

SEMINARIO DE POSGRADO

TÉCNICAS AVANZADAS DE INVESTIGACIÓN SOCIAL

MÓDULO 2.1 A

Análisis de Tablas de Contingencia y Coeficientes de Asociación

DE LAS TABLAS DE CONTINGENCIA AL ANÁLISIS DE ASOCIACIÓN MULTIVARIADO

**¿CÓMO ANALIZAR Y EVALUAR HIPÓTESIS
CAUSALES O DE COVARIACIÓN ENTRE
VARIABLES CUANDO LAS MISMAS ESTÁN
MEDIDAS EN ESCALA ORDINAL O NOMINAL?**



**ANÁLISIS DE TABLAS DE CONTINGENCIA
TEST DE SIGNIFICANCIA NO PARAMÉTRICOS
MEDIDAS DE ASOCIACIÓN**

MEDIDAS DE ASOCIACIÓN ENTRE VARIABLES

UNA TABLA DE CONTINGENCIA ES UNA DISTRIBUCIÓN EN FILAS Y COLUMNAS EN LA QUE LOS INDIVIDUOS DE UNA POBLACIÓN SE CLASIFICAN EN FUNCIÓN DE PARES DE OBSERVACIONES.

La tabla de contingencia es un método de representar simultáneamente dos características diferentes observados en una misma POBLACIÓN. Las dos variables son x e y , el tamaño de la muestra es n . Las categorías de x se escribirán $x.1, x.2.. X.n$, y las de y , se escribirán $y.1, y.2... y.n$. *Los individuos $X.1 Y.1$ son los que reúnes ambos atributos.* Estos valores en una tabla de doble entrada:

X	Y	X.1	X.2	...	X.n
Y.1		X.1Y.1			
Y.2					
...					
Y.n					

COMPONENTES TABLA DE UNA CONTINGENCIA

DISTRIBUCIONES MARGINALES

DISTRIBUCIONES CONDICIONALES

UN TOTAL POBLACIONAL O MUESTRAL

	VARÓN	MUJER	TOTAL
ACTIVO	28	12	40
INACTIVO	42	18	60
TOTAL	70	30	100

Celdas condicionales

Marginales (de fila)

Marginales (de columna)

N: total poblacional o muestral

TIPO DE ANÁLISIS QUE PERMITE UNA TABLA DE CONTINGENCIA

- ❑ ANÁLISIS DE PERFILES O CARACTERÍSTICAS POBLACIONALES
- ❑ ANÁLISIS DESCRIPTIVO DE GRUPOS O SEGMENTOS DE POBLACIÓN
- ❑ **ANÁLISIS DE ASOCIACIÓN / INDEPENDENCIA Y RELACIÓN ESTADÍSTICA / ANÁLISIS DE PROBABILIDADES**

X / Y	VARÓN	MUJER	TOTAL
ACTIVO	28	12	40
INACTIVO	42	18	60
TOTAL	70	30	100

X / Y	VARÓN	MUJER	TOTAL
ACTIVO	28 casos $P(VyA/N)=28\%$	12 casos $P(MyA/N)=12\%$	40 casos $P(A/N)=40\%$
INACTIVO	42 casos $P(VyI/N)=42\%$	18 casos $P(MyI/N)=18\%$	60 casos $P(I/N)=60\%$
TOTAL	70 casos $P(V/N)=70\%$	30 casos $P(M/N)=30\%$	100 casos $P(N)=100\%$

X / Y	VARÓN	MUJER	TOTAL
ACTIVO	28 casos $P(A/V)=40\%$	12 casos $P(A/M)=40\%$	40 casos $P(A/N)=40\%$
INACTIVO	42 casos $P(I/V)=60\%$	18 casos $P(I/M)=60\%$	60 casos $P(I/N)=60\%$
TOTAL	70 casos $P(V)=100\%$	30 casos $P(M)=100\%$	100 casos $P(N)=100\%$

