



# **Universidad Autónoma del Estado de México**

## **Facultad de Economía**



### **Lic. En Actuaría**

## **Unidad de Aprendizaje: Muestreo**

### **Fundamentos de Muestreo:**

**Muestreo Aleatorio Simple (MAS)**

**Muestreo Sistemático Aleatorio (MSA)**

**M. En C. Rafael Morales Ibarra**

**Febrero 2019**

# Contenido

Tema	Plantilla
<b>Guion explicativo</b>	3
Programa por competencia	4
<b>Muestreo</b>	
Introducción	7
Conceptos	10
Tipos de muestreo	18
Cálculo del tamaño	24
Tipos de errores de muestreo	25
Estimación del error	29
Calculo del tamaño de muestra “n”	31
<b>1. Muestreo Sistemático Aleatorio (MSI)</b>	32
MAS sin reemplazo	33
Estimación del total poblacional	43
Estimación del tamaño de muestra	44
Estimación de una proporción poblacional (p)	56
Calculo del tamaño de muestra “n”	65
<b>2. Muestreo Sistemático Aleatorio (MSA)</b>	75
Ejemplos	79
MSA replicado	84

# Guión explicativo de uso del material

En un curso tradicional de inferencia estadística se estudia cómo realizar estimaciones de los parámetros poblacionales, en función de la información que suministra una muestra, sin embargo hasta ahora, dicha información es limitada, por ello el objetivo de las presentes notas, es el proporcionar la base teórica al estudiante para poder tener un manejo adecuado de los métodos de muestreo más usuales en las disciplinas económico administrativas y en licenciatura de Actuaría y Economía.

Particularmente nos interesa estudiar, cómo debe hacerse la selección de los elementos u observaciones de una muestra. Así como el proceder en esa selección y la definición correcta del tamaño adecuado cuando la población es muy grande, cuando es una población finita y acotada o bien cuando los fenómenos carezcan de un marco muestral definido.

De igual manera el muestreo cobra importancia al ser una herramienta fundamental en la investigación económica y empresarial, dado que permite conocer las características de poblaciones concretas, contrastar hipótesis teóricas sobre ellas o realizar predicciones sobre su evolución. En este proceso un componente de gran ayuda para el investigador, es la planeación correcta de las encuestas.

La primera parte de este material hace un paréntesis para considerar y puntualizar algunos de los elementos y conceptos básicos en el estudio del muestreo, para posteriormente hacer una compilación de los métodos probabilísticos de muestreo, particularmente en este apartado el muestreo aleatorio simple, sistemático, y posteriormente en una segunda sección de plantillas el muestreo estratificado y por conglomerados.

El alumno encontrará las referencias bibliográficas que le permitan abundar con una mayor precisión en la teoría y en ejemplos de aplicación de dichas métodos de muestreo.

# Programa por competencias de la UA: Muestreo


 Universidad Autónoma del Estado de México  
 Secretaría de Docencia  
 Coordinación General de Estudios Superiores  
 Programa Institucional de Innovación Curricular

Programa de Estudios por Competencias

**I. IDENTIFICACIÓN DE LA UNIDAD DE APRENDIZAJE.**

ORGANISMO ACADÉMICO		FACULTAD DE ECONOMÍA				
Programa Educativo: Licenciatura en Actuaría		Área de docencia: Matemáticas y Estadísticas				
Aprobación por los R.H. Consejos Académico y de Gobierno	Fecha: 1 de febrero de 2007	Programa elaborado por: Juan José Lachuga Arizón, Ricardo Rodríguez Marcial, Jesús Salgado Vega, Daniel Havel Cartas				Fech 10 d
Nombre de la Unidad de aprendizaje: MUESTREO						
Clave	Horas de teoría	Horas de práctica	Total de horas	Créditos	Tipo de Unidad de Aprendizaje	Carácter de la Unidad de Aprendizaje
L4225	4	2	6	10	Curso	Obligatorio
Prerrequisitos (Conocimientos Previos): Inferencia Estadística				Unidad de Antecedente	Aprendizaje	Unidad de d Cálculo Act
Programas educativos en los que se imparte: Licenciatura en Actuaría						


 Universidad Autónoma del Estado de México  
 Secretaría de Docencia  
 Coordinación General de Estudios Superiores  
 Programa Institucional de Innovación Curricular

**II. PRESENTACIÓN**

La unidad de aprendizaje busca capacitar al alumno en el estudio de muestreo

**III. LINEAMIENTOS DE LA UNIDAD DE APRENDIZAJE**

DOCENTE	DISCENTE
El profesor deberá cubrir la totalidad de los temas del curso y además debe tener conocimiento de todos los temas, deberá asistir a la totalidad de las clases	El alumno resolverá problemas fundamentales

**IV. PROPÓSITO DE LA UNIDAD DE APRENDIZAJE**

Tanto el profesor como el alumno están comprometidos a llevar un curso según los lineamientos que determine la institución. El alumno deberá comprender los conceptos y utilizar la investigación de operaciones determinística para su formación como

**V. COMPETENCIAS GENÉRICAS**

Aplicar el muestreo a problemas en el área de la actuaría

**VI. ÁMBITOS DE DESEMPEÑO PROFESIONAL**

Salón de clase  
Sala de cómputo


 Universidad Autónoma del Estado de México  
 Secretaría de Docencia  
 Coordinación General de Estudios Superiores  
 Programa Institucional de Innovación Curricular

**VII. ESCENARIOS DE APRENDIZAJE**

Sustantivo

**VIII. NATURALEZA DE LA COMPETENCIA**  
(Inicial, entrenamiento, complejidad creciente, ámbito diferenciado)

Competencia inicial

**IX. ESTRUCTURA DE LA UNIDAD DE APRENDIZAJE**

- Unidad 1. Planeación de la encuesta por muestreo
- Unidad 2. Diseño del cuestionario
- Unidad 3. Diseño de la muestra
- Unidad 4. Tamaño de la muestra
- Unidad 5. Muestreo aleatorio simple
- Unidad 6. Muestreo estratificado
- Unidad 7. Muestreo sistemático
- Unidad 8. Muestreo por conglomerados


 Universidad Autónoma del Estado de México  
 Secretaría de Docencia  
 Coordinación General de Estudios Superiores  
 Programa Institucional de Innovación Curricular

**X.- SECUENCIA DIDÁCTICA**

**UNIDAD 1. PLANEACIÓN DE LA ENCUESTA POR MUESTREO**

- 1.1 Puntos básicos de la planeación de la encuesta
- 1.2 Objetivos de la encuesta
- 1.3 Tamaño de la población
- 1.4 Sesgo en el diseño de la encuesta

**UNIDAD 2. DISEÑO DEL CUESTIONARIO**

- 2.1 Objetivos y limitaciones
- 2.2 Tipo de cuestionario a emplear
- 2.3 Formulación de la encuesta
- 2.4 Zona de estudio
- 2.5 Recolección de información
- 2.6 Prueba piloto
- 2.7 Errores en la aplicación
- 2.8 Corrección del cuestionario
- 2.9 Aplicación
- 2.10 Codificación



**UNIDAD 3. DISEÑO Y TAMAÑO DE LA MUESTRA**

- 3.1 Tamaño de la muestra
- 3.2 Error en la muestra
- 3.3 Diseño y tamaño de la muestra por proporciones
- 3.4 Diseño y tamaño de la muestra por medias

**UNIDAD 4. MUESTREO ALEATORIO SIMPLE**

- 4.1 Principios básicos
- 4.2 Definición de muestreo aleatorio simple
- 4.3 Diseño de muestra por muestreo aleatorio simple
- 4.4 Problemas

**UNIDAD 5. MUESTREO ESTRATIFICADO**

- 5.1 Principios básicos
- 5.2 Definición de muestreo estratificado
- 5.3 Diseño de muestra por muestreo estratificado
- 5.4 Problemas

**UNIDAD 6. MUESTREO SISTEMÁTICO**

- 6.1 Principios básicos
- 6.2 Definición de muestreo sistemático
- 6.3 Diseño de muestra por muestreo sistemático



**6.4 Problemas**

**UNIDAD 7. MUESTREO POR CONGLOMERADOS**

- 7.1 Principios básicos
- 7.2 Definición de muestra por muestreo por conglomeratos
- 7.3 Diseño de muestra por muestreo por conglomeratos
- 7.4 Problemas

