

Universidad Católica del Uruguay
Facultad de Ciencias Humanas

Autotexto

Serie Estadística

Tablas de Contingencia



**Universidad
Católica**

DAMASO A. LARRAÑAGA • URUGUAY

Laboratorio Metodológico

Versión original:

P. Alegre – J. Bogliaccini– F. Rodríguez

Autores revisión 2008:

A. de León – M. Dodel – C. Rafaniello

**Laboratorio de Tratamiento y Análisis de la Información
Facultad de Ciencias Humanas - UCUDAL**

Módulo de Práctica de Análisis

Serie Estadística

Tema: Tablas de contingencia

Descripción

Las tablas constituyen una modalidad de presentación de los datos frecuentemente utilizada. Existen tablas “univariadas”, que presentan la distribución de una sola variable. Otro tipo de tablas permiten presentar información de dos o más variables. Este tipo de tablas se le denomina “bivariadas” (cuando refieren a la presentación de dos variables) o “trivariadas” (cuando refieren a la presentación de información referida a tres variables). Usualmente este tipo de tablas nos permiten establecer relaciones primarias entre las variables a partir de analizar la distribución de la información agregada en los cruces de los respectivos “valores” de cada una de ellas.

A modo de ejemplo, una tabla que representa el cruce de la variable “Estado Civil” (con los valores Casado/ No Casado) y la variable “Edad” (codificada en tramos de edad) permite observar la relación existente entre ambas a partir de la distribución de la frecuencia¹ de los casos en distintos cruces de valores de las variables. Por ejemplo, la concentración de casos en el cruce de los valores “Casados” y en los tramos de edad de 30 hacia adelante, permite establecer una relación entre la edad y el estado civil, en donde “a mayor edad, mayor probabilidad de estar casado”.

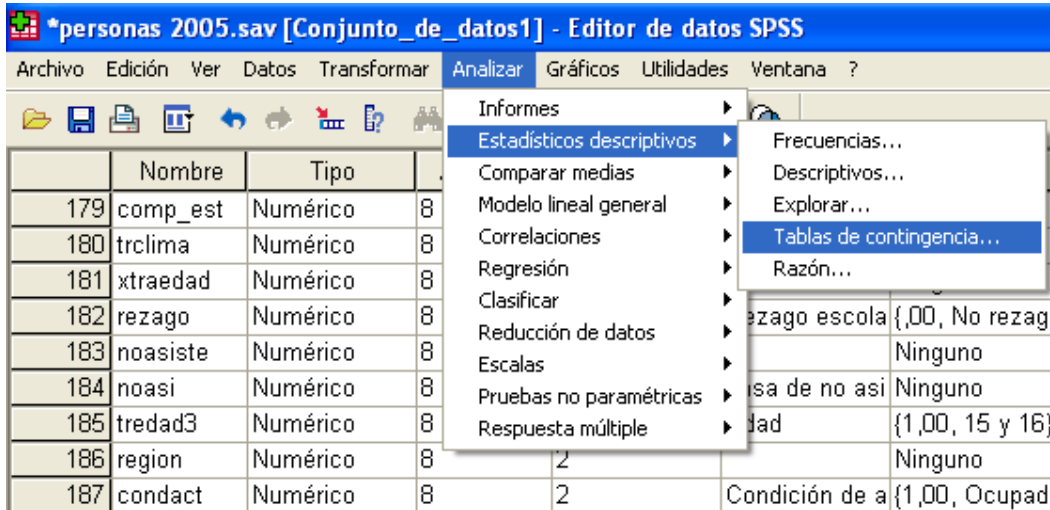
ESTADO CIVIL	EDAD			
	15-29	30-44	45-59	60 y más
Casado				
No Casado				