X / Y	VARÓN	MUJER	TOTAL
ACTIVO	28 casos $P(V/A)=70\%$	12 casos $P(M/A)=30\%$	40 casos $P(A)=100\%$
INACTIVO	42 casos $P(V/I)=70\%$	18 casos $P(M/I)=30\%$	60 casos $P(I)=100\%$
TOTAL	70 casos $P(V/N)=70\%$	30 casos $P(M/N)=30\%$	100 casos $P(N)=100\%$

ASOCIACIÓN ESTADÍSTICA

La idea de asociación / relación estadística entre variables se define en general en oposición al de independencia estadística y se evalúa examinando el sentido y la fuerza de las regularidades empíricas

“Las variables X e Y (sexo y condición de actividad) no están relacionadas si las frecuencias observadas se ajustan a las esperadas bajo el supuesto de independencia estadística. Dicho de otra forma: las frecuencias relativas que poseen el atributo Y_1 (activo) no difieren entre X_1 (hombres) e X_2 (mujeres)”.

FRECUENCIAS ESPERADAS BAJO EL SUPUESTO DE INDEPENDENCIA ESTADÍSTICA

$$Fe_{AyV} = N(A/N) * N(V/N) / 100$$

$$Fe_{AyM} = N(A/N) * N(M/N) / 100$$

$$Fe_{IyV} = N(I/N) * N(V/N) / 100$$

$$Fe_{IyM} = N(I/N) * N(M/N) / 100$$

X / Y	VARÓN	MUJER	TOTAL
ACTIVO	28 (28)	12 (12)	40
INACTIVO	42 (42)	18 (18)	60
TOTAL	70	30	100

INDEPENDENCIA ESTADÍSTICA

Distribución porcentual conjunta de dos variables dicotómicas

$Y \setminus X$	B	B'	Total
A	$\rightarrow P(A/B)$	$P(A/B')$	$P(A)$
A'	$P(A'/B)$	$P(A'/B')$	$P(A')$
Total	100	100	100

Las variables X y Y serán estadísticamente independientes si se cumple que:

$$(1) \quad P(A/B) = P(A/B')$$

UN PROBLEMA DE ASOCIACIÓN ESTADÍSTICA

A MODO DE EJEMPLO

- “La participación en el mercado de trabajo está condicionada por diversos factores económicos, sociales y culturales. [...] La definición de los roles masculinos y femeninos ubica a los varones como principales responsables del sostén económico de los hogares y [...] directamente asociados al mundo laboral [...] Las mujeres [...] como principales responsables de las tareas de reproducción social en el ámbito doméstico”¹.

UN PROBLEMA DE ASOCIACIÓN ESTADÍSTICA A MODO DE EJEMPLO

Hipótesis Nula de Independencia Estadística

“Dentro de la población de 25 a 45 años la tasa de actividad no presentará diferencias por sexo”

Sexo: Varón (V) – Mujer (M)

Condición de Actividad: Activo (A) – Inactivo (I)

V —————→ **I o A**

M —————→ **I o A**

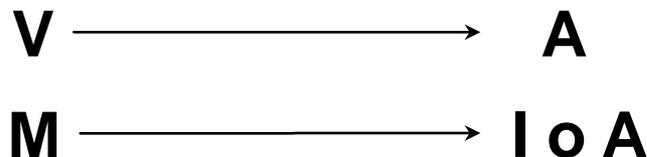
UN PROBLEMA DE ASOCIACIÓN ESTADÍSTICA A MODO DE EJEMPLO

Hipótesis de Trabajo:

- “Dentro de la población de 25 a 45 años los varones tendrán una tasa de actividad significativamente más alta que las mujeres”

Sexo: Varón (V) – Mujer (M)

Condición de Actividad: Activo (A) – Inactivo (I)