**III. DESARROLLO DE LA UNIDAD DE APRENDIZAJE**

UNIDAD DE COMPETENCIA I	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes/ Valor
Unidad 3. Planeación de la encuesta por muestra	Conocer conceptos y antecedentes de planeación de la encuesta por muestra	Conocer definición de encuestas de operaciones	Trabajo en equipo - razonamiento matemático - responsabilidad en el trabajo
<b>ESTRATEGIAS DIDÁCTICAS:</b> Demostración con prácticas, elaboración y solución de problemas en clase y extra clase	<b>RECURSOS REQUERIDOS:</b> pizarrón - proyector de acetatos - Computadora	<b>TIEMPO DESTINADO:</b> 4 horas técnicas y 2 horas prácticas	
<b>CRITERIOS DE DESEMPEÑO I</b>	<b>DESEMPEÑO</b>		<b>EVIDENCIAS</b>
1.1. Puntos básicos de la planeación de la encuesta	Se refiere a que cada alumno conozca y comprenda cada definición de los datos		Conocimiento y manejo correcto de los datos



1.2. Objetivos de la encuesta	Se refiere a que cada alumno conozca y comprenda cada definición del tema	Conocimiento y manejo correcto de los datos
1.3. Tamaño de la población	Se refiere a que cada alumno conozca y comprenda cada definición del tema	Conocimiento y manejo correcto de los datos
1.4. Sesgo en el diseño de la encuesta	Se refiere a que cada alumno conozca y comprenda cada definición del tema	Conocimiento y manejo correcto de los datos

UNIDAD DE COMPETENCIA II	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes
UNIDAD II Diseño del cuestionario	Conocer conceptos básicos de diseño de cuestionario, recolección de datos, validación, formulación, prueba piloto, validación	Conocer definición de	Trabajo en equipo - razonamiento matemático - responsabilidad en el trabajo
<b>ESTRATEGIAS DIDÁCTICAS:</b> Demostración con prácticas, elaboración y solución de problemas en clase y extra clase.	<b>RECURSOS REQUERIDOS:</b> pizarrón - proyector de acetatos - Computadora	<b>TIEMPO DESTINADO:</b> 4 horas técnicas y 2 horas prácticas	
<b>CRITERIOS DE DESEMPEÑO II</b>	<b>DESEMPEÑO</b>		<b>EVIDENCIAS</b>
2.1. Objetivos Y Limitaciones	El alumno resolverá problemas del tema	Que el alumno aprenda tema	
2.2. Tipo De Cuestionario A Emplear	El alumno resolverá problemas del tema	Que el alumno aprenda tema	
2.3. Formulación De La Encuesta	El alumno resolverá problemas del tema	Que el alumno aprenda tema	
2.4. Zona De Estudio	El alumno resolverá problemas del tema	Que el alumno aprenda tema	



2.5. Recolección De Información	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2.6. Prueba Piloto	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2.7. Errores En La Aplicación	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2.8. Corrección Del Cuestionario	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2.9. Aplicación	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2.10. Codificación	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2.11. Integración Y Resultados Del Cuestionario	El alumno resolverá problemas del tema	Que el alumno aprenda tema

UNIDAD DE COMPETENCIA III	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes
UNIDAD III Diseño y tamaño de la muestra	Conocer conceptos básicos de diseño y tamaño de la muestra	Manejar resolver problemas de diseño y tamaño de la muestra	Trabajo en equipo - razonamiento matemático - responsabilidad en el trabajo
<b>ESTRATEGIAS DIDÁCTICAS:</b> Demostración con prácticas, elaboración y solución de problemas en clase y extra clase.	<b>RECURSOS REQUERIDOS:</b> pizarrón - proyector de acetatos - Computadora	<b>TIEMPO DESTINADO:</b> 4 horas técnicas y 2 horas prácticas	
<b>CRITERIOS DE DESEMPEÑO III</b>	<b>DESEMPEÑO</b>		<b>EVIDENCIAS</b>
3.1. Tamaño De La Muestra	El alumno resolverá ejemplos del tema	Conocimiento y manejo de estadísticas	
3.2. Error En La Muestra	El alumno resolverá ejemplos del tema	Conocimiento y manejo de estadísticas	



3.3. Diseño Y Tamaño De La Muestra (por Proporciones)	El alumno resolverá ejemplos del tema	Conocimiento y manejo de estadísticas
3.4. Diseño Y Tamaño De La Muestra Por Medias	El alumno resolverá ejemplos del tema	Conocimiento y manejo de estadísticas

UNIDAD DE COMPETENCIA IV	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes
UNIDAD IV Muestreo aleatorio simple	Conceptos básicos de muestreo aleatorio simple	Manejar y resolver problemas de muestreo aleatorio simple	Trabajo en equipo - razonamiento matemático - responsabilidad en el trabajo
<b>ESTRATEGIAS DIDÁCTICAS:</b> Demostración con prácticas, elaboración y solución de problemas en clase y extra clase.	<b>RECURSOS REQUERIDOS:</b> pizarrón - proyector de acetatos - Computadora	<b>TIEMPO DE STR:</b> 4 horas técnicas y 2 horas prácticas	
<b>CRITERIOS DE DESEMPEÑO IV</b>	<b>DESEMPEÑO</b>		<b>EVIDENCIAS</b>
4.1. Principios Básicos	El alumno resolverá ejercicios del tema	Conocimiento y manejo de estadísticas	
4.2. Definición De Muestreo Aleatorio Simple	El alumno resolverá ejercicios del tema	Conocimiento y manejo de estadísticas	
4.3. Diseño De Muestra Por Muestreo Aleatorio Simple	El alumno resolverá ejercicios del tema	Conocimiento y manejo de estadísticas	
4.4. Problemas	El alumno resolverá ejercicios del tema	Conocimiento y manejo de estadísticas	



UNIDAD DE COMPETENCIA V	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes
UNIDAD V Muestreo estratificado	Conceptos básicos de muestreo estratificado	Manejar y resolver problemas de muestreo estratificado	Trabajo en equipo - razonamiento matemático - responsabilidad en el trabajo
<b>ESTRATEGIAS DIDÁCTICAS:</b> Demostración con prácticas, elaboración y solución de problemas en clase y extra clase.	<b>RECURSOS REQUERIDOS:</b> pizarrón - proyector de acetatos - Computadora	<b>TIEMPO DESTINADO:</b> 4 horas técnicas y 2 horas prácticas	
<b>CRITERIOS DE DESEMPEÑO V</b>	<b>DESEMPEÑO</b>		<b>EVIDENCIAS</b>
5.1. Principios Básicos	El alumno resolverá ejercicios del tema	El alumno deberá	
5.2. Definición De Muestreo Estratificado	El alumno resolverá ejercicios del tema	El alumno deberá	
5.3. Diseño De Muestra Por Muestreo Estratificado	El alumno resolverá ejercicios del tema	El alumno deberá	
5.4. Problemas	El alumno resolverá ejercicios del tema	El alumno deberá	

UNIDAD DE COMPETENCIA VI	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes
UNIDAD VI Muestreo sistemático	Conceptos básicos de muestreo sistemático	Manejar y resolver problemas de muestreo sistemático	Trabajo en equipo - razonamiento matemático - responsabilidad en el trabajo
<b>ESTRATEGIAS DIDÁCTICAS:</b> Demostración con prácticas, elaboración y solución de problemas en clase y extra clase.	<b>RECURSOS REQUERIDOS:</b> pizarrón - proyector de acetatos - Computadora	<b>TIEMPO DESTINADO:</b> 4 horas técnicas y 2 horas prácticas	



CONTENIDOS DE DESEMPEÑO VI	EVIDENCIAS	
	DE DESEMPEÑO	
E.1 Principios Básicos	El alumno recibirá epígrafe de lema	El alumno
E.2 Definición De Muestras Sistemáticas	El alumno recibirá epígrafe de lema	El alumno
E.3 Diseño De Muestra Por Muestras Sistemáticas	El alumno recibirá epígrafe de lema	El alumno
E.4 Prácticas	El alumno recibirá epígrafe de lema	El alumno

UNIDAD DE COMPETENCIA VI	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes
UNIDAD VI. Muestras por conglomerados	Conceptos básicos de	Manejar y resolver problemas de	Tiempo resol resol
<b>ESTRATEGIAS DIDACTICAS</b> Demostración con práctica, elaboración y solución de problemas en clase y extra clase.	<b>RECURSOS REQUERIDOS</b> pizarra proyector de overhead Computadora		<b>TIEMPO</b> 4 horas + 2 horas +
CONTENIDOS DE DESEMPEÑO VI	EVIDENCIAS		
DE DESEMPEÑO			
V.1 Principios Básicos	El alumno recibirá epígrafe de lema	El alumno	
V.2 Definición De Muestras Por Conglomerados	El alumno recibirá epígrafe de lema	El alumno	
V.3 Diseño De Muestra Por Muestras Por Conglomerados	El alumno recibirá epígrafe de lema	El alumno	
V.4 Prácticas	El alumno recibirá epígrafe de lema	El alumno	



## XI. EVALUACIÓN Y ACREDITACIÓN

La evaluación para esta unidad de aprendizaje se cumplirá con el 85% de asistencia para poder presentar los exámenes de suficiencia, aprobándose con calificación mayor a seis puntos.

