

EL USO DE LA ESTADÍSTICA EN LAS ENCUESTAS DE OPINIÓN: RECOMENDACIONES METODOLÓGICAS PARA EVITAR ERRORES

Javier Olaya Ochoa

*Universidad del Valle, Escuela de Ingeniería Industrial y Estadística, Cali.
jolaya@pino.univalle.edu.co*

Rafael A. Klinger Angarita

*Universidad del Valle, Escuela de Ingeniería Industrial y Estadística, Cali.
rakling@pino.univalle.edu.co*

Resumen

El ciudadano de hoy vive cada día más inmerso en indicadores y gráficos que no solamente le informan sino que le permiten tomar decisiones oportunas con errores mesurables. Ya sea que se trate de la administración de una región, de una empresa o de conocer los gustos y la forma de pensar de los ciudadanos, para llegar a la obtención de esos indicadores sobre comportamientos y hábitos de los seres humanos o la medición de fenómenos en cualquier campo del conocimiento, se requiere de la realización de proyectos de investigación, los cuales necesitan de un riguroso y adecuado uso de técnicas estadísticas. El artículo describe algunos de los cuidados que deben tenerse con el uso de la estadística en los estudios de opinión para garantizar su validez.

Palabras clave: Validez estadística, Marco muestral, Error de muestreo, Fracción de muestreo, Sesgo.

Abstract

Today's citizen lives more and more immersed in indices and graphics that not only give him daily information, but also allow him to make timely decisions with measurable accuracy levels. The same could happen, for example, to a private business manager or a public administrator or to a researcher on people's tastes and ways of thinking. Getting a hinge of human beings habits and behaviors requires the previous achievement of a series of research projects with a strict and suitable use of statistical techniques. We describe some of the cautions that should be taken into account when using Statistics in opinion polls, to assure the results validity.

Keywords: Statistical validity, Sampling frame, Sampling error, Sampling fraction, Bias.

Artículo de reflexión
Clasificación JEL: C830

1. Introducción

El uso de las técnicas estadísticas por sí solas no constituye un pasaporte de legitimidad y validez en la realización de un estudio estadístico. Desafortunadamente muchos usuarios de las encuestas de opinión utilizan la Estadística para mostrar que su investigación tiene validez pero desconocen y/o desdeñan el uso cuidadoso de esta.

Por validez y fiabilidad debe entenderse la capacidad de credibilidad que poseen las conclusiones halladas mediante un estudio estadístico, para decirlo en una palabra: el estudio es digno de confianza. Esta confianza desde luego dependerá de las condiciones establecidas y el uso de las técnicas estadísticas dispuestas para el logro de los objetivos planteados.

Los criterios de validez de un estudio estadístico (D'Anaconda, 2004) son básicamente cuatro (adaptados por los autores):

Validez interna: hace referencia al control de explicaciones alternativas en el establecimiento de las relaciones de causalidad, esto es ¿Cuáles son los criterios con los cuales se explica la causalidad de un fenómeno?, ¿Se ha establecido correctamente la causalidad del fenómeno(s) analizado(s)?

Validez externa: Apunta a la externalización de las relaciones observadas, es decir a la generalización de los resultados del estudio hacia la población de la cual se extrajo la muestra analizada, ¿Es la muestra utilizada en el estudio “representativa” de la población?

Validez de conclusión estadística: Tiene que ver con la significatividad de las pruebas estadísticas utilizadas en el estudio. ¿Son los niveles de significación y las pruebas estadísticas utilizadas, las adecuadas para el estudio?

Validez de constructo: Hace referencia a la forma de medir los conceptos teóricos

involucrados en los medios de captación de información. ¿Miden las preguntas establecidas el concepto para el cual fueron diseñadas?

2. El Uso del Muestreo en la Obtención de Información

En una investigación de opinión una vez establecida la unidad que debe ser medida (el tipo de informante) y las características que deben observarse, se hace necesario para la utilización del muestreo en la consecución de la información tendiente a satisfacer unos objetivos preestablecidos dar respuesta a las siguientes preguntas:

¿De cuántas unidades se compone el colectivo (población) del cual se extraerá la muestra?

¿Qué indicador poblacional (parámetro) constituye el objeto central de conocimiento?

¿Qué margen de error (por encima o por debajo) se hace tolerable para la estimación del parámetro?

¿Con qué nivel de confianza se espera el cumplimiento del error preestablecido?

¿Se dispone de información adicional (características correlacionadas con la característica analizada; resultados de estudios anteriores, etc.)?

¿De cuánto dinero se dispone para la toma de la información?

¿Bajo qué restricciones de tipo metodológico, legal y de tiempos se debe realizar el estudio?

De la claridad que se tenga en las respuestas a las preguntas antes mencionadas se establecerá el método de muestreo que se utilizará, el tamaño de la muestra y la forma o estrategia metodológica como serán seleccionadas y medidas las unidades que la componen. La Tabla 1 muestra las principales características de los métodos básicos de muestreo probabilístico que pueden usarse en la obtención de información para la medición. Desde luego la escogencia de uno cualquiera

Tabla 1. Método de muestreo: Procedimientos metodológicos.

