

INSTITUTO SUPERIOR DE AGRONOMIA
ESTATÍSTICA E DELINEAMENTO – 2019-20

15 Novembro 2019

Primeiro Teste

Duração: 2h30

I [4 valores]

Num estudo sobre afinidade de enxertia da variedade de videira Rabigato, foram usados 2 porta-enxertos (designados 99R e 110R). Foram enxertadas e plantadas em viveiro 1885 plantas no porta-enxerto 99R e 1972 plantas no porta-enxerto 110R. No levantamento do viveiro foram contadas as plantas em que a enxertia foi bem sucedida e que estão em condições de serem comercializadas (designadas “vivas”) e as plantas em que a enxertia não foi bem sucedida (designadas “mortas”). Os resultados constam do quadro em baixo. Teste formalmente se é possível concluir que o sucesso da enxertia em viveiro é diferente nos dois porta-enxertos. Justifique pormenorizadamente os seus passos e comente os seus resultados. Com base nas tabelas disponíveis, o que pode afirmar sobre o valor de prova (*p-value*) associado ao valor calculado da estatística do teste?

	vivas	mortas	total plantado
99R	463	1422	1885
110R	690	1282	1972

II [12 valores]

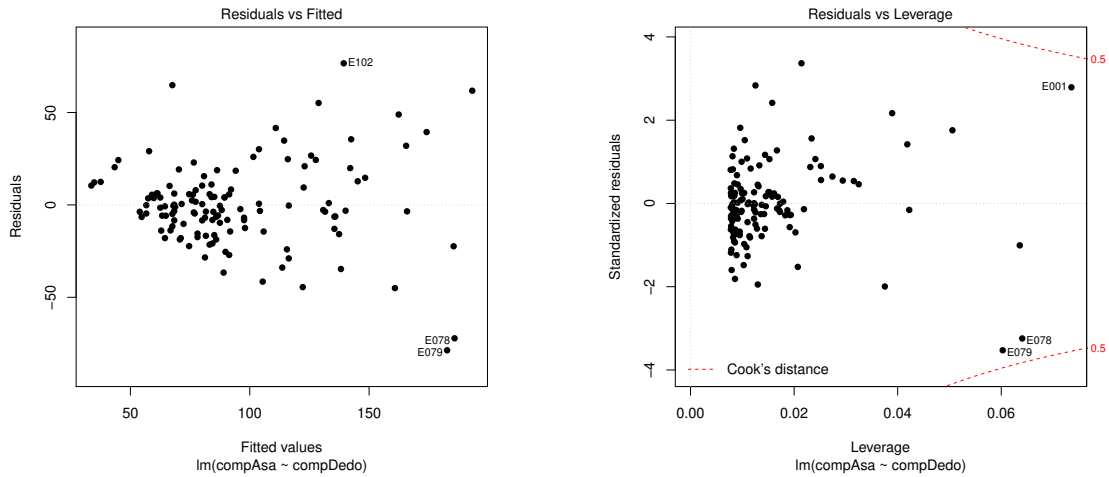
Num estudo sobre morfologia de pássaros recolheram-se dados relativos aos valores médios de várias variáveis morfológicas em 129 espécies de pássaros. Pretende-se modelar o comprimento da asa (variável *compAsa*, em mm). Sete variáveis são potenciais preditores: o comprimento da cauda (variável *compCauda*, em mm); o comprimento do bico (variável *compBico*, em mm); a altura do bico (variável *altBico*, em mm); a largura do bico (variável *largBico*, em mm); a altura do tarso (variável *altTarso*, em mm); o comprimento do dedo média da pata (variável *compDedo*, em mm); e o peso dos pássaros (variável *peso*, em g). Eis alguns indicadores, bem como a matriz de correlações amostrais entre as variáveis:

Indicador	<i>compAsa</i>	<i>compCauda</i>	<i>compBico</i>	<i>altBico</i>	<i>largBico</i>	<i>altTarso</i>	<i>compDedo</i>	<i>peso</i>
min.	44.1	25.2	7.1	2.3	2.7	2.5	6.1	3.1
média	94.2814	78.4318	19.1884	5.9922	7.7643	22.2442	21.6434	54.5558
max.	255.0	259.5	45.5	16.2	18.4	50.0	47.0	500.0
desv.pad.	41.0781	36.3586	8.4208	2.8553	3.2353	8.1706	8.7377	80.3640

