Muestreo Estadístico-SEMANA-1

Raúl Alberto Pérez
raperez1@unal.edu.co
Profesor Asociado - Escuela de Estadística
Universidad Nacional de Colombia, Sede Medellín

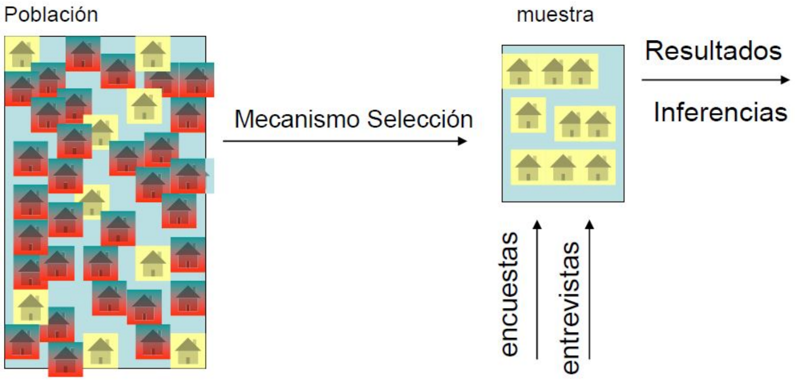
Semestre 2021-I

Introducción al Muestreo

Una necesidad de mucha importancia en las empresas u organismos que desarrollan de una u otra forma actividades al servicio de la comunidad es la información acerca de características poblacionales, sea cual sea el campo de acción de la empresa.

Debido a la imposibilidad en la mayoría de los casos de obtener información con base en la medición de todas las unidades que componen la población (ie. mediante un censo) nos lleva a la selección de una parte de dicha población (ie. seleccionar una muestra), usando distintos métodos apropiados que nos permiten obtener tamaños de muestra representativos para poder hacer inferencias acerca de dichas características poblacionales de interés.

Graficamente



¿Porqué seleccionar una muestra y no toda la población?

Existen situaciones en las cuales es conveniente seleccionar una muestra en lugar de llevar a cabo un censo, como en los siguientes casos:

1. La población es grande y su estudio completo excede los recursos asignados .
2. Existe suficiente homogeneidad en las unidades poblacionales con respecto a lo que se quiere medir, lo cual permite que una muestra adecuada contenga la información de interés necesaria para el estudio.
3. El proceso de selección de la muestra es destructivo (por ejemplo en estudios de control de calidad o en estudios de poblaciones biológicas), lo cual obliga al análisis de sólo una parte de la población.

En 1 y 3 se optimizan recursos y en 2 se evita el despilfarro de recursos.

Adicional a las tres razones mencionadas anteriormente para hacer uso del muestreo estadístico, existe otra razón que tiene que ver con el estricto control que se puede ejercer sobre la recolección de la información, el cual es superior al que se lograría en el caso de la realización de un censo.

Los errores en el estudio, especialmente los conocidos como errores no muestrales (la no-respuesta, las inconsistencias, la codificación errada, entre otros) disminuyen considerablemente cuando se realiza muestreo en lugar de un censo.

Algunos Conceptos Básicos

- Universo o Población: Por universo se entiende el conjunto total de elementos (o individuos) bajo estudio o sobre los cuales se desea extender los resultados obtenidos en la muestra. También se conoce como *población objetivo*.

La definición de la *población objetivo* es una parte importante y a veces difícil de los estudios por muestreo.

- Población de Muestreo: No siempre es posible tener acceso a todos los elementos del universo, teniendo que trabajar entonces con el universo accesible o comúnmente llamado Población de Muestreo, es decir, con el conjunto de elementos susceptibles de ser medidos.

En algunos casos ambas poblaciones objetivo (universo) y la población de muestreo son equivalentes y se habla simplemente de población.

- Población Estadística: Una población estadística es un conjunto de mediciones que se hace sobre todos los elementos de un universo.

Lo anterior implica que el estudio estadístico de un universo puede dar lugar a diferentes poblaciones estadísticas o a lo que comúnmente se conoce como poblaciones multivariadas.

- Unidad Elemental: Una unidad elemental es cada uno de los elementos del universo que se desean medir, por ejemplo: una persona, un vehículo, una planta, etc.
- Unidad de Observación: Las unidades de observación son las unidades que proporcionan la información (ie. las unidades que son medidas).

