

# Índice general

<b>1. Medidas para Variables Ordinales (Parte 1)</b>	<b>3</b>
1.1. Medidas de asociación ordinal por pares . . . . .	3
1.2. Métodos estadísticos de permutación . . . . .	6
1.3. Medida $\tau_a$ de Kendall de Asociación Ordinal . . . . .	7
1.4. Medida $\tau_b$ de Kendall de Asociación Ordinal . . . . .	8
1.5. Medida $\tau_c$ de Stuart de Asociación Ordinal . . . . .	8
1.6. Medida $\gamma$ de Goodman y Kruskal . . . . .	9
1.7. Medidas $d_{yx}$ y $d_{xy}$ de Somers . . . . .	9
1.8. Medidas $d_{y.x}$ y $d_{x.y}$ de Kim . . . . .	10
1.9. Medida $e$ de Asociación Ordinal de Wilson . . . . .	10
1.10. Distribuciones de las Frecuencias Marginales . . . . .	10
<b>Bibliografía</b>	<b>11</b>



# Capítulo 1

## Medidas para Variables Ordinales (Parte 1)

Las medidas de las relaciones entre dos variables de nivel ordinal suelen ser más informativas que las medidas de las relaciones entre simples variables de nivel nominal (categóricas), ya que las categorías disjuntas y ordenadas suelen contener más información que las categorías disjuntas y desordenadas. Las medidas de asociación para dos variables de nivel ordinal suelen ser de dos tipos: las que se basan en las diferencias entre pares, como las medidas  $\tau_a$  y  $\tau_b$  de Kendall y la medida  $\gamma$  de Goodman y Kruskal, y las que se basan en criterios distintos de las diferencias entre pares, como la medida kappa ponderada de Cohen de concordancia entre evaluadores y el análisis ridit de Bross.

En este capítulo se aplican los métodos estadísticos de permutación a una variedad de medidas de asociación diseñadas para variables de nivel ordinal que se basan en todas las comparaciones dos a dos posibles. Se incluyen las medidas de asociación ordinal  $\tau_a$  y  $\tau_b$  de Kendall y  $\tau_c$  de Stuart, las medidas asimétricas  $d_{yx}$  y  $d_{xy}$  de Somers, las medidas  $d_{y.x}$  y  $d_{x.y}$  de Kim, la medida  $e$  de Wilson y el coeficiente de correlación jerárquica de Cureton.

### 1.1. Medidas de asociación ordinal por pares

Una serie de medidas de asociación para dos variables de nivel ordinal se basan en comparaciones por pares de las diferencias entre los rangos. El estadístico de prueba  $S$ , definido por Maurice Kendall en 1938, desempeña un papel importante; este estadístico se expresa a menudo como  $S = C - D$ , donde  $C$  y  $D$  indican el número de pares concordantes y discordantes, respectivamente. Sean dos variables ordinales clasificadas de forma cruzada en una tabla de contingencia  $r \times c$ , donde  $r$  y  $c$  denotan el número de filas y columnas, respectivamente. Sean  $n_{i.}$ ,  $n_{.j}$ , y  $n_{ij}$  los totales de frecuencia marginal de las filas, los totales de las frecuencias marginales de las columnas y el número de individuos en la celda  $ij$ , respectivamente, para  $i = 1, \dots, r$  y  $j = 1, \dots, c$ , y  $N$  denota el número total de individuos en la tabla de contingencia  $r \times c$ , es decir

$$n_{i.} = \sum_{j=1}^c n_{ij} , \quad n_{.j} = \sum_{i=1}^r n_{ij} \quad y \quad N = \sum_{i=1}^r \sum_{j=1}^c n_{ij} .$$

La siguiente tabla muestra una notación convencional para una tabla de contingencia  $r \times c$  para dos variables categóricas,  $x_i$  para  $i = 1, \dots, r$  e  $y_j$  para  $j = 1, \dots, c$ :

Tabla 1.1: Notación para la clasificación cruzada de dos variables categóricas,  $X$  e  $Y$ .