X / Y	VARÓN	MUJER	TOTAL
ACTIVO	28 casos $P(A/V)=40\%$	12 casos $P(A/M)=40\%$	40 casos Dif. %=0
INACTIVO	42 casos $P(I/V)=60\%$	18 casos $P(I/M)=60\%$	60 casos Dif. %=0
TOTAL	70 casos $P(V)=100\%$	30 casos $P(M)=100\%$	100 casos $P(N)=100\%$

X / Y	VARÓN	MUJER	TOTAL
ACTIVO	70 casos $P(A/V)=100\%$	0 casos $P(A/M)=0\%$	70 casos Dif. %= +100
INACTIVO	0 casos $P(I/V)=0\%$	30 casos $P(I/M)=100\%$	30 casos Dif. %=-100
TOTAL	70 casos $P(V)=100\%$	30 casos $P(M)=100\%$	100 casos $P(N)=100\%$

X / Y	VARÓN	MUJER	TOTAL
ACTIVO	70 casos $P(A/V)=100\%$	15 casos $P(A/M)=50\%$	85 casos Dif. %= +50
INACTIVO	0 casos $P(I/V)=0\%$	15 casos $P(I/M)=60\%$	15 casos Dif. %= -50
TOTAL	70 casos $P(V)=100\%$	30 casos $P(M)=100\%$	100 casos $P(N)=100\%$

SEMINARIO DE POSGRADO

**TEST DE HIPÓTESIS DE
INDEPENDENCIA ESTADÍSTICA Y
ANÁLISIS DE ASOCIACIÓN**

ESTADÍSTICO CHI CUADRADO

COEFICIENTES DE ASOCIACIÓN

COEFICIENTE CHI CUADRADO

$$\chi_{\text{exp}}^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \quad 0 \leq \chi_{\text{exp}}^2 \leq \infty$$

CONTINGENCIA CUADRÁTICA MEDIA

PERMITE COMPARACIONES ENTRE MUESTRAS DE DIFERENTE TAMAÑO

$$\chi_{\text{exp}}^2 / n$$

COEFICIENTE CHI-CUADRADO

EJEMPLO

PEA	SEXO		Total
	V	M	
A	100	10	110
I	20	60	80
Total	120	70	190

Frecuencias esperadas

	V	M
A	$(110 \times 120) / 190 = 69,474$	$(110 \times 70) / 190 = 40,53$
I	$(80 \times 120) / 190 = 50,526$	$(80 \times 70) / 190 = 29,47$

COEFICIENTE CHI-CUADRADO

$$\chi_{\text{exp}}^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}} = \frac{(100 - 69,47)^2}{69,474} + \frac{(20 - 50,53)^2}{50,526} + \frac{(10 - 40,53)^2}{40,53} + \frac{(60 - 29,47)^2}{29,47} = 86,47$$

Calculamos el Coeficiente de Contingencia Medio:

$$\chi_{\text{exp}}^2 / n = 86,47 / 190 = 0,455$$

PRUEBA NO PARAMÉTRICA DE INDEPENDENCIA ESTADÍSTICA

LAS PRUEBAS CHI-CUADRADO PARA TABLAS DE CONTINGENCIA EVALÚA SI EXISTE ALGÚN TIPO DE DEPENDENCIA ENTRE LOS VALORES DE DOS O MÁS VARIABLES OBSERVADAS: SI LOS VALORES DE UNA CUALQUIERA DE LAS VARIABLES APORTAN INFORMACIÓN SOBRE LOS VALORES DE LA/S OTRA/S. SUPUESTO QUE ASÍ FUERA RESULTARÁ DE INTERÉS MEDIR EL GRADO Y TIPO DE DEPENDENCIA O ASOCIACIÓN.

Prueba	Descripción
<i>J</i>-cuadrado de Pearson	Determinar si las diferencias entre las frecuencias observadas en la tabla de contingencia corresponde al cruce de los valores de las dos variables y las frecuencias esperadas, supuesto que las variables son independientes, son estadísticamente significativas.

