

INSTITUTO SUPERIOR DE AGRONOMIA
Modelos Matemáticos e Aplicações (2019-20)
Teste – Modelos Lineares Generalizados

11 de Maio, 2020

Duração: 1h30

I [16 valores]

Um estudo publicado numa revista de zoologia¹ inclui dados relativos ao número de espécies de bivalves de água doce existentes em 44 rios dos EUA. Admitindo que o número de espécies (variável `n.especies`) em cada rio segue uma distribuição de Poisson, modelou-se o número de espécies através dum modelo log-linear com quatro preditores numéricos: para cada rio, a área abrangida (variável `area`, em milhas quadradas); a concentração de hidrónio (o catião que resulta do ião hidrogénio em meio aquoso, variável `Hd`); a concentração de sólidos dissolvidos, ou seja, resíduos sólidos (variável `rs`); e a concentração de nitratos (variável `nitratos`). Obtiveram-se os seguintes resultados:

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.545e+00	1.063e-01	23.943	< 2e-16
area	3.129e-05	7.022e-06	4.456	8.33e-06
rs	-2.317e-03	6.184e-04	-3.747	0.000179
Hd	-3.783e-02	1.186e-02	-3.190	0.001424
nitrato	1.906e-04	2.796e-02	0.007	0.994559

Null deviance: 127.527 on 43 degrees of freedom
Residual deviance: 70.676 on 39 degrees of freedom
AIC: 262.7

1. Responda, justificando adequadamente:

- (a) Qual a equação ajustada pelo modelo para o número esperado de espécies de bivalves?
- (b) Discuta formalmente a qualidade global de ajustamento do modelo.
- (c) Diga, justificando com base na teoria assintótica dos estimadores de máxima verosimilhança, se é possível simplificar o modelo sem perda significativa da qualidade de ajustamento.

2. Foi seguidamente ajustado um modelo análogo, mas com apenas um preditor numérico: a `area` dos rios. O Desvio obtido foi 96.192. Teste se a qualidade deste submodelo deve ser considerada inferior à do modelo original.

3. Foi agora ajustado um modelo que apenas difere do modelo da alínea 2 pelo facto de se admitir uma diferente função de ligação: a função identidade. Eis os resultados obtidos:

```
> update(mexilhoes.glm2, family=poisson(link=identity))
```

Coefficients:

(Intercept)	area
7.4695664	0.0005737

Degrees of Freedom: 43 Total (i.e. Null); 42 Residual

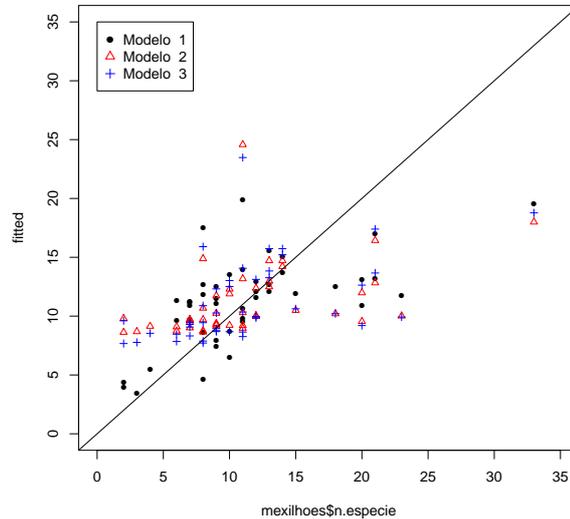
Null Deviance: 127.5

Residual Deviance: 90.81 AIC: 276.8

- (a) Qual o número de espécies esperado num rio abrangendo 1000 milhas quadradas?
- (b) Diga se o valor do desvio deste modelo pode ser directamente comparado com os desvios dos dois modelos anteriores. Comente a sua resposta.

¹J.J. Sepkoski, Jr., M.A. Rex (1974). "Distribution of Freshwater Mussels: Coastal Rivers as Biogeographic Islands," *Systematic Zoology*, Vol. 23, 2, pp. 165-188.

4. O gráfico seguinte indica, no eixo horizontal o valor observado do número de espécies em cada rio, e no eixo vertical os valores ajustados por cada modelo acima referido (identificado pelo número da alínea em que é referido). A recta corresponde à bissetriz $y = x$. Comente o gráfico e discuta as suas implicações para a modelação desejada.



II [4 valores]

Considere o contexto duma Regressão Logística, ajustada com base em n observações independentes da componente aleatória Y , que indica se a cada observação corresponde um êxito (1) ou um fracasso (0).

1. Descreva o modelo, admitindo a existência de p preditores numéricos.
2. Considere agora o Modelo Nulo (sem preditores) correspondente.
 - (a) Mostre que o estimador de Máxima Verosimilhança do único parâmetro do modelo (β_0) é dado pelo logaritmo natural da razão entre o número médio de êxitos e o número médio de fracassos, nas n observações.
 - (b) Indique a probabilidade estimada de êxito, para cada observação, com base neste modelo. Comente.