
Modelos Matemáticos e Aplicações – 2023-24

Exercícios de Modelos Lineares Generalizados

AVISO: O ficheiro `dadosMLG.RData` contém os objectos `tabaco` (Exercício 1), `ratos` (Exercício 2), `Elisa1` (Exercício 3), `Elisa2` (Exercício 5), `flea.beetles` (Exercício 8) e `sangue` (Exercício 10). O ficheiro deve ser carregado para uma sessão do R com o comando `load`.

1. No livro de W.N. Venables e B.D. Ripley, *Modern Applied Statistics with S-Plus* (1994, Springer-Verlag), refere-se uma experiência que estuda a resistência da larva do tabaco *heliothis virescens* a doses de uma substância tóxica. Lotes de 20 traças de cada sexo foram expostas, durante 3 dias, a doses da referida substância, e registou-se o número de indivíduos de cada lote que morria, ou ficava inactivo, no fim desse período de exposição. Os resultados (isto é, o número de mortes) são sintetizados na seguinte tabela, sendo as doses expressas em μg .

Sexo	Dose					
	1	2	4	8	16	32
Machos	1	4	9	13	18	20
Fêmeas	0	2	6	10	12	16

- (a) Crie uma *data frame* contendo os dados e adequada para ajustar modelos com componente aleatória Binomial/n.
 - (b) Construa uma nuvem de pontos com, no eixo horizontal, a variável Dose, e no eixo vertical, a proporção de Mortes em cada lote de 20 indivíduos. Repita, mas agora utilizando cores diferentes para representar os lotes associados a indivíduos de cada Sexo. Comente.
 - (c) Repita a alínea anterior, mas agora associando o eixo horizontal à variável $\log_2(\text{Dose})$. Esta transformação pode justificar-se, tendo em conta que as doses utilizadas duplicam em cada nova situação experimental. Comente.
 - (d) Ajuste uma Regressão Logística aos dados, ignorando as diferenças de sexo, e utilizando como variável preditora $\log_2(\text{Dose})$. Comente os resultados obtidos. Trace, por cima da nuvem de pontos obtida na alínea anterior, a curva estimada para a probabilidade de morte, $p(x)$, onde x indica valores de $\log_2(\text{Dose})$. Discuta o significado do valor do parâmetro estimado b_1 .
 - (e) Repita a alínea anterior, mas utilizando agora um modelo *Probit*. Qual a dosagem a que corresponde uma probabilidade de morte de 50%?
 - (f) Ajuste agora um modelo linear generalizado com componente aleatória adequada e utilizando uma função de ligação log-log do complementar. Comente os resultados.
2. A fim de estudar os efeitos cancerígenos de um produto tóxico em ratos, foram administradas três diferentes doses da substância tóxica (0, 0.45 e 0.75 partes por 10 000) a algumas centenas de ratos, durante um de dois períodos de exposição (16 ou 24 meses). No final do período de exposição verificava-se a existência de tumores nos ratos. Os resultados da experiência foram os seguintes:

Exposição		Dose		
		0	0.45	0.75
16 meses	Ratos com tumores	1	3	7
	Ratos sem tumores	204	301	186
24 meses	Ratos com tumores	20	98	118
	Ratos sem tumores	742	790	469

Os dados encontram-se disponíveis na *data frame* `ratos`. Ajustou-se um Modelo Linear Generalizado adequado para uma componente aleatória dicotómica, com função de ligação *probit*, considerando os preditores `Dose` e tempo de `Exposicao` como variáveis numéricas. Obtiveram-se os seguintes resultados:

```
> summary(ratos.probit.var)
Call:
glm(formula = cbind(com, sem) ~ Dose + Exposicao, family = binomial(probit),
     data = ratos)
[...]
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.8474      0.3948 -12.279 < 2e-16 ***
Dose           1.4344      0.1397  10.269 < 2e-16 ***
Exposicao      0.1229      0.0163   7.538 4.78e-14 ***
---
Null deviance: 198.5347 on 5 degrees of freedom
Residual deviance:  1.3381 on 3 degrees of freedom
AIC: 33.594
Number of Fisher Scoring iterations: 4
```

- (a) Descreva em maior pormenor o tipo de modelo ajustado, indicando a relação considerada entre o surgimento de tumores e as variáveis predictoras.
- (b) Comente a qualidade do ajustamento do modelo aos dados.
- (c) Considera possível simplificar ulteriormente o modelo sem prejuízo significativo na qualidade do ajustamento? Justifique formalmente.
- (d) Com base no modelo ajustado, responda às seguintes questões:
 - i. Para uma dose de 0.75 partes por 10 000 da substância tóxica, qual a proporção esperada de ratos com tumores ao fim de 36 meses de exposição?
 - ii. Qual a dose associada a 50% de ratos com tumor ao fim de 24 meses de exposição?

