



Intervalos de confianza: una aproximación intuitiva para el no-estadístico

Rufino Menchaca-Díaz *

Resumen

La inferencia estadística clásica se establece obteniendo el valor de p mediante el método del contraste de hipótesis o estimando los intervalos de confianza de los resultados puntuales. Los intervalos de confianza permiten calcular el rango esperado de un estimador puntual y se basan en el conocimiento de la variabilidad muestral. En el presente ensayo se ofrece una aproximación intuitiva para la construcción e interpretación de intervalos de confianza para medias, proporciones y razón de momios. Se destacan las aplicaciones prácticas de esta forma de hacer inferencia estadística y el frecuente uso de los intervalos de confianza como requisito en la presentación de resultados en publicaciones arbitradas por pares.

Palabras clave: Análisis de datos, técnicas de estimación, intervalos de confianza.

Summary

Classical statistical inference is mainly based in p value calculation using the hypothesis contrast method, or estimating confidence intervals for punctual results. Confidence intervals represent the expected range of an estimator and are based on calculus about the sample's variability. In this essay an intuitive approach for the construction and interpretation of confidence intervals for means, proportions, and the odds ratio it's presented. The practical issues of this method for statistical inference are stressed and also the frequent use of confidence intervals as a requisite for results presentation in peer-review publications.

Key words: Data analysis, estimation techniques, confidence intervals.

INTRODUCCIÓN

Cuando se realiza una investigación, es usual que se estudie una muestra o subconjunto de la población y no a la población completa debido a muchas razones de factibilidad (*Figura 1*). El estudio de una muestra es suficiente para aproximarse al verdadero parámetro de la población sin tener que estudiar a todos los individuos que la con-

forman. Al estudiar una muestra se obtiene un estimador que no necesariamente es igual al verdadero parámetro de la población, pero, según las teorías del muestreo, debe ser bastante próximo, mientras la muestra sea grande y se haya seleccionado aleatoriamente.^{1,2} Si se efectúan diversas muestras de la misma población, se encontrarán diferencias en los estimadores (*Figura 2*). En la realidad, no es posible estudiar muestras variadas para asegurar que el verdadero parámetro poblacional esté determinado en al menos una de ellas, ni estudiar a toda la población de tal manera que no exista necesidad de hacer inferencia estadística. Sólo se dispone de una muestra, y con base en ésta se realiza el análisis e inferencia estadística de los resultados. Lo que interesa saber es si los resultados observados en la muestra son aplicables a la población de la que procede, es decir, que puedan ser *generalizables* a la población blanco (*Figura 1*). Una forma de generalizar lo observado en una muestra a la población blanco es a través de la construcción de intervalos de confianza de los estimadores (IC).

Los IC determinan la variabilidad esperada o posible del estimador obtenido en una muestra, tomando en cuenta el tamaño de la misma. Para construir un IC a partir de

* Neurólogo del Hospital Ángeles Tijuana. Profesor de Neurociencias, Epidemiología y Bioestadística de la Facultad de Medicina y Psicología de la Universidad Autónoma de Baja California.

Correspondencia:
Dr. Rufino Menchaca-Díaz
Correo electrónico: rufino@uabc.edu.mx

Aceptado: 23-05-2012.

Este artículo puede ser consultado en versión completa en <http://www.medigraphic.com/actamedica>

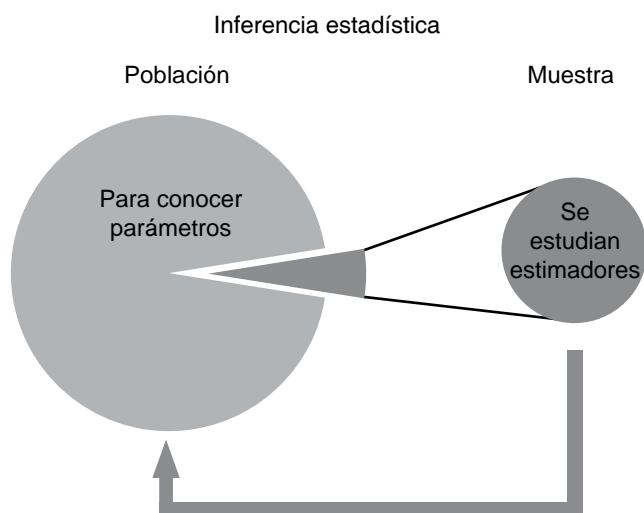


Figura 1. Esquema de la inferencia estadística de una muestra a la población.

