Universidad Autónoma de Estado de México

Facultad de Economía

Licenciatura en Economía

Distribuciones Muestrales

Material didáctico: visual elaborado por Fidelmar Sandoval Durán

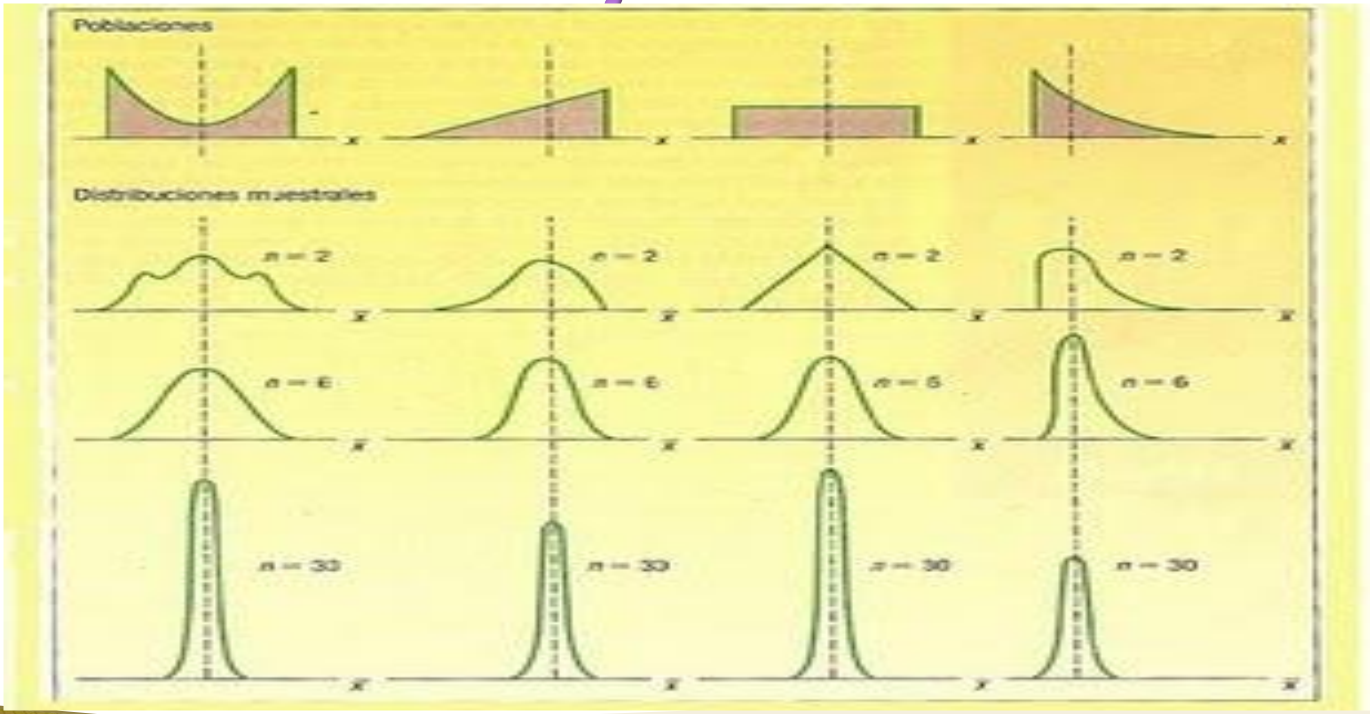
Objetivos

• Entender los conceptos de estimación puntual y estimación por intervalos.
• Calcular las estimaciones para la media poblacional, tanto en el caso en que la desviación
estándar poblacional sea conocida como en el caso de que sea desconocida.
• Calcular las estimaciones (puntuales y por intervalos) para la probabilidad de éxito de una
binomial.
• Saber interpretar correctamente los resultados de las estimaciones por intervalos.

Diapositivas	SUGERENCIAS METODOLÓGICAS Y TECNICAS
1-3	Es recomendable, antes de comenzar la presentación, que comente con el grupo las expectativas frente al tema y para que creen que les servirá.
4-27	Comente con el grupo como cambian las distribuciones muestrales cuando el tamaño de las muestras se y solicíteles que den otros ejemplos. No dilate mucho el comentario, sea preciso y recoja los ejemplos anotándolos en algún lado para que estén presentes durante el trabajo.
28-73	Aquí, además de los ejemplos presentados puede hacer uso de algunas actividades que los lleve a diseñar diagramas.

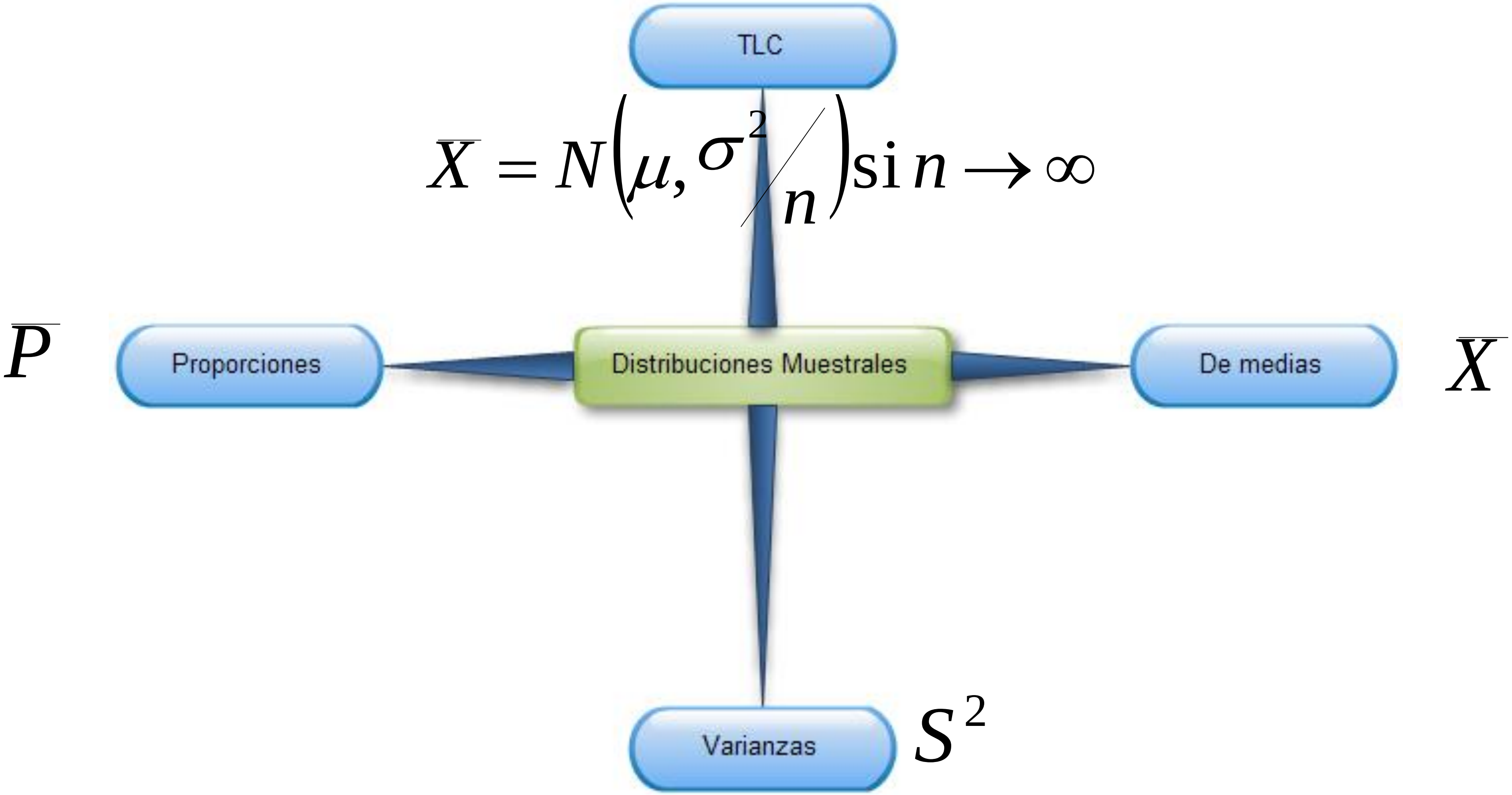
Universidad Autónoma de Estado de México
Facultad de Economía
Licenciatura en Economía

Muestreo y estimación



Material didáctico: visual elaborado por Fidelmar Sandoval Durán

$$\bar{X} = N\left(\mu, \sigma^2/n\right) \text{si } n \rightarrow \infty$$



Contenido

Capitulo/tema	Anderson	Mendenhall	Newbold
7.1	El problema	Introducción	Muestreo de una población
7.2	MAS	7.2 Planes de muestreo y diseños experimentales	Distribución de las medias muestrales en el muestreo. TLC. Intervalos de aceptación
7.3	Estimación puntual	Estadísticos y distribuciones muestrales	Distribución de proporciones muestrales en el muestreo
7.4	Introducción Distribuciones muestrales	TLC	Distribución de las varianzas muestrales en el muestreo.

Capitulo/tema	Anderson	Mendenhall	Newbold
7.5	Distribuciones muestrales de medias	Distribución muestral de medias	
7.6	Distribuciones muestrales de proporciones	Distribuciones muestrales de proporciones	
7.7	Propiedades de los estimadores puntuales	Control estadístico de procesos	
7.8	Otros métodos de muestreo		
7.8	Otros métodos de muestreo		

Términos y Conceptos	A	M	N
Plan de muestreo		255	
Estudio observacional		256	
Estudio experimental		257	
MAS de una población finita	260		
MAS de una población infinita	261		
Parámetro	258	255	
Estadístico	264	260	

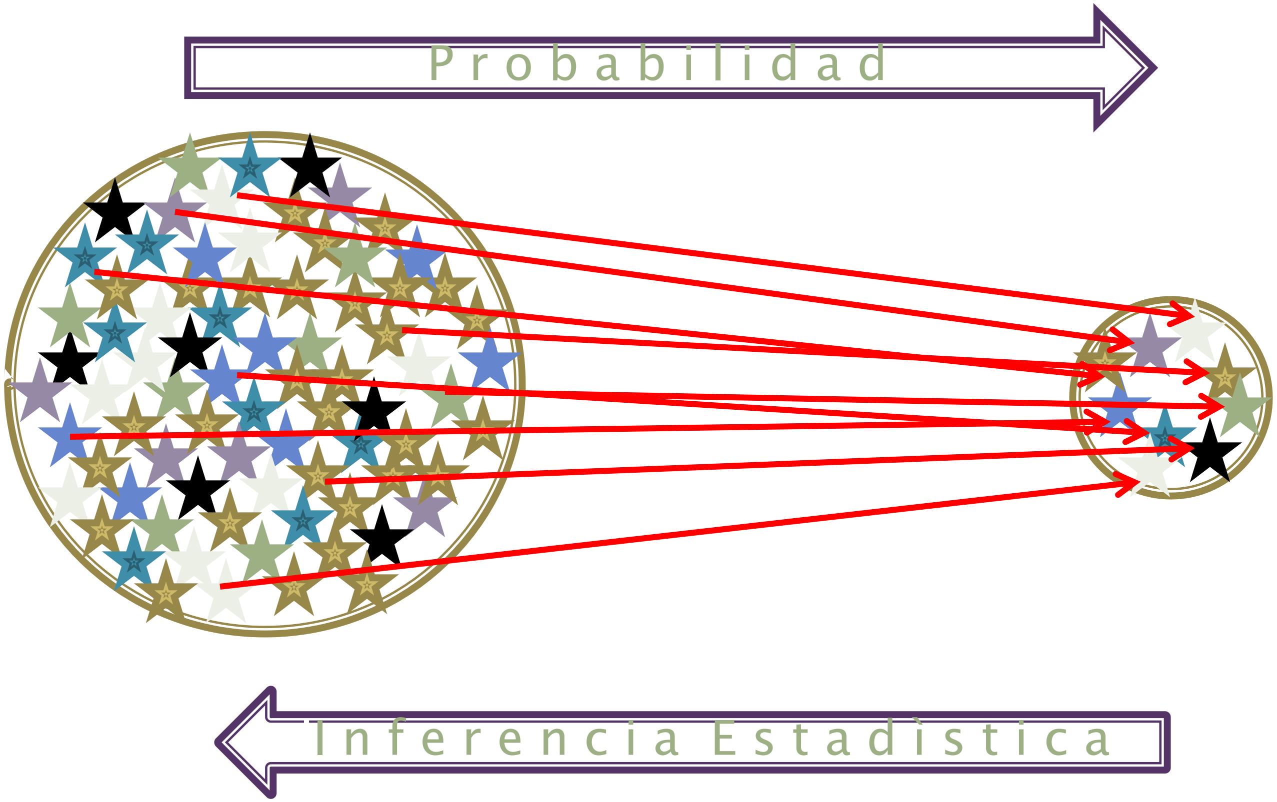
Estimador	265		
Estimación	265		
Distribución muestral	267	260	251
Distribución muestral de medias	270	266	254
Insesgadez	286		
TLC	272	263	260
Eficiencia	287		

