

MODELOS DE REGRESSÃO com apoio computacional

Gilberto A. Paula

Instituto de Matemática e Estatística
Universidade de São Paulo
e-mail:giapaula@ime.usp.br

Prefácio

A área de modelagem estatística de regressão recebeu um grande impulso desde a criação dos modelos lineares generalizados (MLGs) no início da década de 70. O crescente interesse pela área motivou a realização de vários encontros informais no início dos anos 80, a maioria deles na Inglaterra, até que em 1986 foi realizado na cidade de Innsbruck na Áustria o “1st International Workshop on Statistical Modelling” (1st IWSM). Esse encontro tem sido realizado anualmente sendo que o último (27th IWSM) aconteceu em julho de 2012 em Praga, República Checa. O 28th IWSM será realizado em julho de 2013 em Palermo, Itália. No Brasil a área começou efetivamente a se desenvolver a partir de meados da década de 80 e em particular após a 1^a Escola de Modelos de Regressão (1EMR) realizada na Universidade de São Paulo em 1989. As demais escolas ocorreram desde então a cada dois anos sendo que as duas últimas (11EMR e 12EMR) foram realizadas em 2009 e 2011 nas cidades de Recife e Fortaleza, respectivamente.

Este texto começou a ser desenvolvido a partir de 1994 quando a disciplina **Modelos Lineares Generalizados** passou a ser ministrada regularmente no programa de pós-graduação em Estatística do IME-USP. O texto é direcionado para alunos que tenham cursado um semestre de inferência estatística e que tenham conhecimentos de regressão linear. Portanto, trata-se de um segundo curso de modelagem estatística de regressão com um enfoque inferencial básico e várias aplicações. O texto tem sido também utilizado na

disciplina **Tópicos de Regressão** ministrada aos alunos do último ano do Bacharelado em Estatística do IME-USP.

No Capítulo 1 introduzimos a classe dos modelos lineares generalizados juntamente com alguns conceitos básicos. Em seguida discutimos a estimação dos parâmetros, propriedades assintóticas dos estimadores de máxima verossimilhança e a aplicação de alguns testes estatísticos mais conhecidos. Uma revisão abrangente de métodos de diagnóstico é apresentada na sequência, em que definimos pontos de alavanca e discutimos análise de resíduos, métodos de deleção de pontos e influência local, dentre outros procedimentos. Discutimos também a seleção de modelos. Sempre que possível as metodologias são apresentadas em separado para os modelos normais lineares e posteriormente estendidas para toda a classe dos MLGs. O capítulo é concluído com a análise de 6 conjuntos de dados através de MLGs apropriados.

O Capítulo 2 é dedicado aos modelos com resposta gama e resposta normal inversa para a análise de dados assimétricos positivos. Inicialmente abordamos os modelos com resposta gama e apresentamos alguns resultados inferenciais e técnicas de diagnóstico. Três conjuntos de dados são analisados. Em seguida, alguns modelos usualmente aplicados em Econometria são discutidos e um exemplo é apresentado. Em seguida são discutidos modelos com resposta normal inversa, alguns resultados teóricos são apresentados e 2 conjuntos de dados são analisados. No final do capítulo discutimos MLGs duplos, em que a média e a dispersão são ajustados simultaneamente. Apresentamos o processo de estimação conjunto, alguns procedimentos de diagnóstico e um exemplo ilustrativo.

No Capítulo 3 discutimos modelos para a análise de dados binários, com ênfase para os modelos logísticos lineares. Inicialmente uma revisão de procedimentos tradicionais para a análise de tabelas de contingência 2×2 é apresentada. Duas aplicações são descritas nesta primeira parte do capítulo. Em seguida abordamos o modelo logístico linear. Alguns procedimentos são

revisados, tais como seleção de modelos, análise de dados retrospectivos, qualidade do ajuste e técnicas de diagnóstico. Quatro conjuntos de dados são analisados. Discutimos no final do capítulo modelos de dose-resposta, sobre-dispersão e modelos logísticos aplicados na análise de dados emparelhados e mais quatro aplicações são apresentadas.

No Capítulo 4 abordamos alguns modelos para a análise de dados de contagem, com destaque para modelos com resposta de Poisson e modelos com resposta binomial negativa. Inicialmente apresentamos uma revisão de metodologias tradicionais para a análise de tabelas de contingência do tipo 2×2 com dados de contagem. Uma aplicação é apresentada. Em seguida discutimos modelos de Poisson para a análise de dados de seguimento e modelos log-lineares de Poisson. Dois exemplos são apresentados. Na sequência são derivados modelos com resposta binomial negativa para a análise de dados de contagem com sobre-dispersão. Um processo iterativo para a estimação dos parâmetros, resultados assintóticos e metodologias de diagnóstico são apresentados, bem como 3 aplicações. Modelos log-lineares com resposta de Poisson são comparados com modelos log-lineares com resposta multinomial, sendo 2 conjuntos de dados analisados. Finalmente, fazemos uma breve resenha dos modelos com excesso de zeros e discutimos, em particular, os modelos ajustados em zero (modelos de fronteira) e os modelos inflacionados de zeros.

O Capítulo 5 é dedicado aos modelos de quase-verossimilhança e às equações de estimação generalizadas. Iniciamos o capítulo com a introdução do conceito de quase-verossimilhança. Em seguida são apresentados os modelos de quase-verossimilhança para respostas independentes juntamente com o processo de estimação, alguns resultados assintóticos e técnicas de diagnóstico. Três aplicações são apresentadas. Na sequência derivamos as equações de estimação generalizadas para a análise de dados correlacionados não gaussianos. Apresentamos o processo de estimação, alguns resultados assintóticos e metodologias de diagnóstico. Esse subtópico é ilustrado com 3 aplicações.

No apêndice são descritos os conjuntos de dados usados nas aplicações e nos exercícios propostos.

No final de cada capítulo são propostos exercícios teóricos e aplicados, num total de 120 exercícios, e ao longo do texto são apresentados programas especiais e subrotinas em R, particularmente na análise dos exemplos. Procuramos diversificar as aplicações com conjuntos de dados das diversas áreas do conhecimento, tais como Agricultura, Biologia, Ciências Atuariais, Ciências Sociais, Economia, Engenharia, Geografia, Medicina, Nutrição, Pesca e Odontologia. Alguns conjuntos de dados são oriundos de trabalhos desenvolvidos no Centro de Estatística Aplicada (CEA) do IME-USP. A página na Web onde estão disponíveis informações sobre este texto está no endereço: <http://www.ime.usp.br/~giapaula/textoregressao.htm>.

Finalizando, gostaria de agradecer aos alunos que cursaram as disciplinas **Modelos Lineares Generalizados** e **Tópicos de Regressão** e contribuíram com suas sugestões para o aprimoramento dos primeiros manuscritos.

São Paulo, fevereiro de 2013

Gilberto A. Paula

e-mail:giapaula@ime.usp.br

Sumário

Prefácio	i
Sumário	v
1 Modelos Lineares Generalizados	1
1.1 Introdução	1
1.2 Definição	4
1.2.1 Casos particulares	5
1.3 Ligações canônicas	7
1.3.1 Outras ligações	8
1.4 Função desvio	13
1.4.1 Resultados assintóticos	15
1.4.2 Análise do desvio	16
1.5 Função escore e informação de Fisher	20
1.5.1 Escore e Fisher para β	20
1.5.2 Escore e Fisher para ϕ	22
1.5.3 Ortogonalidade	22
1.5.4 Casos particulares	23
1.6 Estimação dos parâmetros	25
1.6.1 Estimação de β	25
1.6.2 Estimação de ϕ	26
1.6.3 Distribuição assintótica	27

1.7	Teste de hipóteses	28
1.7.1	Hipóteses simples	28
1.7.2	Modelos encaixados	31
1.7.3	Modelo de análise de variância	37
1.7.4	Regressão linear simples	38
1.7.5	Hipóteses restritas	39
1.8	Bandas de confiança	40
1.8.1	Modelo normal linear	40
1.8.2	Extensão para os MLGs	40
1.9	Técnicas de diagnóstico: Modelo normal linear	41
1.9.1	Introdução	41
1.9.2	Pontos de alavanca	45
1.9.3	Resíduos	48
1.9.4	Outra interpretação para t_i^*	51
1.9.5	Influência	51
1.9.6	Ilustração	54
1.9.7	Influência local	56
1.9.8	Gráfico da variável adicionada	63
1.9.9	Técnicas gráficas	64
1.10	Técnicas de diagnóstico: Extensão para os MLGs	66
1.10.1	Pontos de alavanca	66
1.10.2	Resíduos	68
1.10.3	Influência	72
1.10.4	Influência local	73
1.10.5	Gráfico da variável adicionada	75
1.10.6	Técnicas gráficas	77
1.11	Seleção de modelos	77
1.11.1	Modelo normal linear	77
1.11.2	Extensão para os MLGs	80

1.12	Aplicações	81
1.12.1	Estudo entre renda e escolaridade	81
1.12.2	Processo infeccioso pulmonar	86
1.12.3	Sobrevivência de bactérias	88
1.12.4	Estudo seriado com ratos	91
1.12.5	Consumo de combustível	95
1.12.6	Salário de executivos	98
1.13	Exercícios	104
2	Modelos para Dados Positivos Assimétricos	115
2.1	Introdução	115
2.2	Distribuição gama	116
2.3	Modelos com resposta gama	119
2.3.1	Qualidade do ajuste	120
2.3.2	Técnicas de diagnóstico	121
2.4	Aplicações	122
2.4.1	Comparação de cinco tipos de turbina de avião	122
2.4.2	Espinhel de fundo	128
2.4.3	Aplicação em seguros	137
2.5	Elasticidade	143
2.5.1	Modelo de Cobb-Douglas	144
2.5.2	Aplicação	145
2.6	Distribuição normal inversa	147
2.7	Modelos com resposta normal inversa	150
2.7.1	Qualidade do ajuste	150
2.7.2	Técnicas de diagnóstico	150
2.8	Aplicações	151
2.8.1	Comparação de cinco tipos de <i>snack</i>	151
2.8.2	Projeção de vendas	159

2.9	Modelagem simultânea da média e da dispersão	161
2.9.1	Estimação	164
2.9.2	Métodos de diagnóstico	166
2.9.3	Aplicação	170
2.10	Exercícios	174
3	Modelos para Dados Binários	185
3.1	Introdução	185
3.2	Métodos clássicos: uma única tabela 2×2	186
3.2.1	Risco relativo	187
3.2.2	Modelo probabilístico não condicional	189
3.2.3	Modelo probabilístico condicional	190
3.2.4	Teste de hipóteses	194
3.3	Métodos clássicos: k tabelas 2×2	197
3.3.1	Estimação da razão de chances comum	198
3.3.2	Testes de homogeneidade	199
3.4	Métodos clássicos: tabelas $2 \times k$	201
3.5	Aplicações	203
3.5.1	Associação entre fungicida e desenvolvimento de tumor	203
3.5.2	Efeito de extrato vegetal	205
3.6	Regressão logística linear	206
3.6.1	Introdução	206
3.6.2	Regressão logística simples	207
3.6.3	Regressão logística múltipla	211
3.6.4	Bandas de confiança	212
3.6.5	Seleção de modelos	213
3.6.6	Amostragem retrospectiva	217
3.6.7	Qualidade do ajuste	218
3.6.8	Técnicas de diagnóstico	220

3.6.9	Aplicações	222
3.6.10	Modelos de dose-resposta	236
3.6.11	Aplicações	238
3.6.12	Estimação da dose letal	244
3.6.13	Modelos de retas paralelas	246
3.6.14	Sobredispersão	249
3.6.15	Modelo logístico condicional	261
3.7	Exercícios	269
4	Modelos para Dados de Contagem	283
4.1	Introdução	283
4.1.1	Métodos clássicos: uma única tabela 2×2	284
4.1.2	Estratificação: k tabelas 2×2	290
4.2	Modelos de Poisson	292
4.2.1	Propriedades da Poisson	292
4.2.2	Modelos log-lineares: k tabelas 2×2	293
4.2.3	Modelos gerais de Poisson	297
4.2.4	Qualidade do ajuste	299
4.2.5	Técnicas de diagnóstico	300
4.2.6	Aplicação	301
4.3	Modelos com resposta binomial negativa	305
4.3.1	Distribuição binomial negativa	305
4.3.2	Modelos de regressão com resposta binomial negativa	307
4.3.3	Qualidade do ajuste	311
4.3.4	Técnicas de diagnóstico	312
4.3.5	Seleção de modelos	313
4.3.6	Aplicações	314
4.3.7	Sobredispersão e quase-verossimilhança	323
4.4	Relação entre a multinomial e a Poisson	327

4.4.1	Modelos log-lineares hierárquicos	330
4.4.2	Aplicações	332
4.5	Modelos de contagem com excesso de zeros	338
4.5.1	Modelos de contagem ajustados em zero	338
4.5.2	Modelos de regressão ajustados em zero	340
4.5.3	Modelos de contagem inflacionados de zeros	341
4.5.4	Modelos de regressão inflacionados de zeros	342
4.6	Exercícios	343
5	Modelos de Quase-Verossimilhança	353
5.1	Introdução	353
5.2	Respostas independentes	357
5.2.1	Estimação	358
5.2.2	Estimador de momentos	358
5.2.3	Função quase-desvio	359
5.2.4	Teste de hipóteses	360
5.2.5	Resíduos	361
5.2.6	Influência	362
5.2.7	Seleção de Modelos	362
5.2.8	Aplicações	362
5.3	Classe estendida	372
5.4	Respostas correlacionadas	375
5.4.1	Estimação	378
5.4.2	Estruturas de correlação	379
5.4.3	Métodos de diagnóstico	380
5.4.4	Seleção de modelos	381
5.5	Exemplos	382
5.5.1	Ataques epilépticos	382
5.5.2	Condição Respiratória	389

5.5.3	Placas dentárias	393
5.6	Exercícios	399
Apêndice		409
Bibliografia		418

Capítulo 1

Modelos Lineares Generalizados

1.1 Introdução

Durante muitos anos os modelos normais lineares foram utilizados na tentativa de descrever a maioria dos fenômenos aleatórios. Mesmo quando o fenômeno sob estudo não apresentava uma resposta para a qual fosse razoável a suposição de normalidade, algum tipo de transformação era sugerida a fim de alcançar a normalidade procurada. Provavelmente a transformação mais conhecida foi proposta por Box e Cox (1964), a qual transforma o valor observado y (positivo) em

$$z = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log y & \text{se } \lambda = 0, \end{cases}$$

sendo λ uma constante desconhecida. O objetivo da transformação de Box e Cox, quando aplicada a um conjunto de valores observados, é produzir aproximadamente a normalidade, a constância de variância e também a linearidade $E(Z) = \eta$, em que $\eta = \beta_1 + \beta_2 x_2 + \cdots + \beta_p x_p$. No entanto, isso raramente ocorre para um único valor de λ (Box e Draper, 1987).

Com o desenvolvimento computacional ocorrido na década de 70, alguns

modelos que exigiam a utilização de processos iterativos para a estimação dos parâmetros começaram a ser mais aplicados, como por exemplo o modelo normal não linear. Todavia, a proposta mais interessante e pode-se dizer inovadora no assunto foi apresentada por Nelder e Wedderburn (1972), que propuseram os modelos lineares generalizados (MLGs). A ideia básica consiste em abrir o leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial uniparamétrica de distribuições, bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta e o preditor linear η . Assim, por exemplo, para dados de contagem, em vez de aplicar-se a transformação \sqrt{y} no sentido de buscar-se a normalidade dos dados e constância de variância, pode-se supor que a distribuição de Y é Poisson e que a relação funcional entre a média de Y e o preditor linear é dada por $\log(\mu) = \eta$. Essa relação funcional é conveniente, uma vez que garante para quaisquer valores dos parâmetros do preditor linear um valor positivo para μ . Similarmente, para proporções, pode-se pensar na distribuição binomial para a resposta e numa relação funcional do tipo $\log\{\mu/(1-\mu)\}$, em que $0 < \mu < 1$ denota a proporção esperada de sucessos.

Nelder e Wedderburn propuseram também um processo iterativo para a estimação dos parâmetros e introduziram o conceito de desvio que tem sido largamente utilizado na avaliação da qualidade do ajuste dos MLGs, bem como no desenvolvimento de resíduos e medidas de diagnóstico. Inúmeros trabalhos relacionados com modelos lineares generalizados foram publicados desde 1972 bem como a implementação dos MLGs em alguns *softwares*. Neste texto as saídas e gráficas foram desenvolvidos no software R (<http://CRAN.R-project.org>).

Os modelos de quase-verossimilhança, que estendem a ideia dos MLGs

para situações mais gerais incluindo dados correlacionados, foram propostos por Wedderburn (1974). Os modelos de dispersão (Jørgensen, 1983) ampliam o leque de opções para a distribuição da variável resposta. Liang e Zeger (1986) estendem os modelos de quase-verossimilhança propondo as equações de estimação generalizadas (EEGs) que permitem o estudo de variáveis aleatórias correlacionadas não gaussianas. Os modelos não lineares de família exponencial (Cordeiro e Paula, 1989 e Wei, 1998) admitem preditor não linear nos parâmetros. Temos ainda os modelos aditivos generalizados (Hastie e Tibshirani, 1990; Green e Silverman, 1994) que supõem preditor linear formado também por funções semiparamétricas e os modelos lineares generalizados mistos (Breslow e Clayton, 1993 e McCulloch e Searle, 2001) que admitem a inclusão de efeitos aleatórios gaussianos no preditor linear. Recentemente, Lee e Nelder (1996, 2001) estenderam o trabalho de Breslow e Clayton propondo modelos lineares generalizados hierárquicos em que o preditor linear pode ser formado por efeitos fixos e efeitos aleatórios não gaussianos. Muitos desses resultados são discutidos no livro de Lee, Nelder e Pawitan (2006). Extensões de MLGs para séries temporais, análise de dados de sobrevivência, modelos de espaço de estado e outros modelos multivariados são descritas, por exemplo, em Fahrmeir e Tutz (2001). Os modelos aditivos generalizados de localização, escala e forma propostos por Rigby e Stasinopoulos (2005) contemplam as diversas extensões dos MLGs. Referências de texto no assunto são os livros de McCullagh e Nelder (1989) e Cordeiro (1986).

Neste capítulo introduzimos os modelos lineares generalizados e apresentamos vários resultados relacionados com estimação, teste de hipóteses, métodos de diagnóstico e seleção de modelos na classe dos MLGs. Algumas aplicações são apresentadas no final do capítulo

1.2 Definição

Sejam Y_1, \dots, Y_n variáveis aleatórias independentes, cada uma com função densidade ou função de probabilidades na forma dada abaixo

$$f(y_i; \theta_i, \phi) = \exp[\phi\{y_i\theta_i - b(\theta_i)\} + c(y_i, \phi)]. \quad (1.1)$$

Podemos mostrar sob as condições usuais de regularidade

$$\begin{aligned} E \left\{ \frac{\partial \log f(Y_i; \theta_i, \phi)}{\partial \theta_i} \right\} &= 0 \quad \text{e} \\ E \left[\frac{\partial^2 \log f(Y_i; \theta_i, \phi)}{\partial \theta_i^2} \right] &= -E \left[\left\{ \frac{\partial \log f(Y_i; \theta_i, \phi)}{\partial \theta_i} \right\}^2 \right], \end{aligned}$$

$\forall i$, que $E(Y_i) = \mu_i = b'(\theta_i)$ e $\text{Var}(Y_i) = \phi^{-1}V(\mu_i)$, em que $V_i = V(\mu_i) = d\mu_i/d\theta_i$ é a função de variância e $\phi^{-1} > 0$ ($\phi > 0$) é o parâmetro de dispersão (precisão). A função de variância desempenha um papel importante na família exponencial, uma vez que a mesma caracteriza a distribuição. Isto é, dada a função de variância, tem-se uma classe de distribuições correspondentes, e vice-versa. Para ilustrar, a função de variância definida por $V(\mu) = \mu(1 - \mu)$, $0 < \mu < 1$, caracteriza a classe de distribuições binomiais com probabilidades de sucesso μ ou $1 - \mu$. Uma propriedade interessante envolvendo a distribuição de Y e a função de variância é a seguinte:

$$\sqrt{\phi}(Y - \mu) \rightarrow_d N(0, V(\mu)), \quad \text{quando } \phi \rightarrow \infty.$$

Ou seja, para ϕ grande Y segue distribuição aproximadamente normal de média μ e variância $\phi^{-1}V(\mu)$. Esse tipo de abordagem assintótica, diferente da usual em que n é grande, foi introduzida por Jørgensen (1987).

Os modelos lineares generalizados são definidos por (1.1) e pela parte sistemática

$$g(\mu_i) = \eta_i, \quad (1.2)$$

em que $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ é o preditor linear, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, $p < n$, é um vetor de parâmetros desconhecidos a serem estimados, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ representa os valores de variáveis explicativas e $g(\cdot)$ é uma função monótona e diferenciável, denominada função de ligação. Apresentamos a seguir as distribuições mais conhecidas pertencentes à família exponencial.

1.2.1 Casos particulares

Normal

Seja Y uma variável aleatória com distribuição normal de média μ e variância σ^2 , $Y \sim N(\mu, \sigma^2)$. A função densidade de Y é expressa na forma

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\} = \exp\left[\left\{\frac{1}{\sigma^2}(\mu y - \frac{\mu^2}{2}) - \frac{1}{2}\{\log 2\pi\sigma^2 + \frac{y^2}{\sigma^2}\}\right\}\right],$$

em que $-\infty < \mu, y < \infty$ e $\sigma^2 > 0$. Logo, para $\theta = \mu$, $b(\theta) = \theta^2/2$, $\phi = \sigma^{-2}$ e $c(y, \phi) = \frac{1}{2}\log\phi/2\pi - \frac{\phi y^2}{2}$ obtemos (1.1). Verificamos facilmente que a função de variância é dada por $V(\mu) = 1$.

Poisson

No caso de $Y \sim P(\mu)$, a função de probabilidades fica dada por

$$e^{-\mu} \mu^y / y! = \exp\{y \log \mu - \mu - \log y!\},$$

em que $\mu > 0$ e $y = 0, 1, \dots$. Fazendo $\log \mu = \theta$, $b(\theta) = e^\theta$, $\phi = 1$ e $c(y, \phi) = -\log y!$ obtemos (1.1). Segue portanto que $V(\mu) = \mu$.

Binomial

Seja Y^* a proporção de sucessos em n ensaios independentes, cada um com probabilidade de ocorrência μ . Assumimos que $nY^* \sim B(n, \mu)$. A função de

probabilidades de Y^* fica então expressa na forma

$$\binom{n}{ny^*} \mu^{ny^*} (1-\mu)^{n-ny^*} = \exp \left\{ \log \binom{n}{ny^*} + ny^* \log \left(\frac{\mu}{1-\mu} \right) + n \log(1-\mu) \right\},$$

em que $0 < \mu, y^* < 1$. Obtemos (1.1) fazendo $\phi = n$, $\theta = \log\{\mu/(1-\mu)\}$, $b(\theta) = \log(1+e^\theta)$ e $c(y^*, \phi) = \log\left(\frac{\phi}{\phi y^*}\right)$. A função de variância aqui fica dada por $V(\mu) = \mu(1-\mu)$.

Gama

Seja Y uma variável aleatória com distribuição gama de média μ e coeficiente de variação $\phi^{-1/2}$, denotamos $Y \sim G(\mu, \phi)$. A função densidade de Y é dada por

$$\frac{1}{\Gamma(\phi)} \left(\frac{\phi y}{\mu} \right)^\phi \exp \left(-\frac{\phi y}{\mu} \right) d(\log y) = \exp[\phi\{(-y/\mu) - \log \mu\} - \log \Gamma(\phi) + \phi \log(\phi y) - \log y],$$

em que $y > 0$, $\phi > 0$, $\mu > 0$ e $\Gamma(\phi) = \int_0^\infty t^{\phi-1} e^{-t} dt$ é a função gama. Logo, fazendo $\theta = -1/\mu$, $b(\theta) = -\log(-\theta)$ e $c(y, \phi) = (\phi-1)\log y + \phi \log \phi - \log \Gamma(\phi)$ obtemos (1.1).

Para $0 < \phi < 1$ a densidade da gama tem uma *pole* na origem e decresce monotonicamente quando $y \rightarrow \infty$. A exponencial é um caso especial quando $\phi = 1$. Para $\phi > 1$ a função densidade assume zero na origem, tem um máximo em $y = \mu - \mu/\phi$ e depois decresce para $y \rightarrow \infty$. A χ_k^2 é um outro caso especial quando $\phi = k/2$ e $\mu = k$. A distribuição normal é obtida fazendo $\phi \rightarrow \infty$. Isto é, quando ϕ é grande $Y \sim N(\mu, \phi^{-1}V(\mu))$. Temos que $\phi = E^2(Y)/\text{Var}(Y)$ é o inverso do coeficiente de variação de Y ao quadrado, ou seja, $\phi = 1/(CV)^2$, em que $CV = \sqrt{\text{Var}(Y)}/E(Y)$. A função de variância da gama é dada por $V(\mu) = \mu^2$.

Normal inversa

Seja Y uma variável aleatória com distribuição normal inversa de média μ e parâmetro de precisão ϕ , denotada por $Y \sim \text{NI}(\mu, \phi)$ e cuja função densidade é dada por

$$\frac{\phi^{1/2}}{\sqrt{2\pi y^3}} \exp \left\{ -\frac{\phi(y - \mu)^2}{2\mu^2 y} \right\} = \exp \left[\phi \left\{ -\frac{y}{2\mu^2} + \frac{1}{\mu} \right\} - \frac{1}{2} \left\{ \log(2\pi y^3 / \phi) + \frac{\phi}{y} \right\} \right],$$

em que $y > 0$, $\mu > 0$. Fazendo $\theta = -\frac{1}{2\mu^2}$, $b(\theta) = -(-2\theta)^{1/2}$ e $c(y, \phi) = \frac{1}{2} \log\{\phi / (2\pi y^3)\} - \frac{\phi}{2y}$ obtemos (1.1). A normal inversa se aproxima da normal quando $\phi \rightarrow \infty$. Ou seja, para ϕ grande temos que $Y \sim N(\mu, \phi^{-1}V(\mu))$. A função de variância fica aqui dada por $V(\mu) = \mu^3$.

Na Tabela 1.1 apresentamos um resumo dessas distribuições.

Tabela 1.1

Principais distribuições pertencentes à família exponencial.

Distribuição	$b(\theta)$	θ	ϕ	$V(\mu)$
Normal	$\theta^2/2$	μ	σ^{-2}	1
Poisson	e^θ	$\log \mu$	1	μ
Binomial	$\log(1 + e^\theta)$	$\log\{\mu/(1 - \mu)\}$	n	$\mu(1 - \mu)$
Gama	$-\log(-\theta)$	$-1/\mu$	$1/(CV)^2$	μ^2
N.Inversa	$-\sqrt{-2\theta}$	$-1/2\mu^2$	ϕ	μ^3

1.3 Ligações canônicas

Supondo ϕ conhecido, o logaritmo da função de verossimilhança de um MLG com respostas independentes pode ser expresso na forma

$$L(\beta) = \sum_{i=1}^n \phi\{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n c(y_i, \phi).$$

Um caso particular importante ocorre quando o parâmetro canônico (θ) coincide com o preditor linear, isto é, quando $\theta_i = \eta_i = \sum_{j=1}^p x_{ij} \beta_j$. Nesse caso,

$L(\boldsymbol{\beta})$ fica dado por

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \phi \left\{ y_i \sum_{j=1}^p x_{ij} \beta_j - b \left(\sum_{j=1}^p x_{ij} \beta_j \right) \right\} + \sum_{i=1}^n c(y_i, \phi).$$

Definindo a estatística $S_j = \phi \sum_{i=1}^n Y_i x_{ij}$, $L(\boldsymbol{\beta})$ fica então reexpresso na forma

$$L(\boldsymbol{\beta}) = \sum_{j=1}^p s_j \beta_j - \phi \sum_{i=1}^n b \left(\sum_{j=1}^p x_{ij} \beta_j \right) + \sum_{i=1}^n c(y_i, \phi).$$

Logo, pelo teorema da fatorização a estatística $\mathbf{S} = (S_1, \dots, S_p)^T$ é suficiente minimal para o vetor $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$. As ligações que correspondem a tais estatísticas são chamadas de ligações canônicas e desempenham um papel importante na teoria dos MLGs. As ligações canônicas mais comuns são dadas abaixo.

Distribuição	Normal	Binomial	Poisson	Gama	N. Inversa
Ligação	$\mu = \eta$	$\log \left\{ \frac{\mu}{1-\mu} \right\} = \eta$	$\log \mu = \eta$	$\mu^{-1} = \eta$	$\mu^{-2} = \eta$

Uma das vantagens de usarmos ligações canônicas é que as mesmas garantem a concavidade de $L(\boldsymbol{\beta})$ e consequentemente muitos resultados assintóticos são obtidos mais facilmente. Por exemplo, a concavidade de $L(\boldsymbol{\beta})$ garante a unicidade da estimativa de máxima verossimilhança de $\boldsymbol{\beta}$, quando essa existe. Para ligações não canônicas Wedderburn (1976) discute condições para a existência da concavidade de $L(\boldsymbol{\beta})$.

1.3.1 Outras ligações

Ligação proibito

Seja μ a proporção de sucessos de uma distribuição binomial. A ligação proibito é definida por

$$\Phi^{-1}(\mu) = \eta,$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão.

Ligação complemento log-log

A distribuição do valor extremo (logaritmo da exponencial) tem função densidade dada por

$$f(y) = \exp\{y - \exp(y)\},$$

em que $-\infty < y < \infty$. Logo, a função de distribuição acumulada fica dada por

$$F(y) = 1 - \exp\{-\exp(y)\}.$$

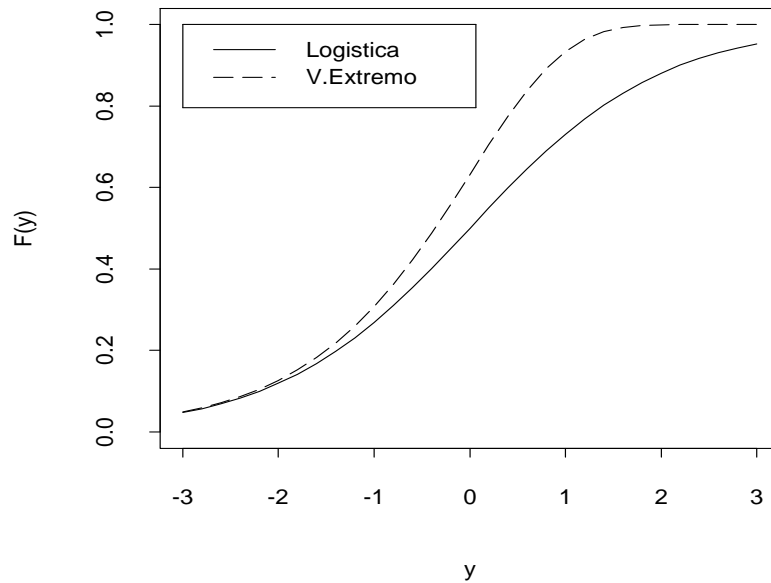


Figura 1.1: Função de distribuição acumulada das curvas logística e do valor extremo.

O modelo binomial com ligação complemento log-log é definido tal que

$$\mu = 1 - \exp\{-\exp(\eta)\},$$

ou, equivalentemente,

$$\log\{-\log(1 - \mu)\} = \eta.$$

A ligação logito é definida de forma similar. A função densidade da distribuição logística é dada por

$$f(y) = \frac{\exp(y)}{\{1 + \exp(y)\}^2},$$

em que $-\infty < y < \infty$. Daí segue que a função de distribuição acumulada fica expressa na forma

$$F(y) = \frac{e^y}{(1 + e^y)}.$$

O modelo logístico binomial é obtido substituindo $F(y)$ por μ e y por η na expressão acima. Como no caso binomial o parâmetro de interesse sempre é uma probabilidade, fica muito razoável que funções de distribuições acumuladas sejam utilizadas para gerarem novas ligações e conseqüentemente novos modelos. Na Figura 1.1 apresentamos a $F(y)$ da distribuição logística e da distribuição do valor extremo para valores de y variando no intervalo $[-3, 3]$. Temos que a curva logística é simétrica em torno de $F(y) = 1/2$, enquanto que a curva do valor extremo apresenta comportamentos distintos para $F(y) \leq 1/2$ e $F(y) > 1/2$.

Ligação de Box-Cox

Uma classe importante de ligações, pelo menos para observações positivas, é a classe de ligações de Box-Cox definida por

$$\eta = (\mu^\lambda - 1)/\lambda,$$

para $\lambda \neq 0$ e $\eta = \log \mu$ para $\lambda \rightarrow 0$. A ideia agora é aplicarmos a transformação de Box-Cox, definida na Seção 1.1, na média da variável resposta

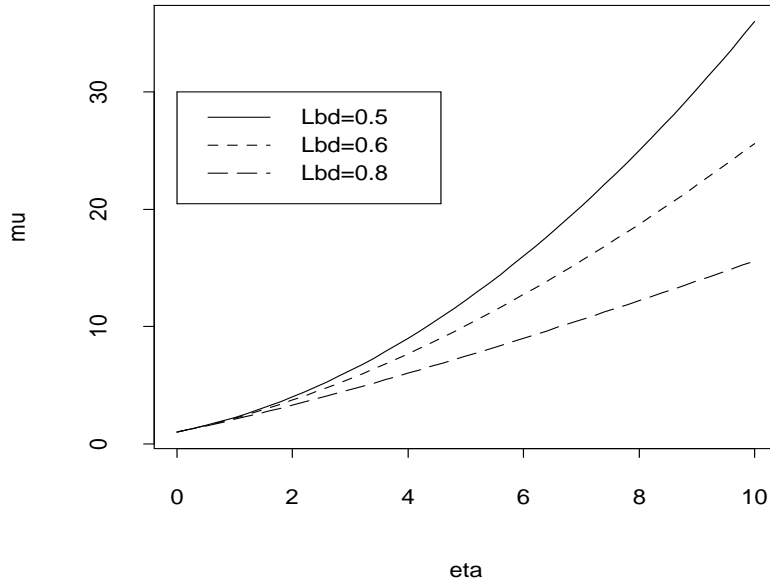


Figura 1.2: Ligação de Box-Cox para alguns valores de λ .

ao invés de transformarmos a própria variável resposta. Temos na Figura 1.2 o comportamento de μ para alguns valores de λ e para η variando no intervalo $[0, 10]$.

Ligação de Aranda-Ordaz

Uma outra transformação importante foi proposta por Aranda-Ordaz (1981) para dados binários. A transformação é dada por

$$\eta = \log \left\{ \frac{(1 - \mu)^{-\alpha} - 1}{\alpha} \right\},$$

em que $0 < \mu < 1$ e α é uma constante desconhecida. Quando $\alpha = 1$ temos a ligação logito $\eta = \log\{\mu/(1-\mu)\}$. Quando $\alpha \rightarrow 0$ temos $\{(1-\mu)^{-\alpha} - 1\}/\alpha \rightarrow \log(1-\mu)^{-1}$ de modo que $\eta = \log\{-\log(1-\mu)\}$ e obtemos portanto a ligação

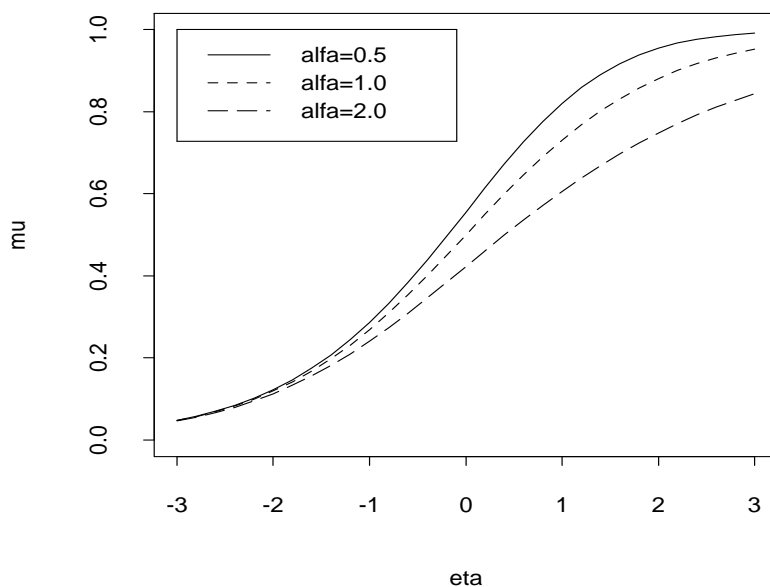


Figura 1.3: Ligação de Aranda-Ordaz para alguns valores de α .

complemento log-log. Na Figura 1.3 temos o comportamento de μ para alguns valores de α . Em muitas situações práticas o interesse pode ser testarmos se o modelo logístico é apropriado, $H_0 : \alpha = 1$, contra a necessidade de uma transformação na ligação, $H_1 : \alpha \neq 1$.

Os MLGs são ajustados no aplicativo R através do comando `glm`. Para ilustrarmos uma aplicação, vamos supor que temos interesse em ajustar um modelo de Poisson com ligação canônica e que a variável resposta é denotada por `resp` com variáveis explicativas `cov1` e `cov2`. Podemos mandar os resultados do ajuste para um arquivo (objeto no R), por exemplo com nome `fit.poisson`, através do comando

```
fit.poisson = glm( resp ~ cov1 + cov2, family=poisson).
```

Com o comando

```
summary(fit.poisson)
```

podemos obter um resumo dos resultados do ajuste.

1.4 Função desvio

Sem perda de generalidade, vamos supor que o logaritmo da função de verossimilhança seja agora definido por

$$L(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^n L(\mu_i; y_i),$$

em que $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$. Para o modelo saturado ($p = n$) a função $L(\boldsymbol{\mu}; \mathbf{y})$ é estimada por

$$L(\mathbf{y}; \mathbf{y}) = \sum_{i=1}^n L(y_i; y_i).$$

Ou seja, a estimativa de máxima verossimilhança de μ_i fica nesse caso dada por $\tilde{\mu}_i = y_i$. Quando $p < n$, denotamos a estimativa de $L(\boldsymbol{\mu}; \mathbf{y})$ por $L(\hat{\boldsymbol{\mu}}; \mathbf{y})$. Aqui, a estimativa de máxima verossimilhança de μ_i será dada por $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$, em que $\hat{\eta}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$.

A qualidade do ajuste de um MLG é avaliada através da função desvio

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\{L(\mathbf{y}; \mathbf{y}) - L(\hat{\boldsymbol{\mu}}; \mathbf{y})\},$$

que é uma distância entre o logaritmo da função de verossimilhança do modelo saturado (com n parâmetros) e do modelo sob investigação (com p parâmetros) avaliado na estimativa de máxima verossimilhança $\hat{\boldsymbol{\beta}}$. Um valor pequeno para a função desvio indica que, para um número menor de parâmetros, obtemos um ajuste tão bom quanto o ajuste com o modelo saturado. Denotando por $\hat{\theta}_i = \theta_i(\hat{\mu}_i)$ e $\tilde{\theta}_i = \theta_i(\tilde{\mu}_i)$ as estimativas de máxima verossimilhança de θ para os modelos com p parâmetros ($p < n$) e saturado ($p = n$), respectivamente, temos que a função $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ fica, alternativamente, dada por

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \{y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i))\}.$$

Apresentamos a seguir a função desvio para alguns casos particulares. O desvio no R sai com o nome *deviance* após o ajuste do modelo e o número de graus de liberdade correspondente é dado por $n - p$. Denotamos $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n d^2(y_i; \hat{\mu}_i)$ em que $d^2(y_i; \hat{\mu}_i)$ será denominado componente do desvio não escalonado.

Normal

Aqui $\theta_i = \mu_i$, logo $\tilde{\theta}_i = y_i$ e $\hat{\theta}_i = \hat{\mu}_i$. O desvio fica portanto dado por

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \{y_i(y_i - \hat{\mu}_i) + \hat{\mu}_i^2/2 - y_i^2/2\} = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2,$$

que coincide com a soma de quadrados de resíduos.

Poisson

Nesse caso temos $\theta_i = \log \mu_i$, o que implica em $\tilde{\theta}_i = \log y_i$ para $y_i > 0$ e $\hat{\theta}_i = \log \hat{\mu}_i$. Assim,

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \{y_i \log(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)\}.$$

Se $y_i = 0$ o i -ésimo termo de $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ vale $2\hat{\mu}_i$. Resumindo, temos o seguinte resultado para o modelo de Poisson:

$$d^2(y_i; \hat{\mu}_i) = \begin{cases} 2\{y_i \log(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)\} & \text{se } y_i > 0; \\ 2\hat{\mu}_i & \text{se } y_i = 0. \end{cases}$$

Binomial

No caso binomial em que assumimos $Y_i \sim B(n_i, \mu_i)$, $i = 1, \dots, k$, obtemos $\tilde{\theta}_i = \log\{y_i / (n_i - y_i)\}$ e $\hat{\theta}_i = \log\{\hat{\mu}_i / (1 - \hat{\mu}_i)\}$ para $0 < y_i < n_i$. Logo, o desvio assume a seguinte forma:

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^k [y_i \log(y_i / n_i \hat{\mu}_i) + (n_i - y_i) \log\{(1 - y_i / n_i) / (1 - \hat{\mu}_i)\}].$$

Todavia, quando $y_i = 0$ ou $y_i = n_i$, o i -ésimo termo de $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ vale $-2n_i \log(1 - \hat{\mu}_i)$ ou $-2n_i \log \hat{\mu}_i$, respectivamente. Portanto, os componentes do desvio no caso binomial assumem as seguintes formas:

$$d^2(y_i; \hat{\mu}_i) = \begin{cases} y_i \log(y_i/n_i \hat{\mu}_i) + (n_i - y_i) \log\{(1 - y_i/n_i)/(1 - \hat{\mu}_i)\} & \text{se } 0 < y_i < n_i; \\ -2n_i \log(1 - \hat{\mu}_i) & \text{se } y_i = 0; \\ -2n_i \log \hat{\mu}_i & \text{se } y_i = n_i. \end{cases}$$

Gama

No caso gama, $\tilde{\theta}_i = -1/y_i$ e $\hat{\theta}_i = -1/\hat{\mu}_i$. Assim, segue que o desvio (quando todos os valores são positivos) pode ser expresso na forma

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \{-\log(y_i/\hat{\mu}_i) + (y_i - \hat{\mu}_i)/\hat{\mu}_i\}.$$

Se algum componente de y_i é igual a zero o desvio fica indeterminado. McCullagh e Nelder (1989) sugerem substituir $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ nesse caso por

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\phi C(\mathbf{y}) + 2\phi \sum_{i=1}^n \log \hat{\mu}_i + 2\phi \sum_{i=1}^n y_i/\hat{\mu}_i,$$

em que $C(\mathbf{y})$ é uma função arbitrária, porém limitada. Podemos, por exemplo, usar $C(\mathbf{y}) = \sum_{i=1}^n y_i/(1 + y_i)$.

Normal inversa

Para esse caso $\tilde{\theta}_i = -1/2y_i^2$ e $\hat{\theta}_i = -1/2\hat{\mu}_i^2$. A função desvio fica então dada por

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / (y_i \hat{\mu}_i^2).$$

1.4.1 Resultados assintóticos

Embora seja usual compararmos os valores observados da função desvio com os percentis da distribuição qui-quadrado com $n - p$ graus de liberdade,

em geral $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ não segue assintoticamente uma χ^2_{n-p} . No caso binomial quando k é fixo e $n_i \rightarrow \infty$ para cada i , $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ segue sob a hipótese de que o modelo é verdadeiro uma χ^2_{k-p} . Isso não vale quando $n \rightarrow \infty$ e $n_i \mu_i (1 - \mu_i)$ permanece limitado. Para o modelo de Poisson, quando $\mu_i \rightarrow \infty$ para todo i , temos que $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \chi^2_{n-p}$. No caso normal, como é conhecido para σ^2 fixo, $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \sigma^2 \chi^2_{n-p}$. Lembre que $E\{\chi_r^2\} = r$, assim um valor do desvio próximo de $n - p$ pode ser uma indicação de que o modelo está bem ajustado. Em geral, para os casos em que $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ depende do parâmetro de dispersão ϕ^{-1} , o seguinte resultado (Jørgensen, 1987) para a distribuição nula da função desvio pode ser utilizado:

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \chi^2_{n-p}, \quad \text{quando } \phi \rightarrow \infty.$$

Isto é, quando a dispersão é pequena, fica razoável compararmos os valores observados de $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ com os percentis da χ^2_{n-p} . Em particular, para o caso normal linear, o resultado acima diz que $\sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / \sigma^2 \sim \chi^2_{n-p}$ quando $\sigma^2 \rightarrow 0$. No caso do modelo gama, o desvio estará bem aproximado por uma qui-quadrado com $n - p$ graus de liberdade à medida que o coeficiente de variação ficar próximo de zero.

1.4.2 Análise do desvio

Vamos supor para o vetor de parâmetros $\boldsymbol{\beta}$ a partição $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$, em que $\boldsymbol{\beta}_1$ é um vetor q -dimensional enquanto $\boldsymbol{\beta}_2$ tem dimensão $p - q$ e ϕ é conhecido (ou fixo). Portanto, podemos estar interessados em testar as hipóteses $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ contra $H_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}$. As funções desvio correspondentes aos modelos sob H_0 e H_1 serão denotadas por $D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0)$ e $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$, respectivamente, em que $\hat{\boldsymbol{\mu}}^0$ é a estimativa de máxima verossimilhança sob H_0 . A estatística da

razão de verossimilhanças fica nesse caso dada por

$$\xi_{RV} = \phi \{D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})\}, \quad (1.3)$$

isto é, a diferença entre dois desvios. Como é conhecido, sob a hipótese nula, $\xi_{RV} \sim \chi_q^2$ quando $n \rightarrow \infty$. De forma similar, podemos definir a estatística

$$F = \frac{\{D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})\}/q}{D(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n-p)}, \quad (1.4)$$

cujas distribuição nula assintótica é uma $F_{q,(n-p)}$ quando o denominador de (1.4) é uma estimativa consistente de ϕ^{-1} (ver, por exemplo, Jørgensen, 1987). A vantagem de utilizarmos (1.4) em relação a (1.3) é que a estatística F não depende do parâmetro de dispersão. O resultado (1.4) também é verificado quando $\phi \rightarrow \infty$ e n é arbitrário. Quando ϕ é desconhecido a estatística da razão de verossimilhanças assume uma expressão diferente de (1.3). A estatística F acima fica, no caso normal linear, reduzida à forma conhecida dada abaixo

$$F = (qs^2)^{-1} \left\{ \sum_{i=1}^n (y_i - \hat{\mu}_i^0)^2 - \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 \right\},$$

em que $s^2 = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / (n-p)$ é o erro quadrático médio do modelo com p parâmetros. A forma da estatística F dada em (1.4) pode ser obtida, em particular, quando testamos uma hipótese de igualdades lineares num modelo de regressão normal linear. Como ilustração, vamos supor o modelo

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\epsilon},$$

em que $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, \mathbf{X} é uma matriz $n \times p$, \mathbf{I}_n é a matriz identidade de ordem n , \mathbf{W} é aqui uma matriz $n \times q$, ambas de posto completo, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^T$. Vamos supor as hipóteses

$$H_0 : \mathbf{C}\boldsymbol{\theta} = \mathbf{0} \quad \text{contra} \quad H_1 : \mathbf{C}\boldsymbol{\theta} \neq \mathbf{0},$$

em que $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ e \mathbf{C} é uma matriz $k \times (p + q)$ de posto completo. O acréscimo na soma de quadrados de resíduos devido às restrições em H_0 é dado por

$$\text{ASQ}(\mathbf{C}\boldsymbol{\theta} = \mathbf{0}) = (\mathbf{C}\hat{\boldsymbol{\theta}})^T \{ \mathbf{C}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{C}^T \}^{-1} (\mathbf{C}\hat{\boldsymbol{\theta}}),$$

em que $\hat{\boldsymbol{\theta}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y}$ e $\mathbf{Z} = (\mathbf{X}, \mathbf{W})$. A estatística F para testarmos H_0 fica então dada por

$$F = \frac{\text{ASQ}(\mathbf{C}\boldsymbol{\theta} = \mathbf{0})/k}{D(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n - p - q)},$$

em que $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ é o desvio do modelo completo com $p + q$ parâmetros e $\text{ASQ}(\mathbf{C}\boldsymbol{\theta} = \mathbf{0}) = D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})$, com $D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0)$ sendo o desvio do modelo sob H_0 . Portanto, F assume a forma

$$F = \frac{\{D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})\}/k}{D(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n - p - q)},$$

e segue, sob H_0 , uma distribuição $F_{k, (n-p-q)}$. No caso de testarmos $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ contra $H_1 : \boldsymbol{\gamma} \neq \mathbf{0}$, a matriz \mathbf{C} tem dimensão $q \times (p + q)$ com a i -ésima linha tendo o valor 1 na posição $p + i$ e zeros nas demais posições. Essa formulação pode também ser aplicada quando testamos a inclusão de novas covariáveis num modelo de regressão normal linear.

Para ilustrarmos o uso das diferenças de desvios para hipóteses em modelos encaixados, vamos supor um MLG com dois fatores, A e B. O fator A com $n(A)$ níveis e o fator B com $n(B)$ níveis. Descrevemos na Tabela 1.2 os possíveis testes envolvendo os dois fatores. Em particular, se o interesse é testarmos a inclusão do fator B dado que o fator A já está no modelo, devemos comparar a diferença $\phi\{D(\mathbf{y}; \hat{\boldsymbol{\mu}}_A) - D(\mathbf{y}; \hat{\boldsymbol{\mu}}_{A+B})\}$ com os níveis críticos da distribuição qui-quadrado com $\{n(B) - 1\}$ graus de liberdade. Alternativamente, podemos comparar o valor observado da estatística F correspondente com os níveis da distribuição F com $\{n(B) - 1\}$ e $\{n - n(A) - n(B) + 1\}$

graus de liberdade. No caso normal linear construímos a tabela **ANOVA** utilizando a estatística F no lugar da diferença entre desvios. A vantagem disso é o fato do parâmetro de dispersão ϕ^{-1} não precisar ser estimado. Através do comando `anova()` o R fornece uma tabela **ANODEV** para os ajustes colocados como objetos. Por exemplo, vamos supor que os objetos `fit1.reg`, `fit2.reg` e `fit3.reg` correspondam aos ajustes de um MLG com um, dois e três fatores, respectivamente. Então, o comando

```
anova(fit1.reg,fit2.reg,fit3.reg)
```

fornece uma tabela **ANODEV** comparando os três fatores.

Tabela 1.2

Análise do desvio (ANODEV) supondo dois fatores na parte sistemática.

Modelo	Desvio	Diferença	G.L.	Testando
Constante	D_0			
		$D_0 - D_A$	$n(A) - 1$	A ignorando B
		$D_0 - D_B$	$n(B) - 1$	B ignorando A
+A	D_A			
		$D_A - D_{A+B}$	$n(B) - 1$	B A ignorando AB
+B	D_B			
		$D_B - D_{A+B}$	$n(A) - 1$	A B ignorando AB
+A+B	D_{A+B}			
		$D_{A+B} - D_{AB}$	$\{n(A) - 1\} \times$ $\{n(B) - 1\}$	AB A + B
+A+B+AB	D_{AB}			

Como aplicação do **ANODEV**, vamos considerar o exemplo descrito na Seção 1.12.2 em que um modelo logístico linear é ajustado para explicar a ocorrência de câncer de pulmão em pacientes com processo infeccioso pulmonar. A parte sistemática do modelo é representada abaixo

$$1 + \text{SEXO} + \text{IDADE} + \text{HL} + \text{FF},$$

em que 1 denota a presença de intercepto no modelo, SEXO (1:feminino, 0:masculino), IDADE (em anos) e HL e FF são dois fatores com 4 níveis

cada um representando a intensidade de dois tipos de célula. Na Tabela 1.3 resumimos alguns resultados.

Tabela 1.3
Análise do desvio referente ao exemplo sobre processo infeccioso pulmonar.

Modelo	Desvio	Diferença	G.L.	Testando
Constante	236,34	-	-	-
+ SEXO	235,20	1,14	1	SEXO
+ IDADE	188,22	46,98	1	IDADE SEXO
+ HL	162,55	25,67	3	HL SEXO + IDADE
+ FF	157,40	5,15	3	FF SEXO + IDADE + HL

Para calcularmos os níveis descritivos das diferenças apresentadas na Tabela 1.3, usamos o comando `pchisq(dv,q)` do R. Por exemplo, para calcularmos o nível descritivo referente ao efeito do fator SEXO, fazemos

```
1 - pchisq(1.14,1)
```

e obtemos $P = 0,285$. Similarmente, para testarmos a inclusão de FF dado que já temos no modelo $1 + \text{SEXO} + \text{IDADE} + \text{HL}$, fazemos

```
1 - pchisq(5.15,3)
```

e obtemos $P = 0,1611$, indicando que o fator FF é não significativo a 10%.

1.5 Função escore e informação de Fisher

1.5.1 Escore e Fisher para β

Vamos considerar a partição $\theta = (\beta^T, \phi)^T$ e denotarmos o logaritmo da função de verossimilhança por $L(\theta)$. Para obtermos a função escore para o

parâmetro β calculamos inicialmente as derivadas

$$\begin{aligned}
\partial L(\theta)/\partial \beta_j &= \sum_{i=1}^n \phi \left\{ y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\beta_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
&= \sum_{i=1}^n \phi \{ y_i V_i^{-1} (d\mu_i/d\eta_i) x_{ij} - \mu_i V_i^{-1} (d\mu_i/d\eta_i) x_{ij} \} \\
&= \sum_{i=1}^n \phi \left\{ \sqrt{\frac{\omega_i}{V_i}} (y_i - \mu_i) x_{ij} \right\},
\end{aligned}$$

em que $\omega_i = (d\mu_i/d\eta_i)^2/V_i$. Logo, podemos escrever a função escore na forma matricial

$$\mathbf{U}_\beta(\theta) = \frac{\partial L(\theta)}{\partial \beta} = \phi \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}),$$

em que \mathbf{X} é uma matriz $n \times p$ de posto completo cujas linhas serão denotadas por $\mathbf{x}_i^T, i = 1, \dots, n$, $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$ é a matriz de pesos, $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$, $\mathbf{y} = (y_1, \dots, y_n)^T$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$.

Para obtermos a matriz de informação de Fisher precisamos das derivadas

$$\begin{aligned}
\partial^2 L(\theta)/\partial \beta_j \partial \beta_\ell &= \phi \sum_{i=1}^n (y_i - \mu_i) \frac{d^2 \theta_i}{d\mu_i^2} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{i\ell} \\
&+ \phi \sum_{i=1}^n (y_i - \mu_i) \frac{d\theta_i}{d\mu_i} \frac{d^2 \mu_i}{d\eta_i^2} x_{ij} x_{i\ell} - \phi \sum_{i=1}^n \frac{d\theta_i}{d\mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{i\ell},
\end{aligned}$$

cujos valores esperados ficam dados por

$$\begin{aligned}
E \{ \partial^2 L(\theta)/\partial \beta_j \partial \beta_\ell \} &= -\phi \sum_{i=1}^n \frac{d\theta_i}{d\mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{i\ell} \\
&= -\phi \sum_{i=1}^n \frac{(d\mu_i/d\eta_i)^2}{V_i} x_{ij} x_{i\ell} \\
&= -\phi \sum_{i=1}^n \omega_i x_{ij} x_{i\ell}.
\end{aligned}$$

Logo, podemos escrever a informação de Fisher para β na forma matricial

$$\mathbf{K}_{\beta\beta}(\theta) = E \left\{ -\frac{\partial^2 L(\theta)}{\partial \beta \partial \beta^T} \right\} = \phi \mathbf{X}^T \mathbf{W} \mathbf{X}.$$

Em particular, para ligação canônica ($\theta_i = \eta_i$), essas quantidades tomam formas simplificadas

$$\mathbf{U}_\beta = \phi \mathbf{X}^T (\mathbf{y} - \boldsymbol{\mu}) \quad \text{e} \quad \mathbf{K}_{\beta\beta} = \phi \mathbf{X}^T \mathbf{V} \mathbf{X},$$

respectivamente. Se particionarmos o vetor de parâmetros tal que $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$, a função escore e a matriz de informação de Fisher ficam para o parâmetro $\boldsymbol{\beta}_1$, respectivamente, dadas por $\mathbf{U}_{\beta_1} = \phi \mathbf{X}_1^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$ e $\mathbf{K}_{\beta_1\beta_1} = \phi \mathbf{X}_1^T \mathbf{W} \mathbf{X}_1$.

1.5.2 Escore e Fisher para ϕ

A função escore para o parâmetro ϕ fica dada por

$$\begin{aligned} \mathbf{U}_\phi(\boldsymbol{\theta}) &= \frac{\partial \mathbf{L}(\boldsymbol{\theta})}{\partial \phi} \\ &= \sum_{i=1}^n \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n c'(y_i, \phi), \end{aligned}$$

em que $c'(y_i, \phi) = dc(y_i, \phi)/d\phi$. Para obtermos a informação de Fisher para ϕ temos que calcular $\partial^2 \mathbf{L}(\boldsymbol{\theta})/\partial \phi^2 = \sum_{i=1}^n c''(y_i, \phi)$, em que $c''(y_i, \phi) = d^2 c(y_i, \phi)/d\phi^2$. Assim, a informação de Fisher para ϕ fica dada por

$$\mathbf{K}_{\phi\phi}(\boldsymbol{\theta}) = - \sum_{i=1}^n \mathbf{E}\{c''(Y_i, \phi)\}.$$

1.5.3 Ortogonalidade

Temos que $\partial^2 \mathbf{L}(\boldsymbol{\theta})/\partial \boldsymbol{\beta} \partial \phi = \sum_{i=1}^n \sqrt{\omega_i V_i^{-1}} (y_i - \mu_i) \mathbf{x}_i$. Portanto, verificamos facilmente que $\boldsymbol{\beta}$ e ϕ são ortogonais, isto é, $\mathbf{K}_{\beta\phi}(\boldsymbol{\theta}) = \mathbf{E}[-\partial^2 \mathbf{L}(\boldsymbol{\theta})/\partial \boldsymbol{\beta} \partial \phi] = \mathbf{0}$. Logo, segue que a matriz de informação de Fisher para $\boldsymbol{\theta}$ é bloco diagonal sendo dada por $\mathbf{K}_{\theta\theta} = \text{diag}\{\mathbf{K}_{\beta\beta}, \mathbf{K}_{\phi\phi}\}$. A função escore para $\boldsymbol{\theta}$ fica dada por $\mathbf{U}_\theta = (\mathbf{U}_\beta^T, \mathbf{U}_\phi^T)^T$. Discutimos a seguir alguns casos particulares.

1.5.4 Casos particulares

Normal

A função de variância no caso normal é dada por $V(\mu) = 1$ ($d\mu/d\theta = 1$). Logo, $\omega = (d\theta/d\eta)^2$. Em particular para ligação canônica ($\theta = \eta$), obtemos $\omega = 1$. Assim,

$$\mathbf{U}_\beta = \sigma^{-2} \mathbf{X}^T (\mathbf{y} - \boldsymbol{\mu}) \quad \text{e} \quad \mathbf{K}_{\beta\beta} = \sigma^{-2} \mathbf{X}^T \mathbf{X},$$

como é conhecido. Temos ainda o resultado

$$U_\phi = \sum_{i=1}^n (y_i \mu_i - \frac{\mu_i^2}{2}) + \sum_{i=1}^n c'(y_i, \phi),$$

em que $c'(y_i, \phi) = 1/2\phi - y_i^2/2$. Daí segue que $c''(y_i, \phi) = -1/2\phi^2$ e portanto $K_{\phi\phi} = -\sum_{i=1}^n E\{c''(Y_i, \phi)\} = n/2\phi^2$.

Poisson

Aqui a função de variância é dada por $V(\mu) = \mu$. Logo, $\omega = \mu(d\theta/d\eta)^2$. Para ligação canônica ($\log \mu = \eta$) os pesos são as próprias médias, isto é $\omega = \mu$. Em particular, para ligação raiz quadrada ($\sqrt{\mu} = \eta$), obtemos $\omega = 4$. Assim, $\mathbf{U}_\beta = \mathbf{X}^T \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$ e $\mathbf{K}_{\beta\beta} = \mathbf{X}^T \mathbf{X}$.

Binomial

No caso binomial, a função de variância é definida por $V(\mu) = \mu(1 - \mu)$, em que $0 < \mu < 1$. Portanto, temos $\omega = \mu(1 - \mu)(d\theta/d\eta)^2$. Por convenção assumimos que $\omega = n\mu(1 - \mu)(d\theta/d\eta)^2$ e $\phi = 1$. No caso de ligação canônica ($\log\{\mu/(1 - \mu)\} = \eta$) os pesos são as variâncias das binomiais, isto é $\omega = n\mu(1 - \mu)$. As matrizes \mathbf{U}_β e $\mathbf{K}_{\beta\beta}$ ficam nesse caso dadas por

$$\mathbf{U}_\beta = \mathbf{X}^T (\mathbf{y} - \boldsymbol{\mu}) \quad \text{e} \quad \mathbf{K}_{\beta\beta} = \mathbf{X}^T \mathbf{V} \mathbf{X},$$

em que \mathbf{X} é uma matriz $k \times p$, $\mathbf{y} = (y_1, \dots, y_n)^T$, $\boldsymbol{\mu} = (n_1\mu_1, \dots, n_k\mu_k)^T$ e $\mathbf{V} = \text{diag}\{n_1\mu_1(1 - \mu_1), \dots, n_k\mu_k(1 - \mu_k)\}$.

Gama

Para o caso gama $V(\mu) = \mu^2$. Logo, $\omega = \mu^2(d\theta/d\eta)^2$. Em particular, para um modelo log-linear ($\log\mu = \eta$), obtemos $d\mu/d\eta = \mu$, o que implica em $\omega = 1$. Assim, $\mathbf{U}_\beta = \phi\mathbf{X}^T\mathbf{V}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$ e $\mathbf{K}_{\beta\beta} = \phi\mathbf{X}^T\mathbf{X}$, similarmente ao caso normal. Para ligação canônica, $\omega = \mu^2$. A função escore para ϕ fica dada por

$$U_\phi = - \sum_{i=1}^n \left(\frac{y_i}{\mu_i} + \log\mu_i \right) + \sum_{i=1}^n c'(y_i, \phi),$$

em que $c'(y_i, \phi) = \log y_i + \log\phi + 1 - \psi(\phi)$ e $\psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$ é a função digama. Daí segue que $c''(y_i, \phi) = 1/\phi - \psi'(\phi)$ e portanto $\mathbf{K}_{\phi\phi} = - \sum_{i=1}^n \mathbf{E}\{c''(Y_i, \phi)\} = n\{\phi\psi'(\phi) - 1\}/\phi$, em que $\psi'(\phi) = d\psi(\phi)/d\phi$ é a função trigama.

Normal inversa

Nesse caso a função de variância é dada por $V(\mu) = \mu^3$. Assim, $\omega = \mu^3(d\theta/d\eta)^2$. Pode ser muito razoável aplicarmos aqui um modelo log-linear, uma vez que as respostas são sempre positivas. No entanto, diferente dos modelos log-lineares com resposta de Poisson, os pesos aqui são inversamente proporcionais às médias, isto é $\omega = \mu^{-1}$. Em particular para ligação canônica, $\omega = \mu^3$, e portanto $\mathbf{U}_\beta = \phi\mathbf{X}^T(\mathbf{y} - \boldsymbol{\mu})$ e $\mathbf{K}_{\beta\beta} = \phi\mathbf{X}^T\mathbf{V}\mathbf{X}$. Temos ainda o resultado

$$U_\phi = \sum_{i=1}^n \left(\frac{1}{\mu_i} - \frac{y_i}{2\mu_i^2} \right) + \sum_{i=1}^n c'(y_i, \phi),$$

em que $c'(y_i, \phi) = 1/2\phi - 1/2y_i$. Daí segue que $c''(y_i, \phi) = -1/2\phi^2$ e portanto $\mathbf{K}_{\phi\phi} = - \sum_{i=1}^n \mathbf{E}\{c''(Y_i, \phi)\} = n/2\phi^2$.

1.6 Estimação dos parâmetros

1.6.1 Estimação de β

O processo iterativo de Newton-Raphson para a obtenção da estimativa de máxima verossimilhança de β é definido expandindo a função escore \mathbf{U}_β em torno de um valor inicial $\beta^{(0)}$, tal que

$$\mathbf{U}_\beta \cong \mathbf{U}_\beta^{(0)} + \mathbf{U}_\beta^{\prime(0)}(\beta - \beta^{(0)}),$$

em que \mathbf{U}_β' denota a primeira derivada de \mathbf{U}_β com respeito a β^T , sendo $\mathbf{U}_\beta'(0)$ e $\mathbf{U}_\beta^{(0)}$, respectivamente, essas quantidades avaliadas em $\beta^{(0)}$. Assim, repetindo o procedimento acima, chegamos ao processo iterativo

$$\beta^{(m+1)} = \beta^{(m)} + \{(-\mathbf{U}_\beta')^{-1}\}^{(m)} \mathbf{U}_\beta^{(m)},$$

$m = 0, 1, \dots$. Como a matriz $-\mathbf{U}_\beta'$ pode não ser positiva definida, a aplicação do método escore de Fisher substituindo a matriz $-\mathbf{U}_\beta'$ pelo correspondente valor esperado $\mathbf{K}_{\beta\beta}$ pode ser mais conveniente. Isso resulta no seguinte processo iterativo:

$$\beta^{(m+1)} = \beta^{(m)} + \{\mathbf{K}_{\beta\beta}^{-1}\}^{(m)} \mathbf{U}_\beta^{(m)},$$

$m = 0, \dots$. Se trabalharmos um pouco o lado direito da expressão acima, chegaremos a um processo iterativo de mínimos quadrados ponderados

$$\beta^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)}, \quad (1.5)$$

$m = 0, 1, \dots$, em que $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$. A quantidade \mathbf{z} desempenha o papel de uma variável dependente modificada, enquanto \mathbf{W} é uma matriz de pesos que muda a cada passo do processo iterativo. A convergência de (1.5) ocorre em geral num número finito de passos, independente dos valores iniciais utilizados. É usual iniciarmos (1.5) com $\boldsymbol{\eta}^{(0)} = \mathbf{g}(\mathbf{y})$.

Apenas como ilustração, para o caso logístico binomial, obtemos $\omega = n\mu(1-\mu)$ e variável dependente modificada dada por $z = \eta + (y - n\mu)/n\mu(1 - \mu)$. Lembrando, para o modelo normal linear não é preciso recorrermos ao processo iterativo (1.5) para a obtenção da estimativa de máxima verossimilhança. Nesse caso, $\hat{\beta}$ assume a forma fechada

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

Observamos que o lado direito de (1.5) não depende de ϕ . Portanto, para obtermos $\hat{\beta}$ não é preciso conhecermos ϕ .

1.6.2 Estimação de ϕ

Igualando a função escore U_ϕ a zero chegamos à seguinte solução:

$$\sum_{i=1}^n c'(y_i, \hat{\phi}) = \frac{1}{2} D(\mathbf{y}; \hat{\mu}) - \sum_{i=1}^n \{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)\},$$

em que $D(\mathbf{y}; \hat{\mu})$ denota o desvio do modelo sob investigação. Verificamos que a estimativa de máxima verossimilhança para ϕ nos casos normal e normal inversa, igualando U_ϕ a zero, é dada por

$$\hat{\phi} = \frac{n}{D(\mathbf{y}; \hat{\mu})}.$$

Para o caso gama, a estimativa de máxima verossimilhança de ϕ sai da equação

$$2n\{\log \hat{\phi} - \psi(\hat{\phi})\} = D(\mathbf{y}; \hat{\mu}).$$

A equação acima pode ser resolvida diretamente pelo R através do comando `require(MASS)` (Venables e Ripley, 1999). Como ilustração, vamos supor que os resultados do ajuste sejam guardados em `fit.model`. Então, para encontrarmos a estimativa de máxima verossimilhança de ϕ com o respectivo erro padrão aproximado devemos usar os comandos

```
require(MASS)
gamma.shape(fit.model).
```

Um outro estimador consistente para ϕ (de momentos) que não envolve processo iterativo é baseado na estatística de Pearson, sendo dado por

$$\hat{\phi} = \frac{(n-p)}{\sum_{i=1}^n \left\{ \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \right\}}.$$

A suposição aqui é que $\hat{\beta}$ tem sido consistentemente estimado. O R solta a estimativa $\hat{\phi} = (n-p)/D(\mathbf{y}; \hat{\mu})$ que não é consistente para ϕ .

1.6.3 Distribuição assintótica

Para mostrarmos que $\hat{\beta}$ e $\hat{\phi}$ são assintoticamente normais e independentes vamos usar os resultados abaixo

$$E(\mathbf{U}_\theta) = \mathbf{0} \quad \text{e} \quad \text{Var}(\mathbf{U}_\theta) = \mathbf{K}_{\theta\theta},$$

com as funções escore de β e ϕ sendo, respectivamente, expressas nas formas $\mathbf{U}_\beta = \sum_{i=1}^n \mathbf{U}_{i\beta}$, em que

$$\mathbf{U}_{i\beta} = \phi \sqrt{\omega_i V_i^{-1}} (y_i - \mu_i) \mathbf{x}_i \quad \text{e} \quad \mathbf{U}_\phi = \sum_{i=1}^n \mathbf{U}_{i\phi},$$

com $\mathbf{U}_{i\phi} = \{y_i \theta_i - b(\theta_i)\} + c'(y_i, \phi)$. Portanto, para n grande, temos que $\mathbf{U}_\theta \sim N_{p+1}(\mathbf{0}, \mathbf{K}_{\theta\theta})$. Em particular, assintoticamente $\mathbf{U}_\beta \sim N_p(\mathbf{0}, \mathbf{K}_{\beta\beta})$ e $\mathbf{U}_\phi \sim N(0, \mathbf{K}_{\phi\phi})$ e \mathbf{U}_β e \mathbf{U}_ϕ são independentes.

Expandindo $\mathbf{U}_{\hat{\theta}}$ em série de Taylor em torno de θ obtemos

$$\mathbf{U}_{\hat{\theta}} \cong \mathbf{U}_\theta + \mathbf{U}'_\theta(\hat{\theta} - \theta),$$

em que $\mathbf{U}'_\theta = \partial \mathbf{U}_\theta / \partial \theta^T$. Assim, como $\hat{\theta}$ é o estimador de máxima verossimilhança de θ temos que $\mathbf{U}_{\hat{\theta}} = \mathbf{0}$ e daí segue a relação

$$\hat{\theta} \cong \theta + (-\mathbf{U}'_\theta)^{-1} \mathbf{U}_\theta.$$

Supondo que para n grande $-\mathbf{U}'_\theta \cong \mathbf{K}_{\theta\theta}$ (para ligação canônica $\mathbf{K}_{\beta\beta} = -\mathbf{U}'_\beta$), então obtemos

$$\hat{\boldsymbol{\theta}} \cong \boldsymbol{\theta} + \mathbf{K}_{\theta\theta}^{-1} \mathbf{U}_\theta,$$

ou seja, para n grande $\hat{\boldsymbol{\theta}} \sim N_{p+1}(\boldsymbol{\theta}, \mathbf{K}_{\theta\theta}^{-1})$. Como $\mathbf{K}_{\theta\theta} = \text{diag}\{\mathbf{K}_{\beta\beta}, \mathbf{K}_{\phi\phi}\}$ então assintoticamente segue que $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \mathbf{K}_{\beta\beta}^{-1})$ e $\hat{\phi} \sim N(0, \mathbf{K}_{\phi\phi}^{-1})$ e $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$ são independentes. Demonstrações mais rigorosas desses resultados podem ser encontradas, por exemplo, em Fahrmeir e Kaufmann (1985) e Sen e Singer (1993, Cap. 7).

1.7 Teste de hipóteses

1.7.1 Hipóteses simples

Buse (1982) apresenta de uma forma bastante didática a interpretação geométrica dos testes da razão de verossimilhanças, escore e Wald para o caso de hipóteses simples. Apresentamos a seguir as generalizações para os MLGs. Vamos supor, inicialmente, a seguinte situação de hipóteses simples:

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^0 \text{ contra } H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}^0,$$

em que $\boldsymbol{\beta}^0$ é um vetor p -dimensional conhecido e ϕ é também assumido conhecido.

Teste da razão de verossimilhanças

O teste da razão de verossimilhanças, no caso de hipóteses simples, é usualmente definido por

$$\xi_{RV} = 2\{L(\hat{\boldsymbol{\beta}}) - L(\boldsymbol{\beta}^0)\}.$$

Essa estatística pode também ser expressa, para os MLGs, como a diferença entre duas funções desvio

$$\xi_{RV} = \phi\{D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})\},$$

em que $\hat{\boldsymbol{\mu}}^0 = \mathbf{g}^{-1}(\hat{\boldsymbol{\eta}}^0)$, $\hat{\boldsymbol{\eta}}^0 = \mathbf{X}\boldsymbol{\beta}^0$. Em particular, para o caso normal linear, obtemos $\xi_{RV} = \{\sum_{i=1}^n (y_i - \hat{\mu}_i^0)^2 - \sum_{i=1}^n (y_i - \hat{\mu}_i)^2\}/\sigma^2$.

Teste de Wald

O teste de Wald é definido, nesse caso, por

$$\xi_W = [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0]^T \hat{\text{Var}}^{-1}(\hat{\boldsymbol{\beta}}) [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0],$$

em que $\hat{\text{Var}}(\hat{\boldsymbol{\beta}})$ denota a matriz de variância-covariância assintótica de $\hat{\boldsymbol{\beta}}$ estimada em $\hat{\boldsymbol{\beta}}$. Para os MLGs, $\hat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \mathbf{K}^{-1}(\hat{\boldsymbol{\beta}})$. Assim, a estatística de Wald fica reexpressa na forma

$$\xi_W = \phi[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0]^T (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X}) [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0].$$

Em particular, para o caso de $p = 1$, o teste de Wald é equivalente ao teste t^2 usual

$$\xi_W = \frac{(\hat{\beta} - \beta^0)^2}{\hat{\text{Var}}(\hat{\beta})}.$$

Um problema com a estatística de Wald, especialmente quando $\boldsymbol{\eta}(\boldsymbol{\beta})$ é não linear em $\boldsymbol{\beta}$, é a dependência de ξ_W com a parametrização utilizada. Isto é, duas formas diferentes e equivalentes para $\boldsymbol{\eta}(\boldsymbol{\beta})$, podem levar a diferentes valores de ξ_W .

Teste de escore

O teste de escore, também conhecido como teste de Rao, é definido quando $\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}) = \mathbf{0}$ por

$$\xi_{SR} = \mathbf{U}_\beta(\boldsymbol{\beta}^0)^T \hat{\text{Var}}_0(\hat{\boldsymbol{\beta}}) \mathbf{U}_\beta(\boldsymbol{\beta}^0),$$

em que $\hat{\text{Var}}_0(\hat{\boldsymbol{\beta}})$ denota que a variância assintótica de $\hat{\boldsymbol{\beta}}$ está sendo estimada sob H_0 . Para os MLGs temos que

$$\xi_{SR} = \phi^{-1} \mathbf{U}_{\beta}(\boldsymbol{\beta}^0)^T (\mathbf{X}^T \hat{\mathbf{W}}_0 \mathbf{X})^{-1} \mathbf{U}_{\beta}(\boldsymbol{\beta}^0),$$

em que $\hat{\mathbf{W}}_0$ é estimado sob H_0 , embora tenha a forma do modelo em H_1 . A estatística de escore pode ser muito conveniente em situações em que a hipótese alternativa é bem mais complicada do que a hipótese nula. Nesses casos, somente seria necessário estimarmos os parâmetros sob H_1 quando o modelo em H_0 fosse rejeitado. Novamente, ilustrando o caso normal linear, temos que a estatística de escore fica expressa na forma $\xi_{SR} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0)^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0) / \sigma^2$. Observamos que, nesse caso, as estatísticas ξ_{RV} e ξ_W coincidem com ξ_{SR} .

Teste F

A estatística F, que foi definida em (1.4), assume a seguinte forma para o caso de hipóteses simples:

$$F = \frac{\{D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})\} / p}{D(\mathbf{y}; \hat{\boldsymbol{\mu}}) / (n - p)},$$

que para $\phi \rightarrow \infty$ e sob H_0 segue uma $F_{p, (n-p)}$. Esse resultado vale também para $n \rightarrow \infty$ quando colocamos no denominador da estatística F uma estimativa consistente para ϕ^{-1} . Uma propriedade interessante das estatísticas ξ_{RV} , ξ_{SR} e F é o fato de serem invariantes com reparametrizações. Isso pode ser muito útil na construção de regiões de confiança para os parâmetros. A estatística F tem a vantagem adicional de não depender do parâmetro de dispersão ϕ^{-1} . Como essa estatística pode ser obtida diretamente de funções desvio, talvez seja a mais conveniente para uso prático. Assintoticamente e sob a hipótese nula, temos que ξ_{RV} , ξ_W e $\xi_{SR} \sim \chi_p^2$.

Uma região assintótica de confiança para β baseada no teste de Wald e com coeficiente de confiança $(1 - \alpha)$, é dada por

$$[\beta; (\hat{\beta} - \beta)^T (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X}) (\hat{\beta} - \beta) \leq \phi^{-1} \chi_p^2(1 - \alpha)],$$

em que $\chi_p^2(1 - \alpha)$ denota o percentil $(1 - \alpha)$ de uma distribuição qui-quadrado com p graus de liberdade. Como essa região pode depender da parametrização utilizada quando η é não linear (ver, por exemplo, Ratkowsky, 1983), pode ser mais conveniente, nesses casos, construirmos a região utilizando uma das estatísticas invariantes. Em particular, se a estatística da razão de verossimilhanças for escolhida, a região assintótica fica dada por

$$[\beta; 2\{L(\hat{\beta}) - L(\beta)\} \leq \chi_p^2(1 - \alpha)].$$

Se estamos interessados num subconjunto β_1 q -dimensional, a região assintótica de confiança utilizando as estatísticas de Wald e da razão de verossimilhanças ficam, respectivamente, dadas por

$$[\beta; (\hat{\beta}_1 - \beta)^T \hat{\text{Var}}^{-1}(\hat{\beta}_1)(\hat{\beta}_1 - \beta) \leq \phi^{-1} \chi_q^2(1 - \alpha)]$$

e

$$[\beta; 2\{L(\hat{\beta}) - L(\beta, \hat{\beta}_2(\beta))\} \leq \chi_q^2(1 - \alpha)],$$

em que β é aqui q -dimensional e $\hat{\beta}_2(\beta)$ é a estimativa de máxima verossimilhança de β_2 dado β (ver, por exemplo, Seber e Wild, 1989).

1.7.2 Modelos encaixados

ϕ conhecido

Vamos supor novamente a partição $\beta = (\beta_1^T, \beta_2^T)^T$ definida na Seção 1.4.2 e as seguintes hipóteses: $H_0 : \beta_1 = \beta_1^0$ contra $H_1 : \beta_1 \neq \beta_1^0$. Para esse caso temos

$$\xi_{RV} = \phi\{D(\mathbf{y}; \hat{\mu}^0) - D(\mathbf{y}; \hat{\mu})\},$$

em que $\hat{\mu}^0$ é a estimativa de máxima verossimilhança do MLG com parte sistemática $\eta = \hat{\eta}_1^0 + \eta_2$, em que $\hat{\eta}_1^0 = \sum_{j=1}^q x_j \beta_j^0$ e $\eta_2 = \sum_{j=q+1}^p x_j \beta_j$. A quantidade $\hat{\eta}_1^0$ desempenha o papel de um *offset* (parte conhecida no preditor linear), conforme a nomenclatura de modelos lineares generalizados. Para ilustrarmos a utilização do *offset*, vamos supor um modelo de Poisson com ligação log-linear, resposta **resp**, covariáveis **cov1** e **cov2** e *offset* dado por **logt0**. Para ajustarmos o modelo e armazenarmos os resultados em **fit1.poisson** devemos fazer

```
fit1.poisson = glm(resp ~ cov1 + cov2 + offset(logt0),
family= poisson).
```

Esse tipo de recurso é muito utilizado em estudos de seguimento em que cada indivíduo é observado durante um tempo diferente (vide Exemplo 1.12.4). Como ilustração, vamos supor um MLG com distribuição normal inversa, ligação canônica e preditor linear dado por $\eta = \beta_1 + \beta_2 \text{cov}_2 + \beta_3 \text{cov}_3$ e que o interesse é testarmos $H_0 : \beta_2 = b$, em que b é uma constante diferente de zero, contra $H_1 : \beta_2 \neq b$. Os ajustes correspondentes a H_0 e H_1 são, respectivamente, dados por

```
fit1.ni = glm( resp ~ cov3 + offset(b*cov2),
family=inverse.gaussian)
```

```
fit2.ni = glm( resp ~ cov2+cov3, family=inverse.gaussian).
```

Logo, de (1.4), a estatística F para testarmos $H_0 : \beta_2 = b$ contra $H_1 : \beta_2 \neq b$ fica dada por

```
d1 = deviance(fit1.ni)
d2 = deviance(fit2.ni)
F = (d1 - d2)/(d2/(n-3)).
```

Em particular, o *offset* desaparece para $b = 0$. O ajuste, nesse caso, fica simplesmente dado por

`fit1.ni = glm(resp ~ cov3, family=inverse.gaussian).`

Teste de Wald

Para testarmos H_0 , a estatística de Wald fica expressa na forma

$$\xi_W = [\hat{\beta}_1 - \beta_1^0]^T \hat{\text{Var}}^{-1}(\hat{\beta}_1) [\hat{\beta}_1 - \beta_1^0],$$

em que $\hat{\beta}_1$ sai do vetor $\hat{\beta} = (\hat{\beta}_1^T, \hat{\beta}_2^T)^T$. Usando resultados conhecidos de álgebra linear, mostramos que a variância assintótica de $\hat{\beta}_1$ é dada por

$$\text{Var}(\hat{\beta}_1) = \phi^{-1} [\mathbf{X}_1^T \mathbf{W}^{1/2} \mathbf{M}_2 \mathbf{W}^{1/2} \mathbf{X}_1]^{-1},$$

em que \mathbf{X}_1 sai da partição $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, sendo portanto $n \times q$, \mathbf{X}_2 é $n \times (p-q)$, $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{H}_2$ e $\mathbf{H}_2 = \mathbf{W}^{1/2} \mathbf{X}_2 (\mathbf{X}_2^T \mathbf{W} \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{W}^{1/2}$ é a matriz de projeção ortogonal de vetores do \mathcal{R}^n no subespaço gerado pelas colunas da matriz $\mathbf{W}^{1/2} \mathbf{X}_2$. Em particular, no caso normal linear, temos as simplificações $\mathbf{H}_2 = \mathbf{X}_2 (\mathbf{X}_2^T \mathbf{X}_2)^{-1} \mathbf{X}_2^T$ e $\text{Var}(\hat{\beta}_1) = \sigma^2 [\mathbf{X}_1^T (\mathbf{I}_n - \mathbf{H}_2) \mathbf{X}_1]^{-1}$.

Teste de escore

A função escore pode ser expressa na forma $\mathbf{U}_\beta = \phi^{1/2} \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{r}_P$, em que $\mathbf{r}_P = \phi^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$ é conhecido como resíduo de Pearson. Observamos que \mathbf{r}_P tem a mesma distribuição de \mathbf{Y} , no entanto, $E(\mathbf{r}_P) = \mathbf{0}$ e $\text{Var}(\mathbf{r}_P) = \mathbf{I}_n$. O teste de escore é definido por

$$\xi_{SR} = \mathbf{U}_{\beta_1}(\hat{\beta}^0)^T \hat{\text{Var}}_0(\hat{\beta}_1) \mathbf{U}_{\beta_1}(\hat{\beta}^0),$$

em que $\mathbf{U}_{\beta_1}(\beta) = \partial L(\beta) / \partial \beta_1 = \phi \mathbf{X}_1^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$, $\hat{\beta}^0 = (\beta_1^{0T}, \hat{\beta}_2^{0T})^T$ e $\hat{\beta}_2^0$ é a estimativa de máxima verossimilhança de β_2 sob o modelo com parte sistemática $\boldsymbol{\eta} = \hat{\boldsymbol{\eta}}_1^0 + \boldsymbol{\eta}_2$, isto é, sob H_0 , em que $\hat{\boldsymbol{\eta}}_1^0 = \mathbf{X}_1 \beta_1^0$ e $\boldsymbol{\eta}_2 = \mathbf{X}_2 \beta_2$.

Se trabalharmos um pouco mais a expressão para $\text{Var}(\hat{\beta}_1)$, chegaremos ao seguinte resultado:

$$\text{Var}(\hat{\beta}_1) = \phi^{-1}(\mathbf{R}^T \mathbf{W} \mathbf{R})^{-1},$$

em que $\mathbf{R} = \mathbf{X}_1 - \mathbf{X}_2 \mathbf{C}$ e $\mathbf{C} = (\mathbf{X}_2^T \mathbf{W} \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{W} \mathbf{X}_1$. Aqui \mathbf{C} é uma matriz $n \times q$ cuja j -ésima coluna é o vetor de coeficientes da regressão linear (com pesos \mathbf{W}) da j -ésima coluna de \mathbf{X}_1 sobre \mathbf{X}_2 . Assim, \mathbf{R} pode ser interpretado como sendo uma matriz $n \times q$ de resíduos. A j -ésima coluna de \mathbf{R} corresponde aos resíduos ordinários da regressão linear (com pesos \mathbf{W}) da j -ésima coluna de \mathbf{X}_1 sobre \mathbf{X}_2 . Assim, o teste de escore fica reexpresso na forma (vide Cordeiro, Ferrari e Paula, 1993)

$$\xi_{SR} = \hat{\mathbf{r}}_{P_0}^T \hat{\mathbf{W}}_0^{1/2} \mathbf{X}_1 (\hat{\mathbf{R}}_0^T \hat{\mathbf{W}}_0 \hat{\mathbf{R}}_0)^{-1} \mathbf{X}_1^T \hat{\mathbf{W}}_0^{1/2} \hat{\mathbf{r}}_{P_0},$$

com as quantidades $\hat{\mathbf{r}}_{P_0}$, $\hat{\mathbf{W}}_0$ e $\hat{\mathbf{R}}_0$ sendo avaliadas em $\hat{\beta}^0$.

Para ilustrarmos o cálculo da estatística de escore, vamos supor um MLG com preditor linear dado por $\eta = \beta_1 + \beta_2 \text{cov}_2 + \beta_3 \text{cov}_3 + \beta_4 \text{cov}_4$ e que o interesse é testarmos $H_0 : \beta_3 = \beta_4 = 0$. As matrizes \mathbf{X}_1 e \mathbf{X}_2 serão então dadas por $\mathbf{X}_1 = [\text{cov}_3, \text{cov}_4]$ e $\mathbf{X}_2 = [1, \text{cov}_2]$. Se temos um modelo de Poisson, por exemplo com ligação canônica, então como já vimos $\omega = \mu$. Logo, $\hat{\mathbf{W}}_0 = \text{diag}\{\hat{\mu}_1^0, \dots, \hat{\mu}_n^0\}$, em que $\hat{\mu}_1^0, \dots, \hat{\mu}_n^0$ são os pesos sob H_0 , ou seja, os pesos do modelo ajustado de Poisson com preditor linear $\eta = \beta_1 + \beta_2 \text{cov}_2$. Portanto, precisamos apenas fazer esse ajuste e daí computamos $\hat{\mathbf{W}}_0$, $\hat{\mathbf{R}}_0$, $\hat{\mathbf{r}}_{P_0}$ e finalmente ξ_{SR} . Chamando no R os pesos por `w`, $\hat{\mathbf{W}}_0$ por `W`, $\hat{\mathbf{r}}_{P_0}$ por `rp` e $\hat{\mathbf{R}}_0$ por `R`, os passos para o cálculo de ξ_{SR} são dados abaixo

```
X1 = cbind(cov3 , cov4)
X2 = cbind(1 , cov2)
fit.poisson = glm( resp ~ cov2, family=poisson)
rp = resid(fit.poisson, type='‘pearson’')
```

```

w = fit.poisson$weights
W = diag(w)
A = solve(t(X2)%*%W%*%X2)
C1 = A%*%t(X2)%*%W%*%cov3
C2 = A%*%t(X2)%*%W%*%cov4
C = cbind(C1 , C2)
R = X1 - X2%*%C
SR = solve(t(R)%*%W%*%R)
SR = t(rp)%*%sqrt(W)%*%X1%*%SR%*%t(X1)%*%sqrt(W)%*%rp.

```

Em particular, para o caso normal linear, $\mathbf{C} = (\mathbf{X}_2^T \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{X}_1$ e $\mathbf{r}_P = (\mathbf{y} - \boldsymbol{\mu})/\sigma$. Logo, $\xi_{SR} = \sigma^{-2}(\mathbf{y} - \hat{\boldsymbol{\mu}}^0)^T \mathbf{X}_1 (\mathbf{R}^T \mathbf{R})^{-1} \mathbf{X}_1^T (\mathbf{y} - \hat{\boldsymbol{\mu}}^0)$, em que $\mathbf{R} = \mathbf{X}_1 - \mathbf{X}_2 (\mathbf{X}_2^T \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{X}_1 = (\mathbf{I}_n - \mathbf{H}_2) \mathbf{X}_1$. Aqui, também as estatísticas da razão de verossimilhanças e de Wald coincidem com a estatística de escore. Isso em geral vale para o modelo normal linear.

A estatística de Wald fica, analogamente ao caso anterior, dada por

$$\xi_W = \phi[\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0]^T [\hat{\mathbf{R}}^T \hat{\mathbf{W}} \hat{\mathbf{R}}] [\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0].$$

O cálculo de $\hat{\mathbf{R}}$ segue os mesmos passos descritos para o cálculo do teste de escore, com a única diferença de que os pesos sairão do ajuste do modelo com todos os parâmetros. As mudanças nos comandos são as seguintes:

```

fit1.poisson = glm( resp ~ cov2 + cov3 + cov4,
family=poisson)
w = fit1.poisson$weights
W = diag(w).

```

Sob H_0 e para grandes amostras, temos que ξ_{RV} , ξ_W e $\xi_{SR} \sim \chi_q^2$.

ϕ desconhecido

No caso de ϕ ser desconhecido e o interesse for testarmos $H_0 : \beta_1 = \beta_1^0$ contra $H_1 : \beta_1 \neq \beta_1^0$, as estatísticas ξ_{RV} , ξ_{SR} e ξ_W assumem formas diferentes daquelas apresentadas para o caso de ϕ ser conhecido. Em particular, denotamos por $\hat{\phi}^0$ e $\hat{\phi}$ as estimativas de máxima verossimilhança de ϕ sob H_0 e H_1 , respectivamente. Para facilitarmos a notação da estatística ξ_{RV} usamos o resultado $c(y, \phi) = d(\phi) + \phi a(y) + u(y)$ válido para algumas distribuições da família exponencial dada em (1.1) (por exemplo normal, gama e normal inversa), em que $a(\cdot)$, $d(\cdot)$ e $u(\cdot)$ são funções diferenciáveis. Assim, a estatística da razão de verossimilhanças fica expressa na forma

$$\xi_{RV} = 2\{\hat{\phi}t(\hat{\mu}) - \hat{\phi}^0t(\hat{\mu}^0)\} + 2n\{d(\hat{\phi}) - d(\hat{\phi}^0)\},$$

em que $t(\mu) = \sum_{i=1}^n \{y_i \theta_i - b(\theta_i) + a(y_i)\}$ e $\theta_i = \theta(\mu_i)$. Para o modelo gama, por exemplo, temos que $t(\mu) = \sum_{i=1}^n \{\log(y_i/\mu_i) - y_i/\mu_i\}$ e $d(\phi) = \phi \log \phi - \log \Gamma(\phi)$. A estatística de Wald fica, por sua vez, dada por

$$\begin{aligned} \xi_W &= [\hat{\beta}_1 - \beta_1^0]^T \hat{V}_{\text{ar}}^{-1}(\hat{\beta}_1) [\hat{\beta}_1 - \beta_1^0] \\ &= \hat{\phi} [\hat{\beta}_1 - \beta_1^0]^T (\hat{\mathbf{R}}^T \hat{\mathbf{W}} \hat{\mathbf{R}}) [\hat{\beta}_1 - \beta_1^0]. \end{aligned}$$

Já a estatística de escore assume a forma

$$\begin{aligned} \xi_{SR} &= \mathbf{U}_{\beta_1}(\hat{\theta}^0)^T \hat{V}_{\text{ar}_0}(\hat{\beta}_1) \mathbf{U}_{\beta_1}(\hat{\theta}^0) \\ &= \hat{\mathbf{r}}_{P_0}^T \hat{\mathbf{W}}_0^{1/2} \mathbf{X}_1 (\hat{\mathbf{R}}_0^T \hat{\mathbf{W}}_0 \hat{\mathbf{R}}_0)^{-1} \mathbf{X}_1^T \hat{\mathbf{W}}_0^{1/2} \hat{\mathbf{r}}_{P_0}, \end{aligned}$$

em que $\hat{\mathbf{r}}_{P_0} = \sqrt{\hat{\phi}^0} \hat{\mathbf{V}}_0^{-1}(\mathbf{y} - \hat{\mu}^0)$ e $\hat{\theta}^0 = (\hat{\beta}^{0T}, \hat{\phi}^0)^T$ é a estimativa de máxima verossimilhança de θ sob H_0 . As três estatísticas seguem assintoticamente e sob H_0 uma distribuição χ_q^2 .

1.7.3 Modelo de análise de variância

Vamos supor o modelo de análise de variância balanceado com um fator e dois grupos

$$g(\mu_{ij}) = \alpha + \beta_i,$$

em que $i = 1, 2$, $j = 1, \dots, m$, $\beta_1 = 0$, $\beta_2 = \beta$ e ϕ é conhecido. Vamos considerar as hipóteses $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$. Aqui \mathbf{X}_2 é um vetor $2m \times 1$ de 1's enquanto \mathbf{X}_1 é um vetor $2m \times 1$ com 0's nas m primeiras posições e 1's nas m restantes. Daí segue que $\mathbf{X}_2^T \mathbf{W} \mathbf{X}_2 = m(\omega_1 + \omega_2)$, $\mathbf{X}_2^T \mathbf{W} \mathbf{X}_1 = m\omega_2$, $C = \omega_2/(\omega_1 + \omega_2)$ e conseqüentemente

$$\mathbf{R}^T \mathbf{W} \mathbf{R} = \frac{m\omega_1\omega_2}{(\omega_1 + \omega_2)},$$

em que ω_1 e ω_2 são os pesos correspondentes aos dois grupos. A estatística de escore fica então dada por

$$\xi_{SR} = \frac{2}{m} \left(\sum_{j=1}^m \hat{r}_{P_{2j}}^0 \right)^2,$$

em que $\hat{r}_{P_{2j}}^0$, $j = 1, \dots, m$, são os resíduos estimados de Pearson, sob H_0 , correspondentes ao segundo grupo, sendo dados por $\hat{r}_{P_{2j}}^0 = \phi^{1/2}(y_{2j} - \hat{\mu}^0)/\hat{V}_0^{1/2}$. Em particular, sob a hipótese nula, $\hat{\mu}^0 = \bar{y}$. Assim, obtemos a simplificação

$$\xi_{SR} = \frac{\phi m}{2\hat{V}_0} (\bar{y}_1 - \bar{y}_2)^2, \quad (1.6)$$

em que \bar{y}_1 e \bar{y}_2 são as médias amostrais correspondentes aos dois grupos e $\hat{V}_0 = V(\bar{y})$ é a função de variância sob a hipótese nula¹.

Similarmente, podemos mostrar que a estatística de Wald fica dada por

$$\xi_W = \frac{\phi m \hat{\omega}_1 \hat{\omega}_2}{(\hat{\omega}_1 + \hat{\omega}_2)} \hat{\beta}^2, \quad (1.7)$$

¹no caso binomial tomar $\bar{y}_i = y_i/m$ e $V(\bar{y}) = \bar{y}(1 - \bar{y})$

em que $\hat{\beta}$ denota a estimativa de máxima verossimilhança de β . Mostramos na Tabela 1.4 como ficam as expressões das estatísticas ξ_{SR} e ξ_W para alguns casos da família exponencial.

Tabela 1.4
Expressões para as estatísticas de escore e de Wald.

Distribuição	ξ_{SR}	ξ_W
Normal	$\frac{m}{2\sigma^2}(\bar{y}_1 - \bar{y}_2)^2$	$\frac{m}{2\sigma^2}\hat{\beta}^2$
Poisson	$\frac{m}{2\bar{y}}(\bar{y}_1 - \bar{y}_2)^2$	$\frac{m\bar{y}_1\bar{y}_2}{(\bar{y}_1 + \bar{y}_2)}\hat{\beta}^2$
Binomial	$\frac{2m}{y(2m-y)}(y_1 - y_2)^2$	$\frac{\hat{\beta}^2}{m} \frac{y_1(m-y_1)y_2(m-y_2)}{y_1(m-y_1)+y_2(m-y_2)}$
Gama	$\frac{\phi m}{2\bar{y}^2}(\bar{y}_1 - \bar{y}_2)^2$	$\frac{\phi m(\bar{y}_1\bar{y}_2)^2}{(\bar{y}_1^2 + \bar{y}_2^2)}\hat{\beta}^2$
Normal inversa	$\frac{\phi m}{2\bar{y}^3}(\bar{y}_1 - \bar{y}_2)^2$	$\frac{\phi m(\bar{y}_1\bar{y}_2)^3}{(\bar{y}_1^3 + \bar{y}_2^3)}\hat{\beta}^2$

1.7.4 Regressão linear simples

Vamos supor agora um MLG com parte sistemática na forma linear simples

$$g(\mu_i) = \alpha + \beta x_i, \quad i = 1, \dots, n,$$

e as hipóteses $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$ com ϕ conhecido. Nesse caso obtemos $R_j = (x_j \sum_{i=1}^n \omega_i - \sum_{i=1}^n \omega_i x_i) / \sum_{i=1}^n \omega_i$ e $\mathbf{R}^T \mathbf{W} \mathbf{R} = \sum_{i=1}^n \omega_i R_i^2$. Consequentemente, $\hat{R}_{0j} = x_j - \bar{x}$ e $\hat{\mathbf{R}}_0^T \hat{\mathbf{W}}_0 \hat{\mathbf{R}}_0 = \hat{\omega}_0 \sum_{i=1}^n (x_i - \bar{x})^2$. Aqui, também obtemos $\hat{\mu}^0 = \bar{y}$.

A estatística de escore fica portanto dada por

$$\xi_{SR} = \frac{\phi}{\hat{V}_0} \frac{\{\sum_{i=1}^n x_i (y_i - \bar{y})\}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (1.8)$$

em que $\hat{V}_0 = V(\bar{y})$.

Similarmente, obtemos para a estatística de Wald

$$\xi_W = \phi \hat{\beta}^2 \sum_{i=1}^n \hat{\omega}_i \hat{R}_i^2, \quad (1.9)$$

em que $\hat{\beta}$ é a estimativa de β sob H_1 .

1.7.5 Hipóteses restritas

Podemos ter interesse, em algumas situações práticas, em testarmos hipóteses na forma de igualdades lineares, isto é, $H_0 : \mathbf{C}\beta = \mathbf{0}$ contra $H_1 : \mathbf{C}\beta \neq \mathbf{0}$, em que \mathbf{C} é uma matriz $k \times p$ de posto completo. A estimativa de máxima verossimilhança sob a hipótese alternativa coincide com a estimativa de máxima verossimilhança irrestrita $\hat{\beta}$, no entanto, obtermos a estimativa de máxima verossimilhança sob H_0 pode ser mais complexo, requerendo o uso de algum procedimento iterativo. Nyquist (1991) propõe um processo iterativo para a obtenção da estimativa de máxima verossimilhança em MLGs com parâmetros restritos na forma $\mathbf{C}\beta = \mathbf{0}$. O processo iterativo é dado abaixo

$$\beta_c^{(m+1)} = \tilde{\beta}^{(m+1)} - (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{C}^T \{ \mathbf{C} (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{C}^T \}^{-1} \mathbf{C} \tilde{\beta}^{(m+1)},$$

$m = 0, 1, \dots$, em que $\tilde{\beta}^{(m+1)}$ é (1.5) avaliado na estimativa restrita $\beta_c^{(m)}$. A matriz de variância-covariância assintótica de $\hat{\beta}_c$ é dada por

$$\text{Var}(\hat{\beta}_c) = \phi^{-1} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} [\mathbf{I}_n - \mathbf{C}^T \{ \mathbf{C} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{C}^T \}^{-1} \mathbf{C} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}].$$

Os testes estatísticos assumem formas similares aos testes do caso irrestrito. Em particular, quando ϕ é conhecido, o teste da razão de verossimilhanças fica dado por

$$\xi_{RV} = \phi \{ D(\mathbf{y}; \hat{\mu}^0) - D(\mathbf{y}; \hat{\mu}) \},$$

em que $\hat{\mu}^0$ denota aqui a estimativa de máxima verossimilhança de μ sob $H_0 : \mathbf{C}\beta = \mathbf{0}$. Já, o teste de escore, assume a forma

$$\xi_{SR} = \phi^{-1} \mathbf{U}_\beta(\hat{\beta}_c)^T (\mathbf{X}^T \hat{\mathbf{W}}_0 \mathbf{X})^{-1} \mathbf{U}_\beta(\hat{\beta}_c),$$

em que $\hat{\mathbf{W}}_0$ é aqui avaliado em $\hat{\beta}_c$. Finalmente, o teste de Wald fica dado por

$$\begin{aligned}\xi_W &= [\mathbf{C}\hat{\beta} - \mathbf{0}]^T [\hat{\text{Var}}(\mathbf{C}\hat{\beta})]^{-1} [\mathbf{C}\hat{\beta} - \mathbf{0}] \\ &= \phi \hat{\beta}^T \mathbf{C}^T [\mathbf{C}(\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{C}^T]^{-1} \mathbf{C} \hat{\beta}.\end{aligned}$$

Sob H_0 e para grandes amostras, as estatísticas ξ_{RV} , ξ_W e ξ_{SR} seguem uma distribuição χ_k^2 . A distribuição nula assintótica dos testes acima para o caso $H_0 : \mathbf{C}\beta = \mathbf{0}$ contra $H_1 : \mathbf{C}\beta \geq \mathbf{0}$, é uma mistura de distribuições do tipo qui-quadrado. Fahrmeir e Klinger (1994) discutem esse tipo de teste em MLGs (ver também Paula, 1997).

1.8 Bandas de confiança

1.8.1 Modelo normal linear

Uma banda de confiança de coeficiente $1 - \alpha$ pode ser construída para $\mu(\mathbf{z}) = \mathbf{z}^T \beta, \forall \mathbf{z} \in \mathbb{R}^p$ (ver, por exemplo, Casella e Strawderman, 1980). Temos que $\hat{\beta} - \beta \sim N_p(\mathbf{0}, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1})$. Logo, uma banda de confiança de coeficiente $1 - \alpha$ para a média $\mu(\mathbf{z}), \forall \mathbf{z} \in \mathbb{R}^p$, fica dada por

$$\mathbf{z}^T \hat{\beta} \pm \sigma \sqrt{c_\alpha} \{\mathbf{z}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{z}\}^{1/2}, \forall \mathbf{z} \in \mathbb{R}^p,$$

em que c_α é tal que $Pr\{\chi_p^2 \leq c_\alpha\} = 1 - \alpha$. É importante observarmos que \mathbf{z} é um vetor $p \times 1$ que varia livremente no \mathbb{R}^p enquanto \mathbf{X} é uma matriz fixa com os valores das variáveis explicativas.

1.8.2 Extensão para os MLGs

Uma banda assintótica de confiança de coeficiente $1 - \alpha$ pode ser também construída para $\mu(\mathbf{z}) = g^{-1}(\mathbf{z}^T \beta), \forall \mathbf{z} \in \mathbb{R}^p$ (Piegorisch e Casella, 1988) generalizando os resultados da seção anterior. Assintoticamente temos que

$\hat{\beta} - \beta \sim N_p(\mathbf{0}, \phi^{-1}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1})$. Logo, uma banda assintótica de confiança de coeficiente $1 - \alpha$ para o preditor linear $\mathbf{z}^T \beta, \forall \mathbf{z} \in \mathbb{R}^p$, fica dada por

$$\mathbf{z}^T \hat{\beta} \pm \sqrt{\phi^{-1} c_\alpha} \{\mathbf{z}^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{z}\}^{1/2}, \quad \forall \mathbf{z} \in \mathbb{R}^p,$$

em que c_α é tal que $Pr\{\chi_p^2 \leq c_\alpha\} = 1 - \alpha$. Aplicando a transformação $g^{-1}(\cdot)$ podemos, equivalentemente, encontrar uma banda assintótica de confiança de coeficiente $1 - \alpha$ para $\mu(\mathbf{z})$, dada por

$$g^{-1}[\mathbf{z}^T \hat{\beta} \pm \sqrt{\phi^{-1} c_\alpha} \{\mathbf{z}^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{z}\}^{1/2}] \quad \forall \mathbf{z} \in \mathbb{R}^p.$$

Lembramos que \mathbf{z} é um vetor $p \times 1$ que varia livremente no \mathbb{R}^p , enquanto \mathbf{X} é uma matriz fixa com os valores das variáveis explicativas. As quantidades \mathbf{W} e ϕ devem ser estimadas consistentemente.

1.9 Técnicas de diagnóstico: Modelo normal linear

1.9.1 Introdução

Uma etapa importante na análise de um ajuste de regressão é a verificação de possíveis afastamentos das suposições feitas para o modelo, especialmente para o componente aleatório e para a parte sistemática do modelo, bem como a existência de observações discrepantes com alguma interferência desproporcional ou inferencial nos resultados do ajuste. Tal etapa, conhecida como análise de diagnóstico, tem longa data, e começou com a análise de resíduos para detectar a presença de pontos aberrantes e avaliar a adequação da distribuição proposta para a variável resposta. Uma referência importante nesse tópico é o artigo de Cox e Snell (1968) em que é apresentada uma forma bastante geral de definir resíduos.

Belsley, Kuh e Welsch (1980) e Cook e Weisberg (1982) discutem a padronização de resíduos para o caso normal linear. Pregibon (1981) propõe o componente do desvio como resíduo na classe dos modelos lineares generalizados e sugere uma padronização que é comprovada por Cordeiro (1982)(ver também McCullagh, 1987) que usa as aproximações propostas por Cox e Snell (1968). McCullagh (1987) apresenta uma outra forma de padronização para o componente do desvio em que procura corrigir os efeitos de assimetria e curtose. Atkinson (1981) propõe a construção por simulação de Monte Carlo de uma banda de confiança para os resíduos da regressão normal linear, a qual denominou envelope, e que permite uma melhor comparação entre os resíduos e os percentis da distribuição normal padrão. Williams (1984,1987) discute, com base em estudos de simulação, a aproximação da forma padronizada proposta por Pregibon (1981) encontrando fortes evidências de concordância entre a distribuição empírica do componente do desvio padronizado e a distribuição normal padrão para vários MLGs. Williams (1987) também discute a construção de envelopes em MLGs. Davison e Gigli (1989) estendem a proposta de Cox e Snell (1968) e definem uma forma geral de padronização para o componente do desvio, mesmo quando a função de distribuição acumulada não é expressa em forma fechada. Fahrmeir e Tutz (1994) estendem o trabalho de McCullagh (1987) para modelos mais gerais, não pertencentes à família exponencial de distribuições. Paula (1995) apresenta uma forma padronizada para o componente do desvio em MLGs com parâmetros restritos e verifica, através de estudos de simulação, forte concordância na maioria dos modelos estudados entre a distribuição empírica do resíduo padronizado e a distribuição normal padrão, generalizando para parâmetros restritos os resultados de Williams. De Souza e Paula (2002) usam o método proposto por Davison e Gigli (1989) a fim de obterem uma forma padronizada para o

componente do desvio em modelos de regressão von Mises, os quais têm sido aplicados na análise de dados circulares. A construção de envelopes com o resíduo proposto é também discutida no trabalho. Svetliza e Paula (2003), Ortega, Paula e Bolfarine (2007), Leiva et al. (2007) e Barros, Paula e Leiva (2008) estudam a distribuição empírica do resíduo componente do desvio, respectivamente, em modelos não lineares com resposta binomial negativa e modelos para análise de dados de sobrevivência log-gama generalizados e Birnbaum-Saunders com dados censurados.

Um outro tópico importante na análise de diagnóstico é a detecção de observações influentes, isto é, pontos que exercem um peso desproporcional nas estimativas dos parâmetros do modelo. Durante a década de 70 surgiram várias propostas relacionadas com a influência das observações nas estimativas dos coeficientes do modelo normal linear. O estudo da diagonal principal da matriz de projeção $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}$ apresentada por Hoaglin e Welsch (1978), em que \mathbf{X} denota a matriz modelo, motivou a definição de pontos de alavanca que receberam esse nome por terem um peso desproporcional no próprio valor ajustado. Esses pontos em geral são remotos no subespaço gerado pelas colunas da matriz \mathbf{X} , ou seja, têm um perfil diferente dos demais pontos no que diz respeito aos valores das variáveis explicativas. Dependendo da localização, tais pontos podem exercer forte influência nas estimativas dos coeficientes da regressão. Extensões da definição de pontos de alavanca para modelos normais não lineares são dadas em St. Laurent e Cook (1992). Wei, Hu e Fung (1998) estendem a definição de pontos de alavanca para modelos bastante gerais cuja variável resposta seja contínua. Nessa generalização são incluídos outros métodos de estimação, além de máxima verossimilhança, e outros enfoques tais como enfoque bayesiano. Paula (1999) discute pontos de alavanca em modelos normais lineares com parâmetros restritos, com

extensões para os MLGs.

A deleção de pontos talvez seja a técnica mais conhecida para avaliar o impacto da retirada de uma observação particular nas estimativas da regressão. A distância de Cook(1977), originalmente desenvolvida para modelos normais lineares, foi rapidamente assimilada e estendida para diversas classes de modelos. Por exemplo, Moolgavkar, Lustbader e Venzon (1984) estendem a metodologia para regressão não linear com aplicações em estudos emparelhados, Ross (1987) discute a geometria da deleção de casos em regressão não linear, Cook, Peña e Weisberg (1988) comparam o afastamento pela verossimilhança com medidas tradicionais de deleção de pontos tais como a distância de Cook e o $DFFITS_i$, esse último proposto por Belsley, Kuh e Welsch (1980) e Paula e Peres (1988) discutem a deleção de pontos em MLGs com parâmetros restritos. Davison e Tsai (1992) e Cordeiro e Paula (1992) estendem a metodologia para modelos cuja distribuição não pertence à família exponencial de distribuições. Galea, Riquelme e Paula (2000) investigam a metodologia em modelos elípticos multivariados. Algumas referências nesse tópico são, dentre outras, os textos de Cook e Weisberg (1982), Atkinson (1985) e Chattergee e Hadi (1988) e Montgomery et al. (2012). Um problema que pode ocorrer com a deleção individual de pontos é o que se denomina *masking effect* ou seja, deixar de detectar pontos conjuntamente discrepantes. Embora esse procedimento de deleção múltipla de pontos não seja muito popular, provavelmente em virtude do custo computacional envolvido, existem vários procedimentos robustos para a detecção de pontos discrepantes, muitos dos quais com um custo computacional relativamente baixo (ver, por exemplo, Fung, 1993; Peña e Yohai, 1999). Como em geral esses procedimentos têm sido desenvolvidos para modelos lineares, abre-se uma perspectiva de pesquisas em classes mais abrangentes, tais como os MLGs.

Contudo, uma das propostas mais inovadoras na área de diagnóstico em regressão foi apresentada por Cook (1986) que propõe avaliar a influência conjunta das observações sob pequenas mudanças (perturbações) no modelo ou nos dados, ao invés da avaliação pela retirada individual ou conjunta de pontos. Essa metodologia, denominada influência local, teve uma grande receptividade entre os usuários e pesquisadores de regressão, havendo inúmeras publicações no assunto em que a metodologia é aplicada em classes particulares de modelos ou estendida para situações mais gerais.

As metodologias descritas acima serão inicialmente derivadas para o modelo normal linear clássico sendo estendidas em seguida para os MLGs. Vamos considerar, portanto, o modelo de regressão normal linear

$$y_i = \beta_1 + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \epsilon_i,$$

$i = 1, \dots, n$, em que os erros ϵ_i 's são variáveis aleatórias independentes normalmente distribuídas de média zero e variância constante σ^2 .

1.9.2 Pontos de alavanca

O resíduo para a i -ésima observação pode ser definido como uma função do tipo $r_i = r(y_i, \hat{\mu}_i)$ que procura medir a discrepância entre o valor observado e o valor ajustado da i -ésima observação. O sinal de r_i indica a direção dessa discrepância. A definição mais usual de resíduo é dada por $r_i = y_i - \hat{\mu}_i$ (resíduo ordinário), todavia há outras formas de definir resíduo que veremos mais adiante. O vetor de resíduos ordinários é definido por $\mathbf{r} = (r_1, \dots, r_n)^T$. Logo, da regressão normal linear segue que $\mathbf{r} = \mathbf{y} - \hat{\boldsymbol{\mu}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$, em que $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ é a matriz de projeção ortogonal de vetores do \mathcal{R}^n no subespaço gerado pelas colunas da matriz \mathbf{X} .

A matriz \mathbf{H} é simétrica e idempotente e é conhecida como matriz *hat*, uma vez que faz $\hat{\boldsymbol{\mu}} = \mathbf{H}\mathbf{y}$. Por ser idempotente, temos que $\text{posto}(\mathbf{H}) =$

$\text{tr}(\mathbf{H}) = \sum_{i=1}^n h_{ii} = p$. O elemento $h_{ii} = \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i$ desempenha um papel importante na construção de técnicas de diagnóstico. Adicionalmente, temos que $\frac{1}{n} \leq h_{ii} \leq \frac{1}{c}$ (ver, por exemplo, Cook e Weisberg, 1982), em que c é o número de linhas de \mathbf{X} idênticas a \mathbf{x}_i^T . O i -ésimo valor ajustado fica então dado por

$$\hat{y}_i = h_{ii} y_i + \sum_{j \neq i} h_{ji} y_j, \quad (1.10)$$

e pelo fato da matriz \mathbf{H} ser idempotente

$$\sum_{j \neq i} h_{ij}^2 = h_{ii}(1 - h_{ii}).$$

Em particular, $h_{ii} = 1$ implica em $\hat{y}_i = y_i$, todavia a recíproca não é necessariamente verdadeira. Logo, para valores altos de h_{ii} predomina na expressão (1.10) a influência de y_i sobre o correspondente valor ajustado. Assim, é muito razoável utilizar h_{ii} como uma medida da influência da i -ésima observação sobre o próprio valor ajustado. Temos também que $h_{ii} = \partial \hat{y}_i / \partial y_i$, ou seja, h_{ii} corresponde à variação em \hat{y}_i quando y_i é acrescido de um infinitésimo.