		Procedimiento Metodológico	Observaciones
Todas las unidades o elementos de la población objeto de estudio tienen una probabilidad diferente de cero para pertenecer a la muestra Todas las combinaciones posibles de unidades elementales tienen igual probabilidad en la muestra	Muestreo Aleatorio Simple (MAS)	1) Determinar un marco de muestreo adecuado 2) Se asigna un número a cada unidad poblacional de 1 a N. 3) Mediante una tabla de números aleatorios o un procedimiento equivalente se seleccionan las unidades que van a constituir la muestra de trabajo.	• Cada unidad se elige en forma independiente de los demás. • Garantiza igual probabilidad de inclusión en la muestra para cualquier unidad. • Muy eficiente en poblaciones homogéneas (muestra representativa) • Es la base para diseños muestrales complejos • El procedimiento de selección de las unidades es fácil. • Se requiere un listado de todas las unidades que conforman la población, lo cual puede ser muy inconveniente cuando esta es muy grande. • Puede producir alto costo en la ubicación de la unidad seleccionada debido al tiempo invertido en la ubicación de la unidad.
	Sistemático (MSIS)	1) Determinar un marco de muestreo adecuado (lista de unidades del universo) o, establecer la ruta que estas unidades deben seguir. 2) Seleccionar la primera unidad aleatoriamente en un intervalo entre 1 y k.(cociente de dividir N entre n) 3) Seleccionar cada unidad restante después de la primera a saltos cada k unidades siguiendo la lista o la ruta.	• El primer elemento de la muestra se selecciona aleatoriamente, los restantes n-1 quedan determinados por el primero. • Todas las unidades poblacionales deben pertenecer a la ruta elegida o marco muestral. • La selección de las unidades es sencilla • La selección puede en muchos casos ser más representativa que la obtenida mediante un MAS. • Cuando la población aparece en el marco o la ruta ordenada en forma aleatoria, este método se asimila al MAS. • El procedimiento de selección de las unidades no es costoso. • Si hay algún tipo de sesgo en el marco (lista) o ruta, que esté afectando a la primera unidad seleccionada puede que el resto de unidades queden afectadas debido a la periodicidad de selección. • El cálculo de la varianza de las estimaciones no es fácil de estimar adecuadamente. • Da buena precisión cuando las unidades poblacionales están en el marco o ruta, ordenadas según la variable observada.
	Estratificado (MAE)	1) Determinar un marco muestral adecuado para cada estrato. 2) Dividir el universo o población objeto de estudio en grupos (estratos) internamente homogéneos. 3) Seleccionar dentro de cada estrato un grupo de individuos en forma aleatoria 4) Las fracciones de muestreo (representatividad) en cada estrato pueden ser diferentes según la necesidad.	• Utiliza el MAS o el MSIS como soporte. • Requiere información auxiliar de una o varias variables preferiblemente correlacionadas con la variable objeto de estudio. Estas variables servirán de base para la estratificación de la población de unidades. • Garantiza la representatividad de todas las clases de unidades establecidas (estratos). • Se reduce el error de muestreo. • La selección se hace de grupos homogéneos • Si los estratos son bastante homogéneos será posible una muestra pequeña y por lo tanto habrá más economía. • Puede ser difícil la ubicación o conformación de estratos homogéneos en algunos casos. • Se requiere un conocimiento previo de la población seleccionada (más variables). • Si la incidencia es baja (pocos con la características analizadas) puede ser difícil encontrar la muestra en algún estrato. • Puede ser costosa la consecución de los marcos muestrales (listas) de unidades en cada estrato. • Puede demandar más tiempo que los métodos anteriores. • Muy útil cuando la población es heterogénea. • Es más costoso que el MAS y el MSIS.
	Conglomerados (MCON)	1) Determinar marco muestral adecuado. 2) Dividir el universo en grupos (conglomerados) de igual o diferente tamaño de unidades elementales 3) Los conglomerados deben ser altamente heterogéneos. 4) Se selecciona una muestra de conglomerados basada en el marco. 5) Dentro de cada conglomerado puede realizarse censo o a su vez seleccionar una muestra de ellos (muestreo bietápico).	• Utiliza el MAS o el MSIS como soporte. • No se requiere de listados para todas las unidades poblacionales (elementales). Requiriéndose solo para aquellas que pertenecen al conglomerado seleccionado. • Ahorra dinero y tiempo al permitir la concentración de las unidades en áreas próximas (conglomerados). • Los costos se reducen si los conglomerados están bien definidos. • Puede exigir tratamientos estadísticos más complejos. • Hay pérdida de precisión en las estimaciones. • Si los conglomerados son heterogéneos habrá una buena representación de la población. • Puede ser difícil la ubicación de cada unidad elemental en el conglomerado. • Muy impreciso cuando los conglomerados son homogéneos. • Por lo general aumentar el número de conglomerados seleccionados, disminuye la varianza del estimador

de los métodos dependerá de qué tanto él se adecua a las necesidades concretas y específicas de la situación investigada (Rojas et al., 1998).

El método que se escoja para la obtención de la muestra en una población debe estar diseñado sobre el conocimiento de esta. Es aconsejable para la misma situación plantear varios escenarios y decidir sobre la base de costo-beneficio y practicidad (Martínez, 2003).

La confianza que debemos tener en el muestreo probabilístico como herramienta básica para la obtención de información con miras a inferir sobre una población está fundamentada en lo siguiente:

- Todo trabajo con muestras requiere un conocimiento previo del universo o población de la cual se extraerá la muestra. En éste sentido la muestra es contextual.
- La similitud entre las unidades que conforman una población objeto de estudio no son propiedades dadas sino distinciones hechas por el investigador pues es este quien decide que observar sobre la base de lo conocido.
- Si $X_1, X_2, \dots, X_n$ son n variables aleatorias independientes y distribuidas idénticamente, vale decir todas tienen igual función de densidad en el caso continuo (variables) o igual función de probabilidad en el caso discreto (atributos) y todas ellas poseen la misma media μ y la misma varianza σ^2 , finitas, entonces se cumple la Ecuación 1 para la suma de las n variables $SUM_n = \sum_{i=1}^n X_i$.

$$\lim_{n \rightarrow \infty} \left\{ Pr \left(a \leq \frac{SUM_n - n\mu}{\sigma/\sqrt{n}} \leq b \right) \right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-u^2/2} du \quad (1)$$

Lo anterior quiere decir que la variable aleatoria media muestral $\bar{X}_n = SUM_n/n$, se distribuye normal, esto es $\bar{X}_n \approx N(\mu, \sigma^2/n)$. Dicho más claramente cuando el tamaño de una muestra aumenta, la distribución de la media muestral se comporta como una normal sin importar la distribución poblacional de la cual se extrae la muestra. El resultado anterior

es uno de los fundamentos más importantes del muestreo probabilístico y recibe el nombre de *Teorema del Límite Central*.

