

# ME720 - Modelos Lineares Generalizados

## Parte 12 - Modelos Multinomiais

Profa. **Larissa Avila Matos**

## Modelos de resposta multinomial

Nas aulas anteriores vimos modelos lineares generalizados (MLGs) para variáveis de resposta binária que assumem um componente aleatório binomial.

MLGs para variáveis com resposta em várias categorias assumem um componente aleatório multinomial.

A regressão logística multinomial pode ser vista como uma extensão do modelo logístico binário, em situações nas quais a variável dependente tem múltiplas categorias.

Iremos ver, generalizações da regressão logística para variáveis de resposta multinomial.

Modelos separados estão disponíveis para variáveis de resposta nominais e para variáveis de resposta ordinais.

## Respostas Nominais

## Respostas Nominais: Modelo logito categórico

Para variáveis de resposta em escala nominal com  $c$  categorias, os modelos logísticos com várias categorias descrevem simultaneamente as log-chances para todos os pares  $\frac{c(c-1)}{2}$  de categorias. Dada uma certa escolha de  $c - 1$ , o restante é redundante.

Nós construímos um modelo logístico multinomial pareando cada categoria da resposta com uma categoria de referência (categoria  $c$ ), usando

$$\log \left( \frac{\pi_{i1}}{\pi_{ic}} \right), \log \left( \frac{\pi_{i2}}{\pi_{ic}} \right), \dots, \log \left( \frac{\pi_{ic-1}}{\pi_{ic}} \right).$$

O logito da  $j$ -ésima categoria-referência,  $\log \left( \frac{\pi_{ij}}{\pi_{ic}} \right)$ , é o logito da probabilidade condicional

$$\text{logito}(\mathbb{P}(Y_{ij} = 1 | Y_{ij} = 1 \text{ ou } Y_{ic} = 1)) = \log \left( \frac{\pi_{i1}}{\pi_{ic}} \right).$$

Seja  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  os valores das variáveis explicativas para o  $i$ -ésimo sujeito, e  $\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp})'$  os parâmetros para o  $j$ -ésimo logito.

**Modelo logito multinominal (casela de referência):**

$$\log \left( \frac{\pi_{ij}}{\pi_{ic}} \right) = \mathbf{x}_i \boldsymbol{\beta}_j = \sum_{k=1}^p \beta_{jk} x_{ik}, \quad j = 1, \dots, c-1.$$

Este modelo descreve simultaneamente o efeito de  $\mathbf{x}$  nos  $c-1$  logitos. Os efeitos variam de acordo com a resposta pareada com a referência.

As  $c-1$  equações determinam equações para logitos com outros pares de categorias de resposta, uma vez que

$$\log \left( \frac{\pi_{ia}}{\pi_{ib}} \right) = \log \left( \frac{\pi_{ia}}{\pi_{ic}} \right) - \log \left( \frac{\pi_{ib}}{\pi_{ic}} \right) = \mathbf{x}_i (\boldsymbol{\beta}_a - \boldsymbol{\beta}_b).$$

Como em outros modelos, normalmente  $x_{i1} = 1$  para o coeficiente de intercepto, que também difere para cada logito.

O modelo trata a variável resposta como escala nominal, no seguinte sentido: se o modelo se mantém e as categorias são permutadas de alguma maneira, o modelo ainda se mantém com a permutação correspondente dos efeitos.

Podemos expressar modelo logito multinominal (casela de referência) diretamente em termos das probabilidades de resposta  $\pi_{ij}$ .

**Modelo logito multinominal (casela de referência):**

$$\pi_{ij} = \frac{\exp \mathbf{x}_i \boldsymbol{\beta}_j}{1 + \sum_{h=1}^{c-1} \exp \mathbf{x}_i \boldsymbol{\beta}_h},$$

com  $\boldsymbol{\beta}_c = 0$ . (Os parâmetros são iguais a zero para uma categoria de referência por motivos de identificação)

O modelo acima é derivado do fato que

$$\frac{\pi_{ij}}{\pi_{ic}} = \exp \left( \sum_{k=1}^p \beta_{jk} x_{ik} \right) = \exp(\mathbf{x}_i \boldsymbol{\beta}_j),$$

uma vez que,  $\sum_{j=1}^c \pi_{ij} = 1$ , para cada  $j$ .

Para  $c = 2$ , este modelo é simplesmente o modelo logística para dados binários.

## Distribuição multinomial



## Distribuição multinomial

Seja  $\mathbf{Y} = (Y_1, \dots, Y_c)$ , onde  $Y_j$  representa o número de vezes que a  $j$ -ésima categoria ocorre em  $N$  independentes ensaios.

Além disso, seja  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_c)$ , com  $\pi_j$  = probabilidade do  $j$ -ésimo evento ocorrer em qualquer ensaio.

Definindo  $\mathbf{Y} = (Y_1, \dots, Y_c)$ ,  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_c)$  e lembrando que  $Y_c = N - \sum_{i=1}^{c-1} Y_i$  e  $\pi_c = 1 - \sum_{i=1}^{c-1} \pi_i$ , a f.d.p. de  $\mathbf{Y}$  é dada por

$$\begin{aligned} f(\mathbf{y}; \boldsymbol{\pi}) &= \frac{N!}{y_1! \dots y_c!} \pi_1^{y_1} \pi_2^{y_2} \dots \pi_c^{y_c} \\ &= \exp \left[ \sum_{i=1}^c y_i \log(\pi_i) + \log \left( \frac{N!}{y_1! \dots y_c!} \right) \right] \\ &= \exp \left[ \sum_{i=1}^{c-1} y_i \log \left( \frac{\pi_i}{\pi_c} \right) + N \log(\pi_c) + \log \left( \frac{N!}{y_1! \dots y_c!} \right) \right]. \end{aligned}$$

Portanto, essa distribuição pertence a família exponencial multivariada com parâmetro de escala, ou seja,

$$f(\mathbf{y}; \boldsymbol{\theta}, \phi) = \exp \left[ \frac{\mathbf{y}'\boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\phi)} + c(\mathbf{y}, \phi) \right],$$

onde  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{c-1})$ , com

$$\blacksquare \theta_i = \log \left( \frac{\pi_i}{\pi_c} \right) = \log \left( \frac{\pi_i}{1 - \sum_{i=1}^{c-1} \pi_i} \right),$$

$$\blacksquare b(\theta_i) = -N \log \left( 1 - \sum_{i=1}^{c-1} \pi_i \right),$$

$$\blacksquare a(\phi) = 1, \text{ e}$$

$$\blacksquare c(\mathbf{y}, \phi) = \log \left( \frac{N!}{y_1! \dots y_c!} \right).$$

## MLG multivariado

Como vimos, a distribuição multinomial é membro da FE multivariada, então o modelo logito categórico é um MLG multivariado.

