

Algunas propiedades deseables en los estimadores que miden el parámetro involucrado en una investigación

Rafael Klinger Angarita, Javier Olaya Ochoa

Profesores de la Escuela de Ingeniería Industrial y Estadística - Facultad de Ingeniería - Universidad del Valle.

RESUMEN

El presente artículo pretende poner en conocimiento de los usuarios de la estadística, en especial aquellos que realizan investigaciones por muestreo, la importancia de dos propiedades fundamentales de los estimadores como lo son: **La precisión y la exactitud** y de otro lado realizar un resumen de los principales errores no muestrales y sus fuentes, mostrando su incidencia en las estimaciones mediante los llamados sesgos inducidos, los cuales conjuntamente con los errores de muestro son los responsables del **error total de muestreo**. Muchas veces el inadecuado control de los errores no muestrales hace que el error total se deba primordialmente a los sesgos y no al error de muestreo, haciéndose este último culpable de circunstancias y procedimientos que no le pertenecen. Un acertado diseño muestral, acompañado de una adecuada y minuciosa ejecución de todo el proyecto de investigación por muestreo aunque no conlleve a la ausencia total de los errores no muestrales, sí permitirá una buena disminución de estos, mejorando la exactitud de las estimaciones. La metodología del muestreo provee procedimientos científicos para la mejora de la precisión en las estimaciones más no así para la mejora de su exactitud, cuya responsabilidad en buena parte, recae en los sesgos generados por los errores no muestrales.

Variabilidad del muestreo y factores que determinan el tamaño de una muestra

La naturaleza del muestreo lleva intrínseca la variabilidad de las estimaciones obtenidas mediante cualquier diseño muestral. En este sentido, es apenas lógico que de muestra a muestra se presente variabilidad en las estimaciones sobre el parámetro de una población objeto de estudio en una investigación por muestreo.

De acuerdo con Cochran¹, una de las expresiones usadas en la determinación del tamaño de una muestra para la estimación de la media poblacional $\bar{X}$ de la característica X analizada viene dada por:

$$n_0 = \frac{Z_{\alpha/2}^2 \cdot S_x^2}{\delta^2} \quad (1)$$

Basándonos en la expresión anterior podemos conformar el siguiente cuadro que nos permite conocer el tipo de relación existente entre el tamaño de la muestra n_0 y cada uno de los elementos que la definen, cuando consideramos el resto de ellos constante. Aquí, n_0 , recibe el nombre de primera aproximación al tamaño de muestra y resulta de suponer que la población objeto de estudio es significativamente grande con respecto al tamaño de la muestra, vale decir en la práctica según Cochran² ocurre que:

$$f = \left(\frac{n_0}{N} \right) < 5\%$$

Tabla 1. Elementos que definen el tamaño de una muestra para la estimación de la media poblacional

Cuasi-varianza S_x^2	Máximo error tolerable δ	Nivel de Confianza (1- α).100%	Comportamiento del tamaño de muestra n_0	Relación existente
Grande	-	-	Grande	Directa
-	Pequeño	-	Grande	Inversa
-	-	Grande	Grande	Directa

¹ Cochran William, Técnicas de muestreo, México, ed. CECSA, 1980, pág. 109.

² Ibid, pág. 49.

Nota: El nivel de confianza está representado en la ecuación (1) por $Z_{\alpha/2}$.

En toda investigación por muestreo es deseable por supuesto que las estimaciones obtenidas mediante el uso del estimador se parezcan o estén lo más cerca posible del parámetro que se está estimando y para ello el investigador determina el máximo error tolerable (δ), que permite tal consideración. En el caso de la estimación de $\bar{X}$ por ejemplo el investigador al determinar δ establece la siguiente desigualdad:

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

En ella, está claro que las estimaciones dadas por el estimador $\hat{X}_n$ pueden ser menores, iguales o mayores que $\bar{X}$, esto nunca se sabrá, siendo lo único deseable que la diferencia entre los dos valores (estimación y parámetro) no vaya más allá del valor δ al tomar una muestra cualquiera. Claro está que existe un riesgo de que esto no suceda cuyo valor es de $100\alpha\%$. Así las cosas por ejemplo si se tiene un nivel de confianza del 95% que se cumpla la ecuación 2 al tomar una muestra, lo que ello significa es que si repitiéramos el muestreo en iguales circunstancias 100 veces esperaríamos que 95 de ellas cumplan con dicha desigualdad y 5 no.

Dos propiedades fundamentales de un estimador

En la práctica se dice que un estimador es "**preciso**", cuando las estimaciones que arroja para diferentes muestras son parecidas entre sí o lo que es lo mismo se parecen a su valor promedio (valor esperado del estimador).

Para algunos estimadores, puede suceder, que al tomar todos los posibles valores dados por él, el promedio de ellos no coincida con el parámetro investigado y en estas situaciones se dice que el estimador es "**sesgado**", mientras que en caso contrario será insesgado.

De acuerdo con lo anterior, podemos realizar la siguiente clasificación:

Tabla 2. Clasificación de los estimadores

Varianza de las estimaciones	Sesgo del estimador	Tipo de Estimador (distribución)
Pequeña	No tiene	Preciso e insesgado
Pequeña	Si tiene	Preciso y sesgado
Grande	No tiene	Impreciso e insesgado
Grande	Si tiene	Impreciso y sesgado

De acuerdo con la tabla 2, es deseable que el estimador que usemos en nuestra investigación sea preciso e insesgado. El cumplimiento de estas propiedades nos dará tranquilidad a la hora de establecer la estimación de nuestro parámetro, aunque como veremos se hace necesario además, controlar cualquier tipo de error no muestral para que nuestra estimación este cerca del parámetro, con alta probabilidad.

La distribución de un estimador:

Error cuadrático medio, varianza y sesgo

Para concretar un poco más estas propiedades de los estimadores, supongamos que se tiene una población de N unidades estadísticas y de ellas se extrae todas las muestras posibles de tamaño n sin reemplazo,

obteniéndose en total $K = \binom{N}{n}$ muestras.

