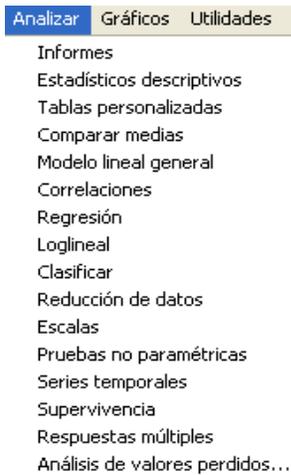


MANUAL SIMPLIFICADO DE ESTADÍSTICA APLICADA VIA SPSS

Medidas de tendencia central

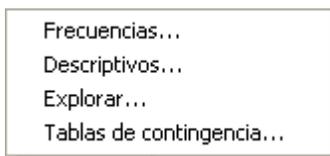
Menú Analizar: Los comandos del menú Analizar (Estadística) ejecutan los procesamientos estadísticos. Sus comandos están organizados dentro de sub-menú de acuerdo al tipo de análisis a ejecutar. Exhibe algunos de los comandos de operaciones de la estadística descriptiva e inferencial.



Los comandos del menú Analizar (estadísticas) están disponibles sólo si hay un archivo de datos activo - abierto.

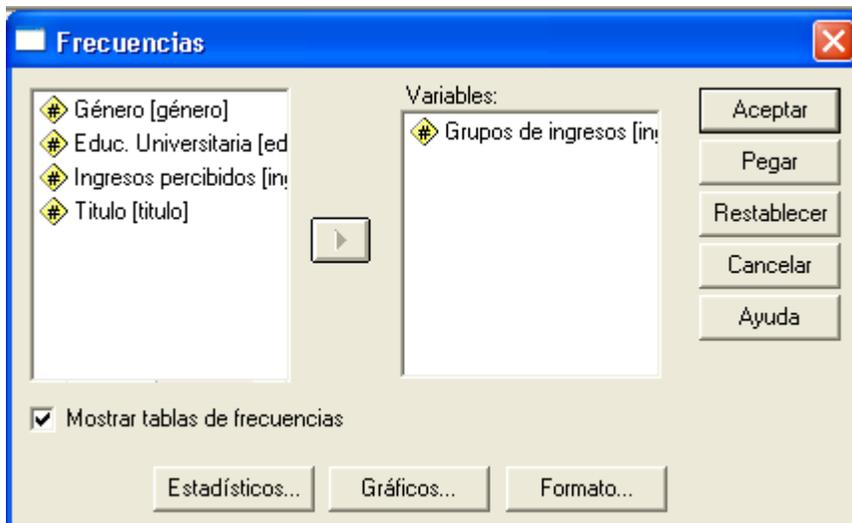
Descripción de los comandos

El sub-menú **Estadísticos Descriptivos** incluye



Frecuencias (estadísticos de frecuencias)

Para obtener estadísticos descriptivos y de resumen opcionales para las variables numéricas, pulse en Analizar → Estadísticos descriptivos → frecuencias. Esto abrirá el Cuadro de Diálogo (Frecuencias), como se muestra en el siguiente Cuadro de Diálogo.



CALCULO DE ESTADÍSTICOS



Estadísticos		
Ingresos percibidos		
N	Válidos	1100
	Perdidos	0
Media		26064,20
Desv. típ.		6967,98
Mínimo		7200
Máximo		65500
Percentiles	25	21000,00
	50	26000,00
	75	30375,00

ANÁLISIS DE DATOS

VALORES DE PERCENTILES. Puede elegir una o más de las siguientes alternativas:

Cuarteles. Muestra los percentiles 25, 70 y 75.

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

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

DISPERSIÓN: Puede elegir una o más de las siguientes alternativas.

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

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

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

Mínimo: El valor más pequeño.

Máximo: El valor más grande.

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

MEDIDAS DE TENDENCIA CENTRAL: Puede elegir una o más de las siguientes alternativas.

Media: La media aritmética (promedio).

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

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

Suma: la suma de todos los valores.

DISTRIBUCIÓN: puede elegir una o más de las siguientes alternativas:

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

Curtois: estimación del grado en que las observaciones se agrupan alrededor del punto central, también se muestra el error típico del estadístico de curtois. Para datos agrupados, también se encuentra disponible la siguiente opción:

MEDIDAS DE TENDENCIA CENTRAL.

