



Lectura crítica en pequeñas dosis

El significado de los intervalos de confianza

M. Molina Arias

Publicado en Internet:
2-abril-2013

Manuel Molina Arias:
mma1961@gmail.com

Servicio de Gastroenterología. Hospital Infantil Universitario La Paz. Madrid. España.
Grupo de Trabajo de Pediatría Basada en la Evidencia, AEP/AEPap.
Editor de www.cienciasinseso.com

Resumen

Palabras clave:

- Intervalo de confianza
- Significación estadística
- Importancia clínica

Cuando precisamos conocer el valor de una variable en la población, debemos estimarla a partir de los datos obtenidos de muestras extraídas de esa población. Los intervalos de confianza nos permiten aproximar, una vez calculado el valor de la variable en la muestra, entre qué rango de valores se encuentra el valor real inaccesible de la variable en la población, con un grado de incertidumbre que podemos determinar.

Además, los intervalos de confianza aportan información añadida que se pierde cuando únicamente se valora la probabilidad de significación estadística, ya que nos permiten estudiar la precisión de la estimación del estudio y la relevancia de los resultados desde el punto de vista clínico.

The meaning of confidence intervals

Abstract

Key words:

- Confidence intervals
- Statistical significance
- Clinical significance

Every time we need to know the value of a variable in a population we have to estimate it from data obtained from samples drawn from that population. Confidence intervals allow us to approximate, once calculated the value of the variable in the sample, to the range of values that includes the unreachable actual variable value in the population, always within a degree of uncertainty that we can determine.

Moreover, confidence intervals provide additional information that is usually lost when we only assess statistical significance, allowing us to study the accuracy of the estimate variable and the clinical relevance of the results.

Imaginemos que queremos saber el valor medio del colesterol en sangre de la población de nuestra ciudad, en la cual viven cuatro millones de habitantes. ¿Hay modo de conocer este dato? ¡Claro!, solo tenemos que hacer cuatro millones de extracciones, analizar las muestras y hacer el cálculo. Pero el problema es evidente: no resulta práctico intentar medir el colesterol a toda la población. Así que decidimos tomar una muestra al azar, por ejemplo, de 100 personas, y determinamos su colesterol en sangre. Obtenemos una media (m) de 236 mg/dl con una desviación típica (s) de 41 mg/dl.

La cuestión, claro está, es que 236 es el valor medio de nuestra muestra, pero nosotros lo que queremos saber es el valor medio de la población de la que procede la muestra. Seguramente estará próximo a ese valor, pero lo más probable es que sea diferente. ¿Hay alguna forma de solucionar el problema?

Por desgracia, no existe forma de conocer el valor exacto en la población (a no ser que midamos el colesterol en los cuatro millones, sin excepción), pero sí podemos obtener un valor dentro de un rango determinado, aunque siempre con un grado

de incertidumbre que, eso sí, podemos elegir nosotros.

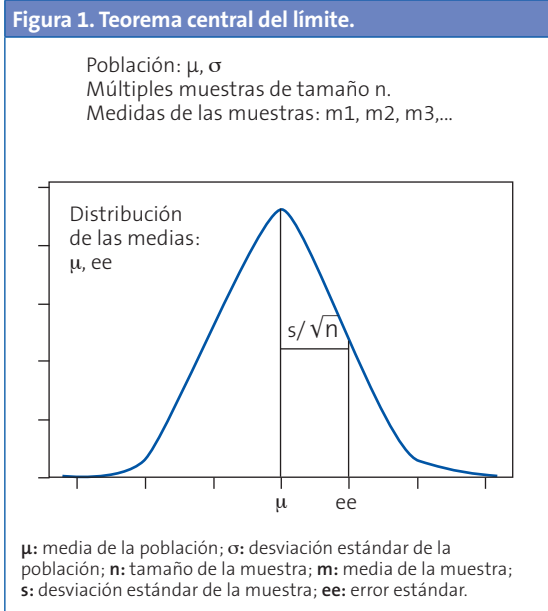
Vamos a suponer que repetimos el experimento con otras nueve muestras elegidas al azar para obtener las diez parejas de *m* y *s* de la **Tabla 1**.

Si calculamos la media de todas las medias obtendremos un valor que se aproximará al valor medio real de la población, tanto más cuanto más medias sumemos a la distribución. En este caso, el valor medio es 235 y su desviación estándar 2,4. Y aquí es cuando empieza la parte interesante, porque no hace falta repetir el experimento un número enorme de veces. Sabemos que, cuando el número de medias es grande, esta distribución de medias sigue una curva normal cuya media es igual a la de la población (μ) y cuya desviación estándar es igual a la desviación estándar de la población dividida entre la raíz cuadrada del tamaño de la muestra (**Fig. 1**). Este cociente se conoce con el nombre de error estándar de la media (*ee*) y es equivalente al valor de la desviación estándar de la distribución de medias que calculamos previamente.

