

ME720 - Modelos Lineares Generalizados

Parte 14 - Modelos para Dados de Contagem - Binomial Negativa

Profa. **Larissa Avila Matos**

MLGs Binomiais Negativos

Na distribuição de Poisson temos que a variância é igual à média.

Na prática, as observações de contagem geralmente exibem variabilidade superior à predita pela Poisson.

Esse fenômeno é chamado de sobredispersão (*overdispersion*).

Sobredispersão para um MLGs de Poisson

Um motivo comum para sobredispersão é a heterogeneidade: em níveis fixos das variáveis explicativas, a média varia de acordo com os valores de variáveis não observadas.

Outra limitação grave dos MLGs de Poisson é que, como a variação de Y deve ser igual à média, em uma média fixa, a variação não pode diminuir à medida que variáveis explicativas adicionais entram no modelo.

A sobredispersão não é um problema em modelos lineares que assumem Y normalmente distribuído, pois essa distribuição possui um parâmetro de variação separado para descrever a variabilidade.

Para distribuições de Poisson e binomial, no entanto, a variação é uma função da média. Sobredispersão é comum na modelagem de contagens.

Suponha que o modelo para a média tenha a correta função de ligação e o correto preditor linear, mas a distribuição verdadeira da resposta tenha mais variabilidade do que a Poisson.

Então os estimadores de ML dos parâmetros do modelo, assumindo uma resposta de Poisson, ainda são consistentes, convergindo em probabilidade para os valores dos parâmetros, mas os erros padrão são muito pequenos.

Extensões do MLG de Poisson que possuem um parâmetro extra são melhores para sobredispersão (veremos mais adiante).

Binomial Negativa

Um modelo de mistura é uma maneira flexível de explicar a sobredispersão.

Em uma configuração fixa das variáveis explicativas realmente observadas, dada a média λ , suponha que a distribuição de Y seja $\text{Poisson}(\lambda)$, mas λ em si varia por causa de covariáveis não medidas.

Seja $\mu = \mathbb{E}(\lambda)$.

Então incondicionalmente,

$$\mathbb{E}(Y) = \mathbb{E}[\mathbb{E}(Y|\lambda)] = \mathbb{E}(\lambda) = \mu,$$

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|\lambda)] + \text{Var}[\mathbb{E}(Y|\lambda)] = \mathbb{E}(\lambda) + \text{Var}(\lambda) = \mu + \text{Var}(\lambda) > \mu.$$

Este é um exemplo importante de um modelo de mistura para dados de contagem.

Suponha que

1 dado λ , Y tem distribuição Poisson(λ), e

2 λ tem distribuição Gama.

Temos que,

$$f(y|\lambda) = \frac{e^{-\lambda} \lambda^y}{y!}.$$

Além disso, lembre-se de que a distribuição Gama possui $\mathbb{E}(\lambda) = \mu$ e $\text{Var}(\lambda) = \frac{\mu^2}{k}$ para um parâmetro de forma $k > 0$, portanto o desvio padrão é proporcional à média, e

$$f(\lambda; \mu, k) = \frac{\left(\frac{k}{\mu}\right)^k}{\Gamma(k)} e^{-\frac{k\lambda}{\mu}} \lambda^{k-1}, \quad \lambda \geq 0.$$

Marginalmente, essa mistura produz a distribuição binomial negativa para Y .