- Bajo las mismas premisas del *Teorema del Límite Central* expuestas anteriormente, se puede afirmar que se cumple la expresión 2.

$$\lim_{n \rightarrow \infty} Pr\{|\bar{X}_n - \mu| > \varepsilon\} = 0 \quad (2)$$

Lo cual significa que a medida que el tamaño de muestra se incrementa, la diferencia entre la media muestral (estimación) y la media poblacional verdadera (parámetro) disminuye, es decir se obtiene cada vez más una mejor precisión (nos acercamos más a la verdad). Esta última propiedad de las variables aleatorias se llama *Ley de los Grandes Números*.

- El tamaño de una muestra no depende sustancialmente del tamaño de la población esto es, dos poblaciones ampliamente diferentes en cuanto a su tamaño pueden requerir el mismo tamaño de muestra (Pérez, 2000). Este hecho puede verificarse observando que si la población es muy grande con respecto a la muestra (infinita) entonces la llamada fracción de muestreo $f = n/N$ es despreciable
- El tamaño de una muestra depende de varios factores que se presentan en la Tabla 2. Esta dependencia indica a las claras la sensibilidad del tamaño de muestra para la obtención de estimaciones sobre parámetros desconocidos:

Tabla 2. Factores que determinan el tamaño de muestra.

Factor	Dependencia
Variabilidad de la población	DIRECTA
Precisión deseada	INVERSA
Confianza establecida para que se cumpla la precisión deseada	DIRECTA

De acuerdo con la Tabla 2, el tamaño de una muestra crece cuando la población de la cual se desea extraer es muy heterogénea, vale decir los valores que las unidades poblacionales toman al medir la variable objeto de estudio son muy diferentes entre sí o lo que es lo mismo difieren de su media global sustancialmente.

De otro lado el tamaño de una muestra crece si el investigador desea una precisión bastante alta, esto es, él desea que la media muestral se parezca mucho a la media poblacional, no importa si la primera es mayor o menor que esta última. Lo que a él le importa es que la diferencia entre las dos no pase de un determinado valor llamado error de muestreo. Entre otras cosas el investigador nunca sabrá si la media muestral sobreestima o subestima a la media poblacional ya que esta última jamás se conocerá con precisión absoluta. Y, finalmente, si el investigador desea depositar mucha confianza en sus estimaciones (alta probabilidad de que su estimación no supere el error de muestreo establecido), esta situación se verá reflejada también en el tamaño de la muestra.

El análisis anterior de la sensibilidad del tamaño de la muestra se plantea sobre la base de la influencia de un factor manteniendo constantes el resto.

- En cuanto a los costos resultantes tanto en unidades físicas (destrucción de la unidad, uso de tecnología sofisticada, necesidad de personal altamente capacitado, etc.) como en tiempo (cantidad de unidades medidas y procesamiento de la información obtenida de ellas), se tiene que aceptar que obviamente el muestreo es la mejor alternativa.

Los aspectos anteriores constituyen la importancia y credibilidad del muestreo probabilístico y se verifican en presencia de un proceso de muestreo serio y bien administrado que evite los sesgos de selección, medición y estimación, los cuales hacen que una investigación por muestreo no adquiera la suficiente validez para la toma de decisiones sobre sus resultados.

3. Algunos Problemas Frecuentes que Violan la Validez de un Estudio Estadístico

Definición de la Población Objeto de Estudio

La población objeto de estudio es aquella que el investigador define de acuerdo con el problema estudiado y las restricciones de tipo legal, económico, tiempo, y lugar (Lohr, 2000). Puede afirmarse que si la población no tiene la suficiente cobertura y no está correctamente definida según las características que deben cumplir las unidades de estudio y/o de muestreo, las conclusiones establecidas mediante el estudio tendrán validez para una población que muy posiblemente no es la que corresponde al problema analizado; estamos así en presencia de fallas en la validez externa.

La figura 1, resume los elementos básicos que el investigador debe tener presente en la definición de la población objeto de estudio.

. Un ejemplo sencillo podría plantearse de la siguiente manera. En la ciudad de Cali (delimitación geográfica) se realiza una investigación durante el mes de enero de 2005 (delimitación temporal) mediante la cual se mide la intención de voto para las próximas elecciones a la alcaldía con una encuesta para mujeres mayores de 18 años que están residiendo en la ciudad mínimo desde hace 1 año (delimitación por cumplimiento de características). Si posteriormente (mes de marzo) en el mismo año se vuelve a medir dicha intención no serían comparables los resultados o las conclusiones si se realizan cambios en las características que delimitan la población así por ejemplo, si ahora se dice que la encuesta se aplicará a mujeres mayores de 18 años que residan en la ciudad desde hace mínimo un año pero que hayan votado en las pasadas elecciones para alcalde de la ciudad. En este caso la comparación de la intención de voto para la alcaldía de las mujeres residentes en Cali realizada en el mes de enero no es comparable con la realizada u obtenida en el mes de marzo; precisamente por estar realizadas sobre poblaciones diferentes.

