Modelos de regressão para dados politômicos

Prof. Caio Azevedo

Exemplo 17: Estudo sobre hábitos alimentares de jacarés

- O conjunto de dados foi extraído de Agresti (2007).
- 59 jacarés foram aleatoriamente selecionados do Lago George, Flórida, EUA.
- De cada um deles foi medido o comprimento (em metros) e também o volume do principal tipo de alimento existente no estômago.
- Classificação do tipo de alimento: peixe (P), invertebrado (I) e outros (O).

Exemplo 17 (cont.)

- Invertebrados: cobras, insetos aquáticos, lagostim.
- Outros: anfíbios, mamíferos, material vegetal, pedras e outros detritos, tartarugas e inclusive restos de jacarés bebês.
- Objetivo: verificar se existe alguma influência do comprimento dos jacarés nos hábitos alimentares.
- **Observação: estamos considerando que as categorias (P,I,O) são nominais**.

Exemplo 17 (cont.)

- Variáveis resposta: vetor que indica o tipo de alimento encontrado no estômago do jacaré (P, I, O).
- Temos assim, um vetor aleatório trinomial (Bernoulli trivariada), que assume valor 1 para a categoria à qual pertence o tipo de alimento encontrado no estômago do jacaré e 0, para as outras.
- Categorias: 1- P; 2-I; 3-O
- Variável explicativa: comprimento do jacaré.

Banco de dados (versão 1)

jacaré	comprimento	tipo de alimento
1	1,24	I
2	1,30	I
3	1,30	I
4	1,32	P
5	1,32	P
⋮	⋮	⋮
57	3,68	O
58	3,71	P
59	3,89	P

Banco de dados (versão 2)

jacaré	comprimento	tipo de alimento		
		I	P	O
1	1,24	1	0	0
2	1,30	1	0	0
3	1,30	1	0	0
4	1,32	0	1	0
5	1,32	0	1	0
⋮	⋮	⋮		
57	3,68	0	0	1
58	3,71	0	1	0
59	3,89	0	1	0

Distribuição trinomial

- Dizemos que $\boldsymbol{Y} = (Y_1, Y_2)' \sim \text{trinomial}(m, p_1, p_2)$, em que, $m \in 1, 2, ...;\ 0 \leq y_i \leq m; i = 1, 2,\ 0 \leq \sum_{i=1}^{2} y_i \leq m;\ 0 < \sum_{i=1}^{2} p_i < 1,$ se sua fdp for dada por:

$$
\begin{aligned}
P(\boldsymbol{Y} = \boldsymbol{y}) &= P(Y_1 = y_1, Y_2 = y_2) = \frac{m!}{y_1! y_2! (m - y_1 - y_2)!} \\
&\times \quad p_1^{y_1} p_2^{y_2} (1 - p_1 - p_2)^{m - y_1 - y_2} \mathbb{1}_{\{0,1,2,\ldots,m\}}(y_1) \\
&\times \quad \mathbb{1}_{\{0,1,2,\ldots,m-y_1\}}(y_2) \\
&= \frac{m!}{y_1! y_2! (m - y_1 - y_2)!} p_1^{y_1} p_2^{y_2} (1 - p_1 - p_2)^{m - y_1 - y_2} \\
&\times \quad \mathbb{1}_{\{0,1,2,\ldots,n\}}(y_2) \mathbb{1}_{\{0,1,2,\ldots,m-y_2\}}(y_1).
\end{aligned}
$$

Distribuição trinomial

- A distribuição trinomial é uma extensão (bivariada) da distribuição binomial.
- Propriedades:
  - $Y_i \sim \text{binomial}(m, p_i), i = 1, 2.$
  - $Y_i | Y_j = y_j \sim \text{binomial}\left(m, \dfrac{p_i}{1 - p_j}\right), i, j \in \{1, 2\}, i \neq j.$
  - $\boldsymbol{Y} \sim \text{FE}_K(\boldsymbol{\theta})$ (família exponencial k-paramétrica), veja aqui e aqui.
- Mais detalhes, veja as duas referências acima e aqui e aqui.

Modelo de regressão (logitos de referência)

$$\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})' \overset{ind.}{\sim} \text{trinomial}(1, \boldsymbol{p}_i), \boldsymbol{p}_i = (p_{i1}, p_{i2})',$$

$$\ln(p_{i1}/p_{i3}) = \beta_{01} + \beta_{11}x_i \quad ; \quad \ln(p_{i2}/p_{i3}) = \beta_{02} + \beta_{12}x_i,$$

$$p_{i3} = 1 - p_{i1} - p_{i2} \qquad i = 1, 2, ..., 59$$

- Y_{ij} : 1 se o i-ésimo indivíduo pertence à j-ésima (j=1,2,3) categoria e 0 caso contrário.
- $x_i = x_i^* - \overline{x}$: comprimento do i-ésimo jacaré centrado na média dos valores observados $\overline{x} = \frac{1}{59}\sum_{i=1}^{59} x_i^*$, em que x_i^* é o comprimento do i-ésimo jacaré.

Interpretação dos parâmetros

- β_{1j} : parâmetro associado ao incremento (positivo/negativo) na probabilidade do indivíduo i pertencer à categoria j para o aumento em uma unidade no comprimento.
- $\ln(p_{ij}/p_{i3}), j = 1, 2$ representa o log da chance do jacaré i pertencer à categoria j em relação à pertencer a categoria 3 (referência).
- Note que quaisquer outras chances são obteníveis a partir das duas anteriores, com efeito:

$$
\begin{aligned}
\ln(p_{i1}/p_{i2}) &= \ln\left(\frac{p_{i1}/p_{i3}}{p_{i2}/p_{i3}}\right) = \ln(p_{i1}/p_{i3}) - \ln(p_{i2}/p_{i3}) \\
&= (\beta_{01} - \beta_{02}) + (\beta_{11} - \beta_{12})x_i.
\end{aligned}
$$

