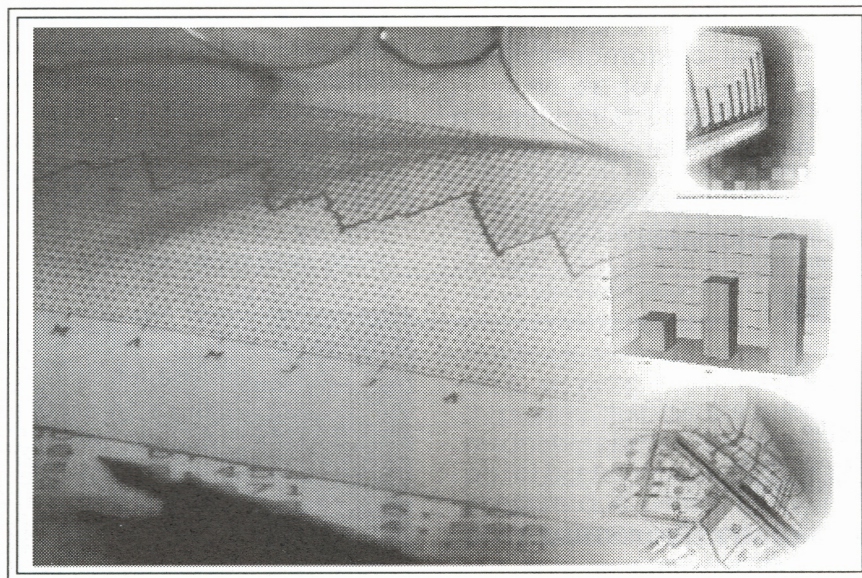


Factores que determinan el tamaño de una muestra

Un análisis de su sensibilidad a través del estimador de la media poblacional en el Muestreo Aleatorio Simple (MAS)



Rafael Klinger Angarita

Estadístico. Profesor Asistente. Universidad del Valle.
rakling@pino.univalle.edu.co

La fase correspondiente a la determinación del tamaño de la muestra para la realización de una investigación estadística es de suma importancia para garantizar la validez externa de la investigación. El presente artículo muestra los elementos fundamentales que determinan un tamaño de muestra y su sensibilidad frente a los cambios experimentados por éstos últimos. Saber de qué depende el tamaño de una muestra y su sensibilidad, es clave para que el usuario del muestreo pueda determinar con mayor objetividad un tamaño de muestra. El presente análisis se realiza tomando como ejemplo la estimación de la media poblacional en el muestreo aleatorio simple (MAS).

Palabras clave: Parámetro poblacional, estimador, estimación, error de muestreo, confiabilidad y cuasivarianza poblacional, efecto de diseño (Deff).

1. Introducción

¿Qué tamaño de muestra se requiere para estimar la media poblacional sobre una característica de una población de unidades estadísticas previamente determinada?

¿Qué margen de error se debe usar para realizar tal estimación?

¿Con qué nivel de confianza se debe esperar que se cumpla el margen de error determinado?

Si se dispone de un presupuesto muy limitado para la toma de información en las unidades

muestreadas, ¿qué tamaño de muestra se debe esperar?

Si paralelamente se desea satisfacer un presupuesto C para el trabajo de campo (toma de información muestral) y una varianza para el estimador de la media poblacional, ¿qué tamaño de muestra se debe asumir?

Preguntas como las anteriores preocupan al investigador antes de definir y estructurar el trabajo de campo mediante el cual recogerá la información necesaria para la obtención de los indicadores que darán respuesta a los objetivos de su investigación.

La técnica del muestreo nos permite acopiar información sobre una parte de la población (muestra) que se analizará en una investigación. Esta población por lo general es de gran tamaño y puede estar integrada por unidades manufacturadas ó por unidades naturales. Debido al carácter multipropósito que está presente en cualquier estudio se requiere medir varias características en las unidades objeto de investigación, y como parte de éste supongamos que la principal característica que se debe analizar es X (variable) y de ella en concreto se desea conocer μ_X , es decir el promedio que resultaría si se pudiera medir la característica X en todas las unidades poblacionales.

