Tema 4: Estimación por intervalo (Intervalos de Confianza)

(a partir del material de A. Jach (`http://www.est.uc3m.es/ajach/`) y A. Alonso (`http://www.est.uc3m.es/amalonso/`))

- Planteamiento del problema: IC para la media de una población normal con varianza conocida
- Método de la cantidad pivotal
- IC's en una población:
  - Población normal
  - Población no normal pero con muestras de tamaño grande
- IC's en dos poblaciones independientes:
  - Poblaciones normales
  - Poblaciones no normales pero con muestras de tamaño grande
- Determinación del tamaño de muestra

Planteamiento del problema

Ejemplo 1.

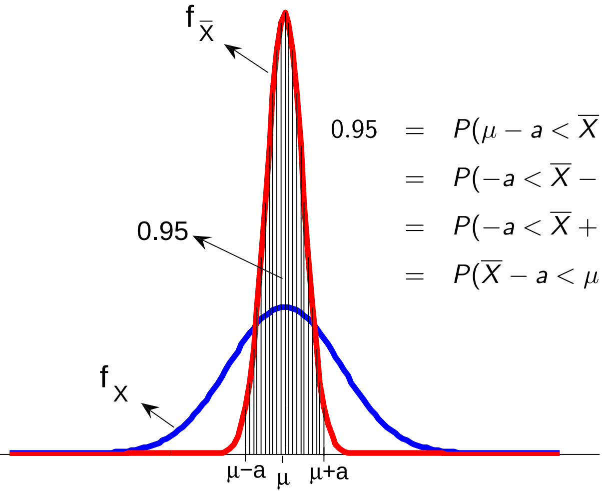
Supongamos que la altura (en cm) de los estudiantes de la UC3M es una v.a. X con distribución $N(\mu, 5)$. Con el objetivo de estimar μ se toma una m.a.s. de 100 estudiantes y se obtiene $\overline{x} = 156.8$.

Sabemos que si tomamos otra muestra distinta, el valor de $\overline{x}$ va a ser distinto.

¿Cómo de "fiable" es el resultado obtenido?

La estimación por intervalo (o construcción de intervalos de confianza) consiste en acotar el valor del parámetro entre dos valores aleatorios con alguna garantía expresada en términos de probabilidad.

Planteamiento del problema



$$
\begin{aligned}
0.95 &= P(\mu - a < \overline{X} < \mu + a) && \\
&= P(-a < \overline{X} - \mu < a) && \text{restando } \mu \\
&= P(-a < \overline{X} + \mu < a) && x < y \Leftrightarrow -x > -y \\
&= P(\overline{X} - a < \mu < \overline{X} + a) && \text{sumando } \overline{X}
\end{aligned}
$$

Intervalos de confianza

Definición 1.

Sea $X_1, \ldots, X_n$ una m.a.s. de una v.a. X cuya distribución depende de un parámetro θ. Un estimador por intervalos de confianza de θ al $(1-\alpha)100\%$, es una función que a cada muestra particular $(x_1, \ldots, x_n)$ le asigna un intervalo $(T_1(x_1, \ldots, x_n), T_2(x_1, \ldots, x_n))$, de forma que

$$P\left(T_1(X_1, \ldots, X_n) < \theta < T_2(X_1, \ldots, X_n)\right) = 1 - \alpha$$

Para cada muestra particular, $(T_1(x_1, \ldots, x_n), T_2(x_1, \ldots, x_n))$ es un intervalo de confianza.

El nivel de significación de un IC es α $\quad (\alpha = 0.05, 0.025, 0.01, \ldots)$.

El nivel de confianza de un IC es $1-\alpha$ $\quad (1-\alpha = 0.95, 0.975, 0.99, \ldots)$.

Observación:

$$(T_1(X_1, \ldots, X_n), T_2(X_1, \ldots, X_n)) \neq (T_1(x_1, \ldots, x_n), T_2(x_1, \ldots, x_n))$$

IC para la media de una pobación normal

Sea $X \sim N(\mu, \sigma)$ una v.a., donde σ es conocida y queremos estimar μ.

Sea $X_1, \ldots, X_n$ una m.a.s. de tamaño n de X. Sabemos que

$$\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Observación: la distribución de $\overline{X}$ es exacta, y no asintótica, ya que la distribución de X es normal. Esto es cierto para cualquier tamaño de muestra (no estamos aplicando el TCL).

