

UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE MATEMÁTICA - DEPARTAMENTO DE ESTATÍSTICA
DISCIPLINA – MODELOS LINEARES GENERALIZADOS

Lista 5

1. Vinte e uma pessoas tomaram um curso de formação. No início do curso, eles fizeram um teste para avaliar sua aptidão inicial. Eles foram alocados aleatoriamente para um dos três diferentes métodos de treinamento e no final do curso um outro teste foi realizado para obter o escore alcançado. Estamos interessados em comparar métodos.

Training	Method 1		Method 2		Method 3	
	Before	After	Before	After	Before	After
	3	6	4	8	3	6
	1	4	5	9	2	7
	3	5	5	7	2	7
	1	3	4	9	3	7
	2	4	3	8	4	8
	1	3	1	5	5	1
	4	6	2	7	4	7

- Construa um ou mais gráficos desses dados com auxílio do programa R de forma que uma descrição visual seja possível com ajuda dos mesmos. Faça um breve comentário
- Defina um modelo linear generalizado, com a devida notação, que você julgue ser o mais adequado para analisar esses dados
- Explique brevemente o que cada parâmetro representa
- Entre com os dados no R como dataframe e ajuste o modelo que você definiu no item c. Apresente os comandos do R e também as estimativas dos parâmetros correspondente do seu modelo sem esquecer de interpretá-los.
- Comente brevemente sobre os pressupostos do modelo utilizado

2. Utilizando os dados do relatório de um estudo em Agresti e Finlay (1997) sobre um estudo na Florida que tinha objetivo de investigar a relação entre saúde mental e diversas variáveis explicativas usando uma amostra aleatória de 40 indivíduos. O desfecho de interesse é um índice de deficiência mental que incorpora medidas de ansiedade e depressão. As variáveis preditoras consideradas nesse estudo são: índice de status sócio-econômico (SES) e um escore relacionados aos eventos da vida, que combina o número e a gravidade de vários eventos estressantes na vida desses adultos. Os dados estão no formato stata “*afMentalHealth.dta*”

- Monte uma matriz de dispersão com as variáveis de interesse e faça um breve comentário sobre a relação entre cada par de variáveis.
- Primeiramente, ajuste um MLG (ligação identidade) simples do índice de deficiência mental e o escore relacionado aos eventos da vida. Interprete o coeficiente. Além disso, qual a proporção de variação entre os indivíduos no índice da saúde mental é explicada pelo índice de eventos de vida? Como é essa proporção relacionada com coeficiente de correlação de Pearson?
- Tomando como base o modelo ajustado no item anterior e verifique a linearidade desta relação

justificando o método utilizado para isso. Em geral, é uma boa ideia centralizar variáveis sobre sua média pois simplifica a interpretabilidade

d) Ajuste um modelo com as duas variáveis preditoras e interprete brevemente a estimativa do coeficiente relacionado a variável eventos da vida, e compare com a estimativa obtida anteriormente

e) Calcule os valores ajustados para esse modelo e calcule o coeficiente de Pearson entre os valores observados e ajustados. Esse número parece familiar?

f) Utilize dois procedimentos para verificar se a função de ligação usada é a mais adequada sem esquecer de descrever adequadamente os métodos utilizados e conclusão.

3. 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

$$Y_{ij} = \beta_0 + \beta_{1i}x_{ij} + \varepsilon_{ij} ; i = 1, 2, 3, 4(\text{grupo}); j = 1, 2, \dots, n_i(\text{unidade experimental})$$

em que ε_{ij} são i.i.d $N(0; \sigma^2)$;

x_{ij} : carga à que o paciente j que apresenta a etiologia cardíaca i foi submetido (conhecido e não aleatório);

β_0 e β_{1i} $i = 1, 2, 3, 4$ possuem as interpretações usuais.

Alguns resultados relativos à análise descritiva e ao ajuste do modelo encontram-se nas Tabelas 1 e 2 e na Figura 1. Considere um nível de significância $\alpha = 0.05$.

Responda:

a) Baseando-se na Figura 1, você acha que o modelo ajustado é apropriado para responder as perguntas de interesse? Justifique, adequadamente, sua resposta.

b) Com base na tabela 1, o que você pode concluir em relação a influência da carga utilizada na esteira ergométrica no consumo de oxigênio ao longo dos grupos. Justifique.

c) Com base nos resultados da Tabela 2, o que você pode concluir com respeito a nulidade de cada um deles? E com relação as possíveis igualdades entre os coeficientes angulares? Justifique.

d) Com base em todos os resultados (relativos às estimativas dos parâmetros e testes de hipótese) proponha um candidato à modelo reduzido escrevendo-o nas formas escalar e matricial, com todas as suposições pertinentes, e indique quais seriam os próximos passos da análise. Escreva na forma $H_0: C\beta = 0$ vs $H_1: C\beta \neq 0$ a hipótese que define o modelo reduzido em termos do modelo original.

