

Universidade Federal do Ceará

Departamento de Estatística e Matemática Aplicada

Prof. : Juvêncio S. Nobre

CC0290 - Modelos de Regressão I - 2022.2

Lista de exercícios # I

Distribuição: 13/09/2022

1. Considere  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^\top$  o estimador de mínimos quadrados em um modelo de regressão linear simples. Defina o resíduo da  $i$ -ésima observação como

$$\hat{e}_i := y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, \quad i = 1, \dots, n.$$

Mostre que

i)  $\mathbb{E}[\hat{y}_i] = \mu_{x_i} = \beta_0 + \beta_1 x_i.$

ii)  $\text{Var}[\hat{y}_i] = \sigma^2 h_{ii},$  em que

$$h_{ii} := \left\{ \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right\}.$$

iii)  $\text{Cov}(y_i, \hat{y}_i) = \sigma^2 h_{ii}.$

iv)  $\mathbb{E}[\hat{e}_i] = 0.$

v)  $\text{Var}[\hat{e}_i] = \sigma^2(1 - h_{ii}).$

vi)  $\text{Cov}(\hat{e}_i, \hat{e}_j) = -\sigma^2 h_{ij},$  em que

$$h_{ij} := \left\{ \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}} \right\}.$$

2. Utilizando o método de **multiplicadores de Lagrange**, prove o teorema de Gauss-Markov.

3. Considere o MRLS e todas as suas suposições. Discuta sobre a homoscedasticidade dos resíduos ordinários. Considerando o item v) da Questão 1, argumente em que situações a variância do  $i$ -ésimo resíduo é máxima e mínima, respectivamente.

4. Considerando adicionalmente que  $e_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ , determine a distribuição conjunta dos EMQ  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^\top$  e do vetor de resíduos  $\hat{\mathbf{e}}$ .

5. Considerando a suposição de normalidade, obtenha um intervalo de confiança de nível  $1 - \alpha$  para  $\sigma^2$  usando o método da quantidade pivotal. Discuta como poderia ser obtido o intervalo de comprimento mínimo.

6. Considerando a suposição de normalidade, obtenha os EMV dos parâmetros do MRLS. Calcule o viés do EMV de  $\sigma^2$  e mostre que ele converge para zero quando  $n \rightarrow \infty$ .
7. Considere o MRLS e todas as suas suposições, inclusive a de normalidade. Mostre que o EMQ de  $\beta$ ,  $\hat{\beta}$ , e SQRes são independentes.
8. Considere o MRLS e todas as suas suposições, mostre que

$$\begin{aligned}\mathbb{E}[\text{SQT}] &= (n-1)\sigma^2 + \beta_1^2 S_{xx} \\ \mathbb{E}[\text{SQReg}] &= \sigma^2 + \beta_1^2 S_{xx} \\ \mathbb{E}[\text{SQRes}] &= (n-2)\sigma^2.\end{aligned}$$

9. Considere o MRLS e todas as suas suposições, inclusive a de normalidade. Mostre que sob  $\mathcal{H}_0 : \beta_1 = 0$ , temos que

$$\begin{aligned}\frac{\text{SQT}}{\sigma^2} &\sim \chi_{(n-1)}^2 \\ \frac{\text{SQRes}}{\sigma^2} &\sim \chi_{(n-2)}^2 \\ \frac{\text{SQReg}}{\sigma^2} &\sim \chi_{(1)}^2,\end{aligned}$$

com SQRes e SQReg independentes.

10. Considere o MRLS e todas as suas suposições, inclusive a de normalidade. Mostre que o MINQUE de  $\sigma^2$  (QMRes) é um estimador fracamente consistente.
11. Suponha que desejamos ajustar um MRLS. A *região de interesse* é tal que  $-1 \leq x_i \leq 1$ ,  $i = 1, \dots, n$ . Discuta como as observações devem ser escolhidas de forma a obter o menor erro-padrão possível para  $\hat{\beta}_1$ .
12. Considere o MRLS com **intercepto nulo** e todas as suas suposições, inclusive a de normalidade.

