

## ANÁLISIS E INTERPRETACIÓN DE DATOS

### Abstract

In social studies are frequently required to analyze the behavior of a data set to determine facts or which are characteristic of interest. Much of the work is basically analyze information before us, in the first instance, in a descriptive way.

### Resumen

En los estudios sociales se necesita frecuentemente analizar el comportamiento de un conjunto de datos con el objeto de determinar hechos o característica que son de interés. Gran parte del trabajo consiste básicamente en analizar una información que se nos presenta, en primera instancia, en forma descriptiva.

El término análisis aquí lo vamos a entender, en una doble perspectiva, una como un proceso de búsqueda de elementos aspectos o configuraciones que caractericen y contextualicen los procesos o hechos estudiados. Y otra, como *una simple lectura* de unos resultados obtenidos. Lo anterior está relacionado con el hecho que el análisis de los datos depende de una serie de elementos, entre los cuales, los más importantes a considerar son:

- a. Los objetivos del estudio
- b. La hipótesis de investigación
- c. El nivel de medición de las variables

Otra consideración a tener en cuenta, cuando se trabaja con un cúmulo de información asociada a una población, son los interrogantes que surgen con relación al comportamiento de los datos, especialmente en lo que respecta a la distribución y asociación de éstos. Briones<sup>1[1]</sup> señala al respecto:

*De manera normal, la primera tarea del análisis cuantitativo consiste en obtener una información básica sobre las variables contenidas en el estudio, como es el caso cuando la información se ha recogido con un cuestionario estructurado con un número más bien elevado de preguntas.*

*Según la naturaleza de las variables, en cuanto a su nivel de medición, esta primera información proporcionará distribuciones de frecuencia, medidas de tendencia central (medias aritméticas, medianas), medidas de dispersión y otras. Todas ellas -o las que el investigador considere más importante- proporcionarán una descripción global del colectivo estudiado como también permitirá orientar el análisis subsiguiente sobre la base de las situaciones más relevantes que hayan podido generar.<sup>2[2]</sup>*

En las lecciones anteriores se ha descrito en qué consiste la medición y la recolección de datos; los diversos tipos de instrumentos que se utilizan con este objetivo, y como se codifica y sistematiza. En esta lección y en las siguientes se presentan y ejemplifican los métodos estadísticos más utilizados para su análisis, tales como las técnicas paramétricas y las no-paramétricas, haciendo énfasis en la lógica y oportunidad de su uso. Implícitamente lo que se pretende responder es la pregunta metodológica básica:

*¿Cuál estadístico se debe usar y qué significa su resultado?*

Hoy en día, el cálculo de los estadísticos es realizado por programas informáticos especializados de manejo y gestión de datos. Y, por ello, en este texto se da mayor

---

<sup>1[1]</sup> Briones, Guillermo, Métodos y técnicas avanzadas de investigación aplicadas a la educación y a las ciencias sociales, Módulo 4, Análisis e interpretación, PIIE, Santiago de Chile, ICFES, Bogotá, 1988, p. 15.

<sup>2[2]</sup> *Ibid.*, p. 23.

importancia a la lógica y oportunidad del uso de estos estadísticos en relación con los objetivos y con el tipo de información recolectado. Lo importante es saber qué técnicas se deben utilizar cuando se desea describir un colectivo o una situación o, encontrar una relación entre unas variables o, hacer algunas inferencias estadísticas. Como hemos adelantado, nuestras ejemplificaciones estarán referidas al programa SPSS.

### **Análisis descriptivo de los datos**

La primera tarea del analista es, a partir de una información confiable y válida, hacer una descripción de las características de las unidades de análisis, una segunda tarea, si es el caso, hace referencia a la comparación de sub-grupos en relación con una o más características que son importantes en los objetivos formulados por el investigador. Esta secuencia de tareas en el campo del análisis de los datos tiene como objetivo encontrar:

- 1) Aspectos que caractericen, en términos cuantitativos, a los colectivos o procesos, en los cuales se han definido, problemas de investigación –descripción de variables-.
- 2) Clasificaciones, tipologías, diferencias o comparaciones entre ellos.
- 3) Relaciones entre variables sociológicas -asociaciones y correlaciones.

### **Descripción de variables**

#### **La distribución de frecuencias.**

Un concepto básico del análisis descriptivo es la distribución de frecuencias que consiste en "el ordenamiento o clasificación de los valores observados en una variable de acuerdo con su magnitud numérica. Permite identificar al investigador la forma como ciertos puntos o características están distribuidos".

#### **Elementos que conforman una distribución de frecuencias.**

La distribución de frecuencias consiste en una tabla que organiza los datos en clases, es decir, en grupos de valores que describen una característica de los datos y tiene la siguiente estructura:

- Categorías de la variable
- Frecuencias: número de casos en cada categoría
- Frecuencias relativas: porcentajes de casos en cada categoría
- Frecuencias acumuladas: porcentaje de casos que se van acumulando en cada categoría desde la más baja hasta la más alta.

Al analizar la distribución de los valores de una variable, el investigador está interesado en:

- Hacer una lectura global de los datos y analizar su distribución: su frecuencia absoluta o relativa; sus medidas de tendencia central, etc.
- ¿Cómo varían las categorías dentro de la variable?
- Buscar una explicación de la variabilidad de la característica, propiedad o variable.  
¿Qué factores están asociados a esta variabilidad?

## Procedimiento SPSS

El submenú <**Frecuencias**> es uno de los más utilizados del programa SPSS dado que no sólo sirve para el cálculo de estadísticas, sino también para **visualizar** y **depurar datos** (eliminación de códigos inválidos, modificación de datos mal digitados).

El submenú <**Frecuencias**> elabora una tabla de frecuencias en donde indica cada código que aparece como respuesta, el número de veces que aparece el porcentaje que presentan esas apariciones con respecto al total de casos.

Uno de los productos en *el visor de resultados* de <Frecuencias> es *una tabla distribución de frecuencias* que muestra *las frecuencias relativas* que son los porcentajes de casos en cada categoría. Las distribuciones de frecuencias relativas pueden presentarse en forma de histogramas o gráficas de otro tipo, tales como polígonos de frecuencias<sup>3[3]</sup>.

## Ejercicio de autocomprobación

La inseguridad es un miedo urbano que permeó todas las esferas. La sospecha nació como mecanismo de defensa y ya, muchos ni siquiera confían en el vecino. Los más intrépidos dicen no sentir ese temor. Para los expertos es claro que las políticas represivas solo brindan sensación de seguridad momentánea, que lo que hay que atacar es el fondo del problema.

Un sondeo realizado recientemente, mostró que un 46.4 % de los habitantes de la capital antioqueña se sienten inseguros, mientras un 53.6 % afirmó sentirse seguros. Entre los delitos que más desvelan a los medellinenses están los robos, el terrorismo, el secuestro, las

---

<sup>3[3]</sup> Los polígonos de frecuencias relacionan las puntuaciones con sus respectivas frecuencias, Se utilizan con variables de tipo intervalos.

violaciones y los homicidios. El sondeo evidenció que un 35.7% de los interrogados, se sienten inseguros por el fenómeno de la indigencia. Las personas encuestadas ven como una solución el aumento de pie de fuerza, pero también la del fomento de programas de cultura ciudadana.

### Ficha técnica

El sondeo de opinión entrevistó, vía telefónica a 322 personas, mayores de 16 años, de diferentes estratos económicos y sexo. Un 46.4% de los encuestados fueron hombres y un 53.6% mujeres. De estos un 32.1% pertenece a los estratos 1 y 2; el 35.7% al 3; el 21.4% al 4 y el 1.7% a los estratos 5 y 6. El sondeo tiene un margen de error de 4%.

### Procedimiento SPSS

- En la barra principal de Menús del SPSS, seleccione el menú <Archivos>
- Active la opción <Abrir> y elija el submenú <Datos >
- Seleccione el archivo **temor ciudadano.sav** en el CD-ROM
- En la barra principal de Menús, seleccione el menú: < Analizar> -->>>  
<Estadísticos descriptivos> -->>> <Frecuencias<sup>4[4]</sup>>
- En el cuadro de diálogo Frecuencias, se selecciona la variable a describir y se transfiere a la sección Variables
- Seleccione simultáneamente las variables “robos” y “secuestr”

En el Visor de resultados aparece:

### Tabla de frecuencia

---

<sup>4[4]</sup> Con la opción <Frecuencias> aparece como resultado en el Visor una **Tabla de Frecuencias**<sup>4[4]</sup> en donde se lista cada categoría o código que aparece como respuesta, el número de veces que aparece y el porcentaje que representa esas apariciones con respecto al total de casos.

### **Análisis de los datos**

De acuerdo a la información del sondeo de opinión sobre “**la Inseguridad, un miedo urbano**”, se puede apreciar que los robos – en cualquiera de sus manifestaciones- ocupan el primer lugar (61%) en las percepciones de los medellinenses; el secuestro le sigue en orden de importancia entre estos habitantes (50.4%).

### **Análisis estadístico de los datos**

Una segunda aproximación al análisis de los datos, se centra en el uso de estadísticos descriptivos. Cuando se describe un conjunto de variables se trata de sintetizar sus características en unos índices numéricos que representen adecuadamente a ese conjunto de características. En este texto, básicamente las técnicas que se expondrán serán las medidas de tendencia central, valores percentiles<sup>5[5]</sup>, dispersión y distribución.

---

<sup>5[5]</sup> Valores percentiles. Los valores de una variable cuantitativa que dividen los datos ordenados en grupos, de forma que un porcentaje de los casos se encuentre por encima y otro porcentaje se encuentre por debajo.

Para lograrlo el analista que ha sometido la información a un proceso de cuantificación debe hacerse tres preguntas básicas:

1. *¿Cómo se presentan los datos? ¿Cuál es la ubicación o localización de los datos?*

Lo que se pretende es describir ciertas características de un conjunto de datos. El procedimiento más utilizado, para dar respuesta a este interrogante, son las *medidas de tendencia central*, que permiten resumir el comportamiento de un conjunto de datos alrededor de un punto prefijado tal como:

La Moda; la Media; la Mediana

2. *¿Los datos respecto a sus puntos de agrupación y localización están concentrados o dispersos?*

Para determinar la concentración de los datos o su dispersión alrededor de la media, en una distribución, se usa frecuentemente los conceptos de:

- Rango de variación
- Valores percentiles
- Desviación estándar o típica

3. *¿Existe relación en el comportamiento de las variables?*

Para establecer la relación, es decir, cuánto de una variable es explicada por las demás, cuyas observaciones han sido registradas en una escala de intervalo o de razón –continua-, se utiliza:

- Coeficientes de asociación entre variables cualitativas
- Coeficientes de correlación entre variables cuantitativas
- Los modelos de regresión



### **Nota metodológica**

Un criterio básico que hay que tener en cuenta al seleccionar la prueba estadística es el nivel de medición en que fueron medidos los datos<sup>6[6]</sup>.

La dirección del análisis o la búsqueda de contextualizantes se hace dentro de ciertas *orientaciones* que son señaladas generalmente por los objetivos del estudio.

En el análisis de los datos, estamos en el campo de las técnicas, en cambio en la interpretación estamos en el campo de la conceptualización teórica.

---

<sup>6[6]</sup> El tipo de escala de medición empleado para obtener los valores de una variable es determinante en la selección de los estadísticos de análisis de los datos.

### **Estadísticas para medir la localización de un conjunto de datos**

Una tarea básica del análisis de los datos consiste en obtener una descripción global de las características de los sujetos o procesos – objetos del estudio-. Esta primera información mostrará cómo se distribuyen los valores de la variable. Frecuentemente se necesita resumir descriptivamente un conjunto de valores numéricos para lo cual es apropiado recurrir a medidas de tendencia central. Todas estas estadísticas darán una descripción global de la población estudiada y permitirán orientar el análisis. Existen 3 medidas que pueden ser asociadas a este concepto y ellas son: La Moda, la Mediana y la Media.

