

Nota estadística

Investigación, poblaciones y muestra

Jorge Camacho-Sandoval

Usualmente, los investigadores en ciencias de la salud están interesados en una población, formada por un conjunto de individuos, con uno o más atributos comunes. Por ejemplo, la población de pacientes de *Diabetes Mellitus* no insulino dependientes, de ambos sexos, con edades comprendidas entre 18 y 60 años, residentes en el área de atracción del Hospital México. Más específicamente, los investigadores suelen estar interesados en determinadas características o **parámetros** de esa población, como el nivel medio de hemoglobina glicosilada, el nivel medio de glicemia preprandial, la proporción de individuos que utilizan insulina, etcétera.

Aún cuando interesan los valores de ciertas características en la población, o **parámetros**, generalmente es imposible estudiar la población en su conjunto. Esa imposibilidad surge por razones de costo, tiempo o disponibilidad de recursos humanos y técnicos, porque la población es infinita o muy grande, no está bien delimitada en término de espacio y tiempo, o por otras razones. La solución es, entonces, estudiar una parte de la población o **muestra**, y luego generalizar a la población, las conclusiones obtenidas a partir del estudio de la muestra.

¿Qué implicaciones tiene utilizar muestras en vez de estudiar la población completa?

Una primera implicación es que no se puede conocer el valor de la característica de interés en la población o **parámetro**. Lo que se obtiene a partir de las mediciones que se toman en los individuos de la muestra, es un valor que suponemos similar o aproximado al parámetro, llamado **estimado**.

Surge otro problema: ¿Qué pasa si en vez de una, se toman varias muestras? La experiencia dice que probablemente el valor del estimado sea diferente en cada una de las muestras. Entonces, el estimado se comporta como una variable, es decir, su valor cambia en razón de los elementos que componen la muestra; como los individuos de la muestra se seleccionan al azar, el estimado se comporta como una variable aleatoria.

Aunque con frecuencia se utilice una sola muestra, se debe tener presente que esta es una de infinitas muestras posibles y, en consecuencia, se tiene un estimado de un infinito número de posibles estimados del parámetro de interés. ¿Es confiable entonces el estimado? El sentido común dice que será más confiable el estimado con menor variabilidad, aunque otras propiedades del estimado también son importantes. Eso hace necesario disponer de alguna medida de la variabilidad de los estimados.

El estimado es una variable aleatoria, y si hubiera muchas muestras y sus respectivos estimados, se podría calcular la varianza de los estimados y su raíz cuadrada, la desviación estándar del estimado, más conocida como **error estándar** o **error típico** del estimado. En la vida real hay una o pocas muestras, de modo que: ¿Cómo obtener el error típico del estimado? Se puede hacer a partir de la desviación estándar de los datos de la muestra de la cual proviene el estimado, ya que el error estándar del estimado es el resultado de dividir la desviación estándar de la muestra por la raíz cuadrada del tamaño de la muestra. Se puede ver un ejemplo en el recuadro.

Hay dos maneras de reducir el error estándar de un estimado: la primera consiste en incrementar el tamaño de la muestra, ya que el error estándar es inversamente proporcional a la raíz del tamaño de la muestra. De esta propiedad se puede deducir una regla general: **a mayor tamaño de muestra, menor variabilidad del estimado**.

Es evidente la importancia de disponer de un tamaño de muestra adecuado para conseguir estimados con menor variabilidad y, en consecuencia, conclusiones más confiables, por ello se dedicará una nota especial al tamaño de muestra y los factores que lo determinan.

Profesor, Maestría en Epidemiología, Postgrado en Ciencias Veterinarias, UNA.

Correspondencia:
Correo electrónico:
jcamacho@ice.co.cr

ISSN 0001-6002/2007/49/1/11-12
Acta Médica Costarricense, ©2007
Colegio de Médicos y Cirujanos

La otra forma de reducir el error estándar de un estimado es si se tiene menor variabilidad de los datos. Esto algunas veces es posible, por ejemplo, utilizando diseños de muestreo estratificados si en la población existen subpoblaciones o estratos más homogéneos para la variable de interés, por ejemplo: por sexo, edad u otro factor. Eso significa que además del tamaño de muestra, la forma como se selecciona la muestra influye en la precisión de los estimados.

Como se mencionó, los estimados obtenidos de distintas muestras probablemente serán diferentes, dado que las muestras estarán compuestas, al menos parcialmente, por distintos elementos seleccionados de forma aleatoria. Esto conduce a un problema básico de la investigación y la estadística: *¿cómo determinar si las diferencias que se observan entre 2 ó más grupos o muestras, para una variable determinada, se deben al azar o a otros factores que distinguen los grupos, como sexo, edad, etnia o tratamiento recibido?*

Dicho en otras palabras, cuando se comparan promedios, proporciones, prevalencias o cualquier otro estimado de 2 ó más grupos, lo normal y frecuente es que se noten diferencias. El problema está en determinar si las diferencias observadas se deben exclusivamente a que los grupos estén formados por distintos elementos, es decir, si obedecen al azar, o si existe efecto de otras características diferentes entre grupos, por ejemplo, algún tratamiento aplicado. Eso será tema de una próxima nota.

La varianza de la edad es:

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} = \frac{3404.4}{10 - 1} = 378.27 \text{ años}^2$$

La desviación estándar de la edad es:

$$S = \sqrt{378.27} = 19.45 \text{ años.}$$

La edad promedio es:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{346}{10} = 34.6 \text{ años}$$

El *error estándar del promedio* es:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{19.45}{\sqrt{10}} = 6.15 \text{ años.}$$

Figura 1. Fórmulas estadísticas correspondientes.

Referencias

1. Azzimonti-Renzo, J. C. 2003. Bioestadística aplicada a bioquímica y farmacia. 2ª Edición. Editorial Universitaria, Argentina.

Cuadro 1. Edad promedio de pacientes de la enfermedad de Castleman en casos reportados en el Hospital Calderón Guardia 1990-2002*		
Caso	Edad (x _i)	(x _i - \bar{x}) ²
1	31	12.96
2	15	384.16
3	13	466.56
4	17	309.76
5	22	158.76
6	72	1398.76
7	56	457.96
8	48	179.56
9	32	6.76
10	40	29.16
Suma	346	3404.4
Promedio (\bar{x})	34.6	

* Fuente: Fiedler-Velásquez, E.; Gourzong-Taylor, C.; Goyenaga-Hernández, P. & López-Castro, J. 2005. Enfermedad de Castleman. Reporte de 14 casos de 1990-2002 en el "Hospital Dr. Rafael Ángel Calderón Guardia", dos con presentación inusual y revisión de bibliografía. *Acta Médica Costarricense* 47 (3): 137-143.