Comparar vec medias 2 poblaciones

Sunday, March 6, 2022 8:13 PM

Comparar vec medias 2 poblaciones

Se puede usar el estadístico T^2 para evaluar la igualdad de los vecs. de medias.

Ej: ♂ vs. ♀

muestra 1 = $\vec{X}_{11} \dots \vec{X}_{1n_1}$ $\quad \bar{\vec{x}} = \frac{1}{n_1}\sum_{j=1}^{n_1} x_{1j}$

Pobla° 1 obs 1

$$S_1 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(\vec{X}_{1j} - \bar{\vec{X}}_1)'(\vec{X}_{1j} - \bar{\vec{X}}_1)$$

Muestra 2: $\vec{X}_{21}, \dots, \vec{X}_{2n_2}$

$\vec{\mu}_1$: Media en pobla° 1
$\vec{\mu}_2$: Media en pobla° 2.
} $H_0 = \vec{\mu}_1 - \vec{\mu}_2$

Suposiciones

1. $X_{11}, \dots, X_{1n_1}$ es una m.a de una pobla° p-variada con media μ_1, Cov Σ_1.
2. $X_{21}, \dots, X_{2n_2}$ es una m.a de una pobla° p-variada con media μ_2, Cov Σ_2.
3. $X_{11}, \dots, X_{1n_1}$ son indep con $X_{21}, \dots, X_{2n_2}$

- si n-p grande, es suficiente con esas 3 suposiciones.
- n-p NO es grande:
  4. Pobla°es normales
  5. $\Sigma_1 = \Sigma_2$.

Para hacer la P.H

$H_0: \mu_1 - \mu_2 = \delta_0$ vs $H_A: \mu_1 - \mu_2 \neq \delta_0$

Podemos usar la dist estadística de $\bar{X}_1 - \bar{X}_2$ a δ_0

$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$

$Cov(\bar{X}_1, \bar{X}_2) = 0$ (son indep).

$cov(\bar{X}_1 - \bar{X}_2) = \frac{1}{n_1}\Sigma + \frac{1}{n_2}\Sigma \longrightarrow$ Asumiendo $\Sigma_1 = \Sigma_2$.

$= \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\Sigma$

Spooled a la estima° de Σ cuando hay 2 poblaciones.

$T^2 = (\bar{X}_1 - \bar{X}_2 - \delta_0)'\left(\left(\frac{1}{n_1} + \frac{1}{n_2}\right)S_{pooled}\right)^{-1}(\bar{X}_1 - \bar{X}_2 - \delta_0) > c^2$ depende de la ~ de T^2

Teorema

Si $X_{11}, \dots, X_{1n_1}$ es una m.a de tamaño n_1
$X_{21}, \dots, X_{2n_2}$ es una m.a de tamaño n_2,
de $\sim N_p(\mu_1, \Sigma)$ y $N_p(\mu_2, \Sigma)$ respectivam.

Entonces,

$$T^2 = (\bar{X}_1 - \bar{X}_2 - \delta_0)'\left(\left(\frac{1}{n_1} + \frac{1}{n_2}\right)S_{pooled}\right)^{-1}(\bar{X}_1 - \bar{X}_2 - \delta_0)$$

$$\frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p, n_1+n_2-p-1}$$

$$P\left(T^2 \leq \frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p, n_1+n_2-p-1}(\alpha)\right) = 1 - \alpha$$

I.C. simultáneos T^2

sea $c^2 = \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p, n_1+n_2-p-1}(\alpha)$

$a'(\bar{X}_1 - \bar{X}_2) \pm c\sqrt{a'\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{pooled} \cdot a}$ $\quad$ $1-\alpha$ contendrá el val. real de $a'(\bar{X}_1 - \bar{X}_2)$

I. Bonferroni con confianza $(1-\alpha)$

simultáneos para la variable i

$$(\bar{x}_{1i} - \bar{x}_{2i}) \pm t_{n_1+n_2-2}\left(\frac{\alpha}{2p}\right)\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right) S_{ii\,pooled}}$$

Caso $\Sigma_1 \neq \Sigma_2$. → Se necesita n-p grande.

