

EVIDENCIAS EN PEDIATRÍA

Toma de decisiones clínicas basadas en las mejores pruebas científicas
www.evidenciasenpediatria.es

Fundamentos de medicina basada en la evidencia

Evaluación de la precisión de las pruebas diagnósticas (2). Variables continuas

Ochoa Sangrador C¹, Molina Arias M²

¹Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.

²Servicio de Gastroenterología. Hospital Infantil Universitario La Paz. Madrid. España.

Correspondencia: Carlos Ochoa Sangrador, cochoas2@gmail.com

Palabras clave en inglés: accuracy, repeatability, diagnostic test, reliability of diagnostic tests, intraclass correlation coefficient, Bland and Altman plot.

Palabras clave en español: precisión, reproducibilidad, pruebas diagnósticas, fiabilidad de las pruebas diagnósticas, coeficiente de correlación intraclase, método de Bland y Altman.

Fecha de recepción: 1 de septiembre de 2017 • **Fecha de aceptación:** 4 de septiembre de 2017

Fecha de publicación del artículo: 6 de septiembre de 2017

Evid Pediatr. 2017;13:45.

CÓMO CITAR ESTE ARTÍCULO

Ochoa Sangrador C, Molina Arias M. Evaluación de la precisión de las pruebas diagnósticas (2). Variables continuas. Evid Pediatr. 2017;13:45.

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín de novedades en <http://www.evidenciasenpediatria.es>

Este artículo está disponible en: <http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2017;13:45>.

©2005-17 • ISSN: 1885-7388

Evaluación de la precisión de las pruebas diagnósticas (2). Variables continuas

Ochoa Sangrador C¹, Molina Arias M²

¹Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.

²Servicio de Gastroenterología. Hospital Infantil Universitario La Paz. Madrid. España.

Correspondencia: Carlos Ochoa Sangrador, cochoas2@gmail.com

INTRODUCCIÓN

En documentos previos de esta serie hemos abordado cómo evaluar la validez de una prueba diagnóstica. También hemos revisado cómo medir su precisión o fiabilidad. Hasta ahora hemos mostrado los métodos de medición de la precisión para datos discretos nominales (índice kappa) y ordinales (índice kappa ponderado). Ahora abordaremos los métodos correspondientes a datos continuos: la desviación estándar intrasujetos, el coeficiente de correlación intraclase y el método de Bland y Altman.

VARIABLES CONTINUAS

Desviación estándar intrasujetos

Cuando el resultado de una prueba se mide en una escala continua, podemos estimar el error de medición calculando

la variabilidad existente entre medidas repetidas en los mismos sujetos. El parámetro que mejor refleja dicha variabilidad es la desviación estándar intrasujetos (excluyendo la observada entre sujetos). Para calcularlo necesitamos una serie de sujetos a los que se les realicen al menos dos mediciones. En la tabla 1 se presentan los resultados de dos mediciones repetidas de bilirrubina transcutánea en recién nacidos ictericos¹. La desviación estándar intrasujetos puede calcularse fácilmente usando un programa que realice análisis de la varianza (ANOVA). El ANOVA descompone la variación que hay entre el conjunto de mediciones (estimada a través de la diferencia de cada valor respecto la media global al cuadrado) en varios componentes: la variación entre las mediciones de los diferentes sujetos (entre las filas de la tabla 1) y la variación residual, que en una ANOVA de un factor corresponde a la variación entre las mediciones de cada sujeto (entre las columnas de la tabla 1).

En la tabla 2 podemos ver el ANOVA para los datos de la tabla 1. El parámetro denominado cuadrados medios de los

TABLA 1. Resultados de dos mediciones repetidas de bilirrubina transcutánea (Jaundice-Meter 101, Minolta Air Shields), en la cara anterior del tórax en 20 recién nacidos ictericos. Datos extraídos de un estudio más amplio³

Sujetos	1.ª medición	2.ª medición	Diferencia	Media
1	14	16	-2	15,0
2	14	14	0	14,0
3	17	17	0	17,0
4	14	15	-1	14,5
5	15	14	1	14,5
6	18	19	-1	18,5
7	16	16	0	16,0
8	12	12	0	12,0
9	19	19	0	19,0
10	9	10	-1	9,5
11	15	16	-1	15,5
12	18	18	0	18,0
13	17	18	-1	17,5
14	15	15	0	15,0
15	9	9	0	9,0
16	14	14	0	14,0
17	17	18	-1	17,5
18	18	18	0	18,0
19	20	20	0	20,0
20	10	11	-1	10,5

