Tema 7. Estimación por intervalos

En este tema:

- Intervalos de Confianza: definición e interpretación frecuentista.
- Intervalos de confianza para medias y varianzas en poblaciones normales: casos de una y dos poblaciones.
- Intervalos de confianza en muestras grandes.
- Determinación del tamaño muestral.

Estimador por intervalos de confianza

Sea $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ una muestra aleatoria de una población X con función de distribución $F_{\boldsymbol{\theta}}$ donde $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ es un vector de parámetros. Un estimador por intervalos de confianza de θ_i al nivel de confianza $1 - \alpha$ es una función que a la muestra $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ le hace corresponder un intervalo
$(T_1(\boldsymbol{x}), T_2(\boldsymbol{x})) = \left(T_1(x_1, x_2, \ldots, x_n), T_2(x_1, x_2, \ldots, x_n)\right)$ que satisface:

$$P\left(\theta_i \in (T_1(\boldsymbol{X}), T_2(\boldsymbol{X}))\right) = 1 - \alpha$$

para cada $\theta_i \in \Theta_i$

- Notar que $(T_1(\boldsymbol{X}), T_2(\boldsymbol{X})) \neq (T_1(\boldsymbol{x}), T_2(\boldsymbol{x}))$
- Se dice $(T_1(\boldsymbol{x}), T_2(\boldsymbol{x}))$ es un intervalo de confianza de θ_i

Estimador por intervalos de confianza

Ejemplo Sea $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ una muestra aleatoria simple de una población $X \sim \mathcal{N}(\mu, \sigma^2)$ con σ^2 conocida. Hallar un estimador por intervalos de confianza para la media, μ.

- Sabemos que $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
- Por tanto $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
- Entonces $P\Big(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\Big) = 1-\alpha$

Un intervalo de confianza para μ es

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Otros intervalos son

$$\left(\bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}}, +\infty\right)$$

$$\left(-\infty, \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right)$$

Cantidad pivotal

Sea $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ una muestra aleatoria de una población X con función de distribución $F_{\boldsymbol{\theta}}$ donde $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ es un vector de parámetros. Una cantidad pivotal para θ_i es una función $C(X_1, X_2, \ldots, X_n, \theta_i)$ tal que su distribución no depende de θ_i.

Ejemplo

- Sabemos que $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
- Entonces, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ es una cantidad pivotal para μ

Método de la cantidad pivotal

Si tenemos una cantidad pivotal, para construir intervalos de confianza se puede utilizar el siguiente procedimiento:

- Obtener la distribución de la cantidad pivotal
- Obtener C_1 y C_2 tales que $P\left(C_1 < C(X_1, X_2, \ldots, X_n, \theta_i) < C_2\right) = 1 - \alpha$
- Despejar θ_i de las desigualdades $C_1 < C(X_1, X_2, \ldots, X_n, \theta_i) < C_2$, esto es, lograr $T_1(X_1, X_2, \ldots, X_n) < \theta_i < T_2(X_1, X_2, \ldots, X_n)$

Finalmente, $\left(T_1(x_1, x_2, \ldots, x_n), T_2(x_1, x_2, \ldots, x_n)\right)$ es un intervalo para θ_i.

Método de la cantidad pivotal

Ejemplo Supongamos que los rendimientos de las acciones de la empresa `SEGURA.SL` siguen una distribución normal de media μ euros y varianza $\sigma^2 = 1$. Se toma una m.a.s. de $n = 20$ rendimientos y se tiene

