¿Qué significa el error absoluto en muestreo poblacional?

FRANCISCA TAPIA ALARCÓN[1], NICOLÁS ESPINOZA RIVEROS[1], JOSÉ RUIZ-TAGLE MATURANA[1,2*]

[1]INFORMÁTICA BIOMÉDICA, ESCUELA DE SALUD, DUOC UC
[2]NÚCLEO MILENIO PARA LA EVALUACIÓN Y ANÁLISIS DE POLÍTICA DE DROGAS (NDP)

Introducción

En el contexto de estimar un parámetro a través de intervalos de confianza, una estrategia útil es definir *a priori* el tamaño muestral necesario para obtener una precisión deseada. Existen diferentes estrategias en la literatura que permiten hacer esto, sin embargo, poco se discute respecto a qué significa en la práctica esta "precisión deseada". El objetivo de este artículo es discutir brevemente a qué se refiere el error absoluto y cómo podemos implementarlo en una investigación adecuadamente, y a pesar de que existen ciertas "fórmulas mágicas" que se enseñan en algunas carreras del área de salud, es fundamental entenderlas para tomar decisiones metodológicas que afectarán directamente a nuestros resultados.

Es frecuente observar en la investigación sanitaria tamaños muestrales obtenidos mediante la siguiente fórmula:

$$n = \frac{z^2_{1-\alpha/2} P(1-P)}{E_0^2} \tag{1}$$

La fórmula (1) busca establecer un tamaño de muestra adecuado para estimar la proporción poblacional (P), algo útil en estudios transversales cuyo interés es describir una prevalencia[1]. Esta fórmula está compuesta por tres elementos: el valor crítico de la distribución normal estándar para el cuantil $1-\alpha/2$, $z_{1-\alpha/2}$, cuyo valor es 1,96 cuando se trabaja a un 95 % de confianza (nivel de confianza más frecuente, donde el nivel de confianza es $1-\alpha$); un valor hipotético para la proporción poblacional P que deseamos estimar, generalmente obtenida a través de un estudio piloto o estudios previos; el error absoluto E_0, o error máximo admisible [1] o margen de error [2], que en la práctica corresponde al valor más alto que estamos dispuestos a tolerar para el error de muestreo esperado, es decir, la diferencia entre lo que informa la muestra (estimación) y lo que realmente sucede en la población (parámetro) [3].

El uso inadecuado que se le da a la fórmula (1), al realizar test de hipótesis o ajustar modelos de regresión, está fuera del objetivo de este trabajo, pero la explicación es bastante sencilla y es que en ningún momento se está considerando la magnitud del efecto que buscamos ni la probabilidad de cometer un error de tipo II [4]. Es importante mencionar que, para que esta fórmula tenga respaldo de la teoría estadística, es decir, para poder extrapolar los resultados hacia la población con relativa seguridad, es necesario que las unidades de análisis sean estadísticamente independientes, lo cual se favorece cuando la muestra es extraída de manera aleatoria. Otro detalle importante radica en que (1) está diseñada cuando se espera, se sabe o se estima que el tamaño de la población es muy grande con respecto al tamaño de la muestra (población infinita), y en caso contrario es necesario hacer un pequeño ajuste al tamaño de muestra. Este ajuste es particularmente útil cuando nuestros recursos son limitados (99,9 % de las veces), ya que el tamaño de muestra requerido es menor al que obtendríamos utilizando la fórmula convencional de población infinita. No obstante, la justificación teórica radica en que, en poblaciones finitas, el hecho de que la población sea limitada significa que la selección de un individuo afecta la probabilidad de selección de los demás. Esto reduce la variabilidad de la muestra en comparación con lo que se esperaría en una población infinita. La corrección de población finita ajusta el error estándar para tener en cuenta esta reducción en la variabilidad, por lo que cumple con estimaciones más precisas en muestras relativamente grandes respecto al tamaño de la población y hace más imprecisas a aquellas con muestras pequeñas [1].