Frecuentemente se necesita resumir un conjunto de valores numéricos para lo cual es apropiado recurrir a las medidas de tendencia central. Estas medidas, en términos coloquiales, hacen referencia a los llamados promedios, aunque este concepto no tiene un significado estadístico preciso.

Existen realmente 3 medidas que pueden ser asociadas a este concepto de promedio y ellas son: la media, la mediana y la moda.

La media es probablemente la medida de tendencia central más utilizada. La Media se calcula sumando los puntajes de una distribución de valores y dividiéndolos por el número total de los puntajes.

La mediana: es el número medio en un conjunto de valores ordenados en forma creciente o decreciente. Veamos con un ejemplo en que consiste esta medida:

Ejemplificación: Calcular *la mediana* en la siguiente serie de 25 números ordenados

1 2 2 3 3 3 3 3 3 4 4 4 4 5 5 5 7 8 10 10 15 25 30 100 150

El valor medio es 4 - valor que ocupa el puesto 13 en la serie.

La media es mucho más sensible a los valores extremos que lo que es la Mediana. La Media está fuertemente influida por los valores extremos o marginales. La Mediana, al contrario, es menos determinada por estos valores extremos.

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

La moda hace referencia a la ocurrencia considerada más típica del hecho; es simplemente el valor que se presenta más frecuentemente.

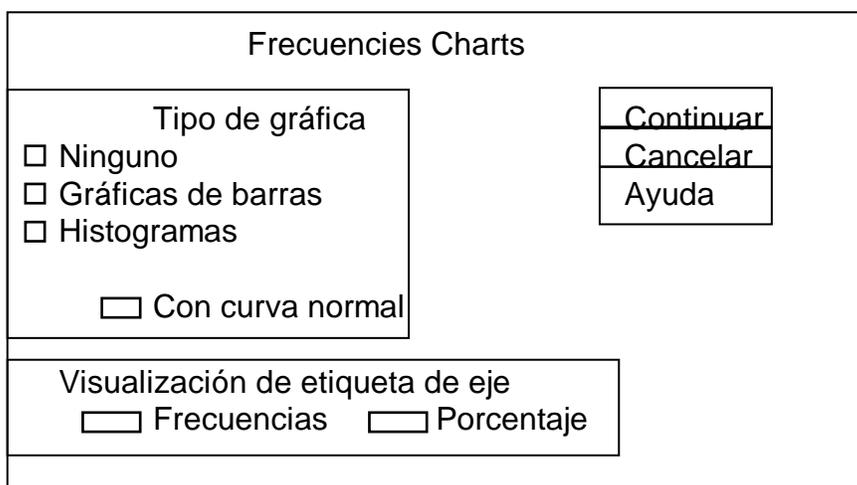
En síntesis, con la ayuda de estos estadísticos se puede hacer el siguiente análisis y presentación de los datos:

El uso de la Media, la Mediana o la Moda depende del nivel de medición utilizado; si se está resumiendo **datos nominales**, sólo la Moda es utilizable. Estas tres medidas de tendencia central son utilizables con **datos a nivel de intervalos o numéricos**.

GRAFICAS DE FRECUENCIA.

Para obtener gráficas de barras o histogramas, pulse en Gráficas del Cuadro de Diálogo Frecuencias. Esto abrirá el Cuadro de Diálogos Frecuencias: Gráficas, como se muestra en la figura siguiente.

Cuadro de Diálogo Frecuencias: Gráficas



Frecuencias Charts

Tipo de gráfica

Ninguno

Gráficas de barras

Histogramas

Con curva normal

Continuar

Cancelar

Ayuda

Visualización de etiqueta de eje

Frecuencias Porcentaje

Tipo de gráfica: puede elegir una de las siguientes alternativas:

Ninguna gráfica: este es el valor predeterminado.

Gráfica de barras): La escala viene determinada por la frecuencia mayor de las categorías dibujadas.

Histograma: Los histogramas se encuentran disponibles solo para variables numéricas. El número de intervalos trazados es de 21 (o menor si la amplitud de valores es menor que 21).

With Normal Curve (Con curva normal): Esta opción superpone una curva normal sobre los histogramas.

OBTENCIÓN DE TABLAS DE CONTINGENCIA

El procedimiento de tablas de contingencia produce tablas de dos factores hasta n factores y estadísticos relacionados para variables numéricas y categóricas. Además de las frecuencias de las celdas, puede obtener porcentajes de las celdas, valores esperados y residuos.