¹ Ver autotexto Análisis Univariado

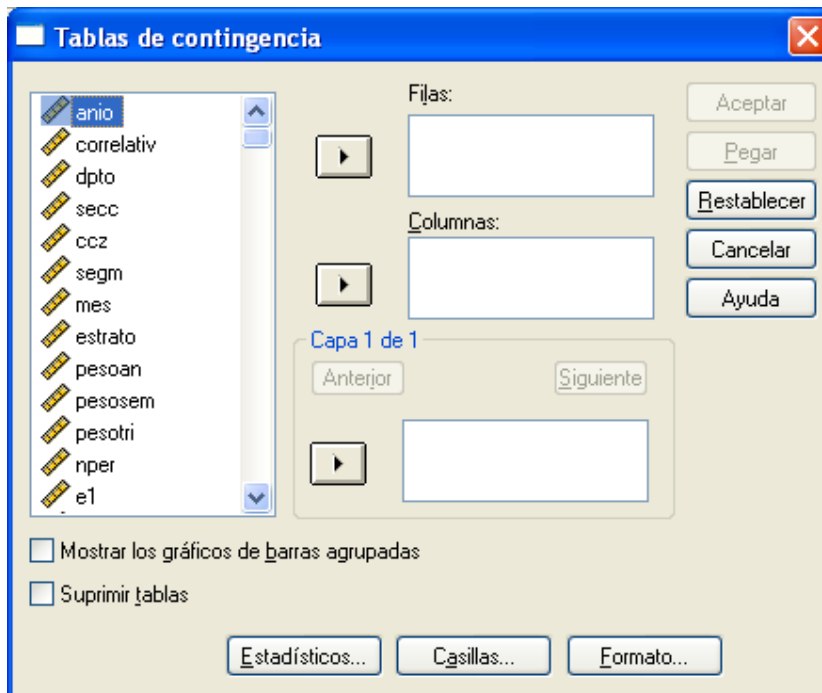
Rutina

A. Análisis Bi-variado por medio de Tablas de Contingencia

Para ir a Tablas de contingencia debemos seleccionar la opción *Analizar/ Estadísticos descriptivos/ Tablas de contingencia*



Allí se abre el menú general:



En la columna de la izquierda tenemos todas las variables de la base de datos para ingresar. A la derecha de la pantalla tenemos el espacio en donde se

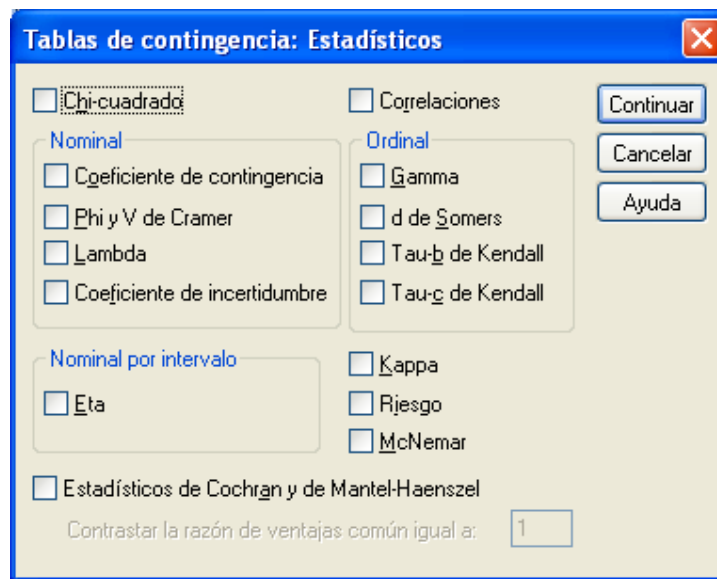
ingresan aquellas variables que están en la fila y debajo de ella el espacio en donde se ingresan las variables que irán en la columna.

Usualmente la variable independiente se coloca en la “columna” mientras la variable dependiente se inserta en la “fila”.

Debajo de ellas aparece un tercer espacio, el cual permite ingresar variables que especifican el cruce de datos de las dos variables anteriores. Por ejemplo, si tenemos sexo en la columna y estado civil en la fila, y seleccionamos en el tercer espacio el departamento, tendremos en la tabla el número de hombres y mujeres casados, divorciados, viudos, etc, en Artigas, Salto, Paysandú, Rivera y así por todos los departamentos.

Al costado del recuadro aparecen las opciones clásicas en cualquier función que ejecutamos en el SPSS: la opción que ratifica la operación “Aceptar”, la opción de pegar la sintaxis de la operación que pedimos al programa “Pegar”, la opción que deselectiona las variables seleccionadas “Restablecer”, la opción de cancelar la operación “Cancelar” y la opción de ayuda “Ayuda”.

Debajo del recuadro aparecen las opciones “Estadísticos”, “Casillas” y “Formato”.