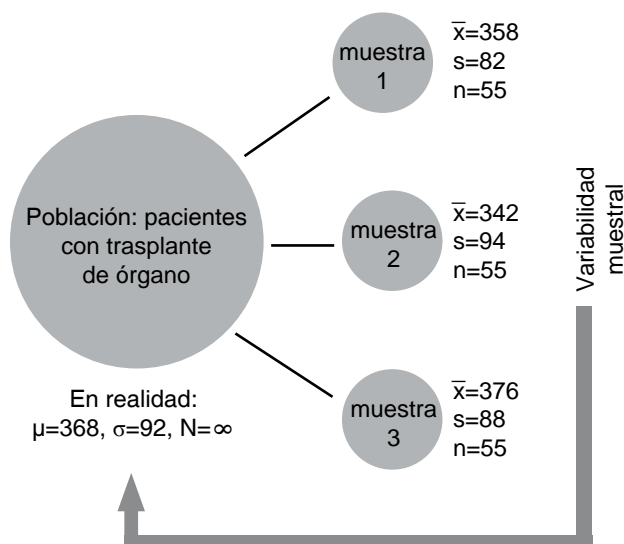


Figura 2. Distribución de los estimadores en el muestreo (datos hipotéticos).

un estimador se requiere: 1) Determinar el nivel de confianza que se desea alcanzar, usualmente del 95 o 99% y 2) Una estimación del error estándar, que representa la variabilidad muestral. En muestras grandes se obtiene escasa variabilidad muestral, lo que condiciona un error estándar pequeño e IC estrechos; en muestras pequeñas la variabilidad muestral es mayor, con error estándar grande e IC muy amplios. A continuación se presentan algunos ejemplos prácticos, usando la curva de distribución normal para el cálculo de IC y que nos sirven para ampliar estos conceptos.

IC PARA MEDIAS

En una investigación clínica es frecuente que se pretenda conocer una media o promedio de una variable cuantitativa o numérica. La media que se estima en una muestra está sujeta a la *variabilidad muestral* (Figura 2). Si se estudian muestras distintas del mismo tamaño, se espera observar ciertas diferencias entre ellas. La variabilidad esperada será mayor si el tamaño de las muestras disminuye o menor si el tamaño de las muestras aumenta. Con cada incremento del tamaño de las muestras, la variabilidad muestral disminuye hasta el punto en que el estimador se encuentre sobre el verdadero parámetro poblacional y la variabilidad muestral sea nula. En la realidad, se tiene oportunidad de estudiar una sola muestra y con base en ésta se calcula la variabilidad posible del estimador. Para realizar el cálculo del IC y conocer la variabilidad posible del estimador se requiere: 1) Establecer el nivel de confianza que se desea observar y 2) Conocer el error estándar del estimador,

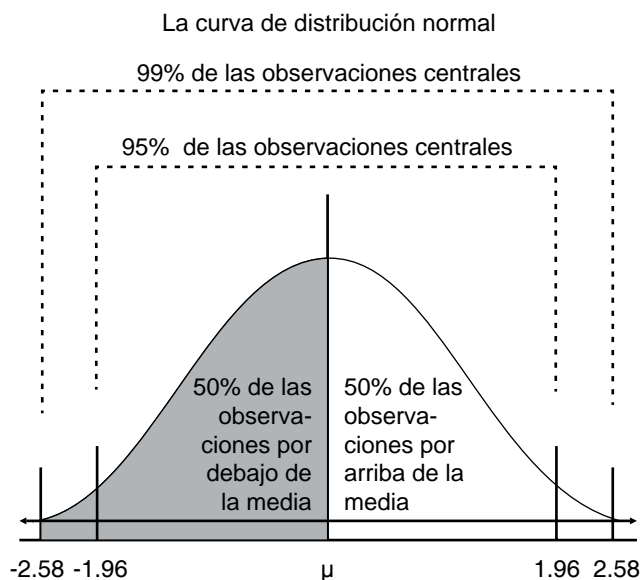


Figura 3. Esquema de la curva de distribución normal con valores de z.

que en este caso es una media. Para asignar el nivel de confianza en el cálculo, se emplean los valores de z de la curva de distribución normal; a una confianza del 95% se emplea el valor de z de ± 1.96 , mientras que a una confianza del 99% se emplea un valor de z de ± 2.58 (Figura 3). Bajo ciertas condiciones también se pueden usar los valores t de la distribución t de Student. Las fórmulas para

Cuadro I. Fórmulas usadas para el cálculo de los IC en los ejemplos.

