los resultados. Para reforzar la teoría, expone casos de investigaciones reales, para que el lector los tome como referencia.

Este libro está orientado a atender aplicaciones de interés común para analistas de negocios, académicos o investigadores dedicados a la comprensión y solución de fenómenos complejos. Los modelos multivariados permiten plantear preguntas específicas de considerable complejidad posibilitando desarrollar investigaciones teóricamente significativas (i.e. validez estadística). Se pueden integrar las correlaciones entre múltiples variables y a la vez se pueden estudiar estadísticamente los efectos aislados de esas variables. Por ejemplo, en un análisis de mercado no se pueden considerar consumidores homogéneos y caracterizados por un número reducido de variables demográficas. En su lugar, se deben desarrollar estrategias para atraer de manera diferenciada a los diferentes segmentos de clientes con diversas características demográficas en un entorno con múltiples restricciones (legales, económicas, competitivas, tecnológicas, etc.). Como concluyeron Pearson y Spearman en sus estudios, las cosas a veces no son tan sencillas. A través de las técnicas de análisis multivariable se pueden examinar adecuadamente las relaciones múltiples para llegar a una comprensión integral y realista en el proceso de toma de decisiones.

La teoría involucrada en las técnicas multivariadas fue desarrollada mucho antes de la aparición de las computadoras. Sin em-

bargo, hasta el momento cuando tuvo disponible el poder computacional para realizar cálculos complejos, fue que llegó a desarrollarse ampliamente este tipo de técnicas. Los analistas de información han contribuido al aumento en el uso y la aceptación de la estadística multivariable en los negocios y en la administración. En la comunidad académica, disciplinas de todos los campos han adoptado las técnicas multivariantes para desarrollar sus investigaciones empíricas. Incluso para personas con sólida preparación cuantitativa, la disponibilidad de software de análisis multivariable ha facilitado el desarrollo de la disciplina.

El objetivo del análisis multivariable es medir, explicar y predecir el grado de relación teórica entre las variables objeto de estudio. El investigador especifica las variables teóricamente, mientras que los coeficientes de las variables son determinadas a través de la técnica multivariable. El resultado de los coeficientes de las variables es un conjunto de valores que representa la combinación óptima que mejor se adapta al objeto del análisis. Obviamente, las técnicas de análisis multivariable son herramientas: cada una sirve para una determinada cosa, la forma para determinar valores óptimos se obtienen de manera distinta en cada prueba, y el resultado obtenido se interpreta de diferente forma. En regresión múltiple, por ejemplo, el valor óptimo se determina de tal forma que se obtenga la mejor correlación con la variable que se está buscando estimar; en el análisis discriminante, por otro lado, el valor teórico se obtiene de tal manera que se diferencien de forma máxima los grupos de observación, mientras que en el análisis factorial, los valores teóricos se forman para obtener una representación óptima de las estructuras subyacentes (i.e. correlación).

Dependiendo del tipo de investigación, existen diferentes tipos de herramientas estadísticas para analizar la información. En cualquier caso, es fundamental seleccionar las variables adecuadas para el análisis de los datos. El análisis será univariable o multivariable, dependiendo de la cantidad de variables que se vayan a estudiar, una o varias, respectivamente. El objeto de estudio de este libro es precisamente aquellas técnicas que permiten trabajar con dos o más variables.

La planificación de una investigación significa que en cada una de las etapas del proceso se deben tomar las decisiones apropiadas de forma holística. La definición del problema, el diseño de instrumento de medición y el diseño experimental de pruebas se afectan unos a otros. Por ejemplo, el diseño de un instrumento de medición está condicionado por las etapas previas —definición de problema e hipótesis—, y a la vez esto influye en el análisis de datos de las etapas posteriores —diseño experimental, aplicación de pruebas e interpretación de resultados—. Lo paradójico es que en la definición del problema de investigación es necesario considerar el método de análisis de datos que será utilizado, ya que el tipo de

preguntas del instrumento será desarrollado de acuerdo con la técnica de datos que facilite el análisis de los resultados del estudio.

Así entonces, podemos definir el *análisis multivariable* como el conjunto de técnicas estadísticas que de forma simultánea miden, explican y predicen las relaciones existentes entre los elementos que conforman un fenómeno.

El análisis multivariable incluye una gran diversidad de técnicas, por tanto resulta crucial seleccionar la técnica adecuada que se adapte a las pretensiones del estudio. El análisis multivariable es un cuerpo de conocimientos de naturaleza plural e interdisciplinaria. Todo ello se refleja en la diversidad de materias científicas (psicología, sociología, estadística, economía, derecho, cibernética, etc...) en las que es aplicada, así como en la variedad de las técnicas de investigación de operaciones involucradas (programación, simulación, teoría de juegos, etc...).

El objetivo de cualquier tipo de investigación es el soporte y apoyo al proceso de toma de decisiones. Existe una serie de razones o circunstancias que dificultan la actividad anterior:

1. El elevado grado de incertidumbre en el entorno.
2. La gran cantidad de variables y el volumen de información a considerar.
3. La dinamicidad, continuidad y rapidez con que cambia la información (i.e. los datos).

Los decisores de cada área de conocimiento requieren información específica que a su vez permita activar el proceso de toma de decisión. El análisis multivariable para poder aplicarse requiere datos, ya que éstos constituyen el objeto de estudio de la investigación. Así entonces, es pertinente hablar del rol que tienen las variables así como de las escalas de medida que estas tendrán. De igual modo es necesario un análisis previo de los datos para detectar puntos ausentes y extremos.

Dependiendo de las características de la investigación es posible encontrar dos clases principales, la *investigación no experimental* y la *investigación experimental*. La investigación no experimental es la que se realiza sin manipular deliberadamente las variables independientes (variables que se analizan en función a la variable dependiente). Además aquí no se construye ninguna situación sino que se observan situaciones ya existentes y no provocadas por el investigador.

La investigación también tiene clasificaciones atendiendo el tratamiento de los fenómenos en el tiempo, y atendiendo el tipo de relaciones entre las variables involucradas en la investigación. En cuanto al tratamiento de los fenómenos en el tiempo, existe la *investigación transeccional*, que es aquella cuya verificación y medición

se realiza en un solo punto en el tiempo. Por otro lado, se tiene la *investigación longitudinal*, que es aquella investigación del tipo evolutivo que tiene el objetivo de analizar cambios a través del tiempo del tipo tendencia. Respecto a la relación entre las variables, la investigación puede ser *investigación descriptiva*, si acaso se mide y describe una variable específica " χ_n ", o también puede ser *investigación correlacional*, llamada también *investigación causal*, si se mide o describe una relación del tipo " $\chi_1 - \chi_2$ ".

Etapas para la aplicación del análisis multivariable

A continuación se da una la explicación más detallada de cada etapa:

1. **Planteamiento del problema de investigación, objetivos y técnica multivariable apropiada:** definir de forma clara y en términos conceptuales una situación que se pretende resolver, así como de la definición de los objetivos a alcanzar. La descripción del problema debe proporcionar indicios suficientes para determinar qué datos son los que participan en la comprensión del fenómeno.
 - a. Se definen los conceptos e identificar las relaciones a investigar mediante un modelo conceptual.
 - b. Se debe especificar los conceptos dependientes e independientes.
 - c. Se identifican las ideas o temas de interés.
 - d. Después de identificar si el problema requiere de un método de dependencia o interdependencia, entonces se procede a elegir la técnica específica basada en las características de medición.

En esta fase es fundamental el conocimiento que el investigador tenga sobre el tema ya sea a nivel de teoría o como experiencia práctica. Aquí se deben definir los conceptos que serán representados mediante variables así como la suposición de existencia de relaciones de dependencia o interdependencia entre ellas. Para la medición de las variables se define su escala de medida, sea métrico o no métrico con la finalidad de poder hacer la elección de la técnica de análisis multivariable más adecuada.

2. **Análisis de supuestos:** tiene que ver con contar con los requisitos mínimos para que los resultados de la prueba elegida sea confiable. Los supuestos tienen que ver con la muestra de datos; tipo, disponibilidad y rango; y con los principios estadísticos y conceptuales aplicables al modelo seleccionado.

- a. **Supuestos de muestra:** tiene que ver con la disponibilidad de la cantidad de datos mínima suficiente y necesaria para poder generalizar los resultados a la población total. Esto es muy importante porque algunas pruebas estadísticas requieren más datos que otras a efecto de que los resultados obtenidos sean confiables.
- b. **Tipo, disponibilidad y rango:** tiene que ver con garantizar que los datos son del tipo adecuado para la prueba seleccionada, pues habrá pruebas que trabajen bien con datos de razón, pero no con datos de intervalo; teniendo los datos del tipo correcto, se debe proceder con la identificación de datos ausentes y aislados, así como datos aberrantes o fuera de rango, cuya inclusión en el procesamiento puede afectar la confiabilidad de los resultados.
- c. **Principios estadísticos y conceptuales:** consiste en evaluar que los datos cumplan con los supuestos tanto estadísticos como conceptuales. Aquí nos referimos a las pruebas de normalidad, linealidad, independencia de los términos de error e igualdad de las varianzas (homocedasticidad). Antes de cualquier estimación del modelo, se debe asegurar que se cumplan los supuestos estadísticos y los conceptuales.

Antes de aceptar el modelo estimado como válido es recomendable preguntarse hasta qué punto es aplicable a la población total. En este sentido es deseable disponer de una segunda muestra de validación, la cual permite repetir el análisis y obtener resultados que apoyen el sentido de generalización hacia la población.

3. **Estimación del modelo multivariable:** satisfechos los supuestos, se procede a la estimación del modelo multivariable. Aquí se evalúa el ajuste del modelo para averiguar si se consiguen los niveles aceptables de *significancia*, entendiendo esto como la amplia probabilidad de que los resultados obtenidos no sean producto del azar.

Se validan las relaciones propuestas y con ello la validación de la significación práctica. El modelo se volverá a especificar nuevamente en un proceso iterativo con la finalidad de mejorar los niveles de ajuste y/o explicación.

Independientemente del nivel de ajuste alcanzado, se debe identificar si los resultados están siendo afectados por casos u observaciones atípicas que puedan estar provocando inestabilidad en los estimadores del modelo. Con ello se asegura la robustez del modelo. Casos influyentes o atípicos deberían ser analizados por separado y en su caso omitidos de la muestra según corresponda.

4. **Interpretación de los resultados:** aquí se identifica y se analiza la naturaleza de las relaciones multivariadas del modelo resultante. La interpretación de los efectos individuales de las variables se realiza examinando los *coeficientes del modelo* (i.e. las betas de la regresión o las cargas factoriales del análisis factorial).

Esta fase de interpretación puede conducir a re-especificaciones adicionales en las variables y/o en la formulación del modelo.

5. **Validación del modelo multivariable:** antes de aceptar los resultados, se debe someterlos a un diagnóstico que asegure el grado de generalidad de los resultados.

La validación del modelo es hacia la demostración de la generalidad de los resultados al conjunto de la población. El objetivo es comprobar la evidencia empírica de las relaciones multivariadas de los datos muestrales y que esto permita generalizar teóricamente hacia la población.

Instrumento de medición

Tipos de escalas de medida

El análisis de los datos implica la separación, identificación y medida de la variación en un conjunto de variables. Para ello existen dos tipos de datos: los no métricos (*datos cualitativos*) y los métricos (*datos cuantitativos*). Los datos no métricos son atributos, características o propiedades categóricas que identifican o describen a un objeto. Describen diferencias en tipo o clase indicando la presencia o ausencia de una característica o propiedad. Las propiedades son discretas porque tienen una característica peculiar que excluye a todas las demás características. Por ejemplo, si uno es hombre no puede ser mujer. Por el contrario, las medidas de datos métricos son de tal forma que los sujetos pueden ser identificados por diferencias de grado, cantidad o magnitud.

Los *datos* en esencia son los valores que toman las variables, las cuales a su vez son magnitudes que representan distintos conceptos o atributos de los objetos de estudio. La precisión de tal medición depende directamente de la escala de medida seleccionada.

Los tipos de escalas de medida más común en estadística multivariable distingue cuatro básicas: la Nominal, la Ordinal, la de Intervalo y la de Razón. Cada una de ellas tiene mayor precisión que la anterior. Las dos primeras son escalas no métricas o cualitativas, puesto que buscan identificar en su medición una determinada cualidad o propiedad. Las dos últimas son escalas métricas o cuantitativas, capaces de reflejar diferencias de grado o cantidad.

Escala nominal. Las medidas no métricas pueden tener escala nominal o también llamada categórica. Una medida con una escala nominal, asigna números para etiquetar o identificar un objeto. Por tanto, el número asignado a un objeto no tiene un significado cuantitativo sino sólo para indicar la presencia o ausencia de un atributo o característica.

Los ejemplos de datos con escala nominal incluyen el sexo, la religión o el partido político de una persona. Para trabajar con estos datos, el analista puede asignar números a cada categoría, por ejemplo, 2 para mujeres y 1 para hombres. Más específico, la escala nominal permite categorizar, ya que su escala es mutuamente excluyente y exhaustiva. O se es una cosa, o se es otra.

Los números que se asignan a cada categoría no encierran ningún significado ni relación de orden. Son simples etiquetas que se asocian a cada posible medición. Un ejemplo de medida en escala nominal es por ejemplo la variable “sexo”, a la que podría asignarse un 1 si el encuestado fuese mujer y un 2 si fuese hombre. Otro ejemplo puede ser la variable “estado civil”, un 1 si el sujeto fuese soltero, un 2 si fuera casado, un 3 si es viudo, y un 4 al ser separado o divorciado. Por presentar únicamente dos categorías posibles, la variable “sexo” se le nombra como *variable dicotómica*. Por su parte, la variable “estado civil”, con más de dos categorías, se le menciona como *politómica*.

Escala ordinal. Adicionalmente, las medidas no métricas pueden también ser de escala ordinal. La escala ordinal tiene un nivel superior de precisión en la medida que con respecto a las variables de escala nominal. Las variables pueden ser ordenadas o clasificadas con escalas ordinales en relación a la cantidad del atributo poseído. Cada caso puede ser comparado con otro en términos de una relación de “mayor que” o “menor que”. Por ejemplo, los diferentes niveles de satisfacción del consumidor. Los números utilizados en escalas ordinales no son cuantitativos, solo indican posiciones relativas en series ordenadas. No se conoce la diferencia exacta entre los puntos de la escala. La escala Ordinal también permite representar categorías, pero en este caso los números asignados, sí guardan una relación de orden. Los números mayores estarán indicando una mayor magnitud del atributo que se esté midiendo. No obstante, este tipo de escala, no llega a indicar cuánto mayor es tal magnitud; por esa razón carecen de carácter cuantitativo. Un ejemplo de escala Ordinal sería la variable “edad” midiéndose de la siguiente manera: 1 si es inferior a 20 años, 2 si está comprendida entre 21 y 35, 3 si lo está entre 36 y 60, y 4 si es mayor que 60.

Escalas de intervalo y de razón. Las escalas métricas pueden ser de intervalo y de razón. Estas métricas proporcionan el nivel más alto de medida de precisión, permitiendo realizar casi todas las operaciones matemáticas. Estas dos escalas tienen unidades constantes de medida. Por tanto cualquier diferencia entre dos puntos

adyacentes en cualquier parte de la escala es igual. La única diferencia entre la escala de intervalo y la de razón es que la *escala de intervalo* tiene un cero arbitrario y la *escala de razón* su cero es absoluto. Un ejemplo de una escala de intervalo es la temperatura. La temperatura tiene un punto de cero arbitrario, pero el cero no indica una cantidad cero o ausencia de temperatura, dado que podemos registrar temperaturas por debajo del punto cero de esa escala. Más aún, no es posible decir que un valor cualquiera situado en un intervalo de la escala es un múltiplo de cualquier otro punto de la escala (i.e. 40°C, no es dos veces más caluroso que 20°C). En una escala de Intervalo la diferencia entre dos niveles consecutivos es constante a lo largo de toda la escala. Sin embargo, esta escala carece de un cero absoluto, lo que impide afirmar que un valor determinado sea el doble, el triple o, en general, un múltiplo de otro valor de la misma escala.

Finalmente, la escala métrica de razón es la forma superior de medida en cuanto a precisión, dado que poseen las ventajas de todas las escalas anteriores y además un punto de cero absoluto. Las medidas de escala de razón permiten el uso de todas las operaciones matemáticas. Además, las medidas de este tipo de escala permiten ser expresadas en términos de múltiplos cuando se relaciona un punto con otro de la escala (i.e. 100 kilos es dos veces más pesado que 50 kilos). La escala de razón supera el inconveniente mencionado para la escala de Intervalo, pues presenta un cero absoluto conocido que hace posible realizar con estas medidas todas las operaciones matemáticas. Variables como la altura, el peso, ingresos, gastos, inversiones, etc. son de este tipo.

En general, es importante identificar la escala de medida de cada variable utilizada para asegurar que no se estén utilizando datos no métricos como si fueran métricos. Más aún, la escala de medida es crucial para determinar cual técnica multivariable es la más conveniente para los datos, considerando que algunas sean variables dependientes y otras independientes.

El que una variable sea del tipo métrico o no métrico, tiene un interés especial en el análisis multivariable por cuanto se permita o no la aplicación de diferentes técnicas. Así, por ejemplo, la regresión lineal múltiple o el análisis factorial requieren variables métricas. Aún para el caso de la aplicación de modelos de regresión lineal múltiple, una variable no métrica puede ser convertida en métrica a través del uso de variables ficticias binarias. Para ello se requiere contar con un número de variables binarias igual al número de categorías de la variable no métrica menos uno. Suponga la variable no métrica “medios de transporte” de tres categorías: 1 = autobús, 2 = tren y 3 = avión. Para convertir dicha variable en métrica, esto podría efectuarse por medio de dos variables ficticias, **F1** y **F2**.

Así entonces:

- Valor de la variable no métrica igual a 1 (Autobús), entonces:
F1 = 0, F2 = 0
- Valor de la variable no métrica igual a 2 (Tren), entonces:
F1 = 0, F2 = 1
- Valor de la variable no métrica igual a 3 (Avión), entonces:
F1 = 1, F2 = 0

El concepto de escala se utiliza para hacer referencia a un determinado instrumento de medición. Los elementos básicos que forman un instrumento de medición se denominan ítems. Un *ítem* podemos definirlo como un hecho u opinión que se pregunta a una persona con el objeto de provocar una respuesta relacionada con el atributo evaluado. Por tanto un instrumento de medición está compuesto por una serie de ítems que miden uno o varios atributos de interés, y que a la vez proporcionan información en relación con el atributo que se mide.

Hasta hace algún tiempo, sólo era posible realizar mediciones físicas en los objetos de estudio en las ciencias naturales. Por tanto en nuestro caso, el objetivo fundamental es cuantificar las sensaciones en las ciencias sociales a partir de un referente físico conocido. Las escalas de medición aparecieron en 1928 con Thurstone. En 1932 Rensis Likert propuso un método de escalas con el objetivo de simplificar el procedimiento de Thurstone. Más adelante, Louis Guttman (1944) presenta un método basado en un escalograma de ítems.

Las escalas de Thurstone, Likert y Guttman se les conoce como escalas diferenciales, sumativas y acumulativas respectivamente. La *escala de Thurstone (escala diferencial)* busca medir las diferencias que se presentan entre los ítems que componen la escala. La *escala de Likert (escala sumativa)* se obtiene a partir de la suma (o promedio) de las respuestas que integran los ítems que componen la escala. Por su parte, la *escala de Guttman (escala acumulativa)* es, como su nombre lo indica, acumulativa, ya que busca medir la velocidad (marginalidad) así como la consistencia con la cual un determinado ítem difiere con respecto a otro ítem dentro de la misma escala.

En las escalas de medición de Thurstone y Guttman se procura que los ítems sean diferentes entre ellos ya que cada ítem representa diferentes niveles de intensidad en el atributo. En el caso de Likert los ítems son como replicas (i.e. intentos) que sirven para medir de diversas formas el mismo atributo.

Construcción del instrumento

La definición de los atributos (i.e. variables o constructos) así como su escala es lo primero y lo más importante que se debe tomar en cuenta antes de desarrollar un instrumento de medición.

La cantidad de ítems (i.e. replicas o intentos) que se utilicen para medir una variable, debe estar racionalizado para cubrir con la menor cantidad de intentos el dominio completo de la variable que se busca medir. No todos los instrumentos de medición se construyen con el mismo grado de conocimiento sobre lo que se requiere medir. En algunos casos se puede partir de una teoría existente, sin embargo en otros casos es posible que el investigador parta de una idea propia acerca de lo que es necesario medir. El primer caso consiste en acudir a escalas ya construidas que persigan mediciones u objetivos similares mientras que en el segundo caso se recurre principalmente al conocimiento y la creatividad del investigador.

Para el caso de una escala de medición ya construida pueden darse dos posibilidades: aprovechar todos los ítems de la escala original u ocupar sólo algunos de ellos. En otro caso puede ocurrir que una escala ya existente se adecue a los objetivos planteados pero haya sido diseñada para un contexto cultural distinto; aquí se intentaría adaptar dicha escala. Se pueden generar ítems a partir de la opinión de personas con un conocimiento específico respecto al objeto a medir. En cualquier caso, se trata de un punto de partida ya que es responsabilidad del investigador documentarse sobre la variable que busca medir, buscando información en todas las fuentes que tenga a su alcance con el objetivo de completar la definición de su objeto de medición.

Algunas sugerencias al elaborar ítems son las siguientes:

1. La estructura gramatical y el vocabulario de los ítems que sean enunciados, debe adecuarse a las características propias de las personas que responderán a la medición. En redacción se le llama selección y combinación: seleccionar las palabras menos ambiguas y mejor entendidas (no es lo mismo encomiar que celebrar), y acomodarlas en un orden que facilite su comprensión (no es lo mismo “el cartero trajo una carta para mí”, que “para mí una carta el cartero trajo”).
2. La redacción debe facilitar al máximo la comprensión de lo que realmente se busca medir.
3. Los enunciados de los ítems deben ser cortos, claros y usando un lenguaje simple y directo.
4. Cada ítem debe preguntar (i.e. medir) una sola idea facilitando con ello la obtención de las respuestas.

5. El significado de los ítems no debe dejar dudas o ambigüedades.
6. En la medición de actitudes, una misma idea puede expresarse de forma positiva o negativa. Por el motivo anterior, la recomendación es que en el instrumento de medición se tenga un balance comparable, entre elementos redactados en sentido favorable y desfavorable respecto al objeto que se esté midiendo. De hecho, lo anterior permite identificar errores voluntarios o involuntarios, en las respuestas obtenidas de los instrumentos de medición.
7. Deben evitarse expresiones que contengan negaciones (más aún las “doble negaciones”), ya que éstas introducen confusión en su entendimiento.
8. Siempre que una medición busque evaluar diferentes facetas en una misma variable, conviene disponer de tantos ítems como sea necesario para describir la presencia de dichas diferencias.
9. No es conveniente redactar ítems en forma absolutista de tal manera que las respuestas siempre sean las mismas y con ello afectar la precisión (i.e. la variabilidad) en las mediciones del instrumento. Para ello se debe evitar el uso de redacciones que incluyan adverbios como “siempre”, “todos” o “nadie”.

Un instrumento con una apariencia adecuada puede motivar a las personas para contestar los ítems y facilitar que lo hagan de manera sincera. Las instrucciones del instrumento redactadas en un lenguaje claro y sencillo es otro elemento importante a considerar. Cuando buscamos la *validez del instrumento* de medición específicamente nos referimos a que el instrumento mida los elementos para los cuales se le diseñó. La validez de contenido tiene que ver con que los ítems que están siendo medidos sean relevantes y pertinentes con respecto a las variables que se buscan medir. Esto significa que el ítem realmente este midiendo la información de la variable que se desea medir.

Por otro lado, la *validez de estructura interna* mide que las relaciones empíricas en los ítems del instrumento de medición, se mantenga en consonancia con el modelo teórico esperado. La validez de un instrumento de medición también puede fundamentarse en la comprobación de que las mediciones obtenidas por grupos de personas con características conocidas se ajustan a lo esperado. Puede resultar interesante identificar la *validez de convergencia*, esto es, que las mediciones obtenidas con el instrumento estén relacionadas con las obtenidas mediante otro tipo de escala.

En otros casos es posible que el objetivo sea probar que los ítems recojan la información necesaria sobre los aspectos que se desea evaluar, de manera que se cubra todo el dominio, es decir, todas las facetas de la variable que se pretende medir. Se trata de comprobar empíricamente las hipótesis sobre la dimensionalidad y la agrupación de los ítems. La estructura interna suele evaluarse utilizando técnicas de Análisis Factorial, las cuales permitan evaluar el ajuste del modelo a las respuestas observadas. Por ejemplo, supongamos que se pretende medir si una variable "A" está o no relacionada con la variable género. Para ello entonces basta con revisar si en la sub-muestra de hombres se identifica una estructura factorial distinta que en la sub-muestra de mujeres. Por tanto aquí estaríamos identificando si los mismos ítems generan respuestas distintas en ambos grupos.

Escalas de respuesta

Cada uno de los ítems, debe tener una escala de respuestas ordenada para facilitar su medición. A dicha escala se le suele denominar *escala de respuesta*.

En las escalas de tipo Likert se utilizan escalas de respuesta graduada con varias categorías ordenadas, que van de un extremo a otro y que pueden expresar calidad (muy bueno/muy malo), cantidad (mucho/nada), frecuencia (siempre/ nunca), consentimiento (acuerdo/desacuerdo), importancia (muy importante/sin importancia), probabilidad (probablemente sí/probablemente no), etc. Todas estas escalas permiten indicar el grado de acuerdo con un enunciado (i.e. con el ítem). Las categorías en un instrumento de medición conviene que se muestren ordenadas en forma simétrica, esto puede ser de mayor a menor o al revés. Para evitar confusiones, se recomienda mantener el orden de las categorías igual para todos los ítems del instrumento de medición, indistintamente de que éstos sean favorables o desfavorables. También es conveniente que la escala de categorías de respuesta combinen referentes verbales y numéricos indicando con ello el valor ordinal del grado de acuerdo, por ejemplo:

- (1) Muy en desacuerdo.
- (2) En desacuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) De acuerdo
- (5) Muy de acuerdo

Partiendo del ejemplo anterior, los valores numéricos de la escala de respuesta también pueden hacer énfasis visual en la bipolaridad del acuerdo haciendo que la escala vaya desde (-2) pasando por (0) y llegando hasta (+2). En general, la escala Likert de respuesta de cinco opciones es la más habitual.

Se recomienda utilizar escalas de respuestas que contengan al menos cinco categorías con el fin de optimizar la precisión (i.e. la variabilidad numérica) del instrumento de medición. Sin embargo, entre mayor sea el número de categorías más difícil es la medición y el entendimiento por parte de los encuestados de lo que realmente se está midiendo. De igual manera es recomendable que todos los ítems que sean utilizados para medir una variable tengan la misma escala de ordenamiento.

En relación a la simetría en la escala de medición, esto nos lleva a la conveniencia o no de que la cantidad de categorías de respuesta sean par o impar. Una escala con una cantidad de categorías impar, supone la capacidad para medir un punto de indiferencia (ni de acuerdo, ni en desacuerdo). En una escala con una cantidad de categorías par, se obtiene deliberadamente un diseño para que cada persona esté obligada a decidirse hacia un lado o hacia el otro. La escala impar facilita la respuesta de las personas indecisas y con ello se evita el que se dejen ítems sin responder. Aunque hay razones a favor y en contra para un diseño de escala par o impar, en general la decisión es mucho más dependiente del tipo de variable que se busca medir.

Pruebas de validez y fiabilidad

El error de medida es el grado en que los valores que están siendo medidos no son representativos de los valores verdaderos. El error de medida tiene múltiples fuentes, que van desde errores de registro en la entrada de datos hasta la imprecisión o incapacidad para hacer la medición.

Un ejemplo de la incapacidad anterior se refiere a la imposición de una escala de puntuación con muchas opciones. De hecho se sabe que los encuestados sólo pueden responder con precisión a una puntuación con relativamente pocas escalas. En ocasiones simplemente no es posible tener precisión, dependiendo del tipo de preguntas o también dependiendo del tipo de encuestados o de ambos.

En conclusión: en la medición de las variables siempre se debe considerar algún grado de error de medida. El valor medido incluye tanto la medida verdadera así como el ruido o error de medida. Por tanto, el cálculo de una correlación, está afectado por el error de medida, causando debilidad y pérdida de precisión. El objetivo es entonces reducir el error de medida.

Al valorar el grado de error de una medida se deben considerar dos conceptos: su validez y su fiabilidad.

La *validez* es el grado en que la medida representa lo que se supone está midiendo. Si se está midiendo "A" nunca se pregunta por "B". La validez requiere tener un conocimiento apropiado de lo que se va a medir para entonces realizar la medida tan precisa como sea posible evitando alguna falta de entendimiento o interpretación. Más importante que la precisión en las respuestas del encuestado, la clave está en que el investigador haga las preguntas correctas.

- a) La prueba de validez, en una investigación, tiene varios elementos que se desarrollan a continuación:
- b) La *validez de la escala* nos dice el nivel en que esta mide lo que se pretende medir y la validación consiste en llevar a cabo un estudio científico. En este proceso se trata de especificar hipótesis sobre el atributo que se desea medir, y recoger los datos necesarios para evaluar las hipótesis especificadas y así analizar e interpretar los datos obtenidos.
- c) La *validez de contenido* de una escala se refiere a la correspondencia entre el atributo que se pretende medir y el contenido de la muestra de ítems que compone la escala. Una escala presenta validez de contenido si los ítems que la componen son relevantes y además representativos del atributo definido.
- d) La *relevancia de un ítem* sucede cuando su contenido permite recoger información relacionada directamente con el atributo que se desea medir.
- e) La *representatividad de un ítem* sucede cuando se obtiene información suficiente de todos los aspectos que se desea evaluar, es decir, considerando todos los elementos considerados importantes del atributo a estudiar y así poder cubrir satisfactoriamente todo el dominio que se pretende medir. Esta prueba se puede llevar a cabo mediante un panel de expertos.

Cuando algunos de los elementos anteriores no se cumplen (relevancia y representatividad) se dice que el contenido representado por los ítems de la escala están sesgados

La validez sobre el proceso de respuesta nos dice que cuando las personas contestan el cuestionario o instrumento de investigación, se lleva a cabo un proceso lógico que no debe predisponer ninguna respuesta por haber contestado alguna pregunta anterior, etc.

Con respecto a la estructura interna, la validez se refiere a que los ítems deben estar organizados de la forma prevista en la defini-

ción del atributo que se pretende medir. Esta prueba se puede medir a través del análisis de factores.

Por otro lado, la *fiabilidad* es el grado en que la variable observada mide el valor verdadero y está libre de error. A menor error de medida mayor fiabilidad. Si la misma medida se realiza repetidas veces (i.e. con replicas), las medidas más fiables tendrán mayor consistencia que las medidas menos fiables. El uso de mediciones aditivas tipo escala ponderada, es útil para integrar varias mediciones o replicas. Aquí varias mediciones se unen para integrar una sola variable compuesta, llamada también *constructo*.

El objetivo es integrar varias mediciones, indicadores o replicas para medir una variable. Cada indicador mide diferentes facetas de la misma variable. El uso de indicadores múltiples permite tener una especificación más precisa (i.e. más fiable) de las variables a medir. De la manera anterior, la fiabilidad no se deja a una única respuesta sino a la respuesta típica de un conjunto de indicadores relacionados. El fundamento teórico descansa en que indicadores múltiples reflejan con mayor fiabilidad la medición verdadera para una variable que al hacerlo mediante una única respuesta. Por tanto se debe trabajar para aumentar la validez y la fiabilidad en las mediciones. Lo anterior asegurará una medición efectiva de las variables de interés. La presencia del error de medida es garantía segura de problemas más adelante al aplicar el análisis multivariable.

Para probar la validez en la escala y el contenido de los ítems en un instrumento de medición siempre será recomendable conocer la opinión de los demás, más específicamente de la población objetivo a ser medida. Se trata de personas de la población incumbente que no hayan intervenido en la redacción de los ítems y que sus intereses no sean tendenciosos. Aquí se busca identificar no solo problemas de redacción y entendimiento sino sobre todo relacionados con la falta de relevancia y de representatividad de los ítems.

La relevancia de los ítems tiene que ver con las facetas y dominios de las variables que se pretenden medir. Lo anterior aporta una visión objetiva e inequívoca acerca de la calidad de los ítems ya que permite diferenciar las opiniones unánimes de las que son discutibles. Las primeras serán consideradas como valoraciones fiables de los ítems ya que esto se fundamenta en las respuestas obtenidas al administrar la escala a una muestra de personas. Por tanto la fiabilidad de los ítems depende no sólo de las características de relevancia y precisión en la redacción, sino también de la representatividad de la muestra de personas que los responde. Sobra decir entonces por qué es importante que la muestra mediante la cual se va a probar la validez de contenido del instrumento sea lo más representativa posible de la población.

La fiabilidad de un instrumento de medición se refiere a la precisión de las puntuaciones que ésta ofrece. La fiabilidad consiste

en cuantificar en qué medida son replicables las mediciones obtenidas. Para medir lo anterior existen los siguientes enfoques generales:

1. **Test-Retest:** consiste en aplicar el instrumento de medición a la misma muestra de personas en dos ocasiones, separadas por un intervalo de tiempo prudencial y luego calcular la correlación de las puntuaciones obtenidas. Para que el coeficiente obtenido pueda ser interpretado como un coeficiente de fiabilidad conviene asegurar que las dos aplicaciones sean comparables. Es recomendable que el intervalo sea lo suficientemente largo para evitar que las respuestas de la segunda ocasión sean recordadas de la primera ocasión. Sin embargo, se debe evitar un intervalo de tiempo demasiado largo ya que la variable que se está midiendo podría ya haber cambiado.
2. **Formas equivalentes:** esta estrategia consiste en aplicar dos formas equivalentes del mismo instrumento de medición y correlacionar las puntuaciones obtenidas en ambas formas. Se utilizan dos instrumentos equivalentes: miden las mismas variables, con el mismo tipo de ítems, la misma apariencia, etc., pero con enunciados redactados de forma distinta. Con lo anterior se resuelve el inconveniente del diseño Test-Retest, pero se añade la necesidad de diseñar dos formas equivalentes del mismo instrumento.
3. **Mitades:** aquí se aplica un solo instrumento de medición a una muestra de participantes en una sola sesión. Para ello el instrumento se divide en dos partes que son consideradas equivalentes, por ejemplo ítems pares e ítems impares. Definidas las dos mitades se miden las dos puntuaciones y luego el coeficiente de correlación entre las dos es interpretado como coeficiente de fiabilidad.
4. **Consistencia interna (alfa de Cronbach):** aquí se prueba la hipótesis de que los ítems sean equivalentes. El coeficiente " α " constituye el coeficiente por mucho más comúnmente utilizado para cuantificar el nivel de fiabilidad de un instrumento de medición.

Prueba de consistencia interna (alfa Cronbach)

Una vez que los ítems ya están debidamente codificados puede realizarse la evaluación y prueba de consistencia interna mediante el indicador de alfa de Cronbach " α ".

La *consistencia interna* mide el grado de relación existente entre los ítems que componen la escala unidimensional de la variable que se está midiendo.

Una de las características de las escalas sumativas (i.e. Likert), es que la medición total de la variable (i.e. constructo) se calcula a partir de la suma (o promedio) de las respuestas proporcionadas de todos los ítems. El fundamento lógico de lo anterior para resumir la información en un único indicador (unidimensional), es que, de hecho, todos los ítems de la escala pretenden medir la misma variable. La prueba de consistencia interna, alfa de Cronbach “ α ”, para los ítems de una variable (i.e. un constructo) constituye una evidencia empírica a favor de la fiabilidad tanto de la escala como de la medición en forma integral.

La prueba de consistencia interna parte de la matriz de correlaciones entre los ítems. Así entonces, puede considerarse que un ítem es consistente con los demás de la escala si todas sus correlaciones con los demás ítems son positivas y moderadas.

El coeficiente alfa (α), propuesto por Lee J. Cronbach (1916-2001), permite resumir esta información de una manera sencilla. Este indicador es, sin duda, el más objetivo y utilizado para evaluar la consistencia interna de los ítems en un constructo o variable. La formula es:

$$\bar{R} = \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^k r_{ij}}{(k-1)^2 - k} \quad \text{donde : } k = \# \text{ de Variables,}$$

$$\alpha = \frac{k \bar{R}}{1 + \bar{R}(k-1)}$$

Coeficiente de Confiabilidad Alfa de Cronbach ($\alpha \geq 80\%$ es aceptable).

Donde k es el número de ítems de la escala de medición del constructo.

Bajo el supuesto de que todos los ítems negativos han sido recodificados en el mismo sentido de respuesta que los positivos, el coeficiente alfa debería tomar siempre valores positivos entre 0 y 1. Más aún, si todos los ítems de un instrumento son independientes entre sí (i.e. que estén midiendo variables diferentes), sus correlaciones y por tanto sus covarianzas, serían iguales a cero. En tal caso, como puede observarse en las fórmulas anteriores, el coeficiente alfa alcanzaría su valor mínimo, es decir cero, lo que indicaría una ausencia total de consistencia interna entre los ítems. Si, por el contra-

rio, todos los ítems fuesen completamente redundantes, sus correlaciones serían iguales a uno y por tanto sus covarianzas tomarían el valor máximo posible. En este caso el valor del coeficiente alfa (α) sería igual a uno (i.e. 100%), lo que indicaría la máxima consistencia posible entre los elementos de la escala. Un valor de alfa cercano al 100% no necesariamente es el objetivo para una prueba de consistencia interna ya que esto indicaría que todos los ítems son redundantes, es decir, miden la misma faceta del constructo y por tanto son replicas innecesarias. Así entonces lo óptimo es que cada ítem tenga su propia aportación a la escala numérica. Se recomiendan entonces valores de alfa (α) entre 75% y 90% (Nunnally 1978).

Selección de los ítems

En cualquier instrumento de medición, la elaboración de los ítems requiere que su contenido haga referencia únicamente a las facetas incluidas dentro de los límites del dominio de la variable que se busca medir. Lo anterior contribuye a asegurar la validez en el contenido de los enunciados que se aplican en la medición así como de la escala de las opciones que se esperan como respuestas.

Independientemente de su origen, es relevante considerar la conveniencia en cuanto a la cantidad idónea de ítems para la medición de una variable. No todos los ítems inicialmente considerados serán adecuados para la medición, ya que habrá que someterlos a criterios de calidad (i.e. alfa de Cronbach) en fases posteriores. Al final únicamente quedarán aquellos ítems que hayan demostrado ser válidos y fiables.

Además de la cantidad, lo que también debe quedar asegurado es la representatividad de los ítems en relación a las diferentes facetas que se buscan cubrir en la medición de la variable. En la medición de una variable, considerando todas sus facetas, no se trata de incluir demasiados ítems (i.e. intentos o replicas) sino de hacerlo de manera racional incluyendo todas sus facetas pero con la menor cantidad posible de ítems. Esto quiere decir que se requiere hacer un esfuerzo de análisis y síntesis por parte del investigador pero con la ventaja de ahorrar tiempo y esfuerzo en el momento de la recolección y medición de los datos. En resumen, se debe pretender contar con el número mínimo, suficiente y necesario de ítems que hayan demostrado ser válidos y fiables para medir a las variables.

El análisis de contribución de cada ítem en la construcción del indicador numérico de una variable es un recurso muy común. De hecho, mediante el análisis de alfa de Cronbach, es posible identificar cuales ítems aportan mayor consistencia interna y cual no. Lo anterior, es con la finalidad de evaluar cuales ítems en su caso sean eliminados por completo de la escala de medición. Para ello solo es necesario analizar el mejor alfa de Cronbach que pueda ser obtenido excluyendo combinatoriamente cada ítem. Si al calcular el alfa de

Cronbach se obtiene un mejor indicador que el actual, esto sería evidencia suficiente para prescindir definitivamente del ítem en cuestión. Lo anterior también aplica en el sentido inverso para aquellos casos en donde un ítem no debiera ser eliminado debido a que aporta relevantemente a la consistencia interna. Aquí lo que en realidad se está midiendo es la correlación de cada ítem en relación al resto.

Decidir cuáles ítems van a formar parte de la escala definitiva y cuáles habría que eliminar con base a la aportación de la consistencia interna (α de Cronbach), es una tarea metódica, en la que paso a paso, debe evaluarse la contribución de cada ítem a la consistencia de la escala. La selección de ítems con base en los criterios de consistencia interna es pertinente siempre que la variable que estamos midiendo sea unidimensional. Si estamos midiendo una variable multidimensional, no sería justificada la eliminación de ítems que tengan una baja correlación con respecto al resto de los ítems. Lo anterior es importante ya que posiblemente se podría tratar de un ítem que aporta información diferente (i.e. de otra faceta de la variable) y que además es relevante y complementa al resto de los ítems, es decir, una dimensión distinta a la que miden los otros ítems.

En general, cuando estamos midiendo dos dimensiones distintas (i.e. facetas diferentes) en una misma variable, puede ocurrir que las dos dimensiones estén relacionadas entre ellas o, por el contrario, que constituyan dimensiones independientes.

El *análisis factorial* (i.e. componentes principales) constituye una técnica adecuada para explorar la dimensionalidad en un variable. Mediante las saturaciones factoriales es posible seleccionar aquellos ítems que correlacionen con un subgrupo particular de ítems y que estén midiendo la misma faceta, es decir el mismo factor. Todo se resume en identificar el vector de cargas factoriales que muestre el subgrupo de ítems mutuamente correlacionados y por tanto dejando el resto como mutuamente excluyentes. Hecho lo anterior, los ítems mutuamente correlacionados estarán midiendo la misma faceta (i.e. dimensionalidad) de la variable que esté siendo medida. Lo anterior permitirá la medición de cada faceta por separado ya que entonces cada puntuación estará determinada por la suma de los ítems que pertenecen a cada dimensión correspondiente. Así pues, el análisis de consistencia interna alfa de Cronbach podrá ser realizado para cada dimensión por separado después de haber hecho el análisis factorial correspondiente.

Más aún y de manera general, el análisis factorial permite analizar las relaciones que mantiene un grupo de ítems entre sí extrayendo entonces una cantidad menor de factores y permitiendo el cálculo de las cargas factoriales entre los ítems y sus factores correspondientes. Las cargas factoriales son las correlaciones de Pearson entre los ítems y los factores extraídos, de manera que nos in-

forman que tan fuerte es la relación que tiene cada ítem con cada uno de los factores extraídos. Generalmente el objetivo es eliminar aquellos ítems que dificultan la obtención de una estructura interna sencilla y fácil de interpretar. En otras palabras, se toma la decisión de quitar aquellos ítems que no estén relacionados con ninguno de los factores extraídos por considerarlos irrelevantes o peor aún fuera de contexto.

Adicional a los enfoques anteriores, podemos diferenciar diversos tipos de variables. Algunas variables permiten clasificar a los objetos en grupos conocidos. En este tipo de casos es importante evitar que la muestra esté compuesta principalmente por objetos muy homogéneos, de lo contrario difícilmente será posible hallar relaciones significativas entre las variables. Si una variable en cuestión tiene mediciones con poca variabilidad intrínseca es entonces más difícil que pueda compartir alguna covarianza con otras variables. Si no hay efecto de covariabilidad entre las variables en un experimento, pues esto irá en perjuicio de la falta de correlación entre las variables.

Así pues, la validación de un instrumento de medición requiere la obtención de datos mediante el uso de una muestra relativamente heterogénea. Para ello, si lo que se requiere es comprobar dicha heterogeneidad entre los elementos de una muestra, es decir probar la existencia de distintos grupos, se puede entonces utilizar la técnica de *análisis de varianza*. Más aún si lo que se busca es evaluar en qué medida las mediciones obtenidas mediante un instrumento sirve para estimar a qué grupo pertenece cada uno de los casos, entonces se puede recurrir al *análisis discriminante*. Si los criterios que se desean predecir tienen carácter cuantitativo, un *análisis de regresión* sería el tipo de análisis sugerido.

Tamaño de la muestra y ponderación.

Como regla general podemos decir que la muestra a la cual se va a aplicar un instrumento de medición debería estar formada, al menos, por el doble de personas que la cantidad de ítems incluidos en el instrumento. Lo anterior es debido a que el tamaño de la muestra está relacionado con el error aleatorio de estimación de los ítems. Además, si no hay garantías de representatividad en la muestra, será indispensable asegurarse de que los resultados obtenidos sean estables, es decir, que no dependan de las características de la muestra utilizada.

Es trascendente tomar una buena decisión para definir el tamaño de una muestra ya que éste tendrá un efecto decisivo en la dispersión de la distribución muestral. El problema consiste en que en este momento no se cuenta con información de la distribución muestral y en muchas ocasiones tampoco de la población, por lo

cual es necesario realizar ciertas suposiciones. Vamos a mostrar la fórmula para el tamaño de muestra mediante el siguiente ejemplo numérico:

$$\frac{S^2}{V^2} = \frac{\text{Varianza de la Muestra} = p(1-p)}{\text{Varianza de la Población} = (\text{Máximo error estandar})^2}$$

Donde: $p = \text{probabilidad de ocurrencia para el evento}$

$$n = \frac{\frac{p(1-p)}{es^2}}{1 + \frac{p(1-p)}{N * es^2}} = \frac{\frac{0.1 * 0.9}{1.5\%^2}}{1 + \frac{0.1 * 0.9}{1176 * 1.5\%^2}} = \frac{400}{1 + \frac{400}{1176}} = 298$$

Junto con el procedimiento del tamaño de muestra se presentan algunas variantes tales como las siguientes:

1. *Muestreo de dominios*: el tamaño de cada muestra se ha decidido por separado en función a las características de cada dominio.
2. *Muestreo estratificado no proporcional*: cada estrato cuenta con una participación en la muestra que puede diferir de su participación en la población.

Ejemplo de un procedimiento para muestreo estratificado:

Estrato # 1	Población = 53	Muestra = 53*298/1176 = 13
2	109	Muestra=109*298/1176 = 28
-----	-----	-----
Estrato # n	51	Muestra = 51*298/1176 = 13
	N = Total Población = 1176	N = Total Muestra = 298

Esto es medido a partir de la distancia estandarizada de la distribución de probabilidad asociada a ese nivel de seguridad ($Z \alpha/2$) y al máximo error estándar permitido. Conforme aumenta la seguridad, es decir, $Z \alpha/2$ se hace más grande entonces aumenta el tamaño; y conforme aumenta la precisión, también aumenta el tamaño.

Estimación de la variación en la población: un aumento en la variación de la población tiene un efecto en el mismo sentido en el

tamaño de la muestra requerida, ya que se espera entonces valores muestrales con mayor dispersión. Por tanto, el tamaño de la muestra, debe aumentar para seguir manteniendo los mismos niveles de precisión y seguridad.

Existen situaciones en el que los elementos de una muestra no tienen todos el mismo peso o importancia relativa, es decir, donde no se ha respetado el principio de que todas las unidades de la población deben contar con la misma probabilidad de pertenecer a la muestra. Como resultado, unos grupos estarán sub-representados y por consecuencia otros lo estarán en exceso.

El proceso de ponderación consiste en volver a poner las cosas en su sitio. Una muestra pequeña, puede resultar en (1) poca potencia estadística o (2) un sobre ajuste de los datos pero con poca capacidad predictiva para generalizar hacia la población. Para el caso de muestras grandes, digamos arriba de 200 encuestados, se debe tener cuidado con la significancia estadística de los resultados para asegurarse que también se tiene una significación práctica debido al aumento de la potencia estadística como consecuencia del tamaño muestral. Por el contrario, se requiere un examen más riguroso de los datos debido a la influencia de casos atípicos sobre todo cuando se usan muestras relativamente pequeñas. En cualquier caso, es importante valorar la significancia práctica de los resultados dependiendo del tamaño de muestra utilizada.

Inferencia estadística en el análisis multivariable

Los datos para iniciar un análisis multivariable son valores que presentan las k variables objeto del análisis, las cuales contienen observaciones para n sujetos o casos.

El hecho de que sean dos las dimensiones permite representar los datos en una matriz con n filas y k columnas. A partir de dichos datos es posible construir nuevas matrices de covarianzas y/o de correlaciones. A diferencia de las matrices anteriores, este nuevo tipo de arreglos son cuadrados, es decir tienen k filas y k columnas. Algunas técnicas de análisis multivariable requieren medidas de proximidad, semejanza o distancia entre cada par de objetos o variables. Para ello se emplean coeficientes que miden cuán similares o distintas son entre sí dichos objetos. De hecho, dichas medidas de similitud o distancia son diferentes en función del carácter métrico o no métrico de las variables seleccionadas.

La mayoría de las técnicas de análisis multivariable tienen como objetivo la búsqueda de la combinación óptima de las variables implicadas en el análisis. Dicha combinación busca identificar las interrelaciones que existen entre esas variables o explicar el comportamiento de otra variable para fines predictivos o clasificatorios. Tal combinación óptima es el *modelo propuesto*, es decir, la

mezcla de variables que se adapta a los hechos observados de la muestra.

Una *muestra* es esencialmente un subconjunto de casos extraídos de una población incumbente. El objetivo de la muestra es conocer el modelo que mejor explica la realidad de la población y poder entonces generar conclusiones. Para ello se emplea la estadística inferencial, la cual permite inferir valores de características poblacionales en términos de probabilidad.

Dado que la inferencia se sustenta en la información de una muestra, es entonces pertinente entender la representatividad de una muestra, esto significa qué tamaño debe tener y cuáles son los métodos apropiados para su obtención. El objetivo final de la *estadística inferencial* es estimar un parámetro (o varios parámetros) de la población. Lo anterior tiene que ver con identificar la diferencia entre la estimación del parámetro y el verdadero valor del parámetro.

En general, todas las técnicas multivariantes, excepto el análisis cluster, se basan en la inferencia estadística para hacer uso de muestras e inferir hacia la población. Si partimos de un censo de toda la población, entonces la inferencia estadística no es necesaria, ya que cualquier diferencia o relación es verdadera y existe. Casi nunca se tiene posibilidad de acceder a un censo; por tanto, es necesario deducir inferencias de la población a partir de una muestra. Para interpretar las inferencias estadísticas, el investigador debe especificar los niveles aceptables de error estadístico.

En el análisis multivariable, la inferencia se lleva a cabo a partir de la prueba de hipótesis, es decir, se plantea una hipótesis sobre el valor el parámetro que se busca probar y según la información de la muestra se acepta o se rechaza la hipótesis.

La probabilidad de equivocarse al rechazar una hipótesis (i.e. en este caso la hipótesis nula), cuando se debería de aceptar, se conoce como *error tipo I* y requiere que no supere un cierto límite denominado nivel de significancia (α). La probabilidad de equivocarse al aceptar una hipótesis, cuando se debería de rechazar, se conoce como *error tipo II* (β), de aquí entonces se deriva la potencia estadística de la prueba la cual es equivalente a $(1 - \beta)$ la cual es la probabilidad de acertar al aceptar la hipótesis de investigación.

En cualquier prueba de hipótesis, se desea minimizar el error tipo I (α) y al mismo tiempo maximizar la potencia estadística $(1 - \beta)$. Sin embargo, estos conceptos están interrelacionados, pues cuando uno disminuye el otro también lo hace. Por tanto, el investigador debe elegir, en cuál de las dos hipótesis, es decir la nula o la de investigación, se incurre en un mayor costo al cometer una equivocación y así entonces elegir, entre minimizar el valor (α) de significancia o maximizar la potencia $(1 - \beta)$. En cualquier caso, para mi-

minimizar la probabilidad de un error se requiere usar muestras lo suficientemente grandes.

El enfoque más común es minimizar el nivel de error tipo I (i.e. alfa α). El error tipo I es la probabilidad de rechazar la hipótesis nula cuando es cierta. Especificando un nivel alfa, el investigador fija los márgenes admisibles de error especificando la probabilidad de concluir que la significación existe cuando en realidad no existe. El error de Tipo II es la probabilidad de fallar en rechazar la hipótesis nula cuando es realmente falsa. La probabilidad ($1 - \beta$), es denominado la potencia del test de inferencia estadística. *Potencia* es la probabilidad de rechazar correctamente la hipótesis nula cuando debe ser rechazada.

Aunque la especificación alfa establece el nivel de significación estadística aceptable, es el nivel de potencia el que dicta la probabilidad de éxito en la búsqueda de las diferencias si es que realmente existen. Los errores de tipo I y tipo II están inversamente relacionados, a medida que el error de tipo I se hace más restrictivo (se acerca a cero), el error de tipo II aumenta. Al disminuir el error de tipo I también se reduce el poder de la prueba estadística.

El analista tiene que conseguir un equilibrio entre el nivel de alfa y la potencia resultante. La potencia estadística no está sólo en función de ($1 - \beta$). Realmente está determinada por los siguientes tres factores:

1. **Alfa (α):** a medida que alfa se reduce, la potencia decrece. Dado que se reduce la probabilidad de encontrar un efecto incorrecto, la probabilidad de encontrar correctamente un efecto también disminuye. Se sugieren niveles alfa del 5% al 1%.
2. **Tamaño del efecto:** un efecto grande es más probable de encontrar que un efecto pequeño y por tanto, afecta a la potencia de la prueba estadística. Las diferencias respecto de la media se determinan en términos de desviaciones estándar, así que un efecto tamaño de 0,5 indica que la diferencia respecto de la media es la mitad de la desviación estándar.
3. **Tamaño de la muestra:** una muestra mayor siempre producirá una mayor potencia del test estadístico. El tamaño de la muestra afecta a la prueba estadística al hacerla insensible en el caso de muestras pequeñas, o demasiado sensible para el caso de muestras grandes.

Las relaciones entre alfa (α) y potencia estadística ($1 - \beta$), deben diseñarse para alfa de cuando mucho 5% para poder entonces alcanzar niveles de potencia estadística de al menos 80%. Al anticipar que los efectos van a ser pequeños, se debe entonces diseñar el

estudio con muestras mucho mayores, y con niveles de alfa menos restrictivos (del 5% al 10%). El investigador debe atender no sólo a la significación estadística de los resultados sino también a su significación práctica. La *significación práctica* se refiere a que los resultados deben tener un efecto demostrable que justifique las implicaciones teóricas. Por ejemplo, un modelo de regresión puede tener significancia estadística, pero su capacidad predictiva puede ser baja.

Hipótesis

Una *hipótesis* es una proposición que nos permite concluir el estado que guarda un fenómeno, o establecer relaciones existente entre los elementos de un fenómeno, con el fin de explicarlo o predecirlo.

En términos generales se presentan tres tipos de hipótesis, a continuación se muestran junto con sus ejemplos:

1. *Hipótesis estadística de estimación:*

$H_1: \chi > 200$ (La media muestral es mayor a 200)

$H_1: \chi = 200$ (La media muestral es igual a 200)

Supongamos que mediante la aplicación de Estadística Inferencial buscamos hacer inferencias a partir de una muestra. Tenemos entonces:

- α = Probabilidad de rechazar una hipótesis verdadera (Error Tipo I) = (5% al 1%). Esto es tener al menos un 95% de significancia.
- β = Probabilidad de aceptar una hipótesis falsa = Error Tipo II
- Suponga entonces una hipótesis que diga que el promedio de horas diarias que los niños estudian es igual a 3.0 horas.

i. $H_1: \chi = 3.0$

- Nivel de Significancia = $\alpha = 2.5\%$
(hipótesis de 2 colas)
- Esto corresponde a una " $Z_{critica} = 1.96$ "
- Muestra $n = 312$ niños tenemos: $\bar{x} = 2.9$ hrs. y $s = 1.2$ hrs.
- $$Z = \frac{x - \bar{x}}{s / \sqrt{n}} = \frac{2.9 - 3.0}{1.2 / \sqrt{312}} = -1.47$$

- Si $Z_{Calc} < Z_{critica}$ se acepta H_i a un nivel de significancia del 5%.

2. Hipótesis estadística de correlación:

$H_i: r_{(x,y)} > 75\%$ (Hay correlación entre las variables $x - y$)

$H_o: r_{(x,y)} = 0\%$ (NO hay correlación entre las variables $x - y$)

3. Hipótesis estadística de diferencia de medias:

$H_i: \chi_1 \neq \chi_2$

(La media del grupo 1 es diferente a la media del grupo 2)

$H_o: \chi_1 = \chi_2$

(La media del grupo 1 es igual a la media del grupo 2)

Supongamos que mediante la aplicación de Estadística Inferencial buscamos hacer inferencias a partir de dos muestras para probar la diferencia de 2 grupos. Tenemos entonces:

- Aplicación de la Prueba "t" para probar si "**2 GRUPOS**" difieren entre sí de manera significativa en lo referente a sus medias.
- $H_i: \chi_1 \neq \chi_2$ (La media del grupo 1 es diferente a la media del grupo 2)
- $H_o: \chi_1 = \chi_2$ (La media del grupo 1 es igual a la media del grupo 2)
- La prueba se aplica sobre UNA SOLA variable. Si hay diferentes variables, entonces se aplicarán varias pruebas "t". Normalmente en un experimento con 2 grupos tenemos que uno es al que se le aplica cierto tratamiento y el otro es el grupo control.

$$t_{Calc} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}} \quad \text{donde : } gl = N_1 + N_2 - 2$$

- Si $gl \geq 120$, entonces la distribución "t" se aproxima a la distribución normal.
- Si $T_{Calc} > T_{critica}$ se acepta H_i al nivel de significancia del 1-5%.

Análisis inicial y pruebas en los datos

La detección de problemas ocultos en los datos afecta en la obtención de resultados lógicos y consistentes. Dicho problema puede existir en los datos ausentes y en los casos atípicos (i.e. *outliers*). De igual modo es fundamental la comprobación de ciertas hipótesis iniciales necesarias para el análisis multivariable: (1) la normalidad, (2) la linealidad y (3) la homocedasticidad.

El análisis de *datos ausentes*, es el proceso de depuración de datos en donde el investigador se enfrenta con falta de información o inclusive información incorrecta en determinadas observaciones y/o variables. Esto lo conocemos como datos ausentes. Hay diversas razones acerca del por qué se presentan los datos ausentes tales como errores de captura, errores de codificación o simplemente errores o negativas del encuestado a responder. Lo anterior, afecta en el tamaño de la muestra disponible para el análisis así como en la explicación de los resultados. Por tanto es necesario depurar estos casos. Una opción es suprimir aquellas variables y/o casos que peor se comportan respecto al resto de la muestra.

El investigador debe comparar lo que gana con la exclusión de esta información versus lo que pierde. Se debe siempre procurar que los datos estén completos y se mantengan válidos. Otra alternativa es la estimación de valores ausentes empleando relaciones conocidas entre los valores válidos de las variables y/o de los casos de la muestra (i.e. aplicar procedimientos de interpolación).

A partir de una muestra, es relativamente fácil e importante detectar la existencia de observaciones que no siguen el mismo comportamiento del resto (i.e. *datos atípicos*). Los casos atípicos se pueden deber a errores en el procedimiento de medición, aunque también pueden ser consecuencia de un evento extraordinario que se manifiesta en dicha observación. Este acontecimiento anormal puede tener o no una causa para su explicación. La detección de los casos atípicos en una perspectiva univariable, puede ser hecho mediante la observación de aquellos casos que caigan fuera de ciertos límites de dispersión asumiendo una distribución normal.

Sin embargo, la detección de *outliers* en un análisis multivariable, es decir evaluar una observación tomando en consideración un conjunto de variables, es mucho más complejo. Una alternativa para ello es mediante la aplicación de la *distancia de Mahalanobis*, la cual es una métrica de distancia de cada observación en un espacio multidimensional respecto del centro medio de las observaciones. Otras alternativas para detectar *outliers* son el análisis discriminante el cual permite diferenciar entre los casos atípicos y el resto de observaciones.

El investigador debe decidir qué es lo más apropiado: si incluir toda la información con estas perturbaciones o eliminarlas definitivamente del análisis. La decisión no es trivial y debe ante todo estar fundamentada, ya que los casos atípicos pueden contener información representativa de un segmento pero a la vez pueden distorsionar seriamente las pruebas estadísticas. La decisión de lo anterior dependerá mayormente del alcance teórico que el investigador busque dar a sus hallazgos. Si las variables explicativas que el investigador está utilizando para dar explicación teórica a los efectos del fenómeno, no logran dar explicación a estos casos atípicos, entonces se tienen dos alternativas: (**1**) o se amplía el marco teórico

con mayor diversidad de variables explicativas, (2) o se limita el alcance de la investigación a cierto rango espectral del fenómeno, excluyendo los casos atípicos y reportando la imposibilidad de hacer extrapolaciones fuera del alcance investigado.

En el proceso del análisis inicial de los datos, el último paso consiste en comprobar que se cumplan las hipótesis iniciales que son requeridas para poder iniciar el análisis multivariable. Es necesario que los datos cumplan ciertos supuestos:

- a) Normalidad
- b) Linealidad y
- c) Homocedasticidad.

Supuesto de normalidad

El *supuesto de normalidad* exige que los datos a ser procesados tengan un perfil de distribución normal. Si este supuesto no se cumple, entonces todas las pruebas estadísticas resultantes no serán válidas, ya que el uso de la prueba t y prueba F lo requieren.

La normalidad multivariable implica que las variables individuales sean normales en un sentido univariable y que sus combinaciones también lo sean. Un caso en la que todas las variables univariadamente sean normales por supuesto contribuirá a obtener normalidad multivariable, aunque no lo garantiza. La normalidad multivariable requiere de mayor contrastación. En general, las muestras grandes tienden a disminuir los efectos indeseables de la no normalidad.

La *prueba z* de la distribución normal, utiliza un valor calculado para compararlo contra el valor de z crítico. Si el primero es mayor que el segundo, entonces se rechaza la hipótesis de que la distribución de datos sea normal. El valor crítico de una distribución z , está basado en el nivel de significación que se busque conseguir. Por ejemplo, un valor calculado que exceda a 2.58 indica que podemos rechazar el supuesto de normalidad a un nivel de probabilidad del 1%.

Un valor crítico habitualmente utilizado es 1.96 el cual corresponde a un nivel de significancia del 5%. Otra prueba de normalidad es el test de *Kolmogorov-Smirnov*. Este calcula el nivel de significación para las diferencias respecto a una distribución normal.

Las pruebas de significancia son menos útiles en muestras pequeñas (i.e. menores de 30) y muy sensibles para grandes muestras (i.e. mayores a 1000). Existen soluciones para enfrentar la no normalidad de los datos. En aquellos casos en las que las distribuciones no sean normales será necesario aplicar transformación en los datos.

Supuesto de linealidad y errores no correlacionados

Dentro de las especificaciones para poder llevar a cabo una regresión lineal múltiple es que exista una relación lineal estocástica entre las variables dependientes e independientes y el error del modelo. Esto es, cuando la relación que existe entre la variable dependiente e independientes es constante a lo largo de todos sus valores

La linealidad indica que el modelo a probar predice de manera lineal los valores de las variables dependientes a partir de las variaciones observadas en las variables independientes.

Técnicas como la regresión múltiple, el análisis factorial y los sistemas de ecuaciones estructurales se basan en medidas de correlación. Esto significa que es condición necesaria la existencia de asociaciones lineales entre las variables para obtener coeficientes de correlación y modelos que las representen.

Para medir la relación lineal entre dos variables basta con examinar los residuos después de efectuar un análisis de regresión. Estos residuos reflejan la variación no explicada por el modelo respecto a la variable dependiente. En caso de que la relación de los datos no sea lineal, se debe hacer una transformación de una o de ambas variables para conseguir la linealidad. Por ejemplo es posible efectuar transformaciones mediante la aplicación de una raíz cuadrada, o mediante el uso de variables polinómicas, para representar la parte no lineal de la relación. Los errores de predicción no son triviales, sino son el punto de partida para diagnosticar la validez de los resultados obtenidos y por tanto la oportunidad para identificar las relaciones que puedan seguir estando aún sin explicar.

Una de las pruebas para la comprobación de la linealidad del modelo, es la prueba gráfica de humo de cigarro, algunos paquetes estadísticos manejan esa opción. La idea es poder apreciar una inclinación de izquierda a derecha en ascenso.

La predicción en las técnicas de dependencia no es perfecta. Sin embargo, ante todo se procura que los errores de predicción no estén auto correlacionados. Por ejemplo, si se identifica que los errores son positivos y negativos alternativamente, entenderíamos que hay alguna relación sistemática no explicada por el conjunto de las variables independientes. En tal caso, no se puede tener certeza que los errores de predicción sean independientes de los niveles que estamos intentando predecir en la variable dependiente. Por tanto, sería muy probable que se esté dejando fuera del modelo una variable que esté afectando los resultados del análisis de dependencia.

La presencia de autocorrelación se puede deber debido a la existencia de datos de diferentes grupos. Si se analizaran los errores separadamente en ambos grupos, estos efectos se mantendrían

constantes y por tanto no habría influencia en la estimación de la relación. Pero si se combinan las observaciones de ambos grupos, entonces la relación estimada se vuelve un compromiso entre los dos tipos de grupos. Evidentemente este problema se debe a que una causa sin especificar (i.e. una variable ausente) está influyendo en la estimación de la relación. Por tanto, es necesario identificar la causa de la diferencia entre los grupos y el problema se soluciona con la inclusión de la variable ausente. Como se puede apreciar, la clave es que el investigador identifique el efecto no especificado y lo incluya de manera apropiado en el modelo.

El objetivo final del supuesto de linealidad es asegurar la capacidad del coeficiente de correlación para medir adecuadamente las relaciones entre las variables. En el caso de relaciones no lineales, se procede a transformar una o ambas variables para conseguir la linealidad o alternativamente, crear variables adicionales para representar los componentes no lineales.

La linealidad es un supuesto en todas las técnicas multivariantes basadas en medidas de correlación, esto incluye la regresión múltiple, regresión discreta, análisis factorial y modelos de ecuaciones estructurales. Dado que las correlaciones representan la asociación lineal entre las variables, los efectos no lineales no estarán presentes en el valor de correlación. Como resultado, es siempre importante identificar todas las variables y sus relaciones para identificar cualquier razón que este provocando la falta de linealidad que pueda impactar en la correlación. De no asegurar lo anterior, cualquier relación no lineal quedará incluida en los residuos. Una vez más, si se detecta una relación no lineal, la solución es transformar las variables involucradas para conseguir la linealidad.

Las *transformaciones de los datos* para cubrir los supuestos de normalidad y homocedasticidad ofrecen un medio para modificar el valor de las variables para cubrir dos objetivos: (1) corregir cualquier incumplimiento de los supuestos estadísticos de las técnicas multivariantes, o (2) mejorar la relación (correlación) entre las variables.

La transformación de los datos puede basarse en razones teóricas de acuerdo a la naturaleza de los datos. En general, el investigador puede proceder por aproximación, ponderando la mejora en la medida de la correlación frente a la necesidad de realizar transformaciones adicionales. Para distribuciones del tipo plano, la transformación más común es la inversa (i.e. $1/X \rightarrow 1/Y$). Las distribuciones asimétricas pueden ser transformadas empleando la raíz cuadrada o logaritmos. En cualquier caso, los datos transformados deberían ser contrastados para ver si se ha logrado el efecto deseado.

Supuesto de homocedasticidad

La *homocedasticidad*, consiste en verificar que la varianza de los errores sea constante. El objetivo es conseguir una dispersión de la varianza de la variable dependiente que sea equiparable a lo largo del rango de valores de la variable independiente. El caso en que este supuesto no se cumple, se denomina *heterocedasticidad*. Lo anterior suele deberse a muestras en las que aparecen asimetrías importantes en los valores de las variables ya sea porque estas toman valores muy altos o muy bajos respecto a la media.

Para verificar este supuesto, se puede hacer una revisión gráfica utilizando un “*scatterplot*” disponible en algunos productos de software para análisis estadístico, en donde en el eje de las Y tendremos la regresión estándar residual, y en el eje de las X tendremos la regresión estandarizada del valor estimado. Si los puntos de la gráfica se agrupan, es decir están reunidos en un punto, se dice que su varianza es constante e igual en todas las observaciones, es *homocedástica*. Si por el contrario, la gráfica nos reporta puntos dispersos o separados, entonces se dice que es heterocedástica.

La solución de problemas de Heterocedasticidad en algunos tipos de variables se puede resolver al convertir estas en variables especiales llamadas *variables dummy*, llamadas también *variables ficticias*. La prueba de homocedasticidad, puede ser hecha mediante el *test de Levene* para evaluar si la varianza de la variable dependiente permanece constante.

La homocedasticidad es un supuesto respecto a las relaciones de dependencia entre las variables. Se refiere al supuesto de que las variables dependientes exhiban iguales niveles de varianza a lo largo del rango de las variables predictoras (i.e. variables independientes). La propiedad de homocedasticidad es necesaria debido a que la varianza de la variable dependiente que se está explicando en el modelo de dependencia no debe centrarse sólo en un rango limitado de los valores independientes.

Este concepto de igual dispersión de la varianza a lo largo de las variables independientes aplica por igual cuando las variables sean métricas o no métricas. En cada una de las técnicas multivariantes, el propósito es el mismo: asegurar que la varianza usada en la explicación y predicción se disperse a través del rango de valores. Para que esta relación se capte apropiadamente, la dispersión (i.e. la varianza) de los valores de la variable dependiente debe mantenerse constante para los valores de la variable independiente. Los problemas de varianza desigual surgen en parte debido al tipo de variables incluidas en el modelo. Por ejemplo, a medida que una variable aumenta su escala (i.e. cuando las unidades van desde cero a millones), el rango de valores en las variables dependientes puede que su variación no sea en la misma magnitud. Esto provoca que las

estimaciones sean mejores a ciertos niveles de la variable independiente que a otros.

El efecto negativo de la heterocedasticidad frecuentemente está relacionado con el tamaño de la muestra. Por ejemplo, el impacto de la heterocedasticidad depende de los tamaños de la muestra asociados con los grupos de menor o mayor varianza. En la regresión múltiple ocurren este tipo de efectos en distribuciones donde existan un número desproporcionado de encuestados en ciertos rangos de la variable independiente. En general, los problemas de heterocedasticidad pueden solucionarse igual que a la falta de normalidad, a través de transformaciones en los datos. De hecho, en muchas ocasiones la heterocedasticidad es el resultado de la no normalidad. Por tanto la corrección de la no normalidad resuelve igualmente la dispersión de la varianza.

Índice del capítulo

- Alfa*, 26
- alfa de Cronbach*, 18
- análisis de regresión*, 22
- análisis de varianza*, 22
- análisis discriminante*, 22
- análisis factorial*, 21
- análisis multivariable*, 2, 5
- análisis multivariante*, 2
- Charles Spearman*, 2
- coeficiente de correlación*, 2
- coeficientes del modelo*, 8
- consistencia interna*, 18
- Consistencia interna*, 18
- constructo*, 17
- datos*, 8
- datos atípicos*, 29
- datos ausentes*, 29
- datos cualitativos*, 8
- datos cuantitativos*, 8
- distancia de Mahalanobis*, 29
- error tipo I*, 25
- error tipo II*, 25
- escala acumulativa*, 11
- escala de Guttman*, 11
- escala de intervalo*, 9
- escala de Likert*, 11
- escala de razón*, 9
- escala de respuesta*, 14
- escala de Thurstone*, 11
- escala diferencial*, 11
- Escala nominal*, 8
- Escala ordinal*, 9
- escala sumativa*, 11
- estadística inferencial*, 25
- fiabilidad*, 16
- Formas equivalentes*, 18
- heterocedasticidad*, 33
- hipótesis*, 27
- Hipótesis estadística de correlación*, 28
- Hipótesis estadística de diferencia de medias*, 28
- Hipótesis estadística de estimación*, 27
- homocedasticidad*, 33
- investigación causal*, 6
- investigación correlacional*, 6
- investigación descriptiva*, 6
- investigación experimental*, 5
- Investigación longitudinal*, 6
- investigación no experimental*, 5
- Investigación transeccional*, 6
- ítem*, 11

Karl Pearson, 2
Kolmogorov-Smirnov, 30
Mitades, 18
modelo propuesto, 25
muestra, 25
Muestreo de dominios, 23
Muestreo estratificado no proporcional, 23
Planteamiento del problema, 6
Potencia, 26
prueba z, 30
relevancia de un ítem, 16
representatividad de un ítem, 16
significación práctica, 27
significancia, 7
supuesto de normalidad, 30
Tamaño de la muestra, 26
Tamaño del efecto, 26
test de Levene, 33
Test-Retest, 18
transformaciones de los datos, 32
validez, 16
validez de contenido, 16
validez de convergencia, 13
validez de estructura interna, 13
validez de la escala, 16
validez del instrumento, 13
variable dicotómica, 9
variables, 6
variables dummy, 33
variables ficticias, 33

Capítulo 2:

Técnicas de análisis multivariable

Contenido

Clasificación de técnicas de análisis multivariable	38
Análisis de la varianza	39
Análisis de varianza con un factor	41
Análisis de varianza de dos factores con una sola muestra por grupo.....	43
Análisis de varianza de dos factores con varias muestras por grupo.....	44
Análisis de varianza de tres factores con varias muestras por grupo.....	45
Modelo de regresión lineal múltiple	47
Introducción al análisis de regresión múltiple.....	48
Supuestos de la regresión para mínimos cuadrados ordinarios.....	50
Medición del ajuste del modelo de regresión.....	51
Fundamentos estadísticos del modelo de regresión	53
Significación del modelo de regresión	59
Validación del modelo	61
Identificación de puntos atípicos e influyentes	62
La multicolinealidad	65
Índice del capítulo	67

Clasificación de técnicas de análisis multivariable

La selección de la técnica de análisis multivariable es crucial debido a la existencia de un conjunto numeroso de éstas. Podemos plantear dos categorías principales según su uso:

1. **Técnicas para el análisis de dependencia:** aplicable cuando una o varias variables dependientes van a ser explicadas por un conjunto de variables independientes que actúan como predictoras.
2. **Técnicas para el análisis de Interdependencia:** técnicas que otorgan la misma consideración a todas las variables del estudio sin distinguir entre dependientes e independientes y que tienen como objetivo descubrir las interrelaciones entre ellas para propósitos de clasificación.

Un análisis de dependencia puede definirse como aquel en el que una variable o conjunto de variables son identificadas como las *variables dependientes* y van estas a ser explicadas por otras variables conocidas como *variables independientes*. Si se tiene una única variable dependiente que es métrica, la técnica apropiada es el *análisis de regresión múltiple*. Si la única variable dependiente es no métrica (categórica), entonces se pueden utilizar el *análisis discriminante* o los *modelos de regresión discreta*. Cuando el problema implica varias variables dependientes y varias variables independientes entonces debemos elegir la técnica *ecuaciones estructurales*.

Un análisis de interdependencia es aquel en que ninguna variable o grupo de variables es definido como independiente o dependiente. En el análisis de interdependencia, las variables no son clasificadas como dependientes o independientes, sino todas las variables son analizadas simultáneamente. El *análisis factorial* es un ejemplo de un análisis de interdependencia. Si se está analizando la estructura de las variables, entonces el análisis factorial es la técnica apropiada. Si los casos se van a agrupar para representar una estructura, entonces se usa el *análisis cluster*.

Se pueden tener restricciones dependiendo de los tipos de datos, aunque los inconvenientes se pueden resolver de forma sencilla. Por ejemplo, el uso de variables ficticias es una manera de transformar datos no métricos en datos métricos. Aquí se asignan unos y ceros al sujeto, dependiendo de si cuenta o no con cierta característica. Los datos no métricos pueden ser transformados a través de una variable ficticia para usarlos en el análisis factorial o en el análisis cluster según corresponda.

A continuación se presenta un resumen de la clasificación de las técnicas multivariadas de acuerdo al tipo de dependencia o interdependencia según corresponda:

Técnicas multivariable para el análisis de dependencia:

1. Análisis de varianza
2. Modelos de regresión y correlación múltiple
3. Modelos de regresión discreta
4. Análisis discriminante
5. Modelos de ecuaciones estructurales

Técnicas multi-variable para el análisis de interdependencia:

1. Análisis factorial y componentes principales
2. Análisis cluster

A través de la aplicación de una sola técnica multivariable, difícilmente se asegurará la mejor predicción. Cualquiera que sea la técnica multivariable empleada, es necesario asegurar no sólo un modelo significativo sino también en que este sea representativo de la población. El objetivo no es identificar el modelo que ofrezca el mejor ajuste para la muestra sino en desarrollar el modelo que mejor describa a la población.

Análisis de la varianza

El *análisis de varianza* es una técnica estadística para explorar las relaciones entre varias variables independientes (i.e. tratamientos) y una variable dependiente.

Se diseña un experimento manipulando varias variables de tratamiento para comprobar hipótesis respecto a la varianza de las respuestas obtenidas en la variable dependiente. El análisis de la varianza o *ANOVA* es una técnica caracterizada por el empleo de una variable dependiente de carácter métrico y varias independientes no métricas que actúan como predictoras. Formalmente tenemos:

$Y = f(X_1, X_2, \dots, X_n)$, donde Y es métrica y las X son no métricas

Se utiliza para analizar si diversas muestras proceden de la misma población con la misma media. Las muestras se identifican mediante los valores de cada una de las variables independiente. El ANOVA mide la significancia estadística para probar hipótesis de diferencias de grupos entre las medias que la variable dependiente tiene de cada una de las distintas muestras (i.e. grupos).

El *análisis multivariable de la varianza (MANOVA)* es una extensión del análisis (univariable) de la varianza, que se aplica cuando se presenta un grupo de variables dependientes relacionadas entre sí.

Formalmente tenemos:

$G(Y_1, Y_2, \dots, Y_m) = f(X_1, X_2, \dots, X_n)$, donde las Y son métricas y las X son no métricas.

Por ejemplo, si se desea estudiar el progreso en física y en química de dos grupos de alumnos que utilicen distinto libro de texto, es decir diferentes tratamientos representados a través de cada variable independiente, entonces se puede aplicar un análisis ANOVA por separado para cada una de las materias. Sin embargo, si se sospecha que el avance de los alumnos en una materia y en la otra, está correlacionada, entonces se debería aplicar un análisis MANOVA.

El análisis de la varianza se basa en la descomposición de la *suma de cuadrados* total de los elementos muestrales X_{ij} respecto a su media global a través de dos términos:

1. Suma de cuadrados entre las muestras: mide la influencia del factor en la variabilidad.
2. Suma de cuadrados dentro de las muestras: mide la influencia del error muestral en la variabilidad.

El análisis de varianza permite evaluar el efecto de dos o más variables independientes sobre una variable dependiente. Evalúa el efecto por separado de cada variable independiente así como el efecto conjunto. Suponga un análisis ANOVA con tres variables independientes (i.e. tres tratamientos diferentes), entonces la tabla típicamente incluiría los siguientes elementos:

- Efectos principales: $(X_1, X_2, X_3) - (F_1, F_2, F_3) - (S_1, S_2, S_3)$
- Interacción en 2 direcciones:
 $(X_1X_2, X_1X_3, X_2X_3) - (F_1F_2, F_1F_3, F_2F_3) - (S_1S_2, S_1S_3, S_2S_3)$
- Interacción en 3 direcciones: $(X_1X_2X_3) - (F_1F_2F_3) - (S_1S_2S_3)$
- Probar si **tres o más grupos** difieren entre sí de manera significativa en sus medias:
 - $H_i: \chi_1 \neq \chi_2 \neq \chi_3$
(Los grupos difieren significativamente entre sí)

- $H_0: \chi_1 = \chi_2 = \chi_3$
(Los grupos no difieren significativamente entre sí)
- La prueba **F** compara la variabilidad entre los grupos que están siendo comparados versus la variabilidad encontrada al interior de cada uno de los grupos. Es decir, se comparan las 2 posibles fuentes de variación.
- Lo que se espera es baja variabilidad intra-grupo y alta variabilidad entre-grupos.

$$F_{Calc} = \frac{\text{Suma}^2 \text{ Entre - Grupos} / \text{GL Entre - Grupos}^{(k-1)}}{\text{Suma}^2 \text{ Intra - Grupos} / \text{GL Intra - Grupos}^{(n-k)}}$$

donde : $k = \#$ de Grupos, $n =$ tamaño de la muestra

- Si $F_{Calc} > F_{critica}$ se acepta H_i al nivel de significancia del 1-5%.

Análisis de varianza con un factor

$$ST = \sum (X_{ij} - \bar{X})^2 = \sum (\bar{X}_i - \bar{X})^2 + \sum (X_{ij} - \bar{X}_i)^2 = SCA + SCE = SA + SE$$

SCA = Suma de Cuadrados debido al factor (**A**) = **SA**

SCE = Suma de Cuadrados debido al Error Muestral = **SE**

SCT = Suma de Cuadrados Total = **ST**

El coeficiente **SA** mide la influencia de los efectos debido a la presencia del factor **A** en la heterogeneidad de los valores de X_{ij} . Por su parte el coeficiente **SE** mide la influencia de los errores aleatorios **Eij** en la heterogeneidad de los valores de X_{ij} . Por tanto, se obtiene la siguiente ecuación aditiva para la explicación de los elementos X_{ij} .

$$X_{ij} = \mu + A_i + E_{ij}$$

Donde: $A_i = \mu_i - \mu$

$$E_{ij} = X_{ij} - \mu_i$$

Se deduce que el valor de un elemento X_{ij} es igual a la suma de los siguientes efectos:

- μ = el resultado medio global de todos los tratamientos más...
- A_i = el efecto añadido por la aplicación del tratamiento i del factor A más...
- E_{ij} = el error aleatorio del caso j al aplicarse el tratamiento i .

Los *grados de libertad* para un factor, se calculan como el número total de tratamientos incluidos en el factor menos uno. En general, el *cuadrado medio* se calcula como sigue:

$$\text{Cuadrados Medio} = \text{Suma de Cuadrados} \div \text{Grados de Libertad}$$

$$SA_m = SA / (a - 1), \quad SE_m = SE / (N - a), \quad ST_m = ST / (N - 1)$$

Observe también que:

- *Suma de Cuadrados del Error* =
 - *Suma de Cuadrados Total* – *Suma de Cuadrados del factor A*
 - $SE = ST - SA$
- *Grados de Libertad del Termino de Error* =
 - *Grados de Libertad Total* – *Grados de Libertad del factor A*
 - $GL_E = GL_T - GL_A$

La prueba de contraste mediante el estadístico F sería como sigue:

$H_0: \mu_i = \mu \rightarrow$ La presencia del factor A NO afecta en la heterogeneidad de los valores de X_{ij}

$H_a: \mu_i \neq \mu \rightarrow$ La presencia del factor A SI afecta en la heterogeneidad de los valores de X_{ij}

Si $SA_m \div SE_m > F(\alpha)$ se rechaza H_0

Si $SA_m \div SE_m \leq F(\alpha)$ se acepta H_0

Al rechazar la H_0 , lo que significa es que al menos uno de los tratamientos incluidos en el factor A es diferente del resto. En otras palabras, lo que se está rechazando es la igualdad simultánea de las medias μ_i de todos los tratamientos. Para averiguar cuál es el tratamiento del factor A que es diferente, se aplica entonces el estadístico T de diferencias de medias para hacer contrastes de dos en dos.

Análisis de varianza de dos factores con una sola muestra por grupo

Se tiene ahora la siguiente ecuación aditiva para la explicación de los elementos X_{ij} .

$$X_{ij} = \mu + A_i + B_j + E_{ij}$$

Donde:

$$A_i = \mu_i - \mu$$

$$B_j = \mu_j - \mu$$

$$E_{ij} = X_{ij} - \mu_{ij}$$

(esto es el valor observado versus el valor verdadero)

Se deduce que el valor de un elemento X_{ij} es igual a la suma de los siguientes efectos:

- μ = el resultado medio global de todos los tratamientos de los factores **A** y **B** más...
- A_i = el efecto añadido por la aplicación del tratamiento **i** del factor **A** más...
- B_j = el efecto añadido por la aplicación del tratamiento **j** del factor **B** más...
- E_{ij} = el error aleatorio al aplicarse el tratamiento **i** del factor **A** y el tratamiento **j** del factor **B**.
- **$SE = ST - SA - SB$, $GLE = GLT - GLA - GLB$**

Los Cuadrados Medios se calculan como sigue:

$$SA_m = SA / (a - 1)$$

$$SB_m = SB / (b - 1)$$

$$SE_m = SE / GLE$$

$$ST_m = ST / (N - 1)$$

Las pruebas de contraste mediante el estadístico F serían como sigue:

$H_a: \mu_i \neq \mu \rightarrow$ La presencia del factor A afecta en la heterogeneidad de los valores de X_{ij}

$H_b: \mu_j \neq \mu \rightarrow$ La presencia del factor B afecta en la heterogeneidad de los valores de X_{ij}

Si $SA_m \div SE_m > F(\alpha)$ se acepta H_a

Si $SB_m \div SE_m > F(\alpha)$ se acepta H_b

Análisis de varianza de dos factores con varias muestras por grupo

Se tiene la siguiente ecuación aditiva para la explicación de los elementos X_{ij} .

$$X_{ij} = \mu + A_i + B_j + I_{ij} + E_{ij}$$

Donde: $A_i = \mu_i - \mu$

$$B_j = \mu_j - \mu$$

$$I_{ij} = \mu_{ij} - \mu - A_i - B_j$$

$$E_{ij} = X_{ij} - \mu_{ij}$$

(esto es el valor observado versus el valor verdadero)

El valor de un elemento X_{ij} es igual a la suma de los siguientes efectos:

- μ = el resultado medio global de todos los tratamientos de los factores A y B más...
- A_i = el efecto añadido por la aplicación del tratamiento i del factor A más...
- B_j = el efecto añadido por la aplicación del tratamiento j del factor B más...
- I_{ij} = el efecto añadido por la interacción de los tratamientos i y j de los factores A y B más...
- E_{ij} = el error aleatorio al aplicarse el tratamiento i del factor A y el tratamiento j del factor B .
- $SE = ST - SA - SB - SI$,
- $GL_E = GL_T - GL_A - GL_B - GL_I$

Los cuadrados medios se calculan como sigue:

$$SA_m = SA / (a - 1)$$

$$SB_m = SB / (b - 1)$$

$$SI_m = SI / [(a - 1) (b - 1)]$$

$$SE_m = SE / GL_E,$$

$$ST_m = ST / (N - 1)$$

Las pruebas de contraste mediante el estadístico **F** serían como sigue:

$H_a: \mu_i \neq \mu \rightarrow$ La presencia del factor **A** SI afecta en la heterogeneidad de los valores de X_{ij}

$H_b: \mu_j \neq \mu \rightarrow$ La presencia del factor **B** SI afecta en la heterogeneidad de los valores de X_{ij}

$H_{ab}: \mu_{ij} \neq \mu, I_{ij} \neq 0 \rightarrow$ La interacción de los factores **AB** SI afecta en la heterogeneidad de X_{ij}

Si $SA_m \div SE_m > F(\alpha)$ se acepta H_a

Si $SB_m \div SE_m > F(\alpha)$ se acepta H_b

Si $SI_m \div SE_m > F(\alpha)$ se acepta H_{ab}

Análisis de varianza de tres factores con varias muestras por grupo

Se tiene ahora la siguiente ecuación aditiva para la explicación de los elementos X_{ijk} .

$$X_{ijk} = \mu + A_i + B_j + C_k + I_{ij} + E_{ijk}$$

Donde: $A_i = \mu_i - \mu$

$$B_j = \mu_j - \mu$$

$$I_{ij} = \mu_{ij} - \mu - A_i - B_j$$

$$C_k = \mu_k - \mu$$

$$E_{ijk} = X_{ijk} - \mu_{ijk}$$

(esto es el valor observado versus el valor verdadero)

Se deduce que el valor de un elemento X_{ijk} es igual a la suma de los siguientes efectos:

- μ = el resultado medio global de todos los tratamientos de los factores A , B y C más...
- A_i = el efecto añadido por la aplicación del tratamiento i del factor A más...
- B_j = el efecto añadido por la aplicación del tratamiento j del factor B más...
- I_{ij} = el efecto añadido por la interacción de los tratamientos i y j de los factores A y B más...
- C_k = el efecto añadido por la aplicación del tratamiento k del factor C más...
- E_{ijk} = el error aleatorio al aplicarse el tratamiento A_i , B_j y C_k .
- $SE = ST - SA - SB - SI - SC$
- $GL_E = GL_T - GL_A - GL_B - GL_I - GL_C$

Los cuadrados medios se calculan como sigue:

$$SA_m = SA / (a - 1)$$

$$SB_m = SB / (b - 1)$$

$$SC_m = SC / (c - 1)$$

$$SI_m = SI / [(a - 1)(b - 1)]$$

$$SE_m = SE / GL_E$$

$$ST_m = ST / (N - 1)$$

Las pruebas de contraste mediante el estadístico F serían como sigue:

$H_a: \mu_i \neq \mu \rightarrow$ La presencia del factor A si afecta en la heterogeneidad de los valores de X_{ijk}

$H_b: \mu_j \neq \mu \rightarrow$ La presencia del factor B si afecta en la heterogeneidad de los valores de X_{ijk}

$H_{ab}: \mu_{ij} \neq \mu, I_{ij} \neq 0 \rightarrow$ La interacción de los factores AB si afecta en la heterogeneidad de X_{ijk}

$H_c: \mu_k \neq \mu \rightarrow$ La presencia del factor C si afecta en la heterogeneidad de los valores de X_{ijk}

Si $SA_m \div SE_m > F(\alpha)$ se acepta H_a

Si $SB_m \div SE_m > F(\alpha)$ se acepta H_b

Si $SI_m \div SE_m > F(\alpha)$ se acepta H_{ab}

Si $SC_m \div SE_m > F(\alpha)$ se acepta H_c

Modelo de regresión lineal múltiple

El *modelo de regresión lineal múltiple* es el método de análisis utilizado para modelos de investigación con una única variable métrica dependiente relacionada con una o más variables métricas independientes. Su objetivo es predecir los cambios en la variable dependiente en respuesta a los cambios presentados en las variables independientes.

La regresión lineal permite analizar la relación que existe entre una variable dependiente métrica y varias variables independientes también métricas. Formalmente tenemos:

$Y = f(X_1, X_2, \dots, X_n)$, donde Y es métrica y las X son métricas

Con un análisis de regresión múltiple se busca determinar la combinación lineal de variables independientes cuyos cambios marginales son los mejores predictores en los cambios experimentados por la variable dependiente. Dicha combinación lineal se obtiene a través de la minimización de la suma de los cuadrados de los errores de regresión (i.e. esto es mínimos cuadrados ordinarios). Cada error de regresión es la diferencia entre el valor real de la variable dependiente versus su valor estimado por el modelo mediante la combinación lineal de las variables independientes.

En general en cualquier modelo de regresión múltiple, es importante evitar omitir una variable predictora, ya que esto generaría un error de especificación en el modelo. Sin embargo también se debe evitar insertar variables indiscriminadamente. Las variables siendo irrelevante en el modelo de predicción, habitualmente aumentan la capacidad de ajuste del modelo a la muestra de datos pero a costa de la capacidad predictiva del modelo. Esto ocasiona que el modelo en general sea menos generalizable para la población. Además, las variables irrelevantes pueden afectar los efectos verdaderos de las variables relevantes debido a la multicolinealidad.

La *multicolinealidad* representa el grado en donde el efecto de una variable puede ser explicado por otras variables. A medida que aumenta la multicolinealidad, se pierde capacidad para identificar el efecto individual de cualquier variable. Por tanto, incluir variables que no sean relevantes incrementa la multicolinealidad, disminuye la capacidad predictiva del modelo y disminuye la significancia estadística del modelo debido a la pérdida de grados de libertad.

Las interrelaciones entre las variables implican que los resultados sean específicos no sólo para la muestra sino también generalizables a la población. Se debe asegurar que existan observaciones suficientes por cada parámetro que se esté estimando para evitar que el modelo quede sobreajustado por los datos de la muestra.

Introducción al análisis de regresión múltiple

El *análisis de regresión múltiple* es una técnica estadística que puede utilizarse para analizar la relación entre una única variable dependiente y varias variables independientes (*predictores*).

El objetivo del análisis de regresión múltiple es usar las variables independientes, cuyos valores son conocidos para predecir la única variable dependiente. Cada variable predictora tiene un coeficiente el cual indica su contribución relativa a la predicción conjunta del modelo. Al calcular los coeficientes, el procedimiento del análisis de regresión asegura la máxima capacidad de predicción del modelo a partir del conjunto de variables independientes. Estos coeficientes permiten la interpretación del impacto de cada variable independiente en la capacidad predictiva del modelo hacia la variable dependiente.

El análisis de regresión se utiliza únicamente cuando las variables dependientes como las independientes son métricas. Sin embargo, es posible incluir datos no métricos para las variables independientes transformando las variables ordinales o nominales en variables ficticias. En el caso de la variable dependiente sería necesario el uso de una variable binaria y la aplicación de la técnica de regresión discreta.

Llamaremos a \mathbf{X} la variable predictora, y llamaremos \mathbf{Y} a la variable de respuesta o variable dependiente. Como regla general. Los modelos de regresión sólo son válidos dentro del rango de las variables regresoras contenidas en los datos medidos. Así, el modelo general de regresión lineal queda como:

$$Y = B_0 + B_1 * X_1 + \dots B_n * X_n + E$$

Todo análisis de regresión pretende llegar a ser tan bueno como lo sean los datos sobre los que se desarrolla el modelo. Básicamente, existen tres métodos para la recolección de los datos:

1. Uso de datos históricos.
2. Medición y observación.
3. Diseño experimental.

La mejor estrategia de recolección de datos es hacer un experimento diseñado donde se puedan metódicamente hacer variar las mediciones en las variables independientes y observar la respuesta obtenida en la variable

dependiente. Esta estrategia se denomina diseño de experimentos. Esta busca separar los efectos de cada factor (i.e. variable predictora) sobre la respuesta observada en la variable dependiente.

Los valores que se hacen variar para cada factor (i.e. variable independiente) que se ajustan en el diseño experimental se llaman *niveles*. Generalmente se utiliza una cantidad pequeña de niveles para cada factor, digamos dos o tres niveles por cada factor.

Una *combinación de tratamientos* es una combinación específica de niveles para cada uno de los factores. Un *diseño experimental* consiste así en una serie de mediciones (i.e. replicas) hechas para cada uno de las combinaciones de niveles entre los factores (i.e. para cada tipo de tratamiento).

Los modelos de regresión se usan en general para cualquiera de los siguientes objetivos:

1. Para la estimación de los parámetros del modelo.
2. Como modelo de predicción y estimación.
3. Como sistema de control.

Un resultado positivo en un modelo de regresión no significa la prueba definitiva de que haya una relación de causa efecto entre las variables. Un modelo de regresión puede ser útil como medio de interpolación dentro del intervalo de las variables regresoras que se usan para ajustar el modelo. Sin embargo, se debe tener cuidado al extrapolar fuera de ese intervalo. En general, mientras más grande sea la extrapolación, es mayor la probabilidad de que se presente un error de estimación en los resultados arrojados por el modelo de predicción. Aunque pueda existir una marcada relación empírica entre las variables, no puede considerarse esto como una prueba definitiva de que las variables regresoras y la respuesta estén relacionadas en forma de causa y efecto.

Para establecer de manera definitiva la causalidad entre los regresores y la respuesta, debe existir una base ajena a las mediciones de los datos de la muestra. Esto significa que la relación de casualidad debe estar apoyada a través de la teórica y la lógica. En efecto, el análisis de regresión contribuye a confirmar la relación de causa-efecto a partir de nuestras teorías, pero no puede ser esto el único fundamento.

Es importante recordar que el análisis de regresión, como técnica de análisis de datos, es tan solo una parte de un objetivo mucho más amplio el cual consiste en contribuir a la toma de decisiones. El diseño de un modelo de regresión es un proceso iterativo. El modelo va evolucionando a medida que se descubren mejoras en la estructura del modelo y fallas en la medición de los datos (i.e. datos atípicos). En general, para el análisis de regresión se requiere el uso inteligente y diestro de la tecnología computacional.

Supuestos de la regresión para mínimos cuadrados ordinarios

Los supuestos para la aplicación de un Modelo de Regresión mediante el método de *mínimos cuadrados ordinarios* (MCO) son los siguientes:

1. La relación entre la respuesta y los regresores es lineal. Se define una relación lineal estocástica entre la variable dependiente y las variables explicativas. Las variables explicativas (i.e. independientes) del modelo son deterministas.
2. El número de observaciones relativas a las variables explicativas del modelo debe ser al menos el número de coeficientes a estimar multiplicado por 5 ($n \geq 5k$).
3. El modelo se supone correctamente especificado y sin errores, esto es, no hay omisión de variables relevantes ni inclusión de variables irrelevantes.
4. No debe existir correlación entre las variables explicativas del modelo (multicolinealidad).
5. No debe existir correlación entre las variables explicativas y los residuales del modelo los cuales deben ser aleatorios.
6. Los residuales del modelo E son variables aleatorias independientes y se distribuyen normalmente. Además estando estandarizados, estos tienen media cero y varianza constante.
7. Los residuales E de observaciones diferentes no deben estar correlacionados (i.e. no-autocorrelación).
8. Se adopta la hipótesis de permanencia estructural, esto significa que los parámetros permanecen constantes para el periodo de observación y de predicción.

El incumplimiento a los supuestos anteriores pueden producir un modelo inestable, esto quiere decir que el uso de muestras distintas de la población, puede conducir a un modelo totalmente diferente. En general, no se pueden detectar dichos incumplimientos a través de los estadísticos t , F o R^2 . El diagnóstico del cumplimiento de los supuestos anteriores se realiza a través del estudio de los residuales del modelo. El residual es una medida de la variabilidad de la variable de respuesta que no puede ser explicada por el modelo de regresión.

El análisis de los residuales es una forma eficaz de descubrir diversos tipos de inadecuación del modelo. En general, es mejor trabajar con los residuales estandarizados para determinar observaciones que sean atípicas o valores extremos. Los residuales estandarizados tienen media cero y varianza aproximadamente unitaria. Por tanto, un residual estandarizado grande (i.e. mayor que tres) indica que se trata de un valor atípico potencial.

Los estimados obtenidos por el método de mínimos cuadrados ordinarios son aquéllos que minimizan la suma de las diferencias al cuadrado

para el total de las observaciones. Esto es entre los valores observados de la variable dependiente versus los estimados por el modelo de regresión. La capacidad predictiva de un modelo, está determinado por el poder que este tiene para estimar el valor que tomará la variable dependiente fuera del periodo de observación (i.e. fuera de la muestra seleccionada). Además de tener un papel importante en la potencia estadística del modelo de regresión, el tamaño muestral también afecta a la generalización de los resultados. Dicha afectación es de acuerdo al ratio de observaciones entre las variables independientes, es decir:

Cantidad de observaciones / Cantidad de variables predictoras ≥ 5

Una norma de uso general es que este ratio nunca debería caer por debajo de cinco. Lo anterior significa que deben existir al menos cinco observaciones para cada variable independiente. Conforme cae este ratio por debajo de cinco, se puede incurrir en el riesgo de sobre ajustar el valor teórico de los coeficientes betas de las variables predictoras respecto a los datos de la muestra. Lo anterior provoca que dichos resultados sean sólo específicos para la muestra y por tanto con falta de generalización hacia otras muestras de la población. Aunque el ratio mínimo es 5, en realidad el nivel deseado está entre 15 y 20 observaciones por cada variable independiente.

Medición del ajuste del modelo de regresión

Conocido como el error estándar de la estimación o también como la *suma de cuadrados de los errores (SCE)*, esta medida es, la desviación estándar de los errores de predicción. La suma de cuadrados de los errores es el error estadístico. Se trata de una variable aleatoria que define en qué medida el modelo no ajusta exactamente a los datos. Este error estadístico está definido por el efecto que las variables predictoras (i.e. las variables independientes) no pueden estimar sobre el resultado de la variable dependiente. Complementariamente, la medida del éxito predictivo se denomina *suma de cuadrados de la regresión (SCR)*. La suma de ***SCR + SCE*** es igual a la suma de cuadrados total (***SCT***). Otra forma de expresar este nivel de precisión predictiva es mediante el *coeficiente de determinación R^2* . Dicho coeficiente se calcula como sigue:

$$\begin{aligned} R^2 &= \text{Suma de Cuadrados de la Regresión} / \text{Suma de Cuadrados Total} \\ &= SCR / SCT \end{aligned}$$

El coeficiente de determinación **R^2** es utilizado para probar la hipótesis de que la cantidad de variación explicada por el modelo de regresión es mayor que la variación explicada por la media (i.e. probar **$R^2 > 0$**). A la raíz cuadrada de **R^2** se le denomina *coeficiente de correlación (r)*. Cuando se utiliza el coeficiente de correlación para evaluar la relación entre las variables dependientes e independientes, el signo del coeficiente de correlación

($+r$, $-r$) denota la pendiente de la línea de regresión. Sin embargo, el impacto de la relación se representa a través del coeficiente de determinación R^2 . Por ejemplo una $R^2 = 0.75$, indica que el 75% de la variación en la variable dependiente está explicada por la variable independiente.

En general, el coeficiente de determinación R^2 está influenciado por el número de variables predictoras (i.e. independientes) así como por el tamaño muestral. Se sugiere un mínimo de 5 observaciones por cada variable independiente. De ser posible aumentar esta relación hasta quince observaciones. El tamaño muestral tiene un impacto directo en la potencia estadística del modelo de regresión.

La *potencia de la regresión múltiple* se refiere a la probabilidad de detectar como estadísticamente significativo un nivel específico de R^2 . Muestras pequeñas (i.e. menos de 20 observaciones), sólo pueden detectar relaciones muy fuertes con cierto grado de certidumbre. Por otro lado, muestras muy grandes (i.e. más de 1,000 observaciones), provocan que los tests de significación estadística sean demasiado sensibles y a la vez ocasionan que casi cualquier relación se manifieste como estadísticamente significativa. En general, al usar muestras muy grandes se necesario asegurar no sólo el criterio de significación estadística sino también el de significación práctica.

En general, si se agrega una variable a un modelo de regresión, la suma de cuadrados de la regresión aumenta y la suma de cuadrados de los errores disminuye. Se debe entonces identificar si el aumento de la suma de cuadrados de la regresión es suficiente para garantizar la conveniencia estadística acerca del uso del regresor adicional en el modelo. La inclusión de más regresores de lo estrictamente necesario, aumenta la varianza de los coeficientes betas que más adelante se estimarán. Por tanto, se debe tener cuidado de incluir sólo regresores que realmente tengan significancia explicativa sobre la variabilidad de la variable dependiente.

Durante la selección de variables de un modelo es útil contar con un procedimiento que pueda evitar un sobreajuste del modelo, esto es, agregar variables que son innecesarias. Para ello, existe el *coeficiente R^2 ajustado* el cual penaliza la adición de variables innecesaria y que en realidad no están siendo útiles para mejorar la capacidad predictiva del modelo. En general, R^2 aumenta siempre que se agrega un regresor al modelo, independientemente del tamaño de la contribución de esa variable. Por tanto, no es trivial determinar si un aumento en la R^2 realmente es significativo o no. Por tanto es mejor utilizar el estadístico R^2 ajustado.

El coeficiente R^2 ajustado se hace más pequeño a medida que también va disminuyendo el ratio de # observaciones por cada variable independiente. El coeficiente R^2 ajustado es particularmente útil para comparar diferentes modelos de regresión con diferentes variables independientes o diferentes tamaños muestrales. El coeficiente de determinación ajustado (i.e. R^2 ajustado) compensa el sobre ajuste obtenido al incluir mayor cantidad de variables predictoras de las que son realmente necesarias para ajustar correctamente una muestral.

Fundamentos estadísticos del modelo de regresión

Mediante el método de mínimos cuadrados ordinarios se obtienen los mejores estimadores lineales insesgados para los coeficientes de B_i .

$$\sum_{i=1}^n (y_i - \bar{y}_i)^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SS_T = SS_M + SS_R$$

La expresión anterior se denomina “*Identidad fundamental del análisis de varianza del modelo de regresión*” y expresa lo siguiente:

La variabilidad total de las observaciones es igual a la variabilidad explicada por el modelo de regresión más la variación residual que queda sin explicar. Esto es igual a la suma de cuadrados del modelo de regresión más la suma de cuadrados residuales.

De la expresión anterior podemos derivar ahora el coeficiente de determinación R^2 como sigue:

$$R^2 = SS_M \div SS_T = 1 - SS_R \div SS_T$$

Donde: $0 \leq R^2 \leq 1$

El coeficiente de determinación R^2 mide la proporción de la variabilidad de la variable dependiente que es explicado por el conjunto de regresores incluidos en el modelo. Los valores de R^2 cercanos a la unidad implican que la mayor parte de la variabilidad en la variable dependiente está determinado por el modelo de regresión. El valor de R^2 siempre se incrementa cada vez que la cantidad de regresores p incluidos en el modelo también aumenta. Por tanto, si R^2_p es el coeficiente R^2 al utilizar una cantidad de p regresores, entonces podemos afirmar que R^2_p se incrementa al aumenta el valor de p y es máximo cuando $p = k$, es decir, cuando se incluyen todos los regresores disponibles en el modelo. Si se toma este criterio, podemos entonces intentar encontrar el valor óptimo de p de tal manera que se busque incluir solo aquel grupo de regresores que realmente agregue un valor significativo en el modelo hasta el punto en donde una variable adicional ya no es útil porque produce un incremento marginal en el coeficiente R^2 . Este efecto relacionado con los rendimientos decrecientes en el coeficiente R^2 se puede verificar en la siguiente gráfica:

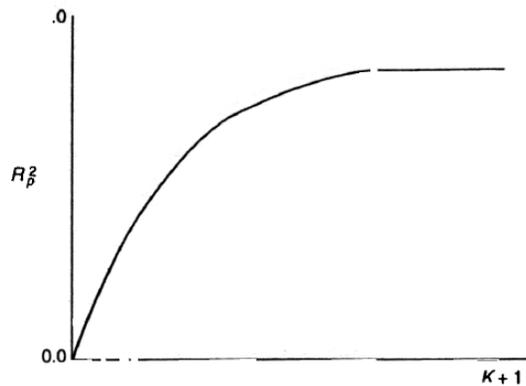


Figura 2.1: Gráfica del coeficiente R^2_p en función del número de regresores p .

Encontrar el valor óptimo de p implica identificar entonces cuales son los regresores que realmente agregan un valor significativo en el modelo. Se puede verificar entonces que existe una cantidad extremadamente grande de combinaciones para ser probados. La cantidad de combinaciones es igual a C^k_{p-1} lo cual equivale a k combinaciones por cada posible modelo que incluya un subconjunto de p regresores. Del coeficiente R^2 podemos derivar ahora R^2_{Ajust} para ello se calcula como sigue:

$$R^2_{Ajust} = 1 - \frac{SS_R / (n - k)}{SS_T / (n - 1)} = 1 - \frac{CM_R}{CM_T}$$

Donde:

CM_R = cuadrado medio de Residuales

CM_T = cuadrado medio de los Totales

Como se puede apreciar arriba, R^2_{Ajust} se comporta de una manera muy diferente a R^2 . En el último caso se puede afirmar que cada vez que se agrega un regresor al modelo, el valor de R^2 siempre se incrementa cada vez que la cantidad de regresores k incluidos en el modelo también aumenta. Sin embargo esto no significa que estos regresores incluidos tengan realmente un valor estadístico o práctico para el modelo. En el caso de R^2_{Ajust} es distinto ya que el denominador se mantiene constante para cualquier cantidad de regresores k . Así entonces el valor de R^2_{Ajust} solo aumentará cuando al agregar un regresor al modelo, si y solo si esa inclusión reduce el cuadrado medio residual.

En un análisis de regresión, al comparar diversos modelos de regresión rivales, es claro que el coeficiente R^2_{Ajust} ofrece ventajas en su interpretación en relación al coeficiente R^2 . Sin embargo la prueba máxima e indiscutible para identificar cuando un modelo de regresión es estadísticamente mejor que otro es la prueba de significancia F . La prueba estadística F se aplica a un modelo en su conjunto y tiene el objetivo de comprobar en el plano estadístico, si significativamente existe una relación lineal causa y efecto entre la variable dependiente y el conjunto de regresores $(X_1 \dots X_k)$

incluidos en el modelo incumbente que se está sometiendo a la prueba de hipótesis. En efecto, se trata de una prueba general o global de la adecuación del modelo en su conjunto. Por tanto las Hipótesis pertinentes para la aplicación de este procedimiento serían:

$$H_0: B_1 = B_2 = \dots = B_k = 0$$

$$H_1: B_j \neq 0 \text{ al menos para uno de los regresores "j".}$$

Es fácil verificar que si se rechaza H_0 y por tanto se acepta H_1 , esto significa que estamos afirmando que al menos uno de los regresores ($X_1 \dots X_k$) incluidos en el modelo sometido a la prueba de hipótesis contribuye al modelo y por tanto a la explicación de la variable dependiente en forma estadísticamente significativa. También es fácil comprobar que dicho procedimiento de prueba es una generalización del análisis de varianza visto en los capítulos anteriores. Es decir se sigue manteniendo la igualdad básica del análisis de varianza:

$$SS_T = SS_M + SS_R$$

Es decir, la variabilidad total es igual a la variabilidad explicada por el modelo de regresión más la variación residual que queda sin explicar. Dicha igualdad de varianzas también se aplica de manera similar para el análisis de los grados de libertad. Es decir los grados de libertad total, es igual a, los grados de libertad debidos al modelo de regresión más los grados de libertad del término de error.