## XII. REFERENCIAS

1. Cochran W. "Técnicas De Muestras". Editorial Ceesa, 14 Edición, 1995, México, D. F.
2. Litinger, Ch. "La Encuesta Por Muestras: Teoría Y Práctica". Editorial Ceesa, 16a Edición, 1994, México, D. F.
3. Mandelkern, W. "Elementos De Muestras". Grupo Editorial Bessaméxico, 1987, México, D. F.
4. Das, Raj. "Teoría Del Muestreo". Fondo De Cultura Económica, 1988, México, D. F.
5. Lohr, S. "Muestreo: Diseño Y Análisis". International Thomson Editores, 2000, México, D. F.
6. Holguín, F. Hayashi, C. "Elementos De Muestras Y Correlación". Tercera Universidad, Segunda Edición, 1977, México.
7. Kuestenberg, O. "Análisis D E Planificación Urbana Matemática Y Muestreo". Editorial Limusa, Primera Reimpresión, 1994
8. Waples, R. "Probabilidad Y Estadística Para Ingenieros". Editorial Prentice Hall, Sexta Edición, 1995, México, D. F.

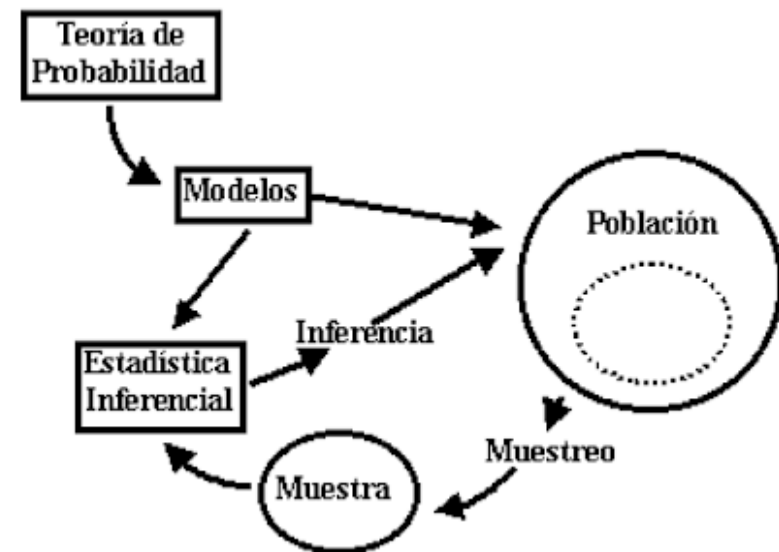
# M U E S T R E O

## 1. Introduction

El muestreo se emplea para aproximarse al conocimiento de la realidad, esto es habitual en la investigación científica.

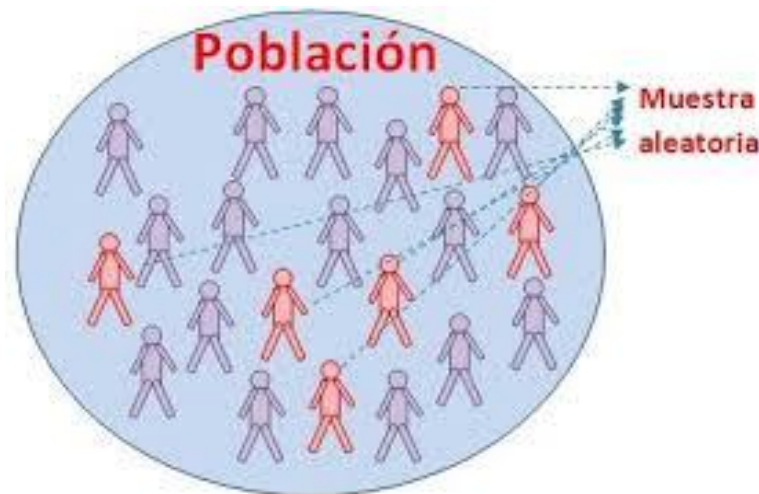


Para que un muestreo sea válido y confiable, debe estar fundamentado a los requerimientos que exige la teoría del muestreo, que a su vez se soporta de la teoría de probabilidades, y de la inferencia estadística.



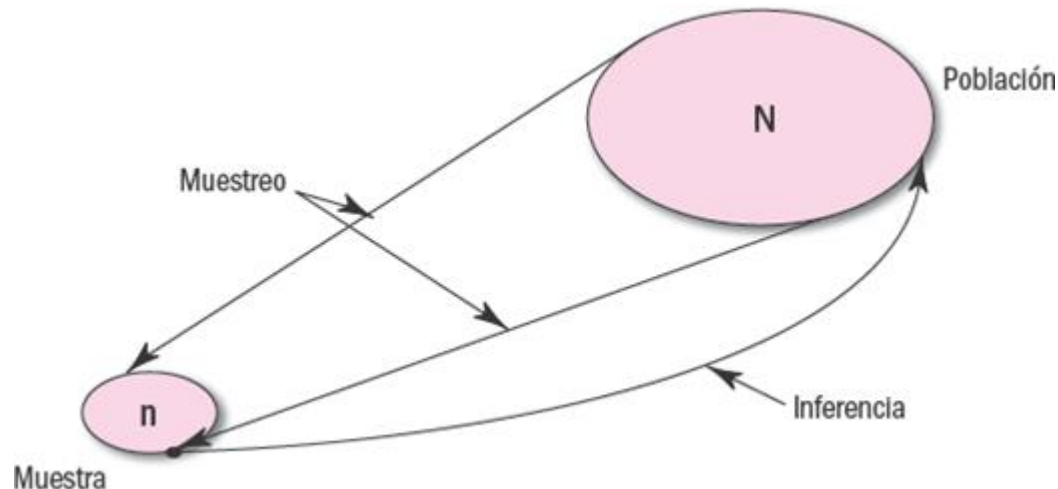
# Defición: muestreo

Procedimiento por el cual extrae, de una población un subconjunto de elementos (muestra) con el objetivo de generalizar los resultados hacia toda la población.



# Conceptos

- Población: Conjunto de cosas, elementos, personas que comparte una característica en común.  
Por ejemplo: Unidades: Personas, Familias, Viviendas, Escuelas, Organizaciones, Artículos de Prensa.
- Muestreo: conjunto de operaciones encaminadas a determinar una muestra, su tamaño y demás características necesarios para identificar a los elementos que la forman.



- Población de estudio: conjunto de elementos sobre los que se toma la muestra.
- Muestra: subconjunto de elementos de la población elegidos para estudiar y así tratar de inferir características de la población.
- Censo: relación completa de los elementos de una población.
- Sesgo: error específico de la muestra por falta de representatividad.

- Error muestral: es el imputable al estudio de una parte de la población o muestra.
- Error no muestral: es el que se produce en toda la investigación como consecuencia de definiciones conceptuales incorrectas, de fallos en los instrumentos de medida, en la entrevista o en el desarrollo del trabajo de campo.

Marco muestral: listado que identifica a los elementos de la población objetivo.

Elemento: cada una de las unidades sobre las que interesa obtener información.

Unidad muestral: unidad seleccionada de la población para la aplicación de la técnica de investigación; contiene los elementos de la población que pueden formar parte de la muestra.

Parámetro: medida de una característica determinada de una población. Si esta medida está referida a una muestra entonces se denomina estadístico. La diferencia entre ambos valores es el error muestral.

Estimador: es el valor muestral utilizado para inferir un valor poblacional.

Un estimador insesgado: es un estimador cuya esperanza matemática es el parámetro poblacional que estima. Se dice que un estimador es consistente si al sustituir el tamaño de la muestra por el del total de la población la estimación coincide con el parámetro poblacional.

Distribución muestral: Es la representación de los valores de los estadísticos (la media, por ejemplo) con sus frecuencias.

# ¿Porque se recurre al muestreo?

- i. Menor costo de costos ( tiempo, recurso humano, material y equipo)
- ii. Al disminuir el número de casos disminuyen también los errores asociados a la manipulación de los datos.
- iii. Se puede generalizar los resultados a la Población, si se ha tenido un rigor teorico para seleccionar la muestra.

