

Modelos de regressão para dados discretos (parte 1): dados binários

Prof. Caio Azevedo

Motivação

- As metodologias, incluindo os modelos de regressão, vistas até agora, são apropriadas para análise de dados categorizados.
- Veremos como analisar situações nas quais a variável resposta é discreta mas os dados não estão categorizados (e as vezes não podem ser categorizados).
- No entanto, como veremos algumas das metodologias podem ser aplicadas para a análise de dados categorizados, em algumas situações.

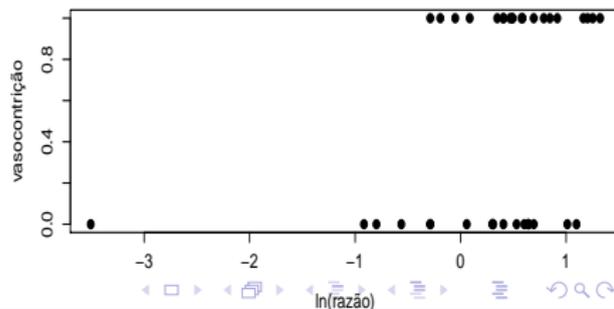
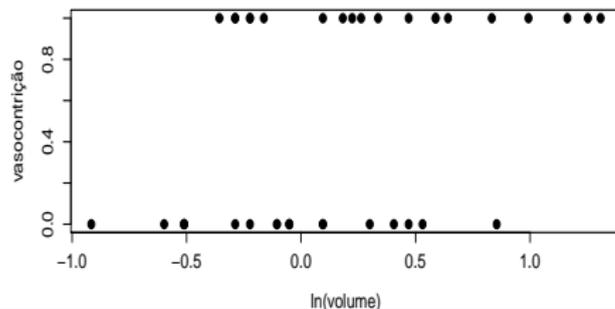
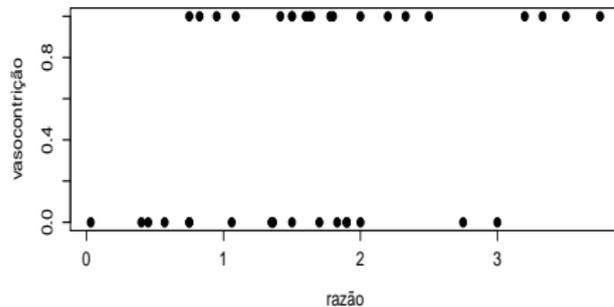
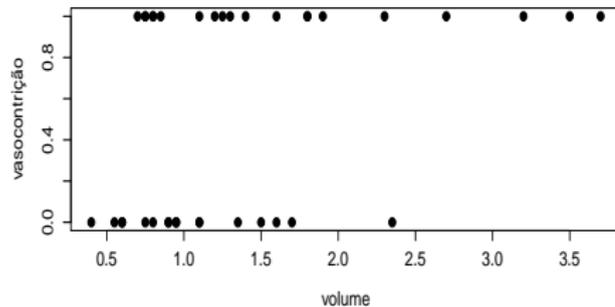
Exemplo 11: Estudo sobre vasoconstricção

- Consideraremos os dados sobre um estudo de vasoconstricção (veja Paula (2013) e referências nela constantes).
- Nesse estudo, foram medidos de 3 pacientes o volume e a razão de ar inspirado, como também a ocorrência ou não de vasoconstricção (contração de vasos sanguíneos) na pele dos dedos da mão. O primeiro paciente contribuiu com 9 observações, o segundo com 8 e o terceiro com 22.

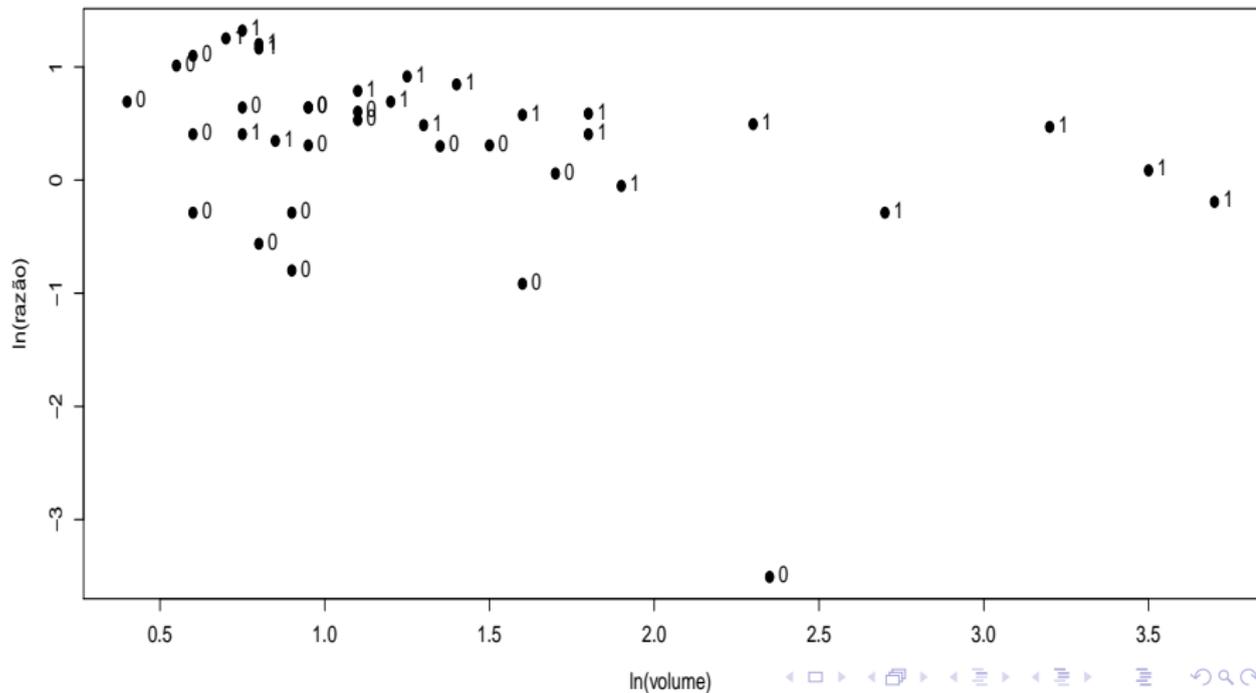
Exemplo 11 (cont.)

- Objetivo: verificar se a quantidade de ar (volume e razão, variáveis explicativas) influenciam a ocorrência de vasoconstrição (resposta).
- Seja Y_i a variável aleatória que assume valor 1, se ocorreu vasoconstrição no i -ésimo paciente e 0, caso contrário.
- As vezes é mais apropriado trabalhar com o \ln (logaritmo natural) das variáveis explicativas (para, por exemplo, medir melhor o impacto de cada uma na variável resposta, principalmente se esta não for contínua).

Gráficos de dispersão individuais



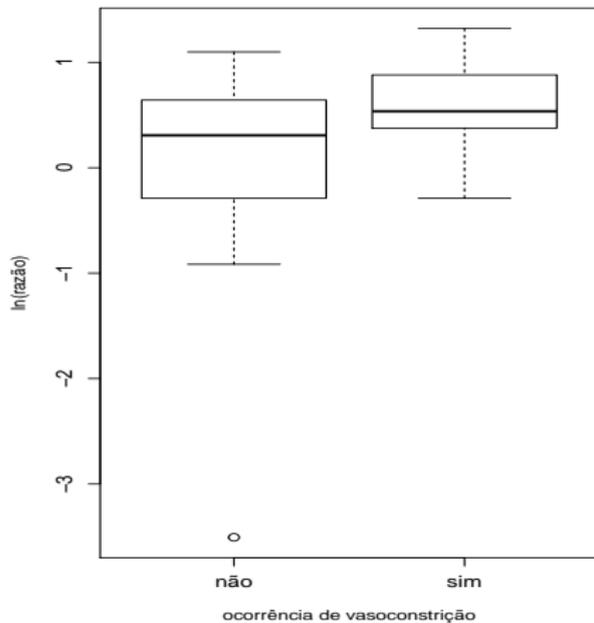
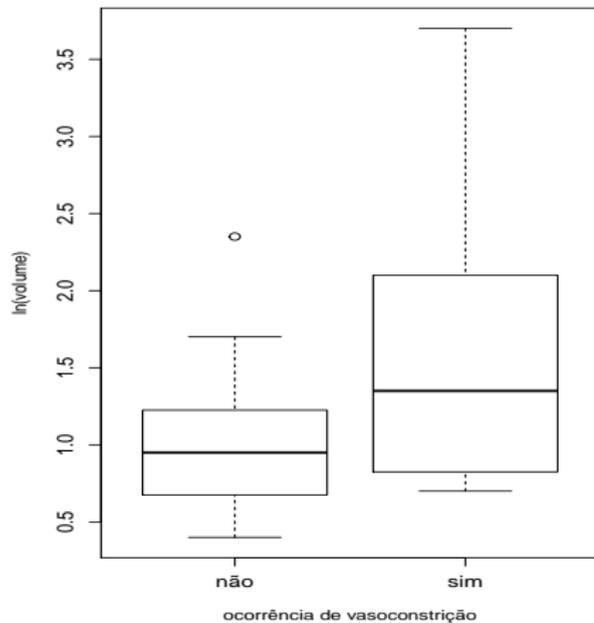
Gráficos de dispersão: $\ln(\text{razão}) \times \ln(\text{volume})$



Medidas resumo $\ln(\text{razão})$ e $\ln(\text{volume})$

Medida resumo	lnvolume		lnrazão	
	Resposta			
	0	1	0	1
Média	-0,06	0,37	0,05	0,58
Mediana	-0,05	0,30	0,31	0,54
DP	0,45	0,54	1,03	0,46
Var.	0,20	0,29	1,07	0,22
$ \text{CV}(\%) $	723,00	147,00	2223,00	81,00
Min.	-0,92	-0,36	-3,51	-0,29
Max.	0,85	1,31	1,10	1,30

Box plots das variáveis explic. em função resposta



Modelo de regressão (geral) para dados binários

$$Y_i \stackrel{ind.}{\sim} \text{Bernoulli}(p_i)$$

$$F^{-1}(p_i) = \sum_{j=1}^p \beta_j x_{ji} \rightarrow p_i = F\left(\sum_{j=1}^p \beta_j x_{ji}\right), i = 1, 2, \dots, n$$

- Y_i : ocorrência (1) ou não (0) de algum evento.
- x_{ji} : valor da variável explicativa j associada ao indivíduo i ; β_j : parâmetro associado ao impacto de cada covariável na probabilidade de ocorrência do supracitado evento.
- $F(\cdot)$: função de distribuição acumulada de alguma variável aleatória (contínua) com suporte em \mathcal{R} . $F^{-1}(\cdot)$ é conhecida como função de ligação. $\eta_i = \sum_{j=1}^p \beta_j x_{ji}$ (preditor linear)
- Modelo com intercepto: $x_{1i} = 1, \forall i$.

Funções de ligação

- Escolhas para $F(\cdot)$:

- Distribuição normal padrão: $p_i = F(\eta_i) = \Phi(\eta_i) = \int_{-\infty}^{\eta_i} e^{-x^2/2} dx$
(regressão probito). Assim, $F^{-1}(p_i) = \Phi^{-1}(p_i)$ (obtenível de forma numérica, função `qnorm` (R), por exemplo).

- Distribuição logística padrão:

$$p_i = F(\eta_i) = \int_{-\infty}^{\eta_i} \frac{e^{-x}}{(1 + e^{-x})^2} dx = \frac{1}{1 + e^{-\eta_i}} \text{ (regressão logística).}$$

$$\text{Assim, } F^{-1}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right)$$

Funções de ligação

- Distribuição Cauchy padrão:

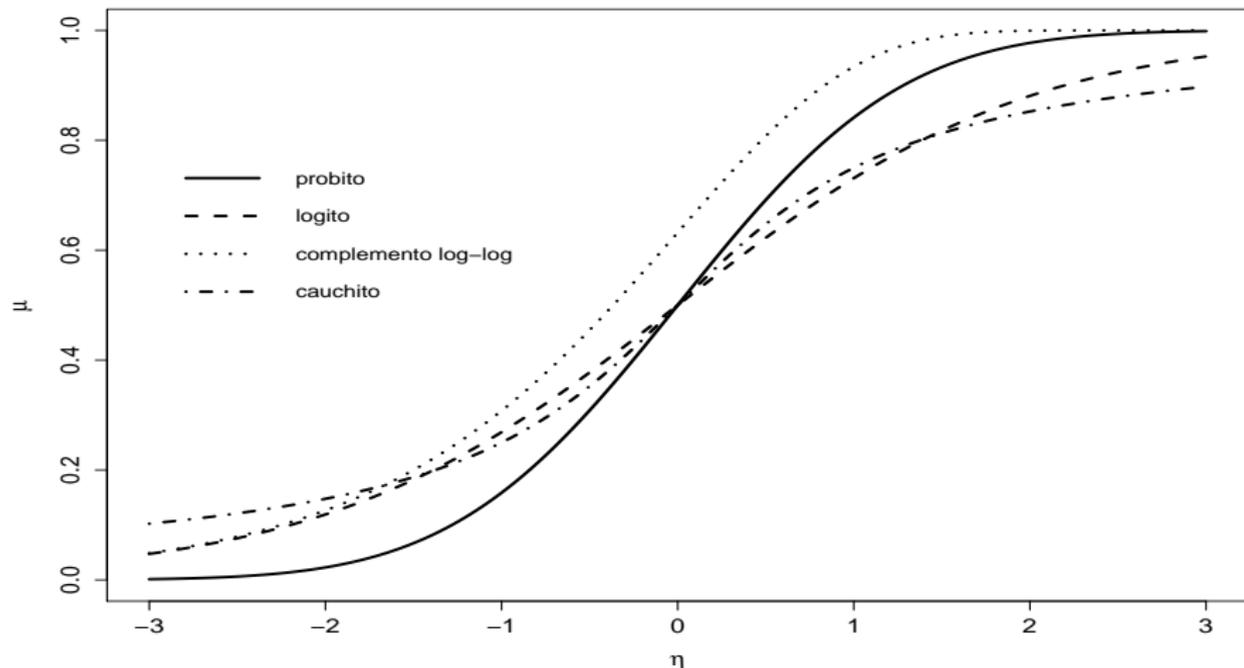
$$p_i = F(\eta_i) = \int_{-\infty}^{\eta_i} \frac{1}{\pi(1+x)^2} dx = \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \text{ (regressão cauchito). Assim, } F^{-1}(p_i) = \tan(\pi(p_i - 0.5))$$

- Distribuição valor extremo padrão:

$$p_i = F(\eta_i) = \int_{-\infty}^{\eta_i} e^{x-e^x} dx = 1 - e^{-e^{\eta_i}} \text{ (regressão extremito).}$$

Assim, $F^{-1}(p_i) = \ln(-(\ln(1 - p_i)))$

Funções de ligação para médias no intervalo (0,1)



Modelo de regressão para os dados de vasoconstricção

$$Y_i \stackrel{ind.}{\sim} \text{Bernoulli}(p_i)$$
$$\text{logito}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$
$$\rightarrow p_i = \frac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}}}, i = 1, 2, \dots, n$$

- Y_i : ocorrência (1) ou não (0) de vaso constricção.
- x_{1i} : logaritmo natural do volume de ar inspirado da i -ésima observação; x_{2i} : logaritmo natural da razão de ar inspirado da i -ésima observação.
- $F(\cdot)$: corresponde à fda de uma distribuição logística padrão (portanto o nome regressão logística). Nesse caso, o $\text{logito}(\cdot)$ é a função de ligação.

Modelo de regressão para os dados de vasoconstricção

- Interpretação dos parâmetros. Defina $l(p_i) = \text{logito}(p_i)$.
- Se $x_{1j} = x_{2j} = 0$, então $p_i = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$.
- Defina $l_1(p_{i+1}) = \beta_0 + \beta_1(x_{1i} + 1) + \beta_2 x_{2i}$ e
 $l_1(p_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$. Então
 $l_1(p_{i+1}) - l_1(p_i) = \beta_1 \rightarrow \frac{p_{i+1}/(1 - p_{i+1})}{p_i/(1 - p_i)} = e^{\beta_1}$ (**razão de chances em relação à primeira covariável**).
- Analogamente, defina $l_2(p_{i+1}) = \beta_0 + \beta_1 x_{1i} + \beta_2(x_{2i} + 1)$ e
 $l_2(p_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$. Então
 $l_2(p_{i+1}) - l_2(p_i) = \beta_2 \rightarrow \frac{p_{i+1}/(1 - p_{i+1})}{p_i/(1 - p_i)} = e^{\beta_2}$ (**razão de chances em relação à segunda covariável**).

Inferência para o modelo

- Defina $\eta_i = \sum_{j=1}^p \beta_j x_{ji} = \mathbf{X}'_i \boldsymbol{\beta}$, em que \mathbf{X}'_i é a i -ésima linha da matriz \mathbf{X} e $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, em que $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)'$ e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$.

Em que:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}.$$

Inferência para o modelo

- Verossimilhança

$$L(\beta) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i}$$

- Logverossimilhança.

$$l(\beta) = \sum_{i=1}^n [y_i \ln p_i + (1 - y_i) \ln(1 - p_i)] \quad (1)$$

Estimação paramétrica (ligação canônica)

- Temos que encontrar o vetor de derivadas de $l(\beta)$ com relação à β ,

$$\mathbf{S}(\beta) = \frac{\partial l(\beta)}{\partial \beta} \text{ em que}$$

$$\mathbf{S}(\beta) = \begin{bmatrix} \frac{\partial l(\beta)}{\partial \beta_1} \\ \frac{\partial l(\beta)}{\partial \beta_2} \\ \vdots \\ \frac{\partial l(\beta)}{\partial \beta_p} \end{bmatrix} \quad (2)$$

- Depois, devemos resolver o sistema de equações

$$\left\{ \mathbf{S}(\tilde{\boldsymbol{\beta}}) = \mathbf{0}_{(p \times 1)} \right. , \quad (3)$$

em que $\mathbf{0}_{(p \times 1)}$ é um vetor de zeros de dimensão $(p \times 1)$.

- Naturalmente, espera-se que a solução encontrada seja ponto de máximo no espaço $(p+1)$ (matriz Hessiana tem de ser negativa definida, o que não é suficiente).
- Para obtermos $\mathbf{S}(\boldsymbol{\beta})$ temos duas opções. Derivar $\mathbf{S}(\boldsymbol{\beta})$ diretamente com relação ao vetor $\boldsymbol{\beta}$ ou obter a derivada com relação à cada parâmetro $\beta_j, j = 1, \dots, p$ e deduzir a forma matricial (equação (4)) dessas derivadas.
- Vamos considerar a primeira opção.

