

Modelos de regressão para dados de contagem (parte 4) (superdispersão)

Prof. Caio Azevedo

Superdispersão (ou sobredispersão)

- Quando a variável resposta apresenta variância maior do que aquela imposta pelo modelo probabilístico/regressão.
- Exemplos 4 (análise descritiva/inferencial) e 12 (análise descritiva/análise inferencial).
- Quando o modelo de regressão define grupos (covariáveis qualitativas, como nos Exemplo 4 e 12) podemos comparar as médias amostrais com as variâncias amostrais.
- Contudo, mesmo quando a abordagem acima não é factível, podemos tentar identificar a presença de tal problema através de análises de diagnóstico (resíduos).

Cont.

- Do ponto de vista inferencial, quando o desvio $D(\mathbf{y}; \tilde{\boldsymbol{\mu}})$ é maior que o número de graus de liberdade $(n-p)$, pode haver indícios de sobredispersão.
- Isso pode ser avaliado mais precisamente pelo nível descritivo do teste de ajustamento através do desvio (utilizando de sua aproximação pela distribuição de qui-quadrado, quando pertinente).
- O gráfico de envelopes e os de diagnóstico também podem fornecer indícios da existência de superdispersão.
- Diferentes circunstâncias, entretanto, podem causar um valor alto para o desvio. Algumas delas representam uma sobredispersão aparente (veja slide seguinte).

Cont.

- Presença de pontos aberrantes e/ou ausência de covariáveis relevantes e/ou de algum termo na parte sistemática (η_i), incorreta especificação da função de ligação.
- A superdispersão também pode ser causada por: existência de subgrupos com diferentes distribuições, dependência entre as observações, características latentes (não observáveis diretamente) presentes nas unidades experimentais, fatores não controlados no experimento entre outros.
- Medidas de diagnóstico, vistas anteriormente, são ferramentas importantes para detectarmos o fenômeno.

Consequências da presença superdispersão (Hinde e Demétrio (1998))

- Os erros-padrão estimados (a partir do modelo) podem ser (muito) subestimados conseqüentemente, podemos avaliar incorretamente a importância dos parâmetros (significância).
- Podem ocorrer (grandes) mudanças no valor do desvio as quais podem conduzir à seleção de modelos extremamente complexos.
- Finalmente, podemos selecionar modelos inapropriados e as previsões podem ser (falsamente) “precisas” (menor comprimento dos intervalos de confiança).

Modelagem da super-dispersão

- Há várias formas de modelar superdispersão, veja [aqui](#) e [aqui](#) .
- Consideraremos uma abordagem que se utiliza variáveis latentes.
- Suponha que para um conjunto fixo $\mathbf{x} = (x_1, \dots, x_p)'$ de valores de variáveis explicativas, $Y|z$ (resposta) tem média z e variância z .
- No entanto Z , que é não observável, varia nas unidades amostrais, (ou seja, temos (Y_i, Z_i) , $i=1,2,\dots,n$) com \mathbf{x} fixo, de modo que $E(Z) = \mu$.

Modelagem da super-dispersão

- Assim

$$\mathcal{E}(Y) = \mathcal{E}(\mathcal{E}(Y|Z)) = \mathcal{E}(Z) = \mu$$

$$\mathcal{V}(Y) = \mathcal{E}(\mathcal{V}(Y|Z)) + \mathcal{V}(\mathcal{E}(Y|Z)) = \mu + \mathcal{V}(Z)$$

- Podemos assumir que $Y|Z = z \sim \text{Poisson}(z)$ e $Z \sim \text{gama}(\mu, \phi)$ (sob a parametrização adotada no curso).
- Assim $\mathcal{E}(Z) = \mu$ e $\mathcal{V}(Z) = \mu^2/\phi$. Isto implica que $\mathcal{E}(Y) = \mu$ e $\mathcal{V}(Y) = \mu + \mu^2/\phi$.

Modelagem da super-dispersão

- Note que se $\phi \rightarrow \infty$ então $\mathcal{V}(Y) \rightarrow \mathcal{E}(Y) = \mu$ (à rigor teremos o modelo de Poisson).
- Além disso,

$$f(y|z) = \frac{e^{-z} z^y}{y!} \mathbb{1}_{\{0,1,2,\dots\}}(y),$$

e

$$g(z) = \frac{1}{\Gamma(\phi)} \left(\frac{z\phi}{\mu} \right)^\phi e^{-\frac{\phi z}{\mu}} \frac{1}{z} \mathbb{1}_{(0,\infty)}(z).$$

Modelagem da super-dispersão

- Portanto, a fdp de Y é dada por

$$\begin{aligned} f(y) &= \int_0^\infty f(y|z)g(z)dy = \frac{1}{y!\Gamma(\phi)} \left(\frac{\phi}{\mu}\right)^\phi \int_0^\infty e^{-z(1+\phi/\mu)} z^{\phi+y-1} dz \\ &= \frac{1}{y!\Gamma(\phi)} \left(\frac{\phi}{\mu}\right)^\phi \frac{\Gamma(\phi+y)}{(1+\phi/\mu)^{\phi+y}} = \frac{\Gamma(\phi+y)}{\Gamma(y+1)\Gamma(\phi)} \frac{\phi^\phi}{\mu^\phi} \frac{\mu^{\phi+y}}{(\mu+\phi)^{\phi+y}} \\ &= \frac{\Gamma(\phi+y)}{\Gamma(y+1)\Gamma(\phi)} \left(\frac{\phi}{\mu+\phi}\right)^\phi \left(\frac{\mu}{\mu+\phi}\right)^y \mathbb{1}_{\{0,1,2,\dots\}}(y) \end{aligned}$$

- Assim, temos que $Y \sim \text{BN}(\mu, \phi)$. Neste caso, a distribuição BN (binomial negativa) não pertence à parametrização da FE adotada para a **construção dos MLG**.

Modelo de regressão binomial-negativo

- Sejam $Y_i, i = 1, 2, \dots, n$, $Y_i \stackrel{ind.}{\sim} \text{BN}(\mu_i, \phi)'$.
- Temos que $\mathcal{E}(Y_i) = \mu_i$ e $\mathcal{V}(Y_i) = \mu_i + \frac{\mu_i^2}{\phi}$.
- $g(\mu_i) = \eta_i \rightarrow \mu_i = g^{-1}(\eta_i)$, $\eta_i = \mathbf{X}_i\boldsymbol{\beta}$, $g(\cdot)$ é uma função de ligação apropriada, inversível e duplamente diferenciável.
- Comumente, $g(\mu_i) = \ln(\mu_i)$. Outras escolhas $g(\mu_i) = \mu_i, \sqrt{\mu_i}, 1/\mu_i$.

Estimação por MV

- Logverossimilhança (seja $\theta = (\beta', \phi)'$).

$$\begin{aligned}l(\theta) &= \sum_{i=1}^n \left\{ \ln \left[\frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \right] + \phi \ln(\phi) + y_i \ln(\mu_i) \right. \\ &\quad \left. - (\phi + y_i) \ln(\mu_i + \phi) \right\} \\ &= \sum_{i=1}^n \left\{ \ln [\Gamma(\phi + y_i)] - \ln [\Gamma(y_i + 1)] - \ln [\Gamma(\phi)] + \phi \ln(\phi) \right. \\ &\quad \left. + y_i \ln(\mu_i) - (\phi + y_i) \ln(\mu_i + \phi) \right\}.\end{aligned}$$

Estimação por MV

- Vetor escore para β (cada componente)

$$\begin{aligned} S(\beta_j) &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{\phi + y_i}{\phi + \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} x_{ji} - \frac{\phi + y_i}{\phi + \mu_i} \frac{\partial \mu_i}{\partial \eta_i} x_{ji} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{\phi (\partial \mu_i / \partial \eta_i)}{\mu_i (\phi + \mu_i)} (y_i - \mu_i) x_{ji} \right\} = \sum_{i=1}^n \omega_i f_i^{-1} (y_i - \mu_i) x_{ji}. \end{aligned}$$

- Em que $\omega_i = (\partial \mu_i / \partial \eta_i)^2 / (\mu_i^2 \phi^{-1} + \mu_i)$ e $f_i = \partial \mu_i / \partial \eta_i$

Estimação por MV

- Matricialmente, temos

$$S(\beta) = \mathbf{X}'\mathbf{W}\mathbf{F}^{-1}(\mathbf{y} - \boldsymbol{\mu}),$$

em que \mathbf{X} é a matriz de planejamento do modelo (conforme visto anteriormente), $\mathbf{W} = \text{diag}\{\omega_1, \dots, \omega_n\}$, $\mathbf{F} = \text{diag}\{f_1, \dots, f_n\}$, $\mathbf{y} = (y_1, \dots, y_n)'$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$.

