

TOPICOS DE INFERENCIA ESTADÍSTICA: EL METODO INDUCTIVO Y EL PROBLEMA DEL TAMAÑO DE LA MUESTRA

TOPICS OF STATISTICAL INFERENCE: INDUCTIVE METHOD AND THE PROBLEM OF SAMPLE SIZE

Bruno E. Vargas Biesuz

**Instituto de Investigacion en Ciencias Económicas y Financieras
Universidad La Salle - Bolivia**

bruvarbi@gmail.com

RESUMEN

El artículo explica lo que se debe entender como "Método Inductivo" de investigación y cómo se debe calcular el tamaño de una muestra, cuando se usa este método.

ABSTRACT

This paper deals about the meaning of the "Inference" in science and the way to calculate the size of the sample.

PALABRAS CLAVE Inferencia, Promedio, Población, muestra aleatoria, muestra, Varianza.

KEYWORDS Inference, Mean, Populations, Random Sampling, Sample, Variance.

I. INTRODUCCION

La estadística moderna se aplica actualmente en una muy grande gama de ciencias como "un método científico de análisis" [1] Posiblemente sea el método preferido para la comprobación de la

validez de los nuevos conocimientos, dado su rigor metodológico y su sustento matemático. Esta ciencia, puede ser muy compleja en su marco teórico que está basado principalmente en modelos matemáticos, sin embargo puede ser bastante accesible en su aplicación práctica.

En este contexto, uno de los problemas que tienen quienes tratan de aplicar la estadística inferencial en las investigaciones, es la determinación del tamaño de la muestra. Este trabajo aborda este problema y trata precisamente de ser una guía para quienes utilizan este método de análisis científico para sus indagaciones y deben resolver el asunto concerniente al tamaño de la muestra que se debe tomar, a partir de la cual se realizarán las inferencias y las pruebas de hipótesis planteadas. Sin embargo antes de analizar el problema planteado, considero necesario explicar en qué consiste el método inductivo o inferencial que usa la estadística.

II. OBJETIVOS

El objetivo del presente artículo, es la revisión del tema específico que se presenta a quienes realizan investigación científica, usando el método inductivo o inferencial y presentar respuestas prácticas, en términos de expresiones matemáticas o fórmulas, para calcular el tamaño de una muestra aleatoria.

III. POBLACION, MUESTRA Y ERROR

La esencia de la estadística tiene que ver con el hecho de que los investigadores tratan de conocer las características de una determinada población o colectivo. La población "es el conjunto total de objetos o personas que van a estudiarse" [2] y, puede ser de cualquier tamaño. Por ejemplo un ingeniero industrial estará interesado en estudiar las fábricas de textiles de todo el país y, obtener ciertas características de este colectivo como ser: cuántos trabajadores en promedio están contratados por fábrica, el monto promedio de inversiones en bienes de capital por trabajador, la proporción de hombres o mujeres que trabajan en la industria, etc. A un psicólogo le interesaría estudiar el comportamiento de los adolescentes de una ciudad y, analizar ciertas variables como: el número de horas promedio que miran televisión, el número de libros que en promedio leen por año o cualquier otra característica.

Lo sorprendente del caso es que, la técnica estadística permite obtener las características buscadas (p.e. promedios), sin necesidad de estudiar a toda la población. **¡Será suficiente obtener y estudiar una pequeña parte de la población y a partir del estudio de esa pequeña parte, la cual se llama muestra, se obtienen aproximadamente los mismos resultados;**

Esta técnica estadística se usa por varias razones, entre las más importantes:

- o Es menos costoso estudiar una muestra que a toda la población.
- o Lleva menos tiempo estudiar una muestra que a toda la población.

Debe quedar claro sin embargo que los resultados o características que se obtengan de la muestra, **pueden ser iguales a los de la población o bastante aproximados.** Esto nos lleva al concepto de "error" estadístico. Es decir esta técnica científica, admite que la característica de la muestra (p.e. promedio muestral \bar{X}), es igual o se aproxima bastante a la característica de la población (p.e. promedio poblacional μ). En otras palabras: el promedio de la población (μ), es igual al promedio de la muestra (\bar{X}), más o menos cierto error: $\mu = \bar{X} \pm \text{un cierto error}$

$$\mu = \bar{X} \pm \text{un cierto error}$$

El asunto fundamental aquí es, ¿qué tan pequeño es ese error?, ¿Qué certeza se tiene de estar cerca del dato correcto? Las respuestas a estas preguntas dependen de que la muestra haya sido obtenida de forma aleatoria, es decir por sorteo, donde todos los elementos o componentes de la población hayan tenido la misma posibilidad de haber sido elegidos y de un tamaño adecuado de muestra, problema que trata de resolver este artículo.

III.1 INDUCCION Y DEDUCCION

Una vez explicados los conceptos de población, muestra y error, pasamos a explicar el significado y esencia del método inductivo o inferencial que usa la estadística en su análisis.

