



**TEC.AZUAY**  
INSTITUTO UNIVERSITARIO

# FUNDAMENTOS DE ESTADÍSTICA

UNIDAD 1:  
CONCEPTOS BÁSICOS

# CONCEPTOS BÁSICOS

## Introducción:

Este capítulo cubre aspectos básicos relacionados con los conceptos fundamentales de Estadística. En la primera sección se revisa la división de la Estadística. Posteriormente, se estudia la población y muestra. Finalmente, se realiza problemas propuestas de nuestro.

# CONCEPTOS BÁSICOS

## ¿Qué es la estadística?

La estadística es la ciencia que trata de la recolección, organización, presentación, análisis e interpretación de datos numéricos con el fin de realizar una forma de decisión más efectiva.

## TIPOS DE ESTADÍSTICA

La Estadística se divide en dos ramas, las mismas que se encuentran interrelacionadas dado que la una es la base y fuente para la otra.

- **ESTADÍSTICA DESCRIPTIVA:** es aquella que hace referencia al conjunto de métodos para organizar, resumir y presentar los datos de manera informativa; es como una fotografía de una realidad determinada.

Ejemplo: Si un profesor desea saber las calificaciones de sus alumnos, puede optar por usar la estadística deductiva o descriptiva, y calcular las calificaciones haciendo un redondeo del grupo completo.

- **ESTADÍSTICA INFERENCIAL:** corresponde al conjunto de métodos utilizados para saber algo acerca de una población, basándose en una muestra; es decir, sus métodos permiten como su nombre lo dice hacer inferencia de una población a partir de los datos y resultados obtenidos de una muestra.

Ejemplo: Si el profesor de igual forma desea saber las calificaciones mas altas de sus estudiantes, en este caso puede optar la estadística inferencial, al tomar todas las calificaciones y seleccionar las mas altas.

# CONCEPTOS BÁSICOS

- **POBLACIÓN:** Es el conjunto de las unidades de análisis (sujetos u objetos) que se van a investigar.
- **PARÁMETRO:** Son los valores numéricos que corresponden a las características de la población.
- **MUESTRA:** es una parte de la población, la misma que debe cumplir con la característica de representatividad, es decir debe reunir en lo posible todas las características de la población.

Las ventajas de utilizar una muestra es que se puede inferir los resultados de la población ahorrando tiempo y recursos; siempre y cuando la muestra sea representativa de la población.

## Ejemplo:

**Población:** Profesores de Secundaria de la Provincia del Azuay.

**Muestra:** Profesores de Bachillerato de la Ciudad de Cuenca perteneciente a la provincia del Azuay

# CONCEPTOS BÁSICOS

- **Tamaño muestral:** Es el número de elementos u observaciones que se denota por  $n$  ó  $N$ ; es la sumatoria de las frecuencias absolutas.
- **Dato:** Cada uno de los individuos, entes abstractos, cosas, sujetos, que integran una población o universo determinado; dicho de otra manera, cada valor observado de la variable, según la presentación de los mismos en una encuesta.

**Encuesta:** Se denomina encuesta al conjunto de preguntas especialmente diseñadas y pensadas para ser dirigidas a una muestra de población o población total, que se considera por determinadas circunstancias funcionales al trabajo, representativa de esa población, con el objetivo de conocer la opinión de las personas sobre determinadas cuestiones.

# CONCEPTOS BÁSICOS

- **VARIABLE:** Es una característica cualitativa o cuantitativa, que puede tomar diferentes valores para cada uno de los elementos de la población.

Las Variables se clasifican en:

- **VARIABLES CUANTITATIVAS:** Son aquellas cuyas características representan una cantidad, es decir se expresan mediante números. A su vez las variables cuantitativas se clasifican en discretas y continuas:

Las **variables discretas** son aquellas que representan números enteros, ya que su naturaleza no permita que se subdivide, generalmente son variables que se cuentan por ejemplo: Cantidad de hijos, número de mascotas, etc.

Las **variables continuas** en cambio son aquellas que pueden subdividirse o encontrarse entre dos valores fijos; por ejemplo: peso, distancia, tiempo, etc., hacen referencia a variables que pueden medirse.

- **VARIABLES CUALITATIVAS:** son aquellas variables que representan características o cualidades que no pueden ser medidas con números, se escriben literalmente.