Las especificaciones mínimas son:

- Una variable de fila numérica o categórica.
- Una variable de columna numérica o categórica

Para obtener tablas de contingencia y estadísticos relacionados así como medidas de asociación, elija de los menús:

Analizar → **Estadísticos descriptivos** → **Tablas de contingencia**

Esto abrirá el Cuadro de Diálogos **Tablas de Contingencia**, como se muestra en la siguiente ventanilla.

Tablas de contingencia: Todas las estadísticas que calculan Tablas de contingencia son medidas de asociación o, en su defecto, de independencia; pero la negación de una lleva a la afirmación de la otra; si dos variables no son independientes están asociadas.

De forma opcional, puede seleccionar una o más capas de variables de control. Se genera una tabla de contingencia independiente para cada categoría de la variable independiente y variable de control. Por ejemplo, si tiene una variable de fila, una variable de columna y una variable de control con dos categorías, podrá obtener dos tablas de contingencia.

TABLA DE CONTINGENCIA.

La Tabla 1, es una **tabla de contingencia** de sexo y edad. El número de casos con cada combinación de valores de las dos variables aparece en una **celda** en la tabla, junto con varios porcentajes. Estas entradas en las celdas proporcionan información acerca de las relaciones entre las dos (2) variables.

En el gráfico, la edad se le llama **variable de columna**, ya que cada estado aparece en una columna de la tabla. De modo similar, al sexo de los encuestados se le llama la **variable de fila**. Con las cuatro (4) categorías de la variable de columna y las dos de la de fila hay ocho (8) celdas en la tabla.

CÁLCULO POR FILA

SEXO		EDAD				Filas Total
		<20 1.	20-29 2.	40-49 3.	50 4.	
Femenino	1	2 8.3%	8 33.3%	12 50.0%	2 8.3%	24 63.2%
Masculino	2	2 14.3%	5 35.7%	4 28.6%	3 21.4%	14 36.8%
Total		10.5	34.2	42.1	13.2	100.0%

CÁLCULO POR COLUMNA

Sexo		Edad				Filas Total
		<20 1.	20-29 2	40-49 3	50 4.00	
Femenino	1	50.0%	61.5%	75.8%	40.0%	63.2%
Masculino	2	50%	38.5%	25.0%	60.0%	36.8%
Total		10.5	34.2	42.1	13.2	100.0%

CÁLCULO POR EL 100% (TOTAL DE CASOS)

Sexo		Edad				Row Total
		<20 1.	20-29 2	40-49 3	50 4.	
Femenino	1	5.3%	21.1%	31.6%	5.3%	63.2%
Masculino	2	5.3%	13.2%	10.5%	7.9%	36.8%
Total		10.5	34.2	42.1	13.2	100.0%

CONTENIDO DE LAS CELDAS Y MARGINALES.

La primera entrada en las celdas es el número de casos, la **frecuencia**, de dicha celda. La segunda anotación de la tabla es el **porcentaje de fila**. Es el porcentaje de todos los casos de una fila que se encuentra en una celda determinada.

El **porcentaje de columna**, el tercer elemento de cada celda, es el porcentaje de todos los casos de una columna que se encuentran en una celda.

La última anotación de la celda es **porcentaje de tabla**. El número de casos de una celda se expresa como un porcentaje del número total de casos de la tabla.

Los números situados a la derecha y debajo de la tabla se conocen como "**marginales**". Son las frecuencias y porcentajes de las variables de la fila y de columna tomadas por separado.

ELECCIÓN DE PORCENTAJES

Los porcentajes de fila, columna y total muestran distintos tipos de información, por lo que resulta extremadamente importante elegir la dirección de la lectura de los porcentajes, con sumo cuidado entre ellos.

En este ejemplo, el porcentaje de columna muestra la distribución de hombres y mujeres en cada una de las categorías de respuesta. Esto no proporciona información directa.

El porcentaje de fila nos dice el porcentaje de hombres y mujeres. La interpretación de esta comparación no se verá afectada por la desigualdad numérica de hombres y mujeres en el estudio.

