EAS201a - Inferencia Estadística

Escuela de Administración UC

Material de Apoyo

Ayudantía 8

Tópicos de la Ayudantía

- Test de Hipótesis
- Ejercicios
- Ejercicios Propuestos

Test de Hipótesis

Introducción

- Se considera la estimación del parámetro desconocido θ, de tal manera que deba pertenecer a un cierto espacio paramétrico Ω
- Suponga que $\Omega = \Omega_0 \cup \Omega_1$ donde Ω_0 y Ω_1 son disjuntos (Partición de Ω)
- Se define la Hipótesis Nula: H_0 como la hipótesis de que $\theta \in \Omega_0$
- Se define la Hipótesis Alternativa: H_1 como la hipótesis de que $\theta \in \Omega_1$
- Un contraste de hipótesis es aquel que a partir de evidencia muestral buscar rechazar o aceptar la hipótesis planteada por el investigador

Test de Hipótesis

Introducción

- Suponga que $X_1, \ldots, X_n$ constituye una m.a de una población $f(x|\theta)$, donde el parámetro θ debe pertenecer al espacio paramétrico Ω. Sea Ω_0 y Ω_1 una partición de Ω. Se quiere contrastar las siguientes hipótesis:

$$\begin{aligned} H_0 &: \quad \theta \in \Omega_0 \\ H_1 &: \quad \theta \in \Omega_1 \end{aligned}$$

- Si el conjunto Ω_i sólo puede contener un valor de θ, entonces la hipótesis H_i es una Hipótesis Simple
- Si el conjunto Ω_i contiene más de un valor de θ, entonces la hipótesis H_i es una Hipótesis Compuesta

- Toda prueba de hipótesis conlleva dos tipos de Errores.
- Un Error de Tipo I consiste en rechazar la hipótesis nula H_0 cuando es verdadera.
- Un Error de Tipo II consiste en no rechazar H_0 cuando H_0 es falsa.

$$\begin{aligned}\alpha(\theta) &= P(\text{Error tipo I}) = P(\text{Rechazar} H_0|H_0)\\ \beta(\theta) &= P(\text{Error tipo II}) = P(\text{No rechazar} H_0|H_1)\end{aligned}$$

- $\alpha(\theta)$ y $\beta(\theta)$ son inversamente proporcionales

- La Región Crítica corresponde a la región donde se rechazará H_0.
- Para obtener una región crítica, se debe previamente especificar el máximo valor de $\alpha(\theta)$ tolerable, es decir la máxima probabilidad tolerable para el Error de Tipo I. Este valor se conoce como el nivel de significancia o riesgo del test.
- Se denotará por α y se calcula como

$$\alpha = \max_{\theta \in \Theta_0} \alpha(\theta)$$

- La Potencia del Test se define por

$$\pi(\theta) \quad = \quad P(\text{Rechazar} H_0|\theta)$$

- La idea de un test es maximizar esta potencia una vez que α es fijado.
- Así para un α fijo uno debiera escoger la región de rechazo con mayor potencia del test.

Test de Hipótesis

Ejercicio 1

Suponga que $X_1, ..., X_n$ es una m.a de una distribución Normal con media μ es desconocida y cuya varianza es 1. Suponga además que μ_0 es un número específico y que se quiere contrastar las siguiente hipótesis

$$\begin{aligned} H_0 : &\quad \mu = \mu_0 \\ H_1 : &\quad \mu \neq \mu_0 \end{aligned}$$

Considere el procedimiento de prueba tal que se rechaza H_0 cuando $|\overline{X}_n - \mu_0| \geq c$. Determine el valor de c para que el nivel de significancia de la prueba sea de α.

Se quiere contrastar las siguientes hipótesis:

$$\begin{aligned} H_0 : &\quad \mu = \mu_0 \\ H_1 : &\quad \mu \neq \mu_0 \end{aligned}$$

donde se rechaza H_0 si $|\overline{X}_n - \mu_0| \geq c$.

