

**Taller #3
Muestreo
Econometría 06169**

**Profesores: Julio César Alonso
Diego Yépez**

Notas:

- Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 24 de agosto.

INSTRUCCIONES:

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

En el archivo anexo (T3-02-04.xls en la Hoja1) se reporta el ingreso mensual para los 5000 jefes de hogar de un pequeño municipio. Esta información se presenta por las 6 estratos socioeconómicos presentes en este municipio. (El ingreso está medido en miles de pesos). A partir de esta información responda:

1. Nuestra variable de interés es el ingreso medio de los jefes de hogar del municipio y se desea poder contar con un intervalo de confianza para ésta que con un 99% de confianza no sea más amplio que \$150.000. De acuerdo a esta información, y empleando una MEA responda:
 - a. Calcule la media poblacional por estrato y para todo el municipio.
 - b. ¿Cuál sería el tamaño total de la muestra y la distribución de la muestra por estrato? (suponga que no conoce la población). Explique claramente como llega a su resultado.
 - c. Seleccione una muestra aleatoria y calcule la media, y un intervalo de confianza del 99%.
 - d. Comente los resultados obtenidos y compárelos con los valores poblacionales.
2. Continuando con la pregunta anterior, suponga que el gobierno municipal está planeando otorgar un subsidio de educación para aquellos jefes de hogar que pertenecen al estrato 1 y la mitad de menores ingresos de los individuos que pertenecen al estrato 2. Pero existe un problema: no es fácil determinar el estrato del que proviene cada individuo, aunque determinar su ingreso si es mucho más sencillo. Determine un esquema por medio del cual el gobierno municipal pueda garantizar que el mayor número de individuos pueda acceder al subsidio. Explique claramente los supuestos y los cálculos que realice.

Una empresa de medición de opinión pública lo ha contratado a usted para asesorarlos en una nueva encuesta que debe desarrollar para una empresa del sector petrolero. La empresa desea conocer cuál es el grado de aceptación de su política salarial, pero como ésta posee 10000 empleados, sería muy costoso consultar la opinión de todos los empleados. (En el archivo anexo (T3-02-04.xls en la Hoja2) se encuentra la información

de todos los empleados de la compañía, la variable Aprob toma el valor de uno si el empleado aprueba la política y cero en caso contrario.) De acuerdo a esta información, responda las siguientes preguntas:

3. Calcule:

- a. La proporción poblacional de aprobación de la política salarial de la firma.
 - b. Determine el tamaño óptimo de la muestra si se desea tener un error de más o menos 1% y un grado de confiabilidad del 99% para la proporción de aprobación estimada por medio de una MAS simple sin reposición. (suponga que no conoce la población). Explique claramente como llega a su resultado.
 - c. Genere una MAS sin reposición y estime la proporción de aprobación.
 - d. Calcule un intervalo de confianza para la proporción de aprobación con un 99% de confiabilidad.
 - e. Comente sus resultados comparándolo con los valores poblacionales.
4. (Empleando la misma información de la pregunta anterior) La empresa petrolera posee 100 diferentes plantas alrededor de la nación. La empresa desea conocer el ingreso medio por trabajador para su firma. Lastimosamente todas las plantas de la firma se encuentran todas en lugares muy apartados y de difícil acceso, por eso un diseño MAS sin o con reposición sería muy costoso.
- a. Calcule la media poblacional
 - b. Escoja el mejor plan de muestreo para cumplir su objetivo.
 - c. Determine el tamaño de la muestra óptimo para asegurar un intervalo de confianza de una longitud no mayor a 200 mil pesos con un nivel de confianza del 99%. para el ingreso medio por trabajador (suponga que no conoce la población). Explique claramente como llega a su resultado
5. (Empleando la misma información de la pregunta anterior) conteste
- a. ¿Qué tan bueno es el tipo de muestreo que usted ha empleado? Discuta, recuerde que en la práctica usted no podrá conocer la población para realizar algún tipo de comparación.
 - b. Compare sus estimaciones con el valor real. Discuta sus resultados.

**Taller #3
Muestreo
Respuestas Sugeridas
Econometría 06169**

Profesores: Julio César Alonso

Notas:

- o Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- o Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 24 de agosto.

