MUESTREO ESTADÍSTICO
MÉTODOS BÁSICOS

• Rafael Klinger Angarita •



Universidad del Valle
Programa Editorial

El propósito de este libro es presentar de una manera amigable y sencilla los diferentes conceptos involucrados en la *teoría del muestreo probabilístico* desde una óptica clásica. Siendo el muestreo una de las formas básicas de la estadística para obtener información con el fin de satisfacer necesidades establecidas en proyectos de investigación de diferente índole y cobertura, se constituye en el mundo de hoy con más fuerza en una metodología científica para la toma de decisiones mediante el uso de la inferencia estadística. Basta con considerar el tamaño de los colectivos objeto de estudio, la dificultad en tiempo y dinero para llegar a todos los elementos que los componen y, en algunos casos, los riesgos por destrucción que se tendrían para poder observarlos (medirlos), así como la dificultad de su acceso, situaciones que obligan en cualquier campo del conocimiento a la toma de una muestra probabilística de dicho colectivo.

Tres aspectos son de suma importancia en la obtención de una muestra probabilística desde el punto de vista práctico: en primer lugar, la forma como serán obtenidas las unidades que conformarán la muestra, la cual debe estar regida por la aleatoriedad para evitar sesgos de selección; en segundo lugar, el tamaño de dicha muestra soportada sobre la base de supuestos teóricos para la distribución (comportamiento) del estimador asociado al parámetro investigado, el error asumido con la estimación debido a la diferencia entre la estimación y el parámetro objeto de estimación, y el nivel de confianza en términos de probabilidad que se desea tenga el cumplimiento de dicho error; y, en tercer lugar, los costos involucrados en la obtención de dicha muestra, pues nadie niega la importancia de trabajar con muestras económicas.

Universidad del Valle

Programa Editorial

MUESTREO ESTADÍSTICO

MÉTODOS BÁSICOS



Colección Ingeniería

Klinger Angarita, Rafael
Muestreo estadístico : métodos básicos / Rafael Klinger Angarita. 1a. ed.—Cali : Programa Editorial Universidad del Valle, 2024
440 páginas ; 17cm x 24 cm. -- (Colección Ingeniería – Libro de texto)
1. Estadística - 2. Métodos de investigación - 3. Muestreo
519.53 cd 22 ed.
K65
Universidad del Valle - Biblioteca Mario Carvajal

Universidad del Valle
Programa Editorial
Título: Muestreo estadístico: métodos básicos
Autor: Rafael Klinger Angarita
ISBN: 978-958-507-095-0
ISBN-pdf: 978-958-507-096-7
ISBN-epub: 978-958-507-099-8
DOI: 10.25100/peu.5070950
Colección: Ingeniería
Primera edición

© Universidad del Valle
© Rafael Klinger Angarita

Diseño de carátula: Ángela María Arboleda Mera
Diagramación: Artes Gráficas del Valle S.A.S.

Este libro, o parte de él, no puede ser reproducido por ningún medio sin autorización escrita de la Universidad del Valle.

El contenido de esta obra corresponde al derecho de expresión del autor y no compromete el pensamiento institucional de la Universidad del Valle, ni genera responsabilidad frente a terceros. El autor es el responsable del respeto a los derechos de autor y del material contenido en la publicación, razón por la cual la Universidad no puede asumir ninguna responsabilidad en caso de omisiones o errores.

Cali, Colombia, enero de 2024

MUESTREO ESTADÍSTICO
MÉTODOS BÁSICOS

• Rafael Klinger Angarita •



Colección Ingeniería

RAFAEL KLINGER ANGARITA

Profesor titular de la Escuela de Ingeniería Industrial y Estadística de la Universidad del Valle. Asesor y consultor en el campo de la estadística aplicada con especialidad en muestreo.

PRÓLOGO

El propósito de este libro es el de presentar de una manera amigable y sencilla, los diferentes conceptos involucrados en la teoría del muestreo probabilístico desde una óptica clásica.

El muestreo es uno de los métodos básicos de la estadística para obtener información que permita satisfacer necesidades establecidas en proyectos de investigación de diferente índole y cobertura, se constituye actualmente en una metodología científica que ha tomado fuerza en la toma de decisiones mediante el uso de la inferencia estadística. Basta con considerar el tamaño de los colectivos objeto de estudio, la dificultad en tiempo y dinero para llegar a todos los elementos que los componen y en algunos casos, los riesgos por destrucción a la que se llegaría para poder observarlos (medirlos), así como la dificultad de su acceso, situaciones que obligan en cualquier campo del conocimiento a la elección de una muestra probabilística de tal conjunto.

Tres aspectos son de suma importancia en la obtención de una muestra probabilística desde el punto de vista práctico: en primer lugar, la forma como serán obtenidas las unidades que conformarán la muestra, la cual debe estar regida por la aleatoriedad con el propósito de evitar sesgos de selección; en segundo lugar, el tamaño de dicha muestra soportada sobre la base de supuestos teóricos para la distribución (comportamiento) del estimador asociado al parámetro investigado, el error asumido con la estimación debido a la diferencia entre la estimación y el parámetro objeto de estimación y el nivel de confianza en términos de probabilidad que se desea tenga el cumplimiento de dicho error; y, en tercer lugar, los costos involucrados en la obtención de dicha muestra, pues nadie niega la importancia de trabajar con muestras económicas.

El libro está dirigido a estudiantes y profesionales de la estadística y demás áreas del conocimiento que posean elementos de matemáticas básicos, para entender las demostraciones presentadas en algunos de los capítulos, se presentan los diferentes aspectos relacionados con la *teoría del muestreo*, así como los diferentes métodos para la obtención de una muestra probabilística, colocando especial cuidado a lo largo del mismo en su aplicabilidad a través de ejemplos alusivos a situaciones reales. El lector podrá mediante el uso del MINITAB 14 o algún otro software de que disponga, aplicar los diferentes conceptos presentados de una manera práctica al usar la base de datos que se anexa, y que contiene 2600 registros de 17 variables. Así mismo, podrá poner en práctica los conceptos aprendidos mediante la solución de los ejercicios que aparecen al final de cada capítulo.

El libro está integrado por cinco capítulos: el primer capítulo tiene como misión presentar la importancia de hacer una buena elección de la muestra en términos generales y el papel que ella juega en todos los aspectos de una investigación sobre la base de datos estadísticos. El segundo capítulo está dedicado al muestreo aleatorio simple, al ser este método la base sobre la cual los demás métodos cobran vida, pues todos ellos desembocan en un muestreo aleatorio simple. El tercer capítulo presenta la sensibilidad de una muestra a través del estudio de los diferentes aspectos que la definen, siendo lo presentado en este, la base metodológica para la obtención de una muestra en cualquiera de los métodos presentados en los restantes capítulos.

El cuarto capítulo está dedicado al muestreo aleatorio estratificado, el cual requiere que las unidades estén agrupadas (o se puedan agrupar) de acuerdo con otras variables conocidas, para producir estimaciones más homogéneas en cada grupo y por ende estimaciones poblacionales más precisas con tamaños de muestra menores a los obtenidos mediante un muestreo aleatorio simple. Las unidades al interior de cada grupo llamado estrato, deben ser lo más homogéneas entre sí, respecto de la variable analizada y altamente heterogéneas de un grupo a otro para conformar la muestra final como el agregado de las muestras de cada grupo en que se divida la población.

Finalmente, en el quinto capítulo se referencian los cuidados y el tipo de muestras que se manejan en el campo de la ingeniería y de los recursos naturales. A manera de ilustración, se presentan algunos dispositivos usados para el muestreo de materiales plantas y animales; recursos fundamentales para la aplicación de la ingeniería.

Son varios los aspectos que se deben tener presentes a la hora de usar un método específico de muestreo: la presencia o no de un marco muestral y su bondad; el tipo de estimador involucrado en la investigación sobre el

parámetro, los costos, la forma como se encuentran en el marco muestral las unidades investigadas y la disponibilidad o no de información auxiliar sobre ellas. El lector debe tener presente que en muchas oportunidades será posible usar diferentes métodos para la estimación del mismo parámetro con el mismo error y nivel de confianza asumido, desde luego, en todo caso debe usarse el menos costoso y más práctico.

Todos los métodos pueden ser mezclados para mejorar las estimaciones solo que las expresiones para los estimadores y sus varianzas pueden hacerse muy extensas y complejas, pero su aplicación se facilita después de realizar un buen algoritmo en Excel, o mediante el uso de software estadístico; en todo caso, el diseño muestral que se adopte debe procurar el menor costo posible, la menor varianza deseable para el estimador.

El autor

CONTENIDO

CAPÍTULO 1
MUESTREO ESTADÍSTICA E INVESTIGACIÓN 15
Introducción 15
Soportes de una investigación de calidad 16
La investigación y el modelo estadístico 17
Características que diferencian el censo y el muestreo 22
Datos estadísticos 23
La validez de un estudio estadístico 25
El muestreo: una necesidad en el mundo de hoy 28
Los métodos de muestreo 34
Caracterizando la población: censo o muestreo 46
Escalas de medición de datos y muestreo 48
Marcos muestrales 50
Insumos básicos para la elaboración de un plan de muestreo 56
Parámetros y estimadores 58
Errores en la investigación por muestreo. 66
Formas básicas de obtención de información 72
Etapas de una encuesta por muestreo 77
Ejercicios 84

CAPÍTULO 2
MUESTREO ALEATORIO SIMPLE (MAS) 89
Introducción 89
Definición del muestreo aleatorio simple (MAS) 89
Hipótesis sobre las que se fundamenta el MAS 91
Formas de realizar un MAS 91
Muestreo aleatorio simple sin reemplazo 92
Procedimiento para seleccionar una muestra aleatoria simple sin reemplazo 94
Muestreo aleatorio simple sin reemplazo para variables 96
Muestreo aleatorio simple sin reemplazo para atributos 124

Proporciones y totales en subpoblaciones 139
Muestreo aleatorio simple con reemplazo 142
El método bootstrap 155
Ejercicios 158