Para aproximarse al valor de μ_X (media poblacional), llamado **parámetro poblacional**, se utiliza una función que depende de los valores obtenidos al medir X en las unidades muestreadas, llamada **estima-**

dor y representada por $\hat{X}$ (media muestral), cuyos valores reciben el nombre de **estimaciones**. Al depender el estimador de las unidades muestreadas, es lógico obtener estimaciones mayores que μ_X , en cuyo caso estaremos sobrestimando la media poblacional ó estimaciones menores que μ_X , en donde lógicamente estaremos subestimando la media poblacional. Es natural entonces que el investigador desee que su estimación esté lo más cerca del parámetro, no importa si lo subestima o lo sobrestima ya que en la práctica nunca lo podrá saber.

Por otro lado el investigador al tomar una muestra

de entre las $k = \binom{N}{n}$ posibles, desea saber, cuál

será el nivel de confianza para que se cumpla el máximo alejamiento permitido entre el parámetro y la estimación obtenida por él con base en la muestra seleccionada.

La Probabilidad y la Inferencia Estadística permiten establecer expresiones matemáticas que reflejan el comportamiento del tamaño de muestra en dependencia con la variabilidad de la pobla-

ción $S_X^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_X)^2$, la cual llamaremos

en lo sucesivo cuasivarianza poblacional; el nivel de confianza deseado $(1-\alpha)100\%$ y el máximo error o alejamiento de μ_X permitido (δ). Igual-

mente es deseable en muchas oportunidades establecer la relación entre el tamaño de la muestra y el presupuesto de que se dispone para la realización de las mediciones o trabajo de campo.

Es importante a la hora de definir un tamaño de muestra para la estimación de un parámetro, en este caso la media poblacional; tener presente los factores que afectan el tamaño de la muestra para de esta forma decidir dicho tamaño, de acuerdo con las necesidades de la investigación y la importancia de las conclusiones.

En lo que sigue se mostrará la importancia de los factores que determinan el tamaño de una muestra y para ello se hará siempre referencia a la estimación de la media poblacional.

2. Relación entre el tamaño de la muestra y el tamaño de la población

La inferencia Estadística establece que al considerar todas las K posibles muestras aleatorias sin reemplazo de tamaño n con igual probabilidad de selección, de una población cuyo número de uni-

dades es N , el estimador $\hat{X}$, llamado la media muestral es una variable aleatoria cuyos indicadores de centramiento y variabilidad son respectivamente:

$$\text{Esperanza matemática: } E(\hat{X}) = \frac{\sum_{i=1}^K \hat{X}_i}{K} = \mu_X$$

$$\text{Varianza: } V(\hat{X}) = \frac{(1-f)}{n} \cdot S_X^2 \quad (1)$$

El término $f = \frac{n}{N}$ que aparece en la expresión

(1), es llamado fracción de muestreo e indica qué tan grande es el tamaño de la muestra comparado con el tamaño de la población; en otras palabras, determina el número de unidades poblacionales que están siendo representadas por cada unidad muestral. En la práctica se dice que la fracción de muestreo es insignificante y que por lo tanto el factor $(1-f)$ es despreciable, cuando $f < 5\%$.

Como se dijo antes al estimar la media poblacional μ_X mediante el uso del estimador, cometemos el llamado **error de muestreo**, el cual desde luego deseamos que sea pequeño, este deseo se consigue mediante la siguiente relación fundamental, llamada **relación de precisión**:

$$\left| \hat{\bar{X}} - \mu_x \right| \leq \delta \quad (2)$$

En ella, δ será el error de muestreo máximo que el investigador está dispuesto a tolerar para dicha diferencia. Note que en cierta manera dar una diferencia entre el parámetro y el estimador es una manera de cuantificar qué tan exacto se desea ser con dicha estimación.

Ahora, como el investigador encontrará su estimación mediante una sola muestra (subgrupo de unidades poblacionales), él deseará medir qué tan confiable será dicha estimación para medir la media poblacional. Esta confianza se mide en términos de probabilidad y significa la seguridad con que la muestra le permite al investigador encontrar una estimación que cumpla con la relación de precisión dada en (2). La relación que involucra la precisión con el nivel de confianza viene dada por:

$$P_r \left(\left| \hat{\bar{X}} - \mu_x \right| \leq \delta \right) \leq (1 - \alpha) \quad (3)$$

Si en esta expresión que podríamos llamar *relación de precisión y de confianza*, por ejemplo, asumimos que $(1 - \alpha) \cdot 100 = 95\%$, estaremos esperando tener un riesgo del 5% y una certeza del 95% de que la estimación obtenida mediante el tamaño de muestreo definido, cumpla con la relación de precisión establecida.

