

# Evaluación de la concordancia inter-observador en investigación pediátrica: Coeficiente de Kappa

JAIME CERDA L.<sup>1</sup>, LUIS VILLARROEL DEL P.<sup>2</sup>

1. Pediatra, Programa de Especialidad en Salud Pública. Departamento de Salud Pública, Facultad de Medicina, Pontificia Universidad Católica de Chile.
2. Doctor en Estadística. Departamento de Salud Pública, Facultad de Medicina, Pontificia Universidad Católica de Chile.

## ABSTRACT

### Evaluation of the interobserver concordance in pediatric research: the Kappa Coefficient

Numerous pediatric research protocols are designed for assessing the degree of concordance between two observers; in other words, to determine the extent of their agreement. In order to determine the interobserver concordance, a frequently used statistical tool is available: the Kappa Coefficient ( $\kappa$ ). The present article explains the theoretical background of this coefficient, the methodology employed for its calculation and the way in which its value is correctly interpreted. In simple terms, Kappa Coefficient ( $\kappa$ ) corresponds to the proportion of concordances observed among the total number of observations, having excluded all random concordances. Kappa Coefficient ( $\kappa$ ) adopts a value between -1 y +1, being the strongest degree of interobserver concordance equal to +1. On the contrary, a value of  $\kappa = 0$  reflects that the observed concordance is precisely the one that is expected by chance. The interpretation of Kappa Coefficient ( $\kappa$ ) is performed by correlating its value with a qualitative scale, which includes six level of strength of agreement ("poor", "slight", "fair", "moderate", "substantial" and "almost perfect"), simplifying its comprehension. (**Key words:** Kappa Coefficient, statistics, research, pediatrics).

Rev Chil Pediatr 2008; 79 (1): 54-58

## RESUMEN

Numerosos protocolos de investigación en pediatría son diseñados para conocer el grado de concordancia entre dos observadores, es decir, hasta qué punto ambos coinciden en su medición. Para determinar el grado de concordancia inter-observador se dispone de una herramienta estadística frecuentemente utilizada, el coeficiente kappa ( $\kappa$ ). El presente artículo explica el fundamento teórico de este coeficiente, su metodología de cálculo y la forma en que se interpreta correctamente su valor, ejemplificando estos conceptos mediante una investigación real. En términos simples, el coeficiente kappa ( $\kappa$ ) corresponde a la proporción de concordancias observadas sobre el total de observaciones, habiendo excluido las concor-

Trabajo recibido el 23 de noviembre de 2007, devuelto para corregir el 07 de enero de 2008, segunda versión el 11 de enero de 2008, aceptado para publicación el 17 de enero de 2008.

Correspondencia a:  
Jaime Cerda Lorca  
E-mail: jcerda@med.puc.cl

dancias atribuibles al azar. El coeficiente kappa ( $\kappa$ ) toma valores entre -1 y +1; mientras más cercano a +1, mayor es el grado de concordancia inter-observador. Por el contrario, un valor de  $\kappa = 0$  refleja que la concordancia observada es precisamente la que se espera a causa exclusivamente del azar. La interpretación del coeficiente kappa se realiza correlacionando su valor con una escala cualitativa que incluye seis niveles de fuerza de concordancia (“pobre”, “leve”, “aceptable”, “moderada”, “considerable” y “casi perfecta”), simplificando la comprensión del mismo.

(**Palabras clave:** coeficiente kappa, estadística, investigación, pediatría).

Rev Chil Pediatr 2008; 79 (1): 54-58

## Introducción

Independientemente del diseño de investigación, la validez de un estudio puede verse severamente afectada si se utilizan mediciones poco fiables. Una importante fuente de error de medición es producto de la variabilidad inter-observador, cuya magnitud es posible de estimar a través de los llamados estudios de concordancia, los cuales tienen como objetivo estimar hasta qué punto dos observadores coinciden en su medición. Estadísticamente, la manera de abordar este problema depende de la naturaleza de los datos. Cuando éstos son de tipo categórico, el test más frecuentemente empleado es el test de kappa<sup>1</sup>, cuyo coeficiente homónimo refleja la fuerza de la concordancia entre dos observadores. Mediante el siguiente ejemplo, tomado de una investigación real, se explicará en términos simples el fundamento teórico de este test, la metodología de cálculo del coeficiente kappa y su forma correcta de interpretación.