Supondo que todos os pontos exerçam a mesma influência sobre os valores ajustados, podemos esperar que h_{ii} esteja próximo de $\frac{\text{tr}(\mathbf{H})}{n} = \frac{p}{n}$. Convém então examinar, por exemplo, aqueles pontos tais que $h_{ii} \geq \frac{2p}{n}$, que são conhecidos como pontos de alavanca ou de alto *leverage* e geralmente estão localizados em regiões remotas no subespaço gerado pelas colunas da matriz \mathbf{X} . Esses pontos podem ser também informativos com relação a $\hat{\beta}$.

Uma outra maneira de entendermos h_{ii} é construindo a matriz Jacobiana de alavancas (ver, por exemplo, St. Laurent e Cook, 1993; Paula, 1999) quando a i -ésima observação é perturbada de modo que o novo valor observado seja dado por $y_i(b) = y_i + b$, em que b é uma constante real. O novo

vetor de valores ajustados fica dado por

$$\hat{\mathbf{y}}(b) = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}(b),$$

em que $\mathbf{y}(b) = (y_1, \dots, y_{i-1}, y_i + b, y_{i+1}, \dots, y_n)^T$. A matriz Jacobiana de alavancas é definida por

$$\mathbf{J}(b) = \lim_{b \rightarrow 0} \frac{1}{b} \{\hat{\mathbf{y}}(b) - \hat{\mathbf{y}}\},$$

e representa a variação no vetor de valores ajustados sob uma variação infinitesimal no i -ésimo valor observado. Podemos verificar que

$$\mathbf{J}(b) = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{f} = \mathbf{H} \mathbf{f},$$

em que \mathbf{f} é um vetor $n \times 1$ de zeros com o valor 1 na i -ésima posição. Portanto, temos que h_{ii} representa a variação no valor predito da i -ésima observação quando o valor observado é acrescido de um infinitésimo.

Para ilustrarmos como são obtidos os valores h_{ii} no R, vamos supor um modelo normal linear de variável resposta `resp`, fatores A e B e covariáveis `cov1` e `cov2` e que os resultados do ajuste são armazenadas em `fit.model`. Esse modelo pode ser ajustado de duas formas

```
fit.model = lm( resp ~ A + B + cov1 + cov2)
```

ou, alternativamente, como um MLG

```
fit.model = glm( resp ~ A + B + cov1 + cov2,
family=gaussian).
```

É claro que a primeira maneira é mais simples. Para gerarmos a matriz modelo (incluindo a constante) fazemos

```
X = model.matrix( ~ A + B + cov1 + cov2).
```

Assim, temos em \mathbf{X} a matriz modelo correspondente. O cálculo da matriz de projeção \mathbf{H} pode ser feito seguindo os passos descritos abaixo

```
H = X%*%solve(t(X)%*%X)%*%t(X).
```

Logo, podemos obter h_{ii} extraindo os elementos da diagonal principal de \mathbf{H}

```
h = diag(H).
```

Outras maneiras mais ágeis de extrairmos os elementos h'_{ii} s de uma regressão linear são através dos comandos

```
h = lm.influence(fit.model)$hat
```

```
h = hat(X,T).
```

Para construirmos um gráfico de índices para h_{ii} , a fim de detectarmos pontos de alavanca, fazemos

```
plot(h, xlab='Indice', ylab='Alavanca').
```

1.9.3 Resíduos

Dos resultados descritos na seção anterior segue que $E(\mathbf{r}) = (\mathbf{I}_n - \mathbf{H})E(\mathbf{Y}) = \mathbf{0}$ e $\text{Var}(\mathbf{r}) = \sigma^2(\mathbf{I}_n - \mathbf{H})$. Isto é, r_i tem distribuição normal de média zero e variância $\text{Var}(r_i) = \sigma^2(1 - h_{ii})$. Além disso, a covariância entre r_i e r_j , $i \neq j$, fica dada por $\text{Cov}(r_i, r_j) = -\sigma^2 h_{ij}$.

Como os r'_i s têm variâncias diferentes, devemos expressá-los em forma padronizada a fim de conseguirmos uma comparabilidade entre os mesmos. Uma definição natural seria dividirmos r_i pelo respectivo desvio padrão. Obtemos assim o resíduo studentizado

$$t_i = \frac{r_i}{s(1 - h_{ii})^{1/2}}, i = 1, \dots, n,$$

em que $s^2 = \sum_{i=1}^n r_i^2 / (n - p)$.

No entanto, como r_i não é independente de s^2 , t_i não segue uma distribuição t de Student como se poderíamos esperar. Cook e Weisberg (1982) mostram que $t_i^2 / (n - p)$ segue uma distribuição beta com parâmetros $\frac{1}{2}$ e $(n - p - 1)/2$. Logo, temos que $E(t_i) = 0$, $\text{Var}(t_i) = 1$ e $\text{Cov}(t_i, t_j) =$

$-h_{ij}/\{(1-h_{ii})(1-h_{jj})\}^{1/2}, i < j$. O problema da dependência entre r_i e s^2 pode ser contornado substituindo s^2 por $s_{(i)}^2$, o erro quadrático médio correspondente ao modelo sem a i -ésima observação. O índice (i) indica que a i -ésima observação foi excluída. Podemos mostrar que

$$\frac{(n-p)s^2}{\sigma^2} = \frac{(n-p-1)s_{(i)}^2}{\sigma^2} + \frac{r_i^2}{\sigma^2(1-h_{ii})},$$

e daí segue pelo teorema de Fisher-Cochran (ver, por exemplo, Rao, 1973, p.185) a independência entre $s_{(i)}^2$ e r_i^2 . Além disso, obtemos

$$(n-p-1)s_{(i)}^2 = \sum_{j=1}^n r_j^2 - \frac{r_i^2}{(1-h_{ii})}$$

e daí segue, após alguma álgebra, que

$$s_{(i)}^2 = s^2 \left(\frac{n-p-t_i^2}{n-p-1} \right). \quad (1.11)$$

Assim, fica fácil mostrarmos que o novo resíduo studentizado

$$t_i^* = \frac{r_i}{s_{(i)}\{1-h_{ii}\}^{1/2}}$$

segue uma distribuição t_{n-p-1} . Se ainda substituimos (1.11) na expressão acima mostramos que t_i^* é uma transformação monótona de t_i ,

$$t_i^* = t_i \left(\frac{n-p-1}{n-p-t_i^2} \right)^{1/2}.$$

O resíduo t_i pode ser calculado pela sequência de comandos

```
lms = summary(fit.model)
s = lms$sigma
r = resid(lms)
ti = r/(s*(1-h)^.5).
```

Logo, o resíduo t_i^* fica dado por

$\text{tsi} = \text{ti} * ((n-p-1)/(n-p-\text{ti}^2))^{\cdot 5}.$

Devemos substituir n e p pelos respectivos valores numéricos.

Várias quantidades do modelo linear ajustado podem ser obtidas diretamente no R através do uso de algumas funções apropriadas (ver Spector, 1994), as quais são úteis na aplicação das técnicas de diagnóstico. Resumimos na Tabela 1.5 alguns casos.

Tabela 1.5
Quantidades úteis para diagnóstico obtidas no R.

Símbolo	Descrição	Função	Elemento
\mathbf{h}	Alavanca	<code>lm.influence()</code>	hat
$\hat{\boldsymbol{\beta}}$	Coefficientes	<code>coef()</code>	
\mathbf{r}	Resíduos	<code>resid()</code>	
s	Desvio padrão amostral	<code>summary()</code>	sigma
$\mathbf{s}_{(i)}$	Desvio padrão sem observação i	<code>lm.influence()</code>	sigma
$\hat{\boldsymbol{\beta}}_{(i)}$	Coefficiente sem observação i	<code>lm.influence()</code>	coef
$(\mathbf{X}^T \mathbf{X})^{-1}$	Covariância de $\hat{\boldsymbol{\beta}}$ sem s^2	<code>summary()</code>	cov.unscaled

Como ilustração, vamos supor um ajuste com resultados no objeto `fit.model` e que o interesse é obtermos a estimativa do desvio padrão da variável resposta Y_i sem considerarmos a i -ésima observação. Aplicamos em R a função `lm.influence(fit.model)$sigma` e obtemos um vetor de dimensão n com todas as estimativas do desvio padrão de Y_i excluindo cada observação correspondente. Como veremos a seguir o resíduo t_i^* pode ser interpretado como uma estatística para avaliarmos se a i -ésima observação é aberrante.

1.9.4 Outra interpretação para t_i^*

Vamos supor que o i -ésimo ponto é suspeito de ser aberrante. Essa hipótese pode ser testada através do modelo

$$y_j = \beta_1 + \beta_2 x_{2j} + \dots + \beta_p x_{pj} + \omega_j \gamma + \epsilon_j, \quad (1.12)$$

$j = 1, \dots, n$, em que $\omega_j = 1$ para $j = i$ e $\omega_j = 0$ em caso contrário.

Usando resultados da Seção 1.4.2 podemos mostrar que, sob a hipótese $H_0 : \gamma = 0$, o acréscimo na soma de quadrados de resíduos é dado por

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \hat{\gamma}^2(1 - h_{ii}),$$

em que $\hat{\gamma} = r_i(1 - h_{ii})^{-1}$ e $r_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$. Assim, uma vez que $D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) = (n - p)s^2$, a estatística F para testarmos $H_0 : \gamma = 0$ contra $H_1 : \gamma \neq 0$ fica dada por

$$F = \frac{\hat{\gamma}^2(1 - h_{ii})}{\left\{ (n - p)s^2 - \frac{r_i^2}{(1 - h_{ii})} \right\} / (n - p - 1)}.$$

Assintoticamente a estatística F segue, sob H_0 , distribuição qui-quadrado com 1 grau de liberdade. Trabalhando um pouco a expressão acima chegamos ao seguinte resultado:

$$F = \frac{r_i^2(n - p - 1)}{s^2(1 - h_{ii})(n - p - t_i^2)} = t_i^{*2}.$$

Portanto, valores altos para $|t_i^*|$ indicam, significativamente, que o i -ésimo ponto é aberrante.

1.9.5 Influência

Vamos supor ϕ conhecido. Então, o logaritmo da função de verossimilhança fica agora expresso na forma

$$L_\delta(\boldsymbol{\beta}) = \sum_{j=1}^n \delta_j L_j(\boldsymbol{\beta}), \quad (1.13)$$

em que $L_j(\boldsymbol{\beta})$ denota o logaritmo da função de verossimilhança correspondente à j -ésima observação e δ_j é um tipo de perturbação, definida tal que $0 \leq \delta_j \leq 1$. Quando $\delta_j = 1, \forall j$, significa que não há perturbação no modelo e quando $\delta_j = 0$ significa que a j -ésima observação foi excluída.

A estimativa de mínimos quadrados para $\boldsymbol{\beta}$ fica, supondo a estrutura (1.13), dada por

$$\hat{\boldsymbol{\beta}}_\delta = (\mathbf{X}^T \boldsymbol{\Delta} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Delta} \mathbf{y},$$

em que $\boldsymbol{\Delta} = \text{diag}\{\delta_1, \dots, \delta_n\}$. Em particular, quando apenas a i -ésima observação é perturbada, isto é, quando $\delta_i = \delta$ e $\delta_j = 1$ para $j \neq i$, obtemos

$$\hat{\boldsymbol{\beta}}_\delta = \hat{\boldsymbol{\beta}} - \frac{(1 - \delta)r_i}{\{1 - (1 - \delta)h_{ii}\}} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i. \quad (1.14)$$

Para $\delta = 0$, significando que o i -ésimo ponto foi excluído, (1.14) fica expressa na forma simplificada

$$\hat{\boldsymbol{\beta}}_{(i)} = \hat{\boldsymbol{\beta}} - \frac{r_i}{(1 - h_{ii})} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i, \quad (1.15)$$

que é bastante conhecida da regressão normal linear (ver, por exemplo, Cook e Weisberg, 1982).

A medida de influência mais conhecida é baseada na região de confiança de coeficiente $(1 - \alpha)$ para o parâmetro $\boldsymbol{\beta}$,

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (\mathbf{X}^T \mathbf{X}) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq ps^2 F_{p, (n-p)}(1 - \alpha),$$

que para o caso de $p = 2$ é um elipsóide no \mathcal{R}^2 centrado em $\hat{\boldsymbol{\beta}}$. Tal medida, conhecida como distância de Cook, é definida por

$$D_\delta = \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_\delta)^T (\mathbf{X}^T \mathbf{X}) (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_\delta)}{ps^2}, \quad (1.16)$$

e mede quanto a perturbação $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^T$ afasta $\hat{\boldsymbol{\beta}}_\delta$ de $\hat{\boldsymbol{\beta}}$, segundo a métrica $\mathbf{M} = \mathbf{X}^T \mathbf{X}$. Por exemplo, se $D_\delta > F_{p, (n-p)}(1 - \alpha)$, significa que

a perturbação está deslocando o contorno do elipsóide para um contorno correspondente a um nível de significância menor do que α .

Em particular, quando o i -ésimo ponto é excluído, a distância de Cook fica expressa na forma

$$\begin{aligned} D_i &= \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T (\mathbf{X}^T \mathbf{X}) (\hat{\beta} - \hat{\beta}_{(i)})}{ps^2} \\ &= \left\{ \frac{r_i}{s(1 - h_{ii})^{1/2}} \right\}^2 \frac{h_{ii}}{(1 - h_{ii})} \frac{1}{p} \\ &= t_i^2 \frac{h_{ii}}{(1 - h_{ii})} \frac{1}{p}. \end{aligned}$$

Portanto, D_i será grande quando o i -ésimo ponto for aberrante (t_i grande) e/ou quando h_{ii} for próximo de um. A distância de Cook pode ser calculada da seguinte maneira:

$$di = (ti^2) * h / (p * (1 - h)).$$

A distância D_i poderá não ser adequada quando r_i for grande e h_{ii} for pequeno. Nesse caso, s^2 pode ficar inflacionado e não ocorrendo nenhuma compensação por parte de h_{ii} , D_i pode ficar pequeno. Uma medida supostamente mais apropriada foi proposta por Belsley, Kuh e Welsch (1980), sendo definida por

$$\begin{aligned} DFFITS_i &= \frac{|r_i|}{s_{(i)}(1 - h_{ii})^{1/2}} \left\{ \frac{h_{ii}}{(1 - h_{ii})} \right\}^{1/2} \\ &= |t_i^*| \left\{ \frac{h_{ii}}{(1 - h_{ii})} \right\}^{1/2}. \end{aligned}$$

O $DFFITS_i$ é calculado conforme abaixo

$$dfit = abs(tsi) * (h / (1 - h))^{.5}.$$

Como o valor esperado de h_{ii} é $\frac{p}{n}$, é razoável darmos mais atenção àqueles pontos tais que

$$DFFITS_i \geq 2 \left\{ \frac{p}{(n - p)} \right\}^{1/2}.$$

Aparentemente D_i e $DFFITS_i$ seriam medidas de influência competitivas, uma vez que $DFFITS_i$ parece ser mais adequada para avaliar a influência nas estimativas dos coeficientes de um ponto aberrante com h_{ii} pequeno. No entanto, como mostram Cook, Peña e Weisberg (1988) D_i e $DFFITS_i$ medem coisas diferentes. Ambas podem ser expressas a partir da medida mais geral de influência denominada afastamento pela verossimilhança (*likelihood displacement*) proposta por Cook e Weisberg (1982). A medida D_i mede essencialmente a influência das observações nos parâmetros de posição, enquanto $DFFITS_i$ tem o propósito de medir a influência das observações nos parâmetros de posição e escala. Como é pouco provável que um ponto com r_i alto e h_{ii} pequeno seja influente nas estimativas dos coeficientes, o uso de D_i não compromete a detecção de observações influentes. Cook, Peña e Weisberg observam também que $DFFITS_i$ não é uma medida completa de influência nos parâmetros de posição e escala simultaneamente, podendo falhar em algumas situações. Uma medida mais geral nesse caso é proposta pelos autores.

Atkinson (1985) propôs uma outra medida de influência que é um aperfeiçoamento do $DFFITS_i$,

$$A_i = \left\{ \frac{(n-p)}{p} \frac{h_{ii}}{(1-h_{ii})} \right\}^{1/2} |t_i^*|.$$

Aqui, quando o experimento for balanceado, isto é, todos os h'_{ii} s forem iguais, obtemos $A_i = |t_i^*|$. A vantagem de A_i é que a mesma pode ser utilizada em gráficos normais de probabilidades.

1.9.6 Ilustração

As Figuras 1.4a-1.4d ilustram as diferenças entre pontos aberrantes, de alavanca e influentes. Na Figura 1.4a temos os pontos alinhados sem nenhum

tipo de perturbação. Na Figura 1.4b perturbamos o ponto #3 fazendo-o aberrante. Note que a exclusão do mesmo (reta pontilhada) altera apenas o intercepto, isto é, os valores ajustados. É um ponto que não está muito afastado dos demais, logo tem um valor para h_{ii} relativamente pequeno. Já na Figura 1.4c, perturbamos o ponto #5 de modo que o mesmo fique mais afastado no subespaço gerado pelas colunas da matriz \mathbf{X} . É um ponto de alavanca, todavia a eliminação do mesmo não muda praticamente nada nas estimativas dos parâmetros. Como é um ponto com h_{ii} relativamente alto, as variâncias dos valores ajustados dos pontos próximos ao mesmo serão maiores do que as variâncias dos valores ajustados correspondentes aos demais pontos. Finalmente, na Figura 1.4d, perturbamos novamente o ponto #5 fazendo-o agora influente e também alavanca. O mesmo, além de mudar a estimativa da inclinação da reta ajustada, continua mais afastado do que os demais.

As possíveis situações discutidas acima, quando detectadas num ajuste de regressão, devem ser examinadas cuidadosamente antes de qualquer decisão. Encontrarmos razões que expliquem o fato dos pontos terem um comportamento atípico com relação aos demais pontos podem ajudar-nos a entendermos melhor a relação entre as variáveis explicativas e o fenômeno sob investigação como também a traçarmos uma estratégia de utilização do modelo ajustado, que não necessariamente implica na eliminação de tais pontos que deve ser o último recurso a ser utilizado.

Mudanças na distribuição postulada para a variável resposta, inclusão ou retirada de variáveis explicativas ou mesmo transformações de variáveis explicativas podem atenuar a influência de observações atípicas. Outra alternativa é a aplicação de procedimentos robustos (vide, por exemplo, Venables e Ripley, 1999, Cap.8; Montgomery et al., 2012, Cap.15). Uma estratégia

interessante é diferenciar mudanças numéricas de mudanças inferenciais nas estimativas quando as observações atípicas não são consideradas no ajuste.

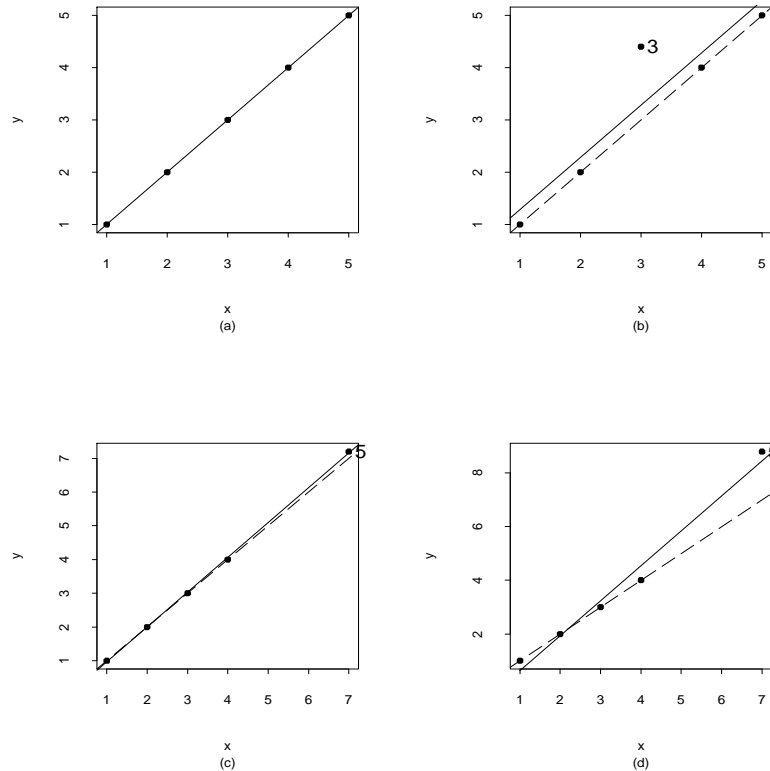


Figura 1.4: Ilustração de pontos aberrantes, influentes e de alavanca.

1.9.7 Influência local

Um dos métodos mais modernos de diagnóstico foi proposto por Cook (1986). A ideia básica consiste em estudar o comportamento de alguma medida particular de influência segundo pequenas perturbações (**influência local**) nos dados ou no modelo. Isto é, verificar a existência de pontos que sob

modificações modestas no modelo causam variações desproporcionais nos resultados.

Podemos, por exemplo, querer avaliar a influência que pequenas mudanças nas variâncias das observações causam nas estimativas dos parâmetros. Nesse caso, podemos utilizar a distância de Cook como medida de referência. Por outro lado, se o interesse é estudarmos a influência local das observações no ajuste, a sugestão de Cook é perturbarmos as covariáveis ou a variável resposta e utilizarmos alguma medida adequada para quantificarmos a influência das observações. Como ilustração, vamos supor que uma variável explicativa que representa uma distância particular é perturbada localmente e detectamos através de uma medida de influência que pontos com distâncias altas produzem variações acentuadas na medida adotada. Isso sugere que a variável explicativa sob estudo é bastante sensível para valores altos, podendo não ser uma boa preditora nesses casos.

Mais de 400 artigos foram publicados no assunto nos últimos 20 anos. Por exemplo, na classe de erros normais, Lawrence (1988) investiga a aplicação de influência local em modelos lineares com parâmetros na transformação da resposta, Beckman, Nachtsheim e Cook (1987) apresentam estudos de influência em modelos de análise de variância com efeito misto, Tsai e Wu (1992) investigam influência local em modelos autoregressivos de 1a. ordem e modelos heteroscedásticos e Paula (1993) aplica influência local em modelos lineares com restrições nos parâmetros na forma de desigualdades lineares. Saindo da classe de erros normais temos, por exemplo, o trabalho de Pettitt e Bin Daud (1989) que investigam influência local em modelos de Cox com riscos proporcionais, Escobar e Meeker (1992) adaptam influência local numa classe paramétrica de modelos para análise de sobrevivência, O'Hara Hines, Lawless e Carter (1992), Kim (1995) e Pan, Fang e von Rosen (1997) aplicam

métodos de influência local em regressão multivariada. Mais recentemente, Galea, Paula e Bolfarine (1997), Liu (2000), Galea, Paula e Uribe-Opazo (2003), Osorio, Paula e Galea (2007) e Russo, Paula e Aoki (2009) apresentam estudos de influência local em modelos de contornos elípticos, enquanto Kwan e Fung (1998) aplicam a metodologia em análise fatorial, Gu e Fung (1998) em análise de correlação canônica, Paula (1996) em modelos próprios de dispersão e Ortega, Bolfarine e Paula (2003) em modelos log-gama generalizados com dados censurados. Svetliza e Paula (2003) discutem influência local em modelos com resposta binomial negativa. Esses últimos modelos têm sido muito usados para corrigir problemas de sobredispersão, frequentemente encontrados em modelos com resposta de Poisson. Galea, Leiva e Paula (2004), Leiva et al. (2007) e Barros, Paula e Leiva (2008) aplicam a metodologia de influência local em modelos Birnbaum-Saunders comumente utilizados em confiabilidade e análise de dados de sobrevivência. Uma discussão interessante a respeito do uso de influência local é apresentada por Fung e Kwan (1997). Os autores mostram que o afastamento pela verossimilhança é uma medida de influência invariante com mudanças de escala nos dados, fato que não ocorre com outras medidas de influência propostas.

Curvatura normal

Para formalizar o método de influência local vamos denotar por $L(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança do modelo postulado e $\boldsymbol{\theta}$ um vetor r -dimensional. No caso de MLGs podemos ter $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ e $r = p + 1$ ou simplesmente $\boldsymbol{\theta} = \boldsymbol{\beta}$ quando ϕ for conhecido. Seja $\boldsymbol{\delta}$ um vetor $q \times 1$ de perturbações, restritas a um conjunto aberto $\boldsymbol{\Omega} \subset \mathbb{R}^q$. Em geral temos $q = n$. As perturbações são feitas no logaritmo da verossimilhança de modo que o mesmo assume a forma $L(\boldsymbol{\theta}|\boldsymbol{\delta})$. Denotando o vetor de não perturbação

por $\boldsymbol{\delta}_0$, assumimos que $L(\boldsymbol{\theta}|\boldsymbol{\delta}_0) = L(\boldsymbol{\theta})$. A fim de verificarmos a influência das perturbações na estimativa de máxima verossimilhança $\hat{\boldsymbol{\theta}}$, consideramos o afastamento pela verossimilhança

$$LD(\boldsymbol{\delta}) = 2\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_\delta)\},$$

em que $\hat{\boldsymbol{\theta}}_\delta$ denota a estimativa de máxima verossimilhança sob o modelo $L(\boldsymbol{\theta}|\boldsymbol{\delta})$. Com a definição acima temos que $LD(\boldsymbol{\delta}) \geq 0$.

A ideia de influência local consiste basicamente em estudarmos o comportamento da função $LD(\boldsymbol{\delta})$ em torno de $\boldsymbol{\delta}_0$. O procedimento procura selecionar uma direção unitária $\boldsymbol{\ell}$, $\|\boldsymbol{\ell}\| = 1$, e então estudar o gráfico de $LD(\boldsymbol{\delta}_0 + a\boldsymbol{\ell})$ contra a , em que $a \in \mathbb{R}$. Esse gráfico é conhecido como linha projetada. Em particular, temos que $LD(\boldsymbol{\delta}_0) = 0$, assim $LD(\boldsymbol{\delta}_0 + a\boldsymbol{\ell})$ tem um mínimo local em $a = 0$. Cada linha projetada pode ser caracterizada por uma curvatura normal $C_\ell(\boldsymbol{\theta})$ em torno de $a = 0$. Essa curvatura é interpretada como sendo o inverso do raio do melhor círculo ajustado em $a = 0$. Uma sugestão é considerarmos a direção $\boldsymbol{\ell}_{\max}$ que corresponde à maior curvatura denotada por $C_{\ell_{\max}}$. Por exemplo, o gráfico de $\boldsymbol{\ell}_{\max}$ contra a ordem das observações pode revelar quais observações que sob pequenas perturbações exercem uma influência desproporcional em $LD(\boldsymbol{\delta})$. Cook(1986) usa conceitos de geometria diferencial para mostrar que a curvatura normal na direção $\boldsymbol{\ell}$ assume a forma

$$C_\ell(\boldsymbol{\theta}) = 2|\boldsymbol{\ell}^T \boldsymbol{\Delta}^T \ddot{\mathbf{L}}_{\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}}^{-1} \boldsymbol{\Delta} \boldsymbol{\ell}|,$$

em que $-\ddot{\mathbf{L}}_{\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}}$ é a matriz de informação observada de Fisher enquanto $\boldsymbol{\Delta}$ é uma matriz $r \times q$ com elementos $\Delta_{ij} = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\delta}) / \partial \theta_i \partial \delta_j$, avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\delta} = \boldsymbol{\delta}_0$, $i = 1, \dots, r$ e $j = 1, \dots, q$.

Temos que o máximo de $\boldsymbol{\ell}^T \mathbf{B} \boldsymbol{\ell}$, em que $\mathbf{B} = \boldsymbol{\Delta}^T (-\ddot{\mathbf{L}}_{\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}})^{-1} \boldsymbol{\Delta}$, corresponde ao maior autovalor (em valor absoluto) de \mathbf{B} . Portanto, $C_{\ell_{\max}}$ corresponde ao maior autovalor da matriz \mathbf{B} e $\boldsymbol{\ell}_{\max}$ denota o autovetor correspondente.

Assim, o gráfico de $|\ell_{\max}|$ contra a ordem das observações pode revelar aqueles pontos com maior influência na vizinhança de $\text{LD}(\delta_0)$. Tais pontos podem ser responsáveis por mudanças substanciais nas estimativas dos parâmetros sob pequenas perturbações no modelo ou nos dados. Seria, portanto, prudente olharmos com mais cuidado esses pontos a fim de entendermos melhor a influência dos mesmos e consequentemente tentarmos propor uma forma segura de usarmos o modelo ajustado. Quando $C_{\ell_{\max}}$ não for muito maior do que o segundo autovalor, pode ser informativo olharmos também os componentes do segundo autovetor. É provável, nesse caso, que o segundo autovetor destaque algum tipo de influência particular das observações nas estimativas. O maior autovalor da matriz \mathbf{B} pode ser obtido pelo comando abaixo

```
Cmax = eigen(B)$val[1].
```

De forma similar, o autovetor correspondente padronizado e em valor absoluto é obtido com os comandos

```
lmax = eigen(B)$vec[,1]
```

```
lmax = abs(lmax).
```

Gráficos alternativos, tais como de C_{ℓ_i} contra a ordem das observações, em que ℓ_i denota um vetor $n \times 1$ de zeros com um na i -ésima posição têm sido sugeridos (ver, por exemplo, Lesaffre e Verbeke, 1998; Zhou e Zhang, 2004). Nesse caso devemos padronizar $C_i = C_i / \sum_{j=1}^n C_j$. Uma sugestão é olharmos com mais atenção aqueles pontos tais que $C_i > 2\bar{C}$, em que $\bar{C} = \sum_{j=1}^n C_j / n$. Por outro lado, se o interesse está num subvetor θ_1 de $\theta = (\theta_1^T, \theta_2^T)^T$, então a curvatura normal na direção ℓ fica dada por $C_\ell(\theta_1) = 2|\ell^T \Delta^T (\ddot{\mathbf{L}}_{\theta\theta}^{-1} - \mathbf{B}_1) \Delta \ell|$, sendo

$$\mathbf{B}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{L}}_{\theta_2\theta_2}^{-1} \end{pmatrix},$$

com $-\ddot{\mathbf{L}}_{\theta_2\theta_2}$ denotando a informação de Fisher observada para θ_2 . O gráfico

do maior autovetor de $\Delta^T(\ddot{\mathbf{L}}_{\hat{\theta}\hat{\theta}}^{-1} - \mathbf{B}_1)\Delta$ contra a ordem das observações pode revelar os pontos com maior influência local em $\hat{\theta}_1$. Outras formas de extração de gráficos são propostas por Poon e Poon (1999).

Ponderação de casos

Para ilustrar uma aplicação particular consideramos o modelo normal linear com σ^2 conhecido e esquema de perturbação ponderação de casos, em que

$$L(\beta|\delta) = -\frac{1}{2\sigma^2} \sum_{i=1}^n \delta_i (y_i - \mathbf{x}_i^T \beta)^2$$

com $0 \leq \delta_i \leq 1$. A matriz Δ nesse caso fica dada por $\mathbf{X}^T \mathbf{D}(\mathbf{r})/\sigma^2$ em que $\mathbf{D}(\mathbf{r}) = \text{diag}\{r_1, \dots, r_n\}$ com $r_i = y_i - \hat{y}_i$. Logo, desde que $\ddot{\mathbf{L}}_{\beta\beta} = -\sigma^{-2}(\mathbf{X}^T \mathbf{X})$ a curvatura normal na direção unitária ℓ fica dada por

$$C_\ell(\beta) = \frac{2}{\sigma^2} |\ell^T \mathbf{D}(\mathbf{r}) \mathbf{H} \mathbf{D}(\mathbf{r}) \ell|,$$

com $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$. Portanto, ℓ_{\max} é o autovetor correspondente ao maior autovalor (em valor absoluto) da matriz $\mathbf{B} = \mathbf{D}(\mathbf{r}) \mathbf{H} \mathbf{D}(\mathbf{r})$. Se tomarmos a direção ℓ_i correspondente à i -ésima observação, a curvatura normal assume a forma simplificada $C_i = \frac{2}{\sigma^2} h_{ii} r_i^2$. Os gráficos de índices de ℓ_{\max} e C_i podem revelar aquelas observações mais sensíveis ao esquema de perturbação adotado.

Cálculos similares para σ^2 desconhecido levam ao seguinte $\Delta = (\Delta_1^T, \Delta_2^T)^T$ em que $\Delta_1 = \mathbf{X}^T \mathbf{D}(\mathbf{r})/\hat{\sigma}^2$ e $\Delta_2 = \mathbf{r}^{(2)T}/2\hat{\sigma}^4$ com $\mathbf{r}^{(2)T} = (r_1^2, \dots, r_n^2)$ e $-\ddot{\mathbf{L}}_{\theta\theta} = \text{diag}\{\mathbf{X}^T \mathbf{X}/\hat{\sigma}^2, n/2\hat{\sigma}^4\}$. Logo, a curvatura normal na direção unitária ℓ fica dada por

$$C_\ell(\theta) = \frac{2}{\hat{\sigma}^2} |\ell^T \{\mathbf{D}(\mathbf{r}) \mathbf{H} \mathbf{D}(\mathbf{r}) + \mathbf{r}^{(2)} \mathbf{r}^{(2)T}/2n\hat{\sigma}^2\} \ell|.$$

Quando o interesse é verificarmos a influência local das observações na estimativa de um coeficiente particular β_1 devemos considerar a curvatura

normal $C_\ell(\beta_1) = 2|\ell^T \mathbf{B} \ell|$, em que

$$\mathbf{B} = \mathbf{D}(\mathbf{r})\mathbf{X}\{(\mathbf{X}^T\mathbf{X})^{-1} - \mathbf{B}_1\}\mathbf{X}^T\mathbf{D}(\mathbf{r})$$

sendo $\mathbf{B}_1 = \text{diag}\{\mathbf{0}, (\mathbf{X}_2^T\mathbf{X}_2)^{-1}\}$ com \mathbf{X}_2 saindo da partição $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$. Aqui \mathbf{X}_1 é um vetor $n \times 1$ correspondente à variável explicativa sob estudo e \mathbf{X}_2 é uma matriz $n \times (p-1)$ correspondente às demais variáveis explicativas. Cook (1986) mostra que ℓ_{\max} , nesse caso, assume a forma

$$\ell_{\max}^T = \left(\frac{v_1 r_1}{\sqrt{C_{\ell_{\max}}}}, \dots, \frac{v_n r_n}{\sqrt{C_{\ell_{\max}}}} \right),$$

em que v_1, \dots, v_n são os resíduos ordinários da regressão linear de \mathbf{X}_1 sobre as colunas de \mathbf{X}_2 , ou seja, o vetor $\mathbf{v} = (v_1, \dots, v_n)^T$ é dado por $\mathbf{v} = (\mathbf{I}_n - \mathbf{H}_2)\mathbf{X}_1$, $\mathbf{H}_2 = \mathbf{X}_2(\mathbf{X}_2^T\mathbf{X}_2)^{-1}\mathbf{X}_2^T$. Aqui, a matriz \mathbf{B} tem posto $m = 1$. Logo, há apenas um autovalor diferente de zero. Nesse caso, podemos tanto utilizar o procedimento descrito acima para calcularmos ℓ_{\max} como obtê-lo diretamente sem precisar calcularmos a matriz \mathbf{H}_2 . Como ilustração, vamos supor que os resultados do ajuste estão armazenados em `fit.model`. Para extrairmos o vetor \mathbf{r} precisamos fazer

```
r = resid(fit.model).
```

Se o modelo tem as covariáveis `cov1` e `cov2` além dos fatores `A` e `B`, o vetor ℓ_{\max} correspondente, por exemplo à covariável `cov1`, sai de

```
fit = lm( cov1 ~ A + B + cov2 - 1)
v = resid(fit)
lmax = v*r
tot = t(lmax)%*%lmax
lmax = lmax/sqrt(tot)
lmax = abs(lmax).
```

1.9.8 Gráfico da variável adicionada

Vamos supor novamente o modelo de regressão dado em (1.12), em que ω é agora uma variável adicional qualquer. Definindo $\mathbf{Z} = (\mathbf{X}, \omega)$, podemos mostrar facilmente que a estimativa de mínimos quadrados de $\theta = (\beta^T, \gamma)^T$ é dada por $\hat{\theta} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y}$. Em particular obtemos, após alguma álgebra, que

$$\hat{\gamma} = \frac{\omega^T (\mathbf{I}_n - \mathbf{H}) \mathbf{y}}{\omega^T (\mathbf{I}_n - \mathbf{H}) \omega} = \frac{\omega^T \mathbf{r}}{\omega^T (\mathbf{I}_n - \mathbf{H}) \omega}.$$

Isto é, $\hat{\gamma}$ é o coeficiente da regressão linear passando pela origem do vetor de resíduos $\mathbf{r} = (\mathbf{I}_n - \mathbf{H}) \mathbf{y}$ sobre o novo resíduo $\mathbf{v} = (\mathbf{I}_n - \mathbf{H}) \omega$, dado por

$$\begin{aligned} \hat{\gamma} &= (\mathbf{v}^T \mathbf{v})^{-1} \mathbf{v}^T \mathbf{r} \\ &= \{\omega^T (\mathbf{I}_n - \mathbf{H}) (\mathbf{I}_n - \mathbf{H}) \omega\}^{-1} \omega^T (\mathbf{I}_n - \mathbf{H}) (\mathbf{I}_n - \mathbf{H}) \mathbf{y} \\ &= \frac{\omega^T (\mathbf{I}_n - \mathbf{H}) \mathbf{y}}{\omega^T (\mathbf{I}_n - \mathbf{H}) \omega}. \end{aligned}$$

Portanto, um gráfico de \mathbf{r} contra \mathbf{v} pode fornecer informações sobre a evidência dessa regressão, indicando quais observações que estão contribuindo para a relação e quais observações que estão se desviando da mesma. Esse gráfico, conhecido como gráfico da variável adicionada, pode revelar quais pontos que estão influenciando (e de que maneira) a inclusão da nova variável no modelo.

Para ilustrarmos a construção do gráfico da variável adicionada, vamos supor novamente o modelo com duas covariáveis e dois fatores. O gráfico da variável adicionada para avaliarmos a influência das observações no coeficiente de `cov1`, pode ser construído com os comandos

```
fit = lm( resp ~ cov2 + A + B)
r = resid(fit)
fit1 = lm( cov1 ~ cov2 + A + B)
v = resid(fit1)
```

```
plot(v,r, xlab= ‘‘residuo v’’, ylab= ‘‘residuo r’’).
```

1.9.9 Técnicas gráficas

Geralmente para detectarmos pontos suspeitos de serem aberrantes e/ou influentes recorreremos aos seguintes gráficos: (i) pontos aberrantes, o gráfico de t_i^* contra a ordem das observações; (ii) heteroscedasticidade, o gráfico de t_i^* contra \hat{y}_i (valores ajustados); (iii) pontos influentes, gráficos de D_i , C_i , $DFFITS_i$, h_{ii} ou $|\ell_{\max}|$ contra a ordem das observações; (iv) falta de algum termo extra, gráfico de t_i^* contra \hat{y}_i ou contra as covariáveis que estão ou não foram incluídas no modelo, (v) correlação entre as observações, gráfico de t_i^* contra o tempo ou contra a ordem em que há suspeita de correlação, e (vi) afastamentos da normalidade, gráfico normal de probabilidades. Esse último é o gráfico de $t_{(i)}^*$ contra os valores esperados das estatísticas de ordem da normal padrão, $Z'_{(i)}$ s. Temos que

$$E(Z_{(i)}) \cong \Phi^{-1} \left(\frac{i - 3/8}{n + 1/4} \right),$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da $N(0, 1)$. Há também o gráfico meio-normal de probabilidades, definido como sendo o gráfico de $|t_{(i)}^*|$ contra os valores esperados de $|Z_{(i)}|$. Temos a aproximação

$$E(|Z_{(i)}|) \cong \Phi^{-1} \left(\frac{n + i + 1/2}{2n + 9/8} \right).$$

Em particular, o gráfico de A_i contra $E(|Z_{(i)}|)$ pode ser indicado para detectarmos simultaneamente pontos aberrantes e/ou influentes. O gráfico normal de probabilidades com a reta ajustada pode ser construído com os comandos dados abaixo

```
qqnorm(tsi , ylab= ‘‘Residuo Studentizado’’)
```

`qqline(tsi)`.

O comando `qqline()` traça uma reta unindo os pontos formados pelo primeiro e terceiro quartis dos resíduos e da distribuição normal padrão. Devido à dificuldade de avaliarmos se o gráfico normal de probabilidades se afasta efetivamente da reta ajustada, a construção de um tipo de banda de confiança para os resíduos pode ser muito útil para detectarmos afastamentos sérios da normalidade. Esse gráfico pode também ser informativo sobre a existência de pontos discrepantes ou mesmo sobre a falta de homogeneidade de variâncias. Todavia, como a distribuição conjunta dos resíduos $t_{(i)}^*$'s é bastante complicada e o uso simples das variâncias dos t_i^* 's para a construção de tais bandas pode introduzir algum viés no cálculo do coeficiente de confiança, Atkinson (1985) sugere a construção de um tipo de banda de confiança através de simulações, a qual denominou **envelope**. O procedimento consiste basicamente em gerarmos resíduos que tenham média zero e matriz de variância-covariância $(\mathbf{I}_n - \mathbf{H})$. Descrevemos o método nos passos seguintes:

1. Geramos n observações $N(0, 1)$ as quais são armazenadas em $\mathbf{y} = (y_1, \dots, y_n)^T$;
2. Ajustamos \mathbf{y} contra \mathbf{X} e obtemos $r_i = y_i - \hat{y}_i$, $i = 1, \dots, n$. Temos que $E(r_i) = 0$, $\text{Var}(r_i) = 1 - h_{ii}$ e $\text{Cov}(r_i, r_j) = -h_{ij}$;
3. Obtemos $t_i^* = r_i / \{1 - h_{ii}\}^{1/2}$, $i = 1, \dots, n$;
4. Repetimos os passos (1)-(3) m vezes. Logo, teremos os resíduos gerados t_{ij}^* , $i = 1, \dots, n$ e $j = 1, \dots, m$.
5. Colocamos cada grupo de n resíduos em ordem crescente, obtendo $t_{(i)j}^*$, $i = 1, \dots, n$ e $j = 1, \dots, m$;

6. Obtemos os limites $t_{(i)I}^* = \min_j t_{(i)j}$ e $t_{(i)S}^* = \max_j t_{(i)j}^*$. Assim, os limites correspondentes ao i -ésimo resíduo serão dados por $t_{(i)I}^*$ e $t_{(i)S}^*$.

A sugestão de Atkinson (1985) é gerar $m = 19$ vezes. Desse modo, a probabilidade do maior resíduo de um envelope particular exceder o limite superior fica sendo $\cong 1/20$. Adaptamos um programa descrito em Everitt (1994) para geração dos envelopes de um modelo de regressão normal linear considerando $m = 100$. Para rodarmos o programa é preciso apenas colocarmos o modelo ajustado em `fit.model`. Daí, devemos fazer

```
source('envel_norm')
```

em que `envel_norm` é o nome do arquivo externo onde deve estar o programa para geração dos envelopes (ver Apêndice B).

1.10 Técnicas de diagnóstico: Extensão para os MLGs

1.10.1 Pontos de alavanca

Como já foi mencionado na Seção 1.9.2 a ideia principal que está por trás do conceito de ponto de alavanca (ver, por exemplo, Hoaglin e Welsch, 1978; Cook e Weisberg, 1982; Emerson, Hoaglin e Kempthorne, 1984; St. Laurent e Cook, 1992 e Wei, Hu e Fung, 1998) é de avaliarmos a influência de y_i sobre o próprio valor ajustado \hat{y}_i . Essa influência pode ser bem representada pela derivada $\partial \hat{y}_i / \partial y_i$ que coincide, como foi visto na Seção 1.9.2, com h_{ii} no caso normal linear. Wei, Hu e Fung (1998) propuseram uma forma bastante geral para obtenção da matrix $(\partial \hat{\mathbf{y}} / \partial \mathbf{y}^T)_{n \times n}$ quando a resposta é contínua e que pode ser aplicada em diversas situações de estimação. No caso de MLGs para ϕ conhecido a matriz $\partial \hat{\mathbf{y}} / \partial \mathbf{y}^T$ pode ser obtida da forma geral

$$\widehat{\text{GL}} = \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}^T} = \{\mathbf{D}_\beta (-\ddot{\mathbf{L}}_{\beta\beta})^{-1} \ddot{\mathbf{L}}_{\beta y}\}_{|\hat{\beta}},$$

em que $\mathbf{D}_\beta = \partial \boldsymbol{\mu} / \partial \boldsymbol{\beta}$, $\ddot{\mathbf{L}}_{\beta\beta} = \partial^2 \mathbf{L}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T$ e $\ddot{\mathbf{L}}_{\beta y} = \partial^2 \mathbf{L}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} \partial \mathbf{y}^T$. Temos que

$$\mathbf{D}_\beta = \mathbf{N}\mathbf{X} \quad \text{e} \quad \ddot{\mathbf{L}}_{\beta y} = \phi \mathbf{X}^T \mathbf{V}^{-1} \mathbf{N},$$

em que $\mathbf{N} = \text{diag}\{d\mu_1/d\eta_1, \dots, d\mu_n/d\eta_n\}$. Se substituirmos $-\ddot{\mathbf{L}}_{\beta\beta}$ pelo seu valor esperado $\phi(\mathbf{X}^T \mathbf{W} \mathbf{X})$, obtemos aproximadamente

$$\widehat{\mathbf{GL}} = \hat{\mathbf{N}} \mathbf{X} (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \hat{\mathbf{N}}.$$

Assim, o elemento $\widehat{\mathbf{GL}}_{ii}$ pode ser expresso na forma

$$\widehat{\mathbf{GL}}_{ii} = \hat{\omega}_i \mathbf{x}_i^T (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i,$$

em que $\omega_i = (d\mu_i/d\eta_i)^2/V_i$. Em particular, para ligação canônica em que $-\ddot{\mathbf{L}}_{\beta\beta} = \phi(\mathbf{X}^T \mathbf{V} \mathbf{X})$ obtemos exatamente $\widehat{\mathbf{GL}} = \hat{\mathbf{V}} \mathbf{X} (\mathbf{X}^T \hat{\mathbf{V}} \mathbf{X})^{-1} \mathbf{X}^T$.

Outra definição de pontos de alavanca que tem sido muito utilizada na classe dos MLGs, embora não coincida exatamente com a expressão acima, exceto no caso de resposta contínua e ligação canônica, é construída fazendo uma analogia entre a solução de máxima verossimilhança para $\hat{\boldsymbol{\beta}}$ num MLG e a solução de mínimos quadrados de uma regressão normal linear ponderada. Considerando a expressão para $\hat{\boldsymbol{\beta}}$ obtida na convergência do processo iterativo dado em (1.5), temos que

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{W}} \hat{\mathbf{z}},$$

com $\hat{\mathbf{z}} = \hat{\boldsymbol{\eta}} + \hat{\mathbf{W}}^{-1/2} \hat{\mathbf{V}}^{-1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}})$. Portanto, $\hat{\boldsymbol{\beta}}$ pode ser interpretado como sendo a solução de mínimos quadrados da regressão linear de $\hat{\mathbf{W}}^{1/2} \hat{\mathbf{z}}$ contra as colunas de $\hat{\mathbf{W}}^{1/2} \mathbf{X}$. A matriz de projeção da solução de mínimos quadrados da regressão linear de $\hat{\mathbf{z}}$ contra \mathbf{X} com pesos $\hat{\mathbf{W}}$ fica dada por

$$\hat{\mathbf{H}} = \hat{\mathbf{W}}^{1/2} \mathbf{X} (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{W}}^{1/2},$$

que sugere a utilização dos elementos \hat{h}_{ii} da diagonal principal de $\hat{\mathbf{H}}$ para detectarmos a presença de pontos de alavanca nesse modelo de regressão normal linear ponderada. Essa extensão para MLGs foi proposta por Pregibon (1981). Podemos verificar facilmente que $\hat{h}_{ii} = \widehat{\mathbf{GL}}_{ii}$, ou seja, para grandes amostras $\widehat{\mathbf{GL}}$ e $\hat{\mathbf{H}}$ coincidem. No caso de ligação canônica essa igualdade vale para qualquer tamanho amostral. Como em geral \hat{h}_{ii} depende de $\hat{\mu}_{ii}$ sugerimos para detectar pontos de alavanca o gráfico de \hat{h}_{ii} contra os valores ajustados.

Moolgavkar, Lustbader e Venzon (1984) estendem a proposta de Pregibon para modelos não lineares e sugerem o uso dos elementos da diagonal principal da matriz de projeção no plano tangente à solução de máxima verossimilhança $\boldsymbol{\mu}(\hat{\boldsymbol{\beta}})$ para detectar pontos de alavanca. Hosmer e Lemeshow (1989) mostram, contudo, que o uso da diagonal principal da matriz de projeção $\hat{\mathbf{H}}$ deve ser feito com algum cuidado em regressão logística e que as interpretações são diferentes daquelas do caso normal linear.

1.10.2 Resíduos

A definição de um resíduo studentizado para os MLGs pode ser feita analogamente à regressão normal linear como veremos a seguir. Todavia, não necessariamente as propriedades continuam valendo. Assim, torna-se importante a definição de outros tipos de resíduo cujas propriedades sejam conhecidas ou pelo menos estejam mais próximas das propriedades de t_i^* .

Uma primeira proposta seria considerarmos o resíduo ordinário da solução de mínimos quadrados da regressão linear ponderada de $\hat{\mathbf{z}}$ contra \mathbf{X} , que é definido por

$$\mathbf{r}^* = \hat{\mathbf{W}}^{1/2}[\hat{\mathbf{z}} - \hat{\boldsymbol{\eta}}] = \hat{\mathbf{V}}^{-1/2}(\mathbf{y} - \hat{\boldsymbol{\mu}}).$$

Se assumirmos que $\text{Var}(\mathbf{z}) \cong \hat{\mathbf{W}}^{-1}\phi^{-1}$, temos aproximadamente $\text{Var}[\mathbf{r}^*] \cong \phi^{-1}(\mathbf{I}_n - \hat{\mathbf{H}})$. Logo, podemos definir o resíduo padronizado

$$t_{S_i} = \frac{\phi^{1/2}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - \hat{h}_{ii})}},$$

em que h_{ii} é o i -ésimo elemento da diagonal principal da matriz \mathbf{H} . Fica fácil mostrarmos que $\mathbf{r}^* = (\mathbf{I}_n - \hat{\mathbf{H}})\hat{\mathbf{W}}^{1/2}\hat{\mathbf{z}}$, isto é, $\hat{\mathbf{H}}$ desempenha o papel de matriz de projeção ortogonal local, como na regressão normal linear em que \mathbf{W} é identidade.

No entanto, na prática, $\hat{\boldsymbol{\eta}}$ não é fixo nem conhecido, bem como \mathbf{z} não segue distribuição normal. Uma implicação desse fato é que as propriedades de t_i^* não são mais verificadas para t_{S_i} . Williams (1984) mostra através de estudos de Monte Carlo que a distribuição de t_{S_i} é em geral assimétrica, mesmo para grandes amostras.

Outros resíduos cujas distribuições poderiam estar mais próximas da normalidade têm sido sugeridos para os MLGs. Por exemplo, o resíduo de Anscombe

$$t_{A_i} = \frac{\phi^{1/2}\{\psi(y_i) - \psi(\hat{\mu}_i)\}}{\hat{V}^{1/2}(\hat{\mu}_i)\psi'(\hat{\mu}_i)},$$

em que $\psi(\cdot)$ é uma transformação utilizada para normalizar a distribuição de Y . Para os MLGs essa transformação é definida por

$$\psi(\mu) = \int_0^\mu V^{-1/3}(t)dt.$$

Em particular, para os principais MLGs o resíduo de Ascombe é descrito na tabela dada abaixo.

Distribuição					
	Normal	Binomial	Poisson	Gama	N. Inversa
$\psi(\mu)$	μ	$\int_0^\mu t^{-1/3}(1-t)^{-1/3}dt$	$\frac{3}{2}\mu^{2/3}$	$3\mu^{1/3}$	$\log\mu$

Contudo, os resíduos mais utilizados em modelos lineares generalizados são definidos a partir dos componentes da função desvio. A versão padronizada (ver McCullagh, 1987; Davison e Gigli, 1989) é a seguinte:

$$t_{D_i} = \frac{d^*(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}} = \frac{\phi^{1/2} d(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}},$$

em que $d(y_i; \hat{\mu}_i) = \pm \sqrt{2} \{y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i))\}^{1/2}$. O sinal de $d(y_i; \hat{\mu}_i)$ é o mesmo de $y_i - \hat{\mu}_i$. Williams (1984) verificou através de simulações que a distribuição de t_{D_i} tende a estar mais próxima da normalidade do que as distribuições dos demais resíduos. McCullagh (1987, p. 214) mostra para os MLGs que a distribuição de probabilidades de

$$\frac{d^*(Y_i; \mu_i) + \rho_{3i}/6}{\sqrt{1 + (14\rho_{3i}^2 - 9\rho_{4i})/36}}$$

é aproximadamente $N(0, 1)$, em que ρ_{3i} e ρ_{4i} são os coeficientes de assimetria e curtose de $\partial L(\eta_i)/\partial \eta_i$, respectivamente, e $d^*(Y_i; \mu_i)$ é o i -ésimo componente do desvio $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ avaliado no parâmetro verdadeiro. Podemos mostrar usando resultados de Cox e Snell (1968) que $E\{d^*(Y_i; \mu_i)\} = 0$ e $\text{Var}\{d^*(Y_i; \mu_i)\} = 1 - h_{ii}$, em que os termos negligenciados são de $O(n^{-1})$. Esses resultados reforçam o uso da padronização $\sqrt{1 - \hat{h}_{ii}}$ para $d^*(y_i; \hat{\mu}_i)$.

Um quarto resíduo foi definido por Williams (1987) e pode ser interpretado como uma média ponderada entre t_{S_i} e t_{D_i} ,

$$t_{G_i} = \text{sinal}(y_i - \hat{\mu}_i) \{(1 - \hat{h}_{ii})t_{D_i}^2 + \hat{h}_{ii}t_{S_i}^2\}^{1/2}.$$

Williams (1987) verificou também através de simulações e para alguns MLGs que t_{G_i} tem esperança ligeiramente diferente de zero, variância excedendo um, assimetria desprezível e alguma curtose.

O R solta os resíduos $d_i = d(y_i; \hat{\mu}_i)$ e \hat{r}_{P_i} sem o termo $\phi^{1/2}$. Precisamos, portanto, para padronizá-los, calcular os correspondentes \hat{h}'_{ii} s bem como extrair $\hat{\phi}$ nos casos em que $\phi \neq 1$. Inicialmente, ilustramos como calcular \hat{h}_{ii} .

Vamos supor um modelo com duas covariáveis e dois fatores e que os resultados do ajuste são armazenados em `fit.model`. A matriz \mathbf{X} é obtida com um dos comandos abaixo

```
X = model.matrix( ~ cov1 + cov2 + A + B)
X = model.matrix(fit.model).
```

Em \mathbf{V} podemos armazenar a matriz $\hat{\mathbf{V}}$. Os elementos da diagonal principal de \mathbf{V} devem ser obtidos dos valores ajustados do modelo, os quais por sua vez são extraídos através do comando `fitted(fit.model)`. Como exemplo, a matriz com as funções de variância estimadas seria obtida para um modelo de Poisson da forma seguinte:

```
V = fitted(fit.model)
V = diag(V).
```

Em particular, a matriz $\hat{\mathbf{W}}$ também depende dos valores ajustados, no entanto, como é a matriz de pesos, podemos obtê-la diretamente fazendo

```
w = fit.model$weights
W = diag(w).
```

Assim, uma vez obtida a matriz $\hat{\mathbf{W}}$ podemos obter os elementos \hat{h}_{ii} com os comandos

```
H = solve(t(X)%*%W%*%X)
H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h = diag(H).
```

Armazenando em `fit` a estimativa $\hat{\phi}$ (o R solta $\hat{\phi}^{-1}$), os componentes do desvio e os resíduos studentizados são obtidos da seguinte maneira:

```
rd = resid(fit.model, type= ‘‘deviance’’)
td = rd*sqrt(fi/(1-h))
rp = resid(fit.model, type= ‘‘pearson’’)
rp = sqrt(fi)*rp
```

$$\text{ts} = \text{rp}/\text{sqrt}(1 - \text{h}).$$

Lembrando que para ligações canônicas \mathbf{W} e \mathbf{V} coincidem.

Por fim, temos o resíduo quantílico (Dunn e Smyth, 1996) que é definido por

$$r_{q_i} = \Phi^{-1}\{F(y_i; \hat{\boldsymbol{\theta}})\},$$

em que $\Phi(\cdot)$ e $F(y_i; \boldsymbol{\theta})$ denotam, respectivamente, as Fda's da $N(0, 1)$ e da distribuição postulada para a resposta, $i = 1, \dots, n$. Para n grande os resíduos r_{q_1}, \dots, r_{q_n} são independentes e igualmente distribuídos $N(0, 1)$. Assim, o gráfico normal de probabilidades para o resíduo quantílico é recomendado para avaliar afastamentos da distribuição postulada para a resposta bem como a presença de observações aberrantes.

1.10.3 Influência

Supondo ϕ conhecido, o afastamento pela verossimilhança quando eliminarmos a i -ésima observação é denotado por

$$\text{LD}_i = 2\{\text{L}(\hat{\boldsymbol{\beta}}) - \text{L}(\hat{\boldsymbol{\beta}}_{(i)})\},$$

sendo portanto uma medida que verifica a influência da retirada da i -ésima observação em $\hat{\boldsymbol{\beta}}$. Não sendo possível obtermos uma forma analítica para LD_i , é usual utilizarmos a segunda aproximação por série de Taylor em torno de $\hat{\boldsymbol{\beta}}$. Essa expansão leva ao seguinte resultado:

$$\text{LD}_i \cong (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \{-\ddot{\mathbf{L}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}})\}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}).$$

Substituindo $-\ddot{\mathbf{L}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}})$ pelo correspondente valor esperado e $\boldsymbol{\beta}$ por $\hat{\boldsymbol{\beta}}_{(i)}$, obtemos

$$\text{LD}_i \cong \phi(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^T (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}). \quad (1.17)$$

Assim, teremos uma boa aproximação para LD_i quando $L(\boldsymbol{\beta})$ for aproximadamente quadrática em torno de $\hat{\boldsymbol{\beta}}$.

Como em geral não é possível obtermos uma forma fechada para $\hat{\boldsymbol{\beta}}_{(i)}$, a aproximação de um passo tem sido utilizada (ver, por exemplo, Cook e Weisberg, 1982), que consiste em tomarmos a primeira iteração do processo iterativo pelo método escore de Fisher quando o mesmo é iniciado em $\hat{\boldsymbol{\beta}}$.

Essa aproximação, introduzida por Pregibon (1981), é dada por

$$\hat{\boldsymbol{\beta}}_{(i)}^1 = \hat{\boldsymbol{\beta}} + \{-\ddot{\mathbf{L}}_{\beta\beta}(\hat{\boldsymbol{\beta}})\}^{-1} \mathbf{L}_{(i)}(\hat{\boldsymbol{\beta}}),$$

em que $\mathbf{L}_{(i)}(\boldsymbol{\beta})$ é o logaritmo da função de verossimilhança sem a i -ésima observação. Substituindo novamente $-\ddot{\mathbf{L}}_{\beta\beta}(\hat{\boldsymbol{\beta}})$ por $\mathbf{K}(\hat{\boldsymbol{\beta}})$ obtemos

$$\hat{\boldsymbol{\beta}}_{(i)}^1 = \hat{\boldsymbol{\beta}} - \frac{\hat{r}_{P_i} \sqrt{\hat{\omega}_i \phi^{-1}}}{(1 - \hat{h}_{ii})} (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i \quad (1.18)$$

e, finalmente, substituindo a expressão acima em (1.17) temos que

$$LD_i \cong \left\{ \frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})} \right\} t_{S_i}^2.$$

A distância de Cook aproximada fica facilmente obtida com o comando

$$LD = \mathbf{h} * (\mathbf{ts}^2) / (1 - \mathbf{h}).$$

A validade da aproximação de um passo tem sido investigada por alguns pesquisadores. A constatação é que a mesma em geral subestima o verdadeiro valor de LD_i , no entanto é suficiente para chamar a atenção dos pontos influentes.

1.10.4 Influência local

A metodologia de influência local pode ser facilmente estendida para a classe de MLGs. Em particular, se considerarmos ϕ conhecido e perturbação de

casos em que $L(\boldsymbol{\beta}|\boldsymbol{\delta}) = \sum_{i=1}^n \delta_i L_i(\boldsymbol{\beta})$ com $0 \leq \delta_i \leq 1$, então a matriz $\boldsymbol{\Delta}$ assume a forma

$$\boldsymbol{\Delta} = \sqrt{\phi} \mathbf{X}^T \hat{\mathbf{W}}^{1/2} \mathbf{D}(\hat{\mathbf{r}}_P),$$

em que $\mathbf{D}(\hat{\mathbf{r}}_P) = \text{diag}\{\hat{r}_{P_1}, \dots, \hat{r}_{P_n}\}$ e $\hat{r}_{P_i} = \sqrt{\phi}(y_i - \hat{\mu}_i)/\sqrt{\hat{V}_i}$ é o i -ésimo resíduo de Pearson estimado. Assim, se substituirmos $-\ddot{\mathbf{L}}_{\beta\beta}$ por $\phi(\mathbf{X}^T \mathbf{W} \mathbf{X})$ temos que a curvatura normal na direção unitária $\boldsymbol{\ell}$ assume a forma

$$C_{\boldsymbol{\ell}}(\boldsymbol{\beta}) = 2|\boldsymbol{\ell}^T \mathbf{D}(\hat{\mathbf{r}}_P) \hat{\mathbf{H}} \mathbf{D}(\hat{\mathbf{r}}_P) \boldsymbol{\ell}|.$$

Se o interesse é calcularmos a curvatura normal na direção $\boldsymbol{\ell}_i$ da i -ésima observação, então podemos avaliar o gráfico de índices de $C_i = 2\hat{h}_{ii}\hat{r}_{P_i}^2$.

Em particular, o vetor $\boldsymbol{\ell}_{max}$ para avaliarmos a influência local das observações nas estimativas dos parâmetros é o autovetor correspondente ao maior autovalor da seguinte matriz $n \times n$:

$$\mathbf{B} = \mathbf{D}(\hat{\mathbf{r}}_P) \hat{\mathbf{H}} \mathbf{D}(\hat{\mathbf{r}}_P).$$

Para obtermos $\boldsymbol{\ell}_{max}$, a maneira mais simples é construirmos a matriz \mathbf{B} e extrairmos o seu autovetor correspondente ao maior autovalor. Os comandos são os seguintes:

```
B = diag(rp)%% H %% diag(rp)
Cmax = eigen(B)$val[1]
lmax = eigen(B)$vec[,1]
lmax = abs(lmax).
```

Por outro lado, se temos interesse em detectar observações influentes na estimativa de um coeficiente particular, associado por exemplo à variável explicativa \mathbf{X}_1 , o vetor $\boldsymbol{\ell}_{max}$ fica dado por

$$\boldsymbol{\ell}_{max}^T = \left(\frac{v_1 \hat{r}_{P_1}}{\sqrt{C_{\ell_{max}}}}, \dots, \frac{v_n \hat{r}_{P_n}}{\sqrt{C_{\ell_{max}}}} \right),$$

em que v_1, \dots, v_n são agora obtidos da regressão linear de \mathbf{X}_1 contra as colunas de \mathbf{X}_2 com matriz de pesos $\hat{\mathbf{V}}$, isto é $\mathbf{v} = \hat{\mathbf{V}}^{1/2}\mathbf{X}_1 - \hat{\mathbf{V}}^{1/2}\mathbf{X}_2(\mathbf{X}_2^T\hat{\mathbf{V}}\mathbf{X}_2)^{-1}\mathbf{X}_2^T\hat{\mathbf{V}}\mathbf{X}_1$.

Para ligação não canônica os resultados continuam valendo desde que a matriz observada de Fisher seja substituída pela matriz de informação de Fisher.

1.10.5 Gráfico da variável adicionada

Apresentamos a seguir a versão do gráfico da variável adicionada para os MLGs. Vamos supor um MLG com p parâmetros, β_1, \dots, β_p , ϕ conhecido, e que um parâmetro adicional γ está sendo incluído no modelo. O interesse é testarmos $H_0 : \gamma = 0$ contra $H_1 : \gamma \neq 0$.

Seja $\boldsymbol{\eta}(\boldsymbol{\beta}, \gamma)$ o preditor linear com $p + 1$ parâmetros, isto é

$$\boldsymbol{\eta}(\boldsymbol{\beta}, \gamma) = \mathbf{X}^T\boldsymbol{\beta} + \gamma\mathbf{Z}.$$

A função escore para γ é dada por

$$U_\gamma = \frac{\partial L(\boldsymbol{\beta}, \gamma)}{\partial \gamma} = \phi^{1/2}\mathbf{Z}^T\mathbf{W}^{1/2}\mathbf{r}_P,$$

em que $\mathbf{Z} = (z_1, \dots, z_n)^T$. De resultados anteriores temos que

$$\text{Var}(\hat{\gamma}) = \phi^{-1}[\mathbf{Z}^T\mathbf{W}^{1/2}\mathbf{M}\mathbf{W}^{1/2}\mathbf{Z}]^{-1},$$

em que $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$. Logo, $\text{Var}(\hat{\gamma}) = \phi^{-1}(\mathbf{R}^T\mathbf{W}\mathbf{R})^{-1}$ com $\mathbf{R} = \mathbf{Z} - \mathbf{X}\mathbf{C}$ e $\mathbf{C} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Z}$.

Portanto, a estatística de escore para testarmos $H_0 : \gamma = 0$ contra $H_1 : \gamma \neq 0$ fica dada por

$$\xi_{SR} = (\hat{\mathbf{r}}_P^T\hat{\mathbf{W}}^{1/2}\mathbf{Z})^2/(\mathbf{Z}^T\hat{\mathbf{W}}^{1/2}\hat{\mathbf{M}}\hat{\mathbf{W}}^{1/2}\mathbf{Z}),$$

em que $\hat{\mathbf{W}}$, $\hat{\mathbf{r}}_P$ e $\hat{\mathbf{M}}$ são avaliados em $\hat{\beta}$ (sob H_0). Sob H_0 , $\xi_{SR} \sim \chi_1^2$ quando $n \rightarrow \infty$.

Wang (1985) mostra que a estatística de escore acima coincide com a estatística F de uma regressão linear ponderada para testar a inclusão da variável \mathbf{Z} no modelo. Nessa regressão linear, o gráfico da variável adicionada é formado pelos resíduos $\hat{\mathbf{r}}_P$ e $\mathbf{v} = \phi^{1/2}(\mathbf{I}_n - \hat{\mathbf{H}})\hat{\mathbf{W}}^{1/2}\mathbf{Z}$. O resíduo \mathbf{v} pode ser obtido facilmente após a regressão linear ponderada (com pesos $\hat{\mathbf{W}}$) de \mathbf{Z} contra \mathbf{X} . Temos que $\hat{\gamma} = (\mathbf{v}^T \mathbf{v})^{-1} \mathbf{v}^T \mathbf{r}$.

Logo, o gráfico de $\hat{\mathbf{r}}_P$ contra \mathbf{v} pode revelar quais observações estão contribuindo mais na significância de γ . A principal dificuldade para construirmos o gráfico da variável adicionada em MLGs é a obtenção do resíduo \mathbf{v} , uma vez que o resíduo $\hat{\mathbf{r}}_P$ é obtido facilmente como já vimos anteriormente. Para ilustrarmos o cálculo de \mathbf{v} num modelo particular, vamos supor que temos duas covariáveis e dois fatores e que o interesse é construirmos o gráfico da variável adicionada correspondente à covariável `cov1`. Precisamos inicialmente ajustar o modelo com os dois fatores e a outra covariável e calcular a matriz $\hat{\mathbf{W}}$ cujos valores serão armazenados em \mathbf{W} . Lembrando que $\hat{\mathbf{W}}$ é a matriz estimada de pesos. Supondo, por exemplo, que temos um modelo de Poisson com ligação canônica, os passos para construirmos o gráfico são os seguintes:

```
fit.poisson = glm( resp ~ cov2 + A + B, family=poisson)
w = fit.poisson$weights
W = diag(w)
rp = resid(fit.poisson, type = 'pearson')
X = model.matrix(fit.poisson)
H = solve(t(X)%*%W%*X)
H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
```



```
v = sqrt(W)%%cov1 - H%%sqrt(W)%%cov1
plot(v, rp, xlab='Residuo v', ylab='Residuo rp').
```

1.10.6 Técnicas gráficas

As técnicas gráficas mais recomendadas para os MLGs são as seguintes: (i) gráficos de t_{D_i} contra a ordem das observações, contra os valores ajustados e contra as variáveis explicativas, ou contra o tempo ou alguma ordem em que há suspeita de correlação entre as observações; (ii) gráfico normal de probabilidades para t_{D_i} com envelope, (iii) gráfico de \hat{z}_i contra $\hat{\eta}_i$ para verificarmos a adequação da função de ligação (uma tendência linear indica adequação da ligação) e (iv) gráficos de LD_i , C_i ou $|\ell_{max}|$ contra a ordem das observações. Os envelopes, no caso de MLGs com distribuições diferentes da normal, são construídos com os resíduos sendo gerados a partir do modelo ajustado (ver, por exemplo, Williams, 1987). No Apêndice B são relacionados programas para gerar envelopes em alguns MLGs.

1.11 Seleção de modelos

1.11.1 Modelo normal linear

Existem vários procedimentos para a seleção de modelos de regressão, embora nenhum deles seja consistente, ou seja, mesmo para amostras grandes selecione com probabilidade um as variáveis explicativas com coeficiente de regressão não nulo. Os procedimentos mais conhecidos são os métodos que consideram todas as regressões possíveis, tais como maior R_p^2 , menor s_p^2 , C_p e **Press**, bem como os métodos sequenciais **forward**, **backward**, **stepwise**, **AIC** e **BIC** (ver, por exemplo, Neter et al., 1996, Cap. 8; Montgomery et al., 2012, Cap.10), além de outros métodos que usam computação intensiva.

Alguns desses métodos serão descritos brevemente a seguir.

Método forward

Iniciamos o método pelo modelo $\mu = \alpha$. Ajustamos então para cada variável explicativa o modelo

$$\mu = \alpha + \beta_j x_j, \quad (j = 1, \dots, q).$$

Testamos $H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$. Seja P o menor nível descritivo dentre os q testes. Se $P \leq P_E$, a variável correspondente entra no modelo. Vamos supor que X_1 tenha sido escolhida. Então, no passo seguinte ajustamos os modelos

$$\mu = \alpha + \beta_1 x_1 + \beta_j x_j, \quad (j = 2, \dots, q).$$

Testamos $H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$. Seja P o menor nível descritivo dentre os $(q - 1)$ testes. Se $P \leq P_E$, a variável correspondente entra no modelo. Repetimos o procedimento até que ocorra $P > P_E$.

Método backward

Iniciamos o procedimento pelo modelo

$$\mu = \alpha + \beta_1 x_1 + \dots + \beta_q x_q.$$

Testamos $H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$ para $j = 1, \dots, q$. Seja P o maior nível descritivo dentre os q testes. Se $P > P_S$, a variável correspondente sai do modelo. Vamos supor que X_1 tenha saído do modelo. Então, ajustamos o modelo

$$\mu = \alpha + \beta_2 x_2 + \dots + \beta_q x_q.$$

Testamos $H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$ para $j = 2, \dots, q$. Seja P o maior nível descritivo dentre os $(q - 1)$ testes. Se $P > P_S$, então a variável correspondente sai do modelo. Repetimos o procedimento até que ocorra $P \leq P_S$.

Método stepwise

É uma mistura dos dois procedimentos anteriores. Iniciamos o processo com o modelo $\mu = \alpha$. Após duas variáveis terem sido incluídas no modelo, verificamos se a primeira não sai do modelo. O processo continua até que nenhuma variável seja incluída ou seja retirada do modelo. Geralmente adotamos $0,15 \leq P_E, P_S \leq 0,25$. Uma sugestão seria usar $P_E = P_S = 0,20$.

Método de Akaike

O método proposto por Akaike (1974) basicamente se diferencia dos procedimentos anteriores por ser um processo de minimização que não envolve testes estatísticos. A ideia básica é selecionarmos um modelo que seja parcimonioso, ou em outras palavras, que esteja bem ajustado e tenha um número reduzido de parâmetros. Como o logaritmo da função de verossimilhança $L(\boldsymbol{\beta})$ cresce com o aumento do número de parâmetros do modelo, uma proposta razoável seria encontrarmos o modelo com menor valor para a função

$$AIC = -2L(\hat{\boldsymbol{\beta}}) + 2p,$$

em que p denota o número de parâmetros. No caso do modelo normal linear podemos mostrar que AIC fica expresso, quando σ^2 é desconhecido, na forma

$$AIC = n \log\{D(\mathbf{y}; \hat{\boldsymbol{\mu}})/n\} + 2p,$$

em que $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2$.

Método de Schwartz

O método de Schwartz (1978) consiste em maximizar $L(\boldsymbol{\beta})$ também minimizando o número de coeficientes da regressão, porém com uma penalização diferente. O método é equivalente a minimizar a função abaixo

$$\text{BIC} = -2L(\hat{\boldsymbol{\beta}}) + p \log(n).$$

Para a regressão linear múltipla tem-se que $\text{BIC} = n \log\{D(\mathbf{y}; \hat{\boldsymbol{\mu}})/n\} + p \log(n)$.

1.11.2 Extensão para os MLGs

Os métodos de seleção de modelos descritos na seção anterior podem ser estendidos diretamente para os MLGs. Algumas observações, contudo, são necessárias. Nos casos de regressão logística e de Poisson o teste da razão de verossimilhanças, pelo fato de ser obtido pela diferença de duas funções desvio, aparece como o mais indicado. Para os casos de regressão normal, normal inversa e gama o teste F, por não exigir a estimativa de máxima verossimilhança do parâmetro de dispersão, é o mais indicado. Isso não impede que outros testes sejam utilizados.

Já o método de Akaike pode ser expresso numa forma mais simples em função do desvio do modelo. Nesse caso, o critério consiste em encontrarmos o modelo tal que a quantidade abaixo seja minimizada

$$\text{AIC} = D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) + 2p,$$

em que $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ denota o desvio do modelo e p o número de parâmetros. Os métodos `stepwise` e de Akaike estão disponíveis no R. O método `stepwise` está disponível apenas para modelos normais lineares. O comando `stepwise` é definido por `stepwise(Xvar, resposta)`, em que `Xvar` denota a matriz com os valores das variáveis explicativas e `resposta` denota o vetor com as respostas.

Para rodarmos o critério de Akaike precisamos usar antes o comando `require(MASS)`. Uma maneira de aplicarmos o critério de Akaike é partindo do maior modelo cujos resultados são guardados no objeto `fit.model`. Daí, então, devemos usar o comando `stepAIC(fit.model)`.

1.12 Aplicações

1.12.1 Estudo entre renda e escolaridade

O conjunto de dados descrito na Tabela 1.6, extraído do censo do IBGE de 2000, apresenta para cada unidade da federação o número médio de anos de estudo e a renda média mensal (em reais) do chefe ou chefes do domicílio. Esses dados estão também armazenados no arquivo **censo.dat**. Para lermos os dados no R e colocá-los num arquivo com o mesmo nome externo, devemos fazer

```
censo.dat = scan('censo.dat', what=list(uf=' ',
escolar=0, renda=0)).
```

Propomos inicialmente um modelo normal linear simples em que Y denote a renda e X a escolaridade. O modelo fica portanto dado por

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \dots, 27,$$

com a suposição de que $\epsilon_i \sim N(0, \sigma^2)$, sendo os erros mutuamente independentes. Estamos supondo que a amostra da Tabela 1.6 é um corte transversal, isto é, as informações sobre renda e escolaridade das unidades da federação são referentes a um determinado instante do tempo. Nesse caso, os erros são devidos à variabilidade da renda (dada a escolaridade) nos diversos instantes do tempo. Assumimos que a relação funcional entre y_i e x_i é a mesma num determinado intervalo do tempo.

Tabela 1.6
Escolaridade e renda média
domiciliar no Brasil em 2000.

RR	5,7	685	AP	6,0	683
AC	4,5	526	RO	4,9	662
PA	4,7	536	AM	5,5	627
TO	4,5	520	PB	3,9	423
MA	3,6	343	RN	4,5	513
SE	4,3	462	PI	3,5	383
BA	4,1	460	PE	4,6	517
AL	3,7	454	CE	4,0	448
SP	6,8	1076	RJ	7,1	970
ES	5,7	722	MG	5,4	681
SC	6,3	814	RS	6,4	800
PR	6,0	782	MT	5,4	775
GO	5,5	689	MS	5,7	731
DF	8,2	1499			

As estimativas dos parâmetros (erro padrão) são dadas por $\hat{\alpha} = -381,28$ (69,40) e $\hat{\beta} = 199,82$ (13,03), indicando que o coeficiente angular da reta é altamente significativo. Essa estimativa pode ser interpretada como o incremento esperado na renda média domiciliar de uma unidade da federação se o tempo de escolaridade médio domiciliar naquela unidade for acrescido de um ano.

A estimativa de σ é dada por $s = 77,22$, enquanto que o coeficiente de determinação foi de $R^2 = 0,904$. O ajuste do modelo e a exibição dos resultados podem ser obtidos com os comandos abaixo

```
attach(censo.dat)
fit1.censo = lm(renda ~ escolar)
summary(fit1.censo).
```

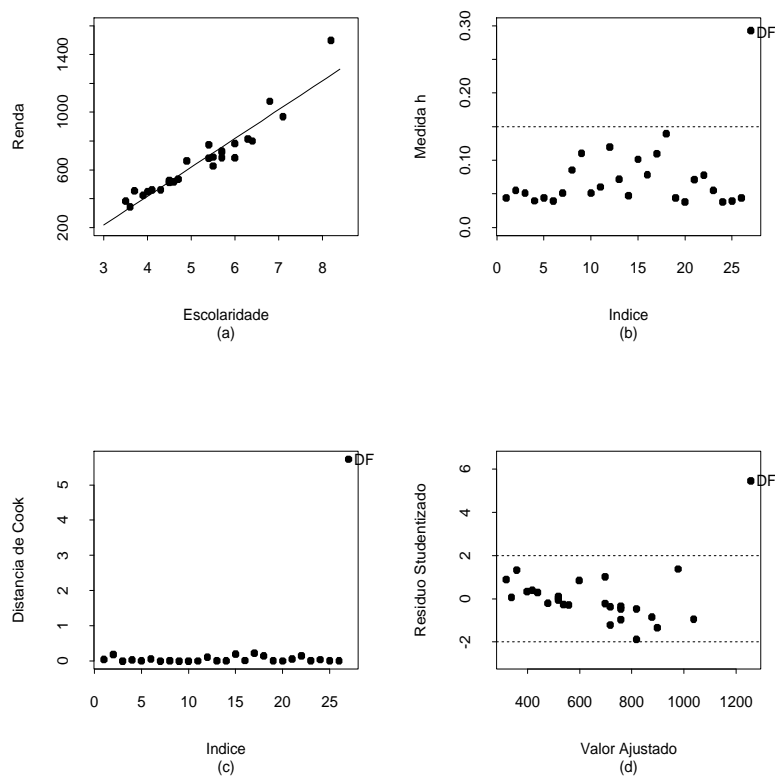


Figura 1.5: Reta ajustada do modelo aditivo-normal e gráficos de diagnóstico para o exemplo sobre renda e escolaridade.

Ou, alternativamente, transformando o arquivo **censo.dat** num arquivo do tipo **data.frame**, através dos comandos

```
censo.dat = data.frame(censo.dat)
fit1.censo = lm(renda ~ escolar, data=censo.dat)
summary(fit1.censo).
```

Pela Figura 1.5 onde são apresentados alguns gráficos de diagnóstico, além da reta ajustada aos dados, notamos uma forte discrepância do Distrito Federal que aparece como ponto de alavanca, influente e aberrante. Além

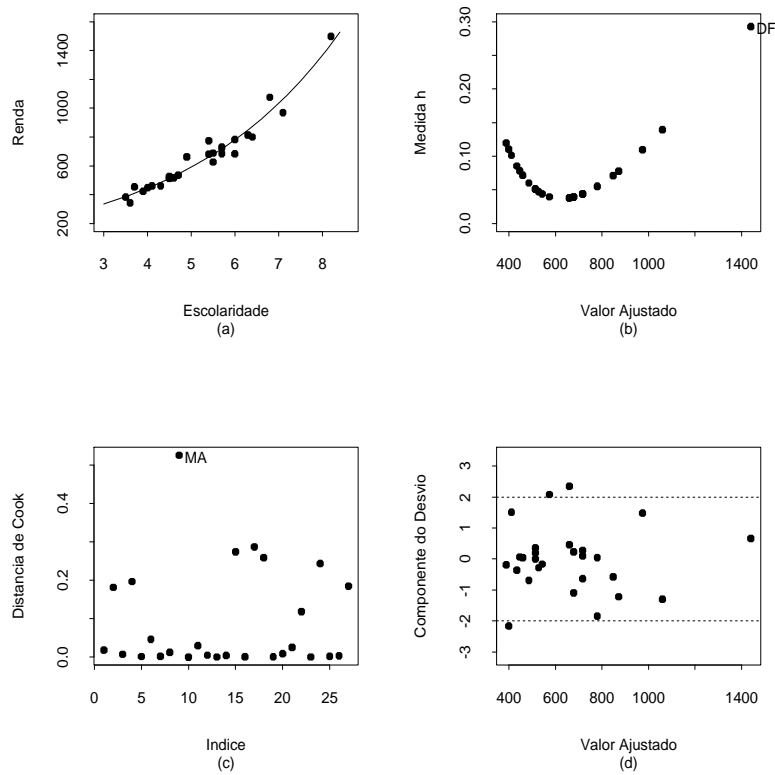


Figura 1.6: Curva ajustada do modelo multiplicativo-gama e gráficos de diagnóstico para o exemplo sobre renda e escolaridade.

disso, notamos pela Figura 1.5d indícios de heteroscedasticidade, ou seja, um aumento da variabilidade com o aumento da escolaridade. Isso pode também ser notado na Figura 1.5a. Assim, podemos propor um modelo alternativo, por exemplo, com efeitos multiplicativos conforme dado abaixo

$$\mu_i = e^{\alpha + \beta x_i} e^{\epsilon_i}, \quad i = 1, \dots, 27,$$

com a suposição de que $\epsilon_i \sim G(1, \phi)$, sendo os erros mutuamente independentes. Podemos ajustar esse modelo no R através dos comandos

```
fit2.censo = glm(renda ~ escolar, family=Gamma(link=log))
summary(fit1.censo).
```


Tabela 1.7

Estimativas de algumas quantidades com todos os pontos e quando as observações mais discrepantes são excluídas do modelo gama.

Estimativa	Com todos os pontos	Excluído DF	Excluído MA	Excluídos DF e MA
$\hat{\alpha}$	4,98 (0,068)	5,00 (0,078)	5,03 (0,067)	5,06 (0,077)
$\hat{\beta}$	0,28 (0,013)	0,27 (0,015)	0,27 (0,012)	0,26 (0,015)
$\hat{\phi}$	192(52)	188(52)	223(62)	223(63)

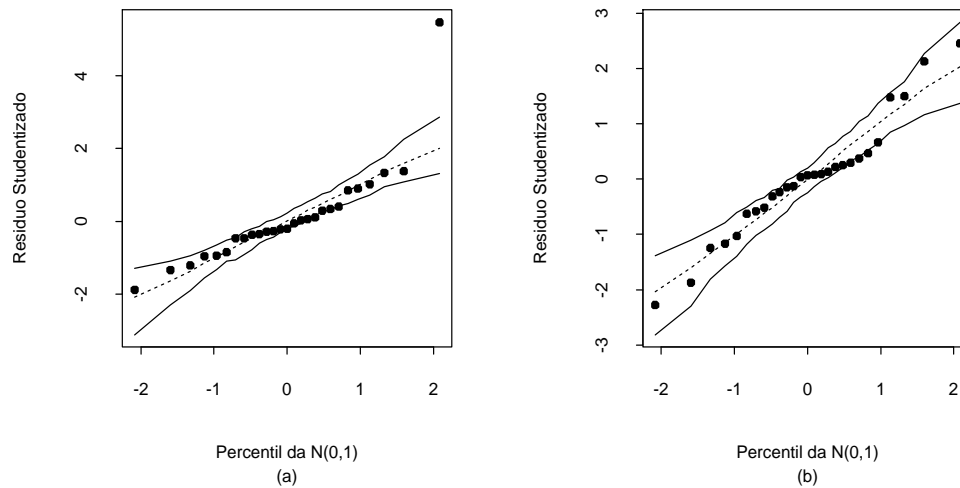


Figura 1.7: Gráficos normais de probabilidades para os modelos ajustados aditivo-normal (a) e multiplicativo-gama (b) aos dados sobre renda e escolaridade.

Na Figura 1.6 temos o ajuste do modelo gama aos dados bem como alguns gráficos de diagnóstico que destacam DF como ponto de alavanca e MA como ponto influente.

Na Tabela 1.7 temos uma análise confirmatória e verificamos poucas variações nas estimativas dos parâmetros com a eliminação dessas unidades da

federação. Finalmente, na Figura 1.7 temos os gráficos normais de probabilidades para os modelos com efeitos aditivos (Figura 1.7a) e com efeitos multiplicativos (Figura 1.7b) e notamos uma melhor acomodação e distribuição dos pontos dentro do envelope gerado no segundo caso. Pelo valor da estimativa do parâmetro de dispersão concluímos que o modelo gama aproxima-se bem de um modelo normal heteroscedástico.

Portanto, o modelo final ajustado fica dado por

$$\hat{y} = e^{4,98+0,28x}.$$

Desse modelo podemos extrair a seguinte interpretação: $e^{\hat{\beta}} = e^{0,28} = 1,32(32\%)$ é o aumento relativo esperado para a renda se aumentarmos em 1 ano a escolaridade média.

1.12.2 Processo infeccioso pulmonar

Utilizaremos agora os dados referentes a um estudo de caso-controle realizado no Setor de Anatomia e Patologia do Hospital Heliópolis em São Paulo, no período de 1970 a 1982 (Paula e Tuder, 1986) (ver arquivo **canc3.dat**). Um total de 175 pacientes com processo infeccioso pulmonar atendido no hospital no período acima foi classificado segundo as seguintes variáveis: Y, tipo de tumor (1: maligno, 0: benigno); IDADE, idade em anos; SEXO (0: masculino, 1: feminino); HL, intensidade da célula histiócitos-linfócitos (1: ausente, 2: discreta, 3: moderada, 4: intensa) e FF, intensidade da célula fibrose-frouxa (1: ausente, 2: discreta, 3: moderada, 4: intensa). Para ler os dados do arquivo **canc3.dat** e armazená-los num arquivo do R com o mesmo nome, fazemos

```
canc3.dat = scan('canc3.dat', what=list(tipo=0, idade=0, sexo=0,
hl=0, ff=0)).
```

Devemos informar o sistema que as variáveis SEXO, HL e FF são qualitativas, isto é, devemos transformá-las em fatores. Os comandos são os seguintes:

```
attach(canc3.dat)
sexo = factor(sexo)
sexo = C(sexo,treatment)
hl = factor(hl)
hl = C(hl,treatment)
ff = factor(ff)
ff = C(ff,treatment).
```

O comando `C(sexo,treatment)`, que é optativo, cria uma variável binária que assume valor zero para o sexo masculino e valor um para o sexo feminino. Analogamente, o comando `C(hl,treatment)` cria variáveis binárias para os níveis discreto, moderado e intenso do fator HL. O mesmo faz o comando `C(ff,treatment)` para o fator FF. Essa maneira de transformarmos todo fator de k níveis em $k - 1$ variáveis binárias, denominado casela de referência, é padrão em MLGs, porém pode não ser a modelagem mais conveniente em outras situações de interesse prático. A casela de referência seria, nesses dois casos, o nível ausente.

Vamos considerar, como exemplo, a aplicação do modelo logístico apenas com os efeitos principais

$$\Pr\{Y = 1 \mid \eta\} = \{1 + \exp(-\eta)\}^{-1},$$

em que $\eta = \beta_1 + \beta_2 \text{IDADE} + \beta_3 \text{SEXO} + \sum_{i=1}^4 \beta_{4i} \text{HL}_i + \sum_{i=1}^4 \beta_{5i} \text{FF}_i$, com SEXO, HL_i e FF_i sendo variáveis binárias correspondentes aos níveis de SEXO, HL e FF, respectivamente. Assumimos que $\beta_{41} = \beta_{51} = 0$. Uma observação importante é que devido ao fato da amostragem ter sido retrospectiva, o uso do modelo acima para fazermos previsões somente é válido se corrigirmos a estimativa da constante, β_1 (ver, por exemplo, McCullagh e

Nelder, 1989, p. 113). Discutimos isso na Seção 3.6.6. Para ajustarmos o modelo acima, os passos são dados abaixo

```
fit1.canc3 = glm( tipo ~ sexo + idade + hl + ff,
family=binomial)
summary(fit1.canc3).
```

Tabela 1.8

Estimativas dos parâmetros referentes ao modelo logístico ajustado aos dados sobre processo infeccioso pulmonar.

Efeito	Estimativa	Efeito	Estimativa	Efeito	Estimativa
Constante	-1,850(1,060)	HL(2)	-0,869(0,945)	FF(2)	-0,687(0,502)
Sexo	0,784(0,469)	HL(3)	-2,249(0,968)	FF(3)	-1,025(0,525)
Idade	0,065(0,013)	HL(4)	-3,295(1,466)	FF(4)	0,431(1,123)

As estimativas dos parâmetros (erro padrão aproximado) são apresentadas na Tabela 1.8. O desvio do modelo foi de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 157,40$ (166 graus de liberdade), indicando um ajuste adequado. Como podemos observar, há indícios de que a chance de processo infeccioso maligno seja maior para o sexo feminino do que para o sexo masculino. Notamos também que a chance de processo maligno aumenta significativamente com a idade e há indicações de que tanto para a célula FF quanto para HL a chance de processo maligno diminui à medida que aumenta a intensidade da célula. Esse exemplo será reanalisado no Capítulo 3.

1.12.3 Sobrevivência de bactérias

Descrevemos na Tabela 1.9 (Montgomery, Peck e Vining, 2001, pgs. 201-202) o número de bactérias sobreviventes em amostras de um produto alimentício segundo o tempo (em minutos) de exposição do produto a uma temperatura de $300^{\circ}F$.

Tabela 1.9*Número de bactérias sobreviventes e tempo de exposição.*

Número	175	108	95	82	71	50	49	31	28	17	16	11
Tempo	1	2	3	4	5	6	7	8	9	10	11	12

Na Figura 1.8a apresentamos o gráfico do número de bactérias sobreviventes contra o tempo de exposição. Notamos uma tendência decrescente e quadrática. Supondo que as amostras do produto enlatado submetidos à temperatura de $300^{\circ}F$ têm o mesmo tamanho, podemos pensar, em princípio, que $Y_i \sim P(\mu_i)$, com Y_i denotando o número de bactérias sobreviventes na i -ésima amostra $i = 1, \dots, n$. Como para μ_i grande é razoável assumirmos que Y_i segue uma distribuição aproximadamente normal (ver Seção 4.2.1), propomos inicialmente os seguintes modelos:

$$y_i = \alpha + \beta \text{tempo}_i + \epsilon_i \quad \text{e}$$

$$y_i = \alpha + \beta \text{tempo}_i + \gamma \text{tempo}_i^2 + \epsilon_i,$$

em que $\epsilon_i \sim N(0, \sigma^2)$. As estimativas dos parâmetros são apresentadas na Tabela 1.10. Pelos gráficos de envelope (Figuras 1.8b e 1.8c) notamos indícios de que a distribuição dos erros pode estar incorretamente especificada. A maioria dos resíduos assume valor negativo. Notamos a presença de um ponto aberrante, observação #1. Uma outra tentativa seria aplicar à resposta a transformação raiz quadrada que é conhecida no caso da Poisson como estabilizadora da variância além de manter a aproximação normal (ver Seção 4.2.1). Logo, podemos pensar em adotar os seguintes modelos alternativos:

$$\sqrt{y_i} = \alpha + \beta \text{tempo}_i + \epsilon_i \quad \text{e}$$

$$\sqrt{y_i} = \alpha + \beta \text{tempo}_i + \gamma \text{tempo}_i^2 + \epsilon_i,$$

em que $\epsilon_i \sim N(0, \sigma^2)$. As estimativas dos parâmetros são apresentadas na Tabela 1.10.

Tabela 1.10

Estimativas de algumas quantidades para os modelos com resposta transformada ajustados aos dados sobre sobrevivência de bactérias.

Parâmetro	Linear- Y	Quadrático- Y	Linear- \sqrt{Y}	Quadrático- \sqrt{Y}
α	142,20(11,26)	181,20(11,64)	12,57(0,38)	13,64(0,51)
β	-12,48(1,53)	-29,20(4,11)	-0,82(0,05)	-1,27(0,18)
γ		1,29(0,31)		0,04(0,01)
R^2	86,9%	95,5%	96,1%	97,8%

Notamos uma melhora na qualidade do ajuste, particularmente no segundo caso. Porém, ainda há indícios pelos gráficos de envelope (Figuras 1.8d e 1.8e) de violação nas suposições para os modelos, além da presença da observação #1 como ponto aberrante. Decidimos, então, propor um modelo log-linear de Poisson em que assumimos

$$Y_i \sim P(\mu_i) \text{ e } \log \mu_i = \alpha + \beta \text{tempo}_i.$$

As estimativas dos parâmetros são apresentadas na Tabela 1.11.

Tabela 1.11

Estimativas dos parâmetros do modelo de Poisson ajustado aos dados sobre sobrevivência de bactérias.

Parâmetro	Estimativa	E/E.Padrão
α	5,30	88,34
β	-0,23	-23,00
Desvio		8,42 (10 g.l.)

Pelo gráfico de envelope (Figura 1.8f) não há evidências de que o modelo esteja mal ajustado. Notamos também que a observação #1 foi acomodada dentro do envelope gerado. Parece, portanto, que esse último modelo é o que

melhor se ajusta aos dados dentre os modelos propostos. O modelo ajustado fica então dado por

$$\hat{\mu}(x) = e^{5,30-0,23x},$$

em que x denota o tempo de exposição. Logo, se diminuirmos de uma unidade o tempo de exposição a variação no valor esperado fica dada por

$$\frac{\hat{\mu}(x-1)}{\hat{\mu}(x)} = e^{0,23} = 1,259.$$

Ou seja, o número esperado de sobreviventes aumenta aproximadamente 25,9%.

1.12.4 Estudo seriado com ratos

O exemplo a seguir provém de um estudo seriado com um tipo de tumor maligno para avaliar a influência da série (passagem do tumor) na morte (caquexia) de um certo tipo de rato (ver Paula, Barbosa e Ferreira, 1989; Paula et al., 1992). Os dados estão descritos no arquivo **canc4.dat**. Um total de 204 animais teve o tumor inoculado num determinado momento da série. Para cada animal, além do grupo de passagem, foram observadas as variáveis presença de MASSa tumoral, caquexia e o tempo de observação (em dias). Esses dados são resumidos na Tabela 1.12. Para inserirmos os dados diretamente no R e armazená-los no arquivo **canc4a.dat**, devemos fazer

```
canc4a.dat = scan(what=list(obs=0,rd=0))
```

```
1:  6 2597 13 3105 8 2786
```

```
2: 12 1613 3 411 1 232.
```

Agora, precisamos introduzir os fatores grupo de passagem e massa tumoral

```
fnames = list(gp=c('P0-P6', 'P7-P18', 'P19-P28'),
mt=c('sim', 'nao')).
```

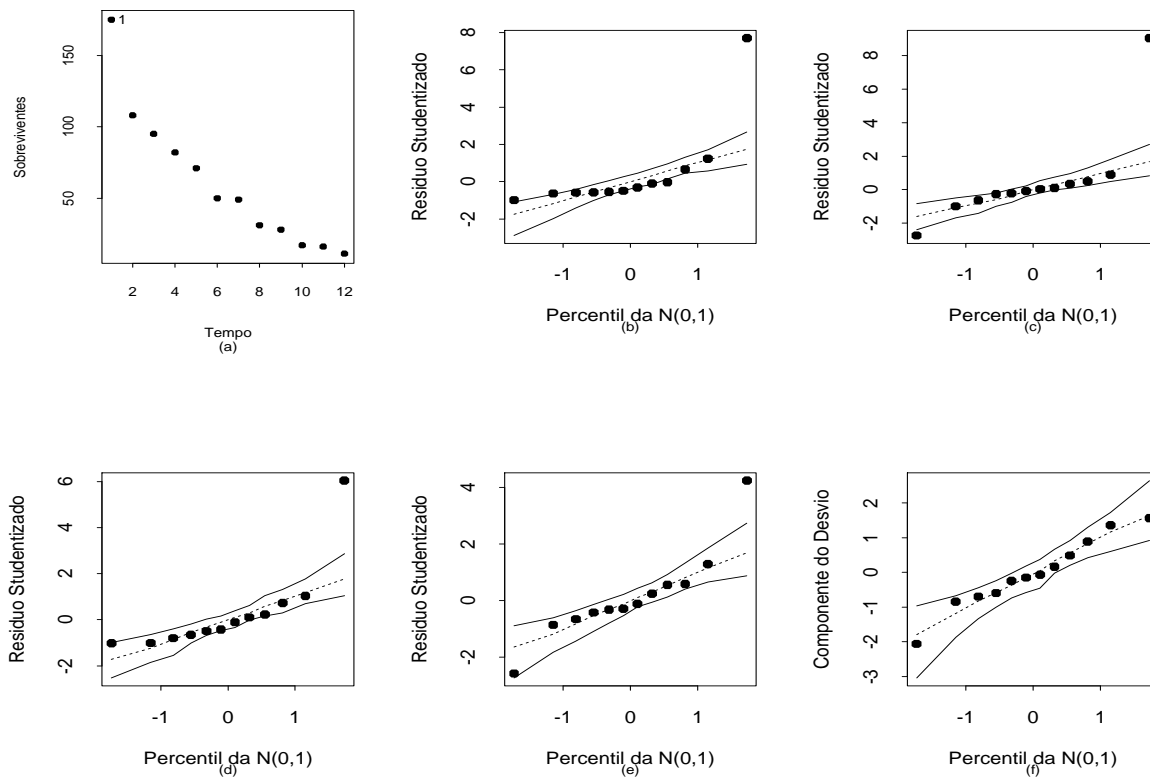


Figura 1.8: Diagrama de dispersão e gráficos normais de probabilidades referentes aos modelos ajustados aos dados sobre sobrevivência de bactérias.

Para informarmos o sistema a ordem em que os dados foram lidos, podemos usar o comando `fac.design`. Em seguida, fazemos o emparelhamento

```
rato.design = fac.design(c(3,2), fnames, rep=1)
attach(canc4a.dat)
rato.df = data.frame(obs,rd,rato.design).
```

As informações completas sobre os dados estão armazenadas no arquivo `rato.df`. Para uma verificação basta bater

```
rato.df.
```

Podemos agora (opcionalmente) criar uma matriz modelo no padrão dos MLGs


```
attach(rato.df)
gp = C(gp,treatment)
mt = C(mt,treatment).
```

Tabela 1.12

Número de ratos caquéticos (O) e ratos dias de observação (R-D) segundo o grupo de passagem e o desenvolvimento de massa tumoral.

Massa tumoral		Grupo de passagem		
		P0-P6	P7-P18	P19-P28
Sim	O	6	13	8
	R-D	2597	3105	2786
Não	O	12	3	1
	R-D	1613	411	232

Vamos supor que O_{ij} , o número de ratos caquéticos no nível i de massa tumoral e grupo de passagem j , segue uma distribuição de Poisson de média $\lambda_{ij}t_{ij}$, $i = 1, 2$ e $j = 1, 2, 3$. Temos que λ_{ij} denota a taxa de caquexia (número médio de mortes por unidade de tempo) e t_{ij} o total de ratos-dias no nível (i, j) . Consideramos inicialmente o modelo de Poisson com parte sistemática

$$\log \lambda_{ij} = \alpha + \beta_i + \gamma_j,$$

em que $\beta_1 = 0$ e $\gamma_1 = 0$, que equivale à suposição de tempos exponenciais como será visto na Seção 4.2.2. Com essa notação, α será o efeito correspondente à classe P0 – P6 com desenvolvimento de massa, β_2 a diferença entre os efeitos dos grupos sem e com o desenvolvimento de massa tumoral, γ_2 a diferença entre os efeitos das classes P7 – P18 e P0 – P6 e γ_3 a diferença entre os efeitos das classes P19 – P28 e P0 – P6. Em particular, quando expressamos os valores esperados de mortes para t_{ij} na forma de um modelo log-linear, teremos um *offset* dado por $\log t_{ij}$. Ou seja, o modelo que iremos

ajustar no R tem parte sistemática dada por $\log\mu_{ij} = \log t_{ij} + \alpha + \beta_i + \gamma_j$. Logo, precisamos definir o *offset* no ajuste.

Os passos são os seguintes:

```
logt0 = log(rd)
canc4a.fit = glm( obs ~ gp + mt + offset(logt0),
family=poisson)
summary(canc4a.fit).
```

As estimativas dos parâmetros (erro padrão aproximado) são dadas por $\hat{\alpha} = -5,875$ (0,312), $\hat{\gamma}_2 = 0,334$ (0,365), $\hat{\gamma}_3 = -0,040$ (0,434) e $\hat{\beta}_2 = 0,860$ (0,343). O desvio do modelo é dado por $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 0,84$ com 2 graus de liberdade. Pelas estimativas acima notamos que o fator grupo de passagem não é significativo. O ajuste do modelo sem esse efeito levou às estimativas $\hat{\alpha} = -5,750$ (0,192) e $\hat{\beta}_2 = 0,802$ (0,315) com um desvio de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 1,99$ (4 graus de liberdade). Logo, o teste da razão de verossimilhanças para testar $H_0 : \gamma_2 = \gamma_3 = 0$ vale $1,99 - 0,84 = 1,15$ com 2 graus de liberdade, o que implica em não rejeitarmos a hipótese H_0 . Assim, o modelo adotado inclui somente o efeito massa tumoral. Temos que β_2 é significativamente diferente de zero. A estimativa $\hat{\beta}_2 = 0,802$ indica que os ratos que desenvolvem massa tumoral (tumor maligno) sobrevivem mais do que os ratos que não desenvolvem o tumor! Esse resultado pode parecer em princípio contraditório, todavia devemos lembrar que todos os ratos tiveram tumor inoculado mas nem todos desenvolveram massa tumoral. Assim, pode ser razoável pensarmos que aqueles ratos que não desenvolveram massa tumoral na verdade teriam resistido muito para que a mesma não se desenvolvesse, levando os mesmos a algum tipo de esgotamento e conseqüentemente a um tempo médio de vida menor do que o tempo médio dos ratos em que o tumor se desenvolveu.

Uma maneira alternativa de avaliarmos a suposição de distribuição de Poisson para O_{ij} com média $\lambda_{ij}t_{ij}$ é através da inclusão do termo $\log t_{ij}$ como covariável, em vez de *offset*. Isto é, supomos a parte sistemática

$$\log \mu_{ij} = \alpha + \delta \log t_{ij} + \beta_i + \gamma_j$$

. Assim, podemos testar $H_0 : \delta = 1$ contra $H_1 : \delta \neq 1$. A não rejeição de H_0 indica que a suposição de distribuição de Poisson para O_{ij} parece ser razoável. No exemplo acima obtemos $\hat{\delta} = 1,390(0,439)$, o que nos leva a não rejeitarmos H_0 .

1.12.5 Consumo de combustível

No arquivo **reg2.dat**(Gray, 1989) são apresentadas as siglas dos 48 estados norte-americanos contíguos juntamente com as seguintes variáveis: taxa (taxa do combustível no estado em USD), licença (proporção de motoristas licenciados), renda (renda percapita em USD), estradas (ajuda federal para as estradas em mil USD) e consumo (consumo de combustível por habitante). O interesse nesse estudo é tentar explicar o consumo de combustível pelas variáveis taxa, licença, renda e estradas. O modelo proposto é o seguinte:

$$y_i = \alpha + \beta_1 \text{taxa}_i + \beta_2 \text{licenca}_i + \beta_3 \text{renda}_i + \beta_4 \text{estradas}_i + \epsilon_i,$$

em que y_i denota o consumo anual de combustível (por habitante) no i -ésimo estado, enquanto ϵ_i são variáveis aleatórias independentes normalmente distribuídas de média zero e variância σ^2 . Ajustamos o modelo acima no **R** e mandamos os resultados para o objeto **fit1.reg2**. Daí então aplicamos o método de Akaike para selecionarmos o submodelo com menor AIC. Para tal, aplicamos os comandos

```
require(MASS)
```

`stepAIC(fit1.reg1).`

A variável estradas foi eliminada. Os resultados do modelo selecionado são apresentados na Tabela 1.13.

Tabela 1.13

Estimativas dos parâmetros referentes ao modelo normal linear ajustado aos dados sobre consumo de combustível.

Efeito	Estimativa	E/E.Padrão
Constante	307,33	1,96
Taxa	-29,48	-2,78
Licença	1374,77	7,48
Renda	-0,07	-4,00
R ²	0,675	
s	8,12	

Portanto, podemos dizer que para cada aumento de uma unidade na renda, o consumo médio de combustível diminui 0,07 unidades. Para cada aumento de 1% na porcentagem de motoristas licenciados o consumo médio de combustível aumenta 13,75 unidades, e para cada aumento de 1% no imposto do combustível o consumo médio diminui 29,48 unidades.

Na Figura 1.9 temos alguns gráficos de diagnóstico e como podemos notar há um forte destaque para o estado de WY, que aparece como influente (Figura 1.9b) e aberrante (Figura 1.9c). Outros estados, tais como CT, NY, SD, TX e NV (Figura 1.9a) aparecem como remotos no subespaço gerado pelas colunas da matrix **X**, embora não sejam confirmados como influentes. Não há indícios pela Figura 1.9d de heteroscedasticidade.

Pelo gráfico de envelope (Figura 1.10a) não há indícios fortes de afastamentos da suposição de normalidade para os erros, apesar da influência no

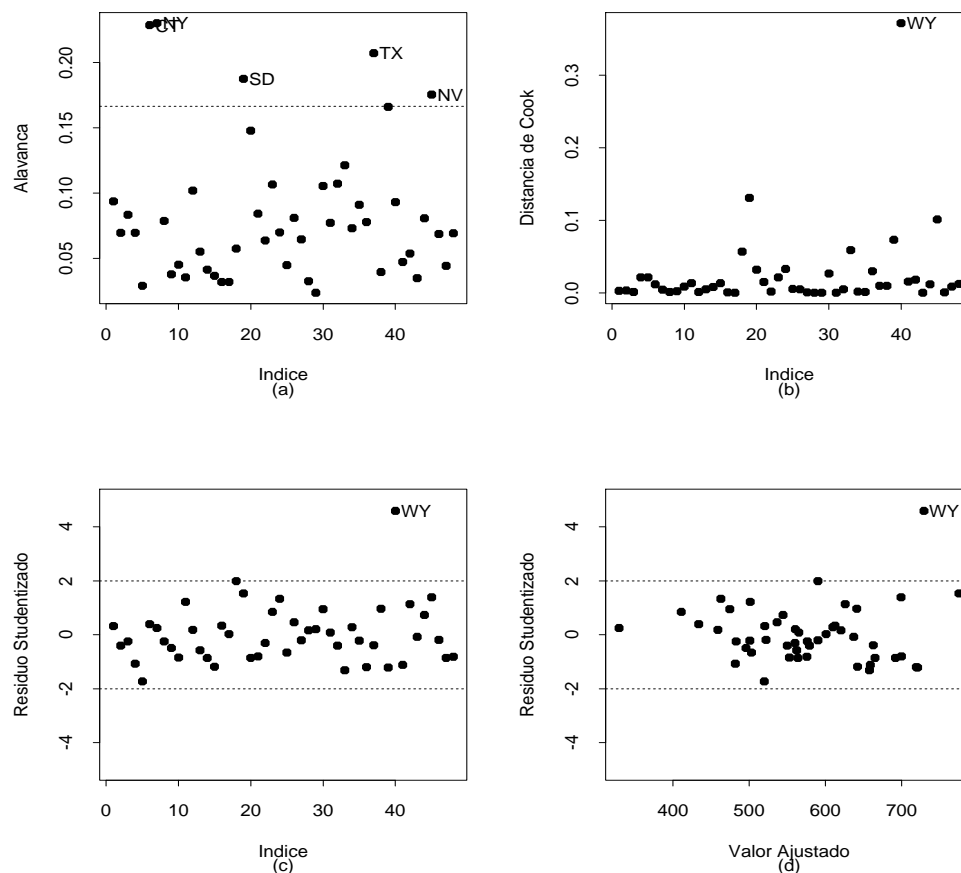


Figura 1.9: Gráficos de diagnóstico referentes ao modelo normal linear ajustado aos dados sobre consumo de combustível.

gráfico do estado de WY. O gráfico de envelope sem esse estado (Figura 1.10b) confirma essa suposição.

Analisando os dados referentes ao estado de WY notamos que o mesmo tem uma taxa de 7% (abaixo da média de 7,67%), uma renda percapita anual de USD 4345 (ligeiramente acima da média de USD 4241,83), uma proporção de motoristas licenciados de 0,672 (acima da média de 0,570), porém um consumo médio de combustível muito alto 968 (quando a média nacional era de 576,77). Talvez as longas distâncias do estado tenham obrigado os

motoristas a um consumo alto de combustível. A eliminação desse estado muda substancialmente algumas estimativas, embora não mude a inferência. A estimativa da variável licença cai 13,2%, a estimativa do intercepto aumenta 27,8%, o s^2 cai 17,1% e o R^2 aumenta 4,1%. As demais estimativas não sofrem grandes variações. A inclusão de alguma variável que leve em conta a densidade demográfica de cada estado, conforme sugerido por Gray (1989), poderia explicar melhor o estado de WY pelo modelo proposto.

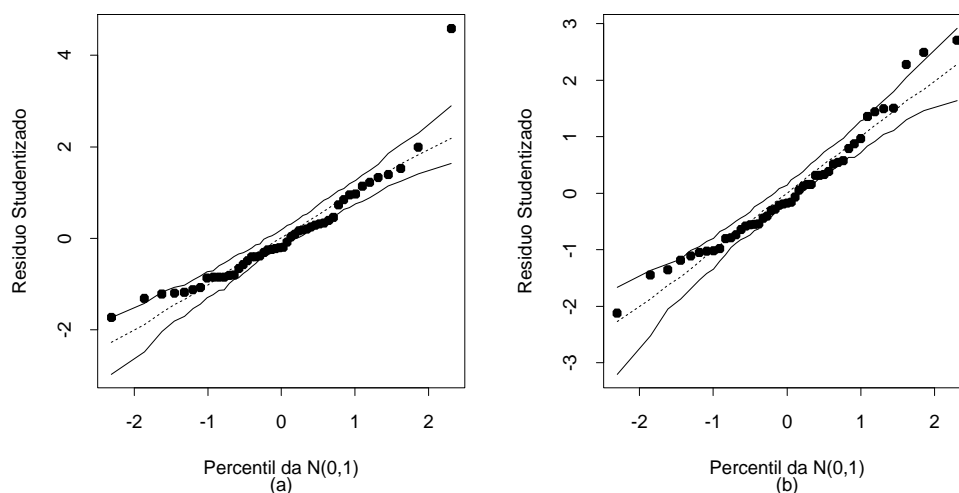


Figura 1.10: Gráficos normais de probabilidades com todos os pontos (a) e sem o estado de WY (b), referentes ao modelo normal linear ajustado aos dados sobre consumo de combustível.

1.12.6 Salário de executivos

Vamos considerar agora os dados descritos no arquivo **salary.dat** sobre o salário anual (em mil USD) de uma amostra aleatória de 220 executivos (145 homens e 75 mulheres)(Foster, Stine e Waterman, 1998, pgs. 180-188). O salário será relacionado com as seguintes variáveis explicativas: sexo (1:

masculino; 0: feminino), anos de experiência no cargo e posição na empresa (varia de 1 a 9), quanto maior o valor mais alta a posição. Apresentamos a seguir a comparação dos salários médios segundo o sexo.

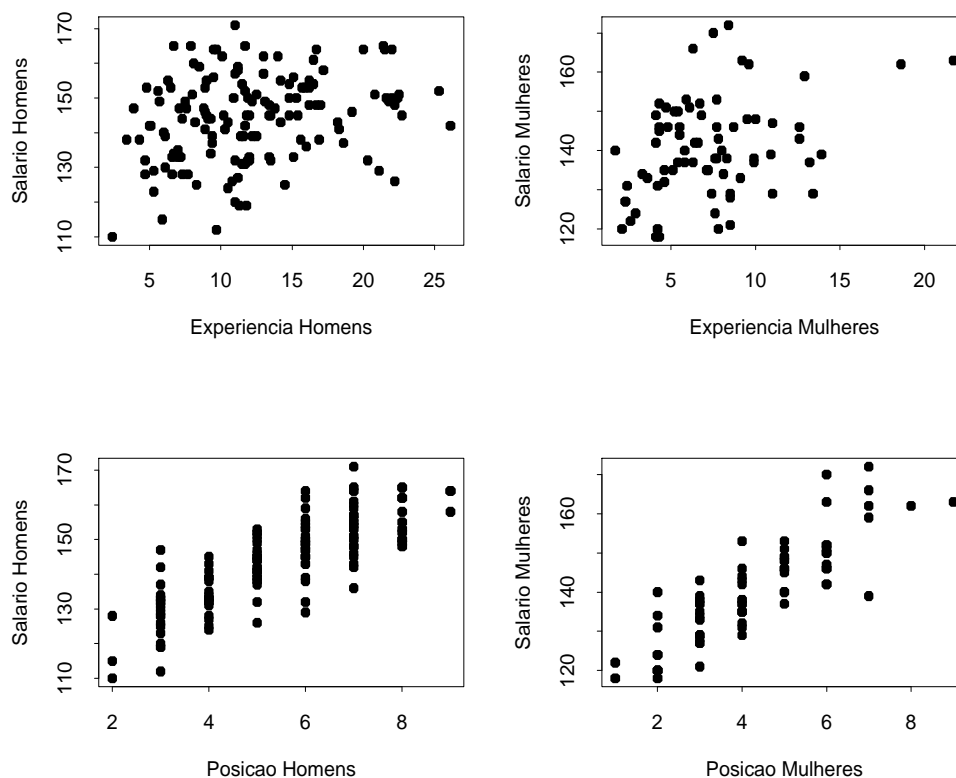


Figura 1.11: Gráficos de dispersão entre salário e as variáveis explicativas experiência e posição.

