Estimación por intervalos dos poblaciones

Jessica Nathaly Pulzara Mora
jessica.pulzara@udea.edu.co

Departamento de ingeniería de sistemas

UNIVERSIDAD
DE ANTIOQUIA

I.C. para la diferencia de medias

I.C diferencia de medias

Casos:

- Caso I: Poblaciones No normales, varianzas conocidas.
- Caso II: Poblaciones No normales, varianzas desconocidas.
- Caso IIIa: Poblaciones normales, varianzas desconocidas iguales.
- Caso IIIb: Poblaciones normales, varianzas desconocidas diferentes.

Caso 1

Sean $X_{11}, X_{12}, \ldots, X_{1n}$ y $X_{21}, X_{22}, \ldots, X_{2n}$ dos muestras que vienen de dos poblaciones independientes NO normales, con n_1 y n_2 grande, con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2 conocidas. Se desea construir un intervalo para $\mu_1 - \mu_2$. Tengamos en cuenta estos resultados:

$$E[\bar{X}_1 - \bar{X}_2] = E[\bar{X}_1] - E[\bar{X}_2]$$

$$V[\bar{X}_1 - \bar{X}_2] = V[\bar{X}_1] + V[\bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Por el TLC

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Entonces un I.C aproximado al $(1-\alpha)100\,\%$ para $\mu_1 - \mu_2$ es:

$$\bar{X}_1 - \bar{X}_2 \mp z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Caso 2

Sean $X_{11}, X_{12}, \ldots, X_{1n}$ y $X_{21}, X_{22}, \ldots, X_{2n}$ dos muestras que vienen de dos poblaciones independientes NO normales, con n_1 y n_2 grande, con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2 desconocidas.

Se pueden usar S_1^2 y S_2^2 ya que n_1 y n_2 son grandes, entonces:

$$\bar{X}_1 - \bar{X}_2 \mp z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Ejemplo

Dos instituciones educativas tienen dos métodos distintos para inscribir a sus alumnos. Se desea comparar el tiempo promedio que les toma a los estudiantes realizar el trámite de inscripción. En cada universidad se anotaron los tiempos de inscripción para 100 alumnos al azar. Las medias y las desviaciones estándar fueron las siguientes:

$$\bar{X}_1 = 50.2 \quad \bar{X}_2 = 52.9 \quad S_1 = 4.8 \quad S_2 = 5.4$$

Las universidades son distintas, independientes. ¿Existe diferencia real entre los tiempos medios para cada universidad?

Solución:

$$(-4.12, -1.28)$$

Como el intervalo no contiene al cero, los tiempos medios de las universidades son diferentes.

Ejemplo

Se hicieron pruebas de resistencia a la tensión a dos tipos distintos de varillas para alambres y se obtuvieron los siguientes resultados, en $kg\ mm^2$:

Grado	Tamaño muestral	Media muestral	Desviación Estándar
AISI-1064	$n = 129$	107.6	1.3
AISI-1078	$m = 129$	123.6	2.0

¿Indican estos datos que la resistencia promedio real del grado 1078 supera a la del grado 1064 en más de $10 kg\ mm^2$?

Solución:

$$(15.59, 16.41)$$

Como el intervalo se encuentra a la derecha de 10, $\mu_2 - \mu_1 > 10$. Por lo tanto, se confirman la sospechas con un 95 % de confianza.

Caso 3

Sean $X_{11}, X_{12}, \ldots, X_{1n}$ y $X_{21}, X_{22}, \ldots, X_{2n}$ dos muestras que vienen de dos poblaciones independientes normales, Si σ_1 y σ_2 son desconocidas, es necesario usar la distribución muestral que tenga en cuenta esa situación.
Aquí se distinguen dos casos:

- $\sigma_1 = \sigma_2$
- $\sigma_1 \neq \sigma_2$

Caso 3 varianzas iguales

Varianzas desconocida, $\sigma_1 = \sigma_2$

Por el TLC,

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

El I.C al $100(1 - \alpha)$ para $\mu_1 - \mu_2$ es:

$$\bar{X}_1 - \bar{X}_2 \mp t_{\alpha/2}(n_1 + n_2 - 2)S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Caso 3 varianzas diferentes