Al suponer que la distribución probabilística de $\hat{\bar{X}}$, tiene un comportamiento normal y apoyándonos en el Teorema del Límite Central, si dividimos la expresión (3) por la raíz cuadrada de la varianza

del estimador, vale decir por $\sqrt{V(\hat{\bar{X}})}$, obtenemos la expresión:

$$P_r \left(\frac{\left| \hat{\bar{X}} - \mu_x \right|}{\sqrt{V(\hat{\bar{X}})}} \leq \frac{\delta}{\sqrt{V(\hat{\bar{X}})}} \right) < (1 - \alpha) \quad (4)$$

De la expresión anterior y bajo el supuesto de normalidad, utilizando la distribución normal estándar, obtenemos:

$$\frac{\delta}{\sqrt{V(\hat{\bar{X}})}} = Z_{\alpha/2} \therefore V(\hat{\bar{X}}) = \left(\frac{\delta}{Z_{\alpha/2}} \right)^2 \quad (5)$$

Lo que permite la expresión (5), es cuantificar la varianza que el investigador desea tengan sus estimaciones mediante el uso del estimador $\hat{\bar{X}}$, para el tamaño de la muestra que finalmente adoptará. Note que precisamente sus dos componen-

tes son el error de muestreo (δ) y la precisión, reflejada ésta última por $Z_{\alpha/2}$, ambos determinados por el investigador.

Igualando las expresiones (1) y (5), por decir de alguna manera igualando la teoría con lo deseado, se tiene la *ecuación fundamental del tamaño de muestra*, la cual viene expresada por:

$$\frac{(1 - f)}{n} \cdot S_x^2 = \left(\frac{\delta}{Z_{\alpha/2}} \right)^2 \quad (6)$$

Mediante esta ecuación el investigador podrá finalmente calcular el tamaño de la muestra requerido para la estimación del parámetro media poblacional. En efecto realizando unos pocos cálculos al despejar la incógnita n , de la ecuación (6), se tiene:

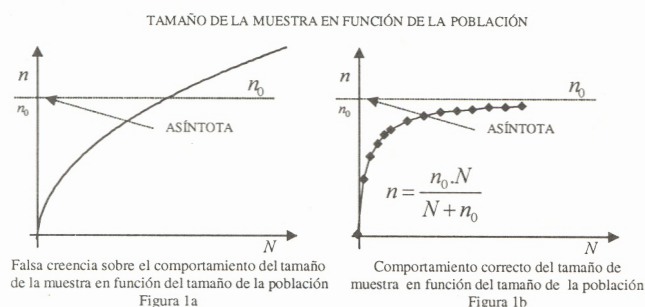
$$n = \frac{\left\{ \frac{Z_{\alpha/2}^2 \cdot S_x^2}{\delta^2} \right\}}{1 + \frac{1}{N} \left\{ \frac{Z_{\alpha/2}^2 \cdot S_x^2}{\delta^2} \right\}} = \frac{n_0}{1 + \frac{n_0}{N}} \quad (7)$$

En la expresión (7), el término n_0 , recibe el nombre de primera aproximación al tamaño de muestra, el cual es por decirlo así, la aproximación gruesa; que como puede verse en caso de que el tamaño de la población sea muy grande, entonces no será necesario realizar ningún ajuste de corrección por población finita siendo en este caso $n \cong n_0$.

Al observar la expresión $n = \frac{n_0}{1 + \frac{n_0}{N}}$, notamos que

hay una relación entre el tamaño de muestra n y el tamaño de la población N , pero la mayoría de los usuarios no estadísticos del muestreo tienen la falsa creencia de que el crecimiento de la muestra es directamente proporcional al tamaño de la población (comportamiento de regla de tres directa) ver figura 1a. Para ellos es un contrasentido, que dos poblaciones ampliamente diferentes en su tamaño, requirieran la misma cantidad de unidades muestrales para la realización de las estimaciones de sus parámetros.

La figura 1b muestra la verdadera relación entre el tamaño de la muestra y el tamaño de la población de la cual ha sido extraída



La figura 1b nos permite apreciar lo siguiente:

- El tamaño de la muestra crece a medida que N aumente pero este crecimiento es asintótico con respecto a n_0 . Aunque la población sea inmensamente grande «infinita», el tamaño de muestra máximo no pasará del valor n_0 .
- Utilizar un tamaño de muestra n mayor a la primera aproximación n_0 , aunque satisface las condiciones establecidas para la estimación de μ_X , es un tamaño de muestra ineficiente ya que se incurre en gastos innecesarios al medir las unidades extras.
- Se aprecia claramente que un tamaño de muestra n , puede satisfacer las necesidades establecidas para la estimación de μ_X en dos poblaciones de tamaños N_1 y N_2 muy diferentes entre sí.

