

6. Análisis de la información

Objetivos

Familiarizar al usuario con algunos instrumentos de análisis de la información.

Una vez planteado el problema que se va a investigar y se ha definido el marco de referencia teórico, conceptual y metodológico que guiará el proceso de investigación, es necesario considerar cómo analizar la información obtenida.

El analista social debe saber qué tipo de técnicas son más pertinentes para el manejo de la información disponible y cómo analizar los resultados obtenidos. La importancia de esto se basa en la idea de que el analista no debe usar ciegamente los instrumentos estadísticos, es decir, debe ser capaz de juzgar su pertinencia en relación con el problema de estudio, y saber leer y evaluar la calidad de los resultados. Generalmente, el analista ciego considera que la objetividad de los resultados están avalados por el carácter que estos tienen, cuando es un producto casi mágico de un proceso de sistematización, lo cual conduce a muchos errores.

En los estudios sociales se necesita con frecuencia analizar el comportamiento de un conjunto de datos con el objeto de determinar hechos o características 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* se va a entender en una doble perspectiva: como un proceso de búsqueda de elementos, aspectos o configuraciones que caractericen y contextualicen los procesos estudiados, y como una simple lectura de los resultados obtenidos.

Lo anterior está relacionado que el análisis de los datos, depende de una serie de elementos, entre los cuales los más importantes a considerar son:

- Los objetivos del estudio.
- La hipótesis de investigación.
- La información disponible.

Otra consideración es, cuando se trabaja con una 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. Guillermo Briones 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 en 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 frecuencias, 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 aparecer.¹

Procedimiento para la sistematización de la información

En este texto se utilizará el programa de manejo y gestión de datos SPSS®, éste es sobretodo, un programa de procesamiento estadístico y su mérito fundamental radica en la amplitud y el rigor de las pruebas estadísticas que contiene. A medida que se van desarrollando las técnicas estadísticas de análisis de datos, se muestra el procedimiento correspondiente, en una sección dentro del texto, que se denomina *Procedimiento SPSS®*.

Para manejar el SPSS® a continuación se describen los elementos que lo constituyen, sus funciones y los procedimientos básicos para la sistematización de la información.

1 Briones, Guillermo, *Evaluación interactiva. Serie Métodos y Técnicas avanzadas de investigación social aplicadas a la educación y las ciencias sociales*. Bogotá: Procesos Editoriales ICFES. 1989. Pág. 23

Elementos del SPSS® y sus funciones

El programa SPSS® para Windows® consta de *Ventanas, Barras de herramientas, Cuadros de diálogos y Menús* que facilitan la acción del usuario. Aquí se incluye una breve descripción de estos conceptos para facilitar su manejo.

Para iniciar el programa, haga <double clic> con el botón izquierdo del ratón sobre el icono del SPSS® (véase figura 6.1)



Figura 6.1 Icono del SPSS®

Fuente: SPSS® para Windows®

Manejo del Editor de datos SPSS®

El Editor de datos proporciona un método práctico² para la creación y edición de archivos de datos. La ventana Editor de datos se abre automáticamente cuando se inicia una sesión. Ver figura 6.2.

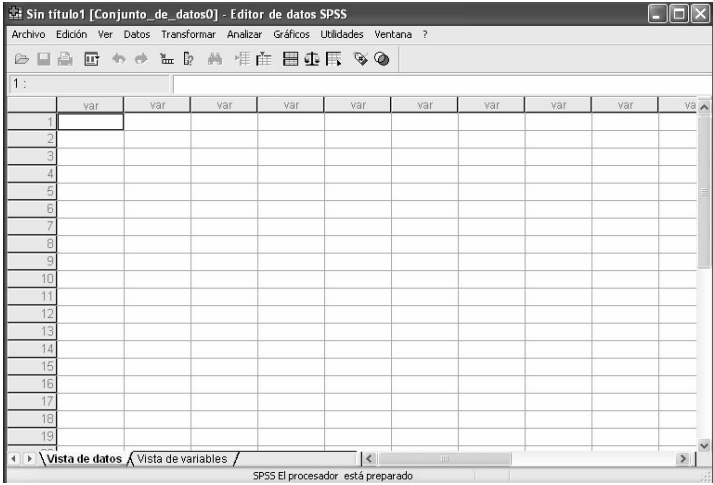


Figura 6.2 Cuadro de diálogo Vista de datos

Fuente: SPSS®

2 Al estilo de las hojas de cálculo compuesta por una matriz de filas, columnas y celdas.

El SPSS® organiza el trabajo en varias ventanas. Cuando se accede al programa aparece la ventana del Editor de datos SPSS compuesta por:

- Menú principal del SPSS®
- Barra de herramientas e iconos
- Solapas de vista de datos y variables
- Cuadros de diálogo

La pantalla principal del SPSS® presenta una barra con diez menús desplegables tipo cascada. Cada menú tiene las siguientes funciones:

Archivo. Crea un nuevo archivo de SPSS®, abre un archivo ya existente o lee archivos de bases de datos ya creados.

Edición. Modifica o copia textos de las ventanas de resultados.

Ver. Visualiza la barra de estado y la barra de herramientas.

Datos. Realiza cambios globales en los archivos de datos de SPSS®, tales como la fusión o división de archivos, la transformación de variables y casos, o la creación de subgrupos de casos para su análisis.

Transformar. Opera cambios en variables seleccionadas en el archivo de datos y calcula nuevas variables basadas en los valores de las ya existen.

Analizar. Selecciona los diferentes procedimientos estadísticos que se desean utilizar, como tablas de contingencia, análisis de varianza, correlación y regresión lineal, etc.

Gráficos. Crea gráficas de barras, de sectores, histogramas, gráficos de dispersión y otras de alta resolución y color. Algunos procedimientos estadísticos también generan gráficas.

Utilidades. Visualiza información acerca del contenido de los archivos de datos y sus variables, abre un índice de los comandos del SPSS®.

Ventana. Dispone, selecciona y controla los atributos de las diferentes ventanas del SPSS®.

Ayuda. Abre una ventana de ayuda que contiene información acerca del uso de diversas operaciones del SPSS®.

Para seleccionar las opciones del menú, se procede de la siguiente manera: llevar el puntero del ratón a la Barra de menús y señalar uno de ellos. Al darle <clik> se despliegan hacia abajo mostrando las opciones o submenús de trabajo.

A través de otras ventanas temporales, el SPSS® interactúa con el analista para completar información necesaria que sirva para desarrollar los procesos necesarios para el análisis de los datos, también para dar ayuda sobre un procedimiento específico (*ventanas de ayuda*) o, en otros casos para dar mensajes de error o advertencias (*ventanas de mensajes*).

Cada una de las ventanas tiene una barra de menús desplegables y una barra de iconos. Tanto la barra de menús como la de iconos son distintas en cada ventana, aunque algunos elementos de los menús y algunos iconos están disponibles en varias de ellas.

Barras de herramientas. Cada ventana del SPSS® tiene su propia barra de herramientas que proporciona acceso a las tareas más frecuentes.



Figura 6.3 Barra de herramientas del SPSS®

Fuente: SPSS®

Los iconos de la barra de herramientas, de izquierda a derecha, significan lo siguiente:

Abrir archivo; Guardar archivo; Imprimir; Recuperar Cuadros de diálogo; Deshacer; Rehacer; Ir a gráfico; Ir a caso; Variables; Buscar; Insertar caso; Insertar variable; Segmentar archivo; Ponderar casos; Seleccionar casos; Etiquetas de valor; Usar conjuntos.

En el editor de datos SPSS®, en la parte inferior de la pantalla hay dos solapas:

- Vista de datos. Esta vista muestra los valores de datos reales o las etiquetas de valor definidas. Es la hoja de cálculo adonde se digitan los datos.

- Vista de variables. Esta vista muestra la información de definición y configuración de las variables, que incluye las etiquetas de la variable definida y de valor, tipo de dato (por ejemplo, cadena, fecha y numérico), los valores perdidos definidos por el analista nivel de medida (nominal, ordinal o de escala) y el nivel de medida (nominal, ordinal o de escala).

Cuadros de diálogo. Estos se utilizan para seleccionar variables y opciones de procedimientos para el análisis de los datos. Cada Cuadro de diálogo de un procedimiento tiene varios componentes básicos:

- Listado de variables de origen contenidas en el archivo de datos de trabajo.
- Listado de variables de destino que indica las variables seleccionadas para el análisis.
- Botones de comandos que dan instrucciones al programa para que ejecute un procedimiento.
- Botones de funciones estadísticas, gráficos y formato.

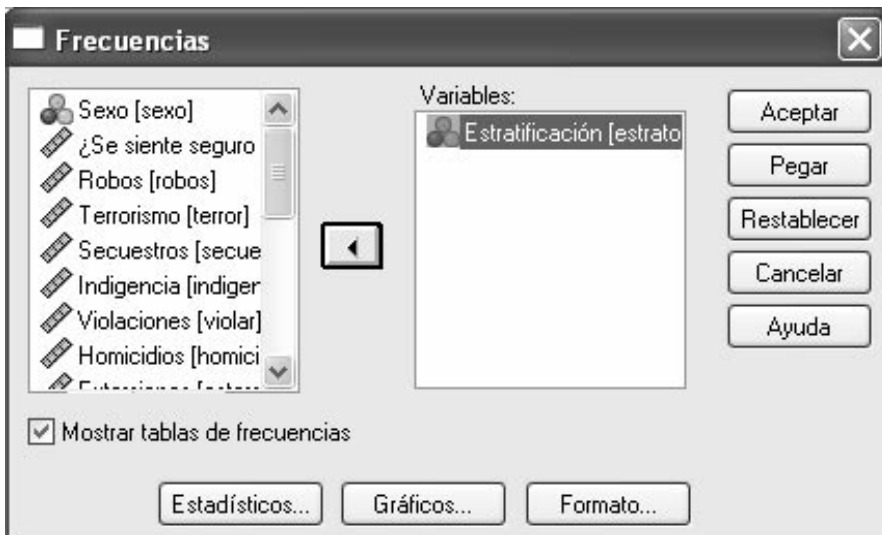


Figura 6.4. Cuadro de diálogo Frecuencias

Fuente: SPSS®

Existen cinco botones de comando en el Cuadro de diálogo que tienen las siguientes funciones:

- *Aceptar*. Ejecuta el procedimiento seleccionado. Después de seleccionar las variables que desee y de elegir cualquier especificación adicional, dé un <clic> en *Aceptar* para ejecutar el procedimiento. Este botón también cierra el Cuadro de diálogo.
- *Pegar*. Genera una sintaxis de comandos a partir de las selecciones en el Cuadro de diálogo y la pega en una ventana de sintaxis.
- *Restablecer*. Anula la selección de cualquier variable de la(s) lista(s) de variables seleccionadas y restablece todas las especificaciones del Cuadro de diálogo principal, al modo predeterminado.
- *Cancelar*. Cancela cualquier cambio en los valores del Cuadro de diálogo desde la última vez que se abrió, y cierra el cuadro de diálogo.
- *Ayuda*. Lleva a la ventana de *Ayuda* del programa.
- *Continuar*. Acepta y ejecuta lo definido en este cuadro.

Configuración de la estructura de la matriz de variables

Cuando se comienza a trabajar con el SPSS® automáticamente se está en el Editor de datos. En la parte inferior de la matriz hay dos solapas: Vista de datos (para digitar los datos codificados) y Vista de variables (para definir o configurar las variables). Por lo tanto, para definir la configuración de las variables, presione en la solapa Vista de variables³. En la Vista de Variables cada línea define una variable y las columnas son los diversos aspectos de la variable que se deben configurar.

El Cuadro de diálogo Vista de variables (Variable View), permite nominar las variables, definir su tipo, definir la etiqueta o descripción de la variable, delimitar el número de dígitos que ocupa el nombre de la variable, precisar su ancho y el número de decimales (especificar si se va a trabajar con datos con decimales o no), detallar los valores o categorías de respuesta de la variable y delimitar los valores perdidos.

3 Se puede cambiar de Vista de datos a Vista de variables o viceversa simplemente tecleando Ctrl. + T.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos
1	Sexo	Numérico	8	2		Ninguno	Ninguno
2	Estrato	Numérico	8	2		Ninguno	Ninguno
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							

Figura 6.5 Editor de datos SPSS®

Fuente: SPSS®

Escriba la palabra Sexo en la columna “Nombre⁴” para nominar la primera variable. Aunque lo escriba en mayúsculas SPSS® lo dejará siempre en minúsculas. Cuando se desplace a la columna “Tipo” observará dos cosas: primero que SPSS automáticamente la define como una variable numérica con una anchura de 8 dígitos de los cuales 2 son decimales. Como la variable Sexo no tiene valores decimales hay que situarse, con el ratón, en la casilla Decimales y digitar el valor 0.

Los datos de una variable numérica sólo pueden ser números, pero los datos de una variable alfanumérica pueden ser letras, números, o letras y números. La anchura de la variable y su número de decimales también se puede configurar. El tipo de variable Cadena se refiere a variables alfanuméricas, como puede ser la variable “Dirección del entrevistado”.

4 Conocida como nombre corto de la variable.

Una variable SPSS® (como se llama a una variable en el programa SPSS) puede tener ocho letras como máximo, debe empezar por una letra, puede contener cualquier letra o número y no debe contener ningún carácter especial como el guión, la coma o el signo de interrogación; sí puede contener acentos, el signo de subrayado, el signo \$ o el signo @. No puede haber dos variables con el mismo nombre.

A continuación, en la columna “Etiqueta” se escribe la descripción de la variable⁵. En esta columna sí se distingue entre mayúsculas y minúsculas. La etiqueta de variable añade una descripción a la variable que se empleará en el visor cuando se use esa variable. Por ejemplo, la variable Sexo tiene dos categorías Femenino y Masculino cuyos valores se registran tal como aparece en el Cuadro de diálogo siguiente.

Para definir los valores de la variable Sexo sitúese en la columna “Valores” y presione dando un <clíc> sobre él.

Aparece el Cuadro de diálogo Etiquetas de valor.



Figura 6.6 Etiquetas de valor

Fuente: SPSS®

5 Las etiquetas de valores se utilizan para describir los valores de variables cualitativas.

En la casilla Valor, teclear el número que representa el código, a continuación moverse con el puntero del ratón hasta la casilla de Etiqueta de valor y escribir la etiqueta, y después presionar con un <clik> el botón <Añadir>. Se procede así para cada uno de los valores que se quieran etiquetar. Finalmente presionar dando un <clik> sobre el botón <Aceptar>. Las etiquetas de valores se emplean para describir los valores de variables cualitativas.

En la casilla Perdidos especificar en el caso que no se logre información o no se quiera incluir un tipo de información, por ejemplo, No sabe o No responde. Se utiliza el procedimiento de asignar un valor que indica la falta de datos o de exclusión.

En la casilla Columnas se controla el ancho de la columna de esa variable.

En la casilla Medición hay que definir con qué tipo de variable se está trabajando. Allí se puede especificar el nivel de medida como Escala (datos numéricos de una escala de intervalo o de razón), Ordinal o Nominal. Los datos nominales y ordinales pueden ser de cadena (alfanuméricos) o numéricos.

Nominal. Una variable puede ser tratada como nominal cuando sus valores representan categorías clasificatorias. Por ejemplo, el sexo biológico de las personas: Femenino o Masculino.

Ordinal. Una variable puede ser tratada como ordinal cuando sus valores representan categorías clasificatorias con alguna ordenación intrínseca. Por ejemplo los niveles de satisfacción con un servicio, que vayan desde muy insatisfecho hasta muy satisfecho.

Escala. Una variable puede ser tratada como de escala cuando sus valores representan categorías ordenadas con una métrica con significado, por lo que son adecuadas las comparaciones de distancia entre valores. Son ejemplos de variables de escala: la edad en años cumplidos y los ingresos familiares en pesos.

Digitación de los datos

Una vez definida la estructura o configuración de la matriz de variables en el Editor de datos, se comienza la digitación de los códigos contenidos en una hoja de codificación o directamente del cuestionario.

La matriz del editor de datos es un rectángulo formado por filas y columnas, lo cual conforma un conjunto de celdas⁶. Cada fila representa un caso o una observación, es así como, cada individuo que responde a un cuestionario es un caso. Cada columna representa una variable o característica que está siendo analizada. La celda es la intersección del caso y la variable. Las celdas contienen sólo valores de datos. Cada una de ellas contiene un solo valor de una variable para un caso. Para variables numéricas, las celdas en blanco se convertirán en valores perdidos por el sistema e incidirán en el cálculo estadístico, por lo cual se recomienda definir el valor que se quiere omitir.

Procedimiento SPSS®

- Para digitar la información disponible dar un <clic> en la solapa *Vista de datos* y comience el procedimiento de digitación, de acuerdo al siguiente procedimiento.
- Digite el valor codificado en la primera celda y así sucesivamente. Cada vez que digite un valor presione sobre la tecla Enter o Intro.
- Una vez digitado todos los datos de la primera columna – variable sexo – se continúa con la siguiente columna – variable *Estrato* – situándose en la casilla del caso 1 y así sucesivamente.

Grabación de los datos digitados

Para grabar los datos digitados se debe situar en el menú Archivos y luego el submenú <Guardar>. Si es un archivo nuevo, escriba el nombre en la celdilla al lado de *Nombre del archivo* y seleccione en <Guardar como> especificando que es un archivo de tipo: SPSS® (con la extensión sav). Para terminar, elija dónde se va a grabar la información – Unidad C; o Unidad A.

Análisis de los datos obtenidos

Analizar datos cuantitativos y cualitativos con el SPSS® es un procedimiento estandarizado y fácil. Una vez configurada la matriz de datos, digitados y depurados los valores en el archivo de trabajo, desde el punto de vista metodológico, se pueden sugerir algunos criterios a seguir:

6 La dimensión del archivo de datos se determina por el número de casos y variables.

Primero, revisar el plan de análisis para asegurar cuál es la estrategia de manejo de datos más adecuada para resolver el problema de investigación.

Segundo, diseñar una tabla descriptiva de las variables en juego: breve descripción de las variables; su nombre en el editor de texto, y la codificación de los valores de esas variables.

Tercero, desarrollar un plan preliminar de análisis de los datos mediante la herramienta *distribución de frecuencias*, que permite la presentación de los resultados y describir las tendencias más generales encontradas. Esto puede ayudar al analista a continuar su análisis combinando otras estrategias.

Cuarto, seleccionar un procedimiento de los menús para transformación y gestión de los datos, calcular estadísticos o crear un gráfico.

Quinto, ejecutar el procedimiento y analizar los resultados. Estos se muestran en el Visor de resultados.

El visor de resultados⁷

Esta gráfica contiene los resultados en forma de texto de la sesión de trabajo con el SPSS®. Esto incluye cualquier resultado estadístico no gráfico, que puede ser, una tabla de contingencia o una matriz de correlaciones. Al realizar un procedimiento de *distribución de frecuencias* con la variable *Tipo* del archivo *hábitos.sav*⁸, en el visor se puede ver que el 51.2% de los estudiantes son de la Facultad de Ciencias Sociales y el 48.8% son estudiantes de la Facultad de Veterinaria y Zootecnia

Tipo de estudiantes según programa académico.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Ciencias Sociales	44	51,2	51,2	51,2
	Medicina Veterinaria y Zootecnia	42	48,8	48,8	100,0
	Total	86	100,0	100,0	

Figura 6.7 Visor de resultados

Fuente: SPSS®

⁷ Output SPSS Viewer

⁸ Ver archivo de trabajo que contiene la información necesaria para el ejercicio en el Anexo. Para desarrollar el ejercicio presentado y los siguientes se debe definir la estructura de la matriz de variables en la ventana Vista de variables y a continuación, digitar los datos registrados en la ventana Vista de datos del Editor de datos.

Análisis descriptivo de los datos

La primera tarea del analista es hacer, - a partir de una información confiable y válida-, una descripción de las características de las unidades de análisis. La segunda tarea, si es el caso, hace referencia a la comparación de subgrupos 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:

- Aspectos que caractericen, en términos cuantitativos o cualitativos, a los colectivos sociales, hechos o procesos, los cuales se han definido en el problema de investigación y en la descripción de variables.
- Clasificaciones, tipologías, diferencias o comparaciones entre las variables.
- Establecer las relaciones entre las variables o características de los sujetos de estudio.

La descripción de las variables puede efectuarse mediante la distribución de frecuencias. Ésta consiste en el ordenamiento o clasificación de los valores observados en una variable, de acuerdo con su magnitud numérica expresada generalmente en términos porcentuales. Permitiendo al investigador no sólo depurar los datos, sino también identificar cómo ciertos puntos o características están distribuidos.

La tabla que representa la distribución de frecuencias 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 los valores de las categorías dentro de la variable?

- Buscar una explicación de la variabilidad de las características - propiedades de la variable.
- ¿Qué factores están asociados a esta variabilidad?

La distribución de frecuencias proporciona también estadísticos y representaciones gráficas que son una herramienta básica para comenzar a analizar los datos. Es uno de los más utilizados del programa SPSS®, dado que no sólo sirve para el cálculo de estadísticos, sino también para visualizar y depurar datos (eliminación de códigos inválidos, modificación de datos mal digitados).

Uno de los productos en el Visor de resultados del submenú Frecuencias, es una *Tabla de frecuencias*, allí 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. Las distribuciones de frecuencias relativas pueden presentarse en forma de histogramas o gráficas de otro tipo, tales como polígonos de frecuencias que relacionan las puntuaciones con sus respectivas frecuencias.

Ejemplo. La inseguridad es un miedo urbano que permeó todas las esferas. El temor y la sospecha nace 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 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 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 el fomento de programas de cultura ciudadana.⁹

La ficha técnica del sondeo de opinión es la siguiente: se entrevistó, vía telefónica, a 322 personas, mayores de dieciséis 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 del 4%.

9 Violencia en Medellín. Periódico, El Colombiano. 3 de noviembre de 2002. Pág. 6

- Procedimiento SPSS®

Para obtener tablas de frecuencias se debe seguir el siguiente procedimiento:

- En la Barra principal de menús del SPSS®, seleccione el menú <Archivo >
- Active el submenú <Abrir> y elija la opción <Datos>.
- Seleccione allí el archivo *temorciudadano.sav*¹⁰
- En la barra principal de menú del SPSS®, seleccione el menú <Analizar>¹¹
- Seleccione el submenú <Estadísticos descriptivos>¹²
- Seleccione la opción <Frecuencias>
- A continuación aparecerá el siguiente Cuadro de diálogo Frecuencias (Véase figura 6.8)
- En el Cuadro de diálogo Frecuencias, seleccione las variables a describir - Robos y secuestros. Transfíeralas a la ventanilla interna Variables.
- Dé un <clik> en el botón <Estadísticos> para obtener, si es el caso, estadísticos para las variables cuantitativas.
- A continuación, dar un <clik> en el botón <Gráficos> para obtener gráficos de barras, de sectores e histogramas.
- El botón <Formato> permite organizar la tabla de frecuencias.
- Proceda a presionar el botón <Aceptar>

Con esta última opción, aparece en el Visor de resultados, una *Tabla de frecuencias*:

10 Ver archivos de trabajo anexos al final del texto. Para desarrollar el ejercicio presentado se debe definir la estructura de la matriz de variables en la ventana Vista de variables y a continuación, digitar los datos registrados en la ventana Vista de datos del Editor de datos.

11 Analyze.

12 Descriptives Statistics.

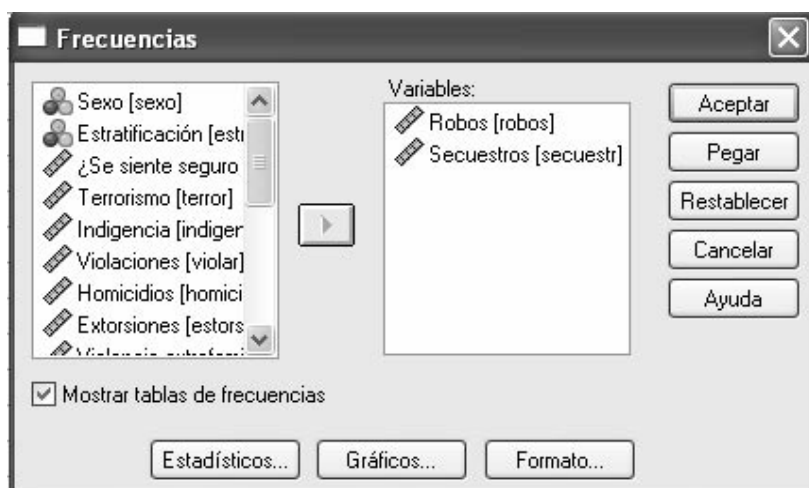


Figura 6.8 Cuadro de diálogo Frecuencias

Fuente: temorc Ciudadano.sav¹³

Tabla de frecuencia

Robos

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Si	19	63,3	63,3	63,3
	No	11	36,7	36,7	100,0
	Total	30	100,0	100,0	

Secuestros

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Si	16	53,3	53,3	53,3
	No	14	46,7	46,7	100,0
	Total	30	100,0	100,0	

Fuente: Violencia en Medellín Periódico, El Colombiano. 3 de noviembre de 2002. Pág. 6

13 Ver archivos de trabajo anexos al final del texto

- Análisis de los datos

La lectura de los datos obtenidos permite afirmar, que de acuerdo con la información del sondeo de opinión sobre la inseguridad ciudadana, se puede apreciar que los robos –en cualquiera de sus manifestaciones– ocupan el primer lugar (63,3%) en las percepciones de los medellinenses, y que el secuestro le sigue en orden de las apreciaciones de inseguridad entre estos habitantes (53,3%).

Análisis estadístico de los datos

La dirección del análisis o la búsqueda de contextualizantes, se hacen dentro de ciertas *orientaciones* que son señaladas generalmente por los objetivos del estudio. En el análisis de los datos, se está en el campo de las técnicas; en cambio, en la interpretación, están en el campo de la conceptualización teórica.

A continuación de la descripción de los datos, se lleva cabo el análisis estadístico de éstos. Cuando se describe un conjunto de variables, se trata de sintetizar sus características en unos índices numéricos que las representen adecuadamente. En este texto, las técnicas que se exponen son las medidas de tendencia central, valores percentiles¹⁴, dispersión y distribución.

Para lograr el análisis estadístico de los datos, el investigador 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. Los procedimientos más utilizados 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: *la moda, la media y 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 usan frecuentemente los conceptos:

14 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 por debajo.

- Rango o amplitud de la variación: Es la diferencia entre la puntuación mayor y la menor. Se supone que cuanto más grande sea el rango, mayor será la dispersión de los datos de una distribución.
 - Valores percentiles: Es un índice de posición de un caso en su grupo o en su percentil, que indica el porcentaje de casos que tienen puntuaciones inferiores o iguales a la suya. Se obtiene con el comando Frecuencias.
 - Desviación estándar o típica: Es el promedio de desviación de las puntuaciones con respecto a la media aritmética. Cuanto mayor es la dispersión de los datos alrededor de la media aritmética mayor es la desviación estándar.
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 medición continua, se utilizan entre otras:
- Coeficientes de asociación entre variables cualitativas: unos de los más utilizados son los Coeficientes de Asociación por Rangos Ordenados de Spearman y Tau de Kendall. Los coeficientes rho de Spearman, simbolizados como r_s y el de Tau de Kendall como t , son medidas de correlación para variables cualitativas.
 - Coeficientes de correlación entre variables cuantitativas: es un valor originado por una prueba estadística para analizar la relación entre dos variables medidas en un nivel por intervalos o de razón. Se simboliza por r minúscula
 - Pruebas de regresión lineal o múltiple.

Selección de la prueba estadística¹⁵

Un criterio básico que hay que tener en cuenta al seleccionar la prueba estadística es *el nivel de medición* de los datos¹⁶. En la tabla 6.1 se definen sus relaciones.

15 Existen dos tipos de pruebas estadísticas: pruebas paramétricas y pruebas no paramétricas. Estas últimas se emplean para comprobar la existencia de diferencias entre grupos que han sido medidos con variables ordinales.

16 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.

Tabla 6.1 Los niveles de medición y las estadísticas apropiadas a cada nivel

Nivel de medición	Relaciones definidas	Pruebas estadísticas apropiadas	Tipos de estadísticas apropiadas
Nominal	Equivalencia	No paramétricas	Moda Distribución de frecuencias Coeficiente de contingencia
Ordinal	Equivalencia de mayor a menor		Mediana Percentiles Rho de Spearman
Intervalo	Equivalencia de mayor a menor	Paramétricas	Media aritmética Desviación estándar
Proporción	Proporción continua		Correlación del Pearson Regresión múltiple Análisis multivariado, etc.

Fuente: Siegel Sydney. Non-parametric statistics for the behavioral sciences, Nueva York: McGraw-Hill Book. Co. 1978. Pág. 51.

Estadísticos 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 que son objeto del estudio. Esta primera información muestra 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 medidas estadísticas dan una descripción global de la población estudiada y permitirán orientar el análisis. Existen tres 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. Por ejemplo, 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 (véase tabla 6.2)

Tabla 6.2 Distribución según grupos de edades

Grupo A. Estudiantes de sexo femenino

Edad (X)	Frecuencias f(x)
9	10
8	20
7	30
6	20
Total	80

La pregunta básica que se hace para 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 siete años, debido a que la mayoría declaró esa edad; luego la *moda* es igual a siete 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 continua o colectiva en dos partes iguales y sirve para medir el punto que está al centro de la distribución.

Un ejemplo es calcular la mediana en la información que representa el promedio de días de estadía de los pacientes en el hospital universitario que se muestra en la tabla 6.3.

Tabla 6.3 Número de días de hospitalización según paciente

Paciente	Nº días	Paciente	Nº días	Paciente	Nº días
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	7	25	150
8	3	17	8	26	
9	3	18	8	27	

La mediana es igual a cuatro: valor que ocupa el lugar trece en el listado. Este valor se puede interpretar así: la mitad de los pacientes tienen una permanencia de menos de cuatro 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 cuatro. Obsérvese 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 cien o más días. La mediana, al contrario, está menos determinada por éstos.