La opción “Estadísticos” contiene los diferentes tipos de tests estadísticos que pueden realizarse con los datos de las tablas. Entre ellos se encuentran distintas medidas de independencia y asociación para diferentes valores de medición de las variables (Nominales, Ordinales, Intervalos). A saber:

Chi cuadrado²: Es un test estadístico para datos medidos en nivel discreto o categóricos. Es usado generalmente para tablas bi-variadas de 2X2, es decir con variables que contienen dos categorías. El chi-cuadrado permite determinar hipótesis relacionadas a la existencia de relación entre los valores de las variables.

² Ver autotexto Chi cuadrado

Correlations: Corresponde a test de asociación que permiten conocer la intensidad de una relación entre dos variables.

Coeficientes de tipo nominal: Coeficientes que permite medir la magnitud de la asociación estadística (en caso de existir) testeada por el chi-cuadrado. Estos coeficientes son utilizados para variables de tipo nominal. Incluye a su vez coeficientes de asociación direccionales de nivel nominal que no se basan en el chi-cuadrado (Lambda, Uncertainly coefficient).

Dentro de los coeficientes de tipo nominal podemos encontrar:

Coeficiente de contingencia: Coeficiente que extiende su rango entre 0 y 1. El valor 0 indica ninguna asociación entre las filas y las columnas, y el valor 1 indican un alto grado asociación entre las variables. El valor máximo posible depende del número de filas y de columnas en una tabla.

Phi y V de Cramer: Phi es una medida de asociación basada en chi-cuadrado que implica dividir chi-cuadrado por el tamaño de muestra y tomar la raíz cuadrada del resultado. V de Cramer es una medida de la asociación basada en chi-cuadrado.

Lambda: Una medida de la asociación que refleja la reducción proporcional en el error cuando los valores de la variable independiente se utilizan para predecir valores de la variable dependiente. Un valor de 1 significa que la variable independiente predice perfectamente la variable dependiente. Un valor de 0 significa que la variable independiente no es de ninguna ayuda al predecir la variable dependiente.

Coeficiente de incertidumbre. Una medida de la asociación que indica la reducción proporcional en el error cuando los valores de una variable se utilizan para predecir valores de la otra variable. Por ejemplo, un valor de 0.83 indica que el conocimiento de una variable reduce error en valores que predicen de la otra variable por el 83%. El programa calcula las versiones simétricas y asimétricas del coeficiente de la incertidumbre.

Coeficientes de tipo ordinal: Coeficientes que permite medir la magnitud de la asociación estadística (en caso de existir) testeada por el chi-cuadrado. Estos coeficientes son utilizados para variables de tipo ordinal.

Dentro de los coeficientes de tipo ordinal podemos encontrar:

Gamma: Una medida simétrica de la asociación entre dos variables ordinales que se extiende entre -1 y 1. Los valores cerca de un valor absoluto de 1 indican una relación fuerte entre las dos variables. Los valores cerca de cero indican poco o nada de relación. Para las tablas bidireccionales, se exhiben los gammas de orden cero. Para las tablas trivariadas o superiores, se exhiben los gammas condicionales.

D de Somers. Una medida de la asociación entre dos variables ordinales que se extiende a partir de -1 a 1. Los valores cerca de un valor absoluto de 1 indican una relación fuerte entre las dos variables, y los valores cerca de 0 indican poco o nada de relación entre las variables. D de Somers es una extensión asimétrica de gamma que diferencia solamente en la inclusión del número de los pares no atados en la variable independiente. Una versión simétrica de esta estadística también se calcula.

Tau-b de Kendall. Una medida no paramétrica de correlación para las variables ordinales o alineadas que toman en cuenta los empates. El signo del coeficiente indica la dirección de la relación, y su valor absoluto indica la fuerza, con valores absolutos más grandes indicando relaciones más fuertes. Los valores posibles se extienden a partir de -1 a 1, pero un valor de -1 o +1 se puede obtener solamente de las tablas cuadradas.

Tau-c de Kendall. Una medida no paramétrica de la asociación para las variables ordinales que no hace caso de lazos. El signo del coeficiente indica la dirección de la relación, y su valor absoluto indica la fuerza, con valores absolutos más grandes indicando relaciones más fuertes. Los valores posibles se extienden a partir de -1 a 1, pero un valor de -1 o +1 se puede obtener solamente de las tablas cuadradas.