INSTRUCCIONES:

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

En el archivo anexo (T3-02-04.xls en la Hoja1) se reporta el ingreso mensual para los 5000 jefes de hogar de un pequeño municipio. Esta información se presenta por las 6 estratos socioeconómicos presentes en este municipio. (El ingreso está medido en miles de pesos). A partir de esta información responda:

1. Nuestra variable de interés es el ingreso medio de los jefes de hogar del municipio y se desea poder contar con un intervalo de confianza para ésta que con un 99% de confianza no sea más amplio que \$150.000. De acuerdo a esta información, y empleando una MEA responda:
 - a. Calcule la media poblacional por estrato y para todo el municipio.

Los resultados se reportan en la Tabla 1.

Tabla 1. Medias Poblacionales por Estrato y Municipal

	Número de Individuos	Ponderación por Estrato	Media	Varianza Poblacional
Estrato 1	1,474	0.295	\$ 193.71	12,707.10
Estrato 2	1,269	0.254	\$ 574.64	16,669.90
Estrato 3	1,004	0.201	\$ 1,028.12	25,986.84
Estrato 4	759	0.152	\$ 1,349.63	7,267.36
Estrato 5	252	0.050	\$ 1,634.20	21,224.32
Estrato 6	242	0.048	\$ 3,313.13	861,255.15
Total Municipal	5,000	1	856.9872	36,138,057,726

- b. ¿Cuál sería el tamaño total de la muestra y la distribución de la muestra por estrato? (suponga que no conoce la población). Explique claramente como llega a su resultado.

Noten que dado que no se cuenta con información de la población, será necesario realizar una prueba piloto. En la Tabla 2 se presentan el resumen de la prueba piloto

Tabla 2. Resumen de la Prueba Piloto

	Número de Individuos en la Pob.	Ponderación por Estrato (a partir de la Población)	Media	Varianza	Tamaño muestra Piloto
Estrato 1	1,474	0.295	\$ 207.19	11,613.44	10
Estrato 2	1,269	0.254	\$ 536.56	17,033.80	10
Estrato 3	1,004	0.201	\$ 1,039.99	29,057.27	10
Estrato 4	759	0.152	\$ 1,383.20	7,858.23	10
Estrato 5	252	0.050	\$ 1,573.53	19,368.70	10
Estrato 6	228	0.046	\$ 3,421.19	1,085,911.65	10
Total Municipal	4,986	1	851.37061	54,043,170,657	60

A partir de los resultados de la muestra piloto, podemos calcular el tamaño de la muestra piloto por medio de la siguiente fórmula:

$$n = \frac{z_{\alpha/2}^2 \left(\sum_{h=1}^H W_h S_H \right)^2}{\delta^2} \left(1 + \frac{1}{N} \left(\frac{z_{\alpha/2}^2 \left(\sum_{h=1}^H W_h S_H \right)^2}{\delta^2} \right) \right) \tag{1.1}$$

En este caso tenemos que $n=33.38$, es decir aproximadamente 34. Ahora por medio de la siguiente expresión se puede determinar el tamaño de la muestra por estrato:

$$n_h = n \frac{W_h S_H}{\sum_{h=1}^H W_h S_H} \tag{1.2}$$

Los resultados se reportan en la Tabla 3.

Tabla 3 Muestra por estrato.

De acuerdo a la Fórmula	Tamaño de la Muestra (redondeado)
Estrato 1	7.000
Estrato 2	7.000
Estrato 3	7.000
Estrato 4	3.000
Estrato 5	2.000
Estrato 6	10.000
Total	36.00

c. Seleccione una muestra aleatoria y calcule la media, y un intervalo de confianza del 99%.

A partir de la muestra aleatoria se obtienen los resultados reportados en la Tabla 4. y los intervalos se calculan a partir de la siguiente fórmula:

$$\bar{y}_h \pm t_{\frac{\alpha}{2}, n-1} s_h \tag{1.3}$$

donde $S_{y_h} = \sqrt{\left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}}$.

Tabla 4. Estimaciones a partir de la Muestra Estratificada.