Usar la mediana conjuntamente con la media es un buen recurso metodológico para analizar el comportamiento de un conjunto de datos. Las estadísticas económicas, por ejemplo, generalmente presentan la *mediana de ingresos* de un grupo de personas antes que *la media de los ingresos*, ya que el nivel de estos, pueden 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 bajos o altos ingresos.

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

- *Media*. Llamada también *media aritmética*, en el lenguaje cotidiano, se le conoce como promedio aritmético. Es la medida de localización o tendencia más utilizada. Con ella se calcula el promedio de los datos. Por ejemplo: de acuerdo con las edades declaradas en la Tabla 6.2 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 formula:

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

Estadísticos para analizar la dispersión¹⁷ 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 importante determinar el grado de variabilidad –diseminación o dispersión– de los datos con respecto a esos promedios. Para tales efectos, las medidas de dispersión, tales como el rango, la varianza¹⁸ 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 o amplitud*: 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 sea el rango, mayor será la dispersión de los datos en una distribución.

Ejemplo: se analiza el número de consultas diarias atendidas por el personal médico de los consultorios A y B del Servicio Médico de una Universidad (véase tabla 6.4)

Tabla 6.4 Consultas diarias en los consultorios A y B

A		B	
Médico	Nº. consultas	Médico	Nº. consultas
A	10	F	28
B	20	G	29
C	30	H	30
D	40	I	31
E	50	J	32
		K	
X= 30		X = 30	

En ambos consultorios, el número promedio de consultas es de treinta pacientes por día. Como se aprecia, 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

17 Variabilidad de los valores con respecto a la media aritmética.

18 Variabilidad promedio de un determinado valor de una población.

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 (Un mínimo de 10 consultas y un máximo de 50 consultas). En el consultorio B, el “rango” de consultas es sólo de 4 (Un mínimo de 28 consultas y un máximo de 32).

- *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 segunda característica, en importancia, que se emplea para describir variables. Tiene la desventaja que los valores extremos en el conjunto de datos distorsionan su valor.

Por ejemplo, si la media de edad es 17.033 años, con una desviación estándar de 2,148, mientras que el rango es 31 con un mínimo de 14 años y un máximo de 45 años, se puede concluir que el 95% de los casos se encuentra dentro de la desviación estándar respecto a la media.

Manejo de estadísticos con el programa SPSS®

En los capítulos anteriores se han presentado algunas técnicas de manejo estadístico, ejemplos simplificados y muestras de tamaño menor. Pero en la vida real, los analistas, se enfrentan a problemas mucho más complejos que requieren de grandes poblaciones muestrales y del análisis de múltiples características, conocimientos y opiniones (variables). En la actualidad los programas de computación especializados facilitan el manejo y análisis de la información.

Todas las técnicas de análisis estadístico que se mencionan en éste capítulo se encuentran en el programa SPSS®. Por tal razón, aquí sólo se hace referencia a lo que el autor 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”¹⁹.

Procedimientos generales para la selección de las técnicas estadísticas en el SPSS®

Tipos de análisis estadísticos

19 Briones, Guillermo. Op cit. Pág. 21

- Análisis Univariante: técnicas empleadas para analizar los cambios de una variable en cada observación realizada
- Análisis Bivariante: técnicas empleadas para analizar simultáneamente los cambios de dos variables en cada observación realizada
- Análisis Multivariante: técnicas empleadas para analizar simultáneamente los cambios de diversas variables.

En el SPSS®, las técnicas básicas de análisis estadístico están agrupadas en la Barra principal de los menús, en el menú desplegable <Analizar>. Al dar <clíc> sobre este menú aparecerá desplegadas las opciones de operaciones de estadísticos descriptivos, relacionales e inferenciales (véase figura 6.9).

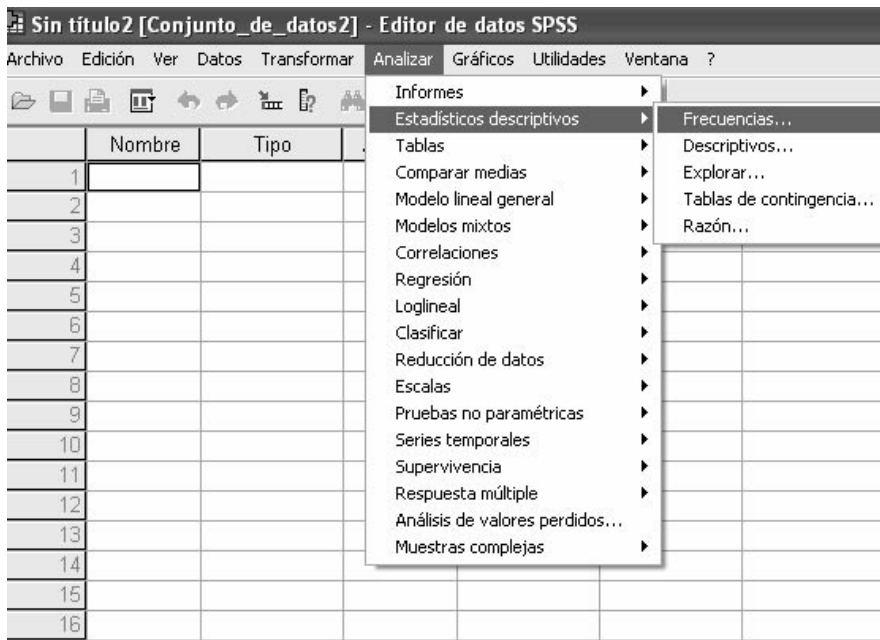


Figura 6.9 Menú desplegable de opciones de operaciones estadísticas

Fuente: SPSS®

Descripción estadística de una variable

El procedimiento SPSS® <Estadísticos descriptivos> muestra estadísticos de resumen univariados.

En esta primera fase, en el submenú <Estadísticos descriptivos>, se le da un <clic> sobre él. A continuación se despliega la ventanilla que aparece en la figura 6.10.

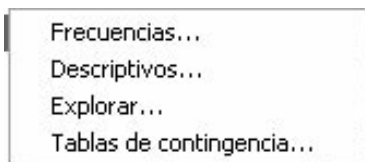


Figura 6.10 Submenú de la opción *Estadísticos descriptivos*

Fuente: SPSS®

La opción <Frecuencias> proporciona estadísticos y representaciones gráficas que son de gran utilidad para explorar y describir el comportamiento de tipos de variables. Es la más utilizada del programa SPSS®, dado que no sólo sirve para el cálculo de estadísticos, 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 con respecto al total de casos.

A continuación, el Cuadro de diálogo *Frecuencias Estadísticas*: presenta las siguientes opciones de manejo de estadísticos:

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 diez. De modo opcional, puede solicitar un número de grupos diferente. Introduzca un entero positivo entre dos y cien. Por ejemplo: si introduce cuatro, se mostrarán los cuartiles.
 - *Percentil*. Valores en los percentiles especificados por el usuario. Introduzca un valor de percentil entre cero y cien 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*²⁰. Diferencia entre los valores mayor (máximo) y el menor (mínimo).
 - *Mínimo*. Muestra el valor más pequeño.
 - *Máximo*: Muestra el valor más grande.
 - *Error típico de la media*. Es 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.
 - *Moda*. El valor que ocurre más frecuentemente. Si diversos valores están empatados en la frecuencia más alta, únicamente aparecerá el valor más pequeño de la variable.
 - *Suma*. La suma de todos los valores.
4. *Formas de 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.

Ejemplo. Perfil de los ingresos de los egresados de cursos de informática.

- Procedimiento SPSS®

- En la barra principal de menú del SPSS®, seleccione el menú < Archivos >
- Active la opción <Abrir>, elija el submenú <Datos>
- Seleccione el archivo *ingresos-sena.sav*
- En la barra de menú, elija el menú < Analizar >, luego <Estadísticos descriptivos>

20 Rango

– *Selecione < Frecuencias>*.²¹ (Véase figura 6.11)

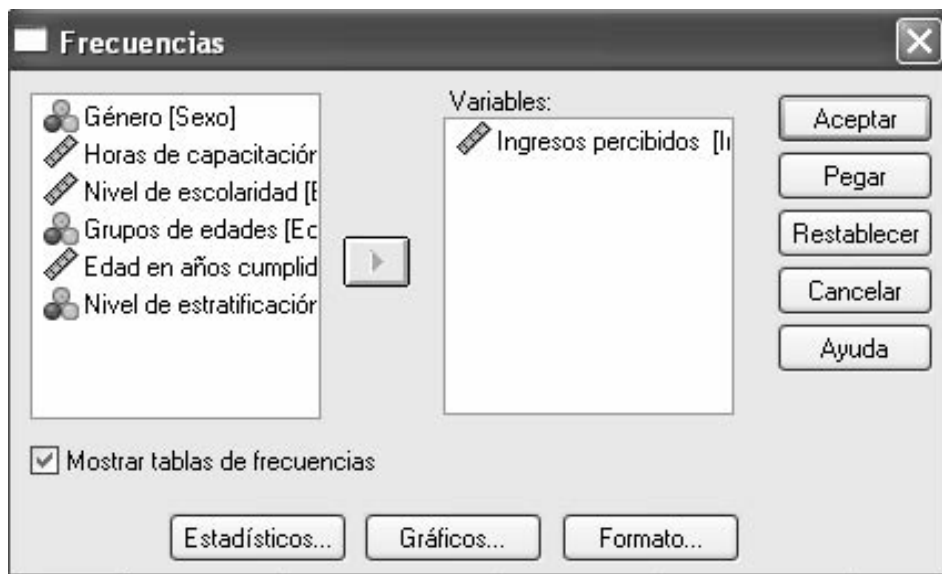


Figura 6.11 Cuadro de diálogo de *Frecuencias*

Fuente: SPSS®

Si se quiere calcular algunos estadísticos, dé un <clíc> en el botón <Estadísticos> y aparece a continuación el Cuadro de diálogo Frecuencias Estadísticos.

Proceda a seleccionar los estadísticos, por ejemplo, Cuartiles, Desviación típica, Mínimo, Máximo, Media.

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

Al dar un <clíc> en el botón <Continuar>, se retorna al Cuadro de diálogo Frecuencias, allí nuevamente se debe dar un <clíc> en el botón <Aceptar>.

A continuación aparece el siguiente Visor de resultados (véase figura 6.13)

21 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.



Figura 6.12 Cuadro de diálogo estadísticos a utilizar en el submenú *Frecuencias*

Fuente: SPSS®

Estadísticos

Ingresos percibidos

N	Válidos	20
	Perdidos	0
Media		828,00
Mediana		930,00
Desv. típ.		403,219
Mínimo		200
Máximo		1300
Percentiles	25	440,00
	50	930,00
	75	1200,00

Figura 6.13 Tabla de resultados estadísticos en el visor

Fuente: SPSS®

- Análisis de los datos

A partir de la información obtenida de los veinte encuestados, se puede concluir que el promedio de ingresos declarados es de \$828.000; igualmente se observa que existe una gran dispersión en la distribución de los ingresos –\$200.000 como mínimo y \$1.300.000 como máximo—. Se puede afirmar, a partir de los datos obtenidos en la entrevista, que el 25% de los encuestados tiene ingresos iguales o inferiores a \$440.000 y que el 50% tiene ingresos iguales o superiores a \$930.000.

Análisis estadístico descriptivo con la herramienta Cubos OLAP

El procedimiento Cubos OLAP (siglas del inglés On-Line Analytic Processing), Procesamiento analítico interactivo calcula totales, medias y otros estadísticos univariantes para variables de resumen continuas²² dentro de las categorías de una o más variables categóricas de agrupación. En la tabla se creará una nueva capa para cada categoría de cada variable de agrupación²³.

Ejemplo: Calcular el promedio, el mínimo, el máximo y la desviación estándar de la edad de los entrevistados según: *Sexo, Estrato Socioeconómico y Municipio de residencia*.

Las variables de resumen deben ser cuantitativas (variables continuas medidas en una escala de intervalo o de razón) y las variables de agrupación deben ser categóricas. Los valores de las variables categóricas pueden ser numéricos.

Procedimiento SPSS® para obtener cubos OLAP

- En la barra principal de menú del SPSS®, seleccione el menú < Archivos >
- Active la opción <Abrir>, elija el submenú <Datos> < archivo TV_FINAL.sav²⁴
- En la Barra de menús, seleccione el menú < Analizar >
- A continuación elija el submenú <Informes>²⁵

22 Variables cuantitativas

23 Las variables de agrupación deben ser categóricas – cualitativas–.

24 Base de datos del estudio “Prácticas culturales y televisivas en el Área Metropolitana del Valle del Aburrá”. Jaime Ruiz R. Miguel Aigner A. Efrén Barrera R. Centro de Estudios de Opinión de la Universidad de Antioquia –CEO- y Comisión Nacional de Televisión, Teleantioquia, TeleMedellín y Canal U. Medellín, 2004.

25 Reports

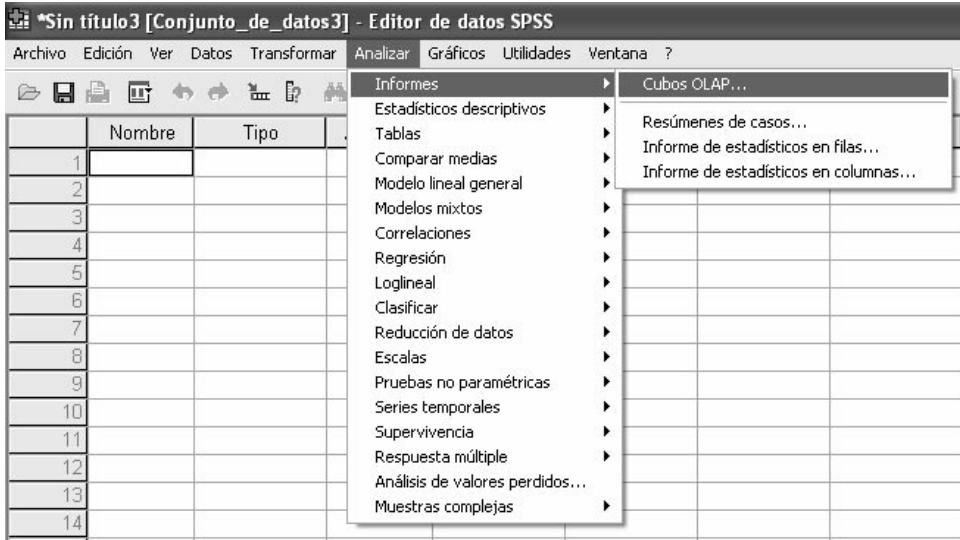


Figura 6.14 Editor de datos SPSS®

Fuente SPSS®

- Dé un <clik> en la opción <Cubos OLAP>.
- Aparece el Cuadro de diálogo Cubos OLAP



Figura 6.15 Cuadro de diálogo Cubos OLAP

Fuente SPSS®

- Seleccione una o más variables de resumen continuas²⁶. en la lista fuente de variables.
- Seleccione la variable *Edad* y trasládelas a la ventanilla Variables de resumen:²⁷
- Seleccione una o más variables de agrupación categóricas. Elija las variables *Sexo*, *Municipio*, *Estrato Socioeconómico* y trasládelas a la ventanilla Variables de Agrupación:²⁸
- Dé un <clik> en <Cubos OLAP: Estadísticos>

Seleccione los siguientes estadísticos de resumen: Media, Mínimo, Máximo, Desviación típica.

Se puede cambiar el orden de aparición de los estadísticos de subgrupo. El orden en el que aparecen en la lista Estadísticos de casilla es el mismo orden que presentarán en los resultados. También se muestran estadísticos de resumen para cada variable a través de todas las categorías.

Debe seleccionar una o más variables de agrupación para poder seleccionar estadísticos de resumen.

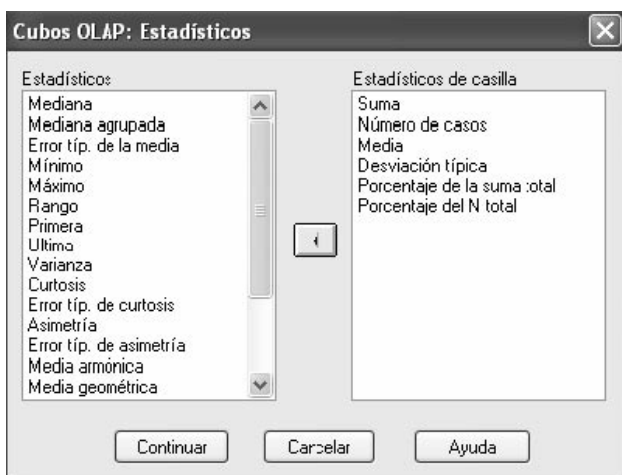


Figura 6.16 Cuadro de diálogo Cubos OLAP: Estadísticos

Fuente SPSS®

26 Continuas, escalas o cuantitativas.

27 Summary Variables (s).

28 Grouping Variables (s).

En el Visor de resultados aparecerán las siguientes tablas:

Resumen del procesamiento de los casos

	Casos					
	Incluidos		Excluidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Edad * Sexo * Municipio * Estrato socioeconómico	2341	100,0%	0	,0%	2341	100,0%

Cubos OLAP

Sexo: Total

Municipio: Total

Estrato socioeconómico: Total

	Suma	N	Media	Desv. típ.	% de la suma total	% del total de N
Edad	72925	2341	31,15	16,923	100,0%	100,0%

Al darle doble <clíc> sobre <Cubos OLAP> aparecerá lo siguiente:

Cubos OLAP

Sexo	Total						▼
Municipio	Total						▼
Estrato socioeconómico	Total						▼
	Suma	N	Media	Desv. típ.	% de la suma total	% del total de N	
Edad	72925	2341	31,15	16,923	100,0%	100,0%	

En el comando Total Sexo, dé un <clíc> sobre el botón <▼>

A continuación se abrirá una casilla con las categorías de la variable Sexo:

Hombre - Mujer.

De acuerdo a los objetivos de su análisis seleccione una categoría:

Cubos OLAP

Sexo	Total		▼			
Municipio	Hombre					
Estrato socioeconomico	Mujer					
	Total					
	Suma	N	Media	Desv. típ.	% de la suma total	% del total de N
Edad	72925	2341	31,15	16,923	100,0%	100,0%

En el comando Total Municipio, dé un <clíc> sobre el botón <▼>

A continuación se abrirá una casilla con las categorías de la variable Municipio: Todos los municipios del Área Metropolitana. De acuerdo a los objetivos de su análisis seleccione un municipio.

Cubos OLAP

Sexo	Total		▼			
Municipio	Total		▼			
Estrato socioeconomico	Caldas		▲			
	Copacabana					
	Envigado		Desv. típ.			
	Girardota		% de la suma total			
	Itagui		% del total de N			
	Medellín					
	La Estrella					
	Total					
	Suma	N	Media	Desv. típ.	% de la suma total	% del total de N
Edad	72925	2341	31,15	16,923	100,0%	100,0%

En el comando Total Estrato Socioeconómico, dé un <clíc> sobre el botón <▼>

A continuación se abrirá una casilla con las categorías de la variable Estrato Socioeconómico: Las seis categorías de estrato socioeconómico. De acuerdo a los objetivos de su análisis seleccione uno de los estrato socioeconómico, por ejemplo, si se quiere obtener información sobre el estrato cuatro, selecciónelo marcándolo.

Cubos OLAP

Sexo		Total	▼			
Municipio		Total	▼			
Estrato socioeconómico		Total	▼			
	Suma	N	1	esv. típ.	% de la suma total	% del total de N
Edad	72925	2	2	16,923	100,0%	100,0%
			3			
			4			
			5			
			6			
			Total			

Los tres Cuadros de diálogos anteriores le permitirán calcular los estadísticos señalados entre la variable resumen *Edad* y los grupos definidos por las variables de agrupación *Sexo*, *Municipio* y *Estrato Socioeconómico*.

Técnicas de representaciones gráficas de las distribuciones de frecuencias

Las gráficas o figuras constituyen uno de los medios más usados y más difundidos para la presentación y análisis global de la información tanto cuantitativa como cualitativa, debido a que las ideas presentadas gráficamente son entendidas con mayor facilidad que las explicaciones textuales o numéricas.

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

- *Histogramas de frecuencias*. Son gráficos que presentan la información de cada una de las categorías de la variable en forma de rectángulos proporcionales. Se utilizan generalmente para representar datos de una variable discreta.
- *Polígonos de frecuencias*. Son gráficas en forma de una serie de líneas rectas conectadas entre sí y que unen unos puntos medios o de intervalo a lo

largo del eje horizontal. Se utilizan para representar distribuciones de frecuencias propias de un nivel de medición por intervalos.

- *Gráficas 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.
- *Gráficas de círculos o de pastel*. En este tipo de gráficas, el área de cada una de los sectores refleja la importancia de la categoría que representa. Se trata de una gráfica en dos dimensiones. El gráfico de pastel consiste en un círculo el cual se divide en tantos sectores como categorías se tienen, de manera que el área que corresponde a cada categoría sea proporcional a su importancia relativa o porcentual.

El uso de figuras o gráficas tiene ciertas limitaciones que es conveniente tener en cuenta:

- No pueden incluir tanta información como las tablas. Sólo se puede presentar a la vez una cantidad limitada de datos, ya que demasiada información o variedad atenta contra su utilidad, dificultando su lectura.
- En las figuras o gráficas se representan valores aproximados mientras, que las tablas permiten incluir la información con exactitud.
- La presentación tabular y gráfica es complementaria al texto, pero no lo reemplaza. Sirve para ejemplificar, para ilustrar el texto, 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®

Para la presentación de gráficas el procedimiento es el siguiente:

- En la Barra de menús del SPSS®, seleccione el menú <Gráficos>
- Si se desea hacer una barra, se elije el menú <Barras> y luego <Gráficos de barras> y allí se selecciona el tipo de grafico, que con el programa SPSS® pueden ser de nueve tipos.

También se pueden obtener gráficas con algunas opciones del menú <Analizar>, seleccionando <Estadísticos descriptivos>, luego <Frecuencias> y allí el botón <Gráficos>.

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 social exige estudiar, por ejemplo, cómo un proceso está asociado a otros o cómo varía en función de otros. Cada día, los analistas e interventores sociales toman decisiones que están basadas sobre predicciones de eventos futuros. Para hacer estos pronósticos, estos 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.

En la sección anterior, se expusieron diversos procedimientos que permiten la descripción de colectivos. En el actual apartado y el siguiente, se presentan diversas técnicas en el que se analizan determinaciones y relaciones entre variables.

A título de ejemplo, se pretenden estudiar, las diferencias de opinión que tienen personas de distinto sexo acerca de cuáles serían las soluciones al problema de la inseguridad ciudadana. A continuación se examina, 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 surge entonces, en el momento en que queramos confirmar nuestra afirmación: *¿cómo comprobar la relación entre las dos variables?* Es decir, ¿cómo una variable explica, determina o 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*.

Análisis de datos tabulados

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 involucra-

das, 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 (relación original). 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.

Sin embargo, puede surgir la pregunta, ¿porque 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 y verticales – y es el analista quién debe decidir, a partir de los objetivos del estudio, cuál es la dirección correcta para realizar el cálculo y la lectura de los porcentajes en la tabla.

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*.

La lectura o análisis 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 ser del 100% en la base de la variable independiente.

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 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*, no tiene sentido considerar como variable independiente a *las alternativas de solución* y parecería sin sentido pensar que esta variable pueda influir al *género*. Sin duda, aquí la variable independiente es el *género*, del cual se ha hipotetizado que afecta o determina la variable dependiente (*alternativas de solución*).

Si el sentido común no 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?

Tablas de contingencia y medidas de asociación

Una de las técnicas más frecuentemente utilizadas en la investigación social es la clasificación de datos en tablas, según uno o más criterios (variables).

Las tablas de contingencia constituyen una herramienta básica para el estudio de la relación existente entre dos variables categóricas.

Un ejemplo de tabulación cruzada cómo se clasifican los datos de dos variables: *género y preferencia por un tipo de solución al problema de la inseguridad ciudadana*.

Tabla 6.5 Alternativas desde la perspectiva de género para disminuir la delincuencia en la ciudad de Medellín

	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
Total	100	100	100

* No sabe, no responde

Procedimiento SPSS®

Muchos procedimientos de SPSS® generan resultados en forma de tablas. La opción de Tablas de SPSS®, ofrece funciones especiales diseñadas para generación de informes personalizados. Muchas de las funciones personalizadas son especialmente útiles para el análisis de datos cuantificados.

El objetivo de esta sección es familiarizarse con la estructura básica de una tabla en SPSS® y los términos que se utilizan para describir los diferentes elementos estructurales que se pueden utilizar en una tabla.

Básicamente, se pueden generar dos tipos:

- Tablas pivote o tablas de contingencia.
- Tablas personalizadas.

El procedimiento Tablas de contingencia crea tablas de clasificación doble y múltiple y además, proporciona una serie de pruebas y medidas de asociación para las tablas.

En el SPSS®, el submenú <Tablas de contingencia>²⁹ 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. Por supuesto, los estadísticos que se calculan tienen validez sólo para las variables adecuadas. Un 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 continua.

Todos los estadísticos que se calculan en una Tablas de contingencia,³⁰ son *medidas de asociación* o, en su defecto, *de independencia*; pero la negación de una lleva la afirmación de la otra; si dos variables no están asociadas, son independientes, o si dos variables no son independientes, están asociadas.

Un ejemplo es sistematizar y analizar la información del estudio sobre la apreciación de los colombianos sobre la crisis política que vive el país.

- En la Barra de menús del SPSS®, seleccione el menú <Archivos>
- Active la opción <Abrir> y elija el submenú <Datos>.
- Seleccionar archivo *situación-2000.sav*³¹
- Nuevamente en la Barra de menús, elija el menú <Analizar> y allí seleccione el submenú <Estadísticos descriptivos>.

29 Se emplea para hallar simultáneamente distribuciones de frecuencias de dos variables y la fuerza de la asociación entre variables.

30 Tabulación cruzada (Crosstabs).

31 Ver archivos de trabajo anexos al final del texto.

– Seleccione <Tablas de contingencia> y aparece la siguiente figura:

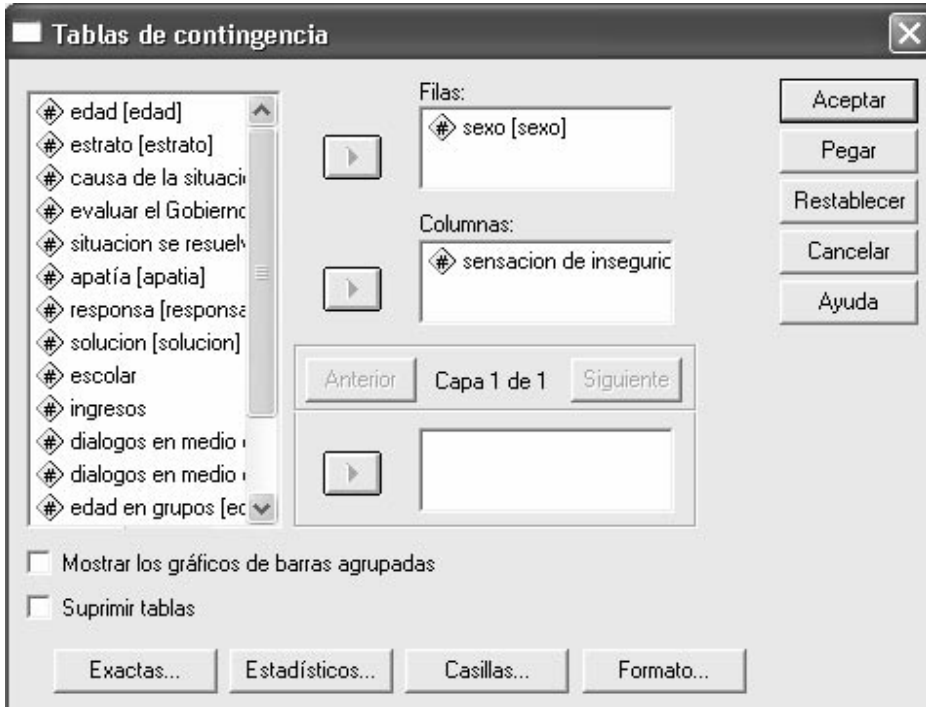


Figura 6.17 Cuadro de diálogo *Tablas de contingencia*

Fuente: SPSS®

- Seleccione una variable independiente de la base de datos *situación-2000.sav*, por ejemplo *sexo* y “crúcela” con la variable *sensación de inseguridad*.
- Proceda a cruzar las variables con el siguiente procedimiento:
- Transfiera la variable independiente *sexo* a la ventanilla *Filas (rows)*:
- Transfiera la variable dependiente *sensación de inseguridad* a la ventanilla *Columnas (Columns)*.
- En el botón *Casillas (Cells)* defina, marcando con \surd si el cálculo de los datos se va a hacer por filas, por columnas o sobre el total de casos.

A continuación, en el Visor de resultados aparecerá la tabla 6.6:

Tabla 6.6 Tabla de contingencia sexo * sensación de inseguridad

			Sensación de inseguridad		Total
			Sentirse inseguro	Sentirse seguro	
sexo	Femenino	Recuento % de sexo	11 91,7	1 8,3	12 100,0
	Masculino	Recuento % de sexo	4 30,8	9 69,2	13 100,0
Total		Recuento % de sexo	15 60,0	10 40,0	25 100,0

- Análisis de los datos

Del total de personas entrevistadas de sexo femenino, una mayoría manifiesta sentirse insegura (91,7%); las personas de sexo masculino, en cambio, manifiestan en menor medida dicha temor: sólo un 30,8% de los entrevistados se siente inseguro.

Controles estadísticos 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*³² que pueden ser consideradas causas de los cambios de la variable dependiente. El concepto de *variable interviniente* o *de control* presupone que la relación entre dos variables está mediada por otra (u otras) que transportan los posibles efectos de la primera.

Por ejemplo, se procede determinar si un programa de alcoholismo rehabilita alcohólicos. Para esto se debe estar seguro que la rehabilitación es debida al programa y no a otros factores intervinientes o asociados.

