Unidad 8
Test de Hipótesis

Parte 4 : Test para dos poblaciones

En esta Unidad estudiaremos:

01
Definiciones
Hipótesis estadística.
Hipótesis Nula y Alternativa.
Test de hipótesis.

02
Metodología
Paso a paso para realizar un test.

03
Test para una población
Test para la media.
Test para la proporción..

04
Test para dos poblaciones
Diferencia de medias.
Medias pareadas.
Diferencia de proporciones

05
Test e IC
Relación entre los dos métodos de estimación.

06
Errores
Que se pueden cometer al hacer un test.

04

Test para dos poblaciones

Diferencia de medias.

Medias pareadas.

Diferencia de proporciones

Prueba de Hipótesis para 2 poblaciones

Frecuentemente se presentan situaciones en la que es necesario comparar 2 poblaciones o 2 subpoblaciones de una población.

En general se desea comparar las medias, las varianzas de dos poblaciones o bien las proporciones de cierto atributo.

Comparación de medias de dos poblaciones:

Pueden presentarse diferentes casos:

- Varianzas conocidas
- Varianzas desconocidas: Varianzas iguales
  Varianzas diferentes

En este curso vamos sólo a estudiar el caso de comparación de **2 medias de poblaciones con varianza desconocidas e iguales.**

Metodología para hacer un test

1. Definir H_0 y H_1.

2. Definir el estadístico a utilizar y su distribución de probabilidad bajo H0. la forma general de este estadístico será:

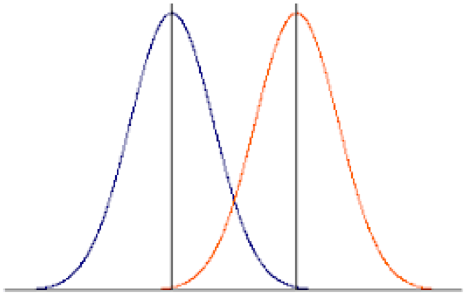
$$\text{Estadístico} = \frac{Estimador - Parámetro\ bajo\ H_0}{Desviación\ estándar\ del\ estimador}$$

3. Dado el nivel de significación α, determinar la región crítica y con los datos tomar la decisión.

Test de medias en poblaciones Normales independientes

Para realizar esta prueba se necesitan las siguientes hipótesis:

$X_1, X_2, \ldots X_n$ iid como X, $X \sim N(\mu_{X}, \sigma)$
$Y_1, Y_2, \ldots Y_m$ iid como Y, $Y \sim N(\mu_{Y}, \sigma)$ Independientes entre sí y σ común.



Test de medias en poblaciones Normales independientes

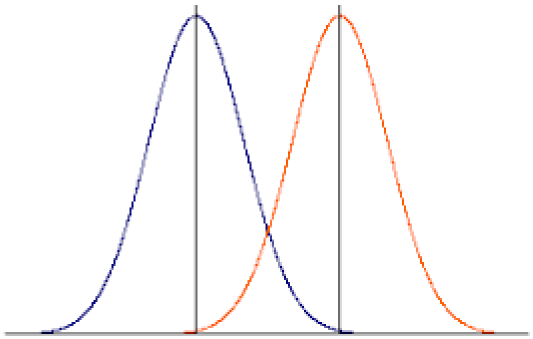
1. Definir H_0 y H_1.

H_0: $\mu_X = \mu_Y$

H_1: $\mu_X \neq \mu_Y$

H_1: $\mu_X > \mu_Y$

H_1: $\mu_X < \mu_Y$



Test de medias en poblaciones independientes

2. Estadístico pivote y su distribución de probabilidades bajo H_0.

$$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{(n+m-2)}$$

Donde $s_p^2 = \dfrac{(n-1)s_X^2+(m-1)s_Y^2}{n+m-2}$

es el cuadrado del estimador de σ

Test de medias en poblaciones independientes

3. Dado el nivel de significación α, determinar la región crítica y con los datos tomar la decisión.

- **Alternativa de dos colas: H_1: $\mu_X \neq \mu_Y$**

Dado α, se tiene $t_{\text{crítico}} = t_{1-\alpha/2}$, la región de rechazo es: $|t_{\text{observado}}| > t_{\text{crítico}}$

- **Alternativa cola derecha: H_1: $\mu_X > \mu_Y$**

Dado α, se tiene $t_{\text{crítico}} = t_{1-\alpha}$, la región de rechazo es: $t_{\text{observado}} > t_{\text{crítico}}$

