Estadística Aplicada

Unidad 4: Estimación y Prueba de Hipótesis

Ricardo Jesús Largaespada Fernández

Ingeniería de Sistemas, DACTIC, UNI

13 de Noviembre, 2025

Agenda

¿Qué es un intervalo de confianza para la media?

- Tenemos una población con media desconocida μ.
- Tomamos una muestra aleatoria $X_1, \ldots, X_n$.
- Calculamos la media muestral $\bar{X}$ (estimación puntual de μ).
- Un **intervalo de confianza** produce un rango de valores donde es razonable que esté μ, con nivel de confianza $(1-\alpha)100\,\%$.

Supuestos: media con varianza conocida

- X es una variable aleatoria con distribución normal $N(\mu, \sigma^2)$.
- La varianza poblacional σ^2 es **conocida**.
- La muestra $X_1, \ldots, X_n$ es aleatoria simple.
- Queremos un IC de $(1-\alpha)100\,\%$ para μ.

Ricardo Jesús Largaespada Fernández

Distribución de la media muestral

- La media muestral:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Si $X \sim N(\mu, \sigma^2)$, entonces

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

- El estadístico estandarizado:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

IC para μ con varianza conocida

A partir de

$$P\left(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}\right) = 1-\alpha,$$

sustituimos $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$:

$$P\left(\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

$$\boxed{\bar{x} \pm z_{1-\alpha/2}\,\frac{\sigma}{\sqrt{n}}}$$

Usamos la distribución normal estándar Z.

Ejemplo: consumo de lubricantes (datos)

Ejemplo. La cantidad de litros de lubricantes que se consumen en un mes en una cooperativa de transporte se modela como $X \sim N(\mu, 81)$ $(\sigma = 9)$.

Se toma una muestra de $n = 4$ meses y se obtienen (en litros):

$$1200,\ 1250,\ 1195,\ 1190.$$

- Estimar μ (estimación puntual).
- Construir un IC del 90 % para μ.

Ejemplo: consumo de lubricantes (solución IC)

- Media muestral:

$$\bar{x} = \frac{1200 + 1250 + 1195 + 1190}{4} = 1208.75 \text{ litros.}$$

- Varianza poblacional conocida: $\sigma = 9$.
- Nivel de confianza 90 % $\Rightarrow \alpha = 0.10$, $z_{1-\alpha/2} = z_{0.95} = 1.645$.

$$1208.75 - 1.645\frac{9}{2} \leq \mu \leq 1208.75 + 1.645\frac{9}{2}$$

$$1201.34 \leq \mu \leq 1216.15.$$

El consumo promedio mensual está entre 1201.34 y 1216.15 litros con confianza del 90 %.

Error máximo permisible

Del IC

$$\bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}},$$

el **ancho** del intervalo es

$$\text{L.S.} - \text{L.I.} = 2\varepsilon,$$

donde ε es el **error máximo permisible**.

Comparando:

$$\varepsilon = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

Tamaño de muestra deseado

De

$$\varepsilon = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}},$$

despejamos n:

$$n = \left(\frac{z_{1-\alpha/2}\, \sigma}{\varepsilon} \right)^2 .$$

Nos permite elegir n para un error máximo ε dado un nivel de confianza y una σ conocida.

Ejemplo: tamaño de muestra para lubricantes

En el ejemplo anterior se obtuvo un intervalo muy ancho. Se desea reducir la longitud del intervalo a 10 litros manteniendo el mismo nivel de confianza (90 %).

- Ancho deseado: $2\varepsilon = 10 \Rightarrow \varepsilon = 5$.
- $\sigma = 9$, $z_{1-\alpha/2} = 1.645$.

$$n = \left(\frac{1.645 \cdot 9}{5}\right)^2 \approx 8.77 \approx 9.$$

Se necesitan al menos 9 meses de muestra para un ancho de 10 litros.