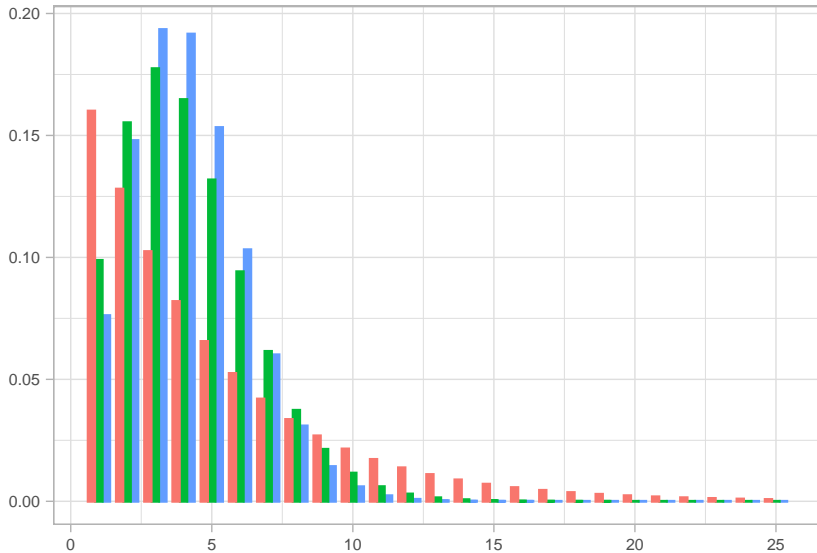
Sua função de probabilidade é dada por

$$f(y; \mu, k) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \left(\frac{\mu}{\mu+k} \right)^y \left(\frac{k}{\mu+k} \right)^k, \quad y = 0, 1, 2, \dots$$

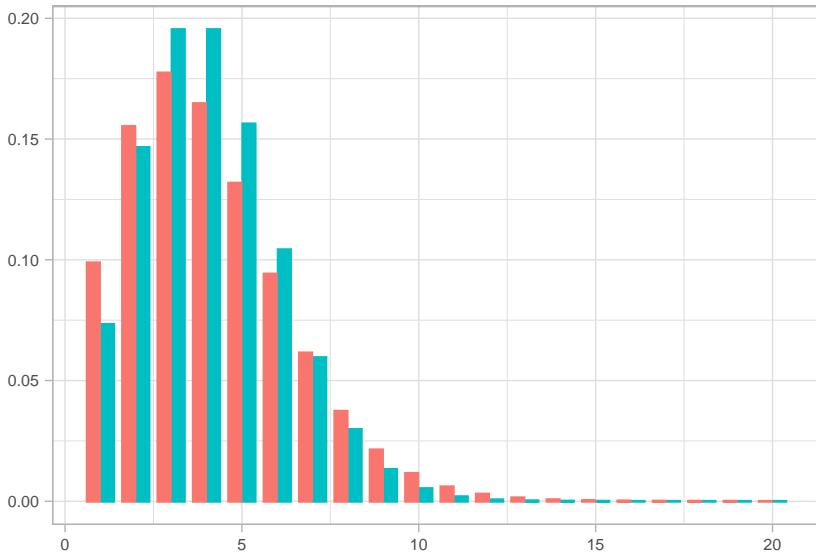
Quando k é fixo, esta distribuição é membro da família exponencial com parâmetro de escala apropriada para variáveis discretas, com parâmetro natural $\log\left(\frac{\mu}{\mu+k}\right)$,

$$\mathbb{E}(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \frac{\mu(\mu+k)}{k} = \frac{\mu^2}{k} + \mu.$$

BN BinNeg(4,1) BinNeg(4,10) BinNeg(4,100)



Distr ■ BinNeg(4,10) ■ Poisson(4)



Podemos, similarmente, supor que

1 $Y|\lambda \sim \text{Poisson}(\lambda)$, e

2 λ tem distribuição Gama;

onde $\mathbb{E}(\lambda) = \mu$ e $\text{Var}(\lambda) = \mu^2\gamma$, ou seja, $\gamma = \frac{1}{k}$.

Então, temos que

$$\mathbb{E}(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \mu + \gamma\mu^2.$$

O índice $\gamma > 0$ é um tipo de parâmetro de dispersão. Quanto maior o valor de γ , maior a sobredispersão em relação a Poisson.

Como $\gamma \rightarrow 0$, $\text{Var}(Y) \rightarrow \mu$ e a distribuição binomial negativa converge para o Poisson.

