



DOBLE GRADO EN  
MATEMÁTICAS Y ESTADÍSTICA

---

TRABAJO FIN DE GRADO

---

*Medidas de Asociación  
para Variables  
Nominales*

---

Alumna: Gloria Vizcaíno Castaño  
Tutor: Juan Manuel Muñoz Pichardo

Sevilla, Junio de 2022



# Índice general

Prólogo . . . . .	III
Resumen . . . . .	V
Abstract . . . . .	VI
Índice de Figuras . . . . .	VII
Índice de Tablas . . . . .	IX
<b>1. Introducción</b>	<b>1</b>
1.1. Definiciones . . . . .	2
1.1.1. Definición de Medición . . . . .	2
1.1.2. Definición de Asociación . . . . .	2
1.2. Dimensiones de la asociación . . . . .	3
1.2.1. Nivel de Medición . . . . .	3
1.2.2. Simetría y Asimetría . . . . .	3
1.2.3. Asociación unidireccional y bidireccional . . . . .	3
1.2.4. Modelos de interpretación . . . . .	3
1.2.5. Clasificación cruzada . . . . .	4
1.2.6. Correlación, asociación y concordancia . . . . .	4
1.3. Criterios para las medidas de asociación . . . . .	5
1.4. Grado de Asociación . . . . .	6
1.5. Elección de la Medida de Asociación . . . . .	7
<b>2. Medidas para Variables Nominales</b>	<b>11</b>
2.1. Valores de probabilidad hipergeométrica . . . . .	11
2.2. Medidas $\lambda_a$ y $\lambda_b$ de Goodman y Kruskal . . . . .	14
2.3. Medidas $t_a$ y $t_b$ de Goodman y Kruskal . . . . .	16
2.4. Una prueba asimétrica de homogeneidad . . . . .	18
2.5. Medidas de concordancia . . . . .	20
2.5.1. Medida de concordancia de Robinson . . . . .	22
2.5.2. Medida de concordancia $\pi$ de Scott . . . . .	23
2.5.3. Medida de concordancia $\kappa$ de Cohen . . . . .	24
2.5.4. Aplicación con Varios Jueces . . . . .	26
2.6. Test Q de McNemar para el cambio . . . . .	28
2.7. Test Q de Cochran para el cambio . . . . .	29
2.8. Una medida sobre el Tamaño del Efecto para el test Q de Cochran . . . . .	30
2.8.1. Una medida del Tamaño del Efecto corregida por el azar . . . . .	32
2.8.2. Ventajas de la medida $\mathfrak{R}$ del tamaño del efecto . . . . .	33
2.9. Medida de asociación $d_N^c$ de Leik y Gove . . . . .	33
2.9.1. Tabla de Contingencia Observada . . . . .	35

2.9.2. Tabla de Contingencia Esperada . . . . .	35
2.9.3. Tabla de Contingencia Maximal . . . . .	36
2.9.4. Cálculo de la $d_N^c$ de Leik y Gove . . . . .	37
2.9.5. Un test de permutación para $d_N^c$ . . . . .	38
2.10. Un problema de ocupación de la matriz . . . . .	38
2.11. Prueba de probabilidad exacta de Fisher . . . . .	38
2.12. Análisis de tablas de $2 \times 2 \times 2$ . . . . .	38
<b>3. Título del Capítulo</b>	<b>39</b>
3.1. Primera sección . . . . .	39
<b>4. Título del Capítulo</b>	<b>41</b>
4.1. Primera sección . . . . .	41
<b>A. Apéndice: Título del Apéndice</b>	<b>43</b>
A.1. Primera sección . . . . .	43
<b>B. Apéndice: Título del Apéndice</b>	<b>45</b>
B.1. Primera sección . . . . .	45
<b>Bibliografía</b>	<b>47</b>

# Prólogo

Escrito colocado al comienzo de una obra en el que se hacen comentarios sobre la obra o su autor, o se introduce en su lectura; a menudo está realizado por una persona distinta del autor.

También se podrían incluir aquí los agradecimientos.



# Resumen

La evaluación de la relación y la intensidad de la misma entre dos o más variables aleatorias es un objetivo presente en cualquier estudio estadístico. Para ello se han propuesto una amplia gama de medidas estadísticas adaptadas a la naturaleza de las variables analizadas.

El objetivo de este trabajo se centra en recopilar las medidas de asociación entre variables nominales, incluyendo la definición teórica e inferencia estadística sobre las mismas.

Asimismo, el trabajo incluye su implementación en R y/o el uso de librerías de R, con una ilustración sobre datos reales. Además, con objeto de ilustrar la aplicabilidad de las mismas, incluye referencias sobre trabajos científicos recientes en los que se han utilizado estas medidas.

# Abstract

The evaluation of the relationship and its intensity between two or more random variables is an objective present in any statistical study. For this purpose, a wide range of statistical measures adapted to the nature of the variables analyzed have been proposed.

The aim of this project is to compile measures of association between nominal variables, including their theoretical definition and statistical inference.

Also, the report includes their implementation in R and/or the use of R libraries, with an illustration on real data. In addition, in order to illustrate their applicability, it includes references to recent scientific studies in which these measures have been used.



# Índice de figuras



# Índice de tablas

2.1.	Tabla 1: Notación tablas de contingencia 2x2 . . . . .	12
2.2.	Tabla 2: Notación tablas de contingencia 4x3 . . . . .	12
2.3.	Tabla 3: Notación para la clasificación cruzada de dos variables categóricas, A_j con j=1,...,c y B_i con i=1,...,r . . . . .	14
2.4.	Notación para la clasificación cruzada de dos variables categóricas, A_j para j=1,...,g y B_i para i=1,...,r . . . . .	19
2.5.	Clasificación cruzada cxc con proporciones en celdas . . . . .	25
2.6.	Notación de una tabla de clasificación cruzada 2x2 para el test Q de McNemar	28



# Capítulo 1

## Introducción

ESTO TENGO QUE MEJORARLO CUANDO TENGA MAS IDEA DEL TEMA:

(En este Trabajo Fin de Grado, se usará un enfoque de permutación para la medición de la asociación estadística, definida ampliamente para incluir medidas de correlación, asociación y ¿concordancia?(agreement).

Aunque existe una gran cantidad de métodos para medir la magnitud de la asociación entre dos variables, existe una dificultad considerable para interpretar y comparar las distintas medidas, ya que a menudo difieren en su estructura, lógica e interpretación.

Así pues, las distintas medidas de asociación desarrolladas a lo largo de los años constituyen una mezcla de enfoques lógicos, estructurales y de interpretación.

Es conveniente clasificar las distintas medidas de asociación por el nivel de medición para el que fueron diseñadas originalmente y para el que son más apropiadas, reconociendo que algunas medidas pueden ser adecuadas para más de un nivel de medición, especialmente las numerosas medidas originalmente diseñadas para el análisis de tablas de contingencia  $2 \times 2$ , en las que el nivel de medición es a menudo irrelevante.

Además de la consideración de la estructura, la lógica y la interpretación, un inconveniente importante de las medidas de asociación es la determinación de la probabilidad de la medida obtenida bajo la hipótesis nula.

Existen dos enfoques principales para determinar de probabilidad para las medidas de asociación: el modelo poblacional de Neyman-Pearson y el modelo de permutación de Fisher-Pitman.

El modelo poblacional está plagado de suposiciones que rara vez se cumplen en la práctica y que a menudo son inapropiados para los niveles inferiores de medición, por ejemplo, independencia, normalidad, homogeneidad de la varianza. . .

De aquí en adelante se usará casi exclusivamente el modelo de permutación ya que está libre de cualquier suposición de distribución, no requiere un muestreo aleatorio, es completamente dependiente de los datos, proporciona valores de probabilidad exactos y es ideal para el análisis de muestras pequeñas.)

## 1.1. Definiciones

Dado que el título de este Trabajo de Fin de Grado es “Medidas de Asociación para Variables Nominales”, es conveniente, en primer lugar, definir “Medición” y “Asociación”.

### 1.1.1. Definición de Medición

La mejor manera de describir con precisión los acontecimientos y las relaciones entre ellos es mediante la medición. La medición ha sido una característica fundamental de la civilización humana desde sus inicios. Así, la medición es la aplicación de las matemáticas a los acontecimientos, el uso de números para designar objetos y acontecimientos y sus relaciones.

Más formalmente, la medición es el proceso de la asignación de los fenómenos empíricos a un sistema numérico.

Se pueden distinguir cuatro niveles o escalas de medición: nominal, ordinal, de intervalo y de razón:

- El nivel **nominal** de medición no mide cantidades, simplemente clasifica los acontecimientos en una serie de categorías no ordenadas y se agrupan los acontecimientos que tienen características comunes. Ejemplos de clasificaciones nominales son el género, el tipo de sangre o el estado civil.
- La esencia del nivel **ordinal** de medición es que emplea las características de “mayor que” ( $>$ ) o “menor que” ( $<$ ). Las relaciones ( $>$ ) y ( $<$ ) no son reflexivas ni simétricas, pero sí transitivas. Ejemplos de escalas ordinales son el orden de nacimiento, el rango académico o las escalas Likert (Muy de acuerdo, De acuerdo, Neutral, En desacuerdo, Totalmente en desacuerdo).
- Las escalas de nivel de **intervalo** introducen otra dimensión en el proceso de medición y ordenan los eventos en intervalos de igual apariencia. En las escalas de intervalo, no hay un punto cero absoluto: si hay un valor de cero, éste se define arbitrariamente. Las temperaturas medidas en grados Fahrenheit o Centígrados son ejemplos tradicionales de medición de intervalos.
- Las escalas de **razón** son escalas que no sólo incorporan todas las características de una escala de intervalo, sino que tienen puntos cero absolutos, lo que permite la construcción de relaciones significativas. Ejemplos de escalas de intervalo son el tiempo, la edad, la altura o los grados Kelvins (0 Kelvins es el cero absoluto, definido como la ausencia de movimiento molecular).