- $GL_T = n - 1$: se pierde un grado de libertad debido a la aplicación de
$$\sum_{i=1}^n (y_i - \bar{y}_i)^2$$
- $GL_M = k$: se requiere el uso de tantos grados de libertad como cantidad de parámetros o regresores k se tengan que calcular.
- $GL_R = GL_T - GL_M$: los grados de libertad residuales es igual a los grados de libertad total menos los grados de libertad debidos al modelo de regresión. Se pierde un grado de libertad adicional debido a que se debe añadir el cálculo del coeficiente B_0 .
- Propiedad Aditiva: $GL_T = GL_M + GL_R \rightarrow n - 1 = k + (n - k - 1)$

A partir lo anterior, podemos calcular el estadístico **F** de la siguiente manera:

$$F_{Calc} = \frac{SS_M / k}{SS_R / (n - k - 1)} = \frac{CM_M}{CM_R}$$

$CM_M = \text{cuadrado medio del Modelo de Regresión}$

$CM_R = \text{cuadrado medio de los Residuales}$

Si el valor de F_{Calc} es suficientemente grande, al menos uno de los coeficientes B_j será diferente de cero. Se pueden comprobar algunas similitudes y diferencias entre el cálculo del coeficiente R^2_{Ajust} versus el coeficiente F_{Calc} . Ambos coeficientes consideran el uso de sus cuadrados medios. Sin embargo el coeficiente R^2_{Ajust} mide un cociente de calidad de las variaciones residuales entre las variaciones totales. A diferencia, el coeficiente F_{Calc} mide un cociente tipo índice de las variaciones del modelo de regresión (i.e. la variación explicada por el modelo) entre las variaciones residuales (i.e. la variación no explicada).

En un análisis de regresión, además de obtener los estimados de los coeficientes de regresión también es importante identificar la varianza de dichos estimadores. De hecho la varianza de dichos estimadores tiene dos importantes propósitos:

1. La prueba de hipótesis para la significancia estadística de dichos estimadores. Esto tiene el objetivo de comprobar en el plano estadístico, si significativamente existe una relación lineal causa y efecto entre la variable dependiente y cualesquiera de las variables independientes ($X_1... X_k$) que se estén sometiendo a la prueba de hipótesis particular.
2. El cálculo de los intervalos de confianza para los estimadores.

Para una prueba de hipótesis típica acerca del estimador de un regresor tendríamos $H_0: B_j = 0$. Si $T_{Calc} > T_{Critica}$ entonces querría decir que no se debe eliminar el regreso X_j del modelo debido a su significancia estadística. El estadístico requerido para comprobar dicha hipótesis sería la prueba T_{Calc} del siguiente modo:

$$T_{Calc} = \frac{\hat{B}_j}{\sqrt{\sigma_{B_j}^2}}$$

En general, hay muchos paquetes de software para desarrollar un modelo de regresión. Sin embargo quisiéramos en este momento destacar la herramienta que provee Microsoft Excel para ello. Mencionamos entonces a la formula Matricial de Microsoft Excel denominada "ESTIMACION.LINEAL". Los parámetros de dicha fórmula son los siguientes:

$$\text{Rango.Celdas.Valor} = \text{Estimacion.lineal}(\text{Ry}, \text{Rx}, B_1, B_2)$$

Donde:

- Ry = Rango para los valores conocidos de la variable dependiente Y .
- Rx = valores conocidos de la(s) variable(s) independiente(s) X .
- B_1 = Verdadero / Falso (modelo con/sin Coeficiente Constante B_0)
- B_2 = Verdadero / Falso (resultados con/sin información estadística)

Los resultados matriciales de la formula anterior al ejecutarse en Microsoft Excel genera una tabla del siguiente tipo:

	A	B	C	D	E	F
1	m_n	m_{n-1}	...	m_2	m_1	b
2	se_n	se_{n-1}	...	se_2	se_1	se_b
3	r_2	se_y				
4	F	d_f				
5	ssreg	ssresid				

- Fila # 1 ($m_n, m_{n-1}, \dots, m_1, b$): estas son las celdas en donde se calculan los coeficientes de los regresores. Nótese que va en orden descendente desde B_n hasta B_0 .
- Fila # 2: éstas son las celdas en donde se calculan los errores estándar para los estimadores de los regresores. Nótese también que estos valores corresponden directamente al denominador utilizado la fórmula para $T_{calc} = \frac{\hat{B}_j}{\sqrt{\sigma_{B_j}^2}}$
- Fila # 3, Columna # 1: esta celda es para calcular el coeficiente de determinación R^2 . Aquí se mide el ajuste de los valores Y estimados versus los valores de Y reales. Si R^2 es 1, significa que hay una correlación perfecta en la muestra, es decir, no hay diferencia entre el valor de Y estimado y el valor Y real. En el otro extremo, si el coeficiente de determinación R^2 es 0, significa que la ecuación de regresión no es útil para predecir un valor Y .
- Fila # 3, Columna # 2: en esta celda se calcula el error estándar para la estimación de Y .
- Fila # 4, Columna # 1: en esta celda se calcula el estadístico F_{calc} . La prueba F es para determinar si la relación observada entre las variables dependientes e independientes se produce por el azar o por la casualidad de las variables independientes.
- Fila # 4, Columna # 2: en esta celda se calcula los grados de libertad correspondientes a la suma de cuadrados de los residuales.
- Fila # 5, Columna # 1: en esta celda se calcula la suma de cuadrados del modelo de regresión. Esto es la variabilidad explicada por el modelo.
- Fila # 5, Columna # 2: en esta celda se calcula la suma de cuadrados residuales. Esto es la variabilidad no explicada por el modelo.

Es importante mencionar que para poder ingresar formulas de tipo matricial en Microsoft Excel es necesario considerar dos puntos:

1. Se debe seleccionar un rango de celdas que sea acorde con el tamaño de la matriz de resultados que se busca obtener. Para nuestra fórmula de "Estimacion.Lineal", se puede verificar que el rango

siempre será igual a 5 FILAS y a $K+1$ COLUMNAS. Donde k es igual a la cantidad de regresores en el modelo de regresión.

2. Se debe ingresar la formula a través de la secuencia de teclas CTRL + MAYUS + ENTER. Esto es a diferencia de lo convencional que es a través del ENTER.

Suponga que se compara una $F_{Calc} = 459.75$ con respecto a una $F_{Crítica} = 4.53$ para un $\alpha = 5\%$. Puesto que $F_{Calc} \gg F_{Crítica}$ tenemos entonces que es extremadamente improbable que un valor F tan elevado se produzca por azar, sino más bien debido a la explicación del modelo. En otras palabras, la hipótesis nula de que no hay relación entre las variables independientes y la variable dependiente debemos rechazarla cuando $F_{Calc} > F_{Crítica}$. Se dispone de dos formulas en Microsoft Excel para obtener la probabilidad de que se produzca un valor F_{Calc} o un valor de T_{Calc} al azar. Dichas formulas son: (1) *DISTR.F* y (2) *DISTR.T*

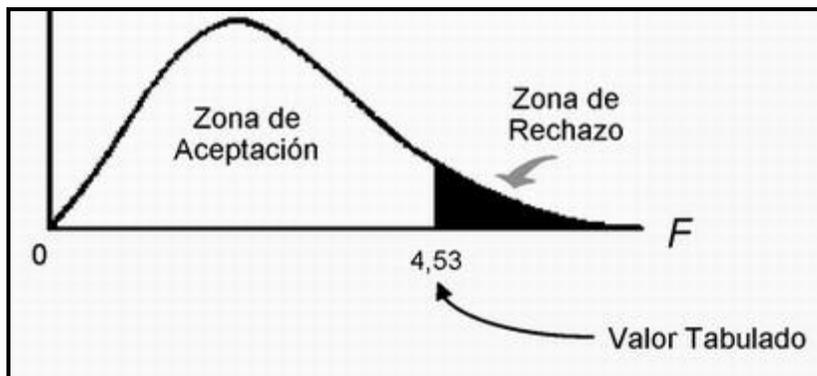


Figura 2.2: Zona de aceptación y rechazo del estadístico F.

Ya hemos anteriormente comentado que el coeficiente R^2_{Ajust} es mejor indicador que R^2 con la finalidad de comparar la calidad de ajuste en modelos rivales. Más específicamente tenemos que el numerador del coeficiente R^2_{Ajust} corresponde al término CM_R (i.e. **cuadrado medio de Residuales**) el cual a su vez está afectado por la cantidad de regresores en el modelo. Así entonces podemos decir que el término CM_R es un indicador útil para comparar la calidad del ajuste entre modelos rivales. El que tenga menor valor residual CM_R será mejor modelo. Esto anterior no es la excepción al momento de comparar un modelo con ordenada al origen (i.e. coeficiente constante) versus uno sin ordenada al origen.

Significación del modelo de regresión

Aunque la medición de **R²** no puede disminuir si se agrega al modelo una nueva variable predictora, eso no quiere decir necesariamente que el nuevo modelo sea mejor que el anterior. A menos que la suma de cuadrados residuales en el nuevo modelo se reduzca en una cantidad igual al error cuadrático medio original. Para medir lo anterior se utiliza la prueba del estadístico **F** como sigue:

$$F_{Calc} = \frac{\text{Suma Cuadrados de la Regresión} / \text{Grados de libertad de la Regresión}}{\text{Suma de Cuadrados Total} / \text{Grados de libertad de los Residuales}}$$

Donde:

Grados de libertad del Modelo = # de Coeficientes Estimados (incluyendo la constante) = *k*

Grados de libertad Residuales = Tamaño Muestral - Grados de libertad del Modelo = *n - k*

$$F_{Calc} = \frac{SCR / k}{SCT / n - k}$$

En un modelo de regresión, si el valor calculado de **F** es lo suficientemente grande, lo anterior significa que al menos uno de los coeficiente del modelo (i.e. **B_k**) es diferente de cero y estadísticamente significativo. Más aún, si el cociente de la varianza explicada (i.e. SCR) respecto a la varianza total, es decir la varianza alrededor de la media (i.e. SCT) es un valor alto; entonces lo anterior quiere decir que el modelo de regresión tiene un resultado significativo respecto a la explicación de la variable dependiente.

Por ejemplo, si el ratio **F_{Calc}** de un modelo fuera igual a 18.0, lo anterior significaría que podemos rechazar la hipótesis nula de que los efectos en la variable dependiente sean producidos al azar en lugar de estar explicados por el conjunto de variables independientes incluidas en el modelo teórico. En otras palabras, a través del modelo teórico podemos explicar 18 veces más variación que cuando utilizamos la media. Lo anterior no es muy probable que ocurra por azar ya que dicha probabilidad sería menor al 5% o al 1% de las veces según el caso.

Una vez determinado que al menos uno de los regresores tiene impacto significativo en la explicación de la variable dependiente, entonces la pregunta lógica es ¿cuál(es) de ellos sirve(n)? La prueba de significación

estadística para los coeficientes estimados de un análisis de regresión es indispensable cuando el modelo ajustado se basa en una muestra de la población y no es un censo. Para cada muestra extraída de una población, es posible que se obtengan valores de estimación diferentes en los coeficientes. Por tanto, el objetivo es que el modelo de regresión sea útil como predictor no sólo para la muestra incumbente sino en general de toda la población. Para muestras pequeñas, los coeficientes estimados variarán ampliamente de muestra a muestra. Pero a medida que el tamaño muestral aumenta, las muestras se hacen más representativas de la población y la variación en los coeficientes estimados se espera que sean más pequeñas.

La variación resultante de los coeficientes estimados se denomina el error estándar de los coeficientes. La significación estadística de los coeficientes de regresión proporciona una estimación probabilística respecto a si los coeficientes estimados a lo largo de un gran número de muestras de un cierto tamaño serán diferentes de cero. Muestras grandes no garantizan que los coeficientes sean diferentes a cero, pero si aseguran que la prueba sea más precisa.

En general, es inapropiado comparar en forma directa los coeficientes de regresión, dado que la magnitud de cada uno de los coeficientes B_j refleja la escala de medida del regresor X_j . Por tanto es incorrecto usar la magnitud de las B_j como medida de la importancia relativa del regresor X_j . Para este tipo de análisis comparativo lo correcto es utilizar los regresores y la variable de respuesta estandarizada para que así se obtengan coeficientes de regresión adimensionales.

A esos coeficientes adimensionales se les suele llamar *coeficientes estandarizados de regresión*. Añadiendo en este tema, si los datos utilizados para desarrollar el modelo no incluyeran observaciones para valores de cero en las variables independientes, entonces querría decir que el coeficiente constante B_0 está fuera de los datos y actúa sólo como un sobre ajuste en el modelo. Por tanto, no sería necesario probar que dicho coeficiente B_0 fuera diferente de cero para el modelo de regresión y por tanto puede ser excluido a priori.

Los contrastes de significación de los coeficientes de regresión proporcionan una valoración empírica de su verdadero impacto. Lo anterior contribuye a asegurar que los impactos medidos en los coeficientes de regresión puedan ser generalizables a otras muestras de la misma población. No debemos olvidar que los coeficientes estimados en el análisis de regresión son específicos para la muestra utilizada en la estimación. Sin embargo dichos coeficientes pueden variar de una muestra a otra. Esto indica la necesidad validar el análisis de regresión sobre muestra(s) diferente(s). Aunque es de esperar que los coeficientes cambien de muestra a muestra, la hipótesis a probar es que generalmente estos coeficientes mantienen la relación y su magnitud de impacto en otras muestras. Así entonces se logra demostrar que el modelo es generalizable para cualquier muestra obtenida de la misma población.

La creación de un intervalo de confianza para el modelo de predicción es importante. Dado que no podemos conseguir predicciones perfectas de la variable dependiente, debemos estimar el rango de valores que la variable a predecir puede tomar y no tan solo en una estimación puntual. A partir de la estimación puntual, podemos calcular entonces el rango de los valores a predecir basándose en una medida de los errores de predicción de la variable dependiente. Por ejemplo, añadiendo 1.96 desviaciones estándar con respecto a la media de un coeficiente beta, se define entonces un rango que incluye el 95% del intervalo para el coeficiente de una variable y con ello establecer los límites superiores e inferiores de la estimación.

Validación del modelo

Antes de que un modelo de regresión se utilice para lo cual fue desarrollado se debe hacer cierta evaluación de su validez. Hay una diferencia entre la comprobación de la adecuación del modelo y su validación.

- La comprobación de la adecuación del modelo incluye el análisis de residuales, la prueba de falta de ajuste y la identificación de observaciones atípicas.
- Por otro lado, la validación del modelo se concentra en la determinación de si el modelo funcionará bien en su ambiente pretendido de operación.

En general, no hay seguridad de que el modelo que proporciona el mejor ajuste a los datos observados sea un buen modelo para la predicción. Pueden existir factores influyentes desconocidos durante la etapa de diseño del modelo que afectarán en forma importante las nuevas observaciones, haciendo inútiles las predicciones. Además, los usuarios potenciales del modelo podrían llegar a conclusiones equivocadas en la interpretación de los signos y las magnitudes de los coeficientes. La validación del modelo permite contar con una seguridad, tanto para el desarrollador como para el usuario. Para validar un modelo de regresión existen tres clases de procedimientos:

1. Análisis de los coeficientes del modelo y de los valores estimados. Esto es revisar los resultados empíricos versus la experiencia o la teoría existente. Se deben analizar los coeficientes del modelo de regresión para determinar si son estables y si son razonables sus signos y magnitudes. Los coeficientes que tengan signos inesperados, o que tengan valores absolutos demasiado grandes, suelen indicar que el modelo es inadecuado ya sea porque faltan incluir algunos regresores o debido a la presencia de multicolinealidad.
2. Revisión de los factores de inflación de varianza (i.e. análisis de multicolinealidad). Si cualquier factor de inflación de varianza es mayor que 10, es muy probable que el coeficiente correspondiente este mal estimado o sea inestable debido a la alta correlación entre los regresores.

3. Recolección de nuevos datos con los cuales identificar el desempeño del modelo. El método más eficaz para validar un modelo de regresión con respecto a su eficiencia de predicción es reunir nuevos datos y comparar en forma directa las predicciones del modelo con esos nuevos datos. Si el modelo produce buenas estimaciones para los nuevos datos, el usuario tendrá más confianza en el modelo. Se recomienda contar con al menos 20 observaciones nuevas para tener una evaluación fiable de la eficiencia de predicción del modelo. En casos en donde se hayan desarrollado, con los mismos datos, dos o más modelos de regresión alternativos (i.e. modelos rivales), la comparación de la eficiencia de predicción de ellos con los datos nuevos puede ser la base de la selección del modelo final.
4. División de los datos, es decir, apartar algunos de los datos originales y usar esas observaciones para investigar el desempeño del modelo en predicción. Ciertamente, en muchos casos no es posible reunir nuevos datos para fines de validación. Un procedimiento razonable es dividir los datos disponibles en dos partes, el primer grupo serán los datos de estimación y el segundo será el de los datos de predicción. Los datos de estimación se usan para formar el modelo de regresión y a continuación los datos de predicción se usan para estudiar la capacidad predictiva del modelo. Se puede intercambiar el procedimiento y con ello obtener un segundo modelo rival. Con ello se podría entonces aplicar una validación para determinar cuál de los dos modelos puede predecir mejor al otro.

En general para los dos últimos procedimientos de validación, esto puede ser dividiendo la muestra en dos sub-muestras. La primera submuestra para estimar los parámetros del modelo y la segunda submuestra para estimar la precisión predictiva. El uso final pretendido del modelo determina cuál es la metodología adecuada de validación. Sin embargo, como el desarrollador no controla el uso del modelo, se recomienda la aplicación de todas las validaciones posibles especialmente aquella en la cual se prueban los modelos rivales con datos nuevos.

Identificación de puntos atípicos e influyentes

La disposición de los valores de las variables independientes tiene un papel importante en el ajuste por mínimos cuadrados. Si bien todos los puntos tienen una ponderación igual para el cálculo de los coeficientes, no obstante éste está mucho más influenciado por los valores atípicos presente las variables independientes. Un valor atípico es una observación extrema. En general, un residual cuyo valor absoluto sea bastante mayor que los demás, digamos de tres a cuatro desviaciones estándar respecto a la media, esto es indicativo de la presencia de valores atípicos. Los residuales estandarizados es la mejor forma para identificar los puntos atípicos potenciales. Los valores atípicos se deben investigar con la finalidad de encontrar una razón de su comportamiento extraordinario. Un punto de influencia, o valor

influyente, es aquel que tiene un impacto notable sobre los coeficientes del modelo, porque "jala" al modelo de regresión en su dirección. Las observaciones influyentes en un sentido amplio son todas aquellas observaciones que tienen un efecto desproporcionado sobre los resultados de la regresión. En general, para la detección de puntos influyentes se utiliza el análisis de residuos identificando aquellas mediciones con residuos estandarizados mayores a 2 desviaciones estándar.

Suponga un caso con dos observaciones influyentes en las que estas únicas dos definen casi por completo toda la relación observada, ya que sin ellas no surge ninguna pauta real en todo el resto de los datos. Esto significa que un pequeño subconjunto de los datos puede ejercer una influencia desproporcionada sobre los coeficientes del modelo. En un caso extremo, los estimados de los coeficientes pueden llegar a depender más del pequeño subconjunto influyente de puntos que de la mayor parte de los datos. Esto es indeseable; ya que se prefiere un modelo de regresión que sea representativo de la mayor parte de las observaciones en la muestra y no que sea un resultado de solo un pequeño subconjunto de las mediciones. El caso en el cual un porcentaje pequeño de los datos tiene un gran impacto sobre el modelo, no es de aceptación general para la definición de un modelo de regresión. Los valores atípicos pueden tener efectos de moderados a graves sobre el modelo de regresión. Por tanto, es importante identificar estos puntos influyentes y evaluar su impacto sobre el modelo.

Es posible el uso del análisis cluster para determinar el conjunto de observaciones influyentes en un modelo de regresión. El análisis cluster es una técnica para identificar grupos de observaciones semejantes. El procedimiento consiste en definir una medida de similaridad entre las observaciones para después clasificar las observaciones en grupos con base a sus similaridades.

El objetivo del análisis cluster es agrupar n datos al menos en dos grupos. El primer grupo será de tamaño m el cual estará compuesto de observaciones típicas y el segundo grupo de tamaño k estará compuesto por datos potencialmente influyentes.

El análisis de puntos influyentes es importante en la medida en la que la generalización de los resultados (i.e. coeficiente de las variables) puede verse alterado debido a un número relativamente pequeño de observaciones. Otros casos corresponden al registro incorrecto de los datos y otros más a la falla del instrumento de medición.

Los datos influyentes o atípicos se presentan debido a alguna de las tres condiciones siguientes:

1. Un error de medición en las observaciones.
2. Una observación excepcional válida aunque solo explicable por una situación extraordinaria. Se debe a un evento desacostumbrado pero explicable.
3. Una observación excepcional sin una explicación evidente.

Para el caso de un error de medición en la observación (i.e. caso 1) la solución puede ser simplemente eliminar la observación. Lo mismo sucede con observaciones válidas pero excepcionales (i.e. caso 2) a menos que las variables que reflejan la situación extraordinaria se incluyan en la ecuación de regresión. En cualquier caso, el valor atípico se debería corregir (si es posible) o eliminar del conjunto de datos. Si esos puntos atípicos son en realidad mediciones incorrectas, se deberían entonces eliminar de la muestra. Por el contrario, si las mediciones son correctas, es entonces necesario identificar la naturaleza de su explicación y por tanto incorporar las variables predictoras que pudieran estar faltando. Es claro que el eliminar valores atípicos es conveniente desde la perspectiva del ajuste de los mínimos cuadrados, sin embargo, se debe contar con una fuerte evidencia no estadística (i.e. desde el fundamento teórico) de que el valor atípico es innecesario para los efectos del estudio, para luego entonces descartarlo.

En general, los datos atípicos son observaciones que tienen grandes valores residuales. Si acaso esos puntos atípicos son en realidad mediciones "incorrectas", se deberían entonces eliminar de la muestra. Por el contrario, si las mediciones son correctas, es entonces necesario identificar la naturaleza de su explicación y por tanto incorporar las variables predictoras que pudieran estar faltando. A veces el valor atípico es una observación extraordinaria, pero perfectamente factible.

Las observaciones influyentes no son necesariamente "incorrectas" tal que tengan que ser omitidas. Puede ser peligroso eliminar estos puntos sólo para "mejorar el ajuste del modelo". Por ejemplo, si el valor atípico es un punto de respuesta especialmente interesante (por ejemplo, bajo costo o alto rendimiento), sería muy importante conocer el valor de los regresores correspondientes cuando se observó esta respuesta. En general, el análisis de identificación de los valores atípicos da como resultado mejoras en el modelo y nuevos conocimientos acerca de los factores cuyo efecto sobre la respuesta se desconocía antes. Casos como éste requieren con frecuencia de la estimación de los parámetros del modelo con alguna técnica que tenga menor influencia de este tipo de puntos que el que se obtiene mediante el método de mínimos cuadrados ordinarios.

En términos prácticos, el efecto de los valores atípicos sobre el modelo de regresión se puede comprobar con facilidad eliminándolos y volviendo a ajustar la ecuación de regresión. Hecho lo anterior, se podrá identificar que tan sensibles son los valores de los coeficientes de regresión o de los estadísticos t , F , R^2 a los valores atípicos que fueron eliminados. De hecho, para la prueba de falta de ajuste de un modelo de regresión es conveniente contar con varias observaciones, (i.e. replicas), de la respuesta de la variable dependiente para los valores de las variables independientes. Estas replicas deben ser verdaderas y no sólo mediciones copiadas.

El diagnóstico de puntos influencia es parte fundamental de las herramientas de quien construye modelos. El objetivo es ofrecer un mejor entendimiento de los datos e identificar cuáles observaciones merecen un

análisis por separado. Qué tan importante es el análisis de los puntos de influencia, esto dependerá de la cantidad de estos que hayan sido identificados así como de su impacto sobre el modelo. Lo anterior es particularmente cierto si se obtiene un resultado inesperado y se sospecha que pueda deberse a este tipo de observaciones.

La cuestión de si se deben desechar las observaciones influyentes es muy similar a la pregunta acerca de si es correcto descartar los valores atípicos. En general, si se confirma que hay un error de medición, o si el punto de la muestra realmente es inválido o no es parte de la población de interés, pues entonces será apropiado descartar la observación de la muestra. La decisión entre eliminar una observación y retenerla conlleva a la necesidad de considerar una técnica de estimación que no sea tan sensible a los puntos influyentes como lo es el método de mínimos cuadrados ordinarios.

Las técnicas de estimación basadas en regresión robusta, tienen el objetivo de disminuir la importancia relativa de las observaciones en proporción con la magnitud o influencia residual. Lo anterior quiere decir que una observación muy influyente recibirá menos peso comparativamente que en el caso cuando se aplicara el método de MCO. Se han desarrollado métodos de Regresión Robusta para tratar específicamente el impacto de los datos atípicos.

La multicolinealidad

El uso e interpretación de un modelo de regresión depende de los estimados de los coeficientes individuales de regresión. Si no existe relación lineal entre los regresores, se dice que éstos son *regresores ortogonales*.

Cuando los regresores son ortogonales la interpretación de los coeficientes de un modelo se puede hacer con relativa facilidad. Desafortunadamente, en la mayor parte de los casos los regresores no son completamente ortogonales. La falta de ortogonalidad no es grave, sin embargo, el otro extremo sí lo es. En algunos casos los regresores tienen una relación lineal casi perfecta y en esos casos las inferencias basadas en el modelo de regresión pueden ser erróneas. Cuando se presenta una fuerte dependencia lineal entre los regresores, se dice que existe el problema de *multicolinealidad*. Lo anterior provoca que la inclusión de una variable independiente, mejore la capacidad de predicción no sólo sobre la variable dependiente sino también sobre el nivel de correlación obtenido en el resto de las variables independientes.

En general existen dos fuentes principales de multicolinealidad:

1. El método de recolección de datos empleado. Se ha sugerido la recolección de datos adicionales como el mejor método para combatir la multicolinealidad. Los datos adicionales se deben reunir en una forma diseñada para de manera efectiva y con el menor esfuerzo eliminar la multicolinealidad. Lo anterior significa identificar aquellos valores que estén ausentes del espacio de medición. Por ejemplo, ir a medir aquellos puntos donde la X_1 sea pequeña y la X_2 sea

grande o viceversa. Desafortunadamente no siempre es posible coleccionar más datos ya sea por restricciones económicas o porque el proceso que se estudia ya no está disponible para su muestreo. Aun cuando haya más datos disponibles, puede ser inadecuado usarlos ya que la medición de las variables regresoras puede ir mucho más allá de la región de interés en la investigación. Es justo decir que la recolección de datos adicionales no es una solución del todo viable para el problema de multicolinealidad ya que los nuevos puntos de datos pueden resultar atípicos en el proceso que se estudia y por tanto influir negativamente en la capacidad predictiva del modelo.

2. Un modelo sobre definido. Un *modelo sobre definido* es aquel en el cual se tienen más variables regresoras que observaciones. Para eliminar la multicolinealidad en este contexto lo recomendable es redefinir el modelo en términos de un conjunto menor de regresores. Para ello se recomienda realizar a manera de pre-proceso un Análisis factorial tipo componentes principales con la finalidad de determinar cuáles regresores se deben omitir del modelo.

En general un modelo puede parecer adecuado para ser usado en estimaciones del tipo interpolación, esto es dentro del intervalo de donde se tomaron los datos originales. Sin embargo, cuando se trata de hacer extrapolaciones el modelo no funciona igual. Una causa de que esto pueda ocurrir en mayor o menor medida es la multicolinealidad. Si se requiere que un modelo extrapole correctamente, se requiere entonces obtener buenos estimados de los coeficientes individuales. En general, cuando se tiene presencia de multicolinealidad, los estimados de los coeficientes de regresión determinados por mínimos cuadrados ordinarios pueden ser muy malos.

En general, la multicolinealidad ocurre cuando dos regresores muy correlacionados se usan al mismo tiempo en el modelo de regresión. En estos casos la re-especificación del modelo de regresión puede aminorar el impacto de la multicolinealidad. Un método muy usado para especificar nuevamente el modelo es la eliminación por completo de una de las variables que esté ocasionando la multicolinealidad.

En efecto y aunque paradójico, la eliminación de una variable puede ser efectivo para reducir la multicolinealidad pero puede dañar el poder predictivo del modelo. Se debe tener cuidado al seleccionar la variable a omitir, no hay seguridad de que el modelo final muestre menor grado de multicolinealidad que la que había en los datos originales. Otro método para especificar nuevamente el modelo es redefinir los regresores. Por ejemplo, si X_1 , X_2 y X_3 son casi linealmente dependientes, se puede aplicar una función tal como $X_T = X_1 * X_2 * X_3$ o también $X_T = (X_1 + X_2) / X_3$; lo anterior con la finalidad de preservar el contenido de información de los regresores originales.

Índice del capítulo

- análisis cluster*, 38
- análisis de regresión múltiple*, 38, 48
- análisis de varianza*, 39
- análisis discriminante*, 38
- análisis factorial*, 38
- análisis multivariable de la varianza*, 39
- ANOVA, 39
- coeficiente de correlación*, 51
- coeficiente de determinación*, 51
- coeficiente R2 ajustado*, 52
- coeficientes estandarizados de regresión*, 60
- combinación de tratamientos*, 49
- cuadrado medio*, 42
- diseño de experimentos*, 49
- diseño experimental*, 49
- ecuaciones estructurales*, 38
- grados de libertad*, 42
- MANOVA, 39
- MCO, 50
- mínimos cuadrados ordinarios*, 50
- modelo de regresión lineal múltiple*, 47
- modelo sobre definido*, 66
- modelos de regresión discreta*, 38
- multicolinealidad*, 47, 65
- niveles*, 49
- potencia de la regresión múltiple*, 52
- predictores*, 48
- r*, 51
- R2*, 51
- regresores ortogonales*, 65
- SCE, 51
- SCR, 51
- suma de cuadrados*, 40
- suma de cuadrados de la regresión*, 51
- suma de cuadrados de los errores*, 51
- Técnicas para el análisis de dependencia*, 38
- Técnicas para el análisis de Interdependencia*, 38
- variables dependientes*, 38
- variables independientes*, 38

Capítulo 3:

Las herramientas estadísticas para la investigación

Contenido

Herramientas estadísticas para la investigación	70
Introducción	70
Principales errores en la utilización de las herramientas estadísticas.	71
Solución Propuesta: Estrategia general de aplicación de las herramientas estadísticas.	73
Modelo Propuesto: Metodología de aplicación de las herramientas.	76
Estudio de los conjuntos de observaciones.	77
La distribución Normal de un conjunto de observaciones.	80
Otras distribuciones estadísticas de datos.....	82
Tamaño de las muestras.	83
Diseño de instrumentos de recolección de datos.....	85
Confiabilidad del Instrumento por medio del Coeficiente Alfa de Cronbach.	92
Relaciones entre variables: Regresión Lineal Simple y Múltiple.	98
Análisis para identificación de clusters o subconjuntos de elementos relacionados dentro de una población.	107
Análisis de datos para comprobación de comportamiento normal prueba de Kolmogorov-Smirnov.....	109
Comparación de dos conjuntos de observaciones: pruebas t-Student y Mann-Whitney.....	111
Comparación de más de dos conjuntos de observaciones. ANOVA.....	120
Conclusiones.	125
Referencias	126
Índice del capítulo	128

Herramientas estadísticas para la investigación

En este capítulo abordaremos la estrategia general para la utilización de las herramientas estadísticas necesarias para la investigación en las ciencias sociales y administrativas. La cuantificación y la comparación de conjuntos de datos son una tarea constante y agotadora pero sin duda una de las más relevantes en cualquier estudio. El fundamento de las conclusiones se basa en el hecho de que podemos demostrar por medio de la observación y la medición directa los conceptos que se expresan y concluyen (Kish, L. 1995).

Es indispensable el manejo adecuado de las herramientas para que las conclusiones tengan validez y la investigación sea reconocida por su metodología y aportación científica. El avance de la ciencia nos ha dotado de técnicas, herramientas e instrumentos que por su abundancia y variedad amenazan con confundir al investigador, por lo que es de sumo cuidado cubrir a detalle este capítulo para poder identificar la naturaleza de cualquier estudio y posteriormente la estrategia más recomendable a utilizar (Combe, S. A. 1979).

Una vez que el investigador cumple sus objetivos básicos surgen a menudo un sinnúmero de alternativas que pueden utilizarse para revisar por diferentes caminos los resultados obtenidos pero hay que recordar que lo más importante es cubrir lo básico de una manera simple, sencilla, eficiente y correcta.

Introducción

La investigación en las ciencias sociales muy frecuentemente requiere la utilización de herramientas estadísticas sofisticadas que permitan comparar, relacionar y contrastar mediciones que se presentan de los diferentes fenómenos observados (Cortés, F. y Rubalcava, 1991). Desde el planteamiento mismo de los problemas hasta la prueba de la hipótesis y las conclusiones, los estudios echan mano de las herramientas del cálculo estadístico. Para hablar de dichas herramientas es muy importante primeramente plantear el escenario y observar todo el contexto en el que se utilizan ya que frecuentemente se incurre en errores y conclusiones equivocadas debido al contexto en el que se aplican los instrumentos. Por ejemplo: Una técnica de comparación de medias de grupos que requiere que los datos tengan una distribución normal (Pértiga, S., Fernández, P. 2001) puede arrojar un resultado erróneo para grupos que no cumplen con un comportamiento normal, haciéndonos concluir que los grupos tienen medias iguales aunque no sea así, la técnica de comparación no detectará el hecho de que se haya cubierto el requisito de demostrar con una prueba que los datos tienen una distribución normal (Ferrand Aranaz, M. 2005). Errores de este tipo tienen que ver con el conocimiento de todo el contexto en el cual el investigador se desempeña.

Debido a lo anterior, procederemos al estudio del contexto global de aplicación de las técnicas de acuerdo a los diferentes tipos de estudio,

haciendo énfasis en las relaciones, requerimientos y condiciones que guardan las diferentes herramientas y cálculos.

Iniciaremos mencionando los principales problemas que se presentan y estableciendo las bases para el estudio de los conjuntos de observaciones, posteriormente la clasificación de los estudios de acuerdo a la cantidad de conjuntos de observaciones que un estudio utiliza, adicionalmente cubriremos las estrategias para utilizar las técnicas y cálculos estadísticos y finalmente una descripción de cada técnica y/o cálculo estadístico.

Principales errores en la utilización de las herramientas estadísticas.

Uno de los errores más comunes en la utilización de la estadística es asumir que los datos cumplen requerimientos de distribución, magnitud y otras propiedades que es necesario demostrar explícitamente mediante pruebas numéricas diseñadas para ello y que no se pueden pasar por alto, pues invalidan todo el estudio un riesgo definitivamente muy serio para cualquier tipo de investigación (Altman, D. A. 1997).

Otro problema que se presenta con frecuencia es la confusión de conceptos en el cálculo de los parámetros, por ejemplo confundir los parámetros estimados para una población con los parámetros medidos para una muestra que aunque análogos son muy distintos. Es este otro error de contexto que conduce a un error de cálculo muy notorio.

En otra modalidad podemos también encontrar estudios que utilizan técnicas muy sofisticadas cuando sus demostraciones pueden hacerse de una forma directa y simple calculando parámetros sencillos y fáciles de utilizar.

Es en parte la abundancia de técnicas y procedimientos lo que fomentan los tipos de errores que anteriormente se mencionan, otra razón para estos errores es la falta de pericia en el manejo de las herramientas estadísticas que como cualquier otra disciplina exige conocimiento y habilidad en el manejo de las facilidades que brinda.

A continuación mostramos la tabla 3.1 en la que se presenta un resumen de los principales tipos de errores y las causas que los originan:

Problema	Causa
Asumir que los datos cumplen requisitos que es necesario demostrar.	No considerar el contexto estadístico.
Confusión de conceptos por ejemplo confundir muestra con población.	Falta de conocimiento de las bases de la estadística.
Utilizar técnicas muy sofisticadas cuando solo se requiere medición de parámetros sencillos.	Falta de análisis de herramientas estadísticas sofisticadas generalmente por la sobreabundancia de las mismas.

Tabla 3.1 Problemas más frecuentes en el uso de las herramientas estadísticas

Para prevenir estos errores proponemos una forma simple de revisión que comprenda los siguientes puntos:

1. ¿El estudio realizado requiere investigar la distribución estadística de los datos que maneja?
2. ¿Hay necesidad de transformar o codificar los datos?
3. ¿Cuántas muestras se manejan y existe más de una población o universo que se va a estudiar?
4. ¿Existe más de una forma de demostrar la hipótesis?

Todo esto se simplifica si planteamos una metodología para llevar a cabo el plan del estudio estadístico de una investigación. Esta es la hipótesis de este capítulo y plantea un modelo en el que se basa la metodología mencionada. En resumen la hipótesis de este capítulo puede declararse como sigue:

Los errores estadísticos pueden prevenirse e incluso evitarse si se sigue una metodología de trabajo para seleccionar y aplicar las herramientas estadísticas.

Solución Propuesta: Estrategia general de aplicación de las herramientas estadísticas.

Para poder sugerir un camino a seguir para un estudio es necesario hacer una clasificación de los estudios o investigaciones de acuerdo al número de conjuntos de observaciones que involucran y a su vez de los tipos de hipótesis alternas que pueden surgir de un estudio o investigación (Combe, S. A. 1979).

Para establecer un marco de referencia primero explicaremos cómo se organiza la información de un estudio, lo cual se muestra en la tabla 3.2. Cada conjunto de observaciones contiene mediciones para una o más variables y cada variable puede formar un constructo, aunque un constructo puede ser formado por más de una variable. Los detalles se pueden observar en la tabla 3.2 que es un patrón muy común en cualquier investigación.

# de Obs.	Constructo 1		Constructo 2			Constructo 3
	Medición Variable 1	Medición Variable 2	Medición Variable 3	Medición Variable 4	Medición Variable 5	Medición Variable 6
1	Valor de la medición de la obs. # 1 para la variable 1			Valor de la medición de la obs. # 1 para la variable 4		
2						
3						
4						
5			Valor de la medición de la obs. # 5 para la variable 3			
6						
7						
8						

Tabla 3.2 Conjunto de Observaciones, Un constructo de puede formar de una o más variables. El contenido de la tabla son los valores de las mediciones para cada instancia observación-variable

El valor de una variable puede utilizarse de manera directa o bien sufrir una transformación antes de utilizarlo en la tabla, este pre proceso de los valores de las variables comúnmente se denomina estandarización y se lleva a cabo para facilitar los cálculos e identificar tendencias (Altman, D.G. y Bland, J.M.1996).

Proponemos también la clasificación de los estudios dependiendo del número de conjuntos de observaciones (esto es la cantidad de tablas como la 3.2) que involucran nos ayudará para establecer el tipo de hipótesis al-

terna que surge de cada estudio y prever las distintas técnicas que se pueden tomar.

Considerando lo anterior podemos clasificar los estudios en:

1. Estudios que involucran un solo conjunto de observaciones, estos estudios son tipo encuesta y cuentan con una medición a una muestra de una población.
2. Estudios que involucran dos conjuntos de observaciones, estos estudios son de tipo experimental o cuasi experimental y consisten de dos mediciones llevadas a cabo sobre una o dos poblaciones.
3. Estudios que involucran más de dos conjuntos de observaciones, estos estudios pueden ser también de tipo experimental o cuasi experimental, o también pueden ser de tipo encuesta con varias aplicaciones y están formados por más de dos mediciones llevadas a cabo sobre uno o varias poblaciones.

Para una definición clasificación formal de los estudios podemos recurrir a Sierra Bravo (1986) que explica detalladamente los modelos de Campbell y Stanley (1966).

Estos tres tipos de estudio originan en general tres estrategias distintas de solución.

Pasamos a establecer los tipos de hipótesis que pueden surgir de cada uno de ellos y tenemos que:

Para un estudio del tipo 1 que involucra un solo conjunto de observaciones la hipótesis pueden seguir alguno de los siguientes patrones:

1. Dos variables (o constructos) o más de un conjunto de observaciones están relacionadas en forma causal o asociativa. Denominaremos a este tipo de hipótesis como: 1.1 Regresión o Correlación
2. Existe uno o más subconjuntos de observaciones que se distinguen de los demás por sus valores en una población dada. Llamaremos a este tipo de hipótesis: 1.2 Clusters

Para un estudio del tipo 2 que involucra a dos conjuntos de observaciones la hipótesis pueden seguir alguno de los siguientes patrones:

1. El primer conjunto de observaciones muestra una media $>$ al segundo conjunto para una o varias variables específicas. Nombraremos este tipo de hipótesis como: 2.1 Comparación de Medias

2. El primer conjunto de observaciones muestra una media $<$ al segundo conjunto para una o varias variables específicas. Es el mismo tipo de hipótesis anterior: 2.1 Comparación de Medias
3. El primer conjunto de observaciones muestra una media \neq al segundo conjunto para una o varias variables específicas. Es el mismo tipo de hipótesis anterior: Tipo de hipótesis 2.1 Comparación de Medias
4. El primer conjunto de observaciones muestra una media $=$ al segundo conjunto para una o varias variables específicas. Es el mismo tipo de hipótesis anterior: Tipo de hipótesis 2.1 Comparación de Medias
5. El primer conjunto de observaciones tiene una distribución estadística \neq al segundo conjunto para una o varias variables específicas. Denominaremos a este tipo de hipótesis: 2.2 Comparación de Distribuciones Estadísticas

Para un estudio del tipo 3 que involucra más de dos conjuntos de observaciones la hipótesis pueden seguir alguno de los siguientes patrones:

1. Todas las medias de todos los conjuntos de observaciones son iguales para una o varias variables específicas. Nombraremos a este tipo de hipótesis 3.1 Análisis de Varianzas (ANOVA por sus siglas en inglés)
2. Todas las medias de todos los conjuntos de observaciones son diferentes entre sí para una o varias variables específicas. Es el mismo tipo de hipótesis anterior: tipo de hipótesis 3.1 Análisis de Varianzas (ANOVA por sus siglas en inglés)
3. Las medias de todos los conjuntos de observaciones muestran una serie de relaciones específicas entre sí para una o varias variables específicas. Es el mismo tipo de hipótesis anterior: tipo de hipótesis 3.1 Análisis de Varianzas (ANOVA por sus siglas en inglés)

La clasificación de estudios o investigaciones propuesta anteriormente abarca la mayoría de los estudios de las ciencias sociales y de ahí podemos extraer en resumen una manera muy simple de clasificar las hipótesis de este tipo de estudio, que quedan como se muestra en la tabla 3.3

Tipo De Estudio	Clasificación	Tipo de Hipótesis
Tipo 1 que involucra un solo conjunto de observaciones	1.1	Regresión o Correlación
	1.2	Clusters
Tipo 2 que involucra a dos conjuntos de observaciones	2.1	Comparación de Medias
	2.2	Comparación de Distribuciones Estadísticas
Tipo 3 que involucra más de dos conjuntos de observaciones	3.1	Análisis de Varianzas (ANOVA)

Tabla 3.3 Clasificación de las hipótesis de los distintos tipo de estudio

Modelo Propuesto: Metodología de aplicación de las herramientas.

Utilizando las clasificaciones anteriores proponemos la siguiente metodología para la aplicación de las técnicas estadísticas necesarias en los estudios:

1. Diseñar, validar perfeccionar y aplicar el instrumento que se utilizará en la recolección de datos.
2. Clasificar el estudio de acuerdo a su tipo 1, 2 ó 3, ver la sección anterior
3. Clasificar la hipótesis atendiendo a la tabla 3.3 y la sección anterior
4. Seleccionar la técnica estadística que se aplicara en el estudio de acuerdo a la tabla 3.3.

El modelo nos proporciona una guía rápida para establecer un rumbo en la utilización de las herramientas estadísticas marca la dirección hacia que técnicas son las clásicas que se utilizan para resolver el tipo de problema que estamos planteando, el investigador puede enriquecer los resultados del modelo completando y diversificando la comprobación de la hipótesis pero sin pasar por alto las demostraciones básicas de la misma ya que de ser así se corre el riesgo de cometer los errores mencionados anteriormente.

En lo que resta del capítulo procederemos a proporcionar el material indispensable para la aplicación de las herramientas, empezando por establecer las definiciones y conceptos básicos en la siguiente sección.

Estudio de los conjuntos de observaciones.

Definimos un conjunto de observaciones como un conjunto de mediciones de una o más variables que forman uno o más constructos como lo podemos observar en la tabla 3.2.

Partimos de las siguientes definiciones estadísticas:

Una *población* son todos los elementos u observaciones que forman un universo.

Una *muestra* es un subconjunto o parte de una población (Johnson, R.1996).

El diagrama de la figura 3.1 muestra la relación entre estas dos definiciones. Se puede tener más de una muestra para una población dada. Un conjunto de observaciones como el que se mostró en la tabla 3.2 puede pertenecer a una población o a una muestra, el investigador debe de especificar correctamente a quien pertenece, pues los cálculos estadísticos que se llevan a cabo sobre una población o una muestra son conceptualmente similares pero matemáticamente diferentes.

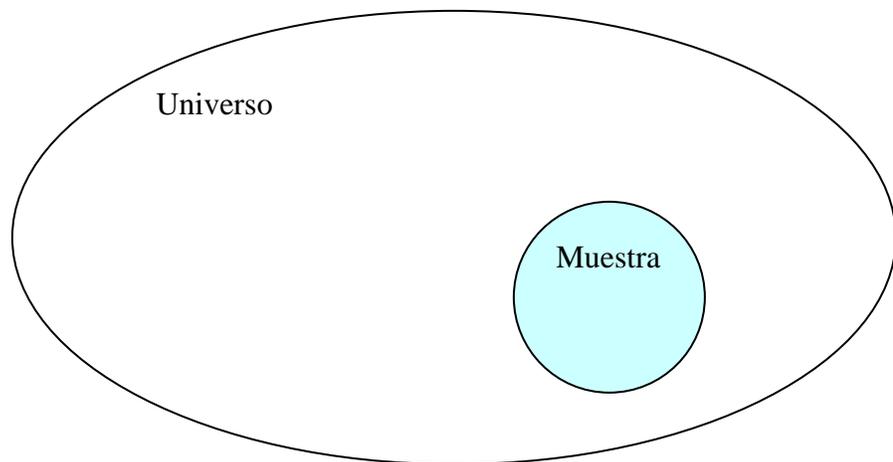


Figura 3.1 Universo y muestra de un conjunto de observaciones.

Los parámetros son valores que se obtienen de hacer ciertos cálculos previamente definidos con las mediciones de una población y nos ayudan a

establecer las características de la misma (Johnson, R.1996). Para definir un universo o población asumiendo que tenemos una distribución normal de los datos, la cual explicaremos más adelante, requerimos los siguientes parámetros:

- μ Media
- σ Desviación estándar
- Π Proporción
- N Tamaño o número de elementos

Los parámetros son desconocidos en la mayoría de los casos, especialmente cuando las poblaciones son muy grandes, por ejemplo los habitantes de una ciudad o de un país. Pueden ser estimados a través de los estadígrafos, los estadígrafos son las medidas equivalentes de las muestras. En una muestra tenemos los siguientes estadígrafos (Jonson, R.1996):

- \bar{X} Media
- S Desviación estándar
- P Proporción
- n Tamaño o número de elementos en la muestra

Si tomáramos varias muestras de la población y a cada una de ellas le calculáramos la media, dichas medias (una para cada muestra) tendrían una desviación estándar a la cual se le llama Error Estándar o SE por sus siglas en inglés para estimar el SE contamos con la siguiente ecuación:

$$SE = S/\sqrt{n} \quad \text{Ecuación 3.1}$$

Este cálculo es muy importante, ya que permite estimar la media μ de la población utilizando un intervalo de confianza, esto es una probabilidad de que la media tenga en realidad el valor calculado. Si el intervalo de confianza es Z la media poblacional μ viene dada por:

$$\mu = \bar{X} \pm Z*SE \quad \text{Ecuación 3.2}$$

Dónde SE está definido en la ecuación 3.1

El valor de la Z para diferentes % de certidumbre está dado en la tabla 3.4 (Lind, D.A., Mason, R.D. 2001)

% de certidumbre	Valor de Z
90	1.64
95	1.96
99	2.33

Tabla 3.4 valores de Z para diferentes porcentajes de certidumbre, estos valores se obtienen de las tablas de la distribución normal.

El % de certidumbre más utilizado en las ciencias sociales es el de 95% esto es que si deseamos estimar la media poblacional con un 95 % de probabilidades de estar en lo cierto la ecuación 3.2 queda como:

$$\mu = \bar{X} \pm 1.96 * SE \quad \text{Ecuación 3.3}$$

Y si deseamos utilizar la desviación estándar de la muestra queda de la siguiente manera:

$$\mu = \bar{X} \pm 1.96 * S / \sqrt{n} \quad \text{Ecuación. 1.4}$$

En realidad lo que las ecuaciones 3.3 y 3.4 estiman, no es un valor, sino un rango, ya que al tomar el signo + se obtiene el rango superior y al tomar el signo - el rango inferior, como se advierte en el siguiente ejemplo:

Si tenemos que de una población se ha tomado una muestra de 25 elementos y se ha medido la media siendo esta 80 con una desviación estándar de 7 podemos estimar que el rango en el cual se localiza la media real de la población es:

$$\mu_{\text{inferior}} = 80 + 1.96 * 7 / \sqrt{25} = 82.744$$

$$\mu_{\text{superior}} = 80 - 1.96 * 7 / \sqrt{25} = 77.256$$

Con un 95% de probabilidades de estar en lo cierto podemos decir que la media de la población está entre 77.256 y 82.744

Si deseamos que el % de probabilidades de estar en lo cierto sea de 99% tenemos que:

$$\mu_{\text{inferior}} = 80 + 2.33 * 7 / \sqrt{25} = 83.262$$

$$\mu_{\text{superior}} = 80 - 2.33 * 7 / \sqrt{25} = 76.738$$

El rango aumenta en ambos lados.

Por el contrario si deseamos que el % de probabilidades de estar en lo cierto sea de 90% tendremos:

$$\mu_{\text{inferior}} = 80 + 1.64 * 7 / \sqrt{25} = 82.296$$

$$\mu_{\text{superior}} = 80 - 1.64 * 7 / \sqrt{25} = 77.704$$

Y observamos que el rango se hace más estrecho.

Otra característica que define un universo o conjunto de observaciones es la distribución estadística que presentan sus datos que puede ser Normal o de otro tipo (Jonson, R.1996), los parámetros anteriores son representativos de una distribución normal.

La distribución Normal de un conjunto de observaciones.

La distribución normal fue detectada por primera vez por el francés Abraham de Moivre (1667-1754). Tiempo después, Carl Friedrich Gauss (1777-1855) elaboró la ecuación de la curva por lo tanto también se le conoce como "campana de Gauss" (Levi, R.I. y Rubin, D.S. 1996) . Una distribución normal queda totalmente definida por dos parámetros su media μ y su desviación estándar σ (Altman, D. A. 1997). Con esta notación, la densidad de curva es determinada por a la ecuación 3.5:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad -\infty < x < \infty$$

Ecuación 3.5

Los cálculos relacionados con una distribución normal se denominan cálculos de estadística paramétrica. Un ejemplo de esta distribución se muestra en la figura 3.2

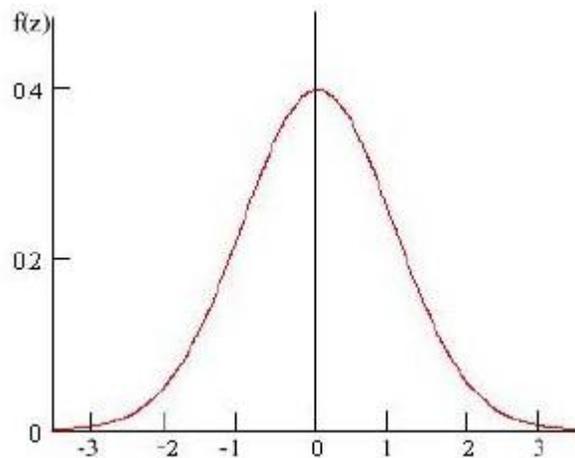


Figura 3.2 Densidad de distribución normal estándar

En el eje vertical se puede apreciar la probabilidad de que la medición de un elemento tome el valor del eje horizontal. Por ejemplo la probabilidad que el valor de la medición sea 0 es de aproximadamente 0.4 o sea el 40% de las veces.

Por otro lado tenemos que el área bajo la curva delimitada por dos valores x_1 y x_2 donde $x_1 < x_2$ es la probabilidad de que la variable x esté dentro ese rango o sea que se cumpla la condición: $x_1 \leq x \leq x_2$.

Existe una gran cantidad de fenómenos naturales que presentan el modelo normal, podemos citar algunos tales como (Altman, D. G. y Bland, J.M. 1995):

Propiedades morfológicas: tallas, pesos, envergaduras, diámetros, perímetros de personas, animales, plantas, etc.

Propiedades fisiológicas: efecto de una misma dosis de un fármaco, o de una misma cantidad de nutrientes.

Propiedades Sociológicas: venta de ciertos productos en una región, preferencia de una moda en un sector de la población, preferencias políticas en un sector socio-económico.

Características psicológicas: mediciones de habilidades, personalidad, cociente intelectual, grado de adaptación a un medio.

Errores y Defectos cometidos al medir o al fabricar ciertas piezas o al llevar a cabo cierta tarea.

Sin embargo para demostrar que un fenómeno presenta una distribución normal se cuenta con varias pruebas estadísticas (Pértiga, S. y Fernández, P. 2001), siendo una de las más utilizadas la prueba de Kolmogorov-Smirnov. Lo que en realidad hace esta prueba es medir la distancia entre la curva de distribución de los datos dados y una curva normal hipotética, al

igual que todas las prueba estadísticas la hipótesis nula se rechaza si el valor del estadístico obtenido no rebasa el requerido para aceptar la hipótesis alterna.

Otras prueba ampliamente utilizada para probar la distribución normal de los datos son las de Anderson Darling o Ryan Joiner también y la de Shapiro Wilk por mencionar algunas (Pértiga, S. y Fernández, P. 2001).

Cuándo se van a utilizar parámetros normales como la media, la varianza y la desviación estándar es indispensable probar primero la normalidad de los datos, de lo contrario será necesario utilizar estadística y comparaciones no paramétricas como lo veremos más adelante.

Otras distribuciones estadísticas de datos.

Naturalmente que los datos no solo pueden presentar una distribución normal, existen otras distribuciones que también se presentan en los procesos (Velasco, F. y Hernández, S.2007).

La distribución Poisson desarrollada por Dennis Poisson (1781-1840), es una distribución discreta que se aproxima a una curva de una distribución normal con un sesgo a la izquierda (Montgomery, D. C. y Runger, G.C. 1996), el ser discreta le permite medir números de eventos y es la segunda distribución más importante en estadística.

Los siguientes fenómenos presentan una distribución Poisson:

Llamadas telefónicas que llagan a un conmutador

La demanda de servicios para un servidor en internet

La demanda de servicios en una institución asistencial por parte de los pacientes

El arribo de clientes a una fila de espera de un servicio bancario

El arribo de los camiones y automóviles a la caseta de cobro

El número de accidentes en un cruce

Los fenómenos anteriores son todos discretos o sea que pueden ser medidos por una variable entera que presenta valores de 0, 1, 2, 3, 4, 5 pero no de fracciones.

La ecuación que define una distribución de la probabilidad de Poisson (Montgomery, D. y C. , Runger, G.C. 1996) es la 3.6.

$$P_{\mu}(x) = \frac{\mu^x}{x!} e^{-\mu} \quad \text{Ecuación 3.6}$$

Se aproxima a la normal cuando μ es grande y se cumple que:

$$P_{\mu}(x) \approx G_{x,\sigma}(x).$$

Con media $x = \mu$ y $\sigma = \sqrt{\mu}$

En ocasiones veremos la ecuación 3.6 escrita de otra manera como se muestra a continuación:

$$P(x = k) = e^{-\lambda} * \frac{\lambda^k}{k!} \quad \text{Ecuación 3.7}$$

Que se interpreta de la siguiente manera:

$P(x=k)$ es la probabilidad de ocurrencia cuando la variable discreta x toma un valor finito k , $\lambda = \text{Lambda}$ es la ocurrencia promedio por unidad (tiempo, volumen, área, o cualquier otra unidad que se trate). La constante e es la base de los logaritmos naturales y su valor aproximado es 2.711828, finalmente k es el número de éxitos por unidad.

La distribución de Gumbel es la que presentan valores extremos (Raynal, J.A., Raynal, M.E. 2004) por ejemplo si seleccionamos el día más lluvioso de cada año durante varios años, el día más frío del año durante varios años o el día más caluroso, la probabilidad de que estos valores superen un valor determinado dado presenta una distribución de Gumbel. Mayores detalles pueden encontrarse en (Chow, V.T., Maidment, D.R. y Mays, L.W. 1993 p.397)

Cuándo los datos no presentan una distribución definida (Wanielista, M. 1997), como las citadas anteriormente, podemos entonces utilizar estadística no paramétrica, es así como es posible estudiar conjuntos de observaciones, compararlos y obtener conclusiones.

Tamaño de las muestras.

Una vez que se tiene una población es necesario determinar el tamaño de la muestra válido para estimar una media de un grupo o conjunto de observaciones, recordando que la media y su desviación estándar determinan una distribución normal.

Contamos con la ecuación 3.4:

$$\mu = \bar{X} \pm 1.96 * S/\sqrt{n} \quad \text{Ecuación 3.4}$$

Dicha ecuación permite calcular con un 95% de certeza la media de un grupo normalmente distribuido, por lo tanto:

Si definimos una tolerancia de error en la media de tal forma que

$$\mu = \bar{X} + \text{tolerancia} \quad \text{Ecuación 3.8}$$

Podemos combinar esta ecuación con la 1.4 obteniendo la siguiente igualdad:

$$\bar{X} \pm \text{tolerancia} = \bar{X} \pm 1.96 * S/\sqrt{n} \quad \text{ya que}$$

$$\mu \text{ de la Ecuación 3.4} = \mu \text{ de la Ecuación 3.8}$$

Y de aquí podemos encontrar que:

tolerancia = $1.96 * S/\sqrt{n}$ despejando obtenemos la ecuación 3.9 que define el tamaño de la muestra válida para calcular una media de una población (Lohr, S. L. 1999).

$$n = 1.96^2 * S^2 / \text{tolerancia}^2 \quad \text{Ecuación 3.9}$$

Recordando que la constante 1.96 nos proporciona un 95% de certeza y tenemos la tabla 3.1 para obtener otros valores de la certidumbre con la que se calcula la muestra (Rossi, P. H., Wright, J.D. y Anderson, A. B. 1983). Recordemos también que S es la desviación estándar de la muestra y la tolerancia es un valor que representa los límites del error permitido en unidades de la muestra. Si nuestro estudio utiliza cuestionarios o encuestas este sería el número de cuestionarios o encuestas a aplicar.

Por ejemplo si deseamos calcular el tamaño de la muestra para calcular la media de una población de la que hemos investigado tomando una muestra previa que tiene una desviación estándar 4 y que podemos permitirnos una tolerancia de 2 substituyendo en la ecuación 3.9 obtenemos que el tamaño de la muestra viene dado por

$$n = (1.96*4)^2 / 2^2 = 15.36$$

Nótese que el problema radica en conocer la desviación estándar de una muestra previa, pues no se conoce el tamaño ideal de la muestra todavía por lo tanto no hay manera de calcular la desviación estándar en forma exacta.

¿Cómo podemos entonces conocer la desviación estándar de una muestra antes de calcular el tamaño de una muestra? Las sugerencias son las siguientes (Lehmann, D. R., Gupta, S. y Steckel, J.H. 1998):

Utilizar resultados de investigaciones previas o utilizar muestras piloto.

Diseño de instrumentos de recolección de datos.

La mayoría de las veces recolección de datos en las ciencias de la salud, sociales y administrativas se lleva a cabo por medio de cuestionarios o encuestas. La literatura reconoce varios tiempos involucrados para un cuestionario (Procter & Gamble 1990):

1. El momento de su diseño, preparación y redacción.
2. El tiempo de perfeccionamiento y afinación.
3. La etapa de aplicación.
4. La etapa de interpretación y análisis de la información obtenida.

EL MOMENTO DE SU DISEÑO, PREPARACIÓN Y REDACCIÓN.

Es necesario cubrir estos cuatro tiempos en forma adecuada para que la investigación obtenga resultados válidos. Empezando por el momento de diseño en el cual se toman las siguientes decisiones:

Constructos que se van a medir: Cada constructo puede medirse con una o más preguntas a cada pregunta se le llama también ítem que es la unidad básica de información. La combinación de los valores de los ítems proporcionan la medida de un constructo un constructo pasa a representar una variable a la cual se le tiene que detectar como primer paso su distribución estadística y posteriormente medir los parámetros que definen la distribución, por ejemplo si tiene distribución normal habrá que calcular media, desviación estándar para ese constructo. En ocasiones es necesario llevar a cabo estos estudios sobre cada ítem para validar los resultados obtenidos y posteriormente calcular los valores de los constructor.

Método a utilizar para aplicar el cuestionario: Electrónico, personal o entrevista e impreso son los más comunes, cada uno de ellos tiene ventajas y desventajas y depende contexto de la investigación cual es el más adecuado.

Escala a utilizar para codificar la información obtenida: Como se obtendrán y calificarán las respuestas, en forma numérica continua (números con fracciones decimales), numérica discreta (números con valores enteros), alfanumérica (letras o números por ejemplo la identificación de un grupo)

Se recomienda un número de ítems aproximado de 6 para cada constructo y el número total de ítems puede variar desde 10 hasta 90 siempre y cuando se abarquen todos los constructor. Si un cuestionario abarca más de un constructo se dice que es multidimensional.

Una vez que se han definido los constructos, el método y la escala es necesario redactar las preguntas, existen algunas recomendaciones ampliamente reconocidas para la redacción de preguntas (Arribas, M. 2004), las cuales se ilustran a continuación:

1.-La definición de cada ítem debe de ser completa:

Pregunta Incorrecta:

¿Ha tenido algún accidente automovilístico? _____

Pregunta correcta:

¿Cuántos accidentes automovilísticos de cualquier tipo, graves o leves ha tenido?
0 ___ 1 ___ 2 ___ 3 ___ 4 ___ 5+ ___

2.-Evitar el mezclar ítems en una sola pregunta:

Pregunta Incorrecta:

El aprendizaje depende tanto del maestro como del alumno ¿cómo considera la participación de maestros y alumnos en la universidad?

poca ___ regular ___ suficiente ___ excelente ___

Pregunta Correcta:

El aprendizaje depende de dos grupos principales maestros y alumnos, considerando esta universidad:

¿Cómo califica la participación de los maestros?

poca ___ regular ___ suficiente ___ excelente ___

¿Cómo califica la participación de los alumnos?

poca ___ regular ___ suficiente ___ excelente ___

3.-Utilizar preguntas cortas y fáciles de comprender KISS iniciales en inglés para *Keep it short and simple* o lo que es lo mismo consérvelo corto y sencillo:

Pregunta Incorrecta:

¿Cuál es su opinión de los procesadores Intel Duo adquiridos recientemente adquirido por la administración para el proceso de información contable y administrativa?

mala ___ regular ___ buena ___ excelente ___

Pregunta Correcta:

¿Cuál es su opinión del equipo de cómputo recientemente adquirido?

mala ___ regular ___ buena ___ excelente ___

4.-No utilizar preguntas que inducen una tendencia natural o estereotipo:

Pregunta Incorrecta:

¿Está de acuerdo con prohibir expresar opiniones correo electrónico entre los estudiantes en los laboratorios?

La mayoría de las personas tenemos una tendencia a favorecer los conceptos de libertad y alejarnos de las prohibiciones. Palabras como castigo, represión, sufrimiento, fraude, inducen respuestas que se alejan de ellas.

5.- El uso de la pregunta ¿Por qué? Es señal de una inducción a la respuesta:

Pregunta Incorrecta:

El servicio de entrega es considerado muy malo, ¿Por qué?

Se retrasa mucho ____

Pierde los paquetes ____

Maltrata los paquetes ____

Otros _____

Pregunta Correcta:

¿Qué aspectos se pueden mejorar en el servicio de entrega?

Tiempo Entrega ____

Garantía ____

Conservación de los paquetes ____

Otros (especifique) _____

6.- No utilizar preguntas en forma negativa:

Pregunta Incorrecta:

No debe de permitirse el uso de celulares durante los exámenes finales:

Si ____ No ____

Pregunta Correcta:

¿Marque su opinión acerca del uso de los celulares durante los exámenes finales?

Puede permitirse su uso ____ No debe utilizarse ____ Me abstengo de opinar ____

7.- Evitar preguntas que tengan una alternativa que difícilmente pueda ser rehusada:

Pregunta Incorrecta:

¿Cuenta con la experiencia y los conocimientos necesarios para conservar su trabajo?

Si ____ No ____

Pregunta Correcta:

¿Cómo considera su preparación y experiencia actual para cumplir con su trabajo?

Puede ser mejorada ___

regular pero se puede mejorar ___

buena ___

excelente ___

8.- No pedirle al encuestado un esfuerzo para recordar.

Pregunta Incorrecta:

¿Cuántas veces a utilizado el sistema de localización de estudiantes los últimos tres meses y cuál ha sido el resultado obtenido?

En cuanto a la escala utilizada es importante considerar que si la escala es una variable continua es más sencillo representarla por rango para facilitar la respuesta y agilizar su análisis. Por ejemplo

¿Cuáles son sus ingresos mensuales?

De 0 a \$4999 ___

De \$5000 a \$14,999 _____

De \$15,000 a \$19,999 ___

Más de \$20,000_____

Por otro lado generalmente la selección se utiliza para responder a una pregunta que identifica un grupo o subgrupo y la respuesta se representa por una variable alfanumérica. Por ejemplo:

¿Cuál es su estado civil?

Soltero(a) ___ Casado(a) ___ Divorciado(a) ___ Separado(a) ___ Viudo(a) ___

Internamente la respuesta puede almacenarse como:

Soltero(a)="S" Casado(a)="C" Divorciado(a)="D" Separado(a)="P" Viudo(a)="V"

Los valores discretos se pueden utilizar por rangos como para los valores continuos o bien se pueden utilizar selecciones como para las variables alfanuméricas.

Por ejemplo:

¿Cuántos personas dependen de usted?

0 ___ 1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___ Mas de 10 ___

O bien:

¿Cuántos personas dependen de usted?

0 ____ de 1 a 3 ____ de 4 a 7 ____ de 8 a 10 ____ Mas de 10 ____

El sentido de las escalas numéricas siempre puede transformarse para su análisis o bien estandarizarse por lo que en las preguntas la escala puede aparecer como el investigador lo requiera. Por ejemplo si la base de la escala es 5 y unas preguntas estan orientadas en sentido negativo o sea el 5 es la calificación menor, entonces la respuesta a la pregunta puede representarse como $(5 - \text{valor de la respuesta} + 1)$ de esta manera 1 es el valor más pequeño 5 el valor más alto la formula es:

$$\text{Valor transformado} = \text{Base} - \text{valor} + 1$$

El valor mínimo calculado será uno y el máximo será el valor de la Base.

EL TIEMPO DE PERFECCIONAMIENTO Y AFINACIÓN.

En este tiempo ya se cuenta con un borrador del instrumento y se procede a probarlo para ello es necesario una prueba o aplicación piloto del cuestionario, es importante seleccionar una muestra que se parezca a la muestra final, debe estar formada entre 30 y 50 personas u observaciones, hay cuestionarios que se aplican por familia o matrimonio y no por persona.

Se debe de validar para asegurarse que el instrumento es confiable y a la vez válido (Miquel, S., Bigné, E., Lévy, J.P., Cuenca, A.C., Miquel, M. 1997) y (Arribas, M. 2004).

La **confiabilidad** tiene que ver con la consistencia, con la independencia del tiempo y el hecho que aplicado por diferentes observadores el instrumento arroje resultados similares en condiciones similares de aplicación. Es la confiabilidad un grado de precisión del instrumento.

Un factor importante a medir es la **consistencia** interna del instrumento, esto tiene que ver con el nivel que los diferentes ítems están relacionados entre sí. Para medir la consistencia se cuenta con e coeficiente **alfa de Cronbach** es un método estadístico muy utilizado. Sus valores oscilan entre 0 y 1. Por debajo de este valor se considera que no hay una buena consistencia interna de los ítems, el valor máximo aceptable es de 0.90 por arriba de este valor hay indicación de redundancia entre los ítems valores entre 0.80 y 0.90 son los esperados para considerar una buena consistencia. (Oviedo, H.C. y Campo-Arias, A. 2005)

Si un mismo investigador aplica el cuestionario dos veces a la misma muestra pero en dos tiempos distintos podemos observar la **estabilidad temporal** del cuestionario también conocida como confiabilidad test-retest. Para medirla se utiliza el **coeficiente de correlación intra clase** [CCI], para variables continuas y evaluaciones temporales distantes un 70% indica un resultado aceptable.

Para medir la **concordancia** entre observadores contamos con mediante el **porcentaje de acuerdo y el índice Kappa**. Para poder medir la concordancia es necesario aplicar el mismo cuestionario a la misma muestra por dos evaluadores diferentes.

VALIDEZ

La **validez** indica si el instrumento mide realmente lo que pretende medir, o sea aquello para lo que fue diseñado. Existen varios tipos de validez.

Validez de contenido. Esta tipo de validez nos indica si el instrumento mide lo que trata de medir se basa en la opinión de investigadores expertos que evalúan el cuestionario para asegurar que mida lo que debe medir. Debido a que se basa en un análisis subjetivo, no hay cálculo valido para cuantificar esta validez que es totalmente cualitativa.

Validez de constructo. Representa el grado en que el instrumento refleja la teoría del fenómeno o del concepto que mide. Para calcular esta validez se utilizan el **análisis factorial y la matriz multirrasgo-multimétodo**.

Validez de criterio. Es una comparación de la puntuación obtenida por cada sujeto con la puntuación obtenida por otro instrumento que mide lo mismo y que se utiliza como referencia. No siempre se cuenta con dichos instrumentos. Dependiendo del tipo de variables, utilizaremos para las variables numéricas los coeficientes de correlación de Pearson y para las variables cualitativas el cálculo de la sensibilidad y especificidad.

La tabla 3.5 muestra un resumen de las herramientas utilizadas para evaluar el instrumento.

Tipo de Evaluación	Aspecto Evaluado	Herramienta
Confiabilidad	Consistencia	Alfa de Cronbach
Confiabilidad	Estabilidad Temporal	Coefficiente de Correlación Intra Clase CCI
Confiabilidad	Concordancia	Porcentaje de Acuerdo Índice de Kappa
Validez	De Contenido	Característica Cualitativa sin cálculo
Validez	Constructo	Análisis factorial y la Matriz Multirrasgo-Multimétodo
Validez	Criterio	Comparación con puntajes obtenidos con otro instrumento por medio de Coeficientes de Correlación de Pearson (variables numéricas) Num y Cálculo de la Sensibilidad y Especificidad (variables cualitativas)

Tabla 3.5 Resumen de herramientas a utilizar para validar el instrumento

Para poder afinar el instrumento pueden llevarse a cabo varias aplicaciones piloto eliminando y ajustando los ítems y la formación de los constructos de acuerdo a los resultados que se obtengan de las distintas herramientas que evalúan el instrumento, es en esta etapa cuándo se prevé posibles problemas posteriores con la inconsistencia o invalidez de los datos obtenidos en el estudio, es en este momento cuándo se puede obtener una garantía de la calidad, confiabilidad y validez de la investigación por lo que cualquier tiempo que se le dedique a esta etapa será aprovechado de una manera muy importante conforme avance la investigación.

LA ETAPA DE APLICACIÓN.

En esta etapa el instrumento ha sido evaluado y esta listo para ser aplicado, el cálculo de la muestra se ha llevado a cabo utilizando la aplicación piloto o bien haciendo muestras piloto para el cálculo de la desviación estándar requerida en la ecuación 1.9 por lo tanto todo lo que resta es llevar a cabo la aplicación del cuestionario ya sea por medios electrónicos, directos o por medios impresos.

Al recabar los resultados estos serán convertidos a medios magnéticos y posteriormente transformados a las escalas requeridas por el estudio.

LA ETAPA DE INTERPRETACIÓN Y ANÁLISIS DE LA INFORMACIÓN OBTENIDA.

Finalmente llegamos al proceso de los resultados obtenidos el cual depende del tipo de hipótesis del estudio normalmente los resultados requieren un pre proceso antes de poder ser analizados, es aquí donde se obtienen los valores de los constructos, se dividen en grupos los datos y se transforman los valores para utilizar escalas estándares todo ello puede probarse en la aplicación piloto, actualmente se cuenta con distintos paquetes estadísticos de software de diferentes tipos, desde los más reconocidos y populares como el SPSS y SAS, hasta los más sencillas hojas de cálculo el investigador utiliza el software de acuerdo a su presupuesto y a su contexto.

Para terminar esta sección mostramos en la tabla 3.6 contiene un listado no exhaustivo pero si representativo de software de tipo estadístico cuya licencia es gratuita tipo GNU-GPL tomada del sitio web *Free Software Directory* de la dirección: <http://directory.fsf.org/category/mathstat/>

Paquete de Software	Descripción
Apophenia	Biblioteca de cálculos estadísticos en lenguaje C
Autoclass	Detección automática de <i>clusters</i> o subconjuntos
Bc	Lenguaje algebraico interactivo
PSPP	Paquete estadístico con cálculos básicos muy similar al SPSS pero mucho más reducido
Ploticus	Paquete para obtener gráficas
Plotutils	Paquete para obtener gráficas
PyChem	Análisis de multi variables con interfase para graficar
SalSat	Análisis estadístico especialmente diseñado para investigación en psicología
Statist	Cálculos estadísticos
StatistX	Programa estadístico basado en GTK (Es un paquete de interfase gráfica de usuario, manejo de ventanas)
Vilno	Transformación de datos, preparación de datos para ser analizados
mcsim	Software de simulación para el diseño, análisis y calibración de modelos matemáticos

Tabla 3.6 Paquetes de software para el pre-proceso y proceso de los datos estadísticos.

Confiabilidad del Instrumento por medio del Coeficiente Alfa de Cronbach.

El coeficiente de Alfa de Cronbach evalúa la consistencia interna del cuestionario, esto es toma un constructo o dimensión y los ítems que lo forman y utilizando los valores de las respuestas de esos ítems lleva a cabo un análisis para calcular un solo número llamado coeficiente de confiabilidad de Cronbach el cual nos indica la consistencia interna de dicho constructo. Un valor de entre 0.8 y 0.9 se considera satisfactorio (Oviedo, H.C. y Campo-Arias, A. 2005).

Requisitos para poder utilizar el alfa de Cronbach (Muñiz, J. 2003):

1.- Que el cuestionario este formado por ítems que se suman para proporcionar una puntuación global del constructo.

2.- Que todos los ítems midan la propiedad en cuestión en la misma dirección, si acaso se usa dirección inversa en algunos ítems, será necesario transformarlos antes de utilizar el coeficiente de alfa de Cronbach.

Existen dos procedimientos para calcular el coeficiente de Alfa de Cronbach (definido como α) el primero requiere el cálculo de la varianza (cuadrado de la desviación estándar) y el segundo el cálculo de las correlaciones de Pearson (Briones, G. 2001).

Tomemos el primer procedimiento, la fórmula del coeficiente es:

$$\alpha = \frac{k}{(k-1)} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right)$$

Ecuación 3.10

Dónde:

k .- es el número de ítems que componen el constructo.

σ_i^2 .-Varianza de cada ítem

σ_x^2 .-Varianza Total: Se obtiene una suma de todos los ítems para el constructo considerando cada observación, generándose una nueva columna con la suma para el constructo. Posteriormente se calcula la varianza de esa columna.

Por ejemplo si los 4 ítems que se muestran en la figura 3.3 forman un constructo y se han aplicado tres encuestas con las respuestas que muestra la figura.

	A	B	C	D	E
1					
2		Var 1	Var 2	Var 3	Var 4
3		2	2	3	3
4		3	3	5	7
5		2	3	3	5
6					

Figura 3.3 Ejemplo de datos para calcular el coeficiente de confiabilidad Alfa de Cronbach

Utilizando una hoja de cálculo podemos aplicar la Ecuación 3.10 para calcular el coeficiente como muestra la figura 3.4.

	A	B	C	D	E	F
1						
2		Var 1	Var 2	Var 3	Var 4	Suma por Constructo
3		2	2	3	3	10
4		3	3	5	7	18
5		2	3	3	5	13
6	Varianza Por Ítem	0.333333333	0.333333333	1.333333333	4	16.33333333
7						
8		Varianza de la columna =VAR(B3:B5)		Suma del renglon =SUMA(B3:E3)		Varianza Total =VAR(F3:F5)
9						
10						
11						
12						
13	Suma varianzas =	6		Alfa=	0.843537415	
14						
15		Suma de las varianzas =SUMA(B6:E6)		Cálculo del coeficiente =(4/3)*(1 - (B13/F6))		
16						

Figura 3.4 Cálculo del coeficiente aplicando la ecuación 3.10 en una hoja de cálculo

Es posible utilizar el SPSS sobre los mismos datos y obtener la siguientes tablas, el procedimiento para obtener estos resultados es el siguiente:

1. Abrir desde el menú principal opción *File* abrir el archivo de datos
2. Desde *Data View* una vez que los datos están cargados, seleccionar la opción *Analyze* y de ahí la opción *Scale*.
3. Desde la opción *Scale* seleccionar *Reliability Analysis*
4. De la ventana resultante seleccionar las variables que contienen los valores de los ítems y pasarlos a la ventana pequeña de la derecha. Conservar o seleccionar modelo Alfa
5. Oprimir el botón *Statistics* y de las nuevas opciones seleccionar "*Scale if item deleted*" oprimir el botón *Continue*
6. Oprimir el botón *OK*

Y se obtendrán los resultados mostrados.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.844	.929	4

Tabla 3.7 Resultado del SPSS al calcular el alfa de Cronbach de los datos mostrados en la figura 3.3

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.417	2.333	5.000	2.667	2.143	1.435	4
Item Variances	1.500	.333	4.000	3.667	12.000	3.000	4

Tabla 3.8 Resultado del SPSS al calcular el alfa de Cronbach de los datos mostrados en la figura 3.3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
V1	11.33	12.333	.904	.	.811
V2	11.00	13.000	.721	.	.846
V3	10.00	9.000	.866	.	.722
V4	8.67	4.333	.961	.	.808

Tabla 3.9 Resultado del SPSS al calcular el alfa de Cronbach de los datos mostrados en la figura 3.3

En la última tabla obtenida por el SPSS se muestra el resultado de volver a calcular el coeficiente si se elimina el ítem del renglón respectivo, en este caso se ve que es posible eliminar el ítem V2 y se mejoraría el resultado final. Esta acción permitiría mejorar la confiabilidad del cuestionario.

Existe un segundo procedimiento para calcular el coeficiente alfa de Cronbach se basa en el promedio de las correlaciones de la matriz de Pearson y su corma de cálculo se muestra en la ecuación 3.11

$$\alpha = (k \cdot P) / (1 + P \cdot (k - 1)) \text{ Ecuación 3.11}$$

Dónde:

k .- es el número de ítems que componen el constructo.

P .- es el promedio de las correlaciones lineales entre cada uno de los ítems

Para los mismos datos de la figura 3.3, se puede calcular la matriz de correlación de Pearson utilizando el SPSS siguiendo el procedimiento a continuación:

1. Abrir desde el menú principal opción *File* abrir el archivo de datos

2. Desde *Data View* una vez que los datos están cargados, seleccionar la opción *Analyze* y de ahí la opción *Correlate*.
3. Desde la opción *Correlate* seleccionar *Bivariate*
4. De la ventana resultante seleccionar las variables que contienen los valores de los ítems y pasarlos a la ventana pequeña de la derecha.
5. Oprimir botón *OK*

Y se obtendrán los resultados mostrados.

Correlations

		V1	V2	V3	V4
V1	Pearson Correlation	1	.500	1.000(**)	.866
	Sig. (2-tailed)		.667	.000	.333
	N	3	3	3	3
V2	Pearson Correlation	.500	1	.500	.866
	Sig. (2-tailed)	.667		.667	.333
	N	3	3	3	3
V3	Pearson Correlation	1.000(**)	.500	1	.866
	Sig. (2-tailed)	.000	.667		.333
	N	3	3	3	3
V4	Pearson Correlation	.866	.866	.866	1
	Sig. (2-tailed)	.333	.333	.333	
	N	3	3	3	3

** Correlation is significant at the 0.01 level (2-tailed).

Tabla 3.10 Matriz de correlación de Pearson para los datos del ejemplo

Estos datos pueden pasarse a una hoja de cálculo para hacer los cálculos finales como se muestra en la figura 3.5

	A	B	C	D
1				
2		0.5		
3		1		
4		0.866		
5		0.5		
6		0.5		
7		0.866		
8		1		
9		0.5		
10		0.866		
11		0.866		
12		0.866		
13		0.866		
14				
15	Suma	9.196		
16	Promedio	0.76633333		
17				
18	Alfa =	0.92917046		
19				
20				
21				
22				

Correlaciones de la matriz de Perason ignorando la diagonal, que solo son valores 1

Suma de todas las correlaciones =SUMA(B2:B13)

Promedio de las correlaciones valor P de la Ec. 1.11

Valor calculado para el Alfa de Cronbach = $(4 * B16) / (1 + B16*(4 - 1))$

Figura 3.5 Cálculo del Alfa de Cronbach mediante el procedimiento de correlaciones

Nótese que el resultado coincide con el resultado del SPSS para datos estandarizados.

Estos procedimientos (Thomas, H.P. 2004) son los más utilizados para llevar a cabo el análisis de confiabilidad mediante el análisis de consistencia interna de un instrumento.

Relaciones entre variables: Regresión Lineal Simple y Múltiple.

Existen dos tipos de Hipótesis:

1. Causales: Los valores de una o más variables (llamadas variables independientes) determinan los valores de una segunda variable (llamada variable dependiente)
2. Asociativas: Los valores de una o más variables covarían o se correlacionan con los valores de una segunda variable.

Primero discutiremos la regresión que tiene su aplicación en las hipótesis causales .

Las hipótesis causales requieren demostrar la relación causal entre dos variables (Maletta, H. 1995) y Diez Medrano, J.(1997) por ejemplo una hipótesis de la forma:

“Los ingresos personales determinan el nivel de estudios de una persona, a mayores ingresos mayor nivel de estudios, los ingresos personales determinan el nivel de estudios de una persona”.

Muestra una variable independiente, en el ejemplo el nivel de ingresos y una variable dependiente en el ejemplo el nivel de estudios. Podemos asumir a **Y** como variable dependiente y a **X** como variable independiente. Para demostrar este tipo de hipótesis se recurre a establecer un modelo que puede comprobarse estadísticamente.

Los modelos pueden ser lineales (se ajustan a una recta) o no lineales (se ajustan a una curva). El modelo más sencillo que existe es

$$\mathbf{y = m x} \quad \text{Ecuación 3.12}$$

Dónde **m** es una constante numérica. En este caso se puede tener modelos con **m** positiva que involucra una relación directamente proporcional o una **m** negativa que involucra una relación inversamente proporcional.

Si al modelo más simple agregamos una constante **b** obtenemos la ecuación general de la recta que es:

$$\mathbf{y = m x + b} \quad \text{Ecuación 3.13}$$

m es la pendiente de la recta y **b** el punto de intersección de la recta con el eje de **y** cuando **x = 0**.

Si para el ejemplo hemos codificados respuestas de un cuestionario y tenemos que los ingresos pueden tomar valores de 1 al 20 (dividiendo en 20 rangos los valores reales posibles) y que el nivel de estudios puede ir de

1 a 15. Al aplicar un cuestionario en 20 personas obtenemos las respuestas mostradas en la tabla 3.11.

Num. Observación	Nivel De Estudios	Ingresos
1	3	3
2	4	4
3	5	3
4	4	5
5	5	6
6	8	4
7	7	5
8	5	6
9	8	7
10	10	9
11	14	10
12	13	14
13	12	12
14	15	16
15	15	18
16	13	19
17	13	20
18	14	18
19	15	19
20	14	20

Tabla 3.11 Tabla de observación del ejemplo para demostrar la hipótesis:
 “Los ingresos personales determinan el nivel de estudios de una persona, a mayores ingresos mayor nivel de estudios”.

Podemos obtener una gráfica en el SPSS, utilizando el siguiente procedimiento:

1. Seleccionar la opción *Graphs*
2. Seleccionar *Legacy Dialogs*
3. Seleccionar *Scatter/Dot*
4. Seleccionar la opción *Simple Scatter* y definiendo la variable Nivel de Estudios como eje **Y** y la variable Ingresos como el eje **X**.

Se obtiene la figura 3.6 este procedimiento es la primera exploración obligada entre dos variables.

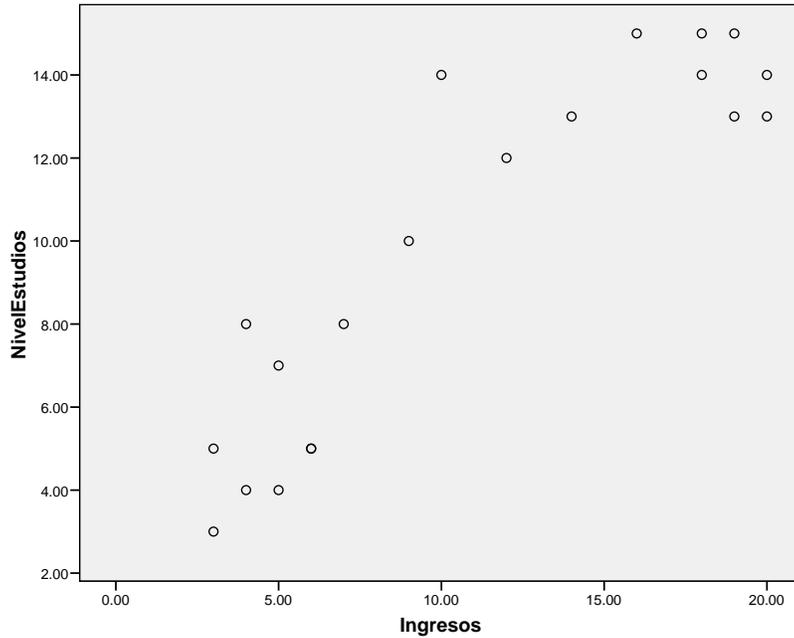


Figura 3.6 Los datos del ejemplo graficados.

Para explorar si hay alguna relación que involucre una regresión simple, esto es una línea recta entre las dos variables, podemos recurrir a calcular m y b para estos datos utilizando el SPSS de la siguiente manera:

1. Seleccionar la opción *Analyze*
2. Seleccionar *Regression*
3. Seleccionar *Curve Estimation*
4. Como sospechamos que hay una regresión simple, marcamos la opción *Linear* colocando en la ventana de la variable dependiente el Nivel de Estudios y en la ventana de la variable Independiente los Ingresos

Para obtener las tabla 3.12 y la figura 3.7 que muestra una línea recta por entre los puntos.

Model Summary and Parameter Estimates

Dependent Variable: NivelEstudios

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	.813	78.157	1	18	.000	3.164	.613

The independent variable is Ingresos.

Tabla 3.12 Resultados para los datos del ejemplo

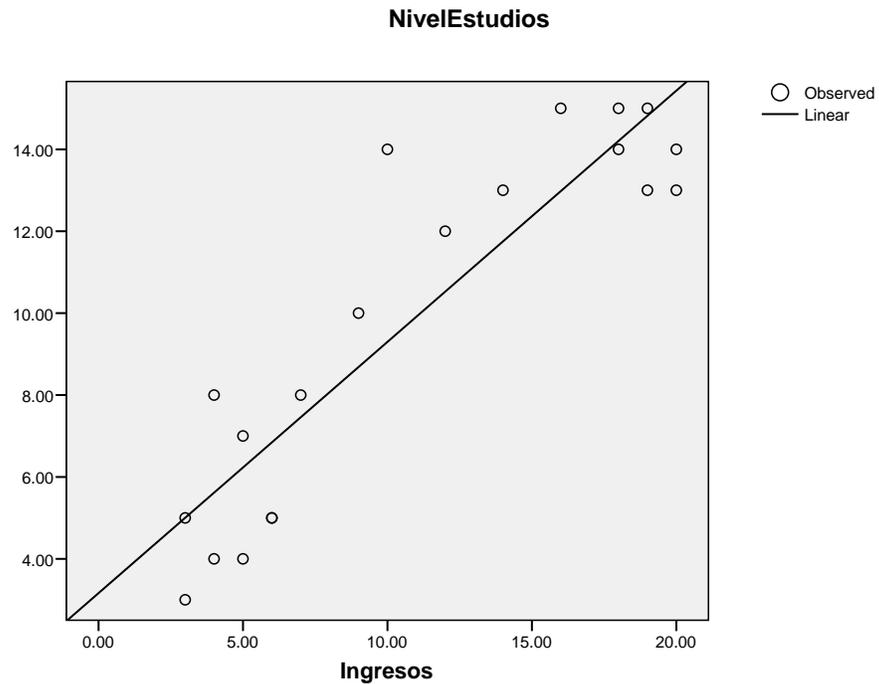


Figura 3.7 Línea recta que modela la relación entre las variables.

De la tabla 3.12 bajo el encabezado de *Parameter Estimates* encontramos los valores de $m = .613$ y $b = 3.164$ substituyendo en la ecuación 3.13 podemos obtener la ecuación de éste modelo que queda de la siguiente forma:

$$y = 0.613 x + 3.164 \text{ Ecuación 3.14}$$

Dónde x representa el nivel de ingresos y y el nivel de estudios, la recta de la figura 3.7 representa la ecuación encontrada.

Para determinar los valores de m y b utilizamos un algoritmo que minimiza la distancia elevada al cuadrado de la recta del modelo a cada uno de los puntos dados, por lo que se le llama algoritmo de mínimos cuadrados. Para minimizar los cuadrados se iguala a cero la derivada de la expresión que define las distancias elevadas al cuadrado este es un procedimiento de cálculo diferencial elemental.

No solo existen modelos lineales, podemos repetir el análisis pero ahora marcando en el paso 4 del procedimiento anterior la opción cuadrática obtenemos la tabla 3.12 y la figura 3.8

Model Summary and Parameter Estimates

Dependent Variable: NivelEstudios

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	.813	78.157	1	18	.000	3.164	.613	
Quadratic	.883	64.110	2	17	.000	-1.300	1.711	-.047

The independent variable is Ingresos.

Tabla 3.13 Aproximación cuadrática de los datos del ejemplo

Observamos que **R** cuadrada aumenta a 0.883, lo que representa una mejor aproximación y la ecuación queda como:

$$y = -0.047x^2 + 1.711x - 1.300 \quad \text{Ecuación 3.15}$$

Hemos encontrado un modelo cuadrático más sofisticado, no lineal que se aproxima mejor a los datos. Como se aprecia en la figura 3.8

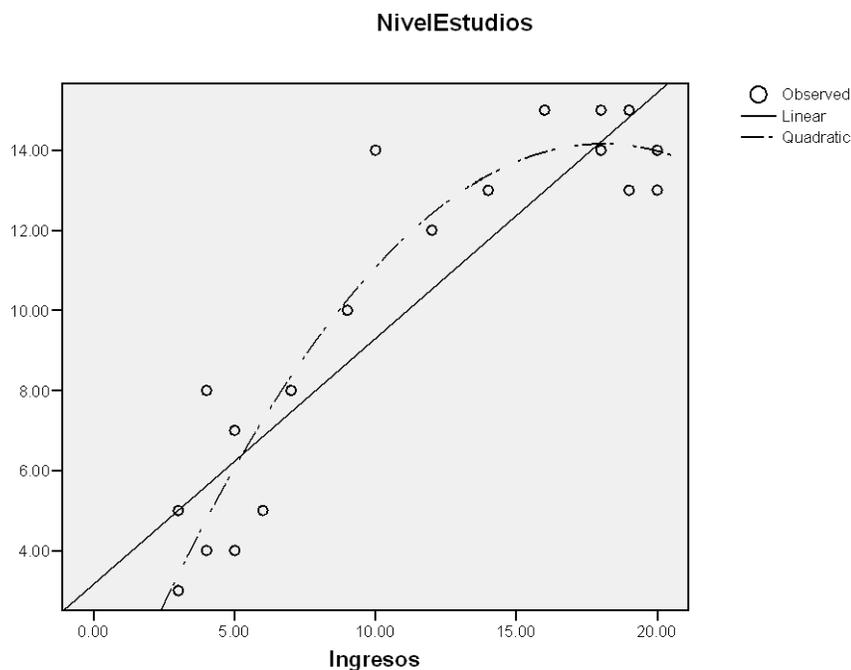


Figura 3.8 Aproximaciones lineal y aproximación cuadrática

Se pueden seleccionar todos los modelos disponibles en el paso 4 y detectar cual es el modelo que tiene mejor valor (mayor) de **R**² en este caso encontramos que el mejor modelo es el de la cúbica, como se aprecia en la tabla 3.14 el cual arroja una **R**² cuadrada de 0.897 y una ecuación

$$y = -.005 x^3 + 0.112 x^2 + 0.157 x + 2.770 \text{ Ecuación 3.16}$$

Model Summary and Parameter Estimates

Equation	Model Summary					Parameter Estimates				
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3	
Linear	.813	78.157	1	18	.000	3.164	.613			
Quadratic	.883	64.110	2	17	.000	-1.300	1.711	-.047		
Cubic	.897	46.226	3	16	.000	2.770	.157	.112	-.005	

Dependent Variable: NivelEstudios
The independent variable is Ingresos.

Tabla 3.14 Resultados para los modelos Lineal, cuadrático y Cúbico del ejemplo

Regresión Múltiple

El caso anterior es conocido también como regresión simple porque la variable dependiente se explica solo con una variable independiente, pero hay ocasiones en que son varios los factores que explican una variable a esto se le llama regresión múltiple (Mora y Araujo, M 1984).

Para explicar una variable dependiente no solo se cuenta con una variable independiente que la determine, sino que puede haber más de una variable involucrada. En el ejemplo hipotético anterior podemos considerar que aparte de medir el constructo de Ingresos también podemos medir el constructo de nivel de educación artística infantil, queriendo decir que no solo los ingresos intervienen para que una persona tenga un nivel de estudios elevado, sino también la educación artística adquirida en su infancia. Asumiendo que en este ejemplo tenemos los datos mostrados ahora en la tabla 3.15.

Num. Observación	Nivel De Estudios	Ingresos	Educación Artística adquirida en la infancia
1	3	3	1
2	4	4	1
3	5	3	1
4	4	5	2
5	5	6	2
6	8	4	3
7	7	5	3
8	5	6	3
9	8	7	4
10	10	9	4
11	14	10	4
12	13	14	4
13	12	12	5
14	15	16	5
15	15	18	4
16	13	19	5
17	13	20	5
18	14	18	4
19	15	19	5
20	14	20	5

Tabla 3.15 Ejemplo ampliado a dos variables independientes.

Como se puede observar la educación artística adquirida en la infancia tiene una escala del 1 al 5 y es una variable más involucrada en el modelo.

Para considerar esta variable tenemos el modelo de regresión lineal múltiple siguiente:

$$y = \beta + \beta_1 x + \beta_2 z \quad \text{Ecuación 3.17}$$

Para determinar los valores de los coeficientes β , β_1 y β_2 se cuenta con un sistema de ecuaciones lineales originadas en el hecho de que cada observación de la tabla 3.15 debe cumplir la ecuación mostrada del modelo.

Por otra parte observamos que para el modelo lineal en el SPSS podemos seleccionar varias variables independientes que expliquen la variable dependiente, seguimos el procedimiento descrito para obtener la tabla 3.16 que nos muestra el modelo obtenido.

Coefficients(a)

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta	B	Std. Error
1 (Constant)	1.136	1.035		1.098	.288
Ingresos	.354	.109	.521	3.239	.005
EducacionArtistica	1.386	.493	.452	2.812	.012

a Dependent Variable: NivelEstudios

Tabla 3.16 Resultados para la regresión lineal múltiple de dos variables

Los valores t indican la importancia de cada variable en la explicación de la variable independiente valores por encima de +2 o por debajo de -2 son significativos. Los coeficientes β , β_1 y β_2 de la columna *Unstandardized Coefficients* en este caso los valores son $\beta = 1.136$, $\beta_1 = 0.354$ y $\beta_2 = 1.386$ por el valor de la columna t podemos ver que la variable más representativa del modelo es la de Ingresos. La ecuación del modelo quedaría como:

$$y = 1.36 + 0.354 x + 1.386 z \text{ Ecuación 3.18}$$

Algo extremadamente importante es que para que esto sea valido las variables de la tabla 3.15 deben de tener una distribución normal, para poder demostrar la normalidad de los datos es necesario aplicar una prueba de normalidad, como lo mostraremos más adelante.

Los supuestos del modelo de regresión son los siguientes (Levi, R.I. y Rubin, D.S. 1996):

- Normalidad
- Linealidad
- Homoscedasticidad (Homogeneidad de varianzas)
- Multicolinealidad y singularidad

Normalidad.- Los datos deben de mostrar una distribución normal, puede aplicarse para ello la prueba de Kolmogorov Smirnov

Linealidad.- Se asume una distribución lineal entre las variables

Homoscedasticidad.- Las variancias de los valores de la variable dependiente (datos del estudio), para cada posible combinación de niveles de la variable x, son iguales; es decir, la variancia de los residuales (errores de estimación del modelo) es constante.

La multicolinealidad significa que las variables independientes están correlacionadas. Si una variable independiente puede ser predicha por otra en forma exacta como por ejemplo si la altura de una persona tiene dos variables independientes que la explican: peso en libras y peso en kilos.

Estos dos predictores son redundantes, ya que el peso es único independiente de si se mide con libras o kilos, por lo tanto una de las variables es redundante ya que entre ellas presentan multicolinealidad, singularidad es exactamente lo opuesto, una variable no depende no es equivalente a la otra.

Las hipótesis asociativas, indican que una variable se correlaciona con otra como se explicó al inicio de esta sección. Para demostrar lo anterior se cuenta con la matriz de correlación de Pearson, que es un matriz que tiene una columna para cada variable y un renglón para cada variable, de tal manera que la diagonal está ocupada por valores 1.0 y el triangulo superior de la diagonal es una imagen en el espejo del triangulo inferior. Entre más cercano a 1 ó -1 sea el valor de un coeficiente la correlación entre ambas variables, la de la columna y la del renglón la correlación es más fuerte, el signo explica el sentido de la correlación para las variables redundantes se llega al 1. La tabla 3.17.

	Nivel De Estudios	Ingresos	Educación Artística adquirida en la infancia
Nivel De Estudios	1.0	0.902	0.891
Ingresos	0.902	1.0	0.842
Educación Artística adquirida en la infancia	0.891	0.842	1.0

Tabla 3.17 Matriz de correlación de Pearson para el ejemplo

La matriz de correlación solo se utiliza cuándo se quiere demostrar que una variable varía con otra (relación asociativa), pero no cuándo una variable determina otra (relación causal). La naturaleza de la hipótesis toca definirla al investigador (Maletta, H.1995).

El procedimiento para obtener esta matriz a partir del SPSS es el siguiente:

1. Seleccionar la opción *Analyze*
2. Seleccionar *Coorrelate*
3. Seleccionar *Curve Bivariate* y pasar a la ventana de la izquierda las variables que van a formar la matriz de correlación y oprimir OK

Análisis para identificación de clusters o subconjuntos de elementos relacionados dentro de una población.

Cuando intentamos identificar una propiedad en una población y aunque esperemos detectarla no podemos confirmar su presencia, es el momento de intentar formar subconjuntos de dicha población para dividirla y estudiar los subconjuntos por separado. Por ejemplo supongamos que estudiamos el área de construcción de las edificaciones de una ciudad y esperamos que exista una distribución normal en esta población pero las pruebas de normalidad no lo confirman, podemos utilizar el tipo de edificación para clasificarlas por tipo (identificar subconjuntos) y de esta manera estudiar cada subconjunto por separado.

Observemos la tabla 3.18 aparece a continuación:

No. Observación	Área Edificada	Tipo de Edificio
1	400	1
2	1500	2
3	2000	2
4	1750	2
5	500	1
6	3500	2
7	4000	2
8	800	2
9	600	1
10	900	2
11	400	1
12	700	2
13	350	1
14	2750	2
15	3000	2
16	500	1
17	3500	2
18	3200	2
19	4500	2
20	275	1
21	5000	2
22	370	1
23	6000	2

Tabla 3.18 Área de construcción y tipos de edificios en una ciudad, ejemplo de datos

El área está dada en metros cuadrados y el tipo de construcción es 1 para casa habitación 2 para edificios de oficinas o vivienda.

Si intentamos probar la normalidad en toda la población la prueba de Kolomgorov Smirnov calcula una significancia de 0.227 que indica cómo se verá más adelante que no existe normalidad en los datos como un todo.

Sin embargo si definimos dos grupos y corremos la misma prueba para cada grupo obtenemos una significancia de 0.839 para el tipo de edificación 1 y 0.995 para el tipo de edificación 2 indicando claramente que la distribución es una distribución normal para cada tipo por separado, por lo que los tipos deberán ser estudiados por separado (Cortés, F. y Rubalcava, R. M. 1987). Este es un ejemplo claro de la necesidad de dividir la población en subconjuntos también llamados *clusters*.

Para hacer un estudio en el SPSS de grupos podemos seguir el siguiente procedimiento:

1. Seleccionar la opción *Data*
2. Seleccionar *Split File* pasar a la ventana de la derecha la variable que pensamos define los grupos en este caso el Tipo de Edificación, y activar la opción *Organize output by groups* y oprimir OK
3. Llevar a cabo la prueba de normalidad o cualquier otra prueba y esta será producida por separado para cada tipo de edificación.

Una hipótesis planteada en un estudio puede asegurar que la población se divide en dos o más grupos, de ser así, la existencia de una característica que esté presente en cada grupo por separado, pero que no esté presente en la población total nos llevará a demostrar la existencia de los *clusters* (Luque M. 2000) tal como se mostró en el ejemplo anterior.

No siempre contamos con la facilidad de tener en una sola variable la clasificación de los *clusters*, por lo que es necesario contar con un algoritmo que nos permite “calcular” la o generar la variable mediante la cual haremos la clasificación (Ferrand Aranaz, M. 2005). Se recurre al siguiente procedimiento:

1. Seleccionar una o más variables para cada observación utilizando estas variables se mide la similitud entre las observaciones
2. Una vez medida la similitud se agrupan en grupos homogéneos internamente y diferentes entre si, para ello se genera una variable clasificatoria.
3. Los estudios se llevan a cabo dividiendo la muestra en grupos atendiendo a la variable clasificatoria.

Por ejemplo si el ejemplo anterior no contara con la variable tipo de edificio, sino con las variables: Num. de habitantes, No. De habitaciones y tuviéramos la tabla 3.19 en lugar de la 1.18.

No. Observación	Área Edificada	No. habitaciones	No. Ocupantes
1	400	4	3
2	1500	20	15
3	2000	40	20
4	1750	27	25
5	500	5	2
6	3500	39	30
7	4000	35	20
8	800	25	19
9	600	6	1
10	900	21	20
11	400	8	4
12	700	20	20
13	350	3	6
14	2750	30	25
15	3000	35	30
16	500	5	4
17	3500	50	30
18	3200	50	35
19	4500	60	40
20	275	3	1
21	5000	59	45
22	370	4	2
23	6000	80	30

Tabla 3.19 Área de construcción, y datos alternos para calcular el tipo de edificio

Se puede observar que utilizando el número de habitaciones y el número de ocupantes se puede deducir el tipo de edificio. La regla es sencilla, si el número de habitantes es mayor de 9 y el número de habitaciones es 8 el tipo de edificio es 2 y de lo contrario es 1.

Con esto podemos llegar a los resultados planteados por la tabla 3.18

Análisis de datos para comprobación de comportamiento normal prueba de Kolmogorov-Smirnov.

Como hemos visto el requerimiento de que los datos sigan una distribución normal se presenta en los diferentes análisis estadísticos de las secciones anteriores. Una de las pruebas más ampliamente utilizadas en la práctica es la prueba Kolmogorov-Smirnov (Altman, D. A. 1997). El fundamento de ésta prueba radica en comparar la función de distribución acumulada de los datos observados con la de una distribución normal, midiendo la máxima distancia entre ambas curvas. Como en cualquier prueba de hipó-

tesis, la hipótesis nula se rechaza cuando el valor obtenido es mayor a un cierto valor crítico que se obtiene desde una tabla de probabilidad.

El fundamento de la prueba es el siguiente:

Sea x una variable de tipo continuo.

Hipótesis:

H_0 . x sigue la distribución normal

H_1 . x no sigue la distribución normal

Comparar la distribución empírica F_n de la muestra con la distribución normal F si esta comparación revela diferencias significativas, se rechaza la prueba la Hipótesis H_0

El estadístico de prueba esta y región crítica, respectivamente son:

$$D = \sup_x |F_n(x) - F(x)| \quad \text{Ecuación 3.18}$$

El SPSS ofrece este procedimiento y proporciona el valor de la prueba y el valor de p asociado a ella. La prueba Kolmogorov-Smirnov, considera menos importantes las observaciones extremas y por la tanto es menos sensible a las desviaciones que normalmente se producen en las puntas (Pértiga, S., Fernández, P. 2001).

Si utilizamos la tabla 3.15 y ejecutamos esta prueba el procedimiento sería el siguiente:

1. Seleccionar la opción *Analyze*
2. Seleccionar Nonparametric Tests
3. Seleccionar *1-Sample K-S* y pasar a la ventana de la izquierda las variables que van a ser analizadas, normalmente variables numéricas continuas y seleccionar *Test Distribution: Normal* y oprimir OK

Se obtienen los resultados como se muestra en la tabla 1.20

One-Sample Kolmogorov-Smirnov Test

		NivelEstudios	Ingresos	EducacionArtistica
N		20	20	20
Normal Parameters(a,b)	Mean	9.8500	10.9000	3.5000
	Std. Deviation	4.39228	6.45552	1.43270
Most Extreme Differences	Absolute	.213	.177	.236
	Positive	.165	.177	.148
	Negative	-.213	-.164	-.236
Kolmogorov-Smirnov Z		.954	.792	1.057
Asymp. Sig. (2-tailed)		.322	.557	.213

a Test distribution is Normal.
 b Calculated from data.

Tabla 3.20 Resultados de la prueba de normalidad

La interpretación es la siguiente si el valor de la significancia (ultimo renglón de la tabla) es mayor que 0.05 indica que los datos se distribuyen normalmente, si es menor a dicho numero los datos no tienen distribución normal y no pueden ser tratados como tal, en éste ejemplo las tres variables en cuestión tienen una distribución normal.

Se proporciona un juego de resultados para cada variable analizada, calculando los parámetros que definen la distribución normal.

Comparación de dos conjuntos de observaciones: pruebas t-Student y Mann-Whitney.

Frecuentemente se requiere comparar dos grupos de observaciones, en un experimento por lo regular se tiene un grupo de control y otro experimental y los resultados se comparan comparando las medias, para ello se recurre a probar la normalidad y aplicar la comparación por medio de la prueba t-Student si los datos son normales o bien utilizar una prueba de estadística no paramétrica como la de Mann-Whitney (Altman, D.G. y Bland, J.M. 1996).

Para llevar a cabo la comparación de las medias de dos poblaciones que presentan una distribución estadística normal el procedimiento a seguir propuesto se basa en una hipótesis de grupos en la cual se procede como se detalla a continuación:

Revisar que hay un comportamiento normal en los datos utilizando la prueba Kolmogorov-Smirnov y podemos así asegurar el funcionamiento de la prueba t.

Comparar las poblaciones comparando la media de los dos grupos utilizando la prueba t para datos normales. A continuación se plantean los

fundamentos de una hipótesis de grupos con fines de comparación por equivalencia:

El problema consiste en comparar dos grupos de observaciones existentes que presentan distribución normal, como por ejemplo: producción de dos fabricas. O bien un experimento diseñado para probar algo nuevo:

población 1: situación habitual

población 2: nuevo tratamiento.