Derivadas matriciais úteis

- Alguns resultados:

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{A}, (\mathbf{A}', \text{ se } \mathbf{A} \text{ for um vetor linha})$$

$$\frac{\partial \mathbf{x}'\mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}', (\mathbf{A}, \text{ se } \mathbf{A} \text{ for um vetor coluna})$$

$$\frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}$$

$$\frac{\partial \mathbf{x}'\mathbf{Ax}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x}, (2\mathbf{Ax}, \text{ se } \mathbf{A} \text{ for simétrica})$$

Inferência para o modelo

■ Vetor escore

$$\mathbf{s}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}) = \sum_{i=1}^n \left(\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i} \right) \frac{\partial p_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \left(\frac{y_i - p_i}{p_i(1-p_i)} \right) \frac{\partial p_i}{\partial \boldsymbol{\beta}}$$

em que $\frac{\partial p_i}{\partial \boldsymbol{\beta}} = \frac{\partial F(\eta_i)}{\partial \boldsymbol{\beta}}$ é um vetor, ou seja,

$$\frac{\partial F(\eta_i)}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \frac{\partial F(\eta_i)}{\partial \beta_1} \\ \frac{\partial F(\eta_i)}{\partial \beta_2} \\ \vdots \\ \frac{\partial F(\eta_i)}{\partial \beta_p} \end{bmatrix}$$

Inferência para o modelo

- Pela regra da cadeia e pelo fato de $F(\cdot)$ ser uma fda, temos que

$$\frac{\partial F(\eta_i)}{\partial \beta} = \frac{\partial F(\eta_i)}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta} = f(\eta_i) \mathbf{X}_i,$$

em que $f(\cdot)$ é a fdp associada à $F(\cdot)$ e $\frac{\partial \eta_i}{\partial \beta} = \mathbf{X}_i$, pois

$$\frac{\partial \eta_i}{\partial \beta_k} = \frac{\partial \sum_{j=1}^p \beta_j x_{ji}}{\partial \beta_k} = x_{ki}, k = 1, 2, \dots, p.$$

Inferência para o modelo

- Logo

$$\mathbf{S}(\boldsymbol{\beta}) = \sum_{i=1}^n \left((y_i - p_i) \frac{f(\eta_i)}{p_i(1-p_i)} \right) \mathbf{X}_i = \mathbf{X}' \mathbf{V} (\mathbf{y} - \boldsymbol{\mu})$$

em que $\mathbf{y} = (y_1, \dots, y_n)'$, $\boldsymbol{\mu} = (p_1, \dots, p_n)'$ e

$\mathbf{V} = \text{diag}(f(\eta_1)/(p_1(1-p_1)), \dots, f(\eta_n)/(p_n(1-p_n)))$ (exercício).

- Podemos ainda escrever $\mathbf{S}(\boldsymbol{\beta}) = \sum_{i=1}^n h_i(\boldsymbol{\beta}) g_i(\boldsymbol{\beta}) \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}$, em que

$$h_i(\boldsymbol{\beta}) = y_i - p_i \text{ e } g_i(\boldsymbol{\beta}) = \frac{f(\eta_i)}{p_i(1-p_i)}.$$

- Podemos notar que o sistema de equações (3) não tem, em geral, solução explícita pois, via de regra, μ_i é uma função não linear de β (quando se considera a função de ligação identidade, é possível obter solução explícita).
- Há várias opções de algoritmos numéricos para resolução de sistemas não lineares: Newton-Raphson, Escore de Fisher (EF), Nelder-Mead, BFGS entre outros.
- Por simplicidade e por apresentar, em geral, um excelente desempenho, consideraremos o algoritmo EF.

- Para isso precisaremos da informação de Fisher (ela também será útil na obtenção de resultados assintóticos):

$$I(\beta) = -\mathcal{E}(\mathbf{H}(\beta)) = -\mathcal{E}\left(\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'}\right)$$

em que

$$\mathbf{H}(\beta) = \begin{bmatrix} \frac{\partial^2 l(\beta)}{\partial \beta_1^2} & \frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_p} \\ \cdot & \frac{\partial^2 l(\beta)}{\partial \beta_2^2} & \cdots & \frac{\partial^2 l(\beta)}{\partial \beta_2 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdot & \frac{\partial^2 l(\beta)}{\partial \beta_p^2} \end{bmatrix}$$

Inferência para o modelo

- A matriz Hessiana é dada por $H(\beta) = \frac{\mathbf{S}(\beta)}{\partial\beta'} = \frac{I(\beta)}{\partial\beta\partial\beta'}$.
- A matriz de informação de Fisher é dada por $I(\beta) = -\mathcal{E}(\mathbf{H}(\beta))$ em que $\mathcal{E}(\cdot)$ é calculada em termos da distribuição de $\mathbf{Y} = (Y_1, \dots, Y_n)'$. Note, ainda, que $\mathcal{E}(h_i(\beta)) = \mathcal{E}(Y_i - p_i) = 0$.
- Pela regra da cadeia, temos que

$$\begin{aligned} \mathbf{H}(\beta) &= \sum_{i=1}^n \left[h_i(\beta) \frac{\partial\eta_i}{\partial\beta} \left(\frac{\partial g_i(\beta)}{\partial\beta} \right)' + h_i(\beta) g_i(\beta) \frac{\partial\eta_i}{\partial\beta\partial\beta'} \right. \\ &\quad \left. + g_i(\beta) \frac{\partial\eta_i}{\partial\beta} \left(\frac{\partial h_i(\beta)}{\partial\beta} \right)' \right] \end{aligned}$$

Inferência para o modelo

- Assim,

$$\begin{aligned} I(\beta) &= -\mathcal{E} \left\{ \sum_{i=1}^n \left[h_i(\beta) \frac{\partial \eta_i}{\partial \beta} \left(\frac{\partial g_i(\beta)}{\partial \beta} \right)' + h_i(\beta) g_i(\beta) \frac{\partial \eta_i}{\partial \beta \partial \beta'} \right. \right. \\ &\quad \left. \left. + g_i(\beta) \frac{\partial \eta_i}{\partial \beta} \left(\frac{\partial h_i(\beta)}{\partial \beta} \right)' \right] \right\} \\ &= -\sum_{i=1}^n g_i(\beta) \frac{\partial \eta_i}{\partial \beta} \left(\frac{\partial h_i(\beta)}{\partial \beta} \right)' = \sum_{i=1}^n g_i(\beta) \frac{\partial \eta_i}{\partial \beta} \left(\frac{\partial p_i(\beta)}{\partial \beta} \right)', \end{aligned}$$

pois $\frac{\partial h_i(\beta)}{\partial \beta} = -\frac{\partial p_i(\beta)}{\partial \beta}$, lembrando que $h_i(\beta) = (y_i - p_i)$.

Inferência para o modelo

- Portanto,

$$I(\beta) = \mathbf{X}' \mathbf{W} \mathbf{X},$$

em que $\mathbf{W} = \text{diag}(f(\eta_1)^2/(p_1(1-p_1)), \dots, f(\eta_n)^2/(p_n(1-p_n)))$

- Se $p_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$ (regressão logística), então

$$\mathbf{S}(\beta) = \mathbf{X}'(\mathbf{y} - \boldsymbol{\mu})$$

e

$$I(\beta) = \mathbf{X}' \mathbf{D} \mathbf{X}$$

(exercício), em que $\mathbf{D} = \text{diag}(p_1(1-p_1), \dots, p_n(1-p_n))$.

Inferência para o modelo

- Independente da escolha de $F(\cdot)$, o sistema de equações $\mathbf{S}(\hat{\beta}) = \mathbf{0}$ não tem solução explícita e algum método de otimização numérica, como o algoritmo escore de Fisher, deve ser utilizado para obter-se as estimativas de MV.

Algoritmo escore de Fisher

- Seja $\beta^{(0)}$ uma estimativa inicial de β (chute inicial), então faça

$$\beta^{(t+1)} = \beta^{(t)} + \mathbf{I}^{-1}(\beta^{(t)})\mathbf{S}(\beta^{(t)}), t = 1, 2, \dots \quad (4)$$

até que algum critério de convergência seja satisfeito, como

$$|l(\beta^{(t+1)}) - l(\beta^{(t)})| < \epsilon, \epsilon > 0,$$

em que $l(\cdot)$ é a logverossimilhança (equação (1)).

Algoritmo escore de Fisher

- A equação (4) pode ser reescrita como

$$\boldsymbol{\beta}^{(t+1)} = \left(\mathbf{X}' \mathbf{W}^{(t)} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{W}^{(t)} \mathbf{z}^{(t)},$$

em que $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-1/2} \mathbf{D}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$.

- Estimativas iniciais: é usual considerarmos $\boldsymbol{\eta}^{(0)} = F^{-1}(\boldsymbol{p}^{(0)})$, em que $\boldsymbol{p} = (p_1, \dots, p_n)$, e $p_i^{(0)}$ = proporção de sucessos na amostra, se $y_i = 1$ e $p_i^{(0)}$ = proporção de fracassos na amostra, caso contrário.

Mais sobre inferência

- Para n suficientemente grande, $\hat{\beta} \approx N_p(\beta, I^{-1}(\beta))$. Na prática, trabalhamos com $I^{-1}(\tilde{\beta})$, em que $\tilde{\beta}$ é a emv de β , obtida através do algoritmo Escore de Fisher.
- Defina, $\hat{\sigma}_i^2$: o i -ésimo elemento da diagonal principal de $I^{-1}(\hat{\beta})$ e $\tilde{\sigma}_i^2$: o i -ésimo elemento da diagonal principal de $I^{-1}(\tilde{\beta})$.
- Assim, hipóteses do tipo $H_0 : \beta_i = \beta_0$ vs $\beta_i \neq \beta_0$, podem ser testadas através da estatística $Z_t = \frac{\hat{\beta}_i - \beta_0}{\sqrt{\hat{\sigma}_i^2}}$, rejeitando-se H_0 quando $p\text{-valor} \leq \alpha$, $p\text{-valor} \approx 2P(Z \geq |z_t| | H_0)$, $Z \sim N(0, 1)$ e $z_t = \frac{\tilde{\beta}_i - \beta_0}{\sqrt{\tilde{\sigma}_i^2}}$, em que $\hat{\beta}_i$ é o estimador de MV de β_i e $\tilde{\beta}_i$ a respectiva estimativa.

Mais sobre inferência

- $IC(\beta_i; \gamma) = \left[\hat{\beta}_i - z_{(1+\gamma)/2} \sqrt{\hat{\sigma}_i^2}; \hat{\beta}_i + z_{(1+\gamma)/2} \sqrt{\hat{\sigma}_i^2} \right]$, em que $P(Z \leq z_{(1+\gamma)/2}) = \gamma$.
- Hipóteses do tipo $H_0 : \mathbf{C}\beta = \mathbf{M}$ vs $H_1 : \mathbf{C}\beta \neq \mathbf{M}$ podem ser testadas através da estatística

$$Q_t = \left(\mathbf{C}\hat{\beta} - \mathbf{M} \right)' \left(\mathbf{C}\mathbf{I}^{-1}(\hat{\beta})\mathbf{C}' \right)^{-1} \left(\mathbf{C}\hat{\beta} - \mathbf{M} \right).$$

Sob H_0 e para n suficientemente grande, $Q_t \approx \chi_c^2$, em que c é o número de linhas de \mathbf{C} .

- Assim, rejeita-se H_0 se $p - \text{valor} \leq \alpha$, em que

$p - \text{valor} \approx P(X \geq q_t | H_0)$, em que $X \sim \chi_c^2$

$$q_t = \left(\mathbf{C}\tilde{\beta} - \mathbf{M} \right)' \left(\mathbf{C}\mathbf{I}^{-1}(\tilde{\beta})\mathbf{C}' \right)^{-1} \left(\mathbf{C}\tilde{\beta} - \mathbf{M} \right).$$

Utilizando o programa R

- função *glm* (disponível quando na instalação básica do R).
- Sua sintaxe se assemelha bastante com a da função *lm* (na forma de definir a estrutura de regressão)

(com intercepto)

```
result <- glm(Y ~ X1 + X2 + ... + Xp,
```

```
family=binomial(link="logit"));
```

```
summary(result)
```

(sem intercepto)

```
result <- glm(Y ~ -1 + X1 + X2 + ... + Xp,
```

```
family=binomial(link="logit"));
```

```
summary(result)
```

Resumo dos comandos básicos

Distribuição (family)	Função de ligação (link)
binomial	logit (logito), probit (probit), cauchit (cauchito), log , cloglog (complemento log-log)

Distribuição (family)	Função de ligação (link)
identity	$g(\mu) = \mu$
log	$g(\mu) = \ln(\mu)$
inverse	$g(\mu) = \frac{1}{\mu}$
sqrt	$g(\mu) = \sqrt{\mu}$
logit	$g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$
probit	$g(\mu) = \Phi^{-1}(\mu)$; $\Phi^{-1}(\cdot)$ inversa da cdf da $N(0, 1)$
cauchito	$g(\mu) = F^{-1}(\mu)$; $F^{-1}(\cdot)$ inversa da cdf da Cauchy(0, 1)
cloglog	$g(\mu) = \ln(-\ln(1 - \mu))$

Outros comandos

Quantidade	Comando
$\tilde{\beta}$	<code>coef(result)</code>
$\widetilde{Cov}(\hat{\beta})$	<code>vcov(result)</code>

- Para ver os resultados disponibilizados pela função *glm* faça `names(fit.model)` ou `names(summary(fit.model))`.

Voltando ao conjunto de dados (os testes se referem à nulidade de cada parâmetro)

Parâmetro	Estimativa	EP	IC(95%)	Estat. Z_t	p-valor
β_0	-2,87	1,32	[-5,46 ; -0,29]	-2,18	0,0295
β_1	5,17	1,86	[1,52 ; 8,83]	2,78	0,0055
β_2	4,56	1,83	[0,96 ; 8,16]	2,48	0,0131

Todos os parâmetros são significativos.

Probabilidades e valores preditos

- Probabilidades de ocorrência de vasoconstrição preditas:

$$\tilde{p}_i = \frac{e^{\tilde{\beta}_0 + \tilde{\beta}_1 x_{1i} + \tilde{\beta}_2 x_{2i}}}{1 + e^{\tilde{\beta}_0 + \tilde{\beta}_1 x_{1i} + \tilde{\beta}_2 x_{2i}}}$$

- Ocorrências de vasoconstrição preditas: simula-se u , $U \sim U(0, 1)$, se $\tilde{p}_i \geq u$, então $\tilde{Y}_i = 1$, caso contrário, $\tilde{Y}_i = 0$.
- Podemos comparar tanto as probabilidades quanto os valores preditos com os valores observados. Atenção: no caso dos valores preditos note que, como estamos gerando somente uma réplica, tal análise deve ser considerada com cautela.

- Como gerar intervalos de confiança para

$$p_i = \frac{1}{1 + e^{-\eta_i}} = \frac{1}{1 + e^{-\sum_{j=1}^p x_{ji}\beta_j}}?$$

- Método delta. Estimador $\hat{p}_i = \frac{1}{1 + e^{-\sum_{j=1}^p x_{ji}\hat{\beta}_j}}$.

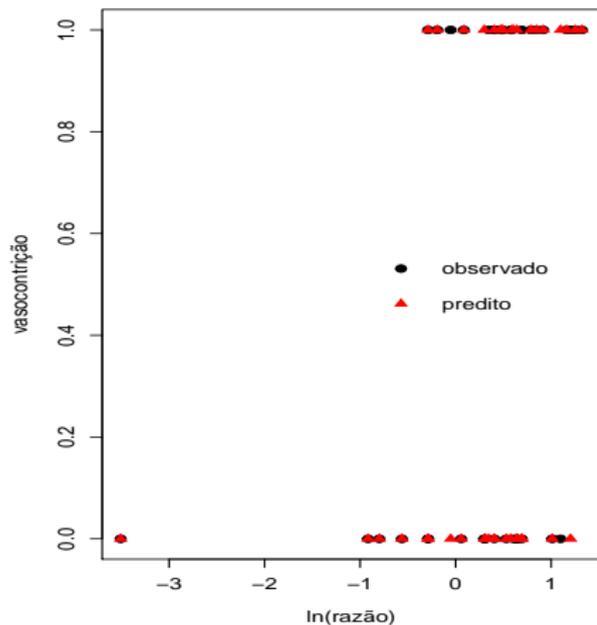
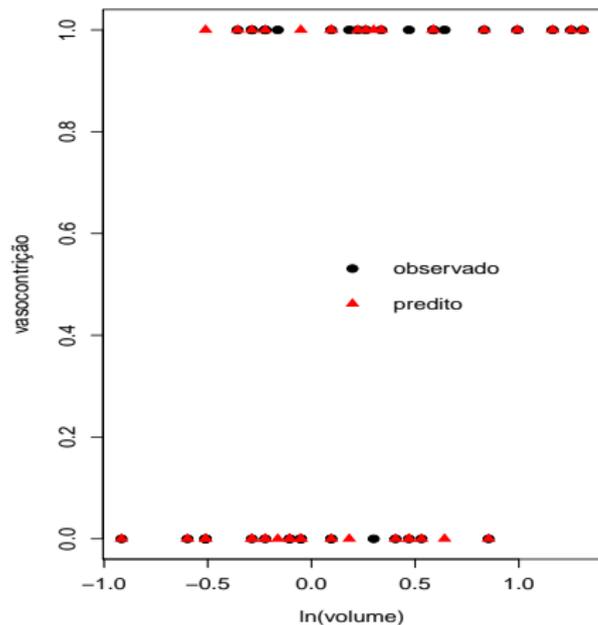
- Para n suficientemente grande, temos que $\hat{p}_i \approx N(p_i, \sigma_{p_i}^2)$, em que $\sigma_{p_i}^2 = \Psi'_\beta \Sigma_\beta \Psi_\beta$, em que

$$\Psi'_\beta = \begin{bmatrix} \frac{\partial p_i}{\partial \beta_1} & \frac{\partial p_i}{\partial \beta_2} & \cdots & \frac{\partial p_i}{\partial \beta_p} \end{bmatrix}$$

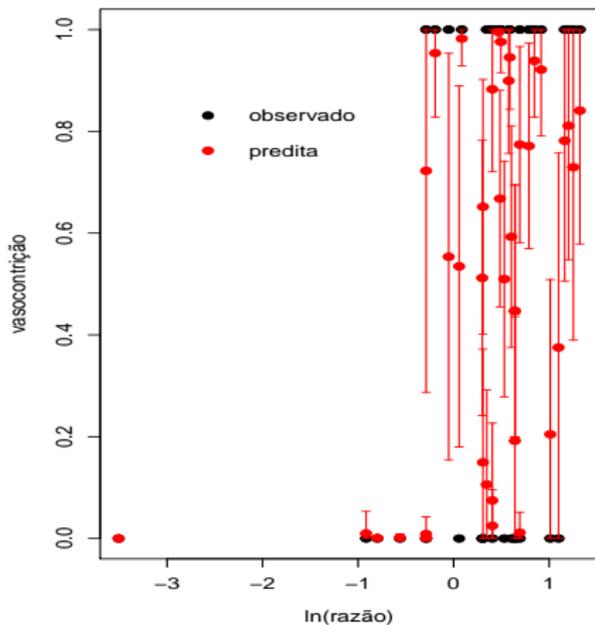
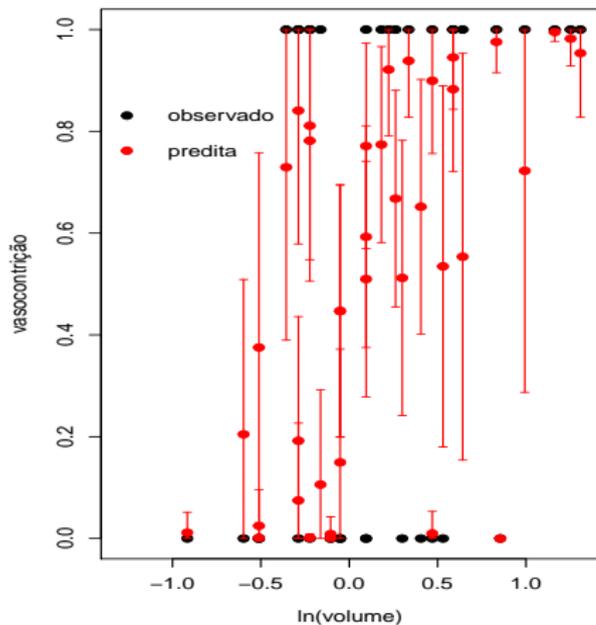
- Assim $IC(p_i, \gamma) \approx [\hat{p}_i - z_{(1+\gamma)/2} \hat{\sigma}_{p_i}^2; \hat{p}_i + z_{(1+\gamma)/2} \hat{\sigma}_{p_i}^2]$, em que $\hat{\sigma}_{p_i}^2 = \hat{\Psi}'_\beta \hat{\Sigma}_\beta \hat{\Psi}_\beta$ e $P(Z \leq z_{(1+\gamma)/2}) = \frac{1+\gamma}{2}, Z \sim N(0, 1)$.

