Comparación de medias multivariadas

Lo visto antes lo podemos usar para comparar varias medias multivariadas

Comparación de 2 medias

$\longrightarrow$ se usa en estudios clínicos.

x_{j1} : respuesta en el j-ésimo item en el tratamiento 1

x_{j2} : respuesta en el j-ésimo item en el tratamiento 2

Definimos: $D_j = x_{j1} - x_{j2}$, $j = 1, \dots, n$. $D_j \sim N(\delta, \sigma_d^2)$

$$t = \frac{\overset{\text{vector promedio}}{\overline{D}} - \delta_0}{s_d / \sqrt{n}} \sim t_{n-1}$$

$H_0 : \delta = 0$, $H_a : \delta \neq 0$, Rechazamos H_0 si:

$$|t| > t_{n-1}(\alpha/2)$$

I. Confianza de δ :

$$\bar{d} \pm t_{n-1}(\alpha/2) \cdot \frac{s_d}{\sqrt{n}}$$

x_{1j1} (tratamiento, variable) Variable 1 con trat. 1

x_{1j2} Variable 2 con trat. 1

$\vdots$

x_{1jp} Variable p con trat. 1

Entonces:

$$\left.\begin{aligned} D_{j1} &: x_{1j1} - x_{2j1} \\ &\vdots \\ D_{jp} &: x_{1jp} - x_{2jp} \end{aligned}\right] \rightarrow D_j = \begin{pmatrix} D_{j1} \\ \vdots \\ D_{jp} \end{pmatrix}$$

$$E(D_j) = \delta = \delta \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_p \end{pmatrix}, \quad Cov(D_j) = \Sigma_d$$

Sea $D_1 \dots D_n$ indep. asumimos que son $N_p(\delta, \Sigma_d)$ (por el TLC)

entonces $$T^2 = n(\bar{D} - \delta)' S_d^{-1} (\bar{D} - \delta)$$

tiene distribución:

$$\frac{(n-1)p}{(n-p)} F_{p,n-p}$$ ← si $(n-p)$ es pequeño y asumimos $D_1, \dots, D_n$ Normales.

Si $(n-p)$ es grande:

$$T^2 \overset{aprox.}{\sim} \chi^2_p$$

Dados los valores observados $d_j = \begin{pmatrix} d_{j1} \\ d_{j2} \\ \vdots \\ d_{jp} \end{pmatrix}$ $j = 1, \dots, n$

Podemos evaluar $H_0: \delta = 0$ $H_a: \delta \neq 0$ nivel α

usando:

si H_0 es cierta

$$T^2 = n \bar{d}' S^{-1} \bar{d}' \bar{d} \sim \frac{(n-1)p}{(n-p)} \cdot F_{p,n-p}$$

Rechazamos H_0 si $T^2 > \frac{(n-1)p}{(n-p)} F_{p,n-p}$.

Una región de confianza con conf. $1-\alpha$ de δ son todos los δ tq.

$$n(\bar{d} - \delta)' S_d^{-1} (\bar{d} - \delta) \leq \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

Los I.C. simultaneos con confianza $(1-\alpha)$ para los δ_i's están dados por:

$$\underbrace{\bar{d}_i}_{\substack{\text{promedio} \\ \text{de la desviación} \\ \text{en la variable } i}} \pm \sqrt{\frac{(n-1)p}{(n-p)} \cdot F_{p,n-p}(\alpha)} \cdot \sqrt{\frac{S^2_{d_i}}{n}}$$

→ i-ésimo elemento en la diagonal de la matriz S_d

→ desviación estándar

Si $(n-p)$ es grande:

$$\frac{(n-1)}{(n-p)} \pm F_{p,n-p}(\alpha) \approx \chi^2_p(\alpha)$$

Intervalos de Bonferroni Simultaneos con Conf. $(1-\alpha)$ son:

$$\text{para} \quad \delta_i \longrightarrow \bar{d}_i \pm t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{d_i}^2}{n}}$$

→ Obs: $\bar{d}$, S_d y T^2 se pueden calcular de $\bar{X}$ y de S.