Puesto que siempre es posible intercambiar las filas y las columnas de cualquier tabla, no pueden proporcionarse reglas generales acerca de cuando utilizar los porcentajes de fila y columna. Los porcentajes que se van a utilizar depende de la naturaleza de las dos variables. Si una de las dos variables está bajo control experimental, se le llamará **variable independiente**. Se supone que la variable independiente afectará la respuesta, o **variable dependiente** si las variables pueden calificarse como dependientes o independientes, las siguientes directrices pueden ser muy útiles:

ESTADÍSTICOS DE LA TABLA DE CONTINGENCIA

Aunque un examen de los distintos porcentajes de fila y columna en una tabla de contingencia es un primer paso útil en el estudio de las relaciones entre dos variables, los porcentajes de fila y de columna no permiten la cuantificación o comprobación de dicha relación. Para esa finalidad es útil considerar distintos índices, **que miden el grado de la asociación**, así como las pruebas estadísticas de la hipótesis de que no existe asociación.

Prueba Chi cuadrado de independencia

La hipótesis de que dos variables en una tabla de contingencia **son independientes entre sí** es a menudo de interés.

Un estadístico utilizado frecuentemente para probar la hipótesis de que las variables de fila y de columna son independientes, es el **Chi Cuadrado de Pearson**.

MEDIDAS DE ASOCIACIÓN

En muchas situaciones de búsqueda, son **el grado y la naturaleza de la dependencia** entre las variables las que despiertan más interés. A los índices que intentan cuantificar la relación entre las variables en una tabla de contingencia se les llama **medidas de asociación**.

Ninguna medida resume de forma adecuada todos **los tipos posibles de asociación**. Las medidas varían en su interpretación y en la forma en que definen la asociación perfecta e intermedia. Estas medidas también se diferencian en la forma en que las afectan varios factores como los marginales. Por ejemplo, muchas medidas son *sensibles a las marginales* ya que están influidas por las distribuciones marginales de las filas y columnas. Tales medidas reflejan la información acerca de las marginales además de la información acerca de la asociación.

MEDIDAS ORDINALES

Aunque las relaciones entre las variables ordinales puedan examinarse utilizando medidas nominales; otras medidas reflejan la información adicional disponible a partir de la ordenación. La consideración de las clases de relaciones que pueden existir entre dos variables ordenadas conduce a la noción de dirección de relación y al concepto de **correlación**. Las variables se correlacionan de forma positiva si

los casos con valores menores para una variable también tienden a tener valores menores para la otra y, los casos con valores altos para una también tienden a ser altos para la otra. Las variables correlacionadas negativamente muestran la relación opuesta: cuanto mayor sea la primera variable, menor tendrá a ser la segunda.

El **coeficiente de correlación de Spearman** es una medida de correlación utilizada habitualmente entre **dos variables ordinales**. Para todos los casos, los valores de cada una de la variables se clasifican de menor a mayor y el coeficiente de correlación de Pearson se calcula en base a los rangos

