Matemática 3

PRACTICA 6

Una población está formada por la totalidad de las observaciones en las cuales se tiene cierto interés.

Una muestra es un subconjunto de observaciones seleccionadas de una población.

Las variables aleatorias X1 , X2 , ... , Xn constituyen una muestra aleatoria de tamaño n de una v.a. X si X1 , X2 , ... , Xn son independientes e idénticamente distribuidas

Parámetro: θ

Algunos de los parámetros poblacionales más usuales que estimaremos son: μ , λ , σ^2 y p

Se estiman parámetros poblacionales.

Estimador : $\widehat{\Theta}$ = f (X1 , X2 , ... , Xn)

El estimador de un parámetro θ, tambien se lo llama **estadístico** ya que es una función de la muestra aleatoria.

Estimación: $\hat{\theta}$= f (x1 , x2 , ... , xn)

x1 , x2 , ... , xn son los **valores numéricos** obtenidos de la muestra aleatoria.

El estimador es la fórmula y la estimación es el número (es en función de la observación). Hay que diferenciarlos bien.

Un estimador (insesgado) de μ es la media muestral y se puede anotar también como $\hat{\mu}$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Un estimador (insesgado) de σ^2 es la varianza muestral y un estimador (insesgado) de σ es la desviación estándar muestral. Se pueden anotar también como $\widehat{\sigma^2}$ y $\hat{\sigma}$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

Estimadores insesgados

Un **estimador puntual** $\hat{\theta}$ es un estimador insesgado del parámetro θ si

E $\widehat{\Theta}$= θ cualquiera sea el valor verdadero de θ. Esto quiere decir que la distribución del estimador está centrada en el parámetro verdadero θ.

- Sesgo del estimador: b ($\widehat{\Theta}$) = E ($\widehat{\Theta}$) − θ
- Magnitud del sesgo: |b ($\widehat{\Theta}$)| = |E($\widehat{\Theta}$) – θ|

Notar que si un estimador es insesgado entonces su sesgo es cero

Error Cuadrático Medio

ECM ($\widehat{\Theta}$) = V($\widehat{\Theta}$) + [b($\widehat{\Theta}$)]^2

- El error cuadrático medio es un criterio importante para comparar estimadores.
- Si $\widehat{\Theta}_1$y $\widehat{\Theta}_2$ son dos estimadores para un parámetro θ, $\widehat{\Theta}_1$ será mejor estimador de θ si ECM $\widehat{\Theta}_1$< ECM $\widehat{\Theta}_2$

Consistencia de estimadores puntuales

Teorema: $\widehat{\Theta}_n$ es un estimador consistente de θ si:

$$\lim_{n\to\infty} E(\widehat{\Theta}_n) = \theta \text{ y } \lim_{n\to\infty} V(\widehat{\Theta}_n) = 0$$

TP N°6 Estimación Puntual

Ejercicio 1

1) Suponga que se tiene una muestra aleatoria de tamaño n tomada de una población X , que $E(X_i) = \mu$ y $V(X_i) = \sigma^2$. Sean

$$\overline{X}_1 = \frac{1}{1-n}\sum_{n=1}^{n-1} X_i \qquad \overline{X}_2 = \frac{1}{n}\sum_{n=1}^{n} X_i$$

Dos estimadores de μ .

a) ¿Cuál es el mejor estimador de μ? Explique su elección.

b) ¿Los estimadores son consistentes?

a) Para responder a la pregunta tenemos que calcular el ECM de cada estimador y el menor de ellos será el mejor estimador para μ .

$$E(\overline{X}_1) = E\left(\frac{1}{n-1}\sum_{i=1}^{n-1} X_i\right) \underset{\text{Linealidad}}{=} \frac{1}{n-1}E\left(\sum_{i=1}^{n-1} X_i\right) = \frac{1}{n-1}\sum_{i=1}^{n-1} E(X_i) = \frac{1}{n-1}(n-1)E(X_i) =$$

$$= \frac{1}{(n-1)}(n-1)\mu = \mu$$

$$E(\overline{X}_1) = \mu \rightarrow b(\overline{X}_1) = 0$$

$$V(\overline{X}_1) = V\left(\frac{1}{n-1}\sum_{i=1}^{n-1} X_i\right) \underset{\text{Prop. de varianza}}{=} \frac{1}{(n-1)^2}V\left(\sum_{i=1}^{n-1} X_i\right) \underset{\text{Independencia}}{=} \frac{1}{(n-1)^2}\sum_{i=1}^{n-1} V(X_i)$$

$$= \frac{1}{(n-1)^2}(n-1)V(X_i) = \frac{1}{(n-1)^2}(n-1)\sigma^2 = \frac{\sigma^2}{(n-1)}$$

Entonces:

$$ECM(\overline{X}_1) = V(\overline{X}_1) + [b(\overline{X}_1)]^2 = \frac{\sigma^2}{(n-1)}$$

$$E(\overline{X}_2) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}nE(X_i) = \frac{1}{n}n\mu = \mu$$

$$E(\overline{X}_2) = \mu \rightarrow b(\overline{X}_2) = 0$$

$$V(\overline{X}_2) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}nV(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

Entonces:

$$ECM(\overline{X}_2) = V(\overline{X}_2) + [b(\overline{X}_2)]^2 = \frac{\sigma^2}{n}$$

Conclusión:

Como $ECM(\overline{X}_2) < ECM(\overline{X}_1)$ entonces $\overline{X}_2$ es el mejor estimador de μ

b) Aplicando el teorema

$$\lim_{n\to\infty} E(\overline{X}_1) = \lim_{n\to\infty} \mu = \mu$$

$$\lim_{n\to\infty} V(\overline{X}_1) = \lim_{n\to\infty} = \frac{\sigma^2}{(n-1)} = \lim_{n\to\infty} \frac{\sigma^2}{n\underbrace{\left(1-\frac{1}{n}\right)}_{\to 1}} = \lim_{n\to\infty} = \frac{\sigma^2}{n} = 0$$

$\therefore \overline{X}_1$ es un estimador consistente para μ

$$\lim_{n\to\infty} E(\overline{X}_2) = \lim_{n\to\infty} \mu = \mu$$

$$\lim_{n\to\infty} V(\overline{X}_2) = \lim_{n\to\infty} = \frac{\sigma^2}{n} = 0$$

$\therefore \overline{X}_2$ es un estimador consistente para μ

TP N°6 Estimación Puntual

Ejercicio 3

Sea X_1, X_2, ..., X_n una muestra aleatoria de tamaño n.

a) Demuestre que $\overline{X}^2$ es un estimador sesgado de μ^2
b) Determine la magnitud del sesgo de este estimador
c) ¿Qué sucede con el sesgo a medida que aumenta el tamaño n de la muestra?

a) Para resolver el problema recordemos la siguiente expresión de la varianza:

$$V(X) = E(X^2) - [E(X)]^2$$

En nuestro caso:

$$V(\overline{X}) = E(\overline{X}^2) - [E(\overline{X})]^2$$

Despejando:

$$E(\overline{X}^2) = V(\overline{X}) + [E(\overline{X})]^2$$

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \underset{\text{Prop. de varianza}}{=} \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) \underset{\text{Independencia}}{=} \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2} nV(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \underset{\text{Linealidad}}{=} \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} nE(X_i) = \frac{1}{n} n\mu = \mu$$

$$E(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2$$

$\therefore \overline{X}^2$ es un estimador sesgado de μ^2 ya que $E(\overline{X}^2) \neq \mu^2$

b) La magnitud del sesgo será:

$$|b(\overline{X}^2)| = \left|\frac{\sigma^2}{n} + \mu^2 - \mu^2\right| = \frac{\sigma^2}{n}$$

c) El sesgo tiende a cero cuanto el tamaño de la muestra aumenta

$$\lim_{n\to\infty} b(\overline{X}^2) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

Métodos de Estimación Puntual

Sea X1,X2,…,Xn muestra aleatoria de una V.A X. Asumimos que se conoce la familia de distribuciones a la que pertenece X.

Sean θ1,…, θr R parámetros desconocidos de la distribución de X.

Definimos los ***momentos de orden k de una variable aleatoria*** como:

$$\mu_k = E(X^k) = \sum_{x_i \in R_X} x_i^k p(x_i) \qquad (k = 0,1,2,\ldots) \qquad \text{Si } X \text{ es discreta}$$

$$\mu_k = E(X^k) = \int^{+\infty} x^k f(x)dx \qquad (k = 0,1,2,\ldots) \qquad \text{Si } X \text{ es continua,}$$

Definimos : $E(X^k)$ es el k esimo momento poblacional ($k = 1, 2, 3, \ldots$) si $k=1$ tenemos $E(X)$

$\frac{1}{m}\sum_{i=1}^{m} x_i^k$ es el k-esimo momento muestral ($k=1,2,3,\ldots$)

si k es 1 tenemos $\overline{X}$

se puede probar que $P\left(\left|\frac{1}{m}\sum_{i=1} x_i^k - E(x^k)\right| > \varepsilon\right) \xrightarrow[m\to\infty]{} 0$

El método de los momentos, para estimar $\theta_1, \theta_2, \dots, \theta_r$ consiste en plantear:

$$\begin{cases} \frac{1}{m}\sum_{i=1}^{m} x_i = E(x) \\ \frac{1}{m}\sum_{i=1}^{m} x_i^2 = E(x^2) \\ \vdots \\ \frac{1}{m}\sum_{i=1}^{m} x_i^r = E(x^r) \end{cases}$$

más que una sistema de r ecuaciones con r incognitas se lo resuelve y se encuentran estimadores $\theta_1, \theta_2, \dots, \theta_r$ $(E(x), E(x^2), \dots, E(x^r))$ fueron expresados en función de $\theta_1, \theta_2, \dots, \theta_r$

Ej) sea $x_1, x_2, \dots, x_n$ muestra aleatoria de una v.a X donde $X \sim N(\mu, \sigma^2)$ hallar estimadores puntuales para μ y σ^2 por el método de los momentos.