Si:

$$X_{n\times 2p} = \left[\begin{array}{cccc|cccc} x_{111} & x_{112} & \dots & x_{11p} & x_{121} & \dots & & x_{12p} \\ \vdots & & & & \vdots & & & \\ x_{n11} & \dots & & x_{n1p} & x_{n21} & \dots & & x_{n2p} \end{array}\right]$$

fila tratamiento

variable

otra forma de escribir:

$$\bar{X} = (\bar{x}_{11}, \dots \bar{x}_{1p} \,\vdots\, \bar{x}_{21}, \dots, \bar{x}_{2p}) = (\bar{X}_1 | \bar{X}_2)$$

$$S = \left(\begin{array}{c|c} S_{11} & S_{12} \\ \hline S_{21} & S_{22} \end{array}\right)$$

Si definimos:

$$C_{p\times 2p} = \left(\begin{array}{ccc|ccc} 1 & & 0 & -1 & & 0 \\ & \ddots & & & \ddots & \\ 0 & & 1 & 0 & & -1 \end{array}\right)$$

$$d_j = C X_j \,, \quad \bar{d} = C\bar{X}$$

$$S_d = C S C'$$

→ ventaja Ahorramos cálculos

entonces:

$$T^2 = n\,\bar{x}\,C'\,(C S C')^{-1}\,C\bar{X}$$

Comparación de Varios tratamientos. Supongamos que se desea comparar q tratamientos con una variable respuesta.

Cada item recibe cada tratamiento una vez.

La j-ésima obs:

$$X_j = \begin{pmatrix} x_{j1} \\ \vdots \\ x_{jq} \end{pmatrix}$$

donde x_{ji} es la respuesta al i-ésimo tratamiento del j-ésimo item.

Para comparar tratamientos podemos usar matrices de contraste sobre los componentes de $\mu = E(X_j)$ $\boxed{\text{ej:}}$

$$\begin{pmatrix} \mu_1 - \mu_2 \\ \vdots \\ \mu_1 - \mu_q \end{pmatrix} = \underbrace{\begin{array}{c} \begin{array}{ccccc} \mu_1 & \mu_2 & & & \\ \downarrow & \downarrow & \downarrow & \downarrow \cdots & \downarrow \end{array} \\ \begin{pmatrix} 1 & -1 & 0 & 0 & \\ 1 & 0 & -1 & 0 & \\ \vdots & 0 & 0 & -1 & \\ & \vdots & \vdots & & \ddots \\ 1 & 0 & 0 & & -1 \end{pmatrix} \end{array}}_{\text{matriz de contraste}} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_q \end{pmatrix} = C\mu$$

Podemos evaluar la "igualdad" de tratamiento evaluando $C\mu = \vec{0}$

En este caso las medias muestrales son $C\bar{x}$ y la matriz cov. muestral es CSC'
estad. de prueba:

$$T^2 = n\,(C\bar{x})'(CSC')^{-1}(C\bar{x})$$

Ahora suponga una población $N_q(\mu, \Sigma)$ sea C una matriz de contraste. Si se tiene

$$H_0: C\mu = 0 \quad vs \quad H_a: C\mu \neq 0 \quad \text{nivel } \alpha$$

Rechazamos H_0 si:

$$T^2 = n\,(C\bar{x})'\big(\overbrace{CSC'}^{\text{matriz de covarianzas y varianzas de } C\bar{x}}\big)^{-1} C\bar{x} > \frac{(n-1)(q-1)}{(n-q+1)} F_{q-1,\, n-q+1}(\alpha)$$

Una región de confianza (con conf. $1-\alpha$) para $C\mu$ se construye con todos los valores de $C\mu$ t.q.

$$n\,(C\bar{x} - C\mu)'(CSC')^{-1}(C\bar{x} - C\mu) \leq \frac{(n-1)(q-1)}{(n-q+1)} F_{q-1,\, n-q+1}(\alpha)$$

Los intervalos de confianza simultáneos para $C\mu$ se obtienen

$$\underset{\substack{\uparrow \\ \text{fila de } C}}{c_i'}\mu = c_i'\bar{x} \pm \sqrt{\frac{(n-1)(q-1)}{(n-q+1)} F_{q-1,n-q+1}(\alpha)}\sqrt{\frac{c_i' S c_i}{n}}$$