CAPÍTULO 3
ELEMENTOS BÁSICOS PARA DETERMINAR EL TAMAÑO DE LA MUESTRA 177

Introducción 177
Ecuaciones fundamentales para la determinación de n 177
Estimaciones adelantadas de la cuasivarianza poblacional. 179
Estimación adelantada de S^2 por conocimiento parcial de la distribución de la característica analizada (conjetura) 180
Tamaño de muestra para variables 183
Tamaño de muestra para estimar la media poblacional $\overline{Y}$ con un error relativo R. 186
Tamaño de muestra para estimar el total Y con un error relativo R 188
El estimador de la media en presencia de la suma de dos muestras diferentes de la misma población 193
Tamaño de muestra para atributos 196
Errores asumidos por las estimaciones de los parámetros en variables que no definen el tamaño de la muestra 206
La sensibilidad de la varianza del estimador. 212
Tamaño de muestra en subdivisiones poblacionales o dominios de estudio para variables numéricas 212
Sensibilidad del tamaño de muestra 224
Costos y tamaño de muestra 230
Ajustes al tamaño de muestra debidos a la no respuesta y los costos de trabajo de campo 235
Ejercicios 236

CAPÍTULO 4
MUESTREO ALEATORIO ESTRATIFICADO 253

Introducción 253
Hipótesis fundamentales del muestreo aleatorio estratificado (MAE) . . . 254
Representación gráfica del método 254
Procedimiento para seleccionar una MAE 255
Razones para el uso del muestreo estratificado 255
Tres preguntas que se deben contestar antes de realizar un MAE. 256
Muestreo aleatorio estratificado para variables 257
Muestreo aleatorio estratificado para atributos 287

¿Qué hacer cuando la asignación en uno de los estratos supera el tamaño del estrato? 315
Comparando el MAS con el MAE 319
La ganancia en la estimación de la media debida a la estratificación de una población 323
La construcción de los estratos 329
Dominios de estudio en el MAE 342
Postestratificación. 347
Ejercicios 354

CAPÍTULO 5
MUESTREO E INGENIERÍA 375

Introducción 375
Glosario de términos relacionados con el muestreo para prácticas de laboratorio 376
Aspectos que deben considerarse en el tratamiento de un problema en ingeniería que implique el análisis de muestras de materiales 378
¿Qué es y qué contiene un plan de muestreo? 379
Cuidados en la manipulación de las muestras 379
Tipos de muestras industriales. 380
Métodos y equipos para la toma de muestras 381
Almacenamiento y transporte de las muestras. 391
Pretratamiento de las muestras sólidas 392
Algunas limitaciones del pretratamiento de una muestra. 395
Incertidumbre y tamaño de muestra probabilística para laboratorio 396
Posibles fuentes de error para los resultados con base en muestras para laboratorio 397
Aproximación estadística para la obtención de muestras representativas de materiales o residuos peligrosos en suelos 397
Diseños de muestreo probabilístico sobre espacio y tiempo para ser usados en investigación medioambiental. 399
Muestras representativas para el control de calidad del agua 401
Muestras representativas para la calidad del aire 401
Muestras de suelo para análisis agropecuarios 402
Tamaño de muestras homogenéas y hetrogéneas en el laboratorio 404
Muestreo en biología animal y vegetal 405
Capturando y marcando las unidades muestreadas para determinar un tamaño poblacional 408
Ejercicios 411

APÉNDICES 417
Tabla a. Números aleatorios 418
Tabla b. Distribución normal estandarizada 421
Tabla c. Distribución *t-student* 423

BIBLIOGRAFÍA 435

CAPÍTULO 1

MUESTREO ESTADÍSTICA E INVESTIGACIÓN

INTRODUCCIÓN

El profesional y el ciudadano de hoy están inmersos cada vez más en resultados de encuestas y estudios estadísticos. Se puede decir que buena parte de las decisiones investigativas están basadas en datos estadísticos que provienen del análisis, observación o medición de unas cuantas unidades estadísticas, llevadas al laboratorio o seleccionadas para ser encuestadas.

El propósito de la estadística es obtener información acerca de las poblaciones involucradas en una investigación. Esta información que se obtiene de las poblaciónes se resume en indicadores de tendencia central, de variabilidad y mediante gráficos. Existen dos estrategias posibles para la recopilación de información inherente a las unidades investigadas, que son: examinar todas las unidades de la población (censo) o examinar, según unos planes establecidos con anterioridad, ciertas unidades de la población (muestra) y suponer que los resultados obtenidos son representativos de toda la población. Utilizando la primera de las estrategias, se puede describir "perfectamente" el comportamiento de las características analizadas en la investigación, mientras que con la segunda estrategia se hace necesario extrapolar las conclusiones obtenidas en la muestra a toda la población. Esta extrapolación (*inferencia*) siempre supone un riesgo, por lo que cabría preguntarnos entonces ¿el porqué de su amplio uso hoy en día frente a la "seguridad" que supone el censo?

No es difícil pensar que existen ocasiones en las que no queda otra alternativa que la selección de una muestra para la viabilidad de realización de una investigación debido a:

- La población es tan grande que excede las posibilidades del investigador.
- La población es suficientemente homogénea como para que cualquier muestra dé una buena representación.
- El proceso de medida hace que se destruya la unidad investigada.

Aparte de los tres casos anteriormente expuestos, existen otras razones que pueden hacer ventajoso el muestreo, como son:

- Costo reducido.
- Mayor rapidez.
- Mayor exactitud debido a un mejor control.

En resumen, la decisión óptima entre muestra y censo consiste en minimizar la pérdida total en que se incurre, por una parte, con respecto a los recursos empleados (tiempo, dinero, recursos y esfuerzos) y, por otra, el error en que se pueda incurrir al usar una muestra y la probabilidad de cometerlo.

Para hacer muestreo se requiere del conocimiento de la teoría básica del muestreo y de su correcta aplicación. Este capítulo relaciona de una forma sencilla y amena, la necesidad y la importancia del muestreo, así como algunos de los elementos básicos sobre los cuales se sustenta su aplicación que serán la base del estudio de los diferentes métodos de muestreo en los siguientes capítulos; se ilustran igualmente los principios fundamentales de una buena muestra que permiten a un investigador garantizar la validez de las conclusiones halladas.

SOPORTES DE UNA INVESTIGACIÓN DE CALIDAD

Todo investigador en cualquier campo del conocimiento debe tener presente que sus realizaciones investigativas han de tener una buena calidad, lo cual significa que estas estén soportadas por los siguientes pilares: estar basadas en el trabajo de otros investigadores, ser repetitivas, poder generalizarse a otras situaciones, usar razonamientos lógicos y estar vinculadas a una teoría, generar nuevas preguntas y ser de naturaleza cíclica, ser incremental y, finalmente, deben emprenderse con el fin de mejorar la sociedad. La investigación estadística, tiene como hipótesis de trabajo *afirmaciones relativas a distribuciones de variables aleatorias* y el trabajo alrededor de estas hipótesis está soportado por los fundamentos anteriormente expuestos. La figura 1.1 resume el significado de cada uno de los principios que soportan una investigación de calidad.

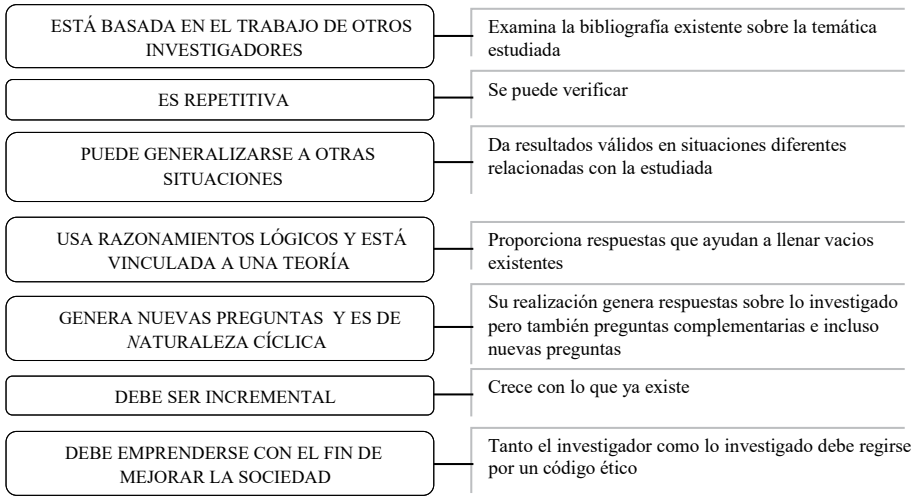


Figura 1.1. **La investigación y el modelo estadístico**

LA INVESTIGACIÓN Y EL MODELO ESTADÍSTICO

El trabajo estadístico es mucho más que una acumulación de datos para tomar decisiones frente a una situación analizada. Siempre se requerirá de la implementación de un modelo estadístico para analizar la realidad en cuestión; pero el uso de dicho modelo debe permitirle al investigador tomar decisiones correctas. Es decir, acordes con la realidad, lo cual requiere de alguna o varias de las acciones mostradas en la figura 1.2.

En cada investigación se debe considerar (Méndez, 1992) un modelo estadístico, el cual debe ser construido a partir de los siguientes aspectos:

- Las conjeturas derivadas de las preguntas o teorías que proponga el investigador.
- La información sobre la forma como serán colectados los datos, llamado propiamente el *diseño* de investigación.
- La información adicional o las conjeturas que existan sobre la naturaleza del fenómeno que se investiga o modela.

Para realizar una investigación se parte siempre de la realidad y, mediante el análisis de esta, se puede formular el *diseño de la investigación* (estructura de la forma como serán colectados los datos), pasando luego a los *modelos estadísticos*, a través de los cuales finalmente se lleva a cabo el *análisis de la información* recogida o colectada.