**Moda.** Se utiliza para medir la localización de los datos en una *escala nominal*. Es aquel valor que se repite con más frecuencia en un conjunto de datos.

### **Ejercicio de autocomprobación**

Se está interesado en estudiar los conocimientos y actitudes frente a la sexualidad en un grupo de niñas de un Colegio. Obviamente, es importante para efecto del análisis, la edad de los sujetos en estudio.

#### **Tabla Distribución según grupos de edades.**

##### **Grupo A. alumnas mujeres**

Edad (X)	Frecuencia (f)	f(x)
9	1	9
8	2	16
7	3	21
6	2	12
Total	8	58

La pregunta básica que se hace en el intento de describir globalmente este grupo o colectivo de estudio es ¿cuál es la edad más representativa en el grupo? A simple vista se puede afirmar que la edad representativa es 7 años, ya que la mayoría declaró esa edad; luego *la Moda* es igual a 7 ya que es el valor de la distribución que aparece con mayor frecuencia.

**Mediana.** Es una medida de tendencia central que divide a cualquier distribución o colectivo en 2 partes iguales o por la mitad y sirve para medir el punto que está al centro de la distribución. Veamos con una ejemplificación en que consiste esta medida.

### Ejercicio de autocomprobación

Calcular la mediana en la información que representa el promedio de días de estadía de los pacientes en el hospital universitario:

**Tabla. Número de días de hospitalización según paciente.**

Paciente	Nº días	Paciente	Nº días	Paciente	Nº
1	1	10	4	19	10
2	2	11	4	20	10
3	2	12	4	21	15
4	3	13	4	22	25
5	3	14	5	23	30
6	3	15	5	24	100
7	3	16	5	25	150
8	3	17	7		
9	3	18	8		

### Análisis de los datos

La mediana es igual a 4: valor que ocupa el lugar 13 en el listado. Este valor se puede interpretar así: la mitad de los pacientes tienen una permanencia de menos de 4 días de hospitalización y la otra mitad más de cuatro días.

El número promedio de días de hospitalización de un paciente típico (Media) es de 16.36 días. La Mediana es igual a 4, observe que son dos valores estadísticos muy diferentes.

La Media es mucho más sensible a los valores extremos que la Mediana. En este caso, la media está fuertemente influida por dos pacientes que han estado hospitalizados 100 o más días. La mediana, al contrario, está menos determinada por estos.

### Nota metodológica

Usar la Mediana conjuntamente con la Media es un buen recurso metodológico para analizar el comportamiento de un conjunto de datos, por ejemplo, las estadísticas económicas generalmente presentan la mediana de ingresos de un grupo de personas antes que "la media de los ingresos", ya que el nivel de ingresos puede aparecer artificialmente alto o bajo debido a la presencia de un pequeño segmento de estas personas que pueden ser ubicadas o clasificadas en los extremos, por sus bajísimos o altísimos ingresos.

Para calcular la mediana se tienen que ordenar los datos y determinar cuál que está a la mitad. Sirve para datos propios de los niveles de medición ordinal, por intervalos y de razón o proporcionales. No es lógico usar ésta medida con variables nominales, porque en este nivel no hay jerarquías, no existe la noción de un escalafón superior o inferior.

**La Media.** Llamada también media aritmética es la medida de localización o tendencia más utilizada. Con ella se calcula el promedio de los datos. En el lenguaje cotidiano, se le conoce como promedio aritmético.

Ejemplificación: De acuerdo a las edades declaradas en la Tabla 6.1 podemos decir que el promedio es de 7.25 años.

La media aritmética se calcula sumando los puntajes de una distribución de valores y dividiéndolos por el número de puntajes u observaciones. Este concepto se resume en la siguiente fórmula:

$$\text{Cálculo de la Media (X)} = \frac{\sum f X}{n} = \frac{n_1 + n_2 + \dots + n_8}{\sum} = \frac{58}{8} = 7.25$$

### **Estadísticas para analizar la dispersión o variabilidad de un conjunto de datos.**

En el análisis de unas características, se necesita algo más que la distribución de los valores de unas respuestas preestablecidas (frecuencias) o el cálculo de unos promedios; también es necesario determinar el grado de variabilidad - diseminación –dispersión- de los datos con respecto a ese promedio. Para tales efectos las medidas de dispersión, tales como el rango, la varianza y la desviación estándar son de extrema utilidad. Estas medidas de variabilidad indican la dispersión y distancias de los valores en la escala de medición.

**Rango.** Se define como la "diferencia o distancia" entre los valores del dato mayor y el dato menor. Este estadístico mide la dispersión de los valores en una serie. Cuando más amplio grande sea el rango, mayor será la dispersión de los datos en una distribución.

### **Ejercicio de autocomprobación**

Supongamos que estamos analizando el número de consultas diarias atendidas por el personal médico de los consultorios del Servicio Médico de la Universidad.

**Tabla N° de consultas diarias en los 2 consultorios.**

<b>Consultorio A</b>		<b>Consultorio B</b>	
<b>Médico</b>	<b>N° Consultas</b>	<b>Médico</b>	<b>N° Consultas</b>
A	10	F	28
B	20	G	29
C	30	H	30
D	40	I	31
E	50	J	32
	<b>X= 30</b>		<b>X = 30</b>

### **Análisis de los datos**

En ambos consultorios el número promedio de consultas es de 30 pacientes por día. Pero como se logra apreciar, el número de pacientes atendidos por cada médico es diferente.

La distribución de las consultas no es idéntica en los dos consultorios: en el Consultorio A, existe una distribución desigual de las consultas, los médicos D y E atienden de dos a cinco veces más pacientes de los que atienden sus colegas A y B.

El rango de consultas en el Consultorio A es de 40 (10 el mínimo y 50 el máximo). En el Consultorio B el "rango" de consultas es de sólo de 4 (32 consultas menos 28 consultas).

**Desviación estándar o típica.** Estadísticamente, cuando se mide la dispersión lo que se está midiendo es el alejamiento de los valores en relación con la Media. La desviación estándar es la medida de dispersión más usada y se interpreta con relación a la media. Tiene la desventaja que los valores extremos en el conjunto de datos distorsionan su valor.

### **Manejo de estadísticos con el programa SPSS**

En las unidades anteriores se han presentado algunas técnicas de manejo estadístico ejemplificaciones simplificadas y muestras de tamaño menor. Pero en la vida real, como analistas, nos enfrentamos a problemas mucho más complejos que requieren de grandes poblaciones muestrales y del análisis de múltiples propiedades o variables. Los programas de computación especializados facilitan el manejo y análisis de la información.

Todas las técnicas de análisis estadístico que se mencionan en ésta unidad didáctica se encuentran en el programa SPSS. Por tal razón aquí sólo se hacen referencias a lo que Briones señala como *“la lógica y la oportunidad de su uso en relación con los objetivos de*

*investigación y con la naturaleza de la información recogida, y no, en el desarrollo de fórmulas estadísticas<sup>7[7]</sup>».*

A medida que se van desarrollando las técnicas estadísticas, se muestra el procedimiento correspondiente, en una sección dentro de la Unidad, que se denomina ***Procedimiento SPSS***.

**Procedimientos generales para la selección de las técnicas estadísticas en el análisis de datos.**

Las técnicas de análisis estadístico están agrupadas en el SPSS, en la *Barra Del Menús Principal*, bajo el menú **<Analizar>**.

Al dar clic sobre este menú aparecerá el listado de técnicas disponibles:

---

<sup>7[7]</sup> Guillermo Briones, Loc. Cit., p. 21



## ANALIZAR

Las opciones del menú Analizar están disponibles sólo si hay un archivo de datos activo o abierto. Estas opciones del menú Analizar ejecutan los procesamientos estadísticos y están organizadas dentro de submenú de acuerdo al tipo de análisis a ejecutar.

El Menú Analizar presenta las siguientes opciones de operaciones de la estadística descriptiva, relacional e inferencia<sup>8[8]</sup>.

En esta primera fase, nos detendremos en la opción “Estadísticos descriptivos” dando un clic sobre la opción. A continuación, se desplegará una ventanilla que incluye:

---

<sup>8[8]</sup> Las opciones o comandos del menú Analizar están disponibles sólo si hay un archivo de datos activo - abierto.

La opción Frecuencias es la más utilizada del programa SPSS, dado que no sólo sirve para el cálculo de estadísticas, sino también para depurar datos. Básicamente elabora una tabla de frecuencias en donde indica cada categoría que aparece como respuesta, el número de veces que aparece y el porcentaje que representan esas apariciones son respecto al total de casos. Además, con esta opción de estadísticos de frecuencias, se examinan las variables mediante cuatro aspectos descriptivos.

Para obtener estadísticos descriptivos y de resumen opcionales para las variables numéricas utilice el siguiente procedimiento SPSS:

- En la barra principal de Menús del SPSS, seleccione el menú *<Archivos>*
- Active la opción *<Abrir>* elija el submenú *<Datos >*
- Seleccione el archivo profesional-2003.sav en el CD-ROM
- En la barra de Menús, elegir el menú *<Analizar<sup>9[9]</sup>>* *<Estadísticos descriptivos>*
- Seleccione *<Tablas de contingencia>* → → *Frecuencias...*

Lo anterior abrirá el cuadro de diálogo: Frecuencias, como se muestra en la siguiente figura:

---

<sup>9[9]</sup> Cada elemento de la barra de Menú, tiene un menú desplegable. Para acceder al menú respectivo use el puntero del ratón u oprima simultáneamente la tecla ALT más la letra subyacente de la opción.



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

[ceo@catios.udea.edu.co](mailto:ceo@catios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

Si usted quiere calcular algunos estadísticos, *de un clic* en el botón Estadísticos y aparecerá a continuación, el cuadro de diálogos Frecuencias: Estadísticos: proceda a seleccionar los estadísticos, por ejemplo, Cuartiles, Desviación típica, Mínimo, Máximo, Media

Al dar un clic en el botón **continuar**, se retorna al cuadro de diálogo Frecuencias. Allí nuevamente se da un clic en el botón **Aceptar**.

A continuación aparecerá el siguiente visor de resultados:



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

[ceo@carios.udea.edu.co](mailto:ceo@carios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

### **Análisis de los datos**

A partir de la información obtenida de 1100 encuestados, se puede concluir que el promedio de ingresos anuales declarados es de € 26.064; igualmente se observa que existe una gran dispersión en la distribución de los ingresos - € 7.200 como mínimo y € 65.500 como máximo. Se puede afirmar, a partir de los datos que el 25 por ciento de los encuestados tiene ingresos iguales o inferiores a € 21.000 y que el 50 por ciento tiene ingresos iguales o superiores a € 26.000.

### **Nota metodológica**

El cuadro de diálogo **Frecuencias: Estadísticos** presenta cuatro opciones de manejo de estadísticos de localización o distribución:

**1. Valores de percentiles.** Puede elegir una o más de las siguientes opciones:

**Cuartiles.** Muestra los percentiles 25, 70 y 75.

**Puntos de corte para n grupos iguales** muestra los valores del percentil que divide la muestra en grupos de casos de igual tamaño. El número de grupos predeterminado es de 10. De modo opcional, puede solicitar un número de grupos diferente. Introduzca un entero positivo entre dos y cien por ejemplo, si introduce 4, se mostrarán los cuartiles.

**Percentil:** Valores en los percentiles especificados por el usuario. Introduzca un valor de percentil entre cero (0) y cien (100) y dé un clic en el botón Añadir.

**2. Dispersión:** Puede elegir una o más de las siguientes opciones:

**Desviación estándar o típica:** Estimación de la cuantía en que se diferencian las observaciones respecto a la media, expresada en las mismas unidades que los datos.

**Varianza:** Estimación de la cuantía en que se diferencian las observaciones respecto a la media, es igual al cuadrado de la desviación típica.

