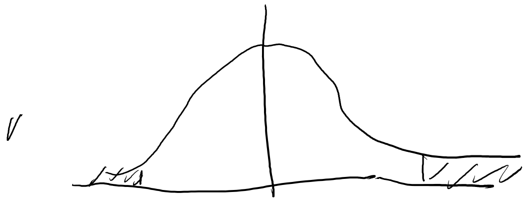
Inferencia sobre el vector de medias

viernes, 25 de febrero de 2022 8:40

Recordemos

P.H respecto a una media poblacional μ_0



nivel α

- Si $X_1, \dots X_n$ es una muestra aleatoria de una ~~po~~ población normal, entonces el estadístico de prueba es

$$t = \frac{(\bar{X} - \mu_0)}{s/\sqrt{n}} \rightarrow S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})$$

$$\bar{X} = \frac{1}{n} \sum x_i$$

Si H_0 es cierta, t tiene una distribución t-student.

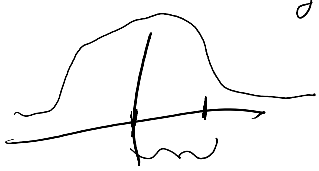
Rechazamos H_0 si $t \in RR \rightarrow |t| > t_{\alpha/2} \rightarrow$ dos colas

$$\left.\begin{array}{l} t > t_\alpha \\ t < -t_\alpha \end{array}\right\} \text{una cola}$$

Rechazamos para $|t|$ grande $\Rightarrow$ Rechazamos para t^2 grande

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \rightarrow t^2 = n(\bar{x} - \mu_0)(S^2)^{-1}(\bar{x} - \mu_0) \text{ Distancia estadística}$$

Es decir, t^2 grande $\Rightarrow$ Hay mucha distancia entre $\bar{x}$ y μ_0. Entonces rechazamos la hipótesis nula



$$n(\bar{x} - \mu_0)(S^2)^{-1}(\bar{x} - \mu_0) > t^2_{n-1}(\alpha/2)$$

- Si H_0 no se rechaza $\Rightarrow \mu_0$ es un valor posible de μ

• Si H_0 no se rechaza $\Rightarrow$ μ_0 es ... μ

- Recordemos que no rechazar H_0 con un nivel α es equivalente a que el valor μ_0 NO esté en el intervalo de confianza $1-\alpha$

- Valores μ_0 en I.C son aquellos para los cuales no se rechaza $H_0: \mu = \mu_0$.  I.C

- Los límites de los I.C son variables aleatorias

- Probabilidad de que el intervalo de confianza contenga a μ es de $1-\alpha$, Entonces si tomo muchas muestras, el $(1-\alpha)\cdot 100\%$ contendrá a μ.

Generalizamos a p variables.

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \quad \text{vector de medias}$$

notación matriz de varianzas y cov.

$$T^2 = (\bar{X} - \mu_0)' \left(\frac{1}{n} \underset{\downarrow}{S}\right)^{-1} (\bar{X} - \mu_0) = \qquad \bar{X} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \end{pmatrix}$$

$$= n (\bar{X} - \mu_0)' (S^{-1}) (\bar{X} - \mu_0)$$

$$\bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j \qquad S = \frac{1}{n-1} \sum (X_j - \bar{X})(X_j - \bar{X})'$$

Observaciones:

1) T^2 se llama T^2 de Hotelling.

2) $\frac{1}{n}S$ es la estimación de la covarianza de X.

3) Si T^2 es grande $\Rightarrow$ $\bar{X}$ estará lejos de μ_0 $\Rightarrow$
$\Rightarrow$ Se rechaza la H_0.