En resumen:

Población 1 con media μ_1 desviación estándar σ_1 de la que tomamos la muestra 1 cuyo tamaño es n , con media \bar{X} desviación estándar muestral \hat{S}_1

Población 2 con media μ_2 desviación estándar σ_2 de la que tomamos la muestra 2 cuyo tamaño es m , con media \bar{Y} desviación estándar muestral \hat{S}_2

La varianza en ambos grupos debe ser estudiada como un primer problema, esto es, es necesario determinar si la varianza es igual en ambas muestras para lo cual se plantea la hipótesis nula: $H_0: \sigma_1 = \sigma_2$ con la hipótesis alternativa $H_a: \sigma_1 \neq \sigma_2$.

Para ello utilizamos la prueba de la razón de varianzas o de Levene. Con varianzas iguales y siguiendo una distribución normal en ambas poblaciones la prueba espera que la razón de varianzas siga una distribución F de Snedecor con parámetros $(n-1)$ y $(m-1)$ el valor estadístico de la prueba de Levene puede ser calculado como se muestra en la siguiente ecuación (Moral, 2007):

$$F = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}{\frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2} = \frac{\hat{S}_1^2}{\hat{S}_2^2}$$

Ecuación 3.19

Se acepta la hipótesis nula reconociendo que las dos varianzas de los dos grupos de observaciones son iguales si el p-valor es mayor a 0.05 por otra parte se rechaza la hipótesis nula si el p-valor calculado es menor o igual 0.05 indicando que las varianzas son diferentes y que la variabilidad en ambos grupos es diferente.

Para la prueba t Student esto quiere decir que el estadístico a calcular variará ligeramente en función de las variabilidades muestrales.

El estadístico de la t Student a utilizar viene dado por la ecuación siguiente:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)\hat{S}_1^2 + (m-1)\hat{S}_2^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Ecuación 3.20

Otra forma de llegar a estos resultados es utilizando el cálculo de intervalos de confianza para el rango de la respuesta media en ambos grupos de observaciones. El intervalo de confianza constituye una medida de la incertidumbre con la que se estima esa diferencia a partir de la muestra, permitiendo valorar tanto la significación estadística como la magnitud de esa diferencia. En el caso de asumir la misma variabilidad en ambos grupos mediante el test de Levene ($p \geq 0.05$), el intervalo de confianza vendrá dado como:

$$(\bar{X} - \bar{Y}) \pm t_{0,975}^{n+m-2} \sqrt{\frac{(n-1)\hat{S}_1^2 + (m-1)\hat{S}_2^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}$$

Ecuación 3.21

La expresión $t_{0,975}^{n+m-2}$ representa el valor que la distribución t de Student con $n+m-2$ grados de libertad deja a su derecha el 2.5% de los datos. El intervalo de confianza para la diferencia de medias indica un rango de valores entre los que es posible encontrar el valor real de la diferencia entre las mediciones de ambos grupos.

Al considerar el valor cero dentro del intervalo estamos indicando que no se dispone de evidencia para demostrar que la variable que estamos comparando es distinta en ambos grupos de observaciones. Si asumimos distintas varianzas en los grupos ($p < 0.05$), el intervalo de confianza se expresará como:

$$(\bar{X} - \bar{Y}) \pm t_{0,975}^f \sqrt{\frac{\hat{S}_1^2}{n} + \frac{\hat{S}_2^2}{m}}$$

Ecuación 3.22

Una serie de ejemplos de aplicaciones se pueden encontrar en (Moral, 2007; Mendenhall, 1997; Montgomery, 1996)

El SPSS proporciona una facilidad para comparar dos grupos utilizando la prueba t-Student, como ejemplo consideremos los datos de la tabla 3.21 que muestra los salarios netos de los empleados de cuatro diferentes organizaciones.

No. Observación	Empresa 1	Empresa 2	Empresa 3	Empresa 4
1	16537.00	26430.00	15054.00	27995.00
2	12056.00	23679.00	19522.00	32861.00
3	32893.00	26316.00	22246.00	29831.00
4	25405.00	18379.00	13739.00	37749.00
5	24473.00	19573.00	13716.00	32266.00
6	22415.00	21404.00	15525.00	33493.00
7	27004.00	31807.00	23835.00	24137.00
8	20666.00	22358.00	20570.00	27693.00
9	33201.00	29179.00	22190.00	27410.00
10	17082.00	26390.00	17404.00	29242.00
11	19393.00	21413.00	25311.00	38521.00
12	27589.00	30140.00	17878.00	29122.00
13	27984.00	25360.00	25398.00	30422.00
14	24851.00	22972.00	19021.00	31184.00
15	21647.00	21320.00	21500.00	27219.00
16	26001.00	31040.00	15540.00	25640.00
17	15952.00	24319.00	24935.00	32044.00
18	26360.00	27730.00	21781.00	27133.00
19	22310.00	20791.00	20149.00	31895.00
20	20651.00	24449.00	17034.00	28846.00

Tabla 3.21 Salarios netos de los empleados de cuatro diferentes empresas, la empresa se muestra como una columna de la tabla

Es posible ubicar la empresa como un campo dentro de la tabla reduciendo el número de columnas y aumentando el número de observaciones como se presenta en la tabla 3.22 ambas tablas la 1.21 y la 1.22 muestran la misma información pero organizada de diferente manera, para el SPSS es más conveniente la tabla 3.22

No. Observación	Salario Neto	Empresa
1	16537.00	1
2	12056.00	1
3	32893.00	1
4	25405.00	1
5	24473.00	1
6	22415.00	1
7	27004.00	1
8	20666.00	1
9	33201.00	1
10	17082.00	1
11	19393.00	1
12	27589.00	1
13	27984.00	1
14	24851.00	1
15	21647.00	1
16	26001.00	1
17	15952.00	1

18	26360.00	1
19	22310.00	1
20	20651.00	1
21	26430.00	2
22	23679.00	2
23	26316.00	2
24	18379.00	2
25	19573.00	2
26	21404.00	2
27	31807.00	2
28	22358.00	2
29	29179.00	2
30	26390.00	2
31	21413.00	2
32	30140.00	2
33	25360.00	2
34	22972.00	2
35	21320.00	2
36	31040.00	2
37	24319.00	2
38	27730.00	2
39	20791.00	2
40	24449.00	2
41	15054.00	3
42	19522.00	3
43	22246.00	3
44	13739.00	3
45	13716.00	3
46	15525.00	3
47	23835.00	3
48	20570.00	3
49	22190.00	3
50	17404.00	3
51	25311.00	3
52	17878.00	3
53	25398.00	3
54	19021.00	3
55	21500.00	3
56	15540.00	3
57	24935.00	3
58	21781.00	3
59	20149.00	3
60	17034.00	3
61	27995.00	4
62	32861.00	4
63	29831.00	4
64	37749.00	4
65	32266.00	4
66	33493.00	4
67	24137.00	4

68	27693.00	4
69	27410.00	4
70	29242.00	4
71	38521.00	4
72	29122.00	4
73	30422.00	4
74	31184.00	4
75	27219.00	4
76	25640.00	4
77	32044.00	4
78	27133.00	4
79	31895.00	4
80	28846.00	4

Tabla 3.22 equivalente a la tabla 3.21 pero con menos columnas y más observaciones

La pregunta obligada sería cual empresa ofrece mejores salarios a sus empleados?

En esta sección nos limitaremos a comparar dos empresas, iniciamos por comprar la empresa 1 con la 2, para ellos seguimos el siguiente procedimiento en el SPSS:

Revisar Comportamiento Normal por Empresa:

1. Seleccionar la opción *Data*
2. Seleccionar *Split File* pasar a la ventana de la derecha la variable que pensamos define los grupos en este caso empresa, y activar la opción *Organize output by groups* y oprimir OK
3. Seleccionar la opción *Analyze*
4. Seleccionar *Nonparametric Tests*
5. Seleccionar *1-Sample K-S* y pasar a la ventana de la izquierda las variables que van a ser analizadas, en este caso salario y seleccionar *Test Distribution: Normal* y oprimir OK

Se obtienen los resultados resumidos en la tabla 1.23

Empresa	Media	Desviación Estándar	Significancia
1	23223.500	5464.7329	.996
2	24752.450	3857.4426	.977
3	19617.400	3793.2328	.972
4	30235.150	3659.9732	.976

Tabla 3.23 con resultados de la prueba de normalidad Kolmogorov-Smirnov para los salarios de cada una de las empresas.

Como se puede observar los salarios de cada una de las empresas tienen una distribución normal indicando que podemos utilizar la t-Student para compararlos. Es necesario demostrar que su varianza es igual, lo cual se hace simultáneamente con la prueba t.

El procedimiento de SPSS se muestra a continuación:

Primero se procede a nulificar la división del archivo, pues ya no es necesaria

1. Seleccionar la opción *Data*
2. Seleccionar *Split File* pasar a la ventana de la izquierda la variable que pensamos define los grupos en este caso empresa, y desactivar la opción *Organize output by groups*, activar la opción *Analyze all cases, do not create groups* y oprimir OK.
3. Seleccionar la opción *Analyze*
4. Seleccionar: *Compare Means*
5. Seleccionar *Independent-Samples T test* y pasar a la ventana de la derecha (*Test variable(s)*) la variable salario y en la parte inferior pasar a *Grouping Variable* la variable agrupadora en este caso empresa. Definir los grupos oprimiendo el botón *Define Groups* en este caso definir un grupo para cada empresa, o sea 1 y 2 para comparar las empresas 1 y 2 oprimir OK

Se obtienen los resultados resumidos en la tabla 1.24

		Independent Samples Test								
		Levene's Test for Equality of Variances		Independent Samples Test					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
salario	Equal variances assumed	1.986	.167	-1.022	38	.313	-1528.950	1495.7134	-4556.86	1498.963
	Equal variances not assumed			-1.022	34.168	.314	-1528.950	1495.7134	-4568.05	1510.154

Tabla 3.24 Resultados de comparar las empresas 1 y 2

La interpretación de los resultados se lleva a cabo siguiendo los lineamientos expuestos con anterioridad:

Primero es necesario revisar si las varianzas son o no iguales, eso lo podemos deducir observando la Significancia bajo La prueba de Levene dónde de lee un valor de .167 que es mayor a 0.05 y por lo que se acepta la hipótesis nula las dos varianzas son iguales, por lo tanto los valores útiles son los del primer renglón en el que la significancia etiquetada como Sig. (2-tailed) es de .313 que es mayor a 0.05 permite esto establecer que los dos grupos tienen una media igual. O se las dos empresas de estos grupos tienen los mismos salarios.

Comprando las empresas 2 y 3 obtenemos la tabla 3.25

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper	
salario	Equal variances assumed	.000	.985	4.245	38	.000	5135.0500	1209.7206	2686.099	7584.001
	Equal variances not assumed			4.245	37.989	.000	5135.0500	1209.7206	2686.076	7584.024

Tabla 3.25 Resultados de comparar las empresas 2 y 3

Las varianzas son iguales observando la Significancia bajo la prueba de Levene hay un valor de .985 que es mayor a 0.05 y por lo que se acepta la hipótesis nula las dos varianzas son iguales, los valores a interpretar son los del primer renglón en el que la significancia etiquetada como Sig. (2-tailed) es de .000 que es menor a 0.05 permite establecer que los dos grupos tienen una media distinta y el signo positivo de la diferencia de la media etiquetada como Mean Difference indica que el primer grupo tiene una media mayor al segundo. Por lo que podemos deducir que la empresa del primer grupo o sea la número 2 tiene mejores salarios que la del segundo grupo que es la empresa número 3.

Por otra parte si la prueba de normalidad falla, podemos utilizar la prueba de estadística no paramétrica de Mann-Whitney también llamada prueba U.

Su funcionamiento se basa en calcular un estadístico que se distribuye según una normal de media 0 y desviación 1. Para calcularlo se compara cada una de las observaciones de uno de los grupos a comparar con todas las observaciones del otro grupo, asignando un valor de 1 en caso de que la observación del primer grupo sea superior a la observación del segundo grupo, un valor de 0.5 en caso de empate o un valor de 0 en el otro caso. Al terminar de comparar todas las observaciones se suman los valores calculados y se mide dicho estadístico probando así la hipótesis planteada.

El SPSS permite llevar a cabo esta prueba, podemos comparar la empresa 1 y la 2 con el siguiente procedimiento:

1. Seleccionar la opción *Analyze*
2. Seleccionar: Nonparametric tests

3. Seleccionar *2 Independent Samples ...* y pasar a la ventana de la derecha (*Test Variable List*) la variable salario y en la parte inferior pasar a *Grouping Variable* la variable agrupadora en este caso empresa. Definir los grupos oprimiendo el botón *Define Groups* en este caso definir un grupo para cada empresa, o sea 1 y 2 para comparar las empresas 1 y 2 oprimir OK

Se obtienen los resultados resumidos en la tabla 1.26

Test Statistics(b)

	salario
Mann-Whitney U	169.000
Wilcoxon W	379.000
Z	-.839
Asymp. Sig. (2-tailed)	.402
Exact Sig. [2*(1-tailed Sig.)]	.414(a)

a Not corrected for ties.

b Grouping Variable: empresa

Tabla 3.26 Resultados de comparar las empresas 1 y 2 con la prueba U de Mann_Whitney.

La significancia etiquetada como *Asymp. Sig. (2-tailed)* con valor de 0.402 que es mayor a 0.05 indica que las medias son iguales, valores menores a 0.05 indican medias distintas. Por ejemplo al compara la empresa 2 con la 3 se obtiene la tabla 3.27

Test Statistics(b)

	salario
Mann-Whitney U	73.000
Wilcoxon W	283.000
Z	-3.435
Asymp. Sig. (2-tailed)	.001
Exact Sig. [2*(1-tailed Sig.)]	.000(a)

a Not corrected for ties.

b Grouping Variable: empresa

Tabla 3.27 Resultados de comparar las empresas 2 y 3 con la prueba U de Mann_Whitney.

Al comparar la empresa 2 con la 3 se obtiene una significancia de .001 que indica que las medias son diferentes.

Estas pruebas solo permiten comprar dos grupos, cuándo el número de grupos es mayor a dos es necesario recurrir al Análisis de Varianzas ANOVA por sus siglas en inglés que cubriremos en la siguiente sección.

Comparación de más de dos conjuntos de observaciones. ANOVA.

Para comparar más de dos medias se utiliza el análisis de varianza, ANOVA por sus siglas en inglés. Utilizar repetidamente la prueba t-Student no es correcto dado que la probabilidad de encontrar un estadístico significativo por azar aumentaría dado a la repetición de la prueba que se traduce en una multiplicación de la probabilidad. ANOVA permite comparar varias medias en distintas situaciones (Abraira, A. y Perez De vargas, A. 1996). Sigamos el siguiente razonamiento:

Por ejemplo supongamos que estudiamos las mediciones de aprovechamiento en distintas universidades que aplican distintos métodos de enseñanza las diferencias observadas en las mediciones se pueden clasificar en dos tipos:

1. Las que tiene su origen en el método de enseñanza
2. Las que se deben a otras causas como la posición social de los estudiantes que no están siendo estudiadas por el modelo propuesto

Si se cuenta en el ejemplo con un número k de universidades en el ejemplo ANOVA considera ambos tipos de variaciones en sus cálculos. Y plantea las siguientes hipótesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots = \mu_k$$

$$H_a : \mu_i \neq \mu_j \text{ para alguna } i \neq j$$

Si las variaciones observadas entre universidades diferentes son importantes respecto de las variaciones obtenidas dentro de las universidades se rechaza la hipótesis nula, en cambio si no son importantes las variaciones generadas por los distintos métodos en las distintas universidades se aceptará la hipótesis nula. El planteamiento de los cálculos es como sigue:

La distribución utilizada es una F considerando los siguientes grados de libertad: Para el numerador: $k-1$ y para el denominador $n-k$, n es la cantidad total de observaciones, la suma de observaciones de las k muestras o sea $n=n_1+n_2+\dots+n_k$ donde n_i es la población de la muestra i ésima muestra (Blalock, H. 1997).

La aceptación y el rechazo de la hipótesis nula queda definida por la distribución F como lo muestra la figura 3.9

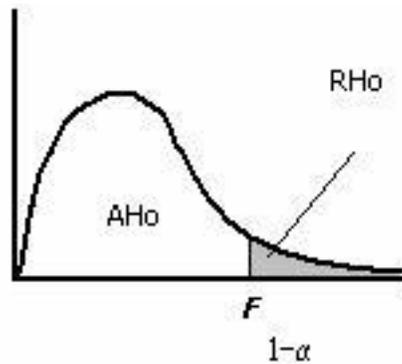


Figura 3.9 Aceptación y Rechazo de la hipótesis nula para ANOVA

Para calcular el estadístico F se cuenta con la ecuación 3.22

$$F_e = \frac{\frac{SCE}{k-1}}{\frac{SCD}{n-k}} \quad \text{Ecuación 3.22}$$

Dónde SCE es la suma cuadrática Externa (entre un grupo y otro)

SCD Suma cuadrática dentro de los grupos (Interna) y quedan definidas como:

$$SCE = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \dots + n_k(\bar{x}_k - \bar{x})^2$$

Ecuación 3.23

Dónde

$$\bar{x} = \frac{\sum_{i=1}^k \bar{x}_i \cdot n_i}{n} \quad \text{x promedio ponderada Ecuación 3.24}$$

$$SCD = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + \dots + (n_k - 1)S_k^2$$

Si el $F_e > F_t$ entonces se rechaza la hipótesis nula y existen diferencias significativas entre las k medias.

Si el $F_e < F_t$ entonces se acepta la hipótesis nula y no existen diferencias significativas entre las k medias.

Para fines prácticos se asume como la significancia de la prueba ($1 - F_t$) un valor de .05, o menor indica diferencias en las medias y se rechaza la hipótesis nula, un valor mayor a .05 indica que los grupos son iguales.

Para llevar a cabo los cálculos de ANOVA contamos con la facilidad que proporciona el SPSS que en su menú de *Análisis* cuenta con la opción de *ANOVA de un Factor*

Para que ANOVA pueda ser válida es necesario cumplir con los siguientes requisitos (Blalock, H. 1997):

- Los datos dentro de cada grupo deben mostrar normalidad, lo cual puede hacerse utilizando la prueba de Kolmogorov Smirnov que ya se discutió previamente
- Homoscedasticidad, que quiere decir Varianzas poblacionales iguales. Recordemos la prueba de Levene que automáticamente se lleva a cabo en la prueba t-Student en el SPSS
- Muestreo aleatorio.

Para demostrar el cálculo supongamos que contamos con datos de 3 universidades que utilizan tres métodos diferentes de enseñanza, por ejemplo: por casos, práctico y tradicional. Para cada una de ellas tenemos 10 observaciones como lo muestra la tabla 3.28A, se mide el aprovechamiento en forma de % del 0 al 100.

No. Observación	Por Casos	Práctico	Tradicional
1	74.58	98.60	73.30
2	99.21	64.60	39.44
3	42.81	59.41	89.81
4	50.64	72.59	55.10
5	62.73	58.43	68.73
6	62.67	59.67	88.74
7	62.48	82.30	54.58
8	79.80	30.83	30.49
9	56.03	66.40	55.26
10	88.36	87.69	64.02

Tabla 3.28A Datos para comparar métodos de enseñanza en universidades

Podemos codificar los métodos como 1, 2 y 3 respectivamente y obtener la tabla 3.28B para facilitar su análisis.

No. Observación	Aprovechamiento	Método
1	74.58	1
2	99.21	1
3	42.81	1
4	50.64	1
5	62.73	1
6	62.67	1
7	62.48	1
8	79.80	1
9	56.03	1
10	88.36	1
11	98.60	2
12	64.60	2
13	59.41	2
14	72.59	2
15	58.43	2
16	59.67	2
17	82.30	2
18	30.83	2
19	66.40	2
20	87.69	2
21	73.30	3
22	39.44	3
23	89.81	3
24	55.10	3
25	68.73	3
26	88.74	3
27	54.58	3
28	30.49	3
29	55.26	3
30	64.02	3

Tabla 3.28B Datos para comparar métodos de enseñanza codificando el método

Primero iniciemos por comprobar que los datos siguen una distribución normal, para ello dividimos el archivo atendiendo a la columna método como ya se vio anteriormente, luego procedemos a aplicar la prueba de Kolmogorov Smirnov. El procedimiento es el en el SPSS es el siguiente:

1. Seleccionar la opción *Data*
2. Seleccionar *Split File* pasar a la ventana de la derecha la variable que define los grupos en este caso método, y activar la opción *Organize output by groups* y oprimir OK.
3. Seleccionar la opción *Analyze*
4. Seleccionar *Nonparametric Tests*

5. Seleccionar *1-Sample K-S* y pasar a la ventana de la izquierda las variables que van a ser analizadas, en este caso aprovechamiento y seleccionar *Test Distribution: Normal* y oprimir OK

Se obtienen los resultados resumidos en la tabla 1.29

Método	Media	Desviación Estándar	Significancia
1	67.9310	17.43853	.733
2	68.0520	18.79739	.798
3	61.9470	19.19313	.476

Tabla 3.29 Resultados de la prueba de normalidad Kolmogorov-Smirnov para los grupos de aprovechamiento de las universidades que usan distintos métodos de aprendizaje.

Con este cálculo esto estamos seguros de que los datos cumplen el primer requisito, ya que columna significancia lo indica ya que todos sus valores son mayores que .05

El cálculo de la prueba de Levene se solicita en el proceso de ANOVA siguiendo el procedimiento del SPSS que se muestra a continuación: Este procedimiento asume que previamente se hizo una división del archivo como se indica en el procedimiento anterior.

1. Seleccionar la opción *Data*
2. Seleccionar *Split File* pasar a la ventana de la izquierda la variable que define los grupos en este caso método y desactivar la opción *Organize output by groups*, activar la opción *Analyze all cases, do not create groups* y oprimir OK.
3. Seleccionar la opción *Analyze*
4. Seleccionar *Compare Means*
5. Seleccionar *One-Way ANOVA...* y pasar a la ventana de la derecha (*Dependent List*) la variable aprovechamiento y en la parte inferior pasar a *Grouping Variable* la variable agrupadora en este caso método. Oprimir el botón *Options* para seleccionar bajo *Statistics* la opción *Descriptive* y la opción *Homogeneity of variance test* y oprimir el botón *Continue* para salir de estas opciones, es en esta última ventana dónde se solicita la prueba de Levene para homogeneidad de varianzas, de otra manera el cálculo no se hará.
6. Oprimir el botón de OK para obtener los resultados.

Las tablas 1.30, 1.31 y 1.32 muestran los resultados obtenidos.

Descriptives

aprovechamiento

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	10	67.9310	17.43853	5.51455	55.4562	80.4058	42.81	99.21
2.00	10	68.0520	18.79739	5.94426	54.6052	81.4988	30.83	98.60
3.00	10	61.9470	19.19313	6.06940	48.2171	75.6769	30.49	89.81
Total	30	65.9767	18.07650	3.30030	59.2268	72.7265	30.49	99.21

Tabla 3.30 Descriptivas estadísticas por grupo.

Test of Homogeneity of Variances

aprovechamiento

Levene Statistic	df1	df2	Sig.
.033	2	27	.967

Tabla 3.31 Prueba de Homogeneidad de varianzas de Levene significancia mayor .05 indica homogeneidad de varianzas esto es homoscedasticidad.

ANOVA

aprovechamiento

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	243.646	2	121.823	.356	.704
Within Groups	9232.386	27	341.940		
Total	9476.032	29			

Tabla 3.32 El Análisis de Varianzas ANOVA muestra que los tres grupos tienen una media igual ya que la significancia es mayor que .05.

La prueba es válida porque hay normalidad, homoscedasticidad y la significancia calculada en ANOVA es .704 valor que es mayor a .05

Nótese como el cumplir con los requisitos de la prueba es indispensable para los resultados que sean válidos.

Conclusiones.

En este capítulo se han mostrado las herramientas más importantes y ejemplos básicos de su aplicación, así como su estrategia de uso para la demostración de los diferentes tipos de hipótesis de investigación.

Se han ilustrado los cálculos con ejemplos y procedimientos concretos para el uso del SPSS en dichos cálculos, desde luego que las pruebas y herramientas se encuentran implementadas en una gran cantidad de paquetes estadísticos, el investigador puede seleccionar el que mejor se acomode a su estudio, habilidades y presupuesto, de lo que no hay duda, es

que existen paquetes de licencia GNU-GPL gratuitos para llevar a cabo los cálculos y obtener resultados casi de manera inmediata, por lo que no hay razón de dedicar esfuerzo y trabajo manual a la tarea del cálculo de los resultados.

La sobre abundancia de herramientas actúa como un catalizador de errores, éste capítulo trata de contrarrestarla proporcionando una estrategia clara de selección de herramientas, este capítulo es la vista a la cancha de fútbol, antes de ver el balón en nuestros pies.

Referencias

1. Abraira, A., Perez De vargas, A.(1996), *Métodos Multivariantes en Bioestadística*. Madrid: Ed. Centro de Estudios Ramón Areces.
2. Altman, D. G., Bland, J.M. (1995), Statistics notes: The normal distribution. *BMJ* 1995 (310), 298-298
3. Altman, D.G., Bland, J.M.(1996), Statistics notes: Transforming data. *BMJ* 1996 (312), 770-770
4. Altman, D. A.(1997), *Practical statistics for medical research*. 1th ed., repr. 1997; London: Chapman &Hall
5. Arribas, M. (2004) Diseño y Validación de Cuestionarios. *Matronas Profesión* 5 (17) 23-29
6. Blalock, H.(1997) *Estadística Social*. México: FCE
7. Briones, G. (2001) *Métodos y Técnicas para la investigación en las ciencias sociales*. México: Trillas
8. Campbell, D.T. y Stanley, J.C.(1966), *Experimental and Quasi Experimental Social Research Design*; N. York: Rand McNally Co. Academia Press
9. Cortés, F. y Rubalcava, R. M. (1987) *Métodos estadísticos aplicados a la investigación en ciencias sociales. Análisis de asociación*. México: El Colegio de México, CES.
10. Cortés, F. y Rubalcava, R. M. (1991) *Consideraciones sobre el uso de la Estadística en ciencias sociales: estar a la moda o pensar un poco*. México: El Colegio de México, CES.
11. Chow, V.T., Maidment, D.R. y Mays, L.W. (1993) *Hidrología Aplicada*. México: McGrw-Hill
12. Combe, S. A. (1979) *La Construcción de las teorías Científicas*. Buenos Aires: Nueva Visión
13. Diez Medrano, J.(1997), *Métodos de análisis causal*. Madrid: Cuadernos Metodológicos 3 / CIS.

14. Ferrand Aranaz, M.(2005) *SPSS para Windows. Programación y análisis estadístico*. Madrid: McGraw-Hill.
15. Jonson, R.(1996) *Elementary Statistics*. Pacific Grove: Duxbury
16. Kish, L. (1995) *Diseño estadístico para la investigación*. Madrid: Siglo XXI Editores, 1995.
17. Lehmann, D. R., Gupta, S. y Steckel, J.H. (1998) *Marketing Research*. Reading, Massachusetts: Addison-Wesley
18. Levi, R.I. y Rubin, D.S.(1996) *Estadística para Administradores*. New York: Prentice Hall
19. Lind, D.A., Mason, R.D. (2001) *Estadística para Administración y Economía*. Madrid: Ed. Irwin McGraw-Hill
20. Lohr, S. L. (1999) *Sampling: Design and Analysis*. Pacific Grove: Duxbury Press.
21. Luque M. (2000): *Técnicas de análisis de datos en investigación de mercados*. Madrid: Pirámide.
22. Maletta, H.(1995) *Análisis estadístico de hipótesis y modelos*. Buenos Aires: Departamento de Computación para ciencias sociales. USAL.
23. Mendenhall, W. (1997). *Probabilidad y estadística para ingeniería y ciencias*. New York: Ed. Prentice-Hall.
24. Miquel, S., Bigné, E., Lévy, J.P., Cuenca, A.C., Miquel, M. (1997) *Investigación de Mercados*. Madrid:McGraw Hill.
25. Montgomery, D. C. , Runger, G.C. (1996). *1.Probabilidad y Estadística aplicadas a la ingeniería*. New York: Ed. McGraw-Hill.
26. Mora y Araujo, M (1984): *El análisis de los datos en la investigación social*. Ed. Nueva Visión, 1984, Bs. As.
27. Moral, I. (2006). *Comparación de Medias, Métodos Estadísticos para Enfermería Nefrológica*, Consultado el 31 de Jul. 2009 de <http://www.seden.org/>
28. Muñiz, J. (2003) *Teoría Clásica de los Tests*. Madrid, España: Ediciones Pirámide
29. Oviedo, H.C., Campo-Arias, A. (2005) Aproximación al uso del coeficiente de alfa de Cronbach. *Revista Colombiana de Psiquiatría* 35 (4) 572-580
30. Pértiga, S., Fernández, P.(2001). La distribución Normal. *Cad Aten Primaria* 2001 (8), 268-274
31. Pértiga, S., Fernández, P.(2001). Representación gráfica en el análisis de datos. *Cad Aten Primaria* 2001 (8), 112-117

32. Procter & Gamble (1990) *Procter & Gamble 150 años de éxito en Marketing*. Bogotá: Editorial Norma S.A.
33. Raynal, J.A., Raynal, M.E.(2004), Cálculo de los límites de confianza de la Distribución de Probabilidad de valores extremos tipo i para dos poblaciones. *Información Tecnológica* 15 (1),87-94
34. Rossi, P. H., Wright, J.D. y Anderson, A. B.(1983) *Handbook of Survey Research*. London: Academic Press.
35. Sierra Bravo, R. (1986), *Tesis Doctorales y Trabajos de Investigación Científica*; Madrid, España: Thompson Editores
36. Thomas, H.P. (2004) *Pruebas Psicológicas*. México: El Manual Moderno
37. Velasco, F., Hernández, S.(2007). Teoría de valores extremos: Una introducción. *Revista de Ciencias Básica UJAT* 6(1), 10-16
38. Wanielista, M. (1997) *Hydrology and Water Quality Control*. (2nd. ed.). New York: Ed. Wiley

Índice del capítulo

alfa de Cronbach, 21, 24, 25, 26, 27, 59
Anderson Darling, 14
 ANOVA, 7, 8, 51, 52, 53, 54, 56, 57
clusters, 24, 39, 40
confiabilidad, 21, 23, 24, 25, 27, 29
Constructo, 5, 22
Gumbel, 15
hipótesis, 2, 4, 5, 6, 7, 8, 14, 23, 30, 31, 38, 40, 42, 43, 44, 50, 52, 53, 54, 57, 59
Hipótesis, 30

Kappa, 22
Kolmogorov-Smirnov, 13, 41, 42, 43, 49, 56
Mann-Whitney, 43, 50, 51
Normal, 12, 42, 43, 48, 56, 59
Poisson, 14
Regresión, 6, 8, 30, 35
Ryan Joiner, 14
Shapiro Wilk, 14
SPSS, 23, 24, 26, 27, 29, 31, 32, 36, 38, 40, 42, 45, 46, 48, 49, 50, 54, 55, 56, 57, 59
t-Student, 43, 45, 49, 52, 54

Capítulo 4:

Factores críticos de éxito de e-Gobierno en México

Contenido

¿Qué se encontrará en este caso?.....	130
Temática de la investigación	130
Técnicas utilizadas en el caso	130
Marco teórico	131
Definición de gobierno electrónico.....	131
Eficacia de los portales ¿Cómo medirla?.....	132
Estudios antecedentes de la investigación	135
Análisis y declaración del problema	143
Experimento	145
Preguntas de investigación.....	145
Hipótesis	146
Tamaño de la muestra	149
Diseño del instrumento.....	150
Recopilación de datos.....	153
Operacionalización de las variables.....	159
Diseño experimental	162
Planteamiento estadístico de la hipótesis	162
Aplicación de pruebas estadísticas.....	164
Comprobación de hipótesis	185
Conclusiones y recomendaciones	186
Respuestas a las preguntas de investigación	186
Conclusiones	187
Recomendaciones.....	188
Bibliografía	188
Índice del capítulo	195

¿Qué se encontrará en este caso?

Temática de la investigación

Mucho se ha analizado ya el conjunto de elementos que participan en un portal de gobierno electrónico, entre ellos aspectos técnicos que inciden en el desempeño, aspectos de estética, aspectos de contenido, aspectos que tienen que ver con transacciones en línea, etcétera. Se califica a los portales, pero no hay estudios que respondan a la pregunta, ¿qué hace que los usuarios visiten los portales gubernamentales?

El caso analiza las mediciones que se han hecho en todo el mundo respecto a la eficacia de los portales de gobierno electrónico, y explora el grado en que dichos factores ya medidos inciden en las visitas que los usuarios hacen a los portales.

Este caso puede servir de modelo para aquellas investigaciones en las cuales se desee saber cuáles son los factores críticos de éxito para un fenómeno. Una particularidad también de esta investigación es que no se aplicaron encuestas que pudieran agregar elementos subjetivos a la medición, sino que todos los datos fueron observados por el equipo investigador; en ese sentido, se propone una metodología para codificar y levantar información de campo.

Técnicas utilizadas en el caso

En el presente caso se utilizaron las siguientes pruebas y técnicas:

1. Recopilación directa intrasubjetiva de Galtung, para observación de campo.
2. Prueba Alfa de Cronbach.
3. Prueba Kolmogorov-Smirnov.
4. Matriz de correlación de Spearman y Pearson.
5. Coeficiente de correlación (r).
6. Coeficiente de correlación ajustado (r^2).
7. Factores de inflación de varianza (FIV).
8. Prueba F.
9. Análisis de varianza (ANOVA).
10. Coeficientes de regresión.
11. Selección de variables Backward.
12. Estimaciones mínimo cuadráticas.

Marco teórico

Definición de gobierno electrónico

Como organización el gobierno cuenta con recursos públicos y limitados para el desarrollo de su actividad; por ser recursos públicos, su disposición debe ser transparente, y por ser recursos limitados deben ser administrados de manera eficiente.

Una de las mayores tendencias actuales es generar alternativas más sencillas y menos costosas del quehacer del gobierno. Una de esas tendencias es la automatización de procesos y servicios gubernamentales; en un inicio la mejora consistía simplemente en automatizar procesos internos que apoyaran a los órganos gubernamentales en el desarrollo de su trabajo, sin llegar al consumidor final. Entendiendo a estos como las personas, organizaciones, instituciones, empresas, otros gobiernos. Esto era comprensible en virtud de que no existía un medio público y económico para la transmisión de datos entre el consumidor y el gobierno, e implementar redes privadas hubiera sido financieramente inviable. Con el surgimiento y la masificación de Internet el medio público de transmisión de datos entre consumidor y gobierno es una realidad, iniciándose el desarrollo alternativas tecnológicas para brindar servicios gubernamentales, dando lugar al concepto de gobierno electrónico.

Existen muchas definiciones de gobierno electrónico, de entre las cuales aportamos las siguientes por considerar que engloban los elementos que comparten la mayoría de las definiciones.

La OCDE (2003) define *gobierno electrónico* como el uso de las tecnologías de información y comunicaciones, particularmente Internet, como una herramienta para alcanzar mejores gobiernos. Holmes (2003) define al gobierno electrónico como el uso de la tecnología de la información, en particular Internet, para ofrecer servicios públicos de una manera más adecuada, conveniente, orientada a los consumidores, económica y en general más efectiva. Otra definición de gobierno electrónico (CDT, 2002) sostiene que es el uso de tecnologías de información y comunicaciones para transformar el gobierno, haciéndolo más accesible, efectivo y responsable. Bonham y Seifert (2003) definen gobierno electrónico como el uso de tecnología de la información, especialmente Internet, para liberar información gubernamental, y en algunos casos, servicios, a los ciudadanos, negocios, y otras dependencias gubernamentales. Según Kost y McClure, (2006) en un informe preparado para Gartner Group gobierno electrónico es la continua optimización de liberación de servicios, participación constitutiva y gobernabilidad, por la transformación de relaciones internas y externas a través de la tecnología, de Internet, y de nuevos medios.

Después de analizar las definiciones anteriores, defino gobierno electrónico como la opción informática de continua optimización, basada en

Internet, para la exposición de información y la prestación de servicios gubernamentales, orientados al consumidor.

Eficacia de los portales ¿Cómo medirla?

Existen en la administración dos principios fundamentales que rigen el quehacer de toda organización: la constante búsqueda de optimización en el uso de los recursos siempre limitados (principio de eficiencia), y el aseguramiento de estar realizando las acciones más adecuadas para el logro de un fin claramente determinado (principio de eficacia). La meta última es tener una eficacia eficiente.

El gobierno electrónico, tema central de este caso, representa la apuesta de innovación tecnológica más importante para la mayoría de las Entidades Federativas en México. El mejoramiento de la presencia en Internet de los gobiernos estatales es evidente e indiscutible, y ese hecho puede distraernos del hecho que la innovación conlleva riesgos inherentes a la falta de experiencia, no sólo en labores de implementación, sino también de falta de enfoque.

¿Cómo se decide el contenido y la funcionalidad del portal gubernamental en México? Todos los estados tendrán, ciertamente, una respuesta, estructurada o no. Puede ser la aplicación de las mejores prácticas de diseño, puede ser la imitación de otros portales, puede ser incluso una exigencia provocada por una Ley. Otra pregunta interesante: decidido el contenido y la funcionalidad ¿para qué?

Esas simples preguntas abren el debate respecto a los métodos y las razones que derivan en lo que uno ve en los portales de los gobiernos estatales. Como en todo sistema económico, se espera que los recursos aplicados en un servicio, en este caso el gobierno electrónico, sean acordes al beneficio o aprovechamiento que de él se obtiene. El tema ya no solamente tiene que ver con la eficiencia en el desempeño profesional de los servidores públicos encargados de los portales, sino de su responsabilidad legal de acuerdo a las leyes de responsabilidades de los servidores públicos.

Invertir en costosos servicios en Internet que nadie utiliza representa un ejercicio deficiente de recursos que ya no es sencillo mantener fuera del ojo público; hoy más que nunca los ciudadanos pueden cuestionar respecto al ejercicio de los recursos públicos y los beneficios obtenidos, en atención a las leyes de acceso a la información.

El problema radica en que los gobiernos en la actualidad no disponen de una justificación sólida para explicar sus iniciativas en materia de gobierno electrónico, ni pueden sustentar por qué decidieron la implementación de una funcionalidad en lugar de otra.

Esa necesidad de certidumbre y argumentos es lo que persigue cubrir la presente investigación, aportando conocimiento nuevo que permita, en primera instancia, señalar qué es lo más importante en un portal gubernamental en México, y en segunda instancia, estimar los resultados que cabe

esperarse de las iniciativas de gobierno electrónico mediante la aplicación dirigida de esfuerzos.

Desde luego, surge la duda: ¿cómo saber si un portal está siendo útil?

Aunque hay varias mediciones que pudieran servirnos para evaluar el éxito de un portal gubernamental, optamos por aquél uno de naturaleza cuantitativa y observable. Según la Organización para la Cooperación y el Desarrollo Económico ([OCDE], 2005), el número de visitas a los portales gubernamentales, referidos a veces como *hints*, son un indicativo del éxito de un portal gubernamental en el sentido que refleja la intención de los ciudadanos por obtener servicios y productos desde el portal de Internet. Este indicador tiene la ventaja de ser cuantitativo, además de ser provisto de forma predeterminada por las plataformas de publicación de páginas (*Web Servers*).

Pareciera obvio que los mejores portales son aquellos que provocan el mayor número de visitas, pues ello equivaldría a atender a un número mayor de servicios de atención al ciudadano. También parecería obvio que los mejores portales estuvieran mejor posicionados en los estudios comparativos (*rankings*).

La realidad nos dice otra cosa. Tomamos como referencia el documento *Estudio Comparativo de Portales Gubernamentales* (Gutiérrez, 2005), que ofrece unos resultados se repiten con pequeñas variaciones en prácticamente en todos los estudios de su clase; se asigna a cada portal una posición dentro de una lista en donde se dice cuál es el mejor y cuál es el peor, atendiendo a las características del portal.

En la figura 4.1 se compara la posición obtenida por cada uno de los portales gubernamentales de acuerdo al estudio anteriormente citado, y se compara con la posición que hubieran obtenido los estados si se hiciera una lista ordenada de los mismos portales, atendiendo al volumen de visitas que obtuvieron en 2005.

Hay casos serios de divergencia: el portal de Chiapas, por ejemplo, tiene el lugar número 3 en el estudio comparativo, pero en la lista de portales respecto a las visitas que reciben, tiene la posición 18, por lo que podría suponerse que el portal es muy bueno pero el número de visitas es muy bajo; un excelente portal que nadie visita. Caso contrario, Nayarit, que en el estudio resulta ser el peor al obtener la posición 32, en la lista de los más visitados está dentro de los 10 mejores.

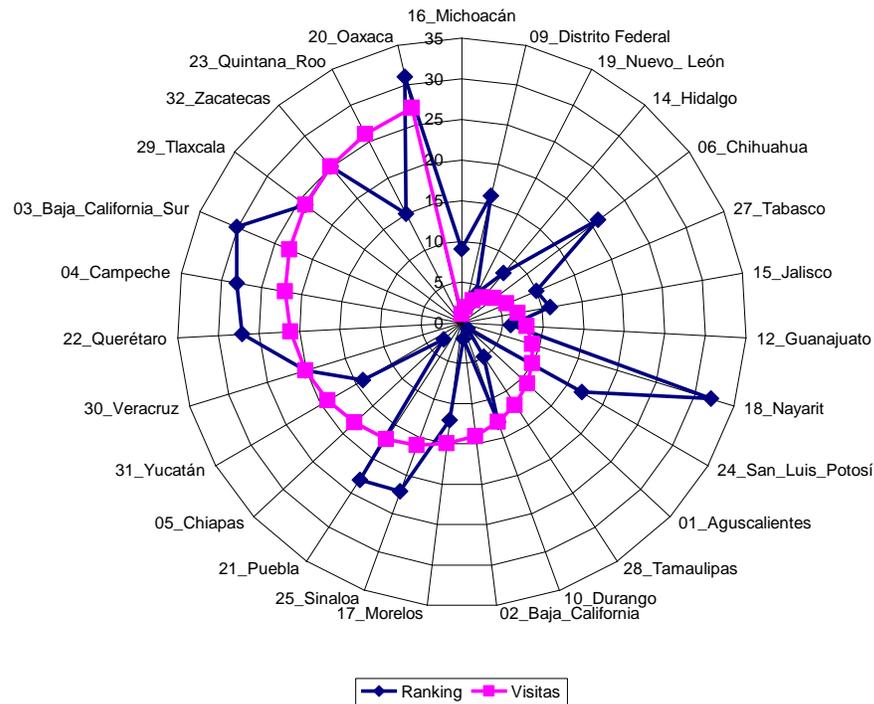


Figura 4.1. Divergencia entre visitas a los portales y su posición en el ranking de portales gubernamentales estatales. Desarrollo propio.

Quien haya estado en reuniones nacionales de organizaciones como *CIAPEM* (Comité de Informática de la Administración Pública Estatal y Municipal, A.C.), que organiza estudios comparativos de los portales gubernamentales año con año, percibirá la preocupación de los administradores de los portales por obtener una buena posición en los estudios comparativos.

A falta de otros indicadores de eficiencia y eficacia, se esfuerzan por cubrir los aspectos observados por organismos calificadores independientes. Es una tendencia, y obtener un buen posicionamiento no es necesariamente bueno, en el sentido de que pueden estar persiguiéndose de manera colectiva aspectos que no derivan en un mayor uso del portal por parte de los ciudadanos.

Resulta predecible concluir que los administradores de los portales deberían tener una gran preocupación por orientar sus esfuerzos a aquellos aspectos que incrementen el número de visitas a los portales, y evitar aquellos que las reducen. El problema es que hasta el momento nadie ha estudiado las relaciones causa efecto entre los aspectos de contenido y funcionalidad de los portales, y sus repercusiones en el volumen de visitas que reciben.

La única forma de estimular las visitas a los portales gubernamentales es conociendo qué factores, de manera individual o combinada, son los que explican dichas visitas. La presente investigación gira en torno a esa

pregunta elemental: ¿qué factor o factores son los que explican que los ciudadanos visiten un portal gubernamental?

La respuesta a esa pregunta es un gran tema olvidado en medio de otras preocupaciones. La funcionalidad de los portales gubernamentales ha estado restringida en gran medida por las posibilidades que ofrece la tecnología disponible. Estas limitantes hacen que el foco de preocupación hasta el momento haya sido crear plataformas estables y de funcionamiento consistente, dejando a un lado la finalidad última de los gobiernos, que es llevar a la ciudadanía la información que les interesa conocer, y brindarles servicios sencillos y confiables de la manera más económica posible.

El presente estudio busca cambiar el foco de atención, revelando qué factores son los que verdaderamente explican el número de visitas a los portales. No sólo analiza los aspectos técnicos del gobierno electrónico, sino que considera también aquellos factores relacionados con el contenido y el medio ambiente.

Estudios antecedentes de la investigación

El gobierno electrónico se vislumbra como la más grande área de oportunidad de los gobiernos en los próximos años.

La importancia del tema es abordado por diversos estudios a nivel mundial que analizan la forma en que el gobierno electrónico ha evolucionado. Algunos de esos estudios pretenden encontrar la fórmula para incrementar las posibilidades de éxito de las iniciativas de gobierno electrónico, así como reducir la incertidumbre respecto a su implantación.

Dentro de los estudios más importantes respecto al tema se encuentra el realizado por las Naciones Unidas ([NU], 2001), denominado *Benchmarking e-government: a global perspective*, que señala el grado de avance de las iniciativas de gobierno electrónico en todos los países miembros, realizando un estudio comparativo y documentando las mejores prácticas y experiencias a nivel mundial. Por otro lado, el estudio realizado por la OCDE (2003), denominado *The e-government imperative*, detalla los beneficios que se esperan con la implantación del gobierno electrónico, la forma en que se debe implementar y las barreras que se pueden encontrar en el proceso; los estudios realizados por la Comunidad Europea, bajo la serie *e-Government Observatory*, en donde se documenta la historia del gobierno electrónico, las estrategias y la infraestructura, de forma particular, de Suecia, Reino Unido, Portugal, Polonia, Malta, Luxemburgo, Lituania, Estonia, España, Eslovenia, Eslovaquia, Dinamarca, Chipre, República Checa, Bélgica, Austria y Alemania, recopilan las diversas formas en que los gobiernos enfrentan las iniciativas de gobierno electrónico; como esos estudios hay muchos otros.

Analizando el resultado de dichos estudios podemos concluir dos cosas: que cada país implementa el gobierno electrónico tomando como refe-

rencia prácticas generales de diseño y arquitectura de servicios en línea, y que la implementación de las mismas prácticas en diferentes países producen diferentes resultados.

El tema del gobierno electrónico genera tanto interés porque constituye una alternativa más rápida y barata de divulgar información, procesar trámites, reducir la corrupción, establecer contacto con el ciudadano y servirle, así como incrementar la participación de este último en los actos de gobierno (Holmes, 2001).

Por otro lado, lo que no se puede medir no se puede mejorar, y el gobierno electrónico no es la excepción. Antes de 2002 no tenía sentido medir el desempeño de los portales de gobierno electrónico, dado que los cambios tecnológicos hubieran provocado que los aspectos a observar cambiaran año con año. Sólo a partir del 2002 se pudieron considerar como serias las encuestas y los análisis comparativos entre portales (Kost y McClure, 2006).

Dichos estudios tuvieron un doble efecto. En primer lugar, redujeron el alto índice de expectativas que se tenía respecto al gobierno electrónico; al hacer evidentes las discrepancias y coincidencias entre los portales gubernamentales, se pudieron identificar quién iba más allá de lo posible. Esto trajo consigo la estandarización de expectativas y representó el inicio del proceso de madurez de las iniciativas de gobierno electrónico. En segundo lugar, los estudios se constituyeron como marco de referencia de lo que debía o no implementarse en un portal; si algún factor se estaba midiendo implicaba que era importante, por lo cual podía concluirse de forma lógica que mejorar el puntaje del portal en un rubro determinado llevaría a mejorar la calidad general del portal.

Junto con los estudios para evaluar y comparar portales de gobierno electrónico surgieron también manuales y guías que explicaban cómo desarrollarlos e implementarlos de la mejor manera.

Dada la corta historia del gobierno electrónico, no es de extrañar que dichos manuales y guías sean el producto de procesos de prueba y error, poco estructurados, ejecutados por personas que incluso no tenían la preparación adecuada para el desarrollo e implementación de ese tipo de iniciativas tecnológicas (Carbo y Williams, 2004).

En el mejor de los casos los proyectos de implementación de gobierno electrónico tienen como base documentos que de forma sistemática y estructurada exponen las mejores prácticas de los ambientes Web. Entre los más relevantes e influyentes podemos citar a la *Guía para desarrollo de sitios Web*, desarrollado por el Gobierno de Chile (2003), el manual *Guidelines for UK government websites, Framework for local government* desarrollados por el Gobierno del Reino Unido (2005), e incluso algunos de índole técnica, por ejemplo el manual *Standards and architectures for e-government applications*, desarrollado por el gobierno alemán (KBSt, 2003) y el manual *Core techniques for web content accessibility guidelines 1.0* de la W3C (2000), éste último en calidad de organismo que regula los aspectos formales y técnicos de la Web.

Los manuales anteriormente citados basan sus postulados en experiencias previas documentadas y definen lineamientos respecto a la forma en que se deben implementar los portales de gobierno electrónico, con el fin de reproducir la experiencia. Este hecho les otorga un carácter eminentemente empírico y práctico.

Los manuales de mejores prácticas, las guías de desarrollo e implementación, así como las evaluaciones y análisis comparativos, constituyen las fuentes documentales a partir de las cuales se conceptualiza un portal de gobierno electrónico respecto a su forma y contenido.

Las fuentes documentales sugieren fórmulas, en su mayoría rígidas, que especifican un conjunto de características que deben estar presentes para lograr éxito de las iniciativas de gobierno electrónico. Algunas de las características pueden aplicarse de forma genérica, mientras que otras sólo deberán ser aplicadas si el contexto y el tipo de audiencia lo ameritan (Carbo y Williams, 2004). Los conjuntos de características se agrupan, dando lugar a lo que se conoce como factores críticos de éxito del gobierno electrónico; en consecuencia, una alta presencia de características proporcionará buenas calificaciones respecto al factor que las agrupa. Se supone que de estar presentes las características sugeridas, estas inciden en el éxito del gobierno electrónico. La figura 4.2 muestra una esquematización de la forma en que usualmente se llega a determinar factores críticos de éxito para el análisis de gobierno electrónico.

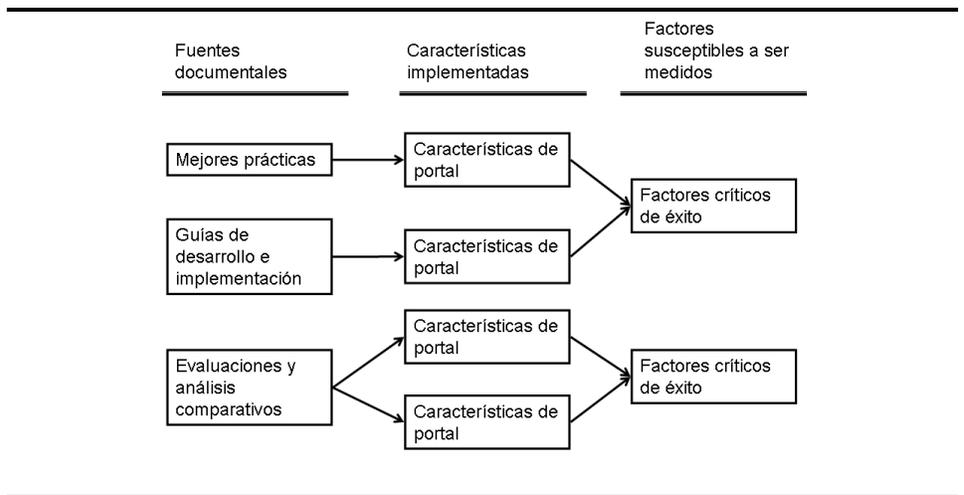


Figura 4.2. Origen de los factores críticos de éxito de los portales gubernamentales. Desarrollo propio.

El fundamento empírico de los manuales y mejores prácticas del gobierno electrónico hacen de su aplicación una estrategia poco recomenda-

ble en algunos casos, dadas las diferencias de contexto que se da entre los portales; tratar de aplicar las mejores prácticas que funcionan en un entorno citadino de riqueza y desarrollo puede no producir efectos positivos en un entorno de naturaleza rural con alto índice de pobreza.

Es posible que los portales, pretendiendo la concordancia con las fuentes documentales incluyan características que no aplican para un público objetivo determinado. Asumir que todas las recomendaciones incluidas en las fuentes documentales funcionarán con la misma efectividad en diferentes contextos sería como afirmar que los ciudadanos son homogéneos, tanto en sus actividades, costumbres, expectativas de vida y su interacción con el gobierno. El problema se hace más profundo si consideramos que muchas de las fuentes documentales tienen un fundamento empírico basado en la experiencia de países con culturas y escenarios socio-económicos diferentes.

Si las iniciativas de gobierno electrónico toman como base las guías y mejores prácticas internacionales, se debe tener conciencia que los lineamientos y sugerencias pueden haber sido determinados de forma empírica, y la experiencia considerada como válida puede no ser aplicable en el contexto en que se desea implementar. Tómese en cuenta también que entre países hay brechas de conocimiento, y las prácticas de un país pueden estar sesgadas por intereses ajenos al proyecto tales como modas administrativas o productos tecnológicos en promoción. En cualquiera de estos escenarios, la eficacia del portal se compromete.

Para evitar este tipo de ineficiencias, resulta necesario que los portales integren sólo aquellas características que aportan valor a su público objetivo. Ya no sólo se trata de implementar bien aquellas mejores prácticas documentadas, al menor costo (eficiencia), sino que es necesario entrar al análisis que nos lleve a concluir cuáles de todas las mejores prácticas deben ser implementadas, por haber demostrado que contribuyen al incremento en las visitas al portal (eficacia). En materia de gobierno electrónico, se requiere que los organismos estatales tengan una eficacia eficiente.

Los tipos de estudio realizados sobre gobierno electrónico en el mundo, son los siguientes:

1. *Benchmarking*. Tienen como finalidad comparar los sitios en torno a ciertas características específicas, ofreciendo como producto final qué sitios son los mejores y los peores, considerando las características evaluadas. Es el típico estudio de rankings. A este tipo de estudios pertenecen la mayoría de los premios que se otorgan a los sitios gubernamentales.
2. Análisis de riesgo. Tiene como finalidad analizar el grado de riesgo que implica habilitar servicios o exponer información en la Internet. Su producto final es un veredicto respecto a la conveniencia de habilitar un servicio o publicar una información. Un estudio de este tipo podría ser el analizar si conviene publicar la nómina de servidores públicos, haciendo referencia al nombre de las personas.

3. *Análisis costo-beneficio*. Tiene como finalidad analizar si es rentable, para el gobierno y para los ciudadanos, en términos de esfuerzo, tiempo, traslados, etc. la implementación de servicios o la exposición de información. Este tipo de análisis es el que se hace para justificar la conveniencia de colocar kioscos para la generación de actas de nacimiento en centros comerciales, por ejemplo.
4. *Análisis de estándares*. Tienen como finalidad valorar si un sitio se apega a un conjunto de reglas especificadas por organismos reconocidos. Un ejemplo típico es la exploración para comprobar si un sitio cumple con todos los lineamientos de la W3C para el uso de portales por parte de los discapacitados.
5. *Guías de construcción*. Son estudios que dicen la forma de construir e implementar un portal de gobierno electrónico. Generalmente constituyen una recopilación de mejores prácticas. Un ejemplo de este tipo de estudios es el *Guía para desarrollo de sitios Web*, desarrollado por el Gobierno de Chile (2003).
6. *Modelos de madurez*. Son estudios que proponen etapas o niveles en los cuales se puede encontrar un portal gubernamental, atendiendo a características específicas. El estudio de Naciones Unidas (2002) denominado *Benchmarking e-government: a global perspectiva*, ofrece un ejemplo de este tipo de estudios.
7. *Modelos de factores*. Son estudios que analizan el comportamiento que presentan los portales gubernamentales en cuanto a ciertos factores específicos. Un ejemplo de este tipo de estudios es un análisis del contenido de información de un portal, en los últimos 4 años.
8. *Mejores prácticas*. Estos estudios documentan las prácticas que han comprobado producir buenos resultados en su aplicación, en un contexto dado. Son prácticas generales, no específicas. Tienen la desventaja que algunas mejores prácticas para un país pueden no ser aplicables en otro con una estructura social y nivel de riqueza diferente.
9. *Análisis de experiencias*. Estos estudios documentan la experiencia de las iniciativas de gobierno electrónico, y los resultados que se obtienen. Este tipo de estudios no busca estimular la imitación de las prácticas, simplemente las exponen. Ejemplos de este tipo de estudios son los realizados por la Comunidad Europea, bajo la serie *e-Government Observatory*.
10. *Relaciones entre factores*. Este tipo de estudio analiza factores dispersos que componen a un portal de gobierno electrónico,

agrupando características para facilitar su análisis. Este tipo de estudios son muy útiles en el sentido que llevan a la mejor comprensión de los portales gubernamentales y su constitución. Este tipo de estudios comienza a proliferar, sobre todo en el ámbito académico.

11. *Análisis de causa efecto*. Estos estudios son los que analizan diferentes factores en función a los efectos que dichos factores tienen sobre otros. Generalmente son análisis científicos de correlación, que buscan explicar fenómenos relacionados con las iniciativas de gobierno electrónico. Este tipo de análisis comienzan a realizarse en todo el mundo, pero aún son escasos.
12. *Estudios teóricos*. Estos estudios son los que tratan propuestas teóricas que persiguen la eficiencia y eficacia del gobierno electrónico, suponiendo resultados respecto a la aplicación de marcos generales de conocimiento y técnicas probadas en otros ámbitos. Un ejemplo de esto puede ser la aplicación de modelos de calidad, que si bien no fueron diseñados específicamente para su aplicación en iniciativas de gobierno electrónico, pueden aplicarse en las mismas, esperando resultados positivos.

La figura 4.3 muestra los diferentes tipos de estudio que se han realizado en torno al gobierno electrónico en el mundo.

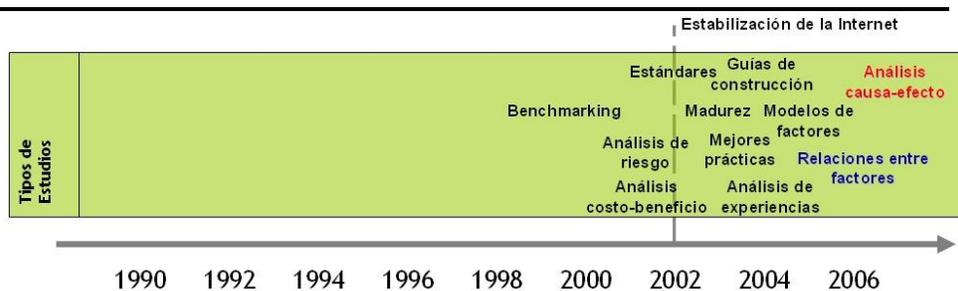


Figura 4.3. Línea de tiempo de los tipos de estudio relativos al gobierno electrónico. Desarrollo propio.

Para el presente trabajo de investigación se analizaron muchos estudios relativos al gobierno electrónico. De todos ellos se privilegiaron aquellos estudios realizados por organismos reconocidos, que además aplicarán métodos científicos para la realización de sus estudios. Se seleccionaron aquellos que mostraban ser más serios, tanto en su temática, profundidad y manejo de fuentes. También fueron considerados algunos estudios eminen-

temente teóricos que proponían algunos factores a ser considerados, que a juicio y experiencia del investigador, ameritaban ser tomados en cuenta.

La selección de los estudios a considerar como antecedentes tienen las siguientes características comunes: **a)** señalan factores observables de alguna manera, que pueden incidir en la eficacia y eficiencia del gobierno electrónico, y **b)** los factores observables son aplicables al contexto mexicano.

En la figura 4.4 se puede identificar, de arriba hacia abajo, la participación que tiene el análisis de estudios antecedentes en la selección de factores que serán considerados en la investigación. El último paso mostrado en la figura, tiene que ver ya con las pruebas estadísticas que deberán llevarse a cabo sobre los factores seleccionados.

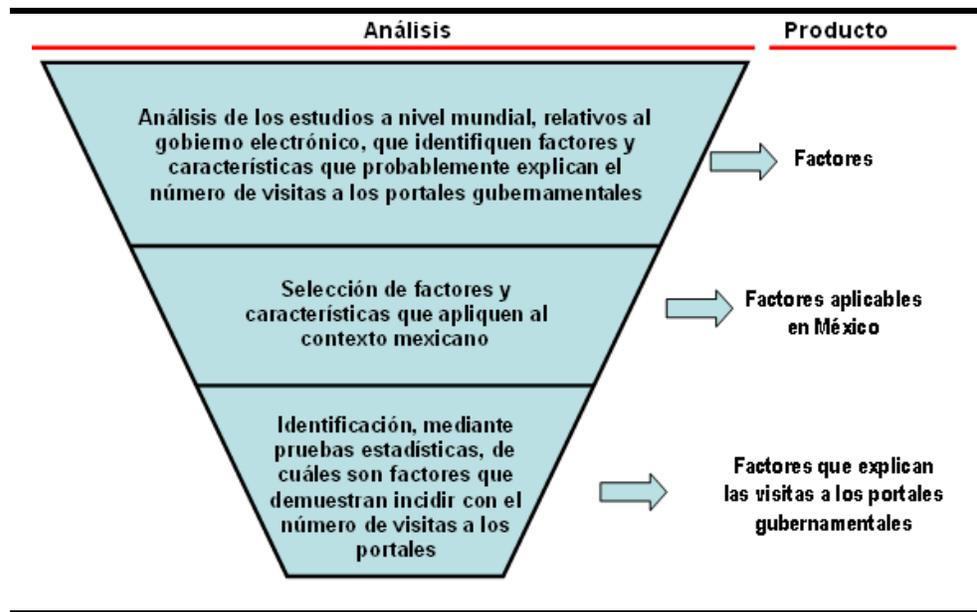


Figura 4.4. Proceso de identificación de factores que explican el número de visitas a los portales gubernamentales en México. Desarrollo propio.

Se decidió tomar en cuenta, para la selección de factores aplicables a la investigación, los estudios mostrados en la figura 4.5.

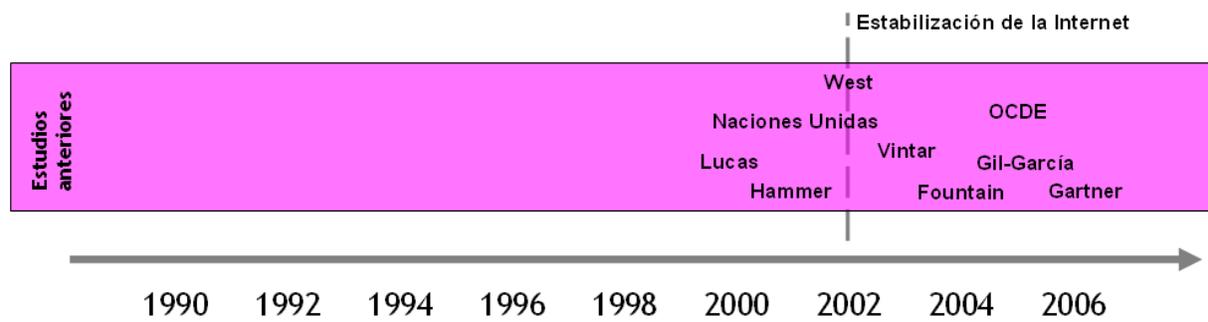


Figura 4.5. Línea de tiempo. Estudios pertinentes relativos al gobierno electrónico. Desarrollo propio.

De cada uno de los estudios analizados, se identificaron y seleccionaron aspectos observables que nos permitieran conocer los factores que explican las visitas a los portales gubernamentales.

Se trataron de cubrir las siguientes categorías de información:

1. Métrica de eficacia, que nos permite evaluar el grado en que un portal gubernamental es utilizado o no.
2. Información de contenido, que nos permite considerar como factor el grado en que el portal gubernamental contiene información que puede ser del interés de los visitantes.
3. Información de funcionalidad, que nos permite considerar como factor el grado en que los visitantes pueden desarrollar trámites ante el gobierno, desde el portal gubernamental.
4. Información de diseño, que nos permite considerar como factor la forma en que el portal gubernamental presenta los contenidos y la funcionalidad, a fin de hacer más simple y segura la experiencia del usuario.
5. Información de medio ambiente, que nos permite considerar como factor las condiciones de medio ambiente de los estados, que siendo ajenas al portal, pueden influir en el uso que se da del mismo.

Es muy importante mencionar que sólo se consideraron estas categorías por ser las aplicables al grado de madurez de los portales observados en la investigación, mismos que se encuentran entre el nivel Realzado/Mejorado (*Enhanced*), caracterizado por información actualizada con regularidad, y el nivel Interactivo (*Interactive*), porque los usuarios pueden descargar formularios, contactar funcionarios, realizar citas y requerimientos (ONU, 2002). Se dejaron fuera aspectos como servicios que requieren pago en línea, voto en línea, controles administrativos para áreas on-line, etcétera.

Se procedió a la selección de variables manejadas en la investigación, y se determinó utilizar una de tipo dependiente, Visitas al portal (Y), y 7 de tipo independiente: Soporte informativo (X1), Soporte a trámites (X2), Aplicación de mejores prácticas (X3), Situación socio-económica (X4), Densidad poblacional (X5), Mediana de la edad (X6), y Escolaridad media (X7) de la población.

Análisis y declaración del problema

El gobierno electrónico es un medio alternativo a través del cual las organizaciones del sector público pueden brindar servicios e información a los ciudadanos, utilizando la Internet como plataforma de exposición de contenidos y realización de transacciones (OCDE, 2003).

El gobierno electrónico plantea un marco de trabajo moderno para los gobiernos, lo que explica la proliferación de iniciativas en México, al menos en el ámbito Federal y Estatal.

Este entusiasmo por el gobierno electrónico, sin embargo, no se ha manifestado en la promulgación de Leyes que regulen su actividad y composición. Ante la ausencia de legislación al respecto, queda al arbitrio de cada gobierno decidir qué características deben ser incluidas en su portal, y el orden en que deben ser incluidas.

No olvidemos que los gobiernos tienen como principal obligación el perseguir el bien común (Martínez, 2000), y es deseable que lo hagan de la manera más eficiente posible; en ese sentido, se debe privilegiar la implementación en Internet aquellos servicios y productos informativos que la ciudadanía busca encontrar en dicho canal; *contrario sensu* se debe evitar la implementación de aquellos servicios y productos informativos que no son del interés de los ciudadanos, al menos desde la plataforma electrónica.

Para entender un poco el problema que la situación plantea, debemos ver a los portales gubernamentales como gasto público. De la misma manera que un gobierno no aplicaría dinero del erario en la construcción de una obra pública que nadie necesita, de la misma manera no se deberían aplicar esfuerzos ni recursos en el diseño y desarrollo de herramientas en Internet que no serán utilizados. En términos abstractos, estamos hablando de la misma cosa: empleo de recursos públicos que pretenden la máxima utilidad para el ciudadano. El problema radica en que, a la fecha, no se ha demostrado científicamente qué factores son los que realmente explican la visita de las personas los portales gubernamentales en México; en ese sentido, cualquier decisión, aunque con buenas intenciones, no deja de ser especulativa.

A nivel estatal, el uso de los portales gubernamentales no es obligatorio. Tal como sucede con los portales privados y comerciales, la decisión de los ciudadanos por utilizar los portales gubernamentales es un acto voluntario al que se verán inducidos en la medida que la alternativa en Internet provea elementos de conveniencia, tales como ahorros de tiempo total de

proceso en su interacción con el gobierno, ahorro de dinero, o comunidad al suprimir la necesidad de desplazamientos, entre otros.

Una forma cuantitativa de medir si los ciudadanos están encontrando elementos de conveniencia en la alternativa que el gobierno les ofrece para informarse y recibir servicios públicos desde su plataforma de gobierno electrónico, es contabilizar el número de visitas al portal (OCDE, 2005). Desafortunadamente, en México no existe información suficiente y confiable para señalar cuáles son los elementos de conveniencia que provocan un mayor número de visitas a los portales. En ese sentido, se asume que los visitantes utilizan la alternativa en Internet por alguna razón, pero se desconoce de forma particular qué es lo que buscan.

El hecho de que los gobiernos desconozcan cuál es el elemento o condición que provoca que los ciudadanos visiten a los portales gubernamentales, los deja sin posibilidades de establecer una estrategia que considere el orden en que deben ser aplicados los esfuerzos en materia de gobierno electrónico, dado que no saben si aquello en lo que se están aplicando tiene o no realmente un impacto positivo en el número de visitas que serán realizadas al portal.

Actualmente, decidir qué elementos incluir en un portal gubernamental, así como el orden en que se deben incluir dichos elementos, está determinado más por el cumplimiento de mejores prácticas internacionales que por el conocimiento de las necesidades de los ciudadanos.

No se han realizado en México estudios científicos que revelen las relaciones causa-efecto entre los elementos de diseño y contenido que son integrados en los portales gubernamentales y el número de visitas que los mismos producen. Tampoco se ha demostrado si factores de medio ambiente, tales como la edad de la población, su situación socio-económica y su nivel de escolaridad, influye en el número de visitas a los portales gubernamentales.

En ese sentido, los esfuerzos en materia de gobierno electrónico en México no tienen bases científicas, y su eficacia es cuestionable. Los gobiernos y sus funcionarios pueden incurrir en responsabilidad, incluso legal, por el hecho de disponer recursos del erario sin poder demostrar el beneficio público obtenido de las iniciativas de gobierno electrónico.

La declaración formal del problema es la siguiente:

“Se desconoce cuáles son los factores que explican la cantidad de visitas a los portales gubernamentales en México; tal desconocimiento evita que se puedan planear estrategias de gobierno electrónico eficaces, que privilegien el desarrollo inicial de los aspectos que estimulan las visitas al portal.”

Experimento

Un *experimento* es un procedimiento través del cual se pretenden resolver dudas de conocimiento, lo que nos lleva a proponer una o varias hipótesis relacionadas con un fenómeno que es necesario comprobar (confirmar o verificar), mediante la manipulación de datos contenidos en variables que presumiblemente están relacionados, aplicando para ello técnicas estadísticas que permiten identificar y cuantificar las causas de un efecto.

Esta simple definición nos deja claro varias cosas respecto a un experimento:

- | | |
|--|---|
| 1. Pretender resolver dudas de conocimiento | a. Planteamiento de preguntas de investigación |
| 2. Es un procedimiento para comprobar una o varias hipótesis relacionadas con un fenómeno | a. Planteamiento de hipótesis de investigación que pretendan responder las preguntas de investigación |
| 3. Se requieren datos que serán manipulados para la obtención de conclusiones. | a. Determinación del tamaño muestra.
b. Diseño del Instrumento
c. Recopilación de Datos. |
| 4. Los datos deben ser traducidos a variables | a. Operacionalización de Variables |
| 5. Se deben aplicar técnicas estadísticas para identificar y cuantificar las causas de un efecto (diseño experimental) | a. Planteamiento estadístico de la hipótesis
b. B. aplicación de pruebas estadísticas
c. Comprobación de hipótesis
d. Desarrollo de conclusiones |

Preguntas de investigación

El análisis del marco teórico y los estudios antecedentes nos dejan en posición de plantearnos dudas de conocimiento respecto a lo que no se ha investigado, o lo que se ha investigado a una profundidad distinta a la necesaria, o bajo un contexto que no es específicamente el que nos interesa.

Las dudas que serán la fuente para la propuesta de hipótesis, se les llama *preguntas de investigación*. Para este caso, las preguntas de investigación son las siguientes.

PREGUNTA CENTRAL DE INVESTIGACIÓN

- P1: Los estudios a nivel mundial, relativos al gobierno electrónico, señalan como factores a considerar para medir la madurez y eficacia de los portales de gobierno electrónico, al soporte informativo, al soporte a trámites y a la aplicación de mejores prácticas mundiales; también señalan como relevantes a las condiciones de medio ambiente, tales como la situación socio-económica, la densidad poblacional, la edad, y la escolaridad; ¿son esos factores estadísticamente significativos para explicar el número de visitas a los portales gubernamentales en México?

PREGUNTAS ESPECÍFICAS DE INVESTIGACIÓN

- P2: De los factores a considerar para medir la madurez y eficacia de los portales de gobierno electrónico sugeridos por los estudios realizados a nivel mundial, ¿cuáles de ellos, de manera individual o conjunta, explican en mayor medida el número de visitas a los portales gubernamentales en México?
- P3: Resuelta la pregunta específica de investigación anterior, ¿en qué grado explican el número de visitas a los portales gubernamentales en México los factores identificados como positivamente más influyentes?

Hipótesis

Las dudas de conocimiento nos llevan a plantearnos *hipótesis* que, de ser comprobadas, constituirán conocimiento nuevo, obtenido científicamente, que de respuesta a las preguntas de investigación.

HIPÓTESIS PARA EL MODELO COMPLETO

Relacionadas con la duda respecto a qué factores de los propuestos por la teoría y los estudios antecedentes podemos considerar para la investigación, dada la disponibilidad y aplicación en el contexto mexicano, podemos plantear las siguientes hipótesis:

- Ho1: El número de Visitas al portal gubernamental (**Y**) no es explicado por los factores, Soporte informativo (**X1**), Soporte a trámites (**X2**), Aplicación de mejores prácticas (**X3**), Situación socio-económica (**X4**), Densidad poblacional (**X5**), Mediana de la edad (**X6**), y Escolaridad media (**X7**) de la población, de manera conjunta.
- Ha1: El número de Visitas al portal gubernamental (**Y**) es explicado por los factores, Soporte informativo (**X1**), Soporte a trámites (**X2**), Aplicación de mejores prácticas (**X3**), Situación socio-económica (**X4**), Densidad poblacional (**X5**), Mediana de la edad (**X6**), y Escolaridad media (**X7**) de la población, de manera conjunta.

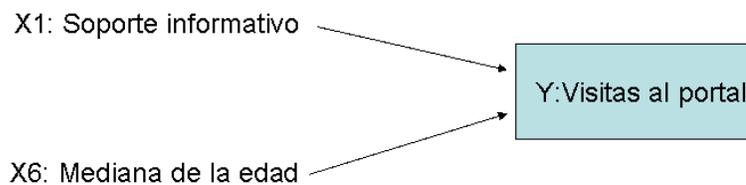


HIPÓTESIS PARA EL MODELO ÓPTIMO

El modelo completo (que involucra a todos los factores), puede no ser el óptimo para explicar las visitas al portal. No toda la información disponible y aplicable al contexto mexicano es necesariamente importante. Relacionadas con la selección de factores que de una manera más eficiente determinen a la variable dependiente, las hipótesis son las siguientes.

Ho2: El número de Visitas al portal gubernamental (Y) no es explicado en mayor medida por los factores Soporte informativo (X1) y Mediana de la edad (X6), de manera conjunta.

Ha2: El número de Visitas al portal gubernamental (Y) es explicado en mayor medida por los factores Soporte informativo (X1) y Mediana de la edad (X6), de manera conjunta.



HIPÓTESIS DE ESTIMACIÓN

Seleccionados los factores que de mejor manera explican a las visitas a los portales, se realiza una estimación del grado en que lo hacen.

Ho3: El Contenido informativo (X1) en conjunto con la Edad de la población (X6) de la Entidad Federativa, explican el 35% o menos de las Visitas al portal gubernamental (Y) de dicha Entidad.

Ha3: El Contenido informativo (**X1**) en conjunto con la Edad de la población (**X6**) de la Entidad Federativa, explican más del 35% de las Visitas al portal gubernamental (**Y**) de dicha Entidad.

Habiendo revisado preguntas de investigación y las hipótesis, podemos integrarlos en la Tabla 4.1, para revisar su congruencia. Se recomienda elaborar una tabla de congruencia que nos permita verificar de manera simple que las hipótesis tienen que ver con las preguntas de investigación; en caso de que se tengan objetivos de investigación, agregarlos también a la tabla, para ver que todo esté relacionado.

Preguntas de investigación	Hipótesis de la investigación
<p>1. P₁: Los estudios a nivel mundial, relativos al gobierno electrónico, señalan como factores a considerar para medir la madurez y eficacia de los portales de gobierno electrónico, al soporte informativo, al soporte a trámites y a la aplicación de mejores prácticas mundiales; también señalan como relevantes a las condiciones de medio ambiente, tales como la situación socio-económica, la densidad poblacional, la edad, y la escolaridad; ¿son esos factores estadísticamente significativos para explicar el número de visitas a los portales gubernamentales en México? ¿Cuáles de esos factores de manera individual o conjunta, explican en mayor medida el número de visitas a los portales gubernamentales en México?</p> <p>2. P₂: ¿De los factores a considerar para medir la madurez y eficacia de los portales de gobierno electrónico sugeridos por los estudios realizados a nivel mundial, cuáles de ellos, de manera individual o conjunta, explican en mayor medida el número de visitas a los portales gubernamentales en México?</p> <p>3. P₃: Resuelta la pregunta específica de investigación anterior, ¿en qué grado explican el número de visitas a los portales gubernamentales en México los factores identificados como positivamente más influyentes?</p>	<p>1. Ha₁: El número de Visitas al portal gubernamental (Y) es explicado por los factores, Soporte informativo (X₁), Soporte a trámites (X₂), Aplicación de mejores prácticas (X₃), Situación socio-económica (X₄), Densidad poblacional (X₅), Mediana de la edad (X₆), y Escolaridad media (X₇) de la población, de manera conjunta.</p> <p>2. Ha₂: El número de Visitas al portal gubernamental (Y) es explicado en mayor medida por los factores Soporte informativo (X₁) y Mediana de la edad (X₆), de manera conjunta.</p> <p>3. Ha₃: El Contenido informativo (X₁) en conjunto con la Edad de la población (X₆) de la Entidad Federativa, explican más del 35% de las Visitas al portal gubernamental (Y) de dicha Entidad.</p>

Tabla 4.1.

Matriz de congruencia entre preguntas de investigación e hipótesis. Desarrollo propio.

Tamaño de la muestra

En la presente investigación se intentó un trabajo de tipo censal, pero a falta de respuesta de algunos estados respecto al valor de la variable dependiente, no fue posible llegar a ese nivel de recopilación de datos. En ese sentido, se opta por el cálculo de *muestra estadística aleatoria simple*, considerando una población finita y conocida.

La población total (**N**) es de 32 elementos, siendo uno por cada uno de los estados. Sólo se pudieron recopilar en trabajo de campo 27 elementos (**n**) de los 32 posibles, lo que constituye el 84.37% de la población.

Para obtener **Z** (*coeficiente de confianza*), tomamos en cuenta el nivel de significancia deseado (**α**) para la investigación, que es de 5% (0.05).

Lo que se buscaría en seguida es el valor de **Z** para una variable aleatoria **z** tal que el área simétrica bajo la curva normal desde **-Z** hasta **Z** sea igual a **1-α** (0.95).

Se busca en la *Tabla de distribución normal* (Ritchey, 2002) el valor de **P** (**α/2**); siendo **P**=0.025, le corresponde un valor de **Z** de 1.96.

La variabilidad positiva (**p**) se maneja de manera conservadora, dado que los estudios antecedentes no sugieren algún valor aproximado del parámetro; se considera un valor de 50% (0.50), dando un lugar a un valor de la *variabilidad negativa* (**q**) sea 50% también (**1-p**).

La *precisión* (**E**) que deseamos para el cálculo de la muestra es del 10%, distribuido en ambos extremos de la densidad de distribución.

Para calcular el tamaño de la muestra, utilizamos la siguiente fórmula (Ritchey, 2002):

$$n = \frac{Z^2 pqN}{NE^2 + Z^2 pq}$$

$$n = \frac{(1.96^2)(0.5 * 0.5)(32)}{(32 * 0.10^2) + (1.96^2 * (0.5 * 0.5))}$$

$$n = \frac{30.7328}{1.2804} = 24.02$$

Dadas las especificaciones, una muestra de 24 elementos nos garantiza un nivel de confianza del 95%, tomando en cuenta una precisión del 10% y una proporción esperada del 50%.

En ese sentido, el número de elementos disponibles ($n=27$), excede la muestra requerida por la investigación, mejorando la precisión, dado que el tamaño de la muestra nos permite alcanzar un error aceptado (E) de 7.45%, en lugar del 10% tomado inicialmente para el cálculo de la muestra.

El despeje de la fórmula para el cálculo de la muestra aleatoria simple nos permite calcular E de la siguiente manera:

$$E = \sqrt{\frac{\left(\frac{Z^2 pqN}{n}\right) - (Z^2 pq)}{N}}$$

$$E = \sqrt{\frac{\left(\frac{(1.96^2)(0.5 * 0.5)(32)}{27}\right) - (1.96^2 * (0.5 * 0.5))}{32}}$$

$$E = \sqrt{\frac{\left(\frac{30.7328}{27}\right) - (0.9604)}{32}} = 0.074551$$

Diseño del instrumento

Diseño del instrumento

El diseño del instrumento es quizá la parte más laboriosa de la investigación, pues implica determinar el mínimo de datos suficientes y necesarios para poder llegar a conclusiones. El arte está en no recopilar información excesiva, ni tampoco insuficiente.

En el presente caso se buscó cumplir con dos premisas:

1. Que las variables representaran a las categorías de información que los estudios señalan como importantes: Eficacia, Contenido, Funcionalidad, Diseño, y Medio ambiente.

2. Que las variables estuvieran soportadas teóricamente, es decir, que no se incluyeran variables al arbitrio de investigador.

COBERTURA DE LAS CATEGORÍAS DE INFORMACIÓN

Las categorías de información quedaron cubiertas con las variables involucradas en las hipótesis. Vea la tabla 4.2.

Variable / Categoría	Métrica de eficacia	Información de contenido	Información de funcionalidad	Información de diseño	Información de medio ambiente
Visitas al portal (Y)					
Soporte informativo (X ₁)					
Soporte a trámites (X ₂)					
Aplicación de mejores prácticas (X ₃)					
Situación socio-económica (X ₄)					
Densidad poblacional (X ₅)					
Mediana de la edad (X ₆)					
Escolaridad media (X ₇)					

Tabla 4.2
 Cobertura de categorías de información con variables seleccionadas. Desarrollo propio.

FUNDAMENTO TEÓRICO DE LAS VARIABLES

Es importante verificar también que todas las variables tengan un sustento teórico. En caso de que las variables sean propuestas por nosotros, hay que hacer la anotación.

La tabla 4.3 muestra el fundamento teórico de las variables seleccionadas para la presente investigación, considerando los estudios antecedentes analizados.

Variables, Indicadores y datos seleccionados	Lucas (1999)	Naciones Unidas (2002)	Hammer (2001)	West (2005)	Vintar, Kunstelj, Dečman, Berčič (2003)	Fountain (2004)	OCDE (2005)	Gutiérrez (2005)	Gil-García (2006)	Gartner (2006)	Felipe Ramírez (Investigador)
(Y) Visitas al portal							■				
andamos com(X1) Soporte informativo				■	■			■		■	
(X2) Soporte a trámites							■				
(X3) Aplicación de mejores prácticas	■			■		■		■		■	
(X4) Situación socio – económica		■							■		
(X5) Densidad poblacional		■									
(X6) Mediana de la edad		■									
(X7) Escolaridad media		■									

Tabla 4.3

Fundamento teórico de las variables seleccionadas en la investigación. Desarrollo propio.

Recopilación de datos

Una vez que comprobamos que todas las categorías de información están cubiertas, debemos determinar la forma de proporcionar valor a las variables.

Para la obtención de datos utilizados en las pruebas estadísticas de la presente investigación, se recurrió a los siguientes métodos:

- a) Solicitudes formales de información.
- b) Observación de campo.
- c) Consulta a fuentes formales de información estadística.

SOLICITUDES FORMALES DE INFORMACIÓN

Para obtener la información de la variable Visitas al portal (Y), se realizó la petición formal mediante carta a los Administradores de los portales gubernamentales de los estados de la República Mexicana, entre el 1 de Marzo y el 31 de Abril de 2006. La petición puntual fue responder a las siguientes preguntas:

“1.- ¿Cuál es el número visitas (Hints) registradas al portal del gobierno del estado, del 1 de enero al 31 de diciembre de 2005?

(Utilizar cifra reportada por las herramientas de administración del Web Server)

2.- ¿Qué motor de procesamiento Web utilizan en el portal del Gobierno del Estado?”

La primera pregunta tenía como finalidad conocer el número de visitas al portal, utilizado como variable dependiente en la investigación. La segunda pregunta se utilizó para verificar la validez de la fuente para la primera pregunta. Se encontró que los estados utilizaban como motor Web de procesamiento Internet uno de los siguientes dos productos: Internet Information Server (plataforma Microsoft, páginas ASP), y Apache (plataforma abierta, páginas PHP). Ambos productos ofrecen herramientas de administración que proporcionan el número de **hints** por periodo, significando lo mismo en ambos productos. Con esta petición se recopilaron 17 datos, de 32 posibles (**n=17, N=32**), del 1 de Marzo al 8 de Junio de 2006.

Posteriormente, la revista Política Digital, el Comité de Informática de la Administración Pública Estatal y Municipal, A.C. (CIAPEM), y el Gobierno

del Estado de Nuevo León, llevaron a cabo la Primera Reunión Nacional de Administradores de Portales Gubernamentales, el 8 de junio de 2006, en la Ciudad de Monterrey, donde se aprovechó el foro para la solicitud personal de la información de los estados que no habían atendido el requerimiento por escrito. Con este esfuerzo se obtuvieron del 8 de Junio al 31 de Agosto de 2006, 6 datos más, sumando 23 datos de 32 posibles (**n=23, N=32**).

Del 1 de septiembre al 31 de Diciembre de 2006 se estuvieron realizando llamadas telefónicas y enviado correo electrónicos a los Administradores de los portales gubernamentales de los estados de la República que no hubieran aportado información, con lo que se llegó a 26 datos obtenidos. El último dato se recibió en Febrero de 2007, dándose por cerrada la recepción de datos, obteniéndose 27 datos de 32 posibles (**n=27, N=32**).

OBSERVACIONES DE CAMPO

Para obtener la información de las variables Soporte informativo (**X1**), Soporte a trámites (**X2**) y Aplicación de mejores prácticas (**X3**), se utilizaron técnicas de recopilación de información en campo.

La metodología que se utilizó para la recopilación de datos en campo fue de observación directa, de tipo intrasubjetiva (Galtung, 1966). De acuerdo a esa técnica, el un grupo investigador observa y recoge datos mediante su propia observación, partiendo del supuesto que diferentes observaciones deben producir los mismos resultados.

Para la recopilación de datos se utilizó un instrumento, en donde se colocaron preguntas de sí o no (presencia / ausencia), que a fin de cuentas se procesaban para la determinación del valor de la variable.

A continuación se muestra la sección del instrumento que se encargaba de obtener información observada, relativa a la variable (**X1**) Soporte informativo.

(X1) v **SOPORTE INFORMATIVO:** Esta variable indica la capacidad y el compromiso del portal para proveer información a los visitantes.

(X11) i **Contenido informativo:** Este indicador muestra el grado en que el portal provee la información que de acuerdo a la Ley Federal de Transparencia y Acceso a la Información, es obligatorio que toda unidad gubernamental reporte.

Código / tipo	Item	SI	NO	Valor
(X1101) m	¿El portal ofrece como información la Constitución Política del Estado?	<input type="radio"/>	<input type="radio"/>	
(X1102) m	¿El portal ofrece como información las Convocatorias a licitación?	<input type="radio"/>	<input type="radio"/>	
(X1103) m	¿El portal ofrece como información los Organigramas y funciones de las dependencias?	<input type="radio"/>	<input type="radio"/>	
(X1104) m	¿El portal ofrece como información los Presupuestos aprobados?	<input type="radio"/>	<input type="radio"/>	
(X1105) m	¿El portal ofrece como información la Cuenta Pública?	<input type="radio"/>	<input type="radio"/>	
(X1106) m	¿El portal ofrece como información el Balance general?	<input type="radio"/>	<input type="radio"/>	
(X1107) m	¿El portal ofrece como información el estado de pérdidas y ganancias?	<input type="radio"/>	<input type="radio"/>	
(X1108) m	¿El portal ofrece como información la Nómina de servidores públicos?	<input type="radio"/>	<input type="radio"/>	
(X1109) m	¿El portal ofrece como información la Relación de pagos a contratistas y proveedores?	<input type="radio"/>	<input type="radio"/>	
(X1110) m	¿El portal ofrece como información el estado que guardan los Trámites?	<input type="radio"/>	<input type="radio"/>	
(X1111) m	¿El portal ofrece como información la Difusión artística y cultura al día?	<input type="radio"/>	<input type="radio"/>	

(X12) i **Marco legal de acceso a la información:** Este indicador muestra el compromiso que el Gobierno Estatal de proporcionar información, en términos del marco legal que lo regule, y las instituciones a través de las cuales pueda hacerse valer la Ley. (Revisar la legislación vigente en el Estado, y Ley Orgánica del Poder Ejecutivo)

Código / tipo	Item	SI	NO	Valor
(X1201) m	¿Se dispone de Ley de transparencia y acceso a la información?	<input type="radio"/>	<input type="radio"/>	
(X1202) m	¿Se dispone de un organismo para resolver asuntos de transparencia y acceso a la información?	<input type="radio"/>	<input type="radio"/>	

Variables, Indicadores y datos seleccionados	Lucas (1999)	Naciones Unidas (2002)	Hammer (2001)	West (2005)	Vintar, Kunstej, Decman, Bercic (2003)	Fountain (2004)	OCDE (2005)	Gutiérrez (2005)	Gil-García (2006)	Gartner (2006)	Felipe Ramírez (Investigador)
El portal ofrece como información la Cuenta Pública											
El portal ofrece como información el Balance general											
El portal ofrece como información el estado de pérdidas y ganancias											
El portal ofrece como información la Nómina de servidores públicos											
El portal ofrece como información la Relación de pagos a contratistas y proveedores											
El portal ofrece como información el estado que guardan los Trámites											
El portal ofrece como información la Difusión artística y cultura al día											
Se dispone de Ley de transparencia y acceso a la información											
Se dispone de un organismo para resolver asuntos de transparencia y acceso a la información											

Tabla 4.4
Selección de variables. Desarrollo propio.

La recopilación de datos se llevó a cabo del 1 de Enero al 31 de Marzo de 2006, por 3 observadores (**01**, **02**, y **03**).

El investigador explicó a los observadores la mecánica de recopilación. Todas las preguntas del instrumento HP01 son valores binarios, en los que se debía anotar si las características señaladas estaban o no presentes en el portal gubernamental. Un ejemplo sería: (**X1101**) m “El portal ofrece como información la Constitución Política del Estado”. En aquellos casos que dentro del contenido del portal estuviera publicada la Constitución Política del Estado, se marcaba en el instrumento la opción “SI”, codificada con valor 1, y si no estaba publicada, se marcaba en el instrumento la opción “NO”, codificada con valor de 0. Dichos valores son utilizados de acuerdo a la formulación de las variables del instrumento.

Se utilizaron 3 observadores con la intención de minimizar al máximo la posibilidad de una observación errónea; cada observador actuó como revisor de los otros dos. Las observaciones de los portales gubernamentales se realizaban en las mismas fechas, y en caso de que las respuestas de los 3 observadores no fueran idénticas, en una sesión conjunta con el investigador se repetía la observación, y se dictaminaba la respuesta correcta de forma colegiada.

CONSULTA A FUENTES FORMALES

Para obtener la información de las variables de medio ambiente, a saber, Situación socio - económica (**X4**), Densidad poblacional (**X5**) y Mediana de la edad (**X6**), y Escolaridad media (**X7**), se consultaron datos del Instituto Nacional de Geografía e Informática, tomando como base los resultados del *XII Censo General de Población y Vivienda, 2000* (INEGI, 2001) y el *II Censo de Población y Vivienda 2005* (INEGI, 2006).

Los detalles metodológicos de la obtención de la información consultada del INEGI, puede ser revisados en la página Web del instituto www.inegi.gob.mx.

Los datos considerados son los siguientes:

1. Para Situación socio - económica (**X4**), se consideró el informe elaborado por el Centro de Estudios de las Finanzas Públicas de la H. Cámara de Diputados, con datos del INEGI, *Sistema de Cuentas Nacionales, Producto Interno Bruto por Entidad Federativa (pesos por habitante a precios de 1993); Censo de Población y Vivienda 1995* y el *Censo General de Población y Vivienda 2000 y 2005*.
2. Para Densidad poblacional (**X5**) se consideró el informe *Densidad de población por entidad federativa, 2000*.
3. Para Mediana de la edad (**X6**) se consideró el informe *Edad mediana por entidad federativa según sexo, 2005*.

4. Para Escolaridad media (X7) se consideró el informe *Promedio de escolaridad de la población de 15 y más años por entidad federativa según sexo, 2000*.

Operacionalización de las variables

De nada sirve que recopilemos datos si no está en la forma y la cantidad en la que las pruebas estadísticas las requieren para poderlas procesar. Se conoce como operacionalización de variables al procedimiento a través del cual se traduce la información recopilada en los instrumentos, en valores proporcionados a las variables, para su tratamiento estadístico.

Es muy probable que a partir de varias preguntas en el instrumento, se obtenga el valor a una variable, a lo que se le da el nombre de *constructo*. Matemáticamente debe encontrarse la manera de traducir las respuestas codificadas y generar un valor que sea determinado de la misma forma para todas las ocurrencias de datos recopilados, cuidando que el valor obtenido sea el más recomendable para las pruebas estadísticas a utilizar.

En la sección anterior se revisó la porción del instrumento en donde se hacían las observaciones relativas a (X1) Soporte informativo. Desde luego, no se puede procesar el conjunto de sí o no mediante un tratamiento estadístico: es necesario traducirlo a valores manejables, con los que podamos operar.

Cada ítem del instrumento deberá estar codificado. En el caso, responder SI, es igual a 1, y NO, es igual a 0, por ejemplo. Se entiende por *valor codificado* al valor equivalente que se obtiene de interpretar una respuesta del instrumento. Por ejemplo en un instrumento se pueden tener para una pregunta las respuestas alto, medio, bajo que codificado podría ser un valor estadísticamente procesable por ejemplo 1, 0.5 y 0, respectivamente. Por otro lado, el valor operacional es el resultado de procesar o interpretar uno o más valores codificados, para obtener el valor al asignar una variable utilizada en pruebas estadísticas. Si un ítem es el único que representa el valor de una variable, el valor codificado es el valor operacional. Si dos o más ítems representan el valor de una variable, se trata de un constructo, y es necesario definir con claridad la formulación a través de la cual las respuestas codificadas, en su conjunto, darán lugar al valor operacional.

Veamos el ejemplo de la operacionalización de la variable (X1) Soporte informativo.

(X1) Soporte informativo

La variable independiente llamada Soporte Informativo, identificada como X1, es un variable continua que representa la capacidad y el compromiso del portal para proveer información y contenidos a los visitantes. Su valor es entre 0% y 100%.

Los indicadores que permiten el cálculo de esta variable son:

- X11 (i) Contenido informativo: Este indicador muestra el grado en que el portal provee la información que de acuerdo a la Ley Federal de Transparencia y Acceso a la Información, es obligatorio que toda unidad gubernamental reporte. Este indicador proporciona el 50% del valor de la variable.
- X12 (i) Marco legal de acceso a la información: Este indicador muestra el compromiso que el gobierno estatal de proporcionar información, en términos del marco legal que lo regule, y las instituciones a través de las cuales pueda hacerse valer la Ley. Para obtener esta información, es necesario consultar la legislación estatal vigente, ya sea en el Congreso Estatal, o en recopilaciones nacionales; es necesario inclusive consultar la Ley Orgánica del Poder Ejecutivo de los estados. Este indicador proporciona el 50% del valor de la variable.

Cada uno de estos indicadores obtiene su valor a través de un cálculo que involucra varios ítems. Dichos ítems son en realidad valores binarios (0,1), recopilados por el instrumento de trabajo de campo, e indican si el portal tiene o no una característica o contenido específico.

La formulación de la variable es la siguiente:

(X1) v-Soporte informativo	
(X11) i- Contenido in- formativo	(X1101) m-El portal ofrece como información la Constitución Política (Valor binario, 0=NO, 1=SI)
	(X1102) m-El portal ofrece como información las Convocatorias a licitación (Valor binario, 0=NO, 1=SI)
	(X1103) m-El portal ofrece como información los Organigramas y funciones de las dependencias (Valor binario, 0=NO, 1=SI)
	(X1104) m-El portal ofrece como información los Presupuestos aprobados (Valor binario, 0=NO, 1=SI)
	(X1105) m-El portal ofrece como información la Cuenta Pública (Valor binario, 0=NO, 1=SI)
	(X1106) m-El portal ofrece como información el Balance general (Valor binario, 0=NO, 1=SI)
	(X1107) m-El portal ofrece como información el estado de pérdidas y ganancias (Valor binario, 0=NO, 1=SI)
	(X1108) m-El portal ofrece como información la Nómina de servidores públicos (Valor binario, 0=NO, 1=SI)
	(X1109) m-El portal ofrece como información la Relación de pagos a contratistas y proveedores (Valor binario, 0=NO, 1=SI)
	(X1110) m-El portal ofrece como información el estado que guardan los Trámites (Valor binario, 0=NO, 1=SI)
(X12) i-Marco legal de acceso a la información	(X1111) m-El portal ofrece como información la Difusión artística y cultura al día (Valor binario, 0=NO, 1=SI)
	(X1201) m-Se dispone de Ley de transparencia y acceso a la información (Valor binario, 0=NO, 1=SI)
	(X1202) m-Se dispone de un organismo para resolver asuntos de transparencia y acceso a la información (Valor binario, 0=NO, 1=SI)

Código	Tipo	Fórmulas (Variable <- Indicadores, Indicadores <- Items)	Unidad de medida
X1	Variable	$X1 = (X11 + X12) * 100$	Proporción de características presentes
X11	Indicador	$X11 = ((X1101 + X1102 + X1103 + X1104 + X1105 + X1106 + X1107 + X1108 + X1109 + X1110 + X1111)/11)*(1/2)$	Proporción de características presentes
X12	Indicador	$X12 = ((X1201 + X1202)/2)*(1/2)$	Proporción de características presentes

Diseño experimental

Planteamiento estadístico de la hipótesis

Así como los datos no pueden ser utilizados en la forma en que fueron recopilados, así las hipótesis no pueden ser utilizadas tal y como se plantearon de manera descriptiva.

Las hipótesis deben ser expresadas en términos estadísticos para poder dirigir el tratamiento estadístico. Una *hipótesis estadística* no puede simplemente decir que una variable determina a un fenómeno, sino que debe especificar con toda precisión el grado en que la hipótesis supone la determinación.

En cuanto a la hipótesis que sostiene que el número de Visitas al portal gubernamental (**Y**) es explicado por los factores, Soporte informativo (**X1**), Soporte a trámites (**X2**), Aplicación de mejores prácticas (**X3**), Situación socio-económica (**X4**), Densidad poblacional (**X5**), Mediana de la edad (**X6**), y Escolaridad media (**X7**) de la población, de manera conjunta, se ajustó el siguiente *modelo de regresión lineal múltiple*.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + \varepsilon_i$$

Con $i = 1, 2, \dots, 27$.

El contraste de la hipótesis se llevará a cabo considerando los coeficientes del modelo.

$$\mathbf{H_{01}}: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

H₀₁: No todas las β son 0

Si el valor de los coeficientes no es igual a cero, se rechaza la hipótesis nula en favor de la hipótesis alternativa, que señala que al menos uno de los factores determina de manera significativa a la variable dependiente.

En cuanto a la hipótesis que sostiene que el número de Visitas al portal gubernamental (**Y**) es explicado en mayor medida por los factores Soporte informativo (**X1**) y Mediana de la edad (**X6**), de manera conjunta, se ajustó el siguiente modelo de regresión lineal múltiple.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_6 X_{6i} + \varepsilon_i \quad \text{Con } i = 1, 2, \dots, 27.$$

El contraste de la hipótesis se llevará a cabo considerando los coeficientes del modelo.

$$\mathbf{H_{02}}: \beta_1 = \beta_6 = 0$$

H₀₂: No todas las β son 0

Si el valor de los coeficientes no es igual a cero, se rechaza la hipótesis nula en favor de la hipótesis alternativa, que señala que al menos uno de los factores determina de manera significativa a la variable dependiente.

Las hipótesis de subíndice 1 y 2, constituyen la parte de la investigación que hacen del estudio, un *estudio de tipo correlacional-causal*.

En cuanto a la hipótesis que sostiene que los factores Soporte informativo (**X1**) y Mediana de la edad (**X6**), de manera conjunta, explican más del 35% del número de Visitas al portal gubernamental (**Y**), se aprovecha el modelo de regresión lineal aplicado a la anterior hipótesis, para el cálculo del coeficiente de determinación.

Para el contraste de la hipótesis, se realizarán pruebas de regresión, de tal manera que, para un modelo compuesto por X1 y X6:

$$\mathbf{H_{03}}: R^2 \text{ ajustado} \leq 0.35$$

$$\mathbf{H_{a3}}: R^2 \text{ ajustado} > 0.35$$

Si el coeficiente de determinación ajustado (**R²** ajustado) no es menor o igual a 0.35, entonces se rechaza la hipótesis nula, en favor de la hipótesis alternativa que sostiene que el modelo seleccionado explica más del 35% de la variable dependiente.

La hipótesis de subíndice 3 permite considerar al estudio como del tipo *estudio cuantitativo-correlacional*.

Aplicación de pruebas estadísticas

A continuación se enumeran, en orden de aplicación, los análisis y pruebas estadísticas elegidas para contrastar las hipótesis de la investigación.

Una de las aportaciones de la presente investigación es la metodología descrita en este punto, misma que puede ser desarrollada considerando incluso factores distintos a los considerados en la presente investigación.

Es importante aclarar que sólo se trata describir qué pruebas se realizarán y en qué orden; corresponde al capítulo siguiente ofrecer los resultados de la aplicación de las pruebas.

1. Inicialmente, se realiza un análisis del marco teórico con el fin de identificar aquellos factores que, de acuerdo a otros estudios y de acuerdo a la experiencia propia, son considerados importantes para explicar el número de visitas a los portales gubernamentales.
 - a. El modelo completo pretende incluir factores representativos de las siguientes categorías: a) Información, b) Funcionalidad, c) Diseño y d) Medio ambiente.
 - b. Sólo se deben considerar factores que en apreciación del investigador, están acordes al grado de madurez de los portales a observar. Tómese como referencia el estudio de la ONU (2002) que especifica los niveles de madurez y sus características.
 - c. Se diseña un modelo al que denominamos completo, compuesto por los factores seleccionados para su consideración en la investigación.
2. Para aquellos factores que deban ser observados en campo, se desarrolla un instrumento de toma de datos en campo.
 - a. Los items del instrumento son de valor nominal, dicotómico y binario, donde 0 es la ausencia de una característica observada, y 1 su presencia.

- b. Los aspectos involucrados en el instrumento deben ser observables por el investigador, directamente de los portales.
3. Se realiza un levantamiento, preliminar o definitivo, de datos. Mediante el procedimiento de observación de campo ya descrito (Galtung, 1966)
4. Obtenidos los datos de campo, y en caso de que para ello se haya recurrido a un instrumento diseñado en la investigación, es necesario verificar la consistencia interna de dicho instrumento, para lo cual se ejecuta la prueba Alfa de Cronbach.
 - a. El *Alfa de Cronbach* es un índice de consistencia interna que sirve para comprobar si el instrumento hace mediciones estables y consistentes.
 - b. El estadístico que arroja es el coeficiente Alfa de Cronbach, que toma valores entre 0 y 1; si se obtiene un valor superior a 0.80 es suficiente para rechazar la **H₀** de que el instrumento no es consistente.
 - c. En caso de no rechazarse la hipótesis nula, es necesario diseñar nuevamente el instrumento. Ir al punto 2.
5. Se realiza la prueba *Kolmogorov-Smirnov* para determinar si los datos se ajustan a una distribución normal.
 - a. La prueba *K-S* es una prueba de bondad de ajuste, que compara la distribución de los datos observados, contra una distribución teórica.
 - b. El estadístico que arroja es la **Z** calculada, que entre mayor sea, es más significativa. Al aplicar esta prueba, debe revisarse el valor de significancia (**p**); podemos decir que si **p** adquiere un valor inferior a 0.05, se rechaza la **H₀** de que la distribución observada se ajusta a la distribución teórica (distribución normal).
6. Se desarrollan matrices de correlación utilizando diferentes técnicas. Se utilizan *pruebas no paramétricas (Spearman)*, si es que K-S detectó que uno o todos los factores no tienen una distribución normal. Se utilizan *pruebas paramétricas (Pearson)*, si es que K-S detectó que uno o todos los factores no tienen distribución normal. Se utilizan las pruebas de correlación para determinar si las variables seleccionadas están correlacionadas, y en qué grado.
 - a. Las pruebas de correlación permiten medir la magnitud de la relación entre dos variables.

- b. Lo recomendable es utilizar pruebas paramétricas (Pearson), en lugar de pruebas no paramétricas (Spearman), dado que tienen más potencia; desafortunadamente, dichas pruebas tienen requisito de normalidad.
 - i. Tal como sostiene Pagano (2006), las pruebas paramétricas son, en todos los casos, igual o más potentes que las pruebas no paramétricas; esto se debe a que trabajan directamente con los valores de la muestra, y no con rangos.
 - ii. En afán de no sacrificar los resultados de un método más robusto para aquellas variables que no tenían problemas con las restricciones de normalidad, se aplicaron las dos pruebas; al momento del análisis de la información, sin embargo, para las variables que no presentan normalidad, sólo se consideran los resultados de las pruebas no paramétricas.
 - c. En caso de que alguno de los factores seleccionados no presente una distribución normal, se tomarán en cuenta para dicha variable los resultados de las pruebas no paramétricas. Para las variables que sí presenten distribución normal, utilizaremos las cifras arrojadas por las pruebas paramétricas.
 - d. El estadístico utilizado por estas técnicas el coeficiente de correlación (r), que de ser positivo representa una relación directamente proporcional, mientras que si es negativo, una inversamente proporcional.
 - e. Al aplicar esta prueba, debe revisarse el valor de significancia (p); podemos decir que si p adquiere un valor inferior a 0.05, se rechaza la H_0 de que las variables no están relacionadas entre sí.
7. En caso de que algunas variables muestren una alta correlación entre ellas (generalmente r^2 superior a 0.90, con alta significancia); en ese caso, es necesario descartar la colinealidad entre variables. Para ello se ejecuta la prueba de factores de inflación de varianza (**FIV**).
- a. La prueba indica si existe multicolinealidad, es decir, que dos o más variables presenten una alta correlación lineal.
 - b. Es importante descartar la multicolinealidad, pues uno de los requisitos de las regresiones lineales múltiples es su ausencia.

- c. El estadístico de la prueba **FIV** es el *Factor de Inflación de Varianza*; un valor de **FIV** mayor a 10 nos permite rechazar **H₀** de que no existe multicolinealidad.
- d. En caso de que no se rechace la hipótesis nula, de las variables que presentan colinealidad se debe dejar sólo uno.

Los análisis y pruebas hasta este momento, se representan en la figura 4.6.