Estimação por MV

- Para o parâmetro ϕ , temos que

$$S(\phi) = \sum_{i=1}^n \left[\Psi(\phi + y_i) - \Psi(\phi) - \frac{y_i + \phi}{\phi + \mu_i} + \ln(\phi) - \ln(\mu_i + \phi) + 1 \right],$$

em que $\Psi(\cdot)$ é a função digama (vista anteriormente).

- Podemos notar que o sistema
$$\begin{cases} \mathbf{S}(\tilde{\beta}) = \mathbf{0}_{(p \times 1)} \\ S(\tilde{\phi}) = 0 \end{cases}$$
 não apresenta solução analítica. Assim, algum método numérico deverá ser empregado e, nesse caso, utilizaremos o algoritmo Escore de Fisher semelhante ao que fizemos para os MLG.

Estimação por MV

■ Matriz Hessiana

$$\begin{aligned}\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l} &= \frac{\partial}{\partial \beta_l} \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} x_{ji} - \frac{\phi + y_i}{\phi + \mu_i} \frac{\partial \mu_i}{\partial \eta_i} x_{ji} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i^2} x_{ji} \left[\frac{\partial^2 \mu_i}{\partial \eta_i^2} \mu_i x_{li} - x_{li} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \right] \right. \\ &\quad \left. - x_{ji} \frac{\phi + y_i}{(\phi + \mu_i)^2} \left[\frac{\partial^2 \mu_i}{\partial \eta_i^2} (\phi + \mu_i) x_{li} - \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 x_{li} \right] \right\} \\ &= \sum_{i=1}^n \left\{ \frac{\phi + y_i}{(\phi + \mu_i)^2} - \frac{y_i}{\mu_i^2} \right\} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 x_{ji} x_{jl} \\ &\quad + \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} - \frac{\phi + y_i}{\phi + \mu_i} \right\} \frac{\partial^2 \mu_i}{\partial \eta_i^2} x_{ji} x_{jl}\end{aligned}$$

Estimação por MV

■ Informação de Fisher

$$\begin{aligned} I(\beta_j, \beta_l) &= -\mathcal{E} \left(\sum_{i=1}^n \left\{ \frac{\phi + Y_i}{(\phi + \mu_i)^2} - \frac{Y_i}{\mu_i^2} \right\} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 x_{ji} x_{jl} \right. \\ &\quad \left. + \sum_{i=1}^n \left\{ \frac{Y_i}{\mu_i} - \frac{\phi + Y_i}{\phi + \mu_i} \right\} \frac{\partial^2 \mu_i}{\partial \eta_i^2} x_{ji} x_{jl} \right) \\ &= -\sum_{i=1}^n \left\{ \frac{1}{(\phi + \mu_i)} - \frac{1}{\mu_i} \right\} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 x_{ji} x_{jl} \\ &= \sum_{i=1}^n \left\{ \frac{\phi}{(\phi + \mu_i) \mu_i} \right\} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 x_{ji} x_{jl} = \sum_{i=1}^n \left\{ \frac{(\partial \mu_i / \partial \eta_i)^2 \phi}{(\phi + \mu_i) \mu_i} \right\} x_{ji} x_{jl} \\ &= \sum_{i=1}^n \omega_i x_{ji} x_{jl} \end{aligned}$$

Estimação por MV

- Matricialmente, temos $I(\beta, \beta) = \mathbf{X}'\mathbf{W}\mathbf{X}$.
- Por outro lado, Lawless (1987) mostra que a componente da informação de Fisher para ϕ pode ser expressa na forma:

$$I(\phi, \phi) = \sum_{i=1}^n \left\{ \sum_{j=0}^{\infty} (\phi + j)^{-2} P(Y_i \geq j) - \phi^{-1} \mu_i / (\mu_i + \phi) \right\},$$

e que β e ϕ são ortogonais, ou seja, $I(\beta, \phi) = \mathbf{0}_{(p \times 1)}$.

Estimação por MV

- Assim, embora a Informação de Fisher seja bloco diagonal, ou seja,

$$I(\boldsymbol{\beta}, \phi) = \begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{0}_{(p \times 1)} \\ \mathbf{0}_{(1 \times p)} & I(\phi, \phi) \end{bmatrix},$$

o algoritmo escore de Fisher, nesse caso, terá de ser conduzido, simultaneamente, para os dois parâmetros, pois \mathbf{W} depende de ϕ , lembrando que $\omega_i = (\partial\mu_i/\partial\eta_i)^2/(\mu_i^2\phi^{-1} + \mu_i)$.

- O algoritmo Escore de Fisher (AEF) é definido como: Sejam $\boldsymbol{\beta}^{(0)}$ e $\phi^{(0)}$ estimativas iniciais de $\boldsymbol{\beta}$ e ϕ (chute inicial), respectivamente, então faça

$$\begin{aligned}
 & \begin{bmatrix} \boldsymbol{\beta}^{(t+1)} \\ \phi^{(t+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}^{(t)} \\ \phi^{(t)} \end{bmatrix} \\
 + & \begin{bmatrix} \mathbf{X}' \mathbf{W}^{(t)} \mathbf{X} & \mathbf{0}_{(p \times 1)} \\ \mathbf{0}_{(1 \times p)} & I(\phi^{(t)}, \phi^{(t)}) \end{bmatrix}^{-1} \\
 \times & \begin{bmatrix} \mathbf{X}' \mathbf{W}^{(t)} (\mathbf{F}^{(t)})^{-1} (\mathbf{y} - \boldsymbol{\mu}^{(t)}) \\ \sum_{i=1}^n \left[\Psi(\phi^{(t)} + y_i) - \Psi(\phi^{(t)}) - \frac{y_i + \phi^{(t)}}{\phi^{(t)} + \mu_i^{(t)}} + \ln(\phi^{(t)}) - \ln(\mu_i^{(t)} + \phi^{(t)}) + 1 \right] \end{bmatrix}
 \end{aligned}$$

$t = 0, 1, 2, \dots$, até que algum critério de convergência seja satisfeito, por exemplo $\|\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)}\| < \epsilon$ para algum $\epsilon > 0$ e $\boldsymbol{\theta} = (\boldsymbol{\beta}', \phi)'$.

Cont.

- O algoritmo Escore de Fisher (AEF) pode ser também escrito como:

$$\begin{aligned}\beta^{(t+1)} &= \left(\mathbf{X}' \mathbf{W}^{(t)} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{W}^{(t)} \mathbf{y}^{*(t)} \\ \phi^{(t+1)} &= \phi^{(t)} + I_{\phi}(\phi^{(t)}, \phi^{(t)})^{-1} \left(S(\phi^{(t)}) \right)\end{aligned}$$

$t = 0, 1, 2, \dots$, até que algum critério de convergência seja satisfeito,
em que

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

Cont.

- (Cont.) em que

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

é a informação de Fisher observada, que corresponde ao valor simétrico da Hessiana ($H(\phi, \phi)$) (podemos utilizar a informação de Fisher esperada $I(\phi, \phi)$) embora esta possa trazer uma certa instabilidade numérica (?) e $\Psi'(\cdot)$ é a função trigama.

Cont.