$$5.29,\ 3.66,\ 5.71,\ 6.62,\ 4.30,\ 5.85,\ 6.25,\ 3.40,\ 3.55,\ 5.57,$$

$$4.60,\ 5.69,\ 5.81,\ 5.71,\ 6.29,\ 5.66,\ 6.19,\ 3.79,\ 4.98,\ 4.84$$

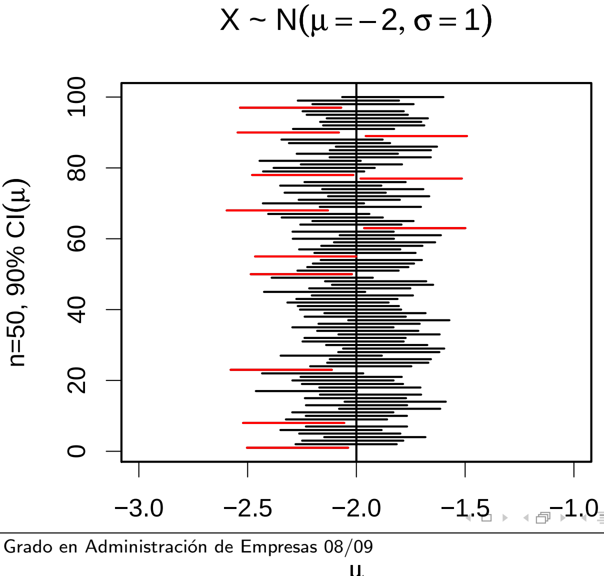
a) Calcular un intervalo de confianza al 90% para el rendimiento promedio de esta empresa.

$$\bar{x} = \frac{1}{20}\Big(5.29 + 3.66 + \cdots + 4.84\Big) = 5.188$$

$$\left(\bar{x} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = \left(5.188 \mp 1.645 \times \frac{1}{\sqrt{20}}\right) = (4.6678, 5.7082)$$

- ¿$P(\mu \in (4.6678,\ 5.7082))$?
- ¿$\mu \in (4.6678,\ 5.7082)$?

Interpretación frecuentista del intervalo de confianza



Cantidades pivotales en la distribución normal

Lema de Fisher Sean $X_1, X_2, \ldots, X_n$ v.a. i.i.d. $\mathcal{N}(\mu, \sigma^2)$. Entonces

- $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
- $\frac{(n-1)}{\sigma^2} s^2 \sim \chi^2_{n-1}$
- $\bar{X}$ y s^2 son independientes

Corolario

- $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ es una cantidad pivotal para μ

Útil si conocemos σ^2.

- $\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$ es una cantidad pivotal para μ
- $\frac{(n-1)}{\sigma^2} s^2$ es una cantidad pivotal para σ^2

I.C. en la distribución normal

Intervalos de confianza de nivel $1-\alpha$ para μ

- Con σ conocida

$$I = \left(\bar{x} \mp z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right)$$

- Con σ desconocida

$$I = \left(\bar{x} \mp t_{n-1;\alpha/2}\,\frac{s}{\sqrt{n}}\right)$$

Intervalo de confianza de nivel $1-\alpha$ para σ^2

$$I = \left(\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}\right)$$

I.C. en la distribución normal

Ejemplo Supongamos que los rendimientos de las acciones de la empresa `SEGURA.SL` siguen una distribución normal de media μ euros y varianza σ^2, ambas desconocidas.

a) Calcular un intervalo de confianza al 90% para el rendimiento promedio de esta empresa

$$\begin{aligned}
\bar{x} &= 5.188 \\
s^2 &= \frac{1}{19}\left((5.29-5.188)^2+(3.66-5.188)^2+\cdots+(4.84-5.188)^2\right) \\
&= 0.9929 \\
I &= \left(\bar{x} \mp t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}\right) = (5.188 \mp 1.729\sqrt{0.9929/20}) \\
&= (4.6432, 5.7328)
\end{aligned}$$

- ¿$P(\mu \in (4.6432, 5.7328))$?
- ¿$\mu \in (4.6432, 5.7328)$?