### Ejemplo

Schonhaut y colaboradores<sup>2</sup> evaluaron la concordancia interobservador del diagnóstico nutricional formulado por profesores de colegios fiscales y por el equipo de salud de un consultorio, ambos ubicados en la comuna de Colina (región metropolitana). Para ello, un grupo de profesores pesó y midió a 413 alumnos de primer año básico, calculó el índice de masa corporal (IMC) de cada alumno y formuló un diagnóstico nutricional, basándose en la referencia del CDC-NCHS. Por otro lado, el equipo de salud del consultorio local repitió el procedimiento en los mismos niños, de modo que cada uno de los menores contó con dos diag-

nósticos nutricionales: uno formulado por el grupo de profesores y otro por el equipo de salud del consultorio. La tabla 1 resume los diagnósticos formulados por ambos grupos. El número al interior de cada celda corresponde al total de niños que recibieron el diagnóstico señalado en su correspondiente columna por parte de los profesores y el diagnóstico señalado en su correspondiente cada fila por parte del equipo de salud del consultorio. Por ejemplo, el total de niños que recibieron el diagnóstico de “eutrofia” por parte de los profesores y de “bajo peso” por parte del equipo de salud corresponde a 12 niños; por su parte, el total de niños que recibieron el diagnóstico de “sobrepeso” por parte de los profesores y de “eutrofia” por parte del equipo de salud corresponde a 37 niños.

Las celdas a, b, c y d (i.e. diagonal principal) corresponden a las concordancias en el diagnóstico nutricional formulado por los profesores y por el equipo de salud, por el contrario, el resto de las celdas (i.e. todas aquellas fuera de la diagonal principal) corresponden a las discordancias en el diagnóstico nutricional formulado por ambos evaluadores. A partir de estos datos, es posible calcular un primer coeficiente de concordancia (al cual se denominará arbitrariamente alfa ( $\alpha$ )), correspondiente al cociente entre la sumatoria de las observaciones concordantes y el total de observaciones. En el ejemplo,  $\alpha = (8 + 212 + 41 + 49)/413 = 0,75$ . El coeficiente  $\alpha$  tiene la desventaja de sobreestimar el grado de concordancia entre profesores y el equipo de salud debido a que no contempla el hecho de que una fracción de la concordancia es atribuible al azar. Para superar esta desventaja, Cohen<sup>3</sup> introdujo en 1960 el coeficiente kappa, el cual incorpora en su fórmula una corrección que excluye la concordancia

atribuible al azar, permitiendo una estimación más precisa de la concordancia genuina.

### Fundamento teórico, cálculo e interpretación del coeficiente Kappa ( $\kappa$ )

El coeficiente kappa refleja la concordancia inter-observador y puede ser calculado en tablas de cualquier dimensión, siempre y cuando se contrasten dos observadores (para la evaluación de concordancia de tres o más observadores se utiliza el coeficiente kappa de Fleiss, cuya explicación excede el propósito del presente artículo). El coeficiente kappa puede tomar valores entre -1 y +1. Mientras más cercano a +1, mayor es el grado de concordancia inter-observador, por el contrario, mientras más cercano a -1, mayor es el grado de discordancia inter-observador. Un valor de  $\kappa = 0$  refleja que la concordancia observada es precisamente la que se espera a causa exclusivamente del azar<sup>1</sup>.

Un requisito *sine qua non* para calcular el coeficiente kappa es definir para cada celda de la diagonal principal (i.e. celdas concordantes) el número teórico de observaciones atribuibles

al azar, las cuales serán denominadas en el ejemplo a', b', c' y d'. El valor de a' corresponde al producto entre el número total de niños con diagnóstico nutricional de "bajo peso" según el equipo de salud (n = 24) y el número total de niños con diagnóstico nutricional de "bajo peso" según los profesores (n = 18), dividido por el total de niños participantes en el estudio (n = 413). Este procedimiento se repite para todas las celdas concordantes. Posteriormente, se suman estos valores teóricos, siendo su resultado equivalente al total de concordancias atribuibles al azar. En el ejemplo,  $(a' + b' + c' + d') = (1,05 + 154,44 + 13,00 + 10,02) = 178,51$ . La tabla 2 resume los valores teóricos calculados para el ejemplo y su metodología de cálculo.