Interpretação dos parâmetros

- Além disso, temos que

$$p_{ij}/p_{i3} = e^{\beta_{0j}+\beta_{1j}x_i} \rightarrow p_{ij} = e^{\beta_{0j}+\beta_{1j}x_i} p_{i3}, j = 1, 2,$$

assim,

$$\begin{aligned} p_{i1} + p_{i2} + p_{i3} &= 1 \rightarrow p_{i1}/p_{i3} + p_{i2}/p_{i3} + 1 = 1/p_{i3} \\ \rightarrow p_{i3} &= \frac{1}{p_{i1}/p_{i3} + p_{i2}/p_{i3} + 1}. \end{aligned}$$

- Logo,

$$p_{ij} = \frac{e^{\beta_{0j}+\beta_{1j}x_i}}{1 + e^{\beta_{01}+\beta_{11}x_i} + e^{\beta_{02}+\beta_{12}x_i}}, j = 1, 2; p_{i3} = \frac{1}{1 + e^{\beta_{01}+\beta_{11}x_i} + e^{\beta_{02}+\beta_{12}x_i}}.$$

Interpretação dos parâmetros

- Se, $x_i = 0$ (comprimento igual ao comprimento médio da amostra), então $p_{ij} = \dfrac{e^{\beta_{0j}}}{1 + e^{\beta_{01}} + e^{\beta_{02}}}, j = 1, 2$
- Chances (entre pertencer e não pertencer a cada categoria)

$$p_{ij}/(1 - p_{ij}) = \frac{e^{\beta_{0j} + \beta_{1j} x_i}}{1 + e^{\beta_{0j'} + \beta_{1j'} x_i}}, j \neq j' \in \{1, 2\},$$

  podem ser estimadas a partir dos valores de $p_{ij}, j = 1, 2, 3$.
- Se $\beta_{1j} > 0$ quanto maior seu valor, maior a probabilidade do indivíduo pertencer à categoria j, ocorrendo o contrário se $\beta_{1j} < 0$.
- **Exercício: obter as razões de chances**.

Distribuição Multinomial

- Dizemos que um vetor aleatório:

  1. $\boldsymbol{Y} = (Y_1, ..., Y_k)' \sim \text{multinomial}_k(m, \boldsymbol{p}^*)$, $\boldsymbol{p}^* = (p_1, ..., p_k)'$, $\sum_{i=1}^{k} p_i = 1, p_i \in (0,1), \forall i,\ \sum_{i=1}^{k} y_i = m,\ y_i \in \{0, 1, ..., m\}$, ou
  2. $\boldsymbol{Y} = (Y_1, ..., Y_{k-1})' \sim \text{multinomial}_k(m, \boldsymbol{p}), \boldsymbol{p} = (p_1, ..., p_{k-1})'$, $\sum_{i=1}^{k-1} p_i < 1, p_i \in (0,1), \forall i,\ \sum_{i=1}^{k-1} y_i < m,\ y_i \in \{0, 1, ..., m\}, \forall i.$

- Se sua fdp (em que A é o suporte da distribuição, $p_k = 1 - \sum_{i=1}^{k-1} p_i, y_k = m - \sum_{i=1}^{k-1} y_i$) é dada por (próxima página)

Distribuição Multinomial

- (Cont.)

$$
\begin{aligned}
P(\boldsymbol{Y} = \boldsymbol{y}) &= \frac{m!}{\prod_{i=1}^{k} y_i!}\left(\prod_{i=1}^{k} p_i^{y_i}\right)\mathbb{1}_A(\boldsymbol{y}) \\
&= \frac{m!}{\prod_{i=1}^{k} y_i!}\left(\prod_{i=1}^{k} p_i^{y_i}\right)\mathbb{1}_{\{0,1,\ldots,m\}}(y_1)\mathbb{1}_{\{0,1,\ldots,m-y_1\}}(y_2)\times \\
\ldots & \times\mathbb{1}_{\left\{0,1,\ldots,m-\sum_{i=1}^{k-2} y_i\right\}}(y_{k-1}).
\end{aligned} \tag{1}
$$

- Existem outras formas (inclusive em relação às funções indicadoras) de apresentar a fdp da multinomial.

Distribuição Multinomial

- Algumas propriedades
  - $Y_i \sim \text{binomial}(m, p_i)$.
  - $Y_i|Y_j = y_j \sim \text{binomial}\left(m - y_j, \dfrac{p_i}{1-p_j}\right), i \neq j$.
  - $Cov(Y_i, Y_j) = -m(m-1)p_i p_j$.
  - $Corre(Y_i, Y_j) = -\sqrt{\dfrac{p_i p_j}{(1-p_i)(1-p_j)}}$.
  - $\displaystyle\sum_{x_1=0}^{n}\sum_{x_2=0}^{n-x_1}\cdots\sum_{x_{k-1}=0}^{n-\sum_{i=1}^{k-2}x_i}\frac{n!}{x_1!x_2!\ldots\left(n-\sum_{i=1}^{k-1}x_i\right)!}\left(\prod_{i=1}^{k-1}a_i^{x_i}\right)a_k^{x_k} = \left(\sum_{i=1}^{k}a_i\right)^n$.
  - $\boldsymbol{Y} \sim \text{FE}_K(\boldsymbol{\theta})$ (família exponencial k-paramétrica), veja aqui e aqui.
- Mais detalhes, veja as duas referências acima e aqui e aqui.

Modelo de regressão geral (logito de referência)

$$
\begin{aligned}
\boldsymbol{Y}_i & \overset{ind.}{\sim} \text{multinomial}_k(m_i, \boldsymbol{p}_i), \boldsymbol{p}_i = (p_{i1}, p_{i2}, ..., p_{i(k-1)})', \\
\ln(p_{ij}/p_{iJ}) & = \sum_{r=1}^{p} x_{ri}\beta_{rj} = \eta_{ij}, i = 1, 2, ..., n, j = 1, 2, ...k-1
\end{aligned}
$$

- $\boldsymbol{Y}_i$ é um vetor de tamanho $k-1$, retirando-se a categoria J, a qual é a categoria de referência. Ou seja, $\boldsymbol{Y}_i = (Y_{i1}, ..., Y_{ik})'$ retirando-se a variável Y_{iJ}, analogamente para o vetor $\boldsymbol{p}_i = (p_{i1}, ..., p_{ik})'$. Seja ainda $A = \{1, 2, ..., k\} - \{J\}$.
- Y_{ij} : 1 se o i-ésimo indivíduo pertence à j-ésima ($j \in A$) categoria e 0 caso contrário.
- x_{ri} : valor da $r-$ésima covariável associada ao i-ésimo indivíduo.