I.C. en la distribución normal

Ejemplo

b) Calcular un intervalo de confianza al 90% para la varianza del rendimiento.

$$I = \left(\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}} \right)$$

$$= \left(\frac{19 \times 0.9929}{30.14}, \frac{19 \times 0.9929}{10.12} \right) = (0.6259, 1.8641)$$

- ¿$P(\sigma^2 \in (0.6259, 1.8641))$?
- ¿$\sigma^2 \in (0.6259, 1.8641)$?
- ¿$1 \in (0.6259, 1.8641)$?

I.C. en muestras grandes - 1

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes siendo el tamaño de muestra suficientemente grande ($n > 30$). Entonces la distribución de $\bar{X}$ es aproximadamente una $\mathcal{N}(\mu, \frac{s^2}{n})$.

- $\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$ es una cantidad pivotal para μ
- Intervalos de confianza de nivel $1-\alpha$ para μ (con $n > 30$)

$$I = \left(\bar{x} \mp z_{\alpha/2}\,\frac{s}{\sqrt{n}}\right)$$

I.C. en muestras grandes - 1

Ejemplo Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes siendo el tamaño de $n > 30$ de una Bernoulli(p), entonces la distribución de $\bar{X} = \hat{p}$ es aproximadamente una $\mathcal{N}(p, \frac{p(1-p)}{n})$.

- $\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ es una cantidad pivotal para p
- Intervalos de confianza de nivel $1-\alpha$ para p (con $n > 30$)

$$I = \left(\hat{p} \mp z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

I.C. en muestras grandes - 1

Ejemplo Un partido político pretende conocer la intención de voto de cara a las próximas elecciones. Para ello encarga un sondeo sobre un total de 230 personas, de las que 69 contestan que le votarían.

a) Hallar un intervalo de confianza al 90% para la verdadera proporción poblacional, indicando las hipótesis asumidas.

Hipótesis asumidas: M.A.S de una Bernoulli(p), y n grande.

$$\begin{aligned}
\widehat{p} &= \frac{69}{230} = 0.3 \\
\alpha &= 0.1 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow \\
z_{\alpha/2} &= 1.645 \\
I &= \left(\hat{p} \mp z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \\
&= (0.3 \mp 1.645 \times 0.0302) = (0.3 \mp 0.049679) \\
&= (0.2503, 0.3497)
\end{aligned}$$

Cálculo del nivel de confianza

Ejemplo

b) Si el intervalo resultó ser $(0.243, 0.357)$, ¿cuál fue el nivel de confianza elegido?

- La longitud del intervalo $= 0.357 - 0.243 = 0.114$
- $$\begin{aligned} 0.114 &= 2z_{\alpha/2}\sqrt{\frac{0.3(1-0.3)}{230}} \Rightarrow \\ z_{\alpha/2} &= 1.8874 \end{aligned}$$
- $$\begin{aligned} \alpha/2 &= 1 - 0.9706 \Rightarrow \\ 1-\alpha &= 0.9412 \end{aligned}$$

Se trata de un intervalo al 94.12%

Cálculo del tamaño muestral

Ejemplo
c) Si el partido quisiera un intervalo de confianza al 90% cuya longitud no excediera de 0.15, ¿cuál sería el tamaño muestral necesario?

- La longitud del intervalo longitud $= 2z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$
- El intervalo de confianza alcanzará su longitud máxima cuando $\widehat{p} = 0.5$, de modo que calcularemos n para este caso que es el más desfavorable. En ese caso, $\text{longitud } = 2z_{\alpha/2}\sqrt{\frac{1}{4n}}$
- $1 - \alpha = 0.9 \Rightarrow \alpha/2 = 0.05 \Rightarrow z_{\alpha/2} = 1.645$
- $0.15 = 2 \times 1.645\sqrt{\frac{1}{4n}} \Rightarrow \sqrt{n} = \frac{1.645}{0.15} = 10.967 \Rightarrow n = 10.967^2 = 120.28$

Por tanto se tomaría $n = 121$.

Cálculo de error de estimación

El error de una estimación por intervalos de confianza de nivel $1-\alpha$ es la semiamplitud del intervalo obtenido.

