Taller #2
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 17 de Febrero.

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 se reporta el ingreso mensual para los 1500 jefes de hogar de una comunidad. (El ingreso está medido en miles de pesos). A partir de esta información responda:

1. De acuerdo a esta información responda:
   a. ¿Cuántas MAS sin reposición de tamaño 20 se pueden generar? ¿Cuál es la probabilidad de que un individuo sea seleccionado y que una muestra sea seleccionada?
   b. Nuestra variable de interés es el ingreso medio de los jefes de hogar de la comunidad y se desea poder contar con un intervalo de confianza para ésta que con un 99% de confianza no sea más amplio que $300.000. Determine el tamaño de la MAS sin reposición necesaria para garantizar esto.
   c. ¿Cuál sería el tamaño de la muestra si el nivel de confianza deseado es del 90%?
2. Genere aleatoriamente 100 MAS sin reposición (para un nivel de confianza del 99% y la longitud del IC expresada anteriormente) y calcule:
   a. La media muestral para cada una de las 100 muestras generadas y grafique su histograma. (Esta corresponde a la distribución de la media muestral)
   b. Calcule el promedio y la desviación estándar de las 100 medias muestrales calculadas en la parte a. de esta pregunta. Compare los resultados con lo que sugiere la teoría estadística.
   c. Grafique el histograma de los ingresos de los jefes de hogar de la población. Compárelo con el histograma calculado en la parte a. Explique las razones de cualquier similitud o diferencia que encuentre.
3. Ahora suponga que no conoce la varianza poblacional. Y por tanto usted debe realizar una prueba piloto para determinar dicha varianza. Realice una prueba piloto con 20 individuos de la población y :
   a. Determine el tamaño de la MAS sin reposición necesaria para garantizar las mismas condiciones de la parte 1b.

   b. Genere una MAS sin reposición con ese tamaño y genere un IC del 99% para la media de ésta.

4. Empleando la misma información, responda:
   a. ¿Cuántas MAS *con* reposición de tamaño 20 se pueden generar? ¿Cuál es la probabilidad de que un individuo sea seleccionado y que una muestra sea seleccionada?
   b. Determine el tamaño de la MAS *con* reposición necesaria para garantizar las mismas condiciones de la parte 1b y suponiendo que no conoce la varianza poblacional (puede emplear los resultados de su prueba piloto de la pregunta 3).
   c. ¿Cuál sería el tamaño de la muestra si el nivel de confianza deseado es del 90%?
5. Genere aleatoriamente 10 MAS *con* reposición (para un nivel de confianza del 99% y la longitud del IC expresada anteriormente) y calcule:
   a. La media muestral para cada una de las 10 muestras generadas y grafique su histograma. (Esta corresponde a la distribución de la media muestral)
   b. Calcule el promedio y la desviación estándar de las 10 medias muestrales calculadas en la parte a. de esta pregunta. Compare los resultados con lo que sugiere la teoría estadística.
   c. Grafique el histograma de los ingresos de los jefes de hogar de la población. Compárelo con el histograma calculado en la pare a. Explique las razones de cualquier similitud o diferencia que encuentre.
   d. Finalmente, a partir de cualquiera de las MAS *con* reposición generadas, encuentre un intervalo de confianza del 99% para la media.

Taller #2
Respuestas Sugeridas
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 17 de Febrero.

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 se reporta el ingreso mensual para los 1500 jefes de hogar de una comunidad. (El ingreso está medido en miles de pesos). A partir de esta información responda:

1. De acuerdo a esta información responda:
   a. ¿Cuántas MAS sin reposición de tamaño 20 se pueden generar? ¿Cuál es la probabilidad de que un individuo sea seleccionado y que una muestra sea seleccionada?