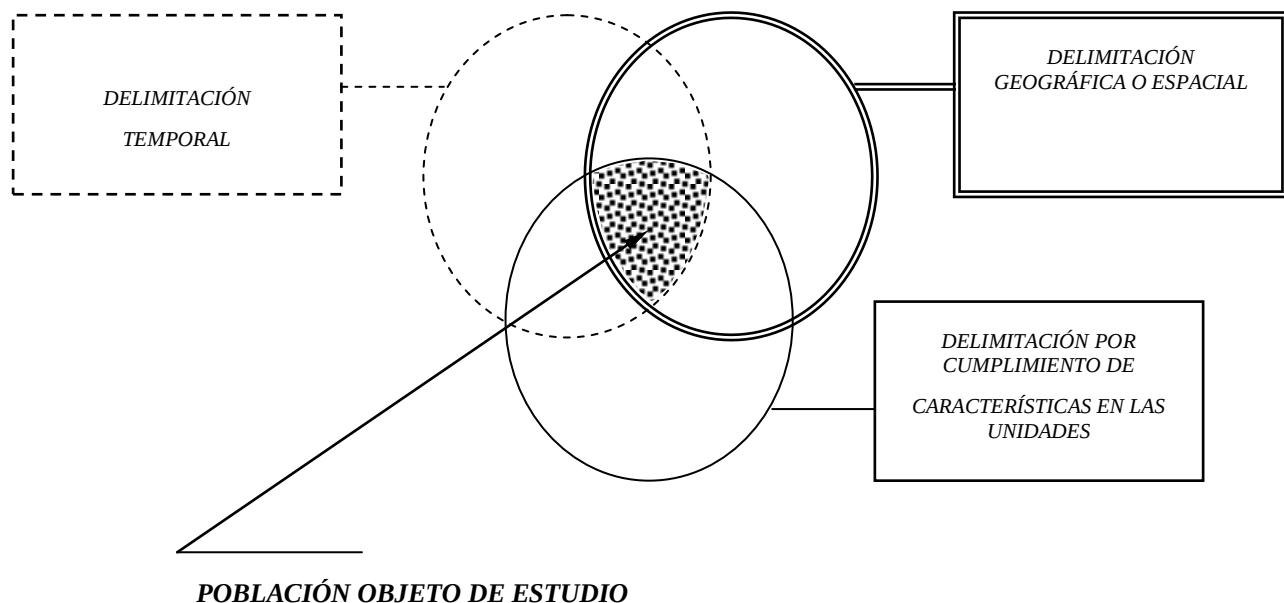


Figura 1. Elementos básicos que definen la población objeto de estudio.

Así mismo es entendible que sería muy impropio decir que las conclusiones obtenidas en el estudio de enero realizado en Cali se pueden aplicar para otra ciudad del Valle del Cauca o de la nación; o que estas pueden aplicarse a los hombres mayores de 18 residentes desde hace mínimo un año en la ciudad de Cali o que pueden aplicarse a cualquier otro mes del año. La delimitación de la población objeto de estudio basada en el cumplimiento por parte de las posibles unidades objeto de estudio de características específicas, exige que el investigador sea muy preciso en la definición de estas hasta el punto que podría decirse que tales características deben estar altamente correlacionadas con aquellas sobre las cuales se desea información para encontrar estimaciones a los parámetros involucrados.

El método de muestreo y el marco muestral

Cuando se desea obtener conclusiones sobre una población es bueno tener presente que dadas las imposibilidades de tiempo, costos, accesibilidad, etc. que se tienen para medir o encuestar la unidades poblacionales, se debe recurrir al uso del muestreo y este debe ser probabilístico para poder

determinar el margen de error y el nivel de confianza con los que se están obteniendo los resultados. Una muestra como es lógico inducirá siempre un error, el cual se denomina error de muestreo, pero este puede delimitarse o medirse solo en los muestreos probabilísticos.

Por otro lado debe tenerse presente que una muestra grande no necesariamente es “representativa” para la población de la cual se extrae si esta no ha seguido una metodología basada en una estructura de aleatoriedad correctamente diseñada para la obtención de ella. En este sentido por ejemplo una muestra de 500 estudiantes de una universidad puede ser muy poco representativa de los estudiantes matriculados el semestre 01 de 2008 si está constituida únicamente por estudiantes de la Facultad de Ingeniería. Muy posiblemente una muestra aleatoria de 200 estudiantes repartida en las diferentes facultades programas y semestres es mucho más representativa de todos los estudiantes de la universidad ya que esta última al ser aleatoria asigna a cada estudiante matriculado en la universidad investigada una probabilidad diferente de cero para pertenecer a la muestra de estudio. En este sentido una muestra es

“representativa” si se obtiene en forma aleatoria manejando un margen de error especificado y un nivel de confianza determinado. Si el método mediante el cual se seleccionan las unidades de la muestra es aleatorio este hecho impide que el investigador cometa sesgo de selección (unidades con probabilidad cero de pertenecer a la muestra) el cual afecta la validez externa de la investigación haciendo que las conclusiones obtenidas mediante el análisis de la muestra no se puedan extrapolar o extender a toda la población de la cual se extrajo.

Por otro lado para la obtención de la muestra, el investigador deberá basarse en un marco muestral (medio físico) (Särndal et al, 2004) para la selección de las unidades que conformaran la muestra. No es difícil pensar que este marco deberá estar exento de errores de cobertura (omisiones, duplicaciones, etc.), por ejemplo para el cálculo de una muestra de estudiantes de la universidad matriculados en el

semestre 02 de 2008 en el Programa de Ingeniería Industrial un excelente marco muestral estará constituido por la lista de ellos reportada por la oficina de Registro Académico de la universidad investigada para dicho período.

Las inexactitudes de los marcos muestrales afectan la validez externa de la investigación que se realiza. La Figura 2, muestra la situación ideal de la relación que debe existir entre la población objeto de estudio y el marco muestral utilizado. Lo anterior plantea que para cada unidad objeto de estudio existe una única unidad de muestreo aunque para cada unidad muestral puedan existir varias unidades objeto de estudio; lo importante es que el marco muestral cubra en un porcentaje superior al 98% la población investigada y no contenga unidades repetidas ni extrañas (Ospina, 2001).

La Figura 3, relaciona los principales defectos de un marco muestral.