A distribuição binomial negativa tem uma extensão muito maior que a Poisson.

Por exemplo, a moda da Poisson é a parte inteira da média e é igual a 0 somente quando $\mu < 1$.

A binomial negativa também é unimodal, mas a moda é 0 quando $\gamma \geq 1$ e, caso contrário, é a parte inteira de $\mu(1 - \gamma)$. A moda pode ser 0 para qualquer μ .

Para essa parametrização temos que

$$\begin{aligned}f(y; \mu, \gamma) &= \frac{\Gamma(y + \frac{1}{\gamma})}{\Gamma(\frac{1}{\gamma})\Gamma(y + 1)} \left(\frac{\mu}{\mu + \frac{1}{\gamma}} \right)^y \left(\frac{\frac{1}{\gamma}}{\mu + \frac{1}{\gamma}} \right)^{\frac{1}{\gamma}}, \quad y = 0, 1, 2, \dots \\&= \frac{\Gamma(y + \frac{1}{\gamma})}{\Gamma(\frac{1}{\gamma})\Gamma(y + 1)} \left(\frac{\mu\gamma}{1 + \mu\gamma} \right)^y (1 + \mu\gamma)^{-\frac{1}{\gamma}}.\end{aligned}$$

Então,

$$f(y; \mu, \gamma) = \exp \left[\log \left(\frac{\Gamma(y + \frac{1}{\gamma})}{\Gamma(\frac{1}{\gamma})\Gamma(y + 1)} \right) + y \log \left(\frac{\mu\gamma}{1 + \mu\gamma} \right) - \frac{1}{\gamma} \log(1 + \mu\gamma) \right].$$

Modelos de regressão com resposta binomial negativa

Vamos supor então que Y_1, \dots, Y_n são variáveis aleatórias independentes tais que $Y_i \sim \text{BN}(\gamma^{-1}, \mu_i)$.

A função de probabilidades de Y_i fica dada por

$$f(y_i; \mu_i, \gamma) = \exp \left[\log \left(\frac{\Gamma(y_i + \frac{1}{\gamma})}{\Gamma(\frac{1}{\gamma})\Gamma(y_i + 1)} \right) + y_i \log \left(\frac{\mu_i \gamma}{1 + \mu_i \gamma} \right) - \frac{1}{\gamma} \log (1 + \mu_i \gamma) \right],$$

onde $\mathbb{E}(Y_i) = \mu_i$ e $\text{Var}(Y_i) = \mu_i + \gamma \mu_i^2$.

Similarmente aos MLGs assumimos a parte sistemática dada por $g(\mu_i) = \eta_i = \mathbf{x}_i' \boldsymbol{\beta}$, em que $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ contém valores das variáveis explicativas, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ é um vetor de parâmetros desconhecidos e $g(\cdot)$ é a função de ligação.

Como nos modelos de Poisson as ligações mais utilizados são:

- logarítmica: $g(\mu_i) = \log(\mu_i)$,
- raiz quadrada: $g(\mu_i) = \sqrt{\mu_i}$, e
- identidade: $g(\mu_i) = \mu_i$.

MLGs binomiais negativos geralmente usam a função de ligação log, como nos modelos lineares de Poisson, em vez da função de ligação canônica (por motivos de interpretação).

OBS.:

Por simplicidade, deixamos que o parâmetro de dispersão γ seja a mesma constante para todas as n observações, mas iremos tratar como desconhecido, bem como a variação nos modelos normais.

