

## Conceptos prácticos en MBE y epidemiología clínica

Medwave 2014;14(1):5892 doi: 10.5867/medwave.2014.01.5892

# Estadística para aterrorizados: interpretando intervalos de confianza y valores p

Statistics for the faint of heart – how to interpret confidence intervals and p values

**Autores:** Eva Madrid Aris<sup>(1,2)</sup>, Felipe Martínez Lomakin<sup>(1,3)</sup>

**Filiación:**

<sup>(1)</sup>Centro de Investigaciones Biomédicas, Escuela de Medicina, Universidad de Valparaíso, Chile

<sup>(2)</sup>Departamento de Salud Pública, Escuela de Medicina, Universidad de Valparaíso, Chile

<sup>(3)</sup>MSc Programme in Evidence-Based Healthcare, University of Oxford, Reino Unido

**E-mail:** [eva.madrid@uv.cl](mailto:eva.madrid@uv.cl)

**Citación:** Madrid E, Martínez F. Statistics for the faint of heart – how to interpret confidence intervals and p values. *Medwave* 2014;14(1):5892 doi: 10.5867/medwave.2014.01.5892

**Fecha de envío:** 7/1/2014

**Fecha de aceptación:** 22/1/2014

**Fecha de publicación:** 28/1/2014

**Origen:** solicitado

**Tipo de revisión:** con revisión por cuatro pares revisores externos, a doble ciego

### Introducción

Según el diccionario de la Real Academia Española, el término “significativo” denota simplemente la presencia de algo “que tiene importancia por representar o significar algo”. Sin embargo, en estadística, la significancia representa un concepto distinto: es una afirmación específica respecto a qué tan probable es que algo se deba al azar<sup>1,2</sup>. El confundir ambos conceptos es sencillo y frecuentemente visto en artículos de investigación biomédica<sup>3</sup>. En este artículo, expondremos brevemente las principales herramientas utilizadas para hacer inferencia estadística, las estimaciones de valores p (*p-values*) e intervalos de confianza.

### El valor p

Tradicionalmente los valores p (de probabilidad) han sido utilizados para evaluar si los resultados son explicables por el azar. Para realizar esta estimación, conceptualmente se realiza una prueba de hipótesis, en la que se decide aceptar una de dos hipótesis mutuamente excluyentes. Conocemos como **hipótesis nula (H0)** a aquella que representa lo tradicionalmente aceptado hasta el momento en que se realiza la prueba de hipótesis, habitualmente que no existe asociación entre las variables o que ésta es explicable por el azar. La **hipótesis alterna (H1)** en cambio, es aquella que plantea que ambas variables están asociadas.

La definición de valor p es la probabilidad de observar un estadístico tan extremo como el de la muestra, **asumiendo que la hipótesis nula es cierta**<sup>4</sup>. Es decir, coloquialmente, cuál es la probabilidad de observar lo que se dio en el estudio asumiendo que ambas variables no

están asociadas. Por consenso, hemos fijado un valor umbral de probabilidad del 5% que es conocido como **nivel de significación** o  **$\alpha$**  para decidir cuál hipótesis es la cierta. Si la probabilidad es baja (<5%) significaría que lo observado es muy infrecuente, por lo que se opta por rechazar la hipótesis nula y asumir que las variables están asociadas. Esto es lo que se conoce como algo *estadísticamente significativo*, que nada tiene que ver con que los hallazgos sean *clínicamente relevantes*.

Por ejemplo, si un ensayo detectara que un tratamiento disminuye la incidencia de una complicación en un 2% con un valor p de 0,01, quiere decir que, por azar, esta disminución en complicación ocurre en un 1% de los casos. Dado que es muy infrecuente, damos el cariz de *significativo* a lo evaluado y asumimos que la intervención estuvo asociada a la diferencia. No obstante, el que una reducción del 2% sea clínicamente relevante, dependerá, entre otras cosas, de lo que estemos midiendo. Reducir la mortalidad en 2% será probablemente interesante en cualquier escenario, pero una disminución en la incidencia una reacción adversa menor, como cefalea, puede no justificar grandes costos o reacciones adversas derivadas del nuevo tratamiento.

