

INSTITUTO SUPERIOR DE AGRONOMIA
Modelos Matemáticos e Aplicações (2018-19)

Teste Modelos Lineares Generalizados e Modelos Lineares Mistos

11 de Junho, 2019

Duração: 2h30

I [10 valores]

Em 2004 realizou-se em Merelim um estudo sobre características de 596 dentes de alho. Foram analisadas três características numéricas dos dentes: *Altura*, *Largura* e *Espessura*, todas em cm. Para cada dente de alho, foi ainda classificada a cor do tegumento (a película que envolve os dentes de alho), numa de quatro categorias: branco (associado ao número 1), castanho (3), vermelho (4) e violeta (5). Eis algumas estatísticas de síntese:

```
> summary(AlhoDentes)
  Altura      Largura      Espessura      CorTegum
Min.   :1.600  Min.   :0.70  Min.   :0.800  1: 20
1st Qu.:2.900  1st Qu.:1.10  1st Qu.:1.375  3: 77
Median :3.100  Median :1.30  Median :1.600  4:299
Mean   :3.124  Mean   :1.35  Mean   :1.541  5:200
3rd Qu.:3.400  3rd Qu.:1.50  3rd Qu.:1.700
Max.   :4.100  Max.   :3.90  Max.   :2.500

> round(cor(AlhoDentes[,4]), d=3)
           Altura Largura Espessura
Altura    1.000  -0.094  0.397
Largura   -0.094  1.000  0.073
Espessura 0.397  0.073  1.000
```

1. Ajustou-se uma Regressão Logística relacionando a cor vermelha do tegumento com as variáveis numéricas observadas. Seguidamente indicam-se alguns dos resultados obtidos. Os pontos dispostos ao longo duma curva no gráfico relacionam as *probabilidades* estimadas com os diferentes valores da combinação linear dos preditores obtida com o ajustamento do modelo, enquanto que os pontos *dispostos ao longo de rectas horizontais* indicam, para cada valor da componente sistemática no eixo horizontal, se a cor observada do tegumento era, ou não, vermelha.

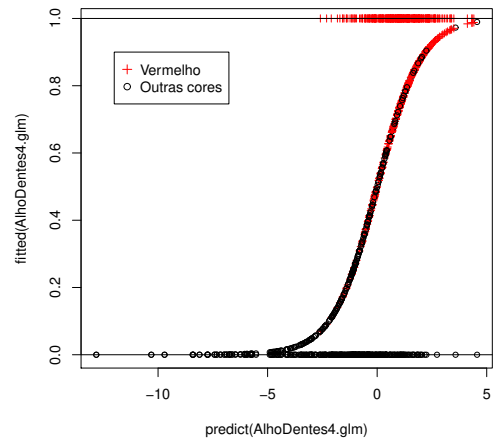
```
> summary(AlhoDentes4.glm)
```

```
Call:
glm(formula = (AlhoDentes$CorTegum == "4") ~ .,
     family = binomial, data = AlhoDentes)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.0236  -0.7623   0.1583   0.7653   2.3085
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.44800    1.06956  -0.419  0.675
Altura       2.26906    0.33517   6.770 1.29e-11
Largura     -5.18047    0.50260 -10.307 < 2e-16
Espessura    0.05124    0.40571   0.126  0.899
```

```
---
Null deviance: 826.22 on 595 degrees of freedom
Residual deviance: 555.61 on 592 degrees of freedom
AIC: 563.61
```



- (a) Comente a qualidade do modelo agora ajustado. Em particular, teste se o modelo agora ajustado tem um ajustamento significativamente melhor do que um modelo sem preditores, que apenas atribui uma probabilidade constante a que um dado dente de alho tenha tegumento vermelho.
- (b) Qual a probabilidade estimada de um dente de alho com 4cm de altura, 1cm de largura e 2cm de espessura ter tegumento de cor vermelha? Indique as contas que efectuar.
- (c) Interprete o valor da estimativa -5.18047 , associada ao preditor *Largura*.
- (d) O menor resíduo do desvio observado foi -3.0236 . Indique, justificando:
 - i. Se está, ou não, associado a um dente com tegumento vermelho.

- ii. Qual a probabilidade estimada pelo modelo para que o respectivo tegumento fosse vermelho.
 - (e) Um investigador ajustou seguidamente um modelo análogo, mas retirando o preditor **Espessura**.
 - i. Justifique, com base na informação disponível até aqui, se considera que se trata duma boa opção.
 - ii. O valor do Critério de Informação de Akaike deste novo modelo resultou ser 561.63. Comente-o.
 - iii. Comente a seguinte afirmação: “*A altura e espessura dos dentes de alho determinam a cor vermelha do tegumento*”.
2. Considere agora o modelo nulo, ou seja, uma Regressão Logística que, a cada dente de alho, associa uma probabilidade constante de ter tegumento vermelho. Mostre que o estimador de máxima verosimilhança para o (único) parâmetro do modelo, β_0 , é dado por $\hat{\beta}_0 = \ln\left(\frac{\bar{y}}{1-\bar{y}}\right)$, onde \bar{y} indica a média das n observações da variável resposta. Mostre ainda que a probabilidade constante estimada resultante é \bar{y} . Comente.

