Teoria da Regressão

$2^{\underline{a}}$ Lista de Exercícios

1. Suponha que as variáveis x_i e y_i estejam relacionadas de acordo com o modelo de regressão linear

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \ldots, n$$

em que os ϵ_i representam erros aleatórios independentes e identicamente distribuidos com média zero e variância σ^2. Obtenha os estimadores de mínimos quadrados α e β e suas respectivas variâncias. Se você tem interesse em testar que $H_0 : \alpha = 0$, que suposição adicional terá que ser feita ?

2 Considere as seguintes funções das variáveis aleatórias Y_1, Y_2, Y_3 e Y_4:

$$\begin{aligned} W_1 &= \tfrac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4) \\ W_2 &= \tfrac{1}{2}(Y_1 + Y_2) - \tfrac{1}{2}(Y_2 + Y_4) \end{aligned}$$

(a) Expresse as relações acima em notação matricial

(b) Obtenha as esperança do vetor $\mathbf{W} = (\mathbf{W_1}, \mathbf{W_2})^{\mathbf{t}}$ em termos das esperança de Y_1, Y_2, Y_3 e Y_4.

(c) Obtenha a matriz de covariâncias do vetor $\mathbf{W}$ em termos das covariâncias de Y_1, Y_2, Y_3 e Y_4.

3. Considere o modelo linear de posto completo

$$\mathbf{y}_{(n\times 1)} = \mathbf{1}\beta_0 + \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim N(0, \sigma^2\mathbf{I})$$

em que $\mathbf{y} \in I\!R^n$ e $\boldsymbol{\beta}_1 \in I\!R^{p-1}$

Mostre que

(a) Os estimadores de mínimos quadrados de β_0 e β_1 são dados por :
$\hat{\beta}_0 = \bar{y} - \bar{\mathbf{x}}^t\hat{\boldsymbol{\beta}}_1$ e $\hat{\boldsymbol{\beta}}_1 = (\mathbb{X}^t\mathbb{X})^{-1}\mathbb{X}^t\mathbf{y}$ onde $\bar{\mathbf{x}}^t = \frac{1}{n}\mathbf{1}^t\mathbf{X}_1$ e $\mathbb{X} = \mathbf{X}_1 - \mathbf{1}\bar{\mathbf{x}}^t$.

(b) $\mathbf{1}^T\hat{\boldsymbol{\epsilon}} = 0$ onde $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$

(c) $\dfrac{SQRes}{\sigma^2} \sim \chi^2_{(n-p)}$.

(d) $\dfrac{SQReg}{\sigma^2} \sim \chi^2_{(p, \frac{1}{2\sigma^2}\boldsymbol{\beta}^t\mathbf{X}^t\mathbf{X}\boldsymbol{\beta})}$.

(e) $\dfrac{SQM}{\sigma^2} \sim \chi^2_{(1, \frac{1}{2n\sigma^2}(\mathbf{1}^t\mathbf{X}\boldsymbol{\beta})^2)}$.

(f) $\dfrac{SQReg_M}{\sigma^2} \sim \chi^2_{(p-1, \frac{1}{2\sigma^2}\boldsymbol{\beta}_1^t\mathbb{X}^t\mathbb{X}\boldsymbol{\beta}_1)}$.

2. Como ficariam os itens (a) e (b) acima se o modelo fosse sem o intecepto, isto é $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$?.

3. Derive a distribuição conjunta de $\hat{\beta}_0$ e $\hat{\boldsymbol{\beta}}_1$

4. Faça uma decomposição de SQReg em duas componentes SQ1 e SQ2 de modo que :

(a) SQReg=SQ1+SQ2

(b) SQ1 e SQ2 e SQRes independentes duas a duas

(c) $\frac{SQ1}{\sigma^2} \sim \chi^2_{(1,\lambda_1)}$ com $\lambda_1 = 0 \Leftrightarrow \beta_0 = 0$. Que hipótese a estatística $\frac{SQ2}{\sigma^2}$ poderia ser utilizada para testar ?

5.(Narula e Stangenhaus, 1988, p. 32, Paula, 2004) No arquivo imoveis.dat são apresentados dados relativos a uma amostra de 27 imóveis. Na ordem são apresentados os valores das seguintes variáveis: (i) imposto do imóvel (em 100 dolares), (ii) área do terreno (em 1000 pés quadrados), (iii) área construída (em 1000 pés quadrados), (iv) idade da residência (em anos) e (v) preço de venda do imóvel (em 1000 dolares).

(a) Construa os gráficos de dispersão convinientes.

(b) Ajuste um modelo normal linear do preço de venda contra as demais variáveis selecionando as variáveis explicativas que contribuem significamente para o modelo. Interprete os diferentes parâmetros do modelo.

(c) estime os parâmetros do modelo e apresente os respectivos erros padrões.

(d) avalie a qualidade do ajuste do modelo

(e) Construa o intervalo de confiança para preço médio do imóvel segundo um conjunto de valores das explicativas a sua escolha.

(f) Apresente um conclusão que evite o jargão estatístico.