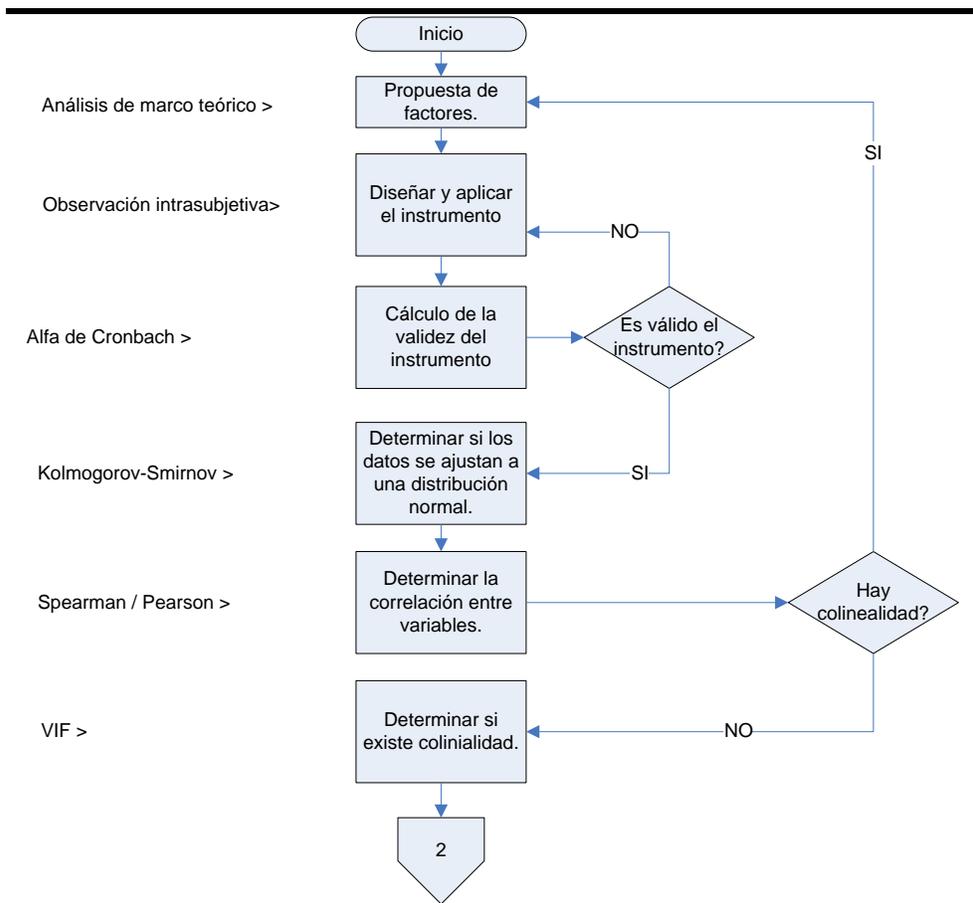


Figura 4.6. Secuencia lógica de análisis y pruebas estadísticas. Primera parte. Desarrollo propio.

8. Con las variables que se hayan seleccionado, se propone un *modelo de regresión lineal múltiple*, dando lugar a lo que llamamos modelo completo.
 - a. Un modelo de regresión lineal múltiple nos permite describir la relación entre una variable dependiente y diferentes variables independientes.
 - b. El estadístico de esta prueba es la *F calculada*. El valor crítico de esta prueba se determina en base a tablas de valores críticos de la *distribución de razón F*, atendiendo el nivel de significancia preferido (0.05 ó 0.01, generalmente), y los grados de libertad del numerador (**k-1**) y denominador (**n-k**) para la prueba.
 - c. Es necesario revisar el valor de **p**, pues en caso de ser menor a 0.05, nos permite rechazar **H₀** que nos dice que dos conjuntos de datos son distintos.
9. Para validar la significancia de los resultados obtenidos con el modelo de regresión lineal múltiple, se aplica un *análisis de varianza (ANOVA)*.
 - a. El análisis de varianza (ANOVA) sirve para comparar si los valores de un conjunto de datos numéricos son significativamente distintos a los valores de otro o más conjuntos de datos. El procedimiento para comparar estos valores está basado en la varianza global observada en los grupos de datos numéricos a comparar.
 - b. El estadístico de esta prueba es la **F** calculada. El valor crítico de esta prueba se determina en base a tablas de valores críticos de la distribución de razón **F**, atendiendo el nivel de significancia preferido (.05 ó .01, generalmente), y los grados de libertad del numerador (**k-1**) y denominador (**n-k**) para la prueba.
 - c. Es necesario revisar el valor de **p**, pues en caso de ser menor a 0.05, nos permite rechazar **H₀** que nos dice que dos conjuntos de datos son distintos.
10. Se parte del supuesto que, si el modelo completo explica a la variable dependiente, probablemente un sub-conjunto de dicho modelo puede también determinar a la variable dependiente, incluso de manera más óptima que el modelo completo.
 - a. Se estiman los *coeficientes de regresión (r)* para determinar la significancia de cada una de las variables del modelo, y saber si todas las variables contribuyen a la explicación de la variable dependiente o no.

- b. Los coeficientes de regresión indican el número de unidades en que se modifica la variable dependiente por efecto del cambio de una variable o viceversa, en una unidad de medida.
 - c. El coeficiente de regresión puede ser positivo, negativo o nulo. Es positivo cuando las variaciones de la variable independiente son directamente proporcionales a las variaciones de la variable dependiente. Es negativo, cuando las variaciones de la variable independiente son inversamente proporcionales a las variaciones de las variables dependientes. Es nulo si entre la variable dependiente e independiente no hay relación alguna.
 - d. Entre mayor sea el valor del estadístico, mayor el grado en que se explica la variable dependiente.
11. Si se encuentra que no todas las variables contribuyen a la significancia del modelo, se hace una selección de variables utilizando la técnica **Backward** (Hacia atrás).
- a. Se toman como base las variables que componen el modelo completo, y se proponen nuevos modelos formados por sub-conjuntos de variables, a los que denominaremos modelos **ajustados** o reducidos.
12. Se realiza un análisis de varianza sobre los modelos ajustados, para ver su significancia estadística.
- a. Si alguno de los modelos ajustados explica a la variable dependiente, y muestra un mejor comportamiento que el resto de los modelos ajustados, a dicho modelo le denominaremos óptimo.
13. De manera informativa, una vez seleccionado el modelo óptimo, se procede a realizar una estimación que nos diga cuál es el impacto que tienen en la variable dependiente, las variaciones en los valores de las variables que componen el modelo óptimo.

La figura 4.7 muestra de manera gráfica la secuencia de pruebas enumerada con anterioridad. Esta figura es la continuación del flujo iniciado en la figura 4.6.

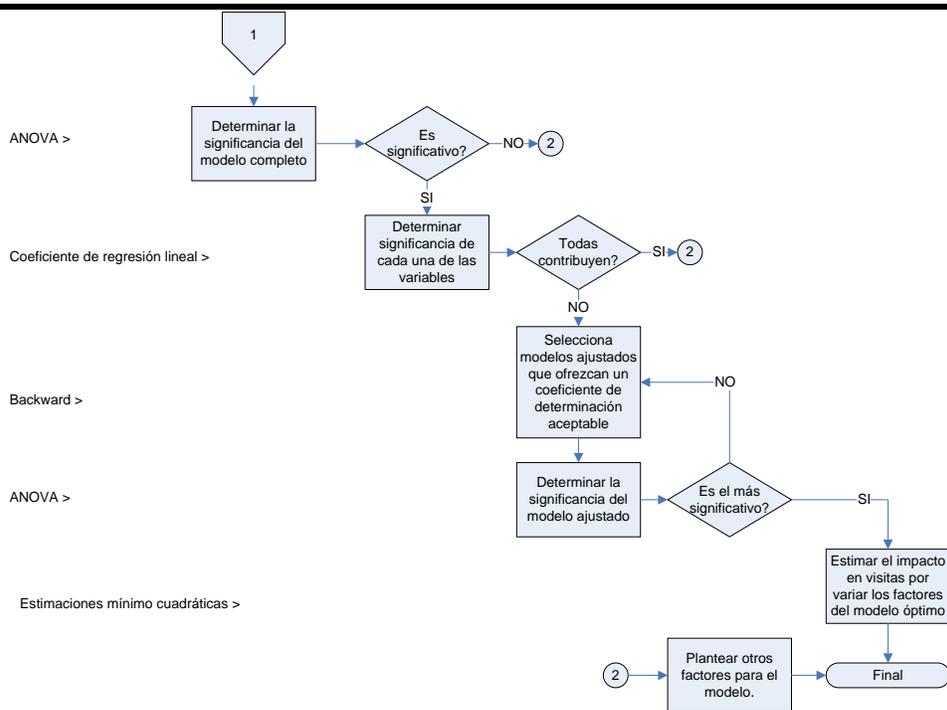


Figura 4.7. Secuencia lógica de análisis y pruebas estadísticas. Segunda parte. Desarrollo propio.

TRATAMIENTO ESTADÍSTICO

Se analizaron múltiples estudios antecedentes con el fin de encontrar qué aspectos han sido considerados por otros investigadores para medir la eficiencia de los portales gubernamentales, su grado de usabilidad y su contribución al servicio que brindan los gobiernos a los ciudadanos. Hecho un inventario de factores, se procedió a seleccionar o proponer un mínimo de factores dentro de los cuales pudiéramos encontrar aquellos que explican el número de visitas.

Como los más relevantes, encontramos los siguientes 10 estudios: Lucas (1999); Naciones Unidas (2002); Hammer (2001); West (2005); Berčič, Dečman, Kunstelj y Vintar, (2003); Fountain (2004); OCDE (2003); Gutiérrez (2003); Gil-García (2006); Kost y McClure (2006); también se hicieron propuestas de esta misma investigación.

A partir de todos los estudios, se seleccionaron 92 aspectos importantes a observar; se privilegiaron aquellos aspectos que no fueran sujetos a interpretación, con el fin de que el estudio se fundamentara solamente en datos duros, cuantitativos y observables en análisis de campo por el investigador. Se cuidó además de incluir sólo aquellos factores que están acordes al nivel de madurez observado de forma preliminar en los portales gubernamentales sujetos a investigación, que es entre Enriquecido e Interactivo (ONU, 2002).

Los aspectos seleccionados permiten la integración cuantitativa de 7 factores, cada uno de ellos constituyéndose como una variable independiente en la investigación.

NIVEL DE SIGNIFICANCIA Y ZONA DE RECHAZO

El nivel de significancia o riesgo considerado para todo el estudio es del 5%.

Se deberán considerar como valores críticos, los otorgados por la tabla de *valores críticos de la distribución F* (Lind, Mason, Marchal, 2001). Se deberá tomar en consideración el número de elementos de la muestra (n) y el número de tratamientos o grados de libertad (k).

Teóricamente, esa información nos ofrecería una zona de rechazo como muestra la figura 4.8.

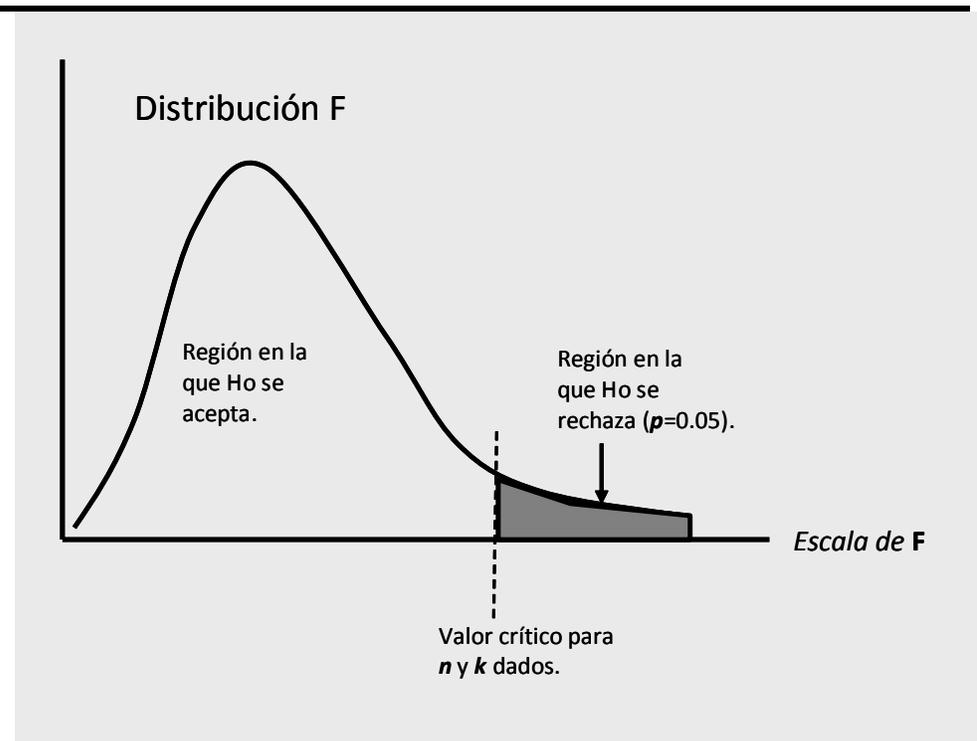


Figura 4.8. Nivel de significancia y zona de rechazo de la investigación. Desarrollo propio.

ESTADÍSTICA DESCRIPTIVA

Con el propósito de conocer la naturaleza de los datos almacenados en las variables, se utilizó estadística descriptiva, obteniéndose tanto *medidas de centralidad* (media, mediana), como *medidas de variabilidad* (varianza, desviación estándar, rango).

Dadas las cifras obtenidas, no se encuentran elementos para suponer que no podemos aplicar con éxito las pruebas estadísticas que hemos seleccionado para nuestro análisis (véase la tabla 4.5).

Variable	Valor Mínimo	Valor Máximo	Media	Mediana	Desviación estándar
Y = Visitas al portal	88420	3800000	1333332	1039231	1029224.4
X ₁ = Soporte informativo	0.18	100.00	51.25	50.00	35.4333
X ₂ = Soporte a trámites	0.00	12.00	2.56	0.00	3.4789
X ₃ = Mejores Prácticas	28.17	69.88	49.79	48.29	11.9319
X ₄ = Socio - Económicos	6.48	36.94	14.80	12.54	7.0681
X ₅ = Densidad poblacional	6.00	5799.00	290.00	53.00	1103.4845
X ₆ = Mediana de la edad	20.00	29.00	23.93	24.00	1.6155
X ₇ = Escolaridad Media	5.30	9.40	7.18	7.20	0.8842

n = 27

Tabla 4.5.

Estadísticas descriptivas. Desarrollo propio.

El origen de la información presentada en esta tabla fue el resultado de procesar los datos obtenidos en la investigación en el software SPSS versión 13.

PRUEBA DE VALIDEZ DEL INSTRUMENTO

Para corroborar la consistencia interna del instrumento a través del cual se realizó el levantamiento de datos en campo, se utilizó el *Alfa de Cronbach*.

Son dos las variables involucradas en el instrumento, de las cuales dos de ellas adquieren su valor a través de constructos: (**X1**) Soporte Informativo y (**X3**) Aplicación de mejores prácticas.

Se calculó el Alfa de Cronbach para los 87 items que componen el instrumento, obteniendo un valor de 0.858. Según Ritchey (2002), un valor de Alfa de Cronbach superior a 0.800 implica que el instrumento es consistente.

A continuación se muestra el resultado obtenido de procesar los datos obtenidos en la investigación en el software SPSS versión 13.

A continuación se muestra la salida obtenida del programa, tal y como la presenta.

Reliability Statistics

Cronbach's Alpha	N of Items
.858	87

PRUEBAS DE NORMALIDAD DE LOS DATOS (KOLMOGOROV - SMIRNOV)

Para revisar la normalidad de los datos se aplicó la prueba de *Kolmogorov - Smirnov*. Se encontró que solamente las variables Soporte a trámites (X2) y Densidad poblacional (X5) no tienen distribución normal ($p = .001$ y $p = .023$, respectivamente).

En el caso de esas dos variables, se rechaza la hipótesis nula que indica que la distribución observada se ajusta a la distribución teórica (normal).

Véase la tabla, tabla 4.6.

Variable	K-S Z_{cal}	Valor de p
Y = Visitas al portal	0.86	.456
X ₁ = Soporte informativo	0.91	.373
X ₂ = Soporte a trámites	1.49	.023
X ₃ = Mejores Prácticas	0.53	.942
X ₄ = Socio - Económicos	0.75	.633
X ₅ = Densidad poblacional	2.35	.001
X ₆ = Mediana de la edad	1.09	.188
X ₇ = Escolaridad Media	0.50	.965

n = 27

Tabla 4.6.
Prueba Kolmogorov - Smirnov para normalidad.
Desarrollo propio.
El origen de la información presentada en esta tabla fue el resultado de procesar los datos obtenidos en la investigación en el software SPSS versión 13.

COEFICIENTES DE CORRELACIÓN ENTRE VARIABLES

Se procedió a aplicar pruebas paramétricas (*Pearson*) y no paramétricas (*Spearman*) sobre los datos recopilados.

Respecto al análisis no paramétrico (*Spearman*), solamente las variables Soporte informativo (**X1**) y Soporte a trámites (**X2**) mostraron correlación directa y significativa con Visitas al portal (**Y**) ($p = .017$, $p = .046$, respectivamente).

Respecto al análisis paramétrico (*Pearson*), solamente las variables Soporte informativo (**X1**), la Densidad poblacional (**X5**) y la Edad (**X6**) se correlacionan de manera positiva y significativa con la variable dependiente, Visitas al portal (**Y**) ($p = .013$, $p = .014$, $p = .018$, respectivamente).

Respecto a las variables Densidad poblacional (**X5**) y Soporte a trámites (**X2**), que no presentaron distribución normal, sólo deben considerarse los resultados de la prueba de *Spearman*, ya que la prueba de *Pearson* es menos confiable en ausencia de normalidad.

Véase el apéndice G, tabla G3.

	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
Y	1	.457 (.017)	.388 (.046)	-.230 (.249)	.188 (.348)	.280 (.157)	.305 (.122)	.136 (.500)
X ₁	.470 (.013)	1	.010 (.961)	.234 (.241)	.387 (.046)	.053 (.792)	.052 (.796)	.169 (.400)
X ₂	.181 (.366)	-.057 (.779)	1	-.082 (.685)	.329 (.094)	.196 (.326)	.273 (.169)	.392 (.043)
X ₃	-.154 (.443)	.212 (.288)	-.145 (.471)	1	.195 (.329)	.025 (.901)	.111 (.582)	.022 (.913)
X ₄	.372 (.056)	.373 (.055)	.134 (.505)	.124 (.537)	1	-.159 (.429)	.552 (.003)	.716 (.001)
X ₅	.466 (.014)	.143 (.476)	.034 (.866)	-.035 (.864)	.613 (.001)	1	-.035 (.862)	.028 (.889)
X ₆	.452 (.018)	.044 (.829)	.131 (.516)	.023 (.908)	.710 (.001)	.624 (.001)	1	.611 (.001)
X ₇	.285 (.149)	.183 (.361)	.304 (.123)	-.067 (.738)	.788 (.001)	.506 (.007)	.772 (.001)	1

Tabla 4.7

Coefficientes de correlación, paramétricos y no paramétricos, y su significancia. Desarrollo propio.

El valor de p, bilateral, se presenta entre paréntesis, n = 27

Sección triangular superior corresponde a pruebas no paramétricas (Spearman), y en la sección triangular inferior corresponde a pruebas paramétricas (Pearson).

Podemos concluir varias cosas interpretando la correlación entre variables: la variable Soporte informativo (X1) es la que se mostró más consistente, pues presenta una correlación positiva y significativa en ambas pruebas. Sólo presentan correlación positiva y significativa, aunque en sólo una de las pruebas, las variables Soporte a trámites (X2) y Mediana de la edad (X6).

Véase, tabla 4.5.

	Spearman (No paramétrico)				Pearson (Paramétrico)			
	Tipo de correlación	Coefficiente	p	Significativo	Tipo de correlación	Coefficiente	p	Significativo
X ₁ = Soporte informativo	Positivo	0.45	0.0	SI	Positivo	0.47	0.0	SI
		7	17			0	13	
X ₂ = Soporte a trámites	Positivo	0.38	0.0	SI	Positivo	0.18	0.3	SI
		8	46			1	66	
X ₃ = Mejores Prácticas	Negativo	0.23	0.2	NO	Negativo	0.15	0.4	NO
		0	49			4	43	
X ₄ = Socio - Económicos	Positivo	0.18	0.3	NO	Positivo	0.37	0.0	SI
		8	48			2	56	
X ₅ = Densidad poblacional	Positivo	0.28	0.1	NO	Positivo	0.46	0.0	SI
		0	57			6	14	
X ₆ = Mediana de la edad	Positivo	0.30	0.1	NO	Positivo	0.45	0.0	SI
		5	22			2	18	
X ₇ = Escolaridad	Positivo	0.13	0.5	NO	Positivo	0.28	0.1	SI
Media		6	00			5	49	

Tabla 4.8 Interpretación de los coeficientes de correlación, paramétricos y no paramétricos, entre las variables independientes y la variable dependiente. Desarrollo propio.

El origen de la información presentada en estas tablas fue el resultado de procesar los datos obtenidos en la investigación en el software SPSS versión 13.

MODELO DE REGRESIÓN LINEAL MÚLTIPLE PARA 7 VARIABLES (MODELO COMPLETO)

Con el propósito de contrastar la hipótesis de que el Y es función de al menos una de las siguientes variables: Soporte informativo (X_1), Soporte a trámites (X_2), Mejores prácticas (X_3), los Elementos socio-económicos (X_4), la Densidad poblacional (X_5), la Mediana de la edad (X_6) y la Escolaridad media (X_7), se ajustó el siguiente modelo de regresión lineal múltiple.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + \varepsilon_i \quad \text{Con } i = 1, 2, \dots, n.$$

Se realizaron análisis de regresión con el fin de contrastar las hipótesis individuales $H_0: \beta_j = 0$ contra la hipótesis alternativa bilateral, $H_1: \beta_j \neq 0$, con $j = 1, \dots, 7$.

Se tiene que sólo las variables X_1 (Soporte informativo) y X_6 (Mediana de la edad) proporcionan valores de p significativos, 0.001 y 0.011 respectivamente. En ambos casos, la T calculada es positiva.

Véase tabla 4.6A.

Fuente de Variación	Suma de Cuadrados	G de L	Cuadrado Medio	F_{ca} ₁	Valor de p
Regresión	1.7226E+13	7	2.4608E+12	4.53	.004
Residual	1.0316E+13	19	5.4296E+11		
Total c	2.7542E+13	26			

$$R^2 = .625, R^2 \text{ ajustado} = .487$$

Tabla 4.9
Análisis de Varianza para el modelo completo. Desarrollo propio.

El origen de la información presentada en esta tabla fue el resultado de procesar los datos obtenidos en la investigación en el software SPSS versión 13.

Para conformar los resultados obtenidos por el modelo de regresión, se realizó un análisis de varianza (ANOVA), y se tiene que el modelo completo es significativo (**Fcal** = 4.53, **p** = .004).

El valor crítico de **F**, obtenido de la tabla de valores críticos, es de 2.60, considerando un nivel de confianza (α) de 5%, siendo los grados de libertad del numerador $(k-1) = 6$, y los grados de libertad del denominador $(n-k) = 20$.

Dado que **Fcal** = 4.53, es mayor al valor crítico que corresponde al modelo, se rechaza la hipótesis nula que establece que los coeficientes del modelo, sin incluir la constante (**β_0**), son iguales a cero.

APLICACIÓN DE LA TÉCNICA BACKWARD PARA GENERACIÓN DE MODELOS AJUSTADOS

La técnica de selección de variables “Backward” (o hacia atrás) se empleó para buscar un modelo de regresión lineal que mejor describa el comportamiento de Visitas al portal (**Y**).

Este procedimiento descarta aquellas combinaciones de variables que no ofrecen un coeficiente de determinación adecuado. A partir de los modelos ajustados que propone Backward, podemos revisar cuál de todos los modelos ofrece una mayor significancia.

El orden en que fueron saliendo las variables fue el siguiente:

1. Situación socio-económica (**X4**)
2. Densidad poblacional (**X5**)
3. Soporte a trámites (**X2**)
4. Escolaridad media (**X7**),
5. Mejores prácticas (**X3**)

Subsisten en el modelo óptimo dos variables: Soporte informativo (**X1**), y Mediana de la edad (**X6**).

Véase la, tabla 4.6B, para ver los modelos ajustados propuestos por el procedimiento, así como los coeficientes de determinación de cada uno.

Modelo	R ²	R ² ajus .	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
1	.6254	.4874	√	√	√	√	√	√	√
2	.6223	.5089	√	√	√	X	√	√	√
3	.5970	.5010	√	√	√	X	X	√	√
4	.5502	.4684	√	X	√	X	X	√	√
5	.4783	.4102	√	X	√	X	X	√	X
6	.4075	.3582	√	X	X	X	X	√	X

√ = variable incluida en el modelo de RLM, X =
variable excluida en modelo de RLM

Tabla 4.10.
Coeficientes de Determinación, para los modelos
ajustados en el procedimiento "Backward" o hacia
atrás. Desarrollo propio.

El origen de la información presentada en esta tabla fue el resultado de procesar los datos obtenidos en la investigación en el software SPSS versión 13.

Dado que *Stepwise* (por pasos) es la técnica más generalmente utilizada, no quisimos omitir su ejecución, con la finalidad de contrastarla contra los resultados que obtuvimos con Backward.

A continuación se muestra la salida obtenida del programa SPSS, para el análisis Stepwise, tal y como la presenta.

Regression

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	X1_Soporte Informativo	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .060).
2	x6_edad	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .060).

a. Dependent Variable: Y_UsoPortal

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.470 ^a	.221	.190	926285.934
2	.638 ^b	.408	.358	824561.512

a. Predictors: (Constant), X1_SoporteInformativo

b. Predictors: (Constant), X1_SoporteInformativo, x6_edad

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6E+012	1	6.092E+012	7.100	.013 ^a
	Residual	2E+013	25	8.580E+011		
	Total	3E+013	26			
2	Regression	1E+013	2	5.612E+012	8.254	.002 ^b
	Residual	2E+013	24	6.799E+011		
	Total	3E+013	26			

a. Predictors: (Constant), X1_SoportelInformativo

b. Predictors: (Constant), X1_SoportelInformativo, x6_edad

c. Dependent Variable: Y_UsPortal

De forma complementaria, se procesó también la técnica *Forward*.

A continuación se muestra la salida obtenida del programa SPSS, para el análisis Forward, tal y como la presenta.

Regression**Variables Entered/Removed^a**

Model	Variables Entered	Variables Removed	Method
1	X1_SoportelInformativo	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
2	x6_edad	.	Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: Y_UsPortal

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.470 ^a	.221	.190	926285.934
2	.638 ^b	.408	.358	824561.512

a. Predictors: (Constant), X1_Soportelnf ormativ o

b. Predictors: (Constant), X1_Soportelnf ormativ o, x6_edad

ANOVA^c

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6E+012	1	6.092E+012	7.100	.013 ^a
	Residual	2E+013	25	8.580E+011		
	Total	3E+013	26			
2	Regression	1E+013	2	5.612E+012	8.254	.002 ^b
	Residual	2E+013	24	6.799E+011		
	Total	3E+013	26			

a. Predictors: (Constant), X1_Soportelnf ormativ o

b. Predictors: (Constant), X1_Soportelnf ormativ o, x6_edad

c. Dependent Variable: Y_UsoPortal

Se realizó el mismo procedimiento utilizando los métodos Stepwise y Forward, mismos que ofrecieron los mismos resultados.

Respecto a por qué se utilizó Backward, siendo que Stepwise es más común, la respuesta es que para los fines de la investigación, resultaba importante ver el orden en que el procedimiento descartaba los factores, a fin de sugerir el orden en que los esfuerzos de gobierno electrónico deben ser aplicados.

Otro argumento para preferir Backward, es que según Kunter, Nachtsheim, Neter (2004), para muestras pequeñas y pocos grados de libertad, Backward es más eficiente que Stepwise.

ANÁLISIS DE VARIANZA PARA DETERMINAR LA SIGNIFICANCIA DE LOS MODELOS AJUSTADOS

Se aplicó el análisis de varianza a los modelos ajustados propuestos, mismos que se generaron con la técnica Backward, esto con el fin de determinar su significancia.

Puede constatarse que el modelo óptimo es el que mejor valor de F mostró ($F_{cal} = 8.25$, $p = .002$). Ver la tabla 4.7.

Modelo	Fuente de Variación	Suma de Cuadrados	G de L	Cuadrado Medio	F_{cal}	Valor de p
1	Regresión	1.7226E+13	7	2.4608E+12	4.53	.004
	Residual	1.0316E+13	19	5.4296E+11		
	Total c	2.7542E+13	26			
2	Regresión	1.7138E+13	6	2.8564E+12	5.49	.002
	Residual	1.0404E+13	20	5.2018E+11		
	Total c	2.7542E+13	26			
3	Regresión	1.6442E+13	5	3.2884E+12	6.22	.001
	Residual	1.11E+13	21	5.2857E+11		
	Total c	2.7542E+13	26			
4	Regresión	1.5152E+13	4	3.7881E+12	6.73	.001
	Residual	1.239E+13	22	5.6316E+11		
	Total c	2.7542E+13	26			
5	Regresión	1.3173E+13	3	4.391E+12	7.03	.002
	Residual	1.4369E+13	23	6.2474E+11		
	Total c	2.7542E+13	26			
6	Regresión	1.1224E+13	2	5.6121E+12	8.25	.002
	Residual	1.6318E+13	24	6.799E+11		
	Total c	2.7542E+13	26			

Tabla 4.11
Tablas de ANOVA para los modelos ajustados en el procedimiento "Backward" o hacia atrás. Desarrollo propio.

SELECCIÓN DEL MODELO ÓPTIMO

El concepto tomado como elemento para la selección del modelo óptimo es la significancia mostrada por el análisis de varianza al que fue sometido cada uno de ellos.

El modelo de regresión lineal seleccionado por esta técnica y que resulta estadísticamente significativo ($F_{cal} = 8.25$, $p = .002$) es el que considera a Visitas al portal (Y) en función del Soporte informativo ($X1$) y la Mediana de la edad ($X6$) de los habitantes del estado.

Aquí se tiene que el 40.75% de la variación existente en el Visitas al portal (Y) se explica por el modelo de regresión lineal múltiple seleccionado (R^2).

Con el propósito de analizar la estabilidad del modelo se emplearon dos técnicas adicionales de selección de variables, Forward o Hacia delante y Stepwise o Por pasos. Ambos métodos indican que el mejor modelo es el seleccionado por el procedimiento Backward o Hacia atrás.

Véase la, tabla 4.12, que muestra un resumen de los modelos ajustados propuestos por el procedimiento Backward, así como la significancia que muestra cada uno, de acuerdo a la prueba ANOVA a la que fue sometido cada uno de los modelos.

Modelo	Variable	Coeficientes de determinación		ANOVA	
		R ²	R ² Ajustado	F _{cal}	Valor de p
1	X ₁ Soporte informativo X ₂ Soporte a trámites X ₃ Mejores prácticas X ₄ Socio – Económicos X ₅ Densidad poblacional X ₆ Mediana de la edad X ₇ Escolaridad media	0.6254	0.4874	4.53	0.004
2	X ₁ Soporte informativo X ₂ Soporte a trámites X ₃ Mejores prácticas X ₅ Densidad poblacional X ₆ Mediana de la edad X ₇ Escolaridad media	0.6223	0.5089	5.49	0.002
3	X ₁ Soporte informativo X ₂ Soporte a trámites X ₃ Mejores prácticas X ₆ Edad promedio X ₇ Escolaridad media	0.5970	0.5010	6.22	0.001
4	X ₁ Soporte informativo X ₃ Mejores prácticas X ₆ Mediana de la edad X ₇ Escolaridad media	0.5502	0.4684	6.73	0.001
5	X ₁ Soporte informativo X ₃ Mejores prácticas X ₆ Mediana de la edad	0.4783	0.4102	7.03	0.002
6	X ₁ Soporte informativo X ₆ Mediana de la edad	0.4075	0.3592	8.25	0.002

Tabla 4.12.
Tabla de comparación de modelos ajustados propuestos por el procedimiento Backward considerando su significancia. Desarrollo propio.
El origen de la información presentada en estas tablas fue el resultado de procesar los datos obtenidos en la investigación en el software SPSS versión 13.

ESTIMACIONES MÍNIMO CUADRÁTICAS PARA LOS MODELOS AJUSTADOS O REDUCIDOS

Además, al incrementarse una unidad el Soporte informativo (**X1**), se espera que el Visitas al portal (**Y**) se incremente en 13,112 unidades, cuando la Mediana de la edad (**X6**) permanece constante. Al incrementarse la Mediana de la edad (**X6**) en un año, se espera que las Visitas al portal (**Y**) se incrementen en 275,295 unidades, permaneciendo constante el Soporte informativo (**X1**).

Véase la tabla 4.9, que muestra las estimaciones mínimo cuadráticas y su significancia, para cada uno de los modelos ajustados, propuestos por Backward.

Modelo	Variable	Coficiente	Error Estándar	Estadístico t_{cal}	Valor de p
1	(Constante)	-5234831.8	3244934.4	-1.61	.123
	X ₁ Soporte informativo	17825.3	4750.8	3.75	.001
	X ₂ Soporte a trámites	71805.1	45085.6	1.59	.128
	X ₃ Mejores prácticas	-24002.4	13057.3	-1.84	.082
	X ₄ Socio – Económicos	-16586.7	41358.6	-0.40	.693
	X ₅ Densidad poblacional	217.2	180.8	1.20	.244
	X ₆ Mediana de la edad	464323.9	165070.5	2.81	.011
	X ₇ Escolaridad media	-593608.0	337003.3	-1.76	.094
2	(Constante)	-4617146.0	2795556.7	-1.65	.114
	X ₁ Soporte informativo	17134.2	4333.4	3.95	.001
	X ₂ Soporte a trámites	72762.6	44067.7	1.65	.114
	X ₃ Mejores prácticas	-25150.6	12469.4	-2.02	.057
	X ₅ Densidad poblacional	193.4	167.1	1.16	.261
	X ₆ Mediana de la edad	453849.1	159535.0	2.84	.010
	X ₇ Escolaridad media	-665423.9	279436.4	-2.38	.027
	3	(Constante)	-6486252.2	2299891.9	-2.82
X ₁ Soporte informativo		17896.6	4317.4	4.15	.000
X ₂ Soporte a trámites		69219.7	44314.3	1.56	.133
X ₃ Mejores prácticas		-26626.5	12503.6	-2.13	.045
X ₆ Edad promedio		534406.8	144693.6	3.69	.001
X ₇ Escolaridad media		-659679.6	281636.6	-2.34	.029
4		(Constante)	-6256731.8	2369113.0	-2.64
	X ₁ Soporte informativo	17038.0	4420.2	3.85	.001
	X ₃ Mejores prácticas	-28155.2	12866.7	-2.19	.040
	X ₆ Mediana de la edad	493666.1	146907.1	3.36	.003
	X ₇ Escolaridad media	-514474.8	274414.9	-1.87	.074

5	(Constante)	-4900618.0	2376128.0	-2.06	.051
	X ₁ Soporte informativo	14784.0	4480.1	3.30	.003
	X ₃ Mejores prácticas	-23480.3	13295.0	-1.77	.091
	X ₆ Mediana de la edad	277743.9	96057.0	2.89	.008
6	(Constante)	-5925385.0	2403765.7	-2.47	.021
	X ₁ Soporte informativo	13112.3	4568.1	2.87	.008
	X ₆ Mediana de la edad	275295.2	100197.6	2.75	.011

Tabla 4.13.
Estimaciones mínimo cuadráticas y su significancia, para los modelos ajustados en el procedimiento “Backward”o hacia atrás. Desarrollo propio.

Comprobación de hipótesis

CONTRASTE DE HIPÓTESIS RELACIONADA CON EL MODELO COMPLETO

Relacionado con la hipótesis del modelo completo, tenemos que en virtud de que la *F calculada* es mayor que la *F tabulada* (**Fcal** = 4.53, **Ftab**=2.6), se rechaza la hipótesis nula que establece que los coeficientes del modelo, sin incluir la constante (**β0**), son iguales a cero, esto en favor a la hipótesis alternativa que sostiene que la variable dependiente, Visitas al portal gubernamental (**Y**), es explicada en cierto grado por las variables independientes, Soporte informativo (**X1**), Soporte a trámites (**X2**), Aplicación de mejores prácticas (**X3**), Situación socio-económica (**X4**), Densidad poblacional (**X5**), Mediana de la edad (**X6**) y Escolaridad media (**X7**). Consultar el apéndice G, tabla G6.

El valor crítico de **F**, obtenido de la tabla de valores críticos de la razón **F**, es de 2.60 (**Ftab**), considerando un nivel de confianza (**α**) de 5%, siendo los grados de libertad del numerador (**k-1**) = 6, y los grados de libertad del denominador (**n-k**) = 20.

CONTRASTE DE HIPÓTESIS RELACIONADA CON EL MODELO ÓPTIMO

Relacionado con la hipótesis del modelo completo, tenemos que en virtud de que la *F calculada* es mayor que la *F tabulada* (**Fcal** = 8.25, **Ftab**=4.24), se rechaza la hipótesis nula que establece que los coeficientes del modelo, sin incluir la constante (**β0**), son iguales a cero, esto en favor a la hipótesis alternativa que sostiene que la variable dependiente, Visitas al portal gubernamental (**Y**), es explicada en cierto grado por las variables independientes, Soporte informativo (**X1**) y Mediana de la edad (**X6**). Consultar el apéndice , tabla4.7.

El valor crítico de **F**, obtenido de la tabla de valores críticos de la razón **F**, es de 4.24 (**F_{tab}**), considerando un nivel de confianza (α) de 5%, siendo los grados de libertad del numerador (**k-1**) = 1, y los grados de libertad del denominador (**n-k**) = 25.

CONTRASTE DE HIPÓTESIS RELACIONADA CON LA ESTIMACIÓN

Dado que el coeficiente de determinación ajustado para el modelo óptimo (**R²** ajustado) es igual a 0.3592, se rechaza la hipótesis nula que sostiene que **R²** ajustado es cero, en favor de la hipótesis alternativa que sostiene que el modelo seleccionado explica más del 35% de la variable dependiente.

Conclusiones y recomendaciones

En esta parte se exponen las conclusiones que, con fundamento científico se desprenden de la presente investigación.

Respuestas a las preguntas de investigación

Las preguntas de investigación fueron respondidas de la siguiente manera.

- P1:** Los estudios a nivel mundial, relativos al gobierno electrónico, señalan como factores a considerar para medir la madurez y eficacia de los portales de gobierno electrónico, al soporte informativo, al soporte a trámites y a la aplicación de mejores prácticas mundiales; también señalan como relevantes a las condiciones de medio ambiente, tales como la situación socio-económica, la densidad poblacional, la edad, y la escolaridad; ¿son esos factores estadísticamente significativos para explicar el número de visitas a los portales gubernamentales en México?
- R1:** Sí lo son, aunque no de manera óptima. (Respuesta obtenida al rechazar **Ho1**, en favor de **Ha1**).
- P2:** ¿De los factores a considerar para medir la madurez y eficacia de los portales de gobierno electrónico sugeridos por los estudios realizados a nivel mundial, cuáles de ellos, de manera individual o conjunta, explican en mayor medida el número de visitas a los portales gubernamentales en México?
- R2:** El soporte informativo (**X1**) en conjunto con la mediana de la edad (**X6**). (Respuesta obtenida al seleccionar el modelo óptimo, y rechazar **Ho2**, en favor de **Ha2**)

- P3: Resuelta la pregunta específica de investigación anterior, ¿en qué grado explican el número de visitas a los portales gubernamentales en México los factores identificados como positivamente más influyentes?
- R3: En un 35.92%. (Respuesta obtenida al rechazar **Ho3**, en favor de **Ha3**).

Conclusiones

La comprobación de las hipótesis nos permite determinar que la selección inicial de variables que componen el modelo completo fue acertada, en el sentido que se descubrió que efectivamente existe una correlación entre el modelo conformado por las variables independientes y la variable dependiente.

Se seleccionaron factores que fueron pertinentes, evitándose factores que por el grado de madurez de los portales gubernamentales en México, sólo complicaran el análisis, sin aportar información real de la situación.

Como hallazgo importante de la investigación, a partir del modelo completo, se obtiene un modelo óptimo que constituye la aportación al conocimiento de la investigación, al descubrir que en México el número de visitas a los portales gubernamentales depende de la información que el sitio expone a los ciudadanos (variable **X1**, Soporte informativo), en conjunto con la edad promedio de la población (variable **X6**, Mediana de la edad).

Estas variables de forma conjunta explican el 35.92% de las visitas a los portales gubernamentales.

Al incrementarse una unidad el Soporte informativo (**X1**), se espera que las Visitas al portal (**Y**) se incrementen en 13,112 unidades, cuando la Mediana de la edad (**X6**) permanece constante.

Al incrementarse la Mediana de la edad (**X6**) en un año, se espera que le Visitas al portal (**Y**) se incremente en 275,295 unidades, permaneciendo constante el Soporte informativo (**X1**).

Uno de los aspectos que más se cuidan al momento de desarrollar un portal gubernamental, es que cumpla con las mejores prácticas recomendadas a nivel internacional, en términos de diseño y distribución de páginas Web. Sin embargo, el estudio nos arroja que dicho factor no es significativo.

Aspectos de medio ambiente que en el concepto popular pudieran ser considerados importantes, por ejemplo la escolaridad o la situación socio-económica, no fueron significativos en el presente estudio.

Recomendaciones

De las dos variables que explican el número de las visitas a los portales, la que puede ser atendida de inmediato es el Soporte informativo. Se recomienda que los gobiernos de los estados mejoren la información contenida en su portal; siendo uno de los rubros críticos, se recomienda disponer de un staff que garantice que la información contenida es actualizada, es consistente, y acorde a las necesidades de comunicación de la entidad.

Los progresos en dicha estrategia pueden medirse utilizando los instrumentos propuestos en el presente estudio. Esto les puede mejorar en el corto plazo las visitas a su portal.

Respecto a las variables de medio ambiente, es complicado manipularlas en favor de las iniciativas de gobierno electrónico, aparte de no ser significativas. Lo que sí podemos atender es la mejora de los factores que no son de medio ambiente, la recomendación es que los gobiernos fijen una prioridad de esfuerzos de la siguiente manera:

1. Primero: Mejorar el contenido de información del portal.
2. Segundo: Mejorar la funcionalidad del portal, ofreciendo más trámites.
3. Tercero: Mejorar el diseño con la aplicación de mejores prácticas, dando prioridad a aquellas que permitan un mejor aprovechamiento del soporte informativo, por ejemplo, búsquedas inteligentes de información, que es un aspecto de diseño.

Es muy importante mencionar que la aportación metodológica está presente, al igual que la aportación teórica. Respecto a la aportación práctica, los gobiernos de los estados cuentan ya con un estudio para referir y justificar las inversiones en los portales gubernamentales, especificando el beneficio estimado que se pretende obtener por las iniciativas que impulsan.

Asimismo, pueden utilizar las métricas e instrumentos para medir el soporte informativo y la aplicación de mejores prácticas, y diseñar programas de mejora continua para sus portales.

Bibliografía

1. Agencia Interamericana para la Cooperación y el Desarrollo. (2003, 14-15 de Enero). *E-Government Applications in Latin America: An overview*. Recuperado el 24 de octubre de 2004, de <http://unpan1.un.org/intradoc/groups/public/documents/caricad/unpan007530.pdf>
2. Aibar, E. (2004, marzo, pp 10, 21, 39, 41-45). *Gobierno Electrónico y Servicios Públicos*. Recuperado el 18 de octubre de 2006, de http://www.uoc.edu/in3/pic/esp/pdf/PIC_Generalitat_esp.pdf

3. Andrea Di Maio, A., McClure, D., Harris, R. (2007). *A New E-Government Assessment Framework*. Recuperado el 15 de noviembre de 2006, de http://www.gartner.com/DisplayDocument?ref=g_search&id=502580&subref=simplesearch
4. Arveson, P. (1998, marzo, pp 1-2). *What is the Balanced Scorecard?*. Recuperado el 3 de marzo de 2006, de <http://www.balancedscorecard.org/basics/bsc1.html>
5. Asociación Mexicana de Internet. (2004). *Hábitos de los usuarios de Internet en México 2003*. Recuperado el 24 de marzo de 2006, de <http://www.amipci.org.mx>
6. Ayala, J. (2004). *Mercado, Elección Pública e Instituciones*. México, D.F., México: Miguel Angel Porrúa.
7. Benchmarking E-government A Global Perspective. (2002, pp 10-12, 22, 25-27, 49). *Assessing thr Progress of the UN Member States*. Recuperado el 15 de marzo de 2006, de <http://unpan1.un.org/intradoc/groups/public/documents/UN/UNPAN021547.pdf>
8. Benko, Catherine; McFarlan, Warren. (2003). *Connecting the dots: aligning projects with objectives in unpredictable times*. Estados Unidos: Harvard Business Scholl Press.
9. Berčič, B., Dečman, M., Kunstelj, M. y Vintar, M. (2003, pp 4, 6, 8). *Integral e-government development indicators*. Recuperado el 24 de febrero de 2006, de <http://unpan1.un.org/intradoc/groups/public/documents/UN/TC/UNPAN009502.pdf>
10. Bickerton, Pauline; Bickerton, Matthew; Simpson-Holley, Kate. (2000). *Ciberestrategia*. México: Prentice Hall.
11. Bonham G.M., Seifert J. (2003). *The Transformative Potential of E-Government in Transitional Democracies*. *Public Management. Electronic journal, Issue Nº 2*. Recuperado el 19 de Mayo de 2006, de [http://e-journal.spa.msu.ru/images/File/2003/bonham\(1\).pdf](http://e-journal.spa.msu.ru/images/File/2003/bonham(1).pdf)
12. Carbo, T., y Williams, J. G. (2004, pp 95-97). *Models and Metrics for Evaluating Local Electronic Government Systems and Services*. Recuperado el 10 de febrero de 2006, de <http://www.ejeg.com/volume-2/volume2-issue-2/v2-i2-art3.htm>
13. Cassidy, Anita. (1998). *A practical guide to information systems strategic planning*. Estados Unidos: St. Lucie Press.

14. Cassidy, Anita; Guggenberg, Keith. (2001). A practical guide to information systems process improvement. Estados Unidos: St. Lucie Press.
15. Cazemier, Jaques; Oberveck, Paul; Peters, Loukm. (2000). *Securiy management. The stationary office*. Reino Unido: The starionary office.
16. Center for Democracy & Technology and InfoDev. (2002). *The e-Government Handbook for Developing Countries*. Recuperado el 10 de febrero de 2006, de <http://www.cdt.org/egov/handbook/2002-11-14egovhandbook.pdf>.
17. Cornish, Randy. (2003). *.NET Solutions architectures*. Estados Unidos: QUE.
18. Deitch, A., West, D. (2004, enero). *Foro de Mejores Prácticas de Gobierno Electrónico de las Américas – Desarrollo de Estrategias de Gobierno Electrónico en Chile, Canadá y Brasil*. Recuperado el 17 de octubre de 2004, de <http://icamericas.net/index.php?module=htmlpages&func=display&pid=5832>
19. Devaraj, Servanan; Kohli, Rajiv. (2002). *The IT payoff: mesasuring the business value of information technology investments*. Estados Unidos: Prentice Hall.
20. Finch, B. (2002). *Cómo desarrollar un Plan de Negocios*. Barcelona, España: Editorial Gedisa, S.A.
21. Fountain, J.E. (2004, julio, pp 2, 4-8, 33). *Prospects for the Virtual State*. Recuperado el 17 de mayo de 2006, de <http://www.j.u-tokyo.ac.jp/coeps/040710.pdf>
22. Fraga, E., Kreizman, G. (2003, 13 de febrero). *Best Practices in County E-Government*. Recuperado el 24 de octubre de 2004, de <http://ww.cira.state.tx.us/Docs/docs/bestpractices.pdf>
23. Gallardo Cervántes, Juan. (1998). *Formulación y evaluación de proyectos de inversión, un enfoque de sistemas*. México: McGraw Hill.
24. Galtung, Johan. (1966). *Teoría y métodos de la investigación social (Tomo I)*. Argentina: Editorial Universitaria de Buenos Aires.
25. Gant, D.B., Gant, J.P. y Jhonson, C. L. (2002, enero, pp 11, 13, 16-19, 40). *State Web Portals: Delivering and Financing E-Service*. Recuperado el 4 de mayo de 2006, de http://www.aboutweb.org/file/gov_portal1.pdf
26. Gardner, Christopher. (2000). *The valuation of information technology: a guide for strategy development, valuation, and financial planning*. Estados Unidos: John Wiley and Sons.

27. Gil-García, J.R. (2006, pp 1-5). *Enacting State Websites: A Mixed Method Study Exploring E-Government Success in Multi-Organizational Settings*. Recuperado el 19 de mayo de 2006, de directamente del autor.
28. Gil-García, J.R. y Martínez-Moyano, I.J. (2005, abril, p.5). *Exploring E-government Evolution: The Influence of Systems of Rules on Organizational Action*. Recuperado el 19 de mayo de 2006, de directamente del autor.
29. Gil-García, J.R. y Sandoval, A. R. (2006, pp 368-369). *E-Government Portals in Mexico*¹. Recuperado el 19 de mayo de 2006, de directamente del autor.
30. Gobierno de Chile. (2003). *Guía para el desarrollo de sitios Web*. Chile: Ministerio Secretaría General de Gobierno.
31. Gobierno del Reino Unido. (2005). *Guidelines for uk government websites - framework for local government*. Reino Unido: Cabinet Office.
32. Gómez de Silva, Guido. (2001). *Breve diccionario etimológico de la lengua española*. México: FCE.
33. Greenstein, Marilyn; Vasarhely, Miklos. (2004). *Electronic commerce. Security, risk management, and control (2ª ed.)*. Estados Unidos: McGraw Hill.
34. Groeneveld, Kees; Hoogerbrug, Johan. (2002). *La nueva economía de servicios*. México: Prentice Hall.
35. Gutiérrez, I. (2003, noviembre, pp 1-3). *La ciencia de diseñar portales de clase mundial: diseño con usabilidad y enfocado en el usuario*. Recuperado el 15 de enero de 2006, de http://www.axitia.com/IGa0004-0311-pol_dig-la_ciencia_de_disenar_portales.pdf
36. Gutiérrez, I. (2005, octubre, pp 5-6). *Estudio Comparativo de Portales Estatales Gubernamentales*. Recuperado el 5 de mayo de 2006, de http://www.axitia.com.mx/AX0261-03-Benchmark_portales_estatales-gratis.pdf
37. Hammer, M. (2001, agosto). *Making eGovernment Work*. Recuperado el 24 de octubre de 2004, de <http://www.nga.org/cda/files/EGOVHAMMER.pdf>
38. Hance, Olivier. (1996). *Leyes y negocios en Internet*. México: McGraw Hill
39. Heeks, R. (2003, marzo, pp 1-2). *eGovernment for Development Succes and Failure Rates of eGovernment in Developing/Transitional Countries: Overview*. Recuperado el 10 de abril de 2006, de <http://www.egov4dev.org/sfoverview.htm>

40. Holmes, Douglas. (2003). *e-Gob. Estrategias para el uso eficiente de Internet en el gobierno*. México: McGraw Hill.
41. Instituto Nacional de Estadística, Geografía e Informática. (2001). *Situación de la informática en México*. Recuperado el 12 de octubre de 2006 de http://www.inegi.gob.mx/inegi/contenidos/espanol/prensa/Boletines/Boletin/Comunicados/Especiales/2001/Marzo/cp_28.pdf
42. Instituto Nacional de Estadística, Geografía e Informática. (2001). *XII Censo general de población y vivienda 2000*. Recuperado el 12 de octubre de 2006, de <http://www.inegi.gob.mx>
43. Instituto Nacional de Estadística, Geografía e Informática. (2006). *II Censo de población y vivienda 2005*. Recuperado el 12 de octubre de 2006, de <http://www.inegi.gob.mx>
44. Instituto de Investigaciones Diccionario de Derecho Administrativo. (2003). *Diccionario de Derecho Administrativo*. México, D.F., México: Porrúa.
45. Instituto de Investigaciones Jurídicas. (1999). *Constitución Política de los Estados Unidos Mexicanos comentada (14ª. ed., Tomo 1)*. México, D.F., México: Porrúa.
46. Instituto de Investigaciones Jurídicas. (1999). *Constitución Política de los Estados Unidos Mexicanos comentada (14ª. ed., Tomo 2)*. México, D.F., México: Porrúa.
47. Instituto de Investigaciones Jurídicas. (1999). *Diccionario Jurídico Mexicano (13ª. ed., Tomo 2)*. México, D.F., México: Porrúa
48. KBSt (2003, diciembre, pp 14, 23, 31, 37, 39, 45). *Standards and Architectures for e-government Applications*. Recuperado el 13 de febrero de 2006, de http://www.willydev.net/Descargas/WillyDev_Saga_2_0_en_final.pdf
49. Keen, Jack. (2002). *Making technology investments profitable: ROI roadmap to better business cases*. Estados Unidos: John Wiley and Sons.
50. Kirakowski, Jurek; Claridge, Nigel. (2003, julio). *Quality Framework for UK Government Website Design: Usability issues for government websites*. Recuperado el 1 de junio de 2006, de <http://archive.cabinetoffice.gov.uk/e-government/docs/qualityframework/pdf/quality.pdf>
51. Kost, J., McClure, D. (2006). *Strategic planning in government*. Recuperado el 1 de junio de 2006, de <http://oit.ohio.gov/IGD/planning/ITPlanningWorkshop/Strategic%20Planning%20in%20Government.pdf>

52. Kunter, Nachtsheim, Neter. (2004). *Applied Linear Regression Models*. Estados Unidos: McGraw Hill.
53. Lazarsfeld, Paul. (1955). *The language of social research*. Estados Unidos: The free press.
54. Lee, John; Ben-Natan, Ron. (2002). *Integrating service level agreement: optimizing your OSS for SLA delivery*. Estados Unidos: John Wiley and Sons.
55. Lind, Douglas; Mason, Robert; Marchal, Williams. (2001). *Estadística para administración y economía*. México: McGraw Hill.
56. Lucas, Henry C. (1999). *Information technology and the productivity paradox: assessing the value of investing in IT*. Estados Unidos: Oxford University Press.
57. Lucas, Henry C. (1999). *Information technology and the productivity paradox: the search for value*. Estados Unidos: Oxford University Press.
58. Martínez, R. (2000). *Derecho Administrativo 1er. y 2º. cursos (4ª. ed.)*. México, D.F., México: Oxford University Press.
59. Microsoft (2005). *Biblioteca electrónica de consulta Microsoft Encarta 2005*.
60. McConnell, Steve. (1996). *Desarrollo y gestión de proyectos informáticos*. España: McGraw Hill.
61. Momentum Research Group of Cunningham Communication. (2000, julio, pp 5, 9-10, 18, 26-27). *Benchmarking the eGovernment Revolution*. Recuperado el 8 de abril de 2006, de http://www.nicusa.com/pdf/EGOV_Benchmark.pdf
62. Moschella, David. (2003). *Customer-driven IT: how users are shaping technology industry growth*. Estados Unidos: Harvard Business Scholl Press.
63. Murch, Richard. (2001). *Project management. Best practices for IT professionals*. Estados Unidos: Prentice Hall.
64. Murphy, Tony. (2002). *Achieving business value from technology: a practical guide for today's executive*. Estados Unidos: John Wiley and Sons
65. Naciones Unidas. (2002). *Benchmarking e-Government: A global perspective*. Estados Unidos: Naciones Unidas.
66. National Governors Association. (2001, 26 de julio). *Building Better eGovernment: Tools for Transformation*. Recuperado el 24 de octubre de 2004, de <http://www.nga.org/center/egovernment/>

67. Organización para la Cooperación y el Desarrollo Económico. (2003). *OCDE e Government Studies: The e-Government imperative*. Estados Unidos: OCDE.
68. Organización para la Cooperación y el Desarrollo Económico. (2005). *OCDE e Government Studies: Mexico*. Estados Unidos: OCDE.
69. Office of Government Commerce. (2000). *Best Practice for Service Support [Mejores Prácticas para Servicios de Soporte]*. Londres, Inglaterra: The stationary office..
70. Office of Government Commerce. (2000). *Service Support (IT infrastructure series)*. Reino Unido: The stationary office.
71. Olías de Lima, Blanca. (2001). *La nueva gestión pública*. México: Prentice Hall.
72. OMEBA Diccionario Jurídico. (2008). *Teoría general del estado*. Recuperado el 7 de Noviembre de 2008, de <http://www.omeba.com>.
73. O'Neill, E., Lavoie, B. y Bennett, R. (2003). *Trends in the Evolution of the Public Web 1998 – 2002. D-Lib Magazine, Volume 9 Number 4* Recuperado el 12 de abril de 2006, de <http://www.dlib.org/dlib/april03/lavoie/04lavoie.html>
74. ORACLE. (2002). *e-government: Conectando Ciudadanos, Empleados y Proveedores*. Recuperado el 12 de Febrero de 2006, de <http://unpan1.un.org/intradoc/groups/public/documents/un/unpan006998.pdf>
75. Osborne, David; Gaebler, Ted. (1995). *La reinención del gobierno*. España: Paidós.
76. Osborne, David; Plastrik, Meter. (2003). *Herramientas para transformar el gobierno*. España: Paidós.
77. Pagano, Robert R. (2006). *Estadística para las ciencias del comportamiento*. (7ª. Ed.). México: Thompson.
78. Payne, Adrian. (1993). *La esencia de la mercadotecnia de servicios*. México: Prentice Hall.
79. Ramírez, F. (2005). *Aprenda Practicando ASP.NET*. Monterrey, Nuevo León., México: Aprenda Practicando Ediciones.
80. Rapp, Stan; Martin, Chuck. (2002). *Max-e-Marketing*. México: McGraw Hill.
81. Real Academia Española. (1992). *Diccionario de la Real Academia Española 1992*.
82. Remenyi, Dan. (2000). *Effective seasurement and management of IT costs and benefits*. Estados Unidos: Butterworth - Heinemann.

83. Ritchey, Ferris J. (2002). *Estadística para las ciencias sociales*. México: McGraw Hill.
84. Schmelkes, Corina. (1998). *Manual para la presentación de anteproyectos e informes de investigación (tesis)*. México : Oxford University Press.
85. Simsion Bowles & Associates. (1998, agosto). *Online Service Delivery*. Recuperado el 17 de octubre de 2004, de <http://egov.vic.gov.au/pdfs/online.pdf>
86. Strassman, Paul. (1999). *Information productivity: Assessing information management cost of U.S. corporations*. Estados Unidos: Information Economics Press.
87. Tardugno, Anthony; DiPasquale, Thomas; Matthews, Robert. (2000). *IT Services. Costs, metrics, benchmarking, and marketing*. Estados Unidos: Prentice Hall.
88. The Working Group on E-Government in the Developing World. (2002, abril, p.14). Roadmap for e-Government in the developing world. Recuperado el 17 de noviembre de 2004, de <http://www.pacificcouncil.org/pdfs/e-gov.paper.f.pdf>
89. Turner, M. (2004) *Global technology trends in e-government..* Recuperado el 5 de mayo de 2006, de <http://www.comnet-it.org/egovernment/E-GovernmentTechnologyTrends-GTEC-2004.pdf>
90. Vassos, Tom. (1996). *Estrategias de mercadotecnia en Internet*. México: QUE.
91. W3C. *Core techniques for web content accessibility guidelines 1.0*. Recuperado el 7 de febrero de 2006, de <http://www.w3.org/TR/WCAG10-CORE-TECHS/core-techniques.pdf>
92. W3C. (1999, marzo, p.2) *User Agent Accessibility Guidelines..* Recuperado el 7 de febrero de 2006, de <http://www.w3.org/TR/1999/WAI-USERAGENT-19990331>
93. West, D.M. (2005, septiembre, pp 4,8-13). *State and Federal E-Government in the United States*. Recuperado el 12 de febrero de 2006, de <http://www.insidepolitics.org/egovt05us.pdf>

Índice del capítulo

Alfa de Cronbach, 165
Alfa de Cronbach., 172

Análisis costo-beneficio, 139
Análisis de causa efecto, 140

- Análisis de estándares*, 139
Análisis de experiencias, 139
Análisis de riesgo, 138
análisis de varianza (ANOVA), 168
Benchmarking, 138
Benchmarking e-government: a global perspectiva, 139
Benchmarking e-government: a global perspective, 135
CIAPEM, 134
coeficiente de confianza, 149
coeficientes de regresión, 168
constructo, 159
Core techniques for web content accessibility guidelines, 136
distribución de razón F, 168
e-Government Observatory, 135
e-Government Observatory, 139
Estudio Comparativo de Portales Gubernamentales, 133
estudio cuantitativo-correlacional, 164
estudio de tipo correlacional-causal, 163
Estudios teóricos, 140
experimento, 145
F calculada, 168, 185
F tabulada, 185
Factor de Inflación de Varianza, 167
FIV, 167
Forward, 180
gobierno electrónico, 131
Guía para desarrollo de sitios Web, 136
Guías de construcción, 139
Guidelines for UK government websites, Framework for local government, 136
hints, 133
hipótesis, 146
hipótesis estadística, 162
Kolmogorov – Smirnov, 173
Kolmogorov-Smirnov, 165
K-S, 165
medidas de centralidad, 172
medidas de variabilidad, 172
Mejores prácticas, 139
modelo de regresión lineal múltiple, 162, 168
Modelos de factores, 139
Modelos de madurez, 139
muestra estadística aleatoria simple, 149
Pearson, 174
precisión, 149
preguntas de investigación, 145
pruebas no paramétricas, 165
pruebas paramétricas (Pearson), 165
rankings, 133
Relaciones entre factores, 139
Spearman, 165, 174
Standards and architectures for e-government applications, 136
Stepwise, 178
Tabla de distribución normal, 149
The e-government imperative, 135
valor codificado, 159
valores críticos de la distribución F, 171
variabilidad negativa, 149
Web Servers, 133

Análisis multivariante en la administración de inventarios

Contenido

Introducción al caso de estudio.....	198
Planteamiento del problema de investigación.....	198
Antecedentes.....	198
Declaración del problema y propósito del estudio.....	199
Objetivos Específicos de la Investigación.....	200
Justificación y Limitaciones.....	201
Antecedentes e Investigación bibliográfica.....	202
Teoría enfocada a la estratificación de inventarios.....	202
Sistema de punto de re-orden de inventarios.....	202
Control con corrección anticipante.....	203
Técnicas de control dirigidas al futuro.....	203
La teoría de Gantt.....	204
La teoría de PERT.....	204
Obsolescencia, su naturaleza e importancia.....	205
La teoría del ciclo de vida del producto.....	209
Servicio de Inventario.....	212
Prácticas administrativas.....	213
Solución propuesta e hipótesis.....	218
Definición teórica de las Variables.....	218
Comprobación de la hipótesis.....	219
Instrumento de Medición, recolección y diseño de la base de datos.....	221
Diseño de la Investigación.....	221
Población, marco muestral y muestra.....	222
Selección empírica de los ítems.....	225
Instrumento de Investigación.....	226
Evaluación de las propiedades Psicométricas de la escala.....	226
Método de Análisis.....	227
Desarrollo del Diseño experimental.....	229
Construcción de Variables para el sector de la industria de la Transformación.....	231
Modelo de Regresión Lineal Múltiple.....	235
Resultados y discusión.....	243
Modelo con Variables ponderadas para el sector Transformación.....	244
Coeficientes de las variables independientes y su prueba T.....	245
Modelo con Variables escaladas para el sector Transformación.....	247
Aceptación de las variables.....	247
Coeficiente de determinación del modelo.....	247
Valor y significancia de la prueba F.....	248
Coeficientes de las variables independientes y su prueba T.....	249
Conclusiones y líneas de investigación.....	251
Bibliografía.....	253
Índice del capítulo.....	255

Introducción al caso de estudio

El presente trabajo de Investigación trata acerca de la problemática en torno a la obsolescencia de los inventarios. Para ello, se propone de acuerdo a los objetivos de la ciencia, la predicción y el control (Abrahamson 1983), determinar los factores que hacen que un producto obsoleto se encuentre en el inventario de las empresas, considerando a la obsolescencia en los inventarios (OI), (Chaneski 2000, Aichlmayr 2002, Masters 1991) como un producto en desuso, buscando ser explicado por tres factores principales:

el de la obsolescencia en el mercado, por la preferencia del consumidor, medido a través del ciclo de vida del producto (CVP) (Solomon (1999), Birou, Faucett, Magnan (1998), Ryan (1985), Levit (1965)).

el de la incertidumbre tanto en el pronóstico de la demanda como de la confiabilidad en la cadena de suministro, medido a través del Servicio de inventario (SI) Muller (2004), Aichlmayr (2002).

el generado por las prácticas administrativas (PA) utilizadas por las empresas como un factor interno (IMR 2002, Masters 1991).

El método utilizado es el cuantitativo, con un modelo multivariable de regresión lineal múltiple bajo el enfoque positivista. Utilizando el paquete estadístico SPSS. Al final se hace un análisis de los resultados y sus conclusiones.

Planteamiento del problema de investigación

Antecedentes

La administración de inventarios es una de las funciones más importantes dentro de una empresa, porque el inventario es una inversión que afecta el servicio al cliente, al capital de trabajo, y aunque está regularmente operado por la gente de operaciones y producción, está íntimamente ligado a las áreas de Mercadotecnia y finanzas. (Biblioteca práctica de negocios Mc Graw Hill, 1987).

Por otro lado, también está la falta de compatibilidad de metas en la organización, entre finanzas, producción y mercadotecnia. Mientras uno pretende mantenerlos en el nivel más bajo con el máximo de rotación, los otros desean dar un mejor servicio y corridas de producción más grandes buscando tener un mínimo de demanda insatisfecha. Esta relación encontrará su equilibrio de acuerdo a la posición del mercado que se tenga tomando en cuenta el ciclo de vida del producto

(CVP), a la disponibilidad de recursos y a las restricciones de lote mínimo dictadas por las prácticas administrativas (PA) y a la situación particular que se establezca sobre el servicio de inventario(SI). (Biblioteca de negocios McGraw Hill 1987).

El propósito de los inventarios es el de complementar la capacidad de predicción de la demanda, debe mantener el equilibrio entre lo que se necesita y lo que se procesa, también puede ser por fluctuaciones en la demanda, donde una reserva de inventario a la mano supone protección. Puede ser por la inestabilidad del suministro, donde protege por la falta de confiabilidad en el proveedor, y por último por la protección de precios en donde la compra acertada de inventario en los momentos adecuados ayuda a evitar el impacto de la inflación de costos. (Muller, 2004)

En la búsqueda de anticipar escenarios futuros (prevención) para estar preparados con productos para servir la demanda con un inventario suficiente (control), nacen varias teorías, unas orientadas al pasado y a buscar algún patrón de repetición, pudiendo estar orientadas a su valor, su estacionalidad o por su rotación. (Biblioteca práctica de negocios McGraw Hill 1987) y por último, las orientadas a su control dirigido hacia el futuro, que van enlazando eventos programados o encadenados en aras de llegar a un resultado final.

Declaración del problema y propósito del estudio

Una de las inversiones principales de las compañías es precisamente los inventarios de producto terminado. La obsolescencia en los inventarios (OI) por su parte, hace que disminuyan sus rendimientos dado su efecto de baja rotación. La pérdida del grado de liquidez de esos inventarios hace necesario en ocasiones el tener que hacer inversiones adicionales para poder vender lo mismo aun a un precio inferior. Esos incrementos de inventario, aunado a la pérdida del valor de mercado, pueden deberse a cambios en la preferencia del consumidor, medido a través del ciclo de vida del producto (CVP), o a la mala estimación de la demanda, desplazamiento por mejoras tecnológicas, fallas en la cadena de suministro medido a través del servicio de inventario (SI), incrementos en los inventarios por duplicidad de pedidos, fallas en el conteo de los inventarios, etc. generados por las prácticas administrativas de la empresa en particular.

En una investigación realizada por el Consejo de Gestión Logística (Masters 1991), obtuvo como resultado un costo de obsolescencia del orden del 1% del Costo total del inventario.

Enunciado del Problema:

En un mercado global que cambia a un ritmo más acelerado cada vez, se vuelve crítico el fenómeno de obsolescencia en los inventarios de las empresas, por el menor desempeño que éste ocasiona a los inventarios, en un mercado competitivo.

Objetivos

Objetivo General:

Determinar cuáles son los factores que contribuyen a generar inventarios de productos obsoletos en las empresas mexicanas.

Preguntas de Investigación

De acuerdo a Creswell (2003) en los estudios cuantitativos donde se emplea la herramienta de estudio o encuesta, se utilizan las preguntas de investigación para enfocar el propósito del estudio. Las preguntas de investigación que se pretende contestar son las siguientes:

1. ¿Cuáles son los factores que impactan en los inventarios de productos obsoletos en las empresas mexicanas?
2. ¿El ciclo de vida del producto (CVP) afecta positivamente a promover o generar la obsolescencia en los inventarios?
3. ¿El servicio de inventario (SI) (Inventario de seguridad y dependencia en los inventarios) afecta positivamente a promover o generar la obsolescencia en los inventarios?
4. ¿Las prácticas de administración de inventarios (PA) utilizadas ayudan a disminuir la obsolescencia en los inventarios (OI.)?

Objetivos Específicos de la Investigación

- a) Comprobar si existe una relación positiva entre CVP y la obsolescencia en los inventarios
- b) Comprobar si existe una relación positiva entre SI y la obsolescencia en los inventarios.
- c) Comprobar si existe una relación negativa entre PA y la obsolescencia en los inventarios.

Justificación y Limitaciones

Importancia y Justificación del Estudio

Esta investigación pretende de una manera sistemática, enfocar la atención hacia procesos que vayan orientados a atacar el problema de la obsolescencia en los inventarios. A veces pasan inadvertidos los efectos del fenómeno obsolescencia, causando pérdidas en el valor realizable de los inventarios, dificultando el proceso de reposición de productos. Los esfuerzos principales de las empresas están orientados a realizar la venta de los productos vigentes, dejando para después los que están estorbando, los inventarios de productos obsoletos.

Mediante esta investigación podremos encontrar información valiosa sobre el manejo de productos obsoletos en los inventarios de producto terminado de las empresas mexicanas. Podremos dirigir los esfuerzos hacia las actividades o variables que más hayan resultado significativas para mitigar los efectos del fenómeno de obsolescencia en los inventarios.

Por lo anterior, se centra la investigación en medir y comprobar que las variables seleccionadas son los factores que explican a la variable dependiente obsolescencia en los inventarios.

Delimitaciones

De acuerdo a los censos económicos 2004 de la INEGI, se consideró una población de 257 empresas grandes de Nuevo León, consideradas así por contar con más de 251 empleados de la Industria de la Transformación. Se realizó una prueba piloto para afinar el instrumento final de investigación.

Limitaciones

Debido a la baja respuesta por el tipo de preguntas se tomó la decisión de desarrollar el instrumento final con preguntas de percepción

Las encuestas fueron llevadas a cabo telefónicamente, por lo que no se pudo aprovechar el glosario que venía dentro del cuestionario.

Antecedentes e Investigación bibliográfica

Historia de administración de inventarios

Existen varias teorías en la administración de inventarios, unas, su fuente principal son las bases históricas, como los puntos de re-orden, otras, clasifican a los productos como la del ABC, y otras más, van orientadas hacia un encadenamiento de procesos lógicos hacia el futuro como el PERT, Camino Crítico y Gantt. En donde se pretende ver escenarios futuros y rutas posibles de solución.

Las primeras teorías estaban enfocadas al proceso de producción en serie, no importaba qué pensara el consumidor, entonces se hablaba de lotes óptimos de producción buscando la eficiencia de las máquinas, después vino la competencia en donde para determinados tamaños de mercado asociados con su participación, fue menester el tener inventarios para poder surtir lotes parciales de producción o demanda y su punto de re-orden era igual al tamaño de pedido aceptable por la gente de producción (Muller 2004).

Teoría enfocada a la estratificación de inventarios.

Esta teoría nace de la necesidad de concentrar los esfuerzos hacia el manejo de los productos en inventario de acuerdo a unos parámetros seleccionados de orden general, estos podrán ser catalogados ya sea por su valor, antigüedad, peso, etc.

El *método ABC* nace en 1907 cuando el sociólogo y economista Vilfredo Pareto, hace su aportación con la ley que hoy lleva su nombre, la *ley de Pareto*; en donde relacionaba que en la generalidad de las cosas el 20% tiene el 80% del valor, a lo que llamaba minoría vital, y al 80% de las cosas que valían el 20%, sería la mayoría trivial. Esta teoría puede aplicarse a cualquier valor, por ejemplo, si lo que queremos ver es rotación, los productos que representan el 80% de la venta, son el 20% de mayor movimiento, y el 80% de los productos restantes, representan tan solo el 20 % de las ventas (Muller 2004). Esta teoría nos ayuda a centrar nuestra atención bajo el criterio de que se puede segmentar el inventario para trabajar con aquello que más nos interese poner en control.

Sistema de punto de re-orden de inventarios

El sistema de *punto de re-orden* se destaca por tener una demanda independiente que funciona a través su demanda

histórica, el objetivo es satisfacer las necesidades de los clientes (Muller 2004). Busca patrones de repetición esperando se repitan en el futuro. Dentro de las teorías de punto de reorden también se consideran los tiempo de ordenar, tiempo de entrega para ser considerado como producto listo para atender a la demanda.

Control con corrección anticipante

El rezago en el proceso de control administrativo demuestra que, para ser eficaz, el control debe dirigirse al futuro. Ello ilustra el problema de emplear únicamente retroalimentación procedente de la producción de un sistema y de medir esta producción como medio de control. Demuestra asimismo la deficiencia de los datos históricos, como los recibidos de los informes de contabilidad. Una de las dificultades implicadas por el uso de los datos históricos es quien les hace saber en noviembre a los administradores de una empresa, que perdieron dinero en octubre (o incluso en septiembre) a causa de algo que se hizo en julio (Koontz 1988).

Lo que *la teoría de control con corrección anticipante* propone, es que los administradores necesitan para un control eficaz un sistema que les indique, con tiempo para emprender acciones correctivas, que surgirán problemas si no hacen algo al momento. La retroalimentación de la producción de un sistema no es suficiente para el control. Equivale a poco más que un aviso postmortem, y hasta ahora nadie ha encontrado un remedio para modificar el pasado. (Koontz 1988)

El control dirigido al futuro es sumamente menospreciado en la práctica, debido sobre todo a la gran dependencia de los administradores de datos contables estadísticos para efectos de control. Claro que en ausencia de medios de previsión, la referencia a la historia es indudablemente mejor que nada. (Koontz y Bradspies, 1972)

En cierto sentido, un sistema de control con corrección anticipante es en realidad un tipo de sistema de retroalimentación. Sin embargo, la retroalimentación de información ocurre en la parte de insumos del sistema, de modo que pueda ser posible hacer correcciones antes de que se vea afectada la producción del sistema (Koontz 1988)

Técnicas de control dirigidas al futuro

En la búsqueda por planear el futuro, surgen teorías encaminadas a eslabonar los procesos futuros de acuerdo a un programa. Así nacen las teorías de Gantt y Pert.

La teoría de Gantt

Henry Laurence Gantt (1861- 1919) ingeniero norteamericano, amplió la obra de Frederick Winslow Taylor (promotor de la organización científica del trabajo 1856- 1915), con quien colaboró desarrollando el aspecto social de la organización del trabajo

Lo que Henry L Gantt, identificó fue que las metas totales del programa deben considerarse como una serie de planes derivados interrelacionados, que las personas pueden entender y seguir, planear el futuro (Biblioteca Práctica de Negocios 1987)

La teoría de PERT

La planeación de redes, PERT (*Program Evaluation and Review Technique*) por sus siglas en inglés, sirve para analizar redes de tiempo y sus eventos.

PERT se desarrolló en la oficina de proyectos de la marina de los Estados Unidos, su primera asignación fue la planeación y el control del sistema Polaris de armas de 1958, después tuvo una ramificación CPM, método de camino crítico. (Biblioteca Práctica de Negocios 1987)

Consiste en ver los procesos y enlazarlos de tal manera que los primeros son necesarios para la ejecución de los siguientes, y así tratamos de minimizar el tiempo para concluir el último de los procesos.

Podemos apreciar que las teorías, unas orientadas a enfocar la relevancia, la oportunidad de la información para la toma de decisiones, el explicar el pasado buscando su repetición en el futuro, el sistematizar los pasos consecutivos dando un orden lógico, viendo el camino crítico en donde un nodo puede frenar un proceso en donde se debe agilizar colectivamente. Ninguna de estas técnicas de control dirigidas al futuro contempló la obsolescencia, cuando la rotación disminuía, consideraba un desajuste en la predicción, un desorden en el sistema de compras, siempre buscando hacia adentro la explicación.

Podemos observar en la tabla 1, como se desarrolla el marco teórico, partiendo de un nivel básico, hasta un nivel integrador de los elementos que se pretende relacionar

Mapa del Marco Teórico

Nivel del Marco	Variable Dependiente	Variable Independiente I	Variable Independiente II	Variable Independiente III
Teórico	Obsolescencia	Prácticas Administrativas	Servicio de Inventario	C.V.P.
Nivel 1 5%	La Obsolescencia, su naturaleza e importancia	Historia de la administración de inventarios Pareto CCA Gant y PERT	Tipos de Producto <ul style="list-style-type: none"> • Comodities • Tecnológicos • Necesidades Sistema de punto de reorden	General Levitt
Nivel 2 15%	La obsolescencia y sus repercusiones en la optimalidad	IMR	Producto y su rentabilidad Las ventas y los rendimientos decrecientes	Ciclo de Vida del Producto, Solomon CDE Birou
Nivel 3 80%	La obsolescencia como indicador de eficiencia de inventarios.	Mejores prácticas en la actualidad y sus repercusiones	Teoría del comportamiento CVP. Demanda Insatisfecha vs. lenta rotación	Competencia y Globalización
Nivel 4	Factores que influyen en la generación de inventarios obsoletos en las empresas			

Tabla 5.1 Mapa del Marco Teórico. Elaboración propia

Obsolescencia, su naturaleza e importancia.

Si buscamos el significado de la palabra *obsolescencia*, encontramos cualidad de obsolescente; Obsolescente, Que está volviéndose obsoleto o que está cayendo en desuso de acuerdo al Diccionario de la Lengua Española de la Real Academia Española Tomo II (2001). Sin embargo, en la mayoría de los casos se le relaciona con cambios tecnológicos en donde ocasiona la discontinuación del producto, o de sus refacciones (Solomon 1999), otros con dañado o no vendible por maltrato de empaque, etc.

El modelo de obsolescencia que se presenta, se propone dentro de un espectro más amplio, el del desuso. De tal manera que si un producto no se está desplazando o utilizando entonces diremos que se está quedando obsoleto.

Siguiendo con la teoría de inventarios, diríamos entonces que el inverso de la rotación de inventarios es el grado de obsolescencia que está adquiriendo el inventario en un momento dado, y la obsolescencia es circunstancial, es decir, de

acuerdo a esta definición de obsolescencia, un producto que fue obsoleto, puede dejar de serlo, se le puede encontrar un reuso, u otro mercado que si lo demande, de acuerdo al flujo o desplazamiento en los inventarios, no necesariamente llegando a ser una situación permanente. (OI) (Chaneski 2000)

A continuación se describen los tipos de obsolescencia involucrados en el estudio y su comparación. La obsolescencia se puede dar por discontinuación, por indiferencia o por factores Internos:

Obsolescencia por discontinuación, es lo que impide su funcionamiento de manera abrupta, obsolescencia permanente. (Programado) (Fishman, Gandal y Shy 1993)

Obsolescencia por indiferencia, se da cuando el consumidor opta, prefiere, elige, el comprar el producto nuevo y pagar por él, a pesar de que cuenta ya con uno "similar", anterior, en perfecto estado. (Gradual) Obsolescencia No permanente (Provocado) (Hadley 1961)

Obsolescencia por factores internos (PA), la falta de prácticas administrativas del control de inventarios adecuadas, como pueden ser por su clasificación, identificación, rastreo o ubicación, políticas de compra, por énfasis del departamento de ventas (gradual o abrupta) Permanente o no permanente. (No intencionado)

Dado los tipos de obsolescencia mencionados anteriormente, podemos decir que la naturaleza de un producto perteneciente a la oferta de productos, debe ser llamado: necesitado, en alguna medida (frecuencia contra el tiempo) por la demanda existente. Si esta oferta es suministrada desde un sistema de inventarios, lo podemos relacionar con la rotación o desplazamiento del producto a través del tiempo. Mientras exista un flujo o desplazamiento de los productos, estos, no serán obsoletos. Cuando su rotación se hace más lenta (no deseada), decimos que el producto está en el proceso de quedar obsoleto.

Clasificación de la obsolescencia de acuerdo a su naturaleza

Obsolescencia	Velocidad	Tipo	Motivo	Gestor	Variable
Tecnología (Fishman, Gandal y Shy (1993))	Inmediata	Permanente	Programado	Proveedor	Controlada*
Preferencia Harvard Business Review(1959) Hadley(1961)	Gradual	No Permanente	Provocado	Mkting.	Controlada* No Controlada
Factores Internos Chaneski (2000) Aichlmayr (2002)	Inmediata o Gradual	Permanente o No Permanente	No Intencionado	Cualquier Depto. Interno	No Controlada

Tabla 5.2 Clasificación de la obsolescencia de acuerdo a su naturaleza Elaboración propia.
 Cuando se habla de variable controlada a la obsolescencia, se refiere al control de la decisión de hacer obsoleto un producto en su fabricación, no en cuanto al control de la salida de esos productos del inventario.

En la tabla, podemos ver que hay obsolescencia no permanente, esto quiere decir que los productos son funcionales pero no deseados o no disponibles para una demanda de productos particular a un nivel dado de precio. Posiblemente, puedan ser deseados o disponibles a otro nivel de precios o en otro mercado o que fomente el desplazamiento (venta). Entonces dejará de ser obsoleto para la nueva situación.

En 1959 el Harvard Business Review (anónimo), nos da el punto de vista sobre el tema de la obsolescencia como una variable controlada y provocada para incentivar la compra superficial (no necesaria). “La planeación de la obsolescencia” Lo manejan como un problema ético. Fomentando la obsolescencia por “preferencia inducida”. Aquí cabe hacer mención, que los autores en 1959 suponían tener el control porque tenían el control en los lanzamientos de los productos, y al ir innovando, iban incentivando la demanda para generar el consumismo. Fomentaban la obsolescencia, más no la controlaban.

En 1961 Hadley desarrolló un modelo sencillo, que utiliza dos elementos tratando de generalizar su aplicación para la obtención de un modelo óptimo de inventario final. Aquí se maneja como búsqueda el encontrar una variable no controlada, obsolescencia.

(Fishman, Gandal y Shy 1993), nos dan otro punto de vista sobre el mismo tema. Aquí los autores hablan de manejarla como un catalizador para la velocidad del desarrollo tecnoló-

gico. Un equilibrio en el manejo de la obsolescencia planeada. Si el costo del desarrollo tecnológico es suficientemente bajo, el equilibrio involucraría la continua introducción de nuevas tecnologías y la producción de productos no duraderos. De igual manera, si el costo del desarrollo tecnológico es suficientemente alto, el único equilibrio es en el que los avances tecnológicos producen productos duraderos. La obsolescencia planeada (provocada).

Para poder comprender de una manera más clara el rol que juega el producto obsoleto en los inventarios, es necesario verlo desde diferentes ángulos, el del mercado desde el ciclo de vida del producto (CVP), su enlace con el quehacer del día a día de la empresa y su característica holística en la estructura organizacional a través del tiempo. El de las prácticas administrativas (PA), y el de los requerimientos de inventario adicional para compensar las distancias del proveedor, largos tiempos de espera y los inventarios de protección o seguridad dada una demanda de productos incierta y variable provista a través de un servicio de inventario (SI). De esta manera, podrá el área de compras frenar pedidos con información proporcionada por ventas, el área de logística o de inventarios para el cuidado de todos los productos en vías a quedar obsoletos.

Es ahí donde se debe contemplar el volumen mínimo deseado, que garantice por el resto de la vida del producto, no caer en una no-venta por falta de producto, teniendo inventarios sanos con rotaciones aceptables.

La importancia en el manejo de la obsolescencia en inventarios, es que se pueda prevenir un deterioro mayor a un objetivo trazado y sobre él basarse para la toma de decisiones. La idea general es el de minimizar los productos de lento movimiento o nulo movimiento en aras de mejorar el desempeño de los productos en inventario. Por ello la importancia de saber su naturaleza y "comportamiento" para poder programar con algún grado de certidumbre su salida, sin incurrir en altos costos de mantener, ni llenarnos de productos sin movimiento en inventario. Las variables a considerar son el ciclo de vida del producto (CVP), las prácticas administrativas (PA) y el servicio de inventario (SI).

La obsolescencia en los inventarios es una consecuencia de los cambios en el entorno de los productos que la componen. Pueden deberse estos cambios a factores tecnológicos (Fishman, Gandal y Shy 1993), pueden deberse a cambios en la moda, por preferencia (Hadley 1961), o también pueden deberse a factores internos o gubernamentales. Lo importante no es evitar que los productos queden obsoletos, sino más bien, que esto no suceda estando en el inventario de las em-

presas o al menos minimizarlo. Por ello, si entendemos el proceso de cómo llegan a quedar obsoletos los productos en inventario podremos encontrar cuales factores favorecen o disminuyen los efectos del fenómeno obsolescencia en los inventarios(OI) buscando las mejores prácticas, actividad económica, tamaño de la empresa, el enfoque de administración de inventarios utilizado, para no estar vulnerable por completo al fenómeno obsolescencia.

La teoría del ciclo de vida del producto

La teoría del ciclo de vida del producto (Levitt 1965). Nace buscando explicar el fenómeno de cambio en la demanda de los productos, para contestar el porqué un producto es demandado de diferente manera a través del tiempo.

En 1965 Theodore Levitt introdujo el término *ciclo de vida del producto*, en donde armonizaba el producto, inventario y la venta a través del tiempo, con el esfuerzo dirigido de la organización hacia el mayor aprovechamiento de dichas variables de acuerdo a la etapa del ciclo en donde se encontraba. Para ello consideró cuatro etapas dentro de su ciclo de vida:

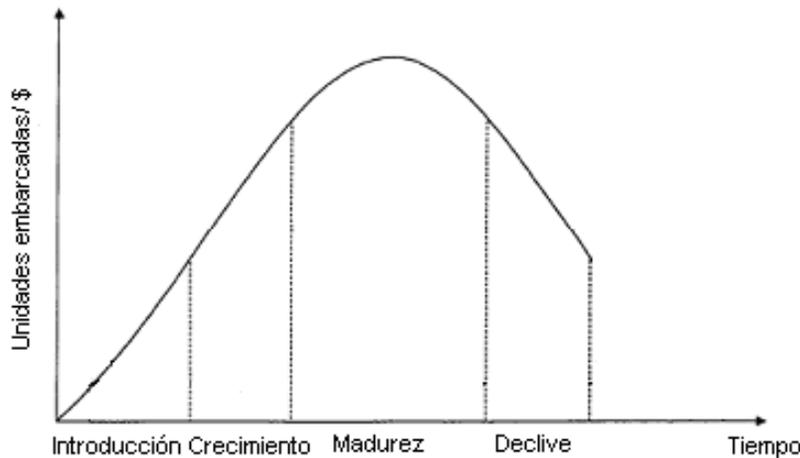
Etapa de Introducción, en ella se consideraba que él producto no se podía mantener por si solo, requería impulso económico y de mercadotecnia para poder escalar a las siguientes etapas.

Etapa del crecimiento, en donde se iba abriendo el mercado al producto, el producto ya pagaba su operación aunque aún no se recuperaba su inversión inicial, en esta etapa ya empiezan a aparecer los competidores buscando una parte de ese nuevo mercado

Etapa de madurez. En ella, ya se obtuvo el volumen de ventas y utilidades óptimo, ya no va a crecer considerablemente, hay que procurar mantener ese nivel y estar muy atentos al menor indicio de cambio de etapa

Etapa de declive, en ella está la salida del producto, en donde ya no es demandado como antes, hay toda una infraestructura que dio servicio a un producto que está saliendo del mercado.

Curva del ciclo de vida del producto de Levitt 1965



Gráfica 5.1 Curva del ciclo de vida del producto de Levitt (1965)

Muchas compañías han fracasado por tener una mala planeación de inventarios, entre ellos la no consideración del ciclo de vida del producto completo, en donde sin darse cuenta, van perdiendo el flujo de dinero. Conforme transcurre el tiempo al estar manejando producto obsoleto, se va haciendo más lenta la venta, y sin darnos cuenta, tenemos un producto más, en su etapa de declive en los inventarios.

En 1985, Ryan considera una nueva etapa, la del *fin de la vida del producto*, *EOL (End of life)* por sus siglas en Inglés. Es Ryan quien establece un fenómeno a considerar que hasta ese momento no se estaba tomando en cuenta, el comportamiento de las diferentes variables al final del camino y sus implicaciones. Entre ellas, qué hacer con los productos no requeridos en inventario, dejar claro que es un tema a considerar, el impacto de los inventarios en su etapa de fin de la vida.

Por otro lado, (Birou et al 1998) desarrolla una teoría en donde propone a un común denominador estratégico (CDE) que sirva para encaminar los esfuerzos de los diferentes departamentos hacia un objetivo superior de la organización que permita, mediante el saber en la etapa en donde se encuentra el producto, cual es el papel que a cada quien le corresponde. El problema se dio en la última etapa, ya que se desarticulaban las sinergias que se habían creado. En su estudio, se comenta la necesidad de estudiar esa etapa en donde se perdió el CDE.

En su estudio, Birou et al (1998), (ver gráfica 4), hicieron una encuesta al personal de diferentes departamentos, 159 ejecutivos de producción, 126 de compras y 133 de distribución. Demostraron que mientras el producto pasa por su desarrollo, crecimiento y madurez, existe una gran participación por lo general de todas las áreas de la organización, en las que manejan los sistemas logísticos, integración interdepartamental por casi todo el proceso. Una vez, que llega a la declinación, y más allá, el confinamiento del producto etiquetado como obsoleto, ya no se administra, se pierde la armonía y se encuentra el desinterés generalizado por atenderlo en esta difícil etapa de su vida.

Es en esta parte, donde nos encontramos dentro de este estudio del *común denominador estratégico (CDE)*, las recomendaciones que hacen en sus conclusiones. Sugieren que se trabaje más el tema de los inventarios obsoletos dado que en la etapa de declive se pudo apreciar un desentendimiento entre los departamentos internos, donde el CDE ya no funcionó como el catalizador orientado a resultados.

(Solomon 1999) propone el *PAO (pronóstico de años a la obsolescencia)* como una ubicación de proximidad a un estado de obsolescencia sobre los inventarios en el mercado de partes de alta tecnología, considerando que la obsolescencia se da por discontinuación, no por desuso. Mide el *grado de obsolescencia (descontinuación)* de los productos de acuerdo a su venta para determinar la etapa del ciclo de vida en el que se encuentra, y estimar cuanto tiempo le queda de vida para que se descontinúe y pare por completo su flujo. Llegando a su obsolescencia. (Solomon 1999)

De tal suerte podemos ver que de acuerdo a la distribución de los puntos a lo largo de la curva, sabemos en qué etapa del ciclo de vida se encuentra el producto. Si los puntos se encuentran dentro de la primera mitad, significa que se tiene una gran confianza de pronóstico, y que además el riesgo a discontinuarse es bajo. Por otro lado, si observamos puntos en ambas mitades de la gráfica, su confianza de predicción es mediana, al igual que su riesgo a discontinuarse. Por último, si se encuentran los puntos en la segunda mitad de la gráfica. Entonces se tiene una alta confianza con una alta probabilidad a discontinuarse.

Con lo anterior, podemos ver como graficando las ventas de los productos en inventario podemos predecir de una forma general si el producto está próximo a entrar en su proceso de obsolescencia de acuerdo a su ciclo de vida.

Servicio de Inventario

El *servicio de inventario* es aquel que apoya para satisfacer a la demanda de productos en un mercado, para ello, deberá de ser suficiente, considerando algún pronóstico de ventas, y también variables tales como tamaño de lote de pedido, tiempo de entrega y el inventario de seguridad dada la incertidumbre en los procesos que pudiera encontrar así como la venta (Muller 2004).

De acuerdo a la rama industrial, a lo competido del mercado, al proceso productivo, a la distancia de los proveedores o insumos y lo crítico que pueda ser la entrega oportuna del producto, cobra relevancia el nivel del inventario de seguridad propuesto. Leenders, Fearon, England (2004) profundizan sobre las consideraciones de la demanda insatisfecha de productos en inventario. La demanda insatisfecha se presenta cuando hay demanda para un producto dado pero que no se puede servir o vender esa unidad por falta de inventario.

El *inventario de seguridad* es un elemento que puede contraponerse con la eficiencia en la búsqueda de contar con una mejor relación de rotación de inventarios ya que por un lado sirve para apoyar la creciente demanda y la imagen de un servicio confiable que atrae más venta, pero por otro lado, puede no incrementar la venta y puede disminuir la rotación de inventarios.

“Hacia las etapas más recientes, el inventario se enfoca más al mercado, necesidades, amenazas de desabasto, prevención. Se está volviendo una actividad más interrelacionada con otras áreas de la organización, como por ejemplo mercadotecnia o ventas, ya no solamente compras” Leenders, et al (1992)

Un efecto sobre la posición competitiva es que una empresa no puede competir a menos que envíe los productos terminados o los servicios cuando se necesitan y al precio que el cliente considera justo. Leenders, et al (1992) por ello es que se deba de contar con un *inventario de servicio* que considere la disminución de la demanda insatisfecha.

Es importante el hacer revisiones de los proveedores de la empresa periódicamente para evaluar su desempeño y su confiabilidad, tanto en el tiempo de entrega como en una calidad acordada en sus productos. El no hacerlo, incidirá en volúmenes mayores de inventario para subsidiar las ineficiencias de los proveedores (Inventario de Seguridad)

Para tratar de minimizar lo anterior, es necesario dentro de dicha revisión el tener más de un proveedor disponible, de no encontrarse, será necesario el desarrollar nuevos proveedo-

res para satisfacer la necesidad de contar con proveedores eficientes.

El servicio de inventario está directamente relacionado con la planeación estratégica, Peter Drucker la define como “el proceso continuo de hacer presentes, sistemáticamente, las decisiones capitalistas (toma de riesgos) con el mayor conocimiento de su futuro; organizando sistemáticamente los esfuerzos necesarios para llevar a cabo estas decisiones; y determinando los resultados de las decisiones contra las expectativas, por medio de una retroalimentación sistemáticamente organizada.” Así es, una estrategia es un plan de acción para cumplir una meta, en el caso del servicio de inventario es el satisfacer a la demanda,

Para efectos del presente estudio la definición de servicio de inventario que utilizaremos es de acuerdo a Muller (2004), Aichlmayr (2002) el servicio de inventario como servicio a la demanda de productos ofrecidos por una empresa, considerando un inventario de seguridad para garantizar dicho servicio y las consideraciones particulares de tiempos de entrega.

Prácticas administrativas

Sobre las *prácticas administrativas* y de las recomendaciones realizadas por los diferentes autores de la revisión de la literatura tenemos a los siguientes:

Marco teórico sobre las prácticas administrativas de Inventarios

Autor	Aportación
a)R.S. Alexander (1964)	Consideración de un responsable dedicado a la desincorporación de productos Obsoletos en Inventario.
b)Chaneski (2000), c)Aichlmayr (2002)	Parámetros de medición
d)Encuesta IMR (2001)	Información empírica de prácticas desarrolladas por un grupo de empresas de los Estados Unidos.
e)Leenders Fearon y England (2004)	Consideración de un puesto clave para la administración de inventarios

Tabla5.3 Marco teórico sobre las Prácticas Administrativas de Inventarios. Elaboración propia

a) *R.S. Alexander (1964)*

Establece que el problema es que no se corre el proceso de eliminación o desincorporación de productos a menos que se establezca claramente quien es el responsable, definir actividades y políticas. La definición de responsabilidades debe de asignarse al menos considerando las siguientes actividades en el proceso:

La selección de productos que son candidatos para eliminarlos;

Recopilación de información acerca de ellos analizando la información;

Tomar decisiones acerca de la eliminación; y

Si es necesario, remover los productos condenados de la línea.

b) *Chaneski (2000)*

Propuso una mecánica para definir la obsolescencia de los productos en inventario de acuerdo a su desplazamiento:

Si un producto no se desplaza en 18 meses, y la parte no es de recién lanzamiento, y no se cuenta con órdenes abiertas de surtimiento, entonces deshazte del 50%.

Una vez que se descartó, si no tiene uso en los siguientes 6 meses y no hay ordenes abiertas, entonces retira otro 50% de la existencia.

Una vez terminada la segunda ronda de reducción de inventarios, si no se usa el material en los siguientes 6 meses, entonces retira la existencia en su totalidad.

c) *Aichlmayr (2002)*

Refiere a Schreibfeder quien menciona que un producto con una rotación de $1 \frac{1}{2}$ en el año es de lento movimiento. Y los que no tuvieron ningún movimiento el año anterior son los inventarios muertos. Recomienda que no más del 10 al 15% del total del inventario sea muerto y del 15 al 20% de lento movimiento.

Recomienda que los pedidos de los de mayor movimiento deban de ser más pequeños y constantes, más que grandes y largos tiempos para pedirlos.

Para el caso de empresas con almacenes en diferentes ciudades, se puede ver la demanda por zona y tipificar los almacenes de acuerdo a la mezcla particular de la zona a la que atiende.

En referencia a lo que maneja Schreiberfeder, aquí no solo vamos a reaccionar con productos muertos, sino también retirar a los que estén agonizando, la idea es mantener en inventario solo aquellos productos que nos favorezcan en la tarea de optimizar el inventario (mayores utilidades)

d) *Encuesta IMR (2001)*

En base a los resultados arrojados por el estudio de reducción de inventarios IMR (2001), se pudo destacar las siguientes prácticas desarrolladas por un grupo de empresas en los Estados Unidos en el 2001

10 maneras de reducir inventarios:

1. Conducen revisiones periódicas 65%
2. Analizan uso y tiempos de entrega 50%
3. Reducción de inventarios de seguridad 42%
4. Utilizan la regla del 80/20 37%
5. Mejoran los ciclos de conteo 37%
6. Repropian al proveedor 34%
7. Re-determinan las cantidades a ordenar 31%
8. Mejoran los pronósticos para los artículos A y B 23%
9. Dan itinerarios a los proveedores 22%
10. Implementan un nuevo software de inventarios 21%

En una encuesta realizada por IMR, se pudo destacar que el 70% de las compañías entrevistadas comentaron que están tomando mediciones para identificar, encontrar y remover sus inventarios obsoletos. Las siguientes son iniciativas que se encontraron a través de los resultados:

1. Haz a alguien responsable del inventario de obsoletos.
2. Periódicamente revisa determinados artículos que son obsoletos
3. Solicitar la utilización de materiales maduros para evitar la obsolescencia
4. Mejor administración del ciclo de vida de los productos.
5. Revisar las reglas de compras en ciertos productos
6. Evita la propiedad de los inventarios
7. Introduce tecnología para encontrar a los candidatos de inventarios obsoletos

8. Mejor coordinación interna entre departamentos.
9. Utilizar una bodega de un tercero como soporte “buffer”
10. Desarrolla un plan de transición para la introducción de nuevos productos
11. Lento, con reducciones estables en contra de soporte de productos con un ciclo largo. Identifícalo y deshazte de él.

Lo importante según este estudio, es no esperarse a su nula movilidad, el responsable del inventario debe estar midiendo el grado de obsolescencia, no solo determinar cuales ya son productos obsoletos y retroalimentar a los demás departamentos involucrados en el sistema.

e) *Leenders, Fearon y England (2004)*

“El problema de la responsabilidad por la administración de la disposición de material de una organización es bastante difícil de contestar. En las grandes compañías, en donde hay cantidades sustanciales de desechos, materiales obsoletos, excedentes o de desperdicio, puede estar justificado un departamento separado. El gerente de ese departamento puede depender del gerente general o del gerente de producción”. Leenders, et al (2004) pp. 434

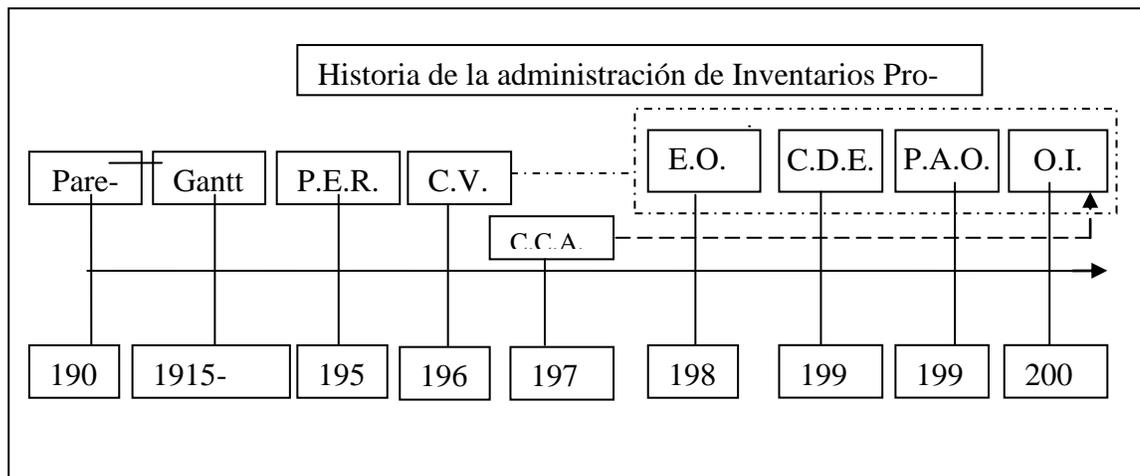
Una de las actividades de la *administración de inventarios* es la de monitorear constantemente el inventario de productos obsoletos (IMR, Sep, 2003) para ir midiendo los cambios en la demanda por *sku* (*stock keeping unit*) con control con corrección anticipante (Koontz 1988), avisando desde adentro del sistema de inventarios, de los cambios inmediatos sobre la demanda, autorregulando el sistema, controlando las entradas, compras, material en proceso y programación de la producción. Promoviendo la salida haciendo ajustes en la política de comercialización, precio, descuento y segmentación de mercado. Y la de inventarios de productos obsoletos, que puede ser creado como resultado del proceso natural “ciclo de vida del producto” interacción con el mercado. Factores externos como preferencias o tecnología, o bien, también por encontrarnos inventarios con vicios comerciales, como falta de promoción y falta de planeación en su lanzamiento (obsolescencia por factores internos).

En un principio, orientados a cuales deben de ser los productos a los que les debemos dar más atención (Pareto), con esto, satisfacíamos el *qué* es lo que debemos cuidar. Poste-

riormente nacen teorías del **cómo** debemos cuidar y vigilar dichos inventarios a través de la planeación del tiempo (Gantt) y de la organización de esos eventos encadenados (P.E.R.T.) buscando la eficiencia.

Por último, nacen las teorías del **porqué** del comportamiento de los inventarios a través del ciclo de vida del producto (CVP). A este ciclo de vida del producto se le han ido incorporando nuevas teorías, como su quinta etapa, donde ya se considera la existencia del fin de la vida del producto (EOL), luego ese patrón bien definido de su ciclo, se trata de utilizar para obtener un común denominador estratégico (CDE) para armonizar las actividades departamentales con el ciclo de vida de los productos. Luego la consideración de la obsolescencia vista como discontinuación y tratar de obtener a través del ciclo de vida del producto un pronóstico de años para la obsolescencia (PAO).

Se propone un nuevo elemento dentro de las teorías de administración de inventarios, que considere en el conjunto de teorías al fenómeno de obsolescencia en los inventarios. Ya que como hemos visto, existen diferentes enfoques y esfuerzos orientados a reducir o eliminar los productos obsoletos en los inventarios de las empresas. EL enfoque particular que en este estudio se considera es el de la obsolescencia por desuso, que puede medirse a través de la rotación de inventarios. Lo que permite que pueda ser o no permanente, dando oportunidad para que a través de prácticas administrativas, conocer su ciclo de vida y observando las necesidades de cumplimiento con ciertos mecanismos del servicio de inventarios, se busque el reducir o eliminar la obsolescencia, que de alguna manera siempre está presente en los inventarios (OI).



Grafica 5.2 Elaboracion Propia

Solución propuesta e hipótesis

Definición teórica de las Variables

De acuerdo al marco teórico revisado, en la sección anterior, a continuación se presenta la tabla que componen a la variable dependiente e independientes con su referencia teórica.

Variable	Definición	Referencia
O.I.	Obsolescencia en los Inventarios, Variable dependiente que mide el desempeño de sus inventarios por su Utilización (rotación).	Chaneski (2000), Aichlmayr (2002) (Masters 1991)
C.V.P	Ciclo de Vida del Producto, que es la venta a través del tiempo de un producto, considerando un comportamiento general predecible	Levitt (1965), Birou (1998), Solomon (1999), Ryan (1985)
S.I.	Servicio de Inventario como servicio a la demanda de productos ofrecidos por una empresa, considerando un inventario de seguridad para garantizar dicho servicio y las consideraciones particulares de tiempos de espera.	Muller (2004), Aichlmayr (2002)
P.A.	Prácticas Administrativas, como políticas internas sobre el manejo administrativo de los inventarios en cuanto a los controles de inventario, estándares de monitoreo, etc.	IMR (2002), Masters (1991)

Tabla 5.4 Definición teórica de las Variables. Elaboración propia.

Hipótesis general de investigación

Los factores que impactan en la generación de inventarios de productos obsoletos en las empresas mexicanas son el ciclo de vida del producto, el servicio de inventario y las prácticas administrativas.

Hipótesis alternas

- Ha1 Existe una relación positiva entre CVP y la obsolescencia en los inventarios
- Ha2 Existe una relación positiva entre SI y obsolescencia en los inventarios
- Ha3 Existe una relación negativa entre PA y la obsolescencia en los inventarios

La hipótesis Ha1 lo que pretende es confirmar la relación que tiene el ciclo de vida del producto con respecto al comportamiento de la variable obsolescencia en los inventarios, que a más avanzado se encuentre el producto dentro de su ciclo, existe una relación de aparición o acentuado del fenómeno obsolescencia.

La hipótesis Ha2 busca encontrar una relación positiva entre la variable servicio de inventario y obsolescencia en los inventarios. Donde el servicio de inventario contempla el uso de inventarios de seguridad (adicionales) para enfrentar eventualidades tales como des-abasto, demoras, etc. que puedan afectar el proceso de la producción o venta, según sea el caso. Estos inventarios adicionales de protección son posibles generadores de baja rotación en los inventarios tendiendo a crear inventarios de productos obsoletos, si esto es así, entonces encontraremos una relación positiva con la variable obsolescencia en los inventarios, relación negativa entre las prácticas administrativas de inventarios (PA) y la obsolescencia.

La hipótesis Ha3 busca encontrar una relación negativa entre la (OI) y (PA), esto es, a una mayor presencia de la variable prácticas administrativas, encontraremos un decremento en los inventarios obsoletos en las empresas.

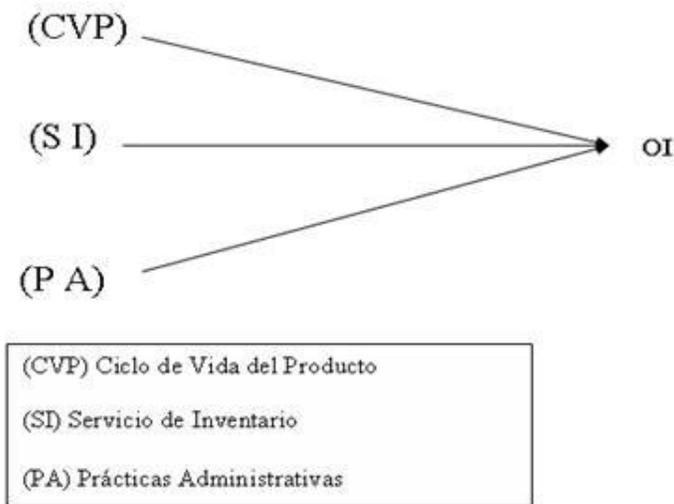
Comprobación de la hipótesis

Se elaboro un instrumento de investigación en forma de encuesta con 32 preguntas para llevar a cabo el levantamiento de los datos, .Se utilizó un análisis multivariado, con el empleo del paquete estadístico SPSS,

Estimando una ecuación mediante el método de regresión múltiple con las variables OI, CVP, SI y PA, arriba citadas. Buscamos el poder observar el impacto y signo de los coeficientes de relación con el que están afectando a la variable dependiente, y de esa manera, poder determinar si su afectación es positiva o negativa, así como la magnitud de su relación y así comprobar las hipótesis de investigación.

Podemos observar la dirección de las flechas de izquierda a derecha, esto nos indica la relación que llevan las variables. Las variables del lado izquierdo son las variables independientes y la de la derecha es la variable dependiente.

Modelo propuesto para la prueba de la hipótesis en formato grafico.



Grafica 5.3 Modelo de obsolescencia en los inventarios.
Elaborada por el Autor

Modelo propuesto para la prueba de la hipótesis en formato matemático.

