MANOVA y regresión lineal

Wednesday, March 9, 2022 7:23 AM

MANOVA y regresión lineal

Voy a comparar m vec. de medias

$x_{lj} = \mu_l + \tau_l + \varepsilon_{lj}$ (todos vecs.) → obs j de la pobla° l

τ_l → efecto del tratamiento en l.

ε_{lj} indep $N_p(0, \Sigma)$

Cada componente de x_{lj} satisface el modelo univariado.

$$x_{lj} = \bar{x} + (\bar{x}_l - \bar{x}) + (x_{lj} - \bar{x}_l)$$

μ τ_l $\hat{\varepsilon}_{lj}$

Hipótesis:

$H_0: \tau_1 = \dots = \tau_g = 0$ (no hay efecto en ninguna pobla°.

Fuente vano°	SSP	gl	Wilk Λ
Tratamiento	$B = \sum_{l=1}^{g} n_l (\bar{x}_l - \bar{x})(\bar{x}_l - \bar{x})'$	$g-1$	$\lvert W\rvert / \lvert T\rvert$
Residual	$W = \sum_{l=1}^{g} \sum_{j=1}^{n_l} (x_{jl} - \bar{x}_l)(x_{jl} - \bar{x}_l)'$	$n-g$	
Total	$T = W + B$	$n-1$	

$$\Lambda^* = \frac{|SSCP_e|}{|SSCP_e + SSCP_h|} \qquad \Lambda^* = \frac{|W|}{|B+W|}$$

Si > F , rechazo H_0.

Intervalos de C. simultáneos.

Hay p vars

Hay $\frac{g(g-1)}{2}$ diferencias

En Bonferroni:

$$t_{n-g}\left(\frac{\alpha}{2m}\right) \quad \text{con } m = \frac{pg(g-1)}{2}$$

Sean $n = \sum_{l=1}^{g} n_l$ con confianza al menos $1-\alpha$

$\tau_{li} - \tau_{ki}$ pertenece a:

$$\bar{x}_{li} - \bar{x}_{ki} \pm t_{n-g}\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{w_{ii}}{n-g}\left(\frac{1}{n_l} + \frac{1}{n_k}\right)}$$

Con τ_{li} el efecto del tratam l en la var i

Regresión lineal → método de predic°

$Y = \beta_0 + \beta_1 x$

$Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \dots$ (varias vars.)

Método estadístico para predecir vals de una o varias vars respuesta, dados los vals de una o varias vars predictoras.

Regr lin clásicas:

Sean $z_1, \dots, z_r$, r vars predictoras.

Sea Y una var respuesta (presuntam relacionada)

ε_j: Y val en el mercado

z_1 = Zona, z_2 = Área z_3 = Antigüedad ...

Hay q' limpiar los datos

El modelo clásico, supone q' Y se compone de media q' depende de las z_i + un error aleatorio e.

Obs:

1. Los vals de las vars indep (z_i) se toman fijos.
2. La v.a respuesta y el error son v.a

Tenemos:

$$Y = \beta_0 + \beta_1 z_1 + \dots + \beta_r z_r + \underbrace{\varepsilon}_{\text{error}}.$$

Si hay n obs indep de Y asoc. a respectivos vals de c/

z_i, entonces:

$$y_1 = \beta_0 + \beta_1 z_{11} + \cdots + \beta_r z_{1r} + \varepsilon_1$$
$$\vdots$$
$$y_n = \beta_0 + \beta_1 z_{n1} + \cdots + \beta_r z_{nr} + \varepsilon_n$$

z_{ab}: val que toma en la obs a la var b.

Se asume

1. $E[\varepsilon_i] = 0$
2. $\text{var}(\varepsilon_i) = \sigma^2 \rightarrow$ cte
3. $\text{cov}(\varepsilon_j, \varepsilon_k) = 0 \quad \forall j \neq k$ (son indep)

$E[y_i] = \beta_0 + \beta_1 z_{i1} + \cdots + \beta_r z_r$

Matricial:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & \cdots & z_{1r} \\ \vdots & \vdots & & \vdots \\ 1 & z_{n1} & & z_{nr} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \rightarrow \underset{n\times 1}{Y} = \underset{n\times(r+1)}{Z} \underset{(r+1)\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

- $E[\varepsilon] = 0$
- $\text{cov}(\varepsilon) = \sigma^2 I \rightarrow$ las cov son 0.

- C/ col de Z son las n obs de las vars predictoras
- C/ fila de Z son los vals de las vars predictoras en la obs j.

Estimador de mínimos cuadrados

Construir estimaciones de β ($\beta_0 + \cdots + \beta_r$) usando Z y Y (datos) para hacer predicciones.

Sea b vals de β, las clfs:

$$y_i - b_0 - b_1 z_{i1} - \cdots - b_r z_{ir}$$

y_i: val observado en y; $b_0 + \cdots + b_r z_{ir}$: predic° si b es correcto.

Busco b que minimice las sumas cuadráticas.

$$S(b) = \sum_{i=1}^{n} (y_i - b_0 - \cdots - b_r z_{ir})^2 = (Y - Zb)'(Y - Zb) = \text{escalar}$$

Las entradas de b q' minimizan $S(b)$ se obtienen con mínimos cuadrados (estima° de min cuadrados de β) y se denotan $\hat{\beta}$.

obs:

1. $\hat{\beta}$ son los vals. "+ consistentes" con rela° lineal de Y y Z.
2. $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 z_{i1} - \cdots - \hat{\beta}_r z_{ir} \quad i = 1, \ldots, n$

   ↳ residuos.

Teo:

Sea Z con rango completo
El estimador de mínimos cuadrados de β está dado por:

$$\hat{\beta} = (Z'Z)^{-1} Z' Y$$

$\hat{Y} = Z\hat{\beta}$ → con base en esto, luego estimo el error

$\hat{Y} = Z\hat{\beta} = HY$ con $H = Z(Z'Z)^{-1}Z'$

$\hat{\varepsilon} = Y - \hat{Y} = Y - Z\hat{\beta} = Y - HY$

$= (I - Z(Z'Z)^{-1}Z')Y = (I - H)Y$