DISTRIBUCIÓN CHI-CUADRADO

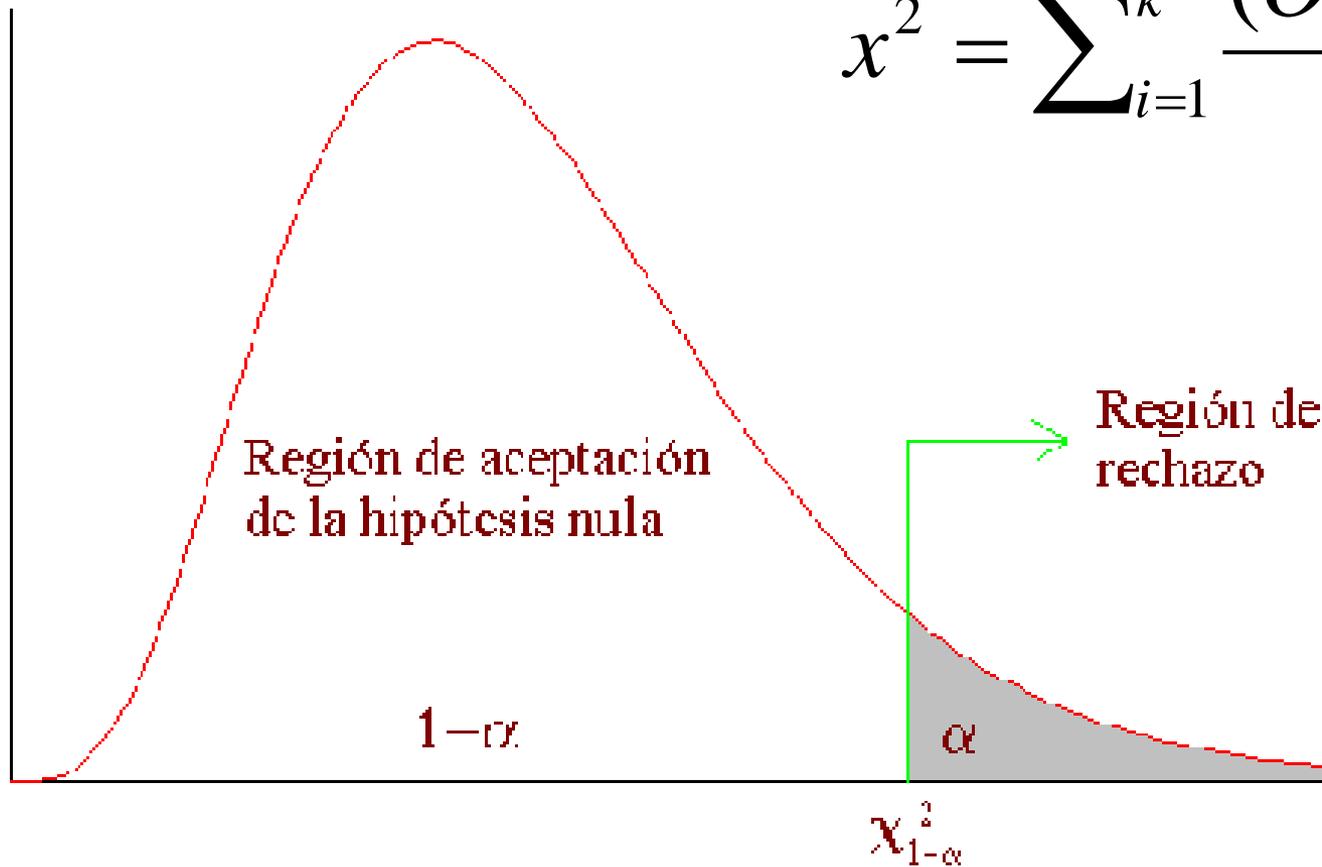
 χ^2

- Nunca adopta valores menores de 0
- Es asimétrica positiva
- Es en realidad una familia de curvas, en función de los llamados “grados de libertad”. Es decir, hay una distribución chi-cuadrado con 1 gl, una distribución chi-cuadrado con 2 gl, etc. (Nota: Los grados de libertad son siempre números positivos.)
- A medida que aumentan los grados de libertad, la distribución se hace más y más simétrica.**

Se usa para pruebas de bondad de ajuste (para comparar las puntuaciones predichas con las observadas), entre otras.