3. Dependencias del tamaño de la muestra

Para apreciar la sensibilidad del tamaño de muestra n , frente a los factores que la definen, analizaremos en detalle la expresión que define la primera aproximación en la estimación de μ_X . Esto es:

$$n_0 = f(Z_{\alpha/2}, \delta, S_X^2) = \frac{Z_{\alpha/2}^2 \cdot S_X^2}{\delta^2} \quad (8)$$

Aunque los factores que definen la expresión (8), pueden variar simultáneamente para producir una primera aproximación n_0 , en lo que sigue vamos a suponer constantes dos de los tres factores y sometemos a cambio el factor restante, con el objeto de ver el comportamiento de la primera aproximación según éste último factor.

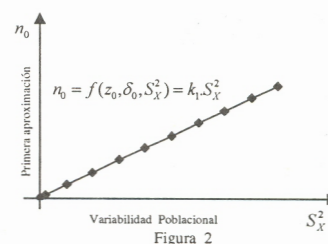
La tabla siguiente muestra las posibilidades de análisis que se abordarán para la función n_0 dependiendo del estado inicial de sus componentes.

Estado inicial de los factores que definen n_0			Función resultante
$Z_{\alpha/2}$	δ	S_X^2	$n_0 = f(Z, \delta, S_X^2)$
Constante	Constante	Variable	$n_0 = k_1 \cdot S_X^2$
Constante	Variable	Constante	$n_0 = k_2 / \delta^2$
Variable	Constante	Constante	$n_0 = k_3 \cdot Z^2$

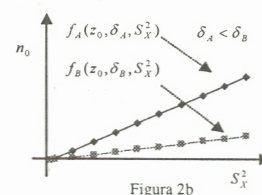
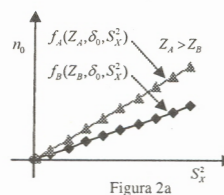
- Sensibilidad de n_0 frente a la variabilidad S_X^2 de la característica analizada X , en la población.

Como se sabe la heterogeneidad de los valores que toma la característica X en la población, se mide básicamente por la cuasivarianza S_X^2 . Estamos interesados entonces en el comportamiento de n_0 a medida que ésta aumenta, es decir a medida que los datos para una determinada población son cada vez más heterogéneos, cuando suponemos fijos el error de muestreo δ y el nivel de confianza $(1-\alpha)100\%$.

PRIMERA APROXIMACIÓN DEL TAMAÑO DE LA MUESTRA EN FUNCIÓN DE LA VARIABILIDAD DE LA POBLACIÓN



LA PRIMERA APROXIMACIÓN EN FUNCIÓN DE LA VARIANZA DE LA POBLACIÓN. DOS SITUACIONES DE INTERÉS



Como podemos apreciar, la figura 2 nos permite dilucidar que si una población es muy heterogénea en los valores que asumen sus unidades para la característica X analizada, entonces el tamaño de muestra que se requiere para estimar la media poblacional para un nivel de confianza y un error de muestreo establecidos será grande, debido precisamente a una relación directamente proporcional entre la primera aproximación de la muestra y la cuasivarianza poblacional.

De la figura 2a podemos deducir que la primera aproximación crece más en presencia de un incremento en la cuasivarianza, cuando se desea un nivel de confianza alto, mientras que ella crece también rápidamente cuando el error de

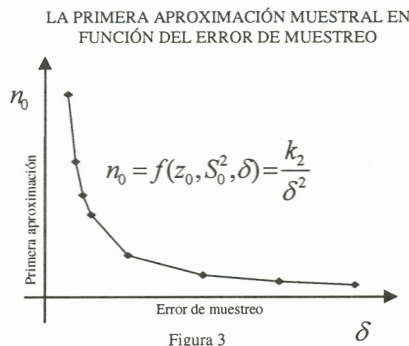
muestreo se hace a su vez más pequeño, figura 2b. El investigador debe esperar una muestra grande cuando el error de muestreo sea pequeño y la variabilidad de la población alta e idénticamente cuando el nivel de confianza sea grande y la variabilidad poblacional también lo sea.