Desde el punto de vista estadístico, las mediciones a nivel de intervalo y de razón suelen tratarse juntas y, en general, se denominan simplemente mediciones de nivel de intervalo.

### 1.1.2. Definición de Asociación

Aunque hay muchas formas de definir la asociación, quizás la más sencilla y útil sea: se dice que dos variables están asociadas cuando la distribución de los valores de una variable difiere para diferentes valores de la otra variable.

Además, si un cambio en la distribución de los valores de una variable no provoca un cambio en la distribución de los valores de la otra variable, se dice que las variables son

independientes. Hay que tener en cuenta que casi todas las discusiones sobre asociación implican una comparación de subgrupos.

Así, la independencia se da cuando los subgrupos de variables no difieren, y cuando los subgrupos difieren, se da la asociación.

## 1.2. Dimensiones de la asociación

Hay que tener en cuenta varias dimensiones a la hora de medir la asociación:

### 1.2.1. Nivel de Medición

Como hemos visto anteriormente, pueden ser: variables de nivel nominal (categóricas), de nivel ordinal (clasificadas) y de nivel de intervalo. Además, en algunos casos, se consideran mezclas de los tres niveles de medición: variables de nivel nominal y ordinal, de nivel nominal y de intervalo, y de nivel ordinal e intervalo.

Los ejemplos de medidas de asociación de nivel nominal incluyen las medidas simétricas basadas en chi-cuadrado como la  $\phi^2$  de Pearson, la  $T^2$  de Tschuprov, la  $V^2$  de Cramér y la  $C$  de Pearson, así como las medidas asimétricas  $t_a$  y  $t_b$  de Goodman y Kruskal y el coeficiente de concordancia kappa ( $\kappa$ ) no ponderado de Cohen.

### 1.2.2. Simetría y Asimetría

Una medida de asociación puede ser asimétrica, con variables independientes y dependientes bien definidas, dando lugar a dos índices que miden la fuerza de la asociación dependiendo de la variable que se considere dependiente; o simétrica, dando lugar a un único índice de fuerza de asociación.

Ejemplos de medidas asimétricas de asociación incluyen las diferencias porcentuales simples y las medidas  $t_a$  y  $t_b$  de asociación nominal de Goodman y Kruskal. Algunos ejemplos de medidas de asociación simétricas son las medidas basadas en chi-cuadrado para variables de nivel nominal, como la  $\phi^2$  de Pearson, la  $T^2$  de Tschuprov, la  $V^2$  de Cramér y la  $C$  de Pearson.

### 1.2.3. Asociación unidireccional y bidireccional

Las medidas de asociación pueden cuantificar la asociación unidireccional entre variables basándose en la medida en que una variable implica a la otra, pero no a la inversa. Por otro lado, la asociación bidireccional o mutua se refiere a la medida en que las dos variables se implican mutuamente. Todas las medidas asimétricas son medidas de asociación unidireccional, y algunas medidas simétricas son medidas de asociación unidireccional.

Un ejemplo de asociación unidireccional es la diferencia porcentual simple. Los ejemplos de asociación mutua son las medidas estándar basadas en el chi-cuadrado:  $\phi^2$  de Pearson,  $T^2$  de Tschuprov,  $V^2$  de Cramér y  $C$  de Pearson.

### 1.2.4. Modelos de interpretación

Las medidas de asociación se basan en diferentes modelos, como el de máxima corrección, el de corrección por azar y el de reducción proporcional del error. Aunque estos modelos

no son exhaustivos ni se excluyen mutuamente, la tipología proporciona un importante esquema de clasificación.

Algunos ejemplos de medidas de asociación con corrección máxima son la  $\phi^2$  de Pearson, la  $T^2$  de Tschuprov, la  $V^2$  de Cramér y la  $C$  de Pearson. Ejemplos de medidas de asociación corregidas por el azar incluyen la medida de acuerdo  $\pi$  de Scott, la medida  $A$  de Robinson, la medida de la regla del pie de Spearman, el coeficiente  $u$  de Kendall y los coeficientes kappa ponderados y no ponderados de Cohen,  $\kappa$  y  $\kappa_w$ . Ejemplos de medidas de asociación con reducción proporcional del error de asociación incluyen las medidas  $\lambda_a$  y  $\lambda_b$  de Goodman y Kruskal de asociación nominal y la medida  $\gamma$  de asociación ordinal de Goodman y Kruskal.

### 1.2.5. Clasificación cruzada

Las medidas de asociación se han construido históricamente para datos clasificados en tablas de contingencia de doble entrada o, alternativamente, en simples listas bivariadas de medidas de respuesta. Además, algunas medidas suelen calcularse en ambos sentidos.

Ejemplos de medidas de asociación para datos organizados en tablas de contingencia incluyen los coeficientes kappa ponderados y no ponderados de Cohen,  $\kappa$  y  $\kappa_w$ , y las medidas habituales basadas en chi-cuadrado, incluyendo la  $\phi^2$  de Pearson, la  $T^2$  de Tschuprov, la  $V^2$  de Cramér y la  $C$  de Pearson. Los ejemplos de medidas de asociación para datos no organizados en una tabla de contingencia incluyen la prueba  $Q$  de Cochran para el cambio, el coeficiente de correlación de rango de Spearman y el coeficiente de correlación de producto-momento de Pearson.

### 1.2.6. Correlación, asociación y concordancia

Las medidas de **asociación** pueden medir de diversas maneras la correlación, la asociación o la concordancia. Muchos autores han tratado de distinguir entre los conceptos de correlación y asociación. Hay dos ámbitos correspondientes al término “asociación”:

- El más general incluye todos los tipos de medidas de asociación entre dos variables en todos los niveles de medición.
- El dominio más restrictivo está reservado a las medidas diseñadas específicamente para medir el grado de relación entre dos variables en los niveles de medición nominal y ordinal.

Así pues, en este trabajo, la asociación se usará de dos maneras. En primer lugar, como un concepto global que incluye medidas de correlación, asociación y concordancia; y en segundo lugar, se utilizará más específicamente como una medida de relación entre dos variables de nivel nominal, dos variables de nivel ordinal o alguna combinación de ambas.

En general, la **correlación** suele referirse a las medidas de covariación derivadas de las ecuaciones de regresión basadas en el método de mínimos cuadrados ordinarios. A menudo, pero no siempre, la correlación simple mide la relación entre dos variables a nivel de intervalo de medida, donde las dos variables se etiquetan normalmente como  $X$  e  $Y$ ; una excepción es, por ejemplo, coeficiente el  $\phi^2$  de Pearson para dos variables binarias. La medida de correlación más usada es el coeficiente de correlación de Pearson al cuadrado.

Las medidas de **concordancia** intentan determinar la identidad de dos variables en cualquier nivel de medición, es decir,  $X_i = Y_i$  para todo  $i$ . Algunos ejemplos de medidas



de concordancia son la medida  $\pi$  de Scott, la medida  $A$  de Robinson, la medida de la regla de Spearman y los coeficientes kappa ponderados y no ponderados de Cohen.

La correlación y la concordancia se suelen confundir, a continuación se muestra un ejemplo para entender las diferencias:

Supongamos que un investigador desea establecer la relación entre los valores observados y los predichos por la regresión,  $y$  e  $\hat{y}$ , respectivamente. La concordancia implica que la relación funcional entre  $y$  e  $\hat{y}$  puede describirse mediante una la recta  $x=y$ . Si por ejemplo obtenemos los pares (1,1), (3,3), (8,8), el coeficiente de correlación de Pearson al cuadrado es  $r_{y,\hat{y}}^2 = 1$  y el porcentaje de concordancia es del 100 %, es decir, los elementos de los tres pares  $(y, \hat{y})$  son iguales. En este contexto, el coeficiente de correlación de Pearson al cuadrado,  $r_{y,\hat{y}}^2$ , también se ha utilizado como medida de concordancia. Sin embargo,  $r_{y,\hat{y}}^2 = 1,00$  implica una relación lineal entre  $y$  e  $\hat{y}$ , donde tanto la el corte con el eje de ordenadas como la pendiente son arbitrarias. Así, aunque la concordancia perfecta se describe con un valor de 1,00, también es cierto que  $r_{y,\hat{y}}^2 = 1,00$  describe una relación lineal que puede o no reflejar una concordancia perfecta, por ejemplo para los valores  $(y, \hat{y})$ : (2, 4), (4, 5), (6, 6), (8, 7), y (10, 8), el coeficiente de correlación de Pearson es  $r_{y,\hat{y}}^2 = 1,00$ , y el porcentaje de concordancia es del 20 %, es decir, sólo un par valores coinciden.

### 1.3. Criterios para las medidas de asociación

Varios investigadores han escrito sobre criterios importantes para las medidas de asociación, sobre todo Costner y Goodman y Kruskal. Sin embargo, esta sección se basa principalmente en los criterios que Weiss consideraba más importantes.

Los criterios importantes para las medidas de asociación incluyen la normalización adecuada, la interpretación, la independencia de las frecuencias marginales y la magnitud (grado o fuerza) de la asociación:

- **Normalización:** Idealmente, los valores de una medida de asociación deberían cubrir el mismo rango que los valores de probabilidad, es decir, de 0 a 1. Además, la medida de asociación debe ser cero cuando las variables son independientes y uno cuando hay una asociación perfecta. Cuando sea conveniente considerar la asociación inversa, entonces menos uno debe representar la asociación negativa perfecta. Las medidas de asociación basadas en la chi-cuadrado (la  $\phi^2$  de Pearson, la  $T^2$  de Tschuprov, la  $V^2$  de Cramér y la  $C$  de Pearson) no suelen tener uno como límite superior, y la razón de probabilidades (odds ratio) tiene un límite superior infinito.

