

Procedimientos para la estimación por intervalos de confianza en las investigaciones biomédicas

Procedures for the confidence interval estimation in biomedical research studies

MSc. Yunier Arpajón Peña, Lic. Ana Ludys Sosa Pérez

Facultad de Estomatología "Raúl González Sánchez". La Habana. Cuba.

RESUMEN

En las investigaciones de las distintas especialidades biomédicas se realizan diariamente informaciones clínicas y epidemiológicas, por lo que es necesario tener bien claro qué es lo que se debe presentar y cómo se tiene que hacer. El objetivo de este artículo es describir las ventajas del cálculo de los intervalos de confianza para la correcta interpretación de los resultados en las investigaciones biomédicas. Se realizó una revisión de libros de texto de bioestadística y estadística para salud pública, además de artículos científicos recogidos en bases de datos como Scielo, MedLine y PubMed, todos publicados en el periodo desde 2004-2014. Para la búsqueda electrónica se utilizaron los descriptores: intervalos de confianza, estadística en salud, estimadores puntuales y contrastes de hipótesis. En la mayoría de los documentos revisados se encontró que la mejor forma de expresar la significación estadística de los resultados es mediante los intervalos de confianza para el 95 %. Aquí se presentan los procedimientos que más se utilizan para su cálculo en investigaciones descriptivas, inferenciales, clínicas o epidemiológicas. La estimación mediante intervalos de confianza es un método sólido y sencillo de aplicar, proporciona mayor robustez a cualquier análisis en una investigación porque evalúa el rango de valores donde posiblemente se encuentre el valor real, y por lo tanto, realiza una mejor interpretación y aplicación de los resultados.

Palabras clave: intervalos de confianza, estadística en salud, estimadores puntuales, contraste de hipótesis.

ABSTRACT

In all biomedical research studies, daily clinical and epidemiological reports are made, so it is necessary to know perfectly well what should be reported and how to do it. The objective of this paper was to describe the advantages of the estimation of the confidence intervals for the right interpretation of results in the scientific research studies. A review of textbooks of biostatistics and statistics for health care was made in addition to scientific articles collected in Scielo, MedLine and PubMed and published from 2004 to 2014. For the search in electronic databases, the used subject headings were confidence intervals, health statistics, point estimators and hypothesis comparisons. In most of the reviewed documents, we found that the best way of presenting the statistical significance of the results is through the 95% confidence intervals. The most used methods for their estimations in descriptive, inferential, clinical and epidemiological research studies were presented in this paper. The confidence intervals estimation is a solid and easy-to-apply method, provides greater robustness to any research analysis since it evaluates the range in which the real value probably lies, and therefore, allows making better interpretation and application of the results.

Keywords: confidence intervals, health statistics, point estimators, hypothesis comparison.

INTRODUCCIÓN

En la especialidad de angiología y cirugía vascular, como en el resto de las ciencias de la salud, se realizan diariamente informaciones de investigaciones tanto clínicas como epidemiológicas, y es importante llegar a la correcta utilización e interpretación de los resultados, dependiendo de cuál metodología, clásica o bayesiana, sea utilizada.²⁻⁴ Entre estos métodos se encuentran la estimación puntual, la estimación por contraste de hipótesis y la estimación por intervalos de confianza.^{5,6}

En la estimación puntual, como su nombre lo indica, el valor de un parámetro poblacional se calcula mediante un único valor que se deriva de dicha muestra, por tanto, no permite especificar las diferentes variaciones de la estimación sobre otras posibles muestras.^{5,6} Es por ello que no es viable derivar una medida que permita determinar con qué grado de certidumbre el valor obtenido en la muestra refleja o infiere el verdadero valor en la población. Para que un estimador puntual sea considerado óptimo, buen estimador, debe cumplir cuatro propiedades:

1. Insesgado
2. Consistente
3. Eficiente
4. Suficiente

Por otra parte, mediante los contrastes de hipótesis, se busca verificar una hipótesis nula (H_0) contra una alternativa (H_1), ambas simples y que indican que una región crítica provoca un error de tipo I con probabilidad 0,05 y un error de tipo II con probabilidad 0,1 mediante la interpretación frecuentista de la probabilidad. Esto quiere decir, que si se repite dicho contraste más veces, conducirán a rechazar incorrectamente H_0 en el 5 % de los casos y aceptar incorrectamente en el 10 % de los casos a H_1 . Esto no implica que no se planteen hipótesis en las investigaciones que lo requieran, sino que su análisis se realiza mediante los valores del intervalo de confianza (IC) y no del valor de p como habitualmente se hace.⁷