Términos y Conceptos	A	M	N
Consistencia	287		
Distribución muestral de proporciones	280	275	272
Muestreo aleatorio simple	260	256	250
Muestreo estratificado	288	257	
Muestreo por conglomerados	289	258	
Muestreo sistemático	289	258	

Términos y Conceptos	A	M	N
Muestreo sistemático	289	258	
Muestreo por conveniencia	290	258	
Muestreo subjetivo (Juicio)	290	258	
Muestreo por cuotas		258	
Intervalo de aceptación			265
Distribución muestral de varianzas			277
Distribución Ji-cuadrado			279

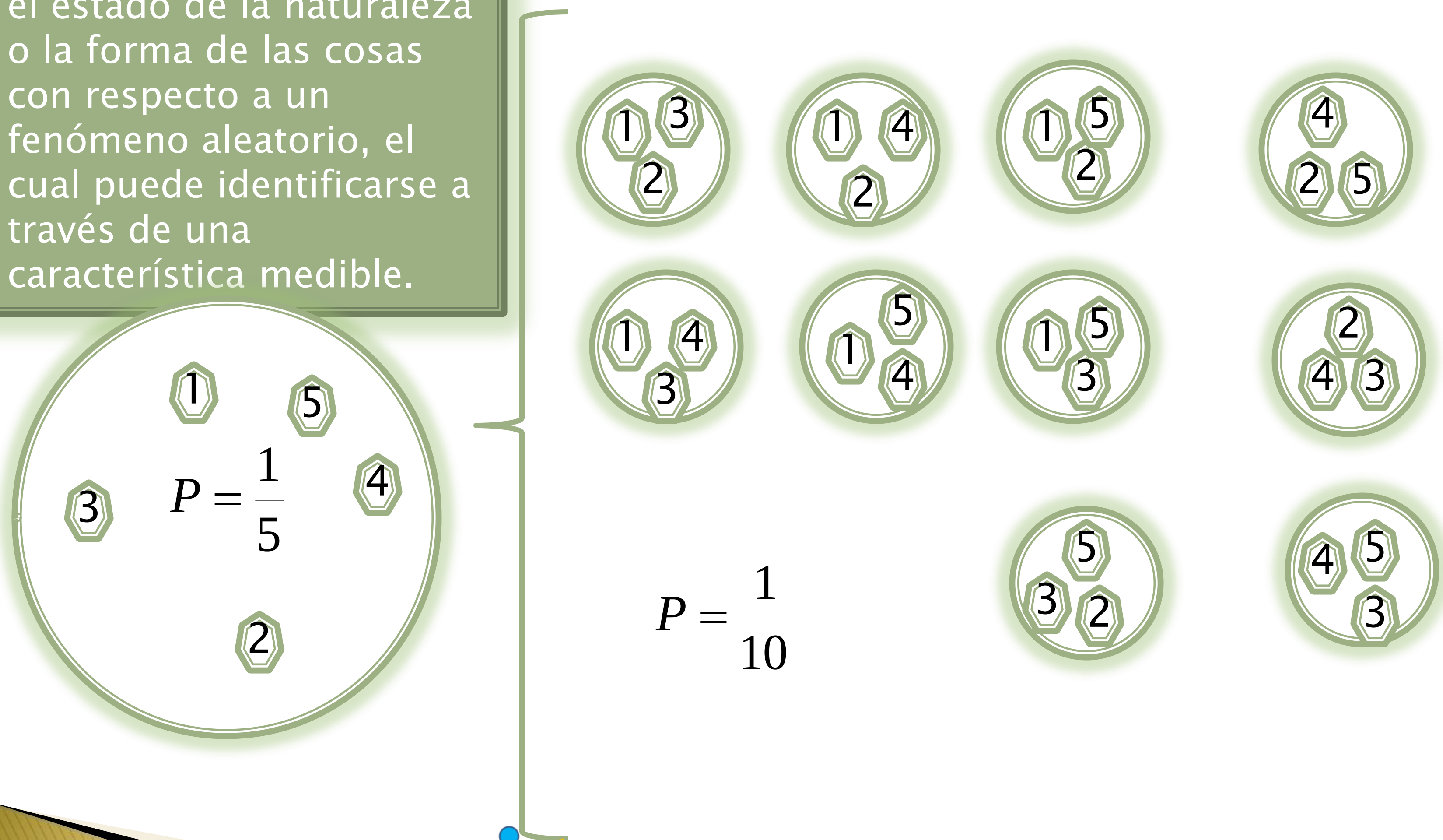
Capitulo/ batería	Anderson	Mendenhall	Newbold
7.1	19, 20, 21	7.1, 7.5, 7.11	7.5, 7.6, 7.7
7.2	50, 51, 52	7.69, 7.75, 7.76	7.21, 7.31, 7.33
7.3			7.63 a 7.71
8.4	10, 12, 16	8.5, 8.7, 8.15	8.4, 8.6, 8.9
8.5	24, 27, 30	8.24, 8.30, 8.37	8.17, 8.19, 8.21
8.6	34, 35, 37	8.40, 8.43, 8.44	8.34, 8.40, 8.41
8.7	45, 47, 49	8.60, 8.75, 8.91	8.47, 8.52, 8.53
9.8	5,6, 7	9.3, 9.6, 9.11	9.4, 9.10, 9.18

Actividad	Portafolios
1	Lista de alumnos inscritos Facultad de Economía 2010B
2	Base de datos. Revisar en www.biestadistica.uma.es/baron/apuntes/
3	Diseño de cuestionario: temática (Problema), Objetivos, variables. Preparar documento de 3 cuartillas sobre "El Diseño de Cuestionarios" de INEGI. Cap. 1-4; pp1-26. http://ctreig.jalisco.gob.mx/normatividad/septiembre_liniamientos/Diseno_Cuestionarios.pdf
4	Elaborar el primer borrador de cuestionario: destacando los subtemas, las preguntas y, las variables
5	Presentar para revisión borrador (firma de avance)



Material didáctico: visual elaborado por Fidelmar Sandoval Durán

Una población representa el estado de la naturaleza o la forma de las cosas con respecto a un fenómeno aleatorio, el cual puede identificarse a través de una característica medible.



$$P = \frac{1}{5}$$

$$P = \frac{1}{10}$$

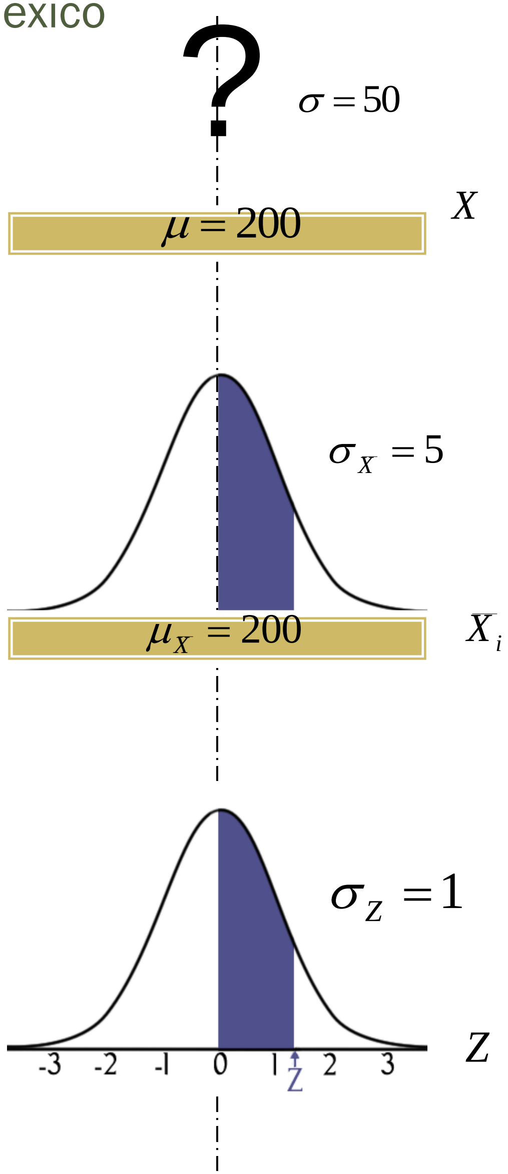
Batería 1

A.7.19 – La media de una población es 200 y su desviación estándar es 50. suponga que se selecciona una MAS de tamaño 100 y se usa $\bar{X}$ para estimar μ .

a) ¿Cuál es la probabilidad de que la diferencia entre la media muestral y la media poblacional no sea mayor que ± 5 ?

b) ¿De que la diferencia entre la media muestral y la media poblacional no sea mayor que ± 10 ?

Datos	Formalización de preguntas
✓. $\mu = 200$ ✓. $\sigma = 50$ ✓. $n = 100$ ✓. Se usa x para estimar μ	a) $P(\lvert \bar{X} - \mu \rvert \leq 5) = ?$ b) $P(\lvert \bar{X} - \mu \rvert \leq 10) = ?$



Teoría

Se desconoce la forma de la población, pero como la muestra es grande se aplica a la distribución muestral el TLC, consecuencia de ello es que ésta tiene una distribución

$$N(\mu_X;\sigma_X) \therefore \mu_X = \mu \text{ y } \sigma_X = \frac{\sigma}{\sqrt{n}}$$

La diferencia de medias no debe ser mayor de 5 en un caso, y de10 en el otro:

$|\bar{X} - \mu| \leq 5 \text{ y } |\bar{X} - \mu| \leq 10$

La probabilidad se determina como el área sobre el intervalo de la DMM correspondiente a la distribución N(0;1).

Desarrollo y operaciones: a)

- $P(|\bar{X} - \mu| \leq 5) =$
- $P(-5 \leq \bar{X} - \mu \leq 5) =$
- $P\left(\frac{-5}{\sigma_X} \leq \frac{\bar{X} - \mu}{\sigma_X} \leq \frac{5}{\sigma_X}\right) =$ Pero $\sigma_X = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{100}} = \frac{50}{10} = 5$
-

A.7.20- Suponga que la desviación estándar de la población es $\sigma = 25$. Calcule el error estándar de la media, $\sigma_{\bar{X}}$, con muestras de tamaño 50, 100, 150 y 200. ¿Qué puede decir acerca del error estándar de la media conforme el tamaño de la muestra aumenta?

