- Métodos computacionales:
  Alejandro Segura

- **Estadística**

  a) Incluir el código Notebook (.ipynb).

  b) Guardar la información en una carpeta llamada Semana15_Nombre1_Nombre2

  c) **Hacer una sola entrega por grupo**.

Contents

List of Figures

1 Estadística

1.1 Intervalos de confianza

1.1.1 Intervalo para la media

Sean $x_1, x_2, ..., x_n$ variables aleatorias independiente con $\mathbb{E}(x_i) = \mu$ y $Var(x_i) < \infty$, entonces:

$$z_n := \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \tag{1}$$

la función de distribución de z_n, converge a una función de distribución normal estándar $z \sim N(0, 1)$.
El intervalo de confianza para la media con varianza conocida está dada por:

$$IC_{1-\alpha} = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{2}$$

Este intervalo es válido para población aproximada normal, u otra distribución poblacional con $n >= 30$.

Cuando no se conoce la varianza poblacional y la muestra es pequeña, se necesita hacer una estimación del intervalo usando la distribución T-Student. El intervalo de confianza para la media está dado por:

$$IC_{1-\alpha} = \left[\bar{x} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{1-\alpha/2,n-1}\frac{S}{\sqrt{n}}\right] \tag{3}$$

donde S es la varianza muestral.

1.1.2 Intervalo de confianza para la diferencia de medias

Para encontrar un intervalo de confianza para la diferencia de medias ($n \leq 30$), se tiene la siguiente expresión:

$$IC_{1-\alpha} = \left[(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{1-\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right] \tag{4}$$

Los grados de libertad se obtienen usando:

$$\nu = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(\frac{s_1^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{s_2^2}{n_2})^2}{n_2 - 1}} \tag{5}$$

Para encontrar un intervalo de confianza para la diferencia de medias ($n > 30$), se tiene la siguiente expresión:

$$IC_{1-\alpha} = \left[(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] \tag{6}$$

1.1.3 Intervalo de confianza para una proporción

Para estimar el intervalo de confianza de un proporción, se tiene:

$$IC_{1-\alpha} = \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < P < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \tag{7}$$

donde se considera la aproximación binomial a la normal si $n\hat{p} \geq 10$ y $n(1-\hat{p}) \geq 10$.

1.1.4 Ejercicios

1. El contenido de botellas de un litro tienen una distribución aproximadamente normal con varianza de 0.15 $litros^2$. Se toma una muestra aleatoria de cajas y se mide el contenido, obteniendo el siguiente resultado.

$$S = [0.974, 0.950, 0.932, 1.104, 1.038, 0.920, 0.935, 0.907, 0.810, 0.915] \tag{8}$$

   a) Construya el intervalo de confianza del 95% para la media población.

$$IC_{95\%} = \Bigg[0.708 < \mu < 1.188\Bigg] \tag{9}$$

   b) Construya el intervalo de confianza del 95% usando el método de Bootstrapping ($N = 50000$ toy experiments).

$$IC_{95\%} = \Bigg[0.903 < \mu < 1.00\Bigg] \tag{10}$$

   Notar que los IC no son consistentes, esto puede indicar que la población que describe a los datos no es normal. Usted puede calcular el IC en el caso de varianza desconocida y comprobar que se obtiene un mejor resultado.

2. Construir un IC de 90% de nivel de confianza para la estatura media de un grupo de estudiantes. Se sabe que la estatura (en cm) es aproximadamente normal y está dada por:

$$S = [175, 176, 180, 164, 170, 170, 181, 169, 163, 190, 170, 171] \tag{11}$$

$$IC_{90\%} = \Bigg[169.28 < \mu < 177.21\Bigg] \tag{12}$$

3. Se tienen dos muestras de la medida de la resistencia eléctrica de un lote de dispositivo electrónicos, las cuales están dadas por:

$$\begin{aligned} S_1 &= [6.0, 5.0, 6.5, 5.0, 4.0, 5.0, 5.0, 5.0, 7.0, 5.5, 4.5] \\ S_2 &= [7.0, 8.0, 8.5, 7.4, 8.9, 6.7, 9.0, 8.4, 7.8, 5.3, 8.1] \end{aligned} \tag{13}$$

   Determine el intervalo al 95% de nivel de confianza. ¿Pertenecen las muestras a la misma población?

$$IC_{95\%} = \Bigg[-3.16 < \mu_1 - \mu_2 < -1.60\Bigg] \tag{14}$$

   Las muestras no pertenecen a la misma población, presenta una desviación a $\sim 4\sigma$.