TABLA 2. Análisis de la varianza de una vía de los datos de la tabla 1

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	
Pacientes	19	371,5000	19,5526	CMp
Residual	20	6,0000	0,3000	CMr
Total	39	377,5000		

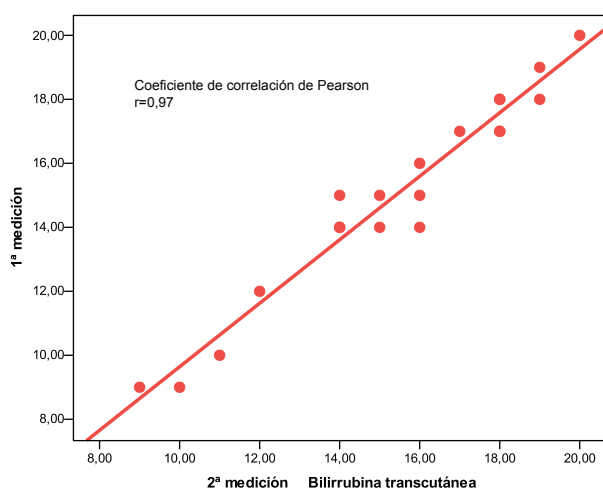
CMp: cuadrados medios de los pacientes; **CMr:** cuadrados medios de los residuos.

residuos (CMr) es la varianza residual o intrasujetos (que depende de las diferencias entre las mediciones repetidas de cada sujeto). Si realizamos la raíz cuadrada de CMr obtendremos la desviación estándar intrasujetos (s_i). La s_i puede calcularse igualmente a partir del ANOVA para estudios con más de dos mediciones por sujeto.

Utilizando la s_i podemos cuantificar el margen de error de nuestras mediciones. Así, podemos estimar que la diferencia entre una medición determinada y el verdadero valor no será mayor de 1,96 veces la s_i en el 95% de las observaciones (asumiendo que siguen una distribución normal, el 95% de las determinaciones caerán en el intervalo entre el valor verdadero más menos 1,96 veces la desviación estándar). También nos permite estimar que la diferencia entre dos mediciones repetidas en un mismo sujeto no superará 2,77 veces la s_i en el 95% de las observaciones^{2,3}. Para nuestro ejemplo, la s_i es 0,54 (raíz cuadrada de 0,3), por lo que la diferencia estimada respecto al valor verdadero será menor de 1,05 ($1,96 \times 0,54$) y la diferencia entre dos mediciones será menor de 1,49 ($2,77 \times 0,54$).

Coeficiente de correlación intraclase

Si solo se realizan dos mediciones por sujeto, la forma más intuitiva de compararlas es representar las parejas de mediciones en un diagrama de puntos, examinando si existe relación lineal entre ambas y calcular su coeficiente de correlación. En la figura 1 se presenta el diagrama de puntos de los

FIGURA 1. Diagrama de puntos y correlación lineal de los datos de la tabla 1

datos de la tabla 1. El coeficiente de correlación de Pearson (r) para estos datos es 0,97 (cuanto más se aproxima a 1, mejor es la correlación).

Sin embargo, la existencia de una fuerte relación lineal con un alto coeficiente de correlación no indica que haya una buena concordancia entre las mediciones, solamente que los puntos del diagrama se ajustan a una recta. El coeficiente de correlación depende en gran manera de la variabilidad entre sujetos, por ello, varía mucho en función de las características de la muestra donde se estima, afectándole especialmente la presencia de valores extremos. Si una de las mediciones es sistemáticamente mayor que otra, el coeficiente de correlación será muy alto, a pesar de que las mediciones nunca concuerden. Estos problemas son evitados utilizando el coeficiente de correlación intraclase.

El coeficiente de correlación intraclase (CCI) estima la concordancia entre dos o más medidas repetidas. El cálculo del CCI se basa en un modelo de ANOVA con medidas repetidas, aplicándose distintas fórmulas en función del diseño y los objetivos del estudio⁴. El escenario más simple es aquel en el que estimamos la variabilidad de las medidas, sin tener en cuenta la variabilidad aportada por los distintos observadores (diseño de una vía con factor aleatorio). Considerando este diseño, y utilizando los resultados del ANOVA, podemos calcular el CCI con la siguiente fórmula:

$$CCI = \frac{CMp - CMr}{CMp + (k - 1)CMr}$$

donde k es el número de observaciones por sujeto, CMp son los cuadrados medios entre pacientes (que depende de las diferencias de las mediciones entre sujetos) y CMr los cuadrados medios de los residuos (que depende de las diferencias entre las mediciones repetidas de cada sujeto).

