

# Índice general

<b>1. Medidas para Variables Nominales</b>	<b>3</b>
1.1. Medidas $\lambda_a$ y $\lambda_b$ de Goodman y Kruskal . . . . .	3
1.2. Medidas $t_a$ y $t_b$ de Goodman y Kruskal . . . . .	6
1.3. Una prueba asimétrica de homogeneidad . . . . .	7
1.4. Medidas de concordancia . . . . .	9
1.4.1. Medida de concordancia $A$ de Robinson . . . . .	10
1.4.2. Medida de concordancia $\pi$ de Scott . . . . .	12
1.4.3. Medida de concordancia $\kappa$ de Cohen . . . . .	12
1.4.4. Aplicación con varios jueces . . . . .	14
1.5. Medida de asociación $d_N^c$ de Leik y Gove . . . . .	16
<b>Bibliografía</b>	<b>22</b>



# Capítulo 1

## Medidas para Variables Nominales

Este capítulo se centra en las medidas de asociación diseñadas para las variables de nivel nominal, pero indagando en los métodos estadísticos de permutación exactos y de Monte Carlo para las medidas de asociación nominal que se basan en criterios distintos del estadístico de prueba chi-cuadrado de Pearson.

En primer lugar, se describen dos medidas asimétricas de asociación de nivel nominal propuestas por Goodman y Kruskal en 1954,  $\lambda$  y  $t$ . A continuación, el coeficiente kappa no ponderado de Cohen,  $\kappa$ , que proporciona una introducción a la medición de la concordancia, en contraste con las medidas de asociación. También se incluye en el capítulo la medida  $d_N^c$  de Leik y Gove de asociación nominal.

Algunas medidas diseñadas para variables de nivel ordinal también sirven como medidas de asociación para variables de nivel nominal cuando  $r$  (número de filas) = 2 y  $c$  (número de columnas) = 2, es decir, una tabla de contingencia  $2 \times 2$ . Otras medidas se diseñaron originalmente para tablas de contingencia  $2 \times 2$  con variables de nivel nominal, entre estas medidas de asociación están las diferencias porcentuales, las medidas  $Q$  e  $Y$  de Yule, los odds ratio y las medidas asimétricas de Somers,  $d_{yx}$  y  $d_{xy}$ .

### 1.1. Medidas $\lambda_a$ y $\lambda_b$ de Goodman y Kruskal

Un problema común al que se enfrentan muchos investigadores es el análisis de una tabla de clasificación cruzada en la que ambas variables son categóricas, ya que las variables categóricas no suelen contener tanta información como las variables de nivel ordinal o intervalo (y por supuesto, las numéricas). Las medidas basadas en la chi-cuadrado, como la  $\phi^2$  de Pearson, la  $T^2$  de Tschuprov, la  $V^2$  de Cramér y la  $C$  de Pearson, son poco útiles en la práctica debido a las dificultades de interpretación.

En 1954, Leo Goodman y William Kruskal propusieron varias medidas nuevas de asociación [Goodman and Kruskal, 1954]. Entre las medidas se encontraban dos medidas de predicción asimétricas de reducción proporcional del error para los análisis de una muestra aleatoria de dos variables categóricas:  $\lambda_a$ , para cuando se considera que  $X$  es la variable dependiente, y  $\lambda_b$ , para cuando se considera que  $Y$  es la variable dependiente.

Estas medidas de reducción proporcional del error consideran como criterio de predicción la moda, es decir, la categoría más frecuente, y como medida del error la tasa de error de clasificación entre las categorías de la variable dependiente.

Sea una tabla de contingencia  $r \times c$  como la representada a continuación:

Tabla 1.1: Clasificación cruzada de dos variables categóricas,  $X$  e  $Y$ , con  $c$  y  $r$  categorías respectivamente

	$X_1$	$X_2$	$\dots$	$X_c$	Total
$Y_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1c}$	$n_{1.}$
$Y_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2c}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$Y_r$	$n_{r1}$	$n_{r2}$	$\dots$	$n_{rc}$	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.c}$	$N$

donde  $X_j$  con  $j = 1, \dots, c$  representan las  $c$  categorías de la variable dependiente  $X$ ,  $Y_i$  con  $i = 1, \dots, r$  denotan las  $r$  categorías para la variable independiente  $Y$ ,  $n_{ij}$  la frecuencia absoluta de celda para  $i = 1, \dots, r$  y  $j = 1, \dots, c$ , y  $N$  es el total observaciones. Denotamos por un punto (.) la suma parcial de todas las filas o todas las columnas, según la posición del (.) en la lista de subíndices. Por lo tanto,  $n_{i.}$  denota el total de la frecuencia marginal de la  $i$ -ésima fila,  $i = 1, \dots, r$ , sumada en todas las columnas, y  $n_{.j}$  indica la frecuencia marginal de la  $j$ -ésima columna,  $j = 1, \dots, c$ , sumada en todas las filas.

Teniendo en cuenta la notación de la *tabla* 1.1, se definen

$$W = \sum_{i=1}^r \max(n_{i1}, n_{i2}, \dots, n_{ic}) \quad y \quad A = \max(n_{.1}, n_{.2}, \dots, n_{.c})$$

Entonces,  $\lambda_a$  (siendo  $X$  la variable dependiente) viene dado por:

$$\lambda_a = \frac{W - A}{N - A}$$

De la misma manera, se definen

$$Z = \sum_{j=1}^c \max(n_{1j}, n_{2j}, \dots, n_{rj}) \quad y \quad B = \max(n_{1.}, n_{2.}, \dots, n_{r.})$$

E igualmente,  $\lambda_b$  (siendo  $Y$  la variable dependiente) viene dado por:

$$\lambda_b = \frac{Z - B}{N - B}$$

Tanto  $\lambda_a$  como  $\lambda_b$  son medidas de reducción proporcional del error. Consideremos  $\lambda_a$  y dos posibles casos:

- Caso 1: Sólo son conocidas las categorías de la variable dependiente  $X$ .
- Caso 2: Son conocidas tanto las categorías de la variable  $X$  como las categorías de la variable independiente  $Y$ .

En el caso 1, se predice con la categoría de la variable dependiente  $X$  que tiene la mayor frecuencia marginal total (moda), que en este caso es  $A = \max(n_{.1}, \dots, n_{.c})$ . Entonces, la probabilidad de error es  $N - A$ ; definimos esto como “errores del primer tipo” o  $E_1$ . En el caso 2, se predice con la categoría de la variable dependiente  $X$  que tiene la mayor frecuencia absoluta (moda) en cada categoría de la variable independiente  $Y$ , que en este caso es, para  $Y = Y_i$ ,  $i = 1, \dots, r$ :

$$W_i = \max(n_{i1}, n_{i2}, \dots, n_{ic}).$$

Por tanto,

$$E_2^{(i)} = n_{i.} - W_i, \quad i = 1, \dots, r$$

y como consecuencia, la probabilidad de error es

$$E_2 = \sum_{i=1}^r (n_{i.} - W_i) = N - W.$$

A estos errores los definimos como “errores del segundo tipo” o  $E_2$ . Entonces,  $\lambda_a$  puede expresarse como

$$\lambda_a = \frac{E_1 - E_2}{E_1} = \frac{N - A - (N - W)}{N - A} = \frac{W - A}{N - A}$$

Como señalaron Goodman y Kruskal en 1954, se observó inmediatamente un problema con las interpretaciones tanto de  $\lambda_a$  como de  $\lambda_b$ . Dado que ambas medidas estaban basadas en los valores modales de las categorías de la variable independiente, cuando los valores modales ocurrían todos en la misma categoría de la variable dependiente,  $\lambda_a$  y  $\lambda_b$  serían cero. Así,  $\lambda_a$  y  $\lambda_b$  son iguales a cero bajo independencia, pero también podían ser iguales a cero para casos distintos de independencia. Esto hace que tanto  $\lambda_a$  como  $\lambda_b$  sean difíciles de interpretar.