32 Escalante. A. Carlos, La medición de las actitudes. Colección Ciencias Sociales. Teorías y Métodos. Bogotá: Tercer Mundo. 1983. Pág. 83

En los diseños cuantitativos, se dice que existe una conexión o relación 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 - práctico que permite asegurar que una de las variables es la determinante de la otra.

Gracias al empleo de controles estadísticos es posible un acercamiento 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 de que la relación encontrada entre la variable independiente y dependiente sea el resultado de otra variable asociada a ellas.

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, se sabrá que la relación entre las variables originales no es tan determinante y lo que explica la situación es la relación secundaria.

La tabla 6.7 hace parte de un trabajo de ejemplificación.

Tabla 6.7 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
No sabe / N. R.	11	8	9
	100%	100%	100%

- Análisis de los datos

En la tabla 6.7 se ilustra que un 38% de las personas clasificadas como hombres se inclinan por la solución del aumento de la dotación policial; en cambio, sólo un 17% de las mujeres mostró igual preferencia por esta solución.

Si se analizan las opiniones de los encuestados, intentando enriquecer el análisis, y se introduce una nueva variable interviniente: *nivel de escolaridad*, se observa que las apreciaciones por el tipo de solución de aumento de la dotación policial varían sustancialmente.

Véase que muestra la tabla 6.8.

Tabla 6.8 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

La lectura comparada de las tablas 6.7 y 6.8 muestra el efecto de la variable interviniente — nivel de escolaridad —. Limitemos ahora el análisis a la primera categoría de la variable *soluciones a la violencia ciudadana*: una primera apreciación señala que el nivel de escolaridad 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 del variable *nivel de escolaridad*.

- Análisis de los datos

Específicamente, 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, incluir una variable 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 que está relacionada tanto con la variable independiente como con la variable dependiente.

Vease otro ejemplo del papel de una variable interviniente, en este caso, se a la relación original la variable “opinión sobre los diálogos de paz en medio de la guerra según sexo de los declarantes, una tercera variable: Edad de los entrevistados (Edad2003).

- Procedimiento SPSS®

- En la barra principal de menús del SPSS®, seleccione el menú <Archivos>
- Active el submenú <Abrir> y elija la opción <Datos>
- Seleccione el archivo *situación-2000.sav*
- En la barra de menús, elija el menú <Analizar> y luego <Estadísticos descriptivos> Seleccione <Tablas de contingencia> (véase figura 6.18)

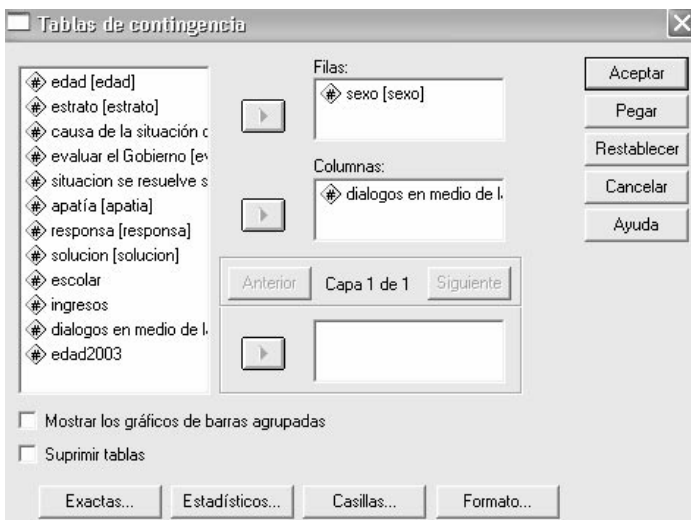


Figura 6.18 Cuadro de diálogo *Tabla de contingencia*

Fuente: SPSS®

- En el Cuadro de diálogo *Tablas de contingencia* seleccione la variable *sexo* y transfírela a la ventanilla *Filas* y la variable *diálogos en medio de la guerra* a la ventanilla *Columnas* y presione el botón <Aceptar>.

En el Visor de resultados, se pueden observar los siguientes datos:

Tabla 6.9 Tabla de contingencia sexo * diálogos en medio de la guerra

		Diálogos en medio de la guerra					total	
		Muy en desacuerdo	desacuerdo	No sabe	acuerdo	Muy de acuerdo		
sexo	femenino	Recuento	4	4	2	1	1	12
		% de sexo	33,3	33,3	16,7	8,3	8,3	100,0
	masculino	Recuento	1	3	3	2	4	13
		% de sexo	7,7	23,1	23,1	15,4	30,8	100,0
Total		Recuento	5	7	5	3	5	25
		% de sexo	20,0	28,0	20,0	12,0	20,0	100,0

Fuente: SPSS®

- Análisis de los datos

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% de las mujeres entrevistadas muestran un mayor desacuerdo con esta posibilidad.

A la relación original planteada – *niveles de favorabilidad, según el sexo, de los diálogos en medio de la guerra* –, 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 variable interviniente, se puede llegar a afirmar que se ha eliminado una posible causa alternativa en la relación original observada. Pero si 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.

Ahora lo que se pretende es estudiar el impacto de una tercera variable considerada aquí como interviniente o de control.

- Procedimiento SPSS®

- En el Cuadro de diálogo, Tablas de contingencia seleccione la variable recodificada *edad2003*³³ y transfírala a la tercera ventanilla –debajo de “Capa 1 de 1” –.
- A continuación dar un <clik> en el botón <Aceptar> (véase figura 6.19)

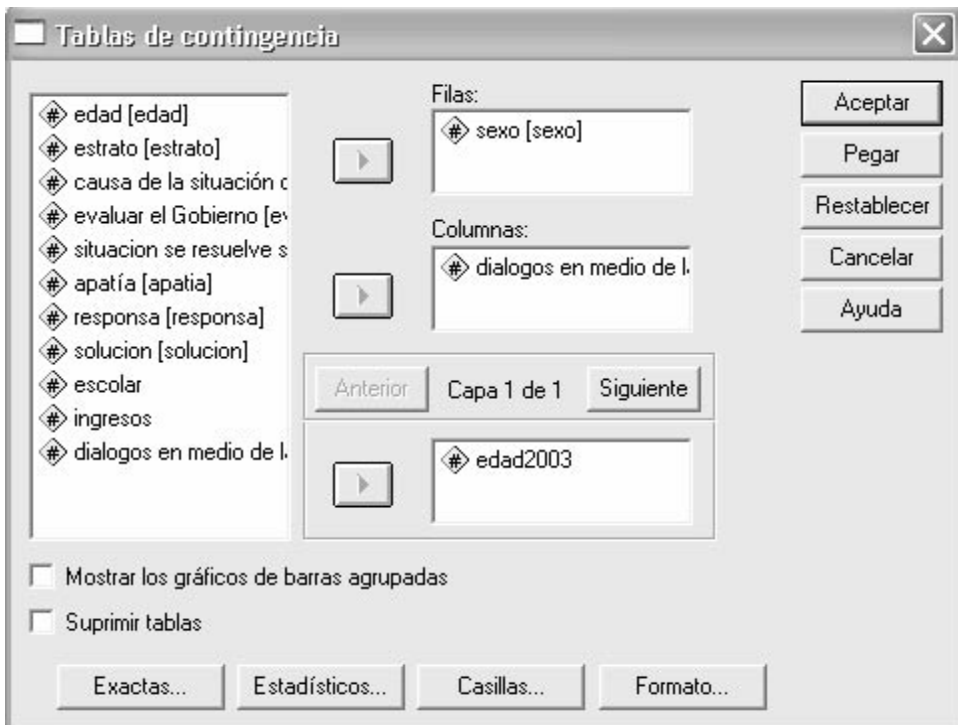


Figura 6.19 Cuadro de diálogo *Tabla de contingencia* con dos variables primarias y una tercera interviniente (*edad2003*)

Fuente: SPSS®

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

33 Categorías y códigos de la variable *edad2003*: jóvenes = 1 y adultos = 2

Tabla 6.10 Tabla de contingencia sexo * Diálogos en medio de la guerra * edad2003

EDAD2003			Diálogos en medio de la guerra				Total	
			muy en desacuerdo	desacuerdo	no sabe	acuerdo		muy de acuerdo
1	sexo femenino	Recuento	3	3		1	7	
		% de sexo	42.9%	42.9%		14.3%	100.0%	
	masculino	Recuento		3	2	1	2	8
		% de sexo		37.5%	25.0%	12.5%	25.0%	100.0%
	Total	Recuento	3	6	2	2	2	15
		% de sexo	20.0%	40.0%	13.3%	13.3%	13.3%	100.0%
2	sexo femenino	Recuento	1	1	2		1	5
		% de sexo	20.0%	20.0%	40.0%		20.0%	100.0%
	masculino	Recuento	1		1	1	2	5
		% de sexo	20.0%		20.0%	20.0%	40.0%	100.0%
	Total	Recuento	2	1	1	1	3	10
		% de sexo	20.0%	10.0%	10.0%	10.0%	30.0%	100.0%

Fuente: SPSS®

- Análisis de los datos

Se puede apreciar que en los jóvenes³⁴ (código1), la relación original persiste; tanto los hombres como las mujeres mantienen sus opiniones. En cambio, los adultos (código 2) manifiestan opiniones diversas: 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.

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.

34 Código 1= personas de treinta y menos años. Código 2 = personas de más de treinta años.

- *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.

Otro procedimiento SPSS® muy útil para el análisis de datos tabulados son las tablas personalizadas que permiten crear tablas de clasificación dobles y múltiples.

Para diseñar una tabla personalizada el programa SPSS® utiliza una interfaz para el generador de tablas, que usa el sistema de arrastrar y colocar, y que permite una representación previa de la tabla, la cual cambia a medida que se seleccionan variables y opciones.

- Procedimiento SPSS®

Para acceder al generador de tablas personalizadas:

- En la Barra de menús seleccionar <Analizar>
- Seleccionar el submenú <Tablas>
- Seleccionar la opción <Tablas personalizadas>³⁵

A continuación aparecerá el generador de tablas personalizadas.

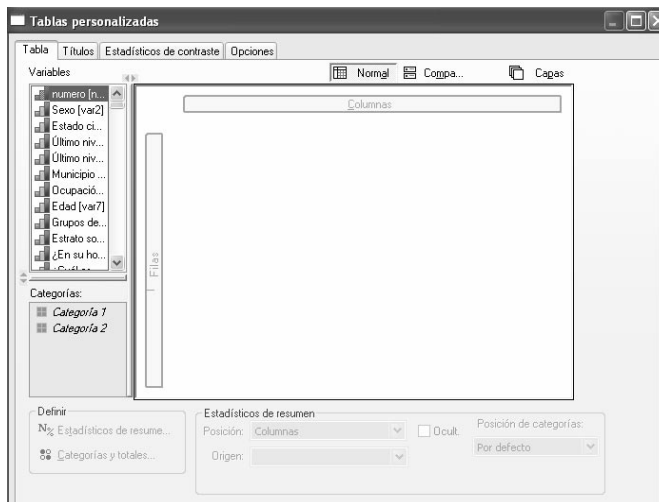


Figura 6.20 Panel generador de tablas personalizadas

Fuente: SPSS®

La figura previa no muestra los valores reales que aparecerían en la tabla, sino sólo el diseño básico de la tabla.

Selección de variables de fila y de columna

Para crear una tabla, sólo tiene que arrastrar y colocar las variables en el sitio donde desea que aparezcan en la tabla.

- Seleccione (dé un <clik> en...) la variable Grupos de Edades en la lista de variables categóricas³⁶ y arrástrela situándola en el área *Filas* del panel interno de generación de tablas.
- Seleccione *Sexo* en la lista de variables y arrástrela sobre el área *Columnas* del panel interno de generación de tablas.

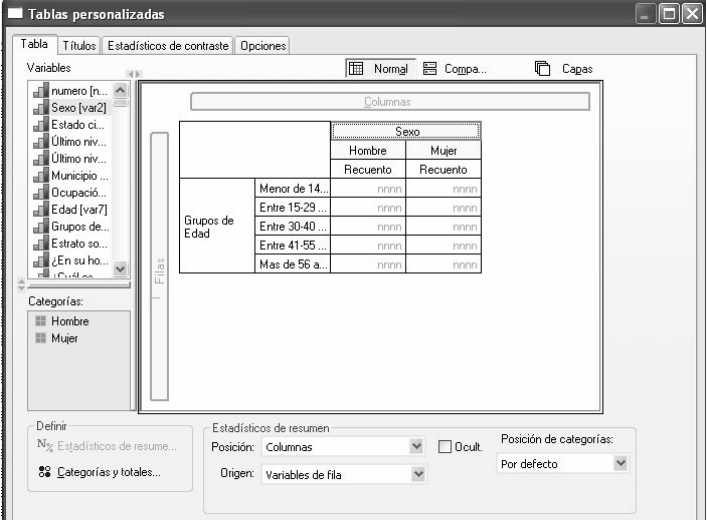


Figura 6.21 Selección de una variable de Columna y de Fila

Fuente: SPSS®

El panel interno de generación de tablas muestra ahora una tabla de contingencia de doble clasificación de Grupos de edades por *Sexo*. Por defecto, los recuentos se muestran en las casillas de las variables categóricas.

36 Nominales - ordinales.

También puede ver los porcentajes de fila, columna y totales. Con el siguiente procedimiento:

- Dar un <clik> con el botón derecho del ratón en la ventanilla Grupos de edades en el panel interno de generación de tablas. A continuación aparecerá un Cuadro de diálogo emergente:

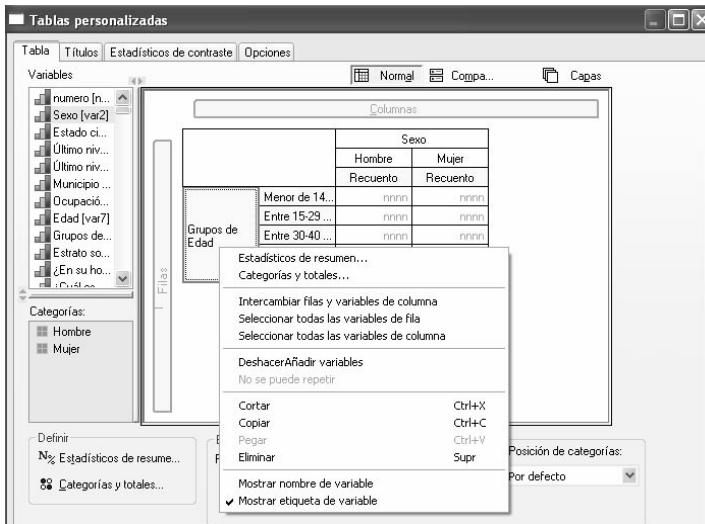


Figura 6.22 Selección en un menú desplegable

Fuente: SPSS®

- Seleccionar *Estadísticos de resumen*³⁷ en el Cuadro de diálogo emergente. A continuación aparecerá el Cuadro de diálogo <Estadísticos de resumen>
- En la ventanilla izquierda superior seleccione la opción <% del N de la fila>³⁸ en la lista Estadísticos
- Dar un <clik> en el botón de flecha para añadirlo a la ventanilla Visualización.

Ahora aparecerán tanto los recuentos como los porcentajes de fila en la tabla.

37 Summary Statistics.

38 % Row.

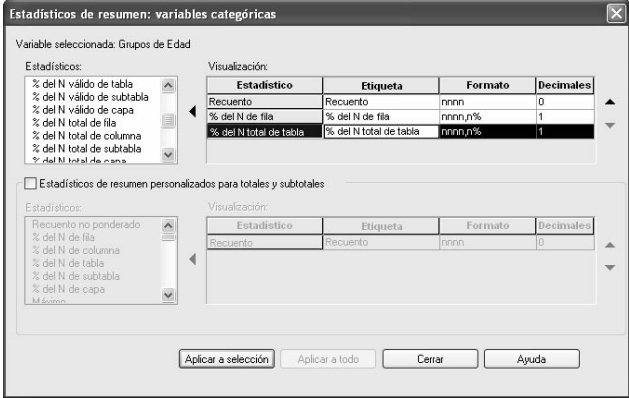


Figura 6.23 Cuadro de diálogo Estadísticos de resumen variables categóricas

Fuente: SPSS®

- Dé un <clik> en el botón <Aplicar a selección> (Apply to Selection) para guardar estas opciones y volver al generador de tablas.

El panel interno de generación de tablas muestra los cambios realizados, mostrando las columnas correspondientes a ambos recuentos y porcentajes de fila.

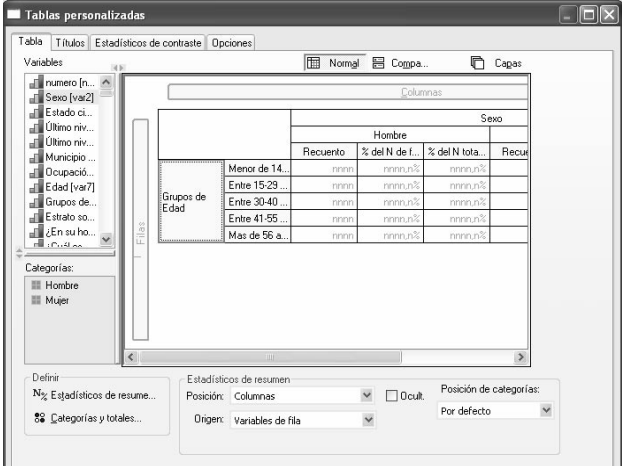


Figura 6.24 Recuentos y porcentajes de fila mostrados en panel interno de generación de tablas Fuente: SPSS®

Fuente: SPSS®

Inserción de totales y subtotales

Para añadir totales y subtotales a una tabla, dé un <clik> con el botón derecho del ratón en Grupos de edades en el panel interno y seleccione <Categorías y totales> en el menú emergente.

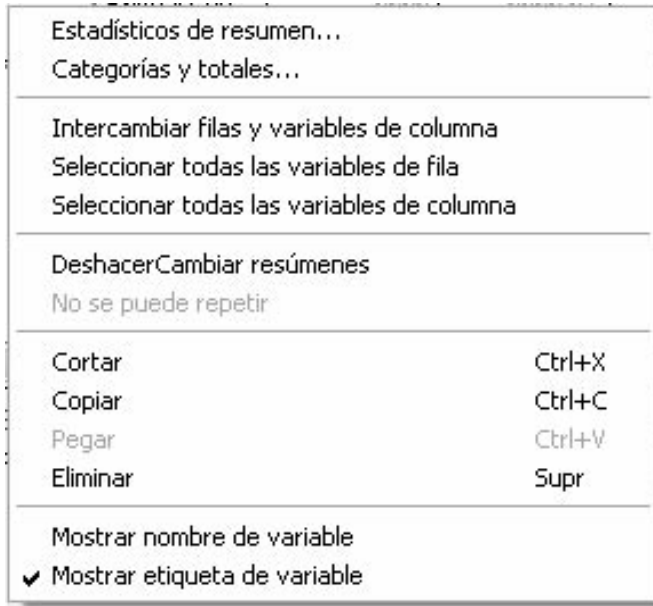


Figura 6.25 Cuadro de diálogo: menú emergente

Fuente SPSS®

En el Cuadro de diálogo Categorías y totales (Categories and Totals), por ejemplo, para que la Tabla presente un subtotal de los Menores de 30 años, seleccione (dé un <clik>) en la categoría Menor de 29 años en la lista de valores.

En la sección Nuevos Valores (Subtotal), en el recuadro interno <Etiquetas> (Labels), nomine el valor escogido, por ejemplo Menores de 30 años. A continuación, dé un <clik> en el botón <Insertar>.

Al hacer esto, se inserta una fila "Menores de 30 años" con el subtotal de las dos primeras categorías de edad.

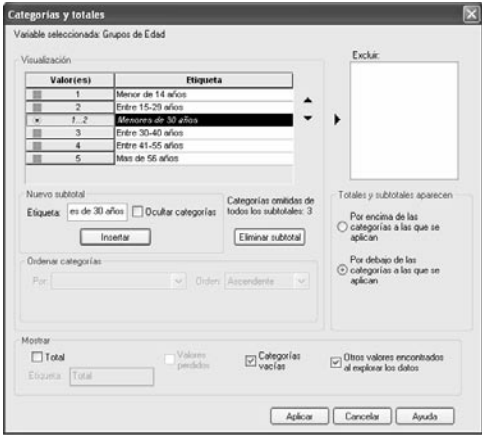


Figura 6.26 Categorías y totales

Fuente: SPSS®

Si es necesario que en la Tabla figuren los totales de los valores siga el siguiente el procedimiento SPSS®:

- En la casilla **Mostrar**, para incluir un total global, seleccione (✓) la casilla <Total del grupo >
- A continuación, dé un <clik> en <Aplicar>.

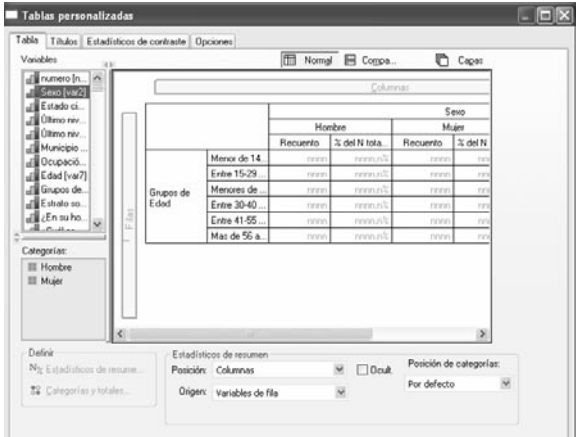


Figura 6.27 Total y subtotales en panel interno de generación de tablas

Fuente: SPSS®

La vista previa del panel interno de generación de tablas ahora incluye las filas correspondientes a los dos subtotales y al total global.

- Dé un <clik> en <Aceptar> para generar la tabla definitiva.

En el Visor de resultados se observará una Tabla con el siguiente formato y resultados

Tabla 6.11 Distribución por grupos de edades según sexo

		Sexo			
		Hombre		Mujer	
		Recuento	% del N total de tabla	Recuento	% del N total de tabla
Grupos de edad	Menor de 14 años	227	9,7 %	220	9,4 %
	Entre 15-29 años	380	16,2 %	371	15,8 %
	Menores de 30 años	607	25,9 %	591	25,2 %
	Entre 30-40 años	223	9,5 %	246	10,5 %
	Entre 41-55 años	238	10,2 %	223	9,5 %
	Más de 56 años	120	5,1 %	93	4,0 %

Fuente: SPSS®

Resumen de variables de escala

Una sencilla tabla de contingencia de dos variables categóricas muestra los recuentos o los porcentajes en las casillas de la tabla, pero también es posible mostrar resúmenes estadísticos de una variable cuantitativa³⁹ (Nº de horas semanales vistas de TV.) en las casillas de la tabla.

- Procedimiento SPSS®

Para volver a acceder al generador de tablas personalizadas siga el siguiente procedimiento:

- Elija en la Barra de menús el menú <Analizar>
- A continuación seleccione el submenú (dé un <clik> en) <Tablas>
- Seleccione la opción <Tablas personalizadas>

39 Variable de escala (Scale).

- Dé un <clik> en <Restablecer> (Reset) para borrar cualquier selección anterior.
- Seleccione la variable Grupos de edades en la ventanilla Variables, arrástrela y colóquela en el área Filas del panel interno de generación de tablas.
- Seleccione *Horas semanales viendo TV* en la lista de variables, arrástrela y colóquela a la derecha de *Grupos de edades* en la dimensión de fila de la tabla.

Aparecerá el siguiente Cuadro de diálogo:

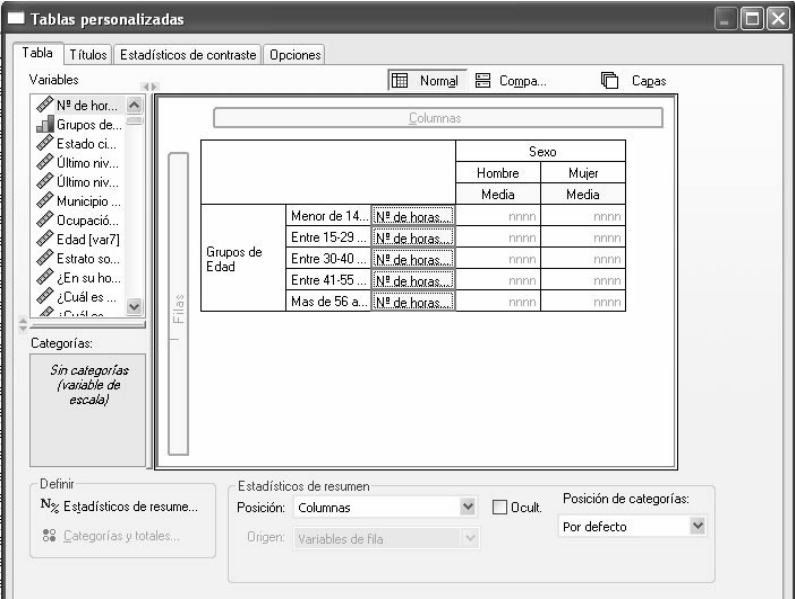


Figura 6.28 Arrastre y colocación de una variable cuantitativa⁴⁰ en la barra de fila

Fuente: SPSS®

Ahora, en lugar de recuentos de categorías, la tabla mostrará la media (promedio) del número de horas de televisión vistas correspondiente a cada categoría de edad.

40 Variable de escala.

Para añadir o cambiar los estadísticos de resumen que aparecen en la tabla siga el siguiente procedimiento:

- Dé un <clik> con el botón derecho del ratón en la variable de escala del panel de generación de tablas
- Seleccione <Estadísticos de resumen> en el menú emergente.

Promedio del número de horas semanales de televisión vistas correspondiente a cada categoría de edad.

			Sexo	
			Hombre	Mujer
			Mean	Mean
Grupos de Edad	Menor de 14 años	Nº de horas semanales ver TV	12	13
	Entre 15-29 años	Nº de horas semanales ver TV	11	12
	Menores 30	Nº de horas semanales ver TV	11	12
	Entre 30-40 años	Nº de horas semanales ver TV	11	12
	Entre 41-55 años	Nº de horas semanales ver TV	11	13
	Mas de 56 años	Nº de horas semanales ver TV	13	14
	Total	Nº de horas semanales ver TV	11	13

Figura 6.29 Promedio del número de horas semanales de televisión vistas correspondiente a cada categoría de edad

Fuente: SPSS®

Mínimo y máximo de horas semanales de televisión vistas correspondiente a cada categoría de edad y sexo.

			Sexo			
			Hombre		Mujer	
			Mínimo	Máximo	Mínimo	Máximo
Grupos de Edad	Menor de 14 años	Nº de horas semanales ver TV	3	21	3	21
	Entre 15-29 años	Nº de horas semanales ver TV	3	42	3	25
	Menores 30	Nº de horas semanales ver TV	3	42	3	25
	Entre 30-40 años	Nº de horas semanales ver TV	1	21	3	25
	Entre 41-55 años	Nº de horas semanales ver TV	1	21	3	21
	Mas de 56 años	Nº de horas semanales ver TV	3	21	3	25
	Total	Nº de horas semanales ver TV	1	42	3	25

Figura 6.30 Mínimo y máximo de horas semanales de televisión vistas correspondiente a cada categoría de edad y sexo

Fuente: SPSS®

Cambio del aspecto de las tablas

Puede cambiar el aspecto de una tabla editando sus propiedades o aplicando el procedimiento <Aspecto de tabla>. Cada aspecto de tabla contiene un conjunto de propiedades de tabla, relativas al aspecto general, los bordes, las propiedades de las notas al pie y las propiedades de las casillas. Puede seleccionar uno de los aspectos de tabla preestablecidos o crear y guardar uno personalizado.

Procedimiento para configurar el aspecto de una tablado de contingencia

Un aspecto de tabla es un conjunto de propiedades que definen la configuración de una tabla. Puede seleccionar un aspecto previamente definido o crear uno propio.

Antes o después de aplicar el procedimiento <aspecto de tabla>, se pueden cambiar los formatos de casillas individuales o de grupos de casillas utilizando las propiedades de casilla. Los formatos de casilla editados se conservan intactos, aunque se aplique un nuevo aspecto de tabla

Para aplicar o guardar un aspecto de tabla.

- Procedimiento SPSS®

- Al activar la opción <Tabla de contingencia Bivariada> aparece el siguiente Cuadro de diálogo y al activar los procedimientos disponibles:



Figura 6.31 Cuadro de diálogo Cruce de variables (Crosstabs)

Fuente: SPSS®

Aparecerá en el Visor de resultados la siguiente tabla

Tabla de contingencia Sexo * Grupos de Edad

			Grupos de Edad					
			Menor de 14 años	Entre 15-29 años	Entre 30-40 años	Entre 41-55 años	Mas de 56 años	Total
Sexo	Hombre	Recuento	227	380	223	238	120	1188
		% de Sexo	19,1%	32,0%	18,8%	20,0%	10,1%	100,0%
	Mujer	Recuento	220	371	246	223	93	1153
		% de Sexo	19,1%	32,2%	21,3%	19,3%	8,1%	100,0%
Total	Recuento		447	751	469	461	213	2341
	% de Sexo		19,1%	32,1%	20,0%	19,7%	9,1%	100,0%

A continuación siga el siguiente procedimiento SPSS®:

- De un doble <clik> en la sección interna de la Tabla.
- Elija en la Barra de menús: la opción <Formato>



Figura 6.32 Tabla pivote de SPSS

Fuente: SPSS®

- Seleccione la opción <Aspectos de Tabla> (Table Looks)

A continuación aparecerá el Cuadro de diálogo Aspecto de la Tabla

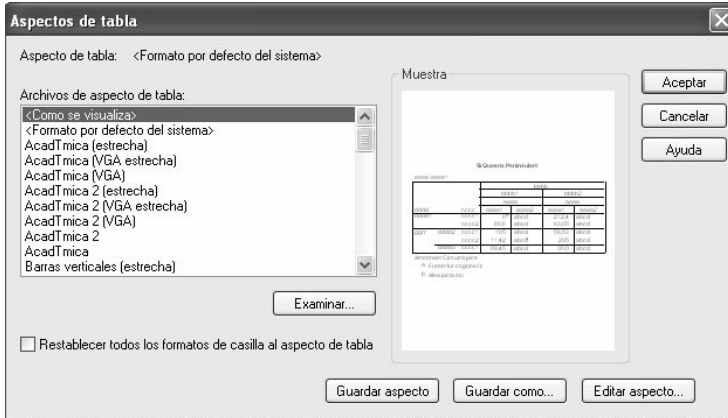


Figura 6.33 Aspectos de tabla

Fuente: SPSS®

- Seleccione un tipo de tabla en la lista de archivos en la ventanilla Archivos de aspecto de tablas (Tables Look Files).
- Para cambiar de color del fondo de la tabla, seleccione la opción <Propiedad de Tablas> (Table Properties).