Fórmula 1:

$$\text{Error estándar de la media } \bar{x} = \frac{s}{\sqrt{n}}$$

Fórmula 2:

$$\text{IC 95\% para la media} = \bar{x} \pm 1.96 * \frac{s}{\sqrt{n}}$$

Fórmula 3:

Error estándar para la diferencia de medias =

$$(\bar{x}_1 - \bar{x}_2) = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Fórmula 4:

Desviación estándar ponderada entre los grupos =

$$s_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Fórmula 5:

IC 95% para la diferencia de medias =

$$(\bar{x}_1 - \bar{x}_2) \pm 1.96 * e.e. (\bar{x}_1 - \bar{x}_2)$$

Fórmula 6:

$$\text{Error estándar de la proporción } p = \sqrt{\frac{p(1-p)}{n}}$$

Fórmula 7:

$$\text{IC 95\% para la proporción } p = p \pm 1.96 * \sqrt{\frac{p(1-p)}{n}}$$

Fórmula 8:

Error estándar de la diferencia de proporciones

$$(p_1 - p_2) = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Fórmula 9:

IC 95% para la diferencia de proporciones =

$$(p_1 - p_2) \pm 1.96 * e.e. (p_1 - p_2)$$

Fórmula 10:

$$\text{Razón de momios } RM = \frac{ad}{bc}$$

Fórmula 11:

$$\text{Error estándar de } \ln RM = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Fórmula 12:

$$\text{IC 95\% para } \ln RM = \ln RM \pm 1.96 * \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

el cálculo del error estándar de la media y del IC para la media se presentan en el *cuadro I*.

En el ejemplo de la *figura 2*, en la muestra 1 existe una media de 358 días con una desviación estándar de 82, en una muestra tamaño $n = 55$, por lo que se obtiene un IC al 95% de la media de 336.33 a 379.67. En estos resultados, la media obtenida en la muestra es de 358 días, pero se anticipa una variabilidad que puede ir desde 336.33 hasta 379.67 días de sobrevivencia del trasplante. En este ejemplo hipotético, la verdadera media poblacional y los otros estimadores muestrales (*Figura 2*) están comprendidos dentro del IC calculado a partir de la muestra 1. En realidad, nunca se llega a conocer si el verdadero parámetro poblacional se encuentra dentro del IC calculado; sólo se puede calcular la variabilidad muestral esperada a partir del único estimador muestral conocido. Si se estudian diversas muestras tamaño $n = 55$ del ejemplo, se espera que el 95% de las mismas presenten estimadores que queden dentro del IC y, con ello, se logre una estimación menos sesgada del posible parámetro poblacional.

Cuando se estudian medias en distintos grupos, se pueden comparar las diferencias entre ellos y determinar si la diferencia es estadísticamente significativa usando IC. Por ejemplo, en un estudio hipotético se analizó el porcentaje promedio de hemoglobina glucosilada (HbA1c) que mostraron 68 pacientes en tratamiento con antidiabéticos orales (*grupo 1*) y 72 pacientes en tratamiento con insulina (*grupo 2*). La media de HbA1c para el *grupo 1* fue de 9.2 (± 2.7), mientras que para el *grupo 2* fue de 8.4 (± 2.2). La diferencia observada entre los grupos es de 0.8 ($\bar{x}_1 - \bar{x}_2$). ¿Es esta diferencia estadísticamente significativa? o por el contrario ¿es la diferencia observada parte de la variabilidad muestral? Para contestar esta pregunta se construye un IC con la diferencia observada. Para construir el IC para la diferencia de medias se requiere: 1) Establecer el nivel de confianza que se desea observar y 2) Conocer el error estándar de la diferencia de medias. Para asignar el nivel

de confianza se emplean los valores de z ya descritos. Bajo ciertas condiciones, también se pueden emplear los valores t de la distribución *t de Student*. Las fórmulas para el cálculo del error estándar de la diferencia de medias y del IC para la diferencia de medias se presentan en el *cuadro I*. Con la información del ejemplo se obtiene un IC 95% para la diferencia de medias de -1.317 a 2.9168.

En este ejemplo, se observa un IC muy amplio, que va desde números negativos (-1.317) hasta números positivos (2.9168), quedando incluido el valor 0 dentro del IC. Con esto, se puede concluir que la diferencia observada en estos resultados no es estadísticamente significativa.