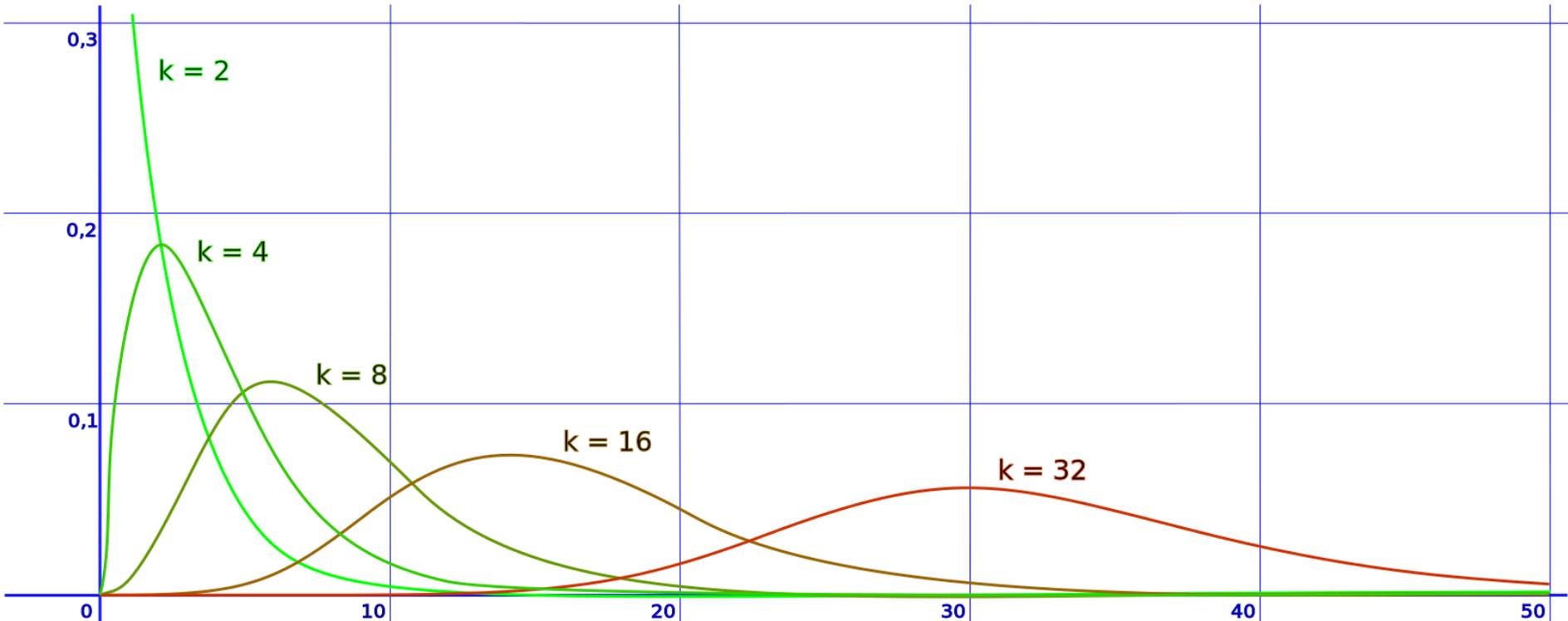
PRUEBA DE HIPÓTESIS CHI-CUADRADA

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$



PRUEBA DE HIPÓTESIS JI-CUADRADA

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$



MEDIDAS DE ASOCIACIÓN

Medida de asociación	Tabla	Escala de Medida	Observaciones
Phi	2 x 2	Nominales	Medidas basadas en chi cuadrado. Toman valores comprendidos entre 0 y 1. Evalúa hipótesis lineales (diagonal principal).
V de Cramer	f x c	Nominales	Son útiles para estimar grados de asociación entre pares de variables, sobre un mismo conjunto de individuos para n filas y columnas.
Lambda	f x c	Nominales	Toma valores entre 0 y 1. Disponen versión asimétrica. Es fácil de interpretar en términos de la proporción que se reduce el error de predicción del valor de una variable a partir de los valores de la otra (pero puede tomar valores muy bajos en tablas con asociación).
Gamma	f x c	Ordinales	Toma valores entre -1 y 1, pasando por 0. Gamma es más fácil de interpretar. Asume relaciones curvilíneas.
Tau b / c de Kendall	f x c	Ordinales	<i>Tau b</i> sólo alcanza valores extremos cuando hay asociación total y f y c son iguales. <i>Tau c</i> tiende a subestimar la relación.

Medidas de asociación para dos variables nominales

Coeficiente phi

- ✚ Medida de asociación para dos variables dicotómicas
- ✚ Basada en el coeficiente ji cuadrado
- ✚ Asume valores entre 0 y 1

Coeficientes Lambdas

- ✚ Basada en reducción del error
- ✚ Interpretación distinta de los anteriores
- ✚ Asume valores entre 0 y 1
- ✚ Proporción en que se reduce el error al predecir los valores de una variable a partir de los de la otra

Coeficiente V de Cramer

- ✚ Extensión de PHI
- ✚ Variables nominales de más de 2 categ
- ✚ Asume valores entre 0 y 1

Coeficiente Kappa

- ✚ Compara los valores de dos variables nominales tales que sus valores pueden ser los mismos