$X \backslash Y$	$Y_1$	$Y_2$	$\dots$	$Y_c$	<b>Total</b>
$X_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1c}$	$n_{1.}$
$X_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2c}$	$n_{2.}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$X_r$	$n_{r1}$	$n_{r2}$	$\dots$	$n_{rc}$	$n_{r.}$
<b>Total</b>	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.c}$	$N$

Si  $X$  e  $Y$  representan las variables de fila y columna, respectivamente, hay  $N(N - 1)/2$  pares de individuos en la tabla que pueden dividirse en cinco tipos de pares mutuamente exclusivos y exhaustivos: pares concordantes, pares discordantes, pares empatados a la variable  $X$  pero no a la variable  $Y$ , pares empatados a la variable  $Y$  pero no a la variable  $X$ , y pares empatados a ambas variables.

Para una tabla de contingencia  $r \times c$ , los pares concordantes (pares de objetos que están clasificados en el mismo orden tanto en la variable  $X$  como en la variable  $Y$ ) 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 (pares de objetos que se clasifican en un orden en la variable  $X$  y en el orden inverso en la variable  $Y$ ) 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,

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

Dados  $C$ ,  $D$ ,  $T_x$ ,  $T_y$  y  $N$ , se suelen definir seis medidas de asociación ordinal, cada una de las cuales tiene el mismo numerador,  $S = C - D$ , pero diferentes denominadores<sup>12</sup>. La primera de estas medidas de asociación por pares fue la  $\tau_a$  de Kendall, que es una medida simétrica de asociación ordinal y más adecuada cuando no hay pares empatados y se define simplemente como la diferencia entre las proporciones de pares concordantes y discordantes, está dada por

$$\tau_a = \frac{C}{\frac{N(N-1)}{2}} - \frac{D}{\frac{N(N-1)}{2}} = \frac{C-D}{\frac{N(N-1)}{2}} = \frac{2S}{N(N-1)} .$$

La  $\tau_b$  de Kendall amplía la  $\tau_a$  para medir la monotonía estricta en las tablas de contingencia y es la más adecuada cuando  $r = c$ . El denominador de la  $\tau_b$  se ajusta para el número de pares empatados tanto a la variable  $X$  como a la variable  $Y$ . La  $\tau_b$  de Kendall viene dada por

$$\tau_b = \frac{S}{\sqrt{(C+D+T_x)(C+D+T_y)}} .$$

La  $\tau_c$  de Stuart modifica la  $\tau_b$  de Kendall para las tablas de contingencia en las que  $r \neq c$  y viene dada por

$$\tau_c = \frac{2mS}{N^2(m-1)} ,$$

donde  $m = \min(r, c)$ . La  $\gamma$  de Goodman y Kruskal es una medida simétrica de monotonía (no estricta) en la que se ignoran los pares empatados de todos los tipos y viene dada por

$$\gamma = \frac{C-D}{C+D} = \frac{S}{C+D} .$$

Las medidas  $d_{yx}$  y  $d_{xy}$  de Somers son asimétricas de asociación ordinal. A diferencia de las cuatro medidas simétricas,  $\tau_a$ ,  $\tau_b$ ,  $\tau_c$  y  $\gamma$ , las medidas  $d_{yx}$  y  $d_{xy}$  de Somers dependen de qué variable,  $Y$  o  $X$ , se considera que es la dependiente. Si variable  $Y$  es la variable dependiente, entonces

$$d_{yx} = \frac{S}{C+D+T_y} , \quad (3.1)$$

y si la variable  $X$  es la variable dependiente, entonces

<sup>1</sup>El número de pares empatados en ambas variables  $X$  e  $Y$  ( $T_{xy}$ ) no se utiliza en ninguna de las seis medidas.

<sup>2</sup>En realidad hay más de seis medidas de asociación ordinal basadas en comparaciones por pares; aquí sólo se tratan las seis medidas más comunes.

$$d_{xy} = \frac{S}{C + D + T_x} . \quad (3.2)$$

Por lo tanto, tanto para  $d_{yx}$  como para  $d_{xy}$ , cuando se produce una diferencia entre valores empatados en la independiente (es decir, un par no empatado) no se refleja como una diferencia entre los correspondientes valores empatados de la variable dependiente, si no que los denominadores de las ecuaciones (3.1) y (3.2) se incrementan en  $T_y$  o  $T_x$ , respectivamente, y los valores de  $d_{yx}$  y  $d_{xy}$  se reducen en consecuencia. Por último, es evidente que la medida  $\tau_b$  de Kendall de asociación ordinal es simplemente la media geométrica de las medidas  $d_{yx}$  y  $d_{xy}$  de Somers, dadas por