Interpretação dos parâmetros

- $p_{ij} = \dfrac{e^{\eta_{ij}}}{\sum\limits_{j \in A} e^{\eta_{ij}}}, \forall i, j$. Se j =J, então $\beta_{rJ} = 0, \forall r$ e, assim $\eta_{rJ} = 0$.
- As outras quantidades têm interpretações análogas ao caso da distribuição trinomial.

Estimação

- Verossimilhança (produto de multinomiais, veja Equação (1))

$$L(\boldsymbol{\beta}) \quad \propto \quad \prod_{i=1}^{n}\left\{\prod_{j=1}^{k} p_{ij}^{y_{ij}}\right\} = \prod_{i=1}^{n}\left\{\prod_{j\in A} p_{ij}^{y_{ij}}\right\} p_{iJ}^{y_{iJ}},$$

- Log-verossimilhança

$$l(\boldsymbol{\beta}) \quad = \quad \sum_{i=1}^{n}\left(\sum_{j\in A} y_{ij}\ln p_{ij} + y_{iJ}\ln p_{iJ}\right) + const.$$

Estimação

- Equações de estimação ($r = 1, 2, ..., p, l \in A$) (componentes do vetor escore ($\boldsymbol{S}(\boldsymbol{\beta})$, $\boldsymbol{\beta} = (\beta_{11}, ..., \beta_{p(k-1)})'$))

$$
\begin{aligned}
S(\beta_{rl}) &= \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_{rl}} = \sum_{i=1}^{n} \left\{ \sum_{j \in A} \left(\frac{y_{ij}}{p_{ij}} \frac{\partial p_{ij}}{\partial \beta_{rl}} \right) + \frac{y_{iJ}}{p_{iJ}} \frac{\partial p_{iJ}}{\partial \beta_{rl}} \right\} \\
&= \sum_{i=1}^{n} \left\{ \sum_{j \in A} \left(\frac{y_{ij}}{p_{ij}} \frac{\partial p_{ij}}{\partial \beta_{rl}} \right) - \frac{y_{iJ}}{p_{iJ}} \sum_{j \in A} \frac{\partial p_{ij}}{\partial \beta_{rl}} \right\} \\
&= \sum_{i=1}^{n} \sum_{j \in A} \left(\frac{y_{ij}}{p_{ij}} - \frac{y_{iJ}}{p_{iJ}} \right) \frac{\partial p_{ij}}{\partial \beta_{rl}}
\end{aligned}
$$

Estimação

- Componentes da Matriz Hessiana $\boldsymbol{H}(\boldsymbol{\beta})$:

$$
\begin{aligned}
H(\beta_{rl}\beta_{r'l'}) \quad &= \quad \sum_{i=1}^{n}\sum_{j\in A}\left(-\frac{y_{ij}}{p_{ij}^{2}}\frac{\partial p_{ij}}{\partial\beta_{r'l'}}+\frac{y_{ij}}{p_{iJ}^{2}}\frac{\partial p_{iJ}}{\partial\beta_{r'l'}}\right)\frac{\partial p_{ij}}{\partial\beta_{rl}} \\
&+ \quad \sum_{i=1}^{n}\sum_{j\in A}\left(\frac{y_{ij}}{p_{ij}}-\frac{y_{iJ}}{p_{iJ}}\right)\frac{\partial^{2}p_{ij}}{\partial\beta_{rl}\partial\beta_{r'l'}}
\end{aligned}
$$

- Componentes da Informação de Fisher $\boldsymbol{I}(\boldsymbol{\beta})$:

$$
I(\beta_{rl}\beta_{r'l'})=\sum_{i=1}^{n}\sum_{j\in A}\left(\frac{1}{p_{ij}}\frac{\partial p_{ij}}{\partial\beta_{r'l'}}-\frac{1}{p_{iJ}}\frac{\partial p_{iJ}}{\partial\beta_{r'l'}}\right)\frac{m_{i}\partial p_{ij}}{\partial\beta_{rl}}
$$

Estimação

- O sistema de equações definido por $S(\widetilde{\beta}_{rl}) = 0, r = 1, ..., p, l \in A$ é não linear e, portanto, não tem solução explícita.
- Contudo, ele pode ser resolvido de forma numérica, via algoritmo de Newton-Raphson ou Escore de Fisher, por exemplo.
- Seja $\widehat{\boldsymbol{\beta}}$ o emv de $\boldsymbol{\beta}$. Em geral, para n suficientemente grande,

$$\widehat{\boldsymbol{\beta}} \approx N_{p(k-1)}(\boldsymbol{\beta}, I^{-1}(\boldsymbol{\beta})).$$

- Com as devidas adaptações, intervalos de confiança podem ser construídos e hipóteses podem ser testadas de forma semelhante ao feito para o modelo binomial (aqui)

Validação dos modelos

- O análogo pode ser feito para a função desvio .
- Por exemplo, os testes $\boldsymbol{C}\beta = \boldsymbol{M}$, podem ser conduzidos de maneira análoga (com as devidas adaptações, veja aqui).
- A análise e diagnóstico pode ser feita para cada uma das (k-1) variáveis do vetor $\boldsymbol{Y}$ as quais, nesse caso, tem distribuição binomial. Também podem ser construídos resíduos e medidas considerando-se todo o vetor $\boldsymbol{Y}$.