# ¿Cómo calcular el tamaño de la muestra?

Factores que influyen, para el tamaño de “n”:

- Presupuesto: Tiempo y recursos disponibles
- Tipo de Muestreo
- Alcance del estudio
- Heterogeneidad de la población
- Margen de Error máximo admisible
- Nivel de confianza de la estimación muestral

# Factores que influyen en el cálculo del tamaño “n”

- El tipo de estudio influye en el tamaño de la muestra:
  - Comparación de Medias:  $(u_1-u_2)$
  - Estimación de Proporciones ( P )
  - Análisis Univariados:  $y=f(x)$
  - Análisis Multivariados:  $y_i=f(x_i)$

# Tipos de muestreos

## Probabilísticos

- Las muestras se seleccionan al azar, no se seleccionan por los investigadores.
- Cada elemento de la población tiene la misma probabilidad de ser elegido.
- Se puede conocer el error muestral, el nivel de confianza y el nivel de precisión de las estimaciones.
- Los resultados se pueden generalizar.
- Es el único método que puede evaluar la representatividad de la muestra.
- Es más caro que el muestreo no probabilística.
- Es, en general, más lento y complicado que el muestreo no probabilística.

# Tipos de Muestreo



# No probabilísticos

## Características

La selección de la muestra no es aleatoria, sino que se basa, en parte, en el juicio del entrevistador o de responsable de la investigación.

- No se basa en ninguna teoría de la probabilidad y, por lo tanto, no es posible calcular la precisión o acotar el error cometido.
- No es posible calcular estos errores ni la confianza de las estimaciones que, además, no siempre se reducen aumentando el tamaño de la muestra.
- En el muestreo no probabilístico los costes y la dificultad del diseño son más reducidos (al no ser necesario disponer de un marco).
- Este muestreo puede dar buenos resultados, pero también apareja el riesgo de proporcionar una información errónea.

❑ **Muestreo de conveniencia.-** Las muestras se seleccionan según un criterio de accesibilidad o comodidad.

- Suele emplearse en centros comerciales, plazas, estaciones de autobuses o de tren, metro, aeropuertos o lugares de gran afluencia pública.
- Se utiliza para obtener un mayor número de cuestionarios completos de forma rápida y económica.
- Recomendado en investigación exploratoria que venga seguida de una investigación adicional en la que se extraiga una muestra probabilística.

❑ **Muestreo de juicios.-** La muestra es elegida por un experto de acuerdo con su criterio, buscando las unidades más representativas.

- Ejemplo: mercados de prueba, presuntos líderes seleccionados en la investigación de conducta de votantes
- Se emplea cuando el tamaño de la muestra es pequeño.

❑ **Muestreo por cuotas** .-el objetivo es asegurar que los subgrupos de una población estén representados en la muestra.

- La selección accidental de sujetos puede producir sesgo.
- La velocidad de recopilación de datos, los menores costes y la comodidad son sus principales ventajas frente al probabilístico.
- Adecuado cuando el investigador sabe que es más probable que un cierto grupo demográfico rehúse colaborar con una encuesta.

❑ **Muestreo de “bola de nieve”**.- Los primeros elegidos como encuestados (probablemente a juicio del investigador) proponen y ayudan a la selección de los restantes de la muestra.

- Técnica empleada para localizar por referencias a miembros de poblaciones peculiares.
- Ventajas: los reducidos tamaños de muestra y los costes.
- Desventaja: Sesgo ya que la persona sugerida por otro miembro de la muestra tiene una probabilidad mayor de ser similar a la primera.

# ¿Cuándo utilizar muestreos probabilísticos y cuándo muestreos No probabilísticos?

## Muestreo Probabilísticos:

- Estimación de Parámetros
- Comprobación de Hipótesis

## Muestreos No Probabilísticos

- Estudios Pilotos
- Estudios Cualitativos
- Investigaciones en poblaciones de difícil registro o localización (marginales, prostitutas, enfermos de VIH, drogas, etc...)



# Cálculo del tamaño de la muestra

# Tipo de errores en muestreo

- Aleatorio
- Muestral
- Sistemático

$$E = Z \sqrt{\frac{P * Q}{n}}$$

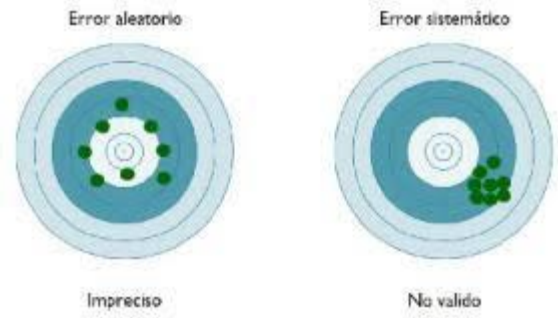
$$V = v \pm e$$

Valor Verdadero en la población

Valor Medido en la muestra

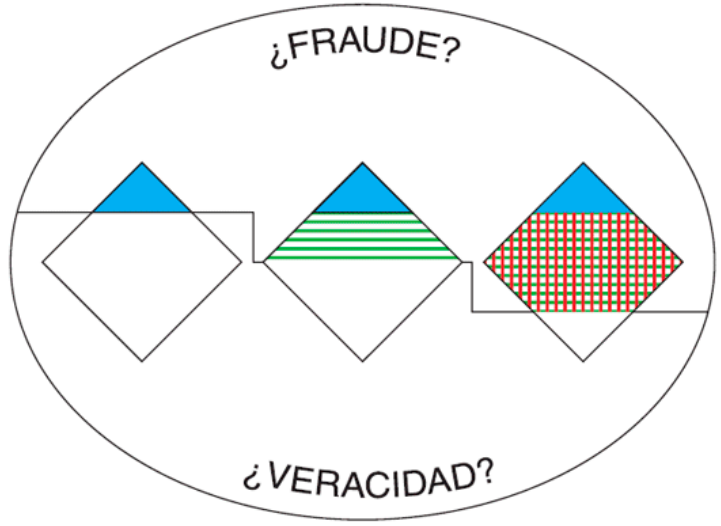
Error

- Error aleatorio:** Se produce como consecuencia de realizar el estudio con una muestra en lugar de estudiar a toda la población y generalizar los resultados que hemos obtenido en nuestra muestra a la población.



- Es un error que se puede cuantificar mediante el cálculo del intervalo de confianza y el nivel de seguridad con el que se da ese intervalo.

- Este tipo de error será de mayor si conforme la muestra sea mas pequeña y disminuirá al aumentar el tamaño muestral , llegando a desaparecer si estudiamos a toda la población..



□ Veracidad    ■ Error aleatorio    ▨ Error sistemático    ▩ Fraude

**Error sistemático.-** Se produce durante el diseño del proyecto de investigación por errores en la selección de los sujetos o recopilación de la información. Suele llamarse, sesgo. No se puede cuantificar y no desaparece aunque estudiemos a toda la población.

**Precisión:** Es el grado de ausencia de error aleatorio. Es decir, cuanto mayor sea el tamaño muestral menor es el papel que juega el azar en nuestra estimaciones (el intervalo de confianza será menor y aumenta la precisión).

- **Validez:** Un estudio tendrá validez si realmente mide lo que pretendemos medir. Esto es, este libre de sesgos o errores sistemáticos.
- **Validez interna:** capacidad que tiene un estudio para generalizar los resultados obtenidos en una muestra a la población experimental.
- **Validez externa:** capacidad que tiene un estudio para poder generalizar los resultados la población diana. En la medida en que utilizemos unos criterios de selección mas restrictivos menor validez externa.

# Error Muestral

- Cuando se extrae una muestra de una población es frecuente que los resultados obtenidos de la muestra no sean exactamente los valores reales de la población.
- El error de muestreo ocurre al estudiar una muestra en lugar de la población total.
- La diferencia entre el valor del parámetro de una población y el obtenido de una muestra recibe el nombre de error muestral.