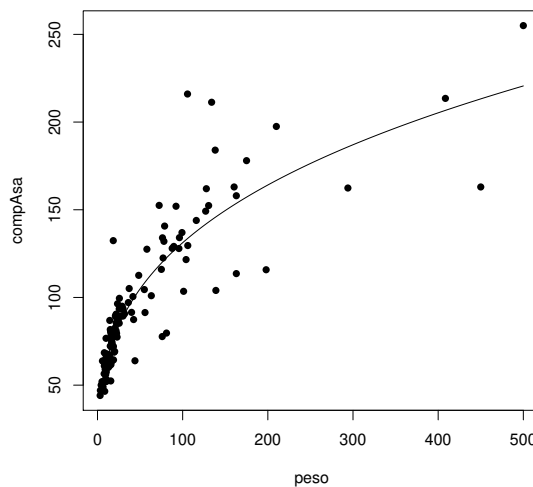
	<i>compAsa</i>	<i>compCauda</i>	<i>compBico</i>	<i>altBico</i>	<i>largBico</i>	<i>altTarso</i>	<i>compDedo</i>	<i>peso</i>
<i>compAsa</i>	1.000	0.817	0.729	0.640	0.786	0.562	0.830	0.809
<i>compCauda</i>	0.817	1.000	0.589	0.539	0.682	0.649	0.731	0.638
<i>compBico</i>	0.729	0.589	1.000	0.532	0.668	0.488	0.651	0.524
<i>altBico</i>	0.640	0.539	0.532	1.000	0.825	0.574	0.693	0.536
<i>largBico</i>	0.786	0.682	0.668	0.825	1.000	0.669	0.759	0.608
<i>altTarso</i>	0.562	0.649	0.488	0.574	0.669	1.000	0.822	0.552
<i>compDedo</i>	0.830	0.731	0.651	0.693	0.759	0.822	1.000	0.835
<i>peso</i>	0.809	0.638	0.524	0.536	0.608	0.552	0.835	1.000

1. Com base na informação disponível identifique, justificando, o melhor preditor para prever o comprimento da asa através duma regressão linear simples. Para a variável que escolheu, calcule e comente:

- (a) a recta de regressão ajustada e o coeficiente de determinação resultante;
 - (b) o valor das três Somas de Quadrados associadas a essa regressão;
 - (c) um intervalo a 95% de confiança para o valor esperado do comprimento da asa em espécies em que o preditor que escolheu toma o valor 15.
2. Uma regressão linear simples do comprimento da asa sobre o comprimento do dedo médio da pata produziu os seguintes gráficos de resíduos e diagnósticos.



- (a) Descreva e comente estes gráficos.
 - (b) Calcule um valor aproximado da distância de Cook correspondente à observação E079 (no canto inferior direito do gráfico da direita). Comente o seu valor.
3. Pretende-se seguidamente estabelecer uma relação entre o comprimento da asa e o peso dos pássaros, cuja nuvem de pontos se encontra no gráfico seguinte. A curva visível no gráfico resultou de ajustar uma regressão linear relacionando os logaritmos de ambas as variáveis, regressão cujos resultados são indicados mais abaixo.



```

Call: lm(formula = log(compAsa) ~ log(peso), data = passaros)
[...]
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.39058    0.04371   77.58  <2e-16
log(peso)    0.32275    0.01245   25.92  <2e-16
---
Residual standard error: 0.1552 on 127 degrees of freedom
Multiple R-squared:  0.8411, Adjusted R-squared:  0.8398
F-statistic:  672 on 1 and 127 DF,  p-value: < 2.2e-16

```

- (a) Teste, e comente, a qualidade de ajustamento do modelo. Em particular, comente a seguinte frase: “*o modelo ajustado explica quase 85% da variabilidade observada nos comprimentos das asas das espécies*”.
- (b) Determine, justificando, a equação da curva indicada no gráfico acima.
- (c) Teste a hipótese de que o comprimento da asa seja proporcional à raiz cúbica do peso. Esta hipótese corresponde a que relação entre as taxas de variação relativa (em relação ao tempo) das duas variáveis?
- (d) Calcule o resíduo usual, na regressão linear ajustada, da observação correspondente ao ponto no canto superior direito do gráfico. Diga, justificando, qual a distância (na vertical) entre o ponto e a curva ajustada, no gráfico acima.

III [4 valores]

1. Considere uma regressão linear simples, ajustada com base em n pares de observações $\{(x_i, y_i)\}_{i=1}^n$, e admita válido o Modelo de Regressão Linear Simples.
 - (a) Deduza a distribuição do estimador do valor esperado de Y quando o preditor toma o valor x , ou seja, de $\hat{\mu}_{Y|x} = \hat{\beta}_0 + \hat{\beta}_1 x$.
 - (b) Determine uma condição simples, envolvendo a soma dos quadrados dos valores x_i , que garanta que o intervalo de confiança para β_1 seja de menor amplitude do que o intervalo de confiança (ao mesmo nível de confiança) para β_0 . A sua condição depende das unidades de medida da variável preditora? Comente.
2. Considere uma regressão linear múltipla, ajustada com base em n observações.
 - (a) Defina a matriz do modelo \mathbf{X} e o seu subespaço das colunas, $\mathcal{C}(\mathbf{X})$.
 - (b) Mostre que são iguais as médias dos n valores observados y_i e dos n valores ajustados \hat{y}_i .
 - (c) A partir do resultado da alínea anterior, mostre que o centro de gravidade da nuvem de pontos pertence necessariamente ao hiperplano de mínimos quadrados.