- **Alternativa cola izquierda: H_1: $\mu_X < \mu_Y$**

Dado α, se tiene $t_{\text{crítico}} = t_{\alpha}$, la región de rechazo es: $t_{\text{observado}} < t_{\text{crítico}}$

Test de medias en poblaciones independientes

3. O bien determinar valor p con los datos y tomar la decisión. El valor p queda determinado por:

- Valor p = $P(|t| > |t_{observado}|)$ si H_1: $\mu_X \neq \mu_Y$
- Valor p = $P(t > t_{observado})$ si H_1: $\mu_X > \mu_Y$
- Valor p = $P(t < t_{observado})$ si H_1: $\mu_X < \mu_Y$

Ejemplo 1

Se lleva a cabo un experimento para comparar el desgaste por abrasivo de dos diferentes materiales laminados.

Se prueban 12 piezas del material X obteniéndose $\bar{X} = 85$ y $s_X = 4$ y 10 piezas del material Y resultando $\bar{Y} = 81$ y $s_Y = 5$.

a) ¿Se puede concluir que tienen el mismo desgaste abrasivo ambos materiales? Considere $\alpha = 0.05$.

b) Si se sabe que el material X no puede tener menor desgaste que el material Y. ¿Se puede concluir que el desgaste del material X es mayor que el del material Y? Considere $\alpha = 0.05$.

Ejemplo 1

Información:

$X_1, X_2, \ldots X_{12}$ iid como X, $X \sim N(\mu_X, \sigma)$

$Y_1, Y_2, \ldots Y_{10}$ iid como Y, $Y \sim N(\mu_Y, \sigma)$ Independientes entre sí y σ común.

$n = 12 \quad \bar{X} = 85 \quad s_X = 4$

$m = 10 \quad \bar{Y} = 81 \quad s_Y = 5$

¿Se puede concluir que tienen el mismo desgaste abrasivo ambos materiales? Considere $\alpha = 0.05$.

Ejemplo 1

1. Definir H_0 y H_1.

H_0: $\mu_x = \mu_y$
H_1: $\mu_x \neq \mu_y$

O equivalentemente:

H_0: $\mu_x - \mu_y = 0$
H_1: $\mu_x - \mu_y \neq 0$

Ejemplo 1

2. Estadístico pivote y su distribución de probabilidades bajo H_0.

$$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{(n+m-2)}$$

Ejemplo 1

3. Dado el nivel de significación α, determinar la región crítica y con los datos tomar la decisión.

$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}$$

$$= \frac{11x4^2+9x5^2}{12+10-2} = 20.05 \quad \Rightarrow \quad s_p = 4.48$$

Entonces:

$$t_{observado} = \frac{(85-81)-0}{4.48\sqrt{\frac{1}{12}+\frac{1}{10}}} = 2.085 \quad \text{y} \quad t_{crítico} = t_{0.975(20)} = 2.09$$

Como: $|t_{observado}| < t_{crítico}$ → No hay evidencia para rechazar H_0.

Concluimos que tienen el mismo desgaste ambos materiales.

Ejemplo 1

Información:

$X_1, X_2, \ldots X_{12}$ iid como X, $X \sim N(\mu_X, \sigma)$
$Y_1, Y_2, \ldots Y_{10}$ iid como Y, $Y \sim N(\mu_Y, \sigma)$ Independientes entre sí y σ común.

$n = 12 \quad \bar{X} = 85 \quad s_X = 4$
$m = 10 \quad \bar{Y} = 81 \quad s_Y = 5$

Si se sabe que el material X no puede tener menor desgaste que el material Y, ¿se puede concluir que el desgaste del material X es mayor que el del material Y? Considere $\alpha = 0.05$.

Ejemplo 1

1. Definir H_0 y H_1.

   $\mathbf{H_0}$: $\mu_X = \mu_Y$

   $\mathbf{H_1}$: $\mu_X > \mu_Y$

O equivalentemente:

$\mathbf{H_0}$: $\mu_X - \mu_Y = 0$

$\mathbf{H_1}$: $\mu_X - \mu_Y > 0$

2. Estadístico pivote y su distribución de probabilidades bajo H_0.

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{(n+m-2)}$$

3. Dado el nivel de significación α, determinar la región crítica y con los datos tomar la decisión.

Región crítica: $t_{observado} > t_{1-\alpha\,(n+m-2)}$

$t_{observado} = 2.085 > t_{0.95(20)} = 1.72$

Hay evidencia para rechazar H_0.