Datos	Formalización de preguntas
▪ $\sigma = 25$ ▪ $n_1 = 50$ ▪ $n_2 = 100$ ▪ $n_3 = 150$ ▪ $n_4 = 200$	▪ $\sigma_{\bar{X}} = ?$ ▪ $n \to \infty \Rightarrow \sigma_{\bar{X}} \to ?$

Desarrollo y operaciones

- $P\left(\frac{-5}{5} \le Z \le \frac{5}{5}\right) =$
- $P(-1 \le Z \le 1) = 1 - 2P(Z < -1)$
- $1 - 2(0.1587) = 1 - 0.3174$
- $= 0.6826$

Resultados y soluciones

a) 0.6826

Desarrollo y operaciones: b)

- $P(|X - \mu| \le 10) =$
- $P(-10 \le X - \mu \le 10) =$
- $P\left(\frac{-10}{\sigma_X} \le \frac{X - \mu}{\sigma_X} \le \frac{10}{\sigma_X}\right) =$
- $P\left(\frac{-10}{5} \le Z \le \frac{10}{5}\right) =$
- $P(-2 \le Z \le 2) = 1 - 2P(Z < -2)$
- $1 - 2(0.0228) = 1 - 0.0456$
- $= 0.9544$

Resultados y soluciones

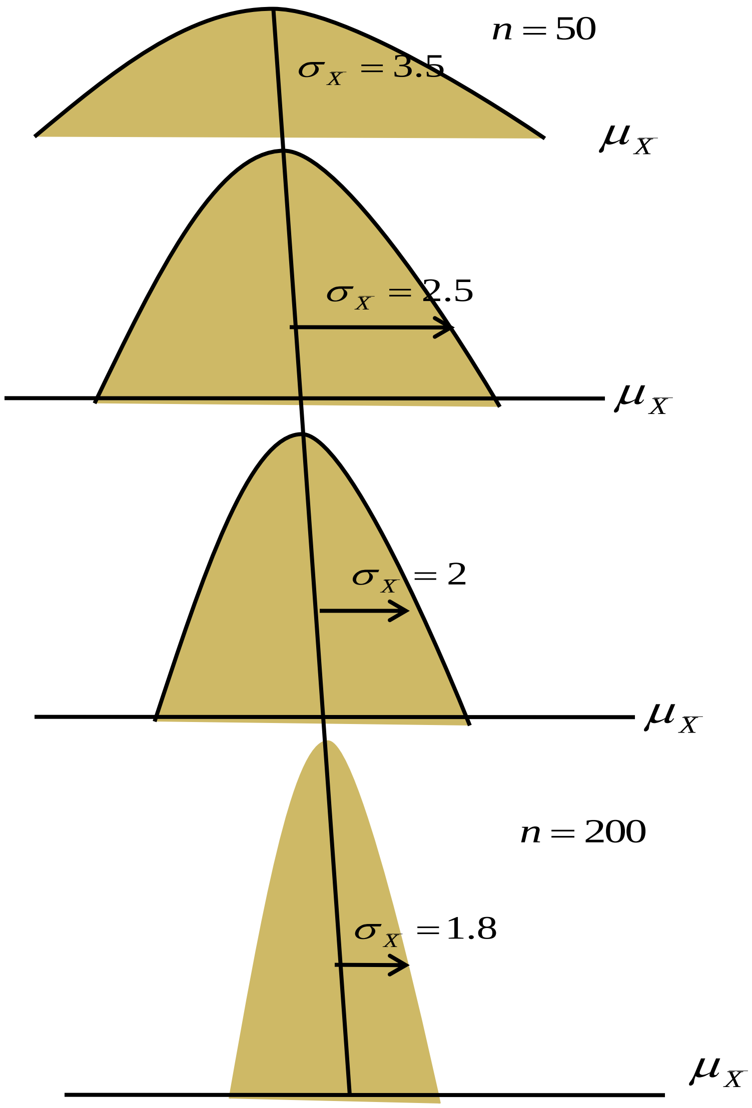
b) 0.9544

Teoría

El error estándar de la media mide la dispersión de las medias muestrales alrededor de la media poblacional. El error estándar relaciona la dispersión de la población con el tamaño de muestra y se define como:

.
$$\sigma_X = \frac{\sigma}{\sqrt{n}}$$
.

. De la fórmula se desprende que σ_X sólo depende del tamaño de muestra, n, ya que σ es una constante. Como n se encuentra en el denominador de la razón, $\frac{\sigma}{\sqrt{n}}$ conforme n se incrementa el denominador aumenta y la razón disminuye.



Desarrollo y operaciones: n=25

- $\sigma_X = \frac{\sigma}{\sqrt{n_1}}$
- $\sigma_X = \frac{25}{\sqrt{50}}$
- $\sigma_X = \frac{25}{7.071}$
- $\sigma_X = 3.54$

Desarrollo y operaciones: n=100

- $\sigma_X = \frac{25}{\sqrt{100}}$
- $\sigma_X = \frac{25}{10}$
- $\sigma_X = 2.5$

Desarrollo y operaciones: n=150

- $\sigma_X = \frac{25}{\sqrt{150}}$
- $\sigma_X = \frac{25}{12.25}$
- $\sigma_X = 2.0$

Desarrollo y operaciones: n=200

- $\sigma_X = \frac{25}{\sqrt{200}}$
- $\sigma_X = \frac{25}{14.14}$
- $\sigma_X = 1.8$

Resultados y soluciones

n	
50	3.5
100	2.5
150	2.0
200	1.8

Resultados y soluciones

De la tabla de resultados se verifica que conforme el tamaño de muestra se incrementa el error estándar disminuye.

A.7.21- Suponga que de una población en la que $\sigma = 25$ se toma una MAS de tamaño 50. Halle el valor del error estándar de la media en cada uno de los caso siguientes (si es necesario use el fcpf).
El tamaño de la población es infinito.
El tamaño de la población es N=50 000
El tamaño de la población es N=5 000
El tamaño de la población es N= 500

Datos	Formalización de preguntas
1) $\sigma = 25$ $n = 50$ $N \to \infty$ $N = 50000$ $N = 5000$ $N = 500$	$\sigma_X = ?$

A.7.21- Suponga que de una población en la que $\sigma = 25$ se toma una MAS de tamaño 50. Halle el valor del error estándar de la media en cada uno de los caso siguientes (si es necesario use el fcpf).

El tamaño de la población es infinito.

El tamaño de la población es N=50 000

El tamaño de la población es N=5 000

El tamaño de la población es N= 500

Datos	Formalización de preguntas
$\sigma = 25$ 1) $n = 50$ $N \to \infty$ $N = 50000$ $N = 5000$ $N = 500$	$\sigma_X = ?$

Teoría

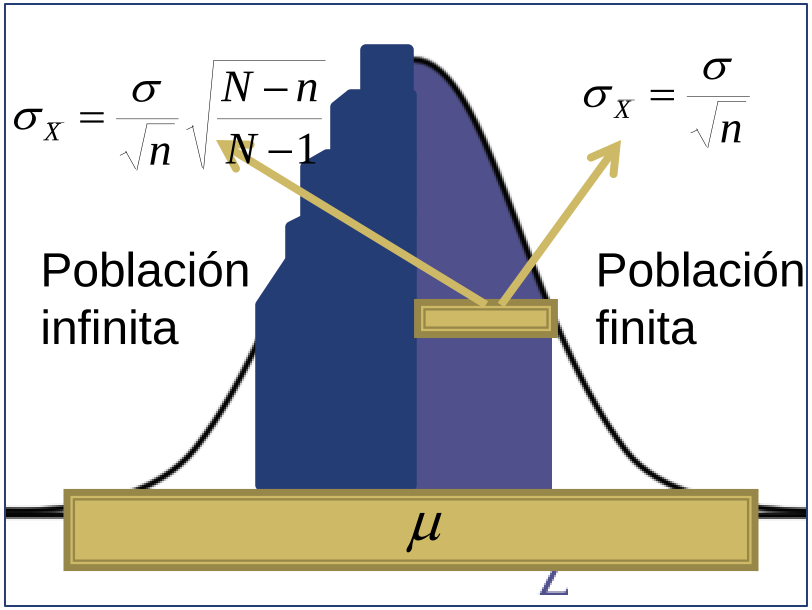
El error estándar de la media mide la dispersión de las medias muestrales alrededor de la media poblacional. El error estándar relaciona la dispersión de la población con el tamaño de muestra y se define como:

.

. $\sigma_X = \dfrac{\sigma}{\sqrt{n}}$

.

Al usar el MAS el error estándar de la distribución de medias depende de si la población es finita o infinita.
Cuando:
. 1) La población es infinita o



2) la población sea finita y el tamaño de muestra sea menor o igual al 5% del tamaño de la población, es decir $n/N \leq 0.05$

.

Se debe usar . $\sigma_X = \dfrac{\sigma}{\sqrt{n}}$

En caso contrario use:

$$\sigma_X = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Material didáctico: visual elaborado por Fidelmar Sandoval Durán

Desarrollo y operaciones: $N \to \infty$

$$\sigma_X = \frac{\sigma}{\sqrt{n}} \qquad \sigma_X = \frac{25}{7.07}$$

$$\sigma_X = \frac{25}{\sqrt{50}} \qquad \sigma_X = 3.54$$

$$\sigma_X = \frac{25}{\sqrt{50}}$$

Desarrollo y operaciones: $N = 5000$

$$f = \frac{n}{N} = \frac{50}{5000} = \frac{5}{500} = 0.01$$

$$f = 0.001 \leq 0.05 \Rightarrow \qquad \sigma_X = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_X = \frac{25}{\sqrt{50}}$$

$$\sigma_X = 3.54$$

Desarrollo y operaciones: $N = 50000$

$$f = \frac{n}{N} = \frac{50}{50000} = \frac{5}{5000} = 0.001$$

$$\sigma_X = \frac{\sigma}{\sqrt{n}}$$

$$f = 0.001 \leq 0.05 \Rightarrow$$

$$\sigma_X = \frac{25}{\sqrt{50}}$$

$$\sigma_X = 3.54$$

Desarrollo y operaciones: $N = 500$

$$f = \frac{n}{N} = \frac{50}{500} = \frac{5}{50} = 0.10$$

$$f = 0.10 > 0.05 \Rightarrow \sigma_X = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\sigma_X = \frac{25}{\sqrt{50}}\sqrt{\frac{500-50}{500-1}}$$

$$\sigma_X = (3.54)(0.9496) = 3.36$$

Resultados y soluciones		
N	$f=n/N$	
		3.54
50000	0.001	3.54
5000	0.01	3.54
500	0.10	3.36

Resultados y soluciones

De la tabla de resultados se observa que muestras relativamente grandes (cuando *f*, la fracción de muestreo es mayor que 0.05) aplica el factor de corrección para población finita (fcpf).

M.7.1- Una población consta de 500 unidades experimentales. Use una tabla de números aleatorios para seleccionar una muestra aleatoria de n= 20 unidades Experimentales. (Sugerencia: puesto que necesita usar números de tres dígitos, asigne dos números de tres dígitos a cada una de las unidades del muestreo. ¿Cuál es la probabilidad de que cada unidad experimental sea seleccionada para incluirla en la muestra?

Datos	Formalización de preguntas
. ▪ $N = 500$ ▪ $n = 20$	seleccionar una muestra aleatoria de 20 unidades: ¿Cuáles son las 20 unidades experimentales que se incluyen en la muestra? . $P(X_i) = ?$

Teoría
Si se selecciona una muestra de n elementos de una población de N elementos por medio de un plan de muestreo en el que cada una de las posibles muestras tiene la misma probabilidad de ser seleccionada, entonces se dice que el muestreo es aleatorio y la muestra resultante es una muestra aleatoria simple, MAS.

U.E	Números Asignados	
1	001	501
2	002	502
3	003	503
4	004	504
5	005	505
499	499	999
500	500	000

Procedimiento

A cada elemento le asignamos un numero (la población es de 500 elementos y requerimos tres dígitos: 001,002,…000).
Se genera una secuencia de 20 (tamaño de muestra) números aleatorios de tres dígitos. Se puede utilizar una tabla de números aleatorios o una calculadora que los genere o bien cualquier software estadístico.
Si usamos la tabla de números aleatorio, debemos seleccionar un punto aleatorio de partida y determinar la secuencia ya sea por fila o columna.

Desarrollo y operaciones:

.Usamos la tabla de números aleatorios de Mendenhall: tabla10 paginas 706-707. Esta tablea presenta grupos de dígitos de cinco en cinco
.Seleccionamos el punto de partida: columna1 y fila 1 y escogemos la fila para seleccionar la secuencia de 20 números.
En la intersección de la fila 1 columna 1 el numero aleatorio que se observa es 10480, este es el punto de partida para seccionar en fila la secuencia:
.10480 15011 01536 02011 81647 91646 69179 14194 62590 …
.
la secuencia de números de tres dígitos es: 104, 801, 501, 101, 536, 020, 118, 164, 791, 646, 691, 791, 419, 462, 590,…
. De estos escogemos los 20 que están identificados para cada elemento que debemos incluir en la muestra: 104, 301 (801-500), 001(501-500), 101, 036 (536-500), 020, 118, 164, 291(791-500),419, 462, 090(590-500), etc.,

20 elementos que debe incluirse en la muestra	
104	191
301	419
001	446
101	259
036	303
020	120
118	220
164	469
291	495
146	209

20 elementos que debe incluirse en la muestra (ordenados)	
001	209
020	220
036	259
101	291
104	301
118	303
120	419
146	446
164	469
191	495

Resultados y soluciones
$P(X_i) = \frac{1}{N} = \frac{1}{500} = 0.0020$

20 elementos que debe incluirse en la muestra	
104	191
301	419
001	446
101	259
036	303
020	120
118	220
164	469
291	495
146	209

20 elementos que debe incluirse en la muestra (ordenados)	
001	209
020	220
036	259
101	291
104	301
118	303
120	419
146	446
164	469
191	495

Resultados y soluciones
$P(X_i) = \frac{1}{N} = \frac{1}{500} = 0.0020$

M.7.5- Cada decima persona. Se obtuvo una muestra aleatoria de opinión publica en un pequeño pueblo, al seleccionar a la decima persona que pasará por la esquina más concurrida en el área del centro. Esta muestra tendrá las características de una muestra aleatoria de los ciudadanos del pueblo? Explique.

