

Variabilidad de las estimaciones o prueba de significación, ¿qué y cuándo?

Dr. Pablo Durán*

¿Con qué grado de precisión se realizan las estimaciones a partir de muestras?

El fin último de todo estudio de investigación, y por tanto del tratamiento de los datos, es alcanzar estimaciones precisas.

El grado de precisión de las estimaciones guarda una relación inversa con el error aleatorio. Este tipo de error determina la variabilidad entre la estimación que se intenta alcanzar (por ejemplo: media poblacional o μ) y la estimación realmente alcanzada (media muestral o \bar{X}), y que es fundamentalmente función del tamaño muestral y de las características del estimador, como la varianza.

En las comunicaciones científicas, frecuentemente, los datos continuos se resumen expresando el valor de la media y su desvío estándar o la comparación de una variable entre dos subgrupos mediante el contraste de sus medias y respectivos desvíos estándar o el valor de riesgo relativo de un grupo en comparación con su grupo control.

Sin embargo, ninguna de estas medidas da cuenta de la precisión con que fueron estimados y no es frecuente la inclusión de datos estadísticos que permitan valorarla.

Una muestra seleccionada aleatoriamente a partir de una población permite estimar el valor de la media poblacional identificada como μ a partir de la media obtenida de los datos de la muestra \bar{X} . Si se obtuvieran diferentes muestras a partir de una población, las medias estimadas en cada una de ellas diferirían entre sí. Uno de los teoremas más importantes de la matemática y, sin duda, el más importante de la estadística es el denominado teorema del límite central.¹ Éste enuncia que las medias de las muestras tomadas de un universo normal se distribuyen

normalmente. Esta afirmación implica que sucesivas muestras brindarán valores de \bar{X} diferentes para un mismo valor de μ .

El parámetro estadístico que permite valorar esta variabilidad, debida –como se mencionó– al error estándar, es el intervalo de confianza. En términos generales, podríamos decir que un intervalo de confianza de 95% implica que si se realiza el mismo estudio bajo las mismas condiciones, por cada 100 muestras aleatorias, en 95 de ellas la media se ubicaría entre los límites inferior y superior del intervalo.

Dicho de un modo más técnico, si se considera que en toda distribución normal la proporción de los valores entre 2 y -2 desvíos estándar es de 95%, los límites del intervalo de confianza de 95% corresponden a 2 y -2 desvíos estándar con respecto al valor medio de la distribución.

¿Cómo varían las estimaciones a partir de diferentes muestras de una misma población?

Para ejemplificar lo anterior, a partir de una serie de 600 valores de índice de masa corporal (IMC) se simuló los datos correspondientes a una población con valor μ de 27,3 y σ 5. Posteriormente se asignó aleatoriamente a cada registro un valor entre 1 y 10, de modo que a partir de los 600 registros (universo) se seleccionaron 10 muestras aleatorias de tamaño variable, según se observa en la *Tabla 1*. Esta tabla muestra cómo el valor de μ se ubica en el punto medio en relación con los valores de \bar{X} de cada una de las muestras simuladas. Se observa que las muestras seleccionadas brindan valores de \bar{X} que varían entre 26,6 y 27,9. Al mismo tiempo, el intervalo de confianza de cada una de las 10 muestras incluye la media poblacional de 27,3. De este modo, independientemente del valor de la media estimada a partir de cada muestra y de su

* Comité Editorial de Archivos Argentinos de Pediatría.

diferencia con respecto a la media μ de la población, en todos los casos ésta se encuentra incluida en todos los intervalos.

Algo similar sucede al comparar las medidas de dos subpoblaciones. Es común observar en estos casos la presentación de cada una de las medias, su desvío estándar y el valor de significación de las diferencias.

Sin embargo, no es frecuente en las publicaciones observar la presentación del valor medio de las diferencias entre las medias, junto con el intervalo de confianza correspondiente.

A partir de la muestra 9 del ejemplo anterior se comparó el valor de IMC según el sexo. Se estimó un valor medio de 29,2 kg/m² en varones y de 26,5 en mujeres (desvío estándar 4,2 y 4,0 respectivamente). Tal diferencia parece relevante desde el punto de vista de su valor, aun cuando su significación estadística se encuentra en el límite (t 1,96; p 0,05). Ante resultados similares es común que algunos autores planteen que si bien no se observó una diferencia estadísticamente significativa, parecería existir una tendencia hacia la observación de valores más elevados de IMC en varones. Sin embargo, al analizar la media e IC 95% de la diferencia, se observa que aquella es de 2,6 (5,3 a -0,5). Tal intervalo de confianza, que incluye el 0, indica que en 95 de cada 100 estudios de características similares, en algunos casos presentarán mayor IMC los varones, en tanto que en otros lo presentarán las mujeres. Esta observación invalida lo anteriormente especulado en términos de la mayor tendencia a valores más elevados de IMC en varones.