- Sensibilidad de n_0 frente a la precisión δ que desea el investigador para la estimación de la media poblacional.

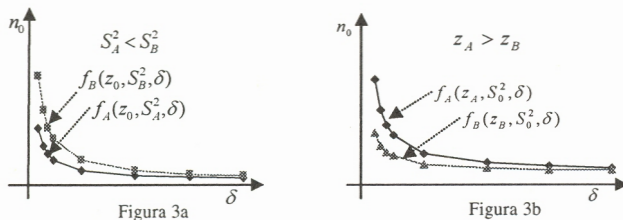
Si se desea ahora establecer cual es el comportamiento del tamaño de la muestra, cuando el error de muestreo aumenta ó disminuye, es decir responder a la pregunta ¿Qué tipo de proporcionalidad existe entre la primera aproximación y la precisión, cuando establecemos de antemano una cuasivarianza poblacional S_x^2

y un nivel de confianza $(1 - \alpha)100\%$?

La respuesta puede obtenerse de las Figuras 3, 3a y 3b.



LA PRIMERA APROXIMACIÓN EN FUNCIÓN DEL ERROR DE MUESTREO DOS SITUACIONES DE INTERÉS



Si el investigador desea una precisión alta (error de muestreo pequeño), el precio que él debe pagar por tal deseo será una muestra grande, debido como puede verse en la figura 3, a que la relación entre la primera aproximación n_0 y el error de muestreo δ es inversamente proporcional.

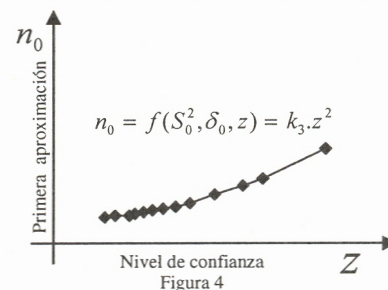
Así mismo con base en la figura 3a se deduce que la primera aproximación decrece más rápi-

damente si la población es más homogénea (S^2 pequeña), e idénticamente decrece rápidamente cuando el nivel de confianza es menor figura 3b.

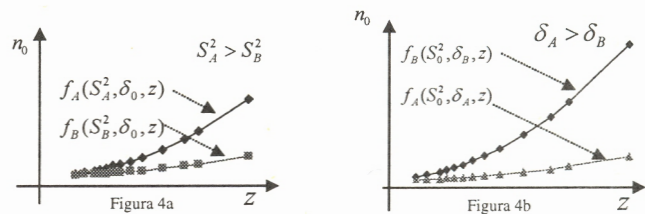
- Sensibilidad de n_0 frente a la confiabilidad $(1 - \alpha)100\%$ que desea el investigador para la estimación de la media poblacional.

Si el investigador desea estar cada vez más seguro de que la estimación hallada para la media poblacional con la muestra establecida, cumpla con el margen de error establecido, se puede en presencia de una cuasivarianza y un error de muestreo fijos analizar el comportamiento del tamaño de la muestra si se incrementa la confianza, en este caso representada por z .

LA PRIMERA APROXIMACIÓN MUESTRAL EN FUNCIÓN DEL NIVEL DE CONFIANZA



LA PRIMERA APROXIMACIÓN EN FUNCIÓN DEL NIVEL DE CONFIANZA. DOS SITUACIONES DE INTERÉS



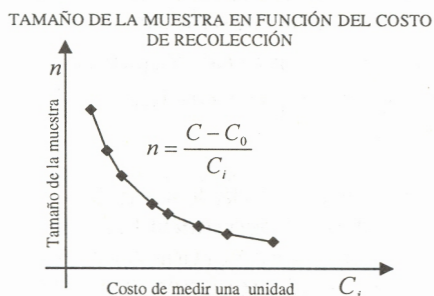
En la figura 4 se puede apreciar, que a medida que se incrementa el nivel de confianza para disminuir el riesgo de que la discrepancia entre la media poblacional y su estimación sea superior al error δ establecido, la primera aproximación se incrementa notablemente. Este hecho como puede verse en la gráfica se da debido a que la relación entre la primera aproximación y el nivel de confianza es de carácter directo. Así mismo se incrementa mucho más la primera aproximación al aumentar el nivel de confianza si la población es mucho más variable (figura 4a) e idénticamente si el error de muestreo es más pequeño (figura 4b).