4. Una moneda equilibrada es lanzada 500 veces. Sea X la variable aleatoria asociada al número de soles que aparecen en los n lanzamientos. Usando la distribución Binomial y la aproximación a la normal, estime:

   a) $\mathbb{P}(X \leq 260.) = 0.8261$ Binomial, 0.8144 Normal.

   b) $\mathbb{P}(230 \leq X \leq 260) = 0.926$ Binomial, 0.926 Normal.

5. Se estudia la intención de voto de un candidato A a la presidencia de un país. Para ello, se toma una muestra aleatoria de 800 habitantes y 325 de ellos responden a la intención de votar por A. Encontrar el intervalo de confianza a 98%, para la proporción de votantes con intensión de voto por el candidato A.

$$IC_{98\%} = \Big[0.365 < P < 0.446 \Big] \tag{15}$$

6. Una empresa de equipo médico quiere saber la proporción de clientes que prefieren su marca. El gerente afirma que el 30% de los clientes del mercado prefieren sus productos. Una muestra de 100 clientes indicó que 24 usan dicha marca. Calcule el intervalo de confianza al 90% e indicar si el gerente tiene la razón.

$$IC_{90\%} = \Big[0.169 < P < 0.310 \Big] \tag{16}$$

   El gerente está en lo correcto dentro de la confianza de la hipótesis nula.

1.1.5 Intervalo para la varianza

El estimador de la varianza muestral está dada por:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{17}$$

Este estadístico es un estimador insesgado de la varianza poblacional puesto que:

$$\mathbb{E}(s^2) = \sigma^2 \tag{18}$$

Con base a los criterios de la ley de los grandes números y suponiendo que las muestras siguen una distribución normal, $X \sim N(\mu, \sigma^2)$. La varianza muestral como variable aleatoria sigue una función de distribución $\chi^2_{n-1}(s)$ de $n-1$ grados de libertad.

$$(n-1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-1} \tag{19}$$

En problemas donde la varianza poblacional no es conocida, la distribución permite encontrar intervalos de confianza para la varianza poblacional; dada la varianza muestral.

1.1.6 Ejercicios

1. Demuestre que $\mathbb{E}(s^2) = \sigma^2$. *Hint:* Recuerde que $Var(x) = \sigma^2$.
2. Genere una muestra de números $x \sim N(0, 9)$ de tamaño 100.
   a) Escriba las funciones para calcular la varianza muestral y χ^2_{n-1}.
   b) Con el método de re-muestreo (Bootstrapping) genere ($N = 50000$ toy experiments) muestras con reemplazo para encontrar la distribución de la varianza.
   c) Dibuje el histograma y la función de distribución de probabilidad χ^2_{n-1}, usando `chi2.pdf(x_,df=n-1)`.
   d) Calcule el valor medio de la distribución usando el histograma y usando `chi2.mean(df=df)`.
   e) Calcule el valor medio de la varianza poblacional (debería ser cercano a 9).

f) Calcule el intervalo de confianza de este estimador a un nivel de confianza del 95% (significancia $\alpha = 0.05$).

$$IC_{95\%} = \left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}} \right] \tag{20}$$

3. A un grupo de individuos en un hospital se sometió a un medicamento experimental, al final se midió el nivel de glóbulos blancos en sangre. En una escala específica los resultados fueron los siguientes:

$$S = [6.0, 6.4, 7.0, 5.8, 6.0, 5.8, 5.9, 6.7, 6.1, 6.5, 6.3, 5.8] \tag{21}$$

Suponiendo que la población es normalmente distribuida, determine un intervalo de confianza al 95% para la varianza poblacional del nivel de colesterol.

$$IC_{95\%} = \left[0.0770 < \sigma^2 < 0.4428 \right] \tag{22}$$

4. Un fabricante de parte de automóvil garantiza que sus baterías duran en promedio 3 años con una desviación estándar de 1 año, si 5 baterías presentan una varianza de 0.815. Realice su estudio al 99% de nivel de confianza.

a) ¿Está la observación de acuerdo con la afirmación del fabricante de que la desviación estándar es de 1 año?

b) ¿Cuál es el p-value de la observación? $p = 0.620$.

c) ¿Cuál es valor límite (crítico) que puede tener la varianza de las baterías para mantener la afirmación del fabricante? $S^2_{up} = 3.715$.

5. Encuentre la probabilidad de que una muestra aleatoria de 30 observaciones de una población normal con varianza $\sigma^2 = 10$. Tenga una varianza muestral:

(a) $\mathbb{P}(S^2 \geq 9.) = 0.379$

(b) $\mathbb{P}(3.5 \leq S^2 \leq 11.) = 0.675$

(c) Calcule el p-value de observar $S^2 = 15$, $p = 0.04$