Coefficiente para variables nominales por intervalos (eta): Coeficiente específico para las tablas bi-variadas que combinan variables nominales con intervalos.

Eta. Una medida de la asociación que se extiende a partir de 0 a 1, donde 0 indicando ninguna asociación entre la fila y las variables y los valores de la columna cerca de 1 que indica un alto grado de la asociación. Eta es apropiado para una variable dependiente medida en una escala de intervalo (p.ej.: el ingreso) y una variable independiente con un número limitado de categorías (p.ej.: el género). Se computan dos valores de eta: uno trata la variable de la fila como la variable de intervalo; el otro trata la variable de la columna como la variable de intervalo.

Otros coeficientes especiales:

Kappa de Cohen. Mide el acuerdo entre las evaluaciones de dos evaluadores cuando ambos están clasificando el mismo objeto. Un valor de 1 indica el acuerdo perfecto. Un valor de 0 indica que el acuerdo es no mejor que al azar. Kappa está solamente disponible para las tablas en las cuales ambas variables utilizan los mismos valores y ambas variables tienen el mismo número de categorías.

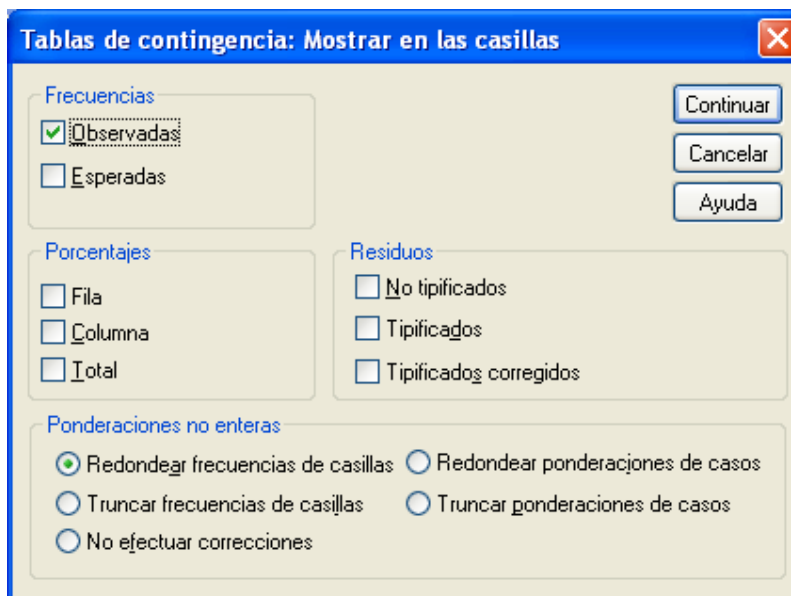
Riesgo. Para tablas de 2 x 2, una medida de la fuerza de la asociación entre la presencia de un factor y la ocurrencia de un acontecimiento. Si el intervalo de la confianza para el estadístico incluye un valor de 1, usted no puede asumir que el factor está asociado al acontecimiento. El cociente de las probabilidades

se puede utilizar como una estimación o riesgo relativo cuando la ocurrencia del factor es infrecuente.

McNemar. Una prueba no paramétrica para dos variables dicotómicas relacionadas. Pone a prueba los cambios en respuestas usando la distribución de chi-cuadrado. Útil para detectar cambios en las respuestas debido a la intervención experimental en diseños "antes-y-después". Para tablas cuadradas más grandes, se calcula la simétrica prueba de McNemar-Bowker.

Cochran y Mantel-Haenszel. El estadístico de Cochran y de Mantel-Haenszel se pueden utilizar para probar la independencia entre dos variables dicotómicas, condicional sobre una o más variables de control (capa). Observe que mientras que otros estadísticos son computados capa por capa, el de Cochran y el de Mantel-Haenszel son computados de una vez para todas las capas.

La opción de *Casillas* presenta:



Frecuencias: Las frecuencias tanto observadas como esperadas. La frecuencia observada representa la frecuencia empírica, es decir la efectivamente existente dada la distribución de los datos. La frecuencia esperada representa la frecuencia esperada o teórica de cada celda de la tabla.