Procedimiento para la realización de un estudio por **Muestreo**

Definición del Problema a Estudiar

Identificar si se relaciona con

- Estimación de una media μ
- Estimación de un total Poblacional τ
- Estimación de una proporción p y el total de individuos con un atributo, A.

Definir claramente:

- Elemento
- Población
- Unidad de Muestreo
- Marco

Definir la metodología de muestreo a ser utilizada

Muestreo Aleatorio Simple(MAS)

Muestreo Sistemático

Muestreo Aleatorio Estratificado (MAE)

Muestreo por conglomerados

Muestreo en Varias Etapas

Seleccionar la muestra

Hacer las estimaciones y calcular los respectivos Intervalos de Confianza

Interesa hacer inferencia acerca de:

1. **Media Poblacional** μ: Promedio de la variable de interés en toda la población.
2. **Total Poblacional** τ: suma de los valores de la variable de interés para todas las observaciones en la población.
3. **Proporción Poblacional** P: Proporción de individuos u objetos en la población que poseen la característica de interés.
4. **Total Poblacional de unidades que tienen un atributo** A: Número total de individuos en la población que cumplen con un atributo.

Investigación o Estudio por Muestreo

Existen cuatro pasos fundamentales en el diseño de una investigación por muestreo, los cuales son: el diseño muestral, las mediciones a tomar, el trabajo de campo y el análisis estadístico a realizar .

- El diseño Muestral: El diseño muestral incluye tanto el plan de muestreo como los procedimientos de estimación.

El plan de muestreo es la metodología utilizada para seleccionar la muestra de la población.

Los procedimientos de estimación son los algoritmos o fórmulas usadas para obtener las estimaciones de los valores poblacionales y su confiabilidad a partir de los datos muestrales.

La selección de un diseño muestral particular debe tener en cuenta qué variables van a medirse, qué estimaciones se requieren, qué niveles de confiabilidad y validez se necesitan y qué restricciones se tienen en cuanto a recursos.

- Las mediciones a tomar : Esta parte generalmente es responsabilidad de expertos en la materia objeto de estudio cuando la investigación se lleva a cabo a través de encuestas, donde su aporte es muy valioso en el diseño de los cuestionarios para obtener la información adecuada.

Un instrumento de recolección de información mal diseñado conduce frecuentemente a un fracaso del estudio, al igual que una mala preparación de los encuestadores.

- Trabajo de Campo: El trabajo de campo comienza una vez los formularios a utilizar han sido elaborados, ensayados y modificados.

En muchos casos es conveniente tomar una muestra piloto pequeña para probar los instrumentos de medición y eliminar los defectos observables del proceso a desarrollar.

La información se debe recolectar de acuerdo al diseño muestral escogido para la investigación.

- Análisis Estadístico: Una vez se posee la información se debe procesar adecuadamente para que puede ser analizada estadísticamente.

Se debe recordar que se están " estimando " características de la población y por lo tanto los resultados están sujetos a errores tanto de muestreo como de medición.

Variables y Parámetros Poblacionales

El objetivo más importante de las investigaciones por muestreo tiene que ver con la estimación de ciertos valores de la distribución de una variable en la población.

Las variables son definidas una vez se conocen las características de interés a estudiar.

Las características que puedan ser medidas cuantitativamente son consideradas como *variables discretas* (número de miembros por hogar, número de automóviles por familia, número de cursos por semestre, etc.) o como *variables continuas* (peso en kg, estatura en cm, ingreso por familia en pesos,rendimiento por hectáreas en lbs, etc.).

Las características que simplemente clasifican o caracterizan sectores de la población se conocen como *atributos o características cualitativas* *(estado civil, religión, género, situación laboral, etc.) y generalmente se tratan con el uso de variables multinomiales.*

- *Parámetros: Los valores que identifican la distribución de una variable en una población estadística se denominan parámetros.*

En forma general los parámetros son representados por la letra: θ *y en casos específicos tendrán su propia representación.*

- *Tamaño Poblacional: El tamaño de la población (población de muestreo) se define como el número total de unidades elementales accesibles y generalmente se denota por la letra* N.

- Variables: Las variables o características a medir se denotan por las letras: $X, Y, Z, \ldots$ y los valores de dichas variables se denotan por: $\quad x, y, z, \ldots$

El valor de la variable Y para la i-ésima unidad elemental, $i = 1, 2, \ldots, N$ se denota por: $\quad y_i$.

