

Pruebas de hipótesis dos poblaciones

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

Departamento de ingeniería de sistemas



**UNIVERSIDAD
DE ANTIOQUIA**

Prueba de hipótesis para la diferencia de medias

Prueba de hipótesis para la diferencia de medias

$$H_0 : \mu_1 - \mu_2 = \delta_0$$

$$H_0 : \mu_1 - \mu_2 \leq \delta_0$$

$$H_0 : \mu_1 - \mu_2 \geq \delta_0$$

$$H_1 : \mu_1 - \mu_2 \neq \delta_0$$

$$H_1 : \mu_1 - \mu_2 > \delta_0$$

$$H_1 : \mu_1 - \mu_2 < \delta_0$$

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 I

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

$$Z_c = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

- Prueba de dos colas: Rechazo H_0 cuando $\{Z_c : |Z_c| > Z_{\alpha/2}\}$
- Prueba de cola superior: Rechazo H_0 cuando $\{Z_c : Z_c > Z_{\alpha}\}$
- Prueba de cola inferior: Rechazo H_0 cuando $\{Z_c : Z_c < -Z_{\alpha}\}$
- Rechazo H_0 si valor $p < \alpha$

Caso II

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

$$Z_c = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0, 1)$$

- Prueba de dos colas: Rechazo H_0 cuando $\{Z_c : |Z_c| > Z_{\alpha/2}\}$
- Prueba de cola superior: Rechazo H_0 cuando $\{Z_c : Z_c > Z_{\alpha}\}$
- Prueba de cola inferior: Rechazo H_0 cuando $\{Z_c : Z_c < -Z_{\alpha}\}$
- Rechazo H_0 si valor $p < \alpha$.

Valor P

$$V_p = \begin{cases} 2[1 - \Phi(|Z_c|)] & \text{Prueba de dos colas: } H_1 : \mu_1 - \mu_2 \neq \delta_0 \\ 1 - \Phi(Z_c) & \text{Prueba de cola superior: } H_1 : \mu_1 - \mu_2 > \delta_0 \\ \Phi(Z_c) & \text{Prueba de cola inferior: } H_1 : \mu_1 - \mu_2 < \delta_0 \end{cases}$$

Rechazo H_0 si valor $p < \alpha$.

Ejemplo

Se informaron los datos resumidos en la tabla sobre la ingesta diaria de calorías tanto para una muestra de adolescentes que dijeron que típicamente no comían comida rápida y otra muestra de adolescentes que dijeron que usualmente comían comida rápida.

| Comen comida rápida | Tamaño de la muestra | Media | Desviación estándar |
|---------------------|----------------------|-------|---------------------|
| No | 663 | 2258 | 1519 |
| Sí | 413 | 2637 | 1138 |

¿Estos datos proporcionan evidencia sólida para concluir que la verdadera ingesta calórica promedio para los adolescentes que normalmente comen comida rápida supera en más de 200 calorías por día la verdadera ingesta promedio para aquellos que normalmente no comen comida rápida? Investiguemos realizando una prueba de hipótesis a un nivel de significancia de 0.05.

Caso III

Sean $X_{11}, X_{12}, \dots, X_{1n}$ y $X_{21}, X_{22}, \dots, X_{2n}$ dos muestras que vienen de dos poblaciones independientes normales, con n_1 y n_2 no necesariamente grandes, con medias μ_1 y μ_2 . Si las varianzas σ_1^2 y σ_2^2 son **desconocidas**, es necesario considerar dos casos:

- Caso a: $\sigma_1^2 = \sigma_2^2$
- Caso b: $\sigma_1^2 \neq \sigma_2^2$

Caso III varianzas iguales

Varianzas desconocidas, $\sigma_1^2 = \sigma_2^2$:

$$T_c = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

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

Caso III varianzas diferentes

Varianzas desconocidas, $\sigma_1^2 \neq \sigma_2^2$:

$$T_c = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t(\nu)$$

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

Caso III

Si $\sigma_1^2 = \sigma_2^2$, entonces $r = n_1 + n_2 - 2$. Si $\sigma_1^2 \neq \sigma_2^2$, entonces $r = \nu$. El criterio de rechazo es:

- Prueba de dos colas: Rechazo H_0 cuando $\{T_c : |T_c| > T_{\alpha/2}(r)\}$
- Prueba de cola superior: Rechazo H_0 cuando $\{T_c : T_c > T_{\alpha}(r)\}$
- Prueba de cola inferior: Rechazo H_0 cuando $\{T_c : T_c < -T_{\alpha}(r)\}$
- Rechazo H_0 si valor $p < \alpha$.

NOTA: para aplicar el CASO III de diferencia de medias en intervalos de confianza o pruebas de hipótesis, *SIEMPRE* es necesario determinar si las varianzas son iguales o diferentes.

Valor P

$$V_p = \begin{cases} 2P(T > |T_c|) & \text{Prueba de dos colas: } H_1 : \mu_1 - \mu_2 \neq \delta_0 \\ P(T > T_c) & \text{Prueba de cola superior: } H_1 : \mu_1 - \mu_2 > \delta_0 \\ P(T < T_c) & \text{Prueba de cola inferior: } H_1 : \mu_1 - \mu_2 < \delta_0 \end{cases}$$

Rechazo H_0 si valor $p < \alpha$.