Ejemplo: (Para la media)

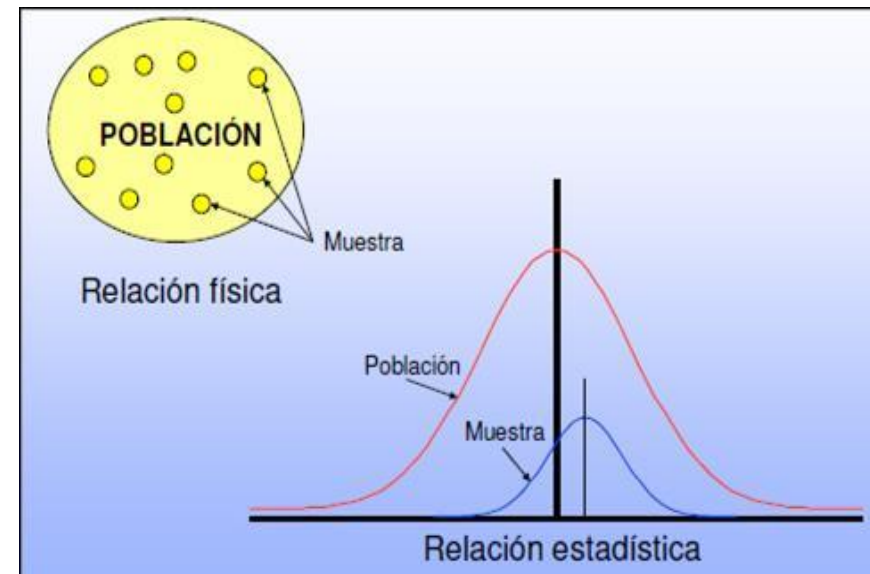
$$\text{Error Muestral} = \bar{x} - \mu$$

Donde:

$\bar{x}$  = Media muestral

$\mu$  = Media poblacional

¡Siempre presente dado que se usa una muestra!



# Estimación del Error

## Tamaño de la muestra para $\mu$

Recuerde que la desviación normal estandarizada  $Z$  de la distribución de muestreo de una v.a  $X$  se define como

$$Z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$$

Operando para "n"

$$n = \left( \frac{Z * \sigma}{e} \right)^2$$

Donde  $e = (\bar{x} - \mu)$

$e$  se define como el error muestral, la diferencia entre la media muestral y la media poblacional que es el error que se admite.

# Estimación del Error

## □ Tamaño de la muestra para P

- Cuando la variable es una proporción

$$e = z \left( \frac{\sqrt{pq}}{n} \right)$$

Donde:

e= error muestral

Z: grado de confianza de la estimación

p: proporción de la muestra para la categoría a examinar

q: 1-p

n: tamaño de la muestra

# Cálculo del tamaño de la “n”

**Para ( $\mu_1 - \mu_2$ )**

**Para estimación P**

$$n = \left[ \frac{zs}{e} \right]^2$$

$$n \cong n - 1 = \frac{z^2 pq}{e^2}$$

I

# **Muestreo Aleatorio Simple (M.A.S.)**

# Muestreo aleatorio simple sin reemplazo

- ❑ Método más usado, la condición es  $p$  cuando se tiene un marco de muestreo que especifique la manera de identificar cada unidad en la población.
- ❑ Además no se tiene conocimiento a priori sobre los posibles valores de  $Y_i$  ni otras mediciones asociadas a  $Y_i$ .
- ❑ En este caso cada unidad se extrae con igual probabilidad, por etapas, y sin reemplazo, hasta tener las  $n$  unidades de la muestra.

□ En la primera extracción, la probabilidad de que se seleccione una de las  $n$  unidades es  $\frac{n}{N}$ .

□ En la segunda extracción la probabilidad de que se seleccione una de las restantes  $n-1$  unidades es:  $\frac{n-1}{N-1}$  y así sucesivamente.

□ En la selección  $k$ , la probabilidad de una unidad  $i$  es  $\frac{n-k+1}{N-k+1}$ .

□ Para estimar  $\bar{Y} = \sum_{i=1}^N Y_i / N$  se obtiene el

promedio de la muestra:  $\bar{y} = \hat{Y} = \sum_{i=1}^n y_i / n \dots (1)$

□ Este es un estimador insesgado (  $E(\bar{y}) = \bar{Y}$  , el promedio de los posibles valores  $\bar{y}$  al tomar muchas muestras es  $\bar{Y}$  ).

□ La varianza de  $\bar{y}$  es:

$$V(\bar{y}) = E(\bar{y} - \bar{Y})^2 = \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}$$

donde 
$$S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$$

□ Nótese que si  $N$  es infinito,  $V(\bar{y}) = \frac{S_y^2}{n}$ ,  
es el resultado que se obtiene  
para poblaciones infinitas.

- $\frac{n}{N}$  es la fracción de muestreo o proporción de la población que se muestrea, y
- $1 - \frac{n}{N}$  es el factor de corrección por de poblaciones finitas (fcp)
- Se puede demostrar que con este proceso de selección, la probabilidad de que cualquier unidad  $u_i$  esté en la muestra es  $\pi_i = \frac{n}{N}$  y la de que ambas una  $u_i$  y una  $u_j$  estén en la muestra es  $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$

- Para estimar el total  $\sum_i^N Y_i = N\bar{Y} = Y$

tenemos:  $\hat{Y} = N\hat{\bar{Y}} = N\bar{y}$

- Además si  $\hat{\theta} \sim N[\theta, V(\hat{\theta})]$  , entonces:

$$P\left[\hat{\theta} - 1.96\sqrt{V(\hat{\theta})} \leq \theta \leq \hat{\theta} + 1.96\sqrt{V(\hat{\theta})}\right] = 0.95$$

- Si no conocemos  $V(\hat{\theta})$  tenemos que estimarla:

$$P\left[\hat{\theta} - 1.96\sqrt{\hat{V}(\hat{\theta})} \leq \theta \leq \hat{\theta} + 1.96\sqrt{\hat{V}(\hat{\theta})}\right] \doteq 0.95$$

- En el caso particular del “*mas*” tenemos:

$$\theta = \bar{Y}, \quad \hat{\theta} = \bar{y} \quad y \quad V(\hat{\theta}) = V(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}$$

□ En el caso particular del “*m.a.s.*” tenemos:

$$\theta = \bar{Y}, \hat{\theta} = \bar{y} \quad y \quad V(\hat{\theta}) = V(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}$$

$$P \left[ \bar{y} - \underbrace{1.96 \sqrt{\left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}}}_{\delta} \leq \bar{Y} \leq \bar{y} + 1.96 \sqrt{\left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}} \right] = 0.95$$

$$\Rightarrow P \left[ |\bar{y} - \bar{Y}| < \delta \right] = 0.95$$

$\delta$  = error absoluto.

□ Despejando  $n$  de  $\delta = 1.96\sqrt{V(\bar{y})}$  se tiene:

$$n = \frac{1}{\frac{\delta^2}{(1.96)^2 S_y^2} + \frac{1}{N}} \doteq \frac{(1.96)^2 S_y^2}{\delta^2}$$

□ Recordemos que:

$$E(y_i - E(y_i))^2 = E(y_i - \bar{Y})^2 = \sigma_y^2 = \frac{\sum_i^N (Y_i - \bar{Y})^2}{N}$$

$$S_y^2 = \frac{N}{N-1} \sigma_y^2, \quad S_y^2 = \frac{\sum (Y_i - \bar{Y})^2}{(N-1)}$$

# Estimación del total poblacional ( $\tau$ )

□ Si lo que se quiere  
estimar es  $Y$ . Entonces

$$\hat{Y} = N\hat{y} = Ny$$

$$V(\hat{Y}) = V(Ny) = N^2V(y) = N^2\left(1 - \frac{n}{N}\right)\frac{S^2}{n}$$

$$P\left[\hat{Y} - 1.95N^2\left(1 - \frac{n}{N}\right)\frac{S^2}{n} \leq Y \leq \hat{Y} + 1.95N^2\left(1 - \frac{n}{N}\right)\frac{S^2}{n}\right] = 0.95$$

# Tamaño de la muestra

- El valor de  $S^2_y$  ó  $\sigma^2_y$  se estima con una *prueba piloto* o bien se “*adivina*” usando tablas, así como del conocimiento previo sobre la población.
- Si se considera que  $\bar{y}$  no se ajusta a la distribución normal, se usa el criterio de fijar la magnitud de la varianza o del coeficiente de variación de  $\bar{y}$  .

Se determina  $n$  para que produzca un coeficiente de variación dado ( $CV_0$ ) usando estimaciones “gruesas” de  $\bar{y}$  y de  $S^2_y$  .