En la práctica, la elección del error absoluto afecta en la cantidad de unidades de muestreo que necesitamos para nuestra investigación. En este caso en particular, al tratarse de un muestreo aleatorio simple, las unidades de muestreo coinciden con las unidades de análisis. Se puede observar en la Figura 1 que mientras menor sea el error absoluto, mayor será el tamaño de muestra requerido. Esto tiene sentido, dado que un menor error absoluto supone un intervalo de confianza más angosto y por tanto una estimación más precisa. Nótese que para obtener una reducción del 50 % en el error absoluto, se necesita un aumento de cuatro veces el tamaño muestral (para $P = 0{,}5$ y $E_0 = 0{,}02$ se requieren $n = 2401$, mientras que para $E_0 = 0{,}01$ se requieren $n = 9604$).

*j.ruiztagle@profesor.duoc.cl

[1]Nótese que el resultado de aplicar (1) tendrá que ser aproximado a un número entero, por ejemplo, redondeándolo.

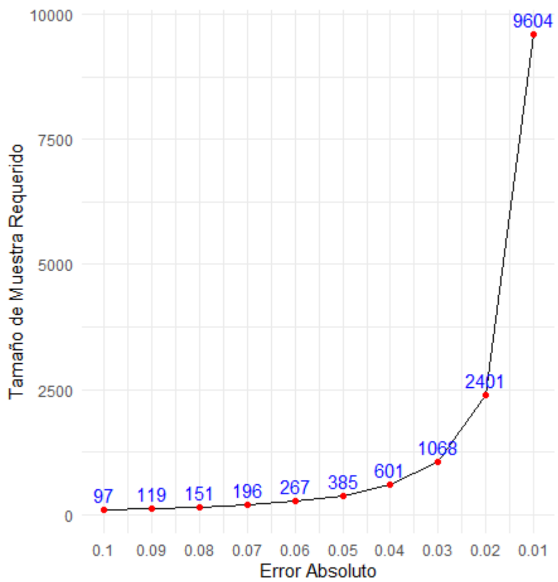

Figura 1: Relación entre error absoluto y tamaño de muestra requerido según fórmula (1).

¿Qué es el error absoluto?

La fórmula (1) se puede modificar para escribirla en términos de la amplitud del intervalo ($A_0 = 2E_0$). Por ejemplo, si quisiéramos obtener el tamaño de muestra mínimo para que la amplitud del intervalo sea de 0,10 (10 %, o un error absoluto $E_0 = 0{,}05$) podríamos utilizar:

$$n = \frac{4z^2_{1-\alpha/2}P(1-P)}{A_0^2} = \frac{4 \cdot 1{,}96^2 \cdot 0{,}5 \cdot 0{,}5}{0{,}1^2} = 384 \tag{2}$$

La fórmula (2) indica que para obtener un intervalo de confianza con un nivel de confianza del 95 % y asumiendo una proporción de 0,5 ("máxima varianza"[2]), necesitaríamos 384 unidades de análisis para que el intervalo tenga una amplitud esperada del 10 %. Esto es equivalente a establecer un error absoluto igual a 0,05, lo que significa que la distancia esperada entre la proporción estimada y los límites del intervalo será de 5 % (se multiplica $\times 100$), por ende, la amplitud del intervalo debería ser el doble del error absoluto. Un error del 5 % es común, y se recomienda en la literatura [5]. Sin embargo, es relevante entender que no aplica para todos los casos [6, 7]. Esto se puede esclarecer aún más cuando se observa a través de un ejemplo. Es necesario recordar que el intervalo de confianza para la proporción se define de la siguiente manera:

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{3}$$

donde $\hat{p}$ es una estimación puntual de la proporción poblacional obtenida con la información de la muestra.

Ejemplos

Supongamos que un equipo de investigación desea estimar la prevalencia de consumo de tabaco en una localidad rural mediante el uso de un cuestionario. Utilizaron la fórmula (2) para determinar que, para estimar la prevalencia de consumo de tabaco, debían entrevistar a 384 personas. Por simplicidad asumiremos que disponían de un listado completo de las unidades que componen la población objetivo, por ende, pudieron realizar un muestreo aleatorio simple. También asumiremos que el tamaño de la muestra es muy pequeño con respecto a la población ("población infinita"). Luego de aplicar el cuestionario, encontraron que la prevalencia de consumo de tabaco en la muestra fue de 50 %, que coincide con el valor que se supuso para P en (2). Aplicando la fórmula (3), obtenemos:

$$0{,}5 \pm 1{,}96\sqrt{\frac{0{,}5(1-0{,}5)}{384}} \tag{4}$$

A partir de (4) podemos estimar con un 95 % de confianza que la prevalencia de consumo de tabaco en la localidad rural se encuentra, aproximadamente, entre 0,45 y 0,55. Este sería un escenario ideal, es decir, el valor de la proporción poblacional definido *a priori* coincide con el valor estimado. En un escenario más realista, puede que luego de aplicar el cuestionario, el equipo de investigación encuentre que la prevalencia de consumo de tabaco fue aproximadamente de 23 %. Claramente, un número muy diferente al que supusieron para P (50 %). Aplicando la fórmula (3), obtenemos:

$$0{,}23 \pm 1{,}96\sqrt{\frac{0{,}23(1-0{,}23)}{384}} \tag{5}$$

En este ejemplo, la proporción de consumo de tabaco en la localidad rural se encuentra entre 0,19 y 0,27. La distancia entre la estimación puntual y el límite del intervalo es menor que la definida *a priori* en (2), el error de estimación fue de un 4 %, ¿por qué ocurre esto? Porque evidentemente el error de estimación es conceptualmente diferente al error absoluto (margen de error, error máximo admisible). Debemos entender el error absoluto como la distancia máxima que se acepta entre la estimación puntual y los límites del intervalo, lo que es equivalente a decir que el ancho del intervalo no debería superar al doble del error absoluto. Sin embargo, no está asegurado que la amplitud del intervalo no sea mayor que la definida previamente mediante el error absoluto que el investigador estuvo dispuesto a aceptar.

Realizamos una simulación con el objetivo de mostrar que no hay garantía de que el intervalo de confianza tendrá una amplitud menor al doble del error absoluto. Para esto, generamos muestras aleatorias con

[2] Suponer que $P = 0{,}5$ resulta en que la varianza $P(1-P)$ tenga su valor más alto.

proporción 0,1 hasta 0,9 con su respectivo tamaño de muestra. Obtuvimos 10.000 simulaciones para cada proporción estimada, manteniendo constante el error absoluto (5 %) y el nivel de confianza (95 %). Todo el detalle del código se encuentra alojado en GitHub (bit.ly/3RZszxo).

En la Figura 2 se puede observar que cuando la proporción estimada es 0,5, no hay ningún intervalo de confianza que sea más ancho que el que se define con el error absoluto. Sin embargo, para valores diferentes a 0,5, en algunos casos más del 50 % de los intervalos tiene una amplitud mayor a la que se desea.

Ocurre algo similar en la Figura 3, donde se observa el comportamiento aleatorio que tiene la amplitud del intervalo. El único escenario en donde la amplitud del intervalo no supera el doble del error absoluto es cuando se obtiene un tamaño de muestra utilizando máxima varianza.

El estudio de Kupper y Hafner [8] examina la adecuación de las fórmulas de tamaño de muestra populares. Las fórmulas de tamaño de muestra basadas en la amplitud de los intervalos de confianza tienden a subestimar el tamaño de la muestra necesario cuando las proporciones son pequeñas o grandes (es decir, diferentes de 0,5). Esto se debe a que la aproximación normal utilizada en estas fórmulas se vuelve menos precisa para proporciones que se alejan de 0,5. Las fórmulas habituales no consideran que los intervalos que se calcularán con ellas serán de amplitud aleatoria y por tanto podrán ser más angostos o extensos que el error absoluto.

Cuando la proporción es exactamente 0,5, el tamaño de muestra calculado es generalmente adecuado porque la varianza de la distribución binomial es máxima y la aproximación normal es más precisa. Para proporciones que no son 0,5, las fórmulas pueden no ser conservadoras, lo que podría llevar a que la amplitud real del intervalo de confianza exceda el error absoluto deseado más a menudo de lo previsto.

Cuando calculamos intervalos de confianza para estimar proporciones, es importante recordar dos cosas. Primero, estos intervalos pueden variar de una muestra a otra debido a la aleatoriedad en la selección de la muestra. Esto significa que la amplitud de los intervalos de confianza no es siempre la misma. Segundo, si la proporción real en la población se aleja mucho del 50 %, la forma en que se distribuyen nuestros estimadores de proporción cambia. Deja de ser normal, y eso afecta la precisión de nuestros métodos estándar que asumen una distribución normal. Por ejemplo, los intervalos de confianza podrían no ser simétricos, lo que afecta cómo interpretamos los resultados.