Ponderación por Estrato	Media	Varianza	Error Estándar	t(0.005, n-1)	Limite		Amplitud del IC	Media	
					Inferior	Superior		Poblacional	
Estrato 1	0.295	\$ 207.50	7.665.17	33.01	3.707	\$ 85.11	\$ 329.90	\$ 244.78	\$ 193.71
Estrato 2	0.254	\$ 591.83	24.693.57	59.23	3.707	\$ 372.24	\$ 811.42	\$ 439.18	\$ 574.64
Estrato 3	0.201	\$ 1,067.86	6,117.56	29.46	3.707	\$ 958.65	\$ 1,177.08	\$ 218.44	\$ 1,028.12
Estrato 4	0.152	\$ 1,398.71	12,523.77	64.48	9.925	\$ 758.71	\$ 2,038.70	\$ 1,279.99	\$ 1,349.63
Estrato 5	0.050	\$ 1,577.88	246.81	11.06	63.656	\$ 873.55	\$ 2,282.20	\$ 1,408.65	\$ 1,634.20
Estrato 6	0.048	\$ 3,349.83	1,043,680.64	316.32	3.250	\$ 2,321.86	\$ 4,377.81	\$ 2,055.95	\$ 794.04
Total	1.00	\$ 735.06	13,721,350,338	26.20	2.72	\$ 663.71	\$ 806.41	\$ 142.70	\$ 735.06

d. Comente los resultados obtenidos y compárelos con los valores poblacionales.

Noten que en todos los casos los intervalos de confianza calculados incluyen la media poblacional. También es importante resaltar que la amplitud del IC para la media poblacional es de \$142.70 mil que es muy cercano a lo esperado según el diseño (\$150 mil).

2. Continuando con la pregunta anterior, suponga que el gobierno municipal está planeando otorgar un subsidio de educación para aquellos jefes de hogar que pertenecen al estrato 1 y la mitad de menores ingresos de los individuos que pertenecen al estrato 2. Pero existe un problema: no es fácil determinar el estrato del que proviene cada individuo, aunque determinar su ingreso si es mucho más sencillo. Determine un esquema por medio del cual el gobierno municipal pueda garantizar que el mayor número de individuos pueda acceder al subsidio. Explique claramente los supuestos y los cálculos que realice.

Noten que se puede diseñar un esquema muy fácil. Si suponemos que la distribución de los ingresos en el estrato dos es simétrica (la mediana muestral es \$528.31 mil muy cercana de la media muestral lo cual permite suponer que la distribución es simétrica) y que los ingresos de los individuos del estrato uno son menores a la media del ingreso del estrato 2, entonces se puede diseñar un esquema tal que las personas con ingresos menores o iguales a \$811.42 mil deberían recibir subsidio. Noten que esto garantiza que aproximadamente la mitad de los individuos del estrato 2 reciban el subsidio.

Una empresa de medición de opinión pública lo ha contratado a usted para asesorarlos en una nueva encuesta que debe desarrollar para una empresa del sector petrolero. La empresa desea conocer cuál es el grado de aceptación de su política salarial, pero como ésta posee 10000 empleados, sería muy costoso consultar la opinión de todos los empleados. (En el archivo anexo (T3-02-04.xls en la Hoja2) se encuentra la información de todos los empleados de la compañía, la variable Aprob toma el valor de uno si el empleado aprueba la política y cero en caso contrario.) De acuerdo a esta información, responda las siguientes preguntas:

3. Calcule:
 - a. La proporción poblacional de aprobación de la política salarial de la firma.

La proporción de aprobación de la política salarial de la firma es 40.7% y la de desaprobación es de 59.3%. Noten que estas proporciones son muy similares.

b. Determine el tamaño óptimo de la muestra si se desea tener un error de más o menos 1% y un grado de confiabilidad del 99% para la proporción de aprobación estimada por medio de una MAS simple sin reposición. (suponga que no conoce la población). Explique claramente como llega a su resultado. El tamaño de la muestra se puede calcular a partir de una muestra piloto de 11 dato y empleando la siguiente fórmula:

$$n = \frac{z_{\alpha/2}^2 \hat{P}\hat{Q}}{\delta^2} \left(\frac{N-1}{N} + \frac{1}{N} \left(\frac{z_{\alpha/2}^2 \hat{P}\hat{Q}}{\delta^2} \right) \right) \tag{1.4}$$

para nuestra prueba piloto se encontró que $\hat{P} = 0.55$ y $\hat{Q} = 0.45$, en este caso tenemos que:

$$n = \frac{(2.58)^2 (0.55)(0.45)}{(0.005)^2} = 8,678.86 \approx 8,679$$

$$n = \frac{10000-1}{10000} + \frac{1}{10000} \left(\frac{(2.58)^2 (0.55)(0.45)}{(0.005)^2} \right)$$