**Amplitud:** Diferencia entre los valores mayor (máximo) y los menor (mínimo).

**Mínimo:** El valor más pequeño.

**Máximo:** El valor más grande.

**E.T. Media:** Error típico de la media. Una estimación de la variabilidad muestral de la media.

**3. Tendencia central:** Puede elegir una o más de las siguientes opciones:

**Media:** La media aritmética o promedio.

**Mediana:** La mediana está definida por el valor por debajo del cual se encuentran la mitad los casos. Si hay un número de casos par, la mediana es la media de los dos casos centrales ascendentes.

**Moda:** el valor que ocurre más frecuentemente. Si diversos valores están empatados en la frecuencia más alta sólo aparecerá el valor más pequeño de la variable.

**Suma:** la suma de todos los valores.

**4. Distribución:** puede elegir una o más de las siguientes opciones:

**Asimetría:** Índice del grado en que la distribución no es simétrica. También se muestra el error típico del estadístico de asimetría.

**Curtosis:** estimación del grado en que las observaciones se agrupan alrededor del punto central, también se muestra el error típico del estadístico de Curtosis.

Para datos agrupados, también se encuentra disponible la siguiente opción:



### **Técnicas de representaciones graficas de las distribuciones de frecuencias.**

Los gráficos o figuras constituyen uno de los medios más usados y más difundidos para la presentación y análisis de la información estadística. Esto se debe al hecho de que las ideas presentadas gráficamente son entendidas con mayor rapidez que las explicaciones textuales o numéricas.

### **Tipos de figuras o gráficos**

La mayoría de las personas han visto, o analizado y hasta construido figuras o gráficos y posiblemente han notado que hay una gran variedad de ellos. Una presentación detallada de todos los tipos de figuras o gráficos excede los propósitos de este texto, por ello sólo se hará referencia a los de mayor uso.

- Histogramas de frecuencia: son figuras que presentan la información de cada una de las categorías de la variable la forma de rectángulos proporcionales. Se utilizan generalmente para representar datos de una variable discreta.
- Polígonos de frecuencias: son gráficos en la forma de una serie de líneas rectas conectadas entre sí y que unen unos puntos medios ó de intervalo a lo largo del eje horizontal. Se utilizan para representar distribuciones de frecuencias propias de un nivel de medición por intervalos.
- Gráficos de barras: Constituyen uno de los tipos más simples y quizás de mayor uso. Resultan de especial utilidad en la presentación de datos cualitativos o cuantitativos. Se emplean tipos muy variados de gráficos de barras, sin embargo todos reúnen ciertas características o requisitos esenciales.
- Gráficos o figura de círculos o de pastel: en este tipo de figuras o gráficos de círculos o de pastel el área de cada una de los sectores refleja la importancia de la categoría que representa. Se trata de un gráfico en dos dimensiones. El gráfico de pastel consiste en un círculo el cual se divide en tantos sectores como categorías se



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

tienen de manera que el área que corresponde a cada categoría sea proporcional a su importancia relativa o porcentual.

### **Limitaciones en el uso de los gráficos o figuras:**

La utilización de figuras o gráficos tiene ciertas limitaciones que es conveniente mencionar:

- No pueden incluir tanta información como las tablas. Sólo se puede presentar a la vez una cantidad limitada de datos ya que demasiada información y/o variedad atenta contra su utilidad, al dificultar su lectura.
- En las figuras o gráficos se representan valores aproximados mientras que las tablas permiten incluir la información con toda exactitud.
- La presentación tabular y gráfica es complementaria al texto pero no lo reemplaza. Por ejemplo, para ilustrar el texto, igualmente para llamar la atención y despertar el interés por los datos que son presentados así como para reforzar las argumentaciones y conclusiones a las cuales se ha llegado.

### **Procedimiento SPSS**

En la barra de Menús del SPSS, seleccione el menú Gráficos<sup>10[10]</sup> » » » Barras »»»  
Gráficos de barras »»» seleccione el tipo de grafico<sup>11[11]</sup>:

También se pueden obtener gráficas con algunas opciones del Menú Analizar → →  
Estadísticos descriptivos → → Frecuencias → → Botón Gráficos

---

<sup>10[10]</sup> Este menú ofrece opciones de figuras tales como: barras, líneas, áreas, sectores, histograma con sus respectivos cuadros de dialogo: Grafico de....

<sup>11[11]</sup> Por ejemplo, el SPSS puede diseñar 9 tipos de gráficos de barra

## **La interrelación entre variables**

En ocasiones, un investigador requiere que su análisis vaya más allá de la mera descripción de una variable; la misma complejidad de la realidad sociedad exige estudiar, por ejemplo, cómo un proceso esta asociado a otros o cómo varía en función de otros.

Cada día, los administradores toman decisiones que están basadas sobre predicciones de eventos futuros. Para hacer estos pronósticos ellos se basan en la relación -intuitiva y calculada- entre lo que se conoce y lo que se va a estimar. Si los administradores pueden determinar cómo está relacionado lo conocido con el evento futuro, pueden ayudar considerablemente al proceso de toma de decisiones. Este es el tema de esta lección: cómo determinar la relación entre variables.

En la unidad anterior, se expusieron diversos procedimientos que permiten la descripción de colectivos. En la presente Unidad y en la siguiente, usted, encontrará diversas técnicas que permiten analizar determinaciones y relaciones entre variables.

### **Ejercicio de autocomprobación**

Supongamos que se estudian las diferencias de opinión que tienen personas de distinto sexo acerca de cuáles serían las soluciones al problema de la inseguridad ciudadana.

Examinemos la relación entre las variables *género* y *posibles soluciones a la violencia*. A partir de ésta se puede formular una hipótesis: con mayor frecuencia que los hombres, las mujeres proponen que un aumento en la dotación policial, es una solución a la violencia que vive la Ciudad.

De hecho, el enunciado anterior señala una relación. El problema metodológico surgirá entonces en el momento en que queramos confirmar nuestra afirmación: *¿Cómo comprobar la relación entre las dos variables -género y soluciones posibles a la inseguridad y violencia-?. Es decir, cómo una explica, determina u afecta a la otra. O, cómo una varía en función de la otra. Una posible solución metodológica es la de construir una tabla de contingencia o de tabulación cruzada.*

### **Tablas de contingencia y medidas de asociación.**

Sigo pensando que este aparte se repite con los párrafos en rojo que están inmediatamente más abajo. Aquí arriba están los ejemplos y más abajo las definiciones.

Una tabla de contingencia es la representación conjunta de la distribución de frecuencias de dos o más variables. A su vez cada una de ellas se subdivide en dos más categorías. Cuando se forman con dos variables reciben el nombre de tablas de 2 entradas. Y cada tabla producto del cruce de más de dos variables se les llama subtablas.

Una de las técnicas más frecuentemente utilizadas en la investigación social es la clasificación de datos en tablas según uno o más criterios (variables). Veamos con un ejemplo de una tabulación cruzada como se clasifican los datos de dos variables: *género y preferencia por un tipo de solución al problema de la inseguridad ciudadana.*

**Tabla: Alternativas desde la perspectiva de género para disminuir la delincuencia en la ciudad de Medellín.**

	<b>Hombres</b>	<b>Mujeres</b>	<b>Total</b>
Aumentar dotación policial	38%	17%	26%
Participación comunitaria	32%	29%	31%



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

Otras soluciones	19%	46%	34%
N.S./ N. R.	11%	8%	9%
Total	100%	100%	100%

### **Análisis de datos tabulados**

La lectura de los datos de una tabla depende de la hipótesis de trabajo, lo cual determina el tipo de cálculo porcentual que se requiere. Los porcentajes en la Tabla son calculados teniendo como base el total de la frecuencia en cada categoría de la variable independiente (género), así la suma de los porcentajes deberá sumar 100 por ciento en la base de la variable independiente. Sin embargo, uno se podrá preguntar, ¿porqué no hacer el cálculo en la dirección de las categorías de la variable dependiente (alternativas de soluciones a la delincuencia)?.

La respuesta es muy sencilla y a la vez muy importante ya que la mayoría de los programas estadísticos de computador generan tablas de contingencia con porcentajes calculados en ambas direcciones –horizontales - verticales y es el Analista quién debe decidir cuál es la dirección correcta en la lectura de la Tabla.

La tabulación cruzada es una técnica estadística también denominada *Tabla de Contingencia*, en la que se presentan los valores de las frecuencias conjuntas de dos o más variables. Cuando se forman con dos variables reciben el nombre de *tablas de dos entradas*; cuando son más de dos las variables involucradas, son tablas de *n entradas*, donde n es el número de variables involucradas. A cada tabla obtenida del cruce de dos o más variables se le conoce como *subtabla*, dado que en realidad es parte de la tabla principal formada por las dos primeras variables. En ella se calcula la distribución de frecuencias de una variable para cada una de las categorías o clases en las que se divide la otra variable con la cual se cruza. Los porcentajes en una tabla cruzada, que expresan frecuencias relativas, pueden calcularse en dirección vertical u horizontal. *Los porcentajes verticales* se calculan sobre el total de elementos de la muestra que pertenecen a la categoría indicada en cada columna. *Los porcentajes horizontales* se calculan sobre el total de cada fila. La frecuencia de cada fila o columna se denomina *frecuencia marginal*.

Se supone que la variable independiente es el factor causal o determinante que tiene el mayor peso en las posibles variaciones de la variable dependiente. La decisión acerca de cuál es *la variable independiente o explicativa* y cuál es la dependiente o explicada parece fácil, pero en ciertas ocasiones tiende a ser de tal complejidad que genera traumatismos, especialmente cuando se está trabajando con programas como el SPSS.

Por ejemplo, en la relación entre "género" y "alternativas de solución", suena sin sentido considerar como variable independiente a las "alternativas de solución" y parecería sin sentido pensar que esta variable pueda influir al "género". Sin lugar a dudas aquí la variable independiente es el *género*, del cual nosotros hemos hipotetizado que afecta o determina la variable dependiente ("alternativas de solución")

Si el sentido común, a veces limitado, no nos ofrece una pista acerca de *cuál es cuál*, un criterio metodológico adoptable es el de establecer la secuencia temporal entre las dos variables:

*¿Cuál de las dos variables precede en el tiempo a la otra?*

### **Procedimiento SPSS**

En el SPSS el submenú <Tablas de contingencia> permite la tabulación cruzada y calcula sus estadísticas. En esta operación, se puede utilizar cualquier tipo de variables: *cuantitativas discretas, cuantitativas continuas, cualitativas nominales, cualitativas ordinales*.

Todas las estadísticas que calcula <Tablas de contingencia> son *medidas de asociación* o, en su defecto, **de independencia**; pero la negación de una lleva la afirmación de la otra; si





UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

dos variables no están asociadas, son independientes, o si dos variables no son independientes están asociadas.

### Ejercicio de autocomprobación

Sistematizar y analizar la información del “estudio sobre la apreciación de los colombianos sobre la crisis política que vive el país”.

- En la barra principal de Menús del SPSS, seleccione el menú <Archivos>
- Active la opción <Abrir> elija el submenú <Datos >
- Seleccionar archivo situación-2000.sav en el CD-ROM
- En la barra de Menús, elegir el menú <Analizar> <Estadísticos descriptivos>
- Seleccionar <Tablas de contingencia>
- Seleccione una variable independiente de la base de datos **Situación-2000.sav**, por ejemplo –*sexo*- y “crúcela” con la variables “sensación de inseguridad”
- Proceda a cruzar las variables con el siguiente procedimiento:
- Transfiera la variable independiente *sexo* a “Filas:”
- Transfiera la variable dependiente *sensación de inseguridad* a “Columnas”
- En el botón Cá\_sillas defina si el cálculo de los datos se va a hacer por filas, columnas o sobre el total de casos



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

A continuación, en el visor de resultados aparecerá la siguiente tabla:

### **Análisis de los datos**

[ceo@carios.udea.edu.co](mailto:ceo@carios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

Del total de personas entrevistadas de sexo femenino, una gran mayoría manifiesta sentirse insegura (91.7%); las personas de sexo masculino, en cambio, manifiestan en menor medida dicho temor: sólo un 30.8% de los entrevistados se siente inseguro.