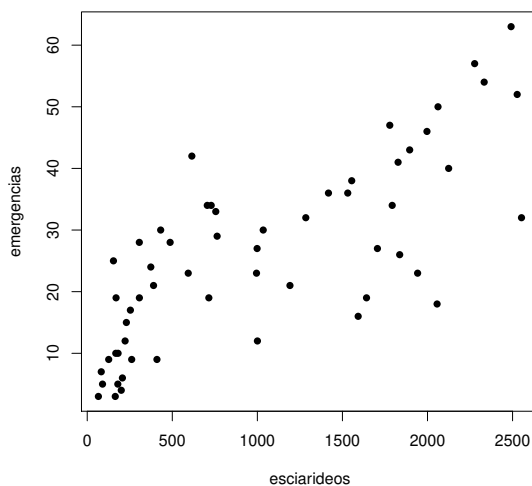
Entretanto, é levantada a objecção de que o baixíssimo número de diferentes valores dos preditores `Dose` e `Exposicao` desaconselha a sua utilização como variáveis numéricas. Decidiu-se assim ajustar um novo modelo, com estes dois preditores considerados como factores. Não se previram efeitos de interacção entre os factores. O ajustamento produziu os seguintes resultados:

```
> summary(ratos.probit.fac)
Call:
glm(formula = cbind(com, sem) ~ as.factor(Dose) + as.factor(Exposicao),
     family = binomial(probit), data = ratos)
[...]
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.9038      0.1561 -18.602 < 2e-16 ***
as.factor(Dose)0.45  0.6880      0.1069   6.435 1.24e-10 ***
as.factor(Dose)0.75  1.0859      0.1081  10.042 < 2e-16 ***
as.factor(Exposicao)24  0.9826      0.1302   7.545 4.52e-14 ***
[...]
```

Null deviance: 198.5347 on 5 degrees of freedom
Residual deviance: 1.0902 on 2 degrees of freedom
AIC: 35.347

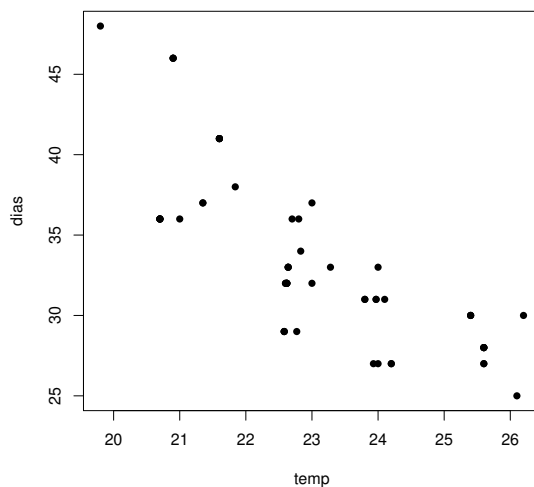
Number of Fisher Scoring iterations: 4

- (e) Descreva em pormenor o modelo agora ajustado. Comente as semelhanças e diferenças com o modelo considerado inicialmente.
 - (f) Qual a probabilidade estimada pelo modelo de um rato ter tumor ao fim de 16 meses, caso não tenha sido exposto ao tóxico? Como é que essa probabilidade estimada se compara com a frequência relativa de tumores nessa situação experimental? Como é que esta probabilidade estimada se compara com a que é obtida a partir do modelo inicial? Comente.
 - (g) É possível estimar a probabilidade de os ratos terem tumores se forem sujeitos a um período de exposição de 36 meses, a partir deste modelo?
 - (h) Com base nos indicadores de qualidade dos ajustamentos disponíveis no enunciado, da discussão efectuada até aqui, e tendo em conta as reservas expressas ao modelo inicialmente considerado, qual dos modelos considera preferível?
 - (i) Ajuste agora um terceiro modelo, considerando *Dose* e *Exposicao* como factores, mas prevendo também efeitos de interacção no modelo. Como explica o facto de o desvio do modelo, e todos os resíduos do desvio, serem nulos? Que implicações decorrem desse facto?
3. A fêmea adulta de um predador coloca os ovos num substrato de terra e aveia com fungo, infestada com mosquitos que servem de alimentação às larvas. Pretende-se perceber relação entre número de larvas de mosquitos presentes no substrato – variável *esciarideos* – e o número de adultos que emergem na geração seguinte (depois de se alimentarem enquanto larva e puparem) – variável *emergencias*. O número de mosquitos foi calculado extrapolando o número de larvas observado numa amostra para o volume total de substrato. Os dados obtidos encontram-se na *data.frame* de nome *Elisa1* e a nuvem de pontos obtida encontra-se no gráfico.