$\theta_1 = \mu \quad \theta_2 = \sigma^2$ sabemos que $E(x) = \mu$ y $E(x^2) = V(x) + (E(x))^2 = \sigma^2 + \mu^2$

$\frac{1}{m}\sum_{i=1}^{m} x_i = E(x)$

$\frac{1}{m}\sum_{i=1}^{m} x_i^2 = E(x^2)$

$$\begin{cases} \frac{1}{m}\sum_{i=1}^{m} x_i = \mu \quad (1) \\ \frac{1}{m}\sum_{i=1}^{m} x_i^2 = \sigma^2 + \mu^2 \quad (2) \end{cases}$$

De (1) $\rightarrow$ $\hat{\mu} = \frac{1}{m}\sum_{i=1} x_i = \overline{X}$

De (2) $\rightarrow$ $\sigma^2 = \frac{1}{m}\sum_{i=1}^{m} x_i^2 - \mu^2 = \frac{1}{m}\sum_{i=1}^{m} x_i^2 - \overline{X}^2$

$$\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m} x_i^2 - \overline{X}^2$$

<u>Observación</u>

$$\widehat{\sigma^2} = \frac{1}{m}\sum_{i=1}^{m} x_i^2 - \overline{X}^2 = \frac{1}{m}\overbrace{\left(\sum_{i=1}^{m} x_i^2 - m\overline{x}^2\right)}^{\sum_{i=1}^{m}(x_i - \overline{x})^2}$$

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \bar{x})^2$$

Notar que

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \bar{x}^2) = \frac{m-1}{m} \underbrace{\left(\frac{1}{m-1} \sum_{i=1}^{m} (x_i - \bar{x})^2 \right)}_{S^2} = \frac{m-1}{m} S^2$$

$$E(\hat{\sigma}^2) = E\left(\left(\frac{n-1}{m}\right) S^2\right) = \left(\frac{m-1}{m}\right) E(S^2) = \left(\frac{m-1}{m}\right) \sigma^2 \neq \sigma^2$$

entonces el método de los momentos no produce en gral estimadores insesgados

Ej) sea $x_1, x_2, \dots, x_n$ muestra aleatoria de una v.a X, $X \sim P(\lambda)$

Hallar estimador puntual de λ por el método de los momentos

$\sigma_1 = \lambda$

$$\frac{1}{m} \sum_{i=1}^{m} x_i = E(X) \Rightarrow \frac{1}{m} \sum_{i=1}^{m} x_i = \lambda \qquad \boxed{\therefore \hat{\lambda} = \bar{X}}$$

Ej) sea $x_1, x_2, \dots, x_n$ muestra aleatoria de una v.a X, $X \sim N(0, \sigma^2)$

Hallar un estimador puntual de σ^2 por el método de los momentos

$$\frac{1}{m} \sum_{i=1}^{m} x_i = E(X) \Rightarrow \frac{1}{m} \sum_{i=1}^{m} x_i = 0$$

$$\frac{1}{m} \sum_{i=1}^{m} x_i^2 = \underbrace{E(X^2) = V(X)}_{} = \sigma^2 \qquad \therefore \boxed{\hat{\theta}_2 = \frac{1}{m} \sum_{i=1}^{m} X_i^2}$$

$$\boxed{\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ V(X) &= E(X^2) - 0 \end{aligned}}$$

Ej) sea $x_1, x_2, \dots, x_n$ muestra aleatoria de una v.a X, donde la densidad de X está dada por

$$f(x) = \begin{cases} \frac{\theta}{2^\theta} x^{\theta - 1} & 0 < x < 2 \\ 0 & c.c \end{cases}$$

Hallar un estimador puntual de θ por el método de los momentos

Observación: supongamos que para hallar θ planteamos $\int_{-\infty}^{\infty} f(x)\, dx = 1$

$$1 = \int_0^2 \frac{\theta}{2^\theta} x^{\theta-1}\, dx = \frac{\theta}{2^\theta} \int_0^2 x^{\theta-1}\, dx =$$

$$= \frac{\theta}{2^\theta} \left(\frac{x^\theta}{\theta} \right) \Big|_0^2 = \frac{\theta}{2^\theta} \frac{2^\theta}{\theta^2} = 1$$

se me va $\theta \Rightarrow$ no puedo saber su valor

Planteamos: $\frac{1}{m}\sum_{i=1}^{m} x_i = E(x)$

$E(x) = \int_0^2 x\left(\frac{\theta}{\theta+1}x^{\theta-1}\right)dx = \frac{\theta}{\theta+1}\int_0^2 x^{\theta}\,dx = \frac{\theta}{\theta+1}\left(\frac{x^{\theta+1}}{\theta+1}\right)\Big|_0^2 = \frac{\theta}{2^{\theta}}\left(\frac{2^{\theta+1}}{\theta+1}\right) = \boxed{\frac{2\theta}{\theta+1}}$

$\therefore \frac{1}{m}\sum_{i=1}^{m} x_i = \frac{2\theta}{\theta+1}$ resuelvo para encontrar θ

$(\theta+1)\bar{x} = 2\theta \Rightarrow \theta\bar{x} + \bar{x} = 2\theta$

$\Rightarrow \bar{x} = 2\theta - \theta\bar{x} \Rightarrow \bar{x} = \theta(2-\bar{x}) \Rightarrow \hat{\theta} = \frac{\bar{x}}{2-\bar{x}}$

Método de Máxima Verosimilitud

Sea X1, X2, …, Xn muestra aleatoria de una v.a X asumimos conocida la familia de distribución a la que pertenece X. Sea θ un parámetro desconocido de la distribución de X.

Supongamos que x1, x2, …, xn es una observación de la muestra aleatoria.

Planteamos:

1) $\mathrm{L}(x_1, x_2, \dots, x_n\,;\theta) = \mathrm{P}(\mathrm{X} = \mathrm{x}_1)\cdot\mathrm{P}(\mathrm{X} = \mathrm{x}_2)\cdots\mathrm{P}(\mathrm{X} = \mathrm{x}_n)$ Si X es una v.a discreta.

$\mathrm{L}(x_1, x_2, \dots, x_n\,;\theta) = f(x_1)\cdot f(x_2)\cdots f(x_n)$ Si X es una v.a continua.

Hallar aquel valor de θ que maximiza lo anterior. Para ello:

2) $\ln[\mathrm{L}(x_1, x_2, \dots, x_n\,;\theta)]$

3) $\frac{\mathrm{d}\ln[\mathrm{L}(x_1,x_2,\dots,x_n\,;\theta)]}{\mathrm{d}\theta} = 0 \rightarrow \theta \rightarrow \hat{\theta}$

Propiedades importantes:

$\ln(a\cdot b) = \ln(a) + \ln(b)$

$\ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b)$

$\ln(a^b) = b\cdot\ln(a)$

$\ln\left(\prod x_i\right) = \sum \ln(x_i)$

$\ln(e) = 1$

$\ln(1) = 0$

Sea $x_1, x_2, \dots, x_n$ muestra aleatoria de una v.a

$X, \quad X \sim B(1,p)$

supongamos que queremos la proporción de defectuosos fabricados por una máquina en un día. tomamos una muestra de m artículos

Para cada artículo se define

$$X_i \begin{cases} 1 & \text{si el artículo es defectuoso} \\ 0 & \text{c.c} \end{cases}$$

$i = 1, 2, \dots, n$

Nota que $X_i \sim B(1,p)$ hallar el emv de p

$$P(X = x_i) = \binom{1}{x_i} p^{x_i} (1-p)^{1-x_i} = p^{x_i} (1-p)^{1-x_i}$$

Planteamos

$$L(x_1, x_2, \dots, x_n, p) = P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdots P(X_n = x_n)$$

$$= p^{x_1}(1-p) \cdots p^{x_n}(1-p) = p^{\sum_{i=1}^{n} x_i} (1-p)^{\sum_{i=1}^{n}(1-x_i)} = p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i}$$

$$\ln L(x_1, x_2, \dots, x_n, p) = \ln\left(p^{\sum x_i}\right) + \ln\left((1-p)^{n \sum x_i}\right) =$$

$$\left(\sum x_i\right) \ln p + \left(m - \sum x_i\right) \ln(1-p)$$

$$\frac{d}{dp} \ln L(x_1, x_2, \dots, x_n, p) = 0 \Rightarrow$$

$$\Rightarrow \left(\sum_{i=1}^{m} x_i\right) \frac{1}{p} + \left(m - \sum_{i=1}^{n} x_i\right) \frac{1}{1-p} (-1) = 0$$

$$\left(\sum_{i=1}^{n} x_i\right) \frac{1}{p} = \left(n - \sum_{i=1}^{n} x_i\right) \frac{1}{1-p} \Rightarrow$$

$$\Rightarrow \frac{1-p}{p} = \frac{m - \sum_{i=1}^{m} x_i}{\sum_{i=1}^{n} x_i} \Rightarrow \frac{1}{p} - 1 = \frac{m}{\sum x_i} - 1 \Rightarrow p = \frac{\sum_{i=1}^{n} x_i}{m} = \bar{X}$$

Ej 5) Si se seleccionan 100 chip fabricados por una compañía y se encuentra que 5 son defectuosos y $\sum_{i=1}^{n} x_i = 5$

La estimación de p es $\bar{x} = \frac{1}{100} \sum_{i=1}^{100} x_i = \frac{5}{100} = 0{,}05$

Propiedades de los EMV

1) Los EMV pueden ser **sesgados**, pero por lo general si $\hat{\theta}$ de un parámetro θ, basado en una muestra aleatoria de tamaño n, entonces $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$, es decir son **asintóticamente insesgado.**
2) Bajo condiciones bastantes generales se puede probar que los EMV son consistentes.
3) Bajo condiciones bastantes generales se puede probar que los EMV **asintóticamente tienen varianza mínima**
4) Los EMV cumplen la **propiedad de invarianza**, es decir:

Si $\hat{\theta}$ es un EMV de un parámetro θ , el EMV es $g(\theta)$ es $g(\hat{\theta})$, si $g(x)$ es una función inyectiva

Propiedades de los EMV, validos para muestras grandes bajo condiciones muy generales se puede probar que:

1) los EMV son asintoticamente insesgados
2) los EMV son consistentes
3) son asintoticamente los de menor varianza
4) Propiedad de invarianza: supongamos que $\hat{\theta}$ es el estimador de θ sea $g(x)$ una función inyectiva. Encontrar EMV de $g(\theta) \to$ es $g(\hat{\theta})$

en el ejemplo anterior se encontra el EMV de σ^2 para la $N(\mu, \sigma^2)$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

si queremos hallar EMV de σ, usamos la prop de invarianza

$\sigma = \sqrt{\sigma^2}$ $\quad \therefore \hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

$g(x) = \sqrt{x}$

5.62) En el ejercicio pide EMV $(1-p)^6$

$n = 100$ $\quad \sum_{i=1}^{100} x_i = 5$ $\quad \hat{p} = \bar{x}$ la estimación de p para la muestra dada $\hat{p} = 0{,}05$

$g(p) = (1-p)^6$ (aca se toman 6 chips y ninguno es defectuoso)

el EMV de $(1-p)^6$ sera $g(\hat{p}) = g(\bar{x}) = (1-\bar{x})^6$ y la estimación para esta prob sera $(1-0{,}05)^6$

pide esta prop en 4 y 7 $\quad (1-p)^6$

en el ejemplo de las lamparitas

Y: "duración en hs de una lamparita tomada al azar"

$x_1, x_2, \dots, x_n$ el EMV de λ $\quad \hat{\lambda} = \frac{1}{\bar{X}}$

Hallar EMV de:

$P(X < 1400) = 1 - e^{-\lambda 1400}$

el EMV será

$\widehat{1 - e^{-\lambda 1400}} = 1 - e^{-\hat{\lambda} 1400} =$

reemplazo λ que encuentre $\leftarrow$ =

TP N°6 Estimación Puntual

Ejercicio 4

El número diario de desconexiones accidentales de un servidor sigue una distribución de Poisson. En cinco días se observan: 2 , 5 , 3 , 3 , 7 desconexiones accidentales.

a) Obtenga el estimador de máxima verosimilitud de λ.¿El estimador es insesgado? ¿Es consistente?.

b) Obtenga la estimación de λ a partir de la muestra dada.

c) Encuentre el estimador de máxima verosimilitud de que ocurrirán 3 o más desconexiones accidentales y encuentre la estimación de dicha probabilidad a partir de la muestra dada.

a) Estimador de Máxima Verosimilitud (EMV)

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad , R_x = \{0,1,2 \dots\}$$

1) $L(x_1, x_2, \dots, x_n, \lambda) \underset{\text{Independencia}}{=} P(X = x_1).P(X = x_1) \dots P(X = x_n) = \left(\frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\right).\left(\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\right) \dots \left(\frac{e^{-\lambda}\lambda^{x_n}}{x_n!}\right) = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod x_i!}$

2) $\ln[L(x_1, x_2, \dots, x_n, \lambda)] = \ln\left(\frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod x_i!}\right) = \ln\left(e^{-n\lambda}\lambda^{\sum x_i}\right) - \ln(\prod x_i!) = \ln\left(e^{-n\lambda}\right) + \ln\left(\lambda^{\sum x_i}\right) - \ln(\prod x_i!) =$

$$-n\lambda + \sum x_i\,(\ln\lambda) - \ln\left(\prod x_i!\right)$$

3) $\frac{d[\ln L(x_1, x_2, \dots, x_n, \lambda)]}{d\lambda} = -n + \frac{\sum x_i}{\lambda} = 0 \rightarrow \lambda = \frac{\sum x_i}{n} = \bar{X} \rightarrow \hat{\lambda} = \frac{\sum x_i}{n} = \bar{X}$

$$E(\overline{X}) = E\left(\frac{\sum X_i}{n}\right) = \frac{1}{n}E\left(\sum X_i\right) = \frac{1}{n}\sum E(X_i) = \frac{1}{n}nE(X_i) = E(X_i) = \lambda \rightarrow \hat{\lambda} \text{ es insesgado}$$

Consistencia:

$$\lim_{n\to\infty} E(\hat{\lambda}) = \lim_{n\to\infty} \lambda = \lambda$$

$$\lim_{n\to\infty} V(\hat{\lambda}) = \lim_{n\to\infty} \frac{\lambda}{n} = 0$$

b) La estimación a partir de la muestra dada será:

$$\hat{\lambda} = \overline{X} = \frac{2+5+3+3+7}{5} = 4$$

c) Estimador de $P(X \geq 3)$:

$$\widehat{P(X \geq 3)} = 1 - \widehat{P(X \leq 2)} = 1 - \widehat{\left[\frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^1}{1!} + \frac{e^{-\lambda}\lambda^2}{2!}\right]} = 1 - \left[\frac{e^{-\overline{x}}\overline{x}^0}{0!} + \frac{e^{-\overline{x}}\overline{x}^1}{1!} + \frac{e^{-\overline{x}}\overline{x}^2}{2!}\right]$$

Prop. Invarianza

Estimación $P(X \geq 3)$:

$$\widehat{P(X \geq 3)} = 1 - \left[\frac{e^{-4}4^0}{0!} + \frac{e^{-4}4^1}{1!} + \frac{e^{-4}4^2}{2!}\right] = 0.761896694$$

TP N°6 Estimación Puntual

Ejemplo 1

Hallar el estimador por el método de los momentos del parámetro p de una v.a. que sigue una distribución geométrica.

Resolucion:

$X \sim G(p) \qquad f(x) = (1-p)^{x-1}p$

Como p es el único parámetro que hay que estimar, planteamos una sola ecuación de momento:

(1) $E(X) = \overline{X}$

Recordando que la esperanza de una distribución geométrica es:

$$E(X) = \frac{1}{p}$$

Reemplazando la expresión anterior en la ecuación (1):

$$\frac{1}{p} = \overline{X}$$

Despejando p:

$$p = \frac{1}{\overline{X}}$$

Finalmente el estimador por el método de los del parámetro p es:

$$\hat{\mathbf{p}} = \frac{\mathbf{1}}{\overline{\mathbf{X}}}$$

TP N°6 Estimación Puntual
Ejemplo 2

Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una v.a X con densidad dada por:

$$f(x_i) = \begin{cases} \frac{1}{\theta} e^{-\frac{x_i}{\theta}} \text{ si } x > 0 \\ 0 \qquad\qquad \text{cc} \end{cases}$$

Hallar el EMV de θ

Resolución

$X_i \sim \text{Exp}\left({}^{1}/_{\theta}\right)$

1) $L(x_1, x_2, \dots, \theta) = f(x_1) \cdot f(x_2) \cdots f(x_n) = \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \cdot \frac{1}{\theta} e^{-\frac{x_n}{\theta}} \cdots \frac{1}{\theta} e^{-\frac{x_n}{\theta}} = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum x_i}$

Independientes

TP N°6 Estimación Puntual
Ejemplo 2

2) $\ln(L) = -\frac{1}{\theta}\sum x_1 - n\ln(\theta)$

3) $\frac{d}{d\theta}\ln(L) = \frac{1}{\theta^2}\sum x_i - \frac{n}{\theta} = 0$

Despejando obtenemos el EMV del parámetro θ:

$$\widehat{\Theta} = \frac{\sum x_i}{n} = \overline{X}$$

TP N°6 Estimación Puntual
Ejemplo 3

Obtener el estimador μ y σ^2 por el método de los momentos de la distribución Normal

Resolución

Como hay que estimar dos parámetros, planteamos las dos primeras ecuaciones de momento:

$(1)\ E(X) = \overline{X}$

$(2)\ E(X^2) = \frac{1}{n}\sum X_i^2$

Recordando que:

$E(X) = \mu$

$V(X) = E(X^2) - [E(X)]^2 \rightarrow E(X^2) = V(X) + [E(X)]^2 = \sigma^2 + \mu^2$

Reemplazando las expresiones anteriores en la ecuación (1) y (2):

$(1)\ \mu = \overline{X}$

$(2)\ \sigma^2 + \mu^2 = \frac{1}{n}\sum X_i^2$

TP N°6 Estimación Puntual
Ejemplo 3

De la ecuación (1) directamente tenemos el estimador de μ:

$\hat{\mu} = \overline{X}$

Reemplazando la ecuación (1) en (2) :

$$\sigma^2 + \overline{X}^2 = \frac{1}{n}\sum X_i^2$$

Despejando obtenemos estimador de σ^2 :

$$\hat{\sigma}^2 = \frac{1}{n}\sum X_i^2 - \overline{X}^2$$

PRACTICA 7

Intervalo de confianza para la media de una distribución normal, varianza conocida

Estadístico Pivote:

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Si $(X_1, X_2, ..., X_n)$ una muestra aleatoria de tamaño n de una v.a. X donde $X \sim N(\mu, \sigma^2)$, σ^2 conocido, un intervalo de confianza para μ de nivel $1-\alpha$ es

$$\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \qquad (8.1)$$

Ejemplo

$z_{\frac{\alpha}{2}}$ es el valor crítico.

$P(x > x) \rightarrow$ si pongo $\frac{\alpha}{2}$ me da $z_{\frac{\alpha}{2}}$

Uso tabla valores críticos o app

acá estadístico pivote del que se parte y distribución

ej) Se sabe que el límite de elasticidad de un tipo particular de barra de acero está normalmente distribuido con $\sigma = 100$. Se toman 25 barras a cada una se le mide el límite de elasticidad y resulta un promedio muestral de 8.439 lb. Calcule un intervalo de confianza de nivel 90% para la verdadera media del límite de elasticidad de la barra de acero

- X: "límite de elasticidad (en lb) de una barra de acero"

$X \sim N(\mu, \sigma^2) \quad \sigma = 100$

X_i: "límite de elasticidad de barra i" $\quad i = 1, 2, ..., 25$

$X_1, X_2, ..., X_{25}$ muestra aleatoria de la v.a. X

$\overline{x} = 8439$ lb $\qquad 1 - \alpha = 0{,}90$

$\left(\overline{x} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}, \overline{x} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}\right) \qquad \alpha = 0{,}10$

$z_{\frac{\alpha}{2}} = z_{0{,}05} = \boxed{1{,}65} \qquad \frac{\alpha}{2} = \boxed{0{,}05}$

$\left(8439 - 1{,}65 \sqrt{\frac{100^2}{25}}, 8439 + 1{,}65 \sqrt{\frac{100^2}{25}}\right) = (8406 ; 8472)$

$L = 8472 - 8406 = 66$

Si tomamos $1 - \alpha = 0{,}92$ si aumenta el nivel de confianza aumenta long.