Es importante tener presente que mientras se está en la fase del *análisis de la información* se deben valorar permanentemente los *modelos estadísticos*. Una vez, terminada esta valoración (certeza de compatibilidad), se realiza una nueva valoración, ahora entre el *análisis de la información* y el *diseño de la investigación o estructura*, lo cual permitirá tener igualmente certeza sobre lo adecuado del *diseño de la investigación o estructura*. El paso siguiente y final permitirá obtener las conclusiones después de contrastar el *análisis de la información* con la *realidad analizada o fenómeno estudiado*, muy seguramente este proceso tendrá que ser realizado varias veces en la misma investigación buscando optimizar al máximo el modelo estadístico planteado. Cuando en el *diseño de la investigación* se usa un muestreo probabilístico, las conclusiones sobre la realidad se llamaran inferencias. El trabajo estadístico requiere mucho más que una acumulación de datos para tomar decisiones (ver figura 1.2).



Figura 1.2. **La investigación y el modelo estadístico**

Elementos del proceso estadístico en una investigación

La figura 1.3 muestra la secuencia de los elementos básicos (representados por cuadros) del proceso estadístico en la investigación y sus relaciones, las cuales se representan por flechas. En este esquema, se supone que el problema objeto de estudio en la investigación que se realiza está completamente definido y que las características o variables involucradas en él, están igualmente definidas y clasificadas.

Toda investigación estadística, por lo regular, involucra básicamente seis elementos fundamentales, a saber:

- La *población objeto de estudio* en la investigación, conformada por las unidades objeto de análisis sobre las cuales se aplicarán las conclusiones de la investigación. Esta población puede ser de diferente naturaleza (individuos, hogares, barrios, parcelas cultivadas, productos manufacturados, entidades bancarias, etc.).
- La *muestra*, se forma con una parte de unidades estadísticas extraídas de la población mediante un muestreo probabilístico y que serán medidas o encuestadas en la investigación. La muestra está relacionada con la población mediante un proceso de muestreo que consiste en la forma como estas unidades serán seleccionadas entre la población,
- los *datos*, son el resultado de la medición o encuesta realizada a las unidades de la muestra. Estos datos muestrales, recopilados midiendo o encuestando las unidades integrantes de la muestra, son organizados o tabulados, dando como resultado
- las *estadísticas descriptivas*, elemento conformado por gráficas e indicadores (medidas de tendencia central, variabilidad, asociación, etc.) que tiene como propósito describir el comportamiento en la muestra de las variables medidas o analizadas.
- La *inferencia estadística*, es la otra parte fundamental de la metodología estadística y mediante ella, aparece en el proceso estadístico el elemento.
- Los *modelos de probabilidad*. Un modelo de probabilidad es un modelo matemático que cumple con determinadas propiedades y sirve de base para hallar la probabilidad de un grupo de unidades estadísticas que posean determinado suceso o aspecto particular con relación a una característica o variable medida en ellos. Estos modelos de probabilidad explican el comportamiento de la variable analizada en la investigación para toda la población de la cual se obtuvo la muestra, permitiendo así el planteamiento de soluciones a los problemas objeto de estudio en dicha población. Llegar al convencimiento y que la variable analizada, utilizando los resultados de la muestra, se comporte según un determinado modelo probabilístico, permite al investigador utilizar dicho modelo para la estimación de parámetros poblacionales.

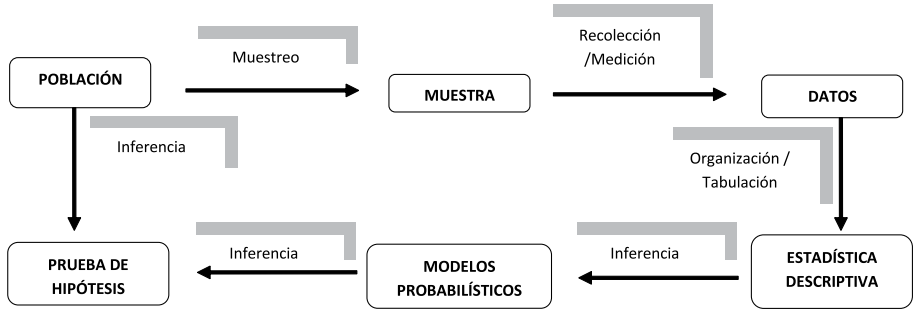


Figura 1.3. **El proceso estadístico en la investigación**

Finalmente, aparece el elemento prueba de hipótesis, el cual está relacionado, de una parte, con los modelos de probabilidad mediante la inferencia estadística y, de otra, con la población mediante el proceso de muestreo. La prueba de hipótesis cierra el círculo del proceso estadístico en la investigación y mediante este elemento se confronta con la realidad el modelo a que se ha llegado en el proceso, al igual que la estimación de sus parámetros.

Objetivo de una encuesta o estudio estadístico

El objetivo de una encuesta o estudio estadístico es el de colectar (reunir) información acerca de una población o muestra, con el propósito de satisfacer necesidades previamente definidas. Existen dos formas básicas de colectar información para propósitos estadísticos: censo y muestreo.

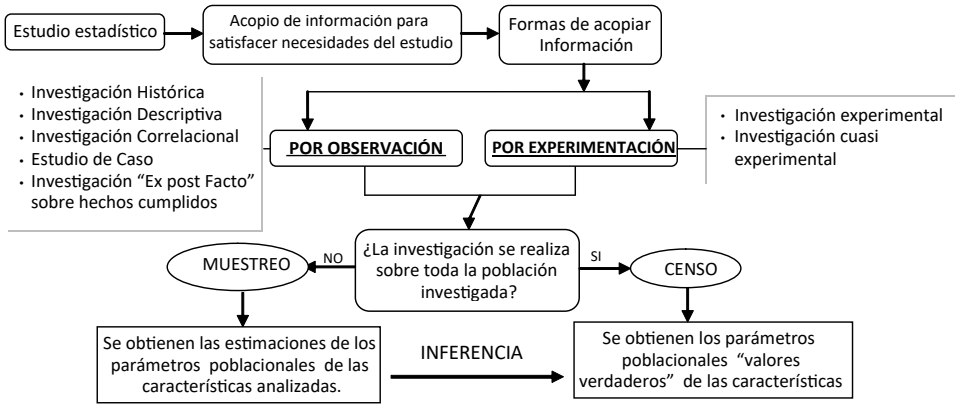


Figura 1.4. **El propósito de un estadístico**

Mediante un censo, observamos o medimos las características de interés en la investigación para todas las unidades que conforman la población objeto de estudio, mientras que si utilizamos un muestreo observamos o medimos las características de interés solo en una parte de dicha población, lo cual permite entonces pensar en la forma como se debe seleccionar esta parte de la población.

La situación anterior induce los llamados métodos de muestreo, que por ahora definiremos como procedimientos científicos y objetivos que permiten seleccionar una parte de la población, a la que investiga. La figura 1.4 presenta un panorama más amplio del propósito de un estudio estadístico.

Tipos de investigación

En la tabla 1.1 se presenta un resumen que define los principales tipos de investigación, estableciendo para cada uno de ellos las características principales (Serie "Aprender a investigar", del ICFES [1999]).

Tabla 1.1. **Los tipos de investigación y sus características**

Tipo de investigación/Definición	Características
Histórica Busca construir el pasado de manera objetiva, con base en evidencias documentales confiables.	• Depende de fuentes primarias y secundarias. • Somete los datos a crítica interna y externa.
Descriptiva Describe características de un conjunto de unidades o áreas de interés.	• Se interesa en describir. • No está interesada en explicar.
Correlacional Determina variaciones de unos factores en relación con otros (covariación).	• Indicada para establecer relaciones estadísticas entre características o fenómenos, pero no conduce directamente a establecer relaciones de causa-efecto entre ellos.
Estudio de caso Estudia intensivamente un sujeto o situación.	• Permite comprender a profundidad lo estudiado. • Sirve para planear después investigaciones más extensas y rigurosas. • No sirve para hacer generalización.
Ex post facto Busca establecer relaciones de causa-efecto, después de que este último ha ocurrido y su causa se ubica en el pasado.	• A partir de un efecto observado, se indaga por su causa en el pasado. • Útil en situaciones en las que no se puede experimentar. • No es muy segura para establecer relaciones causales.

Continúa

Tipo de investigación/Definición	Características
Experimental Es aquella que permite con más seguridad establecer relaciones de causa-efecto.	• Usa grupos experimentales y grupos de control. • El investigador manipula el factor supuestamente causal. • Usa procedimientos al azar para seleccionar y asignar sujetos y tratamientos. • Es artificial y restrictivo.
Cuasi-experimental Estudia relaciones de causa-efecto, pero no en condiciones de control riguroso de todos los factores que pueden afectar el experimento.	• Apropiado en situaciones naturales en donde no es plausible el control experimental riguroso.

CARACTERÍSTICAS QUE DIFERENCIAN EL CENSO Y EL MUESTREO

Tabla 1.2. **Diferencias entre el censo y el muestreo**

Diferencias entre el censo y el muestreo		
Factor analizado	**Censo**	**Muestreo**
Costo	Muy costoso	Más económico
Duración	Demorado	Menos demorado
Procedimiento	Complicado	Más simple
Error muestral*	No tiene	Sí tiene
Error no muestral**	Puede ser muy grande	Mucho menor (más controlado)
Impacto	Afecta a toda la población	Afecta a una parte de la población

* Error muestral: Diferencia entre la realidad (parámetro) y la estimación de dicho parámetro basada en la muestra.

** Error no muestral: Generado por el procedimiento de censo o muestreo, durante la operación de recolección de información como, por ejemplo, preguntas mal formuladas, mediciones mal realizadas, marcos imperfectos e inadecuados, etc.

La tabla 1.2 ilustra las diferencias existentes entre los dos métodos de recolección de información, censo y muestreo para los siguientes factores: costo; duración; procedimiento; error muestral (se definirá más adelante); error no muestral e impacto. Estas diferencias están precisamente determinadas por la naturaleza de cada uno de los métodos, lo que permite caracterizarlos mejor.

DATOS ESTADÍSTICOS

En presencia de cualquier situación problema que surja en la industria o en la naturaleza, debemos decidir si esta amerita estudio (investigación) o no, lo cual dependerá de la magnitud de ella, de qué tan trascendentales sean sus consecuencias y de su duración.