Buscamos un intervalo de confianza para μ al $(1-\alpha)100\%$, es decir, buscamos dos estadísticos $T_1(X_1, \ldots, X_n)$ y $T_2(X_1, \ldots, X_n)$ tales que

$$P\left(T_1(X_1, \ldots, X_n) < \mu < T_2(X_1, \ldots, X_n)\right) = 1 - \alpha$$

IC para la media de una pobación normal

Estandarizando $\overline{X}$, tenemos que $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$. Entonces:

$$1-\alpha = P(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2})$$

$z_{\alpha/2}$ es el valor que verifica que $P(Z > z_{\alpha/2}) = \dfrac{\alpha}{2}$, con $Z \sim N(0,1)$.

Tenemos que:

$$\begin{aligned} 1-\alpha &= P(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = P(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \\ &= P(-\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \\ &= P(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \end{aligned}$$

IC para la media de una pobación normal

Tenemos que:

$$P(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Por tanto, los dos estadísticos que definen el IC al $(1-\alpha)100\%$ para μ en una población normal con σ conocida son:

$$\begin{aligned} T_1(X_1,\ldots,X_n) &= \overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \\ T_2(X_1,\ldots,X_n) &= \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \end{aligned}$$

En la práctica: tendremos una muestra particular $(x_1,\ldots,x_n)$ y un valor particular $\overline{x}$. El IC para μ al $(1-\alpha)100\%$ será:

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

IC para la media de una pobación normal

Ejemplo 1 (cont.)

La altura (en cm) de los estudiantes de la UC3M es una v.a. X con distribución $N(\mu, 5)$. Hemos tomado una m.a.s. de 100 estudiantes y hemos obtienido $\overline{x} = 156.8$.

El IC para μ al 95% será:

$$\left(\overline{x} - z_{0.025}\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x} + z_{0.025}\frac{\sigma}{\sqrt{n}}\right)$$

Como $n = 100$, $z_{0.025} = 1.96$, $\overline{x} = 156.8$ y $\sigma = 5$, obtenemos el intervalo

$$\begin{aligned}\left(156.8 - 1.96\frac{5}{10}\ ,\ 156.8 + 1.96\frac{5}{10}\right) &= (156.8 - 0.98\ ,\ 156.8 + 0.98) \\ &= (155.82\ ,\ 157.78)\end{aligned}$$

Interpretación de un Intervalo de Confianza

Dos interpretaciones de un IC de μ al 95%:

- ANTES DE TOMAR UNA MUESTRA PARTICULAR, la probabilidad de que el IC para la muestra elegida contenga a μ es 0.95.

- Si tomásemos por ejemplo 500 muestras particulares, aproximadamente el 95% de ellas nos daría un IC que contendría a μ y sólo el 5% de esas muestras nos daría un IC que no contendría a μ.

Una interpretación FALSA:

- Es falso afirmar que la probabilidad de que μ esté en un IC concreto, por ejemplo $(155.82\ ,\ 157.78)$, es 0.95.

Interpretación de un Intervalo de Confianza

Ejemplo 2.

En una empresa se sabe que el número mensual de horas extra que realiza un trabajador sigue una distribución normal con una varianza igual a 16.

Mediante un muestreo aleatorio simple se obtienen los datos correspondientes a 15 trabajadores:

$$23 \quad 32 \quad 15 \quad 8 \quad 2 \quad 13 \quad 18 \quad 22 \quad 6 \quad 26 \quad 0 \quad 12 \quad 4 \quad 16 \quad 11$$

Obtener un IC al 90% *para el número medio de horas extra realizadas por un trabajador de la empresa.*

Cantidad Pivotal

En el cálculo de un IC para la media de una pobación normal, hemos utilizado la v.a.

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Su distribución es siempre $N(0,1)$ independientemente del valor de μ.

Es decir, aunque μ sea desconocido, nosotros conocemos su distribución.

$\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ recibe el nombre de pivote o cantidad pivotal.

Definición 2.

Sea $X_1, \ldots, X_n$ una m.a.s. de una v.a. X cuya distribución depende de un parámetro θ. Un pivote o una cantidad pivotal es una función $C(X_1, \ldots, X_n, \theta)$, de la muestra aleatoria y de θ, cuya distribución no depende de θ.

Método de la Cantidad Pivotal

Para construir un IC para θ vamos a buscar una cantidad pivotal, es decir, una función $C(X_1, \ldots, X_n, \theta)$ cuya distribución no depende de θ.

