

INSTITUTO SUPERIOR DE AGRONOMIA  
**ESTATÍSTICA E DELINEAMENTO – 2021-22**

5 Novembro 2021

Primeiro Teste

Duração: 2h00

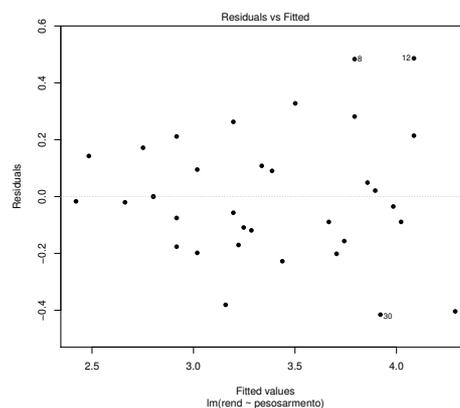
I [16 valores]

Um estudo sobre a variedade de uvas Rabigato, usada na produção de vinhos brancos, pretende relacionar o rendimento (variável **rend**, em kg/planta) com o peso médio por planta dos sarmentos (variável **pesosarmento**, em hectogramas por planta). Dispõem-se de valores das duas variáveis, observados em 34 videiras escolhidas ao acaso, sendo conhecidos os seguintes indicadores:

Variável	Mínimo	Média	Máximo	Desvio Padrão
<b>rend</b>	2.404	3.368	4.572	0.5528155
<b>pesosarmento</b>	1.680	2.425	3.150	0.3978292

$$r_{xy} = 0.9146154$$

1. Considere uma regressão linear do rendimento sobre o peso dos sarmentos.
  - (a) Calcule a recta de regressão ajustada. Interprete o valor calculado do declive da recta.
  - (b) Discuta e teste a qualidade do ajustamento obtido com a recta de regressão.
  - (c) Calcule as três Somas de Quadrados associadas a esta regressão.
  - (d) Qual é o aumento médio previsto no rendimento, associado a um aumento de dez gramas por planta no peso dos sarmentos?
  - (e) Será admissível afirmar que, por cada hectograma por planta adicional de sarmentos, o rendimento médio por planta aumenta, na população, mais de 1 kg? Responda através dum teste de hipóteses adequado, exigindo o ónus da prova à afirmação.
  - (f) Descreva o seguinte gráfico e discuta as suas implicações.



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

Call: `lm(formula = log(rend) ~ log(pesosarmento), data = rabigato)`

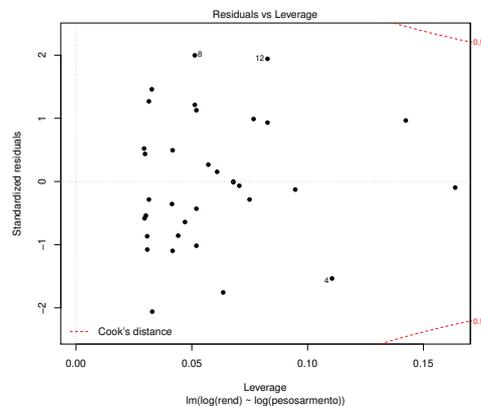
Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.41536	0.05836	7.117	4.48e-08
log(pesosarmento)	0.90085	0.06572	13.707	6.09e-15

---

Residual standard error: 0.06343 on 32 degrees of freedom  
Multiple R-squared: 0.8545, Adjusted R-squared: 0.8499  
F-statistic: 187.9 on 1 and 32 DF, p-value: 6.094e-15

- Comente a seguinte afirmação: “a *logaritmização de ambas as variáveis aumentou a proporção da variabilidade dos rendimentos observados que é explicada pela regressão*”.
- Deduza a relação não linear entre rendimentos e peso dos sarmentos que corresponde à regressão linear agora ajustada.
- Responda, através dum intervalo a 95% de confiança, se considera admissível afirmar que os rendimentos são directamente proporcionais ao peso dos sarmentos.
- Descreva e comente o seguinte gráfico. Sabendo que a média dos log-pesos dos sarmentos observados é 0.8724938, diga, justificando, qual é o peso do sarmento na observação correspondente ao ponto que se encontra mais à direita neste gráfico.



## II [4 valores]

- Considere válido o Modelo de Regressão Linear Simples. Deduza:
  - a distribuição de probabilidades associada a uma observação  $Y_i$  da variável resposta;
  - a variância do estimador do declive da recta.
- Considere a  $i$ -ésima observação utilizada no ajustamento dum Regressão Linear Simples. Seja  $E_i$  o respectivo resíduo (usual),  $D_i$  a sua distância de Cook e  $h_{ii}$  o seu efeito alavanca.
  - Qual a condição que dá ao efeito alavanca o seu menor valor possível?
  - Mostre que a média dos  $n$  efeitos alavanca é dada por  $\bar{h} = \frac{2}{n}$ .
  - Mostre que a estimativa da variância dos erros aleatórios do modelo é dada por  $\frac{E_i^2 h_{ii}}{2D_i(1-h_{ii})^2}$ , qualquer que seja a observação  $i$ .
  - Mostre que, caso a observação  $i$  tenha o valor alavanca médio, a sua contribuição relativa para a Soma de Quadrados dos Resíduos é sempre inferior à sua distância de Cook.