Para determinar el valor de c, se hace fijando el nivel de significancia de la prueba y como Θ_0 contiene sólo un valor μ_0, se tiene que $\alpha(\mu) = \alpha$ (no es función de μ), luego

$$\begin{aligned} \alpha &= P(\text{Rechazar } H_0 | \mu \in \Theta_0) \\ &= P(|\overline{X}_n - \mu_0| \geq c | \mu \in \Theta_0) \\ &= 1 - P(-c \leq \overline{X}_n - \mu_0 \leq c | \mu \in \Theta_0) \qquad (1) \end{aligned}$$

Para cacular dicha probabilidad, debemos conocer la distribución de $\overline{X}_n - \mu_0$ bajo H_0, es decir, cuando $\mu = \mu_0$, ya que Θ_0 contiene sólo el valor de μ_0.

Luego como $X_1, ..., X_n \sim N(\mu, 1)$, entonces $\overline{X}_n \sim N(\mu, \frac{1}{n})$.

Luego Bajo H_0, se tiene que $\mu = \mu_0$.

De esta manera, el estadístico $\overline{X}_n$ distribuye $N(\mu_0, \frac{1}{n})$.
Estandarizando se tiene $\frac{\overline{X}_n - \mu_0}{1/\sqrt{n}} \sim N(0, 1)$

Así podemos calcular α, retomando la ecuación (1):

$$\begin{aligned}
\alpha &= 1 - P(-c \leq \overline{X}_n - \mu_0 \leq c | \mu \in \Theta_0) \\
&= 1 - P(-c \leq \overline{X}_n - \mu_0 \leq c | \mu = \mu_0) \\
&= 1 - P\left(-\frac{c}{1/\sqrt{n}} \leq \frac{\overline{X}_n - \mu_0}{1/\sqrt{n}} \leq \frac{c}{1/\sqrt{n}} \middle| \mu = \mu_0\right)
\end{aligned}$$

Como $\frac{\overline{X}_n - \mu_0}{1/\sqrt{n}}$ Bajo H_0 es N(0,1) se tiene

Test de Hipótesis

Solución Ejercicio 1

$$
\begin{aligned}
\alpha &= 1 - P\left(-\frac{c}{1/\sqrt{n}} \leq Z \leq \frac{c}{1/\sqrt{n}}\right) \\
\alpha &= 1 - \left(\Phi\left(\frac{c}{1/\sqrt{n}}\right) - \Phi\left(-\frac{c}{1/\sqrt{n}}\right)\right) \\
\alpha &= 1 - \left(\Phi\left(\frac{c}{1/\sqrt{n}}\right) - \left(1 - \Phi\left(\frac{c}{1/\sqrt{n}}\right)\right)\right) \\
\alpha &= 1 - \left(2\Phi\left(\frac{c}{1/\sqrt{n}}\right) - 1\right) \\
1 - \frac{\alpha}{2} &= \Phi\left(\frac{c}{1/\sqrt{n}}\right) \\
\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) &= \frac{c}{1/\sqrt{n}}
\end{aligned}
$$

donde $\Phi^{-1}(\alpha)$ corresponde al cuantil que acumula α en una N(0,1).

Test de Hipótesis

Solución Ejercicio 1

Recuerde que en este curso hemos denotado al cuantil que acumula α en una N(0,1) por z_α, luego,

$$z_{1-\frac{\alpha}{2}} = \frac{c}{1/\sqrt{n}} \Rightarrow c = \frac{1}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}$$

De esta manera, hemos determinado explícitamente la región de rechazo del test, en donde se rechaza H_0 si

$$|\overline{X}_n - \mu_0| \geq \frac{1}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}$$

Equivalentemente, se rechaza H_0 si

$$\left| \frac{\overline{X}_n - \mu_0}{\frac{1}{\sqrt{n}}} \right| \geq z_{1-\frac{\alpha}{2}}$$