$\left(8439 \quad 1{,}76 \sqrt{\frac{100^2}{25}}, 8439 + 1{,}76 \sqrt{\frac{100^2}{25}}\right) = (8403{,}8 ; 8474{,}2)$

$L = 70{,}4$ pierdo precisión en estimación

Observaciones

1) $\left(\bar{x} - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}},\ \bar{x} + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right)$ está centrado en $\bar{x}$

2) $L = \bar{x} + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} - \left(\bar{x} - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 2z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}$

Notar que: si dejamos n fijo y aumenta $1-\alpha$ (o disminuye α) entonces $z_{\frac{\alpha}{2}}$ aumenta $\therefore$ L aumenta.

Si $1-\alpha$ se deja fijo y aumentamos n, entonces L disminuye

Se puede plantear: que tamaño n de muestra se necesita para que $L \leq \ell$. ℓ se fija de antemano

+ precisión de estimación

$$L = 2z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq \ell \Rightarrow \boxed{\left(\frac{2z_{\frac{\alpha}{2}}\,\sigma}{\ell}\right)^2 \leq n}$$

Por qué tamaño de muestra se necesita para que el intervalo de nivel 90% se reduzca a la mitad?

$$L = 2 \cdot 1{,}65\sqrt{\frac{100^2}{n}} \leq 33\,\ell$$

$$\underbrace{\left(\frac{2 \cdot 1{,}65 \cdot 100}{33}\right)^2}_{100} \leq n$$

redondeo para arriba

$\therefore n \geq 100$

Error Máximo → L/2

Intervalo de confianza para la media con distribución de desconocida, varianza conocida

Para muestras tomadas de una población normal, o para muestras de tamaño $n \geq 30$, de una población cualquiera, el intervalo de confianza dado anteriormente en (8.1), proporciona buenos resultados.
En el caso de que la población de la que se extrae la muestra no sea normal pero $n \geq 30$, el nivel de confianza del intervalo (8.1) es ***aproximadamente*** $1-\alpha$.
Pero para muestras pequeñas tomadas de poblaciones que no son normales no se puede garantizar que el nivel de confianza sea $1-\alpha$ si se utiliza (8.1).

... que a) debe contener a μ b) debe tener dist. conocida

Caso 1° se conoce σ^2 tomamos el estadístico

Hallamos a tal que $1-\alpha =$ $\quad Z = \dfrac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}$ Notar que $Z \approx N(0,1)$ por T.C.L

$= P(-a \leq Z \leq a) \underset{\text{T.C.L}}{\approx} \Phi(a) - \Phi(-a) =$

$= 2\Phi(a) - 1 \quad \therefore \; a = z_{\frac{\alpha}{2}}$

con un planteo análogo al caso anterior se llega al intervalo

$\left(\bar{x} - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \,;\; \bar{x} + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right)$ Notar que el nivel de confianza es aproximado. $(1-\alpha).100\%$

estadístico tiene dist. aprox a Normal

Escaneado con CamScanner

Intervalo de confianza para la media de una distribución normal, varianza desconocida

Estadístico Pivote:

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}.$$

Se puede probar que la distribución de T es una distribución llamada Student con parámetro n-1

Si $(X_1, X_2, \ldots, X_n)$ una muestra aleatoria de tamaño n de una v.a. X donde $X \sim N(\mu, \sigma^2)$, σ^2 desconocido, un intervalo de confianza para μ de nivel $1-\alpha$ es

$$\left[\overline{X} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right] \qquad (8.2)$$

Ej) un fabricante de polvora desarrollo una nueva formula que se probo con 8 granadas las velocidades iniciales resultantes en pies x seg. fueron:

3005 ; 2009,25 ; 2009,35 ; 2009,65 ; 2009,95 ; 3005 ; 2009,37 ; 2905

Hallar un intervalo de confianza para la media real de las velocidades iniciales para granadas de este tipo de nivel 95%. Suponer que las velocidades iniciales tienen dist. normal

X: "velocidad inicial de una granada tomada al azar con polvora nueva"

$X \sim N(\mu, \sigma^2)$

X_i: "velocidad inicial de granada i" $i = 1, 2, \dots, 8$

$X_1, X_2, \dots, X_8$ muestra aleatoria de va X

$1 - \alpha = 0,95$

$$\left(\bar{X} - t_{\frac{\alpha}{2}, n-1} \sqrt{\frac{S^2}{n}} \; ; \; \bar{X} + t_{\frac{\alpha}{2}, n-1} \sqrt{\frac{S^2}{n}} \right)$$

$$\bar{x} = \frac{1}{8} \sum_{i=1}^{8} x_i = 2959$$

$$S = \sqrt{\frac{1}{8-1} \sum_{i=1}^{8} (x_i - \bar{x})^2} = 39,1$$

$1 - \alpha = 0,95$ $\quad \alpha = 0,05$ $\quad \frac{\alpha}{2} = 0,025$ $\quad t_{0,025,7} = 2,365$

$$\left(2959 - 2{,}365\left(\frac{39{,}1}{\sqrt{8}}\right) ; 2959 + 2{,}365\left(\frac{39{,}1}{\sqrt{8}}\right)\right) =$$

$$= \left(2926{,}3 ; 2991{,}7\right)$$

Observaciones

1) en el caso del intervalo $\left(\bar{X} - t_{\frac{\alpha}{2},n-1}\sqrt{\frac{S^2}{n}} ; \bar{X} + t_{\frac{\alpha}{2},n-1}\sqrt{\frac{S^2}{n}}\right)$

$L = 2\, t_{\frac{\alpha}{2},n-1}\sqrt{\frac{S^2}{n}}$ no se puede despejar

2) en el caso del intervalo $\left(\bar{X} - z_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}} ; \bar{X} + z_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}}\right)$

$L = 2\, z_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}}$ En este caso se puede plantear hallar n tal que $L \leq \ell$ y se despeja n

$$\left(\frac{2\, z_{\frac{\alpha}{2}} \cdot S}{\ell}\right)^2 \leq n$$

Intervalo de confianza para la media con distribución de desconocida, varianza desconocida

Si σ^2 es desconocido y el tamaño de la muestra grande, entonces se puede probar que al reemplazar σ por S, el estadístico

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0,1) \quad \textbf{\textit{aproximadamente}}$$

y puedo construir el intervalo para μ como antes:

$\left[\bar{X} - z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$, pero su nivel es aproximadamente $1-\alpha$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Caso 2°: NO SE CONOCE σ^2

con un planteo análogo al caso 1 se llega al siguiente intervalo

$$\left(\bar{X} - z_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} \, , \, \bar{X} + z_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} \right)$$

con un nivel de $(1-\alpha)\,100\%$ aproximado.

$$z = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}}$$

se puede probar que $z \approx N(0,1)$ por T.C.L y porque S^2 es un estimador consistente de σ^2

Ej) se registraron los tiempos utilizados en la compra para 64 clientes seleccionados al azar en un supermercado y la varianza de los 64 tiempos de compra fueron 33 minutos y 256 minutos respectivamente. estimar la media real μ del tiempo utilizado por los clientes en la compra con un intervalo de confianza del nivel 90%.

media de muestra

varianza de muestra

X: "tiempo de compra en minutos de un cliente tomado al azar en un supermercado"

X_i: "tiempo de compra del cliente i" $i = 1, 2, \dots, 64$

$X_1, X_2, \dots, X_{64}$ muestra aleatoria de la v.a X

$$\bar{x} = \frac{1}{64} \sum_{i=1}^{64} x_i = 33 \qquad S^2 = \frac{1}{63} \sum_{i=1}^{64} (x_i - \bar{x})^2 = 256$$

$1 - \alpha = 0{,}90$

$\alpha = 0{,}10$

$\frac{\alpha}{2} = 0{,}05$

$$\left(\bar{x} - z_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} \, , \, \bar{x} + z_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} \right)$$

$z_{\frac{\alpha}{2}} = z_{0{,}05} = 1{,}645$

$$\left(33 - 1{,}645 \sqrt{\frac{256}{64}} \, , \, 33 + 1{,}645 \sqrt{\frac{256}{64}} \right) =$$

$= (29{,}71 \, ; \, 36{,}29)$ nivel de confianza 90% aprox.

Intervalo de confianza para la diferencia de dos medias con distribución normal, varianzas conocidas

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1), \text{ es decir, tiene distribución normal estandarizada.}$$

Si X_1 y X_2 son dos variables aleatorias ***independientes*** normalmente distribuidas: $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ y suponemos que las varianzas σ_1^2 y σ_2^2 son conocidas. Un intervalo de confianza para la diferencia $\mu_1 - \mu_2$ de nivel $1-\alpha$ es

$$\left[\bar{X}_1 - \bar{X}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\quad \bar{X}_1 - \bar{X}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] \qquad (8.3)$$

Ejemplo) X_{1i} = "Volumen de llenado de botella i llenada con maquina 1"

X_{2i} = "Volumen de llenado de botella i llenada con maquina 2"

$X_{11}, X_{12}, \ldots, X_{1n_1}$ muestra aleatoria X_1

$n_1 = 12$ $\qquad \bar{x}_1 = 30,87$

$X_{21}, X_{22}, \ldots, X_{2n_2}$ muestra aleatoria X_2

$n_2 = 10$ $\qquad \bar{x}_2 = 30,68$

$X_1 \sim N(\mu_1, 0,10^2)$ $\qquad X_2 \sim N(\mu_2, 0,15^2)$

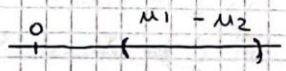
X_1 y X_2 son independientes

$\mu_1 - \mu_2$ de nivel 90%.