□ Así

$$CV_0 = \frac{[V(\bar{y})]^{1/2}}{E(\bar{y})} = \frac{\left\{ \left( 1 - \frac{n}{N} \right) \frac{S_y^2}{n} \right\}^{1/2}}{\bar{Y}}$$

Despejando  $n$ , se obtiene:

$$n = \frac{S_y^2}{(CV_0)^2 \bar{Y}^2 + \frac{S_y^2}{N}}$$

Así:  $\bar{y} \sim N[\bar{Y}, V(\bar{y})]$

$$P\left[\bar{y} - z_{\alpha/2} \sqrt{V(\bar{y})} \leq \bar{Y} \leq \bar{y} + z_{\alpha/2} \sqrt{V(\bar{y})}\right] = 1 - \alpha$$

si  $1 - \alpha = .95$

$$P\left[\bar{y} - 1.96 \sqrt{\left(1 - \frac{n}{N}\right) \frac{S^2}{n}} \leq \bar{Y} \leq \bar{y} + 1.96 \sqrt{\left(1 - \frac{n}{N}\right) \frac{S^2}{n}}\right] = 0.95$$

Si se desea un tamaño de muestra tal que el error de estimación sea inferior a  $\delta$  con una probabilidad de  $1-\alpha$ , esto es:

$$P \left[ |\bar{y} - \bar{Y}| < \delta \right] = 1 - \alpha, \quad \delta = z_{\alpha/2} \sqrt{V(\hat{y})}$$

dividiendo entre  $[V(\bar{y})]^{1/2}$

$$P \left[ \frac{|\bar{y} - \bar{Y}|}{[V(\bar{y})]^{1/2}} < \frac{\delta}{[V(\bar{y})]^{1/2}} \right] = 1 - \alpha$$

De las tablas de la normal estándar,  $Z \sim N(0,1)$ , se obtiene un valor  $z_{\alpha/2}$  tal que

$$P \left[ |Z| < z_{\alpha/2} \right] = 1 - \alpha$$

( $z_{\alpha/2}$  es el valor de  $Z$  obtenido en las tablas que deja un área de  $\alpha/2$  a la derecha de él).

- Como  $\frac{|\bar{y} - \bar{Y}|}{[V(\bar{y})]^{1/2}} \sim N(0,1)$  , hacemos que

$\frac{|\bar{y} - \bar{Y}|}{[V(\bar{y})]^{1/2}}$  sea un valor arbitrario de  $Z$  y que:

$$z_{\alpha/2} = \frac{\delta}{[V(\bar{y})]^{1/2}} = \frac{\delta}{\sqrt{\frac{S_y^2}{n} \left(1 - \frac{n}{N}\right)}} \quad (a)$$

De aquí (a) se despeja  $n$ :

$$n = \frac{1}{\frac{\delta^2}{z_{\alpha/2}^2 S_y^2} + \frac{1}{N}} \doteq \frac{z_{\alpha/2}^2 S_y^2}{\delta^2}$$

si  $\alpha = 0.05$  entonces:

$$n \doteq \frac{(1.96)^2 S_y^2}{\delta^2}$$

Se puede usar  $n' = \frac{z_{\alpha/2}^2 S_y^2}{\delta^2}$  como una primera

aproximación y luego corregir usando  $n = \frac{n'}{1 + \frac{n'}{N}}$

Si no se puede suponer normalidad de la distribución del estimador, se recurre a la desigualdad de Tchebycheff.

## Desigualdad de Chebyshev

Sea  $U$  una variable aleatoria con cualquier distribución y

$$E(U) = \mu_U, \quad V(U) = \sigma_U^2$$

$$\Rightarrow P\left[|U - \mu_U| \geq \lambda \sigma_U\right] \leq \frac{1}{\lambda^2}$$

$$\Rightarrow P\left[|U - \mu_U| \leq \lambda \sigma_U\right] \geq 1 - \frac{1}{\lambda^2}$$

$$\Rightarrow P[U - \lambda\sigma_U \leq \mu_U \leq U + \lambda\sigma_U] \geq 1 - \frac{1}{\lambda^2}$$

$$\Rightarrow P\left[\bar{y} - \lambda\sqrt{V(\bar{y})} \leq \bar{Y} \leq \bar{y} + \lambda\sqrt{V(\bar{y})}\right] \geq 1 - \frac{1}{\lambda^2}$$

$$\lambda = 2 \quad 1 - \frac{1}{\lambda^2} = .75 \qquad \lambda = 3 \quad 1 - \frac{1}{\lambda^2} = .889$$

$$\lambda = 4.4 \quad 1 - \frac{1}{\lambda^2} = .95$$

$$\delta = 4.4\sqrt{V(\bar{y})} \Rightarrow n = \frac{1}{\frac{\delta^2}{(4.4)^2 S^2} + \frac{1}{N}}. \quad (2)$$

En las expresiones anteriores, si tanto  $\delta$  como  $S$  se expresan en por ciento de la media,

$$\delta' = \frac{\delta}{\bar{y}} 100, \quad CV = \frac{S}{\bar{y}} 100 \quad \text{la expresión (2) se}$$

transforma a:

$$n = \frac{1}{\frac{(\delta')^2}{Z_{\alpha/2}^2 (CV)^2} + \frac{1}{N}} = \frac{Z_{\alpha/2}^2 (CV)^2}{(\delta')^2}.$$

Si no se supone normalidad para la distribución de  $\bar{y}$  con confianza del 95%, por la desigualdad de Chebyshev, entonces (2) se transforma a:

$$n = \frac{1}{\frac{(\delta')^2}{(4.4)^2 (CV)^2} + \frac{1}{N}} \doteq \frac{(4.4)(CV)^2}{(\delta')^2}$$

# Estimación de una proporción poblacional (P)

$Y(u_i)$  es una medida o indicador de la *presencia o ausencia de una característica* en la unidad  $u_i$  con valor 1 si la característica está presente y 0 si no es así. En este caso

$\bar{Y} = P =$  *proporción de unidades en la población que tienen la característica*

$$\left( \bar{Y} = \frac{\sum_i^N Y_i}{N} = P \right).$$

$p = \bar{y}$  que es la *proporción* de unidades en la *muestra* con la característica.

El valor de  $S^2_y$  en términos de  $P$  resulta:

$$S^2_Y = \frac{\sum_i^N (Y_i - \bar{Y})^2}{N-1} = NP(1-P) \frac{1}{N-1}$$
$$= \frac{N}{N-1} P(1-P), \quad \sigma^2 = P(1-P)$$

con estimador

$$\hat{S}_y^2 = s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{n\hat{P}}{n-1} (1 - \hat{P}).$$

Con este nuevo valor la expresión resulta:

$$n = \frac{\frac{N}{N-1} (1-P)}{(CV_0)^2 P + \left(\frac{1-P}{N-1}\right)} \doteq \frac{1-P}{P(CV_0)^2} \quad (4)$$

Para usar esta expresión, se estima *a priori* o con una prueba piloto el valor de  $P$  y se fija el  $CV_0$  que se desea.

Si utilizamos la desigualdad de Chebyshev  
tenemos:

$$n = \frac{1}{\frac{\delta^2}{(4.4)^2 \frac{N}{N-1} P(1-P)} + \frac{1}{N}} \doteq \frac{(4.4)^2 \frac{N}{N-1} P(1-P)}{\delta^2}$$

$$n \doteq \frac{(4.4)^2}{\delta^2} = \frac{5}{\delta^2}$$

Nótese que si  $P$  está cercano a cero, el valor de  $n$  *aumenta*.

Esto indica que para estimar la proporción de unidades con una característica rara se requieren *muchas unidades* en la muestra.

Esto es lo contrario de lo que sucede si se usa la aproximación a la normal, en cuyo caso se usa la expresión (4) con

$$S_Y^2 = \frac{NP}{N-1}(1-P)$$

$$\Rightarrow n = \frac{1}{\frac{\delta^2}{z_{\alpha/2}^2 S_y^2} + \frac{1}{N}} = \frac{z_{\alpha/2}^2 S_y^2}{\delta^2}.$$

Si se quiere conocer  $P$ , las  $Y_i$  son 0 ó 1.

$$S_y^2 = \frac{N}{N-1} P(1-P) \doteq P(1-P)$$

$$\Rightarrow n = \frac{z_{\alpha/2}^2 P(1-P)}{\delta^2}$$

Si  $\alpha = .05 \Rightarrow z_{\alpha/2} = 1.96 \doteq 2$  , además como

la varianza de  $\hat{P}$  es máxima cuando  $P = 0.5$ ,  
se usa  $P(1-P) = (.5)(.5) = 0.25$  como margen de  
seguridad

$$\Rightarrow n \doteq \frac{2^2 (.25)}{\delta^2} = \frac{1}{\delta^2}$$

Además, si  $\hat{P} \sim N(P, V(\hat{P}))$  entonces se *debe* reportar el resultado de la estimación de  $P$  con un intervalo de confianza aproximado dado por:

$$P \left[ p - 1.96\sqrt{\hat{V}(p)} \leq P \leq p + 1.96\sqrt{\hat{V}(p)} \right] \cong .95,$$

$$\hat{V}(\hat{p}) = \left( 1 - \frac{n}{N} \right) \frac{N\hat{p}(1 - \hat{p})}{(N - 1)n}$$

# Cálculo de “n”

Así

$$CV_0 = \frac{[V(\bar{y})]^{1/2}}{E(\bar{y})} = \frac{\left\{ \left( 1 - \frac{n}{N} \right) \frac{S_y^2}{n} \right\}^{1/2}}{\bar{Y}}$$

Despejando  $n$ , se obtiene:

$$n = \frac{S_y^2}{(CV_0)^2 \bar{Y}^2 + \frac{S_y^2}{N}}$$

Si  $n$  es "grande" se espera que el Teorema Central del Límite dé una buena aproximación de la distribución de  $\bar{y}$  .