Estas variables también pueden ser de dos tipos: Nominales u ordinales

Las **variables nominales:** son aquellas cuyas características son cualidades no numéricas; por ejemplo: sabores, comidas típicas, entre otras

Las **variables ordinales:** son aquellas que representan un orden determinado; por ejemplo: tallas de ropa (S, L, M), orden de llegada (primero, segundo, tercero); grado de satisfacción (muy satisfecho, satisfecho, poco satisfecho), etc.

# Población y Muestra

La población es el conjunto o colección de objetos al que está referido un estudio estadístico. El vocablo suena a personas, pero una población estadística puede estar constituida por cualquier tipo de elemento, es decir, una población puede estar constituida por personas, pero también por objetos de cualquier tipo de naturaleza. Por ejemplo,

En un estudio sobre la incidencia de cierta enfermedad en un país, la población sería todos los habitantes de dicho país.

En un estudio sobre la calidad de la producción de ciertos componentes para ordenadores, la población serían todos los componentes que se han fabricado.



La Muestra es cualquier subconjunto de una población. Cuando los elementos que componen la muestra están elegidos aleatoriamente y todos los elementos tienen la misma probabilidad de ser elegidos diremos que se trata de una muestra aleatoria simple. Por norma general, en un estudio estadístico hay muchos condicionantes de tipo económico, físico, o de otro tipo que impiden trabajar con todos los elementos de la población, por tanto, se suele recurrir a muestras representativas de la población. Los procedimientos a seguir para la elección de este tipo de muestras se estudian en una parte de la Estadística Inferencial que se denomina muestreo.



# MUESTREO

El muestreo es un proceso o conjunto de métodos para obtener una muestra finita de una población finita o infinita, con el fin de estimar valores de parámetros o corroborar hipótesis sobre la forma de una distribución de probabilidades o sobre el valor de un parámetro de una o más poblaciones.

El muestreo además de ser una ciencia Estadística, es un arte, donde no solo los elementos se seleccionan al azar con una medida de probabilidad, si no que, además, requiere pericia por parte de investigador/a en el diseño de la muestra a la hora de determinar:

- La técnica de selección de elementos de investigación.
- La selección de los estimadores apropiados.
- La elección de un tamaño adecuado de la muestra con precisión (margen de error) y un nivel de confianza aceptable.
- Y el uso de marcos muestrales actualizados.



# TIPOS DE MUESTREO


## ➤ Muestreo probabilístico.



Los métodos de muestreo probabilísticos son aquellos que se basan en el principio de equiprobabilidad. Es decir, aquellos en los que todos los individuos tienen la misma probabilidad de ser elegidos para formar parte de una muestra y, consiguientemente, todas las posibles muestras de tamaño  $n$  tienen la misma probabilidad de ser seleccionadas. Sólo estos métodos de muestreo probabilísticos nos aseguran la representatividad de la muestra extraída y son, por tanto, los más recomendables.

Dentro de los métodos de muestreo probabilísticos encontramos los siguientes tipos:


- **a) Muestreo aleatorio simple:** El procedimiento empleado es el siguiente: 1) se asigna un número a cada individuo de la población y 2) a través de algún medio mecánico (bolas dentro de una bolsa, tablas de números aleatorios, números aleatorios generados con una calculadora u ordenador, etc.) se eligen tantos sujetos como sea necesario para completar el tamaño de muestra requerido. Este procedimiento, atractivo por su simpleza, tiene poca o nula utilidad práctica cuando la población que estamos manejando es muy grande.
- **b) Muestreo aleatorio sistemático:** Este procedimiento exige, como el anterior, numerar todos los elementos de la población, pero en lugar de extraer  $n$  números aleatorios sólo se extrae uno. Se parte de ese número aleatorio  $i$ , que es un número elegido al azar, y los elementos que integran la muestra son los que ocupa los lugares  $i, i+k, i+2k, i+3k, \dots, i+(n-1)k$ , es decir se toman los individuos de  $k$  en  $k$ , siendo  $k$  el resultado de dividir el tamaño de la población entre el tamaño de la muestra:  $k = N/n$ . El número  $i$  que empleamos como punto de partida será un número al azar entre 1 y  $k$ . El riesgo este tipo de muestreo está en los casos en que se dan periodicidades en la población ya que al elegir a los miembros de la muestra con una periodicidad constante ( $k$ ) podemos introducir una homogeneidad que no se da en la población. Imaginemos que estamos seleccionando una muestra sobre listas de 10 individuos en los que los 5 primeros son varones y las 5 últimas mujeres, si empleamos un muestreo aleatorio sistemático con  $k=10$  siempre seleccionaríamos o sólo hombres o sólo mujeres, no podría haber una representación de los dos sexos.