Validação dos modelos

- Resíduo Componente do Desvio:

$$
\begin{aligned}
RCD_{ij} &= \text{sinal}(Y_{ij} - \widehat{p}_{ij})\sqrt{\frac{2}{1-\widehat{h}_{iij}}}\left\{\left[Y_{ij}\ln\left(\frac{Y_{ij}}{m_i\widehat{p}_{ij}}\right)\right.\right. \\
&+ \left.(m_i - Y_{ij})\ln\left(\frac{m_i - Y_{ij}}{m_i(1-\widehat{p}_{ij})}\right)\right]\mathbb{1}_{\{1,2,\ldots,m_i\}}(Y_{ij}) \\
&- \frac{\left(2m_i\left|\ln\left(1-\widehat{p}_{ij}\right)\right|\right)^{1/2}}{\sqrt{1-\widehat{h}_{iij}}}\mathbb{1}_{\{0\}}(Y_{ij}) \\
&+ \left.\frac{\left(2m_i\left|\ln\left(\widehat{p}_{ij}\right)\right|\right)^{1/2}}{\sqrt{1-\widehat{h}_{iij}}}\mathbb{1}_{\{m_i\}}(Y_{ij})\right\}.
\end{aligned}
$$

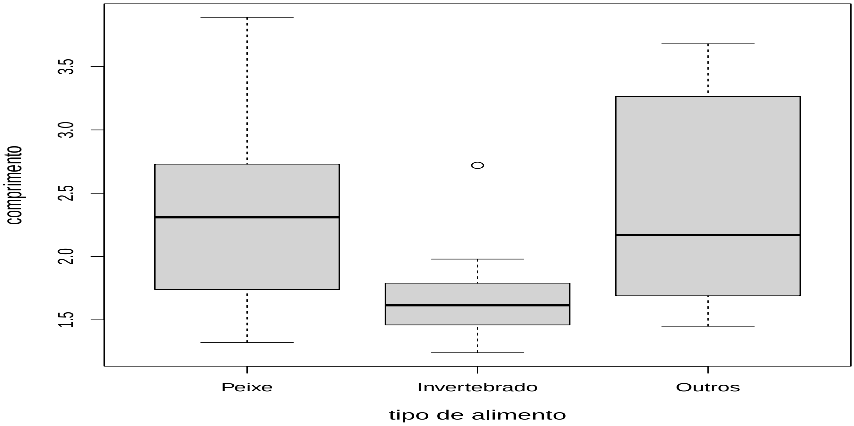
Validação dos modelos

- Resíduo quantílico aleatorizado: repetir o procedimento indicado aqui para cada variável do vetor $\boldsymbol{Y}$.
- Resíduo de Pearson:

$$RP_{ij} = \frac{Y_{ij} - m_i \widehat{p}_{ij}}{\sqrt{m_i \widehat{p}_{ij} \left(1 - \widehat{p}_{ij}\right)}}.$$

- Os critérios de informação e métodos de seleção de modelos são, essencialmente, aqueles apresentados aqui, com as devidas adaptações.
- Pesquisar sobre Análise de influência.
- Utilizaremos a função *mlogit* do pacote homônimo *mlogit*.

Boxplots

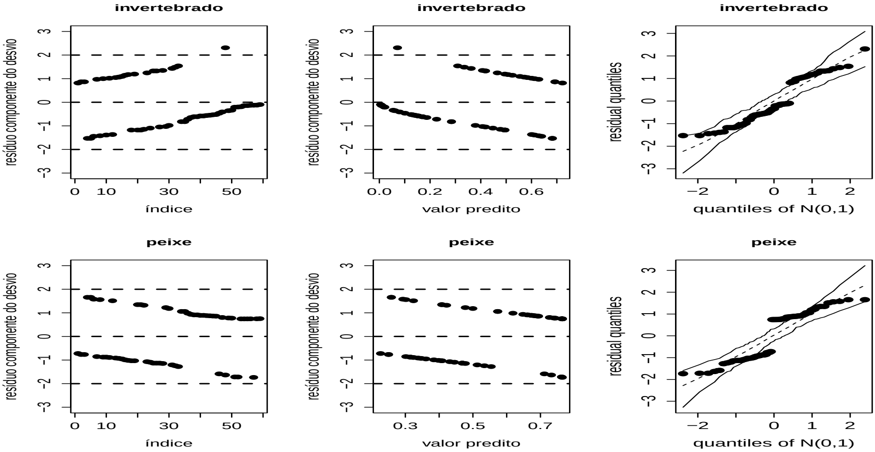


comprimento
1.5
2.0
2.5
3.0
3.5
Peixe
Invertebrado
Outros
tipo de alimento

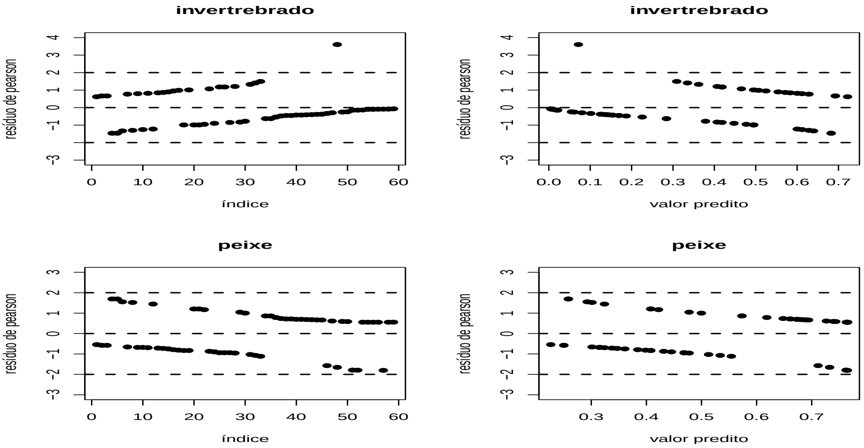
Medidas descritivas

Alimento	Média	Mediana	DP	Var.	CV(%)	Mín.	Máx.
P	2,36	2,31	0,76	0,58	32,23	1,32	3,89
I	1,66	1,61	0,33	0,11	19,70	1,24	2,72
O	2,42	2,17	0,88	0,78	36,44	1,45	3,68