### **Nota metodológica**

Con el procedimiento <Tablas de contingencia> se puede utilizar cualquier tipo de variables: Cuantitativas o cualitativas, sean ordinales o nominales. Por supuesto, las estadísticas que se calculan tienen validez sólo para las variables adecuadas, por ejemplo, el coeficiente de correlación entre sexo y edad no tiene ningún sentido, ya que el sexo es una variable cualitativa nominal y la edad es cuantitativa continúa.

### **Control del efecto de las variables intervinientes en una relación original.**

Uno de los propósitos de los analistas sociales que utilizan los diseños cuantitativos es el de tratar de aislar o circunscribir, en lo posible, aquellos elementos, hechos o factores sociales que median un proceso social. En términos de variables, lo que se trata es de intervenir o aislar aquellas variables intervinientes<sup>12[12]</sup> que pueden ser consideradas causas de los cambios de la variable dependiente.

Supongamos que queremos determinar si un programa de alcoholismo "cura" alcohólicos. Para esto debemos estar seguros que "la curación" es debida al Programa y no a otros factores intervinientes o asociados.

En los diseños cuantitativos, se dice que existe una conexión causal entre dos variables cuando:

- Las dos varían o covarían, esto es, que si una varía la otra también debe hacerlo;
- Que una precede en el tiempo a la otra, es decir, que existe una secuencia temporal entre ellas;
- La relación entre las dos no es el resultado de otra variable que está relacionada a ellas;
- Existe un sólido argumento teórico que permite asegurar que una de las variables es la determinante de la otra.

Gracias al uso de controles estadísticos es posible acercarnos con cierto grado de confiabilidad a cuál es realmente la variable más determinante o causal del proceso que se está analizando y, de paso, eliminar la posibilidad que la relación encontrada entre la variable independiente y dependiente sea el resultado de otra variable asociada a ellas.

---

<sup>12[12]</sup> "El concepto de variable interviniente o de control presupone que la relación entre dos variables esta mediada por otra (u otras) que transportan los posibles efectos de la primera." Escalante, Carlos. "El Problema y la Hipótesis", Módulo 2, ICFES, Bogotá, 1987, p. 83.

Si la relación se mantiene *inmodificada o sostenida* dentro de los subconjuntos, pese a la existencia de la variable interviniente, se puede llegar a afirmar que se ha eliminado una posible causa alternativa en *la relación original* observada. Pero sí la determinación entre las variables independiente y dependiente se mediatiza o desaparece, nosotros sabremos que la relación entre las variables originales no es tan determinante y lo que realmente explica la situación es la relación secundaria.

**Tabla A: Soluciones ante el aumento de la violencia según género**

	Hombres	Mujeres	Total
Aumentar dotación policial	38%	17%	26%
Participación comunitaria	32%	29%	31%
Otras soluciones	19%	46%	34%
N.S./ N. R.	11%	8%	9%
	100%	100%	100%

En la tabla *Soluciones ante el aumento de la violencia según género*, se aprecia que "un 38 por ciento de las personas clasificadas como hombres se inclina por la solución del aumento de la dotación policial, en cambio sólo un 17 por ciento de las mujeres mostró igual preferencia".

Si se analizan las opiniones de los encuestados, al introducir una nueva variable interviniente: *nivel de escolaridad* encontramos que las apreciaciones por el tipo de solución de aumentar de la dotación policial varían sustancialmente.

Veamos que nos muestra la siguiente tabla

**Tabla B: Soluciones ante el aumento de la violencia según género**

	Alto nivel de escolaridad		Bajo nivel de escolaridad	
	Hombres	Mujeres	Hombres	Mujeres
Aumentar dotación policial	36%	31%	27%	9%
Participación comunitaria	32%	29%	30%	40%
Otras soluciones	20%	32%	33%	36%
N.S./N. R.	12%	8%	10%	15%
	100%	100%	100%	100%
	100%			

La lectura comparada de las tablas A y B muestra el efecto de la variable interviniente - *nivel de escolaridad*-. Limitemos ahora nuestro análisis a la primera categorías de la variable *soluciones a la violencia ciudadana*: Una primera apreciación señalaría que es el nivel de escolaridad quién muestra un mayor peso, que la variable género, en la conformación de un tipo de opinión sobre posibles soluciones a la violencia ciudadana.

Una segunda mirada, más bien cualitativa, muestra que las tendencias originales tienden a mantenerse iguales si bien la introducción de la variable interviniente hace variar los guarismos en las dos categorías de la variable nivel de escolaridad.

Específicamente en aquellas personas de sexo masculino que declaran un mayor nivel de escolaridad, siguen mostrando una mayor inclinación que las mujeres por la solución del aumento de la fuerza policial en la ciudad, al igual en las personas clasificadas con un bajo nivel de escolaridad.

En síntesis, la introducción de *una variable de interviniente* puede provocar que *la relación original se sostenga, disminuya o aumente*. En el primer caso, si la relación original se sostiene se puede concluir que la variable interviniente no tiene ningún efecto. En las otras dos situaciones, se puede afirmar que *la variable interviniente* afecta la relación original, y la causa de esto puede ser que ésta antecede tanto como factor causal y temporal a ambas y/o, está relacionada tanto con la variable independiente como con la variable dependiente.

El proceso de agregar nuevas variables intervinientes puede llevar a producir una serie de efectos o situaciones en la relación original de las variables consideradas principales:

- Replicación: la relación original permanece inalterada, así se repitan en diferentes situaciones espacio-temporales;
- Explicación: la relación original es eliminada en los subconjuntos de datos -relación espuria-
- Interpretación: ocurre cuando la relación entre los subconjuntos es reducida o eliminada sin embargo es posible pensar que la variable de control interviene entre la variable independiente y la variable dependiente;
- Especificación: es el término usado para referirse a la situación donde la relación entre los subconjuntos de variables difiere uno de otro; es decir, la relación original puede aumentar sustantivamente en un subconjunto, por ejemplo, entre los hombres y desaparecer entre las mujeres.



### Procedimiento SPSS

- En la barra principal de Menús del SPSS, seleccione el menú *<Archivos>*
- Active la opción *<Abrir>* elija el submenú *<Datos >*
- Seleccionar archivo situación-2000.sav en el CD-ROM
- En la barra de Menús, elegir el menú *<Analizar>* *<Estadísticos  
descriptivos>*
- Seleccionar *<Tablas de contingencia>*
- En el cuadro de diálogo Tablas de contingencia →→ seleccionar la variable *sexo* y transferirla a la ventanilla *Filas* y la variable *diálogos en medio de la guerra* a la ventanilla *Columns*



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

En el visor de resultados, se pueden observar lo siguientes datos:

### **Análisis de los datos**

[ceo@carios.udea.edu.co](mailto:ceo@carios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

Con respecto a la opinión de los hombres, se puede apreciar que muestran una mayor predisposición al diálogo en medio de la guerra, un 46.2% de los entrevistados están de acuerdo; en cambio, el 66,6% por ciento de las mujeres entrevistadas muestran un mayor desacuerdo con esta posibilidad.

A la relación original planteada se le introduce una tercera variable interviniente *edad2003* para responder a la inquietud si es el género o la edad la que realmente explica la situación. Si la relación se mantiene *inmodificada o sostenida* dentro de los subconjuntos, pese a la existencia de esa la variable interviniente, se puede llegar a afirmar que se ha eliminado una posible causa alternativa en *la relación original* observada. Pero, sí la determinación entre las variables independiente y dependiente desaparece, nosotros sabremos que las dos variables no están relacionadas y la que realmente explica la situación es la nueva relación.

En el cuadro de diálogo Tablas de frecuencia, seleccionar la variable recodificada Edad 2003<sup>13[13]</sup> y transferirla a la 3<sup>a</sup> ventanilla –debajo de “Capa 1 de 1”. A continuación dar un clic en el botón Aceptar.

---

<sup>13[13]</sup> Categorías de la variable edad2003: jóvenes = 1 y adultos = 2



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

[ceo@carios.udea.edu.co](mailto:ceo@carios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

A continuación, en el visor de resultados aparecerá la siguiente tabla:

### **Análisis de los datos**

Se puede apreciar que en los jóvenes<sup>14[14]</sup> (código 1), la relación original persiste, tanto los hombres como las mujeres mantienen sus opiniones. En cambio, los adultos (código 2) manifiestan opiniones encontradas: Las mujeres adultas manifiestan un menor desacuerdo y los hombres un menor acuerdo. Se puede concluir, en términos generales, que el género es la variable explicativa.

### **Reagrupamiento de los valores de una variable**

En el supuesto que la información, tanto para la variable independiente como para la variable dependiente, sea de tipo interval, se puede anticipar un gran número de valores para cada variable. Una tabla de contingencia con muchas categorías o valores sería de difícil manejo, por no decir de imposible lectura. Para evitar esto, generalmente se debe

---

<sup>14[14]</sup> Código 1 = personas menores de 30 y menos años y Código 2 = personas de mas de 30 años.

resumir o reagrupar<sup>15[15]</sup> la información, lo que se trata es transformar a un número limitado de categorías (generalmente 8 o 10) lo que evidentemente puede implicar una pérdida de información. En síntesis, la recodificación de valores significa cambiar su valor actual por uno asignado por el analista.

### Procedimiento SPSS

- En la barra principal de Menús del SPSS, seleccione el menú <Archivos>
- Active la opción <Abrir> elija el submenú <Datos >
- Seleccionar archivo situación-2000.sav en el CD-ROM
- En la barra de Menús, elegir el menú <Transformar> →→ Recodificar
- Elegir la opción <En distintas variables>
- Aparece el cuadro de diálogo Recodificar en distintas variables:
- Se selecciona de la ventanilla con el listado de variables, la variable *edad* y se transfiere a la ventanilla central
- La variable aparecerá así en la ventanilla central; edad →?
- Escriba el nuevo nombre de la variable recodificada: *edad2003* en la casilla Nombre de la sección “Variable de resultado”
- De un clic en el botón C
- En la ventanilla central aparecerá así: edad → edad2003

---

<sup>15[15]</sup> Recodificación de los valores en rangos iguales



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

- Definido el nuevo nombre se hace clic en el botón “Valores antiguos y nuevos...”

[ceo@catios.udea.edu.co](mailto:ceo@catios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

- A continuación se abrirá el siguiente cuadro de diálogo:

En este cuadro de diálogo se sigue la siguiente secuencia de procedimientos:

- En la sección Rango situarse en la ventanilla izquierda y digitar el rango mínimo: 1
- En la ventanilla derecha, digitar el rango máximo: 2
- En la ventanilla Valor nuevo situarse en la casilla Valor y escribir el nuevo valor: 1
- Dé un clic en el botón **Añadir**
- Los nuevos valores recodificados aparecerán en la sección **Antiguo → Nuevo**
- Repita los cuatro (4) pasos anteriores con los valores originales y el nuevo.
- Dé un clic en el botón **continuar** y luego en el botón **Aceptar**
- En la matriz de datos aparecerá la nueva variable *edad2003*



## Inferencia estadística.

El objetivo de una investigación no se reduce al trabajo descriptivo solamente, en muchos casos, lo que se pretende es generalizar los resultados obtenidos a partir de una muestra de la población.

