

INSTITUTO SUPERIOR DE AGRONOMIA  
**ESTATÍSTICA E DELINEAMENTO – 2020-21**

4 Novembro 2020

Primeiro Teste

Duração: 2h00

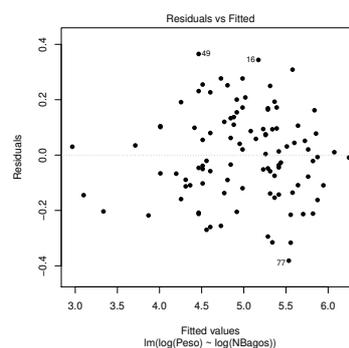
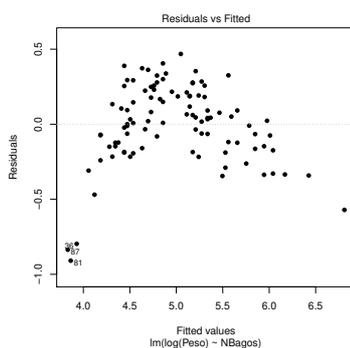
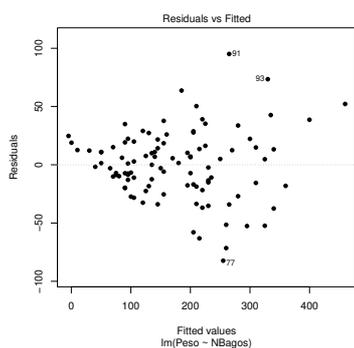
I [15 valores]

Um estudo realizado em 2014 envolvendo a casta Viosinho pretendia modelar o peso dos cachos, uma variável importante na produção. Um robô com câmara percorreu uma vinha experimental do ISA capturando imagens de 100 cachos. Esses cachos foram seguidamente colhidos e pesados (variável **Peso**, em *g*). Nas imagens recolhidas há duas potenciais variáveis predictoras do peso do cacho: o número de bagos visível no cacho (variável **NBagos**) e a área visível do cacho (variável **Area**, em *cm*<sup>2</sup>). Eis alguns indicadores, bem como a matriz de correlações, para as variáveis observadas.

	Peso	NBagos	Area
Mínimo	19.20	10.00	21.92
Máximo	512.00	103.00	210.45
Média	175.84	46.15	101.02
Desvio Padrão	98.24383	18.74355	39.55328

	Peso	NBagos	Area
Peso	1.0000	0.9530	0.8893
NBagos	0.9530	1.0000	0.9217
Area	0.8893	0.9217	1.0000

- Decidiu-se ajustar um modelo de regressão linear simples para prever o peso dos cachos.
  - Diga, justificando, qual a melhor variável predictora e indique a proporção da variabilidade observada nos pesos dos cachos que é explicada por esse modelo. Teste se essa proporção é significativamente diferente de zero. Comente.
  - Ajuste a recta de regressão com o predictor que escolheu na alínea anterior. Quais as unidades de medida e o significado biológico do declive da recta ajustada?
- Decidiu-se modelar o peso dos cachos (**Peso**) como função do número de bagos visível nas imagens (**NBagos**), mas admitiu-se a possibilidade de transformar variáveis. Para ajudar na escolha do modelo construíram-se gráficos de resíduos (usuais) contra valores ajustados para 3 modelos: à esquerda o modelo sem qualquer transformação; no meio logaritmando apenas o **Peso**; à direita logaritmando quer **Peso**, quer **Nbagos**. Discuta as implicações destes gráficos.



- Foi seguidamente ajustada uma regressão linear do logaritmo (natural) do peso do cacho sobre o logaritmo do número de bagos visível em cada cacho, tendo-se obtido os seguintes resultados:

```
Call: lm(formula = log(Peso) ~ log(NBagos), data = viosinho2014)
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.27446	0.13686	-2.005	0.0477
log(NBagos)	1.40715	0.03633	38.736	<2e-16

---

Residual standard error: 0.163 on 98 degrees of freedom  
Multiple R-squared: 0.9387, Adjusted R-squared: 0.9381  
F-statistic: 1500 on 1 and 98 DF, p-value: < 2.2e-16

- (a) Qual o valor estimado da variância dos erros aleatórios do modelo?
  - (b) O intervalo de predição (95%) para uma observação da variável resposta deste modelo, correspondente a um cacho com 50 bagos visíveis, é ] 4.9051, 5.5556 [.
    - i. Qual a estimativa  $\hat{\sigma}_{indiv}$  da variabilidade associada a esta predição?
    - ii. Interprete este intervalo em termos do peso dos cachos (em g).
  - (c) É possível descrever a relação de fundo, na população, afirmando que o quadrado do peso dos cachos é proporcional ao cubo do número de bagos na imagem? Responda através dum intervalo de confiança adequado. Comente.
4. Foi finalmente ajustado um modelo de regressão linear múltipla com as transformações logarítmicas da variável resposta e dos dois preditores referidos. Eis os resultados:

```
Call: lm(formula = log(Peso) ~ log(NBagos) + log(Area), data = viosinho2014)
Coefficients:
(Intercept)  log(NBagos)  log(Area)
-0.6237      1.0366      0.3835
```

- (a) O que se pode afirmar sobre o coeficiente de determinação deste modelo ajustado?
- (b) Sabendo que o cacho com o menor peso tinha, visíveis na sua imagem, 11 bagos e área  $24.22 \text{ cm}^2$ , calcule o respectivo resíduo.
- (c) A qual relação ajustada entre as variáveis originais (não logaritmizadas) é que corresponde o modelo agora ajustado? Justifique. Calcule a diferença entre o peso observado e o previsto por esta relação não linear, para o cacho de menor peso. Comente.

## II [5 valores]

1. No estudo da relação entre o rendimento por planta  $y$  e a densidade  $x$  duma cultura, é por vezes usada a equação  $y = \frac{1}{c+dx}$ . Mostre que um tal modelo corresponde a admitir que a taxa de variação relativa de  $y(x)$  é proporcional a  $y(x)$ . Indique a transformação linearizante que permite estudar este tipo de relação através duma regressão linear simples.
2. Considere o modelo de regressão linear simples, a ajustar com  $n$  pares de observações.
  - (a) Enuncie o modelo.
  - (b) Deduza a distribuição de probabilidades das observações  $Y_i$  da variável resposta.
  - (c) Tendo-se  $\frac{SQRE}{\sigma^2} \sim \chi_{n-2}^2$ , deduza um estimador centrado da variância dos erros aleatórios.
3. Considere uma regressão linear múltipla em contexto descritivo, com  $p$  preditores e ajustada com base em  $n$  conjuntos de observações.
  - (a) O que se entende por matriz do modelo  $\mathbf{X}$  e o seu subespaço das colunas  $\mathcal{C}(\mathbf{X})$ ?
  - (b) Mostre que a matriz de projecção ortogonal sobre  $\mathcal{C}(\mathbf{X})$  é simétrica e idempotente.