**Teoria da Regressão**

$5^{\underline{a}}$ *Lista de Exercícios*

1. Considere o sistema $\mathbf{AX} = \mathbf{y}$ e seja $\mathbf{A}^-$ uma inversa generalizada de $\mathbf{A}$. Mostre que

a) $\mathbf{AA}^-$ é idempotente.

b) Uma condição necessária e suficiente para que $\mathbf{Ax} = \mathbf{y}$ seja consistente é que $\mathbf{AA}^-\mathbf{y} = \mathbf{y}$.

c) Uma solução geral do sistema consistente $\mathbf{Ax} = \mathbf{y}$ é dada por $\mathbf{x} = \mathbf{A}^-\mathbf{y} + (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{z}$, em que $\mathbf{z} \in I\!R$ é um vetor arbitrário; além disso, toda solução do sistema tem essa forma.

2. Considere o modelo $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}$ $i = 1, 2, j = 1, 2, k = 1, 2$.

a)Escreva $\mathbf{X}^T\mathbf{X}$, $\mathbf{X}^T\mathbf{y}$ e as equações normais.

b) Encontre o conjunto de funções estimáveis linearmente independentes.

c)Mostre que $H_0 : \alpha_1 = \alpha_2$ é testável.

3. Utilizando a notação usual, considere o modelo linear

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} E(y_{11}) \\ E(y_{12}) \\ E(y_{21}) \\ E(y_{22}) \end{pmatrix}$$

(i)Expresse os parâmetros do modelo em termos dos valores esperados $E(y_{ij})$.

(ii) Repita o procedimento do item anterior sob as restrições

a)$\alpha_1 = 0$

b)$\alpha_1 + \alpha_2 = 0$

(iii) Repita o procedimento agora sob as reparametrizações

a)$\beta_1 = \mu + \alpha_1, \beta_2 = \mu + \alpha_2$

b)$\beta_1 = \mu, \beta_2 = \mu + \alpha_1$

Interprete os parâmetros em cada caso.

4. Considere o modelo de regressão linear simples $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, em que a variância de $\epsilon_i$ é proporcional a $x_i^2$ , isto é, $Var(\epsilon) = \sigma^2 x_i^2$. Suponha que se use a transformação y' = y/x e x' = 1/x. a) Esta transformação estabiliza a variância ? b) Que relação há entre os parâmetros do modelo original e do modelo transformado ? c) Suponha que nós usamos o método de mínimos quadrados ponderados com pesos $v_i = x_i^2$

5. Uma escola avalia o seu curso através de um questionário com 50 perguntas sobre diversos aspectos de interesse. Cada pergunta tem uma resposta numa escala de 1 a 5, onde a maior nota significa melhor desempenho. Para cada aluno é então encontrada a nota média. Os

dados abaixo representam os resultados obtidos com amostras de alunos dos três períodos de funcionamento da escola.

| Período | | |
|---|---|---|
| Manhã | Tarde | Noite |
| 4,2 | 2,7 | 4,6 |
| 4,0 | 2,4 | 3,9 |
| 3,1 | 2,4 | 3,8 |
| 2,7 | 2,2 | 3,7 |
| 2,3 | 1,9 | 3,6 |
| 3,3 | 1,8 | 3,5 |
| 4,1 | 3,4 | 2,8 |

Faça uma análise de dados (comparando entre as opiniões dos alunos dos 3 periodos) utilizando cada um dos 4 modelos (parametrizações) apresentados em classe ( médias, Posto incompleto, desvios médio, casela de referência). No caso do modelo de Posto incompleto, utilize uma matriz $(\mathbf{X}^t\mathbf{X})^-$ no lugar da $(\mathbf{X}^t\mathbf{X})^{-1}$ de modo que $(\mathbf{X}^t\mathbf{X})(\mathbf{X}^t\mathbf{X})^-(\mathbf{X}^t\mathbf{X}) = (\mathbf{X}^t\mathbf{X})^-$. Essa matriz não é única e é denominada de inversa generaliza de $(\mathbf{X}^t\mathbf{X})$. Mostre que $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^-\mathbf{X}^t\mathbf{y}$ é uma solução do sistema $(\mathbf{X}^t\mathbf{X})\boldsymbol{\beta} = \mathbf{X}^t\mathbf{y}$ e $SQReg = \mathbf{y}^t\mathbf{X}(\mathbf{X}^t\mathbf{X})^-\mathbf{X}\mathbf{y}$ é invariante a escolha da inversa generalizada. Faça uso de pelo menos um pacote estatístico, descrevendo o tipo de modelo(parametrização) utilizado.

5. Os dados (hipotético) abaixo são provenientes de um plano experimental completamente casualizado cujo objetivo é o desenvolvimento de um novo medicamento de alívio de dor. Um experimento envolveu 9 voluntários, e a quantidade de dois ingredientes ativos (fator A e fator B). Na composição do medicamento, os ingredienetes tem variado em 2 níveis para o fator A e 3 níveis para o fator B. Aleatorização tem usada selecionando 3 voluntários para 1 cada um dos 6 tratamentos no estudo (a). Três situações diferentes de estudos aconteceram. Os dados em horas de alivio da dor foram anotados abaixo

a)balanceado

| A | B | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| | 7 | 8 | 2 |
| 1 | 9 | 6 | 4 |
| | 8 | 9 | 2 |
| | 5 | 14 | 20 |
| 2 | 7 | 15 | 22 |
| | 4 | 19 | 23 |

b)desbalanceado

| A | B | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| | 7 | 8 | 2 |
| 1 | 9 | | 4 |
| | 8 | 9 | 2 |
| | 5 | 14 | 20 |
| 2 | 7 | 15 | 22 |
| | | 19 | 23 |

c)incompleto

| A | B | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| | 7 | 8 | 2 |
| 1 | 9 | | 4 |
| | 8 | | |
| | 5 | 14 | |
| 2 | 7 | 15 | |
| | | 19 | |

Para cada caso responda as perguntas abaixo:

(i) Explique a diferença de cada situação.

(ii) Obtenha os efeitos principais A e B.

(iii) Verifique se a interação está presente no modelo.

(iv) Teste os efeitos principais.

(v) Dependendo da resposta do item anterior, construa os intervalos de confiança para as diferenças de médias.

(vi) Construa a tabela da ANOVA e interprete os resultados.