- ✚ Tablas cuadradas
- ✚ Mide el grado de acuerdo entre las dos variables
- ✚ Asume valores entre -1y 1 Valores próximos a 1 : total acuerdo. Valores próximos a -1 : total desacuerdo

Medidas de asociación para variables ordinales

Coeficiente Gamma

- ✚ Medida de asociación para dos variables cualitativas de escala ordinal
- ✚ Asume valores entre -1 y 1
- ✚ Valores próximos a 1 : fuerte asociación positiva: a medida que aumentan los valores de una variable aumentan los de la otra
- ✚ Valores próximos a -1 : fuerte asociación negativa: a medida que aumentan los valores de una variable disminuyen los de la otra
- ✚ 0 indica que no hay relación ni positiva ni negativa aunque puede haber otro tipo de relación.
- ✚ Puede alcanzar valores extremos cuando la asociación no es total

Coeficiente Tau-b de Kendall

- ✚ Extensión del Gamma Asume valores entre -1 y 1
- ✚ Alcanza valores extremos (-1 y 1) cuando la asociación es total
- ✚ Alcanza valores extremos (-1 y 1) sólo cuando las dos variables tienen el mismo número de categorías (la tabla es cuadrada)

Coeficiente Tau-c de Kendall

- ✚ Corrección del tau-b para variables con distinto tipo de categorías
- ✚ Puede subestimar el grado de asociación.

COEFICIENTE GAMMA

II) Variables Ordinales.

	Variable X			Total
Variable Y	a	b	c	a+b+c
	d	e	f	d+e+f
	g	h	i	g+h+i
Total	a+d+g	b+d+h	c+f+i	n

Calcular los pares concordantes: $n_s = a(e+f+h+i) + b(f+i) + d(h+i) + e(i)$

Calcular los pares discordantes: $n_d = g(b+c+e+f) + h(c+f) + d(b+c) + e(c)$

Gamma:
$$\gamma = \frac{n_s - n_d}{n_s + n_d} \quad -1 \leq \gamma \leq +1$$

COEFICIENTE GAMMA

Al trabajar con variables cualitativas ordinales, siempre se debe definir cual variable es la independiente(X) y cual es la variable dependiente(Y). O sea que la "Y" depende de la "X", o lo que es lo mismo la variable "Y" se ve afectada por los cambios de la variable "X".