Así:  $\bar{y} \sim N[\bar{Y}, V(\bar{y})]$

$$P\left[\bar{y} - z_{\alpha/2} \sqrt{V(\bar{y})} \leq \bar{Y} \leq \bar{y} + z_{\alpha/2} \sqrt{V(\bar{y})}\right] = 1 - \alpha$$

si  $1 - \alpha = .95$

$$P\left[\bar{y} - 1.96 \sqrt{\left(1 - \frac{n}{N}\right) \frac{S^2}{n}} \leq \bar{Y} \leq \bar{y} + 1.96 \sqrt{\left(1 - \frac{n}{N}\right) \frac{S^2}{n}}\right] = 0.95$$

Entonces  $\frac{\bar{y} - \bar{Y}}{[V(\bar{y})]^{1/2}}$  se distribuye

aproximadamente como una normal estandarizada (media cero y varianza uno), donde

$$V(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}$$

Si se desea un tamaño de muestra tal que el error de estimación sea inferior a  $\delta$  con una probabilidad de  $1-\alpha$ , esto es:

$$P \left[ |\bar{y} - \bar{Y}| < \delta \right] = 1 - \alpha, \quad \delta = z_{\alpha/2} \sqrt{V(\hat{y})}$$

dividiendo entre  $[V(\bar{y})]^{1/2}$

$$P \left[ \frac{|\bar{y} - \bar{Y}|}{[V(\bar{y})]^{1/2}} < \frac{\delta}{[V(\bar{y})]^{1/2}} \right] = 1 - \alpha$$

De las tablas de la normal estándar,  $Z \sim N(0,1)$ , se obtiene un valor  $z_{\alpha/2}$  tal que

$$P \left[ |Z| < z_{\alpha/2} \right] = 1 - \alpha$$

( $z_{\alpha/2}$  es el valor de  $Z$  obtenido en las tablas que deja un área de  $\alpha/2$  a la derecha de él).

Como  $\frac{|\bar{y} - \bar{Y}|}{[V(\bar{y})]^{1/2}} \sim N(0,1)$  , hacemos que

$\frac{|\bar{y} - \bar{Y}|}{[V(\bar{y})]^{1/2}}$  sea un valor arbitrario de  $Z$  y que:

$$z_{\alpha/2} = \frac{\delta}{[V(\bar{y})]^{1/2}} = \frac{\delta}{\sqrt{\frac{S_y^2}{n} \left(1 - \frac{n}{N}\right)}} \quad (a)$$

De aquí (a) se despeja  $n$ :

$$n = \frac{1}{\frac{\delta^2}{z_{\alpha/2}^2 S_y^2} + \frac{1}{N}} \doteq \frac{z_{\alpha/2}^2 S_y^2}{\delta^2}$$

si  $\alpha = 0.05$  entonces:

$$n \doteq \frac{(1.96)^2 S_y^2}{\delta^2}$$

Se puede usar  $n' = \frac{z_{\alpha/2}^2 S_y^2}{\delta^2}$  como una primera

aproximación y luego corregir usando  $n = \frac{n'}{1 + \frac{n'}{N}}$

Si no se puede suponer normalidad de la distribución del estimador, se recurre a la desigualdad de Chebyshev.

$$\Rightarrow P[U - \lambda\sigma_U \leq \mu_U \leq U + \lambda\sigma_U] \geq 1 - \frac{1}{\lambda^2}$$

$$\Rightarrow P\left[\bar{y} - \lambda\sqrt{V(\bar{y})} \leq \bar{Y} \leq \bar{y} + \lambda\sqrt{V(\bar{y})}\right] \geq 1 - \frac{1}{\lambda^2}$$

$$\lambda = 2 \quad 1 - \frac{1}{\lambda^2} = .75 \qquad \lambda = 3 \quad 1 - \frac{1}{\lambda^2} = .889$$

$$\lambda = 4.4 \quad 1 - \frac{1}{\lambda^2} = .95$$

**II**

**Muestreo Sistemático Aleatorio  
(M.S.A.)**

- ❑ Este tipo de muestreo se utiliza mucho en el control de la calidad en los procesos de líneas de fabricación en donde se opera sin interrupciones.
- ❑ No es un muestreo probabilístico pero si representativo del proceso de producción de la unidades muestrales, generalmente más acucioso que un muestreo CA (Completo la Azar) y sobre todo, más económico.

# Muestreo Sistemático Aleatorio (MSA)

## Introducción

Como ya se había comentado el MASI, es una técnica en la que se elige mediante una observación ( $k$ ) cualquiera al azar, una vez elegida esta se hace una sucesivamente respetando el intervalo o distancia preestablecido.

Método empleado cuando se desea cubrir el rango de las unidades comprendidas en la población.

## Integración de la muestra:

Se hará de manera sistemática a partir de la primera ( $k$ ) según el proceso siguiente.

- Generamos un número aleatorio ( $R$ ). Obtenga su valor entero comprendido en el rango de 1 a  $N$ .
- A partir de este número se obtiene un cierto valor " $k$ " donde  $k = N/n$ . Sumamos  $k$  a  $R$  para obtener el siguiente número que corresponde a segundo elemento muestral seleccionado.
- Procedimiento que se repite para cada número seleccionado, recorriendo así todo el marco muestral de la población.

**Nota:** Si la secuencia proporciona un elemento muestral fuera del rango, se continúa a partir del límite inferior del rango de forma tal que se recorra la imagen de los datos poblacionales.

- Para determinar que elementos conforman la muestra, se procede a asignar un número de identificación a los elementos muestrales.
- Estos serán los elementos que conforman la población.
- mediante números aleatorios se seleccionan uno a uno hasta completar el tamaño de **n**.
- Los números aleatorios se pueden generar mediante un software, calculadora o en última instancia recurrir a las tablas de números aleatorios.
- En caso de repetirse un número ya escogido se desecha y se continúa el proceso.

**Nota:** si los numeros aleatorios son generados por computadora o la calculadora, tendrán valores entre 0 y 1, por lo que deben multiplicarse por un Número  $k$ , tal que proporcione un número entero.

**Ejemplo.** Un investigador desea determinar la calidad del jarabe contenido en la savia de las árboles. El número total de árboles  $N$  es desconocido; por lo tanto es imposible realizar una muestra irrestricta aleatoria de árboles. Como un procedimiento alternativo, el investigador decide usar una muestra sistemática de 1-en-7. Los datos se listan a continuación. Donde el porcentaje del contenido de la savia en los árboles muestreadas. Emplear los datos para estimar  $\mu$ , el contenido de savia promedio de los árboles y establezca un límite para el error de estimación.

Árbol muestreado	Contenido de azúcar en la savia, $y$	$Y$
• 1	82	6724
• 2	76	5776
• 3	83	6889
• .	.	.
• .	.	.
• .	.	.
• 210	84	7056
• 211	80	6400
• 212	79	6241
•		

## Solución

- Una estimación de  $\mu$  esta dada por

$$\bar{y}_{sy} = \frac{\sum_{i=1}^n y_i}{n} = \frac{17064}{212}$$

- Para encontrar un límite para el error de estimación, primero debemos calcular

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_{sy})^2}{n-1} = \frac{48291}{211}$$

- Intuitivamente, podemos suponer que la población de árboles en la finca es aleatoria. Según esta suposición la varianza estimada de esta dada por

$$\sqrt{V(\bar{y}_{sy})} = \sqrt{206}$$

habiendo realizado la muestra de 1 – en – 7 , conocemos N.  
 Suponiendo N = 1484 resulta

$$V_{(s)} = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

Un limite aproximado para el error de estimación esta dado por

$$2 \sqrt{V_{(s)}} = 2.06\%$$

- En resumen, estimamos que el promedio de azúcar contenido en la savia es de 80.5%. Estamos bastante confiados en que el limite para el error de estimación es menor de 2.9%.