Supuestos: media con varianza desconocida

- Población normal $N(\mu, \sigma^2)$.
- σ^2 es **desconocida**.
- Usamos la desviación estándar muestral s.
- Muestra pequeña: $n < 30$.
- Queremos un IC para μ con nivel $(1-\alpha)100\,\%$.

Distribución t de Student

El estadístico

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

tiene distribución **t de Student** con $n-1$ grados de libertad si la población es normal.

Por tanto:

$$P\left(-t_{(1-\alpha/2),\,n-1} \le \frac{\bar{X}-\mu}{s/\sqrt{n}} \le t_{(1-\alpha/2),\,n-1}\right) = 1-\alpha.$$

IC para μ con varianza desconocida

Despejando μ obtenemos:

$$P\left(\bar{X} - t_{(1-\alpha/2),\, n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{(1-\alpha/2),\, n-1}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

$$\boxed{\bar{x} \pm t_{(1-\alpha/2),\, n-1}\, \frac{s}{\sqrt{n}}}$$

Aquí usamos la distribución t de Student.

Ejemplo: promedio de graduados UNI (datos)

Una muestra aleatoria de 18 alumnos graduados en la UNI dio un promedio de 73.3 en todas las asignaturas de su plan de estudio, con desviación estándar muestral $s = 7.8$.

Suponiendo que las notas se distribuyen normalmente, obtener un IC del 98 % para el promedio de graduados.

- $n = 18$.
- $\bar{x} = 73.3$.
- $s = 7.8$.
- $\alpha = 0.02 \Rightarrow t_{(1-\alpha/2),\,17} = t_{0.99,17} \approx 2.567$.

Ejemplo: promedio de graduados UNI (solución)

$$\bar{x} \pm t_{0.99,17} \frac{s}{\sqrt{n}} = 73.3 \pm 2.567 \frac{7.8}{\sqrt{18}}.$$

$$73.3 - 2.567 \frac{7.8}{\sqrt{18}} \leq \mu \leq 73.3 + 2.567 \frac{7.8}{\sqrt{18}}.$$

$$68.581 \leq \mu \leq 78.019.$$

El promedio de notas se encuentra entre 68.581 y 78.019 con confianza del 98 %.

Dos medias, varianzas conocidas

- Dos poblaciones normales independientes:

$$X_1 \sim N(\mu_1, \sigma_1^2), \quad X_2 \sim N(\mu_2, \sigma_2^2).$$

- Varianzas poblacionales σ_1^2 y σ_2^2 conocidas.
- Muestras independientes:

$$\bar{X}_1,\ \bar{X}_2.$$

- Queremos un IC para $\mu_1 - \mu_2$.

Distribución de la diferencia de medias

- La diferencia de medias muestrales:

$$\bar{X}_1 - \bar{X}_2$$

  tiene distribución normal con media

$$\mu = \mu_1 - \mu_2$$

  y varianza

$$\sigma^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

- El estadístico:

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

Ricardo Jesús Largaespada Fernández

IC para $\mu_1 - \mu_2$ (varianzas conocidas)

El IC del $(1-\alpha)100\,\%$ para $\mu_1 - \mu_2$ es:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

$$\boxed{(\bar{x}_1 - \bar{x}_2) - z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + z_{1-}}$$

Ejemplo: Sistemas vs Computación (datos)

Se tomó una muestra de:

- $n_1 = 50$ estudiantes de Ingeniería en Sistemas, con promedio $\bar{x}_1 = 65$ y desviación estándar $\sigma_1 = 6$.
- $n_2 = 48$ estudiantes de Computación, con promedio $\bar{x}_2 = 62$ y desviación estándar $\sigma_2 = 7$.

