

Tamaño apropiado de muestra para obtención de conclusiones válidas en una investigación

Por:

Francisco Javier Jaramillo Álvarez

Resumen

La necesidad de un enfoque estadístico es reconocida en la investigación; pero suele suceder que los investigadores lleguen a conclusiones mal fundamentadas porque basan éstas en muestras de tamaño insuficiente; o, en el mejor de los casos, parten de una muestra representativa porque sustituyen valores en una fórmula que no saben ni qué significa.

Este artículo pretende describir de una manera sencilla, pero rigurosa, las implicaciones que tiene la selección de una muestra representativa al hacer una investigación en cualquier área y, de esta forma, expandir un poco el horizonte que combina el razonamiento estadístico con el razonamiento gerencial e investigativo en el momento de obtener conclusiones.

Palabras clave

Muestra, población, margen de error, nivel de confianza, inferencia estadística.

Abstract

The need of an statistical focus is recognized during research, but often researchers come up to conclusion without foundation due to a lack of samples or in the best of the cases, they use representative samples replacing figures which they have no idea what it means.

The purpose of this article is to describe on a simple, but strict way, the consequences of choosing a sample when conducting a research in any area, and this way expanding the merge of statistical reasoning with managerial and investigative reasoning when obtaining conclusions.

Keywords

Sample, population, sampling error, confidence level, statistical inference.

Introducción

Una parte trascendental de la estadística es la estadística inferencial. El objetivo básico de la estadística inferencial consiste en extraer información aplicable a toda una población con base en los resultados

obtenidos a partir de una muestra; las estimaciones que se hagan deben expresarse mediante un intervalo.

Constantemente vemos en revistas y periódicos afirmaciones como “el 68% de los colombianos apoya la labor del presidente” y generalmente a nosotros no nos han preguntado nada. Esos valores son obtenidos a partir de muestras y con base en conceptos estadísticos puede generalizarse para toda la población colombiana; pero para que ello tenga sustento, la muestra tiene que ser representativa.

Las estadísticas por sí mismas no tienen sentido si no se utilizan racionalmente y si no se relacionan dentro del contexto en el que se trabaja. Es necesario, como punto de partida, entender los conceptos de población y de muestra para lograr comprender mejor su significado en la investigación que se lleva a cabo.

Población

Es el conjunto total de individuos, objetos o medidas del cual queremos obtener conclusiones y que poseen algunas características comunes. Cuando se vaya a llevar a cabo alguna investigación es fundamental delimitar clara y cuidadosamente la población objetivo, pues de ello depende, en gran parte, que las conclusiones extraídas en la investigación sean válidas. Esto implica que deben especificarse muy bien las particularidades deseables y que se restrinjan tiempo y espacio.

McDaniel dice en su libro Investigación de Mercados “no hay reglas específicas que se puedan seguir para definir la población; lo que debe hacer el investigador es aplicar la lógica y el criterio para abordar el problema básico: ¿de quiénes son las opiniones que se necesitan con el fin de satisfacer los objetivos de la investigación?”

Si, por ejemplo, un investigador de mercados quiere hacer un estudio antes de lanzar determinado producto, debe empezar por preguntarse cuál es su público objetivo y, si por ejemplo, su público objetivo son los niños del país, en la población deben estar incluidos los niños de todas las regiones del país, de todos los estratos sociales, de ambos sexos...; pero si su público objetivo son los niños del Valle de Aburrá, la población está formada únicamente por los niños de esa región. No podría concluirse algo relacionado con todos los niños de Colombia, si sólo se encuestó a niños de la ciudad.

Muestra

Es un subconjunto fielmente representativo de la población; debe ser una porción característica, tanto en número como en calidad. Representativa en número implica que no puede ser muy pequeña en relación a la población; representativa en calidad implica que sí refleje las características de la población.

La selección de una muestra ahorra dinero, tiempo y esfuerzo. Además una muestra puede proporcionar tanta o más exactitud que el

correspondiente estudio que pretenda investigar la población completa porque existe un mayor control sobre los errores en la recolección.

Si la muestra no es representativa, los resultados sólo pueden ser aplicados a los elementos muestreados y no resultaría lógico generalizarlos: Por ejemplo, si se hace una encuesta a expertos en un determinado tema, no tendría sentido generalizar los resultados a una población que no sea ese grupo de expertos.

El tipo de muestreo (selección de la muestra) más apropiado estadísticamente es el aleatorio; éste implica la selección al azar y que cada miembro de la población tenga igual oportunidad de ser incluido en la muestra. Si una muestra es no aleatoria, puede contener tendencia; pero si es aleatoria y se hace correctamente, no tendrá tendencia y es por lo tanto relativamente representante de la población. Por supuesto, si se toma una muestra, nunca podemos estar totalmente seguros de que los resultados medidos a partir de la muestra sean también ciertos en la población. No obstante, para efectos prácticos suele ser suficiente si se puede afirmar que el riesgo de desviación no es grande. Tales afirmaciones sólo pueden hacerse si se ha usado una muestra aleatoria.