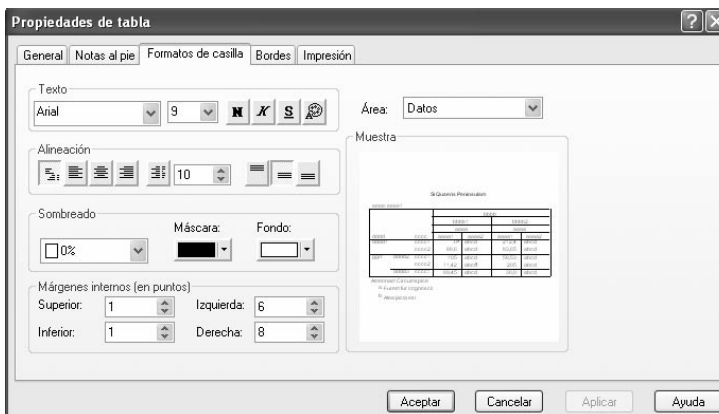


Figura 6.34 Propiedades de la tabla

Fuente: SPSS®

- Seleccione <Formatos de casilla>.
- En la ventanilla <Fondo> seleccione el color de fondo de su gráfico.
- Dé un <clik> en <Aceptar> para aplicar el aspecto escogido de tabla.

Gestión y transformación de los datos

El tratamiento de los datos es la fase del proceso de investigación posterior a la recolección de datos. Por lo general, supone un tratamiento informático de los datos (codificación, digitación, grabación y validación) para su posterior manipulación, tabulación y análisis. En esta sección se abordan los aspectos operativos relativos a la transformación de los datos mediante un proceso de recodificación o reagrupamiento de los valores de una variable.

Al investigador le puede interesar agrupar algunas de las categorías de las variables categóricas, bien porque el número de respuestas en una o más de ellas ha sido muy bajo o no haya habido ninguna respuesta o se desee reducir el número de categorías y así simplificar la manipulación o presentación de los resultados. Igual situación puede darse con los valores de variables continuas, por ejemplo, “*edad en años cumplidos*” que sería conveniente agruparlos en grupos etarios quinquenales o decenales.

En el supuesto de 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 resumir o reagrupar⁴¹ la información. Lo que se trata es de transformar un número limitado de categorías (generalmente ocho o diez). 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> y elija el submenú <Datos>.

41 Recodificación de los valores en rangos iguales.

- Seleccione el archivo situación-2000.sav.
- En la barra de menús, elija el menú <Transformar> y luego <Recodificar>.
- Elija la opción <En distintas variables>.
- Aparece el Cuadro de diálogo Recodificar en distintas variables (véase figura 6.35).

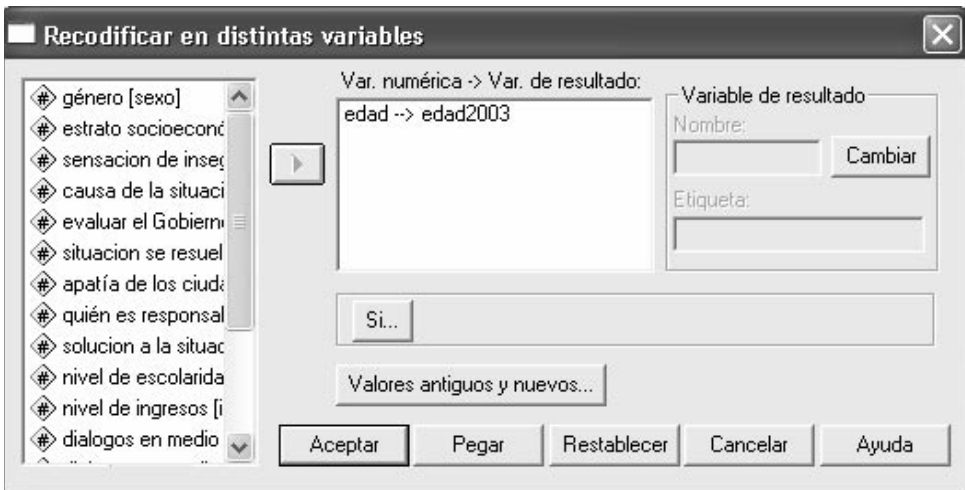


Figura 6.35 Cuadro de diálogo para recodificación con distintas variables

Fuente: SPSS®

- Seleccione en el listado de variables de la ventanilla izquierda, la variable *edad* y transfírala a la ventanilla central, presionando el botón ►.
- 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 <clik> en el botón <Cambiar>. En la ventanilla central aparecerá así: edad →edad2003.
- Definido el nuevo nombre de <clik> en el botón <Valores antiguos y nuevos>. A continuación se abrirá el siguiente Cuadro de diálogo:

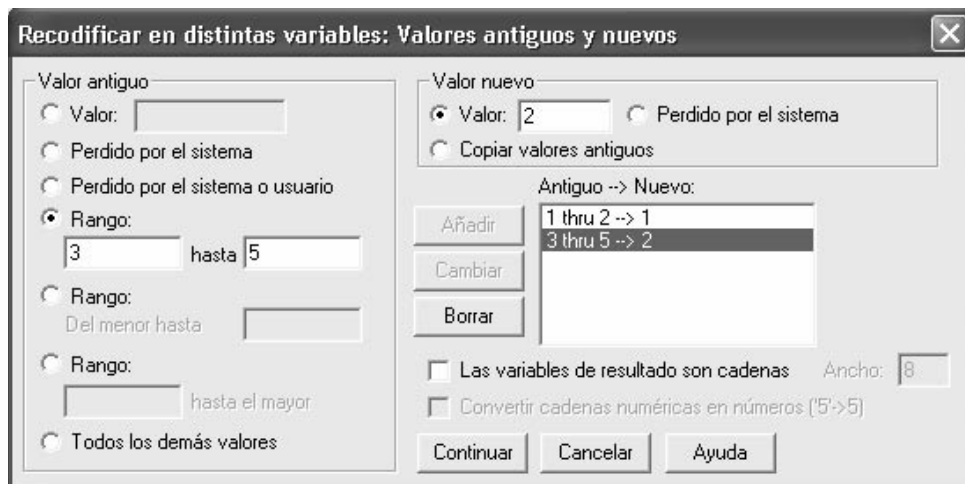


Figura 6.36 Cuadro de diálogo proceso de recodificación de variables

Fuente: SPSS®

En este Cuadro de diálogo se plantea la siguiente secuencia de procedimientos:

- En la sección <Rango>, véase en la ventanilla izquierda y digite el rango mínimo: 1
- En la ventanilla derecha, digite el rango máximo: 2
- En la ventanilla <Valor nuevo>, véase en la casilla Valor y escriba el nuevo valor: 1
- De un <clik> en el botón de comando <Añadir>

Los nuevos valores recodificados aparecerán en la sección Antigo → Nuevo. Repita los cuatro pasos anteriores con los valores originales y el nuevo.

- Dé un <clik> en el botón <Continuar> y luego en el botón <Aceptar>.
- En la Matriz de datos aparecerá la nueva variable *edad2003*.

Gestión y transformación de los archivos

Los archivos de datos no siempre están organizados en la forma requerida para las necesidades específicas del analista. Puede que le interese combinar archivos de datos, organizar los datos en un orden diferente, seleccionar un subconjunto de casos o cambiar la unidad de análisis agrupando casos.

Entre la gama de posibilidades de transformación de archivos que ofrece el programa SPSS® se encuentran las siguientes:

Ordenar datos (Sort Cases). Puede ordenar los casos en función del valor de una o más variables.

Transponer casos y variables (Transpose). El formato de archivo de datos de SPSS lee las filas como casos y las columnas como variables. Permite intercambiar las filas y las columnas para leer los datos en el formato correcto.

Fundir archivos (Merge Files). Puede fundir dos o más archivos de datos. Es posible combinar archivos con las mismas variables pero con casos distintos, o con los mismos casos pero variables diferentes.

Reestructurar datos (Restructure). Puede reestructurar los datos para crear un único caso (registro) a partir de varios casos o crear varios casos a partir de un único caso.

Agregar datos (Aggregate). Puede cambiar la unidad de análisis agregando casos basados en el valor de una o más variables de agrupación.

Identificar casos duplicados (Identify Duplicate Cases). La identificación de los casos duplicados le permite definir prácticamente como quiera lo que se considera *duplicado*

Seleccionar subconjuntos de casos (Split Files). Puede restringir el análisis a un subconjunto de casos o efectuar análisis simultáneos de subconjuntos diferentes.

Seleccionar casos (Select Cases). Esta opción proporciona varios métodos para seleccionar un subgrupo de casos basándose en criterios que incluyen variables y expresiones complejas. También se puede seleccionar una muestra aleatoria de casos.

Ponderar datos (Weight Cases). Puede ponderar los casos para un análisis basado en el valor de una variable de ponderación.

- Procedimiento SPSS®

En la Barra de menús, seleccionar <Datos>: aparecerá un menú desplegable:

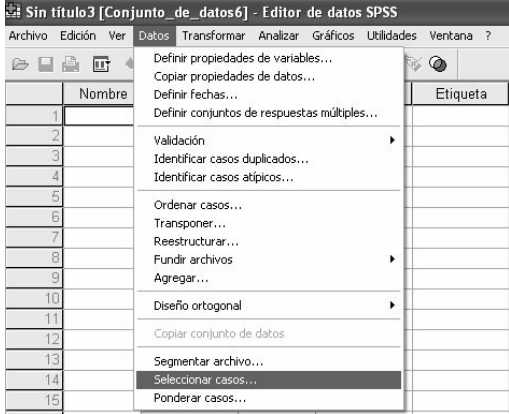


Figura 6.37 Cuadro de diálogo, menú de datos

Fuente: SPSS®

– Seleccionar la opción <Seleccionar caso>. A continuación aparece el Cuadro de diálogo Selección de Casos.

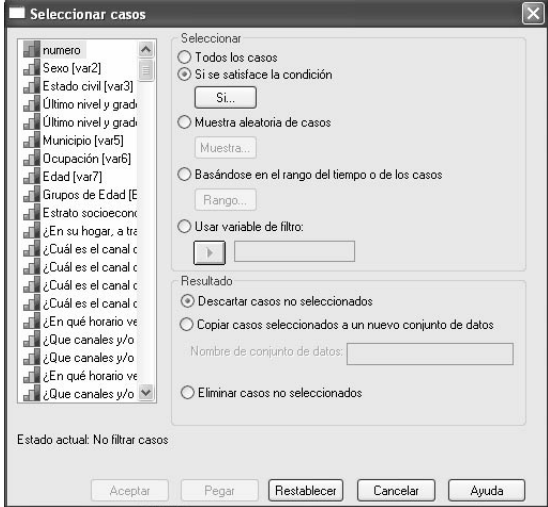


Figura 6.38 Cuadro de diálogo Selección de Casos

Fuente: SPSS®

Opciones en el Cuadro de diálogo Selección de Casos

Todos los casos (All Cases). Desactiva el filtrado y utiliza todos los casos.

Si se satisface la condición (If condition is satisfied). Utiliza una expresión condicional para seleccionar los casos. Si el resultado de la expresión condicional es verdadero, el caso se selecciona. Si el resultado es falso o perdido, entonces el caso no se selecciona.

Muestra aleatoria de casos (Random Samples Cases). Selecciona una muestra aleatoria basándose en un porcentaje aproximado o en un número exacto de casos.

Basándose en el rango del tiempo o de los casos (based on time...). Selecciona los casos basándose en un rango de los números de caso o en un rango de las fechas/horas.

Usar variable de filtro (Use Filter Variable). Utiliza como variable para el filtrado la variable numérica seleccionada del archivo de <datos>. Se seleccionan los casos con cualquier valor distinto del 0 o del valor perdido para la variable seleccionada.

Ejemplo. En el estudio de cultura y prácticas televisivas en el Área Metropolitana del Valle de Aburrá, interesa analizar sólo la información del Municipio de La Estrella.⁴²

Procedimiento SPSS®

- En la Barra de menús seleccionar el menú <Datos>
- A continuación aparecerá un menú emergente elija <Seleccionar casos>
- En el Cuadro de diálogo resalte y traslade con el botón ► la variable Municipio.
- Seleccionar la opción <“If Condition is satisfied”>
- Presione el Botón <IF >

Aparece el Cuadro de diálogo Seleccionar de casos: SI

42 En la base de datos del estudio el Municipio de La Estrella esta identificado con el código 9. Los 10 municipios del Área Metropolitana del Valle del Aburrá están comprendidos en la variable 5 -var5-

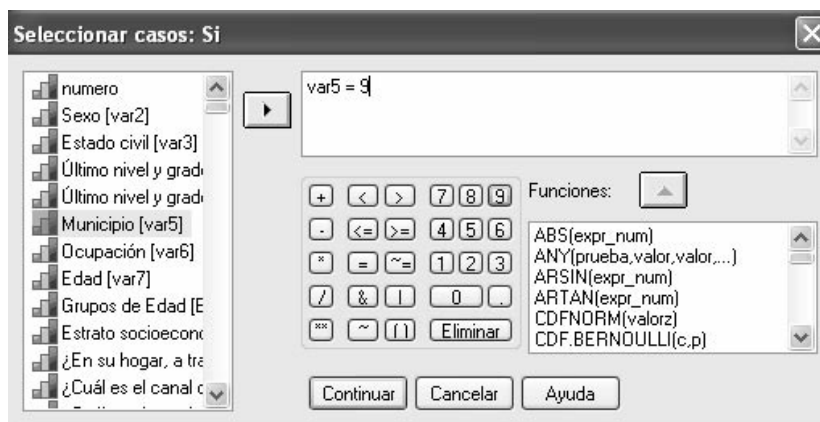


Figura 6.39 Cuadro de diálogo Selección de Casos

Fuente: SPSS®

Seleccionar casos: Si la opción

Este Cuadro de diálogo permite seleccionar subconjuntos de casos utilizando expresiones condicionales. Una expresión condicional devuelve un valor *verdadero*, *falso* o *perdido* para cada caso.

- Seleccione la variable Municipio (var5) en la lista de *variables fuente*.
- Presione el botón < flecha > para introducirla en el cuadro destino a la derecha.
- Complete la sintaxis: var5 = 9 (Municipio de La Estrella)
- Presione el botón <Continuar> (véase figura 6.40)
- En el cuadro interno, marcar (√) en la sección <Eliminar casos no seleccionados>.
- A continuación presionar el Botón <Aceptar>

En el Visor de resultados aparecerá una nueva matriz de datos que sólo incluye a los registros recolectados en el Municipio de La Estrella⁴³.

43 Estudio de Cultura y practicas televisivas en el Área Metropolitana del Valle del Aburra. Centro de Estudio de Opinión Universidad de Antioquia - Comisión Nacional de Televisión, Medellín: 2004.

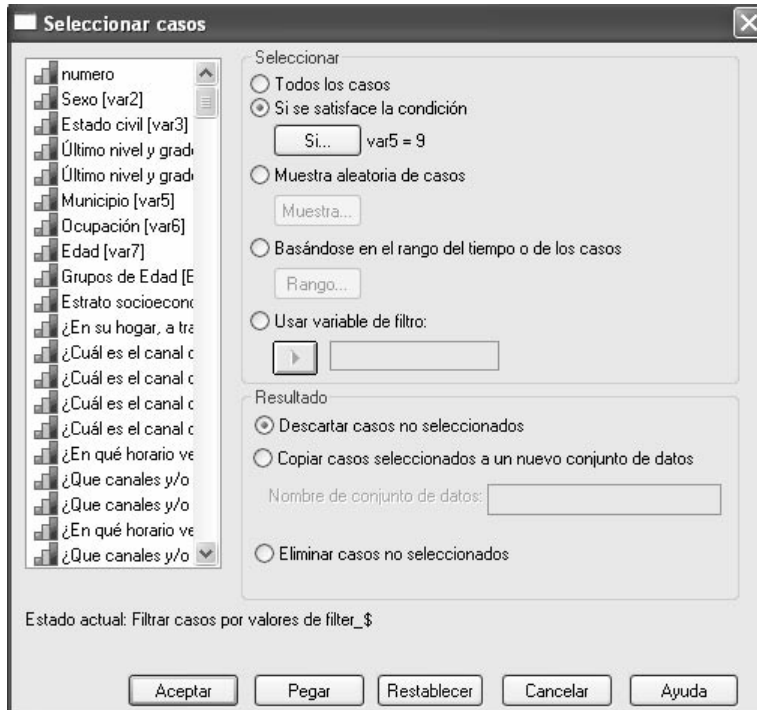


Figura 6.40 Cuadro de diálogo Seleccionar Casos

Fuente: SPSS®

Inferencia estadística

El objetivo de una investigación no se reduce sólo al trabajo descriptivo. 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”.*⁴⁴

44 Hernández S., Roberto; Fernández C. Carlos; Baptista L. Pilar. Metodología de la investigación. México: McGraw-Hill. 1992. Pág. 373.

En esta sección, 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 de 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 seguir:

1. ¿Hay relación entre las variables independiente y dependiente?
2. ¿Cuál es la dirección o forma de la relación?
3. ¿Cuán fuerte es la relación?
4. ¿Es la relación estadísticamente significativa?
5. ¿Es la relación de tipo causal?

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

Se hipotetiza que sólo el 28% 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, cuándo aceptar o rechazar una intuición.

Se hipotetiza un cierto valor para la media de la población de reincidentes (42%). Para probar su validez, 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 del valor hipotético para la media esté correcto. Sin embargo, no en 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.

El ejemplo 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. En casos como éstos, las llamadas pruebas de significación estadística pueden ayudar a responde, con un criterio estadístico, 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 (individuos, grupos, instituciones, hechos, eventos, divisiones político-administrativas o geográficas, etc.).

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, costos 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 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, se 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.

Por otro lado, no se debe olvidar la preocupación — en el ejemplo anterior — acerca de sí la tasa de reincidencia en el programa innovador es realmente concluyente y, por 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 *significancia estadística*, se hace necesario comprender la lógica del muestreo.

Tanto en los problemas de relaciones como en los de diferencias, para contestar a las preguntas: ¿existe verdadera relación entre las variables? o, ¿existe diferencia real entre dos medias? Se debe adoptar un nivel de significación (el máximo nivel de error que se asume en la decisión) y este se compara con la probabilidad del estadístico de contraste. En general, si es menor la probabilidad del estadístico que el nivel de significación, se dice que existen diferencias (o que existe correlación), o bien que se han producido resultados significativos. Existe diferencia significativa entre dos medias cuando la probabilidad de pertenencia a la misma población es menor que un valor pequeño (usualmente el 0.05 o 5%).⁴⁵

En los problemas de diferencias de medias con variables ordinales o cuando las muestras son muy pequeñas se emplean *las pruebas no paramétricas*, que en realidad son pruebas de diferencias entre las medias de los rangos (rangos medios). La Media es la medida adecuada de tendencia central con variables de intervalo o razón, y el rango medio lo es con variables ordinales.

En la estadística paramétrica se trabaja bajo el supuesto de que las muestras se han originado a partir de poblaciones que poseen ciertas distribuciones conocidas, donde cada función de distribución teórica depende de uno o más parámetros poblacionales. El uso eficiente de los métodos paramétricos de estimación y de contrastación requiere, por lo general, la normalidad de la población y la estabilidad de la varianza. Sin embargo, en muchas situaciones, es imposible especificar la forma de la distribución poblacional, y sólo podemos suponer en forma razonable que es continua. El proceso de sacar conclusiones directamente de las observaciones muestrales, sin formar los supuestos con respecto a la forma matemática de la distribución poblacional, se llama teoría no paramétrica. Los análisis estadísticos realizados bajo estos supuestos constituyen la estadística no paramétrica⁴⁶.

45 El valor alfa o probabilidad a elegir puede ser 0.05 y 0.01.

46 César Pérez. Técnicas Estadísticas con SPSS, Madrid: Prentice Hall, 2001. Pág. 273.

Las pruebas no paramétricas se emplean para comprobar la existencia de diferencias entre grupos que han sido medidos con variables ordinales. Estas pruebas comprueban las diferencias entre las medias de las variables ordenadas por rangos, ya que con variables ordinales el rango medio es el estadístico de tendencia central adecuado para comprobar diferencias de tendencia central.⁴⁷

Existen dos tipos generales de técnicas inferenciales: las paramétricas (por ejemplo, las de la sección de diferencia de medias) y las no paramétricas. Ambos tipos tienen como función determinar la probabilidad del estadístico de contraste para compararlo con una probabilidad fija (alfa) y aceptar o rechazar en base a ello la hipótesis experimental. La diferencia entre los dos tipos de técnicas radica en que las pruebas no paramétricas tienen menos supuestos que cumplir.

- Procedimiento SPSS®

- En la Barra de menú seleccionar <Analizar>
- En el menú desplegable, seleccionar <Prueba no paramétricas>
- Aparece un submenú con las siguientes opciones

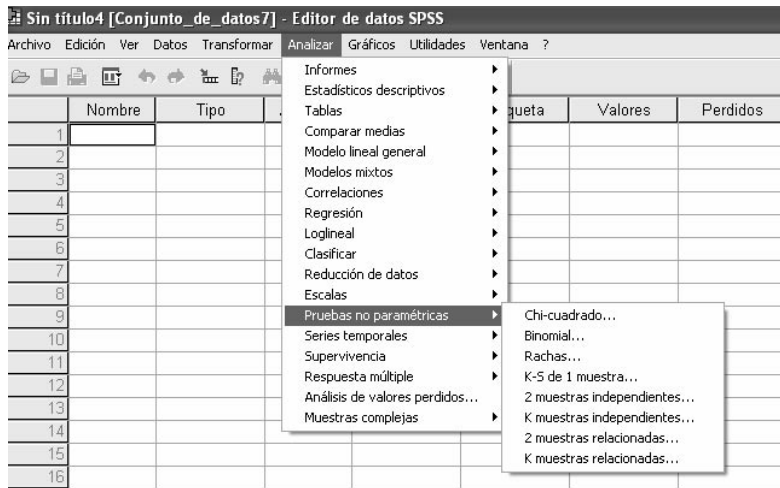


Figura 6.41 Cuadro de dialogo Pruebas no paramétricas

Fuente: SPSS®

47 Camacho, R. Juan, Estadística con SPSS para Windows. Madrid: Ed. RA-MA, 2002. Pág. 233 - 234

Las pruebas no paramétricas (Nonparametric Tests) cuyos procedimientos ofrece SPSS® son:

La prueba Chi-Cuadrado para una muestra,

Prueba Binomial

Prueba de rachas (Run test: Median)

Prueba Kolmogorov-Smirnov para muestras independientes

Prueba para dos muestras independientes (U de Mann-Whitney)

Prueba para varias muestras Independientes (kruskal - Wallis)

Prueba para dos muestras relacionadas (two related simples: Wilcoxon)

Prueba para varias muestras relacionadas (Several related samples: Kendall)⁴⁸

Diferencias de medias

En esta sección se trabaja con una metodología que permite analizar los problemas de diferencias de medias. Básicamente se utiliza cuando se trata de averiguar si dos o más grupos tienen una media significativamente distinta. El programa SPSS® aporta algunos comandos que analizan estas diferencias, entre otros, Comparación de diferencia de medias Means, Pruebas T (T-Test), Oneway, Unianova y GML.

Medias (Means). Comparación de las diferencias de medias. Este comando realiza una descripción de características de las muestras (medias, desviaciones típicas y frecuencias de grupo de una variable dependiente, el nivel de intervalo, dentro de los grupos definidos por una o más variables independientes

Prueba T para muestras independientes o relacionadas (T-TEST). Comparación Simple de Medias. Sirve para comparar dos medias o una media con respecto a la media poblacional. La hipótesis de investigación propone que los dos grupos difieren significativamente entre sí y la hipótesis nula

48 César Pérez. Ídem. Pág. 272 - 308

propone que los grupos no difieren significativamente. Su nivel de medición debe ser de intervalos o razón.

Modelo lineal general (GLM). Técnica estadística multivariante que proporciona un análisis de regresión y un análisis de varianza para una variable dependiente mediante uno o más factores o variables.

ANOVA de un factor (ONEWAY). Prueba de diferencia entre dos o más medias⁴⁹. Es una prueba estadística para analizar si más de dos grupos difieren significativamente entre sí en cuanto a sus medias y varianzas. La prueba “t” se utiliza para dos grupos y el análisis de varianza unidireccional ONEWAY se usa para tres, cuatro o más grupos.

La hipótesis a probar: existe diferencia entre más de dos grupos. La hipótesis de investigación nula (H_0) propone que los grupos difieren significativamente entre sí y la hipótesis alternativa (H_1) propone que los grupos no difieren significativamente.

El nivel de medición requerido es la variable independiente debe ser categórica o cualitativa (nominal o ordinal) y la dependiente de intervalos o razón.

Un ejemplo: En un estudio patrocinado por la Comisión Nacional de Televisión⁵⁰ se requería determinar los perfiles de preferencias de programas televisivos. Uno de los objetivos fue diferenciar los cuatro perfiles o nichos de preferencias de los televidentes (Nicho 1⁵¹ “Los misceláneos”, Nicho 2: “Los culturales”, Nicho 3: “los Light” y Nicho 4: “Los deportivos”) según el sexo de los televidentes.

- Procedimiento SPSS®

- En la Barra de menú principal seleccionar < Analizar >
- Elegir el submenú < Comparar medias > (Compare Means)
- Seleccionar la opción < Anova de un factor > (Oneway Factor)

49 Comparaciones múltiples de Medias.

50 Prácticas culturales y televisivas en el Área Metropolitana del Valle del Aburrá. Jaime Ruiz R. Miguel Aigner A. Efrén Barrera R. Centro de Estudios de Opinión -CEO- de la Universidad de Antioquia y Comisión Nacional de Televisión, Teleantioquia, TeleMedellín y Canal U. Medellín, 2004. Pág. 51 - 85

51 Los nichos o perfiles están representados estadísticamente por los cuatro componentes.

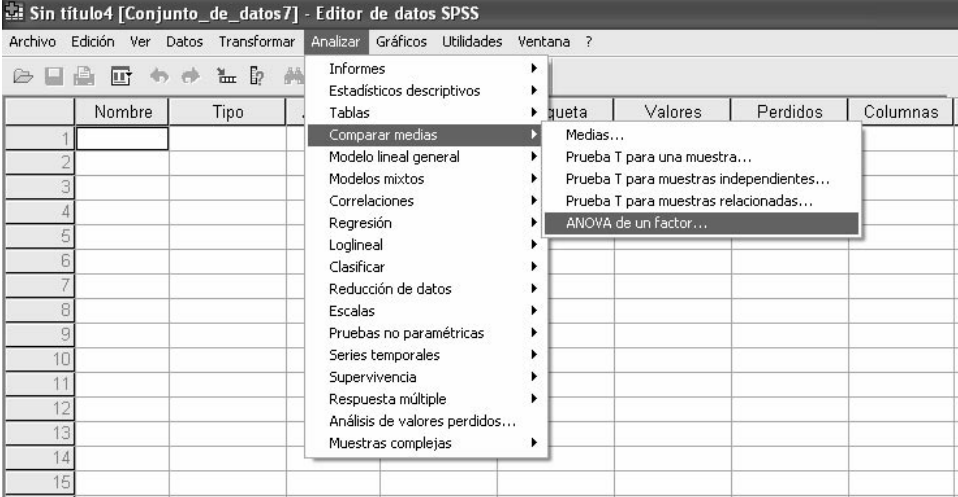


Figura 6.42 Cuadro de diálogo Anova de un factor

Fuente: SPSS®

Aparece el Cuadro de diálogo Anova de un factor (One Way Anova).



Figura 6.43 Cuadro de diálogo Anova de un factor

Fuente: SPSS®

En el Cuadro de diálogo proceda a trasladar a la ventanilla Lista de variables dependientes (Dependent List) los factores 1, 2, 3 y 4 - obtenidos en el proceso de calculo de los Componentes Principales de la información obtenida en el estudio de cultura y prácticas televisivas en el Área Metropolitana del Valle del Aburra⁵².

REGR factor 1: FAC1_1 (primer perfil)

REGR factor 2: FAC1_1 (segundo perfil)

REGR factor 3: FAC1_1 (tercer perfil)

REGR factor 4: FAC1_1 (cuarto perfil)

Como factor seleccionar la variable categórica Sexo

A continuación, en el Cuadro de diálogo Anova de un facto (One Way Anova). Presione el botón <Opciones>. Señale, marcando los procedimientos necesarios, en este ejemplo, de acuerdo a lo que se observa en la figura siguiente y luego presione <Continuar>.