Una vez que se ha decidido estudiar el problema, este debe clasificarse en estudio descriptivo o estudio inferencial. Un estudio será descriptivo cuando su objetivo es describir las características de una población o muestra. De otro lado, se dirá que el estudio es inferencial cuando, basado en los resultados obtenidos mediante una muestra, se infieren por extrapolación algunas características poblacionales. La realización de ambos estudios genera datos estadísticos que pueden ser de naturaleza numérica (variables) o no numérica (atributos) (ver la figura 1.5). Los datos de naturaleza numérica se clasifican a su vez en variables discretas y variables continuas.

Nota: Los datos cualitativos pueden ser numéricos, pero por asignación, nunca por naturaleza.

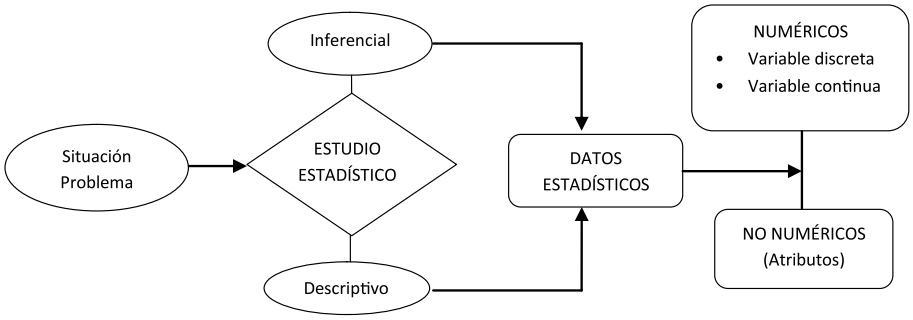


Figura 1.5. **Clasificación de los datos estadísticos**

Ejemplo 1.1

A continuación, se relacionan algunas situaciones que generan datos estadísticos.

Variables discretas

- Número de pernos de aluminio defectuosos por caja. Cada caja contiene 12 unidades.
- Número de árboles muertos en determinada plantación de pino.
- Número de vehículos accidentados en determinada autopista a las 12 del día.
- Número de automóviles que pasan todas las revisiones en cada uno de los turnos de 8 horas en una ensambladora.

- Número de frutos de un árbol.
- Número de remaches defectuosos por m^2 de lámina de una aeronave.
- Número de especímenes (muestra de tamaño o cantidad específica de tierra) contaminados con cianuro en determinada zona de vocación agrícola.
- Número de secciones de 24 pies de tubo cilíndrico de hormigón preesforzado (PCCP), que presentan grietas por esfuerzo en la observación de un grupo de 20 secciones.
- Número de fusibles defectuosos de un grupo de 100 revisados cada hora en un período de 25 horas.
- Número de tubos de rayos catódicos defectuosos encontrados en una muestra de 50, realizada diariamente durante un mes.

Variables continuas

- Peso neto de un determinado producto al ser empacado.
- Longitud de un trozo de fibra de madera macerado y visto al microscopio.
- Longitud de una viga de acero tipo SAE 210.
- Diámetro de un tubo de PVC en pulgadas.
- Contenido de gas carbónico (CO2) en volumen por botella de cerveza.
- Contenido de cemento de una bolsa cuya capacidad se especifica en 50 kg.
- Resistencia al corte en libras de una lámina de acero utilizando soldadura de punto ultrasónica.
- Tiempo de oxidación-inducción en minutos para cierta cantidad de aceite comercial.
- Diámetro de un álamo temblón a la altura de pecho en determinado bosque boreal.
- Concentración de plomo en miligramos/litro para una muestra de agua.

Atributos

- Calidad del agua potable de una ciudad (excelente, buena, regular, mala).
- Variedad de un árbol encontrado en determinada región natural.
- Calificación del daño sufrido por un automóvil como E (estructural) o NE (no estructural), después de ser sometido a un choque a la velocidad de 5 millas por hora.
- Presencia o ausencia de cáncer en un niño que vivió más de ocho años en una casa prefabricada con materiales en cuya fabricación se detecto la presencia de radón.
- Determinar, para un producto terminado, la respectiva máquina en el cual fue llenado y sellado. Hay cinco máquinas llenadoras-selladoras.

- Determinar, para un producto, si la queja del usuario se debe a un aspecto del producto, a la dificultad de adquisición o a la garantía establecida.
- Un concreto de microsílice, reforzado con fibra de acero y mezclado en húmedo, se dice que cumple a satisfacción si falla después de 60 días de ser sometido a una compresión de 9000 psi. Determinar para una muestra de dicho concreto, si cumple o no a satisfacción.
- Determinar si un sello de caucho que resiste más de 3000 libras por pulgada cuadrada fue fabricado mediante un proceso A, B, C o D.
- Clase de metal del cual está hecho determinado eje.

LA VALIDEZ DE UN ESTUDIO ESTADÍSTICO

Para la realización de una investigación de calidad es básico que la realidad investigada quede representada correctamente en la estrategia de investigación utilizada o, de lo contrario, este proceso carecerá de validez; el investigador deberá en todo momento realizar el análisis bidireccional representado en la figura 1.6.

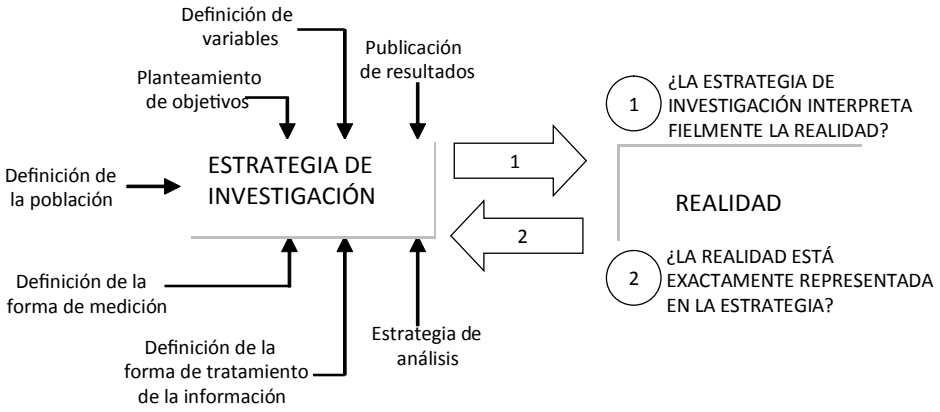


Figura 1.6. **Análisis bidireccional realizado en toda investigación estadística**

Una vez que se ha definido y planteado muy bien el problema objeto de estudio (se tengan claros los objetivos y definiciones así como las preguntas que lo resumen al igual que la forma de medir y resumir las características que le permitirán al investigador dar respuesta a los objetivos planteados), es necesario reflexionar sobre la validez (credibilidad y confianza) de la investigación, pues se quiere que los resultados obtenidos sean de muy buena calidad, para que así mismo las decisiones sobre la población implicada en la investigación estén lo más cercanas posibles a las ideales.

La validez de una investigación estadística tiene las siguientes componentes básicas: validez externa, la cual tiene que ver con la forma de muestrear; la validez de constructo, relacionada con la forma de medir las características analizadas; la validez interna, justificada con la forma de modelar los fenómenos; de las conclusiones, fundamentada por la forma como razona el investigador; y, finalmente, la validez de la comunicación, alimentada por la forma de presentación de los resultados de la investigación. La figura 1.7 muestra las principales fases consecutivas del desarrollo de un proyecto y sus relaciones con los diferentes componentes de la validez.

La validez de un estudio puede mirarse esquemáticamente como un proceso cuya materia prima son los datos, que están sometidos a un proceso (de pensamiento) y del cual saldrá un producto, que es la información que será puesta en contexto para dar solución a interrogantes del problema planteado (Behar *et al.*, 2000).

La validez externa

Este tipo de validez está relacionada con la selección de las unidades estadísticas que serán medidas o encuestadas, aquí la muestra debe ser representativa de la población, de la selección dada, y para ello se necesita que la forma como ella sea seleccionada no tenga sesgos, lo cual se garantiza mediante el uso del muestreo probabilístico.

El otro aspecto que se debe controlar en el estudio para garantizar la validez externa es el de buscar por todos los medios que las mediciones resultantes sean confiables, es decir que no tengan sesgos de ninguna naturaleza. La validez externa da en buena medida la capacidad de generalizar permitiendo la utilización de las técnicas estadísticas, las cuales tienen su base y efectividad en el muestreo probabilístico. El control que se establezca para garantizar la validez externa determina igualmente el diseño de la investigación; así, por ejemplo, para mejorar la validez externa, se debe centrar la atención en el tipo de muestreo y su tamaño, la necesidad de replicar el experimento y el uso de teoría estadística. En la búsqueda para garantizar esta componente de la validez debe tenerse claridad sobre la respuesta a las siguientes preguntas:

a) ¿A quién se aplican los resultados de la investigación?
b) ¿Hasta dónde pueden generalizarse los resultados de la investigación?

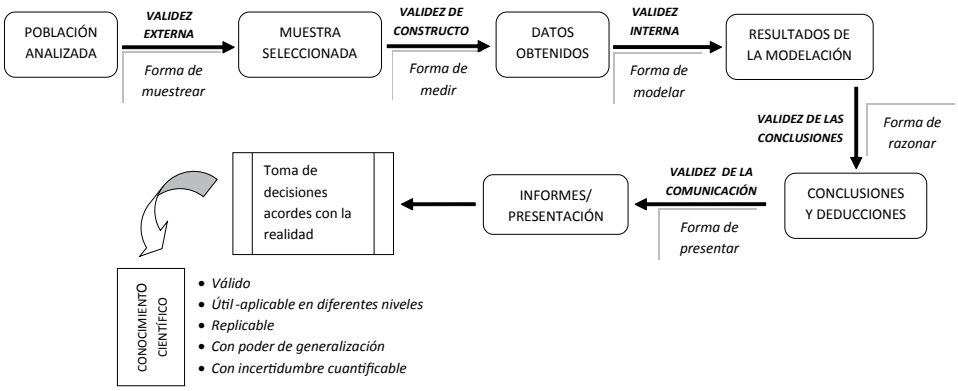


Figura 1.7. **La validez en la generación de conocimiento**

La validez de constructo

Este tipo de validez hace referencia a lo que se desea medir y la forma de medirlo, se hace necesario en este caso tener muy claros los conceptos involucrados en las diferentes preguntas o mediciones que se realizaran.