Si el investigador como es usual involucra el tamaño de muestra con el presupuesto de que se dispone para el trabajo de campo, él debe definir una función para el costo como por ejemplo la función:

$$C = C_0 + C_1 n \quad (9)$$

En donde C es el presupuesto total de que dispone el investigador para la recolección de la información; C_0 es el costo fijo de las actividades o recursos que independientemente del número de unidades muestreadas debe asumirse al realizar el trabajo de campo y C_1 es el costo de medir o encuestar una unidad de la muestra.

Si se asume un presupuesto C , y un costo C_0 determinados, se puede observar el comportamiento del tamaño de la muestra a medida que se incrementa el costo de medir (valor de una encuesta) una unidad muestreada.



Se observa en la figura 5, que cuando el costo de medir una unidad muestreada es muy bajo el tamaño de la muestra es muy grande cuando solo éste tamaño está acotado por el presupuesto C establecido para el trabajo de campo y en caso contrario cuando los costos de medición son supremamente altos (destrucción de la unidad, uso de reactivos muy costosos, unidades de difícil acceso, etc.) el número n de unidades que se pueden muestrear será por obligación pequeño. Los costos son a menudo la limitante más fuerte para la determinación del tamaño de una muestra.

Es importante determinar algunas veces el tamaño de una muestra para minimizar simultáneamente el costo total C para el trabajo de campo y la varianza del estimador para la media poblacional. Para llegar a dicho tamaño de muestra definimos la siguiente función sumando las expresiones (1) y (6) anteriormente planteadas obteniendo:

$$\Delta(n) = V(\hat{X}) + C(n) = \left\{ \left(1 - \frac{n}{N} \right) \frac{S_x^2}{n} \right\} + \{ C_0 + C_1 n \} \quad (10)$$

Si derivamos e igualamos a cero, en virtud de las propiedades del cálculo diferencial, encontramos que la función (10) tiene un mínimo cuando n toma el valor:

$$n = \sqrt{\frac{S_x^2}{C_1}}$$

Esta expresión de la muestra, nos deja ver que su tamaño puede aumentar en presencia simultánea de dos condiciones: de un lado alta heterogeneidad en los valores de X , para las unidades poblacionales y de otro lado poco costo al realizar la medición o encuesta en la unidad muestreada.

Finalmente, hay que destacar que la cuasivarianza

S_x^2 , requiere un estudio independiente ya que constituye un verdadero cuello de botella a la hora de determinar aproximadamente su valor (Estimación adelantada) para poder definir finalmente el tamaño de una muestra ya que como vimos, éste es un factor que afecta directamente el tamaño de una muestra. Existen varias formas para determinar a priori un valor de la cuasivarianza poblacional, entre las que se pueden citar cómo básicas las siguientes: a) Conjetura sobre el rango y la forma aproximada de la distribución de la característica X , b) Mediante el análisis de posibles relaciones con otras características correlacionadas con la variable objeto de análisis y de las cuales se posee información, c) Mediante el conocimiento de investigaciones realizadas en el pasado que involucraron la característica analizada, d) Mediante un muestreo piloto, el cual además nos permite afinar otros aspectos de suma importancia para la investigación, e) Mediante un muestreo en dos pasos, Cox (1952) y Stein (1945)¹. Sea cual fuere el método que se utilice para estimar en forma adelantada la cuasivarianza S_x^2 ,

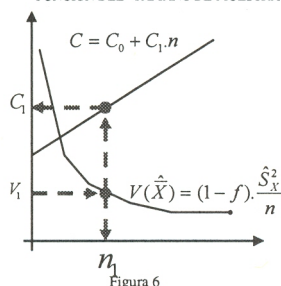
debemos tener el suficiente criterio científico para decidir usarlo.

4. Analizando simultáneamente la precisión alcanzada para la estimación y el costo de recolección en función del tamaño de la muestra

La figura 6 muestra simultáneamente el comportamiento del costo de trabajo de campo en que se incurre y la varianza del estimador alcanzada a medida que se aumenta el tamaño de la muestra. En dicha figura como puede verse, mientras el cos-

¹ Cochran W. Técnicas de muestreo. Cecs. México, 1980- página 111

EL COSTO Y LA VARIANZA DEL ESTIMADOR EN FUNCIÓN DEL TAMAÑO DE MUESTRA



to de trabajo de campo aumenta, la variabilidad alcanzada por las estimaciones muestrales disminuye a medida que se aumenta el número de unidades que conforman la muestra.

