MAE328 - Análise de Regressão
2ª Lista de Exercícios

1. Um pesquisador deseja verificar se um instrumento para medir a concentração de ácido lático no sangue está bem calibrado. Para iso ele tomou 20 amostras de concentrações conhecidas e determinou a respectiva concentração através do instrumento. Como uma análise de regressão poderia auxiliar o pesquisador? Modele o problema acima, especificando as variáveis independente e dependente, e as hipóteses de interesse.

2. Encontre os estimadores de mínimos quadrados para o modelo de 2º grau:
$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon.$$

3. Sejam

X	0	1,5	2	2,5	3	3,5
Y	0,1	0,3	0,9	2,2	4	6

   a) Construa o modelo $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ e determine o coeficiente de explicação.
   b) Ajuste uma parábola aos dados e escreva a ANOVA.
   c) Determine o coeficiente de explicação para o modelo do item b) e compare com o obtido no item a). Justifique a diferença entre os dois valores.

4. No modelo $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, com $\epsilon_i \sim N(0, \sigma^2)$, ϵ_i e ϵ_j independentes para $i \neq j$, suponha que $\beta_0 = 210$, $\beta_1 = 4$ e $\sigma = 5$.
   a) Qual a distribuição de y dado $x = 10, 20, 40$?
   b) Construa um desenho que esquematize este modelo.

5. Acredita-se que a umidade de um produto influencia a densidade final do produto. Num experimento, a umidade foi controlada e a densidade final foi medida resultando os seguintes dados (codificados):

umidade (x)	4,7	5,0	5,2	5,2	5,9	4,7	5,9	5,2	5,2	5,3	5,9	5,6	5,0
densidade (y)	3	4	4	5	10	2	9	3	3	7	6	6	4

   a) Se adotarmos o modelo $y = \beta_0 + \beta_1 x + \epsilon$ qual o significado prático de β_1?
   b) Ajuste o modelo $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ e construa a ANOVA.
   c) Teste as hipóteses $H_0 : \beta_1 = 0$, $H_a : \beta_1 \neq 0$. Que modelo você adotaria para estes dados?

6. Com base em 11 pares de valores das variáveis X e Y obteve-se a equação de regressão $y = 20 - x$, com $R^2 = 0,64$. Se a estimativa não viciada de σ^2 é 4, teste ao nível $\alpha = 0,05$ a hipótese $H_0 : \beta_1 = 0$ contra a alternativa $H_a : \beta_1 \neq 0$.

7. Dado um conjunto de pares de valores X_{ij}, Y_{ij}, $(i = 1, \ldots, m, j = 1, \ldots, n_i)$ ajusta-se um conjunto de retas de m retas paralelas
$$\hat{y}_{ij} = \hat{a}_i + \hat{b}\, x_{ij}$$
Mostre que as estimativas dos parâmetros, de acordo com o método dos mínimos quadrados são dadas por
$$\hat{b} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)}{\sum_{i=1}^{n}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

e

$$\hat{a}_i = \bar{y}_i - b\bar{x}_i$$

em que

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} \quad \text{e} \quad \bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$$

8. Prove que a reta de regressão $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ passa pelo ponto $(\bar{x}, \bar{y})$.

9. Prove que se $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ e $e_i = y_i - \hat{y}_i$; $i = 1, 2, \ldots, n$ então:

   a) $\displaystyle\sum_{i=1}^{n} e_i = 0$,

   b) $\displaystyle\sum_{i=1}^{n} x_i e_i = 0$,

   c) $\displaystyle\sum_{i=1}^{n} \hat{y}_i e_i = 0$,

   d) $\displaystyle\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$,

10. Com base em n observações (x_{ij}, y_{ij}), encontre os estimadores de mínimos quadrados para o modelo:

$$E(y|x) = \beta_0 + \beta_1 \frac{1}{x}$$

e calcule a variância de $\hat{\beta}_1$.

11. Adotou-se o modelo $E(y|x) = \beta x$, na situação em que o verdadeiro modelo era $E(y|x) = \beta_0 + \beta_1 x$. Prove que o estimador de mínimos quadrados de β é viciado para β_1 e determine o seu vício.

12. Com base numa amostra de n observações da variável y, determine o estimador de mínimos quadrados de β_0 no modelo $y_i = \beta_0 + \epsilon_i$. A qual característica populacional corresponde β_0 neste modelo?

13. O que você acha da afirmação: "Para que o método de mínimos quadrados seja válido, é necessário que a distribuição da variável seja normal"?

14. No modelo de regressão linear simples, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, sejam $F = \frac{\text{SSR}}{\text{MSE}}$ e R^2 - coeficiente de explicação do modelo.

   a) Prove que

$$F = \frac{(n-2)R^2}{1-R^2}.$$

   b) Analise o comportamento de F para os possíveis valores de R^2.

15. Um pesquisador deseja estimar o nível de altitude acima do mar através do ponto de ebulição da água. Ele sabia que a altitude poderia ser determinada através da pressão atmosférica, medida através de um barômetro. No seu experimento, realizado em 1850, ele registrou a relação entre pressão e ponto de ebulição. Seu interesse foi motivado pela dificuldade do transporte dos frágeis barômetros da época. As medidas do ponto de ebulição lhe forneceriam uma previsão da pressão e posteriormente uma rápida maneira de prever altitudes.

    Os dados a seguir correspondem a uma amostra de 31 pares de medidas das variáveis TEMP - ponto de ebulição (graus Fahrenheit) e PRES - pressão barométrica (em polegadas de mercúrio).

TEMP	PRES	TEMP	PRES
210,8	29,211	189,5	18,869
210,2	28,559	188,8	18,356
208,4	27,972	188,5	18,507
202,5	24,697	185,7	17,267
200,6	23,726	186,0	17,221
200,1	23,369	185,6	17,062
199,5	23,030	184,1	16,959
197,0	21,892	184,6	16,881
196,4	21,928	184,1	16,817
195,6	21,605	182,4	16,235
193,4	20,480	181,9	16,106
193,6	20,212	181,9	15,928
191,4	19,758	181,0	15,919
191,1	19,490	180,6	15,376
190,6	19,386		

a) Construa o diagrama de dispersão da variável PRES contra TEMP. Uma reta se ajusta bem aos dados?

b) Construa um diagrama da variável $100 \times \log$ PRES contra TEMP e compare com o do item a). (Como os valores de log(PRES) não variam muito, eles foram multiplicados por 100. Isto evita que trabalhemos com valores muito baixos e com pouca variabilidade, sem comprometer os principais aspectos da análise). Este diagrama de dispersão está mais próximo de uma reta que o do item a)?

c) Ajuste o modelo de regressão linear simples de $Y = 100 \times \log(\text{PRES})$ em $X = \text{TEMP}$ e calcule as estatísticas relevantes (estimativas dos parâmetros, testes, tabelas de Análise de Variância, R^2).