MI 416 -Introdução aos Modelos Lineares
Primeiro semestre de 2013
Prova I
Data: 15/04/2013

Nome: ______________________________ RA: ________________

Leia atentamente as instruções abaixo:

- **Coloque seu nome completo e RA em todas as folhas que você recebeu, inclusive nesta.**
- **Utilize somente um dos lados de cada folha.**
- **Leia atentamente cada uma das questões.**
- **Enuncie, claramente, todos os resultados que você utilizar.**
- **Justifique, adequadamente, seus desenvolvimentos, sem, no entanto, escrever excessivamente.**
- **O(a) aluno(a) só poderá sair da sala após as 8h30, mesmo que já tenha finalizado a prova. Após a saída do(a) primeiro(a) aluno(a) não será permitido a entrada de nenhum(a) outro(a) aluno(a).**
- **Não é permitdo empréstimo de material.**
- **Não serão dirimidas dúvidas de quaisquer natureza, após os 20 minutos iniciais.**
- **Resolva a prova, preferencialmente, à caneta, e procure ser organizado(a). Se fizer à lápis, destaque, à caneta, sua resposta.**
- **O(a) aluno(a) deverá portar sua carteira de estudante e apresentá-la, quando for solicitada sua assinatura.**
- **Contestações a respeito da nota, só serão consideradas se estiverem por escrito.**
- **A nota do aluno(a) será $\frac{NP}{NT} \times 10$, em que NP é o número de pontos obtidos na prova e NT é o numero total de pontos da prova.**
- **Use somente um lado de cada folha de resolução.**
- **Os resultados numéricos finais devem ser apresentados com duas casas decimais, apenas.**
- **A prova terá duração de 120 minutos, das 8h00 às 10h00, improrrogáveis.**

Faça uma excelente Prova!!

1. Considere o modelo normal linear homocedástico visto em sala de aula, ou seja:

$$\boldsymbol{Y}_{(n\times 1)} = \boldsymbol{X}_{(n\times p)}\boldsymbol{\beta}_{(p\times 1)} + \boldsymbol{\xi}_{(n\times 1)}, \boldsymbol{\xi} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$$

e defina $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ e $\widehat{\sigma}^2 = \dfrac{1}{n-p}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) = \dfrac{1}{n-p}\boldsymbol{Y}'\left(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)\boldsymbol{Y}$.
Responda os itens abaixo:

a) Obtenha as distribuições de $\widehat{\boldsymbol{\beta}}$ e $\dfrac{(n-p)\widehat{\sigma}^2}{\sigma^2}$ e prove que $\widehat{\boldsymbol{\beta}}$ e $\widehat{\sigma}^2$ são (estatisticamente) independentes (250 pontos).

b) Considere o interesse em se testar as hipóteses $H_0 : \boldsymbol{C}_{(q\times p)}\boldsymbol{\beta} = \boldsymbol{0}_{(q\times 1)}$ vs $H_1 : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{0}$, em que $\boldsymbol{C}$ é uma matriz não aletória e $r(\boldsymbol{C}) = q$. Defina a seguinte estatística de teste:

$$F = \frac{Q^*/q}{\widehat{\sigma}^2/\sigma^2} = \frac{1}{q\widehat{\sigma}^2}\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right)^{-1}\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}\right)$$

em que

$$Q^* = \frac{1}{\sigma^2}\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right)^{-1}\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}\right)$$

Obtenha sua distribuição sob H_0 e sob H_1 (300 pontos).

c) Obtenha a distribuição conjunta do vetor aleatório: $(\widehat{\boldsymbol{Y}}, \boldsymbol{R})'$, em que $\widehat{\boldsymbol{Y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ e $\boldsymbol{R} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$ (150 pontos).