Test de Hipótesis

Ejercicio 2

Suponga que $X_1, ..., X_{25}$ constituye una m.a de una distribución N(μ, σ^2), donde ambos parámetros son desconocidos. Se desea contrastar las siguientes hipótesis

$$
\begin{aligned}
H_0 &: \quad \mu = \mu_0 \\
H_1 &: \quad \mu > \mu_0
\end{aligned}
$$

Considere el estadístico $T = \frac{\overline{X}_{25} - \mu_0}{S/\sqrt{25}}$, donde S es la desviación estándar muestral. Considere el procedimiento de prueba tal que se rechaza H_0 cuando $T \geq 1.32$.

(a) Determine el nivel de significancia del test

(b) Determine la función Potencia del test

(a) El nivel de significancia del test corresponde a

$$\alpha = \max_{\mu \in \Theta_0} \alpha(\mu)$$

Note que Θ_0 contiene solo un valor:$\mu = \mu_0$, luego $\alpha(\mu) = \alpha(\mu_0)$, por lo tanto el nivel de significancia del test se reduce $\alpha = \alpha(\mu_0)$.

$$\begin{aligned} \alpha &= \alpha(\mu_0) = P(\text{Rechazar } H_0 | \mu \in \Theta_0) \\ &= P(T \geq 1.32 | \mu = \mu_0) \end{aligned} \tag{2}$$

Como $X_1, ..., X_{25} \sim N(\mu, \sigma^2)$, se tiene que:

$$\frac{\overline{X}_{25} - \mu}{S/\sqrt{25}} \sim t_{24}$$

Así Bajo H_0: $(\mu = \mu_0)$,

$$\frac{\overline{X}_{25} - \mu_0}{S/\sqrt{25}} \sim t_{24}$$

Luego, volviendo a la ecuación (2)

$$\begin{aligned} \alpha &= P(T \geq 1.32 | \mu = \mu_0) \\ \alpha &= 1 - P(T < 1.32) \quad \text{donde } T \sim t_{24} \\ 1 - \alpha &= P(T < 1.32) \end{aligned}$$

Por lo tanto, el cuantil que acumula $1 - \alpha$ de una t-student con 24 grados de libertad es 1.32, así

$$t^{24}_{1-\alpha} = 1.32$$

Buscando en la tabla se tiene que $1 - \alpha = 0.90 \Rightarrow \alpha = 0.1$

(b) La potencia del test se define como $\pi(\mu) = \alpha(\mu) \quad$ para $\mu \in \Theta_0$, luego $\pi(\mu) = \pi(\mu_0) = 0.1$.
Para $\mu \in \Theta_1$,

$$\begin{aligned} \pi(\mu) &= P(\text{Rechazar } H_0 | \mu \in \Theta_1) \\ &= P(\text{Rechazar } H_0 | \mu > \mu_0) \\ &= P(T \geq 1.32 | \mu > \mu_0) \end{aligned}$$

Note que el estadístico $T = \frac{\overline{X}_{25} - \mu_0}{S/\sqrt{25}}$ bajo $H_1 : (\mu > \mu_0)$ no tiene distribución t_{24}, ya que Bajo H_1, $\frac{\overline{X}_{25} - \mu}{S/\sqrt{25}}$ con $\mu > \mu_0$ distribuye t_{24}

Test de Hipótesis

Solución Ejercicio 2

Luego,

$$\begin{aligned}
\pi(\mu) &= P(T \geq 1.32|\mu > \mu_0) \\
&= P\left(\frac{\overline{X}_{25} - \mu_0}{S/\sqrt{25}} \geq 1.32|\mu > \mu_0\right) \\
&= P\left(\overline{X}_{25} - \mu_0 \geq 1.32\frac{S}{\sqrt{25}}|\mu > \mu_0\right) \\
&= P\left(\overline{X}_{25} \geq \mu_0 + 1.32\frac{S}{\sqrt{25}}|\mu > \mu_0\right) \\
&= P\left(\frac{\overline{X}_{25} - \mu}{S/\sqrt{25}} \geq \frac{\mu_0 - \mu}{S/\sqrt{25}} + 1.32|\mu > \mu_0\right)
\end{aligned}$$