El método o razonamiento inductivo implica la determinación de las características generales de toda una población, a partir del estudio de lo específico, es decir de la muestra. Muchas veces se explica esto diciendo que el método inductivo va de lo particular a lo general, definición poco clara si no se analizan antes los conceptos de población, muestra y error.

III.2 CALCULO DEL TAMAÑO DE MUESTRA

Generalmente, los motivos que llevan a usar el método inductivo o inferencial en las investigaciones y por lo tanto a la obtención de una muestra, son estimar la

media o promedio poblacional (μ), la desviación estándar poblacional (δ) o la proporción poblacional (π). En toda estimación de estos parámetros poblacionales, el investigador se permite o tolera un error máximo “ ϵ ” que es previamente estipulado. Además se debe establecer la confianza de que el error no supere esa cantidad “ ϵ ”, lo cual será denotado por $(1 - \alpha)$, diferencia que llamaremos "coeficiente de confianza".

III.2.1 TAMAÑO DE MUESTRA PARA ESTIMAR LA MEDIA POBLACIONAL (μ), CUANDO SE CONOCE LA DESVIACION ESTÁNDAR (δ).

Lo que se busca es que el tamaño de la muestra "n", sea tal que:

$$\text{Prob } (|X - \mu| < \epsilon) \geq 1 - \alpha$$

Según la desigualdad de Shebyshev:

$$n \geq \frac{\delta^2}{\alpha \cdot \epsilon^2} \quad (1)$$

Según el Teorema Central del Límite:

$$n \geq \max \left\{ z_{\alpha/2}^2 \frac{\delta^2}{\epsilon^2} \mid 30 \right\} \quad (2)$$

TOPICOS DE INFERENCIA ESTADÍSTICA: EL METODO INDUCTIVO Y EL PROBLEMA DEL TAMAÑO DE LA MUESTRA

Ejemplo: Se quiere estimar (μ), promedio poblacional mediante el promedio muestral (X), con un error $\epsilon = 0,01$ y un coeficiente de confianza $(1 - \alpha) = 0,9$. Se conoce que $\delta^2 = 3.1$

De aquí: $\alpha = 0,1$; $(\alpha / 2 = 0,05)$

$Z_{0,05} = 1,64$ según tabla de probabilidad normal estándar

Según la relación (1):

$$n \geq [3.1 / (0,1) (0,01)^2] = 310.000$$

Según la relación (2):

$$n \geq \max [(1,645)^2 (3.1) / (0,01)^2; 30]$$

$$n \geq 83.887$$

Nótese una significativa diferencia entre los resultados de las relaciones (1) y (2), por lo que es más recomendable la última fórmula.

III.2.2 TAMAÑO DE LA MUESTRA PARA ESTIMAR LA MEDIA POBLACIONAL (μ), CUANDO NO SE CONOCE LA DESVIACION ESTÁNDAR (δ).

En este caso existen dos posibilidades de cálculo.

4.2.1 Se estima la Desviación Estándar (δ) con la fórmula empírica :

$$\delta = \frac{1}{4} (\text{Rango})$$

Donde el "Rango" o recorrido de la variable estudiada es la diferencia entre el valor máximo y el valor mínimo.

$$\text{Rango} = (X \text{ max} - X \text{ min})$$

Luego se emplean las relaciones (1) o (2)

Ejemplo: En una fábrica de jabón en polvo, se quiere estimar la producción semanal media (μ), mediante la media muestral (X), deseándose que $|X - \mu| < 4$ Ton. Hay razones para creer que la producción semanal oscila entre 340 y 420 Ton. Calcule el tamaño de la muestra para un coeficiente de confianza $(1 - \alpha) = 0,95$

Datos:

$$\epsilon = 4 \text{ Ton.}$$

$$(1 - \alpha) = 0,95; (\alpha = 0,05; \alpha / 2 = 0,025 ; Z_{0,025} = 1,96)$$

$$\text{Rango} = 420 - 340 = 80 \text{ Ton.}$$

$$\delta = \frac{1}{4} (80) = 20 \text{ Ton.}$$

Por (1):

$$n \geq \delta^2 / \alpha \epsilon^2 = 20^2 / (0,05)(4)^2$$

$$n \geq 500$$

Por (2):

$$n \geq \max (z_{\alpha/2}^2 \delta^2 / \varepsilon^2 ; 30)$$

$$n \geq \max [(1,96)^2 20^2 / 4^2 ; 30]$$

$$n \geq \max (96,04 ; 30)$$

$$n \geq 97$$

En el segundo caso, se sustituye " δ " (desviación estándar poblacional), por " s " (desviación standar muestral o error standar), en las fórmulas (1) y (2).

$$\text{Donde: } s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Por tanto, las nuevas relaciones son:

$$\text{En (1): } n \geq \frac{s^2}{\alpha \cdot \varepsilon^2} \quad (1^1)$$

$$\text{En (2): } n \geq \max \left\{ z_{\alpha/2}^2 \frac{s^2}{\varepsilon^2} \mid 30 \right\} \quad (2^1)$$

Sin embargo, las fórmulas (1¹) y (2¹) contienen una contradicción, ya que para el cálculo de de " s^2 ", se necesita saber el valor de " n ", es decir, se necesita haber tomado una muestra, cuyo valor aún no se sabe y es precisamente lo que se trata de calcular.