- Quantidades ω_i e f_i para algumas funções de ligação:

Ligação	ω_i	f_i
$\ln(\mu_i) = \eta_i$	$\frac{\mu_i}{(\mu_i\phi^{-1} + 1)}$	μ_i
$\mu_i = \eta_i$	$(\mu_i^2\phi^{-1} + \mu_i)^{-1}$	1
$\sqrt{\mu_i} = \eta_i$	$\frac{4}{(\mu_i\phi^{-1} + 1)}$	$2\sqrt{\mu_i}$

Cont.

- As Condições de Regularidade (CR) para o modelo de regressão BN podem ser encontradas [aqui](#).
- Sob as Condições de Regularidade, temos que:

$$|I_0(\phi, \phi) - I(\phi, \phi)| \xrightarrow[n \rightarrow \infty]{P} 0.$$

- Sob as mesmas Condições de Regularidade e para n suficientemente grande temos que $\hat{\beta} \approx N_p(\beta, (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1})$ e $\hat{\phi} \approx N(\phi, I^{-1}(\phi, \phi))$ e $\hat{\beta} \perp \hat{\phi}$ (assintoticamente).
- Defina $\widehat{\mathbf{W}}$ e $I(\widehat{\phi}, \widehat{\phi})$, respectivamente \mathbf{W} e $I(\phi, \phi)$, em que as quantidades desconhecidas são substituídas pelos respectivos emv.

Cont.

- Se $\hat{\beta}_j$ é a j -ésima componente do vetor $\hat{\beta}$ então $\hat{\beta}_j \approx N(\beta_j, \psi_j)$ em que ψ_j é o j -ésimo elemento da diagonal principal de $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$.
- Temos ainda que como $\hat{\phi}$ é um estimador consistente, então $\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\psi}_j}} \approx N(0, 1)$ (sob as condições de regularidade e para n suficientemente grande), em que $\hat{\psi}_j$ é o j -ésimo elemento da diagonal principal de $(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1}$.
- Analogamente, temos que $\frac{\hat{\phi} - \phi}{\sqrt{\widehat{\mathcal{V}}(\hat{\phi})}} \approx N(0, 1)$, em que $\widehat{\mathcal{V}}(\hat{\phi}) = \mathbf{I}^{-1}(\hat{\phi}, \hat{\phi})$.

Inferência

- Portanto, considerando-se $P(X \leq z_{\frac{1+\gamma}{2}}) \approx \frac{1+\gamma}{2}$, $X \sim N(0, 1)$, temos que

$$IC(\beta_j, \gamma) \approx \left[\hat{\beta}_j - z_{\frac{1+\gamma}{2}} \sqrt{\hat{\psi}_j}; \hat{\beta}_j + z_{\frac{1+\gamma}{2}} \sqrt{\hat{\psi}_j} \right],$$

$$IC(\phi, \gamma) \approx \left[\hat{\phi} - z_{\frac{1+\gamma}{2}} \sqrt{\hat{V}(\hat{\phi})}; \hat{\phi} + z_{\frac{1+\gamma}{2}} \sqrt{\hat{V}(\hat{\phi})} \right].$$

- Os intervalos de confiança numéricos são obtidos substituindo-se os estimadores pelas respectivas estimativas.

Inferência

- Hipóteses do tipo $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{M}$ vs $H_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{M}$ podem ser testadas através da estatística (do tipo Wald)

$$Q_t = (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{M})' (\mathbf{C}\hat{\mathcal{V}}(\hat{\boldsymbol{\beta}})\mathbf{C}')^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{M}),$$

em que $\hat{\mathcal{V}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\hat{\mathbf{W}}\mathbf{X})^{-1}$.

Inferência

- Note que $\hat{\beta} \approx N_p(\beta, (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1})$, portanto, devido à esse resultado e algumas propriedades da normal multivariada, temos que se

$$Q_t^* = (\mathbf{C}\hat{\beta} - \mathbf{M})' (\mathbf{C}\mathcal{V}(\hat{\beta})\mathbf{C}')^{-1} (\mathbf{C}\hat{\beta} - \mathbf{M}),$$

então 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} .

- Além disso, $|\widehat{\mathbf{W}} - \mathbf{W}| \xrightarrow[n \rightarrow \infty]{P} \mathbf{0}_{(n \times n)}$ o que implica que $|\widehat{\mathcal{V}}(\hat{\beta}) - \mathcal{V}(\hat{\beta})| \xrightarrow[n \rightarrow \infty]{P} \mathbf{0}_{(p \times p)}$, pois cada componente de $\widehat{\mathcal{V}}(\hat{\beta})$ é uma função contínua da respectiva componente de $\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}$ e também por Slutsky pois $\hat{\phi} \xrightarrow[n \rightarrow \infty]{P} \phi$.

Inferência

- Portanto, sob H_0 e para n suficientemente grande, pelos resultados anteriores $(Q_t^*, \widehat{\mathbf{W}})$, temos que $Q_t \approx \chi_c^2$.
- Assim, rejeita-se H_0 se p – valor $\leq \alpha$, em que

$$p\text{ – valor} \approx P(X \geq q_t | H_0),$$

$$\text{em que } X \sim \chi_c^2 \text{ e } q_t = \left(\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{M} \right)' \left(\mathbf{C}\tilde{\mathbf{V}}(\hat{\boldsymbol{\beta}})\mathbf{C}' \right)^{-1} \left(\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{M} \right).$$

Inferência

- Sob H_1 , temos que $Q_t \approx \chi^2_{(c,\delta)}$ (qui-quadrado não central com c graus de liberdade e parâmetro de assimetria δ), em que

$$\delta = (\mathbf{C}\boldsymbol{\beta} - \mathbf{M})' (\mathbf{C}\mathcal{V}(\hat{\boldsymbol{\beta}})\mathbf{C}')^{-1} (\mathbf{C}\boldsymbol{\beta} - \mathbf{M}).$$

- Uma estimativa de δ é dada por

$$\tilde{\delta} = (\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{M})' (\mathbf{C}\tilde{\mathcal{V}}(\hat{\boldsymbol{\beta}})\mathbf{C}')^{-1} (\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{M}).$$

Diagnóstico do modelo

- A i -ésima componente do desvio do modelo para ϕ fixo é dada por

$$D^*(y_i, \hat{\mu}_i) = 2 \left\{ \phi \ln \left[\frac{\hat{\mu}_i + \phi}{y_i + \phi} \right] + y_i \ln \left[\frac{y_i(\hat{\mu}_i + \phi)}{\hat{\mu}_i(y_i + \phi)} \right] \right\}, \forall y_i \in \{1, 2, \dots\},$$

em que $\hat{\mu}_i$ é o emv de μ_i sob o modelo de regressão.

- Se $y_i = 0$ então

$$\begin{aligned} D^*(y_i, 0) &= 2 \{ \ln f(0; y_i, \phi) - \ln f(0, \hat{\mu}_i, \phi) \} \\ &= 2\phi \ln \{ \phi / (y_i + \phi) \} - 2\phi \ln \{ \phi / (\hat{\mu}_i + \phi) \} \\ &= 2\phi \ln \{ (\hat{\mu}_i + \phi) / (y_i + \phi) \} \\ \rightarrow D^*(0, 0) &= 2\phi \ln \{ (\hat{\mu}_i + \phi) / \phi \}. \end{aligned}$$

Diagnóstico do modelo

- Assim o desvio é dado por

$$D^*(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n \left\{ \left[\phi \ln \left[\frac{\hat{\mu}_i + \phi}{y_i + \phi} \right] + y_i \ln \left[\frac{y_i(\hat{\mu}_i + \phi)}{\hat{\mu}_i(y_i + \phi)} \right] \right] \mathbb{1}_{\{1,2,\dots\}}(y_i) + \phi \ln \{(\hat{\mu}_i + \phi)/\phi\} \mathbb{1}_{\{0\}}(y_i) \right\}.$$

Diagnóstico do modelo

- Na prática, substituímos ϕ por um estimador consistente, ou seja:

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n \left\{ \left[\hat{\phi} \ln \left[\frac{\hat{\mu}_i + \hat{\phi}}{y_i + \hat{\phi}} \right] + y_i \ln \left[\frac{y_i(\hat{\mu}_i + \hat{\phi})}{\hat{\mu}_i(y_i + \hat{\phi})} \right] \right] \mathbf{1}_{\{1,2,\dots\}}(y_i) + \hat{\phi} \ln \left\{ (\hat{\mu}_i + \hat{\phi}) / \hat{\phi} \right\} \mathbf{1}_{\{0\}}(y_i) \right\},$$

por exemplo, podemos usar o respectivo env.

