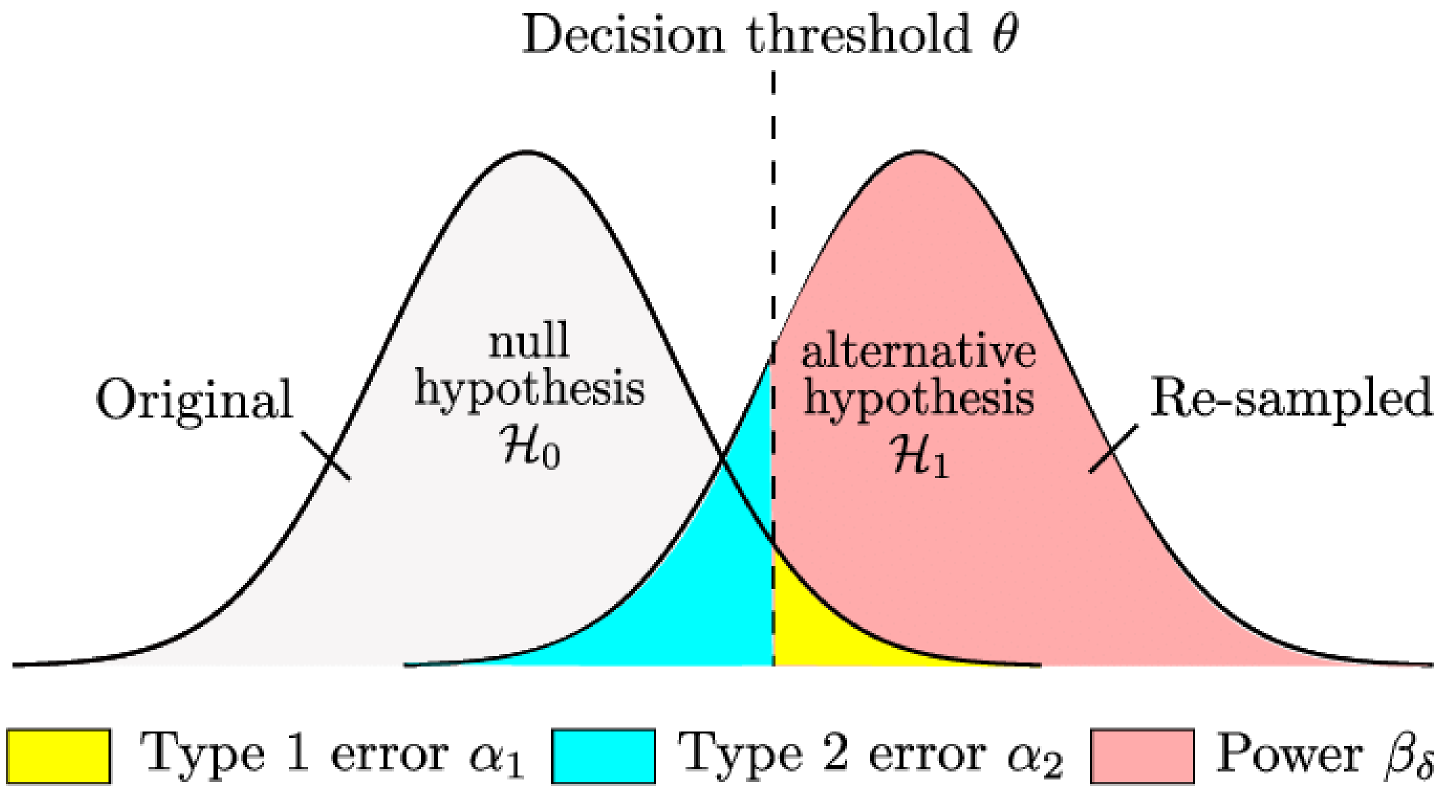
Estadística Computacional



Test de Hipótesis

Braulio Fuentes - Diego Quezada

Temario

- Introducción.
- Hipótesis
- Procedimientos de Test.
    - Test para la media.
    - Test para una proporción de población.
    - Test para la varianza.
    - Valores P.

Introducción

Una *hipótesis estadísticas* es una aseveración o conjetura con respecto a una o más poblaciones. La verdad o falsedad de una hipótesis estadística nunca se sabe con absoluta certidumbre, a menos que examinemos toda la población. En cambio, tomamos una muestra aleatoria de la población de interés, y utilizamos los datos contenidos en esta muestra para proporcionar evidencia que apoye o no la hipótesis.