Para o vetor de respostas  $\mathbf{Y}_i$  do  $i$ -ésimo sujeito, com  $\boldsymbol{\mu}_i = \mathbb{E}(\mathbf{Y}_i)$ , seja  $\mathbf{g}$  um vetor de função de ligação. O MLG multivariado tem a seguinte forma

$$\mathbf{g}(\boldsymbol{\mu}_i) = \mathbf{X}_i \boldsymbol{\beta}, \quad i = 1, \dots, N,$$

onde a linha  $j$  da matriz  $\mathbf{X}_i$  contém os valores das covariáveis para o componente da resposta  $Y_{ij}$ .

Para essa representação, temos que para o  $i$ -ésimo sujeito a f.d.p é dada por

$$f(\mathbf{y}_i; \boldsymbol{\theta}, \phi) = \exp \left[ \sum_{j=1}^{c-1} y_{ij} \log \left( \frac{\pi_{ij}}{\pi_{ic}} \right) + \log (\pi_{ic}) \right],$$

onde  $\pi_{ic} = 1 - \sum_{j=1}^{c-1} \pi_{ij}$  e

$$g_j(\boldsymbol{\mu}_i) = \log \left( \frac{\mu_{ij}}{1 - \sum_{j=1}^{c-1} \mu_{ij}} \right), \quad \boldsymbol{\mu}_i = (\pi_{i1}, \dots, \pi_{i,c-1})'.$$

Temos que,  $\mathbf{X}_i$  é uma matriz de dimensão  $(c-1) \times (c-1)p$ , onde

$$\mathbf{X}_i \boldsymbol{\beta} = \begin{bmatrix} \mathbf{x}_i & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_i & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{x}_i \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{c-1} \end{bmatrix},$$

com  $\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp})'$  e  $\mathbf{0} = (0, \dots, 0)_{1 \times p}$ .

## Ajustando um modelo logito categórico

O ajuste de MV maximiza a função de verossimilhança multinomial sujeita a  $\pi_{ij}$  simultaneamente satisfazendo as  $c - 1$  equações que especifica o modelo.

A contribuição para a log-verossimilhança do  $i$ -ésimo sujeito é

$$\begin{aligned}\ell(\mathbf{y}_i) = \log(L(\mathbf{y}_i)) &= \sum_{j=1}^{c-1} y_{ij} \log \left( \frac{\pi_{ij}}{\pi_{ic}} \right) + \log(\pi_{ic}) \\ &= \sum_{j=1}^{c-1} y_{ij} \log \left( \frac{\pi_{ij}}{\pi_{ic}} \right) - \log \left( \frac{1}{\pi_{ic}} \right).\end{aligned}$$

Sabemos que  $\log \left( \frac{\pi_{ij}}{\pi_{ic}} \right) = \mathbf{x}_i \boldsymbol{\beta}_j = \eta_{ij}$  e  $\pi_{ic} = \frac{1}{1 + \sum_{j=1}^{c-1} \exp(\mathbf{x}_i \boldsymbol{\beta}_j)}$ .

Então,

$$\begin{aligned}\ell(\boldsymbol{\beta}; \mathbf{y}) = \sum_{i=1}^N \ell(\mathbf{y}_i) &= \sum_{i=1}^N \left[ \sum_{j=1}^{c-1} y_{ij} \mathbf{x}_i \boldsymbol{\beta}_j - \log \left( 1 + \sum_{j=1}^{c-1} \exp(\mathbf{x}_i \boldsymbol{\beta}_j) \right) \right] \\ &= \sum_{i=1}^N \sum_{j=1}^{c-1} y_{ij} \mathbf{x}_i \boldsymbol{\beta}_j - \sum_{i=1}^N \log \left( 1 + \sum_{j=1}^{c-1} \exp(\mathbf{x}_i \boldsymbol{\beta}_j) \right).\end{aligned}$$

Portanto,

$$\ell(\boldsymbol{\beta}; \mathbf{y}) = \sum_{i=1}^{c-1} \left[ \sum_{k=1}^p \beta_{jk} \left( \sum_{i=1}^N x_{ik} y_{ij} \right) \right] - \sum_{i=1}^N \log \left( 1 + \sum_{j=1}^{c-1} \exp \left( \sum_{k=1}^p \beta_{jk} x_{ik} \right) \right).$$

Calculando  $\frac{\partial \ell(\beta; \mathbf{y})}{\partial \beta_{jk}}$ , temos que

$$\begin{aligned}\frac{\partial \ell(\beta; \mathbf{y})}{\partial \beta_{jk}} &= \sum_{i=1}^N x_{ik} y_{ij} - \sum_{i=1}^N \frac{1}{1 + \sum_{h=1}^{c-1} \exp(\mathbf{x}_i \beta_h)} \exp(\mathbf{x}_i \beta_j) x_{ik} \\ &= \sum_{i=1}^N \left[ x_{ik} \left( y_{ij} - \frac{\exp(\mathbf{x}_i \beta_j)}{1 + \sum_{h=1}^{c-1} \exp(\mathbf{x}_i \beta_h)} \right) \right] = \sum_{i=1}^N [x_{ik} (y_{ij} - \pi_{ij})].\end{aligned}$$