**c) Muestreo aleatorio estratificado:** Trata de obviar las dificultades que presentan los anteriores ya que simplifican los procesos y suelen reducir el error muestral para un tamaño dado de la muestra. Consiste en considerar categorías típicas diferentes entre sí (estratos) que poseen gran homogeneidad respecto a alguna característica (se puede estratificar, por ejemplo, según la profesión, el municipio de residencia, el sexo, el estado civil, etc.). Lo que se pretende con este tipo de muestreo es asegurarse de que todos los estratos de interés estarán representados adecuadamente en la muestra. Cada estrato funciona independientemente, pudiendo aplicarse dentro de ellos el muestreo aleatorio simple o el estratificado para elegir los elementos concretos que formarán parte de la muestra. En ocasiones las dificultades que plantean son demasiado grandes, pues exige un conocimiento detallado de la población. (Tamaño geográfico, sexos, edades,...). La distribución de la muestra en función de los diferentes estratos se denomina afijación, y puede ser de diferentes tipos:

- Afijación Simple: A cada estrato le corresponde igual número de elementos muestrales.
- Afijación Proporcional: La distribución se hace de acuerdo con el peso (tamaño) de la población en cada estrato.
- Afijación Óptima: Se tiene en cuenta la previsible dispersión de los resultados, de modo que se considera la proporción y la desviación típica. Tiene poca aplicación ya que no se suele conocer la desviación.

**d) Muestreo aleatorio por conglomerados:** Los métodos presentados hasta ahora están pensados para seleccionar directamente los elementos de la población, es decir, que las unidades muestrales son los elementos de la población. En el muestreo por conglomerados la unidad muestral es un grupo de elementos de la población que forman una unidad, a la que llamamos conglomerado. Las unidades hospitalarias, los departamentos universitarios, una caja de determinado producto, etc., son conglomerados naturales. En otras ocasiones se pueden utilizar conglomerados no naturales como, por ejemplo, las urnas electorales. Cuando los conglomerados son áreas geográficas suele hablarse de "muestreo por áreas". El muestreo por conglomerados consiste en seleccionar aleatoriamente un cierto número de conglomerados (el necesario para alcanzar el tamaño muestral establecido) y en investigar después todos los elementos pertenecientes a los conglomerados elegidos.





# MUESTREO NO PROBABILISTICO



# Definición

- Es aquel utilizado en forma empírica, es decir, no se efectúa bajo normas probabilísticas de selección, por lo que sus procesos intervienen opiniones y criterios personales del investigador o no existe norma validada.



# Muestreo sin requerimientos formales

- Es comúnmente utilizado en **observaciones de tipo exploratoria** o de tipo cualitativo.
- La **selección de la muestra es no aleatoria**. Se basa en el juicio de quien realiza el proceso o del responsable de la investigación.
- Los métodos de muestreo no aleatorio **no garantizan la representatividad de la muestra** y por lo tanto **no permiten los procesos de generalización**.



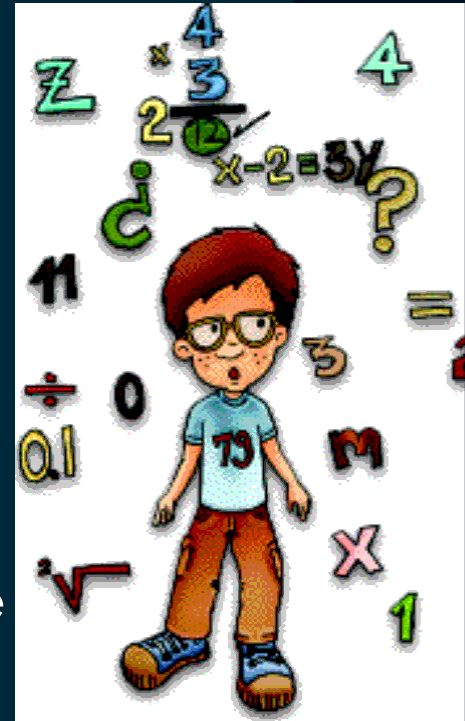
➤ Se caracteriza por no conocerse la probabilidad de que una unidad quede incluida en una muestra, no se puede medir el error de estimación y por lo tanto, no se pueden realizar estimaciones.