2. Os dados analisados são provenientes de um estudo na área de Cardiologia. O objetivo é comparar as curvas de variação do consumo de oxigênio no ponto de compensação respiratório (PCR) em função da carga utilizada na esteira ergométrica para pacientes com diferentes etiologias cardíacas, a saber: grupo 1 - controle (C) ; grupo 2 - chagásicos (CH) ; grupo 3 - idiopáticos (ID) ; grupo 4 - isquêmicos (IS). Inicialmente, foi ajustado o seguinte modelo normal linear (a ordem dos grupos é a descrita anteriormente):

$$Y_{ij} = \beta_{0i} + \beta_{1i}x_{ij} + \xi_{ij}, i = 1, 2, 3, 4(\text{grupo}); j = 1, ..., n_i(\text{unidade experimental})$$

em que

- $\xi_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$;
- x_{ij}: carga à que o paciente j que apresenta a etiologia cardíaca i foi submetido (conhecido e não aleatório).

  Abaixo encontram-se alguns resultados relativos à análise descritiva e ao ajuste do modelo. Considere um nível de significância $\alpha = 0,05$.

Responda os itens:

a) Baseando-se nos gráficos de dispersão, você acha que o modelo é apropriado para responder às perguntas de interesse? Justifique, adequadamente, sua resposta (50 pontos).

b) O que você pode afirmar sobre a validade de TODAS as suposições do modelo (em relação aos erros) para o conjunto de dados em questão, utilizando os resultados da análise residual? Justifique, adequadamente, seus comentários (200 pontos).

c) Com base nos resultados da tabela ANOVA, o que você pode concluir em relação à influência da carga utilizada na esteira ergométrica no consumo de oxigênio ao longo dos grupos? Justifique, adequadamente, sua resposta (100 pontos).

d) Com base nos resultados da tabela com as estimativas dos parâmetros, o que você pode concluir com respeito a nulidade de cada um deles? E com relação à possíveis igualdades entre os parâmetros de mesma natureza (interceptos entre si e coeficientes angulares entre si)? Justifique, adequadamente, suas respostas (250 pontos).

e) Com base em todos os resultados (relativos às estimativas dos parâmetros e testes de hipótese) proponha um candidato à modelo reduzido? Justifique sua resposta. Escreva o modelo reduzido (nas formas escalar e matricial), com todas as suposições pertinentes, e indique quais seriam os próximos passos da análise? (200 pontos)