Hipótesis.

- Se conoce como hipótesis nula, H_0, a la pretensión que inicialmente se supone cierta, **creencia previa**. Mientras aquella aseveración que contradice a esta ultima se concoe como hipótesis alternativa.
- H_0 será rechazada en favor de la hipótesis alternativa sólo si la evidencia muestral sugiere que es falsa.
- Si la muestra no contradice fuertemente a H_0, se dice que no existe suficiente información para rechazar H_0.

> Una prueba de hipótesis es un método de utilizar datos muestrales para decidir si la hipótesis nula debe ser rechazada. 🤓

Procedimientos de Prueba.

Un procedimiento de prueba consiste en:

1. Un **estadístico de prueba**, una función de los datos muestrales en los cuales ha de basarse la decisión.
2. Una **región de critica**, el conjunto de todos los valores estadísticos de prueba por los cuales H_0 será rechazada. El último número que observamos al pasar a la región crítica se llama valor crítico.

> La hipótesis nula *será rechazada* entonces si y sólo si el estadístico de prueba calculado queda en la región de critica. 😵

Errores en el Test de Hipótesis.

Un procedimiento de prueba podría conducir a dos conclusiones erróneas.

- Un *error de tipo I* o nivel de significancia consiste en rechazar H_0 cuando es verdadera (α).
- Un *error de tipo II* implica no rechazar H_0 cuando es falsa. (β)
- Los errores están relacionados, una disminución en la probabilidad de uno produce un incremento en la probabilidad del otro.

- Un aumento en el tamaño muestral n reducirá a α y β de forma simultánea.
- La **potencia** es la probabilidad H_0 sea rechazada cuando la hipótesis alternativa es verdadera ($1-\beta$).

> β es imposible de calcular a menos que tengamos una hipótesis alternativa específica. 💪

Tipos de Prueba.

El método más usado es una prueba de hipótesis con probabilidad fija del error tipo I:

1. Establezca las hipótesis nula y alternativa.
2. Elija un nivel de significancia α fijo.
3. Seleccione un estadístico de prueba adecuado y establezca la región crítica con base en α.
4. A partir del estadístico de prueba calculado, rechace H_0 si el estadístico de prueba está en la región crítica. De otra manera, no rechace H_0.

Pruebas sobre una media de población

Una población normal con σ conocida.

$$\begin{aligned} \text{Hipótesis nula:} \quad & H_0 : \mu = \mu_0 \\ \text{Estadístico de prueba:} \quad & z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \end{aligned}$$

Hipótesis alternativa	Región de rechazo para la prueba de nivel α
$H_a : \mu > \mu_0$	$z \geq z_{1-\alpha}$ (prueba de cola superior)
$H_a : \mu < \mu_0$	$z \leq z_{\alpha}$ (prueba de cola inferior)
$H_a : \mu \neq \mu_0$	$z \geq z_{1-\alpha/2}$ o $z \leq z_{\alpha/2}$ (prueba dos colas)

- **Error de Tipo II**

Hipótesis alternativa	Probabilidad de error de tipo II para una prueba de nivel α
$H_a : \mu > \mu_0$	$\Phi\left(z_{1-\alpha} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
$H_a : \mu < \mu_0$	$1 - \Phi\left(z_{\alpha} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
$H_a : \mu \neq \mu_0$	$\Phi\left(z_{1-\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$

- **Tamaño de la muestra**

  El tamaño de una muestra n con el cual una prueba de nivel α también tiene β con unn valor alternativo μ' es:

$$n = \begin{cases} \left[\dfrac{\sigma(z_{1-\alpha} + z_{1-\beta})}{\mu_0 - \mu'}\right]^2 & \text{para una prueba de una cola} \\ \left[\dfrac{\sigma(z_{1-\alpha/2} + z_{1-\beta})}{\mu_0 - \mu'}\right]^2 & \text{para una prueba de dos cola} \end{cases}$$

Pruebas con muestras grandes

Para este caso usamos el estadístico de prueba:

$$z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

Se utilizará de nuevo la regla empírica $n > 40$ para caracterizar un tamaño de muestra grande.