De esta manera, como explicaron Goodman y Kruskal en 1954:

1.  $\lambda_a$  es indeterminada si y sólo si la población se encuentra en una columna; es decir, aparece en una categoría de la variable  $X$ .
2. En caso contrario, el valor de  $\lambda_a$  se encuentra entre 0 y 1.
3.  $\lambda_a$  es 0 si y sólo si el conocimiento de la clasificación  $Y$  no ayuda a predecir la clasificación  $X$ .
4.  $\lambda_a$  es 1 si y sólo si el conocimiento de un objeto de la categoría  $Y$  especifica completamente su categoría  $X$ , es decir, si cada fila de la tabla de clasificación cruzada contiene como máximo un valor distinto de cero.
5. En el caso de la independencia estadística,  $\lambda_a$ , cuando está determinada, es cero. Lo contrario no tiene por qué ser cierto:  $\lambda_a$  puede ser cero sin que haya independencia estadística.
6.  $\lambda_a$  no cambia con ninguna permutación de filas o columnas.

## 1.2. Medidas $t_a$ y $t_b$ de Goodman y Kruskal

Se consideran dos variables categóricas,  $X$  e  $Y$ , con la variable  $X$  como variable dependiente y la variable  $Y$  como variable independiente. La *tabla* 1.1 proporciona la notación para la clasificación cruzada.

El estadístico  $t_a$  de Goodman y Kruskal también es una medida de la reducción proporcional del error de predicción.

Sin ninguna información de  $Y$ , se predice la categoría de  $X$  aleatoriamente, según las distribuciones de frecuencias marginales. Así, si se predice en la categoría  $X_1$ , la tasa de error es:

$$\frac{N - n_{.1}}{N}.$$

Por tanto, el error esperado es

$$P[error] = \sum_{j=1}^c P[error|X_j]P[X_j] = \sum_{j=1}^c \frac{N - n_{.j}}{N} \frac{n_{.j}}{N} = \frac{1}{N^2} \sum_{j=1}^c (N - n_{.j})n_{.j}.$$

Y por consiguiente, el número de errores esperados del primer tipo viene dado por:

$$E_1 = \frac{1}{N} \sum_{j=1}^c n_{.j}(N - n_{.j}).$$

Asimismo, conociendo que la variable independiente  $Y$  toma la categoría  $Y_i$ , el número de errores esperados será:

$$\frac{1}{n_{i.}} \sum_{j=1}^c n_{ij}(n_{i.} - n_{ij}).$$

En consecuencia, el número de errores esperados del segundo tipo viene dado por:

$$E_2 = \sum_{j=1}^c \sum_{i=1}^r \frac{n_{ij}(n_{i.} - n_{ij})}{n_{i.}}.$$

El estadístico  $t_a$  de Goodman y Kruskal se define, por tanto, como:

$$t_a = \frac{E_1 - E_2}{E_1}$$

Una forma de cálculo eficiente para la  $t_a$  de Goodman y Kruskal viene dada por:

$$t_a = \frac{N \sum_{i=1}^r \sum_{j=1}^c \frac{n_{ij}^2}{n_{i.}} - \sum_{j=1}^c n_{.j}^2}{N^2 - \sum_{j=1}^c n_{.j}^2}.$$

La medida  $t_a$  indica la reducción proporcional del error de predicción dado el conocimiento de la distribución de la variable independiente,  $Y$ , conociendo la distribución de la variable dependiente,  $X$ . Como se define,  $t_a$  es un estimador puntual del parámetro

poblacional  $\tau_a$  de Goodman y Kruskal para la población de la que se obtuvo la muestra de  $N$  casos. Si la variable  $Y$  se considera la variable dependiente y la variable  $X$  la variable independiente, entonces el estadístico de Goodman y Kruskal  $t_b$  y el parámetro poblacional asociado  $\tau_b$  se definen análogamente.

Como  $t_a$  toma valores de 0 a 1, posee una interpretación clara y significativa de reducción proporcional del error [Costner, 1965], y se caracteriza por una alta validez intuitiva [Hunter, 1973]; el estadístico de prueba  $t_a$  plantea dificultades si la hipótesis nula es que  $H_0 : \tau_a = 0$  [Margolin and Light, 1974]. El problema es que la distribución muestral de  $t_a$  no es asintóticamente normal bajo la hipótesis nula. En consecuencia, la aplicabilidad de  $t_a$  de Goodman y Kruskal a las pruebas típicas de hipótesis nulas se ha visto muy limitada.

Aunque la  $t_a$  fue desarrollada por Goodman y Kruskal en 1954, no fue hasta 1963 cuando se estableció la normalidad asintótica para  $t_a$ , y hasta 1972 no se obtuvo la varianza asintótica correcta para  $t_a$ , pero sólo para  $0 < \tau_a < 1$  [Goodman and Kruskal, 1963].

### 1.3. Una prueba asimétrica de homogeneidad

A veces, se necesita determinar si las proporciones de elementos en un conjunto de categorías mutuamente excluyentes son las mismas para dos o más grupos. Cuando se extraen muestras aleatorias independientes de cada uno de los  $g \geq 2$  grupos y luego se clasifican en  $r \geq 2$  categorías mutuamente excluyentes, la prueba adecuada es una prueba de homogeneidad de las  $g$  distribuciones. Por tanto, una de las distribuciones marginales se conoce, es decir, los totales de frecuencia marginal de fila o columna que indican el número de elementos de cada uno de los  $g$  grupos. Esto se denomina muestreo multinomial *producto*, ya que la distribución de muestreo es el producto de  $g$  distribuciones multinomiales y la hipótesis nula es que las  $g$  distribuciones multinomiales son idénticas.

Una prueba de homogeneidad es diferente de una prueba de independencia, en la que se extrae una única muestra y se clasifica en ambas variables. En una prueba de independencia, ambos conjuntos de totales de frecuencias marginales se conocen sólo después de que se hayan recogido los datos. Esto se denomina muestreo multinomial simple, ya que la distribución del muestreo es una distribución multinomial [Böhning and Holling, 1989]. El test de homogeneidad más utilizado es el test chi-cuadrado de Pearson con  $gl = (r - 1)(g - 1)$  grados de libertad.