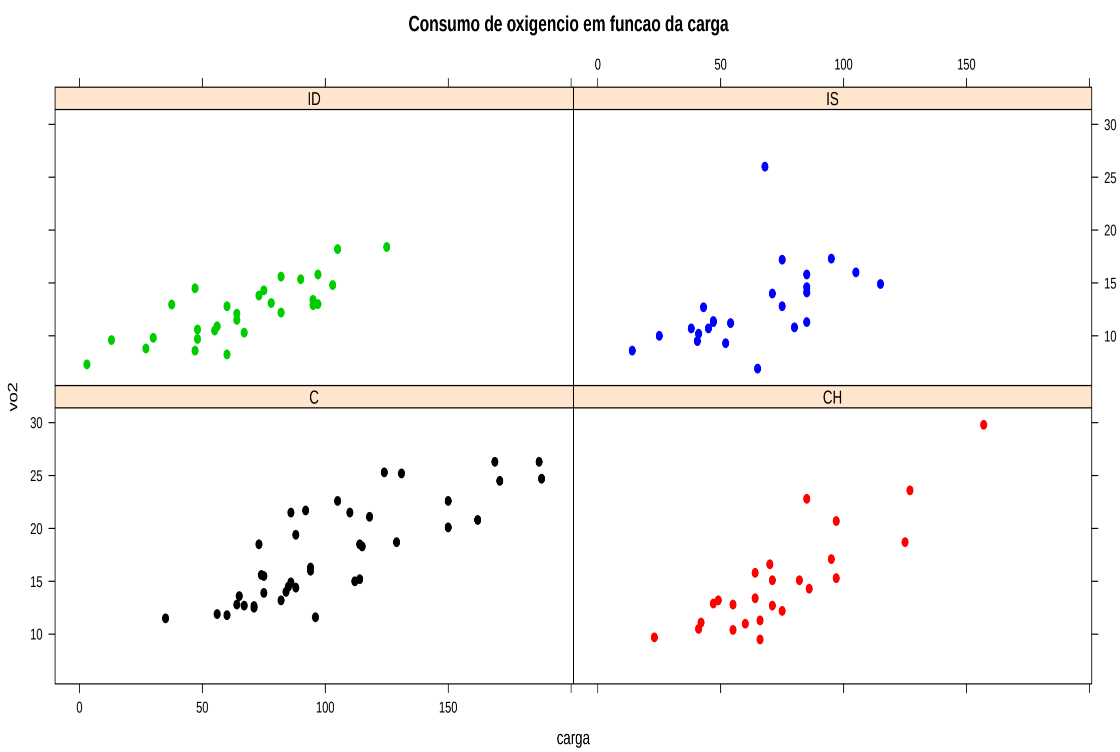

Figura 1: Gráficos de dispersão entre carga utilizada na esteira ergométrica e consumo de oxigênio por etiologia cardíaca.

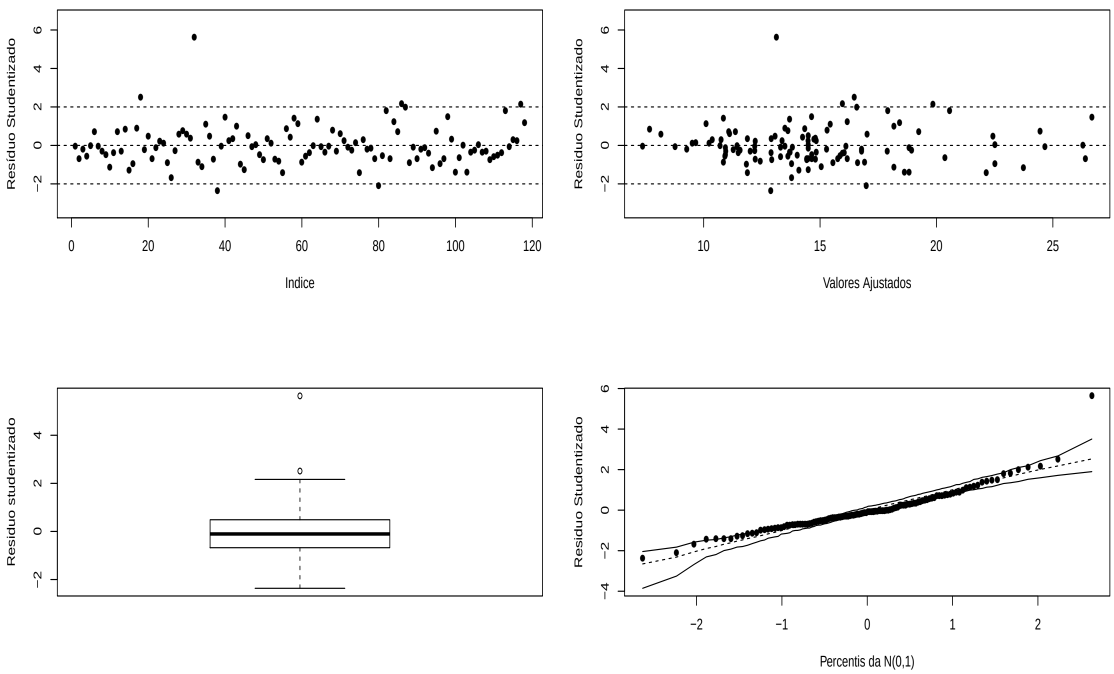

Figura 2: Gráfico de resíduos para o modelo da Questão 2

Tabela 1: Tabela ANOVA para o modelo da Questão 2 (teste para nulidade simultânea dos coeficientes angulares)

FV	G.L.	SQ	QM	Estatística F	p-valor
Modelo	4	1245,89	311,47	44,70	<0,0001
Resíduo	110	766,52	6,97		

Tabela 2: Estimativas (pontuais e intervalares) dos parâmetros do modelo relativo à questão 2