- i) Podemos utilizar a mesma expressão do coeficiente de determinação usada no MRLS usual? Discuta.
- ii) Obtenha a decomposição de ANOVA e obtenha os valores esperados das somas de quadrados correspondentes.
- iii) Mostre que os parâmetros são ortogonais.
- iv) Obtenha a distribuição do EMQ de  $\beta_1$ ,  $\hat{\beta}_1$  e do MINQUE de  $\sigma^2$ .

- v) Prove o teorema de Gauss-Markov para o MRLS de intercepto nulo, i.e., prove que  $\hat{\beta}_1$  é o BLUE de  $\beta_1$ .
- vi) Mostre que  $\frac{\text{SQRes}}{\sigma^2} \sim \chi_{(n-1)}^2$  e que  $\hat{\beta}_1 \perp \text{SQRes}$ .
- vii) Obtenha o IC de comprimento mínimo ao nível de  $(1 - \alpha)$  para  $\beta_1$ , e  $\mu_{x_0} := \beta_1 x_0$ .
- viii) Obtenha o intervalo de predição ao nível de  $(1 - \alpha)$  para  $Y_{x_0}$ .
- ix) Baseado na decomposição em ii), obtenha a tabela de ANOVA para o MRLS com intercepto nulo.
- x) Obtenha uma estatística de teste para testar  $\mathcal{H}_0 : \beta_1 = a$  vs.  $\mathcal{H}_1 : \beta_1 \neq a$ ,  $a \in \mathbb{R}$  especificado.

**13.** Considere o MRLS com **intercepto nulo** e todas as suas suposições. Prove que o EMQ de  $\beta_1$ ,  $\hat{\beta}_1$  e o MINQUE de  $\sigma^2$  são estimadores fracamente consistentes.

**14.** Considere um MRLS e todas as suas suposições. Admita que o intercepto  $\beta_0$  é conhecido.

- i) Determine o EMQ de  $\beta_1$ .
- ii) Obtenha o valor esperado e a variância do estimador obtido no item i).
- iii) O estimador obtido no item i) é o BLUE de  $\beta_1$ ?
- iv) Admitindo normalidade, obtenha a distribuição do EMQ de  $\beta_1$ .

**15.** Considere a seguinte amostra aleatória com 10 observações:

x	y
0	2,5; 3,5
1	1;3
2	2;4
3	0;2
4	0,5 ;1,5

Admite-se que as variáveis  $x$  e  $y$  estão relacionadas de acordo com o MRLS  $y_i = \beta_0 + \beta_1 x_i + e_i$ , com  $e_1, \dots, e_{10}$  representando uma sequência de fontes de variação homoscedásticas com média zero e distribuição normal.

- i) Desenhe o diagrama de dispersão e comente sobre a adequabilidade empírica do MRLS.

- ii) Determine a reta de regressão estimada pelo método de mínimos quadrados e o coeficiente de determinação.
- iii) Considere que a equação obtida é a demanda de rapaduras na cidade de Fortaleza, sendo  $x$  o preço (em reais) e  $y$  a quantidade vendida. Com base nisso, interprete as quantidades obtidas no item anterior.
- iv) Se o produto é vendido na *bodega* do seu *Zezinho* e seu custo médio de produção é constante e igual a R\$ 2, a que preço seu *Zezinho* deve vender as rapaduras para maximizar a sua renda líquida?

**16.** Os dados abaixo referem-se ao faturamento de empresas similares de um mesmo setor industrial nos últimos 15 meses.

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan	Fev	Mar
Vendas	1.0	1.6	1.8	2.0	1.8	2.2	3.6	3.4	3.3	3.7	4.0	6.4	5.7	6.0	6.8

Utilize técnicas de análise de regressão para quantificar o crescimento do faturamento de empresas desse setor ao longo do período observado. Com essa finalidade:

- i) Proponha um modelo adequado, interpretando todos os parâmetros e especificando as suposições.
- ii) Estime os parâmetros do modelo e apresente os resultados numa linguagem não técnica.