El muestreo aleatorio puede hacerse en forma estratificada; ello implica la división de la población en subgrupos según una característica específica; así se garantiza la representatividad en la muestra de todos los grupos. En ese caso, el tamaño elegido se distribuye proporcionalmente, de acuerdo al tamaño de cada estrato.

Eso no significa que muestreos no probabilísticos sean inadecuados en todos los casos. Todo depende del objetivo previsto.

Todo lo expuesto hasta el momento indica que un tamaño de muestra inadecuado conduce a un inevitable desperdicio y desaprovechamiento de recursos. Si es demasiado grande se desperdicia tiempo y dinero y, si es muy pequeña, las conclusiones resultantes no son muy confiables.

Para determinar el tamaño de muestra se debe tener en cuenta si la población es infinita (o muy grande o de un tamaño desconocido) o si es finita (si se puede establecer su tamaño). Siempre que sea posible, se debe determinar el tamaño de la población.

Si la población es infinita, el tamaño de muestra (n) viene dado por:

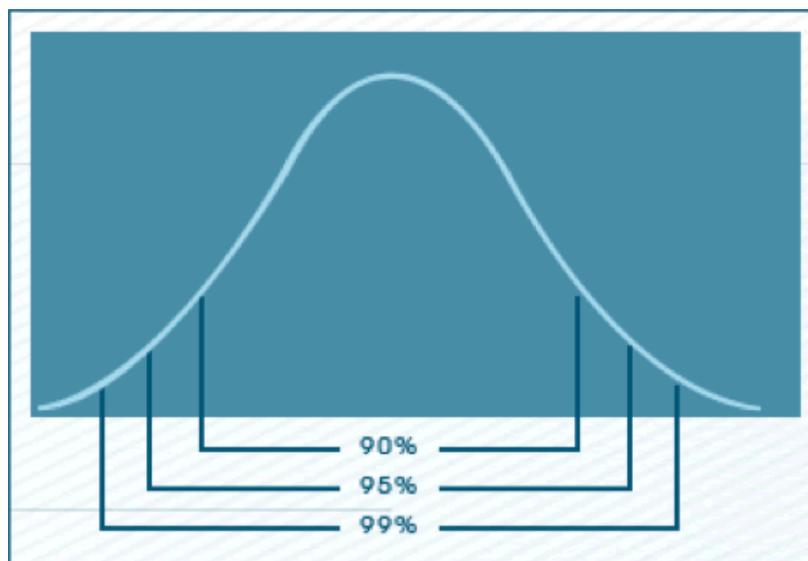
$$n = \frac{Z^2 p(1-p)}{e^2}$$

Un enfoque para seleccionar el tamaño de muestra consiste en maximizar la ecuación, teniendo en cuenta que $p(1-p)$ es máximo cuando $p = 0.5$.

Si la población es finita, la muestra que debe seleccionarse es del siguiente tamaño:

$$n = \frac{Z^2 p(1-p)N}{(N-1)e^2 + Z^2 p(1-p)}$$

En las ecuaciones anteriores Z es el nivel de confianza, p es la probabilidad de ocurrencia, e es el margen de error, N es el tamaño poblacional y n es el tamaño de la muestra. De cada uno de ellos se discutirá ampliamente.



En el gráfico puede observarse fácilmente que, si se aumenta el nivel de confianza, el intervalo de estimación se amplía (aumenta el error). Igualmente, el margen de error también se puede disminuir si se disminuye la confianza en la estimación, pero no es una estrategia muy deseable puesto que se estaría incrementado la posibilidad de obtener una conclusión incorrecta.

Lo ideal es, entonces, que el intervalo establecido tenga un alto nivel de confianza, pero que sea estrecho (con poco error) porque mientras más alto sea el nivel de confianza, más probable es que el intervalo contenga el parámetro y mientras más estrecho es el intervalo más exacta es la estimación. Estas dos propiedades parecen contraponerse y sólo pueden mejorarse si se aumenta el tamaño de la muestra (y eso implica un aumento en costos y tiempo); por lo tanto, el trabajo consiste en “equilibrar” el nivel de confianza, el tamaño de la muestra y el margen de error admitido de modo que se obtenga un intervalo aceptable.