- A função “predict” do R gera estimativas pontuais e intervalares para p_i , conforme descrito acima.

Valores observados e preditos pelo modelo



Probabilidades previstas e valores observados pelo modelo



Perguntas

- Como gerar intervalos de confiança para $\frac{e^{\beta_0}}{1+e^{\beta_0}}$, e^{β_1} e e^{β_2} ?
 - Método delta.
 - Fazer um IC para o parâmetro original e depois calcular o IC para a transformação.
 - Reamostragem.
- Como verificar as suposições do modelo?
 - Estatísticas de qualidade de ajuste.
 - Resíduo componente do desvio.
- Vamos nos concentrar na regressão logística ($F^{-1}(p_i) = \text{logito}(p_i)$).

Intervalos de confiança para funções de interesse

- Sejam $g_1(\boldsymbol{\beta}) \equiv \tau_1 = \frac{e^{\beta_0}}{1+e^{\beta_0}}$, $g_2(\boldsymbol{\beta}) \equiv \tau_2 = e^{\beta_1}$ e $g_3(\boldsymbol{\beta}) \equiv \tau_3 = e^{\beta_2}$.
- Seja $\hat{\boldsymbol{\beta}}$ o estimador de MV de $\boldsymbol{\beta}$. Já vimos que, para n suficientemente grande, $\hat{\boldsymbol{\beta}} \approx N(\boldsymbol{\beta}, \boldsymbol{\Sigma}_\beta)$, em que $\boldsymbol{\Sigma}_\beta = \mathbf{I}^{-1}(\boldsymbol{\beta})$.
- O método delta nos diz que, para n suficientemente grande, $\hat{\tau}_i \approx N(\tau_i, \boldsymbol{\Psi}_i \boldsymbol{\Sigma}_\beta \boldsymbol{\Psi}_i')$, em que

$$\boldsymbol{\Psi}_i = \begin{bmatrix} \frac{\partial}{\partial \beta_0} g_i(\boldsymbol{\beta}) & \frac{\partial}{\partial \beta_1} g_i(\boldsymbol{\beta}) & \frac{\partial}{\partial \beta_2} g_i(\boldsymbol{\beta}) \end{bmatrix}$$

Intervalos de confiança para funções de interesse

- Nesse caso,

$$\Psi_1 = \begin{bmatrix} \frac{e^{\beta_0}}{(1+e^{\beta_0})^2} & 0 & 0 \end{bmatrix}, \quad \Psi_2 = \begin{bmatrix} 0 & e^{\beta_1} & 0 \end{bmatrix},$$

$$\Psi_3 = \begin{bmatrix} 0 & 0 & e^{\beta_2} \end{bmatrix}$$

- Assim $IC(\tau_i, \gamma) = \left[\hat{\tau}_i - z_{(1+\gamma)/2} \sqrt{\hat{\psi}_i}; \hat{\tau}_i + z_{(1+\gamma)/2} \sqrt{\hat{\psi}_i} \right]$, em que $P(Z \geq z_{(1+\gamma)/2}) = \frac{1+\gamma}{2}$ e $\hat{\psi}_i = \hat{\Psi}_i \hat{\Sigma}_\beta \hat{\Psi}_i'$, $Z \sim N(0, 1)$ (lembrando que este é um IC assintótico).

Intervalos de confiança para funções de interesse

Parâmetro	IC (transformação)	IC (método delta)	IC (simul.)
τ_1	[< 0,01 ; 0,43]	[-0,08 ; 0,18]	[< 0,01 ; 0,22]
τ_2	[4,59 ; 6862,99]	[-471,35 ; 826,48]	[21,80 ; 610450285,37]
τ_3	[2,61 ; 3511,02]	[-249,12 ; 440,61]	[9,90 ; 50770836,90]

Neste caso, os IC's obtidos através do método delta, devem ser truncados à esquerda do zero. Exercício: obter os intervalos de confiança através de reamostragem.

Verificação da qualidade de ajuste do modelo

- No modelo em questão, temos, essencialmente, as seguintes suposições a serem avaliadas.
 - Apesar do modelo ser heterocedástico ($\mathcal{V}(Y_i) = p_i(1 - p_i)$), a variância por ele imposta pode ser menor do que a observada (superdispersão) ou maior do que a observada (subdispersão).
 - As observações são independentes.
 - A função de ligação (nesse caso F^{-1} , F uma fda) é apropriada.

Verificação da qualidade de ajuste do modelo

- Função desvio: Seja $l(\boldsymbol{\mu}, \mathbf{y})$ a logverossimilhança do modelo ($\mathcal{E}(\mathbf{Y}) = \boldsymbol{\mu} = F(\boldsymbol{\eta})$). Lembremos que, para o modelo Bernoulli $\mu_i = p_i$.
- Para o modelo saturado $n = p$, ou seja, em que representamos a média de cada observação por ela mesma, o estimador de MV de μ_i é $\hat{\mu}_i = Y_i$. Nesse caso o estimador de $l(\boldsymbol{\mu}, \mathbf{y})$ é dado por $l(\mathbf{Y}; \mathbf{y})$.
- Seja $l(\hat{\boldsymbol{\mu}}, \mathbf{y})$ o estimador de MV da logverossimilhança sob o modelo em estudo, em que $\hat{\boldsymbol{\mu}} = F(\hat{\boldsymbol{\eta}})$ e $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

Verificação da qualidade de ajuste do modelo

- A função desvio (ou simplesmente desvio) é definida por

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \{l(\mathbf{Y}, \mathbf{y}) - l(\hat{\boldsymbol{\mu}}, \mathbf{y})\}$$

- No caso do modelo de regressão Bernoulli e com observações independentes, temos que

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = -2 \sum_{i=1}^n \{ \ln(1 - \hat{\mu}_i) I_{\{0\}}(y_i) + \ln(\hat{\mu}_i) I_{\{1\}}(y_i) \} = \sum_{i=1}^n D(y_i; \hat{\mu}_i).$$

- Contudo, em geral $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ não segue (mesmo assintoticamente) uma distribuição $\chi^2_{(n-p)}$, sob a hipótese de que o modelo em questão é adequado. Em nosso caso temos que $D(\mathbf{y}; \tilde{\boldsymbol{\mu}}) = 29,22$ e $p = 0,7807$ (p-valor assintótico, supondo a aproximação válida), indicando um bom ajuste do modelo.

Verificação da qualidade de ajuste do modelo

- É aconselhável obter um p-valor para a estatística $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ por reamostragem.
- Algoritmo
 - 1 Ajuste o modelo (estime seus parâmetros) por MV e calcule o desvio (desvo).
 - 2 Para $j = 1, \dots, R$, faça
 - Simule n variáveis Bernoulli de acordo com o modelo.
 - Ajuste o modelo considerando as variáveis simuladas anteriormente, e calcule o respectivo desvio (desvs).

3 Assim, $p - \text{valor} = \frac{1}{R} \sum_{j=1}^R \mathbb{1}(\text{desvs}_j \geq \text{desvo})$.

Verificação da qualidade de ajuste do modelo

- Uma outra forma de verificar a qualidade do ajuste do modelo, é através da análise de resíduos.
- Utilizar o resíduo padronizado (semelhante aquele utilizado em modelos de regressão normais lineares), ou seja, $\frac{Y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}}$ não é apropriado.
- Particularmente, esse resíduo não terá distribuição normal (mesmo sob a validade das hipóteses do modelo).

- Podemos perceber que o resíduo studentizado, muito utilizado para verificar a qualidade de ajuste da classe de MRNLH, dificilmente apresentará normalidade assintótica (embora poderia ser usado para verificar a presença de outliers ou problemas na predição dos valores) para os MLG.
- Portanto, a utilização de outro resíduo se faz necessária.
- Paula (2013) apresenta uma revisão muito boa sobre vários resíduos.
- Nos concentraremos no resíduo componente do desvio (RCD).
- Um fato interessante é que, na definição dos MLG, não aparece nenhum tipo de “erro”, como ocorre nos MRNLH.

Introdução

- Um resíduo deve apresentar um comportamento específico quando o modelo está bem ajustado e outro quando o modelo não o estiver.
- O ideal é que, dependendo de qual suposição (ou suposições, p.e., distribuição da variável resposta, independência, função de ligação e forma do preditor linear) não esteja(m) sendo satisfeita(s), alguma mudança específica ocorra em seu comportamento (conforme discutido anteriormente).
- Naturalmente, outras metodologias, para além dos resíduos, podem ser utilizadas para verificar o afastamento de suposições específicas.

- A forma geral do RCD, para a i -ésima observação, é dada por:

$$T_{D_i} = \frac{d(Y_i, \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}}$$

em que

- $d(Y_i, \hat{\mu}_i) = \text{sign}(Y_i - \hat{\mu}_i) \sqrt{2} \sqrt{D(Y_i; \hat{\mu}_i)}$, em que
 $D(Y_i; \hat{\mu}_i) = -[\ln(1 - \hat{\mu}_i)I_{\{0\}}(Y_i) + \ln(\hat{\mu}_i)I_{\{1\}}(Y_i)]$
- \hat{h}_{ii} é o i -ésimo elemento da diagonal principal da matriz
 $\hat{H} = \hat{W}^{1/2} \mathbf{X} (\mathbf{X}' \hat{W} \mathbf{X})^{-1} \mathbf{X}' \hat{W}^{1/2}$, em que \mathbf{X} e \hat{W} são como
 definidas na parte de estimação.

Verificação da qualidade de ajuste do modelo

- Alternativa: resíduo componente do desvio (RCD). Nesse caso, é dado por

$$T_{D_i} = -\frac{(2|\ln(1 - \hat{p}_i)|)^{1/2}}{\sqrt{1 - \hat{h}_{ii}}} I_{\{0\}}(y_i) + \frac{(2|\ln \hat{p}_i|)^{1/2}}{\sqrt{1 - \hat{h}_{ii}}} I_{\{1\}}(y_i)$$

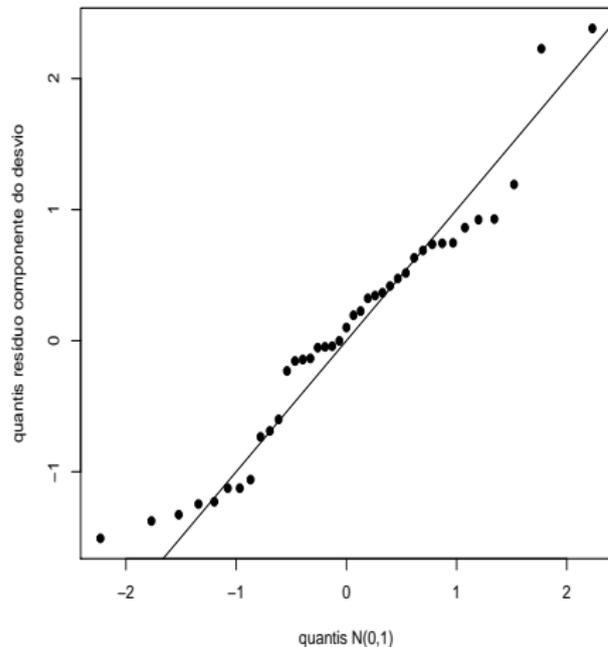
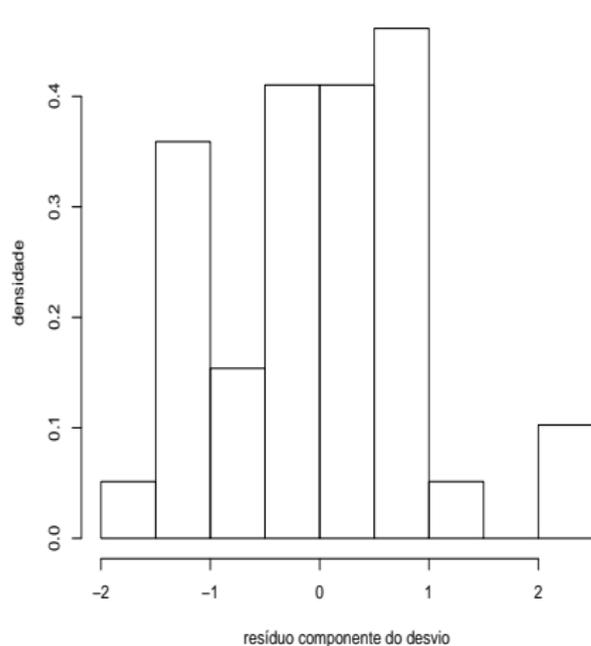
em que $\hat{h}_{ii} = \hat{p}_i(1 - \hat{p}_i)\mathbf{X}'_i (\mathbf{X}'\hat{\mathbf{V}}\mathbf{X})^{-1} \mathbf{X}_i$ e

$$\hat{\mathbf{V}} = \text{diag}(\hat{p}_1(1 - \hat{p}_1), \dots, \hat{p}_n(1 - \hat{p}_n)).$$

- Para n suficientemente grande e sob a validade das suposições do modelo, $T_{D_i} \approx N(0, 1)$ (para o modelo de regressão para dados binários, isso muitas vezes não ocorre).
- Pergunta: construir um histograma e/ou qq-plots para os RCD's é apropriado (suficiente) para avaliar o comportamento dos resíduos?

- Williams (1984) verificou através de simulações que a distribuição de t_{D_i} tende a estar mais próxima da normalidade do que as distribuições de outros resíduos (veja Paula (2013)).
- Utilizando resultados de Cox and Snell (1968), pode-se demonstrar que $\mathcal{E}(D^*(Y_i, \mu_i)) \approx 0$ e $\mathcal{V}(D^*(Y_i, \mu_i)) \approx 1 - h_{ii}$ em que os termos negligenciados são $O(n^{-1})$. Esses resultados reforçam a padronização do RDC por $\sqrt{1 - \hat{h}_{ii}}$.
- A estimativa do RCD é obtida substituindo-se os estimadores nele presentes por suas respectivas estimativas, bem como Y_i pelos valores observados y_i .

Histograma e qq-plot do rcd para o Exemplo 11



Procedimento para se gerar o gráfico de envelopes com o RCD

- 1) Ajuste o modelo de regressão (estima-se os parâmetros do modelo) obtendo-se as estimativas de MV ($\tilde{\beta}$) e calcule o RCD para cada observação, $(t_{D_i}), i = 1, 2, \dots, n$.
- 2) De posse das estimativas de MV, repita os passos (a) e (b) m vezes.
 - a) Simule n variáveis aleatórias ind. Bernoulli(\tilde{p}_i), com $\tilde{p}_i = F(\tilde{\eta}_i)$,
 $\tilde{\eta}_i = \mathbf{X}'_i \tilde{\beta}$.
 - b) Ajuste o modelo de regressão considerando as variáveis simuladas no item a) e obtenha o RCD para cada observação (i) em cada réplica (j).

Procedimento para se gerar o gráfico de envelopes com o RCD

- 3) Ao final teremos uma matriz com os RCD's, ou seja $t_{D_{ij}}^*$, $i=1,\dots,n$, (tamanho da amostra) $j=1,\dots,m$ (réplica).

$$\mathbf{T}_1 = \begin{bmatrix} t_{D_{11}}^* & t_{D_{12}}^* & \dots & t_{D_{1m}}^* \\ t_{D_{21}}^* & t_{D_{22}}^* & \dots & t_{D_{2m}}^* \\ \vdots & \vdots & \ddots & \vdots \\ t_{D_{n1}}^* & t_{D_{n2}}^* & \dots & t_{D_{nm}}^* \end{bmatrix}$$

Procedimento para se gerar o gráfico de envelopes com o RCD

- 4) Dentro de cada amostra, ordena-se, de modo crescente, os RCD's, obtendo-se $t_{D(i)j}^*$ (estatísticas de ordem):

$$\mathbf{T}_2 = \begin{bmatrix} t_{D(1)1}^* & t_{D(1)2}^* & \cdots & t_{D(1)m}^* \\ t_{D(2)1}^* & t_{D(2)2}^* & \cdots & t_{D(2)m}^* \\ \vdots & \vdots & \ddots & \vdots \\ t_{D(n)1}^* & t_{D(n)2}^* & \cdots & t_{D(n)m}^* \end{bmatrix}$$

- 5) Obtem-se os limites $t_{(i)I}^* = \min_{1 \leq j \leq m} t_{D(i)j}^*$ e $t_{(i)S}^* = \max_{1 \leq j \leq m} t_{D(i)j}^*$,
 $j = 1, 2, \dots, m$.

Procedimento para se gerar o gráfico de envelopes com o RCD

- 5) Na prática considera-se $t_{(i)l}^* = \frac{t_{D_{(i)(2)}}^* + t_{D_{(i)(3)}}^*}{2}$ e $t_{(i)s}^* = \frac{t_{D_{(i)(m-2)}}^* + t_{D_{(i)(m-1)}}^*}{2}$ (refinamento das estimativas do mínimo e máximo), em que $t_{D_{(i)(r)}}^*$ é a r -ésima estatística de ordem dentro de cada linha, $i = 1, 2, \dots, n$.