Una población normal con σ desconocida

Hipótesis nula: $H_0 : \mu = \mu_0$

Estadístico de prueba: $t = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$

Hipótesis alternativa	Región de rechazo para la prueba de nivel α
$H_a : \mu > \mu_0$	$t \geq t_{1-\alpha,n-1}$ (prueba de cola superior)
$H_a : \mu < \mu_0$	$t \leq t_{\alpha,n-1}$ (prueba de cola inferior)
$H_a : \mu \neq \mu_0$	$t \geq t_{1-\alpha/2,n-1}$ o $t \leq t_{\alpha/2,n-1}$ (prueba dos colas)

Pruebas con una proporción de población

$$\begin{aligned}
\text{Hipótesis nula:} \quad & H_0 : p = p_0 \\
\text{Estadístico de prueba:} \quad & z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}
\end{aligned}$$

$$\begin{array}{rl}
\text{Hipótesis alternativa} & \text{Región de rechazo para la prueba de nivel } \alpha \\
H_a : p > p_0 & z \geq z_{1-\alpha} \quad \text{(prueba de cola superior)} \\
H_a : p < p_0 & z \leq z_{\alpha} \quad \text{(prueba de cola inferior)} \\
H_a : p \neq p_0 & z \geq z_{1-\alpha/2} \text{ o } z \leq z_{\alpha/2} \quad \text{(prueba dos colas)}
\end{array}$$

- **Error de Tipo II**

Hipótesis alternativa	Probabilidad de error de tipo II $\beta(p')$ para una prueba de nivel α
$H_a : p > p_0$	$\Phi\left[\frac{p_0 - p' + z_{1-\alpha}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$
$H_a : p < p_0$	$1 - \Phi\left[\frac{p_0 - p' + z_{1-\alpha}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$
$H_a : p \neq p_0$	$\Phi\left[\frac{p_0 - p' + z_{1-\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right] - \Phi\left[\frac{p_0 - p' + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

- **Tamaño de la muestra**

  El tamaño de una muestra n con el cual una prueba de nivel α también tiene β con un valor alternativo p' es:

$$
n = \begin{cases} \left[\dfrac{(z_{1-\alpha}\sqrt{p_0(1-p_0)} + z_{1-\beta}\sqrt{p'(1-p')})}{p' - p_0}\right]^2 & \text{para una prueba de una cola} \\ \left[\dfrac{(z_{1-\alpha/2}\sqrt{p_0(1-p_0)} + z_{1-\beta}\sqrt{p'(1-p')})}{p' - p_0}\right]^2 & \text{para una prueba de dos cola} \end{cases}
$$

Pruebas con una varianza de población.

Hipótesis nula: $H_0 : \sigma^2 = \sigma_0^2$

Valor del estadístico de prueba: $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$

Hipótesis alternativa	Región de rechazo para la prueba de nivel α
$H_a : \sigma^2 > \sigma_0^2$	$\chi^2 \geq \chi^2_{1-\alpha,n-1}$ (prueba de cola superior)
$H_a : \sigma^2 < \sigma_0^2$	$\chi^2 \leq \chi^2_{\alpha,n-1}$ (prueba de cola inferior)
$H_a : \sigma^2 \neq \sigma_0^2$	$\chi^2 \geq \chi^2_{1-\alpha/2,n-1}$ o $\chi^2 \leq \chi^2_{\alpha/2,n-1}$ (prueba dos colas)

Tipos de Prueba.

Otro método seguido es una prueba de significancia, aproximación al valor P:

1. Establezca las hipótesis nula y alternativa.
2. Elija un estadístico de prueba adecuado.
3. Calcule el valor P con base en los valores calculados del estadístico de prueba.
4. Utilice el juicio con base en el valor P.

Valores P.

El nivel de significación observado, P, es el nivel de significación más pequeño al cual H_0 sería rechazada.
Una vez que se ha determinado el valor P, la conclusión a un nivel particular α resulta de comparar el valor P con α:

- $$\text{Valor } P \leq \alpha : \text{rechazar } H_0 \text{ al nivel } \alpha.$$
- $$\text{Valor } P > \alpha : \text{no rechazar } H_0 \text{ al nivel } \alpha.$$

Valor de P para una distribución cualquiera.

Dado un estadístico de prueba x que sigue una distribución X, el valor P se calcula de la siguiente forma dependiendo del test:

$$P = \begin{cases} P(X \geq x) & \text{para una test de cola superior} \\ P(X \leq x) & \text{para una test de cola inferior} \\ 2 \cdot P(X \geq x) & \text{para una test de dos colas} \end{cases}$$

> Generalmente, cuando P es un valor muy pequeño se rechaza la hipótesis nula. 🤫