A pesar de que estas dos variantes se han utilizado por mucho tiempo en la actualidad existe una tendencia al uso de los IC, pues refieren el grado de confianza o probabilidad de que, al aplicar repetidamente un procedimiento, el intervalo contenga el parámetro sujeto a análisis. Es decir, expresa la proporción de intervalos que efectivamente incluyen el parámetro. Este es un procedimiento que permite, mediante un análisis estadístico descriptivo o inferencial, determinar si existen diferencias significativas entre muestras o poblaciones.⁷⁻⁹

En otras palabras, el IC describe la variabilidad entre la medida obtenida en un estudio y la medida real de la población (valor real) y se corresponde con un rango de valores, cuya distribución es normal y en el cual se encuentra, con alta probabilidad, el valor real de una determinada variable. Esta alta probabilidad se ha establecido por consenso en 95 %, pero puede calcularse también para el 99 % de confianza o el 90 %. Así, un intervalo de confianza de 95 % (IC_{95 %}) indica que dentro del rango dado se encuentra el valor real de un parámetro con el 95 % de certeza.^{8,9}

A pesar de las grandes ventajas que proporciona el análisis por intervalos de confianza, y que se encuentra muy bien descrita la metodología para su cálculo, aún no está generalizado, debido en parte porque se hace complejo aplicar las fórmulas matemáticas que aparecen en los libros, sin embargo, su determinación se puede realizar mediante programas estadísticos como *Statistic Program Social Sciences (SPSS)*, *Minitab* o *Statistic*.

En este contexto, el propósito de este artículo es describir la utilidad del cálculo de los intervalos de confianza para la correcta interpretación de los resultados en las investigaciones biomédicas.

MÉTODOS

Se realizó una revisión de libros de texto de bioestadística y estadística para salud pública, además de artículos científicos publicados en bases de datos como Scielo, MedLine y PubMed, todos del período entre 2004 y 2014. Para la búsqueda electrónica se utilizaron los descriptores: intervalos de confianza, estadística en salud, estimadores puntuales y contrastes de hipótesis.

RESULTADOS

A diario en las distintas especialidades biomédicas nos podemos encontrar con situaciones problemáticas como la siguiente: Un angiólogo necesita llevar a cabo una investigación donde se recojan los factores de riesgo (r) que inciden en la alta frecuencia de amputaciones de miembros inferiores en una localidad donde habitan 900 personas. Al final de la misma dispone de una gran cantidad de datos a partir de

variables cualitativas y cuantitativas. Él quiere que al presentar los datos estos den la mayor información posible sobre la población.

Para lograr este objetivo el especialista no se puede limitar a los cálculos de estimadores puntuales o de los análisis de contraste de hipótesis, sino que es necesario realizar una estimación por IC para los dos tipos de variables.

El IC es una medida de precisión que permite evaluar dos aspectos de un resultado (estimador puntual):

1. Si existe diferencia estadística significativa (relevancia epidemiológica).
2. Si dicha diferencia es relevante para que sea recomendada a los pacientes (relevancia clínica).⁹⁻¹¹

Cuando se calculan los IC se obtienen dos resultados o extremos: un límite inferior (LI) y un límite superior (LS), y entre estos dos valores se encuentra el estimador puntual que describe la región de aceptación de una hipótesis planteada. Es por ello que se utiliza tanto en la actualidad, pues recoge información más confiable durante su interpretación. La presentación de un IC para el 95 % de confianza en cualquier informe de una investigación, se puede realizar de diferentes formas, sin embargo, la más aceptada es la siguiente: IC_{95 %} (LI; LS).

Para el análisis de la relevancia epidemiológica se observa si el valor del estimador puntual (frecuencia absoluta, frecuencia relativa, media, mediana, moda, riesgo relativo) calculado es estadísticamente significativo respecto al tamaño muestral (n), o sea, si se encuentra dentro del rango calculado de IC. Hay que aclarar que no siempre el valor del estimador puntual coincide con el valor central del IC, lo cual no quiere decir que la investigación no tenga relevancia epidemiológica o clínica.¹²

Para el cálculo del IC se han empleado diversas fórmulas, siendo la más aceptada la planteada por *Wilson* (1927), conocida como método *score* (fórmula 1), en la cual se encuentran implícitos los valores de frecuencia de una variable que sea cero ($p=0$) y también es aplicable para muestras que resultan muy pequeñas:¹²

Fórmula 1: $IC = (A \pm B)/C$

Donde: $A = 2r + z^2$

$$B = z \sqrt{z^2 + 4r(1-r/n)}$$

$$C = 2(n + z^2)$$

Donde:

r : frecuencia del factor o característica a analizar en la población

z : coeficiente de confianza ([tabla](#))

n : tamaño poblacional o de la muestra

Tabla. Valores de los coeficientes de confianza para los niveles de confianza más utilizados en las investigaciones biomédicas

