

MODELO DE APLICACIÓN DE HERRAMIENTAS ESTADÍSTICAS EN LA INVESTIGACIÓN

Dr. Manuel Prieto de Hoyos
Profesor de tiempo completo
Dirección de Carreras de Informática
Facultad de Ingeniería y Arquitectura
Universidad Regiomontana
mprieto@mail.ur.mx

Resumen

Se presenta un modelo para establecer una estrategia de selección y aplicación de herramientas estadísticas en la investigación, clasificando las hipótesis y posteriormente sugiriendo herramientas para comprobar los diferentes tipos de hipótesis, se presentan ejemplos sencillos de los diferentes casos. Las herramientas y/o los cálculos cubiertos son: Distribución Normal aplicaciones generales, Cálculo del tamaño de las muestras, Coeficiente de consistencia interna Alfa de Cronbach, Regresiones lineales simple y múltiple, Análisis de conglomerados (*clusters*), comparación de dos grupos con las pruebas t-Student y Mann Whitney y Comparación de más de dos grupos utilizando el Análisis de Varianza o ANOVA.

1. Introducción

1.1. El problema

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 los resultados de la investigación en proceso, un riesgo definitivamente muy serio para cualquier tipo de estudio (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 1.1 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 cuándo solo se requiere medición de parámetros sencillos.	Falta de análisis de herramientas estadísticas sofisticadas generalmente por la sobre abundancia de las mismas.

Tabla 1.1 Problemas más frecuentes en el uso de las herramientas estadísticas

Los errores disminuyen y los resultados elevan su calidad si planteamos un modelo para llevar a cabo el plan de trabajo estadístico de una investigación. Este es precisamente el objetivo de este estudio, plantear un modelo en el que se base la aplicación de las distintas herramientas estadísticas en una investigación. En resumen la hipótesis de este estudio puede declararse como sigue:

Los errores estadísticos pueden prevenirse e incluso evitarse si se sigue un modelo de trabajo como el que se propone para seleccionar y aplicar las herramientas estadísticas.

1.2. Modelo Propuesto de Aplicación de Herramientas Estadísticas

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

Proponemos también la clasificación de los estudios dependiendo del número de conjuntos de observaciones que involucran esto nos ayudará a establecer el tipo de hipótesis que surge de cada estudio y a seleccionar las distintas técnicas que se pueden aplicar. Considerando lo anterior podemos clasificar los estudios en:

Estudios que involucran un solo conjunto de observaciones, estos estudios son tipo encuesta y cuentan con una medición a una muestra de una población.

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.

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)

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 o 3, ver la sección anterior.
- 3.-Clasificar la hipótesis considerando lo planteado anteriormente.
- 4.-Seleccionar la técnica estadística que se aplicara en el estudio de acuerdo a la clasificación de la hipótesis.

El modelo nos proporciona una guía rápida para establecer un rumbo en la utilización de las herramientas estadísticas marca la dirección hacia que técnicas 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 ya errores mencionados.

2. Las Herramientas y los Cálculos.

2.1. La distribución Normal de un conjunto de observaciones.

Los cálculos relacionados con una distribución normal se denominan cálculos de estadística paramétrica. 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.

2.2. 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.

Tenemos que para una distribución normal podemos calcular 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 y el error estándar es SE , la media poblacional μ viene dada por: $\mu = \bar{X} \pm Z*SE$ (Ecuación 2.1) substituyendo el valor del error estándar por su expresión en función de la desviación estándar de la muestra $SE = S/\sqrt{n}$ y sabiendo que S (la desviación estándar de la muestra) puede fácilmente calcularse. n el tamaño de la muestra es conocido y \bar{X} promedio de la muestra puede también calcularse la ecuación anterior nos queda como: $\mu = \bar{X} \pm 1.96* S/\sqrt{n}$ (Ecuación 2.2)

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 $\mu = \bar{X} + \text{tolerancia}$ (Ecuación 2.3) podemos combinar esta ecuación con la 2.2 obteniendo la siguiente igualdad: $\bar{X} \pm \text{tolerancia} = \bar{X} \pm 1.96* S/\sqrt{n}$ ya que μ de la Ecuación 2.2 = μ de la Ecuación 2.3 Y de aquí podemos encontrar que: $\text{tolerancia} = 1.96* S/\sqrt{n}$ despejando obtenemos la ecuación 2.4 que define el tamaño de la muestra válida para calcular una media de una población (Lohr, S. L. 1999). $n = 1.96^2 * S^2/\text{tolerancia}^2$ (Ecuación 2.4).

Recordando que la constante 1.96 nos proporciona un 95% de certeza y tenemos la tabla 2.1 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.