Imagina que estás midiendo qué tan común es una característica en un grupo grande de personas, como la preferencia por un cierto tipo de música. No puedes preguntarles a todos, así que eliges un pequeño grupo al azar para tener una idea general. Aquí hay dos cosas importantes: primero, la estimación que haces puede variar cada vez que eliges un grupo diferente. Segundo, cuando la característica que estás midiendo es, o muy común, o muy rara (por ejemplo, si casi todos aman u odian un tipo de música), las estimaciones se vuelven menos confiables. Esto es porque asumimos que nuestras medidas siguen un patrón común, pero esto cambia cuando nos alejamos de situaciones donde las preferencias están más equilibradas.

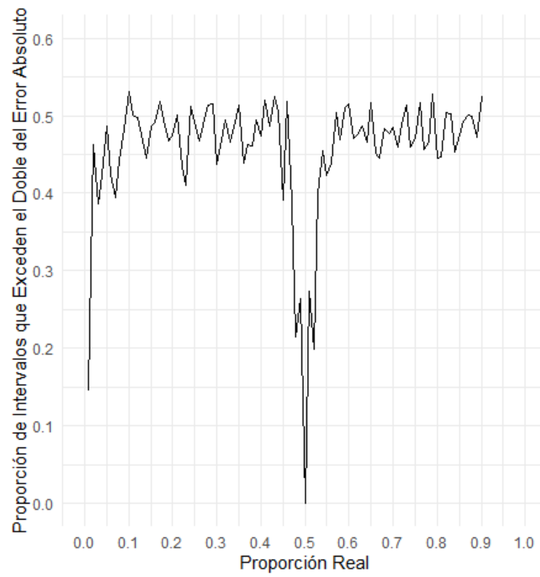

Figura 2: Proporción de intervalos que exceden al doble del error absoluto en función de P real ($E_0 = 0{,}05$, $1 - \alpha = 0{,}95$).

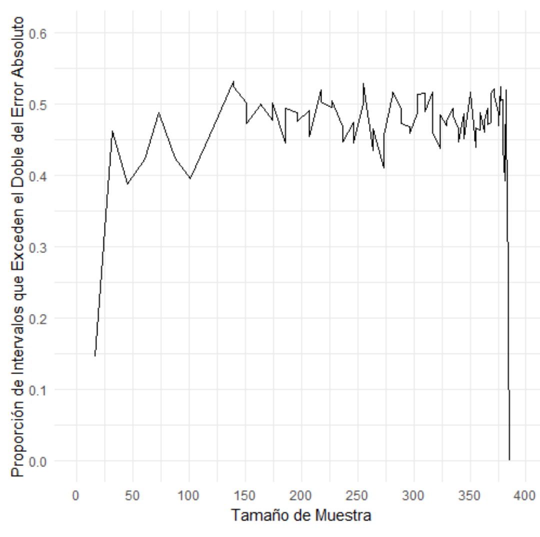

Figura 3: Proporción de intervalos que exceden al doble del error absoluto según tamaño de muestra requerido ($E_0 = 0{,}05$, $1 - \alpha = 0{,}95$).

Esto es realmente problemático porque nunca podremos saber si esto nos está ocurriendo a la hora de realizar el muestreo. Sin embargo, ¿esto necesariamente indica un error? Es importante tener en cuenta que la propia estimación de P inevitablemente estará desviada de la cifra real. En la práctica, no contamos con el valor del parámetro, ni antes ni después de realizar el muestreo. Existe una posibilidad real de que nuestro intervalo no incluya al parámetro, que de manera nominal corresponde al complemento del nivel de confianza. Esto es una consecuencia inherente de la construcción del método de estimación, que en su naturaleza implica un nivel de incertidumbre. Es crucial entender esto dentro del marco estadístico y tenerlo en cuenta a la hora de interpretar los resultados.