Análise do resíduo componente do desvio



Análise do resíduo de Pearson



Ajuste do modelo

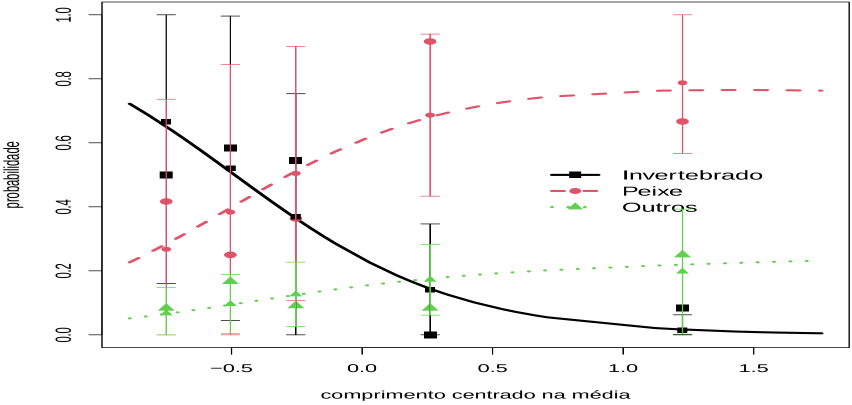
Par.	Est.	EP	IC(95%)	Estat. Z	p-valor
β_{01} (peixe)	1,383	0,422	[0,556;2,211]	3,277	0,0011
β_{02} (invertebrado)	0,445	0,544	[-0,621;1,511]	0,819	0,4131
β_{11} (peixe)	-0,110	0,517	[-1,124;0,903]	-0,213	0,8314
β_{12} (invertebrado)	-2,467	0,900	[-4,229;-0,702]	-2,740	0,0061

Há uma equivalência entre as categorias "peixe" e "outras", em termos do coeficiente angular e entre as categorias "invertebrado" e "outras" em relação ao intercepto.

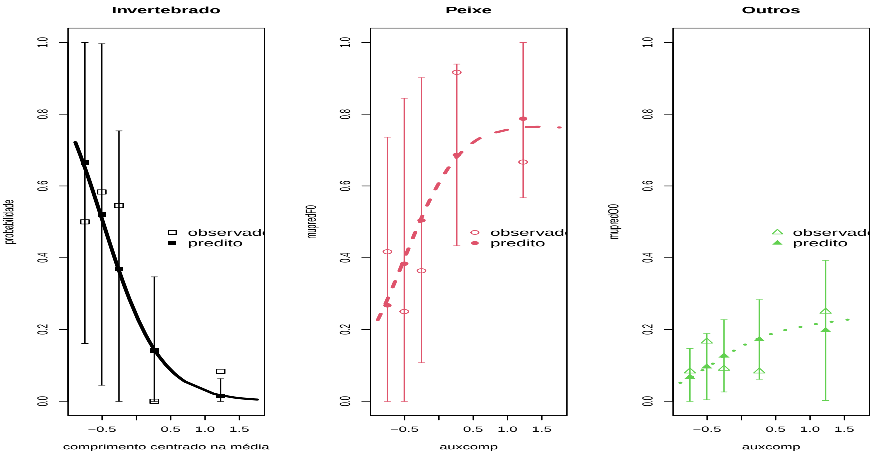
Proporções (probabilidades) estimadas

- Peixe: $p_{P_i} = \dfrac{e^{1,383-0,110x_i}}{1+e^{1,383-0,110x_i}+e^{0,445-2,467x_i}}$
- Invertebrado : $p_{I_i} = \dfrac{e^{0,445-2,467x_i}}{1+e^{1,383-0,110x_i}+e^{0,445-2,467x_i}}$
- Outros : $p_{O_i} = \dfrac{1}{1+e^{1,383-0,110x_i}+e^{0,445-2,467x_i}}$

Proporções observadas e preditas pelo modelo



Proporções observadas e preditas pelo modelo



Desempenho de classificação

Classificação realizada a partir de valores simulados de 59 tetranomiais com parâmetros $\widetilde{p}_{ij}$.

	Verdadeiro		
Predito	P	I	O
P	18	6	4
I	7	14	0
O	6	0	4

Desempenho de classificação

Estatísticas gerais

$$\text{Acurácia (Ac)} : 0{,}6102$$

$$\text{IC}(95\%) : (0.4744;0.7345)$$

$$\text{Taxa de não infomação (TNI)} : 0{,}5254$$

$$\text{P-Value [Acur} > \text{TNI]} : 0{,}1201$$

$$\text{Kappa} : 0{,}3578$$

Desempenho de classificação (por grupo)

	Classe:P	Classe:I	Classe:O
Sensibilidade	0,5806	0,7000	0,5000
Especificidade	0,6429	0,8205	0,8824
Valor preditivo positivo	0,6429	0,6667	0,4000
Valor preditivo negativo	0,5806	0,8421	0,9184
Prevalência	0,5254	0,3390	0,1356
Taxa de detecção	0,3051	0,2373	0,0678
Prevalência de deteçãoo	0,4746	0,3559	0,1695
Acurácia balanceada	0,6118	0,7603	0,6912

Expressões

	Observado	
Predito	Evento	Não-Evento
Evento	A	B
Não Evento	C	D

Fórmulas:

- Acurácia $= (A+D)/(A+B+C+D)$.
- TNI: maior proporção observada entre as classes.
- Kappa: índice de concordância.

Cont.

Fórmulas:

- Sensibilidade = A/(A+C)
- Especifidade = D/(B+D)
- Prevalência = (A+C)/(A+B+C+D)
- Valor preditivo positivo = (Sensibilidade$\times$Prevalência)/((Sensibilidade$\times$Prevalência) + ((1-Especifidade)$\times$(1-Prevalência)))

Cont.

Fórmulas:

- Valor preditivo negativo = (Especifidade $\times$(1-Prevalência))/(((1-Sensibilidade)$\times$Prevalência) + ((Especifidade)$\times$(1-Prevalência)))
- Taxa de detecção = A/(A+B+C+D)
- Prevalência de detecção= (A+B)/(A+B+C+D)
- Acurácia balanceada = (sensibilidade+especificidade)/2

Comentários

- O modelo, aparentemente, apresentou um mal ajuste segundo RCD.
- Contudo, do ponto de vista preditivo apresentou uma ajuste bem satisfatório (probabilidades previstas e observadas) e um pouco satisfatório (usando dados simulados).
- Uma forma de melhorar ajuste seria considerar outras covariáveis.