*"Los datos casi siempre son recolectados de una muestra y sus resultados estadísticos se denominan estadígrafos, por ejemplo, la media o la desviación estándar de la distribución de una muestra son estadígrafos. A las estadísticas de la población o universo se les conoce como parámetros. Los parámetros no son calculados, por que no se recolectan datos de toda la población, pero pueden ser inferidos de los estadígrafos, de ahí el nombre de estadística inferencial."*<sup>16[16]</sup>

La estadística inferencial sirve fundamentalmente para dos procesos: *Pruebas de hipótesis y estimativos de parámetros poblacionales.*

En esta lección, se indican algunas técnicas estadísticas en la que se usan las características de las muestras para probar una hipótesis que se ha formulado de un parámetro de población. Los procedimientos estadísticos a utilizar dependerán, obviamente, del nivel de medición de las variables independiente y dependiente. Cuando las dos variables son nominales u ordinales, generalmente se utilizan las llamadas tablas de contingencia. Si una es nominal u ordinal y la otra de intervalos o proporcional, *los test de diferencias de medias o de análisis de varianza* son las técnicas preferidas. Y cuando ambas variables son de intervalo o proporción, el análisis de regresión o correlación son los más adecuados.

En los procesos de medición de relaciones entre variables y prueba de hipótesis hay cinco preguntas que deben ser tenidas en cuenta: *¿Hay relación entre las variables independiente*

---

<sup>16[16]</sup> Hernández. Roberto et. al, *Metodología de la Investigación*. MacGraw-Hill. México, 1991, p. 373.

*y dependiente? ¿Cuál es la dirección o forma de la relación? ¿Cuán fuerte es la relación?  
¿Es la relación estadísticamente significativa? ¿Es la relación de tipo causal?*

Supóngase que se están evaluando terapias de rehabilitación en pacientes fármaco dependientes. Antecedentes sobre esos procesos de rehabilitación muestran que alrededor del 42% por ciento de los drogadictos que han asumido individualmente su terapia (tratamientos tradicionales) reinciden en estados de drogadicción.

Se hipotetiza que sólo el 28% por ciento de los drogadictos que participan en un programa innovador de rehabilitación reinciden en el consumo de alucinógenos. ¿Es posible, a partir de estas diferencias, afirmar que la metodología innovadora de dicho programa es la más adecuada para el tratamiento de la drogadicción?

No se puede aceptar o rechazar una hipótesis acerca de un parámetro de una población por simple intuición. En lugar de ello, se necesita aprender a decidir objetivamente, bajo las bases de la información muestral, cuando aceptar o rechazar una intuición.

Supóngase que se asume un cierto valor para la media de la población de reincidentes (42%). Para probar su validez de la suposición, se obtienen datos muestrales y se determina la diferencia entre el valor de la hipótesis y el valor real de la media muestral. Luego se juzga si la diferencia es significativa.

Mientras menor sea la diferencia, mayor será la probabilidad de que el valor hipotético para la media esté correcto. Pero, en no todas las ocasiones, la diferencia entre el parámetro hipotético de la población y el estadístico muestral es tan grande que automáticamente se rechace la hipótesis, ni tan pequeña que se acepte inmediatamente. Por tanto, en *pruebas de*

*hipótesis* como en la mayoría de las decisiones significativas, las soluciones precisas son la excepción, no la regla.

La ejemplificación anterior ilustra la necesidad de ir más allá de la sola descripción de las posibles diferencias cuantitativas entre dos resultados -terapias de rehabilitación-. Cabe preguntarse si las diferencias encontradas son un reflejo real o son simplemente *unas diferencias aritméticas no significativas*. Son en casos como éstos, donde las llamadas *pruebas de significación estadística* pueden ayudar a responder a este tipo de interrogantes. Estas herramientas estadísticas son útiles para aceptar o refutar las posibles diferencias o variaciones que se puedan presentar en un conjunto de unidades de análisis.<sup>17[17]</sup>

En síntesis, se trata de obtener conclusiones que sean generalizables a una población a partir de una información muestral resumida en un conjunto de estadísticos llamados inferenciales. Frecuentemente, por razones de tiempo y costo u otras consideraciones, no es posible examinar todas las unidades de análisis que componen el universo del estudio por eso se deben tomar las medidas del caso para que los resultados de la muestra que se observa y se analiza reflejen realmente las características de la población total.

Para lograr este objetivo es necesario conocer algunas características de la población en estudio para determinar el tipo de muestreo que requerimos. Luego, de obtenida de debe efectuar una prueba de hipótesis sobre los estadísticos obtenidos con la muestra (media y desviación estándar), con respecto a los parámetros poblacionales, para determinar su consistencia y estabilidad cuando el experimento se repita, con el fin de hacer algunas generalizaciones.

---

<sup>17[17]</sup> Individuos, grupos, instituciones, hechos, eventos, divisiones político-administrativas o geográficas, etc.

De otro lado, no olvidemos nuestra preocupación –en el ejemplo anterior- acerca de sí la tasa de reincidencia en el programa innovador es realmente concluyente y por lo tanto diferente a las tasas del "programa tradicional". Para responder el interrogante planteado el primer paso a desarrollar será el de definir *la población muestral de ambos grupos de rehabilitación*. Lamentablemente no hay técnica que nos defina una muestra poblacional ciento por ciento confiable; sin embargo, *una prueba de significancia estadística* puede permitirnos estimar la probabilidad de que el tamaño de las diferencias encontradas entre dos grupos se deba al azar. Con el fin de entender el concepto de significancia estadística, se hace necesario comprender la lógica del muestreo.

### **Prueba de hipótesis**

Una vez recogida la muestra probabilística, la importancia de ella radica en que permita predecir el comportamiento de la población. Recordemos que un buen modelo cuantitativo es aquel que logra eliminar la necesidad de explicaciones alternativas para el efecto causado, si es que lo hay, en la variable dependiente, ya que las muestras variarán desde su población por azar. Como en la mayoría de los casos no se conoce la media poblacional verdadera, podemos tener un valor aproximado como resultado del conocimiento o de la experiencia acerca del objeto de estudio.

El procedimiento para la contrastación de hipótesis comprende las siguientes etapas:

- Planteamiento de una hipótesis nula ( $H_0$ ) o hipótesis a contrastar.
- Formulación de una hipótesis alternativa de investigación ( $H_1$ ).
- Elección del riesgo o nivel de significancia que prácticamente es el complemento de lo que se llama nivel de confianza. Normalmente  $\alpha = 0.10$  ó  $0.05$ .
- Contraste de hipótesis. Comparar el valor encontrado del estadístico de prueba obtenido con el valor crítico y se toma *la decisión de aceptar o rechazar la hipótesis nula*.

La prueba de hipótesis comienza con una afirmación o propuesta tentativa, que se hace acerca de un parámetro de una población, llamada hipótesis.<sup>18[18]</sup> Luego se recogen datos muestrales, se producen estadísticos muestrales y se usa esta información para decidir qué tan probable es, que la hipótesis del parámetro de población esté correcta.

### **Selección del test estadístico.**

---

<sup>18[18]</sup> En la estadística inferencial hay dos tipos de hipótesis: Una. es la hipótesis de investigación que señala la posible existencia o relación entre dos hechos. La otra, es la hipótesis nula, que se define para que el analista evalúe su hipótesis de investigación.

Tengamos en cuenta que la hipótesis estadística es una conjetura o suposición concerniente a las características de la población; antes de aceptarla o rechazarla el investigador debe probar su validez. Ante la tarea de probar una hipótesis, se usa una muestra de acuerdo a la teoría de las probabilidades.

El resultado de la prueba conducirá a unos estadísticos, tales como *t de Student* o *el Chi-Cuadrado*, con los cuales el investigador podrá decidir sobre la validez de su propuesta.

### Valores de los niveles de significancia

El objetivo de una prueba de hipótesis no es calcular el valor de la estadística muestral pero sí hacer un juicio sobre la *diferencia* entre la estadística muestral y el parámetro hipotético de la población. Como se ha adelantado, el siguiente paso después de definir la hipótesis de investigación y la hipótesis nula es decidir qué criterio –valor- usar para aceptar o rechazar la hipótesis nula.

Supongamos que obtenemos un estadístico muestral (una media aritmética), entonces debemos decidir con qué porcentaje de confianza este estadístico se acerca al valor del parámetro poblacional. Debe buscarse un porcentaje alto de confianza, es decir, una probabilidad alta, ya que una aparente "cercanía" entre el valor calculado en la muestra y el parámetro puede ser ficticio (errores en la selección de la muestra).

En la investigación social existe la convención de aceptar como porcentaje de confianza válido el nivel de significancia del **0.05**, el cual significa que el analista tiene el 95% de seguridad para generalizar sin equivocarse.

La escogencia del nivel de significación estadística debe hacerse antes de calcular el porcentaje de probabilidad. Si el resultado estadístico muestra una probabilidad más baja, entonces "la hipótesis nula" deberá ser rechazada.

Frecuentemente, en los diseños cuantitativos, los niveles de significancia varían en un rango de 0.10 a 0.001. Si seleccionamos el nivel de significancia de **0.05**, sólo aquellos resultados que *sean iguales o menores al valor escogido, permitirán rechazar la hipótesis nula*. Únicamente reduciéndolo el nivel de significancia, se podrá reducir la probabilidad de rechazar la hipótesis nula cuando deberíamos aceptarla.

### Otros métodos para prueba de hipótesis.

Todos los test estadísticos tienen básicamente el mismo procedimiento para pruebas concernientes a la diferencia entre una media muestral y una media poblacional; la diferencia entre dos medias muestrales, la diferencia entre una proporción poblacional y la diferencia entre dos proporciones muestrales. Las variantes dependen del grado de conocimiento que se tenga sobre los parámetros poblacionales, así:

### Prueba *t* de Students.

Cuando en un estudio, de tipo experimental, se quiere comparar dos medias aritméticas, la prueba que se utiliza es la llamada *t de Student*. La hipótesis en juego permite establecer si existen diferencias significativas entre dos grupos respecto a sus medias<sup>19[19]</sup>. El nivel de medición de las variables debe ser de tipo intervalos, razón o proporcional. Para determinar si el valor "*t*" es significativo, hay que utilizar la Tabla de Distribución *t de Student*. Se trata de comparar el valor calculado con el valor de la Tabla, basándonos en el nivel de confianza elegido.

### Ejercicio de autocomprobación

A un grupo de diez sujetos se les pidió que leyera un texto sencillo y a otro grupo de diez se le pidió que leyera un texto complejo. Se predijo que se recordarían más palabras del texto sencillo que del texto complejo.

**Tabla. N° de palabras recordadas de un texto sencillo y de un texto complejo**

Grupo 1 (texto sencillo)	Grupo 2 (texto complejo)
Puntajes	Puntajes
10	2
5	1
6	7

<sup>19[19]</sup> La prueba "*t*" se puede utilizar para comparar resultados de un pretest con los resultados de un post-test.



	3		4
	9		4
	8		5
	7		2
	5		5
	6		3
	5		4
<b>Media<sub>1</sub>:</b>	<b>6.4</b>	<b>Media<sub>2</sub>:</b>	<b>3.7</b>

**Hipótesis de trabajo:** ¿Existe diferencia significativa entre las medias obtenidas por ambos grupos? – el promedio de 6.4 palabras recordadas del Grupo 1 es estadísticamente diferente al promedio de 3.7 palabras recordadas del Grupo 2. La pregunta de investigación hacia referencia a si era posible afirmar que las diferencias mostradas en la capacidad de recordación era significativamente validas.

Esquemáticamente la contrastación de hipótesis se presenta así:

**H<sub>0</sub>:** *No existen* diferencias estadísticamente significativas de puntajes entre los sujetos del grupo 1 (que leyeron un texto sencillo) y los del grupo 2 (texto complejo).

**H<sub>0</sub>: Puntaje Grupo 1 = Puntaje Grupo 2**

**H<sub>i</sub>:** *existen diferencias* estadísticamente significativas de puntaje entre los sujetos del grupo 1 (que leyeron un texto sencillo) y los del grupo 2 (texto complejo).