Portanto, as equações de log-verossimilhança são dadas por

$$\sum_{i=1}^N x_{ik} y_{ij} = \sum_{i=1}^N x_{ik} \pi_{ij}.$$

Diferenciando novamente, temos que

$$\frac{\partial^2 \ell(\boldsymbol{\beta}; \mathbf{y})}{\partial \beta_{jk} \partial \beta_{jk'}} = - \sum_{i=1}^N x_{ik} x_{ik'} \pi_{ij} (1 - \pi_{ij}),$$

e para  $j \neq j'$ ,

$$\frac{\partial^2 \ell(\boldsymbol{\beta}; \mathbf{y})}{\partial \beta_{jk} \partial \beta_{jk'}} = \sum_{i=1}^N x_{ik} x_{ik'} \pi_{ij} \pi_{ij'}.$$

A matriz da informação consiste em  $(c-1)^2$  blocos de tamanho  $p \times p$ ,

$$-\frac{\partial^2 \ell(\boldsymbol{\beta}; \mathbf{y})}{\partial \beta_j \partial \beta_{j'}} = \sum_{i=1}^N \pi_{ij} [I(j = j') - \pi_{ij'}] \mathbf{x}'_i \mathbf{x}_i,$$

onde  $I(\cdot)$  é a função indicadora. A matriz Hessiana é definida-negativa, portanto a função de log-verossimilhança é côncava e possui um máximo único.



As matrizes de informação observada e esperada são idênticas. Portanto, o método Newton-Raphson é equivalente ao Escore de Fisher, uma consequência de usarmos a função de ligação canônica.

A convergência é geralmente rápida, a menos que pelo menos uma estimativa seja ininita ou não exista.

## Desvio e Inferência para Modelos Multinomiais

Para o modelo logito multinomial com categoria de referência, o estimador de MV,  $\widehat{\beta}$ , tem distribuição normal para amostra suficientemente grande.

Como de costume, os erros padrão são calculados a partir da matriz de informação, onde é a raiz quadrada dos elementos da diagonal principal da inversa da matriz de informação.

Os  $\widehat{\beta}_j$  são correlacionados.

Como vimos, as estimativas  $\widehat{\beta}_a - \widehat{\beta}_b$  dos efeitos no preditor linear para  $\log\left(\frac{\pi_{ia}}{\pi_{ib}}\right)$  não dependem de qual categoria é a referência.

Podemos realizar inferência usando os teste da razão de verossimilhanças, Wald e Escore para MLGs.

A estatística do teste da razão de verossimilhança é igual à diferença nos valores do desvio para a comparação dos modelos.

**Exercício:** Encontrar o desvio para o modelo logito multinomial.

## Exemplo: Alligator Food Choice

O conjunto de dados vem de um estudo de fatores que influenciam a escolha alimentar primária dos jacarés.

O estudo capturou 219 jacarés em quatro lagos da Flórida.

A variável de resposta em escala nominal é o principal tipo de alimento, em volume, encontrado no estômago de um jacaré.

Os dados classificam a principal escolha de alimentos de acordo com o lago de captura e o tamanho do jacaré. A variável tamanho é binária, distinguindo entre jacarés não adultos (comprimento  $\leq 2,3$  metros) e jacarés adultos.

■ Variável Resposta: escolha alimentar primária (5 categorias)

- $y_1$ : Peixe (Fish),
- $y_2$ : Invertebrado (Invertebrate),
- $y_3$ : Réptil (Reptile),
- $y_4$ : Pássaro (Bird),
- $y_5$ : Outros (Other).

■ Covariáveis:

- Tamanho (Size): não adultos (comprimento  $\leq 2,3$  metros) e adultos, onde **size= 1** se não adulto e **size= 0** se adulto;
- Lago (Lake): Hancock (**lake=1**), Ocklawaha (**lake=2**), Trafford (**lake=3**) e George (**lake=4**).

| Lago      | Tamanho    | Escolha alimentar primária |              |        |         |        | Total |
|-----------|------------|----------------------------|--------------|--------|---------|--------|-------|
|           |            | Peixe                      | Invertebrado | Réptil | Pássaro | Outros |       |
| Hancock   | $\leq 2,3$ | 23                         | 4            | 2      | 2       | 8      | 39    |
|           | $> 2,3$    | 7                          | 0            | 1      | 3       | 5      | 16    |
| Ocklawaha | $\leq 2,3$ | 5                          | 11           | 1      | 0       | 3      | 20    |
|           | $> 2,3$    | 13                         | 8            | 6      | 1       | 0      | 28    |
| Trafford  | $\leq 2,3$ | 5                          | 11           | 2      | 1       | 5      | 24    |
|           | $> 2,3$    | 8                          | 7            | 6      | 3       | 5      | 29    |
| George    | $\leq 2,3$ | 16                         | 19           | 1      | 2       | 3      | 41    |
|           | $> 2,3$    | 17                         | 1            | 0      | 1       | 3      | 22    |
| Total     |            | 94                         | 61           | 19     | 13      | 32     | 219   |

```
# Arquivo Alligators.dat em www.stat.ufl.edu/~aa/glm/data  
dados = read.table("Alligators.txt", header=T)  
dados
```

|   | lake | size | y1 | y2 | y3 | y4 | y5 |
|---|------|------|----|----|----|----|----|
| 1 | 1    | 1    | 23 | 4  | 2  | 2  | 8  |
| 2 | 1    | 0    | 7  | 0  | 1  | 3  | 5  |
| 3 | 2    | 1    | 5  | 11 | 1  | 0  | 3  |
| 4 | 2    | 0    | 13 | 8  | 6  | 1  | 0  |
| 5 | 3    | 1    | 5  | 11 | 2  | 1  | 5  |
| 6 | 3    | 0    | 8  | 7  | 6  | 3  | 5  |
| 7 | 4    | 1    | 16 | 19 | 1  | 2  | 3  |
| 8 | 4    | 0    | 17 | 1  | 0  | 1  | 3  |

Ajustamos um modelo logito de categoria com referência para investigar os efeitos do tamanho e do lago na escolha dos alimentos primários.

