Lista 3

1. Um modelo de simulação do fluxo de água quente de *watersheds* é avaliada comparando o fluxo máximo medido em 10 tormentas (cfs) com as previsões de fluxo máximo obtidas do modelo de simulação. Q_0 e Q_p são os fluxos máximos observados e preditos, respectivamente. São consideradas quatro variáveis explanatórias:

   X_1 = área de *watershed* (mt^2);

   X_2 = coeficiente angular médio de *watershed* (em %);

   X_3 = índice de absorção da superfície (0 = completa, 100 = nenhuma absorção);

   X_4 = intensidade máxima de chuva (in/h), calculada em intervalos de meia hora.

   Tabela 1: Dados de *watersheds*

Q_0	Q_p	X_1	X_2	X_3	X_4
28	32	0,03	3,0	70	0,6
112	142	0,03	3,0	80	1,8
398	502	0,13	6,5	65	2,0
772	790	1,0	15,0	60	0,4
2294	3075	1,0	15,0	65	2,3
2484	3230	3,0	7,0	67	1,0
2586	3535	5,0	6,0	62	0,9
3024	4265	7,0	6,5	56	1,1
4179	6529	7,0	6,5	56	1,4
710	935	7,0	6,5	56	0,7

   (a) Use $Y = ln(Q_0/Q_p)$ como variável resposta. Ela poderá tomar valor zero se os fluxos máximos observados e preditos forem semelhantes. Faça uma regressão (por *forward* e *backward*) para determinar se a diferença de Y está relacionada com qualquer das quatro variáveis explanatórias. Inclua o intercepto no modelo.

   (b) Uma informação adicional ao problema sugere que as diferenças entre observação e previsão do fluxo máximo Y poderiam ser próximas de zero, quando os valores das quatro variáveis explanatórias são próximos de zero. Redefina o problema eliminado o intercepto e estime a equação de regressão.

   (c) Calcule novamente a regressão (sem intercepto) utilizando somente X_1 e X_4. Os coeficientes $\hat{\beta}_1$ e $\hat{\beta}_4$ mudam com relação ao item (b)? Explique.

(d) Descreva a mudança nos erros padrão das estimativas dos coeficientes da regressão quando o intercepto é retirado do modelo. E quando X_2 e X_3 são retirados.

(e) Com relação ao modelo obtido em (b), verifique sobre violações dos pressupostos do modelo.

2. Com relação ao exercício anterior, são definidos os seguintes modelos:

$$\begin{aligned} M_1 &: \; y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \\ M_2 &: \; y = \beta_0 + \beta_1 X_1 + \beta_4 X_4 + \varepsilon \end{aligned}$$

Descreva como se comparar os modelos M_1 e M_2 a partir de um teste de hipóteses.

3. Os dados do exercício são `fuel2001`, obtidos no pacote **alr3** do R

```
install.packages(‘‘alr3’’, dependencies=TRUE)
library(alr3)
?fuel2001
```

Dada as variáveis: X_1 = tax, X_2 = dlic, X_3 = ine, X_4 = road e Y = fuel:

(a) Ajuste o modelo de regressão linear múltipla considerando como variáveis explanatórias X_1, X_2, X_3, X_4 e resposta Y.

(b) Faça uma análise de variância do modelo anterior. Identifique quais são as hipóteses que são testadas e estabeleça suas conclusões.

(c) É conveniente incluir X_2 no modelo dado que X_1 já está no modelo?

(d) É conveniente incluir X_2 no modelo dado que X_1 e X_3 estão no modelo?

(e) O que se pode concluir com os itens (c) e (d)? Leve em consideração a correlação das variáveis.

(f) Analise completamente os resíduos dos modelos propostos. O que se pode dizer sobre os pressupostos do modelo de regressão?

4. Os dados abaixo diz respeito a um estudo sobre câncer de cabeça e pescoço (esôfago, laringe, faringe etc.). Sabe-se que o fumo e bebida são fatores de risco para estes tipos de câncer, mas existem outros. Um modelo de regressão logística foi ajustado aos dados para estimar a probabilidade de se ter câncer, e abaixo se encontram os resultados.

Tabela 2: Saída do programa

Parâmetros	**Estimativa**	**P-valor**
Intercepto	-3,66	$< 0,001$
Sexo (1 = F, 0 = M)	0,41	0,47
Fuma1 (0 = nunca, 1 = pouco)	0,98	0,05
Fuma2 (0 = nunca, 1= muito)	2,12	$< 0,001$
Bebe1 (0 = nunca, 1 = pouco)	2,03	$< 0,001$
Bebe2 (0 = nunca, 1 = muito)	2,32	$< 0,001$
Age55 (0 = idade $<$ 55, 1 = idade $>$ 55)	-0,008	0,94
Famcam (0 = sem histórico familiar, 1 = com histórico familiar)	1,56	$< 0,001$
Sexid55 (interação sexo*idade)	1,79	0,006

(a) Construa a matriz de planejamento, X.

(b) Baseado no p-valor, encontre o modelo final fornecido pelo software e interprete-o.

5. Para cada situação abaixo, construa um modelo adequado (definindo explicitamente a matriz X) para analisar os dados quando:

   (a) Y = peso do i-ésimo indivíduo.

   $X_{i1} = 1$ se o i-ésimo indivíduo é negro, 0 c.c.

   X_{i2} = altura do i-ésimo indivíduo.

   (b) Y = Nível de inflamação da artrite

   $X_{i1} = 1$ se o i-ésimo indivíduo pertence ao grupo 1 e 0 c.c.

   $X_{i2} = 1$ se o i-ésimo indivíduo pertence ao grupo 2 e 0 c.c.

   $X_{i3} = 1$ se o i-ésimo indivíduo pertence ao grupo 2 e 0 c.c.

   $X_{i4} = 1$ se o i-ésimo indivíduo recebeu dose de 200ml do medicamento e 0 c.c.

   $X_{i5} = 1$ se o i-ésimo indivíduo recebeu dose de 400ml do medicamento e 0 c.c.

   $X_{i6} = 1$ se o i-ésimo indivíduo recebeu dose de 800ml do medicamento e 0 c.c.

   X_{i7} = idade do i-ésimo indivíduo.