En este caso tenemos que N=1500 y n=20, por tanto podemos tener el siguiente número de MAS sin reposición:

$$\frac{N!}{n!(N-n)!}=\frac{1500!}{20!(1500-20)!}=\frac{1500\cdot1499\cdot\ldots\cdot1481\cdot1480!}{20!(1480)!}$$
$$=\frac{1500\cdot1499\cdot\ldots\cdot1481}{20!}=1{,}20351\text{E+}45$$

La probabilidad de un individuo ser seleccionado es:

$$\frac{n}{N}=\frac{20}{1500}=\frac{2}{150}=\frac{1}{75}=0.13\overline{3}$$

Y la probabilidad de que una muestra sea seleccionada es:

$$\frac{1}{\frac{N!}{n!(N-n)!}}=\frac{1}{1{,}20351\text{E+}45}=8{,}30902\text{E-}46$$

   b. Nuestra variable de interés es el ingreso medio de los jefes de hogar de la comunidad y se desea poder contar con un intervalo de confianza para ésta que con un 99% de confianza no sea más amplio que $300.000. Determine el tamaño de la MAS sin reposición necesaria para garantizar esto.

Noten que en este caso tenemos que:

$$\delta=150$$
$$\alpha=0.01$$
$$z_{\frac{\alpha}{2}}=2.576$$
$$S^2=85764.04944$$

Así, tamaño de la MAS sin reposición necesaria para garantizar las condiciones requeridas es:

$$n=\frac{\frac{z_{\alpha/2}^2S^2}{\delta^2}}{1+\frac{1}{N}\left(\frac{z_{\alpha/2}^2S^2}{\delta^2}\right)}=\frac{\frac{2.576^2\bullet85764.04944}{150^2}}{1+\frac{1}{1500}\left(\frac{2.576^2\bullet85764.04944}{150^2}\right)}=24.871\approx25$$

Es decir, el tamaño de la muestral debe ser de 25.

   c. ¿Cuál sería el tamaño de la muestra si el nivel de confianza deseado es del 90%?

En este caso

$$\alpha=0.1$$
$$z_{\frac{\alpha}{2}}=1.645$$

Por tanto, el tamaño de la MAS sin reposición necesaria para garantizar las condiciones requeridas es:

$$n=\frac{\frac{z_{\alpha/2}^2S^2}{\delta^2}}{1+\frac{1}{N}\left(\frac{z_{\alpha/2}^2S^2}{\delta^2}\right)}=\frac{\frac{1.645^2\bullet85764.04944}{150^2}}{1+\frac{1}{1500}\left(\frac{1.645^2\bullet85764.04944}{150^2}\right)}=10.2421\approx10$$

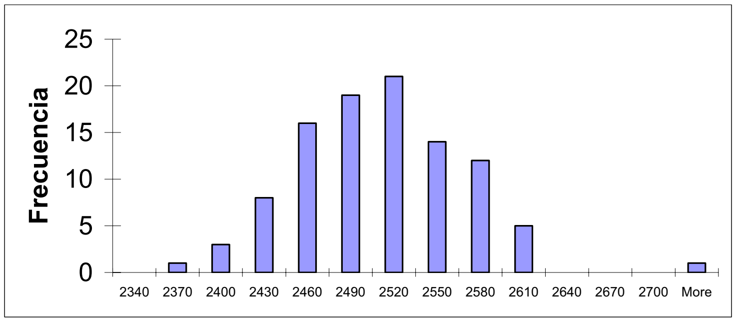
Es decir, el tamaño de la muestral debe ser de 10.

2. Genere aleatoriamente 100 MAS sin reposición (para un nivel de confianza del 99% y la longitud del IC expresada anteriormente) y calcule:

a. La media muestral para cada una de las 100 muestras generadas y grafique su histograma. (Esta corresponde a la distribución de la media muestral)

Noten que los siguientes resultados pueden diferir un poco pues los datos han sido generados aleatoriamente.
El histograma de las medias se puede observar en el Gráfico 1.

Gráfico 1. Histograma de las medias muestrales



Noten que esta distribución muestral del estimador

b. Calcule el promedio y la desviación estándar de las 100 medias muestrales calculadas en la parte a. de esta pregunta. Compare los resultados con lo que sugiere la teoría estadística.