Si en cada muestra aplicamos en general el estimador $\hat{\phi}_n$, obtenemos la distribución de la variable aleatoria $\hat{\phi}_n$ que podemos representar sencillamente por: $\{ \hat{\phi}_i, P_r(\hat{\phi}_i) \}$, en la cual se relaciona cada valor del estimador con su respectiva probabilidad.

Al calcular la variabilidad de las estimaciones con respecto a su valor esperado $E(\hat{\phi}_n)$, obtenemos la **Varianza del Estimador**:

$V(\hat{\phi}_n)$, mientras que si calculamos la variabilidad de ellas con respecto al parámetro ϕ , obtenemos el **Error cuadrático Medio del**

Estimador: $ECM(\hat{\phi}_n)$. Es bueno tener presente que para el caso de la media muestral como estimador de la media poblacional se cumple

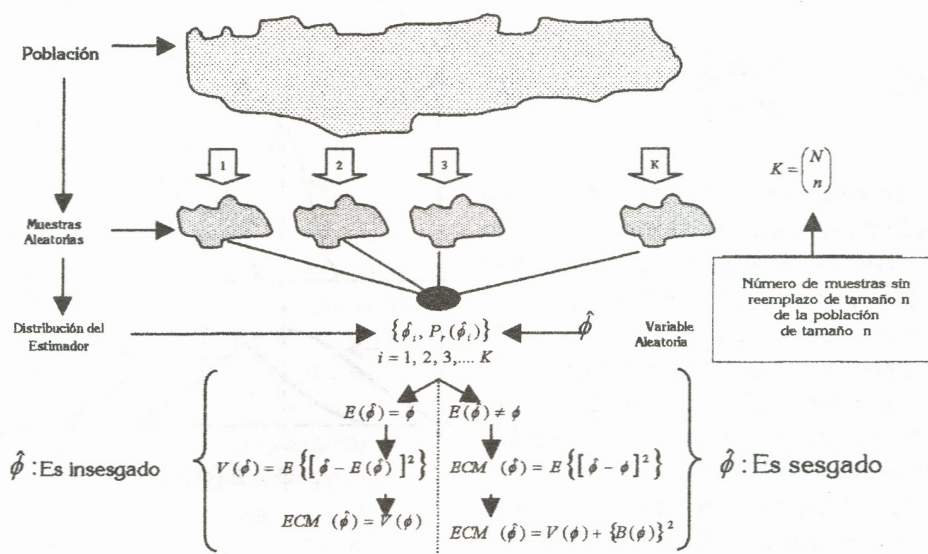


Figura 1. Distribución de un estimador

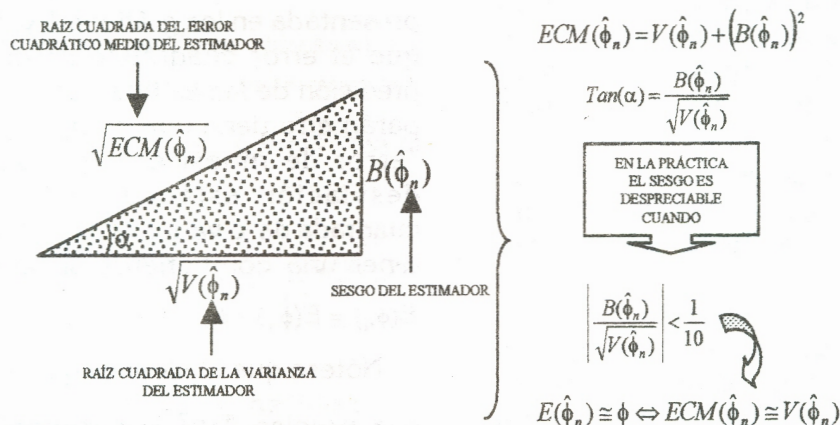


Figura 2. Relación entre la varianza, el ECM y el sesgo de un estimador

que: $V(\hat{X}_n) = ECM(\hat{X}_n)$, precisamente porque es un estimador insesgado.

La figura 1. esquematiza el proceso de obtención de la distribución de un estimador cualquiera y su clasificación como estimador sesgado o estimador insesgado.

La relación entre el ECM de un estimador, su varianza y su sesgo, puede representarse con ayuda del teorema de Pitágoras como lo muestra la figura 2:

Nota³: Si $\hat{\phi}_1$ y $\hat{\phi}_2$, son dos estimadores de ϕ para los cuales se cumple que:

$$ECM(\hat{\phi}_1) = V(\hat{\phi}_1) + B_1^2(\hat{\phi}_1) \text{ y } ECM(\hat{\phi}_2) = V(\hat{\phi}_2) + B_2^2(\hat{\phi}_2)$$

la eficiencia relativa de $\hat{\phi}_1$ respecto de $\hat{\phi}_2$ se define como el cociente:

³ Ruiz-Maya, Luis y Pliego, Martín: Fundamentos de Inferencia Estadística. Madrid, Editorial AC. 1999.

$$ER(\hat{\phi}_1; \hat{\phi}_2) = \frac{ECM(\hat{\phi}_1)}{ECM(\hat{\phi}_2)} \quad (3)$$

La eficiencia relativa (ER), nos dice qué estimador debemos aplicar siendo nuestro deseo que las estimaciones se parezcan lo más posible a nuestro parámetro investigado.

Con frecuencia se establece un cociente entre la varianza de un estimador siguiendo un diseño de muestreo DM_i y la varianza que se deriva del muestreo aleatorio simple (MAS). Este cociente se denomina *efecto de diseño* y permite hacer una estimación bastante buena del tamaño de muestra que se necesitaría con el diseño DM_i manejando la misma precisión establecida para el MAS.

Generalmente, la distribución de un estimador se considera Normal, salvo que la muestra sea pequeña, y se puede mostrar empíricamente que la distribución de un estimador se acerca rápidamente a la distribución Normal a medida que aumenta el tamaño de la muestra.