II [10 valores]

1. Com vista a estudar a variabilidade genética do peso do bolbo (g) em genótipos tradicionais de alho, avaliaram-se 30 genótipos (acessos) quanto a essa característica em 5 locais (Apúlia, Elvas, Faro, Merelim e Vairão). Em cada local foi instalado um ensaio em que os 30 genótipos foram dispostos de acordo com um delineamento experimental totalmente casualizado com 4 repetições (isto é, em cada local existem 4 observações de cada um dos genótipos). Considere o genótipo como um factor de efeitos aleatórios e o local como um factor de efeitos fixos.

- (a) Descreva, em pormenor, o modelo que lhe parece ser adequado para o estudo acima descrito.
- (b) No R, com a função `lmer` do pacote `lme4` executaram-se os seguintes comandos:

```
> alho <- read.table("alho.txt", header=T)
> alholmer1 <- lmer(pesobolbo~local+(1|acesso)+(1|acesso:local), data=alho)

> summary(alholmer1)
Linear mixed model fit by REML ['lmerMod']
Formula: pesobolbo ~ local + (1 | acesso) + (1 | acesso:local)
Data: alho
REML criterion at convergence: 4399.5
Random effects:
 Groups      Name      Variance Std.Dev.
acesso:local (Intercept) 55.57    7.454
acesso      (Intercept) 18.66    4.319
Residual                    60.25    7.762
Number of obs: 600, groups:  acesso:local, 150;  acesso, 30
Fixed effects:
              Estimate Std. Error t value
(Intercept)   36.124      1.725  20.939
localElvas   -12.373      2.170  -5.702
localFaro    -21.992      2.170 -10.135
localMerelim  0.137       2.170  0.063
localVairao  -10.795      2.170  -4.975

> anova(alholmer1)
Analysis of Variance Table
              Df Sum Sq Mean Sq F value
local      4 8902.5  2225.6  36.941

> logLik(alholmer1)
'log Lik.' -2199.77
```

```

> alholmer2 <- lmer(pesobolbo~local+(1|acesso), data=alho)
> logLik(alholmer2)
'log Lik.' -2269.503

> alholmer3 <- lmer(pesobolbo~local+(1|acesso:local), data=alho)
> logLik(alholmer3)
'log Lik.' -2204.556

> ranef(alholmer1)
$'acesso:local'
(Intercept)
A06850:Apulia   -5.84391413
A06850:Elvas    1.85129702
A06850:Faro     4.11438616
A06850:Merelim -1.02270424
A06850:Vairao   6.12453487
A06887:Apulia  -1.87304751
...

$acesso
(Intercept)
A06850  1.75374180
A06887 -0.59092917
A06896  0.74985841
A06902 -0.88856580
A06924  4.17950880
...

```

- i. Quais as estimativas de máxima verosimilhança restrita das componentes de variância do modelo? Se o método de estimação fosse o de máxima verosimilhança, qual seria o resultado esperado nos valores das estimativas dessas componentes de variância? Justifique.
- ii. O que conclui sobre a existência de variabilidade da interação genótipo×local? Enuncie detalhadamente o teste de hipóteses que efectuar.
- iii. O que conclui sobre a existência de variabilidade dos genótipos quanto ao peso do bolbo? Descreva resumidamente o teste de hipóteses que efectuar.
- iv. De acordo com o modelo ajustado, qual o peso do bolbo previsto para o genótipo A06850 em Apúlia?

2. Considere o modelo linear misto em notação matricial:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}.$$

\mathbf{Y} é o vector aleatório $n \times 1$ de observações da variável resposta; \mathbf{X} é a matriz de delineamento $n \times p$ dos efeitos fixos; $\boldsymbol{\beta}$ é o vector $p \times 1$ de efeitos fixos; \mathbf{Z} é a matriz de delineamento $n \times q$ dos efeitos aleatórios; \mathbf{u} é o vector $q \times 1$ de efeitos aleatórios; \mathbf{e} é o vector $n \times 1$ de erros aleatórios. Os vectores \mathbf{u} e \mathbf{e} admitem-se independentes, com distribuição normal multivariada de vector de valores médios nulo e matrizes de covariâncias \mathbf{G} e \mathbf{R} , respectivamente.

- (a) Prove que \mathbf{Y} tem distribuição normal multivariada com vector de valores médios $\mathbf{X}\boldsymbol{\beta}$ e matriz de covariâncias $\mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$.
- (b) Considere um modelo a 1 factor de efeitos aleatórios. Defina as matrizes \mathbf{G} e \mathbf{R} , admitindo efeitos aleatórios independentes e identicamente distribuídos e erros aleatórios independentes e identicamente distribuídos.