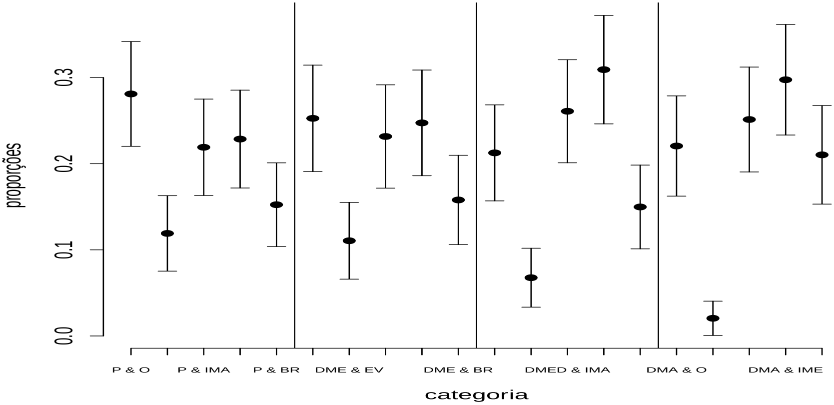
Exemplo 18: Efeito das doses de medicamentos intravenosos em pacientes com trauma hemorrágico subaracnóide

- O conjunto de dados foi extraído de Agresti (2010).
- Corresponde a distribuição de pacientes, que receberam algum tipo de tratamento (dose de medicamento intravenoso), de acordo com o desfecho clínico (óbito (O), estado vegetativo (EV), incapacidade maior (IMA), incapacidade menor (IME), boa recuperação (BR)), de acordo com a escala de "resultado" de Glasgow.
- Os totais de paciente, por cada dose, foram fixados (produto de multinomiais).

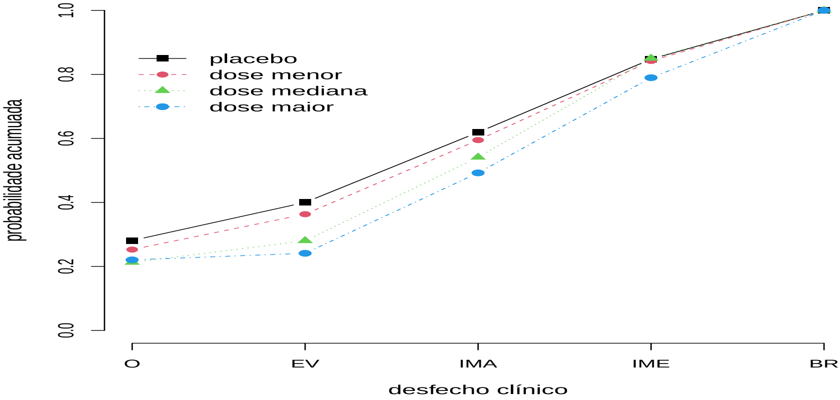
Dados

	Escala de resultado de Glasgow					
tratamento	O	EV	IMA	IME	BR	Total
placebo (P)	59	25	46	48	32	210
dose menor (DME)	48	21	44	47	30	190
dose mediana (DMED)	44	14	54	64	31	207
dose maior (DMA)	43	4	49	58	41	195

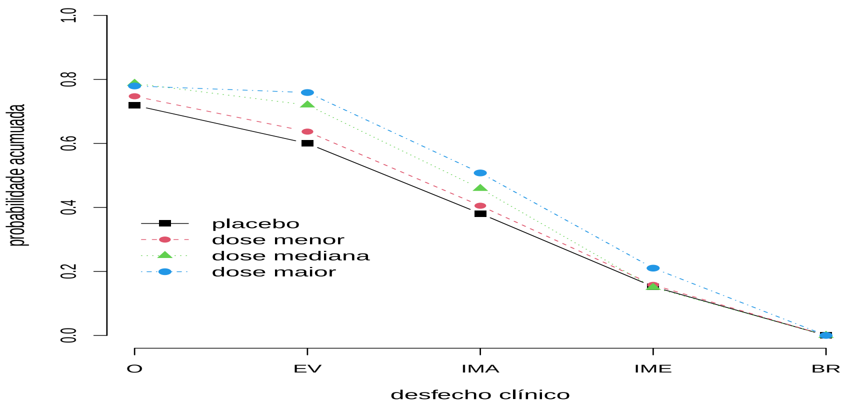
Análise descritiva



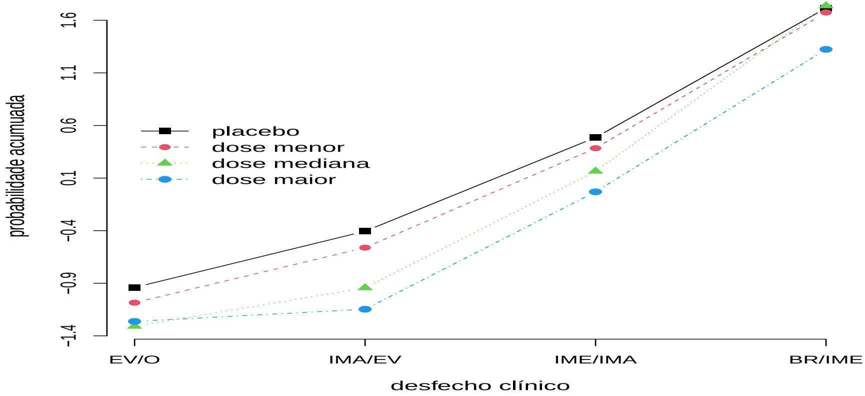
Função de distribuição acumulada



Função de sobrevivência



Logito (FDA/FDS)



Modelo de regressão (logitos cumulativos)

$$\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}, Y_{i5})' \overset{ind.}{\sim} \text{pentanomial}(m_i, \boldsymbol{p}_i)$$

$$\boldsymbol{p}_i = (p_{i1}, p_{i2}, p_{i3}, p_{i4}, p_{i5})', \quad \sum_{j=1}^{5} Y_{ij} = m_i, \sum_{j=1}^{5} p_{ij} = 1,$$

$$\text{logito}(P(Y_{ij} \leq j)) = \ln\left(\frac{P(Y_{ij} \leq j)}{P(Y_{ij} > j)}\right) = \text{logito}(p_{ij})$$

$$= \alpha_j + \beta x_{ij}, x_{ij} = i,$$

$$i = 1(\text{P}), 2(\text{DME}), 3(\text{DMED}), 4(\text{DMA}).$$

$$j = 1(\text{O}), 2(\text{EV}), 3(\text{IMA}), 4(\text{IME}), 5(\text{BR}),$$

$$m_1 = 210, m_2 = 190, m_3 = 207, m_4 = 195.$$

(Cont.) Modelo de regressão (logitos cumulativos)

