

Modelos Lineares

Teste - 17/01/2020

Duração: 2h30m

Mestrado em Estatística

Departamento de Matemática

Nome: _____ Número: _____

Grupo I

1. Considere os seguintes dados:

```
> data(mtcars)
```

Para saber informação sobre os dados:

```
> help(mtcars)
```

Nota: Em todas as questões que utilizar o programa R apresentar os comandos usados.

Usando as variáveis *drat*, *cyl*, *disp*, *hp* e *am* como variáveis explicativas e a variável *mpg* como variável resposta responda às seguintes questões:

- (a) Teste se o modelo com as cinco variáveis explicativas e o modelo só com a variáveis *hp* diferem ou não significativamente quanto ao ajustamento aos dados. Explícite as hipóteses nula e alternativa.
- (b) Ajuste um modelo de regressão aplicando o método backward com teste F. Interprete o significado dos coeficientes de regressão no contexto do problema.
Nas alíneas seguintes, considere o modelo ajustado na alínea (b). (Caso não tenha resolvido considere o modelo com todas as variáveis explicativas)
- (c) Calcule e interprete o coeficiente de determinação obtido para o ajustamento.
- (d) Será admissível considerar que quando todas as variáveis explicativas são nulas, o valor médio previsto para o consumo é também zero? Justifique convenientemente a sua resposta.
- (e) Prever o consumo de combustível do primeiro automóvel. Obter o respetivo intervalo de predição a 90%.
- (f) Verifique os pressupostos do modelo. Identifique as observações alavanca e os pontos influentes. Indique o valor da distância de Cook associado com a primeira observação.

Grupo II

1. Num estudo numa espécie de árvores pretende-se estabelecer relações entre a altura dos troncos das árvores, o respectivo diâmetro à altura do peito e o volume desses troncos. Foram efectuadas medições das variáveis Altura (medida em pés), Diâmetro (medido em polegadas) e Volume (medido em pés cúbicos) em 31 árvores. Eis os valores de algumas estatísticas descritivas elementares, bem como dos coeficientes de correlação entre as variáveis:

	Diâmetro	Altura	Volume
Mean	13.25	76	30.17
Var	9.8479	40.60	270.202796

	Diâmetro	Altura	Volume
Diâmetro	1.0000000	0.5192801	0.9671194
Altura	0.5192801	1.0000000	0.5982497
Volume	0.9671194	0.5982497	1.0000000

Foi ajustado um modelo de regressão linear múltipla para prever os volumes dos troncos, a partir das suas alturas e diâmetro, tendo sido obtidos os seguintes resultados.

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	-57.9877	8.6382	-6.713
Diametro	4.7082	0.2643	17.816
Altura	0.3393	0.1302	2.607

Residual standard error : 3.882 on ?? degrees of freedom

Multiple R – squared : 0.948, *Adjusted R – squared* : ???

F – statistic : ??? on ??? and ??? DF

- Complete os resultados obtidos, indicando os valores em falta (???). Justifique as suas respostas.
- Construir a tabela ANOVA associada a este modelo.
- Diga se é possível simplificar este modelo, obtendo uma regressão linear simples que não seja significativamente pior do que este modelo e indique o respetivo coeficiente de determinação.

2. O modelo seguinte pretende explicar as classificações obtidas pelos alunos de determinada disciplina, lecionada em 52 horas:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

onde Y é a classificação obtida pelo aluno, X_1 é o tempo de estudo da disciplina (em horas), X_2 é o tempo de presença nas aulas (em horas) e $X_3 = 1$ se o aluno repetiu e $X_3 = 0$ caso contrário. Recolhida uma amostra de 30 observações, obteve-se:

$$\text{Modelo : } \hat{Y} = 1.0929 + 0.1589X_1 + 0.1171X_2 - 0.2839X_3 \quad s_Y = 3.43896 \quad F_{obs} = 32.9865$$

$$\widehat{cov(\beta)} = \begin{bmatrix} 1.19839 & -0.01617 & -0.01443 & -0.36365 \\ -0.01617 & 0.00056 & -0.0006 & 0.00265 \\ -0.01443 & -0.0006 & 0.00044 & 0.00394 \\ -0.36365 & 0.00265 & 0.00394 & 0.48899 \end{bmatrix}$$

- Escreva separadamente os modelos ajustados para os alunos repetentes e para os não repetentes.
- Teste se o coeficiente associado com a variável X_3 é igual a zero.
- Teste o ajustamento global do modelo.
- Calcule a variância dos erros.
- Alguns defendem que, para efeitos de classificação, uma hora adicional de estudo é igualmente útil a uma hora de presença nas aulas. Teste esta opinião.
- Qual a classificação esperada que um aluno não repetente, que dedique 60 horas ao estudo da disciplina e assista a todas as aulas?

Grupo III

- Considere o modelo de regressão linear $Y_i = \beta_0 + \beta_1(3X_i^2 - 2) + \epsilon_i$, $i=1,2,3$, onde $x_1 = -1$, $x_2 = 0$ e $x_3 = 1$. Encontre as estimativas de mínimos quadrados dos coeficientes em termos dos valores observados y_1 , y_2 e y_3 .
- Seja dado um Modelo de Regressão Linear Múltipla de p variáveis preditoras, ajustado com base em n observações. Considere o Coeficiente de Determinação ajustado. Mostre que o coeficiente de determinação ajustado varia entre $-\frac{p}{(n-(p+1))}$ e 1.
- Mostrar que a estatística do teste F parcial comparando um modelo de regressão linear com p preditores e o submodelo Nulo (sem preditores) é igual à estatística do teste F de ajustamento global do modelo completo.