Resultados y soluciones
No; Se desconoce la probabilidad de cada elemento y la muestra puede ser no representativa

M. 7.11- Imágenes de resonancia magnética. En un estudio descrito en *American Journal of sports Medicine*, Peter D. Franklin y colaboradores Informaron acerca de la precisión de usar imágenes de resonancia magnética. (IRM) para evaluar las torceduras y desgarres de ligamentos en 35 pacientes De la consulta clínica de uno de los autores se selecciona a pacientes. consecutivos con dolor agudo o crónico de la rodilla quienes estuvieron de acuerdo en participar en el estudio.

a. Describa el plan de muestreo que se utilizó para seleccionar a los participantes en el estudio.
b ¿Qué mecanismo de probabilidad se empleó para seleccionar esta muestra de 35 Individuos con dolor en la rodilla?
c ¿Es posible hacer inferencias validas con los resultados de este estudio? ¿Por qué sí o por qué no?
d Diseñe otro plan de muestreo. ¿Qué cambiaría?

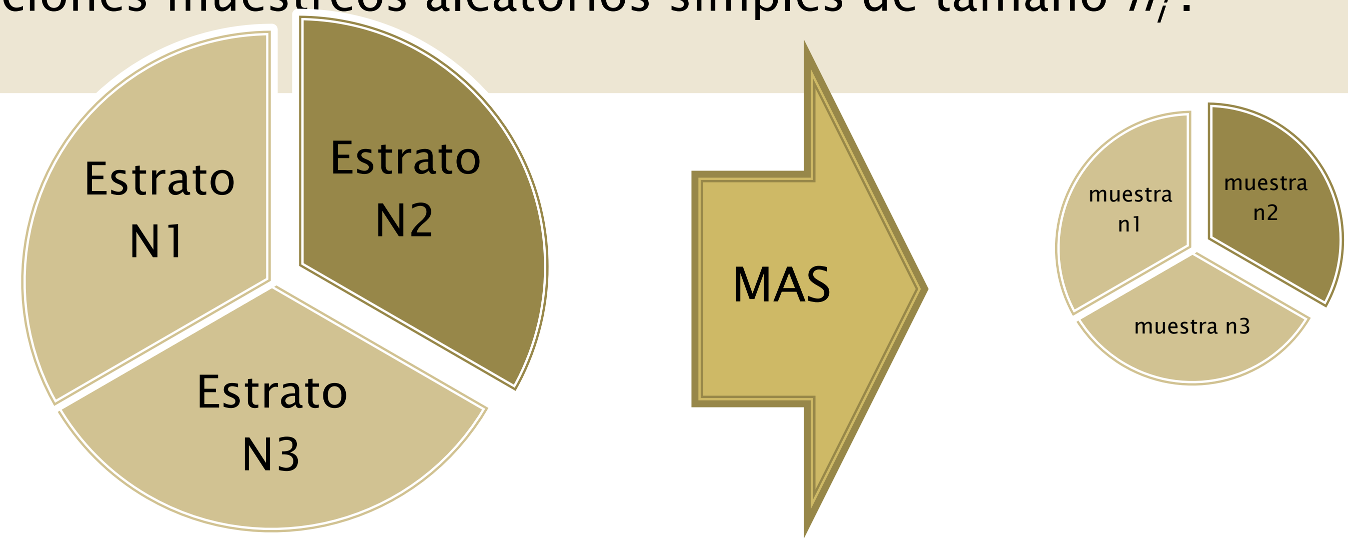
6

Datos	Formalización de preguntas
.n=35 . .	.¿Cuál es el plan de muestreo? . ¿Qué mecanismo de probabilidad se usó? .¿Es posible hacer inferencias validas? .Con esta información qué otro plan de Muestreo se puede efectuar.

Teoría

Los planes de muestreo aleatorios proporcionan una base probabilística para hacer inferencias. Los tres planes son: estratificado, por racimos y sistemático.

El muestreo aleatorio estratificado, MAE requiere seleccionar una MAS de cada una de las varias subpoblaciones a estratos en los que se divide la población de N individuos, atendiendo a criterios que puedan ser importantes en el estudio y realizando en cada una de estas subpoblaciones muestreos aleatorios simples de tamaño n_i .



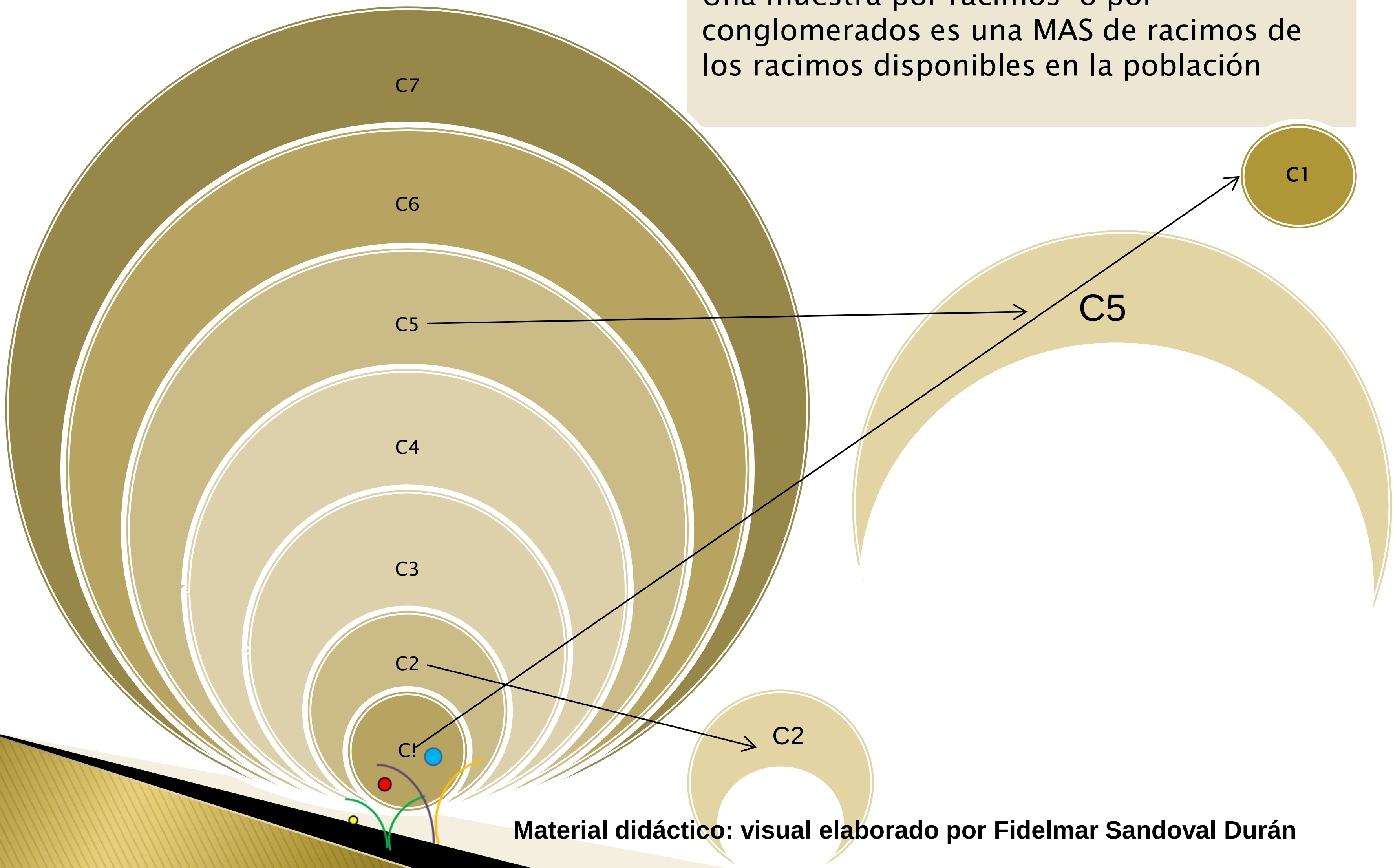
Universidad Autónoma de Estado de México

Facultad de Economía

Licenciatura en Economía

Teoría

Una muestra por racimos o por conglomerados es una MAS de racimos de los racimos disponibles en la población



Material didáctico: visual elaborado por Fidelmar Sandoval Durán

Teoría

Una muestra aleatoria Sistemática 1 en K implica la selecciona aleatoria de uno de los primeros k elementos de una población ordenada y, luego, la selección sistemática de cada k-èsimo elemento.

En un **muestreo aleatorio sistemático** se elige un individuo al azar y a partir de él, a intervalos constantes, se eligen los demás hasta completar la muestra.

Por ejemplo si tenemos una población formada por 100 elementos y queremos extraer una muestra de 25 elementos, en primer lugar debemos establecer el intervalo de selección k, que será igual a 100/25 = 4. A continuación elegimos el elemento de arranque, tomando aleatoriamente un número entre el 1 y el 4, y a partir de él obtenemos los restantes elementos de la muestra: 2, 6, 10, 14,..., 98

Teoría	Teoría
Muestreo de conveniencia. 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 muestreos 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.	Presenta casi siempre sesgos y por tanto debe aplicarse únicamente cuando no existe alternativa. También puede ser útil cuando se pretende realizar una primera prospección de la población o cuando no existe un marco de la encuesta definido. Este tipo de muestreos puede incluir individuos próximos a la media o no, pero casi nunca representará **la variabilidad de la población, que normalmente quedará subestimada**

Teoría
El muestreo por juicio implica decidir quien será o no incluido en la muestra. **En el muestreo por cuota** la muestra debe reflejar la composición de la población con respecto a alguna característica preseleccionada, a menudo tiene un componente no aleatorio en el proceso de selección.

Resultados y soluciones

El plan de muestreo es por conveniencia. Ninguno mecanismo de probabilidad es usado y en consecuencia es imposible realizar indiferencias. Con la información proporcionada sólo es posible el muestreo “accidental” por que nos conviene.

N.7.5- Dada una población de media $\mu = 100$ y varianza $\sigma^2 = 81$ el limite central se aplica cuando el tamaño de la muestra es $n \geq 25$. Se obtiene una muestra aleatoria de tamaño $n = 25$.

a) ¿Cuál son la media y la varianza de la distribución de las medias muestrales en el muestreo?
b) ¿Cuál es la probabilidad de que $\bar{X} > 102$?
c) ¿Cuál es la probabilidad de que $98 \leq \bar{X} \leq 101$?
d) ¿Cuál es la probabilidad de que $\bar{X} \leq 101.5$?

Datos	Formalización de preguntas
• $\mu = 100$ • $\sigma^2 = 81$ • $n = 25$	• $E(\bar{X}) = \mu_{\bar{X}} = ?$ • $V(\bar{X}) = \sigma^2_{\bar{X}} = ?$ • $P(\bar{X} > 102) = ?$ • $P(98 \leq \bar{X} \leq 101) = ?$ • $P(\bar{X} \leq 101.5) = ?$

Teoría

Teorema central del límite (T.C.L.). El teorema establece que la media de una muestra aleatoria, extraída de una población que tiene cualquier distribución de probabilidad, sigue aproximadamente una distribución normal que tiene una media μ y una varianza σ^2/n dado un tamaño de muestra suficientemente grande.

