

## Análisis de Tablas de Contingencia

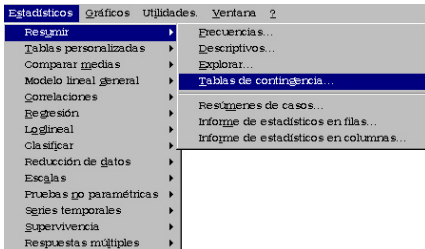


Figura 1

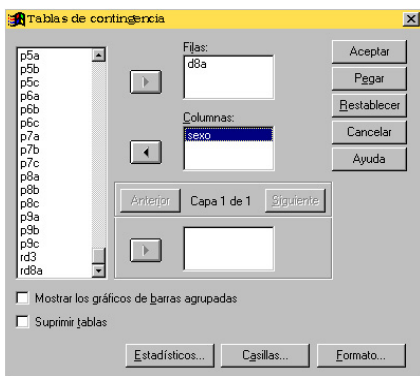


Figura 2

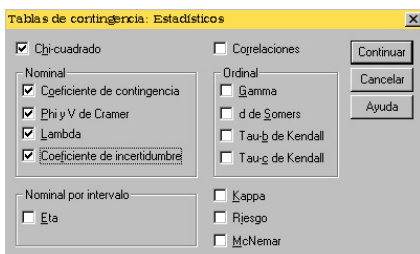


Figura 3

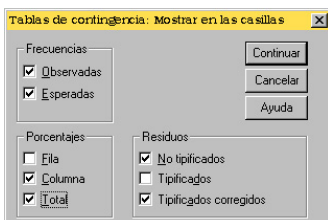


Figura 4

1.- Partimos de dos variables, que suponemos relacionadas, entre las que examinaremos si existe relación, Chi-cuadrado, su fuerza e intensidad, elegir e interpretar la medida de asociación más adecuada para la tabla, la relación entre sus categorías, análisis de los residuos tipificados de Haberman y/o diferencia de proporciones, odds ratio...

2.- Para llevar a cabo este proceso comenzaremos por construir la tabla de las dos variables seleccionadas:

Estadísticos: Resumir: Tablas de Contingencia (*figura 1*): indicar las variables correspondientes: en las filas la dependiente y en las columnas la independiente (*figura2*).

A continuación seleccionar las medidas de asociación más convenientes dependiendo del nivel de medición de las variables. Esta opción se encuentra en el "botón" de abajo del cuadro de dialogo con el nombre de "Estadísticos". Una vez que entramos en esta opción, aparecerán todas las medidas de asociación disponibles y deberemos elegir la más conveniente en función del nivel de medición de las variables (*figura 3*). Para la práctica que nosotros estamos realizando hemos seleccionado todas las medidas de asociación nominales además de la de Chi-cuadrado y poder observar las características de dichas medidas.

Una vez hemos marcados las medidas y para continuar trabajando, le damos al "botón" de continuar y ya en el cuadro de dialogo de "Tablas de Contingencia" marcamos el "botón" de "casillas" o "celdas" según en la versión de SPSS que nos encontremos. Dentro de este cuadro de dialogo (*figura 4*), nos encontramos con tres apartados en los que debemos marcar las opciones que nos interesan.

El primero, son las Frecuencias, y en este, seleccionamos los dos tipos de frecuencias: *Observadas* y *Esperadas*. El segundo, son los Porcentaje, y seleccionamos: *Columnas* y *Total*. Y en tercer lugar, los Residuos, y seleccionamos: *No Tipificados* y *Tipificados Corregidos*.

Con todo esto, ya tenemos todas las opciones necesarias para realizar el Análisis de Contingencia y poder demostrar al final del análisis, si las dos variables con las que trabajamos están o no relacionadas. Posteriormente, y si las variables estaban relacionadas, habrá que demostrar que no se trata de una Relación Espuria, sino que es una relación verdadera. Será un tema que trataremos más adelante.

## A- Estadísticos Nominales.

---

### a.1.- Chi-cuadrado:

La prueba Chi-cuadrado se utiliza para contrastar la Hipótesis Nula;  $H_0$  = las variables x e y son independientes.