Noten lo grande de esta muestra. La razón es que se desea una precisión muy alta y además las proporciones de aceptación y rechazo de la medida son muy cercanas, y por tanto para obtener esa gran exactitud se requiere una muestra muy grande que corresponde al 87% de la población.

c. Genere una MAS sin reposición y estime la proporción de aprobación.

En este caso tenemos que $\hat{P} = 0.40$ y $\hat{Q} = 0.60$.

d. Calcule un intervalo de confianza para la proporción de aprobación con un 99% de confiabilidad.

El intervalo de confianza para la proporción de aprobación con un 99% de confiabilidad está dado por:

$$\hat{P} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{N-n}{N-1} \frac{\hat{P}\hat{Q}}{n}} \tag{1.5}$$

En este caso el intervalo de confianza es:

$$0.40 \pm 2.58(0.002)$$

$$(0.399, 0.409)$$

Noten que la amplitud es de un 1%!!!

e. Comente sus resultados comparándolo con los valores poblacionales.

Es importante puntualizar que la proporción poblacional se encuentra contenido en el intervalo de confianza calculado y este intervalo tiene curiosamente exactamente la misma amplitud diseñada.

4. (Empleando la misma información de la pregunta anterior) La empresa petrolera posee 100 diferentes plantas alrededor de la nación. La empresa desea conocer el ingreso medio por trabajador para su firma. Lastimosamente todas las plantas de la firma se encuentran todas en lugares muy apartados y de difícil acceso, por eso un diseño MAS sin o con reposición sería muy costoso.

a. Calcule la media poblacional

La media poblacional es \$4401.9

b. Escoja el mejor plan de muestreo para cumplir su objetivo.

Noten que en este caso es mejor emplear un muestreo por conglomerado, escogiendo como conglomerado la planta.

- c. Determine el tamaño de la muestra óptimo para asegurar un intervalo de confianza de una longitud no mayor a 200 mil pesos con un nivel de confianza del 99%. para el ingreso medio por trabajador (suponga que no conoce la población). Explique claramente como llega a su resultado

El número de conglomerados a escoger para garantizar las condiciones y empleando una muestra piloto de 5 plantas corresponde a:

$$n_c = \frac{\frac{z_{\alpha/2}^2 S_{congl}^2}{\delta^2 M^2}}{1 + \frac{1}{N_c} \left(\frac{z_{\alpha/2}^2 S_{congl}^2}{\delta^2 M^2} \right)} \tag{1.6}$$

En este caso, necesitamos

$$n_c = \frac{(2.58)^2 \cdot 583,953,290}{100^2 \cdot 100^2} = 38.60 \approx 39$$

$$1 + \frac{1}{100} \left(\frac{(2.58)^2 \cdot 583,953,290}{100^2 \cdot 100^2} \right)$$

Es decir 39 conglomerados, cada uno con 100 empleados.

5. (Empleando la misma información de la pregunta anterior) conteste

- a. ¿Qué tan bueno es el tipo de muestreo que usted ha empleado? Discuta, recuerde que en la práctica usted no podrá conocer la población para realizar algún tipo de comparación.

Note que el método de conglomerados será bueno siempre y cuando la correlación intra-conglomerados sea cercana a cero. En este caso para nuestra muestra de 39 conglomerados tenemos que dicha correlación es de 0.01 relativamente cercana a cero. Lo cual demuestra que el método es adecuado. Además note que este método implicará menos costos que realizar un MAS.

b. Compare sus estimaciones con el valor real. Discuta sus resultados.

A partir de la muestra obtuvimos la siguiente media muestral por individuo: $\bar{y} = 4377.4$, de hecho el correspondiente intervalo de confianza con un nivel de confianza del 99% corresponde a: (4292.969, 4461.907). Noten que la media poblacional (\$4401.9) está contenida en el intervalo.