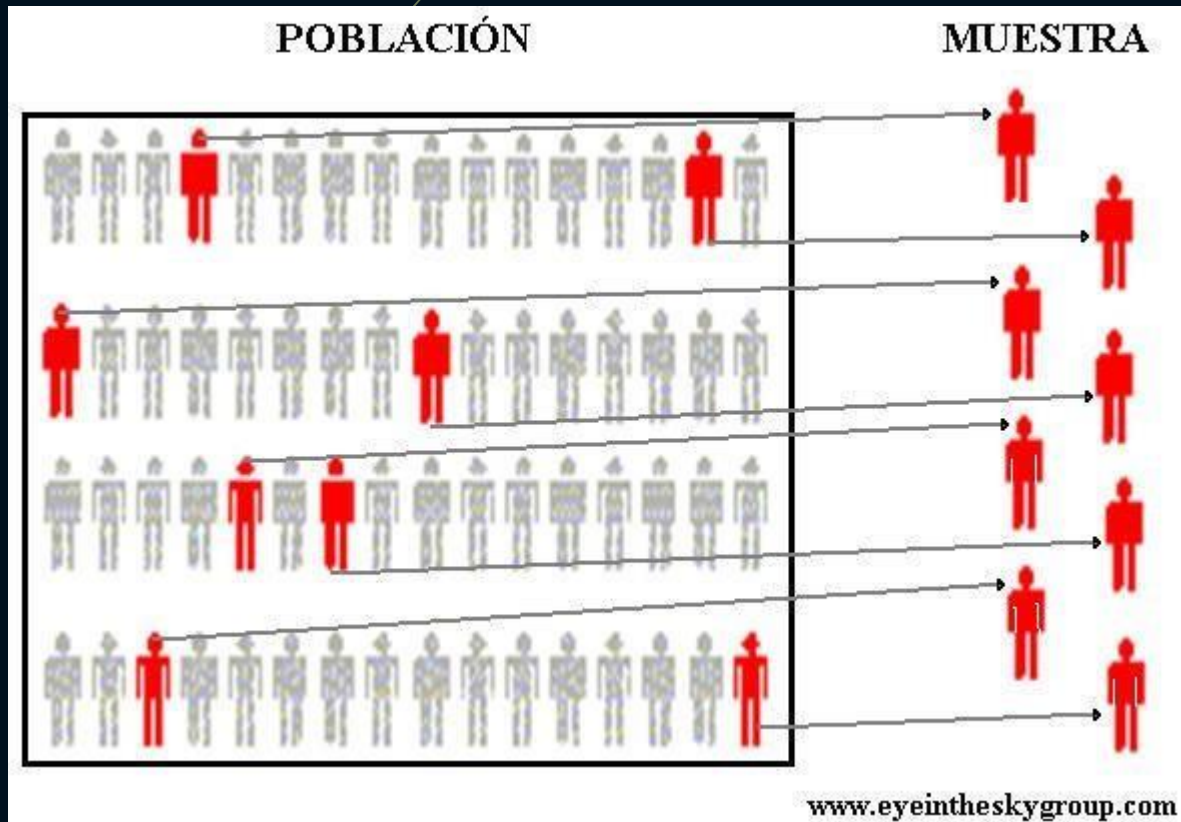


# Características comunes de este tipo de muestreo

- ✓ **No se basa en la teoría de la probabilidad;** por lo tanto, no es posible calcular la Exactitud.
- ✓ **No es posible el cálculo de la confianza** de las estimaciones, la precisión de la muestra resultante.
- ✓ **El investigador elige** los casos que más le interesan con el propósito de lograr **información más nutrida.**
- ✓ Es posible obtener **buenos resultados** siempre y cuando **el investigador tenga un conocimiento claro y preciso** de la población que **quiere observar** y de lo que en ella quiere observar.
- ✓ **Las inferencias** alcanzan planos **lógicos** y hasta de **opinión.**
- ✓ Son los más apropiados para las **investigaciones cualitativas**, ya que en este tipo de investigaciones no se busca la representatividad de los resultados, sino el punto de vista.