Si la significación asociada a este estadístico es  $< \alpha = 0,05$  rechazaremos la hipótesis de independencia. La base de cálculo de Chi-cuadrado son las diferencias entre las frecuencias observadas y las esperadas (en caso de no asociación).

### a.2.- Razón de verosimilitud Chi-cuadrado:

Es un estadístico semejante a Chi-cuadrado, la diferencia es que en este caso el cálculo de la misma se basa en el cociente entre las frecuencias observadas y esperadas, en lugar de evaluar su diferencia como en el caso de Chi-cuadrado.

### a.3.- Phi:

El coeficiente Phi nos permite evaluar el grado de asociación entre dos variables, pero sólo esta normalizada, es decir oscila entre 0 y 1 en tablas de 2X2, en tablas con otras características toma valores  $> 1$ . En estos casos lo recomendable es utilizar el Coeficiente de Contingencia.

### a.4.- Coeficiente de Contingencia:

Esta medida es una extensión de Phi para el caso de tablas mayores a 2X2 pero tampoco es una medida normalizada para este tipo de tablas pues oscila entre 0 para el caso de no asociación y  $C_{max}$  valor que nunca alcanza el 1.

### a.5.- V de Cramer:

La V de Cramer es también una extensión del coeficiente Phi pero en este caso a diferencia del Coeficiente de Contingencia si se encuentra normalizada y oscila entre 0 y 1. El problema de este estadístico es que tiende a subestimar el grado de asociación entre las variables. Valores cercanos a 0 indican no asociación y los próximos a 1 fuerte asociación.

### a.6.- Lambda:

Este estadístico compara el error cometido en la predicción cuando únicamente se consideran los valores de la variable considerada como dependiente, y el error cometido en la predicción de esta a partir del conocimiento de los de la independiente.

Su versión asimétrica se interpreta como la proporción en que se reduce el error al predecir los valores de una de las variables (dependiente) a partir de los de la otra (independiente); mientras que la versión simétrica se interpreta como la proporción en la que se reduce el error al predecir los valores de cualquiera de las dos variables a partir de la tabla de contingencia.

Su cálculo se realiza en base a los valores modales (frecuencias mayores en la distribución), por lo que es una medida que se ve muy influida por el sesgo en las distribuciones marginales (suele ser 0 a pesar de que existe asociación entre las variables). Toma valores entre 0 en el caso de no asociación y 1 en caso de fuerte asociación.

*a.7.- Coeficientes de incertidumbre:*

Es una medida semejante a la anterior en cuanto a su concepción de la asociación de las variables, en relación a la capacidad predictiva y la disminución del error de dicha predicción. La diferencia estriba en su cálculo, ya que en este caso la expresión de estos coeficientes depende de toda la distribución, y no sólo de los valores modales, por lo que sólo toma el valor 0 en casos de total independencia. Así presenta una ventaja sobre Lamda, pero es más difícil de interpretar. Oscila entre 0 y 1. Su versión asimétrica se interpreta como la proporción de incertidumbre reducida al predecir los valores de una variable a partir de los de la otra, mientras que la versión simétrica se interpreta como la proporción de incertidumbre reducida al predecir los valores de cualquiera de las dos variables mediante la tabla de contingencia.

---

*B- Estadísticos Ordinales.*

---

*b.1.- Gamma:*

Gamma es una medida del grado y tipo de asociación entre dos variables cualitativas ordinales que oscila entre -1 (a medida que aumentan los valores de una variable, los de la otra disminuyen) y 1 (a medida que aumentan los valores de una variable los de la otra también aumentan), el 0 indica no asociación. La desventaja de esta medida es que alcanza tales valores aunque la asociación no sea total, es decir, tiende a estar sobrestimada.

*b.2.- D de Somers:*

Es una extensión de Gamma, y se presenta en dos versiones una simétrica y otra asimétrica, al igual que ésta tiene el inconveniente de que puede alcanzar los valores de -1 y 1 aunque no exista asociación total entre las variables.