- Além disso, consideramos como a linha de referência

$$t_{(i)}^* = \frac{1}{m} \sum_{j=1}^m t_{D_{(i)j}}^*, i = 1, 2, \dots, n.$$

Outros gráficos de interesse

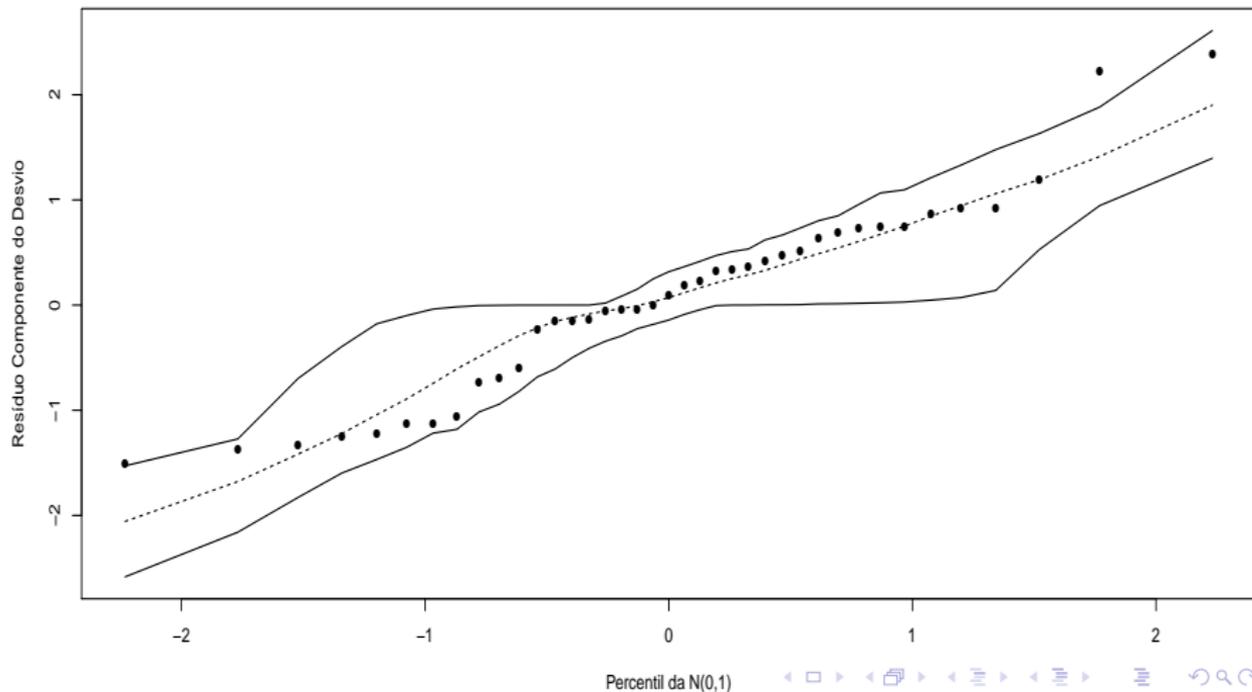
- $t_{D_i} \times$ ordem da observação: pontos aberrantes, heterogeneidade (heterocedasticidade) não capturada pelo modelo.
- $t_{D_i} \times F(\tilde{\eta}_i)$ (valor predito): pontos aberrantes.
- $\tilde{z}_i \times \tilde{\eta}_i$: adequabilidade da função de ligação e do preditor linear (η_i), em que $\tilde{z}_i = \tilde{\eta}_i + \tilde{W}_i^{-1/2} \tilde{D}_i^{-1/2} (y_i - \tilde{\mu}_i)$, em que $\eta_i = \mathbf{X}'_i \tilde{\beta}$ e

$$\mathbf{W} = \text{diag} (f(\tilde{\eta}_1)^2 / (\tilde{p}_1(1 - \tilde{p}_1)), \dots, f(\tilde{\eta}_n)^2 / (\tilde{p}_n(1 - \tilde{p}_n)))$$

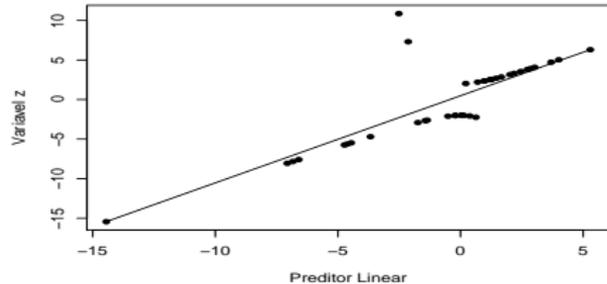
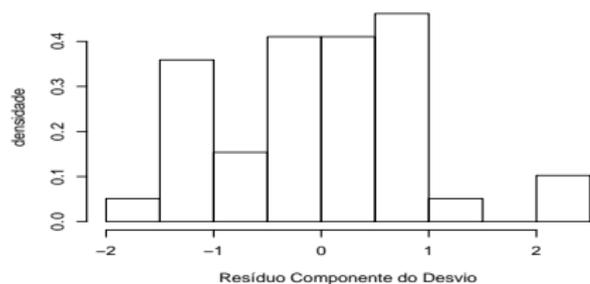
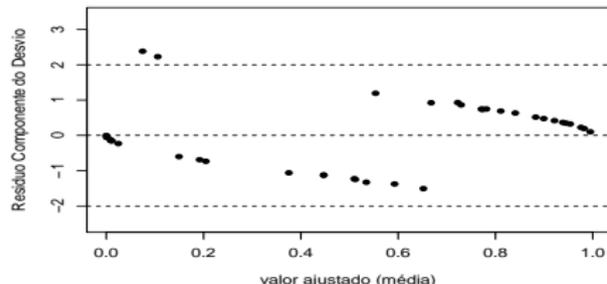
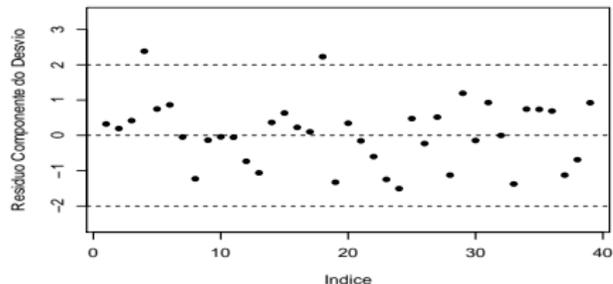
$$\mathbf{D} = \text{diag}(\tilde{p}_1(1 - \tilde{p}_1), \dots, \tilde{p}_n(1 - \tilde{p}_n))$$

Gráficos de envelopes para os RCD's

Gráfico de quantil-quantil normal



Gráficos de diagnóstico para os RCD's



Comentários

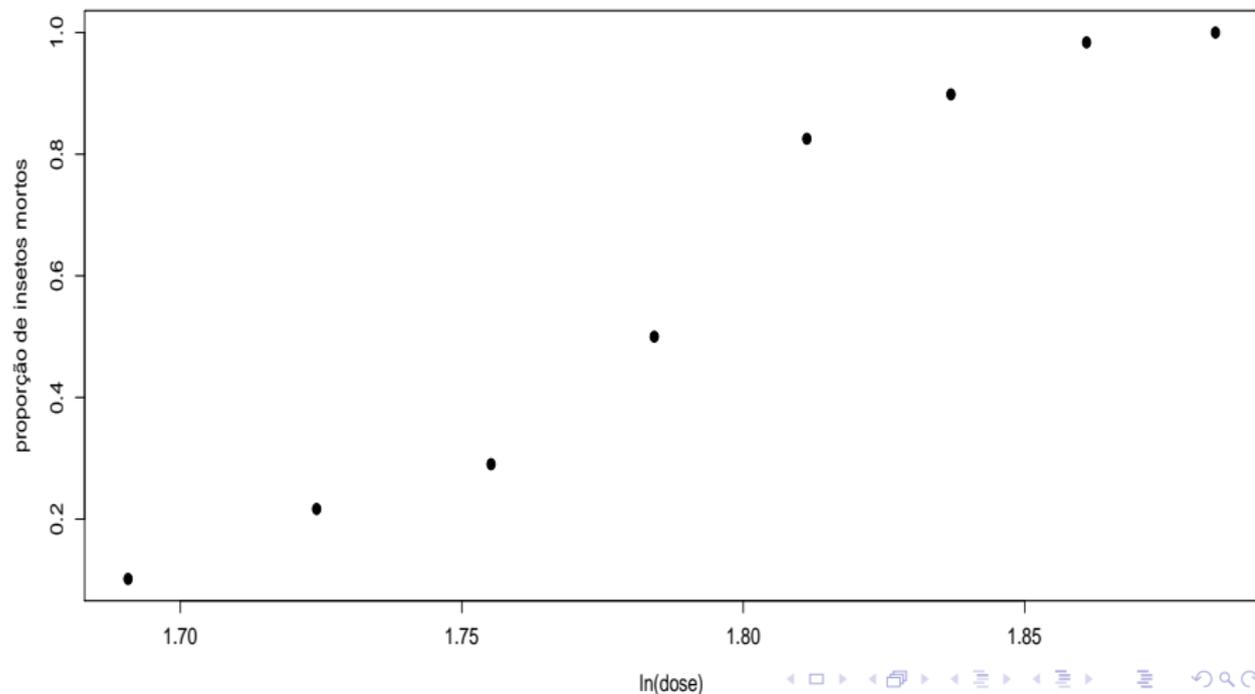
- A análise de diagnóstico indicou que o modelo se ajustou de modo razoável aos dados.
- Para finalizar: apresentar as estimativas pontuais e intervalares de vasoconstricção para diferentes valores do $\log(\text{volume})$ e $\log(\text{razão})$ de interesse do pesquisador (**exercício**).

Exemplo 12: mortalidade de besouros

- Dados relativos ao percentual de besouros mortos quando expostos à diferentes doses de disulfeto de carbono gasoso (CS_2).

Dose: $\log_{10} CS_2$	n° Besouros expostos	n° 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

Gráficos de dispersão



Exemplo 12: mortalidade de besouros

- Modelo 1

$$Y_i \stackrel{ind.}{\sim} \text{binomial}(m_i, p_i)$$
$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i, i = 1, 2, \dots, 8$$

- m_i : número de besouros expostos à dose i de CS_2 .
- Y_i : número de besouros expostos à dose i de CS_2 que morreram.
- x_i : dose (log da concentração de CS_2) à que os besouros do grupo i foram expostos.

Cont. do modelo 1

- Assim, $p_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$.
- β_0 é o logito $\left[\ln \left(\frac{p_i}{1-p_i} \right) \right]$ da proporção de besouros mortos submetidos à uma concentração de 1 unidade de CS_2 . Ou seja, se $x_i = \log_{10}(\text{concent}) = \log_{10}(1) = 0$ então $p_i = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$.

Cont. do modelo 1

- Sejam: $p_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$ e $p_{i+1} = \frac{e^{\beta_0 + \beta_1 (x_i + 1)}}{1 + e^{\beta_0 + \beta_1 (x_i + 1)}}$.
- Assim: $\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i$ e $\ln\left(\frac{p_{i+1}}{1-p_{i+1}}\right) = \beta_0 + \beta_1 (x_i + 1)$.
- Logo: $\ln\left(\frac{p_{i+1}}{1-p_{i+1}}\right) - \ln\left(\frac{p_i}{1-p_i}\right) = \ln\left(\frac{p_{i+1}/(1-p_{i+1})}{p_i/(1-p_i)}\right) = \beta_1$.
- Portanto, $\frac{p_{i+1}/(1-p_{i+1})}{p_i/(1-p_i)} = e^{\beta_1}$ (razão de chances).

Exemplo 12: mortalidade de besouros

■ Modelo 2

$$Y_i \stackrel{ind.}{\sim} \text{binomial}(m_i, p_i)$$
$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(x_i - \bar{x}), \bar{x} = \frac{1}{8} \sum_{i=1}^8 x_i, i = 1, 2, \dots, 8$$

- Neste caso, β_0 é o logito $\left[\ln\left(\frac{p_i}{1-p_i}\right)\right]$ da proporção de besouros mortos submetidos à uma concentração igual à \bar{x} unidades de CS_2 . Ou seja, se $x_i = \frac{1}{8} \sum_{i=1}^8 \log_{10}(\text{concent}_i)$, então $p_i = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$.
- As outras quantidades, incluindo o parâmetro β_1 , possuem as mesmas interpretações que no modelo 1, (substituindo-se x_i por $x_i - \bar{x}$).

Inferência para o modelo

- Defina $\eta_i = \sum_{j=1}^p \beta_j x_{ji} = \mathbf{X}'_i \boldsymbol{\beta}$, em que \mathbf{X}'_i é a i -ésima linha da matriz \mathbf{X} e $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, em que $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)'$ e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$. Assim, temos que $Y_i \stackrel{\text{ind.}}{\sim} \text{binomial}(m_i, p_i)$, $p_i = F(\eta_i)$, $i = 1, \dots, k$.
- Verossimilhança

$$L(\boldsymbol{\beta}) = \prod_{i=1}^k \binom{m_i}{y_i} p_i^{y_i} (1 - p_i)^{m_i - y_i} \propto \prod_{i=1}^k p_i^{y_i} (1 - p_i)^{m_i - y_i}$$

- Logverossimilhança.

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n [y_i \ln p_i + (m_i - y_i) \ln(1 - p_i)] + \text{const.} \quad (5)$$

Inferência para o modelo

- Os desenvolvimentos relativos ao processo de inferência são muito semelhantes àqueles apresentados, considerando-se a distribuição de Bernoulli.
- Por simplicidade, vamos apresentar os resultados somente para o modelo de regressão logístico: $p_i = \frac{e^{\mathbf{X}'_i \boldsymbol{\beta}}}{1 + e^{\mathbf{X}'_i \boldsymbol{\beta}}}$.
- Vetor escore

$$\mathbf{S}(\boldsymbol{\beta}) = \mathbf{X}'(\mathbf{y} - \boldsymbol{\mu}),$$

em que \mathbf{X} é matriz de planejamento, $\mathbf{y} = (y_1, \dots, y_k)'$ e

$$\boldsymbol{\mu} = (p_1, \dots, p_k)'$$

Inferência para o modelo

- Por outro lado, a informação de Fisher é dada por

$$I(\beta) = \mathbf{X}' \mathbf{V} \mathbf{X},$$

em que $\mathbf{V} = \text{diag}(m_1 p_1 (1 - p_1), \dots, m_k p_k (1 - p_k))$

- Novamente, o sistema de equações $\mathbf{S}(\hat{\beta}) = \mathbf{0}$ não tem solução explícita e algum método de otimização numérica, como o algoritmo escore de Fisher, deve ser utilizado para obter-se as estimativas de MV.

Algoritmo escore de Fisher

- Seja $\beta^{(0)}$ uma estimativa inicial de β (chute inicial), então faça

$$\beta^{(t+1)} = \beta^{(t)} + \mathbf{I}^{-1}(\beta^{(t)})\mathbf{S}(\beta^{(t)}), t = 1, 2, \dots \quad (6)$$

até que algum critério de convergência seja satisfeito, como

$$|l(\beta^{(t+1)}) - l(\beta^{(t)})| < \epsilon, \epsilon > 0,$$

em que $l(\cdot)$ é a logverossimilhança (equação (5)).

- A equação (6) pode ser reescrita como

$$\beta^{(t+1)} = \left(\mathbf{X}' \mathbf{V}^{(t)} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{V}^{(t)} \mathbf{z}^{(t)},$$

em que $\mathbf{z} = \boldsymbol{\eta} + \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu})$.

Comentários

- Os resultados anteriores continuam válidos, com pequenas modificações.
- $\mathcal{V}(Y_i) = m_i p_i (1 - p_i)$.
- $I(\beta) = \mathbf{X}' \mathbf{V} \mathbf{X}$, $\mathbf{V} = \text{diag}(m_1 p_1 (1 - p_1), \dots, m_k p_k (1 - p_k))$.

Comentários

- Desvio:

$$\begin{aligned} D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = & 2 \sum_{i=1}^k \left\{ [y_i \ln[y_i / (m_i \hat{p}_i)] + (m_i - y_i) \ln [(1 - y_i / m_i) / (1 - \hat{p}_i)]] \right. \\ & \times \mathbb{1}_{\{1, \dots, (m_i - 1)\}}(y_i) \\ & \left. - 2[m_i \ln(1 - \hat{p}_i)] I_{\{0\}}(y_i) - 2[m_i \ln \hat{p}_i] I_{\{m_i\}}(y_i) \right\}. \end{aligned} \quad (7)$$

Nesse caso, para k fixado e $m_i \rightarrow \infty, i = 1, 2, \dots, k$, sob a hipótese de que o modelo é adequado, $D(\mathbf{y}, \hat{\boldsymbol{\mu}}) \approx \chi^2_{(k-p)}$.

Comentários

- Resíduo componente do desvio (RCD). Nesse caso, é dado por

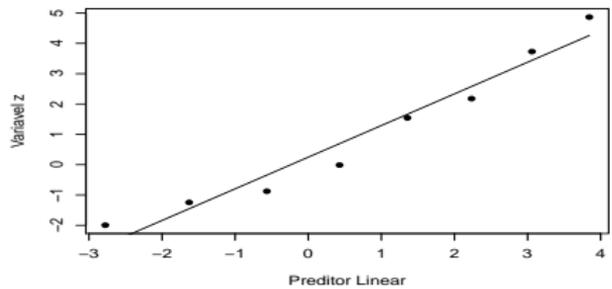
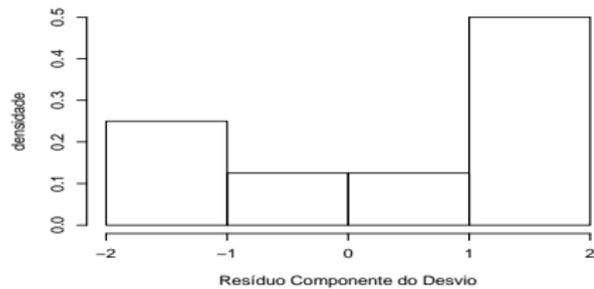
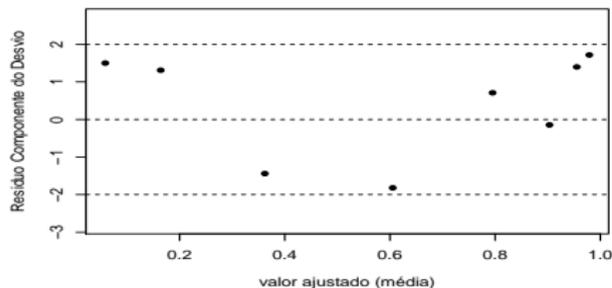
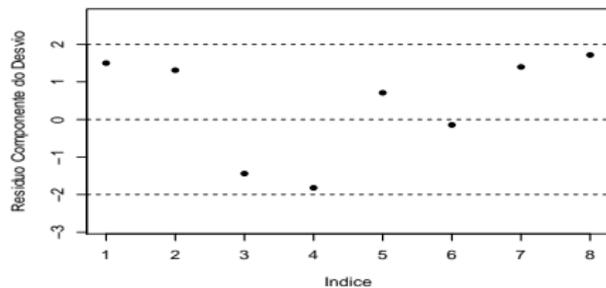
$$\begin{aligned} T_{D_i} &= -\frac{(2m_i |\ln(1 - \hat{p}_i)|)^{1/2}}{\sqrt{1 - \hat{h}_{ii}}} I_{\{0\}}(y_i) + \frac{(2m_i |\ln \hat{p}_i|)^{1/2}}{\sqrt{1 - \hat{h}_{ii}}} I_{\{m_i\}}(y_i) \\ &+ \pm \sqrt{\frac{2}{1 - \hat{h}_{ii}}} \left\{ y_i \ln \left(\frac{y_i}{m_i \hat{p}_i} \right) \right. \\ &+ \left. (m_i - y_i) \ln \left(\frac{m_i - y_i}{m_i - m_i \hat{p}_i} \right) \right\}^{1/2} \mathbb{1}_{\{1, \dots, m_i - 1\}}(y_i). \end{aligned}$$

em que \pm assume o mesmo sinal de $y_i - m_i \hat{p}_i$,

$$\hat{h}_{ii} = m_i \hat{p}_i (1 - \hat{p}_i) \mathbf{X}'_i \left(\mathbf{X}' \hat{\mathbf{V}} \mathbf{X} \right)^{-1} \mathbf{X}_i, \text{ e}$$