Nivel de confianza	90 %	95 %	99 %
Coefficiente de confianza (z)	1,64	1,96	2,57
z^2	2,69	3,84	6,60

Si en el ejemplo planteado anteriormente ($n= 900$) se obtiene que 210 (r) individuos son fumadores (factor a analizar) entonces, el IC_{95 %} para la frecuencia de fumadores sería:

$$A = 2 \times 210 + 3,84 = 423,84$$

$$B = 1,96 \sqrt{3,84 + [4 \times 210 (1-210/900)]} = 1,96 \sqrt{3,84 + [4 \times 210 \times 0,77]}$$

$$B = 1,96 \sqrt{3,84 + 646,80} = 1,96 \times \sqrt{650,54} = 1,96 \times 25,51 = 49,9996$$

$$C = 2 \times (900 + 3,84) = 1\ 807,68$$

Por tanto: IC_{95 %} (LI; LS)

$$IC_{95 \%} = (423 - 49,9996)/1\ 807,68; 423 + 49,9996/1807,68)$$

$$IC_{95 \%} (0,21; 0,26)$$

Esto quiere decir que la probabilidad de que en la población estudiada se encuentren individuos fumadores está entre el 21 y el 26 %. Si cuando se presente el resultado con el estimador puntual, en este caso frecuencia relativa, solamente se podría decir que: la probabilidad de encontrar fumadores en esta población es del 23,3 %.

Entonces: ¿Cómo es que se presenta correctamente este resultado en el informe científico que debe dar el especialista?

Existen estas variantes, sin embargo, la más aceptada es la No. 3:

1. La frecuencia de fumadores es del 23,3 %, con un intervalo de confianza del 95 % que oscila entre 21 y 26 %.
2. La frecuencia de fumadores es del 23,3 % [IC_{95 %} (21; 26)]
3. Se puede afirmar que en la población sujeta a investigación con el 95 % de confiabilidad la frecuencia de fumadores se encuentra ubicada en el intervalo (21; 26).

En el informe que tiene que dar el investigador aparece también el análisis de variables cuantitativas, como por ejemplo las concentraciones de colesterol total sérico. Al tener en cuenta esto, se obtuvo que la concentración media de este parámetro es de 4,25 mmol.L⁻¹, con una desviación estándar poblacional (σ) de 0,97. Aquí la fórmula varía pero el análisis es el mismo (fórmula 2), pues se planteará que con el 95 % de confianza la concentración de colesterol sérico estará entre los valores calculado para el IC_{95%}:

$$\text{Fórmula 2: IC}_{95\%} \left(\bar{x} - \frac{z \cdot \sigma}{\sqrt{n}} ; \bar{x} + \frac{z \cdot \sigma}{\sqrt{n}} \right)$$

Donde:

La \bar{x} promedio representa la concentración media del colesterol.

σ : desviación estándar de la media poblacional de ese parámetro.

z : coeficiente de confianza (tabla)

n : tamaño poblacional o de la muestra

$$\text{Por tanto: IC}_{95\%} \left(4,25 - \frac{1,96 \times 0,97}{\sqrt{900}} ; 4,25 + \frac{1,96 \times 0,97}{\sqrt{900}} \right)$$

$$\text{IC}_{95\%} (4,19; 4,28)$$

Entonces: ¿Cómo es que se presenta correctamente este resultado en el informe científico?

Existen las mismas variantes que se citaron anteriormente:

1. La concentración media de colesterol sérico es 4,25 mmol.L⁻¹, con un intervalo de confianza del 95 % que oscila entre 4,19 mmol.L⁻¹ y 4,28 mmol.L⁻¹.
2. La concentración media de colesterol sérico es 4,25 mmol.L⁻¹ ; IC_{95 %} (21; 26).
3. Se puede afirmar que en la población sujeta a investigación con el 95 % de confiabilidad la concentración media de colesterol sérico se encuentra ubicada en el intervalo (4,19; 4,28).