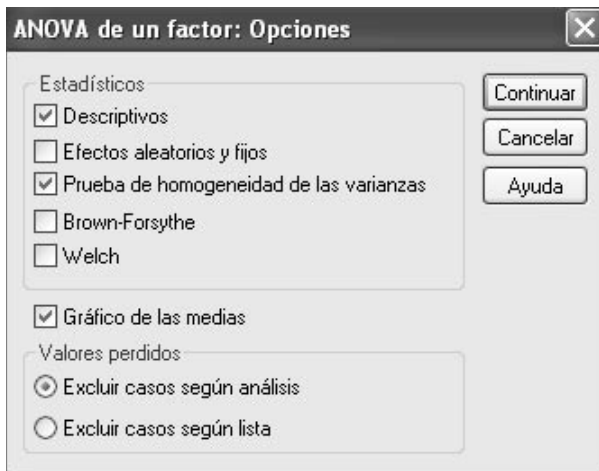


Figura 6.44 Cuadro de dialogo Anova de un factor: opciones

Fuente: SPSS®

52 Prácticas culturales y televisivas en el Área Metropolitana del Valle del Aburrá. Jaime Ruiz R. Miguel Aigner A. Efrén Barrera R. Centro de Estudios de Opinión -CEO- de la Universidad de Antioquia y Comisión Nacional de Televisión, Teleantioquia, TeleMedellín y Canal U. Medellín, 2004. Pág. 51 - 85

En el Visor de resultados aparecerá la siguiente tabla.

Descriptivos

		N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
						Límite inferior	Límite superior		
Componente 21-35 1	Hombre	1188	-,2752996	,97501766	,02828813	-,3308000	-,2197993	-3,98872	2,37224
	Mujer	1153	,2836585	,94499655	,02783015	,2290531	,3382600	-4,09974	2,44819
	Total	2341	,0000000	1,00000000	,02066804	-,0405296	,0405296	-4,09974	2,44819
Componente 21-35 2	Hombre	1188	,1825073	,90018431	,02611700	,1312667	,2337480	-2,54056	2,27245
	Mujer	1153	-,1880475	1,06139360	,03125804	-,2493765	-,1267184	-2,96509	2,63336
	Total	2341	,0000000	1,00000000	,02066804	-,0405296	,0405296	-2,96509	2,63336
Componente 21-35 3	Hombre	1188	,4205995	,89134891	,02586065	,3698619	,4713372	-2,93109	2,48330
	Mujer	1153	-,4333671	,91771798	,02702680	-,4863944	-,3803398	-4,05761	2,19203
	Total	2341	,0000000	1,00000000	,02066804	-,0405296	,0405296	-4,05761	2,48330
Componente 21-35 4	Hombre	1188	,1444581	,99958678	,02900095	,0875592	,2013569	-3,58936	2,79548
	Mujer	1153	-,1488432	,97878785	,02882531	-,2053992	-,0922872	-3,56145	2,99801
	Total	2341	,0000000	1,00000000	,02066804	-,0405296	,0405296	-3,58936	2,99801

Fuente: SPSS®

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
Componente 21-35 1	Inter-grupos	182,810	1	182,810	198,217	,000
	Intra-grupos	2157,190	2339	,922		
	Total	2340,000	2340			
Componente 21-35 2	Inter-grupos	80,343	1	80,343	83,164	,000
	Intra-grupos	2259,657	2339	,966		
	Total	2340,000	2340			
Componente 21-35 3	Inter-grupos	426,703	1	426,703	521,644	,000
	Intra-grupos	1913,297	2339	,818		
	Total	2340,000	2340			
Componente 21-35 4	Inter-grupos	50,335	1	50,335	51,420	,000
	Intra-grupos	2289,665	2339	,979		
	Total	2340,000	2340			

Fuente: SPSS®

-Análisis de datos

Si se analizan las diferencias de las medias según el sexo de los televidentes se aprecian diferencias significativas entre hombres y mujeres. A título de ejemplo, si se analizan las variables (programación) que componen el primer factor que une los concursos, los dramatizados, farándula, películas, reality y telenovelas, se puede apreciar que este factor es básicamente femenino (media = ,2836585). En cambio, el cuarto factor que comprende

básicamente la programación de deportes y Noticias es preferido por los hombres (Media = ,1444581). En este caso la delimitación de las variables (programación) que forman cada factor (perfil del televidente) según sexo del televidente es clara.

En la tabla Anova, lo primero que hay que hacer es observar el valor de F que es igual a 198.217 (valor del Factor 1) $p \leq 0.000$. Como esta probabilidad es menor que el nivel de significancia usual del 5% (0.05) ($0.000 < 0.05$) se concluye que hay diferencias significativa entre las medias de los cuatro grupos de perfiles de televidentes. Si se analizan los resultados en la Tabla Anova, se puede apreciar lo siguiente: la variabilidad entre grupos (Between Groups) es altamente significativa, al ser el valor del nivel de significación⁵³ (Sig.) menor que 0.05.

Prueba de hipótesis

Una vez recogida la muestra probabilística, la importancia de ella radica en que permita predecir el comportamiento de la población. Hay que recordar 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, se puede tener un valor aproximado como resultado del conocimiento o de la experiencia acerca del objeto de estudio.

La prueba de hipótesis comienza con una afirmación o propuesta tentativa llamada *hipótesis*⁵⁴, que se hace acerca de un parámetro de una población. 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.

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

53 Valor del nivel de significancia: Si es menor a 0.05 (5%) se rechaza la H_0 : A diferente a B
Si es mayor a 0.05 se acepta la H_0 : a es diferente a B.

54 En la estadística inferencial hay dos tipos de hipótesis: una, es la hipótesis nula, se define para que el analista evalúe su hipótesis de investigación. La otra, la hipótesis de contrastación, señala la posible existencia o relación entre dos hechos.

- 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.

El objetivo de una prueba de hipótesis, es 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.

Partiendo de un estadístico muestral (una media aritmética); entonces se puede 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 “proximidad” 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, permitirá rechazar la hipótesis nula. Únicamente reduciéndolo, se podrá reducir la probabilidad de rechazar la hipótesis nula cuando deberíamos aceptarla.

Selección de la prueba estadística

Téngase en cuenta que la *hipótesis estadística* es una conjetura o suposición concerniente a las características de la población. Antes de aceptarla o recha-

zarla, 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.

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

- *Prueba t de Student*. 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*⁵⁵. 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.

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

Tabla 6.12 De palabras recordadas de un texto sencillo y de un texto complejo.

Grupo 1 (texto sencillo). Puntajes	Grupo 2 (texto complejo). Puntajes
10	2
5	1
6	7
3	4
9	4
8	5
7	2
5	5
6	3
5	4
Media ₁ : 6.4	Media ₂ : 3

55 La prueba *t* se puede utilizar para comparar resultados de un pretest con los resultados de un post-test.

La hipótesis de trabajo es: ¿existe diferencia significativa entre las medias obtenidas por ambos grupos?

La pregunta de investigación hacía referencia a si era posible afirmar que las diferencias mostradas en la capacidad de recordación eran significativamente validas. A partir de los datos, se puede concluir que las personas que integran ambos grupos tienen procesos de recordación diferentes.

Esquemáticamente la contrastación de hipótesis se presenta así:

H_0 : 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). Así:

H_0 : Puntajes Grupo 1 No difieren Puntajes Grupo 2
estadísticamente

H_i : 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_i : Puntaje Grupo 1 es diferente al Puntaje Grupo 2

- Procedimiento SPSS®

- En la barra de menús, elija el menú <Archivos> y luego el submenú <Abrir> y después la opción <Datos>
- Seleccione el archivo de trabajo: *recordación.sav*⁵⁶
- En la matriz del Editor de datos SPSS crear una primera variable de comparación: *puntajes*, compuesta por los valores obtenidos en el grupo 1 (texto sencillo) y grupo 2 (texto complejo).
- En la matriz del Editor de datos SPSS, crear una segunda variable de comparación: *códigos*, con los siguientes valores consecutivos: 1= sencillo; 2 = complejo (véase figura 6.45).
- Seleccione el menú <Analizar> y elija la opción <Comparar Medias>.
- Elegir el procedimiento <Prueba T para muestras independientes>⁵⁷

56 Ver archivos de trabajo anexos al final del texto.

57 El procedimiento Prueba T para muestras independientes compara medias de dos grupos de casos.

recuerdos - Editor de datos SPSS						
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?						
1 : sencillo		10				
	sencillo	complejo	puntaje	codigo	var	var
1	10	2	10	1		
2	5	1	5	1		
3	6	7	6	1		
4	3	4	3	1		
5	9	4	9	1		
6	8	5	8	1		
7	7	2	7	1		
8	5	5	5	1		
9	6	3	6	1		
10	5	4	5	1		
11	.	.	2	2		
12	.	.	1	2		
13	.	.	7	2		
14	.	.	4	2		
15	.	.	4	2		
16	.	.	5	2		
17	.	.	2	2		
18	.	.	5	2		
19	.	.	3	2		
20	.	.	4	2		

Figura 6.45 Cuadro de diálogo Editor de Datos

Fuente: Archivo de trabajo anexo al final del texto

- A continuación aparecerá el Cuadro de diálogo Prueba T para muestras independientes (véase figura 6.46).

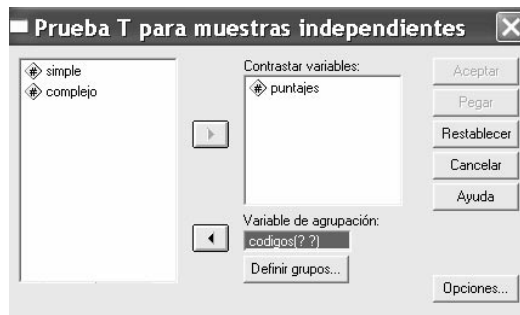


Figura 6.46 Prueba de T para muestras independientes

Fuente: SPSS®

- En primer lugar, seleccione la variable dependiente: *puntajes* y transfírela a la ventana *Contrastar variables*.
- En segundo lugar, seleccione la variable independiente: *códigos* –en donde están los grupos– y transfírela a la ventana *Variable de agrupación*.
- Dé un <clik> en el botón de comando <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.
- Digitar el código 1 en la casilla al lado derecho del Grupo 1.
- Digitar el código 2 en la casilla al lado derecho del Grupo 2 (véase figura 6.47).

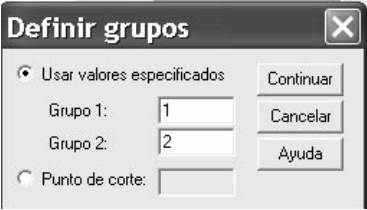


Figura 6.47 Cuadro de diálogo Definir grupos

Fuente: SPSS®

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

Tabla 6.13 Prueba de muestras independientes

	Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias							
	F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error tip. de la diferencia	95% Intervalo de confianza para la diferencia		
								Inferior	Superior	
PUNTAJE Se han asumido varianzas iguales	,423	,524	3,095	18	,006	2,70	,87	,87	4,53	
No se han asumido varianzas iguales			3,095	17,438	,006	2,70	,87	,86	4,54	

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* – es decir, se acepta la H_1 –

Si el nivel de significación obtenido es menor al valor crítico adoptado (0,05), se debe aceptar la H_1 (*las diferencias entre los puntajes son estadísticamente significativas*) es decir se rechaza la hipótesis de igualdad de puntajes.

- Análisis de los datos

A dos grupos de personas se les solicitó que leyeran, a los primeros, un texto sencillo y a los segundos, un texto complejo. Se creía que recordarían más palabras los lectores del texto sencillo que los 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: $3,095 p \leq 0,006$. Al analizar el valor del nivel de significación (p Valor) en la prueba t para la igualdad de medias que es de 0,006, resulta que es *menor al valor crítico elegido 0,05*; luego hay que aceptar que las diferencias son significativamente válidas. Entonces, la diferencia encontrada en los puntajes de ambos grupos de lectura es válida.

Comparación de medidas ordinales

Cuando los datos a comparar en dos subgrupos corresponden a medidas de nivel ordinal como, por ejemplo, puntajes en escalas de actitudes, y se está 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*, se utiliza para comparar dos muestras no relacionadas en la forma que se indica en el ejercicio siguiente.

Se ha aplicado una prueba de “conformismo al aprendizaje” a una muestra de adolescentes de los cuales quince pertenecen a una escuela urbana y veintiuno a una escuela rural. Los resultados se ven en la tabla 6.14.

Tabla 6.14 Puntajes de aceptación al aprendizaje

Estudios Urbanos		Estudios Rurales	
Escolar	Puntaje*	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
		16	39
		17	42
		18	40
		19	26
		20	25
		21	24

* 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 puntaje máximo obtenido).

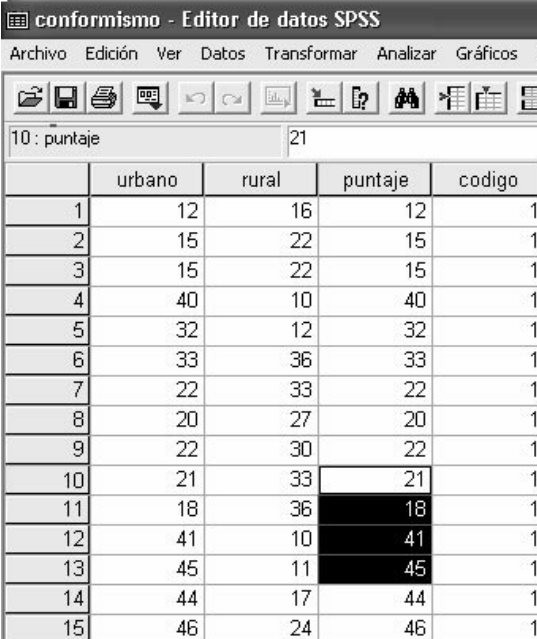
La hipótesis de trabajo es la siguiente:

H_0 : No existen diferencias estadísticamente significativas en el grupo de puntajes entre los estudiantes de escuelas de nivel urbano y los de nivel rural.

H_{ii} : Existen diferencias estadísticamente significativas en el grupo de puntajes entre los estudiantes de escuelas de nivel urbano y los de nivel rural.

- Procedimiento SPSS®

- En la barra de menús, elija el menú <Archivos>, luego <Abrir> y después <Datos>.
- Seleccione el archivo de trabajo *conformismo.sav*⁵⁸.
- En el Editor de datos SPSS® cree una primera variable de comparación: puntajes, compuesta por los valores obtenidos en el grupo 1 (urbano) y grupo 2 (rural)
- Cree una segunda *variable de comparación*: códigos, con los siguientes valores consecutivos: 1= urbano; 2 = rural (véase figura 6.48).



The screenshot shows the SPSS 'conformismo - Editor de datos SPSS' window. The menu bar includes Archivo, Edición, Ver, Datos, Transformar, Analizar, Gráficos, and L. Below the menu is a toolbar with various icons. The main area displays a data table with the following content:

	urbano	rural	puntaje	codigo
1	12	16	12	1
2	15	22	15	1
3	15	22	15	1
4	40	10	40	1
5	32	12	32	1
6	33	36	33	1
7	22	33	22	1
8	20	27	20	1
9	22	30	22	1
10	21	33	21	1
11	18	36	18	1
12	41	10	41	1
13	45	11	45	1
14	44	17	44	1
15	46	24	46	1

Figura 6.48 Editor de Datos. Variables de contraste

Fuente: SPSS®

58 Ver archivos de trabajo anexos al final del texto.

- Seleccione el menú <Analizar>, luego elija el submenú <Pruebas no-paramétricas>
- Seleccione la opción: <2 muestras independientes Mann-Whitney>.

A continuación aparece el Cuadro de diálogo: Prueba para dos muestras independientes.

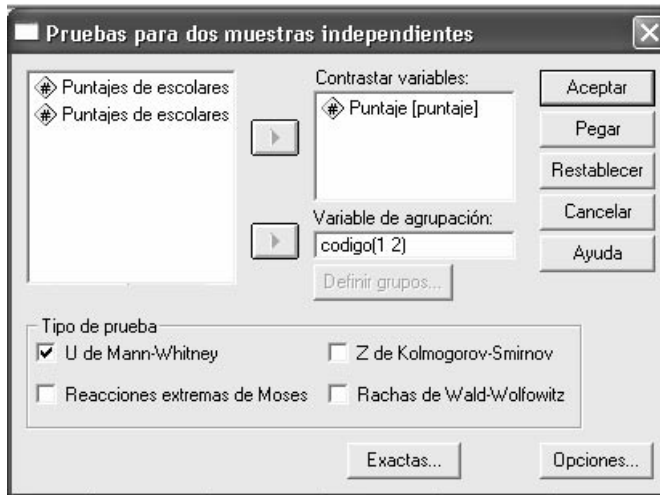


Figura 6.49 Cuadro de diálogo Pruebas para dos muestras independientes

Fuente: SPSS®

- En primer lugar, seleccione la variable dependiente: *puntajes* y transfírela a la ventana *Contrastar variables*.
- En segundo lugar, seleccione la variable independiente: *códigos* – en donde están los grupos – y transfírela a la ventana *Variable de agrupación*.
- En la sección Tipo de prueba: seleccione <U de Mann-Whitney>.
- Dé un <clik> en el botón <Definir grupos>.

A continuación aparecerá el Cuadro de diálogo Definir grupos

- Contrastar variables: seleccionar la variable *puntaje*
- Variable de agrupación: Seleccionar la variable *Código* (??)

- Definir grupos: como se van a comparar los grupos 1 –urbano– y 2 –rural– se digita el código 1 en la casilla al lado derecho del Grupo 1. Igualmente se digita 2 en la casilla al lado derecho del Grupo 2.



Figura 6.50 Cuadro de diálogo de muestras independientes

Fuente: SPSS®

Presione el botón <Continuar> y por último, presione el botón <Aceptar>

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

Tabla 6.15 Prueba de Mann-Whitney

	COD	N	Rango promedio	Suma de rangos
PUNTAJE	Urbano	15	19,67	295,00
	Rural	21	17,67	371,00
	Total	36		

	PUNTAJE
U de Mann-Whitney	140,000
W de Wilcoxon	371,000
Z	-,562
Sig. asintót. (bilateral)	,574
Sig. exacta [2*(Sig. unilateral)]	,590 ^a

Fuente: SPSS®

- 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_i que afirma que las diferencias son significativas entre los estudiantes urbanos y los rurales.

En este caso, como 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_o = A no se diferencia significativamente de B

H_i = A se diferencia significativamente de B

Medidas de relaciones entre variables

En este apartado se examinan las relaciones entre variables. De esta forma se completa, 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, “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”⁵⁹.

Cuando se habla de posibles relaciones entre dos variables, lo que se busca es analizar sí: *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 semicualitativas (variables nominales y ordinales) y con el término correlación la relación entre características cuantitativas*⁶⁰...

Cuando el tipo de medición utilizado es cuantitativo, en el caso específico de variables tipo intervalo y razón o continuas, el rango de valores varía entre

59 Briones, Guillermo. Op. Cit. Pág. 76

60 Ídem

-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 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.

La tabla 6.16 muestra las principales medidas de relación entre variables nominales, ordinales, intervalo y proporcionales.

Tabla 6.16 Cuadro sinóptico de medidas de relaciones simples entre variables

Medidas de relación		
Datos nominales	Datos ordinales	Datos Intervalos o de proporción
Comparación de porcentajes	tau b	
Coefficientes de Contingencia	tau c	
Chi cuadrado	D de Somer	Coefficiente de Pearson
Coefficiente de Goodman-Kruskal	Gamma	
Lambda	Coefficiente Rho de Spearman	

Medidas de asociación entre variables cualitativas

A continuación se examinan algunas *estadísticas no-paramétricas* que permiten medir asociaciones entre variables medidas a nivel nominal y ordinal.

- *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 la no relación entre variables – independencia – y el investigador evalúa si en su caso esto es cierto o no. Lo que es a menudo de interés.

La hipótesis de investigación afirma que la modalidad de estudio es independiente al programa académico en cual está 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 la hipótesis nula – H_0 – se acepta la hipótesis alternativa H_1 , o sea, la no independencia (relación) entre las variables.

La prueba de hipótesis se plantea como:

H_0 : Independencia entre las variables – No relación entre *tipo de estudiante* y *modalidad de estudio* –

H_1 : No independencia entre las variables –relación entre *tipo de estudiante* y *modalidad de estudio* –

- Procedimiento SPSS®

- En la barra de menús, elija el menú <Archivos>, luego elija <Abrir> y después <Datos>.
- Seleccione el archivo de trabajo *habitos-2.sav*⁶¹
- Seleccione el menú <Analizar>.
- Seleccionar la opción <Estadísticos descriptivos>
- Active la opción <Tablas de contingencia>

A continuación aparece el Cuadro de diálogo Tablas de contingencia. (veáse figura 6.51)

- Transfiera a la ventanilla Filas, la variable independiente *tipo*.
- Transfiera a la ventanilla Columnas, la variable dependiente *estudio*.
- Dar <clic> en el botón <Estadísticos>.

61 Ver archivos de trabajo anexos al final del texto

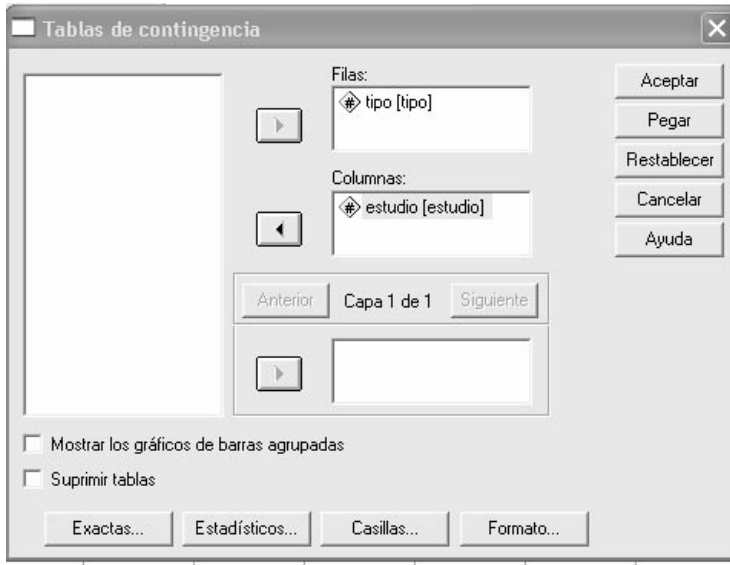


Figura 6.51 Cuadro de diálogo Tablas de contingencia

Fuente: SPSS®

A continuación aparecerá el Cuadro de diálogo *Tablas de contingencia: Estadísticos*:

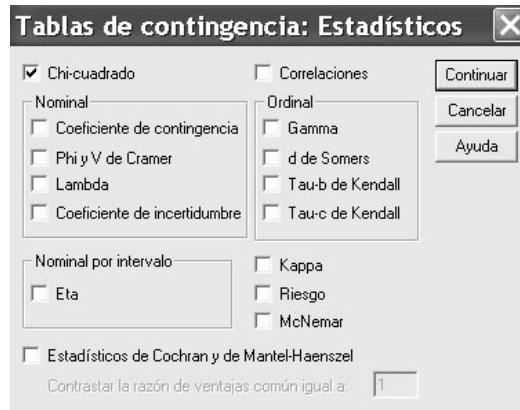


Figura 6.52 Cuadro de diálogo Tablas de contingencia para cálculos estadísticos

Fuente: SPSS®

- Complete los campos requeridos, marcando el estadístico Chi Cuadrado.
- Dar un <click> en el botón <Continuar>

A continuación aparecerá el Cuadro de diálogo:

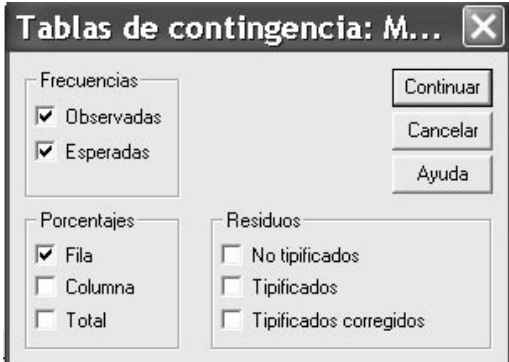


Figura 6.53 Cuadro de diálogo tabla de contingencia: mostrar en las casillas

Fuente: SPSS®

Al completar los campos de acuerdo con el modelo anterior en la ventanilla Frecuencias, seleccione o marque la opción Esperadas.

Dar un <click> en el botón <continuar>

A continuación aparecerá en el Visor de resultados:

Tabla 6.17 Tabla de contingencia tipo * estudio

			estudio			Total
			estudio sistemático diario	periodos irregulares de estudio	combinación de los dos tipos	
tipo ciencias sociales y humanas	Recuento	6	15	23	44	
	% de tipo	13.6%	34.1%	52.3%	100.0%	
Medicina Veterinaria	Recuento	10	8	24	42	
	% de tipo	23.8%	19.0%	57.1%	100.0%	
Total	Recuento	16	23	47	86	
	% de tipo	18.6%	26.7%	54.7%	100.0%	

Fuente: SPSS®

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	3.107 ^a	2	.212
Razón de verosimilitud	3.150	2	.207
Asociación lineal por lineal	.099	1	.753
N de casos válidos	86		

- Análisis de los datos

La pregunta que se plantea 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 sí existe independencia entre ambas variables. El valor del Chi Cuadrado de Pearson de 3.107 $p \leq 0,212$ indica que las variables *tipo de estudiante* y *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, si es mayor – de acuerdo con los resultados es $0,212$ – se debe aceptar que *no hay relación o asociación entre las variables*.

Por tanto, se puede aceptar la hipótesis nula, de la independencia entre la modalidad de estudio y el tipo de estudiantes de las Facultades de Ciencias Sociales y de Medicina veterinaria.

En síntesis, el Chi Cuadrado permite decidir si las variables de tipo cualitativo, son independientes o asociadas. Hay que destacar que la existencia de una asociación entre las variables del estudio medida por el valor logrado no indica que tan fuerte o débil 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.

En este apartado se introduce el fundamento de relación entre variables. Entonces, un investigador debe 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 como *puntajes en estadística*, obtenidos por los estudiantes de la Facultad de Ciencias Sociales.

En un caso como éste, un investigador mide 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, 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.

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.

Se coloca como caso, que a un investigador le interesa saber si existe relación entre la habilidad lectora y la habilidad en el manejo de la herramienta estadística. El objetivo sería más bien estudiar si los estudiantes que obtienen un puntaje alto o bajo en una variable – *capacidad lectora* –, también obtienen puntajes altos o bajos en la otra – *capacidad de manejo de la herramienta estadística* –, véase la tabla 6.18.

La hipótesis de trabajo es la siguiente:

H_0 : No existe asociación entre los puntajes obtenidos en estadística y los puntajes obtenidos en habilidad lectora – *independencia entre las variables*.

H_1 : Existe asociación entre los puntajes obtenidos en estadística y los de habilidad en lectura.

- Procedimiento SPSS®

- En la barra de menús, elija el menú <Archivos>, luego <Abrir> y después <Datos>.

Tabla 6.18 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

Fuente: SPSS®

- Seleccione el archivo de trabajo *nominal.sav*⁶²
- En la barra de menús, seleccione el menú < Analizar>.
- Seleccionar la opción <Estadísticos descriptivos> y luego <Tablas de contingencia>.
- En el Cuadro de diálogo <Tablas de contingencia> seleccione < Estadísticos>
- Seleccione los estadísticos para *variables nominales*: *Coficiente de contingencia C*; *Phi* y *V. de Cramer*.
- En el Cuadro de diálogo *Tablas de contingencia*: *mostrar en las casillas*: en la casilla *Porcentajes*: marcar *filas*.

A continuación, en el Visor de resultados aparece la siguiente tabla:

62 Ver archivos de trabajo anexos al final del texto

Tabla 6.19 Medidas simétricas

		Valor	Sig. aproximada
Nominal por nominal	Phi	2,273	,101
	V de Cramer	,859	,101
	Coefficiente de contingencia	,915	,101
N de casos válidos		12	

a. Asumiendo la hipótesis alternativa.

b. Empleando el error típico asintótico basado en la hipótesis nula.

Fuente: SPSS®

- Análisis de los datos

Como el nivel de significación es mayor al 5% – valor crítico escogido de 0,05 –, se acepta la hipótesis nula – H_0 – es decir, no existe relación entre la habilidad lectora y los puntajes obtenidos en habilidad estadística.

H_0 = No existe asociación entre las variables, si es mayor al valor crítico escogido.

H_1 = Existe asociación entre las variables, si 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.

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 asociación observada. 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%.

Es interesante 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 6.20 Puntajes de participación de estudiantes en deportes y frecuencia de ausentismo.

Sujeto	Variable A	Variable B
	Puntaje en escala de participación deportiva (1 – 10)	Puntaje en escala de frecuencias de ausentismo (1 – 5)
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
11	8	4
12	10	5

La hipótesis de trabajo afirma que una alta participación de los estudiantes en las actividades deportivas está relacionada con bajas tasas de ausentismo académico.

H_o : No existe asociación entre los puntajes obtenidos de participación deportiva y ausentismo académico⁶³ (Independencia entre las variables).

H_i : Existe asociación entre los puntajes en participación deportiva y ausentismo académico.

63 Ausentismo académico hace referencia al abandono ocasional del aula de clase y se le distingue de retiro forzoso, repitencia y otros tales como movilidad, *mortalidad* académica, interrupción de estudios, entre otros.

La tabla 6.21 muestra los resultados de los puntajes en escala de participación deportiva y frecuencia de ausentismo académico en un grupo de estudiantes.

Tabla 6.21 Puntajes en escala de participación deportiva y frecuencia de ausentismo en un grupo de estudiantes.

Sujeto	Puntaje en escala de participación deportiva (1 – 10)	Rango A	Puntaje en escala de frecuencias de ausentismo (1 – 5)	Rango B
1	5	5,0	2	3,5
2	3	2,5	2	3,5
3	7	7,0	4	7,5
4	10	12	5	11
5	9	9,5	4	7,5
6	9	9,5	5	11
7	2	1,0	4	7,5
8	6	6,0	3	5,0
9	3	2,5	1	1,5
10	4	4,0	1	1,5
11	8	8,0	4	7,5
12	10	12	5	11

- Procedimiento SPSS®

- En la barra de menús, elija el menú <Archivos>, luego elija <Abrir> y después <Datos>.
- Seleccione el archivo de trabajo *deportes_ausentismo.sav*⁶⁴
- En la barra de menús, seleccione el menú <Analizar>.
- Elija la opción <Correlaciones> y luego activar opción <Bivariada>

64 Ver archivos de trabajo anexos al final del texto

Aparece el Cuadro de diálogo Correlaciones bivariadas



Figura 6.54 Cuadro de diálogo Correlaciones bivariadas

Fuente: SPSS®

- Elija y transfiera cada una de las variables *rangoa* y *rangob* a la ventanilla Variables:
- Elija el coeficiente de asociación de <Spearman>.
- Seleccione en la opción <Prueba de significación> el procedimiento <Bilateral>.