Cont.

- Sob a hipótese de que o modelo adotado está correto $D(\mathbf{y}, \hat{\boldsymbol{\mu}})$ segue, aproximadamente, para $\hat{\phi}$ e $\hat{\mu}_i$ grandes, uma distribuição qui-quadrado com $(n - p)$ graus de liberdade.
- O resíduo componente do desvio é dado por:

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

em que (próximo slide)

Cont.

- (Cont.)

$$\left\{ \begin{array}{l} \sqrt{2} \operatorname{sinal}(y_i - \hat{\mu}_i) \left\{ \hat{\phi} \ln \left[\frac{\hat{\mu}_i + \hat{\phi}}{y_i + \hat{\phi}} \right] + y_i \ln \left[\frac{y_i(\hat{\mu}_i + \hat{\phi})}{\hat{\mu}_i(y_i + \hat{\phi})} \right] \right\}^{1/2}, \\ \quad \text{se } y_i \in \{1, 2, \dots, \} \\ \sqrt{2} \operatorname{sinal}(y_i - \hat{\mu}_i) \left\{ \hat{\phi} \ln(\hat{\mu}_i + \hat{\phi}) / \hat{\phi} \right\}^{1/2}, \text{ se } y_i = 0 \end{array} \right.$$

Cont.

- (Cont.) e

\hat{h}_{ii} é o i -ésimo elemento da diagonal principal da matriz,

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

ou seja, $\hat{h}_{ii} = \frac{\widehat{\phi} \widehat{\mu}_i}{\widehat{\phi} + \widehat{\mu}_i} \mathbf{X}'_i (\mathbf{X}' \widehat{W} \mathbf{X})^{-1} \mathbf{X}_i$.

- A medida acima, serve como medida de alavancagem, assim como nos **MLG**.

Cont.

- Para a distância de Cook, podemos utilizar

$$LD_i = \frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})^2} RP_i^2,$$

em que $RP_i^2 = \frac{Y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i^2 + \frac{\hat{\mu}_i^2}{\hat{\phi}}}}$.

- Estudos de Monte Carlo desenvolvidos por Svetliza (2002) (ver também Svetliza e Paula, 2003) indicam boa concordância entre o resíduo componente do desvio e a distribuição normal padrão, sob o bom ajuste do modelo.

Voltando ao Exemplo 4: comparação do número de acidentes

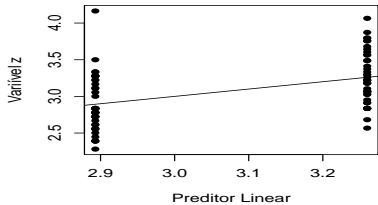
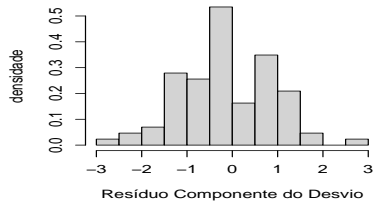
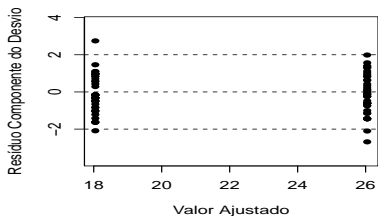
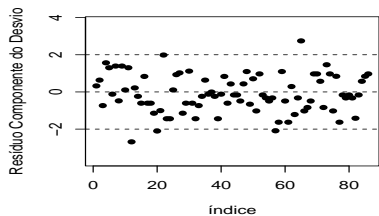
Modelo

$$Y_{ij} \stackrel{\text{ind.}}{\sim} \text{BN}(\mu_i, \phi), i = 1 \text{ (ano de 1961)}, 2 \text{ (ano de 1962)}, j = 1, \dots, 43$$

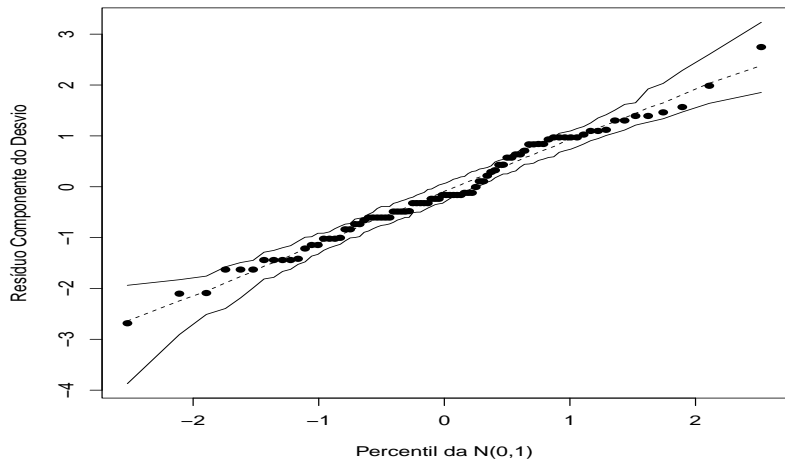
$$\ln \mu_i = \mu + \alpha_i, \alpha_1 = 0$$

- $\mathcal{E}(Y_{ij}) = \mu_i = e^{\mu + \alpha_i}$.
- e^{α_2} : o incremento multiplicativo (positivo ou negativo) da média do ano de 1962 em relação à média do ano de 1961 ($\mu_2 = \mu_1 e^{\alpha_2}$).

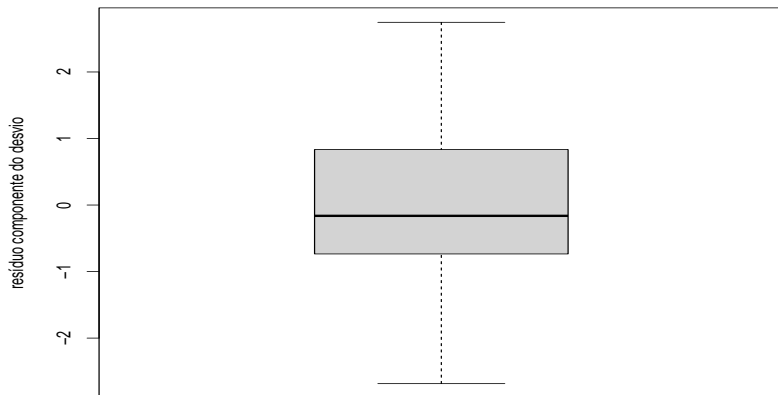
Gráficos de diagnóstico (modelo BN): Exemplo 4



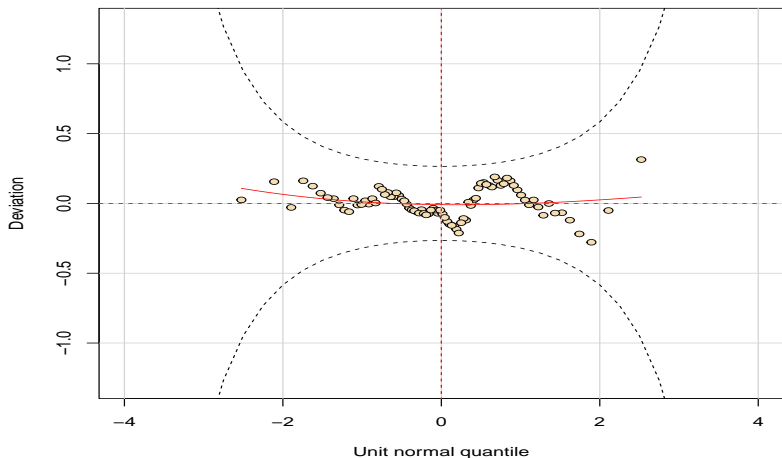
Envelope para os resíduos (modelo BN): Exemplo 4