Además, algunas medidas de asociación de reducción proporcional en el error, como  $\lambda_a$  y  $\lambda_b$  de Goodman y Kruskal, pueden ser cero incluso cuando las dos variables consideradas no son independientes entre sí.

- **Interpretación:** Una medida de asociación debe tener una interpretación significativa, como la reducción proporcional del error probable, la proporción de la varianza explicada o la proporción por encima de lo que cabría esperar por azar. Muchas medidas de asociación carecen notablemente de este aspecto. De hecho, muchas medidas no permiten ninguna interpretación, excepto que un valor más alto indica más asociación que un valor más bajo, e incluso eso es a menudo cuestionable. Las medidas tradicionales basadas en chi-cuadrado ( $\phi^2$  de Pearson,  $T^2$  de Tschuprov,  $V^2$

de Cramér y  $C$  de Pearson) carecen notablemente de una interpretación significativa, excepto para los valores 0 y 1.

- Independencia de las frecuencias marginales: Idealmente, una medida de asociación no debería cambiar con un aumento (disminución) de los totales de frecuencia de filas o columnas; es decir, la medida de asociación debería ser independiente de los totales de frecuencia marginal. Algunas medidas de asociación tienen esta propiedad, como las diferencias porcentuales y los odds ratio, pero muchas otras no.
- Grado de asociación: Los valores de una medida de asociación deben aumentar (disminuir) con el aumento (disminución) de los grados de asociación. Así, cuando las frecuencias de las celdas de una tabla de contingencia indican cambios en la asociación, la medida de asociación debería cambiar de forma acorde.

## 1.4. Grado de Asociación

Las diferentes medidas de asociación evalúan el grado de asociación de diversas maneras. Entre las diversas formas de medir la fuerza de la asociación se encuentran la desviación de la independencia, la magnitud de las diferencias de los subgrupos, las comparaciones por pares, la correspondencia incremental y la concordancia entre variables:

- Desviación de la independencia: Las medidas de asociación que se basan en la desviación de la independencia plantean cómo serían los datos si las dos variables fueran independientes, es decir, que no hubiera asociación, y luego miden el grado en que los datos observados se apartan de la independencia. Ejemplos de medidas de asociación basadas en la desviación de la independencia incluyen las medidas basadas en chi-cuadrado como la  $\phi^2$  de Pearson, la  $T^2$  de Tschuprov, la  $V^2$  de Cramér y la  $C$  de Pearson, así como otras, no basadas en chi-cuadrado, como las medidas de Goodman y Kruskal  $\lambda_a$ ,  $\lambda_b$ , y las medidas de asociación nominal  $t_a$  y  $t_b$ .
- Magnitud de las diferencias entre subgrupos: Dada alguna asociación existente, el grado de asociación puede medirse comparando las proporciones de los subgrupos. Ejemplos de medidas de asociación basadas en las diferencias de subgrupos son las diferencias porcentuales simples, así como otras medidas diseñadas para tablas de contingencia  $2 \times 2$ , como las medidas de asociación  $Q$  y  $Y$  de Yule y los odds ratio.
- Comparaciones por pares: Algunas medidas de asociación se basan en comparaciones por pares donde las diferencias entre las medidas de respuesta se calculan entre todos los pares de mediciones posibles y se dividen en pares concordantes y discordantes. Un par concordante es aquel en el que la dirección de la diferencia con una variable coincide con la dirección de la diferencia con la segunda variable. Un par discordante es aquel en el que la dirección de la diferencia con una variable no es igual a la dirección de la diferencia con la segunda variable. Algunos ejemplos de medidas de asociación ordinal por pares son las medidas  $\tau_a$  y  $\tau_b$  de Kendall, la medida  $\tau_c$  de Stuart, la medida  $\gamma$  de asociación ordinal de Goodman y Kruskal y las medidas asimétricas  $d_{yx}$  y  $d_{xy}$  de asociación ordinal de Somers.
- Correspondencia incremental: El grado de asociación se basa en la medida en que un aumento (disminución) incremental en una variable va acompañado de un aumento (disminución) en la otra variable. Este enfoque se denomina convencionalmente

“correlación” en lugar de “asociación”. Un ejemplo es el coeficiente de correlación producto-momento de Pearson.

- **Concordancia entre variables:** El grado de asociación se mide por el grado en que los valores de una variable discrepan de los valores de la otra variable, por encima de lo esperado por el mero azar. Ejemplos de medidas de asociación basadas en la concordancia incluyen la medida de concordancia  $\pi$  de Scott, la medida  $A$  de Robinson, la medida de la regla del pie de Spearman, el coeficiente  $u$  de Kendall y las medidas kappa ponderada y no ponderada de Cohen,  $\kappa$  y  $\kappa_w$ .

## 1.5. Elección de la Medida de Asociación

A la hora de seleccionar una medida de asociación adecuada, hay que tener en cuenta varios criterios. Liebetrau proporciona algunas directrices para seleccionar la medida más apropiada:

En primer lugar, nos planteamos si las variables son nominales, ordinales, de intervalo o alguna combinación de las tres. Una medida de asociación para variables de nivel nominal (categóricas) no debería depender de categorías ordenadas. Por otro lado, una medida de asociación para variables de nivel ordinal (clasificadas) debería depender de categorías ordenadas. Si se ignora el orden de las categorías, se pierde información. Además, elevar al cuadrado las diferencias entre las categorías ordenadas o los rangos sigue siendo controvertido y debería evitarse. Una medida de asociación (correlación) para variables de nivel de intervalo debería utilizar idealmente toda la información contenida en los datos.

En segundo lugar, ¿se medirá la correlación, la asociación o la concordancia? En general, las medidas asimétricas son adecuadas para la predicción, mientras que las medidas simétricas son apropiadas para la asociación o la correlación, dependiendo del nivel de medición. Las medidas de correlación más utilizadas son el coeficiente de correlación de Pearson, el coeficiente de correlación intraclass de Pearson, el coeficiente de correlación de rango de Spearman, el coeficiente de correlación tetracórico de Pearson y el coeficiente de correlación multiserie de Jaspens. Las medidas de asociación incluyen las medidas basadas en chi-cuadrado como la  $\phi^2$  de Pearson, la  $T^2$  de Tschuprov y la  $V^2$  de Cramér para variables de nivel nominal. Las medidas de concordancia incluyen la  $\pi$  de Scott, la  $A$  de Robinson, la regla del pie de Spearman, la  $\kappa$  de Cohen y la  $u$  de Kendall.

En tercer lugar, ¿se va a utilizar la medida de asociación para hacer inferencia? Bajo el modelo poblacional de Neyman-Pearson, esto requiere conocer el error estándar del estimador y a menudo requiere hacer suposiciones sobre la naturaleza de la población, así como un muestreo aleatorio. Bajo el modelo de permutación de Fisher-Pitman, no se requiere conocer el error estándar, no es necesario el muestreo aleatorio, los supuestos de distribución son irrelevantes y las pruebas de permutación dependen completamente de los datos. Los valores de probabilidad basados en permutaciones pueden ser exactos, basados en el conjunto de referencia de todas las permutaciones posibles de los datos observados, o aproximados, basados en una gran muestra aleatoria extraída del conjunto de referencia con el método Monte Carlo. Gracias a la evolución de la informática, incluso en un pequeño ordenador de sobremesa o portátil, se pueden generar un conjunto de referencia completo de 100.000.000 valores en tan sólo unos minutos, lo que hace que los métodos estadísticos de permutación exacta sean cada vez más populares.

En cuarto lugar, ¿es la medida de asociación sensible a los cambios de frecuencias marginales? Que la medida de asociación no se altere al multiplicar o dividir una o ambas columnas o filas de la tabla de contingencia por cualquier factor arbitrario es una propiedad muy importante de la medida. En general, los valores de una medida de asociación calculados en dos muestras diferentes no pueden compararse si la medida depende de las frecuencias marginales. Prácticamente todas las medidas de asociación para las variables de nivel nominal y ordinal son sensibles a los cambios de frecuencias marginales. Algunas excepciones notables son la odds ratio, las diferencias porcentuales y la medida de asociación  $Q$  de Yule.

En quinto lugar, ¿es la medida de asociación estable ante cambios en el número de categorías? Una medida de asociación estable es aquella en la que el valor no cambia cuando, por ejemplo, el número de categorías se cambia de cinco a cuatro. La medida gamma de asociación ordinal de Goodman y Kruskal es especialmente inestable, mientras que la medida  $\tau_b$  de asociación ordinal de Kendall es relativamente estable. También se puede decir que una medida es estable cuando la medida calculada sobre un número de categorías disjuntas y ordenadas es similar al valor calculado sobre la variable continua subyacente antes de ser dividida en categorías ordenadas.

En sexto lugar, ¿es el valor de la medida de asociación fácilmente interpretable? Algunas medidas de asociación tienen interpretaciones claras y significativas, como las medidas de reducción proporcional del error y las medidas corregidas por el azar. Otras medidas no tienen una interpretación significativa, excepto cuando poseen los valores extremos de 0, 1 o  $-1$ . Por ejemplo, la medida simétrica gamma de Goodman y Kruskal de asociación ordinal y las medidas asimétricas  $d_{yx}$  y  $d_{xy}$  de Somers de asociación ordinal poseen interpretaciones de reducción proporcional en el error, donde los valores positivos indican una mejora proporcional de los errores de predicción con el conocimiento de ambas variables, en comparación de los errores de predicción con el conocimiento de una sola variable. Las medidas kappa no ponderada y kappa ponderada de Cohen para la concordancia y la regla del pie de Spearman poseen interpretaciones corregidas por el azar, en las que los valores positivos indican una concordancia superior a la esperada por el azar y los valores negativos indican una concordancia inferior. En general, las medidas de asociación basadas en el estadístico de prueba chi-cuadrado de Pearson tienen interpretaciones significativas sólo cuando poseen valores de 0 o 1.

