

Lista de Exercícios II

**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: 07/03/2024** 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. Seja  $Y = X\beta + \epsilon$ ,  $\epsilon \sim N_n(0, \sigma^2 I)$ . Se  $X_1$  e  $X_2$  são variáveis independentes e  $Y$  é a dependente de uma amostra aleatória com  $n = 30$  elementos, tal que:

$$X'X = \begin{bmatrix} 8 & 15 & 12 \\ 15 & 33 & 24 \\ 12 & 24 & 24 \end{bmatrix}, \quad X'y = \begin{bmatrix} 57 \\ 121 \\ 99,5 \end{bmatrix} \text{ e } y'y = 464,5$$

- (a) Determine a equação de regressão estimada considerando um MRLM do tipo:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ ;
  - (b) Determine a  $SQE = y'y - \hat{\beta}' X' y$  e apresente a tabela de ANOVA;
  - (c) Realize o teste  $F$  de ajustamento global e conclua.
2. Analogamente ao exercício anterior, considere  $n = 10$ ,  $p = 3$ ,  $y'y = 56$  e as equações normais:

$$\begin{aligned} 4\beta_1 + 2\beta_2 - 2\beta_3 &= 4 \\ 2\beta_1 + 2\beta_2 + \beta_3 &= 7 \\ -2\beta_1 + \beta_2 + 6\beta_3 &= 9 \end{aligned}$$

Encontre os estimadores de Mínimos Quadrados (EMQ) para  $\beta$  e para  $\sigma^2$ .

3. Para se estudar a influência das variáveis capital investido ( $X_1$ ) e gasto em publicidade ( $X_2$ ) no lucro anual ( $Y$ ) de empresas, foram observadas essas variáveis em doze empresas em um mesmo ano. Os seguintes resultados foram registrados, na unidade de R\$ 100 mil.

**Tabela 1.** Lucro anual em função do capital investido e gasto em publicidade.

$Y$	12	13	3	3	11	19	1	14	15	17	2	15
$X_1$	31	16	29	19	27	21	24	11	26	18	12	3
$X_2$	4	5	3	0	2	6	2	3	6	6	1	5

- (a) Ajuste a estes dados um MRLM do tipo:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$  por meio do MMQ e compare com as estimativas obtidas via comando *lm* do *R*;
- (b) Apresente graficamente: matriz de correlações múltiplas, o diagrama de dispersão e o ajuste no hiperplano;
- (c) Realize a ANOVA, apresente o  $R^2$  do ajuste e interprete o seu valor;
- (d) Os dados fornecem evidências suficientes para indicar que o modelo contribui significativamente para a predição de  $Y$  para valores médios das variáveis explicativas? Em caso afirmativo, apresente-o juntamente com seu  $IC$ ;
- (e) Teste as hipóteses  $H_0 : \beta_j = 0$  versus  $H_a : \beta_j \neq 0$ ,  $j = 1, 2$ .

4. Um experimento foi realizado para estudar a relação entre o grau de corrosão de um certo metal,  $Y$ , e o tempo de exposição (em semanas),  $X$ , desse metal à ação da acidez do solo. Foram obtidos os seguintes resultados apresentados na Tabela 2.

**Tabela 2.** Grau de corrosão em função do tempo de exposição (em semanas).

$X$	1	2	3	4	5	6	7	8	9	10
$Y$	0,08	0,18	0,32	0,53	0,88	1,30	1,95	2,80	3,90	4,60

- Ajuste um MRLM a esses dados do tipo:  $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$ ;
  - Apresente graficamente o ajuste;
  - Realize a ANOVA;
  - Os dados fornecem evidências suficientes para indicar que o modelo contribui significativamente para a predição de  $Y$ ?
  - Teste as hipóteses  $H_0 : \beta_1 = 0$  versus  $H_a : \beta_1 \neq 0$ ;
  - Que porcentagem da variação total é explicada pelo modelo?
  - Apresente uma estimativa pontual e intervalar para  $X_0 = 5, 5$ .
5. Para estudar a importância das variações do peso do ovo ( $PO$ ) e do peso corporal ( $PC$ ) sobre o consumo de alimentos ( $Y$ ), em codornas de postura, foram anotados dados de 20 animais, apresentados na Tabela 3.

**Tabela 3.** Medidas de consumo alimentar em função do peso do ovo ( $PO$ ) e peso corporal em codornas de postura.

$Y$	$PO$	$PC$	$Y$	$PO$	$PC$
18,217	11,2	171,3	30,439	11,2	168,1
30,099	11,0	170,6	24,489	9,3	159,1
36,559	11,9	187,6	22,693	9,9	158,2
36,879	11,4	187,1	19,067	10,0	145,3
28,757	10,9	166,3	26,283	10,0	166,2
18,507	9,7	141,3	28,309	10,7	170,1
25,535	10,5	163,5	26,013	11,0	160,2
23,893	10,8	155,2	21,519	9,1	151,6
21,545	8,7	153,5	24,985	10,1	165,0
30,319	12,1	176,1	18,959	9,4	146,1

Apresente:

- As equações das RLS individuais ( $Y$  versus  $PO$  e  $Y$  versus  $PC$ ) e os respectivos gráficos ajustados;
- As equações das regressões quadráticas individuais ( $PO$  versus  $Y$  e  $PC$  versus  $Y$ ) e decida pelo melhor ajuste, individualmente;
- A equação de uma RLM do tipo:  $Y = PO + PC$ ;
- A equação de uma RLM do tipo:  $Y = PO + PO^2 + PC + PC^2 + PO \times PC$ ;
- Apresente o melhor modelo ajustado final, inclusive seu gráfico e conclua.