$$\tau_b = \sqrt{d_{yx}d_{xy}} .$$

## 1.2. Métodos estadísticos de permutación

Para el análisis de permutación exacto de una tabla de contingencia  $r \times c$ , es necesario calcular la medida de asociación ordinal seleccionada para las frecuencias de celdas observadas y enumerar exhaustivamente los  $M$  posibles ordenamientos igualmente probables de los  $N$  individuos en las  $rcc$ eldas, dadas las distribuciones de frecuencias marginales observadas. Para cada ordenamiento en el conjunto de referencia de todas las permutaciones, se calculan una medida de asociación ordinal, digamos  $T$ , y el valor exacto de probabilidad hipergeométrica puntual bajo la hipótesis nula,  $p(n_{ij}|n_{i.}, n_{.j}, N)$ , donde

$$p(n_{ij}|n_{i.}, n_{.j}, N) = \frac{(\prod_{i=1}^r n_{i.}!)(\prod_{j=1}^c n_{.j}!)}{N! \prod_{i=1}^r \prod_{j=1}^c n_{ij}!} ,$$

$n_{ij}$  es la frecuencia de celdas observada para  $i = 1, \dots, r$  y  $j = 1, \dots, c$ ,  $n_{i.}$  es la  $i$ -ésima frecuencia marginal,  $n_{.j}$  es la  $j$ -ésima frecuencia marginal, y  $N$  es el total de todos los valores de  $n_{ij}$  para  $i = 1, \dots, r$  y  $j = 1, \dots, c$ . Si  $T_o$  denota el valor del estadístico de prueba observado, los valores exactos de probabilidad ( $P$ ) a una cola de  $T_o$  son las sumas de los valores  $p(n_{ij}|n_{i.}, n_{.j}, N)$  asociados a los valores  $T$  calculados en los posibles ordenamientos igualmente probables de las frecuencias de las celdas que son iguales o mayores que  $T_o$  cuando  $T_o$  es positivo e iguales o menores que  $T_o$  cuando  $T_o$  es negativo. Así, el valor exacto de la probabilidad hipergeométrica de  $T_o$  cuando  $T$  es positivo viene dado por

$$P = \sum_{k=1}^M \Psi(T_k) p(n_{ij}|n_{i.}, n_{.j}, N) ,$$

donde

$$\Psi(T_k) = \begin{cases} 1 & T_k \geq T_o \\ 0 & c.c. \end{cases}$$

y el valor exacto de la probabilidad hipergeométrica de  $T_o$  cuando  $T$  es negativo viene dado por

$$P = \sum_{k=1}^M \Psi(T_k) p(n_{ij} | n_{i.}, n_{.j}, N) ,$$

donde

$$\Psi(T_k) = \begin{cases} 1 & T_k \leq T_o \\ 0 & c.c. \end{cases} .$$

Cuando el número de posibles ordenamientos de las frecuencias de las celdas es muy grande, las pruebas exactas son poco prácticas y se hacen necesarios los métodos de Monte Carlo. Los métodos estadísticos de permutación de Monte Carlo generan una muestra aleatoria de todos los posibles ordenamientos de las frecuencias de las celdas, extraídos con reemplazo, dadas las distribuciones de frecuencias marginales observadas. Los valores de probabilidad (de cola superior e inferior) del estadístico  $T$  son simplemente las proporciones de los valores  $T$  calculados en los ordenamientos de frecuencias de celdas seleccionadas aleatoriamente que son iguales o mayores que  $T_o$  cuando  $T_o$  es positivo e iguales o menores que  $T_o$  cuando  $T_o$  es negativo. Así, el valor de la probabilidad de remuestreo de Monte Carlo de  $T_o$  cuando  $T$  es positivo viene dado por

$$P(T \geq T_o | H_0) = \frac{\text{número de veces que } T \geq T_o}{L} ,$$

donde  $L$  denota el número de ordenamientos aleatorios de los datos observados.