Sexo	Amostra	Média	E.Padrão
Masculino	145	144,11	1,03
Feminino	75	140,47	1,43

	Diferença	Teste-t	valor-P
Estimativa	3,64	2,06	0,04
E.Padrão	1,77		

O valor-P indica diferença ao nível de 5%. Ignorando as demais variáveis notamos que os executivos ganham em média mais do que as executivas.

Tabela 1.14

Estimativas dos parâmetros referentes ao modelo normal linear ajustado aos dados sobre salário de executivos.

Efeito	Estimativa	E/E.Padrão
Constante	115,262	82,25
Experiência	-0,472	-4,17
Sexo	-2,201	-2,04
Posição	6,710	21,46
R ²	0,71	
s	6,77	

Tabela 1.15

Testes F para avaliar a inclusão de cada interação de 1a. ordem no modelo normal linear ajustado aos dados sobre salário de executivos.

Interação	F-valor	valor-P
Sexo*Exper	1,615	0,20
Sexo*Posição	0,001	0,97
Exper*Posição	7,594	0,00

Pela Figura 1.11 notamos indícios de aumento do salário com o aumento da posição e com o aumento da experiência para ambos os sexos, sugerindo inicialmente um modelo linear do tipo

$$y_i = \alpha + \beta_1 \text{sexo}_i + \beta_2 \text{exper}_i + \beta_3 \text{posic}_i + \epsilon_i,$$

em que y_i denota o salário do i -ésimo executivo da amostra. Vamos supor ainda que ϵ_i são erros independentes e normalmente distribuídos de média zero e variância σ^2 .

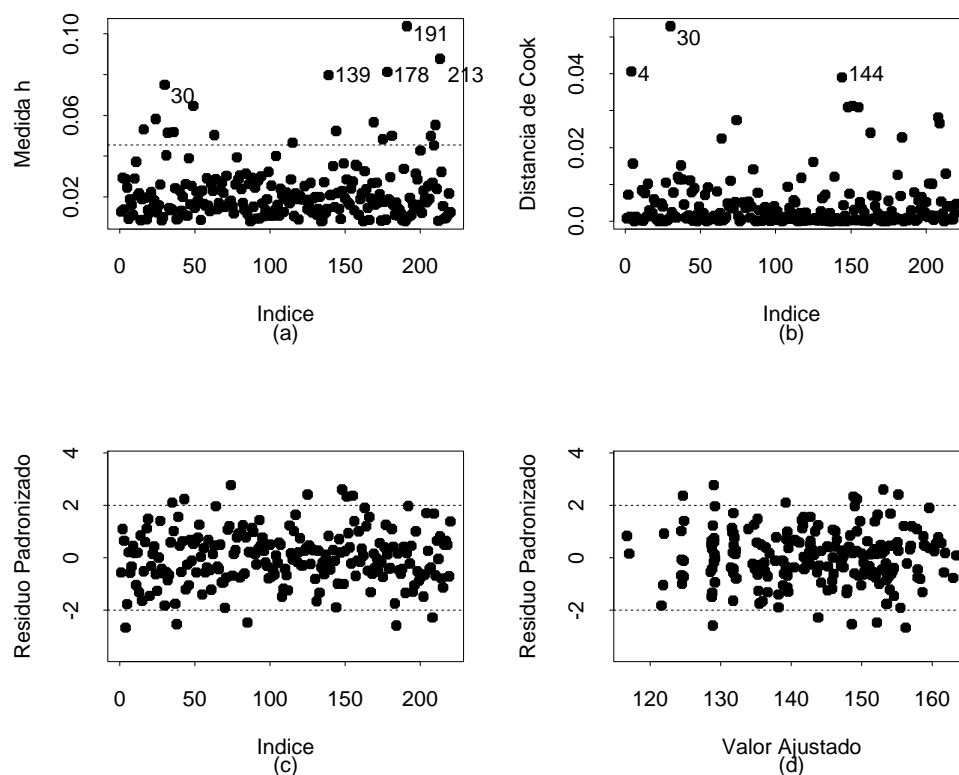


Figura 1.12: Gráficos de diagnóstico referentes ao modelo normal linear ajustado aos dados sobre salário de executivos.

Aplicando o método AIC nenhuma variável é retirada do modelo. As estimativas são apresentadas na Tabela 1.14. Notamos portanto que na presença de experiência e posição as mulheres ganham em média mais do que os homens. Quando essas variáveis são ignoradas, ocorre o contrário, como foi visto na análise descritiva. Experiência e posição são variáveis importantes e devem também permanecer no modelo. Vamos verificar a seguir se é possível incluirmos alguma interação de 1a. ordem. Essas interações são de interesse e também interpretáveis. Por exemplo, presença de interação entre os fatores sexo e experiência significa que a diferença entre os salários médios de executivos e executivas não é a mesma á medida que varia o tempo de

experiência.

Tabela 1.16

*Estimativas dos parâmetros referentes
ao modelo normal linear final ajustado
aos dados sobre salário de executivos.*

Efeito	Estimativa	E/E.Padrão
Constante	108,042	36,48
Experiência	0,336	1,07
Sexo	-2,811	-2,58
Posição	8,096	13,73
Exper*Posição	-0,135	-2,75
R ²	0,72	
s	6,67	

Na Tabela 1.15 são apresentados os valores da estatística F bem como o valor-P correspondente para testar a ausência das interações sexo*experiência, sexo*posição e experiência*posição, respectivamente, dado que estão no modelo constante + sexo + experiência + posição. Portanto, será incluída no modelo apenas a interação **experiência*posição**. As estimativas do modelo final com a interação experiência*posição são apresentadas na Tabela 1.16.

Na Figura 1.12 são apresentados gráficos de diagnóstico. Algumas observações aparecem como pontos de alavanca (Figura 1.12a), três observações são destacadas como possivelmente influentes (Figura 1.12b) e não há indícios de pontos aberrantes e heteroscedasticidade (Figuras 1.12c e 1.12d). Retirando as observações mais influentes, #4 e #30, notamos variações desproporcionais em algumas estimativas, embora não ocorram mudanças inferenciais. Esses pontos causam variações, respectivamente, de -14% e 11% na estimativa do coeficiente de sexo. A observação #4 é uma executiva com salário anual de USD 139 mil, posição 7 e 13,9 anos de experiência (salário

relativamente baixo para essa posição e experiência) e a observação #30 é de um executivo com salário anual de USD 110 mil, posição 2 e 2,4 anos de experiência (menor salário entre os executivos).

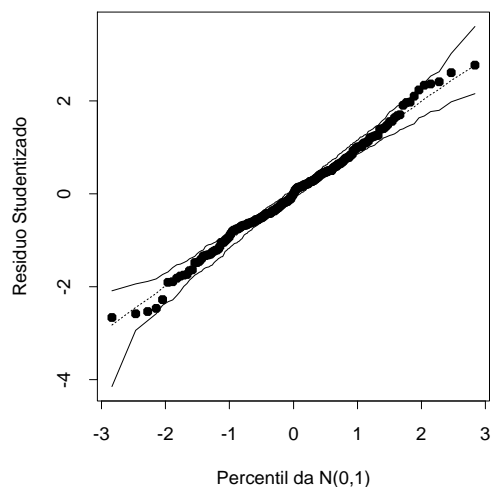


Figura 1.13: Gráfico normal de probabilidades referente ao modelo normal linear final ajustado aos dados sobre salário de executivos.

O modelo final ajustado é portanto dado por

$$\begin{aligned}\hat{y} = & 108,042 + 0,336 \times \text{exper} - 2,811 \times \text{sexo} \\ & + 8,096 \times \text{posic} - 0,135 \times \text{posic} \times \text{exper}.\end{aligned}$$

Por exemplo, desse modelo, qual o salário previsto para executivos com 5 anos de experiência e posição 4?

- Executiva: USD 139,406 mil
- Executivo: USD 136,595 mil.

O modelo ajustado para o grupo de mulheres é dado por

$$\hat{y} = 108,042 + 8,096 \times \text{posic} + (0,336 - 0,135 \times \text{posic}) \times \text{exper}.$$

Examinando a equação acima notamos que as executivas com posição alta e muita experiência tendem a ganhar menos do que executivas com posição alta e menos experiência. Isso quer dizer que aquelas executivas que permaneceram menos tempo no cargo tendem a ganhar mais do que aquelas que ficarem mais tempo no cargo. Mesmo resultado para os homens.

Notamos neste exemplo a importância da interação na interpretação dos resultados. Ignorando as variáveis posição e experiência temos que o salário médio dos executivos é significativamente maior do que o salário médio das executivas. Porém, quando essas variáveis entram no modelo ocorre o contrário, para uma mesma posição e mesma experiência as executivas ganham em média mais. Pela Figura 1.13 notamos que não há indícios de afastamentos da suposição de normalidade para os erros.

1.13 Exercícios

1. Se Y pertence à família exponencial de distribuições, então a função densidade ou função de probabilidades de Y pode ser expressa na forma

$$f(y; \theta, \phi) = \exp[\phi\{y\theta - b(\theta)\} + c(y; \phi)],$$

em que $b(\cdot)$ e $c(\cdot; \cdot)$ são funções diferenciáveis. Supondo ϕ conhecido seja $L(\theta) = \log f(y; \theta, \phi)$ o logaritmo da função de verossimilhança. Se $L(\theta)$ é pelo menos duas vezes diferenciável em θ mostre que

$$E\left(\frac{\partial L(\theta)}{\partial \theta}\right) = 0 \quad \text{e} \quad E\left\{\frac{\partial^2 L(\theta)}{\partial \theta^2}\right\} = -E\left\{\left(\frac{\partial L(\theta)}{\partial \theta}\right)^2\right\}.$$

2. Supondo agora ϕ desconhecido mostre que $E(U_\phi) = 0$ para as distribuições normal, normal inversa e gama.
3. Considere a distribuição estável cuja densidade é dada por

$$f(y; \theta, \phi) = a(y, \phi) \exp[\phi\{\theta(y+1) - \theta \log \theta\}],$$

em que $\theta > 0$, $-\infty < y < \infty$, $\phi^{-1} > 0$ é o parâmetro de dispersão e $a(\cdot, \cdot)$ é uma função normalizadora. Mostre que essa distribuição pertence à família exponencial. Encontre a função de variância. Obtenha a função desvio supondo uma amostra de n variáveis aleatórias independentes de médias μ_i e parâmetro de dispersão ϕ^{-1} .

4. Considere a seguinte função densidade de probabilidade:

$$f(y; \theta, \phi) = \frac{\phi a(y, \phi)}{\pi(1+y^2)^{1/2}} \exp[\phi\{y\theta + (1-\theta^2)^{1/2}\}],$$

em que $0 < \theta < 1$, $-\infty < y < \infty$, $\phi^{-1} > 0$ é o parâmetro de dispersão e $a(\cdot, \cdot)$ é uma função normalizadora. Mostre que essa distribuição pertence à família exponencial. Encontre a função de variância. Obtenha a função desvio supondo uma amostra de n variáveis aleatórias independentes de médias μ_i e parâmetro de dispersão ϕ^{-1} .

5. Mostre que a distribuição logarítmica, com função de probabilidades

$$f(y; \rho) = \rho^y / \{-y \log(1-\rho)\},$$

em que $y = 1, 2, \dots$ e $0 < \rho < 1$, pertence à família exponencial. Calcule μ e $V(\mu)$. Obtenha a função desvio supondo uma amostra de n variáveis aleatórias independentes de parâmetros ρ_i .

6. Supor que $Y_i \stackrel{\text{ind}}{\sim} \text{LG}(\rho_i)$, para $i = 1, \dots, n$, em que $\rho_i = e^\alpha / (1 + e^\alpha)$ e LG denota distribuição logarítmica. Mostre que a variância assintótica

de $\hat{\alpha}$ pode ser expressa na forma $\text{Var}(\hat{\alpha}) = \tau^2(\alpha)/ne^\alpha\{\tau(\alpha) - e^\alpha\}$, em que $\tau(\alpha) = (1 + e^\alpha)\log(1 + e^\alpha)$. Como fica a estatística do teste de escore para testar $H_0 : \alpha = 0$ contra $H_1 : \alpha \neq 0$? Qual a distribuição nula assintótica da estatística do teste?

7. (Paula e Cordeiro, 1986). Suponha o MLG em que $Y_i \sim \text{FE}(\mu_i, \phi)$ e parte sistemática dada por $g(\mu_i; \lambda) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, com λ escalar desconhecido. Encontre as funções escore \mathbf{U}_β e U_λ , as funções de informação de Fisher $\mathbf{K}_{\beta\beta}$, $\mathbf{K}_{\beta\lambda}$ e $\mathbf{K}_{\lambda\lambda}$ e descreva o processo iterativo escore de Fisher para obter a estimativa de máxima verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \lambda)^T$. Como iniciar o processo iterativo? Sugestão de notação: $\boldsymbol{\Lambda} = \partial \boldsymbol{\eta} / \partial \lambda$, em que $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$.
8. Supor o modelo normal linear com parte sistemática dada por $\eta_i = \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2)$. Sabe-se que a correlação amostral entre x_1 e x_2 é dada por $\text{corr}(x_1, x_2) = \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) / (n-1)s_1s_2$, em que s_1 e s_2 são os respectivos desvios padrão amostrais de x_1 e x_2 . Calcule a correlação $\text{corr}(\hat{\beta}_1, \hat{\beta}_2)$. Discuta e tente explicar a relação entre as duas correlações. Use o fato de que $\det(\mathbf{X}^T \mathbf{X})^{-1} > 0$.
9. Suponha o modelo de análise de variância com erros normais

$$y_{ij} = \alpha + \beta_i + \epsilon_{ij},$$

em que $\epsilon_{ij} \sim N(0, \sigma^2)$, $i = 1, \dots, p$ e $j = 1, \dots, n_i$. Supor $\beta_1 = 0$. Mostre que $\text{Var}(r_{ij}) = \sigma^2(1 - 1/n_i)$.

10. Considere o modelo normal linear

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n,$$

em que ϵ_i são erros mutuamente independentes tais que $\epsilon_i \sim N(0, \sigma^2)$. Considere uma nova observação $y(\mathbf{z})$ (que não está na amostra) e que satisfaz $y(\mathbf{z}) = \mathbf{z}^T \boldsymbol{\beta} + \epsilon$, em que $\epsilon \sim N(0, \sigma^2)$. Mostre que um intervalo de confiança de coeficiente $1 - \alpha$ para $y(\mathbf{z})$ pode ser dado por

$$[\hat{y}(\mathbf{z}) \pm t_{n-p}(1 - \frac{\alpha}{2})s\{1 + \mathbf{z}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{z}\}^{1/2}],$$

em que $\hat{y}(\mathbf{z}) = \mathbf{z}^T \hat{\boldsymbol{\beta}}$, $t_{n-p}(1 - \frac{\alpha}{2})$ é o percentil $(1 - \frac{\alpha}{2})$ da distribuição t de Student com $n - p$ graus de liberdade e s^2 é o erro quadrático médio do modelo ajustado.

11. Suponha agora o modelo de regressão normal linear simples

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \dots, n.$$

Mostre a equivalência entre as estatísticas ξ_{RV} , ξ_W e ξ_{SR} para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$. Supor σ^2 conhecido.

12. Um critério tradicional para a validação de modelos de regressão normal linear é através da estatística PRESS, definida por $\hat{\Delta} = \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2$, em que $\hat{y}_{(i)} = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{(i)}$ denota o valor predito para a i -ésima observação quando esta não é considerada no ajuste. O critério é selecionar o ajuste com menor valor para $\hat{\Delta}$. Mostre que

$$\hat{\Delta} = \sum_{i=1}^n \left(\frac{r_i}{1 - h_{ii}} \right)^2,$$

em que $r_i = y_i - \hat{y}_i$ e $h_{ii} = \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i$.

13. Suponha duas populações normais com médias μ_1 e μ_2 , mesma variância, e que amostras independentes de tamanhos n_1 e n_2 foram, respectivamente, obtidas das duas populações. Para o modelo com parte sistemática $\mu_1 = \alpha + \beta$ e $\mu_2 = \alpha - \beta$, mostre que a estatística F para

testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$ pode ser expressa na forma simplificada

$$F = \frac{(n-2) \frac{n_1 n_2}{n} (\bar{y}_1 - \bar{y}_2)^2}{\sum (y_i - \bar{y})^2 - \frac{n_1 n_2}{n} (\bar{y}_1 - \bar{y}_2)^2},$$

em que $\bar{y}, \bar{y}_1, \bar{y}_2$ são as respectivas médias amostrais.

14. (Paula e Sen, 1995). Suponha um MLG com ligação canônica e parte sistemática dada por $g(\mu_{1j}) = \alpha_1 + \beta x_j$ e $g(\mu_{2j}) = \alpha_2 + \beta x_j$, $j = 1, \dots, r$. Interprete esse tipo de modelo. Obtenha a matriz \mathbf{X} correspondente. Como fica o teste de escore para testar $H_0 : \beta = 0$? O que significa testar H_0 ?
15. Sejam Y_{ij} , $i = 1, 2, 3$ e $j = 1, \dots, m$, variáveis aleatórias mutuamente independentes pertencentes à família exponencial tais que $E(Y_{ij}) = \mu_{ij}$, $\text{Var}(Y_{ij}) = V_{ij}\phi^{-1}$ e parte sistemática dada por $g(\mu_{1j}) = \alpha$, $g(\mu_{2j}) = \alpha + \Delta$ e $g(\mu_{3j}) = \alpha - \Delta$. Responda às seguintes questões:
 - (i) como fica a matriz modelo \mathbf{X} ?
 - (ii) O que significa testar $H_0 : \Delta = 0$? Qual a distribuição nula assintótica das estatísticas ξ_{RV} , ξ_W e ξ_{SR} ?
 - (iii) Calcular a variância assintótica de $\hat{\Delta}$, $\text{Var}(\hat{\Delta})$.
 - (iv) Mostre que a estatística do teste de escore para testar $H_0 : \Delta = 0$ contra $H_1 : \Delta \neq 0$ fica dada por

$$\xi_{SR} = \frac{\phi m (\bar{y}_2 - \bar{y}_3)^2}{2\hat{V}_0}.$$

16. Mostre (1.14) e (1.15). Use o seguinte resultado de álgebra linear:

$$(\mathbf{A} + \mathbf{U}\mathbf{V}^T)^{-1} = \mathbf{A}^{-1} - \frac{(\mathbf{A}^{-1}\mathbf{U})(\mathbf{V}^T\mathbf{A}^{-1})}{1 + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}},$$

em que \mathbf{A} é uma matriz não singular e \mathbf{U} e \mathbf{V} são vetores coluna. Mostre primeiro que $\mathbf{X}^T \mathbf{\Delta} \mathbf{X} = \mathbf{X}^T \mathbf{X} - (1-\delta) \mathbf{x}_i \mathbf{x}_i^T$ e $\mathbf{X}^T \mathbf{\Delta} \mathbf{y} = \mathbf{X}^T \mathbf{y} - (1-\delta) \mathbf{x}_i y_i$, em que $\mathbf{\Delta}$ é uma matriz de 1's com δ na i -ésima posição.

17. (Cook e Weisberg, 1982). Suponha o modelo de regressão dado em (1.12). Mostre que $\hat{\gamma} \sim N(\gamma, \sigma^2/(1 - h_{ii}))$. Mostre também que, sob a hipótese $H_1 : \gamma \neq 0$, a estatística F tem uma distribuição $F_{1, (n-p-1)}(\lambda)$, em que $\lambda = \frac{1}{2} \frac{\gamma^2(1-h_{ii})}{\sigma^2}$ é o parâmetro de não centralidade. Comente sobre o poder desse teste para $0 \leq h_{ii} < 1$. Use o resultado: se $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$ então $\mathbf{y}^T \mathbf{y} / \sigma^2 \sim \chi_n^2(\lambda)$, em que $\lambda = \frac{1}{2} \boldsymbol{\mu}^T \boldsymbol{\mu} / \sigma^2$.
18. (Pregibon, 1982). Mostre que a estatística de escore para testar que o i -ésimo ponto é aberrante num MLG com ϕ conhecido e parte sistemática $g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ é dada por $t_{S_i}^2$, em que

$$t_{S_i} = \frac{\sqrt{\phi}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - \hat{h}_{ii})}},$$

sendo $\hat{\mu}_i$, \hat{V}_i e $\hat{h}_{ii} = \hat{\omega}_i \mathbf{x}_i^T (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i$ avaliados em $\hat{\boldsymbol{\beta}}$. Qual a distribuição nula assintótica de $t_{S_i}^2$? Como seria interpretado o gráfico de $t_{S_i}^2$ contra a ordem das observações? Sugestão : chame $\boldsymbol{\eta} = \mathbf{X} \boldsymbol{\beta} + \gamma \mathbf{z}$, em que \mathbf{z} é um vetor $n \times 1$ de zeros com 1 na i -ésima posição, calcule $\text{Var}(\hat{\gamma})$ e U_γ e teste $H_0 : \gamma = 0$ contra $H_1 : \gamma \neq 0$.

19. Mostrar que a expressão para AIC no modelo normal linear com σ^2 desconhecido pode ser escrita na forma equivalente

$$\text{AIC} = n \log\{D(\mathbf{y}; \hat{\boldsymbol{\mu}})/n\} + 2p,$$

em que $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2$.

20. No arquivo **reg3.dat** são descritas as seguintes variáveis referentes a 50 estados norte-americanos: (i) **estado** (nome do estado), (ii) **pop** (população estimada em julho de 1975), (iii) **percap** (renda percapita em 1974 em USD), (iv) **analf** (proporção de analfabetos em 1970), (v) **expvida** (expectativa de vida em anos 1969-70), (vi) **crime** (taxa de criminalidade por 100000 habitantes 1976), (vii) **estud** (porcentagem de estudantes que concluem o segundo grau 1970), (viii) **ndias** (número de dias do ano com temperatura abaixo de zero grau Celsius na cidade mais importante do estado) e (ix) **area** (área do estado em milhas quadradas).

O objetivo do estudo é tentar explicar a variável **expvida** usando um modelo de regressão normal linear dadas as variáveis explicativas **percap**, **analf**, **crime**, **estud**, **ndias** e **dens**, em que $\text{dens} = \text{pop}/\text{area}$.

Inicialmente faça uma análise descritiva dos dados, por exemplo com boxplots de cada uma das variáveis que serão consideradas no estudo e com diagramas de dispersão com as respectivas tendências entre as variáveis explicativas e a variável resposta. Comente essa parte descritiva. Posteriormente, ajuste o modelo de regressão normal linear com todas as variáveis explicativas e através do método **AIC** faça uma seleção de variáveis. Uma vez selecionado o modelo faça uma análise de diagnóstico e apresente as interpretações dos coeficientes estimados do modelo final.

21. (Neter et al., 1996, p. 449). No arquivo **vendas.dat** são descritas informações a respeito das vendas no ano anterior de um tipo de telhado de madeira em 26 filiais de uma rede de lojas de construção. As variáveis estão colocadas na seguinte ordem: (i) **telhados**, total de telhados vendidos (em mil metros quadrados), (ii) **gastos**, gastos

pela loja com promoções do produto (em mil USD), (iii) **clientes**, número de clientes cadastrados na loja (em milhares), (iv) **marcas**, número de marcas concorrentes do produto e (v) **potencial**, potencial da loja (quanto maior o valor maior o potencial). Um dos objetivos do estudo com esse conjunto de dados é tentar prever o número esperado de telhados vendidos dadas as variáveis explicativas. Faça inicialmente uma análise descritiva construindo, por exemplo, os diagramas de dispersão de cada variável explicativa contra a variável resposta **telhados**. Calcule também as correlações entre as variáveis. Use os métodos **stepwise** e **AIC** para selecionar um modelo de regressão normal linear. Se o modelo selecionado for diferente pelos dois métodos, adote algum critério para escolher um dos modelos. Interprete os coeficientes estimados do modelo selecionado. Faça uma análise de diagnóstico para verificar se existem afastamentos sérios das suposições feitas para o modelo e se existem observações discrepantes.

22. (Wood, 1973). No arquivo **reg4.dat** estão os dados referentes à produção de gasolina numa determinada refinaria segundo três variáveis observadas durante o processo e uma quarta variável que é uma combinação das três primeiras. A resposta é o número de octanas do produto produzido. A octanagem é a propriedade que determina o limite máximo que a gasolina, junto com o ar, pode ser comprimida na câmara de combustão do veículo sem queimar antes de receber a centilha vinda das velas. As melhores gasolinas têm uma octanagem alta. Em grandes refinarias, o aumento de um octana na produção de gasolina pode representar um aumento de alguns milhões de dólares no custo final da produção. Assim, torna-se importante o controle dessa variável durante o processo de produção. Use o método **AIC** para selecionar as

variáveis explicativas significativas. Faça uma análise de diagnóstico com o modelo selecionado. Comente.

23. (Narula e Stangenhau, 1988, pgs. 31-33). No arquivo **imoveis.dat** são apresentados dados relativos a uma amostra de 27 imóveis. Na ordem são apresentados os valores das seguintes variáveis: (i) imposto do imóvel (em 100 USD), (ii) área do terreno (em 1000 pés quadrados), (iii) área construída (em 1000 pés quadrados), (iv) idade da residência (em anos) e (v) preço de venda do imóvel (em 1000 USD). Ajuste um modelo normal linear do preço de venda contra as demais variáveis. Use o método AIC para selecionar as variáveis explicativas. Faça uma análise de diagnóstico com o modelo selecionado. Interprete os coeficientes estimados.
24. (Ryan e Joiner, 1994, p. 299). No arquivo **trees.dat** é apresentado um conjunto de dados que tem sido analisado sob diversos pontos de vista por vários pesquisadores (ver, por exemplo, Jørgensen, 1989). As variáveis observadas são o diâmetro (d), a altura (h) e o volume (v) de uma amostra de 31 cerejeiras numa floresta do estado da Pensilvânia, EUA. A relação entre diâmetro, altura e volume de uma árvore depende da forma da mesma e pode-se considerar duas possibilidades

$$v = \frac{1}{4}\pi d^2 h$$

para forma cilíndrica e

$$v = \frac{1}{12}\pi d^2 h$$

para forma cônica. Em ambos os casos a relação entre $\log v$, $\log d$ e $\log h$ é dada por

$$\log v = a + b \log d + c \log h.$$

Supor inicialmente um modelo linear em que $\epsilon \sim N(0, \sigma^2)$. Faça uma análise de diagnóstico e verifique se é possível melhorar o modelo, por exemplo incluindo algum termo quadrático.

25. (Ruppert, 2004). No arquivo **capm.dat** estão os seguintes dados: Tbill (taxa de retorno livre de risco), retorno Microsoft, SP500 (retorno do mercado), retorno GE e retorno FORD de janeiro de 2002 a abril de 2003. Todos os retornos são diários e estão em porcentagem. Faça inicialmente os diagramas de dispersão entre os excessos de retorno ($y_{rt} - r_{ft}$) de cada uma das empresas Microsoft, GE e FORD e os excessos de retorno do mercado ($r_{mt} - r_{ft}$), em que y_{rt} denota o retorno da ação da empresa, r_{mt} é o retorno do mercado e r_{ft} indica a taxa livre de risco durante o t -ésimo período. Posteriormente, ajuste o seguinte modelo de regressão:

$$y_{rt} - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + \epsilon_t,$$

em que $\epsilon_t \sim N(0, \sigma^2)$. Verifique a significância do parâmetro α e compare e interprete as estimativas intervalares para β . Faça uma análise de diagnóstico para cada modelo ajustado.

26. O conjunto de dados descrito na tabela abaixo refere-se a um estudo cujo objetivo foi tentar prever o preço de venda de um imóvel (em mil USD) dada a área total (em mil pés quadrados) numa região de Eugene, EUA (Gray, 1989). Esses dados estão armazenados no arquivo externo **reg1.dat**.

Tente inicialmente ajustar uma regressão normal linear para explicar o preço dada a renda. Faça uma análise de diagnóstico e proponha algum modelo alternativo (se for o caso) a fim de reduzir as eventuais

influências de observações discrepantes bem como afastamentos de suposições feitas para o modelo. Interprete as estimativas obtidas para os coeficientes do modelo proposto.

Área	800	950	910	950	1200	1000	1180	1000
Preço	30,6	31,5	33,3	45,9	47,4	48,9	51,6	53,1
Área	1380	1250	1500	1200	1600	1650	1600	1680
Preço	54,0	54,3	55,2	55,2	56,7	57,9	58,5	59,7
Área	1500	1780	1790	1900	1760	1850	1800	1700
Preço	60,9	60,9	62,4	63,0	64,5	66,0	66,3	67,5
Área	1370	2000	2000	2100	2050	1990	2150	2050
Preço	68,4	68,4	68,7	69,6	70,5	74,7	75,0	75,3
Área	2200	2200	2180	2250	2400	2350	2500	2500
Preço	79,8	80,7	80,7	83,4	84,0	86,1	87,0	90,3
Área	2500	2500	2680	2210	2750	2500	2400	3100
Preço	96,0	101,4	105,9	111,3	112,5	114,0	115,2	117,0
Área	2100	4000						
Preço	129,0	165,0						

Capítulo 2

Modelos para Dados Positivos Assimétricos

2.1 Introdução

A classe de modelos para a análise de dados positivos assimétricos é bastante ampla incluindo distribuições conhecidas para os erros, tais como gama, normal inversa, Weibull, Pareto, log-normal e Birnbaum-Saunders (ver, por exemplo, Leiva, Barros e Paula, 2009). Essas distribuições têm sido particularmente aplicadas na análise de tempos de sobrevivência (ou duração) com forte ênfase nas áreas médica e de engenharia (ver, por exemplo, Lawless, 2003). Todavia, dados positivos assimétricos têm sido também comuns em outras áreas do conhecimento, como por exemplo pesca, meteorologia, finanças, seguros e atuária (ver, por exemplo, Boland, 2007; Jong e Heller, 2008). Um componente importante no estudo de dados de sobrevivência é a possibilidade de incorporação nas análises de observações para as quais não foi possível observar a falha (dados censurados). Com os recentes avanços tecnológicos ocorridos principalmente na fabricação de equipamentos, os tempos até a ocorrência de falhas estão ficando cada vez mais longos, aumentando assim a porcentagem de dados censurados. Isso também pode ser notado na

área médica com os avanços nos tratamentos e medicamentos. Todavia, a inclusão de dados censurados nos modelos envolve um tipo de análise mais específica que está além das metodologias discutidas neste texto. Sugerimos ao leitor mais interessado consultar textos de análise de dados de sobrevivência, como por exemplo os livros de Cox e Oakes (1978), Lawless (2003), Collett (2003), Colosimo e Giolo (2006) e Leiva, Barros e Paula (2009). Portanto, neste capítulo trataremos apenas dados positivos assimétricos não censurados sob modelos com resposta gama e normal inversa, os quais já foram introduzidos no Capítulo 1.

2.2 Distribuição gama

Conforme assumido na Seção 1.2.1 vamos supor que Y é uma variável aleatória com distribuição gama de média μ e coeficiente de variação $\phi^{-1/2}$, denotamos $Y \sim G(\mu, \phi)$, e cuja função densidade é expressa na forma

$$\begin{aligned} f(y_i; \mu, \phi) &= \frac{1}{\Gamma(\phi)} \left(\frac{\phi y}{\mu} \right)^\phi \exp \left(-\frac{\phi y}{\mu} \right) d(\log y) \\ &= \exp[\phi\{(-y/\mu) - \log \mu\} - \log \Gamma(\phi) + \phi \log(\phi y) - \log y], \end{aligned}$$

em que $y > 0$, $\phi > 0$, $\mu > 0$ e $\Gamma(\phi) = \int_0^\infty t^{\phi-1} e^{-t} dt$ é a função gama. Na Figura 2.1 temos a densidade da distribuição gama variando o parâmetro de dispersão para μ fixado. Podemos notar que à medida que ϕ aumenta a distribuição gama fica mais simétrica em torno da média. Pode ser mostrado que à medida que ϕ aumenta Y se aproxima de uma distribuição normal de média μ e variância $\mu^2 \phi^{-1}$. Portanto, a distribuição gama torna-se atrativa para o estudo de variáveis aleatórias assimétricas e também simétricas em que a variância depende de forma quadrática da média. Os momentos centrais

de Y são expressos na seguinte forma:

$$E(Y - \mu)^r = \frac{(r-1)! \mu^r}{\phi^{(r-1)}},$$

para $r = 1, 2, \dots$. Assim, expandindo $\log Y$ em série de Taylor em torno de μ até 2ª ordem obtemos

$$\log Y \cong \log \mu + \frac{1}{\mu}(Y - \mu) - \frac{1}{2\mu^2}(Y - \mu)^2.$$

Portanto, para ϕ grande temos que

$$\begin{aligned} E(\log Y) &\cong \log \mu - \frac{1}{2\mu^2} E(Y - \mu)^2 \\ &= \log \mu - \frac{1}{2\mu^2} \frac{\mu^2}{\phi} \\ &= \log \mu - (2\phi)^{-1} \text{ e} \\ \text{Var}(\log Y) &\cong \phi^{-1}. \end{aligned}$$

Ou seja, a transformação $\log Y$ estabiliza a variância à medida que o coeficiente de variação de Y fica pequeno. Uma outra transformação dada por $3\{(Y/\mu)^{1/3} - 1\}$ se aproxima da distribuição normal padrão no caso gama (vide McCullagh e Nelder, 1989, p. 289).

A função de sobrevivência e a função de risco são quantidades usuais na análise de dados de sobrevivência sendo definidas, respectivamente, por

$$\begin{aligned} S(t) &= Pr\{Y \geq t\} \text{ e} \\ h(t) &= \lim_{\delta \rightarrow 0} \frac{Pr\{t \leq Y < t + \delta | Y \geq t\}}{\delta}. \end{aligned}$$

Em particular, temos que a função de risco pode ser expressa na forma $h(t) = f(t)/S(t)$ com $f(y)$ denotando a função densidade de Y . No caso

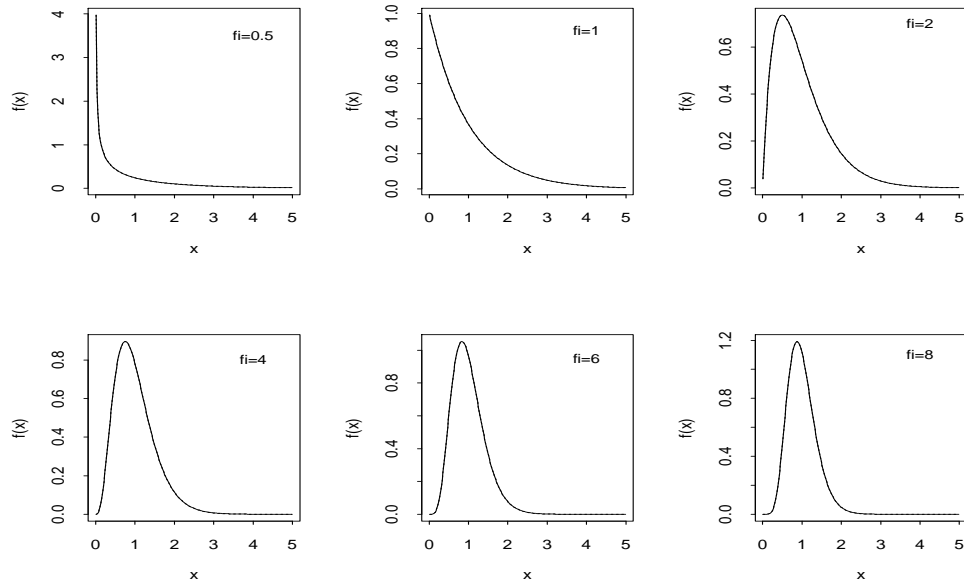


Figura 2.1: Densidades da distribuição gama para alguns valores do parâmetro de dispersão e supondo $\mu = 1$.

da distribuição gama de média μ e parâmetro de dispersão ϕ^{-1} a função de sobrevivência é expressa (ver, por exemplo, Collett, 2003, pgs. 197-198) na forma

$$S(t) = 1 - I_{\lambda t}(\phi),$$

em que $I_{\lambda t}(\phi)$ é a função gama incompleta, dada por

$$I_{\lambda t}(\phi) = \frac{1}{\Gamma(\phi)} \int_0^{\lambda t} u^{\phi-1} e^{-u} du,$$

com $\lambda = \phi/\mu$. A função de risco $h(t)$ para a distribuição gama é crescente para $\phi > 1$ e decrescente para $\phi < 1$. Em particular, quando $t \rightarrow \infty$ temos que $h(t) \rightarrow \lambda$.

2.3 Modelos com resposta gama

Vamos supor que Y_1, \dots, Y_n são variáveis aleatórias independentes tais que $Y_i \sim G(\mu_i, \phi)$. Ou seja, estamos assumindo que essas variáveis possuem médias diferentes e mesmo coeficiente de variação $\phi^{-1/2}$. Ademais, vamos supor que $g(\mu_i) = \eta_i$ com $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ contendo valores de variáveis explicativas e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ sendo o vetor de parâmetros de interesse. As ligações mais usadas no caso gama são identidade ($\mu_i = \eta_i$), logarítmica ($\log \mu_i = \eta_i$) e recíproca ($\mu_i = \eta_i^{-1}$), esta última sendo a ligação canônica. O processo iterativo para estimação de $\boldsymbol{\beta}$, como foi visto na Seção 1.6.1, é dado por

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)},$$

$m = 0, 1, \dots$, variável dependente modificada $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$, $\mathbf{y} = (y_1, \dots, y_n)^T$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$, $\mathbf{V} = \text{diag}\{\mu_1, \dots, \mu_n\}$ e $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$ com $\omega_i = (d\mu_i/d\eta_i)^2/\mu_i$.

É interessante notarmos que sob ligação logarítmica os pesos do processo iterativo para obtenção de $\hat{\boldsymbol{\beta}}$ ficam dados por $\omega_i = \frac{\mu_i^2}{\mu_i^2} = 1$, de modo que o processo iterativo assume a forma simplificada

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{z}^{(m)},$$

em que $\mathbf{z} = (z_1, \dots, z_n)^T$ com $z_i = \eta_i = (y_i - \mu_i)/\mu_i$ e $\mu_i = \exp(\eta_i)$. A variância assintótica de $\hat{\boldsymbol{\beta}}$ fica dada por $\text{Var}(\hat{\boldsymbol{\beta}}) = \phi^{-1} (\mathbf{X}^T \mathbf{X})^{-1}$. Em particular, se as colunas da matriz \mathbf{X} são ortogonais, isto é $\mathbf{X}^T \mathbf{X} = \mathbf{I}_p$, em que \mathbf{I}_p é a matriz identidade de ordem p , então $\text{Var}(\hat{\beta}_j) = \phi^{-1}$ e $\text{Cov}(\hat{\beta}_j, \hat{\beta}_\ell) = 0$, para $j \neq \ell$, ou seja, β_j e β_ℓ são assintoticamente independentes.

Portanto, a ligação logarítmica tem um atrativo especial de possibilitar o desenvolvimento de experimentos ortogonais como são bem conhecidos em

modelos de regressão normal linear. Podemos escolher formas apropriadas para a matriz \mathbf{X} , de modo que $\mathbf{X}^T\mathbf{X} = \mathbf{I}_p$, e assim obtermos estimativas mutuamente independentes e de variância constante para os coeficientes do preditor linear. Myers, Montgomery e Vining (2002, Cap.6) discutem experimentos ortogonais em MLGs e apresentam alguns exemplos. As ligações identidade ($\mu = \eta$), raiz quadrada ($\sqrt{\mu} = \eta$) e arco-seno ($\sin^{-1}\sqrt{\mu} = \eta$) produzem o mesmo efeito em MLGs com resposta normal, Poisson e binomial, respectivamente.

Se aplicarmos para ϕ suficientemente grande a transformação logarítmica na resposta e ajustarmos $E(\log Y_i) = \mathbf{x}_i^T \boldsymbol{\beta}$, estaremos fazendo $\log \mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + 2\phi^{-1}$, ou seja, a menos da constante $2\phi^{-1}$ estaremos obtendo as mesmas estimativas para $\boldsymbol{\beta}$ de um modelo com resposta gama e ligação logarítmica.

2.3.1 Qualidade do ajuste

Como foi visto na Seção 1.4 o desvio de um modelo gama é dado por $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ em que

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \{\log(\hat{\mu}_i/y_i) + (y_i - \hat{\mu}_i)/\hat{\mu}_i\}, \quad (2.1)$$

com $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$ e $\hat{\eta}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$. Pode ser mostrado facilmente para ligação logarítmica que o termo $\sum_{i=1}^n (y_i - \hat{\mu}_i)/\hat{\mu}_i = 0$ se a parte sistemática η_i contém um intercepto. Nesse caso, a função desvio fica dada por $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\phi \sum_{i=1}^n \log(\hat{\mu}_i/y_i)$. Como ϕ é desconhecido devemos estimá-lo, por exemplo, através de máxima verossimilhança que equivale a resolvermos a seguinte equação:

$$2n\{\log \hat{\phi} - \psi(\hat{\phi})\} = D(\mathbf{y}; \hat{\boldsymbol{\mu}}),$$

em que $\psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$ é a função digama (vide Seção 1.6.2), ou então utilizarmos a estimativa consistente $\hat{\phi}^{-1} = (n-p)^{-1} \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / \hat{\mu}_i^2$ que

será discutida no Capítulo 5. Supondo que o modelo postulado está correto temos, para ϕ grande, que o desvio $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ segue distribuição qui-quadrado com $(n - p)$ graus de liberdade. Assim, valores altos para o desvio podem indicar inadequação do modelo ou falta de ajuste.

Quando todas as observações são positivas o desvio $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ deve ser utilizado para avaliar a qualidade do ajuste e estimação de ϕ . Contudo, se pelo menos uma observação for igual a zero $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ torna-se inapropriado. A estimativa para ϕ nesse caso fica indeterminada. Como foi mencionado na Seção 1.4, McCullagh e Nelder (1989) sugerem substituir $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ por

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\phi\{C(\mathbf{y}) + \sum_{i=1}^n \log \hat{\mu}_i + \sum_{i=1}^n y_i / \hat{\mu}_i\},$$

em que $C(\mathbf{y})$ é uma função arbitrária, porém limitada. Se a parte sistemática do modelo contém um intercepto o desvio acima fica dado por $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\phi\{n + C(\mathbf{y}) + \sum_{i=1}^n \log \hat{\mu}_i\}$.

2.3.2 Técnicas de diagnóstico

O resíduo componente do desvio padronizado assume para os modelos gama a forma

$$t_{D_i} = \pm \frac{\sqrt{2\phi}}{\sqrt{1 - \hat{h}_{ii}}} \{\log(\hat{\mu}_i / y_i) - (y_i - \hat{\mu}_i) / \hat{\mu}_i\}^{1/2},$$

em que $y_i > 0$ e \hat{h}_{ii} é o i -ésimo elemento da diagonal principal da matriz $\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$ com $\omega_i = (d\mu_i / d\eta_i)^2 / \mu_i^2$. Em particular quando há um intercepto em η_i o resíduo componente do desvio t_{D_i} assume a forma reduzida

$$t_{D_i} = \pm \frac{\sqrt{2\phi}}{\sqrt{1 - \hat{h}_{ii}}} \{\log(\hat{\mu}_i / y_i)\}^{1/2}.$$

Estudos de simulação indicam que o resíduo t_{D_i} se aproxima da normalidade, particularmente para ϕ grande.

Quando a i -ésima observação é excluída a distância de Cook aproximada fica dada por

$$LD_i = \frac{\phi \hat{h}_{ii}}{(1 - \hat{h}_{ii})^2} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2}.$$

Gráficos de t_{D_i} e \hat{h}_{ii} contra os valores ajustados $\hat{\mu}_i$ como também gráficos de índices de LD_i são recomendados para a análise de diagnóstico.

2.4 Aplicações

2.4.1 Comparação de cinco tipos de turbina de avião

Apresentamos na Tabela 2.1 (ver Lawless 1982, p. 201) os resultados de um experimento conduzido para avaliar o desempenho de cinco tipos de turbina de alta velocidade para motores de avião. Foram considerados dez motores de cada tipo nas análises e foi observado para cada um o tempo (em unidades de milhões de ciclos) até a perda da velocidade. Esses dados estão disponíveis no arquivo **turbina.dat**.

Tabela 2.1
Tempo até a perda da velocidade de cinco tipos de turbina de avião.

Tipo de turbina				
Tipo I	Tipo II	Tipo III	Tipo IV	Tipo V
3,03	3,19	3,46	5,88	6,43
5,53	4,26	5,22	6,74	9,97
5,60	4,47	5,69	6,90	10,39
9,30	4,53	6,54	6,98	13,55
9,92	4,67	9,16	7,21	14,45
12,51	4,69	9,40	8,14	14,72
12,95	5,78	10,19	8,59	16,81
15,21	6,79	10,71	9,80	18,39
16,04	9,37	12,58	12,28	20,84
16,84	12,75	13,41	25,46	21,51

Denotaremos por T_{ij} o tempo até a perda da velocidade para o j -ésimo motor de tipo i , $i = 1, \dots, 5$ e $j = 1, \dots, 10$. Na tabela abaixo são apresentadas as médias, desvios padrão e coeficientes de variação amostrais para os cinco tipos de turbina.

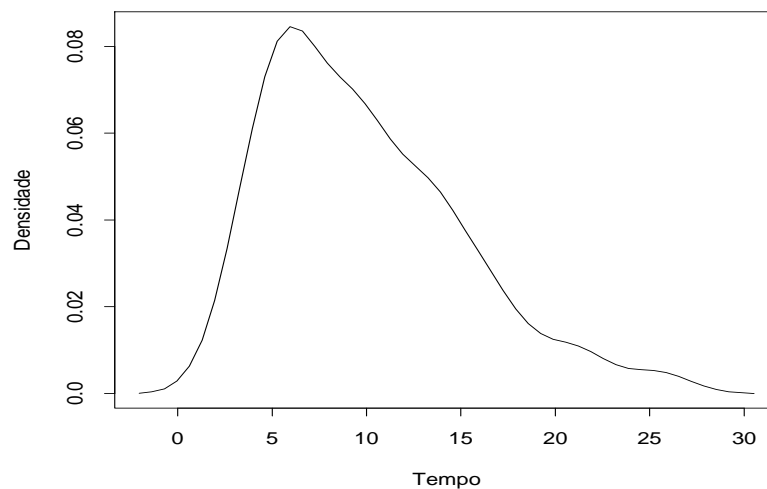


Figura 2.2: Densidade aproximada para o tempo até a perda da velocidade para todos os tipos de turbina de avião.

Estatística	Tipo I	Tipo II	Tipo III	Tipo IV	Tipo V
Média	10,69	6,05	8,64	9,80	14,71
D.Padrão	4,82	2,91	3,29	5,81	4,86
C. Variação	45,09%	48,10%	38,08%	59,29%	33,04%

Ignorando o tipo de turbina temos na Figura 2.2 a densidade aproximada para o tempo até a perda da velocidade. Se assumirmos que $T_{ij} \sim G(\mu, \phi)$ obtemos as estimativas de máxima verossimilhança $\hat{\mu} = 9,98$ (0,73) e $\hat{\phi} =$

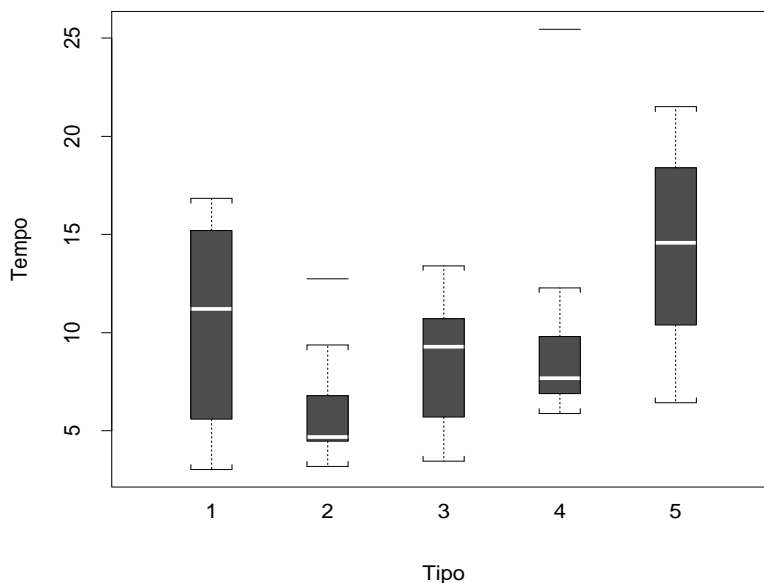


Figura 2.3: Boxplots sobre desempenho dos quatro tipos de turbina de avião.

4,01 (0,77), confirmando pela estimativa de ϕ a assimetria à direita para a distribuição do tempo até a perda da velocidade. Contudo, pelos boxplots correspondentes aos tempos dos cinco grupos (ver Figura 2.3) notamos que as médias parecem diferir com o tipo de turbina, enquanto os coeficientes de variação parecem variar menos do que os desvios padrão. Isso sugere que uma distribuição gama de médias diferentes e coeficiente de variação constante pode ser apropriada para explicar o tempo médio até a perda da velocidade.

Vamos assumir então que T_{ij} segue uma distribuição gama de média μ_i e parâmetro de dispersão ϕ^{-1} . A fim de facilitarmos as interpretações dos resultados ou mesmo fazermos comparações com o modelo normal linear, vamos propor um modelo gama com ligação identidade, sendo a parte sistemática dada por

$$\mu_i = \mu + \beta_i,$$

em que $\beta_1 = 0$. Para ajustarmos o modelo no R devemos fazer o seguinte:

```
turbina = scan("turbina.dat", list(tipo=0, tempo=0))
attach(turbina)
tipo = factor(tipo)
fit1.turbina = glm(tempo ~ tipo, family=Gamma(link=identity))
summary(fit1.turbina).
```

As estimativas de máxima verossimilhança ficam dadas por $\hat{\mu} = 10,693 (1,543)$, $\hat{\beta}_2 = -4,643 (1,773)$, $\hat{\beta}_3 = -2,057 (1,983)$, $\hat{\beta}_4 = -0,895 (2,093)$ e $\hat{\beta}_5 = 4,013 (2,623)$ indicando para o tipo II um tempo médio de sobrevivência significativamente menor do que o tipo I ao nível de 5%. Para o tipo V notamos um tempo médio maior do que o tipo I enquanto que os outros três tipos apresentam tempos médios pouco diferentes do tipo I. Esses resultados confirmam a análise descritiva apresentada na Figura 2.3. O desvio do modelo foi de $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 8,861 \times 5,804 = 51,43$, com 45 graus de liberdade, que leva a $P = 0,236$ e indica um ajuste adequado.

A estimativa de máxima verossimilhança (erro padrão aproximado) do parâmetro de precisão é dada por $\hat{\phi} = 5,804 (1,129)$, indicando que as distribuições dos tempos até a perda da velocidade não devem ser muito assimétricas. Podemos tentar avaliar através de um teste apropriado se os indícios observados pelas estimativas individuais das médias são verificados conjuntamente. Vamos, então, tentar agrupar os tipos I, III e IV. As hipóteses apropriadas são dadas por $H_0 : \beta_4 = \beta_3 = 0$ contra $H_1 : \beta_4 \neq 0$ ou $\beta_3 \neq 0$ do grupo IV. Como $\hat{\phi}$ é relativamente alto podemos aplicar a estatística F dada na Seção 1.7. Sob H_0 obtemos $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 9,091$ para 47 graus de liberdade e sob a hipótese alternativa $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 8,861$ para 45 graus de liberdade. A

estatística F fica dada por

$$\begin{aligned} F &= \frac{(9,091 - 8,861)/2}{8,861/45} \\ &= 0,584, \end{aligned}$$

que leva a $P = 0,562$, ou seja, pela não rejeição de H_0 . As novas estimativas são dadas por $\hat{\mu} = 9,71$ (0,81), $\hat{\beta}_2 = -3,66$ (1,19) e $\hat{\beta}_5 = 5,00$ (2,27). Obtemos $\hat{\phi} = 5,66$ (1,10) e $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 51,47$ para 47 graus de liberdade com $P = 0,30$.

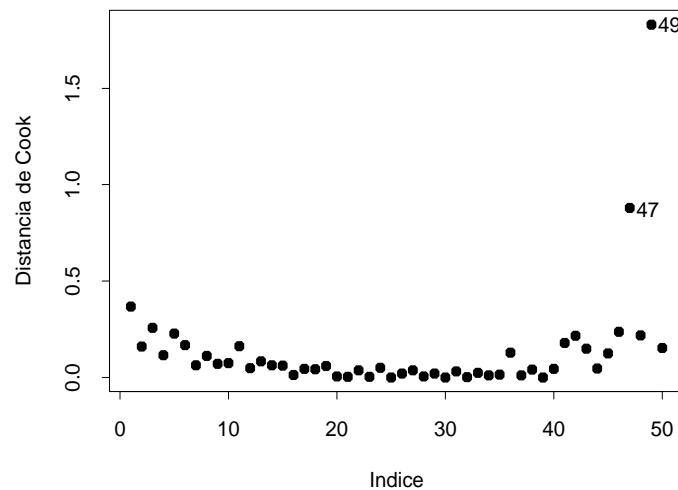


Figura 2.4: Gráfico da distância de Cook aproximada referente ao modelo gama ajustado ao dados sobre desempenho de turbinas de avião.

Na Figura 2.4 temos o gráfico de índices da distância de Cook aproximada. Notamos um forte destaque para a observação #49 seguida da observação # 47 que correspondem, respectivamente, aos valores 25,46 e 12,75 para o tempo até a perda da velocidade de um dos motores de tipo IV e tipo II. O valor 25,46, como mostra o *boxplot* correspondente na Figura 2.3 destoa

dos demais tempos. A eliminação dessa observação aumenta a significância marginal de β_4 , embora esse efeito continue não significativo a 10%. Não há mudanças inferenciais nos demais resultados.

O gráfico normal de probabilidades com envelope para os componentes padronizados do desvio é apresentado na Figura 2.5. Notamos, pelo gráfico, que não há indícios de afastamentos sérios da suposição de distribuição gama para os tempos até a perda da velocidade dos motores, bem como para a suposição de homogeneidade do coeficiente de variação para os cinco grupos. A sequência de comandos para construirmos os gráficos de diagnóstico e normal de probabilidades é descrita no Apêndice B.

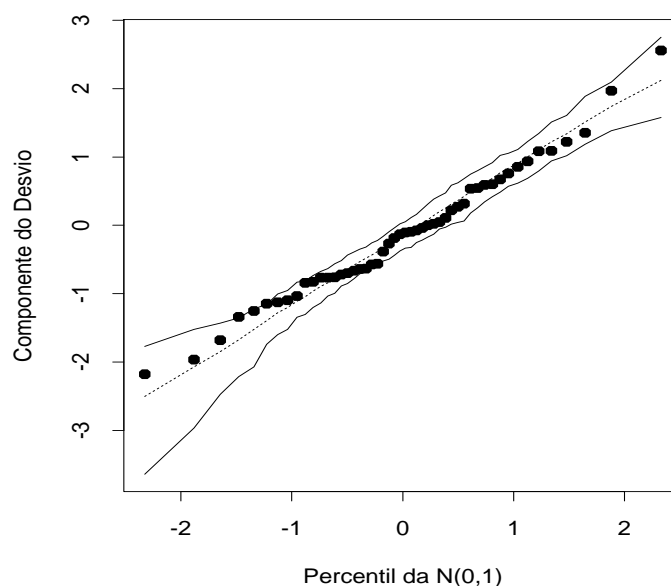


Figura 2.5: Gráfico normal de probabilidades referente ao modelo gama ajustado aos dados sobre desempenho de turbinas de avião.

Portanto, podemos concluir neste exemplo que não existe diferença significativa entre os tipos I, III e IV, enquanto os tipos II e V aparecem de forma

significativa com o menor e maior tempo médio até a perda da velocidade, respectivamente.

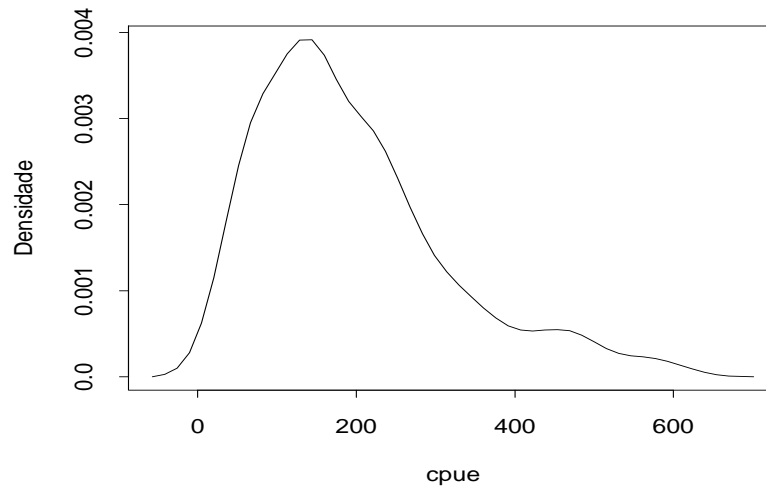


Figura 2.6: Densidade aproximada da cpue para todas as embarcações.

2.4.2 Espinhel de fundo

O espinhel de fundo é definido como um método de pesca passivo, sendo utilizado em todo o mundo em operações de pesca de diferentes magnitudes, da pesca artesanal a modernas pescarias mecanizadas. É adequado para capturar peixes com distribuição dispersa ou com baixa densidade, além de ser possível utilizá-lo em áreas irregulares ou em grandes profundidades. É um dos métodos que mais satisfazem às premissas da pesca responsável, com alta seletividade de espécies e comprimentos, alta qualidade do pescado, consumo de energia baixo e pouco impacto sobre o fundo oceânico. No arquivo **pesca.dat** estão parte dos dados de um estudo sobre a atividade das frotas pesqueiras de espinhel de fundo baseadas em Santos e Ubatuba no litoral

paulista (vide Paula e Oshiro, 2001). A espécie de peixe considerada é o peixe-batata pela sua importância comercial e ampla distribuição espacial. Uma amostra de $n = 156$ embarcações foi analisada no período de 1995 a 1999 sendo 39 da frota de Ubatuba e 117 da frota de Santos. As variáveis consideradas para cada embarcação são as seguintes: frota (Santos ou Ubatuba), ano (95 a 99), trimestre (1 ao 4), latitude¹ (de $23,25^\circ$ a $28,25^\circ$), longitude² (de $41,25^\circ$ a $50,75^\circ$), dias de pesca, captura (quantidade de peixes batata capturados, em kg) e cpue (captura por unidade de esforço, kg/dias de pesca). Um dos objetivos desse estudo é tentar explicar a cpue pelas variáveis frota, ano, trimestre, latitude e longitude. Estudos similares realizados em outros países verificaram que é bastante razoável supor que a cpue tem distribuição assimétrica à direita, como é o caso da distribuição gama (vide, por exemplo, Goni, Alvarez e Adlerstein, 1999).

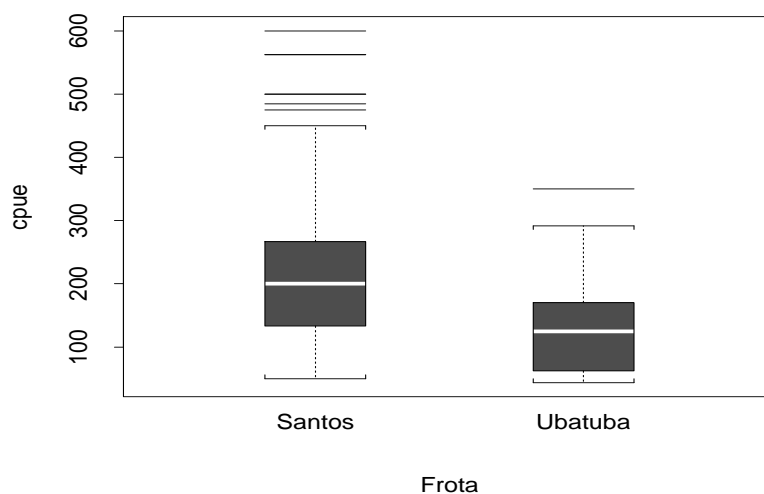


Figura 2.7: Boxplots da cpue segundo a frota.

¹distância ao Equador medida ao longo do meridiano de Greenwich

²distância ao meridiano de Greenwich medida ao longo do Equador

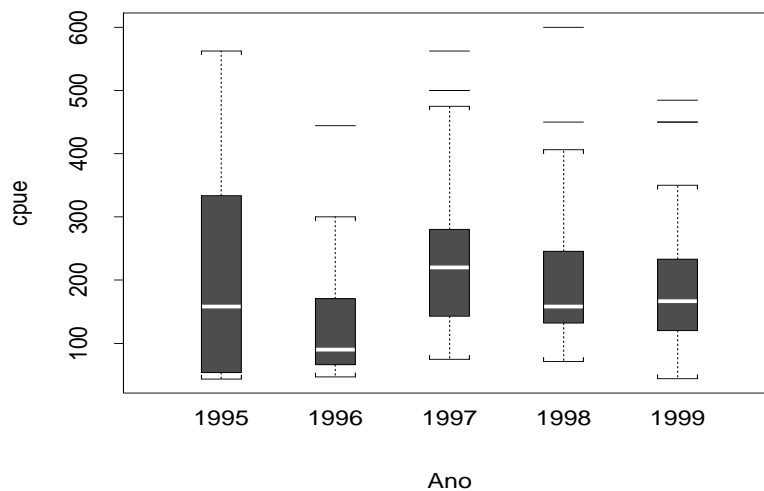


Figura 2.8: Boxplots da cpue segundo o ano.

Para lermos o arquivo **pesca.dat** no R devemos fazer o seguinte:

```
pesca = source("pesca.dat", list(frota=0, ano=0, trimestre=0,
latitude=0, longitude=0, diaspesca=0, captura=0, cpue=0))
frota = factor(frota)
ano = factor(ano)
trimestre = factor(trimestre).
```

Antes de propormos um modelo para tentar explicar a cpue média pelas variáveis explicativas, vamos apresentar uma análise descritiva dos dados. Na Figura 2.6, em que temos a distribuição da cpue para todas as embarcações, podemos notar uma assimetria acentuada à direita confirmando constatações de estudos anteriores. Já nas Figuras 2.7, 2.8 e 2.9 são apresentados os **boxplots** da cpue segundo os fatores frota, ano e trimestre. Notamos uma superioridade da frota de Santos em relação à frota de Ubatuba, porém poucas diferenças entre os níveis dos fatores ano e trimestre, embora o ano de 97

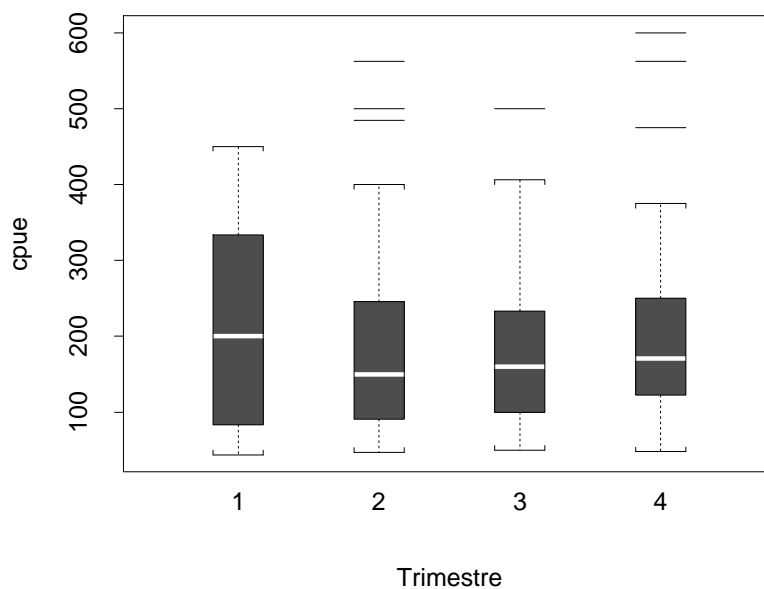


Figura 2.9: Boxplots da cpue segundo o trimestre.

tenha uma mediana um pouco superior aos demais anos.

Frota	Estatística	95	96	97	98	99
Santos	Média	229,37	193,19	262,67	210,29	197,22
	D.Padrão	148,07	132,55	153,60	122,95	103,45
	C. Variação	64,55%	68,61%	58,48%	58,44%	52,45 %
	n	19	8	17	27	46
Ubatuba	Média	47,08	96,09	210,56	174,43	140,85
	D. Padrão	4,73	59,19	77,51	99,16	71,59
	C. Variação	10,05%	61,60 %	36,81%	56,85%	50,83%
	n	3	12	6	5	13

Pela Figura 2.10 notamos que a frota de Santos prefere latitudes e longitudes maiores do que a frota de Ubatuba. Pelos diagramas de dispersão entre cpue e latitude e cpue e longitude apresentados na Figura 2.11 há indícios de um ligeiro crescimento da cpue com a latitude, porém não está bem definida a tendência da cpue com a longitude.

Na tabela dada a seguir são apresentadas as médias, desvios padrão e coeficientes de variação amostrais para as frotas de Santos e Ubatuba referentes ao período 95-99.

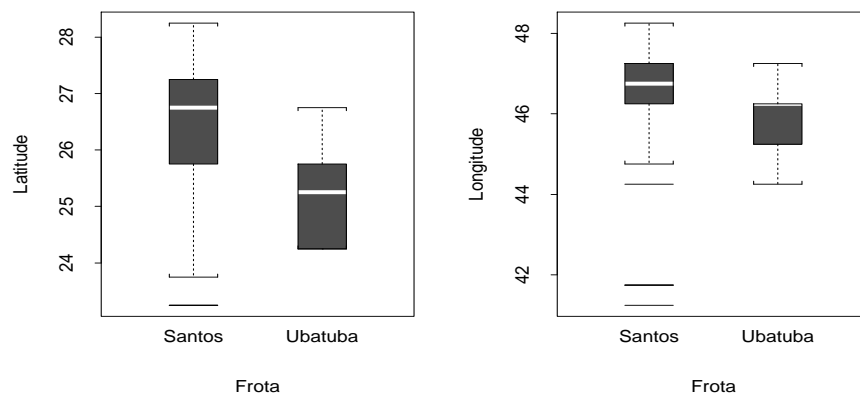


Figura 2.10: Boxplots da latitude e longitude segunda a frota.

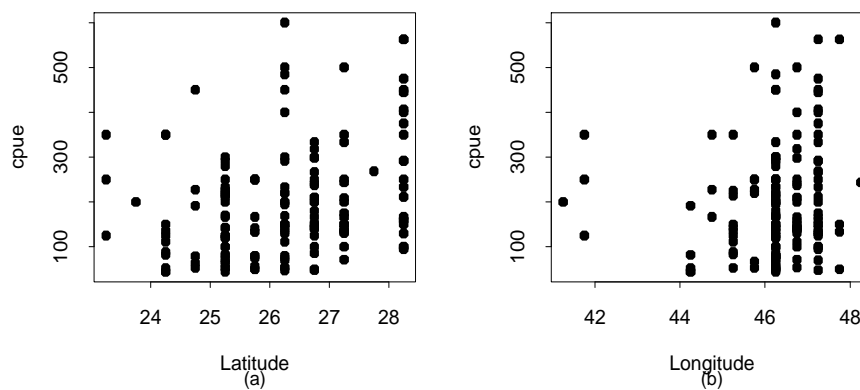


Figura 2.11: Diagramas de dispersão da cpue contra latitude (a) e contra longitude (b).

Portanto, a suposição de coeficiente de variação constante parece ser bas-

tante razoável para a frota de Santos. Para a frota de Ubatuba as estimativas de 95 e 97 destoam das demais, porém devemos levar em conta para essa frota que as amostras são pequenas.

Como todas essas análises são marginais, não levando em conta a presença das outras variáveis, somente através de um modelo apropriado é que poderemos conhecer o efeito de cada variável explicativa na presença das demais. Definimos então Y_{ijkl} como sendo a cpue observada para a i -ésima embarcação da j -ésima frota, (Santos, $j = 1$; Ubatuba $j = 2$), no k -ésimo ano e ℓ -ésimo trimestre ($k, \ell = 1, 2, 3, 4$). Vamos supor que $Y_{ijkl} \sim G(\mu_{ijkl}, \phi)$ com parte sistemática dada por

$$\log \mu_{ijkl} = \alpha + \beta_j + \gamma_k + \theta_\ell + \delta_1 \text{Latitude}_{ijkl} + \delta_2 \text{Longitude}_{ijkl}, \quad (2.2)$$

em que β_j , γ_k e θ_ℓ denotam, respectivamente, os efeitos da j -ésima frota, k -ésimo ano e ℓ -ésimo trimestre. Como estaremos assumindo parametrização casela de referência teremos as restrições $\beta_1 = 0$, $\gamma_1 = 0$ e $\theta_1 = 0$. Temos ainda que latitude_{ijkl} e longitude_{ijkl} denotam, respectivamente, a latitude e longitude da i -ésima embarcação da frota j no k -ésimo ano e trimestre ℓ .

Ajustamos aos dados o modelo gama com parte sistemática dada por (2.2) e aplicando o método de Akaike (vide Seção 1.11.2) retiramos o fator trimestre, permanecendo no modelo os fatores frota e ano além das variáveis quantitativas latitude e longitude. Para ajustarmos o modelo e aplicarmos o método AIC devemos aplicar os seguintes comandos:

```
attach(pesca)
fit1.pesca = glm(cpue ~ frota + ano + trimestre + latitude +
longitude, family=Gamma)
summary(fit1.pesca)
require(MASS)
stepAIC(fit1.pesca).
```

O procedimento **stepAIC** assume que o parâmetro ϕ é constante, ou seja, não muda de um modelo para o outro. Como isso em geral não é satisfeito devemos aplicar algum procedimento alternativo a fim de confirmarmos o modelo escolhido pelo método **AIC**. Fizemos isso através da estatística da razão de verossimilhanças, confirmando a retirada do fator trimestre.

Tabela 2.2

Estimativas dos parâmetros referentes ao modelo gama ajustado aos dados sobre espinhel de fundo.

Efeito	Estimativa	E/E.Padrão
Constante	6,898	3,00
Latitude	0,204	2,81
Longitude	-0,150	-1,97
Frota-Ubatuba	-1,359	-3,68
Ano96	-0,064	-0,26
Ano97	0,141	0,74
Ano98	-0,043	-0,25
Ano99	-0,009	-0,06
FrotaUb*Ano96	0,806	1,77
FrotaUb*Ano97	1,452	3,20
FrotaUb*Ano98	1,502	3,32
FrotaUb*Ano99	1,112	2,76
ϕ	3,67	9,17

O teste da razão de verossimilhanças para incluirmos a interação entre os dois fatores que permaneceram no modelo, frota e ano, foi de $\xi_{RV} = 14,26$ para 4 graus de liberdade e obtemos $P = 0,0065$. Portanto, a interação será incluída no modelo. As estimativas do modelo final que inclui os efeitos principais latitude, longitude, frota e ano além da interação entre ano e frota são apresentadas na Tabela 2.2 e o desvio do modelo foi de $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 162,66$ com 144 graus de liberdade e $P = 0,14$, indicando um modelo bem ajustado. Notamos que à medida que aumenta a latitude aumenta a cpue, ocorrendo tendência contrária à medida que aumenta a longitude. Logo, para

latitudes altas e longitudes baixas (dentro dos limites amostrais), esperamos valores maiores para a captura por unidade de esforço. Com relação a frota e ano, como foi incluída interação entre esses fatores, a interpretação das estimativas deve ser feita com um pouco mais de cuidado. Para isso, exibimos na Figura 2.12 os valores esperados da cpue fixando latitude e longitude nos valores, respectivamente, 26° e 46° . Notamos pela figura que até 96 os valores preditos para a frota de Ubatuba nessas latitude e longitude são bem menores do que os valores preditos para a frota de Santos. Contudo, a partir de 97 as diferenças entre os valores preditos para as duas frotas diminuem. Os valores preditos para a frota de Santos variam pouco no período 95-99, diferentemente dos valores preditos para a frota de Ubatuba.

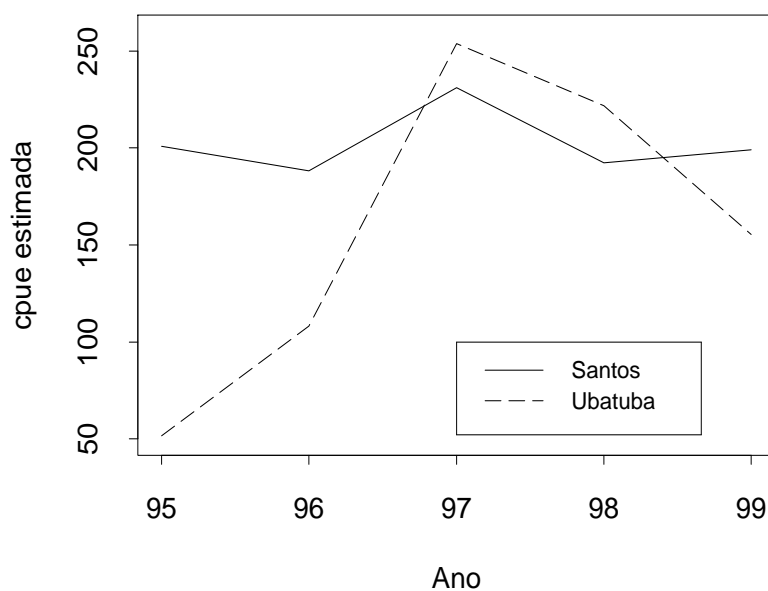


Figura 2.12: Estimativas da cpue média para as frotas de Santos e Ubatuba segundo o ano de operação fixando-se a latitude em 26° e a longitude em 46° através do modelo gama.

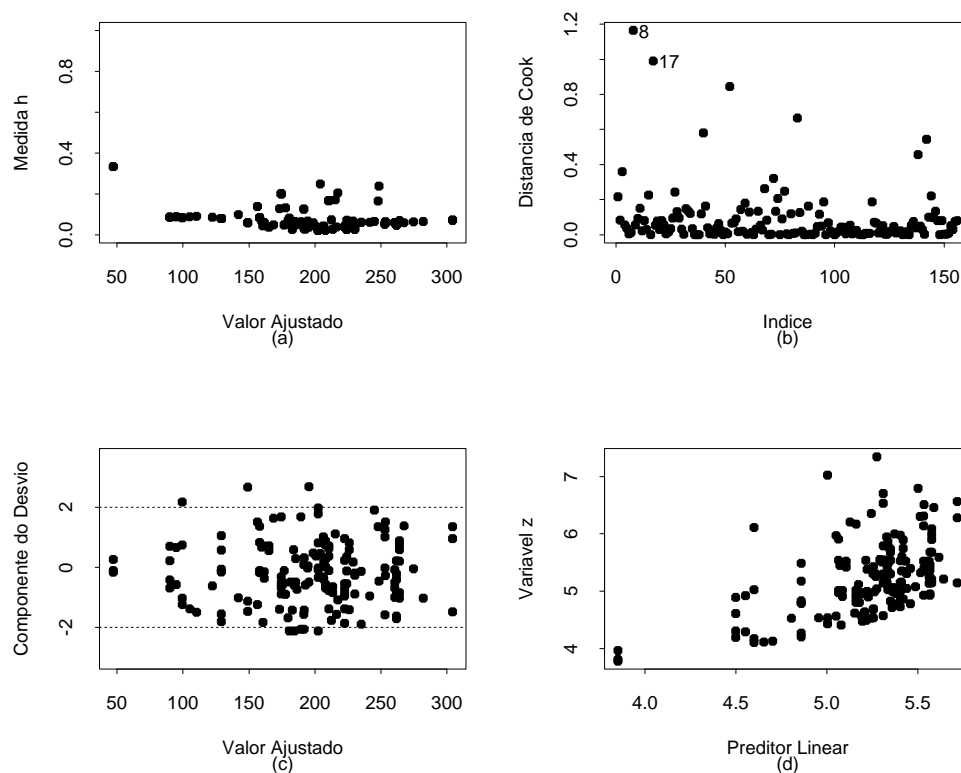


Figura 2.13: Gráficos de diagnóstico referentes ao modelo gama ajustado aos dados sobre espinhel de fundo.

Os gráficos de diagnóstico (Figura 2.13) não apresentam pontos de alavanca ou aberrantes, nem indicações de que a ligação utilizada é inadequada. Contudo, duas observações aparecem com destaque no gráfico de influência, as embarcações #17 e #8. A embarcação #8 é da frota de Ubatuba e obteve uma cpue de 350 numa latitude de $24,25^\circ$ (abaixo da latitude média de $26,22^\circ$) e longitude de $45,25^\circ$ (próxima da longitude média de $46,28^\circ$) no ano de 98. Já a embarcação #17 é da frota de Santos, obteve uma cpue de 450 numa latitude de $24,75^\circ$ e longitude de $46,25^\circ$ no ano de 99. Ambas embarcações alcançaram cpues bastante altas em latitudes relativamente baixas, contrastando com a tendência apresentada pelo modelo. A eliminação

das embarcações individualmente e em conjunto altera algumas estimativas porém não muda as conclusões inferenciais. Todos os efeitos continuam marginalmente significativos ao nível de 10%. O gráfico normal de probabilidades com envelope gerado (Figura 2.14) não apresenta indícios de que a distribuição gama seja inadequada para explicar a cpue.

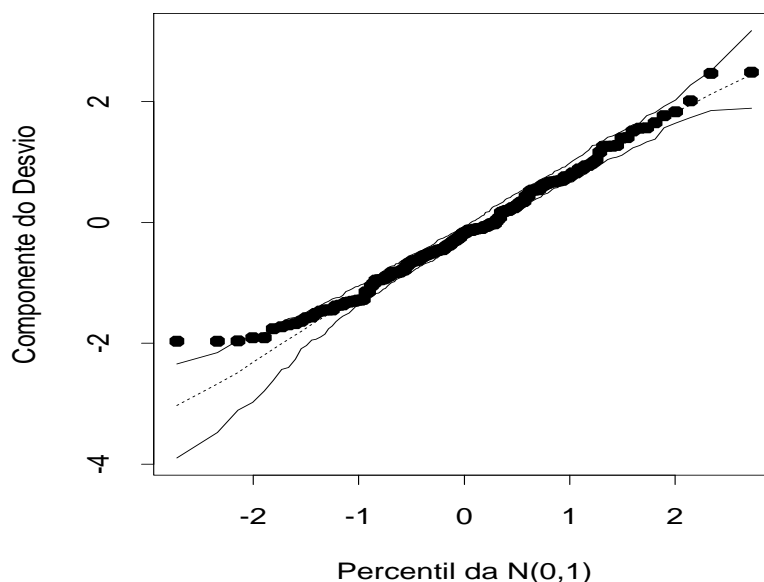


Figura 2.14: Gráfico normal de probabilidades referente ao modelo gama ajustado aos dados sobre espinhel de fundo.

2.4.3 Aplicação em seguros

A fim de ilustrarmos uma aplicação na área de seguros, vamos considerar parte dos dados descritos em de Jong e Heller (2008, pgs. 14-15) referentes aos valores pagos de seguros individuais (em dólares australianos) por danos com acidentes pessoais no período de julho de 1989 a junho de 1999. As análises

serão restritas ao período de janeiro de 1998 a junho de 1999, um total de 769 seguros pagos. Além do valor pago ao segurado serão consideradas as seguintes variáveis explicativas: **legrep**, representação legal (0: não, 1: sim) e **optime**, tempo operacional para pagamento do seguro. Essa última variável assume valores no intervalo (0, 100) e por exemplo um valor 23 significa que 23% dos seguros foram pagos antes do seguro em análise. Como estamos considerando apenas parte dos dados (referentes aos últimos 18 meses), os valores de optime irão variar de 0,1 a 31,9. O subconjunto de dados analisado está descrito no arquivo **insurance.dat**.

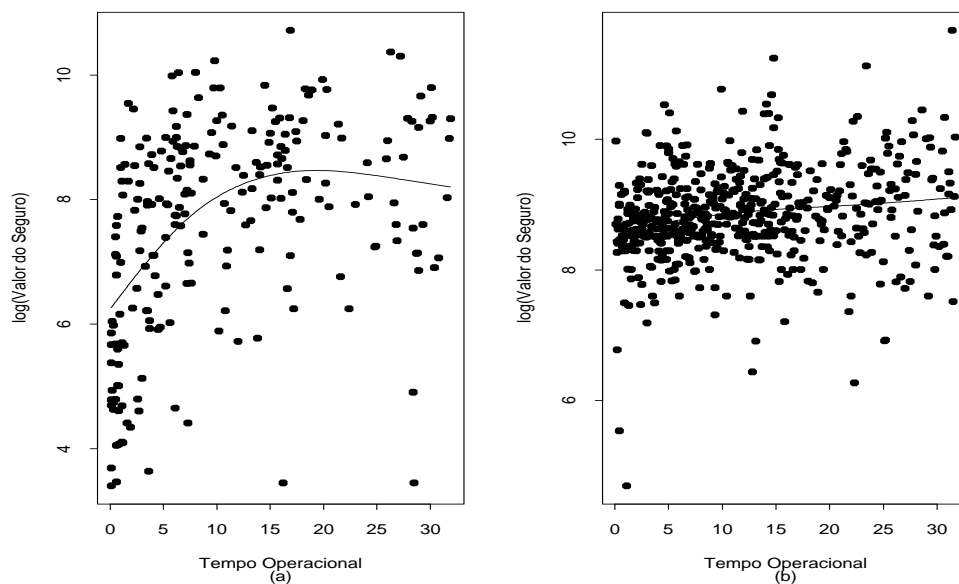


Figura 2.15: Diagrama de dispersão entre o valor pago de seguro e o tempo operacional para os grupos sem representação legal (a) e com representação legal (b).

Na Figura 2.15 temos o diagrama de dispersão entre o logaritmo do valor pago e o tempo operacional para os grupos sem representação legal (Figura

2.15a) e com representação legal (Figura 2.15b) e as tendências ajustadas de forma não paramétrica. Notamos pela Figura 2.15a um crescimento aproximadamente quadrático do logaritmo do valor pago com o tempo operacional, contudo a variabilidade parece ser maior para valores baixos do tempo operacional. Já para os indivíduos com representação legal (Figura 2.15b), notamos que o logaritmo do valor pago cresce linearmente com o tempo operacional enquanto a variabilidade se mantém aproximadamente constante. Notamos também que os valores pagos de seguro são em geral maiores para o grupo com representação legal.

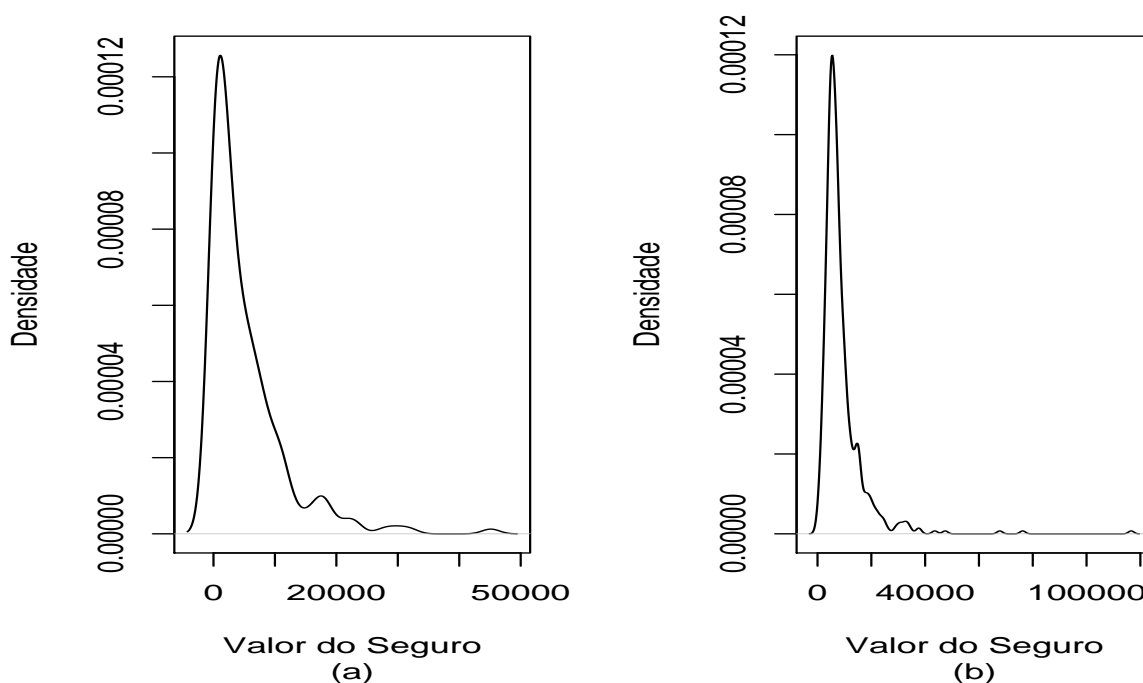


Figura 2.16: Distribuição do valor do seguro para os grupos sem representação legal (a) e com representação legal (b).

Na Figura 2.16 temos a distribuição aproximada do valor pago de seguro

para os dois grupos, sem representação legal (Figura 2.16a) e com representação legal (Figura 2.16b). Em ambos os gráficos podemos notar que a distribuição é fortemente assimétrica à direita, sugerindo distribuições gama ou normal inversa para explicar o valor pago de seguro.

Vamos denotar por Y_{ij} o valor pago de seguro para o j -ésimo indivíduo do i -ésimo grupo ($i = 0$, sem representação legal e $i = 1$ com representação legal) e $j = 1, \dots, n_i$ sendo $n_0 = 227$ e $n_1 = 542$. Conforme sugerido pelas Figuras 2.16a e 2.16b assumiremos inicialmente $Y_{ij} \sim G(\mu_{ij}, \phi_i)$ tais que

$$\begin{aligned}\log \mu_{0j} &= \alpha_0 + \beta_{10} \text{optime}_j + \beta_{20} \text{optime}_j^2 \text{ e} \\ \log \mu_{1j} &= \alpha_1 + \beta_{11} \text{optime}_j.\end{aligned}$$

Tabela 2.3

*Estimativas dos parâmetros referentes
aos modelos com resposta gama ajustados
aos dados sobre seguro.*

Parâmetro	Estimativa	E/E.Padrão
α_0	7,223	44,13
β_{10}	0,204	6,72
β_{20}	-0,005	-5,08
ϕ_0	0,78	12,55
α_1	8,805	140,50
β_{11}	0,023	5,48
ϕ_1	2,22	17,66

As estimativas dos modelos propostos, que foram ajustados separadamente, são descritas na Tabela 2.3. Notamos pelas estimativas que as tendências observadas nas Figuras 2.15a e 2.15b foram confirmadas de forma significativa. Contudo, pelos gráficos normais de probabilidades (Figura 2.17) notamos indícios de afastamentos da distribuição gama para o valor pago de seguro, principalmente para o grupo com representação legal. Para o grupo

sem representação legal notamos que os menores valores do seguro foram superestimados pelo modelo. Os desvios dos dois modelos foram, respectivamente, de $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 270,43$ com 224 graus de liberdade e $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 581,72$ com 540 graus de liberdade. Embora as estimativas de ϕ_0 e ϕ_1 sejam relativamente pequenas, há indícios pelos valores dos desvios que os modelos não estão bem ajustados.

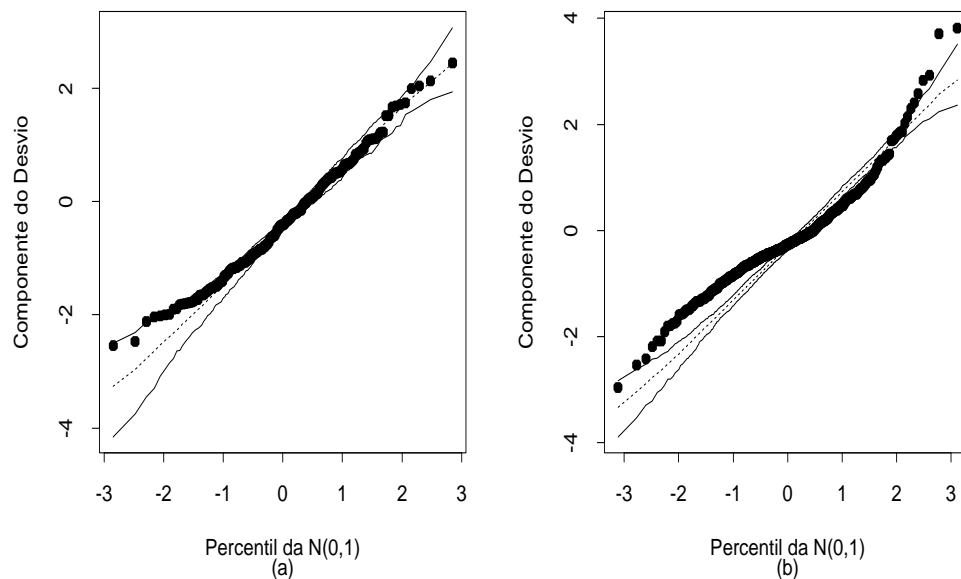


Figura 2.17: Gráfico normal de probabilidades para os modelos com resposta gama ajustados aos dados de seguro para os grupos sem representação legal (a) e com representação legal (b).

Analisando os gráficos da distância de Cook (Figura 2.18a) e resíduo componente do desvio contra os valores ajustados (Figura 2.18b) apenas para o grupo sem representação legal, notamos que não há indícios de observações aberrantes, contudo algumas observações aparecem como possivelmente influentes. Essas observações em geral correspondem a valores altos para o

valor pago de seguro. A eliminação das 10 observações com mais detaque na Figura 2.18a não muda a inferência, todos os coeficientes continuam altamente significativos, porém a estimativa de β_{10} aumenta quase 20%.

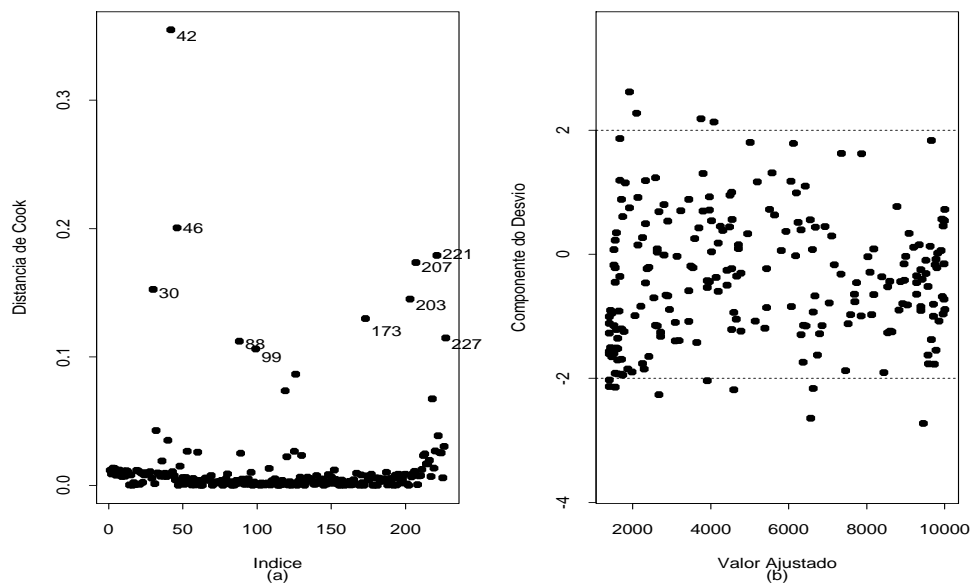


Figura 2.18: Gráficos de diagnóstico para o modelo com resposta gama ajustado aos dados de seguro para o grupo sem representação legal.

Para o grupo com representação legal a utilização de outras ligações ou mesmo outras distribuições são alternativas a fim de tentarmos melhorar a qualidade do ajuste. Paula et al. (2012) comparam ajustes de modelos com resposta gama com modelos com resposta log-Birnbaum-Saunders (log-BS) e log-Birnbaum-Saunders-t (log-BS-t) para explicar o valor pago de seguro para o grupo com representação legal, obtendo um ajuste satisfatório com os modelos log-BS-t. Essa distribuição acomoda melhor as observações aberrantes que correspondem a valores altos de seguro.

2.5 Elasticidade

O modelo log-linear com resposta gama pode ser utilizado para a estimação da elasticidade entre a demanda de um produto e seu preço unitário. Como ilustração, vamos supor que Y denota a demanda e X o preço unitário. É usual em Econometria (ver, por exemplo, Gujarati, 2006, Seção 6.4) assumirmos que

$$Y = \beta_1 x^{\beta_2} e^u, \quad (2.3)$$

sendo u um erro aleatório, em geral assumido $N(0, \sigma^2)$. Isso implica em e^u seguir distribuição log-normal de média $e^{\sigma^2/2}$ e variância $e^{\sigma^2}(e^{\sigma^2} - 1)$. Em vez de uma log-normal podemos assumir, alternativamente, $e^u \sim G(1, \phi)$. Assim, teremos que a média de Y dado x fica dada por

$$\mu(x) = \beta_1 x^{\beta_2},$$

ou seja $\log \mu(x) = \log \beta_1 + \beta_2 \log x$, um MLG com resposta gama e ligação logarítmica.

Para entendermos a elasticidade entre a demanda e o preço do produto, vamos supor que o preço aumente $r \times 100\%$ de modo que o novo preço seja dado por $x_N = (1 + r)x$, para $0 < r < 1$. O novo valor esperado para a demanda fica dado por

$$\mu(x_N) = \beta_1 x_N^{\beta_2}$$

e a razão entre as demandas médias assume a forma

$$\begin{aligned} \frac{\mu(x_N)}{\mu(x)} &= e^{\beta_2 \log(1+r)} \\ &= (1 + r)^{\beta_2}. \end{aligned}$$

Para r pequeno temos a aproximação

$$\frac{\mu(x_N)}{\mu(x)} \cong (1 + r\beta_2),$$

de modo que se o preço aumentar 1% ($r = 0,01$) a demanda aumenta $\beta_2\%$, ou seja,

$$\frac{\mu(x_N)}{\mu(x)} = \left(1 + \frac{\beta_2}{100}\right).$$

O parâmetro β_2 é conhecido como elasticidade entre a demanda e o preço do produto.

2.5.1 Modelo de Cobb-Douglas

O modelo (2.3) pode ser estendido para duas ou mais variáveis explicativas as quais poderão representar outros tipos de preço ou mesmo algum tipo de insumo. Em particular, o modelo de Cobb-Douglas (ver, por exemplo, Gujarati, Exemplo 7.3) considera a seguinte equação para explicar a demanda de um produto pelos insumos de mão de obra e capital:

$$Y = \beta_1 x_2^{\beta_2} x_3^{\beta_3} e^u, \quad (2.4)$$

em que Y denota a demanda, x_2 o valor do insumo de mão de obra, x_3 o valor do insumo de capital e u o erro aleatório. Para x_3 fixado (x_2 fixado) o parâmetro β_2 (β_3) mede a elasticidade parcial entre a demanda e o insumo de mão de obra (capital). A soma $\beta_2 + \beta_3$ mede os retornos de escala, ou seja, se $\beta_2 + \beta_3 = 1$ significa que os retornos são proporcionais, dobrando o uso de insumos a demanda esperada aumenta duas vezes, triplicando os insumos há aumento de três vezes para a demanda esperada, e assim por diante. Se $\beta_2 + \beta_3 < 1$ os retornos de escala serão menores, dobrando os insumos esperamos demanda menor do que o dobro, e se $\beta_2 + \beta_3 > 1$ os retornos de escala serão maiores, dobrando os insumos esperamos que a demanda aumente mais que duas vezes.

Para mostrarmos esses resultados suponha que os novos insumos de mão de obra e de capital sejam dados por $x_{1N} = rx_1$ e $x_{2N} = rx_2$, ou seja,

aumentam r vezes. Assim, a nova demanda esperada será dada por

$$\begin{aligned}\mu(x_{1N}, x_{2N}) &= \beta_1(rx_2)^{\beta_2}(rx_3)^{\beta_3} \\ &= r^{(\beta_2+\beta_3)}\beta_1x_2^{\beta_2}x_3^{\beta_3} \\ &= r^{\beta_2+\beta_3}\mu(x_1, x_2),\end{aligned}$$

em que $\mu(x_1, x_2)$ é a demanda esperada inicial. Logo, se $\beta_2 + \beta_3 = 1$ então $\mu(x_{1N}, x_{2N}) = r\mu(x_1, x_2)$, ou seja, a demanda esperada aumenta r vezes. Por outro lado, se $\beta_2 + \beta_3 < 1$ teremos que $\mu(x_{1N}, x_{2N}) < r\mu(x_1, x_2)$, ou seja, a demanda esperada aumenta menos que r vezes e se $\beta_2 + \beta_3 > 1$ teremos que a demanda esperada aumenta mais que r vezes, $\mu(x_{1N}, x_{2N}) > r\mu(x_1, x_2)$.

Obviamente que existem várias distribuições candidatas para explicar e^u , sendo as distribuições gama e normal inversa as candidatas naturais na classe dos MLGs. Podemos também assumir que $\log u$ tenha distribuição normal. Contudo, somente através de uma análise de diagnóstico é que podemos avaliar a adequação de cada distribuição.

2.5.2 Aplicação

Como ilustração vamos considerar um experimento aleatorizado descrito em Griffiths, Hill e Judge (1993, Seção 11.8.1c) em que a produtividade de milho (libras/acre) é estudada segundo várias combinações de nitrogênio e fósforo (40, 80, 120, 160, 200, 240, 280 e 320 libras/acre). Os dados estão descritos no arquivo **milho.dat**. Nas Figuras 2.19a e 2.19b temos os diagramas de dispersão entre a produtividade de milho e as quantidades de nitrogênio e fósforo, respectivamente, e conforme podemos notar nessas figuras há indícios de uma tendência crescente da produtividade com o aumento dos insumos. Notamos também um aumento da variabilidade com o aumento das quantidades de nitrogênio e fósforo, sugerindo que a suposição de distribuição gama

ou normal inversa para $\log u$ no modelo de Cobb-Douglas pode levar a um ajuste adequado.

Denotaremos por Y_i a produtividade de milho dada a combinação (x_{1i}, x_{2i}) de nitrogênio e fósforo correspondente à i -ésima condição experimental ($i = 1, \dots, 30$). Vamos então supor que $Y_i \sim G(\mu_i, \phi)$ com parte sistemática dada por $\log \mu_i = \alpha + \beta_1 \log x_{1i} + \beta_2 \log x_{2i}$.

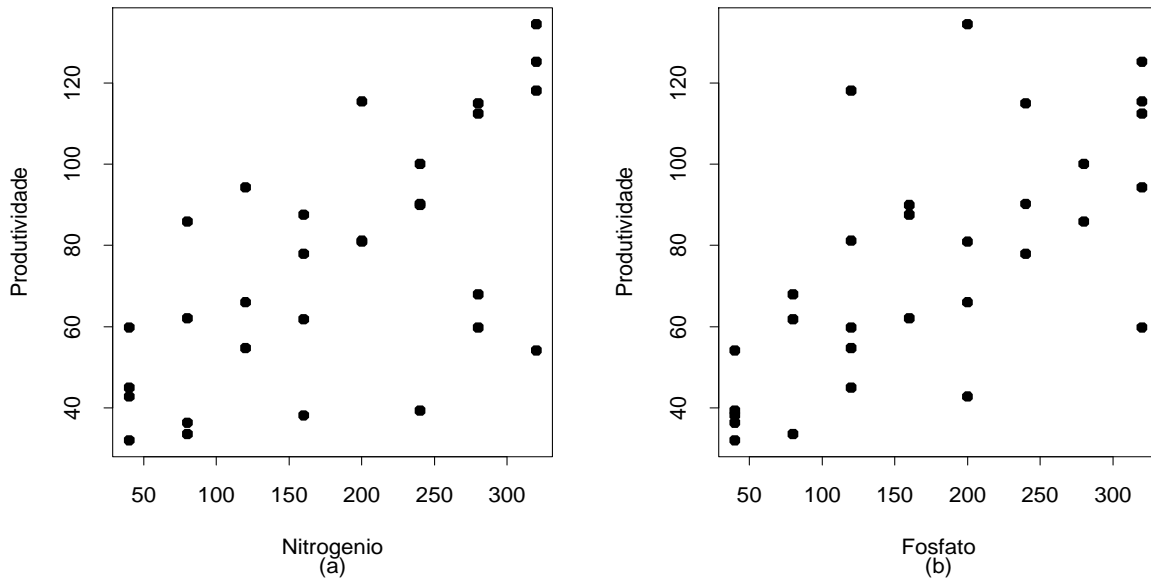


Figura 2.19: Diagramas de dispersão entre a quantidade de nitrogênio e a produtividade de milho (a) e a quantidade de fósforo e a produtividade de milho (b).

As estimativas são descritas na Tabela 2.4, e como podemos notar os coeficientes são altamente significativos, confirmando as tendências observadas na Figura 2.19. Na Figura 2.20a temos o gráfico normal de probabilidades que não apresenta observações aberrantes nem indícios de afastamentos da distribuição assumida para os erros. Pelo gráfico da distância de Cook (Fi-

gura 2.20b) a observação #1 aparece com mais destaque. Reajustando o modelo sem essa observação notamos variações importantes nas estimativas dos coeficientes β_1 e β_2 , respectivamente. Porém a inferência não muda, esses coeficientes continuam sendo altamente significativos. A estimativa de ϕ sugere que a suposição de distribuição log-normal para os erros também pode levar a um ajuste adequado, embora o modelo gama contemple a possibilidade de heteroscedasticidade.

Tabela 2.4

Estimativas dos parâmetros referentes ao modelo de Cobb-Douglas ajustado ao dados sobre produtividade de milho.

Parâmetro	Estimativa	E/E.Padrão
α	0,469	1,67
β_1	0,350	8,30
β_2	0,410	10,07
ϕ	46,59	11,99

A fim de verificarmos como ocorrem os retornos de produtividade de milho com as aplicações de fosfato e nitrogênio vamos encontrar uma estimativa intervalar para $\beta_1 + \beta_2$. Temos que $\hat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2) = 0,003143$, assim uma estimativa intervalar de coeficiente de confiança de 95% fica dada por $[0,35 + 0,41 + 1,96 \cdot \sqrt{0,003143}] = [0,52; 0,87]$ que não cobre o valor 1,0. Portanto, dobrando as aplicações de insumos de nitrogênio e fosfato os retornos esperados de produtividade devem aumentar menos do que duas vezes.

2.6 Distribuição normal inversa

Vamos supor que Y é uma variável aleatória com distribuição normal inversa de média μ e parâmetro de dispersão ϕ^{-1} , denotamos $Y \sim \text{NI}(\mu, \phi)$, cuja

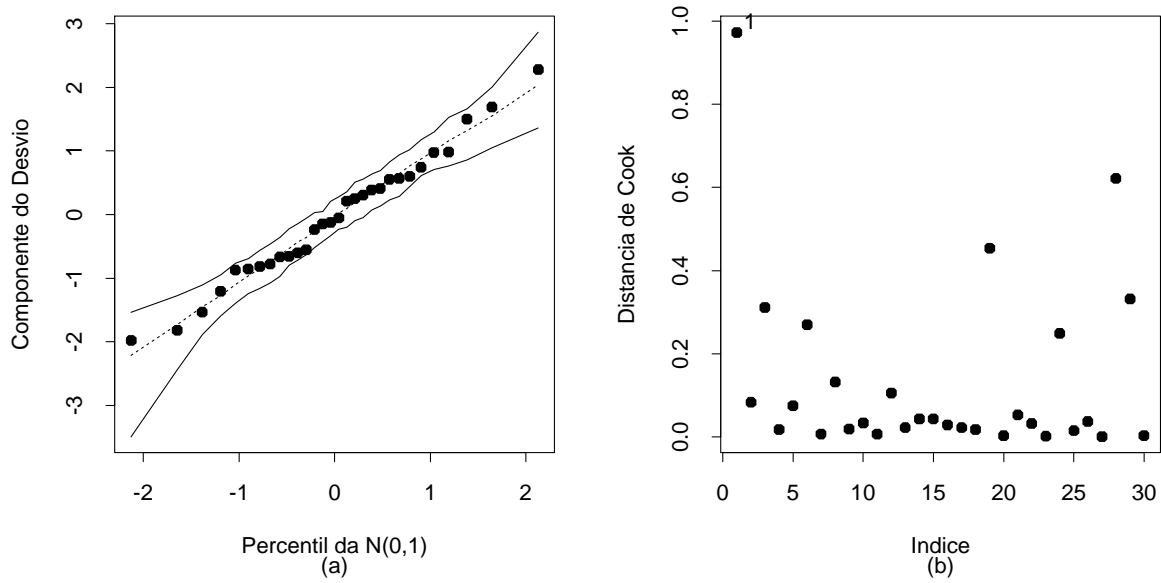


Figura 2.20: Gráfico normal de probabilidades (a) e distância de Cook (b) referentes ao modelo de Cobb-Douglas ajustado aos dados do experimento sobre produtividade de milho.

função densidade é expressa na forma

$$\begin{aligned}
 f(y; \mu, \phi) &= \frac{\phi^{1/2}}{\sqrt{2\pi y^3}} \exp \left\{ -\frac{\phi(y - \mu)^2}{2\mu^2 y} \right\} \\
 &= \exp \left[\phi \left\{ -\frac{y}{2\mu^2} + \frac{1}{\mu} \right\} - \frac{1}{2} \left\{ \log(2\pi y^3 / \phi) + \frac{\phi}{y} \right\} \right],
 \end{aligned}$$

em que $y > 0$, $\mu > 0$. Na Figura 2.21 temos a densidade da distribuição normal inversa variando o parâmetro de precisão para μ fixado. Notamos que à medida que ϕ aumenta a distribuição normal inversa fica mais simétrica em torno da média. Podemos mostrar que à medida que ϕ aumenta Y se aproxima de uma distribuição normal de média μ e variância $\mu^3 \phi^{-1}$. Similarmente à distribuição gama, a normal inversa torna-se atrativa para o estudo de variáveis aleatórias assimétricas e também simétricas em que a variância depende de forma cúbica da média. Uma discussão sobre as su-

posições teóricas para a construção da distribuição normal inversa pode ser encontrada, por exemplo, em Leiva, Barros e Paula (2009, Cap. 2).

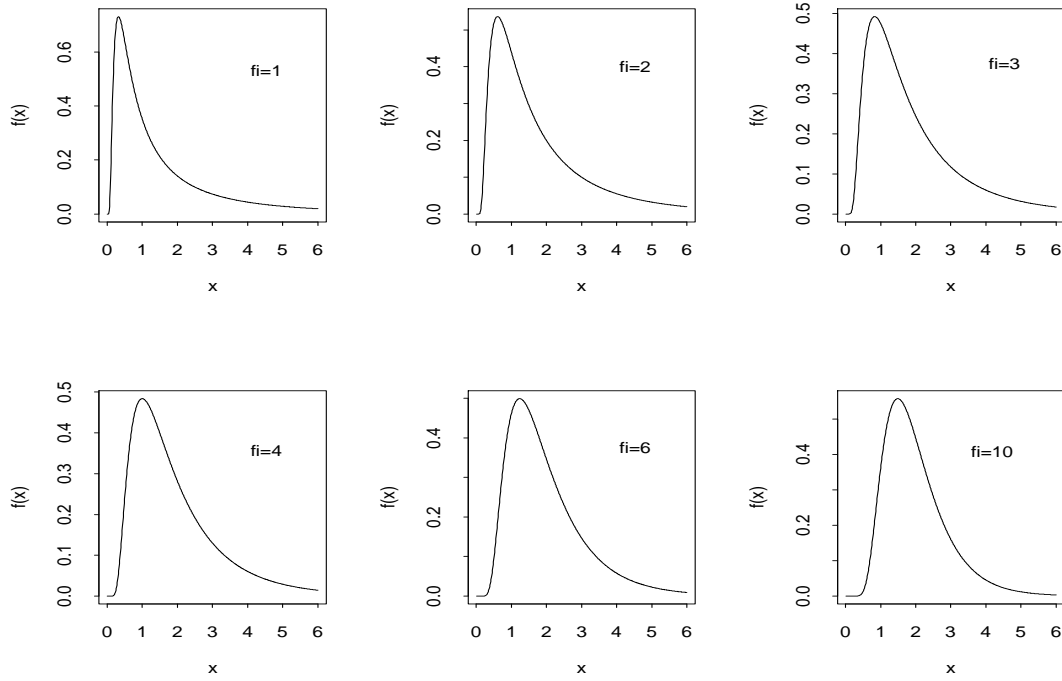


Figura 2.21: Densidades da distribuição normal inversa para alguns valores do parâmetro de dispersão e supondo $\mu = 2$.

A função de sobrevivência da distribuição normal inversa de média μ e parâmetro de dispersão ϕ^{-1} (ver, por exemplo, Collett, 2003, pp. 198-199) é dada por

$$S(t) = \Phi\{(1 - t\mu^{-1})\sqrt{\phi t^{-1}} - \exp(2\phi/\mu)\Phi\{-(1 + t\mu^{-1})\sqrt{\phi t^{-1}}\}.$$

A função de risco é dada por $h(t) = f(t)/S(t)$ em que $f(y)$ denota a função densidade da NI(μ, ϕ).

2.7 Modelos com resposta normal inversa

Sejam Y_1, \dots, Y_n variáveis aleatórias independentes tais que $Y_i \sim \text{NI}(\mu_i, \phi)$. Estamos assumindo que essas variáveis possuem médias diferentes e mesma dispersão ϕ^{-1} . Ademais, vamos supor $g(\mu_i) = \eta_i$ em que $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ com $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ contendo valores de variáveis explicativas e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ sendo o vetor de parâmetros de interesse. As ligações mais usadas no caso normal inversa são identidade ($\mu_i = \eta_i$), logarítmica ($\log \mu_i = \eta_i$) e recíproca quadrática ($\mu_i = \eta_i^{-2}$), esta última sendo a ligação canônica.

2.7.1 Qualidade do ajuste

Como foi visto na Seção 1.4 o desvio de um modelo com resposta normal inversa é dado por $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ em que

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / (y_i \hat{\mu}_i^2), \quad (2.5)$$

com $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$, $\hat{\eta}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$ e $y_i > 0$. Como ϕ é desconhecido devemos estimá-lo, por exemplo através de máxima verossimilhança, cuja solução é dada por $\hat{\phi} = n/D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$. Supondo que o modelo postulado está correto temos, para ϕ grande, que o desvio $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ segue distribuição qui-quadrado com $(n - p)$ graus de liberdade. Assim, valores altos para o desvio podem indicar inadequação do modelo ou falta de ajuste.

2.7.2 Técnicas de diagnóstico

O resíduo componente do desvio padronizado para os modelos com resposta normal inversa assumem a forma

$$t_{D_i} = \frac{\sqrt{2\phi}}{\sqrt{1 - \hat{h}_{ii}}} \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i \sqrt{y_i}},$$

em que $y_i > 0$ e h_{ii} é o i -ésimo elemento da diagonal principal da matriz $\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$ com $\omega_i = (d\mu_i/d\eta_i)^2/\mu_i^3$. Na expressão para t_{D_i} no caso da distribuição normal inversa o sinal do resíduo é o mesmo de $(y_i - \hat{\mu}_i)$. Estudos de simulação indicam que o resíduo t_{D_i} se aproxima da distribuição normal, particularmente para ϕ grande.

Similarmente aos modelos com resposta gama podemos obter uma expressão aproximada para a distância de Cook quando a i -ésima observação é excluída. Essa expressão fica dada por

$$LD_i = \frac{\phi \hat{h}_{ii}}{(1 - \hat{h}_{ii})^2} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2}.$$

Aqui também são sugeridos gráficos de t_{D_i} e \hat{h}_{ii} contra os valores ajustados $\hat{\mu}_i$ e gráficos de índices de LD_i .

2.8 Aplicações

2.8.1 Comparação de cinco tipos de *snack*

Vamos considerar nesta seção parte dos dados de um experimento desenvolvido no Departamento de Nutrição da Faculdade de Saúde Pública da USP em que 5 formas diferentes de um novo tipo de *snack*, com baixo teor de gordura saturada e de ácidos graxos, foram comparados ao longo de 20 semanas. Neste novo produto a gordura vegetal hidrogenada, responsável pela fixação do aroma do produto, foi substituída, totalmente ou parcialmente, por óleo de canola. As formas são as seguintes: A (22% de gordura, 0% de óleo de canola), B (0% de gordura, 22% de óleo de canola), C (17% de gordura, 5% de óleo de canola), D (11% de gordura, 11% de óleo de canola) e E (5% de gordura, 17% de óleo de canola). O experimento foi conduzido de modo que nas semanas pares 15 embalagens de cada um dos produtos A, B, C, D e E fossem analisadas em laboratório e observadas diversas variáveis (ver Paula, de

Moura e Yamaguchi, 2004). Em particular, vamos estudar o comportamento da textura dos produtos através da força necessária para o cisalhamento. Os dados referentes a esta variável estão disponíveis no arquivo **snack.dat**.

Para lermos o arquivo **snack.dat** no R devemos fazer o seguinte:

```
snack = source("snack.dat", list(cisalhamento=0, grupo=0,
semana=0))

grupo = factor(grupo).
```

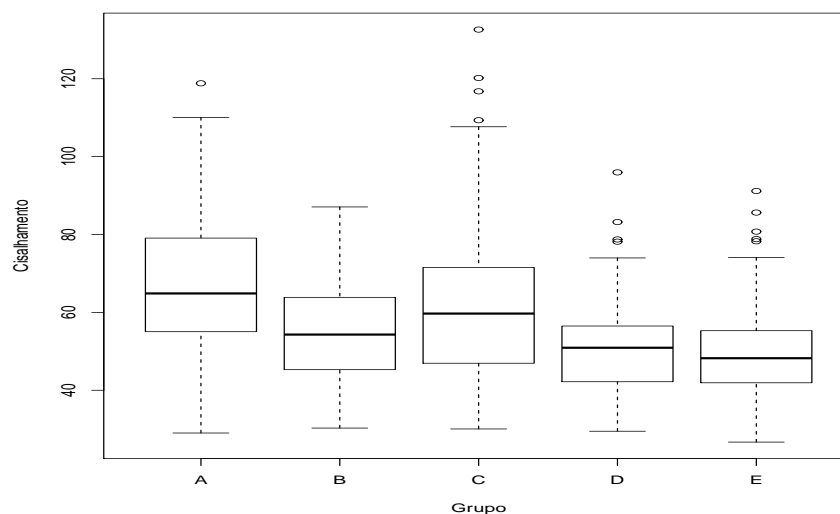


Figura 2.22: **Boxplots** da força de cisalhamento segundo o grupo e para todas as semanas.