- (a) Reproduza a nuvem de pontos dada acima.

-
- (b) Considera adequado um modelo para a variável resposta **emergencias** associado a uma distribuição de Poisson?
- (c) Considera adequado usar a função de ligação canónica para as distribuições de Poisson?
- (d) Ajuste um modelo log-linear e discuta os resultados obtidos. Em particular, a estimativa do parâmetro β_1 é $b_1 = 0.0005248347$. Como se pode interpretar este valor, à luz do problema sob estudo?
- (e) Trace, sobre a nuvem de pontos, a curva ajustada pelo modelo. Comente.
- (f) Calcule intervalos a 95% de confiança para os parâmetros do modelo (β_0 e β_1), usando a teoria assintótica associada aos estimadores de máxima verosimilhança. Comente. Em particular, diga se, com base nestes intervalos, se pode afirmar que a um aumento do número de mosquitos presentes no substrato corresponde um aumento do número médio de adultos na geração seguinte.
4. No módulo **MASS** existe uma tabela de contingências do tipo local \times espécie, contida num objecto de nome **waders**. O conjunto de dados refere-se a frequências de observações de dezanove espécies de aves limícolas (*waders*), em quinze diferentes locais da costa da África Austral (Namíbia e África do Sul).
- (a) Efectue um tradicional teste χ^2 à independência dos factores “locais” e “espécies”, utilizando a estatística de Pearson. (AVISO: O comando para efectuar esse teste no R é o comando **chisq.test**.)
- (b) Crie uma **data frame** adequada para ajustar um MLG aos dados, isto é, uma data frame com três colunas: as contagens, os locais e as espécies respectivas. Utilize o seguinte comando do R:
- ```
> limicolas <- data.frame(obs=as.vector(as.matrix(waders)), local=rep(LETTERS[1:15],19),
 especie=rep(paste("S",1:19,sep=""),each=15))
```
- (c) Considere um modelo log-linear para os dados, com dois factores explicativos (aditivos): **local** e **especie**. Discuta a natureza da equação do modelo. Indique o valor esperado, ao abrigo do modelo, do número de observações da espécie S14, no local C.
- (d) Ajuste o modelo indicado na alínea anterior e discuta a sua qualidade, com base no desvio do modelo. Compare o número de observações da espécie S14, no local C, com o seu valor esperado ajustado. Comente.
- (e) Calcule a soma dos quadrados dos desvios de Pearson do modelo. Compare o valor obtido com o valor da estatística de Pearson do teste  $\chi^2$  da primeira alínea. Comente.
- (f) Interprete o significado da diferença de dois parâmetros do mesmo tipo, por exemplo  $\alpha_4 - \alpha_3$ , onde  $\alpha_i$  indica o efeito do nível  $i$  do factor **local**.
- (g) Construa um intervalo de confiança (assintótico) para  $\alpha_4 - \alpha_3$  e interprete o seu significado.
- (h) Com base nas alíneas anteriores, comente a utilidade do seu modelo.
5. Num estudo sobre controlo de pragas pretende-se modelar, para uma dada espécie de insectos, a relação entre número de dias separando a postura de ovos e a emergência de novos adultos (variável resposta, designada **dias**) e a temperatura do meio ambiente (variável preditora, designada **temp**). Num estudo envolvendo  $n = 57$  repetições, foram recolhidos os dados constantes da **data.frame** **Elisa2**, a que corresponde a seguinte nuvem de pontos.