# Cuando aplicar el muestreo no probabilístico



- ✓ Cuando se pretende estudiar una **población rara** o marginal
- ✓ Cuando **no hay un marco disponible** para propósitos de muestreo
- ✓ Cuando se considera que **no se requieren cifras exactas sobre la representatividad estadística** de los resultados

# Como determinar el tamaño de la muestra



- ✓ En la práctica se obtiene **mediante el juicio y el sentido común** de los investigadores, que se basan en la experiencia.
- ✓ Para determinar el tamaño de la muestra en estos casos, los investigadores se basan en factores como: presupuesto disponible, reglas empíricas y número de subgrupos que se van a analizar.

# Tipos de muestreo no probabilístico

✓ MUESTREO DE CONVENIENCIA.

✓ MUESTREO POR JUICIO U OPINION.

✓ MUESTREO POR CUOTAS.

✓ MUESTREO DE BOLA DE NIEVE.



# Muestreo por conveniencia

- Pretende seleccionar unidades de análisis que cumplen los requisitos de la población objeto de estudio, pero que sin embargo, no son seleccionadas al azar. Se utiliza preferentemente en estudios exploratorios, as pruebas pilotos, también usan con frecuencia éste tipo de muestreo.



# Muestreo por conveniencia

- Se le conoce como selección intencionada, muestreo accidental o muestreo por oportunidad.
- Consiste en la elección por métodos no aleatorios de una muestra cuyas características sean similares a las de la población objetivo.
- En este tipo de muestreo la “representatividad” la determina el investigador de modo subjetivo, siendo este el mayor inconveniente del método ya que no podemos cuantificar la representatividad de la muestra.

# Ventajas

- ▶ Menos costoso
- ▶ No requiere mucho tiempo
- ▶ Fácil de administrar
- ▶ Por lo general asegura alta tasa de participación
- ▶ Posible generalización a sujetos similares



# Desventajas

- ▶ Difícil generalizar a otros sujetos .
- ▶ Menos representativa de una población específica
- ▶ Los resultados dependen de las características únicas de la muestra.
- ▶ Mayor probabilidad de error debido al investigador o influencia de sujetos (sesgos)





# Muestreo por juicio u opinión

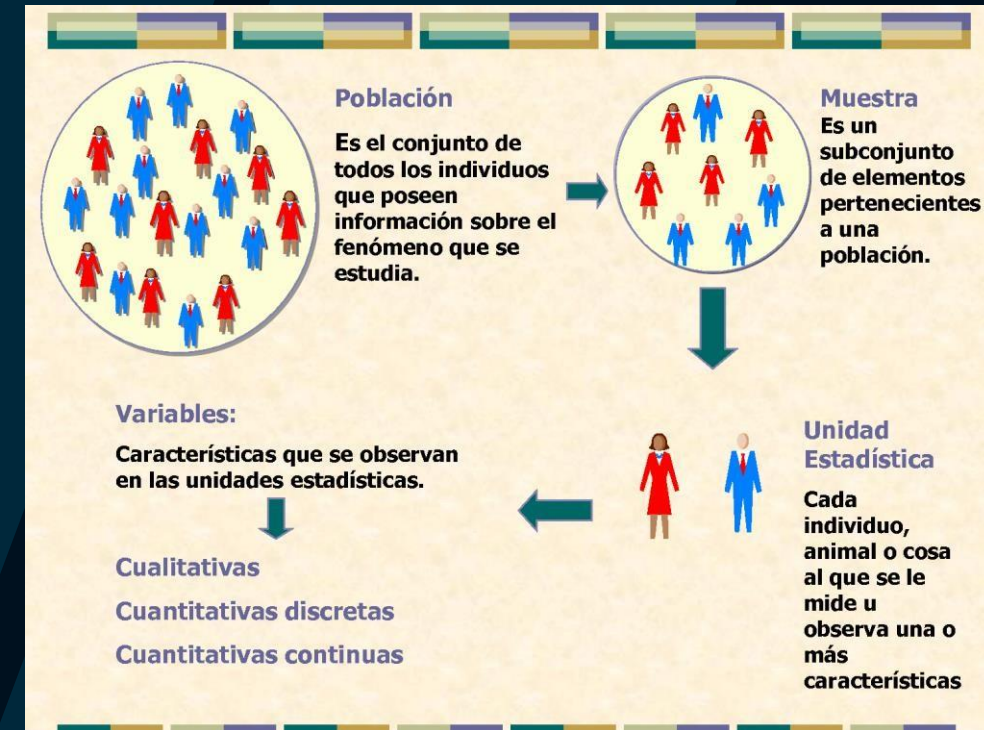
- La muestra se elige sobre la base de conocimientos que el investigador tenga de la población, sus elementos y la naturaleza de los objetivos de la investigación.
- El asunto crítico es la objetividad. Que tanto se puede confiar en el juicio del investigador al seleccionar una muestra.
- Aplica bien para estudios de pre-prueba o prueba piloto para un instrumento. La idea se centra en que el investigador elige la muestra por que los considera los más representativos.
- Se caracteriza este tipo de muestreo por un esfuerzo deliberado de obtener muestras representativas mediante la inclusión en la muestra de sujetos supuestamente típicos.