Test de medias en poblaciones Normales dependientes

Esta prueba se realiza si:

$X_1, X_2, \ldots, X_n$ iid como X, $X \sim N(\mu_X, \sigma)$
$Y_1, Y_2, \ldots, Y_n$ iid como Y, $Y \sim N(\mu_Y, \sigma)$ dependientes entre sí y σ común.

Se usa cuando se tiene una misma población observada en 2 momentos o situaciones diferentes. En este caso se utiliza comparación de Medias Apareadas o Pareadas.

Observación:
Suele usarse la notación:
$X_1^1, X_2^1, \ldots, X_n^1$ m.a. de $X_1 \sim N(\mu_1, \sigma)$ y $X_1^2, X_2^2, \ldots, X_n^2$ m.a. de $X_2 \sim N(\mu_2, \sigma)$

Test de medias pareadas

La muestra aleatoria ahora es una muestra aleatoria de diferencias:

$$D_1, D_2, \ldots, D_n \text{ iid como } D \sim N(\mu_D, \sigma_D), \text{ con } D_i = X_{1i} - X_{2i}$$

O bien n suficientemente grande para aplicar el Teorema Central del Límite, en cuyo caso el estadístico tendrá distribución aproximada.

1. La hipótesis nula es H_0: $\mu_D = \mu_o$.
   Las posibles hipótesis alternativas son: H_1: $\mu_D \neq \mu_o$ H_1: $\mu_D > \mu_o$ H_1: $\mu_D < \mu_o$

2. El estadístico a utilizar es:

$$\frac{\bar{d} - \mu_D}{s_D / \sqrt{n}} \sim t_{n-1}$$

3. El valor p y la región de rechazo se determinan en forma análoga al caso de comparación de una media.

Ejemplo 2

Se desea comparar la efectividad de cierta dieta para adelgazar en jóvenes. Para ello se considera una muestra de 10 jóvenes y se los pesa antes y después de un mes de dieta. Los pesos en kg. son (suponga distribución Normal para los datos y α= 0.05 y que la dieta no permite subir de peso):

Antes:	**85**	**83**	**82**	**81**	**79**	**90**	**88**	**85**	**89**	**91**
Después:	**80**	**78**	**78**	**81**	**75**	**87**	**83**	**80**	**85**	**92**

$$D_1, D_2, \ldots, D_n \text{ iid como } D \sim N(\mu_D, \sigma_D)$$

Diferencias:	**5**	**5**	**4**	**0**	**5**	**3**	**5**	**5**	**4**	**-1**

$\bar{d}$ = 3.5 y s_D = 2.22

Ejemplo 2

1. Definir H_0 y H_1.

$$\mathbf{H_0: \mu_D = 0}$$

$$\mathbf{H_1: \mu_D > 0}$$

O equivalentemente:

$$H_0: \mu_{antes} = \mu_{después}$$

$$H_1: \mu_{antes} > \mu_{después}$$

2. Estadístico pivote y su distribución de probabilidades bajo H_0.

$$\frac{\bar{D}-0}{s_D/\sqrt{n}} \sim t_{(n-1)}$$

3. Dado el nivel de significación α, determinar la región crítica y con los datos tomar la decisión.

$$t_{observado} = \frac{3.5-0}{2.22/\sqrt{10}} = 4.99 \qquad t_{critico} = t_{0.95(9)} = 1.833$$

$$\mathbf{t_{observado} > t_{0.95(9)}} \Rightarrow \text{Se rechaza } H_0$$

Test de diferencia de proporciones en poblaciones Bernoulli

Esta prueba se realiza si:

$X_1, X_2, \ldots X_n$ iid como X, X $\sim B(p_X)$
$Y_1, Y_2, \ldots Y_m$ iid como Y, Y $\sim B(p_Y)$ independientes entre si

Test de diferencia de proporciones en poblaciones Bernoulli

1. Definir H_0 y H_1.

$\boxed{\begin{array}{l} H_0: p_x = p_y \\ H_1: p_x \neq p_y \end{array}}$	O bien	$\boxed{\begin{array}{l} H_0: p_x - p_y = 0 \\ H_1: p_x - p_y \neq 0 \end{array}}$
$H_1: p_x > p_y$		$H_1: p_x - p_y > 0$
$H_1: p_x < p_y$		$H_1: p_x - p_y < 0$