Modelo de Regresión Lineal Múltiple

Modelo: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 - \beta_3 X_3 + \varepsilon$

En donde la $Y =$ la variable dependiente

$\beta_i =$ coeficiente de regresión

$X_i =$ variables independientes

$\varepsilon =$ error

Sustituyendo por las variables para este estudio quedaría:

$$O.I. = \beta_0 + \beta_1 CVP + \beta_2 SI - \beta_3 PA + \varepsilon$$

Donde:

O.I. = Y valor entre 1 – 5 (VarOI14)

C.V.P.=X1 valor entre 4 – 20 (Var CVP)

S.I. =X2 valor entre 3 – 15 (VarSI)

P.A. =X3 valor entre 1– 2 (Var PA)

Instrumento de Medición, recolección y diseño de la base de datos

El método positivista que estamos llevando a cabo es de tipo analítico, cuantitativo, mediante la técnica de análisis de dependencia o técnica exploratoria utilizaremos una regresión múltiple, en donde tenemos una variable dependiente y tres variables independientes que se pretende, expliquen el comportamiento de la variable dependiente.

Diseño de la Investigación

Con el apoyo en la teoría existente, clásica y de estado de arte, se dará el rigor académico a los índices del modelo para su comprobación empírica en la encuesta de campo.

Se utilizó un instrumento piloto en cuadernillo y en formato electrónico para evaluar el tipo de preguntas a realizar y hacer un primer filtro sobre preguntas que no aportan al estudio. Para la prueba piloto se utilizó el muestreo por conveniencia, para el instrumento definitivo se planeó asignar aleatoriamente el listado de empresas dentro de la muestra poblacional

Se llevó un control de seguimiento a las encuestas entregadas pendientes de contestar para agilizar el proceso.

De la información de este proceso, se pudo observar lo siguiente, que los encuestados no contestaron preguntas donde se les pedían datos duros por sentir que eran confidenciales para la empresa no importando que se tratara de un estudio de investigación científica, por lo que se buscará hacer preguntas de percepción en vez de datos duros. El cuestionario les pareció largo por lo que se trató de eliminar preguntas que se consideren no tan relevantes. Dado que se trata del desarrollo de un nuevo instrumento, se procuró manejar mejor los índices y escalas para poder integrar las diferentes variables en un estudio cuantitativo.

Adicionalmente, de la entrevista al panel de expertos, se tomaron sus comentarios en cuenta para el desarrollo del instrumento de investigación.

Se emplearon gráficas de Gantt para dar seguimiento y marcar avances del cumplimiento del proceso. Se utiliza el método histórico para manejar una secuencia lógica de teorías existentes y sus aportaciones a las teorías más actuales.

Población, marco muestral y muestra

Población

La población se compone de las empresas grandes de Nuevo León de acuerdo a la clasificación de los censos 2004 de la INEGI, se compondrá de empresas que tengan 251 o más personas laborando, de los sectores manufacturero y comercio de acuerdo a los Censos Económicos 2004 “Micro, Pequeña, Mediana y Gran Empresa. Estratificación de los establecimientos” de la INEGI en donde el servicio de la demanda sea principalmente atendido a través de su inventario de servicio, no por orden.

Marco muestral

La muestra se obtuvo considerando el tamaño muestral mediante la aplicación de la fórmula para el cálculo del tamaño muestral de intervalo.

$$2$$

$$n = NPQ / ((N-1) (e/z) + PQ)$$

Donde n= tamaño de la población

n= tamaño muestral

P=porcentaje del evento de interés

Q= complemento de P

e= Error tolerable en porcentaje estimado

z= Confianza, expectativa en porcentaje de estimar correctamente

Muestra

La muestra se compone de acuerdo a los Censos Económicos 2004 “Micro, Pequeña, Mediana y Gran Empresa. Estratificación de los establecimientos” de la INEGI. Los Sectores de la muestra son el de manufactura y comercio de Nuevo León (257 y 279 respectivamente) dando un total de 536 empresas, que con un nivel de confianza del 95%, con un error del 9% y suponiendo la mayor variabilidad, (se consideró una P de .5) nos da una muestra de 100 empresas.

Datos e Instrumentación

Se utilizó un instrumento piloto en cuadernillo y en formato electrónico que se utilizó para una evaluación del tipo de preguntas a realizar y hacer un primer filtro sobre preguntas que no aportan al estudio. Para la prueba piloto se utilizó el muestreo por conveniencia. Se llevó un control de seguimiento a

las encuestas entregadas pendientes de contestar para agilizar el proceso.

Resultados de la Prueba Piloto

Debido a la baja respuesta de encuestas contestadas y las que contestaron con muchas omisiones atribuyendo los encuestados a que eran datos de carácter confidencial, al pedir datos duros, se toma la decisión de desarrollar el instrumento final con preguntas de percepción

Se utilizaron las fases implicadas en el desarrollo de escalas de medida (Jean Pierre Lévy Mangin 2003)

1. Delimitación de los objetivos de la escala
2. Elaboración de Ítems
3. Selección teórica de los Ítems
4. Selección empírica de los Ítems
5. Configuración del Instrumento de medida
6. Evaluación de las propiedades Psicométricas de la escala

Fuente Jean-Pierre Levy Mangin (2003)

Delimitación de los objetivos de la escala

En esta encuesta se considera un nivel adecuado de conocimientos por parte del entrevistado.

El objetivo de la investigación es encontrar los factores que afectan a la obsolescencia en los inventarios, su relación, y cual tiene un mayor impacto sobre el fenómeno obsolescencia en los inventarios.

Elaboración de ítems

Para la elaboración de los ítems se consideró “que el contenido de los mismos haga referencia única y exclusivamente a las facetas incluidas dentro de los límites del dominio conceptual del atributo definido y que el proceso que los ítems generarán en las personas que los contesten sea el adecuado. Estos dos aspectos guardan relación con la validez de la escala, en el primer caso, con la validez de contenido, y en el segundo con la validez de proceso de respuesta”. Jean-Pierre Levy Mangin (2003)

La estrategia utilizada para la creación de los ítems fue considerando (de acuerdo a Jean-Pierre Levy Mangin (2003) el que estén relacionados con el atributo que se pretende medir, representando todos los conceptos incluidos en la definición del atributo y el de ser adecuado para la población de encuestados. Se desarrollaron los ítems utilizados en el estudio so-

portados por los índices y variables consideradas dentro del marco teórico.

La redacción de los enunciados se buscó fuera escrita de una manera clara, que tuviera el mínimo de tecnicismos y de fácil lectura. Considera las escalas de respuesta en niveles en forma secuencial de menor a mayor énfasis en el atributo en cuestión.

Selección teórica de los Ítems

Se llevó a cabo entrevistas con expertos para tratar de detectar variables hasta ahora no consideradas, o variables o Ítems que no consideren pertinentes en el estudio para considerar retirarlas. Los expertos son profesionales en el manejo de inventarios en empresas grandes, en México.

Resultados de las entrevistas con experto

Todos consideran a la obsolescencia de los productos en inventarios como un problema importante que requiere atención (relevancia)

Adicionalmente consideran que existe una relación positiva entre la variable tecnología y obsolescencia en los inventarios, de que existe una relación positiva entre competencia y obsolescencia en los inventarios, de que existe una relación negativa entre los productos *comodities* y la obsolescencia en los inventarios, de que existe una relación positiva entre empresas que están con una alta presencia del fenómeno obsolescencia en el mercado y practicas más eficientes para la desincorporación de inventarios de obsoletos

Las razones por las cuales existen inventarios obsoletos en inventario son por cuestiones de mercado (competencia, gustos y preferencias), de tecnología y por el grado de control interno (prácticas administrativas sobre administración de inventarios)

Consideran que la obsolescencia en los inventarios es algo que se debe buscar el medir y controlar hacia el futuro. Por lo que consideran que lo más importante sobre el fenómeno de obsolescencia en los inventarios es encontrar las causas y tratar de pronosticar el fenómeno. Consideran que la obsolescencia en los inventarios en la actualidad solo se registra como un hecho pasado.

Por lo anterior, como primer paso, se pretende determinar las causas validando los factores que en la teoría encontrada, deberían explicar la obsolescencia, las variables de mercado se miden con el ciclo de vida del producto (CVP), competitividad con el inventario de servicio (SI), que además del precio es una variable fundamental para mejorar el servicio pudiendo

ser esta una ventaja competitiva. Sin embargo, suponemos que pueda ser causa para incrementar la obsolescencia de los inventarios de las empresas. Por último la variable de prácticas administrativas (PA), se están tomando las 10 consideradas en el estudio IMR, como las relevantes de uso de Estados Unidos, para observar en México cuales se usan y qué opinan los responsables de las áreas de administración de inventarios sobre ellas.

Selección empírica de los ítems

Se llevó a cabo una prueba piloto en donde se pretendía probar los Ítems, para ver el grado de consistencia interna y estructura interna.

Por el lado de la consistencia interna se evaluó midiendo el coeficiente alpha de Cronbach, el cual resultó dio muy bajo por lo que se requirió un cambio en las escalas buscando una mayor consistencia entre los Ítems y lo que se pretende probar.

Para evaluar la estructura interna se llevó a cabo el análisis de factores

Configuración del Instrumento de medida

Se consideró dentro de la configuración del instrumento de prueba, el orden de presentación de los Ítems, el manejo de apoyos como la tabla explicativa al final del Instrumento, así como las instrucciones del llenado para facilitar el proceso y la comprensión de los enfoques de administración de inventario utilizado.

El formato utilizado fue el de maceta (flowerpot) Hair (2006), utilizando sus once pasos para el desarrollo del instrumento, los cuales se mencionan a continuación.

1. Transformar los objetivos de investigación en objetivos de información
2. Determinar el método adecuado de recolección de datos
3. Determinar la información requerida para cada objetivo
4. Desarrollar formatos de medición de escala/pregunta
5. Evaluación de la medición de Escala/pregunta
6. Establece el formato de Maceta(Flowerpot) y su layout
7. Evaluación del cuestionario y del layout
8. Obtención de la aprobación del cliente
9. Prueba y revisión del cuestionario
10. Finalizar el formato del Cuestionario
11. Implementar el estudio

Instrumento de Investigación

Se desarrolló el nuevo instrumento final, considerando los resultados de la prueba piloto, y tomando dentro de sus conclusiones la de manejar preguntas de percepción, Este nuevo instrumento tiene varios objetivos,

1. Demostrar que la variable dependiente O.I., es explicada por tres variables independientes, la variable ciclo de vida del producto, de servicio de inventario y la de prácticas administrativas.
2. Ver la relación de las variables independientes sobre la variable dependiente
3. Ver el sentido de las relaciones. Saber cuál es su signo y si es significativo
4. Ver las prácticas administrativas utilizadas en México para el tratamiento de la obsolescencia en los inventarios.
5. Saber cuáles son las más utilizadas, cuáles son las menos utilizadas, y cuáles son las que mejor nos alejan del fenómeno obsolescencia en los inventarios.

Evaluación de las propiedades Psicométricas de la escala

Se busca que la escala mida el atributo que se pretende medir (la validez) y que lo haga con precisión (Fiabilidad).

a) Validez

La validez de contenido se refiere a la correspondencia entre el atributo que se pretende medir y el contenido de la muestra de Ítems que compone la escala. Para ello se está apoyando con el panel de expertos.

La validez sobre el proceso de respuesta, se buscó que dentro del llenado del instrumento, se llevara a cabo un proceso lógico que no predisponga ninguna respuesta por haber contestado alguna anterior.

Estructura interna, se pretende comprobar con el análisis de factores.

La relevancia del Ítem se revisará con el panel de expertos

b) *Fiabilidad*

“La fiabilidad se refiere a la precisión de las puntuaciones que esta ofrece” Jean-Pierre Levy Mangin (2003)

Se buscará comprobar la fiabilidad con el coeficiente alpha de Cronbach en el instrumento de investigación.

El instrumento de investigación fue un cuestionario que se manejó en cuadernillo y electrónico vía Internet. Se obtuvo toda la información de un periodo particular para no afectar la medición del nivel de los inventarios al medir productos que se manejan por temporada.

Los datos se obtuvieron mediante un instrumento que mezcla preguntas de percepción utilizando la escala Likert.

Método de Análisis

El método de análisis multivariado que se utilizó fue *regresión lineal múltiple* (Hair, 2006), es decir, estamos manejando una variable dependiente y tres independientes. El paquete utilizado fue el SPSS.

El propósito del modelo es encontrar el grado de la relación entre la variable dependiente y las independientes, es explicativo, no predictivo. Asimismo, veremos el sentido de dicha relación.

Todos los ítems de las variables utilizan la *escala de Likert* con cinco niveles a excepción de la variable P.A. que son ítems con valores de 1 y 2.

Puntos relevantes de la estadística descriptiva

Dentro de los resultados de esta investigación, se preguntó a los encuestados sobre si consideraban que existía una relación de afectación a los inventarios de productos obsoletos, por cada una de las variables independientes, utilizando la siguiente escala:

1. Muy en desacuerdo
2. Desacuerdo
3. Neutral
4. De acuerdo
5. Muy de acuerdo

Para el ciclo de vida del producto contestaron de la siguiente manera:

<u>Variable Ciclo de Vida del Producto</u>	
5	Estar muy en desacuerdo
23	Estar en desacuerdo
8	Neutral
57	De acuerdo
5	Muy de acuerdo
	63.3% De acuerdo
	28.6% En desacuerdo

Para el Servicio de Inventarios contestaron de la siguiente manera:

<u>Variable Servicio de Inventario</u>	
4	Estar muy en desacuerdo
41	Estar en desacuerdo
8	Neutral
39	De acuerdo
6	Muy de acuerdo
	45.9% En desacuerdo
	45.9% De acuerdo

Para las Prácticas Administrativas contestaron de la siguiente manera:

<u>Variable de Prácticas Administrativas</u>	
4	Estar muy en desacuerdo
36	Estar en desacuerdo
5	Neutral
	40.8% En desacuerdo
46	De acuerdo
7	Muy de acuerdo
	54.0% De acuerdo

Como se podrá observar, dadas las respuestas de los encuestados sobre si afectan las variables independientes a la obsolescencia en los inventarios, nos podemos dar cuenta que las opiniones están muy divididas.

En resumen para esta investigación se utilizó el método, positivista, analítico, cuantitativo, mediante la técnica de análisis de dependencia, utilizando la regresión lineal múltiple. Requiere de una prueba piloto para depurar el instrumento, la consulta a un panel de expertos para seleccionar los ítems a utilizar, se desarrolló el Instrumento final considerando el formato de maceta de Hair (2006), la validez de contenido a través del panel de expertos y la fiabilidad a través de la prueba del alpha de Cronbach, utilizando la escala de Likert.

Desarrollo del Diseño experimental

Codificación de variables

De acuerdo al instrumento de Investigación la codificación de variables es como sigue:

Industria

Variable 1 Sector
 Variable 2 Población
 Variable 3 Ventas

Preguntas Generales

Relevancia

Variable 4 Pregunta04	CVP
Variable 5 Pregunta05	PA
Variable 6 Pregunta06	IS

Ponderación cualitativa

Variable 7 Pregunta07	PA > CVP
Variable 8 Pregunta08	PA > IS
Variable 9 Pregunta09	CVP > IS
Variable 10 Pregunta10	PA > CVP + IS
Variable 11 Pregunta11	IS > CVP + PA
Variable 12 Pregunta12	CVP > IS + PA

Variable Obsolescencia **OI**

Item 13	OI13
Item 14	OI14
Item 15	OI15

Variable Ciclo de Vida del Producto **CVP**

Item 16	CVP16
---------	-------

Item 17	CVP17
Item 18	CVP18
Ítem 19	CVP19

Variable Servicio de Inventario (Inventario de Seguridad) SI

Item 20	IS20
Item 21	IS21
Item 22	IS22

Variable Prácticas Administrativas PA

Ítem 23	PA23
Item 24	PA24
Item 25	PA25
Item 26	PA26
Item 27	PA27
Item 28	PA28
Item 29	PA29
Item 30	PA30
Item 31	PA31
Item 32	PA32

Los primeros tres ítems están relacionadas con la Industria

Los siguientes nueve ítems miden la relevancia y ponderación, son confirmatorias.

Se desarrolló el modelo considerando variables que se componen a través de diferentes Ítems.

Construcción de variables

La variable dependiente es el ítem del cuestionario OI14 dado que nos da en forma directa el nivel de obsolescencia en los inventarios de la empresa.

La OI13 maneja la obsolescencia con respecto a la industria donde participa la empresa en forma competitiva y la OI15 Recode nos dice los meses en inventario que consideran necesario para tomarlo como un producto discontinuado.

Sectorización de la muestra

Dentro de las pruebas del modelo, nos encontramos con Heteroscedasticidad clara, se pudo observar mediante la prueba de Homoscedasticidad por su resultado, que la muestra estaba compuesta por dos grupos con datos con distribuciones diferentes. Se decidió utilizar la muestra general del sector transformación para el análisis de los resultados, se

utilizó la información por sectores para eliminar el problema de heteroscedasticidad y poder operacionalizar el modelo.

Como conclusión, a continuación se construye el juego de variables para el modelo de regresión para el sector transformación, evaluando los índices escalados y ponderados.

Construcción de Variables para el sector de la industria de la Transformación.

a) *Construcción de la variable ciclo de vida del producto (CVP)*

El índice que se creó para representar a la variable ciclo de vida del producto (CVP) es con los ítems **CVP16, CVP17, CVP18 y CVP19.**

Prueba de Alpha de Cronbach para el índice CVP de Transformación

sector__INDUSTRIA_Su compañía pertenece al Sector:	Cronbach's Alpha	N of Items
TRANSFORMACION	,741	4

Tabla 5.5 Prueba de Alpha de Cronbach para el índice CVP de Transformación. Elaboración propia.

Podemos observar que el *alfa de Cronbach* es superior a .700, para el caso del sector de transformación, nos da .741. Por lo que se concluye que si tienen consistencia interna, es decir son parte de un mismo grupo, por lo que se permite generar el índice.

De acuerdo al análisis de factores realizado se pudo observar que todos los Ítems están dentro de un solo componente y que su coeficiente de correlación parcial que mide la correlación existente entre los cuatro Ítems *KMO (Kaiser-Meyer-Olkin)* es superior a .5, para el sector de Transformación nos arroja .688.

Prueba de KMO para índice CVP de Transformación

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,688
Bartlett's Test of Sphericity	Approx. Chi-Square	31,534
	df	6
	Sig.	,000

Prueba de KMO para índice CVP de Transformación

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,688
Bartlett's Test of Sphericity	Approx. Chi-Square	31,534
	df	6
	Sig.	,000

- a. sector__INDUSTRIA_Su compañía pertenece al Sector: =
TRANSFORMACION

Tabla 5.6 Prueba de KMO para índice CVP de Transformación.
Elaboración propia.

Una vez revisadas las pruebas se procede a la integración del índice escalado, se utiliza la fórmula para integrar los ítems en una nueva variable:

$$((\text{sum}(\text{var1}, \text{var2}, \text{var3}, \dots) - \text{valor mínimo}) / (\text{valor máximo} - \text{valor mínimo})) * 100$$

La **var1**, **var2**. Corresponde a los nombres con los que se codificaron las variables que se van a agrupar en un índice. El valor mínimo, es el puntaje mínimo integrado por todas las variables o ítems a agrupar. El valor máximo, es el puntaje integrado por todas las variables o ítems a agrupar (**CVP16**, **CVP17**, **CVP18** y **CVP19**). La nueva variable (índice) es **VarCVP**.

De igual manera, se realizó el índice por cargas o ponderado para comparar ambos índices, la fórmula del índice por cargas es:

$$\text{Sum}((\text{var1} * (\text{carga de var1})), (\text{var2} * (\text{carga var2})), (\text{var3} * (\text{carga var3})), \dots) / ((\text{carga var1}) + (\text{carga var2}) + (\text{carga var3}) + \dots). \text{ La nueva variable ponderada es } \mathbf{VarCVP2pond}.$$

b) *Construcción de la variable Servicio de Inventario (SI)*

El índice que se creó para representar a la variable servicio de inventario (SI) para el sector de transformación se incluyeron los ítems **IS20** e **IS21**.

Se utilizaron las pruebas de consistencia interna y de estructura interna, para el primero se considera el alfa de Cronbach y para el segundo el análisis de factores

Prueba de alpha de Cronbach para índice SI, sector Transformación

sector__INDUSTRIA_Su compañía pertenece al Sector:	Cronbach's Alpha	N of Items
TRANSFORMACION	,712	2

Tabla 5.7 Prueba de alpha de Cronbach para índice SI, sector Transformación
Elaboración propia.

Se concluye que si hay consistencia interna para el índice de la variable servicio de inventari, ya que su valor de .712 resultado mayor al mínimo de .700 para aceptarse la prueba para el sector de la Transformación, se permite utilizar el índice.

Podemos observar también que este conjunto de dos ítems, cumple con la prueba KMO al reportar un resultado igual al mínimo para pasar la prueba de .500, para el caso del sector de Transformación.

Prueba de KMO para índice SI, Sector Transformación.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,500
Bartlett's Test of Sphericity	Approx. Chi-Square	12,448
	df	1
	Sig.	,000

a. sector__INDUSTRIA_Su compañía pertenece al Sector: = TRANSFORMACION

Tabla 5.8 Prueba de KMO para índice SI, Sector Transformación. Elaboración propia.

Dado lo anterior se procedió a elaborar un índice con estos ítems,

La fórmula utilizada para la suma de variables es:

$$((\text{sum}(\text{var1}, \text{var2}, \text{var3}, \dots) - \text{valor mínimo}) / (\text{valor máximo} - \text{valor mínimo})) * 100$$

La var1, var2, etc. Corresponde a los nombres con los que se codificaron las variables que se van a agrupar en un índice. El valor mínimo, es el puntaje mínimo integrado por todas las variables o ítems a agrupar. El valor máximo, es el puntaje in-

tegrado por todas las variables o ítems a agrupar (IS20, IS21)
La nueva variable es **VarIS**.

De igual manera, se realizó el índice por cargas para comparar ambos índices, la fórmula del índice por cargas es:

$Sum((var1*(carga\ de\ var1)),(var2*(carga\ var2)),(var3*(carga\ var3)),...)/((carga\ var1)+(carga\ var2)+(carga\ var3)+...)$. La nueva variable ponderada es **VarIS2pond**.

Construcción de la variable Prácticas Administrativas (PA)

El índice que mejor se ajustó para representar a la variable prácticas administrativas (PA8) es con los ítems **PA23, PA24, PA25, PA27, PA28, PA29, PA30 y PA31**.

Se utilizaron las pruebas de consistencia interna y de estructura interna, para el primero se considera el alfa de Cronbach y para el segundo el análisis de factores.

Prueba de alfa de Cronbach para Variable PA en Transformación.

sector__INDUSTRIA_Su compañía pertenece al Sector:	Cronbach's Alfa	N of Items
TRANSFORMACION	,644	8

Tabla 5.9 Prueba de alfa de Cronbach para Variable PA en Transformación. Elaboración propia.

Del análisis de la prueba alfa de Cronbach, podemos decir que la variable PA para la muestra del sector de la Transformación tiene un alfa de Cronbach débil, que aunque cerca, está por debajo del .700

Prueba de KMO para Variable PA en Transformación.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,524
Bartlett's Test of Sphericity	Approx. Chi-Square	65,132
	df	28
	Sig.	,000

a. sector__INDUSTRIA_Su compañía pertenece al Sector: = TRANSFORMACION

Tabla 5.10 Prueba de KMO para Variable PA en Transformación. Elaboración propia.

La prueba KMO para el sector de transformación fue satisfactoria ya que su resultado de .524 estuvo por arriba del .500 que es el mínimo para pasar esta prueba.

El constructo de Prácticas administrativas, tiene 5 dimensiones y como comentamos anteriormente, el alpha de Cronbach es débil.

La nueva variable de prácticas administrativas es **VarPA8**.

Modelo de Regresión Lineal Múltiple

Una vez validadas cada una de las variables procedemos a construir el modelo:

Modelo:
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 - \beta_3 X_3 + \varepsilon$$

En donde la $Y =$ la variable dependiente

$\beta_i =$ coeficiente de regresión

$X_i =$ variables independientes

$\varepsilon =$ error

Todos los ítems de las variables utilizan la escala de Likert con cinco niveles a excepción de la variable P.A. que son ítems con valores de 1 y 2.

O.I. =Y valor entre 1 – 5 (VarOI14)

C.V.P.=X1 valor entre 4 – 20 (Var CVP)

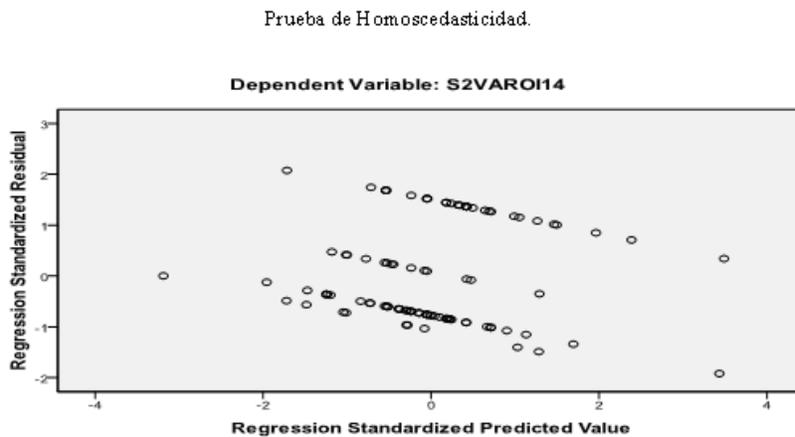
S.I. =X2 valor entre 3 – 15 (VarSI)

P.A. =X3 valor entre 1– 2 (Var PA)

O.I. =
$$\beta_0 + \beta_1 \text{CVP} + \beta_2 \text{SI} - \beta_3 \text{PA} + \varepsilon$$

Supuestos:

- Homoscedasticidad
- Linealidad de la relación
- Normalidad
- No – Autocorrelación (Independencia del término de error)
- Multicolinealidad.



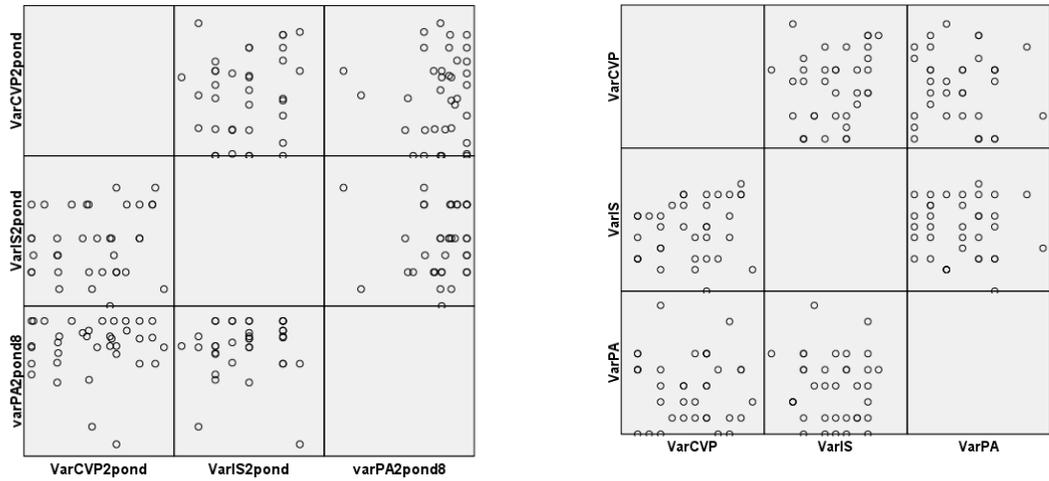
Gráfica 5.4 Prueba de Homoscedasticidad. Elaboración propia

Podemos ver en la gráfica, que se sugiere más de un grupo en los datos. En esta grafica aparecen datos del sector transformación, comercio y servicio. Para eliminar la heteroscedasticidad, se utiliza el sector transformación únicamente.

Linealidad de la relación

De acuerdo a la grafica de humo de cigarro, se puede ver la inclinación de izquierda a derecha en ascenso, esta característica se le atribuye a la linealidad en la relación de los datos, es decir, que la relación de la variable dependiente e independientes es constante a lo largo de todos sus valores.

Prueba de Linealidad a través de gráfica de humo para Transformación.



Prueba de Normalidad para la variable OI14.

		OI14
N		100
Normal Parameters ^{a,b}	Mean	2,64
	Std. Deviation	1,020
	Most Extreme Differences	
	Absolute	,295
	Positive	,295
	Negative	-,209
Kolmogorov-Smirnov Z		2,948
Asymp. Sig. (2-tailed)		,000

a. Test distribution is Normal.

Tabla 5.11 Prueba de Normalidad para la variable OI14.
Elaboración propia.

P-P de Normalidad en la variable dependiente OI14

Prueba de Normalidad para la variable OI14.

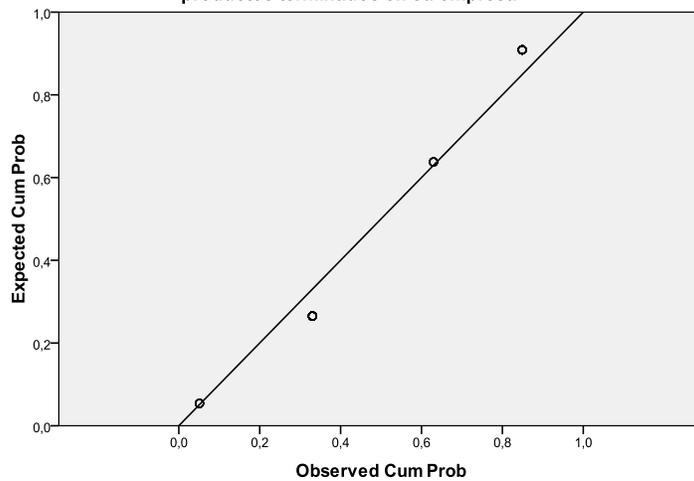
		OI14
N		100
Normal Parameters ^{a,b}	Mean	2,64
	Std. Deviation	1,020
Most Extreme Differences	Absolute	,295
	Positive	,295
	Negative	-,209
Kolmogorov-Smirnov Z		2,948
Asymp. Sig. (2-tailed)		,000

a. Test distribution is Normal.

Tabla 5.11 Prueba de Normalidad para la variable OI14. Elaboración propia.

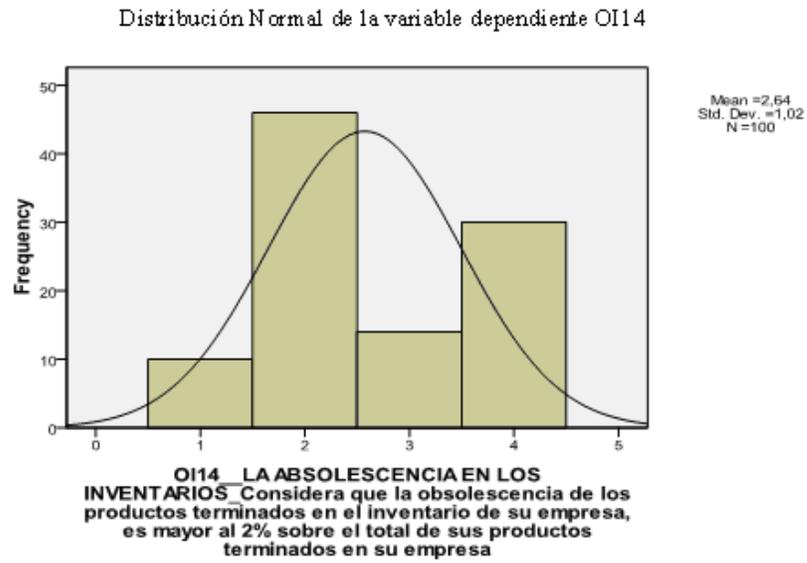
P-P de Normalidad en la variable dependiente OI14

Normal P-P Plot of OI14__LA ABSOLESCENCIA EN LOS INVENTARIOS_ Considera que la obsolescencia de los productos terminados en el inventario de su empresa, es mayor al 2% sobre el total de sus productos terminados en su empresa



Gráfica 5.5 P-P de Normalidad en la variable OI14. Elaboración propia

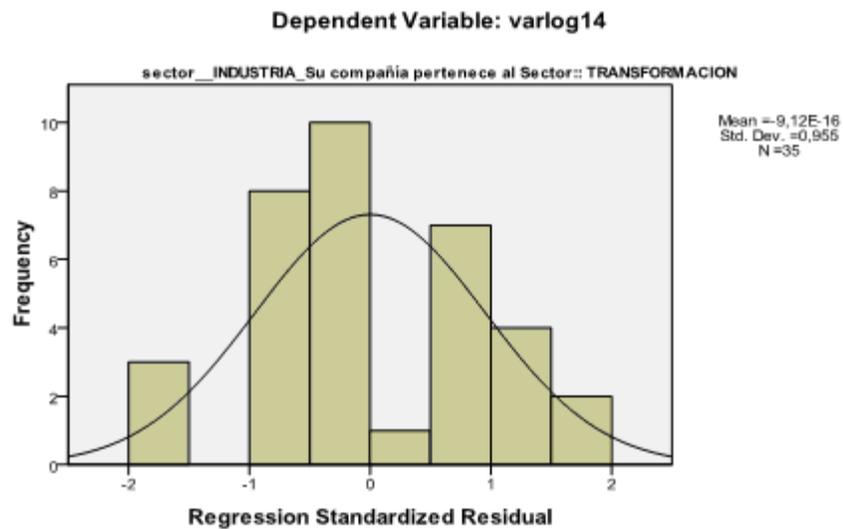
Se transformó la variable OI14 a logarítmica dada la forma de la gráfica



Gráfica 5.6 Distribución Normal de la variable dependiente OI14. Elaboración propia.

La nueva variable de obsolescencia **LogOI14** se normaliza y se adapta mejor a los datos.

Distribución Normal corregida de la nueva variable Log OI14.



Gráfica 5.7 Distribución Normal corregida de la nueva variable Log OI14. Elaboración propia Normalidad en las variables dentro del sector Transformación

Al hacer la prueba de Kolmogorov- Smirnov las tres variables presentan normalidad en el sector Transformación con las variables con índices escalables y ponderados. De acuerdo a la literatura esta prueba se pasa si su valor de significancia es mayor a .05.

Normalidad para las variables ponderadas del Sector Transformación.

		VarCVP2pon d	VarIS2pond	Var- PA2pond8
N		36	36	36
Normal Parameters ^{a,b}	Mean	3,1262	2,8750	1,8061
	Std. Deviation	,76245	,96640	,20746
Most Extreme Differences	Absolute	,123	,183	,193
	Positive	,123	,151	,175
	Negative	-,110	-,183	-,193
Kolmogorov-Smirnov Z		,738	1,100	1,156
Asymp. Sig. (2-tailed)		,648	,178	,138

a. Test distribution is Normal.

b. Calculated from data.

Tabla 5.12 Normalidad para las variables ponderadas del Sector Transformación. Elaboración Propia

Normalidad para las variables escaladas del Sector Transformación.

		VarCVP	VarIS2021	varPA8_2632
N		36	36	36
Normal Parameters ^{a,b}	Mean	53,1250	187,5000	77,7778
	Std. Deviation	19,04529	96,63998	21,15232
Most Extreme Differences	Absolute	,133	,183	,198
	Positive	,127	,151	,147
	Negative	-,133	-,183	-,198
Kolmogorov-Smirnov Z		,799	1,100	1,187
Asymp. Sig. (2-tailed)		,546	,178	,120

a. Test distribution is Normal.

b. Calculated from data.

Tabla 5.13 Normalidad para las variables escaladas del Sector Transformación. Elaboración propia.

No-Autocorrelación

La prueba que se realiza para encontrar la no-autocorrelación, es la *Durbin-Watson*,

Esta prueba revisa el comportamiento sistemático del error, si no existe, el resultado de la prueba debe andar alrededor del valor 2

Para una significancia de .01 con dos variables independientes, los índices ponderados con las cargas de los ítems y el índice sumativo o escalado pasan la prueba.

No autocorrelación en las variables ponderadas del Sector Transformación.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,907 ^a	,824	,818	,40463	
2	,921 ^c	,848	,839	,38053	2,075

- a. Predictors: VarIS2pond
- b. Predictors: VarIS2pond, VarCVP2pond
- c. Dependent Variable: varlog14
- d. Linear Regression through the Origin

Tabla 5.14 No autocorrelación en las variables ponderadas del Sector Transformación. Elaboración propia

No autocorrelación en las variables escaladas del Sector Transformación

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,885 ^a	,782	,776	,44933	
2	,910 ^c	,829	,819	,40430	1,951

- a. Predictors: VarCVP
- b. Predictors: VarCVP, VarIS2021
- c. Dependent Variable: varlog14
- e. Linear Regression through the Origin

Tabla 5.15 No autocorrelación en las variables escaladas del Sector Transformación. Elaboración propia.

Multicolinealidad

Para evaluar la prueba de multicolinealidad se revisa la *Tolerancia* y VIF (Variance Inflator Factor) dentro de la estadística de colinealidad.

Tolerancia = $1 - R^2$,
 VIF = $1 / \text{Tolerancia}$

Donde para valores de VIF arriba de 10 se considera la presencia de alta multicolinealidad

Dado los valores de Tolerancia y VIF, encontrados en el modelo de variables ponderadas, se concluye que existe alta multicolinealidad entre las variables. (Ver tabla 30)

Multicolinealidad en las variables ponderadas del Sector Transformación.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	VarCVP2pond	,143	,081	,486	1,781	,084	,062	16,206
	VarIS2pond	,161	,067	,513	2,418	,021	,102	9,828
	VarPA2pond	-,032	,148	-,063	-,217	,830	,054	18,671

a. Dependent Variable: varlog14

b. Linear Regression through the Origin

Tabla 5.16 Multicolinealidad en las variables ponderadas del Sector Transformación. Elaboración propia.

Dado los valores de Tolerancia y VIF, encontrados en el modelo de variables ponderadas, se concluye que no existe alta multicolinealidad entre las variables al utilizar el método Stepwise en la regresión.

Multicolinealidad con stepwise en las variables ponderadas del Sector Transformación.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	VarIS2pond	,285	,022	,907	12,780	,000	1,000	1,000
2	VarIS2pond	,155	,059	,493	2,626	,013	,126	7,91
	VarCVP2pond	,131	,055	,443	2,361	,024	,126	7,91

a. Dependent Variable: varlog14

b. Linear Regression through the Origin

Tabla 5.17 Multicolinealidad con stepwise en las variables ponderadas del Sector Transformación. Elaboración propia.

Para las variables creadas con el índice escalado, no presenta alta multicolinealidad, ninguna de sus variables del sector Transformación.

Multicolinealidad en las variables escaladas del Sector Transformación.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	VarCVP	,015	,001	,885	11,218	,000	1,000	1,000
2	VarCVP	,009	,002	,517	3,680	,001	,255	3,914
	VarIS2021	,002	,001	,426	3,038	,005	,255	3,914

a. Dependent Variable: varlog14

b. Linear Regression through the Origin

Tabla 5.18 Multicolinealidad en las variables escaladas del Sector Transformación. Elaboración propia.

Resultados y discusión.

Ya una vez revisadas las pruebas de los supuestos para correr una regresión lineal múltiple, analizamos cual índice es mejor para convertir los diferentes ítems en variables, considerando escalado (sumatorio) o ponderado (cargas). Al final, nos lleva a hacer dos modelos:

- a) Modelo con variables ponderadas para el sector Transformación
- b) Modelo con variables escaladas para el sector Transformación

Modelo con Variables ponderadas para el sector Transformación

Aceptación de las variables

Aceptación de variables al modelo.

Model	Variables Entred	Variables Remov	Method
1	VarIS2pond	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
2	VarCVP2pond	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).

a. Dependent Variable: varlog14

b. Linear Regression through the Origin

Tabla 5.19 Aceptación de variables al modelo. Elaboración propia.

Podemos ver que la regresión lineal múltiple elimina a la variable prácticas administrativas (PA) bajo el método *Stepwise*.

Coeficiente de determinación para el modelo.

Coeficiente de determinación del modelo.

Model	R	R Square ^b	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,907 ^a	,824	,818	,40463	
2	,921 ^c	,848	,839	,38053	2,075

a. Predictors: VarIS2pond

b. For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This CANNOT be compared to R Square for models which include an intercept.

c. Predictors: VarIS2pond, VarCVP2pond

d. Dependent Variable: varlog14

e. Linear Regression through the Origin

Tabla 5.20 Coeficiente de determinación del modelo. Elaboración propia.

Podemos observar en los resultados del modelo de regresión lineal múltiple con stepwise que nos arrojó dos modelos, uno con la variable (**SI**) sola, y otro acompañada con el **CVP**. La variable **SI** explica como modelo el 82% de la variabilidad de la variable dependiente obsolescencia en los inventarios **OI**, mientras que el segundo modelo incluyendo la variable ciclo de vida del producto (**CVP**) nos explica el 84%.

Valor y significancia de la prueba F

Podemos observar en la tabla que ambas pruebas F son satisfactorias al estar por debajo de .05 su significancia. Son significativos ambos modelos para la explicación de la variable dependiente **OI**. de las variables ponderadas para el sector Transformación.

Valor y significancia de la prueba F en el modelo.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	26,741	1	26,741	163,332	,000 ^a
	Residual	5,730	35	,164		
	Total	32,471 ^b	36			
2	Regression	27,548	2	13,774	95,122	,000 ^c
	Residual	4,923	34	,145		
	Total	32,471 ^b	36			

a. Predictors: VarIS2pond

b. This total sum of squares is not corrected for the constant because the constant is zero for regression through the origin.

c. Predictors: VarIS2pond, VarCVP2pond

d. Dependent Variable: varlog14

e. Linear Regression through the Origin

Tabla 5.21 Valor y significancia de la prueba F en el modelo. Elaboración propia.

Coefficientes de las variables independientes y su prueba T

Podemos apreciar que las variables aceptadas por los modelos stepwise, son significativas. En el primer modelo la variable **SI** nos da una beta estandarizada de .907 que de acuerdo a Chin (1998) tiene una relación muy fuerte con respecto a la variable dependiente, si la explica. Y su significancia en su prueba t nos dice que si es significativa y apoya al modelo con un valor de .000, menor al .05

Para el segundo modelo, ambas variables tienen betas estandarizadas consideradas de muy fuerte relación con la variable dependiente y sus significancias aceptadas, de .013 para la variable **SI** y de .024 para la variable **CVP**. La VIF está por debajo en ambos modelos de 10 lo que significa que está dentro del rango aceptable de multicolinealidad.

La variable **PA** se rechaza en el modelo con el método stepwise, por un lado, por presentar una alta multicolinealidad, y por el otro, resultó no significativa para el modelo ya que su valor de significancia de la prueba t fue de .970, muy por arriba del límite superior de esta prueba que es .05.

Coefficientes de las variables y su prueba T.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	VarIS2pond	,285	,022	,907	12,780	,000	1,000	1,000
2	VarIS2pond	,155	,059	,493	2,626	,013	,126	7,910
	VarCVP2pond	,131	,055	,443	2,361	,024	,126	7,910

a. Dependent Variable: varlog14

b. Linear Regression through the Origin

Tabla 5.22 Coeficientes de las variables y su prueba T. Elaboración propia.

Modelo con Variables escaladas para el sector Transformación

Aceptación de las variables

Aceptación de variables al modelo.

Model	Variables Entere- red	Variables Remo- ved	Method
1	VarCVP	.	Stepwise (Criteria: Probability-of-F-to- enter <= ,050, Probability-of-F-to- remove >= ,100).
2	VarIS2021	.	Stepwise (Criteria: Probability-of-F-to- enter <= ,050, Probability-of-F-to- remove >= ,100).

a. Dependent Variable: varlog14

b. Linear Regression through the Origin

Tabla 5.23 Aceptación de variables al modelo. Elaboración propia.

Podemos apreciar en la tabla, que de las variables propuestas para el modelo, el método stepwise elimina a la variable **PA**.

Coefficiente de determinación del modelo

Coefficiente de determinación del modelo.

Model	R	R Square ^b	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,885 ^a	,782	,776	,44933	
2	,910 ^c	,829	,819	,40430	1,951

a. Predictors: VarCVP

b. For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This CANNOT be compared to R Square for models which include an intercept.

c. Predictors: VarCVP, VarIS2021

d. Dependent Variable: varlog14

e. Linear Regression through the Origin

Tabla 5.24 .Coeficiente de determinación del modelo. Elaboración propia.

El modelo presenta dos regresiones, en la primera en donde esta **CVP** sola nos da una *R cuadrada* de .782, es decir, que esa sola variable explica el 78.2% de la variable dependiente obsolescencia en los inventarios. En la segunda regresión, están las variables **CVP** y **SI (IS2021)**, ambas en su conjunto arrojan una *R2* o coeficiente de determinación de .829, es decir, que el modelo explica el comportamiento de la variable dependiente, obsolescencia en los inventarios en un 82.9%.

Se confirma la relación explicativa de la variable dependiente.

Valor y significancia de la prueba F.

Valor y significancia de la prueba F en el modelo.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25,405	1	25,405	125,833	,000 ^a
	Residual	7,066	35	,202		
	Total	32,471 ^b	36			
2	Regression	26,914	2	13,457	82,324	,000 ^c
	Residual	5,558	34	,163		
	Total	32,471 ^b	36			

a. Predictors: VarCVP

b. This total sum of squares is not corrected for the constant because the constant is zero for regression through the origin.

c. Predictors: VarCVP, VarIS2021

d. Dependent Variable: varlog14

e. Linear Regression through the Origin

Tabla 5.25 Valor y significancia de la prueba F en el modelo. Elaboración propia.

Podemos observar de acuerdo a la significancia de la prueba F, que el modelo es significativo en su conjunto para explicar a la variable dependiente. Su valor es .000, menor al .05 con el que se pasa la prueba.

Coefficientes de las variables independientes y su prueba T

Coefficientes de las variables y su prueba T

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	VarCVP	,015	,001	,885	11,218	,000	1,000	1,000
2	VarCVP	,009	,002	,517	3,680	,001	,255	3,914
	VarIS2021	,002	,001	,426	3,038	,005	,255	3,914

a. Dependent Variable: varlog14

b. Linear Regression through the Origin

Tabla 5.26 Coeficientes de las variables y su prueba T. Elaboración propia.

Las variables aceptadas sin el método stepwise, son significativas para el modelo a excepción de **PA**, que tuvo un valor de significancia en su prueba t de 773 por lo que se concluye que no es significativa para el modelo, se rechaza la variable **PA**. En el modelo la variable **CVP** (ver tabla 4.26), nos da una beta estandarizada de .517 que de acuerdo a Chin (1998) tiene una relación muy fuerte con respecto a la variable dependiente, si la explica. Y su significancia en su prueba t nos dice que si es significativa y apoya al modelo al obtener un valor de .001, por ser inferior al límite de esta prueba que es .05

La variable SI (Var IS2021), tiene una beta estandarizada de .426 por lo que se considera muy fuerte su relación con la variable dependiente de acuerdo a Chin (1998) y su significancia es aceptada por ser de .005 menor al .05, rango con el que se acepta su significancia

La **VIF** está por debajo en todas sus variables de 10, lo que significa que está dentro del rango aceptable de multicolinealidad.

A través del análisis de los resultados de esta investigación, se destaca el hecho que para el modelo propuesto, buscando explicar a la obsolescencia en los inventarios de las empresas mexicanas, se determinó que, los sectores Transformación y Comercio, se comportan de distinta manera con respecto de las variables de estudio. Se recomienda para futuras investigaciones el manejar una muestra estratificada.

Se analizaron dos tipos de modelo:

- Modelo con variables ponderadas para el sector Transformación

$$\text{Modelo O.I.} = \beta_2 \text{ISpond} + \varepsilon$$

$$R^2 = .824, F = .000, t = .000$$

$$\text{Modelo O.I.} = \beta_1 \text{CVPpond} + \beta_2 \text{ISpond} + \varepsilon$$

$$R^2 = .848, DW = 2.075, F = .000, t = .013, VIF = 7.910$$

$$t = .024, VIF = 7.910$$

- Modelo con variables escaladas para el sector Transformación

$$\text{Modelo O.I.} = \beta \text{CVP} + \varepsilon$$

$$R^2 = .782, F = .000, t = .000,$$

$$\text{Modelo O.I.} = \beta_1 \text{CVP} + \beta_2 \text{SI} + \varepsilon$$

$$R^2 = .829, DW = 1.951, F = 0, t = .001, VIF = 3.914$$

$$t = .005, VIF = 3.914$$

Para el sector Transformación, resultó tener una **R²** más alta el modelo en el que se emplearon variables ponderadas, .848 vs. 829. La variable ponderada de servicio de inventario se ajusto mejor a los datos que la escalada. Y la variable ponderada ciclo de vida del producto, se ajusto mejor a los datos ponderados que en los escalados.

Por el valor de la **R²**, se concluye que el modelo que utiliza las variables ponderadas es el mejor con un 84.8% de explicación de la obsolescencia en los inventarios.

La variable PA aunque fue rechazada por el modelo por no tener significancia en su prueba **t**, mostró sin embargo, signo negativo en su beta, que es el signo esperado en su hipótesis.

Modelo de regresión lineal múltiple para el sector Transformación

$$O.I = .155SI + .131CVP + \varepsilon$$

Se acepta parcialmente la hipótesis general de investigación para el sector Transformación, ya que una de sus variables fue rechazada por no reportar significancia al modelo (**PA**).

Las hipótesis alternas **H1a** y **H2a** se aceptan ya que las variables CVP y SI son significativas y explican el comportamiento de la variable dependiente Obsolescencia en los Inventarios para el sector Transformación.

Conclusiones y líneas de investigación.

Con respecto a las hipótesis de investigación, se declara de la muestra del sector de transformación que:

- a) Se acepta parcialmente la hipótesis general de investigación, ya que se pudieron determinar dos de los factores que contribuyen a generar inventarios de productos obsoletos en las empresas en Nuevo León. Sin embargo la variable Prácticas administrativas, se rechazó ya que de acuerdo a su prueba t, resultó ser no significativa para el modelo.
- b) Se acepta la Hipótesis H1a ya que se demostró que si existe una relación y es positiva entre la variable de ciclo de vida del producto y la obsolescencia en los inventarios, además que de acuerdo a su beta estandarizada, indica tener una relación fuerte con la variable dependiente.
- c) Se acepta la Hipótesis H2a ya que se demostró que si existe una relación, y es positiva entre la variable de servicio de inventario y la obsolescencia en los inventarios en el sector Transformación.
- d) Se rechaza la hipótesis H3a ya que se pudo confirmar que no fue significativa para el modelo, las prácticas administrativas ya que su significancia en la prueba t estuvo por arriba del límite superior de .05.

Implicaciones Teóricas

Se comprobó la relación directa del ciclo de vida del producto, así como de las consideraciones para llevar a cabo un inventario de servicio que incluya dentro de su toma de decisiones, los apoyos que prevengan el des-abasto.

El comportamiento de los datos fue normal, lineal, se pueden seguir haciendo estudios considerando ese comportamiento.

Se desarrollaron dos constructos que representan al ciclo de vida del producto y al servicio de inventarios, así como se determinó su explicación sobre la variable obsolescencia en los inventarios.

El hecho de encontrar cinco dimensiones dentro del constructo prácticas administrativas, implica el abrir esa área de estudio a investigaciones más concretas sobre esa variable y poder seguir con la explicación de la variable obsolescencia en los inventarios para el caso México.

Implicaciones Prácticas

Dado los resultados y conclusiones, lo que se desprende es que la variable del ciclo de vida del producto, debe ser vigilada más de cerca ya que como se ha demostrado a través de esta investigación, es el factor más importante que influye en la generación de productos obsoletos en los inventarios de las empresas mexicanas. Esto significa que ante recursos limitados, el administrador de una empresa, puede dirigir sus esfuerzos hacia la vigilancia del ciclo de vida del producto, la variable que reportó un mayor impacto sobre la obsolescencia en los inventarios.

De igual manera el servicio de inventario es un factor importante, que como se ha demostrado en esta investigación, debe seguirse muy de cerca ya que es un factor que influye en la generación de productos obsoletos en los inventarios de las empresas en Nuevo León.

Futuras líneas de investigación

Tratar de incorporar al modelo, la variable prácticas administrativas (PA), estudiar más su comportamiento buscando tenga un valor de significancia de aceptación.

El aplicar este modelo para grupos más específicos dentro de la Industria de la transformación.

Bibliografía

- Abrahamson Mark (1983), *Social Research Methods*, Englewood Cliffs, EE.UU. Prentice Hall.
- Aichlmayr Mary (2002), *The quick, the dead, and the slow movers, Transportation & Distribution Cleveland Vol. 43*, Iss. 2; p. 38
- Alexander R.S. (1964), *The Death and Burial of "sick" Products*, Journal of Marketing (pre-1986). New York: Apr 1964. Vol. 28, Iss. 000002
- Birou Laura M; Faucett Stanley E. Magnan, Gregory M, (1998), *The product lifecycle: a Tool for functional strategic alignment*, International Journal of Purchasing and materials management. ISSN 1055-6001
- Chaneski Wayne S (2000), *Are you keeping obsolete inventory?* Modern Machine Shop Cincinnati (Oct 2000)
- Creswell John W. (2003), *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, California EE.UU: Sage Publications
- Fishman Arthur, Neil Gandal and Oz Shy (1993), *Planned Obsolescence as an engine of technological progress*. Journal of Industrial Economics
- Groves Robert M, et al. (2004) *Survey Methodology* WILEY ISBN 0-471-48348-6
- Hadley G. (1962) *Univ. de Chicago, Generalizations of the optimal final inventory model*, Management Science Institute (No. 30 Financial Times)
- Hair et al. (2006) *Marketing Research: Within a Changing Information Environment*. N.Y. EE.UU: third edition McGraw Hill International Edition ISBN 0-07-111592-7
- Koontz Harold, (1988), *Administración*. EE.UU. 8va Edición , Mc Graw Hill
- Koontz Harold, 1998, *Administración una perspectiva Global*. EE.UU., Mc Graw Hill, onceava edición ISBN 970-10-2036-7
- Koontz Harold y R.W. Bradspies (1972) *"Managing Through Feedforward Control, Business Horizons vol 15*, No3 June pp 25-36
- Leenders Michiel, Harold E. Fearon, Wilbur B. England (2004) *Administración de Compras y Materiales*, México D.F.: CECSA (1992) Novena reimpresión 2004
- Levitt Theodore (1965) *Exploit the product life cycle*, Harvard Business Review, November – December 1965 pp 81-94

- Levy Mangin Jean-Pierre (2003) *Análisis Multivariante para las Ciencias Sociales*, España: Prentice Hall (2003)
- Masters James M (1991), *A note on the effect of sudden obsolescence on the optimal lot size*. Decision Sciences: Nov/Dec 1991; 22, 5; ABI Inform Global pg. 1180.
- Mc Graw-Hill (1987) *Biblioteca práctica de negocios tomo II*. Administración de operaciones ISBN 968-451-717-3
- Mc Graw Hill (1987) *Biblioteca Práctica de Negocios tomo VII* Mercadotecnia Cáp. 7 Estrategias de la mezcla de productos, Ciclo de vida
- Muller Max, 2005, *Fundamentos de administración de inventarios*, Bogotá Colombia: Grupo Editorial Norma ISBN 958 04 84570
- Ryan Chuck; Riggs Walter E, 1996, *Redefining the product life cycle, the five element product wave*, Business Horizons ISSN 0007-6813
- Solomon Rajeev (1999), *Disertación Doctoral, Life Cycle Forecasting, Mismatch Assessment, and Obsolescence Management*, Faculty of the Graduate school of the University of Maryland at College Park

Índice del capítulo

- administración de inventarios*, 56
- alfa de Cronbach*, 71
- ciclo de vida del producto*, 49
- Común Denominador Estratégico (CDE)*, 51
- descontinuación*, 51
- Durbin-Watson*, 81
- EOL (End of life)*, 50
- escala de Likert*, 67
- fin de la vida del producto*, 50
- grado de obsolescencia*, 51
- inventario de seguridad*, 52
- inventario de servicio*, 52
- KMO (Kaiser-Meyer-Olkin)*, 71
- la teoría de control con corrección anticipante*, 43
- ley de Pareto*, 42
- método ABC*, 42
- obsolescencia*, 45
- PAO (pronóstico de años a la obsolescencia)*, 51
- PERT (Program Evaluation and Review Technique)*, 44
- prácticas administrativas*, 53
- prueba t*, 85
- punto de re-orden*, 42
- R cuadrada*, 88
- regresión lineal múltiple*, 67
- servicio de inventario*, 52
- sku(stock keeping unit)*, 56
- Stepwise.*, 84
- teoría de Gantt*, 44
- Tolerancia*, 82
- VIF (Variance Inflation Factor)*, 82

Capítulo 6:

Análisis y comparación de grupos de trabajo

Contenido

Análisis y Comparación de Grupos de Trabajo.	258
Formación y Análisis de los equipos de trabajo utilizando métodos cuantitativos.	259
Introducción al caso de estudio. El desarrollo de una herramienta generadora de aplicaciones de software y la formación de equipos de trabajo para probarla.....	259
Antecedentes e Investigación Bibliográfica.	262
Propuesta e Hipótesis.	266
Modelo Propuesto para la Hipótesis.	267
Recolección de datos.....	274
Desarrollo y diseño del experimento.....	274
Procedimientos y mediciones utilizadas en el estudio	279
Limitaciones y alcance del estudio.	280
Resultados Obtenidos	281
El Modelo Utilizado.....	287
Análisis de la sensibilidad de los equipos del experimento aplicación del modelo.	287
Comentarios sobre los resultados.....	295
Conclusiones y líneas de investigación futura.....	295
Referencias	296
Índice del capítulo	297

Análisis y Comparación de Grupos de Trabajo.

La primera tarea importante del administrador de un proyecto de software es seleccionar el equipo de trabajo. No solo es una tarea delicada, sino que es crítica para asegurar el éxito del proyecto. El equipo formado debe ser estable y eficiente. Medir su estabilidad y asegurar su eficiencia son los objetivos a cubrir en este capítulo.

Es necesario considerar técnicas cuantitativas que proporcionen una objetividad a prueba de error. La motivación para dedicar a esta tarea el tiempo requerido y efectuarla con lujo de detalle y sumo cuidado se reduce al hecho de que los equipos mal integrados ocasionan proyectos interrumpidos, fuera de presupuesto y de mala calidad, por otro lado los equipos bien integrados sobreviven a los proyectos para seguir produciendo con eficiencia nuevos proyectos y venciendo nuevos retos. El tiempo que se dedica a la formación del equipo se reflejará en los resultados directamente.

Varios autores se han distinguido en este campo, este capítulo se basa en la teoría del auto inventario de perfiles de Belbin (Belbin, 1981), utilizando una versión simplificada de dichos perfiles. La intención es proporcionar una metodología cuantitativa, con fundamentos en estadística básica.

El primer objetivo es para poder comprobar que dos equipos son equivalentes y que por lo tanto son capaces de llevar a cabo proyectos similares en condiciones similares y el segundo objetivo poder proporcionar la capacidad para analizar un equipo y predecir cómo le afecta la partida de uno de sus miembros y que efecto puede tener la llegada de un miembro nuevo al equipo. Todo esto se lleva a cabo mediante la aplicación de cuatro criterios que se han desarrollado basándose en el desempeño de equipos enfocados a desarrollar proyectos de software.

En el caso estudiado se comparan dos grupos que fueron formados por 14 programadores, el primer grupo desarrolló un sistema por métodos tradicionales y el segundo grupo llevó a cabo un experimento en el cual se aplicó la herramienta llamada: Objetos de Software para la Manufactura y la Administración OSMA por sus iniciales desarrollada en base a tecnología de objetos (Prieto, 2008).

La justificación de este módulo tiene su argumento más fuerte en la necesidad de poder predecir el comportamiento de los equipos como un todo para evitar conflictos y fracasos en el desarrollo de proyectos de software. El modelo que se obtiene al aplicar los criterios encontrados en la literatura muestra su eficacia y efectividad en los datos que se mostrarán en este estudio.

Formación y Análisis de los equipos de trabajo utilizando métodos cuantitativos.

La mecánica para formar equipos de trabajo de una manera tradicional se basa en la experiencia y observaciones de las cuales se dispone respecto al trabajo de los candidatos a formar el equipo, la desventaja de este procedimiento es que tiene un alto grado de subjetividad y depende en forma muy importante de la personalidad del tomador de decisiones, dista mucho de ser un método objetivo.

Por otra parte la utilización de perfiles propuesta por Belbin por primera vez en 1981 permite representar las habilidades de los candidatos por medio de una serie de medidas que representan la personalidad de cada miembro, describiremos en detalle cómo se manejan y utilizan estos perfiles y propondremos un mecanismo formado por cuatro criterios para analizar los equipos y predecir posibles conflictos o problemas en base a lo que se ha encontrado por diversos investigadores en el campo del desarrollo de software.

La solución propuesta está basada en el estudio de la matriz que representa al equipo, obteniendo de ella diversos valores estadísticos que permiten su estudio y encuentran los riesgos de presentar diversos problemas típicos de los equipos de trabajo. Para concluir presentamos los resultados y las tendencias de la investigación futura.

Introducción al caso de estudio. El desarrollo de una herramienta generadora de aplicaciones de software y la formación de equipos de trabajo para probarla.

El contexto para aplicar la técnica cuantitativa de integración de equipos de trabajo será el experimento requerido para comprobar los resultados de la herramienta generadora de software basada en Objetos de Software para la Manufactura y la Administración (OSMA) mencionada anteriormente. El proyecto de software resultante consiste en el desarrollo de un sistema que será equivalente a otro ya existente y que fue desarrollado por medio de métodos tradicionales de desarrollo sin utilizar herramienta generadora de software alguna, para comparar los resultados en tiempo y costo de ambos desarrollos.

El experimento busca demostrar que la herramienta OSMA contribuye en una forma substancial al aumento de la productividad del desarrollo de software. Dicho experimento se llevó a cabo siguiendo el siguiente procedimiento:

1. Formar de un equipo de programadores para el grupo experimental equivalente al equipo original.
2. Entrenamiento en la herramienta para el equipo de programadores del grupo experimental.

3. Desarrollar de los mismos programas del grupo de control, midiendo los tiempos de desarrollo.
4. Comparar los resultados utilizando una hipótesis de grupos para analizar el trabajo de los dos equipos de desarrollo.
5. Analizar los equipos de desarrollo para comprobar que además de ser equivalentes técnicamente también lo son en personalidad grupal, a la vez comprar las condiciones en las que se llevaron a cabo ambos desarrollos para comprobar que dichas condiciones son equivalentes, es en éste paso en el cuál es vital comprobar que los dos grupos de trabajo son equivalentes, es aquí donde se utilizó el método propuesto en este capítulo.

El estudio se ha realizado tomando como grupo de control un sistema de información que contiene los objetos de software más utilizados en la manufactura y la administración, dicho sistema ya está siendo utilizado en varias empresas pequeñas y medianas. El software elaborado con la metodología tradicional y con las características en cuanto a funcionalidad de los programas. En éste capítulo se describen a detalle la primera y la última parte del experimento total, que es lo referente a la formación y comparación de los equipos.

Uno de los factores más importantes del experimento es el equipo de trabajo que desarrolla el sistema nuevo equivalente al original, el principal problema es que es necesario garantizar que el nuevo equipo de trabajo es similar al equipo de trabajo que desarrolló el sistema original en el perfil de las habilidades de comunicación interpersonales de sus miembros.

La validez del experimento depende de la integración del nuevo equipo de trabajo y una vez integrado la comprobación del hecho de que sus habilidades son similares o equivalentes a las del equipo de trabajo original. Para poder hacer esto es necesario recurrir a un método cuantitativo, es así como llegamos a éste estudio.

El planteamiento del problema puede establecerse como se muestra a continuación:

Teniendo dos equipos de trabajo es necesario comprobar cuantitativamente que son equivalentes y que ambos tienen habilidades similares de comunicación, coordinación y de relaciones interpersonales. Como punto adicional hemos agregado el requerimiento de poder prevenir el impacto que puede tener el hecho de que un miembro del equipo deje el proyecto y como puede reaccionar el equipo a la llegada de un nuevo miembro que lo sustituya, esto es, hemos agregado la necesidad de analizar la sensibilidad del equipo hacia la personalidad de cada uno de sus miembros.

De esta forma proporcionamos un instrumento que permita orientar al tomador de decisiones en caso de tener que llevar a cabo una modifica-

ción al equipo, evitando que un cambio en su constitución acabe con la eficacia del equipo ya formado.

Los objetivos específicos pueden numerarse como sigue:

1. Desarrollar y probar una metodología cuantitativa de comparación de equipos de trabajo en el contexto del desarrollo de software.
2. Desarrollar y probar una metodología cuantitativa que permita el análisis de sensibilidad de un equipo a la partida de alguno de sus miembros o a la llegada de un nuevo miembro prediciendo el impacto que el evento pueda ocasionar en el comportamiento, la personalidad y el desempeño general del grupo.
3. Establecer las características mínimas básicas suficientes para que un equipo de software sea funcional.

La justificación de este estudio puede fundamentarse en los hallazgos encontrados por Jiang, y Comstock (2007) después de haber analizado más de cuatro mil proyectos de software, dichos hallazgos comprueban que el equipo de trabajo constituye uno de los principales cuatro factores importantes en cualquier proyecto. Esto indica que es uno de los puntos críticos del desarrollo de software.

Cuando se trabaja con equipos surgen preguntas que deben ser respondidas por el administrador, entre otras podemos citar las siguientes:

1. ¿Cuales candidatos a formar un equipo darían los mejores resultados trabajando juntos?
2. ¿Si un dado miembro de un equipo abandona el proyecto durante su realización el resto del equipo formará un equipo funcional y eficiente?
3. ¿Si es necesario agregar un miembro a un equipo y se cuenta con varios candidatos cual sería el más adecuado?
4. ¿Si se requiere prescindir de uno de los miembros del equipo cual de los miembros sería el candidato idóneo?

Este estudio trata de apoyar las respuestas a las preguntas anteriores utilizando argumentos objetivos y cuantitativos.

El estudio de este caso tiene el siguiente alcance: Solo se contemplan los roles de Belbin para cuantificar la personalidad de los equipos, no se utiliza ningún instrumento para medir habilidades técnicas de los equipos, se asume que los candidatos han sido examinados desde el punto de vista técnico y poseen todas las habilidades técnicas necesarias para hacer el trabajo. La comparación de los equipos desde el punto de vista técnico se hace utilizando una variable llamada Tipo de Programador la cual especifi-

ca la experiencia de un programador, solo se mide el número de programadores de un tipo dado y se compara con los programadores del otro equipo con el cual se está comparando.

La sensibilidad de los equipos se analiza desde el punto de vista estadístico básico utilizando para ello los valores que los candidatos han obtenido para cada perfil.

La limitación más importante del estudio es que la estimación original de los perfiles se llevó a cabo por medio de la técnica del auto inventario, esto es: cada miembro del cada equipo opinó a cerca de sí mismo en la mayoría de los casos y en algunos casos para el equipo original cuando hubo ausencia de algún miembro otros miembros que conocieron a la persona y que convivieron con ella respondieron en su lugar.

Otra limitante es que cualquier estudio de esta naturaleza es válido solo para el momento en el cual fue hecho, y depende totalmente del tiempo, pudiendo variar de una manera importante de un momento a otro, puede ser considerado una fotografía de un momento dado.

Uno de los problemas a resolver para el desarrollo de proyectos de software es la integración del equipo de trabajo Parker (1990), éste problema a pesar de no involucrar tecnología es uno de los más riesgosos y de no resolverse satisfactoriamente y darle la importancia que tiene puede paralizar incluso detener los proyectos de desarrollo o instalación de software. Este estudio enfrenta y resuelve el problema de la integración del equipo de trabajo utilizando la teoría de roles.

Antecedentes e Investigación Bibliográfica.

Uno de los problemas a resolver para el desarrollo de proyectos de software es la integración del equipo de trabajo Parker (1990), éste problema a pesar de no involucrar tecnología es uno de los más riesgosos y de no resolverse satisfactoriamente puede paralizar incluso detener los proyectos de desarrollo o instalación de software. Este estudio enfrenta y resuelve el problema de la integración del equipo de trabajo utilizando la teoría de roles.

El problema más común detectado por Belbin es el mismo que se presenta en el desarrollo de software, Belbin le llama síndrome de Apolo, cuando este problema se presenta el ambiente del grupo de trabajo se degrada y se convierte en una competencia agotadora, utilizando la energía del equipo para competir olvidándose de los objetivos del grupo.

Un buen equipo debe tener todos los roles (9) representados y balanceados. El equipo ideal según Belbin es de 6 miembros, otro tamaño, un rango entre 3 a 5 es reportado por Jiang, Naudé, y Comstock (2007) si los roles son 9 es lógico que más de un rol pueden estar en una misma persona. Sin embargo en el desarrollo de software es necesario mantener vigilado el tamaño del equipo, pues como han demostrado Jiang, et. al. (2007), si el equipo es muy grande puede afectar la productividad del grupo.

Rjendran (2005) publicó un estudio realizado sobre equipos de trabajo de Hardware y Software en el que concluye que los perfiles de Belbin bien balanceados contribuyen a la efectividad del equipo, que es la hipótesis fundamental de Belbin. Manning (2006) por su parte concluye que la eficiencia de los equipos de tareas técnicas se ve beneficiada cuando se cuenta con personal introvertido, extremadamente ordenado, detallistas y revolvedores de problemas.

Tipo de tarea

El tipo de tarea a realizar tiene una importancia notable, tal como lo demuestran Higgs, Plewnia, Ploch, (2005) existe una relación directa entre la composición del equipo la complejidad del equipo y la eficiencia de éste. Mediante el estudio de 28 equipos formados por 270 miembros Higos, et al. (2005) establece que:

- La diversidad de los equipos para tareas complejas contribuye a mayor eficiencia
- La diversidad de los equipos para tareas sencillas disminuye la eficiencia

Por lo que si consideramos el desarrollo de software como tarea compleja, que lo es, podemos concluir que la diversidad es un factor importante para la formación de equipos de desarrollo y que la hipótesis de Belbin se ve respaldada por los resultados de Higgs, et al. (2005). Interpretamos aquí la diversidad como el hecho de que todos los perfiles son cubiertos por los miembros del equipo. Este concepto se convierte en un criterio importante para revisar la potencial eficiencia del equipo.

Tamaño del equipo

En cuanto al tamaño del equipo, en el estudio de Jiang, et al. (2007) se analizaron 4,106 proyectos de software llevados a cabo de 1995 al 2005 contenidos en la base de datos *International Software Benchmarking Standards Group* (ISBSG), la cuál contiene información de los proyectos que incluye el tamaño del equipo, la productividad medida por la Razón de Entrega De Productividad Normalizada, en inglés *Normalized Productivity Delivery Rate* (PDR) que es una medida inversa de la productividad, para fines de ilustración y obtener una comparación visual clara, en éste estudio se define la productividad como $100/PDR$. Utilizando datos del estudio de Jiang, et al. (2007), se ha elaborado la gráfica de la Figura 8.1.

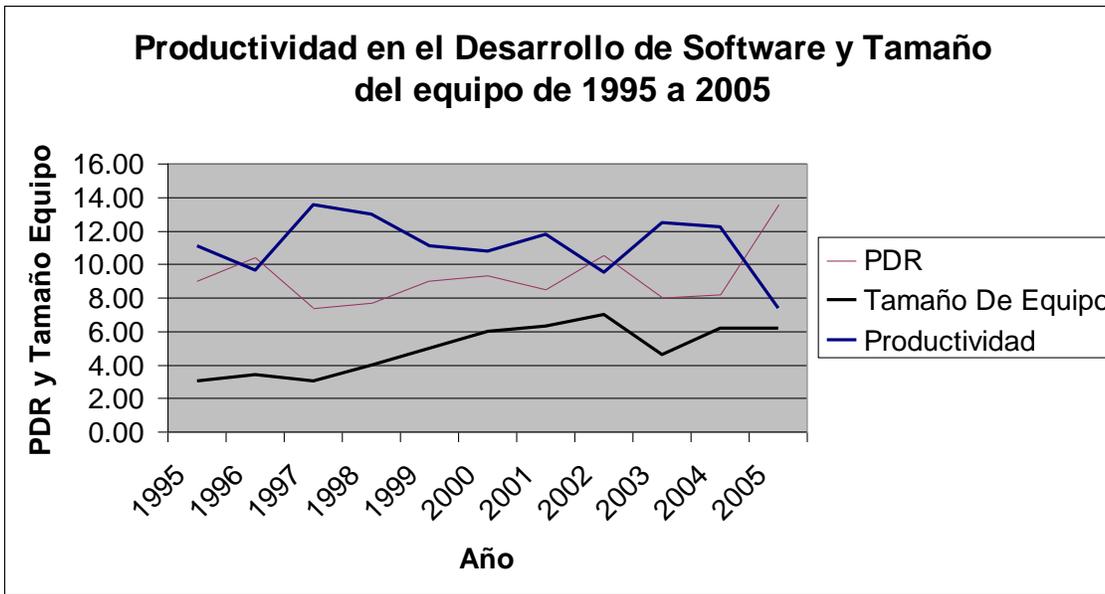


Figura 6.1 Productividad y tamaño del equipo. Considerando $Productividad = 100/PDR$. Fuente: Elaboración propia con datos derivados de de Jiang, et al. (2007)

La línea superior es la productividad y se puede observar como disminuye a medida que el tamaño del equipo (la línea inferior aumenta). Por lo que podemos observar la hipótesis de Jiang, et. al. (May, 2007) que “ en el desarrollo de software los equipos muy grandes tienden a ser más ineficientes ”.

De las investigaciones anteriores podemos deducir las recomendaciones más importantes para integrar los equipos de desarrollo y/o instalación de software:

1.- Sobre el Tamaño: Mantener el tamaño de los equipos entre 3 y 5 personas Jiang, et al. (2007).

2.- Sobre la Diversidad: Cubrir todos los perfiles de Belbin, especialmente aquellos relacionados con habilidades técnicas, disciplinarias y de solución de problemas. Higgs, et al. (2005).

3.- Sobre la Personalidad. Evitar tener varios miembros que tengan acentuado el perfil de líder agudo para evitar el síndrome de Apolo. Belbin (1981).

La tabla 8.1 muestra los nueve perfiles de Belbin, para evaluarlos Belbin desarrollo en 1981 y lo perfeccionó en 1993 un encuesta llamada auto inventario de habilidades, que es un cuestionario que es respondido por el mismo miembro del equipo o por terceros evaluadores que conocen

al miembro del equipo en su ambiente de trabajo, pudiendo ser colegas o amigos. En su libro de 1981 Belbin propone ocho perfiles, el último perfil, el de especialista, fue agregado en su libro de 1993, ajustando también los nombres del resto de los perfiles.

Perfil	Características	Clasificación
TRABAJADOR DE PLANTA	Creador de soluciones novedosas y nuevos paradigmas, gusta de resolver problemas difíciles.	Mental
INVESTIGADOR RECURSOS	Comunicador, desarrollador de contactos, explorador de oportunidades externas al equipo.	Social
COORDINADOR	Confiado, delega responsabilidades, comunica y aclara objetivos, cede la toma de decisiones.	Social
LIDER	Muestra una gran motivación ante la presión, posee gran dinamismo y su eficacia crece con los retos.	De Acción
MONITOR	Analizador de alternativas, revisa las estrategias, juzga y selecciona las mejores opciones.	Mental
TRABAJADOR EN EQUIPO	Amable, político, agradable, excelente conciliador.	Social
IMPLEMENTADOR	Toma conceptos y los convierte en acciones prácticas, es altamente eficiente digno de confianza muestra disciplina y es conservador.	De Acción.
TERMINADOR	Asegurador de la calidad, se entusiasma con revisar el trabajo buscar errores y entregar a tiempo.	De Acción
ESPECIALISTA	Experto en su campo, muestra gran conocimiento y dedicación a su área, es una fuente de conocimiento.	Mental

Tabla 6.1 Tabla de perfiles de Belbin. Fuente: Belbin(1993)

Los perfiles de Belbin fueron clasificados por Belbin atendiendo a la naturaleza de los mismos, esta clasificación se ilustra en la columna de la derecha de la tabla.

La clasificación de los roles o perfiles de equipo, Belbin (1981, 1993) ayuda para entender mejor que representa cada rol y clarifica cada concepto.

Belbin clasifica los roles en Mentales, Sociales y De Acción.

Los roles mentales se caracterizan por contribuir con actividades técnicas y administrativas, son de este tipo los siguientes roles: trabajador de planta, monitor-evaluador y especialista.

Por otra parte los roles sociales tienen que ver con relaciones humanas y de comunicación, estas actividades pueden llevarse a cabo dentro y fuera del equipo, pertenecen a estos roles: investigador de recursos, coordinador y trabajador en equipo.

Los roles de acción se enfocan en las tareas y son: líder, implementador y terminador.

Para poder llevar a cabo una integración de un equipo basándose en los roles es necesario llevar a cabo un procedimiento de evaluación previo, al que se le denomina auto inventario. El auto inventario se forma por una serie de aseveraciones a las cuales el candidato a ser miembro de un equipo les asigna un número de puntos de un total de puntos a distribuir en el conjunto de afirmaciones. Estas se agrupan en siete secciones y para cada sección se permite distribuir 10 puntos entre las aseveraciones que la forman, el candidato asigna los puntos según el mismo se identifique con cada afirmación. En total se distribuyen 70 puntos en todo el proceso del auto inventario, que se denomina así debido a que cada candidato asigna los puntos en forma personal para formar su propio puntaje de perfiles. Cada aseveración está relacionada con un rol o perfil y los puntos asignados a ella se acumulan para dicho rol, de tal manera que al final se cuenta con un total de puntos acumulados para cada rol por candidato. Esto forma el inventario de cada futuro miembro del equipo. Existe un proceso alterno, en el cuál se permite que la distribución de puntos la lleven a cabo una o varias personas que conozcan al candidato.

Una vez que se tiene el perfil de roles de cada candidato a formar parte del equipo, es necesario formar la matriz de los perfiles de candidatos, una de las dimensiones son los candidatos y la otra los perfiles de Belbin.

Existen algunas reglas generales para la formación de los equipos, por ejemplo: es muy importante que al formar equipos se cubran todos los perfiles de una forma balanceada y que los perfiles como el de Agudo, el Planta no se presenten en forma acentuada en más de una o dos personas como máximo para evitar conflictos de liderazgo. En este estudio proponemos un método de análisis cuantitativo de la matriz para que el análisis sea objetivo, dicho método se ha formado con cuatro criterios que aseguran las diferentes recomendaciones de los diferentes autores presentados aquí.

Propuesta e Hipótesis.

Recordando que los objetivos de este estudio son por un lado poder comparar dos equipos de trabajo y por el otro poder analizar los equipos considerando el impacto que un estímulo como la llegada o la salida de un miembro puede ocasionar. Para lograr lo anterior se han considerado y medido las siguientes variables:

1.-Grupo o Equipo que puede ser 1 para grupo de control y 2 para grupo experimental, es en realidad el identificador del equipo de trabajo.

2.-Programador.- Miembro individual del equipo de trabajo que desarrolla el módulo, esta es una identificación personal y se han utilizado las iniciales de los programadores para identificarlos.

3.-Tipo de programador de acuerdo a la siguiente clasificación:

Tipo 1.- Experto, con más de 10 años de experiencia.

Tipo 2.- Intermedio, dos a cuatro años de experiencia.

Tipo 3.- Principiante, entrenado específicamente para el proyecto.

Este dato se utiliza solo para llevar a cabo una comparación general de habilidades técnicas.

4.-Valor de los Perfiles en puntos, tomado utilizando el auto inventario de habilidades antes mencionado. Esta es la variable más importante del estudio y permite considerar al equipo con una personalidad grupal utilizando una matriz, podemos interpretar desde el punto de vista cuantitativo que la personalidad de un programador la representamos por medio de un vector de perfiles y la de un equipo por medio de una matriz de perfiles.

La hipótesis general de investigación pueden expresarse como sigue:

La matriz de un equipo proporciona información cuantitativa que permite comparar a dos equipos el hecho de conocer la matriz del equipo también permite analizar como impacta en la funcionalidad y eficiencia la llegada o partida de un miembro dado al equipo en cuestión.

El procedimiento de solución de la primera parte de la hipótesis se reduce a comprobar la normalidad de los datos y formular una hipótesis de grupos utilizando los valores de los perfiles de los miembros de los equipos.

Para la segunda parte de la hipótesis se requiere un procedimiento más complejo, es necesario desarrollar un conjunto de pruebas cuantitativas que permitan asegurar que un equipo es funcional y eficiente y posteriormente aplicar esos criterios a las matrices resultantes que representan a cada equipo después de los eventos de llegada y/o partida de un miembro dado.

Modelo Propuesto para la Hipótesis.

Podemos entonces considerar que un equipo está formado por miembros, cada miembro es un renglón en una tabla, por lo tanto el número de renglones en tabla queda definido por el número de miembros en el equipo. Cada perfil es representado por una columna, la suma de una co-

lumna representa el grado de cobertura del perfil que representa dicha columna, tal como se muestra la figura.

Miembro	Descripción
TRABAJADOR DE PLANTA	Creador de soluciones novedosas, creador de nuevos paradigmas.
INVESTIGADOR RECURSOS	Comunicador, desarrollador de contactos, explorador de oportunidades
COORDINADOR	Confiado, delega responsabilidades, comunica y aclara objetivos, cede la toma de decisiones
LIDER	Es motivado por la presión, muestra gran dinamismo, su eficacia crece con los retos
MONITOR	Analizador de alternativas, revisa las estrategias, juzga y selecciona las mejores opciones
TRABAJADOR EN EQUIPO	Amable, político, agradable, excelente conciliador.
IMPLEMENTADOR	Toma conceptos y los convierte en acciones prácticas, es altamente eficiente digno de confianza muestra disciplina y es conservador.
TERMINADOR	Asegurador de la calidad, se entusiasma con revisar el trabajo buscar errores y entregar a tiempo.
ESPECIALISTA	Experto en su campo, muestra gran conocimiento y dedicación a su área, es una fuente de conocimiento.

Tabla 6.2 Matriz "E" de descripción de un equipo mediante Fuente: Elaboración propia utilizando los perfiles de Belbin comentados por el autor(1981).

Podemos definir cuantitativamente los criterios para analizar la estabilidad y eficiencia de un equipo como sigue:

Tamaño: Mantener el tamaño de los equipos entre 3 y 5 personas Jiang, et al. (2007). El número de renglones de la tabla debe ser cuando menos 3 y no ser mayor a 5. Este dato ha sido obtenido después de estudiar más de 4,500 proyectos de software, por lo que podemos asumir que el dato tiene su validez comprobada.

Diversidad: Cubrir todos los perfiles, especialmente aquellos relacionados con habilidades técnicas, disciplinarias y de solución de problemas. Higgs, et al. (2005). Podemos interpretar esto de la siguiente manera:

La matriz en su totalidad representa un conjunto de valores con los cuales se cubren los perfiles, a este conjunto podemos calcularle el promedio general y la desviación estándar, si tomamos el promedio de los valores y le restamos la desviación estándar podemos encontrar el valor que representa el límite inferior con el cual un miembro de este equipo cubre un perfil. Cualquier perfil cubierto por debajo de este valor queda fuera del rango mínimo y por lo tanto puede considerarse como no cubierto.

Balanceo. Los valores de las sumas de las columnas, exceptuando la columna de Líder, que es la 4, deben de ser similares entre sí para que estén balanceadas. El balance de los perfiles lo discuten Ayestarán y Aritzeta (2003) en un experimento con 34 equipos. En este modelo utilizaremos la desviación estándar de las sumas de los puntajes individuales obtenidos y

el promedio de los mismos para calcular los límites inferior y superior (sumándole y restándole al valor del promedio el valor de la desviación estándar respectivamente) de cada perfil exceptuando el 4 (que es el perfil del líder) y revisar que los valores estén dentro de esos límites, si un valor sale de dicho límite hacia arriba (siendo mayor que el límite superior) o hacia abajo (siendo menor que el límite inferior) consideramos que dicho valor no está balanceado respecto al resto. Cabe aclarar que en esta interpretación estamos asumiendo que es tan dañino tener una tendencia hacia arriba como hacia abajo, esto es los perfiles que tienden a salir del rango hacia arriba harán que los equipos se entusiasmen con tareas de dichos perfiles descuidando las otras tareas que no caen en este perfil. Por otra parte tener una tendencia hacia abajo indica que los trabajos y actividades pertinentes a ese perfil no se lleven a cabo por ningún miembro del equipo ya que los miembros no se sienten motivados a realizar ese tipo de trabajos. Chong (2007).

Personalidad. Evitar tener varios miembros que tengan acentuado el perfil de líder agudo para evitar el síndrome de Apolo mencionado por Belbin (1981). No más de un renglón debe tener un valor elevado en el perfil de Líder, columna número 4. Para definir cuantitativamente este criterio volvemos a utilizar la desviación estándar y a calcular el límite superior sumándole al promedio la desviación estándar y consideraremos que cualquier valor arriba de éste límite es un valor elevado que puede entrar en conflicto presentando el síndrome de Apolo con otro miembro del equipo que presente esta personalidad, lo ideal sería tener un solo miembro que muestre un valor elevado en esta columna. El ambiente que se presenta cuando hay más de un líder conduce al equipo a un ambiente de competencia generalmente agotadora y tiene un impacto muy negativo en el desempeño del equipo ya que los miembros al competir olvidan los objetivos del equipo substituyéndolos por el objetivo de ganar la competencia artificialmente generada.

En lo sucesivo, llamaremos a los criterios en su conjunto, *TDBP*, siendo el criterio 1, Tamaño; el dos, Diversidad; el 3, Balanceo; y el 4, Personalidad.

Si llamamos E a la matriz que representa el equipo, n el número de renglones y m al número de columnas podemos establecer los siguientes cálculos para comprobar los criterios arriba mencionados y comprobar que el equipo representado por la matriz E es un equipo estable y eficiente:

Criterio 1.- Tamaño: $3 \leq n \leq 5$ y $m=9$ siendo n el número de renglones que es igual al número de miembros del equipo y m el número de columnas que es el número de perfiles que se van a considerar para definir la personalidad de los equipos.

Criterio 2.- Diversidad: Todos los perfiles deben de estar cubiertos.

Si el promedio general de los valores de la matriz queda definido como:

$$V_{\text{Promedio}} = \left(\sum_{i=1}^n \sum_{j=1}^m E(i, j) \right) / n * m$$

Y la desviación estándar σ de todos los valores de la matriz. Puede especificarse como

σ de $E(i, j)$ para toda i desde $i=1$ hasta $i=n$ y toda j desde $j=1$ hasta $j=m$

Entonces el valor mínimo para que un perfil quede cubierto debe estar por encima de $V_{\text{Promedio}} - \sigma$ y la suma de la columna de dicho perfil debe estar por encima de $(V_{\text{Promedio}} - \sigma) * n$, de no ser así el perfil se puede considerar como no cubierto en el equipo representado por la matriz E .

Criterio 3.- Balanceo: Todos los perfiles deben cubrirse de manera similar.

Si $S(j) = \sum_{i=1}^n E(i, j)$ $S(j)$ representará de perfiles de todo el equipo para la columna j . Si también hacemos

$S_{\text{promedio}} = \sum_{j=1}^9 S(j) / 8$ $S(j)$ será el promedio de los valores de las sumas de todas las columnas

Y si $\sigma(j)$ es la desviación estándar de los valores de $S(j)$ entonces debe cumplirse que

para $j=1,2,3,5,6,7,8,9$ $S(j) \leq S_{\text{promedio}} + \sigma(j)$ y que

para $j=1,2,3,5,6,7,8,9$ $S(j) \geq S_{\text{promedio}} - \sigma(j)$

Esto es que todos los valores de $S(j)$ estén dentro del rango que se obtiene del promedio de valores y su desviación estándar.

Podemos decir que si alguna de estas dos últimas condiciones no se cumple para alguna j se considera que el perfil número j no está balanceado y no se cubre correctamente por el equipo. Si está por arriba del rango el perfil está demasiado acentuado en el equipo y si está por debajo del rango el perfil no está suficientemente cubierto en relación a los demás perfiles.

Criterio 4.- Personalidad: Solo debe haber un miembro con perfil de líder acentuado

Si $X(4)_{\text{promedio}} = \sum_{i=1}^n E(i, 4) / n$ y $\sigma(4)$ la desviación estándar de los valores de la columna 4, estos son los valores $E(i, 4)$ para $i=1$ hasta $i = n$, entonces debe cumplirse que

$E(i, 4) \geq X(4)_{\text{promedio}} + \sigma(4)$ se cumple solo para un valor de i .

Si se cumple para más de un valor de i significa que hay mas de una persona con personalidad de líder y por lo tanto existe un posible conflicto en el liderazgo del equipo representado por la matriz E .

Para mayor detalle de los cálculos estadísticos y del fundamento para la utilización de la desviación estándar y el promedio ver Mendenhall (1997) y Montgomery, Runger (1996).

En el modelo desarrollado consideramos que si los criterios no se cumplen se generan puntos de inestabilidad, por ejemplo tomando el criterio 2 referente a la Diversidad consideramos que si no se cubren dos perfiles ese hecho generaría dos puntos de inestabilidad, tomando el criterio 3 referente al balanceo de los perfiles por ejemplo si tres perfiles no están debidamente cubiertos se generarían tres puntos de inestabilidad. El Criterio 4 solo genera puntos de inestabilidad cuándo el número de miembros con el perfil 4 (Líder) es mayor a 1.

Del modelo se puede obtener el cuadro siguiente:

	A	B	C	D	E	F
1						
2						
3	Resumen de Resultados del Análisis de Sensibilidad del Equipo 1					
4				Total	Miembros	
5		Perfiles		Puntos de	Potenciales	
6		No	Perfiles	Inestabilidad	Síndrome	
7		Cubiertos	No Balanceados	Generados	Apolo	
8	Equipo 1					
9	Completo	0	2	2	1	
10						
11	Miembro que					
12	Abandona					
13	Proyecto					
14	CAS	0	3	3	1	
15	HGV	0	1	1	2	
16	JAC	0	3	3	1	
17	JOM	0	2	2	1	
18	KML	0	2	2	1	
19	MPH	0	3	3	1	
20	TLT	0	2	2	1	
21						
22	Conclusion: los miembros CAS, JAC y MPH producirían la mayor inestabilidad al dejar el proyecto					
23	El miembro HGV produciría el menor efecto de inestabilidad al abandonar el proyecto					
24						

Figura 6.2 Resumen del análisis del Equipo 1 Fuente: Elaboración propia.

Por otra parte si se desea comprobar que dos equipos son equivalentes, el procedimiento a seguir propuesto se basa en una hipótesis de grupos en la cual se procede como se detalla a continuación:

1.- Revisar la normalidad de los datos de los perfiles utilizando la prueba Kolmogorov-Smirnov para asegurar el funcionamiento de la prueba t o de lo contrario utilizar estadística no paramétrica y la prueba de Mann-Whitney. En el presente caso de estudio los datos cumplieron con la prueba de normalidad y se utilizó la prueba t como se demuestra más adelante.

2.- Comparar las habilidades sociales y mentales de los equipos comparando el promedio de los dos grupos para cada perfil (columna de la matriz E) utilizando la prueba t para datos normales. A continuación se plantean los fundamentos de una hipótesis de grupos con fines de comparación por equivalencia:

El interés radica en comparar dos poblaciones existentes que presentan distribución normal, como por ejemplo: producción de dos fabricas. O bien un experimento diseñado para probar algo nuevo (como en este caso):

población 1: situación habitual

población 2: nuevo tratamiento (método, producto).

A continuación se muestra el fundamento de la hipótesis de grupos de éste estudio

Población 1 con media μ_1 desviación estándar σ_1 de la que tomamos la muestra 1 cuyo tamaño es n, con media \bar{X} desviación estándar muestral \hat{S}_1

Población 2 con media μ_2 desviación estándar σ_2 de la que tomamos la muestra 2 cuyo tamaño es m, con media \bar{Y} desviación estándar muestral \hat{S}_2

La varianza en ambos grupos debe ser estudiada como un primer problema, esto es, es necesario determinar si la varianza es igual en ambas muestras para lo cual se plantea la hipótesis nula: $H_0: \sigma_1 = \sigma_2$ con la hipótesis alternativa $H_a: \sigma_1 \neq \sigma_2$. El método estadístico utilizado para esta prueba y que nos permite decidir si la varianza en ambos grupos es o no la misma es el test de la razón de varianzas o test de Levene que resuelve este problema. Este método estadístico requiere que las dos poblaciones sigan una distribución normal y tengan igual varianza ($H_0: \sigma_1 = \sigma_2$) bajo estas condiciones se espera que la razón de varianzas siga una distribución F de Snedecor con parámetros (n-1) y (m-1) cuyo valor estadístico puede ser calculado como se muestra en la siguiente ecuación (Moral, 2007):

$$F = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}{\frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2} = \frac{\hat{S}_1^2}{\hat{S}_2^2}$$

Esta prueba permitirá aceptar la hipótesis nula reconociendo que las dos varianzas de los dos grupos de observaciones son iguales o bien rechazar la hipótesis nula admitiendo que las dos varianzas son diferentes. En términos de su p-valor, podemos decir que si el p-valor es menor a 0,05, rechazaremos la hipótesis nula y supondremos que la variabilidad en ambos grupos es sustancialmente distinta (varianzas no homogéneas). Para la prueba t Student esto quiere decir que el estadístico a calcular variará lige-

ramente en función de las variabilidades muestrales. El estadístico de la t Student a utilizar sería el siguiente:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)\hat{S}_1^2 + (m-1)\hat{S}_2^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Una manera alterna de obtener estos resultados es utilizando el cálculo de intervalos de confianza para el rango de la respuesta media en ambos grupos de observaciones. El intervalo de confianza constituye una medida de la incertidumbre con la que se estima esa diferencia a partir de la muestra, permitiendo valorar tanto la significación estadística como la magnitud de esa diferencia. En el caso de asumir la misma variabilidad en ambos grupos mediante el test de Levene ($p \geq 0,05$), el intervalo de confianza vendrá dado como:

$$(\bar{X} - \bar{Y}) \pm t_{0,975}^{n+m-2} \sqrt{\frac{(n-1)\hat{S}_1^2 + (m-1)\hat{S}_2^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}$$

La expresión $t_{0,975}^{n+m-2}$ representa el valor que la distribución t de Student con $n+m-2$ grados de libertad deja a su derecha el 2,5% de los datos. El intervalo de confianza para la diferencia de medias indica un rango de valores entre los que es posible encontrar el valor real de la diferencia entre las mediciones de ambos grupos. Al considerar el valor cero dentro del intervalo estamos indicando que no se dispone de evidencia para demostrar que la variable que estamos comparando es distinta en ambos grupos de observaciones. Si asumimos distintas varianzas en los grupos ($p < 0.05$), el intervalo de confianza se expresará como:

$$(\bar{X} - \bar{Y}) \pm t_{0,975}^f \sqrt{\frac{\hat{S}_1^2}{n} + \frac{\hat{S}_2^2}{m}}$$

Una serie de ejemplos de aplicaciones se pueden encontrar en (Moral, 2007; Mendenhall, 1997; Montgomery, 1996)

Nótese la importancia de que los valores sigan una distribución Normal para poder aplicar la comprobación anterior.

Por otra parte cuándo no se presenta normalidad en los datos o sea que la prueba Kolmogorov-Smirnov demuestra que los datos no son normales, hemos recurrido a la prueba Mann-Whitney, la cual permite comparar dos muestras independientes (tal como la prueba t lo hace para datos con distribución normal) pero la prueba Mann-Whitney es válida para cualquier distribución de la variable X de la población y no es necesario que las poblaciones tengan una distribución normal. Se dice que es una prueba de libre distribución. Además, la hipótesis no se hace sobre un parámetro (μ) y debido a esto es que a este tipo de pruebas se les conocen como pruebas no paramétricas. En el SPSS se el parámetro significativo para ambas pruebas la t Student para grupos independientes y la de Mann-Whitney se reporta como *Asymp. Sig. (2-tailed)* y es 1.00 para grupos idénticos y 0 .00 para grupos totalmente diferentes.

Recolección de datos

La recolección de datos se llevo a cabo utilizando el auto inventario, las aseveraciones se relacionan con el perfil que se evalúa y cada candidato a miembro del equipo asigna en forma personal e individual los puntos a cada aseveración según se sienta cómodo con el contenido de la misma. El instrumento tiene siete secciones y se cuentan con 10 puntos para distribuir en las aseveraciones de cada sección, dando un total de 70 puntos por aplicación. Los 70 puntos se distribuyen entre los nueve perfiles sin que el participante conozca que perfil evalúa cada aseveración. Al final se cuenta con nueve valores que representan el puntaje por perfil para el participante.

El auto inventario actualmente se aplica por Internet en el sitio oficial www.belbin.com proporcionando como resultado los valores de los perfiles y algunos otros reportes complementarios.

El objetivo de éste estudio no es enfocarse a la toma recolección de datos, sino al análisis cuantitativo una vez que se cuenta con la matriz que define los perfiles para cada miembro del equipo.

Desarrollo y diseño del experimento

Para comprobar la hipótesis de éste estudio es necesario que los equipos lleven a cabo un trabajo equivalente, el diseño de un experimento como éste depende en gran parte del presupuesto disponible y de las habilidades de los miembros de los equipos para llevar a cabo el estudio. En el diseño experimental se ha tratado de cumplir con la rigurosidad científica y desde luego limitando el experimento al presupuesto disponible.

Atendiendo los componentes citados por Cresswell (2003, pp. 162-163) para un experimento los podemos enlistar como sigue:

Participantes.- En el presente contexto los participantes son los miembros de los equipos de trabajo que desarrollaron el sistema en forma

original y miembros del equipo de trabajo que utilizaron la herramienta OSMA para desarrollar de nuevo el sistema equivalente

Materiales.-Herramienta OSMA y el software desarrollado

Procedimientos.-Metodologías de desarrollo de software

Mediciones.-Perfiles y resultados obtenidos en números de programas y líneas de código escritas

Otro factor crítico de un experimento además de los anteriores es el proceso de selección de los grupos, que puede ser:

Aleatorio y Conveniente, en este estudio se utilizo la segunda opción por restricciones de tiempo y costo.

Para que un experimento científico pueda ser considerado como tal es necesario cumplir con los siguientes requisitos básicos (Sierra, 1986, p 324):

Los grupos deben de ser muestras aleatorias estadísticamente válidas

Las variables externas deben estar bajo control o claramente explicadas en las limitaciones del estudio.

Las variables internas deben estar bajo control o claramente explicadas en las limitaciones del estudio.

Condiciones que debe cumplir un experimento (Sierra, 1986, p 324).

En este caso de estudio se utilizaron dos grupos uno de control, al cual no se le administra ningún estímulo y otro experimental el cual utiliza OSMA en su desarrollo de software.

Dado a lo anterior, este experimento puede ser clasificado de metodología Cuasi Experimental ya que no reúne todas las condiciones de un experimento propiamente dicho debido a las razones que se enumeran en la siguiente tabla.

- 1.- La selección de los miembros de los grupos no fue aleatoria
- 2.- La selección del trabajo a realizar no fue aleatoria
- 3.- No se controlaron completamente todas las variables externas, tales como el lenguaje de programación, solo se procuró que fueran equivalentes en cierta medida.
- 4.- El estudio se encuentra desfasado en el tiempo, el primer desarrollo se efectuó varios años antes del segundo.

Sin embargo el hecho de que la metodología es Cuasi Experimental no interviene en el estudio de los equipos que llevaron a cabo el experimento que es el objetivo de la presente investigación.

El método experimental tiene dos características muy importantes, la validez y el error, la primera de ellas tiene que ver con la concordancia externa e interna.

La concordancia interna muestra armonía con las variables internas del proceso, por ejemplo en éste estudio todos los programadores muestran la misma tendencia (aumento de productividad) al responder al estímulo del uso de la herramienta. Por su parte la concordancia externa muestra armonía con las variables externas del proceso, en éste caso otras investigaciones han probado efectos similares al utilizar herramientas de apoyo en la programación, por ejemplo el uso de lenguajes de 4ª. Generación han demostrado aumento en la productividad del software según Jiang et al. (2007).

Existe un tipo de validez llamada validez estadística, la cual depende de demostrar estadísticamente que los resultados del experimento pueden aplicarse a un determinado universo o bien que las muestras son estadísticamente validas.

Por otra parte lo opuesto a la validez son los errores, los errores tienen su origen en la planificación, las variables internas o las variables externas o en varios de éstos factores. Los errores de planificación surgen a medida por las limitaciones del estudio.

A continuación se presentan un resumen los diseños de investigaciones cuantitativas más comunes según (Sierra, 1986) utilizando la notación de Campbell y Stanley (1978) divididos en:

Diseños No-Experimentales tipo Encuestas, **R** denota el grupo y **O_i** la observación número *i*

Diseños Pre-Experimentales, **R** denota el grupo y **O_i** la observación número *i* y **X** el estímulo aplicado.

Diseños Experimentales, **R** denota el grupo y **O_i** la observación número *i* y **X** el estímulo aplicado.

Diseños Cuasi-Experimentales, **R** denota el grupo y **O_i** la observación número *i* y **X** el estímulo aplicado.

Modelo	Notación	Descripción
Seccional	R O	Toma una observación en un solo grupo en un solo momento del tiempo

Longitudinal	$O_1 O_2 O_3 \dots O_n$	Dos o más observaciones realizadas al mismo grupo en tiempos diversos
Comparativo Seccional	$R O_1$ $R O'_1$	Dos muestras aleatorias distintas tomadas una sola vez.
Comparativo Longitudinal	$O_1 O_2 O_3 \dots O_n$ $O'_1 O'_2 O'_3 \dots O'_n$	Dos grupos distintos con sucesivas observaciones

Tabla 6.3 Diseños No-Experimentales tipo Encuestas, R denota el grupo y O_i la observación número i .
Fuente: Elaboración propia con datos de Sierra (1986)

Modelo	Notación	Descripción
Un Grupo, Una Observación	$X O$	Un Grupo, Una Observación.
Un grupo con Pre Test y Post Test	$O_1 X O_2$	Dos observaciones, una antes y otra después del estímulo sobre un mismo grupo.
Ex-post-facto	$X O \dots l$ $O \dots l$	La l indica igualación posterior de los grupos, los puntos suspensivos el transcurrir del tiempo entre la actuación de la variable experimental y la observación.

Tabla 6.4 Diseños Pre-Experimentales, R denota el grupo, O_i la observación número i y X denota el estímulo. Fuente: Elaboración propia con datos de Sierra (1986)

Modelo	Notación	Descripción
Grupo de Control con Pre Test y Post Test	$R O_1 X O_2$ $R O'_1 O'_2$	Al grupo de control no se le aplica estímulo el cual solo se aplica al grupo experimental, las mediciones se toman antes y después del estímulo
Grupo de Control sin Pre Test	$R X O_1$ $R O'_1$	Al grupo de control no se le aplica estímulo el cual solo se aplica al grupo experimental, las mediciones se toman solo después del estímulo

Tabla 6.5 Diseños Experimentales, R denota el grupo, O_i la observación número i y X denota el estímulo. Fuente: Elaboración propia con datos de Sierra (1986)

Modelo	Notación	Descripción
Series Cronológicas	$O_1 O_2 O_3 X O_4 O_5 O_6$	Varias medidas antes y después del estímulo sobre un solo grupo.
Muestras Cronológicas	$O_1 X O_2 O_3 X O_4 O_5 X O_6$	Dos observaciones, una antes y otra después del estímulo sobre un mismo grupo. El estímulo se aplica varias veces sobre el mismo grupo

Tabla 6.6 Diseños Experimentales, R denota el grupo, O_i la observación número i y X denota el estímulo. Fuente: Elaboración propia con datos de Sierra (1986)

A continuación se propone un modelo para este estudio, lo que se presenta es la formalización de la notación del modelo experimental. En el modelo se considera el hecho de que el estudio está desfasado en el tiempo y de que en la segunda instancia se aplicó un estímulo que como ya se mencionó es la utilización de la herramienta OSMA que cambia el paradigma del desarrollo de software.

Utilizando la notación ilustrada por Sierra (1986) y desarrollada por Campbell, et al. (1978) se propone en la siguiente tabla, la notación para expresar el diseño cuasi experimental del presente estudio.

$R \ O_1$ $\dots\dots\dots R X O'_1$
<p>Explicación al modelo Cada renglón es una instancia R es el grupo o equipo de trabajo en este caso hay dos uno para cada instancia O₁ y O'₁ son las observaciones que para este caso son los perfiles expresados en forma matricial X es el estímulo en este caso fue el uso de la herramienta generadora de código OSMA Representa el tiempo transcurrido entre una instancia y la otra</p>
<p>La hipótesis del estudio propone que es posible comprobar que los dos equipos son equivalentes y que pueden ser analizados mediante el estudio de la matriz de perfiles que los representa.</p>

Tabla 6.7 Modelo Cuasi Experimental del presente estudio utilizando la notación de Campbell , et al. (1978) Fuente: Elaboración propia.

Procedimientos y mediciones utilizadas en el estudio

Para llevar a cabo este estudio se llevaron a cabo los siguientes procedimientos:

Los participantes del grupo de control fueron los programadores que desarrollaron el sistema original, su selección fue llevada a cabo en forma de contratación de personal revisando que contaran con las habilidades necesarias para el desarrollo.

La selección del grupo experimental se llevó a cabo entre un grupo de estudiantes y varios programadores voluntarios que recibieron una calificación o en el caso de los programadores un entrenamiento a cambio de su trabajo, este criterio cae en el criterio de conveniencia de Cresswell (2003) para métodos cuasi experimentales.

Procedimiento del experimento:

Grupo de Control

El software ya se encontraba terminado y operando, por lo que solo se tomaron mediciones de los perfiles utilizando el auto inventario disponible en el sitio oficial de Belbin en Internet. En los casos en que no se contó con las personas originales, se pidió a terceras personas que trabajaron estrechamente con ellas que contestaran el instrumento de medición.

Grupo Experimental

Desarrollo del Software

Asignación del trabajo

Explicación del trabajo a realizar

Programación.

Entrega de Resultados.

Formas de Medición:

Mediciones de los perfiles de Belbin utilizando el Auto Inventario de Belbin disponible en Internet.

Estímulo:

Uso de la herramienta Osma

Limitaciones y alcance del estudio.

Debido a que el diseño del estudio es cuasi experimental tiene algunas importantes limitaciones las cuales describimos a anteriormente y aclaramos a continuación.

1. Solo se estudiaron dos equipos. El trabajo que desarrollaron ambos equipos fue complejo y de largo plazo.
2. Algunos de los miembros del equipo de control no estuvieron disponibles para llenar el auto inventario de perfiles, pero personas que los conocieron y convivieron con ellos contestaron el inventario en su lugar, asumiendo su personalidad y forma de trabajo en la época en el que desarrollaron el sistema original.

3. La motivación de cada equipo varía en el grupo de control la motivación fue cumplir con un trabajo remunerado y en el grupo experimental la motivación fue cumplir con un trabajo académico escolar.
4. La tecnología de cómo medir los perfiles queda fuera de los objetivos de este estudio por razones de derechos de autor consignados en www.belbin.com a donde se remite al lector para obtener los servicios de medición respectivos.
5. El alcance de este estudio se limita al estudio de los equipos desde un punto de vista cuantitativo tomando como base los valores obtenidos en los perfiles y utilizados en forma matricial.

Resultados Obtenidos

Se llevó a cabo el experimento en el que participó un equipo de programadores similar al del grupo de control. Las similitudes de los equipos de trabajo y su análisis se presentan en este apartado. Se desarrolló el sistema equivalente en funcionalidad al sistema del grupo de control y se obtuvieron los resultados que se presentan a continuación:

Grupo	Tipo De Programador	Numero de Módulos Desarrollados	Promedio Tiempo De Desarrollo Hrs.
1	1	102	23.69
1	2	162	29
1	3	177	30
	Totales	441	
2	1	102	1.30
2	2	162	1.81
2	3	177	2.03
	Totales	441	

Tabla 6.8 Resultados de la prueba clasificando el desarrollo por tipo de programador Fuente: Elaboración propia

Los equipos de trabajo se muestran en la tabla 8.8 clasificando a los programadores por tipo de programador.

ID	GRUPO	TIPO DE PROGRAMADOR	DESCRIPCIÓN
TALT	Control	2	Intermedio, dos a cuatro años de experiencia.
JAC	Control	3	Principiante, entrenado específicamente para el proyecto
CAS	Control	2	Intermedio, dos a cuatro años de experiencia.
JOM	Control	2	Intermedio, dos a cuatro años de experiencia.
MPH	Control	1	Experto, con más de 10 años de experiencia
HGV	Control	3	Principiante, entrenado específicamente para el proyecto
KML	Control	3	Principiante, entrenado específicamente para el proyecto
APG	Experimental	2	Intermedio, dos a cuatro años de experiencia.
GPC	Experimental	3	Principiante, entrenado específicamente para el proyecto
JALL	Experimental	3	Principiante, entrenado específicamente para el proyecto
MPH	Experimental	1	Experto, con más de 10 años de experiencia
RYC	Experimental	2	Intermedio, dos a cuatro años de experiencia.
SGR	Experimental	3	Principiante, entrenado específicamente para el proyecto
JSM	Experimental	2	Intermedio, dos a cuatro años de experiencia.

