

**Universidad Católica del Uruguay
Facultad de Ciencias Humanas**

Autotexto

Serie Estadística

Estadísticos de asociación: Chi cuadrado



**Universidad
Católica**

DAMASO A. LARRAÑAGA • URUGUAY

Laboratorio Metodológico

Versión original:

J. Bogliaccini – M. Cardoso – F. Rodríguez

Autores revisión 2008:

A. de León – M. Dodel – C. Rafaniello

Módulo de Práctica de Análisis

Serie Estadística

Tema: Chi cuadrado

Descripción:

Este autotexto nos permitirá realizar un ejercicio de cálculo de la **prueba de independencia chi-cuadrado**, con el fin de observar en qué medida algunos valores aparecen asociados con el género masculino y otros con el femenino.

Como ejemplo para esta Guía, utilizaremos la Encuesta Continua de Hogares (ECH) que realiza el Instituto Nacional de Estadística (INE). Dicha encuesta se practica en todo el país urbano de forma continua, es decir a lo largo de todo el año, desde hace más de 20 años consecutivos, proporcionando información para el cálculo de indicadores de la actividad laboral, de los ingresos de las personas y los hogares, sobre los años de educación alcanzados por las personas, etc. Para mayor información acerca de la metodología de dicha encuesta puede consultarse:

<http://www.ine.gub.uy/biblioteca/metodologias/ech/metodologiaech.htm>

Prueba de CHI CUADRADO

El chi cuadrado es un test para ser usado con datos medidos en nivel discreto o categórico. Es un test no paramétrico, es decir, un test que no exige que los datos se distribuyan normalmente.

El chi cuadrado trabaja generalmente con tablas 2X2, donde cruzamos comparando una variable que tiene dos categorías (por ejemplo hombre-mujer, blanco-negro, alto-bajo) con otra que también tiene dos categorías. Este tipo de tablas nos da una posibilidad de determinar si es que dos variables están relacionadas de la forma que hemos formulado en la hipótesis. En otras palabras, si existe una relación bivariada.

Por ejemplo si comparamos la variable **sexo (e1)**, con las categorías hombre-mujer con la variable **condición de actividad (conduct)**, con las categorías ocupado y desocupado, planteamos como hipótesis que los hombres y las mujeres se asocian, según el tipo de condición de actividad. El chi cuadrado nos permite verificar si esta hipótesis es o no sustentable, para cada caso.

1. Preparación de las variables para el análisis

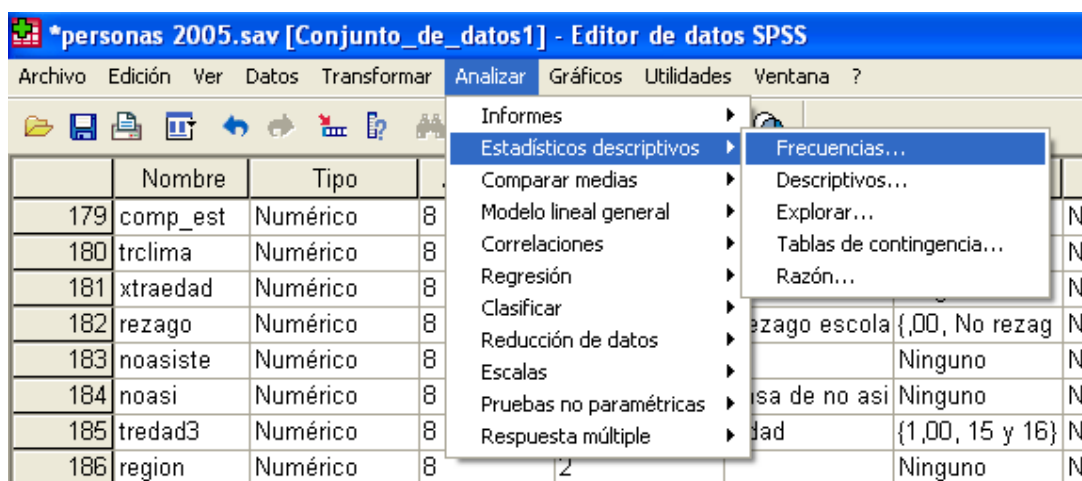
Un primer paso en todo análisis estadístico es la preparación de las variables. Para ello es recomendable tirar frecuencias¹ de las mismas para ver su distribución original y evaluar la necesidad de realizar modificaciones en las mismas.