6. Um experimento foi realizado para se verificar o efeito da temperatura ( $X_1$ ) e da concentração ( $X_2$ ) na produção ( $Y$ ) de um certo processo químico. Os dados obtidos são apresentados na Tabela 4.

**Tabela 4.** *Produção em função da temperatura e concentração.*

$Y$	$X_1$	$X_2$	$Y$	$X_1$	$X_2$
189	80	10	220	80	20
203	100	10	228	100	20
222	120	10	252	120	20
234	140	10	263	140	20
261	160	10	291	160	20
204	80	15	226	80	25
212	100	15	232	100	25
223	120	15	259	120	25
246	140	15	268	140	25
273	160	15	294	160	25

- (a) Ajuste um MRLM que explique  $Y$  e função das variáveis explicativas;  
 (b) Encontre um  $IC$  para o coeficiente da variável concentração no modelo ajustado;  
 (c) Calcule e interprete o valor do coeficiente de determinação para esse modelo;  
 (d) Estime a variância residual do modelo ajustado.
7. Um CEO de um grande banco deseja estudar a relação entre os salários mensais ( $Y$ : salários mínimos) e o tempo de experiência ( $E$ : anos) no cargo de gerente das agências. Outro interesse é saber se existem diferenças quando são levados em conta os salários de homens e de mulheres. Para tal, sugere-se um MRLM do tipo:  $Y = \beta_0 + \beta_1 E + \beta_2 S + \beta_3 ES + \epsilon$ , sendo  $S$ : sexo (0 se mulher e 1 se homem).

Considerando os dados coletados a seguir, apresente um relatório de análise estatística a este CEO:

**Tabela 5.** *Remuneração em função do tempo de experiência, por sexo.*

<b>Mulher:</b>	$Y$	1,93	3,18	2,28	3,13	2,78	3,09	2,65	2,22	2,85	3,23	2,82	1,91	
	$E$	0	17	5	15	9	15	8	5	13	20	11	1	
<b>Homem:</b>	$Y$	2,54	2,57	4,22	4,09	3,61	4,71	3,15	2,99	4,75	4,12	2,36	4,09	4,51
	$E$	6	7	23	20	18	27	11	10	29	23	4	22	25

8. Foi realizado um experimento envolvendo 4 variedades de trigo plantadas em canteiros de mesmo tamanho, tendo sido observadas as seguintes variáveis:

**Prod:** produção de trigo (kg);

**Prec:** precipitação pluviométrica (cm);

**Var:** variedade (1, 2, 3 e 4);

**Fert:** concentração de fertilizante empregado na plantação (1, 2 e 3).

**Tabela 6.** *Produção de trigo em função da precipitação, variedade e concentração de fertilizante.*

Prod	Prec	Var	Fert	Prod	Prec	Var	Fert
15,6	14,4	1	1	14,8	14,4	3	1
16,2	13,2	1	1	13,8	13,2	3	1
17,1	14,8	1	2	14,5	14,8	3	2
16,7	13,6	1	2	14,4	13,6	3	2
15,1	14,0	1	3	15,2	14,0	3	3
17,1	12,8	1	3	14,8	12,8	3	3
18,3	14,4	2	1	15,1	14,4	4	1
17,2	13,2	2	1	16,0	13,2	4	1
17,4	14,8	2	2	16,2	14,8	4	2
18,4	13,6	2	2	16,8	13,6	4	2
19,7	14,0	2	3	17,6	14,0	4	3
17,6	12,8	2	3	18,6	12,8	4	3

- Ajuste um modelo aos dados, considerando todas regressões possíveis (ver Exemplo disponível na página 33 no **Material Didático de Giolo (Apostila-2007) - Análise de Regressão Linear Múltipla**. Utilize todos os critérios para a comparação dos modelos;
  - Ajuste um modelo, selecionando as variáveis significantes, que explique a produção de trigo, utilizando cada um dos métodos vistos (*backward*, *forward* e *stepwise*);
  - Qual ajuste é melhor? Por quê?
9. Utilize o banco de dados *mtcars* do *R* e faça uma análise em um MRLM completa considerando como resposta a variável **mpg** e todas as demais como regressoras sem considerar interações.
10. Utilize o banco de dados *bailarinas* disponível na pasta Banco de Dados da disciplina e ajuste um MRLM (análise completa) considerando como resposta a variável **ped**.