- (a) Ajuste um modelo log-linear aos dados. Em particular,
- Indique todas as opções que fez;
  - Discuta a adequação dum modelo log-linear à relação observada na nuvem de pontos;
  - Trace a curva ajustada sobre a nuvem de pontos.
- (b) Um analista comenta que, sendo a variável resposta *dias* uma contagem de tempo, trata-se na realidade duma variável contínua que foi discretizada. Nesse sentido, sugere que se poderia optar por ajustar um MLG análogo ao da alínea anterior, com apenas uma modificação: considerar que a distribuição da variável resposta é Normal. Discuta o novo MLG ajustado, e em particular:
- Diga porque é que o modelo que ajustou *não* é um modelo *linear*.
  - Explicita a equação da curva ajustada e trace-a sobre a nuvem de pontos. Como se explica que esta curva seja diferente da anterior? E como se explica que seja próxima da anterior?
  - Considere o valor do desvio residual associado a este modelo, e comente o facto de ser substancialmente diferente do obtido ao ajustar o modelo anterior. Comente em particular a seguinte afirmação: “o modelo ajustado na alínea anterior é melhor, uma vez que tem um desvio residual menor”.
- (c) Ajuste agora um *modelo linear* que melhor corresponda ao modelo ajustado na alínea anterior. Em particular,
- Explicita a equação e pressupostos do modelo ajustado, e compare-os com os modelos anteriores.
  - Indique a equação da curva ajustada, e trace-a sobre a nuvem de pontos anterior.
  - Estude os resíduos deste modelo *linear* e comente a validade dos pressupostos do modelo.
  - Sendo um modelo linear um caso particular dum MLG, tem de fazer sentido falar no desvio residual do modelo ajustado nesta alínea. Calcule-o com o auxílio do R e diga, justificando, se o seu valor pode ser comparado com o valor obtido no modelo da alínea anterior, em que também se admitiu uma distribuição Normal da variável resposta.

6. Existem parametrizações alternativas da função densidade da distribuição Gama. A parametrização dada nos acetatos é

$$f(y | \mu, \nu) = \frac{\nu^\nu}{\mu^\nu \Gamma(\nu)} y^{\nu-1} e^{-\frac{\nu y}{\mu}},$$

cujo parâmetro  $\mu$  é o valor esperado da variável e cujo segundo parâmetro  $\nu$  ajuda a construir a variância,  $V[Y] = \frac{\mu^2}{\nu}$ .

- (a) Outra parametrização da densidade Gama é:

$$f(y | \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-\frac{y}{\beta}}.$$

Mostre que se trata duma diferente parametrização da mesma função, com  $\mu = \alpha\beta$  e  $\nu = \alpha$ .

- (b) No livro *Probabilidades e Estatística*, de prof. Bento Murteira (McGraw-Hill de Portugal, 1979), dá-se uma terceira parametrização da Gama, da forma:

$$f(y | n, \gamma) = \frac{\gamma^n}{\Gamma(n)} y^{n-1} e^{-\gamma y}.$$

Identifique as relações entre os parâmetros desta expressão e os das anteriores parametrizações. Relacione o valor esperado e variância nesta parametrização com os da parametrização usada nas aulas.

7. Defina os seguintes conceitos, no contexto de Modelos Lineares Generalizados:

- (a) função de ligação  
(b) resíduo do desvio

8. Dezanove escaravelhos da espécie *Haltica oleracea* e vinte escaravelhos da espécie *Haltica carduorum* foram sujeitos a medições morfométricas em quatro variáveis: a distância do sulco transversal à borda posterior do pró-torax (variável *TG*), o comprimento do élitro (variável *Elytra*), o comprimento do segundo segmento das antenas (variável *Second.Antenna*) e o comprimento do terceiro segmento das antenas (variável *Third.Antenna*). As unidades de todas as variáveis *excepto o comprimento do élitro* são micrómetros (milionésima parte do metro,  $\mu m$ ). O comprimento do élitro é dado em centésimas de milímetro.