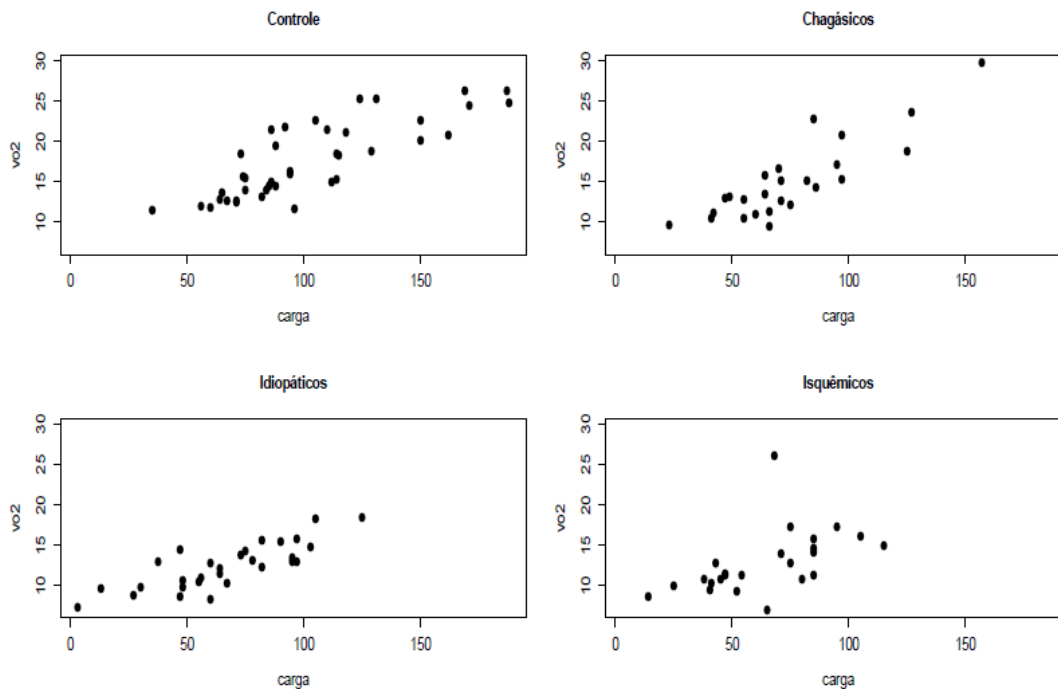


Figura 1: Gráficos de dispersão entre carga utilizada na esteira ergométrica e consumo de oxigênio por etiologia cardíaca.

Tabela 1: Tabela ANOVA para o modelo da Questão 3

FV	G.L.	SQ	QM	Estatística F	p-valor
Modelo	4	1821,81	455,45	65,27	<0,0001
Resíduo	113	788,53	6,98		

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

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	p-valor
β_0	6,671	0,670	[5,345 ; 7,998]	9,967	<0,0001
β_{11}	0,107	0,007	[0,093 ; 0,120]	15,397	<0,0001
β_{12}	0,116	0,010	[0,096 ; 0,136]	11,373	<0,0001
β_{13}	0,085	0,011	[0,063 ; 0,107]	7,794	<0,0001
β_{14}	0,094	0,012	[0,070 ; 0,118]	7,794	<0,0001

4. Explique cada passo justificando com a teoria o programa abaixo para obter alguns gráficos para análise do diagnóstico do modelo

```
# Análise diagnóstico de um MLG – objeto mod2
```

```
X <- model.matrix(fit.model)
n <- nrow(X)
p <- ncol(X)
w <- fit.model$weights
W <- diag(w)
```

```

H <- solve(t(X)%*%W%*%X)
H <- sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h <- diag(H)
ts <- resid(fit.model,type="pearson")/sqrt(1-h)
td <- resid(fit.model,type="deviance")/sqrt(1-h)
di <- (h/(1-h))*(ts^2)
a <- max(td)
b <- min(td)
par(mfrow=c(2,2))
plot(fitted(fit.model),h,xlab="Valores Ajustados", ylab="Medida h",
pch=16)
#identify(fitted(fit.model), h, n=2)
#title(sub="(a)")
#
plot(di,xlab="Indice", ylab="Distancia de Cook",pch=16)
#identify(di, n=1)
#title(sub="(b)")
#
plot(td,xlab="Indice", ylab="Residuo Componente do Desvio",
ylim=c(b-1,a+1), pch=16)
abline(2,0,lty=2)
abline(-2,0,lty=2)
# identify(td, n=1)
# title(sub="(c)")
#
eta = predict(fit.model)
z = eta + resid(fit.model, type="pearson")/sqrt(w)
plot(predict(fit.model),z,xlab="Preditor Linear",
ylab="Variavel z", pch=16)
lines(smooth.spline(predict(fit.model), z, df=2))
#title(sub="(d)")
#-----#

```