Otra consideración que debe tenerse con los valores p está en el método que se emplea para interpretarlos. El valor corte de 5% implica que en 5 de cada 100 casos se podría interpretar que hay asociación entre dos variables siendo que en realidad no existe tal fenómeno, puesto que sigue estando dentro de las posibilidades del azar. Como es casi imposible trabajar con una población completa, nunca se sabrá si la hipótesis nula es cierta con certeza

absoluta. Por este motivo siempre es necesario cotejar los resultados encontrados con otras experiencias científicas para obtener una imagen global lo más fidedigna posible respecto a los efectos de una exposición (que bien puede ser una intervención) en un paciente.

### Intervalos de confianza

Sin importar que tan bien diseñado esté un estudio, el hecho de realizar un muestreo hace susceptible al ensayo a ciertos grados de imprecisión y azar<sup>1</sup>. Lo que somos capaces de apreciar en una muestra o al comparar dos grupos puede no necesariamente representar al universo o población de donde la muestra proviene<sup>5,6</sup>.

Los intervalos de confianza ofrecen una manera de estimar, con alta probabilidad, un rango de valores en el que se encuentra el valor poblacional (o *parámetro*) de una determinada variable<sup>7,8</sup>. Esta probabilidad ha sido fijada por consenso en un 95% en base a supuestos de normalidad, pero rangos entre el 90% y 99% son comúnmente utilizados en la literatura científica. En términos simples, un intervalo de confianza del 95% indica que el valor poblacional se encuentra en un determinado rango de valores con un 95% de certeza. Como regla general, mientras mayor es el tamaño de la muestra, menor es la variabilidad para hacer la estimación del intervalo, lo que lleva a estimadores más precisos<sup>9</sup>. Por el contrario, mientras más certeza se desee respecto a la extrapolación poblacional (por ejemplo, 99%), más amplio será el intervalo.

Estas herramientas también permiten hacer inferencia estadística al excluir un valor crítico que indique la falta de asociación del intervalo. Una interpretación coloquial de esto último es decir que los hallazgos son estadísticamente significativos porque el intervalo "no pasa por el 1". Efectivamente, de tratarse de un indicador cuya fórmula es un cociente, como el riesgo relativo (RR) o el *Odds Ratio* (OR), un valor 1 indica que la frecuencia de un determinado evento fue igualmente presentado tanto en el grupo expuesto como en el que no, por lo que sería el "valor crítico" que debe excluirse para denotar significancia estadística<sup>9-11</sup>. Cuando lo expresado es una resta en el riesgo entre dos grupos (como con la reducción del riesgo absoluto), un valor de 0 pasa a ser el límite para denotar hallazgos significativos, puesto que representaría el punto en que el evento es igualmente probable en ambos grupos.

Supongamos que existe un estudio que detectó que la fibrilación auricular en el postoperatorio se asociaba a mayor mortalidad entre pacientes operados con un RR de 3 (IC 95%: 2-4). Esto significa que en la **muestra**, la presencia de la arritmia *triplicó* la probabilidad de morir en relación a quienes no la tuvieron. Si bien no podemos trabajar con la población, la estimación del intervalo indica que podemos estar 95% seguros que el valor **poblacional** del riesgo relativo estará entre 2 y 4. Como el valor 1 está excluido del rango del intervalo, podemos decir que los hallazgos son *estadísticamente significativos*.

### ¿Intervalo de confianza o valor p?

Hacer pruebas de hipótesis (valores p) o estimar (intervalos de confianza) son técnicas validadas que contribuyen a la precisión de los efectos clínicos de cualquier investigación original<sup>12,13</sup>. Ambas están estrechamente relacionadas, siendo raro que un intervalo de confianza entregue un resultado significativo y un valor p no. No obstante, la información ofrecida por un intervalo de confianza es más detallada que la contenida en valores p, permitiendo al clínico además estimar si el rango de una diferencia entre tratamientos es lo suficientemente grande como para justificar su uso en una patología relevante<sup>7,12,13</sup>.