El coeficiente kappa se construye en base a un cociente, el cual incluye en su numerador la diferencia entre la sumatoria de las concordancias observadas y la sumatoria de las concordancias atribuibles al azar, mientras que su denominador incluye la diferencia entre el total de observaciones y la sumatoria de las concordancias atribuibles al azar. En términos conceptuales, la fórmula del coeficiente kappa puede expresarse de la siguiente forma:

**Tabla 1. Diagnóstico nutricional de 413 niños, según formulación por parte de los profesores vs Equipo de salud<sup>2</sup>**

		Bajo peso	Eutrofia	Profesores Sobrepeso	Obesidad	Total
Equipo de Salud	Bajo peso	8 (a)	12	2	2	24
	Eutrofia	10	212 (b)	37	9	268
	Sobrepeso	0	11	41 (c)	9	61
	Obesidad	0	3	8	49 (d)	60
	Total	18	238	88	69	413

**Tabla 2. Valores teóricos para cada celda de la diagonal principal**

		Bajo peso	Eutrofia	Profesores Sobrepeso	Obesidad	Total
Equipo de Salud	Bajo peso	1,05 (a')	-	-	-	24
	Eutrofia	-	154,4 (b')	-	-	268
	Sobrepeso	-	-	13,00 (c')	-	61
	Obesidad	-	-	-	10,02 (d')	60
	Total	18	238	88	69	413

(a') =  $(24 \times 18 / 413)$ ; (b') =  $(268 \times 238 / 413)$ ; (c') =  $(61 \times 88 / 413)$  y (d') =  $(60 \times 69 / 413)$

$$\kappa = \frac{[(\Sigma \text{ concordancias observadas}) - (\Sigma \text{ concordancias atribuibles al azar})]}{[(\text{total de observaciones}) - (\Sigma \text{ concordancias atribuibles al azar})]}$$

En el ejemplo,  $\kappa = (310-178,51)/(413-178,51) = (131,49/234,49) = 0,56$ . Nótese que el valor de este coeficiente es menor al valor de coeficiente alfa calculado anteriormente ( $\alpha = 0,75$ ). La explicación radica en el hecho de que el coeficiente kappa excluye la concordancia atribuible al azar, a diferencia del coeficiente alfa (el cual sobreestima la magnitud de la concordancia).

La tabla 3 corresponde a la escala utilizada con frecuencia para expresar cualitativamente la fuerza de la concordancia, según Landis y Koch<sup>4</sup>. Se incluyó entre paréntesis la expresión original en inglés.

Retornando al ejemplo, un valor de  $\kappa = 0,56$  permite afirmar que el diagnóstico nutricional formulado por profesores y por el equipo de salud tiene una concordancia de fuerza moderada.

## Discusión

La realización de estudios de concordancia en investigación pediátrica se encuentra ampliamente difundida. Este tipo de diseño de investigación presenta diversas ventajas, destacando su simpleza logística, sencillez de análisis estadístico y una amplia aplicabilidad en escenarios clínicos. Al respecto, dos estudios publicados en la Revista Chilena de Pediatría reflejan fielmente las ventajas anteriormente señaladas.

Urzúa<sup>5</sup> y colaboradores evaluaron el grado de concordancia en la valoración clínica de niños con patologías respiratorias agudas entre educandos de pre y postítulo de pediatría (70 internos y 30 residentes), quienes evaluaron y asignaron el Puntaje de Tal Modificado (PTM) a 167 y 178 niños, respectivamente. Los resultados del estudio demostraron que la concordancia inter-observador en la asignación del PTM fue moderada entre los internos ( $\kappa = 0,52$ ) y solamente aceptable entre los residentes ( $\kappa = 0,23$ ). La investigación resaltó la im-