Varianzas desconocida, $\sigma_1 \neq \sigma_2$

Por el TLC,

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t(\nu)$$

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

El I.C al $100(1 - \alpha)$ para $\mu_1 - \mu_2$ es:

$$\bar{X}_1 - \bar{X}_2 \mp t_{\alpha/2}(\nu)\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Ejemplo

En dos ciudades se llevó a cabo una encuesta sobre el costo de vida para obtener el gasto promedio en alimentación en familias constituidas por cuatro personas. De cada ciudad, se seleccionó aleatoriamente una muestra de 20 familias y se observaron sus gastos semanales en alimentación. Las medias y las desviaciones fueron las siguientes:

$$\bar{X}_1 = 135 \quad \bar{X}_2 = 122 \quad S_1 = 15 \quad S_2 = 10$$

Suponiendo que las dos poblaciones son independientes, con distribución normal cada una, ¿existe una diferencia real entre μ_1 y μ_2?

Solución:

Si las varianzas son iguales:

$$(4.84, 21.16)$$

Como el intervalo no contiene al cero, existen diferencias entre el gasto semanal de las familias en las dos ciudades.

Si las varianzas son diferentes:

$$(4.81, 21.19)$$

Como el intervalo no contiene al cero, existen diferencias entre el gasto semanal de las familias en las dos ciudades.

I.C para el cociente de varianzas

Distribución F

La f.d.p para una variable que se distribuye $F(u, v)$ es:

$$f(x) = \frac{\Gamma\left(\frac{u+v}{2}\right)\left(\frac{u}{v}\right)^{u/2} x^{(u/2)-1}}{\Gamma\left(\frac{u}{2}\right)\Gamma\left(\frac{v}{2}\right)\left[\frac{u}{v}x+1\right]^{(u+v)/2}}$$

donde u son los grados de libertad del numerador y v los grados de libertad del denominador.

Para el caso que nos interesa, se sabe que:

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = F \sim f(n_1 - 1, n_2 - 1)$$

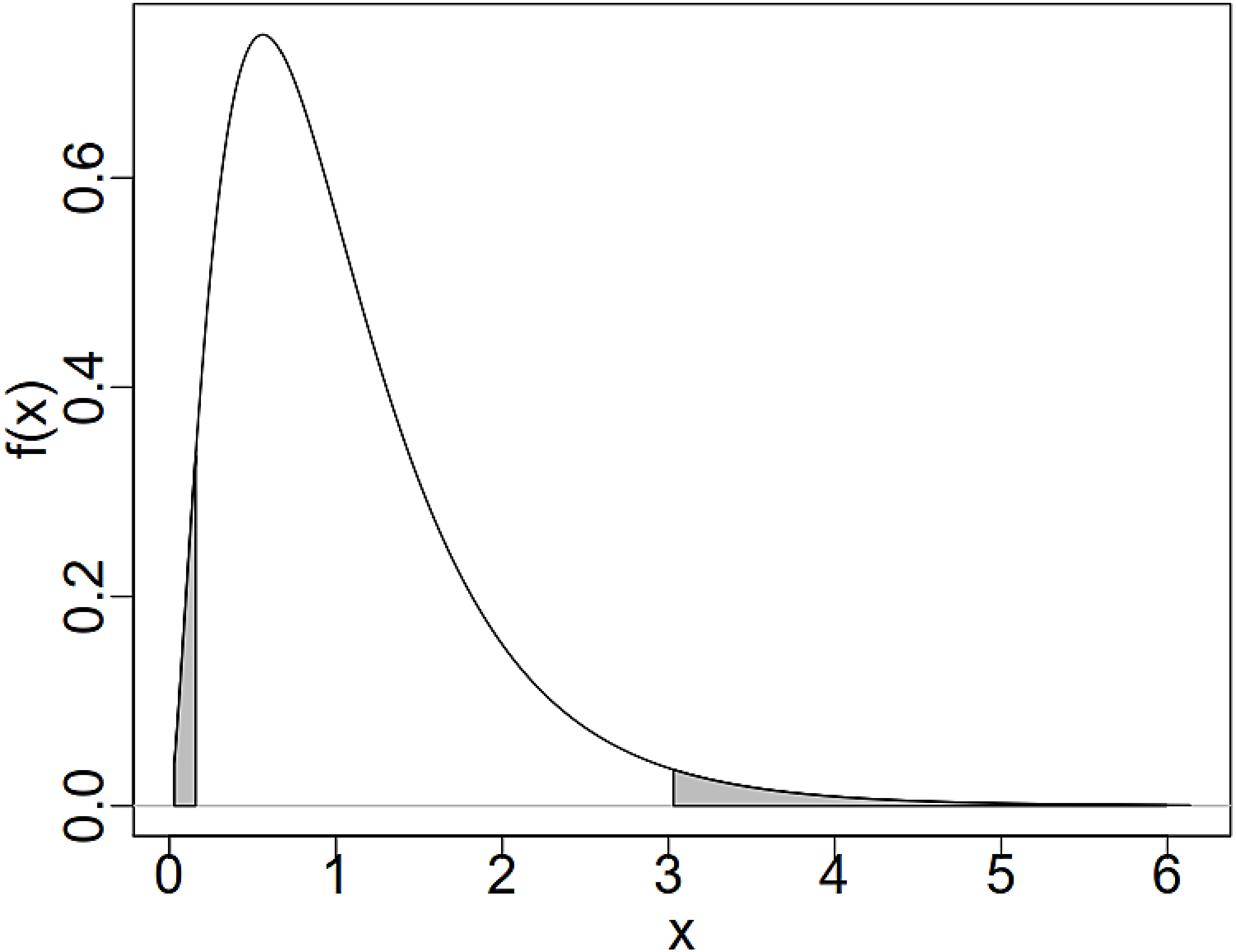
Osea que $u = n_1 - 1$ y $v = n_2 - 1$.

