



EL ÍNDICE KAPPA DE CONCORDANCIA

Rubén Belmonte Coloma

Muchos estudios de las áreas biológicas, psicológicas, sociológicas o politológicas requieren un análisis de concordancia entre variables ordinales, en general este análisis suele quedarse en una simple relación no direccional, mediante la prueba Chi-cuadrado ó la de verosimilitud.

Los avances del análisis de datos en los últimos tiempos hacen énfasis en algunos estadísticos de asociación con dirección, entre ellos, los de concordancia tienen un uso cada vez más frecuente.

En algunos casos, como es el caso en salud pública, se desea observar la concordancia entre varios métodos o evaluadores que clasifican al paciente (o el resultado de una observación) según una serie de posibilidades (categorías) mutuamente excluyentes. Partiremos de un análisis del caso dicotómico donde se desea comparar dos métodos de clasificación, tal como se presenta en el siguiente cuadro:

		Método B		Total
		Positivo	Negativo	
Método A	Positivo	a	c	f1
	Negativo	b	d	f2
Total		c1	c2	n

Intuitivamente la expresion:

$$(a + d) / n$$

Conocida como proporción de coincidencias frente al total de sujetos: puede ser una medida de concordancia, sin embargo, es previsible que encontremos concordancia entre ellos por puro azar, por ejemplo arrojando dos monedas que simulen los métodos es previsible encontrar en promedio del orden de un 50% de coincidencias.

Bajo el supuesto de que el sistema A sea un método establecido con todo el rigor científico y el método B es la opinión probablemente subjetiva de un individuo cualquiera; es previsible encontrar un cierto grado de concordancia debido en parte al azar.

Para salvar el efecto del azar se define el índice de concordancia kappa dado por Cohén (1960) de la siguiente manera:



$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Donde P_o es la proporción de concordancia observada y P_e es la proporción de concordancia esperada por puro azar.

En caso de acuerdo perfecto la concordancia será 1, por lo que $1 - P_e$ representa el margen de acuerdo posible no atribuible al azar. De ese margen nosotros observamos probablemente sólo una parte $P_o - P_e$, salvo que haya acuerdo perfecto $P_o=1$.

Así pues, en caso de concordancia perfecta el valor de kappa es 1; si la concordancia observada es igual a la esperada kappa vale 0; y en el caso de que el acuerdo observado sea inferior al esperado el índice kappa es menor que cero.

Para calcular P_e , la concordancia esperada, el razonamiento es el siguiente: de acuerdo con la tabla anterior la probabilidad de que el método A clasifique a un sujeto como positivo podemos estimarla como f_1/n ; mientras que la correspondiente probabilidad del método B la estimaremos como c_1/n .

Si consideramos que existe independencia entre ambos métodos de clasificación, la probabilidad de que coincidan clasificando al mismo sujeto como positivo será entonces el producto de las dos probabilidades (sucesos independientes). De la misma manera que se calcula los valores esperados para la construcción del estadígrafo Chi-cuadrado.

Aplicando el mismo razonamiento calculamos la probabilidad de que se produzca acuerdo entre los métodos al clasificar a un sujeto como negativo, y entonces la probabilidad de acuerdo para cualquiera de las dos clasificaciones será la suma de ambos valores, esto es:

$$P_e = \frac{f_1 \cdot c_1 \cdot f_2 \cdot c_2}{n^2}$$

La distribución exacta del estadígrafo kappa requiere de un tratamiento que no se presenta en este artículo, sin embargo para muestras multinomiales, el estimador de k es asintóticamente normal y su varianza puede encontrarse en Agresti (1990).

Según Cohén y Swerdlik (1998), el coeficiente Kappa fue propuesto originalmente por Cohén (1960) para el caso de dos evaluadores o dos métodos, por lo que a menudo



se le conoce como Kappa de Cohén, y fue generalizado para el caso de más de dos evaluadores por Fleiss, por lo que a veces también se habla del índice Kappa de Fleiss.

Desde el punto de vista descriptivo Landis y Koch propusieron unos márgenes para valorar el grado de acuerdo en función del índice Kappa que se presentan en la siguiente tabla:

Kappa	Grado de acuerdo
<0	Sin acuerdo
0 - 0,2	Insignificante
0,2 - 0,4	Bajo
0,4 - 0,6	Moderado
0,6 - 0,8	Bueno
0,8 - 1	Muy Bueno

Este índice se puede generalizar para clasificaciones multinomiales (más de dos categorías) y para más de dos evaluadores, siendo similar su interpretación.