### Notas

#### Declaración de conflictos de intereses

Los autores han completado el formulario de declaración de conflictos de intereses del ICMJE traducido al castellano por *Medwave*, y declaran no haber recibido financiamiento para la realización del artículo; no tener relaciones financieras con organizaciones que podrían tener intereses en el artículo publicado, en los últimos tres años; y no tener otras relaciones o actividades que podrían influir sobre el artículo publicado. Los formularios pueden ser solicitados contactando a la autora responsable.

### Referencias

1. Wang EW, Ghogomu N, Voelker CC, Rich JT, Paniello RC, Nussenbaum B, et al. A practical guide for understanding confidence intervals and P values. *Otolaryngol Head Neck Surg*. 2009 Jun;140(6):794-9. | [CrossRef](#) | [PubMed](#) |
2. Fethney J. Statistical and clinical significance, and how to use confidence intervals to help interpret both. *Aust Crit Care*. 2010 May;23(2):93-7. | [CrossRef](#) | [PubMed](#) |
3. McCormack J, Vandermeer B, Allan GM. How confidence intervals become confusion intervals. *BMC Med Res Methodol*. 2013 Oct 31;13:134. | [CrossRef](#) | [PubMed](#) | [PMC](#) |
4. Overholser BR, Sowinski KM. Biostatistics primer: part I. *Nutr Clin Pract*. 2007 Dec;22(6):629-35. | [CrossRef](#) | [PubMed](#) |
5. De Muth JE. Overview of biostatistics used in clinical research. *Am J Health Syst Pharm*. 2009 Jan 1;66(1):70-81. | [CrossRef](#) | [PubMed](#) |
6. Berry EM, Coustère-Yakir C, Grover NB. The significance of non-significance. *QJM*. 1998 Sep;91(9):647-53. | [CrossRef](#) | [PubMed](#) |
7. Candia R, Caiozzi G. Intervalos de confianza. *Rev Med Chil*. 2005 Sep;133(9):1111-5. | [CrossRef](#) | [PubMed](#) |
8. du Prel JB, Hommel G, Röhrig B, Blettner M. Confidence interval or p-value?: part 4 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*. 2009 May;106(19):335-9. | [CrossRef](#) | [PubMed](#) | [PMC](#) |
9. Laing CM, Rankin JA. Odds ratios and confidence intervals: a review for the pediatric oncology clinician. *J Pediatr Oncol Nurs*. 2011 Nov-Dec;28(6):363-7. | [CrossRef](#) | [PubMed](#) |

10. Barratt A, Wyer PC, Hatala R, McGinn T, Dans AL, Keitz S, et al. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. CMAJ. 2004 Aug 17;171(4):353-8. | [CrossRef](#) | [PubMed](#) | [PMC](#) |
11. Siström CL, Garvan CW. Proportions, odds, and risk. Radiology. 2004 Jan;230(1):12-9. | [CrossRef](#) | [PubMed](#) |
12. Montori VM, Kleinbart J, Newman TB, Keitz S, Wyer PC, et al. Tips for learners of evidence-based medicine: 2. Measures of precision (confidence intervals). CMAJ. 2004 Sep 14;171(6):611-5. | [CrossRef](#) | [PubMed](#) | [PMC](#) |
13. Guyatt G, Jaeschke R, Heddle N, Cook D, Shannon H, Walter S. Basic statistics for clinicians: 2. Interpreting study results: confidence intervals. CMAJ. 1995 Jan 15;152(2):169-73. | [PubMed](#) | [PMC](#) | [Link](#) |

**Correspondencia a:**

<sup>(1)</sup>Hontaneda 2664  
Edificio Dr. Bruno Günther Schaffeld  
Valparaíso  
Chile



Esta obra de Medwave está bajo una licencia Creative Commons Atribución-No Comercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, Medwave.