

Lista de Exercícios I

Observação 1: Os exercícios propostos devem ser IMPRESSOS e entregues em formato de relatório técnico;

Observação 2: A lista tem valor 1,0 com **data limite: 21/12/2023** para entrega;

Observação 3: Utilize um nível de significância de 5% quando não foi enunciado e todas análises devem conter interpretações nos, respectivos, contextos enunciados;

Observação 4: Os arquivos de dados, quando enunciados, estarão disponíveis na plataforma Moodle (no tópico Banco de Dados).

1. A reta de regressão obtida pelo método dos mínimos quadrados (MMQ) possui propriedades. Algumas delas necessitam ser demonstradas:

- (a) que

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2;$$

- (b) a soma dos valores observados é igual à soma dos valores estimados, isto é,

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i;$$

- (c) a soma dos resíduos é igual a zero, isto é,

$$\sum_{i=1}^n e_i = 0;$$

- (d) a soma do produto dos valores estimados e resíduos é igual a zero, isto é,

$$\sum_{i=1}^n \hat{y}_i e_i = 0.$$

2. Faça o que se pede:

- (a) Simule $n = 30$ valores para o Modelo de Regressão Linear Simples (MRLS) (*semente 350*) dado por: $Y = -2 + 0,5x + \epsilon$ tal que $X \sim N(0, 1)$ e $\epsilon \sim N(0, \sigma^2 = 9)$;
- (b) Apresente o diagrama de dispersão dos dados;
- (c) Estime os parâmetros do modelo via MMQ e pela forma matricial, isto é, $\hat{\beta} = (X'X)^{-1} X'y$;
- (d) Apresente o histograma dos erros e teste sua normalidade via Shapiro-Wilk.

3. Com o objetivo de se estudar a relação entre Tempo de uma reação química (resposta) e Temperatura, um certo experimento foi realizado. A Tabela 1 a seguir contém os valores das temperaturas, em $^{\circ}C$, e os tempos obtidos, em segundos.

Tabela 1. *Dados experimentais de Temperatura e Tempo.*

Temperatura	Tempo				
20	12,3	11,8	11,5	12,1	11,7
30	11,8	11,5	11,4	11,7	11,2
40	10,9	11,2	10,8	10,6	10,3
50	10,4	9,8	9,5	9,9	9,2
60	9,6	9,0	8,7	8,3	9,1
70	9,1	9,3	8,5	8,6	8,3
80	8,4	8,1	8,1	7,7	7,9

- Apresente o diagrama de dispersão dos dados;
 - Estime a média, desvio-padrão, variância e coeficiente de variação para a resposta (Tempo de reação), considerando-a como normalmente distribuída (Sugestão: Utilize a função *fitdistr* da livreria *MASS* do *R*);
 - Calcule a média, desvio-padrão, variância e coeficiente de variação para a resposta em cada nível de temperatura e responda se há algum indício de heterocedasticidade. Justifique;
 - Estime os parâmetros do MRLS da Temperatura versus Tempo via comando *lm* do *R*;
 - Teste a significância dos parâmetros em nível de 5%;
 - Teste a normalidade nos resíduos;
 - Trace a reta estimada sobre os pontos observados;
 - Construa intervalos de 95% de confiança para os parâmetros estimados;
4. As Tabelas 2 e 3 a seguir, fazem parte da saída de uma Análise de Regressão realizada em um determinado programa estatístico.

Tabela 2. *Saída 1.*

Parameter	Estimate	Std. Error	t value	p
Intercept	83,0740	6,5930	12,60	0,000
<i>x</i>	-1,1848	0,1258	-9,42	0,000

Tabela 3. *Saída 2.*

FV	Df	Sum Sq	Mean Sq	F value	p
<i>x</i>	1	1.021,1	1.021,1	88,68	0,000
Residuals	7	80,6	11,5		
Total	8	1.101,6			

- Escreva a equação da reta ajustada;
- Encontre um intervalo de confiança de 95%, para os coeficientes da reta ajustada;
- Verifique se o modelo contribui para explicar a variável resposta. Está adequado? Justifique;
- Encontre a estimativa da variância residual;

5. Um experimento foi conduzido para avaliar, em coelhos a disponibilidade relativa (DR) do Fósforo existente nos Fosfatos de rocha de Araxá e de Patos de Minas em relação ao Fósforo existente no Fosfato Bicálcico. Os animais foram alimentados com rações contendo níveis crescentes de cada Fosfato e foram anotados os consumos de ração, o que permitiu calcular o consumo de Fósforo em cada unidade experimental. A variável resposta observada foi a resistência do fêmur à quebra, mensurada com dinamômetro (Tabela 4).