### 1.3. Medida $\tau_a$ de Kendall de Asociación Ordinal

La  $\tau_a$  de Kendall es una medida de asociación ordinal, dada por

$$\tau_a = \frac{2S}{N(N-1)} ,$$

se diseñó originalmente para medir la asociación entre dos conjuntos de valores de rangos no empatados, donde los dos conjuntos de puntuaciones de rango se etiquetan habitualmente como  $X$  e  $Y$ , aunque los rangos también pueden representarse en una tabla de contingencia  $r \times c$  donde  $n_{i.} = n_{.j} = 1$  para  $i = 1, \dots, r$  y  $j = 1, \dots, c$ . La  $\tau_a$  de Kendall se presenta a veces como una alternativa al coeficiente de correlación de rango de Spearman. Obsérvese también que, como se supone que no hay datos empatados,  $\tau_a$  también puede estar dado por

$$\tau_a = \frac{C - D}{C + D} .$$

## 1.4. Medida $\tau_b$ de Kendall de Asociación Ordinal

Cuando existen valores empatados, la medida de asociación ordinal  $\tau_a$  de Kendall es muy poco satisfactoria, ya que ignora los dos conjuntos de valores empatados,  $T_x$  y  $T_y$ . Por esta razón, Kendall desarrolló  $\tau_b$  una alternativa a  $\tau_a$ , dada por

$$\tau_b = \frac{S}{\sqrt{(C + D + T_x)(C + D + T_y)}} ,$$

que incorpora los valores empatados a las variables  $X$  e  $Y$  ( $T_x$  y  $T_y$ , respectivamente).

## 1.5. Medida $\tau_c$ de Stuart de Asociación Ordinal

La  $\tau_b$  de Kendall es una medida estrictamente monótona de asociación ordinal, es decir, para cada aumento de categoría en la variable  $X$ , se espera que haya un aumento de categoría en la variable  $Y$ .

Por consiguiente,  $\tau_b$  sólo puede alcanzar límites de  $\pm 1$  para tablas de contingencia en las que  $r = c$  y las distribuciones de frecuencias marginales de filas y columnas sean idénticas. Más concretamente,  $\tau_b$  no puede alcanzar generalmente valores de  $\pm 1$  debido a la desigualdad de Cauchy:

*El cuadrado de la suma de los productos de dos conjuntos será igual o menor que el producto de las sumas al cuadrado de dos conjuntos. Formalmente, para las variables  $X$  e  $Y$ ,*

$$\left(\sum_{i=1}^N x_i y_i\right)^2 \leq \sum_{i=1}^N x_i^2 \sum_{i=1}^N y_i^2 .$$

En consecuencia, el numerador de  $\tau_b$  será igual o menor que el denominador, permitiendo que  $\tau_b$  alcance  $\pm 1$  sólo cuando todas las observaciones se concentren en una de las dos diagonales principales de la tabla de contingencia. Si ninguna frecuencia marginal es cero, esto significa que  $\tau_b$  sólo puede alcanzar  $\pm 1$  para una tabla de contingencia cuadrada con distribuciones de frecuencias marginales idénticas. Es importante señalar que, dado que las categorías están ordenadas, las distribuciones de frecuencias marginales deben ser idénticas, no simplemente equivalentes. Así, distribuciones de frecuencias marginales para filas y columnas de, por ejemplo, 50, 30, 20 y 50, 30, 20, respectivamente, son idénticas, lo que proporciona la posibilidad que  $\tau_b$  sea igual a  $+1$ , y distribuciones de frecuencias marginales para filas y columnas de 50, 30, 20 y 20, 30, 50, respectivamente, son idénticas, por lo que existe la posibilidad de que  $\tau_b$  sea igual a  $-1$ , pero las distribuciones de frecuencia marginal de filas y columnas de fila y columna de 50, 30, 20 y 30, 20, 50, respectivamente, son equivalentes pero no idénticas, y por lo tanto obligan a que la  $\tau_b$  de Kendall sea menor que  $+1$  o mayor que  $-1$ .

Por lo tanto, la  $\tau_b$  de Kendall no es la medida más apropiada de asociación ordinal para la tabla de contingencia  $r \times c$ , con  $r \neq c$ . Para corregir esta limitación, Alan Stuart propuso la  $\tau_c$  para las tablas de contingencia donde  $r \neq c$ , dada por

$$\tau_c = \frac{2mS}{N^2(m-1)} ,$$



donde  $m = \min(r, c)$ . Stuart demostró que si  $N$  es un múltiplo de  $m$  y  $r = c$  con distribuciones marginales de frecuencia idénticas, de modo que todas las observaciones caen en la diagonal de la tabla de contingencia y todas las frecuencias de las celdas son iguales, el valor máximo de la  $S$  de Kendall viene dado por

$$S_{max} = \frac{N^2(m-1)}{2m}.$$

Entonces, si  $N = m$ ,

$$\frac{N^2(m-1)}{2m} = \frac{N^2(N-1)}{2N} = \frac{N(N-1)}{2}.$$

Sin embargo, si  $N$  no es un múltiplo de  $m$ , la expresión de  $S_{max}$  sigue siendo un límite superior que no se puede alcanzar. De ello se concluye que la  $\tau_c$  de Stuart puede alcanzar a veces (y para  $N$  grande, puede alcanzar casi siempre)  $\pm 1$ .