Algunos Parámetros Poblacionales

La descripción básica de una población estadística está dada por sus parámetros principales, los cuales de alguna manera están representando las características de la población. Algunos de estos parámetros son los siguientes:

1. El total poblacional (τ): El total poblacional representa el *tamaño absoluto total de la variable que se está estudiando.*
2. *La Media Poblacional (μ): La media es una medida de tendencia central de todos los valores.*
3. *La Varianza o Cuasivarianza (σ^2): La varianza mide la dispersión o variabilidad de los datos.*
4. *La Desviación Estándar (σ): La desviación estándar también mide la dispersión de los datos, pero es de mayor utilización debido a que está dada en las misma unidades de medida de la variable.*

5. El coeficiente de Variación (CV): Es una medida de *Dispersión relativa o de homogeneidad de los valores.*

Es de gran utilidad para comparar grupos de datos correspondientes a diferentes variables.

6. Cuando se estudian simultáneamente dos poblaciones se tienen *la covarianza (σ_{xy}) y el coeficiente de correlación lineal (ρ_{xy}).*
7. *La proporción poblacional. Cuando la característica a medir es la presencia o no presencia de un atributo, la media poblacional μ se convierte en una Proporción Poblacional (p), que representa la proporción de elementos en la población que poseen el atributo considerado.*

Si el atributo puede clasificarse en k-categorías distintas, se hará uso de diferentes proporciones poblacionales asociadas a cada una de las k-categorías: $p_1, p_2, \ldots, p_k$.

Definición de Algunos Parámetros en Variables Cuantitativas

- Total Poblacional: El total poblacional es la suma de todos los valores de la variable en la población y está dado por:

$$\tau = \sum_{i=1}^{N} y_i$$

- Media Poblacional: La media poblacional es la media aritmética de todos los valores de la variable en la población y está dada por:

$$\mu = \frac{\sum_{i=1}^{N} y_i}{N} = \frac{\tau}{N}$$

- Varianza Poblacional: La Varianza poblacional es el promedio de las desviaciones al cuadrado de todos los valores con respecto a la media y está dada por:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N-1}$$

- Desviación Estándar Poblacional: La desviación estándar poblacional es la raíz cuadrada de la varianza:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N-1}}$$

$B = R_0))$.

$(n = ??)$ t.q $f(.., \hat{\sigma} --)$

- Coeficiente de Variación Poblacional: El coeficiente de variación poblacional es el cociente entre la desviación estándar y la media:

$$CV = \frac{\sigma}{\mu}$$

Este coeficiente se expresa generalmente en forma de porcentaje y se puede considerar como una medida de la homogeneidad de los datos.

Para valores del CV menores al 15% se considera que los datos son homogéneos. Para valores entre 15 y 30% se considera una homogeneidad moderada y para valores superiores al 30% se tiene una homogeneidad significativa o severa.

- Varianza Relativa: La varianza relativa es igual al cuadrado del CV.

$$V_{rel} = (CV)^2 = \frac{\sigma^2}{\mu^2}$$

La varianza relativa se utilizará en la determinación del tamaño de muestra a seleccionar de una población.

- Covarianza: La covarianza entre dos variables X e Y correspondiente a los mismos elementos de una población se define por:

$$\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N-1}$$

- Coeficiente de Correlación: El Coeficiente de Correlación entre dos variables X e Y correspondiente a los mismos elementos de una población se define por:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Se cumple que: $|\rho_{xy}| \leq 1$. Valores altos (negativos o positivos) indican una alta dependencia lineal entre las variables. Valores cercanos a cero indican una débil dependencia lineal.

Nota: Si dos variables son independientes su coeficiente de correlación es cero.

Sin embargo, un coeficiente de correlación igual a cero, no implica necesariamente que las dos variables sean independientes.

Definición de Algunos Parámetros en Variables Cualitativas

En el caso de que la característica de interés sea un atributo, los parámetros de interés se pueden derivar de las mismas fórmulas presentadas anteriormente, definiendo los valores de la variable Y como sigue:

$$y_i = \begin{cases} 1 \text{ si el atributo está presente en la } j\text{-ésima unidad} \\ 0 \quad \text{si no lo está} \end{cases}$$

Usando está variable se tienen las siguientes definiciones.

