

INSTITUTO SUPERIOR DE AGRONOMIA
ESTATÍSTICA E DELINEAMENTO – 2022-23

28 outubro 2022

Primeiro Teste

Duração: 2h00

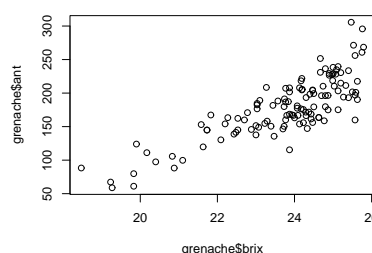
I [16 valores]

Um estudo, sobre a variedade de videira Grenache, pretende modelar o teor de antocianinas no mosto. Trata-se de uma característica cuja análise laboratorial envolve muitos procedimentos. Seria desejável poder modelá-la em função de outra variável com medição mais expedita. Foi sugerido modelar as antocianinas do mosto (variável **ant**, em mg/l) a partir do teor de sólidos solúveis do mosto (grau brix) (variável **brix**). Dispõem-se de valores das duas variáveis, observados em 123 videiras escolhidas ao acaso, sendo conhecidos os seguintes indicadores:

Variável	Mínimo	Média	Máximo	Variância
ant	58.82	179.09	305.64	2034.0242
brix	18.47	23.76	25.79	2.4562

$$cov_{xy} = 57.8842$$

$$r_{xy} = 0.8189$$



1. Descreva em pormenor o modelo de Regressão Linear Simples adequado ao problema sob estudo.
2. Foi, então, ajustada uma regressão linear simples do teor antocianinas do mosto sobre teor de sólidos solúveis (grau brix), tendo-se obtido os seguintes resultados:

Call: `lm(formula = ant ~ brix, data = grenache)`

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept) A	35.755	-10.65	<2e-16	***
brix B	1.501	15.70	<2e-16	***

Residual standard error: 25.99 on 121 degrees of freedom

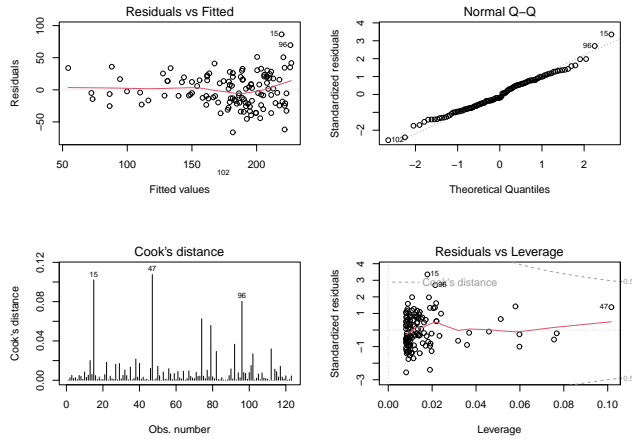
Multiple R-squared: C , Adjusted R-squared: 0.6679

F-statistic: D on 1 and 121 DF, p-value: < 2.2e-16

- (a) Complete a tabela, identificando e calculando os quatro valores em falta (indicados por A, B, C e D). Justifique a sua resposta.
- (b) Verifique que o valor do desvio padrão associado à estimação do declive da recta é 1.501.
- (c) Qual a estimativa da variância dos erros aleatórios do modelo?
- (d) Relativamente à qualidade global do modelo ajustado:
 - i. efectue o teste F de ajustamento global e interprete o valor do coeficiente de determinação R^2 ;
 - ii. mostre que a estatística $F = \frac{QMR}{QMRE}$ do teste de ajustamento global se pode escrever também como

$$F = (n - 2) \cdot \frac{R^2}{1 - R^2}.$$

- (e) Será admissível afirmar que a cada grau brix adicional no mosto corresponde, na população, um aumento médio de 23mg/l no teor de antocianinas? Responda através do teste de hipóteses adequado explicando, em pormenor, todos os seus passos.
- (f) Indique um intervalo a 95% de confiança para o valor esperado do teor de antocianinas em mostos com 22 graus brix.
- (g) Considere os seguintes gráficos:



- i. Descreva os gráficos e discuta as suas implicações para o estudo efectuado.
- ii. Diga, justificando, qual o grau brix na observação correspondente ao ponto 47 (a observação mais à direita no gráfico "Standardized residuals vs Leverage").
- (h) Conhecendo a distribuição de Y_i ao abrigo do modelo de Regressão Linear Simples, mostre que \hat{Y}_i têm a seguinte distribuição:

$$\hat{Y}_i \sim \mathcal{N} \left(\beta_0 + \beta_1 x_i, \sigma^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{(n-1)s_x^2} \right] \right).$$

II [4 valores]

Foi seguidamente ajustada uma regressão linear usando as transformações logarítmicas de ambas as variáveis. Eis os resultados obtidos.

```
Call:lm(formula = log(ant) ~ log(brix), data = grenache)
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-6.2441	0.6145	-10.16	<2e-16 ***
log(brix)	3.5993	0.1941	18.55	<2e-16 ***

```
---
```

```
Residual standard error: 0.1484 on 121 degrees of freedom
```

```
Multiple R-squared: 0.7398, Adjusted R-squared: 0.7377
```

```
F-statistic: 344 on 1 and 121 DF, p-value: < 2.2e-16
```

1. Deduza a relação não linear entre antocianinas e grau brix que corresponde à regressão linear ajustada.
2. Interprete o valor obtido do coeficiente de determinação.
3. Responda, através dum intervalo a 95% de confiança, se considera admissível afirmar que a taxa de variação relativa do teor de antocianinas é três vezes e meia a taxa de variação relativa do grau brix.