**H<sub>i</sub>: Puntaje Grupo 1 es diferente al Puntaje Grupo 2**

### Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos →→→ Abrir →→→ Datos
- Seleccionar el archivo de trabajo recuerdos.sav en el CD-ROM.
- Crear una *1ª variable de comparación: puntajes* compuesta por los valores obtenidos en el grupo 1 (*texto sencillo*) y grupo 2 (*texto complejo*)
- Crear una *2ª variable de comparación: códigos* con los siguientes valores consecutivos: 1= sencillo; 2 = complejo

- Seleccionar el menú Analizar →→→ elegir la opción Comparar Medias
- Elegir la prueba t para muestras independientes
- A continuación, aparecerá el cuadro de diálogo Prueba T para muestras independientes
- En primer lugar hay que seleccionar la variable dependiente: *puntajes* y transferirla a la ventana **Contrastar variables**:
- En segundo lugar, seleccionar la variable independiente: *códigos* – en donde están los grupos- y transferirla a la ventana “Variable de agrupación”

- Dar un clic en el botón Definir grupos
- A continuación aparecerá el cuadro de diálogo Definir grupos
- Contrastar variables: *Puntaje*
- Variable de agrupación: *Código (??)*
- Definir grupos: Como se van a comparar los grupos 1 –sencillo- y 2 – complejo- se escribe 1 en la Casilla al lado derecho del Grupo 1:
- Igualmente se escribe 2 en la casilla al lado derecho del Grupo 2:



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

[ceo@carios.udea.edu.co](mailto:ceo@carios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

A continuación, en el visor de resultados aparecerá la siguiente tabla:

Para aceptar la  $H_0$ , el valor observado del nivel de significancia de “t”-Sig (bilateral) tiene que ser *mayor o igual* que los valores escogidos - **10% (0.10) ó 5% (0.05)**- Para rechazar la  $H_0$ , el valor observado tiene que **ser menor** al *valor crítico escogido* -se acepta la  $H_1$ -.

Si se elige un nivel de significación del **5% (0.05)** y **el nivel de significación bilateral obtenido es mayor**, se debería aceptar *la hipótesis nula* según la cual “no existen *diferencias estadísticamente significativas* entre el grupo 1 y el grupo 2 y *si es menor* al valor crítico se rechaza la hipótesis de igualdad de puntajes -se acepta la  $H_1$ -.

### **Análisis de los datos**

A dos grupos de personas se les pidió que leyeran, a los primeros, un texto sencillo y a los segundos, un texto complejo. Se suponía que recordarían más palabras los lectores del texto sencillo que del texto complejo. Los primeros resultados así lo confirman: el promedio de palabras recordadas por los lectores del texto sencillo -*grupo 1*- fue de 6.4. El promedio de palabras recordadas por los lectores del texto complejo - *grupo 2*, fue de apenas 3.7.

El valor de la prueba t se expresa así **3,095 p ≤0,006**. Al analizar el valor del nivel de significación (p Valor) en la Prueba t para la igualdad de medias que es de **0.006**, es decir, *es menor al valor critico elegido 0,05*, luego hay que aceptar que las diferencias son significativamente válidas. Es decir, la diferencia encontrada en los puntajes de ambos grupos de lectura es válida.

### Comparación de medidas ordinales

Cuando los datos a comparar en dos sub-grupos corresponden a medidas de *nivel ordinal*, como por ejemplo, puntajes en escalas de actitudes y se esta lejos de cumplir con las exigencias de la prueba “t de Student”, se utiliza la *prueba no-paramétrica de Mann-Whitney*, llamada también de *Wilcoxon*, en la forma que se indica en el ejercicio siguiente.

### Ejercicio de autoevaluación:

Se ha aplicado una prueba de “conformismo al aprendizaje” a una muestra de adolescentes de los cuales 15 pertenecen a una escuela urbana y 21 a una escuela rural.

#### Puntajes de aceptación al aprendizaje

Estudiantes Urbanos		Estudiantes Rurales	
Escolar	Puntaje <sup>20[20]</sup>	Escolar	Puntaje
1	12	1	16
2	15	2	22
3	15	3	22
4	40	4	10
5	32	5	12
6	33	6	36
7	22	7	33
8	20	8	27
9	22	9	30
10	21	10	33
11	18	11	38
12	41	12	10
13	45	13	11
14	44	14	17
15	46	15	24

<sup>20[20]</sup> El grado de conformismo y aceptación a los modelos pedagógicos de aprendizaje se midió por un puntaje ascendente de 1 a 50 (donde 1 es el mínimo y 50 es el máximo).



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

16 16

17 39  
17 17

18 42  
18 18

19 40  
19 19

20 26  
20 20

21 25  
21 21

24



### Hipótesis de trabajo

**H<sub>0</sub>:** No existen diferencias estadísticamente significativas en el grupo de puntajes entre los estudiantes de escuelas de nivel *–urbano–* y los de nivel *–rural–*

**H<sub>0</sub>:** Puntajes de estudiantes urbanos = Puntaje de estudiantes rurales

**H<sub>1</sub>:** Existen diferencias estadísticamente significativas en el grupo de puntajes entre los estudiantes de escuelas de nivel *–urbano–* y los de nivel *–rural–*

**H<sub>1</sub>:** Puntajes de estudiantes urbanos son diferentes a puntajes de estudiantes rurales

### Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos →→→ Abrir →→→ Datos
- Seleccionar el archivo de trabajo conformismo.sav.
- En el Editor de datos: Crear una *1ª variable de comparación: puntajes* compuesta por los valores obtenidos en el grupo 1 (urbano) y grupo 2 (rural)
- Crear una *2ª variable de comparación: códigos* con los siguientes valores consecutivos: 1= urbano; 2 = rural

- □
- Seleccionar el menú Analizar →→→ elegir la opción **Pruebas no-paramétricas**
- Seleccionar: **2 muestras independientes Mann-Whitney**<sup>21[21]</sup>
- A continuación aparecerá el cuadro de diálogo: **Prueba para dos muestras Independientes**
- En primer lugar hay que seleccionar la variable dependiente: *puntajes* y transferirla a la ventana “Contrastar variables:”
- En segundo lugar, seleccionar la variable independiente: *códigos* – en donde están los grupos- y transferirla a la ventana *V*ariable de agrupación

---

<sup>21[21]</sup> Se utiliza para comparar dos muestras no relacionadas



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

- En la sección Tipo de prueba: Seleccionar U de Mann-Whitney

- Dar un clic en el botón Definir grupos
- A continuación aparecerá el cuadro de diálogo Definir grupos
- Contrastar variables: *Puntaje*
- Variable de agrupación: *Código (??)*
- Definir grupos: Como se van a comparar los grupos 1 –urbano- y 2 –rural- se escribe 1 en la Casilla al lado derecho del Grupo 1:
- Igualmente se escribe 2 en la casilla al lado derecho del Grupo 2:
  
- Definir grupos: Grupo 1 = 1; Grupo 2 = 2
-



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

[ceo@carios.udea.edu.co](mailto:ceo@carios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

A continuación, en el visor de resultados aparecerá la siguiente tabla:

### **Análisis de los datos**

El valor de la prueba U de Mann-Whitney se expresa así:  $140,000 \geq 0,574$  es decir, es mayor al valor crítico elegido de un 5%, luego hay que aceptar la  $H_1$  que las diferencias son significativas entre los estudiantes urbanos y los rurales.

### **Nota metodológica**

En nuestro caso, se eligió un nivel de significación de **5%** (0.05) y **el nivel de significación obtenido (Asintótica o exacta) es mayor**, se debería aceptar *la hipótesis alternativa* según

la cual “*existen diferencias estadísticamente significativas entre el grupo de puntaje entre los estudiantes de escuelas de nivel urbano y los de nivel rural.*”

$H_0 = A$  no se diferencia significativamente de B

$H_1 = A$  se diferencia significativamente de B

## Medidas de relaciones entre variables

En esta lección examinaremos la relación entre variables. De esta forma completaremos todo el escenario que se refiere a la existencia o no de relaciones entre variables medidas en cualquier nivel de medición. De acuerdo a Briones<sup>22[22]</sup> “Una de las principales funciones de la investigación social -cualquiera que sea el campo disciplinario particular en la cual se efectúa- consiste en buscar relaciones entre el objeto principal de estudio y otros objetos o, si se quiere, entre características principales del objeto de estudio y otras características del mismo o del contexto en el cual se encuentra”.

Cuando se habla de posibles relaciones entre dos variables, lo que se está haciendo es analizar si: Existe o no relación entre las variables, la forma o dirección de la relación y la fuerza de la relación.

También Briones señala que *se acostumbra distinguir con el término "asociación" la relación que se da entre características cualitativas o semi-cualitativas (variables nominales y ordinales) y con el término "correlación" la relación entre características cuantitativas.*<sup>23[23]</sup>

### Nota metodológica

Cuando el tipo de medición utilizado es cuantitativo, en el caso específico de variables tipo interval y razón o continuas, el rango de valores varía entre -1.00, lo que significa la existencia de una asociación perfecta de tipo negativa, hasta +1.00, que señala una asociación positiva perfecta. Tales mediciones se dicen ser "direccionales" y reflejan la

---

<sup>22[22]</sup> Guillermo Briones, Op. Cit., p. 76.

<sup>23[23]</sup> Briones. Op. Cit. p. 76





UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

propiedad de direccionalidad. Con variables cualitativas, a partir del *nivel de significancia*, se podrá determinar la existencia o no de la asociación entre las variables.

**Cuadro sinóptico de medidas de relaciones simples entre variables.**

**Principales medidas de relación de variables nominales, ordinales, intervalo y proporcionales.**

	<b>Datos Nominales Proporcionales</b>	<b>Datos Ordinales</b>	<b>Datos Intervales</b>	<b>Datos</b>
<b>Medidas de relación</b>	comparación de porcentajes Coeficientes de Contingencia Chi - Cuadrado Coeficiente de Goodman-Kruskal Lambda	tau b  Tau c  D de Somer  Gamma		Coeficiente de Pearson         Coeficiente Rho de Spearman

**Medidas de asociación entre variables cualitativas.**

Examinemos algunas *estadísticas no-paramétricas* que permiten medir asociaciones entre variables medidas a nivel nominal y ordinal.

**Coefficiente de contingencia Chi-cuadrado.**

En este caso, el coeficiente Chi-cuadrado no es propiamente un medida de asociación, sino más bien una prueba de independencia. Esta es una prueba de significancia estadística no-paramétrica, utilizada con variables medidas a nivel nominal u ordinal que permite determinar si existe o no, una relación-asociación sistemática entre dos variables.

*¿Qué es la Prueba del Chi Cuadrado de Pearson?* Es una prueba estadística no-paramétrica para **evaluar hipótesis** acerca de la relación entre dos variables cualitativas. Es una prueba que parte del supuesto de “**no relación entre variables**” y el investigador evalúa si en su caso **esto es cierto o no** y esto es a menudo de interés.

### Ejercicio de auto comprobación

La modalidad de estudio es independiente al programa académico en cual esta vinculado el estudiante. Operacionalmente hablando, esta afirmación se puede plantear por ejemplo, en la siguiente pregunta **¿las variables “Tipo de estudiante según Programa” (tipo) y “modalidad de estudio” (estudio) son independientes entre sí?**

Para contestar esta pregunta se hace uso de **la prueba de hipótesis basada en el Chi - Cuadrado**: La **aceptación de  $H_0$**  (hipótesis nula) indica **la independencia** entre las variables. Sí se rechaza  **$H_0$** , se acepta **la hipótesis alternativa  $H_i$** , o sea, **la no independencia (relación)** entre las variables.