Definindo $\boldsymbol{\theta} = (\boldsymbol{\beta}', \gamma)'$ o log da função de verossimilhança fica dado por

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^n \left[\log \left(\frac{\Gamma(y_i + \frac{1}{\gamma})}{\Gamma(\frac{1}{\gamma})\Gamma(y_i + 1)} \right) \right] + \sum_{i=1}^n \left[y_i \log \left(\frac{\mu_i \gamma}{1 + \mu_i \gamma} \right) - \frac{1}{\gamma} \log(1 + \mu_i \gamma) \right].$$

Reescrevendo,

$$\begin{aligned} \ell(\boldsymbol{\theta}; \mathbf{y}) &= \sum_{i=1}^n \left[\log \left(\Gamma \left(y_i + \frac{1}{\gamma} \right) \right) - \log \left(\Gamma \left(\frac{1}{\gamma} \right) \right) - \log(\Gamma(y_i + 1)) \right] \\ &+ \sum_{i=1}^n \left[y_i \log \left(\frac{\mu_i \gamma}{1 + \mu_i \gamma} \right) - \frac{1}{\gamma} \log(1 + \mu_i \gamma) \right]. \end{aligned}$$

Lembrando que μ_i é uma função de $\boldsymbol{\beta}$ através de $\eta_i = g(\mu_i) = \sum_j \beta_j x_{ij}$, com a função de ligação g .

As equações de log-verossimilhança para β , $\frac{\partial \ell(\theta; \mathbf{y})}{\partial \beta} = 0$, são

$$\begin{aligned} \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{\text{Var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) &= \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{\mu_i + \gamma \mu_i^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \\ &= \sum_{i=1}^n w_i d_i^{-1} (y_i - \mu_i) x_{ij} = 0, \quad j = 1, 2, \dots, p, \end{aligned}$$

em que $w_i = \frac{(\partial \mu_i / \partial \eta_i)^2}{\mu_i + \gamma \mu_i^2}$ e $d_i = \frac{\partial \mu_i}{\partial \eta_i}$.

Logo, podemos expressar as equações na forma matricial

$$\mathbf{X}' \mathbf{W} \mathbf{D}^{-1} (\mathbf{y} - \boldsymbol{\mu}),$$

onde \mathbf{X} é a matriz de planejamento do modelo com linhas \mathbf{x}'_i , $i = 1, \dots, n$, $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$, $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$, $\mathbf{y} = (y_1, \dots, y_n)'$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$.

De forma similar a equação de log-verossimilhança para γ , $\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y})}{\partial \gamma} = 0$, fica dada por

$$\sum_{i=1}^n \left[\psi \left(y_i + \frac{1}{\gamma} \right) - \psi \left(\frac{1}{\gamma} \right) - \frac{\left(y_i + \frac{1}{\gamma} \right)}{\left(\mu_i + \frac{1}{\gamma} \right)} + \frac{1}{\gamma^2} \log \left(\frac{\mu_i + \frac{1}{\gamma}}{\frac{1}{\gamma}} \right) + 1 \right] = 0,$$

em que $\psi(\cdot)$ é a função digama.

Para obtermos a matriz de informação precisamos calcular as segundas derivadas parciais, $\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$.

Primeiro calculamos,

$$\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y})}{\partial \beta_j \partial \gamma} = - \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{(1 + \gamma \mu_i)^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right),$$

cujos valores esperados ficam dados por

$$\mathbb{E} \left(\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y})}{\partial \beta_j \partial \gamma} \right) = 0, \quad \text{para cada } j,$$

uma vez que $\mathbb{E}(Y_i) = \mu_i$.

Então, β e γ são parâmetros ortogonais.

Assim, a matriz de informação de Fisher para $\boldsymbol{\theta} = (\boldsymbol{\beta}, \gamma)$ assume a forma bloco diagonal,

$$\mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\gamma\gamma} \end{bmatrix}.$$

Portanto, $\widehat{\boldsymbol{\beta}}$ e $\widehat{\gamma}$ são assintoticamente independentes, e para amostras grandes o erro padrão de $\widehat{\beta}_j$ é o mesmo independente se γ é conhecido ou não.