En el caso de nuestras simulaciones, la medía de las medias es 2495.80716 y la desviación estándar de ellas (error estándar) corresponde a 56.84604779 (mientras que la respectiva varianza es 3231.47315). Noten que el valor poblacional de la media es $\mu = 2496.446282$. Como era de esperarse, en promedio el estimador de la medía está muy cercano al valor real. Por otro lado, la varianza poblacional es $\sigma^2 = 85821.22548$ (es decir la desviación estándar es $\sigma = 292.9525994$).

Noten que según el teorema del limite central implica que si la muestra es lo suficientemente grande, entonces

$$\bar{y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

En otras palabras la distribución de las medias maestrales se aproxima a la normal, tal como se ve en el Gráfico 1y Gráfico 2. Además la media de las 100 medias muestrales está relativamente cerca de la población y la varianza (error estándar) de las medias maestrales

es aproximadamente igual a $\sigma^2/n = 3432.849019$ ($\sigma/\sqrt{n} = 58.59051987$). Es decir lo que predice la teoría se cumple en este caso.

Gráfico 2. Histograma de las medias Muestrales y la distribución Normal.

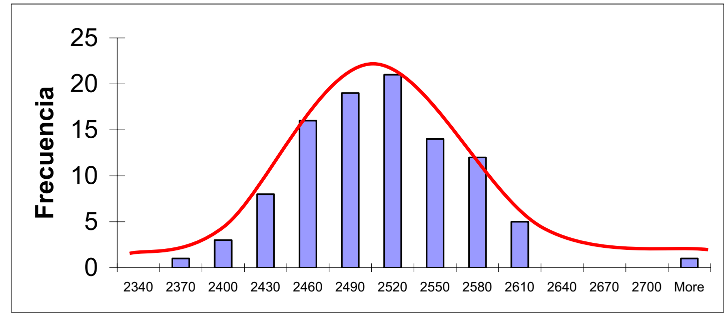


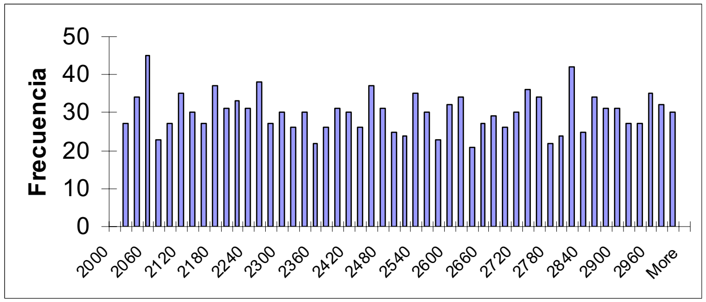
Tabla 1. Resumen de los resultados

Predicción de la Teoría	***Estimado***
$\mu = 2496.446282$	2495.80716
$\sigma/\sqrt{n} = 58.59051987$	56.84604
Distribución Normal	Distrib parecida a la normal

c. Grafique el histograma de los ingresos de los jefes de hogar de la población. Compárelo con el histograma calculado en la parte a. Explique las razones de cualquier similitud o diferencia que encuentre.

En el Gráfico 3 se reporta el histograma de los ingresos de los jefes de hogar de la población. Noten que esta distribución corresponde a una distribución uniforme muy diferente a la distribución que presentan las medias maestrales. Esta diferencia ilustra el Teorema del Limite Central. Recuerden que este Teorema implica que sin importar la distribución de la cual provengan los individuos de una población, la media muestral se distribuirá normalmente.

Gráfico 3. Histograma de los Ingresos de los Individuos de la Población (Distribución Poblacional)



3. Ahora suponga que no conoce la varianza poblacional. Y por tanto usted debe realizar una prueba piloto para determinar dicha varianza. Realice una prueba piloto con 20 individuos de la población y :
   a. Determine el tamaño de la MAS sin reposición necesaria para garantizar las mismas condiciones de la parte 1b.