Notamos pela Figura 2.22, em que são apresentados os **boxplots** da força de cisalhamento segundo o grupo e para todas as semanas, que os grupos A e C possuem os maiores valores, enquanto o grupo C tem valores intermediários e os grupos D e E têm os menores valores. Nota-se uma assimetria à direita na distribuição da força de cisalhamento para todos os grupos. Essas tendências

são confirmadas pela tabela dada a seguir em que são apresentadas as médias, desvio padrão e coeficiente de variação para a força de cisalhamento para cada grupo.

Estatística	Grupo A	Grupo B	Grupo C	Grupo D	Grupo E
Média	66,201	55,294	61,632	51,027	50,257
D.Padrão	18,707	13,143	19,601	10,960	11,402
C. Variação	28,20%	23,80%	31,80%	21,50%	22,70%

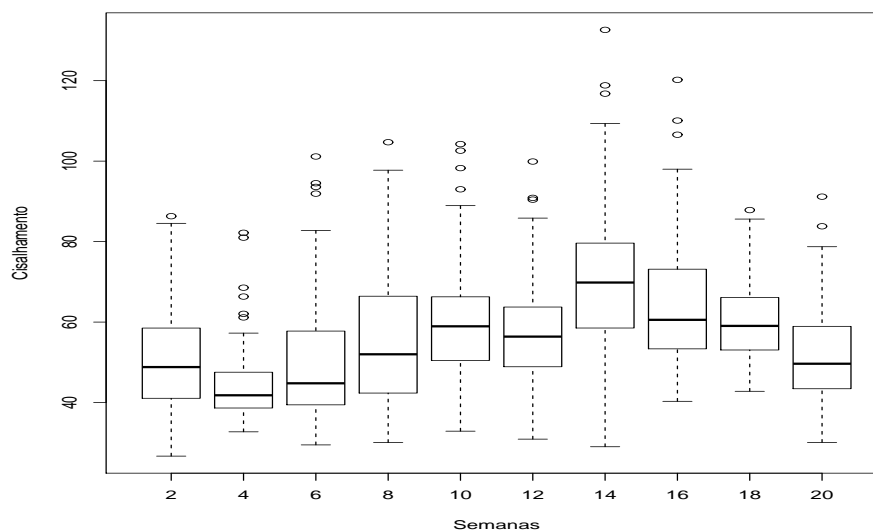


Figura 2.23: **Boxplots** da força de cisalhamento segundo a semana e para todos os grupos.

Já na Figura 2.23, em que são apresentados os **boxplots** para todos os grupos ao longo das 20 semanas, uma tendência crescente é observada até a 14^a semana seguida de um decréscimo até a última semana. Verificamos também, para cada semana, que a distribuição da força de cisalhamento

mostra-se assimétrica à direita sugerindo uma distribuição gama ou normal inversa. Essas tendências são confirmadas pelo gráfico de perfis para a força de cisalhamento (vide Figura 2.24) e pela tabela dada a seguir em que são apresentadas as médias, desvio padrão e coeficiente de variação para a força de cisalhamento para cada semana.

Estatística	Semana 2	Semana 4	Semana 6	Semana 8	Semana 10
Média	50,95	44,66	50,08	55,57	60,15
D.Padrão	13,12	9,76	15,97	16,28	14,72
C. Variação	25,80%	21,90%	31,90%	29,30%	24,50%

Estatística	Semana 12	Semana 14	Semana 16	Semana 18	Semana 20
Média	57,84	71,57	65,18	60,37	52,45
D.Padrão	13,61	20,17	16,95	10,25	12,58
C. Variação	23,50%	28,20%	26,00%	17,00%	24,00%

Assim, denotaremos por Y_{ijk} a força de cisalhamento referente à k -ésima réplica do i -ésimo grupo na j -ésima semana, para $k = 1, \dots, 15$, $j = 2, 4, 6, \dots, 20$ e $i = 1(A), 2(B), 3(C), 4(D)$ e $E(5)$. A fim de compararmos as duas distribuições assimétricas vamos supor que $Y_{ijk} \sim G(\mu_{ij}, \phi)$ e $Y_{ijk} \sim NI(\mu_{ij}, \phi)$ com parte sistemática dada por

$$\mu_{ij} = \alpha + \beta_i + \gamma_1 \text{semana}_j + \gamma_2 \text{semana}_j^2, \quad (2.6)$$

em que $\beta_1 = 0$. Portanto α é o efeito da forma A, controlando pela semana, e $\alpha + \beta_i$ ($i=2,3,4,5$) são os efeitos das demais formas B, C, D e E, respectivamente. Estamos supondo a mesma tendência para os cinco tipos de *snack*. Alternativamente poderíamos incluir interação entre grupo e semana, possibilitando o ajuste de tendências separadas para cada grupo.

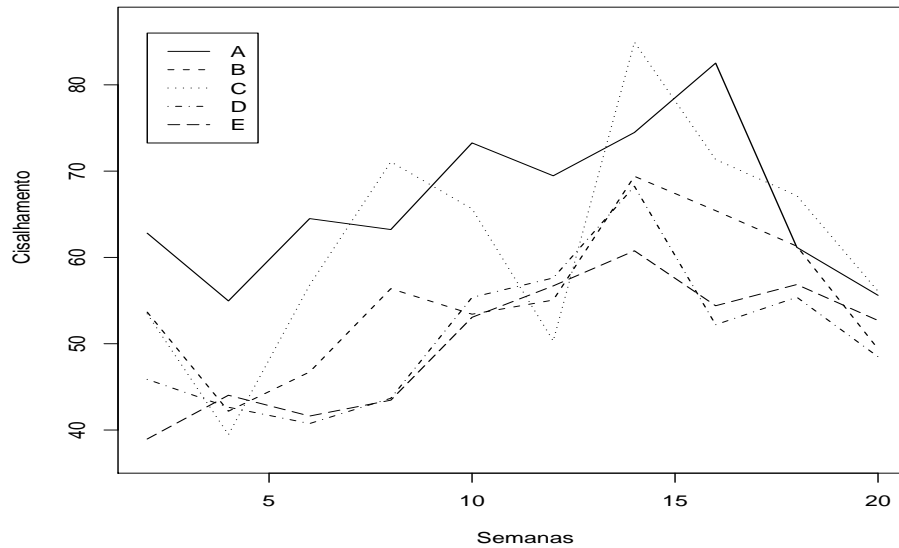


Figura 2.24: Perfis da força de cisalhamento segundo as semanas e os grupos.

Para ajustarmos o modelo (2.6) com resposta normal inversa sem interação devemos fazer o seguinte:

```
s1 = semana
s2 = s1*s1
fit1.snack = glm(cisalhamento ~ grupo + s1 + s2,
family=inverse.gaussian(link=identity))
summary(fit1.snack).
```

Abaixo seguem os comandos para o ajuste com interação

```
fit2.snack = glm(cisalhamento ~ grupo + s1 + s2 + s1*grupo
+s2*grupo, family=inverse.gaussian(link=identity))
summary(fit2.snack).
```

Este é um exemplo em que há uma ligeira superioridade da distribuição normal inversa em relação à distribuição gama. Embora a função de variância

da normal inversa seja cúbica enquanto para a gama temos função de variância quadrática, nem sempre é possível diferenciarmos de forma clara os dois ajustes. Notamos pela Figura 2.25 que o gráfico de resíduos de Pearson contra os valores ajustados apresenta uma tendência sistemática crescente sob o modelo gama, que é amenizada sob o modelo com erros normal inversa. Os dois modelos ajustam-se muito bem aos dados como podemos notar pelo valor do desvio do modelo gama $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 756,87$ (753 g.l.) com $P=0,35$ e pelo gráfico normal de probabilidades para o modelo com resposta normal inversa apresentado na Figura 2.26.

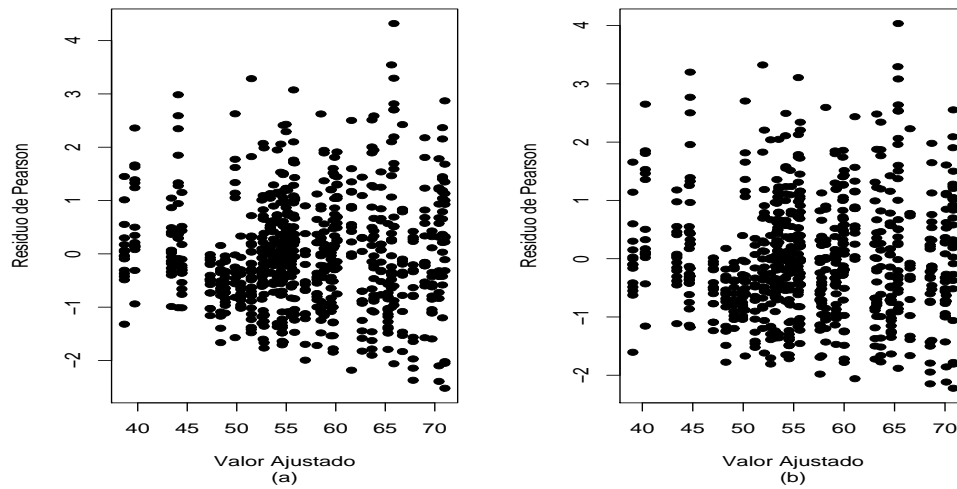


Figura 2.25: Gráficos do resíduo de Pearson contra os valores ajustados referentes aos modelos gama (a) e normal inversa (b) ajustados aos dados sobre *snacks*.

Na Tabela 2.5 são apresentadas as estimativas sob o modelo com resposta normal inversa. Todos os efeitos são altamente significativos, em particular o efeito de semana na forma quadrática. Controlando esse efeito, a maior força média de cisalhamento ocorre com o produto sob a forma A (ausência de óleo

de canola) e a menor força média de cisalhamento ocorre com as formas D e E, confirmando-se as tendências observadas na Figura 2.22.

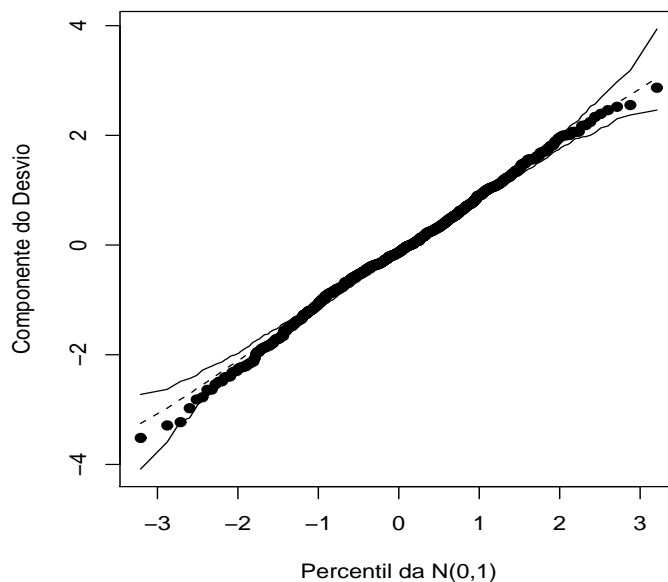


Figura 2.26: Gráfico normal de probabilidades referente ao modelo com resposta normal inversa ajustado aos dados sobre *snacks*.

Na Figura 2.27 temos os valores preditos para os 5 grupos ao longo das 20 semanas. A estimativa do parâmetro de precisão indica que a distribuição da força de cisalhamento em cada grupo, fixando o tempo, é aproximadamente normal. Contudo, a variância depende da média. A forma cúbica para a variância mostrou-se ligeiramente superior à forma quadrática. Outras formas para ajustarmos a variância podem ser testadas, como por exemplo, através de modelos de quase-verossimilhança que serão discutidos no Capítulo 5. O paralelismo entre as curvas apresentadas na Figura 2.27 é devido à não inclusão de interação entre semana e grupo. Alternativamente, poderíamos incluir uma função para cada grupo, ou então, o efeito semana poderia ser

controlado através de funções não paramétricas (ver, por exemplo, Wood, 2006).

Tabela 2.5

Estimativas dos parâmetros referentes ao modelo com resposta normal inversa ajustado aos dados sobre snacks.

Efeito	Estimativa	E/E.Padrão
Constante	50,564	26,32
Grupo B	-10,916	-6,41
Grupo C	-5,459	-3,03
Grupo D	-15,357	-9,42
Grupo E	-16,596	-10,30
Semana	2,727	8,18
Semana ²	-0,091	-5,90
ϕ	1005	-

Algumas observações foram detectadas como possivelmente influentes (vide Figura 2.28): #2 (2ª semana, grupoB), #8 (2ª semana, grupo B), #10 (2ª semana, grupo B), #311 (2ª semana, grupo C), #465 (2ª semana, grupo D) e #744 (última semana, grupo E). Embora os valores preditos para a força de cisalhamento dessas amostras estejam abaixo da média, os valores observados são em geral altos quando comparados com os valores dos grupos e das semanas correspondentes. Também o fato de 5 dessas observações terem ocorrido logo na segunda semana pode ser um indício de alguma dificuldade inicial com o experimento. A eliminação dessas 6 observações do total de 744 observações leva a algumas variações desproporcionais. Por exemplo, as estimativas dos efeitos dos grupos B e C diminuem por volta de 8%. Todavia, não ocorrem mudanças inferenciais importantes.

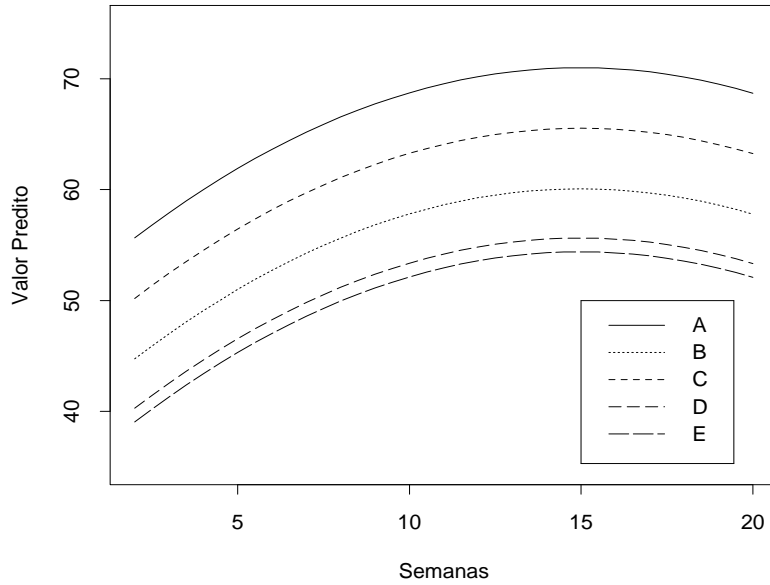


Figura 2.27: Valores preditos para a força média de cisalhamento para as 5 formas de *snacks* através do modelo com resposta normal inversa.

2.8.2 Projeção de vendas

Whitmore (1986) (ver também Faraway, 2006, Cap. 7) apresenta um conjunto de dados em que o valor verdadeiro para as vendas é relacionado com o valor projetado para as vendas de 20 produtos. Os dados estão descritos na Tabela 2.6. Como as vendas tendem a ter uma variabilidade maior para valores mais altos e os dados são positivos, a suposição de erros normais com variância constante não parece ser razoável. Whitmore (1986) sugere uma distribuição normal inversa para a resposta Y_i que denota o total de vendas reais para o i -ésimo produto e que $\mu_i = E(Y_i)$ é relacionada com x_i , o i -ésimo valor projetado, de forma linear $\mu_i = \beta x_i$, ou seja, uma regressão linear passando pela origem. O diagrama de dispersão dos dados é descrito na Figura 2.29, juntamente com as retas ajustadas sob erros gama e normal inversa. Notamos que as retas ajustadas são similares.

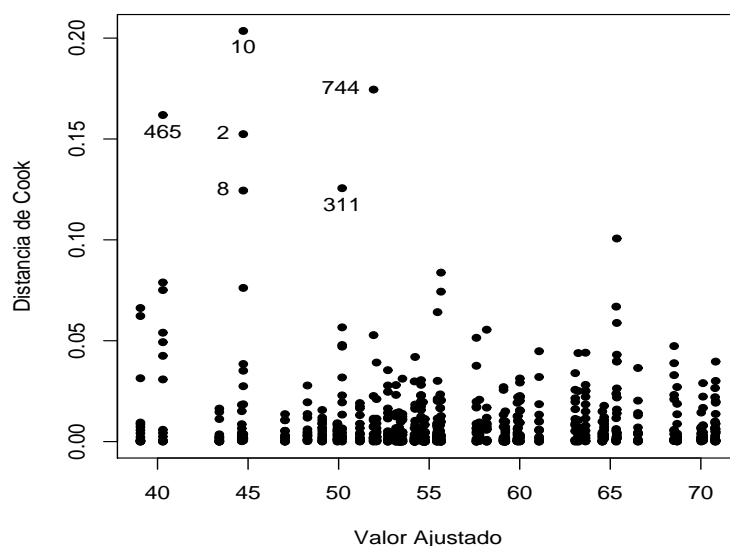


Figura 2.28: Gráfico da distância de Cook contra os valores ajustados referente ao modelo com resposta normal inversa ajustado aos dados sobre *snacks*.

Tabela 2.6
*Vendas projetadas e reais de
vários produtos.*

Projetada	Real	Projetada	Real
5959	5673	527	487
3534	3659	353	463
2641	2565	331	225
1965	2182	290	257
1738	1839	253	311
1182	1236	193	212
667	918	156	166
613	902	133	123
610	756	122	198
549	500	114	99

Como podemos notar pelos gráficos do resíduo de Pearson contra os valores ajustados (Figura 2.30) os dois ajustes não parecem adequados, embora a

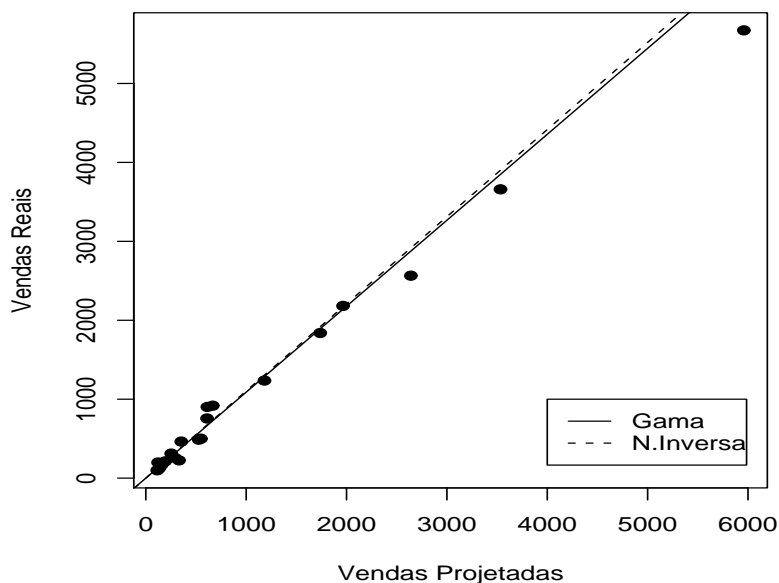


Figura 2.29: Modelos ajustados aos dados sobre vendas projetadas sob erros gama e normal inversa.

suposição de erros gama seja mais razoável do que normal inversa. Por exemplo, as estimativas para o modelo gama são dadas por $\hat{\beta} = 1,089(0,051)$ e $\hat{\phi} = 24,94(7,83)$. Assim, notamos pela estimativa de β que o total projetado de vendas é um bom preditor do total real vendido e a estimativa de ϕ indica que a distribuição do total real de vendas (dado o total projetado) pode ser aproximada por uma normal heteroscedástica.

2.9 Modelagem simultânea da média e da dispersão

Eventualmente o coeficiente de variação pode não ser constante variando com as observações. Smyth (1989) introduziu os modelos lineares generalizados duplos com modelagem conjunta da média e do parâmetro de precisão (ou dispersão) e desenvolveu um processo de estimação baseado no método

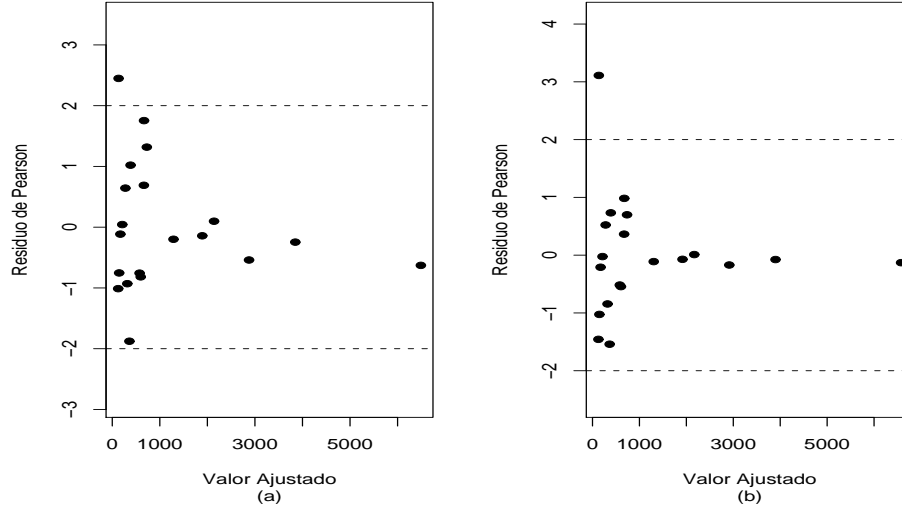


Figura 2.30: Gráficos do resíduo de Pearson contra os valores ajustados referentes aos modelos com resposta gama (a) e normal inversa (b) ajustados aos dados sobre vendas projetadas.

de máxima verossimilhança que descreveremos a seguir. Contudo, outros métodos alternativos de estimação, tais como máxima verossimilhança restrita, foram propostos mais recentemente com o intuito de reduzir o viés das estimativas de máxima verossimilhança, particularmente dos coeficientes do componente de dispersão. Uma discussão a respeito desses métodos pode ser encontrada em Smyth, Huela e Verbyla (2001).

A fim de formalizarmos os MLGs duplos vamos supor que Y_1, \dots, Y_n são variáveis aleatórias independentes com função densidade ou função de probabilidades expressa na forma

$$f(y; \theta_i, \phi_i) = \exp[\phi_i\{y\theta_i - b(\theta_i)\} + c(y, \phi_i)],$$

em que $c(y, \phi_i) = d(\phi_i) + \phi_i a(y) + u(y)$. Essa decomposição, como vimos na Seção 1.7.2, vale somente para as distribuições normal, normal inversa e

gama da família exponencial. Além disso, vamos supor que

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad \text{e} \quad h(\phi_i) = \lambda_i = \mathbf{z}_i^T \boldsymbol{\gamma},$$

em que $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ e $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^T$ contêm valores de variáveis explicativas e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^T$ são os parâmetros a serem estimados.

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$, então o logaritmo da função de verossimilhança fica dado por

$$\begin{aligned} L(\boldsymbol{\theta}) &= \sum_{i=1}^n [\phi_i \{y_i \theta_i - b(\theta_i)\} + d(\phi_i) + \phi_i a(y_i) + u(y_i)] \\ &= \sum_{i=1}^n \{\phi_i t_i + d(\phi_i) + u(y_i)\}, \end{aligned} \quad (2.7)$$

em que $t_i = y_i \theta_i - b(\theta_i) + a(y_i)$. Portanto, se θ_i for fixado a expressão (2.7) coincide com o logaritmo da função de verossimilhança de um modelo da família exponencial com respostas independentes T_1, \dots, T_n (valores observados t_1, \dots, t_n), parâmetros canônicos ϕ_1, \dots, ϕ_n e parâmetro de dispersão igual a 1. Pelas propriedades da família exponencial segue que

$$\mu_{T_i} = E(T_i) = -d'(\phi_i) \quad \text{e} \quad \text{Var}(T_i) = -d''(\phi_i).$$

Essas quantidades são descritas na Tabela 2.7 para as distribuições normal, normal inversa e gama. Os resultados acima podem ser obtidos, alternativamente, aplicando-se condições usuais de regularidade no logaritmo da verossimilhança dado em (2.7).

Conforme observado por Verbyla (1989) para as distribuições normal e normal inversa se chamarmos $D_i = -2T_i$ ($i = 1, \dots, n$) segue que

$$E(D_i) = \phi_i^{-1} \quad \text{e} \quad \text{Var}(D_i) = \frac{E^2(D_i)}{\nu},$$

em que $\nu = \frac{1}{2}$. Portanto, a expressão (2.7) pode ser interpretada para os modelos com resposta normal e normal inversa como um MLG de respostas independentes D_1, \dots, D_n com distribuição gama de médias $\phi_1^{-1}, \dots, \phi_n^{-1}$, respectivamente, e parâmetro de dispersão $\nu^{-1} = 2$. Assim, para θ_i fixado, os parâmetros da dispersão podem ser estimados alternativamente através de um MLG com respostas independentes gama, função de ligação $h(\cdot)$ e parâmetro de dispersão igual a 2.

Tabela 2.7

Derivação de algumas quantidades para distribuições da família exponencial.

	Normal	Normal inversa	Gama
t_i	$y_i \mu_i - \frac{1}{2}(\mu_i^2 + y_i^2)$	$-\{y_i/2\mu_i^2 - \mu_i^{-1} + (2y_i)^{-1}\}$	$\log(y_i/\mu_i) - y_i/\mu_i$
$d(\phi)$	$\frac{1}{2}\log\phi$	$\frac{1}{2}\log\phi$	$\phi\log\phi - \log\Gamma(\phi)$
$d'(\phi)$	$(2\phi)^{-1}$	$(2\phi)^{-1}$	$(1 + \log\phi) - \psi(\phi)$
$d''(\phi)$	$-(2\phi^2)^{-1}$	$-(2\phi^2)^{-1}$	$\phi^{-1} - \psi'(\phi)$

2.9.1 Estimação

A função escore e a matriz de informação de Fisher para β podem ser obtidas facilmente seguindo os passos da Seção 1.5.1. Assim, obtemos

$$\begin{aligned} \mathbf{U}_\beta &= \mathbf{X}^T \Phi \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}) \quad \text{e} \\ \mathbf{K}_{\beta\beta} &= \mathbf{X}^T \Phi \mathbf{W} \mathbf{X}, \end{aligned}$$

em que \mathbf{X} é uma matriz $n \times p$ de linhas \mathbf{x}_i^T ($i = 1, \dots, n$), $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$ com pesos $\omega_i = (d\mu_i/d\eta_i)^2/V_i$, $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$, $\Phi = \text{diag}\{\phi_1, \dots, \phi_n\}$, $\mathbf{y} = (y_1, \dots, y_n)^T$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$.

Para obtermos a função escore para o parâmetro $\boldsymbol{\gamma}$, calculamos inicialmente a derivada

$$\begin{aligned}\partial L(\boldsymbol{\theta})/\partial \gamma_j &= \sum_{i=1}^n \left\{ \frac{d\phi_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \gamma_j} t_i + d'(\phi_i) \frac{d\phi_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \gamma_j} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{1}{h'(\phi_i)} z_{ij} t_i + d'(\phi_i) \frac{1}{h'(\phi_i)} z_{ij} \right\} \\ &= \sum_{i=1}^n \frac{z_{ij}}{h'(\phi_i)} \{t_i + d'(\phi_i)\},\end{aligned}$$

em que $h'(\phi_i) = d\lambda_i/d\phi_i$. Portanto, em forma matricial obtemos

$$\mathbf{U}_\gamma = \mathbf{Z}^T \mathbf{H}_\gamma^{-1} (\mathbf{t} - \boldsymbol{\mu}_T),$$

em que $\mathbf{H}_\gamma = \text{diag}\{h'(\phi_1), \dots, h'(\phi_n)\}$, $\mathbf{t} = (t_1, \dots, t_n)^T$ e $\boldsymbol{\mu}_T = (E(T_1), \dots, E(T_n))^T = (-d'(\phi_1), \dots, -d'(\phi_n))^T$.

Para obtermos a matriz de informação de Fisher para o parâmetro $\boldsymbol{\gamma}$ precisamos das derivadas

$$\begin{aligned}\partial^2 L(\boldsymbol{\theta})/\partial \gamma_j \partial \gamma_\ell &= - \sum_{i=1}^n \frac{z_{ij}}{\{h'(\phi_i)\}^2} \left[d''(\phi_i) h(\phi_i) \frac{d\phi_i}{d\lambda_i} z_{i\ell} - h''(\phi_i) \{t_i + d'(\phi_i)\} \frac{d\phi_i}{d\lambda_i} z_{i\ell} \right] \\ &= - \sum_{i=1}^n \frac{z_{ij} z_{i\ell}}{\{h'(\phi_i)\}^2} \left[d''(\phi_i) - \frac{h''(\phi_i)}{h'(\phi_i)} \{t_i + d'(\phi_i)\} \right]\end{aligned}$$

e cujos valores esperados ficam dados por

$$E \left\{ - \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_\ell} \right\} = - \sum_{i=1}^n \frac{d''(\phi_i)}{\{h'(\phi_i)\}^2} z_{ij} z_{i\ell}.$$

Logo, em forma matricial obtemos

$$\mathbf{K}_{\gamma\gamma} = \mathbf{Z}^T \mathbf{P} \mathbf{Z},$$

em que $\mathbf{P} = \mathbf{V}_\gamma \mathbf{H}_\gamma^{-2}$, $\mathbf{V}_\gamma = \text{diag}\{-d''(\phi_1), \dots, -d''(\phi_n)\}$. Devido à ortogonalidade entre os parâmetros θ_i e ϕ_i , segue diretamente a ortogonalidade entre

β e γ . Assim, a matriz de informação de Fisher para θ é bloco diagonal $\mathbf{K}_{\theta\theta} = \text{diag}\{\mathbf{K}_{\beta\beta}, \mathbf{K}_{\gamma\gamma}\}$.

Similarmente aos MLGs podemos desenvolver um processo iterativo escore de Fisher para encontrar as estimativas de máxima verossimilhança $\hat{\beta}$ e $\hat{\gamma}$. Após algumas manipulações algébricas obtemos o proceso iterativo

$$\beta^{(m+1)} = (\mathbf{X}^T \Phi^{(m)} \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \Phi^{(m)} \mathbf{W}^{(m)} \mathbf{y}^{*(m)} \quad \text{e} \quad (2.8)$$

$$\gamma^{(m+1)} = (\mathbf{Z}^T \mathbf{P}^{(m)} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{P}^{(m)} \mathbf{z}^{*(m)}, \quad (2.9)$$

em que $\mathbf{y}^* = \mathbf{X}\beta + \mathbf{W}^{-1/2} \mathbf{V}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, $\mathbf{z}^* = \mathbf{Z}\gamma + \mathbf{V}_\gamma^{-1} \mathbf{H}_\gamma(\mathbf{t} - \boldsymbol{\mu}_T)$ e $m = 0, 1, 2, \dots$. Conforme mencionado por Smyth (1989) o processo iterativo (2.8)-(2.9) pode ser resolvido alternando-se as duas equações até a convergência. Pode-se iniciar o processo iterativo (2.8) com as estimativas do MLG com ϕ_i comum a todas as observações.

Sob as condições de regularidade apresentadas na Seção 1.6.3 temos que para n grande $\hat{\beta} \sim N_p(\beta, \mathbf{K}_{\beta\beta}^{-1})$ e $\hat{\gamma} \sim N_q(\gamma, \mathbf{K}_{\gamma\gamma}^{-1})$, respectivamente. Além disso, devido à ortogonalidade entre β e γ segue a independência assintótica entre $\hat{\beta}$ e $\hat{\gamma}$.

2.9.2 Métodos de diagnóstico

Faremos nesta subseção uma adaptação de alguns procedimentos de diagnóstico para a classe dos MLGs duplos.

Resíduos

Na classe dos MLGs duplos podemos definir desvios para a média e para a precisão, respectivamente. O desvio para a média assume a mesma expressão da classe dos MLGs em que somente a média é ajustada, com ϕ_i no lugar de ϕ . Denotaremos esse desvio por $D_1^*(\mathbf{y}; \hat{\boldsymbol{\mu}}, \phi) = \sum_{i=1}^n d_1^{*2}(y_i; \hat{\mu}_i, \phi_i)$,

em que $d_1^{*2}(y_i; \hat{\mu}_i, \phi_i) = 2\phi_i[y_i(\tilde{\theta}_i - \hat{\theta}_i) + \{b(\hat{\theta}_i) - b(\tilde{\theta}_i)\}]$. Para ϕ_i grande $\forall i$ o desvio $D_1^*(\mathbf{y}; \hat{\boldsymbol{\mu}}, \boldsymbol{\phi})$ pode ser comparado com os percentis da distribuição qui-quadrado com $(n - p)$ graus de liberdade. Para o modelo normal heteroscedástico o desvio para a média fica dado por $D_1^*(\mathbf{y}; \hat{\boldsymbol{\mu}}, \boldsymbol{\phi}) = \sum_{i=1}^n \sigma_i^{-2}(y_i - \hat{y}_i)^2$. Na prática deveremos substituir ϕ_i por $\hat{\phi}_i = h^{-1}(\hat{\lambda}_i) = \mathbf{z}_i^T \hat{\boldsymbol{\gamma}}$.

O resíduo Studentizado, no modelo normal heteroscedástico, assume a forma

$$t_i^* = \frac{y_i - \hat{y}_i}{\hat{\sigma}_i \sqrt{1 - \hat{h}_{ii}}},$$

em que $\hat{h}_{ii} = \hat{\sigma}_i^2 \mathbf{x}_i^T (\mathbf{X}^T \hat{\boldsymbol{\Phi}} \mathbf{X})^{-1} \mathbf{x}_i$ com $\boldsymbol{\Phi} = \text{diag}\{\sigma_1^{-2}, \dots, \sigma_n^{-2}\}$. Para os demais MLGs duplos o resíduo componente do desvio para a média fica dado por

$$t_{D_{1i}} = \frac{d_1^*(y_i; \hat{\mu}_i, \hat{\phi}_i)}{\sqrt{1 - \hat{h}_{ii}}},$$

em que $d_1^*(y_i; \hat{\mu}_i, \hat{\phi}_i) = \pm \sqrt{d_1^{*2}(y_i; \hat{\mu}_i, \hat{\phi}_i)}$, o sinal continua sendo o mesmo de $(y_i - \hat{\mu}_i)$ e \hat{h}_{ii} é o i -ésimo elemento da diagonal principal da matriz

$$\hat{\mathbf{H}} = \hat{\boldsymbol{\Phi}}^{1/2} \hat{\mathbf{W}}^{1/2} \mathbf{X} (\mathbf{X}^T \hat{\boldsymbol{\Phi}} \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\boldsymbol{\Phi}}^{1/2} \hat{\mathbf{W}}^{1/2},$$

ou seja,

$$\hat{h}_{ii} = \hat{\phi}_i \hat{\omega}_i \mathbf{x}_i^T (\mathbf{X}^T \hat{\boldsymbol{\Phi}} \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i.$$

Sugerimos o gráfico normal de probabilidades para $t_{D_{1i}}$ e o gráfico de $t_{D_{1i}}$ contra os valores ajustados.

Denotaremos por $D_2^*(\mathbf{y}; \hat{\boldsymbol{\phi}}, \boldsymbol{\mu}) = \sum_{i=1}^n d_2^{*2}(y_i; \hat{\phi}_i, \mu_i)$ o desvio para a precisão, em que $d_2^{*2}(y_i; \hat{\phi}_i, \mu_i) = 2[t_i(\tilde{\phi}_i - \hat{\phi}) + \{d(\tilde{\phi}_i) - d(\hat{\phi}_i)\}]$, $\tilde{\phi}_i$ é solução para ϕ_i sob o modelo saturado sendo dada por $d'(\tilde{\phi}_i) = -t_i$. Para os modelos com resposta normal e normal inversa temos que $\tilde{\phi}_i = -(2t_i)^{-1}$. Já para modelos com resposta gama $\tilde{\phi}_i$ é a solução da equação $\{\psi(\tilde{\phi}_i) - \log \tilde{\phi}_i + 1\} = t_i$. Aqui

também para ϕ_i grande $\forall i$ o desvio $D_2^*(\mathbf{y}; \hat{\boldsymbol{\phi}}, \boldsymbol{\mu})$ pode ser comparado com os percentis da distribuição qui-quadrado com $(n - q)$ graus de liberdade.

O resíduo componente do desvio para a precisão fica dado por

$$t_{D_{2i}} = \frac{d_2^*(y_i; \hat{\phi}_i, \hat{\mu}_i)}{\sqrt{1 - \hat{r}_{ii}}},$$

em que $d_2^*(y_i; \hat{\phi}_i, \hat{\mu}_i) = \pm \sqrt{d_2^{*2}(y_i; \hat{\phi}_i, \hat{\mu}_i)}$, o sinal sendo o mesmo de $\{\hat{t}_i + d'(\hat{\phi}_i)\}$ e \hat{r}_{ii} é o i -ésimo elemento da diagonal principal da matriz

$$\hat{\mathbf{R}} = \hat{\mathbf{P}}^{1/2} \mathbf{Z} (\mathbf{Z}^T \hat{\mathbf{P}} \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\mathbf{P}}^{1/2},$$

ou seja,

$$\hat{r}_{ii} = \hat{p}_i \mathbf{z}_i^T (\mathbf{Z}^T \hat{\mathbf{P}} \mathbf{Z})^{-1} \mathbf{z}_i.$$

Note que $p_i = -d''(\phi_i) \{h'(\phi_i)\}^{-2}$. Por exemplo, para ligação logarítmica temos que $h(\phi_i) = \log \phi_i$ então $h'(\phi_i) = \phi_i^{-1}$ e portanto $p_i = -\phi_i^2 d''(\phi_i)$. Assim, para os modelos com resposta normal e normal inversa segue que $p_i = \phi_i^2 (2\phi_i^2)^{-1} = \frac{1}{2}$ e para os modelos com resposta gama $p_i = \phi_i \{\phi_i \psi'(\phi_i) - 1\}$.

Sugerimos o gráfico normal de probabilidades para $t_{D_{2i}}$ e o gráfico de $t_{D_{2i}}$ contra os valores ajustados.

Influência

Para avaliar a sensibilidade das estimativas dos parâmetros que modelam a média podemos usar a medida de influência LD_i definida na Seção 1.10.3 com $\hat{\phi}_i$ no lugar de $\hat{\boldsymbol{\phi}}$, que será definida por

$$LD_i^\beta = \left\{ \frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}} \right\} t_{S_i}^2,$$

em que

$$t_{S_i} = \frac{\sqrt{\hat{\phi}_i} (y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i (1 - \hat{h}_{ii})}}.$$

Gráficos de índices de LD_i^β e \hat{h}_{ii} contra os valores ajustados são recomendados.

Para avaliar a sensibilidade da estimativa $\hat{\gamma}$ quando a i -ésima observação é deletada usaremos uma aproximação de um passo, que é obtida de forma similar à aproximação de uma passo $\hat{\beta}_{(i)}$ descrita na Seção 1.10.3, dada por

$$\hat{\gamma}_{(i)} = \hat{\gamma} - \frac{(\mathbf{Z}^T \hat{\mathbf{P}} \mathbf{Z})^{-1} \mathbf{z}_i \{t_i + d'(\hat{\phi}_i)\}}{h'(\hat{\phi}_i)(1 - \hat{r}_{ii})}, \quad (2.10)$$

em que \hat{r}_{ii} é o i -ésimo elemento da diagonal principal da matriz \mathbf{R} .

Uma medida para avaliar a influência nas estimativas dos parâmetros da precisão fica dada por

$$\begin{aligned} LD_i^\gamma &= (\hat{\gamma}_{(i)} - \hat{\gamma})^T (\mathbf{Z}^T \hat{\mathbf{P}} \mathbf{Z}) (\hat{\gamma}_{(i)} - \hat{\gamma}) \\ &= \left\{ \frac{\hat{r}_{ii}}{1 - \hat{r}_{ii}} \right\} t_{T_i}^2, \end{aligned}$$

em que

$$\begin{aligned} t_{T_i} &= \frac{t_i + d'(\hat{\phi}_i)}{h'(\hat{\phi}_i) \sqrt{\hat{p}_i(1 - \hat{r}_{ii})}} \\ &= \frac{t_i + d'(\hat{\phi}_i)}{\sqrt{-d''(\hat{\phi}_i)(1 - \hat{r}_{ii})}}. \end{aligned}$$

Gráficos de índices de LD_i^γ e \hat{r}_{ii} contra os valores ajustados são recomendados.

Para os modelos com resposta normal e com resposta normal inversa o resíduo t_{T_i} assume a forma

$$t_{T_i} = \frac{t_i + (2\hat{\phi}_i)^{-1}}{(\sqrt{2\hat{\phi}_i})^{-1} \sqrt{1 - \hat{r}_{ii}}},$$

e para modelos com resposta gama obtemos

$$t_{T_i} = \frac{t_i + \{1 + \log \hat{\phi}_i - \psi(\hat{\phi}_i)\}}{\sqrt{\{\psi'(\hat{\phi}_i) - \hat{\phi}_i^{-1}\}(1 - \hat{r}_{ii})}}.$$

Verbyla (1993) apresenta uma aproximação de uma passo para $\hat{\gamma}_{(i)}$ para o caso normal usando um esquema de perturbação específico para modelos normais heteroscedásticos. Para obtermos a aproximação apresentada em (2.10) usamos a ponderação de casos usual para MLGs. Estudos sobre a qualidade da aproximação apresentada em (2.10) ainda não foram desenvolvidos.

2.9.3 Aplicação

Pela análise descritiva apresentada na Seção 2.8.1 sobre o comportamento da força de cisalhamento dos cinco tipos de *snack* ao longo das 20 semanas e também pelo gráfico de perfis para a força de cisalhamento (Figura 2.31) nota-se que o coeficiente de variação não parece ser constante. Assim, a modelagem dupla da média e da precisão pode levar a um ajuste mais satisfatório para o modelo com resposta gama. Dessa forma vamos supor que $Y_{ijk} \sim G(\mu_{ij}, \phi_{ij})$, em que Y_{ijk} denota a força de cisalhamento referente à k -ésima réplica do i -ésimo grupo na j -ésima semana, para $k = 1, \dots, 15$, $j = 2, 4, 6, \dots, 20$ e $i = 1(A), 2(B), 3(C), 4(D)$ e $E(5)$, com parte sistemática dada por

$$\begin{aligned}\mu_{ij} &= \beta_0 + \beta_i + \beta_6 \text{semana}_j + \beta_7 \text{semana}_j^2 \text{ e} \\ \log(\phi_{ij}) &= \gamma_0 + \gamma_i + \gamma_6 \text{semana}_j + \gamma_7 \text{semana}_j^2,\end{aligned}$$

em que $\beta_1 = 0$ e $\gamma_1 = 0$. Portanto β_0 e γ_0 são os efeitos da forma A, controlando-se pela semana, na média e na precisão, respectivamente, enquanto $\beta_0 + \beta_i$ e $\gamma_0 + \gamma_i$ são os efeitos das demais formas B, C, D e E na média e precisão, respectivamente.

O MLG duplo pode ser ajustado no R através dos seguintes comandos:

```
require(dglm)
fit3.snack = dglm(cisalhamento ~ grupo + s1 + s2,
~ grupo + s1 + s2, family=Gamma(link=identity))
```

```
summary(fit3.snack).
```

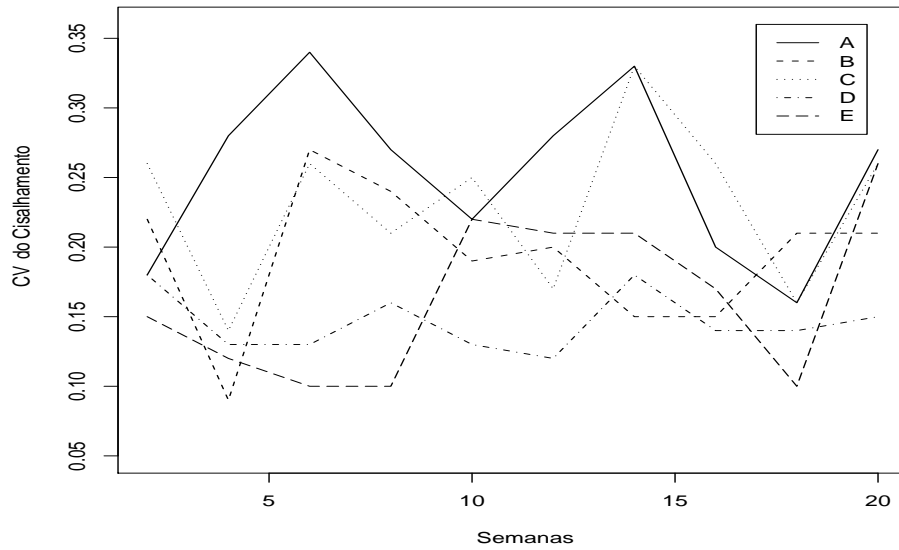


Figura 2.31: Perfis do CV da força de cisalhamento segundo as semanas e os grupos.

Tabela 2.7

Estimativas dos parâmetros referentes ao MLG duplo com resposta gama ajustado aos dados sobre snacks.

Efeito	Média		Dispersão	
	Estimativa	E/E.Padrão	Estimativa	E/E.Padrão
Constante	36,990	11,53	1,560	7,27
Grupo B	-10,783	-6,40	0,477	2,95
Grupo C	-3,487	-1,98	0,050	0,31
Grupo D	-14,829	-9,18	0,815	5,05
Grupo E	-15,198	-9,54	0,817	5,06
Semana	5,198	9,88	0,155	3,91
Semana ²	-0,189	-8,88	-0,005	-2,99

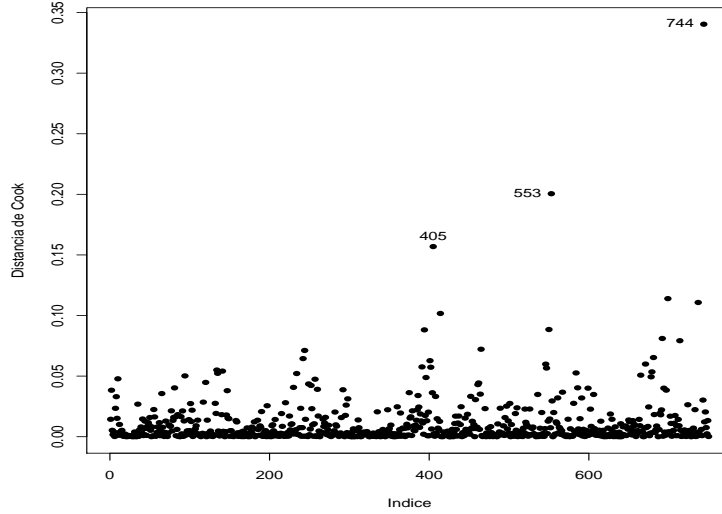


Figura 2.32: Gráfico da distância de Cook para as estimativas dos coeficientes da média referente ao MLG duplo com resposta gama ajustado aos dados sobre *snacks*.

Observarmos que a *library dglm* faz o ajuste de $\log(\phi_i^{-1})$, ou seja da dispersão, sendo necessário fazermos as adaptações nos modelos com resposta gama e normal inversa para obtermos $\log(\phi_i)$, ajuste da precisão. Em particular no caso de modelos normais heteroscedásticos tem-se diretamente o ajuste de $\log\sigma_i^2$, em que σ_i^2 é a variância.

Na Tabela 2.7 são apresentadas as estimativas com os respectivos erros padrão dos parâmetros da média e da dispersão.

Podemos notar pelas estimativas dos parâmetros da média as mesmas tendências observadas na Figura 2.22 para o modelo com resposta normal inversa. O grupo A tem a maior média para a força de cisalhamento enquanto os grupos D e E têm as menores médias. Com relação às estimativas dos parâmetros da dispersão nota-se que a variabilidade (no sentido do coeficiente

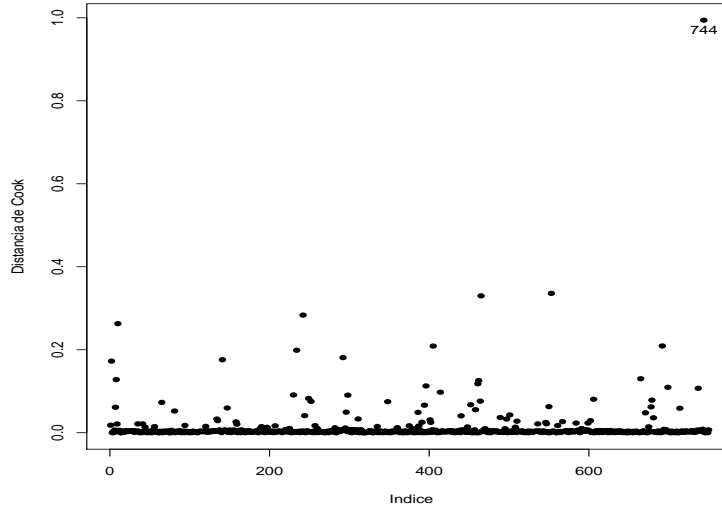


Figura 2.33: Gráfico da distância de Cook para as estimativas dos coeficientes da dispersão referente ao MLG duplo com resposta gama ajustado aos dados sobre *snacks*.

de variação) depende do tempo de forma quadrática e que os grupos A e C apresentam maior variabilidade enquanto os grupos D e E apresentam as menores variabilidades. Essas tendências estão de acordo com os gráficos das Figuras 2.22, 2.23 e 2.24.

Nota-se ainda que os mesmos efeitos que são significativos para os parâmetros da média são também significativos para os parâmetros da dispersão. Apenas três observações, #430, #595 e #744, aparecem como possivelmente influentes nos parâmetros da média e da dispersão, como pode ser observado pelas Figuras 2.32 e 2.33. A eliminação desses pontos não muda a inferência. Pelos gráficos normais de probabilidades para o resíduo componente do desvio para a média e para a dispersão apresentados nas Figuras 2.34 e 2.35, respectivamente, não há indícios de inadequação do MLG duplo.

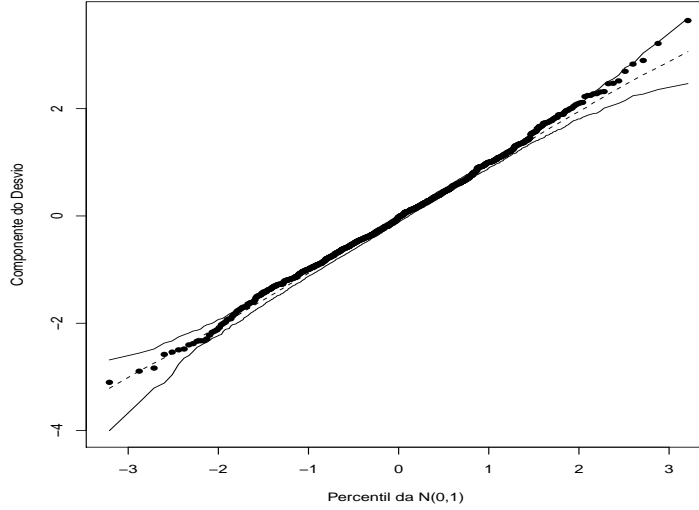


Figura 2.34: Gráfico normal de probabilidades para o resíduo componente do desvio para a média referente ao MLG duplo com resposta gama ajustado aos dados sobre *snacks*.

2.10 Exercícios

1. Seja $Y \sim G(\mu, \phi)$ e considere a variável aleatória $\log Y$. Use a condição de regularidade $E(U_\phi) = 0$ para mostrar que $E(\log Y) = \log \mu - \log \phi + \psi(\phi)$, em que $U_\phi = \partial L(\mu, \phi) / \partial \phi$.
2. Seja $Y \sim NI(\mu, \phi)$ e considere a variável aleatória Y^{-1} . Use a condição de regularidade $E(U_\phi) = 0$ para mostrar que $E(Y^{-1}) = \mu^{-1} + \phi^{-1}$, em que $U_\phi = \partial L(\mu, \phi) / \partial \phi$.
3. Considere o logaritmo da função de verossimilhança dado em (2.7). Obtenha $E(T_i)$ e $\text{Var}(T_i)$ a partir das condições usuais de regularidade

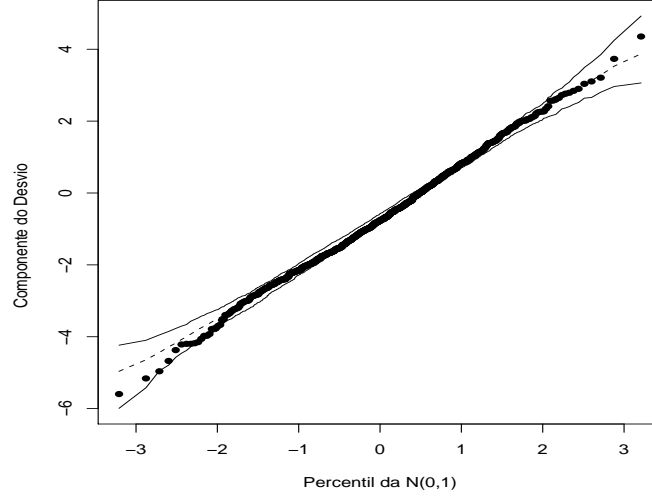


Figura 2.35: Gráfico normal de probabilidades para o resíduo componente do desvio para a dispersão referente ao MLG duplo com resposta gama ajustado aos dados sobre *snacks*.

$$E\{\partial L(\boldsymbol{\theta})/\partial \phi_i\} = 0 \text{ e } E\{\partial^2 L(\boldsymbol{\theta})/\partial \phi_i^2\} = -E\{(\partial L(\boldsymbol{\theta})/\partial \phi_i)^2\}.$$

4. Mostre que o desvio da distribuição gama para o caso i.i.d., ou seja $Y_i \sim G(\mu, \phi)$, é dado por $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2n\phi \log(\bar{y}/\tilde{y})$, em que \tilde{y} é a média geométrica das observações, isto é $\tilde{y} = (\prod_{i=1}^n y_i)^{1/n}$.
5. Sejam Y_1, \dots, Y_n variáveis aleatórias independentes tais que $Y_i \sim \text{NI}(\mu_i, \phi)$ com parte sistemática dada por $g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$. (i) Calcule a estimativa de máxima verossimilhança de ϕ e $\text{Var}(\hat{\phi})$. (ii) Mostre que o critério de Akaike equivale a minimizar

$$\text{AIC} = n \log \left\{ \frac{D(\mathbf{y}; \hat{\boldsymbol{\mu}})}{n} \right\} + 2p,$$

em que $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / (y_i \hat{\mu}_i^2)$.

6. Sejam $Y_i \sim \text{FE}(\mu_1, \phi_1)$, $i = 1, \dots, m$, e $Y_i \sim \text{FE}(\mu_2, \phi_2)$, $i = m + 1, \dots, n$, variáveis aleatórias mutuamente independentes. Encontre a estimativa comum de máxima verossimilhança para ϕ_1 e ϕ_2 sob a hipótese $H_0 : \phi_1 = \phi_2$. Particularize para os casos gama e normal inversa.
7. (Lawless, 1982, p. 338). Na tabela abaixo são apresentados os resultados de um experimento em que a resistência (em horas) de um determinado tipo de vidro foi avaliada segundo quatro níveis de voltagem (em kilovolts) e duas temperaturas (em graus Celsius). Esses dados estão também disponíveis no arquivo **vidros.dat**. Na primeira coluna do arquivo tem-se o tempo de resistência, na segunda coluna a voltagem (1: 200kV, 2: 250kV, 3: 300kV e 4: 350kV) e na terceira coluna a temperatura (1: 170°C e 2: 180°C). Seja Y_{ijk} o tempo de resistência da k -ésima amostra de vidro submetida à i -ésima temperatura e à j -ésima voltagem.

Faça inicialmente uma análise descritiva dos dados, por exemplo apresentando os perfis médios da resistência segundo a voltagem para os dois níveis de temperatura. Cacule também para cada casela algumas medidas descritivas tais como média, desvio padrão e coeficiente de variação. Comente.

O interesse principal desse estudo é comparar as resistências médias, denotadas por μ_{ij} , $i = 1, 2$ e $j = 2, 3, 4$. É usual neste tipo de estudo assumir respostas com alguma distribuição assimétrica. Assim, vamos supor que $Y_{ijk} \sim G(\mu_{ij}, \phi)$. Considere inicialmente uma parametrização tipo casela de referência sem interação, em que $\mu_{11} = \alpha$, $\mu_{1j} = \alpha + \beta_j$, $\mu_{21} = \alpha + \gamma$ e $\mu_{2j} = \alpha + \gamma + \beta_j$, $j = 2, 3, 4$.

Temperatura (°C)	Voltagem(kV)			
	200	250	300	350
170	439	572	315	258
	904	690	315	258
	1092	904	439	347
	1105	1090	628	588
180	959	216	241	241
	1065	315	315	241
	1065	455	332	435
	1087	473	380	455

Verifique se é possível incluir a interação entre voltagem e temperatura. Procure responder com o modelo final de que forma os níveis de voltagem e temperatura afetam o tempo médio de resistência dos vidros. Apresente, por exemplo, os perfis médios ajustados e interprete a estimativa de dispersão. Faça também uma análise de diagnóstico.

8. Sejam Y_i , $i = 1, \dots, n$, variáveis aleatórias i.i.d. com distribuição normal inversa de média μ e parâmetro de precisão ϕ .
 - (i) Como fica a estatística da razão de verossimilhanças para testar $H_0 : \phi = 1$ contra $H_1 : \phi \neq 1$?
 - (ii) Qual a distribuição nula assintótica da estatística do teste?
9. Sejam Y_1, \dots, Y_n variáveis aleatórias independentes tais que $Y_i \sim G(\mu_i, \phi)$ com parte sistemática dada por $\log \mu_i = \beta_0 + \beta_1(x_i - \bar{x})$. Responda aos itens abaixo:
 - (a) como fica a matriz de informação de Fisher para $\boldsymbol{\theta} = (\beta_0, \beta_1, \phi)^T$ e a variância assintótica de $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\phi}$?

- (b) Como fica o teste de escore para testar $H_0 : \phi = 1$ contra $H_1 : \phi \neq 1$?
- (c) Mostre que a estatística do teste de escore para testar as hipóteses $H_0 : \beta_0 = 1, \beta_1 = 0$ contra $H_1 : \beta_0 \neq 1$ ou $\beta_1 \neq 0$ pode ser expressa na forma

$$\xi_{SR} = \frac{\hat{\phi}^0}{e^2} \left[n(\bar{y} - e)^2 + \frac{\{\sum_{i=1}^n (x_i - \bar{x})(y_i - e)\}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right].$$

Qual a distribuição nula assintótica de ξ_{SR} ?

10. Supor Y_1, \dots, Y_n variáveis aleatórias independentes tais que $Y_i \sim \text{FE}(\mu, \phi_i)$ com $\log \phi_i = \alpha + \gamma z_i$. Responda às seguintes questões:
- (i) como fica a matriz modelo \mathbf{Z} ?
 - (ii) Calcule a variância assintótica de $\hat{\gamma}$.
 - (iii) Como fica a estatística de escore para testar $H_0 : \gamma = 0$ contra $H_1 : \gamma \neq 0$? Qual a distribuição nula assintótica da estatística do teste?
11. Supor $Y_i \stackrel{\text{ind}}{\sim} N(\mu, \sigma_i)$, em que $\log(\sigma_i) = \gamma_0 + \gamma_1 z_i$, em que σ_i denota o desvio padrão de Y_i , para $i = 1, \dots, n$. Obter \mathbf{U}_γ e $\mathbf{K}_{\gamma\gamma}$. Como fica a estimação de μ e γ ? Obtenha a estatística do teste da razão de verossimilhanças para testar $H_0 : \gamma_1 = 0$ contra $H_1 : \gamma_1 \neq 0$. Qual a distribuição nula assintótica da estatística do teste?
12. Supor $Y_i \stackrel{\text{iid}}{\sim} \text{NI}(\mu, \phi)$, para $i = 1, \dots, n$, em que $\gamma = \log(\phi)$. Obter a estimativa de máxima verossimilhança $\hat{\gamma}$ (dado $\hat{\phi}$) e $\mathbf{K}_{\gamma\gamma}$. Como fica a estatística do teste de Wald para testar $H_0 : \gamma = 0$ contra $H_1 : \gamma \neq 0$?
13. Sejam Y_1, \dots, Y_n variáveis aleatórias independentes tais que $Y_i \sim \text{NI}(\mu_i, \phi)$ com $\mu_i^{-1} = \alpha$. Encontre $\hat{\alpha}$ e $\text{Var}(\hat{\alpha})$. Como fica a estatística de Wald

para testar $H_0 : \alpha = 1$ contra $H_1 : \alpha \neq 1$? Qual a distribuição nula assintótica da estatística do teste?

14. Supor Y_1, \dots, Y_n variáveis aleatórias independentes tais que $Y_i \sim \text{NI}(\mu_i, \phi)$ e $\sqrt{\mu_i} = \eta_i^{-1}$ com $\eta_i = \alpha + \beta(x_i - \bar{x})$, em que $\bar{x} = n^{-1} \sum_{i=1}^n x_i$, $i = 1, \dots, n$. Responda às seguintes questões:

- (i) como fica a matriz modelo \mathbf{X} ?
- (ii) Calcule as variâncias assintóticas $\text{Var}(\hat{\alpha})$ e $\text{Var}(\hat{\beta})$. Calcule $\text{Cov}(\hat{\alpha}, \hat{\beta})$ e comente.
- (iii) Como fica a estatística de Wald para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$? Qual a distribuição nula assintótica da estatística do teste?

15. Supor $Y_i \stackrel{\text{ind}}{\sim} G(\mu_i, \sigma_i)$, em que $\log(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ e $\log(\sigma_i) = \lambda_i = \mathbf{z}_i^\top \boldsymbol{\gamma}$, em que σ_i denota o coeficiente de variação de Y_i , para $i = 1, \dots, n$. Obter \mathbf{U}_β , \mathbf{U}_γ , $\mathbf{K}_{\beta\beta}$ e $\mathbf{K}_{\gamma\gamma}$ e desenvolva um processo iterativo duplo para obter as estimativas de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$.

16. (Cordeiro, Paula e Botter, 1994). Sejam $Y_i, i = 1, \dots, n$, variáveis aleatórias independentes com distribuição gama de média μ_i e parâmetro de precisão ϕ . Mostre que a estatística da razão de verossimilhanças para testar $H_0 : \phi = 1$ contra $H_1 : \phi \neq 1$ vale

$$\xi_{RV} = 2n[\log \hat{\phi} - \log \Gamma(\hat{\phi}) - (\hat{\phi} - 1)\{1 - \psi(\hat{\phi})\}],$$

em que $\Gamma(\phi)$ é a função gama e $\psi(\phi)$ é a função digama. Use o resultado $\log(\hat{\phi}) - \psi(\hat{\phi}) = \bar{D}/2$, em que $\bar{D} = \sum_{i=1}^n D(y_i; \hat{\mu}_i)/n$ denota o desvio médio do modelo correspondente.

17. Supor Y_{ij} variáveis aleatórias mutuamente independentes tais que $Y_{ij} \sim G(\mu_i, \phi)$ para $i = 1, 2$ e $j = 1, \dots, m$, sendo $\log \mu_1 = \alpha - \beta$ e $\log \mu_2 = \alpha + \beta$. (i) Obtenha a matrix modelo \mathbf{X} . (ii) Expresse em forma fechada as estimativas de máxima verossimilhança $\hat{\alpha}$ e $\hat{\beta}$. (iii) Calcule as variâncias assintóticas $\text{Var}(\hat{\alpha})$ e $\text{Var}(\hat{\beta})$ e mostre que $\text{Cov}(\hat{\alpha}, \hat{\beta}) = 0$. (iv) Como fica o teste de escore para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$? Qual a distribuição nula assintótica da estatística do teste?
18. No arquivo **energy.dat** estão descritos os dados referentes ao consumo de energia em 53 domicílios (Montgomery, Peck e Vining, 2001, pgs. 175-178) em que o total de energia consumido num determinado mês (em kilowatts-hora) é utilizado para explicar a demanda de energia na hora de pico. Faça inicialmente uma análise descritiva dos dados. Use um modelo de regressão normal linear (com erros homocedásticos) para explicar a demanda média no horário de pico através do consumo mensal de energia. Proponha, alternativamente, modelos com erros heteroscedásticos. Compare os ajustes e interprete os coeficientes estimados do modelo escolhido.
19. No arquivo **restaurante.dat** estão descritos os faturamentos anuais bem como os gastos com publicidade (em mil USD) de uma amostra aleatória de 30 restaurantes (Montgomery, Peck e Vining, 2001, pgs. 197-200). O objetivo principal é tentar relacionar o faturamento médio com o gasto com publicidade. Inicialmente faça uma análise descritiva dos dados, em particular o diagrama de dispersão entre as variáveis.
- Tente ajustar inicialmente um modelo de regressão normal linear entre faturamento e gastos e verifique através das técnicas de diagnóstico se existem afastamentos sérios das suposições feitas para o modelo.

Ajustar modelos com erros gama e normal inversa e um modelo normal heteroscedástico. Compare os resultados.

20. (Myers, Montgomery e Vining, 2002, p. 192). A fim de avaliar-se a qualidade de um determinado filme utilizado em máquinas fotográficas, o tempo de duração do filme (em horas) é relacionado com a densidade máxima do filme sob três condições experimentais conforme descrito na tabela abaixo e também no arquivo **dfilme.dat**.

Tempo	D_{max} (72°C)	Tempo	D_{max} (82°C)	Tempo	D_{max} (92°C)
72	3,55	48	3,52	24	3,46
144	3,27	96	3,35	48	2,91
216	2,89	144	2,50	72	2,27
288	2,55	192	2,10	96	1,49
360	2,34	240	1,90	120	1,20
432	2,14	288	1,47	144	1,04
504	1,77	336	1,19	168	0,65

Faça um diagrama de dispersão entre o tempo de duração e a densidade máxima para cada condição experimental e proponha um modelo com resposta gama para ajustar esses dados. Aplique métodos de diagnóstico e interprete as estimativas do modelo selecionado.

21. Considere novamente os dados descritos na Tabela 2.6 sobre a relação entre o valor real e o valor projetado das vendas de 20 produtos. Considere agora um modelo normal heteroscedático tal que

$$y_i = \beta x_i + \epsilon_i \text{ e } \log \sigma_i^2 = \gamma_1 + \gamma_2 x_i,$$

com $\epsilon_i \sim N(0, \sigma_i^2)$. Verifique a adequação desse modelo através de métodos de diagnóstico.

22. O arquivo **claims.dat** contém uma amostra aleatória de 996 apólices de seguros de veículos extraídas do livro de de Jong e Heller (2008) referente ao período 2004-2005. As variáveis do arquivo estão na seguinte ordem : (i) **valorv** (valor do veículo em 10000 dolares australianos), (ii) **expos** (exposição do veículo), (iii) **nsinistros** (número de sinistros no período), (iv) **csinistros** (custo total dos sinistros em dolares australianos), (v) **tipov** (tipo do veículo em 11 categorias), (vi) **idadev** (idade do veículo em 4 categorias), (vii) **sexoc** (sexo do condutor principal), (viii) **areac** (área de residência do condutor principal) e (ix) **idadec** (idade do condutor principal em 6 categorias).

Faça inicialmente uma análise descritiva dos dados e procure agrupar em um número menor de categorias algumas variáveis categóricas. Considere como variável resposta **cmsinistros = csinistros/nsinistros**. Aplique numa primeira etapa modelos com resposta gama e normal inversa com ϕ constante. Faça uma análise de diagnóstico. Numa segunda etapa, se necessário, aplique modelos duplos com resposta gama e normal inversa. Faça também uma análise de diagnóstico. Para o modelo final selecionado interprete os coeficientes estimados.

23. (Feigl e Zelen, 1965). Apresentamos a seguir um conjunto de dados em que pacientes com leucemia foram classificados segundo a ausência ou presença de uma característica morfológica nas células brancas. Pacientes classificados de AG positivo foram aqueles com a presença da característica e pacientes classificados de AG negativo não apresentaram a característica. É apresentado também o tempo de sobrevivência do paciente (em semanas) após o diagnóstico da doença e o número de células brancas (WBC) no momento do diagnóstico. Esses dados estão descritos no arquivo **sobrev.dat**.

Supondo que o tempo de sobrevivência após o diagnóstico segue uma distribuição gama, proponha um modelo para explicar o tempo médio de sobrevivência dados $\log(\text{WBC})$ e $\text{AG}(=1 \text{ positivo}, =0 \text{ negativo})$. Faça uma análise de diagnóstico com o modelo ajustado e interprete as estimativas.

AG Positivo		AG Negativo	
WBC	Tempo	WBC	Tempo
2300	65	4400	56
750	156	3000	65
4300	100	4000	17
2600	134	1500	7
6000	16	9000	16
10500	108	5300	22
10000	121	10000	3
17000	4	19000	4
5400	39	27000	2
7000	143	28000	3
9400	56	31000	8
32000	26	26000	4
35000	22	21000	3
100000	1	79000	30
100000	1	100000	4
52000	5	100000	43
100000	65		

24. No arquivo **rent** do gamlss são descritas 9 variáveis observadas numa amostra aleatória de 1967 unidades habitacionais da cidade de Munich em 1993. Para fins de análise iremos considerar as seguintes variáveis: (i) **R** (valor mensal líquido do aluguel em DM), (ii) **F1** (área útil em m^2), (iii) **A** (ano da construção), (iv) **H** (variável binária referente à existência de aquecimento central, 0: sim, 1: não) e (v) **loc** (qualidade da localização do imóvel, 1: abaixo da média, 2: na média e 3: acima

da média). O arquivo está disponibilizado diretamente no `gamlss`, no entanto é preciso informar que a variável `loc` é categórica através do comando

```
loc=factor(loc).
```

A variável explicativa `A` é considerada contínua. Fazer inicialmente uma análise descritiva dos dados, tais como densidade da variável resposta, boxplots e diagramas de dispersão entre as variáveis explicativas contínuas e a variável resposta. Procure selecionar um modelo gama duplo com ligação logarítmica para explicar o valor médio mensal do aluguel e o coeficiente de dispersão. Fazer uma análise de diagnóstico e interpretar os coeficientes estimados do modelo selecionado.

Capítulo 3

Modelos para Dados Binários

3.1 Introdução

Neste capítulo apresentamos modelos para a análise de dados com resposta binária, isto é, que admite apenas dois resultados. Comumente é chamado de sucesso o resultado mais importante da resposta ou aquele que pretendemos relacionar com as demais variáveis de interesse. É comum encontrarmos situações práticas em que esse tipo de resposta aparece. Como ilustração, seguem alguns exemplos: (i) o resultado do diagnóstico de um exame de laboratório, positivo ou negativo; (ii) o resultado da inspeção de uma peça recém fabricada, defeituosa ou não defeituosa; (iii) a opinião de um eleitor a respeito da implantação do voto distrital, favorável ou contrário; (iv) o resultado de um teste de aptidão aplicado a um estudante, aprovado ou reprovado; (v) o resultado de uma promoção de uma rede de lojas enviando para cada cliente um cupom com desconto, cupom utilizado ou cupom não utilizado num determinado período etc. Há também situações em que apenas duas possibilidades são consideradas de interesse para uma variável contínua, valores menores do que um valor de referência v_0 e valores maiores ou iguais a v_0 . Nesses casos, podemos considerar uma nova variável binária para essas duas

possibilidades. Por exemplo, numa determinada prova de conhecimentos v_0 pode ser a nota mínima para ser aprovado no exame, ou o valor mínimo para um exame de laboratório ser considerado alterado. Assim, variáveis binárias podem surgir naturalmente num experimento ou serem criadas dependendo do interesse do estudo.

Inicialmente, apresentamos neste capítulo uma resenha dos principais métodos clássicos para a análise de tabelas de contingência do tipo 2×2 . Em seguida, introduzimos o modelo de regressão logística para resposta binária e fazemos uma analogia com os métodos tradicionais para tabelas 2×2 . Discutimos também a seleção de modelos logísticos, métodos de diagnóstico, alguns tipos de modelos de dose-resposta, sobredispersão e regressão logística condicional.

3.2 Métodos clássicos: uma única tabela 2×2

Métodos clássicos em tabelas de contingência 2×2 são datados da década de 50. Os primeiros trabalhos foram motivados pelo interesse na inferência de certos parâmetros com grande aplicabilidade na área biomédica, especialmente em Epidemiologia. Vários trabalhos foram publicados durante as décadas de 50 e 60 e até hoje as técnicas desenvolvidas têm sido utilizadas, particularmente na análise descritiva dos dados, antes de um tratamento mais sofisticado através de regressão. Apresentamos nesta seção uma resenha das principais técnicas segundo o ponto de vista inferencial clássico. Embora a metodologia apresentada possa ser aplicada em qualquer área do conhecimento, daremos ênfase para a área biomédica em que tem ocorrido um número maior de aplicações.

3.2.1 Risco relativo

Vamos supor que os indivíduos de uma determinada população sejam classificados segundo um fator com dois níveis, A e B , e a presença ou ausência de uma certa doença, denotados por D e \bar{D} , respectivamente. As proporções populacionais ficam, nesse caso, descritas conforme a tabela abaixo.

Doença	Fator	
	A	B
D	P_1	P_3
\bar{D}	P_2	P_4

Portanto, podemos definir outras quantidades:

$P_1/(P_1 + P_2)$: proporção de indivíduos classificados como doentes no grupo A ;

$P_3/(P_3 + P_4)$: proporção de indivíduos classificados como doentes no grupo B .

A razão entre as duas proporções acima foi denominada por Cornfield (1951) como sendo o risco relativo de doença entre os níveis A e B , ou seja

$$\text{RR} = \frac{P_1/(P_1 + P_2)}{P_3/(P_3 + P_4)} = \frac{P_1(P_3 + P_4)}{P_3(P_1 + P_2)}. \quad (3.1)$$

Cornfield (1951) também notou que se a doença for rara ($P_1 \ll P_2$ e $P_3 \ll P_4$) a quantidade (3.1) assume a forma simplificada

$$\psi = \frac{P_1 P_4}{P_3 P_2}, \quad (3.2)$$

a qual denominou *Odds Ratio*, que para nós será denominada razão de chances. Muitas vezes é comum ψ ser chamado de risco relativo, embora isso

somente seja válido quando P_1 e P_3 forem muito pequenos. A grande vantagem do uso de ψ é a facilidade inferencial tanto na abordagem tradicional como na abordagem através de regressão.

Como em geral a porcentagem de indivíduos doentes é muito menor do que a porcentagem de não doentes, é bastante razoável num estudo cujo objetivo é avaliar a associação entre algum fator particular e uma certa doença, que a quantidade de doentes na amostra seja a maior possível. Assim, a amostragem retrospectiva, em que os indivíduos são escolhidos separadamente nos estratos D e \bar{D} , pode ser mais conveniente do que os demais procedimentos amostrais. Um cuidado, entretanto, devemos ter nesses estudos. É importante que os doentes (casos) sejam comparáveis aos não doentes (controles) segundo outros fatores (fatores potenciais de confundimento), possivelmente associados com a doença. Nos estudos prospectivos, em que a amostragem é feita nos estratos A e B , esse tipo de problema pode ser controlado, embora em geral seja necessário um longo período até a obtenção de um número suficiente de doentes para uma análise estatística mais representativa.

Como as inferências para os estudos retrospectivos e prospectivos são idênticas, trataremos apenas o caso retrospectivo. Assim, assumimos que no estrato D são amostrados n_1 indivíduos e no estrato \bar{D} são amostrados n_2 indivíduos. O número observado de indivíduos com presença de A nos estratos D e \bar{D} será denotado por y_1 e y_2 , respectivamente. Os dados resultantes dessa amostragem podem ser resumidos conforme a tabela abaixo.

Doença	Fator		Total
	A	B	
D	y_1	$n_1 - y_1$	n_1
\bar{D}	y_2	$n_2 - y_2$	n_2

Discutimos nas seções seguintes a abordagem clássica para analisar a tabela acima.

3.2.2 Modelo probabilístico não condicional

Denotaremos por Y_1 e Y_2 o número de indivíduos com presença de A nos estratos D e \bar{D} , respectivamente. Será também assumido que essas variáveis são binomiais independentes de parâmetros (n_1, π_1) e (n_2, π_2) , respectivamente. Logo, a função de probabilidades conjunta de (Y_1, Y_2) fica dada por

$$f(\mathbf{y}; \boldsymbol{\pi}) = \binom{n_1}{y_1} \binom{n_2}{y_2} \pi_1^{y_1} \pi_2^{y_2} (1 - \pi_1)^{n_1 - y_1} (1 - \pi_2)^{n_2 - y_2}, \quad (3.3)$$

em que $\mathbf{y} = (y_1, y_2)^T$ e $\boldsymbol{\pi} = (\pi_1, \pi_2)^T$. Seguindo a notação da seção anterior, temos que $\pi_1 = P_1/(P_1 + P_3)$, $1 - \pi_1 = P_3/(P_1 + P_3)$, $\pi_2 = P_2/(P_2 + P_4)$ e $1 - \pi_2 = P_4/(P_2 + P_4)$. Assim, mostramos que

$$\psi = \frac{P_1 P_4}{P_3 P_2} = \frac{\pi_1 (1 - \pi_2)}{\pi_2 (1 - \pi_1)},$$

e conseqüentemente que $\pi_1 = \pi_2 \psi / \{\pi_2 \psi + 1 - \pi_2\}$. A expressão (3.3) pode então ser expressa apenas em função de (ψ, π_2) ,

$$f(\mathbf{y}; \boldsymbol{\pi}) \propto \exp \left\{ y_1 \log \psi + (y_1 + y_2) \log \left(\frac{\pi_2}{1 - \pi_2} \right) \right\} \frac{(1 - \pi_2)^n}{\{\psi \pi_2 + 1 - \pi_2\}^{n_1}}, \quad (3.4)$$

em que $n = n_1 + n_2$. As estimativas de máxima verossimilhança de π_1 e π_2 são dadas por $\tilde{\pi}_1 = y_1/n_1$ e $\tilde{\pi}_2 = y_2/n_2$, respectivamente. Logo, a estimativa de máxima verossimilhança não condicional de ψ fica $\tilde{\psi} = y_1(n_2 - y_2)/y_2(n_1 - y_1)$. Note que $E(\tilde{\psi}) = \infty$, o que impossibilita qualquer tipo de inferência para pequenas amostras. Por outro lado, para n_1 e n_2 grandes, $\tilde{\psi}$ segue uma distribuição normal de média ψ e variância assintótica

$$\text{Var}_A(\tilde{\psi}) = \psi^2 \left\{ \frac{1}{n_1 \pi_1 (1 - \pi_1)} + \frac{1}{n_2 \pi_2 (1 - \pi_2)} \right\}.$$

Formalmente, podemos dizer que sob condições gerais de regularidade e assumindo que $\frac{n_1}{n} \rightarrow a > 0$, quando $n \rightarrow \infty$, vale o resultado assintótico

$$\sqrt{n}(\tilde{\psi} - \psi) \rightarrow_d N(0, V_I(\psi)),$$

em que $V_I(\psi) = \psi^2\{1/a\pi_1(1 - \pi_1) + 1/(1 - a)\pi_2(1 - \pi_2)\}$. A variância assintótica $V_I(\psi)$ é consistentemente estimada por $n\text{Var}_A(\tilde{\psi})$.

Alguns autores preferem trabalhar com $\log\psi$ em vez de ψ . Assim, podemos mostrar, sob condições gerais de regularidade, que a estimativa não condicional $\log\tilde{\psi}$ segue para grandes amostras uma distribuição normal de média $\log\psi$ e variância assintótica $\text{Var}_A(\log\tilde{\psi}) = \{1/n_1\pi_1(1 - \pi_1) + 1/n_2\pi_2(1 - \pi_2)\}$. De forma equivalente, quando $n \rightarrow \infty$, temos que

$$\sqrt{n}(\log\tilde{\psi} - \log\psi) \rightarrow_d N(0, \psi^{-2}V_I(\psi)).$$

Esse resultado será útil na construção de intervalos de confiança para ψ .

3.2.3 Modelo probabilístico condicional

Devido aos problemas inferenciais com o modelo não condicional para pequenas amostras, a utilização de um modelo condicional, cuja construção será discutida a seguir, tem sido a solução encontrada sob o ponto de vista clássico para fazer inferências a respeito de ψ .

Assim, aplicando o teorema da fatorização para a função de probabilidades (3.4), mostramos que o conjunto de estatísticas $(Y_1, Y_1 + Y_2)$ é suficiente minimal para o vetor de parâmetros $[\log\psi, \log\{\pi_2/(1 - \pi_2)\}]$. Logo, a distribuição de (Y_1, Y_2) condicionada a $Y_1 + Y_2 = m$, deverá resultar numa função de probabilidades que depende apenas do parâmetro de interesse ψ . Essa distribuição resultante (ver Cornfield, 1956) tem sido largamente utilizada em pequenas amostras. Alguns autores questionam, entretanto, o procedimento adotado, uma vez que a estatística $Y_1 + Y_2$ não é ancilar para ψ ; isto é, contém informações a respeito do parâmetro ψ (ver discussão, por exemplo, em Lehnman, 1994, pgs. 546-547).

O condicionamento de (Y_1, Y_2) em $Y_1 + Y_2 = m$ produz o modelo caracterizado pela família de distribuições hipergeométricas não centrais, definida

por

$$f(y_1|m; \psi) = \frac{\binom{n_1}{y_1} \binom{n_2}{m-y_1} \psi^{y_1}}{\sum_t \binom{n_1}{t} \binom{n_2}{m-t} \psi^t}, \quad (3.5)$$

em que $0 < \psi < \infty$ e t varia de $\max(0, m - n_2)$ a $\min(n_1, m)$. Em particular, quando $\psi = 1$, a expressão (3.5) fica reduzida à conhecida distribuição hipergeométrica central, dada por

$$f(y_1|m; \psi = 1) = \frac{\binom{n_1}{y_1} \binom{n_2}{m-y_1}}{\binom{n_1+n_2}{m}},$$

cuja média e variância são, respectivamente,

$$E(1) = E(Y_1|m; \psi = 1) = \frac{mn_1}{n}$$

e

$$V(1) = \text{Var}(Y_1|m; \psi = 1) = \frac{n_1 n_2 (n - m) m}{n^2 (n - 1)}.$$

Para o modelo condicional (3.5) o logaritmo da função de verossimilhança fica dado por

$$L(\psi) \propto y_1 \log \psi - \log \left\{ \sum_t \binom{n_1}{t} \binom{n_2}{m-t} \psi^t \right\}.$$

Denotaremos por $\hat{\psi}$ a estimativa de máxima verossimilhança condicional. Essa estimativa pode ser expressa como a solução positiva da equação $y_1 = E(Y_1|m; \hat{\psi})$. Temos que o momento de ordem r da distribuição condicional, $E(Y_1^r|m; \psi)$ é dado por $E(Y_1^r|m; \psi) = P_r(\psi)/P_0(\psi)$, em que

$$P_r(\psi) = \sum_t t^r \binom{n_1}{t} \binom{n_2}{m-t} \psi^t, \quad r = 1, 2, \dots$$

e $P_0(\psi) = \sum_t \binom{n_1}{t} \binom{n_2}{m-t} \psi^t$. Assim, a equação de máxima verossimilhança para obtermos $\hat{\psi}$ fica reescrita na forma

$$y_1 - \frac{P_1(\hat{\psi})}{P_0(\hat{\psi})} = 0. \quad (3.6)$$

Com o aumento de n_1, n_2, m e $n - m$, fica impraticável obtermos $\hat{\psi}$ através de (3.6), uma vez que essa equação contém polinômios em $\hat{\psi}$ de grau bastante elevado. Uma saída, nesses casos, é resolvermos (3.6) através de métodos numéricos que não requerem a extração das raízes do polinômio $P_1(\psi)P_0^{-1}(\psi)$ (ver McCullagh e Nelder, 1989, p. 256 ; Silva, 1992).

Para ilustrarmos a obtenção de $\hat{\psi}$, consideremos a tabela abaixo.

	A	B	Total
D	1	3	4
\bar{D}	1	2	3

Temos, nesse caso, que $n_1 = 4, n_2 = 3$ e $m = 2$. A distribuição condicional fica então dada por

$$f(y_1|m; \psi) = \binom{4}{y_1} \binom{3}{2-y_1} \psi^{y_1} / \sum_t \binom{4}{t} \binom{3}{2-t} \psi^t,$$

em que o somatório varia no intervalo $0 \leq t \leq 2$. Isso resulta nas probabilidades condicionais

$$\begin{aligned} f(0|m; \psi) &= 3/\{3 + 12\psi + 6\psi^2\} \\ f(1|m; \psi) &= 12\psi/\{3 + 12\psi + 6\psi^2\} \text{ e} \\ f(2|m; \psi) &= 6\psi^2/\{3 + 12\psi + 6\psi^2\}. \end{aligned}$$

A equação $E(Y_1|m; \hat{\psi}) = y_1$ fica então dada por

$$12\hat{\psi} + 12\hat{\psi}^2 = 3 + 12\hat{\psi} + 6\hat{\psi}^2,$$

que é equivalente a $6\hat{\psi}^2 = 3$ ou $\hat{\psi} = 0,707$.

Similarmente ao estimador não condicional, podemos mostrar para grandes amostras que $\hat{\psi}$ segue uma distribuição normal de média ψ e variância assintótica $\text{Var}(\hat{\psi}) = V_A^{-1}(\psi)$, em que

$$V_A(\psi) = \left[\frac{1}{E_A(\psi)} + \frac{1}{n_1 - E_A(\psi)} + \frac{1}{m - E_A(\psi)} + \frac{1}{n_2 - m + E_A(\psi)} \right]^{-1},$$

e $E_A(\psi)$ sai da equação

$$\frac{E_A(\psi)\{n_2 - m + E_A(\psi)\}}{\{n_1 - E_A(\psi)\}\{m - E_A(\psi)\}} = \psi, \quad (3.7)$$

que para ψ fixo resulta numa equação quadrática em $E_A(\psi)$. Mostramos, para $\psi \neq 1$, que a única raiz de (3.7) que satisfaz $\max(0, m - n_2) \leq E_A(\psi) \leq \min(n_1, m)$ é dada por

$$E_A(\psi) = ||r| - s|,$$

em que $r = \frac{1}{2}[n/(\psi - 1) + m + n_1]$ e $s = [r^2 - mn_1\psi/(\psi - 1)]^{1/2}$.

Formalmente, podemos dizer que sob condições gerais de regularidade e assumindo ainda que n_1, n_2, m e $n - m$ são grandes, vale o seguinte resultado:

$$\sqrt{n}(\hat{\psi} - \psi) \rightarrow_d N(0, V_C(\psi)),$$

em que $V_C(\psi) = \lim_{n \rightarrow \infty} n\text{Var}(\hat{\psi})$. Logo, a variância assintótica $V_C(\psi)$ é consistentemente estimada por $n\text{Var}(\hat{\psi})$.

Quando $\psi = 1$, a expressão (3.7) não resulta numa forma quadrática em $E_A(\psi)$. Verificamos facilmente, nesse caso, que

$$E_A(1) = \frac{mn_1}{n}$$

e

$$V_A(1) = \frac{n_1 n_2 m (n - m)}{n^3}.$$

Podemos notar que a média e a variância assintótica de $\hat{\psi}$, quando $\psi = 1$, coincidem praticamente com a média e a variância da distribuição condicional dada em (3.5).

3.2.4 Teste de hipóteses

Testes exatos

Uma vez conhecida a distribuição condicional que depende apenas do parâmetro de interesse ψ , podemos desenvolver testes exatos para pequenas amostras.

Um caso de interesse seria testarmos $H_0 : \psi = \psi_0$ contra $H_1 : \psi < \psi_0$, em que ψ_0 é um valor conhecido. O nível descritivo do teste, isto é, a probabilidade sob H_0 de obtenção de valores tão ou mais desfavoráveis a H_0 (no sentido de H_1) é definido por

$$P_I = \sum_{t \leq y_1} f(t|m; \psi_0),$$

em que o somatório vai de $\max(0, m - n_2)$ até y_1 . Analogamente, para testarmos $H_0 : \psi = \psi_0$ contra $H_1 : \psi > \psi_0$, teremos $P_S = \sum_{t \geq y_1} f(t|m; \psi_0)$. Nesse caso, o somatório vai de y_1 até $\min(n_1, m)$. Para o teste bilateral, $H_0 : \psi = \psi_0$ contra $H_1 : \psi \neq \psi_0$, o nível descritivo é definido por $P = 2\min\{P_I, P_S\}$.

Em particular, quando fazemos $\psi_0 = 1$, estamos objetivamente testando a não existência de associação entre o fator e a doença, sendo o teste resultante conhecido como teste exato de Fisher (ver, por exemplo, Everitt, 1977). Nesse caso, o nível descritivo é obtido computando as probabilidades da distribuição hipergeométrica central.

Podemos também utilizar o modelo condicional (3.5) para a estimação intervalar de ψ . Os respectivos limites de confiança serão baseados em P_I e P_S e denotados por $\hat{\psi}_I$ e $\hat{\psi}_S$, respectivamente. Como ilustração, vamos supor que estamos interessados em construir um intervalo de confiança de coeficiente $(1 - \alpha)$ para ψ . Os limites $\hat{\psi}_I$ e $\hat{\psi}_S$ ficam então, invertendo a região crítica do teste $H_0 : \psi = \psi_0$ contra $H_1 : \psi \neq \psi_0$, determinados pelas equações

$$\frac{\alpha}{2} = \sum_{t \leq y_1} f(t|m; \hat{\psi}_S) \quad \text{e} \quad \frac{\alpha}{2} = \sum_{t \geq y_1} f(t|m; \hat{\psi}_I),$$

que são polinômios de grau elevado em $\hat{\psi}_S$ e $\hat{\psi}_I$ à medida que os tamanhos amostrais crescem, o que praticamente inviabiliza a solução dessas equações.

Nesses casos, a saída é procurarmos intervalos assintóticos.

Voltando à tabela da seção anterior, vamos supor que queremos testar $H_0 : \psi = 1$ contra $H_1 : \psi \neq 1$. Temos então os níveis descritivos $P_I = f(0|m; \psi = 1) + f(1|m; \psi = 1) = 15/21$ e $P_S = f(1|m; \psi = 1) + f(2|m; \psi = 1) = 18/21$ o que leva a $P = 1,0$. Por outro lado, os limites $\hat{\psi}_I$ e $\hat{\psi}_S$ ficam dados por

$$\frac{\alpha}{2} = \sum_{t=0}^1 f(t|m; \hat{\psi}_S) \quad \text{e} \quad \frac{\alpha}{2} = \sum_{t=1}^2 f(t|m; \hat{\psi}_I)$$

que é equivalente, supondo $\alpha = 0,20$, a

$$0,10 = f(0|m; \hat{\psi}_S) + f(1|m; \hat{\psi}_S) \quad \text{e} \quad 0,10 = f(1|m; \hat{\psi}_I) + f(2|m; \hat{\psi}_I),$$

que levam às equações

$$0,10 = \frac{4\hat{\psi}_I + 2\hat{\psi}_I^2}{1 + 4\hat{\psi}_I + 2\hat{\psi}_I^2} \quad (\hat{\psi}_I = 0,0274)$$

e

$$0,10 = \frac{1 + 4\hat{\psi}_S}{1 + 4\hat{\psi}_S + 2\hat{\psi}_S^2} \quad (\hat{\psi}_S = 18,25).$$

Testes assintóticos

Para grandes amostras, n_1, n_2, m e $n - m$ grandes, a distribuição condicional (3.5) se aproxima de uma distribuição normal de média $E_A(\psi)$ e variância

$V_A(\psi)$ (ver Hannan e Harkness, 1963). Esse fato tem sido utilizado para o desenvolvimento de testes assintóticos para testarmos $H_0 : \psi = \psi_0$ contra $H_1 : \psi \neq \psi_0$ ($H_1 : \psi > \psi_0$ ou $H_1 : \psi < \psi_0$). No caso de $H_1 : \psi \neq \psi_0$, utilizamos a estatística qui-quadrado dada abaixo

$$X^2 = \frac{\{y_1 - E_A(\psi_0)\}^2}{V_A(\psi_0)}, \quad (3.8)$$

que sob H_0 segue assintoticamente uma distribuição qui-quadrado com 1 grau de liberdade. Para $H_1 : \psi < \psi_0$ e $H_1 : \psi > \psi_0$, o nível descritivo é dado por

$$P_I = Pr \left\{ Z \leq \frac{y_1 - E_A(\psi_0)}{\sqrt{V_A(\psi_0)}} \right\}$$

e

$$P_S = Pr \left\{ Z \geq \frac{y_1 - E_A(\psi_0)}{\sqrt{V_A(\psi_0)}} \right\},$$

respectivamente, em que Z segue uma distribuição $N(0, 1)$. Em particular, quando $\psi_0 = 1$, a estatística qui-quadrado (3.8) fica reduzida à forma conhecida

$$X^2 = \frac{\left\{y_1 - \frac{mn_1}{n}\right\}^2}{n_1 n_2 m(n-m)/n^3}. \quad (3.9)$$

Um intervalo assintótico de confiança para ψ pode ser obtido utilizando a distribuição assintótica de $\log \tilde{\psi}$. Os limites desse intervalo são dados por

$$\log \tilde{\psi}_I = \log \tilde{\psi} - z_{(1-\alpha/2)} \sqrt{\text{Var}_A(\log \tilde{\psi})}$$

e

$$\log \tilde{\psi}_S = \log \tilde{\psi} + z_{(1-\alpha/2)} \sqrt{\text{Var}_A(\log \tilde{\psi})},$$

em que $z_{(1-\alpha/2)}$ é o percentil $(1 - \alpha/2)$ da distribuição normal padrão. Esses limites podem ser expressos em uma outra forma, levando em conta a

estatística qui-quadrado para testarmos $H_0 : \psi = 1$ contra $H_1 : \psi \neq 1$. Essa estatística é dada por

$$X^2 = \frac{(\log \tilde{\psi})^2}{\text{Var}_A(\log \tilde{\psi})}, \quad (3.10)$$

que segue, para grandes amostras, uma distribuição qui-quadrado com 1 grau de liberdade. Assim, os limites ficam reexpressos nas formas

$$\tilde{\psi}_I = \tilde{\psi}^{(1-z_{(1-\alpha/2)}/X)}$$

e

$$\tilde{\psi}_S = \tilde{\psi}^{(1+z_{(1-\alpha/2)}/X)}.$$

Alguns autores (ver Breslow e Day, 1980, p. 135) têm constatado que para $n_1 = n_2$ a probabilidade de cobertura do intervalo $(\tilde{\psi}_I, \tilde{\psi}_S)$ é em geral menor do que o valor nominal utilizado. Por outro lado, quando n_1 e n_2 são muito diferentes, essa probabilidade de cobertura é superestimada. Uma sugestão, nesses casos, é utilizarmos o valor de X obtido do teste condicional (3.9) em vez do valor obtido do teste não condicional (3.10).

3.3 Métodos clássicos: k tabelas 2×2

Muitas vezes temos interesse em controlar a associação entre dois fatores binários através de um terceiro fator, comumente chamado de fator de confundimento. O principal objetivo com esse tipo de estratificação é eliminarmos ou pelo menos reduzirmos a influência desses fatores na associação de interesse. Uma maneira mais eficiente de controlarmos fatores de confundimento é através da regressão logística, que será discutida na Seção 3.6. Nesta seção, assumiremos apenas um fator de confundimento com k níveis, que são amostrados n_i indivíduos no i -ésimo estrato (n_{1i} casos e n_{2i} controles) e que os mesmos são classificados conforme a tabela 2×2 abaixo.

Doença	Fator		Total
	A	B	
D	y_{1i}	$n_{1i} - y_{1i}$	n_{1i}
\bar{D}	y_{2i}	$n_{2i} - y_{2i}$	n_{2i}

Seguindo a mesma notação das seções anteriores temos que as estimativas não condicional e condicional de ψ_i são, respectivamente, tais que

$$\tilde{\psi}_i = \frac{y_{1i}(n_{2i} - y_{2i})}{y_{2i}(n_{1i} - y_{1i})} \text{ e } y_{1i} - \frac{P_{1i}(\hat{\psi}_i)}{P_{0i}(\hat{\psi}_i)} = 0.$$

As propriedades assintóticas de $\tilde{\psi}_i$ e $\hat{\psi}_i$ são as mesmas de $\tilde{\psi}$ e $\hat{\psi}$ da Seção 3.2, bem como as formas dos testes de hipóteses e da estimação intervalar.

3.3.1 Estimação da razão de chances comum

Um teste de interesse quando há k tabelas de contingência 2×2 é verificarmos a não existência de interação entre os estratos, isto é, verificarmos se a associação entre o fator e a doença não muda de um estrato para o outro. Isso é equivalente a verificarmos se as razões de chances são homogêneas, ou seja, testarmos as hipóteses

$$H_0 : \psi_1 = \dots = \psi_k$$

$$H_1 : \text{ pelo menos dois valores diferentes.}$$

Há várias propostas de estimativas para a razão de chances comum. As estimativas de máxima verossimilhança não condicional e condicional serão denotadas por $\tilde{\psi}$ e $\hat{\psi}$, respectivamente. A primeira estimativa pode ser obtida facilmente através do ajuste de uma regressão logística, enquanto que a segunda é mais complexa do ponto de vista computacional e será omitida.

Duas estimativas não iterativas foram propostas por Mantel e Haenszel (1959) e Wolf (1955), as quais serão denotadas por $\hat{\psi}_{MH}$ e $\hat{\psi}_W$, respectiva-

mente. A estimativa de Mantel-Hanszel é definida por

$$\hat{\psi}_{MH} = \frac{\sum_{i=1}^k y_{1i}(n_{2i} - y_{2i})/n_i}{\sum_{i=1}^k y_{2i}(n_{1i} - y_{1i})/n_i},$$

e pode também ser expressa como uma média ponderada de estimativas não condicionais

$$\hat{\psi}_{MH} = \frac{\sum_{i=1}^k v_i \tilde{\psi}_i}{\sum_{i=1}^k v_i},$$

em que $v_i = y_{2i}(n_{1i} - y_{1i})/n_i$. O estimador de Mantel-Hanszel é consistente e assintoticamente normal com variância assintótica dada por

$$\text{Var}_A(\hat{\psi}_{MH}) = \psi^2 \sum_{i=1}^k a_i \omega_i^{-1} / \left(\sum_{i=1}^k a_i \right)^2,$$

em que $\omega_i = \{n_{1i}\pi_{1i}(1 - \pi_{1i})\}^{-1} + \{n_{2i}\pi_{2i}(1 - \pi_{2i})\}^{-1}$ e $a_i = n_{1i}n_{2i}(1 - \pi_{1i})\pi_{2i}/n_i$. A estimativa de Wolf é dada por

$$\hat{\psi}_W = \exp \left(\frac{\sum_{i=1}^k u_i \log \tilde{\psi}_i}{\sum_{i=1}^k u_i} \right),$$

em que $u_i = \{1/y_{1i} + 1/(n_{1i} - y_{1i}) + 1/y_{2i} + 1/(n_{2i} - y_{2i})\}^{-1}$. Esse estimador é também consistente e assintoticamente normal com variância dada por

$$\text{Var}_A(\hat{\psi}_W) = \psi^2 \omega^{-1},$$

em que $\omega = \omega_1 + \dots + \omega_k$. Para as estimativas $\tilde{\psi}$, $\hat{\psi}_{MH}$ e $\hat{\psi}_W$ de ψ comum assumimos o modelo não condicional para os dados.

3.3.2 Testes de homogeneidade

Suponha que estamos interessados em testar as hipóteses H_0 e H_1 definidas na seção anterior. A estatística da razão de verossimilhanças que assume o produto de $2k$ binomiais independentes é a mais utilizada nesse caso. Do

ponto de vista de análise preliminar dos dados, duas estatísticas têm sido sugeridas. A primeira delas, proposta por Hosmer e Lemeshow (1989, p. 74), é definida abaixo

$$X_{HL}^2 = \sum_{i=1}^k \tilde{\omega}_i (\log \tilde{\psi}_i - \log \hat{\psi}_W)^2,$$

que segue, sob H_0 e assintoticamente (para n_{1i} e n_{2i} grandes, $\forall i$), uma distribuição qui-quadrado com $k - 1$ graus de liberdade. A outra estatística, definida em Breslow e Day (1980, p. 42), é baseada no modelo condicional, sendo dada por

$$X_{BD}^2 = \sum_{i=1}^k \frac{\{y_{1i} - E_{A_i}(\hat{\psi}_{MH})\}^2}{V_{A_i}(\hat{\psi}_{MH})},$$

que também segue, sob H_0 e para grandes amostras, uma distribuição qui-quadrado com $k - 1$ graus de liberdade. A novidade, nesse caso, é a utilização da estatística não iterativa de Mantel-Hanszel no lugar da estimativa condicional $\hat{\psi}$.

Quando a hipótese nula não é rejeitada, um teste imediato é verificarmos a não existência de associação entre o fator e a doença, mantendo apenas o efeito da estratificação. Esse teste, conhecido como teste de Mantel-Hanszel (1959), utiliza a seguinte estatística:

$$X_{MH}^2 = \frac{\{(\sum_{i=1}^k y_{1i} - \sum_{i=1}^k E_{A_i}(1))\}^2}{\sum_{i=1}^k V_{A_i}(1)},$$

que, sob $H_0 : \psi = 1$, segue para grandes amostras (n_i grande $\forall i$ ou para k grande) uma distribuição qui-quadrado com 1 grau de liberdade. Similarmente ao caso de uma única tabela 2×2 , um intervalo assintótico de confiança para ψ com coeficiente de confiança $(1 - \alpha)$ fica dado por

$$(\hat{\psi}_I, \hat{\psi}_S) = \hat{\psi}_{MH}^{(1 \pm z_{(1-\alpha/2)}/X_{MH})},$$

em que $X_{MH} = \sqrt{X_{MH}^2}$.

3.4 Métodos clássicos: tabelas $2 \times k$

A dicotomização de um fator com mais de 2 níveis, a fim de deixarmos mais simples o estudo da associação entre esse fator e uma determinada doença, pode omitir informações relevantes acerca da associação de cada um dos níveis agrupados e a doença em estudo. Assim, sempre que possível, devemos manter para as análises o maior número possível de níveis do fator. Uma tabela resultante, nesse caso, é dada abaixo.

Doença	Fator				Total
	Nível 1	Nível 2	\cdots	Nível k	
D	y_{11}	y_{12}	\cdots	$n_1 - \sum_{i=1}^{k-1} y_{1i}$	n_1
\bar{D}	y_{21}	y_{22}	\cdots	$n_2 - \sum_{i=1}^{k-1} y_{2i}$	n_2

Analogamente ao caso de uma única tabela 2×2 , assumimos que são amostrados n_1 elementos do estrato D e n_2 elementos do estrato \bar{D} e que $(Y_{i1}, \dots, Y_{ik})^T$ segue uma distribuição multinomial de parâmetros $(\pi_{i1}, \dots, \pi_{ik})^T$, com $\pi_{ik} = 1 - \sum_{j=1}^{k-1} \pi_{ij}$, $i = 1, 2$. Comumente, para analisarmos as associações entre os níveis do fator e a doença, definimos um nível do fator como referência, que formará com os demais as razões de chances. Escolhendo o nível 1 como referência, as razões de chances ficam dadas por

$$\psi_1 = 1 \text{ e } \psi_j = \frac{\pi_{1j}\pi_{21}}{\pi_{2j}\pi_{11}}, \quad j = 2, \dots, k,$$

em que ψ_j é a razão de chances entre o nível j e o nível 1 do fator. As análises inferenciais através do uso do modelo multinomial são tratadas em textos correntes de análise de dados categorizados (ver, por exemplo, Agresti, 1990). Aqui, nos concentraremos no estudo do modelo condicional, que é obtido após o condicionamento de $(Y_{i1}, \dots, Y_{ik})^T$, $i = 1, 2$, nas estatísticas suficientes minimais $Y_{1j} + Y_{2j} = m_j$, $j = 1, \dots, k$. O modelo resultante é caracterizado pela distribuição hipergeométrica multivariada não central

que depende apenas dos parâmetros de interesse ψ_1, \dots, ψ_k (ver McCullagh e Nelder, 1989, p. 261). Em particular, a hipótese de ausência de associação completa entre os níveis do fator e a doença é definida por $H_0 : \psi_j = 1, \forall j$, que será avaliada através da distribuição hipergeométrica central k -dimensional, cuja função de probabilidades é o produto de k distribuições hipergeométricas centrais

$$f(\mathbf{y}_1 | \mathbf{m}; \boldsymbol{\psi} = \mathbf{1}) = \prod_{j=1}^k \frac{\binom{n_{1j}}{y_{1j}} \binom{n_{2j}}{m_j - y_{1j}}}{\binom{n_{1j} + n_{2j}}{m_j}}, \quad (3.11)$$

em que $\mathbf{y}_1 = (y_{11}, \dots, y_{1k})^T$, $\mathbf{m} = (m_1, \dots, m_k)^T$ e $\boldsymbol{\psi} = (\psi_1, \dots, \psi_k)^T$. A média, variância e covariância correspondentes à distribuição (3.11) são, respectivamente, dadas por

$$E_j(1) = E(Y_{1j} | m_j; \psi = 1) = \frac{m_j n_1}{n},$$

$$V_j(1) = \text{Var}(Y_{1j} | m_j; \psi = 1) = \frac{n_1 n_2 (n - m_j) m_j}{n^2 (n - 1)}$$

e

$$C_{j\ell} = \text{Cov}(Y_{1j}, Y_{1\ell} | m_j, m_\ell; \psi = 1) = -\frac{m_j m_\ell n_1 n_2}{n^2 (n - 1)}, \quad j \neq \ell,$$

em que $n = n_1 + n_2$. Um teste estatístico para H_0 , que tem sido largamente utilizado para testarmos a homogeneidade de k proporções (Armitage, 1971), é dado por

$$\begin{aligned} X_A^2 &= \frac{(n-1)}{n} \sum_{j=1}^k \{y_{1j} - E_j(1)\}^2 \left\{ \frac{1}{E_j(1)} + \frac{1}{m_j - E_j(1)} \right\} \\ &= (n-1) \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} \sum_{j=1}^k \frac{\{y_{1j} - E_j(1)\}^2}{m_j}, \end{aligned} \quad (3.12)$$

que segue, sob H_0 e para valores grandes de n_1, n_2 e $m_j, \forall j$, uma distribuição qui-quadrado com $k - 1$ graus de liberdade. Entretanto, quando os níveis do fator são quantitativos ou qualitativos ordinais, pode ser mais informativo

o uso de um teste para a tendência do risco da doença com o aumento dos níveis do fator. Como ilustração, vamos supor que há k doses x_j , $j = 1, \dots, k$, associadas aos k níveis do fator. Um teste apropriado é considerarmos a regressão dos desvios $\{y_{1j} - E_j(1)\}$ sobre x_j (Armitage, 1955; Mantel, 1963). A estatística correspondente fica dada por

$$X_{HOM}^2 = \frac{n^2(n-1)[\sum_{j=1}^k x_j \{y_{1j} - E_j(1)\}]^2}{n_1 n_2 \{n \sum_{j=1}^k x_j^2 m_j - (\sum_{j=1}^k x_j m_j)^2\}}, \quad (3.13)$$

que segue, para grandes amostras e sob H_0 , uma distribuição qui-quadrado com $k - 1$ graus de liberdade.

Uma outra maneira de analisarmos a associação entre o fator e a doença é através da amostragem nos k níveis do fator de interesse. Nesse caso, a distribuição resultante é um produto de k binomiais independentes e a hipótese de ausência de associação entre o fator e a doença pode ser avaliada através do ajuste de uma regressão logística, que será discutida na Seção 3.6. Por outro lado, se também forem fixados os totais n_1 e n_2 , a distribuição condicional resultante é uma hipergeométrica não central k -dimensional que sob H_0 fica reduzida a (3.11). Logo, as estatísticas dadas em (3.12) e (3.13) podem ser aplicadas, pelo menos numa análise preliminar dos dados, para avaliarmos a ausência de associação total entre o fator e a doença.

Generalizações de (3.12) e (3.13) para o caso de h estratos são dadas em Breslow e Day (1980, pgs. 148-149).

3.5 Aplicações

3.5.1 Associação entre fungicida e desenvolvimento de tumor

Como ilustração, analisaremos o conjunto de dados apresentado em Innes et al. (1969), referente a um estudo para avaliar o possível efeito cancerígeno

do fungicida Avadex. No estudo, 403 camundongos são observados. Desses, 65 receberam o fungicida e foram acompanhados durante 85 semanas, verificando o desenvolvimento ou não de tumor cancerígeno. Os demais animais não receberam o fungicida (grupo controle) e também foram acompanhados pelo mesmo período, verificando a ocorrência ou não de tumor. Dois fatores potenciais de confundimento, sexo e raça, foram considerados nas análises. Os dados do experimento são resumidos na Tabela 3.1.

Tabela 3.1

Classificação dos camundongos quanto a raça (R1 ou R2), sexo, grupo e ocorrência ou não de tumor cancerígeno.

Estrato	Grupo	Com tumor	Sem tumor	Total
R1-Macho	Tratado	4	12	16
	Controle	5	74	79
	Total	9	86	95
R2-Macho	Tratado	2	14	16
	Controle	3	84	87
	Total	5	98	103
R1-Fêmea	Tratado	4	14	18
	Controle	10	80	90
	Total	14	94	108
R2-Fêmea	Tratado	1	14	15
	Controle	3	79	82
	Total	4	93	97

Em virtude dos valores relativamente altos das marginais das quatro tabelas 2×2 formadas pela combinação dos fatores sexo e raça, procedemos inicialmente uma análise através do modelo não condicional. Temos então, na primeira coluna da Tabela 3.2, as estimativas pontuais das razões de chances de tumor maligno entre o grupo tratado e o grupo controle. Na segunda

coluna apresentamos os intervalos assintóticos de 95% para ψ . Notamos que, embora todas as estimativas sinalizem para uma associação positiva, apenas o primeiro intervalo de confiança não cobre o valor $\psi = 1$, evidenciando associação apenas no primeiro estrato, ao nível de 5%.

O teste de homogeneidade das razões de chances forneceu $X_{BD}^2 = 0,867$ (3 g.l. e $P = 0,833$), indicando fortemente pela não rejeição da ausência de interação entre os estratos. Já o teste de Mantel-Hanszel forneceu $X_{MH}^2 = 8,289$ (1 g.l. e $P = 0,004$), indicando pela rejeição da hipótese de razão de chances comum igual a um, isto é, de que há fortes indícios de associação entre os grupos controle e tratado. As estimativas de ψ comum deram $\hat{\psi}_{MH} = 3,079$ e $\hat{\psi}_W = 3,109$, com intervalo assintótico de confiança de 95% dado por $[1,43; 6,62]$.

Tabela 3.2

Estimativas das razões de chances de tumor cancerígeno nos estratos de camundongos.

Estrato	Estimativa $\tilde{\psi}$	Intervalo assintótico
R1-Macho	4,93	[1,28 ; 18,97]
R2-Macho	4,00	[0,69 ; 23,09]
R1-Fêmea	2,29	[0,64 ; 8,14]
R2-Fêmea	1,88	[0,19 ; 48,87]

3.5.2 Efeito de extrato vegetal

Consideremos agora parte dos dados de um experimento (ver Paula, Sevanes e Ogando, 1988) conduzido para avaliarmos o efeito de diversos extratos vegetais na mortalidade de embriões de *Biomphalaria Glabrata* (hospedeiro da equistossomose). Para o extrato vegetal aquoso frio de folhas de *P. Hyrsiflora* consideramos um total de $k = 7$ grupos sendo que os n_i embriões do i -ésimo grupo foram submetidos a uma dose x_i (ppm) do extrato vegetal,

observando após o 20º dia o número de embriões mortos. Os dados são resumidos na Tabela 3.3. Para aplicarmos o teste de tendência dado em (3.13), devemos considerar que $n = 50 + \dots + 50 = 350$, $n_1 = y_1 + \dots + y_7 = 178$, $n_2 = n - n_1 = 172$ e $m_i = 50$, $\forall i$. Assim, obtemos $E_i(1) = 25,43$ para $i = 1, \dots, 7$. A estatística forneceu o valor $X_{HOM}^2 = 131,82$, que é altamente significativo, indicando uma forte tendência crescente para a proporção de mortes com o aumento da dose.

Tabela 3.3

Distribuição dos embriões segundo os níveis de exposição do estrato vegetal aquoso.

x_i	0	15	20	25	30	35	40
m_i	50	50	50	50	50	50	50
y_i	4	5	14	29	38	41	47

3.6 Regressão logística linear

3.6.1 Introdução

A regressão logística tem se constituído num dos principais métodos de modelagem estatística de dados. Mesmo quando a resposta de interesse não é originalmente do tipo binário, alguns pesquisadores têm dicotomizado a resposta de modo que a probabilidade de sucesso possa ser ajustada através da regressão logística. Isso ocorre, por exemplo, em análise de sobrevivência discreta em que a resposta de interesse é o tempo de sobrevivência, no entanto, em algumas pesquisas, a função de risco tem sido ajustada por modelos logísticos. Tudo isso se deve, principalmente, pela facilidade de interpretação dos parâmetros de um modelo logístico e também pela possibilidade do uso desse tipo de metodologia em análise discriminante.

Embora a regressão logística seja conhecida desde os anos 50, foi através de Cox (1970) (ver também Cox e Snell, 1989) que a regressão logística ficou popular entre os usuários de Estatística. Nesta seção apresentamos alguns resultados relacionados com o modelo logístico linear que completam o que foi apresentado no Capítulo 1, onde vimos esse modelo como um caso particular de modelos lineares generalizados.