O peixe foi a escolha de alimento primário mais frequente, e consideramos como categoria de referência.

Estimamos os efeitos nas chances dos jacarés selecionarem outros tipos de alimentos primários em vez dos peixe.

Seja  $s = 1$  os jacarés com tamanho  $\leq 2,3$  metros e 0 caso contrário Além disso, sejam  $z^H$ ,  $z^O$ ,  $z^T$  e  $z^G$  variáveis indicadoras (dummies) para os lagos ( $z = 1$  para jacarés naquele lago e 0 caso contrário). O modelo com efeitos principais é

$$\log \left( \frac{\pi_{ij}}{\pi_{i1}} \right) = \beta_{j0} + \beta_{j1}s_i + \beta_{j2}z_i^O + \beta_{j3}z_i^T + \beta_{j4}z_i^G, \quad j = 2, 3, 4, 5.$$

No R, a função `vglm` no pacote **VGAM** pode ajustar este modelo.



```
library(VGAM)
fit <- vglm(formula = cbind(y2,y3,y4,y5,y1) ~ size + factor(lake),
            family=multinomial,
            data=dados) # fish=1 é categoria de referência
summary(fit)
```

Call:

```
vglm(formula = cbind(y2, y3, y4, y5, y1) ~ size + factor(lake),
     family = multinomial, data = dados)
```

Pearson residuals:

|   | $\log(\mu[,1]/\mu[,5])$ | $\log(\mu[,2]/\mu[,5])$ | $\log(\mu[,3]/\mu[,5])$ |
|---|-------------------------|-------------------------|-------------------------|
| 1 | 0.0953                  | 0.028205                | -0.54130                |
| 2 | -0.5082                 | 0.003228                | 0.66646                 |
| 3 | -0.3693                 | -0.461102               | -0.42005                |
| 4 | 0.4125                  | 0.249983                | 0.19772                 |
| 5 | -0.5526                 | -0.191149               | 0.07215                 |
| 6 | 0.6500                  | 0.110694                | -0.02784                |
| 7 | 0.6757                  | 0.827737                | 0.79863                 |
| 8 | -1.3051                 | -0.802694               | -0.69525                |

|   | $\log(\mu[,4]/\mu[,5])$ |
|---|-------------------------|
| 1 | -0.7268                 |
| 2 | 1.2589                  |
| 3 | 1.8347                  |
| 4 | -1.3779                 |
| 5 | 0.2790                  |
| 6 | -0.2828                 |
| 7 | -0.3081                 |
| 8 | 0.4629                  |

```

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept):1  -3.2074    0.6387  -5.021 5.13e-07 ***
(Intercept):2  -2.0718    0.7067  -2.931 0.003373 **
(Intercept):3  -1.3980    0.6085  -2.297 0.021601 *
(Intercept):4  -1.0781    0.4709  -2.289 0.022061 *
size:1          1.4582    0.3959   3.683 0.000231 ***
size:2         -0.3513    0.5800  -0.606 0.544786
size:3         -0.6307    0.6425  -0.982 0.326296
size:4          0.3316    0.4482   0.740 0.459506
factor(lake)2:1  2.5956    0.6597   3.934 8.34e-05 ***
factor(lake)2:2  1.2161    0.7860   1.547 0.121824
factor(lake)2:3 -1.3483    1.1635  -1.159 0.246529
factor(lake)2:4 -0.8205    0.7296  -1.125 0.260713
factor(lake)3:1  2.7803    0.6712   4.142 3.44e-05 ***
factor(lake)3:2  1.6925    0.7804   2.169 0.030113 *
factor(lake)3:3  0.3926    0.7818   0.502 0.615487
factor(lake)3:4  0.6902    0.5597   1.233 0.217511
factor(lake)4:1  1.6584    0.6129   2.706 0.006813 **
factor(lake)4:2 -1.2428    1.1854  -1.048 0.294466
factor(lake)4:3 -0.6951    0.7813  -0.890 0.373608
factor(lake)4:4 -0.8262    0.5575  -1.482 0.138378
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: log(mu[,1]/mu[,5]), log(mu[,2]/mu[,5]),
log(mu[,3]/mu[,5]), log(mu[,4]/mu[,5])

Residual deviance: 17.0798 on 12 degrees of freedom

Log-likelihood: -47.5138 on 12 degrees of freedom

Number of Fisher scoring iterations: 5

Warning: Hauck-Donner effect detected in the following estimate(s):
'(Intercept):1'

Reference group is level 5 of the response

```

Os dados são um pouco esparsos, mas o desvio de 17,08 ( $g.l. = 12$ ) não fornece muita evidência contra o modelo de efeitos principais.

O valor  $g.l.$  confirma que modelamos 32 probabilidades multinomiais (4 em cada combinação de tamanho e lago) usando 20 parâmetros (5 para cada logito).

A remoção do tamanho ou lago do modelo de efeitos principais resulta em um ajuste significativamente mais pobre: o desvio aumenta em 21,09 ( $g.l. = 4$ ) na remoção de tamanho e 49,13 ( $g.l. = 12$ ) na remoção de lago.

```
fit1<-vglm(formula=cbind(y2,y3,y4,y5,y1) ~ size,family=multinomial,data=dados)
fit2<-vglm(formula=cbind(y2,y3,y4,y5,y1)~factor(lake),family=multinomial,data=dados)
rbind(cbind(deviance(fit),df.residual(fit)),cbind(deviance(fit1),df.residual(fit1)),
      cbind(deviance(fit2),df.residual(fit2)))
```

|      | [,1]     | [,2] |
|------|----------|------|
| [1,] | 17.07983 | 12   |
| [2,] | 66.21291 | 24   |
| [3,] | 38.16724 | 16   |

A equação para predição das log chances de selecionar invertebrados em vez de peixes é

$$\log \left( \frac{\widehat{\pi_{i2}}}{\widehat{\pi_{i1}}} \right) = -3,207 + 1,458s_i + 2,596z_i^O + 2,780z_i^T + 1,658z_i^G.$$

Para um determinado lago, para pequenos jacarés, a chance estimada de que a escolha de alimentos primários é invertebrados, em vez de peixes, é  $\exp(1,458) = 4,30$  vezes a chance estimada para jacarés grandes.