*b.3.- Tau-b- de Kendall:*

Es una extensión de Gamma pero presenta la ventaja de que alcanza los valores -1 y 1 (sólo para tablas cuadradas) en situaciones de total asociación.

*b.4.- Tau-c- de Kendall:*

Es una corrección de la medida anterior de modo que alcanza los valores -1 y 1 incluso en tablas no cuadradas, el problema que presenta es que tiende a subestimar la asociación real entre las variables.

#### b.5.- Kappa:

Es una medida apropiada cuando se desea comparar el comportamiento de dos variables nominales medidas en la misma escala. Es decir, en que medida se distribuyen del mismo modo o no respecto a una tercera variable. Si las variables están fuertemente relacionadas es porque tienen el mismo comportamiento, en caso contrario no. En definitiva, mide el grado de acuerdo entre los valores de las dos variables y oscila entre -1 y 1.

#### b.6.- Riesgo relativo:

Es una medida semejante a Odds ratio por tanto tiene sus inconvenientes es decir que no está normalizada si no se extrae su logaritmo. De este modo si las variables son independientes su valor es 1.

Una vez seleccionados éstos, elegiremos las cifras y porcentajes que deseamos aparezcan en cada casilla. (Ver explicación cuadro de dialogo Celdas. Figura 4).

Celdas: en general, *valores observados*, porcentajes por *columnas* y *residuos tipificados ajustados*. Recordemos que cuando éstos son mayores a 1,96 (95 NC) o 2,58 (99 NC) son significativos e indican el tipo de relación (según el signo), que existe entre las categorías implicadas.

### Neutralización.

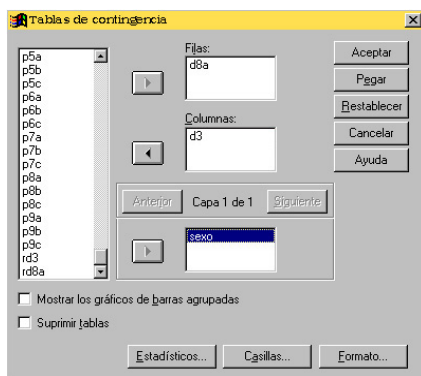


Figura 5

Procesos de neutralización e implicación de terceras variables: Análisis Multivariable de Tablas de Contingencia.

Para llevar a cabo este proceso debemos incluir una variable más en el análisis. Del mismo modo podemos controlar por más de dos variables (figura 5). Para ello superpondremos en diferentes capas las variables que queremos controlar. Si incluimos en la primera capa más de una variable, cada una de ellas se cruzará independientemente con el cruce indicado arriba.

En cuanto al *formato y presentación* de las tablas tenemos varias opciones, en general, es aconsejable pedir los nombres de las *variables* y las *etiquetas* de los valores de las variables. El orden de los valores de la escala puede indicarse en sentido *ascendente* o *descendente*. Además de ello es recomendable pedir que se especifique el *índice* de tablas para facilitar su lectura y manejo.

Una vez hemos seleccionado la variable por la que vamos a neutralizar, que la colocaremos en la última casilla del cuadro de dialogo y que en nuestro ejemplo será la variable sexo (d1), tendremos, en la salida del ordenador, un coeficiente de Pearson para cada categoría de la variable

de control (por la que estamos neutralizando), y para decir que la relación original (en nuestro ejemplo: la relación entre la d8a y d3) no es una relación espuria, el coeficiente de Pearson para cada una de las categorías de la variable de control debe tener una significación por debajo de 0,05, de lo contrario se trataría de una relación espuria.

## Análisis de Tablas de Contingencia

Resumen del procesamiento de los casos

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Practica Religiosa * SEXO	1171	97,6%	29	2,4%	1200	100,0%

Tabla de contingencia Practica Religiosa \* SEXO

			SEXO		Total
			Varon	Mujer	
Practica Religiosa	no practi.	Recuento	218	127	345
		Frecuencia esperada	164,1	180,9	345,0
		% de SEXO	39,1%	20,7%	29,5%
		% del total	18,6%	10,8%	29,5%
		Residual	53,9	-53,9	
		Residuos corregidos	6,9	-6,9	
	algo practi.	Recuento	113	98	211
		Frecuencia esperada	100,4	110,6	211,0
		% de SEXO	20,3%	16,0%	18,0%
		% del total	9,6%	8,4%	18,0%
		Residual	12,6	-12,6	
		Residuos corregidos	1,9	-1,9	