Alguns dos dados obtidos são indicados na tabela seguinte.

|           | Species   | TG       | Elytra   | Second.Antenna | Third.Antenna |
|-----------|-----------|----------|----------|----------------|---------------|
| 1         | oleracea  | 189      | 245      | 137            | 163           |
| 2         | oleracea  | 192      | 260      | 132            | 217           |
| 3         | oleracea  | 217      | 276      | 141            | 192           |
| 4         | oleracea  | 221      | 299      | 142            | 213           |
| (...)     |           |          |          |                |               |
| 18        | oleracea  | 181      | 255      | 146            | 183           |
| 19        | oleracea  | 192      | 287      | 141            | 198           |
| 20        | carduorum | 181      | 305      | 184            | 209           |
| 21        | carduorum | 158      | 237      | 133            | 188           |
| (...)     |           |          |          |                |               |
| 36        | carduorum | 192      | 276      | 154            | 209           |
| 37        | carduorum | 181      | 278      | 149            | 235           |
| 38        | carduorum | 175      | 271      | 140            | 192           |
| 39        | carduorum | 197      | 303      | 170            | 205           |
| -----     |           |          |          |                |               |
| variância |           | 196.888  | 502.7085 | 216.0445       | 341.8313      |
| média     |           | 186.8205 | 279.2308 | 147.5385       | 197.8974      |

*Haltica oleracea*



Pretende-se determinar um modelo que permita identificar a que espécie pertence um dado escaravelho, isto é, pretende-se efectuar uma análise discriminante da espécie. Tendo em atenção a dificuldade em obter medições precisas, dada a pequena dimensão dos animais, considera-se importante que o modelo seja parcimonioso, com o menor número possível de preditores morfométricas.

- (a) Efectuou-se uma Regressão Logística, tomando como ponto de partida as quatro variáveis morfométricas referidas. Obtiveram-se os seguintes resultados.

```
Call: glm(formula = (Species == "carduorum") ~ TG + Elytra + Second.Antenna
+ Third.Antenna, family = binomial, maxit = 50, data=flea.beetles)
```

Coefficients:

|                | Estimate   | Std. Error | z value   | Pr(> z ) |
|----------------|------------|------------|-----------|----------|
| (Intercept)    | -6.237e+02 | 1.869e+06  | -3.34e-04 | 1        |
| TG             | -1.162e+01 | 2.077e+04  | -0.001    | 1        |
| Elytra         | 5.559e+00  | 9.735e+03  | 0.001     | 1        |
| Second.Antenna | 7.634e+00  | 1.757e+04  | 4.34e-04  | 1        |
| Third.Antenna  | 8.133e-01  | 1.411e+04  | 5.77e-05  | 1        |

```
Null deviance: 5.4040e+01 on 38 degrees of freedom
Residual deviance: 4.7616e-10 on 34 degrees of freedom
AIC: 10 Number of Fisher Scoring iterations: 28
```

- i. Descreva completamente o modelo ajustado, enquanto Modelo Linear Generalizado, indicando as suas três componentes fundamentais.
  - ii. Comente a qualidade do modelo proposto para efeitos de identificação da espécie dos escaravinhos. Como se pode explicar o desvio quase nulo do modelo ajustado? Há um problema de sobreparametrização?
  - iii. Interprete o significado do valor 7.634 para a estimativa associada à variável *Second.Antenna*.
  - iv. Com base na informação disponível, diga se é possível simplificar o modelo sem perda significativa na qualidade da discriminação entre espécies efectuada. Em caso afirmativo, qual a primeira variável preditora a ser excluída do modelo por um método de tipo exclusão sequencial?
- (b) Foi utilizado um algoritmo de exclusão sequencial, usando a função `step` do R. Comente os vários passos do algoritmo e indique qual o modelo final.

```
> step(flea.glm.logit)
Start: AIC=10
(Species == "carduorum") ~ TG + Elytra + Second.Antenna + Third.Antenna
```

|                  | Df | Deviance | AIC    |
|------------------|----|----------|--------|
| - Third.Antenna  | 1  | 0.000    | 8.000  |
| - Second.Antenna | 1  | 0.000    | 8.000  |
| <none>           |    | 0.000    | 10.000 |
| - Elytra         | 1  | 10.132   | 18.132 |
| - TG             | 1  | 24.686   | 32.686 |

```
Step: AIC=8
(Species == "carduorum") ~ TG + Elytra + Second.Antenna
```

|                  | Df | Deviance | AIC    |
|------------------|----|----------|--------|
| <none>           |    | 0.0000   | 8.000  |
| - Second.Antenna | 1  | 9.8414   | 15.841 |
| - Elytra         | 1  | 16.6409  | 22.641 |
| - TG             | 1  | 29.7719  | 35.772 |