Tabla 6.9 Equipos de trabajo participantes

Una vez que se obtuvieron los perfiles de los equipos, se procedió al análisis estadístico descriptivo de los mismos, el cual se muestra en la tabla 6.9

		I D	TRA B PL	INV RE C	CO- ORD	LI- DER	MONI- TOR	TRAB EQUI- PO	IM- PLEM.	TER M.	ESP.	
1	1	TALT	1	3	15	7	8	8	12	4	12	
		2	JAC	11	8	3	11	4	5	7	10	11
		3	CAS	8	5	6	9	6	2	4	10	20
		4	JOM	6	6	6	7	9	7	17	5	7
		5	HGV	2	2	12	17	18	1	3	3	12
		6	MPH	4	10	11	11	3	4	5	11	11
		7	KML	8	4	6	10	6	7	4	10	15
	Total	7	7	7	7	7	7	7	7	7	7	
	Sum Me an		40	38	59	72	54	34	52	53	88	
			5.71	5.43	8.43	10.29	7.71	4.86	7.43	7.5	12.57	
2	1	APG	17	7	7	7	2	4	10	8	8	
		2	GPC	7	7	5	7	5	9	15	10	5
		3	JALL	8	4	3	12	9	6	13	10	5
		4	MPH	4	10	11	11	3	4	5	11	11
		5	RYC	10	11	6	8	8	5	5	7	10
		6	SGR	7	9	8	10	14	4	4	6	8

	7	JSM	4	7	14	12	7	12	7	6	1
Total N	7	7	7	7	7	7	7	7	7	7	7
Sum		57	55	54	67	48	44	59	58	48	
Mean		8.14	7.86	7.71	9.57	6.86	6.29	8.43	8.29	6.86	
Total N	14	14	14	14	14	14	14	14	14	14	14
Sum		97	93	113	139	102	78	111	111	136	
Mean		6.93	6.64	8.07	9.93	7.29	5.57	7.93	7.93	9.71	

Tabla 6.10 Estadísticas descriptivas de la composición de los equipos de trabajo, Gpo 1 es el grupo de control, Gpo 2 grupo experimental. Elaboración propia con datos del análisis con SPSS 15

En el primer paso del análisis de los perfiles de los equipos, se llevó a cabo la prueba de normalidad, obteniendo los datos que se muestran en las tablas 6.10, 6.11 y 6.12 demostrando que los datos tienen una distribución normal tanto para el grupo control como para el grupo experimental.

One-Sample Kolmogorov-Smirnov Test

GRUPO DE CONTROL		TRAB PLTA	INV REC	COORD	LIDER	MONITOR	TRAB EQUIPO	IMPLEM	TERM	ESP
N		7	7	7	7	7	7	7	7	7
Normal Parameters (a,b)	Mean	5.71	5.43	8.43	10.29	7.71	4.86	7.43	7.57	12.57
	Std. Deviation	3.592	2.820	4.276	3.402	4.990	2.673	5.192	3.409	4.036
Most Extreme Differences	Absolute	.166	.134	.286	.274	.255	.217	.251	.333	.271
	Positive	.135	.134	.286	.274	.255	.143	.251	.203	.271
	Negative	-.166	-.112	-.155	-.167	-.172	-.217	-.197	-.333	-.206
Kolmogorov-Smirnov Z		.440	.355	.758	.725	.676	.575	.665	.882	.716
Asymp. Sig. (2-tailed)		.990	1.000	.614	.669	.751	.896	.768	.418	.685

a Test distribution is Normal.

b Calculated from data.

c Equipo = 1

Tabla 6.11 Prueba de normalidad Kolmogorov-Smirnov para el grupo de control. Fuente: Elaboración propia con datos del análisis con SPSS 15

One-Sample Kolmogorov-Smirnov Test

GRUPO EXPERIMENTAL		TRAB PLTA	INV REC	COORD	LIDER	MONITOR	TRAB EQUIPO	IMPLEM	TERM	ESP
N		7	7	7	7	7	7	7	7	7
Normal Parameters (a,b)	Mean	8.14	7.86	7.71	9.57	6.86	6.29	8.43	8.29	6.86
	Std. Deviation	4.451	2.340	3.729	2.225	4.059	3.094	4.315	2.059	3.436
Most Extreme Differences	Absolute	.227	.214	.184	.189	.156	.251	.215	.226	.202
	Positive	.227	.214	.184	.189	.156	.251	.215	.162	.134
	Negative	-.176	-.214	-.103	-.168	-.116	-.230	-.152	-.226	-.202
Kolmogorov-Smirnov Z		.601	.567	.486	.499	.413	.664	.569	.598	.534
Asymp. Sig. (2-tailed)		.863	.905	.972	.965	.996	.770	.902	.867	.938

a Test distribution is Normal.

b Calculated from data.

c Equipo = 2

Tabla 6.12 Prueba de normalidad Kolmogorov-Smirnov para el grupo experimental. Fuente: Elaboración propia con datos del análisis con SPSS 15

Inicialmente los equipos fueron comparados con la prueba t, para comprobar que el experimento había sido realizado en condiciones equivalentes, esto es que los equipos contaban en promedio los mismos tipos de programadores y perfiles, encontrando los resultados que se muestran en la tabla 5.13 y 5.14, donde se muestra que si son equipos equivalentes. Exceptuando por el perfil de especialista que se explica posteriormente en esta sección.

Observando los resultados de la prueba t podemos concluir que los equipos son equivalentes ya que solo el perfil de especialista no tiene significancia aceptable, pero en la tabla 8.9 se puede observar que el valor del perfil en el equipo del grupo de control para el perfil de especialista es de 88 puntos y en el equipo del grupo experimental este mismo perfil muestra solo 48, en otras palabras, se cubrió más ampliamente el perfil de especialista en el grupo de control. Sin embargo para fines del experimento que fue demostrar que la herramienta OSMA debe mejorar la eficiencia del grupo experimental podemos aceptar la equivalencia entre ambos grupos, ya que la diferencia entre ellos no favorece al grupo experimental.

One-Sample Kolmogorov-Smirnov Test

		TRAB PLTA	INV REC	COORD	LIDER	MONITOR	TRAB EQUIPO	IMPLEM	TERM	ESP
AMBOS GRUPOS										
N		14	14	14	14	14	14	14	14	14
Normal Parameters(a,b)	Mean	6.93	6.64	8.07	9.93	7.29	5.57	7.93	7.93	9.71
	Std. Deviation	4.085	2.790	3.872	2.786	4.393	2.875	4.615	2.731	4.665
Most Extreme Differences	Absolute	.182	.122	.204	.157	.205	.150	.237	.276	.169
	Positive	.182	.114	.204	.157	.205	.150	.237	.130	.169
	Negative	-.094	-.122	-.132	-.147	-.114	-.149	-.143	-.276	-.109
Kolmogorov-Smirnov Z		.682	.458	.762	.588	.768	.562	.887	1.033	.633
Asymp. Sig. (2-tailed)		.741	.985	.607	.880	.597	.910	.411	.237	.818

a Test distribution is Normal. b Calculated from data.

Tabla 6.13 Prueba de normalidad Kolmogorov-Smirnov para ambos grupos. Fuente: Elaboración propia con datos del análisis con SPSS 15

Group Statistics

	Equipo	N	Mean	Std. Deviation	Std. Error Mean
TRABAJADOR PLANTA	1	7	5.71	3.592	1.358
	2	7	8.14	4.451	1.682
INVESTIGADOR REC	1	7	5.43	2.820	1.066
	2	7	7.86	2.340	.884
COORDINADOR	1	7	8.43	4.276	1.616
	2	7	7.71	3.729	1.409
LIDER	1	7	10.29	3.402	1.286
	2	7	9.57	2.225	.841
MONITOR	1	7	7.71	4.990	1.886
	2	7	6.86	4.059	1.534
TRABAJADOR EQUIPO	1	7	4.86	2.673	1.010
	2	7	6.29	3.094	1.169
IMPLEMENTADOR	1	7	7.43	5.192	1.962
	2	7	8.43	4.315	1.631
TERMINADOR	1	7	7.57	3.409	1.288
	2	7	8.29	2.059	.778
ESPECIALISTA	1	7	12.57	4.036	1.525
	2	7	6.86	3.436	1.299

Tabla 6.14 Comparación de los dos equipos de trabajo utilizando la prueba t, 1 es el grupo de control y 2 es el grupo experimental. Fuente: Elaboración propia con datos del análisis con SPSS 15

Independent Samples Test

Independent Samples Test	Levene's Test for Equality of Variances		t-test for Equality of Means							
	F	Sig.	t	df	Sig. (2-tailed)	Mean Diff	Std. Error Difference	95% Confidence Interval of the Difference		
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	
TRAB. PLANTA	.016	.903	-	12	.283	-	2.429	2.162	-	2.282
			-	11.488	.284	-	2.429	2.162	-	7.139
INVEST. REC	.247	.628	-	12	.105	-	2.429	1.385	-	.589
			-	11.605	.106	-	2.429	1.385	-	5.446
COORDINADOR	.619	.447	.333	12	.745	.714	2.144	-	3.958	5.387
			.333	11.782	.745	.714	2.144	-	3.968	5.396
LIDER	.197	.665	.465	12	.650	.714	1.536	-	2.633	4.062
			.465	10.341	.652	.714	1.536	-	2.694	4.122
MONITOR	.054	.819	.353	12	.731	.857	2.431	-	4.440	6.155
			.353	11.522	.731	.857	2.431	-	4.465	6.179
TRAB. EQUIPO	.094	.765	-.925	12	.373	-	1.429	1.545	-	1.938
			-.925	11.752	.374	-	1.429	1.545	-	4.795
IMPLEMENTADOR	.105	.752	-.392	12	.702	-	1.000	2.552	-	4.559
			-.392	11.612	.702	-	1.000	2.552	-	6.559
TERMINADOR	8.976	.011	-.475	12	.644	-	1.505	1.505	-	2.565
			-.475	9.863	.645	-	1.505	1.505	-	3.994
ESPECIALISTA	.005	.947	2.852	12	.015	5.714	2.003	-	1.349	10.079
			2.852	11.703	.015	5.714	2.003	-	1.337	10.092

Tabla 6.15 Comparación de los dos equipos de trabajo utilizando la prueba t, 1 es el grupo de control y 2 es el grupo experimental. Fuente: Elaboración propia con datos del análisis con SPSS 15

El Modelo Utilizado

En base a los cuatro criterios TDBP, se ha desarrollado un modelo para el análisis de las matrices de perfiles, el modelo aplica los cálculos y marca las celdas de una hoja de cálculo que representan riesgos para el equipo. Solo es necesario alimentar la identificación (iniciales) de los miembros de los equipos y el puntaje obtenido en cada uno de los perfiles, esto se hace entre las dos franjas como se muestra en la figura a continuación los resultados se obtienen directamente como se ilustra en la siguiente sección.

	A	B	C	D	E	F	G	H	I	J	
1	Equipo 2										
2	Criterio 1	TP	IR	CC	LI	MO	TE	M	TE	ES	
3	Tamaño equipo	7	El tamaño del equipo debe estar entre 3 y 5 miembros							35.00	
4								7.76	3.32	5.00	
5											
6	APG	17	7	7	7	2	4	10	8	8	
7	GPCV	7	7	5	7	5	9	15	10	5	
8	JALL	8	4	3	12	9	6	13	10	5	
9	MPH	4	10	11	11	3	4	5	11	11	
10	RYC	10	11	6	8	8	5	5	7	10	
11	SGR	7	9	8	10	14	4	4	6	8	
12	JSM	4	7	14	12	7	12	7	6	1	
13											
14											
15											
16											
17											

Figura 6.3 Modelo Utilizado Área de alimentación de datos. Fuente: Elaboración propia. El área entre las dos franjas es el área de alimentación de datos, el modelo tiene capacidad para analizar equipos de hasta 10 miembros.

Se puede obtener una copia magnética de éste modelo en www.osmasoftware.com está disponible con licencia en la modalidad de software libre.

Análisis de la sensibilidad de los equipos del experimento aplicación del modelo.

La matriz que define cada equipo fue analizada con un modelo escrito en una hoja de cálculo, dicho modelo aplica los cuatro criterios de sensibilidad de los equipos mencionados en el apartado 8.2.3.1. Mostramos a continuación los resultados del análisis, primero del equipo completo y posteriormente los efectos que se originan al abandonar los distintos miembros de los equipos el proyecto.

A continuación se presenta el modelo aplicado al equipo de control también llamado Equipo 1 :

	A	B	C	D	E	F	G	H	I	J	
1	Equipo 1										
2	Criterio 1	TP	IR	CO	LI	MO	TE	IM	TE	ES	
3	Tamaño equipo	7	El tamaño del equipo debe estar entre 3 y 5 miembros								21.00
4								7.78	4.33	3.00	
5											
6	TALT	1	3	15	7	8	8	12	4	12	
7	JAC	11	8	3	11	4	5	7	10	11	
8	CAS	8	5	6	9	6	2	4	10	20	
9	JOM	6	6	6	7	9	7	17	5	7	
10	HGV	2	2	12	17	18	1	3	3	12	
11	MPH	4	10	11	11	3	4	5	11	11	
12	KML	8	4	6	10	6	7	4	10	15	
13											
14											
15											
16											
17											
18	Criterio 2	Diversidad	Todos los perfiles deben estar cubiertos								
19	Suma	40	38	59	72	54	34	52	53	88	
20	Promedio	6	5	8	10	8	5	7	8	13	
21	DesvStd	3.59	2.82	4.28	3.40	4.99	2.67	5.19	3.41	4.04	
22	LimitSup	9.59	7.82	12.28	13.40	12.99	7.67	12.19	11.41	17.04	
23	LimitInf	2.41	2.18	3.72	6.60	3.01	2.33	1.81	4.59	8.96	
24											
25											
26											
27	Criterio 3	Todos los perfiles debe estar balanceados exceptuando la columna de Lider									
28	Criterio3Sup	28.11	30.11	9.11	0.00	14.11	34.11	16.11	15.11	-19.89	
29	Criterio3Inf	3.61	1.61	22.61	0.00	17.61	-2.39	15.61	16.61	51.61	
30											

Figura 6.4 Modelo Aplicado al Equipo de Control Primera Parte Criterios 1,2 y 3. Fuente: Elaboración propia.

Primeros tres criterios aplicados para el equipo 1, se observa que existen dos perfiles no balanceados. El especialista con tendencias hacia arriba y el de terminador con tendencias a la baja.

33										
34	Criterio 4	Prevención del síndrome de Apolo, solo debe haber un líder natural								
35		Criterio4Sup - Puntaje de lider								
36	TALT	7	6.40							
37	JAC	11	2.40							
38	CAS	9	4.40							
39	JOM	7	6.40							
40	HGV	17	-3.60							
41	MPH	11	2.40							
42	KML	10	3.40							
43										
44										

Figura 6.5 Modelo Aplicado al Equipo de Control Segunda Parte Criterios 4 . Fuente: Elaboración propia.

Criterio 4 de liderazgos para el equipo 1, se observa solo una persona con tendencias de liderazgo, que en la práctica no fue el líder formal, el líder formal del equipo fue MPH, no se presentó síndrome de Apolo.

	A	B	C	D	E	F
1						
2						
3	Resumen de Resultados del Análisis de Sensibilidad del Equipo 1					
4				Total	Miembros	
5		Perfiles		Puntos de	Potenciales	
6		No	Perfiles	Inestabilidad	Síndrome	
7	Equipo 1	Cubiertos	No Balanceados	Generados	Apolo	
8	Completo					
9		0	2	2	1	
10	Miembro que					
11	Abandona					
12	Proyecto					
13	CAS	0	3	3	1	
14	HGV	0	1	1	2	
15	JAC	0	3	3	1	
16	JOM	0	2	2	1	
17	KML	0	2	2	1	
18	MPH	0	3	3	1	
19	TLT	0	2	2	1	
20						
21	Conclusion: los miembros CAS, JAC y MPH producirían la mayor inestabilidad al dejar el proyecto					
22	El miembro HGV produciría el menor efecto de inestabilidad al abandonar el proyecto					
23						

Figura 6.6 Modelo Aplicado al Equipo de Control Resumen de Resultados. Fuente: Elaboración propia.

Análisis de sensibilidad para el equipo 1 en cuanto al abandono del proyecto por parte de un miembro del equipo. Para calcular estos números solo se eliminó el renglón del miembro respectivo del modelo del Equipo 1.

Modelo aplicado al Equipo Experimental:

	A	B	C	D	E	F	G	H	I	J	
1	Equipo 2										
2	Criterio 1	TP	IR	CO	LI	MO	TE	IM	TE	ES	
3	Tamaño equipo	7		El tamaño del equipo debe estar entre 3 y 5 miembros						35.00	
4								7.78	3.32	5.00	
5											
6	APG	17	7	7	7	2	4	10	8	8	
7	GPCV	7	7	5	7	5	9	15	10	5	
8	JALL	8	4	3	12	9	6	13	10	5	
9	MPH	4	10	11	11	3	4	5	11	11	
10	RYC	10	11	6	8	8	5	5	7	10	
11	SGR	7	9	8	10	14	4	4	6	8	
12	JSM	4	7	14	12	7	12	7	6	1	
13											
14											
15											
16											
17											
18	Criterio 2	Diversidad	Todos los perfiles deben estar cubiertos								
19	Suma	57	55	54	67	48	44	59	58	48	
20	Promedio	8	8	8	10	7	6	8	8	7	
21	DesvStd	4.45	2.34	3.73	2.23	4.06	3.09	4.31	2.06	3.44	
22	LimitSup	12.45	10.34	11.73	12.23	11.06	9.09	12.31	10.06	10.44	
23	LimitInf	-3.55	5.66	4.27	7.77	2.94	2.91	3.69	5.94	3.56	
24											
25											
26											
27	Criterio 3	Todos los perfiles debe estar balanceados exceptuando la columna de Lider									
28	Criterio3Sup	1.03	3.03	4.03	0.00	10.03	14.03	-0.97	0.03	10.03	
29	Criterio3Inf	9.28	7.28	6.28	0.00	0.28	-3.72	11.28	10.28	0.28	
30											
31		Datos de los totales		DesvStd=	5.16	Promedio=	52.875				
32				Criterio3Sup	58.03	Criterio3Inf	47.72				
33											

Figura 6.7 Modelo Aplicado al Equipo Experimental Primera Parte Criterios 1,2 y 3. Fuente: Elaboración propia.

En el análisis de los primeros tres criterios aplicados para el equipo experimental, se observa que existen dos perfiles no balanceados el de Implementador con tendencias hacia arriba y el de terminador con tendencias a la baja.

33										
34	Criterio 4	Prevención del síndrome de Apolo, solo debe haber un líder natural								
35		Criterio4Sup - Puntaje de lider								
36	APG	7	5.23							
37	GPCV	7	5.23							
38	JALL	12	0.23							
39	MPH	11	1.23							
40	RYC	8	4.23							
41	SGR	10	2.23							
42	JSM	12	0.23							
43										

Figura 6.8 Modelo Aplicado al Equipo Experimental Segunda Parte Criterios 4 . Fuente: Elaboración propia.

En el análisis del Criterio 4 de liderazgos para el equipo experimental, no se observa ninguna persona con tendencias de liderazgo, el líder formal del equipo fue JALL, no se presentó síndrome de Apolo.

	A	B	C	D	E	F	G
1							
2	Resumen de Resultados del Análisis de Sensibilidad del Equipo 2						
3				Total	Miembros		
4		Perfiles		Puntos de	Potenciales		
5		No	Perfiles	Inestabilidad	Síndrome		
6		Cubiertos	No Balanceados	Generados	Apolo		
7	Equipo 1						
8	Completo						
9		0	2	2	0		
10	Miembro que						
11	Abandona						
12	Proyecto						
13	APG	0	4	4	0		
14	GPCV	0	2	2	0		
15	JALL	0	4	4	1		
16	JSM	0	2	2	1		
17	MPH	0	4	4	2		
18	RYC	0	4	4	0		
19	SGR	0	3	3	0		
20							
21	Conclusion: los miembros APG, JALL, MPH y RYC producirían la mayor inestabilidad al dejar el proyecto						
22	Los miembros GPCV y JSM producirían el menor efecto de inestabilidad al abandonar el proyecto						
23	Al partir MPH generaría un posible conflicto del síndrome de Apolo						
24							

Figura 6.9 Modelo Aplicado al Equipo Experimental
Resumen de Resultados. Fuente: Elaboración propia.

Análisis de sensibilidad para el equipo experimental en cuanto al abandono del proyecto por parte de un miembro del equipo. Para calcular estos números solo se eliminó el renglón del miembro respectivo del modelo del Equipo Experimental.

Las formulas de la hoja de cálculo para los diferentes criterios se muestran en las siguientes figuras:

	A	B
1	Equipo 1	
2	Criterio 1	TP
3	Tamaño equipo	=10- CONTAR.BLANCO(A6:A16)
4		
5		
6	TALT	1
7	JAC	11
8	CAS	8
9	JOM	6
10	HGV	2
11	MPH	4
12	KML	8
13		
14		
15		
16		
17		
18	Criterio 2	Diversidad
19	Suma	=SUMA(B6:B15)
20	Promedio	=REDONDEAR(B19/\$B\$3,0)
21	DesvStd	=DESVEST(DESREF(B6:B15,0,0,\$B\$3,1))
22	LimitSup	=B20+B21
23	LimitInf	=B20-B21
24		

Figura 6.10 Formulas del Criterio 1 tamaño del Equipo. Fuente: Elaboración propia.

Para la revisión del Criterio 1 Tamaño del Equipo, se puede observar en la celda B3 como se cuenta y almacena el número de miembros del equipo

	H	I	J
1			
2	IM	TE	ES
3			=B\$3*\$J\$4
4	=PROMEDIO(\$B\$6:\$J\$15)	= DESVEST(\$B\$6:\$J\$15)	=REDONDEAR(PROMEDIO(\$B\$6:\$J\$15) - DESVEST(\$B\$6:\$J\$15),0)
5			
6	12	4	12
7	7	10	11
8	4	10	20
9	17	5	7
10	3	3	12
11	5	11	11
12	4	10	15
13			
14			
15			
16			

Figura 6.11 Formulas para cálculos del Criterio 2 Diversidad. Fuente: Elaboración propia.

Para analizar el Criterio2 se llevan a cabo los cálculos generales: en la celda J4 se calcula el límite inferior de puntos por celda y en la J3 el limite general inferior de la suma de cada columna.

	A	B	C	D
13				
14				
15				
16				
17				
18	Criterio 2	Diversidad		Todos los perfiles deben estar cubiertos
19	Suma	=SUMA(B6:B15)	=SUMA(C6:C15)	=SUMA(D6:D15)
20	Promedio	=REDONDEAR(B19/\$B\$3,0)	=REDONDEAR(C19/\$B\$3,0)	=REDONDEAR(D19/\$B\$3,0)
21	DesvStd	=DESVEST(DESREF(B6:B15,0,0,\$B\$3,1))	=DESVEST(DESREF(C6:C15,0,0,\$B\$3,1))	=DESVEST(DESREF(D6:D15,0,0,\$B\$3,1))
22	LimitSup	=B20+B21	=C20+C21	=D20+D21
23	LimitInf	=B20-B21	=C20-C21	=D20-D21
24				
25				
26				
27	Criterio 3			Todos los perfiles debe estar balanceados
28	Criterio3Sup	=E\$32 - B19	=E\$32 - C19	=E\$32 - D19
29	Criterio3Inf	=B19 - \$G\$32	=C19 - \$G\$32	=D19 - \$G\$32
30				
31		Datos de los totales		DesvStd=
32				Criterio3Sup

Figura 6.12 Formulas para el Criterio 2 Diversidad. Fuente: Elaboración propia.

Para revisar el Criterio 2 Diversidad en las celdas del renglón 19 se suman los puntos de cada perfil, en el renglón 20 se obtiene el promedio por perfil y se comparan por medio de un formato condicional con los valores J4 y J3 que se explican en la siguiente figura.

	E	F	G
13			
14			
15			
16			
17			
18			
19	=SUMA(E6:E15)	=SUMA(F6:F15)	=SUMA(G6:G15)
20	=REDONDEAR(E19/\$B\$3,0)	=REDONDEAR(F19/\$B	=REDONDEAR(G19/\$B\$3,0)
21	=DESVEST(DESREF(E6:E15,0,0,\$B\$3,1))	=DESVEST(DESREF(F	=DESVEST(DESREF(G6:G15,0,0,\$B\$3,1))
22	=E20+E21	=F20+F21	=G20+G21
23	=E20-E21	=F20-F21	=G20-G21
24			
25			
26			
27			
28	0	=\$E\$32 - F19	=\$E\$32 - G19
29	0	=F19 - \$G32	=G19 - \$G32
30			
31	=DESVESTP(B19,C19,D19,F19,G19,H19,I19,J19)	Promedio=	=SUMA(B19,C19,D19,F19,G19,H19,I19,J19)/8
32	=\$G31 + \$E31	Criterio3Inf	=\$G31 - \$E31
33			
34			

Figura 6.13 Formulas para el Criterio 3 Balanceo.
Fuente: Elaboración propia.

Para revisar el Criterio 3 de balanceo se hacen los cálculos para las sumas por perfil: la celda E31 calcula la desviación estándar de las sumas, la celda G31 calcula el promedio de las sumas y las E32 y G32 los límites inferiores y superiores para aplicar el criterio 3 en las líneas 28 y 29 mediante un formato condicional.

	A	B	C
34	Criterio 4	Prevención del síndrome de Apolo	
35			Criterio4Sup - Puntaje de lider
36	=SI(IGUAL(ESPACIOS(A6),""),"",A6)	=SI(E6=0,"",E6)	=SI(E6>0,\$B\$50-B36,"")
37	=SI(IGUAL(ESPACIOS(A7),""),"",A7)	=SI(F7=0,"",F7)	=SI(F7>0,\$B\$50-B37,"")
38	=SI(IGUAL(ESPACIOS(A8),""),"",A8)	=SI(E8=0,"",E8)	=SI(E8>0,\$B\$50-B38,"")
39	=SI(IGUAL(ESPACIOS(A9),""),"",A9)	=SI(E9=0,"",E9)	=SI(E9>0,\$B\$50-B39,"")
40	=SI(IGUAL(ESPACIOS(A10),""),"",A10)	=SI(E10=0,"",E10)	=SI(E10>0,\$B\$50-B40,"")
41	=SI(IGUAL(ESPACIOS(A11),""),"",A11)	=SI(E11=0,"",E11)	=SI(E11>0,\$B\$50-B41,"")
42	=SI(IGUAL(ESPACIOS(A12),""),"",A12)	=SI(E12=0,"",E12)	=SI(E12>0,\$B\$50-B42,"")
43	=SI(IGUAL(ESPACIOS(A13),""),"",A13)	=SI(E13=0,"",E13)	=SI(E13>0,\$B\$50-B43,"")
44	=SI(IGUAL(ESPACIOS(A14),""),"",A14)	=SI(E14=0,"",E14)	=SI(E14>0,\$B\$50-B44,"")
45	=SI(IGUAL(ESPACIOS(A15),""),"",A15)	=SI(E15=0,"",E15)	=SI(E15>0,\$B\$50-B45,"")
46	Perfil de Lider		
47	Suma	=E19	
48	Promedio	=E20	
49	DesvStd	=E21	
50	LimitSup	=\$B48 + \$B49	
51			

Figura 6.14 Formulas para el Criterio 4 Personalidad de Líder. Fuente: Elaboración propia.

Para aplicar el criterio 4 de personalidad de líder las celdas de los renglones 36 al 45 utilizan los valores promedio, desviación estándar y rango superior de la columna 4 para determinar por medio de un formato condicional cuáles miembros tienen un perfil de líder superior al rango de los datos.

Comentarios sobre los resultados.

Los dos equipos muestran ser similares incluso en los riesgos que presentan cuándo sus miembros abandonan hipotéticamente el proyecto. El modelo permite comprobar cuantitativamente tales similitudes.

Cabe señalar que los resultados coinciden con las observaciones directas de administradores experimentados en el manejo de equipos y que estuvieron cerca de estos proyectos. En este experimento específico no hubo que lamentar fracasos o sobregiros de los proyectos, sin embargo los riesgos en todo trabajo en equipo son intrínsecos al hecho de formar un sistema que se comporta no como la suma de sus miembros sino con una nueva y muy propia personalidad.

Conclusiones y líneas de investigación futura.

Los resultados obtenidos muestran y comprueban la hipótesis de que los equipos son equivalentes, adicionalmente lo que se encontró en los análisis de sensibilidad coincide con las observaciones reales que se ocasionaron al partir algunos miembros del equipo original.

La facilidad de modelar un equipo proporciona al administrador la posibilidad de negociar intercambio de miembros sin poner en riesgo los proyectos, a su vez permite que conozca las debilidades de los equipos y pueda llevar a cabo acciones correctivas antes de que aparezcan los primeros signos de problemas en los proyectos.

La implementación cuantitativa de los cuatro criterios planteados como base tiene su fundamento en la estadística básica que aplicada en el contexto de este estudio muestra su gran poder predictivo el cual nos permite llegar al modelo y de ahí a la simulación obteniendo de esta forma la posibilidad de montar escenarios que en la vida real son costos o imposibles de ensayar, como por ejemplo simular la partida de un miembro de un equipo o la llegada de otro.

Los cuatro criterios aquí planteados resumen en forma cuantitativa el conocimiento con el que se cuenta actualmente respecto al comportamiento de los equipos y proporcionan una directriz para analizar los grupos de trabajo,

Hay que recordar el valor que para una organización tienen los equipos estables y exitosos que no solo logran terminar proyectos con calidad y oportunidad, sino que incluso trascienden los proyectos y en ocasiones no solo trascienden los proyectos sino las organizaciones mismas, la literatura documenta una buena cantidad de casos de este tipo de grupos que son generadores de éxito y riqueza en forma natural Davis, Millburn, Murphy y Woodhouse(1992).

Las líneas de investigación futura apuntan hacia el estudio de un mayor número de equipos de trabajo de software que lleven a cabo proyectos complejos y de larga duración para validar los criterios aquí propuestos y el

modelo obtenido. Establecer indicadores y relaciones cuantitativas entre los perfiles en el equipo y los problemas específicos que se presentan en los proyectos de software.

Problemas específicos tales como:

1. Comunicación con los clientes del proyecto.
2. Calidad e interpretación de los requerimientos del software.
3. Productividad en el desarrollo de software

De tal manera que el modelo se enriquezca al grado de ser capaz de predecir los problemas más frecuentes que un equipo puede mostrar aún antes de entrar en funciones.

El considerar el equipo de trabajo no como un conjunto de individuos, sino como un ente que puede ser comparado en forma integral con otro ente similar, esa es la intención de éste estudio. El grupo de trabajo toma personalidad propia y se puede representar cuantitativamente como si fuera un individuo, considerando y representando los matices de su personalidad grupal.

Referencias

Ayestarán, S. y Aritzeta, A. (2003), Aplicabilidad de la teoría de los roles de equipo de Belbin: un estudio longitudinal comparativo con equipos de trabajo. *Revista de psicología general y aplicada: Revista de la Federación Española de Asociaciones de Psicología*, 56(1), 2003, pp 61-75

Belbin Associates (2007), Belbin Team Roles. Sitio oficial Recuperado el 13 de Noviembre de 2007 de <http://www.belbin.com>.

Belbin, R. M. (1981). *Management Teams: Why They Succeed or Fail*. Butterworth-Heinemann: Oxford.

Belbin, R. M. (1993) *Team Roles at Work*; Butterworth-Heinemann; London, UK: Oxford University Press Oxford.

Campbell, D.T. & Stanley, J.C.(1966), *Experimental and Quasi Experimental Social Research Design*; N. York: Rand McNally Co. Academia Press

Chong, E. (2007). Role balance and team development: A study of team role characteristics underlying high and low performing teams. *Journal of Behavioral and Applied Management*, May, 2007 pp. 202-207.

Cresswell, J. W. (2003). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (Second ed.). Thousand Oaks, CA: Sage Publications.

Davis, J., Millburn, P., Murphy, T. & Woodhouse, M. (1992), *Successful Team Building: How to Create Teams that Really Work*; London :Kogan Page

Higgs, M., Plewnia, U. & Ploch J. (2005) Influence of team composition and task complexity on team performance. *Team Performance Management*; 11(7/8), 227-250

Jiang, Z. & Comstock, C. (2007) The Factors Significant to Software Development Productivity. *Proceedings of World Academy of Science, Engineering and Technology* 2, 160-164

Jiang, Z., Naudé, P., and Comstock, C. (May, 2007) The Variation of Software Development Productivity 1995-2005. *Proceeding of World Academy of Science, Engineering and Technology*,21, 355-359

publications.html.en

Mendenhall, W. (1997). Probabilidad y estadística para ingeniería y ciencias. New York: Ed. Prentice-Hall.

Montgomery, D. C. , Runger, G.C. (1996). 1.*Probabilidad y Estadística aplicadas a la ingeniería*. New York: Ed. McGraw-Hill.

Moral, I. (2006). Comparación de Medias, Metodos Estadísticos para Enfermería Nefrológica, Consultado el 31 de Jul. 2009 de <http://www.seden.org/>

Parker, G. M. (1990), Team Players and Teamwork: The Competitive Business Strategy; Oxford , U.K.; Jossey-Bass

Prieto, M(2008). Generación Automática de Software para Aplicaciones Empresariales: Un nuevo enfoque para el desarrollo de software, *Innovación, Investigación Aplicada: Una visión empresarial*.1,(1),1, 33-37

Prieto, M. (2008). Sitio Oficial de OSMA, Consultado el 10 Nov. 2008 <http://www.osmasoftware.com>

Rajendran, M. (2005). Analysis of team effectiveness in software development teams working on hardware and software environments using Belbin self -perception inventory. *The Journal of Management Development*, 24, 7/8, 738-753.

Sierra Bravo, R. (1986), *Tesis Doctorales y Trabajos de Investigación Científica*; Madrid, España: Thompson Editores

Índice del capítulo

<i>análisis estadístico descriptivo</i> ,	286, 287, 288, 289, 290, 291,
281	294, 295, 296
<i>Balanceo</i> , 267, 269, 293	<i>grupo de control</i> , 259, 265, 276,
<i>Diversidad</i> , 263, 267, 268, 270,	278, 279, 280, 282, 285, 286
292	<i>grupo experimental</i> , 258, 265,
<i>equipo</i> , 257, 258, 259, 260, 261,	276, 278, 279, 282, 283, 285,
262, 263, 264, 265, 266, 267,	286
268, 269, 273, 278, 279, 280,	<i>hipótesis</i> , 259, 262, 263, 266,
	270, 273, 278, 294

metodología cuantitativa, 257,
260

Modelo, 266, 275, 276, 277, 278,
286, 287, 288, 289, 290

OSMA, 257, 258, 273, 274, 277,
278, 286, 297

perfil, 259, 261, 263, 264, 265,
266, 267, 268, 269, 270, 273,
286, 292, 293

Personalidad, 263, 268, 269,
293

Prueba de normalidad, 282,
283, 284

síndrome de Apolo, 261, 263,
268, 288, 290

Tamaño, 262, 263, 267, 268,
291

teoría de roles, 261

Capítulo 7:

Administración del conocimiento y capital intelectual

Contenido

Introducción	300
Desarrollo del caso	301
Generalidades del caso	301
<i>Planteamiento del problema</i>	302
<i>Objetivos específicos y generales</i>	304
<i>Justificación de la investigación</i>	304
Antecedentes e investigación bibliográfica	305
Solución propuesta e hipótesis	320
Metodología y diseño de la investigación	328
Análisis y Discusión de los Resultados	335
Conclusiones y líneas de investigación futura.....	369
Bibliografía	371
Índice del capítulo	374

Introducción

Desde hace poco más de una década, se empezó a generar toda una corriente de pensamiento hacia el valor intangible de las empresas. ¿Qué es?, ¿Cómo identificarlo?, ¿Cómo valorarlo? y sobre todo ¿Qué hacer para que éste siga creciendo?, Para que continúe siendo el motor principal en la generación de la riqueza, dentro de lo que los expertos en la materia han llamado *economía del conocimiento*.

Estamos viviendo en la economía de los intangibles. La *economía del conocimiento* es la nueva realidad. Es una economía basada en la intelectualidad y la técnica, las reglas económicas han sido desafiadas y cambiadas. Las nuevas realidades comerciales, requieren perspectivas e ideas radicalmente nuevas de una economía radicalmente nueva (Edvinsson, 2002). Los factores intangibles como la investigación y desarrollo, secretos comerciales, marcas y capital organizacional están siendo la clave de la competitividad (OECD, 2005). El e-commerce es a la revolución informática lo que el ferrocarril fue a la revolución industrial: un desarrollo totalmente nuevo, sin ningún precedente y del todo inesperado (Drucker, 2002).

Dentro de la literatura que describe el conocimiento organizacional, existen cuatro corrientes, que de alguna manera lo incluyen: Una es la teoría evolucionista, otra la visión basada en los recursos y capacidades que tiene la organización; las últimas dos son las que nos ocupan en éste estudio y se refieren al capital intelectual y la administración del conocimiento. Dentro de las primeras tres, el stock de conocimiento es considerado como un activo intangible dentro de las organizaciones y la última de las cuatro se enfoca en la interacción del conocimiento embebido dentro de la organización en sus tres niveles: individual, de grupo y organizacional. Es decir, el conocimiento que fluye del individuo al grupo y a la organización; así como del grupo a la organización. Los flujos que completan el ciclo para mantener los equilibrios son los que parten de la organización al grupo y al individuo; así como del grupo al individuo.

Consideramos que un factor esencial para que el aprendizaje se presente por consecuencia de los flujos, es el compromiso organizacional de los individuos. Es por esto, que en esta investigación nos enfocamos en el impacto que tiene el compromiso organizacional sobre la administración del conocimiento y el capital intelectual.

La presente investigación está integrada por seis puntos. En el primer punto presentamos el planteamiento del problema, el objetivo general y objetivos específicos de la investigación, las preguntas de la investigación y la justificación.

En el segundo punto, revisamos las diferentes teorías que le dan sustento a la investigación. Por un lado, analizamos lo relativo a los stocks de conocimiento en los diferentes niveles de la organización y sus flujos de entrada y salida que permiten un equilibrio dentro de la organización. Por el otro lado, analizamos la importancia del compromiso organizacional en la administración del conocimiento y el capital intelectual. En particular se destaca el trabajo de Allen & Meyer, cuyo modelo presenta tres componen-

tes: Afectivo, de Continuidad y Normativo, asimismo, se contrastan los diferentes conceptos del Compromiso organizacional de otros investigadores como Kelman, Mowday, Reichers, Becker y algunos más.

En el tercer punto presentamos los siete grupos de hipótesis de la investigación, surgidas de la revisión de la literatura. Esto es, el impacto que tiene el modelo de Compromiso organizacional en el stock de conocimiento individual (IK), stock de conocimiento de grupo (GK), stock de conocimiento organizacional (OK), el flujo de aprendizaje hacia adelante (FF), el flujo de aprendizaje hacia atrás (FB), el capital intelectual (IC) y la administración del conocimiento (KM); así como los modelos surgidos en este caso, de la relación del Compromiso organizacional con el capital intelectual (IC) y la administración del conocimiento (KM).

En el punto cuatro mostramos la metodología que fue utilizada para comprobar las hipótesis planteadas en el punto anterior, el diseño de la muestra, los criterios utilizados para su selección, la técnica de investigación y el instrumento de medición.

En el punto cinco, llevamos a cabo el análisis y discusión de los resultados obtenidos de la presente investigación, para el conjunto muestral. Analizamos la confiabilidad, medias, varianza y correlación de las variables. Analizamos los modelos de regresión lineal múltiple con el detalle de la prueba F, coeficiente de determinación (R^2) así como los valores T y las respectivas significancias. Por último, en el punto final que es el seis, presentamos las conclusiones y recomendaciones derivadas de los resultados del estudio, así como, las líneas de investigación futuras.

Desarrollo del caso

En esta sección presentamos los seis puntos que integran, el caso de estudio.

Generalidades del caso

En el presente punto presentamos el planteamiento del problema y su declaración formal. Detallamos las razones por las cuales se considera estudiar la relación del compromiso organizacional con la administración del conocimiento y el capital intelectual, así como el objetivo general de la investigación y los objetivos específicos.

Adicionalmente, mostramos la justificación de la investigación en términos de conveniencia, relevancia social, implicaciones prácticas, valor teórico y utilidad metodológica.

Planteamiento del problema

El valor de las empresas en el mundo actual dominado por la economía del conocimiento, se explica no sólo por el valor de sus activos tangibles sino también por sus *activos intangibles*. De acuerdo con el índice mundial Morgan Stanley, el promedio del valor de las empresas en las bolsas de valores mundiales es el doble de su valor en libros. En los Estados Unidos el valor de mercado de una empresa es regularmente de dos a nueve veces su valor en libros y cuando las empresas son de uso intensivo de conocimiento, supera diez veces su valor.

Peter Drucker (2002: 52), nos dice que necesitamos información exterior, y tendremos que aprender a conseguirla. Pero es complicado debido a que casi todas las empresas tienen dos sistemas de información: uno está organizado en torno al flujo de datos; el otro, muchísimo más antiguo, en torno al sistema contable. Este último, un sistema de información que tiene quinientos años de antigüedad, está en mal estado. Los cambios que veremos en las tecnologías de información dentro de veinte años, serán nada, en comparación con los cambios que veremos en la contabilidad.

En 1991 el Instituto Norteamericano de Contadores Públicos Certificados organizó un comité especial para analizar el porqué los sistemas contables tradicionales no reflejaban el valor de mercado de las empresas. Tres años después, el comité rindió su informe y encontró que hay áreas importantes que merecen mejorarse:

1) Con el suministro de información, podría mejorarse todo lo relacionado con planes corporativos, oportunidades, riesgos e incertidumbre; 2) Un mejor ajuste de los sistemas externos de información con los sistemas internos de control administrativo e información; y 3) Una nueva discusión de los factores de rendimiento no financiero que crean valor a más largo plazo. Estas recomendaciones descansan en el supuesto de que, los datos de rendimiento no financiero son de valor pertinente y pueden usarse eficazmente por los inversionistas como indicadores principales de futuro rendimiento financiero (Edvinsson & Malone, 1997).

A medida que la economía del conocimiento establece nuevas reglas de negocio, el cambio se va convirtiendo en algo paradójico. Las empresas ofrecen sus productos y/o servicios a costos cada vez más bajos y las compañías con mayor crecimiento son las que poseen activos intangibles.

El valor de los activos intangibles está directamente relacionado con la formación del capital intelectual; es por esto que muchas empresas han llevado a cabo proyectos para la identificación, medición y evaluación, así como su explotación. Algunas de estas son: Dow Chemical Company, que ha desarrollado un modelo de administración del activo intelectual, así como Arthur Andersen y Hughes Space & Communications Company que están gestionando una base de datos de conocimiento que contiene todas las experiencias pasadas en proyectos. En la misma línea Chevron creó un equipo para identificar sus mejores prácticas con el fin de descubrir depósitos de

conocimiento escondidos en toda la compañía; esto dio como resultado, un mapa de recursos de mejores prácticas.

De acuerdo con Tissen, Andriessen & Lekanne (2000: 10), la formación del capital intelectual en la empresa depende de dos importantes ingredientes: el nivel de servicio que ofrece y su intensidad de conocimiento, además del grado en que la empresa utiliza el conocimiento para crear productos o servicios, es decir, de la administración del conocimiento.

Investigaciones recientes indican que la administración del conocimiento depende de la cultura, estrategia, liderazgo y medio ambiente. Adicionalmente otros autores tales como Ulrich (1998), Edvinsson (2002, p. 115), Tissen, Andriessen & Lekanne (2000, p. 153) y Marquardt (1996, p. 72), destacan que también depende del Compromiso organizacional.

El compromiso organizacional, también conocido como el compromiso que tienen los empleados con la organización, ha sido objeto de estudios por diferentes académicos como Meyer & Allen (1997), Rashid, Sambasivan, Johari (2003), Mowday (1998), Reichers (1985), Swailes (2004) y Liou & Nihan (1994) en relación con la satisfacción en el trabajo, rotación que tienen las empresas del empleado, la atención que tiene el empleado hacia su trabajo, el desempeño del empleado dentro de la organización y el éxito de la empresa.

DECLARACIÓN DEL PROBLEMA

De las investigaciones previas que se han revisado de la literatura a nivel mundial, hemos encontrado que el compromiso organizacional ha sido estudiado en relación a la satisfacción en el trabajo, el ausentismo, la rotación en la empresa, la cultura de la organización y la efectividad organizacional. Sin embargo, en las estructuras actuales de las empresas en los que los intangibles más importantes son el capital intelectual y la administración del conocimiento, el compromiso organizacional puede significar un impacto que afecte positivamente a éstos elementos, impacto que desde nuestro punto de vista no ha sido suficientemente estudiado. Debido al beneficio económico que tiene la generación del conocimiento como ventaja competitiva en la economía del conocimiento, consideramos que es importante para las empresas evaluar cómo impacta el compromiso organizacional en la administración del conocimiento y el capital intelectual.

Objetivos específicos y generales

Los objetivos de la investigación que se derivan del planteamiento del problema son presentados a continuación.

OBJETIVOS ESPECIFICOS:

1. Identificar y evaluar cómo impacta el compromiso organizacional en el stock de conocimiento individual.
2. Identificar y evaluar cómo impacta el compromiso organizacional en el stock de conocimiento de grupo.
3. Identificar y evaluar cómo impacta el compromiso organizacional en el stock de conocimiento organizacional.
4. Identificar y evaluar cómo impacta el compromiso organizacional en los flujos de aprendizaje hacia adelante.
5. Identificar y evaluar cómo impacta el compromiso organizacional en los flujos de aprendizaje hacia atrás.
6. Identificar y evaluar cómo impacta el compromiso organizacional en el capital intelectual.
7. Identificar y evaluar cómo impacta el compromiso organizacional en la administración del conocimiento.
8. Apoyar a la alta dirección con indicadores claros, sobre los factores críticos que afectan al capital intelectual y la administración del conocimiento de la organización, en el presente y futuro.

OBJETIVO GENERAL:

La presente investigación tiene como propósito principal contribuir teórica y empíricamente al análisis del impacto que ejerce el compromiso organizacional en la administración del conocimiento y la formación de capital intelectual de las empresas a través de un modelo que permite mejorar la administración de recursos humanos en las empresas a fin de lograr una mejor generación de la riqueza.

Justificación de la investigación

JUSTIFICACIONES TEÓRICAS Y PRÁCTICAS

Cuando hablamos de justificar un trabajo de investigación debemos referirnos a la justificación por conveniencia, justificación por sus relevancia social e implicaciones prácticas; justificación por su valor teórico y por último justificación por su utilidad metodológica (Hernandez, R., Fernández y C., Baptista, P., (2003).

JUSTIFICACIÓN DE CONVENIENCIA

Esta investigación servirá para enriquecer el conocimiento de los elementos del compromiso organizacional que impactan a la administración del conocimiento y el capital intelectual.

JUSTIFICACIÓN POR SU RELEVANCIA SOCIAL E IMPLICACIONES PRÁCTICAS:

El beneficio que obtendremos de la presente investigación será de gran utilidad para la comunidad de investigadores, porque aportará conocimiento teórico-práctico de las interrelaciones de las variables independientes (compromiso organizacional) con las variables dependientes (administración del conocimiento y el capital intelectual).

JUSTIFICACIÓN POR SU VALOR TEÓRICO:

El aporte teórico de la investigación, nos permitirá hacer recomendaciones sobre cuáles serían los programas y/o proyectos que el área de recursos humanos deberá realizar para desarrollar el compromiso en las organizaciones.

JUSTIFICACIÓN METODOLÓGICA:

El modelo presentado apoyará a la alta dirección de las organizaciones a determinar las fortalezas y debilidades en los programas de administración de los recursos humanos.

Antecedentes e investigación bibliográfica

El propósito de este punto es llevar a cabo una revisión de la literatura de investigación sobre aprendizaje organizacional, capital intelectual, administración del conocimiento y compromiso organizacional. El marco teórico implica varios requerimientos. Primero se identifica el fenómeno de interés; segundo las premisas claves o las suposiciones sobresalientes necesarias para ser declaradas; tercero se describe la relación entre los elementos relevantes (Crossan, Lane & White, 1999).

El proceso de aprendizaje organizacional

El término *aprendizaje organizacional* se ha venido utilizando desde la década de los 60's, existiendo diversas definiciones sobre el mismo. Simon (1969) define el término aprendizaje organizacional como el crecimiento de ideas y la reestructuración exitosa de los distintos problemas de la organización, reflejados por los individuos en los elementos estructurales y los resultados de la organización misma (Simon 1969, en Fiol & Lyles, 1985).

Garvin (1993) nos dice que una organización que aprende es una organización que tiene las habilidades para crear, adquirir y transferir conocimiento para modificar su comportamiento y que refleje ese nuevo conocimiento en nuevas perspectivas e ideas; nos dice que las organizaciones que aprenden son expertas en cinco actividades claves: 1) Solución de problemas sistemáticamente, 2) Experimentación con nuevos enfoques, 3) Aprendizaje de su propia experiencia y su historia pasada, 4) Aprendizaje de la experiencia y mejores prácticas de otros y 5) Transferencia del conocimiento a toda la organización rápida y eficientemente.

El aprendizaje se ve afectado por factores contextuales que son: la cultura corporativa que conduce al aprendizaje, la estrategia que permite la flexibilidad, la estructura organizacional que permite la innovación y nuevas ideas, y por último el medio ambiente donde se encuentra inmersa la organización (Fiol & Lyles, 1985).

Fiol & Lyles (1985) señalan dos niveles de conocimiento en relación al aprendizaje. Por un lado, el bajo nivel de conocimiento que lo definen como el aprendizaje enfocado, que puede ser mera repetición o un comportamiento del pasado, por lo general de corto plazo, superficial, temporal, y rutinario. Captura sólo un cierto elemento, ajustándolo a lo que la organización hace: ciclo sencillo y nivel de rutina. Por otro lado, el alto nivel de conocimiento que lo definen como el desarrollo de reglas complejas asociadas a nuevas acciones; desarrollo de un entendimiento de causalidad, aprendizaje que afecta la organización entera, aprendizaje de ciclo completo, e interiorización de normas centrales, marcos de referencia y cambios, véase tabla 7.1.

	Bajo nivel de conocimiento	Alto nivel de conocimiento
Características	Ocurre a través de la repetición, rutinario, control sobre la tarea inmediata con reglas y estructura, de contexto claramente entendido, ocurre en todos los niveles de la organización.	Ocurre a través de perspectivas, ideas y heurísticas, no rutinario, desarrollo de estructuras diferenciadas, reglas (para enfrentar la falta de control), de contexto ambiguo, ocurre en los altos niveles.
Consecuencias	Resultados de comportamientos	Ideas, Heurísticas y conciencia colectiva
Ejemplos	Reglas formales institucionalizadas, ajustes en los sistemas de administración, habilidades para la solución de problemas	Nuevas misiones y definiciones de la dirección, configuración de la agenda habilidades para la solución de problemas, desarrollo de nuevos mitos, historias y cultura.

Tabla 7.1 Niveles de Aprendizaje. Fuente: Fiol & Lyles (1985).

La tabla 7.2 expone las definiciones de los elementos de un sistema de aprendizaje organizacional y nos muestra las definiciones de los stocks de conocimiento individual, de grupo y organizacional, así como los flujos de aprendizaje hacia adelante y atrás.

Índice	Concepto	Definición
IK	Stock de conocimiento individual	La capacidad y motivación individual para hacer el trabajo. Capital Humano.
GK	Stock de conocimiento de grupo	Dinámicas de grupo y entendimiento compartido, aprendizaje en equipo a través del diálogo, conocimiento embebido en interacciones sociales. Capital Relacional.
OK	Stock de conocimiento organizacional	Los sistemas, la estructura, la estrategia, los procedimientos y cultura; el conocimiento embebido en capital estructural, rutinas organizacionales.
FF	Flujo de aprendizaje hacia adelante	Como los individuos transfieren y/o comparten conocimiento al grupo-organización; y como el grupo transfiere y/o comparte conocimiento a la organización.
FB	Flujo de aprendizaje hacia atrás	Como la organización transfiere conocimiento al grupo-individuo; y como el grupo transfiere conocimiento al individuo.

Tabla 7.2: Definiciones de los elementos de un Sistema de Aprendizaje Organizacional
Fuente: Bontis (2000).

Un Sistema de aprendizaje organizacional integra las dimensiones clave del aprendizaje organizacional, lo cual incluye: 1) Múltiples niveles de análisis, 2) La operacionalización conceptual para que pueda ser medido, 3) La integración de los stocks y flujos de conocimiento (Bontis, 2000).

El proceso de aprendizaje organizacional cruza por los diferentes niveles de conocimiento (*stocks*) que existen en la organización. Al respecto conviene analizar cómo los diferentes niveles de stock de conocimiento se interrelacionan unos con otros a través de los flujos de aprendizaje hacia adelante y los flujos de aprendizaje hacia atrás ver figura 3.1. Esta figura nos muestra una matriz de tres niveles, con sus flujos hacia adelante y flujos hacia atrás. El primer nivel mostrado es el individual, que es donde nace el conocimiento, el segundo nivel es el de grupo, donde se dan las interacciones a través del diálogo y la discusión, y el tercer nivel es el organizacional donde se integra e institucionaliza todo el conocimiento generado en los niveles anteriores. Asimismo la figura nos muestra en el eje horizontal la exploración y en el eje vertical nos muestra la explotación del conocimiento. La *exploración* es el proceso de aprendizaje que tienen los individuos, los grupos y la organización. En el nivel del individuo, la exploración es transformada en conocimiento a través de la intuición e interpretación. En el nivel de grupo la exploración es transformada a través de la interpretación del individuo e integración al grupo. En el nivel organizacional la exploración es transformada a través de la integración de los grupos e institucionalización. Por otro lado la explotación, es el flujo hacia atrás, es a través de éste que logra la organización generar la riqueza; es decir la estructura, los

sistemas, los procedimientos y rutinas institucionalizadas es donde se apoyan los grupos e individuos para llevar a cabo su trabajo. Es la organización a través del flujo hacia atrás como logra dictar el rumbo que deberá llevar ésta.

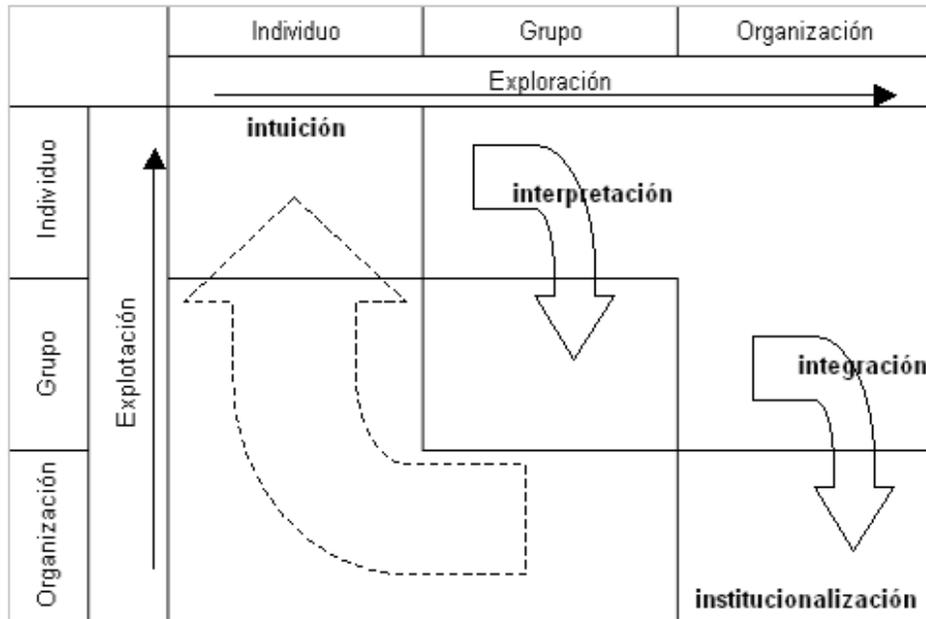


Figura 7.1: Aprendizaje Organizacional como un proceso dinámico. Fuente: Crossan, Lane & White (1999).

STOCKS DE CONOCIMIENTO:

En lo que se refiere a los stocks de conocimiento cabe distinguir tres niveles: individual, de grupo y organizacional.

STOCK DE CONOCIMIENTO INDIVIDUAL (Individual Knowledge)

En el nivel individual, los stocks de conocimiento son los que de manera tácita tienen las personas, así, las personas apoyando sus habilidades en la estructura de la organización, hacen crecer su conocimiento para entregarlo al grupo y a la organización a través de la interpretación e integración.

El stock de conocimiento individual (IK), también es conocido como capital humano, y está compuesto por educación, motivación, creatividad, agilidad mental, actitudes, aptitudes y competencias (Bontis, 2000).

STOCK DE CONOCIMIENTO DE GRUPO (Group Knowledge /GK)

En el nivel de grupo, los stocks de conocimiento son compartidos al grupo por individuos; es una mezcla de conocimiento tácito y explícito, coexisten los dos de una manera complementaria, y es donde se da la interacción de lo que Nonaka (1995) describe como la espiral de la creación del conocimiento, a éste también se le conoce como *capital relacional*.

Teniendo como base fundamental que el conocimiento de un individuo tiene que ser compartido a uno o más individuos, mediante el diálogo y la discusión, entonces podemos decir que, el conocimiento compartido es la base para el conocimiento de grupo (GK). La disciplina del aprendizaje implica dominar las prácticas del diálogo y la discusión, las dos maneras en que conversan los equipos. En el diálogo, existe la exploración libre y creativa de asuntos complejos y sutiles, donde se escucha a los demás y se suspenden las perspectivas propias. En cambio, en la discusión se presentan y defienden diferentes perspectivas y se busca la mejor perspectiva para respaldar las decisiones que se deben tomar. El diálogo y la discusión son potencialmente complementarios (Senge, 1998). El diálogo es uno de los elementos que describe también Nonaka & Takeuchi (1995), como destacado para la espiral de la creación del conocimiento. El cociente intelectual del equipo es potencialmente superior al de los individuos.

Una vez que el aprendizaje de grupo está integrado, entonces el aprendizaje de nivel organizacional empieza.

STOCK DE CONOCIMIENTO ORGANIZACIONAL (Organizational Knowledge /OK)

El último nivel de los stocks de conocimiento es el organizacional, son los stocks de conocimiento que posee la empresa; es todo aquel conocimiento tácito que poseen los individuos y los grupos de la organización, y que ha sido convertido en conocimiento explícito, que los individuos y los grupos han integrado e institucionalizado, transformándolo en capital estructural.

Quizás el aprendizaje organizacional ocurre a través de los individuos, pero sería un error concluir que el aprendizaje no es más que el resultado acumulado del aprendizaje de sus miembros. Las organizaciones no tienen cerebros, pero tienen sistemas cognitivos y memoria. Así como los individuos desarrollan su personalidad, sus hábitos personales y creencias con el paso del tiempo, las organizaciones desarrollan sus visiones del mundo e ideología. Los miembros van y vienen, el liderazgo cambia, pero la memoria de las organizaciones preserva ciertos comportamientos, mapas mentales, normas y valores al paso del tiempo (Fiol & Lyles, 1985).

El enfoque basado en los recursos y capacidades nos dice que los recursos que posee una empresa son más importantes que la manera en que está estructurada la industria a la que pertenece (Schendel, 1994 en Zapata, 2004). Uno de los recursos más importantes y difíciles de copiar es el cono-

cimiento que se transforma en productos y servicios que oferta la organización en el mercado en el que participa.

El conocimiento es una ventaja competitiva sostenible, nos dice Zapata (2004), si se fundamenta en las siguientes premisas:

Que esté basado en la experiencia acumulada más que en la tarea, lo cual implica que contenga una gran cantidad de conocimiento tácito, que hace difícil transferirlo de una empresa a otra e imitarlo por sus competidores.

Que no radique únicamente en los individuos, sino en la forma en que éstos individuos interactúan entre sí.

El stock de conocimiento organizacional (OK), también es definido como capital estructural, que está compuesto por la cultura corporativa, los procesos de administración, las bases de datos, la estructura organizacional, las patentes, las marcas registradas, y las relaciones financieras (Namasivayam & Denizci, 2006).

Los tres niveles de stocks de conocimiento pueden crecer con base a los flujos de aprendizaje.

FLUJOS DE APRENDIZAJE

En lo que se refiere a los flujos de aprendizaje cabe distinguir dos niveles: hacia adelante y hacia atrás.

FLUJO DE APRENDIZAJE HACIA ADELANTE (Flow Forward/FF)

El flujo de aprendizaje hacia adelante (FF), es el conocimiento que genera el individuo y que al momento de compartirlo y/o transferirlo, lo pone en movimiento hacia adelante ya sea que lo comparta y/o transfiera al grupo que pertenece, así como a la organización. También abarca el conocimiento que genera el grupo y lo transfiere y/o comparte con la organización. A esta acción se le conoce como el intercambio entre el conocimiento tácito para convertirse en conocimiento explícito, también se le conoce como el proceso de *exteriorización* (Nonaka & Takeuchi, 1995). El flujo hacia adelante inicia en el individuo, con los procesos de la intuición e interpretación. Al flujo que empieza en el individuo y termina en el grupo es referenciado como el proceso de integración. Al flujo que inicia el individuo y/o grupo y termina en la organización se le llama institucionalización (véase figura 3.1).

En lo que se refiere a la contribución del individuo a la organización y al flujo de aprendizaje hacia adelante inicia con la intuición, ésta es un proceso exclusivo del individuo que se lleva a cabo en un estado inconsciente; son ideas vagas sin acciones específicas; perspectivas que no están claramente definidas, con una conceptualización difusa. En el estado intuitivo las acciones son improvisadas, debido a que se fundamenta más en los senti-

mientos que en la racionalidad (Crossan, Lane & White, 1999), esto es flujo hacia adelante (exploración, véase figura 3.1).

Cuando el individuo realiza acciones de interpretación, se embebe en un proceso a través del cual el individuo toma de manera consciente los elementos que surgieron en la etapa intuitiva, de tal forma que los interpreta para convertir las perspectivas e ideas vagas en algo claro, aterrizado, conceptualizado. El proceso de interpretación individual viene a ser un entendimiento compartido de lo que es posible, interactuando los individuos e intentando llevar a cabo esa posibilidad (Crossan, Lane & White, 1999); en esta etapa es en la que el individuo comparte el conocimiento con el grupo. Esto es claramente una identificación del flujo de conocimiento hacia adelante (exploración, véase figura 3.1).

Un proceso más avanzado en el flujo de aprendizaje es la integración, proceso en el cual se incorpora el conocimiento generado por el individuo al grupo; es a través de la conversación continua entre los miembros de la comunidad y a través de la práctica compartida que se da el pensamiento colectivo desarrollando acciones negociadas. El diálogo y la discusión son las herramientas del grupo, que permiten evolucionar hacia nuevo y profundo entendimiento compartido (Crossan, Lane & White, 1999). Este proceso también es claramente una identificación del flujo hacia adelante (exploración).

Un cuarto elemento del flujo de aprendizaje es el relativo al proceso de institucionalización a través del cual el stock de conocimiento organizacional se incrusta dentro de la empresa u organización. Una organización es algo más que un grupo de individuos; el aprendizaje organizacional es muy diferente de la simple suma del aprendizaje de sus miembros. Quizás las personas pueden entrar o salir de la organización, pero el hecho de que ellos dejen la organización no significa que el conocimiento se irá con ellos. La institucionalización es fijar el aprendizaje dentro de la organización en sistemas, estructuras, estrategias, rutinas, procedimientos, sistemas de información e infraestructura.

En la figura 3.1 en el eje vertical se muestran las flechas del flujo de aprendizaje hacia adelante y es observado como la exploración que una organización realiza. Esta acción es la que permite a la organización mantenerse en equilibrio con respecto a sus flujos de salida.

FLUJO DE APRENDIZAJE HACIA ATRÁS (Flow Backward/FB)

El flujo de aprendizaje hacia atrás (FB), es el conjunto de conocimientos que posee la organización y que trasmite a los grupos e individuos, así como también el conjunto de conocimientos que poseen los grupos y que retroalimenta a los individuos. Estos flujos son considerados como la interacción entre el conocimiento explícito y el conocimiento

tácito. Cuando el conocimiento se transfiere a los individuos y/o grupos, se le conoce también como el *proceso de internalización* (Nonaka & Takeuchi, 1995).

Los flujos de aprendizaje hacia atrás son la explotación del conjunto de conocimientos que posee la organización y los grupos, y esto se hace a través de los individuos (Crossan, Lane & White, 1999). Esta explotación de conocimientos organizacionales y de grupo son las políticas, procedimientos, rutinas de trabajo, la cultura de la organización, la utilización de las patentes, marcas registradas, los programas de capacitación y desarrollo, la utilización de la infraestructura disponible en la organización, tecnologías de información y las bases de datos (Edvinsson, 2002).

Hay una necesidad para asegurar que los sistemas, procedimientos, rutinas embebidas en la organización logren los resultados y el desempeño deseado. Una organización usa sus sistemas para regular sus rutinas del día a día del negocio, explotando el conocimiento que posee.

Edvinsson (2002), nos dice que en el pasado las empresas se sentían satisfechas si podían hacer que $1+1=2$; en la economía del conocimiento donde ha crecido la importancia en la empresa del capital intelectual, el entero es muchas veces más que la suma de sus partes, es decir $1+1=11$. La suma de los stocks de conocimiento a nivel individual, de grupo y de organización representa el capital intelectual de la empresa.

CAPITAL INTELECTUAL (Intellectual Capital /IC)

En una época en que no sólo compañías sino categorías enteras de productos pueden desaparecer de la noche a la mañana, y en que los competidores pueden cambiar a diario sus relaciones y su relativa participación de mercado, los estados de pérdidas y ganancias y los balances generales ofrecen poco más que instantáneas de donde estuvo la compañía (Edvinsson & Malone, 1997). Cuando Henry Ford construyó su primer modelo T en 1908, lo vendió en \$900 dólares, para 1912 se vendía en \$690 dólares, en 1924 y 1925 el precio del coche se redujo a \$295 y \$290 dólares respectivamente. Como consecuencia de la industrialización y estandarización de la producción en masa, se logró bajar el precio del vehículo dramáticamente, Henry Ford no hizo otra cosa que aplicar conocimiento profundo, aumentando el capital intelectual de la empresa.

La exploración de la literatura del capital intelectual, nos da como resultado el presente material que sirve como marco para exponer lo relacionado con el capital intelectual.

Para Edvinsson & Malone (1997), el capital intelectual es el que mantiene a una empresa atractiva y sostenible en su creación del valor. Esto es lo que hace tan valioso el capital intelectual y encuentran dos factores ocultos que son:

Capital Humano: Combinación de conocimientos, destrezas, inventiva y capacidad de las personas para llevar a cabo las tareas encomendadas.

Incluye igualmente los valores de la organización, su cultura y su filosofía. La organización no puede ser propietaria del capital humano.

Capital Estructural: Los equipos, programas, bases de datos, estructura organizacional, patentes, marcas de fábrica y todo lo demás de la capacidad organizacional que sostiene la productividad de sus empleados; es decir todo lo que se queda en la oficina cuando los empleados se van a casa. También incluye el capital relacional que tiene con sus clientes.

$$\text{CAPITAL HUMANO} + \text{CAPITAL ESTRUCTURAL} = \text{CAPITAL INTELECTUAL}$$

Para Stewart (1997), el capital intelectual es el conocimiento, información, propiedad intelectual, experiencia, que pueden ser usados para la generación de riqueza.

De acuerdo con Roos, Roos, Dragonetti y Edvinsson (2001), el capital intelectual son las relaciones con los clientes, los proveedores y los socios aliados, la estructura, los sistemas que permiten que la empresa pueda llevar a cabo sus tareas diarias, así como las competencias, la actitud y agilidad mental que tienen las personas que pertenecen a la organización.

El crecimiento de los stocks de conocimiento que representan el capital intelectual depende de los flujos de aprendizaje, y la dinámica de estos es conocida como la administración del conocimiento.

ADMINISTRACIÓN DEL CONOCIMIENTO (Knowledge Management/KM)

La riqueza fluye directamente de la innovación, no de la optimización, es decir, la riqueza no se obtiene perfeccionando lo conocido, sino aprovechando de forma imperfecta lo desconocido (Kelly, 1997).

Para Andreu & Sieber (1999), la administración del conocimiento (KM) es el proceso que continuamente asegura el desarrollo y la aplicación de todo tipo de conocimientos pertinentes de una empresa, con objeto de mejorar su capacidad de resolución de problemas y así contribuir a la sostenibilidad de sus ventajas competitivas.

Según Bueno (1999), la administración del conocimiento es la función que planifica, coordina y controla los flujos de conocimiento que se producen en la empresa en relación con sus actividades y su entorno, con el fin de crear unas competencias esenciales.

A medida que la economía del conocimiento establece nuevas reglas de negocio, el cambio se va convirtiendo en algo más paradójico, las empresas con mayor crecimiento son aquellas que poseen mayormente activos intangibles y la administración de estos es lo que las lleva al éxito. La admi-

nistración del conocimiento es por consecuencia algo vital para las empresas, veamos cómo lo han definido algunos académicos:

Por su parte Wiig (1997), indica que la administración del conocimiento tiene perspectivas tácticas y operativas; es más detallada que la administración del capital intelectual y se centra en la forma de dar a conocer y administrar las actividades relacionadas con el conocimiento como su creación, captura, transformación y uso. Su función es planificar, implementar y controlar todas las actividades relacionadas con el conocimiento y los programas requeridos para la administración efectiva del capital intelectual.

Davenport & Prusak (1998) nos ofrecen lo que ellos llaman los principios de la administración del conocimiento:

1. El conocimiento se origina y reside en la mente de las personas.
2. Para compartirse el conocimiento se requiere confianza.
3. La tecnología habilita nuevos comportamientos del conocimiento.
4. Para compartirse el conocimiento deberá ser animado y recompensado.
5. El apoyo y los recursos de la administración son esenciales.
6. Las iniciativas de conocimiento deberán empezar con un programa piloto.
7. Las mediciones cualitativas y cuantitativas son necesarias para evaluar las iniciativas.
8. El conocimiento es creativo y deberá ser animado a desarrollarse en formas inesperadas.

De acuerdo con Davenport & Prusak, (1998), dentro de la administración del conocimiento (KM) se encuentran cuatro etapas que son:

- a) La generación del conocimiento.
- b) Codificación, almacenamiento e integración del conocimiento.
- c) Transferencia del conocimiento.
- d) Utilización del conocimiento.

Una administración del conocimiento eficaz permite que los flujos de aprendizaje, alimenten los tres niveles de stocks de conocimiento (exploración y explotación), es decir al capital intelectual.

Una explicación sobre la vinculación entre flujos y stocks de conocimiento la ofrece el profesor Ikujiro Nonaka de la Universidad de Hitotsubashi en Tokio. Nonaka escribió en el Harvard Business Review, Nov-Dec. 1991, "La compañía creadora de conocimiento" donde dice:

"Las compañías exitosas son aquellas que consistentemente crean nuevo conocimiento, y lo distribuyen a través de toda la organización y rápidamente queda transformado en nuevas tecnologías y nuevos productos".

Y declara, la creación del conocimiento debería ser el epicentro de la estrategia corporativa de la compañía.

La creación del conocimiento tiene dos dimensiones: la ontológica y la epistemológica.

Por la parte ontológica, el conocimiento es creado sólo por los individuos, la creación de conocimiento organizacional es un proceso que amplifica organizacionalmente el conocimiento creado por los individuos y lo solidifica como parte de la red de conocimiento de la organización. Este proceso se lleva a cabo en el interior de una creciente comunidad de interacción, la cual atraviesa niveles y fronteras intra e ínter organizacionales.

La dimensión epistemológica, establece las diferencias entre el conocimiento tácito y explícito. El tácito es personal y de contexto específico, difícil de formalizar y comunicar (Polanyi, 1966, en Prusak 1997); Davenport and Prusak (1998) nos dicen que los seres humanos aprenden mejor de historias, este precepto siempre ha sido más intuitivo para quien enseña. El conocimiento explícito o codificado es aquel que puede transmitirse utilizando el lenguaje formal y sistemático Davenport and Prusak (1998).

Habiendo definido las dimensiones de la creación del conocimiento, Nonaka nos dice, que es justamente en la interacción entre el conocimiento tácito y explícito, donde se da la creación del conocimiento, llamándole: la espiral del conocimiento, identificando cuatro etapas de la generación del conocimiento.

Socialización: Es el conocimiento tácito que posee una persona y transmite a través de la socialización con otras personas, de esta manera queda como conocimiento tácito en la persona que lo recibe.

Externalización: Es el conocimiento tácito que posee una persona y lo transmite codificándolo de una manera sistemática, para que todos puedan tener acceso a él.

Combinación: Es el conocimiento explícito, que genera y transmite nuevo conocimiento explícito para toda la organización, a partir de conocimiento explícito.

Internalización: Es el conocimiento explícito existente dentro de la organización y que se transmite a las personas, quedando convertido en conocimiento tácito.

La figura 7.2 nos muestra la espiral del conocimiento, con sus interacciones entre tácito y explícito.



Figura 7.2: Teoría de la Creación del Conocimiento Organizacional (Espiral del Conocimiento). Fuente: Nonaka (1995).

La administración del conocimiento implica stocks de conocimiento individual, de grupo y organizacional, así como flujos de aprendizaje hacia adelante, flujos de aprendizaje hacia atrás (Bontis, 2000). Sin embargo, un elemento de primera importancia para que la administración del conocimiento sea eficaz, es el compromiso de los individuos que forman parte de la organización para poder compartir con el grupo y la organización los conocimientos creados.

Cuando un individuo está comprometido con las metas y objetivos de la organización, se relaciona satisfactoriamente con otros individuos y grupos de la organización y desea permanecer en la empresa, entonces, estará dispuesto a realizar un esfuerzo extra en beneficio de la empresa, por lo tanto generará y compartirá el conocimiento con el grupo y la organización y permitirá que la administración del conocimiento dé mayores frutos en el capital intelectual de la empresa.

COMPROMISO ORGANIZACIONAL (Organizational Commitment / OC)

El compromiso organizacional surge de la intersección de las necesidades que tiene una organización y de la experiencia personal de los individuos que pertenecen a dicha organización. Por un lado, la organización con sus sistemas sociales tiene que conocer sistemáticamente sus necesidades; en lo que se refiere al individuo, es la orientación positiva o negativa, su postura emocional e intelectual hacia las situaciones que presenta la organización (Kanter, 1968).

Para Kanter (1968), el compromiso organizacional es definido como el proceso a través del cual, los intereses del individuo llegan a estar realizados y vinculados a los patrones de comportamiento socialmente organizados, los cuales son vistos como cumplimiento de esos intereses, expresando la naturaleza y necesidades de la persona.

El compromiso organizacional es definido como la actitud y el comportamiento que tiene un individuo dentro de una organización. Está integrado por tres elementos:

- a) La conformidad: es cuando el individuo adopta actitudes y comportamientos para obtener recompensas específicas y evitar castigos específicos.
- b) La identificación: es cuando el individuo adopta actitudes y comportamientos para ser asociado y relacionado satisfactoriamente con otros individuos o grupos.
- c) La interiorización: es cuando el individuo adopta actitudes y comportamientos que son congruentes con el sistema de valores del resto de los individuos.

Estos elementos representan las bases del compromiso organizacional (Kelman, 1958 en Becker, 1992).

Por su parte Mowday (1998), define el compromiso como la fuerza con la que un individuo se identifica e involucra con una organización e identifica tres componentes:

1. Una fuerte convicción y aceptación de las metas y valores de la organización.
2. La disposición para ejercer un esfuerzo considerable en beneficio de la organización.
3. El deseo de permanecer como miembro de la organización.

Es importante reconocer que los dos enfoques con los que se clasifica el compromiso, actitud y comportamiento son complementarios y nos ayudan a clarificar el entendimiento del compromiso organizacional. Por un lado la actitud es la disposición de los individuos en relación con la organización; ello puede ser visto como una inclinación del individuo a considerar las metas y valores de la organización, como si fueran las suyas. En lo que se refiere al enfoque de comportamiento, éste se vincula al proceso en el que el individuo se identifica con la organización, y comparte los problemas de la misma (Meyer & Allen, 1997).

Meyer & Allen (1997), notaron que varias de las definiciones sobre compromiso organizacional reflejaban tres amplios temas. El primero relacionado con la orientación afectiva, el segundo relacionado con los costos económicos y el tercero relacionado con la obligación y/o responsabilidad moral del individuo. Esto es, que el Compromiso ha sido visto como el reflejo de una orientación afectiva hacia la organización, un reconocimiento de los costos económicos asociados al dejar la organización, y una obligación

moral de permanecer con la organización. Al reconocer que cada uno de estos tres conjuntos de definiciones representan una clara y diferente conceptualización del Compromiso, Meyer & Allen proponen un modelo de tres componentes del compromiso organizacional.

Para la propuesta del modelo de tres componentes Allen & Meyer, también observaron algo en común de las varias definiciones de compromiso organizacional; señalaron que la percepción del Compromiso es un estado psicológico que: (a) caracteriza la relación del empleado con la organización, (b) tiene implicaciones en la decisión de continuar siendo miembro de la organización, y (c) que un empleado comprometido es más probable que permanezca en la organización a un empleado que no esté comprometido.

Así, Allen & Meyer (1990), plantean un modelo de compromiso con tres componentes:

- **Afectivo:** La atadura emocional del empleado, identificación e involucramiento con la organización.
- **Normativo:** Sentimientos de obligación para continuar en la organización.
- **Continuidad:** Los costos asociados si se deja la organización.

COMPROMISO AFECTIVO (Affective Commitment / CA)

El compromiso afectivo (CA) se refiere a la atadura emocional del empleado, la identificación e involucramiento con la organización. Empleados con un fuerte compromiso afectivo continúan con la organización porque ellos lo desean así.

De acuerdo con Kanter, (1968), el compromiso afectivo es la atadura que se encuentra en un individuo de afectividad y emoción hacia el grupo con el que se relaciona.

Sheldon, (1971) señala que el compromiso afectivo es la actitud u orientación hacia la organización, con la cual se liga o ata la identidad de la persona.

Para Hall, Schneider, & Nygren, (1970) el compromiso afectivo (CA) es un proceso a través del cual las metas de la organización y las del individuo llegan a ser congruentes o integrarse poco a poco.

Por su parte, para Buchanan, el compromiso afectivo es cuando un individuo muestra un sentimiento afectivo a las metas y valores de la organización, así como a su papel con respecto a éstas, por su propio deseo.

En el caso de Mowday, Porter & Steers, (1982) el compromiso afectivo (CA) es la fuerza relativa que identifica a un individuo y se involucra con una organización en particular.

COMPROMISO DE CONTINUIDAD (CC)

El compromiso de continuidad (CC) se refiere a estar consciente de los costos económicos asociados que implica dejar la organización. Un empleado que su liga principal está basado en el compromiso de continuidad, permanece porque él necesita hacerlo.

Para Kanter (1968) el compromiso de continuidad es el beneficio asociado con la participación continua y los costos asociados al dejar la organización.

De acuerdo con Becker (1992) el compromiso de continuidad es cuando la persona le apuesta al interés y ésta se liga con una actividad constante de su parte.

Por su parte Hrebiniak & Alutto (1972), señalan que el compromiso de continuidad es un fenómeno estructural, el cual ocurre como resultado de las transacciones e interacciones entre el individuo y la organización al apostarle e invertirle tiempo.

COMPROMISO NORMATIVO (CN)

El compromiso normativo (CN), refleja un sentimiento de obligación moral por continuar con la organización. Los empleados con un alto nivel de Compromiso Normativo sienten que ellos están obligados a permanecer con la organización.

De acuerdo con Weiner & Gechman (1977), el compromiso normativo es cuando un individuo tiene un comportamiento o conducta que excede las expectativas relevantes de la lealtad y/o formal del objeto del compromiso.

Weiner (1982) nos dice que el compromiso normativo es cuando la totalidad de la presión de lealtad internalizada en el individuo, actúa de tal forma que entrelaza los intereses del individuo y las metas organizacionales.

Para Marsh & Mannari (1977) el compromiso normativo es cuando un empleado comprometido, considera moralmente correcto permanecer en la organización, sin importar cuánto mejoró su estatus o cuánta satisfacción le haya dado la empresa a lo largo de los años.

Como se ha analizado, el aprendizaje organizacional es un sistema compuesto de dos elementos, uno de ellos estático y el otro dinámico. El capital intelectual es el elemento que identificamos como estático, integrado por los stocks de conocimiento en sus tres diferentes niveles: individual, de grupo y organizacional; por otro lado el elemento dinámico es la administración del conocimiento formado por los flujos de aprendizaje hacia adelante y hacia atrás. El aprendizaje se presenta por la interacción de los stocks de conocimiento entre sí, a través de los flujos hacia adelante y atrás, derivado de la intuición, interpretación, integración e institucionalización.

Algo de vital importancia dentro de las empresas, es que los flujos se mantengan en equilibrio; esto permitirá que el capital intelectual aumente con una eficaz administración del conocimiento.

Pero no basta el proceso de flujo si no se considera a las personas, éstas son la parte más importante del sistema, puesto que es en ellas donde reside el compromiso organizacional, siendo el motor principal para motivarlas a generar, codificar e integrar, así como transferir y/o compartir, y a la vez utilizar el conocimiento generado por ellos mismos y sus compañeros, o el conocimiento existente dentro de la organización.

El compromiso organizacional tiene que ver con la actitud y el comportamiento de los personas en la organización, éste es la fuerza con la que un individuo se identifica e involucra con la organización, aceptando las metas y valores de la empresa haciéndolos parte de él, teniendo la disposición para ejercer un esfuerzo considerable en beneficio de la empresa, resultando un deseo de permanecer como miembro de la organización.

Así, si aumenta el compromiso organizacional se obtendrán efectos positivos en la administración del conocimiento y se incrementará el capital intelectual.

En los próximos puntos analizamos la relación entre compromiso organizacional, la administración del conocimiento y el capital intelectual.

Solución propuesta e hipótesis

El objetivo de este punto es describir el modelo propuesto de la investigación, la hipótesis general y los siete grupos de hipótesis que se desprenden de la hipótesis general. Está integrado por ocho secciones que son: el modelo conceptual propuesto, así como los grupos de hipótesis del impacto del compromiso organizacional (OC) sobre el stock de conocimiento individual (IK), el stock de conocimiento de grupo (GK), el stock de conocimiento organizacional (OK), el flujo de aprendizaje hacia adelante (FF) el flujo de aprendizaje hacia atrás (FB), el capital intelectual (IC) y la administración del conocimiento (KM).

El modelo conceptual propuesto en el figura 7.3 nos permite conocer como se dan las interrelaciones entre las variables dependientes (IK, GK, OK, FF, FB, IC, KM) e independientes (CA, CC, CN). En lo que se refiere a los grupos de hipótesis, se muestra, cómo se alcanzan los objetivos específicos propuestos del estudio, así como las respuestas para cada una de las preguntas planteadas en la investigación. Todo esto se logra a través de la operacionalización de las variables transformándose en siete ecuaciones, una para cada grupo de hipótesis.

Modelo propuesto

Es importante señalar que esta investigación tiene como propósito principal encontrar cuál es el impacto del compromiso organizacional sobre un sistema de aprendizaje organizacional y de esta manera habilitar a la alta gerencia de la organización con la información necesaria para modificar sus estrategias en la administración de recursos humanos, ayudar a la mejor utilización y fluidez de los stocks de conocimiento en todos los niveles de la organización, a fin de mejorar la administración del conocimiento e incrementar el capital intelectual de la organización.

El marco teórico de esta investigación nos da el sustento para lanzar nuestras hipótesis. Esta investigación se propone probar que el compromiso organizacional impacta sobre un Sistema de aprendizaje organizacional.

Como podemos observar en la figura del Modelo de la relación del Compromiso Organizacional con un Sistema de Aprendizaje Organizacional, el modelo conceptual de la investigación, nos permite lanzar de manera general nuestra hipótesis central:

“Existe un impacto positivo del compromiso organizacional en la administración del conocimiento y el capital intelectual. A mayor nivel de compromiso organizacional de los individuos, habrá una mejor administración del conocimiento y por consecuencia el crecimiento del capital intelectual.”

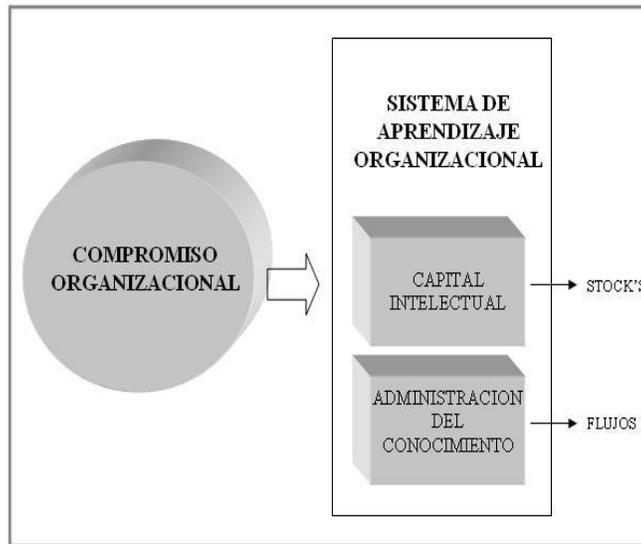


Figura 7.3: Modelo de la relación del Compromiso Organizacional con un Sistema de Aprendizaje Organizacional Fuente: Diseñado por el Dr. Arturo Reyes, en base a revisión de literatura.

El modelo de la relación del compromiso organizacional de Allen & Meyer con los componentes básicos de stocks (capital intelectual) y flujos de aprendizaje de conocimiento (administración del conocimiento) de un

sistema de aprendizaje organizacional de Crossan & Lane & White (véase la figura 3.4), nos permite comprender de una mejor forma, como las variables se relacionan. El modelo nos muestra de manera gráfica como se dan los impactos del compromiso organizacional en cada uno de los componentes básicos del sistema de aprendizaje organizacional.

Dichos impactos son los que dan lugar a los siete grupos de hipótesis, lo que a su vez permite la operacionalización de las variables dependientes que son IK = stock de conocimiento individual, GK = Stock de conocimiento de grupo, OK = Stock de conocimiento organizacional e IC = Capital intelectual que es equivalente al promedio de los tres stocks de conocimiento $IC = (IK+GK+OK)/3$ (Bontis, 2000, p. 30); estas variables se refieren a los stocks. Para el caso de los flujos son:

FF = Flujo de aprendizaje hacia adelante

FB = Flujo de aprendizaje hacia atrás

KM = Administración del conocimiento

KM es igual al promedio de los flujos de entrada y salida del sistema

$$KM = (FF + FB) / 2$$

(Bontis, 2000, p. 30).

Por otra parte, las variables independientes se encuentran agrupadas en el CO=Compromiso Organizacional, refiriéndose éste a tres elementos:

CA=Compromiso afectivo

CC=Compromiso de continuidad

CN=Compromiso normativo

$$CO = (CA+CC+CN)/3.$$

Teniendo como base nuestra hipótesis central, derivamos siete grupos de hipótesis, donde cada uno de estos grupos de hipótesis a su vez está formado por 3 hipótesis básicas, según se explica más adelante.

Hipótesis del Impacto del Compromiso Organizacional sobre los Stocks de Conocimiento

Conocimiento Individual:

El primer grupo de hipótesis es acerca del impacto que tiene el compromiso organizacional en el stock de conocimiento individual (IK), con lo cual se obtiene el primer objetivo de la investigación, y nos habla de:

“Identificar y evaluar cómo impacta el compromiso organizacional en el stock de conocimiento individual”.

Lo anterior da lugar a las hipótesis básicas H1, H2, H3 y la formulación de la primera ecuación de la investigación (véase ecuación No.1), donde se operacionaliza la variable dependiente IK y las variables independientes CA, CC y CN.

Las hipótesis son:

H1: Existe un impacto positivo del compromiso afectivo en el stock de conocimiento individual.

H2: Existe un impacto positivo del compromiso de continuidad en el stock de conocimiento individual.

H3: Existe un impacto positivo del compromiso normativo en el stock de conocimiento individual.

Este grupo de hipótesis se expresa matemáticamente mediante la siguiente ecuación.

Ecuación 1: Impacto del compromiso organizacional en el stock de conocimiento individual.

$$IK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$$

Conocimiento de Grupo:

El segundo grupo de hipótesis se refiere al impacto que tiene el compromiso organizacional en el stock de conocimiento de grupo (GK). Es en este grupo de hipótesis donde se alcanza el segundo objetivo de nuestra investigación que nos plantea lo siguiente: "Identificar y evaluar cómo impacta el compromiso organizacional en el stock de conocimiento de grupo".

En este grupo se integran las hipótesis básicas H4, H5 y H6. La ecuación No. 2 nos muestra la operacionalización de las variables independientes CA, CC, CN y la dependiente GK.

Las hipótesis son:

H4: Existe un impacto positivo del compromiso afectivo en el stock de conocimiento de grupo.

H5: Existe un impacto positivo del compromiso de continuidad en el stock de conocimiento de grupo.

H6: Existe un impacto positivo del compromiso normativo en el stock de conocimiento de grupo.

La ecuación No. 2 es una ecuación de regresión lineal múltiple que nos permite expresar matemáticamente, este grupo de hipótesis.

Ecuación 2: Impacto del compromiso organizacional en el stock de conocimiento de grupo.

$$GK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$$

Conocimiento Organizacional:

El tercer objetivo que perseguimos dentro de la investigación es: “Identificar y evaluar cómo impacta el compromiso organizacional en el stock de conocimiento organizacional”, y éste lo alcanzamos a través del tercer grupo de hipótesis, que se refiere al impacto del compromiso organizacional en el stock de conocimiento organizacional (OK).

Este grupo integra las hipótesis básicas H7, H8 y H9 y operacionaliza la variable dependiente OK y las variables independientes CA, CC y CN.

Las hipótesis son:

H7: Existe un impacto positivo del compromiso afectivo en el stock de Conocimiento Organizacional.

H8: Existe un impacto positivo del compromiso de continuidad en el stock de conocimiento organizacional.

H9: Existe un impacto positivo del compromiso normativo en el stock de conocimiento organizacional.

La expresión matemática para este grupo de hipótesis se observa en la ecuación No. 3.

Ecuación 3: Impacto del compromiso organizacional en el stock de conocimiento organizacional.

$$OK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$$

Hipótesis del Impacto del Compromiso Organizacional sobre los Flujos de Aprendizaje

Flujo de Aprendizaje hacia Adelante:

El cuarto grupo de hipótesis que describe el impacto del compromiso organizacional en los flujos de aprendizaje hacia adelante, responde al objetivo que señala “Identificar y evaluar cómo impacta el compromiso organizacional en los flujos de Aprendizaje hacia Adelante”.

Las hipótesis básicas H10, H11 y H12 son las que conforman este grupo de hipótesis y se operacionalizan a través de las variables independientes CA, CC, CN y la variable dependiente FF.

Las hipótesis son:

H10: Existe un impacto positivo del compromiso afectivo en el flujo de aprendizaje hacia adelante.

H11: Existe un impacto positivo del compromiso de continuidad en el flujo de aprendizaje hacia adelante.

H12: Existe un impacto positivo del compromiso normativo en el flujo de aprendizaje hacia adelante.

Este grupo de hipótesis se puede expresar matemáticamente mediante la siguiente ecuación.

Ecuación 4: Impacto del compromiso organizacional en el flujo de aprendizaje organizacional hacia adelante.

$$FF = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$$

Flujo de Aprendizaje hacia Atrás:

El quinto grupo de hipótesis surge de la relación del compromiso organizacional sobre los flujos de aprendizaje hacia atrás (FB); con este grupo de hipótesis se alcanza el objetivo de “Identificar y evaluar cómo impacta el compromiso organizacional en los flujos de aprendizaje hacia atrás”.

Este grupo integra las hipótesis básicas H13, H14 y H15 de los flujos de retroalimentación del sistema.

Las hipótesis son:

H13: Existe un impacto positivo del compromiso afectivo en el flujo de aprendizaje hacia atrás.

H14: Existe un impacto positivo del compromiso de continuidad en el flujo de aprendizaje hacia atrás.

H15: Existe un impacto positivo del compromiso normativo en el flujo de aprendizaje hacia atrás.

La ecuación No. 5 de regresión lineal múltiple es la expresión matemática de este grupo de hipótesis, con la que se solventa la pregunta mencionada en este punto.

Ecuación 5: Impacto del compromiso organizacional en el flujo de aprendizaje hacia atrás.

$$FB = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$$

La figura del Modelo Hipotético del impacto del Compromiso Organizacional en los componentes del Capital Intelectual y la Administración del Conocimiento, muestra gráficamente la relación de cada una de las variables dependientes con las variables independientes, así como las hipótesis básicas de la H1 hasta la H15 que son identificadas como los impactos que

tienen los elementos del compromiso organizacional en los elementos del sistema de aprendizaje organizacional.

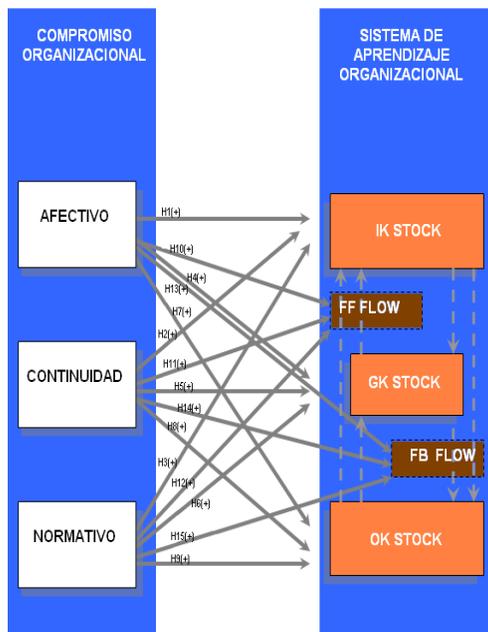


Figura 7.4: Modelo Hipotético del Impacto del compromiso organizacional en los Componentes del Capital Intelectual (Stocks) y la administración del conocimiento (Flujos), Macro Componentes de un sistema de aprendizaje organizacional. Fuente: Elaboración gráfica del Dr. Arturo Reyes en base al marco teórico de la investigación.

Hipótesis del Impacto del Compromiso Organizacional sobre el Capital Intelectual

Como se vio anteriormente, el capital intelectual (IC) es igual al promedio de los tres niveles de stock de conocimiento de una organización; siendo éste la parte que agrupa los stocks, resulta importante para el estudio identificar y evaluar los impactos del compromiso organizacional sobre dicho capital. Se apoya en la ecuación de regresión lineal múltiple surgida del grupo de hipótesis de esta sección.

Las hipótesis básicas resultantes en este grupo son H16, H17 y H18; la variable dependiente IC y las variables independientes CA, CC y CN. Estas variables forman parte de la ecuación No. 6.

Las hipótesis son:

H16: Existe un impacto positivo del compromiso afectivo en el capital intelectual.

H17: Existe un impacto positivo del compromiso de continuidad en el capital intelectual.

H18: Existe un impacto positivo del compromiso normativo en el capital intelectual.

Este grupo de hipótesis se puede expresar matemáticamente mediante la siguiente ecuación.

Ecuación 6: Impacto del compromiso organizacional en el capital intelectual.

$$IC = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$$

Hipótesis del Impacto del Compromiso Organizacional sobre la Administración del Conocimiento

El último grupo de hipótesis se vincula al impacto del compromiso organizacional en la administración del conocimiento (KM). Como ya se estableció en este apartado, la administración del conocimiento (KM) es el promedio de los dos flujos de Aprendizaje (hacia Adelante y hacia Atrás), siendo éste el componente dinámico del sistema de aprendizaje organizacional. Ello tiene particular importancia, debido a que es, en la administración del conocimiento (KM) donde se tiene que dar el impacto positivo del compromiso organizacional para que a su vez el capital intelectual se incremente.

Las hipótesis básicas H19, H20 y H21, integran el grupo de hipótesis que se refieren a la administración del conocimiento (KM) y generan la ecuación No. 7 con la que se le da respuesta a la pregunta No. 7, y se logra el último objetivo planteado en esta investigación.

Las hipótesis son:

H19: Existe un impacto positivo del compromiso afectivo sobre la administración del conocimiento.

H20: Existe un impacto positivo del compromiso de continuidad sobre la administración del conocimiento.

H21: Existe un impacto positivo del compromiso normativo sobre la administración del conocimiento.

Este grupo de hipótesis se puede expresar matemáticamente mediante la siguiente ecuación.

Ecuación 7: Impacto del compromiso organizacional en la administración del conocimiento.

$$KM = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$$

La figura 3.5 representa gráficamente la relación del compromiso organizacional sobre la administración del conocimiento y el capital intelectual, así como las hipótesis básicas de la H16 hasta la H21 que son identificadas como los impactos de los elementos del compromiso organizacional en los dos macro componentes de un sistema de aprendizaje organizacional.

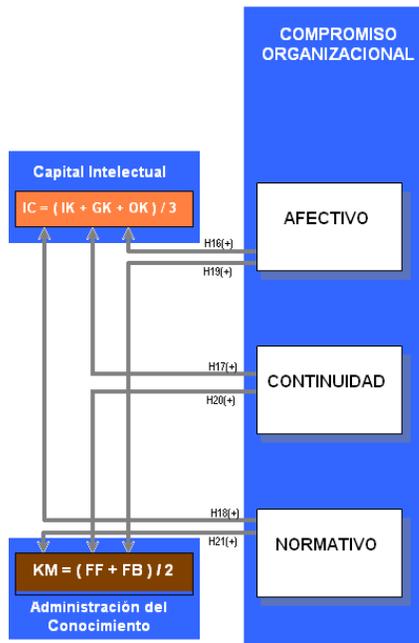


Figura 7.5: Modelo Hipotético del Impacto del compromiso organizacional en la administración del conocimiento y capital intelectual, Macro Componentes de un sistema de aprendizaje organizacional. Fuente: Elaboración gráfica del Dr. Arturo Reyes Valdez en base a marco teórico de la investigación.

Metodología y diseño de la investigación

El propósito de este punto es describir la metodología utilizada en el caso de estudio. Está integrado por cuatro elementos que son el diseño de la muestra, la selección de la muestra, la técnica de investigación y el instrumento de medición utilizado. En el diseño de la muestra se señala cuáles fueron los criterios que se utilizaron para determinar las empresas participantes en el estudio. La selección de la muestra incluye la mecánica instrumentada para captar las empresas participantes en la investigación.

En la técnica de investigación señalamos la forma como se llevo a cabo el estudio. Y por último, se describe el instrumento de medición, así co-

mo el origen e instrumentación de la encuesta utilizada para el levantamiento de la información en las empresas participantes.

Diseño de la muestra

La ciudad de Monterrey tiene un asentamiento industrial que data de finales del siglo XIX y principio del XX. Son de particular importancia las industrias del acero, vidrio, cerveza, química, y manufactura de productos relacionadas con el acero. En los últimos 25 años del siglo XX empiezan a surgir las empresas de servicios como son los bancos, telecomunicaciones, comercio en detalle, comercio de mayoristas, cuidado de la salud, educación y logística.

	Cantidad	Porcentaje
Manufactura	108	54%
Servicios	90	45%
Primario	2	1%
Totales	200	100%

Tabla 7.3: Distribución por sector económico de las 200 empresas más grandes de Monterrey y su área Metropolitana. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez en base a la información proporcionada por la Secretaría de Desarrollo Económico del Gobierno del Estado de Nuevo León.

Según datos del INEGI (2008), el estado de Nuevo León aporta el 7.5% del Producto Interno Bruto nacional “PIB”, se encuentra en tercer lugar después de Distrito Federal con 21.5% del “PIB” y el Estado de México con 9.7% del “PIB”. Por el lugar que ocupa en importancia de la aportación al “PIB” se considera un polo de desarrollo económico dentro de México. Si observamos que en los últimos 25 años del siglo XX aparecen las empresas de servicios y a su vez las de manufactura que utiliza una gran cantidad de conocimiento para llevar a cabo su trabajo, son consideradas, como empresas de conocimiento intensivo o de alto valor en sus activos intangibles, resulta relevante una investigación al respecto.

De acuerdo con la Secretaría de Desarrollo Económico del Gobierno del Estado de Nuevo León (2006), las empresas de servicios representan el 45% del total de las 200 empresas más grandes de Monterrey y su área metropolitana (véase tabla 4.1). Estas empresas en lo general, son consideradas intensivas en capital humano y capital intelectual y en procesos relacionados con la administración del conocimiento. Por lo tanto, resulta relevante para los académicos y la alta gerencia de las empresas, estudiar los factores que puedan afectar las acciones relativas a la administración del conocimiento y el crecimiento del capital intelectual.

El estudio que llevamos a cabo considera el análisis de dichos conceptos para una población surgida de las empresas más grandes de la ciudad de Monterrey y su área Metropolitana. Dicha población es acotada para un nivel organizacional solamente de Dirección, Gerencia y Empleados claves que manejan información fundamental para el funcionamiento de la empresa. La muestra se diseñó sólo para aquellas empresas que se encuentran dentro de Monterrey y su área metropolitana, esta restricción fue por con-

veniencia geográfica, pues en Monterrey y su área Metropolitana se concentran todas las empresas que pertenecen a la población en cuestión. Así mismo, tomando en cuenta la consideración de que Nuevo León participa con el 7.5% del PIB nacional (INEGI, 2008), consideramos que en Monterrey existe una cantidad importante de empresas con las características deseadas para el estudio.

De acuerdo a la información de la Secretaría de Desarrollo Económico del Estado de Nuevo León, México, el universo de las empresas más grandes de Monterrey y su área Metropolitana asciende a 200. La muestra elegida para contrastar las hipótesis planteadas en el modelo conceptual, la seleccionamos de esta población.

Selección y tamaño de la muestra

Para efectos de seleccionar la muestra participante en este estudio, tomamos como base, la información encontrada en la Secretaría de Desarrollo Económico del Gobierno del Estado de Nuevo León (2006). Dicha información muestra una base de datos con las 200 empresas de mayor tamaño de la ciudad de Monterrey y su área Metropolitana.

En virtud de que la población objetivo fueron las personas clave que pertenecen al primer nivel de la organización, considerando que tienen acceso a información vital para el funcionamiento de la empresa. El número estimado fue de 20 personas por empresa. Con esta cantidad de personas clave o de primer nivel que participarían y tomando en cuenta que son 200 empresas la base de datos seleccionada, calculamos como nuestra población total objetivo 4,000 personas.

Una vez definida la población objetivo y su tamaño, definimos el tamaño de la muestra basándonos en los métodos estadísticos que aplican para la técnica de análisis de regresión lineal múltiple, utilizada en esta investigación. En primer término mostraremos el desarrollo matemático para la media muestral (véase la Ecuación 8), posteriormente se mostrará para la varianza muestral (véase la Ecuación 9), y por último la ecuación que se utiliza para obtener el tamaño de muestra óptimo (véase la Ecuación 10).

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

\bar{x} = Media muestral
 x_i = Elemento de la muestra
 n = Total de la muestra

Ecuación No. 8: Obtención de la media muestral de una población.

La ecuación No. 8 nos permitió obtener la media muestral, y es utilizada para hacer inferencias estadísticas de la muestra con respecto a la población.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

s^2 = Varianza muestral de la población
 \bar{x} = Media muestral
 x_i = Elemento de la muestra
 n = Total de la muestra

Ecuación No. 9: Obtención de la varianza muestral de una población.

La ecuación No. 9 nos permitió obtener la varianza muestral, y es utilizada para hacer inferencias estadísticas de la muestra con respecto a la población.

$$n = \frac{Z_{\alpha/2}^2 S^2}{e^2}$$

n = Tamaño de la muestra
 $Z_{\alpha/2}^2$ = Tipificación normal estándar
 S^2 = Varianza muestral de la población
 e^2 = Estimación del error

Ecuación No. 10: Tamaño de la muestra de una población.

La ecuación 10 nos permitió obtener el tamaño de muestra óptimo de la población objetivo. La estimación del error se consideró como $e=0.1$, y el intervalo de confianza que se define para realizar el cálculo de la muestra es de 95%, esto nos permitió conservar un 0.10 de error en ambos extremos de la curva de la distribución normal, esto nos da un valor para $Z_{\alpha/2} = 1.96$. De las variables que utilizamos para obtener el tamaño de la muestra, solamente la varianza muestral es el valor no conocido, por lo tanto, esto nos llevó a realizar un cálculo de la varianza muestral con la muestra tomada durante el segundo semestre del 2006 de la cual obtuvimos 152 personas encuestadas y los resultados de la media muestral y varianza muestral se pueden observar en la tabla No. 4.5. Una vez obtenida la varianza muestral, completamos los valores de las variables involucradas en la obtención del tamaño de la muestra. Por lo tanto, procedimos a realizar el cálculo del tamaño de la muestra de nuestra población objetivo, en donde el resultado arrojado fue de 212.44 personas (véase la ecuación 11), esto lo redondeamos a 213 encuestas; si consideramos una población objetivo de 4000 personas en términos de porcentaje resulta que el tamaño de muestra debió ser del 5.325% del total de la población. En aras de obtener un mejor margen para las condiciones de normalidad y a su vez tratando de disminuir el margen de error, el muestreo continuó durante el primer semestre del 2007 llegando a obtener 265 personas encuestadas, lo cual nos dió un 6.62% del total de la población, esto es 1.295% por encima de lo recomendado como tamaño de muestra óptima, permitiéndonos mejorar las condiciones de normalidad y a su vez disminuir el margen de error.

La invitación para participar en este estudio se llevo a cabo aleatoriamente, a través del contacto con los departamentos de recursos humanos de las empresas. En el proceso de habilitar los canales de comunicación con las empresas, tuvimos el acercamiento con 36 organizaciones, de las cuales 29 aceptaron una entrevista, donde se nos permitió explicar con detalle la manera de cómo podían participar en la investigación y que beneficios obtendrían de ésta. En el caso de las 29 empresas que accedieron a la entrevista, en 9 de ellas las personas que nos recibieron (Director o Gerente), nos hicieron saber que debían consultar con su jefe inmediato la decisión de involucrarse en la investigación, finalmente 20 de las 29 organizaciones aceptaron participar.

El instrumento utilizado para la recolección de la información en las empresas fue una encuesta impresa en papel, entregada en paquetes de 20 encuestas por organización, terminando con un total de 400 encuestas en-

tregadas. El mecanismo de aplicación que utilizamos para la obtención de la información, fue entregar la encuesta a directores y/o gerentes así como también a personas que por su trabajo operan información clave o vital para el funcionamiento de la empresa. El departamento de Recursos Humanos de cada empresa en lo particular, entregó la encuesta a las personas con las características mencionadas, de igual manera las recolectaron, quedando pendiente solamente fijar una fecha para recoger el total de encuestas recuperadas.

La aplicación de las encuestas transcurrió durante el segundo semestre del 2006 y el primer semestre del 2007, la cantidad de empresas que finalmente participaron fueron en total 16, obtuvimos 265 encuestas en total, con un resultado promedio de 16.5 encuestas por empresa. Este tamaño de muestra superó en número, a las 213 encuestas sugeridas para el tamaño de nuestra muestra óptima, como se observó en los cálculos anteriores. El porcentaje de recuperación con respecto a las entregadas fue de 66.25% y en referencia al tamaño de la población objetivo fue de 6.62%.

De la muestra total obtenida de 265 encuestas, la distribución por nivel organizacional fue de 11 directores con un 4.15%, 63 gerentes con un 23.77%, y por último 191 empleados con 72.07% (véase la tabla No. 7.4). De las 16 empresas que participaron en la investigación, en su distribución por sector económico, 6 pertenecen al sector de servicios con un 37.5% y 10 al sector de manufactura con un 62.5% (véase la tabla No. 7.5). En el siguiente punto se abordará el tema de la técnica de investigación utilizada en el estudio.

$$n = \frac{z_{\alpha/2}^2 s^2}{e^2} = \frac{(1.96)^2 (0.553)^2}{(0.10)^2} = \frac{2.1244}{0.01} = 212.44$$

Ecuación No. 11: Obtención del tamaño de la muestra de la población.

Nivel	Cantidad	Porcentaje
Director	11	4.15 %
Gerentes	63	23.77 %
Empleados	191	72.07 %
Totales	265	100%

Tabla No. 7.4: Distribución por nivel organizacional de los encuestados. Fuente: Tabla generada por el Dr. Arturo Reyes Valdez con resultados generados con datos de investigación directa.