Noten que los siguientes resultados pueden diferir un poco pues los datos han sido generados aleatoriamente.

Para generar la muestra piloto se puede emplear una MAS sin reposición de 20 individuos. En este caso, tenemos que $S^2 = 70017.78994$. Entonces el tamaño de la muestra será:

$$n = \frac{\frac{z_{\alpha/2}^2 S^2}{\delta^2}}{1+\frac{1}{N}\left(\frac{z_{\alpha/2}^2 S^2}{\delta^2}\right)} = \frac{\frac{2.576^2 \bullet 70017.78994}{150^2}}{1+\frac{1}{1500}\left(\frac{2.576^2 \bullet 70017.78994}{150^2}\right)} = 20.367 \approx 20$$

   b. Genere una MAS sin reposición con ese tamaño y genere un IC del 99% para la media de ésta.

Para nuestra muestra la media estimada es $\bar{x} = 2463.246559$ y además $S_{\bar{x}} = \sqrt{\left(1-\frac{n}{N}\right)}\frac{S}{\sqrt{n}} = 71.6255445$. Además note que $t_{0.005,24} = 2.796950866$. Por tanto un IC del 99% para la media está dado por:

$$\bar{y} \pm t_{\frac{\alpha}{2}} S_{\bar{y}} = 2463.246559 \pm 200.3331289$$

$$(2262.913, 2663.579688)$$

Noten que en este caso la mitad de la longitud del intervalo de confianza es de aproximadamente \$200.33 relativamente cerca al que queríamos (\$150 mil). Además recuerden que $\mu = 2496.446282$, el cual en efecto está contenido en este intervalo.

4. Empleando la misma información, responda:
   a. ¿Cuántas MAS *con* reposición de tamaño 20 se pueden generar? ¿Cuál es la probabilidad de que un individuo sea seleccionado y que una muestra sea seleccionada?

El número de diferentes MAS con reposición de tamaño 20 está dado por:

$$N^n = 1500^{20} = 3.325E{+}63$$

La probabilidad de que un individuo sea seleccionado es

$$1-\left(\frac{N-1}{N}\right)^n = 1-\left(\frac{1500-1}{1500}\right)^{20} \approx .013$$

La probabilidad de que una muestra sea seleccionada es

$$\frac{1}{N^n} = \frac{1}{3.325E{+}63} \approx .000000000001$$

   b. Determine el tamaño de la MAS *con* reposición necesaria para garantizar las mismas condiciones de la parte 1b y suponiendo que no conoce la varianza poblacional (puede emplear los resultados de su prueba piloto de la pregunta 3).

En este caso tenemos que

$$\delta = 150$$
$$\alpha = 0.01$$
$$z_{\frac{\alpha}{2}} = 2.576$$
$$S^2 = 70017.78994$$

Así, el tamaño de la MAS con reposición está dada por:

$$n = \frac{z_{\alpha/2}^2 S^2}{\delta^2} = \frac{2.576^2 \bullet 70017.78994}{150^2} = 25.29 \approx 25$$

Es decir la muestral debería ser de tamaño 25.

   c. ¿Cuál sería el tamaño de la muestra si el nivel de confianza deseado es del 90%?

En este caso el tamaño de la muestra debería ser:

$$n=\frac{z_{\alpha/2}^{2}S^{2}}{\delta^{2}}=\frac{1.645^{2}\bullet 70017.78994}{150^{2}}=25.29\approx 25$$

Noten que en este caso el tamaño de la muestral es más o menos el mismo que en la parte anterior.