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

Tabla 2.1 valores de Z para diferentes porcentajes de certidumbre, estos valores se obtienen de las tablas de la distribución normal.

2.3. 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). A continuación enlistamos los 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 calculo de la varianza (cuadrado de la desviación estándar) y el segundo el cálculo de las correlaciones de Pearson (Briones, G. 2001). En este estudio usaremos el primer procedimiento. La formula para calcular el coeficiente es:

$$\alpha = \frac{k}{(k-1)} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right) \quad \text{Ecuación 2.5}$$

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. Utilizando una hoja de cálculo podemos aplicar la Ecuación 2.5 para calcular el coeficiente como muestra la figura 2.1.

	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 2.1 Cálculo del coeficiente aplicando la ecuación 2.5 en una hoja de cálculo

Es posible utilizar el SPSS sobre los mismos datos y obtener las 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*, se obtiene el resultado mostrado en la tabla 2.2, el coeficiente es aceptable para todas las preguntas.

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 2.2 Resultado del SPSS al calcular el alfa de Cronbach de los datos mostrados en la figura 2.2

Este procedimiento (Thomas, H.P. 2004) es el más utilizados para llevar a cabo el análisis de confiabilidad mediante el análisis de consistencia interna de un instrumento.

2.4. Relaciones entre variables: Regresión Lineal Simple y Múltiple.

La regresión implica una relación causal entre dos variables .

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”. 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 $y = m x$ (Ecuación 2.6) 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: $y = m x + b$ (Ecuación 2.7) 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 2.3.

Num. Observación	Nivel Estudios	De	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 2.3 Tabla de observación del ejemplo para demostrar la hipótesis del ejemplo

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 2.4 y la figura 2.5 que muestra una línea recta por entre los puntos.

Model Summary and Parameter Estimates

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

Tabla 2.4 Resultados para los datos del ejemplo

De la tabla 2.4 bajo el encabezado de *Parameter Estimates* encontramos los valores de $m = .613$ y $b = 3.164$ substituyendo en la ecuación 2.7 podemos obtener la ecuación de éste modelo que queda de la siguiente forma: $y = 0.613 x + 3.164$ (Ecuación 2.8) dónde x representa el nivel de ingresos y el nivel de estudios.

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 2.5.

Num. Observación	Nivel Estudios	De	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 2.5 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:

$y = \beta + \beta_1 x + \beta_2 z$ (Ecuación 2.9) 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 2.5 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 2.6 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 2.6 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: $y = 1.36 + 0.354 x + 1.386 z$ (Ecuación 2.10)

Algo extremadamente importante es que para que esto sea válido las variables de la tabla 2.6 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.- 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.

Singularidad significa que las variables no presentan multicolinealidad, o sea que no están correlacionadas entre si y que no son redundantes.

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

2.5. 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 y por lógica debería presentarse y sin embargo 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.

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 2.7 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 no existe normalidad en los datos como un todo (un valor de la significancia cercano a 1.0 representa un comportamiento normal de los datos). 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 este 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 ilustra 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” 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 éstas 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.

Si el ejemplo anterior no contara con la variable tipo de edificio, sino con las variables: Num. de habitantes, No. De habitaciones en lugar del tipo de la edificación. Se puede observar que utilizando el número de habitaciones y el número de ocupantes se puede deducir el tipo de edificio. Por ejemplo se puede asumir que: 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 mismos resultados obtenidos con la tabla 2.7

2.6. 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 por ejemplo población 1: situación habitual y 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 2.11

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 2.12

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 2.13

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 2.14

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 2.8 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 2.8 Salarios netos de los empleados de cuatro diferentes empresas, la empresa se muestra como una columna de la tabla

La pregunta obligada sería ¿Cual empresa ofrece mejores salarios a sus empleados? Para hacer el análisis conviene poner a la empresa como un dato y no como una columna. En esta sección nos limitaremos a comparar dos empresas, iniciamos por comprar la empresa 1 con la 2, para ello seguimos el siguiente procedimiento en el SPSS: Iniciamos por 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 2.9

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 2.9 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 2.10

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 2.10 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. Lo que es lo mismo las dos empresas tienen salarios similares. 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

5.- 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 2.11

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 2.11 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.

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.