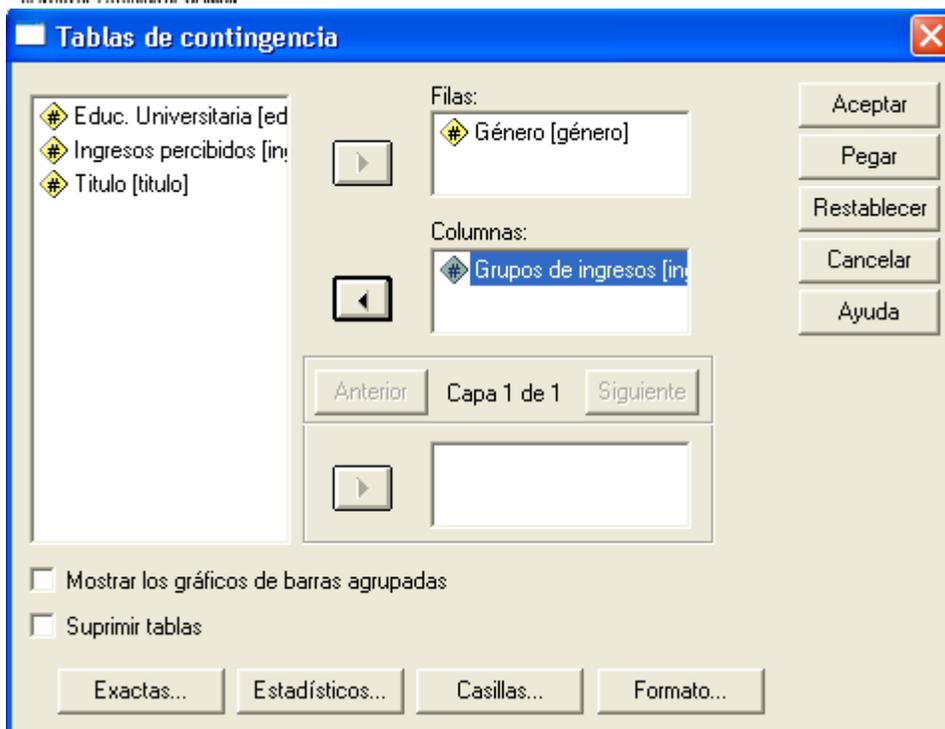
MEDIDAS CON DATOS DE INTERVALOS

Si las dos variables en la tabla se calculan en **una escala de intervalos**, pueden calcularse distintos coeficientes que hacen uso de esta información adicional. Un coeficiente simétrico útil que mide la fuerza de la relación lineal es el **Coefficiente de correlación de Pearson (r)**. Puede tomar valores desde -1 a +1, indicando correlaciones lineales positivas o negativas.

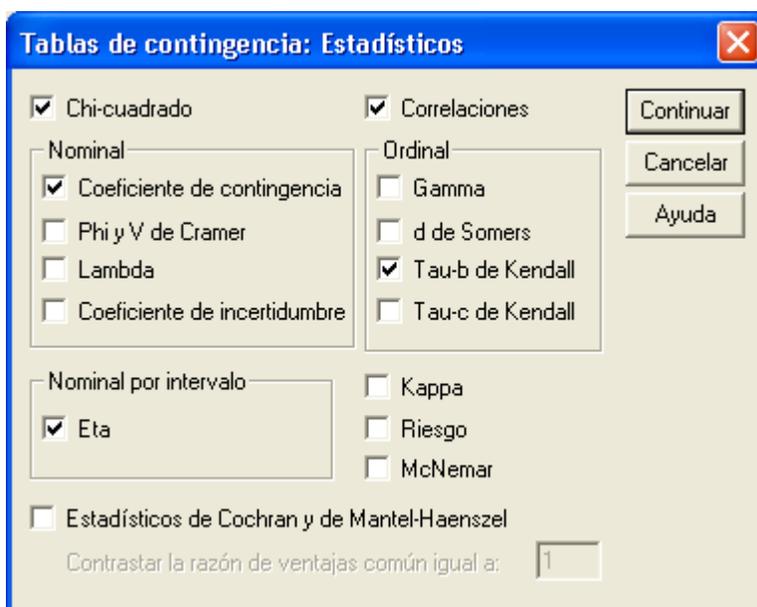
El **coeficiente ETA** es adecuado para datos en que la variable dependiente se calcula en una escala de intervalos y la variable independiente en una escala nominal y ordinal. Cuando se eleva al cuadrado, **ETA** puede interpretarse como la proporción de la variabilidad total de la variable dependiente que puede ser explicada por el conocimiento de los valores de la variable independiente. La medida es asimétrica y no da por puesta una relación lineal entre las variables

Estadísticos de tablas de contingencia

Para obtener estadísticos y medidas de asociación, pulse en Analizar → Estadísticos descriptivos → Tablas de contingencia; aparecerá el Cuadro de Diálogo *Tablas de contingencia*, como se muestra en la próxima figura



En el cuadro de dialogo seleccione Estadísticos y aparecerá el siguiente cuadro de dialogo:



Puede elegir uno o más de los siguientes estadísticos:

El Chi Cuadrado de Pearson, **correlaciones de Pearson** y coeficiente de **Correlación de Spearman**. Estos se encuentran disponibles sólo para datos numéricos.

- **Kappa**. El coeficiente Kappa sólo puede calcularse para tablas cuadradas en las que los valores de fila y de columna sean idénticos
- **Riesgo**. La tasa de riesgo relativo. Este sólo puede calcularse para tablas de 2x2.

Datos Nominales. Las medidas nominales clasifican por ejemplo, católico, protestante, judío; femenino, masculino, Puede elegir una o más de las siguientes alternativas:

- **Coefficiente de contingencia**
- **Phi y V de Cramér**
- **Lambda**. La Lambda Simétrica y la asimétrica y la Tau de Goodman y Kruskal.
- **Coefficiente de incertidumbre**. El coeficiente de incertidumbre simétrico y el asimétrico

NOMINAL POR INTERVALO. Se supone que una variable se calcula en una escala nominal y la otra se calcula en una escala de intervalos

- **ETA**. Se utiliza para valores medidos en una escala nominal.