Sector	Cantidad	Porcentaje
Servicios	6	37.5 %
Manufactura	10	62.5 %
Totales	16	100%

Tabla No. 7.5: Distribución por sector económico. Fuente: Tabla generada por el Dr. Arturo Reyes Valdez con resultados generados con datos de investigación directa.

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.572	3.485	3.938	.453	1.178	.039	10
Item Variances	.553	.398	.704	.306	1.768	.012	10
Inter-Item Covariances	.271	.049	.544	.495	11.054	.019	10
Inter-Item Correlations	.506	.094	.946	.852	10.059	.072	10

Tabla 7.6: Valores de la media muestral, varianza muestral, mínimos y máximos, para el muestreo previo de 152 encuestas. Fuente: Tabla realizada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación directa

Técnica de investigación

En lo que corresponde a nuestra investigación observamos las cuatro técnicas de investigación en las diferentes etapas del estudio. Exploratoria, porque el objetivo fue examinar el impacto del compromiso organizacional en la administración del conocimiento y el capital intelectual, tema que no había sido abordado anteriormente, igualmente es descriptivo porque describe con detalle las propiedades y características de los elementos de la investigación, además es correlacional porque su propósito es analizar la relación de las variables independientes compromiso afectivo (CA), compromiso de continuidad (CC) y compromiso normativo (CN) con las variables dependientes stock de conocimiento individual (IK), stock de conocimiento de grupo (GK), stock de conocimiento organizacional (OK), flujo de aprendizaje hacia adelante (FF), flujo de aprendizaje hacia atrás (FB), capital intelectual (IC) y administración del conocimiento (KM), asimismo también es explicativo porque los objetivos fueron definir y establecer las causas del impacto del compromiso organizacional en la administración del conocimiento y el capital intelectual, es decir, el impacto de las variables independientes en las variables dependientes.

Este estudio se llevo a cabo en un entre el segundo semestre del 2006 y el primer semestre del 2007, la forma de recolección de datos fue una encuesta impresa a través de la cual la percepción de las personas que es cualitativa se transformó en cuantitativa apoyándose en la metodología de Likert, permitiéndonos analizar la información con las herramientas de estadística paramétrica. La evaluación del problema la realizamos mediante el método de análisis de regresión lineal múltiple.

Instrumento de medición

Actualmente son escasos los métodos para la medición de un sistema de aprendizaje organizacional en los diferentes niveles stock de conocimiento individual (IK), stock de conocimiento de grupo, stock de conocimiento organizacional (OK) y sus flujos de aprendizaje hacía adelante y

atrás. Una manera de medirlo es a través de una encuesta donde se capte la percepción de los individuos con respecto al aprendizaje organizacional; en esta investigación utilizamos como una de las herramientas la encuesta propuesta por Bontis (2000).

La herramienta utilizada para medir el compromiso organizacional es el instrumento de medición desarrollado por Meyer & Allen (1997), el cual ha sido aplicado desde su aparición en 1991 hasta el 2006 en más de 23 países, en diferentes contextos. Se han encontrado investigaciones de compromiso organizacional utilizando el instrumento de Meyer & Allen en el sector privado y en el sector público; dentro del sector privado la aplicación se ha realizado en ámbitos de servicios como bancos, universidades, hospitales, telecomunicaciones, y también manufactura; por otra parte, en el sector público se ha observado la aplicación de la encuesta en servicios que presta el gobierno como son la recolección de impuestos, servicios de salud, educación, y sistemas de penitenciaria.

Ambas herramientas mencionadas han sido probadas y validadas respectivamente por sus autores. Por lo tanto, en este estudio nos avocamos a las validaciones de confiabilidad (alfa de Cronbach), de contenido y la doble traducción, de tal manera que tuvimos la confianza de utilizar los instrumentos dentro de los parámetros establecidos para una investigación científica.

El instrumento como un todo fue compuesto por 71 preguntas de las cuales quedan estratificadas en 3 variables independientes y 7 variables dependientes. Cada variable quedó definida por un constructo que es la representación de la variable que participa en el modelo propuesto; a su vez, el constructo contiene una serie de preguntas que están relacionadas con la variable que se desea medir, donde cada pregunta tiene una palabra clave que apunta a obtener información de la variable en cuestión desde diferentes perspectivas. El conjunto de constructos se agrupan de tal manera, que permiten integrar el instrumento de medición.

En lo que se refiere al compromiso organizacional (CO) se integra por tres elementos: compromiso afectivo (CA), compromiso de continuidad (CC) y compromiso normativo (CN). Cada uno de estos constructos se compone de 7 preguntas.

Por otra parte, el sistema de aprendizaje organizacional está integrado por cinco elementos: stock de conocimiento individual (IK), stock de conocimiento de grupo (GK), stock de conocimiento organizacional (OK), flujo de aprendizaje hacia adelante (FF) y el flujo de aprendizaje hacia atrás (FB). Cada uno de estos constructos se compone de 10 preguntas.

Por último, dos macro elementos integran al sistema de aprendizaje organizacional: el capital intelectual IC (stocks) y la administración de conocimiento KM (flujos). El primero, definido como el promedio de la suma de los valores del stock de conocimiento en sus tres diferentes niveles que se encuentran en una organización IK, GK y OK. El segundo macro elemento, está definido por el promedio de la suma de los valores de los flujos de aprendizaje hacia adelante y hacia atrás FF y FB. Estos macro elementos

son considerados variables dependientes y sus valores son obtenidos a partir de los valores de los constructos IK, GK y OK para el caso de IC y los valores de los constructos FF y FB para KM.

En cada pregunta del cuestionario se utilizó la escala Likert de cinco elementos, donde: 1 = Nada de acuerdo, 2 = Poco de acuerdo, 3 = Ligeramente de acuerdo, 4 = De acuerdo y 5 = Totalmente de acuerdo. Cada una de las preguntas tiene el mismo valor.

La confiabilidad de cada constructo, variables independientes y dependientes fue probada con el alfa de Cronbach, involucrando todas las preguntas que integran cada uno de los constructos de las variables estudiadas, obteniendo lo siguiente:

Constructos de las variables independientes compromiso afectivo (CA) $\alpha=0.776$, compromiso de continuidad (CC) $\alpha=0.742$ y compromiso normativo (CN) $\alpha=0.757$. Para los constructos CC y CN algunas preguntas fueron eliminadas, debido a la baja correlación total con el resto de los elementos y por no alcanzar un mínimo de 0.7 (Peterson, 1994; Duhachek, Coughlan, Iacobucci, 2005) para el alfa de Cronbach. Estos valores de alfa de Cronbach fueron congruentes con las alfas de Cronbach reportadas por Meyer y Allen (1997) para CA $\alpha=0.85$, CC $\alpha=0.79$ y CN $\alpha=0.7$, respectivamente, y en el caso de la investigación realizada por Udechukwu (2006) reportó los siguientes valores CA $\alpha=0.792$, CC $\alpha=0.818$ y CN $\alpha=0.834$.

Los constructos para las variables dependientes presentaron los siguientes resultados del stock de conocimiento individual (IK) $\alpha=0.916$, stock de conocimiento de grupo (GK) $\alpha=0.874$, stock de conocimiento organizacional (OK) $\alpha=0.927$, flujo de aprendizaje hacia adelante (FF) $\alpha=0.904$ y flujo de aprendizaje hacia atrás (FB)

$\alpha=0.895$. Estos valores de alfa de Cronbach fueron congruentes con las alfas de Cronbach, reportados en la investigación de Bontis (1999) para IK $\alpha=0.909$, GK $\alpha=0.902$, OK $\alpha=0.904$, FF $\alpha=0.907$ y por último FB $\alpha=0.766$.

Análisis y Discusión de los Resultados

En este capítulo se presenta el análisis y discusión de los resultados obtenidos de la investigación.

En principio presentamos el análisis de confiabilidad (Alfa de Cronbach), los valores de las medias de cada variable involucrada, la Desviación Estándar, la Varianza, la Asimetría, la Curtosis, el rango, el valor mínimo y máximo de éstas y las correlaciones que presentaron el modelo de compromiso organizacional con respecto a la administración del conocimiento y el capital intelectual del conjunto muestral obtenido. Dentro de este apartado se revisan los resultados, por un lado las correlaciones de CA, CC y CN con los componentes del capital intelectual y la administración del conoci-

miento IK, GK, OK, FF y FB; por otro lado las correlaciones de CA, CC y CN con el capital intelectual y la administración del conocimiento.

El siguiente punto presenta el análisis de los modelos de impacto del compromiso organizacional con los componentes del capital intelectual y la administración del conocimiento, integrándose por cinco puntos. Como primer punto analizamos el modelo de impacto $IK = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$, en el segundo punto el modelo que examinamos es $GK = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$, en el tercer punto nos tocó explorar el modelo $OK = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$, para el cuarto punto de este apartado presentamos el análisis del modelo $FF = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$, el quinto y último modelo analizado en este apartado fue $FB = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$. Para todos y cada uno de los modelos anteriormente mencionados revisamos sus coeficientes de impacto, los valores “t” con su respectiva significancia, el valor arrojado “F” del modelo de estimación y su valor de significancia, y por último el coeficiente de determinación “R²”.

Otro punto más, es el análisis de los modelos de estimación del compromiso organizacional con los macro componentes de un sistema de aprendizaje organizacional, el capital intelectual y la administración del conocimiento, conteniendo dos puntos. El primer punto que abordamos para su análisis es el modelo de estimación $IC = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$ en el segundo, el modelo analizado es $KM = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$, para ambos modelos la exploración que llevamos a cabo es similar que en el apartado anterior.

Finalmente presentamos un resumen de los resultados del impacto que tiene el compromiso afectivo (CA), compromiso de continuidad (CC) y el compromiso normativo (CN) sobre el stock de conocimiento individual (IK), stock de conocimiento de grupo (GK), stock de conocimiento organizacional (OK), flujo de aprendizaje hacia adelante (FF), flujo de aprendizaje hacia atrás (FB), capital intelectual (IC) y administración del conocimiento (KM).

Análisis de confiabilidad, medias, varianza y correlación de las variables. Los procesos y cálculos estadísticos para la información de esta investigación, los hicimos con base en el programa estadístico “SPSS”. Apoyándonos en los resultados obtenidos realizamos el análisis de las medias, desviación estándar, varianza, Asimetría, Curtosis y correlaciones de las variables.

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.921	.927	10

Tabla 7.7: Valor de Alfa de Cronbach del instrumento de medición. Fuente: Tabla realizada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación directa.

Como podemos observar en la tabla 7.8 se nos muestra el valor del alfa de Cronbach del instrumento de medición integrado por 10 variables, esta prueba de confiabilidad constata que su uso, cumple con los parámetros estadísticos de confiabilidad, donde obtuvimos un alfa= 0.921 y su valor estandarizado de 0.927,

	Mean	Std. Deviation	Variance	Skewness	Kurtosis	Range	Minimum	Maximum	N
CA	3.99677	0.639836	0.409	-0.392	-0.487	3.00 0	2.000	5.000	265
CC	3.48528	0.789401	0.838	-0.369	0.389	4.00 0	1.000	5.000	265
CN	3.92453	0.834952	0.697	-0.701	0.150	4.00 0	1.000	5.000	265
IK	3.58075	0.695799	0.484	-0.485	0.088	3.40 0	1.600	5.000	265
GK	3.59962	0.628822	0.395	-0.753	0.575	3.10 0	1.600	4.700	265
OK	3.84264	0.726157	0.527	-0.581	0.154	3.60 0	1.400	5.000	265
FF	3.54004	0.723575	0.524	-0.445	0.244	4.00 0	1.000	5.000	265
FB	3.66340	0.705699	0.498	-0.382	0.157	3.40 0	1.600	5.000	265
IC	3.67434	0.612011	0.403	-0.578	0.271	3.43 3	1.567	5.000	265
KM	3.60172	0.674492	0.455	-0.427	0.049	3.65 0	1.300	4.950	265

Tabla 7.8: Valores de la Media por constructo, Desviación Estándar, Varianza, Asimetría y Curtosis.
Fuente: Tabla realizada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación directa.

muy superior al valor sugerido de 0.7 (Peterson, 1994; Duhachek, Coughlan, Iacobucci, 2005), estos resultados sugieren un alto nivel de confiabilidad del instrumento utilizado.

Los valores de las medias para cada una de las variables las mostramos en la tabla 7.9. En principio analizamos los valores de las medias obtenidas en esta investigación para las variables independientes compromiso afectivo (CA), compromiso de continuidad (CC) y compromiso normativo (CN). El valor máximo fue para el compromiso afectivo (CA) igual a 3.99677, el segundo valor en orden descendente fue para el compromiso normativo (CN) con 3.92453 y por último el compromiso de continuidad (CC) con 3.48528.

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.691	3.485	3.997	.511	1.147	.030	10
Item Variances	.499	.375	.697	.323	1.861	.010	10
Inter-Item Covariances	.269	.098	.461	.363	4.701	.011	10
Inter-Item Correlations	.559	.198	.945	.748	4.781	.055	10

Tabla 7.9: Valores de la Media muestral, Varianza muestral, Mínimos y Máximos. Fuente: Tabla realizada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación directa.

Comparando estos valores obtenidos en el presente estudio, con los mostrados por otros investigadores, encontramos que tenemos algunos valores similares, Kaneshiro (2008) presenta en su investigación las siguientes medias, como su valor más alto el compromiso afectivo (CA) con 3.422, en segundo término el compromiso de continuidad (CC) con 3.240 y por último el compromiso normativo (CN) con 3.126; mientras que Patalano (2008) halló al compromiso afectivo (CA) con el valor más alto 3.02, después al compromiso normativo (CN) y compromiso de continuidad (CC) empatados en segundo término presentando un valor de 2.94.

Observamos que el compromiso afectivo (CA) tiene el valor más alto en los tres estudios y el compromiso normativo (CN) tiene el segundo en esta investigación y la de Patalano (2008), mientras que en el trabajo de Kaneshiro (2008) está en tercer término, por último, el compromiso de continuidad (CC) que en el presente estudio está en tercer término, en Kaneshiro (2008) y Patalano (2008) aparece en segundo.

El otro grupo de valores de medias analizado corresponde a los componentes del capital intelectual (Stocks) y la administración del conocimiento (Flujos), estas son las variables dependientes stock de conocimiento individual (IK), stock de conocimiento de grupo (GK), stock de conocimiento organizacional (OK), flujo de aprendizaje hacia adelante (FF) y flujo de aprendizaje hacia atrás (FB).

Por el lado de los stocks de conocimiento tenemos que, el valor más alto correspondió al stock de conocimiento organizacional (OK) con 3.84264, posteriormente fue el stock de conocimiento de grupo (GK) con 3.59962 y por último el stock de conocimiento individual (IK) con 3.58075. En el caso de los flujos de aprendizaje se presenta primero con un valor más alto, el flujo de aprendizaje hacia atrás (FB) con 3.66340 y después el flujo de aprendizaje hacia adelante (FF) con un valor de 3.54004.

De acuerdo con la investigación de Bontis (2000), encontramos que en los stocks de conocimiento, el valor más alto corresponde al stock de conocimiento individual (IK) con 3.6682, el segundo es el stock de conoci-

miento de grupo (GK) con 3.5231, y por último el stock de conocimiento organizacional (OK) con 3.50221. Para los Flujos de Aprendizaje, los valores que se presentan son 3.4075 y 3.1682, para flujo de aprendizaje hacia atrás (FB) y flujo de aprendizaje hacia adelante (FF), respectivamente.

Al contrastar los valores de Stocks y Flujos de la presente investigación con el trabajo de Bontis (2000), podemos observar que es más fuerte el stock de conocimiento individual (IK) en el de Bontis (2000), mientras que esa variable en el presente estudio es la más débil. La variable que en la presente investigación, arrojó el valor más fuerte fue el stock de conocimiento organizacional (OK) y esa misma variable en el trabajo de Bontis (2000) es la más débil, el stock de conocimiento de grupo (GK) permanece igual en ambos trabajos, en segundo lugar. Los Flujos de Aprendizaje se mantienen en el mismo lugar en ambos trabajos, primero el flujo de aprendizaje hacia atrás (FB) y después el flujo de aprendizaje hacia adelante (FF).

Es de particular importancia resaltar que la investigación de Bontis (2000) fue realizada en Canadá, y participaron empresas que pertenecen al sector Financiero de todo el país, estas son empresas que caen en el sector de servicios, mientras que la presente investigación se realizó en México, dentro del estado de Nuevo León, particularmente en Monterrey y su área Metropolitana y participaron empresas del Sector de servicios y de manufactura, predominantemente manufactura (véase tabla 7.5).

En lo referente a los valores de las medias de las variables dependientes capital intelectual (IC) y administración del conocimiento (KM), obtuvimos un valor de 3.67434 para el capital intelectual (IC), con un 2.01% por encima del 3.60172 que correspondió a la administración del conocimiento (KM). Lo cual es indicativo de que, tanto Stocks como Flujos tienen valores muy balanceados.

Una de las condiciones que debe cumplir la información que se usa para los modelos de regresión lineal múltiple es la normalidad de los datos. Es decir, que la distribución de los datos se comporte como una distribución normal, con valores de Asimetría y Curtosis dentro de los límites de -1 a 1. Para probar esto, utilizamos las pruebas de Asimetría y Curtosis, cuyos resultados los podemos observar en la tabla 7.9. Los valores arrojados para la prueba de Asimetría van en un rango que inicia en -0.753 hasta -0.369, en el caso de la Curtosis el rango empieza en -0.487 llegando a 0.575. En base a los resultados obtenidos podemos afirmar que tenemos Asimetría negativa en todas las variables, pero dentro del rango de aceptación, para la Curtosis se presenta en tres variables de manera negativa CA, CC y FB para el resto de las variables encontramos Curtosis positiva, también dentro del rango de aceptación

De acuerdo con la tabla 7.10 el valor de la media muestral del conjunto de las diez variables que participan en el estudio es 3.691, presentando un rango de 0.511, que va desde un mínimo de 3.485 hasta un máximo de 3.997, la varianza muestral queda en 0.499, el rango observado para la varianza muestral es de 0.323 empezando en 0.375 y termina en 0.697.

En la siguiente sección, hacemos referencia al Coeficiente de Correlación Momento del Producto de Pearson de las variables involucradas, el análisis de éste lo dividimos en dos grupos, en el primero detallamos la Correlación de los elementos que integran el compromiso organizacional con los componentes del capital intelectual *stocks* y la administración del conocimiento *flujos*, en el segundo detallamos el análisis de la correlación que arroja el “SPSS”, entre los elementos del compromiso organizacional con el capital intelectual y la administración del conocimiento.

Correlación del compromiso organizacional y el proceso de aprendizaje

Las correlaciones que presentan las variables independientes CA, CC, CN con las variables dependientes IK, GK, OK, FF, FB son muy significativas, presentando uniformemente el valor de 0.01 sin excepción. Por otro lado, estas correlaciones estuvieron en un rango de 0.216 hasta 0.445, todas positivas, en términos generales se puede considerar que tenemos buena correlación con alta significancia (véase la tabla 7.11).

La figura 3.6 nos muestra de forma gráfica, la correlación de las variables en el modelo del impacto que tiene el compromiso organizacional sobre los componentes del capital intelectual *stocks* y la administración del conocimiento *flujos*. Para efectos de entender cuáles correlaciones tienen importancia por su alto valor de correlación y cuáles no la tienen, realizamos el análisis para cada una de ellas.

Obtuvimos los siguientes valores de correlación para la variable compromiso afectivo (CA) con respecto a las variables stock de conocimiento individual (IK), stock de conocimiento de grupo (GK) y stock de conocimiento organizacional (OK), que son los *stocks* de conocimiento de la empresa, y éstos a su vez integran el capital intelectual (IC), son: 0.393, 0.362 y 0.367, respectivamente. Para el caso de los flujos de aprendizaje organizacional hacia adelante (FF) y los flujos de aprendizaje organizacional hacia atrás (FB) los valores desplegados son: 0.392 y 0.375, respectivamente (véase la figura 3.6).

Los valores que presenta la variable compromiso de continuidad (CC), con respecto a las variables stock de conocimiento individual (IK), stock de conocimiento de grupo (GK) y stock de conocimiento organizacional (OK) son: 0.216, 0.198 y 0.264, respectivamente. En las variables que pertenecen a la administración del conocimiento (KM), flujo de aprendizaje hacia adelante (FF) y flujo de aprendizaje hacia atrás (FB), los valores de correlación presentados son: 0.222 y 0.296, respectivamente.

Para la tercera variable independiente que es compromiso normativo (CN), la correlación presentada con respecto a las variables dependientes stock de conocimiento individual (IK), stock de conocimiento de grupo (GK) y stock de conocimiento organizacional (OK) fueron: 0.398, 0.436 y 0.445, respectivamente. Por último tenemos la correlación de CN con las variables dependientes flujo de aprendizaje organizacional hacia adelante (FF) con un valor de 0.369 y flujo de aprendizaje organizacional hacia atrás (FB) reportando un valor de 0.404.

De tal manera que tengamos una visión más clara sobre la baja o alta correlación existente entre las variables, desarrollamos la gráfica que se muestra en la figura 3.7, utilizando la distribución de frecuencias y dividiéndolas en cuatro regiones. En el límite inferior tenemos el valor más bajo de correlación, mientras que en el límite superior tenemos el valor más alto de correlación presentada.

Como podemos observar, los valores de correlación de las variables independientes compromiso afectivo (CA) y compromiso normativo (CN) con sus contrapartes dependientes, son los de más alto valor, el rango de estos valores de correlación ubicados en: 0.362 como mínimo y 0.445 como valor máximo. El valor más alto lo observamos en la correlación de $CN \leftrightarrow OK = 0.445$, el segundo valor más alto lo tenemos en la correlación entre $CN \leftrightarrow GK = 0.436$, la tercera posición es $CN \leftrightarrow FB = 0.404$, la cuarta posición es $CN \leftrightarrow IK = 0.398$, la quinta posición corresponde a $CA \leftrightarrow IK = 0.393$, la sexta posición y última es $CA \leftrightarrow FF = 0.392$. Posteriormente situamos cuatro correlaciones, que aparecen en el siguiente orden, $CA \leftrightarrow FB = 0.375$, $CN \leftrightarrow FF = 0.369$, $CA \leftrightarrow OK = 0.367$ y por último $CA \leftrightarrow GK = 0.362$.

Por otra parte, la correlación de la variable independiente compromiso de continuidad (CC) con sus contrapartes dependientes se ubicó en los valores de más bajo nivel. Se colocaron en el siguiente orden $CC \leftrightarrow FB = 0.296$ y $CC \leftrightarrow OK = 0.264$, mientras que las últimas tres se posicionaron de la siguiente forma: $CC \leftrightarrow FF = 0.222$, $CC \leftrightarrow IK = 0.216$ y por último $CC \leftrightarrow GK = 0.198$.

Como consecuencia del presente análisis, identificamos que sólo las correlaciones de la variable independiente CC con las variables dependientes IK, GK, OK, FF y FB, presentan coeficientes de correlación con valores por debajo del nivel de 0.300. No obstante que se observan con valores de significancia de 0.01 (muy significativo), dentro de los modelos de regresión lineal múltiple que se analizan más adelante la variable CC no presenta coeficientes de impacto. Las correlaciones más altas son las que a su vez tienen valores de coeficiente de impacto con significancia aceptable dentro de los modelos de regresión lineal múltiple, que desarrollamos para dar respuesta a las hipótesis planteadas en esta investigación, el análisis de estos modelos de impacto se verá más adelante.

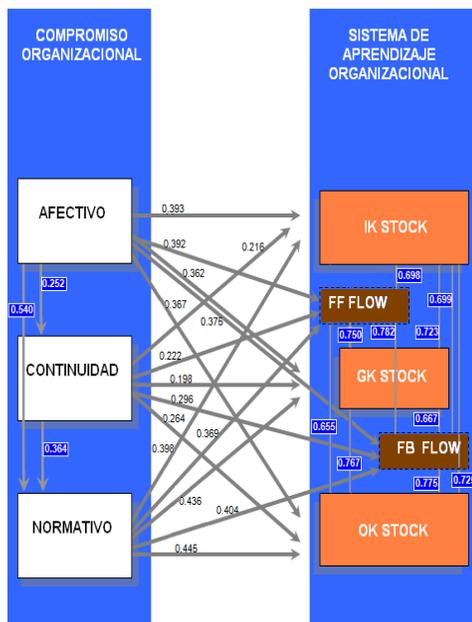


Figura 7.6: Correlación de las variables en el Modelo del Impacto del compromiso organizacional en los Componentes de la administración del conocimiento (Flujos) y capital intelectual (Stocks). Fuente: Elaboración gráfica del Dr. Arturo Reyes Valdez en base a resultados obtenidos del análisis con “SPSS” de datos de investigación directa.

Correlación de las variables del compromiso organizacional con el capital intelectual y la administración del conocimiento

Al igual que en el punto anterior, en éste presentamos de manera gráfica en la figura 7.10, la correlación de las variables en el modelo del impacto que tiene el compromiso organizacional sobre el capital intelectual (IC) y la administración del conocimiento (KM). Las correlaciones presentan valores que van desde 0.254 hasta 0.476, lo cual nos dice que todas tienen una relación lineal positiva, con valor de significancia igual a 0.01 considerándose “Muy significativo”. El análisis correspondiente lo describimos con detalle a continuación.

Los valores de correlación para la variable independiente compromiso afectivo (CA) con respecto a las variables dependientes capital intelectual (IC) y administración del conocimiento (KM) son: 0.418 y 0.407, respectivamente. Por otro lado, tenemos que los valores de correlación que presenta la variable compromiso de continuidad (CC) con respecto a IC y KM, son: 0.254 y 0.274. Por último el valor de la correlación de la variable compromiso normativo (CN) con el capital intelectual (IC) es 0.476 y la correlación de esta variable independiente con su contraparte independiente administración del conocimiento (KM) es de 0.409 (véase la figura 3.9).

Como podemos observar, la correlación más alta corresponde a $CN \leftrightarrow IC = 0.476$ en el siguiente orden descendente a $CA \leftrightarrow IC = 0.418$, $CN \leftrightarrow KM = 0.409$ y $CA \leftrightarrow KM = 0.407$. Posteriormente situamos los coeficientes de correlación $CC \leftrightarrow KM = 0.274$ y $CC \leftrightarrow IC = 0.254$.

Esto es congruente con los resultados encontrados en el punto anterior y a su vez igualmente congruente con los resultados que se obtienen en los modelos de regresión lineal múltiple, donde el coeficiente de CC en los modelos de impacto $IC = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$ y $KM = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$ presentan una significancia con un valor no aceptable.

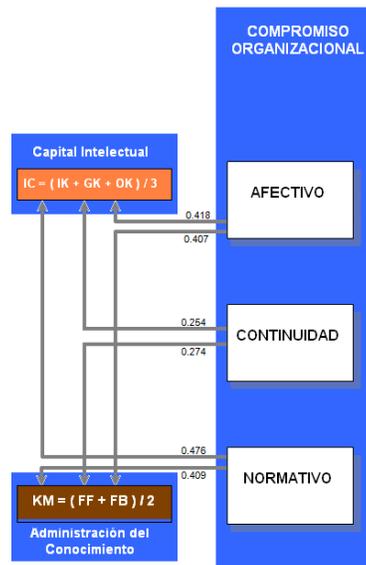


Figura 7.7: Correlación de las variables en el modelo del impacto del compromiso organizacional en la administración del conocimiento y capital intelectual. Fuente: Elaboración gráfica del Lic. Arturo Reyes Valdez en base a resultados obtenidos del análisis con “SPSS” de datos de investigación.

Análisis del modelo de impacto del compromiso organizacional en los componentes del capital intelectual y la administración del conocimiento

El contraste de nuestros primeros cinco grupos de hipótesis planteadas en el punto 3.2.3, se lleva a cabo en este apartado. Basándonos en los resultados de los modelos de regresión lineal múltiple, procedemos a aceptar o rechazar las hipótesis, así como el cumplimiento de los objetivos específicos y encontrar la respuesta al grupo de preguntas planteadas.

Los modelos analizados para el conocimiento estático *stocks* son el impacto del compromiso organizacional en el stock de conocimiento individual $IK = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$, el impacto del compromiso organizacional en el stock de conocimiento de grupo $GK = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$, el impacto del compromiso organizacional en el stock de conocimiento organizacional $OK = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$.

En el caso del conocimiento dinámico *flujos* los modelos son el impacto del compromiso organizacional en los flujos de aprendizaje hacia adelante $FF = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$ y el impacto del compromiso organizacional en los flujos de aprendizaje hacia atrás $FB = \beta_0 + \beta_1 CA + \beta_2 CC + \beta_3 CN$.

La presencia de los coeficientes de impacto de los modelos analizados, tienen diferentes niveles de intensidad. Observamos seis niveles que nos permiten obtener el nivel de intensidad en el que se encuentra cada uno de ellos, estos niveles son: "Imperceptible" con valores en el rango de 0.00 a 0.09, "Perceptible" con valores en el rango de 0.10 a 0.15, "Considerable" con valores en el rango de 0.16 a 0.19, "Importante" con valores en el rango de 0.20 a 0.29, "Fuerte" con valores en el rango de 0.30 a 0.50 y por último el nivel "Muy fuerte" con valores mayores de 0.50.

En conjunto con el nivel de intensidad de los coeficientes de impacto, también se evalúa la significancia estadística en sus diferentes grados, un valor de "t" menor a 1.68 arroja una significancia superior a 0.5 cuyo valor es " + No significativo", para un valor de "t" mayor de 1.68 y menor a 2.33 la significancia tienen un valor de 0.05 tal valor es considerado como " * Significativo", el valor "t" mayor de 2.33 y menor a 3.1 tiene un valor de significancia igual a 0.01 siendo esto " ** Considerablemente significativo" y el último valor de la tabla para "t" es para los mayores a 3.1 con una significancia de 0.001 considerándose " *** Altamente significativo".

De esta forma interpretamos los resultados y por lo tanto, validamos nuestras hipótesis para aceptarlas o rechazarlas.

La figura No. 3.11, muestra gráficamente el valor de los coeficientes de impacto estandarizados de las variables Afectivo y Normativo sobre las variables IK, GK, OK, FF, FB.

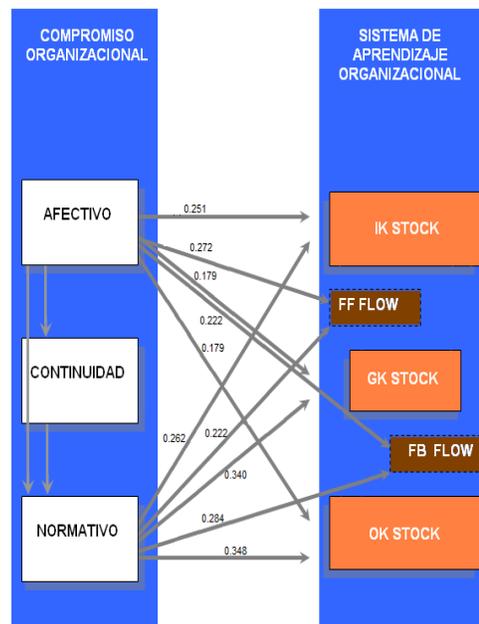


Figura 7.8: Impacto y su significancia estadística del compromiso organizacional en los componentes de la administración del conocimiento flujos y capital intelectual stocks, macro componentes de un sistema de aprendizaje organizacional. Fuente: Elaboración gráfica del Lic. Arturo Reyes Valdez en base a resultados obtenidos del análisis con “SPSS” de datos de investigación.

Conocimiento individual

De acuerdo con el primer grupo de hipótesis, se establece que hay una relación positiva del compromiso organizacional y el stock de conocimiento individual (IK).

Para efectos de probar la misma construimos una ecuación de regresión lineal múltiple $IK=1.633+0.251CA+0.262CN$. El cálculo de regresión que hicimos en el “SPSS”, arroja dos modelos, el primero con un coeficiente de determinación de $R^2=0.158$ y el segundo con un coeficiente de determinación de $R^2=0.203$, donde se observa que el segundo modelo tiene un mayor impacto en la variabilidad de la variable dependiente IK (véase la tabla 7.10).

Para el primer modelo sólo la variable compromiso normativo (CN) presentó un coeficiente de regresión estandarizado con un valor estadísticamente aceptable para la prueba $t=7.032$ con significancia=0.000, en el caso de las variables que fueron excluidas para este primer modelo, donde no observamos coeficientes de regresión fueron las variables compromiso afectivo (CA) y compromiso de continuidad (CC). Ya para el segundo modelo se incrementa el número de variables con parámetros aceptables, cuyo análisis abordaremos con más detalle.

Las variables que presenta el segundo modelo con coeficientes de regresión fueron el compromiso afectivo (CA) y compromiso normativo (CN), únicamente queda fuera de la ecuación final sin coeficiente de regresión la variable de compromiso de continuidad (CC). Los valores de los coeficientes de impacto estandarizados son 0.251 para el compromiso afectivo (CA) con un valor para la prueba $t=3.826$ con significancia=0.000, y 0.262 para el compromiso normativo (CN) con $t=4.000$ y significancia=0.000 (véase la tabla 7.11), los parámetros presentados por estas variables son estadísticamente aceptables. En el caso de la variable compromiso de continuidad (CC) los valores que observamos para el estadístico $t=1.349$ y significancia=0.178 no aceptables. El valor de la constante de regresión $\beta_0=1.633$, nos revela que cuando el resto de las variables de la ecuación tengan un valor igual a cero el stock de conocimiento individual (IK) mantendrá un valor que estará indicado por esta constante. El valor para la prueba "F" para el segundo modelo, también llamada prueba de significancia global es $F=14.640$ con significancia=0.000 que en términos estadísticos son ampliamente aceptables.

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.398 ^a	.158	.155	.639582	.158	49.449	1	263	.000	
2	.450 ^b	.203	.197	.623614	.045	14.640	1	262	.000	1.668

a. Predictors: (Constant), CN

b. Predictors: (Constant), CN, CA

c. Dependent Variable: IK

Tabla 7.10: Sumario de modelos de regresión lineal múltiple para el primer grupo de hipótesis $IK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación.

El efecto de la multicolinealidad no existe, esto es, para el segundo modelo se realizan tres iteraciones presentando en la última iteración un Condition Index= 16.057, valor estadísticamente aceptable.

Una prueba más, con la cual confirmamos que la ecuación de regresión lineal múltiple es válida por haberse comprobado que existe correlación entre las variables dependiente e independientes, la llevamos a cabo con el análisis de la varianza, también citada como ANOVA, ésta alcanzó un valor $F=33.327$ y significancia=0.000, al contrastar con la tabla de valores " $F_{0.01}$ ", para un numerador con grados de libertad = 2 y un denominador con grados de libertad = 262 (véase tabla 7.12), observamos que el valor $F=33.327$ es mayor que el valor crítico $F_{0.01}=4.61$.

Por lo tanto tenemos pruebas suficientes para afirmar que los valores estimados por el modelo econométrico desarrollado son estadísticamente aceptables.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.280	.189		12.053	.000
	CN	.332	.047	.398	7.032	.000
2	(Constant)	1.633	.250		6.526	.000
	CN	.219	.055	.262	4.000	.000
	CA	.273	.071	.251	3.826	.000

a. Dependent Variable: IK

Tabla 7.11: Coeficientes de impacto del modelo de regresión lineal múltiple para el primer grupo de hipótesis $IK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

ANOVA^c

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20.228	1	20.228	49.449	.000 ^a
	Residual	107.584	263	.409		
	Total	127.812	264			
2	Regression	25.921	2	12.961	33.327	.000 ^b
	Residual	101.890	262	.389		
	Total	127.812	264			

a. Predictors: (Constant), CN

b. Predictors: (Constant), CN, CA

c. Dependent Variable: IK

Tabla 7.12: Análisis de varianza ANOVA del modelo de regresión lineal múltiple para el primer grupo de hipótesis $IK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

Es importante considerar que habrá otras variables, que podrían impactar la variable dependiente, pero que no han sido objeto de esta investigación. De acuerdo con los coeficientes de impacto estandarizados, la estimación nos indica que la variable compromiso normativo (CN) es la que tiene mayor influencia sobre el stock de conocimiento individual (IK) con un valor de 0.262 lo que se considera un impacto “Importante” y “Altamente significativo”. La variable compromiso afectivo (CA) es la segunda por su nivel de influencia con un valor de 0.251, asimismo su valor del impacto es igualmente “Importante” y “Altamente significativo”. La estimación no presentó parámetros estadísticamente aceptables para establecer relación alguna entre la variable dependiente y el compromiso de continuidad (CC).

De acuerdo a los resultados obtenidos con el modelo de regresión lineal múltiple $IK=1.633+0.251CA+0.262CN$, el primer objetivo específico de “Identificar y evaluar cómo impacta el compromiso organizacional en el stock de conocimiento individual” fue posible lograrlo. Dejándonos ver que, de los tres elementos que integran el compromiso organizacional, identificamos y evaluamos a los elementos afectivo y normativo como los que tienen un impacto “Importante” y “Altamente significativo” en el stock de conocimiento individual (IK), el segundo de ellos con un mayor impacto que el primero, quedando solamente el elemento de Continuidad sin impacto.

	β	T	Significancia
β_0	1.633	6.526	0.000
CA	0.251	3.826	0.000
CC	-	-	-
CN	0.262	4.000	0.000

Tabla 7.13: Modelo de regresión lineal múltiple para el primer grupo de hipótesis. $IK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$, “ $R^2 = .203$ ”, “ $F=14.640$ ” y “ $Sig=.000$ ”. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

Con esto probamos que de nuestro primer grupo de hipótesis tenemos argumentos válidos para aceptar las hipótesis básicas “H1” y “H3”, así como rechazar la hipótesis “H2”.

Por último, la tabla 7.12 presenta un resumen del modelo analizado, observamos el valor de la constante de regresión, el valor del coeficiente de impacto estandarizado de CA, el valor del coeficiente de impacto estandarizado de CN, todas estas con sus respectivos valores “t” y su significancia. También se presenta el valor del coeficiente de determinación “R²” con su respectivo valor “F” y significancia.

Conocimiento de grupo

De acuerdo al segundo grupo de hipótesis, se establece que hay una relación positiva del compromiso organizacional y el stock de conocimiento de grupo (GK).

Para efectos de probar la misma se construyó una ecuación de regresión lineal múltiple $GK=1.893+0.179CA+0.340CN$. El cálculo de regresión que hicimos en el “SPSS”, arroja dos modelos el primero con un coeficiente de determinación de $R^2=0.190$, el segundo nos dio un coeficiente de determinación de $R^2=0.213$, donde se observa que el segundo modelo tiene un mayor impacto en la variabilidad del stock de conocimiento de grupo (GK) (véase tabla 7.14).

Para el primer modelo sólo la variable compromiso normativo (CN) presentó un coeficiente de regresión estandarizado con un valor estadísticamente aceptable para la prueba $t=7.866$ con significancia=0.000, en el caso de las variables que fueron excluidas para éste primer modelo, donde

no observamos coeficientes de regresión fueron el compromiso afectivo (CA) y el compromiso de continuidad (CC) (véase la tabla 7.14). Ya para el segundo modelo se incrementa el número de variables con parámetros aceptables, veremos el análisis a continuación con más detalle.

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.436 ^a	.190	.187	.566851	.190	61.878	1	263	.000	
2	.462 ^b	.213	.207	.559951	.023	7.522	1	262	.007	1.53

a. Predictors: (Constant), CN

b. Predictors: (Constant), CN, CA

c. Dependent Variable: GK

Tabla 7.14: Sumario de modelos de regresión lineal múltiple para el segundo grupo de hipótesis $GK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación.

Las variables que presenta el segundo modelo con coeficientes de regresión estandarizados fueron el compromiso afectivo (CA) y compromiso normativo (CN), únicamente queda fuera de la ecuación final sin coeficiente de regresión la variable de compromiso de continuidad (CC). Los valores de los coeficientes de impacto estandarizados son 0.179 para el compromiso afectivo (CA) con un valor para la prueba $t=2.743$ con significancia=0.007, y 0.340 para el compromiso normativo (CN) con $t=5.218$ y significancia=0.000 (véase tabla 7.15), los parámetros presentados por éstas variables son estadísticamente aceptables. En el caso de la variable compromiso de continuidad (CC) los valores que observamos son $t=0.210$ y significancia=0.834 no aceptables en base a los criterios definidos. El valor de la constante de regresión $\beta_0=1.893$, cuyo valor tomará el stock de conocimiento de grupo (GK), cuando el resto de las variables de la ecuación tengan un valor igual a cero. El valor para la prueba “F” para el segundo modelo, también llamada prueba de significancia global es $F=7.522$ con significancia=0.007 estadísticos ampliamente aceptables.

El efecto de la multicolinealidad es eliminado, esto es, para el segundo modelo se realizan tres iteraciones presentando en la última iteración un Condition Index= 16.057, valor estadísticamente aceptable.

Para efectos de confirmar que la ecuación de regresión lineal múltiple es válida, realizamos la comprobación de que existe correlación entre las variables dependiente e independientes con la prueba estadística ANOVA (véase tabla 7.16), ésta alcanza un valor $F=35.467$ y significancia=0.000, al contrastar con la tabla de valores “ $F_{0.01}$ ”, para un numerador con grados de libertad = 2 y un denominador con grados de libertad = 262, observamos es mayor que el valor crítico $F_{0.01}=4.61$. Por lo tanto tenemos pruebas suficien-

tes para afirmar que los valores estimados por el modelo econométrico desarrollado son estadísticamente aceptables.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.310	.168		13.778	.000
	CN	.329	.042	.436	7.866	.000
2	(Constant)	1.893	.225		8.428	.000
	CN	.256	.049	.340	5.218	.000
	CA	.176	.064	.179	2.743	.007

a. Dependent Variable: GK

Tabla 7.15: Coeficientes de impacto del modelo de regresión lineal múltiple para el segundo grupo de hipótesis $GK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$.

Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	19.883	1	19.883	61.878	.000 ^a
	Residual	84.507	263	.321		
	Total	104.390	264			
2	Regression	22.241	2	11.121	35.467	.000 ^b
	Residual	82.149	262	.314		
	Total	104.390	264			

a. Predictors: (Constant), CN

b. Predictors: (Constant), CN, CA

c. Dependent Variable: GK

Tabla 7.16: Análisis de varianza ANOVA del modelo de regresión lineal múltiple para el segundo grupo de hipótesis $GK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$.

Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

Los coeficientes de impacto estandarizados de la estimación nos indican que la variable compromiso normativo (CN) es la que tiene mayor influencia sobre el stock de conocimiento de grupo (GK) con un valor de 0.340 lo que se considera un impacto "Fuerte". La variable compromiso afectivo (CA) es la segunda por su nivel de influencia con un valor de 0.179, de igual manera su valor del impacto cae en el rango de "Considerable". La estimación no presentó parámetros estadísticamente aceptables, para establecer relación alguna entre la variable dependiente y el compromiso de continuidad (CC).

En concordancia con los resultados presentados en el modelo de regresión lineal múltiple $GK=1.893+0.179CA+0.340CN$, alcanzamos el segundo objetivo específico de “Identificar y evaluar cómo impacta el compromiso organizacional en el stock de conocimiento de grupo”. Revelando que solo el elemento de continuidad del compromiso organizacional, no tiene impacto, por otro lado el elemento normativo se identifica con un impacto “Fuerte” y “Altamente significativo”, mientras que el elemento afectivo se identifica con un impacto “Considerable” y “Considerablemente significativo”, el primero de ellos con un mayor impacto que el segundo.

De esta manera probamos que de nuestro segundo grupo de hipótesis tenemos argumentos suficientes para aceptar las hipótesis básicas “H4” y “H6” y rechazar “H5”.

Por último, la tabla 7.17 presenta un resumen del modelo analizado, observamos el valor de la constante de regresión, el valor del coeficiente de impacto estandarizado de CA, el valor del coeficiente de impacto estandarizado de CN, todas estas con sus respectivos valores “t” y su significancia. También se presenta el valor del coeficiente de determinación “R²” con su respectivo valor “F” y significancia.

	β	T	Significancia
β_0	1.893	8.428	0.000
CA	0.179	2.743	0.007
CC	-	-	-
CN	0.340	5.218	0.000

Tabla 7.17: Modelo de regresión lineal múltiple para el segundo grupo de hipótesis. $GK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$, “R²= .213”, “F=7.522” y “Sig=.000”. Fuente: Tabla integrado por el Dr. Arturo Reyes Valdez con resultados generados por el software estadísticos SPSS en base a datos de investigación directa.

Conocimiento organizacional

De acuerdo al tercer grupo de hipótesis, se establece que hay una relación positiva entre el compromiso organizacional y el stock de conocimiento organizacional (OK).

La ecuación de regresión lineal múltiple $OK=1.842+0.179CA+0.348CN$ fue desarrollada con la intención de probar el tercer grupo de hipótesis. El cálculo de regresión que hicimos en el “SPSS”, arroja dos modelos, el primero con un coeficiente de determinación de $R^2=0.195$ y el segundo con un coeficiente de determinación de $R^2=0.215$, donde se observa que el segundo modelo presenta un mayor impacto en la variabilidad del stock de conocimiento organizacional (OK) (véase tabla 7.18).

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.445 ^a	.198	.195	.651627	.198	64.843	1	263	.000	
2	.470 ^b	.221	.215	.643541	.023	7.650	1	262	.006	1.552

a. Predictors: (Constant)CN

b. Predictors: (Constant), CN, CA

c. Dependent Variable: OK

Tabla 7.18: Sumario de modelos de regresión lineal múltiple para el tercer grupo de hipótesis $OK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.325	.193		12.063	.000
	CN	.387	.048	.445	8.053	.000
2	(Constant)	1.842	.258		7.135	.000
	CN	.303	.056	.348	5.366	.000
	CA	.203	.074	.179	2.766	.006

a. Dependent Variable: OK

Tabla 7.19: Coeficientes de impacto del modelo de regresión lineal múltiple para el tercer grupo de hipótesis $OK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

En el primer modelo sólo la variable compromiso normativo (CN) presentó un coeficiente de regresión estandarizado con un valor estadísticamente aceptable para la prueba $t=8.053$ con significancia=0.000, en el caso de las variables que fueron excluidas para este primer modelo, donde no observamos coeficientes de regresión fueron las variables compromiso afectivo (CA) y compromiso de continuidad (CC). Ya para el segundo modelo se incrementa el número de variables con parámetros aceptables.

Las variables que presenta el segundo modelo con coeficientes de regresión estandarizados fueron el compromiso afectivo (CA) y compromiso normativo (CN), únicamente queda fuera de la ecuación final sin coeficiente de regresión la variable de compromiso de continuidad (CC). Los valores de los coeficientes de impacto estandarizados son 0.179 para el compromiso afectivo (CA) con un valor para la prueba $t=2.766$ con significancia=0.006, y

0.348 para el compromiso normativo (CN) con $t=5.366$ y significancia=0.000 (véase tabla 7.19), los parámetros presentados por estas variables son estadísticamente aceptables.

ANOVA^c

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	27.534	1	27.534	64.843	.000 ^a
	Residual	111.674	263	.425		
	Total	139.208	264			
2	Regression	30.702	2	15.351	37.067	.000 ^b
	Residual	108.506	262	.414		
	Total	139.208	264			

a. Predictors: (Constant), CN

b. Predictors: (Constant), CN, CA

c. Dependent Variable: OK

Tabla 7.20: Análisis de varianza ANOVA del modelo de regresión lineal múltiple para el tercer grupo de hipótesis $OK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$.

Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

En el caso de la variable compromiso de continuidad (CC) los valores que observamos para el estadístico $t=1.291$ y significancia=0.198 no aceptables en base a los criterios definidos. El valor de la constante de regresión $\beta_0=1.842$, nos revela que cuando el resto de las variables de la ecuación tengan un valor igual a cero el stock de conocimiento organizacional (OK) mantendrá un valor que estará indicado por esta constante. El valor para la prueba “F” para el segundo modelo es $F=7.650$ con significancia=0.006 que en términos estadísticos son ampliamente aceptables.

El efecto de la multicolinealidad no existe, esto es, para el segundo modelo se realizan tres iteraciones presentando en la última iteración un Condition Index= 16.057, valor estadísticamente aceptable.

Una prueba más, con la cual confirmamos que la ecuación de regresión lineal múltiple es válida por haberse comprobado que existe correlación entre las variables dependiente e independientes, la llevamos a cabo con el análisis de la varianza, ésta alcanzó un valor $F=37.067$ y significancia=0.000, al contrastar con la tabla de valores “ $F_{0.01}$ ”, para un numerador con grados de libertad = 2 y un denominador con grados de libertad = 262 (véase tabla 7.20), observamos que el valor $F=37.067$ es mayor que el valor crítico $F_{0.01}=4.61$. Por lo tanto, tenemos pruebas suficientes para afirmar que los valores estimados por el modelo econométrico desarrollado son estadísticamente aceptables.

Los coeficientes de impacto estandarizados de la estimación nos indica que la variable compromiso normativo (CN) es la que tiene mayor in-

fluencia sobre el stock de conocimiento organizacional (OK) con un valor de 0.348 lo que se considera un impacto “Fuerte” y “Altamente significativo”. La variable compromiso afectivo (CA) es la segunda por su nivel de influencia con un valor de 0.179, asimismo su valor del impacto es considerado “Considerable” y “altamente significativo”. La estimación no presentó parámetros estadísticamente aceptables para establecer relación alguna entre la variable dependiente y el compromiso de continuidad (CC).

De acuerdo a los resultados obtenidos con el modelo de regresión lineal múltiple $OK=1.842+0.179CA+0.348CN$, el tercer objetivo específico de “Identificar y evaluar cómo impacta el compromiso organizacional en el stock de conocimiento organizacional” se alcanzó. Dejándonos ver que, de los tres elementos que integran el compromiso organizacional, identificamos y evaluamos al elemento Normativo con un impacto “Fuerte” y “Altamente significativo” en el stock de conocimiento organizacional (OK), mientras que el elemento Afectivo tiene un impacto “Considerable” y también “Altamente significativo”, el primero de ellos con un mayor impacto que el segundo, quedando solamente el elemento de Continuidad sin impacto.

Con esto probamos que de nuestro tercer grupo de hipótesis tenemos argumentos validos para aceptar las hipótesis básicas “H7” y “H9” y rechazar “H8”.

Por último, la tabla 7.21 presenta un resumen del modelo analizado, observamos el valor de la constante de regresión, el valor del coeficiente de impacto estandarizado de CA, el valor del coeficiente de impacto estandarizado de CN, todas estas con sus respectivos valores “t” y su significancia. También se presenta el valor del coeficiente de determinación “R²” con su respectivo valor “F” y significancia.

	β	T	Significancia
β_0	1.842	7.135	0.000
CA	0.179	2.766	0.006
CC	–	–	–
CN	0.348	5.366	0.000

Tabla 7.21: Modelo de regresión lineal múltiple para el tercer grupo de hipótesis. $OK = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$, “R²= .221”, “F=7.650” y “Sig=.006”. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación directa.

Flujo de aprendizaje hacia adelante

De acuerdo al cuarto grupo de hipótesis, se establece que hay una relación positiva entre el compromiso organizacional y el flujo de aprendizaje hacia adelante (FF).

Para efectos de probar la misma construimos una ecuación de regresión lineal múltiple $FF=1.556+0.272CA+0.222CN$. El cálculo de regresión que hicimos en el “SPSS”, arroja dos modelos, el primero con un coeficiente

de determinación de $R^2=0.154$ y el segundo con un coeficiente de determinación de $R^2=0.188$, donde se observa que el segundo modelo tiene un mayor impacto en la variabilidad del flujo de aprendizaje hacia adelante (FF) (véase tabla 7.22).

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.392 ^a	.154	.150	.666940	.154	47.741	1	263	.000	
2	.434 ^b	.188	.182	.654310	.035	11.251	1	262	.001	1.442

- a. Predictors: (Constant), CA
- b. Predictors: (Constant), CA, CN
- c. Dependent Variable: FF

Tabla 7.22: Sumario de modelos de regresión lineal múltiple para el cuarto grupo de hipótesis $FF = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación.

Para el primer modelo sólo la variable compromiso afectivo (CA) presentó un coeficiente de regresión estandarizado con un valor estadísticamente aceptable para la prueba $t=6.909$ con significancia=0.000, en el caso de las variables que fueron excluidas para este primer modelo, donde no observamos coeficientes de regresión fueron las variables compromiso normativo (CN) y compromiso de continuidad (CC). Ya para el segundo modelo se incrementa el número de variables con parámetros aceptables, cuyo análisis abordaremos con más detalle.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.768	.260		6.811	.000
	CA	.443	.064	.392	6.909	.000
2	(Constant)	1.556	.263		5.927	.000
	CA	.308	.075	.272	4.113	.000
	CN	.192	.057	.222	3.354	.001

- a. Dependent Variable: FF

Tabla 7.23: Coeficientes de impacto del modelo de regresión lineal múltiple para el cuarto grupo de hipótesis $FF = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$.

Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

Las variables que presenta el segundo modelo con coeficientes de regresión estandarizados fueron el compromiso afectivo (CA) y compromiso normativo (CN), únicamente queda fuera de la ecuación final sin coeficiente de regresión la variable de compromiso de continuidad (CC). Los valores de los coeficientes de impacto estandarizados son 0.272 para el compromiso afectivo (CA) con un valor para la prueba $t=4.113$ con significancia=0.000, y 0.222 para el compromiso normativo (CN) con $t=3.354$ y significancia=0.001 (véase tabla 7.23), los parámetros presentados por éstas variables son estadísticamente aceptables. En el caso de la variable compromiso de continuidad (CC) los valores que observamos para el estadístico $t=1.038$ y significancia=0.3000 no aceptables en base a los criterios definidos. El valor de la constante de regresión $\beta_0=1.556$, nos revela que cuando el resto de las variables de la ecuación tengan un valor igual a cero el flujo de aprendizaje hacia adelante mantendrá un valor que estará indicado por esta constante. El valor para la prueba "F" del segundo modelo, es $F=11.251$ con significancia=0.001 que en términos estadísticos son ampliamente aceptables.

El efecto de la multicolinealidad no existe, esto es, para el segundo modelo se realizan tres iteraciones presentando en la última iteración un Condition Index= 16.057, valor estadísticamente aceptable.

ANOVA^c

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	21.236	1	21.236	47.741	.000 ^a
	Residual	116.985	263	.445		
	Total	138.220	264			
2	Regression	26.052	2	13.026	30.426	.000 ^b
	Residual	112.168	262	.428		
	Total	138.220	264			

a. Predictors: (Constant), CA

b. Predictors: (Constant), CA, CN

c. Dependent Variable: FF

Tabla 7.24: Análisis de varianza ANOVA del modelo de regresión lineal múltiple para el cuarto grupo de hipótesis $FF = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$.

Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

Una prueba más, con la cual confirmamos que la ecuación de regresión lineal múltiple es válida por haberse comprobado que existe correlación entre las variables dependiente e independientes, la llevamos a cabo con el análisis ANOVA, ésta alcanzó un valor $F=30.426$ y significancia=0.000, al contrastar con la tabla de valores " $F_{0.01}$ ", para un numerador con grados de libertad = 2 y un denominador con grados de libertad = 262 (véase tabla 7.24), observamos que el valor $F=30.426$ es mayor que el valor crítico $F_{0.01}=4.61$. Por lo tanto tenemos pruebas suficientes para afirmar que los valores estimados por el modelo econométrico desarrollado son estadísticamente aceptables.

Los coeficientes de impacto estandarizados de la estimación nos indica que la variable compromiso afectivo (CA) es la que tiene mayor influencia sobre el flujo de aprendizaje hacia adelante (FF) con un valor de 0.272 lo que se considera un impacto "Importante" y "Altamente significativo". La variable compromiso normativo (CN) es la segunda por su nivel de influencia con un valor de 0.222 su valor del impacto es "Importante" y "Altamente significativo". La estimación no presentó parámetros estadísticamente aceptables para establecer relación alguna entre la variable dependiente y el compromiso de continuidad (CC).

De acuerdo a los resultados obtenidos con el modelo de regresión lineal múltiple $FF=1.556+0.272CA+0.222CN$, el cuarto objetivo específico de "Identificar y evaluar cómo impacta el compromiso organizacional en el flujo de aprendizaje hacia adelante" fue posible lograrlo. Dejándonos ver que, de los tres elementos que integran el compromiso organizacional, identificamos y evaluamos al elemento Afectivo como el que tiene un impacto "Importante" y "Altamente significativo" en el flujo de aprendizaje hacia adelante y también identificamos y evaluamos al elemento normativo con un impacto "Importante" y "Altamente significativo", el primero de ellos con un mayor impacto que el segundo, mientras que el elemento de continuidad queda sin impacto.

Con esto probamos el cuarto grupo de hipótesis e inferimos que tenemos los argumentos suficientes para aceptar las hipótesis básicas "H10" y "H12" y rechazar "H11".

Por último, la tabla 7.25 presenta un resumen del modelo analizado, observamos el valor de la constante de regresión, el valor del coeficiente de impacto estandarizado de CA, el valor del coeficiente de impacto estandarizado de CN, todas estas con sus respectivos valores "t" y su significancia. También se presenta el valor del coeficiente de determinación "R²" con su respectivo valor "F" y significancia.

	β	T	Significancia
β_0	1.556	5.927	0.000
CA	0.272	4.113	0.000
CC	-	-	-
CN	0.222	3.354	0.001

Tabla 7.25: Modelo de regresión lineal múltiple para el cuarto grupo de hipótesis. $FF = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$, " $R^2 = .188$ ", " $F=11.251$ " y " $Sig=.001$ ". Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación directa.

Flujo de aprendizaje hacia atrás

De acuerdo al quinto grupo de hipótesis, se establece que hay una relación positiva entre el compromiso organizacional y el flujo de aprendizaje hacia atrás (FB).

Para efectos de probar la misma construimos una ecuación de regresión lineal múltiple $FB=1.743+0.222CA+0.284CN$. El cálculo de regresión que hicimos en el "SPSS", arroja dos modelos, el primero con un coeficiente de determinación de $R^2=0.163$ y el segundo con un coeficiente de determinación de $R^2=0.198$, donde se observa que el segundo modelo tiene un mayor impacto de la variabilidad del flujo de aprendizaje hacia atrás (FB) (véase tabla 7.26).

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.404 ^a	.163	.160	.646790	.163	51.280	1	263	.000	
2	.445 ^b	.198	.192	.634374	.035	11.396	1	262	.001	1.591

a. Predictors: (Constant), CN

b. Predictors: (Constant), CN, CA

c. Dependent Variable: FB

Tabla 7.26: Sumario de modelos de regresión lineal múltiple para el quinto grupo de hipótesis $FB = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación.

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			
1	(Constant)	2.324	.191		12.147	.000
	CN	.341	.048	.404	7.161	.000
2	(Constant)	1.743	.255		6.848	.000
	CN	.240	.056	.284	4.319	.000
	CA	.245	.073	.222	3.376	.001

a. Dependent Variable: FB

Tabla 7.27: Coeficientes de impacto del modelo de regresión lineal múltiple para el quinto grupo de hipótesis $FB = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$.

Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación.

Para el primer modelo sólo la variable compromiso normativo (CN) presentó un coeficiente de regresión estandarizado con un valor estadísticamente aceptable para la prueba $t=7.161$ con significancia=0.000, en el caso de las variables que fueron excluidas para este primer modelo, donde no observamos coeficientes de regresión estandarizados fueron las variables compromiso afectivo (CA) y compromiso de continuidad (CC). Ya para el segundo modelo se incrementa el número de variables con parámetros aceptables; cuyo análisis abordaremos con más detalle.

Las variables que presenta el segundo modelo con coeficientes de regresión estandarizados fueron el compromiso afectivo (CA) y compromiso normativo (CN), únicamente queda fuera de la ecuación final sin coeficiente de regresión la variable de compromiso de continuidad (CC). Los valores de los coeficientes de impacto estandarizados son 0.222 para el compromiso afectivo (CA) con un valor para la prueba $t=3.376$ con significancia=0.001, y 0.284 para el compromiso normativo (CN) con $t=4.319$ y significancia=0.000 (véase tabla 7.27), los parámetros presentados por estas variables son estadísticamente aceptables.

En el caso de la variable compromiso de continuidad (CC) los valores que observamos para el estadístico $t=1.732$ y significancia=0.084 no aceptables en base a los criterios definidos. El valor de la constante de regresión $\beta_0=1.743$, nos revela que cuando el resto de las variables de la ecuación tengan un valor igual a cero el stock de conocimiento individual (IK) mantendrá un valor que estará indicado por esta constante. El valor para la prueba "F" para el segundo modelo es $F=11.396$ con significancia=0.001 que en términos estadísticos son ampliamente aceptables.

ANOVA^c

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	21.452	1	21.452	51.280	.000 ^a
	Residual	110.023	263	.418		
	Total	131.475	264			
2	Regression	26.038	2	13.019	32.351	.000 ^b
	Residual	105.437	262	.402		
	Total	131.475	264			

a. Predictors: (Constant), CN

b. Predictors: (Constant), CN, CA

c. Dependent Variable: FB

Tabla 7.28: Análisis de varianza ANOVA del modelo de regresión lineal múltiple para el quinto grupo de hipótesis $FB = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$.

Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

El efecto de la multicolinealidad no existe, esto es, para el segundo modelo se realizan tres iteraciones presentado en la última iteración un Condition Index=16.057, valor estadísticamente aceptable.

Una prueba más, con la cual confirmamos que la ecuación de regresión lineal múltiple es válida por haberse comprobado que existe correlación entre las variables dependiente e independientes, la llevamos a cabo con el análisis de la varianza, ésta alcanzó un valor $F=32.351$ y significancia=0.000, al contrastar con la tabla de valores " $F_{0.01}$ ", para un numerador con grados de libertad = 2 y un denominador con grados de libertad = 262 (véase tabla 7.28), observamos que el valor $F=32.351$ es mayor que el valor crítico $F_{0.01}=4.61$. Por lo tanto tenemos pruebas suficientes para afirmar que los valores estimados por el modelo econométrico desarrollado son estadísticamente aceptables.

Los coeficientes de impacto estandarizados de la estimación nos indica que la variable compromiso normativo (CN) es la que tiene mayor influencia sobre el flujo de aprendizaje hacia atrás (FB) con un valor de 0.284 lo que se considera un impacto "Importante" y "Altamente significativo". La variable compromiso afectivo (CA) es la segunda por su nivel de influencia con un valor de 0.222 su valor del impacto es igualmente "Importante" y "Altamente significativo". La estimación no presentó parámetros estadísticamente aceptables para establecer relación alguna entre la variable dependiente y el compromiso de continuidad (CC).

	β	T	Significancia
β_0	1.743	6.848	0.000
CA	0.222	3.376	0.001
CC	-	-	-
CN	0.284	4.319	0.000

Tabla 7.29: Modelo de regresión lineal múltiple para el quinto grupo de hipótesis. $FB = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$, “ $R^2 = .198$ ”, “ $F = 11.396$ ” y “ $Sig = .001$ ”. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación directa.

De acuerdo con los resultados obtenidos con el modelos de regresión lineal múltiple $FB = 1.743 + 0.222CA + 0.284CN$, el quinto objetivo específico de “Identificar y evaluar cómo impacta el compromiso organizacional en el flujo de aprendizaje hacia atrás” fue posible lograrlo. Dejándonos ver que, de los tres elementos que integran el compromiso organizacional, identificamos y evaluamos a los elementos afectivo y normativo como los que tienen un impacto “Importante” y “Altamente significativo” en el flujo de aprendizaje hacia atrás, el segundo de ellos con un mayor impacto que el primero, quedando solamente el elemento de continuidad sin impacto.

Con esto probamos el quinto grupo de hipótesis e inferimos que tenemos los argumentos suficientes para aceptar las hipótesis básicas “H13” y “H15” y rechazar “H14”.

Por último, la tabla 7.29 presenta un resumen del modelo analizado, observamos el valor de la constante de regresión, el valor del coeficiente de impacto estandarizado de CA, el valor del coeficiente de impacto estandarizado de CN, todas estas con sus respectivos valores de “t”, y sus significancia. También se presenta el valor del coeficiente de determinación “ R^2 ” con su respectivo valor “F” y sus significancia.

Análisis del modelo de impacto de compromiso organizacional en el capital intelectual y la administración del conocimiento

Para analizar y comprobar las hipótesis contenidas en el sexto y séptimo grupo, desarrollamos este apartado. Apoyándonos en los resultados de los modelos de regresión lineal múltiple, aceptamos o rechazamos las hipótesis. Los últimos objetivos específicos, el sexto y séptimo fueron alcanzados.

Por un lado analizamos el modelo de impacto para el conocimiento estático, el capital intelectual (véase la ecuación 12) $IC = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Por otro lado analizamos el modelo de impacto para el conocimiento dinámico, la administración del conocimiento (véase la ecuación 13) $KM = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$.

La presencia de los coeficientes de impacto estandarizados de los modelos analizados, tienen diferentes niveles de intensidad, al analizar estos niveles, nos permite saber cuál o cuáles de las tres variables que integran el compromiso organizacional tiene mayor influencia sobre el capital intelectual (IC) y la administración del conocimiento (KM).

Asimismo, a la par con el nivel de intensidad de los coeficientes de impacto estandarizados, también se evalúa la significancia estadística en sus diferentes grados, que van desde “+ No significativo” hasta un “*** Altamente significativo”.

De esta forma interpretamos los resultados y por lo tanto, validamos nuestras hipótesis para aceptarlas o rechazarlas.

Ecuación 12: Capital intelectual es el promedio de los tres niveles de conocimiento que existen en una organización.

$$IC = (IK + GK + OK) / 3 \quad IC = \text{Capital Intelectual}$$

Ecuación 13: Administración del conocimiento es el promedio de los flujos de entrada y salida de los tres niveles de conocimiento de la organización.

$$KM = (FF + FB) / 2 \quad KM = \text{Administración del Conocimiento}$$

Capital intelectual

De acuerdo al sexto grupo de hipótesis se establece que hay una relación positiva del compromiso organizacional y el capital intelectual.

Para efectos de probar la misma se estableció una ecuación de regresión lineal múltiple $IC = 1.716 + 0.237CA + 0.344CN$. El cálculo de regresión que hicimos en el “SPSS”, arroja dos modelos, el primero con un coeficiente de determinación de $R^2 = 0.223$ y el segundo con un coeficiente de determinación de $R^2 = 0.262$, donde se observa que el segundo modelo tiene un mayor impacto en la variabilidad del capital intelectual (IC) (véase tabla 7.30).

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.472 _a	.223	.220	.560933	.223	75.299	1	263	.000	
2	.512 _b	.262	.257	.547443	.040	14.121	1	262	.000	1.527

- a. Predictors: (Constant), CN
- b. Predictors: (Constant), CN, CA
- c. Dependent Variable: IC

Tabla 7.30: Sumario de modelos de regresión lineal múltiple para el sexto grupo de hipótesis $IC = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.274	.166		13.709	.000
	CN	.359	.041	.472	8.677	.000
2	(Constant)	1.716	.220		7.815	.000
	CN	.261	.048	.344	5.450	.000
	CA	.235	.063	.237	3.758	.000

- a. Dependent Variable: IC

Tabla 7.31: Coeficientes de impacto del modelo de regresión lineal múltiple para el sexto grupo de hipótesis $IC = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación.

Para el primer modelo sólo la variable compromiso normativo (CN) presentó un coeficiente de regresión estandarizado con un valor estadísticamente aceptable para la prueba $t=8.677$ con significancia=0.000, en el caso de las variables que fueron excluidas para este primer modelo. Donde no observamos coeficientes de regresión fueron el compromiso afectivo (CA) y el compromiso de continuidad (CC). Ya para el segundo modelo se incrementa el número de variables con parámetros aceptables, cuyo análisis abordaremos con más detalle.

Las variables que presenta el segundo modelo con coeficientes de regresión estandarizados fueron el compromiso afectivo (CA) y compromiso

normativo (CN), únicamente queda fuera de la ecuación final sin coeficiente de regresión estandarizado, la variable de compromiso de continuidad (CC). Los valores de los coeficientes de impacto estandarizados son 0.237 para el compromiso afectivo (CA) con un valor para la prueba $t=3.758$ con significancia=0.000, y 0.344 para el compromiso normativo (CN) con $t=5.450$ y significancia=0.000 (véase tabla 7.31), los parámetros presentados por estas variables son estadísticamente aceptables. En el caso de la variable compromiso de continuidad (CC) los valores que observamos son $t=1.295$ y significancia=0.197 no aceptables en base a los criterios definidos. El valor de la constante de regresión $\beta_0=1.716$, nos revela que cuando el resto de las variables de la ecuación tengan un valor igual a cero el stock de conocimiento individual (IK) mantendrá un valor que estará indicado por esta constante. El valor para la prueba "F" para el segundo modelo es $F=14.121$ con significancia=0.000 que en términos estadísticos son ampliamente aceptables.

El efecto de la multicolinealidad es eliminado, esto es, para el segundo modelo se realizan tres iteraciones presentando en la última iteración un Condition Index= 16.057, valor estadísticamente aceptable.

ANOVA^c

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	23.692	1	23.692	75.299	.000 ^a
	Residual	82.752	263	.315		
	Total	106.444	264			
2	Regression	27.924	2	13.962	46.588	.000 ^b
	Residual	78.520	262	.300		
	Total	106.444	264			

a. Predictors: (Constant), CN

b. Predictors: (Constant), CN, CA

c. Dependent Variable: IC

Tabla 7.32: Análisis de varianza ANOVA del modelo de regresión lineal múltiple para el sexto grupo de hipótesis $IC = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$.

Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

Una prueba más, con la cual confirmamos que la ecuación de regresión lineal múltiple es válida por haberse comprobado que existe correlación entre las variables dependiente e independientes, la llevamos a cabo con el análisis de la varianza, ésta alcanzó un valor $F=46.588$ y significancia=0.000, al contrastar con la tabla de valores "F_{0.01}", para un numerador con grados de libertad = 2 y un denominador con grados de libertad = 262 (véase tabla 7.32), observamos que el valor $F=46.588$ es mayor que el valor crítico $F_{0.01}=4.61$. Por lo tanto tenemos pruebas suficientes para afirmar que

los valores estimados por el modelo econométrico desarrollado son estadísticamente aceptables.

Los coeficientes de impacto estandarizados del modelo, nos indica que la variable compromiso normativo (CN) es la que tiene mayor influencia sobre el capital intelectual (IC) con un valor de 0.344 lo que se considera un impacto “Fuerte” y “Altamente significativo”. La variable compromiso afectivo (CA) es la segunda por su nivel de influencia con un valor de 0.237 su valor del impacto es “Importante” y “Altamente significativo”. La estimación no presentó parámetros estadísticamente aceptables para establecer relación alguna entre la variable dependiente y el compromiso de continuidad (CC).

En congruencia con los resultados obtenidos con el modelo de regresión lineal múltiple $IC=1.716+0.237CA+0.344CN$, el sexto objetivo específico de “Identificar y evaluar cómo impacta el compromiso organizacional en el capital intelectual” fue posible conseguirlo. Dejándonos ver que, de los tres elementos que integran el compromiso organizacional, identificamos y evaluamos a los elementos afectivo y normativo como los que tienen un impacto “Importante” para el primero y “Fuerte” para el segundo, la significancia resulto ser “Altamente significativa” para los dos en el capital intelectual (IC), el segundo de ellos con un mayor impacto que el primero, restando únicamente el compromiso de continuidad (CC) sin impacto.

De esta manera contrastamos nuestro sexto grupo de hipótesis, por lo tanto asumimos que nuestros argumentos son validos para aceptar las hipótesis básicas “H16” y “H18”, asimismo rechazar “H17”.

Por último, la tabla 7.33 presenta un resumen del modelo analizado, observamos el valor de la constante de regresión, el valor del coeficiente de impacto estandarizado de CA, el valor del coeficiente de impacto estandarizado de CN, todas estas con sus respectivos valores “t” y su significancia. También se presenta el valor del coeficiente de determinación “R²” con su respectivo valor “F” y significancia.

	β	T	Significancia
β_0	1.716	7.815	0.000
CA	0.237	3.758	0.000
CC	-	-	-
CN	0.344	5.450	0.000

Tabla 7.33: Modelo de regresión lineal múltiple para el sexto grupo de hipótesis. $IC = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$, “R²= 0.262”, “F=14.121” y “Sig=0.000”. Fuente: Tabla integrado por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación directa.

De acuerdo al séptimo grupo de hipótesis, se establece que hay una relación positiva del compromiso organizacional y la administración del conocimiento.

Para efectos de probar la misma se estableció una ecuación de regresión lineal múltiple $KM=1.649+0.262CA+0.268CN$. El cálculo de regresión que hicimos en el "SPSS", arroja dos modelos, el primero con un coeficiente de determinación de $R^2=0.167$ y el segundo con un coeficiente de determinación de $R^2=0.216$, donde se observa que el segundo modelo presenta una mayor variabilidad sobre la administración del conocimiento (KM) (véase tabla 7.34).

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.409 _a	.167	.164	.616610	.167	52.890	1	263	.000	
2	.465 _b	.216	.210	.599478	.049	16.247	1	262	.000	1.476

a. Predictors: (Constant), CN

b. Predictors: (Constant), CN, CA

c. Dependent Variable: KM

Tabla 7.34: Sumario de modelos de regresión lineal múltiple para el séptimo grupo de hipótesis $KM = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.304	.182		12.637	.000
	CN	.331	.045	.409	7.273	.000
2	(Constant)	1.649	.241		6.858	.000
	CN	.216	.053	.268	4.116	.000
	CA	.276	.069	.262	4.031	.000

a. Dependent Variable: KM

Tabla 7.35: Coeficientes de impacto del modelo de regresión lineal múltiple para el séptimo grupo de hipótesis $KM = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación.

Para el primer modelo sólo la variable compromiso normativo (CN) presentó un coeficiente de regresión estandarizado con un valor estadísticamente aceptable para la prueba $t=7.637$ con significancia=0.000, en el caso de las variables que fueron excluidas para este primer modelo, donde no observamos coeficientes de regresión estandarizados fueron el compromiso afectivo (CA) y el compromiso de continuidad (CC). Ya para el segundo modelo se incrementa el número de variables con parámetros aceptables; cuyo análisis abordaremos con más detalle.

Las variables que presenta el segundo modelo con coeficientes de regresión estandarizados fueron el compromiso afectivo (CA) y compromiso normativo (CN), únicamente queda fuera de la ecuación final sin coeficiente de regresión la variable de compromiso de continuidad (CC). Los valores de los coeficientes de impacto estandarizados son 0.262 para el compromiso afectivo (CA) con un valor para la prueba $t=4.031$ con significancia=0.000, y 0.268 para el compromiso normativo (CN) con $t=4.116$ y significancia=0.000 (véase tabla 7.35), los parámetros presentados por estas variables son estadísticamente aceptables. En el caso de la variable compromiso de continuidad (CC) los valores que observamos son $t=1.483$ y significancia=0.139 no aceptables en base a los criterios definidos. El valor de la constante de regresión $\beta_0=1.649$, nos revela que cuando el resto de las variables de la ecuación tengan un valor igual a cero el stock de conocimiento individual (IK) mantendrá un valor que estará indicado por esta constante. El valor para la prueba "F" para el segundo modelo es $F=16.247$ con significancia=0.000 en términos estadísticos son ampliamente aceptables.

ANOVA^c

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20.109	1	20.109	52.890	.000 ^a
	Residual	99.995	263	.380		
	Total	120.104	264			
2	Regression	25.948	2	12.974	36.102	.000 ^b
	Residual	94.156	262	.359		
	Total	120.104	264			

a. Predictors: (Constant), CN

b. Predictors: (Constant), CN, CA

c. Dependent Variable: KM

Tabla 7.36: Análisis de varianza ANOVA del modelo de regresión lineal múltiple para el séptimo grupo de hipótesis $KM = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$.
Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software SPSS en base a datos de investigación.

El efecto de la multicolinealidad es eliminado, esto es, para el segundo modelo se realizan tres iteraciones presentando en la última iteración un Condition Index= 16.057, valor estadísticamente aceptable.

Una prueba más, con la cual confirmamos que la ecuación de regresión lineal múltiple es válida por haberse comprobado que existe correlación entre las variables dependiente e independientes, la llevamos a cabo con el análisis de la varianza, ésta alcanzó un valor $F=36.102$ y significancia=0.000, al contrastar con la tabla de valores “ $F_{0.01}$ ”, para un numerador con grados de libertad = 2 y un denominador con grados de libertad = 262 (véase tabla 7.36), observamos que el valor $F=36.102$ es mayor que el valor crítico $F_{0.01}=4.61$. Por lo tanto tenemos pruebas suficientes para afirmar que los valores estimados por el modelo econométrico desarrollado son estadísticamente aceptables.

Los coeficientes de impacto estandarizado de la estimación nos indican que la variable compromiso normativo (CN) es la que tiene mayor influencia sobre la administración del conocimiento (KM) con un valor de 0.268 lo que se considera un impacto “Importante” y “Altamente significativo”. La variable compromiso afectivo (CA) es la segunda por su nivel de influencia con un valor de 0.262 su valor del impacto es igualmente “Importante” y “Altamente significativo”. La estimación no presentó parámetros estadísticamente aceptables para establecer relación alguna entre la variable dependiente y el compromiso de continuidad (CC).

De acuerdo a los resultados obtenidos con el modelo de regresión lineal múltiple $KM=1.649+0.262CA+0.268CN$, el séptimo objetivo específico de “Identificar y evaluar cómo impacta el compromiso organizacional en la administración del conocimiento” fue posible conseguirlo. Dejándonos ver que, de los tres elementos que integran el compromiso organizacional, identificamos y evaluamos a los elementos afectivo y normativo como los que tienen un impacto “Importante” y “Altamente significativo” en la administración del conocimiento (KM), el segundo de ellos con un impacto ligeramente mayor que el primero, restando únicamente el compromiso de continuidad (CC) sin impacto.

	β	T	Significancia
β_0	1.649	6.858	0.000
CA	0.262	4.031	0.000
CC	-	-	
CN	0.268	4.116	0.000

Tabla 7.37: Modelo de regresión lineal múltiple para el séptimo grupo de hipótesis. $KM = \beta_0 + \beta_1CA + \beta_2CC + \beta_3CN$, “ $R^2= 0.216$ ”, “ $F=16.247$ ” y “ $Sig=0.000$ ”. Fuente: Tabla integrada por el Dr. Arturo Reyes Valdez con resultados generados por el software estadístico SPSS en base a datos de investigación directa.

De esta manera contrastamos nuestro séptimo grupo de hipótesis, por lo tanto, asumimos que nuestros argumentos son validos para aceptar las hipótesis básicas “H19” y “H21”, así como rechazar “H20”.

Por último, la tabla 7.37 presenta un resumen del modelo analizado, observamos el valor de la constante de regresión, el valor del coeficiente de impacto estandarizado de CA, el valor del coeficiente de impacto estandarizado de CN, todas estas con sus respectivos valores "t" y su significancia. También se presenta el valor del coeficiente de determinación "R²" con su respectivo valor "F" y significancia.

Conclusiones y líneas de investigación futura

El estudio de la relación que tiene el compromiso organizacional con el capital intelectual y la administración del conocimiento, fue el tema central de la presente investigación. Para ello se desarrolló un modelo hipotético en base a la revisión de la literatura. Dicho modelo planteó de manera gráfica las relaciones de impacto que ejercen los elementos afectivo, de continuidad y normativo del modelo de compromiso organizacional desarrollado por Allen & Meyer (1990) sobre el stock de conocimiento individual, stock de conocimiento de grupo y stock de conocimiento organizacional, que en conjunto integran el capital intelectual asimismo, se analizó el impacto del compromiso organizacional sobre el flujo de aprendizaje hacia adelante y el flujo de aprendizaje hacia atrás, componentes de la administración del conocimiento.

Se ha comprobado mediante el análisis de regresión lineal múltiple, que los coeficientes de impacto de los modelos matemáticos utilizados son estadísticamente importantes y significativos, con coeficientes de determinación "R²" que van desde 18.8% a 26.2%. El rango de los coeficientes de determinación "R²" obtenidos en la investigación es congruente con los reportados por otros investigadores, es decir, Swailes (2000) reporta un coeficiente de determinación de R²=12.9% en el impacto del compromiso organizacional sobre la Innovación, Malhotra & Mukherjee (2004) demostró en su investigación el impacto que tiene el compromiso organizacional sobre la Satisfacción en el Trabajo con un coeficiente de determinación de R²=14.5%.

Otro resultado corroborado es que, de entre los tres elementos que integran el compromiso organizacional, la correlación más fuerte y significativa se encuentra en el compromiso afectivo y el compromiso normativo, mientras que la correlación más débil se presenta en el compromiso afectivo y el compromiso de continuidad, e igualmente como el caso de los coeficientes de determinación estos valores de correlación son congruentes con los de otros investigadores como Meyer & Allen (1997, p. 24), Malhotra & Mukherjee (2004), Swailes (2000).

Adicionalmente se comprobó que el elemento compromiso de continuidad permanece sin impacto significativo en todos los modelos económicos analizados en esta investigación e igualmente queda sin impacto significativo en las de otros investigadores como Meyer & Allen (1997, p. 24), Malhotra & Mukherjee (2004), Swailes (2000).

Consideramos que las personas con alto nivel de compromiso afectivo (CA) y compromiso normativo (CN) tienen el enorme potencial de crear/generar, apalancar/almacenar, distribuir/compartir y usar conocimiento en beneficio de la organización haciendo crecer el capital intelectual (IC) y la administración del conocimiento (KM). Cada persona comprometida aporta su experiencia, educación, valores, habilidades y agilidad mental para apoyar el incremento de capital intelectual (IC). También animan y motivan para que el conocimiento pueda fluir más dinámica y ágilmente resultando en una mejor administración del conocimiento (KM) dentro de la organización.

Restando solamente las conclusiones, recomendaciones y posibles líneas de investigación del presente estudio en base a los resultados obtenidos del análisis y discusión de estos.

Conclusiones de la investigación

Esta investigación a permitido comprobar que el compromiso organizacional tiene un impacto importante y altamente significativo en los componentes del capital intelectual (IC) y la administración del conocimiento (KM).

En el caso de las 21 hipótesis planteadas en los siete grupos, catorce de ellas fueron aceptadas, estas son: H4 y H7 tuvieron nivel de impacto "Considerable", H1, H3, H10, H12, H13, H15, H16, H19, H21 tuvieron nivel de impacto "Importante" y por último H6, H9 y H18 tuvieron nivel de impacto "Fuerte", todas ellas alcanzaron nivel de significancia "Altamente significativo". Las hipótesis rechazadas sumaron siete, de las cuales sobresale el hecho de participar siempre la variable independiente compromiso de continuidad, refiriéndose ésta a la atadura económica del individuo hacia la organización.

Podemos obtener las siguientes conclusiones de la validación empírica realizada a la muestra obtenida.

En lo referente al conocimiento estático, es decir los stocks de conocimiento y el capital intelectual:

Las organizaciones con alto nivel de compromiso afectivo (CA) y compromiso normativo (CN), tienen altos niveles de stock de conocimiento individual (IK), stock de conocimiento de grupo (GK), stock de conocimiento organizacional (OK) y capital intelectual (IC).

Las organizaciones no tienen impacto significativo del compromiso de continuidad (CC) en el stock de conocimiento individual (IK), stock de conocimiento de grupo (GK), stock de conocimiento organizacional (OK) y su Capital intelectual (IC).

Para lo que se refiere al conocimiento dinámico, es decir los flujos de aprendizaje y la administración del conocimiento.

Las organizaciones con alto nivel de compromiso afectivo (CA) y compromiso normativo (CN), tienen altos niveles de flujo de aprendizaje

hacia adelante (FF), flujo de aprendizaje hacia atrás (FB) y administración del conocimiento (KM).

Las organizaciones no tienen impacto significativo del compromiso de continuidad (CC) en el flujo de aprendizaje hacia adelante (FF), flujo de aprendizaje hacia atrás (FB) y su administración del conocimiento (KM).

En lo general:

Concluimos que el motor de los incentivos económicos para generar, compartir y utilizar el conocimiento dentro de las organizaciones, carece de fuerza en la consecución de estas metas. Lo más importante es el sentimiento afectivo y la obligación moral de las personas hacia la organización.

LINEAS DE INVESTIGACION

El compromiso organizacional puede ser considerado multidimensional, nos dicen Meyer & Allen (1997) en su investigación, donde nos revelan que tenemos, además de los niveles identificados por Reichers (1985) y Becker (1992), un concepto de enfoque multidimensional.

Este concepto de enfoque multidimensional, identificado por Meyer y Allen (1997), lo representan gráficamente como una matriz, donde en el vector del eje horizontal encontramos las variables del modelo desarrollado por ellos, compromiso afectivo (CA), compromiso de continuidad (CC) y el compromiso normativo (CN), por otro lado, en el vector correspondiente al eje vertical localizamos el compromiso organizacional con los enfoques aportados por Reichers (1985) y Becker (1992) y estos son: el Organizacional, Alta gerencia, Individuo, Grupo y Supervisor. Dentro de esta multidimensionalidad del Compromiso, sería interesante estudiar los impactos en el capital intelectual y la administración del conocimiento.

Desde el punto de vista del Liderazgo en las organizaciones, también resulta importante evaluar la influencia que tiene el líder en el desarrollo del capital intelectual y la administración del conocimiento.

Una línea más para un futuro estudio, sería analizar con mayor especificidad el comportamiento en el sector de manufactura y el sector de servicios, de tal modo que sea posible contrastar los resultados para identificar las potenciales diferencias entre ellos.

Bibliografía

Allen, N. & Meyer, J. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization, *Journal of Occupational Psychology*, Vol. 63, p. 1-18.

Becker, T. (1992). Foci and bases of commitment are the distinctions worth making?, *Academy of Management Journal*, Vol. 35, No. 1, p. 232-244.

Bontis, N. (2000). Managing and Organizational Learning System by Aligning Stocks and Flows of Knowledge: An empirical examination of inte-

lectual capital, knowledge management and business performance, Ph.D. Dissertation, London, ON: Ivey school of business.

Buchanan, B. (1974). Building Organizational Commitment: The socialization of managers in work organizations. *Administrative Science Quarterly*, 19, 533-546.

Crossan, M., Lane, H. & White, R. (1999). An organizational learning framework: From intuition to institution, *The Academy of Management Review*, Vol. 24, No. 3, p. 522.

Davenport, H. & Prusak, L. (1998). *Working Knowledge: How organizations manage What they know*, USA: Harvard Business School.

Duhachek, A., Coughlan, A., Iacobucci, D.(2005). Results on the Standard error of the Coefficient Alpha Index of Reability, *Marketing Science*, Vol. 24, No. 2, p. 294.

Drucker, P. (2002). *El management del futuro*, Argentina: Editorial Sudamericana.

Edvinsson, L. & Malone, S. (1997). *El capital intelectual: Como identificar y calcular el valor inexplorado de los recursos intangibles de su empresa*, COLOMBIA: Norma, S.A.

Edvinsson, L. (2002). *Corporate Longitude: What you need to know to navigate the knowledge economy*, GREAT BRITAIN: Prentice Hall, Financial times.

Fiol, C. & Lyles. M. (1985). Organizational Learning, *Academy of Management Review*, 10, 4, p. 803.

Garvin, D. (1993). Building a Learning Organization, *Harvard Business Review*, July-Aug, p. 77-91.

Hall, D., Schneider, B. & Nygren, H. (1970). Personal factors in organizational identification. *Administrative Science Quarterly*, 15, 176-190.

Hernandez, R., Fernandez, C. & Baptista, P. (2003). *Metodología de la investigación*, México: McGraw-Hill Interamericana.

Hrebiniak, L. & Alutto, J. (1972). Personal and Role-Related Factors in The Development of organizational Commitment, *Administrative Science Quarterly*, Ithaca: Dec 1972. Tomo 17, No. 4; p. 555.

Kaneshiro, P. (2008). *Analizing the Organizational Justice, Trust and Commitment Relationship in a Public Organization*, Ph. D. Dissertation, Prescott Valley, Arizona, USA: Graduate Faculty of the Department of Business and Technology Management, Northcentral University.

Kanter, R. M. (1968). Commitment and Social Organization: A study of commitment mechanisms in utopian communities, *American Sociological Review*, Vol. 33, No. 4, p. 499-517.

Kelly, K. (1997). *New Rules for the New Economy, Twelve dependable principles for thriving in a turbulent world*, *Wired*, Sep.

Liou, K. & Nihan, R. (1994). Dimensions of Organizational Commitment in the Public Sector: An Empirical Assessment, *Public Administration Quarterly*, Spring, p. 99.

Malhotra, N. & mukherjee, A. (2004). The relative influence of organizational commitment and job satisfaction on service quality of customer contact employees in banking call centres. *The Journal of Service Marketing*, 18, 2/3, p. 162-174.

Marsh, R. & Mannari, H. (1977). Organizational Commitment and Turnover: A predictive study. *Administrative Science Quarterly*, Vol. 22, p. 57-75.

Marquardt, J. (1996). *Building the learning Organization: A System Approach to Quantum Improvement and Global Success*, USA: McGraw-Hill.

Meyer, J. & Allen, N. (1997). *Commitment in the Workplace: Theory, Research, and Application*, USA: Sage Publications Inc.

Mowday, R. (1998). Reflections on the study and relevance of organizational commitment, *Human Resource Management Review*, Vol. 8, No. 4, p. 387-401.

Mowday, R., Porter, L. & Steers, R. (1982). *Organizational linkages: The psychology of commitment, absenteeism, and turnover*. San Diego, CA: Academy Press.

Meyer, J. & Allen, N. (1997). *Commitment in the Workplace: Theory, Research, and Application*, USA: Sage Publication Inc.

Navasimayam, K. & Denizci, B. (2006). Human Capital in Service Organizations: Identifying Value Drives, *Journal of Intellectual Capital*, Vol. 7, No. 3, p. 381-393.

Nonaka, I. & Takeuchi, H. (1995). *La Organización Creadora de conocimiento: Cómo las compañías japonesas crean la dinámica de la innovación*, USA: Oxford University Press.

OECD (2005). *Annual Report 2005 45 th Anniversary* Paris, France.

Patalano, C. (2008). *A Study of the Relationship Between Generation Group Identification and Organizational Commitment: Generation X vs Generation Y*. Ph.D. Dissertation: H. Wayne Huizenga school of business and Entrepreneurship Nova Southeastern University.

Peterson, R. (1994). A Meta-analysis of Cronbach's coefficient alpha, *Journal of Consumer Research*, Vol. 21, September, p. 381.

Prusak, L. (1997). *Knowledge in Organizations*, USA: Butterworth-Heinemann.

Rashid, Z. & Sambasivan, M. & Johari, J. (2003). The influence of corporate culture and organizational commitment on performance, *The Journal of Management Development*, Vol. 22, No. 7, p. 708-728.

Reichers, A. (1985). A Review and Reconceptualization of Organizational Commitment, *Academy of Management Reviews*, Vol. 10, No. 3, p. 465-476.

Roos, J. & Roos, G. & Dragonetti, C. & Edvinsson, L. (2001). *Capital intelectual: El valor intangible de la empresa*, ESPAÑA: Paidós Ibérica, S.A.

Sheldon, M. (1971). Investments and involvements as mechanisms producing commitment to the organization, *Administrative Science Quarterly*, Vol. 16, p. 143-150.

Stewart, T. (1997). *The Wealth of Knowledge: Intellectual Capital and the Twenty-First-Century Organization*, USA: Random House, Inc.

Swales, S. (2000). Goals, creativity and achievement: commitment in contemporary organizations, *Blackwell Publishers Ltd 2000*, Vol. 9, No. 3, p. 185.

Tissen, R. & Andriessen, D & Lekanne, F. (2000). *El Valor del conocimiento: Para aumentar el rendimiento en las empresas*, ESPAÑA: Penitence Hall, Financial Times.

Ulrich, D. (1998). Intellectual Capital = Competence x Commitment, *Sloan Management Review*, Winter, p. 15-26.

Weiner, Y. (1982). Commitment Organizations: A normative view. *Academy of Management Review* Vol. 7, p. 418-428.

Weiner, Y. & Gechman, A. (1977). Commitment: A behavioral approach to job involvement. *Journal Vocational Behavior*, Vol. 10, p. 47-52.

Wiig, M. (1995). *Knowledge Management Methods: Practical Approaches to Managing Knowledge*, USA: Schema Press, LTD.

Zapata, L. (2004). Los determinantes de la generación y transferencia del conocimiento en pequeñas y medianas empresas del sector de las tecnologías de la información de Barcelona, Tesis Doctoral, Barcelona, ES, Universidad Autónoma de Barcelona.

Índice del capítulo

(group knowledge /gk), 309

(individual knowledge), 308

activos intangibles, 302

administración del

conocimiento, 313

affective commitment / ca), 318

aprendizaje organizacional,
305

capital intelectual, 312

capital relacional, 309

compromiso afectivo, 318

compromiso organizacional, 303
compromiso organizacional, 316
economía del conocimiento, 300
exploración, 307
exteriorización, 310
flow backward/fb, 311
flow foreward/ff, 310
flujo de aprendizaje hacia adelante, 310
flujo de aprendizaje hacia atrás, 311
flujos de aprendizaje, 310
intellectual capital /ic, 312
knowledge management/km, 313
organizational knowledge /ok, 309
organizational commitment / oc, 316
proceso de internalización, 312
stock de conocimiento de grupo, 309
stock de conocimiento individual, 308
stocks de conocimiento, 308
tock de conocimiento organizacional, 309

Capítulo 8:

Disponibilidad de los sistemas computacionales

Contenido

Introducción	378
Planteamiento del problema	379
Justificación de la investigación	380
Objetivos	380
Alcance y limitaciones	380
Antecedentes e investigación bibliográfica.....	381
Principales escuelas del pensamiento	381
Solución propuesta e hipótesis	395
Variable dependiente.....	396
Variables independientes	397
Preguntas moderadoras	402
Hipótesis	403
Hipótesis nula.....	403
Hipótesis de investigación	403
Modelo propuesto para la prueba de la hipótesis.....	404
Elaboración de la encuesta.....	405
Recolección de datos	406
Determinación de la población.....	406
Descripción de la muestra.....	407
Selección de la muestra representativa	408
Trabajo de campo.....	408
Ficha técnica de la investigación	409
Análisis de resultados	409
Análisis descriptivo.....	410
Análisis estadístico.....	413
Aceptación o rechazo de la hipótesis nula	422
Comprobación de las hipótesis de investigación	423
Contribuciones.....	424
Comentarios generales	426
Líneas futuras de investigación.....	426
Referencias	428
Índice del capítulo	431

Introducción

En México como en el resto del mundo, las grandes, medianas y pequeñas empresas que manejan sus transacciones de negocio a través de sistemas computacionales, en algún momento han dejado de operar a causa de la no disponibilidad de sus recursos computacionales. Esta problemática aparte de repercutir en pérdidas financieras, daño a las marcas, pérdida de imagen, confianza de los clientes y socios de negocio, puede llegar a generar hasta la salida del mercado para algunas empresas.

Por otro lado, es generalmente aceptado que en oposición a la *Era Industrial*, hoy vivimos en la *Era de la Información y del Conocimiento*, por lo que uno de los valores más importantes que poseen las organizaciones, aparte de los seres humanos, es la información que es mayormente almacenada, procesada o generada a través de sistemas computacionales.

Considerando lo anterior, la presente investigación se sitúa en el campo de la Seguridad de Información. Su objetivo general es determinar los principales factores que influyen en el mejoramiento de la disponibilidad de los sistemas computacionales, así como obtener una caracterización de los niveles de riesgo, vulnerabilidades, fortalezas, amenazas, y principales problemas de seguridad de información que enfrentan las empresas de los sectores industrial, comercial o de servicios.

Desde el invento de las computadoras hasta nuestros días, se han venido manejando conceptos nuevos en el ámbito de las *Tecnologías de Información (TI)* y *Procesos de Negocio (PN)*, así como en la *Seguridad de Información (SI)*. Sin embargo, con frecuencia las personas que estamos relacionadas con la tecnología computacional nos preguntamos hasta dónde debemos de invertir en herramientas de seguridad computacional y en personal administrativo y técnico para que una organización minimice el riesgo e sufrir ataques de personas internas y externas.

Cano (2006) indica que la SI es subjetiva, es decir, la organización y sus ejecutivos de alto nivel plantearán y autorizarán la estructura organizacional y las herramientas de seguridad relacionadas a la SI de acuerdo a su percepción de ésta. Cada uno de los ejecutivos o especialistas en SI tiene una manera de comprender y entender la SI. Un especialista de una organización bancaria tiene diferente visión de un especialista de seguridad de una organización comercializadora o de una organización manufacturera. Cada uno de ellos comprende la realidad de la seguridad según su exposición al riesgo.

Se puede pensar que la seguridad es intangible, es decir, no se puede medir. Existen los mecanismos o herramientas de seguridad instalada, la cultura en seguridad, las políticas, los procedimientos, personas contratadas altamente calificadas con certificaciones internacionales y metodologías de vanguardia, pero esto no es suficiente. Podemos darnos cuenta que todos los días hay noticias de diferentes fuentes, así como encuestas de seguridad de información que hacen organismos reconocidos tanto nacional como internacionalmente. Sólo hay que mirar a nuestro alrededor a las instituciones financieras, comerciales, bancarias, manufactureras y hasta

las mismas de tecnología computacional y encontraremos que frecuentemente sus esquemas de seguridad computacional han sido atacados por personas internas o externas, algunas veces teniendo éxito y otras no.

Rainer, Marshall, Kenneth, y Montgomery (2007) consideran que posiblemente lo más importante que las organizaciones deben comprender hoy en día, es que no es la Era Industrial, sino la Era de la Información y del Conocimiento, y que en su gran mayoría la información de las organizaciones, pequeñas, medianas ó grandes, es procesada a través de medios computacionales, por lo tanto, debe tenerse conciencia del valor de la información que se genera a través de los sistemas computacionales. Estos autores son de la opinión que hoy las organizaciones conocen que las tecnologías de información son esenciales no únicamente para las operaciones diarias, sino también para obtener ventajas estratégicas y competitivas. La importancia de la tecnología computacional significa que la SI también es muy importante. Las brechas en la SI puede que repercutan en pérdidas financieras, daño a las marcas, pérdida de confianza de los clientes, pérdida de confianza de los socios de negocio, pudiendo provocar en algún momento que la organización hasta pueda salir del mercado. Por lo anterior, la no disponibilidad de los recursos computacionales por falta de controles administrativos y herramientas computacionales, o por fallas de estrategia o arquitectura de seguridad de información, crea tremendos retos para accionistas, directores de empresas, gerentes y profesionales que están involucrados en la salvaguarda de los recursos computacionales y muy particularmente en los responsables de la seguridad de información.

Planteamiento del problema

Es fácil verificar que como en el resto del mundo, en México las organizaciones que manejan sus transacciones de negocio a través de sistemas computacionales han dejado de operar a causa de fallas en sus diferentes recursos computacionales.

De igual manera, las empresas que manejan sus transacciones de negocio a través de sistemas computacionales no siempre cuentan con la estructura organizacional adecuada, herramientas de seguridad computacional, seguridad lógica, física y ambiental, cultura en seguridad computacional y un clima organizacional aceptable para hacer frente a los diferentes problemas de SI que propician la no disponibilidad de los recursos computacionales.

Considerando la problemática anterior, y como se trataba de llevar este caso a un terreno específico y real, nos dimos a la tarea de buscar una experiencia que nos permitiera desarrollar un trabajo de campo.

Justificación de la investigación

Todo tipo de organizaciones (industriales, comerciales o de servicios) cuyas transacciones estén siendo procesadas mediante sistemas computacionales —compras, ventas, inventarios, recursos humanos, nóminas, mercadotecnia, e-business, etc., — se verán ampliamente beneficiadas con una investigación ya que identifica y evalúa las variables que afectan la disponibilidad de los recursos computacionales.

Objetivos

OBJETIVO GENERAL

Identificar los principales factores que influyen en el mejoramiento de la disponibilidad de los sistemas computacionales, así como obtener una caracterización de los niveles de riesgo, vulnerabilidades, fortalezas, amenazas, y principales problemas de seguridad de información que enfrentan las empresas de los sectores industrial, comercial o de servicios.

OBJETIVOS ESPECÍFICOS

- a. Establecer y definir los factores clave que determinan la disponibilidad de los sistemas computacionales.
- b. Utilizando las herramientas de análisis estadístico encontrar aquellos factores que expliquen mejor la disponibilidad de los sistemas computacionales.
- c. Elaborar una amplia investigación bibliográfica sobre el estado del arte, que permita fundamentar el marco teórico que sustente la elaboración de la investigación.

Alcance y limitaciones

Esta investigación sólo tomará en cuenta el estudio y análisis de la información referente a la no disponibilidad de los recursos informáticos derivada de la problemática de los ataques internos y externos y de que los recursos informáticos hubiesen sido alterados por estos incidentes. Los recursos que podrían ser afectados se han clasificado en telecomunicaciones, hardware, software, bases de datos, sistemas aplicativos, sistemas para la toma de decisiones, entre otros. Asimismo, se consideraron aquellos elementos que aporten criterios con los cuales se puedan realizar juicios valorativos respecto al papel que juega la SI ante éste tipo de eventos.

La principal limitante para realizar ésta investigación ha sido la dificultad para la obtención de información; ya que algunos directores o gerentes de TI y PN o responsables de la SI no contestaron la encuesta para no revelar que sus empresas habían tenido problemas en aspectos de SI. No obstante, de las 110 encuestas enviadas a los diferentes ejecutivos de TI y PN, fueron contestadas 88 y con base en estas observaciones se realizó el estudio.

Antecedentes e investigación bibliográfica

Principales escuelas del pensamiento

En 1951, IBM distribuye a diferentes empresas y universidades las primeras computadoras para usos comerciales y de administración, pero hasta el año 1964 se presentó el primer caso judicial por delito a través de una computadora, es decir 13 años después (Beekman, 1995). A partir de esta fecha hasta nuestros días se han suscitado una serie de delitos informáticos y con esto surgieron pensadores, temas y certificaciones con el objetivo de colocar una serie de controles para minimizar los riesgos en SI.

Vallabhaneni (2000) establece que la SI tiene tres propiedades o requerimientos importantes que son: integridad, confiabilidad y disponibilidad. También comenta que existen tres aspectos importantes a tener en cuenta para contrarrestar los problemas de seguridad. El primero, ser reactivo, es decir, colocar detectores de intrusos, firewall, antivirus, entre otras herramientas de seguridad. El segundo aspecto, ser proactivo, es decir, colocar scanner de vulnerabilidad, analizador de tráfico, firewall para aplicaciones, entre otras. El tercer aspecto, contar con políticas, y lo más importante, planes de contingencia, entre otras medidas. Hay que tomar en cuenta medidas, como capacitar empleados, realizar auditorías en TI y PN así como SI, contar con mecanismos legales, reglamentos internos, pólizas de seguros, mecanismos tecnológicos de SI, así como mecanismos de autenticación y encriptación.

Herold (2005), propone que todas las personas necesitan conocer acerca de las amenazas, capacitación y motivación referente a la seguridad de información. La motivación ha sido olvidada del lenguaje de la seguridad de información, se ha demostrado que la seguridad de información es parte de la ejecución el trabajo. El poder cae en la aplicación real de la teoría de la educación, métodos, y prácticas de enseñanza de alertar sobre la seguridad a través de la capacitación.

Peltier (2006), explica que existen conceptos básicos de planeación estratégica, sobre estrategias de seguridad y tecnología de información. Es importante contar con lineamientos definidos sobre estrategias de seguridad de información. La definición de lineamientos de la estrategia de información también es importante ya que provén la metas de seguridad a largo plazo y que estas deben ser efectivas para proteger los recursos informáticos

Thompson (2007), comenta que las nuevas tecnologías han sido inventadas para innovar las estrategias de los negocios. La web ha sido nombrada como el lugar de mercadotecnia por los investigadores de mercado. La web provee acceso completo a la información. Tiene un crecimiento anual de más del 500% en el mundo, es ideal para llevar los negocios hasta el rincón más apartado del mundo. Esta tendencia tecnológica es inminente y representa una apertura sin precedentes.

Whitman y Mattord (2007), describen a detalle las políticas y procedimientos de seguridad y los lineamientos referentes a la seguridad de información. Enfatizan de cómo los documentos y estándares de seguridad son elementos clave en los procesos de negocio y que nunca deberíamos olvidarlos. Insisten que las políticas, estándares y procedimientos deberían existir únicamente para soportar los objetivos de negocio o los requerimientos de la misión, ellos son elementos que ayudan a la ejecución de las políticas administrativas. Hay que contemplar que la seguridad de información debe estar integrada dentro de todos los aspectos de los procesos del negocio.

Hasta aquí nos hemos propuesto a explorar teóricamente la evolución del porque las organizaciones dejan de operar sus sistemas computacionales. A partir de ahora explicaremos teóricamente los cinco variables que están relacionados a este caso de estudio. Para ello será necesario explicar los términos que los soportan y de esta manera explorar la relación que se establece entre estos conceptos desde la teoría y cómo influyen en la disponibilidad computacional.

PRIMERA VARIABLE: EO EN TI Y LA FUNCIÓN DE SI

De 20 años a la fecha, los negocios a nivel mundial han adoptado a las TI como un recurso para aumentar la eficiencia y productividad (Harley, et al. 2006). La adopción de estas tecnologías ha tenido implicaciones significativas para el rediseño de las estructuras organizacionales. Gold, Malhotra, Arvind y Segars (2001) expresan que la estructura organizacional es importante en el apalancamiento de la arquitectura computacional.

Dell y Grayson (1998) explican que de hecho la optimización de compartir el conocimiento dentro de una área funcional puede muchas veces sub-optimizar la efectividad. Esencialmente es importante que las estructuras organizacionales sean diseñadas con flexibilidad (en opción a la rigidez) para que fomenten la eficiencia y eficacia y la colaboración a través del personal dentro de la organización.

Palmer, Dunford y Akin (2005) comentan que la implicación de estos enfoques, es que las reacciones adversas de los empleados son el resultado de una inherente aversión al cambio, preferencia por evitar la incertidumbre, un apego a la forma establecida de hacer las cosas, una falta de entendimiento de que el cambio es necesario, pérdida de interés del empleado, o resultado de cansancio por el cambio. Menos frecuentemente, se concede que las reacciones adversas al cambio podrían estar limitadas a preocupaciones legítimas por el impacto del cambio en el desarrollo y bienestar individual u organizacional.

Cabe hacer mención, que en el año 2003 los responsables de TI y PN se inclinaron por separar la función del área de TI y PN de la de SI. Sin embargo, de acuerdo al artículo de Scott (2007) se concluye que el departamento de TI pretende controlar la seguridad nuevamente. La razón por la cual el Oficial Principal de Información (Chief Information Officer CIO) y el Oficial Principal de Seguridad (CSO) siempre han defendido la separación de TI y PN de la de SI, es por el clásico problema del zorro en la casa de la

gallina. El inconveniente se da cuando el CIO controla, por ejemplo, un proyecto dedicado a la innovación del uso de TI y la seguridad al mismo tiempo. Esto provocaría un serio conflicto de intereses entre el CIO y el CSO, mismo que podría provocar atrasos e incrementar costos. Entonces ¿donde debería estar ubicada el área de SI y de quien dependería?

Actualmente en muchas organizaciones se están rediseñando algunos puestos y funciones acorde a los nuevos requerimientos en las áreas relacionadas con TI y PN y SI, algunos de esos puestos están desapareciendo y otros están estableciéndose de acuerdo a las necesidades de las tecnologías. Asimismo, los ejecutivos de alto nivel no están tan involucrados en las funciones de los responsables de la SI, es por eso que en algunas organizaciones todavía dependen estas funciones del área de TI y PN y en otras organizaciones por regulaciones legales o de control dependen de un área independiente. Pero para que los ejecutivos de alto nivel puedan ubicar a esta área (Nonaka y Takeuchi, 1995) desarrollaron una nueva estructura organizacional, que puede dar entrada a la creación de aumentar el conocimiento eficientemente dentro de la organización.

En general, esta es una combinación de una estructura organizacional formal y una no jerárquica, por lo que pueden separarse las áreas y sus funciones sin ningún problema. Esto es particularmente el caso para la última generación de las áreas de TI y PN, así como para las de SI, la cual integra información dentro de las empresas y a través de límites funcionales. Mientras más organizaciones adoptan estas tecnologías, es cada vez más importante un entendimiento de los factores que influyen en las respuestas gerenciales al cambio tecnológico, tanto en términos teóricos como prácticos.