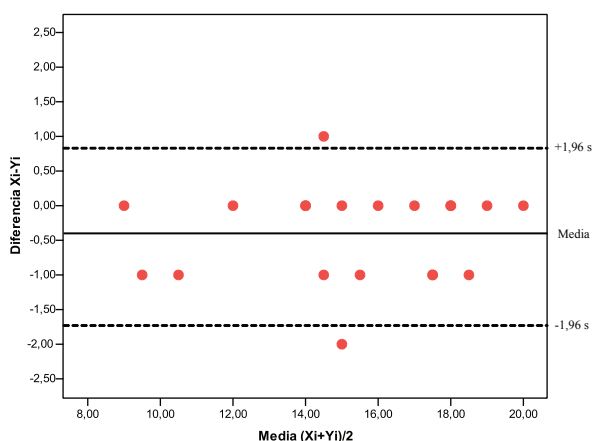
Con los datos del ANOVA de la tabla 2 el CCI será:

$$CCI = \frac{19,55 - 0,30}{19,55 + (2 - 1)0,30} = 0,96$$

En nuestro ejemplo, apenas hay diferencias entre el CCI y el coeficiente de correlación de Pearson (r). Si el CCI fuera mucho menor que r , habría que pensar que existe un cambio sistemático entre una medida y otra, lo que podría estar causado por un efecto de aprendizaje. En este caso, las mediciones no se habrían realizado en las mismas circunstancias, por lo que no se darían las condiciones para realizar un estudio de fiabilidad⁵.

Método de Bland y Altman

Un método alternativo para analizar la concordancia entre dos observaciones repetidas que se miden en una escala continua es el método gráfico descrito por Bland y Altman⁶. Consiste en representar en un diagrama de puntos la diferencia entre los pares de mediciones contra su media (figura 2). Los puntos tienden a agruparse alrededor del cero para las

FIGURA 2. Método de Bland y Altman con los datos de la tabla 1

diferencias entre las dos mediciones, de forma que cuanto mayor sea el grado de dispersión alrededor del cero, peor será el acuerdo entre los dos métodos. Una forma de valorarlo sería dibujar las líneas horizontales en el nivel de diferencia máxima que puede ser tolerable desde el punto de vista clínico y ver si los puntos, o la mayoría de ellos, se agrupan entre estas dos líneas horizontales. De forma alternativa, se puede estimar la desviación estándar de las diferencias y los intervalos entre los que cabe esperar que se encuentre el 95% de ellas.

Este método también permite examinar la magnitud de las diferencias y su relación con la magnitud de la medición.

Cuando la variabilidad en las medidas no es constante, sino que cambia al aumentar o disminuir la magnitud de la medida, el cálculo se complica⁷. Si existe correlación significativa entre las diferencias y las medias, la variabilidad no será constante (puede haber un acuerdo aceptable en determinado intervalo de valores, pero no en otros). En ese caso, puede intentarse realizar transformaciones logarítmicas de los datos o analizar la variabilidad por separado para varios intervalos de valores, aunque siempre tendremos que cuestionarnos la validez de la medición en ese intervalo.

BIBLIOGRAFÍA

1. Ochoa Sangrador C, Marugán Isabel VM, Tesoro González R, García Rivera MT, Hernández Calvo MT. Evaluación de un instrumento de medición de la bilirrubina transcutánea. *An Esp Pediatr.* 2000;52:561-8.
2. Bland JM, Altman DG. Measurement error. *BMJ.* 1996;312:1654.
3. Altman DG, Bland JM. Comparing several groups using analysis of variance. *BMJ.* 1996;312:1472.
4. Fleiss JL. *The design and analysis of clinical experiments.* Nueva York: John Wiley & Sons 1986. p. 1-32.
5. Bland JM, Altman DG. Measurement error and correlation coefficients. *BMJ.* 1996;313:41-2.
6. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307-10.
7. Bland JM, Altman DG. Measurement error proportional to the mean. *BMJ.* 1996;313:106.