- Para una variable $F \sim f(u, v)$ la tabla reporta el cuantil superior:

$$P(F > f(u, v)) = \alpha$$

- Esta distribución no tiene simetría. Se cumple lo siguiente:

$$f_{1-\alpha}(u, v) = \frac{1}{f_\alpha(v, u)}$$



0.0
0.2
0.4
0.6
f(x)
0
1
2
3
4
5
6
x

IC para el cociente de varianzas

Sean $X_{11}, X_{12}, \ldots, X_{1n}$ y $X_{21}, X_{22}, \ldots, X_{2n}$ dos muestras que vienen de dos poblaciones independientes normales. El intervalo de confianza para el cociente de las varianzas σ_1^2 y σ_2^2

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{S_1^2}{S_2^2} f_{1-\alpha/2}(n_1 - 1, n_2 - 1), \frac{S_1^2}{S_2^2} f_{\alpha/2}(n_1 - 1, n_2 - 1) \right)$$

$$\Longleftrightarrow \frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{S_1^2}{S_2^2} \frac{1}{f_{\alpha/2}(n_2 - 1, n_1 - 1)}, \frac{S_1^2}{S_2^2} f_{\alpha/2}(n_1 - 1, n_2 - 1) \right)$$

Ejemplo

Suponga que usted esta frente a un proceso de llenado de cajas de cereal. Se tomó una muestra de las cajas con un contenido de 850 g y otra muestra de las cajas de 475 g. A continuación, se presenta una tabla que muestra la variabilidad en el llenado de los dos tamaños de cajas:

Cajas de 850 g	$n_1 = 4$	40.917
Cajas de 475 g	$n_2 = 15$	16.714

¿Existe diferencia en la variabilidad del llenado?

I.C para la diferencia de proporciones

I.C para la diferencia de proporciones

Sean $X_{11}, X_{12}, \ldots, X_{1n}$ e $X_{21}, X_{22}, \ldots, X_{2n}$ dos muestras que vienen de cierta población, con n_1 y n_2 **grandes**.

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

Ejemplo

Considere el proceso de manufactura de unos rodamientos. 85 de ellos se fabricaron de la forma tradicional, y otros 85 se fabricaron utilizando un acabado superficial distinto. Del primer grupo resultaron 10 defectuosos,

y del segundo grupo, resultaron 8 defectuosos. ¿Es la proporción de defectos del nuevo acabado superficial menor a la proporción con el método tradicional?