**17.** O custo da produção de um lote de uma certa peça depende do número de peças produzidas, ou seja, do tamanho do lote. Em uma amostra de dez lotes, observou-se os seguintes resultados:

Tamanho	5	10	15	20	25	30	35	40	45	50
Custo (R\$)	65	120	210	260	380	450	510	555	615	660

- i) Faça um diagrama de dispersão e comente.
- ii) Ajuste um MRLS para estes dados e interprete os resultados.
- iii) Apresente o quadro de ANOVA e interprete-o.

**18.** Considere o seguinte conjunto de dados construído por Anscombe (1973, Graphs in statistical analysis, *The American Statistician*), dispostos na Tabela 1, que ilustra a importância da utilização das técnicas de diagnóstico.

- i) Faça um diagrama de dispersão, para cada conjunto de dados, e comente.
- ii) Ajuste um MRLS para estes dados e interprete os resultados.
- iii) Usando o coeficiente de determinação, você concluiria que o melhor ajuste é obtido para qual conjunto de dados?
- iv) Baseado nos itens i) e iii) comente sobre a fragilidade do coeficiente de determinação para avaliar a qualidade de ajuste.

Tabela 1: Conjunto de dados Anscombe (1973)

A		B		C		D	
<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>
10,00	8,04	10,00	9,14	10,00	7,46	8,00	6,58
8,00	6,95	8,00	8,14	8,00	6,77	8,00	5,76
13,00	7,58	13,00	8,74	13,00	12,74	8,00	7,71
9,00	8,81	9,00	8,77	9,00	7,11	8,00	8,84
11,00	8,33	11,00	9,26	11,00	7,81	8,00	8,47
14,00	9,96	14,00	8,10	14,00	8,84	8,00	7,04
6,00	7,24	6,00	6,13	6,00	6,08	8,00	5,25
4,00	4,26	4,00	3,10	4,00	5,39	19,00	12,50
12,00	10,84	12,00	9,13	12,00	8,15	8,00	5,56
7,00	4,82	7,00	7,26	7,00	6,42	8,00	7,91
5,00	5,68	5,00	4,74	5,00	5,73	8,00	6,89

**19.** Considere o modelo

$$y_i = x_i^\beta e_i, \quad i = 1, \dots, n.$$

Que transformação dos dados e que suposição sobre os *erros/fonte de variação* você faria para ajustar esse modelo por meio da metodologia de regressão linear simples? Quem é o EMQ de  $\beta$ ?

**20.** Considere  $\beta_{yx}$  ( $\beta_{xy}$ ) a inclinação referente ao MRLS de  $y$  versus  $x$  ( $x$  versus  $y$ ). Mostre que os EMQ são tais que  $\hat{\beta}_{yx}\hat{\beta}_{xy} = r_{xy}^2$ , em que  $r_{xy}$  representa o coeficiente de correlação linear de Pearson entre  $x$  e  $y$ .

**21.** (Gujarati, 2000) A tabela abaixo fornece dados sobre a taxa de demissão por 100 empregados ( $y$ ) e a taxa de desemprego ( $x$ ) na indústria dos EUA, no período de 1960-1972.

**Obs.:** A taxa de demissão se refere às pessoas que se afastaram de seu emprego voluntariamente.

Tabela 2: Taxas de demissão e de desemprego na indústria americana, 1960 -1972.

Ano	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
y	1,3	1,2	1,4	1,4	1,5	1,9	2,6	2,3	2,5	2,7	2,1	1,8	2,2
x	6,2	7,8	5,8	5,7	5,0	4,0	3,2	3,6	3,3	3,3	5,6	6,8	5,6

- i) Faça um diagrama de dispersão e comente.
  - ii) Ajuste um MRLS para estes dados e interprete os resultados.
  - iii) Calcule o coeficiente de regressão linear e o coeficiente de determinação.
  - iv) Utilizando somente os dados do período 1966-1978, foi ajustado um MRLS, obtendo:  $\hat{y}_i = 3,1237 - 0,1714x_i$ , com  $\widehat{EP}(\hat{\beta}_1) = 0,0210$  e  $r^2 = 0,8575$ . Se estes resultados são diferentes daqueles que você obteve em ii), como você explicaria a diferença?
- 22.** Faça todas as questões do Cap.2 do livro do Hoffman (2006) e as questões 2.7, 2.13 e 2.15 do Livro do Montgomery et al. (2012).