¹ Ver autotexto Análisis univariado

En este caso es necesario lograr que ambas variables: sexo (e1) y condición de actividad (conduct) tengan únicamente dos categorías. Será necesario retocar la variable de condición de actividad.

Ud. deberá corregir la variable de condición de actividad mediante la selección de CASOS PERDIDOS (missing cases). De este modo, los casos perdidos quedarán fuera del análisis. Deberá eliminar los casos en que lógicamente no es posible una condición de actividad, tales son: los menores de 14 años y los inactivos. Para el caso de los menores de 14 años un procedimiento equivalente y conceptualmente más correcto es el de filtrar² la base de datos para que únicamente tome a aquellos que son mayores de 14 años.

De la misma forma las categorías: desocupado c/seguro, desocupados s/seguros y buscan por 1era vez deberían reagruparse en una única categoría denominada “desocupados”.



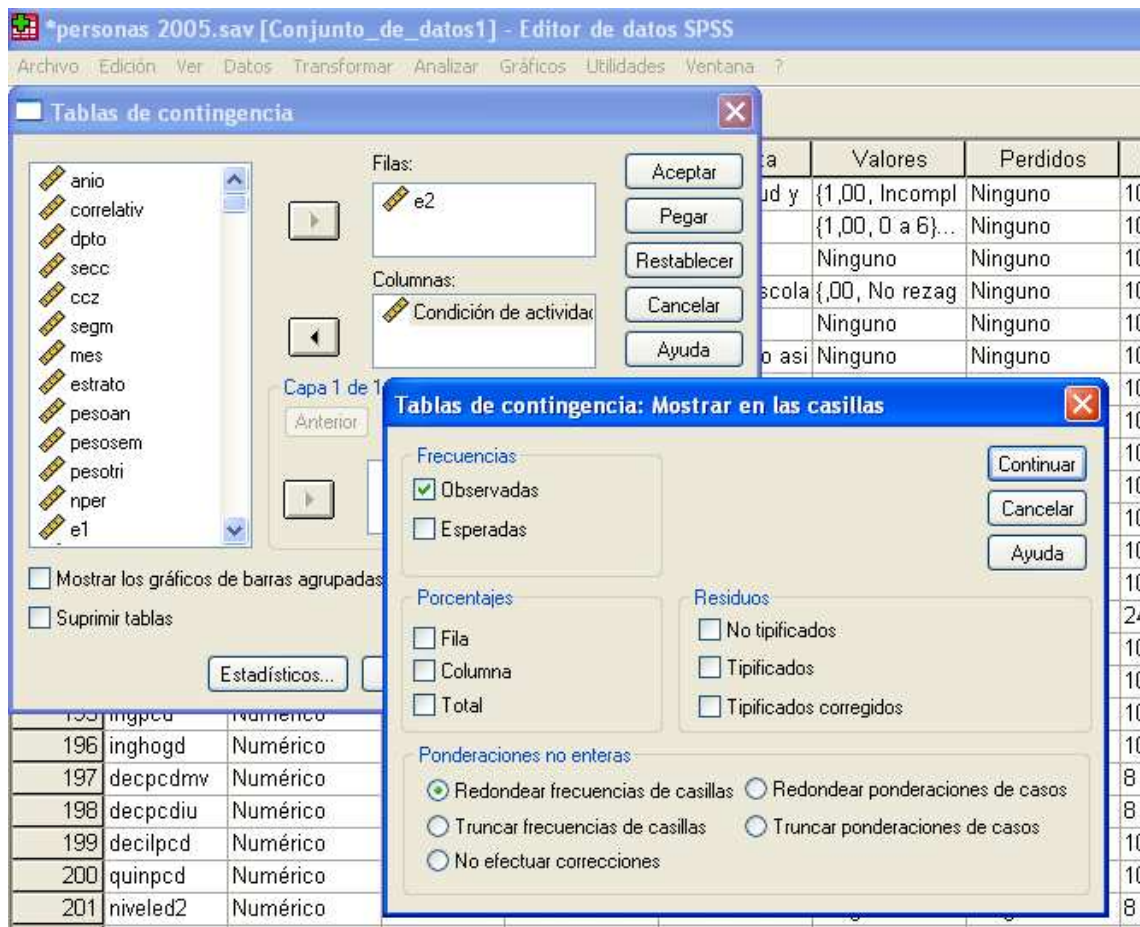
2. Tablas de contingencia

Tirar tablas de contingencia³ es necesario dado que el uso del **chi-cuadrado** se rige por ciertas restricciones. En caso de no satisfacer las variables las restricciones que plantea la prueba, será necesario volver a modificar la presentación de las mismas.