Se garantiza que una buena pregunta o medición tiene validez cuando:

a) Las respuestas a la pregunta por "individuos iguales" son parecidas (fiables): el instrumento mide lo que se quiere, es decir es fiable.
b) Las respuestas a la pregunta por "individuos iguales" no deben presentar tendencias (sesgos) que las alejen de la verdad.

La validez interna

Esta se refiere a qué tan apropiado para la investigación es el análisis que se está realizando, qué tan válidas son las comparaciones que se hacen y a la existencia o no en el estudio de los posibles factores de confusión que necesariamente invalidan las conclusiones obtenidas.

Los factores de confusión son variables que se encuentran presentes en las muestras o en las unidades que van a compararse, de tal forma que su presencia modifica la relación de causalidad buscada o contrastada mediante el estudio. Así, por ejemplo, al comparar dos estudiantes respecto de su rendimiento escolar, al ser sometidos al mismo método de enseñanza, es necesario que para el caso de ambos los siguientes aspectos estén controlados: sexo, salud, alimentación, edad, horas de estudio, profesor, materiales de apoyo, relación familiar, etc. De lo contrario, alguno de los aspectos ejerce influencia en uno de los dos estudiantes, invalidando así la fuerza de la comparación de ellos respecto del método de enseñanza. La validez interna intenta entonces garantizar que las comparaciones se hagan entre individuos y situaciones comparables.

La validez de las conclusiones
La falta de validez en las conclusiones tiene que ver con el uso de herramientas de análisis inapropiadas por parte del investigador para el diseño planteado, o simplemente sacar conclusiones equivocadas, alterar o desconocer los supuestos metodológicos, así como aplicar equivocadamente los conceptos de la metodología estadística.

La validez de la comunicación
La validez de la comunicación tiene que ver con la forma y el contenido de los informes realizados por el investigador: escritura, gráficas, cuadros etc.

EL MUESTREO: UNA NECESIDAD EN EL MUNDO DE HOY

Uno de los problemas que afronta el mundo de hoy es que los conjuntos de cosas son demasiado grandes y resulta supremamente difícil conseguir la información que sobre ellos se requiere para tomar decisiones. Basta con analizar unas pocas situaciones para darse cuenta de la importancia de la metodología del muestreo en la actualidad, ante la imposibilidad de medir o encuestar a todas las unidades sobre las cuales se desea actuar (figura 1.8).

Ejemplo 1.2
- Situación 1: Se desea planificar el desarrollo de una gran ciudad y para ello se debe realizar un estudio socioeconómico en su área metropolitana y en sus áreas rurales de influencia.
- Situación 2: Se desea controlar la calidad de los productos manufacturados por una gran empresa, con el fin de garantizar los estándares de calidad de dichos productos para ser competitivos en el mercado.
- Situación 3: Se desea conocer la producción total de determinado producto agrícola en una amplia extensión territorial compuesta por una gran cantidad de predios con vocación agrícola.

Como se muestra, en los tres ejemplos anteriores sería muy dispendioso, costoso y demorado obtener y procesar toda la información que resultaría de medir las características deseadas en cada elemento de estos conjuntos (censar), haciendo realmente imposible tal procedimiento.

Lo anterior pone de manifiesto que es mucho más práctico desde todo punto de vista tomar una parte de estos conjuntos (muestra) y con sus medidas hacerse una idea (estimar) de lo que se quiere saber sobre todo el conjunto (población).

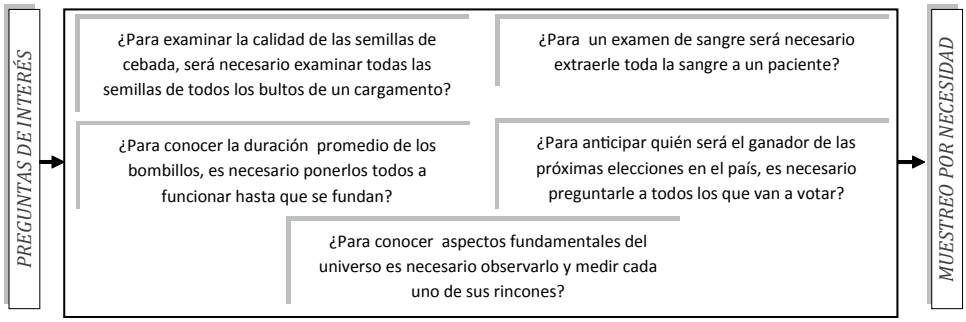


Figura 1.8. **Algunas situaciones que obligan el uso del muestreo**

Es bueno tener presente que al tener dos poblaciones muy diferentes en su tamaño, por dar un ejemplo: N_1 de 3000 unidades y N_2 de 10000 unidades de las mismas, esto no implica que al requerir para un correcto estudio de 80 unidades en la población #1, entonces se necesite, debido al tamaño de la #2, más de 80 para ella. Es posible que una unidad más (81), tomada para las dos poblaciones, no agregue diferencia significativa a lo que ya se sabía sobre las poblaciones con las 80 unidades estudiadas. Así mismo, se debe tener presente la homogeneidad de las unidades de una población respecto a las características que se desean estudiar, este hecho hace que necesariamente el número de unidades o muestra requerida para el estudio sea relativamente pequeño. Sobre estos conceptos se volverá más adelante.

¿Por qué muestrear?

Son muchos los factores que determinan en una investigación sobre una población particular el uso de la metodología del muestreo para llevarla a cabo. Entre las razones que obligan a ello, se pueden enumerar las siguientes:

- La población objeto de estudio es infinita o muy grande.
- La observación o medición de la característica analizada pone en peligro la existencia de la unidad misma.
- El costo de la medición.
- La dificultad de ubicación de las unidades objeto de estudio.
- El tiempo necesario para la realización del trabajo de campo (ubicación y medición de las unidades objeto de estudio).
- La ausencia de marcos muestrales específicos en los cuales se puedan ubicar detalladamente las unidades investigadas.

Una vez determinada la necesidad de utilizar el muestreo para la realización de una investigación, se deben balancear, de una parte, la precisión deseada (diferencia entre la estimación lograda por el muestreo y el verdadero valor del parámetro investigado), situación que será definida más adelante, como la *ecuación de precisión*. Es lógico pensar que entre más unidades de la población se observen más cerca estará la estimación del parámetro; la otra parte integrante del balance son los costos totales de medición, los cuales son directamente proporcionales a la cantidad de unidades elementales investigadas. La figura 1.9 combina el comportamiento de la precisión con el comportamiento de los costos mostrando un balance que haga eficiente la investigación.

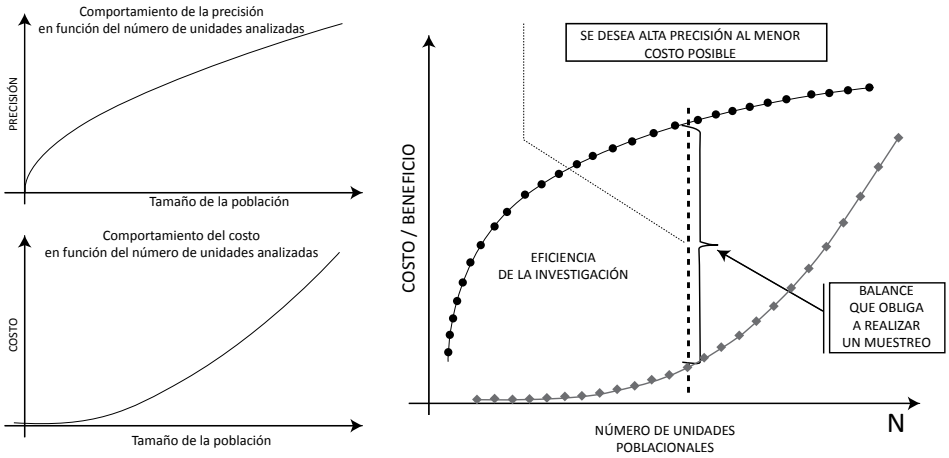


Figura 1.9. **Balance para la obtención de una buena muestra**

Cuando se realiza un muestreo, de acuerdo con la estructura de la realidad analizada frente a una situación particular, si ella es desconocida, la función del muestreo es la de diagnosticar (estimación de parámetros y visualización del comportamiento de su distribución), por el contrario, si la situación es conocida (o se tiene algún supuesto de su valor o comportamiento), la función del muestreo será la de verificar un diagnóstico, lo que equivale a la realización de pruebas de hipótesis basadas en las muestras (ver la figura 1.10).

Muestreo e inferencia

El objetivo de la inferencia estadística es precisamente el de obtener o formular predicciones o decisiones sobre una población, con base en la información obtenida de una *muestra* seleccionada por métodos probabilísticos.

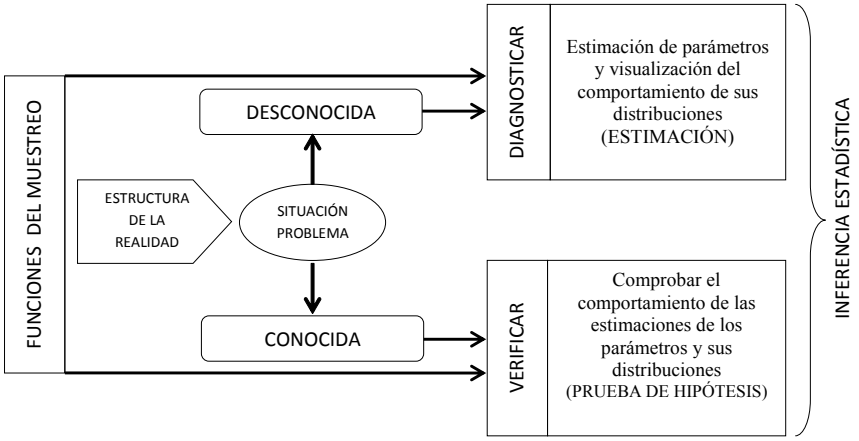


Figura 1.10. **Funciones del muestreo**

Suponga que en una fábrica de llantas para vehículos livianos se desea controlar el peso de los neumáticos de las llantas de código A0001, y para ello se toma una muestra de 100 neumáticos. Se desea saber si el peso de los neumáticos procede de una población normal y cuáles son los parámetros de dicha población, igualmente se desea conocer ¿qué fracción de neumáticos serán rechazados en el mercado por exceso y por defecto si se sabe que el peso correcto del neumático debe ser $1500\ grs \leq X \leq 1600\ grs$?