En séptimo lugar, ¿la medida de asociación tiene en cuenta los valores empatados? Algunas medidas de asociación requieren ajustes complicados para acomodar los valores empatados, mientras que otras incorporan los valores empatados sin ningún ajuste. Por ejemplo, el coeficiente de correlación de rango de Spearman, desarrollado originalmente en 1904, requería complejos ajustes para los valores empatados. Hoy en día, los investigadores tienen los conocimientos suficientes para calcular simplemente el coeficiente de correlación producto-momento de Pearson sobre los rangos observados, que ajusta automáticamente cualquier valor empatado. Sin embargo, hay muchas otras medidas de asociación que requieren ajustes complicados para los valores empatados, algunos de los cuales son muy cuestionables.

En octavo lugar, ¿se generalizará fácilmente la medida de asociación para datos multivariantes? El análisis multivariante es cada vez más importante en la investigación, por lo que las medidas de asociación que se adaptan a los datos multivariantes son extremadamente útiles. Algunas medidas se generalizan fácilmente a las estructuras multivariantes, otras son más difíciles de generalizar y algunas son imposibles. Durante años, los investigadores

intentaron generalizar la medida kappa de Cohen de concordancia a más de dos variables. En todas las ocasiones, la generalización presentaba problemas. Finalmente, en 2008 se encontró una solución y la medida kappa de Cohen puede ahora manejar cualquier número de variables con cualquier tipo de función de ponderación. Otro ejemplo es la medida de asociación ordinal de Spearman desarrollada en 1906 para dos conjuntos de clasificaciones, que se estableció finalmente una generalización para cualquier número de variables 92 años después, en 1998.

En noveno lugar, ¿para qué tipo de asociación asume la medida de asociación su valor extremo? Algunas medidas de asociación asumen sus valores extremos, por ejemplo,  $+1$  en casos de asociación débil, como es el caso del estadístico gamma de Goodman y Kruskal. Otras medidas de asociación asumen sus valores extremos sólo en el caso de una asociación perfecta. Por ejemplo, la medida  $\tau_b$  de Kendall de asociación ordinal asume un valor de  $+1$  sólo cuando existe una asociación extremadamente fuerte.

En décimo lugar, ¿es la medida de asociación fácil de calcular? Algunas medidas de asociación son notoriamente difíciles de calcular. Una de las más difíciles es la medida de correlación tetracórica de Pearson para tablas de contingencia  $2 \times 2$ . Otra es la medida de asociación nominal  $d_N^c$  de Leik y Gove. Por otro lado, la medida de la regla del pie de Spearman se diseñó específicamente para facilitar el cálculo y la mayoría de las medidas de asociación basadas en chi-cuadrado requieren muy poco esfuerzo para calcularlas.



# Capítulo 2

## 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 (el tema 3 del libro analiza esos criterios, ¿los añado?).

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 incluyen en el capítulo las pruebas  $Q$  de McNemar y Cochran, que miden el grado en que las medidas de respuesta cambian con el tiempo, la medida  $d_N^c$  de Leik y Gove de asociación nominal, y una solución al problema de ocupación de la matriz propuesta por Mielke y Siddiqui.

La prueba de probabilidad exacta de Fisher es la prueba de permutación ideal para las tablas de contingencia. Mientras que la prueba exacta de Fisher suele limitarse a las tablas de contingencia de  $2 \times 2$ , para las que se concibió originalmente, en este capítulo la prueba exacta de Fisher se extiende a las tablas de contingencia de  $2 \times c$ ,  $3 \times 3$ ,  $2 \times 2 \times 2$  y otras más grandes.

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  $d_{xy}$ .

### 2.1. Valores de probabilidad hipergeométrica

Los métodos estadísticos de permutación exacta, especialmente cuando se aplican a las tablas de contingencia, dependen en gran medida de los valores de probabilidad hipergeométricos. En esta sección, se hace una breve introducción a los valores de probabilidad hipergeométricos ilustrando su cálculo e interpretación. Para las tablas de contingencia de  $2 \times 2$ , el cálculo de los valores de probabilidad hipergeométricos es fácil de demostrar. Consideremos la siguiente tabla de contingencia  $2 \times 2$ :

Tabla 2.1: Tabla 1: Notación tablas de contingencia 2x2

	$A_1$	$A_2$	<b>Total</b>
$B_1$	$n_{11}$	$n_{12}$	$R_1$
$B_2$	$n_{21}$	$n_{22}$	$R_2$
<b>Total</b>	$C_1$	$C_2$	$N$

donde  $n_{11}, \dots, n_{22}$  denotan las cuatro frecuencias absolutas,  $R_1$  y  $R_2$  denotan los totales de las frecuencias marginales de cada fila,  $C_1$  y  $C_2$  denotan los totales de las frecuencias marginales de cada columna y

$$N = \sum_{i=1}^2 \sum_{j=1}^2 n_{ij}$$

Dado que la tabla de contingencia que aparece en la *Tabla 1* es una tabla  $2 \times 2$  y, en consecuencia tiene sólo un grado de libertad, la probabilidad de cualquier frecuencia de celda constituye la probabilidad de toda la tabla de contingencia. Por lo tanto, el valor de la probabilidad del punto hipergeométrico para la celda que contiene  $n_{11}$  viene dada por:

$$p(n_{11}|R_1, C_1, N) = \binom{C_1}{n_{11}} \binom{C_2}{n_{12}} \binom{N}{R_1}^{-1} = \binom{R_1}{n_{11}} \binom{R_2}{n_{21}} \binom{N}{C_1}^{-1} = \frac{R_1!R_2!C_1!C_2!}{N!n_{11}!n_{12}!n_{21}!n_{22}!}$$

El cálculo de los valores de probabilidad hipergeométricos para las tablas de contingencia  $r \times c$  es más complejo que para las tablas de contingencia simples de  $2 \times 2$ . Consideremos la tabla de contingencia  $4 \times 3$ :

Tabla 2.2: Tabla 2: Notación tablas de contingencia 4x3

	$A_1$	$A_2$	$A_3$	<b>Total</b>
$B_1$	$n_{11}$	$n_{12}$	$n_{13}$	$R_1$
$B_2$	$n_{21}$	$n_{22}$	$n_{23}$	$R_2$
$B_3$	$n_{31}$	$n_{32}$	$n_{33}$	$R_3$
$B_4$	$n_{41}$	$n_{42}$	$n_{43}$	$R_4$
<b>Total</b>	$C_1$	$C_2$	$C_3$	$N$

donde  $n_{11}, \dots, n_{43}$  denotan las 12 frecuencias de las celdas frecuencias absolutas,  $R_1, \dots, R_4$  denotan los totales de frecuencia marginal de las cuatro filas,  $C_1, C_2$ , y  $C_3$  denotan los totales de las frecuencias marginales de las tres columna y

$$N = \sum_{i=1}^4 \sum_{j=1}^3 n_{ij}$$

Cuando sólo hay dos filas, como en el ejemplo anterior de  $2 \times 2$ , cada valor de probabilidad de columna es binomial, pero con cuatro filas cada valor de probabilidad de columna es



multinomial. Es bien sabido que un valor de probabilidad multinomial puede obtenerse de una serie interconectada de expresiones binomiales. Por ejemplo, para la columna  $A_1$  de la *Tabla 2*:

$$\binom{C_1}{n_{11}} \binom{C_1 - n_{11}}{n_{21}} \binom{C_1 - n_{11} - n_{21}}{n_{31}} = \frac{C_1!}{n_{11}!(C_1 - n_{11})!} \frac{(C_1 - n_{11})!}{n_{21}!(C_1 - n_{11} - n_{21})!} \frac{(C_1 - n_{11} - n_{21})!}{n_{31}!(C_1 - n_{11} - n_{21} - n_{31})!}$$

Para la columna  $A_2$ :

$$\binom{C_2}{n_{12}} \binom{C_2 - n_{12}}{n_{22}} \binom{C_2 - n_{12} - n_{22}}{n_{32}} = \frac{C_2!}{n_{12}!(C_2 - n_{12})!} \frac{(C_2 - n_{12})!}{n_{22}!(C_2 - n_{12} - n_{22})!} \frac{(C_2 - n_{12} - n_{22})!}{n_{32}!(C_2 - n_{12} - n_{22} - n_{32})!}$$

y para la distribución de frecuencias marginales de las filas:

$$\binom{N}{R_1} \binom{N - R_1}{R_2} \binom{N - R_1 - R_2}{R_3} = \frac{N!}{R_1!(N - R_1)!} \frac{(N - R_1)!}{R_2!(N - R_1 - R_2)!} \frac{(N - R_1 - R_2)!}{R_3!(N - R_1 - R_2 - R_3)!}$$

Por consiguiente, para tablas de contingencia  $r \times c$  se tiene:

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

De esta forma, la ecuación anterior puede generalizarse fácilmente a tablas de contingencia multidireccionales más complejas.

Mientras que esta sección ilustra el cálculo de un valor de probabilidad puntual hipergeométrico, para una prueba de permutación exacta de una tabla de contingencia  $r \times c$  es necesario calcular la medida de asociación seleccionada para las frecuencias de celdas observadas y, a continuación, enumerar exhaustivamente todos los posibles ordenamientos igualmente probables de los  $N$  objetos en las  $rc$  celdas, dadas las distribuciones de frecuencia marginal observadas.