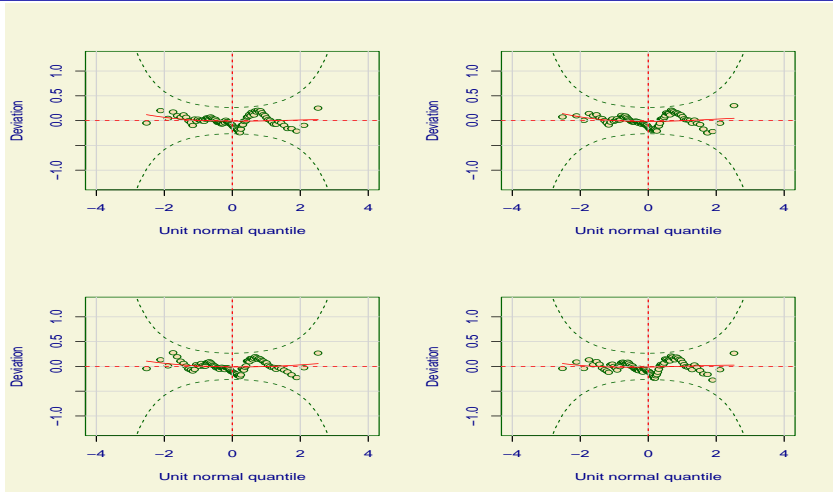
Box plot para os resíduos (modelo BN): Exemplo 4



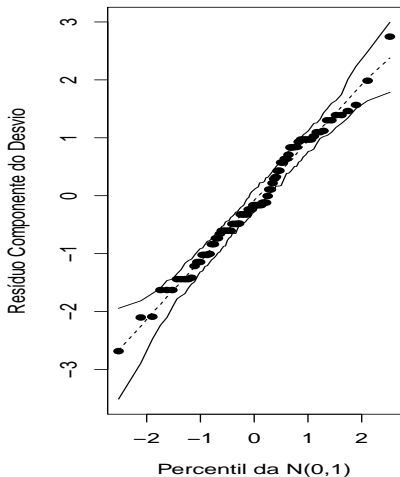
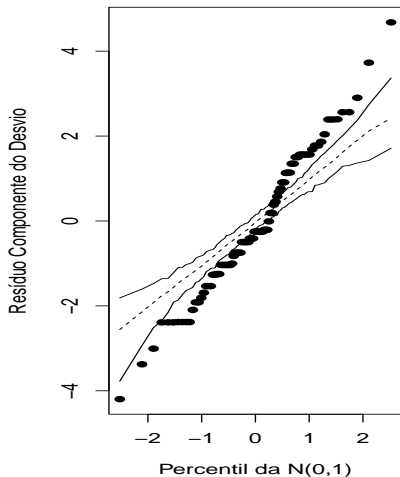
Worm plot para os resíduos (modelo BN): Exemplo 4



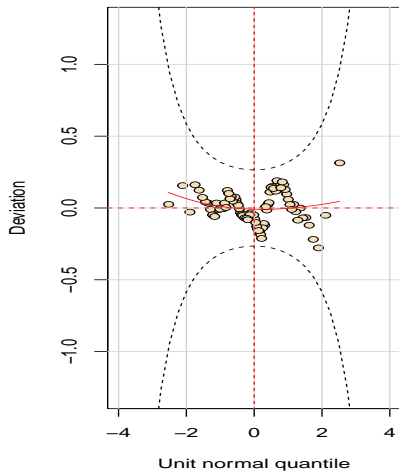
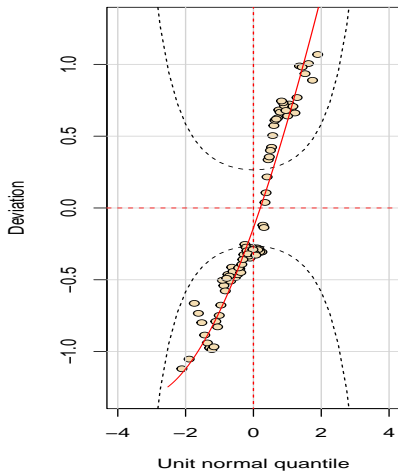
Worm plots para os resíduos (modelo BN): Exemplo 4



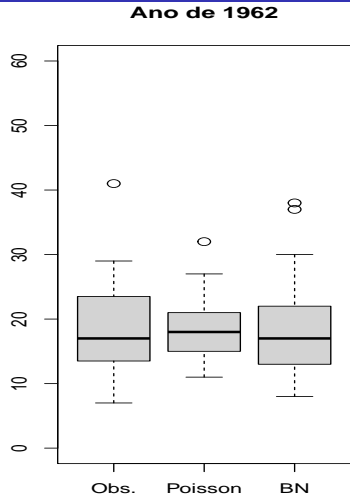
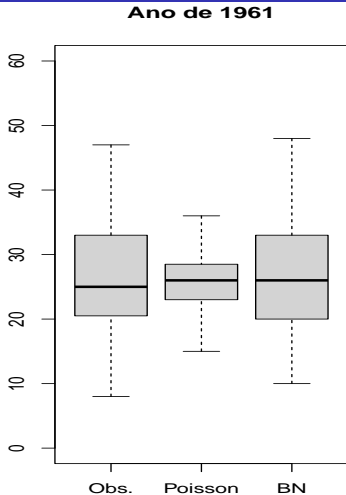
Gráficos de envelopes: Poisson e binomial negativo



Worm plots: Poisson e binomial negativo



Distribuições previstas e observadas (boxplot)



Estatísticas de ajuste e comparação de modelos

Modelo	AIC	BIC	AICc	SABIC	HQCIC	CAIC	p-valor
Poisson	656,94	661,85	657,09	655,54	658,92	663,85	< 0,0001
BN	595,37	602,73	595,66	593,27	598,33	605,73	0,3996

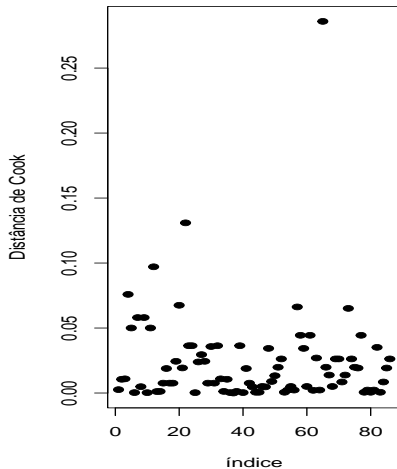
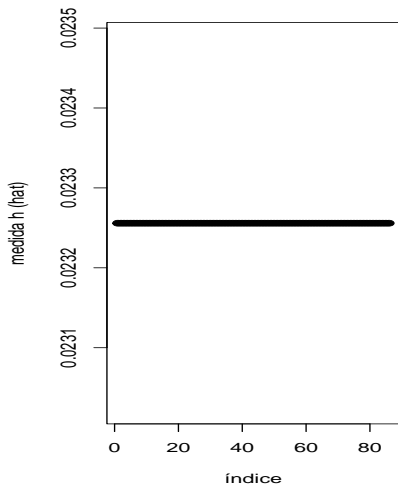
Estimativas dos parâmetros dos modelos

Modelo	Par.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
Poisson	β_0	3,26	0,03	[3,20 ; 3,32]	109,10	< 0,0001
	β_1	-0,37	0,05	[-0,46 ; -0,28]	-7,86	< 0,0001
BN	β_0	3,26	0,05	[3,16 ; 3,36]	62,31	< 0,0001
	β_1	-0,37	0,08	[-0,52 ; -0,22]	-4,79	< 0,0001
	ϕ	12,61	3,04	[6,65 ; 18,57]	-	-

Comentários

- O modelo BN se ajusto bem aos dados e muito melhor do que o modelo de Poisson, segundo a análise de resíduos.
- Adicionalmente, do ponto de vista dos critérios de informação bem como em relação à predição, o modelo BN se ajustou muito melhor aos dados.
- Percentualmente, os erros-padrão aumentaram bastante no modelo BN em relação ao modelo Poisson.
- Escolheremos o modelo BN.