- Y_{ij} : quantidade de indivíduos submetidos à dose i que apresentaram o desfecho clínico j.
- x_{ij} : valor da dose i, relativa ao desfecho clínico j.
- $P(Y_{ij} \leq j) = \dfrac{e^{\alpha_j + \beta x_{ij}}}{1 + e^{\alpha_j + \beta x_{ij}}}$.
- Logo

$$
\begin{aligned}
P(Y_{ij} = j) &= P(Y_{ij} \leq j) - P(Y_{ij} \leq j - 1) \\
&= \frac{e^{\alpha_j + \beta x_{ij}}}{1 + e^{\alpha_j + \beta x_{ij}}} - \frac{e^{\alpha_{j-1} + \beta x_{i(j-1)}}}{1 + e^{\alpha_{j-1} + \beta x_{i(j-1)}}},
\end{aligned}
$$

(Cont.) Modelo de regressão (logitos cumulativos)

- (Cont.) Note que

$$P(Y_{ij} \leq 0) = 0, P(Y_{ij} \leq 5) = 1 \,(\text{última categoria}),$$

e, além disso

$$-\infty < \alpha_1 < \alpha_2 < \alpha_3 < \alpha_4 < \infty.$$

- Ou seja, os interceptos tem de ter uma ordenação crescente, são um total $k-1$ (em que k é o número de categorias) e sempre

$$P(Y_{ij} \leq 0) = 0 \text{ e } P(Y_{ij} \leq k) = 1.$$

(Cont.) Modelo de regressão (logitos cumulativos)

- Para o mesmo nível j da resposta e diferentes níveis i da covariável:

$$
\begin{aligned}
\text{logito}(P_{i'j}) - \text{logito}(P_{ij}) &= \beta(x_{i'j} - x_{ij}) \rightarrow \\
\frac{P_{i'j}/(1 - P_{i'j})}{P_{ij}/(1 - P_{ij})} &= e^{\beta(x_{i'j} - x_{ij})}.
\end{aligned}
$$

- Para o mesmo nível i da covariável e diferentes níveis j da resposta:

$$
\begin{aligned}
\text{logito}(P_{ij'}) - \text{logito}(P_{ij}) &= \alpha_{j'} - \alpha_j \rightarrow \\
\frac{P_{ij'}/(1 - P_{ij'})}{P_{ij}/(1 - P_{ij})} &= e^{\alpha_{j'} - \alpha_j}.
\end{aligned}
$$

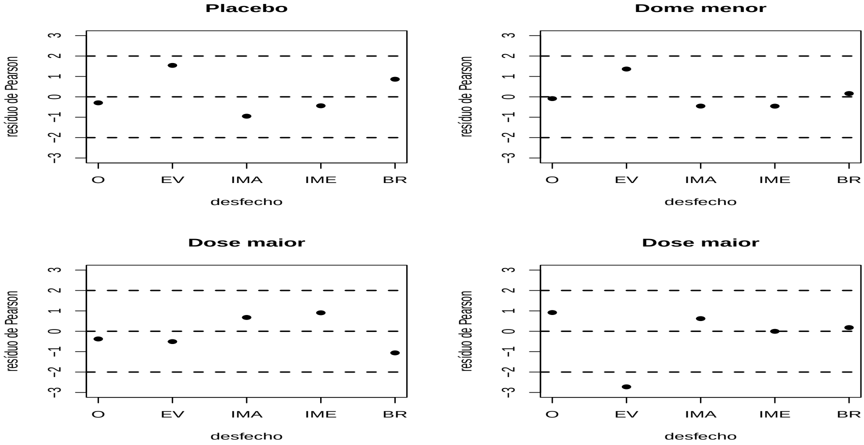
Comentários e ajuste do modelo

- Os processos de ajuste (estimativa dos parâmetros), verificação da qualidade do ajuste e seleção de comparação de modelos são bem semelhantes ao modelo anterior.
- Contudo, as expressões, eventualmente, necessitarão de ser adaptadas.
- Exercício, verificar como ficam as expressões supramencionadas.
- Utilizaremos a função *vglm* do pacote homônimo *VGAM*

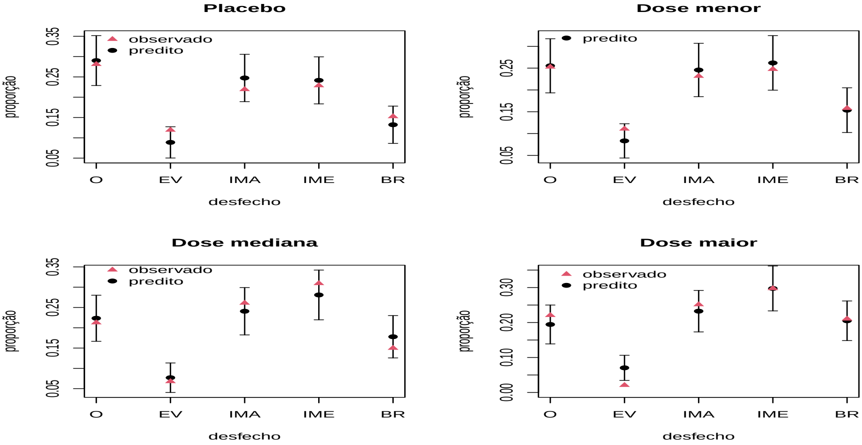
Ajuste do modelo

- Teste de aderência de qui-quadrado (comparando as frequências populacionais (amostra) com as populacionais (preditas) com as observadas).
- p-valores:
  - Placebo: 0,9192.
  - Dose menor: 0,9734.
  - Dose mediana: 0,9283.
  - Dose maior: 0,8751.