O efeito estimado é impreciso, pois o intervalo de confiança de Wald de 95% é dado por

$$\exp[1,458 \pm 1,96 * (0,396)] = (1,98, 9,34).$$

Os efeitos do lago indicam que as chances estimadas de que a escolha de alimentos primários é invertebrados em vez de peixes são relativamente maiores nos lagos Ocklawaha, Trafford e George do que no lago Hancock.

As estimativas dos parâmetros do modelo geram as probabilidades ajustadas. Por exemplo, a probabilidade estimada de que um jacaré grande no lago George tenha invertebrados como a principal opção alimentar é

$$\pi_{i2} = \frac{e^{-3,207+1,658}}{1 + e^{-3,207+1,658} + e^{-2,072+1,243} + e^{-1,398-0,695} + e^{-1,078-0,826}} = 0,14.$$

As probabilidades estimadas de (invertebrados, répteis, pássaros, outros, peixes) para grandes jacarés naquele lago são (0,14; 0,02; 0,08; 0,10; 0,66).

```
fitted(fit)
```

|   | y2         | y3         | y4          | y5         | y1        |
|---|------------|------------|-------------|------------|-----------|
| 1 | 0.09309880 | 0.04745657 | 0.070401523 | 0.25373963 | 0.5353035 |
| 2 | 0.02307168 | 0.07182461 | 0.140896287 | 0.19400964 | 0.5701978 |
| 3 | 0.60189675 | 0.07722761 | 0.008817482 | 0.05387208 | 0.2581861 |
| 4 | 0.24864518 | 0.19483742 | 0.029416085 | 0.06866281 | 0.4584385 |
| 5 | 0.51683852 | 0.08876722 | 0.035894709 | 0.17420051 | 0.1842990 |
| 6 | 0.19296122 | 0.20239954 | 0.108225068 | 0.20066164 | 0.2957525 |
| 7 | 0.41285579 | 0.01156654 | 0.029671169 | 0.09380245 | 0.4521040 |
| 8 | 0.13967784 | 0.02389871 | 0.081067366 | 0.09791362 | 0.6574425 |

## Respostas Ordinais

## Respostas Ordinais: Modelos logístico e probito cumulativo

Para variáveis de resposta ordinal, os modelos têm termos que refletem características ordinais, como uma tendência monótona, em que as respostas tendem a cair em categorias mais altas (ou mais baixas) à medida que o valor de uma variável explicativa aumenta.

Tais modelos são mais parcimoniosos do que modelos para respostas nominais, porque potencialmente eles têm muito menos parâmetros.

Veremos os modelos logístico e probito para respostas ordinais.

## Modelo logito cumulativo: chances proporcionais

Seja  $Y_i$  a categoria da resposta do  $i$ -ésimo sujeito. Ou seja,  $Y_i = j$  significa que  $Y_{ij} = 1$  e  $Y_{ik} = 0$  para  $k \neq j$ , para  $c$  multinomiais indicadores.

Para usar a ordenação por categoria, podemos expressar o modelo em termos das probabilidades cumulativas,

$$\mathbb{P}(Y_i \leq j) = \pi_{i1} + \dots + \pi_{ij}, \quad j = 1, \dots, c.$$

Os *logitos cumulativos* são os logitos dessas probabilidades cumulativas,

$$\begin{aligned} \text{logito}[\mathbb{P}(Y_i \leq j)] &= \log \left( \frac{\mathbb{P}(Y_i \leq j)}{1 - \mathbb{P}(Y_i \leq j)} \right) \\ &= \log \left( \frac{\pi_{i1} + \dots + \pi_{ij}}{\pi_{i,j+1} + \dots + \pi_{ic}} \right), \quad j = 1, \dots, c-1. \end{aligned}$$

Cada logito cumulativo usa todas as  $c$  categorias de resposta.



Um modelo para  $\text{logito}[\mathbb{P}(Y_i \leq j)]$  sozinho é um modelo logístico comum para resposta binária, onde categorias de 1 a  $j$  representam “sucesso” e categorias de  $j + 1$  a  $c$  representam “fracassos”.

Um modelo parcimonioso que usa simultaneamente todos os  $(c - 1)$  logits é dado a seguir.

**Modelo logito cumulativo:**

$$\text{logito}[\mathbb{P}(Y_i \leq j)] = \alpha_j + \mathbf{x}_i \beta, \quad j = 1, \dots, c - 1.$$

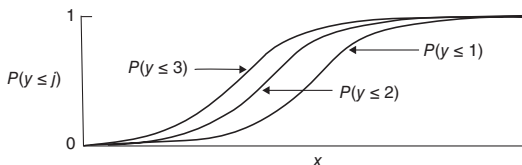
Cada logito cumulativo tem seu próprio intercepto.

Os  $\{\alpha_j\}$  estão aumentando em  $j$ , uma vez que  $\mathbb{P}(Y_i \leq j)$  aumenta em  $j$  para qualquer  $\mathbf{x}_i$  fixo, e o logito é uma função crescente de  $\mathbb{P}(Y_i \leq j)$ .

É usado uma notação separada para o intercepto  $(\alpha_j)$  no preditor linear, pois eles dependem de  $j$ , mas os outros efeitos não.

Este modelo afirma que os efeitos ( $\beta$ ) das variáveis explicativas são os mesmos para cada logito cumulativo.

Para uma única variável explicativa contínua  $x$ , a Figura a seguir representa o modelo quando temos 4 categorias ( $c = 4$ ). As curvas para  $j = 1, 2$  e 3 têm exatamente a mesma forma e não se cruzam.