Tabla de contingencia Practica Religiosa \* SEXO

		SEXO		Total
		Varon	Mujer	
practi.	Recuento	125	149	274
	Frecuencia esperada	130,3	143,7	274,0
	% de SEXO	22,4%	24,3%	23,4%
	% del total	10,7%	12,7%	23,4%
	Residual	-5,3	5,3	
	Residuos corregidos	-,7	,7	
más practi.	Recuento	60	144	204
	Frecuencia esperada	97,0	107,0	204,0
	% de SEXO	10,8%	23,5%	17,4%
	% del total	5,1%	12,3%	17,4%
	Residual	-37,0	37,0	
	Residuos corregidos	-5,7	5,7	
muy practi	Recuento	41	96	137
	Frecuencia esperada	65,2	71,8	137,0
	% de SEXO	7,4%	15,6%	11,7%
	% del total	3,5%	8,2%	11,7%
	Residual	-24,2	24,2	
	Residuos corregidos	-4,4	4,4	
Total	Recuento	557	614	1171
	Frecuencia esperada	557,0	614,0	1171,0
	% de SEXO	100,0%	100,0%	100,0%
	% del total	47,6%	52,4%	100,0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asint. (bilateral)
Chi-cuadrado de Pearson	81,258 <sup>a</sup>	4	,000
Razón de verosimilitud	83,040	4	,000
Asociación lineal por lineal	77,465	1	,000
N de casos válidos	1171		

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 65,17.

Medidas direccionales

			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Nominal por nominal	Lambda	Simétrica	,093	,023	3,909	,000
		Practica Religiosa dependiente	,027	,020	1,325	,185
		SEXO dependiente	,190	,038	4,535	,000
	Tau de Goodman y Kruskal	Practica Religiosa dependiente	,019	,004		,000 <sup>c</sup>
		SEXO dependiente	,069	,014		,000 <sup>c</sup>
	Coeficiente de incertidumbre	Simétrica	,031	,007	4,674	,000 <sup>d</sup>
		Practica Religiosa dependiente	,023	,005	4,674	,000 <sup>d</sup>
		SEXO dependiente	,051	,011	4,674	,000 <sup>d</sup>

a. No asumiendo la hipótesis nula.

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

c. Basado en la aproximación chi-cuadrado.

d. Probabilidad del chi-cuadrado de la razón de verosimilitud.



Medidas simétricas

		Valor	Sig. aproximada
Nominal por nominal	Phi	,263	,000
	V de Cramer	,263	,000
	Coeficiente de contingencia	,255	,000
N de casos válidos		1171	

a. No asumiendo la hipótesis nula.

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

## Análisis de Tablas de Contingencia II

### Relación Original o de partida.

Resumen del procesamiento de los casos

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Practica Religiosa * E. Civil Recodificada	1171	97,6%	29	2,4%	1200	100,0%

Tabla de contingencia Practica Religiosa \* E. Civil Recodificada

			E. Civil Recodificada				Total
			solteros	casados	Pareja, sepa, Divor	viudo	
Practica Religiosa	no practi.	Recuento	130	191	9	15	345
		Frecuencia esperada	87,2	224,2	7,7	25,9	345,0
		% de E. Civil Recodificada	43,9%	25,1%	34,6%	17,0%	29,5%
		% del total	11,1%	16,3%	,8%	1,3%	29,5%
		Residual	42,8	-33,2	1,3	-10,9	
		Residuos corregidos	6,3	-4,5	,6	-2,7	
	algo practi.	Recuento	64	129	4	14	211
		Frecuencia esperada	53,3	137,1	4,7	15,9	211,0
		% de E. Civil Recodificada	21,6%	17,0%	15,4%	15,9%	18,0%
		% del total	5,5%	11,0%	,3%	1,2%	18,0%
		Residual	10,7	-8,1	-,7	-1,9	
		Residuos corregidos	1,9	-1,3	-,4	-,5	

