

ANEXO VI.

DETALLE ESTADÍSTICO

El análisis de variables categóricas⁸ mediante tablas cruzadas o de contingencia, es uno de los más empleados en ciencias sociales para estudiar la existencia de relación entre variables nominales.⁹ Para ello, se han seguido los siguientes pasos:

1. Obtención de una tabla de contingencia donde cada entrada representa una variable categórica. En esta tabla, compuesta por “*f*” filas y “*c*” columnas, se refleja la frecuencia de la muestra (número o porcentaje de casos).

Se han obtenido, en algunos casos, frecuencias inferiores a 5¹⁰ sobre todo en los ítems denominados NS/NC (no sabe/no contesta). En estos casos se recomienda reducir las categorías utilizadas para obtener frecuencias más elevadas, no obstante, se ha optado por mantener el número de clases de respuesta y sus respectivas frecuencias por dos motivos:

- 1) El número de categorías empleado no es elevado, entre tres y cuatro.
- 2) Suponemos que, en determinados casos, puede existir relación entre las respuestas NS/NC de las dos variables analizadas en las tablas de contingencia.

Por otro lado, no se ha considerado conveniente plantear como “no válidos” a los individuos que no han respondido a alguna de las preguntas, ya que eso reduciría el tamaño muestral, aumentaría el error y se perdería representatividad en la muestra.

Dado que la tabla de contingencia es un instrumento necesario, pero no suficiente, para conocer la existencia de relación entre variables categóricas, se calcula, a continuación, el estadístico Chi-cuadrado, cuyos resultados establecen la independencia o asociación entre dichas variables.

⁸ Las variables categóricas son aquellas cuyos valores indican categorías o son etiquetas alfanuméricas o nombres.

⁹ Las variables nominales son variables de tipo cualitativo y discontinuo que, mediante caracteres, identifican categorías o tipos de respuesta.

¹⁰ Suele asumirse que, si existen frecuencias esperadas menores que 5, éstas no deben superar el 20 por ciento del total de frecuencias esperadas. En el caso de que ese porcentaje supere el 20 por ciento, el estadístico de Pearson debe ser interpretado con cautela.

2. A partir de la tabla de contingencia se realiza un contraste de hipótesis empleando el Test de Chi-cuadrado de Pearson:

H_0 = ambas variables son independientes

H_1 = ambas variables no son independientes

mediante el estadístico:

$$\chi^2 = \sum_{j=1}^c \sum_{i=1}^f \frac{\left(n_{ij} - n_i \frac{n_j}{n} \right)^2}{n_i \frac{n_j}{n}} \quad \text{donde,}$$

- c y f: número de columnas y filas de la tabla de contingencia
- n_i : frecuencia de la fila i
- n_j : frecuencia de la columna j
- n_{ij} : frecuencia de la fila i y la columna j
- n: frecuencia absoluta o nº total de casos

(1.1.)

A través de este estadístico (1.1.) se comparan las frecuencias observadas (n_{ii}) con las frecuencias esperadas o teóricas ($n_i(n_j/n)$), donde los grados de libertad vienen determinados por:

$$gl = (f-1)(c-1) \quad (1.2.)$$

Y el criterio de decisión establecido es el siguiente:

Aceptación de H_0 cuando

$$\chi^2 < \chi_{\alpha}^2 (f-1)(c-1) \quad (1.3.)$$

Rechazo de H_0 en caso contrario

Se rechaza la hipótesis nula (H_0), aceptando por tanto la existencia de relación entre ambas variables, si el valor del estadístico obtenido supera el valor reflejado en la

tabla de distribución de la Chi-cuadrado para $(f-1)(c-1)$ grados de libertad y un nivel de significación de α (que es del 5% para la investigación que nos ocupa).

3. En el caso de aceptación de una relación entre ambas variables, se aplican los Coeficientes Phi [ϕ] (para variables dicotómicas) y de Cramer (en el caso de variables no dicotómicas) para estudiar el grado de asociación de las variables nominales.¹¹

Coeficiente ϕ : este coeficiente se obtiene de la siguiente manera

donde,

$$\phi = \sqrt{\chi^2 / N}$$

- χ^2 : estadístico Chi-cuadrado de Pearson
- N : total de observaciones

(1.4.)

V de Cramer: queda definido por

$$V = \sqrt{\frac{\chi^2}{N \cdot m}}$$

donde,

- χ^2 : estadístico Chi-cuadrado
- N : total de la muestra
- m : se define como “mínimo $(f-1; c-1)$.” Lo forma el menor valor entre el número de filas menos 1 o el número de columnas menos 1

(1.5.)

Siguiendo a N. K. Malhotra (2004), la V de Cramer es una versión modificada del coeficiente de correlación de *phi* (ϕ) y se usa en tablas de más de 2×2 , coincidiendo, por tanto, los valores de ambos ($V = \phi$) en el caso de dichas tablas. Los valores de estos coeficientes varían entre 0 y 1 –aunque en tablas en las que una de las variables tiene más de dos niveles, *phi* puede tomar valores mayores que 1-. Una mayor proximidad a 1 indica un mayor grado de asociación entre las variables, correspondiendo el 0 a una ausencia de asociación y el 1 a una asociación perfecta entre atributos.

¹¹ Al ser medidas de asociación nominales, sólo informan sobre el grado de asociación existente, y no de la dirección o naturaleza de dicha asociación.