- Total Poblacional: El número total de elementos en la población que poseen el atributo considerado es:

$$A = \sum_{i=1}^{N} y_i \; . \qquad (= \tau)$$

- Proporción Poblacional: La proporción poblacional de elementos con el atributo considerado es:

$$P = \frac{\sum_{i=1}^{N} y_i}{N} = \frac{A}{N}$$

$\frac{\Sigma}{N}$ $(= \mu)$

- Varianza Poblacional: La Varianza poblacional se define como:

$$\sigma^2 = \frac{N}{N-1} PQ$$

$\Sigma(y_i - P)^2$

$= \Sigma(y_i^2 - 2Py_i + P^2)$

$\cdots = \frac{N}{N-1} PQ$

donde $Q = 1 - P$, es la proporción de elementos de la población que no poseen el atributo considerado.

$Q = (1-P)$

- Desviación Estándar Poblacional: La desviación estándar poblacional es la raíz cuadrada de la varianza:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{N}{N-1}}\sqrt{PQ}$$

- Coeficiente Variación Poblacional: El coeficiente de variación poblacional:

$$CV = \frac{\sigma}{P} = \sqrt{\frac{N}{N-1}}\sqrt{\frac{Q}{P}}$$

  Este coeficiente de variación para variables cualitativas no puede considerarse siempre como una medida de la homogeneidad de los datos.

Cuando $P \leq 0.5$ el CV toma valores mayores o iguales al 100%, lo cual es difícil de interpretar para homogeneidad.

Este CV se utilizará para la determinación de tamaños muestrales.

Muestra, Marco Muestral y Estadísticos

La muestra seleccionada para estimar los parámetros de la población puede ser una muestra *probabilística* o una muestra *no-probabilística*.

En las muestras probabilísticas cada elemento de la población tiene una probabilidad conocida, diferente de cero, de formar parte de la muestra.

Lo anterior permite obtener, a partir de dicha muestra, estimaciones válidas desde el punto de vista estadístico-matemático, lo cual no es posible en muestras no-probabilísticas.

En el muestreo *no-probabilístico* puede haber influencia de la persona o personas que seleccionan la muestra o simplemente se realiza atendiendo a razones de comodidad.

Por ejemplo, si hacemos una encuesta telefónica por la mañana, las personas que no tienen teléfono o que están trabajando, no podrán formar parte de la muestra.

Marco Muestral

El marco muestral es cualquier material o mecanismo (listas, mapas, registros, etc.) que permita delimitar o identificar en forma apropiada los elementos de la población de donde se selecionará la muestra. Por ejemplo: Listado de todas las manzanas, de todos los intervalos de tiempo, de todas las parcelas, etc.

Algunas propiedades que debe cumplir un marco muestral son:

1. Las unidades en el marco deben identificarse a través de algún código, de 1 hasta N, donde N-es el número de unidades de muestreo.
2. Cada unidad debe tener factibilidad de ser hallada si es seleccionada en la muestra.
3. El marco debe estar organizado de una forma sistemática (por ejemplo las unidades ordenadas en forma geográfica o por tamaño).

4. Cuando se requiere estimación para *dominios* (o subpoblaciones), el marco debe especificar el dominio al cual pertenece cada unidad.
5. Todo elemento de la población de interés debe estar presente solamente una vez en el marco.
6. Ningún elemento ajeno a la población de interés debe estar presente en el marco.
7. Todo elemento de la población de interés debe aparecer en el marco.

En investigaciones de gran escala, como por ejemplo en las Encuestas Nacionales de Hogares, el diseño muestral, es un diseño muestral multietápico, donde se requieren diferentes *marcos muestrales* para cada una de las etapas establecidas, por ejemplo: *selección de municipios del país, selección de sectores dentro de los municipios, selección de manzanas dentro de los sectores, y selección de hogares dentro de las manzanas*.

En este caso, las unidades muestrales reciben el nombre de unidades primarias, secundarias, etc. dependiendo de la etapa respectiva en la que pueden ser seleccionadas.

Las unidades finales de muestreo (las que proporcionan la información en la encuesta) comprenden generalmente más de una unidad elemental, por ejemplo, un hogar está formado por más de una persona.

Estadísticos y Estimadores

Un *Estadístico* es una función de los *valores muestrales*, por lo que a su vez es una *variable aleatoria* ya que su valor cambia de muestra a muestra con los resultados del proceso de selección aleatoria.

Los estadísticos pueden ser calculados con fines meramente descriptivos o para estimar los parámetros poblacionales.

Cuando se utilizan para estimar los parámetros poblacionales reciben el nombre de *Estimadores*.