Como $T_1 = \frac{\overline{X}_{25} - \mu}{S/\sqrt{25}}$ bajo H_1 tiene distribución t_{24}, se tiene que

$$\pi(\mu) \quad = \quad 1 - F_{T_1}\left(\frac{\mu_0 - \mu}{S/\sqrt{25}} + 1.32\right)$$

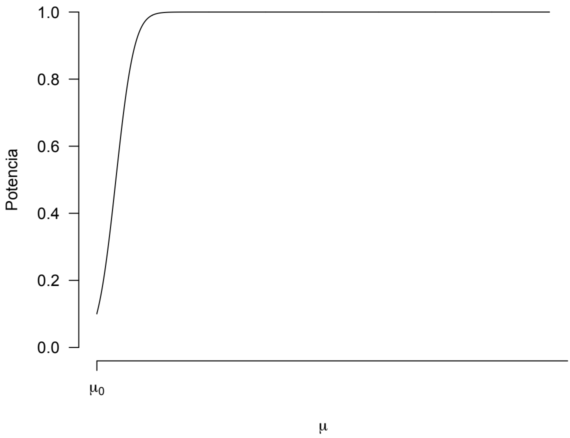
Por lo tanto, la Función Potencia del test está dada por

$$\pi(\mu) = \begin{cases} 0.1 & \mu = \mu_0 \\ 1 - F_{T_1}\left(\frac{\mu_0-\mu}{S/\sqrt{25}} + 1.32\right) & \mu > \mu_0 \end{cases}$$

Cuyo gráfico es el siguiente:

Test de Hipótesis

Solución Ejercicio 2



Una mezcla de combustible y cemento se utiliza para techar y ésta debe tener una resistencia a la comprensión de más de 1300. La mezcla no se utilizará a menos que haya una evidencia experimental que indique de manera concluyente que se ha cumplido la especificación de resistencia. Supongamos que la resistencia a la comprensión esta mezcla distribuye normal con $\sigma = 60$. Represente con μ la media poblacional de resistencia, y se considera una muestra de tamaño $n = 20$,

(a) ¿Cuáles son las hipótesis nula y alternativas adecuadas?

(b) Considere el procedimiento de prueba utilizando el estadístico $\overline{X}$ y la región de rechazo $\overline{X} \geq 1331.26$. ¿Cuál es la distribución del estadístico de prueba cuando H_0 es verdadera? ¿Cuál es la probabilidad de un error de tipo I? ¿Cuál es el nivel de significancia del test?

Test de Hipótesis

Ejercicio 3

(c) ¿Cuál es la distribución del estadístico de prueba cuando $\mu = 1.350$? Mediante el procedimiento de prueba usado en la parte *b*),¿cuál es la probabilidad de que la mezcla se considere no satisfactoria cuando $\mu = 1350$?

(d) ¿Cómo cambiaría el procedimiento de prueba de la parte *b*) para obtener una prueba con nivel de significancia 0.05?

Test de Hipótesis

Solución Ejercicio 3

(a) Sea μ la media poblacional de resistencia. Se quiere testear si

$$
\begin{aligned}
H_0 &: \quad \mu \leq 1300 \\
H_1 &: \quad \mu > 1300
\end{aligned}
$$

(b) Sea $X_1, ..., X_{20}$ la m.a de resistencias que distribuye Normal$(\mu, 60^2)$. Sea $\overline{X}$ el estadístico de prueba del test, el cual distribuye Normal$(\mu, \frac{60^2}{20})$.