3.6.2 Regressão logística simples

Vamos considerar inicialmente o modelo logístico linear simples em que $\pi(x)$, a probabilidade de “sucesso” dado o valor x de uma variável explicativa qualquer é definida tal que

$$\log \left\{ \frac{\pi(x)}{1 - \pi(x)} \right\} = \alpha + \beta x, \quad (3.14)$$

em que α e β são parâmetros desconhecidos. Esse modelo poderia, por exemplo, ser aplicado para analisar a associação entre uma determinada doença e a ocorrência ou não de um fator particular. Seriam então amostrados, independentemente, n_1 indivíduos com presença do fator ($x=1$) e n_2 indivíduos com ausência do fator ($x=0$) e $\pi(x)$ seria a probabilidade de desenvolvimento da doença após um certo período fixo. Dessa forma, a chance de desenvolvimento da doença para um indivíduo com presença do fator fica dada por

$$\frac{\pi(1)}{1 - \pi(1)} = e^{\alpha + \beta},$$

enquanto que a chance de desenvolvimento da doença para um indivíduo com ausência do fator é simplesmente

$$\frac{\pi(0)}{1 - \pi(0)} = e^{\alpha}.$$

Logo, a razão de chances fica dada por

$$\psi = \frac{\pi(1)\{1 - \pi(0)\}}{\pi(0)\{1 - \pi(1)\}} = e^{\beta},$$

dependendo apenas do parâmetro β . Mesmo que a amostragem seja retrospectiva, isto é, são amostrados n_1 indivíduos doentes e n_2 indivíduos não doentes, o resultado acima continua valendo. Essa é uma das grandes vantagens da regressão logística, a possibilidade de interpretação direta dos coeficientes como medidas de associação. Esse tipo de interpretação pode ser estendido para qualquer problema prático.

Vamos supor agora que temos dois estratos representados por x_1 ($x_1 = 0$ estrato 1, $x_1 = 1$ estrato 2) e que são amostrados do estrato 1 n_{11} indivíduos com presença do fator e n_{21} indivíduos com ausência do fator e n_{12} e n_{22} , respectivamente, do estrato 2. A probabilidade de desenvolvimento da doença será denotada por $\pi(x_1, x_2)$, com x_2 ($x_2=1$ presença do fator, $x_2 = 0$ ausência do fator). Temos aqui quatro parâmetros a serem estimados, $\pi(0, 0)$, $\pi(0, 1)$, $\pi(1, 0)$ e $\pi(1, 1)$. Logo, qualquer reparametrização deverá ter no máximo quatro parâmetros (modelo saturado).

Vamos considerar então a seguinte reparametrização:

$$\log \left\{ \frac{\pi(x_1, x_2)}{1 - \pi(x_1, x_2)} \right\} = \alpha + \gamma x_1 + \beta x_2 + \delta x_1 x_2,$$

em que γ representa o efeito do estrato, β o efeito do fator e δ a interação entre estrato e fator. Para entendermos melhor essa reparametrização, vamos calcular as razões de chances em cada estrato

$$\psi_1 = \frac{\pi(0, 1)\{1 - \pi(0, 0)\}}{\pi(0, 0)\{1 - \pi(0, 1)\}} = e^\beta$$

e

$$\psi_2 = \frac{\pi(1, 1)\{1 - \pi(1, 0)\}}{\pi(1, 0)\{1 - \pi(1, 1)\}} = e^{\beta+\delta}.$$

Assim, a hipótese de homogeneidade das razões de chances ($H_0 : \psi_1 = \psi_2$) é equivalente à hipótese de não interação ($H_0 : \delta = 0$). Portanto, a ausência de interação entre fator e estrato significa que a associação entre o fator e a

doença não muda de um estrato para o outro. Contudo, pode haver efeito de estrato. Como ilustração nesse caso, vamos supor que não rejeitamos a hipótese $H_0 : \delta = 0$. Assim, o logaritmo da chance de desenvolvimento da doença fica dado por

$$\log \left\{ \frac{\pi(x_1, x_2)}{1 - \pi(x_1, x_2)} \right\} = \alpha + \gamma x_1 + \beta x_2,$$

ou seja, é o mesmo nos dois estratos a menos da quantidade γ . Isso quer dizer que mesmo não havendo interação entre os dois estratos (razão de chances constante), as probabilidades de desenvolvimento da doença podem estar em patamares diferentes. Num estrato essas probabilidades são maiores do que no outro estrato. Essas interpretações podem ser generalizadas para três ou mais tabelas.

Aplicação

Como ilustração, vamos considerar novamente o exemplo descrito na Seção 3.5.1, supondo que agora temos apenas os estratos macho e fêmea. Os dados são resumidos na Tabela 3.4 e no arquivo **camundongos.dat**.

Tabela 3.4

Classificação de camundongos segundo sexo, grupo e ocorrência de tumor.

Tumor	Macho		Fêmea	
	Tratado	Controle	Tratado	Controle
Sim	6	8	5	13
Não	26	158	28	159
Total	32	166	33	172

Denotamos por $\pi(x_1, x_2)$ a probabilidade de desenvolvimento de tumor dados x_1 ($x_1=1$ macho, $x_1=0$ fêmea) e x_2 ($x_2=1$ tratado, $x_2=0$ controle). Para testarmos a hipótese de ausência de interação ($H_0 : \delta = 0$) comparamos o desvio do modelo sem interação $D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) = 0,832$ com os percentis

da distribuição qui-quadrado com 1 grau de liberdade (temos que o desvio do modelo saturado é zero). O nível descritivo obtido é dado por $P = 0,362$, indicando pela não rejeição da hipótese de homogeneidade das razões de chances. Assim, ajustamos o modelo sem interação. As estimativas resultantes são apresentadas na Tabela 3.5.

Tabela 3.5

*Estimativas dos parâmetros do modelo
logístico ajustado aos dados sobre
ocorrência de tumor em camundongos.*

Efeito	Estimativa	E/E.Padrão
Constante	-2,602	-9,32
Estrato	-0,241	-0,64
Tratamento	1,125	2,81

Os níveis descritivos dos testes para $H_0 : \beta = 0$ e $H_0 : \gamma = 0$ são, respectivamente, dados por $P = 0,005$ e $P = 0,520$, indicando fortemente pela presença de associação entre a exposição ao fungicida e o desenvolvimento de tumor e que as probabilidades de desenvolvimento de tumor não são diferentes entre os dois estratos.

Temos que $\hat{\psi} = e^{\hat{\beta}}$, logo um intervalo assintótico de confiança para ψ com coeficiente $(1 - \alpha)$, terá os limites

$$(\hat{\psi}_I, \hat{\psi}_S) = \exp\{\hat{\beta} \pm z_{(1-\alpha/2)} \sqrt{\text{Var}(\hat{\beta})}\}.$$

Para o exemplo acima e assumindo um intervalo de 95%, esses limites ficam dados por $[1,403; 6,759]$.

O valor observado da variável explicativa no modelo logístico dado em (3.14) pode representar o valor de alguma variável quantitativa qualquer como, por exemplo, a dose ou a log-dose de uma determinada droga. Nesse caso, faz sentido calcularmos a chance de um indivíduo que recebeu a dose

x^* , ser curado, em relação a um outro indivíduo que recebeu a dose x . A razão de chances de cura, entre os dois níveis, fica dada por

$$\psi_{(x^*-x)} = \frac{\pi(x^*)\{1 - \pi(x)\}}{\pi(x)\{1 - \pi(x^*)\}} = \exp\{\beta(x^* - x)\}.$$

Portanto, $\log\psi_{(x^*-x)}$ é proporcional à diferença entre as duas doses. Se $\beta > 0$, temos que a chance de cura aumenta com o aumento da dose e se $\beta < 0$ ocorre o contrário. Essa interpretação pode ser estendida para qualquer variável explicativa quantitativa.

3.6.3 Regressão logística múltipla

Vamos considerar agora o modelo geral de regressão logística

$$\log \left\{ \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right\} = \beta_1 + \beta_2 x_2 + \cdots + \beta_p x_p,$$

em que $\mathbf{x} = (1, x_2, \dots, x_p)^T$ contém os valores observados de variáveis explicativas. Como vimos na Seção 1.6.1, o processo iterativo para obtermos $\hat{\beta}$ pode ser expresso como um processo iterativo de mínimos quadrados reponderados

$$\beta^{(m+1)} = (\mathbf{X}^T \mathbf{V}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{(m)} \mathbf{z}^{(m)},$$

em que $\mathbf{V} = \text{diag}\{\pi_1(1 - \pi_1), \dots, \pi_n(1 - \pi_n)\}$, $\mathbf{z} = (z_1, \dots, z_n)^T$ é a variável dependente modificada, $z_i = \eta_i + (y_i - \pi_i)/\pi_i(1 - \pi_i)$, $m = 0, 1, \dots$ e $i = 1, \dots, n$. Para dados agrupados (k grupos), substituímos n por k , $\mathbf{V} = \text{diag}\{n_1\pi_1(1 - \pi_1), \dots, n_k\pi_k(1 - \pi_k)\}$ e $z_i = \eta_i + (y_i - n_i\pi_i)/\{n_i\pi_i(1 - \pi_i)\}$. Assintoticamente, $n \rightarrow \infty$ no primeiro caso e para $\frac{n_i}{n} \rightarrow a_i > 0$ no segundo caso, $\hat{\beta} - \beta \sim N_p(\mathbf{0}, (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1})$.

Uma interpretação interessante pode ser dada para as razões de chances quando temos $(q-1)(q \leq p)$ das $(p-1)$ variáveis explicativas do tipo binário. Como ilustração, vamos supor $q = 4$ e que x_2 ($x_2 = 1$ presença, $x_2 = 0$

ausência) e x_3 ($x_3 = 1$ presença, $x_3 = 0$ ausência) representam dois fatores. Vamos supor ainda que $x_4 = x_2x_3$ representa a interação entre os dois fatores. O modelo fica então dado por

$$\log \left\{ \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right\} = \beta_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \sum_{j=5}^p x_j\beta_j.$$

Denotaremos por ψ_{ij} a razão de chances entre um indivíduo na condição $(x_2 = i, x_3 = j)$ em relação a um indivíduo na condição $(x_2 = 0, x_3 = 0)$, para $i, j = 0, 1$, supondo que os dois indivíduos têm os mesmos valores observados para as demais $(p - 4)$ variáveis explicativas. Assim, podemos mostrar facilmente que

$$\psi_{10} = \exp(\beta_2), \quad \psi_{01} = \exp(\beta_3) \quad \text{e} \quad \psi_{11} = \exp(\beta_2 + \beta_3 + \beta_4).$$

Portanto, testarmos a hipótese $H_0 : \beta_4 = 0$ (ausência de interação) é equivalente a testarmos a hipótese de efeito multiplicativo $H_0 : \psi_{11} = \psi_{10}\psi_{01}$. Em particular, se x_3 representa dois estratos ($x_3 = 0$, estrato 1; $x_3 = 1$, estrato 2), a razão de chances no primeiro estrato entre presença e ausência do fator fica dada por $\psi_{10} = \exp(\beta_2)$, enquanto que no segundo estrato essa razão de chances vale $\psi_{11}/\psi_{01} = \exp(\beta_2 + \beta_4)$. Logo, testarmos $H_0 : \beta_4 = 0$ equivale também a testarmos a hipótese de homogeneidade das razões de chances nos dois estratos.

3.6.4 Bandas de confiança

Como foi visto na Seção 1.8.2 uma banda assintótica de confiança de coeficiente $1 - \alpha$ pode ser construída para $\pi(\mathbf{z}), \forall \mathbf{z} \in \mathbb{R}^p$ (ver também Piegorisch e Casella, 1988). Assintoticamente $\hat{\beta} - \beta \sim N_p(\mathbf{0}, (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1})$. Logo, uma banda assintótica de confiança de coeficiente $1 - \alpha$ para o preditor linear $\mathbf{z}^T \beta, \forall \mathbf{z} \in \mathbb{R}^p$, fica dada por

$$\mathbf{z}^T \hat{\beta} \pm \sqrt{c_\alpha} \{ \mathbf{z}^T (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{z} \}^{1/2}, \quad \forall \mathbf{z} \in \mathbb{R}^p,$$

em que c_α é tal que $Pr\{\chi_p^2 \leq c_\alpha\} = 1 - \alpha$. Aplicando a transformação logito podemos, equivalentemente, encontrar uma banda de confiança de coeficiente $1 - \alpha$ para $\pi(\mathbf{z})$, dada por

$$\frac{\exp[\mathbf{z}^T \hat{\boldsymbol{\beta}} \pm \sqrt{c_\alpha} \{\mathbf{z}^T (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{z}\}^{1/2}]}{1 + \exp[\mathbf{z}^T \hat{\boldsymbol{\beta}} \pm \sqrt{c_\alpha} \{\mathbf{z}^T (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{z}\}^{1/2}]}, \quad \forall \mathbf{z} \in \mathbb{R}^p.$$

É importante observarmos que \mathbf{z} é um vetor $p \times 1$ que varia livremente no \mathbb{R}^p , enquanto \mathbf{X} é uma matriz fixa com os valores das variáveis explicativas.

3.6.5 Seleção de modelos

Uma vez definido o conjunto de covariáveis (ou fatores) a ser incluído num modelo logístico, resta saber qual a melhor maneira de encontrarmos um modelo reduzido que inclua apenas as covariáveis e interações mais importantes para explicarmos a probabilidade de sucesso $\pi(\mathbf{x})$. Esse problema poderia ser resolvido pelos métodos usuais de seleção de modelos discutidos na Seção 1.11. Contudo, a questão de interpretação dos parâmetros é crucial num modelo logístico, implicando que uma forma puramente mecânica de seleção pode levar a um modelo sem sentido e de difícil interpretação. Particularmente, a inclusão de certas interações impõe a permanência no modelo de seus respectivos efeitos principais de ordem inferior, na ótica do princípio hierárquico. Muitas vezes, variáveis consideradas biologicamente importantes não devem ser deixadas de lado pela sua falta de significância estatística. Assim, a seleção de um modelo logístico deve ser um processo conjugado de seleção estatística de modelos e bom senso.

Método stepwise

Um dos métodos mais aplicados em regressão logística é o método *stepwise*. O método, como foi visto na Seção 1.11, baseia-se num algoritmo misto de

inclusão e eliminação de variáveis explicativas segundo a importância das mesmas de acordo com algum critério estatístico. Esse grau de importância pode ser avaliado, por exemplo, pelo nível de significância do teste da razão de verossimilhanças entre os modelos que incluem ou excluem as variáveis em questão. Quanto menor for esse nível de significância tanto mais importante será considerada a variável explicativa. Como a variável mais importante por esse critério não é necessariamente significativa do ponto de vista estatístico, devemos impor um limite superior P_E (os valores usuais estão no intervalo $[0, 15; 0, 25]$) para esses níveis descritivos, a fim de atrairmos candidatos importantes em princípio à entrada.

Dado que a inclusão de novas variáveis explicativas num modelo pode tornar dispensáveis outras variáveis já incluídas, faremos a verificação da importância dessas variáveis confrontando os seus respectivos níveis com um limite superior P_S . As variáveis explicativas com um nível descritivo maior do que P_S serão assim candidatas à remoção.

Descrevemos a seguir uma variante desse algoritmo usada por Hosmer e Lemeshow (1989, Cap. 3) (ver também Silva, 1992). A etapa inicial começa com o ajustamento do modelo apenas com o intercepto e é completada pelos passos seguintes:

1. construímos testes da razão de verossimilhanças entre o modelo inicial e os modelos logísticos simples formados com cada uma das variáveis explicativas do estudo. O menor dos níveis descritivos associados a cada teste será comparado com P_E . Se P_E for maior, incluimos a variável referente àquele nível e passamos ao passo seguinte; caso contrário, paramos a seleção e adotamos o último modelo;
2. partindo do modelo incluindo a variável explicativa selecionada no passo anterior, introduzimos individualmente as demais variáveis. Cada

um desses novos modelos é testado contra o modelo inicial desse passo. Novamente, o menor valor dos níveis descritivos é comparado com P_E . Se for menor do que P_E , implica na inclusão no modelo da variável correspondente e a passagem ao passo seguinte. Caso contrário, paramos a seleção;

3. comparamos o desvio do modelo logístico contendo as variáveis selecionadas nos passos anteriores com os desvios dos modelos que dele resultam por exclusão individual de cada uma das variáveis. Se o maior nível descritivo dos testes da razão de verossimilhanças for menor do que P_S , a variável explicativa associada a esse nível descritivo permanece no modelo. Caso contrário, ela é removida. Em qualquer circunstância, o algoritmo segue para o passo seguinte;
4. o modelo resultante do passo anterior será ajustado, no entanto, antes de tornar-se o modelo inicial da etapa 2 (seleção de interações de primeira ordem entre as variáveis explicativas incluídas), avaliamos a significância de cada um dos coeficientes das variáveis selecionadas, por exemplo através de um teste de Wald. Se alguma variável explicativa não for significativa podemos excluí-la do modelo;
5. uma vez selecionadas as variáveis explicativas “mais importantes”, ou os efeitos principais, damos entrada na etapa 2 com o passo 1 que agora envolve apenas interações de primeira ordem entre as variáveis selecionadas, e assim por diante.

É comum que algumas variáveis explicativas ou interações de interesse ou com algum significado no estudo sejam mantidas no modelo desde o início, mesmo que não sejam significativas. É também comum que a seleção de

interações seja feita dentre aquelas de interesse ou com algum significado no problema.

Uma desvantagem do procedimento descrito pelos passos 1-5 é de exigir as estimativas de máxima verossimilhança em cada passo, o que encarece o trabalho computacional, particularmente quando há muitas variáveis explicativas (ou fatores). Alguns autores têm sugerido aproximações para esse processo de seleção. O aplicativo científico BMDP (Dixon, 1987) usa aproximações lineares nos testes da razão de verossimilhanças. Peduzzi, Hardy e Holford (1980) apresentam uma variante desse método baseada no uso da estatística de Wald.

Método de Akaike

Um procedimento mais simples para selecionarmos variáveis explicativas num modelo logístico é através do método de Akaike descrito na Seção 1.11. Uma sugestão é primeiro fazermos uma seleção dos efeitos principais e depois num segundo passo, das interações de 1^a ordem. Para ilustrarmos uma aplicação do método, vamos supor que as respostas binárias estejam armazenadas em **resp** e as variáveis explicativas sejam denotadas por **var1**, **var2** e **var3**. O ajuste do modelo logístico apenas com os efeitos principais pode ser realizado através dos comandos

```
ajuste <- glm(resp ~ var1 + var2 + var3, family=binomial).
```

A seleção dos efeitos principais pode ser realizada pelos comandos

```
require(MASS)
stepAIC(ajuste).
```

Eventualmente algumas variáveis explicativas selecionadas podem não ser significativas marginalmente e a retirada das mesmas do modelo poderá ser confirmada através de algum teste estatístico apropriado, como por exemplo

o teste da razão de verossimilhanças. A inclusão de interações de 1^a ordem pode ser feita individualmente dentre aquelas interações de interesse ou de fácil interpretação.

3.6.6 Amostragem retrospectiva

Em muitas situações práticas, especialmente no estudo de doenças raras, pode ser mais conveniente a aplicação de uma amostragem retrospectiva em que um conjunto de n_1 casos (indivíduos com $y = 1$) e n_2 controles (indivíduos com $y = 0$) é selecionado aleatoriamente e classificado segundo os valores de $\mathbf{x} = (x_1, \dots, x_p)^T$. Esse tipo de planejamento é muitas vezes motivado por questões econômicas ligadas ao custo e a duração do experimento. A amostragem retrospectiva assim constituída levaria diretamente a um modelo para $Pr(\mathbf{X} = \mathbf{x}|y)$, ao contrário dos dados prospectivos que estão associados ao modelo $\pi(\mathbf{x}) = Pr(Y = y|\mathbf{x})$. Como o desenvolvimento de um modelo para $Pr(\mathbf{X} = \mathbf{x}|y)$ pode ficar muito complexo à medida que o valor \mathbf{x} envolve um número maior de variáveis explicativas, particularmente contínuas, a proposta de uma abordagem alternativa através da especificação de um modelo para $Pr(Y = y|\mathbf{x})$, de modo a induzirmos um modelo para $Pr(\mathbf{X} = \mathbf{x}|y)$, tem sido bastante utilizada.

Vamos supor então um modelo logístico linear para explicarmos $\pi(\mathbf{x}) = Pr(Y = 1|\mathbf{x})$. Mostraremos a seguir que a probabilidade $\pi(\mathbf{x})$, a menos de uma constante adicionada ao intercepto do modelo, coincide com a probabilidade $\pi^*(\mathbf{x}) = Pr(Y = 1|\mathbf{x}, Z = 1)$ se a seleção amostral não depende de \mathbf{x} , em que Z é uma variável indicadora da classificação amostral (ver, por exemplo, Armitage, 1971). Denotaremos

$$\begin{aligned}\gamma_1 &= Pr(Z = 1|Y = 1) \text{ e} \\ \gamma_2 &= Pr(Z = 1|Y = 0),\end{aligned}$$

em que γ_1 é a probabilidade de um caso ser selecionado e γ_2 é a probabilidade de um controle ser selecionado da população global. Estamos supondo que γ_1 e γ_2 não dependem de \mathbf{x} . Portanto

$$\begin{aligned}\pi^*(\mathbf{x}) &= Pr(Y = 1|x, Z = 1) \\ &= \frac{Pr(Z = 1|Y = 1)Pr(Y = 1|\mathbf{x})}{\sum_{y=0,1} Pr(Z = 1|Y = y)Pr(Y = y|\mathbf{x})},\end{aligned}$$

que pode ser expressa em função de $\pi(\mathbf{x})$, ou seja

$$\begin{aligned}\pi^*(\mathbf{x}) &= \frac{\gamma_1 \pi(\mathbf{x})}{\gamma_2 \{1 - \pi(\mathbf{x})\} + \gamma_1 \pi(\mathbf{x})} \\ &= \frac{\frac{\gamma_1}{\gamma_2} \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right]}{1 + \frac{\gamma_1}{\gamma_2} \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right]}.\end{aligned}$$

Assim, obtemos

$$\pi^*(\mathbf{x}) = \frac{e^{\log\{\gamma_1/\gamma_2\} + \eta}}{1 + e^{\log\{\gamma_1/\gamma_2\} + \eta}},$$

em que $\eta = \sum_{j=1}^p x_j \beta_j$.

Portanto, se fazemos uma amostragem retrospectiva e ajustamos um modelo logístico como se fosse uma amostragem prospectiva, os coeficientes devem coincidir desde que a seleção tenha sido feita independente de \mathbf{x} . Se, no entanto, temos interesse particular em estimar $\pi(\mathbf{x})$, isto é, fazer predições dado \mathbf{x} , devemos corrigir a constante do modelo ajustado, obtendo o novo intercepto

$$\hat{\beta}_1 = \hat{\beta}_1^* - \log(\gamma_1/\gamma_2),$$

em que $\hat{\beta}_1^*$ é o intercepto do modelo ajustado. Apresentamos um exemplo ilustrativo na próxima seção.

3.6.7 Qualidade do ajuste

Como vimos na Seção 1.4, quando o número de grupos k é fixo num experimento binomial e $\frac{n_i}{n} \rightarrow a_i > 0$ quando $n \rightarrow \infty$, o desvio $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ segue sob

a hipótese do modelo adotado ser verdadeiro uma distribuição qui-quadrado com $(k - p)$ graus de liberdade. Esse resultado não vale quando $n \rightarrow \infty$ e $n_i \pi_i (1 - \pi_i)$ fica limitado. Nesse caso, Hosmer e Lemeshow (1989) sugerem uma estatística alternativa para avaliação da qualidade do ajuste. Essa estatística é definida comparando o número observado com o número esperado de sucessos de g grupos formados. O primeiro grupo deverá conter n'_1 elementos correspondentes às n'_1 menores probabilidades ajustadas, as quais serão denotadas por $\hat{\pi}_{(1)} \leq \hat{\pi}_{(2)} \leq \dots \leq \hat{\pi}_{(n'_1)}$. O segundo grupo deverá conter os n'_2 elementos correspondentes às seguintes probabilidades ajustadas $\hat{\pi}_{(n'_1+1)} \leq \hat{\pi}_{(n'_1+2)} \leq \dots \leq \hat{\pi}_{(n'_1+n'_2)}$. E assim, sucessivamente, até o último grupo que deverá conter as n'_g maiores probabilidades ajustadas $\hat{\pi}_{(n'_1+\dots+n'_{g-1}+1)} \leq \hat{\pi}_{(n'_1+\dots+n'_{g-1}+2)} \leq \dots \leq \hat{\pi}_{(n)}$. O número observado de sucessos no primeiro grupo formado será dado por $O_1 = \sum_{j=1}^{n'_1} y_{(j)}$, em que $y_{(j)} = 0$ se o elemento correspondente é fracasso e $y_{(j)} = 1$ se é sucesso. Generalizando, obtemos $O_i = \sum_{j=n'_1+\dots+n'_{i-1}+1}^{n'_1+\dots+n'_i} y_{(j)}$, $2 \leq i \leq g$. A estatística é definida por

$$\hat{C} = \sum_{i=1}^g \frac{(O_i - n'_i \bar{\pi}_i)^2}{n'_i \bar{\pi}_i (1 - \bar{\pi}_i)},$$

em que

$$\bar{\pi}_1 = \frac{1}{n'_1} \sum_{j=1}^{n'_1} \hat{\pi}_{(j)} \quad \text{e} \quad \bar{\pi}_i = \frac{1}{n'_i} \sum_{j=n'_1+\dots+n'_{i-1}+1}^{n'_1+\dots+n'_i} \hat{\pi}_{(j)},$$

para $2 \leq i \leq g$. Hosmer e Lemeshow sugerem a formação de $g = 10$ grupos de mesmo tamanho (aproximadamente), de modo que o primeiro grupo contenha n'_i elementos correspondentes às $[n/10]$ menores probabilidades ajustadas e assim por diante até o último grupo com n'_{10} elementos correspondentes às $[n/10]$ maiores probabilidades ajustados. Quando não há empates, isto é, $n_i = 1, \forall i$, fica relativamente fácil montarmos os 10 grupos com tamanhos aproximadamente iguais. No entanto, quando há empates, pode ser

necessário que dois indivíduos com a mesma configuração de covariáveis sejam alocados em grupos adjacentes a fim de que os grupos formados não tenham tamanhos muito desiguais. Hosmer e Lemeshow verificaram através de simulações que a distribuição nula assintótica de \hat{C} pode ser bem aproximada por uma distribuição qui-quadrado com $(g - 2)$ graus de liberdade.

3.6.8 Técnicas de diagnóstico

Estudos de simulação (ver, por exemplo, Williams, 1984) têm sugerido o resíduo t_{D_i} para as análises de diagnóstico em modelos lineares generalizados, uma vez que o mesmo tem apresentado nesses estudos propriedades similares às daquelas do resíduo t_i^* da regressão normal linear. Em particular, para os modelos binomiais, esse resíduo é expresso, para $0 < y_i < n_i$, na forma

$$t_{D_i} = \pm \sqrt{\frac{2}{1 - \hat{h}_{ii}}} \left\{ y_i \log \left(\frac{y_i}{n_i \hat{\pi}_i} \right) + (n_i - y_i) \log \left(\frac{n_i - y_i}{n_i - n_i \hat{\pi}_i} \right) \right\}^{1/2},$$

em que o sinal é o mesmo de $y_i - \hat{y}_i$. Quando $y_i = 0$ ou $y_i = n_i$, o componente do desvio padronizado toma as formas

$$t_{D_i} = -\frac{\{2n_i|\log(1 - \hat{\pi}_i)|\}^{1/2}}{\sqrt{1 - \hat{h}_{ii}}} \quad \text{e} \quad t_{D_i} = \frac{\{2n_i|\log \hat{\pi}_i|\}^{1/2}}{\sqrt{1 - \hat{h}_{ii}}},$$

respectivamente. O resíduo Studentizado t_{S_i} , também utilizado para avaliarmos a presença de observações aberrantes mesmo tendo em geral distribuição assimétrica acentuada, toma a forma

$$t_{S_i} = \frac{1}{\sqrt{1 - \hat{h}_{ii}}} \frac{(y_i - n_i \hat{\pi}_i)}{\{n_i \hat{\pi}_i (1 - \hat{\pi}_i)\}^{1/2}}.$$

Para medirmos a influência das observações nas estimativas dos coeficientes, utilizamos a distância de Cook aproximada dada por

$$LD_i = \frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})^2} \frac{(y_i - n_i \hat{\pi}_i)^2}{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}.$$

Hosmer e Lemeshow (1989) observam que \hat{h}_{ii} depende das probabilidades ajustadas $\hat{\pi}_i$, $i = 1, \dots, k$, e consequentemente os resíduos t_{S_i} e t_{D_i} e a medida de influência LD_i também dependem. Temos que

$$h_{ii} = n_i \pi_i (1 - \pi_i) \mathbf{x}_i^T (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{x}_i,$$

com $\mathbf{V} = \text{diag}\{n_1 \pi_1 (1 - \pi_1), \dots, n_k \pi_k (1 - \pi_k)\}$. Hosmer e Lemeshow mostram através de um estudo numérico que o comportamento de \hat{h}_{ii} numa regressão logística pode ser muito diferente do comportamento de \hat{h}_{ii} na regressão linear para uma mesma matrix modelo \mathbf{X} .

Tabela 3.6
Possíveis valores para algumas medidas de diagnóstico segundo as probabilidades ajustadas.

Medida	Probabilidade ajustada				
	0,0-0,1	0,1-0,3	0,3-0,7	0,7-0,9	0,9-1,0
$t_{S_i}^2$	grande ou pequeno	moderado	moderado ou pequeno	moderado	grande ou pequeno
LD_i	pequeno	grande	moderado	grande	pequeno
\hat{h}_{ii}	pequeno	grande	moderado ou pequeno	grande	pequeno

A Tabela 3.6 descreve os possíveis valores de algumas medidas de diagnóstico em função das probabilidades ajustadas. A medida \hat{h}_{ii} pode ser interpretada de maneira similar à medida h_{ii} da regressão normal linear para $0,1 \leq \hat{\pi}_i \leq 0,9$. No entanto, quando $\hat{\pi}_i$ é pequena ou alta, \hat{h}_{ii} fica em geral pequeno o que pode dificultar a detecção de pontos que estejam mais afastados no subespaço gerado pelas colunas da matrix \mathbf{X} . A sugestão, portanto, são os gráficos de $t_{S_i}^2$, $t_{D_i}^2$ e LD_i contra as probabilidades ajustadas $\hat{\pi}_i$. Esses gráficos podem ser informativos a respeito do posicionamento dos pontos aberrantes e influentes com relação às probabilidades ajustadas. Os gráficos dessas quantidades contra \hat{h}_{ii} podem ser complementares, pelo menos para

verificarmos se as tendências apresentadas na Tabela 3.11 são confirmadas para o modelo ajustado.

Outros gráficos recomendados em regressão logística são os gráficos da variável adicionada e de $|\ell_{max}|$ contra $\hat{\pi}_i$.

3.6.9 Aplicações

Processo infeccioso pulmonar

Voltemos agora ao exemplo discutido na Seção 1.12.2 em que 175 pacientes com processo infeccioso pulmonar foram classificados de acordo com as variáveis tipo de tumor, sexo, idade, nível de HL e nível de FF. Para simplicidade das análises, iremos reagrupar os níveis de HL e FF de modo que os níveis de intensidade “ausente” e “discreto” sejam agora considerados como intensidade “baixa” e os níveis “moderado” e “intenso” sejam agora de intensidade “alta” (ver Tabela 3.7).

Tabela 3.7

Descrição das novas variáveis referentes ao exemplo sobre processo infeccioso pulmonar.

Variável	Descrição	Valores
Y	Processo Infecioso	1:maligno 0:benigno
IDADE	Idade	em anos
SEXO	Sexo	0:masculino 1:feminino
HL	Intensidade de Histiócitos-linfócitos	1:alta 0:baixa
FF	Intensidade de Fibrose-frouxa	1:alta 0:baixa

Nesse estudo os pacientes foram amostrados retrospectivamente, sendo que os controles (processo benigno) foram formados por uma amostra de 104

pacientes de um grupo de 270, enquanto que os casos (processo maligno) foram todos os pacientes diagnosticados com processo infeccioso pulmonar maligno durante o período da pesquisa. Portanto, seguindo a notação da Seção 3.6.6, temos que $\gamma_1 = 1$ e $\gamma_2 = 104/270$ ¹.

Aplicaremos a seguir o método de seleção *stepwise* proposto por Hosmer e Lemeshow (1989). Na etapa 1 consideraremos apenas os efeitos principais. Adotaremos $P_E = 0,20$ (nível para inclusão de covariáveis) e $P_S = 0,25$ (nível para eliminação de covariáveis).

No passo 1 incluímos a variável explicativa IDADE, uma vez que o nível descritivo dessa variável foi o menor dentre os níveis descritivos das demais variáveis explicativas e também foi menor do que P_E . No passo seguinte incluímos a variável explicativa HL, e agora com duas variáveis incluídas no modelo verificamos se é possível eliminar uma das duas. O maior nível descritivo é da IDADE que encontra-se na Tabela 3.8 na linha de referência do passo 2. O nível descritivo dessa variável não é superior a P_S , logo mantemos a IDADE no modelo. Seguindo essa lógica, encontramos os menores níveis descritivos em cada passo como sendo o elemento da diagonal principal de cada passo. No passo 3, por exemplo, entra a variável explicativa SEXO que tem o menor nível descritivo que por sua vez é menor do que P_E . Dado que SEXO entra no modelo verificamos se uma das duas variáveis já incluídas no modelo pode sair. Assim, no mesmo passo 3, notamos que o maior nível descritivo (em asterisco) corresponde à variável explicativa HL que não deve sair do modelo pois o nível descritivo não é maior do que P_S . Seguindo essa mesma lógica todos os efeitos principais são incluídos no modelo. Em resumo, o modelo resultante na etapa 1 é o modelo com todos os efeitos principais.

De forma análoga procedemos a etapa 2, cujos níveis descritivos para

¹Estamos supondo que a razão $\gamma_1/\gamma_2 = 270/104$ vale também se as amostras tivessem sido feitas diretamente da população

tomada de decisão em cada passo encontram-se na Tabela 3.9. Por exemplo, no passo 1, entra a interação entre IDADE e HL que tem o menor nível descritivo que por sua vez é menor do que P_E . Não verificamos nessa etapa se algum efeito principal deve sair do modelo mesmo que fique não significativo com a inclusão das interações. Isso pode ser reavaliado após a seleção do modelo final. No passo 4, por exemplo, notamos que a interação entre IDADE e FF não entra no modelo pois o nível descritivo correspondente é maior do que P_E . Assim, como essa interação não entra no modelo, não é preciso verificar a retirada das demais interações já incluídas no modelo. Concluimos então que apenas três interações de primeira ordem serão incluídas no modelo. Essas interações são IDADE * HL, HL * FF e SEXO * FF.

Na etapa 3 nenhuma interação de segunda ordem foi selecionada, uma vez que o menor nível descritivo dos testes de inclusão foi menor do que P_E . Assim, o modelo resultante contém os efeitos principais e três interações de primeira ordem.

Tabela 3.8
*Níveis descritivos referentes à etapa 1
do processo de seleção stepwise.*

Passo	IDADE	HL	SEXO	FF
1	0,000	0,000	0,288	0,001
2	0,000	0,000	0,100	0,003
3	0,000	0,000*	0,050	0,125
4	0,000	0,000	0,072*	0,183
5	0,000	0,000	0,072	0,183*

O desvio do modelo foi de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 146,22$ (167 graus de liberdade), indicando um ajuste adequado. As Figuras 3.1a-3.1d apresentam alguns gráficos de diagnóstico. Na Figura 3.1a temos o gráfico de \hat{h}_{ii} contra os valores ajustados e notamos dois pontos com maior destaque, #6 e #69. No gráfico

de resíduos t_{D_i} , Figura 3.1b, a maioria dos pontos cai dentro do intervalo $[-2,2]$, com apenas duas observações, #21 e #172, fora do intervalo, porém muito próximas aos limites. Já o gráfico de influência LD_i destaca novamente a observação #69 e a observação #172. O paciente #172 é do sexo feminino, tem processo maligno, idade 55 anos e níveis altos para HL e FF. Pelos resultados das estimativas seria mais provável esperarmos de um paciente com esse perfil um processo benigno. O paciente #69 é também do sexo feminino, tem 78 anos, níveis altos para HL e FF e não tem processo maligno. Aqui seria um pouco menos provável processo benigno para o paciente. Perfil parecido tem o paciente #6. Já o paciente #21 tem processo benigno, 82 anos, é do sexo feminino e tem nível alto para HL e baixo para FF. Seria mais provável nesse caso processo maligno para o paciente.

Tabela 3.9

Níveis descritivos referentes à etapa 2 do processo de seleção stepwise.

Passo	IDA*HL	HL*FF	SEX*FF	IDA*FF	IDA*SEX	HL*SEX
1	0,013	0,014	0,059	0,056	0,657	0,063
2	0,023	0,027	0,060	0,231	0,218	0,099
3	0,028*	0,005	0,012	0,234	0,275	0,176
4				0,208	0,403	0,794

Finalmente, temos na Figura 3.1d o gráfico normal de probabilidades para o resíduo t_{D_i} e não notamos nenhum indício de que a distribuição utilizada seja inadequada. Retirando cada uma das observações destacadas pelos gráficos de diagnóstico notamos mudança inferencial quando a observação #172 é excluída, a interação SEXO * FF deixa de ser significativa. Ou seja, a significância da interação SEXO * FF é induzida pela observação #172. Logo, essa interação deve ser retirada do modelo.

As estimativas dos parâmetros do modelo final sem a interação SEXO * FF

bem como os valores padronizados pelos respectivos erros padrão aproximados encontram-se na Tabela 3.10.

Como temos interesse em estudar a associação entre o tipo de processo infeccioso pulmonar e as covariáveis histológicas HL e FF, formamos algumas razões de chances envolvendo essas covariáveis. Como ilustração, a razão de chances de processo infeccioso maligno entre um paciente no nível alto de HL e um paciente no nível baixo de HL, que denotaremos por ψ_{HL} , supondo que os pacientes tenham o mesmo sexo, idade e nível de FF, é estimada por

$$\hat{\psi}_{HL} = \exp\{-5,371 + 0,061\text{IDADE} + 2,255\text{FF}\}.$$

Tabela 3.10

Estimativas dos parâmetros referentes ao modelo logístico ajustado aos dados sobre processo infeccioso pulmonar.

Efeito	Parâmetro	Estimativa	E/E.Padrão
Constante	β_1^*	-1,247	-1,36
IDADE	β_2	0,038	2,23
HL	β_3	-5,371	-3,34
SEXO	β_4	0,765	1,60
FF	β_5	-2,090	-2,36
IDADE*HL	β_6	0,061	2,18
HL*FF	β_7	2,255	2,11

Logo, podemos concluir que a chance de processo maligno é maior para pacientes com nível baixo de HL do que para pacientes com nível alto de HL, quando ambos estão no nível baixo de FF e também tenham a mesma idade. Por outro lado, quando ambos estão na categoria alta de FF, $\hat{\psi}_{HL}$ fica maior do que um após a idade de 52 anos (aproximadamente), indicando uma chance maior de processo maligno para pacientes no nível alto de HL após essa idade.

Analogamente, denotaremos por ψ_{FF} a razão de chances de processo infeccioso maligno entre um paciente com nível alto de FF e um paciente com nível baixo de FF. Supondo que os pacientes são semelhantes nas demais covariáveis esse parâmetro é estimado por

$$\hat{\psi}_{FF} = \exp\{-2,090 + 2,255\text{HL}\}.$$

Dessa expressão podemos deduzir que a chance de processo maligno é maior para pacientes com intensidade baixa de FF do que para pacientes com intensidade alta de FF, isso no grupo de pacientes com intensidade baixa de HL. Ocorre o contrário no grupo de pacientes com intensidade alta de HL.

Na comparação dos pacientes com relação ao sexo temos que a razão de chances de processo infeccioso pulmonar entre pacientes do sexo feminino e masculino é estimada por $\hat{\psi}_{FM} = \exp(0,765) \cong 2,15$.

Se temos interesse em prever $Pr\{Y = 1|\mathbf{x}\}$, probabilidade de um paciente da população com um determinado conjunto de valores para as covariáveis estar com processo infeccioso maligno, devemos antes estimar β_1 fazendo a correção

$$\hat{\beta}_1 = \hat{\beta}_1^* - \log(270/104) = -1,247 - 0,954 = -2,201.$$

Tabela 3.11

Discriminação do modelo logístico ajustado aos dados sobre processo infeccioso pulmonar.

Classificação	Classificação pelo modelo	
	Benigno	Maligno
Correta		
Benigno	81	23
Maligno	13	58

A regressão logística tem múltiplas utilidades, entre as quais a possibilidade de também ser utilizada em análise discriminante quando há apenas dois

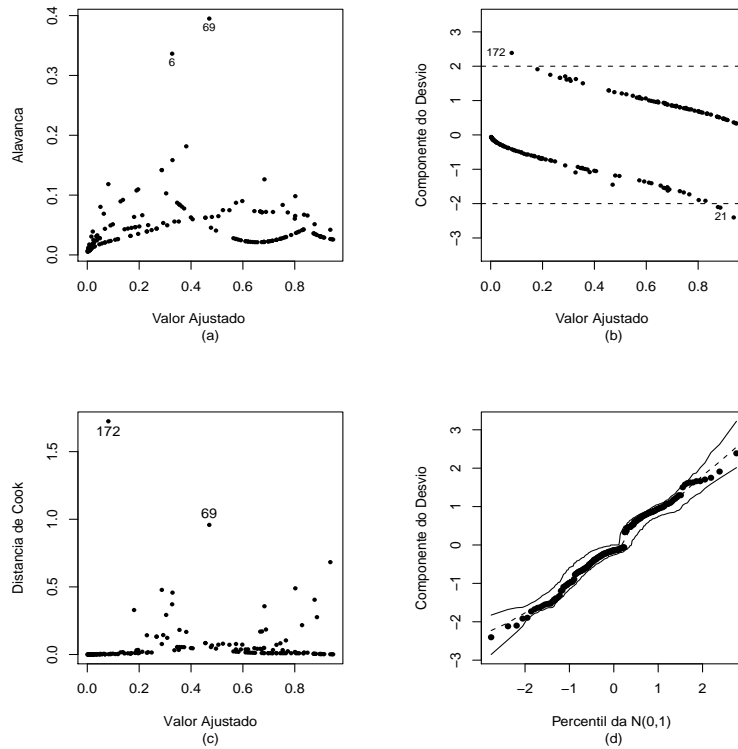


Figura 3.1: Gráficos de diagnóstico referentes ao modelo logístico ajustado aos dados sobre processo infeccioso pulmonar.

grupos para serem discriminados. O objetivo aqui é encontrarmos um modelo ajustado que melhor discrimine os dois grupos. Como aproximadamente 21% dos 341 pacientes foi diagnosticado com processo maligno podemos verificar qual a taxa de acertos do modelo ajustado. Um critério seria classificarmos com processo maligno todo indivíduo com probabilidade ajustada de pelo menos 0,21. Caso contrário o indivíduo seria classificado com processo benigno. A Tabela 3.11 apresenta a discriminação feita pelo modelo ajustado aos dados sobre processo infeccioso pulmonar. Temos que a taxa de acertos é de $139/175 = 0,795$ (79,5%).

Ocorrência de vaso-constricção

Como outra aplicação, vamos considerar os dados de um experimento desenvolvido para avaliar a influência da quantidade de ar inspirado na ocorrência de vaso-constricção na pele dos dedos da mão (Finney, 1978; Pregibon, 1981). Os dados do experimento são descritos na Tabela 3.12 e também no arquivo **pregibon.dat**. A resposta, nesse exemplo, é a ocorrência ($Y = 1$) ou ausência ($Y = 0$) de compressão de vasos e as covariáveis são o logaritmo do volume e o logaritmo da razão de ar inspirado.

Tabela 3.12

Dados do experimento sobre a influência da razão e do volume de ar inspirado na ocorrência de vaso-constricção da pele dos dedos da mão.

Obs	Volume	Razão	Resposta	Obs.	Volume	Razão	Resposta
1	3,70	0,825	1	20	1,80	1,800	1
2	3,50	1,090	1	21	0,40	2,000	0
3	1,25	2,500	1	22	0,95	1,360	0
4	0,75	1,500	1	23	1,35	1,350	0
5	0,80	3,200	1	24	1,50	1,360	0
6	0,70	3,500	1	25	1,60	1,780	1
7	0,60	0,750	0	26	0,60	1,500	0
8	1,10	1,700	0	27	1,80	1,500	1
9	0,90	0,750	0	28	0,95	1,900	0
10	0,90	0,450	0	29	1,90	0,950	1
11	0,80	0,570	0	30	1,60	0,400	0
12	0,55	2,750	0	31	2,70	0,750	1
13	0,60	3,000	0	32	2,35	0,030	0
14	1,40	2,330	1	33	1,10	1,830	0
15	0,75	3,750	1	34	1,10	2,200	1
16	2,30	1,640	1	35	1,20	2,000	1
17	3,20	1,600	1	36	0,80	3,330	1
18	0,85	1,415	1	37	0,95	1,900	0
19	1,70	1,060	0	38	0,75	1,900	0
				39	1,30	1,625	1

Vamos supor para a i -ésima unidade experimental que $Y_i \sim \text{Be}(\pi_i)$, em que

$$\log \left\{ \frac{\pi_i}{1 - \pi_i} \right\} = \beta_1 + \beta_2 \log(\text{volume})_i + \beta_3 \log(\text{razão})_i,$$

com π_i denotando a probabilidade de ocorrência de vaso-constricção.

As estimativas dos parâmetros são descritas na Tabela 3.13 e como podemos notar as variáveis explicativas $\log(\text{volume})$ e $\log(\text{razão})$ são altamente significativas. O desvio do modelo foi de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 29,36$ (com 36 graus de liberdade), indicando um ajuste adequado. As Figuras 3.2a-3.2d descrevem alguns dos gráficos sugeridos acima bem como o gráfico normal de probabilidades com envelope para o resíduo t_{D_i} . Na Figura 3.2a temos o gráfico de \hat{h}_{ii} contra os valores ajustados e podemos notar que a observação #31 é destacada mais do que as restantes.

Tabela 3.13

Estimativas dos parâmetros do modelo logístico ajustado aos dados sobre vaso-constricção.

Parâmetro	Estimativa	E/E.Padrão
β_1	-2,875	-2,18
β_2	5,179	4,85
β_3	4,562	2,49

Na Figura 3.2b temos o gráfico de LD_i contra os valores ajustados e notamos duas observações mais discrepantes, #4 e #18, cujos valores ajustados são menores do que 0,11. Uma tendência similar é exibida na Figura 3.2c onde temos o gráfico de $t_{S_i}^2$ contra os valores ajustados. A eliminação da observação #4 levou às novas estimativas $\hat{\beta}_1 = -5,204(2,17)$, $\hat{\beta}_2 = 7,452(2,93)$ e $\hat{\beta}_3 = 8,465(3,246)$ com variação, respectivamente, de -81%, 64% e 63%. O desvio do modelo reduziu para $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 22,42$ (35 g.l.), variação de 24%. Resultado parecido ocorreu com a eliminação da observação #18. Nesse caso obtemos $\hat{\beta}_1 = -4,757(2,008)$, $\hat{\beta}_2 = 6,879(2,718)$ e $\hat{\beta}_3 = 7,669(2,937)$

com variação, respectivamente, de -66%, 48% e 51%. O desvio caiu para $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 23,58$ (35 g.l.), redução de 20%. Mesmo com as variações desproporcionais não houve mudança inferencial. Esses resultados indicam que os pontos #4 e #18 são influentes e aberrantes. Note que para os dois casos houve ocorrência de ar inspirado, porém o valor do volume e da razão são relativamente baixos contrariando a tendência observada pelo modelo ajustado. O gráfico normal de probabilidades para o resíduo t_{D_i} (Figura 3.2d) não fornece indícios de afastamentos sérios da suposição de distribuição binomial para a resposta. Podemos notar a maioria dos pontos dentro do envelope gerado.

Tabela 3.14

Quantidades usadas para o cálculo da estatística \hat{C} referente ao modelo logístico ajustado aos dados sobre vaso-constricção.

Grupo	Observações	O_i	n'_i	$\bar{\pi}_i$
1	7,9,10,11,32	0	5	0,0024
2	4,18,21,26,30	2	5	0,0459
3	12,13,22,28,38	0	5	0,2737
4	8,19,23,29,37	1	5	0,5113
5	6,24,31,33,39	3	5	0,6728
6	5,15,34,35,36	5	5	0,7956
7	3,14,20,25,27	5	5	0,8974
8	1,2,16,17	4	4	0,9766

Apresentamos na Tabela 3.14 os grupos formados com as observações da Tabela 3.12 para o cálculo da estatística \hat{C} proposta por Hosmer e Lemeshow (1989). Foram formados sete grupos com cinco observações cada e um grupo

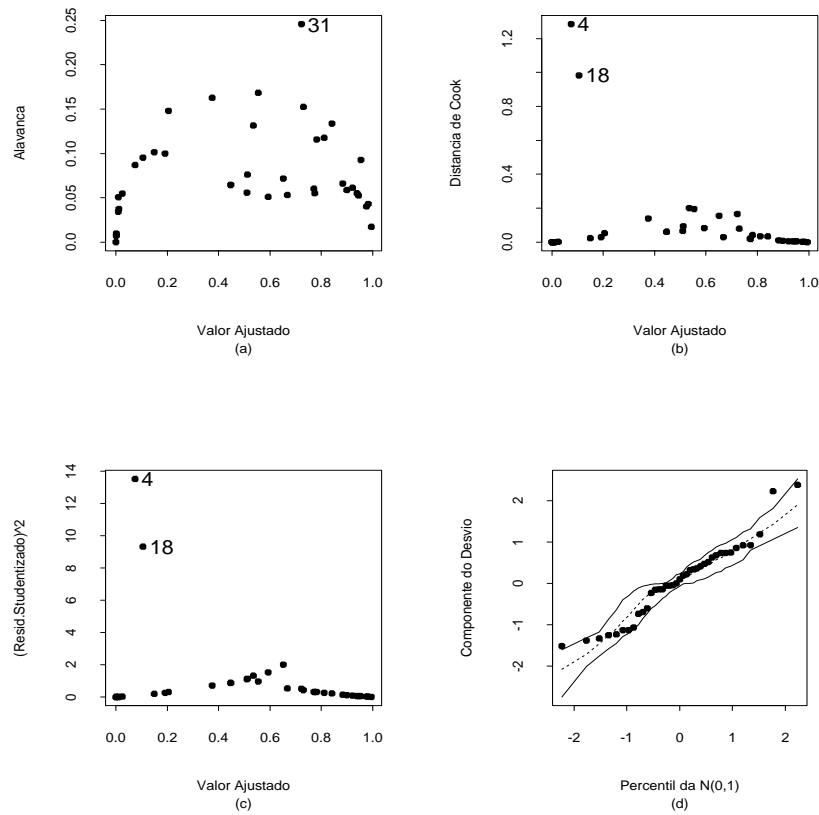


Figura 3.2: Gráficos de diagnóstico referentes ao modelo logístico ajustado aos dados sobre vaso-constricção.

com quatro observações. Os termos para o cálculo de \hat{C} são dados abaixo

$$\begin{aligned}
 \hat{C} &= 0,0120 + 14,3157 + 1,8842 + 1,9391 \\
 &+ 0,1203 + 1,2846 + 0,5716 + 0,0958 \\
 &= 20,2233,
 \end{aligned}$$

cujo nível descritivo para uma qui-quadrado com 6 graus de liberdade é dado por $P = 0,0025$, indicando que o ajuste não é adequado. Por outro lado, se eliminamos as observações #4 e #18, obtemos $\hat{C} = 5,9374$, que leva ao nível descritivo $P = 0,4302$. Portanto, as duas observações destacadas pelas

análises de diagnóstico têm grande influência na falta de ajuste detectada pela estatística \hat{C} . O programa para a geração do envelope da Figura 3.2d é descrito no Apêndice B. Assumimos que os resultados do ajuste estão disponíveis em `fit.model`.

Preferência de consumidores

Para ilustrarmos uma terceira aplicação com resposta binária vamos analisar parte dos dados descritos no arquivo **prefauto.dat** sobre a preferência de consumidores americanos com relação a automóveis. Uma amostra aleatória de 263 consumidores foi considerada. As seguintes variáveis foram observadas para cada comprador: preferência do tipo de automóvel (1: americano, 0: japonês), idade (em anos), sexo (0: masculino; 1: feminino) e estado civil (0: casado, 1: solteiro). Para maiores detalhes ver Foster, Stine e Waterman (1998, pgs. 338-339). Na Tabela 3.15 temos a distribuição da preferência do comprador segundo o sexo e estado civil, respectivamente.

Tabela 3.15

Distribuição da preferência do comprador de automóvel segundo o sexo e o estado civil.

	Masculino	Feminino
Americano	61 (42,4%)	54 (45,4%)
Japonês	83 (57,6%)	65 (54,6%)
Total	144	119
	Casado	Solteiro
Americano	83 (48,8%)	32 (34,4%)
Japonês	87 (51,2%)	65 (65,6%)
Total	170	93

Podemos notar que para ambos os sexos a maior preferência é por carro japonês. Dentre os casados há pequena vantagem por carro japonês. Con-

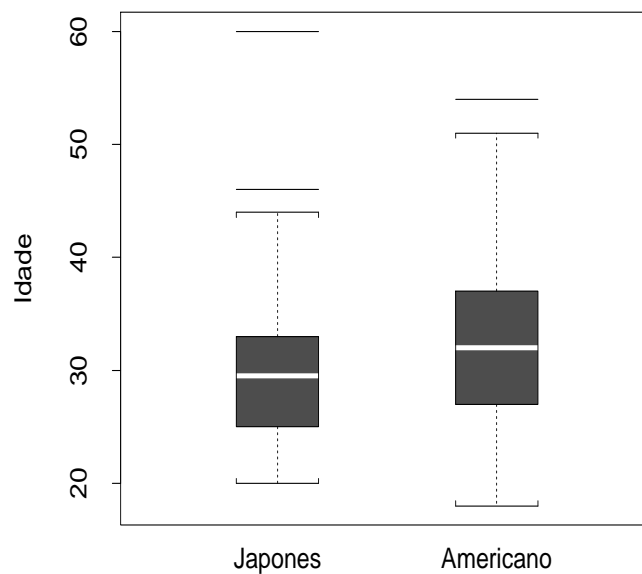


Figura 3.3: Boxplots da idade dos compradores de automóveis japonês e americano.

tudo, essa preferência é bem mais acentuada entre os solteiros. Pelos boxplots da Figura 3.3 notamos que a idade mediana dos compradores de automóvel americano é ligeiramente superior à idade mediana dos compradores de automóvel japonês.

Tabela 3.16

Estimativas dos parâmetros referentes ao modelo logístico ajustado aos dados sobre preferência de compradores.

Efeito	Estimativa	E/E.Padrão
Constante	-1,600	-2,31
Idade	0,049	2,30
E.Civil	-0,526	-1,94

Denotamos por Y_i a preferência com relação ao tipo do automóvel pelo i -ésimo comprador (1: americano, 0: japonês). Vamos supor inicialmente

um modelo logístico sem interação em que $Y_i \sim \text{Be}(\pi_i)$ com

$$\log \left\{ \frac{\pi_i}{1 - \pi_i} \right\} = \beta_1 + \beta_2 \text{idade}_i + \beta_3 \text{sexo}_i + \beta_4 \text{ecivil}_i,$$

sendo π_i a probabilidade do i -ésimo comprador preferir automóvel americano. Aplicando o método AIC retiramos a variável sexo do modelo. As estimativas dos parâmetros do modelo final sem interação são descritas na Tabela 3.16.

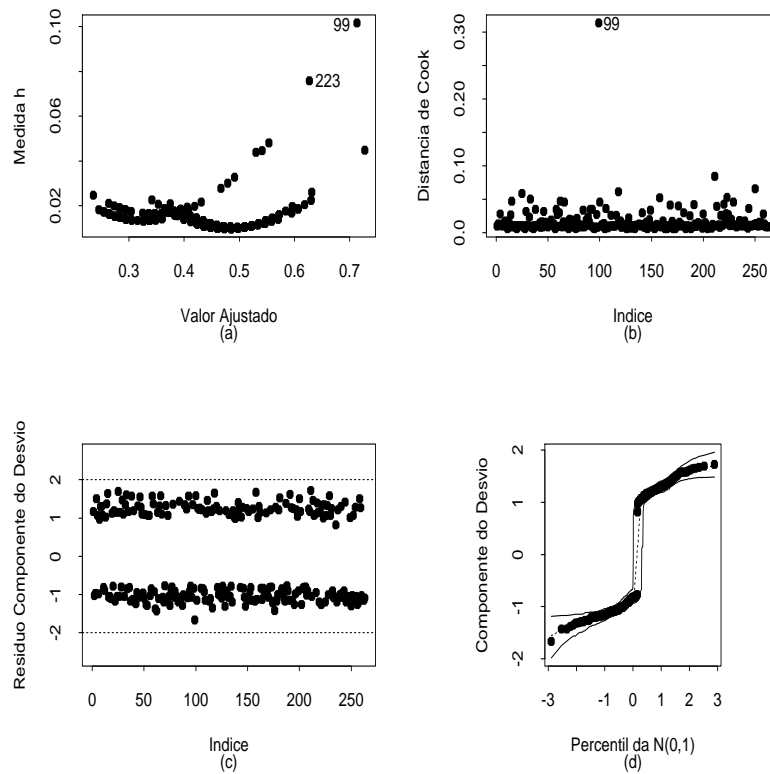


Figura 3.4: Gráficos de diagnóstico referentes ao modelo logístico ajustado aos dados sobre preferência de consumidores.

Não foi significativa a inclusão no modelo da interação entre a idade e o estado civil do comprador. Assim, temos que a preferência por automóvel

americano aumenta com a idade do comprador. Com relação ao estado civil notamos que os casados preferem mais carro americano do que os solteiros. Essa razão de chances (entre casados e solteiros) por carro americano pode ser estimada por $\hat{\psi} = \exp(0,526) = 1,69$. Nos gráficos de diagnóstico da Figura 3.4 duas observações são destacadas como possivelmente pontos de alavanca e influentes, #99 (idade de 60 anos, solteira e prefere carro japonês) e #223 (idade de 54 anos, solteiro e prefere carro americano). Porém, não há indícios de observações aberrantes nem de afastamentos sérios da suposição de distribuição de Bernoulli para a resposta. Na Tabela 3.17 são apresentadas as estimativas dos parâmetros sem essas duas observações. Embora ocorram algumas variações desproporcionais não há mudança inferencial.

Tabela 3.17

Estimativas dos parâmetros referentes ao modelo logístico ajustado aos dados sobre preferência de consumidores sem as observações #99 e #223.

Efeito	Estimativa	E/E.Padrão	Variação
Constante	-1,942	-2,65	-21,4%
Idade	0,060	2,65	22,4%
E.Civil	-0,474	-1,72	9,9%
Efeito	Estimativa	E/E.Padrão	Variação
Constante	-1,463	-2,07	8,6%
Idade	0,045	2,05	-8,1%
E.Civil	-0,550	-2,02	-4,6%

3.6.10 Modelos de dose-resposta

O modelo logístico é frequentemente utilizado em Toxicologia no estudo do comportamento de determinados medicamentos, que é medido pela probabilidade $\pi(x)$ de algum efeito produzido pelo medicamento em estudo, segundo a dose (ou a log-dose) x aplicada. Essa probabilidade pode ser escrita pela

expressão geral

$$\pi(x) = \int_{-\infty}^x f(u)du, \quad (3.15)$$

em que $f(\cdot)$ representa uma função densidade de probabilidade, também conhecida como função de tolerância. Como vimos na Seção 1.3.1, alguns candidatos naturais para $f(u)$ são as funções de densidade da normal padrão, da distribuição logística e da distribuição do valor extremo, as quais levam aos modelos probito, logístico e complementar log-log, respectivamente. Utilizamos o preditor linear $\eta = \beta_1 + \beta_2 x$ no lugar de x em (3.15) a fim de ampliarmos o leque de opções para $\pi(x)$.

Os modelos de dose-resposta visam não somente a predição da probabilidade de sucesso $\pi(x)$ para uma dosagem específica x , mas também a determinação da dosagem necessária para atingirmos uma probabilidade de sucesso p . Essa dosagem é chamada de dose letal. A notação usual para uma dose letal de $100p\%$ é dada por DL_{100p} . Logo,

$$p = \pi(\beta_1 + \beta_2 DL_{100p}), \quad 0 < p < 1.$$

A dose letal mais comum em Toxicologia é a dose mediana (DL_{50}), embora em certos casos sejam também de interesse doses extremas, tais como DL_1 ou DL_{99} . Devemos observar que hoje em dia modelos de dose-resposta são definidos em várias áreas do conhecimento, em que a dose pode ser a idade, o peso, a resistência de um material etc.

Supondo o modelo logístico com preditor linear $\eta = \beta_1 + \beta_2 x$, a estimativa de máxima verossimilhança de DL_{100p} fica, pela propriedade de invariância, dada por

$$\widehat{DL}_{100p} = d(\hat{\beta}) = \frac{1}{\hat{\beta}_2} \left[\log \left(\frac{p}{1-p} \right) - \hat{\beta}_1 \right],$$

em que $\hat{\beta}$ é a estimativa de máxima verossimilhança de $\beta = (\beta_1, \beta_2)^T$.

A variância assintótica de $\widehat{\text{DL}}_{100p}$ pode ser obtida após uma aproximação de primeira ordem por série de Taylor de $d(\hat{\boldsymbol{\beta}})$ em torno de $\boldsymbol{\beta}$, levando ao seguinte:

$$\text{Var}_A[\widehat{\text{DL}}_{100p}] = D(\boldsymbol{\beta})^T (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} D(\boldsymbol{\beta}),$$

em que

$$D(\boldsymbol{\beta}) = \partial d(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} = \left[\frac{-1}{\beta_2}, \frac{1}{\beta_2^2} \left\{ \beta_1 - \log \left(\frac{p}{1-p} \right) \right\} \right]^T.$$

Lembramos que $(\mathbf{X}^T \hat{\mathbf{V}} \mathbf{X})^{-1}$ contém as variâncias e covariância estimadas de $\hat{\beta}_1$ e $\hat{\beta}_2$. Portanto, um intervalo de confiança assintótico de coeficiente $(1 - \alpha)$ para DL_{100p} fica dado por

$$\widehat{\text{DL}}_{100p} \pm z_{(1-\alpha/2)} \sqrt{\text{Var}_A[d(\hat{\boldsymbol{\beta}})]}.$$

3.6.11 Aplicações

Exposição de besouros

Tabela 3.18

*Mortalidade de besouros expostos
a disulfeto de carbono gasoso.*

Dose $\log_{10} C S_2$	Besouros expostos	Besouros mortos
1,6907	59	6
1,7242	60	13
1,7552	62	18
1,7842	56	28
1,8113	63	52
1,8369	59	53
1,8610	62	61
1,8839	60	60

Em Bliss (1935) (ver também Silva, 1992) encontramos uma situação típica para o ajuste de um modelo logístico de dose-resposta. O estudo

baseia-se no comportamento de besouros adultos à exposição de disulfeto de carbono gasoso (CS_2) durante cinco horas. Os resultados obtidos a partir dos 481 besouros expostos segundo diferentes doses são apresentados na Tabela 3.18 e no arquivo **besouros.dat**. Ajustando um modelo logístico do tipo $\text{logit}\{\pi(x)\} = \beta_1 + \beta_2 x$ aos dados, em que x denota a dose de CS_2 , obtemos as estimativas $\hat{\beta}_1 = -60,72(5,18)$, $\hat{\beta}_2 = 34,27(2,91)$ e $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -15,04$. O desvio do modelo foi de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 11,23$ para 6 graus de liberdade, o que leva a um nível descritivo de $P = 0,0815$, indicando um ajuste razoável. O gráfico de envelope descrito na Figura 3.5 confirma essa falta de ajuste. Talvez a inclusão de um termo quadrático ou mesmo o ajuste de um modelo logístico não linear (ver Silva, 1992) possam melhorar a qualidade do ajuste.

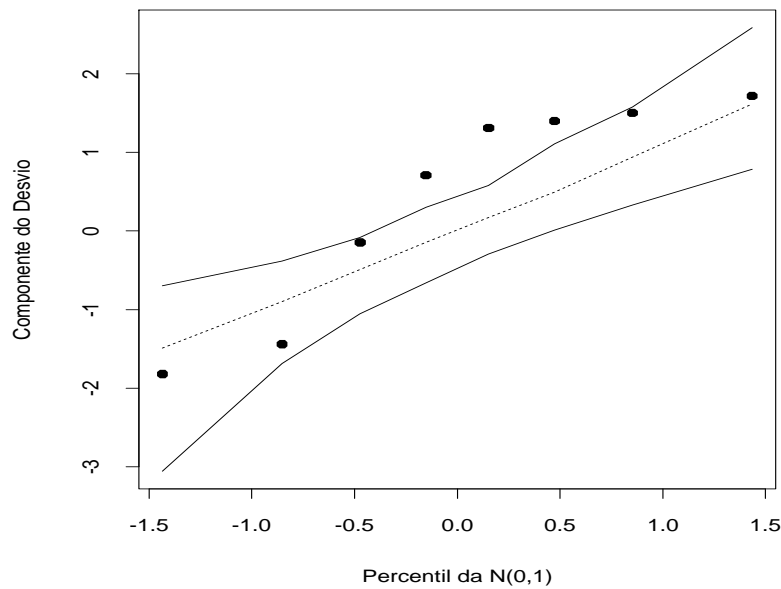


Figura 3.5: Gráfico normal de probabilidades referente ao modelo logístico ajustado aos dados sobre exposição de besouros.

Uma vez conhecida a covariância assintótica entre $\hat{\beta}_1$ e $\hat{\beta}_2$, podemos calcu-

lar a variância assintótica de DL_{100p} para alguns valores de p e consequentemente os intervalos assintóticos de confiança. Em particular, para $p = 0,50$, obtemos

$$\begin{aligned}\widehat{DL}_{50} &= \frac{1}{\hat{\beta}_2} \left[\log \left(\frac{0,5}{1-0,5} \right) - \hat{\beta}_1 \right] \\ &= -\frac{\hat{\beta}_1}{\hat{\beta}_2} = \frac{60,72}{34,27} \\ &= 1,772.\end{aligned}$$

Um intervalo de confiança assintótico de 95% para DL_{50} fica então dado por

$$\begin{aligned}&1,772 \pm 1,96 \sqrt{(-0,029, -0,052)^T (\mathbf{X}^T \hat{\mathbf{V}} \mathbf{X})^{-1} \begin{pmatrix} -0,029 \\ -0,052 \end{pmatrix}} \\ &= 1,772 \pm 1,96 \sqrt{0,0001488} \\ &= [1,764; 1,780].\end{aligned}$$

A Figura 3.6 descreve a curva ajustada e as frequências observadas. Como podemos observar os pontos abaixo de $\hat{\pi}(x) = 0,50$ parecem mais mal ajustados do que os pontos com resposta estimada acima desse valor. Isso sugere que um modelo binomial com ligação complemento log-log poderia ser mais apropriado. A parte sistemática desse modelo fica expressa na forma

$$\log\{-\log(1 - \pi(x))\} = \beta_1 + \beta_2 x,$$

em que x denota a dose de CS_2 . As estimativas paramétricas ficam dadas por $\hat{\beta}_1 = -39,57(3,24)$, $\hat{\beta}_2 = 22,04(1,80)$ e $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -5,82$. O desvio do modelo caiu para $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 3,45$ com 6 graus de liberdade, que leva a um nível descritivo de $P = 0,751$. Logo, não rejeitamos o modelo. O gráfico

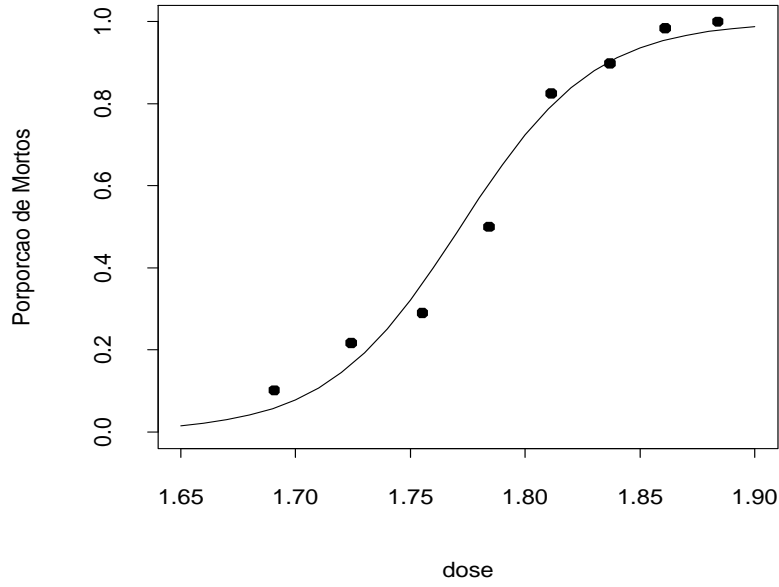


Figura 3.6: Modelo logístico ajustado à proporção de besouros mortos.

da curva ajustada (Figura 3.7a) e o gráfico normal de probabilidades (Figura 3.7b) confirmam essa indicação de modelo bem ajustado.

Para o modelo com ligação complemento log-log a estimativa de máxima verossimilhança de DL_{100p} fica dada por

$$\widehat{DL}_{100p} = d(\hat{\beta}) = \frac{1}{\hat{\beta}_2} \left[\log\{-\log(1-p)\} - \hat{\beta}_1 \right],$$

para a qual obtemos a variância assintótica

$$\text{Var}_A[\widehat{DL}_{100p}] = D(\beta)^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} D(\beta),$$

em que

$$D(\beta) = \partial d(\beta) / \partial \beta = \left[\frac{-1}{\beta_2}, \frac{1}{\beta_2^2} \{ \beta_1 - \log(-\log(1-p)) \} \right]^T,$$

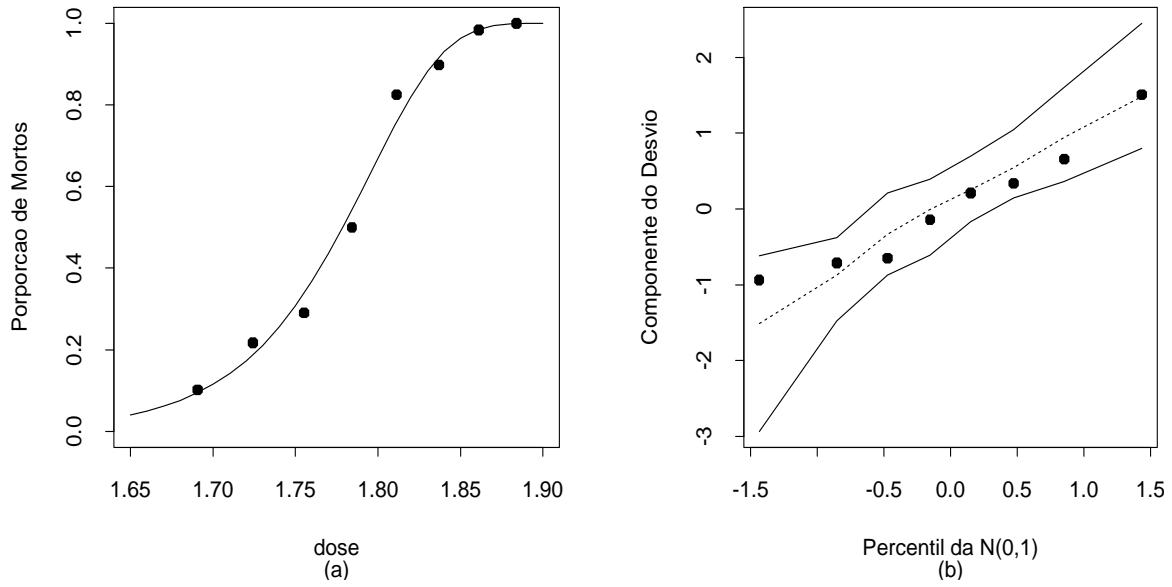


Figura 3.7: Curva ajustada para a proporção de besouros mortos (a) e gráfico normal de probabilidades sob o modelo complementar log-log (b).

com \mathbf{W} sendo uma matriz diagonal de pesos dados por $\omega_i = n_i \pi_i^{-1} (1 - \pi_i) \log^2(1 - \pi_i)$ $i = 1, \dots, 8$. Em particular, para $p = 0,50$, obtemos

$$\begin{aligned} \widehat{DL}_{50} &= \frac{1}{\hat{\beta}_2} \left[\log\{-\log(1 - 0,5)\} - \hat{\beta}_1 \right] \\ &= \frac{1}{22,04} (-0,3665 + 39,57) \\ &= 1,779. \end{aligned}$$

Logo, um intervalo assintótico de 95% para DL_{50} fica dado por

$$\begin{aligned} &1,779 \pm 1,96 \sqrt{(-0,0454, -0,0807)^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \begin{pmatrix} -0,0454 \\ -0,0807 \end{pmatrix}} \\ &= 1,779 \pm 1,96 \sqrt{0,00001606} \\ &= [1,771; 1,787]. \end{aligned}$$

Podemos notar que as estimativas intervalares para DL_{50} são praticamente as mesmas sob os dois modelos ajustados.

Garotas de Varsóvia

Os problemas de dose-resposta não se esgotam em Toxicologia. Milecer e Szczotka (1966) investigam a idade do início da menstruação em 3918 garotas de Varsóvia. Para 25 médias de idade foram observadas a ocorrência ($Y = 1$) ou não ($Y = 0$) do início de períodos de menstruação nas adolescentes. Os dados desse estudo são apresentados na Tabela 3.19 e no arquivo **meninas.dat**. Adotamos o modelo logístico linear

$$\log \left\{ \frac{\pi(x)}{1 - \pi(x)} \right\} = \beta_1 + \beta_2 x,$$

em que $\pi(x) = Pr\{Y = 1|x\}$ e x denota a idade média. As estimativas de máxima verossimilhança deram $\hat{\beta}_1 = -21,23(0,769)$, $\hat{\beta}_2 = 1,63(0,059)$ e $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -0,045$. Na Figura 3.8 são apresentadas a curva ajustada e as frequências observadas. O desvio do modelo foi de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 26,80$ (23 graus de liberdade) para um nível descritivo de $P = 0,264$, indicando um ajuste adequado.

A estimativa da idade mediana de início do período de menstruação fica portanto dada por

$$\widehat{DL}_{50} = \frac{21,23}{1,63} = 13,02,$$

com o seguinte intervalo assintótico de confiança de 95%:

$$13,02 \pm 1,96\sqrt{0,004524} = [12,89; 13,15].$$

Pelo gráfico de envelope descrito na Figura 3.9a notamos que os resíduos apresentam uma tendência sistemática dentro do envelope gerado, sugerindo a inclusão de um termo quadrático na parte sistemática do modelo. O ajuste de um modelo com parte sistemática dada por

$$\eta(x) = \beta_1 + \beta_2 x + \beta_3 x^2$$

forneceu as seguintes estimativas: $\hat{\beta}_1 = -30,96(5,24)$, $\hat{\beta}_2 = 3,12(0,78)$ e $\hat{\beta}_3 = -0,06(0,03)$ com desvio $D(\mathbf{y}, ; \hat{\boldsymbol{\mu}}) = 23,40$ (22 graus de liberdade) para um nível descritivo de $P = 0,38$. O gráfico de envelope descrito na Figura 3.9b confirma a adequação do modelo com termo quadrático.

Tabela 3.19
Ocorrência do início da menstruação em garotas de Varsóvia.

Idade	Número de garotas		Idade	Número de garotas	
	Menstruadas	Entrevistadas		Menstruadas	Entrevistadas
9,21	0	376	13,08	47	99
10,21	0	200	13,33	67	106
10,58	0	93	13,58	81	105
10,83	2	120	13,83	88	117
11,08	2	90	14,08	79	98
11,33	5	88	14,33	90	97
11,58	10	105	14,58	113	120
11,83	17	111	14,83	95	102
12,08	16	100	15,08	117	122
12,33	29	93	15,33	107	111
12,58	39	100	15,58	92	94
12,83	51	108	15,83	112	114
			17,53	1049	1049

Stukel (1988) (ver também Silva, 1992) mostra que o uso de um modelo logístico não linear pode melhorar substancialmente a qualidade do ajuste dos modelos de dose-resposta apresentados nesta seção.

3.6.12 Estimação da dose letal

Intervalos de confiança aproximados para a dose letal DL_{100p} podem ser construídos utilizando a variância assintótica para \widehat{DL}_{100p} , conforme descrito na

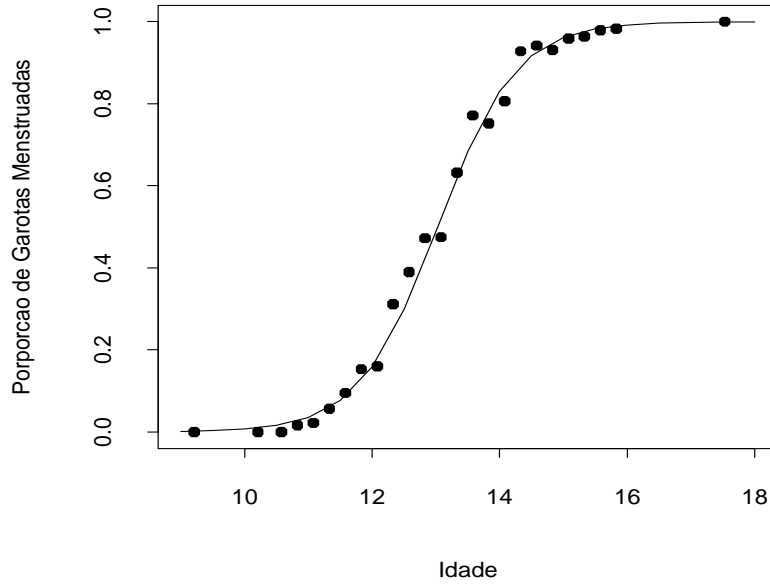


Figura 3.8: Curva ajustada pelo modelo logístico linear para a proporção de garotas de Varsóvia.

seção anterior. Há, contudo, um outro método que é baseado no teorema de Fieller (1954) e será descrito a seguir. Chamamos $\rho = \frac{\beta_0}{\beta_1}$, em que β_0 e β_1 são estimados por $\hat{\beta}_0$ e $\hat{\beta}_1$ e assumimos que essas estimativas são normalmente distribuídas com médias β_0 e β_1 , variâncias v_{00} e v_{11} e covariância v_{01} . Definimos a função $\hat{\psi} = \hat{\beta}_0 - \rho\hat{\beta}_1$. Então, se $\hat{\beta}_0$ e $\hat{\beta}_1$ são estimativas não viesadas de β_0 e β_1 , obtemos $E(\hat{\psi}) = 0$. A variância de $\hat{\psi}$ fica, portanto, dada por

$$v = \text{Var}(\hat{\psi}) = v_{00} + \rho^2 v_{11} - 2\rho v_{01}. \quad (3.16)$$

Desde que $\hat{\beta}_0$ e $\hat{\beta}_1$ são normalmente distribuídos, então $\hat{\psi}$ também é normalmente distribuído. Consequentemente, a variável

$$(\hat{\beta}_0 - \rho\hat{\beta}_1)/\sqrt{v}$$

segue uma distribuição normal padrão. Assim, um intervalo assintótico de confiança para ρ com coeficiente $(1 - \alpha)$ é formado pelos valores de ρ tais

que

$$|\hat{\beta}_0 - \rho\hat{\beta}_1| \leq z_{(1-\alpha/2)}\sqrt{v}.$$

Os limites desse intervalo de confiança saem da equação quadrática

$$\hat{\beta}_0^2 + \rho^2\hat{\beta}_1^2 - 2\rho\hat{\beta}_0\hat{\beta}_1 - z_{(1-\alpha/2)}^2 v = 0,$$

que, após algumas manipulações algébricas e usando (3.16), fica dada por

$$(\hat{\beta}_1^2 - z_{(1-\alpha/2)}^2 v_{11})\rho^2 + (2v_{01}z_{(1-\alpha/2)}^2 - 2\hat{\beta}_0\hat{\beta}_1)\rho + \hat{\beta}_0^2 - v_{00}z_{(1-\alpha/2)}^2 = 0.$$

Portanto, as raízes da equação acima formam os limites inferior e superior do intervalo de confiança para ρ . Basta chamarmos $\rho = -\beta_1/\beta_2$ e aplicarmos os resultados acima para encontrarmos um intervalo assintótico de coeficiente $(1 - \alpha)$ para DL_{50} .

3.6.13 Modelos de retas paralelas

Modelos de retas paralelas são comumente aplicados na área de Farmacologia para compararmos a eficiência de drogas do mesmo tipo, ou seja, com ação similar (ver, por exemplo, Finney, 1971; Collett, 1991). Nesses estudos, o interesse principal é compararmos as potências entre as drogas definindo uma droga particular como nível base ou droga padrão. Para aplicarmos esses modelos em experimentos com respostas binárias assumimos que Y_{ijk} , o efeito produzido pela j -ésima dose correspondente à i -ésima droga no k -ésimo indivíduo, $i = 1, \dots, g$, $j = 1, \dots, d_i$ e $k = 1, \dots, n_{ij}$, segue uma distribuição de Bernoulli com probabilidade de sucesso π_{ij} definida tal que

$$g(\pi_{ij}) = \alpha_i + \beta \log x_{ij}, \quad (3.17)$$

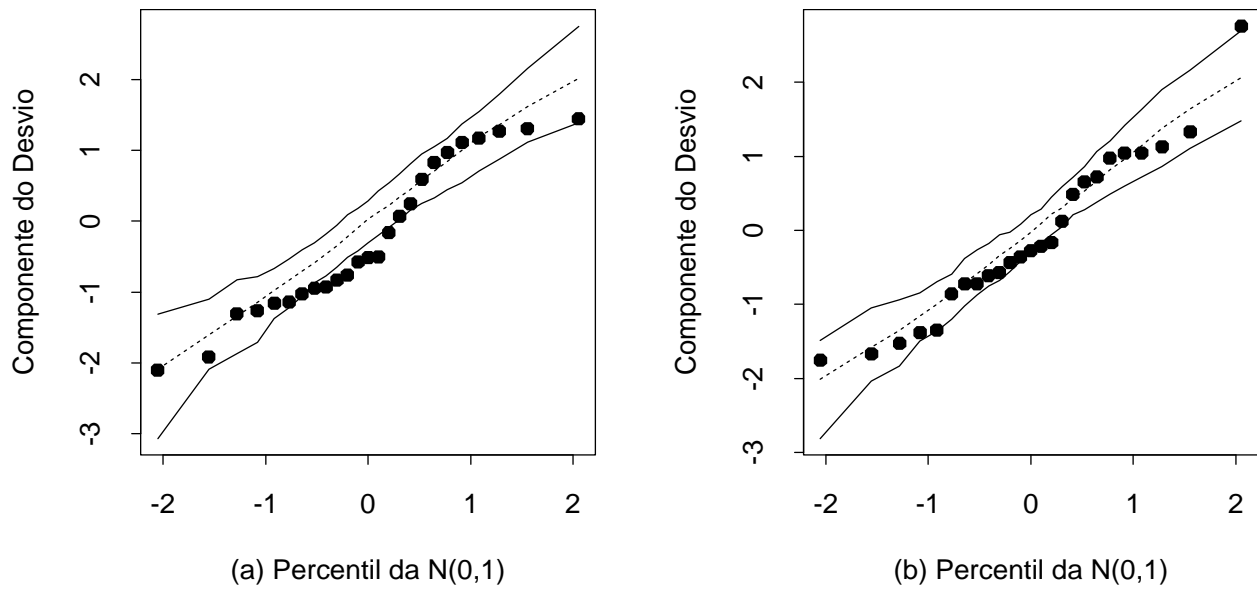


Figura 3.9: Gráficos normais de probabilidades para os modelos logístico com componente sistemática linear (a) e não linear (b) ajustados aos dados sobre garotas de Varsóvia.

e que as variáveis Y_{ijk} 's são mutuamente independentes. Se tomarmos a primeira droga como padrão, a potência ρ_i da i -ésima droga com relação à primeira é definida por

$$\log \rho_i = (\alpha_i - \alpha_1) / \beta,$$

$i = 1, \dots, g$. Essa suposição leva à seguinte relação:

$$g(\pi_{ij}) = \alpha_1 + \beta \log \rho_i x_{ij},$$

isto é, x unidades da droga i têm o mesmo efeito que $\rho_i x$ unidades da primeira droga.

Aplicação

A Tabela 3.20 resume os resultados de um experimento (ver Collett, 1991) em que três inseticidas são aplicados num determinado tipo de inseto e é

verificado o número de sobreviventes para cada dose aplicada. Esses dados estão também descritos no arquivo **insetic.dat**.

Tabela 3.20

Mortalidade de insetos segundo as doses de três inseticidas.

Inseticida	Dose mg/cm^2					
	2,00	2,64	3,48	4,59	6,06	8,00
DDT	3/50	5/49	19/47	19/50	24/49	35/50
γ -BHC	2/50	14/49	20/50	27/50	41/50	40/50
DDT + γ -BHC	28/50	37/50	46/50	48/50	48/50	50/50

Ajustando o modelo (3.17) com ligação logito aos dados, obtemos as estimativas $\hat{\alpha}_1 = -4,555(0,361)$, $\hat{\alpha}_2 = -3,842(0,333)$, $\hat{\alpha}_3 = -1,425(0,285)$ e $\hat{\beta} = 2,696(0,214)$, com desvio dado por $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 21,282$, para 14 graus de liberdade, $P = 0,0946$. Isso quer dizer que o ajuste do modelo de retas paralelas parece ser razoável.

Temos, portanto, os seguintes ajustes para as três drogas:

$$\begin{aligned} \log \left\{ \frac{\hat{\pi}_1(x_j)}{1 - \hat{\pi}_1(x_j)} \right\} &= -4,555 + 2,696 \log x_j \quad (\text{DDT}); \\ \log \left\{ \frac{\hat{\pi}_2(x_j)}{1 - \hat{\pi}_2(x_j)} \right\} &= -3,842 + 2,696 \log x_j \quad (\gamma\text{-BHC}) \quad \text{e} \\ \log \left\{ \frac{\hat{\pi}_3(x_j)}{1 - \hat{\pi}_3(x_j)} \right\} &= -1,425 + 2,696 \log x_j \quad (\text{DDT} + \gamma\text{-BHC}), \end{aligned}$$

para $j = 1, \dots, 6$. Notamos, pelas estimativas, que há um aumento de potência quando as drogas DDT e γ -BHC são misturadas. Em particular, a potência da mistura com relação às drogas DDT e γ -BHC é estimada, respectivamente, por $\hat{\rho}_1 = \exp\{(-1,425 + 4,555)/2,696\} = 3,19$ e $\hat{\rho}_2 = \exp\{(-1,425 + 3,842)/2,696\} = 2,45$.

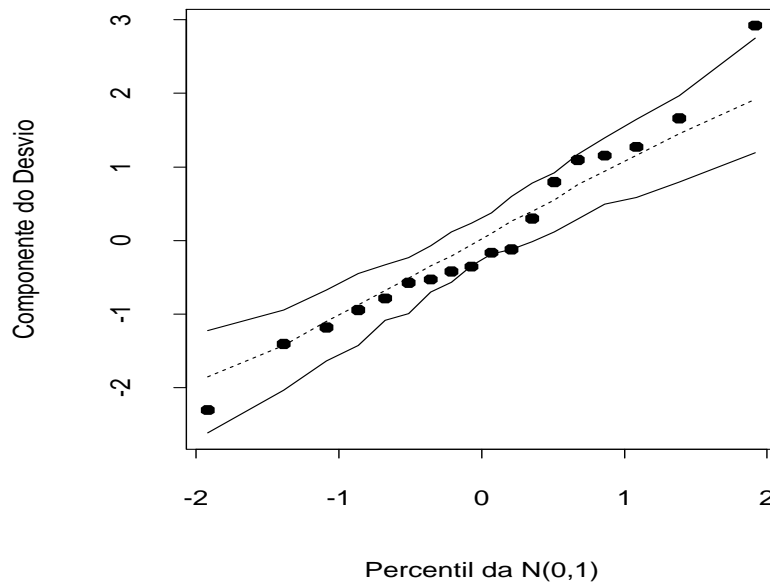


Figura 3.10: Gráfico normal de probabilidades para o modelo logístico de retas paralelas ajustado aos dados sobre três tipos de inseticida.

Pelo gráfico normal de probabilidades (Figura 3.10), notamos que todos os resíduos caem dentro do envelope gerado. No entanto, parece haver uma tendência no gráfico, uma vez que os resíduos negativos apresentam-se ligeiramente abaixo da média enquanto os resíduos positivos apresentam-se ligeiramente acima. Isso pode ser um indício de sobredispersão, isto é, que as réplicas (para cada dose e cada inseticida) não são totalmente independentes. Em Collett (1991, Cap. 6) há uma discussão sobre o assunto. Apresentaremos a seguir uma abordagem para esse tipo de problema.

3.6.14 Sobredispersão

Sobredispersão ou variação extrabinomial é um fenômeno comum que ocorre na modelagem de dados binários agrupados e cuja ocorrência é caracterizada quando a variação observada excede aquela assumida pelo modelo (ver, por

exemplo, Hinde e Demétrio, 1998). Em particular em regressão logística, quando o desvio $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ é maior que o número de graus de liberdade $(g - p)$, pode haver indícios de sobredispersão, em que g é o número de grupos. Isso pode ser avaliado mais precisamente pelo nível descritivo do teste de ajustamento comparando $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ com os percentis da distribuição qui-quadrado com $(g - p)$ graus de liberdade.

Diferentes circunstâncias, entretanto, podem causar um valor alto para o desvio. Algumas delas representam uma sobredispersão aparente. Por exemplo, alguns pontos aberrantes podem aumentar substancialmente o valor do desvio e a simples eliminação desses pontos pode reduzir as evidências de sobredispersão. Outra causa aparente de sobredispersão é a ausência de algum termo extra na parte sistemática do modelo. Medidas de diagnóstico são ferramentas importantes para detectarmos o fenômeno. Em síntese, há duas possíveis causas de sobredispersão: correlação entre as réplicas binárias ou variação entre as probabilidades de sucesso de um mesmo grupo. Do ponto de vista prático é difícil distinguirmos entre os dois casos, contudo, como veremos a seguir, os procedimentos estatísticos para tratarmos o problema podem ser os mesmos.

Caso I

Vamos supor inicialmente a existência de g grupos de modo que para o i -ésimo grupo sejam observadas n_i repetições de uma variável aleatória $Y_{ij} \sim \text{Be}(\pi_i)$ (Bernoulli com probabilidade de sucesso π_i). O número total de sucessos no i -ésimo grupo será definido por

$$Y_i = Y_{i1} + \cdots + Y_{in_i}.$$

Segue que $E(Y_{ij}) = \pi_i$ e $\text{Var}(Y_{ij}) = \pi_i(1 - \pi_i)$. Vamos supor adicionalmente a existência de correlação entre as repetições do i -ésimo grupo. Logo,

$$\text{Var}(Y_i) = \sum_{j=1}^{n_i} \text{Var}(Y_{ij}) + \sum_{j=1}^{n_i} \sum_{k=1, k \neq j}^{n_i} \text{Cov}(Y_{ij}, Y_{ik}).$$

Se essa correlação é constante, $\text{Corr}(Y_{ij}, Y_{ik}) = \delta$ para $j \neq k$, então teremos que $\text{Cov}(Y_{ij}, Y_{ik}) = \delta\pi_i(1 - \pi_i)$. Daí obtemos

$$\begin{aligned} \text{Var}(Y_i) &= \sum_{j=1}^{n_i} \pi_i(1 - \pi_i) + \sum_{j=1}^{n_i} \sum_{k=1, k \neq j}^{n_i} \delta\pi_i(1 - \pi_i) \\ &= n_i\pi_i(1 - \pi_i) + n_i(n_i - 1)\delta\pi_i(1 - \pi_i) \\ &= \sigma_i^2 n_i\pi_i(1 - \pi_i), \end{aligned}$$

em que $\sigma_i^2 = 1 + (n_i - 1)\delta$. Se é exigido que $\sigma_i^2 > 0$, então devemos ter

$$1 + (n_i - 1)\delta > 0,$$

que implica em $\delta > -1/(n_i - 1)$. Portanto, haverá a restrição

$$-\frac{1}{n_i - 1} \leq \delta \leq 1.$$

Assim, δ assumirá valores negativos apenas para n_i pequeno. Caso contrário, δ assumirá valores positivos. Logo, teremos em geral $\text{Var}(Y_i) > n_i\pi_i(1 - \pi_i)$ (sobredispersão).

Caso II

Vamos supor agora que p_i representa a probabilidade de sucesso nas respostas do i -ésimo grupo tal que $E(p_i) = \pi_i$ e $\text{Var}(p_i) = \delta\pi_i(1 - \pi_i)$, $\delta \geq 0$. Temos portanto um modelo de efeito aleatório, que reduz ao modelo usual de efeito fixo se tomarmos $\delta = 0$. Assumimos ainda que $Y_{ij}|p_i \sim \text{Be}(p_i)$ de onde segue que $E(Y_{ij}|p_i) = p_i$ e $\text{Var}(Y_{ij}|p_i) = p_i(1 - p_i)$. Daí obtemos

$$E(Y_i) = E\{E(Y_i|p_i)\} = n_i\pi_i$$

e

$$\begin{aligned}
\text{Var}(Y_i) &= \text{E}\{\text{Var}(Y_i|p_i)\} + \text{Var}\{\text{E}(Y_i|p_i)\} \\
&= n_i\pi_i(1-\pi_i)(1-\delta) + n_i^2\delta\pi_i(1-\pi_i) \\
&= n_i\pi_i(1-\pi_i)\{1+(n_i-1)\delta\},
\end{aligned}$$

que coincidem com os resultados obtidos para o primeiro caso. No entanto aqui temos a restrição $\delta \geq 0$.

Estimação

A estimação de δ tem sido discutida em vários contextos. No primeiro caso, por exemplo, δ pode ser consistentemente estimado por

$$\tilde{\delta} = \sum_{i=1}^g \sum_{\ell' < \ell} \hat{r}_{P_{i\ell}} \hat{r}_{P_{i\ell'}} / (N - p), \quad (3.18)$$

em que $\hat{r}_{P_{i\ell}} = (y_{i\ell} - \hat{\pi}_i) / \sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}$ é o resíduo de Pearson estimado, $N = \frac{1}{2} \sum_{i=1}^g n_i(n_i - 1)$ e $\hat{\pi}_i$ é a estimativa de máxima verossimilhança de π_i supondo $\delta = 0$. Podemos, contudo, estimar β e δ simultaneamente através de um processo iterativo. Uma proposta é o uso de equações de estimação generalizadas (Liang e Zeger, 1986) as quais serão discutidas no Capítulo 5. As novas estimativas, denotadas por $\hat{\beta}_G$ e $\hat{\delta}$, saem do sistema de equações

$$\sum_{i=1}^g \{1 + (n_i - 1)\hat{\delta}\}^{-1} \mathbf{x}_i(y_i - n_i\hat{\pi}_i) = \mathbf{0}.$$

Dada uma estimativa inicial para δ , que pode ser $\tilde{\delta}$, temos o seguinte processo iterativo para obter $\hat{\beta}_G$:

$$\beta^{(m+1)} = \beta^{(m)} + \left\{ \sum_{i=1}^g \omega_i^{(m)} \mathbf{x}_i \mathbf{x}_i^T \right\}^{-1} \sum_{i=1}^g \{ \omega_i^{(m)} \mathbf{x}_i (y_i - n_i \pi_i^{(m)}) / n_i \pi_i^{(m)} (1 - \pi_i^{(m)}) \}, \quad (3.19)$$

$m = 0, 1, 2 \dots$, em que $\omega_i = n_i \pi_i (1 - \pi_i) / \{1 + (n_i - 1) \hat{\delta}\}$. O processo iterativo (3.19) é alternado com (3.18) até chegarmos à convergência. Podemos mostrar que o estimador $\hat{\beta}_G$ é consistente e assintoticamente normal. A variância assintótica de $\hat{\beta}_G$ é dada por

$$\text{Var}(\hat{\beta}_G) = \left\{ \sum_{i=1}^g \omega_i \mathbf{x}_i \mathbf{x}_i^T \right\}^{-1}.$$

Há também uma proposta de variância assintótica robusta no caso da estrutura de correlação ter sido definida incorretamente, que é dada por

$$\text{Var}(\hat{\beta}_G) = \left\{ \sum_{i=1}^g \omega_i \mathbf{x}_i \mathbf{x}_i^T \right\}^{-1} \left\{ \sum_{i=1}^g \nu_i \mathbf{x}_i \mathbf{x}_i^T \right\} \left\{ \sum_{i=1}^g \omega_i \mathbf{x}_i \mathbf{x}_i^T \right\}^{-1},$$

em que $\nu_i = \{1 + (n_i - 1) \hat{\delta}\}^{-2} \sum_{\ell, \ell'} (y_{i\ell} - \pi_i)(y_{i\ell'} - \pi_i)$. Um desvio corrigido para esse modelo fica dado por $D(y; \hat{\mu}_G) = \sum_{i=1}^g \{1 + (n_i - 1) \hat{\delta}\} d_i^2$, em que d_i^2 é o i -ésimo componente do desvio de um modelo binomial, avaliado em $\hat{\beta}_G$.

Apresentamos a seguir os procedimentos para resolvermos (3.19) no R. Inicialmente iremos propor uma função `corpearson` para obtermos (3.18). Denotaremos os vetores $(y_1/n_1, \dots, y_g/n_g)^T$, $(y_1, \dots, y_g)^T$ e $(n_1, \dots, n_g)^T$ por `fr`, `yt` e `nt`, respectivamente, e o número de parâmetros por `npar`. A função é definida por

```
corpearson = function(fr, yt, nt, npar) {
  nt1 = 0.5*sum(nt*(nt-1))
  sum1 = (0.5*yt*(yt-1) - fr*(nt-1)*yt +
  0.5*fr*fr*nt*(nt-1))/(fr*(1-fr))
  sum1 = sum(sum1)
  rho = sum1/(nt1-npar)
  rho }
```

Vamos supor que temos duas variáveis explicativas representadas por `x1` e `x2` sem intercepto e que os resultados do ajuste do modelo supondo inde-

pendência sejam colocados em `fit.model`. Em `fit.gee` são armazenados os resultados do processo iterativo dado em (3.19) e vamos supor 10 iterações.

Seguem os comandos

```
fit.model = glm(resp ~ x1 + x2 - 1, family=binomial)
eta = predict(fit.model)
fr = fitted(fit.model)
rr = corpearson(fr, yt, nt, npar)
i = 1
while(i <= 10) {
fit.gee = glm(resp ~ x1 + x2 -1, family=binomial, start=
mu = exp(eta)/(1 + exp(eta)),
maxiter = 1,
weights = 1/(1 + (nt - 1)*rr))
eta = predict(fit.gee)
fr = fitted(fit.gee)
rr = corpearson(fr, yt, nt, npar)
i = i + 1 }.
```

A estimativa final da correlação está armazenada em `rr`. Para rodarmos os programas descritos acima no R colocamos inicialmente a função `corpearson` num arquivo externo, por exemplo denominado `corr.s`, e executamos o mesmo através do comando abaixo

```
source('corr.s').
```

Então a função `corpearson` estará instalada. Em seguida devemos fazer o mesmo para ajustarmos o modelo colocando os demais comandos num arquivo externo, por exemplo denominado `super.s`, fazendo o seguinte:

```
source('super.s').
```

Teste de ausência de sobredispersão

Podemos ter interesse particular em testar a hipótese de ausência de sobredispersão $H_0 : \delta = 0$ contra $H_1 : \delta > 0$. Como o conhecimento da distribuição de Y_{ij} é bastante complexo sob a hipótese alternativa, o que inviabilizaria a aplicação de testes tradicionais tais como razão de verossimilhanças, Wald e escore, propomos a aplicação de um teste tipo escore que requer apenas o conhecimento dos dois primeiros momentos de Y_{ij} e a estatística do teste é avaliada sob a hipótese nula (modelo binomial de respostas independentes). A estatística do teste (ver Paula e Artes, 2000) assume a forma

$$\xi_S = \frac{\sum_{i=1}^g \hat{M}_i}{\sqrt{\sum_{i=1}^g \hat{M}_i^2}},$$

em que $\hat{M}_i = \sum_{\ell < \ell'} \hat{r}_{P_{i\ell}} \hat{r}_{P_{i\ell'}}$ de modo que H_0 seja rejeitada quando $\xi_S > z_{(1-\alpha)}$. Podemos mostrar que essa estatística corresponde à forma padronizada (sob H_0) de $\tilde{\delta}$. Para calcularmos ξ_S propomos a função abaixo em que `fr` denota os valores ajustados sob a hipótese nula.

```
escore = function(fr,yt,nt) {  
  sum1 = (0.5*yt*(yt-1) - fr*(nt-1)*yt +  
  0.5*fr*fr*nt*(nt-1))/(fr*(1-fr))  
  sum2 = sum(sum1*sum1)  
  sum1 = sum(sum1)  
  escore = sum1/sqrt(sum2)  
  escore }.
```

Quase-Verossimilhança

Uma outra possibilidade de estudarmos o fenômeno de sobredispersão é através do uso do modelo beta-binomial em que $Y_i | v \sim B(n_i, v)$ enquanto v

segue uma distribuição beta. Podemos mostrar que a variância de Y_i é dada por

$$\text{Var}(Y_i) = n_i \pi_i (1 - \pi_i) \{1 + (n_i - 1) \delta\},$$

em que π_i e δ dependem dos parâmetros da distribuição beta. A estimação de δ é bastante complexa nesse caso requerendo o uso de métodos iterativos e de integração numérica (ver, por exemplo, Collett, 1991, Cap. 6).

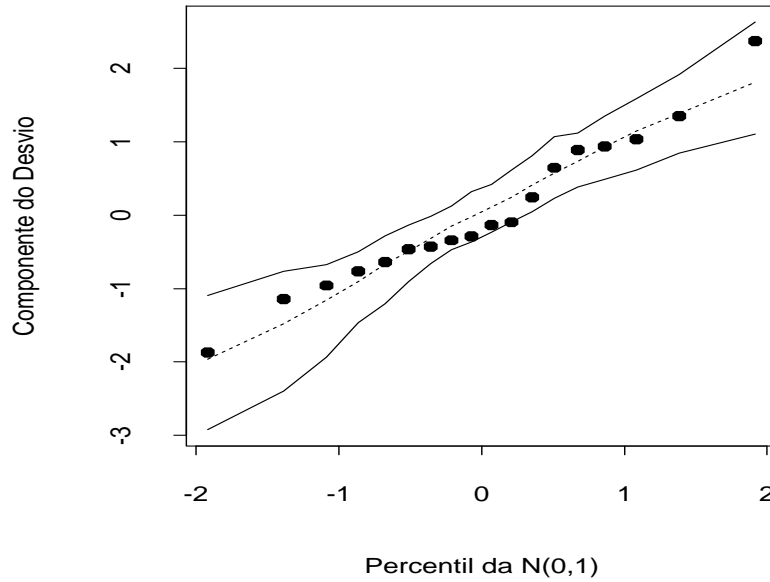


Figura 3.11: Gráfico normal de probabilidades para o resíduo corrigido para o modelo logístico de retas paralelas ajustado aos dados sobre três tipos de inseticida.

Podemos ainda supor $\sigma_i^2 = \phi^{-1}$, estimarmos ϕ consistentemente dos dados ou do modelo ajustado e substituírmos a estimativa obtida nas quantidades que envolvem ϕ . Quando n_i é grande, $\forall i$, podemos estimar ϕ diretamente do desvio

$$\hat{\phi}^{-1} = \frac{D(\mathbf{y}; \hat{\boldsymbol{\mu}})}{g - p}.$$

No caso de n_i pequeno, para algum i , recomendamos a estimativa abaixo

$$\hat{\phi}^{-1} = \frac{1}{g-p} \sum_{i=1}^g \frac{(y_i - n_i \hat{\pi}_i)^2}{n_i \hat{\pi}_i (1 - \hat{\pi}_i)},$$

em que p denota o número de parâmetros da parte sistemática do modelo e $\hat{\pi}_1, \dots, \hat{\pi}_g$ são as probabilidades ajustadas nos g grupos. Sob a hipótese de que o modelo é verdadeiro, essa estimativa é também consistente para ϕ . Essa opção é um caso particular de modelos de quase-verossimilhança que serão discutidos no Capítulo 5.

No exemplo da seção anterior, envolvendo a comparação de três inseticidas, temos um total de 18 grupos com probabilidades ajustadas $\hat{\pi}_i(x_j)$, $i = 1, 2, 3$ e $j = 1, \dots, 6$. Como $n_i = 50$ para a maioria dos grupos e próximo a esse valor para os demais grupos, podemos estimar ϕ consistentemente através de

$$\hat{\phi}^{-1} = \frac{D(\mathbf{y}; \hat{\boldsymbol{\mu}})}{g-p} = \frac{21,282}{14} = 1,52.$$

Algumas quantidades que envolvem ϕ deverão ser corrigidas,

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \hat{\phi}^{-1}(\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1},$$

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \hat{\phi} D(\mathbf{y}; \hat{\boldsymbol{\mu}}) \text{ e}$$

$t_{D_i}^* = \sqrt{\hat{\phi}} t_{D_i}$. O novo gráfico normal de probabilidades, agora com $t_{D_i}^*$, é apresentado na Figura 3.11 e não apresenta indícios de afastamentos sérios das suposições feitas para o modelo. É importante observarmos que o novo resíduo $t_{D_i}^*$ não corresponde ao componente do desvio de nenhum modelo particular. Nos modelos de quase-verossimilhança a distribuição da resposta é em geral desconhecida e o uso de $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ deve ser encarado de forma descritiva.

Aplicação

Collett (1991, Seção 6.9) descreve um experimento com duas espécies de *rotifers*, um tipo microscópico de invertebrado aquático. O objetivo do experimento é determinar a densidade relativa para cada uma das espécies. Foi utilizado um método indireto que consiste em centrifugar os animais em recipientes com densidades relativas de uma determinada substância e então utilizar uma regressão logística para ajustar a proporção de *rotifers* que permanece suspensa segundo a densidade relativa. A densidade relativa de cada espécie pode ser estimada pela DL_{50} , que nesse caso representa a densidade relativa da substância que deixa suspenso 50% de *rotifers*.

Seja Y_{ij} o número de animais da i -ésima espécie que permanecem suspensos num recipiente com densidade relativa d_j da solução, onde foram colocados n_{ij} *rotifers*. Assumimos inicialmente que $Y_{ij} \sim B(n_{ij}, \pi_{ij})$, $i = 1, 2$ e $j = 1, \dots, 20$, em que

$$\log \left\{ \frac{\pi_{ij}}{1 - \pi_{ij}} \right\} = \alpha_i + \beta_i d_j.$$

Na Tabela 3.21 e no arquivo **rotifers.dat** são apresentados para cada espécie a densidade relativa da substância, o número de *rotifers* expostos e o número de *rotifers* em suspensão. Para a espécie *Polyathra* as estimativas de máxima verossimilhança são dadas por $\hat{\alpha}_1 = -109,72(5,22)$ e $\hat{\beta}_1 = 105,67(5,02)$, enquanto que para a espécie *Keratella* obtemos $\hat{\alpha}_2 = -114,35(4,03)$ e $\hat{\beta}_2 = 108,75(3,86)$. Embora essas estimativas sejam altamente significativas, o desvio do modelo $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 434,02$ (36 graus de liberdade) indica para um ajuste inadequado. O gráfico normal de probabilidades descrito na Figura 3.12 confirma a sobredispersão. Segundo Collett (1991, Cap. 6) a sobredispersão nos dados pode ter sido causada por uma possível má distribuição dos animais nos recipientes, uma vez que *roti-*

fers mais jovens são menos densos do que os mais maduros. Collett (1991) propõe um modelo logístico com efeito aleatório para ajustar a proporção de animais em suspensão e consegue uma redução substancial no valor do desvio. Vamos assumir, alternativamente, o modelo proposto na Seção 3.6.14, que com uma adaptação de notação corresponde a assumirmos $E(Y_{ij}) = n_{ij}\pi_{ij}$ e $\text{Var}(Y_{ij}) = n_{ij}\pi_{ij}(1 - \pi_{ij})\{1 + (n_{ij} - 1)\delta\}$, em que δ denota a correlação intraunidade experimental.

Tabela 3.21

Distribuição de rotifers das duas espécies.

Densidade	Polyarthra major		Keratella cochlearis	
	Suspensos	Expostos	Suspensos	Expostos
1,019	11	58	13	161
1,020	7	86	14	248
1,021	10	76	30	234
1,030	19	83	10	283
1,030	9	56	14	129
1,030	21	73	35	161
1,031	13	29	26	167
1,040	34	44	32	286
1,040	10	31	22	117
1,041	36	56	23	162
1,048	20	27	7	42
1,049	54	59	22	48
1,050	20	22	9	49
1,050	9	14	34	160
1,060	14	17	71	74
1,061	10	22	25	45
1,063	64	66	94	101
1,070	68	86	63	68
1,070	488	492	178	190
1,070	88	89	154	154

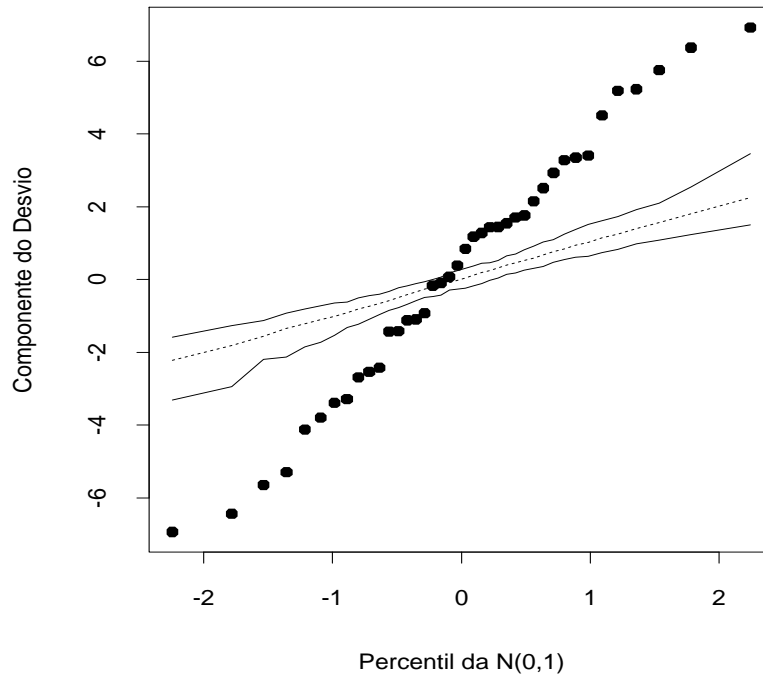


Figura 3.12: Gráfico normal de probabilidades do modelo logístico ajustado aos dados sobre *rotifers*.

Usando o processo iterativo dado na seção anterior obtemos as novas estimativas $\hat{\alpha}_1 = -90,64(13,18)$, $\hat{\beta}_1 = 87,22(12,66)$, $\hat{\alpha}_2 = -117,25(14,91)$, $\hat{\beta}_2 = 111,45(14,21)$ e $\hat{\delta} = 0,0815$. Pela Figura 3.13 notamos que exceto a observação #16, que corresponde a uma unidade experimental com baixa proporção de *rotifers*, 10/22, para uma densidade alta, os demais resíduos permanecem no intervalo $[-2,2]$ e não apresentam nenhuma tendência sistemática contra os valores ajustados. A aplicação da estatística ξ_S para testar $H_0 : \delta = 0$ contra $H_1 : \delta > 0$ forneceu o valor $\xi_S = 3,126$, com nível descritivo $P = 0,0009$, indicando fortemente pela rejeição da hipótese nula. Portanto, há indícios de sobredispersão nos dados.

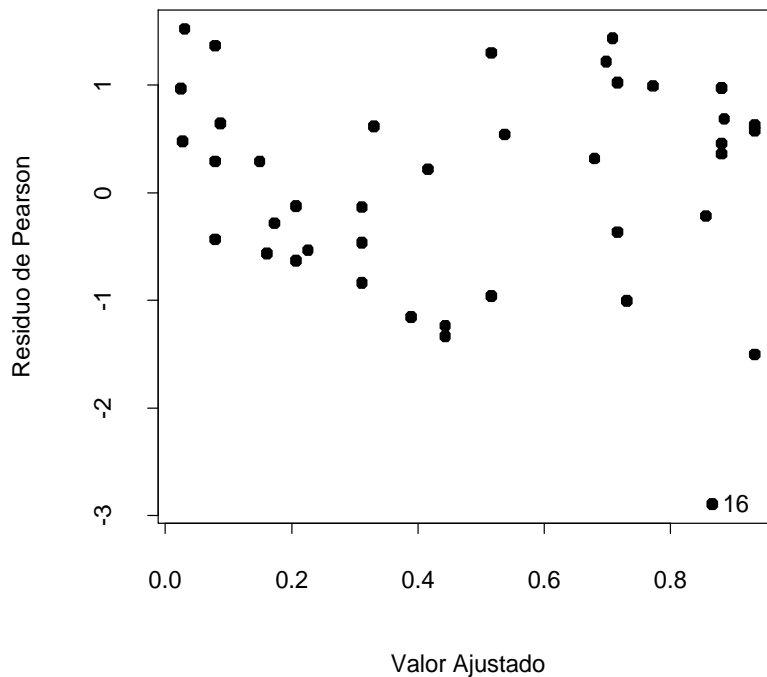


Figura 3.13: Gráfico de resíduos de Pearson contra os valores ajustados para o modelo logístico de sobredispersão ajustado aos dados sobre *rotifers*.