Então a matriz de covariância de $\widehat{\boldsymbol{\beta}}$ é

$$\text{Var}(\widehat{\boldsymbol{\beta}}) = \left(\mathbf{X} \widehat{\mathbf{W}} \mathbf{X} \right)^{-1},$$

onde $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$, com $w_i = \frac{(\partial \mu_i / \partial \eta_i)^2}{\text{Var}(Y_i)}$.

As estimativas de máxima verossimilhança para β e γ podem ser obtidas através de um algoritmo de mínimos quadrados ponderados, aplicando o método escore de Fisher, a partir das equações de log-verossimilhança para β e do método de Newton-Raphson para obter $\hat{\gamma}$ desenvolvido a partir das equações de log-verossimilhança para γ , os quais são descritos abaixo,

$$\beta^{(t+1)} = (\mathbf{X}'\mathbf{W}^{(t)}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}^{(t)}\mathbf{y}^{*(t)}$$

$$\gamma^{(t+1)} = \gamma^{(t)} - (\ddot{\ell}_{\gamma\gamma}^{(t)})^{-1} \mathbf{u}_{\gamma}^{(t)},$$

para $t = 0, 1, 2, \dots$, em que

$$\mathbf{y}^* = \mathbf{X}\beta + \mathbf{D}^{-1}(\mathbf{y} - \boldsymbol{\mu});$$

e $\mathbf{u}_{\gamma} = \frac{\partial \ell(\beta, \gamma; \mathbf{y})}{\partial \gamma}$ e $\ddot{\ell}_{\gamma\gamma} = \frac{\partial^2 \ell(\beta, \gamma; \mathbf{y})}{\partial \gamma \partial \gamma}$ (para mais detalhes ver Paula, G.A. (2013)).

Desvio MLGs binomiais negativos

O desvio para MLGs binomiais negativos é

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - \left(y_i + \frac{1}{\hat{\gamma}} \right) \log \left(\frac{1 + \hat{\gamma} y_i}{1 + \hat{\gamma} \hat{\mu}_i} \right) \right].$$

O desvio para MLGs binomiais negativos é próximo ao desvio para MLGs de Poisson quando $\hat{\gamma}$ é perto de 0.

Quantidades w_i e d_i para algumas funções de ligações.

Ligação	w_i	d_i
$\log(\mu_i) = \eta_i$	$\frac{\mu_i}{(1+\mu_i\gamma)}$	μ_i
$\mu_i = \eta_i$	$(\mu_i + \mu_i^2\gamma)^{-1}$	1
$\sqrt{\mu_i} = \eta_i$	$\frac{4}{(1+\mu_i\gamma)}$	$2\sqrt{\mu_i}$

Exercício: conferir essas quantidades.

Comparando MLGs: Poisson e Binomial Negativa

Como podemos comparar MLGs de Poisson e binomiais negativos que possuem as mesmas variáveis explicativas, para determinar se o modelo binomial negativo tem um melhor ajuste que o Poisson?

Uma comparação informal pode ser baseada nos valores do AIC.

Para um teste formal de significância, podemos testar $H_0 : \gamma = 0$, pois a Poisson é um caso limite da binomial negativo quando $\gamma \downarrow 0$.

Como γ é positivo, $\gamma = 0$ na fronteira do espaço do parâmetro. Assim, a estatística da razão de verossimilhança não possui distribuição assintótica qui-quadrado sob H_0 . Em vez disso, é uma mistura de uma distribuição de ponto único em 0 (que ocorre quando $\hat{\gamma} = 0$) e uma qui-quadrado com $g.l. = 1$.

Modelo Binomial Negativo com Variação Proporcional à Média

Uma parametrização alternativa para a binomial negativa é escrever a densidade Gama com $k\mu$ como parâmetro de forma,

$$f(\lambda; \mu, k) = \frac{\left(\frac{k}{\mu}\right)^k}{\Gamma(k)} e^{-\frac{k\lambda}{\mu}} \lambda^{k-1}, \quad \lambda \geq 0,$$

então $\mathbb{E}(\lambda) = \mu$ e $\text{Var}(\lambda) = \frac{\mu}{k}$.