Porcentajes: Los porcentajes pueden ser presentados por fila, columna y/o sobre el total de la tabla.

Es importante considerar qué porcentaje es el que se selecciona en función de cómo ha sido construido el cuadro. Así, si seguimos la convención y ubicamos la variable independiente en las columnas, deberemos de pedir un porcentaje de columna.

Residuos: Presenta los residuos tanto estandarizados como no estandarizados, así como el ajuste estandarizado. Los residuos no

estandarizados permiten ver las diferencias entre las frecuencias observadas y las teóricas. Los residuos estandarizados toman la diferencia ajustada a la raíz cuadrada. Por último, el ajuste estandarizado toma la diferencia ajustada a la raíz cuadrada y considerando la estimación de error estándar.

Finalmente la opción de formato de tabla permite ordenar las filas de la tabla en orden ascendente o descendente.

Ejemplo

Seleccionando del menú *Analizar*, "*Estadísticos descriptivos*", llegamos a la opción "*Tablas de contingencia*". Allí seleccionamos dos variables: sexo y estado civil. Colocamos en la fila, el estado civil y en la columna la variable sexo. Dentro de la opción "*Casillas*", mantenemos las frecuencias esperadas y le pedimos los porcentajes por columna. Le damos "*Aceptar*", y el programa nos da la salida u oput.

El siguiente cuadro nos brinda información de la cantidad absoluta y porcentajes de hombres y mujeres según su estado civil, seleccionando los porcentajes por columna. El cuadro nos permite conocer la composición del estado civil en cada sexo, esto es, que cantidad de hombres y mujeres hay casados, en unión libre, en divorciados, etc.

ESTADO CIVIL * SEXO Crosstabulation

			SEXO		Total
			HOMBRES	MUJERES	
ESTADO CIVIL	CASADO	Count	818	820	1638
		% within SEXO	33,5%	31,6%	32,5%
	UNION LIBRE	Count	232	237	469
		% within SEXO	9,5%	9,1%	9,3%
	DIVORCIADO	Count	62	93	155
		% within SEXO	2,5%	3,6%	3,1%
	SEPARADO	Count	44	65	109
		% within SEXO	1,8%	2,5%	2,2%
	VIUDO	Count	79	298	377
		% within SEXO	3,2%	11,5%	7,5%
	SOLTERO	Count	1210	1081	2291
		% within SEXO	49,5%	41,7%	45,5%
Total		Count	2445	2594	5039
		% within SEXO	100,0%	100,0%	100,0%

El siguiente cuadro nos brinda información, referida a las mismas variables, pero seleccionando esta vez los porcentajes por fila. Para reconfigurar los porcentajes es necesario volver al menú tablas de contingencia, ir a "casillas" y seleccionar porcentajes por fila. Como se puede ver, el cambio en los porcentajes permite generar un nuevo tipo de información. En este caso nos

permite conocer la cantidad porcentual de hombres y mujeres que existe en cada categoría de estado civil. Dicho en otros términos, de los casados que porcentaje son hombres y que porcentaje son mujeres, de los que están en unión libre que porcentaje son hombres y cuantas son mujeres y así con el resto de las categorías de estado civil.

ESTADO CIVIL * SEXO Crosstabulation

			SEXO		Total
			HOMBRES	MUJERES	
ESTADO CIVIL	CASADO	Count	818	820	1638
		% within ESTADO CIVIL	49,9%	50,1%	100,0%
	UNION LIBRE	Count	232	237	469
		% within ESTADO CIVIL	49,5%	50,5%	100,0%
	DIVORCIADO	Count	62	93	155
		% within ESTADO CIVIL	40,0%	60,0%	100,0%
	SEPARADO	Count	44	65	109
		% within ESTADO CIVIL	40,4%	59,6%	100,0%
	VIUDO	Count	79	298	377
		% within ESTADO CIVIL	21,0%	79,0%	100,0%
	SOLTERO	Count	1210	1081	2291
		% within ESTADO CIVIL	52,8%	47,2%	100,0%
Total	Count		2445	2594	5039
	% within ESTADO CIVIL		48,5%	51,5%	100,0%

Agregando una tercera variable, nos permite ver esta relación especificada por un tercer factor. En este caso, podemos analizar el porcentaje de hombres y mujeres bajo distinta condición de estado civil (en este caso, codificada en una variable "Casados/ No Casados") en distintos departamentos, codificados por región (Montevideo/ Interior). Como puede verse en la imagen, los porcentajes por columna nos permiten conocer la composición relativa de hombres y mujeres dentro de cada estado civil a nivel regional.