Tabla de contingencia Practica Religiosa \* E. Civil Recodificada

		E. Civil Recodificada				Total
		solteros	casados	Pareja, sepa, Divor	viudo	
practi.	Recuento	57	190	7	20	274
	Frecuencia esperada	69,3	178,1	6,1	20,6	274,0
	% de E. Civil Recodificada	19,3%	25,0%	26,9%	22,7%	23,4%
	% del total	4,9%	16,2%	,6%	1,7%	23,4%
	Residual	-12,3	11,9	,9	-,6	
	Residuos corregidos	-1,9	1,7	,4	-,2	
más practi.	Recuento	31	150	3	20	204
	Frecuencia esperada	51,6	132,6	4,5	15,3	204,0
	% de E. Civil Recodificada	10,5%	19,7%	11,5%	22,7%	17,4%
	% del total	2,6%	12,8%	,3%	1,7%	17,4%
	Residual	-20,6	17,4	-1,5	4,7	
	Residuos corregidos	-3,6	2,8	-,8	1,4	
muy practi	Recuento	14	101	3	19	137
	Frecuencia esperada	34,6	89,0	3,0	10,3	137,0
	% de E. Civil Recodificada	4,7%	13,3%	11,5%	21,6%	11,7%
	% del total	1,2%	8,6%	,3%	1,6%	11,7%
	Residual	-20,6	12,0	,0	8,7	
	Residuos corregidos	-4,3	2,3	,0	3,0	
Total	Recuento	296	761	26	88	1171
	Frecuencia esperada	296,0	761,0	26,0	88,0	1171,0
	% de E. Civil Recodificada	100,0%	100,0%	100,0%	100,0%	100,0%
	% del total	25,3%	65,0%	2,2%	7,5%	100,0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asint. (bilateral)
Chi-cuadrado de Pearson	70,502 <sup>a</sup>	12	,000
Razón de verosimilitud	72,105	12	,000
Asociación lineal por lineal	48,524	1	,000
N de casos válidos	1171		

a. 3 casillas (15,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 3,04.

*Detección de Relaciones Espurias*

Resumen del procesamiento de los casos

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Practica Religiosa * E. Civil Recodificada	1171	97,6%	29	2,4%	1200	100,0%

**Análisis de Tablas de Contingencia**  
Técnicas de Investigación Cuantitativa y Cualitativa en Criminología

SEXO				solteros	casados	Pareja, sepa, Dívor	viudo	Total
Varón	Practica Religiosa	no practi.	Recuento	79	129	6	4	218
			Frecuencia esperada	66,9	142,5	4,3	4,3	218,0
			% de E. Civil Recodificada	46,2%	35,4%	54,5%	36,4%	39,1%
			% del total	14,2%	23,2%	1,1%	,7%	39,1%
			Residual	12,1	-13,5	1,7	-,3	
			Residuos corregidos	2,3	-2,5	1,1	-,2	
		algo practi.	Recuento	40	70	0	3	113
			Frecuencia esperada	34,7	73,8	2,2	2,2	113,0
			% de E. Civil Recodificada	23,4%	19,2%	,0%	27,3%	20,3%
			% del total	7,2%	12,6%	,0%	,5%	20,3%
			Residual	5,3	-3,8	-2,2	,8	
			Residuos corregidos	1,2	-,9	-1,7	,6	
		practi.	Recuento	32	86	4	3	125
			Frecuencia esperada	38,4	81,7	2,5	2,5	125,0
			% de E. Civil Recodificada	18,7%	23,6%	36,4%	27,3%	22,4%
			% del total	5,7%	15,4%	,7%	,5%	22,4%
			Residual	-6,4	4,3	1,5	,5	
			Residuos corregidos	-1,4	,9	1,1	,4	
		más practi.	Recuento	16	43	0	1	60
			Frecuencia esperada	18,4	39,2	1,2	1,2	60,0
			% de E. Civil Recodificada	9,4%	11,8%	,0%	9,1%	10,8%
			% del total	2,9%	7,7%	,0%	,2%	10,8%
			Residual	-2,4	3,8	-1,2	-,2	
			Residuos corregidos	-,7	1,1	-1,2	-,2	
		muy practi.	Recuento	4	36	1	0	41
			Frecuencia esperada	12,6	26,8	,8	,8	41,0
			% de E. Civil Recodificada	2,3%	9,9%	9,1%	,0%	7,4%
			% del total	,7%	6,5%	,2%	,0%	7,4%
			Residual	-8,6	9,2	,2	-,8	
			Residuos corregidos	-3,0	3,1	,2	-,9	
	Total		Recuento	171	364	11	11	557
			Frecuencia esperada	171,0	364,0	11,0	11,0	557,0
			% de E. Civil Recodificada	100,0%	100,0%	100,0%	100,0%	100,0%
			% del total	30,7%	65,4%	2,0%	2,0%	100,0%