Se debe precisar que mientras más pequeña sea la distancia entre el límite inferior y el límite superior (intervalos más estrechos), más preciso es, y por ende menos probable de que pueda contener el valor nulo. Del mismo modo, si se trabaja con menor seguridad, el intervalo también será más estrecho, y viceversa de nuevo; pero utilizar un IC como mero sustituto del valor p no es correcto. Así no se saca ningún partido: para eso ya está el valor p . Por tanto a la hora de informar un resultado se hace necesario también informar el valor p para el análisis realizado.^{12,13}

Si se desea determinar si es verdadera la diferencia observada entre dos grupos, se calcula el intervalo de confianza de 95 % de la diferencia entre sus respectivas medias. Si el intervalo abarca el valor cero, no se puede descartar que no haya una diferencia; si no lo abarca, la probabilidad de que se esté observando una diferencia que en realidad no existe se considera remota. La misma lógica se aplica al calcular el intervalo de un riesgo relativo, solo que en estos casos el valor 1 es el que indica la ausencia de una diferencia porque se trata de una proporción.¹³ No obstante, si los

intervalos de confianza solo se usaran de esta manera, entonces no se diferenciarían en nada de las pruebas de significación. Se puede observar, nuevamente, que no solo ofrecen mucha más información que los valores p , sino que abarcan a las pruebas de significación.

Ahora cabría preguntarse: ¿existirán algunos valores dentro de un intervalo de confianza que puedan tener relevancia práctica o clínica? La mayoría de las veces, si el mismo contiene valores de relevancia práctica, se situarán más bien cerca de alguno de sus límites, y ahí es donde se debe mirar para matizar más la importancia de un estudio y de su resultado. Dos son las situaciones que se pueden encontrar:¹⁴

1. Que el intervalo de confianza contenga el valor nulo (0 para una diferencia, 1 para una razón como el riesgo relativo) y, por lo tanto, que el resultado no sea estadísticamente significativo. En este caso, habría que mirar hacia el límite superior del intervalo de confianza y preguntar si ese límite superior tiene relevancia. Si la tiene, el estudio, pese a no ser concluyente (algunos los llaman negativos) al no presentar significación estadística, debe ser considerado como no definitivo, pues quizás de haber manejado mayores tamaños de muestra se obtendrían conclusiones de importancia. Valdría la pena, pues, volver sobre la cuestión con un mejor diseño. Si ese límite superior no tiene relevancia práctica, el estudio, aparte de ser negativo en cuanto a significación estadística, puede ser considerado como definitivo en cuanto a que el asunto investigado seguramente carece de importancia real.¹⁴
2. Que el intervalo de confianza no contenga el valor nulo y, por lo tanto, se está ante un resultado que posiblemente no se deba al azar. En este caso, se debe dirigir la mirada hacia el límite inferior del intervalo de confianza y preguntar de nuevo si ese límite tiene relevancia. Si la tiene, el estudio, además de ser concluyente o positivo en cuanto a significación estadística, será también definitivo en cuanto a su importancia práctica pues, aún en el peor de los casos que representa ese límite inferior, hay significación práctica además de significación estadística. Si ese límite inferior no presenta relevancia clínica a pesar de la significación estadística, no se puede considerar como definitivo el resultado. Otra vez valdría la pena volver sobre la cuestión con otro enfoque que pudiera aclarar la duda.¹⁴

CONSIDERACIONES FINALES

Como se ha descrito, la determinación de los intervalos de confianza es un método sólido y sencillo de aplicar y le da mayor robustez a cualquier análisis en una investigación. Por tanto si se desea lograr que los resultados a exponer sean más exactos, hay que tener en cuenta además, la elección del análisis estadístico y del diseño de estudio y dentro de este último, el proceso de muestreo, que continúa constituyendo la fuente de sesgo altamente prevalente en la literatura de investigación.¹²⁻¹⁴

En este mismo orden de ideas, se enfatiza que cuando se calcula un intervalo de confianza, se expresa solamente una forma de incertidumbre, la cual es el resultado de la naturaleza finita de la muestra estudiada. Cuando se emplea una muestra que sea representativa, en la que se incluye una selección de las unidades definidas por el investigador de manera aleatoria, la discusión sobre la generalización de los resultados sería más o menos sencilla.¹⁴ Pero si por alguna razón la muestra no es

representativa de la población relevante, tanto la estimación puntual (por ejemplo, la proporción específica de aquellos que se han amputado, por decir, 43 %) como los intervalos de confianza a calcular (por decir, de 31 a 58 %) serán afectados por esta situación. Si la muestra específica está sesgada hacia la población que se pretende estudiar (por ejemplo, cierto segmento puede estar más motivado para llenar cuestionarios o para participar en estudios y es una muestra autoseleccionada), todo lo que se calcule estará sesgado. Ninguna de las herramientas estadísticas tales como las estimaciones puntuales, los intervalos de confianza o las pruebas de hipótesis, puede corregir tal sesgo. Y más aún, no se debería postular que tales herramientas superan el sesgo. En otras palabras, el diseño de todo el proceso de investigación debe controlar las fuentes de posibles sesgos.^{10,15}

Finalmente, se debe recordar que la estimación de los intervalos no da una información en términos absolutos, pues también solo ofrece una probabilidad de contener la cantidad en estudio. De esta manera, los intervalos son los límites que tienen una alta probabilidad de que ocurra un determinado valor poblacional y esta probabilidad es de una extensión del 95 % u otro límite crítico.