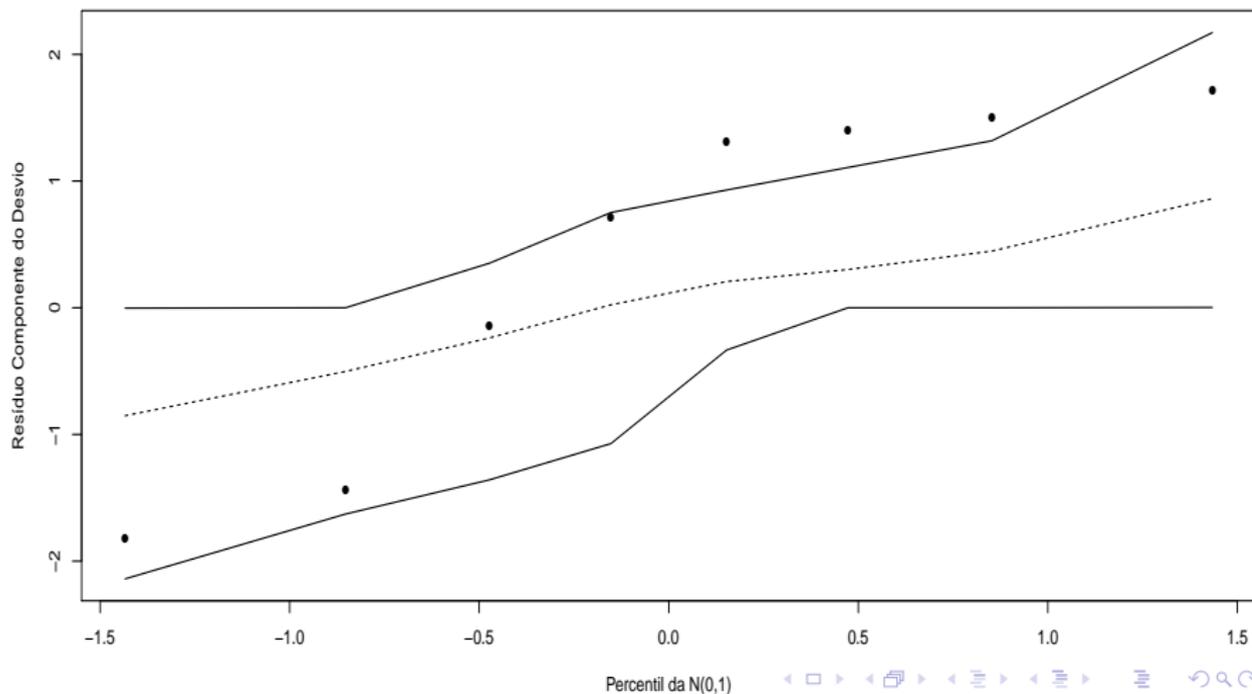
$$\hat{\mathbf{V}} = \text{diag}(m_1 \hat{p}_1 (1 - \hat{p}_1), \dots, m_k \hat{p}_k (1 - \hat{p}_k)).$$

Gráficos de diagnóstico para os RCD's



Gráficos de envelopes para os RCD's

Gráfico de quantil-quantil normal



Estimativas dos parâmetros (os testes se referem à nulidade de cada parâmetro)

Parâmetros	Estimativa	EP	IC(95%)	Estat. Z_t	p-valor
β_0	-60,72	5,18	[0,47 ; 1,01]	-11,72	<0,0001
β_1	34,27	2,91	[28,56 ; 39,98]	11,77	< 0,0001

Todos os parâmetros são significativos. Além disso, $D(\mathbf{y}; \tilde{\boldsymbol{\mu}}) = 11,23$, para $k - p = 8 - 2 = 6$ graus de liberdade, o que leva à um p-valor = 0,08145, o que sugere um ajuste apenas razoável.

Estimativas das proporções de insetos mortos

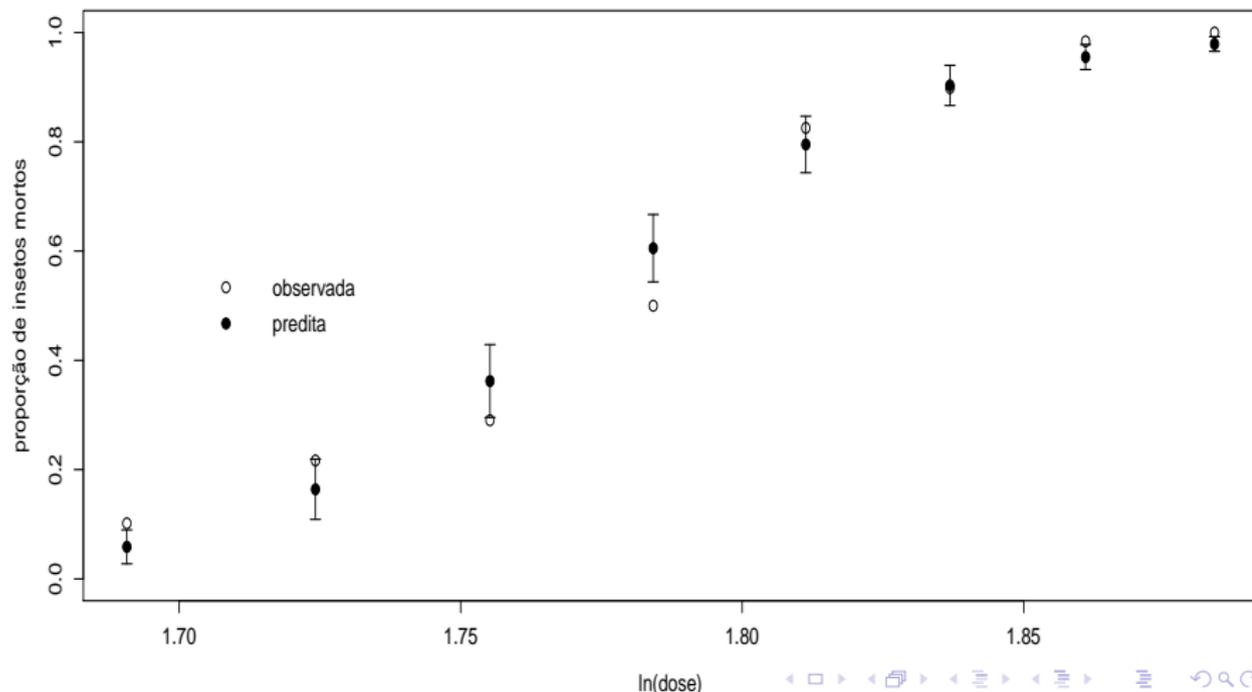
- A proporção de insetos mortos submetidos à dose x_i predita pelo modelo é dada por $\hat{p}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}$.
- Pelo método delta, para $m_i, i = 1, 2, \dots, 8$, suficientemente grandes, temos que $\hat{p}_i \approx N(p_i, \Psi_i \Sigma_\beta \Psi_i')$, em que

$$\Psi_i = \begin{bmatrix} \frac{\partial}{\partial \beta_0} p_i & \frac{\partial}{\partial \beta_1} p_i \end{bmatrix}$$

e $p_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$. Pode-se provar que $\frac{\partial}{\partial \beta_0} p_i = p_i(1 - p_i)$ e $\frac{\partial}{\partial \beta_1} p_i = p_i(1 - p_i)x_i$ (**exercício**).

- Assim $IC(p_i, \gamma) = \left[\hat{p}_i - z_{(1+\gamma)/2} \sqrt{\hat{\psi}_i}; \hat{p}_i + z_{(1+\gamma)/2} \sqrt{\hat{\psi}_i} \right]$, em que $P(Z \geq z_{(1+\gamma)/2}) = \frac{1+\gamma}{2}$, $\hat{\psi}_i = \hat{\Psi}_i \hat{\Sigma}_\beta \hat{\Psi}_i'$ e $Z \sim N(0, 1)$ (lembrando que esse IC é assintótico).

Proporções observadas \times proporções previstas pelo modelo



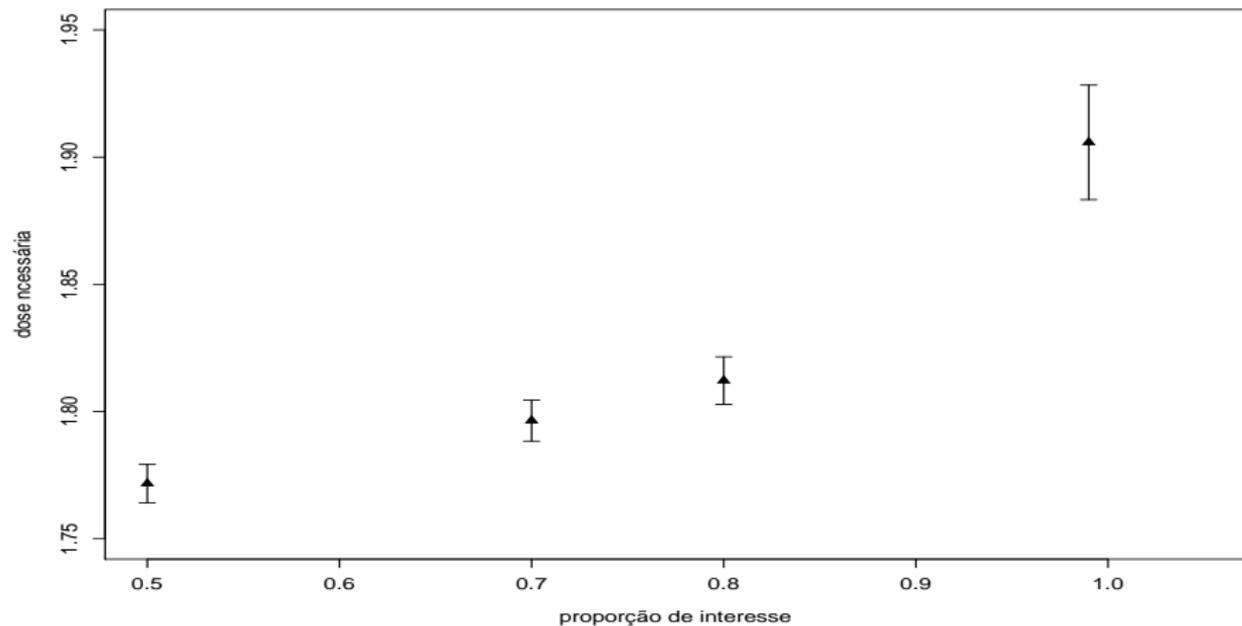
Comentários

- A análise de diagnóstico indicou que o modelo não se ajustou bem aos dados, portanto ele não pode ser utilizado para analisar os dados.
- Isso ocorreu, possivelmente, devido à função de ligação.
- Alternativas de análise: utilizar o mesmo modelo com outra função de ligação, p.e., baseada na distribuição normal assimétrica ou t assimétrica.

Estimativas da dose letal

- Sabemos que a probabilidade (proporção) estimada de insetos mortos para uma certa dose (x) é dada por $\tilde{\mu} = F(\tilde{\beta}_0 + \tilde{\beta}_1 x)$.
- A notação usual para uma dose letal de 100p% é DL_{100p} . Dessa forma a proporção de sucessos para essa dose é dada por $\tilde{p} = F(\tilde{\beta}_0 + \tilde{\beta}_1 DL_{100p})$.
- Portanto o estimador para DL_{100p} é dado por $\widehat{DL}_{100p} = \frac{1}{\widehat{\beta}_1} [F^{-1}(p) - \widehat{\beta}_0]$. No caso do modelo de regressão logística temos que: $\widehat{DL}_{100p} = \frac{1}{\widehat{\beta}_2} \left[\ln \left(\frac{p}{1-p} \right) - \widehat{\beta}_1 \right]$.
- Exercício: obter a distribuição assintótica do emv da dose letal.

Estimativas da dose letal



Voltando ao Exemplo 7: efeitos de certos fatores na sobrevivência de recém-nascidos

idade	N. de cigarros	Sobrevivência		
		Não	Sim	Total
<30	< 5	$74(\theta_{(1)11})$	$4327(\theta_{(1)12})$	4401
	5+	$15(\theta_{(1)21})$	$499(\theta_{(1)22})$	514
30+	< 5	$55(\theta_{(2)11})$	$1741(\theta_{(2)12})$	1796
	5+	$5(\theta_{(2)21})$	$135(\theta_{(2)22})$	140

Cada linha corresponde à uma distribuição binomial.

Proporções amostrais

idade	N. de cigarros	Sobrevivência	
		Não	Sim
<30	< 5	0,02	0,98
	5+	0,03	0,97
30+	< 5	0,03	0,97
	5+	0,04	0,97

Modelo

- Modelo

$$Y_{(i)j1} \stackrel{ind.}{\sim} \text{binomial}(m_{(i)j}, \theta_{(i)j1})$$
$$\ln \left(\frac{\theta_{(i)j1}}{1 - \theta_{(i)j1}} \right) = \mu_i + \alpha_{(i)j}, i = 1, 2, j = 1, 2, \alpha_{(i)1} = 0, i = 1, 2.$$

- $m_{(i)j}$: número total de recém nascidos de mães que fumam uma quantidade j de cigarros por dia e que pertencem ao grupo i da idade.
- $Y_{(i)j1}$: número de recém nascidos que vieram à óbito, de mães que fumam uma quantidade j de cigarros por dia e que pertencem ao grupo i da idade.
- $\beta = (\mu_1, \mu_2, \alpha_{(1)2}, \alpha_{(2)2})'$. Note que este modelo é saturado ($n = p$).

Logitos, parâmetros e quantidades observadas

$$\ln \left(\frac{\theta_{(1)11}}{1 - \theta_{(1)11}} \right) = \mu_1 \Rightarrow \theta_{(1)11} = \frac{e^{\mu_1}}{1 + e^{\mu_1}}$$

$$\ln \left(\frac{\theta_{(1)21}}{1 - \theta_{(1)21}} \right) = \mu_1 + \alpha_{(1)2} \Rightarrow \theta_{(1)21} = \frac{e^{\mu_1 + \alpha_{(1)2}}}{1 + e^{\mu_1 + \alpha_{(1)2}}}$$

$$\ln \left(\frac{\theta_{(2)11}}{1 - \theta_{(2)11}} \right) = \mu_2 \Rightarrow \theta_{(2)11} = \frac{e^{\mu_2}}{1 + e^{\mu_2}}$$

$$\ln \left(\frac{\theta_{(2)21}}{1 - \theta_{(2)21}} \right) = \mu_2 + \alpha_{(2)2} \Rightarrow \theta_{(2)21} = \frac{e^{\mu_2 + \alpha_{(2)2}}}{1 + e^{\mu_2 + \alpha_{(2)2}}}$$

Além disso, $m_{(1)1} = 4401$, $m_{(1)2} = 514$, $m_{(2)1} = 1796$, $m_{(2)2} = 140$ e $y_{(1)11} = 74$, $y_{(1)21} = 15$, $y_{(2)11} = 55$, $y_{(2)21} = 5$. **Exercício: escrever as razões de chance de interesse em função dos parâmetros β , interpretando os parâmetros $(\alpha_{(1)2}, \alpha_{(2)2})'$.**

Hipóteses de interesse

- Ausência de independência, entre sobrevida e n. de cigarros, para cada uma das subpopulações (idade).

$$H_0 : \begin{cases} \theta_{(1)11} = \theta_{(1)21} \\ \theta_{(2)11} = \theta_{(2)21} \end{cases} \leftrightarrow \begin{cases} \alpha_{(1)2} = 0 \\ \alpha_{(2)2} = 0 \end{cases}$$

vs H_1 : há pelo menos uma diferença

- Como testar as hipóteses acima? Através de testes individuais de nulidade, testes do tipo $\mathbf{C}\beta = \mathbf{M}$, teste da razão de verossimilhanças, análise do desvio (quando os modelos envolvidos são não saturados).

Testes da Razão de verossimilhanças

- Vamos supor para o vetor de parâmetros β a partição $\beta = (\beta_1', \beta_2')'$, em que β_1 e β_2 são vetores de dimensão $q \times 1$ e $(p - q) \times 1$, respectivamente.
- Desejamos testar $H_0 : \beta_1 = \mathbf{0}$ vs $H_1 : \beta_1 \neq \mathbf{0}$.
- Sejam $D(\mathbf{y}; \hat{\mu}^{(0)})$ e $D(\mathbf{y}; \hat{\mu})$ os desvios dos modelos: ajustados sob H_0 e irrestrito, respectivamente. Lembremos que $D(\mathbf{y}; \hat{\mu}) = 2 \{l(\mathbf{Y}, \mathbf{y}) - l(\hat{\mu}, \mathbf{y})\}$.

Testes da Razão de verossimilhanças

- Defina a seguinte estatística:

$$Q_{RV} = D(\mathbf{y}; \hat{\boldsymbol{\mu}}^{(0)}) - D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = -2(l(\hat{\boldsymbol{\mu}}, \mathbf{y}) - l(\hat{\boldsymbol{\mu}}^{(0)}, \mathbf{y})) \text{ (exercício).}$$

Sob H_0 , para n suficientemente grande, $Q_{RV} \approx \chi_q^2$

- Assim, rejeita-se H_0 se p -valor $\leq \alpha$, em que

p -valor $\approx P(X \geq q_{RV} | H_0)$, em que $X \sim \chi_q^2$

$$q_{RV} = D(\mathbf{y}; \tilde{\boldsymbol{\mu}}^{(0)}) - D(\mathbf{y}; \tilde{\boldsymbol{\mu}}).$$

Análise do desvio

- Baseados no teste da RV, podemos ainda definir um outro procedimento para testar as hipóteses $H_0 : \beta_1 = \mathbf{0}$ vs $H_1 : \beta_1 \neq \mathbf{0}$.

- A estatística $Q_{AD} = \frac{(D(\mathbf{y}; \hat{\boldsymbol{\mu}}^{(0)}) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})) / q}{D(\mathbf{y}; \hat{\boldsymbol{\mu}}) / (n - p)}$ sob H_0 e para n suficientemente grande, é tal que $Q_{AD} \approx F_{(q, n-p)}$

- Note que só podemos utilizar esta abordagem para modelos não saturados ($n > p$).