Y, sin darnos cuenta, acabamos de enunciar el teorema central del límite, que es la base para el cálculo de intervalos de confianza (IC), de intervalos predictivos, de tamaños de muestras y, prácticamente, de todo el contraste de hipótesis en estadística. ¿Y qué objeto tiene todo esto? Pues la magia está en que sabemos que, en una distribución

Tabla 1. Distribución de media (m) y desviación estándar (s) de los valores de colesterol sérico (mg/dl) en las diez muestras de 100 personas obtenidas al azar de nuestra población

Muestra	m	s
1	236	41
2	234	43
3	238	44
4	233	44
5	235	43
6	233	38
7	235	43
8	233	43
9	240	47
10	235	37
Valor medio de las diez muestras	235	2,4



normal, el 95% de la población se encuentra en el intervalo de su media $\pm 1,96$ veces su desviación estándar (**Fig. 2**), lo que nos permite calcular entre qué valores estará el 95% de las medias de todas las muestras que hemos elegido al azar. El rango entre esos dos valores es lo que se denomina IC:

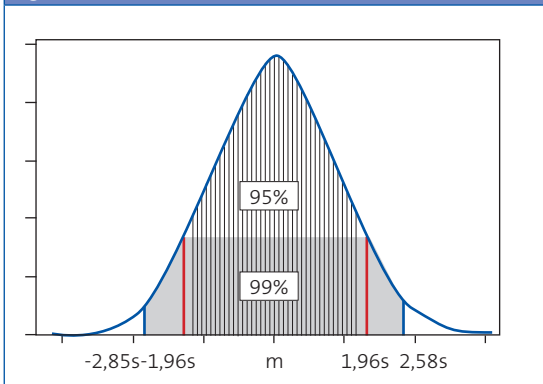
$$IC_{95\%} = m \pm 1,96 \frac{s}{\sqrt{n}}$$

Así que nos olvidamos de seleccionar múltiples muestras. Tomamos, por ejemplo, la primera de las diez, con *m*=236 y *s*=41 y calculamos su IC del 95%:

$$IC_{95\%} = 236 \pm 1,96 \frac{41}{\sqrt{100}} = \text{de } 228 \text{ a } 244 \text{ mg/dl}$$

¿Qué utilidad tiene este IC del 95%? Pues, dicho de una forma simple, es el que nos permite estimar entre qué valores está el valor inaccesible real de la población a partir del que podemos obtener de nuestra muestra, con una probabilidad de equivocarnos del 5%. Si queremos afinar más, podemos calcular el IC del 99% multiplicando el *ee* por 2,58 (en una distribución normal el 99% de la población se encuentra en el intervalo $m \pm 2,58 s$) y tendremos una probabilidad de error del 1%. Además, el IC nos

Figura 2. Distribución normal



informa sobre la precisión de nuestra estimación, ya que no será lo mismo estimar que el valor medio del colesterol sérico de la población sea de 228 a 244 mg/dl que de 200 a 270 mg/dl.

Aunque la fórmula del IC del 95% es siempre la misma ($m \pm 1,96 ee$), el cálculo del ee no es siempre tan sencillo como en el caso de la media. Con algunos estadísticos, la fórmula puede ser bastante complicada, pero existen programas informáticos y calculadoras en Internet, como las de la red CASPe (www.redcaspe.org/drupal/?q=node/30) o las de Infodoctor (www.infodoctor.org/rafabravo/herramientas.htm), que permiten calcular los IC con facilidad.

Imaginemos que, ahora que ya sabemos cuál es el colesterol medio de nuestros vecinos, queremos compararlo con el de otra ciudad (B), para ver si existen diferencias. Repetimos el proceso y obtenemos un valor medio de 242 mg/dl. ¿Quiere esto decir que en la ciudad B tienen el colesterol más alto? Pues, en realidad, no lo sabemos, porque lo único que tenemos hasta ahora es el resultado de una muestra de la ciudad, por lo que las diferencias que hemos observado pueden deberse al mero azar. Así que calculamos su IC del 95%, que nos permite estimar que el valor medio de colesterol sérico en B está entre 235 y 251 mg/dl. Y ahora que tenemos las estimaciones de las dos poblaciones procedemos a comparar sus IC.

Si los IC no se solapan querrá decir que el mínimo valor de B será siempre más alto que el máximo de