$1-\alpha = 0,90$ $\qquad \alpha = 0,10$ $\qquad \frac{\alpha}{2} = 0,05$ $\qquad z_{\frac{\alpha}{2}} = z_{0,05} = 1,65$

$$\left[30,87 - 30,68 - 1,65\sqrt{\frac{0,10^2}{12} + \frac{0,15^2}{10}},\ 30,87 - 30,68 + 1,65\sqrt{\frac{0,10^2}{12} + \frac{0,15^2}{10}}\right] =$$

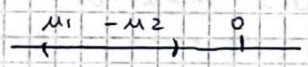
$$= [0,09837 ; 0,28620]$$



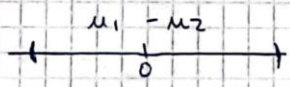
$\mu_1 - \mu_2 > 0$ con 90% de confianza

$\therefore$ $\mu_1 > \mu_2$ con 90% de confianza (porque el intervalo es positivo)

si quiero $\mu_1 - \mu_2 < 0$ $\therefore$ $\mu_1 < \mu_2$ con 90% de confianza



si el 0 está dentro no podemos decir nada



si pide que $L \leq l$ obtener el n de esto (hay que suponer que $n_1 = n_2 = n$)

$$\boxed{n \geq \left(\frac{2 z_{\frac{\alpha}{2}}}{l}\right)^2 (\sigma_1^2 + \sigma_2^2)}$$

Intervalo de confianza para la diferencia de dos medias con distribución desconocida (muestras grandes), varianzas desconocidas

Si σ_1 y σ_2 son ***desconocidos***, $n_1 \geq 30$ y $n_2 \geq 30$, entonces se puede probar que al reemplazar σ_1 por S_1 y σ_2 por S_2, el estadístico

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx N(0,1) \quad \textit{aproximadamente}$$

y puedo construir el intervalo para $\mu_1 - \mu_2$ como antes:

$$\left[\overline{X}_1 - \overline{X}_2 - z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \;,\; \overline{X}_1 - \overline{X}_2 + z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right], \qquad (8.4)$$

pero su nivel es aproximadamente $1-\alpha$

Intervalo de confianza para la diferencia de dos medias con distribución normal, varianzas desconocidas e iguales

Estadístico Pivote:

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{tiene distribución Student con } n_1 + n_2 - 2 \text{ grados de libertad}$$

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

Si X_1 y X_2 son dos variables aleatorias ***independientes*** normalmente distribuidas: $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$ y suponemos que las varianzas σ_1^2 y σ_2^2 son ***desconocidas e iguales***, es decir $\sigma_1 = \sigma_2 = \sigma$

Un intervalo de confianza para la diferencia $\mu_1 - \mu_2$ de nivel $1-\alpha$ es

$$\overline{X}_1 - \overline{X}_2 - t_{\alpha/2, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \;;\; \overline{X}_1 - \overline{X}_2 + t_{\alpha/2, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad (8.5)$$

Intervalo de confianza para la diferencia de dos medias con distribución normal, varianzas desconocidas y distintas

Estadístico Pivote:

$$T^* = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Se puede probar que T* tiene aproximadamente una distribución Student con v grados de libertad donde

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

si v no es entero, ***se toma el entero más próximo a*** v

Si X_1 y X_2 son dos variables aleatorias ***independientes*** normalmente distribuidas: $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ y suponemos que las varianzas σ_1^2 y σ_2^2 son ***desconocidas*** y ***distintas***

Un intervalo de confianza para la diferencia $\mu_1 - \mu_2$ de nivel ***aproximadamente*** $1-\alpha$ es

$$\left[\overline{X}_1 - \overline{X}_2 - t_{\frac{\alpha}{2},v}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},\quad \overline{X}_1 - \overline{X}_2 + t_{\frac{\alpha}{2},v}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right] \qquad (8.6)$$

Donde

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

Intervalo de confianza para muestras pareadas

Ahora X1i, X2i son dos mediciones diferentes sobre la misma unidad muestral i. Osea ahora X1, y X2 no son independientes.

Consideramos una tercera muestra:

$D_1 = X_{11} - X_{21}$, $D_2 = X_{12} - X_{22}$, ..., $D_n = X_{1n} - X_{2n}$

Entonces $D_1, D_2, \ldots, D_n$ es una muestra aleatoria de la va $D = X_1 - X_2$

Notar que $E(D) = E(X_1 - X_2) = E(X_1) - E(X_2) = \mu_1 - \mu_2$

$\sigma_D^2 = V(D) = V(X_1 - X_2) = \sigma_1^2 + \sigma_2^2 - 2\,cov(X_1, X_2)$

σ_D^2 es desconocida

entonces que decir que dif es normal

Si la variable $D = X_1 - X_2 \sim N(\mu_D, \sigma_D^2)$ entonces tomamos el estadístico pivote

$$T = \frac{\overline{D} - \mu_D}{\sqrt{\frac{S_D^2}{n}}} \sim t_{n-1}$$

se llega al intervalo:

$$\left(\bar{D} - t_{\frac{\alpha}{2},n-1}\sqrt{\frac{S_D^2}{n}} \; , \; \bar{D} + t_{\frac{\alpha}{2},n-1}\sqrt{\frac{S_D^2}{n}}\right)$$

Ejemplo) Las siguientes son las pérdidas semanales prom. de horas hombre debidas a accidentes en 10 plantas industriales antes y después de poner en operación cierto programa de seguridad

	1	2	3	4	5	6	7	8	9	10
antes	45	73	46	124	33	57	83	34	26	17
después	36	60	44	119	35	51	77	29	24	11

Hallar un intervalo de confianza de nivel 95% para $\mu_1 - \mu_2$ asumiendo que la diferencia de los datos es normal. ¿Se puede decir que el programa es efectivo?

X_{1i} = "pérdida de hs hombre antes del programa de seguridad en planta i" $i = 1, 2, \ldots, 10$

$X_{11}, X_{12}, \ldots X_{1n}$ muestra aleatoria de la X_1

X_{2i} = "pérdida de hs hombre después del programa de seguridad en planta i" $i = 1, 2, \ldots, 10$

$X_{21}, X_{22}, \ldots X_{2n}$ = muestra aleatoria de la X_2

$D = X_1 - X_2 \sim N(\mu_D, \sigma_D^2)$

$\mu_D = \mu_1 - \mu_2$

$1 - \alpha = 0{,}95 \qquad \alpha = 0{,}05 \qquad \alpha = 0{,}025 \qquad t_{0{,}9,9} = 2{,}262$

D: 9 13 2 5 −2 6 6 5 2 6

↓

tendría que ir abajo del cuadro → la tabla dada

$\bar{D} = 5{,}2 \qquad S_D = 4{,}08$

$$\left[5{,}2 - 2{,}262\left(\frac{4{,}08}{\sqrt{10}}\right) ; 5{,}2 + 2{,}262\left(\frac{4{,}08}{\sqrt{10}}\right)\right] =$$

$= (2{,}2815 \; ; \; 8{,}1184)$ $\qquad \mu_1 - \mu_2 > 0$ con una confianza de 95%.

∴ el programa es efectivo con una confianza de 95%.

Intervalo de confianza para la varianza de una distribución normal

Estadístico Pivote:

$$X = \frac{(n-1)S^2}{\sigma^2}$$

Se puede probar que X tiene una distribución llamada ji-cuadrado con n-1 grados de libertad

Si $(X_1, X_2, \ldots, X_n)$ es una muestra aleatoria de una v.a. X, donde $X \sim N(\mu, \sigma^2)$, un intervalo de confianza para σ^2 de nivel $1-\alpha$ es

$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} ; \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} \right) \qquad (8.8)$$

<u>Observación</u>: un intervalo de confianza para σ de nivel $1-\alpha$, es $\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}} ; \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}} \right)$

Ejemplo) aplicar

X: "volumen de elevado en ingresos de una botella de determinado consumo al azar"

$X \sim N(\mu, \sigma^2)$

X_i: "volumen de elevado de la botella i"

$i = 1, 2, \ldots, 20$

$x_1, x_2, \ldots, x_{20}$ muestra aleatoria de $v.a.$ X

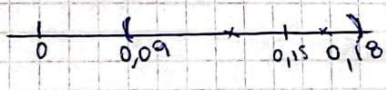
$S^2 = 0{,}0153$ $\qquad 1-\alpha = 0{,}95$ $\qquad \frac{\alpha}{2} = 0{,}025$

$b \rightarrow \chi^2_{\frac{\alpha}{2},19} = \chi^2_{0{,}025,19} = 32{,}9$

$a \rightarrow \chi^2_{1-\frac{\alpha}{2},19} = \chi^2_{0{,}975,19} = 8{,}91$

$$\frac{(20-1)\cdot}{32{,}9} \; ; \; \frac{(20-1)\cdot 0{,}0153}{8{,}91} = (0{,}00884 \; , \; 0{,}0326)$$

el intervalo para σ $\left(\sqrt{0{,}00884} \; ; \; \sqrt{0{,}0326}\right) = (0{,}09 \; ; \; 0{,}180)$



no podemos decir si el condicionante se queda triangulo

no se puede afirmar que $\sigma < 0{,}15$

Intervalo de confianza para el cociente de varianzas de dos distribuciones normales

Estadístico Pivote:

$$F = \frac{S_2^2 \times \sigma_1^2}{S_1^2 \times \sigma_2^2}$$

Se puede probar que F tiene una distribución llamada Fisher con n2-1 y n1-1 grados de libertad.

