Inferencia Estadística

Profesor(es): Jarnishs Beltran
Ayudante: Pablo Rivera

Ayudantía N°2
Otoño 2020

Ejercicios

i) Se desea estimar la **calificación promedio** de los estudiantes de la PSU, para ello se define los siguientes criterios: Nivel de confianza 95 %, desviación estándar de 16, 44 y error máximo tolerable de 5. Calcule.

a) En base a lo anterior, ahora suponga $N = 5000$

b) Ahora se desea estimar la **proporción** de los estudiantes de la UDP, con un nivel de aprendizaje de excelencia, para ello se define los siguientes criterios: nivel de confianza 95 %, proporción de estudiantes de $(0,10$ a $0,15)$ como referencia, error máximo tolerable del 5 %

c) En base a los datos de b), ahora suponga $N = 5000$

Solution:

$$n = (\frac{1{,}96 \cdot 16{,}44}{5})^2 = 41{,}53 \simeq 42$$

a) Con N=5000 (ósea, finito)

$$n = \frac{1{,}96^2 \cdot 5000 \cdot 16{,}44^2}{16{,}44^2 1{,}96^2 + (5000 - 1) \cdot 5^2} = 41{,}19 \simeq 42$$

b)

$$n = \frac{1{,}96^2 \cdot 0{,}15 \cdot 0{,}85}{0{,}05^2} = 195{,}9 \simeq 196$$

c)

$$n = \frac{1{,}96^2 \cdot 5000 \cdot 0{,}15 \cdot 0{,}85}{1{,}96^2 \cdot 0{,}15 \cdot 0{,}85 + (5000 - 1) \cdot 5^2} = 189$$

Recordar que debe aproximarse arriba.

ii) Se desea estimar el **ingreso promedio** de las familias que viven en la comuna de Maipú, para ello se definen los siguientes criterios: Nivel de confianza 95 %, desviación estándar de 908, 07 y error máximo tolerable de 200. Calcule.

a) En base a los datos anteior, ahora suponga $N = 5000$.

b) Ahora se desea estimar la **proporción** de las familias que viven en la comuna de Maipú y que itenen un ingreso alto, para ello se definen los siguientes criterios: nivel de confianza 95 %, proporción de familias con ingreso alto 0, 10 y error máximo tolerable del 8 %.

c) En base a los datos en b), ahora suponga $N = 5000$

Solution:

$$n = (\frac{1{,}96 \cdot 908{,}07}{200})^2 = 79{,}19 \simeq 80$$

a)

$$n = \frac{1{,}96^2 \cdot 50000 \cdot 908{,}07^2}{908{,}07^2 1{,}96^2 + (50000-1) \cdot 200^2} = 79{,}01 \simeq 80$$

b)

$$n = (\frac{1{,}96 \cdot 0{,}10 \cdot 0{,}90}{0{,}08^2} = 54$$

c)

$$n = \frac{1{,}96^2 \cdot 50000 \cdot 0{,}10 \cdot 0{,}90}{1{,}96^2 \cdot 0{,}10 \cdot 0{,}90 + (50000-1) \cdot 0{,}08^2} = 54$$

iii) Para estudiar la imagen de los diferentes políticos, se pide a los encuestados que los evalúen en una escala (continua) de 0 a 10 puntos. Si se acepta que la desviación típica de esta variable es de 1,5 puntos, ¿cuántos casos se necesitan para que la semiamplitud del intervalo de confianza (e) al 95 % de la media poblacional sea de $0,05$ puntos? Responda.

a) Dado que 3458 casos son demasiados, se rebaja la ambición de conocimiento desde una semiamplitud de 0.05 puntos hasta 0.25 puntos. ¿Cuál es ahora el tamaño necesario? ¿Cómo podríamos realizar este cálculo del tamaño muestral en R? [Recuerde que lo primero es realizar la instalación y la carga del paquete. Hint: *samplingbook*, *sample.size.mean*].

Solution:

$$n = (\frac{Z_{\alpha/2} \cdot \sigma}{e})^2 = (1{,}96 \cdot 1{,}50{,}05)^2 \simeq 3457{,}44 \rightarrow 3458 casos$$

Nota técnica: En los cálculos de tamaño muestral el número obtenido siempre se redondea al alza para alcanzar el objetivo especificado

a) En R, existen diversos paquetes para el cálculo del tamaño muestral. El paquete samplingbook contiene funciones aplicables a la estimación de un parámetro; por ejemplo la función sample.size.mean realiza el cálculo para la estimación de una media. [Recuerde que lo primero es realizar la instalación y la carga del paquete].