## 1.6. Medida $\gamma$ de Goodman y Kruskal

En 1954 Goodman y Kruskal desarrollaron una nueva medida de asociación para dos variables de nivel ordinal que denominaron gamma ( $\gamma$ ). Gamma es una medida de asociación ordinal de reducción proporcional al error que se basa únicamente en los pares no empatados,  $C$  y  $D$ , y viene dada por

$$\gamma = \frac{S}{C+D} = \frac{C-D}{C+D} = \frac{C}{C+D} - \frac{D}{C+D}.$$

Por tanto, de la expresión se deduce que  $\gamma$  es simplemente la diferencia entre las proporciones de los pares iguales y no iguales, ignorando todos los pares empatados, es decir,  $T_x$ ,  $T_y$  y  $T_{xy}$ .

Hay un problema potencial con  $\gamma$  que fue reconocido inmediatamente por Goodman y Kruskal: Gamma es inestable en varios “puntos de corte”, es decir,  $\gamma$  tiende a aumentar a medida que las categorías de una tabla de contingencia se colapsan porque no tiene en cuenta los pares empatados y el número de pares empatados aumenta a medida que la tabla se colapsa. Además,  $\gamma$  es una medida de asociación ordinal monótona (no estricta), es decir, para cada aumento (disminución) de la categoría ordenada en la variable  $x$ , la variable  $y$  aumenta (disminuye) o permanece igual.

## 1.7. Medidas $d_{yx}$ y $d_{xy}$ de Somers

En 1962 el sociólogo Robert Somers se opuso a la medida simétrica de asociación ordinal de Goodman y Kruskal,  $\gamma$ , y propuso dos alternativas asimétricas dadas por

$$d_{yx} = \frac{C-D}{C+D+T_y} = \frac{S}{C+D+T_y},$$

donde  $T_y$  denota el número de pares empatados en la variable  $Y$  pero no empatados en la variable  $X$ , y

$$d_{xy} = \frac{C - D}{C + D + T_x} = \frac{S}{C + D + T_x} ,$$

donde  $T_x$  denota el número de pares empatados en la variable  $X$  pero no empatados en la variable  $Y$ .

Observando las ecuaciones anteriores, se ve que Somers incluyó en los denominadores de  $d_{yx}$  y  $d_{xy}$  el número de valores ligados en la variable dependiente:  $T_y$  para  $d_{yx}$  y  $T_x$  para  $d_{xy}$ . La razón para incluir los valores empatados es simplemente que cuando la variable  $Y$  es la variable dependiente ( $d_{yx}$ ), entonces si dos valores de la variable independiente,  $X$ , difieren pero los dos valores correspondientes de la variable dependiente,  $Y$ , no difieren (están empatados), hay evidencia de una falta de asociación y los empates en la variable  $Y$  ( $T_y$ ) deben ser incluidos en el denominador donde actúan para disminuir el valor de  $d_{yx}$ . El mismo razonamiento es válido para la medida  $d_{xy}$  de Somers.

## 1.8. Medidas $d_{y.x}$ y $d_{x.y}$ de Kim

En 1971, Jae-On Kim propuso medidas asimétricas proporcionales de reducción del error de asociación ordinal dadas por

$$d_{y.x} = \frac{C - D}{C + D + T_x} \quad y \quad d_{x.y} = \frac{C - D}{C + D + T_y}$$

A diferencia de las medidas  $d_{yx}$  y  $d_{xy}$  de Somers de asociación ordinal, que ajustan los empates en la variable dependiente; las medidas  $d_{y.x}$  y  $d_{x.y}$  de Kim ajustan los empates en la variable independiente. Es evidente que las medidas  $d_{y.x}$  y  $d_{x.y}$  de Kim son equivalentes a las medidas  $d_{xy}$  y  $d_{yx}$  de Somers, respectivamente.