Análise de Influência (modelo BN)



Análise de Sensibilidade

todas as observações

Modelo	Par.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
BN	β_0	3,26	0,05	[3,16 ; 3,36]	62,31	< 0,0001
	β_1	-0,37	0,08	[-0,52 ; -0,22]	-4,79	< 0,0001

sem a observação -# 65

Modelo	Par.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
BN	β_0	3,26	0,05	[3,16;3,36]	65,09	< 0,0001
	β_1	-0,40	0,07	[-0,54;-0,25]	-5,35	< 0,0001

Comentários

- Nenhuma das duas observações mudaram os resultados apresentados, significativamente.
- Aparentemente, houve redução estatisticamente significativa no número de acidentes de 1961 para 1962.

Exemplo 12: Modelo BN

$$Y_{ijklr} \stackrel{ind.}{\sim} \text{BN}(\mu_{ijkl}, \phi),$$

$i = 1, 2$ (etnia: 1 - aborígene, 2 - não aborígene),

$j = 1, 2$ (sexo : 1- feminino, 2 - masculino),

$k = 1, 2, 3, 4,$ (ano escolar: 1 - 8ª série, 2 - 1º ano, 3 - 2º ano,
4 - 3ºano)

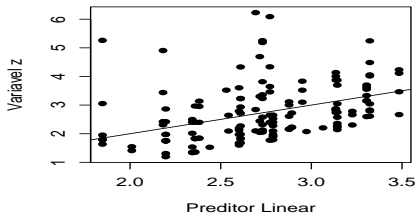
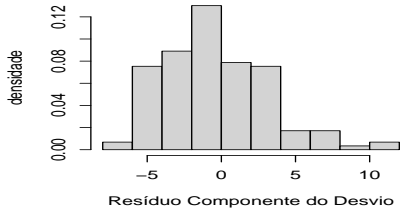
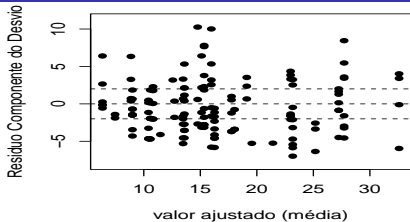
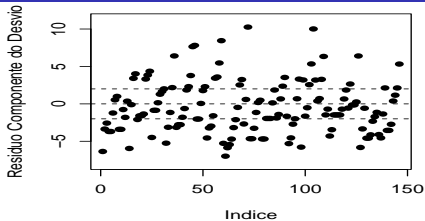
$l = 1, 2$ (desempenho: 1- suficiente, 2 - insuficiente,

$r = 1, 2, \dots, n_{ijkl}$ (unidades amostrais)

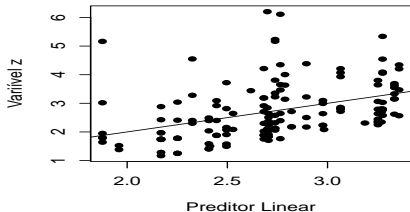
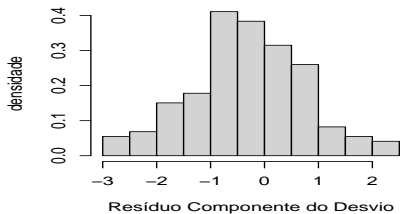
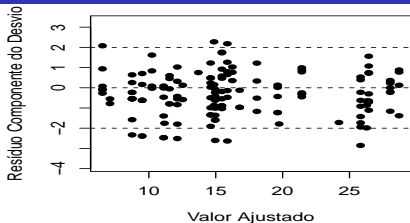
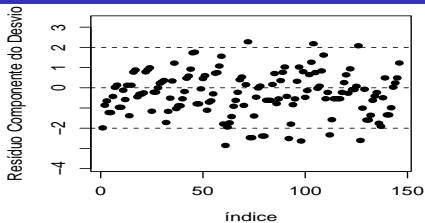
$$\ln \mu_{ijkl} = \alpha + \beta_i + \gamma_j + \delta_k + \lambda_l$$