Cuanto afecta esos cambios a la variable "Y" dependerá del grado de asociación que tengan las variables. El grado de asociación, se observa con un número, que varía entre -1 y +1. Cuanto más próximo a cero esté el coeficiente Gamma más débil es la asociación entre las variables estudiadas (RELACIONES LINEALES O CURVILINEAS / RINCONALES).

Cuanto más próximo a -1 esté el coeficiente Gamma, indicará que al crecer "X" disminuye "Y". Cuanto más próximo a +1 esté el coeficiente Gamma, indicará que al crecer la "X", también crecerá la "Y".

Sin embargo, el coeficiente Gamma también asume máximos valores cuando se cumplen estructuras de datos para relaciones rinconales.

ESTADÍSTICOS DE ANÁLISIS BIVARIADO

A MODO DE EJEMPLO

PEA * Sexo Crosstabulation

			Sexo		Total
			Varón	Mujer	
PEA	Activo	% within PEA	56,5%	43,5%	100,0%
		% within Sexo	95,2%	65,6%	79,6%
		% of Total	45,0%	34,6%	79,6%
	Inactivo	% within PEA	11,1%	88,9%	100,0%
		% within Sexo	4,8%	34,4%	20,4%
		% of Total	2,3%	18,2%	20,4%
Total	% within PEA	47,2%	52,8%	100,0%	
	% within Sexo	100,0%	100,0%	100,0%	
	% of Total	47,2%	52,8%	100,0%	

PEA * Sexo Crosstabulation

			Sexo		Total
			Varón	Mujer	
PEA	Activo	Count	6726	5122	11848
		Expected Count	5588,8	6259,2	11848,0
		Residual	1137,2	-1137,2	
		Std. Residual	15,2	-14,4	
	Inactivo	Count	401	2860	3261
		Expected Count	1538,2	1722,8	3261,0
		Residual	-1137,2	1137,2	
		Std. Residual	-29,0	27,4	
Total	Count	7127	7982	15109	
	Expected Count	7127,0	7982,0	15109,0	

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2029,509 ^b	1	,000
Likelihood Ratio	2258,729	1	,000
Linear-by-Linear Association	2029,375	1	,000
N of Valid Cases	15109		

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 1538,23.

Distribuciones para Tablas de Contingencia y Prueba de Hipótesis Ji cuadrado

ESTADÍSTICOS DE ANÁLISIS BIVARIADO

A MODO DE EJEMPLO

Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Nominal by Nominal	Phi	,367			,000
	Cramer's V	,367			,000
	Contingency Coefficient	,344			,000
Ordinal by Measure of	Gamma	,807	,010	50,109	,000
	Kappa	,292	,006	45,050	,000
N of Valid Cases		15109			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for PEA (Activo / Inactivo)	9,366	8,385	10,462
For cohort Sexo = Varón	4,617	4,207	5,066
For cohort Sexo = Mujer	,493	,481	,505
N of Valid Cases		15109	

Coeficientes de Asociación Tablas

Directional Measures

			Value	Asymp. Std. Error	Approx. T
Nominal by Nominal	Lambda	Symmetric	,154	,010	14,843
		PEA Dependent	,000	,000	.
		Sexo Dependent	,225	,013	14,843
Goodman and Kruskal tau		PEA Dependent	,134	,005	
		Sexo Dependent	,134	,004	
Uncertainty Coefficient		Symmetric	,123	,005	26,532
		PEA Dependent	,143	,005	26,532
		Sexo Dependent	,108	,004	26,532