Una vez que la tengamos, los pasos serán:

1. Obtener la distribución de la cantidad pivotal.
2. Obtener C_1 y C_2 tales que $P\left(C_1 < C(X_1, \ldots, X_n, \theta) < C_2\right) = 1 - \alpha$.
3. Despejar θ de las desigualdades $C_1 < C(X_1, \ldots, X_n, \theta) < C_2$, hasta obtener $P\left(T_1(X_1, \ldots, X_n) < \theta < T_2(X_1, \ldots, X_n)\right) = 1 - \alpha$.

Para una muestra particular $(x_1, \ldots, x_n)$, el IC para θ será

$$\left(\, T_1(x_1, \ldots, x_n)\,,\; T_2(x_1, \ldots, x_n)\,\right)$$

IC's en una población normal

IC para σ^2 en una población normal

Sea $X_1, \ldots, X_n \sim N(\mu, \sigma)$ una m.a.s. con μ y σ desconocidas. Si queremos aplicar el método de la cantidad pivotal para obtener un IC para σ^2 necesitamos una cantidad pivotal, es decir, una función de la m.a.s. y de σ^2 cuya distribución sea conocida y no dependa de σ^2.

Por el Lema de Fisher sabemos que

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$$

donde S^2 es la cuasi-varianza muestral.

La distribución de $\dfrac{n-1}{\sigma^2} S^2$ no depende de σ^2. Por lo tanto

$$C(X_1, \ldots, X_n, \sigma^2) = \frac{n-1}{\sigma^2} S^2$$

es una cantidad pivotal para σ^2.

IC's en una población normal

IC para σ^2 en una población normal (cont.)

Aplicamos el método de la cantidad pivotal a $C(X_1, \ldots, X_n, \sigma^2) = \frac{n-1}{\sigma^2} S^2$.

1. Obtener la distribución de la cantidad pivotal: $\frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$
2. Obtener C_1 y C_2 tales que $P\left(C_1 < C(X_1, \ldots, X_n, \sigma^2) < C_2\right) = 1 - \alpha$.

   La distribución χ^2 es asimétrica. Hay muchos valores C_1 y C_2 que verifican que

   $$1 - \alpha = P(C_1 < \chi^2_{n-1} < C_2)$$

   por ejemplo $C_1 = \chi^2_{n-1;1-\alpha/2}$ y $C_2 = \chi^2_{n-1;\alpha/2}$.

   (Para una distr. χ^2_n, denotamos $\chi^2_{n;\alpha}$ el valor t.q. $P(\chi^2_n > \chi^2_{n;\alpha}) = \alpha$).

   Por lo tanto

   $$1 - \alpha = P(\chi^2_{n-1;1-\alpha/2} < \frac{n-1}{\sigma^2} S^2 < \chi^2_{n-1;\alpha/2})$$

IC's en una población normal

IC para σ^2 en una población normal (cont.)

3. Despejar σ^2:

$$\chi^2_{n-1;1-\alpha/2} < \frac{n-1}{\sigma^2} S^2 < \chi^2_{n-1;\alpha/2}$$

$$\Leftrightarrow \sigma^2 \chi^2_{n-1;1-\alpha/2} < (n-1) S^2 < \sigma^2 \chi^2_{n-1;\alpha/2}$$

$$\Leftrightarrow \sigma^2 < \frac{n-1}{\chi^2_{n-1;1-\alpha/2}} S^2 \quad \text{y} \quad \sigma^2 > \frac{n-1}{\chi^2_{n-1;\alpha/2}} S^2$$

$$\Leftrightarrow \frac{n-1}{\chi^2_{n-1;\alpha/2}} S^2 < \sigma^2 < \frac{n-1}{\chi^2_{n-1;1-\alpha/2}} S^2$$

Por tanto

$$1 - \alpha = P\left(\frac{n-1}{\chi^2_{n-1;\alpha/2}} S^2 < \sigma^2 < \frac{n-1}{\chi^2_{n-1;1-\alpha/2}} S^2 \right)$$

IC's en una población normal

IC para σ^2 en una población normal (cont.)

Tenemos que:

$$P\left(\frac{n-1}{\chi^2_{n-1;\alpha/2}}S^2 < \sigma^2 < \frac{n-1}{\chi^2_{n-1;1-\alpha/2}}S^2\right) = 1-\alpha$$

Por tanto, los dos estadísticos que definen el IC al $(1-\alpha)100\%$ para σ^2 en una población normal son:

$$T_1(X_1,\dots,X_n) = \frac{n-1}{\chi^2_{n-1;\alpha/2}}S^2 \qquad T_2(X_1,\dots,X_n) = \frac{n-1}{\chi^2_{n-1;1-\alpha/2}}S^2$$

En la práctica: tendremos una muestra particular $(x_1,\dots,x_n)$ y un valor particular s^2. El IC para σ^2 al $(1-\alpha)100\%$ será:

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

Distribuciones asociadas con la distribución normal

Definición 3.