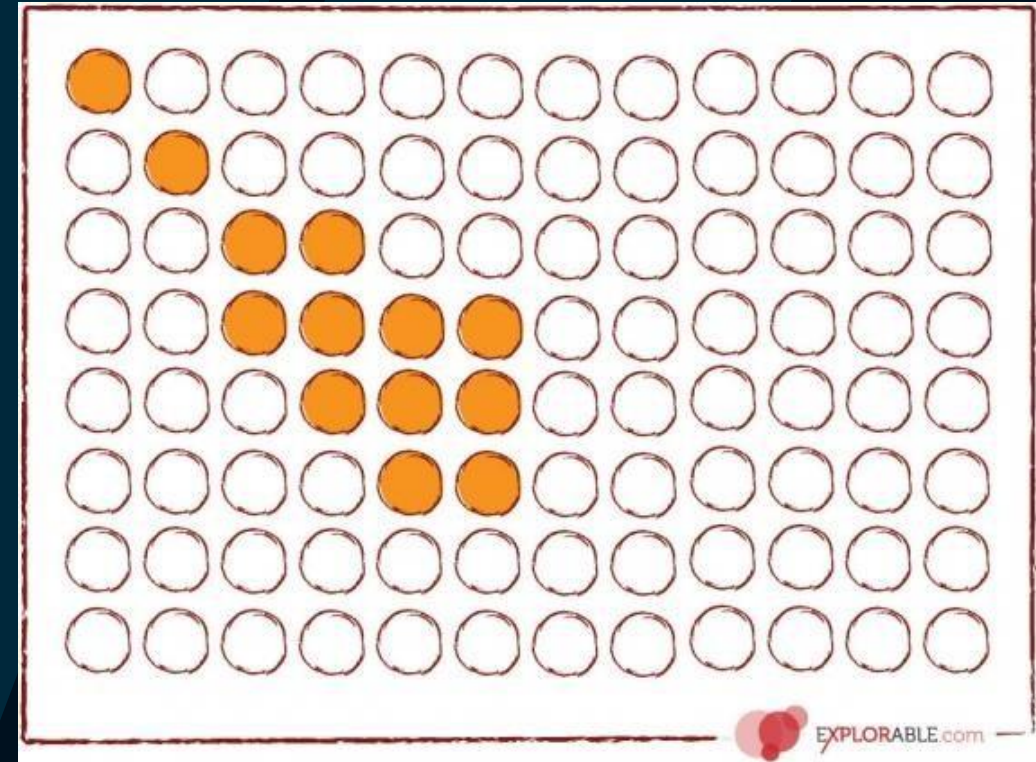
# Muestreo por cuotas

- Es similar al muestreo aleatorio estratificado pero no tiene el carácter de aleatoriedad.
- La población se fracciona en subpoblaciones y de cada población se fija una cuota, que consisten en un número de individuos, que reúnen unas determinadas condiciones.
- Una vez determinada la cuota se eligen los primeros que se encuentren (siendo mejor los más representativos o adecuados) que cumplan con tales condiciones.
- La asignación de las cuotas puede ser proporcional o igual, de la misma manera que en muestreo aleatorio estratificado.
- Es un método muy utilizado en encuestas de opinión.