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	p-valor
$\beta_{01}(C)$	7,17	1,23	[4,72 ; 9,61]	5,81	<0,0001
$\beta_{02}(CH)$	4,42	1,45	[1,55 ; 7,29]	3,05	0,0028
$\beta_{03}(ID)$	7,14	1,24	[4,67 ; 9,61]	5,74	<0,0001
$\beta_{04}(IS)$	7,65	1,48	[4,71 ; 10,58]	5,16	<0,0001
$\beta_{11}(C)$	0,10	0,01	[0,08 ; 0,12]	9,05	<0,0001
$\beta_{12}(CH)$	0,14	0,02	[0,11 ; 0,18]	7,87	<0,0001
$\beta_{13}(ID)$	0,08	0,02	[0,04 ; 0,11]	4,58	<0,0001
$\beta_{14}(IS)$	0,08	0,02	[0,04 ; 0,12]	3,74	0,0003

3. Considere o seguinte modelo de regressão normal linear:

$$Y_i = \beta x_i + \xi, i = 1, 2, ..., n; \xi \overset{ind.}{\sim} N(0, \psi_i)$$

em que x_i e ψ_i são conhecidos e não aleatórios, $\forall i$ e β é desconhecido e não aleatório. Responda os itens.

a) Escreva o modelo acima na forma matricial e interprete o parâmetro β nas duas seguintes situações:

(a) x_i corresponde à uma variável contínua.

(b) $x_i = 0; i = 1, ..., n_1$ e $x_i = 1; i = n_1 + 1, ..., n$, em que $n_1 < n$.
(150 pontos)

b) Obtenha o estimador de mínimos quadrados ponderados de β, ou seja, o valor que minimiza a expressão

$$Q = \sum_{i=1}^{n} \psi_i^{-1} \left(Y_i - \beta x_i\right)^2$$

provando que o estimador obtido, de fato, é ponto de mínimo (250 pontos).

c) Obtenha a distribução exata de $\widehat{\beta}$, o estimador obtido no item b) (150 pontos).

d) Proponha algum teste (estatística do teste e as regiões de aceitação e crítica) para testar $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$, à um nível de significância de α. Sugestão: utilize o item c) (250 pontos).

e) Obtenha a função poder do teste proposto no item d) (200 pontos).

Formulário/Teoremas

1. Se $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então $\boldsymbol{Y} = \boldsymbol{A}_{(q\times p)}\boldsymbol{X} \sim N_q(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}')$.

2. Seja $\boldsymbol{Y} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$ e $\boldsymbol{A}$ uma matriz simétrica. Então

$$\boldsymbol{Y}'\boldsymbol{AY} \sim \chi^2_{[r(\boldsymbol{A})]}$$

se, e somente se $\boldsymbol{A}$ for idempotente.

3. Seja $\boldsymbol{Y} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ e $\boldsymbol{A}$ uma matriz simétrica. Então

$$\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y} \sim \chi^2_{[r(\boldsymbol{A})]}$$

se, e somente se $\boldsymbol{A}\boldsymbol{\Sigma}$ for idempotente.

4. Seja $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{I})$ e $\boldsymbol{A}$ uma matriz simétrica. Então

$$\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y} \sim \chi^2_{[r(\boldsymbol{A}),\delta=\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}]}$$

se, e somente se $\boldsymbol{A}$ for idempotente.

5. Seja $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\boldsymbol{A}$ uma matriz simétrica. Então

$$\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y} \sim \chi^2_{[r(\boldsymbol{A}),\delta=\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}]}$$

se, e somente se $\boldsymbol{A}\boldsymbol{\Sigma}$ for idempotente.

6. Naturalmente, se $\delta = \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} = 0$, as distribuições (dos itens 2-5) passam a ser qui-quadrados centrais.

7. Seja $\boldsymbol{Y} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{A}$ uma matriz simétrica e $\boldsymbol{B}$ uma matriz qualquer. Então $\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}$ e $\boldsymbol{B}\boldsymbol{Y}$ são independentes se, e somente se $\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A} = \mathbf{0}$.

8. Seja $\boldsymbol{Y} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{A}$ e $\boldsymbol{B}$ matrizes simétricas. Então $\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}$ e $\boldsymbol{Y}'\boldsymbol{B}\boldsymbol{Y}$ são independentes se, e somente se $\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B} = \mathbf{0}$.