Sean $X_1, \ldots, X_n$ *n v.a. i.i.d. con distribución* $N(0,1)$*. Entonces, la v.a.*

$$W = \sum_{i=1}^{n} X_i^2$$

tiene distribución χ^2 *con n grados de libertad* ($W \sim \chi_n^2$).

Definición 4.

Sea X *una v.a. con distribución* $N(0,1)$ *y* W *una v.a. con distribución* χ_n^2 *e* **independiente de** X. *Entonces, la v.a.*

$$T = \frac{X}{\sqrt{W/n}}$$

tiene distribución t de Student con n grados de libertad ($T \sim t_n$).

Distribuciones asociadas con la distribución normal

Definición 5.

Sean W_1 una v.a. con distribución $\chi^2_{n_1}$ y W_2 una v.a. con distribución $\chi^2_{n_2}$ **independientes entre sí**. *Entonces, la v.a.*

$$F = \frac{W_1/n_1}{W_2/n_2}$$

tiene distribución F de Snedecor con n_1 y n_2 grados de libertad ($F \sim F_{n_1,n_2}$).

Propiedades:

- $F \sim F_{n_1,n_2} \quad \Rightarrow \quad \frac{1}{F} \sim F_{n_1,n_2}$
- Sea $F_{n_1,n_2;\alpha}$ el valor tal que $P(F_{n_1,n_2} > F_{n_1,n_2;\alpha}) = \alpha$. Se tiene que

$$F_{n_1,n_2;1-\alpha} = \frac{1}{F_{n_2,n_1;\alpha}}$$

IC's en una población normal

$X \sim N(\mu, \sigma)$. $X_1, \ldots, X_n$ m.a.s. de X.

Parámetro	Pivote	IC al $(1-\alpha)100\%$	Comentarios
μ	$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\left(\overline{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right)$	σ conocida
μ	$\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$	$\left(\overline{x} \pm t_{n-1;\alpha/2}\dfrac{s}{\sqrt{n}}\right)$	σ desconocida
σ^2	$\dfrac{n-1}{\sigma^2}S^2 \sim \chi^2_{n-1}$	$\left(\dfrac{n-1}{\chi^2_{n-1;\alpha/2}}s^2\, ,\ \dfrac{n-1}{\chi^2_{n-1;1-\alpha/2}}s^2\right)$	int. asimétrico

IC's en una población NO normal con muestras grandes

X v.a. con $E[X] = \mu$ y $Var[X] = \sigma^2$. $X_1, \ldots, X_n$ m.a.s. de X, n grande (TCL).

Parámetro	Pivote	IC al $(1-\alpha)100\%$	Comentarios
μ	$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \overset{A}{\sim} N(0,1)$	$\left(\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$	σ conocida
μ	$\frac{\overline{X}-\mu}{S/\sqrt{n}} \overset{A}{\sim} N(0,1)$	$\left(\overline{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}\right)$	σ desconocida
p	$\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n}} \overset{A}{\sim} N(0,1)$	$\left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$	$X \sim \mathcal{B}(p)$ $\hat{p} = \overline{x}$
λ	$\frac{\hat{\lambda}-\lambda}{\sqrt{\hat{\lambda}}/\sqrt{n}} \overset{A}{\sim} N(0,1)$	$\left(\hat{\lambda} \pm z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}}\right)$	$X \sim \mathcal{P}(\lambda)$ $\hat{\lambda} = \overline{x}$

IC's en una población NO normal con muestras grandes

X v.a. cuya distribución depende de un parámetro θ. $X_1, \dots, X_n$ m.a.s. de X, n grande.
Si $\hat{\theta}_{MV}$ es el EMV de θ, sabemos que:

$$\hat{\theta}_{MV} \overset{A}{\sim} N\left(\theta, \sqrt{Var.asint.de\ \hat{\theta}_{MV}}\right)$$

Por lo tanto

$$\frac{\hat{\theta}_{MV} - \theta}{\sqrt{Var.asint.de\ \hat{\theta}_{MV}}} \overset{A}{\sim} N(0,1)$$

es una cantidad pivotal para θ.