La proximidad que tiene la distribución de un estimador a la Distribución Normal, depende básicamente de los siguientes aspectos:

- La distribución de la característica X analizada (la forma de la distribución de frecuencias de X en la población muestreada).
- La forma del estimador.
- El diseño muestral (forma y características).
- El tamaño de la muestra.

No es posible o por lo menos fácil suministrar una guía sencilla y precisa para decidir cuando dicha aproximación es suficientemente buena. En la práctica generalmente se asume una distribución Normal pero debemos tener presente que esta situación no será tan cierta, si la muestra es pequeña y la distribución de la población es muy asimétrica.

La figura 3 muestra cuatro tipos de distribución de acuerdo con la variabilidad y la existencia o no del sesgo suponiendo distribución Normal para el estimador

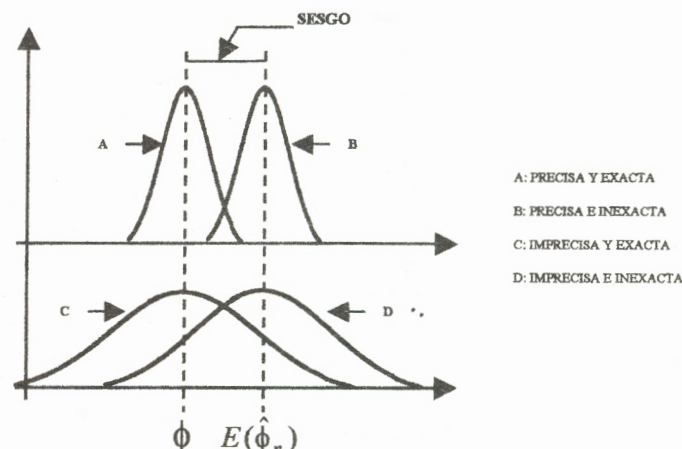


Figura 3. Clasificación de la distribución de un estimador suponiendo normalidad

De la ecuación: $ECM(\hat{\phi}_n) = V(\hat{\phi}_n) + [B(\hat{\phi}_n)]^2$, presentada en los gráficos 1 y 2, se desprende que el error cuadrático medio que mide la precisión de las estimaciones con respecto al parámetro tiene una componente aleatoria o "error de muestreo" que se mide por la desviación estándar del estimador (raíz cuadrada de la varianza del estimador) y puede tener una componente sistemática o sesgo

$$B(\hat{\phi}_n) = E(\hat{\phi}_n) - \phi.$$

Nótese que lo ideal es que el estimador $\hat{\phi}_n$, sea preciso " $V(\hat{\phi}_n)$ pequeña" e insesgado " $B(\hat{\phi}_n) = 0$ ", ya que siempre trabajaremos con una muestra la cual podrá sobrestimar o subestimar el parámetro investigado. Esto nunca se sabrá, pero de lo que sí debemos estar seguros es de que la estimación obtenida esté cerca del parámetro ϕ , con la probabilidad estipulada.

Lo anterior no significa que no sean válidos e importantes los estimadores sesgados, es más en la práctica, seguramente solo de estos dispondremos y por lo tanto debemos procurar que el sesgo de estimador sea despreciable, lo cual, de acuerdo con Cochran⁴, en la práctica significa que:

⁴ Ibid, pág.36.

$$\left| \frac{B(\hat{\phi}_n)}{\sqrt{V(\hat{\phi}_n)}} \right| < \frac{1}{10} \quad (4)$$

El sesgo del estimador puede ser inducido en una investigación por muestreo, bien sea por el tipo de estimador, ya que algunos estimadores son de naturaleza sesgada como es el caso de los estimadores de razón y de regresión para citar dos ejemplos, o el sesgo puede ser inducido en el estimador por la presencia de errores no muestrales. Para hacer claridad sobre sesgo natural y sesgo inducido veamos algunos ejemplos sencillos:

Supongamos que para la característica X analizada se tiene la siguiente población:

$$U = \{u_1(x_1=4), u_2(x_2=5), u_3(x_3=7), u_4(x_4=4)\}$$

Encontrando todas las posibles muestras de tamaño $n=2$, tendríamos la siguiente distribución para el estimador $\hat{X}_n$ de la media poblacional $\bar{X}$:

Muestras	(u_1, u_2)	(u_1, u_3)	(u_1, u_4)	(u_2, u_3)	(u_2, u_4)	(u_3, u_4)
$\hat{X}_n$	9/2	11/2	8/2	12/2	9/2	11/2

En esta distribución no se ha inducido ningún error no-aleatorio y además el estimador es de naturaleza insesgada, por lo tanto:

$$E(\hat{X}_n) = \frac{\sum \hat{X}_i}{6} = 5 = \bar{X} \therefore B(\hat{X}_n) = E(\hat{X}_n) - \bar{X} = 0$$

Ejemplo N° 1 (Sesgo de selección).

Vamos a suponer ahora que la unidad poblacional u_1 , es suplantada por la unidad u_2 , significando así que la primera tiene probabilidad cero de ser seleccionada. En estas circunstancias la distribución del estimador de la media poblacional será:

Muestras	(u_2, u_2)	(u_2, u_3)	(u_2, u_4)	(u_3, u_3)	(u_3, u_4)	(u_4, u_4)
$\hat{X}_n$	10/2	12/2	9/2	12/2	9/2	11/2

Y calculando el valor esperado para el estimador, obtenemos ahora que:

$$E(\hat{X}_n) = \frac{\sum \hat{X}_i}{6} = 5 + \frac{3}{12} \neq \bar{X} \therefore B(\hat{X}_n) = E(\hat{X}_n) - \bar{X} = \frac{3}{12}$$

Ejemplo N° 2 (Sesgo de medición)

Si ahora nuestra unidad u_1 no ha sido medida correctamente o como en el caso de una encuesta no se ha dicho la verdad, la distribución del estimador $\hat{X}_n$, al suponer que el valor de la unidad u_1 , esta alterado en una cantidad $e > 0$ será:

Muestras	(u_1, u_2)	(u_1, u_3)	(u_1, u_4)	(u_2, u_3)	(u_2, u_4)	(u_3, u_4)
$\hat{X}_n$	(9+e)/2	(11+e)/2	(8+e)/2	12/2	9/2	11/2

Encontrando en este caso con respecto al valor esperado del estimador que:

$$E(\hat{X}_n) = \frac{\sum \hat{X}_i}{6} = 5 + \frac{e}{4} \neq \bar{X} \therefore B(\hat{X}_n) = E(\hat{X}_n) - \bar{X} = \frac{e}{4}$$

Ejemplo N°3 (Estimador sesgado por naturaleza). Tomando como base la misma población U , la cual tiene para la característica X una media poblacional $\bar{X} = 5$, vamos ahora a suponer que utilizamos como estimador de la media poblacional, el estimador definido como el rango muestral multiplicado por $3/2$, esto es:

$$\hat{X}_n = \frac{3}{2} (x_{\max} - x_{\min})$$

Encontrando bajo este supuesto que la distribución del estimador de la media poblacional es:

Muestras	(u_1, u_2)	(u_1, u_3)	(u_1, u_4)	(u_2, u_3)	(u_2, u_4)	(u_3, u_4)
$\hat{X}_n$	3/2	9/2	0	6/2	3/2	9/2

Y por lo tanto su valor esperado será:

$$E(\hat{X}_n) = \frac{\sum \hat{X}_i}{6} = \frac{5}{2} \neq \bar{X} \therefore B(\hat{X}_n) = E(\hat{X}_n) - \bar{X} = -\frac{5}{2}$$

Comparando dos estimadores de acuerdo con su naturaleza

En muchas oportunidades estaremos en presencia de dos estimadores para el mismo parámetro ϕ , viéndonos obligados necesariamente a elegir una de las siguientes circunstancias de acuerdo con la naturaleza de los dos estimadores:

a) Ambos insesgados

En este caso para cada uno de ellos $\hat{\phi}_1$ y $\hat{\phi}_2$, el ECM coincide con su respectiva varianza y de este modo, el mejor de ellos será aquel que posea la menor varianza o el menor coeficiente de variación. Ver figura 4.

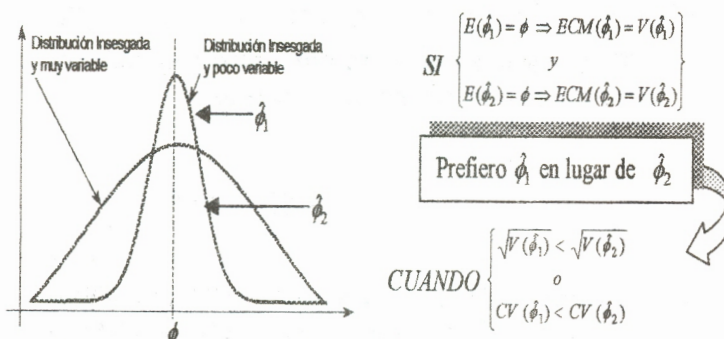


Figura 4 Comparación de dos estimadores insesgados

b) Ambos son sesgados

En este caso preferimos aquel que tenga menor ECM. Pero dado que la estimación de este indicador no siempre es fácil de obtener, se acostumbra utilizar el valor absoluto del cociente entre el sesgo y la raíz cuadrada de la varianza.

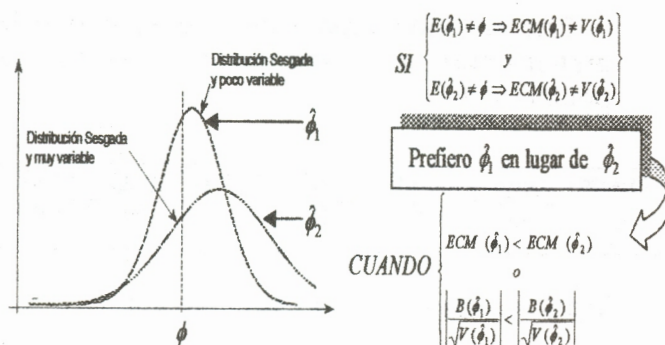


Figura 5. Comparación de dos estimadores sesgados

Debemos tener presente que si ambos estimadores cumplen con la condición establecida por la desigualdad (4) antes mencionada, la comparación se puede realizar como en el caso anterior, pues podemos considerar despreciable el sesgo en ambos estimadores.

c) Uno es insesgado y el otro sesgado

En este caso lo ideal será decidimos por aquel que tenga el menor ECM al igual que en el caso anterior, pero en caso de presentarse imposibilidad para la estimación del ECM, damos la favorabilidad a aquel que tenga menor coeficiente de variación. Ver figura 6.

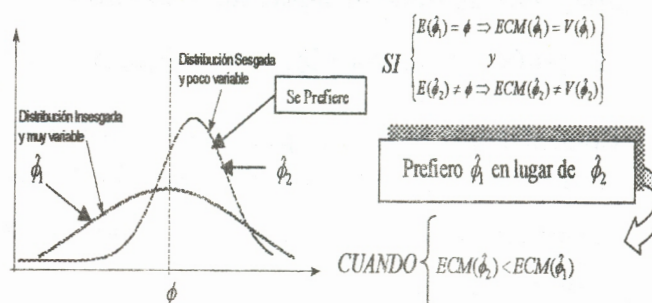


Figura 6. Comparación de un estimador insesgado y otro sesgado

¿Cuándo el sesgo de un estimador es insignificante?

La conclusión mediante la cual despreciamos el sesgo de un estimador, cuando el valor absoluto del cociente entre el sesgo y la raíz cuadrada de la varianza del estimador es menor del 10%, podemos verificarla de la siguiente manera, después de suponer normalidad para la distribución del estimador.