Conclusión

El error absoluto en el cálculo del tamaño muestral es una estimación basada en suposiciones teóricas y sirve como una guía para obtener un tamaño de muestra que, en promedio y bajo ciertas condiciones, debería proporcionar la precisión deseada. Sin embargo, en la práctica, debido a la variabilidad inherente al muestreo y a las limitaciones de las aproximaciones estadísticas, no todas las muestras individuales cumplirán exactamente con este margen de error. Esto no invalida el método, pero resalta la importancia de comprender sus limitaciones y de interpretar los resultados con un entendimiento claro de la variabilidad y la incertidumbre en los datos y los métodos estadísticos.

El error absoluto nos permite controlar la amplitud esperada del intervalo de confianza. Mientras menor sea el error absoluto, mayor será el tamaño muestral requerido y menor será el ancho esperado del intervalo. Por el contrario, mientras mayor sea el error absoluto, menor tamaño de muestra será necesario y mayor será el ancho esperado del intervalo. Mientras más estrecho sea el intervalo, más precisión tendremos en torno a las estimaciones.

Si bien es una medida importante para controlar la precisión de las estimaciones mediante los intervalos de confianza, debemos reconocer sus limitaciones y entender que no es recomendable usarlo en todos los casos, sobre todo cuando se trata de enfermedades o condiciones poco frecuentes en la población (Silva, 2000). Ante este caso es recomendable optar por el error relativo, que razonablemente, se puede predeterminar sin conocer el valor del parámetro que se desea estimar [9].

Otro aspecto importante para considerar, sobre todo ante muestras pequeñas, es que las fórmulas que comúnmente se utilizan para calcular intervalos de confianza no consideran la amplitud aleatoria de las muestras, lo que puede resultar en intervalos más angostos o más extensos que el error absoluto definido, por lo que requieren una corrección para que realmente el ancho del intervalo sea el que deseamos [8]. La recomendación es siempre intentar conocer un valor aproximado del parámetro antes de realizar el muestreo, al definir el margen de error, éste debe tener un sentido teórico, el cual encontramos al pensar en el ancho que tendría el intervalo de confianza. Evidentemente, si nos interesa estimar la proporción de una enfermedad poco prevalente, no deberíamos considerar un ancho del intervalo muy amplio, porque podríamos obtener resultados que no aportan nada en lo práctico debido a su falta de precisión.

Agradecimientos

Agradecemos al Dr. Felipe Medina Marín por su generosa contribución al desarrollo de este artículo.

Referencias

[1] Vivanco, Manuel (2005). *Muestreo Estadistico: Diseño y Aplicaciones*. Editorial Universitaria.

[2] Martínez-González, M.A., Villegas, A., & Faujín-Fajardo, F.J. (2020). *Bioestadística amigable*, 4ª. Elsevier.

[3] Taucher, Erica (1997). *Bioestadística*. Editorial Universitaria.

[4] García-García, J.A., Reding-Bernal, A., & López-Alvarenga, J.C. (2013). Cálculo del tamaño de la muestra en investigación en educación médica. *Investigación en Educación Médica*, *2*(8), 217–224. doi.org/10.1016/S2007-5057(13)72715-7

[5] Bartlett, J.E., Kotrlik, J.W., & Higgins, C.C. (2001). Organizational Research: Determining Appropriate Sample Size in Survey Research. *Information Technology, Learning, and Performance Journal*, *19*(1), 43-50.

[6] Silva, L.C. (2000). Nueva visita al supuesto de máxima indeterminación y al empleo de errores absolutos y relativos. *Gaceta Sanitaria*, *14*(3), 254–257.

[7] Suárez-Gil, P., & Alonso, J.C. (1999). Sobre el supuesto de máxima indeterminación, el tamaño muestral y otras consideraciones sobre muestreo. *Gaceta Sanitaria*, *13*(3), 243–246. https://www.gacetasanitaria.org/es-pdf-S0213911199713610

[8] Kupper, L.L., & Hafner, K.B. (1989). How appropriate are popular sample size formulas? *American Statistician*, *43*(2), 101–105. https://doi.org/10.1080/00031305.1989.10475628

[9] Marrugat, J., Vila, J., & Pavesi, M. (1999). Supuesto de máxima indeterminación: ¿error absoluto o error relativo en el cálculo del tamaño de la muestra? *Gaceta Sanitaria*, *13*(6), 491. doi.org/10.1016/S0213-9111(99)71416-0