Si se tienen dos poblaciones normales e independientes con varianzas desconocidas σ_1^2 y σ_2^2 respectivamente, entonces un intervalo de nivel $1-\alpha$ para el cociente de las dos varianzas $\frac{\sigma_1^2}{\sigma_2^2}$ es

$$\left[\frac{S_1^2}{S_2^2} f_{1-\frac{\alpha}{2},n_2-1,n_1-1} \; ; \; \frac{S_1^2}{S_2^2} f_{\frac{\alpha}{2},n_2-1,n_1-1}\right] \qquad (8.9)$$

Si el 1 no esta en el intervalo, y $1 < \sigma_1^2 / \sigma_2^2$ $\sigma_2^2 < \sigma_1^2$

Intervalo de confianza para una proporción

Intervalo para una proporción

Se toma una muestra de n artículos

$$X_i = \begin{cases} 1 & \text{si el artículo } i \text{ es defectuoso} \\ 0 & \text{cc} \end{cases}$$

$X_1, X_2, \ldots, X_n$ ma de X

$X \sim B(1, p)$ $\qquad p = P(\text{defectuoso})$

Si $X \sim B(n, p)$ $\quad E(X) = np \quad V(X) = np(1-p)$

$X \sim B(1, p) \quad E(X) = p \quad V(X) = p(1-p)$

se quiere un intervalo de confianza para p de nivel $(1-\alpha) \cdot 100\%$

1) $\hat{p} = \overline{X}$

$$\frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \quad s^2$$

2) $$z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

si la muestra es grande entonces por T.C.L $z \tilde{\sim} N(0,1)$

$n\hat{p} \geq 10 \qquad n(1-\hat{p}) \geq 10$

$\therefore \; n\hat{p} = n\overline{X} = n \frac{1}{n} \sum x_i = \sum_{i=1}^{n} x_i \geq 10$ cantidad de aciertos mayor a 10

$n - \sum x_i \geq 10$ cantidad de no aciertos ≥ 10

numero de exitos y de fracasos por lo menos 10

$$\left[\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \; ; \; \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

Si nos necesito hallar el valor de n de manera tal que la longitud L sea menor que un valor determinado podemos

a) tomar muestra preliminar, con ella estimar p con $\hat{p}$ y despejar n

$$L = 2 z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq l \Rightarrow n \geq \left(\frac{2 z_{\frac{\alpha}{2}}}{l} \right)^2 \hat{p}(1-\hat{p})$$

b) si no hay preliminar, acotamos $\hat{p}(1-\hat{p}) \leq 0{,}5 \times (1 - 0{,}5)$

$$L \leq 2 z_{\frac{\alpha}{2}} \sqrt{\frac{0{,}5(1-0{,}5)}{n}} \leq l \Rightarrow n \geq \left(\frac{z_{\frac{\alpha}{2}}}{l} \right)^2$$

Intervalo de confianza para la diferencia de dos proporciones

pivote:

$$z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \approx N(0,1)$$

como antes

$$z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}(1-\hat{p}_2)}{n_2}}} \approx N(0,1)$$

intervalo

$$\left[\hat{p}_1 - \hat{p}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{p_2(1-\hat{p}_2)}{n_2}} \right]$$

PRACTICA 8

P-Valor

Si $P - \text{valor} > 0.05$ entonces no se rechaza H_0 con nivel de significancia 0.05

Si $P - \text{valor} < 0.05$ entonces se rechaza H_0 con nivel de significancia 0.05

Cálculo de P-valor para Z

Si las hipótesis son $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$:

$P - \text{valor} = P(|Z| > |z_0|)$

Si las hipótesis son $H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$:

$P - \text{valor} = P(Z > z_0)$

Si las hipótesis son $H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$:

$P - \text{valor} = P(Z < z_0)$

Cálculo de P-valor para t

Si las hipótesis son $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$:
$P - valor = P(|T| > |t_0|)$

Si las hipótesis son $H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$:
$P - valor = P(T > t_0)$

Si las hipótesis son $H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$:
$P - valor = P(T < t_0)$

Cálculo de P-valor para χ^2

Si las hipótesis son $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$:
$P - valor = 2 \cdot minimo[P(X < x_0), P(X > x_0)]$

Si las hipótesis son $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 > \sigma_0^2$:
$P - valor = P(X > x_0)$

Si las hipótesis son $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 < \sigma_0^2$:
$P - valor = P(X < x_0)$

ntervalo de confianza y Test de Hipótesis

Regla de decisión:

Si $\theta_0 \in IC(\theta) = (\widehat{\Theta} - \varepsilon ; \widehat{\Theta} + \varepsilon)$ entonces no rechazo H_0
Si $\theta_0 \notin IC(\theta) = (\widehat{\Theta} - \varepsilon ; \widehat{\Theta} + \varepsilon)$ entonces rechazo H_0

1) Test de hipótesis sobre la media de una distribución normal, varianza conocida

Planteo de la hipótesis:

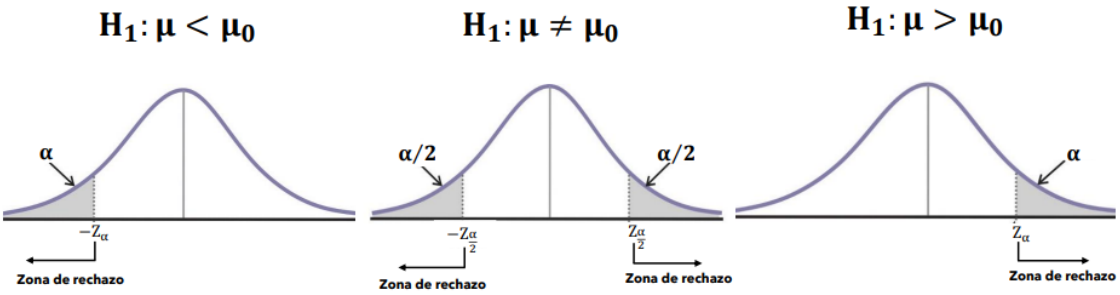
$$H_0: \mu \begin{cases} = \mu_0 \\ = \mu_0 \\ = \mu_0 \end{cases} \quad \text{vs} \quad H_1: \mu \begin{cases} > \mu_0 \\ \neq \mu_0 \\ < \mu_0 \end{cases}$$

¿Cuándo utilizo este Test?

- $X \sim N(\mu, \sigma^2)$
- σ^2 conocido

Estadístico de prueba:

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \text{ bajo } H_0$$



Hipótesis alternativa	Criterio de rechazo
$H_1: \mu \neq \mu_0$	$\lvert Z \rvert > Z_{\frac{\alpha}{2}}$
$H_1: \mu > \mu_0$	$Z > Z_\alpha$
$H_1: \mu < \mu_0$	$Z < -Z_\alpha$

2) Test de hipótesis sobre la media de una distribución normal, varianza desconocida

Planteo de la hipótesis:

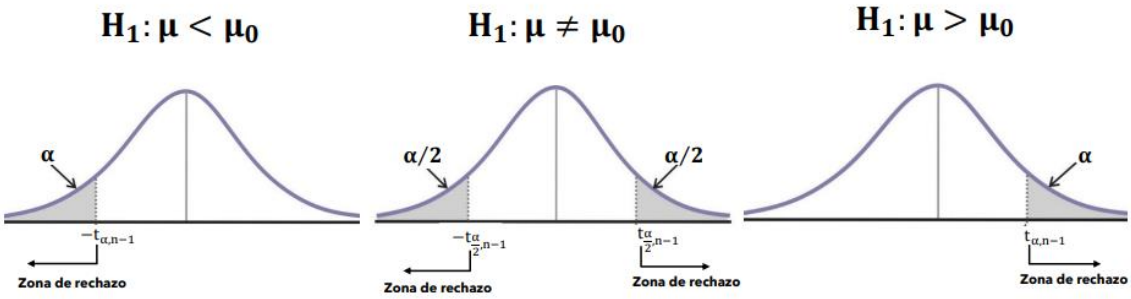
$$H_0: \mu \begin{cases} = \mu_0 \\ = \mu_0 \\ = \mu_0 \end{cases} \quad \text{vs} \quad H_1: \mu \begin{cases} > \mu_0 \\ \neq \mu_0 \\ < \mu_0 \end{cases}$$

¿Cuándo utilizo este Test?

- $X_i \sim N(\mu, \sigma^2)$
- σ^2 desconocido

Estadístico de prueba:

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \text{ bajo } H_0$$



Hipótesis alternativa	Criterio de rechazo
$H_1: \mu \neq \mu_0$	$\lvert T \rvert > t_{\frac{\alpha}{2},n-1}$
$H_1: \mu > \mu_0$	$T > t_{\alpha,n-1}$
$H_1: \mu < \mu_0$	$T < -t_{\alpha,n-1}$

3) Test de hipótesis sobre la media, varianza desconocida para muestras grandes

Planteo de la hipótesis:

$$H_0: \mu \begin{cases} = \mu_0 \\ = \mu_0 \\ = \mu_0 \end{cases} \quad \text{vs} \quad H_1: \mu \begin{cases} > \mu_0 \\ \neq \mu_0 \\ < \mu_0 \end{cases}$$

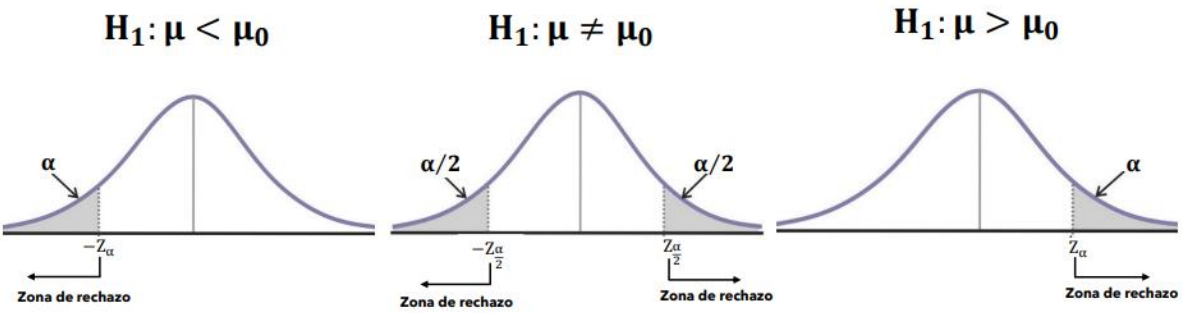
¿Cuándo utilizo este Test?