Para cada ordenamiento en el conjunto de referencia de todas las permutaciones de las frecuencias de las celdas, se calcula una medida de asociación,  $T$ , y el valor exacto de la probabilidad puntual hipergeométrica,  $p(n_{ij}|R_i, C_j, N)$  para  $i = 1, \dots, r$  y  $j = 1, \dots, c$ . Sea  $T_0$  el valor del estadístico de prueba observado, es decir, la medida de asociación, entonces, el valor de probabilidad exacto de  $T_0$  es la suma de la probabilidad de los puntos hipergeométricos asociados a los valores de  $T$  calculados en todos los posibles ordenamientos de las frecuencias de las celdas que son iguales o mayores que  $T_0$ .

Cuando el número de disposiciones posibles de las frecuencias es muy grande, las pruebas exactas son poco prácticas y se hacen necesarios los métodos estadísticos de permutación de Monte Carlo. Dadas las distribuciones de frecuencias marginales observadas, los métodos estadísticos de permutación de Monte Carlo generan una muestra aleatoria de todas las posibles ordenaciones de las frecuencias de las celdas, extraídas con reemplazo. El valor de la probabilidad a dos colas del remuestreo es simplemente la proporción de los

valores  $T$  calculados en los ordenamientos seleccionados aleatoriamente que son iguales o mayores que  $T_0$ . En el caso del remuestreo de Monte Carlo los valores de probabilidad hipergeométricos no están implicados, simplemente se necesita la proporción de los valores de las medidas de asociación (valores  $T$ ) iguales o mayores que el valor de la medida de asociación observada ( $T_0$ ).

## 2.2. 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 de intervalo. Las medidas habituales de asociación 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, han demostrado ser menos que satisfactorias debido a las dificultades de interpretación.

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

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

Tabla 2.3: Tabla 3: Notación para la clasificación cruzada de dos variables categóricas,  $A_j$  con  $j=1, \dots, c$  y  $B_i$  con  $i=1, \dots, r$

$B \backslash A$	$A_1$	$A_2$	$\dots$	$A_c$	Total
$B_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1c}$	$n_{1.}$
$B_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2c}$	$n_{2.}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$B_r$	$n_{r1}$	$n_{r2}$	$\dots$	$n_{rc}$	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.c}$	$N$

donde  $A_j$  con  $j = 1, \dots, c$  representan las  $c$  categorías de la variable dependiente  $A$ ,  $B_i$  con  $i = 1, \dots, r$  denotan las  $r$  categorías para la variable independiente  $B$ ,  $n_{ij}$  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 ( $\cdot$ ) la suma parcial de todas las filas o todas las columnas, según la posición del ( $\cdot$ ) en la lista de subíndices. Si el ( $\cdot$ ) está en la primera posición del subíndice la suma es sobre todas las filas y si el ( $\cdot$ ) está en la segunda posición de subíndice, la suma es sobre todas las columnas. 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 3*, se definen

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

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

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

De la misma manera, se definen

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

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

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

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 disjuntas de la variable dependiente A.
- Caso 2: Son conocidas tanto las categorías disjuntas de la variable A como las categorías disjuntas de la variable independiente B.

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

$$W = \sum_{i=1}^r \max(n_{i1}, n_{i2}, \dots, n_{ic})$$

La probabilidad de error es entonces  $N - W$ ; y lo definimos como “errores del segundo tipo” o  $E_2$ . Entonces,  $\lambda_a$  puede expresarse como

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

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, mientras que  $\lambda_a$  y  $\lambda_b$  sean iguales a cero bajo independencia,  $\lambda_a$  y  $\lambda_b$  también podían ser iguales a cero para casos distintos de independencia. Esto hace que tanto  $\lambda_a$  como  $\lambda_b$  sean difíciles de interpretar; en consecuencia,  $\lambda_a$  y  $\lambda_b$  rara vez se encuentran en estudios de la asociación de variables categóricas.

Así, 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 A.

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  $B$  no ayuda a predecir la clasificación  $A$ .
4.  $\lambda_a$  es 1 si y sólo si el conocimiento de un objeto de la categoría  $B$  especifica completamente su categoría  $A$ , 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.

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

Como se ha señalado anteriormente, en 1954 Leo Goodman y William Kruskal propusieron varias medidas de asociación nuevas. Entre las medidas había una medida de predicción asimétrica de reducción proporcional del error,  $t_a$ , para el análisis de una muestra aleatoria de dos variables categóricas. Considere dos variables desordenadas cruzadas,  $A$  y  $B$ , con la variable  $A$  como variable dependiente y la variable  $B$  como variable independiente. La *tabla 3* proporcionaba la notación para la clasificación cruzada.

El estadístico  $t_a$  de Goodman y Kruskal es una medida de la reducción relativa del error de predicción en la que se definen dos tipos de errores. El primer tipo es el error de predicción basado únicamente en el conocimiento de la distribución de la variable dependiente, denominado “error del primer tipo” ( $E_1$ ) y que consiste en el número esperado de errores al predecir las  $c$  categorías de la variable dependiente ( $a_1, \dots, a_c$ ) a partir de la distribución observada de las marginales de la variable dependiente ( $n_{.1}, \dots, n_{.c}$ ). El segundo tipo es el error de predicción basado en el conocimiento de las distribuciones de las de la variable independiente y de la dependiente, denominado “error del segundo tipo” ( $E_2$ ) y que consiste en el número o los errores esperados al predecir las  $c$  categorías de la variable dependiente ( $a_1, \dots, a_c$ ) a partir del conocimiento de las  $r$  categorías de la variable independiente ( $b_1, \dots, b_r$ ).

Para ilustrar los dos tipos de error, se considera la predicción de la categoría  $A_1$  a partir de sólo el conocimiento de su distribución marginal,  $n_{.1}, \dots, n_{.c}$ . Claramente,  $n_{.1}$  de los  $N$  casos totales están en la categoría  $a_1$ , pero se desconoce exactamente qué  $n_{.1}$  de los  $N$  casos. La probabilidad de identificar incorrectamente uno de los  $N$  casos de la categoría  $a_1$  viene dada por:

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

Dado que se requieren  $n_{.1}$  clasificaciones de este tipo, el número de clasificaciones incorrectas esperadas es

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

y, para todas las  $c$  categorías de la variable  $A$ , el número de errores esperados del primer tipo viene dado por:

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

Asimismo, para predecir  $n_{11}, \dots, n_{1c}$  a partir de la categoría independiente  $B_1$ , la probabilidad de clasificar incorrectamente uno de los  $n_{1.}$  casos de la celda  $n_{11}$  es:

$$\frac{n_{1.} - n_{11}}{n_{1.}}$$

Dado que se requieren  $n_{11}$  clasificaciones de este tipo, el número de clasificaciones incorrectas es

$$\frac{n_{11}(n_{1.} - n_{11})}{n_{1.}}$$

y, para todas las  $cr$  celdas, 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:

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

Un valor calculado de  $t_a$  indica la reducción proporcional del error de predicción dado el conocimiento de la distribución de la variable independiente  $B$  por encima del conocimiento de la distribución de la variable dependiente  $A$ . 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  $B$  se considera la variable dependiente y la variable  $A$  la variable independiente, entonces el estadístico de prueba de Goodman y Kruskal  $t_b$  y el parámetro poblacional asociado  $\tau_b$  se definen análogamente.

Como el parámetro  $\tau_a$  toma valores de 0 a 1, posee una interpretación clara y significativa de reducción proporcional del error, y se caracteriza por una alta validez intuitiva; el estadístico de prueba  $t_a$  plantea dificultades si la hipótesis nula es que  $H_0 : \tau_a = 0$ . 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 que se estableció la normalidad asintótica para  $t_a$  y hasta 1972 no se obtuvo la varianza asintótica correcta para  $t_a$ , pero sólo para  $0 < \tau_a < 1$ .

En 1971, Richard Light y Barry Margolin desarrollaron  $R^2$ , una técnica de análisis de la varianza para las variables de respuesta categóricas, llamada CATANOVA para CAtegorical ANalysis Of VAriance. Light y Margolin aparentemente no sabían que el  $R^2$  era idéntico al  $t_a$  de Goodman y Kruskal y que habían resuelto el viejo problema de probar  $H_0 : \tau_a = 0$ . La igualdad entre  $R^2$  y  $t_a$  fue reconocida por primera vez por Särndal en 1974 y posteriormente discutida por Margolin y Light, donde demostraron que  $t_a(N-1)(r-1)$  se distribuía según una chi-cuadrado con  $(r-1)(c-1)$  grados de libertad bajo  $H_0 : \tau_a = 0$  cuando  $N \rightarrow \infty$ .

## 2.4. 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 distribuciones  $g$ . En una prueba de homogeneidad, una de las distribuciones marginales se conoce antes de recoger los datos, 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 de *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 bastante 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. La prueba de homogeneidad más utilizada es la prueba chi-cuadrado de Pearson con  $gl = (r-1)(g-1)$  grados de libertad. La prueba chi-cuadrado de homogeneidad de Pearson supone la hipótesis nula de que no hay diferencia en las proporciones de sujetos en un conjunto de categorías mutuamente excluyentes entre dos o más poblaciones.

La prueba de homogeneidad chi-cuadrado de Pearson es una prueba simétrica, que produce un solo valor para una tabla de contingencia  $r \times g$ . Por el contrario, una prueba asimétrica arroja 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.

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. Aunque a veces es necesario reducir el nivel de medición cuando no se pueden cumplir los requisitos de distribución, en general no es aconsejable utilizar una prueba estadística que descarte información importante.