Esta contradicción se salva calculando " S^2 " a partir de una "muestra piloto", cuyo tamaño podría ser treinta ($n = 30$). Luego se sigue el siguiente criterio de decisión:

- a) Se calcula el valor de n_1 , si $n_1 \leq 30$ se acepta el valor de $n = 30$
- b) Si $n_1 > 30$ se amplía el tamaño de la muestra, se calcula el nuevo valor de " s " y con este se determina n_2 , si $n_2 \leq n_1$ se acepta n_1
- c) Si $n_2 > n_1$ se amplía la muestra, calculando nuevamente " s ", y así sucesivamente.

III.2.3 TAMAÑO DE LA MUESTRA PARA ESTIMAR LA VARIANZA POBLACIONAL (δ^2)

Se pretende calcular el tamaño de la muestra " n " de modo que:

$$\text{Prob} (|P - \pi| < \varepsilon) \geq 1 - \alpha$$

Usando la desigualdad de Chebyshev se obtiene:

$$n \geq \frac{1}{4 \alpha \varepsilon^2} \quad (4)$$

Empleando el Teorema Central del Límite, se obtiene:

$$\mu = \bar{X} \pm \text{un cierto error}$$

TOPICOS DE INFERENCIA ESTADÍSTICA: EL METODO INDUCTIVO Y EL PROBLEMA DEL TAMAÑO DE LA MUESTRA

La primera fórmula (4), da generalmente un valor demasiado grande, en cambio la formula (5), gracias a la aproximación del teorema Central del Límite, da generalmente un valor de "n" bastante menor, por lo que es preferida.

Ejemplo: Se quiere estimar " π " mediante "P" con $\epsilon = 0,02$ y $(1 - \alpha) = 0,95$, de donde $\alpha = 0,05$; $\alpha/2 = 0,025$. $Z_{0,025} = 1,96$

$$\begin{aligned} \text{Según (4):} \quad n &\geq \frac{1}{4 \alpha \epsilon^2} \\ &= \frac{1}{(4)(0,05)0,02^2} = 12.500 \end{aligned}$$

$$\begin{aligned} \text{Según (5):} \quad n &\geq \max \left\{ \frac{z_{\alpha/2}^2}{4\epsilon^2} \mid 30 \right\} \\ n &\geq \max \left\{ \frac{1,96^2}{(4)0,02^2} \mid 30 \right\} \\ n &\geq \max \{2401 \mid 30\} \\ n &\geq 2401 \end{aligned}$$

III.2.5 FORMULA DE MUNCH Y ANGELES.

[3] Otra posibilidad para estimar el tamaño de muestra para proporciones o porcentajes de población, cuando la población (N) es muy grande y la variable aleatoria estudiada es binomial o dicotómica, es la fórmula de Lourdes Munch y Ernesto Angeles:

$$n = \frac{z^2 N p q}{N \epsilon^2 + z^2 p q}$$

Donde:

Z = valor estándar de la distribución probabilística normal para el nivel de confianza establecido (si 95 % $\Rightarrow z = 1,96$)

ϵ = error probable

p = probabilidad de éxito

q = probabilidad de fracaso

N = Población o universo

n = tamaño de la muestra

Si no se cuenta con información sobre la probabilidad de éxito o fracaso, se recomienda usar $p = q = 0,5$

Por otra parte, el error (ϵ) máximo recomendable debe ser 0.1

Si la población es finita, se recomienda usar el Muestreo Aleatorio Simple y la siguiente fórmula:

$$n = \frac{z^2 N p q}{(N - 1) \epsilon^2 + z^2 p q}$$

BIBLIOGRAFIA

[1] Thomas H. Wonnacott & Ronald J. Wonnacott. (1979). Fundamentos de Estadística Para Administración y Economía. Editorial Limusa Mexico

[2] Thomas H. Wonnacott & Ronald J. Wonnacott. (1991). Estadística Básica Práctica. Editorial Limusa Mexico

[3] www.monografias.com

VI AGRADECIMIENTO:

El autor agradece las explicaciones, aclaraciones y cooperación brindadas en su momento, por el Ingeniero Luís Crespo Ostría, destacado académico, a quien tuvimos el privilegio de contar en la planta docente de la Universidad La Salle, por lo que el presente artículo está dedicado a su memoria.

Asimismo, se agradece al Licenciado Dindo Valdez en la atención recibida en la revisión del artículo.

Recibido: 10/08/2013

Aceptado: 03/02/2014