Sea $\hat{\phi}_n$ el estimador insesgado de ϕ para el cual $\hat{\phi}_n \approx N(E(\hat{\phi}_n), V(\hat{\phi}_n))$. Estandarizando tenemos:

$$Z = \frac{\hat{\phi}_n - E(\hat{\phi}_n)}{\sqrt{V(\hat{\phi}_n)}} = \frac{\hat{\phi}_n - \phi}{\sqrt{V(\hat{\phi}_n)}} \approx N(0, 1) \text{ para la cual}$$

sabemos que: $P_r(Z \leq 2.0) = 0.9772$

Ahora consideremos que el estimador $\hat{\phi}_n$ es sesgado, en cuyo caso tenemos:

$$Z = \frac{\hat{\phi}_n - \phi}{\sqrt{V(\hat{\phi}_n)}} = \frac{\hat{\phi}_n - E(\hat{\phi}_n) + B(\hat{\phi}_n)}{\sqrt{V(\hat{\phi}_n)}} = \frac{\hat{\phi}_n - E(\hat{\phi}_n)}{\sqrt{V(\hat{\phi}_n)}} + \frac{B(\hat{\phi}_n)}{\sqrt{V(\hat{\phi}_n)}} = Z^* + \frac{B(\hat{\phi}_n)}{\sqrt{V(\hat{\phi}_n)}}$$

Y al intentar calcular la $P_r(Z \leq 2.0)$, llegamos a la expresión:

$$P_r(Z \leq 2.0) \Leftrightarrow P_r\left(Z^* + \frac{B(\hat{\phi}_n)}{\sqrt{V(\hat{\phi}_n)}} \leq 2.0\right) = P_r\left(Z^* \leq 2.0 - \frac{B(\hat{\phi}_n)}{\sqrt{V(\hat{\phi}_n)}}\right) = ?$$

Para la cual usando el paquete estadístico MINITAB, obtenemos los valores de la tabla siguiente:

Tabla 3. Insignificancia del sesgo cuando $\left| \frac{B(\hat{\phi})}{\sqrt{V(\hat{\phi})}} \right| < \frac{1}{10}$

$A = \left \frac{B(\hat{\phi})}{\sqrt{V(\hat{\phi})}} \right < \frac{1}{10} = 0.1$	$P_r(Z^* \leq 2.0 - A)$
0.00	0.977250
0.02	0.976148
0.04	0.975002
0.06	0.973810
0.08	0.972571
0.10	0.971283
0.20	0.964070
0.30	0.955435
0.40	0.945201

MUCHA SIMILITUD

Mediante los cálculos anteriores podemos concluir que la distorsión introducida por el sesgo del estimador no es importante cuando:

$$\left| \frac{B(\hat{\phi}_n)}{\sqrt{V(\hat{\phi}_n)}} \right| < 0.1$$

En las investigaciones por muestreo, los errores de medición y de selección inducen

sesgos en las estimaciones deteriorando la proximidad de ellas al parámetro investigado.

En general se dice que "Sesgo es el efecto que priva a un resultado estadístico de representatividad al distorsionarlo sistemáticamente"⁵, esta definición lo que significa matemáticamente es que sesgo es todo aquello que induce a que el promedio de las estimaciones sobre un parámetro no coincida con el parámetro.

De acuerdo con Kish⁶-Se usan ampliamente los términos *exactitud* y *precisión* para separar los efectos del sesgo. En general, la precisión se refiere a los errores pequeños y variables; a veces denota únicamente el valor inverso de la varianza; en cualquiera de los casos, excluye los efectos de los sesgos. La exactitud se refiere a los errores totales pequeños, e incluye los efectos del sesgo. En un diseño preciso debe haber errores variables pequeños; pero un diseño exacto, además de ser preciso, ha de tener un sesgo de cero o de valor muy pequeño. Un diseño de encuesta con sesgo grande será preciso si sus errores variables son pequeños, aunque no será exacto-

De acuerdo con esta apreciación podemos decir que *El error total de estimación* (ET) en que se incurre en cualquier investigación por muestreo, depende de dos aspectos: *El error de muestreo* (Raíz cuadrada de la varianza del estimador) y *El sesgo de estimación*. La relación entre estos elementos puede representarse mediante un triángulo rectángulo para el cual, de acuerdo con el teorema de Pitágoras:

$$(ET(\hat{\phi}_n))^2 = V(\hat{\phi}_n) + (B(\hat{\phi}_n))^2 \quad (5)$$

⁵ Kendall, Maurice y Buckland, William: Diccionario de Estadística. Madrid. Ediciones Pirámide, S.A. 1980.

⁶ Kish Leslie, Muestreo de encuestas, México, ed trillas, 1979, pág. 588.

Mirando el efecto del sesgo y el error de muestreo en el error total de estimación al usar dos estimadores de un mismo parámetro.

La figura 7 muestra la posible incidencia del sesgo y el error de muestreo en el error total de estimación cuando tratamos de estimar un parámetro mediante dos estimadores diferentes.

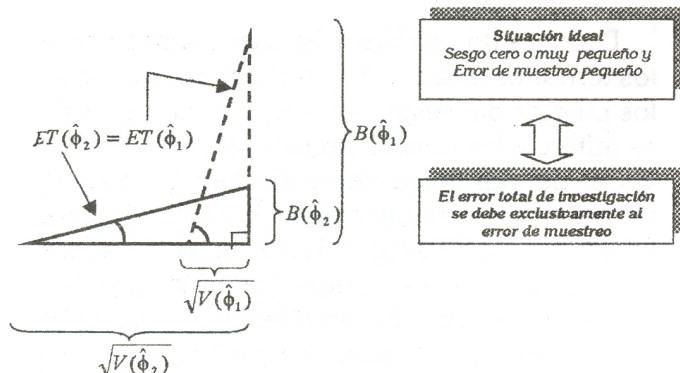


Figura 7. Efecto del sesgo en el error total de estimación

La figura 7, muestra el mismo error total para los estimadores $\hat{\phi}_1$ y $\hat{\phi}_2$, pero en el caso del estimador $\hat{\phi}_1$, dicho error está explicado sustancialmente por el sesgo de estimación aunque el error de muestreo está controlado (poca variabilidad en el muestreo), mientras que para el estimador $\hat{\phi}_2$, el sesgo está controlado pero la variabilidad del muestreo no, haciendo que el error total se deba fundamentalmente al error de muestreo.