**Figure 6.1** Cumulative logit model with the same effect of  $x$  on each of three cumulative probabilities, for an ordinal response variable with  $c = 4$  categories.

Figura ilustrativa do livro texto (Agresti, A. (2015)).

Este modelo trata a variável de resposta como escala ordinal no seguinte sentido:

- Se o modelo se mantém e a ordem das categorias das respostas é invertida, o modelo ainda se mantém com uma mudança no sinal dos  $\beta$ ; no entanto, o modelo não precisa se sustentar se as categorias das respostas forem permutadas de qualquer outra maneira.

O modelo logito cumulativo satisfaz

$$\begin{aligned} \text{logito}[\mathbb{P}(Y_j \leq j | \mathbf{x}_i = \mathbf{u})] &- \text{logito}[\mathbb{P}(Y_j \leq j | \mathbf{x}_i = \mathbf{v})] \\ &= \log \left( \frac{\mathbb{P}(Y_j \leq j | \mathbf{x}_i = \mathbf{u}) / \mathbb{P}(Y_j > j | \mathbf{x}_i = \mathbf{u})}{\mathbb{P}(Y_j \leq j | \mathbf{x}_i = \mathbf{v}) / \mathbb{P}(Y_j > j | \mathbf{x}_i = \mathbf{v})} \right) \\ &= (\mathbf{u} - \mathbf{v})\beta. \end{aligned}$$

As chances de que as respostas  $\leq j$  em  $\mathbf{x}_i = \mathbf{u}$  são  $\exp[(\mathbf{u} - \mathbf{v})\boldsymbol{\beta}]$  vezes as chances em  $\mathbf{x}_i = \mathbf{v}$ .

Uma razão de chances de probabilidades cumulativas é chamada de *razão de chances cumulativa*.

O log da razão de chances cumulativa é proporcional à distância entre  $\mathbf{u}$  e  $\mathbf{v}$ .

Para cada  $j$ , a chance de que  $Y_i \leq j$  é  $\exp(\beta_k)$  para o aumento de uma unidade em  $x_{ik}$ , fixando as outras variáveis explicativas.

A mesma constante de proporcionalidade se aplica a todos os  $c - 1$  logitos cumulativos; isto é, o efeito é  $\beta_k$ , não  $\beta_{jk}$ . Esta propriedade de um efeito comum para todas as probabilidades cumulativas é denominada *chances proporcionais*.

## Modelo probito cumulativo

Como nos MLGs binários, outras funções de ligação são possíveis para as probabilidades cumulativas. Seja  $G^{-1}$  uma função de ligação que é a inversa da f.d.a contínua  $G$ . O modelo de ligação cumulativo é dado a seguir

**Modelo de ligação cumulativo:**

$$G^{-1}[\mathbb{P}(Y_i \leq j)] = \alpha_j + \mathbf{x}_i\beta, \quad j = 1, \dots, c-1.$$

Assim como no modelo logito cumulativo, cada ligação da probabilidade cumulativa tem seu próprio intercepto.

O modelo de probito cumulativo é o modelo de ligação cumulativo que usa a f.d.a. de uma Normal padrão ( $\Phi(\cdot)$ ) para  $G$ .

**Modelo de probito cumulativo:**

$$\Phi^{-1}[\mathbb{P}(Y_i \leq j)] = \alpha_j + \mathbf{x}_i\beta, \quad j = 1, \dots, c-1.$$

## Ajustando um modelo de ligação cumulativo

Seja  $Y_i$  a categoria da resposta do  $i$ -ésimo sujeito. Ou seja,  $Y_i = j$  significa que  $Y_{ij} = 1$  e  $Y_{ik} = 0$  para  $k \neq j$ , para  $c$  multinomiais indicadores, a função de verossimilhança multinomial para modelo de ligação cumulativo

$$G^{-1}[\mathbb{P}(Y_i \leq j)] = \alpha_j + \mathbf{x}_i \beta$$

$$\prod_{i=1}^N \left( \prod_{j=1}^c \pi_{ij}^{y_{ij}} \right) = \prod_{i=1}^N \left( \prod_{j=1}^c [\mathbb{P}(Y_i \leq j) - \mathbb{P}(Y_i \leq j-1)]^{y_{ij}} \right)$$

vista como função de  $(\{\alpha_j\}, \beta)$ , onde  $\mathbb{P}(Y_i \leq 0) = 0$ . E, a função de log-verossimilhança é

$$\ell(\alpha, \beta; \mathbf{y}_i) = \sum_{i=1}^N \sum_{j=1}^c y_{ij} \log [G(\alpha_j + \mathbf{x}_i \beta) - G(\alpha_{j-1} + \mathbf{x}_i \beta)].$$

Seja  $g$  a derivada de  $G$ , ou seja, o f.d.p. correspondente à f.d.a.  $G$ , e seja  $\delta_{jk}$  uma função idicadora, ou seja,  $\delta_{jk} = 1$  se  $j = k$  e  $\delta_{jk} = 0$  caso contrário. Então as equações de log-verossimilhança são dadas por

$$\frac{\partial \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{y})}{\partial \beta_k} = \sum_{i=1}^N \sum_{j=1}^c y_{ij} x_{ik} \frac{g(\alpha_j + \mathbf{x}_i \boldsymbol{\beta}) - g(\alpha_{j-1} + \mathbf{x}_i \boldsymbol{\beta})}{G(\alpha_j + \mathbf{x}_i \boldsymbol{\beta}) - G(\alpha_{j-1} + \mathbf{x}_i \boldsymbol{\beta})} = 0,$$

e

$$\frac{\partial \ell(\boldsymbol{\beta}; \mathbf{y})}{\partial \alpha_k} = \sum_{i=1}^N \sum_{j=1}^c y_{ij} \frac{\delta_{jk} g(\alpha_j + \mathbf{x}_i \boldsymbol{\beta}) - \delta_{j-1,k} g(\alpha_{j-1} + \mathbf{x}_i \boldsymbol{\beta})}{G(\alpha_j + \mathbf{x}_i \boldsymbol{\beta}) - G(\alpha_{j-1} + \mathbf{x}_i \boldsymbol{\beta})} = 0.$$

A matriz hessiana é bastante extensa, uma vez que  $\boldsymbol{\alpha}$  e  $\boldsymbol{\beta}$  não são ortogonais.