Para esta parametrização, a mistura produz uma distribuição binomial negativa com

$$\mathbb{E}(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \frac{\mu(1+k)}{k}.$$

Agora, a variância é linear em vez de quadrática em μ . Corresponde a uma inflação na variância da Poisson, convergindo para a variância da Poisson quando $k \rightarrow \infty$.

As duas parametrizações da binomial negativa às vezes são denotadas por NB1 (linear) e NB2 (quadrático).

Somente o modelo NB2 se enquadra na estrutura tradicional do MLG, sendo expressável como uma distribuição da família exponencial com parâmetro de escala, e é muito mais usada.

Diferentemente do modelo NB2, para o modelo NB1, β e k não são parâmetros ortogonais, e $\hat{\beta}$ não é um estimador consistente quando o modelo da média é válido, mas a distribuição verdadeira não é binomial negativa (Cameron e Trivedi 2013).

Lee e Nelder (1996) apresentaram o ajuste de MV do modelo para os modelos NB1.

Exemplo: Estudantes australianos

Venables e Ripley(1999) apresentam os resultados de um estudo sociológico desenvolvido na Austrália com 146 estudantes de 8a série e ensino médio com o objetivo de comparar a ausência na escola segundo os seguintes fatores:

- ano que o estudante está cursando (1: 8a série, 2: 1o ano do ensino médio, 3: 2o ano do ensino médio, 4: 3o ano do ensino médio);
- etnia (0: aborígine, 1: não aborígine);
- desempenho escolar (0: insuficiente, 1: suficiente); e
- sexo (0: masculino, 1: feminino).

Dados:

```
library(MASS)
dados<-quine
head(dados)
```

	Eth	Sex	Age	Lrn	Days
1	A	M	F0	SL	2
2	A	M	F0	SL	11
3	A	M	F0	SL	14
4	A	M	F0	AL	5
5	A	M	F0	AL	5
6	A	M	F0	AL	13

```
str(dados)
```

```
'data.frame':  146 obs. of  5 variables:
 $ Eth : Factor w/ 2 levels "A","N": 1 1 1 1 1 1 1 1 1 1 ...
 $ Sex : Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 2 2 ...
 $ Age : Factor w/ 4 levels "F0","F1","F2",...: 1 1 1 1 1 1 1 1 2 2 ...
 $ Lrn : Factor w/ 2 levels "AL","SL": 2 2 2 1 1 1 1 1 2 2 ...
 $ Days: int  2 11 14 5 5 13 20 22 6 6 ...
```

Denotamos por Y_{ijklm} o número de faltas num determinado período referente ao m -ésimo aluno, cursando o i -ésimo ano, de etnia j , com desempenho escolar k e pertencente ao l -ésimo sexo, em que $i = 1, 2, 3, 4$, $j, k, l = 1, 2$ e $m = 1, \dots, 144$.

Vamos supor que $Y_{ijklm} \sim \text{BN}(\gamma^{-1}, \mu_{ijkl})$, em que

$$\log(\mu_{ijkl}) = \alpha + \beta_i + \gamma_j + \delta_k + \theta_l,$$

com $\beta_1 = 0$, $\gamma_1 = 0$, $\delta_1 = 0$ e $\theta_1 = 0$.

Assim, temos um modelo casela de referência em que β_2 , β_3 e β_4 denotam os incrementos do primeiro, segundo e terceiro ano do ensino médio, respectivamente, em relação à 8a série, γ_2 é a diferença entre os efeitos do grupo não aborígine com relação ao grupo aborígine, δ_2 denota a diferença entre os efeitos dos grupos com desempenho suficiente e insuficiente e θ_2 é a diferença entre os efeitos do sexo feminino e masculino.

```
fit.bn = glm.nb( Days ~ Eth + Sex + Age + Lrn, data = dados)
summary(fit.bn)
```