$$\beta_1 = \gamma_1 = \delta_1 = \lambda_1 = 0$$

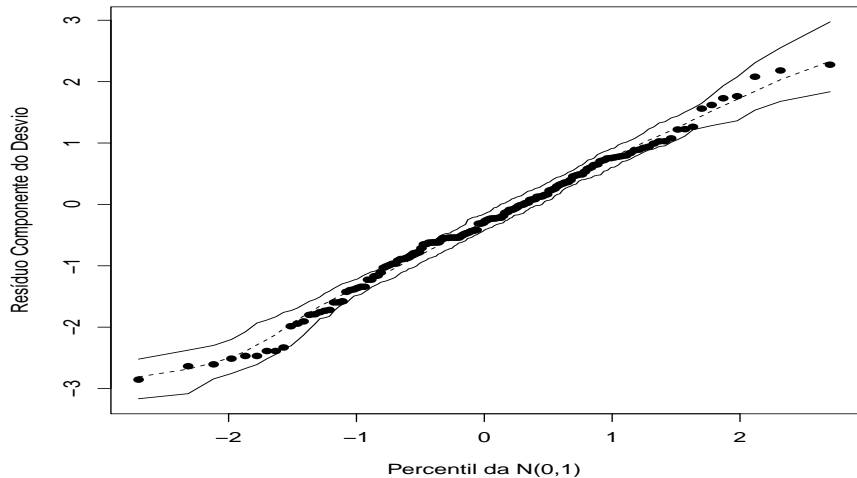
Gráficos de diagnóstico: modelo Poisson



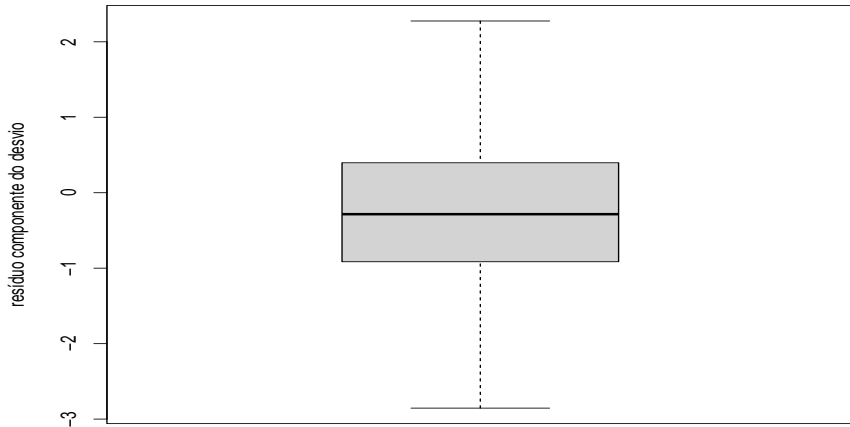
Gráficos de diagnóstico: modelo BN



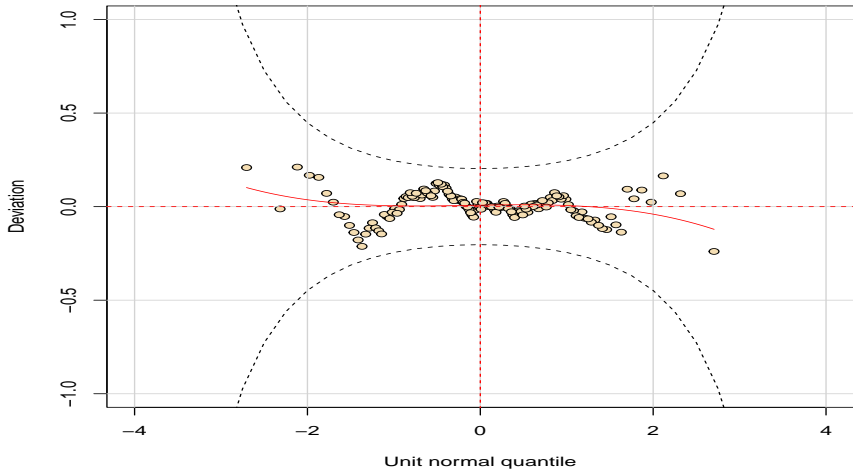
Gráficos de envelope: modelo BN



Boxplot do RCD modelo BN



Wormplot do RQA: modelo BN



Wormplots do RQA: modelo BN

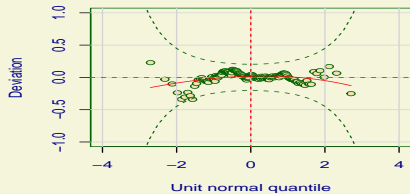
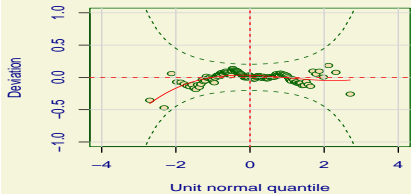
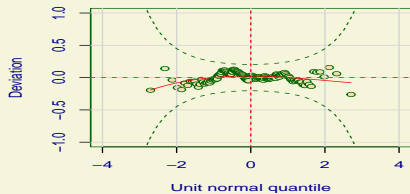
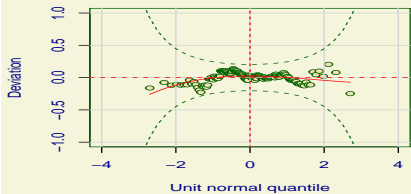
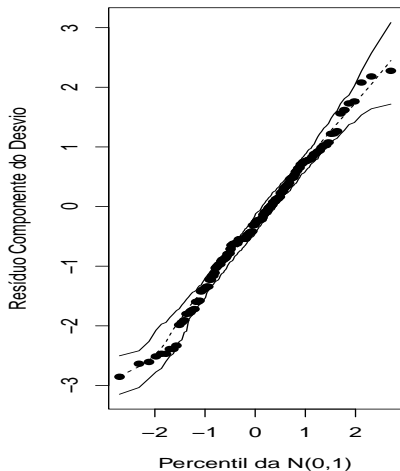
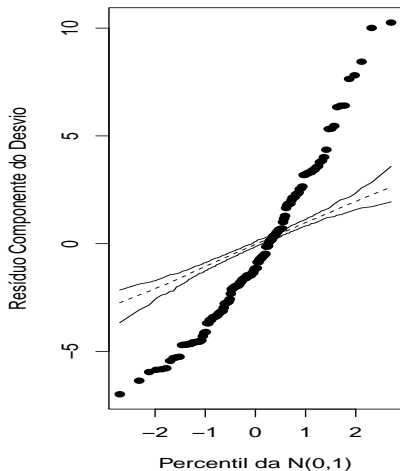
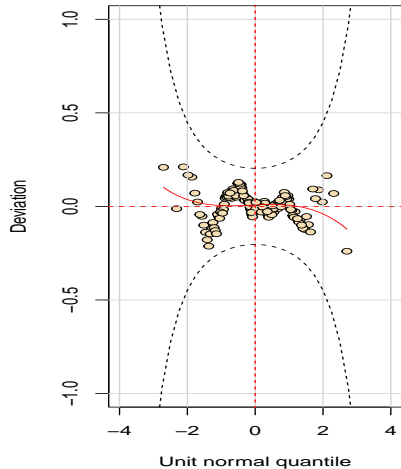
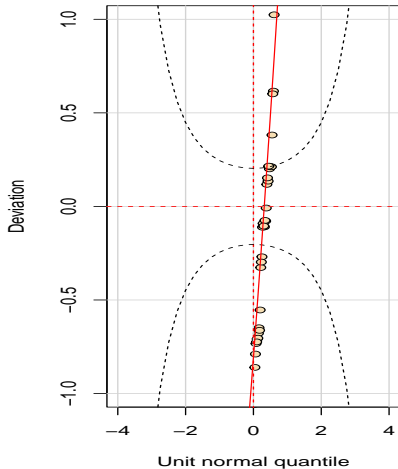


Gráfico de envelope: modelos Poisson e BN



Worm plots: modelos Poisson e BN



Estatísticas de ajuste e comparação de modelos

Modelo	AIC	BIC	AIC _c	SABIC	HQCIC	CAIC	p-valor
Poisson	2299,18	2320,07	2300,00	2297,92	2307,67	2327,07	< 0,0001
BN	1107,15	1128,04	1107,96	1105,89	1115,64	1135,04	0,0477

Comentários

- O modelo BN se ajusto bem aos dados e muito melhor do que o modelo de Poisson, segundo a análise de resíduos.
- Adicionalmente, do ponto de vista dos critérios de informação o modelo BN se ajustou muito melhor aos dados.
- Escolheremos o modelo BN.

Estimativas do modelo BN

Par.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
α	2,89	0,23	[2,45 ; 3,34]	12,67	<0,0001
β_2 (não-aborígene)	-0,57	0,15	[-0,87 ; -0,27]	-3,71	0,0002
γ_2 (masculino)	0,08	0,16	[-0,23 ; 0,40]	0,51	0,6067
δ_2 (1º ano)	-0,45	0,24	[-0,92 ; 0,02]	-1,87	0,0614
δ_3 (2º ano)	0,09	0,24	[-0,37 ; 0,55]	0,37	0,7092
δ_4 (3º ano)	0,36	0,25	[-0,13 ; 0,84]	1,44	0,1507
λ_2 (insuficiente)	0,29	0,19	[-0,07 ; 0,66]	1,57	0,1172
ϕ	1,27	0,16	[0,96 ; 1,59]	-	-

Estimativas do modelo BN

- O teste $C\beta = M$ para testar $H_0 : \gamma_2 = \lambda_2 = 0$ vs H_1 : há pelo menos uma diferença, indicou que “sexo” e “desempenho” não são significativos, $q_t = 2,54$ (0,2808).
- Portanto, ajustaremos um modelo sem esses dois fatores (modelo BN2).

Exemplo 12: Modelo BN2

$$Y_{ikr} \stackrel{ind.}{\sim} \text{BN}(\mu_{ik}, \phi),$$

$i = 1, 2$ (etnia: 1 - aborígene, 2 - não aborígene),

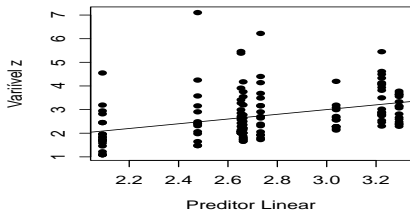
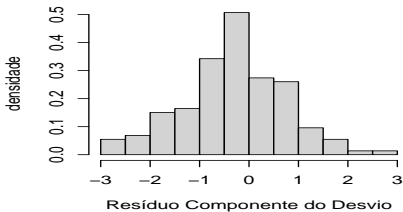
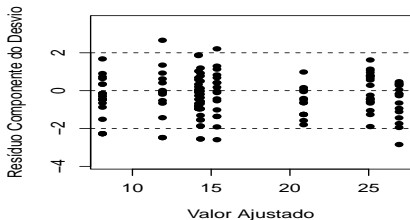
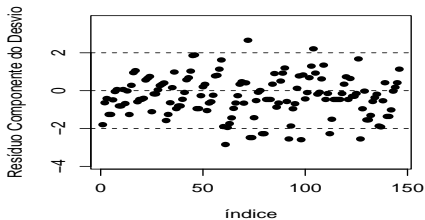
$k = 1, 2, 3, 4,$ (ano escolar: 1 - 8ª série, 2 - 1º ano, 3 - 2º ano,
4 - 3º ano)

$r = 1, 2, \dots, n_{ik}$ (unidades amostrais)