As equações de log-verossimilhanças podem ser resolvidas usando Escore de Fisher ou o método de Newton-Raphson. Os valores de SE diferem um pouco para os dois métodos, uma vez as matrizes de informações esperadas e observadas não são as mesmas para este modelo.

## Exemplo: Mental Impairment (Comprometimento mental)

Os dados da Tabela a seguir são baseados em um estudo de saúde mental para uma amostra aleatória de residentes adultos de Alachua County, Flórida.

O comprometimento mental é uma variável ordinal, com categorias:

- = 1, se bom,
- = 2, se formação suave de sintomas,
- = 3, se formação moderada de sintomas,
- = 4, se comprometido.

O estudo relacionou o comprometimento mental ( $Y = \text{impair}$ ) a várias variáveis explicativas, duas das quais usaremos aqui.



O índice de eventos da vida ( $x_1 = \text{life}$ ) é uma medida composta do número e gravidade de eventos importantes da vida que ocorreram ao sujeito nos últimos três anos, como o nascimento de um filho, um novo emprego, um divórcio ou uma morte na família.

Nesta amostra,  $x_1$  tem uma média de 4,3 e desvio padrão de 2,7.

O status socioeconômico ( $x_2 = \text{ses}$ ) é medido como uma variável binária

■ = 1, se alto; e

■ = 0, se baixo.

Dados:

| Ind. | Compr.<br>mental | $x_1$ | $x_2$ | Ind. | Compr.<br>mental | $x_1$ | $x_2$ |
|------|------------------|-------|-------|------|------------------|-------|-------|
| 1    | Bom              | 1     | 1     | 21   | Suave            | 1     | 9     |
| 2    | Bom              | 1     | 9     | 22   | Suave            | 0     | 3     |
| 3    | Bom              | 1     | 0     | 23   | Suave            | 1     | 3     |
| 4    | Bom              | 1     | 4     | 24   | Suave            | 1     | 1     |
| 5    | Bom              | 1     | 3     | 25   | Moderado         | 0     | 0     |
| 6    | Bom              | 0     | 2     | 26   | Moderado         | 1     | 4     |
| 7    | Bom              | 0     | 1     | 27   | Moderado         | 0     | 3     |
| 8    | Bom              | 1     | 3     | 28   | Moderado         | 0     | 9     |
| 9    | Bom              | 1     | 3     | 29   | Moderado         | 1     | 6     |
| 10   | Bom              | 1     | 7     | 30   | Moderado         | 0     | 4     |
| 11   | Bom              | 0     | 1     | 31   | Moderado         | 0     | 3     |
| 12   | Bom              | 0     | 2     | 32   | Comprometido     | 1     | 8     |
| 13   | Suave            | 1     | 5     | 33   | Comprometido     | 1     | 2     |
| 14   | Suave            | 0     | 6     | 34   | Comprometido     | 1     | 7     |
| 15   | Suave            | 1     | 3     | 35   | Comprometido     | 0     | 5     |
| 16   | Suave            | 0     | 1     | 36   | Comprometido     | 0     | 4     |
| 17   | Suave            | 1     | 8     | 37   | Comprometido     | 0     | 4     |
| 18   | Suave            | 1     | 2     | 38   | Comprometido     | 1     | 8     |
| 19   | Suave            | 0     | 5     | 39   | Comprometido     | 0     | 8     |
| 20   | Suave            | 1     | 5     | 40   | Comprometido     | 0     | 9     |

```
# Arquivo Metal.dat em www.stat.ufl.edu/~aa/glm/data  
dados = read.table("Mental.txt", header=T)  
head(dados)
```

|   | impair | ses | life |
|---|--------|-----|------|
| 1 | 1      | 1   | 1    |
| 2 | 1      | 1   | 9    |
| 3 | 1      | 1   | 0    |
| 4 | 1      | 1   | 4    |
| 5 | 1      | 1   | 3    |
| 6 | 1      | 0   | 2    |

```
fit <- vglm(impair ~ life + ses, family=cumulative(parallel=TRUE), data=dados)
# parallel=TRUE impõe estrutura proporcional de chances
summary(fit)
```

Call:

```
vglm(formula = impair ~ life + ses, family = cumulative(parallel = TRUE),
     data = dados)
```

Pearson residuals:

|                    | Min    | 1Q      | Median  | 3Q     | Max   |
|--------------------|--------|---------|---------|--------|-------|
| logitlink(P[Y<=1]) | -1.568 | -0.7048 | -0.2102 | 0.8070 | 2.713 |
| logitlink(P[Y<=2]) | -2.328 | -0.4666 | 0.2657  | 0.6904 | 1.615 |
| logitlink(P[Y<=3]) | -3.688 | 0.1198  | 0.2039  | 0.4194 | 1.892 |

Coefficients:

|               | Estimate | Std. Error | z value | Pr(> z )   |
|---------------|----------|------------|---------|------------|
| (Intercept):1 | -0.2819  | 0.6231     | -0.452  | 0.65096    |
| (Intercept):2 | 1.2128   | 0.6511     | 1.863   | 0.06251 .  |
| (Intercept):3 | 2.2094   | 0.7171     | 3.081   | 0.00206 ** |
| life          | -0.3189  | 0.1194     | -2.670  | 0.00759 ** |
| ses           | 1.1112   | 0.6143     | 1.809   | 0.07045 .  |

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2]),  
logitlink(P[Y<=3])

Residual deviance: 99.0979 on 115 degrees of freedom

Log-likelihood: -49.5489 on 115 degrees of freedom

Number of Fisher scoring iterations: 5

No Hauck-Donner effect found in any of the estimates

Exponentiated coefficients:

|  | life      | ses       |
|--|-----------|-----------|
|  | 0.7269742 | 3.0380707 |

O modelo logito cumulativo na forma de chances proporcionais com efeitos principais tem ajuste de ML dado por

$$\text{logito}[\hat{\mathbb{P}}(Y_i \leq j)] = \widehat{\alpha}_j - 0,319 x_{i1} + 1,111 x_{i2}.$$

A probabilidade cumulativa estimada, iniciando em “bom” e terminando em deficiência mental na escala de comprometimento, diminui à medida que os eventos da vida aumentam e é maior no nível mais alto do status socioeconômico.