3.6.15 Modelo logístico condicional

Em alguns estudos de caso e controle ou de seguimento o número de estratos formados pode ser relativamente grande. Isso ocorre em particular nos estudos emparelhados de caso e controle, em que a influência de fatores suspeitos de confundimento é controlada através de emparelhamentos de casos com controles, segundo alguns níveis desses fatores. Para cada emparelhamento temos um estrato. Assim, se é adotado um modelo logístico linear, além dos parâmetros correspondentes aos efeitos incluídos no modelo, temos um parâmetro (intercepto) para cada estrato. Nos casos de estratos com poucas observações, o número de parâmetros pode ser da mesma ordem do número total de observações, o que em geral leva a estimativas viesadas (ver Cox e

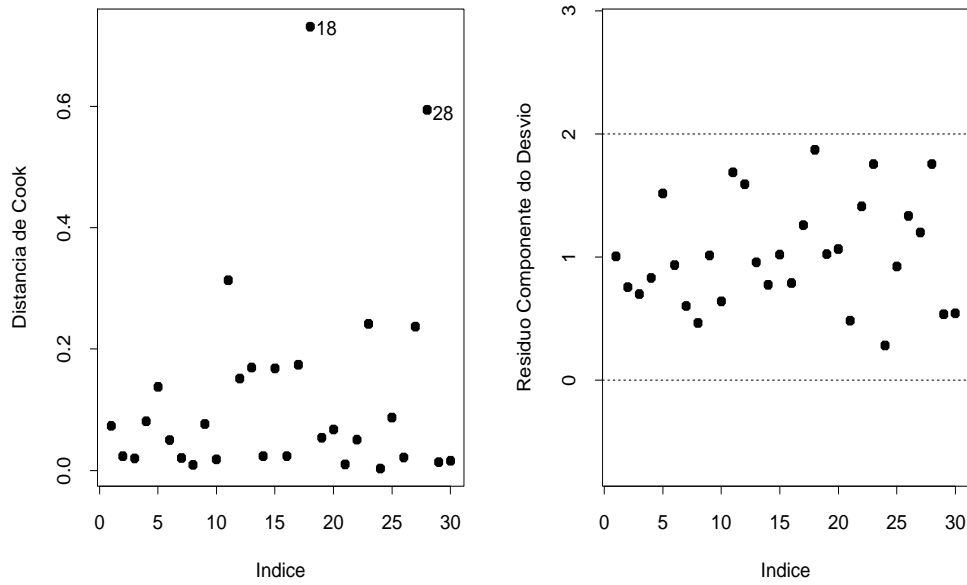


Figura 3.14: Gráficos de diagnóstico para o modelo logístico condicional aplicado aos dados sobre diabetes.

Hinkley, 1974, p. 292).

Como ilustração, vamos supor um estudo de caso e controle com k emparelhamentos do tipo 1:1 (1 caso por 1 controle) segundo os níveis de um fator binário de exposição representado pela variável X ($X = 1$ presença da exposição, $X = 0$ ausência da exposição). Denotamos por $Y_i(x)$ o resultado da resposta para o indivíduo do i -ésimo estrato com $X = x$ ($Y_i(x) = 1$ caso, $Y_i(x) = 0$ controle). Vamos supor que $Y_i(x) \sim \text{Be}\{\pi_i(x)\}$, em que

$$\log \left\{ \frac{\pi_i(x)}{1 - \pi_i(x)} \right\} = \alpha_i + \beta x.$$

A razão de chances de ser caso entre o indivíduo exposto e o indivíduo não exposto no i -ésimo estrato fica dada por

$$\psi = \frac{\pi_i(1)/\{1 - \pi_i(1)\}}{\pi_i(0)/\{1 - \pi_i(0)\}} = \exp(\beta)$$

sendo, portanto, constante ao longo dos estratos.

Para eliminarmos os parâmetros α_i 's podemos trabalhar com a distribuição condicional de $Y_i(1)$ dado $Y_i(1) + Y_i(0) = m$. Essa distribuição foi discutida na Seção 3.2.3. A função de probabilidades pode ser expressa na forma

$$f(a|m; \psi) = \frac{\binom{1}{a} \binom{1}{m-a} \psi^a}{\sum_{t=u}^v \binom{1}{t} \binom{1}{m-t} \psi^t},$$

em que $a = 0, 1$ e $m = 0, 1, 2$. É fácil mostrar que $f(a|0; \psi) = f(a|2; \psi) = 1$, havendo portanto informação a respeito de ψ somente nos estratos em que $Y_i(1) + Y_i(0) = 1$. A função de probabilidades nesse caso é definida para $a = 0$ e $a = 1$, sendo as probabilidades dadas por

$$f(0|1; \psi) = 1/(1 + \psi)$$

e

$$f(1|1; \psi) = \psi/(1 + \psi).$$

Se definirmos para o i -ésimo estrato duas novas variáveis binárias X_{1i} e X_{2i} representando, respectivamente, o nível de exposição do caso e do controle, poderemos expressar as probabilidades condicionais na forma

$$f(a|1, \psi) = \frac{\exp(x_{1i} - x_{2i})\beta}{1 + \exp(x_{1i} - x_{2i})\beta},$$

em que $a = 0, 1$. Assim, para k estratos, a função de verossimilhança conjunta condicional, que depende apenas de β e será denotada por $\ell(\beta)$, assume a forma

$$\ell(\beta) = \prod_{i=1}^k \left[\frac{\exp\{(x_{i1} - x_{i2})\beta\}}{1 + \exp\{(x_{i1} - x_{i2})\beta\}} \right].$$

Temos que a expressão acima coincide com a função de verossimilhança de uma regressão logística com k sucessos em k ensaios, com uma única co-variável com valores observados $z_i = x_{i1} - x_{i2}$, $i = 1, \dots, k$, e passando pela origem.

Generalizando para p covariáveis e supondo ainda emparelhamentos 1:1, teremos o modelo

$$\log \left\{ \frac{\pi_i(\mathbf{x})}{1 - \pi_i(\mathbf{x})} \right\} = \alpha_i + \mathbf{x}^T \boldsymbol{\beta},$$

em que $\mathbf{x} = (x_1, \dots, x_p)^T$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ e $\pi_i(\mathbf{x}) = Pr\{Y_i = 1|\mathbf{x}\}$, $i = 1, \dots, k$. Se observamos no i -ésimo estrato os valores $\mathbf{x}_{i1} = (x_{i11}, \dots, x_{i1p})^T$ para o caso e os valores $\mathbf{x}_{i2} = (x_{i21}, \dots, x_{i2p})^T$ para o controle, a função de verossimilhança conjunta condicional assume a forma geral (ver, por exemplo, Breslow e Day, 1980, p. 205; Hosmer e Lemeshow, 1989, Cap. 7)

$$\ell(\boldsymbol{\beta}) = \prod_{i=1}^k \left[\frac{\exp\{(\mathbf{x}_{i1} - \mathbf{x}_{i2})^T \boldsymbol{\beta}\}}{1 + \exp\{(\mathbf{x}_{i1} - \mathbf{x}_{i2})^T \boldsymbol{\beta}\}} \right].$$

Logo, a estimação de $\boldsymbol{\beta}$ pode ser feita através do ajuste de uma regressão logística com k sucessos em k ensaios, com valores observados das covariáveis dados por $z_{ij} = x_{i1j} - x_{i2j}$, $i = 1, \dots, k$ e $j = 1, \dots, p$ e passando pela origem. Devemos observar que embora algumas quantidades da regressão logística condicional para estudos emparelhados do tipo 1:1 coincidam com as quantidades de uma regressão logística não condicional passando pela origem, tais como estimativas dos parâmetros e erros padrão assintóticos, as distribuições dos modelos são diferentes. No primeiro caso temos o produto de hipergeométricas independentes enquanto que no segundo caso temos o produto de binomiais independentes. Isso pode refletir na obtenção de alguns resultados, como por exemplo, geração de envelope para o resíduo componente do desvio que usa a distribuição da resposta no processo de geração dos dados.

Técnicas de diagnóstico

Moolgavkar, Lustbader e Venzon (1985) e Pregibon (1984) têm mostrado que a maioria das técnicas usuais de diagnóstico do modelo logístico não condicional podem ser estendidas para o modelo logístico condicional. Como

a variável resposta no modelo logístico condicional sempre assume o valor 1, o resíduo componente do desvio é sempre positivo, sendo dado por

$$t_{D_i} = \frac{\sqrt{2}|\log \hat{\pi}_i|}{\sqrt{1 - \hat{h}_{ii}}},$$

em que

$$\hat{\pi}_i = \frac{\exp(\mathbf{z}_i^T \hat{\boldsymbol{\beta}})}{1 + \exp(\mathbf{z}_i^T \hat{\boldsymbol{\beta}})} \quad \text{e} \quad \hat{h}_{ii} = \hat{\pi}_i(1 - \hat{\pi}_i) \mathbf{z}_i^T (\mathbf{Z}^T \hat{\mathbf{V}} \mathbf{Z})^{-1} \mathbf{z}_i.$$

Os gráficos de t_{D_i} e \hat{h}_{ii} contra os valores ajustados $\hat{\pi}_i$ podem revelar emparelhamentos discrepantes com algum tipo de influência nos resultados do modelo.

De forma similar, a distância de Cook no caso emparelhado fica dada por

$$LD_i = \frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})^2} \hat{r}_{P_i}^2,$$

em que

$$\hat{r}_{P_i} = \frac{1 - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}}$$

é o resíduo de Pearson. Temos que \hat{r}_{P_i} assume sempre valores não negativos. O gráfico de LD_i contra os valores ajustados $\hat{\pi}_i$ pode revelar aqueles emparelhamentos com maior influência nas estimativas dos parâmetros. A geração de envelope, contudo, somente pode ser feita através do modelo logístico condicional.

Para ilustrarmos o ajuste no R, vamos supor um estudo com $k = 20$ emparelhamentos do tipo 1:1 e que foram observados os valores de duas covariáveis $V1$ e $V2$. Os valores observados dos casos serão armazenados nos objetos `v11` e `v12` e os valores observados dos controles nos objetos `v21` e `v22`. O ajuste segue os seguintes passos:

```
resp < rep(1, times=20)
```

```
z1 < v11 - v21
```

```
z2 < v12 - v22
```

```
fit.cond < glm(resp ~ z1+z2 - 1, family=binomial).
```

Podemos analisar `fit.cond` em geral da mesma forma que analisamos a saída de um modelo logístico linear. Por exemplo, as estimativas e os erros padrão, como foi mostrado acima, coincidem com as estimativas e os erros padrão obtidos pelo modelo logístico condicional.

Aplicação

Como aplicação, discutimos a seguir um estudo cujo objetivo foi avaliar o efeito da obesidade, do histórico familiar e de atividades físicas no desenvolvimento de diabetes não dependente de insulina. 30 indivíduos não diabéticos foram emparelhados com 30 indivíduos diabéticos não dependentes de insulina pela idade e pelo sexo. A obesidade foi medida através do índice de massa corporal (IMC), que é definida como sendo o peso (em kg) dividido pela altura (em metros quadrados). O histórico familiar com diabetes (HF) e as atividades físicas (ATF) foram tratadas como sendo variáveis binárias (HF=1 presença, HF=0 ausência; ATF=1 presença, ATF=0 ausência). Os dados são descritos em Lee (1991, p. 312) e reproduzidos na Tabela 3.22 e estão também no arquivo **diabetes.dat**. Denotaremos por x_{i11} , x_{i12} e x_{i13} , respectivamente, o valor da massa corporal (IMC), histórico familiar (HF) e atividades físicas (ATF) para o i -ésimo indivíduo diabético e por x_{i21} , x_{i22} e x_{i23} os valores dessas variáveis para o i -ésimo indivíduo não diabético. A função de verossimilhança do modelo logístico condicional será dada por

$$\ell(\boldsymbol{\beta}) = \prod_{i=1}^{30} \left\{ \frac{\exp(z_{i1}\beta_1 + z_{i2}\beta_2 + z_{i3}\beta_3)}{1 + \exp(z_{i1}\beta_1 + z_{i2}\beta_2 + z_{i3}\beta_3)} \right\},$$

em que $z_{i1} = x_{i11} - x_{i21}$, $z_{i2} = x_{i12} - x_{i22}$ e $z_{i3} = x_{i13} - x_{i23}$.

Tabela 3.22
Emparelhamento de 30 diabéticos não dependentes de insulina (casos) e 30 não diabéticos (controles).

Par	Casos			Controles		
	IMC	HF	ATF	IMC	HF	ATF
1	22,1	1	1	26,7	0	1
2	31,3	0	0	24,4	0	1
3	33,8	1	0	29,4	0	0
4	33,7	1	1	26,0	0	0
5	23,1	1	1	24,2	1	0
6	26,8	1	0	29,7	0	0
7	32,3	1	0	30,2	0	1
8	31,4	1	0	23,4	0	1
9	37,6	1	0	42,4	0	0
10	32,4	1	0	25,8	0	0
11	29,1	0	1	39,8	0	1
12	28,6	0	1	31,6	0	0
13	35,9	0	0	21,8	1	1
14	30,4	0	0	24,2	0	1
15	39,8	0	0	27,8	1	1
16	43,3	1	0	37,5	1	1
17	32,5	0	0	27,9	1	1
18	28,7	0	1	25,3	1	0
19	30,3	0	0	31,3	0	1
20	32,5	1	0	34,5	1	1
21	32,5	1	0	25,4	0	1
22	21,6	1	1	27,0	1	1
23	24,4	0	1	31,1	0	0
24	46,7	1	0	27,3	0	1
25	28,6	1	1	24,0	0	0
26	29,7	0	0	33,5	0	0
27	29,6	0	1	20,7	0	0
28	22,8	0	0	29,2	1	1
29	34,8	1	0	30,0	0	1
30	37,3	1	0	26,5	0	0

As estimativas de máxima verossimilhança (erro padrão aproximado) são dadas por $\hat{\beta}_1 = 0,090(0,065)$, $\hat{\beta}_2 = 0,968(0,588)$ e $\hat{\beta}_3 = -0,563(0,541)$, cujos níveis descritivos são, respectivamente, dados por 0,166, 0,099 e 0,298, indicando indícios de efeito significativo apenas para o histórico familiar.

Na Figura 3.14 são apresentados alguns gráficos de diagnóstico em que podemos notar a influência das observações #18 e #28 como possivelmente influentes nas estimativas dos parâmetros. A eliminação do emparelhamento #18 não muda os resultados inferenciais embora aumente a significância do histórico familiar. Já a eliminação do emparelhamento #28 muda os resultados inferenciais uma vez que o índice de massa corporal passa a ser significativo ao nível de 10%. Nesse emparelhamento o controle tem histórico familiar e atividade física enquanto o caso não apresenta as duas características. Além disso, o caso tem um índice de massa corporal menor do que o controle.

Emparelhamento 1:M

Para emparelhamentos do tipo 1:M ($M \geq 2$) e k estratos a função de verossimilhança (ver, por exemplo, Breslow e Day, 1980; Cordeiro e Paula, 1989b) para $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ fica dada por

$$\ell(\boldsymbol{\beta}) = \Pi_{i=1}^k \{ \exp(\mathbf{x}_{i0}^T \boldsymbol{\beta}) / \sum_{\ell=0}^M \exp(\mathbf{x}_{i\ell}^T \boldsymbol{\beta}) \}, \quad (3.20)$$

cujo logaritmo assume a forma

$$L(\boldsymbol{\beta}) = \log \ell(\boldsymbol{\beta}) = \sum_{i=1}^k [\mathbf{x}_{i0}^T \boldsymbol{\beta} - \log \{ \sum_{\ell=0}^M \exp(\mathbf{x}_{i\ell}^T \boldsymbol{\beta}) \}], \quad (3.21)$$

em que $\mathbf{x}_{i0} = (x_{i01}, \dots, x_{i0p})^T$ denota os valores observados para o caso e $\mathbf{x}_{i\ell} = (x_{i\ell 1}, \dots, x_{i\ell p})^T$ denota os valores observados para o ℓ -ésimo controle.

A função de verossimilhança (3.21) coincide com a função de verossimilhança do modelo de regressão de Cox (Cox, 1972; Cox e Oakes, 1974) quando

não há ocorrência de empates. Isso permite que os modelos logísticos condicionais para emparelhamentos 1:M ($M \geq 2$) sejam ajustados através de programas desenvolvidos para o modelo de Cox.

3.7 Exercícios

1. Os dados abaixo são de um estudo de seguimento cujo objetivo foi avaliar a associação de duas técnicas cirúrgicas, A e B , e a ocorrência de problemas graves pós-operatórios segundo duas faixas de idade.

Problema	Faixa I		Faixa II	
	A	B	A	B
Sim	6	7	7	4
Não	14	23	9	12

Obtenha um intervalo assintótico de confiança de 95% para a razão de chances em cada estrato. Teste a hipótese de homogeneidade das razões de chances. Comente.

2. A tabela abaixo resume um estudo de caso e controle em que foram considerados como casos 200 homens adultos diagnosticados com câncer de esôfago num hospital de uma determinada comunidade. Os controles foram uma amostra de 775 homens adultos escolhidos aleatoriamente da lista de eleitores da comunidade. Esses dois grupos foram classificados segundo os níveis alto (mais de 80g/dia) e baixo (até 80g/dia) do fator Exposição ao Alcool.

	Alto	Baixo	Total
Caso	96	104	200
Controle	109	666	775
Total	205	770	975

Verifique, através de um teste apropriado, se há associação entre o fator de exposição e a doença. Encontre um intervalo de confiança assintótico para a razão de chances. Indique as suposições utilizadas e interprete os resultados.

3. (Day e Byar, 1979). Suponha $Y_{ij} \sim B(n_{ij}, \pi_{ij})$ mutuamente independentes, $i, j = 1, 2$ com as probabilidades π_{ij} sendo definidas por

$$\log \left\{ \frac{\pi_{i1}}{1 - \pi_{i1}} \right\} = \alpha_i - \Delta \quad \text{e} \quad \log \left\{ \frac{\pi_{i2}}{1 - \pi_{i2}} \right\} = \alpha_i + \Delta.$$

Interprete α_1 , α_2 e Δ . Mostre que o teste de escore para testar $H_0 : \Delta = 0$ contra $H_1 : \Delta \neq 0$, coincide com o teste de Mantel-Hanszel (X_{MH}^2) para testar $H_0 : \psi = 1$ contra $H_1 : \psi \neq 1$, em que $\psi = \pi_{i2}(1 - \pi_{i1})/\pi_{i1}(1 - \pi_{i2})$, $i = 1, 2$.

4. Supor um modelo binomial quadrático de dose-resposta, em que $Y_i \stackrel{\text{ind}}{\sim} B(n_i, \pi_i)$, com $g(\pi_i) = \alpha + \beta x_i + \gamma x_i^2$, $i = 1, \dots, k$. Como fica expressa a estimativa \widehat{DL}_{100p} ? E a variância assintótica de \widehat{DL}_{100p} ?
5. Suponha o modelo logístico com ligação de Aranda-Ordaz. Desenvolva um processo iterativo para estimar $(\beta^T, \alpha)^T$ e escreva um programa em R. Aplique esse processo iterativo para ajustar os dados do exemplo sobre a exposição de besouros descrito na Seção 3.6.11. Assuma $\eta = \beta_1 + \beta_2 x$. É α significativamente diferente de um? Com fica o resíduo componente do desvio? E o desvio? Houve melhora na qualidade do ajuste? Tente gerar envelope.
6. Supor um modelo binomial quadrático de dose-resposta, em que $Y_i \stackrel{\text{ind}}{\sim} B(n_i, \pi_i)$, com $g(\pi_i) = \eta_i = \alpha + \beta x_i + \gamma x_i^2$, $i = 1, \dots, k$. Assumir que $\partial^2 \eta / \partial x^2 < 0$, $\forall x$. Qual a solução para $\partial \eta / \partial x = 0$? Denotando essa solução por x_0 , interprete e encontre uma estimativa intervalar para x_0 .

7. (Paula, Sevanes e Ogando, 1988). Os conjuntos de dados apresentados nos arquivos **dose1.dat**, **dose2.dat** e **dose3.dat** são provenientes de um experimento de dose-resposta conduzido para avaliar a influência dos extratos vegetais “aquoso frio de folhas”, “aquoso frio de frutos” e de um extrato químico, respectivamente, na morte de um determinado tipo de caramujo. Para cada conjunto, ajuste um modelo logístico linear simples e um modelo complementar log-log linear simples. Para o melhor ajuste (use envelopes como critério), encontre um intervalo assintótico de 95% para a dose letal DL_{50} , construa as bandas de confiança e verifique se há indícios de sobredispersão aplicando um teste apropriado.
8. (Collett, 1991, p.127). Os dados abaixo são provenientes de um experimento desenvolvido para avaliar a germinação de um determinado tipo de semente segundo três condições experimentais: nível da temperatura ($21^{\circ}C$, $42^{\circ}C$ e $62^{\circ}C$); nível da umidade (baixo, médio e alto) e temperatura da germinação ($11^{\circ}C$ e $21^{\circ}C$). A tabela abaixo apresenta o número de sementes que germinaram após cinco dias para cada 100 sementes submetidas a cada condição experimental.

Temperatura da Germinação	Nível da Umidade	Nível da Temperatura		
		$21^{\circ}C$	$42^{\circ}C$	$62^{\circ}C$
$11^{\circ}C$	baixo	98	96	62
$11^{\circ}C$	médio	94	79	3
$11^{\circ}C$	alto	92	41	1
$21^{\circ}C$	baixo	94	93	65
$21^{\circ}C$	médio	94	71	2
$21^{\circ}C$	alto	91	30	1

Assuma um modelo logístico para explicar o número de sementes que germinaram. Aplique o método AIC para selecionar um modelo consi-

derando interações de 1ª ordem. Interprete os resultados. Faça uma análise de resíduos com o modelo selecionado. Esses dados estão descritos no arquivo **sementes.dat**.

9. Mostre que a variância assintótica do estimador de máxima verossimilhança não condicional da razão de chances numa tabela 2×2 é dada por

$$\text{Var}_A(\tilde{\psi}) = \psi^2 \left\{ \frac{1}{n_1\pi_1(1-\pi_1)} + \frac{1}{n_2\pi_2(1-\pi_2)} \right\}.$$

Lembre que: sob condições gerais de regularidade, os estimadores de máxima verossimilhança são assintoticamente normais e não viesados com variância assintótica igual à inversa da matriz de informação de Fisher.

10. A tabela abaixo descreve o resultado de um experimento em que vários pacientes foram submetidos a um de quatro níveis de exposição de um tratamento particular e foi observado, após 12 meses, se o paciente foi curado ou não curado.

Resultado	Nível de Exposição			
	E1	E2	E3	E4
Curado	20	16	12	5
Não-Curado	80	84	48	20

Seja Y_i o número de pacientes curados dentre os n_i submetidos ao nível de exposição E_i . Suponha que $Y_i \sim B(n_i, \pi_i)$, $i = 1, \dots, 4$. Tome o nível E1 como nível de referência e teste a hipótese de homogeneidade das razões de chances contra a alternativa de razões de chances diferentes. Sugestão: use a estatística X_A^2 .

11. Sejam Y_1 e Y_2 variáveis aleatórias independentes tais que $Y_1 \sim B(n_1, \pi_1)$ e $Y_2 \sim B(n_2, \pi_2)$. Seja $RR = \pi_1/\pi_2$ o risco relativo. (i) Expresse a

função de probabilidades conjunta de Y_1 e Y_2 em função de (RR, π_2) , (ii) encontre as estimativas de máxima verossimilhança \widehat{RR} e $\hat{\pi}_2$, (iii) como fica a matriz de informação de Fisher para (RR, π_2) ? e a variância assintótica de \widehat{RR} ? (iv) Desenvolva o teste da Wald para testar $H_0 : RR = 1$ contra $H_1 : RR \neq 1$. Qual a distribuição nula assintótica do teste?

12. Sejam Y_1, \dots, Y_n variáveis aleatórias mutuamente independentes tais que $Y_i \sim B(m, \pi_i)$, em que $\log\{\pi_i/(1 - \pi_i)\} = \alpha$. (i) Encontre a estimativa de máxima verossimilhança de α . (ii) Calcule $\text{Var}(\hat{\alpha})$. (iii) Como fica o teste da razão de verossimilhanças para testar $H_0 : \alpha = 0$ versus $H_1 : \alpha \neq 0$? Qual a distribuição nula assintótica da estatística do teste?

13. Considere o modelo logístico de dose-resposta em que $Y_i \sim B(m, \pi_i)$, $i = 1, \dots, k$, com parte sistemática dada por

$$\log \left\{ \frac{\pi_i}{1 - \pi_i} \right\} = \alpha + \beta x_i,$$

Expresse o logaritmo da função de verossimilhança em função da dose letal $\psi = \text{DL}_{50}$ e de β . Encontre a função escore $U_\psi = \partial L(\psi, \beta) / \partial \psi$. Considere agora as hipóteses $H_0 : \psi = a$ contra $H_1 : \psi \neq a$. Como fica o teste de escore para testar H_0 contra H_1 ? Qual é a distribuição nula assintótica da estatística do teste? Sugestão: para facilitar a notação expresse a variância assintótica de $\hat{\psi}$ em função das quantidades $v_{00} = \text{Var}(\hat{\alpha})$, $v_{11} = \text{Var}(\hat{\beta})$ e $v_{01} = \text{Cov}(\hat{\alpha}, \hat{\beta})$.

14. Supor $Y_{ij} \stackrel{\text{ind}}{\sim} \text{Be}(\mu_i)$, $0 < \mu_i < 1$, para $i = 1, 2$ e $j = 1, 2, \dots, r$, em que $\arcsen(\sqrt{\mu_1}) = \alpha + \Delta$ e $\arcsen(\sqrt{\mu_2}) = \alpha - \Delta$. Como fica a matriz \mathbf{X} ? Obtenha $\hat{\alpha}$ e $\hat{\Delta}$ e as respectivas variâncias assintóticas,

além de $\text{Cov}(\hat{\alpha}, \hat{\Delta})$. Obtenha a estatística do teste de escore para testar $H_0 : \Delta = 0$ contra $H_1 : \Delta \neq 0$. Qual a distribuição nula assintótica da estatística do teste? Use o resultado: $\frac{d}{dx} \arcsen\{u(x)\} = \frac{1}{\sqrt{1-u^2}} \frac{du}{dx}$.

15. (Lawless, 1982, p.389; Efron, 1988). Vamos considerar agora uma aplicação de regressão logística em análise de sobrevivência. Seja $\pi_i(t)$ a probabilidade de um equipamento do tipo i falhar no intervalo $I_t = (t - 1, t]$ dado que o mesmo não falhou até o tempo $t - 1$. Seja Y_{it} o número de falhas no intervalo I_t e seja n_{it} o número de equipamentos que não falharam até o tempo $t - 1$ no i -ésimo grupo. Assumiremos que $Y_{it} \sim B(n_{it}, \pi_i(t))$ e que as falhas são independentes. Ajustar um modelo logístico do tipo

$$\log \left\{ \frac{\pi_i(t)}{1 - \pi_i(t)} \right\} = \alpha_i + \beta_i t + \gamma_i t^2 \quad (3.22)$$

ao seguinte conjunto de dados:

Tempo	Tipo A		Tipo B		Tipo C	
	n_{1t}	y_{1t}	n_{2t}	y_{2t}	n_{3t}	y_{3t}
1	42	4	50	6	48	11
2	38	3	44	11	37	10
3	35	3	32	10	27	12
4	31	5	22	8	15	8
5	26	6	12	6	6	4

Apresente o gráfico com as curvas ajustadas e os valores observados. Tente selecionar um submodelo apropriado. Verifique a adequação do modelo adotado através do gráfico normal de probabilidades com envelope utilizando o resíduo t_{D_i} . Interprete os resultados. Os dados estão descritos no arquivo **equipamentos.dat**.

16. (Hosmer e Lemeshow, 1989, Cap.7). No arquivo **matched.dat** estão os dados de um estudo de caso-controle com emparelhamentos do tipo 1:1, em que os casos foram mulheres com diagnóstico confirmado de tumor benigno na mama e os controles de mulheres sadias diagnosticadas no mesmo hospital e período dos casos. A variável de emparelhamento foi a idade da paciente na época da entrevista **AGMT**. Escolha três variáveis do arquivo mencionado e verifique através de uma regressão logística condicional a associação entre as variáveis escolhidas e o diagnóstico da doença (sim=1, não=0) representado pela variável **FNDX**. Interprete as estimativas dos parâmetros do modelo ajustado. Faça uma análise de diagnóstico e gere envelope. Observeção: caso você escolha alguma variável com observações perdidas, exclua das análises as pacientes correspondentes.
17. Vamos considerar agora uma aplicação de regressão logística em transportes. Seja $\pi_i(t)$ a probabilidade de um caminhão do tipo i ser desativado durante o ano t dado que o mesmo não foi desativado durante o ano $t-1$. Assuma que durante o ano t foram desativados y_{it} caminhões dentre os n_{it} existentes no começo do ano, $i = 1, 2$ e $t = 1, \dots, k$. Suponha que $Y_{it} \sim B(n_{it}, \pi_i(t))$ e que são mutuamente independentes. Considere o modelo

$$\log \left\{ \frac{\pi_1(t)}{1 - \pi_1(t)} \right\} = \gamma_t \quad \text{e} \quad \log \left\{ \frac{\pi_2(t)}{1 - \pi_2(t)} \right\} = \gamma_t + \beta.$$

O que significa testar $H_0 : \beta = 0$? Qual é a matriz X do modelo? Como fica $\text{Var}(\hat{\beta})$? Mostre que a estatística de escore para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$ pode ser expressa na forma

$$\xi_{SR} = \left\{ \sum_{t=1}^k \left(y_{2t} - \frac{y_{1t}n_{2t}}{n_t} \right) \right\}^2 / \sum_{t=1}^k \frac{y_{1t}n_{1t}n_{2t}(n_t - y_t)}{n_t^3},$$

em que $n_t = n_{1t} + n_{2t}$ e $y_t = y_{1t} + y_{2t}$. Qual é a distribuição nula assintótica de ξ_{SR} ?

18. Sejam Y_1, \dots, Y_k variáveis aleatórias independentes tais que a função de probabilidades de Y_i seja dada por

$$f(y_i; \psi_i) = \frac{\binom{1}{y_i} \binom{1}{1-y_i} \psi_i^{y_i}}{\sum_{t=0}^1 \binom{1}{t} \binom{1}{1-t} \psi_i^t},$$

em que $y_i = 0, 1$. Supor a parte sistemática $\log \psi_i = \beta$. (i) Encontre a estimativa de máxima verossimilhança de β ; (ii) encontre a informação de Fisher para β ; (iii) como fica o teste de escore para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$? Qual a distribuição nula assintótica do teste? (iv) Expresse o resíduo $r_i = (y_i - \hat{\mu}_i) / \sqrt{\hat{\text{Var}}(Y_i)}$ em função de y_i e $\hat{\beta}$; (v) Como você faria para gerar valores de Y_i da distribuição dada acima?

19. (Agresti, 1990, p.253). Considere a tabela abaixo em que um grupo de gestantes fumantes foi classificado segundo os fatores: idade (< 30 ou 30 ou $+$), número de cigarros consumidos por dia (< 5 ou 5 ou $+$), tempo de gestação (≤ 260 dias ou > 260 dias) e a situação da criança (sobreviveu ou não sobreviveu).

Idade	No. de cigarros	Duração da Gestação	Sobrevivência	
			Não	Sim
< 30	< 5	≤ 260	50	315
		> 260	24	4012
	5+	≤ 260	9	40
		> 260	6	459
30+	< 5	≤ 260	41	147
		> 260	14	1594
	5+	≤ 260	4	11
		> 260	1	124

Ajustar um modelo logístico linear em que a resposta é a sobrevivência da criança (sim=1, não=0). Considere até interações de 1ª ordem. Verifique a adequação do modelo e interprete os resultados através de estimativas intervalares para as razões de chances. Esses dados estão descritos no arquivo **gestantes.dat**.

20. (Everitt, 1994). Os dados do arquivo **leuce.dat** referem-se a um estudo com 51 pacientes adultos, previamente diagnosticados com um tipo agudo de leucemia, que receberam um tipo de tratamento e foi verificado após um certo período a eficiência ou não do tratamento. Algumas variáveis explicativas pré-tratamento foram também observadas. As variáveis em estudo são as seguintes: (i) idade do paciente na época do diagnóstico (em anos), (ii) mancha diferencial da doença (em %), (iii) infiltração na medula (em %), (iv) células com leucemia na medula (em %), (v) malignidade da doença ($\times 10^3$), (vi) temperatura máxima antes do tratamento ($\times 10^\circ F$), (vii) tratamento (1: satisfatório, 0: não satisfatório), (viii) tempo de sobrevivência após o diagnóstico (em meses) e (ix) situação (1: sobrevivente, 0: não sobrevivente). Considere um modelo logístico linear para explicar a probabilidade de eficiência do tratamento dadas as seis variáveis explicativas. Seleccionar as variáveis explicativas bem como as interações de primeira ordem através do método *stepwise*. Usar $P_E = P_S = 0,20$. Fazer uma análise de diagnóstico com o modelo selecionado e interpretar algumas razões de chances. Calcular a estatística de Hosmer-Lemeshow para avaliar a qualidade do ajuste do modelo selecionado.
21. (Agresti, 1990, pgs. 122-123). Cinquenta e quatro indivíduos considerados idosos são submetidos a um exame psiquiátrico para avaliar a ocorrência ou não de sintoma de caduquice. Acredita-se que o escore

obtido num exame psicológico feito previamente esteja associado com a ocorrência ou não do sintoma. Os dados são apresentados abaixo (**score**: escala no exame psicológico e **resp**: ocorrência (resp=1) ou não ocorrência (resp=0) do sintoma). Esses dados estão descritos no arquivo **caduquice.dat**.

Score	Resp	Score	Resp	Score	Resp	Score	Resp	Score	Resp
9	1	7	1	7	0	17	0	13	0
13	1	5	1	16	0	14	0	13	0
6	1	14	1	9	0	19	0	9	0
8	1	13	0	9	0	9	0	15	0
10	1	16	0	11	0	11	0	10	0
4	1	10	0	13	0	14	0	11	0
14	1	12	0	15	0	10	0	12	0
8	1	11	0	13	0	16	0	4	0
11	1	14	0	10	0	10	0	14	0
7	1	15	0	11	0	16	0	20	0
9	1	18	0	6	0	14	0		

Ajustar um modelo logístico para explicar a probabilidade de ocorrência do sintoma em função do escore. Interpretar os resultados. Calcule a estatística de Hosmer-Lemeshow. Faça uma análise de diagnóstico com o modelo ajustado.

22. (McCullagh e Nelder, 1989, pgs. 128-135). No arquivo **grahani.dat** estão os dados referentes à distribuição de duas espécies de lagarto (grahani e opalinus) segundo quatro fatores: (i) período do dia (manhã, meio-dia, tarde), (ii) comprimento da madeira (curta, comprida), (iii) largura da madeira (estreita, larga) e (iv) local de ocupação (claro, escuro). Suponha que o número de lagartos encontrados da espécie grahani tenha distribuição binomial.

- (i) Proponha um modelo logístico (sem interação) para explicar a proporção de lagartos da espécie *grahani*. Ajuste o modelo e verifique através do teste da razão de verossimilhanças quais efeitos são significativos ao nível de 10%.
 - (ii) Verifique separadamente se cada interação de primeira ordem pode ser incluída no modelo ao nível de 5%. Construa o ANODEV.
 - (iii) Interprete os resultados tentando falar de uma forma não técnica sobre as preferências dos dois tipos de lagarto.
23. (Neter et al., 1996, pgs. 582-584). Em um estudo para investigar a incidência de dengue numa determinada cidade da costa mexicana, um total de 196 indivíduos, escolhidos aleatoriamente em dois setores da cidade, respondeu às seguintes perguntas: (i) **idade**, idade do entrevistado (em anos), (ii) **nivel**, nível sócio-econômico (**nivel**=1, nível alto; **nivel**=2, nível médio; **nivel**=3, nível baixo) e (iii) **setor**, setor da cidade onde mora o entrevistado (**setor**=1, setor 1; **setor**=2, setor 2) e (iv) **caso**, se o entrevistado contraiu (**caso**=1) ou não (**caso**=0) a doença recentemente. Um dos objetivos do estudo é tentar prever ou explicar a probabilidade de um indivíduo contrair a doença dadas as variáveis explicativas **idade**, **nivel** e **setor**. Os dados estão descritos no arquivo **dengue.dat**. Tente selecionar um modelo através da aplicação do método AIC considerando interações de 1ª ordem. Faça uma interpretação do modelo selecionado (através de razões de chances) e faça uma análise de diagnóstico do mesmo. Verifique a qualidade do ajuste através da estatística de Hosmer-Lemeshow.
24. (McCullagh e Nelder, 1989, p.144). No arquivo **olhos.dat** são apresentados dados referentes a 78 famílias com pelo menos seis filhos cada

uma. Na primeira coluna tem-se a classificação dos olhos dos pais segundo a cor (1: ambos claros, 2: ambos castanhos, 3: ambos escuros, 4: claro e castanho, 5: claro e escuro e 6: castanho e escuro), na segunda coluna a classificação dos olhos dos avós segundo a cor (1: todos claros, 2: todos castanhos, 3: todos escuros, 4: três claros e um castanho, 5: três claros e um escuro, 6: um claro e três castanhos, 7: um escuro e três castanhos, 8: um claro e três escuros, 9: um castanho e três escuros, 10: dois claros e dois castanhos, 11: dois claros e dois escuros, 12: dois castanhos e dois escuros, 13: dois claros, um castanho e um escuro, 14: um claro, dois castanhos e um escuro e 15: um claro, um castanho e dois escuros), na terceira coluna tem-se o número de filhos na família e na última coluna o número de filhos com olhos claros. Seja Y_i o número de filhos com olhos claros pertencentes à i -ésima família. Assuma inicialmente que $Y_i \sim B(n_i, \pi_i)$, $i = 1, \dots, 78$. Resolver os itens abaixo.

- (i) Ajustar inicialmente um modelo logístico linear apenas com o fator ‘cor dos olhos dos pais’. Construir gráficos de resíduos. Identificar os pontos aberrantes. Quais as mudanças nos resultados com a eliminação desses pontos. Há indícios de sobredispersão? Ajustar um modelo de quase-verossimilhança com e sem os pontos aberrantes. Comente.
- (ii) Incluir agora o fator ‘cor dos olhos dos avós’. Refazer todos os passos acima. Comente os resultados.

25. (Morgan, 1992, p.90). A tabela abaixo descreve os resultados de um experimento em que a toxicidade de três concentrações (R-rotenine, D-deguelin e M-mistura, essa última como uma mistura das duas pri-

meiras) é investigada. As concentrações foram testadas em insetos e observado para cada dose o número de insetos mortos. Os dados estão descritos no arquivo **morgan.dat**.

Concentração	Dose	Expostos	Mortos
R	0,41	50	6
R	0,58	48	16
R	0,71	46	24
R	0,89	49	42
R	1,01	50	44
D	0,71	49	16
D	1,00	48	18
D	1,31	48	34
D	1,48	49	47
D	1,61	50	47
D	1,70	48	48
M	0,40	47	7
M	0,71	46	22
M	1,00	46	27
M	1,18	48	38
M	1,31	46	43
M	1,40	50	48

Suponha inicialmente o modelo $\log\{\pi_i(x)/(1 - \pi_i(x))\} = \alpha_i + \beta_i x$, $i = 1, 2, 3$, em que $\pi_i(x)$ é a proporção esperada de insetos mortos sob a concentração i e dose x . Faça uma análise de diagnóstico e verifique se há indícios de sobredispersão aplicando um teste apropriado. Teste a hipótese de paralelismo com todos os pontos e sem as observações discrepantes. Comente.

26. No arquivo **pulso.dat** são descritas as variáveis pulsação em repouso (1: normal, 2: alta), hábito de fumar (1: sim, 2: não) e peso (em kg) de

92 adultos do sexo masculino. Ajuste um modelo logístico linear para explicar a probabilidade de pulsação alta dadas as demais variáveis. Faça uma análise de diagnóstico. Apresente as curvas ajustadas para cada grupo de hábito de fumar com as respectivas bandas de confiança de 95%.

Capítulo 4

Modelos para Dados de Contagem

4.1 Introdução

Neste capítulo apresentamos alguns métodos para a análise de dados de contagem. Inicialmente são apresentados os principais métodos tradicionais e em seguida discutimos a modelagem através de regressão. Duas situações de interesse são consideradas. Na primeira delas, muito comum em estudos de seguimento, as unidades amostrais são classificadas segundo os níveis de categorias, tais como sexo, faixa etária, tipo de tratamento etc, e são acompanhadas por um período fixo pré-estabelecido ou até a ocorrência de um determinado evento. Temos, portanto, um tempo particular de observação para cada unidade amostral, o qual deverá ser incorporado nas análises. Na segunda situação, o interesse é estudarmos o número de ocorrências de um evento particular segundo os níveis de categorias, de modo que seja possível construirmos uma tabela típica de contingência. Aqui, a suposição de distribuição de Poisson para o número de ocorrências do evento em cada configuração de níveis das categorias leva a resultados equivalentes à suposição de distribuição multinomial para as caselas da tabela de contingência formada.

Assim, muitas tabelas de contingência que seriam originalmente analisadas através de um modelo log-linear multinomial podem ser analisadas, alternativamente, por um modelo log-linear de Poisson. A vantagem disso é o fato do modelo log-linear de Poisson ser ajustado mais facilmente do que o modelo log-linear multinomial, além da possibilidade de todos os procedimentos desenvolvidos para os MLGs serem diretamente estendidos para o modelo log-linear de Poisson. Não discutimos, contudo, aspectos particulares na análise de tabelas de contingência, tais como testes ou modelos multinomiais mais específicos.

Discutimos também neste capítulo o fenômeno de sobredispersão que pode ocorrer com dados de contagem quando a variância da variável resposta é maior do que a média. Nesses casos, a suposição de distribuição de Poisson para a resposta é inadequada sendo necessário o uso de modelos alternativos. O modelo de quase-verossimilhança com parâmetro de dispersão leva às mesmas estimativas do modelo de Poisson, porém corrige a variabilidade das estimativas. Daremos, contudo, atenção especial aos modelos com resposta binomial negativa, os quais permitem uma análise mais completa dos dados do que os modelos de quase-verossimilhança. Finalmente, abordamos de forma sucinta os modelos de Poisson e binomial negativo com excesso de zeros.

4.1.1 Métodos clássicos: uma única tabela 2×2

Considere inicialmente a tabela abaixo resultante de um estudo de seguimento (em que indivíduos expostos e não expostos são acompanhados ao longo do tempo por um período fixo ou até a ocorrência de um evento).

	E	\bar{E}
Casos	y_1	y_2
Pessoas-Tempo	t_1	t_2

Vamos assumir que Y_1 e Y_2 seguem, respectivamente, distribuição de Poisson com parâmetros λ_1 e λ_2 , em que λ_1 é a taxa média de casos (por unidade de tempo) no grupo exposto e λ_2 é a taxa média de casos no grupo não exposto. O parâmetro de interesse nesse tipo de estudo é a razão entre as taxas, denotada por $\psi = \frac{\lambda_1}{\lambda_2}$. O objetivo principal é fazermos inferências a respeito de ψ .

A função de probabilidades conjunta de (Y_1, Y_2) fica então dada por

$$\begin{aligned} f(\mathbf{y}; \boldsymbol{\lambda}) &= \frac{e^{-\lambda_1 t_1} (\lambda_1 t_1)^{y_1}}{y_1!} \frac{e^{-\lambda_2 t_2} (\lambda_2 t_2)^{y_2}}{y_2!} \\ &= \exp\{-\psi \lambda_2 t_1 - \lambda_2 t_2 + y_1 \log \psi + (y_1 + y_2) \log \lambda_2 \\ &\quad + y_1 \log t_1 + y_2 \log t_2 - \log y_1! - \log y_2!\}, \end{aligned}$$

em que $\mathbf{y} = (y_1, y_2)^T$ e $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$. Portanto, pelo teorema da fatorização temos que as estatísticas $(Y_1, Y_1 + Y_2)$ são suficientes minimais para (ψ, λ_2) . Logo, condicionando em $Y_1 + Y_2 = m$, obtemos uma distribuição que depende apenas de ψ , isto é

$$\begin{aligned} f(a|m; \psi) &= Pr\{Y_1 = a \mid Y_1 + Y_2 = m\} \\ &= \binom{m}{a} \pi^a (1 - \pi)^{(m-a)}, \end{aligned}$$

em que $\pi = \psi t_1 / \{t_2 + \psi t_1\} = \psi / \{t_2/t_1 + \psi\}$, sendo π a probabilidade de um caso ter sido exposto. Equivalentemente, temos que

$$\psi = \frac{\pi t_2}{(1 - \pi) t_1}.$$

Aqui o interesse é testarmos $H_0 : \psi = 1$ contra $H_1 : \psi \neq 1$, que é equivalente a testarmos $H_0 : \pi = \pi_0$ contra $H_1 : \pi \neq \pi_0$, em que $\pi_0 = t_1/(t_1 + t_2)$.

O nível descritivo exato para testarmos a hipótese H_0 contra H_1 é dado por $P = 2\min\{P_I, P_S\}$, em que

$$P_I = \sum_{x=0}^a \binom{m}{x} \pi_0^x (1 - \pi_0)^{(m-x)}$$

e

$$P_S = \sum_{x=a}^m \binom{m}{x} \pi_0^x (1 - \pi_0)^{(m-x)}.$$

Podemos usar o resultado abaixo (ver, por exemplo, Leemis e Trivedi, 1996) para expressarmos a distribuição condicional de Y_1 dado $Y_1 + Y_2 = m$ em função de uma distribuição $F_{u,v}$, ou seja uma distribuição F com u e v graus de liberdade. Supondo $Y \sim B(n, p)$, temos que

$$Pr(Y \geq y) = Pr\{F_{2y, 2(n-y+1)} < (n-y+1)p/y(1-p)\}, \quad (4.1)$$

com $0 < p < 1$. Daí temos, sob $H_0 : \pi = \pi_0$, que

$$\begin{aligned} P_I &= 1 - \sum_{x=a+1}^m \binom{m}{x} \pi_0^x (1 - \pi_0)^{(m-x)} \\ &= 1 - Pr\left\{F_{u,v} < \frac{(m-a-1+1)\pi_0}{(a+1)(1-\pi_0)}\right\} \\ &= 1 - Pr\{F_{u,v} < bt_1/(a+1)t_2\}, \end{aligned}$$

com $b = m-a$, $u = 2(a+1)$ e $v = 2b$. Similarmente, obtemos sob $H_0 : \pi = \pi_0$, que

$$P_S = Pr\{F_{u,v} < (b+1)t_1/at_2\},$$

com $u = 2a$ e $v = 2(b+1)$. De (4.1) temos que os limites exatos de confiança para p , para um coeficiente de confiança $(1 - \alpha)$, são tais que

$$\frac{\alpha}{2} = \sum_{t \geq y} Pr(Y = t; \hat{p}_I) = Pr(Y \geq y; \hat{p}_I)$$

e

$$\frac{\alpha}{2} = \sum_{t \leq y} Pr(Y = t; \hat{p}_S) = 1 - Pr(Y \geq y + 1; \hat{p}_S).$$

Logo, usando (4.1) obtemos

$$\hat{p}_I = \frac{1}{1 + \frac{n-y+1}{yF_{2y,2(n-y+1)}(\alpha/2)}}$$

e

$$\hat{p}_S = \frac{1}{1 + \frac{n-y}{(y+1)F_{2(y+1),2(n-y)}(1-\alpha/2)}},$$

em que $F_{u,v}(\alpha/2)$ denota o percentil $\alpha/2$ de uma distribuição F com u e v graus de liberdade. Portanto, temos para π , fazendo $y = a$ e $m = a + b$, o limite inferior exato de confiança

$$\begin{aligned} \hat{\pi}_I &= \frac{1}{1 + \frac{b+1}{aF_{u,v}(\alpha/2)}} \\ &= aF_{u,v}(\alpha/2) / \{b + 1 + aF_{u,v}(\alpha/2)\}, \end{aligned}$$

em que $u = 2a$ e $v = 2(b + 1)$. De forma análoga obtemos o limite superior exato

$$\begin{aligned} \hat{\pi}_S &= \frac{1}{1 + \frac{b}{aF_{u,v}(1-\alpha/2)}} \\ &= aF_{u,v}(1 - \alpha/2) / \{b + aF_{u,v}(1 - \alpha/2)\}, \end{aligned}$$

em que $u = 2(a + 1)$ e $v = 2b$. A estimativa de máxima verossimilhança para ψ considerando a distribuição não condicional (produto de Poissons independentes) fica dada por

$$\tilde{\psi} = \frac{\tilde{\lambda}_1}{\tilde{\lambda}_2},$$

em que $\tilde{\lambda}_1 = y_1/t_1$ e $\tilde{\lambda}_2 = y_2/t_2$. Portanto, obtemos $\tilde{\psi} = y_1t_2/y_2t_1$. Se, por outro lado, utilizamos a distribuição condicional, $B(m, \pi)$, temos que a

estimativa de máxima verossimilhança de ψ fica dada por

$$\hat{\psi} = \frac{\hat{\pi}t_2}{(1 - \hat{\pi})t_1},$$

em que $\hat{\pi} = y_1/m$ e $(1 - \hat{\pi}) = y_2/m$. Logo, $\hat{\psi}$ fica expresso de forma análoga ao caso não condicional. A explicação desse fato, que não ocorre nos estudos de caso e controle com respostas binomiais, é que a estatística $Y_1 + Y_2$, além de ser suficiente para λ_2 , é também ancilar para ψ , isto é, não contém qualquer informação acerca de ψ . No caso do produto de duas binomiais independentes, $Y_1 + Y_2$ é suficiente para π_2 , no entanto, não é ancilar para ψ . Uma consequência desse fato é que a estimativa de máxima verossimilhança condicional não coincide com a estimativa não condicional.

Aplicação

Vamos considerar, como aplicação, os dados apresentados em Boice e Monson (1977) referentes a um estudo de seguimento com dois grupos de mulheres com tuberculose, um grupo exposto a radiação e o outro grupo não exposto, sendo observado ao longo do tempo o desenvolvimento ou não de câncer de mama. Os resultados desse estudo são resumidos na Tabela 4.1.

Tabela 4.1
*Casos de câncer de mama em mulheres
com tuberculose.*

	Radiação	
	Exposto	Não Exposto
Casos	41	15
Pessoas-anos	28010	19017

Temos, portanto, que $a = 41$, $b = 15$, $t_1 = 28010$ e $t_2 = 19017$. Os níveis descritivos correspondentes ao teste exato para testar $H_0 : \psi = 1$ contra

$H_1 : \psi \neq 1$ ficam dados por

$$P_I = 1 - Pr\{F_{84,30} < 0,526\} = 0,988$$

e

$$P_S = Pr\{F_{82,32} < 0,575\} = 0,024,$$

e obtemos o nível descritivo $P = 0,048$ que indica, para um nível de significância de 5%, pela rejeição de H_0 . Isso quer dizer que há indícios de que mulheres com tuberculose e expostas a radiação têm uma chance maior de desenvolvimento de câncer de mama do que mulheres não expostas com a mesma doença. Uma estimativa pontual de máxima verossimilhança para ψ fica dada por $\hat{\psi} = \frac{0,732 \times 19017}{0,268 \times 28010} = 1,85$ e um intervalo exato de confiança de 95% para π tem os limites

$$\begin{aligned}\hat{\pi}_I &= 41 \times F_{82,32}(0,025) / \{16 + 41 \times F_{82,32}(0,025)\} \\ &= 0,597 \text{ e} \\ \hat{\pi}_S &= 41 \times F_{84,30}(0,975) / \{15 + 41 \times F_{84,30}(0,975)\} \\ &= 0,838.\end{aligned}$$

Desses limites obtemos os limites exatos de confiança para ψ

$$\begin{aligned}\hat{\psi}_I &= \frac{\hat{\pi}_I t_2}{(1 - \hat{\pi}_I) t_1} = \frac{0,597 \times 19017}{(1 - 0,597) \times 28010} \\ &= 1,007 \text{ e} \\ \hat{\psi}_S &= \frac{\hat{\pi}_S t_2}{(1 - \hat{\pi}_S) t_1} = \frac{0,838 \times 19017}{(1 - 0,838) \times 28010} \\ &= 3,512.\end{aligned}$$

Temos que o intervalo $[1,007; 3,512]$ não cobre o valor $\psi = 1$, como era esperado.

4.1.2 Estratificação: k tabelas 2×2

Se os dados são estratificados segundo um fator com k níveis, cada tabela resultante pode ser expressa na forma abaixo.

	E	\bar{E}
Casos	y_{1i}	y_{2i}
Pessoas-Tempo	t_{1i}	t_{2i}

Temos aqui as suposições $Y_{1i} \sim P(\lambda_{1i}t_{1i})$ e $Y_{2i} \sim P(\lambda_{2i}t_{2i})$, $i = 1, \dots, k$. Consequentemente, a distribuição condicional de Y_{1i} dado $Y_{1i} + Y_{2i} = m_i$ é uma $B(m_i, \pi_i)$, em que $\pi_i = \psi_i / \{t_{2i}/t_{1i} + \psi_i\}$, ou equivalentemente

$$\psi_i = \frac{\pi_i t_{2i}}{(1 - \pi_i) t_{1i}}.$$

Se temos interesse em testar a homogeneidade das razões de taxas $H_0 : \psi_1 = \dots = \psi_k$ contra a alternativa de pelo menos duas diferentes, a estimativa comum $\hat{\psi}$, sob H_0 , sai do sistema de equações

$$\sum_{i=1}^k y_{1i} = \hat{\psi} \sum_{i=1}^k m_i / \{\hat{\psi} + t_{2i}/t_{1i}\},$$

que tem no máximo uma raiz positiva. Alternativamente, de forma análoga aos estudos de caso e controle, podemos construir uma versão da estimativa de Mantel-Haenszel dada por

$$\hat{\psi}_{MH} = \frac{\sum_{i=1}^k y_{1i} t_{2i} / t_i}{\sum_{i=1}^k y_{2i} t_{1i} / t_i},$$

em que $t_i = t_{1i} + t_{2i}$. Segundo Breslow e Day (1987), $\hat{\psi}_{MH}$ é consistente e assintoticamente normal com variância assintótica estimada por

$$\hat{\text{Var}}_A(\hat{\psi}_{MH}) = \frac{\hat{\psi}_{MH} \sum_{i=1}^k t_{1i} t_{2i} m_i / t_i^2}{\left\{ \sum_{i=1}^k \frac{t_{1i} t_{2i} m_i}{t_i (t_{1i} + \hat{\psi}_{MH} t_{2i})} \right\}^2}.$$

A estatística sugerida para testar H_0 é definida por

$$X^2 = \sum_{i=1}^k \left\{ \frac{(y_{1i} - \hat{y}_{1i})^2}{\hat{y}_{1i}} + \frac{(y_{2i} - \hat{y}_{2i})^2}{\hat{y}_{2i}} \right\},$$

em que $\hat{y}_{1i} = m_i \hat{\pi}_i$, $\hat{y}_{2i} = m_i(1 - \hat{\pi}_i)$ e

$$\hat{\pi}_i = \frac{\hat{\psi}_{MH}}{t_{2i}/t_{1i} + \hat{\psi}_{MH}}.$$

A distribuição nula assintótica de X^2 é uma qui-quadrado com $k - 1$ graus de liberdade. Quando a hipótese de homogeneidade das razões de chances não é rejeitada, podemos testar a hipótese de associação entre o fator e a doença levando em conta o efeito de estrato. Isso equivale a testarmos $H_0 : \psi = 1$ contra $H_1 : \psi \neq 1$. O teste qui-quadrado apropriado é dado por

$$\begin{aligned} X^2 &= \frac{\{\sum_{i=1}^k y_{2i} - \sum_{i=1}^k E(Y_{2i}|m_i, \psi = 1)\}^2}{\sum_{i=1}^k \text{Var}(Y_{2i}|m_i, \psi = 1)} \\ &= \frac{\{\sum_{i=1}^k y_{2i} - \sum_{i=1}^k m_i t_{1i}/(t_{1i} + t_{2i})\}^2}{\sum_{i=1}^k m_i t_{1i} t_{2i}/(t_{1i} + t_{2i})^2}. \end{aligned} \quad (4.2)$$

A distribuição nula assintótica de X^2 , quando $\frac{n_i}{n} \rightarrow a_i > 0$ fazendo $n \rightarrow \infty$, em que $n = n_1 + \dots + n_k$, é uma χ_1^2 .

Temos que a variância assintótica de $\log(\hat{\psi}_{MH})$ é estimada por

$$\hat{\text{Var}}_A\{\log(\hat{\psi}_{MH})\} = \hat{\psi}_{MH}^{-2} \hat{\text{Var}}_A(\hat{\psi}_{MH}).$$

Assim, um intervalo assintótico de confiança com coeficiente $(1 - \alpha)$ para $\log \psi$ fica dado por $\log(\hat{\psi}_{MH}) \pm z_{(1-\alpha/2)} \hat{\psi}_{MH}^{-1} \{\hat{\text{Var}}_A(\hat{\psi}_{MH})\}^{1/2}$ o que implica nos limites de confiança superior e inferior dados abaixo

$$\begin{aligned} \hat{\psi}_I &= \hat{\psi}_{MH} \exp\{-z_{(1-\alpha/2)} \hat{\psi}_{MH}^{-1} \sqrt{\hat{\text{Var}}_A(\hat{\psi}_{MH})}\} \text{ e} \\ \hat{\psi}_S &= \hat{\psi}_{MH} \exp\{-z_{(1-\alpha/2)} \hat{\psi}_{MH}^{-1} \sqrt{\hat{\text{Var}}_A(\hat{\psi}_{MH})}\}. \end{aligned}$$

Esse intervalo deve ser construído quando a aplicação da estatística (4.2) levar à rejeição da hipótese $H_0 : \psi = 1$.

4.2 Modelos de Poisson

4.2.1 Propriedades da Poisson

Vamos supor que $Y \sim P(\lambda)$ cuja função de probabilidades é dada por

$$Pr(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, 2, \dots$$

Podemos mostrar (ver, por exemplo, McCullagh e Nelder, 1989, p. 195) que quando $\lambda \rightarrow \infty$

$$(Y - \lambda)/\sqrt{\lambda} \rightarrow_d N(0, 1).$$

Em outras palavras, para λ grande temos que Y segue aproximadamente uma distribuição normal de média λ e desvio padrão $\sqrt{\lambda}$. Se queremos, no entanto, aplicar um modelo normal linear para explicar λ , teremos o inconveniente do desvio padrão depender da média, o que inviabiliza o uso de um modelo normal linear homocedástico. Uma maneira de contornarmos esse problema é através da aplicação de uma transformação na resposta Y de modo a alcançarmos a normalidade e a constância de variância, mesmo que aproximadamente. Nesse sentido, temos que se Y é Poisson, segue quando $\lambda \rightarrow \infty$ o seguinte resultado:

$$\{\sqrt{Y} - E(\sqrt{Y})\} \rightarrow_d N(0, 1/4).$$

Portanto, quando λ é grande, a variável aleatória $2\{\sqrt{Y} - E(\sqrt{Y})\}$ segue aproximadamente uma distribuição $N(0, 1)$. Assim, se temos uma amostra aleatória Y_1, \dots, Y_n tal que $Y_i \sim P(\lambda_i)$ e queremos explicar λ_i através de variáveis explicativas, podemos propor para λ_i grande, $\forall i$, o modelo normal linear abaixo

$$\sqrt{Y_i} = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i,$$

em que $\epsilon_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$. Isso foi feito na Seção 1.12.3 no exemplo sobre sobrevivência de bactérias.

4.2.2 Modelos log-lineares: k tabelas 2×2

Como foi visto no Capítulo 1, os modelos log-lineares são recomendados para a análise de dados de contagem, mesmo quando o tempo de observação não é o mesmo para cada unidade amostral. Em particular, se temos um conjunto de k tabelas 2×2 , uma modelagem possível para a taxa média por unidade de tempo em cada casela é supormos que $Y_{1i} \sim P(\lambda_{1i}t_{1i})$ e $Y_{2i} \sim P(\lambda_{2i}t_{2i})$, $i = 1, \dots, k$, mutuamente independentes e com a seguinte parte sistemática:

$$\begin{aligned}\log \lambda_{11} &= \alpha, \\ \log \lambda_{21} &= \alpha + \beta, \\ \log \lambda_{1i} &= \alpha + \gamma_i \text{ e} \\ \log \lambda_{2i} &= \alpha + \beta + \gamma_i + \delta_i,\end{aligned}$$

para $i = 2, \dots, k$. Portanto, temos a reparametrização $(\lambda_{11}, \lambda_{21}, \dots, \lambda_{1k}, \lambda_{2k}) \rightarrow (\alpha, \beta, \gamma_2, \delta_2, \dots, \gamma_k, \delta_k)$. A razão de taxas na i -ésima tabela fica definida por $\psi_i = \lambda_{2i}/\lambda_{1i} = \exp(\beta + \delta_i)$, com $\delta_1 = 0$. Assim, testarmos $H_0 : \psi_1 = \dots = \psi_k$ é o mesmo que testarmos na nova parametrização $H_0 : \delta_2 = \dots = \delta_k = 0$, o que significa não haver interação entre as tabelas. Devemos lembrar que γ_i é o efeito da i -ésima tabela com relação à primeira tabela. Logo, testarmos $H_0 : \gamma_2 = \dots = \gamma_k$, dado que $\delta_i = 0$, significa testarmos a ausência de efeito de estrato.

Aqui t_{ij} denota o total de unidades de tempo na casela (i, j) , $i = 1, 2$ e $j = 1, \dots, k$. Assim, temos que $\log \mu_{ij} = \log t_{ij} + \log \lambda_{ij}$, em que $\log t_{ij}$ desempenha o papel de um *offset*. Pela propriedade de que os totais marginais $Y_{1i} + Y_{2i}$ são estatísticas suficientes para os parâmetros $\lambda_{21}, \dots, \lambda_{2k}$ e ancilares para ψ_1, \dots, ψ_k , devemos esperar que as estimativas de máxima verossimilhança não condicionais $\hat{\psi}_i = \exp(\hat{\beta} + \hat{\delta}_i)$, $i = 1, \dots, k$, coincidam com as estimativas condicionais.

Uma maneira de verificarmos se é razoável a suposição de distribuição de Poisson nas unidades de tempo é tratarmos $\log T_{ij}$ como sendo uma variável explicativa, isto é, ajustarmos o modelo com parte sistemática dada por $\log \mu_{ij} = \theta \log t_{ij} + \log \lambda_{ij}$. Assim, ao testarmos $H_0 : \theta = 1$ contra $H_1 : \theta \neq 1$, a não rejeição de H_0 indica que a suposição de distribuição de Poisson nas unidades de tempo não é inadequada. Como veremos a seguir isso significa que os tempos têm distribuição exponencial.

Relação com a exponencial

O logaritmo da função de verossimilhança do modelo de Poisson para a análise de k tabelas 2×2 é dado por

$$L(\boldsymbol{\lambda}) \propto \sum_{i=1}^2 \sum_{j=1}^k (y_{ij} \log \lambda_{ij} - \lambda_{ij} t_{ij}), \quad (4.3)$$

em que $\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{21}, \dots, \lambda_{k1}, \lambda_{k2})^T$. Temos, portanto, para cada casela (i, j) um estudo de seguimento em que as unidades amostrais foram observadas um total de t_{ij} unidades de tempo. Sem perda de generalidade, vamos supor que $t_{ij} = N$ e que nesse substrato foram acompanhadas I unidades amostrais cujos tempos de observação foram, respectivamente, N_1, N_2, \dots, N_I . Faremos $u_\ell = 1$ se o evento sob estudo ocorrer para a ℓ -ésima unidade amostral antes de um tempo pré-fixado T . Quando o evento não ocorrer para a ℓ -ésima unidade amostral durante o período de estudo ($u_\ell = 0$) dizemos que há censura, sendo aqui o tempo de observação dado por $N_\ell = T$. Vamos supor ainda que a taxa de ocorrência do evento, que é definida por

$$\xi = \lim_{\Delta t \rightarrow 0} \frac{Pr\{\text{o evento ocorrer em } (t, t + \Delta t)\}}{\Delta t},$$

dado que o evento não ocorreu até o tempo t , permanece constante durante o período de observação. Finalmente, assumimos que as ocorrências são

independentes entre as unidades amostrais. Sob essas condições, mostramos que a distribuição conjunta das variáveis (N_ℓ, u_ℓ) , $\ell = 1, \dots, I$, é um produto de I exponenciais independentes de parâmetro ξ . Se o evento ocorrer antes do tempo T para a ℓ -ésima unidade amostral ($N_\ell < T, u_\ell = 1$) a mesma contribui com o fator $\xi e^{-\xi N_\ell}$ na função de verossimilhança. Caso contrário ($N_\ell = T, u_\ell = 0$), o fator é dado por $e^{-\xi T}$. O logaritmo da função de verossimilhança conjunta fica então dado por

$$\begin{aligned} L(\xi) &= \sum_{\ell=1}^I (u_\ell \log \xi - N_\ell \xi) \\ &= \log \xi \sum_{\ell=1}^I u_\ell - \xi \sum_{\ell=1}^I N_\ell. \end{aligned} \quad (4.4)$$

Se considerarmos que para a casela (i, j) o evento ocorreu y_{ij} vezes, as unidades amostrais foram observadas um total de t_{ij} unidades de tempo e a taxa de ocorrência do evento é λ_{ij} , então (4.4) fica reexpressa na forma

$$L(\lambda_{ij}) = y_{ij} \log \lambda_{ij} - \lambda_{ij} t_{ij},$$

que coincide com o termo geral da expressão (4.3). Portanto, a suposição de modelo de regressão log-linear de Poisson com *offset* $\log t_{ij}$ equivale à suposição de tempos exponenciais para as unidades amostrais. No entanto, é importante ressaltarmos que as inferências exatas para ξ no modelo exponencial são bastante complexas em virtude da ocorrência de censura (ver discussão, por exemplo, em Breslow e Day, 1987, p. 132). Já os resultados assintóticos são equivalentes àqueles obtidos para o modelo de Poisson.

Aplicação

A Tabela 4.2 resume os resultados de um estudo de seguimento em que doutores Britânicos foram acompanhados durante a década de 50 e observado,

em particular, a ocorrência de mortes por câncer de pulmão segundo o consumo médio diário de cigarros e a faixa etária. Esses dados estão disponíveis no arquivo **breslow.dat**. Denotamos por Y_{ij} o número de mortes para o i -ésimo nível de consumo e j -ésima faixa etária, $i, j = 1, \dots, 4$. Vamos supor que $Y_{ij} \sim P(\lambda_{ij}t_{ij})$, em que λ_{ij} é a taxa média de mortes por unidade de tempo para o consumo i e faixa etária j . O modelo saturado nesse caso é dado por

$$\log \lambda_{ij} = \alpha + \beta_i + \gamma_j + \delta_{ij},$$

em que $\beta_1 = 0$, β_i é o efeito da i -ésima classe de consumo de cigarros com relação à classe de não fumantes, $i = 2, 3, 4$, $\gamma_1 = 0$, γ_j é o efeito da j -ésima faixa etária com relação à faixa etária de 40 – 49 anos e δ_{ij} denota a interação entre faixa etária e consumo de cigarros, em que $\delta_{i1} = \delta_{1j} = 0$, para $i, j = 1, \dots, 4$.

Tabela 4.2

Número de casos de morte por câncer de pulmão e pessoas-anos de observação em doutores Britânicos segundo a faixa etária e o consumo médio diário de cigarros.

Consumo médio diário de cigarros		Faixa Etária			
		40-49	50-59	60-69	70-80
0	mortes	0	3	0	3
	p-anos	33679	21131,5	10599	4495,5
1-9	mortes	0	1	3	3
	p-anos	6002,5	4396	2813,5	1664,5
10-30	mortes	7	29	41	45
	p-anos	34414,5	25429	13271	4765,5
+ 30	mortes	3	16	36	11
	p-anos	5881	6493,5	3466,5	769

O teste de ausência de interação, $H_0 : \delta_{ij} = 0, \forall ij$, contra a alternativa de pelo menos um parâmetro diferente de zero forneceu $\xi_{RV} = 11,91$ (9 graus de liberdade) que equivale a um nível descritivo $P = 0,218$. Adotamos, portanto, um modelo sem interação.

Tabela 4.3

Estimativas dos parâmetros do modelo log-linear de Poisson para explicar a taxa média de morte de doutores Britânicos com câncer de pulmão.

Efeito	Parâmetro	Estimativa	E/E.Padrão
Constante	α	-11,424	-22,44
C(1-9)	β_2	1,409	2,53
C(10-20)	β_3	2,866	6,86
C(+30)	β_4	3,758	8,80
F(50-59)	γ_2	1,769	5,10
F(60-69)	γ_3	2,897	8,62
F(70-80)	γ_4	3,791	11,12

As estimativas são apresentadas na Tabela 4.3. Notamos claramente que as estimativas são significativamente diferentes de zero e que há fortes indícios de um aumento (exponencial) da taxa média de mortes com o aumento da faixa etária e/ou com o aumento do consumo médio diário de cigarros. O ajuste do modelo com $\log T_{ij}$ como variável explicativa forneceu a estimativa de máxima verossimilhança $\hat{\theta} = 1,839(0,610)$. O teste de Wald para testarmos $H_0 : \theta = 1$ contra $H_1 : \theta \neq 1$ forneceu o valor $\xi_{RV} = 1,89$, cujo nível descritivo é dado por $P = 0,17$, indicando que o modelo pode ser ajustado com $\log t_{ij}$ como sendo *offset*.

4.2.3 Modelos gerais de Poisson

Vamos supor agora que Y_i são variáveis aleatórias independentes distribuídas tais que $Y_i \sim P(\mu_i)$, $i = 1, \dots, n$, com parte sistemática dada por $g(\mu_i) = \eta_i$,

em que $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ contém valores de variáveis explicativas e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é um vetor de parâmetros desconhecidos. As ligações mais utilizados são logarítmica ($g(\mu_i) = \log \mu_i$), raiz quadrada ($g(\mu_i) = \sqrt{\mu_i}$) e identidade ($g(\mu_i) = \mu_i$). O processo iterativo para estimação de $\boldsymbol{\beta}$, como foi visto na Seção 1.6.1, é dado por

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)},$$

$m = 0, 1, \dots$, variável dependente modificada $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$, $\mathbf{y} = (y_1, \dots, y_n)^T$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$, $\mathbf{V} = \text{diag}\{\mu_1, \dots, \mu_n\}$ e $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$ com $\omega_i = (d\mu_i/d\eta_i)^2/\mu_i$. Em particular temos $\omega_i = \mu_i$ para ligação logarítmica, $\omega_i = 4$ para ligação raiz quadrada e $\omega_i = \mu_i^{-1}$ para ligação identidade.

No caso das unidades experimentais serem observadas em tempos distintos t_i 's e for assumido que $Y_i \sim P(\lambda_i t_i)$, $i = 1, \dots, n$, a parte sistemática do modelo para ligação logarítmica fica dada por

$$\log \mu_i = \log t_i + \mathbf{x}_i^T \boldsymbol{\beta},$$

em que $\log t_i$ desempenha papel de *offset* e isso deve ser informado ao sistema. Outra possibilidade é incluirmos os tempos t_i 's como valores da variável explicativa $\log T_i$. Nesse caso, a parte sistemática assume a forma

$$\log \mu_i = \theta \log t_i + \mathbf{x}_i^T \boldsymbol{\beta}.$$

O teste de $H_0 : \theta = 1$ contra $H_1 : \theta \neq 1$ verifica se $\log t_i$ deve ser incluído no modelo como *offset*. A não rejeição da hipótese nula significa a suposição de tempos exponenciais nas unidades experimentais.

O estimador de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ é consistente, eficiente e tem distribuição assintótica dada por

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim N_p(\mathbf{0}, (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}),$$

portanto, assintoticamente, $\text{Var}(\hat{\beta}) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$.

4.2.4 Qualidade do ajuste

A função desvio de um modelo de Poisson supondo $y_i > 0, \forall i$, é definida por

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \{y_i \log(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)\}.$$

Porém, se $y_i = 0$, o i -ésimo termo de $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ fica dado por $2\hat{\mu}_i$.

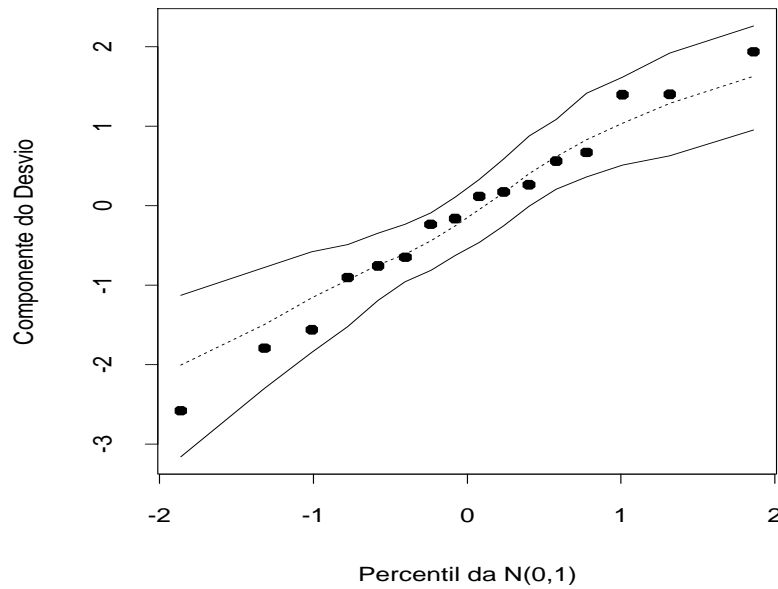


Figura 4.1: Gráfico normal de probabilidades referente ao modelo log-linear de Poisson ajustado aos dados sobre morte por câncer de pulmão de doutores Britânicos.

Em particular, para ligação logarítmica e se o modelo inclui uma constante na parte sistemática, mostramos que $\sum_{i=1}^n (y_i - \hat{\mu}_i) = 0$, ficando a função desvio reexpressa na forma $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n y_i \log(y_i / \hat{\mu}_i)$. Logo, se

particionamos o vetor de parâmetros tal que $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$, em que $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ são subvetores de dimensão $p - q$ e q , respectivamente, a estatística da razão de verossimilhanças para testarmos $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ contra $H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$ em modelos log-lineares fica dada por

$$\begin{aligned}\xi_{RV} &= D(\mathbf{y}; \hat{\boldsymbol{\mu}}_0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}}) \\ &= 2 \sum_{i=1}^n y_i \log(\hat{\mu}_{0i}/\hat{\mu}_i).\end{aligned}$$

Sob H_0 e para grandes amostras $\xi_{RV} \sim \chi_q^2$. Os resultados assintóticos para os modelos de Poisson valem tanto para p fixo e $n \rightarrow \infty$ como para n fixo e $\mu_i \rightarrow \infty, \forall i$.

4.2.5 Técnicas de diagnóstico

Um dos resíduos mais recomendados para modelos com resposta de Poisson é o componente do desvio padronizado, que para $y_i > 0$, fica dado por

$$t_{D_i} = \pm \frac{\sqrt{2}}{\sqrt{1 - \hat{h}_{ii}}} \{y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)\}^{1/2},$$

em que h_{ii} é o i -ésimo elemento da diagonal principal da matriz de projeção $\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$. Quando $y_i = 0$ o resíduo componente do desvio padronizado assume a forma $t_{D_i} = \pm \sqrt{2\hat{\mu}_i}/\sqrt{1 - \hat{h}_{ii}}$.

Estudos de simulação (ver Williams, 1984) mostram que em geral a distribuição de t_{D_i} não se afasta muito da distribuição normal padrão, podendo ser usadas nas análises de diagnóstico as mesmas interpretações da regressão normal linear. Em particular, a construção de envelopes é fortemente recomendada para t_{D_i} .

A Figura 4.1 apresenta o gráfico normal de probabilidades para o resíduo t_{D_i} correspondente ao modelo ajustado aos dados da Tabela 4.2. Como podemos notar, todos os resíduos caíram dentro do envelope gerado sem apre-

sentarem nenhuma tendência sistemática, indicando que a suposição de distribuição de Poisson parece ser bastante razoável. O programa utilizado para gerarmos o gráfico de envelopes é apresentado no Apêndice B. Os resultados do modelo ajustado devem ser colocados no arquivo `fit.model`.

4.2.6 Aplicação

Como ilustração vamos considerar os dados apresentados em Neter et al. (1996, p. 613) sobre o perfil dos clientes de uma determinada loja oriundos de 110 áreas de uma cidade. O objetivo do estudo é relacionar o número esperado de clientes em cada área com as seguintes variáveis explicativas em cada área: número de domicílios (em mil), renda média anual (em mil USD), idade média dos domicílios (em anos), distância ao concorrente mais próximo (em milhas) e distância à loja (em milhas). Portanto, a área é a unidade experimental. Esses dados estão também descritos no arquivo `store.dat`.

Tabela 4.4
*Estimativas dos parâmetros do modelo log-linear
de Poisson ajustado aos dados sobre perfil
de clientes.*

Efeito	Parâmetro	Estimativa	E/E.Padrão
Constante	α	2,942	14,21
Domicílio	β_1	0,606	4,27
Renda	β_2	-0,012	-5,54
Idade	β_3	-0,004	-2,09
Dist1	β_4	0,168	6,54
Dist2	β_5	-0,129	-7,95

Na Figura 4.2 são apresentados os diagramas de dispersão entre o número de clientes (variável resposta) e as variáveis explicativas renda e idade média, distância ao concorrente mais próximo (dist1) e distância à loja (dist2).

Indícios mais evidentes de relação linear podem ser observados entre a resposta e as distâncias dist1 e dist2. Ou seja, há indícios de que o número de clientes aumenta à medida que a distância ao concorrente mais próximo aumenta e a distância à loja diminui.

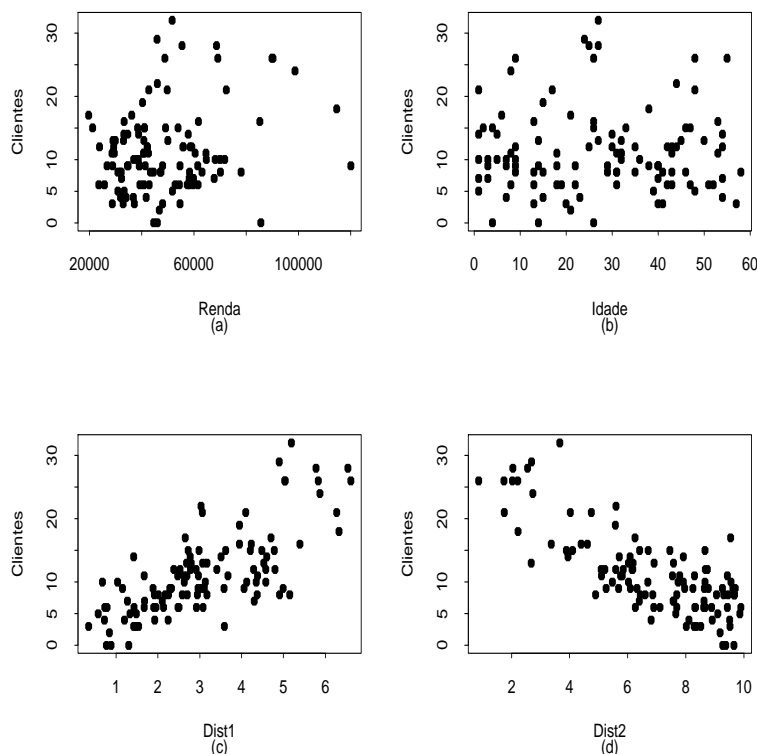


Figura 4.2: Diagramas de dispersão entre o número de clientes que visitaram a loja e algumas variáveis explicativas.

Denotamos por Y_i o número de clientes da i -ésima área que foram à loja no período determinado. Vamos supor que $Y_i \sim P(\mu_i)$ com parte sistemática dada por

$$\log \mu_i = \alpha + \beta_1 \text{domic}_i + \beta_2 \text{renda}_i + \beta_3 \text{idade}_i + \beta_4 \text{dist1}_i + \beta_5 \text{dist2}_i.$$

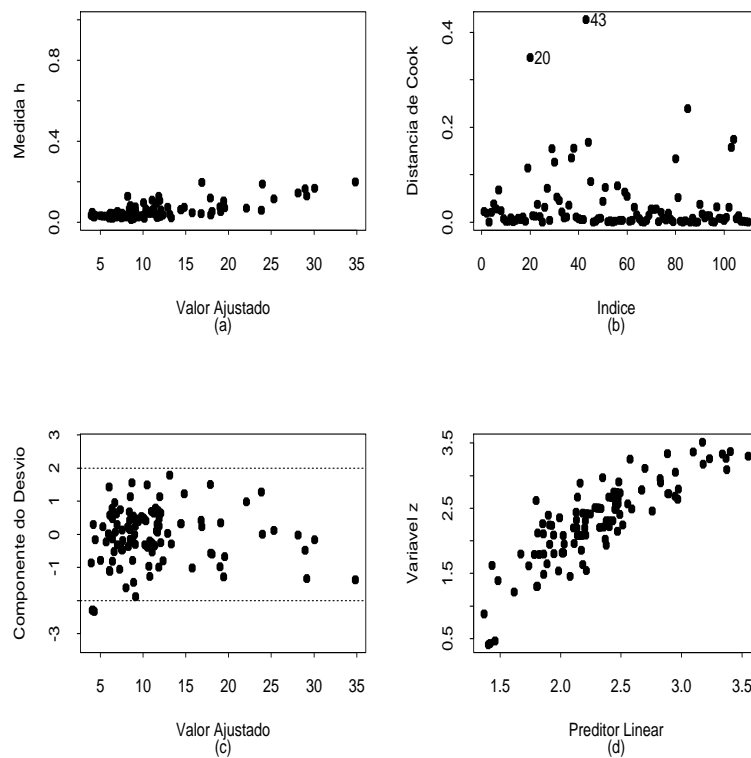


Figura 4.3: Gráficos de diagnóstico referentes ao modelo log-linear de Poisson ajustado aos dados sobre perfil de clientes.

Temos que a variável número de domicílios (*domic*) deve ser incluída no modelo uma vez que as áreas não têm o mesmo número de domicílios. As estimativas dos parâmetros são apresentadas na Tabela 4.4 e como podemos notar todas as estimativas são altamente significativas. O desvio do modelo foi de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 114,98$ (104 graus de liberdade) que equivale a um nível descritivo $P = 0,35$ indicando um ajuste adequado. Notamos pela tabela que o número esperado de clientes na loja cresce com o aumento do número de domicílios na área e da distância ao concorrente mais próximo, porém diminui com o aumento da renda média e da idade média dos domicílios bem como da

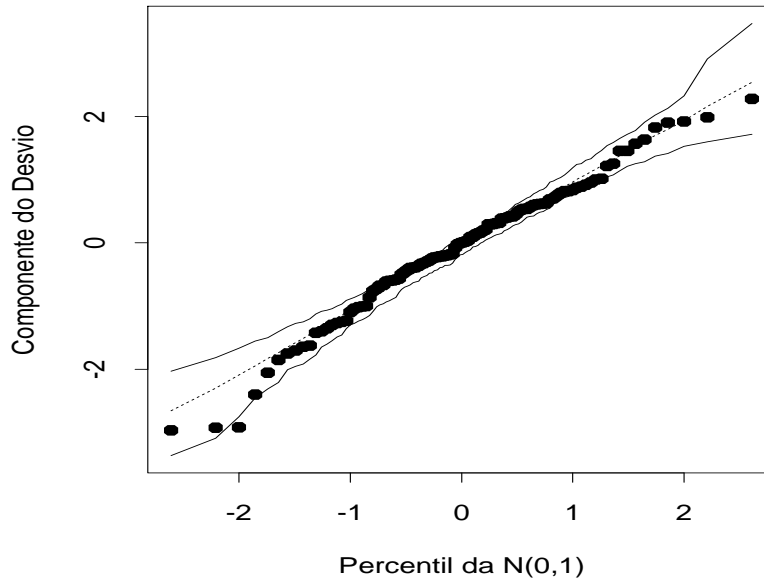


Figura 4.4: Gráfico normal de probabilidades referente ao modelo log-linear de Poisson ajustado aos dados sobre perfil de clientes.

distância da área à loja. Isso sugere que deve ser uma loja de conveniência.

Podemos fazer algumas interpretações. Por exemplo, se aumentarmos em 1 mil USD a renda média dos domicílios de uma determinada área esperamos aumento relativo no número de clientes que irão à loja de $\exp(-0,012) = 0,988$. Ou seja, decréscimo de 1,2%. Por outro lado, se a distância ao concorrente mais próximo aumentar em uma milha esperamos aumento relativo no número de clientes de $\exp(0,168) = 1,183$. Ou seja, aumento de 18,3%.

Dentre as observações destacadas pelos gráficos de diagnóstico (ver Figura 4.3), apenas as áreas #20 e #43 apresentam algumas variações desproporcionais nas estimativas dos parâmetros, porém não houve mudança inferencial. Pela Figura 4.3d não há indícios de que a ligação utilizada seja inapropriada e o gráfico de envelope (Figura 4.4) não apresenta indicações de afastamentos

sérios da suposição de distribuição de Poisson para o número de clientes.

4.3 Modelos com resposta binomial negativa

4.3.1 Distribuição binomial negativa

O fenômeno de sobredispersão, similarmente ao caso de dados com resposta binária discutido na Seção 3.6.14, ocorre quando é esperada uma distribuição de Poisson para a resposta, porém a variância é maior do que a resposta média. Uma causa provável desse fenômeno é a heterogeneidade das unidades amostrais que pode ser devido à variabilidades interunidades experimentais. Isso pode ser visto, por exemplo, supondo que para um conjunto fixo $\mathbf{x} = (x_1, \dots, x_p)^T$ de valores de variáveis explicativas, $Y|z$ tem média z e variância z , no entanto Z , que é não observável, varia nas unidades amostrais com \mathbf{x} fixo, de modo que $E(Z) = \mu$. Então,

$$E(Y) = E[E(Y|Z)] = E[Z] = \mu \text{ e}$$

$$\begin{aligned} \text{Var}(Y) &= E[\text{Var}(Y|Z)] + \text{Var}[E(Y|Z)] \\ &= \mu + \text{Var}(Z). \end{aligned}$$

Podemos, adicionalmente, supor que $Y|z$ tem distribuição de Poisson com média z e função de probabilidades denotada por $f(y|z)$ e que Z segue uma distribuição gama de média μ e parâmetro de dispersão $k = \phi\mu$ cuja função de densidade será denotada por $g(z; \mu, k)$.

Temos $E(Z) = \mu$ e $\text{Var}(Z) = \mu^2/k$ de modo que $E(Y) = \mu$ e $\text{Var}(Y) = \mu + \mu^2/k = \mu(1 + \phi)/\phi$. Assim, as funções densidades $f(y|z)$ e $g(z; \mu, k)$ assumem as seguintes formas:

$$f(y|z) = \frac{e^{-z} z^y}{y!} \text{ e } g(z; \mu, k) = \frac{1}{\Gamma(k)} \left(\frac{zk}{\mu} \right)^k e^{-\frac{kz}{\mu}} \frac{1}{z}.$$

Logo, Y tem função de probabilidades dada por

$$\begin{aligned} Pr\{Y = y\} &= \int_0^\infty f(y|z)g(z; \mu, k)dz \\ &= \frac{1}{y!\Gamma(k)} \left(\frac{k}{\mu}\right)^k \int_0^\infty e^{-z(1+k/\mu)} z^{k+y-1} dz. \end{aligned}$$

Fazendo a transformação de variável $t = z(1 + \frac{k}{\mu})$ temos que $\frac{dz}{dt} = (1 + \frac{k}{\mu})^{-1}$. Então,

$$\begin{aligned} Pr\{Y = y\} &= \frac{1}{y!\Gamma(k)} \left(\frac{k}{\mu}\right)^k \left(1 + \frac{k}{\mu}\right)^{-(k+y)} \int_0^\infty e^{-t} t^{k+y-1} dt \\ &= \frac{\Gamma(y+k)\phi^k}{\Gamma(y+1)\Gamma(k)(1+\phi)^{y+k}} \\ &= \frac{\Gamma(y+k)}{\Gamma(y+1)\Gamma(k)} \left(\frac{\phi}{1+\phi}\right)^k \left(\frac{1}{1+\phi}\right)^y \\ &= \frac{\Gamma(y+k)}{\Gamma(y+1)\Gamma(k)} (1-\pi)^k \pi^y, \quad y = 0, 1, 2, \dots, \end{aligned}$$

em que $\pi = 1/(1 + \phi)$. Portanto, Y tem distribuição binomial negativa de média μ e parâmetro de forma k .

Podemos, similarmente, supor que $Y|z \sim P(z)$ e que $Z \sim G(\mu, \phi)$, em que ϕ não depende de μ . Nesse caso $E(Z) = \mu$ e $\text{Var}(Z) = \mu^2/\phi$ de onde segue que $E(Y) = \mu$ e $\text{Var}(Y) = \mu + \mu^2/\phi$. Temos então que

$$f(y|z) = \frac{e^{-z} z^y}{y!} \quad \text{e} \quad g(z; \mu, \phi) = \frac{1}{\Gamma(\phi)} \left(\frac{z\phi}{\mu}\right)^\phi e^{-\frac{\phi z}{\mu}} \frac{1}{z}.$$

A função de probabilidades de Y fica dada por

$$\begin{aligned} Pr\{Y = y\} &= \int_0^\infty f(y|z)g(z; \mu, \phi)dz \\ &= \frac{1}{y!\phi} \left(\frac{\phi}{\mu}\right)^\phi \int_0^\infty e^{-z(1+\phi/\mu)} z^{\phi+y-1} dz. \end{aligned}$$

Fazendo a transformação de variável $t = z(1 + \frac{\phi}{\mu})$ temos que $\frac{dz}{dt} = (1 + \frac{\phi}{\mu})^{-1}$.

Daí segue que

$$\begin{aligned}
 \Pr\{Y = y\} &= \frac{1}{y!\Gamma(\phi)} \left(\frac{\phi}{\mu}\right)^\phi \left(1 + \frac{\phi}{\mu}\right)^{-(\phi+y)} \int_0^\infty e^{-t} t^{\phi+y-1} dt \\
 &= \frac{\Gamma(\phi + y) \mu^y \phi^\phi}{\Gamma(\phi) \Gamma(y + 1) (\mu + \phi)^{\phi+y}} \\
 &= \frac{\Gamma(\phi + y)}{\Gamma(y + 1) \Gamma(\phi)} \left(\frac{\mu}{\mu + \phi}\right)^y \left(\frac{\phi}{\mu + \phi}\right)^\phi \\
 &= \frac{\Gamma(\phi + y)}{\Gamma(y + 1) \Gamma(\phi)} (1 - \pi)^\phi \pi^y, \quad y = 0, 1, 2, \dots,
 \end{aligned}$$

com $\pi = \mu/(\mu + \phi)$. Portanto, neste caso Y também segue distribuição binomial negativa de média μ e parâmetro de forma ϕ . Denotamos $Y \sim \text{BN}(\mu, \phi)$. Podemos mostrar (ver, por exemplo, Jørgensen, 1996, p. 96) que

$$\frac{1}{\sqrt{\phi}}(Y - \mu) \rightarrow_d N(0, \pi/(1 - \pi)^2), \quad \text{quando } \phi \rightarrow \infty.$$

Podemos obter também aproximações da binomial negativa para a Poisson e gama.

4.3.2 Modelos de regressão com resposta binomial negativa

Vamos supor então que Y_1, \dots, Y_n são variáveis aleatórias independentes tais que $Y_i \sim \text{BN}(\mu_i, \phi)$. A função de probabilidades de Y_i fica dada por

$$f(y_i; \mu_i, \phi) = \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1) \Gamma(\phi)} \left(\frac{\mu_i}{\mu_i + \phi}\right)^{y_i} \left(\frac{\phi}{\mu_i + \phi}\right)^\phi, \quad y_i = 0, 1, 2, \dots$$

Temos que $E(Y_i) = \mu_i$ e $\text{Var}(Y_i) = \mu_i + \mu_i^2/\phi$. Similarmente aos MLGs assumimos parte sistemática dada por $g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, em que $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ contém valores de variáveis explicativas, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é um vetor de parâmetros desconhecidos e $g(\cdot)$ é a função de ligação. Como

nos modelos de Poisson as ligações mais utilizados são logarítmica ($g(\mu_i) = \log \mu_i$), raiz quadrada ($g(\mu_i) = \sqrt{\mu_i}$) e identidade ($g(\mu_i) = \mu_i$).

Definindo $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ o logaritmo da função de verossimilhança fica dado por

$$L(\boldsymbol{\theta}) = \sum_{i=1}^n \left[\log \left\{ \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \right\} + \phi \log \phi + y_i \log \mu_i - (\phi + y_i) \log(\mu_i + \phi) \right],$$

em que $\mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$. A fim de obtermos a função escore para $\boldsymbol{\beta}$ calculamos inicialmente as derivadas

$$\begin{aligned} \partial L(\boldsymbol{\theta}) / \partial \beta_j &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{d\mu_i}{d\eta_i} x_{ij} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{\phi(d\mu_i/d\eta_i)}{\mu_i(\phi + \mu_i)} (y_i - \mu_i) x_{ij} \right\} \\ &= \sum_{i=1}^n \omega_i f_i^{-1} (y_i - \mu_i) x_{ij}, \end{aligned}$$

em que $\omega_i = (d\mu_i/d\eta_i)^2 / (\mu_i^2 \phi^{-1} + \mu_i)$ e $f_i = d\mu_i/d\eta_i$. Logo, podemos expressar a função escore na forma matricial

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \mathbf{X}^T \mathbf{W} \mathbf{F}^{-1} (\mathbf{y} - \boldsymbol{\mu}), \quad (4.5)$$

em que \mathbf{X} é a matriz modelo com linhas \mathbf{x}_i^T , $i = 1, \dots, n$, $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$, $\mathbf{F} = \text{diag}\{f_1, \dots, f_n\}$, $\mathbf{y} = (y_1, \dots, y_n)^T$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$. De forma similar a função escore para ϕ fica dada por

$$U_\phi(\boldsymbol{\theta}) = \sum_{i=1}^n [\psi(\phi + y_i) - \psi(\phi) - (y_i + \phi) / (\phi + \mu_i) + \log\{\phi / (\phi + \mu_i)\} + 1], \quad (4.6)$$

em que $\psi(\cdot)$ é a função digama.

Para obtermos a matriz de informação de Fisher calculamos as derivadas

$$\begin{aligned}\partial^2 L(\boldsymbol{\theta})/\partial\beta_j\partial\beta_\ell &= -\sum_{i=1}^n \left\{ \frac{(\phi + y_i)}{(\phi + \mu_i)^2} - \frac{y_i}{\mu_i^2} \right\} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij}x_{i\ell} \\ &\quad + \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \right\} \frac{d^2\mu_i}{d\eta_i^2} x_{ij}x_{i\ell},\end{aligned}$$

cujos valores esperados ficam dados por

$$\begin{aligned}E\{\partial^2 L(\boldsymbol{\theta})/\partial\beta_j\partial\beta_\ell\} &= -\sum_{i=1}^n \frac{\phi(d\mu_i/d\eta_i)^2}{(\phi + \mu_i)} x_{ij}x_{i\ell} \\ &= -\sum_{i=1}^n \omega_i x_{ij}x_{i\ell}.\end{aligned}$$

Logo, podemos expressar a informação de Fisher para $\boldsymbol{\beta}$ em forma matricial

$$\mathbf{K}_{\beta\beta}(\boldsymbol{\theta}) = E\left\{-\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right\} = \mathbf{X}^T \mathbf{W} \mathbf{X}.$$

Lawless(1987) mostra que a informação de Fisher para ϕ pode ser expressa na forma

$$K_{\phi\phi}(\boldsymbol{\theta}) = \sum_{i=1}^n \left\{ \sum_{j=0}^{\infty} (\phi + j)^{-2} Pr(Y_i \geq j) - \phi^{-1} \mu_i / (\mu_i + \phi) \right\},$$

e que $\boldsymbol{\beta}$ e ϕ são parâmetros ortogonais. Assim, a matriz de informação de Fisher para $\boldsymbol{\theta}$ assume a forma bloco diagonal

$$\mathbf{K}_{\theta\theta} = \begin{bmatrix} \mathbf{K}_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & K_{\phi\phi} \end{bmatrix}.$$

As estimativas de máxima verossimilhança para $\boldsymbol{\beta}$ e ϕ podem ser obtidas através de um algoritmo de mínimos quadrados ponderados, aplicando o método score de Fisher, a partir de (4.5) e do método de Newton-Raphson para obter $\hat{\phi}$ desenvolvido a partir de (4.6), os quais são descritos abaixo

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{y}^{*(m)}$$

e

$$\phi^{(m+1)} = \phi^{(m)} - \{\mathbf{U}_{\phi}^{(m)} / \ddot{\mathbf{L}}_{\phi\phi}^{(m)}\},$$

para $m = 0, 1, 2, \dots$, em que

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{F}^{-1}(\mathbf{y} - \boldsymbol{\mu})$$

é uma variável dependente modificada e

$$\ddot{\mathbf{L}}_{\phi\phi} = \sum_{i=1}^n \{\psi'(\phi + y_i) + (y_i - 2\mu_i - \phi)/(\phi + \mu_i)^2\} + n\phi^{-1}\{1 - \phi\psi'(\phi)\}.$$

Tabela 4.5

Quantidades ω_i e f_i para algumas ligações.

Ligação	ω_i	f_i
$\log \mu_i = \eta_i$	$\mu_i / (\mu_i \phi^{-1} + 1)$	μ_i
$\mu_i = \eta_i$	$(\mu_i^2 \phi^{-1} + \mu_i)^{-1}$	1
$\sqrt{\mu_i} = \eta_i$	$4 / (\mu_i \phi^{-1} + 1)$	$2\sqrt{\mu_i}$

Os dois procedimentos são aplicados simultaneamente até a convergência. Podemos encontrar as estimativas de máxima verossimilhança $(\hat{\boldsymbol{\beta}}^T, \hat{\phi})^T$ pela aplicação do comando `library(MASS)` do R. Como ilustração, vamos supor um modelo log-linear com resposta binomial negativa `resp` e covariáveis `cov1` e `cov2`. Devemos acionar os seguintes comandos no R:

```
library(MASS)
fit.bn = glm.nb( resp ~ cov1 + cov2 ).
```

No objeto `fit.bn` estarão os resultados do ajuste. Outras ligações, além da ligação logarítmica, podem ser usadas com a distribuição binomial negativa. Por exemplo, para o ajuste de um modelo com resposta binomial negativa e ligação identidade se `resp` é considerada resposta e `cov1` e `cov2` são consideradas variáveis explicativas, devemos fazer o seguinte:

```
library(MASS)
```

`fit.bn = glm.nb(resp ~ cov1 + cov2, link=identity).`

A Tabela 4.5 apresenta as expressões para ω_i e f_i para algumas ligações usuais em modelos com resposta binomial negativa.

Usando os mesmos argumentos da Seção 1.6 temos que para n grande $\hat{\beta}$ segue distribuição aproximadamente normal p -variada de média β e matriz de variância-covariância $\mathbf{K}_{\beta\beta}^{-1}$, ou seja, para n grande $\hat{\beta} \sim N_p(\beta, \mathbf{K}_{\beta\beta}^{-1})$. Similarmente para n grande $\hat{\phi} \sim N(\phi, K_{\phi\phi}^{-1})$. Além disso, $\hat{\beta}$ e $\hat{\phi}$ são assintoticamente independentes.

4.3.3 Qualidade do ajuste

A função desvio assumindo ϕ fixo fica dada por

$$D^*(\mathbf{y}; \hat{\mu}) = 2 \sum_{i=1}^n \left[\phi \log \left\{ \frac{\hat{\mu}_i + \phi}{y_i + \phi} \right\} + y_i \log \left\{ \frac{y_i(\hat{\mu}_i + \phi)}{\hat{\mu}_i(y_i + \phi)} \right\} \right],$$

em que $\hat{\mu}_i = g^{-1}(\mathbf{x}_i^T \beta)$. Quando $y_i = 0$ o i -ésimo componente da função desvio $D^*(\mathbf{y}; \hat{\mu})$ fica dado por

$$\begin{aligned} d^{*2}(y_i, \hat{\mu}_i) &= 2\{\log f(0; y_i, \phi) - \log f(0; \hat{\mu}_i, \phi)\} \\ &= 2\phi \log\{\phi/(y_i + \phi)\} - 2\phi \log\{\phi/(\hat{\mu}_i + \phi)\} \\ &= 2\phi \log\{(\mu_i + \phi)/(\hat{\mu}_i + \phi)\} \\ &= 2\phi \log\{(\hat{\mu}_i + \phi)/\phi\}. \end{aligned}$$

Portanto, os componentes do desvio no caso binomial negativo assumem as seguintes formas:

$$d^{*2}(y_i; \hat{\mu}_i) = \begin{cases} 2 \left[\phi \log \left\{ \frac{(\hat{\mu}_i + \phi)}{(y_i + \phi)} \right\} + y_i \log \left\{ \frac{y_i(\hat{\mu}_i + \phi)}{\hat{\mu}_i(y_i + \phi)} \right\} \right] & \text{se } y_i > 0; \\ 2\phi \log \left\{ \frac{(\hat{\mu}_i + \phi)}{\phi} \right\} & \text{se } y_i = 0. \end{cases}$$

Sob a hipótese de que o modelo adotado está correto $D^*(\mathbf{y}; \hat{\mu})$ segue para ϕ grande e μ_i grande, $\forall i$, uma distribuição qui-quadrado com $(n - p)$ graus de liberdade.

Vamos supor agora a partição $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$ em que $\boldsymbol{\beta}_1$ é um vetor q -dimensional enquanto $\boldsymbol{\beta}_2$ tem dimensão $p - q$ e que ϕ é fixo ou conhecido. O teste da razão de verossimilhanças para testarmos $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ contra $H_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}$ reduz, neste caso, à diferença entre dois desvios

$$\xi_{RV} = D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}),$$

em que $\hat{\boldsymbol{\mu}}^0$ e $\hat{\boldsymbol{\mu}}$ são, respectivamente, as estimativas de $\boldsymbol{\mu}$ sob H_0 e H_1 . Para ϕ desconhecido o teste da razão de verossimilhanças fica expresso na seguinte forma:

$$\begin{aligned} \xi_{RV} = & 2 \sum_{i=1}^n [\log\{\Gamma(\hat{\phi} + y_i)\Gamma(\hat{\phi}^0)/\Gamma(\hat{\phi}^0 + y_i)\Gamma(\hat{\phi})\} + \hat{\phi} \log\{\hat{\phi}/(\hat{\phi} + \hat{\mu}_i)\} \\ & - \hat{\phi}^0 \log\{\hat{\phi}^0/(\hat{\phi}^0 + \hat{\mu}_i^0)\} + y_i \log\{\hat{\mu}_i(\hat{\phi}^0 + \hat{\mu}_i^0)/\hat{\mu}_i^0(\hat{\phi} + \hat{\mu}_i)\}], \end{aligned}$$

em que $\hat{\phi}^0$ e $\hat{\phi}$ são as estimativas de máxima verossimilhança de ϕ sob H_0 e H_1 , respectivamente. Para n grande e sob H_0 temos que $\xi_{RV} \sim \chi_q^2$.

4.3.4 Técnicas de diagnóstico

Fazendo uma analogia com os MLGs a matriz de projeção \mathbf{H} assume aqui a seguinte forma:

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}.$$

O i -ésimo elemento da diagonal principal de \mathbf{H} fica dado por

$$h_{ii} = \frac{(d\mu_i/d\eta_i)^2}{(\mu_i\phi^{-1} + \mu_i)} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i.$$

Em particular, para os modelos log-lineares h_{ii} fica dado por

$$h_{ii} = \frac{\phi\mu_i}{(\phi + \mu_i)} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i,$$

em que $\omega_i = \phi\mu_i/(\phi + \mu_i)$. Como \hat{h}_{ii} deverá depender de $\hat{\mu}_i$, gráficos de \hat{h}_{ii} contra os valores ajustados são mais informativos do que os gráficos de \hat{h}_{ii} contra a ordem das observações.

Estudos de Monte Carlo desenvolvidos por Svetliza (2002) (ver também Svetliza e Paula, 2003) indicam boa concordância entre o resíduo componente do desvio

$$t_{D_i} = \frac{d^*(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}}$$

com a distribuição normal padrão, em que

$$d^*(y_i; \hat{\mu}_i) = \begin{cases} \pm\sqrt{2} \left[\phi \log \left\{ \frac{\hat{\mu}_i + \phi}{y_i + \phi} \right\} + y_i \log \left\{ \frac{y_i(\hat{\mu}_i + \phi)}{\hat{\mu}_i(y_i + \phi)} \right\} \right]^{1/2} & \text{se } y_i > 0; \\ \pm\sqrt{2} \left[\phi \log \left\{ \frac{(\hat{\mu}_i + \phi)}{\phi} \right\} \right] & \text{se } y_i = 0. \end{cases}$$

Para extrairmos a quantidade $d_i^*(y_i; \hat{\mu}_i)$ do objeto `fit.bn` devemos fazer o seguinte:

```
d = resid(fit.bn, type= "deviance").
```

Uma versão da distância de Cook aproximada é dada por

$$LD_i = \frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})^2} \hat{r}_{P_i}^2,$$

em que $r_{P_i} = (y_i - \mu_i) / \sqrt{\text{Var}(Y_i)}$ e $\text{Var}(Y_i) = \mu_i + \mu_i^2 / \phi$. A quantidade r_{P_i} é obtida no R através do comando

```
rp = resid(fit.bn, type="pearson").
```

O gráfico de LD_i contra as observações ou valores ajustados pode revelar pontos influentes nas estimativas $\hat{\beta}$ e $\hat{\phi}$. Svetliza (2002) desenvolveu as expressões matriciais para a obtenção de ℓ_{max} para $\hat{\beta}$ e $\hat{\phi}$.

4.3.5 Seleção de modelos

Similarmente aos modelos lineares generalizados, pelo critério de Akaike devemos encontrar um submodelo para o qual a quantidade abaixo seja minimizada

$$AIC = D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) + 2p.$$

Devemos acionar no R os seguintes comandos:

```
library(MASS)
stepAIC(fit.model).
```

Aqui `fit.model` denota o objeto com o modelo binomial negativo ajustado.

4.3.6 Aplicações

Estudantes australianos

Venables e Ripley(1999, Caps. 6 e 7) apresentam os resultados de um estudo sociológico desenvolvido na Austrália com 146 estudantes de 8^a série e ensino médio com o objetivo de comparar a ausência na escola segundo os seguintes fatores: ano que o estudante está cursando (1: 8^a série, 2: 1^o ano do ensino médio, 3: 2^o ano do ensino médio, 4: 3^o ano do ensino médio), etnia (0: aborígine, 1: não aborígine), desempenho escolar (0: insuficiente, 1: suficiente) e sexo (0: masculino, 1: feminino). Para obtermos esses dados no R devemos acionar o comando `library(MASS)` e em seguida `quine`. Uma cópia desses dados está disponível no arquivo **quine.dat**. Denotamos por Y_{ijklm} o número de faltas num determinado período referentes ao m -ésimo aluno, cursando o i -ésimo ano, de etnia j , com desempenho escolar k e pertencente ao ℓ -ésimo sexo, em que $i = 1, 2, 3, 4$, $j, k, \ell = 1, 2$ e $m = 1, \dots, 144$. Vamos supor que $Y_{ijklm} \sim \text{BN}(\mu_{ijkl}, \phi)$, em que

$$\log \mu_{ijkl} = \alpha + \beta_i + \gamma_j + \delta_k + \theta_\ell,$$

com $\beta_1 = 0$, $\gamma_1 = 0$, $\delta_1 = 0$ e $\theta_1 = 0$. Assim, temos um modelo casela de referência em que β_2 , β_3 e β_4 denotam os incrementos do primeiro, segundo e terceiro ano do ensino médio, respectivamente, em relação à 8^a série, γ_2 é a diferença entre os efeitos do grupo não aborígine com relação ao grupo

aborígine, δ_2 denota a diferença entre os efeitos dos grupos com desempenho suficiente e insuficiente e θ é a diferença entre os efeitos do sexo feminino e masculino.

Tabela 4.6

Estimativas de máxima verossimilhança referentes ao modelo log-linear binomial negativo ajustado aos dados sobre ausência escolar de estudantes australianos.

Efeito	Modelo 1	E/E.Padrão	Modelo 2	E/E.Padrão
Intercepto	2,895	12,70	2,628	10,55
Etnia	-0,569	-3,72	0,131	0,38
Sexo	0,082	0,51		
Ano2	-0,448	-1,87	0,178	0,56
Ano3	0,088	0,37	0,827	2,61
Ano4	0,357	1,44	0,371	1,11
Desemp	0,292	1,57		
Etn*Ano2			-0,991	-2,26
Etn*Ano3			-1,239	-2,78
Etn*Ano4			-0,176	-0,38
ϕ	1,275	7,92	1,357	7,80

Na Tabela 4.6 temos as estimativas de máxima verossimilhança com os respectivos erros padrão aproximados. O desvio do modelo ajustado (modelo 1) foi de $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 167,95$ (139 graus de liberdade). Notamos que os fatores sexo e desempenho escolar não são significativos a 10%, sendo portanto retirados do modelo. Contudo, notamos a necessidade de inclusão da interação ano*etnia no novo modelo. O valor da estatística da razão de verossimilhanças nesse caso é de $\xi_{RV} = 11,16$ ($P = 0,0109$). As novas estimativas são também apresentadas na Tabela 4.6. O desvio do novo modelo (modelo 2) foi de $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 167,84$ (138 graus de liberdade). A Figura 4.5 apresenta as médias ajustadas do modelo final. Podemos notar que o grupo não aborígine tem em geral um n^o médio menor de dias ausentes. A maior média é observada para estudantes do grupo aborígine cursando o 2^o do ensino médio

e o menor valor médio é observado para estudantes do grupo não aborígene cursando o 1º do ensino médio.

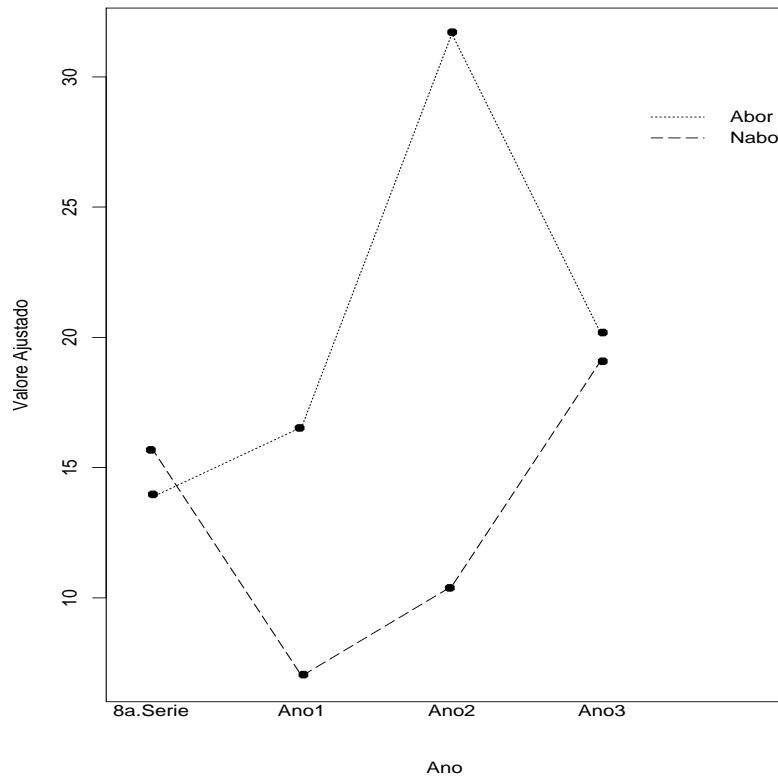


Figura 4.5: Valores médios estimados pelo modelo log-linear binomial negativo ajustado aos dados sobre ausência escolar de estudantes australianos.

Verificamos também, neste estudo, como fica o ajuste através de um modelo log-linear de Poisson. Temos nas Figura 4.6a e 4.6b os gráficos normais de probabilidades para os dois ajustes e notamos uma clara superioridade do modelo log-linear com resposta binomial negativa. O modelo log-linear de Poisson apresenta fortes indícios de sobredispersão com os resíduos cruzando o envelope gerado. Isso é justificado pelo valor do desvio $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 1597,11$

(138 graus de liberdade).

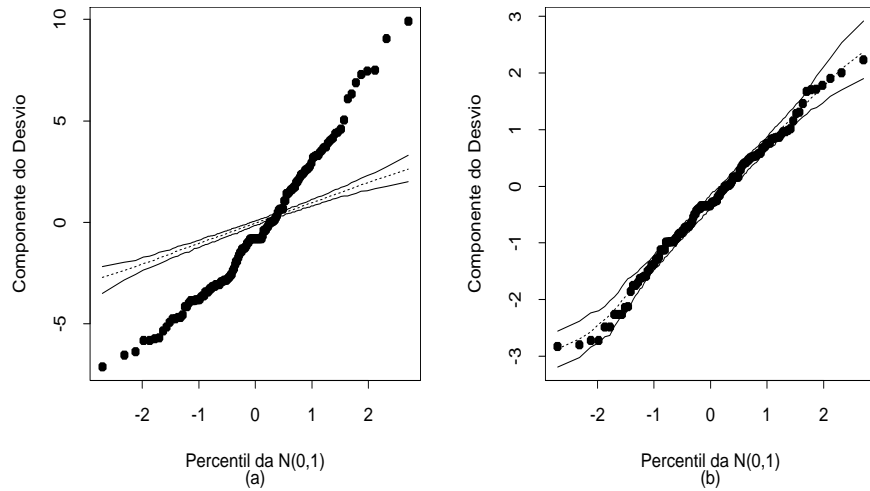


Figura 4.6: Gráficos normais de probabilidades referentes ao modelo log-linear de Poisson (a) e ao modelo log-linear binomial negativo (b) ajustados aos dados sobre ausência escolar de estudantes australianos.

Nas Figuras 4.7a a 4.7d são apresentados alguns gráficos de diagnóstico. Na Figura 4.7a em que são apresentados os valores de \hat{h}_{ii} nenhum dos 8 grupos formados são destacados como alavanca. Já pela Figura 4.7b notamos pelo menos três pontos com mais destaque como influentes em $\hat{\beta}$, são os alunos #72, #104 e #36. Os três alunos têm vários dias ausentes, respectivamente, 67, 69 e 45. O aluno #72 é não aborígene e estava cursando a 8ª série. O aluno #104 é também não aborígene, porém estava cursando o 3º ano, enquanto o aluno #36 é aborígene e estava também cursando a 8ª série. Pela Figura 4.7c notamos dois pontos com mais destaque como aberrantes, #98 e #61. Esses alunos não tiveram faltas, estavam cursando o 3º ano, um é aborígene (#61) e o outro (#98) é não aborígene. Em geral os pontos aberrantes desse exemplo referem-se a alunos sem nenhuma falta. A retirada desses pontos

não altera os resultados inferenciais. Finalmente, a Figura 4.7d indica que a escolha da ligação logarítmica não parece ser inadequada.

