

Ventajas del uso de estimaciones por intervalos para proporciones poblacionales

Advantages of the use of interval estimations for population proportions

Dra. Ana Carlina López Himely, MSc. Silvio Faustino Soler Cárdenas

Universidad de Ciencias Médicas de Matanzas. Matanzas, Cuba.

RESUMEN

En este trabajo los autores argumentan la necesidad de incluir el cálculo de estimaciones por intervalos para proporciones poblacionales. Se discute un ejemplo real donde se pone de manifiesto la insuficiencia del intervalo estándar para las proporciones, a pesar de su uso generalizado en las investigaciones. Se propone el uso del intervalo de Wilson.

Palabras clave: estimaciones por intervalos, proporciones poblacionales.

ABSTRACT

In this article, the authors argue the necessity of including the calculation of interval estimations for population proportions. A real example is discussed the insufficiency of the standard interval for proportions is stat, in spite of its generalized use in researches. The Wilson's interval is proposed.

Key words: interval estimations, population proportions.

INTRODUCCIÓN

La Estadística es, sin lugar a dudas, una disciplina metodológica que en los últimos decenios ha sido aplicada con gran profusión en la investigación en salud. El conjunto de sus métodos abarca las llamadas medidas estadísticas, tablas y gráficos que resultan muy útiles para la descripción de fenómenos empíricos. Comprende además las pruebas de hipótesis estadísticas y procedimientos de estimación de parámetros que se aplican cuando es preciso tomar decisiones bajo condiciones de incertidumbre.

Tanto la práctica investigativa como la publicación de artículos científicos evidencian que con alguna frecuencia se omite la necesaria aplicación de procedimientos estadísticos, o bien se aplican inadecuadamente.

En este informe, los autores argumentan la necesidad de incluir estimaciones por intervalos para proporciones poblacionales.

DESARROLLO

Para ejemplificar las ideas principales, se tiene en cuenta la siguiente situación: se realizó un estudio parasitológico,¹ para evaluar la proporción (o el porcentaje) de pacientes de un hospital que están infectados con determinadas especies de parásitos: giardia duodenalis, entamoeba coli, endolimax nana y necátor americanus. La determinación exacta de esas proporciones (parámetros desconocidos) necesitaría estudiar a todos los pacientes; pero esto, por razones económicas y de tiempo no es aconsejable. En situaciones como esta, el procedimiento usual consiste en calcular las proporciones en una muestra de pacientes y generalizar los resultados al conjunto de todos los casos del hospital (universo de estudio).

En términos estadísticos, lo anterior se expresa diciendo que se seleccionó una muestra del universo de estudio y se calcularon estimaciones puntuales de ciertos parámetros de interés.

Una estimación puntual,² de un parámetro es un número calculado a partir de una muestra que expresa aproximadamente el valor desconocido de un parámetro poblacional.

La tabla 1 muestra los resultados obtenidos a partir de 76 pacientes tomados como muestra.

Tabla 1. Estimaciones puntuales (76 pacientes)

Especie	No.	Proporción	Porcentaje
Giardia duodenalis	10	0,178	17,8
Entamoeba coli	4	0,071	7,1
Endolimax nana	3	0,054	5,4
Necátor americanus	1	0,018	1,8

La tercera columna de la tabla 1 contiene las estimaciones puntuales de las proporciones para cada especie (a veces se usan porcentajes, conforme aparece en la última columna de la tabla).

Por ejemplo, se puede decir que la proporción de pacientes portadores de Giardia duodenalis se encuentra “alrededor de 0,178”. Una interpretación similar se puede hacer para los restantes parásitos.

Nótese la falta de precisión de la expresión “alrededor de 0,178”: la misma es consecuencia directa de que los resultados se basan en una muestra de pacientes. Las estimaciones puntuales siempre proporcionan valores aproximados de parámetros, como consecuencia del error de muestreo.

Ahora bien, ¿es posible cuantificar ese error de muestreo? Una solución satisfactoria la proporciona otro tipo de estimación.

Un intervalo de confianza (estimación por intervalos) es un recorrido de valores, basados en una muestra tomada de una población, en el que cabe esperar que se encuentre el verdadero valor de un parámetro poblacional con cierto grado de confianza.³

Existen varias fórmulas para el cómputo de intervalos de confianza, algunas son de uso restringido pues dependen del tamaño de la muestra del estudio y su uso inadecuado proporciona en ocasiones resultados absurdos. Este es el caso del llamado “Intervalo Estándar” que, lamentablemente, se aplica con frecuencia en la práctica investigativa y también en libros de texto de Estadística y Bioestadística. Es probable que eso se deba a la simplicidad algebraica de su fórmula:

$$p \pm z \sqrt{\frac{p(1-p)}{n}}$$

La tabla 2 muestra los intervalos de confianza aplicando el intervalo estándar. A modo de ejemplo se puede decir que, aunque se desconozca la proporción poblacional de pacientes con Giardia duodenalis, se puede afirmar, con un 95% de probabilidad, que esa proporción se encuentra entre 0,078 y 0,279.