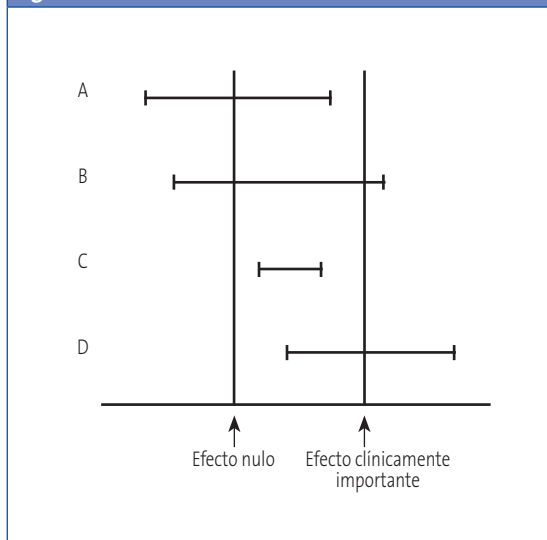
nuestra ciudad, con lo que podremos concluir que los habitantes de B tienen el colesterol más alto que nosotros, siempre con una probabilidad de error del 5%. Pero, en nuestro ejemplo, los intervalos se solapan, lo que quiere decir que no podemos descartar, con una probabilidad de error menor del 5%, que las diferencias se deban al azar, con lo que asumiremos que las dos ciudades tienen de media un valor de colesterol semejante.

Otra forma de enfocar el problema habría sido utilizar el teorema central del límite de forma inversa y calcular la probabilidad de que la diferencia se deba al azar. Si esta probabilidad es menor del 5% ($p < 0,05$) la rechazaríamos y concluiríamos que la diferencia es real, mientras que si $p > 0,05$ no podríamos descartar que la diferencia observada fuese fruto de la casualidad por proceder de muestras aleatorias y no de las poblaciones globales.

Pero el valor de p , mucho más utilizado que los IC, tiene algunos defectos desde el punto de vista práctico. Si nos fijamos en la fórmula del error estándar podemos ver que el tamaño de la muestra (n) está en el denominador dentro de una raíz cuadrada. Esto quiere decir que la precisión del estudio (la amplitud del IC) varía de forma inversamente proporcional con el cuadrado del tamaño de la muestra o, dicho de forma más sencilla, que para aumentar la precisión diez veces necesitamos una muestra 100 veces más grande. Por esto, el valor de la p depende del tamaño de la muestra, por lo que, en muchas ocasiones, obtener una p estadísticamente significativa es cuestión de aumentar lo suficiente el tamaño muestral.

Esto nos puede llevar a que pequeñas diferencias, que pueden ser irrelevantes desde el punto de vista clínico, alcancen significación estadística simplemente porque el tamaño muestral es elevado. Y, en sentido contrario, podemos despreciar diferencias que pueden ser clínicamente importantes por no ser estadísticamente significativas. Para comprenderlo mejor, veámoslo con un ejemplo.

Supongamos que queremos comparar el efecto de un nuevo fármaco broncodilatador con el del tratamiento actual, para lo cual medimos la diferencia de efecto de los dos sobre las pruebas de función

Figura 3. Valoración de los intervalos de confianza

respiratoria, asumiendo como clínicamente importantes diferencias del efecto superiores al 10%. Consideremos los cuatro estudios de la **Fig. 3**.

En el gráfico se representan los IC del 95% de los cuatro estudios, así como las líneas de efecto nulo (no diferencias entre los dos tratamientos) y del efecto clínicamente importante. Solo los estudios cuyo IC del 95% no cruce la vertical del efecto nulo tendrán significación estadística ($p < 0,05$).

El estudio A no tiene significación estadística (el IC incluye el valor nulo) y, además, clínicamente no parece importante.

El estudio B tampoco es estadísticamente significativo, pero clínicamente podría ser importante, ya que el límite superior del intervalo cae en la zona de relevancia clínica. Si aumentásemos la precisión del estudio (aumentando el tamaño de la muestra), es posible que el intervalo pudiese estrecharse y quedar por encima del nivel nulo, alcanzando significación estadística. Pensemos por un momento

que estuviésemos realizando un estudio de toxicidad y medimos mortalidad. Aunque la diferencia no fuese estadísticamente significativa, el estudio del IC del 95% probablemente desaconsejaría el uso del fármaco hasta disponer de estudios más precisos.

Los estudios C y D alcanzan significación estadística, pero solo los resultados del D son clínicamente importantes. El estudio C mostraría una diferencia, pero su impacto clínico y, por tanto, su interés son mínimos. Si el nuevo fármaco es más caro o tiene un perfil de toxicidad superior al fármaco actual, probablemente no sea conveniente indicar su uso, a pesar de que se muestre superior al tratamiento estándar desde el punto de vista de significación estadística.

Vemos, pues, cómo un enfoque más simple que considere solo el valor de la p puede llevarnos a despreciar estudios que pueden proporcionar información importante desde el punto de vista clínico. Y al contrario, podemos dar como bueno un ensayo con $p < 0,05$ aunque sus resultados sean clínicamente poco relevantes.

Por este motivo es recomendable favorecer el uso de los IC frente a la cifra de significación estadística, ya que los intervalos aportan información adicional sobre precisión de la estimación de los resultados y sobre el posible impacto clínico de los mismos.

CONFLICTO DE INTERESES

El autor declara no presentar conflictos de intereses en relación con la preparación y publicación de este artículo.

ABREVIATURAS

ee: error estándar de la media • IC: intervalo de confianza
• m: media • s: desviación típica.