Dada uma observação no eventos de vida, no nível alto do status socioeconômico, a chance estimada de comprometimento mental abaixo de qualquer nível fixo é  $e^{1.111} = 3,0$  vezes as chances estimadas no nível baixo do status socioeconômico.

O intervalo de confiança de 95% de Wald para esse efeito é  $\exp[1,111 \pm 1,96 * (0,614)] = (0,91, 10,12)$ . Um efeito de status socioeconômico nulo é plausível, mas o efeito status socioeconômico também pode ser muito forte.

Para nos ajudar a interpretar os efeitos, podemos estimar as probabilidades das categorias da resposta.

Primeiro, considere o efeito do status socioeconômico. Nos eventos da vida médio de 4,3,

- $\hat{P}(y = 1) = 0,37$  no **ses** alto (isto é,  $x_2 = 1$ ;) e
- $\hat{P}(y = 1) = 0,16$  no **ses** baixo ( $x_2 = 0$ ).

Em seguida, considere o efeito de eventos da vida. Para o status socioeconômico alto,

- $\hat{P}(y = 1)$  muda de 0,70 para 0,12 entre o mínimo da amostra de 0 e o máximo de 9 eventos de vida.

Para o status socioeconômico baixo,

- $\hat{P}(y = 1)$  muda de 0,43 para 0,04.

Comparar 0,70 a 0,43 no mínimo dos eventos de vida e 0,12 a 0,04 no máximo fornece uma descrição adicional do efeito status socioeconômico.

O efeito da amostra é substancial para cada preditor. A saída a seguir mostra as probabilidades estimadas da categoria de resposta para alguns sujeitos da amostra. Indivíduos (como o sujeito 40) com **ses** baixo e **life** relativamente alto têm uma probabilidade estimada relativamente alta de serem deficientes mentais.

```
fitted(fit)[c(1,2,3,4,37,38,39,40),]
```

|    | 1          | 2         | 3          | 4          |
|----|------------|-----------|------------|------------|
| 1  | 0.62491696 | 0.2564211 | 0.07131461 | 0.04734732 |
| 2  | 0.11502166 | 0.2518290 | 0.24398492 | 0.38916438 |
| 3  | 0.69621443 | 0.2146331 | 0.05428144 | 0.03487106 |
| 4  | 0.39028333 | 0.3502169 | 0.14495617 | 0.11454363 |
| 37 | 0.17402808 | 0.3103139 | 0.23352889 | 0.28212916 |
| 38 | 0.15166796 | 0.2918524 | 0.23993358 | 0.31654610 |
| 39 | 0.05557719 | 0.1522439 | 0.20761621 | 0.58456270 |
| 40 | 0.04102571 | 0.1191429 | 0.18048190 | 0.65934948 |

```
fit2 <- vglm(impair~life+ses+life:ses,family=cumulative(parallel=TRUE),data=dados)
# modelo com interação
summary(fit2)
```

Call:

```
vglm(formula = impair ~ life + ses + life:ses, family = cumulative(parallel = TRUE),
     data = dados)
```

Pearson residuals:

|                    | Min    | 1Q      | Median  | 3Q     | Max   |
|--------------------|--------|---------|---------|--------|-------|
| logitlink(P[Y<=1]) | -1.393 | -0.7139 | -0.2172 | 0.9084 | 2.262 |
| logitlink(P[Y<=2]) | -2.758 | -0.4862 | 0.2781  | 0.7218 | 1.797 |
| logitlink(P[Y<=3]) | -3.364 | 0.1347  | 0.2062  | 0.3795 | 2.344 |

Coefficients:

|               | Estimate | Std. Error | z value | Pr(> z )   |
|---------------|----------|------------|---------|------------|
| (Intercept):1 | 0.09807  | 0.81102    | 0.121   | 0.90375    |
| (Intercept):2 | 1.59248  | 0.83717    | 1.902   | 0.05714 .  |
| (Intercept):3 | 2.60660  | 0.90966    | 2.865   | 0.00416 ** |
| life          | -0.42045 | 0.19031    | -2.209  | 0.02715 *  |
| ses           | 0.37090  | 1.13022    | 0.328   | 0.74279    |
| life:ses      | 0.18131  | 0.23611    | 0.768   | 0.44255    |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2]),  
logitlink(P[Y<=3])

Residual deviance: 98.5044 on 114 degrees of freedom

Log-likelihood: -49.2522 on 114 degrees of freedom

Number of Fisher scoring iterations: 5

No Hauck-Donner effect found in any of the estimates

Exponentiated coefficients:

|  | life      | ses       | life:ses  |
|--|-----------|-----------|-----------|
|  | 0.6567529 | 1.4490350 | 1.1987822 |



Quando adicionamos um termo de interação, isso sugere que o efeito dos eventos de vida pode ser mais fraco em um status socioeconômico mais alto.

No entanto, o valor não é significativamente melhor (p-valor = 0,44).

```
# p-valor para o teste da razão de verossimilhança comparando o modelo co, interação  
1 - pchisq(2*(logLik(fit2)-logLik(fit)),df=df.residual(fit)-df.residual(fit2))
```

```
[1] 0.4410848
```

- Paula, G.A. (2013). Modelos de Regressão com Apoio Computacional.
- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Wiley series in probability and statistics.
- Faraway, J. J. (2006). *Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman and Hall/CRC.
- Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2.  
<https://CRAN.R-project.org/package=stargazer>