Así las cosas, podemos entonces clasificar los errores en las investigaciones por muestreo en dos clases: *Errores muestrales* y *Errores no-muestrales*, estos últimos causantes de los sesgos inducidos. La figura 8 nos ayuda a la caracterización de cada uno de ellos.



Figura 8. Caracterización de los errores en las investigaciones por muestreo

Los errores no muestrales son tan importantes como los errores muestrales, pero debido a la poca atención que se les presta tiran al traste la mejor investigación por muestreo, al punto que debido a tal desconocimiento se le atribuye al muestreo errores que no le pertenecen.

En términos generales, un error no muestral es la diferencia entre el verdadero valor para una característica analizada en una unidad investigada y su valor reportado en la hoja de recolección de información o encuesta. Las causas o fuentes que determinan estos errores se pueden clasificar en tres como se muestra en la figura 9:

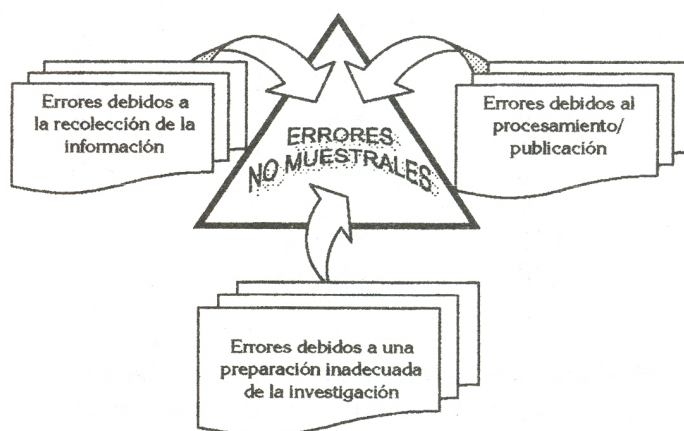


Figura 9. Principales fuentes de los errores no muestrales

Los errores no muestrales debidos a una preparación inadecuada de la investigación, pueden a su vez clasificarse en: a) Errores que

generan procedimientos sesgados debido a la medida de una característica, a la selección de una unidad o a la estimación de un parámetro. b) los que se generan mediante el uso de instrumentos sesgados, que pueden a su vez provenir de la conformación o mal uso de tablas de números aleatorios; cuestionarios con preguntas mal redactadas o diseñadas, instrucciones equivocadas impartidas a los recolectores y del uso equivocado de un marco de muestreo o de sus deficiencias. Ver figura 10.

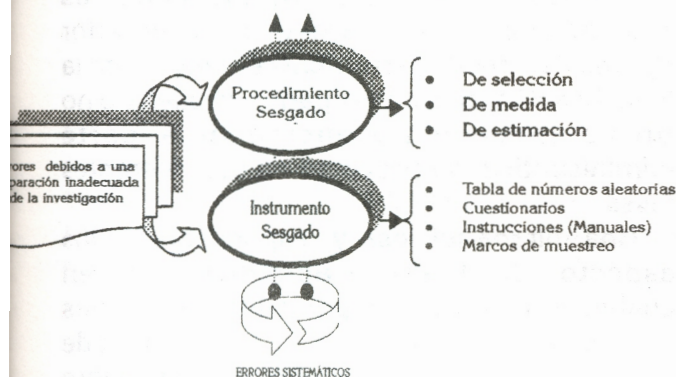


Figura 10. Clasificación básica de los errores debidos a una preparación inadecuada de la investigación:

Los errores no muestrales debidos a la recolección de la información, pueden clasificarse en tres a saber: a) los debidos a los defectos del marco de muestreo, que pueden darse por el uso de marcos ineficientes o de un mal uso de estos. b) los llamados DATOS FALTANTES, que son un tipo especial de errores y que ocurren cuando para una unidad estadística investigada no se tiene la información solicitada total o parcialmente "no se puede medir" (ejemplo en el caso de las encuestas: La unidad no se puede ubicar, la unidad no quiere contestar, etc.) y c) los debidos a los errores en las respuestas o mediciones, los cuales pueden ser: humanos (ejemplo: fatiga, falta de interés, etc.); de temporada (ejemplo: festividades, vacaciones, etc.) o de tendencia (la memoria del entrevistado, la experiencia del encuestado). Ver figura 11.