Desarrollo y operaciones:

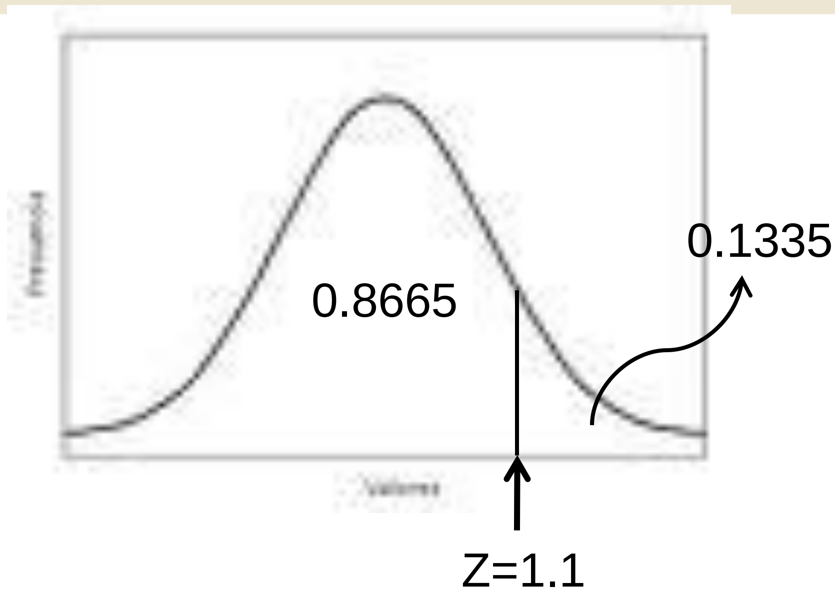
$$E(\bar{X}) = \mu_{\bar{X}} = \mu = 100$$

$$V(\bar{X}) = \sigma^2_{\bar{X}} = \sigma^2/n = 81/25 = 3.24$$

$$P(\bar{X} > 102) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{102-100}{\sqrt{3.24}}\right) =$$

$$= P(Z > 1.11) = 1 - F(z = 1.11) =$$

$$= 1 - 0.8665 = 0.1335$$



Desarrollo y operaciones:

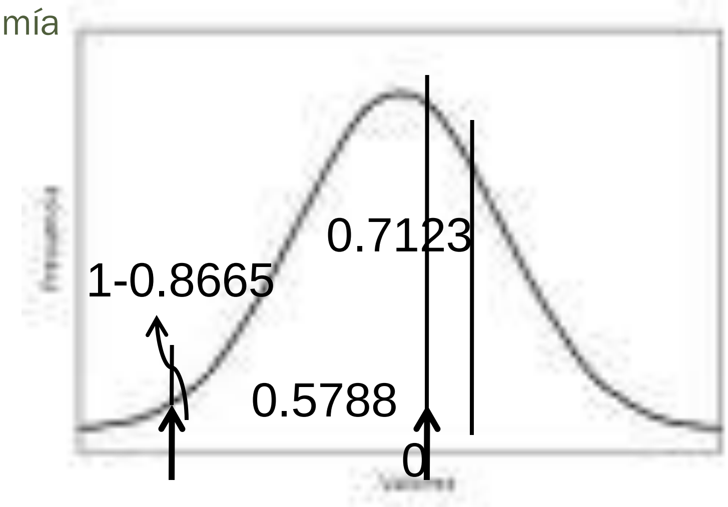
$$P(98 \le \bar{X} \le 101) = P\left(\frac{98-\mu}{\sigma/\sqrt{n}} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{101-\mu}{\sigma/\sqrt{n}}\right) =$$

$$= P\left(\frac{98-100}{9/\sqrt{25}} \le Z \le \frac{101-100}{9/\sqrt{25}}\right) =$$

$$= P(-1.11 \le Z \le 0.56) =$$

$$= F(z=0.56) - (1 - F(z=1.11)) =$$

$$= 0.7123 - (1 - 0.8665) = 0.5788$$



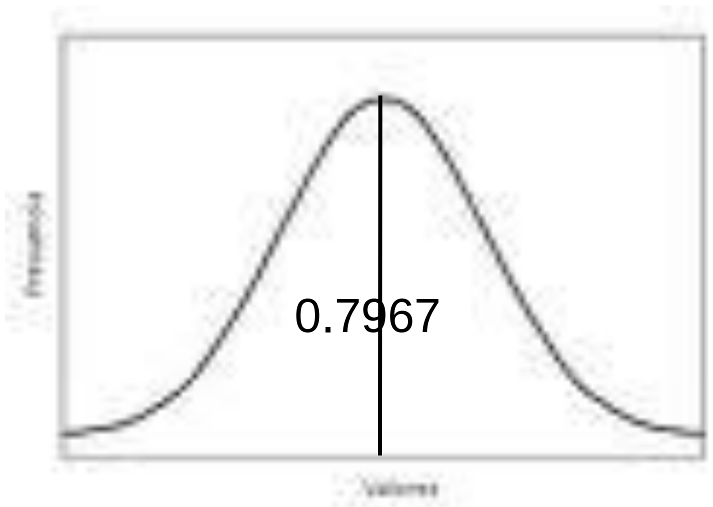
Z=-1.11 Z=0.56

Desarrollo y operaciones:

$$P(\bar{X} \le 101.5) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{10.5-\mu}{\sigma/\sqrt{n}}\right) =$$

$$= P\left(Z \le \frac{101.5-100}{9/\sqrt{25}}\right) = P(Z \le 0.83) =$$

$$= F(z=0.83) = 0.7967$$



Z=0.83

Material didáctico: visual elaborado por Fidelmar Sandoval Durán

N.7.6- Dada una población de media $\mu = 100$ y varianza $\sigma^2 = 900$ el limite central se aplica cuando el tamaño de la muestra es $n \geq 25$. Se obtiene una muestra aleatoria de tamaño $n = 30$.

a) ¿Cuál son la media y la varianza de la distribución de las medias muestrales en el muestreo?

b) ¿Cuál es la probabilidad de que $\bar{X} > 109$?

c) ¿Cuál es la probabilidad de que $96 \leq \bar{X} \leq 110$?

d) ¿Cuál es la probabilidad de que $\bar{X} \leq 107$?

Datos	Formalización de preguntas
• $\mu = 100$ • $\sigma^2 = 900$ • $n = 30$	• $E(\bar{X}) = \mu_{\bar{X}} = ?$ • $V(\bar{X}) = \sigma^2_{\bar{X}} = ?$ • $P(\bar{X} > 109) = ?$ • $P(96 \leq \bar{X} \leq 110) = ?$ • $P(\bar{X} \leq 107) = ?$

Teoría

Teorema central del límite (T.C.L.). El teorema establece que la media de una muestra aleatoria, extraída de una población que tiene cualquier distribución de probabilidad, sigue aproximadamente una distribución normal que tiene una media y una varianza dado un tamaño de muestra suficientemente grande.

Desarrollo y operaciones:

$$E(\bar{X}) = \mu_{\bar{X}} = \mu = 100$$

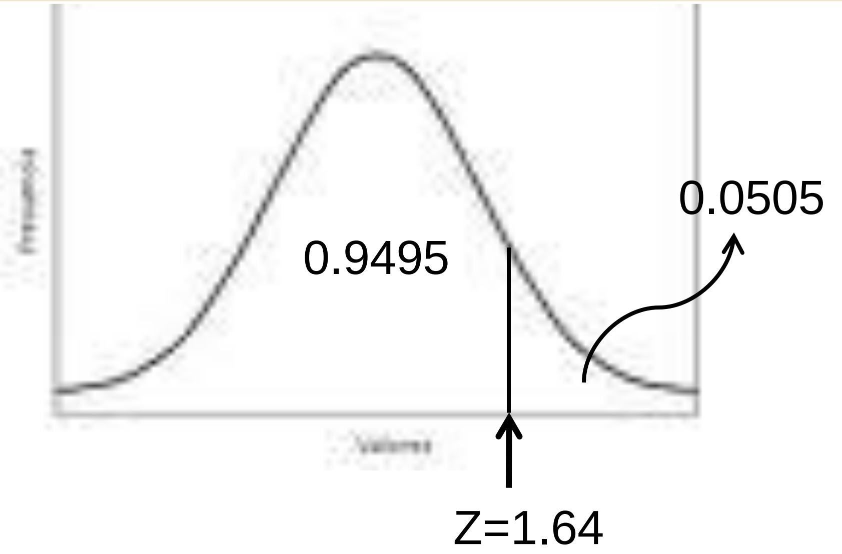
$$V(\bar{X}) = \sigma^2_{\bar{X}} = \sigma^2 / n = 900/30 = 30$$

$$\Rightarrow \sigma = \sqrt{V(\bar{X})} = \sqrt{30} = 5.48$$

$$P(\bar{X} > 109) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{109 - 100}{5.48}\right) =$$

$$= P(Z > 1.64) = 1 - F(z = 1.64) =$$

$$= 1 - 0.9495 = 0.0505$$



Desarrollo y operaciones:

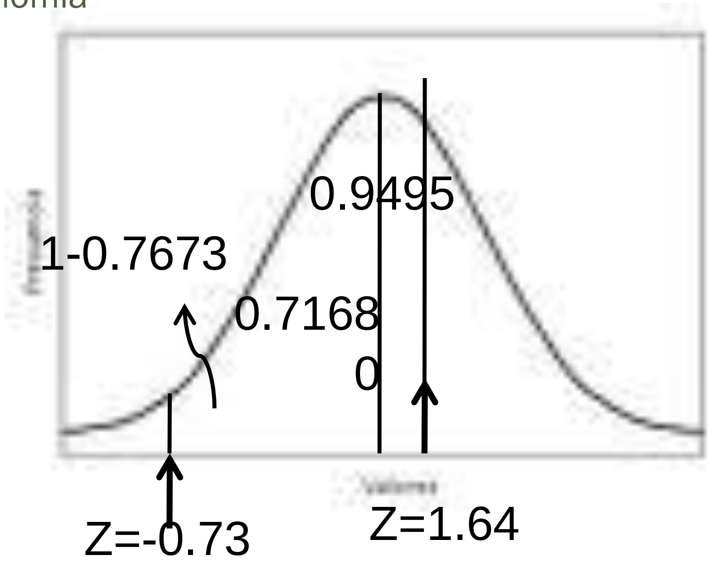
$$P(96 \le \bar{X} \le 109) = P\left(\frac{96-\mu}{\sigma/\sqrt{n}} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{109-\mu}{\sigma/\sqrt{n}}\right) =$$

$$= P\left(\frac{96-100}{5.48} \le Z \le \frac{109-100}{5.48}\right) =$$

$$= P(-0.73 \le Z \le 1.64) =$$

$$= F(z=1.64) - (1 - F(z=0.73)) =$$

$$= 0.9495 - (1-0.7673) = 0.7168$$

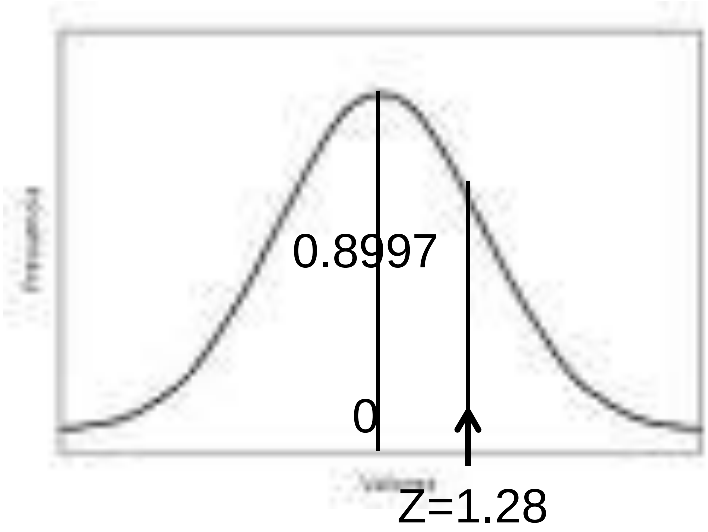


Desarrollo y operaciones:

$$P(\bar{X} \le 107) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{107-\mu}{\sigma/\sqrt{n}}\right) =$$

$$= P\left(Z \le \frac{107-100}{5.48}\right) = P(Z \le 1.28) =$$

$$= F(z=1.28) = 0.8997$$



N.7.7- Dada una población de media $\mu = 200$ y varianza $\sigma^2 = 625$ el limite central se aplica cuando el tamaño de la muestra es $n \geq 25$. Se obtiene una muestra aleatoria de tamaño $n = 25$.

a) ¿Cuál son la media y la varianza de la distribución de las medias muestrales en el muestreo?

b) ¿Cuál es la probabilidad de que $\bar{X} > 209$?

c) ¿Cuál es la probabilidad de que $198 \leq \bar{X} \leq 211$?

d) ¿Cuál es la probabilidad de que $\bar{X} \leq 202$?