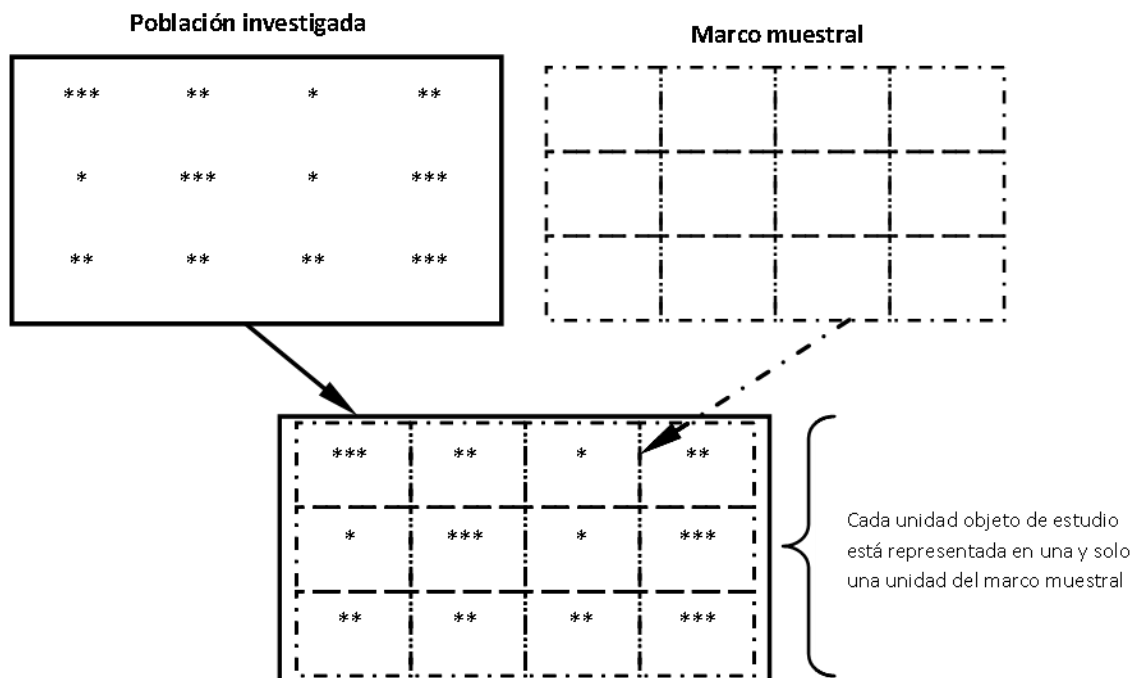


Figura 2. Relación entre el marco muestral y población estudiada.

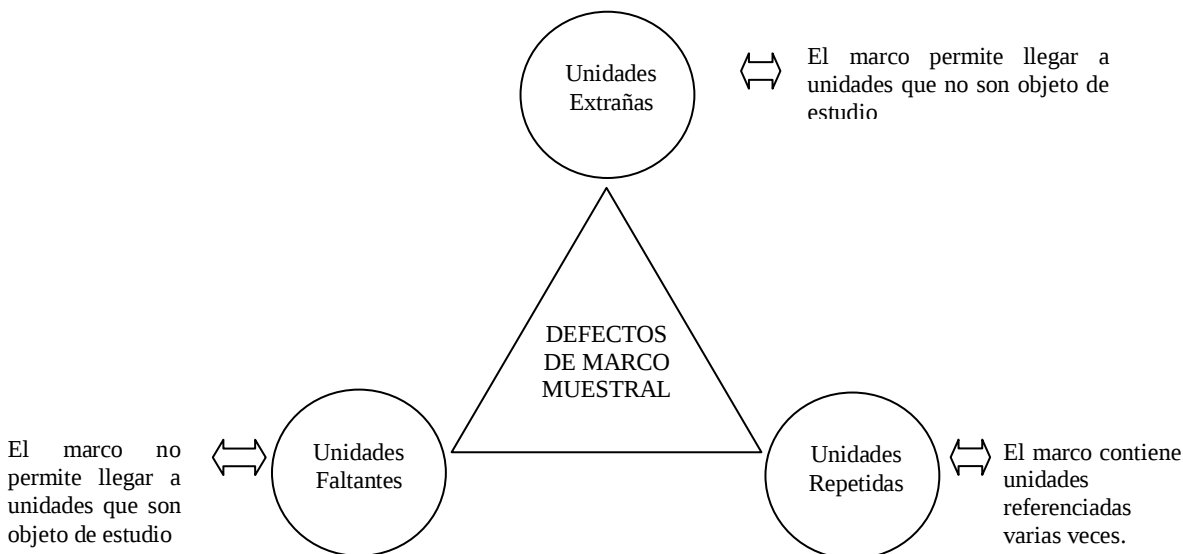


Figura 3. principales defectos del marco muestral

El tamaño de la muestra, su validación (tamaños de dominios de estudio)

En una encuesta de opinión es posible que se desee la obtención de estimaciones para la población entera y para algunas subdivisiones de ella según alguna característica predeterminada. En esta situación de acuerdo con Cochran, páginas 115 y 116 (Cochran, 1980); se pueden presentar dos circunstancias:

- a) Cuando las subdivisiones de la población para las cuales se desean estimaciones se pueden identificar previamente al trabajo de campo. En estas condiciones el investigador puede determinar un tamaño de muestra n_i para cada subdivisión, así por ejemplo si el investigador desea estimar la media poblacional $\bar{X}_i$ para cada i -ésima subdivisión poblacional con varianza deseada para dicha estimación igual a S_i^2 constante para cada subdivisión, puede utilizar respectivamente las expresiones que se muestran en la Tabla 3 para estimar los tamaños de muestra.

Tabla 3. Tamaños de muestra para estimar la media de cada subdivisión, cuando las subdivisiones se pueden estimar previamente al trabajo de campo.