Los estimadores se designaran por: $\hat{\theta}$.

Al valor que toma un estimador en una muestra específica se le denomina *estimación*.

Los principales *Estadísticos (o Estimadores)* usados frecuentemente para la estimación de *parámetros* son los siguientes:

- Total muestral: El total muestral es la suma de todos los valores de la variable Y en la muestra:

  $$\hat{\tau} = \sum_{i=1}^{n} y_i$$

  donde n-es el tamaño de la muestra.

- Media Muestral: La media muestral es la media aritmética de todos los valores de la variable Y en la muestra:

  $$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\hat{\tau}}{n}$$

- Varianza Muestral: La Varianza muestral es el promedio de las desviaciones al cuadrado de todos los valores en la muestra con respecto a la media muestral:

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}$$

- Desviación Estándar Muestral: La desviación estándar muestral es la raíz cuadrada de la varianza muestral:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}}$$

- Coeficiente de Variación Muestral: El coeficiente de variación muestral es el cociente entre la desviación estándar y la media muestral:

$$cv = \frac{s}{\overline{y}}$$

Para el caso de variables cualitativas dicotónicas se definen como sigue, con una definición apropiada de la variable Y:

- Total Muestral: El número total de elementos en la muestra que poseen el atributo considerado es:

$$a = \sum_{i=1}^{n} y_i$$

- Proporción Muestral: La proporción muestral de elementos en la muestra con el atributo considerado es:

$$p = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{a}{n}$$

Varianza Muestral: La Varianza Muestral se define como:

$$s^2 = \frac{n}{n-1} pq$$

donde $q = 1 - p$, es la proporción de elementos de la muestra que no poseen el atributo considerado.

Desviación Estándar Muestral: La desviación estándar muestral es la raíz cuadrada de la varianza:

$$s = \sqrt{s^2} = \sqrt{\frac{n}{n-1}} \sqrt{pq}$$

Coeficiente Variación Muestral: El coeficiente de varaición muestral se expresa como:

$$cv = \frac{s}{p} = \sqrt{\frac{n}{n-1}} \sqrt{\frac{q}{p}}$$

Distribuciones Muestrales

A la distribución de todas las estimaciones de un parámetro basadas en **todas las muestras posibles** que pueden ser generadas por el plan muestral particular se le llama Distribución Muestral del Estimador.

Algunas características de la distribución de un estimador se definen de manera similar a las características de una población finita.

- Media de la Distribución de un Estimador: La media de la distribución de un estimador $\hat{\theta}$, se define por:

$$E[\hat{\theta}] = \sum_{i=1}^{v} \hat{\theta}_i \pi_i$$

  donde: v: número total de valores distintos tomados por el estimador; $\hat{\theta}_i$: i-ésima estimación diferente del parámetro $\hat{\theta}$ y π_i: es la probabilidad de que el estimador tome el valor $\hat{\theta}_i$. La probabilidad π_i-es igual a la frecuencia relativa de las estimaciones por ser una variable discreta.

- Varianza de la Distribución de un Estimador: La varianza de la distribución de un estimador $\hat{\theta}$, se define por:

$$Var[\hat{\theta}] = \sum_{i=1}^{v} (\hat{\theta}_i - E[\hat{\theta}])^2 \pi_i.$$

- Desviación Estándar de la Distribución de un Estimador: La desviación estándar de la distribución de un estimador $\hat{\theta}$, se denomina frecuentemente *Error Estándar de la Estimación* y se define por:

$$E.E[\hat{\theta}] = \sqrt{Var[\hat{\theta}]}$$

- Coeficiente de Variación de la Distribución de un Estimador: El coeficiente de Variación de la distribución de un estimador $\hat{\theta}$, se define por:

$$C.V[\hat{\theta}] = \frac{E.E[\hat{\theta}]}{E[\hat{\theta}]}$$

- Error Cuadrático Medio de un Estimador: El error cuadrático medio de un estimador $\hat{\theta}$ es una medida de dispersión con respecto al verdadero parámetro poblacional y se define por:

$$ECM[\hat{\theta}] = E(\hat{\theta} - \theta)^2 = \sum_{i=1}^{v} (\hat{\theta}_i - \theta)^2 \pi_i$$

si el estimador es insesgado, ie. si $E[\hat{\theta}] = \theta$, entonces:

$$ECM[\hat{\theta}] = E(\hat{\theta} - \theta)^2] = E(\hat{\theta} - E[\hat{\theta})^2] = Var[\hat{\theta}]$$

