

### COEFICIENTES DE CORRELACIÓN Y RESIDUALES ESTANDARIZADOS

Existen ciertas pruebas que permiten conocer, una vez que se ha establecido que hay asociación, qué tan fuerte es ésta. Entre estas medidas se encuentran el cálculo de el *coeficiente Phi* y el *coeficiente V de Cramer*.

Adicionalmente, y antes de establecer la fuerza de la asociación vale la pena saber qué tanto difiere un valor observado de uno esperado bajo la suposición de la  $H_0$ . Para estos casos sirve el cálculo de *residuales estandarizados*.

### RESIDUALES ESTANDARIZADOS

Un residual estandarizado es una medida del grado en que una frecuencia observada difiere del valor que sería esperado bajo la hipótesis de la  $H_0$ . Computacionalmente es

$$r = \frac{O - E}{\sqrt{E}}$$

(Un residual es igual al valor observado menos el valor esperado entre la raíz cuadrada del valor esperado)

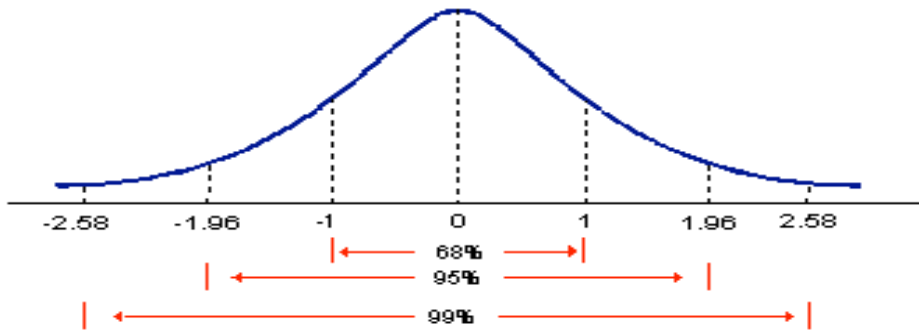
Ahora bien, cuando los grados de libertad son iguales a 1 (uno) normalmente se aplica una corrección de continuidad de 0.5, lo cual implica que:

$$r = \frac{|O - E| - 0.5}{\sqrt{E}}$$

El valor de los residuales puede ser positivo o negativo. Los residuales positivos muestran que hay más gente en esas celdas de lo que se esperaría si no hubiera asociación. En otras palabras, el valor observado es mayor al valor esperado ( $O > E$ ).

Los residuales negativos indican que hay menos gente en esas celdas de la que se esperaría si no hubiera asociación, o sea que los valores observados son menores que los esperados ( $O < E$ ).

Ahora bien, se puede saber si un residual es significativo o no comparándolo con un valor previamente determinado (convencionalmente de 1.96 o -1.96). Si el residual es más grande que 1.96 o menos que -1.96 es significativo, mientras que cualquier valor entre -1.96 y 1.96 no es significativo. ¿Qué representan estos valores en términos de una curva normal?



1. Residuales estandarizados en una curva normal

### EL COEFICIENTE V DE CRAMER

Como se mencionó al inicio, cuando los resultados de la *Ji Cuadrada* permiten establecer una asociación entre variables vale la pena investigar el grado o fuerza de asociación.

El coeficiente V de Cramer es una medida, basada en la prueba de la *Ji Cuadrada*, del grado de asociación o relación entre dos variables.

Se usa exclusivamente con datos categóricos de escalas nominales y sirve para diferentes tablas de contingencia, *independientemente de su tamaño*.

El procedimiento parte del cálculo de las frecuencias esperadas, esto es las frecuencias que se esperaría que ocurrieran si no existiera asociación entre las dos variables.

Ahora bien, mientras más grande sea la diferencia entre los valores esperados y los valores observados, más alto es el grado de asociación entre las dos variables, y en consecuencia, más grande el valor del coeficiente V de Cramer.

Computacionalmente,

$$V = \sqrt{(\chi^2/n) (m-1)}.$$

(V equivale a la raíz cuadrada de la *Ji Cuadrada* dividida por el tamaño de la muestra -*n*- por *m*, que es el valor más pequeño del número de columnas o número de filas menos uno)

El coeficiente V de Cramer tiene un valor máximo de 1 (uno) -asociación perfecta- o igual a 0 (cero) cuando las variables son independientes.

### EL COEFICIENTE PHI ( $\Phi$ )

Es una medida -basada en la prueba de la *Ji Cuadrada*- que evalúa la asociación o relación entre dos variables categóricas de una escala nominal para tablas 2 x 2.

Computacionalmente,

$$\phi = \frac{\sqrt{\chi^2}}{n}$$

(Phi es igual a la raíz cuadrada del valor de la Ji Cuadrada dividido por  $n$  -el número total de casos)

Toma el valor de 0 (cero) si no hay asociación entre las variables y el valor de 1 (uno) si hay una asociación perfecta. Entre estos dos extremos, como en el caso del valor de  $p$ , hay la posibilidad de establecer grados de asociación, tal como se observa en las siguientes tablas

**A. Valores para diferentes coeficientes      B. Interpretación de coeficientes de correlación**

<i>Correlation Coefficients</i>	<i>Interpretation</i>
-1.0 to -0.7	Strong negative association
-0.7 to -0.3	Weak negative association
-0.3 to +0.3	Little or no association
+0.3 to +0.7	Weak positive association
+0.7 to +1.0	Strong positive association

Valor del coeficiente	Interpretación
.8 to 1.0	Very strong
.6 to .8	Strong
.4 to .6	Moderate
.2 to .4	Weak
.0 to .2	Very weak

Cabe aclarar que, la mayoría de los coeficientes de correlación tienen los mismos valores de interpretación (de -1 a 1), en consecuencia se puede utilizar los mismos márgenes de valores para uno u otro coeficiente, como se aprecia en los cuadros anteriores. Asimismo, lo más importante no es que los valores sean negativos o positivos sino que tan cerca se encuentran de los valores extremos.

**Recomendación final**

Cuando el valor de la Ji Cuadrada se ubica en uno de los dos niveles de significación menos rigurosos [.05 y .025] no es recomendable hacer los cálculos de coeficientes de correlación, ni de porcentajes de desviación y residuales estandarizados, los cuales pueden aportar información relevante cuando la asociación entre las variables es fuerte, esto es cuando el resultado estadístico se encuentra en los niveles de significación más rigurosos [0.01, 0.005 y 0.001]. En estos casos sí se recomienda llevar a cabo el análisis completo.