Para cada subdivisión i	Para la población entera
$n_i = S_i^2 / V_0$	$n = \sum_i n_i = \sum_i S_i^2 / V_0$

Aquí debe tenerse presente que S_i^2 es la cuasivarianza de la subdivisión i -ésima; que se tiene idea del tamaño poblacional de cada subdivisión y que no se considera la corrección por población finita en cada subdivisión. Adicionalmente como el cálculo del tamaño de muestra en cada subdivisión requiere de la estimación por adelantado de la cuasivarianza de la subdivisión podemos conocer el tamaño de muestra que se requeriría para toda la población con idéntica varianza deseada después de estimar la cuasivarianza poblacional S^2 , usando la expresión básica del análisis de varianza que se presenta en la Ecuación 3.

$$(N-1)S^2 = \sum_i (N_i-1)S_i^2 + \sum_i N_i(\bar{Y}_i - \bar{Y})^2 \quad (3)$$

Si S_i^2 para cada una de las k subdivisiones es aproximadamente igual a la cuasivarianza poblacional, entonces el tamaño de muestra para encontrar las estimaciones en cada subdivisión poblacional con el requerimiento de una varianza deseada V_0 para el estimador, puede aproximarse como k veces el tamaño de muestra requerido para la población entera.

- b) Si la subdivisión de la población a la que pertenece una unidad estadística sólo se sabe en el momento de encuestarla en el trabajo de campo, en este caso si se conocen las proporciones π_i para las unidades que pertenecen a cada subdivisión, se determina una muestra global n , y de acuerdo con ello, el valor esperado para el tamaño de cada subdivisión será el que aparece en la Ecuación 4.

$$E(n_i) = n\pi_i \Rightarrow V_0 = \frac{S_i^2}{E(n_i)} = \frac{S_i^2}{n\pi_i} \quad (4)$$

$$\therefore n = \frac{S_i^2}{\pi_i V_0}$$

La Ecuación 4 permite encontrar el tamaño de la muestra requerido para cada subdivisión. Así mismo el tamaño de muestra necesario para toda la población lo ajustamos mediante la Ecuación 5.

$$n \cong \text{máximo} \left(\frac{S_i^2}{\pi_i V_0} \right) \quad (5)$$

Igual que en la situación a) se supone que es despreciable la fracción de muestreo en cada subdivisión. En caso de un desconocimiento en las proporciones π_i de las subdivisiones en la población se procederá a estimarlas en forma adelantada. Idénticamente si para alguna subdivisión dicha proporción es muy pequeña (ejemplo clases extremas en la

distribución de la característica analizada) se debe asumir una varianza deseada V_0 más pequeña en dicha subdivisión para aumentar la posibilidad de tener suficientes unidades de esta subdivisión.

El método de recolección de la información y la redacción de las preguntas

En las encuestas de opinión, el instrumento por excelencia para la recolección de la información suministrada por el investigado es la encuesta o cuestionario. De acuerdo con Fernando García Córdoba (García, 2002, p. 32), para conformarse como herramienta útil a la investigación, el cuestionario deberá cumplir mínimamente los requisitos de validez, confiabilidad, comparabilidad y adaptabilidad, haciéndose entonces muy necesario realizar las pruebas correspondientes para el cumplimiento de dichos requisitos antes de su implementación.

Se deben realizar las siguientes preguntas a la hora de caracterizar un buen cuestionario:

¿El cuestionario permite la obtención de información necesaria y suficiente según los propósitos de la investigación?

¿El cuestionario toma en cuenta las necesidades y reacciones del informante sin mortificarlo?

¿El cuestionario genera un ambiente favorable para la realización de la encuesta?

¿El cuestionario mantiene el interés del encuestado?

¿El cuestionario permite precisar la información que se requiere del informante permitiendo además tipificarla en unidades manejables?

Las anteriores preguntas entre otras, podrán contestarse muy apropiadamente mediante la realización de un estudio piloto (encuesta piloto). A pesar del tiempo y el dinero en que se incurra

para el proyecto de investigación no debe desdeñarse la realización del estudio piloto y mucho menos realizarlo en forma descuidada. La siguiente gráfica resume la importancia de su realización para un estudio de opinión.

Realización del estudio piloto

Podría decirse que un estudio piloto es en buena parte una ejecución en pequeño del proyecto de investigación y por tal razón, él permitirá verificar aspectos de logística y suministrará en general conocimiento y precisión de todos los aspectos involucrados en el proyecto a más de ayudar a obtener estimaciones adelantadas para la definición misma del tamaño final de la muestra. La Figura 4, ilustra las bondades de su realización.

Dada la importancia de un estudio piloto, éste debe realizarse con la misma rigurosidad que se utilizará en el estudio final hasta el punto de que si no hay cambios sustanciales para la encuesta; las unidades ya encuestadas en el estudio piloto podrán ser parte de la muestra final.

La tasa de no respuesta

Dentro de los inconvenientes a los que se ven frecuentemente abocados los investigadores al realizar encuestas de opinión, está la ausencia de información y/o la inconsistencia de la misma. La presencia en la investigación de estos inconvenientes genera lo que suele llamarse la no respuesta, y su control permite evitar o reducir los sesgos de las estimaciones.