La prueba de hipótesis se plantea como:

**$H_0$ : Independencia** entre las variables - **Tipo de estudiante y modalidad de estudio-**

**$H_i$ : No independencia** entre las variables -**Tipo de Estudiante y modalidad de estudio-**

### Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos →→→ Abrir →→→ Datos
- Seleccionar el archivo de trabajo *habitos2003.sav* en el CD-ROM
- Seleccionar el menú Analizar →→→ Seleccionar la opción <Estadísticos descriptivos> →→→ Activar opción <Tablas de contingencia>

A continuación aparecerá el cuadro de diálogo Tablas de contingencia:

- Transferir a la ventanilla Filas, la variable independiente *tipo*
- Transferir a la ventanilla Columnas, la variable dependiente *estudio*
- Dar clic en el botón *Estadísticos...*
- A continuación aparecerá el cuadro de diálogo *Tablas de contingencia: Estadísticos*
- Complete los campos requeridos: marcando el estadístico Chi-cuadrado
- Dar clic en el botón Continuar



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

De un clic sobre el botón *Continuar*

A continuación aparecerá el cuadro de diálogo siguiente:

[ceo@catios.udea.edu.co](mailto:ceo@catios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

Al completar los campos de acuerdo al modelo anterior: Marcar la opción “Esperadas”  
A continuación aparecerá en el Visor de resultados la siguiente tabla:

### **Análisis de los datos.**

La pregunta que se planteo es ¿existe relación entre el tipo de estudiante según programa académico y la modalidad de estudio que utilizan? El contraste de Chi-cuadrado comprueba que si existe independencia entre ambas variables. El valor del Chi cuadrado de Pearson de **3.107**  $p \leq 0,212$  indica que “el tipo de estudiante” y “la modalidad de estudio” son independientes ya que la probabilidad de que una muestra aleatoria arroje un valor de Chi cuadrado de al menos dicha magnitud es de **0.212**. Esta probabilidad también se conoce como **nivel de significación observado** de la prueba.

Si la probabilidad es suficientemente pequeña (habitualmente menor que **0.05 ó 0.10**), se rechazará la hipótesis de que *las dos variables son independientes*. Pero, *es mayor*, de acuerdo a los resultados es **0.212**, se debe aceptar que *no hay relación o asociación entre las variables*.

Por lo tanto, se puede aceptar la hipótesis nula, de la independencia<sup>24[24]</sup> entre la modalidad de estudio y el tipo de estudiantes (de las facultades de Ciencias Sociales y de Medicina Veterinaria).

### **Nota metodológica**

En síntesis, el Chi-cuadrado permite decidir si las variables, de tipo cualitativo, son independientes o asociadas, pero no indica que tan fuerte es esa relación.

Conjuntamente con Chi cuadrado existen otras medidas para evaluar si las variables incluidas en una tabulación están asociadas, veamos algunas de ellas:

- Comparación de porcentajes: Es una medida de asociación que consiste en comparar los porcentajes en una tabla, en el cual una de las variables es considerada como causal o independiente y sirve como base del cálculo de los porcentajes.
- Lambda: es una medida de asociación para cruces de variables que se usa comúnmente con variables medidas a nivel nominal.

---

<sup>24[24]</sup> No existe asociación.

### **Asociación entre variables nominales**

En esta unidad se introduce la noción de relación entre variables en contraste con los diseños experimentales que predicen diferencias entre las condiciones experimentales. Es decir, un investigador podría estar interesado no tanto en el desempeño de los dos grupos preseleccionados, sino más bien en observar cómo todo el conjunto del rango de los puntajes en una variable, por ejemplo **los puntajes de capacidad lectora**, se relaciona con todo el conjunto del rango de puntajes en otra variable, por ejemplo los puntajes que se representan **puntajes en estadística** obtenidos por los estudiantes de la Facultad de Administración de Empresas.

En un caso como éste un investigador mediría los puntajes de las dos variables en un solo grupo de sujetos, con el fin de investigar hasta qué punto las características individuales de cada uno sujeto en cada variable se correlacionan con su desempeño en la otra variable. La pregunta por resolver es si la habilidad en lectura y la habilidad en estadísticas “*van juntas*” (*asociadas*), o si no hay ninguna conexión entre los dos tipos de habilidades.

### **Ejercicio de autocomprobación**

Supongamos que a un investigador le interesa saber si existe relación entre la habilidad lectora y la habilidad en estadística. El objetivo sería más bien estudiar si los estudiantes que obtienen un puntaje alto o bajo en una variable, también obtienen puntajes altos o bajos en la otra.



**TABLA. Puntajes en estadística y en lectura**

Estudiante	Puntaje de Estadística	Puntaje de lectura
1	9	2
2	10	3
3	12	1
4	6	1
5	11	4
6	9	1
7	12	5
8	16	8
9	13	5
10	10	3
11	13	6
12	14	7

**Hipótesis de trabajo:**

**H<sub>0</sub>:** No existe asociación entre los puntajes obtenidos en estadística y los puntajes obtenidos en habilidad lectora *-independencia entre las variables-*

**H<sub>1</sub>:** Existe asociación entre los puntajes obtenidos en estadística y los de habilidad en lectura.

**Nota metodológica:** Para datos nominales (clasificatorios) se utilizan los siguientes coeficientes: *coeficiente de contingencia, coeficiente Phi, coeficiente V. de Cramer, coeficiente Lambda y coeficiente de incertidumbre.* Finalmente, digamos aquí que cuando se trata de muestras, la posible asociación debe ser sometida a una prueba de significación estadística.

### Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos →→→ Abrir →→→ Datos
- Seleccionar el archivo de trabajo nominal.sav en el CD-ROM
- En la barra de Menús, seleccionar el menú Analizar
- Elija la opción **<Estadísticos descriptivos>** → **<Tablas de contingencia>**
- En el cuadro de diálogo **Tablas de contingencia** →→ Estadísticos:
- Seleccionar los estadísticos para **variables nominales**:
- *Coefficiente de contingencia C; Phi y V. de Cramer.*
- En el cuadro de diálogo *Tablas de contingencia: mostrar en las casillas:*  
Porcentajes: *seleccionar filas*
- A continuación, en el visor de resultados aparecerá la siguiente tabla:

### Análisis de los datos

Como el nivel de significación es mayor al 5% -valor crítico escogido-, se acepta la  $H_0$ , es decir, que no existe relación entre la habilidad lectora y los puntajes obtenidos en habilidad estadística.



**Nota metodológica**

$H_0$  = no existe asociación entre las variables                      sí es mayor al valor crítico  
escogido

$H_1$  = existe asociación entre las variables                      sí es menor al valor crítico escogido

**Asociación entre variables ordinales: Coeficiente de asociación por rangos de Spearman.**

Determina la asociación entre dos variables medidas en escalas ordinales. Una forma de medir el grado de asociación entre dos variables medidas a nivel ordinal la proporciona el coeficiente de asociación de Spearman o de rangos.

**Ejercicio de autocomprobación**

Suponga que nos interesa averiguar si hay alguna relación entre la participación de los estudiantes en deportes al aire libre -medida sobre una escala desde 1 (nunca juega) hasta 10 (siempre juega) y la frecuencia con que los mismos estudiantes se ausentan –medida en una escala desde 1 (con frecuencia – hasta 5 (rara vez)).

Tabla: Puntajes de participación de estudiantes en deportes y frecuencia de ausentismo.

	<b>Variable A</b>	<b>Variable B</b>
<b>Sujeto</b>	<b>Puntaje en escala de participación deportiva (1 – 10)</b>	<b>Puntaje en escala de frecuencia de ausentismo (1 – 5)</b>
1	5	2
2	2	3
3	3	7
4	10	5
5	9	4
6	9	5
7	2	4
8	6	3
9	3	1
10	4	1



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

11            8  
12

10            4

5

**Hipótesis de trabajo:** La hipótesis de trabajo afirma que una alta participación de los estudiantes en los actividades deportivas esta relacionada con bajas tasas de ausentismo académico.

*H<sub>0</sub>: No existe asociación entre los puntajes obtenidos de participación deportiva y ausentismo académico (Independencia entre las variables).*

*H<sub>i</sub>: existe asociación entre los puntajes en participación deportiva y ausentismo académico.*

Sujeto	Puntaje en escala de participación deportiva (1 – 10)	Rango A <sup>25[25]</sup>	Puntaje en escala de frecuencia de ausentismo (1 – 5)	Rango B
1	5	5	2	3.5
2	3	2.5	2	3.5
3	7	7	4	7.5
4	10	12	5	11
5	9	9.5	4	7.5
6	9	9.5	5	11
7	2	1	4	7.5
8	6	6	3	5
9	3	2.5	1	1.5
10	4	4	1	1.5
11	8	8	4	7.5
12		10		
		12		
		5		

11

<sup>25[25]</sup> Crear rangos para ambas variables

### Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos →→→ Abrir →→→ Datos
- seleccionar el archivo de trabajo deportes.sav en el CD-ROM
- En la barra de Menús, seleccionar el menú Analizar
- Elija la opción <Correlaciones> →→→ Activar opción <Bivariada>
- Aparece el cuadro de diálogo Correlaciones bivariadas:

- Seleccionar y transferir cada una de las variables *rangoa* y *rangob* a la ventanilla Variables:
- Seleccionar **Spearman**<sup>26[26]</sup> » » » seleccionar Prueba de significación: » » Bilateral

---

<sup>26[26]</sup> El estadístico Rho de Spearman es una prueba no paramétrica que mide el grado de asociación entre los puntajes obtenidos en dos variables e indica *el nivel de significación de la*



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

---

*asociación observada*. Debe usarse cuando los datos experimentales se miden en una escala ordinal. Para que Rho pueda considerarse significativo, debe ser *igual o mayor* que los valores críticos escogidos - **0.10 ó del 0.05** o, lo que es lo mismo adoptar el nivel de significación del 10% o del 5%.

[ceo@catios.udea.edu.co](mailto:ceo@catios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775



A continuación, en el visor de resultados aparecerá la siguiente tabla:

**Análisis de los datos:** el nivel de significación estadística encontrado es de  $0.003$ , es decir, menor al valor crítico escogido ( $.05$ ), por lo tanto, no existe asociación entre los puntajes que miden la participación deportiva y el ausentismo académico.

### **Análisis de regresión.**

El análisis de regresión y correlación permite mostrar cómo deberíamos determinar tanto la naturaleza como la fuerza de la relación que se puede existir entre dos variables. También es posible predecir el valor de una variable desconocida basados en observaciones pasadas de esa u otra variable.

Lo anterior exige, en muchos casos, el uso de una técnica que permita usar toda la información lograda referente a un juego de variables y al mismo tiempo, describir la relación entre dos o más variables en una forma simple y precisa; ante tal requerimiento, el análisis de regresión es la técnica más apropiada.

### **Regresión lineal**

Esta es una técnica estadística que permite analizar la relación entre una variable dependiente – criterio- y un a o más variables – predictoras- El nivel de medición de las variables tiene que ser de intervalo o razón. A su vez, las variables del estudio deben ser, una considerada como independiente y la otra como dependiente -decisión de tipo teórico-metodológica-.

Permite resumir los datos y cuantificar la naturaleza y fuerza de la relación entre variables. También se puede usar ésta técnica para **predecir** o estimar nuevos valores de la variable dependiente de acuerdo a los valores observados en la variable independiente.

### **Ejercicio de autocomprobación**

Supongamos que necesitamos analizar el impacto de un programa de capacitación en sistematización en el nivel de ingresos de un grupo de egresados del Servicio Nacional de Aprendizaje; se cuenta con información acerca del número de horas de capacitación

adquiridos en sistematización y el monto de los ingresos de los egresados (población muestral)

### Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos →→→ Abrir →→→ Datos
- Seleccionar el archivo de trabajo ingresos.sav en el CD-ROM
- En la barra de Menús, seleccionar el menú Analizar
- En el menú Analizar » » » seleccione la opción Regresión » » » elija Lineal

- Seleccionar y transferir la variable *ingresos mensuales* a la ventanilla *Dependiente*.
- Seleccionar la o las variables predictoras o independientes –horas de capacitación– a la ventanilla *Independiente* y dar clic en el botón *Aceptar*

### **Nota metodológica**

El primer valor es *la variable dependiente ingresos mensuales percibidos* y se trasfiere a la casilla Dependiente: -el segundo, es *la variable independiente horas - número de horas de capacitación en sistematización-* y se trasfiere a la cásilla Independientes:

A continuación, en el visor de resultados aparecerá la siguiente tabla:

### **Análisis de los datos**

Lo primero que interesa conocer es la fuerza de la correlación entre la variable predictora *horas (número de horas de capacitación)* y la variable criterio *Ingresos (ingresos percibidos)*. Esto se mide con el *coeficiente de correlación R cuadrado*-  $R^2 = .898$ , ello indica que el 89,8 por ciento de la varianza de la variable ingresos está predicha por la variable horas de capacitación.

