INSTITUTO SUPERIOR DE AGRONOMIA

ESTATÍSTICA E DELINEAMENTO – 2017-18

3 de Novembro de 2017 **PRIMEIRO TESTE** **Duração: 2h30**

I [3,5 valores]

Num ensaio sobre a variedade de macieira Bravo de Esmolfe em três diferentes terrenos, comparou-se a produtividade (em número de frutos, um ano após a plantação) de macieiras de três diferentes proveniências, a fim de determinar se haveria uma eventual relação entre proveniência e tipos de terreno. As proveniências eram a Beira Litoral, a Beira Interior e Trás-os-Montes. A fim de garantir a comparabilidade de resultados, foram demarcados três diferentes terrenos com igual dimensão e igual número de *árvores* de cada proveniência. Observou-se o número de *frutos* de cada proveniência, num total de 1262 frutos, com os resultados indicados na tabela seguinte:

Proveniência	Terreno 1	Terreno 2	Terreno 3
Beira Litoral	85	137	186
Beira Interior	76	112	214
Trás-os-Montes	67	140	245

1. Qual a probabilidade estimada de um fruto escolhido ao acaso ser: (i) duma árvore proveniente de Trás-os-Montes; e (ii) proveniente do Terreno 1?

2. Sabendo que o valor calculado da estatística de Pearson adequada ao problema sob estudo é $X^2_{calc} = 10.305$, identifique e concretize os passos do teste que considere adequado para testar se há relação entre proveniência e tipos de terreno. Utilize os níveis de significância: $\alpha = 0.05$ e $\alpha = 0.01$. Comente. O que pode dizer sobre o valor de prova (*p-value*) correspondente?

3. Qual a contribuição dos frutos provenientes de Trás-os-Montes, no Terreno 1, para o valor calculado da estatística de teste?

II [11,5 valores]

Um estudo sobre vigor vegetativo da casta Touriga Nacional, realizado em Nelas, avaliou o peso médio da lenha de poda (variável `peso`, em *kg*) por videira, e o número médio de lançamentos por videira (variável `Nlancamentos`), em 238 parcelas de terreno. Eis alguns indicadores obtidos:

Variável	Mínimo	Máximo	Média	Variância
`Nlancamentos`	1.750	22.000	6.049	???
`peso`	0.0630	1.5200	0.3258	0.0639068

Foi decidido ajustar uma recta de regressão para modelar a variável `peso` a partir do número de lançamentos. Foram obtidos os seguintes resultados:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.001227   0.013809   0.089    0.929
Nlancamentos 0.053667   0.001884    ???    <2e-16
---
Residual standard error: 0.1203 on ??? degrees of freedom
Multiple R-squared:  0.7746,   Adjusted R-squared:  0.7737
F-statistic: ??? on 1 and ??? DF,  p-value: < 2.2e-16
```

1. Complete a tabela de resultados da regressão, indicando como obtém os quatro valores em falta e diga, justificando, qual a variância amostral do número de lançamentos (variável `Nlancamentos`).
2. Discuta, de forma pormenorizada, a qualidade do modelo de regressão linear ajustado. Comente.
3. Será admissível considerar que a recta de regressão populacional passa na origem? Responda através dum intervalo de confiança adequado.
4. Construa um intervalo de predição (95%) para uma observação de `peso`, numa parcela individual cujo número médio de lançamentos seja 10. **Nota:** Se não resolveu o ponto 1, use o valor $s_x^2 = 17$.
5. Foi seguidamente ajustado um modelo linear relacionando os logaritmos das duas variáveis referidas, tendo-se obtido os seguintes resultados:

```
Call: lm(formula = log(peso) ~ log(Nlancamentos), data = poda)
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       -2.96320    0.05508   -53.8   <2e-16
log(Nlancamentos)  1.00085    0.03208    31.2   <2e-16
---
Residual standard error: 0.2871 on 236 degrees of freedom
Multiple R-squared:  0.8049,    Adjusted R-squared:  0.804
F-statistic: 973.4 on 1 and 236 DF,  p-value: < 2.2e-16
```

   (a) Comente a seguinte afirmação: "*Comparando os coeficientes de determinação dos dois modelos ajustados, conclui-se que este modelo explica melhor a variabilidade observada nos pesos da lenha de poda*".

   (b) A que tipo de relação não linear entre `peso` e `Nlancamentos` corresponde a regressão linear agora ajustada? Calcule a equação da curva ajustada, relacionando as duas variáveis originais. Qual a relação estimada entre as taxas de variação relativas de `peso` e `Nlancamentos`?

   (c) Comente, com base num teste de hipóteses adequado, a seguinte afirmação: "*os dois modelos ajustados sugerem que o peso médio da lenha de poda é aproximadamente proporcional ao número médio de lançamentos por videira*".

III [5 valores]

Seja dado o modelo de regressão linear simples, ajustado com n pares de observações $\{(x_i, Y_i)\}_{i=1}^n$.

1. Partindo do resultado $\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \cap t_{n-2}$, construa um intervalo a $(1-\alpha) \times 100\%$ de confiança para a ordenada na origem da recta populacional.
2. Com base nos resultados disponíveis, relativos aos estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$, mostre que a variância do estimador do valor esperado de Y, dado $X = x$, é dada por $V[\hat{\mu}_{Y|x}] = \sigma^2 \left[\frac{1}{n} + \frac{(x-\overline{x})^2}{(n-1)\, s_x^2}\right]$.
3. Formule e resolva a equação diferencial a que corresponde uma relação exponencial entre Y (variável resposta) e X (preditor).
4. Com base nos resultados deduzidos nas aulas, mostre que a distribuição de probabilidades da covariância amostral $cov_{xY} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(Y_i - \overline{Y})$ é $\mathcal{N}(\beta_1 s_x^2 \, , \, \frac{\sigma^2}{n-1} s_x^2)$.