2.7. 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. (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 estadísticas:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \dots = \mu_k \quad H_a : \mu_i \neq \mu_j \text{ para alguna } i \neq j$$

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. 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 como se muestra a continuación: Para calcular el estadístico F se cuenta con la ecuación 2.15

$$F_e = \frac{\frac{SCE}{k-1}}{\frac{SCD}{n-k}} \quad \text{Ecuación 2.15}$$

Dónde SCE es la suma cuadrática Externa (entre un grupo y otro). SCD Suma cuadrática dentro de los grupos (Interna) y quedan definidas como:

$$SCE = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \dots + n_k(\bar{x}_k - \bar{x})^2 \quad \text{Ecuación 2.16}$$

$$\bar{x} = \frac{\sum_{i=1}^k \bar{x}_i n_i}{n}$$

Dónde \bar{x} x promedio ponderada Ecuación 2.17

$$SCD = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + \dots + (n_k - 1)S_k^2 \quad \text{Ecuación 2.18}$$

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 2.12, 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 2.12 Datos para comparar métodos de enseñanza en universidades

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 2.13

Método	Media	Desviación Estándar	Significancia
1	67.9310	17.43853	.733
2	68.0520	18.79739	.798
3	61.9470	19.19313	.476

Tabla 2.13 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 2.14, 2.15 y 2.16 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 2.14 Descriptivas estadísticas por grupo.

Test of Homogeneity of Variances

aprovechamiento			
Levene Statistic	df1	df2	Sig.
.033	2	27	.967

Tabla 2.15 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 2.16 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.

3. Conclusiones.

El modelo facilita la selección de las herramientas estadísticas de acuerdo a las características del problema.

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, no hay que olvidar que afortunadamente 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, es la estrategia, la combinación de herramientas y la secuencia de las pruebas lo que representa la verdadera problemática estadística.

La sobre abundancia de herramientas actúa como un catalizador de errores, éste estudio trata de contrarrestarla proporcionando un compendio a manera de resumen de ejemplos para facilitar la selección de herramientas, esta investigación es la vista panorámica de la cancha de fútbol, antes de ver el balón en nuestros pies.

REFERENCIAS

Abraira, A., Perez De vargas, A.(1996), *Métodos Multivariantes en Bioestadística*. Madrid: Ed. Centro de Estudios Ramón Areces.

Altman, D. G., Bland, J.M. (1995), Statistics notes: The normal distribution. *BMJ* 1995 (310), 298-298

Altman, D.G., Bland, J.M.(1996), Statistics notes: Transforming data. *BMJ* 1996 (312), 770-770

Altman, D. A.(1997), *Practical statistics for medical research*. 1th ed., repr. 1997; London: Chapman &Hall

- Blalock, H.(1997) *Estadística Social*. México: FCE
- Briones, G. (2001) *Métodos y Técnicas para la investigación en las ciencias sociales*. México: Trillas
- Campbell, D.T. y Stanley, J.C.(1966), *Experimental and Quasi Experimental Social Research Design*; N. York: Rand McNally Co. Academia Press
- 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.
- Combe, S. A. (1979) *La Construcción de las teorías Científicas*. Buenos Aires: Nueva Visión
- Diez Medrano, J.(1997), *Métodos de análisis causal*. Madrid:Cuadernos Metodológicos 3 / CIS.
- Ferrand Aranaz, M.(2005) *SPSS para Windows. Programación y análisis estadístico*. Madrid: McGraw-Hill.
- Levi, R.I. y Rubin, D.S.(1996) *Estadística para Administradores*. New York: Prentice Hall
- Lohr, S. L. (1999) *Sampling: Design and Analysis*. Pacific Grove: Duxbury Press.
- Luque M. (2000): *Técnicas de análisis de datos en investigación de mercados*. Madrid: Pirámide.
- Maletta, H.(1995) *Análisis estadístico de hipótesis y modelos*. Buenos Aires: Departamento de Computación para ciencias sociales. USAL.
- 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.
- Mora y Araujo, M (1984): *El análisis de los datos en la investigación social*. Ed. Nueva Visión, 1984, Bs. As.
- Moral, I. (2006). Comparación de Medias,*Metodos Estadísticos para Enfermería Nefrológica*, Consultado el 31 de Jul. 2009 de <http://www.seden.org/>
- Muñiz, J. (2003) *Teoría Clásica de los Tests*. Madrid,España: Ediciones Pirámide
- Oviedo, H.C., Campo-Arias, A. (2005) Aproximación al uso del coeficiente de alfa de Cronbach. *Revista Colombiana de Psiquiatría* 35 (4) 572-580
- Pértiga, S., Fernández, P.(2001). La distribución Normal. *Cad Aten Primaria* 2001 (8), 268-274
- Rossi, P. H., Wright, J.D. y Anderson, A. B.(1983) *Handbook of Survey Research*. London: Academic Press.
- Sierra Bravo, R. (1986), *Tesis Doctorales y Trabajos de Investigación Científica*; Madrid, España: Thompson Editores
- Thomas, H.P. (2004) *Pruebas Psicológicas*. México: El Manual Moderno