DATOS ORDINALES. Las medidas ordinales Puede elegir una o más de los siguientes estadísticos:

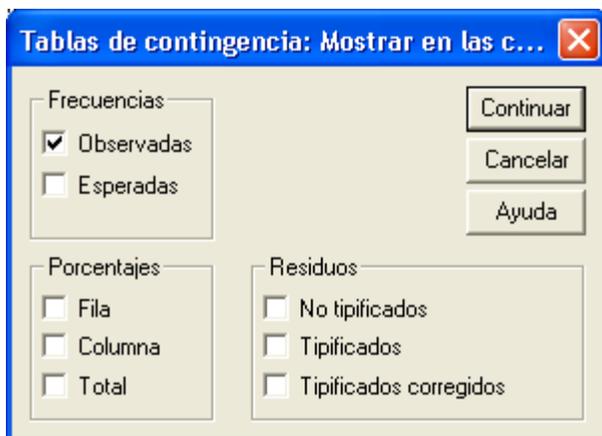
- **Gamma**.
- **D de Somers**: simétrica y asimétrica.
- **Tau b de Kendall**
- **Tau c de Kendall**

VISUALIZACIÓN EN LAS CELDAS DE LAS TABLAS DE CONTINGENCIA

La tabla de contingencia predeterminada sólo muestra el número de casos en cada celda. También puede ver los porcentajes de fila, columna y totales, valores esperados y residuos.

Para cambiar la visualización de la celda, pulse en Celdas del Cuadro de Diálogo Tablas de contingencia.

Esto abrirá el cuadro Tablas de Contingencia): Visualizar en celdas. Como se muestra en la siguiente figura.



Puede elegir cualquier combinación de visualizaciones en las celdas. Se origina seleccionando los porcentajes de fila, columna y total, además de la frecuencia observada predeterminada. Debe seleccionarse al menos una opción.

Frecuencias. Puede elegir una o más de las siguientes opciones

- **Observadas.** Frecuencias observadas. Esta es la opción predeterminada.
- **Esperadas.** Frecuencias esperadas. El número de casos esperado en cada celda si las dos variables de la subtabla son estadísticamente independientes.
- **Porcentajes.** Puede elegir una o más de las siguientes opciones:
 - **Fila.** El número de casos en cada celda expresado como un porcentaje de todos casos de dicha fila
 - **Columna.** El número de casos en cada celda expresado como un porcentaje de todos los casos de subtabla

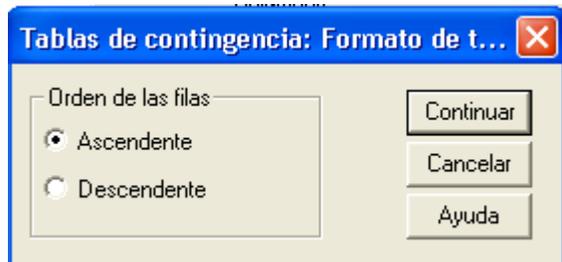
- **Total.** El número de casos en cada celda expresado como un porcentaje de todos los casos de subtabla

Residuos. Puede elegir una o más de las siguientes opciones:

- **No tipificado.** El valor de frecuencia de celda observada menos el valor esperado.
- **Tipificados.** Residuos tipificados
- **Tipificados corregidos.** Residuos tipificados corregidos

FORMATO DE LAS TABLAS DE CONTINGENCIA

Puede modificar el formato de tablas pulsando en Formato) del Cuadro de Diálogo Tablas de contingencia. Esto abrirá el Cuadro de Diálogo Tablas de contingencia: Formato de tablas, como se muestra en la siguiente figura.



MEDIDAS DE DISPERSIÓN.

EL RANGO.

Este estadístico mide la dispersión de los valores en una serie. Veamos con un ejercicio como se expresa esta medida: Dos conjuntos de valores pueden presentar diferentes distribuciones y tener idénticas medias aritméticas. Por ejemplo, supongamos que estamos analizando la atención media diaria de los médicos de los dos consultorios del Servicio de Salud de la Universidad:

Nº DE CONSULTAS DIARIAS EN LOS 2 CONSULTORIOS.

Consultorio A			Consultorio 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
X = 30			X = 30	

En ambos casos **el número medio**¹ de consultas es de 30 pacientes por día. Pero como se logra apreciar la distribución de las consultas no es idéntica en los dos consultorios: en el Consultorio A la distribución desigual de las consultas; los médicos D y E atienden de dos a cinco veces más pacientes de los que atienden sus colegas A y B.