Datos	Formalización de preguntas
$\mu = 200$	$E(\bar{X}) = \mu_{\bar{X}} = ?$
$\sigma^2 = 625$	$V(\bar{X}) = \sigma^2_{\bar{X}} = ?$
$n = 25$	$P(\bar{X} > 209) = ?$
	$P(198 \leq \bar{X} \leq 211) = ?$
	$P(\bar{X} \leq 202) = ?$

Teoría

Teorema central del límite (T.C.L.). El teorema establece que la media de una muestra aleatoria, extraída de una población que tiene cualquier distribución de probabilidad, sigue aproximadamente una distribución normal que tiene una media μ y una varianza σ^2/n dado un tamaño de muestra suficientemente grande.

Desarrollo y operaciones:

$$E(\bar{X}) = \mu_{\bar{X}} = \mu = 200$$

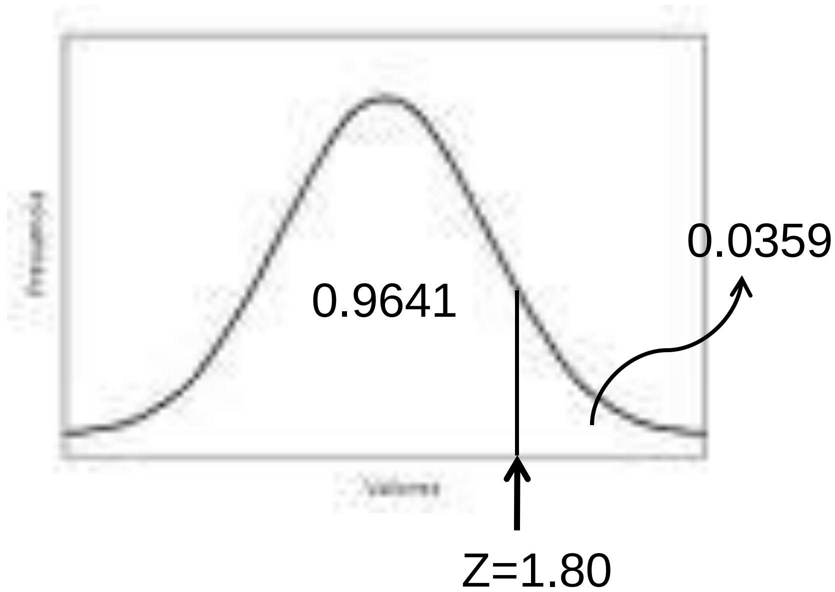
$$V(\bar{X}) = \sigma^2_{\bar{X}} = \sigma^2/n = 625/25 = 25$$

$$\Rightarrow \sigma = \sqrt{V(\bar{X})} = \sqrt{25} = 5.00$$

$$P(\bar{X} > 209) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{209-200}{5.00}\right) =$$

$$= P(Z > 1.80) = 1 - F(z = 1.80) =$$

$$= 1 - 0.9641 = 0.0359$$



Desarrollo y operaciones:

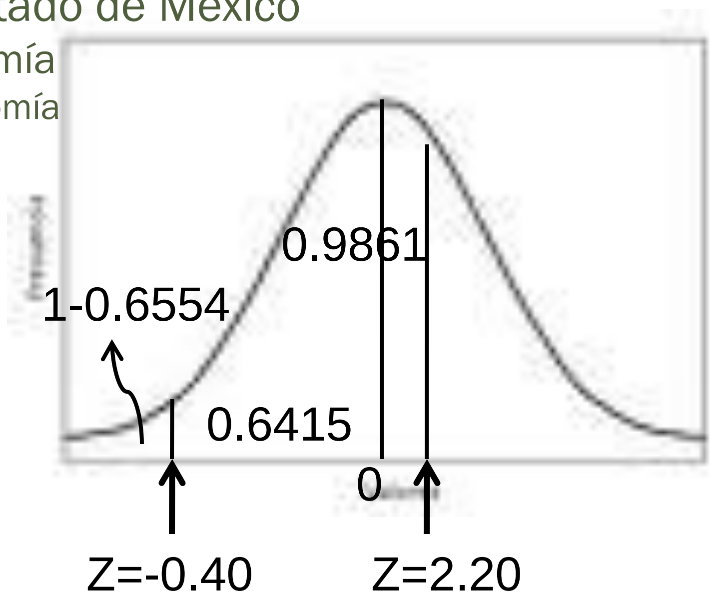
$$P(196 \le \bar{X} \le 211) = P\left(\frac{198-\mu}{\sigma/\sqrt{n}} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{211-\mu}{\sigma/\sqrt{n}}\right) =$$

$$= P\left(\frac{198-200}{5.00} \le Z \le \frac{211-200}{5.00}\right) =$$

$$= P(-0.40 \le Z \le 2.20) =$$

$$= F(z = 2.20) - (1 - F(z = 0.40)) =$$

$$= 0.9861 - (1 - 0.6554) = 0.6415$$

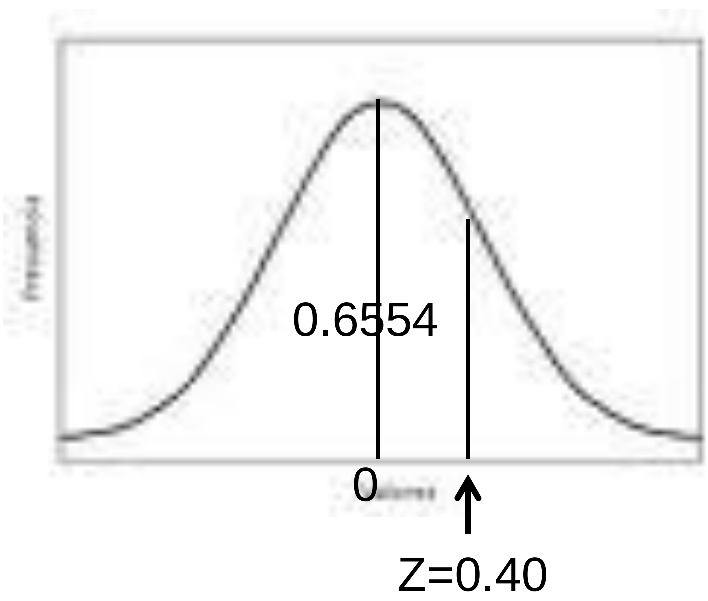


Desarrollo y operaciones:

$$P(\bar{X} \le 202) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{202-\mu}{\sigma/\sqrt{n}}\right) =$$

$$= P\left(Z \le \frac{202-200}{5.00}\right) = P(Z \le 0.40) =$$

$$= F(z = 0.40) = 0.6554$$



Batería 2

A.7.50 – cerca de 28% de las empresas tiene como propietario una mujer. Responda estas preguntas con base en una muestra de 240 empresas.

a) Muestre la distribución muestral de $\hat{p}$, la proporción muestral de las empresas propiedad de una mujer.

b) ¿Cuál es la probabilidad de que la proporción muestral esté a no más de ± 0.04 de la proporción poblacional.

c) ¿Cuál es la probabilidad de que la proporción muestral esté a no más de ± 0.02 de la proporción poblacional.

Datos	Formalización de preguntas
. P=0.28 . n=240 .	.Distribución muestral de proporciones . $P(\hat{p} \pm 0.04) = ?$ . . $P(\hat{p} \pm 0.02) = ?$

10

Teoría

Teorema central del límite (T.C.L.). El teorema establece que la proporción de una muestra aleatoria, extraída de una población que tiene distribución no normal de probabilidad, sigue aproximadamente una distribución muestral normal que tiene una media P y una desviación estándar $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ dado un tamaño de muestra suficientemente grande.

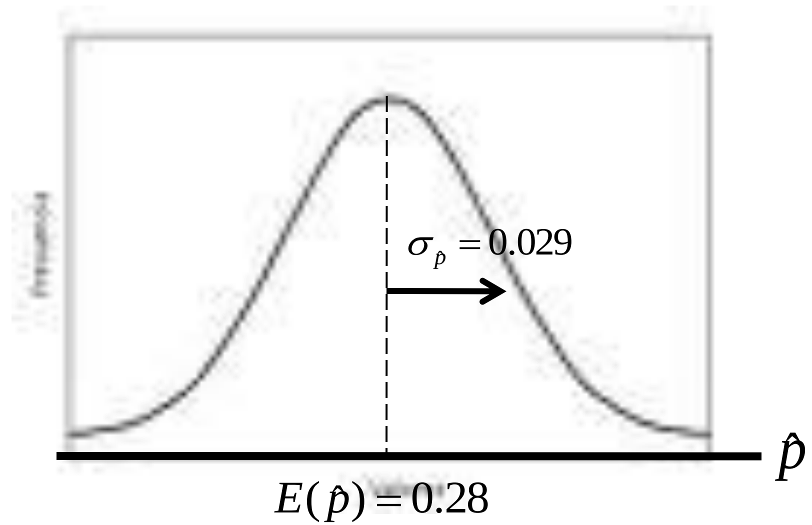
Una distribución Binomial se aproxima mediante una distribución normal siempre que la muestra sea suficientemente grande para satisfacer las dos condiciones siguientes : $n\hat{p} \geq 5 y n(1-\hat{p}) \geq 5$

Desarrollo y operaciones:

.Muestra grande n=240

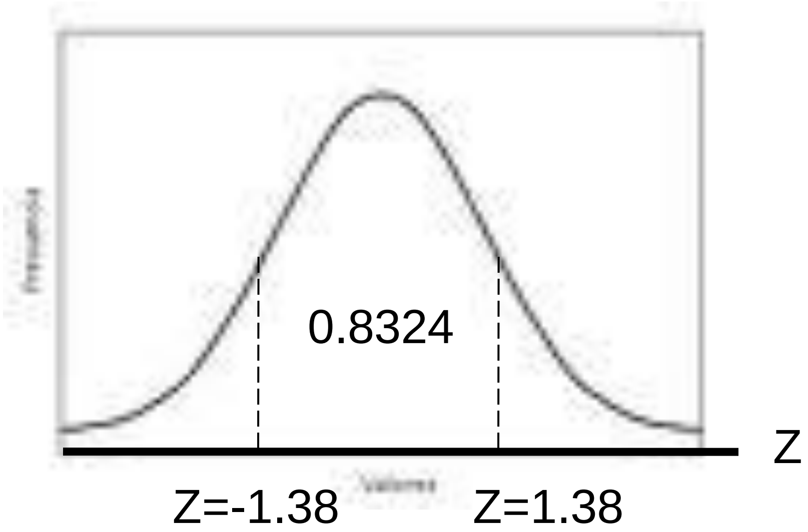
$$n\hat{p} = (240)(0.28)$$

- $n\hat{p} = 67.2$
- $n(1-\hat{p}) = (240)(0.72)$
- $n(1-\hat{p}) = 172.8$
- $E(\hat{p}) = 0.28$
- $\sigma_{\hat{p}} = \sqrt{\dfrac{(0.28)(0.72)}{240}}$
- $\sigma_{\hat{p}} = 0.029$



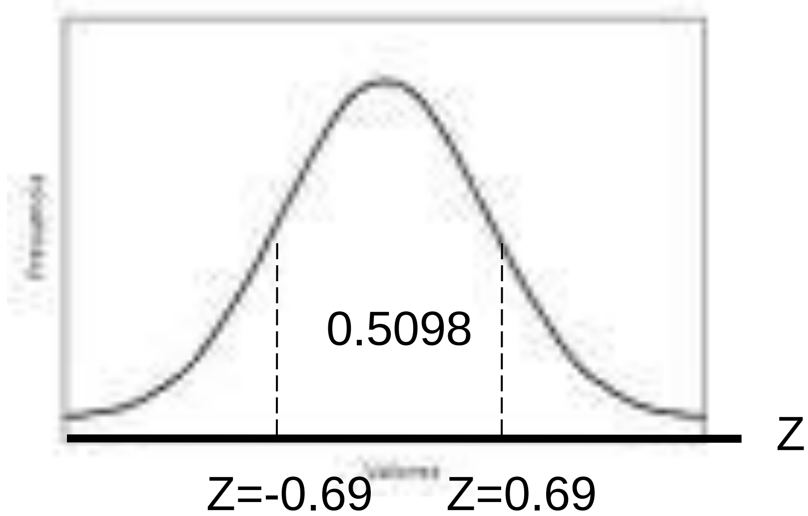
Desarrollo y operaciones:

- $\hat{p} \pm ME$
- $\therefore ME = Z\sigma_{\hat{p}}$
- $Z = \dfrac{ME}{\sigma_{\hat{p}}} = \dfrac{0.04}{0.029} = 1.38$
- $P(-1.38 < Z < 1.38) =$
- $F(Z = 1.38) - (1 - F(-1.38)) =$
- $0.9162 - 0.0838 = 0.8324$



Desarrollo y operaciones:

- $\hat{p} \pm ME$
- $\therefore ME = Z\sigma_{\hat{p}}$
- $Z = \dfrac{ME}{\sigma_{\hat{p}}} = \dfrac{0.02}{0.029} = 0.69$
- $P(-0.69 < Z < 0.69) =$
- $F(Z = 0.69) - (1 - F(-0.69)) =$
- $0.7549 - 0.2451 = 0.5098$



A.7.51 – Una empresa de investigación de mercados realiza encuestas telefónicas con una tasa de respuesta de 40% de acuerdo con la experiencia. ¿Cuál es la probabilidad de que en una muestra de 400 números telefónicos 150 personas cooperen y respondan las preguntas? En otras palabras, ¿cuál es la probabilidad de que la proporción muestral sea al menos de 150/400=0.375?