- σ^2 desconocido
- $n \geq 30$

Estadístico de prueba:

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \approx N(0,1) \text{ bajo } H_0$$

El nivel de significancia de este test es aproximado



Hipótesis alternativa	Criterio de rechazo
$H_1: \mu \neq \mu_0$	$\lvert Z \rvert > Z_{\frac{\alpha}{2}}$
$H_1: \mu > \mu_0$	$Z > Z_\alpha$
$H_1: \mu < \mu_0$	$Z < -Z_\alpha$

4) Test de hipótesis sobre la diferencia de dos medias, varianzas conocidas

Planteo de la hipótesis:

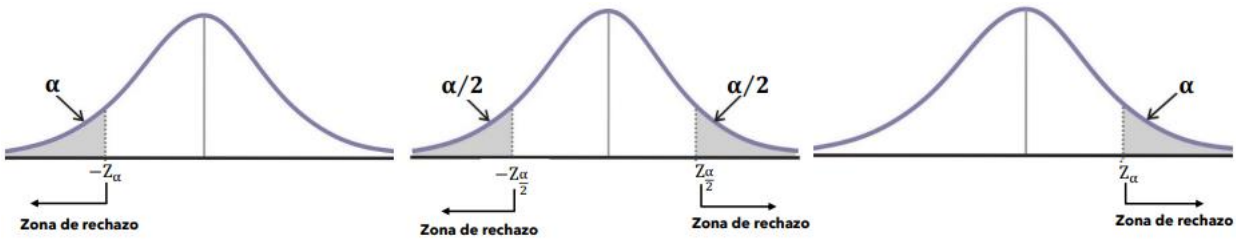
$$H_0: \mu_1 - \mu_2 \begin{cases} = \Delta_0 \\ = \Delta_0 \\ = \Delta_0 \end{cases} \quad \text{VS} \quad H_1: \mu_1 - \mu_2 \begin{cases} > \Delta_0 \\ \neq \Delta_0 \\ < \Delta_0 \end{cases}$$

¿Cuándo utilizo este test?

- X_{1i} y X_{2j} Independientes
- $X_{1i} \sim N(\mu_1, \sigma_1^2)$ y $X_{2j} \sim N(\mu_2, \sigma_2^2)$
- σ_1^2 y σ_2^2 conocidos

Estadístico de prueba:

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1) \text{ bajo } H_0$$



Hipótesis alternativa	Criterio de rechazo
$H_1: \mu_1 - \mu_2 \neq \Delta_0$	$\lvert Z \rvert > Z_{\frac{\alpha}{2}}$
$H_1: \mu_1 - \mu_2 > \Delta_0$	$Z > Z_\alpha$
$H_1: \mu_1 - \mu_2 < \Delta_0$	$Z < -Z_\alpha$

5) Test de hipótesis sobre la diferencia de dos medias, varianzas desconocidas para muestras grandes

Planteo de la hipótesis:

$$H_0: \mu_1 - \mu_2 \begin{cases} = \Delta_0 \\ = \Delta_0 \\ = \Delta_0 \end{cases} \quad \text{vs} \quad H_1: \mu_1 - \mu_2 \begin{cases} > \Delta_0 \\ \neq \Delta_0 \\ < \Delta_0 \end{cases}$$

¿Cuándo utilizo este test?

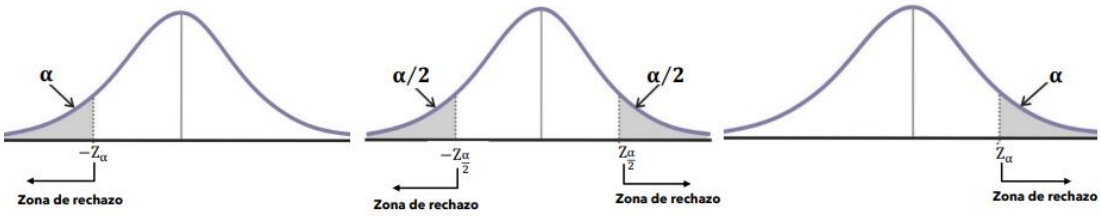
- X_{1i} y X_{2j} Independientes
- X_{1i} y X_{2j} Distribución desconocida
- $n_1 \geq 30$ y $n_2 \geq 30$
- σ_1^2 y σ_2^2 desconocidos

Estadístico de prueba:

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \approx N(0,1) \text{ bajo } H_0$$

El nivel de significancia de este test es aproximado

$\mathbf{H_1: \mu_1 - \mu_2 < \Delta_0}$ $\mathbf{H_1: \mu_1 - \mu_2 \neq \Delta_0}$ $\mathbf{H_1: \mu_1 - \mu_2 > \Delta_0}$



Hipótesis alternativa	Criterio de rechazo
$H_1: \mu_1 - \mu_2 \neq \Delta_0$	$\lvert Z \rvert > Z_{\frac{\alpha}{2}}$
$H_1: \mu_1 - \mu_2 > \Delta_0$	$Z > Z_\alpha$
$H_1: \mu_1 - \mu_2 < \Delta_0$	$Z < -Z_\alpha$

6) Test de hipótesis sobre la diferencia de dos medias, varianzas desconocidas e iguales

Planteo de la hipótesis:

$$H_0: \mu_1 - \mu_2 \begin{cases} = \Delta_0 \\ = \Delta_0 \\ = \Delta_0 \end{cases} \quad \text{vs} \quad H_1: \mu_1 - \mu_2 \begin{cases} > \Delta_0 \\ \neq \Delta_0 \\ < \Delta_0 \end{cases}$$

¿Cuándo utilizo este test?

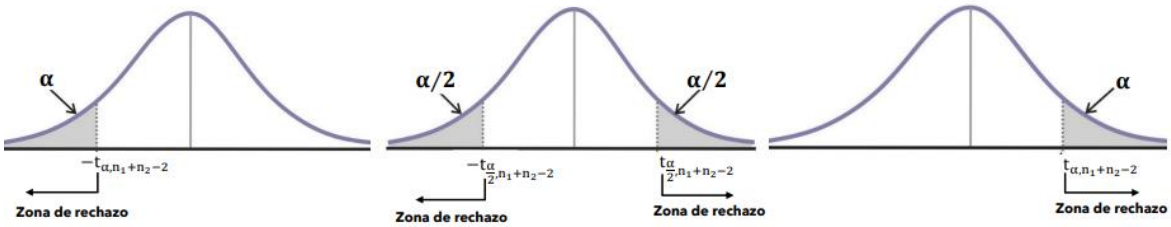
- X_{1i} y X_{2j} Independientes
- $X_{1i} \sim N(\mu_1, \sigma_1^2)$ y $X_{2j} \sim N(\mu_2, \sigma_2^2)$
- σ_1^2 y σ_2^2 desconocidos pero iguales $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (el enunciado lo tiene que decir)

Estadístico de prueba:

$$T = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2} \text{ bajo } H_0$$

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

$H_1: \mu_1 - \mu_2 < \Delta_0$ $\qquad$ $H_1: \mu_1 - \mu_2 \neq \Delta_0$ $\qquad$ $H_1: \mu_1 - \mu_2 > \Delta_0$



Hipótesis alternativa	Criterio de rechazo
$H_1: \mu_1 - \mu_2 \neq \Delta_0$	$\lvert T \rvert > t_{\frac{\alpha}{2},n_1+n_2-2}$
$H_1: \mu_1 - \mu_2 > \Delta_0$	$T > t_{\alpha,n_1+n_2-2}$
$H_1: \mu_1 - \mu_2 < \Delta_0$	$T < -t_{\alpha,n_1+n_2-2}$

7) Test de hipótesis sobre la diferencia de dos medias, varianzas desconocidas distintas

Planteo de la hipótesis:

$$H_0: \mu_1 - \mu_2 \begin{cases} = \Delta_0 \\ = \Delta_0 \\ = \Delta_0 \end{cases} \text{ VS } H_1: \mu_1 - \mu_2 \begin{cases} > \Delta_0 \\ \neq \Delta_0 \\ < \Delta_0 \end{cases}$$

¿Cuándo utilizo este test?

- X_{1i} y X_{2j} Independientes
- $X_{1i} \sim N(\mu_1, \sigma_1^2)$ y $X_{2j} \sim N(\mu_2, \sigma_2^2)$
- σ_1^2 y σ_2^2 desconocidos y distintos $\sigma_1^2 \neq \sigma_2^2$ (el enunciado no dice nada al respecto)

Estadístico de prueba:

$$T^* = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx t_\upsilon \text{ bajo } H_0$$

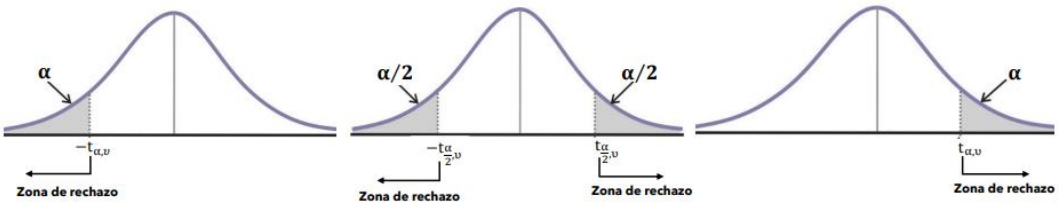
El nivel de significancia de este test es aproximado

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

si ν no es entero, ***se toma el entero más próximo a*** ν

Gráficos de la región de rechazo

$H_1: \mu_1 - \mu_2 < \Delta_0$ $\qquad$ $H_1: \mu_1 - \mu_2 \neq \Delta_0$ $\qquad$ $H_1: \mu_1 - \mu_2 > \Delta_0$



Hipótesis alternativa	Criterio de rechazo
$H_1: \mu_1 - \mu_2 \neq \Delta_0$	$\lvert T^* \rvert > t_{\frac{\alpha}{2},\upsilon}$
$H_1: \mu_1 - \mu_2 > \Delta_0$	$T^* > t_{\alpha,\upsilon}$
$H_1: \mu_1 - \mu_2 < \Delta_0$	$T^* < -t_{\alpha,\upsilon}$

8) Test de hipótesis sobre la diferencia de dos medias, datos de a pares

Planteo de la hipótesis:

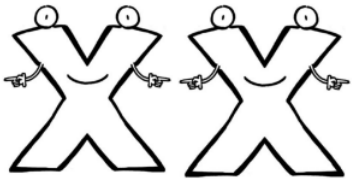
$$H_0: \mu_1 - \mu_2 \begin{cases} = \Delta_0 \\ = \Delta_0 \\ = \Delta_0 \end{cases} \text{ vs } H_1: \mu_1 - \mu_2 \begin{cases} > \Delta_0 \\ \neq \Delta_0 \\ < \Delta_0 \end{cases}$$

¿Cuándo utilizo este test?