En el caso de más de dos categorías, además del índice de concordancia global puede ser interesante determinar el grado de concordancia específico en alguna de las categorías (o en todas), lo que equivale a convertir el resultado posible en dos únicas respuestas. De esta manera para cada una de las categorías vamos convirtiendo la tabla original en tablas 2x2 y podemos entonces calcular el valor del correspondiente índice kappa como si se tratara de una variable dicotómica.

La gran utilización del índice de concordancia Kappa se debe probablemente tanto a la facilidad de cálculo, como a su clara interpretación; no obstante, tiene sus problemas y limitaciones.

El principal problema de esta medida de concordancia radica en que está pensada para clasificaciones nominales, en las que no existe un orden de graduación entre las diferentes categorías. El índice Kappa hasta ahora descrito únicamente tiene en consideración si hay o no acuerdo, esto es si se “clasifica o no al sujeto en la misma categoría.

Si deseamos tener en cuenta el hecho de que estamos manejando variables ordinales para calcular una medida de concordancia, existen diferentes posibilidades. La más sencilla es calcular individualmente la concordancia en cada categoría, pero de esta



forma seguimos sin ponderar el nivel de desacuerdo global según esa clasificación ordinal.

Otro enfoque más global consiste en asignar un “peso” a las diferentes posibilidades de desacuerdo, de tal manera que se considere como más importante un desacuerdo entre categorías alejadas que entre las próximas. Este peso variará entre 0 (acuerdo, misma categoría) y 1 (desacuerdo con categorías extremas).

El problema surge a la hora de determinar esos “pesos”, ya que el valor de concordancia obtenido será diferente según los pesos utilizados. Existen los pesos más utilizados (lineales o bicuadrados) y que suelen proporcionar por defecto los programas de computación y que son detallados a continuación.

Índice Kappa con “pesos”

Una solución que puede verse como intermedia entre las (un único kappa global o K kappas individuales para cada categoría), pero que, en general, sólo tiene sentido para variables ordinales, es el denominado Kappa ponderado, también propuesto por Cohén, en el cual se asignan unos pesos para cuantificar la importancia relativa entre los desacuerdos.

La idea de este índice ponderado es asignar a cada celda de la tabla un peso w_{ij} comprendido entre 0 y 1 que represente la importancia del desacuerdo. Dando el máximo peso al acuerdo perfecto, y pesos proporcionalmente menores según la importancia del desacuerdo

$$w_{ii} = 1 \quad 0 \leq w_{ij} < 1 \quad \forall i \neq j$$

Además, obviamente $w_{ij} = w_{ji}$.

Las proporciones ponderadas de acuerdos observados y esperados se definen:

$$P_{o(w)} = \frac{\sum_{i=1}^K \sum_{j=1}^K w_{ij} X_{ij}}{N} \quad P_{e(w)} = \frac{\sum_{i=1}^K \sum_{j=1}^K w_{ij} X_i \cdot X_j}{N^2}$$

y a partir de aquí, el Kappa ponderado

$$\kappa_w = \frac{P_{o(w)} - P_{e(w)}}{1 - P_{e(w)}}$$



Nótese que, en el caso extremo, si se definen los pesos como:

$$w_{ij} = 0 \quad \forall i \neq j; \quad w_{ii} = 1$$

el Kappa ponderado coincide con el Kappa global.

La principal ventaja del Kappa ponderado reside en la posibilidad de cuantificar diferentes grados de desacuerdo. Los valores de los pesos dependerán, en cada caso, de la importancia que se conceda a cada desacuerdo, hay que tener presente, sin embargo, que ello añade cierta dificultad a su interpretación: si en dos estudios diferentes, se calcula el Kappa ponderado con dos sistemas de pesos distintos, es difícil realizar comparaciones entre ellos. Los más usados en este sentido son, por su sencillez, los denominados pesos lineales, propuestos inicialmente por Cohén

$$w_{ij} = 1 - \frac{|i - j|}{k - 1}$$

y los denominados pesos bicuadrados, propuestos posteriormente por Fleiss y Cohen

$$w_{ij} = 1 - \frac{(i - j)^2}{(k - 1)^2}$$

Pesos lineales y bicuadrados para 3 categorías:

Categoría	Lineal (wl)			Bicuadrado (wb)		
	1	2	3	1	2	3
1	1	0,5	0	1	0,75	0
2	0,5	1	0,5	0,75	1	0,75
3	0	0,5	1	0	0,75	1

Bibliografía:

Agresti Alan. 1990. Categórica! Data Analysis. John Wiley and Sons New York.

Cohén J. Swerdlik M. 2000 Pruebas y Evaluación Psicológicas Mc Graw Hill

México.

Cohén J. 1968 Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or parcial credit Psychol Bull.