5. Genere aleatoriamente 10 MAS *con* reposición (para un nivel de confianza del 99% y la longitud del IC expresada anteriormente) y calcule:
   a. La media muestral para cada una de las 10 muestras generadas y grafique su histograma. (Esta corresponde a la distribución de la media muestral)

En este caso tenemos que la distribución de la media muestral está dado por el siguiente histograma (ver Gráfico 4) Noten que aunque hay pocas realizaciones, estas parecen seguir una distribución normal (ver Gráfico 5)

Gráfico 4. Histograma de las medias muestrales

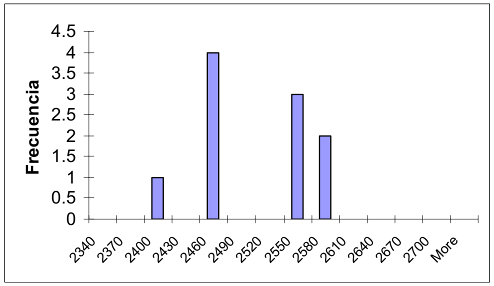
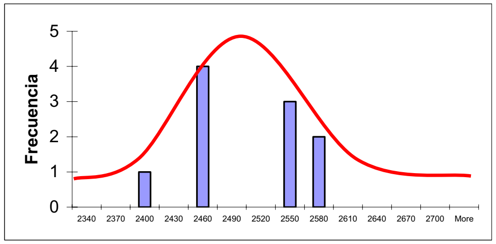


Gráfico 5. Histograma de las medias Muestrales y la distribución Normal.



   b. Calcule el promedio y la desviación estándar de las 10 medias muestrales calculadas en la parte a. de esta pregunta. Compare los resultados con lo que sugiere la teoría estadística.

En este caso tenemos que la medía de las medias es 2488.763573 y la desviación estándar de ellas (error estándar) corresponde a 61.51316722 (mientras que la respectiva varianza es 3783.869741). Recuerden que el valor poblacional de la media es $\mu=2496.446282$. Como era de esperarse, en promedio el estimador de la medía está muy cercano al valor real. Además, la varianza poblacional es $\sigma^{2}=85821.22548$ (es decir la desviación estándar es $\sigma=292.9525994$)

Noten que según el teorema del limite central implica que si la muestra es lo suficientemente grande, entonces

$$\bar{y}\sim N\left(\mu,\frac{\sigma^{2}}{n}\right)$$

En otras palabras la distribución de las medias maestrales se aproxima a la normal, tal como se ve en el Gráfico 5. Además la media de las 10 medias maestrales está relativamente cerca de la población y la varianza (error estándar) de las medias muestrales es aproximadamente igual a $\sigma^{2}/n=3432.849019$ ($\sigma/\sqrt{n}=58.59051987$). Es decir lo que predice la teoría se cumple en este caso.

   c. Grafique el histograma de los ingresos de los jefes de hogar de la población. Compárelo con el histograma calculado en la pare a. Explique las razones de cualquier similitud o diferencia que encuentre.

En el Gráfico 3 se reportó el histograma de los ingresos de los jefes de hogar de la población. Como se anotó anteriormente, esta distribución corresponde a una distribución uniforme muy diferente a la distribución que presentan las medias maestrales. Esta diferencia ilustra el Teorema del Limite Central. Recuerden que este Teorema implica que sin importar la distribución de la cual provengan los individuos de una población, la media muestral se distribuirá normalmente.

   d. Finalmente, a partir de cualquiera de las MAS *con* reposición generadas, encuentre un intervalo de confianza del 99% para la media.

Noten que los siguientes resultados pueden diferir un poco pues los datos han sido generados aleatoriamente.

Para nuestra muestra la media estimada es $\bar{x} = 2476.806543$ y además $S_{\bar{x}} = 63.17826707$. Además note que $t_{0.005,24} = 2.796950866$. Por tanto un IC del 99% para la media está dado por:

$$\bar{y} \pm t_{\frac{\alpha}{2}} S_{\bar{y}} = 2476.806543 \pm 176.7065088$$

$$(2300.100, 2653.513)$$

Noten que en este caso la mitad de la longitud del intervalo de confianza es de aproximadamente $176.7 relativamente cerca al que queríamos ($150 mil). Además recuerden que $\mu = 2496.446282$, el cual en efecto está contenido en este intervalo.