### **Nota metodológica**

El coeficiente  $R^2$  puede variar entre 0 y 1. Cuando  $R^2 = 1$  quiere decir que el 100% de la variabilidad de la variable criterio es explicada por la variable predictora. Cuando  $R^2 = 0$  significa que no hay ninguna relación lineal entre la variable predictora y las variables criterios.

### **Coefficientes de correlación.**

A menudo, los analistas sociales están interesados en descubrir cuál o cuáles variables están más estrechamente asociadas a una determinada variable dependiente, en este caso los coeficientes de correlación se convierten en la herramienta estadística adecuada. En síntesis, el análisis de correlación es una herramienta que se puede usar para describir el grado en el que una variable está relacionada con otra.

Hasta ahora sólo ha interesado saber si hay o no una relación entre variables (si es significativa o no) pero nada se sabe si la relación obtenida es fuerte o débil -fuerza de la relación-. Cuando existe la relación, es conveniente preguntarse: ¿Cuál es su fuerza? Analicemos las formas más utilizadas para detectar la fuerza de la relación entre variables

### **Correlación de Pearson**

El coeficiente de correlación de Pearson  $r$  mide básicamente la naturaleza de la relación entre dos variables cuantitativas. No solo sirve para indicar si hay o no relación, sino también para indicar *la fuerza –grado- y dirección* de la misma. Consiste en una prueba "paramétrica" que mide el grado de correlación entre *los puntajes* obtenidos en dos variables e indica el nivel de significación de la correlación observada. La prueba de Pearson requiere de una medición de los datos en una escala interval o de razón.

El coeficiente de correlación lineal  $r$  refleja el grado de relación entre dos variables y lo expresa con un número que va desde "-1" (correlación negativa perfecta) pasa por "0" (no existe correlación) hasta +1 (correlación positiva perfecta). En síntesis, puede tomar valores comprendidos entre -1 y 1. Si el coeficiente es positivo expresa *una covariación* de las variables en el mismo sentido, y un coeficiente negativo, lo contrario.

La fuerza o intensidad de la relación puede ser débil o fuerte o inclusive puede no existir. Generalmente se acepta el criterio según la cual, si la relación entre las variables se acerca a cero (0), indica debilidad y si se acerca a uno 1 sería indicativo de una alta asociación o correlación.

### Ejercicio de autocomprobación

Mediante una prueba se ha medido la habilidad<sup>27[27]</sup> en comprensión y lectura y el uso de las herramientas estadísticas en un grupo de alumnos del curso de Introducción a la Investigación social.

Se ha hipotetizado que existe una correlación positiva entre la habilidad en comprensión y lectura de textos y la habilidad en la utilización de la herramienta estadística, es decir, que los alumnos que tienen un buen desempeño en comprensión y lectura de textos lo tendrán también en el uso de la estadística, como herramienta de análisis.

**Tabla: Puntajes obtenidos en la prueba de lectura y comprensión y estadística social.**

Alumno	Variable A <sup>28[28]</sup>	Variable B <sup>29[29]</sup>
1	30	20
2	50	10
3	60	20
4	73	12
5	87	16
6	24	7
7	44	9
8	35	18

<sup>27[27]</sup> Instrumentos de medición generan puntajes cuantitativos que miden las capacidades de uso de las herramientas.

<sup>28[28]</sup> Variable A: puntaje obtenido en prueba que miden el grado de utilización de la herramienta estadística como elemento de análisis.

<sup>29[29]</sup> Variable B: puntaje obtenido en prueba de comprensión y lectura de textos.



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

9	97	19
10	62	13
11	49	6
12	93	10



### Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos →→→ Abrir →→→ Datos
- Seleccionar el archivo de trabajo *nominal.sav* en el CD-ROM
- En la barra de Menús, seleccionar el menú Analizar → →Correlaciones
- Seleccionar → → Bivariadas → → el cuadro de diálogo: Correlaciones bivariadas.
- Seleccionar y transferir las variables independiente y dependiente a la ventanilla Variables
- En la sección Coeficientes de correlación elegir **Pearson**
- En la sección Prueba de significación elegir **Bilateral**
- Al presionar el botón **Aceptar** aparecen los resultados en el Visor

Resultados en el Visor:

### Análisis de los datos.

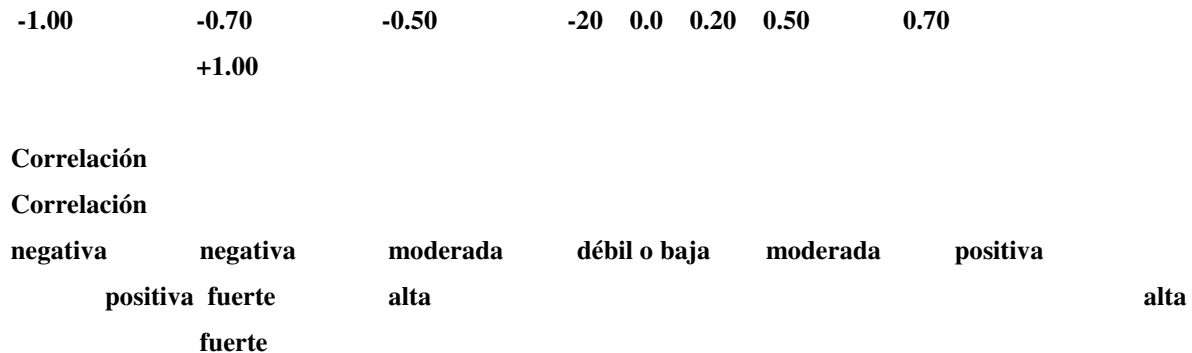


UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

Se puede concluir que existe una correlación significativa entre las dos variables: Habilidad en estadística social y habilidad en comprensión y lectura, ya que el coeficiente de correlación entre ellas sí es significativo –  $r = 0,857$   $p \leq 0,000$

### Nota metodológica

En el gráfico siguiente se observa una posición relativa de un coeficiente  $r$ :



### Algunos comentarios sobre el uso de la regresión y correlación.

Una pregunta básica tiene que ver con la oportunidad del uso de las técnicas de regresión. Más adelante, en esta lección, cuando hablemos de **regresión múltiple**, volveremos sobre este interrogante. Por lo pronto, digamos que raramente un analista social trabaja en el estudio del impacto de una sola variable. El foco de atención generalmente se centra en el impacto o determinación de dos o más variables independientes sobre un hecho o proceso social. En casos como estos, la técnica más adecuada es la de regresión múltiple.

Acerca de la conveniencia de usar las técnicas de correlación o las de análisis de regresión, es evidente que esta última tiene mucho más que ofrecer para un analista social. Fundamentalmente porque motiva al investigador a considerar más cuidadosamente la dirección causal entre las variables, por ejemplo, se deberá tener claro cuál es la variable independiente y cuál es la dependiente. También, el resultado del cálculo de una regresión:

[ceo@carios.udea.edu.co](mailto:ceo@carios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

el coeficiente de regresión tendrá en términos generales, un significado más sustantivo que un coeficiente de correlación.

### **Análisis multivariado.**

Existen algunas técnicas estadísticas que han sido desarrolladas para estudiar problemas complejos desde una perspectiva multivariada, es decir, considerar en el análisis, todas las posibles interrelaciones entre las variables, para lograr un mayor entendimiento y explicación de un proceso. En síntesis, el análisis multivariado es aquel que analiza la relación entre varias variables independientes y, al menos una variable dependiente.

### **Regresión múltiple.**

En lecciones anteriores describimos las técnicas de relación entre variables. La discusión se centró en el análisis de dos variables, por ejemplo, el efecto de una variable independiente en una dependiente. Sin embargo, la utilidad del modelo de regresión es mejor percibido a través del uso de la técnica de regresión múltiple. Esta herramienta estadística permite analizar la naturaleza de una variable dependiente a partir de 2 o más variables independientes o descriptoras. Sirve para predecir el valor de una variable dependiente conociendo el valor y la influencia de las variables independientes incluidas en el modelo.

Permite evaluar el valor individual de cada variable independiente en la variable dependiente mientras se controla tal influencia por el impacto de otras variables al igual que la influencia de todas las variables independientes combinadas.

Otra de las ventajas de la regresión múltiple sobre el análisis de tablas de continencia se basa en la cantidad de variables de control o intervinientes que se pueden utilizar; en las tablas de continencia, se limita el uso a una o dos variables intervinientes; en cambio, con la regresión múltiple no existe tal limitación.

### **Ejercicio de autoevaluación**

Se aplicó una encuesta a los alumnos destinada a conocer la opinión sobre sus cursos y el desempeño de los profesores que los dictan. Se les pidió que evaluaran tanto al profesor como al curso de acuerdo a la siguiente escala de calificaciones:

Muy deficiente 1      deficiente 2      aceptable 3      bueno 4      excelente 5

Objetivo: Estudiar la relación entre la evaluación del profesor de un curso realizada por sus alumnos y cinco características del profesor y del curso: Dominio de los temas del curso, empleo de recursos didácticos, respeto en el trato con los alumnos, evaluación de ejercicios y exámenes, evaluación del interés de los contenidos.

### Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos →→→ Abrir →→→ Datos
- Seleccionar el archivo de trabajo encuesta al estudiante sobre el curso y el profesor.sav (el archivo de trabajo esta en el CD-ROM)
- En la barra de Menús, seleccionar el menú Analizar →→ Regresión →→→ Lineal
- Seleccionar y transferir las variables predictoras o independientes: dominio de los temas, empleo de recursos didácticos, respeto en el trato con estudiantes, evaluación de ejercicios y exámenes y evaluación del interés de los contenidos a la ventanilla “Independientes:” y dar un clic sobre el botón Aceptar
- Dar un clic sobre el botón Estadísticos
- A continuación aparecerá el cuadro de diálogo Regresión lineal: Estadísticos
- Seleccionar y marcar las opciones: Estimaciones; Descriptivos y Correlaciones parcial y semiparcial
- Dar un clic sobre el botón **Continuar**



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

[ceo@carios.udea.edu.co](mailto:ceo@carios.udea.edu.co)

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

A continuación, en el visor de resultados aparecerá la siguiente tabla:

En la tabla *Estadísticos descriptivos* se muestran los promedios de la evaluación estudiantil y la desviación estándar o típica de cada una de las seis variables, lo que permite visualizar los promedios de evaluaciones y los casos extremos.



### **Análisis de los datos**

El peso o influencia que tiene cada variable independiente sobre la variable dependiente se muestra en la tabla correlaciones. En términos generales, se puede afirmar que hay algunas correlaciones significativas entre las variables predictoras o independientes – *dominio del tema y evaluación de ejercicios y exámenes* con la evaluación del profesor.

A continuación en el Visor de resultados aparecen los coeficientes que miden la correlación de todas las variables independientes en la variable dependiente – *calificación del profesor-*

### **Análisis de los datos**

Un  $R^2 = 763$  indica que el 76 % de la variación total de los promedios de evaluación se explica por estas 5 variables que se usaron. Se puede concluir que el  $R^2$  es significativo, ya que F es igual a **12,249**  $p \leq 0.000$

### **Nota metodológica**

Para utilizar la técnica de análisis de regresión múltiple es necesario que las variables estén medidas en escalas de intervalos o proporcionales.