Ejemplo Sea $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ una muestra aleatoria simple de una población $X \sim \mathcal{N}(\mu, \sigma^2)$ con $\sigma^2 = 5^2$. ¿cuál debe ser el mínimo tamaño muestral para que el error de estimación no sea superior a 0.5 con nivel de confianza del 95%?

- Sabemos que la semiamplitud el intervalo es $z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
- Imponemos que $z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq 0.5$
- Sustituyendo los valores $1.96 \frac{5}{\sqrt{n}} \leq 0.5$

Por tanto, $384.16 \leq n \Rightarrow n_{\text{min}} = 385$.

Cantidades pivotales en dos pobl. con distribución normal

Proposición Sean $X_1, X_2, \ldots, X_{n_1}$ v.a. i.i.d. $\mathcal{N}(\mu_1, \sigma_1^2)$ e $Y_1, Y_2, \ldots, Y_{n_2}$ v.a. i.i.d. $\mathcal{N}(\mu_2, \sigma_2^2)$ e independientes entre si. Entonces

- $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0,1)$
- Si $\sigma_1 = \sigma_2$

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2} \text{ donde}$$

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

- Si $\sigma_1 \neq \sigma_2$

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = t_f \text{ donde}$$

$$(entero)\ f \approx \frac{(s_1^2/n_1+s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$$

Cantidades pivotales en dos pobl. con distribución normal

Colorario

- $$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \text{ es una cantidad pivotal para } \mu_1-\mu_2$$

  Útil si conocemos σ_1 y σ_2

- $$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{s_p^2(\frac{1}{n_1}+\frac{1}{n_2})}} \text{ es una cantidad pivotal para } \mu_1-\mu_2$$

  Válida si $\sigma_1=\sigma_2$

- $$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \text{ es una cantidad pivotal para } \mu_1-\mu_2$$

  Válida si $\sigma_1\neq\sigma_2$

I.C. en dos poblaciones con distribución normal

Intervalos de confianza de nivel $1-\alpha$ para $\mu_1 - \mu_2$

- σ_1 y σ_2 conocidas

$$I = \left(\bar{x} - \bar{y} \mp z_{\alpha/2} \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} \right)$$

- σ_1 y σ_2 desconocidas e iguales

$$I = \left(\bar{x} - \bar{y} \mp t_{n_1+n_2-2;\alpha/2}\, s_p \sqrt{1/n_1 + 1/n_2} \right)$$

- σ_1 y σ_2 desconocidas y diferentes

$$I = \left(\bar{x} - \bar{y} \mp t_{f;\alpha/2} \sqrt{s_1^2/n_1 + s_2^2/n_2} \right)$$

I.C. en dos poblaciones con distribución normal

Proposición

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$$

Corolario

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \text{ es una cantidad pivotal para el cociente } \frac{\sigma_1^2}{\sigma_2^2}$$

Intervalos de confianza de nivel $1-\alpha$ para σ_1^2/σ_2^2

$$I = \left(\frac{s_1^2/s_2^2}{F_{n_1-1,n_2-1;\alpha/2}}, \frac{s_1^2}{s_2^2} F_{n_2-1,n_1-1;\alpha/2} \right)$$

I.C. en dos poblaciones con distribución normal

Ejemplo Queremos ver mediante un intervalo de confianza si los países emergentes tienen en media la misma inflación que los países desarrollados. Para eso, tomamos la inflación en 10 países emergentes y en 10 países desarrollados

Desarrollados	3.99	4.07	3.70	1.79	5.30	3.47	2.39	3.33	4.14	3.11
Emergentes	4.73	5.01	5.07	4.66	4.49	4.00	4.33	5.14	3.15	3.46

a) Calcular un intervalo de confianza para el cociente de varianzas

b) ¿Es posible calcular un I.C. para la diferencia de medias? ¿a qué conclusiones llegamos?

c) Comentar los supuestos que hay que hacer.