4) En este caso si H_0 es cierta:

$$T^2 \sim \frac{(n-1)}{(n-p)} F_{p,n-p}$$

Entonces

$$\alpha = P\Big(\underbrace{T^2 > \frac{(n-1)p}{(n-p)} F_{p,n-p}}_{\text{rechazar } H_0} \;\Big|\; \underbrace{H_0 \text{ es cierta}}_{\text{siendo cierta}} \Big)$$

$\alpha \downarrow$ error Tipo 1
Rechazar H_0 siendo cierta

$$\alpha = P\left(n(\bar{X}-\mu_0)'(S^{-1})(\bar{X}-\mu_0) > \frac{(n-1)p}{(n-p)} F_{p,n-p} \;\Big|\; H_0 \text{ es cierta} \right)$$

Si tenemos una P. H. tal que
$H_0: \mu = \mu_0 \quad H_a: \mu \neq \mu_0$ nivel α

Rechazaremos si

$$T^2 = n(\bar{X}-\mu_0)'(S^{-1})(\bar{X}-\mu_0) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

Ejemplo

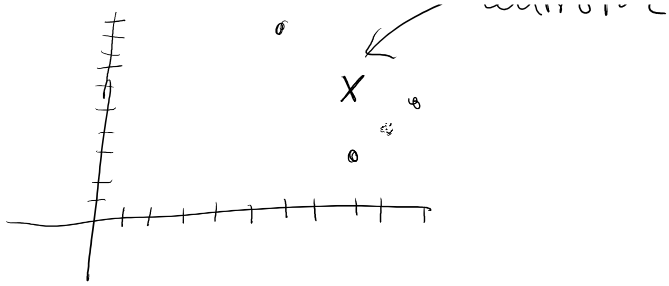
$n=3$ $\downarrow$ observaciones
$p=2$ $\rightarrow$ variables

centroide

X

Observaciones variables

$$X = \begin{pmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{pmatrix} \begin{matrix} \leftarrow P1 \\ \leftarrow \\ \end{matrix}$$



$$\bar{X} = \begin{pmatrix} 8 \\ 6 \end{pmatrix} \qquad S = \begin{pmatrix} 4 & -3 \\ -3 & 9 \end{pmatrix} \Rightarrow \text{Comprobar}$$

$$H_0 : \mu_0 = \begin{pmatrix} 9 \\ 5 \end{pmatrix}$$

$$S^{-1} = \frac{1}{4 \cdot 9 - (9)} \begin{pmatrix} 9 & 3 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix}$$

$$T^2 = n\,(\underset{2\times 1}{\bar{X}} - \underset{2\times 1}{\mu_0})'\,\underset{(2\times 2)}{(S^{-1})}\,(\underset{2\times 1}{\bar{X}} - \underset{2\times 1}{\mu_0}) \overset{\text{comprobar}}{=} 7/9$$

$$F_{2,1}(0.05) = 199{,}51$$

$$\frac{(3-1)\cdot 2}{3-2}\, 199{,}51 = 798.04$$

$$T^2 > \frac{(n-1)p}{(n-p)} F \ ? \Rightarrow \text{No rechazamos } H_0.$$

$$\Downarrow$$

$$\mu = \begin{pmatrix} 9 \\ 5 \end{pmatrix}$$

Observación: T^2 es invariante bajo cambios casl ej de las medidas de X.

$$Y = \underset{p\times p}{C}\,\underset{p\times 1}{X} + \underset{p\times 1}{d}$$

C ↓ matriz constante

d ↘ vector constante

constante

Dado que $Y = CX + d$

$$\mu_Y = C\mu + d \qquad S_Y = CSC'$$
$$\bar{Y} = C\bar{X} + d \qquad , \qquad \Sigma_Y = C\Sigma C'$$

$$T^2 = n(\bar{Y} - \mu_{Y_0})'(S_Y)^{-1}(\bar{Y} - \mu_{Y_0}) =$$

$$= n(\bar{X} - \mu_0)'(S^{-1})(\bar{X} - \mu_0)$$

Método de razón de verosimilitud

$$\max L(\mu, \Sigma) = \frac{1}{(2\pi)^{np/2} |\hat{\Sigma}|^{n/2}} \cdot e^{-np/2}$$

