

Índice general

1. Introducción	3
1.1. Definiciones	3
1.1.1. Definición de Medición	4
1.1.2. Definición de Asociación	4
1.2. Dimensiones de la asociación	4
1.2.1. Nivel de Medición	5
1.2.2. Simetría y Asimetría	5
1.2.3. Asociación unidireccional y bidireccional	5
1.2.4. Clasificación cruzada	5
1.2.5. Correlación, asociación y concordancia	5
1.3. Criterios para las medidas de asociación	6
1.4. Grado de Asociación	7
Bibliografía	9

Capítulo 1

Introducción

Aunque existen una gran cantidad de métodos para medir la magnitud de la asociación entre dos variables, hay grandes dificultades 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 2×2 , en las que el nivel de medición es a veces irrelevante.

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

Existen dos enfoques principales para determinar los p-valores de 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, algunas veces, son inapropiadas; 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.

Por tanto, en este Trabajo Fin de Grado, se usará el enfoque de permutación para la medición de la asociación estadística, definida ampliamente para incluir medidas de correlación, asociación y concordancia.

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

Según Cowles, la medición es la mejor manera de describir con precisión los acontecimientos y las relaciones entre ellos [Cowles, 2001]. 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 a los fenómenos empíricos de 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.

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.

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.

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.

1.2.4. 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 para ambos casos.

1.2.5. 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 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 . 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 π 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 [Berry et al., 2010].

((aquí podría meter gráficos para visualizar el ejemplo))

1.3. Criterios para las medidas de asociación

Varios investigadores han escrito sobre criterios importantes para las medidas de asociación, sobre todo Costner [Costner, 1965] y Goodman y Kruskal [Goodman and Kruskal, 1954]. Sin embargo, esta sección se basa principalmente en los criterios que Weiss [Weiss, 1968] 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.
- **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.
- **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.
- **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.
- **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”.
- **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.

Bibliografia

Kenneth J. Berry, Janis Johnston, and Paul W. Jr. Mielke. *The Measurement of Association: A Permutation Statistical Approach*. 2010.

Herbert L. Costner. *Criteria for Measures of Association*. American Sociological Review, 1965.

M. Cowles. *Statistics in Psychology: An Historical Perspective*. 2 edition, 2001.

L.A. Goodman and W.H. Kruskal. Measures of association for cross classifications. *J. Am. Stat. Assoc.*, 49, 1954.

R.S. Weiss. *Statistics in Social Research: An Introduction*. Wiley, 1968.