Ejemplo de R

```
# Instalación y carga de 'samplingbook'
> install.packages('samplingbook')
> library(samplingbook)

# Aplicación al Ejemplo 1.1. (e es la semi-amplitud, S es
# la sigma (σ) y level es la confianza)
> sample.size.mean(e=0.05, S=1.5, level = 0.95)
Sample size needed: 3458
```

iv) Para conocer el porcentaje de votos de un partido político, con una semiamplitud total del intervalo de confianza al 95 % igual a 0,5 % (Amplitud del 1 %) ¿Cuántos casos se necesitan?

a) Decididamente 38416 son demasiados casos, por lo que una vez más se rebaja la ambición de conocimiento desde una semi-amplitud de 0,5 puntos (0,5 % = 0,005) hasta 2,5 puntos (2,5 % = 0,025). Calcule el nuevo tamaño muestra [Recuerde que lo primero es realizar la instalación y la carga del paquete. Hint: *samplingbook*, *sample.size.mean*].

Solution:

$$n = (\frac{Z_{\alpha/2}}{2e})^2 = (\frac{1{,}96}{2 \cdot 0{,}005})^2 \simeq 38416\, casos$$

a) En R, la instrucción sample.size.prop del paquete samplingbook realiza el cálculo del tamaño necesario para el intervalo de confianza de una probabilidad. [Instálese y cárguese el paquete si no se hizo previamente]

Ejemplo de R

```
#  Aplicación  al  Ejemplo  1.2  (e  es  la  semi-amplitud  del  IC  y
# level es la confianza)
> sample.size.prop(e=0.005, level = 0.95)
Sample size needed: 38415
```

Formulario

- Media Muestral Sea una muestra de tamaño n (x1, x2, ... , xn) de una población de tamaño N.
  • Media poblacional: μ
  • Varianza poblacional: σ
  • Media muestral: $\bar{x} = \frac{x_1+..+x_n}{n}$
  también es una variable aleatoria; se puede mostrar que:
  • Esperanza $E[\bar{x}] = \mu$
  • Varianza $Var[\bar{x}] = \frac{\sigma^2}{n}$
  • Desviación típica: $SD[\bar{x}] = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

  En caso de que $n > 0{,}05$
  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

- Teorema Central del límite

$$Y = \sum ni = 1X_i$$

  sigue aproximadamente una distribución normal

$$\bar{X} = \frac{1}{n}\sum ni = 1X_i = \frac{1}{n}Y$$

- Distribución de la media muestral

$$Y = \sum ni = 1X_i$$

  Por tamaño muestral es grande, por el teorema central del limite se tiene:

$$\bar{X} \approx N(\mu; \frac{\sigma}{\sqrt{n}})$$

- Distribución muestral de la proporción
  Proporción muestral $\bar{p}$ es el estimador puntual de la proporción poblacional p

$$\bar{o} = \frac{x}{n}$$

  El valor esperado $E(\bar{p}) = p$.
  • Desviación estándar $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$(Población finita)

  • Desviación estándar $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$ (Población infinita)

  • Selección igual a la media.

- Distribución Chi-Cuadrado

$$\chi^2_k + \chi^2_n = \chi^2_{k+n}$$

  • Por el teorema central del limite, cuando $k > 30$:

$$\sqrt{2\chi^2_k} \backsim N(\sqrt{2k-1}; 1)$$

- Distribución t de Student

$$X = \frac{Y}{\sqrt{\frac{1}{n}\sum ni = 1Z_i}}$$

Media: 0

Varianza: $\frac{n}{n-2}$

- Distribución F

$$F_{n;m} = \frac{\frac{1}{n}\sum n1Y_i^2}{\frac{1}{n}\sum m1Z_i^2}$$

Media: $\frac{m}{m-2}$ Varianza: $\frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$

- Tamaño de la muestra (T. central del limite)
  - Tamaño de una muestra para estimación de la "media"
  - Varianza conocida – población infinita (o desconocida)

$$Z^2_{\alpha/2}\frac{\sigma^2}{e^2}$$

  - Varianza conocida – población finita

$$Z^2_{\alpha/2}\frac{\sigma^2 N}{Z^2_{\alpha/2}\sigma^2 + e^2(N-1)}$$

Donde e es el error porcentual y $Z^2_{\alpha/2}$ es el nivel de confianza.

Tamaño de una muestra para estimación de la "proporción"

  - Varianza conocida – población infinita (o desconocida)

$$Z^2_{\alpha/2}\frac{p(1-p)}{e^2}$$

  - Varianza conocida – población finita

$$Z^2_{\alpha/2}\frac{p(1-p)N}{Z^2_{\alpha/2}p(1-p) + e^2(N-1)}$$

Donde e es el error porcentual y $Z^2_{\alpha/2}$ es el nivel de confianza.