# Muestreo por bola de nieve

- Apropiado cuando es difícil localizar a los miembros de una población. (Desamparados, Trabajadores foráneos, Indocumentados)
- Para llevarlo a cabo se reúnen los datos de los pocos miembros de la población objetivo que se puedan localizar y se les pide información necesaria para ubicar a otros miembros que conozcan de esa población.
- La bola de nieve se refiere a la acumulación que resulta de que cada sujeto localizado proponga a otros.



# Cuándo utilizar el muestreo no probabilístico

- Este tipo de muestreo puede ser utilizado cuando se quiere mostrar que existe un rasgo determinado en la población.
- También se puede utilizar cuando el investigador tiene como objetivo hacer un estudio cualitativo, piloto o exploratorio.
- Se puede utilizar cuando es imposible la aleatorización, como cuando la población es casi ilimitada.
- Se puede utilizar cuando la investigación no tiene como objetivo generar resultados que se utilicen para hacer generalizaciones respecto de toda la población.
- También es útil cuando el investigador tiene un presupuesto, tiempo y mano de obra limitados.
- Esta técnica también se puede utilizar en un estudio inicial que será llevado a cabo nuevamente utilizando un muestreo probabilístico aleatorio.

# Conclusión

- ▶ A diferencia del muestreo probabilístico, la muestra no probabilística no es un producto de un proceso de selección aleatoria. Los sujetos en una muestra no probabilística generalmente son seleccionados en función de su accesibilidad o a criterio personal e intencional del investigador.
- ▶ La desventaja del método de muestreo no probabilístico es que no se toman pruebas de una porción desconocida de la población. Esto implica que la muestra puede representar a toda la población con precisión o no. Por lo tanto, los resultados de la investigación no pueden ser utilizados en generalizaciones respecto de toda la población.

# DETERMINACIÓN DE UNA MUESTRA

Para determinar la muestra es necesario considerar primero cual es nuestro universo. Se entiende por universo al total de elementos que reúnen ciertas características homogéneas, los cuales son objeto de una investigación.

## Estimar una proporción:

Si deseamos estimar una proporción, debemos saber:

- El nivel de confianza (1- $\alpha$ ). El nivel de confianza prefijado da lugar a un coeficiente ( $Z_\alpha$ ). Para una confianza del 95% = 1.96, para una confianza del 99% = 2.58.
- La precisión que deseamos para nuestro estudio.
- Una idea del valor aproximado del parámetro que queremos medir (en este caso una proporción). En caso de no tener dicha información utilizaremos el valor  $p = 0.5$  (50%).

### Población Finita

$$n = \frac{N * Z_\alpha^2 * p * q}{e^2 * (N - 1) + Z_\alpha^2 * p * q}$$

### Población Infinita

$$n = \frac{Z_\alpha^2 * p * q}{e^2}$$



# EJEMPLOS

¿A cuántas personas tendríamos que estudiar para conocer la preferencia de sistema operativo que utiliza?

Confianza = 95%; Precisión = 3%: Proporción esperada = asumamos que puede ser próxima al 5%; si no tuviésemos ninguna idea de dicha proporción utilizaríamos el valor  $p = 0.5$  (50%) que maximiza el tamaño muestral:

$$n = \frac{Z_{\alpha}^2 * p * q}{e^2}$$

$$n = \frac{1.96^2 \times 0.05 \times 0.95}{0.03^2}$$

Si la población es finita, es decir conocemos el total de la población y deseásemos saber cuántos del total tendremos que estudiar la respuesta sería:

$$n = \frac{N * Z_{\alpha}^2 * p * q}{e^2 * (N - 1) + Z_{\alpha}^2 * p * q}$$

¿A cuántas personas tendría que estudiar de una población de 15.000 habitantes para conocer la preferencia de la marca de una computadora portátil?

Seguridad = 95%; Precisión = 3%; proporción esperada = asumamos que puede ser próxima al 5%; si no tuviese ninguna idea de dicha proporción utilizaríamos el valor  $p = 0.5$  (50%) que maximiza el tamaño muestral.

$$n = \frac{15000 \times 1.96^2 \times 0.05 \times 0.95}{0.03^2 \times (15000 - 1) + 1.96^2 \times 0.05 \times 0.95} = 200$$



# MUCHAS GRACIAS

---

Ing. Héctor  
Mejía V, Mgtr.

Correo electrónico  
[hector.mejia@tecazuay.edu.ec](mailto:hector.mejia@tecazuay.edu.ec)