Ejemplo 2. Una muestra sistemática de 1 en 6 es obtenida de una lista de votantes registrados para estimar la proporción de votantes que están a favor de la emisión de bonos propuesta. Los resultados codificados de esta encuesta de elección previa se muestran en la tabla siguiente. Estime la proporción ( $p$ ) de los  $N= 5,775$  votantes registrados que están a favor de la emisión de bonos propuesta y obtenga un límite para el error de estimación.

Votante	Respuesta
4	1
10	0
16	1
.	
.	
.	
5760	0
5766	0
5772	

$$\sum_{I=1}^{962} \lambda_i = 652$$

# Solución

- La proporción muestral es:

$$p_{sy} = \frac{\sum_{i=1}^{962} y_i}{962} = \frac{652}{962} = 0.678$$

- Puesto que N es grande y varios puntos de inicio aleatorio fueron seleccionados en la extracción de la muestra sistemática, podemos suponer que

$$\hat{V}(\hat{p}_{sy}) = \frac{\hat{p}_{sy} \hat{q}_{sy} (N-n)}{n-1} \left( \frac{N-n}{N} \right)$$

- Proporciona una buena estimación de

- El límite de error de estimación es

$$\hat{V}(\hat{p}_{sy})$$

$$2\sqrt{\hat{V}(\hat{p}_{sy})} = 2\sqrt{\frac{\hat{p}_{sy} \hat{q}_{sy} (N-n)}{n-1} \left( \frac{N-n}{N} \right)}$$

$$= 2\sqrt{\left( \frac{0.678 \times 0.322 \times 5775}{961} \right) \left( \frac{5775}{5775} \right)} = 0.028$$

- Por lo tanto se estima que 67.8% de los votantes registrados favorecen a la emisión de bonos propuesta.
- Con un error de estimación menor al 2.8%.

# MSA Replicado

- ❑ Es el MASI con iteraciones. Es decir, es la selección de más de una muestra sistemática.
- ❑ Por ejemplo: 10 muestras sistemáticas de 1 en 50, cada una conteniendo 6 mediciones, podría ser obtenidas en aproximadamente en el mismo tiempo que una muestra sistemática de 1 en 5 conteniendo 60 mediciones.
- ❑ Ambos procedimientos producen 60 mediciones para estimar la media poblacional  $\mu$ . La bondad del procedimiento de muestreo replicado permite estimar al emplear el cuadrado de las desviaciones de las  $n_s = 10$  medias muestrales individuales alrededor de su media.  $V(\bar{y}_{.sy})$
- ❑ El promedio de las 10 medias muestrales permitirá estimar  $\mu$ .

Ejemplo 3. Un parque estatal cobra la admisión por automóvil en lugar de por persona, y un funcionario del parque quiere estimar el número promedio de personas por automóvil para un día efectivo en particular durante el verano. El funcionario sabe por experiencia que entrarán al parque alrededor de 400 automóviles y quiere muestrear 80 de ellos. Para obtener una estimación de la varianza, utiliza el muestreo sistemático replicado con 10 muestras de 8 automóviles de cada una. usando los datos que se presentan en la tabla adjunta estime el número promedio de personas por automóvil y establezca un límite para el error de estimación

Punto de inicio aleatorio	Segundo elemento	Tercer elemento	Cuarto elemento	Quinto elemento	Sexto elemento	Séptimo elemento	Octavo elemento $\bar{y}_i$	
2(3)	52(4)	102(5)	152(3)	202(6)	252(1)	302(4)	352(4)	3.75
5(5)	55(3)	105(4)	155(2)	205(4)	255(2)	305(3)	355(4)	3.38
13(6)	63(4)	113(6)	163(7)	213(2)	263(3)	313(2)	363(7)	2.88
26(4)	76(5)	126(7)	176(4)	226(2)	276(6)	326(2)	376(6)	4.50
31(7)	81(6)	131(4)	181(4)	231(3)	281(6)	331(7)	381(5)	5.25
35(3)	85(3)	135(2)	185(3)	235(6)	285(5)	335(6)	385(8)	4.50
40(2)	90(6)	140(2)	190(5)	240(5)	290(4)	340(4)	390(5)	4.12
45(2)	95(6)	145(3)	195(6)	245(4)	295(4)	345(5)	395(4)	4.25
46(6)	96(5)	146(4)	196(6)	246(3)	296(3)	346(5)	396(3)	4.38

## Solución

□ Para una muestra sistemática  $k = \frac{N}{n} = \frac{400}{80} = 5$

por lo tanto para  $n_s = 10$  muestras

$$k' = 10k = 10(5) = 50$$

□ Los siguientes 10 números aleatorios entre el 1 y el 50 son:

13,35,2,40,26,7,31,45,5,46

Los automóviles con estos números forman los puntos de inicio aleatorio para las muestras sistemáticas.

□ La cantidad  $\bar{y}_1$  de la tabla es el promedio para la primera hilera,  $\bar{y}_2$  es el promedio de la segunda hilera, y así sucesivamente. La estimación de  $\mu$  es

$$\bar{\mu} = \frac{1}{n_s} \sum_{i=1}^{n_s} y_i = \frac{1}{10} (3.75 + 3.38 + \dots + 4.38) = 4.16$$

puede establecerse la siguiente identidad

$$\sum_{i=1}^{n_s} (\bar{y}_i - \bar{\mu})^2 = \sum_{i=1}^{n_s} y_i^2 - \frac{1}{n_s} \left( \sum_{i=1}^{n_s} \bar{y}_i \right)^2$$

sustituyendo, se obtiene

$$\sum_{i=1}^{n_s} (\bar{y}_i - \bar{\mu})^2 = 177.410 - \frac{1}{10}(1733.06) = 4.104$$

□ Varianza estimada de  $\hat{\mu}$  es:

$$\hat{\mu} \pm 2\sqrt{\hat{V}(\hat{\mu})} = 4.163 \pm 2\sqrt{0.0365} = 4.16 \pm 0.38$$

□ Estimación de  $\mu$  con un limite para el error de estimación es

$$\hat{V}(\hat{\mu}) = \left(\frac{N-n}{N}\right) \frac{\sum_{i=1}^{n_s} (\bar{y}_i - \hat{\mu})^2}{n_s(n_s-1)} = \left(\frac{400-80}{400}\right) \left[\frac{4.104}{10(9)}\right] = 0.0365$$

Observe que la mejor estimación del promedio de personas por automóvil es 4.16. y con un error de estimación menor a 0.38 con probabilidad de 0.95

# Bibliografía

1. Cochran, William. (1985) "Técnicas de Muestreo". Compañía Editorial Continental, S.A. México.
2. Downie, M. (1973) "Métodos Estadísticos Aplicados". Harper & Row Publishers INC. México.
3. Lewis, Alvin. (2000) "Bioestadística". Compañía Editorial Continental, S.A. México. S/F.
4. Neter y Otros.(2011), "Fundamentos de Estadística para Economía y Negocios". Compañía Editorial Continental, S.A. México. S/F.
5. Martín Pliego, F. y Ruiz-Maya, L. (1995) *Estadística II: Inferencia*. Madrid: AC.
6. Méndez, I., & Quintana, C. R. H. (2007). Muestreo: Respuesta aleatorizada.  
En:<http://www.dpye.iimas.unam.mx/finales2007/MuestreoRespuestaAleatoriada>. Especialidad en Estadística Aplicada. IIMAS, UNAM.

6. Méndez, I., Eslava, G., & Romero, P. (2004). Conceptos Básicos de Muestreo. Cd México: IIMAS, UNAM.
7. Levy, P. S., Lemeshow, S. (1991), Sampling of Populations Methods and Applications. John Wiley & Sons.
8. Pérez, L. C. (2000). Técnicas de muestreo estadístico. Teoría, práctica y aplicaciones informáticas. México, DF: Alfaomega-Rama.
9. Rendón, S. G. (1997). Métodos estadísticos. Muestreo, diseños experimentales, estadística no paramétrica. México, DF: Universidad Autónoma de Chapingo.
10. Scheaffer, R. L., Mendenhall, W., & Lyman, O. (1987). Elementos de muestreo. México, DF: Grupo Editorial Iberoamérica.
11. Stevenson, William. (1981) "Estadística para Administración y Economía". Harla. México.