El test de homogeneidad chi-cuadrado de Pearson es un test simétrico, que produce un solo valor para una tabla de contingencia  $r \times g$ . Por el contrario, un test asimétrico produce dos valores dependiendo de la variable que se considere dependiente. Como señala Berkson, si las diferencias son todas en una dirección, una prueba simétrica como la chi-cuadrado es insensible a este hecho [Berkson, 1938].

Una prueba simétrica de homogeneidad, por su naturaleza, excluye la información conocida sobre los datos: qué variable es la independiente y qué variable es la dependiente.

Se consideran las variables  $A$  y  $B$ , con  $B$  la variable dependiente. Sean  $B_1, \dots, B_r$  las  $r \geq 2$  categorías de la variable dependiente,  $A_1, \dots, A_g$  las  $g \geq 2$  categorías de la variable independiente,  $n_{ij}$  la frecuencia de celda en la  $i$ -ésima fila y  $j$ -ésima columna,  $i = 1, \dots, r$  y  $j = 1, \dots, g$ , y  $N$  el tamaño total de la muestra. Sean  $n_{1.}, \dots, n_{r.}$  las frecuencias marginales

de la variable  $B$  y  $n_{.1}, \dots, n_{.g}$  las frecuencias marginales totales de la variable  $A$ . La clasificación cruzada de las variables  $A$  y  $B$  se muestra en la tabla siguiente:

Tabla 1.2: Clasificación cruzada de dos variables categóricas,  $A$  y  $B$ , con  $g$  y  $r$  categorías respectivamente

	$A_1$	$A_2$	$\dots$	$A_g$	Total
$B_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1g}$	$n_{1.}$
$B_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2g}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$B_r$	$n_{r1}$	$n_{r2}$	$\dots$	$n_{rg}$	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.g}$	$N$

Aunque nunca se ha propuesto como prueba de homogeneidad, la prueba asimétrica  $t_b$ , introducida por primera vez por Goodman y Kruskal en 1954 [Goodman and Kruskal, 1954], es una alternativa atractiva a la prueba simétrica de homogeneidad, la chi-cuadrado.

Si el parámetro poblacional se denota por  $\tau_b$ , se puede suponer que si  $\tau_b = 0$ , entonces las distribuciones de  $B$  en las subpoblaciones definidas por las categorías de  $A$  son homogéneas.

Así, si se desea realizar el contraste  $\begin{cases} H_0 : \tau_b = 0 \\ H_1 : \tau_b \neq 0 \end{cases}$ , el estadístico de la prueba viene dado por la versión muestral:

$$t_b = \frac{N \sum_{j=1}^g \sum_{i=1}^r \frac{n_{ij}^2}{n_{.j}} - \sum_{i=1}^r n_{i.}^2}{N^2 - \sum_{i=1}^r n_{i.}^2}$$

Si se considera que la variable dependiente es  $A$ , el estadístico de la prueba  $\begin{cases} H_0 : \tau_b = 0 \\ H_1 : \tau_b \neq 0 \end{cases}$  viene dado por:

$$t_a = \frac{N \sum_{i=1}^r \sum_{j=1}^g \frac{n_{ij}^2}{n_{i.}} - \sum_{j=1}^g n_{.j}^2}{N^2 - \sum_{j=1}^g n_{.j}^2}.$$

El estadístico de prueba  $t_b$  toma valores entre 0 y 1;  $t_b$  es 0 si y sólo si hay homogeneidad sobre las  $r$  categorías de la variable dependiente,  $B$ , para todos los  $g$  grupos, y  $t_b$  es 1 si y sólo si el conocimiento de la variable  $A_j$ , con  $j = 1, \dots, g$ , determina completamente el conocimiento de la variable  $B_i$  para  $i = 1, \dots, r$ . Del mismo modo, el estadístico de prueba  $t_a$  es 0 si y sólo si existe homogeneidad sobre las  $g$  categorías de la variable dependiente,  $A$ , para todos los  $r$  grupos, y  $t_a$  es 1 si y sólo si el conocimiento de la variable  $B_i$ , con  $i = 1, \dots, r$ , determina completamente el conocimiento de la variable  $A_j$  para  $j = 1, \dots, g$ .

Aunque no existe una equivalencia general para los estadísticos de prueba  $t_b$  o  $t_a$  con  $\chi^2$ , existen ciertas relaciones en condiciones especiales: Si  $g = 2$ , entonces  $\chi^2 = N t_b$ , y



si  $g > 2$  y  $n_{.j} = N/g$  para  $j = 1, \dots, g$ , entonces  $\chi^2 = N(g-1)t_b$ . Del mismo modo, si  $r = 2$ ,  $\chi^2 = Nt_a$ , y si  $r > 2$  y  $n_{.i} = N/r$  para  $i = 1, \dots, r$ , entonces  $\chi^2 = N(r-1)t_a$ . De lo anterior, se deduce que si  $r = g = 2$ , entonces  $t_b = t_a = \chi^2/N$ , que es el coeficiente de contingencia al cuadrado de la media de Pearson,  $\phi^2$ . Por último, cuando  $N \rightarrow \infty$ , se tiene que  $t_b(N-1)(r-1)$  y  $t_a(N-1)(g-1)$  se distribuyen según una chi-cuadrado con  $(r-1)(g-1)$  grados de libertad.

Existen tres métodos para determinar la distribución de probabilidades del estadístico de prueba  $t_b$  o  $t_a$ : procedimientos exactos, de remuestreo de Montecarlo y asintóticos. A continuación se recogen algunos comentarios sobre estos métodos, considerando sólo  $t_b$  (los métodos son análogos para  $t_a$ ):

- Valores exactos de probabilidad: Bajo la hipótesis nula,  $H_0 : \tau_b = 0$ , cada uno de los  $M$  posibles ordenamientos de los  $N$  elementos sobre las  $rg$  categorías de la tabla de contingencia es igualmente probable con distribuciones marginales fijas. Para cada ordenamiento de los datos observados en el conjunto de referencia de todos los ordenamientos posibles, se calcula el estadístico de prueba deseado. El valor exacto de la probabilidad de un estadístico  $t_b$  observado es la suma de las probabilidades puntuales hipergeométricas asociadas a valores mayores o iguales a  $t_b$ .
- Métodos de remuestreo: Una prueba exacta no es práctica desde el punto de vista computacional, excepto para muestras bastante pequeñas. Un método alternativo que evita las exigencias computacionales de una prueba exacta es una aproximación a través de remuestreo sobre permutaciones. Bajo la hipótesis nula,  $H_0 : \tau_b = 0$ , se generan y estudian un subconjunto aleatorio de Monte Carlo de todos los posibles ordenamientos igualmente probables de los datos observados. Para cada ordenamiento seleccionado al azar de los datos observados, se calcula el estadístico del test. La probabilidad de remuestreo de Montecarlo de un estadístico  $t_b$  observado es simplemente la proporción de los valores seleccionados aleatoriamente de  $t_b$  iguales o mayores que el valor observado de  $t_b$ .
- Valores asintóticos de probabilidad: Bajo la hipótesis nula,  $H_0 : \tau_b = 0$ , cuando  $N \rightarrow \infty$ ,  $t_b(N-1)(g-1)$  se distribuye según una chi-cuadrado con  $(r-1)(g-1)$  grados de libertad. El valor de la probabilidad asintótica es la proporción de la chi-cuadrado apropiada igual o mayor que el valor observado de  $t_b(N-1)(g-1)$ .