**Tabla 3. Valoración del coeficiente kappa (Landis y Koch, 1977)<sup>4</sup>**

Coefficiente kappa	Fuerza de la concordancia
0,00	Pobre ( <i>Poor</i> )
0,01 - 0,20	Leve ( <i>Slight</i> )
0,21 - 0,40	Aceptable ( <i>Fair</i> )
0,41 - 0,60	Moderada ( <i>Moderate</i> )
0,61 - 0,80	Considerable ( <i>Substantial</i> )
0,81 - 1,00	Casi perfecta ( <i>Almost perfect</i> )

portancia de revisar los objetivos y métodos de los programas docentes, con el fin de alcanzar entre los educandos un mayor nivel de acuerdo en la valoración clínica de los enfermos. Si se considera que la falta de concordancia en la valoración clínica de pacientes generalmente conduce a planteamientos diagnósticos y terapéuticos inadecuados, este tipo de investigación reviste una alta importancia clínica. Por otro lado, Pinheiro<sup>6</sup> y colaboradores evaluaron la concordancia en el diagnóstico nutricional formulado a partir de tres estándares antropométricos internacionales aplicados a 116 niños con Síndrome de Down, demostrando que su concordancia diagnóstica según la relación peso/edad fluctúa entre leve y aceptable ( $\kappa = 0,13-0,31$ ). La demostración de falta de concordancia en el diagnóstico nutricional formulado a partir de estos estándares motivó a los investigadores a sugerir el uso de solamente uno de ellos en niños con Síndrome de Down (i.e. el más adecuado para discriminar déficits y excesos), con el fin último de mejorar el planteamiento de diagnósticos nutricionales y así poder establecer las estrategias pertinentes.

Desde un punto de vista estadístico y metodológico, es importante aclarar dos conceptos. En primer lugar, el valor del coeficiente kappa no proporciona indicación alguna de la precisión de dicha estimación, siendo esencial conocer su variabilidad, con el fin último de plantear tests de hipótesis y construir intervalos de confianza para kappa. Las hipótesis nula ( $H_0$ ) y alternativa ( $H_1$ ) de estos tests son  $H_0: \kappa = 0$  y  $H_1: \kappa \neq 0$ . Su resultado indica si el valor del coeficiente kappa es o no estadísticamente significativo, con un determinado nivel de confianza, permitiendo la realización de inferencia estadística<sup>1</sup>. En segundo lugar, el coeficiente kappa

no aporta información alguna sobre la cualidad de la medición realizada por los observadores, pues está diseñado únicamente para estimar la magnitud de la concordancia entre ambos. Por ejemplo, dos observadores pueden presentar una concordancia diagnóstica “casi perfecta” ( $\kappa \neq 0,80-1,00$ ), sin embargo, ambos pueden estar “concordantemente equivocados” respecto al diagnóstico real.

Existen numerosos programas computacionales capaces de calcular el coeficiente kappa, sin embargo, su estimación es útil solamente en manos de un investigador familiarizado con su significado estadístico e interpretación. En consecuencia, es altamente recomendable construir tablas de contingencia en base a los datos reportados en artículos de investigación y calcular manualmente el valor del coeficiente kappa. Esta práctica seguramente aportará al lector una mayor comprensión de los resultados de la investigación.

## Referencias

- 1.- *López de Ullibarri I, Pita S*: Medidas de concordancia: el coeficiente kappa. *Cad aten primaria* 1999; 6: 169-71. Disponible en [www.fisterra.com](http://www.fisterra.com) [consultado el 01/10/07].
- 2.- *Schonhaut L, Rodríguez L, Pizarro T, Kohn J, Merino D, López A, et al*: Concordancia en el diagnóstico nutricional según coeficiente de masa corporal, entre el equipo de salud y las escuelas de la comuna de colina. *Rev Chil Pediatr* 2004; 75: 32-5.
- 3.- *Cohen J*: A coefficient of agreement for nominal scales. *Educational and psychological measurement* 1960; 20: 37-46.
- 4.- *Landis J, Koch G*: The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.
- 5.- *Urzúa S, Duffau G, Zepeda G, Sagredo S*: Estudio de concordancia clínica en educandos de pre y postítulo en pediatría: puntaje de tal. *Rev Chil Pediatr* 2002; 73: 471-7.
- 6.- *Pinheiro A, Urteaga C, Canete G, Atalah E*: Evaluación del estado nutricional en niños con síndrome de Down según diferentes referencias antropométricas. *Rev Chil Pediatr* 2003; 74: 585-9.