- Sesgo de un Estimador: El Sesgo de un estimador $\hat{\theta}$, denotado por $B[\hat{\theta}]$, se define como la diferencia entre la media de la distribución muestral y el valor verdadero del parámetro desconocido, es decir:

$$B[\hat{\theta}] = E[\hat{\theta}] - \theta$$

- Estimador Insesgado: Un estimador $\hat{\theta}$ se dice que es *Insesgado* si su valor esperado es igual al parámetro, es decir:

$$E[\hat{\theta}] = \theta \quad \longrightarrow \quad B[\hat{\theta}] = 0.$$

Se puede verificar la siguiente relación entre el ECM, la varianza y el sesgo de un estimador:

$$ECM[\hat{\theta}] = Var[\hat{\theta}] + B^2[\hat{\theta}]$$

- Precisión de un Estimador: La precisión de un estimador $\hat{\theta}$ se refiere a qué tan lejos se encuentra un valor particular estimado del verdadero valor del parámetro y se acostumbra medir haciendo uso del ECM del estimador o de su raíz cuadrada.

Entre menor sea el ECM mayor es la precisión de la estimación.

Si el estimador es insesgado, la precisión sería la varianza=ECM o el error-estándar, ie. la raíz cuadrada de la varianza.

- Covarianza entre dos Estimadores: La covarianza entre dos estimadores distintos $\hat{\theta}_1$ y $\hat{\theta}_2$ de un parámetro $\hat{\theta}$, con medias $E[\hat{\theta}_1]$ y $E[\hat{\theta}_2]$ respectivamente, se define como:

$$Cov[\hat{\theta}_1, \hat{\theta}_2] = E[(\hat{\theta}_1 - E[\hat{\theta}_1])(\hat{\theta}_2 - E[\hat{\theta}_2])] = E[\hat{\theta}_1\hat{\theta}_2] - E[\hat{\theta}_1]E[\hat{\theta}_2]$$

- Coeficiente de Correlación entre dos Estimadores:

El coeficiente de correlación entre dos estimadores distintos $\hat{\theta}_1$ y $\hat{\theta}_2$ de un parámetro $\hat{\theta}$, con errores estándar dados por $EE[\hat{\theta}_1]$ y $EE[\hat{\theta}_2]$ respectivamente, se define como:

$$\rho_{\hat{\theta}_1,\hat{\theta}_2} = \frac{Cov[\hat{\theta}_1, \hat{\theta}_2]}{EE[\hat{\theta}_1]EE[\hat{\theta}_2]}$$

Eficiencia Relativa de un Estimador

Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores distintos de un parámetro $\hat{\theta}$, con varianzas dadas por $Var[\hat{\theta}_1]$ y $Var[\hat{\theta}_2]$ respectivamente.

Si $Var[\hat{\theta}_2] > 0$, entonces se define la *Eficiencia Relativa* (EFR) de $\hat{\theta}_1$ con respecto a $\hat{\theta}_2$ como sigue:

$$EFR[\hat{\theta}_1, \hat{\theta}_2] = \frac{Var[\hat{\theta}_1]}{Var[\hat{\theta}_2]}$$

Según sea la $EFR[\hat{\theta}_1, \hat{\theta}_2]$ inferior, igual o superior a la unidad, se dira que $\hat{\theta}_1$ es más, igual o menos eficiente que $\hat{\theta}_2$.

EJEMPLO: DISTRIBUIÓN MUSTRAL DE $\hat{\theta} = \bar{Y}$

Suponga que su población está conformada por cuatro, 4, **Unidades de Muestreo**, cuyos valores son conocidos:

$$\{1, 4, 7, 10\}.$$

Y se seleccionan muestras de tamaño **dos**, 2.