El "rango" de consultas en este consultorio es 40 (10 el mínimo y 50 el máximo). En el otro consultorio el "rango" de consultas es de sólo de 4 (32 consultas - 28 consultas).

DESVIACIÓN ESTÁNDAR.

Generalmente es necesario ir más allá del rango para mirar cada valor en relación con valor medio. Cuando se mide la dispersión, estadísticamente lo que se está midiendo son los valores en relación con la Media.

Varianza: Otra medida de dispersión asociada con la Media y la desviación estándar es la Varianza.

¹ **Media aritmética.**

Medidas de relaciones entre variables.

A veces, sin embargo, los investigadores no están interesados en predecir diferencias en la conducta como resultado de una variable independiente. En cambio, quieren investigar si las variables están asociadas entre sí. Tome de nuevo el ejemplo de la relación entre la habilidad para la ortografía y los puntajes en lectura. Un investigador podría predecir que en *todo el rango* de la habilidad para la ortografía y de la habilidad para la lectura, los estudiantes que tienen buena ortografía también tenderán a obtener buenos puntajes en una prueba de lectura, mientras que los estudiantes atrasados en ortografía lo estarán también en lectura. Esto podría ser útil si quisiera hacer un examen de ortografía para escoger aquellos estudiantes que posiblemente van a tener buen rendimiento en lectura.

Pruebas no paramétricas

Cuando usarla:

La prueba del **Chi Cuadrado** es apropiada cuando los datos obtenidos en su estudio son de **nivel nominal**. Esto significa que en lugar de poder medir los puntajes de los sujetos, sólo es posible asignar los sujetos a una o más *categorías*. Esta medida no es propiamente una medida de asociación pero permite a partir de la idea que se puede definir "una asociación cero" en términos de la igualdad lograda entre la frecuencias observadas y las frecuencias esperadas.

Nota Metodológica:

Para calcular la medida de Chi-Cuadrado se sigue el siguiente procedimiento: el valor de la frecuencia esperada en cada celda de la matriz de una tabla de un cruce de variables es restado del valor de la frecuencia observada en esa celda y el valor resultante elevado al cuadrado.

Tablas de contingencia (Cruce de variables)

El subprograma tablas de contingencia elabora *tablas de contingencia* y calcula sus estadísticas.

En este subprograma se puede utilizar cualquier tipo de variables: *Cuantitativas Discretas, Cuantitativas Continuas, Cualitativas Nominales, Cualitativas Ordinales*. Por supuesto, las estadísticas que se calculan tienen validez sólo para las variables adecuadas; por ejemplo, el coeficiente de correlación entre sexo y edad

no tiene ningún sentido, ya que sexo es cualitativa nominal y edad es cuantitativa continua; pero el coeficiente de correlación entre estatura y peso sí tiene validez, porque ambas son cuantitativas, y con ese coeficiente se puede determinar si existe o no relación entre esas variables, con un nivel de significancia dado.

Todas las estadísticas que calcula la Tabla de contingencia 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 llevará si dos variables no son independientes están asociadas. *Así, el punto por considerar es cuál de las estadísticas tiene significado dado el tipo de las variables*

Las tablas de contingencia también se conocen como *tablas cruzadas*. Cuando se forman con dos variables reciben el nombre de *tablas de dos entradas*; cuando son más de dos las variables involucradas, son tablas de *n* entradas, donde *n* es el número de variables involucradas. A cada tabla obtenida del cruce de dos o más variables se le conoce como *subtabla*, dado que en realidad es parte de la tabla principal formada por las dos primeras variables.

Análisis de Regresión

La *regresión* es una técnica basada en los mínimos cuadrados que permite analizar la relación entre dos variables (*regresión simple*), o entre una variable dependiente y varias independientes (*regresión múltiple*).

La regresión sirve para analizar relaciones funcionales; es decir, relaciones en las que los valores de una variable están en función de los de otra u otras.

El caso de la regresión múltiple es diferente; con valores de las variables independientes se puede estimar el valor de la variable dependiente; pero un valor de la variable dependiente no sirve para estimar los valores de las variables independientes. En todo caso sólo se puede estimar uno, el resto tendría que declararse. También se obtiene el coeficiente de correlación (simple o múltiple) que indica la magnitud de la relación que guardan las variables incluidas en el análisis.