- Assim, rejeita-se H_0 se $p - \text{valor} \leq \alpha$, em que

$p - \text{valor} \approx P(X \geq q_{AD} | H_0)$, em que $X \sim F_{(q, n-p)}$

$$q_{AD} = \frac{(D(\mathbf{y}; \tilde{\boldsymbol{\mu}}^{(0)}) - D(\mathbf{y}; \tilde{\boldsymbol{\mu}})) / q}{D(\mathbf{y}; \tilde{\boldsymbol{\mu}}) / (n - p)}.$$

Ajuste do modelo

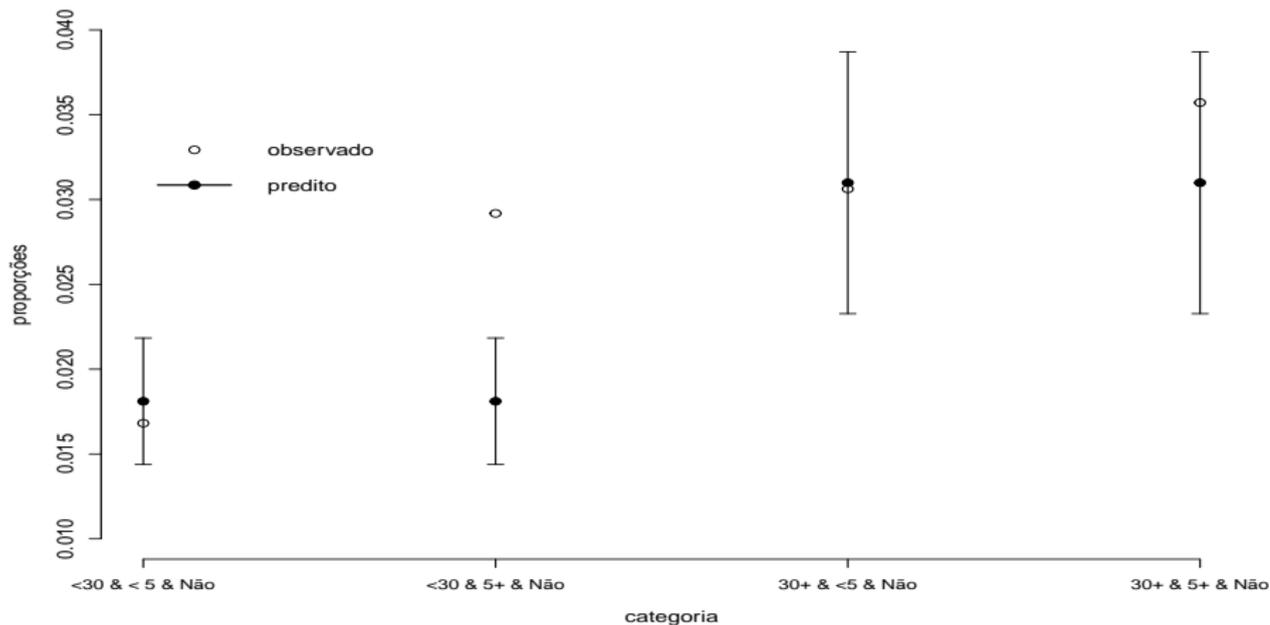
Parâmetro	Estimativa	EP	Estat. Z_t	p-valor
μ_1	-4,069	0,117	-34,70	$< 0,0001$
$\alpha_{(1)2}$	0,564	0,2871	1,96	0,0495
μ_2	-3,455	0,137	-25,23	$< 0,0001$
$\alpha_{(2)2}$	0,159	0,4756	0,33	0,7381

Há uma significância marginal da hipótese alternativa. Temos poucas observações para realizarmos uma análise de resíduos. Alternativa, considerar a decomposição das observações binomiais nas contagens 0 e 1.

Outros testes

- Teste $C\beta = M$, $q_t = 3,97$, $p - \text{valor} = 0,1373$. **Exercício:** encontrar as matrizes C e M .
- Teste da RV, $q_{RV} = 3,53$, $p - \text{valor} = 0,1714$.
- Assim, não rejeitamos a hipótese de independência simultânea.
- Ajustou-se, então, um modelo reduzido (em que $\alpha_{(1)2} = \alpha_{(2)2} = 0$).

Proporções observadas e previstas pelo modelo reduzido



Exemplo 13: preferência de consumidores com relação à marcas de carros

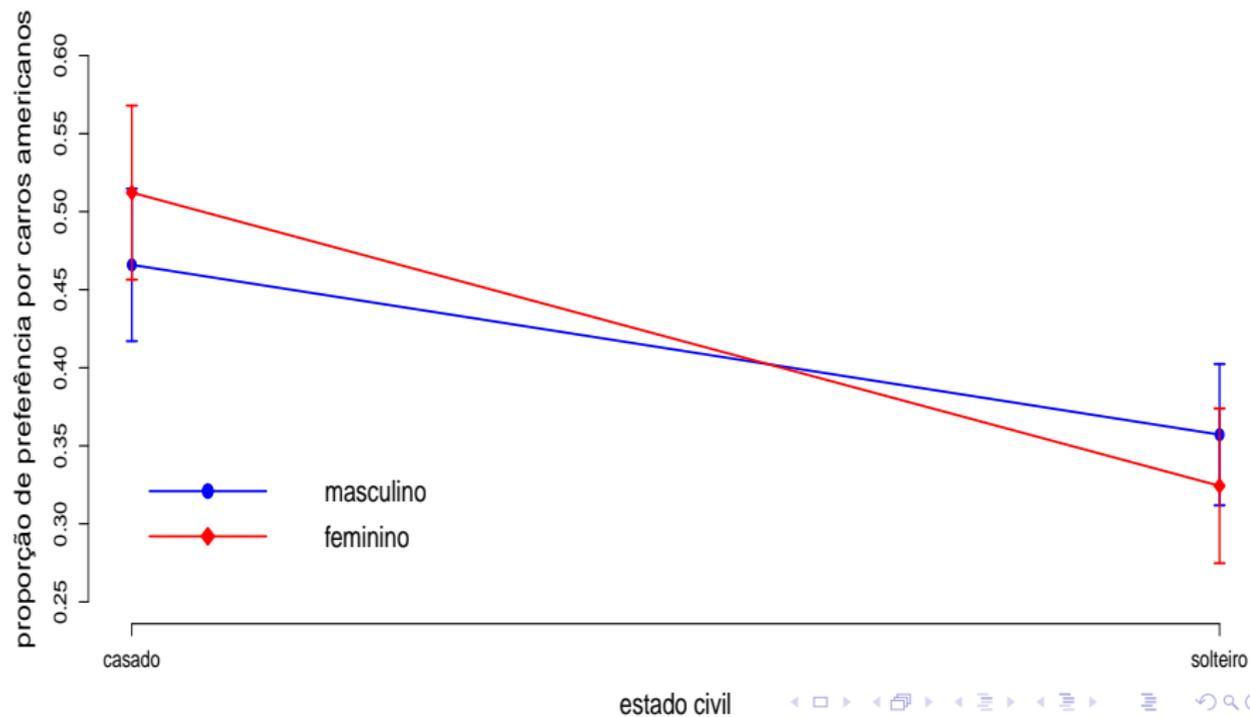
- Uma amostra aleatória de 263 consumidores foi considerada.
- As seguintes variáveis foram observadas para cada comprador: preferência do tipo de automóvel (1: americano, 0: japonês), idade (em anos), sexo (0: masculino; 1: feminino) e estado civil (0: casado, 1: solteiro).
- Variável resposta: preferência do tipo de automóvel.
- Para maiores detalhes ver Foster, Stine e Waterman (1998, pgs. 338-339).

Análise descritiva

Os percentuais foram calculados dentro de cada categoria de gênero e estado civil (os percentuais dentro de cada linha somam 100%).

	preferência	
gênero	japonês	americano
masculino	57,64	42,36
feminino	54,62	45,38
estado civil		
casado	51,18	48,82
solteiro	65,59	34,41

Gráficos de perfis



Modelo

■ Modelo 1

$$Y_{ijk} \stackrel{ind.}{\sim} \text{Bernoulli}(\theta_{ij})$$
$$\ln \left(\frac{\theta_{ij}}{1 - \theta_{ij}} \right) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}, i = 1, 2, j = 1, 2, k = 1, 2, \dots, n_{ij}$$
$$\alpha_1 = \beta_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = 0, \forall i, j.$$

- n_{ij} : número total de consumidores pertencentes ao i -ésimo gênero (1: masculino, 2: feminino) e ao j -ésimo estado civil (1: casado, 2: solteiro), $n_{11} = 88, n_{12} = 56, n_{21} = 82, n_{22} = 37$.
- Y_{ijk} : 1 se o k -ésimo consumidor pertencente ao i -ésimo gênero e ao j -ésimo estado civil prefere carros americanos e 0, caso ele prefira carros japoneses.

Modelo

- $\beta = (\mu, \alpha_2, \beta_2, (\alpha\beta)_{22})'$.

- Logitos

$$\ln\left(\frac{\theta_{11}}{1 - \theta_{11}}\right) = \mu \Rightarrow \theta_{11} = \frac{e^\mu}{1 + e^\mu}$$

$$\ln\left(\frac{\theta_{21}}{1 - \theta_{21}}\right) = \mu + \alpha_2 \Rightarrow \theta_{21} = \frac{e^{\mu + \alpha_2}}{1 + e^{\mu + \alpha_2}}$$

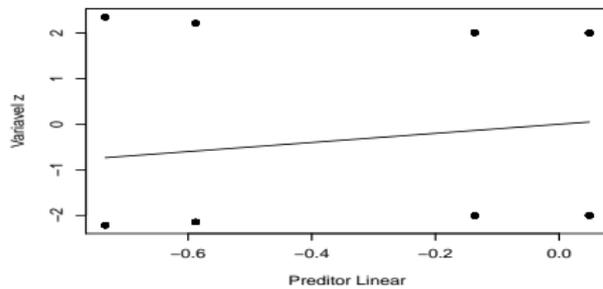
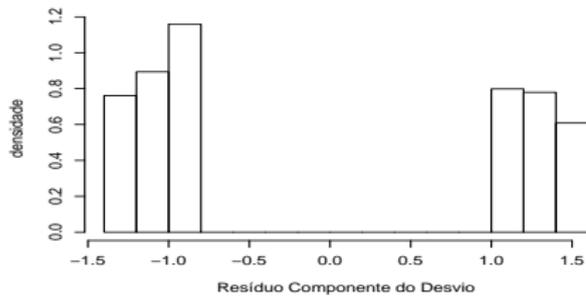
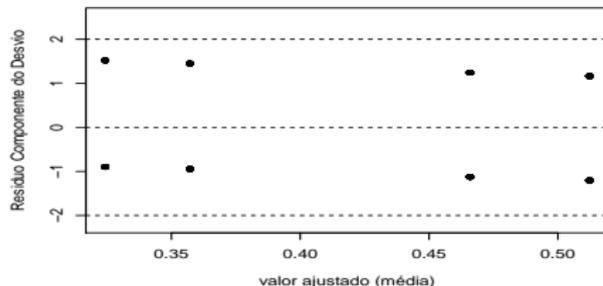
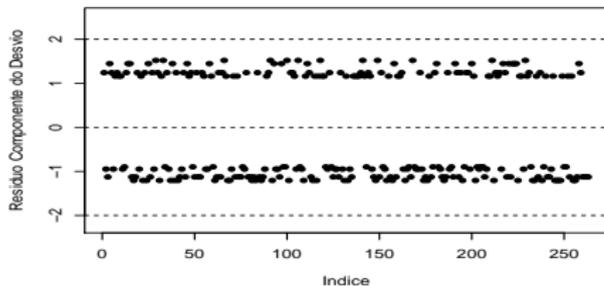
$$\ln\left(\frac{\theta_{12}}{1 - \theta_{12}}\right) = \mu + \beta_2 \Rightarrow \theta_{12} = \frac{e^{\mu + \beta_2}}{1 + e^{\mu + \beta_2}}$$

$$\ln\left(\frac{\theta_{22}}{1 - \theta_{22}}\right) = \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22} \Rightarrow \theta_{22} = \frac{e^{\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}}}{1 + e^{\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}}}$$

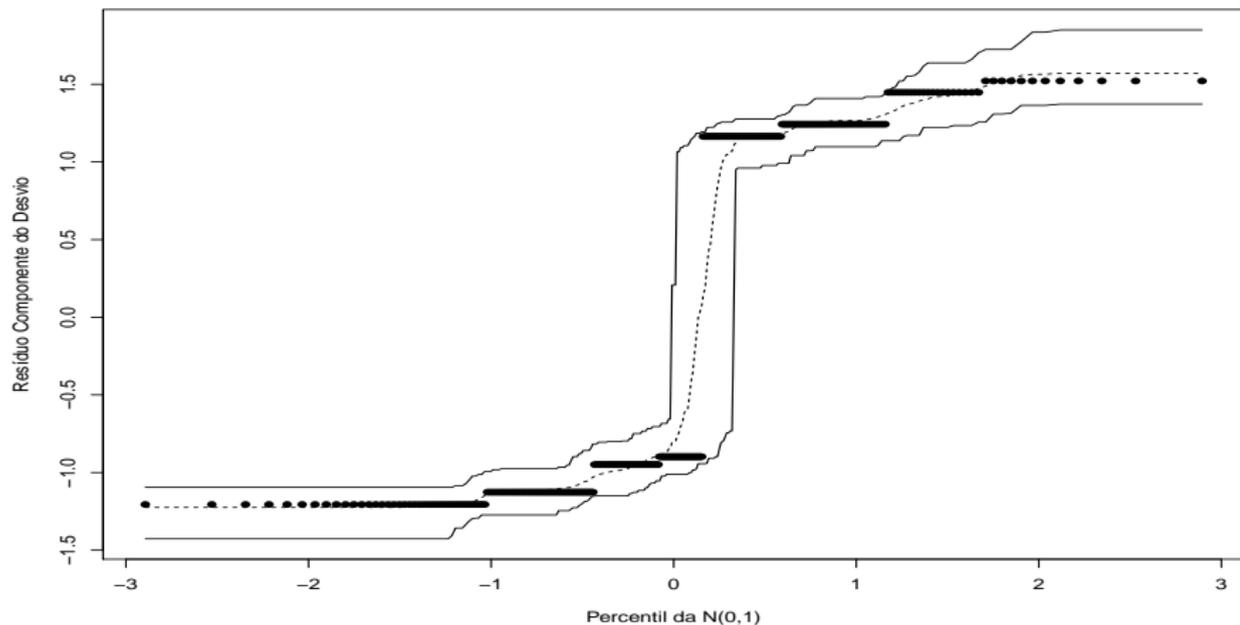
Modelo

- Os parâmetros seguem as interpretações usuais, mas agora em termos das probabilidades e das razões de chances.
- **Exercício: provar que o parâmetro $(\alpha\beta)_{22}$ está relacionado com a presença de interação entre os fatores.**
- **Exercício: interprete os parâmetros $(\alpha_2, \beta_2)'$ em termos de razões de chances, dado a presença de interação.**
- **Exercício: provar que os parâmetros $(\alpha_2, \beta_2)'$ estão relacionados com a presença dos efeitos dos seus respectivos fatores, dado a ausência de interação.**

Gráficos de diagnóstico para o RCD: modelo 1



Gráficos de envelopes para o RCD: modelo 1

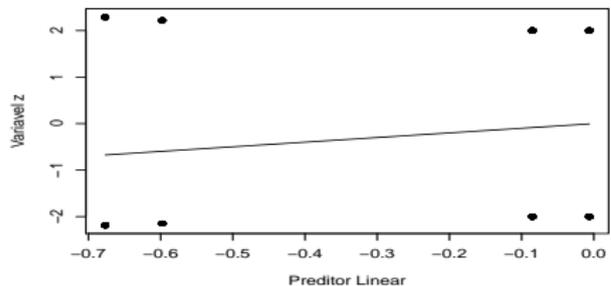
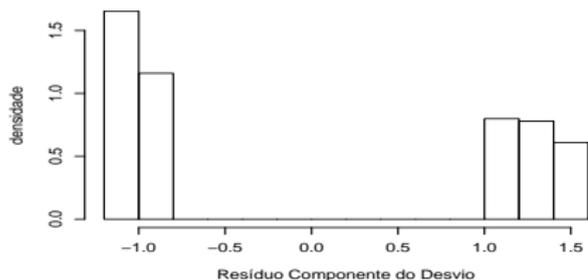
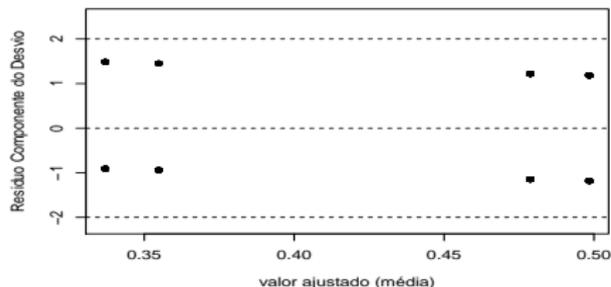
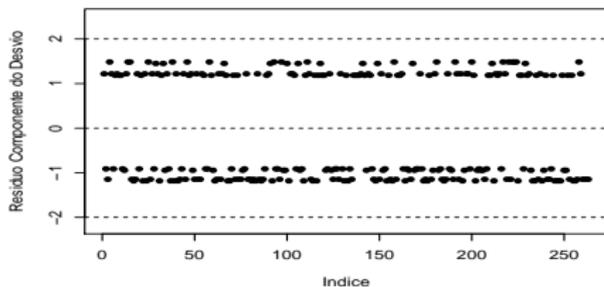


Ajuste do modelo completo

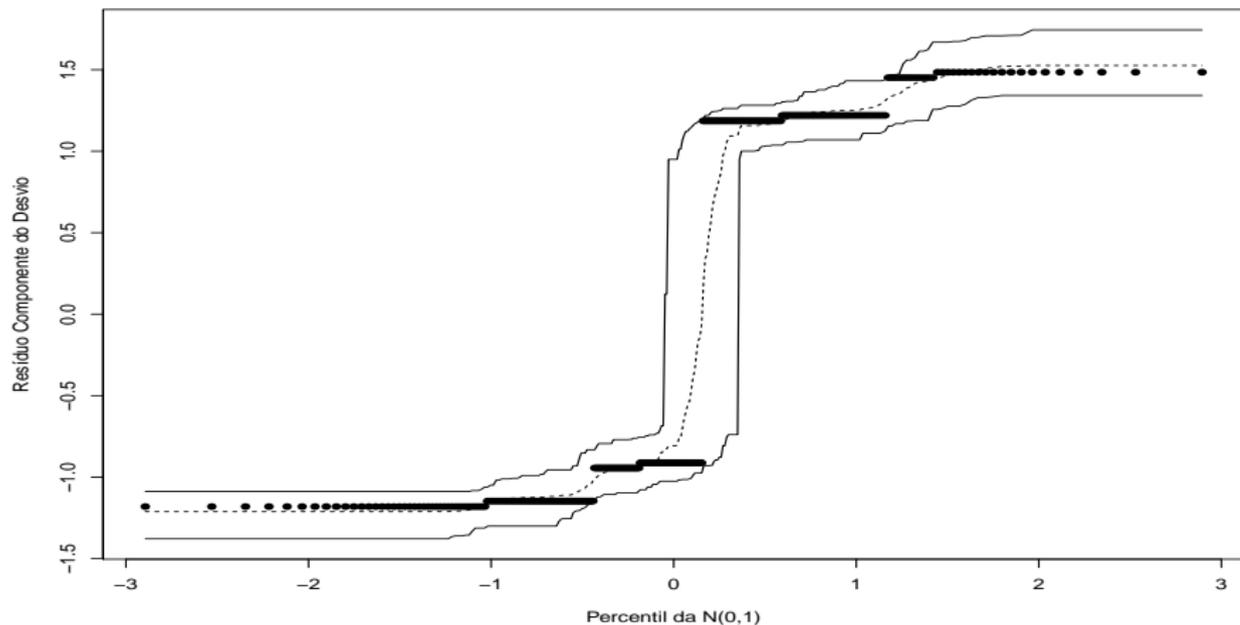
Parâmetro	Estimativa	EP	IC(95%)	Estat. Z_t	p-valor
μ	-0,137	0,214	[-0,56 ; 0,28]	-0,639	0,5228
α_2	0,185	0,307	[-0,42 ; 0,79]	0,603	0,5465
β_2	-0,451	0,351	[-1,14 ; 0,24]	-1,284	0,1991
$(\alpha\beta)_{22}$	-0,332	0,5437	[-1,40 ; 0,73]	-0,610	0,5420

Aparentemente, nenhum coeficiente é significativo. Entretanto, vamos explorar o modelo um pouco melhor retirando, num primeiro momento, a interação (modelo 2).