Luego, Bajo H_0, $\mu \leq 1300$, se tiene que $\overline{X} \sim N(\mu, \frac{60^2}{20})$

La probabilidad de Error de Tipo I, está dada por

$$
\begin{aligned}
\alpha(\mu) &= P(\text{ Rechazar } H_0 | \mu \leq 1300) \\
&= P(\overline{X} \geq 1331.26 | \mu = 1300) \\
&= P\left(\frac{\overline{X} - \mu}{60/\sqrt{20}} \geq \frac{1331.26 - \mu}{60/\sqrt{20}} \right) \\
&= 1 - \Phi\left(\frac{1331.26 - \mu}{60/\sqrt{20}} \right)
\end{aligned}
$$

Test de Hipótesis

Solución Ejercicio 3

El nivel de significancia del test es

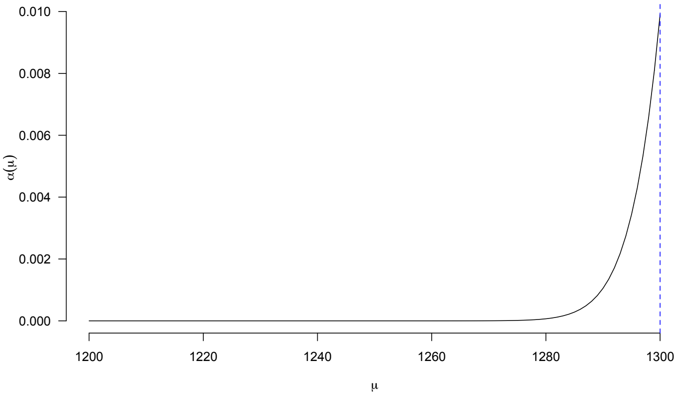
$$\begin{aligned}\alpha &= \max_{\mu \leq 1300} \alpha(\mu) \\ &= \max_{\mu \leq 1300} 1 - \Phi\left(\frac{1331.26 - \mu}{60/\sqrt{20}}\right)\end{aligned}$$

La cual es máxima para $\mu = 1300$ (Ver el gráfico en la siguiente slide). Luego α está dado por

$$\begin{aligned}\alpha &= 1 - \Phi\left(\frac{1331.26 - 1300}{60/\sqrt{20}}\right) \\ &= 1 - \Phi(2.329) = 0.0099\end{aligned}$$

Test de Hipótesis

Solución Ejercicio 3



(c) Bajo H_1, $\mu > 1300$ el estadístico tiene distribución

$$\overline{X} \sim N\left(\mu, \frac{60^2}{20}\right)$$

Luego en el caso particular que nos preguntan cuando $\mu = 1350$

$$\overline{X} \sim N\left(1350, \frac{60^2}{20}\right)$$

Ahora nos piden la probabilidad de Error de Tipo II,

$$\begin{aligned} \beta(\mu) &= P(\text{No Rechazar } H_0 | \mu > 1300) \\ &= P(\overline{X} < 1331.26 | \mu > 1300) \end{aligned}$$

Como $\overline{X}$ Bajo H_1 distribuye $N\left(\mu, \frac{60^2}{20}\right)$, con $\mu > 1300$, se tiene que

$$\begin{aligned} \beta(\mu) &= P\left(\frac{\overline{X}-\mu}{60/\sqrt{20}} < \frac{1331.26-\mu}{60\sqrt{20}} \Big| \mu > 1300\right) \\ &= \Phi\left(\frac{1331.26-\mu}{60\sqrt{20}}\right) \end{aligned}$$

En el caso particular que nos preguntan para $\mu = 1350$, se tiene

$$\begin{aligned}\beta(1350) &= \Phi\left(\frac{1331.26 - 1350}{60\sqrt{20}}\right) \\ &= \Phi(-1.39) \\ &= 1 - \Phi(1.39) \\ &= 1 - 0.9177 \\ &= 0.0823\end{aligned}$$