## 1.4. Medidas de concordancia

La medición de la concordancia es un caso especial de medición de la asociación entre dos o más variables. Muchos problemas de investigación estadística requieren medir la concordancia, en lugar de la asociación o la correlación. Los índices de concordancia miden el grado en que un conjunto de medidas de respuesta son idénticas a otro conjunto, es decir, concuerdan.

En 1957, Robinson [Robinson, 1957] explicó que la concordancia estadística requiere que los valores emparejados sean idénticos, mientras que la correlación sólo requiere que los valores emparejados estén descritos por alguna función matemática. Por tanto, la concordancia es una medida más restrictiva que la correlación.

Según Berry, una medida de acuerdo entre evaluadores debería, como mínimo, incorporar siete atributos básicos [Berry and Mielke, 1988]:

- Una medida de concordancia debe ser corregida por el azar, es decir, cualquier coeficiente de concordancia debe reflejar la cantidad de concordancia que excede lo que se esperaría por el azar.
- Una medida de concordancia entre evaluadores posee una ventaja añadida si es directamente aplicable a la evaluación de la fiabilidad.
- Varios investigadores han comentado la simplicidad de la distancia Euclídea para las medidas de concordancia entre evaluadores, señalando que la elevación al cuadrado de las diferencias entre los valores de las escalas es, en el mejor de los casos, cuestionable, aunque reconocen que las diferencias al cuadrado permiten interpretaciones más claras de los coeficientes [Fleiss, 1973][Krippendorff, 1970].
- Toda medida de acuerdo debe tener una base estadística. Una medida de concordancia sin una prueba de significación adecuada está muy limitada en su aplicación a situaciones prácticas de investigación. Los análisis asintóticos son interesantes y útiles, en condiciones de muestras grandes, pero suelen tener una utilidad práctica limitada cuando el tamaño de las muestras es pequeño.
- Una medida de concordancia que sirva para datos multivariantes tiene una ventaja decisiva sobre las medidas de acuerdo univariantes. Así, si un observador localiza un conjunto de datos en un espacio  $r$ -dimensional, una medida de concordancia multivariante puede determinar el grado en que un segundo observador localiza el mismo conjunto de datos en el espacio  $r$ -dimensional definido.
- Una medida de concordancia debe ser capaz de analizar los datos en cualquier nivel de medición. La medida kappa de Cohen para la concordancia entre evaluadores es, actualmente, la medida de concordancia más utilizada. Se han establecido extensiones de la kappa de Cohen a datos clasificados de forma incompleta por Iachan [Iachan, 1984], a datos ordinales totalmente clasificados y a datos de intervalo por Berry y Mielke en 1988 [Berry and Mielke, 1988].
- Una medida de concordancia debe ser capaz de evaluar la información de más de dos calificadores o jueces. Fleiss propuso una medida de concordancia para múltiples calificadores en una escala nominal [Fleiss, 1971]. Landis y Koch consideraron la concordancia entre varios calificadores en términos de una opinión mayoritaria [Landis and Koch, 1977]. Light se centró en una extensión de la medida kappa de Cohen [Cohen, 1960] de concordancia entre calificadores a múltiples calificadores que se basaba en la media de todos los valores kappa por pares [Light, 1971].

### 1.4.1. Medida de concordancia $A$ de Robinson

Una de las primeras medidas de concordancia máxima corregida fue desarrollada por W.S. Robinson en 1957 [Robinson, 1957]. Supongamos que hay  $k = 2$  jueces y califican independientemente  $N$  objetos. Robinson propuso una nueva medida de concordancia basada en el coeficiente de correlación intraclase que denominó  $A$ . Sean dos conjuntos de valoraciones con  $N$  pares de valores,  $(X_{1i}, X_{2i})$  con  $i = 1, \dots, N$ . Robinson definió  $A$  como

$$A = 1 - \frac{D}{D_{max}},$$

donde  $D$  (Desacuerdo) viene dado por:

$$D = \sum_{i=1}^N (X_{1i} - \bar{X}_i)^2 + \sum_{i=1}^N (X_{2i} - \bar{X}_i)^2$$

y  $X_{1i}$  = valor de  $X_1$  para el  $i$ -ésimo par de valoraciones,  $X_{2i}$  = valor de  $X_2$  para el  $i$ -ésimo par de valoraciones,  $\bar{X}_i$  = la media de  $X_1$  y  $X_2$  para el  $i$ -ésimo par de valoraciones, es decir,  $\bar{X}_i = \frac{1}{2}(X_{1i} + X_{2i})$ .

Para encontrar una medida de concordancia relativa, más que absoluta, Robinson estandarizó  $D$  por su rango de variación posible, dado por:

$$D_{max} = \sum_{i=1}^N (X_{1i} - \bar{X})^2 + \sum_{i=1}^N (X_{2i} - \bar{X})^2,$$

donde la media viene dada por:

$$\bar{X} = \frac{\sum_{i=1}^N X_{1i} + \sum_{i=1}^N X_{2i}}{2N}$$

### Coefficiente de correlación intraclase

El coeficiente de correlación intraclase ( $r_I$ ) entre  $N$  pares de observaciones sobre dos variables es, por definición, el momento producto ordinario de Pearson (interclase) entre  $2N$  pares de observaciones, de los cuales los primeros  $N$  son las observaciones originales, y los segundos  $N$  las observaciones originales con  $X_{1i}$  sustituyendo a  $X_{2i}$  y viceversa, con  $i = 1, \dots, N$  [Fisher, 1934]:

$$r_I = \frac{N \sum_{i=1}^N X_{1i} X_{2i} - \sum_{i=1}^N X_{1i} \sum_{i=1}^N X_{2i}}{\sqrt{[N \sum_{i=1}^N X_{1i}^2 - (\sum_{i=1}^N X_{1i})^2][N \sum_{i=1}^N X_{2i}^2 - (\sum_{i=1}^N X_{2i})^2]}}$$

Para el caso de dos variables, las relaciones entre el coeficiente de concordancia de Robinson y el coeficiente de correlación intraclase vienen dadas por:

$$r_I = 2A - 1; A = \frac{r_I + 1}{2}$$

Por tanto, en el caso de dos variables, la correlación intraclase es una función lineal simple del coeficiente de concordancia.

Para  $k > 2$  conjuntos de valoraciones, las relaciones entre el coeficiente de correlación intraclase y la  $A$  de Robinson no son tan simples, pero sigue siendo una función lineal del coeficiente de concordancia, y vienen dadas por:

$$r_I = \frac{kA - 1}{k - 1}; A = \frac{r_I(k - 1) + 1}{k}.$$

De las expresiones anteriores se observa que el valor del coeficiente intraclase no depende sólo de  $A$  sino también de  $k$ , el número variables. El rango de  $A$  de Robinson siempre incluye los valores desde 0 hasta 1, independientemente del número de observaciones. Por lo tanto, las comparaciones entre los coeficientes de concordancia basados en diferentes números de variables son equiparables. El límite superior del coeficiente de correlación intraclase es siempre la unidad, pero su límite inferior es  $-1/(k-1)$ . Para  $k = 2$  variables, el límite inferior de  $r_I$  es  $-1$ , pero para  $k = 3$  variables es  $-1/2$ , para  $k = 4$  es  $-1/3$ , para  $k = 5$  el límite inferior es  $-1/4$ , y así sucesivamente.

