Clase 12/06/2023

En esta clase desarrollaremos el tema de Pruebas de hipótesis.

Hipótesis estadística.

Una hipótesis estadística es un enunciado (o afirmación) que concierne a las características (valores de los parámetros forma de distribución de las observaciones) de una población.

Prueba de hipótesis

Una prueba de hipótesis (o test de hipótesis) es un procedimiento que tiene por fin establecer una regla de decisión que permita, basada en los resultados de una muestra, realizar una elección entre dos hipótesis estadísticas.

Hipótesis nula (H_0) e hipótesis alternativa (H_1)

La hipótesis bajo la cual nos fija a priori un valor particular se llama hipótesis nula y se denota por H_0. Cualquier otra hipótesis que difiera de H_0 la llamamos hipótesis alternativa. y se denota por H_1.

Importante: Es la hipótesis nula H_0 la que se somete a prueba y todo el procedimiento de la prueba se hace considerando que esta hipótesis es verdadera.

Lo que desarrollaremos a continuación serán reglas de decisión que nos conducirán a la aceptación o rechazo de la hipótesis nula H_0.

Definición 1: Dada una muestra aleatoria simple $X_1, X_2, \ldots, X_n$ con distribución F_X donde alguna de las componentes de F_X es desconocida (un parámetro o toda la distribución) realizar una prueba de hipótesis es tomar una decisión entre la hipótesis H_0 y H_1. La toma de decisión consiste en dos posibilidades "rechazo H_0" o "no rechazo H_0".

Observación: La realidad (lo que realmente pasa) es H_0 o H_1, pero no sabemos cual de las dos es cierta. A partir de la información que nos suministra la M.A.S $(X_1, \ldots, X_n)$ se trata de tomar una decisión sobre si H_0 es cierta o H_1 es cierta, pero podemos cometer errores en la decisión.

realidad

	H_0	H_1
rechazo H_0	error	correcto
no rechazo H_0	correcto	error.

Procedimiento general para realizar una prueba de hipótesis

1.- Plantear H_0 y H_1.

2.- Elegir un valor α y construir una región crítica de nivel α es decir

$$\sup_{H_0} P(RC) = \alpha$$

3.- Tomar la decisión (si la muestra cumple la región crítica rechazo H_0 y si no la cumple RC no rechazo H_0.

Instrumentemos un ejemplo.

Vamos a construir una prueba de hipótesis que busca dilucidar si una M.A.S, $X_1, \ldots, X_n$ proviene de una distribución que tiene media μ.

1. etapa: Formulación de las hipótesis
Vamos a probar H_0 vs H_1 cuando

$$\begin{cases} H_0 : m = \mu_0 \\ H_1 : m \neq \mu_0 \end{cases}$$

2. etapa. Determinar el estadístico discriminante de la prueba

(a) Se determina el estadístico que conviene para esta prueba. Para esta prueba $\bar{X}_n$ parece el más indicado.

(b) Se determina la distribución de $\bar{X}_n$ cuando consideramos que la hipótesis H_0 es cierta.

Para la determinación de la distribución se nos presentan dos casos (4)

1.- Primer caso. El valor de n es grande por ejemplo $n \geq 30$ o la varianza σ^2 es conocida.

Sabemos que bajo H_0 $\bar{X}_n$ ~~sigue~~ posee una distribución $\stackrel{d}{=} N(\mu_0, \frac{\sigma^2}{n})$

Entonces definimos el estadístico de la prueba por

$$T = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$$

Y sabemos que bajo H_0 $T \stackrel{d}{=} N(0,1)$.

T será entonces el estadístico discriminante de la prueba

2.- Segundo caso: La muestra aleatoria simple es de tamaño pequeño y además la varianza σ^2 es desconocida. Entonces si S^2 es la varianza empírica el estadístico de la prueba ahora es

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n-1}}$$ si suponemos que $X_j \stackrel{d}{=} N(\mu_0, \sigma^2)$

Sabemos que bajo H_0 $T \stackrel{d}{=} t_{n-1}$ (t de Student con $n-1$ grados de libertad)

3er etapa: Determinación de la región crítica a partir de T que será la zona de rechazo de nivel α.

para el caso en que conocemos la varianza $(n \geq 30)$ para sabemos que

$$P(|T| \geq c_{\alpha/2}) = \alpha$$

Donde $c_{\alpha/2}$ se define como el número positivo tal que $Z = N(0,1)$

$$P(T \geq c_{\alpha/2}) = \alpha/2$$

esto implica que $P(|Z| \geq c_{\alpha/2}) = \alpha.$

y luego Rechazamos Ho en la región

$$RC_\alpha = (-\infty, -c_{\alpha/2}] \cup [c_{\alpha/2}, \infty)$$

En el caso en que $n < 30$ y σ^2 es desconocida Construimos $t_{\alpha/2}$

$$P(-t_{\alpha/2} \leq t_{n-1} \leq t_{\alpha/2}) = 1-\alpha$$

luego la región crítica será

$$RC_\alpha = (-\infty, -t_{\alpha/2}] \cup [t_{\alpha/2}, \infty)$$

decimos en este caso que la prueba es bilateral.