Prueba de hipótesis para el cociente de varianzas

Prueba de hipótesis para el cociente de varianzas

Sean $X_{11}, X_{12}, \dots, X_{1n}$ y $X_{21}, X_{22}, \dots, X_{2n}$ dos muestras que vienen de unas poblaciones independientes normales, con n_1 y n_2 no necesariamente grandes.

Supongamos que queremos probar las siguientes hipótesis:

$$H_0 : \sigma_1 = \sigma_2$$

$$H_1 : \sigma_1 \neq \sigma_2$$

$$H_0 : \sigma_1 \leq \sigma_2$$

$$H_1 : \sigma_1 > \sigma_2$$

$$H_0 : \sigma_1 \geq \sigma_2$$

$$H_1 : \sigma_1 < \sigma_2$$

Estadístico de prueba:

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

Recuerde que:

$$f_{1-\alpha/2}(n_2 - 1, n_1 - 1) = \frac{1}{f_{\alpha/2}(n_1 - 1, n_2 - 1)}$$

Criterio de rechazo

Estadístico de prueba:

$$F_c = \frac{S_1^2}{S_2^2}$$

Criterio de rechazo:

- Prueba de dos colas: Rechazo H_0 cuando

$$\begin{aligned} & \{F_c : F_c > f_{\alpha/2}(n_1 - 1, n_2 - 1) \text{ ó } F_c < f_{1-\alpha/2}(n_1 - 1, n_2 - 1)\} \\ \Leftrightarrow & \left\{ F_c : F_c > f_{\alpha/2}(n_1 - 1, n_2 - 1) \text{ ó } F_c < \frac{1}{f_{\alpha/2}(n_2 - 1, n_1 - 1)} \right\} \end{aligned}$$

- Rechazo H_0 si valor $p < \alpha$

Estadístico de prueba:

$$F_c = \frac{S_1^2}{S_2^2}$$

- Prueba de cola superior: Rechazo H_0 cuando

$$\{F_c : F_c > f_{\alpha}(n_1 - 1, n_2 - 1)\}$$

- Prueba de cola inferior: Rechazo H_0 cuando

$$\{F_c < f_{1-\alpha}(n_1 - 1, n_2 - 1)\}$$

Nota: recuerde que, por ejemplo, $f_{1-\alpha}(n_1 - 1, n_2 - 1)$ es un área hacia la derecha.

Ejemplo

Dos proveedores fabrican un engrane plástico utilizado en una impresora láser. El interés está en la resistencia al impacto del engrane, la cual se mide en pie-lb. Una muestra aleatoria de 10 engranes suministrados por el proveedor *A* arroja una resistencia promedio de 290 pie-lb, con una desviación estándar de 12 pie-lb.

Del proveedor *B*, se toma una muestra de 15 engranes. Esta muestra arroja una resistencia promedio de 308.5 pie-lb y una desviación estándar de 15 pie-lb. Se sabe que las resistencias al impacto se comportan de manera normal para este tipo de engranes.

¿Puede afirmarse que los engranes del proveedor *A* tienen una resistencia media inferior a los engranes del proveedor *B*? Use una significancia del 5 %.

Prueba de hipótesis para la diferencia de proporciones

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 \neq p_2$$

$$H_0 : p_1 \leq p_2$$

$$H_1 : p_1 > p_2$$

$$H_0 : p_1 \geq p_2$$

$$H_1 : p_1 < p_2$$

Prueba de hipótesis para la diferencia de proporciones

Sean $X_1 \sim \text{Binom}(n_1, p_1)$ y $X_2 \sim \text{Binom}(n_2, p_2)$. Si n_1 y n_2 son grandes, entonces el estadístico de prueba es:

$$Z_c = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad \hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

- Prueba de dos colas: Rechazo H_0 cuando $\{Z_c : |Z_c| > Z_{\alpha/2}\}$
- Prueba de cola superior: Rechazo H_0 cuando $\{Z_c : Z_c > Z_{\alpha}\}$
- Prueba de cola inferior: Rechazo H_0 cuando $\{Z_c : Z_c < -Z_{\alpha}\}$
- Rechazo H_0 si valor $p < \alpha$.

Valor P

$$V_p = \begin{cases} 2[1 - \Phi(|Z_c|)] & \text{Prueba de dos colas: } H_1 : p_1 \neq p_2 \\ 1 - \Phi(Z_c) & \text{Prueba de cola superior: } H_1 : p_1 > p_2 \\ \Phi(Z_c) & \text{Prueba de cola inferior: } H_1 : p_1 < p_2 \end{cases}$$

Rechazo H_0 si valor $p < \alpha$.

Ejemplo

Un coordinador universitario sospecha que la deserción estudiantil es más común hoy en día. Un estudio del año 2005 reportó que, de 400 estudiantes seleccionados aleatoriamente, 130 habían dejado sus universidades. Otro estudio de la misma naturaleza se realizó en 2015, y mostró que 225 estudiantes desertaron, de una muestra aleatoria de 600. ¿Es cierta la conjetura del coordinador?