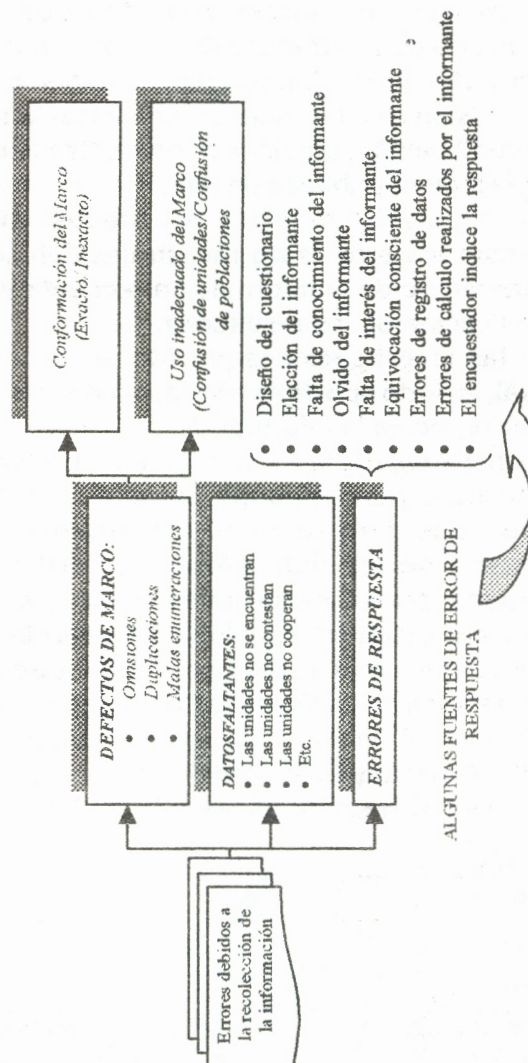


Figura 11. Errores no muestrales debidos a la recolección de la información

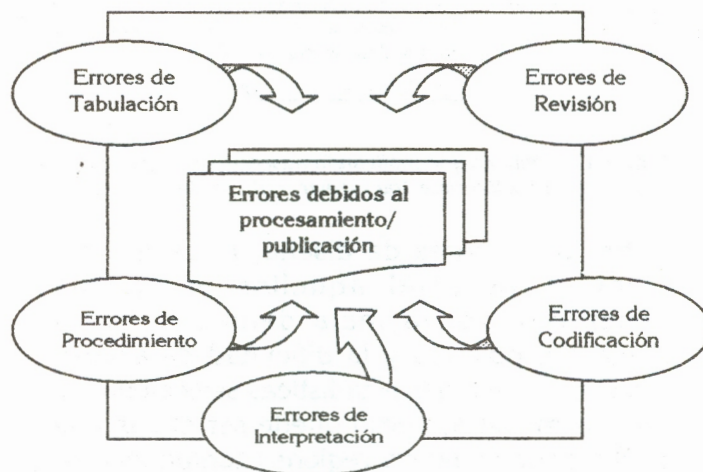


Figura 12. Fuentes de errores no muestrales debidos al procesamiento y/o publicación de resultados

Los errores no muestrales debidos al *tratamiento de la información recogida y a la publicación de resultados*, pueden deberse a: mala revisión de cuestionarios, equivocaciones en la codificación, tabulación, procesamiento, interpretación, publicación, etc.

La figura 12 muestra las fuentes que alimentan los errores no muestrales debidos al tratamiento de la información recogida y a la publicación de los resultados.

En las investigaciones por muestreo y en general, en todo tipo de encuesta, puede ocurrir que no se recoja la información para todas las unidades investigadas o esta información resulte nula para algunas preguntas. Este hecho, que perjudica notablemente la obtención de los indicadores estadísticos planeados dentro de la investigación, está motivado básicamente por los siguientes cinco factores: a) los motivados por el encuestador, b) los debidos al diseño de la investigación, c) los debidos al encuestado, d) los debidos al tipo de encuesta y e) los debidos a la forma de recolección de la información. Ver figura 13.

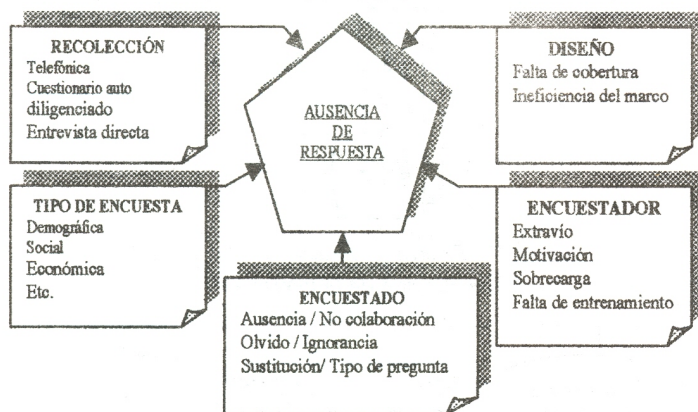


Figura 13. Factores que motivan la ausencia de respuesta en una investigación muestral mediante el uso de encuestas

En los factores de diseño, merecen una mención especial aquellos problemas generados por la mala definición de la población objetivo y la dificultad de acceso físico a las unidades estadísticas seleccionadas para la muestra, como puede ser el caso de unidades ubicadas en regiones contaminadas, regiones selváticas y regiones en conflicto militar o social.

A manera de conclusión

Todo investigador al realizar un proyecto por muestreo debe valorar juiciosamente en cuanto al diseño muestral los siguientes aspectos: Método de muestreo; tipo de parámetro que se investiga; la función de los costos involucrada; la obtención o tenencia de información adicional; el tamaño de la población investigada; la variabilidad de la característica o características investigadas; los niveles de confianza y los errores máximos tolerables (distancia permitida entre las estimaciones y el parámetro). La anterior rigurosidad puede derrumbarse invalidando la investigación, si el investigador, desdeña o no presta la debida atención a la parte administrativa del proyecto en sus diferentes fases.

Una mirada juiciosa y responsable de los aspectos de diseño y administración en cualquier investigación por muestreo son los que determinan la precisión y la exactitud de las estimaciones involucradas en ella. Sobre todo el investigador debe controlar la varianza del estimador, ya que esta controla la longitud del intervalo de confianza que contiene al parámetro y el sesgo, por que este hace que las estimaciones no estén cerca del parámetro, esto es, reducen la exactitud de las estimaciones.

Glosario de términos:

Cuasivarianza:

División de la suma de los cuadrados de desviaciones de un grupo de datos muestrales o poblacionales con respecto a su media, entre el número de ellos menos uno. Al igual que la varianza poblacional es un indicador de variabilidad. La cuasivarianza poblacional y la varianza poblacional coinciden cuando la población es muy grande o infinita.

Exactitud:

Un estimador es exacto cuando puede producir estimaciones sobre un parámetro carentes de sesgo

Estimador:

Regla o método de estimación de un parámetro para una población definida. Es

usualmente expresado como una función de valores muestrales. En el muestreo probabilístico, el estimador es una variable aleatoria, cuya distribución es de gran importancia para valorar la fiabilidad de la estimación a la que él conduce.

