**Teoria da Regressão**

*Prova 2.*

1. Considere o modelo de regressão com $p$ termos especificado corretamente, incluindo o intercepto. Utilizando as suposições usuais sob $\boldsymbol{\epsilon}$, mostre que
a)$Var(\hat{\mathbf{y}}) = \sigma^2\mathbf{H}$;
b)$Cov(\hat{\boldsymbol{\epsilon}}, Y) = (I - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)\sigma^2$

2. Responda as questões selecionadas abaixo do livro Modelos de Regressão com apoio computacional(referência n.7 da ementa) - Capítulo 1 seção 1.11 questões 11, 18 e 19.
3. Em 48 estados americanos foram registradas as seguintes variáveis taxa(taxa de combustível no estado), licença(proporção de motorista licenciados), renda (renda per-capita), estradas (ajuda federal para as estradas) e consumo (consumo de combustível por habitante). O interesse é explicar o consumo de cmbustível pelas variáveis taxa, licença, renda e estradas.

**(a)** Construa os gráficos de dispersão convinientes.

**(b)** Ajuste um modelo normal linear para explicar o consumo contra as demais variáveis selecionando as variáveis explicativas que contribuem significamente para o modelo. Interprete os diferentes parâmetros do modelo.

**(c)** Estime os parâmetros do modelo e apresente os respectivos erros padrões.

**(d)** Avalie a qualidade do ajuste do modelo.

**(e)** Construa o intervalo de confiança para consumo médio e vitórias segundo um conjunto de valores das explicativas a sua escolha.

**(f)** Apresente um conclusão que evite o jargão estatístico.

**(g)** Faça uma análise de resíduo e diagnóstico completa.

| | | | | | |
|---|---|---|---|---|---|
| ME | 9.00 | 0.525 | 3571 | 1976 | 541 |
| VT | 9.00 | 0.580 | 3865 | 1586 | 561 |
| RI | 8.00 | 0.544 | 4399 | 431 | 410 |
| NY | 8.00 | 0.451 | 5319 | 11868 | 344 |
| PA | 8.00 | 0.529 | 4447 | 8577 | 464 |
| IN | 8.00 | 0.530 | 4391 | 5939 | 580 |

| | | | | | |
|---|---|---|---|---|---|
| MI | 7.00 | 0.574 | 4817 | 6930 | 525 |
| MN | 7.00 | 0.608 | 4332 | 8159 | 566 |
| MO | 7.00 | 0.572 | 4206 | 8508 | 603 |
| SD | 7.00 | 0.724 | 4716 | 5915 | 865 |
| KS | 7.00 | 0.663 | 4593 | 7834 | 649 |
| MD | 9.00 | 0.511 | 4897 | 2449 | 464 |
| WV | 8.50 | 0.551 | 4574 | 2619 | 460 |
| SC | 8.00 | 0.548 | 3448 | 5399 | 577 |
| FL | 8.00 | 0.563 | 4188 | 5975 | 574 |
| TN | 7.00 | 0.518 | 3640 | 6905 | 571 |
| MS | 8.00 | 0.578 | 3063 | 6524 | 577 |
| LA | 8.00 | 0.487 | 3528 | 3495 | 487 |
| TX | 5.00 | 0.566 | 4045 | 17782 | 640 |
| ID | 8.50 | 0.663 | 3635 | 3274 | 648 |
| CO | 7.00 | 0.626 | 4449 | 4639 | 587 |
| AZ | 7.00 | 0.603 | 4300 | 3635 | 632 |
| NV | 6.00 | 0.672 | 5215 | 2302 | 782 |
| OR | 7.00 | 0.623 | 4296 | 4083 | 610 |
| NH | 9.00 | 0.572 | 4092 | 1250 | 524 |
| MA | 7.50 | 0.529 | 4870 | 2351 | 414 |
| CT | 10.00 | 0.571 | 5342 | 1333 | 457 |
| NJ | 8.00 | 0.553 | 5126 | 2138 | 467 |
| OH | 7.00 | 0.552 | 4512 | 8507 | 498 |
| IL | 7.50 | 0.525 | 5126 | 14186 | 471 |
| WI | 7.00 | 0.545 | 4207 | 6580 | 508 |
| IA | 7.00 | 0.586 | 4318 | 10340 | 635 |
| ND | 7.00 | 0.540 | 3718 | 4725 | 714 |
| NE | 8.50 | 0.677 | 4341 | 6010 | 640 |
| DE | 8.00 | 0.602 | 4983 | 602 | 540 |
| VA | 9.00 | 0.517 | 4258 | 4686 | 547 |
| NC | 9.00 | 0.544 | 3721 | 4746 | 566 |
| GA | 7.50 | 0.579 | 3846 | 9061 | 631 |
| KY | 9.00 | 0.493 | 3601 | 4650 | 534 |
| AL | 7.00 | 0.513 | 3333 | 6594 | 554 |
| AR | 7.50 | 0.547 | 3357 | 4121 | 628 |
| OK | 6.58 | 0.629 | 3802 | 7834 | 644 |
| MT | 7.00 | 0.586 | 3897 | 6385 | 704 |
| WY | 7.00 | 0.672 | 4345 | 3905 | 968 |
| MN | 7.00 | 0.563 | 3656 | 3985 | 699 |
| UT | 7.00 | 0.508 | 3745 | 2611 | 591 |
| WA | 9.00 | 0.571 | 4476 | 3942 | 510 |
| CA | 7.00 | 0.593 | 5002 | 9794 | 524 |

Estes dados podem ser encontrados em http://www.ime.usp.br/∼giapaula/reg2.dat.

4. 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.

5. 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.

6. 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.

7. 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 $v_i = x_i^2$

8.Os dados a seguir correspondem ao preço de casa (em milhares de reais) em três regiões de uma Cidade

| Período | | |
|---|---|---|
| $R_1$ | $R_2$ | $R_3$ |
| 74 | 76 | 45 |
| 68 | 81 | 49 |
| 77 | 76 | 55 |
| 67 | 80 | 67 |
| 59 | 83 | 69 |
| 69 | 87 | 70 |
| 71 | 65 | 75 |

Faça uma análise de dados (comparando entre os preços das 3 regiões) utilizando cada um dos 4 modelos (parametrizações) apresentados em classe ( médias, Posto incompleto, desvios médio, casela de referência). Caracterize $\mathbf{y}$, $\mathbf{X}$ e interprete os parâmetros. 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.

9. 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.

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

10. No modelo

$$Y_1 = \theta_1 + \theta_2 - 2\theta_3 + \epsilon_1$$

$$Y_2 = \theta_1 - \theta_2 + \epsilon_2$$

$$Y_3 = \theta_2 - \theta_3 + \epsilon_3$$

com $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$ $i, j = 1, 2, 3$. Prove que

a)$\theta_1 - \theta_3$ é estimável

b)$\theta_1$ não é estimável

c) Admitindo que $\theta_1 + \theta_2 + \theta_3 = 0$, prove que $\theta_1$ se torna estimável.