---

```
Call: glm(formula = (Species == "carduorum") ~ TG + Elytra +
 Second.Antenna, family = binomial, data = flea.beetles, maxit = 50)
```

```
Coefficients:
 (Intercept) TG Elytra Second.Antenna
 -968.93 -19.46 9.37 13.91
```

```
Degrees of Freedom: 38 Total (i.e. Null); 35 Residual
Null Deviance: 54.04
Residual Deviance: 3.846e-10 AIC: 8
```

- (c) Independentemente da sua resposta no ponto anterior, decidiu-se ajustar um modelo com apenas duas variáveis preditoras. O melhor modelo desse tipo resultou ser o que deixava de fora as medições relativas às antenas. Alguns dos resultados respectivos são indicados de seguida.

```
Call: glm(formula = (Species == "carduorum") ~ TG + Elytra,
 family = binomial, maxit = 50, data=flea.beetles)
```

```
Coefficients:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) 10.1559 12.8285 0.792 0.4286
TG -0.4271 0.1792 -2.384 0.0171 *
Elytra 0.2505 0.1038 2.413 0.0158 *
```

```

Null deviance: 54.0398 on 38 degrees of freedom
Residual deviance: 9.8414 on 36 degrees of freedom
AIC: 15.841 Number of Fisher Scoring iterations: 8
```

- i. Teste formalmente se este modelo e o modelo inicial diferem significativamente.
  - ii. Diga quais as probabilidades para cada espécie previstas pelo modelo agora ajustado, no caso de um escaravelho com  $TG = 200$  e  $Elytra = 250$ . Qual a espécie a que associaria um tal escaravelho?
- (d) Decidiu-se agora experimentar uma diferente função de ligação, e em particular a função de ligação log-log do complementar, utilizando apenas os dois preditores referidos no ponto 8c. Os resultados obtidos foram agora os seguintes:

```
Call: glm(formula = (Species == "carduorum") ~ TG + Elytra,
 family = binomial(link = "cloglog"), maxit = 50)
```

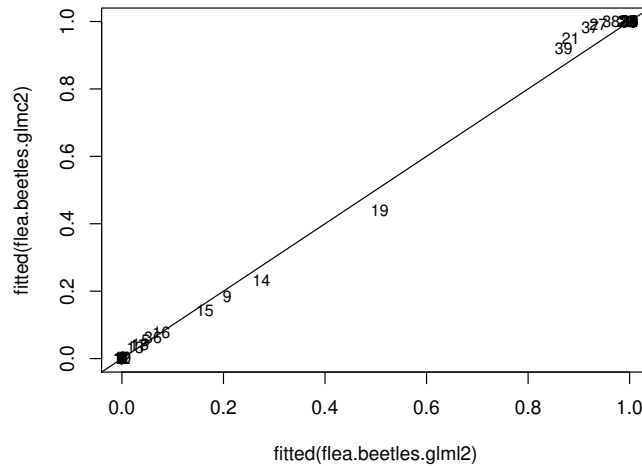
```
Coefficients:
 Estimate Std. Error z value Pr(>|z|)
(Intercept) 7.78272 7.75729 1.003 0.3157
TG -0.33889 0.13206 -2.566 0.0103 *
Elytra 0.19769 0.07766 2.546 0.0109 *
```

```

Null deviance: 54.0398 on 38 degrees of freedom
Residual deviance: 8.7522 on 36 degrees of freedom
AIC: 14.752 Number of Fisher Scoring iterations: 12
```

- i. O seguinte gráfico indica as probabilidades ajustadas por cada modelo, com os valores relativos ao modelo com função ligação log-log do complementar no eixo vertical e os relativos à ligação canónica no eixo horizontal. Comente os resultados. Discuta, em particular, o indivíduo 19.