EL ESTUDIO PILOTO

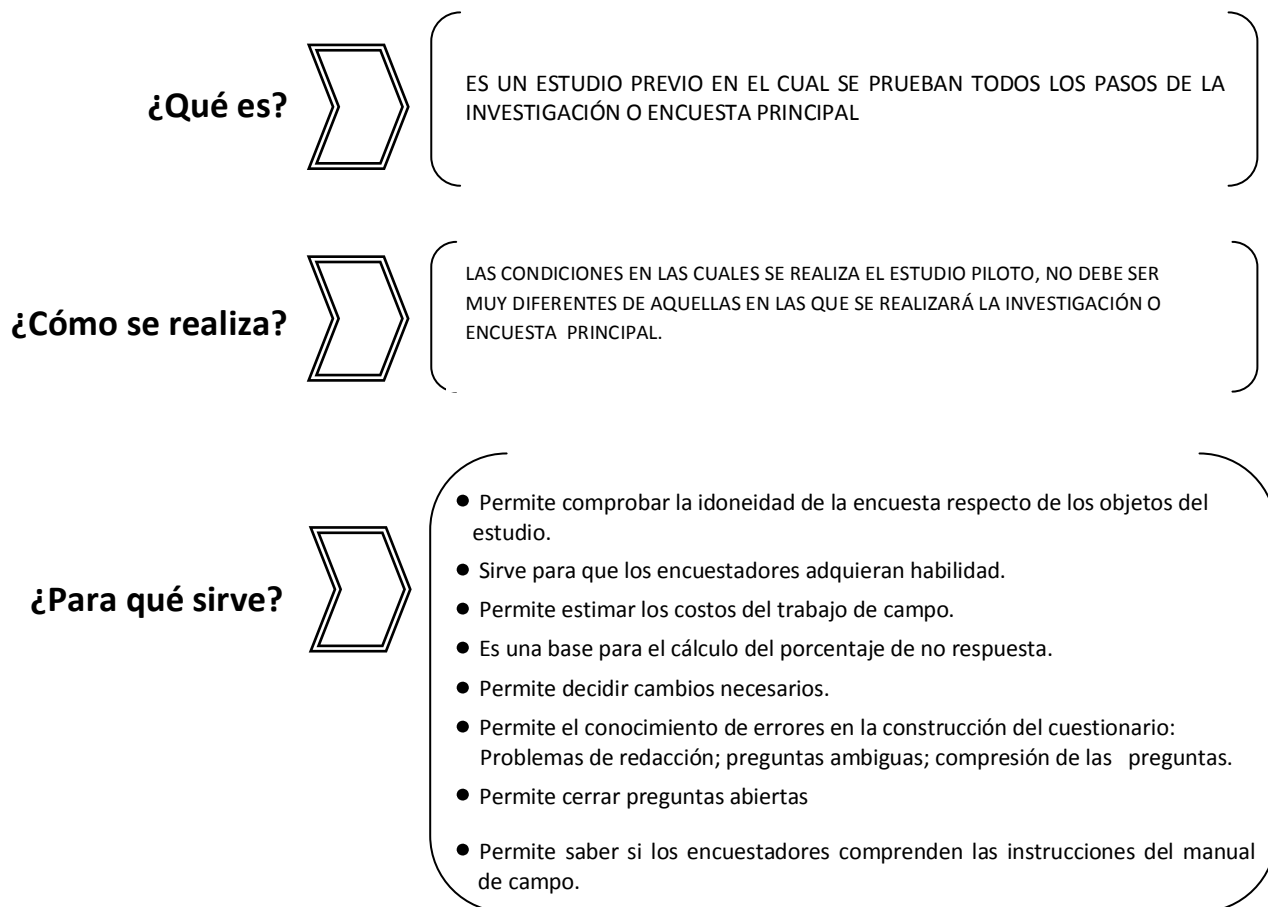


Figura 4. Diagrama metodológico del estudio piloto.

La no respuesta se presenta cuando algunas de las unidades investigadas (encuestadas) no contestan una o varias de las preguntas formuladas en el cuestionario diseñado para tal efecto. Dentro de las principales causas de la no respuesta se pueden citar las siguientes: Defectos de marco que hacen que algunas unidades objeto de estudio no puedan identificarse; Problemas de localización que hacen que en el momento de realizada la encuesta algunas unidades investigadas no están presentes; Incapacidad del encuestado que impide que este proporcione la información como es debido y finalmente, la falta de cultura estadística en el encuestado haciendo que este se rehúse a colaborar con el estudio.

Para que una investigación mediante encuestas mantenga una baja presencia de la no respuesta se deben implementar las siguientes acciones:

a) realizar nuevas visitas (revisitar) la unidad seleccionada en horas más apropiadas para lograr la información del encuestado, b) envió de recordatorios por correo o Internet aclarando al informante la importancia de su colaboración, c) realización de submuestreo en el grupo de las personas seleccionadas que no han respondido la encuesta y hacer todos los esfuerzos posibles para lograr que los integrantes de la submuestra suministren la información solicitada. d) El uso de métodos de respuesta aleatorizada para las preguntas que afectan la intimidad del encuestado. En este sentido la encuesta anónima y auto-administrada garantizan la cooperación y la veracidad del encuestado.

En cuanto a los datos faltantes se pueden realizar las siguientes estrategias: a) tomar los registros que contienen algún valor faltante y descartarlos realizando el análisis con el resto, b) no descartar ningún registro y procesar la información sin realizar ningún tipo de ajuste para la no respuesta total, c) realización de procedimientos de ponderación para incrementar los pesos de los respondientes (con problemas) para que estos representen a los no respondientes a nivel poblacional, d) realización de procedimientos de imputación, mediante los cuales se llenan los valores faltantes de las variables que presenten dichos problemas.

Sea cual fuere el procedimiento a seguir, éste debe estar plenamente justificado; en resumen una investigación mediante encuesta puede presentar uno o varios de los siguientes problemas en la toma de información sobre las características investigadas: inconsistencia en los valores asignados; no presencia de información para la característica ó valor extraño (atípico) en la casilla correspondiente a la característica, por tal razón se hace necesario en todo momento ser cuidadoso con la información que integra la base de datos que será procesada para la obtención de las diferentes estimaciones.

La comparación de cifras y los test de significancia (un mismo estudio; estudios diferentes)

El objetivo central de cualquier estudio de opinión es la estimación de parámetros y la realización de pruebas de hipótesis sobre dominios de estudio de interés dentro de una misma población para la toma de decisiones sobre la población investigada o para la comparación de la población investigada con otras poblaciones que le son comparables. El que sean o no comparables dos poblaciones con base en los estimadores hallados en cada una, está fundamentada en la concordancia que exista entre las dos poblaciones respecto de los elementos que las definen en cada caso.