Las respuestas a estas preguntas son obtenidas mediante el uso de la inferencia estadística, la cual a su vez se alimenta de los resultados obtenidos mediante la aplicación de la estadística descriptiva a los datos que fueron obtenidos de la muestra seleccionada.

Suponga también que la empresa necesita revisar el proceso de fabricación y especialmente los parámetros que inciden en el peso del neumático tales como: longitud, ancho y espesor, para que la calidad mejore y así disminuir el consumo de material y por ende el costo por neumático. Se tiene entonces que la decisión que se tomará sobre la producción (población total de neumáticos) depende fundamentalmente de los resultados de la inferencia, los cuales a su vez deben tener como base una muestra probabilística correctamente diseñada.

La figura 1.11 muestra la clasificación de las preguntas claves que debe hacerse todo investigador a la hora de inferir con base en una muestra.

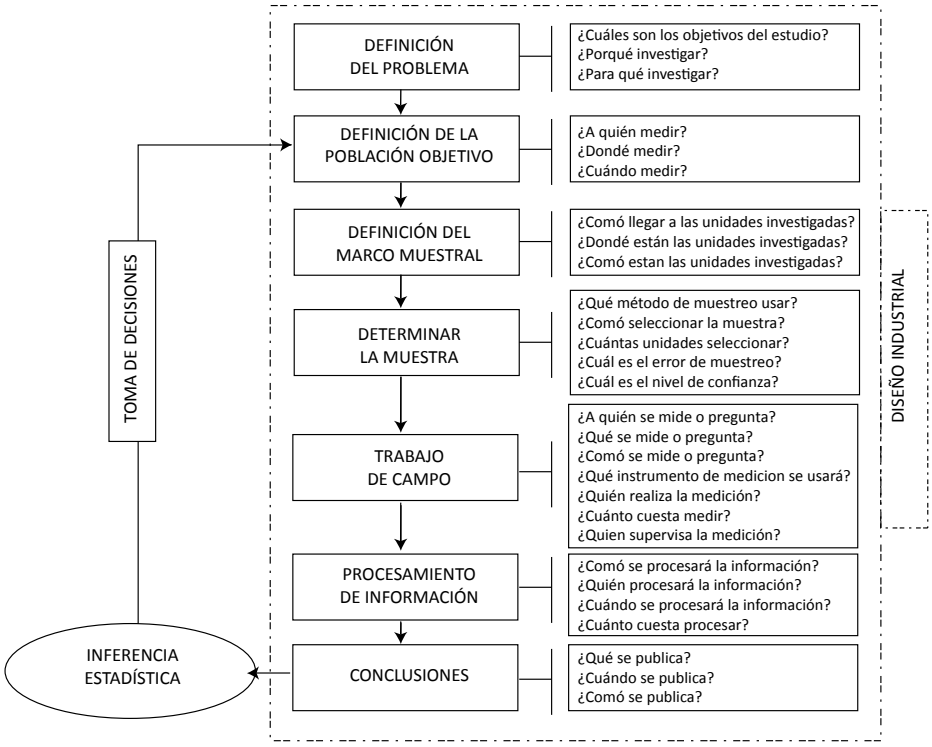


Figura 1.11. **Preguntas básicas a la hora de inferir**

Muestreo: teoría y empirismo

Del análisis de los datos procedentes de un estudio estadístico aplicado a una determinada población, obtenidos mediante una *muestra* de unidades, se generan interrogantes, los cuales son a su vez motivo para el surgimiento de nuevas teorías que darán respuesta a dichos interrogantes.

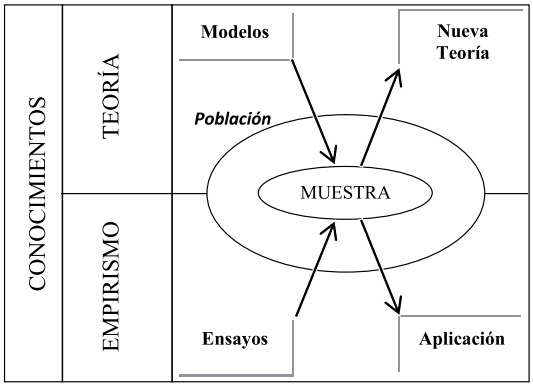


Figura 1.12. **La teoría y el empirismo en la generación de conocimiento**

Sin embargo, los modelos teóricos generados mediante investigación teórica son implementados a través del análisis de datos, produciendo aplicaciones concretas de la teoría. Existe de esta manera un estrecho vínculo entre la teoría y el empirismo a través del análisis de datos, los cuales son obtenidos mediante la aplicación correcta de la teoría del muestreo. La relación entre la teoría y el empirismo focalizada en el análisis de datos puede apreciarse en la figura 1.12.

Características deseables de una investigación por muestreo

Cuando se realiza una investigación por muestreo, es deseable que ella cumpla con unas características básicas, estas se relacionan y explican someramente en la figura 1.13.

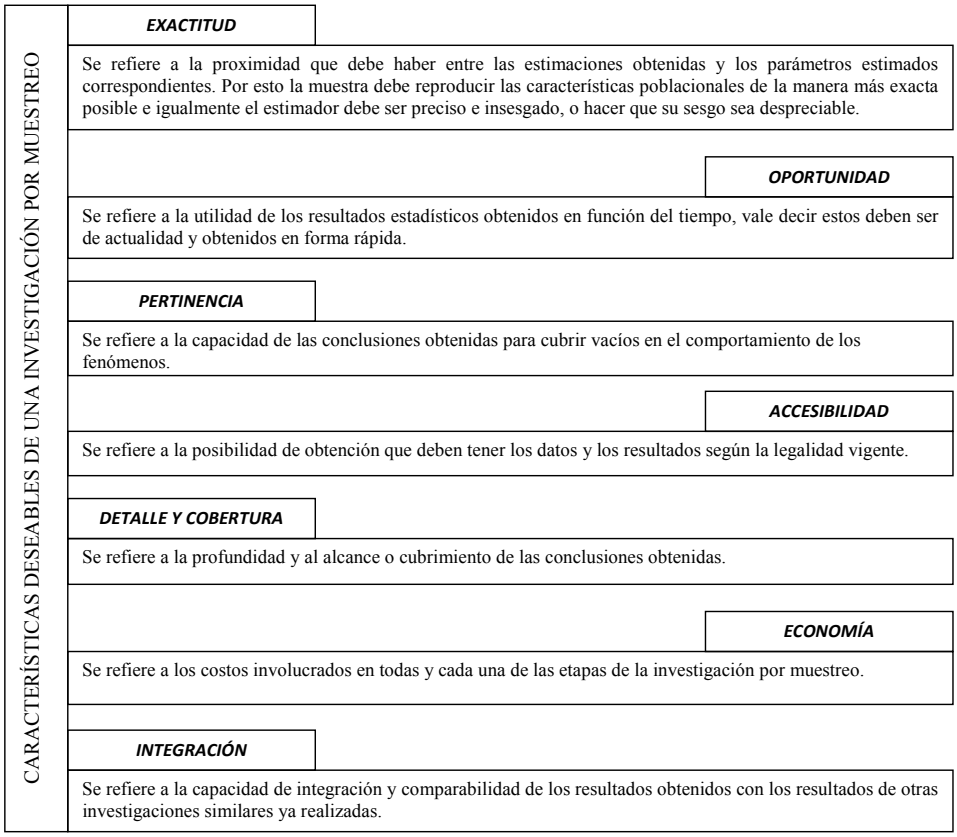


Figura 1.13. **Características deseables de una investigación por muestreo**

LOS MÉTODOS DE MUESTREO

Gracias a las muestras, el ser humano adquiere conocimientos para desarrollarse en su vida cotidiana y científica, mucho de lo que hoy sabemos ha sido deducido mediante el análisis de una muestra "típica", esta se define como aquella en donde su contenido de información reproduce en lo fundamental las tendencias, variaciones y características generales de la población que representa.

La tipicidad de una muestra se puede lograr mediante un adecuado procedimiento de selección de las unidades que la conformaran. En este sentido, se dice que un procedimiento de selección de las unidades de la muestra puede basarse o no en mecanismos aleatorios mediante los cuales se conozca con anticipación la probabilidad que cada unidad poblacional pertenezca o no a la muestra.

El hecho de que la selección de una muestra esté basada en el cálculo de probabilidades permitirá la aplicación de las herramientas de la inferencia estadística, haciendo que las conclusiones obtenidas del tratamiento de la información recogida de esta tengan carácter científico y puedan aplicarse a toda la población de la cual se extrajo la muestra.

En caso de que la selección de las unidades muestrales esté basada o no en el cálculo de probabilidades, el muestreo se clasifica en *muestreo probabilístico* o *muestreo no probabilístico* (para saber más ver Sierra, 1994).

Muestreo probabilístico

En estos métodos, todas las unidades seleccionadas para la muestra tienen una probabilidad conocida de inclusión en ella y su selección se realiza mediante procesos aleatorios. Un proceso de selección aplicado a un conjunto de datos o unidades se considera aleatorio si este da a cada uno oportunidad igual de ser elegido, esto puede lograrse mediante el uso de una tabla de números aleatorios, dispositivos manuales (papeletas marcadas o fichas/baloteras, dados) o dispositivos electrónicos mediante el uso de software, etc.