Tabla de contingencia Practica Religiosa \* Estado Civil Recodificada \* SEXO

			Estado Civil Recodificada					
SEXO			solteros	casados	Pareja, sepa, Divor.	Viudo	Total	
Mujer	Practica Religiosa	no practi.	Recuento	51	62	3	11	127
			Frecuencia esperada	25,9	82,1	3,1	15,9	127,0
			% de RD3	40,8%	15,6%	20,0%	14,3%	20,7%
			% del total	8,3%	10,1%	,5%	1,8%	20,7%
			Residual	25,1	-20,1	-,1	-4,9	
			Residuos corregidos	6,2	-4,2	-,1	-1,5	
		algo practi.	Recuento	24	59	4	11	98
			Frecuencia esperada	20,0	63,4	2,4	12,3	98,0
			% de RD3	19,2%	14,9%	26,7%	14,3%	16,0%
			% del total	3,9%	9,6%	,7%	1,8%	16,0%
			Residual	4,0	-4,4	1,6	-1,3	
			Residuos corregidos	1,1	-1,0	1,1	-,4	
		practi.	Recuento	25	104	3	17	149
			Frecuencia esperada	30,3	96,3	3,6	18,7	149,0
			% de RD3	20,0%	26,2%	20,0%	22,1%	24,3%
			% del total	4,1%	16,9%	,5%	2,8%	24,3%
			Residual	-5,3	7,7	-,6	-1,7	
			Residuos corregidos	-1,2	1,5	-,4	-,5	
		más practi.	Recuento	15	107	3	19	144
			Frecuencia esperada	29,3	93,1	3,5	18,1	144,0
			% de RD3	12,0%	27,0%	20,0%	24,7%	23,5%
			% del total	2,4%	17,4%	,5%	3,1%	23,5%
			Residual	-14,3	13,9	-,5	,9	
			Residuos corregidos	-3,4	2,8	-,3	,3	
		muy practi	Recuento	10	65	2	19	96
			Frecuencia esperada	19,5	62,1	2,3	12,0	96,0
			% de RD3	8,0%	16,4%	13,3%	24,7%	15,6%
			% del total	1,6%	10,6%	,3%	3,1%	15,6%
			Residual	-9,5	2,9	-,3	7,0	
			Residuos corregidos	-2,6	,7	-,2	2,3	
Total		Recuento	125	397	15	77	614	
		Frecuencia esperada	125,0	397,0	15,0	77,0	614,0	
		% de RD3	100,0%	100,0%	100,0%	100,0%	100,0%	
		% del total	20,4%	64,7%	2,4%	12,5%	100,0%	

Pruebas de chi-cuadrado

SEXO		Valor	gl	Sig. asint. (bilateral)
Varon	Chi-cuadrado de Pearson	21,773 <sup>a</sup>	12	,040
	Razón de verosimilitud	27,525	12	,006
	Asociación lineal por lineal	5,924	1	,015
	N de casos válidos	557		
Mujer	Chi-cuadrado de Pearson	53,119 <sup>b</sup>	12	,000
	Razón de verosimilitud	50,013	12	,000
	Asociación lineal por lineal	22,865	1	,000
	N de casos válidos	614		

a. 10 casillas (50,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es ,81.

b. 5 casillas (25,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 2,35.