### 1.4.2. Medida de concordancia $\pi$ de Scott

Una de las primeras medidas de concordancia corregida por el azar<sup>1</sup> fue introducida por William Scott en 1955 [Scott, 1955]. Supongamos que dos jueces o calificadores clasifican independientemente cada una de las  $N$  observaciones en una de las  $c$  categorías. Las clasificaciones resultantes pueden mostrarse en una tabla de contingencia  $c \times c$ , con las frecuencias absolutas en cada celda. Sea  $n_i$  la frecuencia marginal de la  $i$ -ésima fila,  $i = 1, \dots, r$ ; sea  $n_j$  la frecuencia marginal de la  $j$ -ésima columna,  $j = 1, \dots, c$  y sea

$$N = \sum_{i=1}^r \sum_{j=1}^c n_{ij}$$

la frecuencia total de la tabla. El coeficiente de concordancia de Scott para los datos de nivel nominal viene dado por:

$$\pi = \frac{p_o - p_e}{1 - p_e}$$

donde

$$p_o = \frac{1}{N} \sum_{i=1}^c n_{ii} \quad y \quad p_e = \frac{1}{4N^2} \sum_{k=1}^c (n_{.k} + n_{k.})^2 .$$

Aunque la  $\pi$  de Scott es interesante desde una perspectiva histórica, esta medida ha caído en desuso. Basada en proporciones conjuntas, la  $\pi$  de Scott asume que los dos jueces tienen la misma distribución de respuestas. La medida  $\kappa$  de Cohen no hace esta suposición y, en consecuencia, ha surgido como la medida preferida de concordancia entre evaluadores corregida por el azar para dos jueces/calificadores.

### 1.4.3. Medida de concordancia $\kappa$ de Cohen

Actualmente, la medida más popular de concordancia entre dos jueces o calificadores es la medida corregida por el azar, propuesta por primera vez por Jacob Cohen en 1960 y denominada kappa [Cohen, 1960]. La kappa de Cohen mide la magnitud de la concordancia entre  $b = 2$  observadores en la asignación de  $N$  objetos a un conjunto de  $c$  categorías disjuntas y desordenadas. En 1968, Cohen propuso una versión de kappa que permitía ponderar las  $c$  categorías [Cohen, 1968]. Mientras que el kappa original (no ponderado) no distinguía entre magnitudes de desacuerdo, el kappa ponderado incorporaba la magnitud

---

<sup>1</sup>“Corregida por el azar” significa “quitar del acuerdo global”, es decir, el acuerdo que se obtendría si las valoraciones se hicieran aleatoriamente.

de cada desacuerdo y proporcionaba un crédito parcial para las discordancias cuando la concordancia no era completa. El enfoque habitual consiste en asignar pesos a cada par de desacuerdos, con pesos mayores que indican un mayor desacuerdo.

Tanto en el caso no ponderado como en el ponderado, kappa es igual a +1 cuando se produce una concordancia perfecta entre dos o más jueces, 0 cuando la concordancia es igual a la esperada en condiciones de independencia, y negativo cuando la concordancia es inferior a la esperada por azar. En esta sección se estudiará únicamente la kappa no ponderada, ya que es la que se usa normalmente para datos categóricos no ordenados.

Supongamos que dos jueces o calificadores clasifican independientemente cada una de las  $N$  observaciones en una de las  $c$  categorías desordenadas, exhaustivas y mutuamente excluyentes. Las clasificaciones resultantes pueden mostrarse en una clasificación cruzada  $c \times c$  con proporciones para las entradas de las celdas:

Tabla 1.3: Clasificación cruzada  $c \times c$  con proporciones en celdas

Fila\Columna	1	2	...	c	Total
1	$p_{11}$	$p_{12}$	$\dots$	$p_{1c}$	$p_{1.}$
2	$p_{21}$	$p_{22}$	$\dots$	$p_{2c}$	$p_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
c	$p_{c1}$	$p_{c2}$	$\dots$	$p_{cc}$	$p_{c.}$
Total	$p_{.1}$	$p_{.2}$	$\dots$	$p_{.c}$	$p_{..}$

donde  $p_{i.}$  denota la proporción marginal de la  $i$ -ésima fila,  $i = 1, \dots, c$ ;  $p_{.j}$  denota la proporción marginal de la  $j$ -ésima columna,  $j = 1, \dots, c$  y  $p_{..} = 1, 00$ . Usando esta notación, el coeficiente kappa no ponderado de Cohen para datos de nivel nominal viene dado por:

$$\kappa = \frac{p_o - p_e}{1 - p_e},$$

donde

$$p_o = \sum_{i=1}^c p_{ii} \quad y \quad p_e = \sum_{i=1}^c p_{i.} p_{.i}.$$

Considerando los pesos

$$w_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases},$$

se tiene que

$$p_o = 1 - \sum_{i=1}^c \sum_{j=1}^c w_{ij} p_{ij} \quad y \quad p_e = 1 - \sum_{i=1}^c \sum_{j=1}^c w_{ij} p_{i.} p_{.j}.$$

Por tanto:

$$\kappa = 1 - \frac{1 - p_o}{1 - p_e} = 1 - \frac{\sum_{i=1}^c \sum_{j=1}^c w_{ij} p_{ij}}{\sum_{i=1}^c \sum_{j=1}^c w_{ij} p_{i.} p_{.j}}.$$

El estimador de este coeficiente basado en frecuencias, el coeficiente kappa de Cohen muestral, puede definirse en términos de valores de frecuencia absoluta como

$$\kappa = \frac{\sum_{i=1}^c n_{ii} - \sum_{i=1}^c E_{ii}}{N - \sum_{i=1}^c E_{ii}},$$

donde  $n_{ii}$  denota el valor de frecuencia de celda observado en la diagonal principal de la tabla de concordancia  $c \times c$  y  $E_{ii}$  denota un valor de frecuencia de celda esperado en la diagonal principal, es decir,

$$E_{ii} = \frac{n_{i.} n_{.i}}{N}, \quad i = 1, \dots, c.$$

En la configuración de la tabla anterior,  $p_o$  es la proporción observada de observaciones en las que los jueces están de acuerdo,  $p_e$  es la proporción de observaciones para las que se espera una concordancia es la proporción de concordancia esperada por el azar,  $p_o - p_e$  es la proporción de concordancia más allá de la esperada por azar,  $1 - p_e$  es la proporción máxima posible de concordancia más allá de lo esperado por el azar, y el estadístico de la prueba kappa de Cohen es la proporción de concordancia entre los dos jueces, una vez eliminada la concordancia por azar.