Propiedades del muestreo probabilístico

Los métodos de selección probabilística cumplen con las siguientes propiedades básicas:

- Dada una población de tamaño N, es siempre posible definir un conjunto de k muestras de tamaño n ($n < N$).

$$S_i = \{u_1, u_2, u_3, .., u_n\}; \quad i = 1, 2, 3, ..., k$$

Donde:
S_i: Muestra i.
u_j: Elemento j de la muestra.
k: Todas las muestras posibles.

- A cada una de las k muestras es posible asignarle una probabilidad conocida de selección:

$$P_r(S_i)=\pi_i, \quad \forall\, i=1,2,3,...,....,k, \Rightarrow \begin{cases} \pi_i \geq 0, \ \forall\, i \\ \sum_{i=1}^{k} \pi_i = 1 \end{cases}$$

- El proceso de selección de cada una de las muestras es aleatorio y depende de su probabilidad π_i.
- El proceso de selección de una muestra garantiza que todo elemento de la población tenga una probabilidad diferente de cero de pertenecer a una muestra seleccionada
- El procedimiento para obtener cualquier estimación con base en la muestra, debe estar definido y conducir a un único valor respecto de la misma.

Nota:
Se llama diseño muestral al conjunto $D=\left\{S_i,\ P_r(S_i);\ \hat{\theta}(S_i)\right\},\ i=1,2,3,...,k$ donde:

S_i: Muestra *i*-ésima de tamaño n.
$P_r(S_i)$: Probabilidad de seleccionar la muestra i.
$\hat{\theta}$: Estimador del parámetro θ.

Las anteriores propiedades permiten que el investigador conozca el grado de representatividad de la muestra seleccionada y estime el error cometido al realizar inferencias sobre la población con base en su muestra.

Fundamentos del muestreo probabilístico
El muestreo probabilístico o científico tiene como base los siguientes enunciados o principios: dos axiomas, un teorema y dos leyes.

- ***Axiomas***

Los axiomas son:

Axioma 1
La muestra, como se ha definido antes, es un conjunto de eventos casos o unidades que son tomadas o extraídas de una población siguiendo determinados procedimientos, los cuales buscan que las unidades muestreadas

representen fielmente lo más que se pueda la población que representa. En este sentido, el diseño óptimo de una muestra, en particular la determinación previa de su tamaño, solo puede conseguirse a partir del conocimiento de la población.

De acuerdo con lo anterior, podemos establecer el siguiente axioma:

Todo trabajo con muestras para obtener inferencias requiere del conocimiento previo del universo o población de donde se extraen dichas muestras.

Axioma 2

Una de las etapas básicas de una encuesta por muestreo es la determinación o definición de la población objeto de estudio en la investigación. En la definición de la población objeto de estudio, es el propio investigador quien determina cuáles de las características presentes en las unidades del universo permiten definir la población objeto de estudio como tal. Lo anterior significa que (a) la propiedad de ser población objeto de estudio no está dada, si no que es declarada por el investigador y (b) para definir o determinar una población objeto de estudio, se requiere al menos de una característica, la cual debe aparecer en todas las unidades que la conforman. Estas características permiten tener claridad sobre si una unidad en particular pertenece o no a la población objeto de estudio.

De acuerdo con lo anterior se puede establecer el siguiente axioma:

La similitud entre las unidades que conforman una población objeto de estudio son distinciones hechas por el investigador.

- ***Teorema***

La propiedad más importante, entre todas las propiedades que tiene la distribución de la media muestral de datos, es el teorema del límite central.

Teorema del límite central

Si $X_1, X_2, X_3, ..., X_n$ son n variables aleatorias independientes, distribuidas idénticamente (todas tienen la misma función de probabilidad en el caso discreto o función de densidad en el caso continuo) con media μ y varianza σ^2 finitas, entonces para la suma SM_n se cumple que:

$$SM_n = X_1 + X_2 + X_3 + ...+ X_n = \sum_{i=1}^{n} X_i \,, \quad \lim_{n \to \infty} P_r\left(a \le \frac{SM_n - n.\mu}{\sigma.\sqrt{n}} \le b \right) = \frac{1}{\sqrt{2.\pi}} \int_a^b e^{-t^2/2} dt$$

Es decir que la variable aleatoria $\overline{X}_n = \frac{SM_n}{n} \cong N(\mu, \sigma^2/n)$, cuando n se hace muy grande. "A medida que aumenta el tamaño de la muestra, la media muestral, se distribuye normal, sin importar la distribución de la población, de la cual procede la muestra".

Para efectos prácticos, este hecho generalmente ocurre cuando $n \geq 30$. La figura 1.14 muestra el significado del teorema del límite central para tres distribuciones, muy comunes en el trajinar estadístico como lo son las distribuciones: normal ($\mu = 0$, $\sigma = 1$), uniforme ($\mu = 0.5$, $\sigma = 0.29$) y exponencial negativa ($\mu = 1$, $\sigma = 1$).

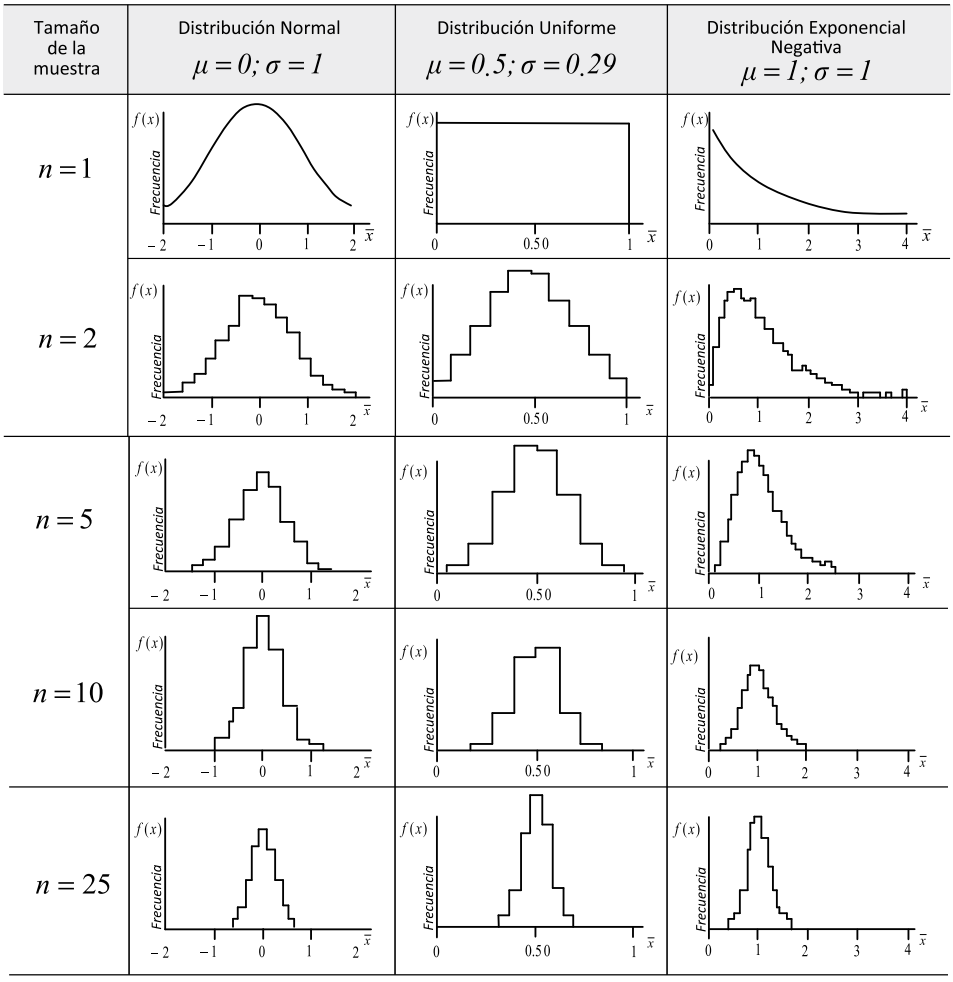


Figura 1.14. **Teorema del límite central y la distribución de la media muestral**
Fuente: Mendenhall (1990).

- ***Leyes***

La primera ley de los grandes números

Si $X_1, X_2, X_3, ..., X_n$ son n variables aleatorias independientes, distribuidas idénticamente (todas tienen la misma función de probabilidad en el caso discreto), con valor esperado y varianza iguales, respectivamente a:

$$E(X_i) = P,\ V(X_i) = P.(1-P) = P.Q,\quad X_i = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix}$$

La variable $SM_n = X_1 + X_2 + X_3 + ... + X_n = \sum_{i=1}^{n} X_i; n \geq 1$ puede tomar los valores 0, 1, 2,… n y tiene distribución binomial con $E(SM_n) = n.P$ y $V(SM_n) = n.P.(1-P) = n.P.Q$. Entonces si c es un número real positivo, tal que $\lim_{n\to\infty} P[|(SM_n/n) - P| < c]=1$ Es decir que cuando n aumenta, la variable SM_n/n tiende cada vez a P. La variable $(SM_n/n)= \hat{P}$ es la frecuencia relativa de valores 1, llamada comúnmente "proporción muestral". En estos términos se dice que la variable $(SM_n/n)= \hat{P}$ es ***consistente***, significando con ello que a medida que n aumenta, $\hat{P}$ se acerca cada vez más a la proporción poblacional P.

La segunda ley de los grandes números

Si X_i ; $i \geq 1$ es una secuencia de variables idénticamente distribuidas, con media y varianza respectivas iguales a $E(X_i) = \mu$ y $V(X_i) = \sigma^2$. Al considerar la variable $SM_n = X_1 + X_2 + X_3 + ... + X_n = \sum_{i=1}^{n} X_i; n \geq 1$ con media igual a:

$$\overline{X}_n = \frac{SM_n}{n} = \frac{X_1 + X_2 + ... + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Se cumple que para $\forall \in > 0 : \lim_{n\to\infty} P[|\overline{X}n - \mu| < \in] = 1$; significando con ello que la media muestral es ***consistente***, es decir que al aumentar el tamaño de la muestra, la media muestral se acerca cada vez más a la media poblacional.