En otras palabras, tomar una muestra más grande de lo necesario para alcanzar los resultados deseados es derrochar los recursos, mientras que las muestras muy pequeñas conducen a resultados inválidos. El objetivo principal es obtener tanto una exactitud deseable como un nivel apropiado de confianza con mínimos costos y gasto de recursos. Debido a esto, al calcular el tamaño de muestra es importante que el investigador se pregunte qué precisión deben tener sus estimaciones.

Probabilidad de ocurrencia (p):

Puede ser hallada según un estudio similar anterior, de lo contrario se fija como 0.5.

Tamaño de la población (N)

La universidad colombiana no puede asumir una posición de indiferencia ante los Fijado al iniciar el estudio. Si es muy grande o es muy difícil

contarlo, no se considera y se aplica la ecuación para poblaciones infinitas.

Es importante anotar que todo lo anterior está basado en la distribución Normal o distribución de Gauss, que es muy confiable mientras la muestra no sea muy pequeña. Si la población es muy pequeña, podría pensarse en evaluar todos y cada uno de sus elementos.

A continuación se ilustra la temática expuesta en este artículo a través de dos ejemplos que detallan la manera de seleccionar el tamaño apropiado de muestra en casos específicos:

1. Se está realizando una investigación sobre “Perfil y percepción del turista extranjero que visita la ciudad de Medellín”. Una de las etapas del estudio comprende a los turistas que vinieron a Medellín durante la celebración de la feria de Colombiatex.

Según datos suministrados por los directivos de la feria, la población de turistas esperada era de 1300 extranjeros. Si en los resultados se quiere tener un nivel de confianza de 90% y un margen de error de 5%, ¿cuál debe ser el tamaño de la muestra a utilizar?

$$n = \frac{1.645^2 * 0.5 * 0.5 * 1300}{1299 * 0.05^2 + 1.645^2 * 0.5 * 0.5} = 225 \text{ turistas (17.3\%)}$$

Puede observarse al reemplazar en la ecuación que, si el margen de error se disminuye a 3%, el tamaño de muestra debe aumentarse a 477.

2. Se está realizando una investigación de mercados para lanzar un nuevo producto. ¿De qué tamaño debe ser la muestra que se seleccione si se quiere tener un nivel de confianza de 90% y un margen de error de 5%?

La población objetivo se puede considerar infinita si es muy grande. Cuando la población puede considerarse infinita, se necesitan muestras proporcionalmente inferiores con relación a la población (no tan grandes).

$$n = \frac{1.645^2 * 0.5 * 0.5}{0.05^2} = 271 \text{ personas}$$

Si el margen de error se disminuye a 3%, el tamaño de muestra debe aumentarse a 752 personas.

Conclusión

Los resultados de las pruebas estadísticas no deben seguirse ciegamente; obviamente, es necesario combinar el análisis estadístico con la experiencia en el campo. “Las técnicas estadísticas son inservibles a menos que se combinen con el apropiado conocimiento del tema al que se aplican y con la experiencia previa; son un complemento al conocimiento del tema, nunca un sustituto”.

Bibliografía

McDANIEL, Carl y GATES, Roger. Investigación de mercados. Sexta edición. México: Editorial Thomson, 2005. 617p.c

KINNEAR, Thomas y TAYLOR, James. Investigación de mercados: un enfoque aplicado. 4a edición. México: McGraw-Hill Interamericana S.A., 1998. 874p.

ANDERSON, David; SWEENEY, Dennys y WILLIAMS, Thomas. Estadística para administración y economía. 10a edición. México: CENGAGE Learning, 2008. 1056p.

Pérez, Adriana. Castañeda, Javier. Gil, Jacky (2000) Tamaño de la muestra en análisis de sobrevida. Revista Colombiana de Estadística. <http://ideas.repec.org/a/col/000163/004229.html> Consultado el 4 de mayo de 2009.

Estadística para no estadísticos. César Augusto Gutiérrez Villafuerte.

Lima, Febrero de 2008

http://www.epiredperu.net/epired/eventos/eve_socimep-redaccion08/_socimep_redaccion08_21.pdf Consultado el 2 de mayo de 2009.

Algunas técnicas de muestreo estadístico. Dianelys Munguía Álvarez, Cuba:

http://www.cfg.rimed.cu/revista/_publications/vol2_no2/article61.pdf

Consultado el 20 de mayo de 2009.

Razonamiento estadístico para decisiones gerenciales. _ Consultado el 15 de mayo de 2009.

*** Francisco Javier Jaramillo Álvarez**

- Ingeniero de Alimentos, egresado de la Corporación Universitaria Lasallista, Medellín, Colombia
- Especialista en estadística, Universidad Nacional de Colombia
- Diplomado en Formación en tutoría virtual, OEA
- Docente Institución Universitaria CEIPA.
- francisco.jaramillo@ceipa.edu.co