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 2.4: Notación para la clasificación cruzada de dos variables categóricas,  $A_j$  para  $j=1, \dots, g$  y  $B_i$  para  $i=1, \dots, r$

$B \backslash A$	$A_1$	$A_2$	$\dots$	$A_g$	Total
$B_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1g}$	$n_{1.}$
$B_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2g}$	$n_{2.}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$B_r$	$n_{r1}$	$n_{r2}$	$\dots$	$n_{rg}$	$n_{r.}$
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, es una alternativa atractiva a la prueba simétrica de homogeneidad, la chi-cuadrado. El estadístico de la prueba viene dado por:

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

donde  $B$  es la variable dependiente y el parámetro poblacional asociado se denota como  $\tau_b$ . Si se considera que la variable dependiente es  $A$ , el estadístico de la prueba viene dado por:

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

y el parámetro poblacional asociado es  $\tau_a$ .

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  $\chi^2 = Nt_b$ , 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 el valor de la probabilidad del estadístico de prueba  $t_b$  o  $t_a$  calculado: procedimientos exactos, de remuestreo de Montecarlo y asintóticos. Explicaciones considerando sólo  $t_b$ , pero 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 categorías  $rg$  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 de la probabilidad exacta de un estadístico de prueba  $t_b$  observado es la suma de los valores de probabilidad puntuales hipergeométricos asociados a valores mayores o iguales a  $t_b$ .
- Valores de probabilidad 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 de permutación por remuestreo. Bajo la hipótesis nula,  $H_0 : \tau_b = 0$ , las pruebas de permutación por remuestreo 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 de prueba deseado. El valor de la probabilidad de remuestreo de Montecarlo de un estadístico de prueba  $t_b$  observado es simplemente la proporción de los valores seleccionados aleatoriamente de  $t_b$  iguales o mayores que el valor observado de  $t_b$ .
- Valores de probabilidad asintótica: Bajo la hipótesis nula,  $H_0 : \tau_b = 0$ , como  $N \rightarrow \infty$ ,  $t_b(N-1)(g-1)$  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  $t_b(N-1)(g-1)$ .

## 2.5. Medidas de concordancia

La medición de la concordancia es un caso especial de medición de la asociación entre dos o más variables. Una serie de 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.

Su uso surge se asignan objetos a un conjunto de categorías desordenadas y disjuntas. En 1957, Robinson explicó que la concordancia estadística requiere que los valores emparejados sean idénticos, mientras que la correlación sólo requiere que los valores emparejados estén vinculados por alguna función matemática. Por tanto, la concordancia es una medida más restrictiva que la correlación. Robinson argumentó que la distinción entre acuerdo y correlación lleva a la conclusión de que una estimación lógicamente correcta de la



fiabilidad de una prueba viene dada por el coeficiente de correlación intraclase en lugar del coeficiente de correlación de Pearson (interclase) y que el concepto de acuerdo, en lugar de correlación, es la base adecuada de la teoría de la fiabilidad.

Una medida de acuerdo entre evaluadores debería, como mínimo, incorporar siete atributos básicos:

- 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. Varios investigadores han defendido las medidas de concordancia corregidas por el azar, como Brennan y Prediger, Cicchetti, Showalter y Tyrer, Cohen, Conger y Krippendorff. Aunque algunos otros han argumentado en contra, por ejemplo, Armitage, Blendis y Smyllie, y Goodman y Kruskal, los partidarios de las medidas de concordancia corregidas por el azar superan ampliamente a los detractores.
- Una medida de concordancia entre evaluadores posee una ventaja añadida si es directamente aplicable a la evaluación de la fiabilidad. Robinson, en particular, hizo hincapié en que la fiabilidad no podía medirse simplemente mediante alguna función de la correlación producto-momento de Pearson, y argumentó que el concepto de concordancia debería ser la base de la teoría de la fiabilidad, no la correlación.
- 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. Además, Graham y Jackson señalaron que la elevación al cuadrado de las diferencias entre los valores, es decir, la ponderación cuadrática, da como resultado una medida de asociación, no de concordancia. Por lo tanto, la distancia Euclídea es una propiedad deseada para las medidas de concordancia entre evaluadores.
- 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 y a datos categóricos continuos por Conger. Una extensión de la medida de acuerdo kappa de Cohen a datos ordinales totalmente clasificados y a datos de intervalo fue proporcionada por Berry y Mielke en 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. Landis y Koch consideraron la concordancia entre varios calificadores en términos de una opinión mayoritaria. Light se centró en una extensión de la medida kappa de Cohen de concordancia entre calificadores a múltiples calificadores que se basaba en la media de todos los valores kappa por pares. Lamentablemente, la medida propuesta por Fleiss dependía de la proporción media de calificadores que estaban de acuerdo en la clasificación de cada observación y la formulación de Landis y Koch se vuelve computacionalmente prohibitiva si el número de observadores o el número de categorías de respuesta es grande. Además, la extensión de kappa propuesta por Fleiss no se redujo a una kappa de Cohen cuando el número de calificadores era de dos.

### 2.5.1. Medida de concordancia de Robinson

Una de las primeras medidas de concordancia máxima corregida fue desarrollada por W.S. Robinson en 1957. Supongamos que hay  $k = 2$  jueces y califican independientemente  $N$  objetos. Robinson argumentó que la correlación producto-momento de Pearson (interclase) calculada entre las calificaciones de dos jueces era una medida inadecuada de concordancia porque mide el grado en que los valores emparejados de las dos variables son proporcionales, cuando se expresan como desviaciones de sus medias, en lugar de ser idénticos. 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. Robinson definió  $A$  como

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

donde  $D$  (de 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.

Robinson observó que, por sí misma,  $D$  no es una medida muy útil porque implica  $X_1$  y  $X_2$ . 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

Es bien sabido que 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  $i = 1, \dots, N$ :

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

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 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.

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

Una de las primeras medidas de concordancia corregida por el azar fue introducida por William Scott en 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  $c \times r$ , 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

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

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

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

En esta configuración,  $p_o$  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 por azar,  $p_o - p_e$  es la proporción de concordancia más allá de la esperada por el azar,  $1 - p_e$  es la máxima proporción posible de concordancia más allá de la esperada por el azar, y  $\pi$  de Scott es la proporción de concordancia entre los dos jueces, una vez eliminada la concordancia por azar.

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.

### 2.5.3. 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. 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. Mientras que el kappa original (no ponderado) no distinguía entre magnitudes de desacuerdo, el kappa ponderado incorporaba la magnitud 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 con proporciones para las entradas de las celdas:

Tabla 2.5: Clasificación cruzada cxc con proporciones en celdas

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

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

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

El kappa de Cohen también puede definirse en términos de valores de frecuencia absoluta, lo que hace que los cálculos sean algo más sencillos. Así,

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

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

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

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.

### 2.5.4. Aplicación con Varios Jueces

La medida  $\kappa$  de Cohen de la concordancia entre evaluadores corregida por el azar fue originalmente 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  $c$  cortes. Sean  $n_{ijk}$ ,  $R_i$ ,  $C_j$  y  $S_k$  las frecuencias esperadas de las celdas y las frecuencias marginales de las filas, las columnas y los cortes para  $i, j, k = 1, \dots, c$ . La frecuencia total viene dada por:

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

El estadístico de la prueba kappa no ponderada de Cohen para una tabla de contingencia de tres vías 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:

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

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 el valor exacto de la probabilidad 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 de probabilidad  $\kappa_o$  exacto 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. Para una tabla de contingencia de tres dimensiones con  $r$  filas,  $c$  columnas y  $s$  cortes 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 cortes  $IR, IC$  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 cruce  $i = j = k = 1$ , 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

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

## 2.6. Test Q de McNemar para el cambio

En 1947, el psicólogo Quinn McNemar propuso una prueba de cambio derivada de la prueba  $t$  de pares emparejados para proporciones. Una aplicación típica es analizar respuestas binarias, codificadas con 0s y 1s, en  $g = 2$  periodos de tiempo para cada uno de los  $N \geq 2$  sujetos, como Éxito y Fracaso, Sí y No, De acuerdo y Desacuerdo, o A favor y en contra. Si las cuatro celdas se identifican como en la siguiente tabla:

Tabla 2.6: Notación de una tabla de clasificación cruzada 2x2 para el test Q de McNemar

Tiempo 1\Tiempo 2	Favor	Contra	Total
Favor	$A$	$B$	$A + B$
Contra	$C$	$D$	$C + D$
Total	$A + C$	$B + D$	$N$

entonces la prueba de McNemar para el cambio viene dada por:

$$Q = \frac{(B - C)^2}{B + C},$$

donde  $N = A + B + C + D$  y  $B$  y  $C$  representan las dos celdas de cambio, es decir, de Favor a Contra y de Contra a Favor.

Alternativamente, la prueba Q de McNemar puede considerarse como un test de bondad de ajuste de una chi-cuadrado con dos categorías, donde las frecuencias observadas,  $O_1$  y  $O_2$ , corresponden a las celdas  $B$  y  $C$ , respectivamente, y las frecuencias esperadas,  $E_1$  y  $E_2$ , vienen dadas por  $E_1 = E_2 = \frac{B+C}{2}$ , es decir se espera que la mitad de los sujetos



cambien en una dirección (por ejemplo, de Favor a Contra) y la otra mitad en la otra dirección (por ejemplo, de Contra a Favor), bajo la hipótesis nula de que no hay cambios del Tiempo 1 al Tiempo 2.