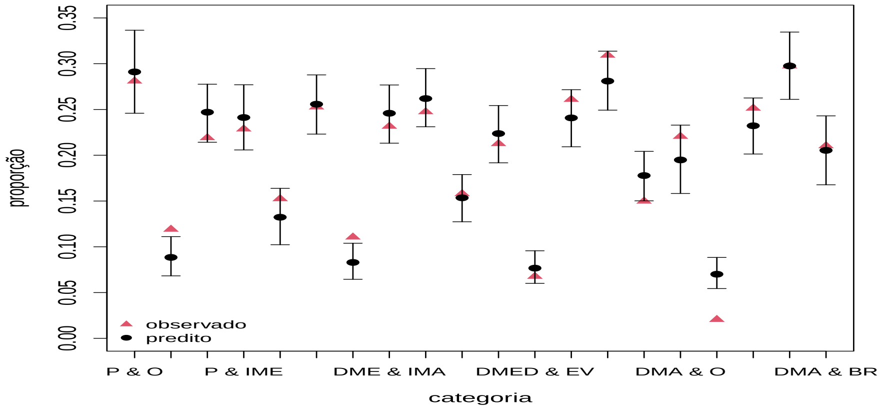
Análise residual (Resíduo de Pearson)



Análise preditiva (assintótica)



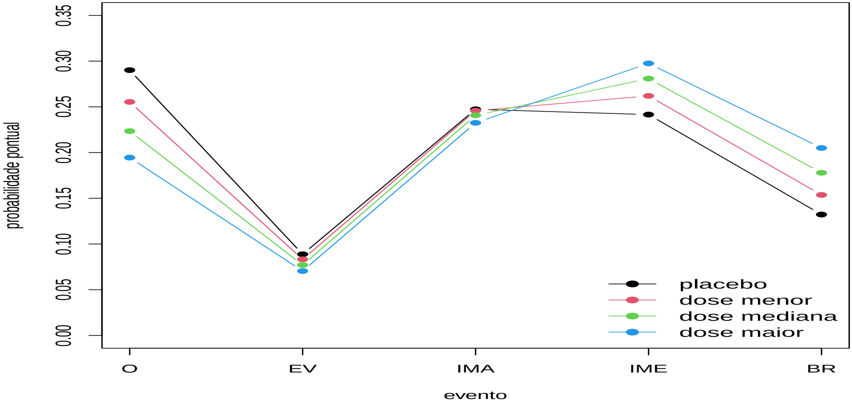
Análise preditiva (numérica)



Estimativas dos parâmetros modelo

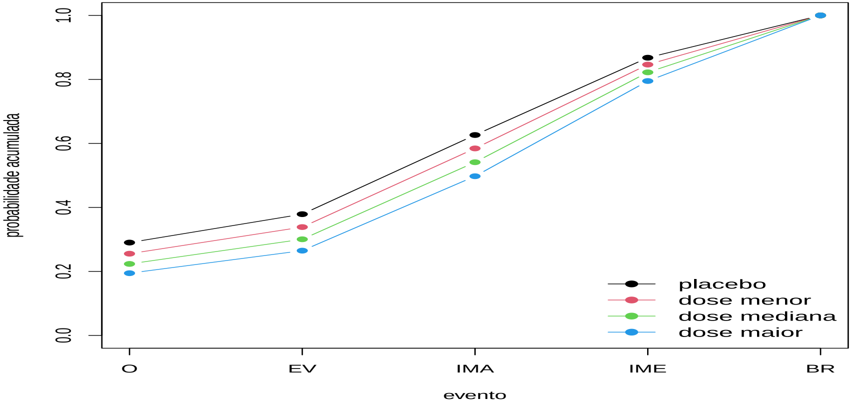
Parâmetro	Estimativa	EP	IC(95%)	Estat. Z	p-valor
α_1	-0,72	0,16	[-1,03 ; -0,41]	-4,53	$< 0,0001$
α_2	-0,32	0,16	[-0,63 ; -0,01]	-2,04	0,0417
α_3	0,69	0,16	[0,38 ; 1,00]	4,38	$< 0,0001$
α_4	2,06	0,17	[1,72 ; 2,40]	11,84	$< 0,0001$
β	-0,18	0,06	[-0,29 ; -0,07]	-3,12	0,0018

FDA predita

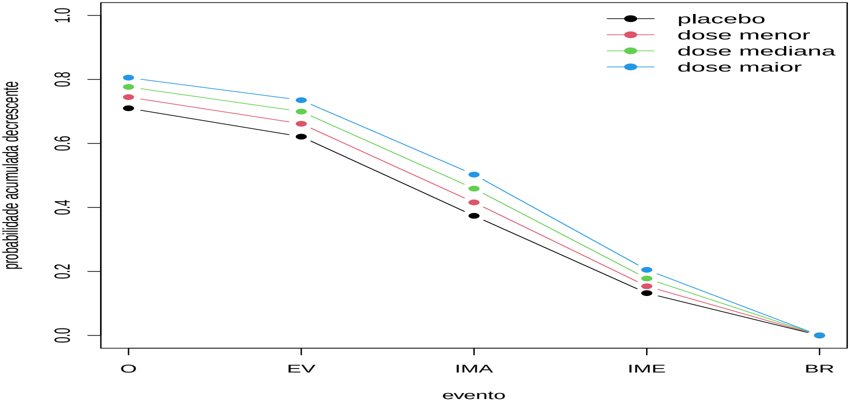


probabilidade pontual
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
O
EV
IMA
IME
BR
evento
placebo
dose menor
dose mediana
dose maior

FDA predita


probabilidade acumulada
0.0
0.2
0.4
0.6
0.8
1.0
O
EV
IMA
IME
BR
evento
placebo
dose menor
dose mediana
dose maior

FDS predita



placebo
dose menor
dose mediana
dose maior
probabilidade acumulada decrescente
0.0
0.2
0.4
0.6
0.8
1.0
O
EV
IMA
IME
BR
evento

Comentários

- O modelo se ajustou de forma satisfatória aos dados.
- Os resultados sugerem um efeito positivo (maior probabilidade de recuperação) à medida que a dose do tratamento aumenta.
- Outras referências:
  - Análise de dados categorizados e R: aqui.
  - Análise de MLG's multivariados e R: aqui, aqui, aqui.
  - Análise residual para modelos de regressão multinomiais: aqui, aqui.