IC's al $(1-\alpha)100\%$:

$$\left(\hat{\theta}_{MV} \pm z_{\alpha/2}\sqrt{Var.asint.de\ \hat{\theta}_{MV}}\right)$$

IC's en dos poblaciones independientes

Ejemplo 3.

Se quiere comparar la inflación de los países emergentes con la de los países desarrollados. Para ello, se han recogido los datos de la inflación en 9 países emergentes y en 11 países desarrollados.

D	3,99	4,07	3,70	1,79	5,30	3,47	2,39	3,33	4,14	3,11	3,26
E	4,73	5,01	5,07	4,66	4,49	4	4,33	5,14	3,15		

Podemos realizar la comparación obteniendo IC's para la diferencia de medias y para el cociente de varianzas.

IC's en dos poblaciones normales independientes

$X \sim N(\mu_1, \sigma_1)$, $Y \sim N(\mu_2, \sigma_2)$ indep. $X_1, \ldots, X_{n_1}$ m.a.s. de X, $Y_1, \ldots, Y_{n_2}$ m.a.s. de Y.

Parámetro	Pivote	IC al $(1-\alpha)100\%$	Coment.
$\mu_1 - \mu_2$	$\dfrac{\overline{X} - \overline{Y} - \mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$\left(\overline{x} - \overline{y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$	σ_1, σ_2 conocidas
$\mu_1 - \mu_2$	$\dfrac{\overline{X} - \overline{Y} - \mu_1 - \mu_2}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$	$\left(\overline{x} - \overline{y} \pm t_{n_1+n_2-2;\frac{\alpha}{2}}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$	$\sigma_1 = \sigma_2$ descon.
$\mu_1 - \mu_2$	$\dfrac{\overline{X} - \overline{Y} - \mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_f$	$\left(\overline{x} - \overline{y} \pm t_{f;\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$	$\sigma_1 \neq \sigma_2$ descon.
σ_1^2/σ_2^2	$\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$	$\left(\dfrac{s_1^2/s_2^2}{F_{n_1-1,n_2-1;\frac{\alpha}{2}}}, \dfrac{s_1^2}{s_2^2}F_{n_2-1,n_1-1;\frac{\alpha}{2}}\right)$	IC asimét.

donde f es el entero más próximo a $\dfrac{(s_1^2/n_1+s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$ y $S_p^2 = \dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

IC's en dos poblaciones NO normales independientes con muestras grandes

Sean $X \sim \mathcal{B}(p_1)$, $Y \sim \mathcal{B}(p_2)$ independientes.
$X_1, \ldots, X_{n_1}$ m.a.s. de X e $Y_1, \ldots, Y_{n_2}$ m.a.s. de Y, n_1 y n_2 grandes (TCL).

Parámetro	Pivote	IC al $(1-\alpha)100\%$
$p_1 - p_2$	$\dfrac{\hat{p}_1 - \hat{p}_2 - p_1 - p_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \overset{A}{\sim} N(0,1)$	$\left(\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$

donde $\hat{p}_1 = \overline{x}$ y $\hat{p}_2 = \overline{y}$.

IC's en dos poblaciones independientes

Ejemplo 3 (cont.)

Queremos encontrar IC's para la diferencia de medias y el cociente de varianzas de la inflación en los países desarrollados y en los países emergentes.

X = tasa de inflación en los países desarrollados.

$X_1, \dots, X_{n_1}$ m.a.s. de X, n_1=11.

Y = tasa de inflación en los países emergentes.

$Y_1, \dots, Y_{n_2}$ m.a.s. de Y, n_2=9.

¿Qué hipótesis tenemos que hacer?

IC's en dos poblaciones independientes

Ejemplo 3 (cont.)

Como $n_1 = 11$ y $n_2 = 9$ no son lo suficientemente grande para poder aplicar el TCL, **supondremos que la tasa de inflación sigue una distribución normal**. Además **tenemos que suponer que X e Y son independientes**.

$X =$ tasa de inflación en los países desarrollados $\sim N(\mu_1, \sigma_1)$.

$Y =$ tasa de inflación en los países emergentes $\sim N(\mu_2, \sigma_2)$.

Como no nos dicen nada sobre las varianzas, asumimos que $\sigma_1 \neq \sigma_2$ y ambas son desconocidas.

IC al $(1 - \alpha)100\%$ para la diferencia de medias:

$$\left(\overline{x} - \overline{y} \pm t_{f;\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

con f el entero más próximo a $\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$.

IC's en dos poblaciones independientes

Ejemplo 3 (cont.)