4ta <u>etapa</u>: Cálculo del valor de T para la muestra y decisión.

- Calculamos el valor de T para la muestra observada.
- Si el valor observado se encuentra en la región de rechazo rechazamos H_0 al nivel de α. De otra formas aceptamos H_0.

<u>Ejemplo</u>: Supongamos que el contenido del líquido que genera una máquina de café puede modelarse como $L = \mu + Z$ donde $Z \stackrel{d}{=} N(0,1)$
Un cliente inserta 50 peso y la máquina debería entregar exactamente 150 ml de café. Entonces si la máquina entrega + de 150 el propietario no quedará contento y si entrega menos de 150 es el cliente el que se molestaría.

Se toma una muestra

144, 154, 156, 144, 150, 157, 144, 143, 142

Queremos hacer una prueba de Hipótesis

1.- $H_0: \mu = 150$

2.- Ver si es una prueba bilateral. En cualquiera de las dos situaciones estaríamos en problemas esto es $\mu > 150$ o $\mu < 150$ entonces (7)

$H_1: \mu \neq 150$ (bilateral).

3.- Debemos elegir un estadístico.
Como vimos es conveniente tomar el estadístico estandarizado

$$\boxed{Z = \frac{\bar{L}_9 - 150}{5/\sqrt{9}} = -1.067}$$

4. Debemos elegir un nivel; Tomamos $\alpha = 0.05$ en la tabla gaussiana nos da 1.96.
Luego aceptaremos la hipótesis nula.

Ahora debemos introducir dos tipos de errores

(a) Error de primera especie
es ~~rechazo~~ la probabilidad de rechazar H_0 siendo cierta

$$\boxed{P(RC / H_0) = \alpha .}$$

(b) Error de segunda especie

$$\boxed{\beta = P(\text{No rechazo } H_0 / H_1)}$$

(c) La potencia del test o prueba es

$$\boxed{1-\beta = \text{Potencia de la prueba.}}$$

Entonces queremos que α sea pequeño y $1-\beta$ grande.

$$\cancel{P\left(|T| \leq c_{\alpha/2} \,|\, H_1\right) = 2\Phi\left(c_{\alpha/2} + \frac{(\mu_1-\mu_0)}{\sigma/\sqrt{n}}\right) - 1}$$

En el caso de la prueba bilateral de la media con σ^2 conocida podemos calcular la potencia de

$$\left.\begin{array}{l} H_0: \mu = \mu_0 \\ H_1: \mu = \mu_1 \end{array}\right\} \qquad \mu_1 > \mu_0 \text{ (por ejemplo}$$

$$T_n = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \qquad T_n \overset{d}{\cong} N(0,1) \text{ bajo } H_0.$$

pero bajo H_1

$$T_n = \frac{\bar{X}_n - \mu_1 + \mu_1 - \mu_0}{\sigma/\sqrt{n}}$$

$$T_n = \frac{\bar{X}_n - \mu_1}{\sigma/\sqrt{n}} + \sqrt{n}\,\frac{(\mu_1-\mu_0)}{\sigma}$$

$$P_{H_1}\left((-\infty, -c_{\alpha}] \cup [c_{\alpha/2}, \infty) \,|\, H_1\right)$$

$$= P\left(|T_n| > c_{\alpha/2}\right)$$

$$= P\left(T_n > c_{\alpha/2}\right) + P\left(T_n < -c_{\alpha/2}\right)$$

$$= P\left(\frac{\bar{X}_n - \mu_1}{\sigma/\sqrt{n}} > c_{\alpha/2} - \frac{\sqrt{n}\,(\mu_1-\mu_0)}{\sigma}\right) + P\left(\frac{\bar{X}_n-\mu_1}{\sigma/\sqrt{n}} < -c_{\alpha/2} \mp \frac{\mu_1-\mu_0}{\sigma\sqrt{n}}\right)$$

$$= 1 - \Phi\left(c_{\alpha/2} - \frac{\sqrt{n}\,(\mu_1-\mu_0)}{\sigma}\right) + \Phi\left(-c_{\alpha/2} - \frac{\sqrt{n}\,(\mu_1-\mu_0)}{\sigma}\right).$$