Del mismo modo, la estimación de medidas de riesgo (riesgo relativo o razón de probabilidad [*odds ratio*]) requiere de la consideración del intervalo de confianza correspondiente. Sin embargo, en este caso es la inclusión del valor 1 y no 0 lo que indica la ausencia de una diferencia.

¿Cómo expresar la variabilidad de las estimaciones?

Diferentes autores han planteado visiones contrapuestas con respecto a la importancia del valor de significación y el intervalo de confianza. La discusión sobre la importancia de valorar el error aleatorio en las estimaciones se ha dado fundamentalmente

en los últimos 20 años.² La discusión se centró principalmente en la diferencia e importancia de la valoración de pruebas de significación contra la consideración de la variabilidad a partir de la estimación de intervalos de confianza. Para algunos autores,³ la convención de considerar pruebas de significación y valor de p no brinda información del todo relevante y puede conducir a una inadecuada interpretación de los resultados, por no dar cuenta de la variabilidad de las estimaciones. Por el contrario, Fleiss^{4,6} valora la importancia de las pruebas de significación al considerar que la ciencia implica la toma de decisión y que las convenciones en términos de las pruebas de significación (un valor de p inferior a -0,05 o 0,001 permite rechazar la hipótesis nula) facilitan la toma de decisión, sin dejar de considerar posibles limitaciones y potenciales abusos. Otros autores, como Walker,⁷ refutan los argumentos de Fleiss por considerar que el intervalo de confianza brinda por separado información sobre la magnitud de la estimación, independientemente de su precisión. En cambio, el valor de significación no permite tal diferenciación. Poole, por otra parte, presenta otros puntos de vista. Considera que muchas veces los intervalos de confianza se utilizan inadecuadamente como pruebas de significación estadística al valorarse en términos de

TABLA 1. Comparación entre medias, desvíos estándar e intervalos de confianza obtenidos a partir de muestras aleatorias y la población total en una serie de datos de IMC simulados

Muestra	N casos	Media	Intervalo de confianza 95%		Desvío estándar
			Límite inferior	Límite superior	
1	37	26,6	25,1	28,1	4,4
2	61	26,8	25,4	28,2	5,4
5	65	27,0	25,7	28,3	5,2
9	60	2,0	25,9	28,1	4,1
4	71	27,3	26,2	28,4	4,8
Población total	600	27,3	26,9	27,7	5,0
6	64	27,4	25,9	28,8	5,6
3	62	27,5	26,4	28,6	4,3
7	71	27,6	26,3	28,9	5,5
8	80	27,6	26,4	28,8	5,3
10	31	27,9	25,8	29,9	5,6

la inclusión del valor 0 o 1 en la comparación de medias o la estimación de riesgo, respectivamente. Pero al mismo tiempo considera una inadecuada interpretación del intervalo de confianza cuando se toman ambos límites como límite de inclusión, para diferenciar los valores que quedan incluidos o excluidos en el intervalo.

Sobre la base de lo planteado, cabe remarcar que el intervalo de confianza, tanto en la estimación de una media poblacional, en la comparación de medias, como en la estimación de riesgo, permite valorar la variabilidad de las estimaciones. Las pruebas de significación para valorar hipótesis se basan en convenciones estadísticas orientadas a la toma de decisión y, aun en tal caso, la valoración del intervalo de confianza constituye una información invaluable que permite relativizar las conclusiones alcanzadas me-

dante las pruebas de significación pero, por sobre todo, dar cuenta de la variabilidad de las estimaciones. ■

BIBLIOGRAFÍA

1. Snedecor GW, Cochran WG. *Statistical Methods*. 8th ed. Ames, Iowa: Iowa State University Press, 1989.
2. Poole C. Beyond the confidence interval. *Am J Public Health* 1987; 77:195-199.
3. Gardner MJ, Altman DG. Confidence interval rather than p values: estimation rather than hypothesis testing. *BMJ* 1986; 292:746-750.
4. Fleiss JL. Significance test have a role in epidemiologic research: reactions to AM Walker (Different views). *Am J Public Health* 1986; 76:559-560.
5. Fleiss JL. Confidence intervals vs significance test: quantitative interpretation. *Am J Public Health* 1986; 76:587. [Letter].
6. Fleiss JL. Dr. Fleiss responds. *Am J Public Health* 1986; 76:1033-1044. [Letter].
7. Walker AM. Reporting the results of epidemiologic studies. *Am J Public Health (Different views)* 1986; 76:556-558.