## 1.9. Medida $e$ de Asociación Ordinal de Wilson

En 1974, Thomas Wilson propuso otra medida de asociación ordinal que denominó  $e$ . Argumentando que una medida de asociación debería ajustarse para los valores empatados tanto en la variable  $X$  como en la variable  $Y$ , Wilson sugirió una medida simétrica de asociación ordinal dada por

$$e = \frac{C - D}{C + D + T_x + T_y} = \frac{S}{C + D + T_x + T_y} .$$

Como observó Wilson,  $e$  toma los valores de  $\pm 1$  si y sólo si los datos son estrictamente monótonos. Es obvio, a partir de la ecuación anterior, que la  $e$  de Wilson es equivalente a la  $d_{yx}$  de Somers cuando  $T_x = 0$  y es equivalente a la  $d_{xy}$  de Somers cuando  $T_y = 0$ . Además, si  $T_x = 0$  como  $T_y = 0$ , entonces  $e = d_{yx} = d_{xy} = \gamma = \tau_a = \tau_b$ .

## 1.10. Distribuciones de las Frecuencias Marginales

Sea  $C$  el número de pares concordantes,  $D$  el número de pares discordantes,  $T_x$  el número de pares empatados en la variable  $X$  pero no en la variable  $Y$ ,  $T_y$  el número de

pares empatados en la variable  $Y$  pero no en la variable  $X$  y  $T_{xy}$  denota el número de pares empatados tanto en variable  $X$  como en la variable  $Y$ . Entonces, el número total de pares puede dividirse como

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

Obsérvese que

$$\frac{1}{2}(N^2 - \sum_{j=1}^c n_{.j}^2) = C + D + T_x$$

y

$$\frac{1}{2}[\sum_{j=1}^c n_{.j}(n_{.j} - 1)] = T_y + T_{xy} ,$$

donde  $n_{.j}$  indica la frecuencia marginal total de la  $j$ -ésima columna,  $j = 1, \dots, c$ .

Entonces, todos los pares posibles se pueden dividir en términos de los totales de frecuencia marginal como

$$\begin{aligned} \binom{N}{2} &= \frac{1}{2}(N^2 - \sum_{j=1}^c n_{.j}^2) + \frac{1}{2}[\sum_{j=1}^c n_{.j}(n_{.j} - 1)] = \\ &= \frac{1}{2}[N^2 - \sum_{j=1}^c n_{.j}^2 + \sum_{j=1}^c n_{.j}(n_{.j} - 1)] = \frac{1}{2}(N^2 - \sum_{j=1}^c n_{.j}) = \frac{N(N-1)}{2} . \end{aligned}$$

Mientras que la relación dada en la anterior ecuación es en términos de los totales de frecuencia marginal de columna, los mismos resultados pueden obtenerse de los totales de frecuencia marginal de fila de fila, es decir

$$\binom{N}{2} = \frac{1}{2}[N^2 - \sum_{i=1}^r n_{i.}^2 + \sum_{i=1}^r n_{i.}(n_{i.} - 1)] ,$$

donde  $n_{i.}$  indica el total de frecuencia marginal de la fila  $i$ ,  $i = 1, \dots, r$ .

Por lo tanto, como las distribuciones de frecuencias marginales se fijan bajo permutación, los valores de probabilidad exactos de  $\tau_a$  de Kendall,  $\tau_b$  de Kendall,  $d_{yx}$  de Somers y  $d_{xy}$  de Somers se basan totalmente en la distribución de permutación del numerador común,  $S$ . En el caso de la medida de asociación ordinal de Stuart, la fórmula para  $\tau_c$  no incluye ni  $C + D + T_x$  ni  $C + D + T_y$ , sino que utiliza  $m = \min(r, c)$ , que se basa en el número de filas o columnas que se fijan bajo permutación. En consecuencia, el valor de la probabilidad de  $\tau_c$  de Stuart también se basa únicamente en la distribución de permutación del estadístico  $S$ . En el caso de la medida de asociación ordinal de Goodman y Kruskal,  $\gamma$ , no considera que  $T_x$  ni  $T_y$  proporcionen ninguna información utilizable; por lo tanto, su valor de probabilidad difiere ligeramente del valor de valor de probabilidad común para  $\tau_a$  y  $\tau_b$  de Kendall,  $\tau_c$  de Stuart y  $d_{yx}$  y  $d_{xy}$  de Somers.