(d) Queremos que la prueba tenga un nivel de significancia $\alpha = 0.05$. Luego debemos modificar la región de rechazo del test, modificando el punto de corte. Sea c el punto de corte, entonces se rechaza H_0 si $\overline{X} \geq c$. Para determinar c, fijamos el nivel de significancia del test $\alpha = 0.05$, así

$$\alpha = \max_{\mu \leq 1300} \alpha(\mu)$$

Como vimos anteriormente $\alpha(\mu)$ es una función creciente en μ, luego su máximo valor lo alcanza en $\mu = 1300$, luego

$$
\begin{aligned}
\alpha &= P(\overline{X} \geq c | \mu = 1300) \\
0.05 &= P\left(\frac{\overline{X} - 1300}{60/\sqrt{20}} \geq \frac{c - 1300}{60/\sqrt{20}} \middle| \mu \leq 1300 \right) \\
0.05 &= 1 - \Phi\left(\frac{c - 1300}{60/\sqrt{20}} \middle| \mu \leq 1300 \right) \\
0.95 &= \Phi\left(\frac{c - 1300}{60/\sqrt{20}} \middle| \mu \leq 1300 \right)
\end{aligned}
$$

$$
\begin{aligned}
\Phi^{-1}(0.95) &= \frac{c - 1300}{60/\sqrt{20}} \\
z_{0.95} &= \frac{c - 1300}{60/\sqrt{20}} \\
c &= 1300 + z_{0.95}\frac{60}{\sqrt{20}} \\
&= 1300 + 1.64\frac{60}{\sqrt{20}} \\
&= 1322.003
\end{aligned}
$$

Por lo tanto, para un nivel de significancia del test de $\alpha = 0.05$ se rechaza H_0 si

$$\overline{X} \geq 1322.003$$

Test de Hipótesis

Ejercicio 4

Suponga que $X_1, ..., X_n$ constituye una m.a de una distribución $\mathrm{U}(0, \theta)$ y se desea contrastar las siguientes hipótesis

$$\begin{aligned} H_0 : &\quad \theta \geq 2 \\ H_1 : &\quad \theta < 2 \end{aligned}$$

Sea $T = \max\{X_1, ..., X_n\}$. Considere un procedimiento de prueba tal que se rechaza H_0 si $T \leq 1.5$. Determine el nivel de significancia del test

El nivel de significancia del test corresponde a

$$\alpha = \max_{\theta \in \Theta_0} \alpha(\theta)$$

En este caso $\Theta_0 = \{\theta, \, \theta \geq 2\}$

$$\begin{aligned} \alpha(\theta) &= P(\text{Rechazar } H_0 | \theta \in \Theta_0) \\ &= P(T \leq 1.5 | \theta \geq 2) \end{aligned}$$

Para calcular dicha probabilidad debemos conocer la distribución del estadístico T bajo H_0.

Test de Hipótesis

Solución Ejercicio 4

Recuerde que la densidad de T está dada por

$$f_T(t) = nF_X(t)^{n-1} f_X(t), \quad 0 \leq t \leq \theta$$

Luego como $X \sim U(0, \theta)$ se tiene que

$$f_T(t) = n \left(\frac{t}{\theta}\right)^{n-1} \frac{1}{\theta} = n\frac{t^{n-1}}{\theta^n}$$

Así para $\theta \geq 2$,

$$\begin{aligned} \alpha(\theta) &= P(T \leq 1.5 | \theta \geq 2) \\ &= \int_0^{1.5} n\frac{t^{n-1}}{\theta^n}\, dt = \left(\frac{1.5}{\theta}\right)^n \end{aligned}$$

Luego el nivel de significancia del test está dado por

$$\alpha = \max_{\theta \in \Theta_0} \alpha(\theta) = \max_{\theta \geq 2} \left(\frac{1.5}{\theta}\right)^n = \left(\frac{1.5}{2}\right)^n$$

Pues es una función decreciente en θ.