Construir un IC del 95 % para $\mu_1 - \mu_2$ (diferencia real de promedios).

$$\alpha = 0.05, \quad z_{1-\alpha/2} = z_{0.975} = 1.96.$$

Ejemplo: Sistemas vs Computación (solución)

$$(\bar{x}_1 - \bar{x}_2) \pm 1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = (65 - 62) \pm 1.96\sqrt{\frac{6^2}{50} + \frac{7^2}{48}}.$$

$$(65-62)-1.96\sqrt{\frac{36}{50} + \frac{49}{48}} \leq \mu_1-\mu_2 \leq (65-62)+1.96\sqrt{\frac{36}{50} +$$

$$0.414 \leq \mu_1 - \mu_2 \leq 5.586.$$

La diferencia real no puede ser 0, por lo que hay evidencia de diferencia significativa al 95 %.

Dos medias, varianzas desconocidas e iguales

- Dos poblaciones normales independientes con varianzas desconocidas pero iguales: $\sigma_1^2 = \sigma_2^2$.
- Muestras independientes: n_1, $\bar{X}_1$, s_1^2 y n_2, $\bar{X}_2$, s_2^2.
- Queremos un IC para $\mu_1 - \mu_2$.
- Usamos una varianza combinada o **varianza agrupada** s_p^2.

Varianza agrupada y estadístico t

La varianza agrupada se define como:

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

El estadístico:

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

tiene distribución t de Student con $n_1 + n_2 - 2$ grados de libertad.

IC para $\mu_1 - \mu_2$ (varianzas desconocidas e iguales)

El IC del $(1-\alpha)100\,\%$ para $\mu_1 - \mu_2$ es:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{(1-\alpha/2),\,(n_1+n_2-2)}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

$$\boxed{(\bar{x}_1 - \bar{x}_2) - t_{(1-\alpha/2),\,n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 -}$$

Ejemplo: máquinas troqueladoras (datos)

Una empresa tiene dos máquinas troqueladoras de láminas de zinc, ubicadas en Masaya y León. Se desea comparar el **tiempo promedio diario** que estas máquinas **no trabajan**.

Se seleccionan 15 días aleatoriamente para cada máquina:

$$n_1 = 15, \quad \bar{x}_1 = 45, \quad s_1 = 2.5,$$
$$n_2 = 15, \quad \bar{x}_2 = 36, \quad s_2 = 3.$$

Se asume que las varianzas poblacionales son iguales pero desconocidas. Construir un IC del 99 % para $\mu_1 - \mu_2$.

Ejemplo: máquinas (solución)

$$s_p^2 = \frac{(15-1)2.5^2 + (15-1)3^2}{15+15-2} = 7.625, \quad s_p \approx 2.761.$$

$$\alpha = 0.01, \quad t_{(1-\alpha/2),\, n_1+n_2-2} = t_{0.995,\, 28} \approx 2.76326.$$

IC:

$$(\bar{x}_1 - \bar{x}_2) \pm 2.76326 \cdot 2.761 \sqrt{\frac{1}{15} + \frac{1}{15}}.$$

$$45-36-2.76326(2.761)\sqrt{\tfrac{1}{15}+\tfrac{1}{15}} \le \mu_1-\mu_2 \le 45-36+2.763$$

$$6.2138 \le \mu_1 - \mu_2 \le 11.7862.$$

La máquina 1 deja de trabajar entre 6.2138 y 11.7862 minutos más que la máquina 2 (99 % de confianza).

Ricardo Jesús Largaespada Fernández

Proporción de éxitos en una población

- Población binomial con parámetro P (proporción de éxitos).
- Muestra de tamaño n con x éxitos.
- Proporción muestral:

$$\hat{p} = \frac{x}{n}.$$

- Para n grande, la distribución binomial se aproxima a la normal.

IC para una proporción

El IC del $(1-\alpha)100\,\%$ para P es:

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

$$\boxed{\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq P \leq \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Usamos aproximación normal Z.

Ejemplo: bloques defectuosos (datos)

Un fabricante asegura que el porcentaje de bloques defectuosos no es mayor del 5 %.
Una ferretería selecciona $n = 200$ bloques de su inventario y encuentra $x = 19$ defectuosos.