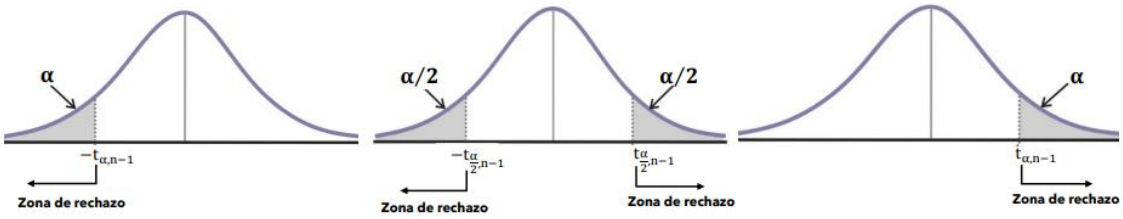
- Las observaciones se dan de a pares y no son independientes
- $D_i \sim N(\mu_D, \sigma_D^2) \quad i = 1,2,\dots,n$
- σ_D^2 desconocido

Estadístico de prueba:

$$T = \frac{\overline{D} - \Delta_0}{S_D/\sqrt{n}} \sim t_{n-1} \text{ bajo } H_0$$



$\mathbf{H_1: \mu_1 - \mu_2 < \Delta_0}$ $\qquad$ $\mathbf{H_1: \mu_1 - \mu_2 \neq \Delta_0}$ $\qquad$ $\mathbf{H_1: \mu_1 - \mu_2 > \Delta_0}$



Hipótesis alternativa	Criterio de rechazo
$H_1: \mu_1 - \mu_2 \neq \Delta_0$	$\lvert T \rvert > t_{\frac{\alpha}{2},n-1}$
$H_1: \mu_1 - \mu_2 > \Delta_0$	$T > t_{\alpha,n-1}$
$H_1: \mu_1 - \mu_2 < \Delta_0$	$T < -t_{\alpha,n-1}$

9) Test de hipótesis sobre la varianza de una distribución normal

Planteo de la hipótesis:

$$H_0: \sigma^2 \begin{cases} = \sigma_0^2 \\ = \sigma_0^2 \\ = \sigma_0^2 \end{cases} \quad \text{vs} \quad H_1: \mu \begin{cases} > \sigma_0^2 \\ \neq \sigma_0^2 \\ < \sigma_0^2 \end{cases}$$

Estadístico de prueba:

$$X = \frac{(n-1)^2 S^2}{\sigma_0^2} \sim \chi^2{}_{n-1} \text{ bajo } H_0$$

¿Cuándo utilizo este test?

- $X_i \sim N(\mu, \sigma^2)$
- Se pida un test para la varianza

Hipótesis alternativa	Criterio de rechazo
$H_1: \sigma^2 \neq \sigma_0^2$	$X > \chi^2_{\frac{\alpha}{2},n-1}$
$H_1: \sigma^2 > \sigma_0^2$	$X > \chi^2_{\alpha,n-1}$
$H_1: \sigma^2 < \sigma_0^2$	$X < \chi^2_{1-\alpha,n-1}$

10) Test de hipótesis sobre la igualdad de dos varianzas

Sean las hipótesis $H_0 : \sigma_1^{\ 2} = \sigma_2^2$ contra $H_1 : \sigma_1^{\ 2} \neq \sigma_2^2$

Tomamos como estadístico de prueba a $F = \dfrac{S_1^2}{S_2^2}$

Vemos que $F = \dfrac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1}$ si $H_0 : \sigma_1^{\ 2} = \sigma_2^2$ es verdadera

Recordando que la distribución Fisher es asimétrica, la regla de decisión es

$$\begin{cases} rechazar\ \ H_0\ si\ \ F > f_{\frac{\alpha}{2}\cdot n_1-1\cdot n_2-1} \quad ó \quad F < f_{1-\frac{\alpha}{2}\cdot n_1-1\cdot n_2-1} \\ \quad aceptar\ \ H_0\ si\ \ f_{1-\frac{\alpha}{2}\cdot n_1-1\cdot n_2-1} \leq F \leq f_{\frac{\alpha}{2}\cdot n_1-1\cdot n_2-1} \end{cases}$$

Si $H_1 : \sigma_1^{\ 2} > \sigma_2^2$ entonces la regla de decisión es $\begin{cases} rechazar\ \ H_0\ si\ \ F > f_{\alpha\cdot n_1-1\cdot n_2-1} \\ aceptar\ \ H_0\ si\ \ F \leq f_{\alpha\cdot n_1-1\cdot n_2-1} \end{cases}$

Si $H_1 : \sigma_1^{\ 2} < \sigma_2^2$ entonces la regla de decisión es $\begin{cases} recahzar\ \ H_0\ si\ \ F < f_{1-\alpha\cdot n_1-1\cdot n_2-1} \\ aceptar\ \ H_0\ si\ \ F \geq f_{1-\alpha\cdot n_1-1\cdot n_2-1} \end{cases}$

Para calcular el p-valor, si el estadístico F tomó el valor f_0, y teniendo en cuenta que no hay simetría en la distribución Fisher, hacemos:

Si $H_1 : \sigma_1^{\ 2} > \sigma_2^2$ entonces $p-valor = P(F > f_0)$

Si $H_1 : \sigma_1^{\ 2} < \sigma_2^2$ entonces $p-valor = P(F < f_0)$

Si $H_1 : \sigma_1^{\ 2} \neq \sigma_2^2$ entonces $p-valor = 2\min\left(P(F < f_0)\ ,P(F > f_0)\right)$

11) Test de hipótesis sobre una proporción

<u>Observaciones</u>:

1- La prueba descrita anteriormente requiere que la proporción muestral esté normalmente distribuida. Esta suposición estará justificada siempre que $np_0 > 10$ y $n(1-p_0) > 10$, donde p_0 es la proporción poblacional que se especificó en la hipótesis nula.

2- También se podía haber tomado como estadístico de prueba a $Z = \dfrac{X - np_0}{\sqrt{np_0(1-p_0)}}$ donde $X \sim B(n,p)$

Planteo de la hipótesis:

$$H_0: p \begin{cases} = p_0 \\ = p_0 \\ = p_0 \end{cases} \quad \text{vs} \quad H_1: p \begin{cases} > p_0 \\ \neq p_0 \\ < p_0 \end{cases}$$

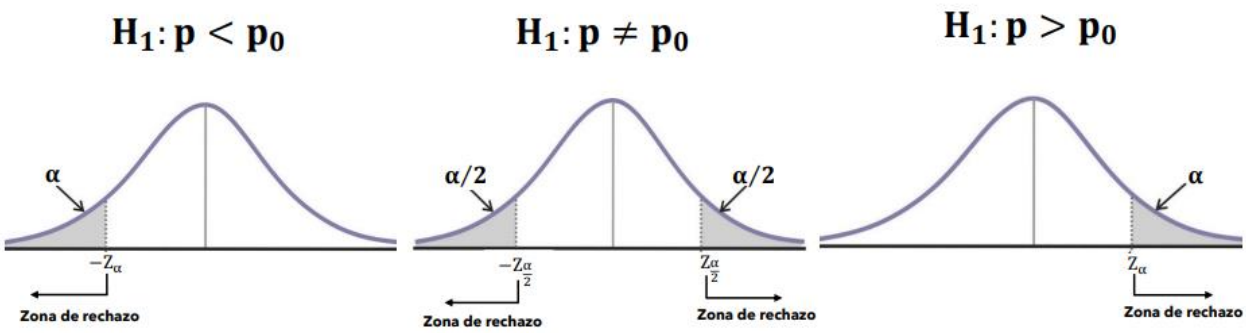
¿Cuándo utilizo este test?

- $X \sim B(n, p)$
- $n \geq 30$

Estadístico de prueba:

$$Z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \approx N(0,1) \text{ bajo } H_0$$

El nivel de significancia de este test es aproximado



Hipótesis alternativa	Criterio de rechazo
$H_1: p \neq p_0$	$\lvert Z \rvert > Z_{\frac{\alpha}{2}}$
$H_1: p > p_0$	$Z > Z_\alpha$
$H_1: p < p_0$	$Z < -Z_\alpha$

12) Tests de hipótesis sobre dos proporciones

Entonces el estadístico de prueba es $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ que bajo $H_0: p_1 - p_2 = 0$ se puede probar que tiene distribución aproximadamente $N(0,1)$

Entonces la regla de decisión es $\begin{cases} rechazar \;\; H_0 \;\; si \;\; |Z| > z_{\frac{\alpha}{2}} \\ aceptar \;\; H_0 \;\; si \;\; |Z| \leq z_{\frac{\alpha}{2}} \end{cases}$ donde $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

Si $H_1: p_1 - p_2 > 0$ entonces la regla de decisión es $\begin{cases} rechazar \;\; H_0 \;\; si \;\; Z > z_\alpha \\ aceptar \;\; H_0 \;\; si \;\; Z \leq z_\alpha \end{cases}$

Si $H_1 : p_1 - p_2 < 0$ entonces la regla de decisión es $\begin{cases} rechazar \;\; H_0 \;\; si \;\; Z < -z_\alpha \\ aceptar \;\; H_0 \;\; si \;\; Z \geq -z_\alpha \end{cases}$

el estadístico de prueba es:

$$\boxed{Z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}}$$

$\sim N(0,1)$ por T.C.L y por ser $\hat{p}$ estimador consistente

CS Escaneado con CamScanner