La comparación de dos poblaciones se realiza básicamente mediante dos procesos estadísticos inferenciales. De una parte mediante la estimación de parámetros involucrando desde luego la variabilidad presente en dichas estimaciones, y de otro lado mediante la realización de pruebas de hipótesis sobre los parámetros poblacionales. Ambos procesos están ligados.

a) Estimación de parámetros poblacionales

En las encuestas de opinión es muy deseable obtener estimaciones por intervalo sobre los parámetros poblacionales y allí debe tenerse presente que cualquiera sea el intervalo de confianza este debe ser lo más estrecho posible (longitud), significando con ello que a más

estrechez más exactitud se tendrá sobre la localización del parámetro poblacional estudiado y por otra parte un buen intervalo de confianza debe construirse involucrando una probabilidad alta para que el parámetro esté contenido entre sus límites. Lógicamente detrás de estas dos propiedades debe tenerse presente que la construcción del intervalo exige del investigador pleno conocimiento de la distribución probabilística que sirve de modelo para interpretar las estimaciones puntuales del parámetro involucrado en la investigación. Dicho conocimiento está en función de la naturaleza de la característica analizada, del conocimiento que se tenga sobre la variabilidad poblacional de dicha característica y del tamaño de muestra utilizado por el investigador para la obtención de sus estimaciones.

b) Verificación de hipótesis sobre parámetros poblacionales

En general un contraste de hipótesis es una regla de decisión mediante la cual el investigador puede optar por uno de dos caminos (dos hipótesis), después de encontrar estimaciones sobre los parámetros involucrados en dichas hipótesis.

Al respecto un investigador a más del conocimiento exigido en la situación a) debe tener claridad sobre las hipótesis involucradas (forma); los posibles valores deseables para la probabilidad de aceptar una hipótesis nula falsa (error tipo II); la de rechazar una hipótesis nula verdadera (error tipo I) y finalmente haber definido muy bien las regiones de rechazo y aceptación para la prueba.

Al hablar de verificación de hipótesis (prueba de hipótesis) se debe tener presente que se trata de un proceso mediante el cual se comparan dos hipótesis. Primero está la hipótesis nula o de investigación (H_0) que se debe comprobar. Esta establece que no existe ninguna diferencia entre dos poblaciones, dos parámetros poblacionales o entre un valor verdadero de algún parámetro y su valor hipotético. Si la información obtenida por el investigador no le permite rechazar la hipótesis nula, él debe concluir que los datos particulares de la muestra usada no dan suficiente evidencia para

que él concluya que la hipótesis nula es falsa; mientras que si la hipótesis nula se rechaza, él debe concluir que los datos particulares de la muestra usada sí le dan suficiente evidencia para concluir que dicha hipótesis es falsa y que por lo tanto existe una segunda hipótesis llamada hipótesis alterna (H_1). Aparentemente los intervalos de confianza y las pruebas de hipótesis pareciera que son formas diferentes de comparación de cifras en estudios por muestreo pero no es así, incluso una prueba de hipótesis puede realizarse desde la comparación de intervalos de confianza.

4. Conclusiones y Recomendaciones

Como se dijo al comienzo, el uso de la Estadística por sí sola no hace que una investigación tenga validez y por esto, todo investigador debe ser muy cuidadoso con el uso de las herramientas estadísticas, las cuales deben estar claras y justificadas a la luz del problema investigado desde el mismo comienzo del proyecto.

Las siguientes son recomendaciones generales para tener en cuenta a la hora de aplicar técnicas estadísticas en un proyecto de investigación mediante encuestas (Scheaffer, 2007).

- Tener claro el problema que se desea investigar y plantearlo en forma de objetivos medibles.
- Delimitar los alcances de la investigación en materia de cobertura geográfica; poblacional, de características y de tiempo a la luz de los costos establecidos.
- Realizar un cronograma con todas las actividades del proyecto con la asignación del responsable.
- Integrar un grupo multidisciplinar de asesoramiento a la dirección del proyecto en el cual el profesional de la Estadística jugará un papel importante y trascendental desde la definición misma del proyecto investigativo.

- Realizar cuadros de salida acordes con los objetivos del proyecto, los cuales permitirán visualizar el uso y la necesidad de las técnicas estadísticas.
 - Establecer un estricto control de calidad en la recolección y el procesamiento de la información obtenida (López, 201).
 - Utilizar el software estadístico adecuadamente (Albridge et al., 2003).
 - Buscar asesoría Estadística desde el comienzo del proyecto de investigación.
- 11. Särndal, C., Swensson, B., Wretman, J. (1992). *Model Assited Survey Sampling*. New York: Springer-Verlag.
 - 12. Scheaffer, R., Mendenhall, W. y Ott, L. (2007) *Elementos de Muestreo*. Thomson; Madrid.

5. Referencias Bibliográficas

1. Aldridge, A. y Levine K. (2003) *Topografía del Mundo Social*. España: Gedisa.
2. Cochran, W. (1980) *Técnicas de Muestreo*. México: CECSA.
3. D'Anacona, M. A. (2004). *Métodos de Encuesta. Teoría y Práctica, Errores y Mejora*. España: Síntesis.
4. García, F. (2002) *El Cuestionario*. México. Limusa.
5. Lohr, S. (2000) *Muestreo*. México: Thomson.
6. López, A. (2001) *Investigaciones de Mercado*. México. CECSA.
7. Martínez, V. (2003) *Diseño de encuestas de opinión*. España: Editorial Ra-Ma.
8. Ospina, D. (2001) *Introducción al Muestreo*. Bogotá: Universidad Nacional de Colombia.
9. Pérez, C. (2000). *Técnicas de Muestreo Estadístico*. México: Alfaomega.
10. Rojas, A., Fernández, J. , Pérez, C. (1998) *Investigar mediante encuestas*. España. Síntesis.