Se dice que: "$\lim_{n\to\infty} \overline{X}n=\mu$ en probabilidad", significando que: cuando n se hace muy grande entonces $\overline{X}_n$ se hace igual a μ.

Principales métodos de muestreo probabilístico

Los métodos probabilísticos de muestreo están soportados por las estructuras del azar, cuyos méritos son dos (Silva, 2000); de una parte, los procedimientos probabilísticos satisfacen la exigencia de eliminar o mitigar la subjetividad en la elección de las unidades que integrarán la muestra y, de otra parte, solo cuando se aplica un método probabilístico se podrá medir el grado de precisión de las estimaciones. A continuación, se explican los métodos de muestreo probabilísticos básicos.

Muestreo aleatorio simple

Este método de muestreo es el método probabilístico más estudiado y gracias a las propiedades que él posee con relación a la estimación de parámetros y los errores de muestreo, se ha convertido en el método base para la realización de los otros métodos probabilísticos de muestreo.

Definimos el método de muestreo aleatorio simple como el método que consiste en seleccionar una muestra de n elementos uno a uno de una población N, de tal forma que cada unidad poblacional tenga una probabilidad idéntica y conocida de pertenecer a la muestra. Aquí, cada elemento se selecciona en forma independiente de los demás y su extracción se realiza mediante un proceso aleatorio, a partir de un marco de muestreo.

Existen dos tipos de muestreo aleatorio: (a) *muestreo con reposición*, cuando una unidad cualquiera que ha sido seleccionada se repone (se devuelve) a la población antes de seleccionar la siguiente unidad, manteniendo constante la estructura de la población; (b) muestreo sin reposición, en cuyo caso la muestra se obtiene elemento por elemento, sin reponer estos (no se devuelven) a la población antes de cada extracción, de este modo cambia en cada extracción la estructura de la población.

Ejemplo 1.3

Suponga que se desea realizar una encuesta al interior de una unidad residencial compuesta por 200 apartamentos, con el objetivo de tomar decisiones sobre el manejo administrativo y de servicios que tiene la unidad. Para ello, se toma el listado de todas las familias propietarias o inquilinas de los apartamentos y después de asignarles un código desde el 001 hasta el 200, se procede a generar tantos números aleatorios comprendidos en este rango, como tamaño tenga la muestra aleatoria de familias que se haya definido.

Muestreo aleatorio estratificado

En este método se divide la población en subpoblaciones o partes llamadas *estratos*, en los cuales se debe buscar la homogeneidad de sus unidades y la heterogeneidad entre grupos. Posteriormente, se selecciona en forma independiente de cada estrato una muestra aleatoria simple y, finalmente, la muestra total será la suma de las muestras de cada uno de los estratos.

Ejemplo 1.4

Los vehículos destinados al servicio de las diferentes dependencias del municipio de Cali, se agrupan en las siguientes categorías o estratos de acuerdo con su naturaleza y uso: categoría 1, buses y camiones; categoría 2, automóviles; categoría 3, maquinaria pesada, y categoría 4, motos. Con el fin de realizar un estudio técnico y de mantenimiento sobre estos vehículos, se realiza un muestreo aleatorio proporcional en cada categoría conformando de esta manera una muestra con participación de todas las categorías. Los vehículos así seleccionados serán enviados a los talleres para su evaluación técnica.

Muestreo aleatorio por conglomerados

En este método, las unidades elementales se encuentran formando grupos de unidades más grandes llamadas *conglomerados* (por razones administrativas, geográficas, o de proceso y manejo). Se selecciona un grupo de conglomerados usando un muestreo aleatorio simple o estratificado, y en cada conglomerado se analizan o miden todas las unidades elementales que la componen o bien una parte de ellas, utilizando nuevamente un muestreo aleatorio simple dentro de cada conglomerado. Es importante tener presente que la eficiencia de este método se presenta cuando al interior de cada grupo o conglomerado hay alta heterogeneidad entre las unidades y al comparar los grupos respecto de sus medias hay alta homogeneidad.

Ejemplo 1.5

En una fábrica de refrescos se desea medir el contenido promedio de líquido por botella para mantener el estándar de calidad. El producto se empaca en canastillas de 12 unidades c/u. Para tal propósito, se toma una muestra aleatoria de 5 cajas de refrescos y se determina en cada una de las botellas de cada caja el contenido exacto de líquido. Este análisis debe ser realizado al comienzo de cada turno en la planta, con el fin de calibrar las máquinas embotelladoras del refresco.

Muestreo sistemático

Las unidades elementales aparecen en listas o presentan una secuencia que permite la selección de una muestra a saltos. Para ello, se elige un punto aleatorio (primera unidad seleccionada) dentro de un grupo de k unidades y posteriormente las restantes unidades se escogen a saltos cada k unidades de la lista o secuencia. Aquí: N es el tamaño de la población; n, el tamaño de la muestra, y k, un entero tal que $N = n.k$.

Ejemplo 1.6

Con el objetivo de medir el nivel de contaminación de metales pesados en un terreno de forma irregular aledaño a una empresa industrial, la cual deposita en él sus desechos industriales, se ha levantado un plano del terreno donde se ha construido una rejilla que permite dividirlo en 50 unidades muestrales de determinada área (estas unidades tienen cómo mínimo la tercera parte de terreno incluido en la rejilla). Se selecciona una muestra de tamaño 10, de 5 en 5, partiendo de la unidad # 2 y siguiendo una ruta estipulada. Posteriormente, una vez en el terreno se selecciona una muestra de lodo de un punto correspondiente a la unidad determinada en el plano. Finalmente, las 10 muestras de lodo son enviadas al laboratorio para su respectivo análisis.

Muestreo multietápico (muestreo con submuestreo)

En este método de muestreo se selecciona una muestra aleatoria simple de conglomerados o unidades primarias (primera etapa), y entre estas unidades se seleccionan aleatoriamente otras unidades secundarias (segunda etapa) y así sucesivamente, hasta completar todas las etapas que se hayan previsto en el diseño para llegar a la unidad que finalmente será encuestada o medida. Lo más frecuente en este método es que haya dos etapas, en cuyo caso hablamos de *muestreo bietápico*.

Ejemplo 1.7

Con el fin de controlar la explotación del atún, en temporada alta en un determinado puerto marítimo, se realiza un muestreo por conglomerados en dos etapas de la siguiente manera: Se seleccionan aleatoriamente 20 embarcaciones pesqueras, de 400 que llegan a dicho puerto para el desembarco del atún. Posteriormente, dentro de cada una de las embarcaciones escogidas se realiza un muestreo aleatorio de algunas de sus bodegas (cada nave tiene entre 5 y 9 bodegas) y se determina el número de unidades de pescado que no tienen el tamaño establecido para su pesca.

Tabla 1.3. **Ejemplos de unidades primarias y secundarias en un muestreo bietápico**

Variable de interés	Unidades primarias UPM	Unidades secundarias UPS
Habitantes por vivienda	Municipios	Viviendas
Cantidad total de carne	Fincas ganaderas	Ganado vacuno
N.º de estudiantes que viven con sus padres	Colegios de secundaria	Estudiantes matriculados
N.º de estudiantes en una ciudad	Colegios	Salones de clase
N.º de mascotas en un municipio	Manzanas	Viviendas

Algunos ejemplos de unidades primarias y sus respectivas unidades elementales (objeto de análisis) o secundarias para la realización de un muestreo bietápico se muestran en la tabla 1.3.

El área de aplicación, la forma y la mezcla de los diferentes tipos de muestreo, generan algunas variedades específicas de métodos probabilísticos (muestreo polifásico, muestreo de paneles, muestreo por áreas, muestreo de selección controlada, muestreo simulado, etc.); se debe tener en cuenta que todos los métodos probabilísticos tienen como base el MAS. La tabla 1.4 presenta un resumen de los métodos básicos de muestreo que serán objeto de estudio en los capítulos siguientes. En las ciencias sociales, el muestreo se ha desarrollado de la mano de los sondeos políticos, lo cual se debe sin duda a que estos son una de las pocas posibilidades que tienen los científicos sociales de descubrir la exactitud de sus cálculos; el día de las elecciones averiguan cuán bien o mal lo hicieron (Babbie, 1999).

Tabla 1.4. **Los métodos de muestreo y sus características**

	Todos los individuos 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)
Procedimiento metodológico	1. Determinar un marco de muestreo adecuado. 2. Asignar un número a cada unidad poblacional de 1 a N. 3. Mediante una tabla de números aleatorios o un procedimiento equivalente seleccionar las unidades que van a constituir la muestra de trabajo.
Observaciones	• Cada elemento se elige en forma independiente de los demás. • Garantiza igual probabilidad de inclusión en la muestra para cualquier individuo. • 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.
	Necesariamente no todas las combinaciones posibles de unidades elementales tienen igual probabilidad
	Sistemático (MSIS)
Procedimiento metodológico	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.
Observaciones	• 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 al primer individuo seleccionado, pueden verse afectadas el resto de unidades 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 ordenada según la variable observada.

Continúa

	Todos los individuos o elementos de la población objeto de estudio tienen una probabilidad diferente de cero para pertenecer a la muestra
	Necesariamente no todas las combinaciones posibles de unidades elementales tienen igual probabilidad
	Estratificado (MAE)
Procedimiento metodológico	1. Determinar un marco muestral adecuado para cada estrato. 2. Dividir el universo o población objeto de estudio en grupos (estratos) internamente homogéneos y entre ellos heterogéneos. 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.
Observaciones	• 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 ella será más económica. • 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 característica analizada) 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.
	Todos los individuos o elementos de la población objeto de estudio tienen una probabilidad diferente de cero para pertenecer a la muestra
	Necesariamente no todas las combinaciones posibles de unidades elementales tienen igual probabilidad.
	Conglomerados (MCON)
Procedimiento metodológico	1. Determinar marco muestral adecuado. 2. Dividir el universo en grupos (clúster) de igual o diferente tamaño de unidades elementales. 3. Los clúster deben ser altamente heterogéneos. 4. Se selecciona una muestra de clúster basada en el marco. 5. Dentro de cada clúster puede realizarse censo o a su vez seleccionar una muestra de ellos.*

Continúa