Williams, Vistakon, Rains, Johnson, Vistakon y Hohnson (2007) son de la opinión que muy a menudo el re-diseño organizacional se enfoca únicamente al cambio de estructura, esto es moviendo personal de un departamento a otro. Hay poca consideración a los lineamientos del negocio o estrategias que afectan las decisiones del diseño organizacional. Si se hace sistemáticamente el diseño organizacional puede ser una ventaja competitiva poderosa, colocando a los mejores talentos en los lugares adecuados que permitan dar los mejores resultados. Pero esto requiere un claro entendimiento de la estrategia del negocio, un óptimo proceso de trabajo y finalmente un adecuado plan y procedimientos para asegurar el éxito a corto, mediano y largo plazo.

Los cambios estructurales les han costado dinero y esfuerzo a las organizaciones y más en las áreas de TI y PN y las de SI. En estas áreas que son de soporte computacional para toda la organización, también se ha complicado la adaptación del ser humano a esos cambios, porque en todos los tiempos ha existido resistencia al cambio. Las TI están cambiando por lo tanto la estructura organizacional también, algunos puestos relacionados a la TI están desapareciendo y algunos se están creando al mismo tiempo que la tecnología está surgiendo.

Palmer, Dunford y Akin (2005) comentan que la implicación de estos enfoques, es que las reacciones adversas de los empleados son el resultado de una inherente aversión al cambio, preferencia por evitar la incertidumbre, un apego a la forma establecida de hacer las cosas, una falta de entendimiento de que el cambio es necesario, pérdida de interés del empleado, o resultado de cansancio por el cambio. Menos frecuentemente, se concede que las reacciones adversas al cambio podrían estar limitadas a preocupaciones legítimas por el impacto del cambio en el desarrollo y bienestar individual u organizacional.

Los ejecutivos de alto nivel deberían de repensar donde ubicar las funciones de SI. Esto es, para que pueda ser un área con independencia para analizar, diseñar e implementar políticas, así como para sugerir herramientas y fomentar la cultura en SI y para minimizar el riesgo, entre otras responsabilidades. Cabe hacer mención, que en la encuesta del año 2003 un CSO comentó que el conflicto era demasiado grande para superarlo, y que cuando un CSO reportara a un director de TI y PN, sería un fracaso. Cada año a partir de ese momento, la tendencia fue que la función de SI fuera ganando autonomía. Se fueron creando más posiciones ejecutivas en las áreas de SI y el poder de toma de decisión se fue pasando al área de SI y alejando de TI y PN y así más grupos en SI reportaron sus tareas a otras áreas distintas de TI y PN, incluyendo el departamento de riesgos y el más significativo, el CIO y el de CSO. Esta tendencia es aún más pronunciada en las grandes compañías pero las medianas y pequeñas ya están adoptando este modelo.

Por lo tanto una no adecuada estructura organizacional en tecnología de información y la función de la SI afecta directamente a la disponibilidad computacional.

SEGUNDA VARIABLE: HERRAMIENTAS DE PROTECCIÓN

Cuando ya son conocidos las vulnerabilidades y los ataques a las que están expuestos los recursos computacionales, es necesario conocer las herramientas de protección disponibles para protegerlos. Mientras algunos controles son evidentes como los de seguridad lógica, física o ambiental, otros no, por lo que hay que investigar que herramientas de protección están en el mercado y proponer su compra y su implementación, si es justificado con un estudio previo de costo beneficio (Lucena, 1999).

Muchas de las vulnerabilidades encontradas son el resultado de la implementación incorrecta de tecnologías, otras son consecuencias de la falta de planeación, pero como más adelante se menciona, la mayoría de las brechas y amenazas en la seguridad computacional son ocasionadas por usuarios que trabajan dentro de la organización y por personal externo, y una de las tareas del responsable de la seguridad es detectarlas y encontrar la mejor manera de que estas disminuyan. Esto puede lograrse a través de la selección e implementación de herramientas de protección de SI.

Otra aparte importante es tener al personal administrativo y técnico competente para que seleccione, implemente y monitoree las herramientas de protección de SI, con base al riesgo que quiera tomar la empresa. ¿Qué

quiere decir esto? Si la organización no invierte en herramientas de seguridad probablemente haya más riesgo de que los recursos computacionales que utiliza como el hardware, software, telecomunicaciones, bases de datos, sistemas operativos, entre otros, puedan ser eliminados, alterados (disminuir o aumentar información) o dañados, ya sea de forma intencional o no.

Como indica Ardita (2001), los costos de las diferentes herramientas de protección se están haciendo accesibles, en general, incluso para las organizaciones más pequeñas. Esto hace que la implementación de mecanismos de seguridad se dé prácticamente en todos los niveles: empresas grandes, medianas, chicas y usuarios finales. Todos pueden acceder a las herramientas que necesitan y los costos (la inversión que cada uno debe realizar) van de acuerdo con el tamaño y potencialidades de las herramientas, considerando que los constantes cambios de la tecnología hacen que para mantener un nivel actualizado de seguridad, se deba invertir permanentemente en las actualizaciones de las herramientas con las que se cuenta o en la compra de otras.

La selección de herramientas de seguridad depende por un lado de la arquitectura de hardware, software y telecomunicaciones, así como del tipo de sistemas operativos, bases de datos, y otros recursos que se quieren proteger; por otro lado, depende de la cantidad de recursos económicos que la organización quiera invertir en estas herramientas de protección y en la contratación del personal para que las opere.

Narasu (2006) sugiere que las empresas deben al menos seleccionar e implementar en sus empresas software's de antivirus, antispyware, filtrado de paquetes o contenido y firewalls (contrafuego). Estas herramientas básicas pueden proveer una convincente y sólida protección contra importantes amenazas. Este conjunto de herramientas actualmente se puede encontrar en un solo paquete llamado suite.

Las vulnerabilidades no son debilidades o ausencias de controles, es reconocer que aún no terminamos de conocer las diferentes deficiencias de las plataformas que administramos así como las aplicaciones que en ellas se instalan. Es claro que el responsable de SI debe ser fruto de un esfuerzo continuo y dedicado para entender y modelar las posibles situaciones de falla y no una posición cómoda que sólo reacciona ante un incidente.

Ser vulnerable es reconocer que tenemos que aprender todos los días, es abrir la mente a nuevas posibilidades, es mirar en la inseguridad de la información la fuente de proponer controles. Renunciar a aprender de la inseguridad, es enterrar el futuro de la seguridad. Es negar la dinámica misma de los desarrollos que se pueden dar, entendiendo las nuevas formas de protección que no son otra cosa que la comprensión de la inseguridad materializada en una técnica, una herramienta o una propuesta.

Finalmente, ningún sistema es invulnerable, por lo que el suponer que en un momento dado se va a realizar un ataque exitoso, es una parte importante de prepararse para ello. Esto significa, que detectar ataques y entenderlos es tan importante como detectar vulnerabilidades y corregirlas. Las acciones correctivas dependen en general de la capacidad que tenga

el administrador para rastrear la forma en que se llevó a cabo el ataque y descubrir él o los culpables. Esto lleva a considerar que la mejor estrategia consiste en tres líneas o normativas:

- a. Revelar la menor cantidad posible de información sobre el sistema operativo, cuentas de usuarios y servicios disponibles, entre otros, tanto a usuarios internos como externos.
- b. Cerrar la mayor cantidad de puertas a vulnerabilidades que sea posible y dejar las que queden abiertas bien vigiladas, pero siempre hay que estar alerta y suponer que los atacantes en un momento dado encontrarán puertas traseras no detectadas por el responsable de la SI y tendrán acceso al sistema, y
- c. Asegurarse de que ningún atacante borre sus huellas, para evitar que los ataques se lleven a cabo de manera clandestina y repetida.

Por lo tanto no tener las herramientas de protección adecuadas podría aumentar el riesgo de que la disponibilidad computacional sea afectada.

TERCERA VARIABLE: SEGURIDAD LÓGICA, FÍSICA Y AMBIENTAL

La Seguridad Lógica consiste en la "aplicación de barreras y procedimientos que resguarden el acceso a los datos y sólo se permita acceder a ellos a las personas autorizadas para hacerlo. Existe un viejo comentario que la seguridad informática dicta que "todo lo que no está permitido debe estar prohibido" y esto es lo que debe asegurar la Seguridad Lógica.

SL también se define como la aplicación de barreras y procedimientos que resguarden el acceso a los recursos computacionales y sólo permita acceder a ellos a las personas autorizadas para hacerlo. De acuerdo a Gollmann (1999) explica que la SI es un concepto o una utopía" y reconoce a SI como la integridad, disponibilidad, confidencialidad y no – repudiación de los activos.

Según Stilianos y Zafar (2007) clasifican a la seguridad computacional en las siguientes categorías de activos: software, comunicaciones, datos, hardware, seguridad lógica ambiental y física, personal y organizacional y administrativa.

Así, la seguridad física y ambiental consiste en la "aplicación de barreras físicas y procedimientos de control, como medidas de prevención y contramedidas ante amenazas a los recursos e información confidencial". Se refiere a los controles y mecanismos de seguridad dentro y alrededor de las áreas restringidas de Tecnología de Información, así como los medios de acceso remoto al y desde el mismo; implementados para proteger el hardware y medios de almacenamiento de datos.

Seguridad física y ambiental tal como Rainer et al. (2007) mostraron diferentes controles de seguridad física y ambiental. Según los estándares internacionales los centros de cómputo o los lugares físicos donde se encuentra el hardware, los periféricos y los recursos humanos relacionados a TI, pueden contar con uno o varios controles de entrada física y ambiental para asegurar que únicamente el personal autorizado le sea permitido el

acceso a áreas restringidas y también que se hayan instalado los controles ambientales para que puedan operar el hardware, software, telecomunicaciones, sistemas, etc., así como se salvaguarde la seguridad de los recursos humanos.

Según el manual del CISA (2007) considera los siguientes controles de seguridad física; puertas con llave, puertas eléctricas, puertas de entrada biométricas, bitácoras electrónicas, credencial con identificación, cámaras de video, guardias de seguridad, puertas dobles, puntos de entrada únicos y controlados, sistemas de alarma, entre otros.

El mas critico factor para la protección de los activos de información es la colocación de las bases para la administración de seguridad de información efectiva. La utilización del comercio electrónico, la pérdida de barreras de la organización por el uso de facilidades remotas de acceso, y de exposiciones de seguridad por dar acceso a perfiles con altos privilegios y que pueden ingresar virus, negación de servicios, conseguir acceso a recursos no autorizados, descubrir y robar identidades sobre el Internet, entre otros, esto se ha convertido en la necesidad para la gestión efectiva de la seguridad de información CISA Review (2008).

Como indican Rainer et al. (2007) la gran cantidad de sistemas y accesos restringidos y al número cada vez mayor de sistemas computacionales, se requiere en cada uno la identificación y contraseña independiente, ya sea para el acceso lógico o físico, por lo tanto, los usuarios pueden tener más de una contraseña y esto reduce la habilidad de de que recuerden la información de accesos requerida. Como resultado, los usuarios tienen que escribir la información, lo cual lo hace vulnerable, o llaman a sus administradores del control. Los administradores del control de accesos por lo general tienen que lidiar con llamadas de los usuarios que han olvidado su contraseña e identificación. Cabe mencionar, que las llamadas por servicio son muy costosas y están en aumento, así como el incremento de los servicios computacionales que ofrecen.

En el mundo actual de los negocios, términos como convergencia, sinergia e integración en seguridad física, lógica y ambiental, son utilizados tan frecuentemente que es fácil pasar por alto situaciones en donde realmente estén ocurriendo, estos sistemas son un caso a considerar.

Debido al resultado en el crecimiento en proyectos de seguridad administrativa y técnica la integración es crucial, cada vez los gerentes y ejecutivos de tecnología de información y procesos de negocio y de SI se dan cuenta que se ven amenazadas sus organizaciones, tanto en instalaciones como en sistemas de información. Adicionalmente muchos ejecutivos buscan obtener ventajas competitivas que puedan ser obtenidos por sus organizaciones a raíz de la integración de múltiples sistemas de seguridad.

El término "Integración de Sistemas de Seguridad" puede tomar una variedad de significados. Históricamente, la mayoría de proyectos de integración de sistemas de seguridad han involucrado las ligas de sistemas de alarma y sistemas de control de acceso y luego la integración de circuitos cerrados de televisión. Por ejemplo, una vez que estos sistemas son conec-

tados una violación del control de acceso provocara que la alarma suene y envíe una señal a la cámara de video para que se active y registre el incidente.

“Integración de Sistemas de Seguridad” también se refiere a la tendencia dentro de un creciente número de organizaciones para mezclar aplicaciones de seguridad física, como controles de acceso y aplicaciones de seguridad lógica tales como programas de identificación biométrica, que permiten a los empleados utilizar la red corporativa dentro de un sistema integral. Los empleados responsables de la seguridad física, así como los responsables de la seguridad de la red y las computadoras pueden compartir el mismo lugar físico.

Mientras más ejecutivos reconozcan la necesidad de instalar sistemas que protejan tanto las instalaciones como sus sistemas de información, bases de datos, software, sistemas para la toma de decisiones, telecomunicaciones, entre otros, este enfoque está ganando mayor aceptación. “La convergencia está sucediendo ahora”

En los últimos años, un tópico de discusión entre los expertos de SI ha sido la liga entre los sistemas de seguridad con los sistemas de recursos humanos. De esta manera cuando una persona ingresa o abandona la compañía, el sistema de seguridad automáticamente se actualizará cuando el departamento de recursos humanos actualice sus registros.

Como parte del movimiento a un enfoque más integral hacia la SI, un número creciente de organizaciones está asignado la responsabilidad tanto de la seguridad lógica como física y ambiental bajo una sola persona, tal como un Oficial Principal de Seguridad.

Es importante estar conciente que por más que las organizaciones hayan instalado los controles de seguridad lógica, física y ambiental acordes a las necesidades de la organización esto no garantiza una afectación por fallas o por alguna persona que intencional o no pueda dejar de operar la organización, por lo tanto se tiene un riesgo de que por alguna causa relacionada esta variable la disponibilidad computacional sea afectada.

CUARTA VARIABLE: CULTURA EN SEGURIDAD DE INFORMACIÓN

En el contexto de tecnología de información, usando el marco de trabajo de los valores de Hofstede examinó y soportó la noción de las investigaciones de diferencias en cultura. “Choi y Choi (2003) investigaron también diferencias culturales entre los Estados Unidos de América y Korea con un enfoque sobre el desempeño administrativo. Mao y Prashant (2006) reforzaron la estructura de investigación inter culturales examinando las diferencias culturales en la aceptación de la tecnología entre Estados Unidos y Turquía. En un estudio inter - cultural entre los Estados Unidos y Uruguay, McCoy y Jones (2005) examinaron la adopción usando el modelo de aceptación de tecnología”. Hofstede (2001)

Las organizaciones invierten fuertemente en las más recientes firewalls, sistemas de detección de instrucciones, y otras tecnologías de seguridad avanzadas; pero las pérdidas por incidentes de seguridad continúan

creciendo cada año. El problema no es con la tecnología de SI en si misma, sino con la falta de concientización en SI a los usuarios. Muy a menudo los profesionales en SI tienen que resolver problemas relacionados con violaciones causadas por los usuarios cuando abren archivos adjuntos enviados por correo electrónico y que contienen virus, por el olvido de hacer respaldos de archivos críticos, por la pérdida de computadoras personales con datos confidenciales o siendo forzados para proporcionar sus claves de acceso dando estas a través de trucos de técnicas de ingeniería social, entre otros problemas.

Por lo anterior, existe documentación establecida de métodos formales para asegurar que la conciencia de seguridad computacional está integrada hacia la cultura de una organización, en la documentación se tienen los planes de entrenamiento de habilidades. Métodos informales para presentar un cambio cultural incluye la mayoría de iniciativas porque esto ocurre tan seguido, bajo tales circunstancias, pequeños y negativos mensajes pueden rápidamente llegar a ser temas mayores.

Ha sido largamente reconocido que la cultura tiene un significativo impacto en las teorías organizacionales. Sería erróneo asumir que las teorías de aceptación de TI predicen igualmente bien en otros ambientes culturales. En un sentido práctico, mientras las compañías multinacionales continúan penetrando mercados extranjeros, la pregunta obligada es de ¿cómo debería ser introducida la cultura en SI en las otras partes del mundo y que respuesta se tendría? (Venetis, 2005).

Partiendo del concepto que plantea Nakagawa (1990) la cultura es aquella parte de las interacciones y experiencias humanas que determinan como el ser humano se siente, actúa y piensa. Es la cultura la que determina el sentido mismo de la visión que tiene el individuo de la realidad.

La cultura ha sido descrita como “algo para hacer con la gente y la calidad y estilo de organización únicos” (Kilmann 1985) “o la manera que se hacen las cosas aquí”. Una definición frecuentemente utilizada es la de (Hofstede, 2001) “La programación colectiva de las mentes que distingue un grupo o categoría de gente de otro grupo”. Hofstede (2001) comenta que existe una íntima relación entre cultura nacional y cultura organizacional, las compañías no pueden desarrollar una cultura organizacional que difiera sustancialmente de los factores culturales prevalecientes en el país en el cual la compañía opera.

La aceptación de la investigación y difusión sobre tecnología es crítica, dado de que nos provee conocimiento de cómo las organizaciones pueden manejarse. Con la globalización, es importante entender la adopción de las TI en otras culturas Mao y Prashant (2006).

Los expertos en SI reconocen que la primera línea de defensa de una organización son los empleados que interactúan con clientes, que procesan su información y la pasan a otros. La mayoría de discusiones de cumplimiento de privacidad y seguridad mencionan que las iniciativas exitosas requieren que las organizaciones realicen cambios culturales drásticos, incluyendo cambios en procesos operativos y conductas, aún estas discu-

siones pasan por alto maneras prácticas en que las organizaciones puedan realizar el cambio cultural para mejorar la SI y privacidad de datos.

Para colocar a las organizaciones en el sendero correcto hacia el cambio exitoso de la cultura es hay que seguir los siguientes pasos:

- a. Integrar a un profesional del cambio al equipo de la seguridad computacional.
- b. Entender la cultura actual de la seguridad.
- c. Determinar el estado futuro.
- d. Descubrir donde los problemas se encuentran.
- e. Establecer el rigor y el apoyo de la alta dirección

Cambio cultural de SI - Área de recursos humanos

Existen otros factores importantes para que el riesgo disminuya aún más, se pueden tener todas las políticas y procedimientos, las herramientas de SI de acuerdo a la estrategia de la empresa, pero si el personal interno no se selecciona bien, se capacita y se concientiza sobre un nuevo cambio de cultura en seguridad computacional tal vez todos los esfuerzos sean en vano.

Swing et al. (2007) explica que en el pasado los departamentos de recursos humanos se concentraban principalmente en el reclutamiento y retención de empleados, se preocupaban por atraer al personal calificado más adecuado o con habilidades específicas y mantener su moral alta. Considerando en la actualidad los aspectos de seguridad computacional el rol de los departamentos de recursos humanos deben evolucionar y crecer. Reclutar y retener al personal deberá ser la meta principal, pero las organizaciones deben tomar en cuenta que el departamento de recursos humanos es la primera línea de defensa contra los ataques maliciosos.

Stanley (2007) hace la siguiente pregunta ¿Cómo se puede constatar si se está seleccionando la persona adecuada para un trabajo? Una mala decisión de contratación puede conducir a resultados desastrosos. Por lo tanto encontrar a la persona correcta para el puesto es crítica para el éxito de una empresa.

“Algunas de los actos ilícitos en SI frecuentemente resultan de acciones de usuarios ya sean deliberadas o accidentales.” Un uso adecuado de políticas claramente establece que es considerada una conducta adecuada para el empleado en relación al uso de una computadora propiedad de la empresa, entre otros aspectos. Sin embargo se ha encontrado que aún en organizaciones con políticas apropiadas no siempre se comunican éstas a los empleados que necesitan seguirlas Ewing, Falcon y McGrane (2007).

Asumir el negocio alrededor de una política de seguridad que ha sido bien concebida, en el lugar, es la primera y más importante acción, es la creación de una cultura donde todos los empleados tienen la necesidad de conocer (qué hacer) habilidades (cómo hacerlo) y actitud (desear hacerlo) acerca de la política de seguridad computacional.

Por lo anterior, todo el personal de la organización debe de transmitirle una cultura en seguridad de información ya que juega un papel importante para la disponibilidad computacional.

Puede decirse que contratar a la persona correcta no es fácil. Muchos factores deben de ser tomados en consideración. En pocas palabras las personas responsables de hacer contrataciones deben mirar a los candidatos como un todo, de tal suerte que puedan realizar una buena decisión de contratación.

Una vez que el equipo de implementación entendió la cultura actual de la organización y conoce que tipo de cultura desea alcanzar, a continuación deberá conectar los puntos aplicando el análisis de deficiencia cultural a su plan general de seguridad/privacidad y estrategia de implementación. Este proceso debería descansar sobre la experiencia del equipo completo de seguridad y privacidad, empezando los ejecutivos de primer nivel. Mantener un registro a través de la implantación, el equipo debería tener en cuenta los resultados deseados en cuanto a cultura. Así como también establecer una matriz de riesgos y soporte con este proceso podría incluir iniciativas de creatividad como compañías de privacidad y seguridad, información continúa acerca de seguridad y privacidad en carácter internas o en comunicados como el intranet, posters, cartelones, incluir en agendas del staff y foros de preguntas y respuestas.

Para asegurar el éxito de una nueva cultura en SI, el programa del cambio debería ser implantado sistemáticamente, más que por etapas. Más aún el programa deberá ser implementado solamente cuando los grupo de SI y privacidad esté listo para empujar el plan y de manera rápida. La implementación deberá incluir un fuerte mensaje de la dirección, acciones visibles que proporcionen un rápido momento hacia el cambio y un seguimiento consistente.

Estudios futuros deberán ser dirigidos hacia la discriminación de la cultura y explicar porqué las teorías de explicación de TI funciona mejor o peor en diferentes culturas. Independientemente de las dimensiones de la cultura de un país en si misma puede haber políticas regulatorias y factores políticos y demográficos también en juego.

QUINTA VARIABLE: CLIMA ORGANIZACIONAL

Ha habido muchos autores que han escrito sobre el clima organizacional y han investigado otras variables, pero es bien sabido que el clima organizacional, es la relación de las personas que trabajan en una organización con la empresa. Si una empresa u organización tiene bien definidos los niveles jerárquicos y descripciones de funciones, los sueldos y salarios son competitivos, se fomenta una satisfacción en el trabajo y trabajo en equipo, así como una aceptable comunicación, la empresa podría llegar a tener éxito.

Para tener un mejor entendimiento del clima organizacional se presentan una serie de definiciones que permitirán visualizar con claridad las implicaciones de estos términos en las organizaciones.

Taylor y Bowers (1972) menciona que el clima organizacional se refiere a la “interacción entre ambiente y variables personales” o que es el resultado de la conducta de miembros de la organización. Pritchard y Karasick (1973) toman un punto de vista sobre el clima, viéndolo como el resultado de las conductas de los miembros de la organización. Ashforth (1985) argumentó que el clima organizacional es una “propiedad conjunta tanto de la organización como del individuo” en otras palabras, el clima es el resultado de la interacción de atributos de la organización y conducta de los individuos. Según Hall (1996) el clima organizacional se define como un conjunto de propiedades del ambiente laboral, percibidas directamente o indirectamente por los empleados, que se supone son una fuerza que influye en la conducta del empleado.

La importancia del clima organizacional es aparente. Es una liga entre el individuo y la organización. Guzley (1992) describe con base en otros autores que “representa creencias generalizadas de los miembros y actitudes acerca de la organización” si estas creencias y actitudes dentro de la organización son favorables puede significar que el individuo estará comprometido con la organización.

Tagiuri (1968) ve al clima como una propiedad de la organización en si misma identificable por características particulares e interpretada por los miembros de la organización que en turno afectan sus actitudes y motivaciones.

Las personas deben de tener un comportamiento individual de compartir el conocimiento respecto a los SI y TI. El conocimiento de los individuos con respecto a los sistemas de información/ tecnología de información (IS/IT) a través de los procesos de compartir el conocimiento, podría gradualmente acumularse y convertirse en un conjunto de conocimientos para una organización, lo cual haría la distinción de entre otros competidores y mejoraría su futuro y desempeño de la empresa. Kwok y Sheng (2005/2006).

Esencialmente los cambios estructurales implican cambios de diseños. Puede ser ello una norma, los cambios estructurales implican cambios en la división y coordinación del trabajo, es decir, en la integración. “...el propósito central de la estructura es coordinar el trabajo repartido en varias formas cómo, se logre esa coordinación (quién la hace y con qué) determina cómo será la organización” Mintzberg (1998). Uno de los métodos que aseguraron, tanto al personal como a los departamentos funcionales (con tareas altamente diferenciadas) coordinar (integración) adecuadamente el manejo interno de estas cooperativas fue la formalización de todos sus procesos. “La formalización involucra el control organizacional sobre el individuo y así tiene un significado ético y político además de ser un componente estructural.” Hall (1996).

En el pasado la mayoría de los directores generales de empresas han venido de áreas de finanzas o mercadotecnia. Sin embargo, últimamente se observó una tendencia que los directores generales pueden emerger de áreas de tecnología o al menos tienen fuerte conocimiento en TI y SI. Esta

tendencia es probable que continúe ya que la educación de TI y SI es necesaria para cualquiera que aspire a dirigir una organización en el futuro. Lo que buscan los directores generales en sus equipos cercanos para dirigir el área de TI y SI refleja el rol cambiante de TI en los negocios. Se empieza con definir los procesos del trabajo y prácticas y se pueden contestar las preguntas claves:

- a. ¿Cómo es realizado el trabajo?
- b. ¿Cómo fluye la información?
- c. ¿Qué es el proceso y las capacidades laterales?
- d. ¿Quién y serán impresionados por cualquier cambio al diseño actual de la organización (las funciones, los departamentos, socios externo/internos de clientes o negocio)?

Sueldos y salarios

Sin duda alguna, el salario es la variable estrella, tanto para la descripción de la actividad laboral como para la explicación y predicción de los comportamientos que en ella se manifiestan. Un empleo no es más que una transacción en la que el trabajador intercambia horas de ocio por una remuneración que le da el ente que lo contrata. El dinero sólo proporciona utilidad de forma indirecta, en la medida que concede poder de compra de otros bienes y servicios que si la proporcionan.

En este sentido, el dinero es visto como un bien que es impersonal y neutral y que sólo tiene significado cuantitativo, además la compensación económica tiene una importancia crítica porque es vista como la variable que iguala o compensa las diferencias entre empleados, en relación con otros aspectos del trabajo.

Cuando los sueldos y salarios no son competitivos en áreas de riesgo como son las de TI y SI se pueden prestar más a que las personas que trabajan en estas áreas sean las que cometan fraude ya que ellas poseen la mayoría del conocimiento de las vulnerabilidades y hay que recordar que personal interno es el que comete la mayoría de los fraudes.

Satisfacción en el trabajo

El concepto es una variable difícil de medir, porque implica que el individuo éste consciente de qué significa esto, también ha sido un área de gran interés para los ejecutivos de primer nivel, para los sociólogos y otras áreas relacionadas, que se han preocupado por los problemas que se presentan en las organizaciones.

En la mayoría de las sociedades industrializadas, las personas generalmente emplean aproximadamente un tercio de su tiempo diario en el trabajo, pero cabe hacer la aclaración que las áreas de TI y SI este tiempo aumenta ya que es un área de servicio interna y que debe de estar siempre disponible.

La toma de decisiones independiente de los empleados en beneficio de la empresa es un satisfactor que puede afectar el comportamiento de los

individuos ya que ellos pueden actuar de manera positiva o negativa. En www.eumed.net/tesis/cgb/c1.pdf comenta que los empleados que están a disgusto desaniman la participación de la fuerza de trabajo.

Los logros y metas son reconocidos por los niveles jerárquicos a través de remuneraciones económicas y reconocimiento emocional, tal vez el sueldo es un determinante importante para la satisfacción en el trabajo y en organizaciones que se tiene un salario base y se tiene un modelo de reconocimiento de bono económico por logros, el comportamiento de los empleados es diferente. Cuando no se tiene este tipo de modelo de bono económico los empleados suelen comportarse y ellos establecen a través de su juicio los resultados propios con los obtenidos por otros sujetos en su medio laboral y se termina la relación laboral, si están o no satisfechos con su trabajo.

Según la teoría de comparación entre necesidades y resultados de Maslow (1954) los individuos cotejan continuamente su estado actual de necesidades con el nivel de satisfacción que desean obtener de sus puestos de trabajo. Cuando no se satisfacen las necesidades aparece un estado de tensión que hace que los trabajadores no cumplan con su trabajo. El entrenamiento continuo ya sea interno como externo es otro de los satisfactores laborales, cuando los empleados se sienten que se les está actualizando en sus propias áreas de trabajo para mejorar los resultados, ellos pueden sentir que la empresa los está preparando y en cualquier momento decidirán si pertenecen a ello o buscan otras oportunidades, a diferencia si no están actualizados el trabajo se presenta como rutinario y poco retador.

Otro de los factores importantes es la motivación dentro de la empresa, un trabajador puede estar dotado generosamente con todo el equipo físico, las capacidades y las herramientas necesarias para realizar sus tareas, pero esto no garantiza que las realice. La motivación, es principalmente, esa "misteriosa" cualidad que lo impulsa a emprender la acción para ejecutar la tarea.

Es un estado interno que actúa o induce a algo, es lo que da energía, dirige, encauza y sostiene las acciones y comportamiento de los empleados (Steers y Porter, 1983).

Así como Gibson (2001) la describe como una necesidad no satisfecha lo que constituye el punto inicial en el proceso de motivación, por cuanto la diferencia de algo dentro del individuo es el primer eslabón de la cadena de suceso que dirige la conducta.

Ahora más que nunca el trabajo en equipo es de suma importancia por la aparición de centros tecnológicos da un nuevo significado en el cual se está construyendo el trabajo por día, por semana o por mes y que un grupo de personas ya sea concentrados en un área o desde su casa hacen el trabajo no importando el lugar donde estén, este grupo se les asigna tareas específicas y comparten información y pueden tener revisiones en el momento que el líder o los líderes así lo decidan, simplemente abriendo la página de Internet donde todos ingresan sus avances, entonces la vida de trabajo es diferente, es decir, los empleados y obreros se pueden contratar

en cualquier tiempo, en cualquier lugar en el mundo. Sin duda las tecnologías emergentes y predominantemente el impacto del Internet sobre el trabajo y las organizaciones, se podrían haber visto como una mayor revolución, quizá la más significativa desde la revolución industrial (Toffler y Toffler, 1980).

Muchinsky (1977) comentó que de las más intangibles variables de la organización es la comunicación. Porqué la comunicación es un fenómeno dinámico y es un concepto difícil de medir. Recientemente, muchos estudios han examinado variables relacionadas al clima organizacional.

Se ha realizado la siguiente sugerencia “los patrones de comunicación utilizados por la organización tienen un impacto inmediato sobre la vida de los individuos dentro de la organización y pueden ser un vital aún actualmente inexplorado aspectos del clima organizacional”.

Es indispensable notar que la comunicación informal es tan importante que más canales formales. Es relativamente fácil la obtención y uso de información atrayendo modificaciones en la estructura social en general. Y ciertos cambios en “Métodos de Trabajo” dentro de las organizaciones y los individuos. El rápido crecimiento en el uso de computadoras y el Internet ha contribuido en numerosos cambios en los métodos de trabajo tradicional y se han abierto nuevos horizontes para la creación de nuevos patrones de trabajos flexibles.

El clima organizacional es en gran medida establecido por las palabras y acciones de la alta dirección y administración de una empresa. Para reforzar el mensaje de la organización el consejo de directores y de alta administración debe servir como modelos a seguir y demostrar un fuerte compromiso con una conducta ética. La pregunta para la organización es siempre ¿Estoy haciendo las cosas como quisiera que todo el personal de mi organización las hiciera? Partiendo del análisis de las teorías y términos sobre clima organizacional y las variables definidas en este estudio podemos concluir que los términos cobran importancia en todas las organizaciones de ellos dependerá la eficacia y productividad de las misma. El clima organizacional rígido influye sobre la desertión del personal calificado de la organización hacia otras fuentes de trabajo.

Por lo tanto si una empresa u organización tiene bien definidos los niveles jerárquicos y descripciones de funciones, los sueldos y salarios son competitivos, se fomenta una satisfacción en el trabajo y trabajo en equipo, así como una aceptable comunicación, la empresa podría disminuir el riesgo de la disponibilidad computacional.

Solución propuesta e hipótesis

En esta sección se describen los significados de la variable dependiente, independientes y moderadoras y su significado de cada una de ellas.

Variable dependiente

Y: Definida como el tiempo que dejó de operar la organización a través de sus recursos computacionales por problemas relacionados a la SI, medido en horas que ha dejado de operar la organización en los últimos 12 meses.

La no disponibilidad de servicios computacionales abarca interrupciones debidas a distintos rubros: telecomunicaciones, hardware, software, errores humanos o ignorancia del personal, soporte y entrega de servicios, seguridad física y ambiental y por los ataques de personal interno o externo, como a continuación se muestra.

Simbología de la variable	No disponibilidad
Y _T	Cuántas HORAS estima usted que la empresa ha dejado de operar en los últimos 12 meses por fallas en los rubros que a continuación se enumeran. Considere centros de cómputo centralizado (áreas corporativas) y centros de cómputo descentralizado (sucursales).
Y ₁	1.Telecomunicaciones (LAN, MAN, PBX, WAN, VAN, VPN, Internet, Redes Inalámbricas, otras):
Y ₂	2.Hardware Mainframes (Computadoras con alta capacidad de procesamiento para realizar trabajos rutinarios) Midrange computers (Computadoras de medio rango) Servers (Servidores) Microcomputers (Computadoras personales) Thin client computers (Computadoras que están en estaciones de trabajo) Notebook/laptop computers (Computadoras que no están en estaciones de trabajo) PDA (Asistentes digitales personalizados) y otros hardware 's.
Y ₃	3.Software (base de datos sistemas operativos, sistemas de seguridad, sistemas transaccionales, sistemas ejecutivos, entre otros)
Y ₄	4. Errores humanos e ignorancia del personal, tanto interno como externo (consultores y/o socios de negocios).
Y ₅	5. Soporte y entrega de los servicios de Tecnología de Información o Sistemas (fallas procesos, respaldos, mala planeación en crecimiento de capacidad, compras de periféricos a destiempo, instalaciones, entre otros).
Y ₆	6.Por seguridad lógica, física y ambiental (falta de controles o carencia en puertas, puertas de entrada biométricas, bitácoras, electrónicas, paneles de control de alarma , detectores de agua, extintores de fuego manejados manualmente, alarmas de fuego, detectores de humo, sistemas de supresión de fuego, pisos, techos y paredes resistentes al fuego, entre otros, que afectaron la continuidad de la operación).
Y ₇	7. Ataques de personal interno y/o externo por diferentes métodos como (caballos de troya, virus, tunneling, gusanos, servicio denegado, spyware, trap door, modificación de contenido de mensajes, sniffers, sabotaje, phishing, scripts, buffer overflow, seguridad física, pharming, ataque de fuerza bruta, ingeniería social, escaneo de puertos, mal código de software , spam, extorsión de información, bomba lógica, botnet, adware, browsing, fingerpints, vandalismo, robo de computadoras, daños por derrame de agua, entre otros).

Tabla 8.1Variable dependiente. Fuente: Elaboración propia

Variables independientes

A continuación se definen las variables independientes establecidas:

X1: Estructura organizacional en TI y la función de la SI

Se refiere a como se integra un departamento de tecnologías de información y específicamente el área de SI en caso de que las organizaciones ya la tengan, asimismo, a qué nivel de madurez se encuentran.

X2 : Herramientas de protección

Se refieren a las herramientas de seguridad para proteger los recursos computacionales. Pueden ser de varios tipos: hardware, software, telecomunicaciones, bases de datos, y sistemas operativos, entre otros.

X3 : Seguridad lógica, física y ambiental

Reglas para acceder tanto los recursos computacionales como la información que es ingresada, procesada y emitida; así como también reglas de control para operar los recursos computacionales.

X4 : Cultura en SI

Es la transmisión, dentro de una organización, de las reglas e ideas referentes a la Seguridad de Información.

X5 : Clima organizacional

Se refiere al sentimiento de los seres humanos que pertenecen a una organización, en relación a cuestiones laborales y a su satisfacción en el trabajo. A continuación se muestran cada una de las tablas de las 5 variables independientes.

Variable X₁	Estructura organizacional en TI y la función de la SI
Simbología de la variable	Pregunta relacionada a la simbología
X _{1.1}	1. El personal que conforma la estructura organizacional de tecnología de Información y Procesos de Negocio (TI y PN) cubre la mayoría de las necesidades de soporte, desarrollo, mantenimiento, telecomunicaciones, según el tamaño, giro y nivel de complejidad de la empresa.
X _{1.2}	2. Todo el personal de TI y PN cuenta con descripciones de puestos donde se le hayan definido sus roles y responsabilidades.

X _{1.3}	3. Cada una de las personas que están en el departamento de TI y PN realizan las tareas que han sido descritas en la descripción de puestos.
X _{1.4}	4. Indique si existe una persona o grupo de personas responsable de la función de Seguridad Computacional. Si su respuesta es NO pasar a la pregunta
X _{1.5}	5. Si existe una persona o grupo de personas, a que nivel de precisión se encuentran definidas las funciones del o los responsables de la Seguridad Computacional.
X _{1.6}	6. Se cuenta con un documento escrito donde se detalle la POLÍTICA o las POLÍTICAS DE SI.
X _{1.7}	7. Si se cuenta con la Política escrita, en qué grado aproximado de avance se lleva.
X _{1.8}	8. Las Políticas de Seguridad Computacional son conocidos por todo el personal de la empresa, y en situaciones cuando se requiera por personal externo como consultores o socios de negocio.
X _{1.9}	9. Como parte de la obligación contractual, los empleados, consultores y socios de negocio, firman la política de seguridad computacional
X _{1.10}	10. En términos generales ¿Cómo considera la estructura organizacional de TI y PN y el Área responsable de la Seguridad Computacional? (de acuerdo al tamaño, complejidad y giro de la empresa).

Tabla 8.2 Variable Estructura organizacional en TI y la función de la SI.
Fuente: Elaboración propia

Variable X₂	Herramientas de protección
Simbología de la variable	Pregunta relacionada a la simbología
X _{2.1}	1. ¿Se cuenta con inventario de todos los activos computacionales, se tienen identificados, se les da mantenimiento y se actualizan de manera oportuna?
X _{2.2}	2. ¿Se tiene identificado, documentado e implementado las reglas de aceptación para el uso de los activos y de la información?
X _{2.3}	3. En caso que esté utilizando alguna de las siguientes

	herramientas ¿qué tan efectivas las considera? 3. Firewalls (Bloqueos para no dejar pasar a personas extrañas por la red).
X _{2.4}	4. Tecnologías para accesos físicos de usuarios (accesos biométricos, accesos con tarjetas, accesos a través de puertas entre otros).
X _{2.5}	5. Tecnologías de antivirus que se encuentren instalados en su empresa.
X _{2.6}	6. Tecnologías para el filtrado de contenidos y filtrados de Localizadores Uniformes de Recursos (Uniform Resource Locator).
X _{2.7}	7. Tecnologías de encriptación y criptografía.
X _{2.8}	8. Si su empresa cuenta con alguno o algunos de los siguientes tipos de comunicaciones (Red de área Local (LAN) Red de área Metropolitana (MAN) Private Branch Exchange (PBX) Redes de área Extendida (WAN) Redes de valor Agregado (VAN) Redes privadas Virtuales (VPN) redes Inalámbricas u otras) ¿cómo considera su nivel de seguridad?
X _{2.9}	9. Si su empresa maneja comercio electrónico, transacciones en línea o información que esté disponible públicamente ¿cómo la considera?
X _{2.10}	10. Si la empresa tiene instalados certificados o firmas digitales, infraestructura de llave pública, infraestructura de llave privada, computadoras de investigación para SI (honeypots) o redes de investigación para asegurar la información (honeynets) ¿cómo considera su implementación, administración y actualización?
X _{2.11}	11. Si se tiene instalado software para escanear vulnerabilidades, monitorearlas y darle seguimiento a los problemas de hackeo ¿cómo considera la administración de este software?
X _{2.12}	12. En términos generales ¿cómo considera las herramientas y medidas de protección?

Tabla 8.3 .Variable Herramientas de protección.
Fuente Elaboración propia

Variable X₃	
Seguridad lógica, física y ambiental	
Simbología de la variable	Pregunta relacionada a la simbología
X _{3.1}	1. ¿Se tiene clasificada la información en términos de su valor, requerimientos legales, sensibilidad, importancia o riesgo para la organización?
X _{3.2}	2. Las copias de los respaldos de información y del software ¿son probados regularmente (plan de contingencias)?.
X _{3.3}	3. La administración de accesos de usuarios en cuanto a su registro, administración de privilegios, de claves de usuario y derechos de acceso.

X _{3.4}	4. ¿Que tan efectivos con los control de acceso a los servicios de telecomunicaciones?
X _{3.5}	5. ¿El control de accesos a los sistemas operativos, se les consideran?
X _{3.6}	6. ¿Los controles de accesos a las aplicaciones y a la información, se le consideran?
X _{3.7}	7. ¿Seguridad de acceso remoto, se le considera?
X _{3.8}	8. ¿La Administración de la Continuidad del Negocio - Plan de Recuperación de desastres- se les considera?
X _{3.9}	9. ¿Mover información con tecnología móvil como discos miniatura, flash drives (universal systems bus USB) se tiene un procedimiento y está restringida?
X _{3.10}	10. ¿Como considera la confiabilidad, disponibilidad e integridad de la información que se maneja en la empresa?
X _{3.11}	11. En términos generales, ¿cómo se considera la seguridad lógica de la empresa?
X _{3.12}	12. ¿cuenta la empresa con adecuados perímetros de seguridad física para proteger áreas restringidas como las de desarrollo, pruebas, e implantación, donde se encuentren los equipos de telecomunicaciones, o sites de cómputo?.
X _{3.13}	13. Según los estándares internacionales la empresa puede contar con uno o varios controles de entrada física para asegurar que únicamente personal autorizado le sea permitido el acceso a áreas restringidas. Puertas con llave Puertas eléctricas Puertas de entrada biométricas Bitácoras electrónicas Credencial con identificación Cámaras de video Guardias de seguridad Puertas dobles Puntos de entrada únicos y controlados Sistema de alarma De acuerdo al tamaño y riesgo del negocio ¿cómo considera los controles físicos que están instalados en su empresa?
X _{3.14}	14. Según los estándares internacionales su empresa puede contar con uno o varios controles ambientales para las áreas restringidas. Paneles de control de alarma Detectores de agua Extintores de fuego manejados manualmente Alarmas de fuego Detectores de humo Sistemas de supresión de fuego Pisos, techos y paredes remitentes al fuego Reguladores de voltaje Generadores de energía interrumpible Materiales de oficina resistente al fuego Planes de evacuación de emergencia Prohibición de comer o beber en áreas restringidas Otras, mencione cuales De acuerdo al tamaño y riesgo del negocio ¿cómo considera los controles ambientales que están instalados en su empresa?.
X _{3.15}	4. En términos generales, ¿cómo se considera la seguridad física y ambiental de la empresa?

Tabla 8.4 Variable Seguridad lógica, física y ambiental..

Fuente Elaboración propia

Variable X₄	Cultura en SI
Simbología de la variable	Pregunta relacionada a la simbología
X _{4.1}	¿Los valores, creencias, lenguaje, nivel educativo, entre otros del personal que forma parte de la empresa? consideran que pueden apoyar al cambio de cultura en seguridad computacional.
X _{4.2}	2. ¿Son respetados los derechos del personal y sus valores?
X _{4.3}	3. ¿Los empleados se sienten seguros de expresar su punto de vista?
X _{4.4}	4. ¿Todos los empleados de la empresa son tratados con respeto?
X _{4.5}	5. ¿Se dan las mismas oportunidades de empleo a mujeres y a hombres sin importar su género, edad, creencias y nivel social?
X _{4.6}	6. ¿Cómo considera en términos la cultura de la empresa?

Tabla 8.5 Variable Cultura en SI.
Fuente Elaboración propia

Variable X₅	Clima organizacional
Simbología de la variable	Pregunta relacionada a la simbología
	Nivel jerárquico y descripción de funciones
X _{5.1}	1. ¿Tienen claras las funciones que se deben de desarrollar en su puesto?
X _{5.2}	2. ¿Se reciben indicaciones claras de parte de los niveles jerárquicos apropiados para llevar a cabo su trabajo?
X _{5.3}	3. ¿Se evalúa al personal a través de objetivos o metas organizacionales previamente establecidas ya sean bimestrales, semestrales o anuales?
	Sueldos y salarios
X _{5.4}	1. ¿Considera que su salario es razonablemente remunerado respecto a la productividad de su trabajo (compensación fija / compensación variable)?
X _{5.5}	2. ¿Considera que los retos que le fueron asignado son adecuados para su capacidad y nivel educacional?
X _{5.6}	3. ¿Está de acuerdo de cómo se manejan las promociones dentro de la empresa?
X _{5.7}	4. ¿Los aumentos que recibe son razonablemente equitativos respecto a la productividad de su trabajo?
	Satisfacción en el trabajo
X _{5.8}	1. ¿Tiene libertad de tomar decisiones independientes en beneficio de la empresa?
X _{5.9}	2. ¿Los logros que superan las metas establecidas por la empresa son reconocidos económicamente y emocionalmente?
X _{5.10}	3. ¿El entrenamiento continuo es una prioridad dentro de la empresa?
X _{5.11}	4. ¿Los diferentes niveles jerárquicos alientan y motivan para que se logren las metas organizacionales?
	Trabajo en equipo
X _{5.12}	1. ¿El trabajar en equipo les da a los miembros un sentido de propósito común?
X _{5.13}	2. ¿El trabajo realizado en equipo es mejor que el trabajo realizado individualmente?
X _{5.14}	3. ¿Cuando se trabaja en equipo se trabaja de una manera más

	creativa y efectiva?
X _{5.15}	4. ¿Cuando se trabaja en equipo se produce un resultado mayor que cuando se trabaja individualmente?
	Comunicación
X _{5.16}	1. ¿Cómo considera la comunicación en la empresa con los niveles superiores, inferiores y con su mismo nivel jerárquico?
X _{5.17}	2. ¿La comunicación organizacional es verdadera y oportuna?
X _{5.18}	3. ¿La información que se comparte es importante y abierta?
X _{5.19}	4. ¿Cómo considera la comunicación verbal y no verbal en la organización?
X _{5.20}	5. ¿Cómo considera en términos generales el clima organizacional de su empresa?

Tabla 8.6 Variable Clima organizacional. .

Fuente Elaboración propia

Preguntas moderadoras

Preguntas moderadoras	
Simbología de la variable	Pregunta relacionada a la simbología
TE 1 2 3 4 5	1. ¿Cuál es el giro de la empresa donde labora? Industria Comercio Servicios Bancario Otros
AE	2. ¿Qué antigüedad tiene operando la empresa?
AIT	3. ¿Qué antigüedad tiene el Área de Informática, de Sistemas, de Tecnología de Información y Procesos de Negocio (o nombre que se le haya asignado al área)?
AST	4. ¿Qué antigüedad tiene el Área de SI, Seguridad Computacional (u otro nombre que se le haya definido)?
E O	5. Tamaño aproximado por número de empleados y obreros Empleados Obreros
H 6.1 H 6.2 H 6.3 H 6.4 H 6.5 H 6.6 H 6.7 H 6.8	6. De las siguientes plataformas de hardware marque ¿Cuál o cuáles están instaladas en su empresa? Mainframes (Computadoras con alta capacidad de procesamiento para realizar trabajos rutinarios) Midrange computers (Computadoras de medio rango) Notebook / laptop computers (Computadoras que no están en estaciones de trabajo) Microcomputers (Computadoras personales) PDA (Asistentes digitales personalizados) Servers (Servidores) Thin client computers (Computadoras que están en estaciones de trabajo) Otro
S 7.1 S 7.2 S 7.3 S 7.4	7. El Software de la empresa es: Desarrollado internamente Comprado Proporciona el servicio empresas de outsourcing Otro

<p>B 8.1 B 8.2 B 8.3 B 8.4 B 8.5 B 8.6 B 8.7 B 8.8</p>	<p>8. ¿Cuál (o cuáles) organización(es) de datos lógica (base de datos) usa su empresa? Jerárquica Red Relacional Multidimensional Orientada a objetos Hipermedia Por archivos Otros</p>
<p>T 9.1 T 9.2 T 9.3 T 9.4 T 9.5 T 9.6 T 9.7 T 9.8 T 9.9</p>	<p>9. ¿Cuál (o cuáles) tipo de redes utiliza la empresa? Red de Área Local (LAN) Red de Área Metropolitana (MAN) Red Branch Exchange (PBX) Red de Área Extendida (WAN) Red de Valor Agregado (VAN) Redes Privadas Virtuales (VPN) Redes Inalámbricas (Wireless Network) Internet (Infraestructura de Internet) Otra</p>

Tabla 8.7 Preguntas moderadoras.
Fuente Elaboración propia

Hipótesis

En este caso de estudio se presentaran dos hipótesis: la nula y alterna, además de orientar el caso de estudio, servirán para confirmar o negar si los resultados de ésta se aproximan a las hipótesis planteadas, guardando siempre las limitaciones y el alcance del estudio.

Hipótesis nula

H₀. La continuidad de operación a través de sistemas computacionales se explica en función de las siguientes variables: estructura organizacional en TI y la función de la SI; herramientas de protección; y condiciones de seguridad lógica, física y ambiental.

Hipótesis de investigación

Hi. La continuidad de operación a través de sistemas computacionales se explica no sólo en función de las siguientes variables: estructura organizacional en TI y la función de la SI; herramientas de protección; condiciones de seguridad lógica, física y ambiental sino también por la cultura en SI y el clima organizacional en la organización.

Modelo propuesto para la prueba de la hipótesis

Para el presente caso de estudio, la variable dependiente es la no disponibilidad de los recursos computacionales (Y1) y las variables independientes serán las relaciones que afectan a la disponibilidad. Es necesario tener en cuenta que las variables son de tipo cualitativo, por lo tanto deben emplearse los estadísticos multivariantes diseñados para el cálculo cuantitativo.

Y: disponibilidad de los recursos computacionales

Xi: Variables independientes

(X1) Estructura organizacional en TI y función de la SI

(X2) Herramientas de protección

(X3) Seguridad lógica, física y ambiental

(X4) Cultura en SI

(X5) Clima organizacional

P (Y): Probabilidad de la ocurrencia de "Y"

El modelo matemático correspondiente, será la demostración de la existencia de una función de probabilidad de la variable de la no disponibilidad de los recursos computacionales respecto a las variables independientes seleccionadas. Simbólicamente

$$P(Y) = f(X1, X2, X3, X4, X5)$$

Para el análisis de datos se utilizó el método estadístico de análisis de regresión múltiple, el cual se considera cuando dos o más variables independientes influyen sobre una variable dependiente, teniendo una fórmula de dependencia funcional del tipo:

$$Y = f(X1, X2, X3, X4, X5)$$

Se presentará primero el método estadístico de análisis de regresión múltiple al desarrollar y explicar el uso de la ecuación de regresión, así como el error estándar múltiple de la estimación. Después se medirá la fuerza de la relación entre las variables independientes, utilizando los coeficientes múltiples de determinación. Para el proceso estadístico se utilizó el paquete informático SPSS (Statistical Package for Social Sciences) versión 12.0.

Para la estimación se utilizó la fórmula:

$$Y' = \beta_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_mx_m + \text{error}$$

No disponibilidad de los sistemas computacionales = f (Estructura organizacional en TI y función de la SI, herramientas de protección, seguridad lógica, física y ambiental, cultura en SI, clima organizacional)

(Y_i) No disponibilidad de los sistemas computacionales = La variable dependiente se define como el porcentaje de tiempo que dejó de operar la empresa a través de sus recursos computacionales, medido en términos de promedio de horas por mes en el último año.

$$(Y_i) = \frac{\text{Tiempo total - horas promedio de paros por mes en el último año.}}{\text{Tiempo total de la operación computacional en el mes}}$$

Elaboración de la encuesta

Después de la selección de las preguntas relacionadas con cada variable se realizó la elaboración de la encuesta. La encuesta definitiva se obtuvo a partir de diversas versiones que fueron modificándose en su contenido inicial hasta conseguir una autorización final que aportase información necesaria según los objetivos del estudio. Se pensó en el diseño que fuese una presentación lógica y ordenada de las diversas preguntas, y sobre todo, un contenido conciso y fácilmente comprensible para las distintas personas que contestaría la encuesta.

La encuesta se construyó a partir de la definición de sesenta y tres preguntas relacionadas a la variable dependiente e independiente y nueve preguntas relacionadas a las preguntas moderadoras, cuyas repuestas permiten deducir todos los indicadores requeridos para la estimación de las variables consideradas según la revisión teórica relacionada con el marco teórico. Se programó la encuesta en el lenguaje html y ésta se subió a un servidor donde los ejecutivos pudieron contestarla en línea. El método utilizado fue el de la escala de Likert (Likert, 1928). Para cada pregunta de la encuesta, exceptuando las preguntas dicotómicas se le pidió al personal entrevistado que contestara seleccionado uno de los cinco niveles de la escala. A cada nivel se le asignó un valor numérico. La escala utilizada fue la siguiente:

Nivel de efectividad	Valor
Poco efectivos	1
Algo efectivos	2
Medianamente efectivos	3
Efectivos	4
Muy efectivos	5

De las 63 preguntas que conformaron el cuestionario, se incluyeron cinco preguntas con categoría dicotómicas, como ya se menciona. Para que facilitara su cumplimiento la encuesta fue dividida en secciones siguiendo una estructura funcional. La estructura de la encuesta se muestra esquemáticamente en la figura 6.1

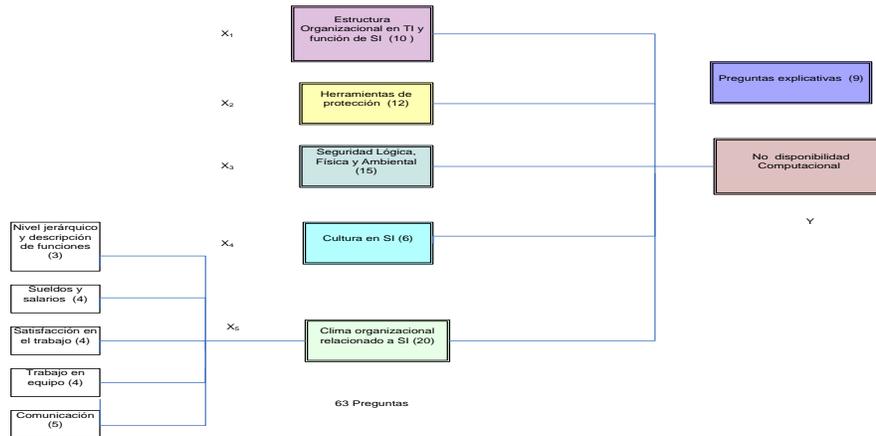


Figura 8.1 Estructura de la encuesta

Recolección de datos

Ya que se tenía la aprobación de la encuesta se procedió a determinar la población, descripción de la muestra, las fuentes de información y el trabajo de campo, así como la selección de la muestra representativa y el envío de la encuesta.

Determinación de la población

Para la determinación de la población objeto de estudio se consultaron dos bases de datos: Una de la Secretaria de Economía del Gobierno Federal y otra del Sistema Empresarial Mexicano.

En una primera etapa, se seleccionaron empresas que tuvieran registrados más de 250 empleados, mismas que son consideradas empresas grandes por ambos organismos. Sin embargo revisando para varias empresas algunas características de interés para nuestro estudio (tales como plataformas de hardware, software, telecomunicaciones y bases de datos entre otros aspectos relacionados a la TI) llegamos a la conclusión de que lo más conveniente para el propósito de nuestra investigación, era utilizar una muestra dirigida restringiéndonos sólo a empresas grandes dedicadas a los sectores Industrial, Servicio y Comercial, ubicados en el Estado de Nuevo León, con lo cual obtuvimos un universo poblacional de 110 empresas que cumplían las características de interés para el estudio.

Descripción de la muestra

El proceso de muestreo se realizó en dos etapas, que de acuerdo a Hernández, Fernández, Baptista, (2000).

- 1) Determinar la muestra sin ajustar
- 2) Calcular la muestra ajustada

Para determinar la muestra sin ajustar, se dividen en dos valores que son aportados por el investigador, considerando las implicaciones significativas que se desprenden del tipo de variable que pretende medir y del margen de error con el que desea proyectar sus resultados. La muestra sin ajustar conocida como n', (ene prima) se evalúa como sigue:

$$n' = \frac{\text{Varianza de la Muestra}}{\text{Varianza de la Población}}$$

Para determinar la muestra del estudio consideramos los siguientes valores

$$n' = \frac{S^2}{V2} = \frac{p(1-p)}{(.015)^2} = \frac{.9(1-.9)}{.000225} = \frac{.09}{.000225} = 400$$

Una vez que hemos determinado la muestra sin ajustar, estamos en condiciones de seguir la segunda etapa, calcular la muestra ajustada. La muestra ajustada se simboliza con la letra n y su fórmula es como sigue:

$$n = \frac{n'}{1 + n' / N} \text{ donde } N = \text{Tamaño de la Población}$$

Así tenemos que:

$$n = \frac{n'}{1 + n' / N} = \frac{400}{1 + 400/110} = \frac{400}{1 + 3.6454} = \frac{400}{4.6363} = 86$$

86 ejecutivos que contesten las encuestas.

Selección de la muestra representativa

Las empresas que fueron seleccionadas pertenecen a los sectores industrial, servicios y comercial. La intención de seleccionar todos los sectores fue de obtener mayor validez externa en la investigación. El destinatario de la encuesta fue el personal de máximo nivel en TI y PN, así como los responsables de SI, en empresas de gran tamaño. Se contactó previamente para explicar el tipo de investigación y solicitar a la persona más idónea para atender el requerimiento de contestar la encuesta.

La investigación tiene un carácter transversal al tratar de establecer cuáles son las posibles razones que justifican los diferentes resultados alcanzados por un conjunto de ejecutivos que están enfocados a la TI y en particular en puestos relacionados a la SI y que según su contestación a la encuesta aseguran que la empresa ha dejado de operar por la no disponibilidad de los recursos computacionales.

La unidad de análisis de este trabajo son los minutos que han dejado de operar las organizaciones por no tener disponibles los recursos computacionales, y las variables contempladas en este estudio son de dos tipos: preguntas moderadoras, relacionadas con aquellos factores generales de las organizaciones y han sido considerados relevantes en el problema planteado según distintas proposiciones teóricas; las otras son variables que se pueden medir los resultados obtenidos por los ejecutivos encuestados.

Trabajo de campo

En este trabajo de investigación se utilizó un método deductivo de carácter cualitativo, las preguntas moderadoras se cuantificaron en relación al tiempo. La fuente de información para el estudio fue una encuesta. Se envió vía Internet a las empresas seleccionadas la liga <http://pro.thunderduck.org/> donde se encontraban las preguntas de la encuesta. Este método de obtención de información se ha considerado más apropiado que otros métodos de campo como la entrevista, dado que facilita la colaboración del encuestado, permiten una mejor elaboración de las respuestas y supone un menor tiempo y costo en su realización.

Una vez obtenida la repuestas vía Internet y después de concentrar y depurar los registros conteniendo las respuestas a las encuestas se procedió a realizar el estudio y análisis orientados hacia la consecución de distintos fines.

Un primer objetivo fue evaluar toda la información obtenida, sintetizando aquellos datos más relevantes y que mejor expresarán las características del perfil del conjunto de organizaciones que han dejado de operar por la no disponibilidad de los recursos computacionales. El primer tratamiento de datos se realizó de manera descriptiva para cada una ocurrencia de la variable dependiente y las preguntas moderadoras, así como también de las posibles interrelaciones que pudieran manifestarse entre ellas.

Ficha técnica de la investigación

Este estudio se realizó durante el período de enero del 2007 a mayo del 2008. En este período se envió la encuesta y también se dio seguimiento a las respuestas, así como se fueron procesando los datos y se definió el método a emplear durante la etapa del análisis. A continuación se presenta la ficha técnica que contiene el resumen de las características del caso práctico realizado..

Unidad de análisis:	
Universo	Organizaciones del Sistema Empresarial Mexicano (sector industria, comercio y servicios) ubicadas en el Estado de Nuevo León y con una plantilla superior a 250 empleados.
Población	110 ejecutivos que contestaran la encuesta
Ámbito geográfico	Estado de Nuevo León
Observaciones contestadas	88
Tamaño de la muestra aceptada	60 casos válidos de un total de 88 encuestas completadas
Tasa de respuesta correctas	68%
Fuentes de información	Encuesta enviada vía Internet
Fecha de trabajo de campo	junio 2007 – mayo 2008
Ámbito temporal de la información	2007-2008
Unidad de análisis	Horas
Sometimiento de la información al análisis	SPSS 12.0.

Tabla 8.8 Ficha técnica de la investigación. Fuente elaboración propia

Análisis de resultados

A lo largo de este estudio de investigación se trató de identificar los factores que determinan o condicionan el éxito de la disponibilidad computacional con relación a la seguridad de información. Desde el planteamiento conceptual, en el presente caso de estudio se ha propuesto identificar los factores que directamente afectan la no disponibilidad. Asimismo, que estos resultaran identificables fácilmente para la explicación de los resultados en relación a: la estructura organizacional en tecnología de información, la función de la seguridad de información, herramientas de protección, seguridad lógica, física y ambiental, cultura en seguridad de información y clima organizacional.

Las respuestas a las preguntas planteadas (costructos) no pueden tener la misma importancia para todos los ejecutivos que contestaron la encuesta, ya que la importancia de cada pregunta es relativa al giro de la organización y al nivel de exposición al riesgo. Así que serán sus riesgos los que determinen las preguntas relevantes para fundamentar el éxito en cada

caso concreto. Por ejemplo, una organización que se dedica a administrar el dinero de sus clientes, tal como un banco, puede influir de manera destacada en los resultados de la encuesta. Es preciso, por lo tanto, adoptar una posición en la que los factores del contexto hacen que unas mismas preguntas sean valiosas para algunos ejecutivos y no lo sean para otros. Considerando lo anterior, se presentan la valoración final y las conclusiones de este proyecto de investigación.

En este apartado se expone el criterio que se siguió para presentar los tres métodos que responden a las hipótesis de investigación, así como su aceptación o rechazo. Se presentan los principales resultados, la contribución de la tesis, las conclusiones que pueden deducirse del estudio de investigación, así como algunas recomendaciones dirigidas a los ejecutivos y a todas las personas interesadas, relacionadas o involucradas con la Tecnología Computacional y la Seguridad de Información. Asimismo se reflexiona sobre las limitaciones de la investigación realizada y se proponen algunas líneas futuras de investigación.

Análisis descriptivo

El objetivo de esta sección es presentar y explicar la información obtenida en la encuesta sobre los distintos aspectos que expresan la realidad de las preguntas moderadoras, así como la variable dependiente y las variables independientes. El tratamiento de la información descriptiva se llevó a cabo de manera univariable y bivariable.

En este apartado se expone el criterio que se siguió para presentar los tres modelos que responden a las hipótesis de investigación, así como su aceptación o rechazo. Se presentan los resultados a detalle tanto para la confiabilidad del estudio como para las preguntas moderadoras, así como para la variable dependiente y las independientes.

Validez de la evaluación general de los modelos

Alfa de Cronbach del primer modelo

Se corrió el primer modelo utilizando las 63 preguntas de la encuesta obteniendo para el Alfa de Cronbach los siguientes resultados:

ΣS_i^2 :	94.56
S_T^2 :	1,684.48
K:	97
$\alpha =$	95.4%

El resultado de la certeza de la consistencia interna para todos los constructos de la encuesta fue del 95.4%, por lo que puede concluirse que se tiene alta o muy alta correlación. Sin embargo al hacer la interpretación de las pruebas estadísticas R² y Anova, no dieron significativas, por lo que se tomó la decisión de ir eliminando preguntas relacionadas a cada variable a través de los resultados del Alfa de Cronbach. Usando la columna "Correc-

ted Item Total Correlation” de la corrida, se analizaron los resultados de cada pregunta y se eliminaron las que tuvieron muy altas correlaciones, dejando las que tuvieron menos correlación. Con dicho análisis se realizó la eliminación de las preguntas quedando únicamente 20. Se muestra el resultado obtenido.

RELIABILITY ANALYSIS - SCALE (ALPHA)

Reliability Coefficients 20 items

Alpha = .9206 Standardized item alpha = .9093

Alfa de Cronbach del segundo modelo

Debido a que al ingresar las 20 preguntas, que quedaron después de la eliminación, no teníamos sumariadas las preguntas por variable, se decidió hacer el análisis del Alfa de Cronbach por variable. Las preguntas relacionadas a cada variable se muestran en la siguiente figura, así como sus valores obtenidos del el Alfa de Cronbach.

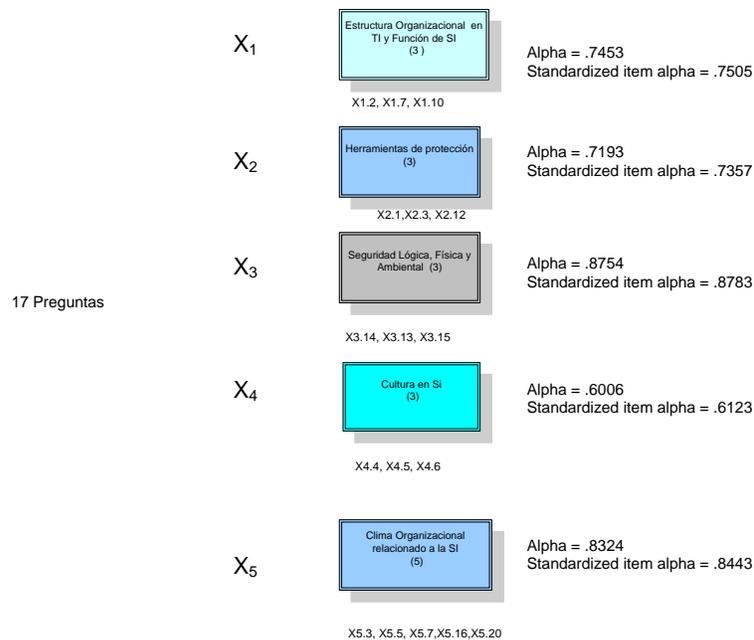


Figura 8.2. Resultados del Coeficiente Alfa de Cronbach por variable

La siguiente tabla resume los valores encontrados.

	X1	X2	X3	X4	X5	SUMATORIA	PROMEDIO
Alpha	0.7453	0.7193	0.8754	0.6006	0.8324	3.773	0.7546
Standardized item alpha	0.7505	0.7357	0.8783	0.6123	0.8443	3.821	0.7642

Tabla 8.9 Resultados del Coeficiente Alfa de Cronbach por variable.

De acuerdo a los valores de la tabla anterior se observa que se tiene un 76.42% donde las variables se relacionan. Esto significa que el modelo tiene una Confiabilidad Sustancial o Marcada, según Ruiz (1998).

Alfa de Cronbach del tercer modelo

Después de haber analizado el segundo modelo, se decidió ingresar al paquete estadístico los valores del Alfa de Cronbach por variable y correr el procedimiento Stepwise. Los resultados obtenidos fueron que sólo la variable X1 (Estructura organizacional en TI y SI) obtuvo valores significativos, por lo que también se corrió el Alfa Cronbach para esta variable. A continuación se muestran los resultados.

Reliability Coefficients 3 items

Alpha = .7453 Standardized item alpha = .7505

De acuerdo a los resultados obtenidos se observa que se tiene un 75.05% donde la variable se relaciona. Esto significa que el modelo tiene una Confiabilidad Sustancial o Marcada, según Ruiz (1998).

Resultados de las preguntas Moderadoras

En relación a las preguntas moderadoras se inició por la revisión de las variables que proyectarán mejor el tipo de la empresa; la antigüedad de la organización; la antigüedad del área de informática de TI y PN, la antigüedad del área de seguridad SI de información, seguridad computacional; tamaño aproximado por número de empleados y obreros; plataformas de hardware; plataformas de software; plataformas de bases de datos; tipo de red o redes que utilizan las empresas. A continuación se presentan un resumen de los resultados más significativos.

- a. El promedio de antigüedad del departamento de informática en las empresas encuestadas fué de 9 años 11 meses.
- b. El promedio de antigüedad del área de SI de las empresas que la tienen es de 4 años y 5 meses. Cabe mencionar que de las 60 empresas que respondieron la encuesta 17 (28%) no han establecido su área de SI, y esto es preocupante porque pueden dejar de operar sus sistemas computacionales.
- c. Se encontró que el hardware más utilizado son "notebook/laptop computers, computadoras que no están en estaciones de trabajo y "servers" (Servidores).
- d. Un 70% de las empresas mencionaron que su software es comprado, sin embargo un 65% comentaron que también era desarrollado internamente.
- e. Las bases de datos del tipo relacional son usadas en un 67% de las empresas encuestadas mientras que las bases de datos del tipo red

las usan un 42%. La menos mencionada en utilización fue la base de datos tipo Hipermedia con sólo 2%.

- f. EL 95% de las empresas encuestadas utilizan redes LAN (Redes de Área Local). Por su parte las redes Inalámbricas (Wireless Network) son utilizadas por un 82%. De acuerdo a la encuesta, el tipo de red menos utilizada fue la VAN (Red de Valor Agregado) ya que su uso fue reportado por sólo el 3% de las organizaciones.
- g. El 27.0% del tiempo total en el que los sistemas computacionales dejaron de operar en las organizaciones fue debido a errores humanos o ignorancia del personal, siendo estas las dos mayores causas entre diversas.

Análisis estadístico

En este trabajo de investigación se trato de averiguar que variables de la no disponibilidad son las que más explican la causa del porque los sistemas computacionales dejan de operar en las organizaciones.

El análisis de regresión lineal fue usado para probar las relaciones entre variables, ya que utiliza una gran gama de modelos de regresión y permite mostrar diferentes modelos para la interpretación de los datos. En este estudio se presentarán tres modelos y se describirán cada uno de ellos y se hará un comparativo para finalmente quedarse con el modelo que más explica la variable dependiente.

Primer Modelo - Descripción del modelo

El primer modelo que se ha sometido prueba, es el que considera la variable dependiente, las 63 preguntas correspondientes a las 5 variables independientes y las 9 preguntas moderadoras.

Las 63 preguntas se sometieron al análisis estadístico. A través del Alfa de Cronbach se analizaron los resultados arrojados por pregunta y así se fueron eliminando si estaban altamente correlacionadas. Se eliminaron 44 preguntas ya que no eran significativas para la evaluación de los factores, quedando la variable dependiente, 15 preguntas relacionadas a las variables independientes y 3 preguntas relacionadas a las preguntas modeladoras y 1 pregunta relacionada la variable dependiente. Las preguntas que fueron seleccionadas finalmente y que mostraban la mejor interpretación se son: LNYT,X1.2,X1.7;X1.10;X2.1,X2.12;X2.3;X3.13,X3.14;X3.15 X4.1,X5.5: X5.3,X5.7;X5.16;X5.20, INDUSTRI, COMERCIO, SERVICIO.

Debido a la gran cantidad de constructos relacionados a las variables, para la explicación individual de las variables se ha sintetizado en cuadros, también se muestran indicadores estadísticos como; R², ANOVA, F, entre otros datos estadísticos.

Resultados de la R²,

Los resultados de la R² de la regresión es de (.578) esto indica que la variancia en la no- disponibilidad de recursos computacionales es explicada

substantialmente por las cinco variables independientes. Es decir, la estructura organizacional, herramientas y medidas de protección, la seguridad lógica, física y ambiental, la cultura computacional y el clima se explica 57.8% por la no disponibilidad de los recursos computacionales.

Model Summary^P

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df 1	df 2	Sig. F Change
1	.760 ^a	.578	.316	1.3077	.578	2.205	18	29	.028

a. Predictors: (Constant), SERVICIO, X5.16, INDUSTRI, X1.2, X5.5, COMERCIO, X1.7, X2.3, X2.1, X4.1, X5.3, X1.10, X3.14, X3.13, X2.12, X5.20, X5.7, X3.15

b. Dependent Variable: LNYT

Tabla 8.10 Resultados de la R y R² -Primer modelo

ANOVA

De acuerdo a los resultados del modelo en su conjunto está soportado significativamente con un valor de la F de 2.205 ($p < .028$). Como a continuación se muestra.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	67.866	18	3.770	2.205	.028 ^a
	Residual	49.589	29	1.710		
	Total	117.456	47			

a. Predictors: (Constant), SERVICIO, X5.16, INDUSTRI, X1.2, X5.5, COMERCIO, X1.7, X2.3, X2.1, X4.1, X5.3, X1.10, X3.14, X3.13, X2.12, X5.20, X5.7, X3.15

b. Dependent Variable: LNYT

Tabla 8.11 Resultados ANOVA- Primer modelo

Por lo anterior, el análisis de varianza, indica que el modelo establece relaciones estadísticamente significativas entre la variable dependiente y las variables independientes a un nivel de significancia del .028 y en este caso el nivel de confianza es de un 95% según los valores críticos.

Resultado de la F

También fue conveniente para comprobar la hipótesis calcular la F 0,05 (teórica) y al 0,01 (teórica) y compararla contra la F calculada dando como resultado el siguiente:

	Valor
F teórica 0.05	1.940
F teórica 0.01	2.570
F calculada	2.205

Tabla 6.12 Resultados de la F - Primer modelo

La F calculada es mayor a la F 0,05 teórica por lo tanto la hipótesis de investigación es aceptada por lo que el modelo es confiable, pero para la f calculada es menor a la F 0,01 teórica por lo tanto la hipótesis de investigación se rechaza.

Significancia del modelo a detalle

Coefficients		Unstandardized Coefficients		Standard	t	Sig.
Model		B	Std. Error	Beta		
1	(Constant)	7.334	1.397		5.250	0.000
	SERVICIO	1.971	0.714	0.573	2.761	0.010
	X5.16	-0.777	0.292	-0.533	-2.663	0.012
	INDUSTRI	-0.475	0.619	-0.119	-0.767	0.449
	X1.2	-0.825	0.271	-0.599	-3.045	0.005
	X5.5	0.018	0.321	0.014	0.055	0.956
	COMERCIO	1.218	0.697	0.304	1.747	0.091
	X1.7	-0.192	0.232	-0.180	-0.827	0.415
	X2.3	-0.233	0.478	-0.108	-0.487	0.630
	X2.1	0.170	0.263	0.119	0.646	0.523
	X4.1	0.342	0.330	0.225	1.037	0.308
	X5.3	-0.557	0.427	-0.277	-1.304	0.202
	X1.10	-0.519	0.416	-0.349	-1.247	0.222
	X3.14	0.450	0.410	0.324	1.099	0.281
	X3.13	-0.071	0.427	-0.048	-0.165	0.870
	X2.12	0.453	0.395	0.275	1.145	0.262
	X5.20	0.320	0.453	0.185	0.705	0.486
	X5.7	0.089	0.460	0.055	0.193	0.848
	X3.15	0.368	0.567	0.257	0.649	0.521

a Dependent Variable: LNYT

Tabla 8.12 Significancia a detalle - Primer modelo

El cuadro anterior presenta el resultado del modelo de coeficientes y el grado de significancia de cada una de ellas. Lo más importante aquí es destacar que en general son pocas las preguntas que tienen un nivel de significancia estadístico adecuado, que normalmente se considera conveniente un nivel del 0.05 o menos, por lo que bajo este criterio solo la pregunta Servicios, X 5.16 y X 1.2 cubrirían este requisito, aunque si se es más flexible y se relaja dicho nivel al 0.09 entonces también la pregunta Comercio la cumpliría este criterio.

Segundo Modelo - descripción del modelo

El segundo modelo fue cuando se sumaron de manera estadística las preguntas correspondientes a cada variable, quedando las variables sumadas como a continuación se muestra.

Preguntas que entraron al análisis		Datos obtenidos de la Columna Corrected Item Total Correlation Alpha
X1	X1.2	.5476
	X1.7	.6339
	X1.10	.5676
X2	X2.1	.5237
	X2.3	.5725
	X2.12	.5665
X3	X3.13	.7927
	X3.14	.6567
	X3.15	.8429
X4	X4.4	.7169
	X4.5	.4949
	X4.6	.5805
X5	X5.3	.5146
	X5.5	.6105
	X5.7	.5875
	X5.16	.7179
	X5.20	.8177

Fórmula que se utilizó para sumarizar las preguntas a cada variable.

$$X_i = \frac{I_1 t_{1i} + I_2 t_{2i} \dots \dots I_n t_{ni}}{i_1 + i_2 \dots \dots i_n}$$

Para i = número de observaciones o encuestas

Cómo se sumaron las pregunta a una variable.

Observaciones	0.5476	0.6339	0.5676	1.7491
	x1.2	x1.7	x1.10	VALOR X1
1	5	5	4	4.68
2	5	4	4	4.31
3	5	5	5	5.00
4	4	0	3	2.23
5	4	4	3	3.68
6	5	4	4	4.31
7	4	5	4	4.36
8	4	3	4	3.64
9	5	4	4	4.31
10	5	4	1	3.34
n	5	5	4	4.68

Ejemplo de sumación de variables

A continuación se presentan los resultados del modelo que resultó de correr los datos sumados a cada variable independiente. Cabe aclarar

que sólo se metieron al modelo la suma de las preguntas significativas relacionadas a cada variable independiente y la variable dependiente.

Segundo modelo

Resultados de la R²

La R2 de la regresión es de (.271) esto indica que la variancia en la no- disponibilidad de recursos computacionales es explicada básicamente por las cinco variables independientes. Es decir, la estructura organizacional, herramientas y medidas de protección, la seguridad lógica, física y ambiental, la cultura computacional y el clima se explica 27.1% por la no disponibilidad de los recursos computacionales.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df 1	df 2	Sig. F Change
1	.520 ^a	.271	.196	1.4407	.271	3.634	5	49	.007

a. Predictors: (Constant), X5SUM, X1SUM, X2SUM, X3SUM, X4SUM

Tabla 8.13 Resultados de la R y R² – Segundo modelo

Lo primero que se puede destacar, es que considerando el total de variables independientes bajo el modelo de regresión lineal propuesto, éstas explican el 27.1% de la variación de la variable independiente.

ANOVA

De acuerdo a los resultados del segundo modelo en su conjunto ésta soportado significativamente con un valor de la F de 3.634 (p < .007). Como a continuación se muestra.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	37.717	5	7.543	3.634	.007 ^a
	Residual	101.711	49	2.076		
	Total	139.428	54			

a. Predictors: (Constant), X5SUM, X1SUM, X2SUM, X3SUM, X4SUM

b. Dependent Variable: LNYT

Tabla 8.14 Resultados de la ANOVA – Segundo modelo

ANOVA

El análisis de varianza, indica que el modelo establece relaciones estadísticamente significativas entre la variable dependiente y las variables independientes a un nivel de significancia del .007.

Resultado de la F

También fue conveniente para comprobar la hipótesis calcular la F 0,01 (teórica) y compararla contra la F calculada dando como resultado el siguiente:

	Valor
	Valor
F teórica 0.05	2.350
F teórica 0.01	3.510
F calculada	3.634

Tabla 8.15 Resultados de la F

Significancia a detalle

Coefficients	Model	Unstandardized Coefficient		Standardizedt	Sig.	% Confidence Interval for		Correlations			Collinearity Statistics		
		B	Std. Error			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF	
	1 (Constant)	6.138	1.213		5.062	0.000	3.701	8.575					
	X1SUM	-0.707	0.245	-0.494	-2.882	0.006	-1.200	-0.214	-0.381	-0.381	-0.352	0.506	1.975
	X2SUM	0.255	0.483	0.085	0.527	0.600	-0.716	1.226	-0.155	0.075	0.064	0.571	1.752
	X3SUM	0.563	0.271	0.362	2.079	0.043	0.019	1.106	-0.091	0.285	0.254	0.490	2.040
	X4SUM	-0.330	0.316	-0.196	-1.046	0.301	-0.965	0.304	-0.354	-0.148	-0.128	0.422	2.369
	X5SUM	-0.363	0.453	-0.176	-0.801	0.427	-1.274	0.548	-0.335	-0.114	-0.098	0.310	3.230

a Dependent Variable: LNYT

Tabla 8.17 Significancia a detalle del segundo modelo

$$Y' = 6.138 + - 0.494 X_{1Sum} + - 0.085 X_{2Sum} + - 0.362 X_{3Sum} + - 0.196 X_{4Sum} + - 0.176 X_{5Sum} + \text{error}$$

Tercer modelo – Stepwise Descripción del modelo

La corrida del tercer modelo se realizó utilizando el procedimiento “Linear Regresión” y “Stepwise” de Regresión Lineal de SPSS.

Tercer Modelo - descripción del modelo

Resultados de la R2

La R2 de la regresión es de (.145) esto indica que la variancia en la no- disponibilidad de recursos computacionales es explicada únicamente por el 14.5% por las cinco variables independientes. Es decir, la estructura organizacional, herramientas y medidas de protección, la seguridad lógica, física y ambiental, la cultura computacional y el clima se explica 14.5% por la no disponibilidad de los recursos computacionales.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df 1	df 2	Sig. F Change
1	.381 ^a	.145	.129	1.4998	.145	8.984	1	53	.004

a. Predictors: (Constant), X1SUM

Tabla 8.17 Resultados de la R y R² . Tercer modelo

ANOVA

De acuerdo a los resultados del modelo segundo en su conjunto ésta soportado significativamente con un valor de la F de 8.984 (p < .004). Como a continuación se muestra.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20.209	1	20.209	8.984	.004 ^a
	Residual	119.219	53	2.249		
	Total	139.428	54			

a. Predictors: (Constant), X1SUM

b. Dependent Variable: LNYT

Tabla 8.18 Resultados de la ANOVA . Tercer modelo

El análisis de varianza, indica que el modelo establece relaciones estadísticamente significativas entre la variable dependiente y las variables independientes a un nivel de significancia del .004.

Resultado de la F

También fue conveniente para comprobar la hipótesis calcular la F 0,01 (teórica) y compararla contra la F calculada dando como resultado el siguiente:

	Valor
F teórica 0.05	4.000
F teórica 0.01	7.080
F calculada	8.984

Tabla 8.19 Resultados de la F

Coefficients	Model	Unstandardized Coefficient		Standardized Coefficient	Sig.	% Confidence Interval for		Correlations	Partial	Part	Collinearity Statistics		
		B	Std. Error			Lower Bound	Upper Bound				Zero-order	Tolerance	VIF
	(Constant)	5.709	0.694		8.227	0.000	4.317	7.101					
	X1SUM	-0.545	0.182	-0.381	-2.997	0.004	-0.909	-0.180	-0.381	-0.381	-0.381	1.000	1.000

a Dependent Variable: LNYT

Tabla 8.20 Resultados de la significancia – Tercer modelo

$$Y' = 5.709 + - 0.381 X1Sum + e$$

(X5) Clima organizacional

Interpretación del modelo que fue seleccionado – Segundo Modelo

De lo anterior podemos concluir, que el segundo modelo es el más adecuado para el caso de estudio, ya que con él se explican dos importantes variables que tienen significancia y que son: (X1) — en TI y función de la SI — y (X3) — Seguridad lógica, física y ambiental —. Ya que la significancia de las dos variables es $< .01$ podemos decir que las variables del modelo son significativas al 99%.

En la tabla anterior puede observarse que la variable Cultura en Seguridad de Información se encuentra en tercer lugar del modelo en nivel de significancia, el Clima Organizacional en el cuarto y las Herramientas de Seguridad en el quinto lugar.

Considerando lo anterior sería importante que para asegurar el éxito de una nueva Cultura en SI, se debería analizar, diseñar, implementar y controlar un programa de cambio, el cual debería ser llevado a cabo sistemáticamente, más que por etapas. Estudios futuros deberían ser dirigidos hacia la discriminación de la cultura y explicar por qué las TI funciona mejor o peor en diferentes culturas. Independientemente de las dimensiones de la cultura de un país en sí misma puede haber políticas regulatorias, y factores políticos y demográficos también en juego y pueden afectar a la cultura.

Así como el Clima organizacional es en gran medida establecido por las acciones de la alta dirección y administración de la empresa, para refor-

zar el mensaje de la organización, el consejo de directores y a la alta administración, deben ser como modelos a seguir y demostrar un fuerte compromiso con una conducta ética.

Según el estudio, las Herramientas de Seguridad son la parte donde ponen mayor atención los responsables de la TI ya que están contratando personal técnico competente para que seleccione, implemente y monitoree las herramientas de SI, con base al riesgo que quiera tomar la empresa.

Comprobación de las hipótesis de investigación

Lo que nos habíamos planteado como hipótesis de investigación fué:

H1. La continuidad de operación a través de sistemas computacionales se explica no sólo en función de la estructura organizacional en tecnología de información y área de seguridad de información, de herramientas y medidas de protección, de condiciones de seguridad lógica, física y ambiental, sino también por la “cultura en seguridad de información” y el “clima organizacional de la empresa”.

De acuerdo a los resultados del modelo, en su conjunto, está soportado significativamente con un valor de la F (Fisher) de 3.634 ($p < .007$) como a continuación se muestra.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	37.717	5	7.543	3.634	.007 ^a
	Residual	101.711	49	2.076		
	Total	139.428	54			

a. Predictors: (Constant), X5SUM, X1SUM, X2SUM, X3SUM, X4SUM

b. Dependent Variable: LNYT

Tabla 8.21 Resultado de ANOVA - Modelo seleccionado

Al realizar la prueba F, se encontró que la $f_c 3.634 > f_t 3.510$ para el nivel de significancia 0.01 y $f_c 3.634 > f_t 2.450$ para el nivel de significancia 0.05. La razón de F calculada que resulta mayor en ambos valores, es significativa tanto al nivel 0.01 como al nivel 0.05, por lo tanto, en ambos niveles es aceptada la hipótesis de investigación.

Asimismo, al realizar las pruebas para cada una de las variables, encontramos que (X1) — Estructura organizacional en TI y función de la SI — y (X3) — Seguridad lógica, física y ambiental — son aceptados, por lo que la continuidad de los sistemas computacionales es son explicadas en función de las siguientes variables: la estructura organizacional en tecnología de información y función de la SI; las herramientas y medidas de protección; de las condiciones de seguridad lógica, física y ambiental; así como de la cultura computacional y el clima organizacional. Por lo que se acepta la hipótesis de investigación.

Análisis de las variables independientes del segundo modelo

Con base a los resultados del segundo modelo, que se presenta a continuación, podemos manifestar las siguientes conclusiones:

Model	Unstandardized Coefficients		Standardized Coefficients		Convertido en min	Hrs.	min	con min	tiempo
	B	Std. Error	Beta						
1	(Constant)	6.138	1.213						
(X1) Estructura organizacional en TI y función de la SI	X1SUM	-0.707	0.245	-0.494	494	8.23	0.23	0.138	8.13
(X2) Herramientas de protección	X2SUM	0.255	0.483	0.085	85	1.42	0.42	0.252	1.25
(X3) Seguridad lógica, física y ambiental	X3SUM	0.563	0.271	0.362	362	6.03	0.03	0.018	6.01
(X4) Cultura en SI	X4SUM	-0.33	0.316	-0.196	196	3.27	0.27	0.162	3.16
(X5) Clima organizacional	X5SUM	-0.363	0.453	-0.176	176	2.93	0.93	0.558	3.56
2	(Constant)	3.719	0.217						
a	Ident Variable: LNYT								

Tabla 8.22 Del análisis de las variables independientes del segundo modelo

Para (X1) — Estructura organizacional en TI y función de la SI. La empresa podría dejar de operar en un promedio hasta en 8 horas y 13 minutos, en un período de tiempo de un año.

Para (X2) — Herramientas de seguridad — Si la empresa tiene bien definidas estas herramientas de seguridad pueden quedar fuera de operación hasta 1 hora 25 minutos, en un período de un año.

Para (X3) — Seguridad lógica, física y ambiental — El riesgo de dejar de operar en 6 horas un minuto, en un período de un año.

Para (X4) — Cultura en Seguridad de Información —. Si existe una mala cultura en seguridad de información la empresa podría dejar de operar por 3 horas con 16 minutos, un período de tiempo de un año.

Para (X5) — Clima Organizacional — Si existe un clima organizacional malo la empresa podría dejar de operar en sus recursos computacionales por 3 horas 56 minutos, en un período de un año.

Aceptación o rechazo de la hipótesis nula

Para aceptar o rechazar la hipótesis nula se realizaron las pruebas estadísticas correspondientes utilizando las 3 variables que intervienen en ella, así como la variable dependiente.

(X1) — Estructura organizacional en TI y función de la SI.

(X2) — Herramientas de seguridad.

(X3) — Seguridad lógica, física y ambiental

Obteniéndose los siguientes resultados:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.448 ^a	.201	.154	1.4779	.201	4.277	3	51	.009

a. Predictors: (Constant), X3SUM, X2SUM, X1SUM

Tabla 8.23 Resultados de la R y R²

Comprobación de las hipótesis de investigación

Lo que nos habíamos planteado como hipótesis de investigación fué:

H1. La continuidad de operación a través de sistemas computacionales se explica no sólo en función de la estructura organizacional en tecnología de información y área de seguridad de información, de herramientas y medidas de protección, de condiciones de seguridad lógica, física y ambiental, sino también por la “cultura en seguridad de información” y el “clima organizacional de la empresa”.

De acuerdo a los resultados del modelo, en su conjunto, está soportado significativamente con un valor de la F (Fisher) de 3.634 ($p < .007$) como a continuación se muestra.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	37.717	5	7.543	3.634	.007 ^a
	Residual	101.711	49	2.076		
	Total	139.428	54			

a. Predictors: (Constant), X5SUM, X1SUM, X2SUM, X3SUM, X4SUM

b. Dependent Variable: LNYT

Tabla 8.24 Significancia a detalle

Al realizar la prueba F, se encontró que la $f_c 3.634 > f_t 3.510$ para el nivel de significancia 0.01 y $f_c 3.634 > f_t 2.450$ para el nivel de significancia 0.05. La razón de F calculada que resulta mayor en ambos valores, es significativa tanto al nivel 0.01 como al nivel 0.05, por lo tanto, en ambos niveles es aceptada la hipótesis de investigación.

Asimismo, al realizar las pruebas para cada una de las variables, encontramos que (X1) — Estructura organizacional en TI y función de la SI — y (X3) — Seguridad lógica, física y ambiental — son aceptados, por lo que la continuidad de los sistemas computacionales son explicadas en función de las siguientes variables: la estructura organizacional en tecnología de información y función de la SI; las herramientas y medidas de protección; de las condiciones de seguridad lógica, física y ambiental; así como de la cultu-

ra computacional y el clima organizacional. Por lo que se acepta la hipótesis de investigación.

Contribuciones

Este caso de estudio “Factores que determinan el grado de disponibilidad de los sistemas computacionales (SC)”, hace varias contribuciones significativas, tanto para la teoría como para el conocimiento del status de las empresas en relación al tema de investigación, asimismo produce recomendaciones de aplicación práctica para el mejor desempeño de las organizaciones que utilizan sistemas computacionales en su operación. Entre los hallazgos y aportaciones se mencionan:

- a. El análisis descriptivo realizado sobre los diversos aspectos sobre TI y PN y SI, permiten afirmar que se trata de un campo teórico y práctico muy dinámico y con actualización constante, ya que al mismo tiempo que la tecnología computacional avanza los riesgos y las amenazas a los que están sometidas las organizaciones también se modifican.
- b. Los aspectos de seguridad de información vienen manifestando en los últimos años intensos y profundos cambios. Henry (2005) menciona que las vulnerabilidades en aplicativos y sistemas operativos, reportados en la revista BugTraq, alcanzan niveles hasta de 100 nuevos métodos de ataque cada mes. Esto nos hace recomendar una actualización continua del personal relacionado del personal relacionado con Seguridad de Información para contrarrestar estos ataques.
- c. Las nuevas tecnologías han sido inventadas para innovar las estrategias de los negocios. Hoy por hoy a la Web se le ha nombrado como el lugar de mercadotecnia por los investigadores. Sin embargo debe considerarse que con la utilización de la Web, se han generado muchos problemas de seguridad computacional. Cada día una nueva noticia destaca la importancia de la seguridad en las redes corporativas que han sido quebrantadas. Los sistemas de información son accesados por individuos no autorizados o las identidades son robadas y usadas para realizar transacciones fraudulentas. Como resultado de ello, tanto los negocios como los gobiernos, están evaluando e implementando nuevos sistemas de administración y manejo de identidades para proveer un acceso lógico más seguro.
- d. Una sólida autenticación para acceso lógico, requiere del uso de múltiples factores de autenticación. La tecnología de tarjetas inteligentes – por lo regular en conjunto con un PIN para abrir la tarjeta – es usada cada vez mas, para ofrecer el segundo o tercer factor crítico de autenticación que hace el acceso lógico más seguro.

- e. Las vulnerabilidades no son debilidades o ausencias de controles, es reconocer que aún no terminamos de conocer las diferentes deficiencias de las plataformas que se administran, así como las aplicaciones que en ellos se instalan. Es claro que esta posición debe ser fruto de un esfuerzo continuo y dedicado para entender y modelar las posibles situaciones de falla, y no una posición cómoda que sólo reacciona ante un incidente. Ser vulnerable, es reconocer que se tiene que aprender todos los días, es abrir la mente a nuevas posibilidades, es mirar en la inseguridad de la información la fuente de proponer controles. Renunciar a aprender de la inseguridad, es enterrar el futuro de la seguridad. Es negar la dinámica misma de los desarrollos que se pueden dar, entendiendo las nuevas formas de protección, que no son otra cosa que la comprensión de la inseguridad materializada en una técnica, una herramienta o una propuesta.
- f. La tecnología computacional es un elemento crucial para movilizar el capital social y para la creación del nuevo conocimiento a través de la liga de sistemas de información y comunicación en una organización. Pueden ser integrados flujos previamente fragmentados de información y conocimiento. Estas ligas pueden también eliminar barreras de comunicación, que naturalmente ocurren entre diferentes partes de la organización. Invertir en una amplia infraestructura, que soporte los varios tipos de conocimiento y comunicación son factores críticos. Las dimensiones tecnológicas que son parte de la efectiva administración del conocimiento incluyen, inteligencia de negocios, colaboración, aprendizaje distribuido, descubrimiento de conocimiento, mapeo de conocimiento, así como seguridad.
- g. Con frecuencia, el cambio es el resultado no sólo de avances en la tecnología sino de cambios que ocurren dentro de la cultura organizacional de la corporación. La tecnología es solamente la mitad de la ecuación que trae valor de la compañía, y la cultura es con mucho la otra mitad. Usualmente la resistencia al cambio se topa con la tradicional actitud de “hemos realizado las cosas de esta manera y han funcionado, así qué para que molestar con los cambios”.
- h. Si una organización no invierte en las herramientas mínimas de seguridad, en personal administrativo y técnico para que esté operando estas herramientas y colocando controles y si sus operaciones están siendo procesadas por sistemas computacionales, ésta puede salir del mercado. Como parte del movimiento hacia un enfoque más integral hacia la seguridad, un número creciente de compañías están asignado las responsabilidades tanto de la seguridad lógica como física bajo una sola persona, tal como un Oficial Principal de Seguridad.

Comentarios generales

Las explicaciones proporcionadas desde las perspectivas y enfoques que se han venido sucediendo con el avance tecnológico computacional y la teoría sobre la seguridad de información, han tenido diferentes enfoques y puntos de vista. La no disponibilidad de los recursos computacionales se ve afectada por dos tipos de personas, unas internas y otras externas, ambas pueden intencionales o no intencionales dañar los recursos computacionales. Cabe aclarar, que el comportamiento tanto de las personas internas como externas, está fuera de cualquier herramienta instalada de seguridad de información o política que se haya implementado, ya que el comportamiento de los seres humanos es impredecible.

Para minimizar el riesgo de un ataque por parte de las personas internas o externas, es indispensable que se analicen, diseñen, implementen y actualicen soluciones administrativas tales como políticas y procedimientos, y que éstas se difundan mediante capacitación y campañas de concientización. De igual manera tales soluciones técnicas como son: antivirus, firewalls, o bloqueo de discos extraíbles, entre otros. Las soluciones técnicas, a través de herramientas deberán ser adecuadamente seleccionadas, implementadas y monitoreadas, de acuerdo a las plataformas de hardware, telecomunicaciones, bases de datos, sistemas operativos, o sistemas en aplicación, entre otras consideraciones. Las herramientas se podrán seleccionar en base a lo que la empresa esté dispuesta a invertir o al nivel de riesgo que están dispuestos a permitir, pero siempre debe de existir la consigna de que el riesgo debe ser minimizado. Si las personas internas toman conciencia, ayudarán a reducir los gusanos, virus, spyware y otras amenazas. Así sería menos probable que se infecten los recursos computacionales. Asimismo, hay que reforzar el seguimiento sobre el comportamiento de algunos usuarios que tienen autoridades especiales y que manejan información confidencial o que tienen los privilegios de modificar, borrar o cambiar información.

Hay que verificar los antecedentes laborales y la trayectoria profesional para todos las área clave de la organización, particularmente parte del departamento de TI y PN y SI. Para estas últimas, deberá reforzarse el monitoreo, debido que éstas áreas están más familiarizadas con las vulnerabilidades sobre aspectos de seguridad computacional, y son las que pueden intencionalmente o no causar más daño.

Líneas futuras de investigación

Las limitaciones intrínsecas a todo proyecto abren distintas líneas de trabajo que pudieran permitir progresar en el propósito de la investigación planteada, más allá de los resultados obtenidos con la labor desarrollada hasta estos momentos. Algunas de las restricciones manifestadas pueden eludirse, al menos parcialmente, con nuevas extensiones en el tratamiento del problema estudiado.

En relación a nuevas líneas de investigación sobre "El Futuro de la Seguridad de Información". Pueden mencionarse:

- a. Una línea de investigación alternativa consistiría en ampliar el escenario o ámbito para la aplicación de la encuesta, concretamente que ésta sea aplicada a nivel nacional o internacional. Esta posibilidad contribuiría a tener un mayor universo de la investigación en varios aspectos, por un lado, permitiría tener una muestra más amplia de empresas u organizaciones que contestaran la encuesta y por otro lado contar con resultados a nivel nacional o global.
- b. Investigar con base a los resultados de esta y otras tesis la construcción de una efectiva y eficiente infraestructura o arquitectura de seguridad de información para hacer que la continuidad computacional relacionada a la seguridad de información no tenga problemas. Es importante la implementación de herramientas tecnológicas soportadas con aspectos administrativos con objeto de disminuir ataques e incrementar la seguridad de información.
- c. Abrir líneas de investigación interdisciplinarias, es decir, involucrar a todos los sectores de la tecnología de información y procesos de negocio, y seguridad de información como por ejemplo, el sector médico, sector aeronáutico, sector de nano-tecnología. Así como también, se deben de realizar líneas de investigaciones interdisciplinarias, es decir, involucrar a todos los sectores de la tecnología de información y procesos de negocio, y seguridad de información como por ejemplo, el sector médico, sector aeronáutico, sector de nano-tecnología, que estos sectores no fueron considerados en la muestra, ya que la seguridad de información en estos sectores es crítica. En un estudio del futuro tecnológico expuesto por Alois Ferscha (2006) sugirió seis grandes grupos que son: informática y comunicaciones; nano - tecnología, seguridad de información, sinergias en los sistemas, sistemas inteligentes y software sistemas intensivos, en los cuales también nos podríamos soportar para las futuras investigaciones.
- d. Desarrollo software para seguridad de información por empresas mexicanas. No obstante, también existen fuertes rezagos por superar: según los resultados obtenidos de la encuesta sobre el software que utilizan las empresas muestran que un 70% del software aplicativo es comprado. Pero cabe mencionar, que tal vez el software para seguridad de información no se esté generando en nuestro país, entonces esta podría abrir otra línea de investigación, que estamos haciendo al respecto para generar software sobre seguridad de información, o se lo estamos dejando a países como China, India o Irlanda para que ellos los realicen.
- e. Métodos de ataque relacionados a la seguridad de información. Otra línea de investigación podría ser, los diferentes métodos de ataque se están saliendo todos los días y que pueden dejar no disponibles los sistemas que usan tecnología computacional, llámese

computadoras, teléfonos celulares, pdas, entre otros y las empresas que están generando productos (software) de seguridad de información, ya que en los próximos años entrarán en una etapa de crecimiento constante debido a que cada día se generan nuevos métodos de ataque y estos software's podrían contrarrestar el riesgo para corporaciones y empresas. Todo virus o método de ataque para las compañías que se dedican a generar software o suite de seguridad se traduce en dólares.

Referencias

Ardita, Julio César. (2001). Security System y ex-Hacker. Entrevista personal realizada el día 15 de enero de 2001 en instalaciones de Cybsec S.A. <http://www.cybsec.com>

Ashforth, B. (1985). Climate formation. Issues and extensions. *Academy of Management Review*.

Beekman, George. (2005). *Introducción a la Informática*. Prentice-Hall.

Cano, J. (2006). Information Insecurity: A dual concept of information security. *Proceeding of 2nd Internacional Conference on E-Global Security*. Londres, UK. Abril.

CISA (2000). *Information Systems Audit and Control Association. Certified Information Systems Auditor. Review Technical Information Manual*.

CISA Review Manual (2005). *The Information Systems Audit and Control Association*.

CISA Review (2009). *Certified Information Security Management, Information Systems Audit and Control Association*.

Dell, C and Grayson, C.(1998). If only we know what we know: identification and transfer of internal best practices. *California Management Review*.

Ewing, John; Falcon, Jeff and McGrane, Kyle. (2007). IT Security: Preventing the March of Madness *Business Communications Review*; p. 30, ABI/INFORM GLOBAL.

Gibson, Ivancevich Donnelly. (2001). *Las Organizaciones: Comportamiento, Estructura, Procesos*. Editorial Mc Graw Hill. Décima Edición. Chile.

Gold, Andrew H., Malhotra, Arvind and Segars. (2001). Knowledge Management: Organizational Capabilities Perceptive. *Journal of Management Information Systems*. ABI/INFORM Global.

Guzley, Ruth M. (1992). *Organizational Climate and Communication Climate, Predictors of Management Communication Quality*, MCQ.

Hall, Richard. (1996). *Organizations: Structures, Processes and Outcomes*.

Harley, Bill, and Wright, Christopher. Hall, Richard. and Dery, Kristone. (2006). *Management Reactions to Technological Change: The Example of Enterprise Resource Planning*. The Journal of Applied Behavioral Science. ABI/INFORM Global.

Herold, Rebecca (2005). *Managing an Information Security and Privacy Awareness and Training Program*. Auerbach Publications; 1 edition

Hofstede, Geert. (2001). *Culture's Consequences y Software of the Mind*.

Hofstede, Geert. (2001). *Culture's Consequences, Comparing Values, Behaviors, Institutions, and Organizations Across Nations* Thousand Oaks CA: Sage Publications. - for every detail on Hofstede's research.

Kilmann, Ralph. (1985). "Five Steps for Closing Culture-Gaps," www.kilmann.com/articles.html.

Lucena López, Manuel José. (1991). *Criptografía y Seguridad en Computadores*. Dpto. de Informática Universidad de Jaén. Edición virtual. España. 1999. <http://www.kriptopolis.org>. Capítulo 2-Página 23.

Mao, En and Prashant, Palvia. (2006). *Testing an Extended Model of IT Acceptance in the Chinese Cultural Context*. Database for Advances in Information Systems. ABI/INFORMATION Global.

Maslow, AH. (1954). *Motivation and Personality*.

Mintzberg H. (1998). *Covert leadership: notes on managing professionals*. Knowledge workers respond to inspiration, not supervision.

Muchinsky, PM. (1977). *Organizational communication: Relationships to organizational climate and job satisfaction* - Academy of Management Journal.

Nakagawa M. A. (1996). "A Closer Look at Culture". Edit. Gránica S.A. p. 6. Citado por Girard, G. / Koch S.J..

Narasu, Rebbapragada. (2006). *All-in-One Security Suites*. Answers questions about security and privacy during her live Web chat.

Narasu, Rebbapragada. (2006). *All-in-One SECURITY*, PC World; 24, 7.

Nonaka, I., and Takeuchi, H. (1995). *The knowledge Creating Company How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press.

Palmer, Lan; Dunford, Richard and Akin, Gib. (2006). *Managing Organizational Change: a Multiple Perspectives Approach*. McGraw-Hill/Irwin.

Peltier, Thomas R (2006). *Principles of Information Security*. Third Edition builds on internationally recognized standards and bodies of knowledge to provide the knowledge and skills that information systems students need for their future roles as business decision-makers.

Pritchard, R. and Karasick, B. (1973) The Effect of Organizational Climate on Managerial Job Satisfaction. *Organizational de la Republic Behavior and Human Performance*.

Purser, Steve. (2004). *A Practical Guide to Managing Information Security* (Artech House Technology Management Library).

Rainer, R Kelly Jr. Marshall, Thomas E. Knappi, Kenneth J. y Montgomery, Gina. (2007). Do Information Security Professionals and Business Managers View Information Security; Issues Differently. *Information Systems Security*; 16,2 ABI/INFORM.

Scott, Berinato. Analysis: 2007 Global State of Information Survey. Analysis: 2007 Global. http://www.csoonline.com/read/100107/fea_innocence.html

Stanley, T.L. (2007). Being able to logically think through and plainly discuss a problem and a resolution is critical. Hire the right person – ABI/INFROM Gobal.

Stanley, T.L. (2007). Hire the right person *Super Vision*; 68, 7; ABI/INFORM Global.

Steers, Richard M. and Porter, Lyman W. (1983). *Motivation and work behavior*. PUB TYPE: Book (ISBN 007060942X). VOLUME/EDITION: 3rd edition.

Stilianos, Vidalis and Zafar, Kazmi. (2007). *Security Through Deception*, *Information Systems Security*, Jan/Feb:

Swing, John; Falcon, Jeff and McGrane, Kyle. (2007). *La tecnología de Redes Privadas Virtuales – Virtual Private Network (VPN)*.

Tagiuri, R., Litwin, G. (1968). *Organizational climate: Explorations of a concept*. Boston: Harvard Business School, Division of Research.

Taylor JR and Bowers DG. (1972). *Survey of Organizations: Supervisory and peer leadership*.

Thompson, S.H. Teo. (2007). *Journal of Global Information Management*, 15(2) 1-117. *Organizational Characteristics, Modes of Internet Adoption and Their Impact: A Singapore Perspective* National University of Singapore, Singapore.

Toffler, Alvin and Toffler, Heidi. (1980). *Creating a New Civilization: The Politics of the Third Wave* (Paperback).

Vallabhaneni, Rao. (2000). *Cisa Examination Textbooks, Volume 1: Theory*, RRV Professional Publications, Illinois.

Venetis, Tom. (2005). *Change is cultural as much as Technological*. *Canadian Plastics*. ABI/INORM Global

Weinert, B. (1987). *Manual de Psicología de la Organización*. Barcelona: Herder.

Whitman, Michael E. y Mattord, Herbert J. (2007). Information Security Policies and Procedures: A Practitioner's Reference, Second Edition illustrates how policies and procedures support the efficient running of an organization. Course Technology; 3 edition.

Williams, T Craig; Vistakon, Rains Johnson Juliet; and Vistakon, Hohnson Johson. (2007). Linking Strategy to Structure: The Power of Systematic Organization Design. Organization Development Journal; 25, 2; ABI /INFORM Global.

Otras páginas web consultadas

www.eumed.net/tesis/cgb/c1.pdf

Índice del capítulo

- (si)*, 378
- CLIMA ORGANIZACIONAL, 391, 397, 401, 402, 404, 420
- CULTURA EN SEGURIDAD DE INFORMACIÓN, 388, 401
- cultura en si, 397, 404, 420
- era de la información y del conocimiento, 379
- era industrial, 378, 379
- estructura organizacional en ti y función de la si, 404
- estructura organizacional en ti y la función de la si, 397
- hardware, 380, 385, 386, 396, 397, 402, 406, 412, 426
- HERRAMIENTAS DE PROTECCIÓN, 384, 397, 398, 399, 404
- oficial principal de seguridad (cso), 382
- oficial principal de información (chief information officer cio), 382
- preguntas moderadoras, 402, 403
- seguridad de información, 378, 397, 401, 410, 420, 422, 424, 427
- seguridad de información. si, 378
- SEGURIDAD LÓGICA, FÍSICA Y AMBIENTAL, 386, 397, 399, 400, 404, 420, 421, 422, 423
- software, 380, 385, 386, 387, 388, 396, 397, 399, 406, 412, 427, 428
- tecnologías de información (ti) y procesos de negocio (pn), 378
- telecomunicaciones, 380, 385, 387, 388, 396, 397, 400, 406, 426