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

Tabla 6.22 Correlaciones no paramétricas

Correlaciones			RANGO A	RANGO B
Rho de Spearman	RANGO A	Coefficiente de correlación	1.000	.780**
		Sig. (bilateral)	.	.003
		N	12	12
	RANGO B	Coefficiente de correlación	.780**	1.000
		Sig. (bilateral)	.003	.
		N	12	12

** . La correlación es significativa al nivel 0,01 (bilateral).

Fuente: SPSS®

- Análisis de los datos

El nivel de significación (bilateral) estadística encontrado es de 0,003, es decir, menor al valor crítico escogido (0,05); por lo tanto, no existe asociación entre los puntajes que miden la participación deportiva y el ausentismo académico.

Análisis de regresión y correlaciones

Hasta ahora, sólo se 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 (la 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

El análisis de regresión y correlación permite mostrar cómo deberíamos determinar tanto la naturaleza como la fuerza de la relación que 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, una técnica que permita utilizar 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 una 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, considerada una como independiente y la otra como dependiente – decisión de tipo teórico metodológica.

Además se puede resumir los datos y predecir o estimar nuevos valores de la variable dependiente de acuerdo con los valores observados en la variable independiente.

También se busca 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 adquiridas en sistematización y el monto de los ingresos de los egresados (población muestral).

- Procedimiento SPSS®

- En la barra de menús, elija el menú <Archivos>, luego <Abrir>, y después <Datos>.
- Seleccione el archivo de trabajo *ingresos-2003.sav*⁶⁵.
- En la barra de menús, seleccione el menú <Analizar>.
- En el menú <Analizar> seleccione la opción <Regresión> y luego elija el procedimiento <Lineal>



Figura 6.55 Cuadro de diálogo Regresión

Fuente: SPSS®

Aparece la ventana Regresión lineal (veáse la figura 6.56)

- Seleccione y transfiera la variable *ingresos mensuales* a la ventanilla Dependiente.
- Seleccione y transfiera la o las variables predictoras o independientes – horas de capacitación – a la ventanilla Independiente y de <clíc> en el botón <Aceptar>

65 Ver archivos de trabajo anexos al final del texto.

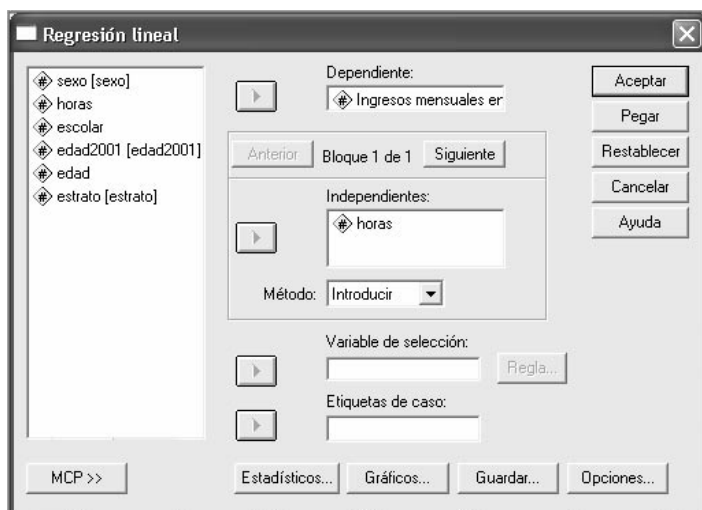


Figura 6.56 Cuadro de diálogo Regresión lineal

Fuente: SPSS®

A continuación, en el Visor de resultados aparecerán las siguientes tablas:

Tabla 6.23 Visor de resultados

Variables introducidas/eliminadas ^b			
Modelo	Variables introducidas	Variables eliminadas	Método
1	HORAS ^a		Introducir

a. Todas las variables solicitadas introducidas
 b. Variable dependiente: Ingresos mensuales en dólares

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,948 ^a	,898	,893	132,05

a. Variables predictoras: (Constante), HORAS

Fuente: SPSS®

- Análisis de 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*, $R^2 = 0,898$; ello indica que el 89,8% de la varianza de la variable ingresos está predicha por la variable horas de capacitación.

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 la *variable criterio*.

Coeficientes 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 para resolver este tipo de inquietudes. El análisis de correlación es una herramienta que se puede usar para describir el grado en el que una variable está relacionada con otra.

Correlación de Pearson

El coeficiente de correlación de Pearson⁶⁶ mide básicamente la naturaleza de la relación entre dos variables cuantitativas. No sólo 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 si es negativo, lo contrario.

La fuerza o intensidad de la relación puede ser débil o fuerte, o incluso puede no existir. Generalmente se acepta el criterio según el cual si la rela-

66 La notación tradicional de un coeficiente de correlación es r minúscula.

ción entre las variables se acerca a cero indica debilidad y si se acerca a uno 1 sería indicativo de una alta correlación.

En la figura siguiente se observa una posición relativa de un coeficiente r:

-1,00	-0,70	-0,50	-20	0,0	0,20	0,50	0,70	+1,00
Correlación negativa alta	negativa fuerte	moderada	débil o baja		moderada	fuerte		Correlación positiva alta

Figura 6.57 Índices de correlación entre variables cuantitativas

Fuente: M. Aignerren. Investigación cuantitativa en Ciencias Sociales. Procesamiento y análisis de datos. Centro de Estudios de Opinión. Medellín: Vieco. 1997. Pág. 143

Un ejemplo, mediante un test⁶⁷ se ha medido la habilidad en comprensión y lectura y el uso de las herramientas estadísticas en un grupo de estudiantes 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, por ello, los estudiantes que tienen un buen desempeño en comprensión y lectura de textos lo tendrán también en el uso de la estadística, como herramienta de análisis.

Los resultados de la tabla 6.24 muestran los puntajes obtenidos en dos pruebas que medían la capacidad de lectura y comprensión y los puntajes obtenidos en el manejo de estadísticos descriptivos. (veáse tabla 6.24)

Correlaciones bivariadas

El procedimiento Correlaciones bivariadas calcula el coeficiente de correlación de Pearson, el rho de Spearman y la tau-b de Kendall con sus niveles de significación. Las correlaciones miden cómo están relacionadas (fuerza y dirección) las variables o los órdenes de los rangos.

67 Instrumentos de medición generan puntajes cuantitativos que miden las capacidades de uso de las herramientas.
 68 Variable A: puntaje obtenido en prueba que miden el grado de utilización de la herramienta estadística como elemento de análisis.
 69 Variable B: puntaje obtenido en prueba de comprensión y lectura de textos.

Tabla 6.24 Puntajes obtenidos en la prueba de lectura y comprensión y estadística social

Alumno	Variable A ⁶⁸	Variable B ⁶⁹
1	30	20
2	50	10
3	60	20
4	73	12
5	87	16
6	24	7
7	44	9
8	35	18
9	97	19
10	62	13
11	49	6
12	93	10

Fuente: SPSS®

El coeficiente de correlación de Pearson es una medida de asociación lineal. Dos variables pueden estar perfectamente relacionadas, pero si la relación no es lineal, el coeficiente de correlación de Pearson no será un estadístico adecuado para medir su asociación.

Ejemplo. ¿Está el número de cursos registrados en un programa académico correlacionado con el promedio de notas obtenido en el semestre? Al analizar los datos del semestre 2005 II, se descubre que el coeficiente de correlación de Pearson (-0,401) es significativo al nivel 0,05. Se puede afirmar que cuantos más cursos registren por semestre, un menor promedio de notas obtendrán. Estas variables están correlacionadas negativamente (-0,401).

Para cada pareja de variables se pueden utilizar: coeficiente de correlación de Pearson, Rho de Spearman, tau-b de Kendall.

Utilizar variables cuantitativas simétricas para el coeficiente de correlación de Pearson y variables cuantitativas o variables con categorías ordenadas para la Rho de Spearman y la tau-b de Kendall.

- Procedimiento SPSS®

- En la barra de menús, elija el menú <Archivos>, luego <Abrir> y después <Datos>.

- Seleccione el archivo de trabajo *nominal.sav*⁷⁰
- En la barra de menús, seleccione el menú <Analizar> y luego <Correlaciones>.
- Seleccione el procedimiento <Bivariadas> y luego el Cuadro de diálogo Correlaciones bivariadas.
- Seleccione y transfiera la variable independiente y la variable dependiente a la ventanilla Variables.
- En la sección Coeficientes de correlación elegir el procedimiento <Pearson>.
- En la sección Prueba de significación elegirle procedimiento <Bilateral>.
- Al presionar el botón de comando <Aceptar> aparecen los siguientes resultados en el Visor. de resultados:

Tabla 6.25 Visor de resultados de Correlaciones

		puntaje en estadística	Puntaje en lectura
puntaje en estadística	Correlación de Pearson	1,000	,857**
	Sig. (bilateral)	,	,000
	N	12	12
Puntaje en lectura	Correlación de Pearson	,857**	1,000
	Sig. (bilateral)	,000	,
	N	12	12

** La correlación es significativa al nivel 0,01 (bilateral).

Fuente: SPSS®

- Análisis de los datos

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 $r = 0,857 p \leq 0,000$

70 Ver archivos de trabajo anexos al final del texto.

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, cuando se trate de *regresión múltiple*, se vuelve 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 con más cuidado la dirección causal entre las variables; una muestra puede ser: tener claro cuál es la variable independiente y cuál es la dependiente. El resultado del cálculo de una regresión: 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 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. Así, el análisis multivariado es aquel que analiza la relación entre varias variables independientes actuando simultáneamente en más de una variable dependiente. Comprende un conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes en el sentido de que hay varias variables medidas para cada individuo u objeto estudiado. Su razón de ser radica en un mejor entendimiento del proceso u hecho social objeto de estudio.

En esta sección se va a dar una breve visión general de dicho conjunto de técnicas exponiendo, brevemente, cuál es su finalidad, ilustrada con ejemplos

Ejemplo. Los expertos en rating televisivo de hoy no pueden seguir las especulaciones clásicas en las que los televidentes eran considerados homogéneos

y caracterizados por algunas variables demográficas. En su lugar, deben desarrollar estrategias que atraigan a numerosos segmentos de televidentes con características demográficas y psicográficas diversas en un mercado con múltiples restricciones (legales, económicas, competitivas, tecnológicas, etc.). Sólo a través del análisis multivariante las relaciones múltiples de este tipo pueden ser *examinadas adecuadamente* para obtener un entendimiento más completo y real del entorno que permita tomar las decisiones más adecuadas sobre la programación más adecuada.

Tipos de técnicas multivariantes

Métodos de Dependencia: suponen que las variables analizadas están divididas en dos grupos: las variables dependientes y las variables independientes. El objetivo de los métodos de dependencia consiste en determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes y de qué forma. Las variables dependientes pueden ser cuantitativas y cualitativas. Si son cuantitativas los métodos generalmente usados son: Análisis de Regresión; Análisis de Supervivencia; MANOVA; Correlación Canónica.

Si las variables dependientes son cualitativas los métodos pueden ser: Análisis Discriminante; Regresión Logística; Análisis Conjoint.

Métodos de Interdependencia: estos métodos no distinguen entre variables dependientes e independientes y su objetivo consiste en identificar qué variables están relacionadas, cómo lo están y por qué. Se trabaja con dos tipos de datos: cualitativos o nominales y cuantitativos.

Si los datos son cuantitativos se pueden utilizar los siguientes procedimientos estadísticos: Análisis de Componentes Principales; Análisis Factorial; Escalas Multidimensionales; Análisis Cluster

Si los datos son cualitativos se pueden utilizar los siguientes procedimientos estadísticos: Análisis de Correspondencia; Modelos Log-lineales; Escalas multidimensionales; Análisis de Cluster.

Métodos Estructurales: suponen que las variables están divididas en dos grupos: el de las variables dependientes y el de las independientes. El objetivo de estos métodos es analizar, no sólo como las variables independientes afectan a las variables dependientes, sino también cómo están relacionadas las variables de los dos grupos entre sí.

Regresión múltiple

En secciones anteriores se describieron 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 dos 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.

La regresión múltiple permite evaluar el valor individual de cada variable independiente en la variable dependiente, mientras se controla tal influencia por el impacto de otras variables de tipo cuantitativas al igual que la influencia de todas las variables independientes combinadas.

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

Por ejemplo, se aplicó una encuesta a estudiantes, 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 con la siguiente escala de calificaciones:

Muy deficiente 1 deficiente 2 aceptable 3 bueno 4 excelente 5

El objetivo del estudio fue estudiar la relación entre la evaluación del profesor de un curso realizada por sus estudiantes y cinco características del profesor y del curso: dominio de los temas, empleo de recursos didácticos, respeto en el trato con los estudiantes, evaluación de ejercicios y exámenes, evaluación del interés de los contenidos.

- Procedimiento SPSS®

- En la barra de menús, elija el menú <Archivos>, luego <Abrir> y después <Datos>.

- Seleccione el archivo de trabajo *encuesta al estudiante sobre el curso y el profesor.sav*⁷¹
- En la barra de menús, seleccione el menú <Analizar>, luego el submenú <Regresión>
- Seleccione la opción <Lineal>.

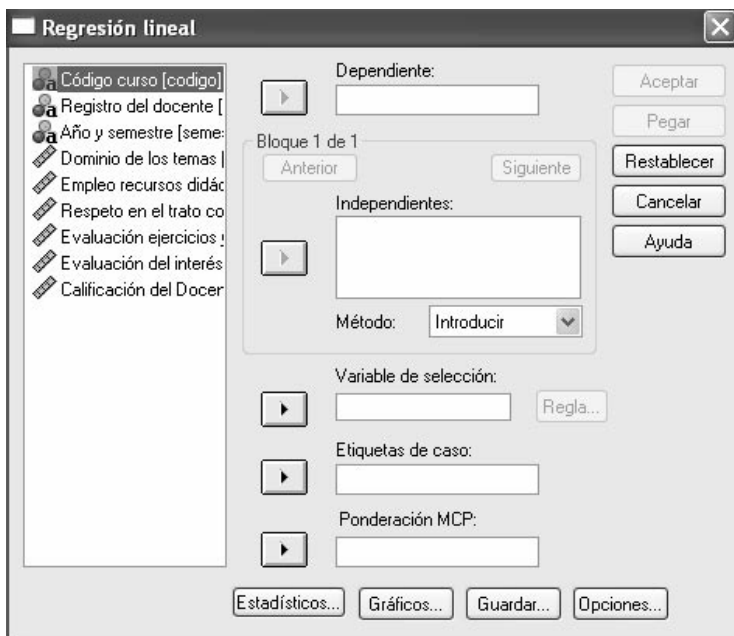


Figura 6.58 Cuadro de diálogo Regresión

Fuente: SPSS®

- Seleccione y transfiera la variable Calificación del docente de 1 a 5 a la ventanilla <Dependiente>
- Seleccione y transfiera las variables: *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>
- Dar un <clik> sobre el botón <Aceptar>.

71 Ver archivos de trabajo anexos al final del texto.

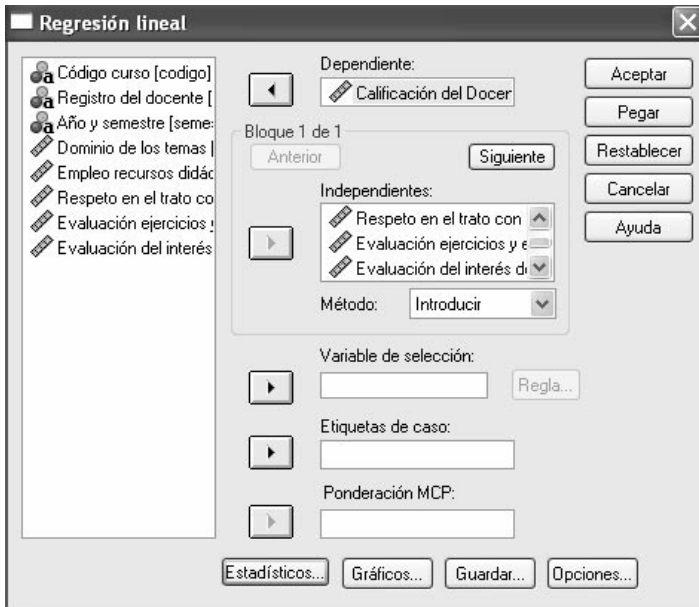


Figura 6.59 Cuadro de diálogo

Fuente: SPSS®

– En el Cuadro de diálogo dar un <clíc> sobre el botón <Estadísticos>.

A continuación aparecerá el Cuadro de diálogo Regresión lineal: Estadísticos

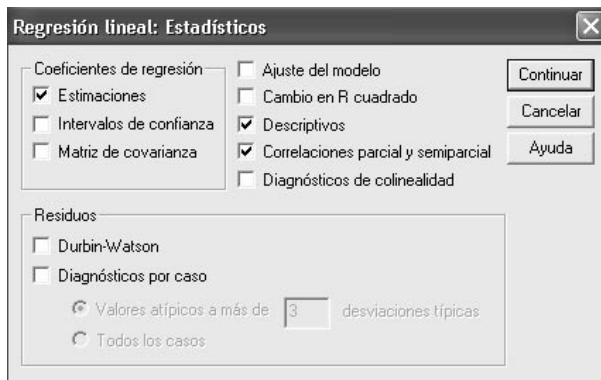


Figura 6.60 Cuadro de diálogo Regresión lineal: Estadísticos

Fuente: SPSS®

- En la ventana Coeficiente de regresión, seleccione marcando las opciones: <Estimaciones>, <Descriptivos> y <Correlaciones parcial y semiparcial>
- A continuación, dar un <clic> sobre el botón <Continuar>

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

Tabla 6.26 Visor de resultados Estadísticos descriptivos

	Media	Desviación típ.	N
Calificación del Docente de 1 a 5	3,84	,55	25
Dominio de los temas	3,92	,49	25
Empleo recursos didácticos	4,24	,72	25
Respeto en el trato con estudiantes	3,92	,49	25
Evaluación ejercicios y exámenes	3,80	,65	25
Evaluación del interés de los contenidos	3,84	,69	25

Fuente: SPSS®

En la tabla 6.26 se señalan 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.

En la tabla 6.27 se observan las correlaciones entre cada una de las variables asociadas a la evaluación docente.

Tabla 6.27 Visor de resultados Correlaciones.

	Calificación del Docente de 1 a 5	Dominio de los temas	Empleo recursos didácticos	Respeto en el trato con estudiantes	Evaluación ejercicios y exámenes	Evaluación del interés de los contenidos
Calificación del Docente de 1 a 5	1,000	,714	,516	,561	,839	,477
Dominio de los temas	,714	1,000	,640	,486	,733	,329
Empleo recursos didácticos	,516	,640	1,000	,640	,642	,583
Respeto en el trato con estudiantes	,561	,486	,640	1,000	,733	,697
Evaluación ejercicios y exámenes	,839	,733	,642	,733	1,000	,488
Evaluación del interés de los contenidos	,477	,329	,583	,697	,488	1,000

Fuente: SPSS®

- Análisis de los datos

El peso o influencia⁷² 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 *Calificación 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*.

Tabla 6.28 Visor de resultados Resumen del Modelo y ANOVA⁷³

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,874 ^a	,763	,701	,30

a. Variables predictoras: (Constante), Evaluación del interés de los contenidos, Dominio de los temas, Respeto en el trato con estudiantes, Empleo recursos didácticos, Evaluación ejercicios y exámenes

ANOVA ^b						
Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	5,617	5	1,123	12,249	,000 ^a
	Residual	1,743	19	9,172E-02		
	Total	7,360	24			

a. Variables predictoras: (Constante), Evaluación del interés de los contenidos, Dominio de los temas, Respeto en el trato con estudiantes, Empleo recursos didácticos, Evaluación ejercicios y exámenes

Fuente: SPSS®

⁷² La llamada fuerza global de la asociación mide el peso entre las variables independientes y dependiente

⁷³ ANOVA: Es una prueba estadística para evaluar el efecto de dos o más variables independientes sobre una variable dependiente. La tabla ANOVA muestra el contraste para comprobar si la relación global es significativa o no.

Un índice de regresión⁷⁴ $R^2=763$ indica que el 76% de la variación total de los promedios de evaluación del profesorado se explica por estas cinco variables que se usaron. Se puede concluir que el R^2 es significativo, ya que $F = 12,249, p \leq 0.001$

Sin embargo, los resultados estadísticos indican que hay una relación más significativa entre las variables *dominio del tema* y *evaluación de ejercicios y exámenes* con la variable *Calificación del Docente*.

Análisis factorial

El análisis factorial es una técnica de análisis multivariante que intenta identificar variables subyacentes o factores, que expliquen la configuración de las correlaciones dentro de un conjunto de variables observadas. Permite identificar factores subyacentes que explican las correlaciones entre un conjunto de variables. Se usa para resumir un gran número de variables con un número más pequeño de variables definidas, denominadas factores. *El análisis factorial se puede utilizar en la reducción de los datos para identificar un pequeño número de factores que explique la mayoría de la varianza observada en un número mayor de variables manifiestas, También puede utilizarse para generar hipótesis relacionadas con los mecanismo causales.*⁷⁵

El problema que se plantea es cómo distinguir la información relevante de la intrascendente; es decir, se le hace necesario dotarse de una herramienta que le permita separar el tronco de las ramas. Esta herramienta es el análisis factorial. Se trata de una técnica adecuada para el caso de variables continuas⁷⁶ altamente correlacionadas.

Ejemplo. Hay procesos culturales como estilo de vida, niveles de satisfacción con una situación u hecho, nivel socioeconómico, que son necesarios conocer, pero que no se pueden medir con una sola variable -pregunta, porque son procesos complejos que se manifiestan a través de una infinidad de situaciones, sentimientos, comportamientos y opiniones. Estas manifestaciones son el resultado de la medición de un conjunto de características, actitudes y comportamientos sociales. El análisis factorial permitirá combinar preguntas de manera que se pueda obtener nuevas variables o factores que no son directamente medibles pero que tienen un significado.

74 La regresión múltiple se simboliza con R cuadrado

75 Manual del usuario de SPSS Base 14.0. Capítulo 30, pág. 499

76 Variables cuantitativas

Tipos de procedimientos de reducción de datos⁷⁷

Análisis Conjunto. Analiza las características de un producto, idea o servicio, determina cuáles son las ideales y cuantifica las preferencias.

Análisis de Correspondencia. Analiza tablas de correspondencia (como tablas de contingencia) para medir mejor las distancias entre categorías o variables.

Análisis de Homogeneidad. Es un procedimiento de escalamiento óptimo análogo al análisis factorial, pero que analiza variables categóricas u ordinales. También se conoce como análisis de correspondencia múltiple.

Análisis No Lineal de Componentes Principales. Realiza análisis no lineales de componentes principales para reducir la dimensionalidad de un grupo de variables categóricas.

Análisis No Lineal de Correlación Canónica. Realiza un análisis no lineal de correlación canónica, para determinar cómo son los conjuntos de variables de uno respecto a otro.

A título de ejemplificación se muestran los objetivos, resultados y procedimientos SPSS® utilizados para determinar los perfiles de preferencias en la programación de televisión: *La Comisión Nacional de Televisión deseaba estudiar los perfiles de preferencias de programas televisivos. De gran importancia es el hecho de lograr identificar preferencias que ayuden a conformar perfiles sobre diferentes aspectos asociados al servicio televisivo. Por ejemplo: interesaba identificar qué géneros desde el punto de vista del usuario, conformaban conglomerados de preferencias, afinidades y demandas programáticas. A partir de la aplicación de técnicas de análisis multivariado se logró identificar por medio de la asociación de componentes principales unas dimensiones o conglomerados tomando como punto partida, los factores subjetivos subyacentes. De estas correlaciones se diseñaron unos perfiles de programación o segmentaciones sicográficas, que admiten conocer de manera más precisa quiénes son, donde viven, que edad tienen, cuáles son sus hábitos de consumo por géneros televisivos, etc., nuestros consumidores de televisión.*⁷⁸

77 Análisis factorial y Escalamiento óptimo.

78 Prácticas culturales y televisivas en el Área Metropolitana del Valle del Aburrá. Jaime Ruiz R., Miguel Aigner A. Efrén Barrera R. Centro de Estudios de Opinión de la Universidad de Antioquia –CEO- y Comisión Nacional de Televisión, Teleantioquia, TeleMedellín y Canal U. Medellín: 2004. Pág. 51 - 85

Para ello, se diseñó una encuesta donde se le pide a cada uno de los entrevistados que valore de uno (1) a cinco (5) si un tipo de programación es o no es muy preferida o muy importante para él. Los encuestados deberán contestar con un 5 si el programa es muy preferido, un 4 si es preferido, un 3 si tiene regular preferencia, un 2 si es poco importante y un 1 si no nada importante o no preferido. Los 15 tipos de programación son: concursos, deportivos, documentales, dramatizados, farándula – humorísticos, educativos, del hogar, infantiles, musicales, noticias, películas, (realitys), religiosos, telenovelas, variedades.

El procedimiento metodológico consiste realizar un análisis factorial que permita extraer unos factores adecuados a los datos que resuman correctamente la información que contienen y que permitan configurar unos perfiles de preferencias de los televidentes sobre la programación ofrecida.

Ejemplo: se trata de calcular los Componentes Principales de la información obtenida en el estudio de cultura y prácticas televisivas en el área metropolitana del Valle del Aburra.

- Procedimiento SPSS®

- Seleccionar en la Barra de Menú el submenú <Analizar>
- Activar el submenú <Reducción de datos> y luego la opción <Análisis factorial>

A continuación aparece el Cuadro de diálogo: Análisis factorial.



Figura 6.61 Cuadro de diálogo Análisis factorial

Fuente: SPSS®

Las variables para el análisis deben ser cuantitativas a nivel de intervalo o razón. Variables nominales u ordinales –categóricas– (como sexo o estado civil) no son adecuadas para el análisis factorial. En el caso del ejemplo anterior, las variables seleccionadas se consideran como cuantitativas porque provienen de la utilización de una escala Likert.

- Presione el botón <flecha> para transferir las 15 variables seleccionadas⁷⁹ a la ventana Variables

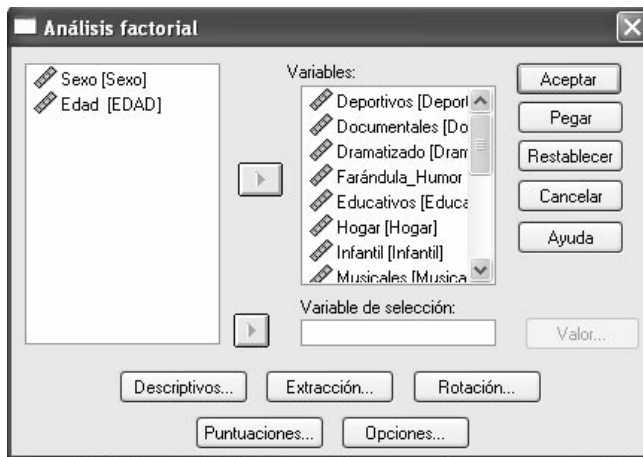


Figura 6.62 Cuadro de diálogo

Fuente: SPSS®

- Para seleccionar los estadísticos descriptivos presione el botón <Descriptivos> Aparece el siguiente Cuadro de diálogo: (veáse figura 6.63)
- Seleccione marcando la opción <Descriptivos univariados> que calcula los siguientes estadísticos: Media, Desviación estándar y el número de casos válidos para cada variable.
- En la ventanilla interna Matriz de correlaciones, seleccione las siguientes matrices de correlación:
- Seleccione marcando la opción <Coeficientes> que muestra la matriz de correlaciones para las variables seleccionadas por el análisis factorial.

⁷⁹ Seleccionar las 15 variables asociadas a *preferencia por género o tipos de programación* en la lista fuente de variables.



Figura 6.63 Cuadro de diálogo

Fuente: SPSS®

- Seleccione, marcando la opción <KMO and Bartlett's of Sphericity>. El Test de KMO señalará el nivel de correlación en el conjunto de variables seleccionadas y si el análisis factorial sería una técnica útil.
 - Presione el botón <Continuar>, para confirmar las opciones seleccionadas
- Para definir el método de extracción de factores presione el botón <Extracción>.

Aparece el siguiente Cuadro de diálogo.

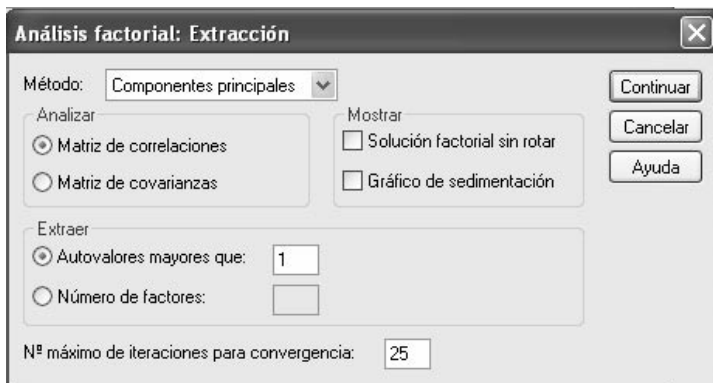


Figura 6.64 Cuadro de diálogo Análisis factorial: Extracción

Fuente: SPSS®

- Seleccione en Método, la opción <Componentes Principales> (ACP)⁸⁰
 - Presione el botón <Continuar>, para confirmar las opciones seleccionadas
- Para definir el método de configuraciones factoriales, presione el botón <Puntuaciones>, aparece el siguiente Cuadro de diálogo.