- Las variables no tendrán más de dos categorías.
- No más del 20% de las celdas pueden tener una frecuencia esperada menor que 5.
- Ninguna celda puede tener una frecuencia esperada menor que 1.

² Ver autotexto Filtrar y Segmentar archivos

³ Ver autotexto Tablas de contingencia



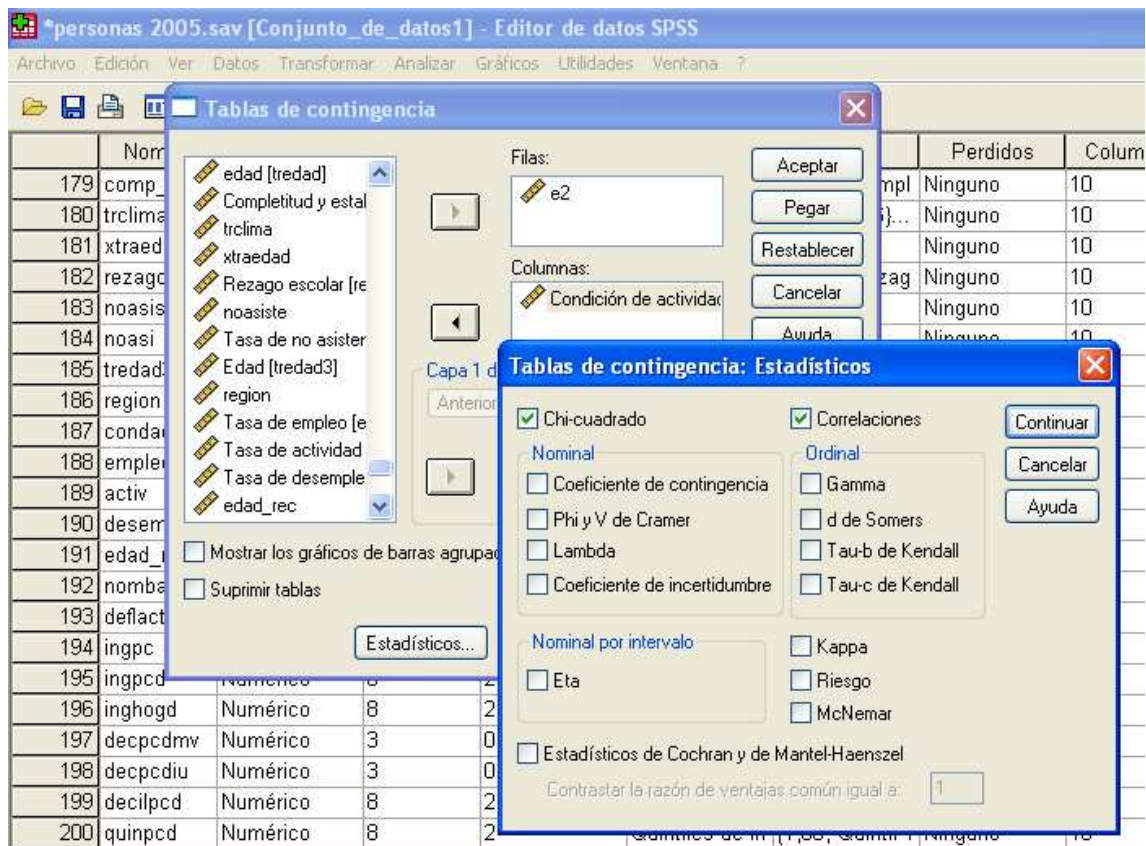
Si existen frecuencias esperadas demasiado pequeñas, estas categorías tienen que ser sacadas del análisis o combinadas con otras categorías para obtener la frecuencia esperada necesaria.