Si los tamaños de las muestras satisfacen que (n_1-p) y (n_2-p) grandes. La $(1-\alpha)\%$ elipsoide de confianza para $\mu_1-\mu_2$ está dada por los vals de $\mu_1-\mu_2$ que satisfacen

$$(\bar{x}_1 - \bar{x}_2 - (\mu_1-\mu_2))' \underbrace{\left(\frac{1}{n_1}S_1 + \frac{1}{n_2}S_2\right)^{-1}}_{\text{matriz}} (\bar{x}_1 - \bar{x}_2 - (\mu_1-\mu_2)) \leq \chi^2_p(\alpha)$$

Lo mismo para las comb. lin. $a'(\mu_1-\mu_2)$, los IC son dados por:

$$a'(\bar{x}_1 - \bar{x}_2) \pm \sqrt{\chi^2_p(\alpha)}\sqrt{a'\left(\frac{1}{n_1}S_1 + \frac{1}{n_2}S_2\right)a}$$

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

MANOVA

ANOVA . Veo si mis poblaciones son iguales o no.
Análisis de varianza → veo si las dif. en medias se deben a sus varianzas o a que son ≠ pob.
la variabilidad entre los grupos o la variabilidad interna de c/ grupo.

Se busca comparar más de 2 poblaciones.

Pob 1 : $x_{11} \dots x_{1n_1}$ → n obs con p vars en la poblaº 1.
⋮
Pob g : $x_{g1} \dots x_{gn_g}$ → " " g.

Buscamos investigar cuáles medias poblacionales son = y si no, cuales son ≠.

Suposiciones:

1: $x_{l1}, \dots, x_{ln_l}$ es una m.a de tamaño n_l de una poblaº con media μ_l $l=(1,\dots,g)$ indep. entre poblaciones (las muestras de c/ poblaº es indep a las demás)

2. Todas las poblaciones tienen la misma varianza Σ.

3. Todas las poblaciones son N_p normales multivar.

ANOVA (univar)

$x_{l1}, \dots, x_{ln_l}$ m.a de $N(\mu_l, \sigma^2)$ $l=(1,\dots,g)$ → muestra indep.

Buscamos evaluar $\mu_1 = \mu_2 = \dots = \mu_g$

$$\mu_l = \underset{\text{media general}}{\mu} + \tau_l$$ → desviaº o efecto asoc. a la poblaº l (tratamiento l)

Para nuestra P.H : $H_0 : \mu_1 = \dots = \mu_g \longleftrightarrow \tau_1 = \dots = \tau_g = 0$.

obs j, pob l

$x_{lj} \sim (\mu + \tau_l, \sigma^2) \Rightarrow x_{lj} = \mu + \tau_l + e_{lj}$ → error aleatorio
↳ $e_{lj} \sim N(0, \sigma^2)$ (son indep entre ellos)

• Tenemos q' fijar q:

$$\sum_{l=1}^{g} n_l \tau_l = 0$$

• Descomponemos c/ obs como:

$$\underset{\text{(obs)}}{x_{lj}} = \underset{\text{media total}}{\bar{x}} + \underset{\text{efecto del trat } \tau_l}{(\bar{x}_l - \bar{x})} + \underset{\text{error: } e_{lj}}{(x_{lj} - \bar{x}_l)}$$

val obs − media en tratam.

Rechazamos H_0 si la contribuº de los tratamientos es grande relativo a los residuos
↳ El tratamiento si está haciendo efecto.

Como medida de tamaño de la contribuº se calculan las sumas de cuadrados de las obs SS.