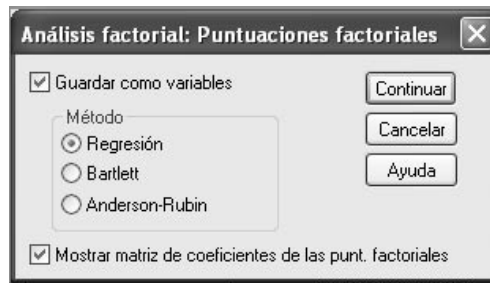


Figura 6.65 Cuadro de diálogo

Fuente: SPSS®

- Seleccione, marcando la opción, <Guardar como variables> que guarda las puntuaciones factoriales como nuevas variables (ver Vista de Datos), creándose una nueva variable para cada factor.
- En la Ventana <Método> seleccionar la opción <Regresión>

Para seleccionar el método de Configuraciones de Opciones, presione el botón <Opciones> y a continuación aparece el siguiente Cuadro de diálogo.

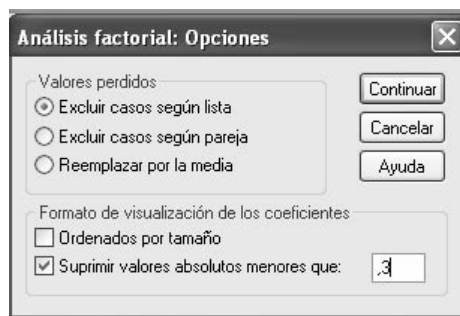


Figura 6.66 Cuadro de diálogo

Fuente: SPSS®

80 Modalidad de análisis factorial que permite la extracción de factores.

- En la ventanilla interna <Formato de visualización de los coeficientes>, seleccionar, marcando, <Suprimir valores absolutos menores que>. Este valor, por defecto es 0.10. Se recomienda cambiar al valor 0.3
- Presione el botón <Continuar>, para confirmar las opciones seleccionadas.

Vuelve a aparecer el Cuadro de diálogo Análisis de Factores completamente definido.

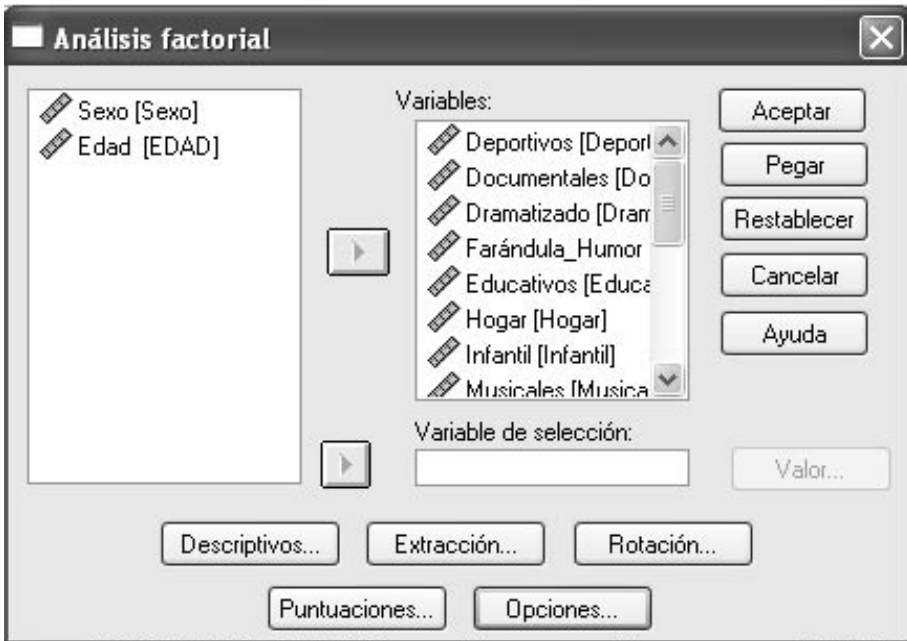


Figura 6.67 Cuadro de diálogo

- Proceda a presionar el Botón <Aceptar> para ejecutar el procedimiento de Análisis Factorial⁸¹.

81 El Análisis Factorial es una técnica que consiste en resumir la información contenida en una matriz de datos con V variables. Para ello se identifican un reducido número de factores F , siendo el número de factores menor que el de variables.

Tabla 6.29 Tabla de Estadísticos Descriptivos

	Media	Desviación estándar	Análisis N
Concursos	2,91	1,454	2234
Deportivos	3,19	1,620	2234
Documentales	3,41	1,515	2234
Dramatizado	2,74	1,415	2234
Farándula_Humor	3,49	1,433	2234
Educativos	3,46	1,449	2234
Hogar	2,72	1,523	2234
Infantil	2,95	1,625	2234
Musicales	3,63	1,441	2234
Noticias	3,82	1,396	2234
Películas	3,98	1,264	2234
Reality	3,16	1,595	2234
Religiosos	2,24	1,455	2234
Telenovelas	3,44	1,550	2234
Varietades	3,04	1,387	2234

Fuente: SPSS®

Análisis de los datos.

Interpretación de los resultados de un Análisis de Componentes Principales (ACP).

La tabla de Estadísticos descriptivos muestra homogeneidad en la escala de medida de las variables del ACP.

Tabla de Matriz de Componentes

La tabla siguiente muestra las comunalidades finales⁸², explicadas por los factores seleccionados. Aparecen 4 componentes finales y, en cada uno de ellos se destacan, por sus valores, las programaciones más relevantes.

82 Contribución o importancia de cada una de las variables a los componentes 1, 2, 3, y 4

Tabla 6.30 Matriz de Componentes

	Componente			
	1	2	3	4
Concursos	,181	-,002	-,012	,229
Deportivos	,029	,170	,430	,355
Documentales	,025	,351	,117	-,154
Dramatizado	,188	-,078	-,121	-,013
Farándula_Humor	,178	-,047	,117	,095
Educativos	,128	,316	,023	-,249
Hogar	,163	,185	-,295	-,166
Infantil	,130	-,006	,175	-,444
Musicales	,133	-,044	,266	-,250
Noticias	,016	,283	,002	,498
Películas	,105	-,069	,453	,013
Reality	,181	-,168	,007	,249
Religiosos	,126	,194	-,313	,116
Telenovelas	,180	-,246	-,142	,112
Variedades	,172	-,006	-,048	-,008

Extraction Method: Principal Component Analysis. Component Scores

Fuente: SPSS®

El estudio identificó 4 nichos de preferencias: Nicho 1⁸³ *Los misceláneos*, Nicho 2: *los culturales*, Nicho 3: *los Light* y Nicho 4: *los deportivos*.

La matriz factorial muestra las puntuaciones factoriales correspondientes a las variables y los factores. En este caso la delimitación de las variables (programación) que forman cada factor (perfil del televidente) es clara. Si se analizan las variables (programación) que componen el primer factor que une los Concursos, Dramatizados, Farándula, Películas, Reality y Telenovelas. El

83 Los nichos o perfiles están representados estadísticamente por los cuatro componentes.

segundo factor comprende las programaciones de tipo educativo y de noticias. Un tercero una los Deportivos, Infantiles y las películas. Finalmente un cuarto factor comprende básicamente la programación de Deportes y Noticias.

Tabla 6.32 Component Matrix(a)

	Component			
	1	2	3	4
Concursos	,601			
Deportivos		,335	,579	,396
Documentales		,691		
Dramatizado	,625			
Farándula _ humor	,593			
Educativos	,425	,622		
Hogar	,541	,364	-,398	
Infantil	,433			-,496
Musicales	,442		,359	
Noticias		,558		,556
Películas	,350		,610	
Reality	,600	-,332		
Religiosos	,421	,383	-,422	
Telenovelas	,599	-,484		
Variedades	,572			

Extraction Method: Principal Component Analysis: a 4 components extracted.

Tabla 6.33 Test de KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,791
Bartlett's Test of Sphericity	Approx. Chi-Square	5690,729
	df	9
	Sig.	,000

Fuente: SPSS®

Se observa que el estadístico KMO vale 0.791, valor muy cercano a la unidad, lo que indica una alta correlación entre los datos y una buena adecuación de los datos a un modelo de análisis factorial.

Glosario

A

Actitud. Reacción o predisposición favorable o desfavorable hacia algo o alguien, que se manifiesta en las creencias o en un comportamiento proyectado.

Aleatorio. Suceso o hecho producido al azar.

Amplitud o rango. Diferencia entre los valores mayor y menor de una variable numérica; el máximo menos el mínimo. También se denomina el recorrido de la variable

Análisis multivariante. Cualquier análisis que considere varias variables juntas.

Análisis de regresión. Técnica estadística multivariable que establece una relación funcional entre una variable dependiente y una serie de variables dependientes. Si sólo hay una variable independiente, se denomina *análisis simple*; si hay varias, es *múltiple*.

Análisis de conglomerados. (cluster). Técnica multivariable para identificar agrupaciones homogéneas de hechos, procesos o personas.

Análisis discriminante. Una técnica multivariable para analizar el valor predictivo de un conjunto de variables independiente.

Archivo. Conjunto de registros o datos relacionados entre sí, los cuales se pueden almacenar en una base de datos.

Asimetría. Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de asimetría igual a 0. Una distribución que tenga una asimetría positiva significativa tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga.

C

Categoría. Es una medida de una variable cuya escala de medición es de tipo nominal u ordinal.

Codificación de los datos. Consiste en asignar códigos alfanuméricos a las respuestas dadas a un cuestionario para poder efectuar el tratamiento estadístico de los datos.

Código. Convención para representar los valores en letras, números y caracteres especiales, con el fin de identificar características o categorías de respuestas.

Coficiente Alfa de Cronbach. Coeficiente que estima la consistencia interna de una escala de medición.

Comportamiento organizacional. Estudio de los individuos y de los grupos dentro del ámbito de la organización.

Curtosis. Medida del grado de concentración en que las observaciones están agrupadas en torno al punto central. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que las observaciones se concentran más y presentan colas más largas que las de una distribución normal. Una curtosis negativa indica que las observaciones se agrupan menos y presentan colas más cortas.

Confiabilidad. Grado en que una prueba proporciona resultados consistentes.

Contraste de hipótesis. Conjunto de procedimientos estadísticos que permiten decidir si una hipótesis se debe aceptar o rechazar.

D

Datos. Se entiende por dato un valor específico de una variable. Por ejemplo, 35 años es un dato de la variable edad. Valores alfanuméricos que procesa el computador para convertirlos en información.

Desviación estándar (Desviación típica). Es una medida de dispersión y/o alejamiento con respecto a la Media de los datos. Es una estimación de la cuantía en que se diferencian las observaciones respecto a la media aritmética. En una distribución normal, el 68% de los casos se encuentra dentro de

una desviación típica respecto a la media y el 95% de los casos se encuentra dentro de dos desviaciones típicas respecto a la media.

Dígito. Cada uno de los números que integran un dato numérico. Por ejemplo: 2003 tiene cuatro dígitos, o sea, cuatro números.

Digitación. Es el registro de los códigos de los valores de las categorías de las variables en un programa especializado para su posterior sistematización y análisis. Por ejemplo, digitación de los datos en el editor de datos del SPSS®.

Diagrama de causa y efecto. Diagrama usado para organizar y mostrar, en figuras o gráficos, las posibles causas de un problema.

Dispersión. Los estadísticos que miden la cantidad de variación o de dispersión en los datos, incluyen: Desviación típica, Varianza, Rango, Mínimo, Máximo y Error típico de la media.

E

Editor de datos SPSS (Data View). Es una matriz conformada por filas y columnas similares a una hoja de cálculo y permite configurar, introducir, editar y presentar datos.

Efecto de halo. Situación que se da cuando el entrevistador permite que una característica u opinión particular contamine la opinión sobre los demás rasgos —respuestas que se esperan.

Encuesta piloto. Aplicación de un cuestionario de prueba a una submuestra con el objetivo de identificar problemas con las preguntas y su nivel de respuesta.

Encuesta de panel. Muestra permanente de unidades de análisis de los cuales se obtiene información periódicamente.

Escala: Una pregunta cerrada para medir actitudes.

Escala de razones. Una escala con intervalos medibles entre sus puntos.

Escala nominal. Una escala que clasifica a las personas en grupos que no tienen un orden numérico. Ejemplo: Femenino - Masculino

Escala ordinal. Una escala que clasifica a las personas en grupos que tienen un orden inherente.

Estándar. Unidad de medida adoptada y aceptada comúnmente como criterio para realizar algún proceso.

Estimación estadística. Procedimiento estadístico que trata de determinar las características de una población a partir de las características obtenidas de una muestra representativa.

E. T. media. Medida de cuánto puede variar el valor de la media de una muestra a otra, extraídas éstas de la misma distribución.

F

Formas de distribución. Asimetría y curtosis son estadísticos que describen la forma y la simetría de la distribución.

G

Grabación de datos. Almacenamiento de los datos en un medio magnético.

Grupo de control. Grupo de unidades de análisis o sujetos que en un estudio experimental no son expuestos a ninguno de los tratamientos cuyos efectos tratan de medirse.

I

Inferencia estadística. Proceso estadístico inductivo que trata de obtener información sobre las características de una población a partir de los datos obtenidos en una muestra.

M

Mínimo. Valor más pequeño de una variable numérica.

Máximo. El mayor valor de una variable.

Muestra. Una proporción de la población seleccionada para un estudio.

Muestra probabilística. Una muestra en la que toda unidad tiene una probabilidad igual y conocida de ser seleccionada. Llamada también muestra al azar.

Muestra intencional o de conveniencia. Se eligen aquellas unidades de análisis que de acuerdo con el investigador resuman las características deseadas o a su alcance.

Manova. Modelo para analizar la relación entre dos o más variables independientes y dos o más variables dependientes.

Media. Una medida de tendencia central. La media aritmética con la cual se calcula el promedio aritmético; la suma dividida por el número de casos.

Mediana. Es una medida que divide a cualquier distribución de datos en dos (2) partes iguales por la mitad y sirve para medir el dato o punto que está en el centro de la distribución.

Moda. Es el valor que ocurre más frecuentemente, que se repite con más frecuencia en un conjunto de datos.

P

Población. Conjunto de todas las unidades, del cuales saca una muestra. También llamado Universo.

R

Regresión lineal. Es un modelo matemático para estimar el efecto de una variable sobre otra.

Regresión múltiple. Técnica estadística que estudia la relación lineal entre una variable dependiente (criterio) y una o más variables independientes (predictoras); el nivel de medida de las variables tiene que ser de intervalo o razón.

S

Submuestra. Conjunto de unidades de análisis —sujetos u objetos— obtenidos de una muestra, de acuerdo a un criterio determinado, para efectuar análisis específicos en ella.

T

Tabla de contingencia. Tabla de doble entrada en la que se presentan los valores de las frecuencias conjuntas de dos o más variables. En el lenguaje operativo del SPSS® se le conocen como Tablas de pivote.

Tabulación cruzada. Técnica estadística que permite calcular la distribución de las frecuencias de una variable para cada una de las categorías de la otra variable. Llamada también cruce de variables.

Tendencia central. Los estadísticos que describen la localización de la distribución, incluyen: Media, Mediana, Moda y Suma de todos los valores.

Tipos de variable. En el programa SPSS®, en la matriz Vista de variables (Variable View), en la columna Medición (measures) hay que definir si la variable es nominal, ordinal o de Escala o continua.

Transformaciones de los datos. Las funciones de transformación del SPSS® permiten preparar los datos para el análisis. Puede crear subconjuntos de datos, combinar categorías, añadir, agregar, funcionar, segmentar y transponer archivos.

U

Unidad de análisis. Es el sujeto – objeto del estudio –. Por ejemplo, la aplicación de un cuestionario a un grupo de 205 estudiantes implica la existencia de 205 casos o unidades de análisis.

V

Validación de los datos. Es la inspección de las respuestas de los cuestionarios con el fin de asegurar que estén correctamente contestadas y que las respuestas sean pertinentes. En el programa SPSS®, se puede utilizar la herramienta <Frecuencias> para validar códigos, datos perdidos y también la herramienta cruce de variables (crosstabs) para ver la coherencia en las respuestas.

Validez. El grado en que una prueba mide lo que se pretende que mida.

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. Los cuartiles (los percentiles 25, 50 y 75) dividen las observaciones en cuatro grupos de igual tamaño. Si desea un número igual de grupos que no sea cuatro, seleccione <Puntos de corte para *n* grupos iguales>. También puede especificar

percentiles individuales (por ejemplo, el percentil 95, el valor por debajo del cual se encuentran el 95% de las observaciones).

Variable. Cualquier característica, propiedad observable que puede variar. En relación con los procedimientos SPSS® es la característica que varía y que se mide u observa de un elemento, una muestra o la población; ejemplos: edad, escolaridad, estado civil, etc.

Variable independiente. Aquella variable que permite explicar, determinar, anteceder otra variable dependiente. También llamadas variable predictiva.

Variable dependiente o explicada. Es la variable explicada o determinada por la variable independiente.

Variable ordinal. Aquella variable de tipo cualitativo que mediante códigos se identifican categorías o tipos de respuesta.

Variables nominales. Aquella variable de tipo cualitativo que mediante códigos se identifican categorías o tipos de respuesta.

Variable continua. Variable cuantitativa que puede tomar cualquier valor dentro de un intervalo.

Variable dicotómica. Es la que únicamente tiene dos opciones de respuesta.

Variable discreta. Aquella variable que dispone de más de dos opciones de respuesta.

Varianza. Medida de dispersión en torno a la media. Es una estimación de la cuantía en que se diferencian las observaciones respecto a la media, es igual al cuadrado de la desviación típica.

Validez. El grado en que las investigaciones miden de hecho lo que dicen estar cuantificando. Grado en que una prueba mide lo que se pretende que mida.

Bibliografía

- Aigner, Miguel, *Investigación cuantitativa en Ciencias Sociales. Procesamiento y análisis de datos*. Universidad de Antioquia, Facultad de Ciencias Sociales y Humanas, Centro Estudios de Opinión. Medellín: Fondo Editorial CEO. 1997.
- _____, *Diseño y análisis de la información cuantificable en la investigación social*. Universidad de Antioquia, Facultad de Ciencias Sociales y Humanas, Centro Estudios de Opinión. Medellín: Fondo Editorial CEO. 1999.
- _____, *El análisis del dato tabulado*. Medellín. Universidad de Antioquia. Facultad de Ciencias Sociales y Humanas, Departamento de Sociología. Medellín: Fondo Editorial CEO. 1991.
- Briones, Guillermo, *Métodos y técnicas de investigación aplicada a la educación y a las ciencias sociales. Curso avanzado de técnicas de investigación social aplicadas a la educación Módulo 2. Tipos de investigaciones y diseños metodológicos*. Bogotá: Corporación Editorial Universitaria. 1988.
- _____, *Métodos y técnicas de investigación aplicada a la educación y a las ciencias sociales. Curso avanzado de técnicas de investigación social aplicadas a la educación. Módulo 2. Técnicas e instrumentos para la recolección de información. Programa Interdisciplinario de investigaciones en educación –PIIE- Instituto Colombiano para el Fomento de la Educación Superior – ICFES- Santiago de Chile: 1988.*
- _____, *Métodos y técnicas avanzadas de investigación aplicadas a la educación y a las ciencias sociales. Curso avanzado de técnicas de investigación social aplicadas a la educación. Módulo 1. Epistemología y metodología de la investigación social. Programa Interdisciplinario de investigaciones en educación –PIIE- Instituto Colombiano para el Fomento de la Educación Superior –ICFES-. Bogotá: 1988.*
- Briones, Guillermo, *Métodos y técnicas avanzadas de investigación aplicadas a la educación y a las ciencias sociales. Módulo 4. Análisis e interpretación de datos cuantitativos*. Bogotá: Corporación Editorial Universitaria de Colombia. 1989.
- Briones, Guillermo, *Evaluación interactiva. Serie Métodos y técnicas avanzadas de investigación social aplicadas a la educación y las ciencias sociales*. Bogotá: Procesos Editoriales ICFES. 1989.
- _____, *Métodos y técnicas de investigación para las ciencias sociales*. Editorial Trillas. México. 1990.
- Barbbie, Earl, *The Practice of Social Research*. Belmont, California: Wadsworth Publishing Co. 1979.

- Blalock, Hubert, Jr., y Blalock B. Ann, *Methodology in social research*. Nueva York: McGraw-Hill Co. 1978.
- Camacho, G. Álvaro y Guzmán B. Álvaro, Colombia. Ciudad y violencia. Bogotá: Ediciones Foro Nacional por Colombia. 1990.
- Camacho Rosales Juan, *Estadística con SPSS para Windows*. Madrid: RA-MA Editorial. 2002.
- Escalante, A. Carlos, *La medición de las actitudes*. Colección Ciencias Sociales. Teorías y Métodos. Bogotá: Ediciones Tercer Mundo. 1983.
- _____, *Serie aprender a investigar. El problema y la hipótesis. Módulo 2*. Instituto Colombiano para el Fomento de la Educación Superior. Bogotá: Procesos Editoriales ICFES. Bogotá: 1987.
- Facundo D., Ángel, *Serie aprender a investigar. Módulo 5. El proyecto de investigación*. Instituto Colombiano para el Fomento de la Educación Superior. Bogotá: Procesos Editoriales ICFES. Bogotá: 1987.
- Ghiglione, Rodolfo, *Las encuestas sociológicas, teoría y práctica*. México: Editorial Trillas. 1989.
- Henríquez A. Guillermo, et. al., *El grupo virtual de discusión, Una alternativa de investigación social en el ciberespacio*. Universidad de Concepción. Facultad de Ciencias Sociales. Departamento de Sociología. Concepción, Chile: 1991.
- Hernández, S. Roberto, Carlos Fernández C y Pilar Baptista L. *Metodología de la investigación*. México: Editorial McGraw - Hill. 1991.
- Hitchens, Christopher; Cándido, Monzón Comp. *Opinión Pública, comunicación y política. Las encuestas electorales*. Madrid: Tecnos. 1996. Pág. 351.
- Hitchens, Christopher, *Encuestas: ¿las malas del paseo? El Tiempo, suplemento Dominical*, Bogotá, 1992.
- Korman Hyman, *Técnicas para investigar la opinión pública*. Department of Sociology. New York: State University of New York at Stony Brook. 1984.
- Mayor, Alberto, *Serie aprender a investigar. La recolección de la información*. Instituto Colombiano para el Fomento de la Educación. Bogotá: Procesos Editoriales ICFES. Bogotá: 1987.
- Noelle, Elizabeth, *Encuestas en la sociedad de masas*, Madrid: Alianza Editorial. 1970.
- Puerta Zapata, Antonio, *Evaluación y sistematización de proyectos sociales: una metodología de la investigación*. Universidad de Antioquia, Facultad de Ciencias Sociales, Medellín: 1996.
- Restrepo, M. Piedad y Aigner A. Miguel, *La evaluación participativa de programas de intervención social y física*. Universidad de Antioquia, Facultad de Ciencias Sociales y Humanas, Centro Estudios de Opinión, Medellín: Fondo Editorial CEO. Medellín 2004.
- Rodríguez, Humberto, *Serie aprender a investigar. Modulo 6. El informe de investigación*. Instituto Colombiano para el Fomento de la Educación. Bogotá: Procesos Editoriales ICFES, Bogotá: 1987.
- Rokeach, Milton, *The Nature of Human Values*. New York: Free Press. 1973.

- Rojas José María y Castillo Luís, Poder local y recomposición campesina. Universidad del Valle. Fondo DRI – CIDSE Universidad del Valle. Cali: CIDSE 1991.
- Rojas José María, La bipolaridad del poder local. Caldosos en el Cauca Indígena. Universidad del Valle. Colección Edición Previa. Cali: Centro Editorial. 1993.
- Rojas, B. Mauricio. Metodología de la investigación. Corporación Universitaria de Ibagué. Centro de Investigaciones, consultoría y extensión a la comunidad. Ibagué: 1999.
- Ruiz, R. Jaime, La teoría social de la ciudad. Elementos para un marco conceptual y de referencia para la investigación urbana. Universidad de Antioquia, Facultad de Ciencias Sociales y Humanas, Centro Estudios de Opinión. Medellín: Fondo Editorial CEO. 1997.
- Ruiz, Jaime; Serna, Alba Lucía y Aigner, Miguel, El Medellín que yo quiero, Medellín. Universidad de Antioquia, Facultad de Ciencias Sociales y Humanas. Centro de Estudios de Opinión, Concejo de Medellín. Medellín: 1991.
- Ruiz, Jaime, Aigner, Miguel, Ochoa, Jaime; De Los Ríos, Héctor; Parra, Carlos, Imágenes de identidad, integración y conflicto entre la comunidad estudiantil de la Universidad de Antioquia, Medellín. Universidad de Antioquia. Facultad de Ciencias Sociales y Humanas. Centro de Estudios de Opinión. Medellín: Editorial Universitaria 1991.
- Ruiz R. Jaime, Vélez C. Beatriz. Medellín. Fronteras invisibles y espacios de guerra, Universidad de Antioquia, Facultad de Ciencias Sociales y Humanas, Centro Estudios de Opinión. Medellín: Fondo Editorial CEO. 2004
- Serna A., Alba Lucía, et al., Módulo 3 Las técnicas de recolección de información. Instituto de Estudios Regionales e Instituto para el Fomento de la Educación Superior. Bogotá: Procesos Editoriales ICFES. Bogotá: 1992.
- Siegel, Sydney, Estadística no-paramétrica aplicada a las ciencias de la conducta, México: Editorial Trillas. 1978.
- Siegel Sydney, Non-parametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Co. 1978.
- Tirón, Eugenio, Pobladores e integración social. Chile. Propositiones 14. Santiago de Chile: Ediciones Sur. Agosto 1997
- Vélez B., Eduardo, Serie aprender a investigar. Módulo 4. El análisis de la información. Instituto para el Fomento de la Educación. Bogotá: Procesos Editoriales ICFES. Bogotá: 1987.
- Webb, Ken.; Hatry, Harry, Handbook of Research Design and Social Measurement. New York: David McKay Company. Inc. 1984.

Anexos: Archivos de trabajo

Información sobre los archivos de trabajo

Cada archivo de trabajo utilizado en este texto, es presentado en una matriz de datos en este Anexo. Estos datos deben ser *traducidos* mediante el Editor de Datos de SPSS¹, el cual proporciona, similar a una hoja de cálculo rectangular, una forma muy amigable de crear y editar un archivo de trabajo.

Un archivo de trabajo SPSS® contiene información descriptiva, como:

- Nombre de las variables.
- Formato de las variables.
- Etiquetas descriptivas de las variables y de las categorías de cada una de ellas.
- Tipos de variables.

La estructura de un archivo de trabajo SPSS® esta compuesta por:

Las filas que son los casos.

Las columnas son las variables o característica que se esta midiendo. Es decir, cada pregunta o ítem de un cuestionario es una variable.

Las celdas contienen los valores únicos que toman las variables.

Nota metodológica. Cada uno de los archivos de trabajo SPSS® referenciados en el texto debe ser creado y editado a partir de la información contenida en las matrices de datos presentada en este Anexo.

1 La ventana Editor de Datos SPSS® se abre automáticamente cuando se inicia una sesión de trabajo con el programa SPSS®.

Archivo de trabajo: Hábitos-2.sav

Descripción: Hábitos o modalidades de estudios de los estudiantes de la Universidad de Antioquia según programa académico.

Objetivo. Analizar la información del estudio “Hábitos de estudios de los estudiantes de las Facultades de Ciencias Sociales y Humanas y de Medicina Veterinaria y Zootecnia”. Se ha hipotetizado que uno de los factores que incide en el rendimiento académico es la modalidad de estudio que utilizan los estudiantes universitarios. Al respecto han surgido hipótesis de trabajo asociadas a la modalidad de estudio: para algunos especialistas existe una estrecha relación entre tipo de estudiante – según la facultad en que estén registrados – y la modalidad de estudio.

Una de las preguntas de investigación es si hay diferencias de hábitos o modalidades de estudio entre los estudiantes de Facultades de Ciencias Sociales y Humanas y los de Medicina Veterinaria y Zootecnia.

Metodología. Mediante la técnica de la encuesta social, se diseñó y aplicó un cuestionario a 86 estudiantes seleccionados por un diseño muestral: 43 estudiantes de la Facultad de Ciencias Sociales y Humanas y otros 43 estudiantes de la Facultad de Medicina Veterinaria y Zootecnia, pidiéndoles que indicarán si su método de estudio correspondía a alguno de los tres siguientes: Estudio sistemático diario; Estudio intenso en períodos irregulares y modalidad de estudio combinado –diario e intenso_

Hoja de codificación

	1	2
1	1 1	2 11
2	1	11
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	2
9	1	2

	1	2
47	2	1
48	2	1
49	2	1
50	2	1
51	2	1
52	2	1
53	2	1
54	2	1

Diseños cuantitativos: análisis e interpretación de la información

10	1	2
11	1	2
12	1	2
13	1	2
14	1	2
15	1	2
16	1	2
17	1	2
18	1	2
19	1	2
20	1	2
21	1	2
22	1	3
23	1	3
24	1	3
25	1	3
26	1	3
27	1	3
28	1	3
29	1	3
30	1	3
31	1	3
32	1	3
33	1	3
34	1	3
35	1	3
36	1	3
37	1	3
38	1	3
39	1	3
40	1	3
41	1	3
42	1	3
43	1	3
44	1	3
45	2	1
46	2	1

55	2	2
56	2	2
57	2	2
58	2	2
59	2	2
60	2	2
61	2	2
62	2	2
63	2	3
64	2	3
65	2	3
66	2	3
67	2	3
68	2	3
69	2	3
70	2	3
71	2	3
72	2	3
73	2	3
74	2	3
75	2	3
76	2	3
77	2	3
78	2	3
79	2	3
80	2	3
82	2	3
83	2	3
84	2	3
85	2	3
86	2	3

Variable 1: Tipo de Estudiante= (1) Ciencias Sociales y Humanas: (2) Medicina Veterinaria y Zootecnia.

Variable 2: Modalidad de Estudio= (1) Estudio Sistemático Diario. (2) Periodos irregulares de intenso estudio (3) combinación de los dos tipos de estudio.