Estimación:

Es el proceso mediante el cual, utilizamos datos muestrales para estimar o realizar pruebas de hipótesis sobre los valores de los parámetros desconocidos en una población.

Error Cuadrático Medio(ECM):

Parámetro que mide en los estimadores sesgados las desviaciones de las estimaciones con respecto al parámetro investigado. Su cuadrado es igual a la suma de la varianza del estimador más el cuadrado del sesgo. Para los estimadores insesgados el ECM coincide con la varianza del estimador.

Error total de Investigación:

Para las investigaciones por muestreo se define como la raíz cuadrada de la suma del cuadrado del error muestral con el cuadrado de los errores no muestrales, medidos estos a través de los sesgos inducidos en la investigación.

Fracción de Muestreo:

Proporción establecida entre el tamaño de la muestra y el tamaño de la población. Suele llamarse también proporción o intensidad del muestreo y básicamente indica cuantas unidades poblacionales están siendo representadas por cada unidad muestral.

Inferencia estadística:

Parte de la Estadística que nos permite hacer juicios respecto de características poblacionales, basándonos en una muestra probabilística de dicha población.

Marco de Muestreo:

Es todo dispositivo material o electrónico en el cual se hallan representadas directa o indirectamente las unidades objeto de estudio.

Muestreo Probabilístico:

Método formal basado en la teoría de la probabilidad que permite la obtención de una muestra representativa. Muestreo y probabilidad están estrechamente ligados y juntos constituyen la base de la Inferencia Estadística.

Parámetro:

Valor que sintetiza la distribución de una población. Uno de los objetivos del análisis estadístico consiste en estimar los parámetros poblacionales a partir de las observaciones muestrales.

Precisión:

Es la variabilidad de las observaciones muestrales con respecto a su valor esperado y se mide por la varianza del estimador.

Población Objetivo:

Es el conjunto de unidades en espacio y tiempo que cumplen con determinadas propiedades definidas en una investigación.

Sesgo:

Es el efecto que priva a un resultado estadístico de representatividad al distorsionarlo sistemáticamente. Se dice que un estimador es de naturaleza sesgada cuando el promedio de sus estimaciones no coincide con el parámetro medido.

BIBLIOGRAFÍA

- AMBROSIO FLÓRES, L., *Muestreo*. Universidad Politécnica de Madrid. 1999.
- AZORIN, F., SÁNCHEZ CRESPO, J. L., *Métodos y Aplicaciones del Muestreo*. Alianza Editorial. España. 1986.
- BAUTISTA, L., *Diseños de Muestreo Estadístico*. Universidad Nacional de Colombia. Colombia 1998.
- COCHRAN, W. G., *Samling Teniques*. John Wiley & Sons. EEUU. 1980.
- DICCIONARIO ESPASA. España. 2001
- FERNÁNDEZ GARCÍA, F. R., MAYOR GALLEGU, J. A., *Muestreo en poblaciones finitas: Curso Básico*. EUB. España. 1995.

- HERNÁNDEZ SAMPIERI, R., FERNÁNDEZ COLLADO, C., BAPTISTA COLLADO, P., *Metodología de la Investigación*. Mc Graw Hill. México. 1991.
- KENDALL, M., BUCKLAND, W., *Diccionario de Estadística*. Ed, Pirámide. Madrid. 1976.
- KLINGER ANGARITA, R., *Elementos Básicos de la Teoría del Muestreo*. Universidad del Valle. Cali. 2001.
- MANUAL DEL EMPADRONADOR Y SUPERVISOR. DANE. Bogotá. 1993.
- MÉNDEZ, R., *Valoración Estadística en la Investigación*. UNAM. México. 1993.
- MENDENHALL, W., *Estadística para Administradores*. Grupo Editorial Iberoamérica. México. 1990.
- METODOS DE EVALUACIÓN DE ERRORES EN LOS CENSOS Y EN ENCUESTAS. DANE. Bogotá.
- MORGANSTEIN, D., *El método de Replicación para la Estimación de Errores de Muestreo*. Eustad. España. 1998.
- MIRAS, J., *Elementos de Muestreo Para Poblaciones Finitas*. INE. España. 1985.
- OTAÑEZ, G., *Metodología del Muestreo Agrícola de Áreas en Colombia*. PNUD/FAO. Colombia. 1988.
- PÉREZ, C., *Técnicas de Muestreo Estadístico*. Grupo Editor Alfaomega. México. 2000
- SARNDAL, C. E. SWENSSON, B., WRETMAN, J., *Model Assisted Survey Sampling*. Springer – Verlag 1992
- SÁNCHEZ CRESPO, J.L., *Curso intensivo de Muestreo en Poblaciones Finitas*. INE Madrid 1980.
- SIERRA BRAVO, R., *Técnicas de Investigación Social*. Editorial Paraninfo. España. 1994.
- SCHEAFFER, R., MENDENHALL, W., OTT, L., *Elementos de Muestreo*. Grupo Editorial Iberoamericano. 1987.
- SHARON, L. L., *Muestreo "Diseño y Análisis"*. International Thomson Editores. México. 2000.
- RAO, J.N.K., *Metodología Estadística para Estimaciones Indirectas en pequeñas áreas*. Eustat. España. 2000.
- RUIZ- MAYA, L., MARTÍN PLIEGO, F., *"Fundamentos de inferencia Estadística"*. Editorial. AC. Madrid. 1999.
- RUEDA GARCÍA, M., ARCOS CEBRIAN, A., *Problemas de Muestreo en Poblaciones Finitas*. GEU. 1998.
- WOLTER, K. M., *Introduction to Variance Estimation*. Springer – Verlag. EEUU. 1985
- WEISS, J. W., WYSOCKI, R. K., *Dirección de Proyectos "Las 5 Fases de su Desarrollo"*. Addison - Wesley Iberoamericana. EEUU. 1994.
- WILBURN, A.J., *Practical Statistical Sampling For Auditors*. Marcel Dekker, Inc. EEUU. 1984.