A disponibilidade relativa é estimada pela razão entre o coeficiente linear de regressão (angular no caso da reta) obtido para o Fosfato de interesse e para o Fosfato Bicálcico, isto é:

$$DR^* = \text{beta}_1^*/\text{beta}_1^{\text{Bicálcico}}$$

Tabela 4. Resistência à quebra de fêmur de coelhos vs níveis de Fósforo.

Araxá		Patos		Bicálcico	
Y	X	Y	X	Y	X
21,28	0,184	23,50	0,195	25	0,18
31,60	0,350	34,59	0,350	39	0,36
32,42	0,516	31,18	0,505	43	0,50
41,74	0,683	43,27	0,661	60	0,65
42,06	0,849	40,36	0,816	63	0,78
53,38	1,015	50,45	0,972	71	0,94

- Ajuste os modelos individuais;
 - Calcule a disponibilidade relativa do Fósforo nos dois Fosfatos de rocha e decida pelo melhor;
 - Trace as retas ajustadas aos dados simultaneamente.
6. Com base nos dados do Instituto Nacional de Estatística de Portugal (INE), o arquivo **cereais** contém a evolução da superfície agrícola (y) utilizada anualmente na produção de cereais para grão (y : **area**, em km^2) em Portugal, no período de 1986 a 2011 (x : **ano**). Faça o que se pede:
- Construa uma nuvem de pontos de superfície agrícola vs. ano e comente;
 - A partir do gráfico obtido do item anterior, sugira um valor para o coeficiente de correlação entre superfície agrícola e ano. Depois, utilize os comandos do *R* para calcular esse mesmo coeficiente de correlação. Comente o seu significado;
 - Ajuste uma reta de regressão de superfície agrícola utilizada sobre anos. Discuta o significado dos parâmetros da reta ajustada, no contexto do problema sob estudo;
 - Comente a qualidade da reta obtida, calculando o respectivo coeficiente de determinação e interpretando o valor obtido;
 - Trace a reta de regressão ajustada em cima da nuvem de pontos e comente;
 - Calcule a Soma de Quadrados Total (SQT), a partir do cálculo da variância amostral de y ;
 - Calcule o valor da Soma de Quadrados da Regressão (SQReg);
 - Calcule a Soma de Quadrados dos Resíduos (SQE), diretamente a partir dos resíduos, e verifique numericamente a relação fundamental da Regressão Linear: $SQT = SQReg + SQE$;
 - Altere as unidades de medida da variável área, de km^2 para hectares ($area \rightarrow area \times 100$). Ajuste novamente a regressão, após efetuar esta alteração. O que aconteceu aos parâmetros estimados e ao coeficiente de determinação R^2 ? Comente;
 - De novo a partir dos dados originais, transforme a variável ano num contador dos anos do estudo ($ano \rightarrow ano - 1985$). Ajuste novamente a regressão, após efetuar esta alteração. O que aconteceu aos parâmetros estimados e ao coeficiente de determinação R^2 ? Comente.

7. Hsuie, Ma e Tsai (1995) estudam o efeito da **razão** molar do ácido sebácico (o regressor) na **viscosidade** intrínseca dos copoliésteres (a resposta).

- (a) Realize uma Análise de RLS completa e apropriada;
- (b) Apresente um gráfico simultâneo com os dados, o ajuste, o IC e o IP;
- (c) Apresente um intervalo de predição para uma viscosidade de 0,95.

Razão	1	0,9	0,8	0,7	0,6	0,5	0,4	0,3
Viscosidade	0,45	0,20	0,34	0,58	0,70	0,57	0,55	0,75

8. Com interesse de investigar a relação linear existente entre X (nível de dose (%) nutricional na ração) e Y (resposta à taxa de crescimento), um experimento foi realizado considerando um Delineamento Inteiramente Casualizado (DIC) em um grupo de animais. Realize uma análise completa para os dados e conclua.

X	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4
Y	2	2	1	1	0,1	1	0,1	0,1	1	1	12	10	14	17	11	7	9	15	8	10