Estimaciones

$$\hat{\Sigma} = \frac{1}{n} \Sigma (x_j - \bar{x})(x_j - \bar{x})' \qquad \mu = \bar{X}$$

$$L(\mu_0, \Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(-\frac{1}{2} \Sigma (x_j - \mu_0)' \Sigma^{-1} (x_j - \mu_0)\right)$$

· Fijamos μ_0:

Podemos variar Σ para encontrar valores más posibles con las obs. que se tienen.

$$\max L(\mu_0, \Sigma) = \frac{1}{(2\pi)^{np/2} |\hat{\Sigma}_0|^{n/2}} e^{-np/2}$$

$$\hat{\Sigma}_0 = \frac{1}{n} \Sigma (x_j - \mu_0)(x_j - \mu_0)$$

Para determinar si μ_0 es un posible valor de μ, se utilizan los maximos de $L(\mu_0, \Sigma)$ y $L(\mu, \Sigma)$:

$$\underset{\substack{\uparrow \\ \text{lambda} \\ \text{de Wilk}}}{\Lambda} = \frac{\max L(\mu_0, \Sigma)}{\max L(\mu, \Sigma)} = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{n/2}$$

Λ es pequeña $\Rightarrow |\hat{\Sigma}_0|$ es grande $\Rightarrow H_0$ es poco probable

$\Rightarrow$ Rechazamos H_0.

$\Rightarrow$ Si $|\hat{\Sigma}_0|$ es grande $\Rightarrow$ μ_0 es muy distante de μ.

Concretamente, $\left.\Lambda < C_\alpha \right\}$ No lo vamos a usar.

Podemos decir que:

$$\boxed{-2\ln(\Lambda) \sim \chi^2_p}$$

Example 4: four psychological tests

x media de las variables

$n = 64$, $p = 4$, $\bar{\mathbf{x}}' = (14.15, 14.91, 21.92, 22.34)$,

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{14} \\ \vdots & & & \vdots \\ x_{64} & & & x_{644} \end{pmatrix}$$

$$\mathbf{S} = \begin{pmatrix} 10.388 & 7.793 & 15.298 & 5.3740 \\ 7.793 & 16.658 & 13.707 & 6.1756 \\ 15.298 & 13.707 & 57.058 & 15.932 \\ 5.374 & 6.176 & 15.932 & 22.134 \end{pmatrix} \quad \& \ \det(\mathbf{S}) = 61952.085$$

Test: $H_o : \boldsymbol{\mu}' = (20, 20, 20, 20)$ versus $H_o : \boldsymbol{\mu}' \neq (20, 20, 20, 20)$

$$\boldsymbol{\Sigma}_o = \frac{1}{n}(\mathbf{X} - \mathbf{1}\mu_o')'(\mathbf{X} - \mathbf{1}\mu_o') = \begin{pmatrix} 44.375 & 37.438 & 3.828 & -8.406 \\ 37.438 & 42.344 & 3.703 & -5.859 \\ 3.828 & 3.703 & 59.859 & 20.187 \\ -8.406 & -5.859 & 20.187 & 27.281 \end{pmatrix}$$

$\det(\boldsymbol{\Sigma}_o) = 518123.8$.

Wilk's Lambda is $\Lambda = (61952.085/518123.8)^{64/2} = 3.047E - 30$, and
Comparing $-2\ln(\Lambda) = 135.92659$ to a χ^2_4 gives p–value $<< .01$.

Tendríamos que buscar χ^2_4 en la tabla y comparar con Λ.

Relación de Λ entre T^2.

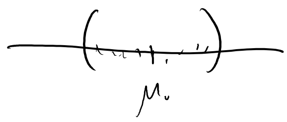
$$(\Lambda)^{2/n} = \left(1 + \frac{T^2}{(n-1)}\right)^{-1}$$

Cuanto más grande es T^2, más pequeño es Λ

Rechazamos H_0, cuando T^2 es grande

| " H_0, cuando Λ es pequeña.

Próxima clase:

Univariante 

Varias variables