Modelos Lineares e Aplicações

Formulário – 2021/2022

Genéricos

$Cov[aX+b,Y]=aCov[X,Y]$

$S_{XX}=\sum_{i=1}^{n}(x_i-\overline{x})^2=\sum_{i=1}^{n}x_i^2-n\overline{x}^2$ $\qquad S_{YY}=\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}y_i^2-n\overline{y}^2$

$S_{XY}=\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\sum_{i=1}^{n}x_iy_i-n\overline{x}\,\overline{y}$

Coeficientes de correlação de Pearson: $r_{xy}=\dfrac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$

$H_0:\rho=0$ vs $H_1:\rho\neq 0$ sendo $ET=\frac{\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}\times r_{xy}\sim t_{n-2}$ e $RC=\{t:|t|>t_{\frac{\alpha}{2};n-2}\}$

Coeficientes de correlação de Spearman: $r_S=1-\frac{6\sum_{i=1}^{n}d_i^2}{n(n^2-1)}$ em que d_i são as diferenças entre as ordens de x_i e y_i

Regressão linear simples

$Y_i=\beta_0+\beta_1x_i+\epsilon_i$

$\hat{\beta}_1=\frac{S_{XY}}{S_{XX}}$ $\qquad \hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{x}$ $\qquad SSE=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$

$H_0:\beta_1=0$ vs $H_1:\beta_1\neq 0$ sendo $ET=\dfrac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2/S_{XX}}}\sim t_{n-2}$ e $RC=\{t:|t|>t_{\frac{\alpha}{2};n-2}\}$

$H_0:\beta_0=0$ vs $H_1:\beta_0\neq 0$ sendo $ET=\dfrac{\hat{\beta}_0}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{\overline{x}^2}{S_{XX}}\right)}}\sim t_{n-2}$ e $RC=\{t:|t|>t_{\frac{\alpha}{2};n-2}\}$

onde $\hat{\sigma}^2=\dfrac{1}{n-2}\left(S_{YY}-\dfrac{S_{XY}^2}{S_{XX}}\right)=\dfrac{SSE}{n-2}$

Coeficientes de determinação: $R^2=\dfrac{S_{XY}^2}{S_{XX}S_{YY}}=r_{xy}^2$

Intervalo de Confiança a 100(1-α)% para $E[Y_0]$: $\hat{E}[Y_0]\pm t_{\frac{\alpha}{2};n-2}\times\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{(\overline{x}-x_0)^2}{S_{XX}}\right)}$

onde $\hat{E}[Y_0]=E[Y|x=x_0]=\hat{\beta}_0+\hat{\beta}_1x_0$

Intervalo de Predição a 100(1-α)% para Y_0: $\hat{Y}_0\pm t_{\frac{\alpha}{2};n-2}\times\sqrt{\hat{\sigma}^2\left(1+\frac{1}{n}+\frac{(\overline{x}-x_0)^2}{S_{XX}}\right)}$

onde $\hat{Y}_0=\hat{\beta}_0+\hat{\beta}_1x_0$

ANOVA: SST=SSR+SSE $\iff$ $\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$

$H_0:\beta_1=\cdots=\beta_p=0$ vs $H_1:\exists\beta_j\neq 0,\ j=1\ldots p$ sendo $ET=\dfrac{MSR}{MSE}\sim F_{(p,n-p-1)}$ e

$RC=\{F:F>F_{\alpha;p,n-p-1}\}$ onde $MSR=\frac{SSR}{p}$ e $MSE=\frac{SSE}{n-p-1}$

Análise de diagnóstico. Observações discordantes.

- outlier, se $|r_i| \geq 3$, onde $r_i \sim$ T_student
- ponto com elevado "leverage", se $|h_{ii}| > \frac{2(p+1)}{n}$
- ponto influente, se distância de Cook $|c_i| > F(0.5, p+1, n-p-1)$ ou $|\text{dffits}_i| > 2\sqrt{\frac{p+1}{n-p-1}}$

Regressão linear múltipla

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \qquad \Longleftrightarrow \qquad Y = X\beta + \epsilon \quad \text{(notação matricial)}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad \Sigma_\beta = \hat{\sigma}^2 (X^T X)^{-1} \quad \text{onde} \quad \hat{\sigma}^2 = \frac{\left(Y-\hat{Y}\right)^T \left(Y-\hat{Y}\right)}{n-p-1} = \frac{SSE}{n-p-1} \quad \text{e} \quad \hat{Y} = X\hat{\beta}$$

$$H_0 : \beta_j = b_j \ \text{ vs } \ H_1 : \beta_j \neq b_j \qquad \text{sendo} \qquad ET = \frac{\hat{\beta}_j - b_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t_{n-p-1} \quad \text{e} \quad RC = \{t : |t| > t_{\frac{\alpha}{2};n-p-1}\}$$

onde C_{jj} é o j-ésimo elemento da diagonal principal da matriz $C = (X^T X)^{-1}$

Coeficientes de determinação: $\quad R^2 = 1 - \dfrac{SSE}{SST} = \dfrac{SSR}{SST}$

Coeficientes de determinação ajustado: $\quad R_a^2 = 1 - \dfrac{n-1}{n-p-1}(1 - R^2)$

Intervalo de Confiança a 100(1-α)% para $E[Y_0]$: $\quad \hat{E}[Y_0] \pm t_{\frac{\alpha}{2};n-p-1} \times \sqrt{\hat{\sigma}^2 \left(\vec{x_0}^T C \vec{x_0}\right)}$

onde $\hat{E}[Y_0] = E[Y|x = \vec{x_0}] = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \cdots + \hat{\beta}_p x_{0p} = \vec{x_0}^T \hat{\beta}$

Intervalo de Predição a 100(1-α)% para Y_0 : $\quad \hat{Y}_0 \pm t_{\frac{\alpha}{2};n-p-1} \times \sqrt{\hat{\sigma}^2 \left(1 + \vec{x_0}^T C \vec{x_0}\right)}$

onde $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \cdots + \hat{\beta}_p x_{0p} = \vec{x_0}^T \hat{\beta}$

Teste de Fisher–parcial

H_0 : r parâmetros são nulos ($r < p$) $\qquad$ vs $\qquad H_1 : Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i$

$$\text{sendo} \qquad ET = F_0 = \frac{n-p-1}{r} \times \frac{SSE(H_0) - SSE(H_1)}{SSE(H_1)} \sim F_{(r,n-p-1)} \quad \text{e} \quad RC = \{F : F > F_{\alpha;r,n-p-1}\}$$