11

Datos	Formalización de preguntas
. P=0.40 . n=400 X=150 . $\hat{p} = 0.375$	$P(\hat{p} \geq 0.375) = ?$

Teoría

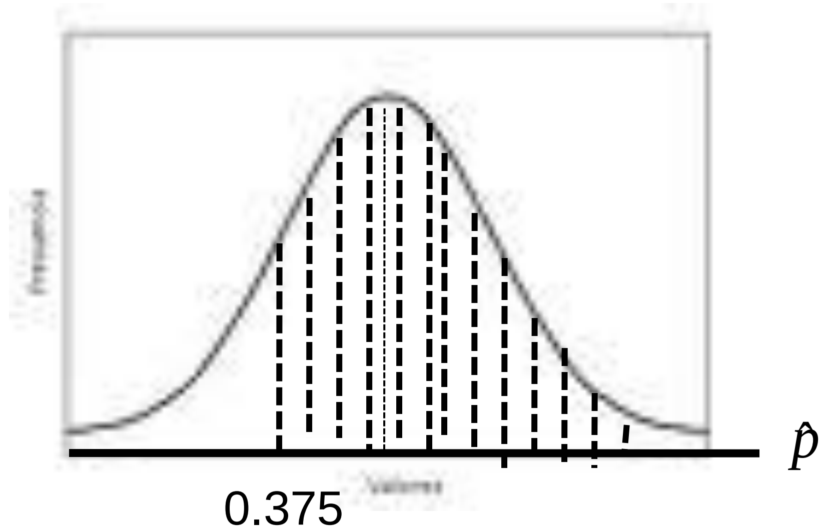
Teorema central del límite (T.C.L.). El teorema establece que la proporción de una muestra aleatoria, extraída de una población que tiene distribución no normal de probabilidad, sigue aproximadamente una distribución muestral normal que tiene una media P y una desviación estándar $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ dado un tamaño de muestra suficientemente grande.

Una distribución Binomial se aproxima mediante una distribución normal siempre que la muestra sea suficientemente grande para satisfacer las dos condiciones siguientes: $n\hat{p} \geq 5 y n(1-\hat{p}) \geq 5$

Desarrollo y operaciones:

.Muestra grande n=400

- $n\hat{p} = (400)(0.375)$
- $n\hat{p} = 150$
- $n(1-\hat{p}) = (400)(0.625)$
- $n(1-\hat{p}) = 250$
- $E(\hat{p}) = 0.375$
- $\sigma_{\hat{p}} = \sqrt{\frac{(0.375)(0.625)}{400}}$
- $\sigma_{\hat{p}} = 0.024$



Material didáctico: visual elaborado por Fidelmar Sandoval Durán

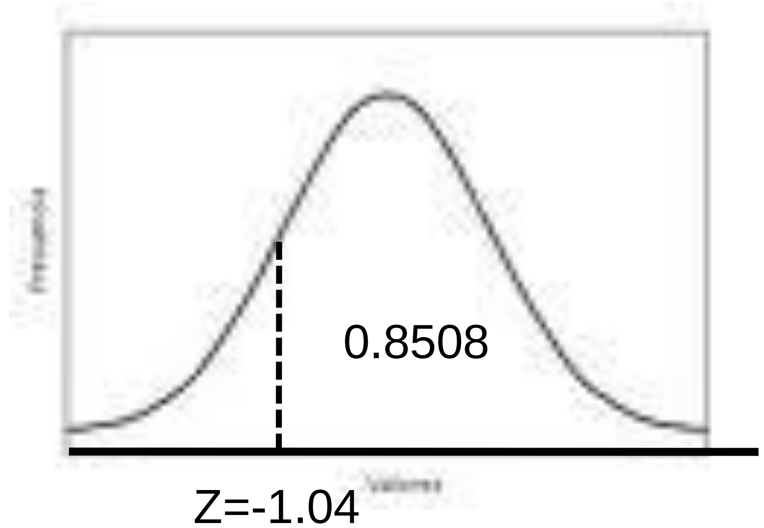
Desarrollo y operaciones:

$$P\left(\frac{\hat{p}-P}{\sigma_{\hat{p}}} \geq \frac{0.375-0.400}{0.024}\right)=$$

$$P(Z \geq -1.04)=$$

$$1-F(Z=-1.04)$$

$$1-0.1492=0.8508$$



$$N = 5; n = 3; N_n = 10$$

No de muestra	Elementos en muestra	Suma	Mediana	Media	varianza	Desviación estándar
1	1,2,3	6	2	2.0	1	1
2	1,2,4	7	2	2.3	2.3	1.5
3	1,2,5	8	2	2.7	4.3	2.1
4	1,3,4	8	3	2.7	2.3	1.5
5	1,3,5	9	3	3.0	4	2.0
6	1,4,5	10	4	3.3	4.3	2.1
7	2,3,4	9	3	3.0	1	1
8	2,3,5	10	3	3.3	2.3	1.5
9	2,4,5	11	4	3.7	2.3	1.5
10	3,4,5	12	4	4.0	1	1
				3.0	0.37	

Ahora, vamos obtener los llamados **Estadísticos de la Muestra**, que constituyen **variables muestrales** con una determinada **distribución muestral (**la media muestral, la varianza muestral, la proporción muestral, etc.,).

Cuando usamos estos **Estadísticos Muestrales** para estimar los parámetros poblacionales; lo fundamentamos teóricamente en la Ley de Grandes Números y el Teorema del Límite Central, dos grandes teoremas de la Estadística.

A continuación, usamos Minitab

Vamos a generar 100 valores en la columna C1, de 1 a 100, que indicará el i-ésimo elemento de la población, de tamaño 100; es decir N = 100.

Pasos:

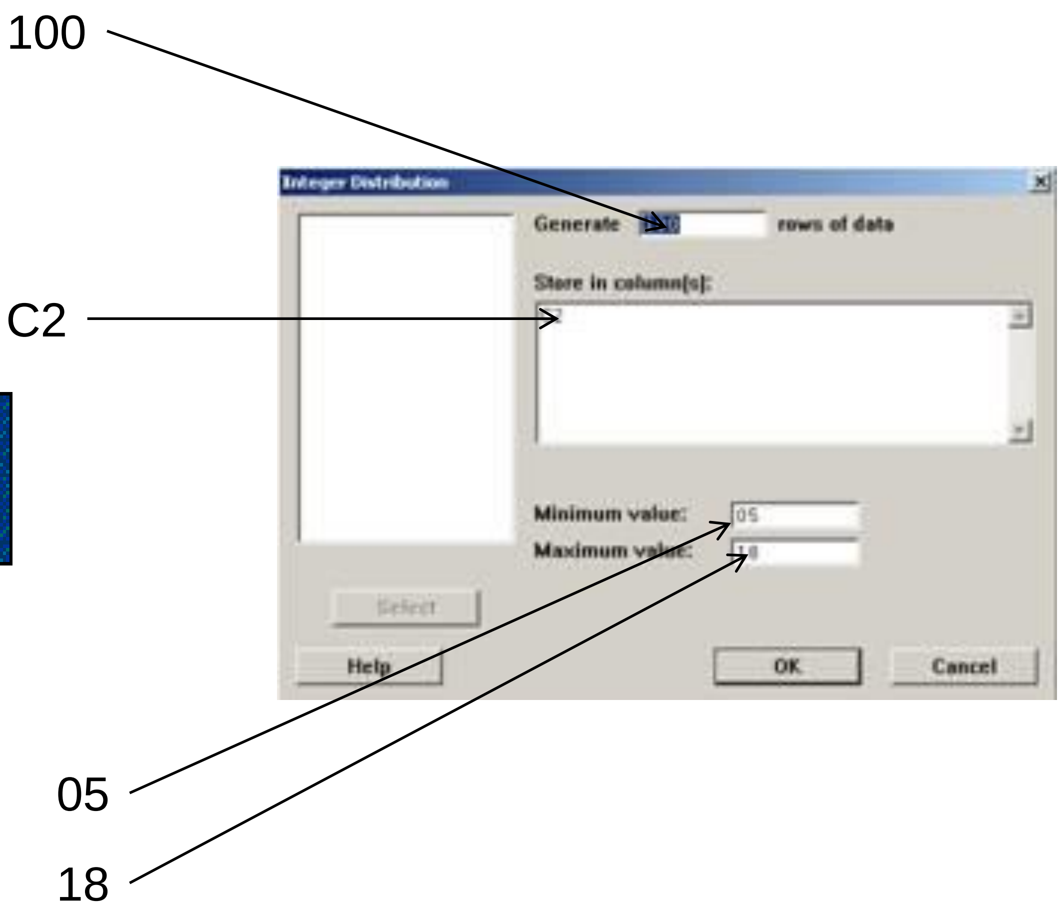
Paso 1: **/Calc/Make patterned data/Simple of set numbers/**

Valor inicial: 1
Valor final: 100.
Incrementos: 1.

Paso 2: En la columna C2 vamos a generar aleatoriamente las notas vigesimales de 100 alumnos de un determinado colegio. Supongamos que las notas van de 05 a 18.

Secuencia:/Calc/Random data/Integer/

En la ventana debemos ingresar los datos



Primera muestra para la columna C3 (Obs 01): La variable está en la columna C2 ó X. La almacenaremos en C3 ú Obs 01.

Descriptive Statistics

```
Variable          N      Mean    Median    TrMean     StDev   SE Mean
X               100    11.920    13.000    11.967     3.892     0.389

MTB > Describe 'Media de muestra';

Variable          N      Mean    Median    TrMean     StDev   SE Mean
Media de         36    11.567    11.700    11.588     1.641     0.273
```

t 3 ***

C1	C2	C3	C4	C5	C6	C7	C8	C9-T	C10
i	X	Obs 01	Obs 02	Obs 03	Obs 04	Obs 05	Media de muestra		
	12	14	18	14	11	14	14.2		*
2	18	12	10	16	11	5	10.8	Media poblacional	11.9200
3	18	18	11	11	14	14	13.6	Media muestral	11.5667
4	11	18	15	6	14	17	14.0		*
5	14	8	13	18	16	8	12.6	Desv. Est. Poblc.	3.8917
6	15	6	14	16	13	9	11.6	Desv. Est. Muest.	1.6406

Secuencia: /Calc/Random data/Sample from columns/Sample 36 rows/

Las otras muestras: Repetir para las otras columnas,C4,C5,C6 y C7, de la misma forma (parte de lo cual se muestra en la figura anterior).

Paso 3: Obtención de la media y desviación estándar de la muestra:

Secuencia:/Stat/Basic statistics/Display descriptive statistics/ La media, 11.92 y la desviación estándar, 3.892, se muestra en las primeras líneas de la salida

Paso 4: Ahora vamos a extraer 5 muestras aleatorias de tamaño 36 que las almacenaremos en las columnas C3 – C7:

Paso 5: Calcular la media de cada una de estas muestras:

Las medias por fila, la que se dispone en la columna C8

Las medias por columna (que no se ve) son: 11.0278, 12.5278, 12.2222, 11.2222 y 10.8333, para las 5 muestras.