I.C. en muestras grandes - 2

Sean $(X_1, X_2, \ldots, X_{n_1})$ e $(Y_1, Y_2, \ldots, Y_{n_2})$ variables aleatorias i.i.d. e independientes entre si, siendo los tamaños de muestra suficientemente grandes $(n_1, n_2 > 30)$. Entonces la distribución de $\bar{X} - \bar{Y}$ es aproximadamente una $\mathcal{N}\left(\mu_1 - \mu_2, \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)$

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\left(\frac{s_1^2}{n} + \frac{s_2^2}{m}\right)^{1/2}} \text{ es una cantidad pivotal para } \mu_1 - \mu_2$$

Intervalos de confianza de nivel $1 - \alpha$ para μ (con $n_1, n_2 > 30$)

$$I = \left(\bar{x} - \bar{y} \mp z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$$

I.C. en muestras grandes - 2

Ejemplo Sean $(X_1, X_2, \ldots, X_{n_1}) \sim$ Bernoulli(p_1) e $(Y_1, Y_2, \ldots, Y_{n_2}) \sim$ Bernoulli(p_2) variables aleatorias i.i.d. e independientes entre si $(n_1, n_2 > 30)$. Entonces la distribución de $\hat{p}_1 - \hat{p}_2$ es aproximadamente una $\mathcal{N}\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)^{1/2}} \text{ es una cantidad pivotal para } p_1 - p_2$$

Intervalos de confianza de nivel $1 - \alpha$ para $p_1 - p_2$ (con $n_1, n_2 > 30$)

$$I = \left(\hat{p}_1 - \hat{p}_2 \mp z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$$

I.C. en muestras grandes - 2

Ejemplo En un sondeo de opinión llevado a cabo en el año 2002 con 1000 individuos en las ciudades de Murcia y Cartagena se encontró que el 55% de ellos eran favorables a la construcción de un nuevo aeropuerto.
Dicho sondeo se repitió en el año 2003 tras una intensa campaña publicitaria, encontrándose esta vez un 60% de individuos favorables a esta opción, entre 1500 encuestados. Se pide:
a) Construir un intervalo de confianza de nivel 95% para la proporción p_1 de individuos favorables a la construcción del nuevo aeropuerto, de entre los encuestados en el año 2002. ¿Podría afirmarse al nivel de significación del 5% que sólo la mitad de los encuestados preferían un nuevo aeropuerto?
b) Planteando los supuestos necesarios, construir un intervalo de confianza al 95% para la diferencia $p_2 - p_1$ de las proporciones de individuos favorables en los años 2003 y 2002. ¿Podría afirmarse al nivel de significación del 5% que la campaña publicitaria en favor de la construcción del nuevo aeropuerto pudo tener efecto?

I.C. en muestras grandes - 2

Ejemplo
a) I.C. de nivel 95% para p_1

$$I = \left(\hat{p}_1 \mp 1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}}\right) = (0.5192,\ 0.5808)$$

Como dicho intervalo no contiene el valor $p = 0.5$ no podemos afirmar al nivel de significación del 5% que sólo la mitad de los encuestados eran favorables al nuevo aeropuerto.

I.C. en muestras grandes - 2

Ejemplo

b) Si suponemos que las muestras de encuestados en los años 2002 y 2003 son independientes, se tendrá I.C. de nivel 95% para $p_2 - p_1$

$$\begin{aligned} I &= \left(\hat{p}_2 - \hat{p}_1 \mp 1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right) \\ &= \left(0.6 - 0.55 \mp 1.96\sqrt{\frac{0.55(0.45)}{1000} + \frac{0.6(0.4)}{1500}}\right) \\ &= (0.05 \mp 0.0396) = (0.0104, 0.0896) \end{aligned}$$

Como este intervalo sólo contiene valores positivos, podemos afirmar al nivel de significación del 5% que la campaña publicitaria en favor del nuevo aeropuerto pudo tener efecto.