Test de diferencia de proporciones en poblaciones Bernoulli

2. Estadístico pivote y su distribución de probabilidades bajo H_0.

$$\frac{\bar{X}-\bar{Y}-(p_X-p_Y)}{\sqrt{pq\left(\frac{1}{n}+\frac{1}{m}\right)}} \approx N(0,1)$$

por TCL

Donde $p = \frac{\sum X_i + \sum Y_i}{n+m} = \frac{n\bar{X}+m\bar{Y}}{n+m}$ y $q = 1 - p$

Test de diferencia de proporciones en poblaciones Bernoulli

3. Dado el nivel de significación α, determinar la región crítica y con los datos tomar la decisión.

- **Alternativa de dos colas: H_1: $p_X \neq p_Y$**

Dado α, se tiene $z_{\text{crítico}} = z_{1-\alpha/2}$, la región de rechazo es: $|z_{\text{observado}}| > z_{\text{crítico}}$

- **Alternativa cola derecha: H_1: $p_X > p_Y$**

Dado α, se tiene $z_{\text{crítico}} = z_{1-\alpha}$, la región de rechazo es: $z_{\text{observado}} > z_{\text{crítico}}$

- **Alternativa cola izquierda: H_1: $p_X < p_Y$**

Dado α, se tiene $z_{\text{crítico}} = z_{\alpha}$, la región de rechazo es: $z_{\text{observado}} < z_{\text{crítico}}$

O bien calcular valor p.

Ejemplo 3

Se desea comparar la proporción de votantes a favor de la instalación de una planta química de dos ciudades que se verían afectadas por su construcción. Se recogen los siguientes datos:

Ciudad A: Total de encuestados 200, a favor: 120

Ciudad B: Total de encuestados 500, a favor: 240

¿Se podría concluir que la proporción de votantes a favor es mayor en la ciudad A que en la B?

$$\widehat{p_A} = \frac{120}{200} = \bar{X} \qquad \widehat{p_B} = \frac{240}{500} = \bar{Y}$$

Usaremos valor p para tomar la decisión.

Ejemplo 3

1. Definir H_0 y H_1

   H_0: $p_A = p_B$

   H_1: $p_A \neq p_B$

2. Estadístico pivote y su distribución de probabilidades bajo H_0.

$$\frac{\widehat{p_A}-\widehat{p_B}-(p_A-p_B)}{\sqrt{pq\left(\frac{1}{n}+\frac{1}{m}\right)}} \approx N(0,1) \qquad \text{por TCL}$$

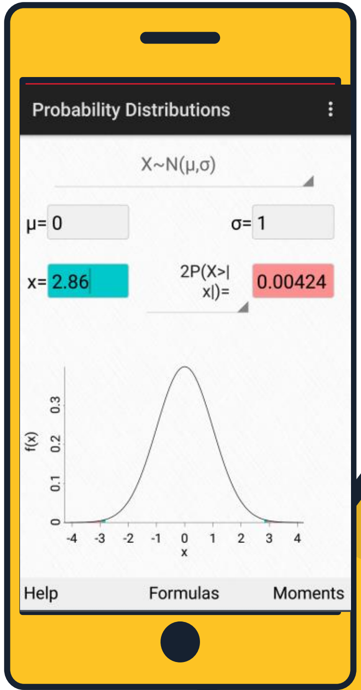
Ejemplo 3

3. Calcular valor p.

$$\widehat{p_A} = \bar{X} = \frac{120}{200} = 0.6 \quad \widehat{p_B} = \bar{Y} = \frac{240}{500} = 0.48 \quad \text{p} = \frac{n\,\bar{X}+m\bar{Y}}{n+m} = 0.514$$

$$z_{observado} = \frac{\bar{X} - \bar{Y} - (0)}{\sqrt{pq\left(\frac{1}{n} + \frac{1}{m}\right)}} = \frac{0.6 - 0.48}{0.042} = 2.86$$

Valor p = P($|Z| > |Z_{observado}|$) = P(| Z| > 2.86)=0.004



Como valor p es pequeño, se rechaza H_0.
La proporción de votantes a favor es diferente en la ciudad A que en la B. Además, como $Z_{observado}$ cae a la derecha en la región de rechazo, podemos decir que la proporción en A es mayor que en B.

Fin de la Parte 4