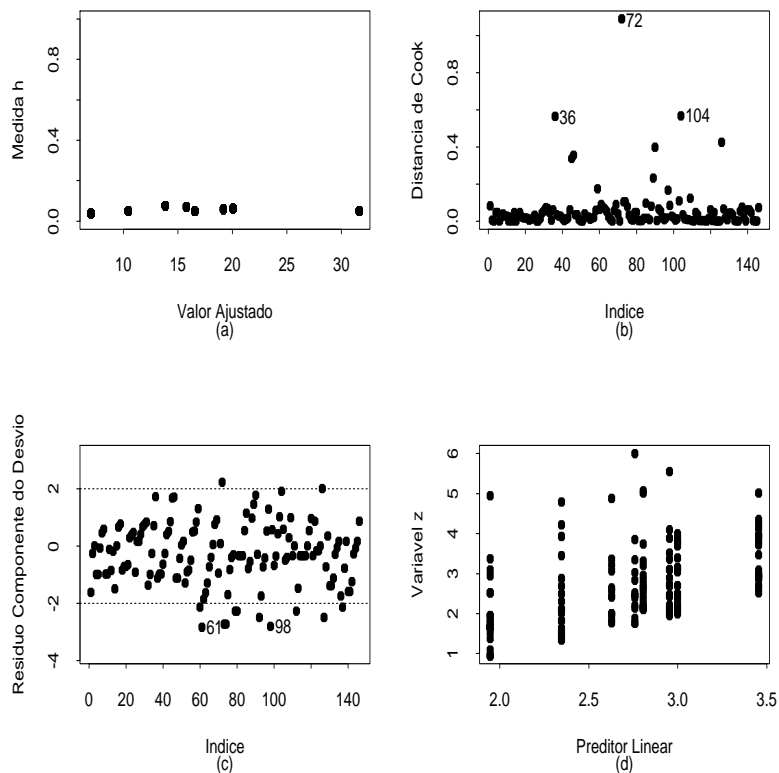


Figura 4.7: Gráficos de diagnóstico referentes ao modelo binomial negativo ajustado aos dados sobre ausência escolar de estudantes australianos.

Demanda de TV a cabo

Na Tabela 4.7 é apresentado um conjunto de dados sobre a demanda de TVs a cabo em 40 áreas metropolitanas dos EUA (Ramanathan, 1993). Esses dados estão também disponíveis no arquivo **tv cabo.dat**. Foram observadas, para cada área, o número de assinantes (em milhares) de TV a cabo (nass),

Tabela 4.7*Demanda de TV a cabo em 40 áreas metropolitanas dos EUA.*

Nass	Domic	Perc	Percap	Taxa	Custo	Ncabo	Ntv
105	350	30,000	9839	14,95	10	16	13
90	255,631	35,207	10606	15	7,5	15	11
14	31	45,161	10455	15	7	11	9
11,7	34,840	33,582	8958	10	7	22	10
46	153,434	29,980	11741	25	10	20	12
11,217	26,621	42,136	9378	15	7,66	18	8
12	18	66,667	10433	15	7,5	12	8
6,428	9,324	68,940	10167	15	7	17	7
20,1	32	62,813	9218	10	5,6	10	8
8,5	28	30,357	10519	15	6,5	6	6
1,6	8	20,000	10025	17,5	7,5	8	6
1,1	5	22,000	9714	15	8,95	9	9
4,355	15,204	28,644	9294	10	7	7	7
78,910	97,889	80,612	9784	24,95	9,49	12	7
19,6	93	21,075	8173	20	7,5	9	7
1	3	33,333	8967	9,95	10	13	6
1,65	2,6	63,462	10133	25	7,55	6	5
13,4	18,284	73,288	9361	15,5	6,3	11	5
18,708	55	34,015	9085	15	7	16	6
1,352	1,7	79,529	10067	20	5,6	6	6
170	270	62,963	8908	15	8,75	15	5
15,388	46,540	33,064	9632	15	8,73	9	6
6,555	20,417	32,106	8995	5,95	5,95	10	6
40	120	33,333	7787	25	6,5	10	5
19,9	46,39	42,897	8890	15	7,5	9	7
2,45	14,5	16,897	8041	9,95	6,25	6	4
3,762	9,5	39,600	8605	20	6,5	6	5
24,882	81,98	30,351	8639	18	7,5	8	4
21,187	39,7	53,368	8781	20	6	9	4
3,487	4,113	84,780	8551	10	6,85	11	4
3	8	37,500	9306	10	7,95	9	6
42,1	99,750	42,206	8346	9,95	5,73	8	5
20,350	33,379	60,966	8803	15	7,5	8	4
23,15	35,5	65,211	8942	17,5	6,5	8	5
9,866	34,775	28,371	8591	15	8,25	11	4
42,608	64,840	65,713	9163	10	6	11	6
10,371	30,556	33,941	7989	20	7,5	8	6
5,164	16,5	31,297	7924	14,95	6,95	8	5
31,150	70,515	44,175	8454	9,95	7	10	4
18,350	42,040	43,649	8429	20	7	6	4

o número de domicílios (em milhares) na área (domic), a porcentagem de domicílios com TV a cabo (perc), a renda per capita (em USD) por domicílio com TV a cabo (percap), a taxa de instalação de TV a cabo (taxa) em USD, o custo médio mensal de manutenção de TV a cabo (custo) em USD, o número de canais a cabo disponíveis na área (ncabo) e o número de canais não pagos com sinal de boa qualidade disponíveis na área (ntv).

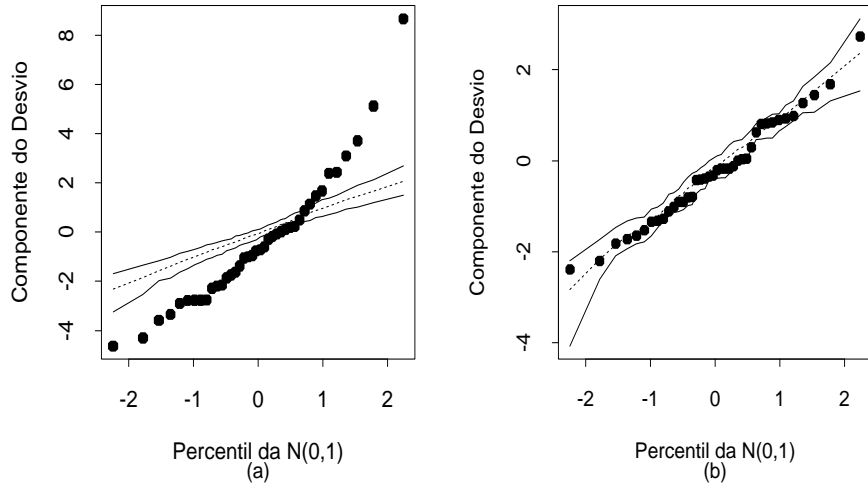


Figura 4.8: Gráficos normais de probabilidades referentes aos modelos log-linear de Poisson(a) e log-linear binomial negativo (b) ajustados aos dados sobre demanda de TV a cabo.

Como são dados de contagem podemos pensar inicialmente num modelo de Poisson em que $nass_i$ denota o número de assinantes na i -ésima região tal que $nass_i \sim P(\mu_i)$, em que

$$\log \mu_i = \alpha + \beta_1 \text{domic}_i + \beta_2 \text{percap}_i + \beta_3 \text{taxa}_i + \beta_4 \text{custo}_i + \beta_5 \text{ncabo}_i + \beta_6 \text{ntv}_i$$

para $i = 1, \dots, 40$. No entanto, o ajuste do modelo forneceu desvio $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 225$ para 33 graus de liberdade indicando fortes indícios de sobredispersão,

que é confirmado pelo gráfico normal de probabilidades da Figura 4.8a. Tentamos então um modelo binomial negativo em que $\text{nass}_i \sim \text{BN}(\mu_i, \phi)$. O gráfico normal de probabilidades (Figura 4.8b) bem como o desvio $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 42,35$ fornecem indícios de ajuste adequado. No entanto, pela Figura 4.9, notamos uma área altamente influente (observação #14) e outra área com moderada influência (observação #1). A área #14 apresenta custos altos de instalação e manutenção de TV a cabo, porém um alto índice de assinantes.

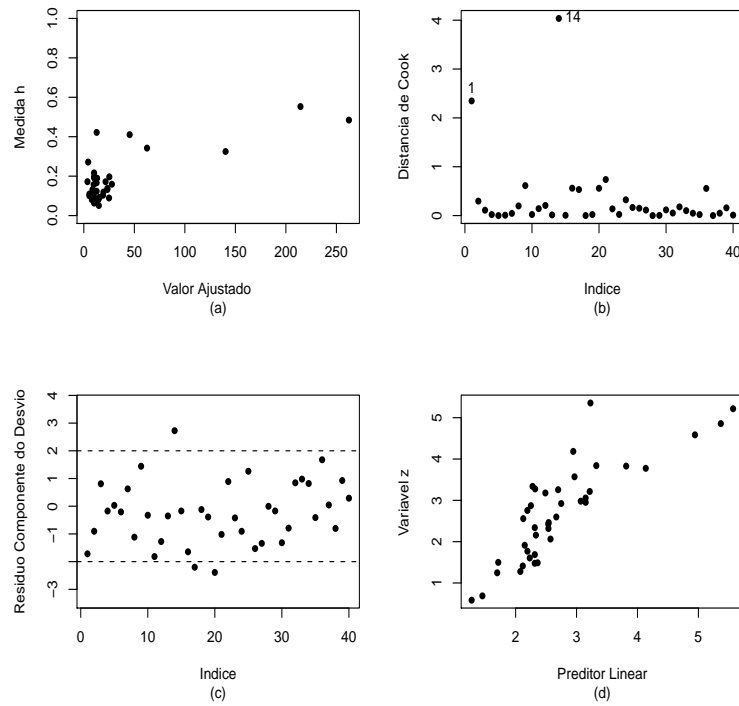


Figura 4.9: Gráficos de diagnóstico referentes ao modelo log-linear binomial negativo ajustado aos dados sobre demanda de TV a cabo.

A área #1 tem um baixo índice de assinantes com grande oferta de canais

a cabo e canais não pagos de boa qualidade. As estimativas dos coeficientes com todos os pontos e eliminando as observações mais discrepantes (1 e 14) são apresentadas na Tabela 4.8. Como podemos observar há indícios de que quatro coeficientes (percap, taxa, ncabo e ntv) são marginalmente não significativos a 10%. Aplicamos então o teste da razão de verossilhanças para testarmos $H_0 : \beta_2 = \beta_3 = \beta_5 = \beta_6 = 0$ contra pelo menos um parâmetro diferente de zero que forneceu o valor $\xi_{RV} = 2,50$ para 4 graus de liberdade ($P=0,64$), indicando pela não rejeição da hipótese nula. Isso significa que as duas observações discrepantes são responsáveis pela significância de três desses coeficientes que aparecem significativos marginalmente com todos os pontos, bem como pelo aumento da sobredispersão uma vez que a estimativa de ϕ cresce com a eliminação das duas áreas.

Tabela 4.8

Estimativas de máxima verossilhança referentes do modelo log-linear binomial negativo ajustado aos dados sobre demanda de TV a cabo.

Efeito	Todos pontos	E/E.Padrão	Sem 1 e 14	E/E.Padrão
Intercepto	2,437	1,99	3,608	3,34
Domic	0,013	8,24	0,014	9,69
Percap	6×10^{-5}	0,42	-2×10^{-6}	-0,01
Taxa	0,041	1,84	0,010	0,50
Custo	-0,207	1,95	-0,266	-2,69
Ncabo	0,067	2,01	0,050	1,63
Ntv	-0,135	1,84	-0,071	-1,02
ϕ	3,311	3,49	5,060	2,89

Portanto, um modelo mais recomendado envolveria apenas as variáveis explicativas domic e custo, ou seja, com parte sistemática dada por $\log \mu_i = \alpha + \beta_1 \text{domic}_i + \beta_4 \text{custo}_i$. O desvio desse modelo fica dado por $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 41,05$ para 35 graus de liberdade ($P= 0,22$), indicando um ajuste adequado. As novas estimativas sem as áreas #1 e # 14 (erro padrão aproximado) ficam dadas por $\hat{\alpha} = 3,620(0,637)$, $\hat{\beta}_1 = 0,015(0,001)$, $\hat{\beta}_4 = -0,242(0,091)$ e

$$\hat{\phi} = 4,54(1,51).$$

Portanto, pela estimativa de β_4 , se o custo mensal de manutenção aumentar de USD 1 esperamos uma redução aproximada de 22% no número de assinantes com TV a cabo na área.

4.3.7 Sobredispersão e quase-verossimilhança

De uma forma geral o fenômeno de sobredispersão sugere que a variância de Y seja dada por $\text{Var}(Y) = \sigma^2 \mu$, em que $\sigma^2 > 1$. Uma maneira mais simples de resolvermos o problema é ajustarmos um modelo log-linear de Poisson aos dados e estimarmos σ^2 separadamente (método de quase-verossimilhança), por exemplo, usando a estimativa proposta por Wedderburn (1974), dada por

$$\hat{\sigma}^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} / (n - p), \quad (4.7)$$

em que $\hat{\mu}_i = \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})$. Algumas quantidades, tais como a matriz de variância-covariância assintótica de $\hat{\boldsymbol{\beta}}$, o desvio, resíduos etc, deverão ser corrigidos de maneira similar ao caso tratado na Seção 3.6.14. Finalmente, podemos pensar na aplicação de modelos mais gerais de quase-verossimilhança que serão discutidos no Capítulo 5.

Aplicação

Como ilustração, vamos considerar os dados descritos na Tabela 4.9 (McCullagh e Nelder, 1989, Seção 6.3.2) e também no arquivo **navios.dat** em que avarias causadas por ondas em navios de carga são classificadas segundo o tipo do navio (A-E), ano da fabricação (1:1960-64, 2:1965-69, 3:1970-74 e 4:1975-79) e período de operação (1:1960-74 e 2:1975-79).

Tabela 4.9

Distribuição de avarias em navios de carga segundo o tipo do navio, ano de fabricação período de operação e total de meses em operação.

Tipo	Ano	Período	Meses	Avarias
A	1	1	127	0
A	1	2	63	0
A	2	1	1095	3
A	2	2	1095	4
A	3	1	1512	6
A	3	2	3353	18
A	4	2	2244	11
B	1	1	44882	39
B	1	2	17176	29
B	2	1	28609	58
B	2	2	20370	53
B	3	1	7064	12
B	3	2	13099	44
B	4	2	7117	18
C	1	1	1179	1
C	1	2	552	1
C	2	1	781	0
C	2	2	676	1
C	3	1	783	6
C	3	2	1948	2
C	4	2	274	1
D	1	1	251	0
D	1	2	105	0
D	2	1	288	0
D	2	2	192	0
D	3	1	349	2
D	3	2	1208	11
D	4	2	2051	4
E	1	1	45	0
E	2	1	789	7
E	2	2	437	7
E	3	1	1157	5
E	3	2324	2161	12
E	4	2	542	1

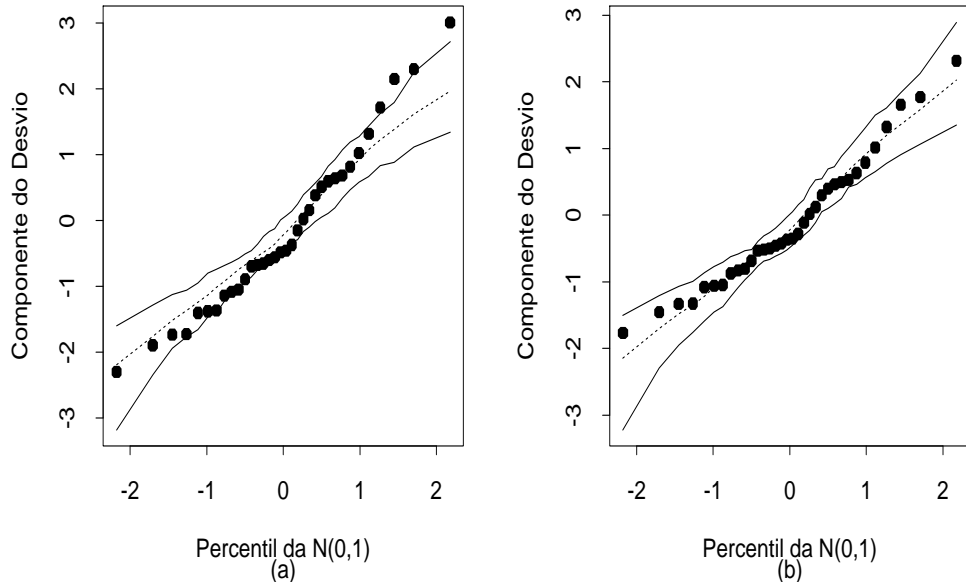


Figura 4.10: Gráficos normais de probabilidades referentes ao modelo log-linear de Poisson (a) e log-linear de quase-verossimilhança (b) ajustados aos dados sobre avarias em navios de carga.

Foi também considerado o tempo em que cada navio ficou em operação (em meses). Inicialmente, sugerimos um modelo log-linear de Poisson com *offset* dado por $\log(\text{meses})$ e efeitos principais. Assim, denotamos por Y_{ijk} o número de avarias observadas para o navio do tipo i , construído no ano j que operou no período k e vamos supor que $Y_{ijk} \sim P(\lambda_{ijk}t_{ijk})$, em que t_{ijk} é o total de meses de operação e λ_{ijk} o número médio esperado de avarias por unidade de tempo. A parte sistemática do modelo é dada por

$$\log \lambda_{ijk} = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)},$$

com as restrições $\beta_{1(1)} = \beta_{2(1)} = \beta_{3(1)} = 0$, para $i = 1, \dots, 5$; $j = 1, \dots, 4$ e $k = 1, 2$, com β_1 , β_2 e β_3 denotando, respectivamente, o efeito de tipo, de ano de construção e período de operação. O desvio do modelo foi de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 38,69$ (25 graus de liberdade) que corresponde a um nível descritivo $P = 0,040$, indicando que o ajuste não está satisfatório.

Pelo gráfico normal de probabilidades, descrito na Figura 4.10a, notamos a maioria dos resíduos próximos dos limites superior e inferior do envelope gerado, sugerindo sobredispersão que nesse caso deve ser devido ao fato de um mesmo navio ter sido observado mais de uma vez. Usando (4.7) obtemos $\hat{\sigma}^2 = 1,69$, e corrigindo o componente do desvio padronizado de modo que

$$t_{D_i}^* = \pm d_i / \hat{\sigma} \sqrt{1 - \hat{h}_{ii}},$$

obtemos um novo gráfico normal de probabilidades descrito na Figura 4.10b, em que os resíduos estão melhor distribuídos dentro do envelope gerado.

Tabela 4.10

*Estimativas dos parâmetros referentes ao modelo
log-linear de quase-verossimilhança ajustado
aos dados sobre avarias em navios de carga.*

Efeito	Estimativa	E/E.Padrão
Constante	-6,406	-22,69
Tipo		
A	0,000	-
B	-0,543	-2,36
C	-0,687	-1,61
D	-0,076	0,20
E	0,326	1,06
Ano		
60-64	0,000	-
65-69	0,697	3,59
70-74	0,818	3,71
75-79	0,453	1,50
Período		
60-74	0,000	-
75-79	0,384	2,50

O novo desvio fica dado por $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = D(\mathbf{y}; \hat{\boldsymbol{\mu}}) / \hat{\sigma}^2 = 38,69 / 1,69 = 22,89$ (25 graus de liberdade), indicando um ajuste adequado. Devemos observar

que tanto o resíduo $t_{D_i}^*$ como o desvio $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ devem ser olhados de maneira meramente descritiva uma vez que em modelos de quase-verossimilhança a distribuição da resposta é em geral desconhecida. As estimativas de máxima verossimilhança e os valores padronizados pelos respectivos erros padrão aproximados, já multiplicados pelo fator $\hat{\sigma}$, são apresentadas na Tabela 4.10. Williams (1987) mostra que o problema de sobredispersão neste exemplo é causado particularmente por duas observações discrepantes e sugere a inclusão da interação tipo*ano com pelo menos uma dessas observações excluídas. Pela Tabela 4.10 notamos que os navios de tipos B e C são aqueles com uma incidência menor de avarias por unidade de tempo. Por outro lado, os navios fabricados de 65 a 74 como também aqueles que operaram de 75 a 79 apresentam uma incidência maior de avarias por unidade de tempo do que os demais.

4.4 Relação entre a multinomial e a Poisson

Vamos supor agora que todas as unidades amostrais são acompanhadas durante o mesmo período e que são classificadas segundo s níveis de exposição e r grupos, conforme descrito abaixo.

Grupo	Exposição				
	E1	E2	E3	...	Es
G1	y_{11}	y_{12}	y_{13}	...	y_{1s}
G2	y_{21}	y_{22}	y_{23}	...	y_{2s}
		...			
Gr	y_{r1}	y_{r2}	y_{r3}	...	y_{rs}

Supondo que $Y_{ij} \sim P(\mu_{ij})$, $i = 1, \dots, r$ e $j = 1, \dots, s$, temos que

$$Pr\{\mathbf{Y} = \mathbf{a} | \sum_{i,j} Y_{ij} = n\} = \frac{n!}{\prod_{i,j} a_{ij}!} \prod_{i,j} \pi_{ij}^{a_{ij}},$$

em que $\pi_{ij} = \mu_{ij}/\mu_{++}$, $\mu_{++} = \sum_{i,j} \mu_{ij}$, $\mathbf{Y} = (Y_{11}, \dots, Y_{rs})^T$ e $\mathbf{a} = (a_{11}, \dots, a_{rs})^T$. Vamos considerar o modelo log-linear de Poisson com parte sistemática dada por $\log \mu_{ij} = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}$, com as restrições $\beta_{1(1)} = \beta_{2(1)} = \beta_{12(1j)} = \beta_{12(i1)} = 0$ para $i = 1, \dots, r$ e $j = 1, \dots, s$. Temos que

$$\begin{aligned} \tau = \mu_{++} &= \sum_{i=1}^r \sum_{j=1}^s \exp\{\alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}\} \\ &= e^\alpha \sum_{i=1}^r \sum_{j=1}^s \exp\{\beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}\}, \end{aligned}$$

e podemos definir as probabilidades

$$\pi_{ij} = \frac{\exp\{\beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}\}}{\sum_{i=1}^r \sum_{j=1}^s \exp\{\beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}\}},$$

em que o total do denominador é invariante com a parametrização utilizada no modelo. Temos que as probabilidades π'_{ij} s não dependem do parâmetro α . Como veremos a seguir, a estimativa de máxima verossimilhança do vetor $\boldsymbol{\beta}$ correspondente ao modelo multinomial coincide com a estimativa de máxima verossimilhança para $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_{12}^T)^T$ referente ao modelo log-linear de Poisson. Se, por exemplo, ajustarmos um modelo multinomial do tipo log-linear aos dados tal que

$$\log \pi_{ij} = \alpha^* + \beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)},$$

teremos, devido à imposição $\sum_{i,j} \pi_{ij} = 1$, que $\exp(\alpha^*) = 1 / \sum_i \sum_j \exp\{\beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}\}$, ou seja, $\alpha^* = \alpha - \log(\tau)$. O que muda é a estimativa do intercepto, embora na prática sempre seja possível obtermos α^* através de α e vice-versa. Para mostrarmos a equivalência das estimativas partiremos da relação abaixo

$$Pr\{\mathbf{Y} = \mathbf{a} | n\} = \frac{Pr\{\mathbf{Y} = \mathbf{a}; Y_{++} = n\}}{Pr\{Y_{++} = n\}},$$

em que $Y_{++} = \sum_{i,j} Y_{ij}$. Denotando $L_{\mathbf{y}|n}(\boldsymbol{\beta}) = \log Pr\{\mathbf{Y} = \mathbf{a}|n\}$, $L_{\mathbf{y}}(\tau, \boldsymbol{\beta}) = \log Pr\{\mathbf{Y} = \mathbf{a}; Y_{++} = n\}$ e $L_{y_{++}}(\tau) = \log Pr\{Y_{++} = n\}$ temos que

$$L_{\mathbf{y}}(\tau, \boldsymbol{\beta}) = L_{y_{++}}(\tau) + L_{\mathbf{y}|n}(\boldsymbol{\beta}), \quad (4.8)$$

em que

$$L_{y_{++}}(\tau) = -\tau + y_{++}\log\tau - \log(y_{++}!)$$

e

$$L_{\mathbf{y}|n}(\boldsymbol{\beta}) = \log n! + \sum_{i,j} a_{ij} \log \pi_{ij} - \sum_{i,j} \log a_{ij}!.$$

Portanto, maximizarmos $L_{\mathbf{y}}(\tau, \boldsymbol{\beta})$ com relação a $\boldsymbol{\beta}$ é equivalente a maximizarmos $L_{\mathbf{y}|n}(\boldsymbol{\beta})$ com relação a $\boldsymbol{\beta}$. Isso quer dizer que as estimativas de máxima verossimilhança para o vetor $\boldsymbol{\beta}$ são as mesmas sob o modelo log-linear multinomial com probabilidades $\pi_{11}, \dots, \pi_{rs}$ e sob o modelo log-linear de Poisson de médias $\mu_{11}, \dots, \mu_{rs}$. As matrizes de segundas derivadas com relação a $\boldsymbol{\beta}$, para os dois modelos, são tais que

$$\frac{\partial^2 L_{\mathbf{y}}(\tau, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \frac{\partial^2 L_{\mathbf{y}|n}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}.$$

Devido à linearidade em (4.8) segue que a matriz de informação observada para $(\tau, \boldsymbol{\beta}^T)^T$ é bloco-diagonal com elementos dados por $-\partial^2 L_{\mathbf{y}}(\tau, \boldsymbol{\beta})/\partial \tau^2$ e $-\partial^2 L_{\mathbf{y}}(\tau, \boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T$, respectivamente. Segue, portanto, que a matriz de informação de Fisher será também bloco-diagonal com os valores esperados das quantidades acima,

$$\mathbf{K}_{\tau\boldsymbol{\beta}} = \begin{bmatrix} E_{\mathbf{y}} \left\{ -\frac{\partial^2 L_{\mathbf{y}}(\tau, \boldsymbol{\beta})}{\partial \tau^2} \right\} & \mathbf{0} \\ \mathbf{0} & E_{\mathbf{y}} \left\{ -\frac{\partial^2 L_{\mathbf{y}}(\tau, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right\} \end{bmatrix}.$$

A variância assintótica de $\hat{\boldsymbol{\beta}}$ fica então dada por

$$\text{Var}_{\mathbf{y}}(\hat{\boldsymbol{\beta}}) = [E_{\mathbf{y}}\{-\partial^2 L_{\mathbf{y}}(\tau, \boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T\}]^{-1}.$$

Palmgren (1981) mostra que $\mathbf{K}_{\tau\beta}$ coincide com a matriz de informação observada sob a restrição $\tau = n$.

Esses resultados podem ser generalizados para quaisquer dimensões de tabelas bem como sob a presença de variáveis explicativas contínuas. A variância assintótica de $\hat{\beta}$ fica no modelo multinomial dada por

$$\text{Var}_{\mathbf{y}|n}(\hat{\beta}) = \left[\mathbb{E}_{\mathbf{y}|n} \left\{ -\frac{\partial^2 L_{\mathbf{y}|n}(\beta)}{\partial \beta \partial \beta^T} \right\} \right]^{-1},$$

coincidindo com a variância assintótica do modelo não condicional sob a restrição $\tau = n$. Contudo, do ponto de vista prático, as variâncias assintóticas de $\hat{\beta}$ devem coincidir uma vez que a estimativa de máxima verossimilhança de τ é dada por $\hat{\tau} = n$.

4.4.1 Modelos log-lineares hierárquicos

Um modelo log-linear é dito hierárquico se dado que uma interação está no modelo, todas as interações de ordem menor como também os efeitos principais correspondentes deverão estar também no modelo. A utilização de tais modelos tem a vantagem de permitir uma interpretação das interações nulas como probabilidades condicionais. Em muitos casos podemos expressar as estimativas dos valores médios em forma fechada, evitando assim a utilização de processos iterativos.

Como ilustração, vamos supor o modelo log-linear apresentado na seção anterior. Podemos mostrar que a hipótese $H_0 : \beta_{12(ij)} = 0, \forall ij$, é equivalente à hipótese de independência na tabela, isto é $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}, \forall ij$. Dado que não há interação, testarmos a ausência de efeito de exposição, isto é testarmos $H_0 : \beta_{1(i)} = 0, i = 1, \dots, r$, é equivalente a testarmos $H_0 : \pi_{1+} = \dots = \pi_{r+} = 1/r$. Finalmente, dado que não há interação, testarmos a ausência de efeito de grupo, isto é testarmos $H_0 : \beta_{2(j)} = 0, j = 1, \dots, s$, é equivalente a testarmos $H_0 : \pi_{+1} = \dots = \pi_{+s} = 1/s$.

Vamos supor agora um modelo log-linear de Poisson com três fatores de r , s e t níveis, respectivamente. Podemos representar a parte sistemática do modelo saturado da seguinte forma:

$$\log \mu_{ijk} = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)} + \beta_{12(ij)} + \beta_{13(ik)} + \beta_{23(jk)} + \beta_{123(ijk)}, \quad (4.9)$$

com as restrições $\beta_{1(1)} = \beta_{2(1)} = \beta_{3(1)} = 0$, $\beta_{12(1j)} = \beta_{12(i1)} = 0$, $\beta_{13(1k)} = \beta_{13(i1)} = 0$, $\beta_{23(1k)} = \beta_{23(j1)} = 0$, $\beta_{123(1jk)} = \beta_{123(i1k)} = \beta_{123(ij1)} = 0$, para $i = 1, \dots, r$; $j = 1, \dots, s$ e $k = 1, \dots, t$. Temos várias classes de modelos hierárquicos que correspondem a situações de interesse na tabela de contingência formada. Uma primeira classe corresponde à hipótese de ausência de interação de segunda ordem, representada por $H_0 : \beta_{123(ijk)} = 0, \forall ijk$, sendo equivalente à hipótese de associação entre dois fatores quaisquer ser constante nos níveis do terceiro. Isso quer dizer, em outras palavras, que a razão de produtos cruzados $\pi_{ijk}\pi_{i'j'k}/\pi_{ij'k}\pi_{i'jk}$, representando a associação entre os níveis (i, j) e (i', j') dos dois primeiros fatores, é constante nos níveis do terceiro fator. Se omitimos no modelo (4.9) a interação de segunda ordem mais uma interação de primeira ordem, dizemos que os dois fatores omitidos correspondentes à interação de primeira ordem são independentes do terceiro fator. Por exemplo, se omitimos $\beta_{123(ijk)}$ e $\beta_{23(jk)}$, $\forall ijk$, ficando com a parte sistemática

$$\log \mu_{ijk} = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)} + \beta_{12(ij)} + \beta_{13(ik)},$$

dizemos que os fatores 2 e 3 são independentes nos níveis do primeiro fator, ou equivalentemente, que

$$\pi_{ijk} = \pi_{ij+}\pi_{i+k}/\pi_{i++}, \quad \forall ijk.$$

Se agora omitimos além de $\beta_{123(ijk)}$ e $\beta_{23(jk)}$ também $\beta_{13(ik)}$, $\forall ijk$, ficando a parte sistemática

$$\log \mu_{ijk} = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)} + \beta_{12(ij)},$$

dizemos que o terceiro fator é independente dos dois primeiros, ou equivalentemente, que

$$\pi_{ijk} = \pi_{ij+}\pi_{++k}, \quad \forall ijk.$$

O modelo apenas com os efeitos principais cuja parte sistemática é dada por

$$\log \mu_{ijk} = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)},$$

equivale à hipótese de independência entre os três fatores, isto é, que

$$\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}, \quad \forall ijk.$$

A Tabela 4.11 resume as três situações de independência para o modelo (4.9).

Tabela 4.11 <i>Algumas interações em modelos log-lineares de Poisson.</i>		
Forma para π_{ijk}	Interação	Interpretação
$\pi_{i++}\pi_{+j+}\pi_{++k}$	nenhuma	fatores mutuamente independentes
$\pi_{ij+}\pi_{++k}$	$\beta_{12(ij)}$	fatores 1 e 2 independentes do fator 3
$\pi_{ij+}\pi_{i+k}/\pi_{i++}$	$\beta_{12(ij)} + \beta_{13(ik)}$	fatores 2 e 3 independentes nos níveis do fator 1

Em muitos desses casos é possível expressarmos as estimativas das probabilidades π_{ijk} 's em forma fechada. Uma análise mais completa de modelos hierárquicos pode ser encontrada, por exemplo, em Cordeiro e Paula (1989b, Cap. 3) e Agresti (1990, Cap. 5).

4.4.2 Aplicações

Associação entre renda e satisfação no emprego

A Tabela 4.12 apresenta o resultado de uma pesquisa com 901 indivíduos (Agresti, 1990, pgs. 20-21) classificados segundo a renda anual e o grau

de satisfação no emprego. Denotamos por Y_{ij} o número de indivíduos pertencentes à classe de renda i com grau de satisfação j . Esses dados estão disponíveis no arquivo **emprego.dat**.

Tabela 4.12
Classificação de indivíduos segundo a renda e o grau de satisfação no emprego.

Renda (US\$)	Grau de Satisfação			
	Alto	Bom	Médio	Baixo
<6000	20	24	80	82
6000-15000	22	38	104	125
15000-25000	13	28	81	113
>25000	7	18	54	92

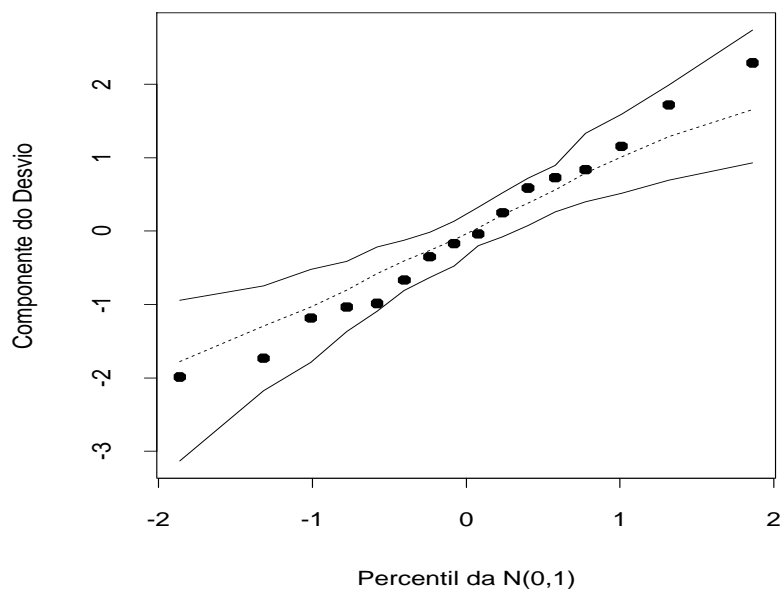


Figura 4.11: Gráfico normal de probabilidades referente ao modelo log-linear de Poisson ajustado aos dados sobre renda e satisfação no emprego.

Tabela 4.13
*Estimativas dos parâmetros do modelo log-linear
de Poisson ajustado ao dados sobre renda e
satisfação no emprego.*

Efeito	Parâmetro	Estimativa	E/E.Padrão
Constante	α	2,651	18,80
Renda 2	$\beta_{1(2)}$	0,338	3,71
Renda 3	$\beta_{1(3)}$	0,132	1,389
Renda 4	$\beta_{1(4)}$	-0,186	-1,81
Grau 2	$\beta_{2(2)}$	0,555	3,49
Grau 3	$\beta_{2(3)}$	1,638	11,87
Grau 4	$\beta_{2(4)}$	1,894	13,93

Vamos supor que $Y_{ij} \sim P(\mu_{ij})$ com parte sistemática inicialmente dada por (modelo saturado)

$$\log \mu_{ij} = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)},$$

em que μ_{ij} denota o número esperado de indivíduos pertencentes à classe de renda i com grau de satisfação j , $\beta_{1(i)}$ denota o efeito renda, $\beta_{2(j)}$ denota o efeito satisfação e $\beta_{12(ij)}$ denota a interação. Temos as restrições $\beta_{1(1)} = \beta_{2(1)} = 0$. O teste da razão de verossimilhanças para testarmos $H_0 : \beta_{12(ij)} = 0, \forall ij$ (ausência de interação) fornece o valor $\xi_{RV} = 12,04$ com nível descritivo $P = 0,21$, indicando pela ausência de interação ou independência entre os dois fatores. Se denotarmos por π_{ij} a proporção de indivíduos na classe de renda i e grau de satisfação j , não rejeitarmos H_0 é equivalente a escrevermos $\pi_{ij} = \pi_{i+}\pi_{+j}, \forall ij$, em que π_{i+} denota a proporção de indivíduos na classe de renda i e π_{+j} denota a proporção de indivíduos com grau de satisfação j . Ou seja, temos independência entre renda e satisfação no emprego. Isso significa que a distribuição do grau de satisfação no emprego é mesma em todas as faixas de renda. A Tabela 4.13 apresenta as estimativas dos parâmetros do modelo com efeitos principais.

Os fatores renda e grau de satisfação são altamente significativos. Notamos pelas estimativas dos parâmetros que há uma proporção maior de indivíduos na classe de renda 2 (6000-15000) e uma proporção menor na classe de renda 4 (>25000). Por outro lado, notamos que a proporção de indivíduos cresce com o aumento do grau de satisfação. O desvio do modelo foi de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 12,04$ (9 graus de liberdade) com nível descritivo de $P = 0,21$, indicando um ajuste adequado. Pelo gráfico normal de probabilidades com o resíduo t_{D_i} (Figura 4.11) não há indícios fortes de que o modelo adotado seja incorreto, embora o fato dos resíduos negativos estarem abaixo da reta mediana e os resíduos positivos ligeiramente acima seja um indício de sobredispersão nos dados.

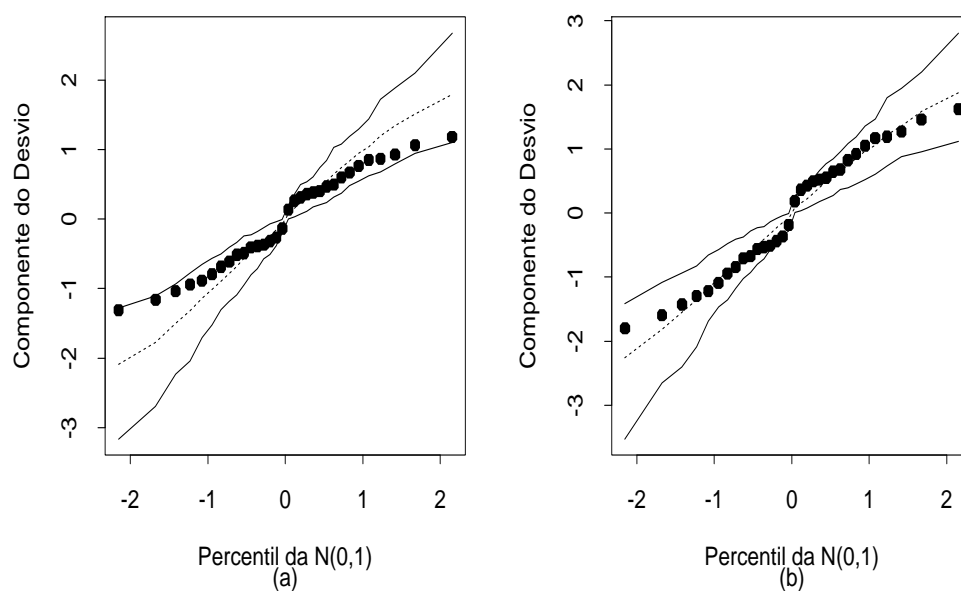


Figura 4.12: Gráficos normais de probabilidades referentes ao modelo log-linear de Poisson (a) e log-linear de quase-verossimilhança (b) ajustados aos dados sobre doença das coronárias.

Doença das coronárias

Tabela 4.14

Distribuição de 1330 pacientes segundo ocorrência de doença das coronárias, nível de colesterol e pressão arterial.

Doença das coronárias	Nível de colesterol	Pressão arterial			
		1	2	3	4
Sim	1	2	3	3	4
	2	3	2	1	3
	3	8	11	6	6
	4	7	12	11	11
Não	1	117	121	47	22
	2	85	98	43	20
	3	119	209	68	43
	4	67	99	46	33

Vamos considerar agora os dados da Tabela 4.14 (Everitt, 1977) referente à classificação de 1330 pacientes segundo três fatores: doença das coronárias (sim ou não), nível de colesterol (1: menor do que 200 mg/100 cc, 2: 200-219, 3: 220-259 e 4: 260 ou +) e pressão arterial (1: menor do que 127 mm Hg, 2: 127-146, 3: 147-166 e 4: 167 ou +). Os dados estão também descritos no arquivo **heart.dat**. Denotamos por Y_{ijk} o número de pacientes nos níveis (i, j, k) dos três fatores: doença das coronárias, nível de colesterol e pressão arterial, respectivamente. Vamos supor que $Y_{ijk} \sim P(\mu_{ijk})$ com parte sistemática inicialmente dada por (modelo saturado)

$$\log \mu_{ijk} = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)} + \beta_{12(ij)} + \beta_{13(ik)} + \beta_{23(jk)} + \beta_{123(ijk)},$$

em que μ_{ijk} denota o número esperado de indivíduos pertencentes aos níveis (i, j, k) , respectivamente, $\beta_{1(i)}$ denota o efeito doença das coronárias, $\beta_{2(j)}$ denota o efeito nível de colesterol, $\beta_{3(k)}$ denota o efeito pressão arterial e $\beta_{12(ij)}$,

$\beta_{13(ik)}$, $\beta_{23(jk)}$ e $\beta_{123(ijk)}$ são as interações de 1ª e 2ª ordens, respectivamente, com as restrições dadas na Seção 4.4.1.

Tabela 4.15

*Resumo do ANODEV referente ao modelo
log-linear de Poisson ajustado aos
dados sobre doença das coronárias.
(D:doença, C:colesterol e P:pressão)*

Efeito	Desvio	g.l.	Diferença	g.l.
D+C+P	78,96	24	-	-
+ D.C	48,51	21	30,45	3
+ D.P	24,40	18	24,10	3
+ C.P	4,77	9	19,63	9

Pela Tabela 4.15 notamos que, segundo o princípio hierárquico, apenas a interação de segunda ordem pode ser eliminada. A inclusão dos efeitos principais é altamente significativa. Dado que os efeitos principais estão no modelo, a inclusão da interação doença*colesterol ($\beta_{12(ij)}$) leva a $\xi_{RV} = 30,45$ (3 graus de liberdade) com $P = 0,00$. Dado que essa interação está no modelo, a inclusão da interação doença*pressão ($\beta_{13(ik)}$) fornece $\xi_{RV} = 24,10$ (3 graus de liberdade) com $P = 0,00$. Finalmente, dadas as duas interações de primeira ordem, a inclusão da interação remanescente, colesterol*pressão, leva a $\xi_{RV} = 19,62$ (9 graus de liberdade) com $P = 0,02$. O desvio do modelo (4.9) sem a interação de segunda ordem é de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 4,77$ (9 graus de liberdade) para um nível descritivo de $P = 0,853$, indicando um ajuste adequado.

A ausência de interação de segunda ordem neste exemplo significa que as razões de chances (entre os níveis de colesterol ou entre os níveis de pressão arterial) são as mesmas nos grupos de doentes e não doentes. Contudo, o gráfico normal de probabilidades descrito na Figura 4.12a indica que os resíduos negativos estão acima da média esperada, ocorrendo o contrário com os resíduos positivos, embora todos sejam em geral pequenos. Isso é

um indício modesto de subdispersão, fenômeno que também pode ocorrer em modelos de Poisson.

Um modelo de quase-verossimilhança similar ao que foi usado no exemplo da Seção 4.2.6 leva à estimativa $\hat{\sigma}^2 = 0,53$. Na Figura 4.12b temos o gráfico normal de probabilidades com o resíduo componente do desvio corrigido pela estimativa de dispersão. Notamos que os resíduos estão melhor distribuídos dentro do envelope gerado. A conclusão deste exemplo é que há associação entre os fatores dois a dois e que essa associação é constante nos níveis do terceiro fator.

4.5 Modelos de contagem com excesso de zeros

4.5.1 Modelos de contagem ajustados em zero

Os modelos de contagem ajustados em zero são também conhecidos como modelos de barreira (ver, por exemplo, Mullaby, 1986). Para formalizá-los vamos supor que Z é uma variável aleatória com função de probabilidades dada por

$$P\{Z = z\} = \begin{cases} \pi & \text{se } z = 0, \\ (1 - \pi) \frac{f_Y(z)}{1 - f_Y(0)} & \text{se } z = 1, 2, \dots, \end{cases}$$

em que $0 < \pi < 1$ e $f_Y(z)$ denota a função de probabilidades de uma variável aleatória Y de contagem, por exemplo, Poisson ou binomial negativa. Portanto, desde que $\sum_{z=1}^{\infty} f_Y(z) = 1 - f_Y(0)$, segue que

$$\begin{aligned} P\{Z \geq 1\} &= (1 - \pi) \sum_{z=1}^{\infty} f_Y(z) / \{1 - f_Y(0)\} \\ &= (1 - \pi) \{1 - f_Y(0)\} / \{1 - f_Y(0)\} \\ &= 1 - \pi. \end{aligned}$$

Logo, $\sum_{y=0}^{\infty} P\{Z = z\} = \pi + (1 - \pi) = 1$. Um exemplo poderia ser Z denotando o número de dias que pacientes dependentes de álcool que estão fazendo tratamento consumiram a bebida. O zero representa os pacientes que ficaram em abstinência no período mas que poderiam ter consumido álcool. Um outro exemplo poderia ser estudar o número de vezes que um idoso visita um médico no período de 1 ano. Os zeros são aqueles idosos que naquele ano não precisaram ir ao médico. Os fatores que explicam a probabilidade de zero podem ser diferentes daqueles que explicam a probabilidade de ocorrência do evento. Se queremos apenas estudar os pacientes que consumiram a bebida no período ou os idosos que visitaram um médico no ano podemos assumir no modelo $\pi = 0$.

Os dois primeiros momentos de Z ficam dados por

$$\begin{aligned} E(Z) &= \sum_{z=1}^{\infty} z(1 - \pi) \frac{f_Y(z)}{\{1 - f_Y(0)\}} \\ &= \frac{(1 - \pi)}{\{1 - f_Y(0)\}} \sum_{z=1}^{\infty} z f_Y(z) \\ &= \frac{E(Y)(1 - \pi)}{\{1 - f_Y(0)\}} \end{aligned}$$

e

$$\begin{aligned} E(Z^2) &= \sum_{z=1}^{\infty} z^2(1 - \pi) \frac{f_Y(z)}{\{1 - f_Y(0)\}} \\ &= \frac{(1 - \pi)}{\{1 - f_Y(0)\}} \sum_{z=1}^{\infty} z^2 f_Y(z) \\ &= \frac{E(Y^2)(1 - \pi)}{\{1 - f_Y(0)\}}. \end{aligned}$$

Daí segue que

$$\text{Var}(Z) = \frac{(1 - \pi)}{\{1 - f_Y(0)\}} \left[E(Y^2) - \frac{E^2(Y)(1 - \pi)}{\{1 - f_Y(0)\}} \right].$$

Iremos denotar $Z \sim \text{ZAP}(\lambda, \pi)$ para o modelo de Poisson ajustado em zero e $Z \sim \text{ZANB}(\lambda, \phi, \pi)$ para o modelo binomial negativo ajustado em zero. Logo, se $Y \sim P(\lambda)$ então $f_Y(y) = e^{-\lambda} \lambda^y / y!$ e em particular $f_Y(0) = e^{-\lambda}$. Para $Y \sim \text{BN}(\lambda, \phi)$ temos que

$$f_Y(y) = \frac{\Gamma(\phi + y)}{\Gamma(y + 1)\Gamma(\phi)} \left(\frac{\lambda}{\lambda + \phi} \right)^y \left(\frac{\phi}{\lambda + \phi} \right)^\phi,$$

em particular $f_Y(0) = \phi^\phi / (\lambda + \phi)^\phi$.

4.5.2 Modelos de regressão ajustados em zero

Vamos supor agora que Z_1, \dots, Z_n são variáveis aleatórias independentes com distribuição de Poisson ou binomial negativa ajustadas em zero. Então,

$$P\{Z_i = z_i\} = \begin{cases} \pi_i & \text{se } z_i = 0, \\ (1 - \pi_i) \frac{f_{Y_i}(z_i)}{\{1 - f_{Y_i}(0)\}} & \text{se } z_i = 1, 2, \dots, \end{cases}$$

para $i = 1, \dots, n$. O logaritmo da função de verossimilhança fica dado por $L = \sum_{i=1}^n \log f_{Z_i}(z_i)$, em que $\log f_{Z_i}(0) = \log \pi_i$ e $\log f_{Z_i}(z_i) = \log(1 - \pi_i) + \log f_{Y_i}(z_i) - \log\{1 - f_{Y_i}(0)\}$ para $z_i = 1, 2, \dots$.

Por exemplo, se assumimos que $Z_i \sim \text{ZAP}(\lambda_i, \pi)$ em que $\lambda_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$ e $\mu_i = E(Z_i)$, então segue que

$$\begin{aligned} \mu_i &= \frac{E(Y_i)(1 - \pi)}{\{1 - f_Y(0)\}} \\ &= \frac{\lambda_i(1 - \pi)}{\{1 - e^{-\lambda_i}\}} \\ &= \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}(1 - \pi)}{[1 - \exp\{-\exp(\mathbf{x}_i^T \boldsymbol{\beta})\}]}. \end{aligned}$$

Portanto,

$$\log \mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \log(1 - \pi) - \log[1 - \exp\{-\exp(\mathbf{x}_i^T \boldsymbol{\beta})\}].$$

4.5.3 Modelos de contagem inflacionados de zeros

Os modelos de contagem inflacionados de zeros (ver, por exemplo, Lambert, 1992) são caracterizados pela ocorrência de zeros em duas situações: (i) zeros que ocorrem segundo uma distribuição de contagem ou (ii) zeros inflados que podem ser zeros estruturais. Por exemplo, podemos estar interessados em estudar o número de dias num período que um indivíduo consumiu um determinado produto. Aqueles indivíduos que não consomem o produto por alguma razão, por exemplo intolerância ao produto, serão tratadas como zeros estruturais e aqueles indivíduos que não consumiram o produto no período, mas podem vir a consumi-lo, como zeros eventuais que serão associados a alguma distribuição de contagem Y . Um outro exemplo poderia ser o estudo do número de vezes que um indivíduo contraiu um tipo de infecção num determinado período. Aqueles indivíduos imunes à infecção seriam tratados como zeros inflados. Nesses casos, a variável aleatória Z tem função de probabilidades expressa na seguinte forma:

$$P\{Z = z\} = \begin{cases} \pi + (1 - \pi)f_Y(0) & \text{se } z = 0, \\ (1 - \pi)f_Y(z) & \text{se } z = 1, 2, \dots, \end{cases}$$

em que $0 < \pi < 1$ e $f_Y(z)$ denota a função de probabilidades de uma variável aleatória Y , por exemplo, Poisson ou binomial negativa. Desde que $\sum_{z=1}^{\infty} f_Y(z) = 1 - f_Y(0)$ obtemos $\sum_{z=0}^{\infty} P\{Z = z\} = \pi + (1 - \pi)f_Y(0) + (1 - \pi)\{1 - f_Y(0)\} = \pi + (1 - \pi) = 1$.

Os dois primeiros momentos de Y ficam dados por

$$\begin{aligned} E(Z) &= \sum_{z=1}^{\infty} z(1 - \pi)f_Y(z) \\ &= (1 - \pi) \sum_{z=1}^{\infty} zf_Y(z) \\ &= (1 - \pi)E(Y) \end{aligned}$$

e

$$\begin{aligned}
E(Z^2) &= \sum_{z=1}^{\infty} z^2(1-\pi)f_Y(z) \\
&= (1-\pi) \sum_{z=1}^{\infty} z^2 f_Y(z) \\
&= (1-\pi)E(Y^2).
\end{aligned}$$

Assim,

$$\begin{aligned}
\text{Var}(Z) &= E(Z^2) - E^2(Z) \\
&= (1-\pi)E(Y^2) - (1-\pi)^2 E^2(Y) \\
&= (1-\pi)\{E(Y^2) - (1-\pi)E^2(Y)\}.
\end{aligned}$$

Iremos denotar $Z \sim \text{ZIP}(\lambda, \pi)$ para a distribuição de Poisson inflacionada de zeros e por $Z \sim \text{ZINB}(\lambda, \phi, \pi)$ para a distribuição binomial negativa inflacionada de zeros.

4.5.4 Modelos de regressão inflacionados de zeros

Vamos supor agora que Z_1, \dots, Z_n são variáveis aleatórias independentes com distribuição de Poisson ou binomial negativa inflacionadas de zeros. Então,

$$P\{Z_i = z_i\} = \begin{cases} \pi_i + (1 - \pi_i)f_Y(0) & \text{se } z_i = 0, \\ (1 - \pi_i)f_Y(z_i) & \text{se } z_i = 1, 2, \dots, \end{cases}$$

para $i = 1, \dots, n$. O logaritmo da função de verossimilhança fica dado por $L = \sum_{i=1}^n \log f_{Z_i}(z_i)$, em que $\log f_{Z_i}(0) = \log\{\pi_i + (1 - \pi_i)f_{Y_i}(0)\}$ e $\log f_{Z_i}(z_i) = \log(1 - \pi_i) + \log f_{Y_i}(z_i)$ para $z_i = 1, 2, \dots$

Por exemplo, podemos supor que $Y_i \sim P(\lambda_i)$ com $\lambda_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$ e $\log\{\pi_i/(1 -$

$\pi_i)\} = \mathbf{u}_i^T \boldsymbol{\gamma}$. Dessa forma segue que $\mu_i = E(Z_i)$ fica expresso como

$$\begin{aligned}\mu_i &= (1 - \pi_i)E(Y_i) \\ &= (1 - \pi_i)\lambda_i \\ &= \left\{ 1 - \frac{e^{\mathbf{u}_i^T \boldsymbol{\gamma}}}{1 + e^{\mathbf{u}_i^T \boldsymbol{\gamma}}} \right\} e^{\mathbf{x}_i^T \boldsymbol{\beta}} \\ &= \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{\{1 + e^{\mathbf{u}_i^T \boldsymbol{\gamma}}\}}.\end{aligned}$$

Isto é,

$$\log \mu_i = \mathbf{x}_i^T \boldsymbol{\beta} - \log\{1 + e^{\mathbf{u}_i^T \boldsymbol{\gamma}}\}.$$

Os modelos de contagem ajustados em zero e inflacionados de zeros podem ajustados, por exemplo, pela biblioteca `pscl` (Zeileis et al. 2008) do R.

4.6 Exercícios

1. Seja Y uma variável aleatória com distribuição binomial negativa, isto é, Y é o número de ensaios até a ocorrência do r -ésimo sucesso, em que π é a probabilidade de sucesso em cada ensaio. Mostre que a função de probabilidades de Y pode ser expressa na forma exponencial. Calcule μ e $V(\mu)$. Use a forma abaixo para a função de probabilidades de Y

$$f(y; \pi, r) = \binom{y-1}{r-1} \pi^r (1-\pi)^{(y-r)},$$

em que $y = r, r+1, \dots$. Obtenha a função desvio supondo uma amostra de n variáveis aleatórias independentes de probabilidades de sucesso π_i .

2. Seja Y variável aleatória com distribuição binomial negativa biparamétrica de média μ e parâmetro de forma ν , cuja função de probabilidades é dada por

$$f(y; \mu, \nu) = \frac{\Gamma(\nu + y)}{\Gamma(y+1)\Gamma(\nu)} \left(\frac{\mu}{\mu + \nu} \right)^y \left(\frac{\nu}{\mu + \nu} \right)^\nu,$$

em que $\mu > 0$, $\nu > 0$ e $y = 0, 1, 2, \dots$. Mostre que para ν conhecido a distribuição de Y pertence à família exponencial de distribuições. Encontrar a função de variância. Obtenha a função desvio supondo uma amostra de n variáveis aleatórias independentes de médias μ_i e parâmetro de dispersão ν .

3. Sejam Y_1 e Y_2 variáveis aleatórias independentes tais que $Y_i \sim P(\lambda_i)$, $i = 1, 2$. Considere a razão de taxas $\psi = \lambda_1/\lambda_2$. Encontre a variância assintótica de $\tilde{\psi}$, $\text{Var}_A(\tilde{\psi})$.
4. (Breslow e Day, 1987). A tabela abaixo apresenta o número de mortes por câncer respiratório e o número de pessoas-anos de observação entre trabalhadores de indústrias siderúrgicas do estado de Montana (EUA) segundo o nível de exposição ao arsênico.

	Nível de Exposição	
	Alto	Baixo
Casos	68	47
Pessoas-Anos	9018	13783

Sejam Y_1 e Y_2 o número de casos observados para o nível alto e baixo de arsênico, respectivamente. Suponha que $Y_i \sim P(\lambda_i t_i)$, em que t_i denota o número de pessoas-anos, $i = 1, 2$. Considere a razão de taxas $\psi = \lambda_1/\lambda_2$. Encontre $\tilde{\psi}$ e um intervalo de confiança exato de 95% para ψ . Com base neste intervalo qual sua conclusão sobre a hipótese $H_0 : \psi = 1$? Informações úteis: $F_{136,96}(0,025) = 0,694$ e $F_{138,94}(0,975) = 1,461$.

5. (Neter et al., 1996, p. 623). No arquivo **geriatra.dat** estão descritos os dados de um estudo prospectivo com 100 indivíduos de pelo menos 65 anos de idade em boas condições físicas. O objetivo do estudo é

tentar relacionar o número médio de quedas num período de seis meses com algumas variáveis explicativas. Os dados estão descritos na seguinte ordem: **quedas** (número de quedas no período), **intervenção** (=0 educação somente, =1 educação e exercícios físicos), **sexo** (=0 feminino, =1 masculino), **balanço** (escore) e **força** (escore). Para as variáveis **balanço** e **força** quanto maior o valor maior o balanço e a força do indivíduo, respectivamente. Tente selecionar um modelo apropriado apenas com os efeitos principais. Interprete os resultados e faça uma análise de diagnóstico.

6. (Breslow e Day, 1987, pgs. 140-142). Os dados do arquivo **canc1.dat** são provenientes de um estudo de seguimento para estudar a associação entre a taxa anual de câncer nasal em trabalhadores de uma refinaria de níquel no País de Gales e algumas variáveis explicativas: idade no primeiro emprego (4 níveis), ano do primeiro emprego (4 níveis) e tempo decorrido desde o primeiro emprego (5 níveis). São também apresentados o número de casos de câncer nasal e o total de pessoas-anos para cada combinação desses três fatores. Proponha um modelo log-linear com resposta de Poisson sendo o número de casos de câncer nasal com *offset* dado por $\log(\text{pessoas-anos})$. Considere inicialmente apenas os efeitos principais dos três fatores: idade no primeiro emprego, ano do primeiro emprego e tempo decorrido desde o primeiro emprego. Verifique se é possível incluir alguma interação de primeira ordem. Faça uma análise de diagnóstico com o modelo final e interprete os resultados, por exemplo, fazendo uma comparação das taxas estimadas de câncer nasal entre os níveis de um mesmo fator.
7. (Hinde, 1982). No arquivo **rolos.dat** são apresentados os dados referentes à produção de peças de tecido numa determinada fábrica. Na

primeira coluna tem-se o comprimento da peça (em metros) e na segunda coluna o número de falhas. Faça inicialmente um gráfico do número de falhas contra o comprimento da peça. Ajuste um modelo log-linear de Poisson apropriado. Faça uma análise de resíduos e verifique se há indícios de sobredispersão. Em caso afirmativo ajuste um modelo de quase-verossimilhança e um modelo log-linear com distribuição binomial negativa. Interprete os resultados pelas razões de médias $\mu(x+1)/\mu(x)$, em que x denota o comprimento da peça.

8. Sejam Y_1, \dots, Y_n variáveis aleatórias independentes tais que $Y_i \sim P(\mu_i)$ e parte sistemática dada por $\eta_i = \sum_{j=1}^p x_{ij}\beta_j$ tal que $\sum_{i=1}^n x_{ij}x_{i\ell} = 0$, para $j \neq \ell$. Ou seja, as colunas da matriz modelo \mathbf{X} são ortogonais. Encontre a ligação que faz com que $\text{Corr}(\hat{\beta}_j, \hat{\beta}_\ell) = 0$, para $j \neq \ell$. Comente sobre as vantagens desse resultado.
9. Considere um experimento em que duas máquinas, M1 e M2, são observadas durante o mesmo período sendo computados para cada uma o número de peças defeituosas produzidas, conforme descrito pelo esquema abaixo.

	M1	M2
P. Defeituosas	y_1	y_2

Suponha que $Y_1 \sim P(\lambda_1)$ e $Y_2 \sim P(\lambda_2)$ e considere o modelo log-linear $\log \lambda_1 = \alpha$ e $\log \lambda_2 = \alpha + \beta$. Obtenha a variância assintótica de $\hat{\beta}$, $\text{Var}_y(\hat{\beta})$, expressando-a em função de α e β . Proponha agora um modelo binomial condicional, dado $Y_1 + Y_2 = m$. Expresse a probabilidade de sucesso π em função de β . Interprete π e encontre a variância assintótica de $\hat{\beta}$, $\text{Var}_{y|m}(\hat{\beta})$. Mostre que as duas variâncias assintóticas

estimadas coincidem e são dadas por

$$\hat{\text{Var}}(\hat{\beta}) = \frac{(1 + e^{\hat{\beta}})^2}{me^{\hat{\beta}}},$$

em que $\hat{\beta}$ é o estimador de máxima verossimilhança de β . Comente.

10. Supor Y_1, \dots, Y_n variáveis aleatórias independentes tais que $Y_i \sim P(\mu_i)$ e seja $\sqrt{\mu_i} = \alpha + \beta(x_i - \bar{x})$, em que \bar{x} é a média amostral de x_1, \dots, x_n .
 (i) Obtenha a matriz modelo \mathbf{X} . (ii) Calcule as variâncias assintóticas $\text{Var}(\hat{\alpha})$ e $\text{Var}(\hat{\beta})$. (iii) Mostre também que $\text{Cov}(\hat{\alpha}, \hat{\beta}) = 0$ e comente.
 (iv) Como fica o teste de escore para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$? Qual a distribuição nula assintótica da estatística do teste?

11. Sejam Y_{ij} variáveis aleatórias mutuamente independentes tais que $Y_{ij} \sim \text{BN}(\mu_i, \nu)$ para $i = 1, 2$ e $j = 1, \dots, m$ com parte sistemática dada por $\mu_1 = \alpha - \beta$ e $\mu_2 = \alpha + \beta$. (i) Como fica a matriz modelo \mathbf{X} ? (ii) Calcule $\text{Var}(\hat{\beta})$ e (iii) mostre que a estatística de escore para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$ pode ser expressa na forma

$$\xi_{SR} = \frac{m\hat{\nu}^0 (\bar{y}_2 - \bar{y}_1)^2}{2\bar{y} (\bar{y} + \hat{\nu}^0)},$$

em que $\bar{y} = (\bar{y}_1 + \bar{y}_2)/2$ e $\hat{\nu}^0$ denota a estimativa de ν sob H_0 .

12. Sejam Y_1, \dots, Y_n variáveis aleatórias independentes tais que $Y_i \sim \text{BN}(\mu_i, \nu)$ com parte sistemática dada por $\log \mu_i = \alpha + \beta(x_i - \bar{x})$ em que $\bar{x} = \frac{\sum x_i}{n}$.
 (i) Como fica a matriz modelo \mathbf{X} ? (ii) Obtenha $\text{Var}(\hat{\beta})$. (iii) Como fica o teste de escore para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$? Qual a distribuição nula assintótica da estatística do teste?
13. Sejam Y_1, \dots, Y_n variáveis i.i.d. tais que $Y_i \sim \text{BN}(\mu, \phi)$, $i = 1, \dots, n$. Como fica o teste da razão de verossimilhanças para testar $H_0 : \phi = 1$

contra $H_1 : \phi \neq 1$? Qual a distribuição nula assintótica da estatística do teste? Como fica a estimativa de μ sob as duas hipóteses?

14. Na tabela abaixo uma amostra de 174 alunos de Estatística Básica no IME-USP foi classificada segundo o curso e o desempenho na disciplina.

Curso	Resultado da Avaliação		
	Aprovado	Reprovado	Reavaliação
Pedagogia	32	16	3
Geografia	32	18	10
Física	35	14	14

Ajustar um modelo log-linear de Poisson para explicar π_{ij} , a proporção de alunos do curso i com resultado j , em que $i, j = 1, 2, 3$. Interprete os resultados e faça uma análise de diagnóstico.

15. (Hand et al., 1994). No arquivo **recrutas.dat** são descritos os resultados de um estudo desenvolvido em 1990 com recrutas americanos referente a associação entre o número de infecções de ouvido e alguns fatores. Os dados são apresentados na seguinte ordem: hábito de nadar (ocasional ou frequente), local onde costuma nadar (piscina ou praia), faixa etária (15-19, 20-25 ou 25-29), sexo (masculino ou feminino) e número de infecções de ouvido diagnosticadas pelo próprio recruta. Verifique qual dos modelos, log-linear de Poisson, quase-verossimilhança ou log-linear binomial negativo, se ajusta melhor aos dados. Utilize métodos de diagnóstico como critério.
16. Supor, por um lado, o modelo log-linear de Poisson em que $Y_i \sim P(\mu_i)$, $i = 1, 2, 3$, em que $\log \mu_1 = \alpha$, $\log \mu_2 = \alpha + \beta_2$ e $\log \mu_3 = \alpha + \beta_3$. Fazendo $\tau = \mu_1 + \mu_2 + \mu_3$ expresse o logaritmo da função de verossimilhança desse modelo em função de (τ, β_2, β_3) . Mostre que a matriz de informação

de Fisher é bloco diagonal $\mathbf{K}_{\tau\beta} = \text{diag}\{\mathbf{K}_\tau, \mathbf{K}_\beta\}$, em que $\boldsymbol{\beta} = (\beta_2, \beta_3)^T$. Por outro lado, sabe-se que a distribuição condicional $\mathbf{Y} = \mathbf{a}|Y_1 + Y_2 + Y_3 = n$, em que $\mathbf{Y} = (Y_1, Y_2, Y_3)^T$ e $\mathbf{a} = (a_1, a_2, a_3)^T$, é multinomial $M(a_1, a_2, a_3; \pi_1, \pi_2, \pi_3)$. Supor o modelo log-linear $\log\pi_1 = \alpha^*$, $\log\pi_2 = \alpha^* + \beta_2$ e $\log\pi_3 = \alpha^* + \beta_3$, em que $\alpha^* = -\log(1 + e^{\beta_2} + e^{\beta_3})$ devido à restrição $\pi_1 + \pi_2 + \pi_3 = 1$. Encontre a matriz de informação de Fisher \mathbf{K}_β^* para $\boldsymbol{\beta} = (\beta_2, \beta_3)^T$ no modelo multinomial. Mostre que as estimativas de máxima verossimilhança para $\boldsymbol{\beta}$ coincidem nos dois modelos log-lineares. Mostre também que $\mathbf{K}_\beta = \mathbf{K}_\beta^*$ quando $\tau = n$, comente.

17. (Bishop, Fienberg e Holland, 1975, p. 143). A tabela abaixo apresenta o resultado de uma pesquisa em que 1008 pessoas receberam duas marcas de detergente, X e M, e posteriormente responderam às seguintes perguntas: maciez da água (leve, média ou forte); uso anterior do detergente M (sim ou não); temperatura da água (alta ou baixa); preferência (marca X ou marca M). Esses dados estão descritos no arquivo **detergente.dat**.

Temperatura	Uso de M	Preferência	Maciez		
			Leve	Média	Forte
Alta	Sim	X	19	23	24
		M	29	47	43
	Não	X	29	33	42
		M	27	23	30
Baixa	Sim	X	57	47	37
		M	49	55	52
	Não	X	63	66	68
		M	53	50	42

Ajustar um modelo log-linear de Poisson para explicar π_{ijkl} , a pro-

porção de indivíduos que responderam, respectivamente, nível de temperatura (i=1 alta, i=2 baixa), uso prévio de M (j=1 sim, j=2 não), preferência (k=1 X, k=2 M) e nível de maciez ($\ell = 1$ leve, $\ell = 2$ médio, $\ell = 3$ forte). Selecionar através do método **AIC** os efeitos principais significativos. Depois incluir apenas as interações significativas de primeira ordem. Interpretar os resultados e fazer uma análise de diagnóstico.

18. Supor que $Y_{ij} \sim P(\mu_{ij})$, para $i = 1, \dots, r$ e $j = 1, \dots, c$, com parte sistemática dada por

$$\log \mu_{ij} = \alpha + \beta_i + \gamma_j,$$

em que $\beta_1 = \gamma_1 = 0$. Supor ainda que os β_i 's referem-se aos efeitos do fator A e os γ_j 's aos efeitos do fator B. Defina um modelo multinomial equivalente e mostre que a representação acima corresponde à independência (no sentido probabilístico) entre os fatores A e B.

19. Seja o modelo trinomial em que $\pi_0 = Pr(Y = 0)$, $\pi_1 = Pr(Y = 1)$ e $\pi_2 = Pr(Y = 2)$ com a restrição $\pi_0 + \pi_1 + \pi_2 = 1$. Suponha que $Y = 0$ se $(Z_0 = 1, Z_1 = 0, Z_2 = 0)$, $Y = 1$ se $(Z_0 = 0, Z_1 = 1, Z_2 = 0)$ e $Y = 2$ se $(Z_0 = 0, Z_1 = 0, Z_2 = 1)$. Note que $Z_0 + Z_1 + Z_2 = 1$. Portanto, a função de probabilidades de (Z_0, Z_1, Z_2) fica dada por

$$g(z_0, z_1, z_2; \pi_0, \pi_1, \pi_2) = \pi_0^{z_0} \pi_1^{z_1} \pi_2^{z_2}.$$

Logo, para uma amostra aleatória de tamanho n a função de probabilidades de $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ pode ser expressa na forma

$$g(\mathbf{y}; \boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2) = \prod_{i=1}^n \pi_{0i}^{z_{0i}} \pi_{1i}^{z_{1i}} \pi_{2i}^{z_{2i}}.$$

É usual considerar a parte sistemática

$$\log \left\{ \frac{\pi_{1i}}{\pi_{0i}} \right\} = \eta_{1i} = \mathbf{x}_i^T \boldsymbol{\beta}_1 \quad \text{e} \quad \log \left\{ \frac{\pi_{2i}}{\pi_{0i}} \right\} = \eta_{2i} = \mathbf{x}_i^T \boldsymbol{\beta}_2$$

sendo que $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$, $\boldsymbol{\beta}_1 = (\beta_{11}, \dots, \beta_{1p})^T$ e $\boldsymbol{\beta}_2 = (\beta_{21}, \dots, \beta_{2p})^T$.

Responda aos itens abaixo:

- (a) Verifique que $\pi_{0i} = \{1 + e^{\eta_{1i}} + e^{\eta_{2i}}\}^{-1}$, $\pi_{1i} = e^{\eta_{1i}} / \{1 + e^{\eta_{1i}} + e^{\eta_{2i}}\}$ e $\pi_{2i} = e^{\eta_{2i}} / \{1 + e^{\eta_{1i}} + e^{\eta_{2i}}\}$.
 - (b) Encontre as funções escore \mathbf{U}_{β_1} e \mathbf{U}_{β_2} de $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$, respectivamente.
 - (c) Encontre a matriz de informação de Fisher para $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$.
 - (d) Desenvolva um processo iterativo para obter a estimativa de máxima verossimilhança de $\boldsymbol{\beta}$. Deixe o processo iterativo em forma matricial. Como iniciá-lo?
 - (e) Como fica o desvio do modelo? E o resíduo componente do desvio?
20. Supor que $Z_i \stackrel{\text{iid}}{\sim} \text{ZAP}(\mu, \pi)$, para $i = 1, \dots, n$. Obtenha as estimativas de máxima verossimilhança $\hat{\mu}$ e $\hat{\pi}$ e a matriz de informação de Fisher para (μ, π) . Sugestão: supor que o total de zeros na amostra segue uma $B(n, \pi)$.
21. Supor que $Z_i \stackrel{\text{iid}}{\sim} \text{ZANBI}(\mu, \nu, \pi)$, para $i = 1, \dots, n$, em que a função de probabilidades de z_i fica dada por

$$f_z(z_i; \mu, \nu, \pi) = \begin{cases} \pi & \text{se } z_i = 0 \\ (1 - \pi) \frac{f_y(z_i; \mu, \nu)}{1 - f_y(0; \mu, \nu)} & \text{se } z_i = 1, 2, \dots, \end{cases}$$

em que $f_y(y_i; \mu, \nu)$ denota a função de probabilidades de uma $\text{BN}(\mu, \nu)$. Supondo $\nu = 1$ obter a estatística da razão de verossimilhanças para testar $H: \mu = 1$ contra $A: \mu \neq 1$?

22. (Lang et al., 1994). No arquivo **nitrofen.dat** estão descritos os dados de um experimento com uma amostra de 50 **C.dubia** (pequeno animal invertebrado aquático de água doce), que foram submetidos a dosagens diferentes do herbicida **Nitrofen**: 0, 80, 160, 235 e 310 mg/ℓ . Para cada nível de **Nitrofen** 10 animais ficaram expostos e foi observado o total de ovos eclodidos após 3 ninhadas. Faça inicialmente uma análise descritiva dos dados, por exemplo um diagrama de dispersão entre o número de ovos eclodidos (**tovos**) contra o nível de exposição do herbicida (**dose**). Compare os ajustes de alguns modelos com resposta de Poisson para explicar o total de ovos eclodidos dado o nível de exposição. Escolha o melhor ajuste através de métodos de diagnóstico. Para o modelo selecionado faça uma interpretação dos coeficientes estimados.
23. Supor que $Z_i \stackrel{\text{iid}}{\sim} \text{ZAP}(\mu, \pi)$, em que $\pi = e^\alpha / (1 + e^\alpha)$, para $i = 1, \dots, n$. Obtenha a estimativa de máxima verossimilhança $\hat{\alpha}$ bem como $\text{Var}(\hat{\alpha})$. Como fica a estatística do teste da razão de verossimilhanças para testar $H_0 : \alpha = 0$ contra $H_1 : \alpha \neq 0$?
24. Supor que Y_i são variáveis aleatórias iid Poisson truncada em zero com função de probabilidades dada por

$$f(y_i; \lambda) = \frac{e^{-\lambda} \lambda^{y_i}}{y_i! (1 - e^{-\lambda})},$$

para $y_i = 1, 2, \dots$, em que $\lambda > 0$ e $i = 1, \dots, n$. Mostre que $U_\lambda = \frac{n}{\lambda} \{\bar{y} - \lambda e^\lambda / (e^\lambda - 1)\}$, obtenha $K_{\lambda\lambda}$ e apresente o procedimento iterativo escore de Fisher para obter $\hat{\mu}$.

Capítulo 5

Modelos de Quase-Verossimilhança

5.1 Introdução

Wedderburn (1974) propôs uma função biparamétrica, denominada função de quase-verossimilhança, que engloba algumas funções de verossimilhança da família exponencial. Todavia, na maioria das situações não é possível através da função de quase-verossimilhança recuperar a verdadeira distribuição da variável resposta. Se Y é a variável aleatória de interesse o logaritmo da função de quase-verossimilhança é definido por

$$Q(\mu; y) = \frac{1}{\sigma^2} \int_y^\mu \frac{y-t}{V(t)} dt,$$

em que $V(t)$ é uma função positiva e conhecida, $-\infty < y, \mu < \infty$ e $\sigma^2 > 0$ é um parâmetro de dispersão. Como temos acima uma integral definida, segue que

$$\begin{aligned} \frac{\partial Q(\mu; y)}{\partial \mu} &= \left. \frac{y-t}{\sigma^2 V(t)} \right|_y^\mu \\ &= \frac{y-\mu}{\sigma^2 V(\mu)}. \end{aligned}$$

Aplicando as condições abaixo de regularidade

$$(i) E \left\{ \frac{\partial Q(\mu; Y)}{\partial \mu} \right\} = 0 \quad e$$

$$(ii) E \left[\left\{ \frac{\partial Q(\mu; Y)}{\partial \mu} \right\}^2 \right] = -E \left\{ \frac{\partial^2 Q(\mu; Y)}{\partial \mu^2} \right\},$$

mostra-se facilmente que $E(Y) = \mu$ e $\text{Var}(Y) = \sigma^2 V(\mu)$. Ou seja, μ é a média da variável resposta e a variância de Y é proporcional a $V(\mu)$, como nos MLGs, embora nem sempre $V(\mu)$ seja uma função de variância. Uma terceira propriedade mostrada por Wedderburn (1974) é a seguinte:

$$(iii) -E \left\{ \frac{\partial^2 Q(\mu; Y)}{\partial \mu^2} \right\} \leq -E \left\{ \frac{\partial^2 L(\mu; Y)}{\partial \mu^2} \right\}.$$

Essa relação mostra que a informação a respeito de μ quando se conhece apenas a relação entre a variância e a média é menor do que a informação a respeito de μ quando se conhece a distribuição da resposta (informação de Fisher). Assim, a quantidade $E\{\partial^2(Q - L)/\partial \mu^2\}$ pode ser interpretada como o ganho quando acrescenta-se ao conhecimento da relação média-variância também o conhecimento da distribuição da resposta.

Dependendo das especificações de σ^2 e $V(\mu)$ poderemos recuperar a distribuição de Y . Abaixo são apresentados alguns exemplos.

Exemplos

NORMAL

Vamos supor $V(t) = 1$ e $-\infty < t, y < \infty$. Logo, o logaritmo da função de quase-verossimilhança fica dado por

$$Q(\mu; y) = \int_y^\mu \frac{y - t}{\sigma^2} dt = -\frac{(y - t)^2}{2\sigma^2} \Big|_y^\mu = -\frac{(y - \mu)^2}{2\sigma^2} + \text{cte},$$

que é proporcional ao logaritmo da função de verossimilhança de uma $N(\mu, \sigma^2)$ para σ^2 conhecido.

POISSON

Vamos supor $V(t) = t$ e $y \geq 0, t > 0$. Logo, obtemos

$$\begin{aligned} Q(\mu; y) &= \int_y^\mu \frac{y-t}{\sigma^2 t} dt \\ &= \frac{1}{\sigma^2} (y \log t - t) \Big|_y^\mu \\ &= \frac{1}{\sigma^2} \{y \log \mu - \mu - y \log y + y\} + \text{cte.} \end{aligned}$$

Se assumirmos $\sigma^2 = 1$ e $y > 0$ temos que $Q(\mu; y)$ é proporcional ao logaritmo da função de verossimilhança de uma $P(\mu)$.

Para $y = 0$ obtemos

$$Q(\mu; y) = \int_0^\mu \frac{-t}{\sigma^2 t} dt = \frac{-t}{\sigma^2} \Big|_0^\mu = -\frac{\mu}{\sigma^2} + \text{cte},$$

que coincide quando $\sigma^2 = 1$ com $\log P(Y = 0)$, em que $Y \sim P(\mu)$.

BINOMIAL

Supor a função $V(t) = t(1-t)$, $0 \leq y \leq 1$ e $0 < t < 1$. O logaritmo da função de quase-verossimilhança fica nesse caso dado por

$$\begin{aligned} Q(\mu; y) &= \int_y^\mu \frac{y-t}{\sigma^2 t(1-t)} dt \\ &= \frac{y}{\sigma^2} \int_y^\mu \frac{1}{t(1-t)} dt - \frac{1}{\sigma^2} \int_y^\mu \frac{1}{(1-t)} dt \\ &= \frac{y}{\sigma^2} \log \left(\frac{t}{1-t} \right) \Big|_y^\mu + \frac{1}{\sigma^2} \log(1-t) \Big|_y^\mu \\ &= \frac{y}{\sigma^2} [\log\{\mu(1-\mu) - \log\{y/(1-y)\}\}] + \frac{1}{\sigma^2} \{\log(1-\mu) - \log(1-y)\} + \text{cte}, \end{aligned}$$

para $0 < y, \mu < 1$.

Para $y = 0$ temos que

$$\begin{aligned}
 Q(\mu; y) &= \int_0^\mu \frac{-t}{\sigma^2 t(1-t)} dt \\
 &= -\frac{1}{\sigma^2} \int_0^\mu \frac{t}{t(1-t)} dt \\
 &= \frac{1}{\sigma^2} \log(1-t) \Big|_0^\mu \\
 &= \frac{1}{\sigma^2} \log(1-\mu) + \text{cte},
 \end{aligned}$$

que para $\sigma^2 = 1$ coincide com $\log P(Y = 0)$, em que $Y \sim Be(\mu)$.

Quando $y = 1$ segue que

$$\begin{aligned}
 Q(\mu; y) &= \int_1^\mu \frac{(1-t)}{\sigma^2 t(1-t)} dt \\
 &= \frac{1}{\sigma^2} \int_1^\mu \frac{1}{t} dt \\
 &= \frac{1}{\sigma^2} \log t \Big|_1^\mu \\
 &= \frac{1}{\sigma^2} \log \mu + \text{cte},
 \end{aligned}$$

que para $\sigma^2 = 1$ coincide com $\log P(Y = 1)$, em que $Y \sim Be(\mu)$.

GAMA

Supor a função $V(t) = t^2$ e $y, t > 0$. O logaritmo da função de quase-verossimilhança fica nesse caso dado por

$$\begin{aligned}
 Q(\mu; y) &= \int_y^\mu \frac{y-t}{\sigma^2 t^2} dt \\
 &= \frac{1}{\sigma^2} (-y/t - \log t) \Big|_y^\mu \\
 &= \frac{1}{\sigma^2} \{-y/\mu - \log \mu + 1 + \log y\} + \text{cte}.
 \end{aligned}$$

Para σ^2 conhecido temos que $Q(y; \mu)$ é proporcional ao logaritmo da função de verossimilhança de uma $G(\mu, \phi)$, em que $\phi = 1/\sigma^2$

FUNÇÃO $V(t) = t^2(1-t)^2$

Suponha $0 < t < 1$ e $0 \leq y \leq 1$. Nesse caso o logaritmo da função de quase-verossimilhança fica dada por

$$\begin{aligned} Q(\mu; y) &= \frac{1}{\sigma^2} \int_y^\mu \frac{y-t}{t^2(1-t)^2} dt \\ &\propto \frac{1}{\sigma^2} [(2y-1)\log\{\mu/(1-\mu)\} - y/\mu - (1-y)/(1-\mu)] + \text{cte.} \end{aligned}$$

A função $Q(\mu; y)$ obtida acima não corresponde a nenhuma função com verossimilhança conhecida. Portanto, apenas para algumas funções de quase-verossimilhança tem-se uma função de verossimilhança correspondente.

Em particular, para as funções $V(t) = t^3$, $t > 0$, $V(t) = t(1+t)$, $t > 0$ e $V(t) = e^{-t}$, $t \in \mathbb{R}$, é possível recuperar distribuições da família exponencial uniparamétrica, bem como definir novos modelos de quae-verossimilhança.

5.2 Respostas independentes

Vamos supor que Y_1, \dots, Y_n são variáveis aleatórias independentes com logaritmo da função de quase-verossimilhança $Q(\mu_i; y_i)$, $i = 1, \dots, n$. O logaritmo da função de quase-verossimilhança correspondente à distribuição conjunta fica dado por

$$Q(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^n Q(\mu_i; y_i). \quad (5.1)$$

Vamos supor ainda que

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}, \quad (5.2)$$

em que $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ contém valores de variáveis explicativas, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ e $g(\cdot)$ é uma função de ligação. Note que os MLGs são um caso particular de (5.1)-(5.2).

5.2.1 Estimação

Denotando $Q(\boldsymbol{\beta}) = Q(\boldsymbol{\mu}(\boldsymbol{\beta}); \mathbf{y})$, podemos mostrar que a função quase-escore para $\boldsymbol{\beta}$ fica expressa na forma

$$\mathbf{U}_\beta = \frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2} \mathbf{D}^T \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}),$$

em que $\mathbf{D} = \partial \boldsymbol{\mu} / \partial \boldsymbol{\beta} = \mathbf{W}^{1/2} \mathbf{V}^{1/2} \mathbf{X}$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$, $\mathbf{y} = (y_1, \dots, y_n)^T$, $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$, $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$ com $\omega_i = (d\mu/d\eta)_i^2 / V_i$ e \mathbf{X} é uma matriz $n \times p$ de linhas \mathbf{x}_i^T , $i = 1, \dots, n$. A matriz de quase-informação para $\boldsymbol{\beta}$ fica dada por

$$\mathbf{K}_{\beta\beta} = -E \left\{ \frac{\partial^2 Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right\} = \frac{1}{\sigma^2} \mathbf{D}^T \mathbf{V}^{-1} \mathbf{D}.$$

A estimativa de quase-verossimilhança para $\boldsymbol{\beta}$ sai da solução da equação $\mathbf{U}_\beta = \mathbf{0}$ que pode ser resolvida pelo método escore de Fisher resultando no seguinte processo iterativo:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \{\mathbf{D}^{(m)T} \mathbf{V}^{-(m)} \mathbf{D}^{(m)}\}^{-1} \mathbf{D}^{(m)T} \mathbf{V}^{-(m)} \{\mathbf{y} - \boldsymbol{\mu}^{(m)}\}, \quad (5.3)$$

$m = 0, 1, 2, \dots$. Note que o processo iterativo (5.3) não depende de σ^2 , no entanto, precisa ser iniciado numa quantidade $\boldsymbol{\beta}^{(0)}$. Mostra-se, sob certas condições de regularidade (vide, por exemplo, McCullagh e Nelder, 1989, p. 333), que $\hat{\boldsymbol{\beta}}$ é consistente e assintoticamente normal com matriz de variância-covariância dada por $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1}$. O parâmetro de dispersão σ^2 deve ser estimado separadamente.

5.2.2 Estimador de momentos

Podemos verificar facilmente que

$$\text{Var} \left\{ \frac{(Y_i - \mu_i)}{\sigma \sqrt{V(\mu_i)}} \right\} = 1,$$

e daí segue

$$\text{Var} \left\{ \frac{(Y_i - \mu_i)}{\sqrt{V(\mu_i)}} \right\} = \sigma^2,$$

e, portanto, um estimador de momentos para σ^2 fica dado por

$$\hat{\sigma}^2 = \frac{1}{(n-p)} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}.$$

5.2.3 Função quase-desvio

É possível definir uma função tipo desvio para os modelos de quase-verossimilhança de forma similar aos MLGs. Sejam $Q(\mathbf{y}; \mathbf{y})$ e $Q(\hat{\boldsymbol{\mu}}; \mathbf{y})$, respectivamente, as funções de quase-verossimilhança do modelo saturado e do modelo sob investigação. A função quase-desvio não escalonada é definida por

$$\begin{aligned} D(\mathbf{y}; \hat{\boldsymbol{\mu}}) &= 2\sigma^2 \{Q(\mathbf{y}; \mathbf{y}) - Q(\hat{\boldsymbol{\mu}}; \mathbf{y})\} \\ &= -2\sigma^2 Q(\hat{\boldsymbol{\mu}}; \mathbf{y}) = -2\sigma^2 \sum_{i=1}^n Q(\hat{\mu}_i; y_i) \\ &= 2 \sum_{i=1}^n \int_{\hat{\mu}_i}^{y_i} \frac{y_i - t}{V(t)} dt, \end{aligned}$$

que não depende de σ^2 . É natural que se compare a função quase-desvio escalonada $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sigma^{-2} D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ com os percentis da distribuição $\chi_{(n-p)}^2$, embora não seja em geral conhecida a distribuição nula de $\sigma^{-2} D(\mathbf{y}; \hat{\boldsymbol{\mu}})$. Apresentamos abaixo a função quase-desvio não escalonada para alguns casos particulares supondo uma única observação.

$V(\mu)$	Componente de $D(\mathbf{y}; \boldsymbol{\mu})$
μ	$-2\{y \log \mu - \mu - y \log y + y\}, y, \mu > 0$
$\mu(1 - \mu)$	$-2[y \log \{\mu/(1 - \mu)\} + \log(1 - \mu) - \log y], 0 < y, \mu < 1$
μ^2	$-2\{1 - y/\mu - \log \mu + \log y\}, y, \mu > 0$

5.2.4 Teste de hipóteses

Seja o vetor paramétrico β particionado tal que $\beta = (\beta_1^T, \beta_2^T)^T$, β_1 e β_2 são subvetores de dimensão q e $p - q$, respectivamente. Suponha que temos interesse em testar $H_0 : \beta_1 = \mathbf{0}$ contra $H_1 : \beta_1 \neq \mathbf{0}$. McCullagh (1983) mostra que também no caso de quase-verossimilhança a diferença entre duas funções quase-desvio funciona como um teste da razão de verossimilhanças. Ou seja, se denotarmos por $D(\mathbf{y}; \hat{\mu}^0)$ a função quase-desvio sob H_0 e por $D(\mathbf{y}; \hat{\mu})$ a função quase-desvio sob H_1 , para n grande e sob H_0 , temos que

$$\frac{1}{\sigma^2} \{D(\mathbf{y}; \hat{\mu}^0) - D(\mathbf{y}; \hat{\mu})\} \sim \chi_q^2,$$

para σ^2 fixo que pode ser estimado consistentemente, como ocorre com os MLGs. Testes tipo Wald e tipo escore são também possíveis de serem desenvolvidos. Usando resultados do Capítulo 1 podemos mostrar que

$$\text{Var}(\hat{\beta}_1) = \{\mathbf{D}_1^T \mathbf{V}^{1/2} \mathbf{M}_2 \mathbf{V}^{1/2} \mathbf{D}_1\}^{-1},$$

em que $\mathbf{M}_2 = \mathbf{I} - \mathbf{H}_2$, $\mathbf{H}_2 = \mathbf{V}^{1/2} \mathbf{D}_2 (\mathbf{D}_2^T \mathbf{V} \mathbf{D}_2)^{-1} \mathbf{D}_2^T \mathbf{V}^{1/2}$, $\mathbf{D}_1 = \mathbf{W}^{1/2} \mathbf{V}^{1/2} \mathbf{X}_1$ e $\mathbf{D}_2 = \mathbf{W}^{1/2} \mathbf{V}^{1/2} \mathbf{X}_2$. Assim, um teste tipo Wald fica dado por

$$\xi_W = \hat{\beta}_1^T \hat{\text{Var}}^{-1}(\hat{\beta}_1) \hat{\beta}_1,$$

em que $\hat{\text{Var}}(\hat{\beta}_1)$ denota que a variância está sendo avaliada em $\hat{\beta}$. Já o teste quase-escore para testar $H_0 : \beta_1 = \mathbf{0}$ contra $H_1 : \beta_1 \neq \mathbf{0}$ fica dado por

$$\xi_{SR} = \mathbf{U}_{\beta_1}(\hat{\beta}^0)^T \hat{\text{Var}}_0(\hat{\beta}_1) \mathbf{U}_{\beta_1}(\hat{\beta}^0),$$

em que

$$\begin{aligned} \mathbf{U}_{\beta_1} &= \frac{\partial Q(\beta)}{\partial \beta_1} \\ &= \frac{1}{\sigma^2} \mathbf{D}_1^T \mathbf{V}^{-1}(\mathbf{y} - \mu), \end{aligned}$$

com todas as quantidades sendo avaliadas em $\hat{\beta}^0 = (\mathbf{0}^T, \hat{\beta}_2^{0T})^T$ e $\hat{\beta}_2^0$ sendo a estimativa de β_2 sob H_0 . Sob H_0 e sob condições usuais de regularidade temos que, para $n \rightarrow \infty$, $\xi_W, \xi_{SR} \sim \chi_q^2$.

5.2.5 Resíduos

O não conhecimento da verdadeira função de verossimilhança de β dificulta o desenvolvimento de alguns métodos de diagnóstico. Tanto o estudo de resíduos como de medidas de influência dependem em geral do conhecimento de $L(\beta)$. O que tem sido proposto em modelos de quase-verossimilhança no sentido de avaliar a qualidade do ajuste são gráficos de resíduos. Uma sugestão (vide McCullagh e Nelder, 1989, Cap. 9) é o gráfico do resíduo de Pearson

$$\hat{r}_{P_i} = \frac{y_i - \hat{\mu}_i}{\hat{\sigma} \sqrt{V(\hat{\mu}_i)}}$$

contra alguma função dos valores ajustados, como por exemplo contra $g(\hat{\mu}_i)$, em que $g(\cdot)$ é a função de ligação. Espera-se uma distribuição aleatória dos resíduos em torno do eixo zero. Tendências diferentes, como por exemplo aumento da variabilidade, podem indicar que a função $V(\mu_i)$ não é adequada. Um outro resíduo que pode também ser utilizado, embora de forma descritiva, é dado por

$$t_{D_i} = \frac{\pm d(y_i; \hat{\mu}_i)}{\hat{\sigma} \sqrt{1 - \hat{h}_{ii}}},$$

em que $d(y_i; \hat{\mu}_i)$ é a raiz quadrada com sinal de $y_i - \hat{\mu}_i$ do i -ésimo componente do quase-desvio $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$, enquanto h_{ii} é o i -ésimo elemento da diagonal principal da matriz

$$\mathbf{H} = \mathbf{V}^{-1/2} \mathbf{D} (\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{V}^{-1/2}.$$

5.2.6 Influência

Uma versão da distância de Cook para os modelos de quase-verossimilhança fica dada por

$$LD_i = \frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})^2} \hat{r}_{P_i}^2,$$

em que \hat{r}_{P_i} é o resíduo de Pearson e \hat{h}_{ii} denota o i -ésimo elemento da diagonal principal da matriz $\hat{\mathbf{H}}$. Gráficos de LD_i contra a ordem das observações ou contra os valores ajustados podem revelar pontos possivelmente influentes nos parâmetros do preditor linear.

5.2.7 Seleção de Modelos

Uma extensão natural do método de Akaike para os modelos de quase-verossimilhança (ver, por exemplo, Pan, 2001) é considerar

$$AIC = -2Q(\hat{\beta}) + 2p,$$

em que $Q(\hat{\beta})$ é a função de quase-verossimilhança avaliada em $\hat{\beta}$.

5.2.8 Aplicações

MOSCA DO CHIFRE

No arquivo **mosca.dat** é apresentado parte dos dados de um experimento desenvolvido para estudar a distribuição do número de ácaros em placas de esterco de gado bovino no estado de S. Paulo (Paula e Tavares, 1992). Essas placas são depósitos de ovos da mosca do chifre (*Haematobia irritans*), uma das pragas mais importantes da pecuária brasileira. Os ácaros são inimigos naturais da mosca do chifre uma vez que se alimentam de ovos e larvas dessas moscas. No arquivo **mosca.dat** tem-se a distribuição do número de ácaros de quatro espécies segundo algumas variáveis de interesse: (i) N , número de

partes da posição da placa onde foram coletados os ácaros, (ii) **Posição**, posição na placa onde foram coletados os ácaros (1: lateral, 0: central), (iii) **Região**, região onde a placa foi coletada (1: São Roque, 2: Pindamonhangaba, 3: Nova Odessa e 4: Ribeirão Preto) e (iv) **Temp**, temperatura no local da coleta (em $^{\circ}C$).

Tabela 5.1

Estimativas dos parâmetros do modelo de quase-verossimilhança com função $V(\mu) = \mu^2$ ajustado aos dados sobre a mosca do chifre.

Efeito	Com todos os pontos		Sem pontos aberrantes	
	Estimativa	E/E.Padrão	Estimativa	E/E.Padrão
Constante	-0,828	-0,74	-2,575	-2,13
Posição	-0,288	-0,64	0,380	0,78
Pinda	-0,424	-0,66	-0,910	-1,31
N. Odessa	-1,224	-1,71	-1,836	-2,36
R. Preto	-2,052	-2,98	-2,589	-3,46
Temp.	0,029	0,67	0,087	1,84
σ^2	5,129		5,913	

Pensou-se inicialmente, como trata-se de dados de contagem, num modelo log-linear de Poisson para explicar o número médio de ácaros segundo as variáveis explicativas. Denotando por Y_{ijk} o número de ácaros coletados na i -ésima posição da k -ésima placa e j -ésima região, vamos supor que $Y_{ijk} \sim P(\mu_{ijk})$, $\mu_{ijk} = N_{ijk}\lambda_{ijk}$, $i = 1, 2$ e $j = 1, \dots, 6$, com N_{ijk} denotando o número de partes na i -ésima posição da k -ésima placa coletada na j -ésima região. A parte sistemática do modelo fica dada por

$$\log \mu_{ijk} = \log N_{ijk} + \log \lambda_{ijk}, \quad (5.4)$$

em que

$$\log \lambda_{ijk} = \alpha + \beta_i + \gamma_j + \delta \text{Temp}_{jk}, \quad (5.5)$$

$\log N_{ijk}$ desempenha papel de *offset*, β_i denota o efeito da posição, γ_j o efeito da região e Temp_{jk} a temperatura na j -ésima região no momento da coleta da k -ésima placa. Temos as restrições $\beta_1 = \gamma_1 = 0$. O desvio do modelo ajustado para a espécie 6 foi de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 318,69$ (96 graus de liberdade) indicando fortes indícios de sobredispersão. Propomos então um modelo de quase-verossimilhança com função dada por $V(\mu_{ijk}) = \mu_{ijk}$. Esse modelo parece também inadequado pelo gráfico de resíduos de Pearson $\hat{r}_{P_{ijk}} = (y_{ijk} - \hat{\mu}_{ijk}) / \hat{\sigma} \sqrt{\hat{\mu}_{ijk}}$ contra $\log \hat{\mu}_{ijk}$ (Figura 5.1).

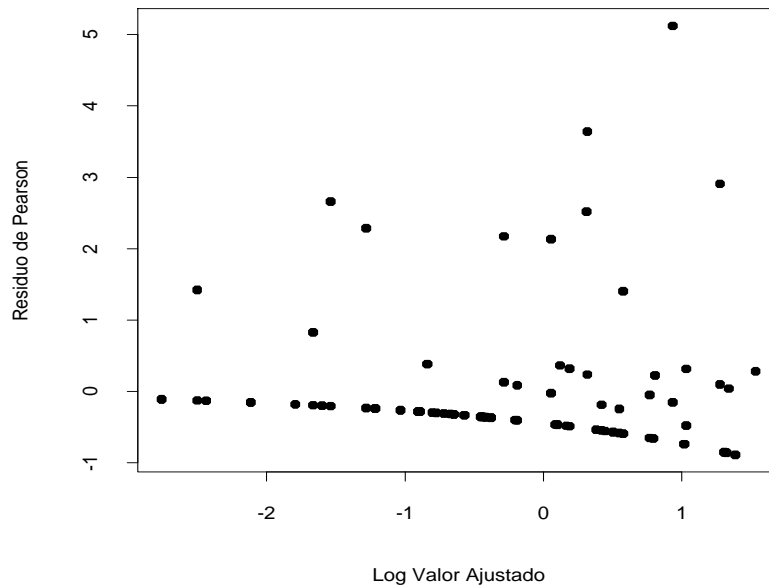


Figura 5.1: Gráfico do resíduo de Pearson contra $\log \hat{\mu}$ para o modelo ajustado com função $V(\mu) = \mu$ aos dados sobre a mosca do chifre.

Nota-se um aumento da variabilidade com o aumento do logaritmo das médias ajustadas, indício de que a variabilidade não foi totalmente controlada. Para ajustar o modelo no R, vamos supor que as variáveis *Posição*, *Região* e *Temp* sejam colocadas em *posicao*, *regiao* e *temp*, respectivamente,

e que $\log N$ denota o logaritmo do número de partes da placa. O número de ácaros será denotado por `acaros`. A sequência de comandos é dada abaixo

```
regiao = factor(regiao)

fit1.mosca = glm(acaros ~ posicao + regiao + temp +
offset(logN), family=quasi(link=log, variance= "mu")).
```

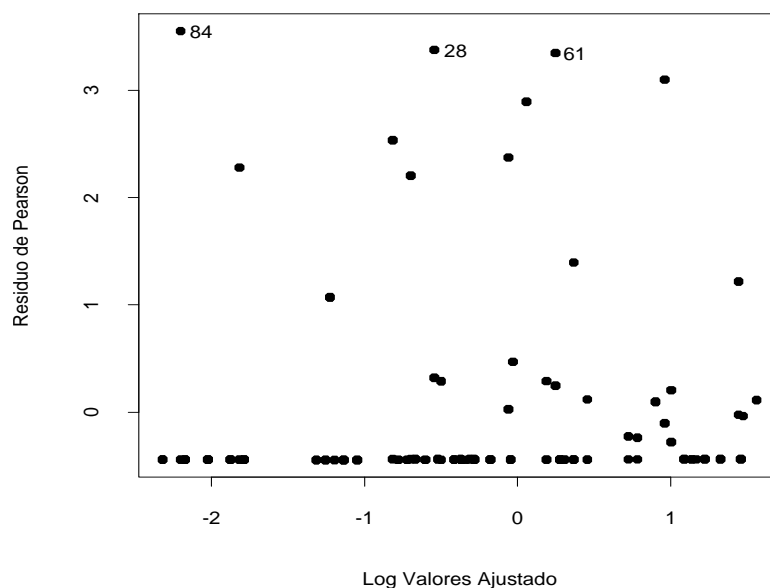


Figura 5.2: Gráfico do resíduo de Pearson contra $\log \hat{\mu}$ para o modelo ajustado com função $V(\mu) = \mu^2$ aos dados sobre a mosca do chifre.

Se colocarmos em `phi` a estimativa do parâmetro de dispersão, o resíduo de Pearson padronizado será obtido pelo comando

```
phi = summary(fit1.mosca)$dispersion
rp = resid(fit.mosca, type = "pearson")/sqrt(phi).
```

No objeto `fit.mosca` estão os principais resultados do ajuste. Propomos agora, a fim de controlar a variabilidade, um modelo de quase-verossimilhança

com função quadrática $V(\mu_{ijk}) = \mu_{ijk}^2$ e parte sistemática dada por (5.4)-(5.5). O gráfico do resíduo de Pearson contra o logaritmo das médias ajustadas (Figura 5.2) parece bastante razoável, embora apareçam 9 placas com valores para $\hat{r}_{P_{ijk}}$ acima de 2. Na Tabela 5.1 apresentamos as estimativas dos parâmetros com todas as placas e também eliminando as placas com resíduos mais aberrantes, #28, #61 e #84.

Os comandos no R para ajustar os dois modelos são dados abaixo

```
fit1.mosca = glm(acaros ~ posicao + regioao + temp +
offset(logN), family=quasi(link=log, variance= "mu^ 2"), maxit=50)

fit2.mosca = glm(acaros ~ posicao + regioao + temp +
offset(logN), family=quasi(link=log, variance= "mu^ 2 "), subset
= -c(28,61,84), maxit=50).
```

Nota-se pelas estimativas dos dois modelos ajustados que Nova Odessa e Ribeirão Preto apresentam um número médio de ácaros bem menor do que as outras duas regiões. Não há indícios de efeito de posição, porém a eliminação das três placas com valores mais aberrantes faz com que o efeito de temperatura fique mais acentuado, havendo indícios de que o número médio de ácaros cresce com o aumento da temperatura.

As placas #28, #61 e #84 têm em comum o fato de apresentarem um número médio de ácaros (por parte de placa) pelo menos duas vezes acima da média em temperaturas relativamente baixas. Essas placas foram coletadas nas regiões de Pindamonhangaba, Nova Odessa e Ribeirão Preto, respectivamente. Assim, é esperado que a eliminação dessas placas reduza o valor das estimativas dos efeitos dessas regiões como também aumente a estimativa do coeficiente da temperatura. A fim de que as 9 placas com resíduos mais aberrantes possam ser melhor ajustadas pode-se tentar outras formas para a função $V(\mu)$, como por exemplo $V(\mu) = \mu^2(1 + \mu)^2$ (vide Paula e Tavares,

1992).

DEMANDA DE TV A CABO

Vamos reanalisar nesta seção o exemplo sobre demanda de TV a cabo discutido no Capítulo 4 sob um enfoque de modelo log-linear com resposta binomial negativa. Proporemos aqui um modelo um pouco diferente. Ao invés de ser ajustado o número médio esperado de assinantes de TV a cabo será ajustada a proporção esperada de assinantes de TV a cabo em cada área. A proporção observada é dada por $\text{Razao} = \text{Nass}/\text{Domic}$. Como $0 \leq \text{Razao} \leq 1$, propomos o seguinte modelo de quase-verossimilhança:

$$\begin{aligned} E(\text{Razao}_i) &= \pi_i \text{ e} \\ \text{Var}(\text{Razao}_i) &= \sigma^2 \pi_i (1 - \pi_i), \end{aligned}$$

em que π_i denota a proporção esperada de assinantes na i -ésima área, $i = 1, \dots, 40$. A parte sistemática do modelo será dada por

$$\log \left\{ \frac{\pi_i}{1 - \pi_i} \right\} = \alpha + \beta_1 \text{Percap}_i + \beta_2 \text{Taxa}_i + \beta_3 \text{Custo}_i + \beta_4 \text{Ncabo}_i + \beta_5 \text{Ntv}_i.$$

Na Figura 5.3 é apresentado o gráfico da distância de Cook contra das observações com destaque para as áreas #5 e #14. A observação #5 corresponde a uma área de renda alta porém com uma proporção pequena de assinantes de TV a cabo, talvez devido aos altos custos de instalação e manutenção. Já a área #14 tem uma proporção alta de assinantes de TV a cabo embora as taxas também sejam altas. Também na Figura 5.3 temos o gráfico do resíduo $\hat{r}_{P_i} = (\text{Razao}_i - \hat{\pi}_i) / \hat{\sigma} \sqrt{\hat{\pi}_i (1 - \hat{\pi}_i)}$ contra o logito dos valores ajustados e como pode-se notar há um ligeiro aumento da variabilidade com o aumento da proporção de áreas com o TV a cabo.

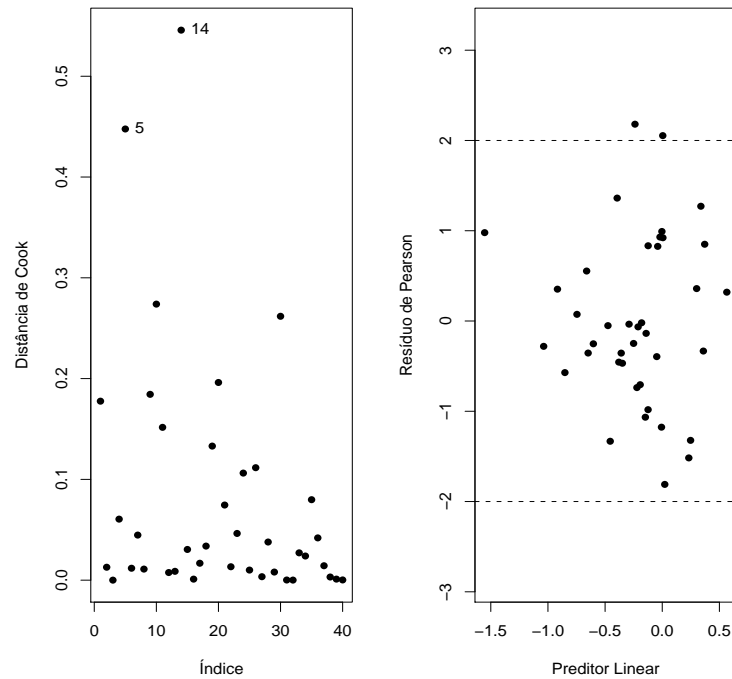


Figura 5.3: Distância de Cook e gráfico do resíduo de Pearson contra o logito de $\hat{\pi}$ para o modelo ajustado com função $V(\pi) = \pi(1 - \pi)$ aos dados sobre demanda de TV a cabo.

Tabela 5.2

Estimativas dos parâmetros do modelo de quase-verossimilhança com função $V(\pi) = \pi(1 - \pi)$ ajustado aos dados sobre demanda de TV a cabo.

Efeito	Com todos os pontos		Sem áreas 5 e 14	
	Estimativa	E/E.Padrão	Estimativa	E/E.Padrão
Intercepto	-2,407	-1,72	-2,440	-1,60
Percap	4×10^{-4}	2,50	4×10^{-4}	2,80
Taxa	0,023	0,93	0,016	0,64
Custo	-0,203	-1,79	-0,252	-2,27
Ncabo	0,073	1,94	0,079	2,22
Ntv	-0,216	-2,61	-0,201	-2,61
σ^2	0,114		0,098	

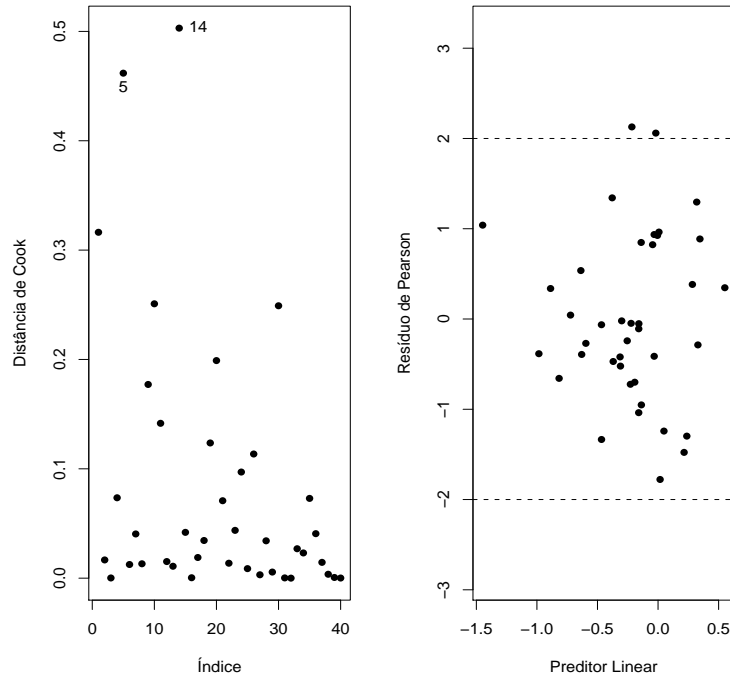


Figura 5.4: Distância de Cook e gráfico do resíduo de Pearson contra o logito de $\hat{\pi}$ para o modelo ajustado com função $V(\pi) = \pi^2(1 - \pi)^2$ aos dados sobre demanda de TV a cabo.

A eliminação dessas duas áreas, como pode ser observado pela Tabela 5.2, não altera os resultados inferenciais (ao nível de 5%) com todas as observações, embora aumente a significância dos coeficientes. Nota-se que apenas o coeficiente da variável **Taxa** parece não ser significativo marginalmente.

Uma tentativa no sentido de tentar reduzir a variabilidade observada na Figura 5.3 é utilizando uma função do tipo $V(\pi) = \pi^2(1 - \pi)^2$. Na Figura 5.4 temos o gráfico da distância de Cook e o gráfico do resíduo de Pearson contra o logito dos valores ajustados supondo $V(\pi) = \pi^2(1 - \pi)^2$. Nota-se comportamentos muito similares àqueles encontrados na Figura 5.3. Assim, podemos assumir para esse exemplo o ajuste com a função $V(\pi) = \pi(1 - \pi)$.

Nota-se, que sob esse ajuste, mais variáveis permanecem no modelo do que sob o ajuste do número esperado de domicílios com TV a cabo com resposta binomial negativa, como foi visto no Capítulo 4.

Para o ajuste do modelo de quase-verossimilhança com $V(\pi) = \pi^2(1 - \pi)^2$ é preciso requerer a *library* `gnm` e usar a família `wedderburn` conforme os comandos dados abaixo

```
require(gnm)

ajuste.tvcabo = glm(razao ~ percap + taxa + custo + ncabo + ntv,
family=wedderburn).
```

Todavia, os resultados com a família `wedderburn` ficaram muito parecidos com aqueles resultados apresentados com a função $V(\pi) = \pi(1 - \pi)$.

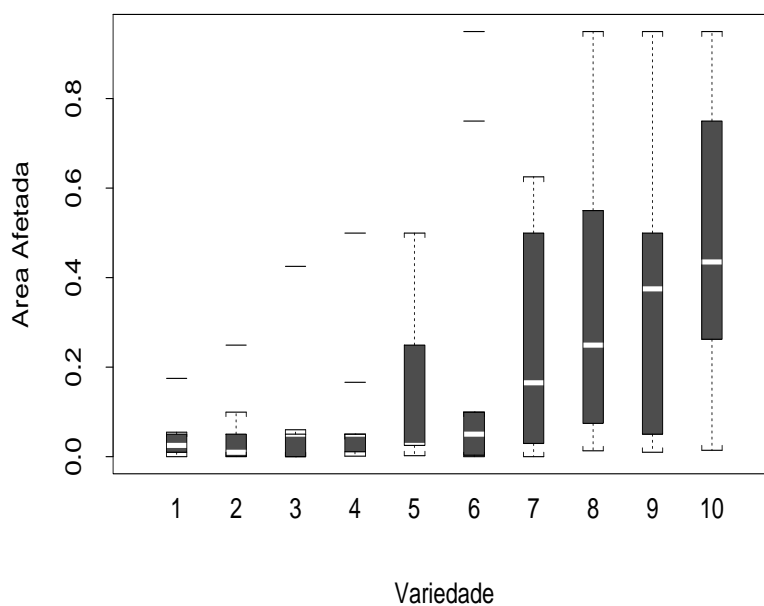


Figura 5.5: Boxplots da proporção da área afetada segundo a variedade para os dados sobre manchas na folha da cevada.

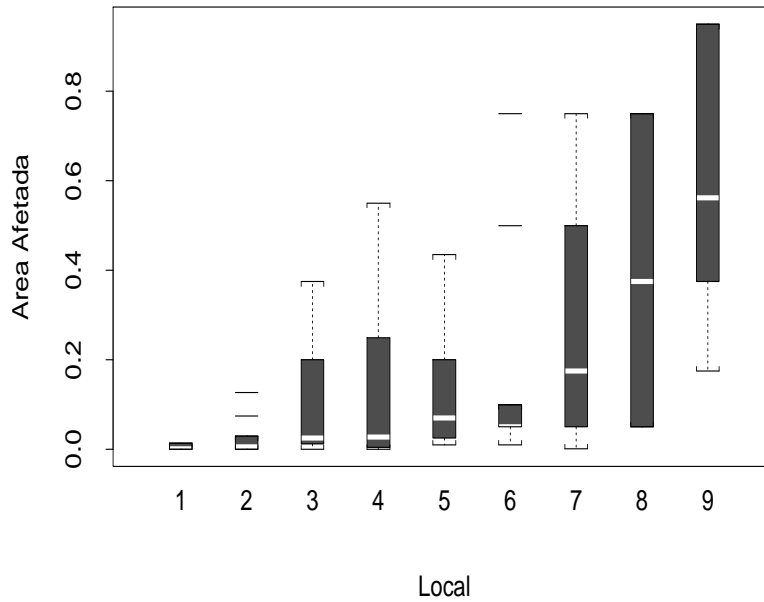


Figura 5.6: Boxplots da proporção da área afetada segundo o local para os dados sobre manchas na folha da cevada.