Una vez verificada y adecuada la presentación de las variables para el análisis en cuestión (en caso de ser necesario) se procede con la prueba de CHI CUADRADO.

3. Prueba CHI CUADRADO

Establecidas las variables que entran en el análisis podemos elegir el estadístico, la prueba de significación que elegiremos. Lo hacemos apretando *ANALIZAR / ESTADÍSTICOS DESCRIPTIVOS / TABLAS DE CONTINGENCIA* y seleccionando el botón “ESTADÍSTICOS”.

Haremos las pruebas **Chi cuadrado** y **Coefficiente de contingencia**. El primero para analizar si tenemos asociación estadísticamente significativa entre las variables, y el segundo para inquirir en la magnitud de la asociación.



Haremos las pruebas **Chi cuadrado** y **Coeficiente de contingencia**. El primero para analizar si tenemos asociación estadísticamente significativa entre las variables, y el segundo para inquirir en la magnitud de la asociación.

Obtenemos las siguientes salidas

Condición de actividad * SEXO Crosstabulation

| Count | | SEXO | | Total |
|------------------------|------------|---------|---------|---------|
| | | HOMBRES | MUJERES | |
| Condición de actividad | Ocupado | 511981 | 395324 | 907305 |
| | Desocupado | 79977 | 103974 | 183951 |
| Total | | 591958 | 499298 | 1091256 |

Chi-Square Tests

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|------------------------------------|------------------------|----|-----------------------|----------------------|----------------------|
| Pearson Chi-Square | 10336.333 ^b | 1 | .000 | | |
| Continuity Correction ^a | 10335.812 | 1 | .000 | | |
| Likelihood Ratio | 10302.870 | 1 | .000 | | |
| Fisher's Exact Test | | | | . | . |
| Linear-by-Linear Association | 10336.324 | 1 | .000 | | |
| N of Valid Cases | 1091256 | | | | |

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 84165.74.

Symmetric Measures

| | | Value | Asymp. Std. Error ^a | Approx. T ^b | Approx. Sig. ^c |
|----------------------|----------------------|---------|--------------------------------|------------------------|---------------------------|
| Interval by Interval | Pearson's R | .097 | .001 | 102.153 | .000 ^c |
| Ordinal by Ordinal | Spearman Correlation | .097 | .001 | 102.153 | .000 ^c |
| N of Valid Cases | | 1091256 | | | |

a. Not assuming the null hypothesis.

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

c. Based on normal approximation.

A la hora de analizar la salida obtenida para Chi cuadrado, lo importante es concentrarse en la columna de "Asymp Sig" a través de la cual podremos saber si existe o no dicha asociación entre ambas variables.

Considerando el valor en dicha columna la lógica es la siguiente: si el valor es menor a 0,05 eso quiere decir a que existe asociación entre ambas variables, mientras que si el valor es mayor a 0,05 entonces no hay asociación entre las variables.

En cuanto a la salida para el coeficiente de contingencia (quien nos determina la fuerza de la asociación en caso de que la misma exista), se deberá observar la columna de "Value" siendo la lógica la siguiente:

-si el valor es igual a 1 eso significa que hay una relación directa entre ambas variables (mientras una aumenta, aumenta también la otra)

-si el valor es igual a -1 eso significa que hay una relación inversa entre ambas variables (mientras una aumenta, la otra disminuye)

-si el valor es 0 esta es la opción neutra, la cual nos confirma que no hay relación entre las variables

Ejercicios

- Considerando el ejemplo del autotexto ¿Es posible afirmar la Hipótesis de asociación entre variables?
- En los casos en que sea posible, por favor planté alguna hipótesis explicativa de la asociación encontrada.
- Selecciona 2 variables de la Encuesta Continua de Hogares a las cuales le sea posible aplicar la prueba de independencia chi-cuadrado y realiza los procedimientos aplicados en esta guía.

Bibliografía de referencia

- Blalock, H. (1966) Estadística Social. FCE. México.
 - Peña, D. Romo, J - Introducción a la Estadística para las Ciencias Sociales. Mc Graw Hill 1997.
 - Mason y Lind - Estadística para administración y economía. Alfaomega 1998. México, D.F. 8ª edición.
-