#### 1.4.4. Aplicación con varios jueces

Originalmente, la medida  $\kappa$  de Cohen de la concordancia entre evaluadores corregida por el azar fue diseñada sólo para  $b = 2$  jueces. En esta sección, se introduce un procedimiento para calcular  $\kappa$  (no ponderado) con múltiples jueces. Aunque el procedimiento es apropiado para cualquier número de  $c \geq 2$  categorías disjuntas y desordenadas y  $b \geq 2$  jueces, la descripción del procedimiento se limita a  $b = 3$  jueces independientes.

Sean  $b = 3$  jueces que clasifican independientemente  $N$  objetos en  $c$  categorías disjuntas y desordenadas. La clasificación puede conceptualizarse como una tabla de contingencia  $c \times c \times c$  con  $c$  filas,  $c$  columnas y  $c$  niveles. Sea  $n_{ijk}$  el número de objetos clasificados en la  $i$ -ésima categoría por el juez 1, en la  $j$ -ésima categoría por el juez 2 y en la  $k$ -ésima categoría por el juez 3 y sean  $R_i$ ,  $C_j$  y  $S_k$  las frecuencias marginales de las filas, las columnas y los niveles para  $i, j, k = 1, \dots, c$ . La frecuencia total viene dada por:

$$N = \sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^c n_{ijk}.$$

Extendiendo la expresión obtenida para 2 jueces, el estadístico kappa no ponderado de Cohen para una tabla de contingencia de tres dimensiones está dado por:

$$\kappa = 1 - \frac{\sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^c w_{ijk} n_{ijk}}{\sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^c w_{ijk} R_i C_j S_k} \quad (2.1),$$

donde  $w_{ijk}$  son los “pesos” de desacuerdo asignados a cada celda para  $i, j, k = 1, \dots, c$ . Para el kappa no ponderado, los pesos de desacuerdo vienen dados por:

$$w_{ijk} = \begin{cases} 0 & i = j = k \\ 1 & \text{c.c} \end{cases}.$$

Obviamente, para el estimador del coeficiente kappa de Cohen poblacional, también se puede plantear el estudio de la significación de la medida, es decir,  $\begin{cases} H_0 : \kappa = 0 \\ H_1 : \kappa \neq 0 \end{cases}$ .

Dada una tabla de contingencia  $c \times c \times c$  con  $N$  objetos clasificados de forma cruzada por  $b = 3$  jueces independientes, una prueba de permutación exacta consiste en generar todos los ordenamientos posibles, igualmente probables, de los  $N$  objetos en las  $c^3$  celdas, preservando las distribuciones marginales de frecuencia. Para cada ordenamiento de las frecuencias de las celdas, se calcula el estadístico kappa no ponderado,  $\kappa$ , y la probabilidad exacta puntual hipergeométrica bajo la hipótesis nula,  $p(n_{ijk}|R_i, C_j, S_k, N)$ , donde

$$p(n_{ijk}|R_i, C_j, S_k, N) = \frac{(\prod_{i=1}^c R_i!)(\prod_{j=1}^c C_j!)(\prod_{k=1}^c S_k!)}{(N!)^{b-1} \prod_{i=1}^c \prod_{j=1}^c \prod_{k=1}^c n_{ijk}!}.$$

Si  $\kappa_o$  denota el valor del estadístico kappa no ponderado observado, el valor exacto de la cola de probabilidad asociada al valor  $\kappa_o$  bajo la hipótesis nula viene dado por:

$$P(\kappa_o) = \sum_{l=1}^M \Psi_l(n_{ijk}|R_i, C_j, S_k, N),$$

donde

$$\Psi_l(n_{ijk}|R_i, C_j, S_k, N) = \begin{cases} p(n_{ijk}|R_i, C_j, S_k) & \text{si } \kappa \geq \kappa_o \\ 0 & \text{c.c} \end{cases}$$

y  $M$  denota el número total de posibles ordenamientos de frecuencia de celdas igualmente probables en el conjunto de referencia de todos los ordenamientos posibles de las frecuencias de las celdas, dado las distribuciones de frecuencias marginales observadas. Cuando  $M$  es muy grande, como es típico de contingencia multidireccional, las pruebas exactas no son prácticas y es necesario recurrir a los procedimientos de remuestreo de Monte Carlo. En estos casos, una muestra aleatoria de los  $M$  posibles ordenamientos igualmente probables de las frecuencias de las celdas proporciona una comparación de los estadísticos de la prueba  $\kappa$  calculados en  $L$  tablas aleatorias multidireccionales con el estadístico de la prueba  $\kappa$  calculado en la tabla de contingencia multidireccional observada.

Un algoritmo eficiente de remuestreo de Monte Carlo para generar ordenaciones aleatorias de frecuencias de celdas para tablas de contingencia multidireccionales con distribuciones de frecuencias marginales fijas fue desarrollado por Mielke, Berry y Johnston en 2007 [Mielke et al., 2007]. Para una tabla de contingencia de tres dimensiones con  $r$  filas,  $c$  columnas y  $s$  niveles, el algoritmo de remuestreo se presenta en 12 sencillos pasos:

PASO 1. Construir una tabla de contingencia  $r \times c \times s$  a partir de los datos observados.

PASO 2. Obtener los totales de frecuencia marginal fija  $R_1, \dots, R_r, C_1, \dots, C_c, S_1, \dots, S_s$ , y el total de frecuencias  $N$ . Establecer un contador de remuestreo  $JL = 0$ , y fijar  $L$  igual al número de muestras deseado.

PASO 3. Establecer el contador de remuestreo  $JL = JL + 1$ .

PASO 4. Establecer los contadores de frecuencia marginal  $JR_i = R_i$  para  $i = 1, \dots, r$ ;  $JC_j = C_j$  para  $j = 1, \dots, c$ ;  $JS_k = S_k$  para  $k = 1, \dots, s$ , y  $M = N$ .

PASO 5. Establecer  $n_{ijk} = 0$  para  $i = 1, \dots, r, j = 1, \dots, c$ , y  $k = 1, \dots, s$ , y establecer los contadores de filas, columnas y niveles  $IR, IC$  e  $IS$  iguales a cero.

PASO 6. Crear las distribuciones de probabilidad acumulada  $PR_i, PC_j$  y  $PS_k$  a partir de los totales de frecuencia marginal ajustados  $JR_i, JC_j$  y  $JS_k$  para  $i = 1, \dots, r, j = 1, \dots, c$ , y  $k = 1, \dots, s$ , donde  $PR_1 = JR_1/M$  y  $PR_i = PR_{i-1} + JR_i/M$  para  $i = 1, \dots, r$ ;  $PC_1 = JC_1/M$  y  $PC_j = PC_{j-1} + JC_j/M$  para  $j = 1, \dots, c$  y  $PS_1 = JS_1/M$  y  $PS_k = PS_{k-1} + JS_k/M$  para  $k = 1, \dots, s$ .

PASO 7. Generar tres números pseudoaleatorios uniformes  $U_r, U_c$  y  $U_s$  sobre  $[0, 1)$  y establecer los índices de fila, columna y nivel  $i = j = k = 1$ , respectivamente.