MANCHAS NA FOLHA DA CEVADA

Esses dados estão descritos em McCullagh e Nelder (1982, Tabela 9.2) e no arquivo **cevada.dat**, em que a incidência de um tipo de mancha é observada na folha da cevada segundo 10 variedades em 9 locais diferentes. A amostra consiste de 90 observações em que a resposta é a área afetada da folha (em proporção) e os fatores são a variedade e o local.

Nas Figuras 5.5 e 5.6 são apresentados os boxplots da área afetada (em proporção) segundo a variedade e local, respectivamente. Nota-se no primeiro gráfico um aumento da mediana da proporção da área afetada e também da dispersão com a variedade. Tendência similar pode ser observada no segundo gráfico. Seja Y_{ij} a proporção da área afetada da folha da cevada correspondente ao i -ésimo local e j -ésima variedade para $i = 1, \dots, 9$ e $j = 1, \dots, 10$. Conforme sugerido por McCullagh e Nelder (1989, Cap. 9) vamos

supor o seguinte modelo de quase-verossimilhança:

$$\begin{aligned} E(Y_{ij}) &= \pi_{ij} \text{ e} \\ \text{Var}(Y_{ij}) &= \sigma^2 V(\pi_{ij}), \end{aligned}$$

com parte sistemática dada por

$$\log \left\{ \frac{\pi_{ij}}{1 - \pi_{ij}} \right\} = \alpha + \beta_i + \gamma_j,$$

em que π_{ij} denota a proporção esperada da área afetada para a j -ésima variedade do i -ésimo local, $\beta_1 = 0$ e $\gamma_1 = 0$.

Nas Figuras 5.7 e 5.8 são apresentados gráficos de diagnóstico para ajustes do modelo de quase-verossimilhança supondo $V(\pi_{ij}) = \pi_{ij}(1 - \pi_{ij})$ e $V(\pi_{ij}) = \pi_{ij}^2(1 - \pi_{ij})^2$, respectivamente. Nota-se ao compararmos os gráficos de resíduos que o segundo ajuste é mais adequado embora algumas observações sejam destacadas como possivelmente influentes. As observações #24, #65 e #76 apresentam proporções amostrais acima das proporções médias amostrais das variedades e locais correspondentes, enquanto a observação #52 tem uma proporção amostral abaixo da proporção média do local correspondente. A eliminação dessas observações não muda a inferência com relação às proporções médias dos locais, porém muda a inferência com relação às menores proporções médias das variedades. Em geral as estimativas de quase-verossimilhança indicam um aumento da proporção esperada da área afetada com o aumento da variedade e do local conforme descrito nos boxplots apresentados nas Figuras 5.5 e 5.6.

5.3 Classe estendida

O logaritmo da função de quase-verossimilhança $Q(\mu; y)$ assume que a função $V(\mu)$ é conhecida, logo a mudança dessa função significa que um novo modelo está sendo definido. No sentido de permitir comparações de diferentes

funções $V(\mu)$ para um mesmo modelo como também possibilitar a obtenção de uma estimativa para o erro padrão assintótico de $\hat{\sigma}^2$, Nelder e Pregibon (1987) propuseram uma (log) quase-verossimilhança estendida, definida por

$$Q^+(\mu; y) = -\frac{1}{2\sigma^2}D(y; \mu) - \frac{1}{2}\log\{2\pi\sigma^2V(y)\},$$

em que $D(y; \mu) = 2 \int_{\mu}^y \{(y-t)/V(t)\}dt$ é o quase-desvio e $\phi = \frac{1}{\sigma^2}$ o parâmetro de dispersão.

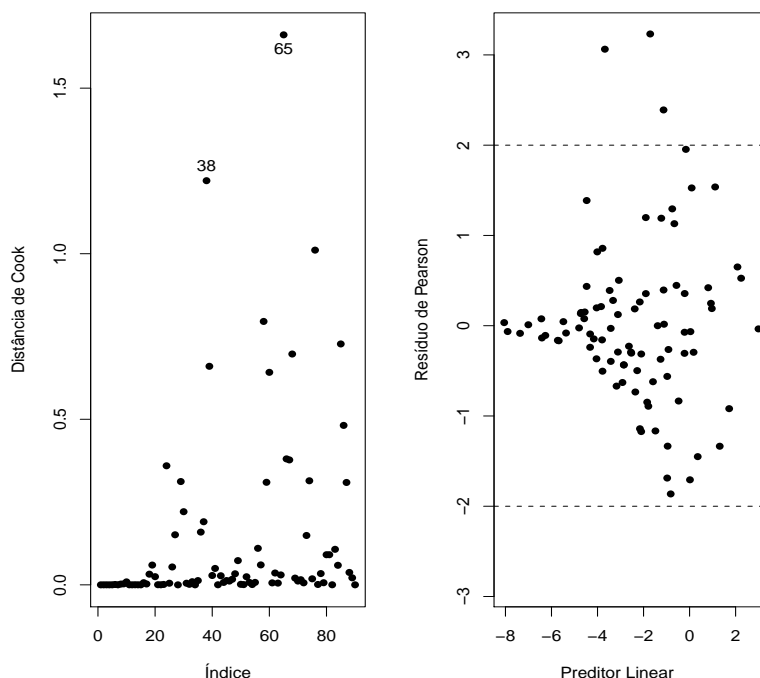


Figura 5.7: Distância de Cook e gráfico do resíduo de Pearson contra o logito de $\hat{\pi}$ para o modelo ajustado com função $V(\pi) = \pi(1 - \pi)$ aos dados sobre manchas na folha da cevada.

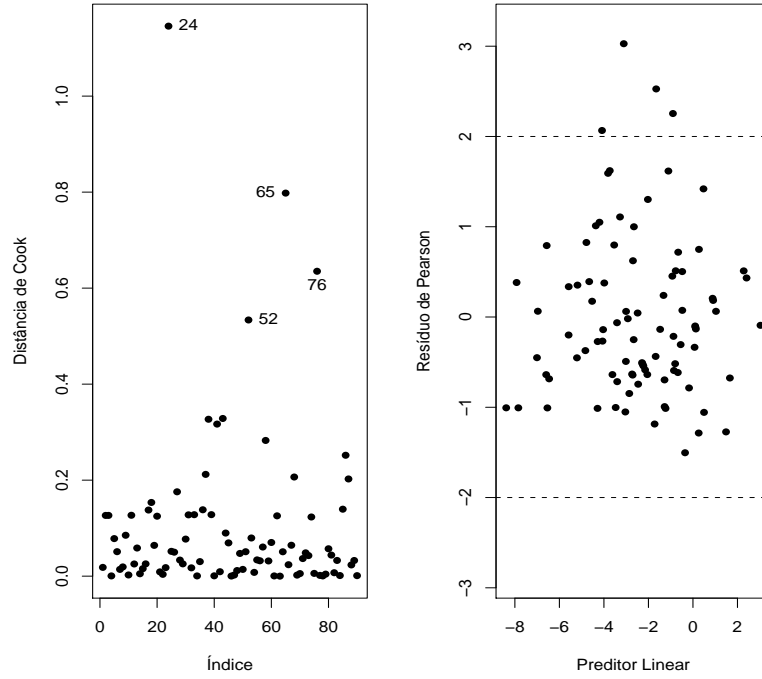


Figura 5.8: Distância de Cook e gráfico do resíduo de Pearson contra o logito de $\hat{\pi}$ para o modelo ajustado com função $V(\pi) = \pi^2(1 - \pi)^2$ aos dados sobre manchas na folha da cevada.

Similarmente a Q , Q^+ não pressupõe que a distribuição completa de Y seja conhecida, mas somente os dois primeiros momentos. A estimativa de β maximizando-se $Q^+(\mathbf{y}; \mu)$, para uma amostra aleatória de tamanho n , coincide com a estimativa de quase-verossimilhança para β , uma vez que Q^+ é uma função linear de Q . A estimativa de ϕ maximizando Q^+ é dada por $\hat{\phi} = D(\mathbf{y}; \hat{\mu})/n$. Portanto, para os casos especiais em que Q^+ corresponde às distribuições normal e normal inversa, $\hat{\phi}$ corresponde à estimativa de máxima verossimilhança de ϕ . Para a distribuição gama, Q^+ difere do logaritmo da função de verossimilhança por um fator dependendo somente de ϕ . Para as

distribuições de Poisson, binomial e binomial negativa, Q^+ é obtida do logaritmo da função de verossimilhança correspondente substituindo qualquer fatorial $k!$ pela aproximação de Stirling $k! \cong (2\pi k)^{1/2} k^k e^{-k}$. Discussões mais interessantes e aplicações da classe estendida são dadas em Nelder e Pregibon (1987).

5.4 Respostas correlacionadas

A fim de estabelecermos a notação a ser utilizada nesta seção, denotaremos por $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ir_i})^T$ o vetor resposta multivariado para a i -ésima unidade experimental, $i = 1, \dots, n$, e assumiremos em princípio que apenas é conhecida a distribuição marginal de Y_{it} , dada por

$$f(y; \theta_{it}, \phi) = \exp[\phi\{y\theta_{it} - b(\theta_{it})\} + c(y, \phi)], \quad (5.6)$$

em que $E(Y_{it}) = \mu_{it} = b'(\theta_{it})$, $\text{Var}(Y_{it}) = \phi^{-1}V_{it}$, $V_{it} = d\mu_{it}/d\theta_{it}$ é a função de variância e $\phi^{-1} > 0$ é o parâmetro de dispersão, em geral desconhecido. Podemos definir um modelo linear generalizado para cada instante t acrescentando a (5.6) a parte sistemática

$$g(\mu_{it}) = \eta_{it}, \quad (5.7)$$

em que $\eta_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta}$ é o preditor linear, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é um vetor de parâmetros desconhecidos a serem estimados, $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})^T$ representa os valores de variáveis explicativas observadas para a i -ésima unidade experimental no tempo t e $g(\cdot)$ é a função de ligação.

A função score e a matrix de informação para $\boldsymbol{\beta}$, ignorando-se a estrutura de correlação intraunidade experimental, ficam, respectivamente, dadas por

$$\mathbf{U}_{\boldsymbol{\beta}} = \phi \sum_{i=1}^n \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \quad (5.8)$$

e

$$\mathbf{K}_{\beta\beta} = \phi \sum_{i=1}^n \mathbf{D}_i^T \mathbf{V}_i \mathbf{D}_i, \quad (5.9)$$

em que $\mathbf{D}_i = \mathbf{W}_i^{1/2} \mathbf{V}_i^{1/2} \mathbf{X}_i$, \mathbf{X}_i é uma matriz $r_i \times p$ de linhas \mathbf{x}_{it}^T , $\mathbf{W}_i = \text{diag}\{\omega_{i1}, \dots, \omega_{ir_i}\}$ é a matriz de pesos com $\omega_{it} = (d\mu_{it}/d\eta_{it})^2/V_{it}$, $\mathbf{V}_i = \text{diag}\{V_{i1}, \dots, V_{ir_i}\}$, $\mathbf{y}_i = (y_{i1}, \dots, y_{ir_i})^T$ e $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{ir_i})^T$. Quando há ligação canônica a função escore e a matriz de informação de Fisher ficam dadas por $\mathbf{U}_\beta = \phi \sum_{i=1}^n \mathbf{X}_i^T (\mathbf{y}_i - \boldsymbol{\mu}_i)$ e $\mathbf{K}_{\beta\beta} = \phi \sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_i \mathbf{X}_i$, respectivamente. O estimador de $\boldsymbol{\beta}$, ignorando-se a estrutura de correlação intraunidade experimental, sai da equação $\mathbf{U}_\beta = \mathbf{0}$. Esse estimador é consistente e assintoticamente normal. Note que podemos supor que a distribuição marginal de Y_{it} é desconhecida assumindo uma função $V(\mu_{it})$ diferente daquela que caracteriza a distribuição de Y_{it} . Nesse caso, teremos um modelo de quase-verossimilhança em cada instante t com função escore e matriz de informação, ignorando-se a estrutura de correlação, dadas por (5.8) e (5.9), respectivamente.

Um tópico de pesquisa importante, que tem interessado a vários pesquisadores, é o desenvolvimento de metodologias para a estimação dos parâmetros de interesse quando os dados são correlacionados e a distribuição marginal não é normal, como é o caso introduzido nesta seção. Uma maneira de resolver o problema é ignorar a estrutura de correlação, como vimos acima, produzindo estimadores consistentes e assintoticamente normais, porém muitas vezes com perda de eficiência. Uma outra maneira, que descreveremos a seguir, é introduzindo alguma estrutura de correlação na função escore, produzindo um novo sistema de equações para estimar $\boldsymbol{\beta}$. A fim de facilitarmos o entendimento dessa metodologia, vamos supor inicialmente que os dados são não correlacionados e que a matriz de correlação correspondente ao i -ésimo grupo é denotada por \mathbf{R}_i . Logo, teremos $\mathbf{R}_i = \mathbf{I}_{r_i}$. A matriz de

variância-covariância para \mathbf{Y}_i , por definição, é dada por

$$\text{Var}(\mathbf{Y}_i) = \phi^{-1} \mathbf{V}_i^{1/2} \mathbf{R}_i \mathbf{V}_i^{1/2}, \quad (5.10)$$

que no caso de dados não correlacionados fica simplesmente dada por $\phi^{-1} \mathbf{V}_i$. A ideia é introduzirmos em (5.10) uma matriz de correlação não diagonal, por exemplo dada por $\mathbf{R}_i(\boldsymbol{\beta})$, com reflexos na função escore que passaria a depender também de $\mathbf{R}_i(\boldsymbol{\beta})$. O inconveniente dessa proposta é o fato da correlação, que é restrita ao intervalo $[-1, 1]$, depender de $\boldsymbol{\beta}$, o que aumentaria a complexidade do processo de estimação. A solução encontrada para contornar esse problema foi dada por Liang e Zeger (1986) que propuseram uma matriz de correlação dada por $\mathbf{R}_i(\boldsymbol{\rho})$, em que $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)^T$ é um vetor de parâmetros de perturbação que não dependem de $\boldsymbol{\beta}$. Ou seja, os parâmetros da matriz de correlação não dependem dos parâmetros de posição.

Para entender melhor essa proposta definimos

$$\boldsymbol{\Omega}_i = \phi^{-1} \mathbf{V}_i^{1/2} \mathbf{R}_i(\boldsymbol{\rho}) \mathbf{V}_i^{1/2},$$

em que $\boldsymbol{\Omega}_i$ é a matriz de variância-covariância de \mathbf{Y}_i se a verdadeira correlação entre os elementos de \mathbf{Y}_i for dada por $\mathbf{R}_i(\boldsymbol{\rho})$. Note que $\mathbf{R}_i(\boldsymbol{\rho})$ é uma matriz $r_i \times r_i$ que depende de um número finito de parâmetros $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)^T$, sendo denominada matriz trabalho. Para estimarmos $\boldsymbol{\beta}$ devemos resolver o seguinte sistema de equações:

$$\mathbf{S}_\beta(\hat{\boldsymbol{\beta}}_G) = \mathbf{0}, \quad (5.11)$$

denominado equações de estimação generalizadas (EEGs), em que

$$\mathbf{S}_\beta(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i).$$

Note que (5.11) reduz-se a $\mathbf{U}_\beta = \mathbf{0}$ quando $\mathbf{R}_i(\boldsymbol{\rho}) = \mathbf{I}_{r_i}$, isto é, quando é ignorada a estrutura de correlação intraunidade experimental. Na verdade $\mathbf{S}_\beta(\boldsymbol{\beta})$

depende também de ϕ e $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)^T$ que são estimados separadamente de $\boldsymbol{\beta}$.

5.4.1 Estimação

O processo iterativo para a estimação de $\boldsymbol{\beta}$, que é uma modificação do método escore de Fisher, é dado por

$$\begin{aligned} \boldsymbol{\beta}_G^{(m+1)} &= \boldsymbol{\beta}_G^{(m)} + \left\{ \sum_{i=1}^n \mathbf{D}_i^{(m)T} \boldsymbol{\Omega}_i^{-(m)} \mathbf{D}_i^{(m)} \right\}^{-1} \times \\ &\quad \left[\sum_{i=1}^n \mathbf{D}_i^{(m)T} \boldsymbol{\Omega}_i^{-(m)} \{ \mathbf{y}_i - \boldsymbol{\mu}_i^{(m)} \} \right], \end{aligned} \quad (5.12)$$

$m = 0, 1, 2, \dots$. As estimativas $\hat{\phi}$ e $\hat{\boldsymbol{\rho}}$ são dadas inicialmente e modificadas separadamente a cada passo do processo iterativo.

Supondo que $\hat{\boldsymbol{\rho}}$ e $\hat{\phi}$ são estimadores consistentes de $\boldsymbol{\rho}$ e ϕ , respectivamente, temos que

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) \rightarrow_d N_p(\mathbf{0}, \boldsymbol{\Sigma}),$$

em que

$$\boldsymbol{\Sigma} = \lim_{n \rightarrow \infty} \left[n \left(\sum_{i=1}^n \mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \right)^{-1} \left\{ \sum_{i=1}^n \mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} \text{Var}(\mathbf{Y}_i) \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \right\} \left(\sum_{i=1}^n \mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \right)^{-1} \right].$$

Se a matriz de correlação $\mathbf{R}_i(\boldsymbol{\rho})$ é definida corretamente, então um estimador consistente para $\text{Var}(\hat{\boldsymbol{\beta}}_G)$ é dado por $\mathbf{H}_1^{-1}(\hat{\boldsymbol{\beta}}_G)$, em que

$$\mathbf{H}_1(\hat{\boldsymbol{\beta}}_G) = \sum_{i=1}^n (\hat{\mathbf{D}}_i^T \hat{\boldsymbol{\Omega}}_i^{-1} \hat{\mathbf{D}}_i),$$

com $\hat{\mathbf{D}}_i$ sendo avaliado em $\hat{\boldsymbol{\beta}}_G$ e $\hat{\boldsymbol{\Omega}}_i$ avaliado em $(\hat{\phi}, \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\beta}}_G)$. Entretanto, se a matriz trabalho $\mathbf{R}_i(\boldsymbol{\rho})$ é definida incorretamente $\mathbf{H}_1^{-1}(\hat{\boldsymbol{\beta}}_G)$ pode ser inconsistente. Um estimador robusto para $\text{Var}(\hat{\boldsymbol{\beta}}_G)$, sugerido por Liang and Zeger

(1986), é dado por

$$\hat{\mathbf{V}}_G = \mathbf{H}_1^{-1}(\hat{\boldsymbol{\beta}}_G) \mathbf{H}_2(\hat{\boldsymbol{\beta}}_G) \mathbf{H}_1^{-1}(\hat{\boldsymbol{\beta}}_G),$$

em que $\mathbf{H}_2(\hat{\boldsymbol{\beta}}_G) = \sum_{i=1}^n \{\hat{\mathbf{D}}_i^T \hat{\boldsymbol{\Omega}}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^T \hat{\boldsymbol{\Omega}}_i^{-1} \hat{\mathbf{D}}_i\}$. O estimador $\hat{\mathbf{V}}_G$ é consistente mesmo se a matriz trabalho for definida incorretamente.

5.4.2 Estruturas de correlação

Não estruturada

Quando a matriz de correlação \mathbf{R}_i é não estruturada teremos $r_i(r_i - 1)/2$ parâmetros para serem estimados. Denotando $\mathbf{R}_i = \{R_{ijj'}\}$, o (j, j') -ésimo elemento de \mathbf{R}_i poderá ser estimado por

$$\hat{R}_{jj'} = \frac{\phi}{n} \sum_{i=1}^n \frac{(y_{ij} - \hat{\mu}_{ij})(y_{ij'} - \hat{\mu}_{ij'})}{\sqrt{\hat{V}_{ij}} \sqrt{\hat{V}_{ij'}}}.$$

Simétrica ou permutável

Neste caso assumimos $\mathbf{R}_i = \mathbf{R}_i(\rho)$, em que o (j, j') -ésimo elemento de \mathbf{R}_i fica dado por $R_{ijj'} = 1$, para $j = j'$, e $R_{ijj'} = \rho$, para $j \neq j'$. Um estimador consistente para ρ fica dado por

$$\hat{\rho} = \frac{\phi}{n} \sum_{i=1}^n \frac{1}{r_i(r_i - 1)} \sum_{j=1}^{r_i} \sum_{j'=1, j' \neq j}^{r_i} \frac{(y_{ij} - \hat{\mu}_{ij})(y_{ij'} - \hat{\mu}_{ij'})}{\sqrt{\hat{V}_{ij}} \sqrt{\hat{V}_{ij'}}}.$$

Autoregressiva AR(1)

Aqui também assumimos $\mathbf{R}_i = \mathbf{R}_i(\rho)$, em que o (j, j') -ésimo elemento de \mathbf{R}_i fica dado por $R_{ijj'} = 1$, para $j = j'$, e $R_{ijj'} = \rho^{|j-j'|}$, para $j \neq j'$. Um estimador consistente para ρ fica dado por

$$\hat{\rho} = \frac{\phi}{n} \sum_{i=1}^n \frac{1}{(r_i - 1)} \sum_{j=1}^{r_i-1} \frac{(y_{ij} - \hat{\mu}_{ij})(y_{i(j+1)} - \hat{\mu}_{i(j+1)})}{\sqrt{\hat{V}_{ij}} \sqrt{\hat{V}_{i(j+1)}}}.$$

Parâmetro de dispersão

O parâmetro de dispersão ϕ^{-1} pode ser estimado consistentemente por

$$\hat{\phi}^{-1} = \frac{1}{(N - p)} \sum_{i=1}^n \sum_{j=1}^{r_i} \frac{(y_{ij} - \hat{\mu}_{ij})^2}{\hat{V}_{ij}},$$

em que $N = \sum_{i=1}^n r_i$. Assim, o processo iterativo (5.12) deve alternar com as estimativas para ρ e ϕ até a convergência.

Testes de hipóteses para β ou para subconjuntos de β podem ser desenvolvidos através de estatísticas tipo Wald com a matriz de variância-covariância estimada $\hat{\mathbf{V}}_G$.

5.4.3 Métodos de diagnóstico

Técnicas de diagnóstico para EEGs podem ser encontradas, por exemplo, em Hardin e Hilbe (2003) e Venezuela et al. (2007) e mais recentemente em Venezuela et al. (2011). Os procedimentos apresentados a seguir foram extraídos de Venezuela et al. (2007).

Resíduos

Aplicando para as EEGs um procedimento similar àquele apresentado na Seção 1.10.2 chega-se ao seguinte resíduo de Pearson:

$$\hat{r}_{P_{ij}} = \frac{\mathbf{e}_{ij}^T \hat{\mathbf{A}}_i^{1/2} (\hat{\mathbf{V}}_i \hat{\mathbf{W}}_i)^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)}{\sqrt{1 - \hat{h}_{ijj}}},$$

para $i = 1, \dots, n$ e $j = 1, \dots, r_i$, em que $\mathbf{A}_i^{1/2} = \phi \mathbf{W}_i^{1/2} \mathbf{R}_i^{-1} \mathbf{W}_i^{1/2}$ é uma matriz de dimensão $r_i \times r_i$, \mathbf{e}_{ij}^T é um vetor de dimensão $1 \times r_i$ de zeros com 1 na j -ésima posição e h_{ijj} é o j -ésimo elemento da diagonal principal da matriz

$$\mathbf{H}_i = \mathbf{A}_i^{1/2} \mathbf{X}_i (\mathbf{X}^T \mathbf{A} \mathbf{X})^{-1} \mathbf{X}_i^T \mathbf{A}_i^{1/2},$$

em que $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_n^T)^T$ tem dimensão $N \times p$ e $\mathbf{A} = \text{diag}\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ tem dimensão $N \times N$ com $N = \sum_{i=1}^n r_i$.

Alavanca

Duas medidas de alavanca são usualmente aplicadas em EEGs. Medida de alavanca referente ao j -ésimo indivíduo do i -ésimo grupo, dada por \hat{h}_{ijj} e medida de alavanca referente ao i -ésimo grupo, definida por

$$\hat{h}_i = \frac{1}{r_i} \sum_{j=1}^{r_i} h_{ijj}.$$

Gráficos de índices para \hat{h}_{ijj} e \hat{h}_i ou contra os valores ajustados são recomendados.

Influência

Uma versão aproximada da distância de Cook para avaliar o impacto da eliminar individual das observações na estimativa $\hat{\beta}_G$ é dada por

$$\text{LD}_{ij} = \frac{\hat{h}_{ijj}}{(1 - \hat{h}_{ijj})} \hat{r}_{P_{ij}}^2.$$

Gráficos de índices para LD_{ij} são recomendados.

5.4.4 Seleção de modelos

Uma proposta de critério para seleção de modelos em EEGs (ver, por exemplo, Hardin e Hilbe, 2003) é dado por

$$\text{QIC} = -2Q(\hat{\beta}_G) + 2\text{tr}(\hat{\mathbf{V}}_G \hat{\mathbf{H}}_{1I}),$$

em que $\hat{\beta}_G$ é a estimativa de quase-verossimilhança para uma matriz específica de correlação $\mathbf{R}_i(\boldsymbol{\rho})$ e \mathbf{H}_{1I} é a matriz \mathbf{H}_1 avaliada sob a estrutura

de independência. Esse critério pode ser aplicado para selecionar submodelos encaixados ou para selecionar a matriz de correlação para um modelo específico.

5.5 Exemplos

5.5.1 Ataques epilépticos

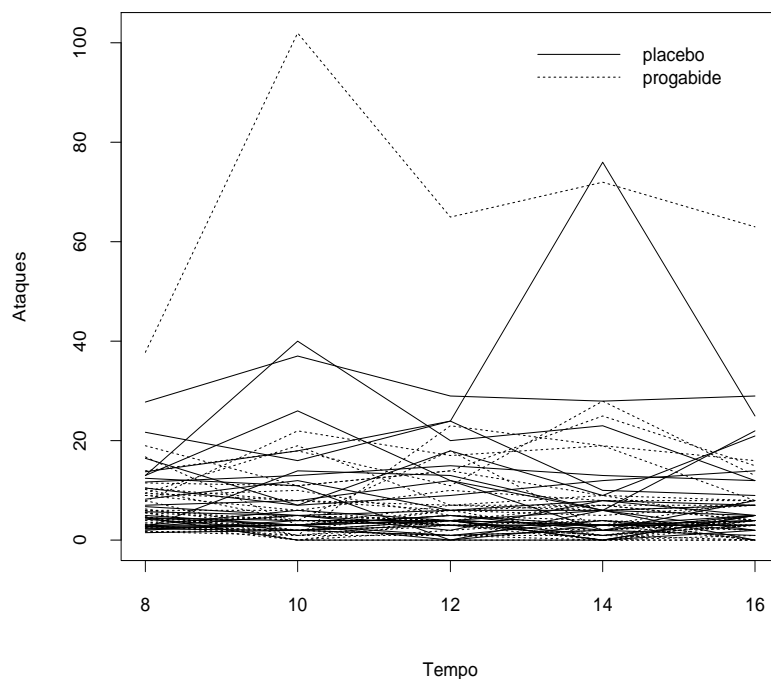


Figura 5.9: Gráfico de perfis com o número de ataques por período de 2 semanas.

No arquivo **ataques.dat** (Diggle, Liang e Zeger, 1994, Seção 8.4) são resumidos os resultados de um ensaio clínico com 59 indivíduos epilépticos os quais foram aleatorizados de modo que cada um recebesse uma droga

antiepiléptica denominada *progabide* ou placebo. Os dados de cada indivíduo consistiram de um número inicial de ataques epiléticos num período de oito semanas antes do tratamento, seguido do número de ataques em cada período de duas semanas, num total de quatro períodos, após o tratamento. O interesse da pesquisa é saber se a droga reduz a taxa de ataques epiléticos.

Para ajustar esses modelos no R usaremos a *library gee*, que deve ser acionada através do comando

```
require(gee).
```

Os ajustes podem ser feitos de forma muito similar aos MLGs desde que os dados estejam descritos de forma apropriada. Existem outras formas de gerar dados longitudinais através de outras subrotinas que facilitam, por exemplo, a elaboração de gráficos de perfis. Nesses casos, será necessário informarmos nos comandos de ajuste como as unidades experimentais estão dispostas e o tipo de correlação intraunidade experimental a ser assumida.

No caso dos ataques epiléticos uma possível distribuição marginal para os dados é a distribuição de Poisson, uma vez que tem-se dados de contagem. Contudo, observando-se a tabela abaixo, onde estão descritos os valores amostrais para a razão variância/média para os 10 grupos experimentais, nota-se um forte indício de sobredispersão sugerindo que o parâmetro de dispersão ϕ não deve ser fixado como sendo igual a um.

	Antes	Per1	Per2	Per3	Per4
Placebo	22,13	10,98	8,04	24,50	7,24
Progradibe	24,76	38,77	16,70	23,75	18,79

Para compararmos o número de ataques epiléticos nos 10 períodos experimentais, devemos padronizar os valores referentes ao período anterior ao tratamento em que os pacientes foram observados por 8 semanas. Assim,

será possível uma comparação com os demais períodos de 2 semanas. Na Figura 5.9 temos o gráfico de perfis com os dois tratamentos. Nota-se que pelo menos um paciente (#49), que foi tratado com a droga *progabide*, apresenta um número alto de ataques antes e depois do tratamento.

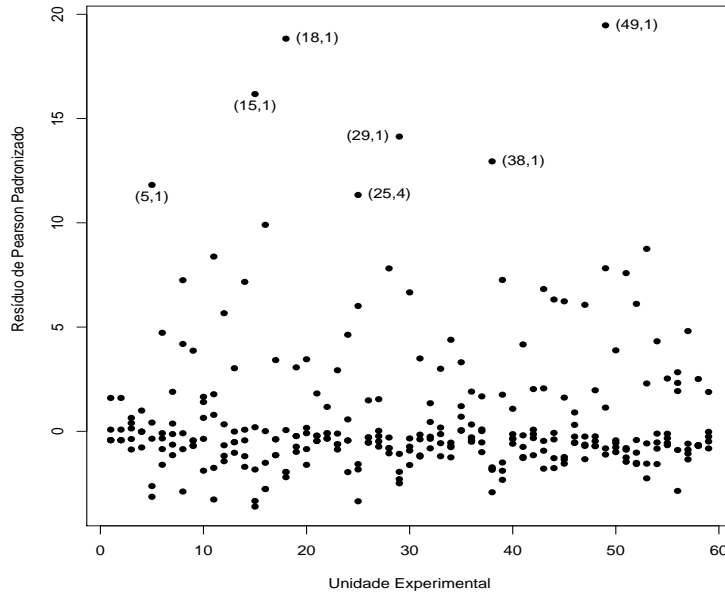


Figura 5.10: Gráfico do resíduo de Pearson referente ao modelo de Poisson com estrutura de correlação permutável ajustado aos dados sobre ataques epiléticos.

Vamos supor então que Y_{ijk} representa o número de ataques epiléticos ocorridos com o k -ésimo indivíduo do i -ésimo grupo no j -ésimo período. Assumimos que $Y_{ijk} \sim P(\lambda_{ij}t_j)$, t_j denota o número de semanas do j -ésimo período, $i = 1, 2$; $j = 0, 1, 2, 3, 4$ e $k = 1, \dots, r_{ij}$, em que $r_{1j} = 28$ (grupo placebo), $r_{2j} = 31$ (grupo tratado), $t_0 = 8$ e $t_1 = t_2 = t_3 = t_4 = 2$. Assumi-

mos também uma estrutura de correlação permutável para cada indivíduo, isto é, $\text{Corr}(Y_{ijk}, Y_{ijk'}) = \rho$, para $k \neq k'$ e (i, j) fixos. A parte sistemática do modelo será dada por

$$\begin{aligned}\log \lambda_{10} &= \alpha, \\ \log \lambda_{1j} &= \alpha + \beta, \\ \log \lambda_{20} &= \alpha + \gamma \text{ e} \\ \log \lambda_{2j} &= \alpha + \gamma + \beta + \delta,\end{aligned}$$

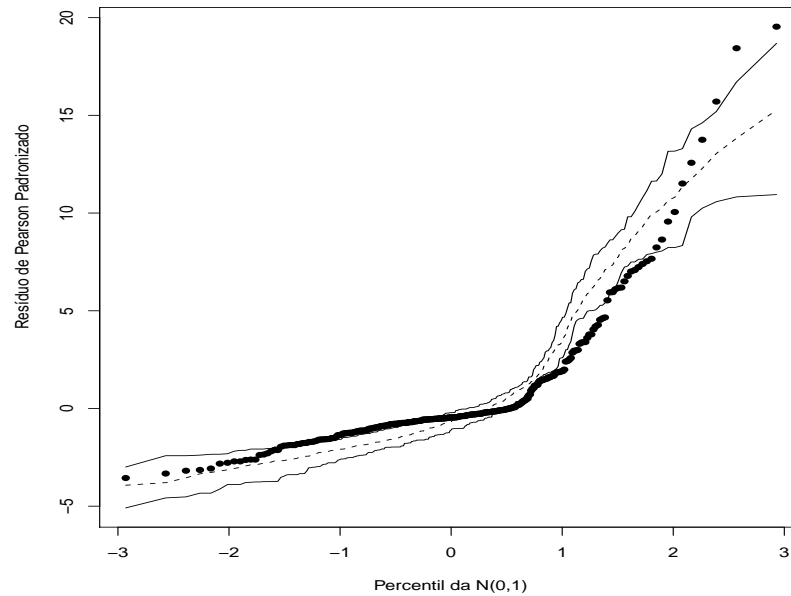


Figura 5.11: Gráfico normal de probabilidades referente ao modelo de Poisson com estrutura de correlação permutável ajustado aos dados sobre ataques epilépticos.

para $j = 1, 2, 3, 4$, em que α denota o nível base, β o efeito de tratamento, γ

o efeito de grupo e δ a interação entre tratamento e grupo. Note que, antes do tratamento, o logaritmo da razão entre as taxas dos dois grupos é dado por

$$\log\{\lambda_{20}/\lambda_{10}\} = \alpha + \gamma - \alpha = \gamma. \quad (5.13)$$

Após o tratamento, o logaritmo da razão entre as taxas fica dado por

$$\log\{\lambda_{2j}/\lambda_{1j}\} = \alpha + \gamma + \beta + \delta - \alpha - \beta = \gamma + \delta. \quad (5.14)$$

Portanto, se o tratamento não é eficaz espera-se que o logaritmo da razão não mude após o tratamento. Logo, avaliar a eficiência do tratamento equivale a testar $H_0 : \delta = 0$ contra $H_1 : \delta \neq 0$.

Tabela 5.3

Estimativas dos parâmetros do modelo log-linear de Poisson aplicado aos dados sobre ataques epilépticos.

Parâmetro	Com todos os pacientes		Sem o paciente #49	
	Estimativa	z-robusto	Estimativa	z-robusto
α	1,347	8,564	1,347	8,564
β	0,112	0,965	0,112	0,965
γ	0,027	0,124	-0,107	-0,551
δ	-0,105	-0,491	-0,302	-1,768
ρ	0,771		0,593	
ϕ^{-1}	19,68		10,53	

Se denotarmos por $\mu_{ij} = E(Y_{ijk})$, a parte sistemática do modelo em função das médias fica dada por

$$\log\mu_{ij} = \log t_j + \log\lambda_{ij},$$

em que $\log t_j$ desempenha o papel de *offset*. Para ajustarmos esse modelo no R deve-se seguir a sequência abaixo de comandos

```
fit1.ataques = gee(ataques ~ grupo + periodo + grupo*periodo +
offset(log(semanas)), id=paciente, family=poisson,
corstr="exchangeable"),
```

em que **grupo** representa o grupo (=0 placebo, =1 progabide), **periodo** representa o período (=0 antes, =1 depois), **semanas** o número de semanas, **paciente** o número do paciente (são 59 pacientes) e **corstr** o tipo de correlação a ser assumida.

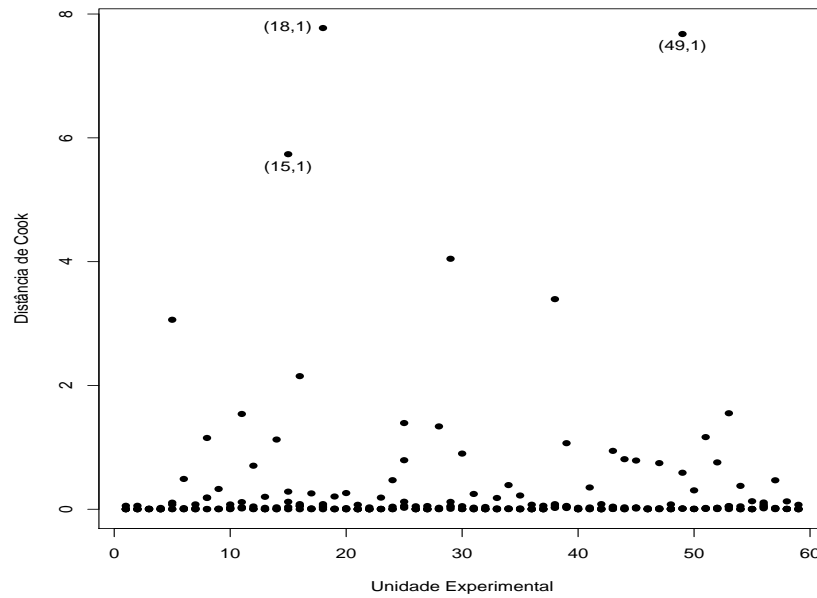


Figura 5.12: Distância de Cook referente ao modelo de Poisson com estrutura de correlação permutável ajustado aos dados sobre ataques epiléticos.

As estimativas dos parâmetros (erro padrão aproximado) são apresentadas na Tabela 5.3. Não há portanto nenhum indício de efeito de tratamento. Para a análise de resíduos vamos considerar o resíduo de Pearson $\hat{r}_{P_{ij}}$ definido na Seção 5.4.3. A geração de envelopes para esse resíduo é um pouco mais complexa do que no caso usual de respostas independentes, uma vez que requer o conhecimento da distribuição conjunta das respostas de cada

indivíduo. No entanto, mesmo quando essa distribuição não é totalmente desconhecida é possível, em alguns casos, gerar a distribuição empírica dos dados (vide, por exemplo, Venezuela et al., 2007).

Nota-se pela Tabela 5.3 que a estimativa do parâmetro de dispersão ϕ^{-1} é muito diferente da suposição de $\phi = 1$ para modelos com resposta de Poisson, sugerindo indícios fortes de sobredispersão. Assim, para uma análise de resíduos mais apropriada deve-se considerar o resíduo de Pearson padronizado $\hat{r}_{P_{ij}}^* = \sqrt{\hat{\phi}} \hat{r}_{P_{ij}}$ cujo gráfico é descrito na Figura 5.10. Nota-se 7 resíduos com valores superiores a 10,0, todos referentes a medidas de diferentes pacientes. O gráfico normal de probabilidades para o resíduo de Pearson padronizado $\hat{r}_{P_{ij}}^*$ (Figura 5.11) mostra alguns afastamentos da suposição de modelo marginal de Poisson, provavelmente devido à sobredispersão que não foi totalmente controlada.

Finalmente, na Figura 5.12, temos a distância de Cook aproximada em que três medidas se destacam. Nota-se novamente uma medida referente ao paciente (#49) cujo perfil destoa na Figura 5.9. Vamos fazer um estudo das estimativas não considerando esse paciente no ajuste. Os comandos em R são dados abaixo:

```
fit2.ataques = gee(ataques ~ grupo + periodo + grupo*perido +
offset(log(semanas)), id=paciente, subset=-c(241, 242, 243, 244,
245), family=poisson, corstr="exchangeable").
```

As novas estimativas (vide Tabela 5.3) indicam evidência de que o tratamento com a droga *progabide* reduz o número médio de ataques epiléticos, ou seja, há mudança inferencial em relação ao modelo com todos os pontos.

5.5.2 Condição Respiratória

Vamos considerar agora um exemplo discutido em Myers, Montgomery e Vining (2002, Seção 6.5) que envolve a comparação de dois tratamentos aplicados em pacientes com problemas respiratórios. Um total de 56 pacientes foi considerado no estudo sendo que 27 receberam o tratamento com uma droga ativa enquanto que os 29 pacientes restantes receberam placebo. Cada paciente foi observado em quatro ocasiões em que mediu-se a condição respiratória (boa ou ruim). Foram também observados o sexo e a idade (em anos) de cada paciente além da pré-existência de um nível base (sim ou não). Apenas como ilustração descrevemos abaixo a incidência do problema respiratório em cada ocasião segundo os dois tratamentos.

	Visita 1	Visita 2	Visita 3	Visita 4
Tratamento	22/27	13/27	5/27	1/27
Placebo	20/29	18/29	21/29	15/29

Nota-se pela tabela acima que na primeira visita há uma incidência alta para ambos os tratamentos de pacientes em condição respiratória ruim, contudo a partir da segunda visita nota-se uma queda acentuada para os pacientes tratados com a droga ativa e pouca variação para os pacientes tratados com placebo. Portanto, há fortes indícios de que a droga reduz a chance de condição respiratória ruim. Os dados completos desse experimento estão descritos no arquivo **respiratorio.dat**.

Vamos denotar por Y_{ij} a condição (=1 ruim, =0 boa) do i -ésimo paciente na j -ésima ocasião, $i = 1, \dots, 56$ e $j = 1, 2, 3, 4$. Como trata-se de resposta binária será assumido marginalmente que $Y_{ij} \sim Be(\pi_{ij})$ com parte sistemática dada por

$$\log \left\{ \frac{\pi_{ij}}{1 - \pi_{ij}} \right\} \alpha + \beta_1 \text{Idade}_i + \beta_2 \text{Trat}_i + \beta_3 \text{Sexo}_i + \beta_4 \text{Base}_i,$$

em que Idade_i denota a idade (em anos), Trat_i (=0 droga ativa, =1 placebo), Sexo_i (=0 feminino, =1 masculino) e Base_i (=0 ausência do nível base, =1 presença do nível base) do i -ésimo paciente. Seguindo a sugestão de Myers, Montgomery e Vining (2002, Seção 6.5) será assumida uma estrutura de correlação AR(1) para as respostas de cada paciente, ou seja, que $\text{Corr}(Y_{ij}, Y_{ij'}) = 1$ para $j = j'$ e $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho^{|j-j'|}$ para $j \neq j'$. Para ajustar esse modelo no R deve-se usar os comandos

```
fit1.respir = gee(condicao ~ idade + trat + sexo + base,
id=paciente, family=binomial, corstr="AR-M", M=1).
```

Tabela 5.4

Estimativas dos parâmetros do modelo logístico aplicado aos dados sobre condição respiratória.

Parâmetro	Correlação AR(1)		Independência	
	Estimativa	z-robusto	Estimativa	z-robusto
α	-0,377	-0,529	-0,404	-0,563
β_1	0,043	3,380	0,048	3,683
β_2	1,001	3,066	1,070	3,254
β_3	-2,003	-2,988	-2,178	-3,207
β_4	0,492	0,586	0,498	0,585
ρ	0,275		0,00	

As estimativas dos parâmetros dos modelos com estrutura AR(1) e independente são apresentadas na Tabela 5.4. Nota-se que as estimativas não diferem muito e os resultados inferências são os mesmos. Isso pode ser explicado pela baixa correlação entre as respostas do mesmo indivíduo, $\hat{\rho} = 0,275$.

Pelas estimativas da Tabela 5.4 pode-se concluir que o resultado da condição respiratória independe do nível base, no entanto depende da idade, do tratamento e do sexo. Por exemplo, há um aumento na chance de condição respiratória ruim com o aumento da idade, conforme esperado. A razão de chances entre sexo feminino e masculino é estimada por $\hat{\psi} = e^{2,003} = 7,41$,

ou seja, as mulheres têm aproximadamente 7,41 vezes a chance dos homens terem o problema. Pacientes que foram tratados com placebo têm $\hat{\psi} = e^{1,001} = 2,72$ vezes a chance dos pacientes que foram tratados com a droga de terem condição respiratória ruim. Em todos os cálculos acima supõe-se que as demais variáveis estão fixadas.

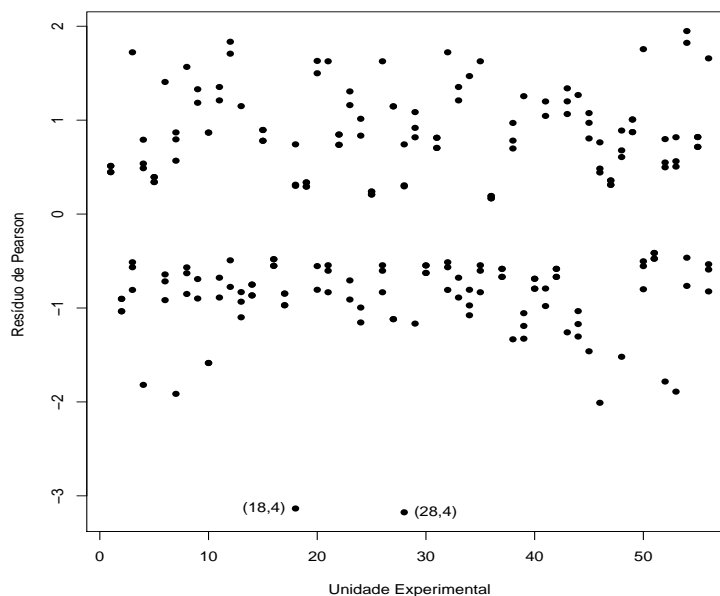


Figura 5.13: Gráfico do resíduo de Pearson referente ao modelo binomial com estrutura de correlação AR(1) ajustado aos dados sobre condição respiratória.

Na Figura 5.13 é apresentado o gráfico do resíduo de Pearson contra a ordem das observações e como podemos observar, com exceção de 2 resíduos referentes a medidas dos pacientes #18 e #28, todos os demais caem no intervalo $[-2,2]$, indicando um bom ajuste do modelo com estrutura de correlação AR(1). O gráfico normal de probabilidades com o resíduo de Pearson

(Figura 5.14) não indica afastamentos da suposição de distribuição marginal Bernoulli com estrutura de correlação AR(1).

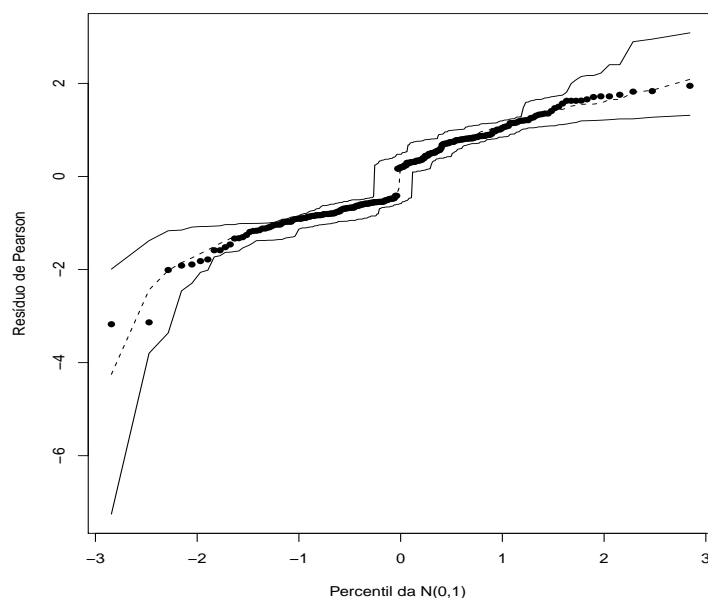


Figura 5.14: Gráfico normal de probabilidades referente ao modelo binomial com estrutura de correlação AR(1) ajustado aos dados sobre condição respiratória.

Já o gráfico da distância de Cook descrito na Figura 5.15 destaca três medidas de pacientes diferentes sendo duas dessas medidas destacadas também no gráfico com o resíduo de Pearson. Contudo, o ajuste sem considerarmos esses três pacientes não causa mudanças inferenciais.

5.5.3 Placas dentárias

Hadgu e Koch(1999) discutem os resultados de um ensaio clínico com 109 adultos voluntários com pré-existência de placa dentária. Nesse estudo os indivíduos foram distribuídos de forma aleatória para receberem um líquido tipo A (34 indivíduos), um líquido tipo B (36 indivíduos) e um líquido controle (39 indivíduos). As placas dentárias de cada indivíduo foram avaliadas e classificadas segundo um escore no início do tratamento, após 3 meses e após 6 meses. Os dados encontram-se no arquivo **rinse.dat**.

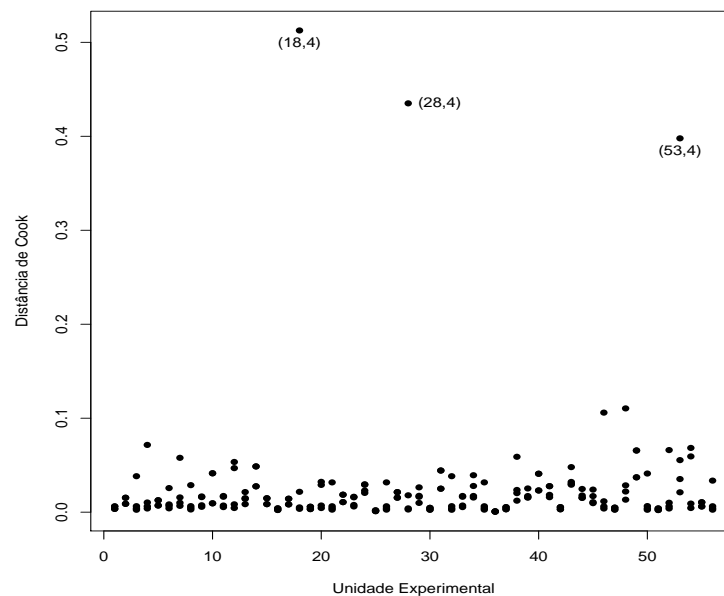


Figura 5.15: Distância de Cook referente ao modelo binomial com estrutura de correlação AR(1) ajustado aos dados sobre condição respiratória.

Tabela 5.5

Medidas resumo para os escores das placas dentárias segundo os tratamentos e períodos de escovação.

	Início	3 Meses	6 Meses
Controle	2,562	1,786	1,738
	(0,343)	(0,700)	(0,595)
	n=39	n=39	n=36
Líquido A	2,568	1,315	1,259
	(0,354)	(0,715)	(0,744)
	n=34	n=34	n=34
Líquido B	2,479	1,255	1,032
	(0,296)	(0,550)	(0,451)
	n=36	n=36	n=36

O objetivo do estudo é verificar se pelo menos um dos novos líquidos reduz o número médio de placas dentárias. Seja Y_{ijk} o escore do k -ésimo indivíduo do i -ésimo grupo (=1 controle, =2 líquido A, =3 líquido B) e j -ésimo período (=1 início do tratamento, =2 após 3 meses, =3 após 6 meses), $k = 1, \dots, n_{ij}$ com $n_{1j} = 39$, $n_{2j} = 34$ e $n_{3j} = 36$. Foram omitidas das nossas análises quatro observações para as quais não foi possível obter o valor do escore. Na Tabela 5.5 descrevemos os valores médios com os respectivos erros padrão para os grupos formados. Nota-se um decréscimo no valor médio após 3 meses de escovação para os três tratamentos, sendo a redução mais acentuada para os líquidos A e B. Nota-se também um aumento da variabilidade. De 3 meses para 6 meses de escovação o decréscimo continua para o escore médio dos grupos que receberam os líquidos A e B, havendo uma redução mais evidente para o grupo tratado com o líquido B. Esse grupo também apresenta as menores variabilidades. Essas tendências podem ser observadas quando são considerados os perfis individuais dos voluntários para os três tipos de líquido

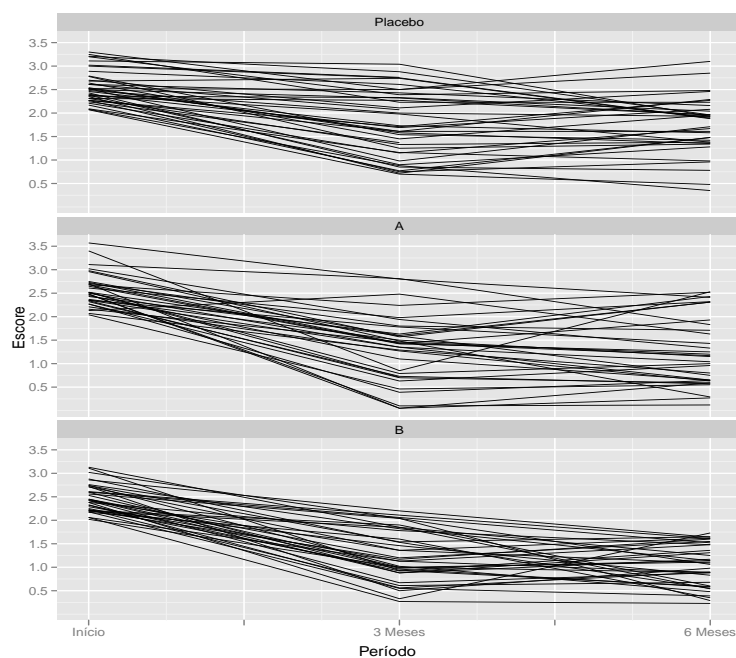


Figura 5.16: Gráfico de perfis para o escore dos voluntários que receberam placebo, líquido tipo A e líquido tipo B referente aos dados sobre placas dentárias.

ao longo do tempo conforme descrito na Figura 5.16.

Tabela 5.6

Estimativas dos parâmetros do modelo log-linear gama aplicado aos dados sobre placas dentárias.

Parâmetro	Estimativa	z-robusto	Parâmetro	Estimativa	z-robusto
α	0,941	44,407	$(\beta\gamma)_{22}$	-0,308	-3,124
β_2	0,002	0,080	$(\beta\gamma)_{32}$	-0,319	-3,835
β_3	-0,033	-1,138	$(\beta\gamma)_{23}$	-0,333	-3,266
γ_2	-0,278	-7,335	$(\beta\gamma)_{33}$	-0,492	-5,792
γ_3	-0,004	-8,321			
ρ	0,38				
ϕ^{-1}	5,68				

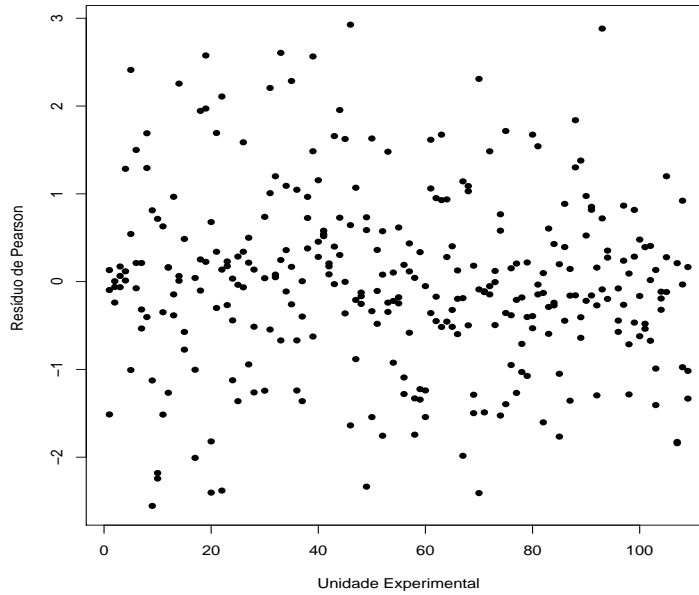


Figura 5.17: Gráfico do resíduo de Pearson referente ao modelo binomial com estrutura de correlação simétrica ajustado aos dados sobre placas dentárias.

Os pesquisadores verificaram após uma análise descritiva dos dados que a distribuição gama é mais apropriada para descrever a resposta do que a distribuição normal. Assim, vamos assumir que $Y_{ijk} \sim G(\mu_{ij}, \phi)$. Seguindo ainda os pesquisadores vamos supor um modelo log-linear com interação entre tratamento e período, porém com uma parametrização um pouco diferente,

$$\log \mu_{ij} = \alpha + \beta_i + \gamma_j + (\beta\gamma)_{ij},$$

em que $(\beta\gamma)_{ij}$ representa a interação entre tratamento e período, sendo β_i e γ_j os efeitos principais. Teremos as restrições $\beta_1 = 0$, $\gamma_1 = 0$ e $(\beta\gamma)_{1j} = (\beta\gamma)_{i1} = 0$, para $i = 1, 2, 3$ e $j = 1, 2, 3$. As estimativas dos parâmetros são descritas na Tabela 5.6 supondo correlação simétrica entre as medidas de um mesmo indivíduo. Nota-se que a estimativa da correlação não é muito alta.

Claramente confirma-se a existência de interação entre período e tratamento. Os líquidos A e B reduzem em média a quantidade de placas dentárias, havendo indícios de uma redução mais acentuada com o líquido B de 3 meses para 6 meses de escovação.

Para ajustar esse modelo no R deve-se usar os comandos

```
tratm = factor(tratm)
mes = factor(mes)
fit1.placas = gee(score ~ + tratm + mes + tratm*mes,
id=voluntar, family=Gamma(link=log), corstr="exchangeable").
```

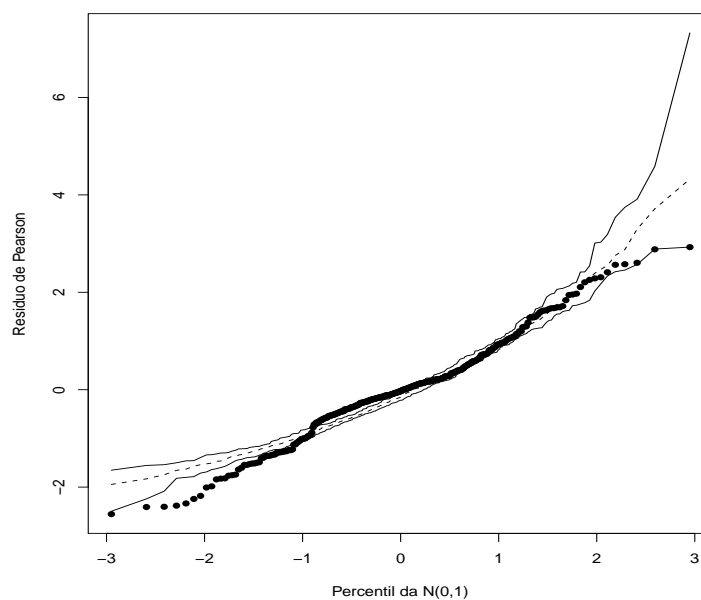


Figura 5.18: Gráfico normal de probabilidades referente ao modelo gama com estrutura de correlação simétrica ajustado aos dados sobre placas dentárias.

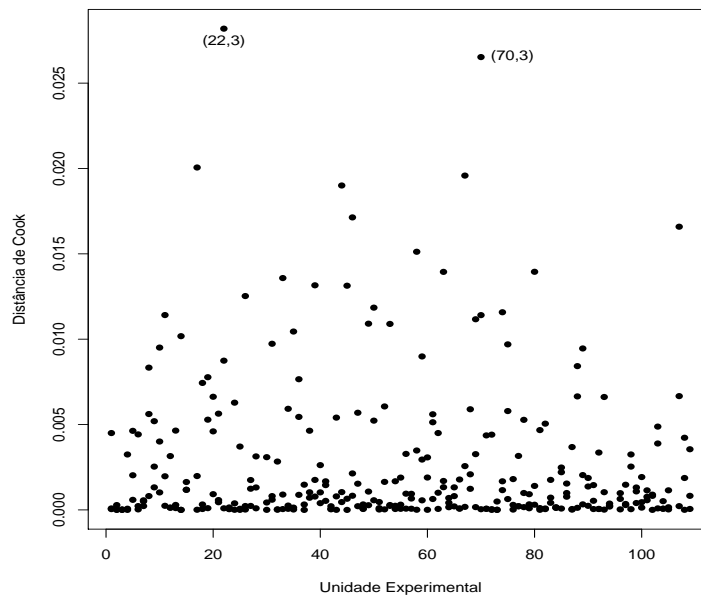


Figura 5.19: Distância de Cook referente ao modelo gama com estrutura de correlação simétrica ajustado aos dados sobre placas dentárias.

A Figura 5.17 descreve o gráfico de índices do resíduo de Pearson. Nota-se uma distribuição simétrica dos resíduos que ficam concentrados no intervalo $[-3,3]$. Pelo gráfico normal de probabilidades com o resíduo de Pearson (Figura 5.18) nota-se alguns afastamentos, em particular para os resíduos com valores negativos mais extremos, indicando uma falta de ajuste nesses casos. Finalmente, tem-se na Figura 5.19 o gráficos de índices da distância de Cook. Destaque para a 3ª medida dos voluntários #22 (líquido B) e #70 (líquido B). Espera-se para ambos os voluntários um decréscimo no escore ao longo do tempo. Todavia, para o voluntário #22 tem-se a sequência 2,56; 2,04 e 0,29, ou seja, uma queda muito acentuada da 2ª medida para a 3ª medida. Já para o voluntário #70 tem-se a sequência 2,38; 0,33 e 1,75, ou seja, um decréscimo muito acentuado da 1ª para a 2ª medida, porém um aumento

após a 2ª medida. Essas tendências que destoam do esperado para o líquido tipo B podem ter elevado o valor da distância de Cook para a 3ª medida desses voluntários. A retirada desses dois voluntários, contudo, altera muito pouco as estimativas e não altera os resultados inferenciais. Cardoso-Neto e Paula (2001) analisaram este exemplo supondo restrições em alguns dos parâmetros e encontraram evidências mais fortes com relação aos resultados obtidos por Hadgu e Koch(1999).

5.6 Exercícios

1. Supor as funções de variância $V(t) = t^3$ e $V(t) = t + t^2/k$ para $t > 0, k > 0$. Encontre para cada caso a função $Q(\mu; y)$ e verifique sob quais restrições as funções encontradas são proporcionais a funções de verossimilhança da família exponencial.
2. Considere a seguinte função de quase-verossimilhança:

$$Q(\mu; y) = \frac{1}{\sigma^2} \int_y^\mu \frac{y - t}{V(t)} dt,$$

em que $V(t) = t(1 + t)$ para $t > 0$. (i) Desenvolva essa função de quase-verossimilhança. (ii) Verifique se é possível recuperar alguma distribuição da família exponencial. Em caso afirmativo qual é a distribuição? (iii) Supor agora uma amostra aleatória de n variáveis aleatórias independentes com função de quase-verossimilhança $Q(\mu_i; y_i)$ dada acima. Como fica a função quase-desvio? (iv) Como estimar σ^2 ?

3. Considere novamente o arquivo **claims.dat** descrito no Capítulo 2 (exercício #20), em que 9 variáveis são observadas para uma amostra aleatória de 996 apólices de seguros de veículos extraída do livro de de Jong e Heller (2008). A variável **expos** (exposição do veículo), que

varia no intervalo $(0,1)$, será considerada agora como variável resposta. Inicialmente, faça uma análise descritiva dos dados e procure agrupar as variáveis categóricas em um número menor de categorias. Aplique modelos de quase-verossimilhança com funções $V(\mu) = \mu(1 - \mu)$ e $V(\mu) = \mu^2(1 - \mu)^2$, em que μ denota o valor esperado para a exposição do veículo, para explicar a variável resposta dadas as demais variáveis explicativas. Para o modelo selecionado faça uma análise de diagnóstico e procure interpretar os coeficientes estimados através de razões de chances.

4. Supor Y_1, \dots, Y_n variáveis aleatórias independentes com logaritmo da função de quase-verossimilhança $Q(\mu_i; y_i)$, $i = 1, \dots, n$. Mostre que as funções escore e de informação para β ficam, respectivamente, dadas por:

$$\mathbf{U}_\beta = \frac{1}{\sigma^2} \mathbf{D}^T \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu})$$

e

$$\mathbf{K}_{\beta\beta} = -E \left\{ \frac{\partial \mathbf{U}(\beta)}{\partial \beta} \right\} = \frac{1}{\sigma^2} \mathbf{D}^T \mathbf{V}^{-1} \mathbf{D}.$$

5. Sejam Y_{ij} variáveis aleatórias tais que $Y_{ij} \sim \text{FE}(\mu_i, \phi)$, $i = 1, 2$ e $j = 1, \dots, m$. A estatística de Wald para testar $H_0 : \mu_1 - \mu_2 = 0$ contra $H_1 : \mu_1 - \mu_2 \neq 0$ é dada por $\xi_W = (\bar{Y}_1 - \bar{Y}_2)^2 / \text{Var}(\bar{Y}_1 - \bar{Y}_2)$. Sob H_0 e para $m \rightarrow \infty$ segue que $\xi_W \sim \chi_1^2$. Calcular $\text{Var}(\bar{Y}_1 - \bar{Y}_2)$ para as seguintes situações:

- (a) supondo que $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho$ para $(j \neq j'; i \text{ fixo})$ e $=0$ em caso contrário;
- (b) supondo que $\text{Corr}(Y_{ij}, Y_{i'j}) = \rho$ para $(i \neq i'; j \text{ fixo})$ e $=0$ em caso contrário;

Para $\mu_1 - \mu_2$ e ϕ fixos e $\rho \geq 0$ discutir o comportamento do poder de ξ_W conforme ρ cresce para as situações (a) e (b). São esperados esses comportamentos? Comente.

6. Supor $Y_{ij} \sim \text{FE}(\mu, \phi)$, para $i = 1, \dots, n$ e $j = 1, \dots, r_i$, $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho$ para $j \neq j'$ (=1 caso contrário), $r_i \geq 2$. Obter $E(Y_i)$ e $\text{Var}(Y_i)$, em que $Y_i = Y_{i1} + \dots + Y_{ir_i}$. Mostre que $-1/(r_{\min} - 1) \leq \rho \leq 1$, comente. Use os resultados $\text{Var}(X + Z) = \text{Var}(X) + \text{Var}(Z) + 2\text{Cov}(X, Z)$ e $\text{Cov}(X, Z) = \rho\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Z)}$.
7. Supor $Y_i \stackrel{\text{ind}}{\sim} Q(\pi_i; y_i)$, em que $E(Y_i) = \pi_i$ e $\text{Var}(Y_i) = \sigma^2\pi_i(1 - \pi_i)$, para $i = 1, \dots, n$, com parte sistemática dada por $\text{arcsen}(\sqrt{\pi_i}) = \beta_0 + \beta_1(x_i - \bar{x})$. Obtenha a matriz de variância-covariância assintótica $\text{Var}(\hat{\beta})$, em que $\beta = (\beta_0, \beta_1)^\top$. Desenvolva uma estatística tipo-escore para testar $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$? Qual a distribuição nula assintótica da estatística do teste? Resultados úteis: $\sum_{i=1}^n (x_i - \bar{x}) = 0$ e $\frac{d}{dx} \text{arcsen}\{u(x)\} = \frac{1}{\sqrt{1-u^2}} \frac{du}{dx}$.
8. Supor o modelo de quase-verossimilhança em que Y_1, \dots, Y_n são variáveis aleatórias independentes tais que $E(Y_i) = \mu_i$ e $\text{Var}(Y_i) = \sigma^2\mu_i^2$ com parte sistemática dada por $\log\mu_i = \beta_0 + \beta_1(x_i - \bar{x})$. Responda aos itens abaixo:
 - (a) como ficam as variâncias assintóticas de $\hat{\beta}_0$ e de $\hat{\beta}_1$?
 - (b) Como fica o teste de Wald para testar $H_0 : \beta_0 = 0$ contra $H_1 : \beta_0 \neq 0$?
 - (c) Proponha um teste tipo escore para testar $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$.

9. Como fica a diferença entre desvios para testar $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$ num modelo de quase-verossimilhança com $V(\mu_i) = \mu_i^2(1 - \mu_i)^2$, $g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ e $\boldsymbol{\beta} = (\beta_1^T, \beta_2^T)^T$?

10. (Park, Shin e Park, 1998). Vamos supor que o vetor de respostas seja agora dado por $\mathbf{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijT})^T$, em que Y_{ijt} denota a resposta para o j -ésimo elemento do i -ésimo grupo no instante t , $i = 1, \dots, g$ e $j = 1, \dots, r_i$. Supor ainda que $E(Y_{ijt}) = \mu_i$, $\text{Var}(Y_{ijt}) = V_i \phi^{-1}$ e que Y_{ijt} pertence à família exponencial. Mostre que dado $\hat{\boldsymbol{\rho}}$ a equação de estimação generalizada para μ_i pode ser expressa na forma $\mathbf{S}(\hat{\mu}_i) = 0$, em que

$$\mathbf{S}(\mu_i) = \sum_{j=1}^{r_i} \mathbf{1}_T^T \mathbf{R}_{ij}(\boldsymbol{\rho})(\mathbf{y}_{ij} - \mu_i \mathbf{1}_T),$$

\mathbf{R}_{ij} é a matriz trabalho para o j -ésimo indivíduo do i -ésimo grupo e $\mathbf{1}_T$ é um vetor $T \times 1$ de uns. Expresse a estimativa de μ_i em forma fechada.

11. Supor que $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ir_i})^T$, $i = 1, \dots, n$, são vetores aleatórios independentes tais que $Y_{ij} \sim \text{Be}(\pi_i)$. Assumir ainda que a matriz trabalho para \mathbf{Y}_i é permutável e que

$$\log \left\{ \frac{\pi_i}{1 - \pi_i} \right\} = \mathbf{x}_i^T \boldsymbol{\beta}.$$

Mostre que, dado $\hat{\rho}$, as EEGs para $\boldsymbol{\beta}$ ficam dadas por

$$\mathbf{S}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_G) = \sum_{i=1}^n \{1 + (r_i - 1)\hat{\rho}\}^{-1} \mathbf{x}_i (y_i - n_i \hat{\pi}_i) = \mathbf{0},$$

em que $y_i = y_{i1} + \dots + y_{ir_i}$. Sugestão: use a relação abaixo

$$\mathbf{R}_i^{-1}(\rho) = (1 - \rho)^{-1} [\mathbf{I}_{r_i} - \rho \{1 + (r_i - 1)\rho\}^{-1} \mathbf{J}],$$

em que \mathbf{J} é uma matriz $r_i \times r_i$ de uns. Como fica o processo iterativo para estimar $\boldsymbol{\beta}$?

12. Supor que $Y_{ij} \sim \text{Be}(\mu)$ para $i = 1, \dots, n$ e $j = 1, \dots, r_i$, em que $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho$ (fixado) para $j \neq j'$ com parte sistemática dada por $\log\left\{\frac{\mu}{1-\mu}\right\} = \beta$. Responda às seguintes questões: (i) como fica a equação de estimação generalizada para estimar β ? (ii) expresse em forma fechada a estimativa $\hat{\beta}_G$ (obtenha inicialmente $\hat{\mu}_G$) e (iii) como fica a variância assintótica (não robusta) de $\hat{\beta}_G$?
13. Supor que $Y_{ij} \sim \text{Q}(\mu, \sigma^2)$ para $i = 1, \dots, n$ e $j = 1, 2$, em que $\text{Var}(Y_{ij}) = \sigma^2\mu^2$, $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho$ para $j \neq j'$ com parte sistemática dada por $\log\mu = \beta$. Responda às seguintes questões: (i) como fica a equação de estimação generalizada para estimar β ? (ii) expresse em forma fechada a estimativa $\hat{\beta}_G$ (obtenha inicialmente $\hat{\mu}_G$) e (iii) como fica a variância assintótica (não robusta) de $\hat{\beta}_G$? Supor que ρ e σ^2 são estimados consistentemente.
14. Considere uma amostra aleatória de n indivíduos que são observados em 2 ocasiões cada um, sendo Y_{ij} a resposta do i -ésimo indivíduo na j -ésima ocasião para $i = 1, \dots, n$ e $j = 1, 2$, com a suposição $Y_{i1} \stackrel{\text{ind}}{\sim} \text{FE}(\mu_1, \phi)$ e $Y_{i2} \stackrel{\text{ind}}{\sim} \text{FE}(\mu_2, \phi)$ e $\rho = \text{Corr}(Y_{i1}, Y_{i2})$ ou seja $\text{Cov}(Y_{i1}, Y_{i2}) = \rho\sqrt{\text{Var}(Y_{i1})}\sqrt{\text{Var}(Y_{i2})}$. A diferença entre as médias amostrais nas duas ocasiões $\bar{Y}_2 - \bar{Y}_1$, em que $\bar{Y}_j = n^{-1} \sum_{i=1}^n Y_{ij}$ para $j = 1, 2$, é utilizada para detectar eventuais diferenças entre as médias μ_2 e μ_1 . Responda às seguintes questões:
- (i) calcule $\text{Var}(\bar{Y}_2 - \bar{Y}_1)$,
 - (ii) chame $\Delta = \mu_2 - \mu_1$ e calcule $P(\Delta - \epsilon < \bar{Y}_2 - \bar{Y}_1 < \Delta + \epsilon) = 1 - \alpha$, $0 < \alpha < 1$ e $\epsilon > 0$, em que $1 - \alpha = P(-z < Z < z)$, $Z \sim \text{N}(0, 1)$,
 - (iii) expresse n em função das quantidades z , ϵ , Δ e ρ e

- (iv) discuta o comportamento de n em função de ρ mantendo-se as demais quantidades fixas.

Supor para n grande $\bar{Y}_2 - \bar{Y}_1 \sim N(\Delta, \text{Var}(\bar{Y}_1 - \bar{Y}_2))$.

15. (Myers, Montgomery e Vining, 2002, p. 231). Um experimento é conduzido para avaliar a dispersão de um pigmento particular numa pintura. Quatro diferentes misturas do pigmento são estudadas. O procedimento consiste em preparar cada mistura e aplicá-la num painel usando três métodos diferentes: pincel, rolo e spray. O experimento é repetido três dias diferentes e a resposta é a porcentagem de reflectância do pigmento. Os dados são descritos na tabela abaixo e no arquivo **mistura.dat**.

Dia	Método	Mistura			
		1	2	3	4
1	1	64,5	66,3	74,1	66,5
	2	68,3	69,5	73,8	70,0
	3	70,3	73,1	78,0	72,3
2	1	65,2	65,0	73,8	64,8
	2	69,2	70,3	74,5	68,3
	3	71,2	72,8	79,1	71,5
3	1	66,2	66,5	72,3	67,7
	2	69,0	69,0	75,4	68,6
	3	70,8	74,2	80,1	72,4

Analise os dados através de equações de estimação generalizadas com estrutura de correlação simétrica. Faça análise de diagnóstico.

16. (Myers, Montgomery e Vining, 2002, Seção 6.5). No arquivo **ratos-gee.dat** estão os dados de um experimento em que 30 ratos tiveram

uma condição de leucemia induzida. Três drogas quimio-terápicas foram utilizadas no tratamento dos animais. Foram coletadas de cada animal a quantidade de células brancas (WBC), a quantidade de células vermelhas (RBV) e o número de colônias de células cancerosas (RESP) em quatro períodos diferentes. Assuma distribuição de Poisson para RESP em cada período e verifique através de um modelo log-linear se existe diferenças significativas entre os três tratamentos considerando WBC e RBC como variáveis explicativas. Compare os resultados supondo estruturas de correlação independente e AR(1). Faça uma análise de diagnóstico.

17. Sejam $Y_{i1} \stackrel{\text{iid}}{\sim} \text{FE}(\mu_1, \phi)$ e $Y_{i2} \stackrel{\text{iid}}{\sim} \text{FE}(\mu_2, \phi)$, em que $\text{Corr}(Y_{i1}, Y_{i2}) = \rho$, para $i = 1, \dots, n$. Para testar $H_0 : \mu_1 - \mu_2 = 0$ contra $H_1 : \mu_1 - \mu_2 \neq 0$ considere a estatística

$$\xi_W = \frac{(\bar{Y}_1 - \bar{Y}_2)^2}{\text{Var}(\bar{Y}_1 - \bar{Y}_2)},$$

que sob H_0 segue assintoticamente distribuição χ_1^2 central. Sob a hipótese alternativa ξ_W segue assintoticamente distribuição $\chi_1^2(\lambda)$, em que $\lambda = (\mu_1 - \mu_2)^2 / 2\text{Var}(\bar{Y}_1 - \bar{Y}_2)$ é o parâmetro de não centralidade. Seja $P_n(\lambda, \rho) = P(\xi_W > c | H_1)$ o poder do teste (probabilidade de rejeitar H_0 quando há diferença entre as médias populacionais). Calcule inicialmente $\text{Var}(\bar{Y}_1 - \bar{Y}_2)$ e discuta o comportamento do poder do teste à medida que varia o coeficiente de correlação linear de Pearson $-1 \leq \rho \leq 1$, supondo fixados μ_1 , μ_2 , ϕ e n . Procure responder para quais situações será necessário um tamanho amostral maior para detectar a mesma diferença (entre as médias populacionais) com a mesma probabilidade.

18. (Myers, Montgomery e Vining, 2002, Seção 6.5). No arquivo **artrite.dat** estão os dados de um ensaio clínico em que 20 pacientes com ar-

trite foram aleatorizados de modo que 10 receberam o medicamento **auronofin** e os outros 10 receberam **placebo**. Foram observadas as variáveis explicativas **gênero** (1: masculino, 0: feminino) e a **idade** do paciente em anos além do **tratamento** (0: placebo, 1: auronofin). Os pacientes foram consultados em 4 ocasiões (1: início, 2: 1 mês, 3: 2 meses e 4: 3 meses) a respeito do seu estado avaliado pelo próprio paciente (1: ruim, 2: regular, 3: bom). Faça inicialmente uma análise descritiva com os dados.

Seja Y_{ij} o estado do i -ésimo paciente na j -ésima ocasião (=1 bom, =0 regular ou ruim) para $i = 1, \dots, 20$ e $j = 1, 2, 3, 4$. Assuma que $Y_{ij} \sim \text{Be}(\pi_{ij})$, em que π_{ij} é a probabilidade do estado ser considerado bom pelo i -ésimo paciente na j -ésima ocasião. Proponha uma EEG para explicar π_{ij} através de uma regressão logística e considerando as estruturas de correlação simétrica e AR(1) entre as ocasiões de um mesmo paciente. Considere no modelo apenas os efeitos principais **tratamento**, **idade**, **gênero** e **ocasião**. Compare os modelos através de métodos de diagnóstico e para o modelo escolhido faça uma interpretação através de razões de chances.

19. No arquivo **Milk** do gamlss são apresentados dados referentes a um experimento longitudinal desenvolvido na Austrália com 79 vacas que foram aleatorizadas segundo três dietas e foi observado semanalmente a quantidade de proteína no leite de cada animal. O objetivo principal do estudo é verificar se há diferenças significativas entre as quantidades médias semanais de proteína sob as três dietas. Os dados estão descritos na seguinte ordem: (i) **protein** (quantidade de proteína), (ii) **Time** (semana), (iii) **Cow** (identificação do animal) e (iv) **Diet** (cevada, cevada+tremoços e tremoços). É preciso informar que a variável **Diet**

é categórica através do comando

```
Diet=factor(Diet).
```

Fazer inicialmente uma análise descritiva dos dados, por exemplo, apresentando os perfis dos animais segundo a quantidade de proteína observada ao longo das semanas e para cada dieta gráficos de densidade e boxplots. Ajustar inicialmente uma equação de estimação generalizada gama com estrutura de correlação do tipo AR(1) e considere o tempo como variável explicativa contínua. Verifique se é possível incluir interação entre **Diet** e **Time**. Faça uma análise de diagnóstico e interprete os resultados do modelo selecionado.

20. No arquivo **gross.dat** estão resumidos os dados de produtividade dos 48 estados norte-americanos contíguos no período de 1970 a 1986. As variáveis estão descritas na seguinte ordem (os recursos estão expressos em milhões de USD): (i) **state**, nome do estado, (ii) **region**, região do estado, (iii) **yr**, ano, (iv) **pcap**, total do capital de empresas públicas, (v) **hwy**, capital das estradas e rodovias, (vi) **water**, capital das empresas de saneamento básico, (vii) **util**, capital das demais empresas públicas, (viii) **pc**, total do capital privado, (ix) **gsp**, produto interno bruto, (x) **emp**, total de empregos e (xi) **unemp**, taxa de desemprego. O objetivo do estudo é tentar relacionar o produto interno bruto de cada estado com as demais variáveis. Faça uma análise descritiva considerando apenas as variáveis, **gsp**, **water** e **yr**.

Supor inicialmente o seguinte modelo de quase-verossimilhança:

$$(i) Y_{ij} \sim Q(\mu_{ij}; y_{ij}), \quad E(Y_{ij}) = \mu_{ij} \text{ e } \text{Var}(Y_{ij}) = \sigma^2 \mu_{ij}^2$$

$$(ii) \log(\mu_{ij}) = \beta_0 + \beta_1 \log(\text{water})_{ij} + \beta_2 \text{yr}_{ij}$$

$$(iii) \text{Corr}(\mathbf{Y}_{ij}) = \mathbf{R}_{ij}(\alpha),$$

em que $\sigma^2 > 0$, $\mathbf{Y}_{ij} = (Y_{ij1}, \dots, Y_{ij17})^\top$. Compare o modelo acima supondo as estruturas de correlação simétrica e AR(1). Para a estrutura selecionada tente melhorar o modelo, por exemplo, incluindo interação. Faça uma análise de diagnóstico do modelo final e interprete os resultados.

Apêndice

Neste apêndice descrevemos os conjuntos de dados usados nos exemplos e nos exercícios propostos. As variáveis são descritas na ordem em que aparecem em cada arquivo.

Capítulo 1

canc3.dat: tipo de tumor (0:benigno, 1:maligno), idade (em anos), sexo (1:masculino, 2:feminino), HL e FF (1:ausente, 2:discreta, 3:moderada, 4:intensa).

canc4.dat: grupo de passagem (0 a 28), presença de massa tumoral (1:sim, 0:não), caquexia (1:sim, 0:não) e tempo de sobrevivência (em dias).

capm.dat: taxa de retorno Tbill, retorno Microsoft, retorno SP500, retorno GE e retorno Ford.

censo.dat: unidade da federação, escolaridade média (anos de estudo) e renda média (em reais).

imoveis.dat: imposto do domicílio (em 100 USD), área do terreno (em 1000 pés quadrados), área construída (em 1000 pés quadrados), idade da residência (em anos) e preço de venda do imóvel (em 1000 USD).

reg1.dat: área (em mil pés quadrados) e preço (em mil USD).

reg2.dat: sigla do estado, taxa do combustível (em USD), porcentagem de motoristas licenciados, renda per capita (em USD), ajuda federal às estradas do estado (em mil USD) e consumo per capita de combustível (em galões por ano).

reg3.dat: nome do estado, população estimada em julho de 75, renda per capita em 74 (em USD), proporção de analfabetos em 70, expectativa de vida 69-70, taxa de criminalidade em 76 (por 100000 habitantes), proporção de estudantes que concluíram o segundo grau em 70, número de dias do ano com temperatura abaixo de zero graus Celsius e área do estado (em milhas quadradas).

reg4.dat: x1, x2, x3, x4, e octanas. A resposta é o número de octanas.

salary.dat: salário anual (em mil USD), sexo, posição na empresa (escore de 1 a 9) e experiência (em anos).

trees.dat: diâmetro (em polegadas), altura (em pés) e volume da árvore (em pés cúbicos).

vendas.dat: total de telhados vendidos (em mil metros quadrados), gastos pela loja com publicidade (em mil USD), número de clientes cadastrados na loja (em milhares), número de marcas concorrentes do produto e potencial da loja.

Capítulo 2

claims.dat: valor do veículo (em 10000 dólares australianos), exposição do veículo, número de sinistros no período, custo total dos sinistros (em dólares australianos), tipo do veículo (em 11 categorias), idade do veículo

(em 4 categorias), sexo do condutor principal, área de residência do condutor principal (em 6 categorias) e idade do condutor principal (em 6 categorias).

dfilme.dat: tempo de duração do filme (em horas) e densidade máxima do filme.

energy.dat: total de energia consumida num mês (em kilowatts-hora) e demanda de energia na hora de pico.

insurance.dat: valor pago do seguro (dolares australianos), representação legal (0:não, 1:sim), mês em que ocorreu o acidente e tempo operacional.

milho.dat: quantidade de nitrogênio, quantidade de fosfato e produtividade de milho (libras/acre).

pesca.dat: frota (Santos e Ubatuba), ano (95 a 99), trimestre (1 a 4), latitude (de 23,25° a 28,25°), longitude (de 41,25° a 50,75°), dias de pesca, captura (quantidade em kg de peixes capturados) e cpue (captura por unidade de esforço).

restaurante.dat: faturamento anual (em mil USD) e gastos com publicidade (em mil USD).

snack.dat: força necessária para o cisalhamento, tipo de snack (1:A, 2:B, 3:C, 4:D, 5:E), número de semanas.

sobrev.dat: número de células brancas, tempo de sobrevivência (em semanas) e característica morfológica (AG=1 positivo, AG=0 negativo).

turbina.dat: tipo de turbina (1 a 5) e tempo de duração do motor (em milhões de ciclos).

vidros.dat: tempo de resistência (em horas), voltagem (1:200, 2:250, 3:300, 4:350) e temperatura (1:170 graus Celsius, 2:180 graus Celsius).

Capítulo 3

besouros.dat: besouros mortos, besouros expostos e dose.

caduquice.dat: escore no exame psicológico, ocorrência de caduquice (1:sim, 0:não).

camundongos: sexo (1:macho, 0:fêmea), tratamento (1:sim, 0:controle), casos e expostos.

dengue.dat: idade (em anos) do entrevistado, nível sócio-econômico (1:alto, 2:médio, 3:baixo), setor da cidade onde mora o entrevistado (1:setor 1, 2:setor 2) e diagnóstico da doença (1:sim, 0:não).

diabetes.dat: massa corporal, histórico familiar (1:presença, 0:ausência) e atividades físicas (1:presença, 0:ausência) para os casos e para os controles, respectivamente.

dose1.dat: dose, caramujos expostos e caramujos mortos.

dose2.dat: dose, caramujos expostos e caramujos mortos.

dose3.dat: dose, caramujos expostos e caramujos mortos.

equipamentos.dat: tempo, número de equipamentos expostos, número de equipamentos que falharam.

gestantes.dat: idade (0:≤ 30, 1:30 ou +), número de cigarros consumidos por dia (0:≤ 5, 1:5 ou +), tempo de gestação (0:≤260 dias, 1:> 260 dias), crianças não sobreviventes e crianças sobreviventes.

grahani.dat: número de lagartos da espécie grahani, total de lagartos, período do dia (1:manhã, 2:meio-dia, 3:tarde), comprimento da madeira (1:curta, 2:cumprida), largura da madeira (1:estreita, 2:larga) e local de ocupação (1:claro, 2:escuro).

insetic.dat: número de insetos mortos, número de insetos expostos, dose do inseticida, inseticida DDT, inseticida γ -DDT e inseticida DDT + γ -DDT (1:presença, 0:ausência).

leuce.dat: idade do paciente (em anos), mancha diferencial da doença, infiltração na medula, células com leucemia, malignidade da doença, temperatura máxima antes do tratamento, tratamento (1:satisfatório, 0:não), tempo de sobrevivência (em meses) e situação (1:sobrevivente, 0:não sobrevivente).

matched.dat: estrato, observação (1:caso, 2:controle), idade da paciente no momento da entrevista (em anos), diagnóstico (1:caso, 0:controle), tempo de escolaridade (em anos), grau de escolaridade (0:nenhum, 1:segundo grau, 2:técnico, 3:universitário, 4:mestrado, 5:doutorado), checkup regular (1:sim, 2:não), idade da primeira gravidez, idade do início da menstruação, número de abortos, número de filhos, peso (em libras), idade do último período menstrual e estado civil (1:casada, 2:divorciada, 3:separada, 4:viúva, 5:solteira). Observações perdidas são denotadas por NA.

meninas.dat: garotas menstruando, garotas entrevistadas e idade média.

morgan.dat: concentração (R, D, M), dose, insetos expostos, insetos mortos.

olhos.dat: cor dos olhos dos pais, cor dos olhos dos avós, número total de filhos e número de filhos com olhos claros.

prefauto.dat: preferência comprador tipo de automóvel (1:americano, 0:japonês), idade do comprador (em anos), sexo do comprador (0:masculino, 1:feminino) e estado civil do comprador (0:casado, 1:solteiro).

pregibon.dat: resposta (1:ocorrência, 0:ausência), volume e razão.

pulso.dat: pulsação em repouso (1:normal, 0:alta), hábito de fumar (1:sim, 2:não) e peso (em kg).

rotifers.dat: densidade, rotifers suspensos, rotifers expostos e espécie (1:Polyarthra, 0:Keratella).

sementes.dat: temperatura da germinação, nível da umidade, nível da temperatura, número de sementes que germinaram.

Capítulo 4

breslow.dat: número de casos de câncer, total de pessoas-anos, número de cigarros por dia (1:não fumante, 2:1-9 cigarros, 3:10-30 cigarros, 4:+30 cigarros) e faixa-etária (1:40-49 anos, 2:50-59 anos, 3:60-69 anos, 4:70-80 anos).

canc1.dat: idade no primeiro emprego com 4 níveis (1:≤20, 2:20-27, 3:27.5-34.9, 4:35+ anos), ano do primeiro emprego com 4 níveis (1:≤1910, 2:1910-1914, 3:1915-1919, 4:1920-1924), tempo decorrido desde o primeiro emprego com 5 níveis (1:0-19, 2:20-29, 3:30-39, 4:40-49, 5:50+ anos), número de casos de câncer e o total de pessoas-anos de observação.

detergente.dat: temperatura da água, uso de M, preferência (X,M), maciez da água, número de pessoas.

emprego.dat: nível de renda (1: < USD 6000, 2: USD 6000-15000, 3: USD 15000-25000, 4: > USD 25000), grau de satisfação (1:alto, 2: bom, 3: médio, 4: baixo) e número de indivíduos.

geriatra.dat: número de quedas no período, intervenção (0:educação somente, 1:educação e exercícios físicos), sexo (0:feminino, 1:masculino), balanço e força.

heart.dat: doença das coronárias (1:sim, 2:não), nível de colesterol (1:menor do que 200 mg/100 cc, 2:200-219, 3:220-259, 4:260 ou +), pressão arterial (1:menor do que 127 mm Hg, 2:127-146, 3:147-166, 4:167 ou +) e número de indivíduos.

navios.dat: tipo do navio (1:A, 2:B, 3:C, 4:D, 5:E), ano da fabricação (1:60-64, 2:65-69, 3:70-74, 4:75-79), período de operação (1:60-74, 2:75-79), tempo de operação (em meses) e número de avarias.

nitrofen: dosagem de nitrofen, total de ovos eclodidos.

quine.dat: etnia (A:aborígine, N:não aborígine), sexo (M:masculino, F: feminino), ano (F0:8^a série, F1:1^a ano ensino médio, F2:2^a ano ensino médio, F3:3^a ano ensino médio), desempenho (SL:baixo, AL:normal) e dias ausentes no ano letivo.

recrutas.dat: hábito de nadar (ocasional, frequente), local onde costuma nadar (piscina, praia), faixa-etária (15-19, 20-25, 25-29), sexo (masculino, feminino) e número de infecções de ouvido.

rolos.dat: comprimento do tecido (em metros) e número de falhas.

store.dat: número de clientes, número de domicílios, renda média anual (em USD), idade média dos domicílios (em anos), distância entre a área e o competidor mais próximo (em milhas) e distância entre a área e a loja (em milhas).

tv cabo.dat: número de domicílios na área (em milhares), porcentagem de domicílios com TV a cabo, renda per capita (em USD) por domicílio com TV a cabo, taxa de instalação de TV a cabo (em USD), custo médio mensal de manutenção de TV a cabo (em USD), número de canais a cabo disponíveis na área e número de canais não pagos com sinal de boa qualidade disponíveis na área.

Capítulo 5

artrite.dat: paciente, ocasião (1:início, 2:1 mês, 3:2 meses, 4:3 meses), gênero (1:masculino, 0:feminino), idade (em anos), tratamento (0:placebo, 1:aurofin), resultado (1:ruim, 2:regular, 3:bom).

ataques.dat: indivíduo, período (1:antes do tratamento, 2:1º período após o tratamento, 3:2º período após o tratamento, 4:3º período após o tratamento), número de semanas em cada período, número de ataques em cada período e tratamento (0:placebo, 1:progabide).

cevada.dat: incidência da mancha (proporção), local (1 a 9) e variedade (1 a 10).

mosca.dat: número de ácaros coletados espécie2, espécie3, espécie6, espécie14, número de partes da placa, posição (1:lateral, 0:central), região (1:São Roque, 2:Pindamonhangaba, 3:Nova Odessa, 4:Ribeirão Preto) e temperatura (em graus Celsius).

mistura.dat: painel, dia, método, mistura, porcentagem de reflectância do pigmento.

ratosgee.dat: animal, período, quantidade de células brancas, quantidade de células vermelhas e número de colônias de células cancerosas.

respiratorio.dat: paciente, tratamento (0:droga ativa, 1:placebo), sexo (0:feminino, 1:masculino), idade (em anos), nível base (0:ausência, 1:presença) e condição do paciente nas visitas (0:boa, 1:ruim).

rinse.dat: voluntário, período (1:início, 2:após 3 meses, 3:após 6 meses), tratamento (1:placebo, 2:rinse A, 3:rinse B) e escore.

Bibliografia

- Agresti A. (1990). *Categorical Data Analysis*. John Wiley, New York.
- Aitkin, M.; Anderson, D. A.; Francis, B e Hinde, J. P. (1989). *Statistical Modelling in Glim*. Clarendom Press, Oxford.
- Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control* **AU-19** 716-722.
- Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data. *Biometrika* **68**, 357-364.
- Armitage, P. (1955). Test for linear trend in proportions and frequencies. *Biometrics* **11**, 375-386.
- Armitage, P. (1971). *Statistical Methods in Medical Research*. Blackwell Scientific Publications, Oxford.
- Atkinson, A. C. (1981). Two graphical display for outlying and influential observations in regression. *Biometrika* **68**, 13-20.
- Atkinson, A. C. (1985). *Plots, Transformations and Regressions*. Oxford Statistical Science Series, Oxford.
- Beckman R. J.; Nachtsheim, C. J. e Cook, R. D. (1987). Diagnostics for mixed-model analysis of variance. *Technometrics* **29**, 413-426.

- Belsley, D. A.; Kuh, E. e Welsch, R. E. (1980). *Regression Diagnostics*. John Wiley, New York.
- Bliss, C. I. (1935). The calculation of the dosage-mortality curve. *Annals of Applied Biology* **22**, 134-167.
- Bishop, Y. M. M.; Fienberg, S. E. e Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge.
- Boice, J. D. e Monson, R. R. (1977). Breast cancer in women after repeated fluoroscopic examinations of the chest. *Journal of the National Cancer Institute* **59**, 823-832.
- Box, G. E. P. e Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society B* **26**, 211-252.
- Breslow, N. E. e Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9-25.
- Breslow N. E. e Day, N. E. (1980). *Statistical Methods in Cancer Research, Vol. I, The Analysis of Case-Control Studies*. IARC Scientific Publications, International Agency for Research on Cancer, Lyon.
- Breslow, N. E. e Day, N. E. (1987). *Statistical Methods in Cancer Research, Vol. II, The Design and Analysis of Cohort Studies*. IARC Scientific Publications, International Agency for Research on Cancer, Lyon.
- Buse, A. (1982). The likelihood ratio, Wald and Lagrange multiplier tests: an expository note. *The American Statistician* **36**, 153-157.

- Cardoso-Neto, J. e Paula, G. A. (2001). Wald one-sided test using generalized estimating equations approach. *Computational Statistics and Data Analysis* **36**, 475-495.
- Casella, G. e Straederman, W. E. (1980). Confidence bands for linear regression with restricted predictor variables. *Journal of the American Statistical Association* **75**, 862-868.
- Chatterjee, S. e Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*. New York: Wiley.
- Collett, D. (1991). *Modelling Binary Data*. Chapman and Hall, London.
- Cook, R. D. (1977). Detection of influential observations in linear regressions. *Technometrics* **19**, 15-18.
- Cook, R. D. (1987). Influence assessment. *Journal of Applied Statistics* **14**, 117-131.
- Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B* **48**, 133-169.
- Cook, R. D.; Peña, D. e Weisberg, S. (1988). The likelihood displacement: A unifying principle for influence measures. *Communications in Statistics, Theory and Methods* **17**, 623-640
- Cook, R. D. e Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.
- Cordeiro, G. M. (1986). *Modelos Lineares Generalizados*. Livro texto de minicurso, VII Simpósio Nacional de Probabilidade e Estatística, UNICAMP, Campinas, SP.

- Cordeiro, G. M. e McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B* **53**, 629-643.
- Cordeiro, G. M.; de P. Ferrari, S. L. e Paula, G. A. (1993). Improved score tests for generalized linear models. *Journal of the Royal Statistical Society B* **55**, 661-674.
- Cordeiro, G. M.; Paula, G. A. e Botter, D. A. (1994). Improved likelihood ratio tests for dispersion models. *International Statistical Review* **62**, 257-274.
- Cordeiro, G. M. e Paula, G. A. (1989a). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika* **76**, 93-100.
- Cordeiro, G. M. e Paula, G. A. (1989b). *Modelos de Regressão para a Análise de Dados Univariados*. Livro texto de minicurso, 17^o Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro.
- Cordeiro, G. M. e Paula, G. A. (1992). Estimation, large-sample parametric tests and diagnostics for non-exponential family nonlinear models. *Communications in Statistics - Simulation and Computation* **21**, 149-172.
- Cornfield, J. (1951). A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast and cervix. *Journal of the National Cancer Institute* **11**, 1269-1275.
- Cornfield, J. (1956). A statistical problem arising from retrospective studies. *In: Proceedings of the Third Berkeley Symposium*, Berkeley, University of California Press, pgs. 133-148.
- Cox, D. R. (1970). *The Analysis of Binary Data*. Methuen, London.

- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B* **74**, 187-220.
- Cox, D. R. e Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.
- Cox, D. R. e Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London.
- Cox, D. R. e Snell, E. J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society B* **30**, 248-275.
- Cox, D. R. e Snell, E. J. (1989). *The Analysis of Binary Data, 2nd Edition*. Chapman and Hall, London.
- Davison, A. C. e Gigli, A. (1989). Deviance residuals and normal scores plots. *Biometrika* **76**, 211-221.
- Davison, A.C. e Tsai, C-L. (1992). Regression model diagnostics. *International Statistical Review* **60**, 337-353.
- Day, N. E. e Byar, D. P. (1979). Testing hypothesis in case-control studies-equivalence of Mantel-Haenszel statistics and logit score tests. *Biometrics* **35**, 623-630.
- de Souza, F. A. M. e Paula, G. A. (2002). Deviance residuals for an angular response. *Australian and New Zealand Journal of Statistics* **44**, 345-356.
- Diggle, P. J.; Liang, K. Y. e Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford University Press.

- Dixon, W. J. (1987). *BMDP Statistical Software*. University of California Press, Berkeley.
- Efron, B. (1988). Logistic regression, survival analysis and the Kaplan-Meier curve. *Journal of the American Statistical Association* **83**, 414-425..
- Emerson, J. D.; Hoaglin, D. C. and Kempthorne, P. J. (1984). Leverage in least squares additive-plus-multiplicative fits for two-way tables. *Journal of the American Statistical Association* **79**, 329-335.
- Escobar, L. A. e Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics* **48**, 507-528.
- Everitt, B. S. (1977). *The Analysis of Contingency Tables*. Chapman and Hall, London.
- Everitt, B. S. (1994). *A Handbook of Statistical Analysis using S-Plus*. Chapman and Hall, London.
- Fahrmeir, L. e Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* **13**, 342-368.
- Fahrmeir, L. e Klinger, J. (1994). Estimating and testing generalized linear models under inequality constraints. *Statistical Papers* **35**, 211-229.
- Fahrmeir, L. e Tutz, G. (1994). *Multivariate Statistical Modelling based on Generalized Linear Models*. Springer, New York.
- Feigl, P. e Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics* **21**, 826-838.

- Finney, D. J. (1971). *Probit Analysis, 3rd. Edition*. Cambridge University Press, Cambridge.
- Finney, D. J. (1978). *Statistical Methods in Biological Assay, 3rd. Edition*. Cambridge University Press, Cambridge.
- Fieller, E. C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society B* **16**, 175-185.
- Fung, W. K. (1993). Unmasking outliers and leverage points: A Confirmation. *Journal of the American Statistical Association* **88**, 515-519.
- Fung, W. K. e Kwan, C. W. (1997). A note on local influence based on normal curvature. *Journal of the Royal Statistical Society B* **59**, 839-843.
- Galea, M.; Paula, G. A. e Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician* **46**, 71-79.
- Galea, M.; Paula, G. A. e Uribe-Opazo, M. (2003). On influence diagnostic in univariate elliptical linear regression models. *Statistical Papers* **44**, 23-45.
- Galea, M.; Riquelme, M. e Paula, G. A. (2000). Diagnostic methods in elliptical linear regression models. *Brazilian Journal of Probability and Statistics* **14**, 167-184.
- Galves, J. A.; Paula, G. A. e Goebbels, M. (1998). Relatório de Análise Estatística sobre o Projeto: *Evolução Temporal da Variação Próclise/Ênclise no Português Clássico*. RAECEA-9810, IME-USP.
- Gray, J. B. (1989). On the use of regression diagnostics. *The Statistician* **38**, 97-105.

- Gu, H. e Fung, W. K. (1998). Assessing local influence in canonical correlation analysis. *Annals of the Institute of Statistical Mathematics* **50**, 755-772.
- Hadgu, A. e Koch, G. (1999). Application of generalized estimating equations to a dental randomized clinical trial. *Journal of Biopharmaceutical Statistics* **9**, 161-178.
- Hand, D. J.; Daly, F.; Lunn, A. D.; McConway, K. J. e Ostrowski, E. (1994). *A Handbook of Small Data Sets*. Chapman and Hall, London.
- Hannan, J. e Harkness, W. (1963). Normal approximation to the distribution of two independent binomials, conditional to the sum. *Annals of Mathematical Statistics* **34**, 1593-1595.
- Hastie, T. e Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Hinde, J. (1982). Compound poisson regression models. In R. Gilchrist Ed., GLIM82, pgs. 109-121. Springer, New York.
- Hinde, J. e Demétrio, C. G. B (1998). Overdispersion: model and estimation. *Computational Statistics and Data Analysis* **27**, 151-170.
- Hoaglin, D. C. e Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *The American Statistician* **32**, 17-22.
- Hosmer, D. W. e Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley, New York.
- Innes, J. R. M.; Ulland, B. M.; Valerio, M. G.; Petrucelli, L.; Fishbein, L.; Hart, E. R.; Pallota, A. J.; Bates, R. R.; Falk, H. L.; Gart, J. J.; Klein,

- M.; Mitchell, I. e Peters, J. (1969). Biossay of pesticides and industrial chemicals for tumorigenicity in mice: A preliminary note. *Journal of the National Cancer Institute* **42**, 1101-1114.
- Jørgensen, B. (1983). Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models. *Biometrika* **70**, 19-28.
- Jørgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society B* **49**, 127-162.
- Jørgensen, B. (1996). *The Theory of Dispersion Models*. Chapman and Hall, London.
- Kim, M. G. (1995). Local influence in multivariate regression. *Communications in Statistics, Theory Methods* **20**, 1271-1278.
- Kwan, C. W. e Fung, W. K. (1998). Assessing local influence for specific restricted likelihood: Applications to factor analysis. *Psychometrika* **63**, 35-46.
- Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley, New York.
- Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* **15**, 209-225.
- Lawrence, A. J. (1988). Regression transformation diagnostics using local influence. *Journal of the American Statistical Association* **84**, 125-141.
- Lee, E. T. (1991). *Statistical Methods for Survival Data Analysis*, Second Edition. John Wiley, New York.

- Lee, Y. e Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society B* **58**, 619-678.
- Lee, Y. e Nelder, J. A. (2001). Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* **88**, 987-1006.
- Leemis, L. M. e Trivedi, K. S. (1996). A comparison of approximate interval estimators for the Bernoulli parameter. *The American Statistician* **50**, 63-68.
- Liang, K. Y. e Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.
- Liu, S. Z. (2000). On local influence for elliptical linear models. *Statistical Papers* **41**, 211-224.
- Mantel, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association* **58**, 690-700.
- Mantel, N. e Haenszel, B. F. (1959). Statistical aspects of the analysis of the data from retrospective studies of disease. *Journal of the National Cancer Institute* **22**, 719-748.
- McCullagh, P. (1983). Quasi-likelihood functions. *Annals of Statistics* **11**, 59-67.
- McCullagh, P. (1987). *Tensor Methods in Statistics*. Chapman and Hall, London.
- McCullagh, P. e Nelder, J. A. (1989). *Generalized Linear Models, 2nd. Edition*. Chapman and Hall, London.

- McCulloch, C. E. e Searle, S. R. (2001). *Linear and Generalized Linear Mixed Models*. Wiley, New York.
- Milicer, H. e Szczotka, F. (1966). Age at menarche in Warsaw girls in 1965. *Human Biology* **38**, 199-203.
- Montgomery, D. C.; Peck, E. A. e Vining, G. G. (2001). *Introduction to Linear Regression Analysis, Third Edition*. John Wiley, New York.
- Myers, R.H.; Montgomery, D. C. e Vining, G. G. (2002). *Generalized Linear Models: With Applications in Engineering and the Sciences*. John Wiley, New York.
- Moolgavkar, S. H.; Lustbader, E. D. e Venzon, D. J. (1984). A geometric approach to non-linear regression diagnostics with application to matched case-control studies. *Annals of Statistics* **12**, 816-826.
- Morgan, B. J. T. (1992). *Analysis of Quantal Response Data*. Chapman and Hall, London.
- Narula, S. C. e Stangenhau, G. (1988). *Análise de Regressão L_1* . Notas de minicurso do VIII Simpósio Nacional de Probabilidade e Estatística, IMPA, Rio de Janeiro, RJ.
- Nelder, J. A. e Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika* **74**, 221-232.
- Nelder, J. A. e Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A* **135**, 370-384.
- Neter, J.; Kutner, M. H.; Nachtsheim, C. J. e Wasserman, W. (1996). *Applied Linear Regression Models*, 3rd Edition. Irwin, Illinois,

- Nyquist, H. (1991). Restricted estimation of restricted generalized linear models. *Applied Statistics* **40**, 133-141.
- O'Hara Hines, R. J.; Lawless, J. F. e Carter, E. M. (1992). Diagnostics for a cumulative multinomial generalized linear model with application to grouped toxicological mortality data. *Journal of the American Statistical Association* **87**, 1059-1069.
- Ortega, E. M. M.; Bolfarine, H. e Paula, G. A. (2003). Influence diagnostic in generalized log-gamma regression models. *Computational Statistics and Data Analysis* **42**, 165-186.
- Palmgren, J. (1981). The Fisher information matrix for log linear models against conditionally on observed explanatory variables. *Biometrika* **68**, 563-566.
- Pan, J. X.; Fang, K. T. e von Rosen (1997). Local influence assessment in the growth curve model with unstructured covariance. *Journal of Statistical Planning and Inference* **62**, 263-278.
- Park, T. P.; Shin, D. W. e Park, C. G. (1998). A generalized estimating equations approach for testing ordered group effects with repeated measurements. *Biometrics* **54**, 1645-1653.
- Paula, G. A. (1993). Assessing local influence in restricted regression models. *Computational Statistics and Data Analysis* **16**, 63-79.
- Paula, G. A. (1995). Influence and residuals in restricted generalized linear models. *Journal of Statistical Computation and Simulation* **51**, 315-352.

- Paula, G. A. (1996). Influence diagnostic in proper dispersion models. *Australian Journal of Statistics* **38**, 307-316.
- Paula, G. A. (1997). *Estimação e Testes em Modelos de Regressão com Parâmetros Restritos*. Livro texto de minicurso da 5ª Escola de Modelos de Regressão, realizada de 26 a 28-02-97 em Campos do Jordão, SP.
- Paula, G. A. (1999). Leverage in inequality constrained regression models. *The Statistician* **48**, 529-538.
- Paula, G. A. e Artes, R. (2000). One-sided test to assess correlation in logistic linear models using estimating equations. *Biometrical Journal* **42**, 701-714.
- Paula, G. A.; Barbosa, L. S. e Ferreira, R. F. G. (1989). Relatório de Análise Estatística sobre o Projeto: *Comportamento Biológico Evolutivo do Tumor KB no Decorrer de suas Passagens Seriadas em Ratos Nude Adultos*. RAE-CEA8904, IME-USP.
- Paula, G. A. e Cordeiro, G. M. (1986). Alguns modelos não-lineares via o Glim. *Atas do VII Simpósio Nacional de Probabilidade e Estatística*, UNICAMP, São Paulo, pp. 204-217.
- Paula, G. A.; Denaro-Machado, L.; Ogata, T. T.; Machado, J. C.; Matta, M. S. e Petrella, S. M. C. N. (1992). Caquexia cancerosa em modelo experimento rato nude atímico/tumor humano KB. *Revista Laes Haes* **76**, 28-30.
- Paula, G. A. e Oshiro, C. H. (2001). Relatório de Análise Estatística sobre o Projeto: *Análise de Captura por Unidade de Esforço do Peixe-Batata na Frota Paulista*. RAE-CEA0102, IME-USP.

- Paula, G. A. e Peres, C. A. (1988). Diagnostics for GLMs with linear inequality parameter constraints. *Communications in Statistics, Theory and Methods* **17**, 4205-4219.
- Paula, G. A. e Tuder, R. M. (1986). Utilização da regressão logística para aperfeiçoar o diagnóstico de processo infeccioso pulmonar. *Revista Ciência e Cultura* **40**, 1046-1050.
- Paula, G. A.; Sevanes, M. e Ogando, M. A. (1988). Relatório de Análise Estatística sobre o Projeto: *Estudo de Plantas Brasileiras com Efeito Moluscicida em Biomphalaria Glabrata*. RAE-CEA8824, IME-USP.
- Paula, G. A. e Sen, P. K. (1995). One-sided tests in generalized linear models with parallel regression lines. *Biometrics* **51**, 1494-1501.
- Paula, G. A. e Tavares, H. R. (1992). Relatório de Análise Estatística sobre o Projeto: *Ácaros Associados ao Esterco Bovino. Subsídios para Controle Biológico da Mosca do Chifre*. RAECEA 9206, IME-USP
- Peduzzi, P. N.; Hardy, R. J. e Holford, T. T. (1980). A stepwise variable selection procedure for nonlinear regression models. *Biometrics* **36**, 511-516.
- Peña, D. e Yohai, V. (1999). A fast procedure for outlier diagnostics in large regression problems. *Journal of the American Statistical Association* **94**, 434-445.
- Pettitt, A. N. e Bin Daud, I. (1989). Case-weight measures of influence for proportional hazards regression. *Applied Statistics* **38**, 51-67.
- Piegorsch, W. W. e Casella, G. (1988). Confidence bands for logistic regression with restricted predictor variables. *Biometrics* **44**, 739-750.

- Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics* **9**, 705-724.
- Pregibon, D. (1982). Score tests in GLIM with applications. *Lecture Notes in Statistics* **14**, 87-97. Springer-Verlag, New York.
- Pregibon, D. (1984). Data analytic methods for matched case-control studies. *Biometrics* **40**, 639-651.
- Ramanathan, R. (1993). *Statistical Methods in Econometrics*. Wiley, New York.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Second Edition. Wiley, New York.
- Ratkowsky, D. A. (1983). *Nonlinear Regression Modelling*. Marcel Dekker, New York.
- Ross, W. H. (1987). The geometry of case deletion and the assessment of influence in nonlinear regression. *Canadian Journal of Statistics* **15**, 91-103.
- Ryan, B. F. e Joiner, B. L. (1994). *Minitab Handbook, Third Edition*. Duxbury Press, Belmont.
- Seber, G. A. F. e Wild, C. J. (1989). *Nonlinear Regression*. John Wiley, New York.
- Sen, P. K. e Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. Chapman and Hall, London.
- Silva, G. L. (1992). *Modelos Logísticos para Dados Binários*. Dissertação de Mestrado, IME-USP.

- Spector, P. (1994). *An Introduction to S and S-Plus*. Duxbury Press, Belmont.
- St. Laurent, R. T. e Cook, R. D. (1992). Leverage and superleverage in nonlinear regression. *Journal of the American Statistical Association*, **87**, 985-990.
- Stukel, T. A. (1988). Generalized logistic models. *Journal of the American Statistical Association* **83**, 426-431.
- Svetliza, C. F. (2002). *Modelos Não-Lineares com Resposta Binomial Negativa*. Tese de Doutorado, IME-USP.
- Svetliza, C. F. e Paula, G. A. (2001). On diagnostics in log-linear negative binomial models. *Journal of Statistical Computation and Simulation* **71**, 231-244.
- Svetliza, C. F. e Paula, G. A. (2003). Diagnostics in nonlinear negative binomial models. *Communications in Statistics, Theory Methods* **32**, 1227-1250.
- Thomas, W. e Cook, R. D. (1990). Assessing influence on predictions from generalized linear models. *Technometrics* **32**, 59-65.
- Tsai, C. H. e Wu, X. (1992). Assessing local influence in linear regression models with first-order autoregressive or heteroscedastic error structure. *Statistics and Probability Letters* **14**, 247-252.
- Venables, W. N. e Ripley, B. D. (1999). *Modern Applied Statistics with S-Plus, Third Edition*. Springer, New York.
- Wang, P. C. (1985). Adding a variable in generalized linear models. *Technometrics* **27**, 273-276.

- Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61**, 439-447.
- Wedderburn, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika* **68**, 27-32.
- Wei, B. C. (1998). *Exponential Family Nonlinear Models*. Lecture Notes in Statistics Vol. 130. Springer, New York.
- Wei, B.C.; Hu, Y.Q. e Fung, W.K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics* **25**, 25-37.
- Williams, D. A. (1984). Residuals in generalized linear models. *In: Proceedings of the 12th. International Biometrics Conference*, Tokyo, pp. 59-68.
- Williams, D. A. (1987). Generalized linear model diagnostic using the deviance and single case deletion. *Applied Statistics* **36**, 181-191.
- Wolf, (1955). On estimating the relationship between blood group and disease. *Annals of Human Genetic* **19**, 251-253.
- Wood, F. S. (1973). The use of individual effects and residuals in fitting equations to data. *Technometrics* **15**, 677-687.