Gráficos de diagnóstico para o RCD: modelo 2



Gráficos de envelopes para o RCD: modelo 2

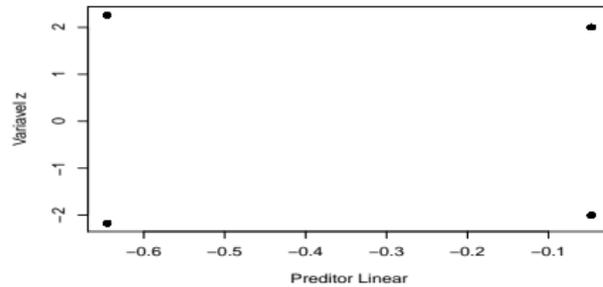
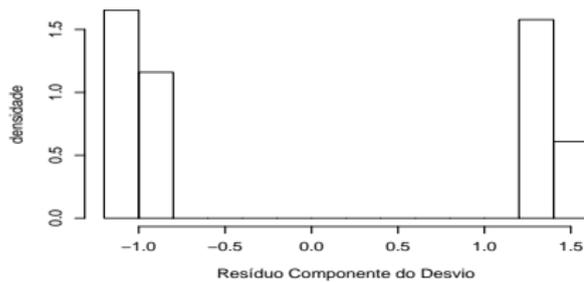
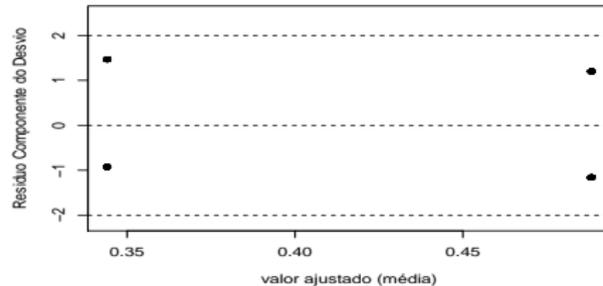
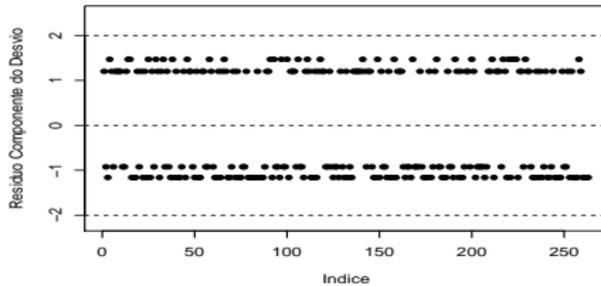


Ajuste do modelo 2

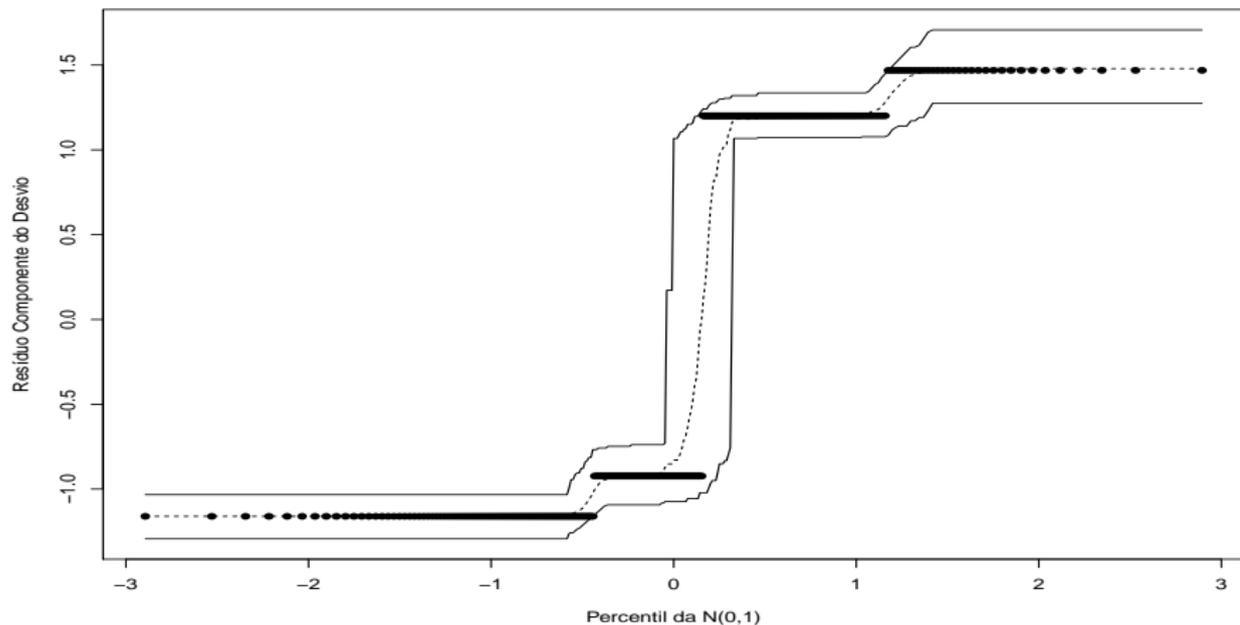
Parâmetro	Estimativa	EP	IC(95%)	Estat. Z_t	p-valor
μ	-0,085	0,196	[-0,47 ; 0,30]	-0,434	0,6642
α_2	0,079	0,253	[-0,42 ; 0,57]	0,312	0,7551
β_2	-0,592	0,268	[-1,12 ; -0,07]	-2,211	0,0270

O fator gênero parece ser não significativo enquanto que o fator estado civil parecer ser significativo. Vamos ajustar um modelo sem gênero (modelo 3).

Gráficos de diagnóstico para o RCD: modelo 3



Gráficos de envelopes para o RCD: modelo 3



Ajuste do modelo 3

Parâmetro	Estimativa	EP	IC(95%)	Estat. Z_t	p-valor
μ	-0,0471	0,153	[-0,35 ; 0,25]	-0,307	0,7590
β_2	-0,5981	0,267	[-1,12 ; -0,08]	-2,242	0,0250

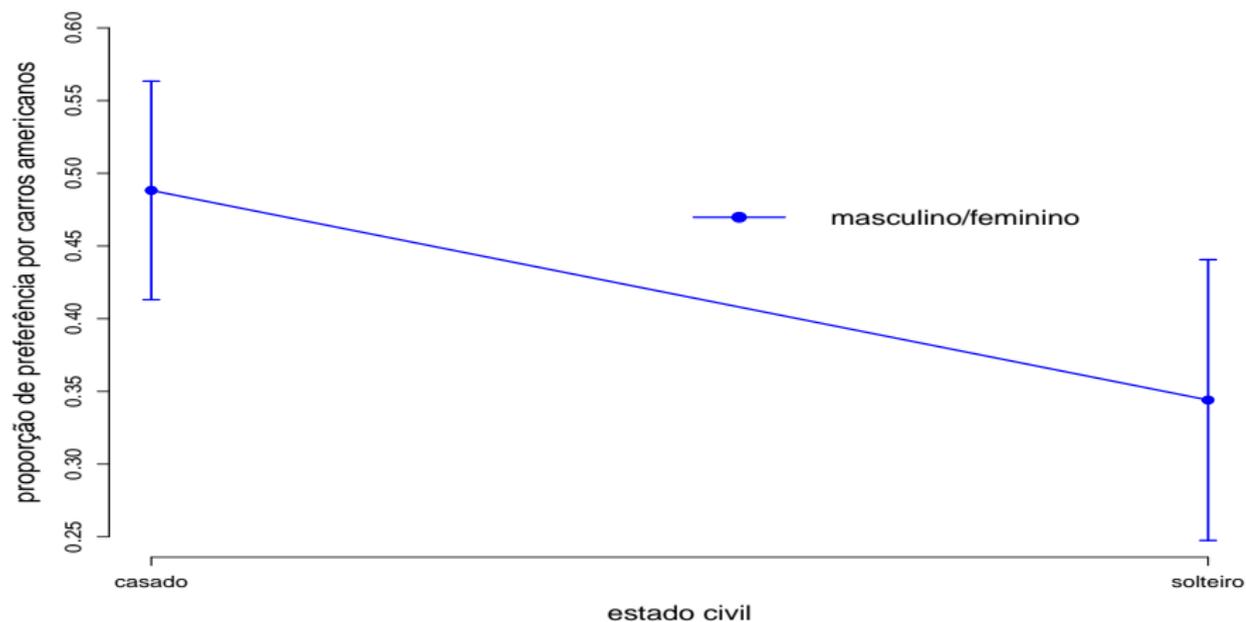
Modelo final: fator estado civil é significativo.

Percentuais preditos pelo modelo final (através do método delta)

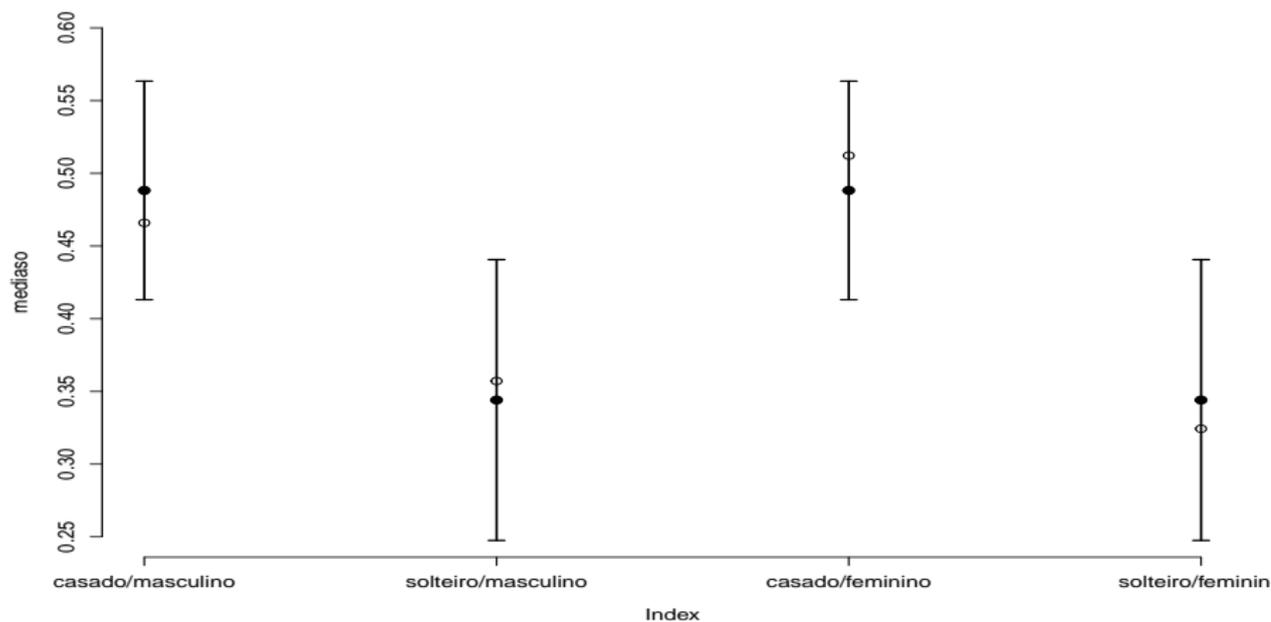
Estado civil	Gênero	Estimativa	EP	IC(95%)
Casado	Masculino	48,82	3,83	[41,31 ;56,34]
Solteiro	Masculino	34,41	4,93	[24,75 ; 44,06]
Casado	Feminino	48,82	3,83	[41,31 ; 56,34]
Solteiro	Feminino	34,41	4,93	[24,75 ; 44,06]

Exercício: obter os resultados acima aplicando o método delta.

Proporções previstas pelo modelo final



Proporções observadas e previstas pelo modelo final



Seleção de modelos

- Vimos como verificar se um determinado modelo se ajusta adequadamente aos dados.
- Uma outra questão de interesse surge quando se dispõe de diversos modelos (que se ajustam adequadamente aos dados) e respondem às perguntas de interesse, e queremos escolher um como o “mais apropriado” .
- Há diversas técnicas disponíveis para este fim.
- Veremos técnicas baseadas em testes de hipótese e comparação de estatísticas de qualidade de ajuste.

Teste da razão de verossimilhanças

- Sejam M_1 e M_2 dois modelos, em que M_1 está encaixado em M_2 , ou seja, o modelo M_1 é um caso particular de M_2 .
- Por exemplo, M_1 é um modelo linear obtido de M_2 , o qual é um modelo quadrático.
- Neste caso temos que

H_0 : **o modelo M_1 é preferível ao modelo M_2** vs H_1 : **o modelo M_2 é preferível ao modelo M_1 .**

Teste da razão de verossimilhanças (cont.)

- Seja $\hat{\theta}_i$ o estimador de máxima verossimilhança obtido sob o modelo i e $\tilde{\theta}_i$ sua respectiva estimativa.
- Denote por $L_i(\hat{\theta})$ e $l_i(\hat{\theta})$ o máximo da verossimilhança e da log-verossimilhança do modelo i , respectivamente, em relação aos estimadores enquanto que $L_i(\tilde{\theta})$ e $l_i(\tilde{\theta})$ são os respectivos máximos avaliados nas estimativas.

Teste da razão de verossimilhanças (cont.)

- A estatística do TRV é dada por $\Delta = \frac{L_1(\hat{\theta}_1)}{L_2(\hat{\theta}_2)}$.
- Rejeita-se H_0 se $\Delta \geq \delta_c$, em que δ_c é um valor crítico adequado.
- Alternativamente, rejeitamos H_0 se

$$\Lambda = -2\ln(\Delta) = -2 \left(l_1(\hat{\theta}_1) - l_2(\hat{\theta}_2) \right) \geq \lambda_c,$$

em que $P(Q \geq \lambda_c) = \alpha$, $Q \approx \chi^2_{(\gamma)}$ e

γ = número de parâmetros do modelo M_2 - número de parâmetros do modelo M_1 .

- Nesse caso, p -valor $\approx P(Q \geq \lambda | H_0)$, em que λ é o valor observado da estatística Λ e $Q \sim \chi^2_{\gamma}$. Assim, rejeita-se H_0 se p -valor $\leq \alpha$.

Estatísticas de comparação de modelos

- O TRV é apropriado na comparação somente de modelos encaixados (o modelo com menor número de parâmetros é um caso particular do modelo com maior número de parâmetros).
- Além disso, ele não leva em consideração (diretamente) o número de parâmetros do modelo (somente na distribuição da estatística).
- Existem várias alternativas, em termos de estatísticas para comparar modelos, que “penalizam” a verossimilhança em relação ao número de parâmetros, tamanho da amostra entre outros fatores.
- Veremos o AIC e o BIC.

Estatísticas de comparação de modelos (cont.)

- O AIC e BIC, para o i -ésimo modelo, são dados, respectivamente, por:

$$AIC_i = -2l_i(\tilde{\theta}_i) + 2k$$

$$BIC_i = -2l_i(\tilde{\theta}_i) + 2k \ln(n)$$

que $l_i(\tilde{\theta}_i)$ denota a log-verossimilhança do i -ésimo modelo avaliada em alguma estimativa (p.e. máxima verossimilhança), k é o número de parâmetros e n é o número de observações.

- Portanto, o modelo que apresentar os menores valores, será o modelo “melhor ajustado” aos dados.

Métodos de seleção “dinâmico” ou automatizados

- Existem métodos que selecionam modelos, fixados alguns critérios, de modo “dinâmico” (automatizado).
- Veremos os métodos “forward”, “backward” e “stepwise”.
- Tais métodos são particularmente úteis quando se dispões de muitas covariáveis e/ou muitos fatores.
- Sem perda de generalidade, vamos considerar um determinado modelo (normal linear, linear generalizado) tal que o preditor linear é dado por

$$\eta_{ij} = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_{ij}$$



Método “forward”

- Primeiramente, ajustamos um modelo com somente o intercepto, ou seja $\eta_{ij} = \beta_0$. Ajustamos então, para cada variável explicativa, um modelo

$$\eta_{ij} = \beta_0 + \beta_j x_{ij}, j = 1, 2, \dots, p - 1$$

- Testa-se $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$, $j=1,2,\dots,p-1$ (usando-se algum teste como o TRV, teste $C\beta$, ou alguma estatística de comparação de modelos). Seja P o menor nível descritivo entre os $p - 1$ testes. Se $P \leq P_E$ a variável correspondente entra no modelo (caso contrário, o processo é interrompido).

Métodos “forward” (cont.)

- Vamor supor que a variável X_1 foi escolhida. Então, no passo seguinte, ajustamos os modelos

$$\eta_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_j x_{ij}, j = 2, \dots, p - 1$$

- Testa-se $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$, $j=2, \dots, p-1$ (usando-se algum teste como TRV, teste $C\beta$, ou alguma estatística de comparação de modelos). Seja P o menor nível descritivo entre os $p - 2$ testes. Se $P \leq P_E$ a variável correspondente entra no modelo. Repetimos o procedimento até que ocorra $P > P_E$.

Método “backward”

- Primeiramente, ajustamos o seguinte modelo:

$$\eta_{ij} = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_{ij}$$

- Testa-se $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$, $j=1,2,\dots,p-1$ (usando-se algum teste como o TRV, teste $C\beta$, ou alguma estatística de comparação de modelos). Seja P o maior nível descritivo entre os $p - 1$ testes. Se $P > P_S$ a variável correspondente sai do modelo (caso contrário, o processo é interrompido).

Método “backward” (cont.)

- Vamos supor que X_1 tenha saído do modelo. Então ajustamos o seguinte modelo

$$\eta_{ij} = \beta_0 + \sum_{j=2}^{p-1} \beta_j x_{ij}$$

- Testa-se $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$, $j=2, \dots, p-1$ (usando-se algum teste como TRV, teste $C\beta$, ou alguma estatística de comparação de modelos). Seja P o maior nível descritivo entre os $p - 2$ testes. Se $P > P_S$ 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 $\eta_{ij} = \beta_0$. Após duas variáveis terem sido incluídas no modelo, verificamos se a primeira sai ou não do modelo.
- O processo continua até que nenhuma variável seja incluída ou retirada do modelo.
- Geralmente adotamos $0,15 \leq P_E, P_S \leq 0,25$. Outra possibilidade é usar $P_E = P_S = 0,20$.