Call:

```
glm.nb(formula = Days ~ Eth + Sex + Age + Lrn, data = dados,
       init.theta = 1.274892646, link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7918	-0.8892	-0.2778	0.3797	2.1949

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.89458	0.22842	12.672	< 2e-16 ***
EthN	-0.56937	0.15333	-3.713	0.000205 ***
SexM	0.08232	0.15992	0.515	0.606710
AgeF1	-0.44843	0.23975	-1.870	0.061425 .
AgeF2	0.08808	0.23619	0.373	0.709211
AgeF3	0.35690	0.24832	1.437	0.150651
LrnSL	0.29211	0.18647	1.566	0.117236

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

(Dispersion parameter for Negative Binomial(1.2749) family taken to be 1)

Null deviance: 195.29 on 145 degrees of freedom

Residual deviance: 167.95 on 139 degrees of freedom

AIC: 1109.2

Number of Fisher Scoring iterations: 1

Theta: 1.275

Std. Err.: 0.161

2 x log-likelihood: -1093.151

O desvio do modelo ajustado foi de 167,95 (139 graus de liberdade).

Notamos que os fatores sexo e desempenho escolar não são significativos a 10%, sendo portanto retirados do modelo.

Então, para o novo modelo retiramos os fatores sexo e desempenho escolar e incluímos a interação ano e etnia.

O novo ajuste é apresentado a seguir.

```
fit.bn1 = glm.nb( Days ~ Eth + Age + Eth:Age, data = dados)
summary(fit.bn1)
```

Call:

```
glm.nb(formula = Days ~ Eth + Age + Eth:Age, data = dados, init.theta = 1.357377778,
link = log)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-2.7367	-0.9489	-0.3137	0.4514	2.1473

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.6280	0.2495	10.535	< 2e-16 ***
EthN	0.1311	0.3455	0.379	0.70436
AgeF1	0.1784	0.3195	0.558	0.57664
AgeF2	0.8267	0.3172	2.606	0.00916 **
AgeF3	0.3708	0.3337	1.111	0.26650
EthN:AgeF1	-0.9916	0.4394	-2.257	0.02403 *
EthN:AgeF2	-1.2392	0.4466	-2.775	0.00552 **
EthN:AgeF3	-0.1763	0.4636	-0.380	0.70380

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

(Dispersion parameter for Negative Binomial(1.3574) family taken to be 1)

Null deviance: 205.56 on 145 degrees of freedom
Residual deviance: 167.84 on 138 degrees of freedom
AIC: 1102.6

Number of Fisher Scoring iterations: 1

Theta: 1.357
Std. Err.: 0.174

2 x log-likelihood: -1084.638

```
fitted(fit.bn1)[1:10]
```

1	2	3	4	5	6	7	8
13.84615	13.84615	13.84615	13.84615	13.84615	13.84615	13.84615	13.84615
9	10						
16.55000	16.55000						

```
X<-cbind(rep(1,length(dados$Days)), ifelse(dados$Eth=="N",1,0),
         ifelse(dados$Age=="F1",1,0),
         ifelse(dados$Age=="F2",1,0),ifelse(dados$Age=="F3",1,0),
         ifelse(dados$Eth=="N",1,0)*ifelse(dados$Age=="F1",1,0),
         ifelse(dados$Eth=="N",1,0)*ifelse(dados$Age=="F2",1,0),
         ifelse(dados$Eth=="N",1,0)*ifelse(dados$Age=="F3",1,0))
betas<-matrix(fit.bn1$coefficients,8,1)

exp(X%*%betas)[1:10,]
```

```
[1] 13.84615 13.84615 13.84615 13.84615 13.84615 13.84615 13.84615
[8] 13.84615 16.55000 16.55000
```



```
cbind(dados[,c(1,3,5)],fitted(fit.bn1),exp(X%*%betas))[15:40,]
```