SEXO * Estado Civil * Region Crosstabulation

% within Estado Civil

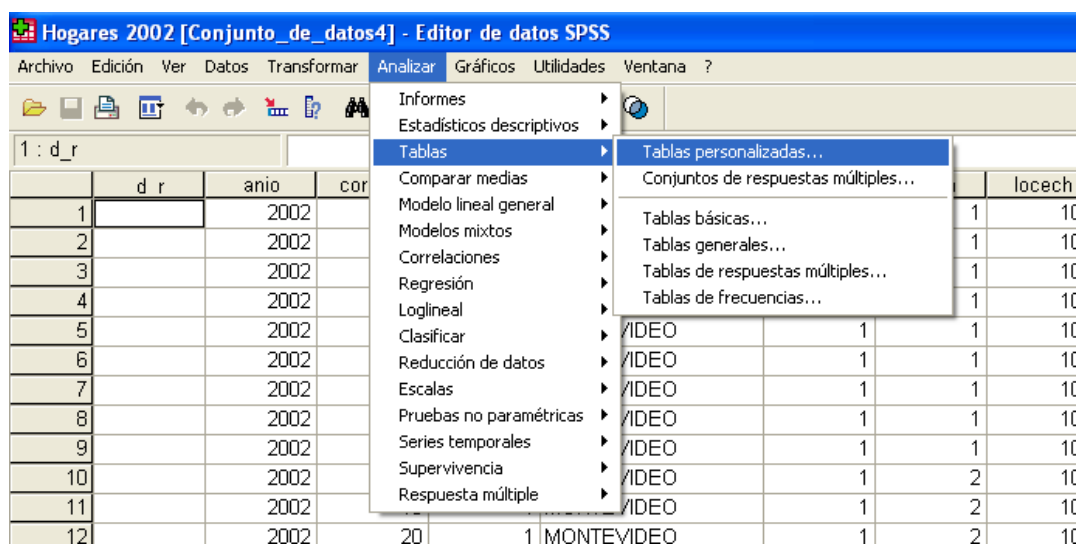
Region			Estado Civil		Total
			Casado	No Casado	
Montevideo	SEXO	HOMBRES	49,4%	48,8%	48,9%
		MUJERES	50,6%	51,3%	51,1%
	Total		100,0%	100,0%	100,0%
Interior	SEXO	HOMBRES	50,0%	47,8%	48,5%
		MUJERES	50,0%	52,2%	51,5%
	Total		100,0%	100,0%	100,0%