La distribución muestral de $\hat{\theta} = \bar{Y}$ para todas las muestras posibles de tamaño **dos**:

muestra		$\bar{y}$	$P[\bar{Y} = \bar{y}]$
1	$\{1, 4\}$	2.5	$\frac{1}{6}$
2	$\{1, 7\}$	4.0	$\frac{1}{6}$
3	$\{1, 10\}$	5.5	$\frac{1}{6}$
4	$\{4, 7\}$	5.5	$\frac{1}{6}$
5	$\{4, 10\}$	7.0	$\frac{1}{6}$
6	$\{7, 10\}$	8.5	$\frac{1}{6}$

DISTRIBUIÓN MUSTRAL DE $\hat{\theta} = \bar{Y}$

La distribuión de mustral de $\bar{y}$ s:

$\bar{Y}_i$	2.5	4.0	5.5	5.5	7.0	8.5
$P[\bar{Y}_i]$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\begin{aligned}
E[\bar{Y}] &= \sum_{i=1}^{6} \bar{y}_i P\left[\bar{Y} = \bar{y}_i\right] \\
&= 2.5\frac{1}{6} + 4\frac{1}{6} + 5.5\frac{1}{6} + 5.5\frac{1}{6} + 7\frac{1}{6} + 8.5\frac{1}{6} \\
&= \frac{1}{6}\left\{2.5 + 4.0 + 5.5 + 5.5 + 7.0 + 8.5\right\} = \frac{11}{2}
\end{aligned}$$

DISTRIBUIÓN MUSTRAL DE $\hat{\theta} = \bar{Y}$

Por otro lado se halla la Varianza de $\bar{Y}$:

$$
\begin{aligned}
V(\bar{Y}) &= \sum_{i=1}^{6}\left(\bar{y}_i - \frac{11}{2}\right)^2 * \frac{1}{6} \\
&= \frac{1}{6}\left[\left(\frac{5}{2} - \frac{11}{2}\right)^2 + \left(4.0 - \frac{11}{2}\right)^2 + \left(5.5 - \frac{11}{2}\right)^2 + \left(5.5 - \frac{11}{2}\right)^2\right. \\
&= \left. + \left(7 - \frac{11}{2}\right)^2 + \left(8.5 - \frac{11}{2}\right)^2\right] \\
&= \frac{1}{6}\left[9 + \frac{9}{4} + \frac{9}{4} + 9\right] = \frac{1}{6}\left[18 + \frac{18}{4}\right] = 3 + \frac{3}{4} \\
&= \frac{15}{4}
\end{aligned}
$$

$$E.E[\bar{Y}] = \sqrt{V(\bar{Y})} = \sqrt{\frac{15}{4}}$$

Muestreo Aleatorio de Poblaciones Finitas

En el muestreo de poblaciones finitas se asume que el número total de elementos en la población es N, de los cuales se seleccionan aleatoriamente n.

Selección aleatoria no es sinónimo de selección al azar.

El término aleatorio implica el usa de un mecanismo de probabilidad bien diseñado en la selección de la muestra. El uso de dicho mecanismo hace distinguir entre las muestras *probabilísticas* y las muestras *subjetivas , a juicio o tipo* las cuales pueden ser algunas veces muestras al azar.

Aunque toda muestra aleatoria es una muestra al azar, lo contrario no siempre se cumple.

Se consideraran únicamente muestras aleatorias (o probabilísticas) las cuales son las que permiten el uso de la teoría estadística.

Muestreo Probabilístico

Algunas condiciones que deben cumplirse para poder hablar de una *muestra probabilística* son las siguientes:

1. Poder definir el conjunto total de muestras posibles $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_T\}$ que pueden seleccionarse de la población de acuerdo con el procedimiento muestral.

2. Conocer para cada una de las muestras posibles la probabilidad $\pi(\mathcal{S})$ de que sea seleccionada.

3. El procedimiento de selección utilizado debe dar a cada elemento de la población una probabilidad de selección distinta de cero.

4. La selección debe ser aleatoria, es decir, el mecanismo de probabilidad diseñado para la selección debe ser tal que cada muestra posible $\mathcal{S}$ tenga la probabilidad de selección asignada previamente, $\pi(\mathcal{S})$.

Muestreo No-Probabilístico

Todo tipo de muestreo que no cumpla con alguna de las condiciones mencionadas anteriormente es un muestreo no-probabilístico.

Algunos ejemplos típicos de muestreo no-probabilistico son los siguientes:

1. La muestra ha sido restringida a la parte de la población que es fácilmente accesible.
2. La muestra se selecciona teniendo en cuenta el azar mas no la aleatoriedad (seleccionar los elementos que estén a la mano).
3. Con una población heterogénea y pequeña, el muestrista inspecciona la población en general y selecciona una muestra pequeña de unidades "tipo" (ie. unidades que a su parecer están cercanas al promedio de la poblacional).
4. La muestra está compuesta esencialmente de voluntarios (cuando el proceso de medición es desagradable o le puede representar problemas a la persona medida).