- $SS_{obs} = SS_{total} = \sum_{l=1}^{g}\sum_{j=1}^{n_l} x_{lj}^2$
- $SS_{mean} = \sum_{l=1}^{g}\sum_{j=1}^{n_l} \bar{x}^2 = \left(\sum_{l=1}^{g} n_l\right)\bar{x}^2$. $\hat{\theta}$ → estimaº.
- $SS_{treatm} = \sum\sum \hat{\tau}_l^2 = \sum\sum(\bar{x}_l - \bar{x})^2$
- $SS_{residue} = \sum\sum \hat{e}_{lj}^2 = \sum\sum(x_{lj} - \bar{x}_l)^2$

- SScorrected = SSobs − SSmean = SS treatm + SS residue

Sea $n = n_1 + \dots + n_l$ (tamaño TOTAL muestra)

Source of variance	SS	gl
Treatment	SS treatm.	$g-1$
Residual	SS residue	$n-g$
Total (corrected	SS corrected	$n-1$

$$F = \frac{SS\ treatm/(g-1)}{SS\ residue/(n-g)} \sim F_{(g-1),(n-g)}(\alpha)$$

Si F grande → rechazo Ho
↳ la fuente de variabilidad de los tratam. es mayor a los residuos.

Ej: Pob 1: 9, 6, 9
Pob 2: 0, 2
Pob 3: 3, 1, 2

$$\begin{array}{l} \text{Pob 1} \\ \text{Pob 2} \\ \text{Pob 3} \end{array} \begin{bmatrix} 9 & 6 & 9 \\ 0 & 2 & \\ 3 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & \\ 4 & 4 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 4 & 4 \\ -3 & -3 & \\ -2 & -2 & -2 \end{bmatrix} + \begin{bmatrix} 1 & -2 & 1 \\ -1 & 1 & \\ 1 & -1 & 0 \end{bmatrix}$$

$SStr = 78$ | $g - 1 = 3 - 1 = 2$
$SSres = 10$ | $n_l - g = (3+2+3) - 3 = 5$.

$$F = \frac{SStr/(g-1)}{SSres/(n_l-g)} = 19.5 > F_{2,5}(0.01) = 13.27$$

↳ El tratam. si tiene efecto
⇒ Rechazo $Ho: \tau_1 = \tau_2 = \tau_3 = 0$

$$\begin{bmatrix} 9 & 0 & 3 \\ 6 & 2 & 1 \\ 9 & \backslash & 2 \end{bmatrix}$$

8 1 2

$$4 \simeq \frac{32}{8} = 4$$

$$\bar{x}_l - \bar{x} \simeq \begin{bmatrix} 4 & -3 & -2 \\ \vdots & \vdots & \vdots \\ \vdots & \backslash & \cdot \end{bmatrix}$$

$$x_{lj} - \bar{x}_l \simeq \begin{bmatrix} 1 & -1 & 1 \\ -2 & 1 & -1 \\ 1 & x & 0 \end{bmatrix}$$

Comparaº de medias con ANOVA.

Ej: Pob 1 = 9, 6, 9
Pob 2 = 0, 2
Pob 3 = 3, 1, 2

$$\bar{X} = \frac{9+6+9+0+2+3+}{8}$$

$\bar{X} = 4.$
$\bar{X}_1 = 24/3 = 8$
$\bar{X}_2 = 2/2 = 1$
$\bar{X}_3 = 6/3 = 2$

$$x_{lj} - \bar{x}_l = \begin{bmatrix} 1 & -2 \\ -1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$F = \frac{SS\ treatment/((}{SS\ residue/(n-g)}$$

$SStreat = \sum\sum(\bar{x}_l -$

$SS\ residue = \sum\sum$

$$F = \frac{78/2}{10/5} = \frac{3}{2}$$

$F_{g-1,n-g}(0.$

Como $F > F_{g-1,}$

de gl haya ig

$SStr = (\bar{x}_i - \bar{\bar{x}})^2$

$= 3(16) + 2(9) + 12$

$= 78$

$SSres = (x_{ij} - \bar{x}_i)^2$

$= 6(1) + 4 = 10$