Sea

$$E = \frac{B + C}{2}$$

el valor esperado en el que, por azar, la mitad de los cambios son de Favor a Contra y la otra mitad son de Contra a Favor. Entonces, la bondad de ajuste de una chi-cuadrado para las dos categorías de cambio viene dada por:

$$\chi^2 = \frac{(B - E)^2}{E} + \frac{(C - E)^2}{E} = \frac{B^2}{E} + \frac{C^2}{E} + 2E - 2B - 2C .$$

Sustituyendo  $E$  por  $(B + C)/2$  se obtiene:

$$\begin{aligned} \chi^2 &= \frac{2B^2}{B + C} + \frac{2C^2}{B + C} + B + C - 2B - 2C = \frac{2B^2}{B + C} + \frac{2C^2}{B + C} - B - C = \\ &= \frac{2B^2 + 2C^2 - B(B + C) - C(B + C)}{B + C} = \frac{B^2 - 2BC + C^2}{B + C} \Rightarrow \\ &\Rightarrow \chi^2 = \frac{(B - C)^2}{B + C} \end{aligned}$$

## 2.7. Test Q de Cochran para el cambio

La variable dicotómica desempeña un gran papel y tiene muchas aplicaciones en la investigación y la medición. Convencionalmente, se asigna un valor de uno a cada elemento de la prueba que un sujeto responde correctamente y un cero a cada respuesta incorrecta.

En 1950, William Cochran publicó un artículo sobre “La comparación de porcentajes en muestras emparejadas” . En este breve pero formativo artículo, Cochran describió una prueba de igualdad de proporciones emparejadas que ahora se utiliza ampliamente en la investigación educativa y psicológica. El emparejamiento puede basarse en las características de diferentes sujetos o en los mismos sujetos bajo diferentes condiciones.

El test  $Q$  de Cochran puede considerarse una extensión del test de McNemar a tres o más condiciones de tratamiento. Supongamos que se observa una muestra de  $N \geq 2$  sujetos en una situación en la que en cada sujeto actúa individualmente bajo cada una de las  $k \geq 1$  condiciones experimentales diferentes. La actuación se puntúa como un éxito (1) o como un fracaso (0). La pregunta de investigación evalúa si la proporción real de éxitos es constante a lo largo de los  $k$  períodos de tiempo.

El estadístico del test  $Q$  de Cochran para el análisis de  $k$  condiciones de tratamiento (columnas) y  $N$  sujetos (filas) viene dado por:

$$Q = \frac{(k-1)(k \sum_{j=1}^k C_j^2 - A^2)}{kA - B}$$

donde

$$C_j = \sum_{i=1}^N x_{ij}$$

es el número de 1s en la  $j$ -ésima de las  $k$  columnas,

$$R_i = \sum_{j=1}^k x_{ij}$$

es el número de 1s en la  $i$ -ésima de las  $N$  filas,

$$A = \sum_{i=1}^N R_i, \quad B = \sum_{i=1}^N R_i^2,$$

y  $x_{ij}$  indica la entrada de la celda de 0 ó 1 asociada a la  $i$ -ésima de las  $N$  filas y la  $j$ -ésima de las  $k$  columnas. La hipótesis nula estipula que cada una de los

$$M = \prod_{i=1}^N \binom{k}{R_i}$$

ordenamientos distinguibles de 1s y 0s dentro de cada una de las  $N$  filas ocurren con igual probabilidad, dado que los valores de  $R_1, \dots, R_N$  son fijos.

## 2.8. Una medida sobre el Tamaño del Efecto para el test Q de Cochran

Las medidas del tamaño del efecto son cada vez más importantes a la hora de informar sobre los resultados de la investigación. Lamentablemente, no existen medidas del tamaño del efecto para muchas de las pruebas estadísticas más comunes. En esta sección, se presenta una medida del tamaño del efecto corregida por el azar para el test  $Q$  de Cochran para proporciones relacionadas.

Consideremos un enfoque alternativo al test  $Q$  de Cochran en el que se aplican  $g$  tratamientos de forma independiente a cada uno de los  $N$  individuos, con el resultado de cada aplicación del tratamiento registrado como 1 o 0, representando cualquier dicotomización adecuada de los resultados del tratamiento, es decir, un diseño de bloques aleatorios en el que los individuos son los bloques y los resultados del tratamiento se registran como 1 o 0. Sean  $x_{ij}$  las medidas de respuesta 1 y 0 registradas para  $i = 1, \dots, N$  y  $j = 1, \dots, g$ . Entonces, el estadístico del test de Cochran puede definirse como

$$Q = \frac{g-1}{2 \sum_{i=1}^N p_i(1-p_i)} [2(\sum_{i=1}^N p_i)(N - \sum_{i=1}^N p_i) - N(N-1)\delta] ,$$

donde

$$\delta = [g \binom{N}{2}]^{-1} \sum_{k=1}^g \sum_{i=1}^{N-1} \sum_{j=i+1}^N |x_{ik} - x_{jk}|$$

y

$$p_i = \frac{1}{g} \sum_{j=1}^g x_{ij} \quad \text{para } i = 1, \dots, N ,$$

es decir, la proporción de valores 1 para el  $i$ -ésimo individuo. Se observa que en esta representación la variación de  $Q$  depende totalmente de  $\delta$ .

En 1979, Acock y Stavig propusieron un valor máximo para  $Q$  dado por:

$$Q_{max} = N(g-1)$$

El valor máximo de  $Q$  de Acock y Stavig fue empleado por Serlin, Carr y Marascuilo para proporcionar una medida del tamaño del efecto de la  $Q$  de Cochran dada por:

$$\hat{\eta}_Q^2 = \frac{Q}{Q_{max}} = \frac{Q}{N(g-1)} ,$$

que estandarizó la  $Q$  de Cochran por un valor máximo. Lamentablemente, el valor de  $Q_{max} = N(g-1)$  defendido por Acock y Stavig sólo se alcanza cuando cada sujeto  $g$ -tupla es idéntico y hay al menos un 1 y un 0 en cada  $g$ -tupla. Así  $\hat{\eta}_Q^2$  es una medida de “máxima corrección” del tamaño del efecto y  $0 \leq \hat{\eta}_Q^2 \leq 1$  bajo estas singulares condiciones.

Supongamos que  $0 < p_i < 1$  para  $i = 1, \dots, N$  ya que  $p_i = 0$  y  $p_i = 1$  no son informativos. Si  $p_i$  es constante para  $i = 1, \dots, N$ , entonces  $Q_{max} = N(g-1)$ . Sin embargo, para la gran mayoría de los casos en los que  $p_i \neq p_j$  para  $i \neq j$ ,  $Q_{max} < N(g-1)$ . Por lo tanto, el uso rutinario de establecer  $Q_{max} = N(g-1)$  es problemático y conduce a resultados cuestionables.

También hay que señalar que  $\hat{\eta}_Q^2$  es un miembro de la familia  $V$  de medidas de asociación nominal basada en el estadístico de prueba  $V^2$  de Cramér dado por:

$$V^2 = \frac{\chi^2}{\chi_{max}^2} = \frac{\chi^2}{N[\min(r-1, c-1)]} ,$$

donde  $r$  y  $c$  denotan el número de filas y columnas de una tabla de contingencia  $r \times c$ . Otros miembros de la familia  $V$  son el  $\phi^2$  de Pearson para tablas de contingencia  $2 \times 2$  y  $T^2$  de Tschuprov para tablas de contingencia  $r \times c$  donde  $r = c$ . Las dificultades para interpretar  $V^2$  se extienden a las de  $\hat{\eta}_Q^2$ .

Wickens observó que la  $V^2$  de Cramér carece de una interpretación intuitiva que no sea como un escalado de chi-cuadrado, lo que limita su utilidad. Asimismo, Costner observó que  $V^2$  y otras medidas basadas en la chi-cuadrado de Pearson carecen de interpretación para valores distintos de 0 y 1, o el máximo, dadas las distribuciones de frecuencia marginal observadas. Agresti y Finlay también señalaron que la  $V^2$  de Cramér es muy difícil de interpretar y recomendaron otras medidas. Blalock señaló que “todas las medidas basadas en el chi cuadrado son de alguna manera arbitrarias por naturaleza, y sus interpretaciones dejan mucho que desear. . . todas ellas dan mayor peso a las columnas o filas que tienen las marginales más pequeñas en lugar de que a las que tienen las marginales más grandes”. Ferguson discutió el problema de utilizar frecuencias marginales idealizadas, y Guilford señaló que medidas como la  $\phi^2$  de Pearson, la  $T^2$  de Tschuprov y la  $V^2$  de Cramér necesariamente subestiman la magnitud de la asociación presente. Berry, Martin y Olson consideraron estas cuestiones con respecto a las tablas de contingencia 2Ö2, y Berry, Johnston y Mielke discutieron con cierto detalle los problemas de utilizar  $\phi^2$  de Pearson,  $T^2$  de Tschuprov y  $V^2$  de Cramér como medidas del tamaño del efecto. Dado que  $\hat{\eta}_Q^2$  es simplemente un caso especial del  $V^2$  de Cramér, presenta los mismos problemas de interpretación.