- ii. Diga qual dos modelos com dois preditores prefere: este, ou o indicado no ponto 8c. Justifique.
  - iii. Diga qual a probabilidade prevista para um indivíduo com valores observados  $TG = 200$  e  $Elytra = 250$ . Compare com o resultado análogo obtido com o modelo do ponto 8c e comente.
9. Considere de novo os dados do Exercício 1. Ajuste um modelo *probit* para modelar a probabilidade de morte, mas desta vez considerando na componente sistemática não apenas a variável numérica  $\log_2(dose)$ , mas igualmente o factor **sexo**.
- (a) Obtenha um único modelo que possa ser interpretado como tendo duas diferentes componentes sistemáticas,  $\beta_0 + \beta_1 \log_2(Dose)$ , uma para machos e outra para fêmeas, cada qual com os seus parâmetros específicos.
  - (b) Ajuste o modelo da alínea anterior aos dados e comente. Pode considerar-se que este modelo é melhor que o modelo ajustado no Exercício 1?
  - (c) Considere agora um terceiro modelo, em que na componente sistemática se admite que o coeficiente da log-dose seja igual para os dois sexos, mas que a constante aditiva pode diferir consoante o sexo. Ajuste o modelo, e compare os resultados obtidos com os dois modelos anteriormente considerados. Comente.
  - (d) Por qual dos três modelos considerados optaria? Justifique.
10. No livro de P. McCullagh e J.A. Nelder, *Generalized Linear Models* (2a. edição, 1989, Chapman & Hall), nas páginas 300-302, é discutido um conjunto de dados onde se mediram tempos de coagulação (em segundos) de sangue, para plasma normal diluído com nove diferentes concentrações de plasma sem a proteína protrombina (do tipo protease serina, produzida no fígado e que, quando activada - gerando a trombina - está associada à coagulação do sangue). Dois diferentes lotes de um agente activador da coagulação foram utilizados. Os dados observados foram os seguintes, e são dados na *data frame* **sangue**.

---

| Concentração | Tempo de coagulação |        |
|--------------|---------------------|--------|
|              | Lote 1              | Lote 2 |
| 5            | 118                 | 69     |
| 10           | 58                  | 35     |
| 15           | 42                  | 26     |
| 20           | 35                  | 21     |
| 30           | 27                  | 18     |
| 40           | 25                  | 16     |
| 60           | 21                  | 13     |
| 80           | 19                  | 12     |
| 100          | 18                  | 12     |

Deseja-se estudar o efeito das diferentes concentrações de plasma sem protrombina sobre os tempos de coagulação. Comece por ignorar os efeitos associadas aos lotes.

- Represente graficamente tempo de coagulação (`tempo`, eixo vertical) contra concentrações de plasma (`conc.plasma`, eixo horizontal), utilizando símbolos e/ou cores diferentes para representar as observações de cada lote. Comente.
- É sugerido que a relação entre as variáveis `tempo` e concentração de plasma sem protrombina (variável `conc.plasma`) é de tipo hiperbólico, ou seja da forma  $tempo = \frac{1}{\beta_0 + \beta_1 \cdot conc}$ . Produza uma representação gráfica adequada para validar visualmente esta proposta. Comente.
- Após um estudo gráfico adequado, conclui-se que a relação mais adequada parece ser do tipo hiperbólico mas sobre os logaritmos das concentrações de plasma, ou seja, da forma  $tempo = \frac{1}{\beta_0 + \beta_1 \ln(conc)}$ . Confirme, produzindo a representação gráfica adequada.
- Para ajustar a relação indicada na alínea anterior, a função de ligação indicada é a função recíproco,  $g(\mu) = \frac{1}{\mu}$ , utilizando como preditor a variável das log-concentrações. Mas permanece de pé a escolha de qual a distribuição a associar à variável-resposta `tempo`. Ajuste dois diferentes MLGs, admitindo:
  - que `tempo` tem distribuição Normal (Nota: No *R*, este ajustamento corresponde a dar o argumento `family=gaussian(link="inverse")` no comando `glm`);
  - que `tempo` tem distribuição Gama (Nota: No *R*, este ajustamento corresponde a dar o argumento `family=Gamma`, não sendo necessário especificar a função de ligação, uma vez que a função recíproco é a função de ligação canónica para a distribuição Gama).

Trace as curvas correspondentes a cada ajustamento por cima da nuvem de pontos de `tempo` (eixo vertical) contra log-concentrações de plasma (eixo horizontal). Comente.

- Compare os ajustamentos obtidos na alínea anterior. Comente, e diga qual a escolha mais adequada para distribuição de `tempo`, tendo em conta a natureza e valores dessa variável resposta, e o conjunto da informação disponível.

Nas alíneas seguintes considere o factor `lote`, com os seus dois níveis.

- Ajuste modelos com componente aleatória Normal e Gama, e função de ligação recíproco, mas prevendo agora que se cruza o preditor quantitativo log-concentração com o factor `lote`.
- Interprete o significado dos parâmetros obtidos, traçando as curvas ajustada para cada lote num gráfico de tempo *vs.* log-concentração.
- Comente a qualidade dos ajustamentos agora obtidos.