Como conclusión, planteamos que el análisis de un intervalo de confianza permite evaluar el rango de valores donde posiblemente se encuentra el valor real, y por lo tanto, realizar una mejor interpretación y aplicación clínica de los resultados.

REFERENCIAS BIBLIOGRÁFICAS

1. Comité Internacional de Editores de Revistas Médicas (ICMJE). Requisitos de uniformidad para manuscritos enviados a revistas biomédicas: Redacción y preparación de la edición de una publicación biomédica. Rev Cubana Salud Pública. 2012 [citado 30 Abr 2014]; 38(2): 300-43. Disponible en: <http://www.scielo.sld.cu/pdf/rcsp/v38n2/spu14212.pdf>
2. Limón J, Rodríguez M, Báez Y, Tlapa D. Evaluación de la robustez del sistema Mahalanobis-Taguchi a diferentes arreglos factoriales. Información Tecnológ. 2011; 22(4): 85-96.
3. Limón J, Rodríguez M, Sánchez J, Tlapa D- Metodología bayesiana para la optimización simultánea de múltiples respuestas. Información Tecnológ. 2012; 23(2): 151-66.
4. Campbell G. Guidance for Industry and FDA Staff Guidance for the use of Bayesian statistics in medical device clinical trials. 2010 [citado 15 Ene 2014]. Disponible en: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf>
5. Feinsinger P. Lo que es, lo que podría ser y el análisis e interpretación de los datos de un estudio de campo. Ecol Bolivia. 2012; 47(1): 1-6.
6. Silva-Ayçaguer LC, Suárez-Gil P, Fernández-Somoano A. 2010. The null hypothesis significance test in health sciences research (1995-2006): statistical analysis and interpretation. BMC Medical Research Methodology. 2010 [citado 15 Ene 2014]; (10). Disponible en: <http://www.biomedcentral.com/1471-2288/10/44>

7. Van Belle G, Lloyd F, Heagerty P, Lumley T. Point and Interval Estimates. En: Balding D, Cressie N, Fisher N, editores. Biostatistics: a methodology for the health sciences. EE. UU.: John Wiley and Sons, Inc.; 2004. p. 85-7.
8. Campos A, Ceballo G, Herazo E. Prevalence of pattern of risky behaviors for reproductive and sexual health among middle- and high-school students. Rev Latino-Am Enfermagem. 2010 [citado 15 Ene 2014]; 18(2). Disponible en: <http://www.dx.doi.org/10.1590/S0104-11692010000200005>
9. Escrig J, Miralles J, Martínez D, Rivadulla I. Intervalos de confianza: por qué usarlos. Cir Esp. 2007;81(3):121-5.
10. Tajer C. Ensayos terapéuticos, significación estadística y relevancia clínica. Rev Argent Cardiol. 2010 [citado 10 Ene 2014]; 78(4). Disponible en: http://www.scielo.org.ar/scielo.php?pid=S1850-37482010000400019&script=sci_arttext
11. Kaul S, Diamond G. Trial and error how to avoid commonly encountered limitations of published clinical trials. J Am Coll Cardiol. 2010;55:415-27.
12. Newcombe R, Merino C. Intervalos de confianza para las estimaciones de proporciones y las diferencias entre ellas. Interdisciplinaria. 2006;23(2):141-54.
13. Clark M. Los valores P y los intervalos de confianza: ¿en qué confiar? Rev Panam Salud Pública. 2004 [citado 21 Ene 2014]; 15(5). Disponible en: <http://www.dx.doi.org/10.1590/S1020-49892004000500001>
14. Fay M. Confidence intervals that match Fisher's exact or Blaker's exact tests. Biostat. 2010;11(2):373-4.
15. Sarria Castro M, Silva Ayçaguer LC. Las pruebas de significación estadística en tres revistas biomédicas en lengua española: una revisión crítica. Rev Panam Salud Pública. 2004;15(5):300-6.

Recibido: 28 de febrero de 2014.

Aprobado: 23 de abril de 2014.

Yunier Arpajón Peña. Facultad de Estomatología "Raúl González Sánchez". Ave. Salvador Allende y Ave de los Presidentes. Plaza de Revolución, La Habana. Cuba. Dirección electrónica: mpcosme@infomed.sld.cu