	Eth	Age	Days	fitted(fit.bn1)	exp(X %*% betas)
15	A	F2	32	31.65000	31.65000
16	A	F2	53	31.65000	31.65000
17	A	F2	57	31.65000	31.65000
18	A	F2	14	31.65000	31.65000
19	A	F2	16	31.65000	31.65000
20	A	F2	16	31.65000	31.65000
21	A	F2	17	31.65000	31.65000
22	A	F2	40	31.65000	31.65000
23	A	F2	43	31.65000	31.65000
24	A	F2	46	31.65000	31.65000
25	A	F3	8	20.06250	20.06250
26	A	F3	23	20.06250	20.06250
27	A	F3	23	20.06250	20.06250
28	A	F3	28	20.06250	20.06250
29	A	F3	34	20.06250	20.06250
30	A	F3	36	20.06250	20.06250
31	A	F3	38	20.06250	20.06250
32	A	F0	3	13.84615	13.84615
33	A	F0	5	13.84615	13.84615
34	A	F0	11	13.84615	13.84615
35	A	F0	24	13.84615	13.84615
36	A	F0	45	13.84615	13.84615
37	A	F1	5	16.55000	16.55000
38	A	F1	6	16.55000	16.55000
39	A	F1	6	16.55000	16.55000
40	A	F1	9	16.55000	16.55000

```
cbind(min(fitted(fit.bn1)),max(fitted(fit.bn1)))
```

```
      [,1] [,2]  
[1,]      7 31.65
```

```
id1<-as.numeric(names(fitted(fit.bn1)[fitted(fit.bn1)==max(fitted(fit.bn1))])[1]);id1
```

```
[1] 14
```

```
dados[id1,]
```

```
      Eth Sex Age Lrn Days  
14      A  M  F2  SL     6
```

```
id2<-as.numeric(names(fitted(fit.bn1)[fitted(fit.bn1)==min(fitted(fit.bn1))])[1]);id2
```

```
[1] 79
```

```
dados[id2,]
```

```
      Eth Sex Age Lrn Days  
79      N  M  F1  SL     0
```

O desvio do novo modelo foi de 167,84 (138 graus de liberdade).

Podemos notar que o grupo não aborígene tem em geral um número médio menor de dias ausentes.

A maior média é observada para estudantes do grupo aborígene cursando o 2o do ensino médio e o menor valor médio é observado para estudantes do grupo não aborígene cursando o 1o do ensino médio.

O valor estimado de γ é $\hat{\gamma} = \frac{1}{1,357} = 0,737$.

```
fit.poisson = glm( Days ~ Eth + Age + Eth:Age, family = poisson, data = dados)
summary(fit.poisson)
```

Call:

```
glm(formula = Days ~ Eth + Age + Eth:Age, family = poisson, data = dados)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-6.947	-2.987	-0.797	1.651	9.554

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.62801	0.07454	35.258	< 2e-16 ***
EthN	0.13110	0.10040	1.306	0.1916
AgeF1	0.17838	0.09261	1.926	0.0541 .
AgeF2	0.82673	0.08447	9.787	< 2e-16 ***
AgeF3	0.37084	0.09312	3.983	6.82e-05 ***
EthN:AgeF1	-0.99157	0.13637	-7.271	3.56e-13 ***
EthN:AgeF2	-1.23923	0.12824	-9.664	< 2e-16 ***
EthN:AgeF3	-0.17627	0.12753	-1.382	0.1669

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2073.5 on 145 degrees of freedom
Residual deviance: 1597.1 on 138 degrees of freedom
AIC: 2201.6

Number of Fisher Scoring iterations: 5

Verificamos neste exemplo, como fica o ajuste através de um modelo log-linear de Poisson.

O modelo log-linear de Poisson apresenta fortes indícios de sobredispersão, isso é justificado pelo valor do desvio de 1597, 11 (138 graus de liberdade).

O AIC para modelo log-linear de Poisson é de 1102,6, enquanto para a binomial negativa é de 2201,6.

- 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.