B. ANALISIS BI VARIADO POR MEDIO DE TABLAS PERSONALIZADAS

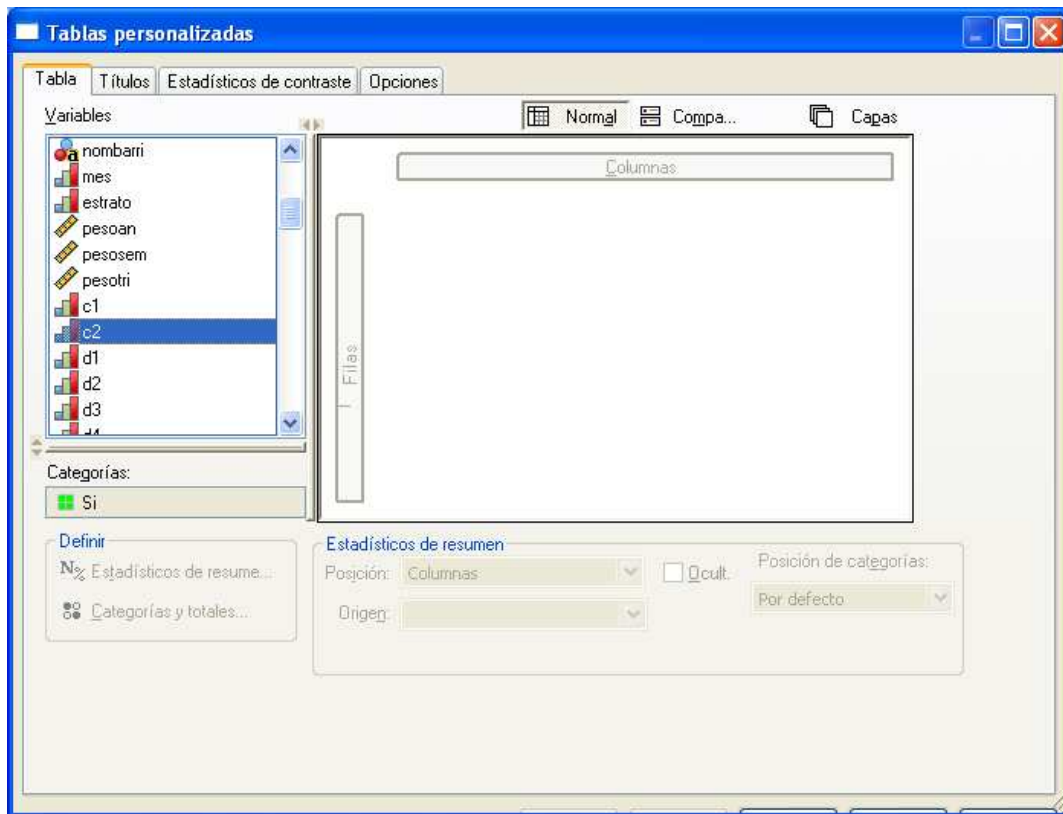
En el SPSS podemos encontrar también (según la licencia que uno tenga) la opción de tablas personalizadas.

Esta opción permite el diseño de tablas con las características que uno desee de una forma mucho más amigable.

Para encontrar esta opción (en el caso de que en nuestro SPSS esté habilitada) se debe ir a “Analizar”, “Tablas”, “Tablas personalizadas”



Allí se abre una venta en la cual aparece lo siguiente:

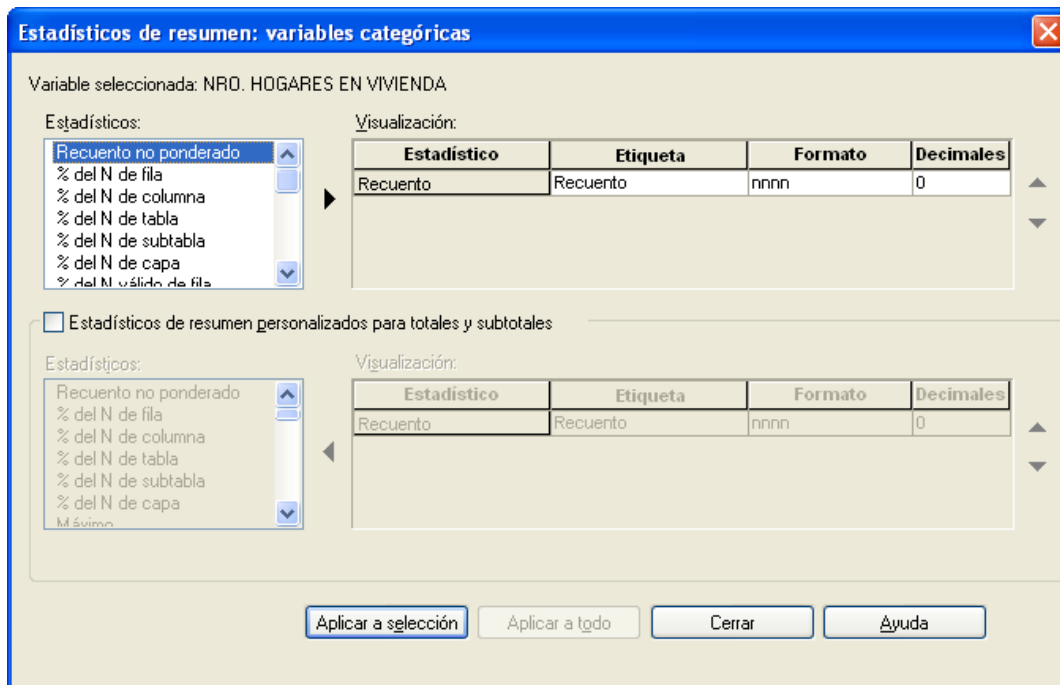


La lógica de este comando es la siguiente:

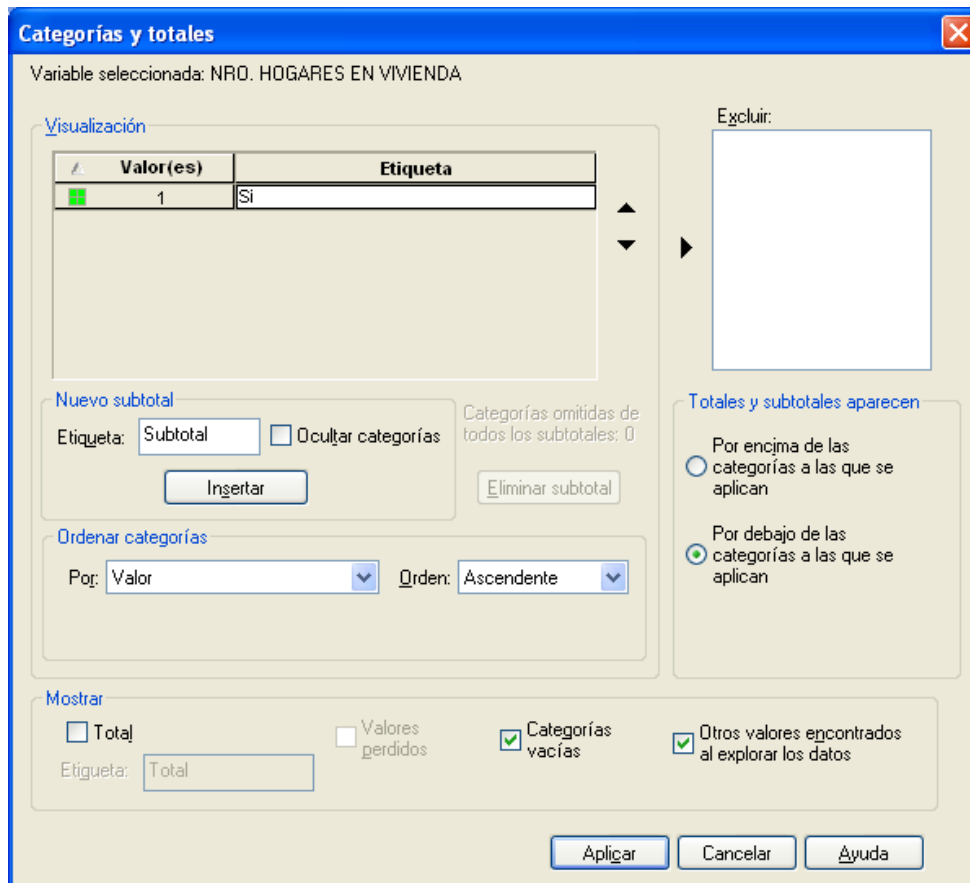
Uno selecciona una variable y la arrastra hasta el lugar en el la quiere colocar, tanto sea columna, fila o capa.

A su vez, en los recuadros inferiores *“Definir”* y *“Estadísticos de resumen”* se presentan una serie de cuadros y opciones muy amigables para el usuario a la hora de personalizar las tablas.

Por ejemplo, se puede optar por porcentajes de fila, columna, o totales para variables nominales y ordinales, y promedio, mediana u otros estadísticos para las intervalas.



También se pueden agregar subtotales en y entre variables y omitir algunos valores, entre otras funciones.



Ejercicios

1. Seleccione las variables Nivel de estudio alcanzado, Edad y Estado Civil.
 2. En base a "Tablas de contingencia" analice:
 - A) En que edad hay mayor cantidad de casados
 - B) Como es la relación entre el nivel educativo y el estado civil entre la población joven (20 a 29 años).
 - C) Agregando la variable región (Montevideo/ Interior), analice la relación entre nivel educativo y edad en distintos tramos, especificando las diferencias encontradas.
-

Bibliografía de referencia

- Bogliaccini, Juan, Cardoso, Manuel; Rodríguez Federico. Auto-texto serie estadística "Coeficientes de Asociación", Laboratorio Metodológico, UCUDAL, 2005.
 - Visauta Vinacua, B. Análisis estadístico con SPSS para Windows, McGraw-Hill, Madrid, 1997.
-