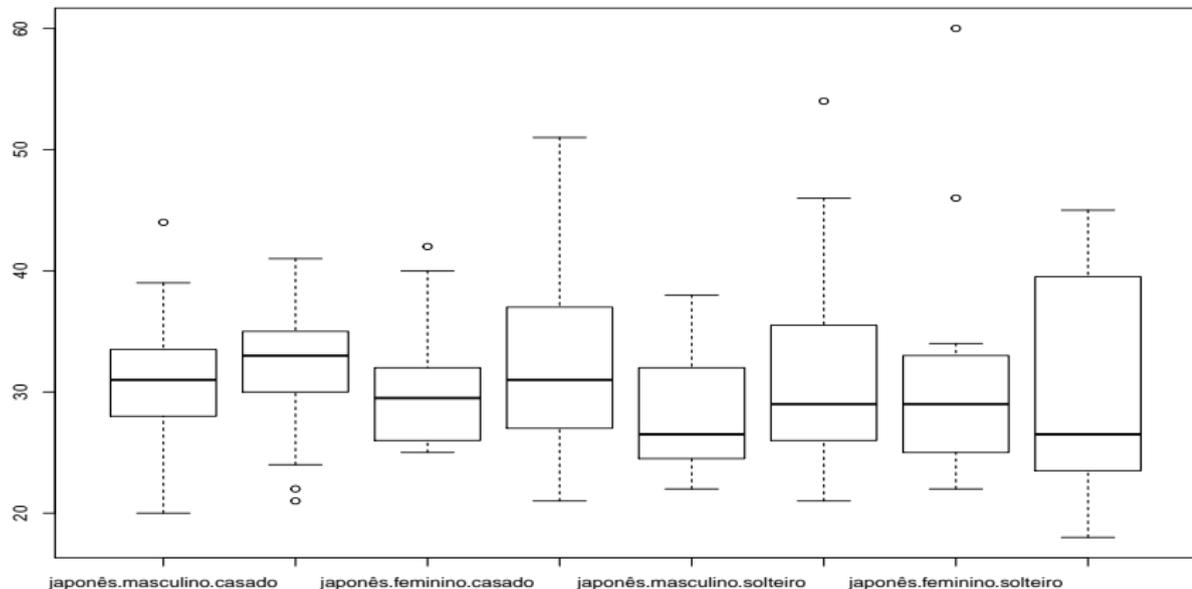
Métodos anteriores usando AIC/BIC

- Para qualquer um dos métodos anteriores, se usarmos alguma estatística de comparação de modelos (como AIC ou BIC), procedemos da seguinte forma
 - Sempre escolhemos o modelo (retirar/incluir a variável) que apresentar o menor valor da estatística.
 - O processo é interrompido quando as estatísticas para todos os modelos possíveis aumentarem em relação ao modelo corrente.
- Observação: as estatísticas AIC e BIC também servem para comparar modelos que difiram em termos da função de ligação e distribuição da variável resposta, entre outras características.

Aplicação no exemplo 13

- Aplicou-se cada um dos três métodos, forward, backward e stepwise, no exemplo anterior, através da estatística AIC.
- As três abordagens escolheram o modelo que contempla somente o intercepto e o fator estado civil.

Utilização da variável idade (boxplots)



Utilização da variável idade

■ Modelo 1:

$$\ln \left(\frac{\theta_{ij}}{1 - \theta_{ij}} \right) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma(x_{ijk} - \bar{x}),$$
$$i = 1, 2, j = 1, 2, k = 1, 2, \dots, n_{ij}$$

$$\alpha_1 = \beta_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = 0, \forall i, j.$$

em que x_{ijk} é a idade do k -ésimo indivíduo do gênero i e do estado civil j e $\bar{x} = \frac{1}{n} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} x_{ijk}$, $n = \sum_{i=1}^2 \sum_{j=1}^2 n_{ij}$. Para $x_{ijk} = \bar{x}$ e/ou para indivíduos com a mesma idade, os parâmetros $(\mu, \alpha_2, \beta_2, (\alpha\beta)_{22})'$ possuem a mesma interpretação anterior, enquanto que γ é o incremento no logito para o aumento em uma unidade da variável idade.

Utilização da variável idade

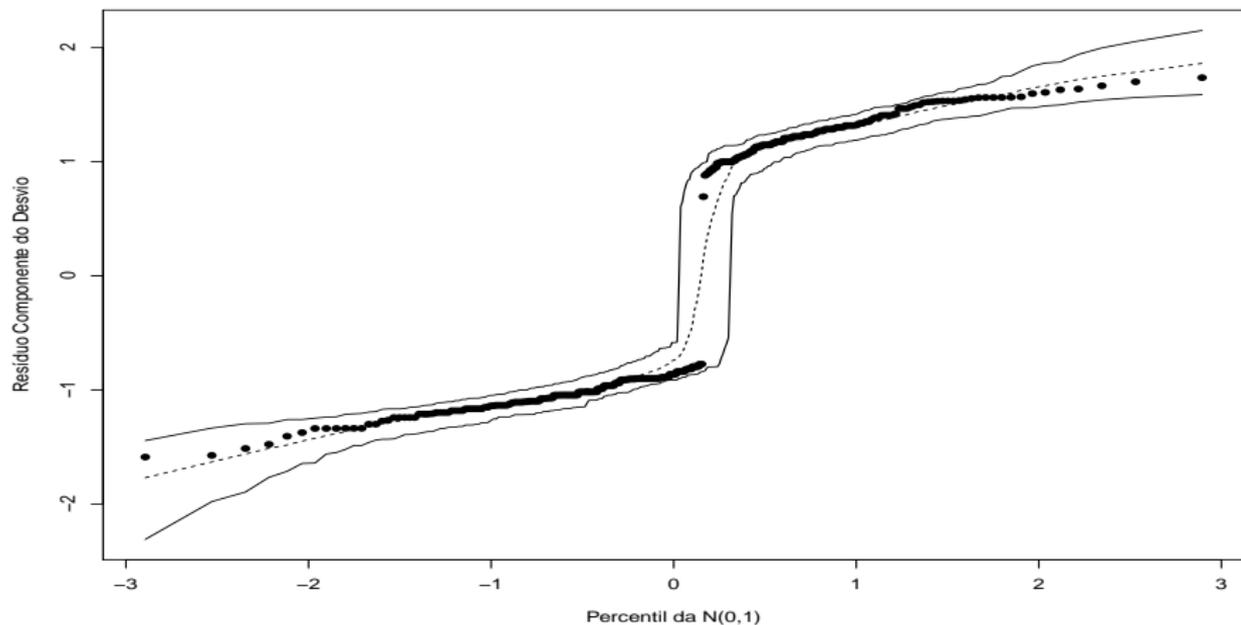
■ Modelo 2:

$$\ln\left(\frac{\theta_{ij}}{1-\theta_{ij}}\right) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{ij}(x_{ijk} - \bar{x}),$$
$$i = 1, 2, j = 1, 2, k = 1, 2, \dots, n_{ij}$$

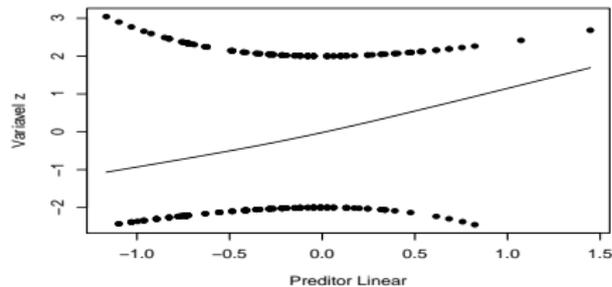
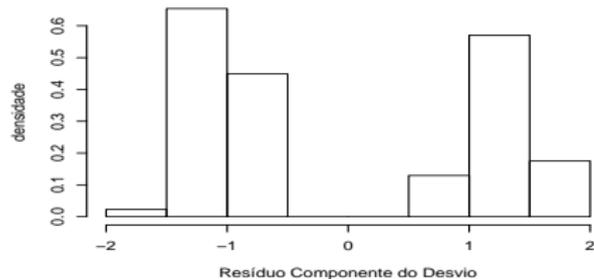
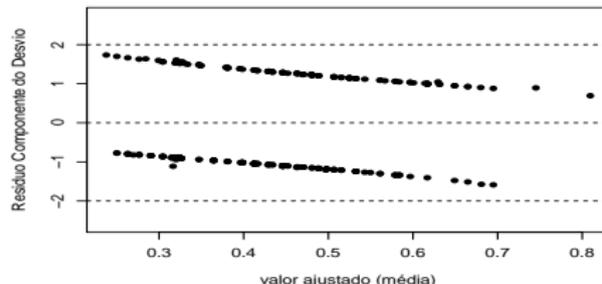
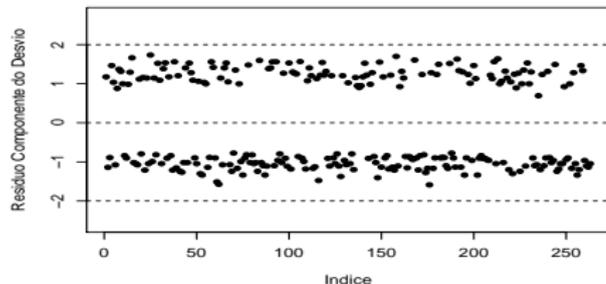
$$\alpha_1 = \beta_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = 0, \forall i, j.$$

Para $x_{ijk} = \bar{x}$ e/ou para indivíduos com a mesma idade e pertencentes ao mesmo grupo, os parâmetros $(\mu, \alpha_2, \beta_2, (\alpha\beta)_{22})'$ possuem a mesma interpretação anterior, enquanto que γ_{ij} continua sendo o incremento no logito para o aumento em uma unidade da variável idade, agora para cada grupo.

Gráficos de envelopes para os RCD's do modelo 2



Gráficos de diagnóstico para os RCD's do modelo 2



Modelos finais (usando o método stepwise)

- Modelo 1:

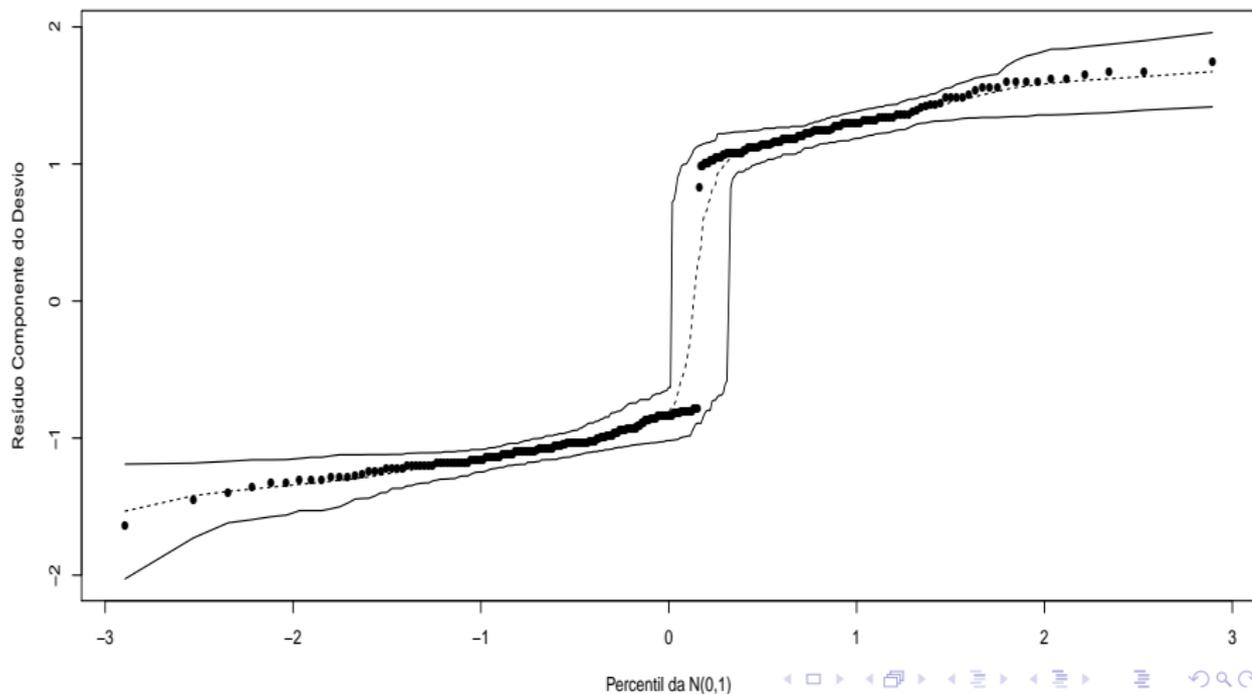
$$\ln \left(\frac{\theta_{ij}}{1 - \theta_{ij}} \right) = \mu + \beta_j + \gamma(x_{ijk} - \bar{x}),$$
$$i = 1, 2, j = 1, 2, k = 1, 2, \dots, n_{ij}$$
$$\beta_1 = 0.$$

- Modelo 2:

$$\ln \left(\frac{\theta_{ij}}{1 - \theta_{ij}} \right) = \mu + \beta_j$$
$$i = 1, 2, j = 1, 2, k = 1, 2, \dots, n_{ij}$$
$$\beta_1 = 0.$$

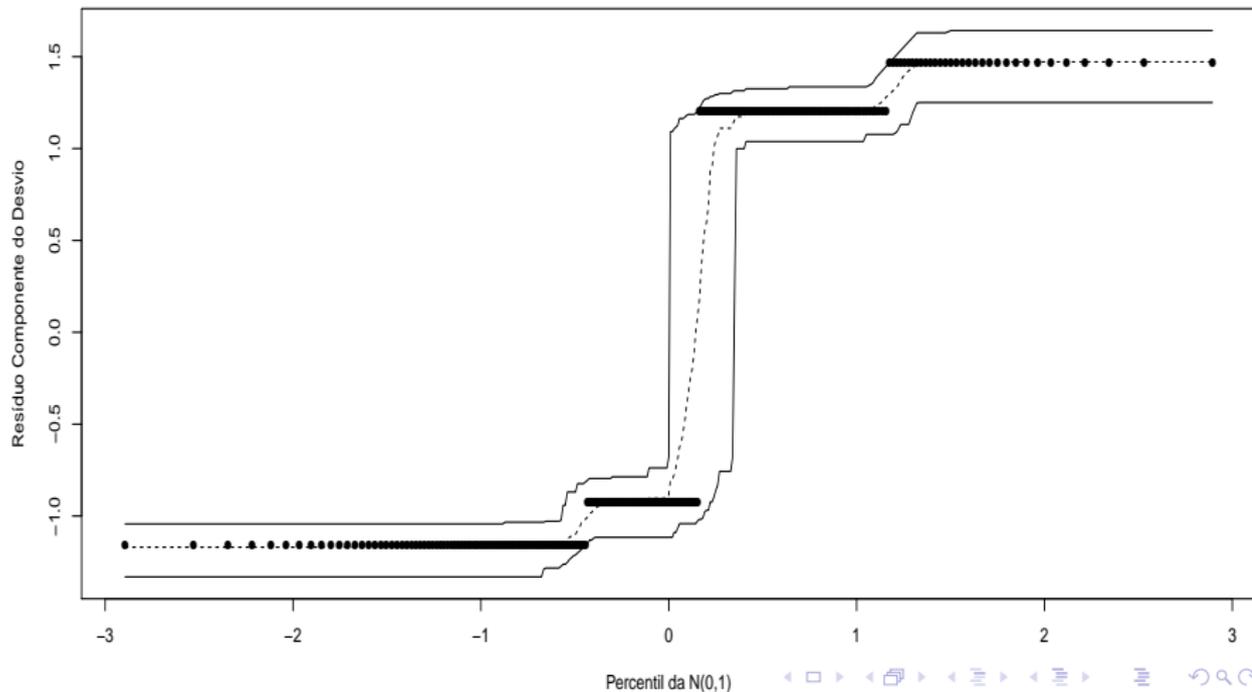
Gráficos de envelopes para os RCD's do modelo final 1

Gráfico de quantil-quantil normal



Gráficos de envelopes para os RCD's do modelo final 2

Gráfico de quantil-quantil normal



Comentários

- As análises de diagnósticos indicaram que os modelos se ajustam bem aos dados.
- Modelo 1: AIC - ; BIC- . Modelo 2
- Vamos comparar os modelos via análise preditiva. Considere a seguinte tabela de contingência que compara os valores observados e preditos (classificação segundo algum critério) das respostas binárias.

Comentários

- VP: verdadeiro positivo, FN: falso negativo, VN: verdadeiro negativo, FP: falso positivo.

Observado	Predito	
	Fracasso (0)	Sucesso (1)
Fracasso (0)	VN	FN
Sucesso (1)	FP	VP

- Duas formas de classificação: simular valores de Bernoullis dada uma estimativa \tilde{p}_i , ou estabelecer algum ponto de corte, digamos c , de sorte que $\tilde{y}_i = 1$, se $\tilde{p}_i > c$ e 0, caso contrário.

Medidas de precisão

- Sensibilidade ou razão de verdadeiro positivo (RVP):

$$RVP = VP / (VP + FN).$$

- Razão de falso negativo (RFN): $RFN = FN / (VP + FN)$

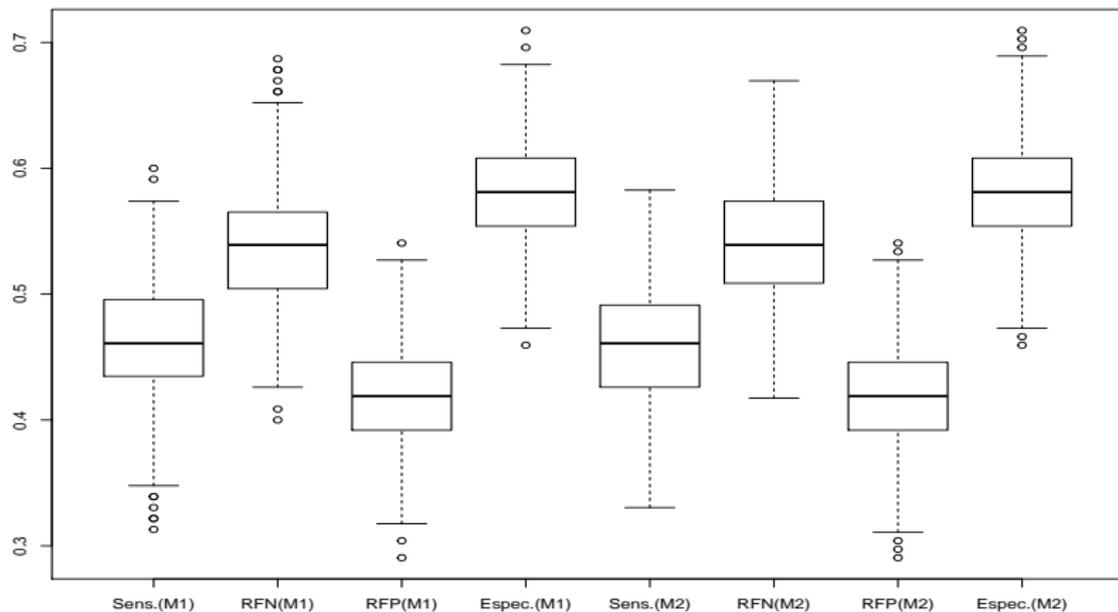
- Razão de falso positivo (RFP): $RFP = FP / (FP + VN)$

- Especificidade ou razão de verdadeiro negativo (RVN):

$$RVN = VN / (FP + VN)$$

- Apresentaremos as duas abordagens: 1- 500 amostras aleatórias de $Y_i \sim \text{Bernoulli}(\tilde{p}_i)$, $i=1,2,\dots,263$ foram simuladas e a sensibilidade, especificidade, RFN e RFP foram calculadas. 2- As curvas ROC foram calculadas, para diferentes pontos de corte.

Boxplots das medidas de precisão



Curvas ROC