La simultaneidad entre el costo y la varianza alcanzada a medida que aumenta n proporciona a su vez tres situaciones importantes aunque sencillas, dependiendo de cuál sea el elemento que defina el tamaño de la muestra para el investigador. Ver figura 7.

Primeramente si el investigador prioriza la varianza (la variabilidad que desea tengan sus estimaciones) éste hecho determinará a su vez el tamaño de muestra necesario para alcanzarla y éste por transitividad el costo necesario que tendrá dicha variabilidad. Cuando el investigador prioriza el costo de investigación éste a su vez determinará el tamaño de muestra y por transitividad la varianza alcanzada. Finalmente si él prioriza el tamaño de muestra de hecho determina separadamente la varianza alcanzada y el costo al que se llega en el trabajo de campo.

Es bueno aclarar que el investigador debe manejar simultáneamente varios elementos y plantear de acuerdo con ellos diferentes escenarios, de entre los cuales de acuerdo con las necesidades de la investigación él escogerá el más adecuado.

5. Conclusiones

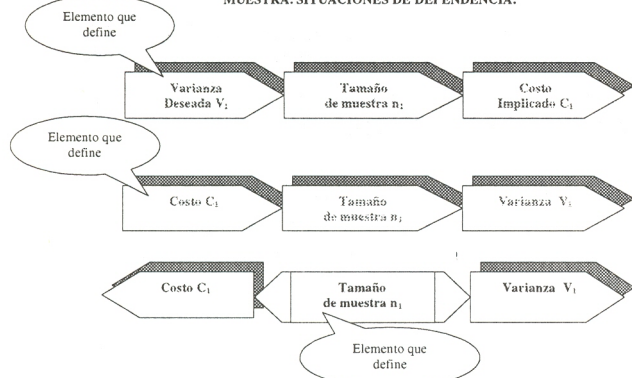
A manera de resumen sobre la determinación del tamaño de una muestra en términos generales se puede resaltar lo siguiente:

- El tamaño de muestra para estimar un parámetro poblacional como la media es un hecho científico que involucra de un lado la teoría estadística y de otro el deseo y las necesidades de la investigación.
- El tamaño de una muestra así como la forma de seleccionarla y el marco de muestreo de que se disponga para su selección, constituyen pilares fundamentales para la validez externa de una investigación, aspecto este, que conjuntamente con la validez interna de la misma, dan credibilidad a la investigación.
- El tamaño de una muestra es directamente proporcional a los siguientes factores: cuasivarianza poblacional S_x^2 y nivel de confianza $(1 - \alpha)100\%$ e inversamente proporcional al error de muestreo δ .
- En muchas oportunidades en las cuales es supremamente costoso medir la unidad estadística, muy seguramente el único criterio para definir su tamaño será precisamente el presupuesto de que se dispone. No obstante esta circunstancia, al igual que en una situación diferente, se hace necesario estimar el margen de error cometido con la muestra utilizada.
- Dependiendo del parámetro investigado; del método de muestreo utilizado; las funciones de costo así como los requisitos deseados y las limitaciones teóricas y de recursos, la obtención del tamaño de muestra será más o menos complejo. En todo caso se debe aplicar siempre el criterio más científico y racional de que se disponga, balanceando las limitaciones teóricas, prácticas, económicas y de orden legal.
- Es aconsejable siempre una vez definido el tamaño de muestra n teórico, definir un nuevo tamaño de muestra basado en el primero, llamado muestra final de trabajo, el cual se calcula mediante la siguiente expresión:

$$n_T = n.(1 + p)$$

En donde p , es el porcentaje de mediciones (encuestas) que presentan inconsistencias, están incompletas o que simplemente no fue posible

COSTO DE TRABAJO DE CAMPO, VARIABILIDAD DEL ESTIMADOR Y TAMAÑO DE MUESTRA. SITUACIONES DE DEPENDENCIA.



medir. Este porcentaje se estima según estudios similares realizados con anterioridad o mediante la encuesta piloto. El utilizar una muestra ajustada de esta forma, permite al investigador asegurarse de que siempre tendrá el número de encuestas n que se requiere de acuerdo con las necesidades de la investigación. Debe tenerse en cuenta que el trabajar con muestras inferiores a las establecidas según los supuestos, acarreará incrementos en el error muestral establecido y/o bajas en la confiabilidad de la estimación del parámetro.