Paso 6: Calcular la **media muestral de medias muestrales:** este valor es 11.5667, (ver celda C10(3) y en la hoja de trabajo, encerrada en una elipse.

$N_n = 4$

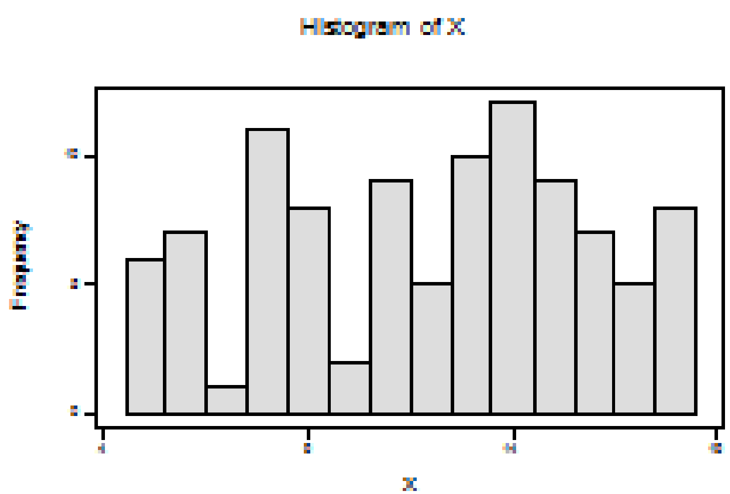
Si tomamos $N_n = 5$ MAS, n = 36, la media muestral de medias muestrales es 11.5667 y está muy cerca de la media poblacional, 11.9200. Del mismo modo, mue de n = 36, encontramos que su media 11.5667 es la misma que la media de las medias muestrales de tamaño 36 y muy cerca de la media poblacional.

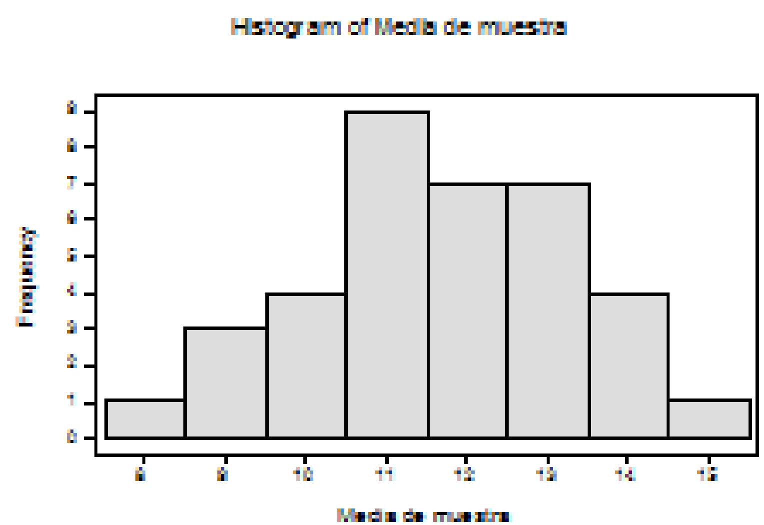
Esto nos dice que la media de medias muestrales podría ser tomada como un buen estadístico capaz de ser considerado como valor de la media poblacional. Como un dato adicional, la desviación poblacional dividida entre la raíz cuadrada del tamaño de la muestra, n = 5, que es igual a 1.74042, nos da un valor muy cercano a la desviación estándar de la media de las medias muestrales.

$\mu = 11.5667$ y $\sigma_X = 1.6406$

Finalmente, la siguiente figura contiene la gráfica de histogramas de los datos poblacionales y la gráfica de las medias muestrales.

Observe que si n = 36, la gráfica (de la derecha) nos da la forma de la campana de Gauss indicándonos que LA DISTRIBUCIÓN MUESTRAL (eso es lo que representa la columna C8) de las medias muestrales tiene una distribución aproximadamente normal, con parámetros

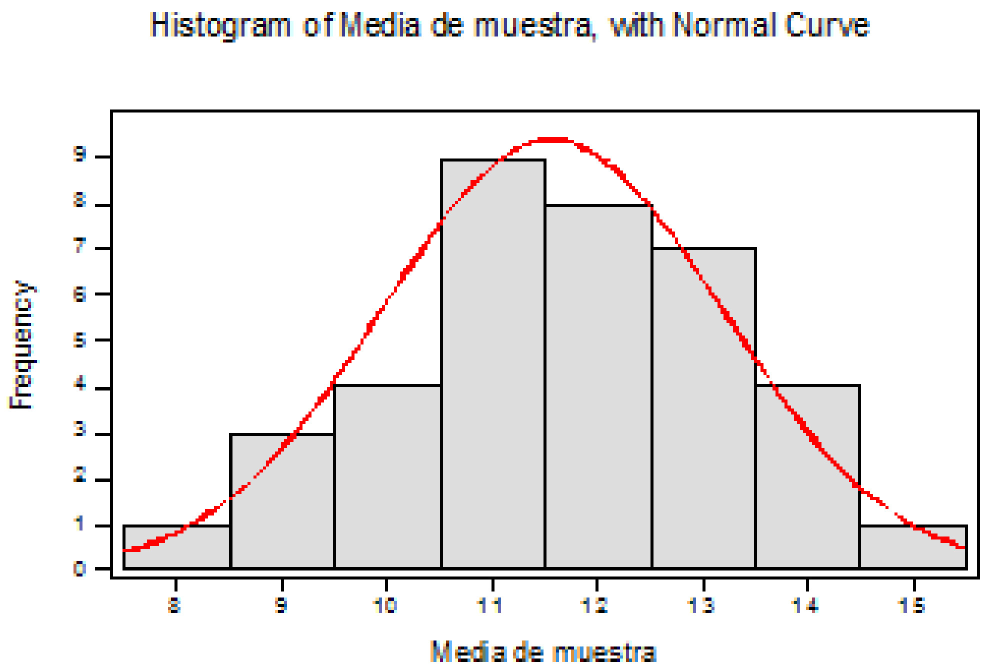




Nota:
Como el valor de la media de cada muestra puede ser considerada como un valor de la media de las medias muestrales, podemos considerar a la media de las medias muestrales como una variable aleatoria, de allí su nombre: Variable aleatoria definida como la Media muestral de medias muestrales.

La siguiente gráfica prueba nuestra afirmación.

Cuando $n \rightarrow \infty \Rightarrow \mu_X \rightarrow \mu$ es decir cuando el tamaño de la muestra se hace bastante grande, la media muestral de medias muestrales se aproxima a la media poblacional. Y la distribución de las medias muestrales tiene un comportamiento normal.



El siguiente trabajo de simulación consiste de una población de 100 alumnos cuyas notas se muestran en la columna C2. Su media (poblacional) y su desviación estándar (poblacional) se muestran a continuación, así como un histograma de frecuencias en modo texto.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
X	100	11.880	12.000	11.867	4.430	0.443

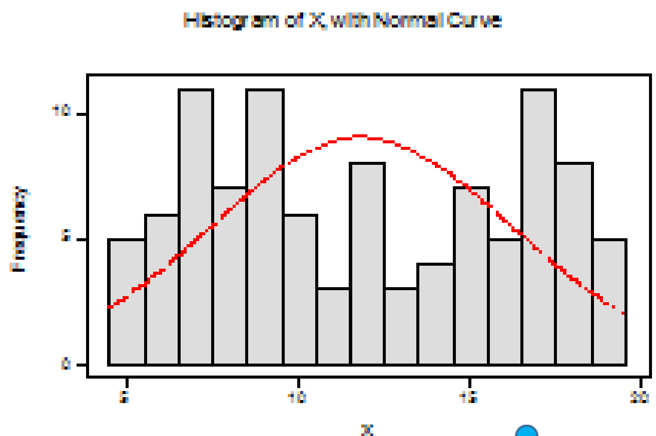
Variable	Minimum	Maximum	Q1	Q3
X	5.000	19.000	8.000	16.000

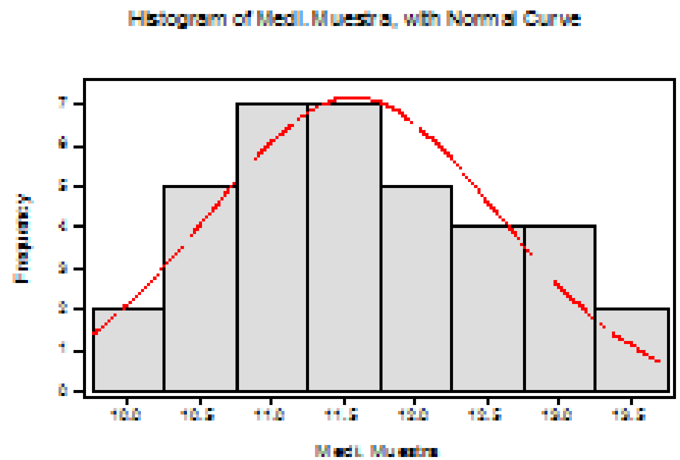
Histograma N=100

Midpoint	Count	
6	11	***********
8	18	******************
10	17	*****************
12	11	***********
14	7	*******
16	12	************
18	19	*******************
20	5	*****

A continuación hemos generado 36 muestras de tamaño 16 (columnas de 3 a 18) y en la columna C19 hemos obtenido las medias de cada muestra. La media muestral de las medias muestrales obtenidas es 11.5747; la desviación de la media muestral de medias es 1.0034. (Este resultado es aproximado a la poblacional y 1.003 = (4.4297 / 4)).

En la siguiente figura apreciamos lo siguiente: la primera gráfica es la distribución poblacional de las notas. Como es lógico, no tenía por qué ser normal. Sin embargo la gráfica de las medias muestrales sí tiene, aproximadamente, un comportamiento normal, con media 11.5747 y una varianza 1.003^2.

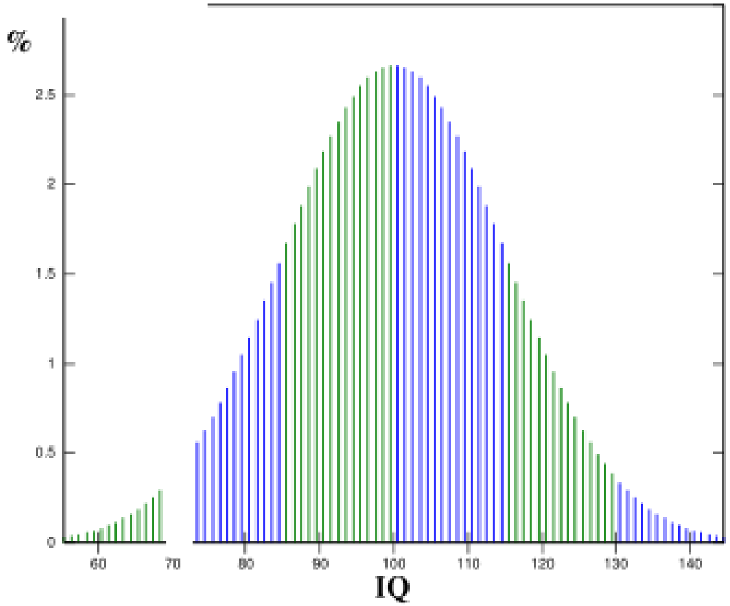




<u>De todo lo dicho, podemos emitir dos conclusiones muy importantes</u>:

La primera: Podemos usar la distribución normal para resolver problemas poblacionales cuya distribución es desconocida o no interesa conocerla. Teorizada esta afirmación, dio origen al Teorema del Límite Central (TLC) en su forma generalizada.

La segunda: Podemos usar los estadísticos de la muestra para realizar estimaciones sobre el comportamiento poblacional. Teorizado esta afirmación, dio origen a la Ley de Grandes Números (LGN).



Bibliografia

1) Lind, D.; Mason, R.; Marchal, W. (2001): "Estadística para Administración y Economía". Ed.Irwin McGraw-Hill.F.

2) Kvanli, A. (2000) "Introduction to Business Statistics" South-Western.
Johnson, R. (1996): "Elementary Statistics". Ed. Duxbury.

1) Levin, R.; Rubin, D. (1996): "Estadística para Administradores". Ed. Prentice Hall.

2) Farber, E. (1995): "A Guide to Minitab". Ed. McGraw-Hill