IC PARA PROPORCIONES

Los IC también pueden emplearse cuando se estudian proporciones. Las proporciones, al igual que las razones y las tasas son el producto del análisis de variables cualitativas o categóricas. En el *cuadro II* se presenta información hipotética de un estudio sobre tabaquismo en adolescentes, en el que la proporción total de fumadores es de 0.273 (27.3%). Esta proporción estimada está sujeta a la variabilidad muestral. Si la muestra hubiese estado conformada por otros sujetos ¿se habría detectado el mismo número de fumadores? Lo más probable es que no. ¿Qué tanta variación cabría esperar del resultado observado? Para contestar esta pregunta se construye un IC para la proporción observada. Para construir el IC para la proporción se requiere: 1) Establecer el nivel de confianza que se desea observar y 2) Conocer el error estándar de la proporción. Para asignar el nivel de confianza en el cálculo se emplean los valores de z como se describió antes. Las fórmulas para el cálculo del error estándar de una proporción y del IC para una proporción se presentan en el *cuadro I*. En el ejemplo que se presenta, existe una proporción de fumadores de 0.273 (27.3%) en una muestra tamaño $n = 860$; por lo tanto, se obtiene un IC al 95% de 0.244 a 0.302.

Cuadro II. Estudios sobre tabaquismo en adolescentes (datos hipotéticos).

		Fuman n (%)		No fuman n (%)		Total n (%)
Hombres	a	153 (32.3)	b	320 (67.7)	a + b	473 (100)
Mujeres	c	82 (21.2)	d	305 (78.8)	c + d	387 (100)
Total	a + c	235 (27.3)	b + d	625 (72.7)	a + b + c + d	860 (100)

En estos resultados, la proporción obtenida en la muestra fue de 0.273, pero se anticipa una variabilidad que va desde 0.244 hasta 0.302 (24.4 a 30.2%). En este ejemplo hipotético, la verdadera proporción poblacional es desconocida.

Cuando se estudian proporciones en distintos grupos se pueden comparar las diferencias entre ellos y determinar si la diferencia es estadísticamente significativa usando IC. Así, en el mismo ejemplo hipotético se observó que en los 860 adolescentes entrevistados, 387 (0.45 ó 45%) eran mujeres y 473 (0.55 ó 55%), varones. De las 387 mujeres, 82 eran fumadoras (0.212 ó 21.2%; p_1) y de los 473 varones, 153 eran fumadores (0.323 ó 32.2%; p_2). La diferencia observada de fumadores entre hombres y mujeres es de 0.111 u 11.1% ($p_1 - p_2$). ¿Es esta diferencia estadísticamente significativa? O por el contrario, ¿es la diferencia observada parte de la variabilidad muestral? Para contestar esta pregunta se construye un IC con la diferencia observada. Para construir el IC para la diferencia de proporciones se requiere: 1) Establecer el nivel de confianza que se desea observar y 2) Conocer el error estándar de la diferencia de proporciones. Para asignar el nivel de confianza se emplean los valores de z , como se mencionó antes. Las fórmulas para el cálculo del error estándar de la diferencia de proporciones y del IC para la diferencia de proporciones, se presentan en el *cuadro II*. Con la información del ejemplo se obtiene un IC al 95% para la diferencia de proporciones de 0.0522 a 0.1698.

En este ejemplo, el IC va de 0.0522 a 0.1698 (5.22 a 16.98%), quedando excluido del rango el valor 0. En la muestra estudiada de 860 adolescentes; 387 mujeres y 474 varones, se observó una diferencia de 11.1% en la proporción de fumadores entre los dos grupos. De acuerdo con la variabilidad en la muestra estudiada, se pueden esperar diferencias que van desde 5.22% hasta 16.98%, siempre con una diferencia por arriba de 0, lo que indica que la proporción de varones fumadores es consistente y significativamente mayor que la proporción de mujeres fumadores, más allá de lo que pudiera ser debido a la variabilidad muestral. Con esto se puede concluir que la diferencia observada en estos resultados es estadísticamente significativa.

IC PARA LA RAZÓN DE MOMIOS (RM)

La RM (*odds ratio*, razón de productos cruzados o razón de disparidad) es una medida de resumen que se utiliza con mucha frecuencia en estudios transversales y de casos y controles. Sirve para describir la asociación numérica entre dos variables dicotómicas. Se basa en el cálculo de momios o razón de probabilidades, que es la probabilidad de tener una condición sobre la proba-

bilidad de no tenerla. En el ejemplo de tabaquismo en adolescentes (*Cuadro II*) se puede estimar la asociación entre el sexo y el tabaquismo al comparar el nivel de exposición al tabaquismo entre los dos grupos mediante la fórmula que aparece en el *cuadro I*. La RM que se obtiene con estos datos es igual a 1.78, lo que significa que el grupo expuesto (varones) tuvo una frecuencia de tabaquismo de 178% con respecto del grupo no expuesto (mujeres), o que los varones tuvieron una frecuencia de ser fumadores de 0.78 (78%) veces más que las mujeres.