- Al especificar un tamaño de muestra n para la estimación de un parámetro poblacional como por ejemplo la media poblacional, debemos seguir los siguientes pasos:

1. Especificar de nivel de precisión o sea el máximo error de muestreo tolerable δ .
2. Especificar el nivel de confianza $(1 - \alpha)100\%$
3. Determinar el valor $Z_{\alpha/2}$ asociado al nivel de confianza en una tabla de distribución normal estándar.
4. Estimar en forma adelantada la cuasivarianza S_x^2 , para la variable analizada X .
5. Encontrar la primera aproximación n_0 mediante la expresión (8) antes vista.
6. Comparar la primera aproximación n_0 , con el tamaño N de la población analizada y realizar el ajuste según la siguiente norma:

$$n = \begin{cases} n_0, & \text{si } \frac{n_0}{N} < 5\% \\ \frac{n_0}{1 + \frac{n_0}{N}}, & \text{si } \frac{n_0}{N} \geq 5\% \end{cases}$$

7. El anterior ajuste recibe el nombre de ajuste por población finita y como puede apreciarse su efecto será reductor sobre la primera aproximación n_0 cuando la fracción de muestreo f , sea mayor o igual al 5%. Recordemos aquí que precisamente es una cota superior para la función tamaño de muestra.

- Cuando se realiza un muestreo complejo en donde las unidades objeto de estudio están agrupa-

das formando conglomerados de igual o diferente tamaño los cuales a su vez aparecen formando estratos, el uso del efecto de diseño *Deff* es de mucha utilidad en el cálculo del tamaño de muestra.

$$Deff = \frac{V(\hat{\bar{X}})_{COM}}{V(\hat{\bar{X}})_{MAS}}$$

Si al usar un MAS para la estimación de μ_x se obtiene el cumplimiento de una varianza específica $V(\hat{\bar{X}})_{MAS} = V_0$, este hecho permite establecer un tamaño de muestra para el cual se obtenga un *Deff* menor que 1, significando con ello que la varianza mediante la utilización del diseño complejo $V(\hat{\bar{X}})_{COM}$ satisface las exigencias de estimación de μ_x .

Bibliografía

1. Cochran, W. Técnicas de Muestreo. Cecs. México, 1980.
2. Fernández F., Mayor J. Muestreo en poblaciones finitas: Curso Básico. EUB. Barcelona 1995.
3. Klinger, R. MUESTREO «Filosofía, Importancia y Necesidad. Univalle, 2003.
4. Klinger, R. Elementos Básicos de la Teoría del Muestreo. Univalle, 2001.
5. Lohr, S. Muestreo: Diseño y Análisis. Thomson. México, 2000.
6. Malhotra, N. Investigación de Mercados «Un enfoque práctico». Prentice Hall. México, 1997.
7. Miras, J. Elementos de muestreo para poblaciones finitas. INE, Barcelona, 2000.
8. Pérez, C. Técnicas de Muestreo Estadístico. Rama. Madrid, 1999.
9. Santos, J., Muñoz A., Juez P y Cortinas P. Diseño de encuestas para estudios de mercado. Editorial Centro de Estudios Ramón Arces, S.A. Madrid, 2003



Universidad
del Valle

Escuela de Ingeniería
Industrial y Estadística

PROGRAMA DE FORMACIÓN PROFESIONAL EN PREGRADO



ESTADÍSTICA

Perfil profesional:

- Dirige, organiza, estructura y administra oficinas o departamentos de estadística en empresas privadas o del estado.
- Asesora en diseño y análisis de experimentos a las instituciones de investigación científica y tecnológica y a los profesionales de otros campos.
- Diseña, ejecuta y analiza encuestas e implementa las técnicas de muestreo necesarias en campos como: estudios de mercados, estudios de existencias, estudios socio-económicos, etc.
- Asesora en la definición y validación de modelos estadísticos – matemáticos para estudios investigativos en los cuales se requiera analizar el comportamiento simultáneo de varios factores aleatorios para la explicación de una problemática existente.
- Es un experto en el uso de bancos de datos con fines estadísticos y excelente asesor en procesamiento y análisis de información mediante el uso de software estadístico altamente especializado.

INFORMES:

Ciudad Universitaria Meléndez
Facultad de Ingeniería
Escuela de Ingeniería Industrial y Estadística
Edificio: 346
Teléfonos: 321 2196
Fax: 331 52 74
A.A. 25360
Cali, Colombia