En nuestro caso, para un IC al 90%:

$$\overline{x} = 3.5045 \qquad s_1^2 = \frac{1}{n_1 - 1}\left(\sum_{i=1}^{n_1} x_i^2 - n_1 \overline{x}^2\right) = 0.8650$$

$$\overline{y} = 4.5089 \qquad s_2^2 = \frac{1}{n_2 - 1}\left(\sum_{i=1}^{n_2} y_i^2 - n_2 \overline{y}^2\right) = 0.3967$$

$$f = \left[\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}\right]_{ent} = \left[\frac{(0.8650/11 + 0.3967/9)^2}{\frac{(0.8650/11)^2}{10} + \frac{(0.3967/9)^2}{8}}\right]_{ent} = [17.4853]_{ent} = 17$$

$$t_{f;\frac{\alpha}{2}} = t_{17;0.05} = 1.740$$

IC al 90% para la diferencia de medias:

$$\begin{aligned}
&\left(\overline{x} - \overline{y} \pm t_{f;\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right) \\
= \; &\left(3.5045 - 4.5089 \pm 1.740 \sqrt{\frac{0.8650}{11} + \frac{0.3967}{9}}\right) \\
= \; &(3.5045 - 4.5089 \pm 1.740 \cdot 0.3503) = (-1.6139\,,\, -0.3949)
\end{aligned}$$

IC's en dos poblaciones independientes

Ejemplo 3 (cont.)

IC al $(1-\alpha)100\%$ para el cociente de varianzas:

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

En nuestro caso, para un IC al 90%:

$$s_1^2 = 0.8650 \qquad s_2^2 = 0.3967$$

$$F_{n_1-1,n_2-1;\frac{\alpha}{2}} = F_{10,8;0.05} = 3.35 \qquad F_{n_2-1,n_1-1;\frac{\alpha}{2}} = F_{8,10;0.05} = 3.07$$

IC al 90% para el cociente de varianzas:

$$\begin{aligned} & \left(\frac{s_1^2/s_2^2}{F_{n_1-1,n_2-1;\frac{\alpha}{2}}}, \frac{s_1^2}{s_2^2} F_{n_2-1,n_1-1;\frac{\alpha}{2}} \right) \\ = \quad & \left(\frac{0.8650/0.3967}{3.35}, \frac{0.8650}{0.3967} 3.07 \right) = (0.6509\,,\,6.6941) \end{aligned}$$

Determinación del tamaño muestral

Definición 6.

Se llama error de estimación por intervalos de confianza *a la mitad de la amplitud del intervalo obtenido.*

Por ejemplo, en el caso de un IC al $(1-\alpha)100\%$ para la media de una población normal con varianza conocida,

$$\left(\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

el error de estimación es $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$.

Observaciones:

- Para n fijo: a mayor nivel de confianza, mayor error (a mayor confianza, menor precisión).
- Para α fijo: a mayor n, menor error (a mayor información, mayor precisión).
- Para un error fijo: a mayor n, mayor nivel de confianza (más información, más confianza).

Determinación del tamaño muestral

Ejemplo 1 (cont.)

X = altura (en cm) de los estudiantes de la UC3M $\sim N(\mu, 5)$.
Para n=100, obtuvimos el siguiente IC al 95% para μ: (155.82 , 157.78).

¿Cuál ha de ser el mínimo tamaño muestral para que el error de estimación no sea superior a 0.5 con un nivel de confianza de 0.95? ¿Y si sólo exigimos un nivel de confianza de 0.90?

Determinación del tamaño muestral

Ejemplo 1 (cont.)

X = altura (en cm) de los estudiantes de la UC3M $\sim N(\mu, 5)$.
Para n=100, obtuvimos el siguiente IC al 95% para μ: (155.82 , 157.78).

¿Cuál ha de ser el mínimo tamaño muestral para que el error de estimación no sea superior a 0.5 con un nivel de confianza de 0.95? ¿Y si sólo exigimos un nivel de confianza de 0.90?

IC al $(1-\alpha)100\%$ para μ: $\left(\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

$\alpha = 0.05 \Rightarrow z_{\alpha/2} = 1.96$:

$$\begin{aligned} Error < 0.5 \quad &\Leftrightarrow \quad z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.96\frac{5}{\sqrt{n}} < 0.5 \\ &\Leftrightarrow \quad 1.96\frac{5}{0.5} < \sqrt{n} \Leftrightarrow n > \left(1.96\frac{5}{0.5}\right)^2 = 384.16 \end{aligned}$$

n tiene que ser al menos 385.