PASO 8. Si  $U_r \leq PR_i$ , entonces  $IR = i, JR_i = JR_i - 1$ , e ir al PASO 9; en caso contrario  $i = i + 1$  y se repite el PASO 8.

PASO 9. Si  $U_c \leq PC_j$ , entonces  $IC = j, JC_j = JC_j - 1$ , e ir al PASO 10; en caso contrario,  $j = j + 1$  y se repite el PASO 9.

PASO 10. Si  $U_s \leq PS_k$ , entonces  $IS = k, JS_k = JS_k - 1$ , e ir al PASO 11; en caso contrario,  $k = k + 1$  y se repite el PASO 10.

PASO 11. Establecer  $M = M - 1$  y  $n_{IR,IC,IS} = n_{IR,IC,IS} + 1$ . Si  $M > 0$ , ir al PASO 4; en caso contrario, obtener el estadístico de prueba requerido.

PASO 12. Si  $JL < L$ , ir al PASO 3; en caso contrario, STOP.

Al terminar el procedimiento de remuestreo, se obtiene la  $\kappa$  de Cohen, como se indica en la ecuación (2.1), para cada una de las  $L$  tablas de contingencia aleatorias de tres dimensiones, dadas las distribuciones de frecuencia marginal fijas. Si  $\kappa_o$  denota el valor observado de  $\kappa$ , entonces bajo la hipótesis nula, el valor de la probabilidad aproximada del remuestreo para  $\kappa_o$  viene dado por:

$$P(\kappa_o) = \frac{1}{L} \sum_{l=1}^L \Psi_l(\kappa),$$

donde

$$\Psi_l(\kappa) = \begin{cases} 1 & \text{si } \kappa \geq \kappa_o \\ 0 & \text{c.c} \end{cases}.$$

## 1.5. Medida de asociación $d_N^c$ de Leik y Gove

En 1971, Robert Leik y Walter Gove propusieron una nueva medida de asociación nominal basada en comparaciones por pares de las diferencias entre las observaciones



[Leik and Gove, 1971]. Sugirieron una medida de asociación de reducción proporcional en el error que fuese corregida para la cantidad máxima real de asociación, dadas las distribuciones de frecuencias marginales observadas.

Consideremos dos variables de nivel nominal que se han clasificado de forma cruzada en una tabla de contingencia  $r \times c$ , donde  $r$  y  $c$  denotan el número de filas y columnas, respectivamente (véase *tabla 1.1*).

Si  $Y$  y  $X$  representan las variables de fila y columna, respectivamente, hay  $N(N-1)/2$  pares de objetos en la tabla que pueden dividirse en cinco tipos de pares exhaustivos y mutuamente excluyentes:

- Pares concordantes, es decir, pares de objetos que están clasificados en el mismo orden tanto en la variable  $X$  como en la variable  $Y$ .
- Pares discordantes, es decir, pares de objetos que se clasifican en un orden en la variable  $X$  y en el orden inverso en la variable  $Y$ .
- Pares empatados en la variable  $Y$  pero que difieren en la variable  $X$ .
- Pares empatados en la variable  $X$  pero que difieren en la variable  $Y$ .
- Pares empatados en ambas variables  $X$  e  $Y$ .

Para una tabla de contingencia  $r \times c$ , los pares concordantes vienen dados por:

$$C = \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} n_{ij} \left( \sum_{k=i+1}^r \sum_{l=j+1}^c n_{kl} \right),$$

los pares discordantes vienen dados por:

$$D = \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} n_{i, c-j+1} \left( \sum_{k=i+1}^r \sum_{l=1}^{c-j} n_{kl} \right),$$

los pares de objetos empatados en la variable  $X$  pero que difieren en la variable  $Y$  vienen dados por:

$$T_x = \sum_{i=1}^r \sum_{j=1}^{c-1} n_{ij} \left( \sum_{k=j+1}^c n_{ik} \right),$$

los pares de objetos empatados en la variable  $Y$  pero que difieren en la variable  $X$  vienen dados por:

$$T_y = \sum_{j=1}^c \sum_{i=1}^{r-1} n_{ij} \left( \sum_{k=i+1}^r n_{kj} \right),$$

y los pares de objetos empatados en la variable  $X$  y en la variable  $Y$  vienen dados por:

$$T_{xy} = \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^c n_{ij} (n_{ij} - 1).$$

Entonces,

$$C + D + T_x + T_y + T_{xy} = \frac{N(N-1)}{2} .$$

Usando la tabla de contingencia observada, se contruye la tabla de contingencia esperada bajo la hipótesis de independencia entre ambas variables de la siguiente forma:

$$E_{ij} = \frac{n_{i.}n_{.j}}{N} \quad \text{con } i = 1, \dots, r \quad \text{y } j = 1, \dots, c .$$

A partir de la tabla de contingencia esperada se obtienen los parámetros  $C'$ ,  $D'$ ,  $T'_x$ ,  $T'_y$  y  $T'_{xy}$  (véase [Berry et al., 2010]):

$$T'_{xy} = \frac{1}{2N^2} \left( \sum_{i=1}^r n_{i.}^2 \right) \left( \sum_{j=1}^c n_{.j}^2 \right) - \frac{N}{2} ; \quad T'_y = \frac{1}{2} \sum_{i=1}^r n_{i.}^2 - \frac{N}{2} - T'_{xy} ;$$

$$T'_x = \frac{1}{2} \sum_{j=1}^c n_{.j}^2 - \frac{N}{2} - T'_{xy} ; \quad C' = D' = \frac{1}{2} \left[ \frac{N(N-1)}{2} - T'_x - T'_y - T'_{xy} \right] .$$

Hay que tener en cuenta que  $C'$ ,  $D'$ ,  $T'_x$ ,  $T'_y$  y  $T'_{xy}$  se calculan sobre los totales de frecuencia marginal de la tabla de contingencia observada, que son invariantes bajo permutaciones.

El estadístico de prueba  $d_N^c$  se basa en tres tablas de contingencia: la tabla de valores observados, la tabla de valores esperados y una tabla de valores máximos que se describirá a continuación.

Un algoritmo para generar un ordenamiento de las frecuencias de las celdas en una tabla de contingencia  $r \times c$  que proporciona el valor máximo de un conjunto estadístico es:

PASO 1: Mantener las frecuencias marginales observadas de una tabla de contingencia  $r \times c$  y eliminar los valores de frecuencias de celdas ( $n_{ij}$ ).

PASO 2: Si algún par de frecuencias marginales, uno de cada conjunto de marginales, son iguales entre sí, introducir ese valor en la tabla como  $n_{ij}$  y restar el valor de los dos totales de frecuencia marginal. Repetir el PASO 2 hasta que no haya dos totales de frecuencia marginal iguales. Si todos los totales de frecuencia marginal se han reducido a cero, ir al PASO 5; de lo contrario, ir al PASO 3.

PASO 3: Observar la frecuencia marginal más grande que queda en cada conjunto e introducir el menor de los dos valores en  $n_{ij}$ . A continuación, restar ese valor (más pequeño) de los dos totales de frecuencias marginales.

PASO 4: Si todos los totales de frecuencia marginal se han reducido a cero, ir al PASO 5; En caso contrario, ir al PASO 2.