Un intervalo de confianza no determina unívocamente el valor desconocido de un parámetro; pero establece un recorrido de máxima probabilidad para el mismo, es

decir, de cierto modo se cuantifica la incertidumbre que produce hacer los cálculos considerando solamente una muestra del universo de estudio.

Por otra parte, nótese que aquellos que corresponden a las especies endolimax nana y necátor americanus, no son interpretables pues contienen números negativos: el intervalo estándar tiene este gran inconveniente y por tanto no es válida su aplicación generalizada.

Tabla 2. Intervalos de confianza (Estándar del 95%)

Especies	Intervalo estándar
Giardia duodenalis	0,078 ----- 0,279
Entamoeba coli	0,004 ----- 0,139
Endolimax nana	-0,005 ----- 0,113
Necátor americanus	-0,017 ----- 0,053

En la literatura especializada se han propuesto variadas formulaciones para la estimación por intervalos de proporciones. En este sentido, baste mencionar a Lanchón,⁴ y Brown,⁵ quienes, respectivamente, presentan tres y seis intervalos de confianza.

Por otra parte Kott,⁶ y Agreste,⁷ destacan las propiedades del llamado “intervalo de confianza de Wilson” y lo recomiendan para las aplicaciones.

La expresión del intervalo de Wilson es:

$$\frac{2np + z^2}{2(z^2 + n)} \pm \frac{z\sqrt{4np(1-p) + z^2}}{2(z^2 + n)}$$

En la expresión anterior la letra “n” representa el tamaño de la muestra, “p” la estimación puntual y “z” el percentil de la distribución normal de probabilidades cuyo valor corresponde a una confiabilidad prefijada.

La tabla 3 contiene los resultados para la fórmula de Wilson. Se aprecia que todos ellos están incluidos en el intervalo abierto (0; 1).

Tabla 3. Intervalos de confianza (Wilson del 95%)

Especies	Intervalo de Wilson
Giardia duodenalis	0,100 ----- 0,298
Entamoeba coli	0,028 ----- 0,169
Endolimax nana	0,019 ----- 0,147
Necátor americanus	0,003 ----- 0,095

CONCLUSIONES

Los intervalos de confianza son mejores que las estimaciones puntuales a los efectos de la descripción estadística de proporciones poblacionales, se recomienda el uso del intervalo de confianza de Wilson.

REFERENCIAS BIBLIOGRÁFICAS

- 1- González Montero Y, Cañete Villafranca R, Machado Cazorla K, et al. Parasitosis intestinal en pacientes internados en el Hospital Provincial Psiquiátrico Docente Antonio Guiteras Holmes. Matanzas, Cuba. Rev Méd Electrón [Internet]. 2014 [citado 22 mayo 2018]; 36(2). Disponible en: <http://www.revmatanzas.sld.cu/revista%20medica/ano%202014/vol2%202014/tema03.htm>
- 2- Reyes H, Almendra F. Problemas al usar la aproximación normal n intervalos de confianza suponiendo datos Bernoulli [Internet]. Colombia: XXV Simposio Internacional de Estadística [citado 22 mayo 2018]; 2015. Disponible en: http://simposioestadistica.unal.edu.co/fileadmin/content/eventos/simposioestadistica/documentos/memorias/MEMORIAS_2015/Comunicaciones/Modelamiento/Reyes_Almendra_Morales_Tajonar_Aproximacion_Normal_Datos_Bernoulli.pdf
- 3- Batista NE, Hernández VJ, Guirado O. Pensamiento estadístico: el valor “p” y los intervalos de confianza. Medcent Electrón [Internet]. 2016 [citado 22 mayo 2018]; 20(2). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30432016000200001
- 4- Lachin JM. Bioestatistical Methods. The Assesment of Relative Risks. 2da. ed. New Jersey: John Wiley& Sons; 2011.
- 5- Brown LD, Cai T, DasGupta A. Interval Estimation for a Binomial Proportion. Statistical Science [Internet]. 2001 [citado 22 mayo 2018]; 16(2): 101–33. Disponible en: <http://epweb2.ph.bham.ac.uk/user/bansil/Files/Papers/Statistics/Binomial-IntervalEstimation.pdf>
- 6- Kot S. A note on Wilson coverage intervals for proportions estimated from complex samples. Survey Methodology [Internet]. 2017 [citado 22 mayo 2018]; 43(2): 235-40. Disponible en: <https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201700254872>
- 7- Agresti A. Categorical Data Analysis. 2da. ed. USA: John Wiley & Sons; 2002.

Recibido: 31/5/18
Aprobado: 7/6/18

Ana Carlina López Himely. Universidad de Ciencias Médicas de Matanzas. Correo electrónico: silviosoler.mtz@infomed.sld.cu

CÓMO CITAR ESTE ARTÍCULO

López Himely AC, Soler Cárdenas SF. Ventajas del uso de estimaciones por intervalos para proporciones poblacionales. Rev Méd Electrón [Internet]. 2018 May-Jun [citado: fecha de acceso]; 40(3). Disponible en:
<http://www.revmedicaelectronica.sld.cu/index.php/rme/article/view/2785/3873>