### 2.8.1. Una medida del Tamaño del Efecto corregida por el azar

Las medidas de tamaño del efecto corregidas por el azar tienen mucho que ver con las medidas corregidas por el máximo. Una medida del tamaño del efecto corregida por el azar es una medida de concordancia entre los  $N$  individuos sobre  $g$  tratamientos, corregida por el azar. Varios investigadores han defendido las medidas del tamaño del efecto corregidas por el azar, como Brennan y Prediger, Cicchetti, Showalter y Tyrer, Conger y Krippendorff. Una medida corregida por el azar es 0 en condiciones de azar, 1 cuando la concordancia entre los  $N$  sujetos es perfecta, y negativa en condiciones de desacuerdo. Algunas medidas bien conocidas corregidas por el azar son el coeficiente de concordancia entre evaluadores de Scott, la medida de concordancia  $u$  de Kendall y Babington Smith, los coeficientes no ponderados y ponderados de concordancia entre evaluadores de Cohen y la medida de la regla de pie de Spearman. Bajo ciertas condiciones, el coeficiente de correlación de orden de rango de Spearman es también una medida de concordancia corregida por el azar, es decir, cuando las variables  $x$  e  $y$  toman valores de 1 a  $N$  sin valores empatados, o cuando la variable  $x$  incluye valores empatados y la variable  $y$  es una permutación de la variable  $x$ , entonces el coeficiente de correlación de orden de rango de Spearman es tanto una medida de correlación como una medida de concordancia corregida por el azar.

Sean  $x_{ij}$  las medidas de respuesta (0, 1) para  $i = 1, \dots, N$  bloques y  $j = 1, \dots, g$  tratamientos, entonces

$$\delta = [g \binom{N}{2}]^{-1} \sum_{k=1}^g \sum_{i=1}^{N-1} \sum_{j=i+1}^N |x_{ik} - x_{jk}|$$

Bajo la hipótesis nula de que la distribución de  $\delta$  asigna igual probabilidad a cada una de las  $M = (g!)^N$  posibles asignaciones de las  $g$  medidas de respuesta dicotómicas a las  $g$  posiciones de tratamiento para cada uno de los  $N$  individuos, el valor medio de  $\delta$  viene dado por

$$\mu_\delta = \frac{2}{N(N-1)} \left[ \left( \sum_{i=1}^N p_i \right) \left( N - \sum_{i=1}^N p_i \right) - \sum_{i=1}^N p_i (1 - p_i) \right],$$

donde

$$p_i = \frac{1}{g} \sum_{j=1}^g x_{ij} \quad \text{para } i = 1, \dots, N.$$

Entonces, una medida del tamaño del efecto corregida por el azar puede definirse como

$$\mathfrak{R} = 1 - \frac{\delta}{\mu_\delta}.$$

### 2.8.2. Ventajas de la medida $\mathfrak{R}$ del tamaño del efecto

Las medidas del tamaño del efecto corregidas por el azar, como  $\mathfrak{R}$ , poseen claras ventajas de interpretación sobre las medidas del tamaño del efecto corregidas por el máximo, como  $\hat{\eta}_Q^2$ . El problema con  $\hat{\eta}_Q^2$  radica en la forma en que se maximiza  $\hat{\eta}_Q^2$ . El denominador de  $\hat{\eta}_Q^2$ ,  $Q_{max} = N(g-1)$ , estandariza el valor observado de  $Q$  para el tamaño de la muestra ( $N$ ) y el número de tratamientos ( $g$ ). Desafortunadamente,  $N(g-1)$  no estandariza  $Q$  para los datos en los que se basa, sino que estandariza  $Q$  en otro conjunto hipotético de datos no observados.

$\mathfrak{R}$  es la alternativa preferida a  $\hat{\eta}_Q^2$  como medida del tamaño del efecto por dos razones:

- Puede lograr un tamaño del efecto de 1 para los datos observados, mientras que esto es a menudo imposible para  $\hat{\eta}_Q^2$ .
- Es una medida del tamaño del efecto corregida por el azar, lo que significa que es 0 en condiciones de azar, 1 cuando la concordancia entre los  $N$  sujetos es perfecta, y negativo en condiciones de desacuerdo.

Por lo tanto, posee una interpretación clara que se corresponde con el coeficiente de concordancia entre evaluadores de Cohen y otras medidas corregidas por el azar que son familiares para la mayoría de los investigadores. Por otro lado otro lado,  $\hat{\eta}_Q^2$  no posee una interpretación significativa, excepto para los valores límite de  $Q = 0$  y  $Q = 1$ .

## 2.9. 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. Insatisfechos con las medidas existentes de asociación nominal, Leik y Gove 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. La nueva medida se denominó  $d_N^c$ , donde  $d$  indicaba el índice, siguiendo otros índices como  $d_{yx}$  y  $d_{xy}$  de Somers; el subíndice  $N$  indicaba la relevancia de  $d$  para una variable dependiente nominal; y el superíndice  $c$  indicaba que la medida estaba corregida por las restricciones impuestas por las distribuciones de frecuencia marginal.

Al igual que  $d_N^c$ , muchas medidas de asociación para dos variables se han basado en comparaciones por pares de las diferencias entre las observaciones. 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. Sean  $n_{i.}$ ,  $n_{.j}$ , y  $n_{ij}$  la frecuencia marginal de la fila  $i$ , la frecuencia marginal de la columna  $j$  y el número de objetos en la celda  $ij$ , respectivamente, para  $i = 1, \dots, r$  y  $j = 1, \dots, c$ , y sea  $N$  el número total de objetos en la tabla de contingencia  $r \times c$ . 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
- pares discordantes
- pares ¿vinculados? en la variable  $y$  pero que difieren en la variable  $x$
- pares ¿vinculados? en la variable  $x$  pero que difieren en la variable  $y$
- pares ¿vinculados? en ambas variables  $x$  e  $y$

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 ¿vinculados? 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 ¿vinculados? 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 ¿vinculados? 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,

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

### 2.9.1. Tabla de Contingencia Observada

A partir de la tabla de contingencia observada se obtienen los parámetros  $C$ ,  $D$ ,  $T_x$ ,  $T_y$  y  $T_{xy}$ .

### 2.9.2. Tabla de Contingencia Esperada

Usando la tabla de contingencia observada, se contruye la tabla de contingencia esperada de la siguiente forma:

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

A partir de la tabla de contingencia esperada se obtienen los parámetros  $C'$ ,  $D'$ ,  $T'_x$ ,  $T'_y$  y  $T'_{xy}$ .

Afortunadamente, existe una forma más eficiente de calcular  $C'$ ,  $D'$ ,  $T'_x$ ,  $T'_y$  y  $T'_{xy}$  sin calcular primero los valores esperados. En primer lugar, dadas las distribuciones de frecuencias marginales de filas y columnas observadas se calcula el número de pares de valores esperados de frecuencia de celdas esperadas vinculados a las variables  $x$  e  $y$ ,

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

A continuación, se calcula el número de pares de valores de frecuencia de celda esperados vinculados a la variable  $y$ ,

$$T'_y = \frac{1}{2} \sum_{i=1}^r n_{i.}^2 - \frac{N}{2} - T'_{xy}$$

Del mismo modo, se calcula el número de pares de valores de frecuencia de celda esperados vinculados a la variable  $x$ ,

$$T'_x = \frac{1}{2} \sum_{j=1}^c n_{.j}^2 - \frac{N}{2} - T'_{xy}$$

Por último, se calcula el número de pares concordantes y discordantes de valores de frecuencia de celdas esperados,

$$C' = D' = \frac{1}{2} \left[ \frac{N(N-1)}{2} - T'_x - T'_y - T'_{xy} \right]$$

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.

### 2.9.3. Tabla de Contingencia Maximal

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, el número de pares concordantes de valores de frecuencia de celda maximizada es

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

el número de pares discordantes de valores de frecuencia celular maximizada es

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

el número de pares de valores de frecuencia de celda maximizados vinculados a la variable  $x$  es

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

el número de pares de valores de frecuencia celular maximizados vinculados a la variable  $y$  es

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



y el número de pares de valores de frecuencia celular maximizada empatados en ambas variables  $x$  e  $y$  es

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

Entonces,

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

#### 2.9.4. Cálculo de la $d_N^c$ de Leik y Gove

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, a la manera de las medidas de asociación de reducción proporcional en el 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 vinculados en la variable  $x$  puede utilizarse en lugar del número de valores vinculados 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 vinculados tanto en  $x$  como en  $y$ . Así,

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

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

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

### **2.9.5. Un test de permutación para $d_N^c$**

Leik y Gove no proporcionaron un error estándar para el estadístico de prueba  $d_N^c$ . Por otro lado, las pruebas de permutación no suponen ni requieren el conocimiento de los errores estándar. 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 por la distribución discreta de permutación de  $T_y$  sólo, 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,  $T_y$ .

## **2.10. Un problema de ocupación de la matriz**

## **2.11. Prueba de probabilidad exacta de Fisher**

## **2.12. Análisis de tablas de $2 \times 2 \times 2$**

# Capítulo 3

## Título del Capítulo

### 3.1. Primera sección



# Capítulo 4

## Título del Capítulo

### 4.1. Primera sección



# Apéndice A

## Apéndice: Título del Apéndice

### A.1. Primera sección





# Apéndice B

## Apéndice: Título del Apéndice

### B.1. Primera sección



# Bibliografía

- JJ Allaire, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. *rmarkdown: Dynamic Documents for R*, 2021. URL <https://CRAN.R-project.org/package=rmarkdown>. R package version 2.11.
- Pedro L. Luque-Calvo. *Escribir un Trabajo Fin de Estudios con R Markdown*, 2017.
- Pedro L. Luque-Calvo. *Cómo crear Tablas de información en R Markdown*, 2019.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL <https://www.R-project.org/>.
- RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2015. URL <http://www.rstudio.com/>.
- Techopedia. "definition - what does business intelligence (bi) mean?". Disponible en <https://www.techopedia.com/definition/345/business-intelligence-bi>, 2017.
- Hadley Wickham, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*, 2021a. URL <https://CRAN.R-project.org/package=ggplot2>. R package version 3.3.5.
- Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. *dplyr: A Grammar of Data Manipulation*, 2021b. URL <https://CRAN.R-project.org/package=dplyr>. R package version 1.0.7.
- Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2021. URL <https://yihui.org/knitr/>. R package version 1.34.