- ¿Debe sospechar de la afirmación del fabricante?
- Construir un IC del 95 % para P.

$$\hat{p} = \frac{19}{200} = 0.095, \quad z_{1-\alpha/2} = z_{0.975} = 1.96.$$

Ejemplo: bloques defectuosos (solución)

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{19}{200} \pm 1.96\sqrt{\frac{(19/200)(181/200)}{200}}.$$

$$0.05436 \leq P \leq 0.1356.$$

$$5.43\,\% \leq P \leq 13.56\,\%.$$

El 5 % (valor asegurado) queda a la izquierda del intervalo, por lo que la ferretería tiene motivos para sospechar de la afirmación.

Diferencia de dos proporciones

- Dos poblaciones binomiales, parámetros P_1 y P_2.
- Muestras independientes:

$$\hat{p}_1 = \frac{x_1}{n_1}, \quad \hat{p}_2 = \frac{x_2}{n_2}.$$

- Para n_1 y n_2 grandes, $\hat{p}_1 - \hat{p}_2$ se aproxima a normal.

IC para $P_1 - P_2$

La varianza aproximada de $\hat{p}_1 - \hat{p}_2$ es

$$\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}.$$

El IC del $(1-\alpha)100\,\%$ para $P_1 - P_2$ es:

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

Ejemplo: duración de llantas (datos)

En un estudio sobre la duración de llantas de dos marcas:

- Marca A: $n_1 = 200$, $x_1 = 34$ llantas que **no** duran los 48 000 km.
- Marca B: $n_2 = 200$, $x_2 = 29$ llantas que **no** duran los 48 000 km.

Obtener un IC del 90 % para $P_1 - P_2$ (diferencia real de proporciones que no duran la garantía).

$$\hat{p}_1 = \frac{34}{200}, \quad \hat{p}_2 = \frac{29}{200}, \quad z_{1-\alpha/2} = z_{0.95} = 1.645.$$

Ejemplo: duración de llantas (solución)

$$(\hat{p}_1 - \hat{p}_2) \pm 1.645\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

$$\left(\frac{34}{200} - \frac{29}{200}\right) \pm 1.645\sqrt{\frac{(34/200)(166/200)}{200} + \frac{(29/200)(171/2}{200}}$$

$$-0.035 \leq P_1 - P_2 \leq 0.085.$$

Como el 0 está dentro del intervalo, no hay diferencia significativa en la duración de las dos marcas (90 % de confianza).

IC para la varianza de una población normal

- Población normal con varianza σ^2.
- Muestra de tamaño n, varianza muestral S^2.
- El estadístico

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

tiene distribución χ^2 con $n-1$ grados de libertad.

IC para σ^2

A partir de los cuantiles $\chi^2_{(1-\alpha/2),\, n-1}$ y $\chi^2_{(\alpha/2),\, n-1}$:

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

Despejando σ^2:

$$\boxed{\frac{(n-1)S^2}{\chi^2_{(\alpha/2),\, n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{(1-\alpha/2),\, n-1}}}$$

Para la desviación estándar basta tomar raíces cuadradas.

Ejemplo: peso de paquetes de semilla (datos)

Los pesos (kg) de 10 paquetes de semilla son:

46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2, 46.0.

- $n = 10$, varianza muestral $S^2 = 0.2862$.
- Nivel de confianza 95 %: $\alpha = 0.05$.
- Cuantiles:

$$\chi^2_{0.975,\,9} = 19.02, \quad \chi^2_{0.025,\,9} = 2.70.$$

Calcular el IC del 95 % para σ.

Ejemplo: peso de paquetes de semilla (solución)

Primero el IC para σ^2:

$$\frac{(10-1)0.2862}{19.02} \leq \sigma^2 \leq \frac{(10-1)0.2862}{2.70}.$$

$$0.1355 \leq \sigma^2 \leq 0.9540.$$

Tomando raíces cuadradas:

$$0.3681 \leq \sigma \leq 0.9767 \text{ kg}.$$

La desviación estándar del peso de los paquetes de semilla está entre 0.3681 y 0.9767 kg con confianza del 95 %.