La razón de momios es otro estimador que se calcula a partir de una muestra. Volviendo a la pregunta sobre la variabilidad muestral, si la muestra de 860 adolescentes hubiese sido conformada por distintos sujetos ¿se mantendría el mismo número de varones y mujeres fumadores y no fumadores? Lo más probable es que no. El número y la proporción de hombres y de mujeres fumadores mostrarán variación y, por ende, la razón de momios también variará ¿Qué variabilidad se puede esperar de la razón de momios de acuerdo al tamaño de la muestra? Para contestar esta pregunta, se construye un IC para la razón de momios. Para construir el IC para la razón de momios se requiere: 1) Establecer el nivel de confianza que se desea observar y 2) Conocer el error estándar de la razón de momios. Para asignar el nivel de confianza en el cálculo se emplean los valores de z , como se mencionó antes. Una desventaja de la razón de momios es que no se apega a la distribución normal, por lo que el cálculo del IC debe realizarse usando una transformación logarítmica, usando el logaritmo natural (\ln) de la razón de momios en lugar de la razón de momios directa. El error estándar del \ln de la RM y el IC del \ln de la RM se obtienen con las fórmulas que aparecen en el *cuadro I*.

En el ejemplo que se presenta, el \ln de 1.78 es 0.5766 y su IC va de 0.2661 a 0.8871. Al regresarlo a las unidades originales mediante su exponencial, resulta en un IC de 1.30 a 2.42. En estos resultados, la razón de momios obtenida en la muestra fue de 1.78, pero se anticipa una variabilidad que va desde 1.30 hasta 2.42. Es importante considerar que el valor 1 quedó fuera del IC calculado, lo que nos habla de que el resultado es consistente y significativamente mayor en hombres que en mujeres. Si el IC de la razón de momios se hubiese extendido y quedara comprendido entre valores menores y valores mayores a 1, es decir, si el valor 1 queda dentro del IC calculado, indicaría que la diferencia observada no es significativa y que puede esperarse que en ocasiones el grupo expuesto tenga menos, igual o mayor frecuencia de la condición que el grupo no expuesto.

CONCLUSIONES

Los IC son una herramienta de la estadística inferencial que se emplea con mucha frecuencia en estudios clínicos. Aplican para cualquier medida de resumen que se obtenga de una muestra y permiten disponer de un estimador que considera la variabilidad muestral, por lo que producen un sesgo menos pronunciado al generalizar el resultado observado de la muestra a la población blanco. En el presente ensayo se revisaron los conceptos básicos sobre los IC para medias, proporciones y razón de momios; sin embargo, cualquier estimador es susceptible de analizarse con IC.

Los IC aparecen como requisito en las guías para la publicación de artículos científicos³⁻⁵ y cada vez es más frecuente que sean solicitados por los revisores de las publicaciones periódicas como evidencia estadística en la presentación de los resultados de estudios clínicos.⁶ En la medicina basada en evidencia, también se consideran como indicadores importantes que los médicos debemos conocer y saber interpretar, para mejorar la evaluación crítica de los artículos de investigación que leemos y tener una visión más amplia y menos sesgada de la literatura médica.⁷

REFERENCIAS

1. Pagano M, Gauvreau K. *Fundamentos de bioestadística*. México, DF: Thomson Learning; 2001.
2. Daniel WW. *Bioestadística. Base para el análisis de las ciencias de la salud*. 4ª ed. México, DF: Limusa; 2002.
3. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: up-sated guidelines for reporting parallel group randomized trials. *Ann Int Med* 2010; 152: 1-8.
4. Elm EV, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP. Declaración de la iniciativa STROBE (Strengthening the Reporting of Observational studies in Epidemiology): directrices para la comunicación de estudios observacionales. *Gac Sanit* 2008; 22: 144-150.
5. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. *Lancet* 1999; 354: 1896-1900.
6. Lang TA, Secic M. *How to report statistics in medicine. Annotated guidelines for authors, editors and reviewers*. 2ª ed. Philadelphia P, USA: American College of Physicians; 2006.
7. Guyatt G, Rennie D, eds. *User's guides to the medical literature. Essentials of evidence-based clinical practice*. Chicago I, USA: American Medical Association; 2002.
8. Altman DG. *Practical statistics for medical research*. London UK: Chapman & Hall/CRC; 1999.