Vista de variables

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	Tipo	N Numérico	2	0	Tipo Estudiante	{1, ciencias sociales y	Ninguno	8	Derecha	Nominal
2	Estudio	N Numérico	2	0	Modalidad estudio	{1, estudio sistemático	Ninguno	8	Derecha	Nominal
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										

Fuente: SPSS®

Archivo de trabajo: Temor ciudadano.sav

Descripción. En un sondeo de opinión la inseguridad, un miedo urbano, se llega a la conclusión que la inseguridad es un miedo urbano que permeó todas las esferas. 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. *Es verdad, no sólo se logra con infraestructura, pero sí se necesita la presencia del Estado para que los delitos tengan control.*

Objetivo. Con los datos aportados por el estudio mencionado realizar un primer diagnóstico sobre las apreciaciones de los medellinenses acerca de la inseguridad.

Metodología. Proceder a crear y configurar una base de datos que servirá como archivo de trabajo para desarrollar el diagnóstico.

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 10.7% a los estratos 5 y 6. El sondeo se realizó los días 26, 27 y 28 de septiembre y tiene una margen de error de 4%.

A continuación se le presenta una Hoja de Codificación con la información del sondeo de opinión titulado *la Inseguridad, un miedo urbano*.

Hoja de codificación

Temor ciudadano.sav

	Sexo	Etrato	Seguro	Robos	Terrorismo	Secuestro	Indigencia	Violación	Homicidio	Extorsión	Violencia	Tránsito	Fuerzas	Cultura	Colaboración	Control	Desarme	Civil	SEXO (sexo) 1: Femenino ; 2: Masculino
1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Estratificación socioeconómica (estrato)
2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Se siente seguro en Medellín? (seguro): SI (1); NO (2)
3	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	De los siguientes delitos, señale cuál le hace sentirse más inseguro?
4	2	3	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	Robos (robos) - SI (1); NO (2)
5	1	3	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	Terrorismos (terrorismo)
6	2	3	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	Secuestro (secuestro)
7	1	5	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	2	Indigencia (indigencia)

20	1	4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Desarmar a la ciudadanía (desarme)- SI (1); NO(2)
21	1	5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Implementar alarmas comunitarias y seguridad (civil)- SI (1); NO(2)
22	2	5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
23	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
24	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
25	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
26	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
27	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
28	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	

Vista de variables

*Temor ciudadano [Conjunto_de_datos1] - Editor de datos SPSS										
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?										
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	sexo	Númérico	8	0	Sexo	{1, femenino}	Ninguno	8	Derecha	Nominal
2	estrato	Númérico	8	0	Estratificación	{1, Bajo.bajo}	Ninguno	8	Derecha	Nominal
3	seguro	Númérico	8	0	¿Se siente seguro en Medellín?	{1, Si}...	Ninguno	8	Derecha	Escala
4	robos	Númérico	8	0	Robos	{1, Si}...	Ninguno	8	Derecha	Escala
5	terror	Númérico	8	0	Terrorismo	{1, Si}...	Ninguno	8	Derecha	Escala
6	secuestr	Númérico	8	0	Secuestros	{1, Si}...	Ninguno	8	Derecha	Escala
7	indigenc	Númérico	8	0	Indigencia	{1, Si}...	Ninguno	8	Derecha	Escala
8	violat	Númérico	8	0	Violaciones	{1, Si}...	Ninguno	8	Derecha	Escala
9	homicidi	Númérico	8	0	Homicidios	{1, No}...	Ninguno	8	Derecha	Escala
10	estorsio	Númérico	8	0	Extorsiones	{1, Si}...	Ninguno	8	Derecha	Escala
11	violenci	Númérico	8	0	Violencia extrafamiliar	{1, Si}...	Ninguno	8	Derecha	Escala
12	transito	Númérico	8	0	Peligro de accidente de tránsito	{1, Si}...	Ninguno	8	Derecha	Escala
13	fuerzas	Númérico	8	0	Aumentar pie de fuerza policial	{1, Si}...	Ninguno	8	Derecha	Escala
14	culturaz	Númérico	8	0	Implementar programas de cultura ciudadana	{1, Si}...	Ninguno	8	Derecha	Escala
15	colabora	Númérico	8	0	Fomentar la colaboración ciudadana	{1, Si}...	Ninguno	8	Derecha	Escala
16	control	Númérico	8	0	Aumentar los controles callejeros	{1, Si}...	Ninguno	8	Derecha	Escala
17	desarme	Númérico	8	0	Desarmar la ciudadanía	{1, Si}...	Ninguno	8	Derecha	Escala
18	civil	Númérico	8	0	Alarmas comunitarias de seguridad	{1, Si}...	Ninguno	8	Derecha	Escala

Fuente: SPSS®

Archivo de trabajo: Ingresos-Sena. sav

Descripción. El impacto de la capacitación laboral tecnológica y los niveles de ingresos de los egresados de los cursos de nuevas tecnologías informáticas.

Objetivo. Analizar la información del estudio *Capacitación laboral tecnológica y niveles de ingresos de los egresados de los cursos de nuevas tecnologías informáticas (cursos de Office, Básico de Internet, Front Page, usos educativos de Internet)*. Describir conceptualmente las tendencias y análisis estadísticos que son posibles observar en la información disponible.

Metodología. Mediante la técnica de la entrevista social, se diseñó y aplicó una guía de entrevista a un grupo de 20 egresados de los cursos de nuevas tecnologías informáticas de una institución tecnológica. Con la información registrada y codificada creó un archivo de trabajo. La matriz de datos del archivo se configuró de acuerdo a la siguiente estructura:

Nombre de la variable	Definición de la variable
Sexo	Sexo observado.
Ingresos	Ingresos declarados (€) por los egresados.
Horas	Horas de capacitación recibida en técnicas informáticas.

Escolar	Número de años de escolaridad alcanzados en la educación formal.
Edad2001	Edad recodificada en grupos etarios: entre 20 y 23 años; entre 24 y 26 años
Edad	Edad en años cumplidos.
Estrato	Estratificación socioeconómica según el tipo de vivienda en que reside.

Hoja de codificación

Sexo	Ingresos	Horas	Escolar	Edad2001	Edad	Estrato
1	240	12	15	2	25	3
1	440	20	15	2	25	3
1	720	24	11	1	21	3
2	960	28	10	1	20	4
2	1200	32	11	2	21	4
2	1300	36	12	1	22	4
2	1200	44	11	1	21	4
2	1300	44	11	1	20	4
1	240	12	14	2	24	3
1	440	20	14	2	24	3
2	720	28	11	1	20	4
2	1200	32	11	1	20	4
2	1300	40	15	1	20	4
1	200	12	15	2	25	3
2	400	20	15	2	24	3
1	500	24	15	2	24	3
1	900	28	11	1	20	4
2	1000	32	11	1	20	4
2	1300	40	10	1	20	4
1	1000	36	11	1	21	4

Vista de variables

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	Sexo	Numérico	8	0	Género	{1, Femenino}...	Ninguno	8	Derecha	Nominal
2	Ingresos	Numérico	8	0	Ingresos percibidos	Ninguno	Ninguno	8	Derecha	Escala
3	Horas	Numérico	8	0	Horas de capacitación	Ninguno	Ninguno	8	Derecha	Escala
4	Escolar	Numérico	8	0	Nivel de escolaridad	Ninguno	Ninguno	8	Derecha	Escala
5	Edad2001	Numérico	1	0	Grupos de edades	{1, menores de 23}	Ninguno	8	Derecha	Nominal
6	Edad	Numérico	8	0	Edad en años cumplidos	Ninguno	Ninguno	8	Derecha	Escala
7	Estrato	Numérico	8	0	Nivel de estratificación	{1, 1}...	Ninguno	8	Derecha	Nominal
8										
9										
10										
11										
12										

Fuente: SPSS®

Archivo de trabajo. Situación-2002.sav

Descripción. La apreciación de los colombianos sobre la situación económica y política que vive el país. El problema de la crisis política y económica se ha instalado en la agenda política en Colombia, en los medios de comunicación y en el ánimo de las personas. A este cuadro se agrega la aparición y el desarrollo de un nuevo fenómeno: la sensación de inseguridad política, económica y ciudadana en el común de la gente. Un fantasma generalizado de precariedad sobrevuela al común de la gente, motivado por la incertidumbre laboral y económica.

Objetivo. El sondeo de opinión trató de medir esa sensación de inseguridad y vulnerabilidad, utilizando un cuestionario con preguntas que inquirían sobre los factores que están asociados a la crisis, como se resolvería, él por que del desinterés de los ciudadanos, quién son los responsables y quienes serían los llamados a solucionar la situación.

Metodología. Ficha técnica. Fecha de recolección: del 16 al 22 de mayo de 2002. Universo. Personas mayores de 18 años residentes en los barrios seleccionados. Metodología de muestreo: muestra aleatoria estratificada en hogares. Tamaño de la muestra: se realizaron 362 entrevistas telefónicas, en hogares de Medellín, por estratos socioeconómicos. La muestra tomada garantiza validez estadística para el universo con un 4% aproximado de margen de error y una confiabilidad del 95%. La población total de Medellín esta considerada en 2.026.798 habitantes, el 45.8% de hombres y el 54.2% mujeres, repartidos en 16 comunas y 5 corregimientos.

Estructura de la matriz de datos

Pregunta	Nombre variable	Categoría respuesta	Código
1	SEXO	Femenino	1
		Masculino	2
2	EDAD	15-19	1
		20-29	2
		30-39	3
		40-49	4
		Mayor de 50	5
3	ESTRATO	Bajo	2
		Medio bajo	3
		Medio	4
		Medio alto	5
		Alto	6
4	CAUSA	Globalización	1
		Alto costo	2
		Dineros narcotráfico	3
		Impunidad	4
		Corrupción	5
		Ingovernabilidad	6
		Guerra interna	7
		Inseguridad	8
		Apertura	9
		No Sabe	10
		No hay situación difícil	11
5	EVALUAR	Malo	1
		Deficiente	2
		Regular	3
		Bueno	4
		Excelente	5

6	RESOLVER		
		Gobierno autoritario	1
		Emergencia económica	2
		Cierre del Congreso	3
		Dialogo con los actores armados	4
		Derogatoria Congreso	5
		Cierre apertura económica	6
		Nuevo pacto social	7
		Alianza empresarios gobierno	8
		Gobierno cívico militar	9
		No sabe	10
		No hay situación difícil	11
7	APATIA		
		No hay apatía o desinterés	1
		Manipulación	2
		Falta compromiso	3
		Carencia líderes	4
		Problemas cotidianos	5
		Conflictos sin fin	6
		Problemas ajenos al ciudadano	7
		Falta cultura ciudadana	8
		Problema de gremios	9
		Escepticismo	10
		No sabe	11
8	RESPONSABLES		
		Los gremios empresariales	1
		El gobierno	2
		Sector financiero	3
		Medios comunicación	4
		Incertidumbre económica	5
		Los políticos	6
		La guerrilla	7
		Dineros del narcotráfico	8
		Paramilitarismo	9
		Escasa participación ciudadana	10
		No sabe	11
		No existe situación difícil	12
		Otros	13

9	SOLUCIÓN	Partidos políticos Ciudadanos Sector económico El gobierno Medios de comunicación Personalidades Poder judicial Organismos internacionales ONG Organizaciones armadas No sabe Otra No existe situación difícil	1 2 3 4 5 6 7 8 9 10 11 12 13
----------	-----------------	--	---

10 GUERRA

MD² (1) D(2) NO SABE (3) A(4) MA³ (5)

11 TREGUA

MD (1) D(2) NO SABE (3) A(4) MA (5)

12 INGRESOS

250.000	306.000	bajos
306.001	999.999	medios -bajos
1.000.000	2.000.000	medios
2.000.001	4.000.000	medios altos

2 Muy en desacuerdo
 3 Muy de acuerdo

Hoja de codificación

SEXO	EDAD	ESTRATO	CAUSA	EVALUAR	RESOLV	APATIA	RESPON	SOLUC	ESCOLAR	INGR.	GUERRA	TREGUA
2	5	3	4,0	2	9	9	1	3	13	890000	1	5
1	2	2	2,0	4	6	3	7	3	16	2300000	2	4
1	5	3	5,0	4	11	9	1	1	13	609000	3	3
2	5	3	6,0	2	3	2	2	6	8	306000	4	2
2	2	3	9,0	2	3	4	12	2	11	406000	5	1
1	1	2	2,0	3	3	7	2	8	6	203000	1	5
1	2	5	11,0	4	8	7	9	2	13	1250000	2	2
2	1	3	1,0	2	11	10	6	3	11	609000	3	3
2	1	3	5,0	2	6	6	2	2	13	610000	4	2
2	2	4	11,0	4	8	5	12	2	11	1000000	5	1
1	1	3	4,0	3	11	1	9	10	13	306000	1	5
2	2	3	5,0	4	8	2	1	2	11	609000	2	4
2	2	3	4,0	1	7	5	2	4	11	609000	3	5
1	2	2	11,0	2	9	7	6	3	13	610000	4	2
2	4	4	5,0	4	7	7	12	2	13	1222000	5	1
2	4	2	5,0	3	5	9	6	13	15	1019000	3	5
1	4	3	4,0	3	6	7	12	13	13	812000	3	5
1	1	2	4,0	5	1	7	8	1	11	812000	1	5
2	2	4	4,0	4	10	3	6	8	13	1000000	2	2
1	3	5	4,0	2	1	9	6	7	15	1650000	1	5
2	2	3	2,0	1	11	2	8	2	11	812000	2	4
1	1	4	6,0	3	6	7	6	6	11	812000	2	4
1	3	4	1,0	4	5	9	7	4	15	1000000	2	4
1	3	6	7,0	4	1	1	5	4	18	3500000	5	5
2	4	6	7,0	4	1	1	7	4	18	4000000	5	5

Vista de variables

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	Sexo	Numérico	2	0	Sexo	{1, femenino}...	Ninguno	8	Derecha	Nominal
2	Edad	Numérico	8	0	Edad	{1, <20}...	Ninguno	8	Derecha	Escala
3	Estrato	Numérico	8	0	Estrato socioeconómico	{2, 2}...	Ninguno	8	Derecha	Nominal
4	Causa	Numérico	8	0	Causa de la situación que vive el país	{1, globalizacio	Ninguno	8	Derecha	Nominal
5	Evaluar	Numérico	8	0	Evaluación del Gobierno	{1, malo}...	Ninguno	8	Derecha	Escala
6	Resolver	Numérico	8	0	Crisis se resuelve si...	{1, gobierno au	Ninguno	8	Derecha	Nominal
7	Apatía	Numérico	8	0	Apatía ciudadana	{1, no hay desi	Ninguno	8	Derecha	Nominal
8	Responsa	Numérico	8	0	Responsables	{1, gremios}...	Ninguno	8	Derecha	Nominal
9	Solucion	Numérico	8	0	Soluciones	{1, partidos}...	Ninguno	8	Derecha	Nominal
10	Escolar	Numérico	2	0	Nivel de escolaridad	Ninguno	Ninguno	8	Derecha	Nominal
11	Ingresos	Numérico	8	0	Nivel de ingresos	Ninguno	Ninguno	8	Derecha	Nominal
12	Guerra	Numérico	1	0	Diálogos en medio de la guerra	{1, muy en des	Ninguno	6	Derecha	Ordinal
13	Tregua	Numérico	1	0	Diálogos en medio de la tregua	{1, muy en des	Ninguno	6	Derecha	Ordinal
14										
15										

Fuente: SPSS®

Archivo de trabajo: Recordación.sav

Descripción. 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.

Objetivo. Comprobar si existe diferencia significativa entre las medias obtenidas por ambos grupos de lectores.

H_0 : 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).

$$H_0: \text{Puntaje Grupo 1} = \text{Puntaje Grupo 2}$$

H_1 : existen diferencias estadísticamente significativas de puntaje entre los sujetos del grupo 1 (que leyeron un texto sencillo) y los del grupo 2 (texto complejo).

$$H_1: \text{Puntaje Grupo 1 es diferente al Puntaje Grupo 2}$$

Hoja de codificación

Tabla: N° de palabras recordadas de un texto sencillo y de un texto complejo

Grupo 1 (Texto sencillo)	Grupo 1 (Texto complejo)
Puntajes	Puntajes
10	2
5	1
6	7
3	4
9	4
8	5
7	2
5	5
6	3
5	4
Media ⁴ : 6.4	Media: 3.7

Visor de variables

	Nombre	Tipo	Anch	Dec	Etiqueta	Valores	Perdidos	Col	Alineació	Medida
1	Sencillo	Num	8	0	Puntaje recordación texto sencillo	Ninguno	Ninguno	8	Derecha	Escala
2	Complejo	Num	8	0	Puntaje recordación texto complejo	Ninguno	Ninguno	8	Derecha	Ordinal
3	Puntaje	Num	8	0	1ª variable de comparación: sencillo+complejo	Ninguno	Ninguno	8	Derecha	Nominal
4	Códigos	Num	8	0	2ª variable de comparación: códigos 1 sencillo - 2 complejo	{1, sencillo}	Ninguno	8	Derecha	Nominal
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										

Fuente: SPSS®

Metodología. Comparación de 2 medias aritméticas. La prueba “t” no relacionada (paramétrica) se usa para diseños experimentales en los cuales se estudia una variable independiente bajo 2 condiciones, cuando hay diferentes sujetos

4 Media aritmética o promedio de los puntajes.

en las dos condiciones. Es idéntica a la prueba no-paramétrica Mann Whitney para diseños no relacionados con dos condiciones experimentales.

- Procedimiento SPSS®

Utilizar el archivo de trabajo Recordación.sav

Crear 1ª variable de comparación: puntajes (sencillo + complejo)

Crear 2ª. Variable de comparación: códigos (etiquetas: 1 = sencillo; 1 = complejo)

Analizar / Seleccionar Comparar Medias / Seleccionar Prueba T para muestras independientes.

Aparece Cuadro de diálogo: Prueba T para muestras independientes.

En la ventanilla Contrastar variables: Agregar variable Puntaje.

En la ventanilla Variable de agrupación: Agregar Código: grupos: Grupo 1 = 1 Grupo 2 = 2

Nota Metodológica: El objetivo principal de la prueba “t” no relacionada es comparar la cantidad de variabilidad debida a las diferencias predichas en puntaje entre los dos grupos. Representa el tamaño de la diferencia entre las medias de los dos grupos.

Para aceptar la H_0 , el valor observado del nivel de significancia de “t” tiene que ser mayor o igual que los valores escogidos (10% (.10) ó 5% (.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 de 10% (.10) o del 5% (.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)

Resultados: Analizar el valor del nivel de significación (P. Valor) en la Prueba T para la igualdad de medias.

Archivo de trabajo: Conformismo

Descripción. Se aplicó un test de “conformismo al aprendizaje” a una muestra de adolescentes escolares de los cuales 15 pertenecen a una institución

escolar urbana y clasificada de nivel socioeconómico alto y 21 pertenecen una institución escolar rural clasificada de nivel socioeconómico medios bajos y bajos. El test evaluaba el conjunto de actitudes en una escala valórica ascendente de 1 a 50.

Objetivo. Mediante un test se pretende medir un conjunto de actitudes de conformismo al aprendizaje.

Hoja de codificación

Conformismo al aprendizaje

URBANO			RURAL		
ESCOLAR	PUNTAJE	EDADURBANA	ESCOLAR	PUNTAJE	EDADRURAL
1	12	7	1	16	7
2	15	7	2	22	7
3	15	7	3	22	7
4	40	14	4	10	10
5	32	12	5	12	7
6	33	8	6	36	14
7	22	8	7	33	15
8	20	8	8	27	11
9	22	9	9	30	12
10	21	10	10	33	13
11	18	8	11	38	17
12	41	14	12	10	6
13	45	14	13	11	7
14	44	14	14	17	6
15	46	14	15	24	13
			16	39	16
			17	42	16
			18	40	17
			19	26	10
			20	25	9
			21	24	11

Vista de variables

	Nombre	Tipo	Anch	Dec	Etiqueta	Valores	Perdidos	Col	Alineació	Medida
1	Sencillo	Num	8	0	Puntaje recordación texto sencillo	Ninguno	Ninguno	8	Derecha	Escala
2	Complejo	Num	8	0	Puntaje recordación texto complejo	Ninguno	Ninguno	8	Derecha	Ordinal
3	Puntaje	Num	8	0	1ª variable de comparación: sencillo+complejo	Ninguno	Ninguno	8	Derecha	Nominal
4	Códigos	Num	8	0	2ª variable de comparación: códigos 1 sencillo - 2 complejo	{1, sencillo}	Ninguno	8	Derecha	Nominal
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										

Fuente: SPSS®

Metodología. Cuando los datos a comparar en dos sub-grupos corresponden a medidas de nivel ordinal, como por ejemplo, puntajes, valores en escalas de actitudes y se está 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.

Hipótesis de trabajo

H_0 : No existen diferencias estadísticamente significativas entre el grupo de puntaje entre los estudiantes de escuelas de nivel alto *-urbano -* y los de nivel bajo *-rural -*.

H_1 : Existen diferencias estadísticamente significativas entre el grupo de puntaje entre los estudiantes de escuelas de nivel alto *-urbano-* y los de nivel bajo *-rural-*

Procedimiento SPSS®

Utilizar el archivo de trabajo conformismo.sav

Crear 1ª variable de comparación: puntajes (urbano + rural)

Crear 2ª. Variable de comparación: códigos (etiquetas: 1 = urbano; 1 = rural)

En la Barra de menús seleccionar Analizar / Seleccionar el menú Pruebas no-paramétricas

En el Cuadro de diálogo seleccionar: 2 muestras independientes Mann - Whitney

En el cuadro de dialogo: 2 muestras Independientes

Contrastar variable: Puntaje Variable de agrupación: Código

Definir grupos: Grupo 1 = 1; Grupo 2 = 2

Definir tipo de prueba: U de Mann Whitney - Z de Kolmogorov- Smirnov

Análisis de los datos en el Visor de resultados. Si se elige un nivel de significación de 10% (.10) o de 5% (.05) y el nivel de significación obtenido (Asintótica o exacta) es mayor, se debería aceptar la hipótesis nula según la cual "no existen *diferencias estadísticamente significativas* entre el grupo de puntaje entre los estudiantes de escuelas de nivel alto *-urbano* - y los de nivel bajo *-rural*".

Archivo de trabajo: Nominal.sav

Descripción. Se supone 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 obtiene un puntaje puntajes altos o bajos en una variable, también obtienen puntajes altos o bajos en la otra.

Objetivo. Se pretende determinar si existe relación entre la habilidad lectora y la habilidad en el manejo de la herramienta estadística por parte de los estudiantes. La hipótesis es que los estudiantes que obtienen un puntaje alto o bajo en una variable, también obtienen puntajes altos o bajos en la otra.

Hoja de codificación

Puntajes en estadística y en lectura

ESTUDIANTE	PUNTAJE EN ESTADÍSTICA	PUNTAJE EN 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

Vista de variables

	Nombre	Tipo	Anch	Decimal	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	Estadística	Númérico	2	0	Puntajes en Estadística	Ninguno	Ninguno	5	Derecha	Ordinal
2	Lectura	Númérico	2	0	Puntajes en lecto escritura	Ninguno	Ninguno	7	Derecha	Ordinal
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										

Fuente: SPSS®

Metodología. En esta ejemplificación se introduce la noción de diseño correlacional 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 representan “puntajes de Estadística obtenidos por los estudiantes.

En un caso como este 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 sujeto en cada variable se correlaciona con su desempeño en la otra variable. La pregunta por resolver es si la habilidad en lectura y la habilidad en Estadística Social *van juntas* (asociadas), o si no hay ninguna conexión entre los dos tipos de habilidad.

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

Hipótesis de trabajo

H_0 : No existe asociación entre los puntajes obtenidos en estadística y los puntajes obtenidos en habilidad lectora (Independencia entre las variables)

H_1 : existe asociación entre la habilidad en estadística y habilidad en lectura.

Procedimiento SPSS®

Utilizar el archivo de trabajo nominal.sav

En la Barra de menús, seleccionar el menú Analizar/ seleccionar la opción Estadísticos descriptivos / seleccionar la opción Tablas de contingencia

En el Cuadro de dialogo Tablas de contingencia: Estadísticos, seleccionar los estadísticos para variables nominales: Coeficiente de contingencia C; Phi y V. de Cramer; Lambda; Coeficiente de incertidumbre.

En el Cuadro de diálogo: tablas de contingencia: Mostrar en las casillas

Porcentajes: seleccionar filas

Análisis de resultados Si el nivel de significación es mayor a 5% (o a 10%), se acepta la H_0 y si el nivel de significancia es menor a 5% (o a 10%) se rechaza al H_0 y se acepta la H_1

Resultados en el Visor de resultados. Analizar los valores observados del nivel de significación (P valor) para aceptar o rechazar la hipótesis nula:

H_0 = no existe asociación entre las variables (sí es mayor a 5%)

H_1 = existe asociación entre las variables (sí es menor a 5%)

Archivo de trabajo: Deportes_ausentismo.sav

Descripción. 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 de la actividad académica -medida en una escala desde 1 (con frecuencia - hasta 5 (rara vez) - Una de las hipótesis de trabajo hace referencia a que una alta participación en los deportes se relaciona con una baja frecuencia de ausentismo académico.

Hipótesis de trabajo

H_0 : No existe asociación entre los puntajes obtenidos de participación deportiva y ausentismo académico (Independencia entre las variables)

H_1 : Existe asociación entre los puntajes en participación deportiva y ausentismo académico.

Metodología. A partir de los datos obtenidos se trata de analizar si existe una relación entre el grado de actividad deportiva y el nivel de ausentismo académico. En este estudio se entrevistaron a 12 estudiantes. Para los efectos de analizar la relación se utilizó la herramienta estadística: Rho de Spearman que es una prueba no paramétrica que mide el grado de correlación entre los puntajes obtenidos en dos variables y que indica el nivel de significación de la correlación 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.

Un segundo paso es convertir las variables originales en variables rangos, tal como se observa en la hoja de codificación.

Hoja de codificación

Deportes_ausentismo.sav

PARTICIPACIÓN	AUSENTISMO	RANGO A	RANGO B
5	2	5,0	3,5
3	2	2,5	3,5
7	4	7,0	7,5
10	5	12,0	11,0
9	4	9,5	7,5
9	5	9,5	11,0
2	4	1,0	7,5
6	3	6,0	5,0
3	1	2,5	1,5
4	1	4,0	1,5
8	4	8,0	7,5
10	5	12,0	11,0

Vista de variables

Procedimiento SPSS®

En la Barra de Menús seleccionar el menú Analizar / Submenú Correlaciones / Bivariada

En el Cuadro de diálogo Correlaciones Bivariadas / relacionar variable rango A con rango B

Escoger el submenú Spearman / seleccionar la opción Prueba de significación bilateral

Archivo de trabajo: Encuesta al estudiante sobre el Curso y el profesor.sav

Descripción. La apreciación de los estudiantes del Programa de Sociología acerca de su carrera.

Objetivo. La información obtenida se orienta específicamente a las apreciaciones de los estudiantes con respecto a los diferentes aspectos que conforman su vida académica y extraacadémica.

Metodología. El análisis de las diferentes dimensiones del estudio se hizo a nivel del estudiante. Se toma en cuenta que entre los muchos factores que inciden en la formación de una opinión, en este caso, de la vida académica y las expectativas presentes y futuras sobre él, se destacan los siguientes: Características generales y sociodemográficas de los estudiantes; imágenes acerca de su cotidianidad académica en el Programa de Sociología; El entorno académico; Expectativas estudiantiles por actividades programas por la Facultad o el Departamento de Sociología; Apreciación sobre el futuro profesional y apreciaciones sobre la información que recibe sobre el programa de Sociología.

Hoja de codificación

Encuesta al estudiante sobre el curso y el profesor

REGISTRO	CONOCER	MÉTODOS	RELACIÓN	EVALUAR	CURSO	CALIFICACIÓN
1302	3	3	3	2	4	3
1302	4	5	4	4	4	4
1302	4	5	4	4	4	4
1302	4	4	5	5	5	5
1302	4	4	4	4	4	4
1302	4	4	4	4	4	4
1302	4	4	4	4	4	4
1302	4	4	4	4	4	4
1302	3	4	4	3	4	3
1302	4	4	4	4	4	4
1302	4	4	4	3	4	3
1302	5	5	4	4	4	4
1302	4	5	4	4	4	4
1302	4	5	4	4	4	4
1302	3	3	3	3	3	3
1302	4	5	4	4	5	4
1302	4	5	4	4	4	4
1302	4	4	4	4	3	3
1302	4	3	3	3	2	4
1302	4	4	4	4	4	4
1302	4	5	5	4	4	4
1302	4	4	4	4	4	4
1302	4	5	4	4	4	4
1302	3	3	3	3	2	3
1302	5	5	4	5	4	5

Vista de variables

encuesta al estudiante sobre el curso y el profesor [Conjunto_de_datos9] - Editor de datos SPSS										
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?										
	Nombre	Tipo	Anch	Decimal	Etiqueta	Valores	Perdido	Col	Alineación	Medida
1	Registro	Cadena	8	0	Registro del docente	Ningun	Ninguno	6	Derecha	Nominal
2	Conocer	Numérico	1	0	Dominio de los temas del curso	Ningun	Ninguno	7	Derecha	Escala
3	Metodos	Numérico	1	0	Empleo recursos didácticos	Ningun	Ninguno	6	Derecha	Escala
4	Relación	Numérico	1	0	Respeto en el trato con estudiantes	Ningun	Ninguno	6	Derecha	Escala
5	Evaluar	Numérico	1	0	Evaluación de ejercicios y exámenes	Ningun	Ninguno	6	Derecha	Escala
6	Curso	Numérico	1	0	Evaluación del interés de los contenidos	Ningun	Ninguno	5	Derecha	Escala
7	Calificación	Numérico	1	0	Calificación del Docente de 1 a 5	Ningun	Ninguno	5	Derecha	Escala
8										
9										
10										
11										
12										
13										
14										
15										

Fuente: SPSS®



Imprenta
Universidad de Antioquia
PENDIENTE de 2006

