

INSTITUTO SUPERIOR DE AGRONOMIA
ESTATÍSTICA E DELINEAMENTO – 2016-17
PRIMEIRO TESTE

31 de Outubro de 2016

Duração: 2h30

I [4 valores]

A casta Tinta Barroca é, nas condições do Douro Superior, sensível ao *stress* abiótico (calor e seca). Uma experiência visou aprofundar o estudo da tolerância da casta a esse *stress* (dizendo-se que há tolerância quando menos de 25% dos cachos secam). Foram seleccionados ao acaso 67 genótipos da casta, cada um dos quais foi ensaiado em 3 parcelas (unidades experimentais). No final, contou-se, para cada genótipo, o número de parcelas onde houve tolerância ao *stress* abiótico. Os resultados são dados no quadro seguinte.

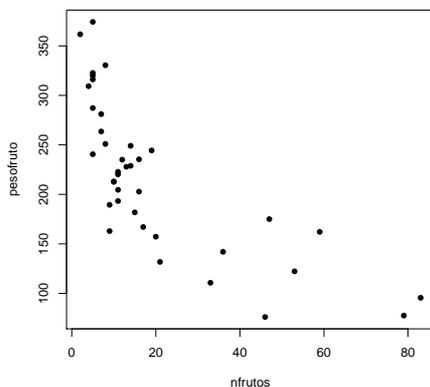
No. de parcelas	0	1	2	3
No. de genótipos	22	26	17	2

Pretende-se saber se é admissível considerar que o número de parcelas onde a casta tolera o *stress* segue uma distribuição Binomial.

1. Indique, justificando, quais os parâmetros apropriados para a distribuição Binomial em causa.
2. Descreva o teste ao ajustamento da distribuição Binomial. Sabendo que o valor calculado da estatística do teste, com base na tabela apresentada, é $X^2_{calc} = 1.0894$, e admitindo válida a distribuição assintótica da estatística do teste, qual seria a sua conclusão?
3. Calcule o valor esperado ao abrigo da hipótese de distribuição Binomial, bem como a parcela da estatística do teste, correspondente ao resultado 3 (em todas as parcelas houve tolerância ao *stress*). Comente.

II [11 valores]

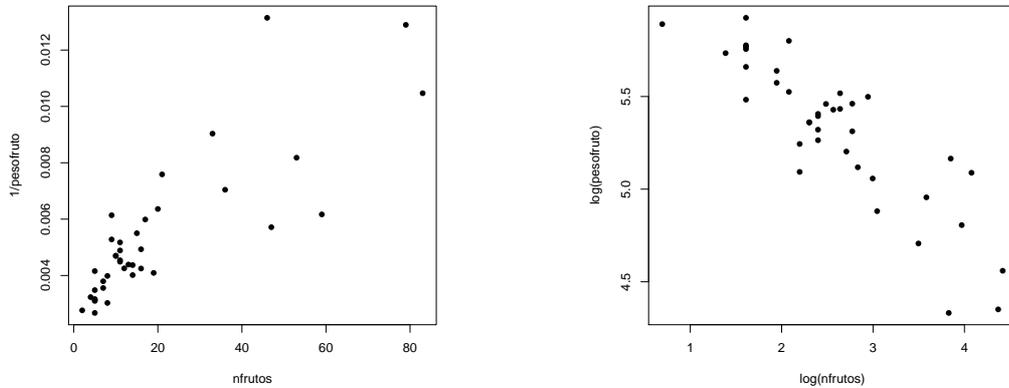
Num estudo de variedades antigas de tomateiro, pretende-se relacionar o número de frutos comerciais por planta (variável preditora **nfrutos**, x) com o peso médio dos frutos da planta (variável resposta **pesofruto**, y , em g). Para tal, foram seleccionados aleatoriamente 39 tomateiros, tendo sido registados os valores de cada variável em cada uma das plantas. Eis a nuvem de pontos obtida, bem como alguns dos valores observados e indicadores relevantes:



obs.	nfrutos	pesofruto
1	5	316.1218
2	9	162.8988
3	4	309.2794
4	17	167.0068
⋮	⋮	⋮
média	19.51282	217.91564
desvio padrão	20.16755	75.77446

$$r_{xy} = -0.744526$$

1. Foram construídos dois gráficos envolvendo transformações das variáveis, e nos quais se observa uma tendência linear de fundo. Em baixo, à esquerda encontra-se um gráfico de $y^* = \frac{1}{y}$ contra $x^* = x$. À direita, encontra-se o gráfico de $y^* = \ln(y)$ contra $x^* = \ln(x)$.



- (a) Admitindo a relação linear de fundo $y^* = b_0 + b_1 x^*$ em cada um destes gráficos indique, justificando algebricamente, qual a relação original entre o peso dos frutos (y) e o número de frutos (x) em cada caso.
- (b) Se tivesse que escolher uma das duas transformações linearizantes, para aplicar o modelo de regressão linear simples para efeitos inferenciais, qual escolheria? Justifique.
2. Independentemente da sua resposta na alínea anterior, decidiu-se aplicar o modelo de regressão linear simples à transformação correspondente ao gráfico da direita, tendo sido obtidos os seguintes resultados:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.32931	0.09869	64.14	< 2e-16
log(nfrutos)	-0.39386	0.03636	-10.83	4.96e-13

Residual standard error: 0.1958 on 37 degrees of freedom

Multiple R-squared: 0.7603, Adjusted R-squared: 0.7538

F-statistic: ??? on 1 and 37 DF, p-value: 4.965e-13

- (a) Discuta pormenorizadamente a qualidade de ajustamento do modelo. Em particular, indique como obtém o valor calculado da estatística do teste de ajustamento global.
- (b) Qual o coeficiente de correlação entre as variáveis logaritimizadas? Justifique.
- (c) Interprete o significado da estimativa -0.39386 , quer em termos da relação linearizada, quer em termos da relação entre as variáveis originais.
- (d) Sabendo que a média dos log- pesos observados foi $\overline{y^*} = 5.3157$ e que a variância dos log- pesos observados foi $s_{y^*}^2 = 0.155667$, justifique que a média e a variância dos log- números de frutos observados são, respectivamente $\overline{x^*} = 2.5735$ e $s_{x^*}^2 = 0.76295$. Caso necessite, pode admitir conhecidos estes valores nas alíneas seguintes.
- (e) Construa um intervalo a 95% de confiança para o declive da recta que, na população, relaciona o log-peso dos frutos e o log-número de frutos. Comente.
- (f) Calcule um intervalo de predição (a 95%) para o peso do fruto (em g) dum tomateiro com 20 frutos. Comente.

III [5 valores]

Responda às seguintes perguntas, admitindo conhecidas as restantes fórmulas deduzidas nas aulas.

1. Considere dado o modelo de regressão linear simples, em contexto inferencial.
 - (a) Mostre que a variância estimada do estimador $\hat{\beta}_1$ é dada pelo quadrado do declive da recta, a dividir pela estatística calculada do teste de ajustamento global, ou seja: $\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{b_1^2}{F_{calc}}$.
 - (b) Mostre que a distribuição do estimador $\hat{\beta}_1 = \sum_{i=1}^n c_i Y_i$ é $\mathcal{N}\left(\beta_1, \frac{\sigma^2}{(n-1)s_x^2}\right)$. Interprete esse resultado. **NOTA:** Pode admitir que $\sum_{i=1}^n c_i x_i = 1$.
2. Considere uma regressão linear simples em contexto descritivo.
 - (a) Mostre que o coeficiente de determinação R^2 é igual ao quadrado do coeficiente de correlação entre a variável preditora x e a variável resposta y .
 - (b) Mostre que o quadrado do coeficiente de correlação entre os n valores observados y_i e os n correspondentes valores ajustados, \hat{y}_i , é também igual ao coeficiente de determinação: $(r_{y\hat{y}})^2 = R^2$.