PASO 5: Establecer los valores  $n_{ij}$  restantes como 0,  $i = 1, \dots, r$  y  $j = 1, \dots, c$ .

Denotamos la doble comilla (") como una suma de pares calculada sobre los valores de frecuencias de las celdas maximizadas. Entonces, se pueden obtener los valores asociados a esta tabla:

$$C'' = \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} n_{ij} \left( \sum_{k=i+1}^r \sum_{l=j+1}^c n_{kl} \right) ; \quad D'' = \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} n_{i,c-j+1} \left( \sum_{k=i+1}^r \sum_{l=1}^{c-j} n_{kl} \right) ;$$

$$T_x'' = \sum_{i=1}^r \sum_{j=1}^{c-1} n_{ij} \left( \sum_{k=j+1}^c n_{ik} \right) ; \quad T_y'' = \sum_{j=1}^c \sum_{i=1}^{r-1} n_{ij} \left( \sum_{k=i+1}^r n_{kj} \right) ;$$

$$T_{xy}'' = \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^c n_{ij} (n_{ij} - 1) .$$

Entonces, de la misma forma que antes,

$$C'' + D'' + T_x'' + T_y'' + T_{xy}'' = \frac{N(N-1)}{2} .$$

Dados los valores observados, esperados y maximizados de  $C$ ,  $D$ ,  $T_x$ ,  $T_y$  y  $T_{xy}$ , los errores del primer tipo ( $E_1$ ) -la variación entre la independencia y la máxima asociación- vienen dados por:

$$E_1 = T_y' - T_y''$$

y los errores del segundo tipo ( $E_2$ ) -la variación entre la tabla observada y la tabla de máxima asociación- vienen dados por:

$$E_2 = T_y - T_y'' .$$

Entonces, siguiendo el patrón de las medidas de asociación de reducción proporcional del error,

$$d_N^c = \frac{E_1 - E_2}{E_1} = \frac{(T_y' - T_y'') - (T_y - T_y'')}{T_y' - T_y''} = \frac{T_y' - T_y}{T_y' - T_y''} .$$

Dado que  $d_N^c$  es una medida simétrica, el número de valores empatados en la variable  $X$  puede utilizarse en lugar del número de valores empatados en la variable  $Y$ . Así pues,

$$d_N^c = \frac{T_x' - T_x}{T_x' - T_x''} .$$

Alternativamente,  $d_N^c$  puede definirse en términos del número de valores empatados tanto en  $X$  como en  $Y$ . Así,

$$d_N^c = \frac{T_{xy}' - T_{xy}}{T_{xy}' - T_{xy}''} .$$

Por último, como los datos son categóricos,  $C$  y  $D$  pueden considerarse agrupados. Por lo tanto,

$$d_N^c = \frac{(C' + D') - (C + D)}{(C' - D') - (C'' + D'')} .$$

Como señalan Leik y Gove, para ayudar a interpretar la relación entre las variables  $X$  e  $Y$ , sería preferible determinar explícitamente el número de pares perdidos por los requisitos marginales de la tabla de contingencia. La asociación puede entonces ser definida dentro de esos límites, permitiendo que el índice alcance la unidad si las frecuencias de las celdas están tan cerca de un patrón perfecto como lo permitan las distribuciones marginales. Así pues, la proporción de casos que se considera es

$$1 - \frac{2(T_x'' + T_y'')}{N(N-1)} .$$

Para obtener un test de la significación de la medida de asociación, se puede aplicar la estrategia de los test de permutaciones, bien exactos o bien aproximados a través de la aplicación del método de Montecarlo. Consideremos la expresión

$$d_N^c = \frac{T'_y - T_y}{T'_y - T''_y} .$$

Es evidente que  $T'_y$  y  $T''_y$  son invariantes bajo permutación. Por lo tanto, la probabilidad de  $d_N^c$  bajo la hipótesis nula puede determinarse sólo por la distribución discreta de permutación de  $T_y$ , que se obtiene fácilmente de la tabla de contingencia observada. Los métodos estadísticos de permutación exacta son muy eficaces cuando sólo se calcula la parte variable del estadístico de prueba definido en cada uno de los  $M$  ordenamientos posibles de los datos observados; en este caso,  $T_y$ .

# Bibliografia

- J. Berkson. Some difficulties of interpretation encountered in the application of the chi-square test. *J. Am. Stat. Assoc.*, 33, 1938.
- Kenneth J. Berry, Janis Johnston, and Paul W. Jr. Mielke. *The Measurement of Association: A Permutation Statistical Approach*. 2010.
- K.J. Berry and P.W. Mielke. A generalization of cohen's kappa agreement measure to interval measurement and multiple raters. *Educ. Psychol. Meas.*, 48, 1988.
- D. Böhning and H. Holling. A monte carlo study on minimizing chi-square distances under the hypothesis of homogeneity or independence for a two-way contingency table. *Statistics*, 20:55–70, 1989.
- J. Cohen. *A coefficient of agreement for nominal scales*, volume 20. 1960.
- J. Cohen. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.*, 70, 1968.
- Herbert L. Costner. *Criteria for Measures of Association*. American Sociological Review, 1965.
- R.A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, 5th ed edition, 1934.
- J.L. Fleiss. Measuring nominal scale agreement among many raters. *Psych. Bull.*, 76, 1971.
- J.L. Fleiss. The equivalence of weighted kappa and the intraclass coefficient as measures of reliability. *Educ. Psychol. Meas.*, 33, 1973.
- L.A. Goodman and W.H. Kruskal. Measures of association for cross classifications. *J. Am. Stat. Assoc.*, 49, 1954.
- L.A. Goodman and W.H. Kruskal. Measures of association for cross classifications, iii: Approximate sampling theory. *J. Am. Stat. Assoc.*, 58, 1963.
- A.A. Hunter. *On the validity of measures of association: The nominal-nominal two-by-two case*. 1973.
- R. Iachan. Measures of agreement for incompletely ranked data. *Educ. Psychol. Meas.*, 44, 1984.

- K. Krippendorff. *Bivariate agreement coefficients for reliability of data*. Borgatta, E.F. (ed.) Sociological Methodology, 1970.
- J.R. Landis and G.G. Koch. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33, 1977.
- R.K. Leik and W.R. Gove. *Integrated approach to measuring association*. Costner, H.L. (ed.) Sociological Methodology, 1971.
- R.J. Light. *Measures of response agreement for qualitative data: Some generalizations and alternatives*, volume 76. Psychol. Bull, 1971.
- B.H. Margolin and R.J. Light. An analysis of variance for categorical data, ii: Small sample comparisons with chi square and other competitors. *J. Am. Stat. Assoc*, 69, 1974.
- P.W. Mielke, K.J. Berry, and J.E. Johnston. Resampling programs for multiway contingency tables with fixed marginal frequency totals. *Psychol. Rep.*, 101, 2007.
- W.S. Robinson. The statistical measurement of agreement. *Am. Sociol. Rev.*, 22, 1957.
- W.A. Scott. Reliability of content analysis: The case of nominal scale coding. *Public Opin. Quart.*, 19, 1955.