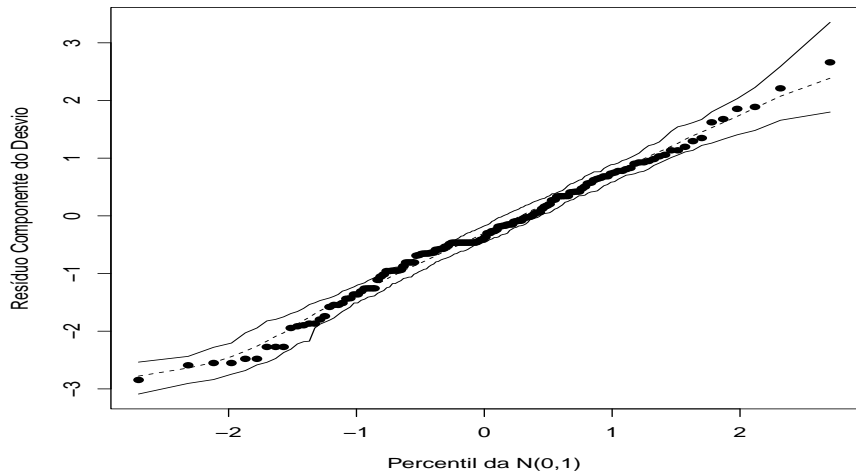
$$\ln \mu_{ik} = \alpha + \beta_i + \delta_k$$

$$\beta_1 = \delta_1 = 0$$

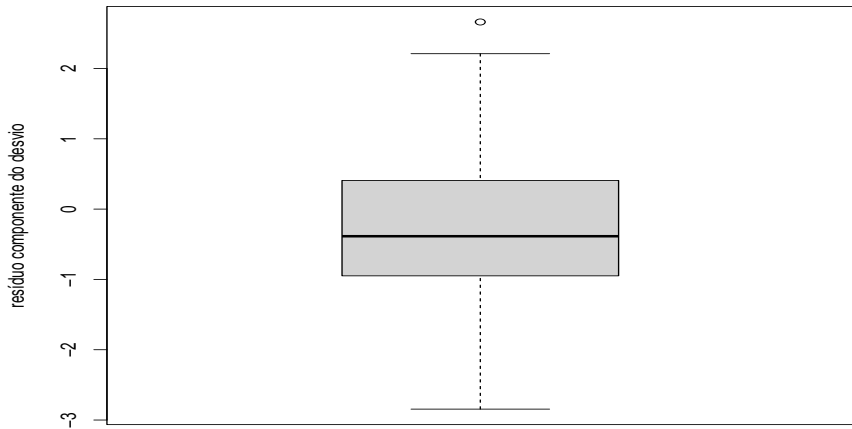
Gráficos de diagnóstico: modelo BN2



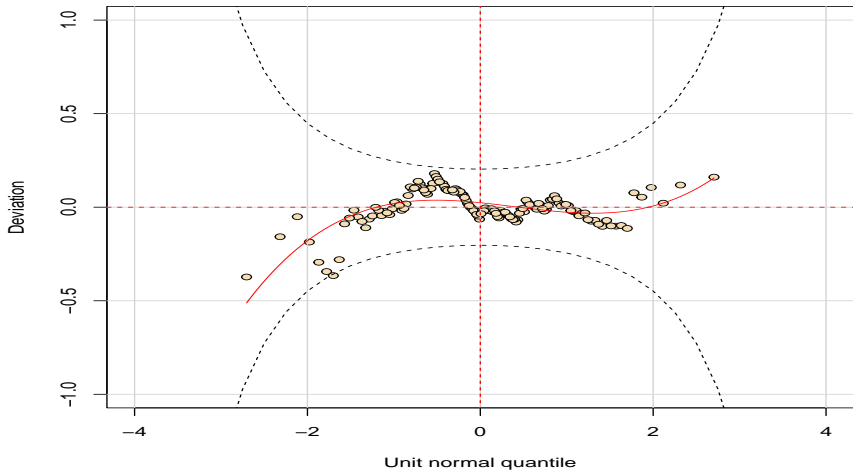
Gráficos de envelope: modelo BN2



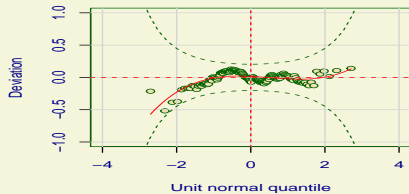
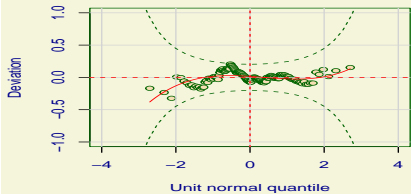
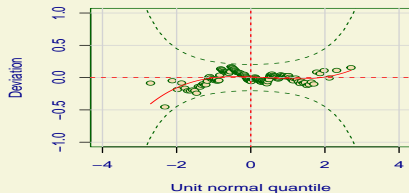
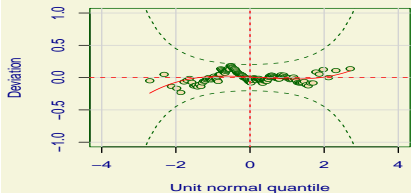
Boxplot do RCD modelo BN2



Wormplot do RQA: modelo BN2



Wormplots do RQA: modelo BN2



Estimativas do modelo BN2

Par.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
α	3,04	0,20	[2,65 ; 3,42]	15,53	< 0,0001
β_2 (não-aborígene)	-0,56	0,15	[-0,86 ; -0,26]	-3,63	0,0003
δ_2 (1º ano)	-0,39	0,23	[-0,83 ; 0,06]	-1,70	0,0900
δ_3 (2º ano)	0,18	0,23	[-0,27 ; 0,64]	0,80	0,4247
δ_4 (3º ano)	0,26	0,24	[-0,22 ; 0,73]	1,06	0,2893
ϕ	1,25	0,16	[0,94 ; 1,56]	-	-

Estimativas do modelo BN2

- O teste $\mathbf{C}\beta = \mathbf{M}$ para testar $H_0 : \delta_2 = \delta_3 = \delta_4 = 0$ vs $H_1 :$ há pelo menos uma diferença, indicou que uma significância (quase) marginal , $q_t = 11,68 (0,0086)$.
- Ajustaremos agora um modelo com esse dois fatores e interação (doravante modelo BNM3).

Exemplo 12: Modelo BN3

$$Y_{ikr} \stackrel{ind.}{\sim} \text{BN}(\mu_{ik}, \phi),$$

$i = 1, 2$ (etnia: 1 - aborígene, 2 - não aborígene),

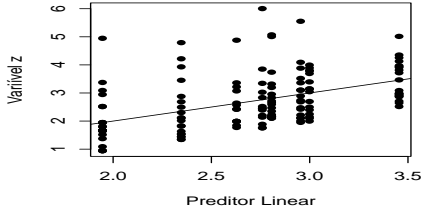
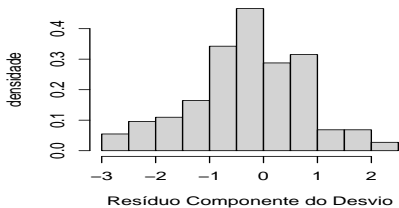
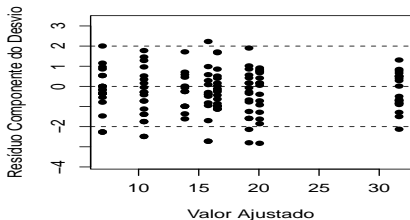
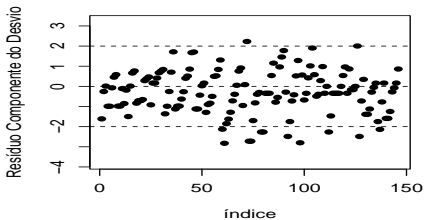
$k = 1, 2, 3, 4,$ (ano escolar: 1 - 8ª série, 2 - 1º ano, 3 - 2º ano,
4 - 3ºano)

$r = 1, 2, \dots, n_{ik}$ (unidades amostrais)

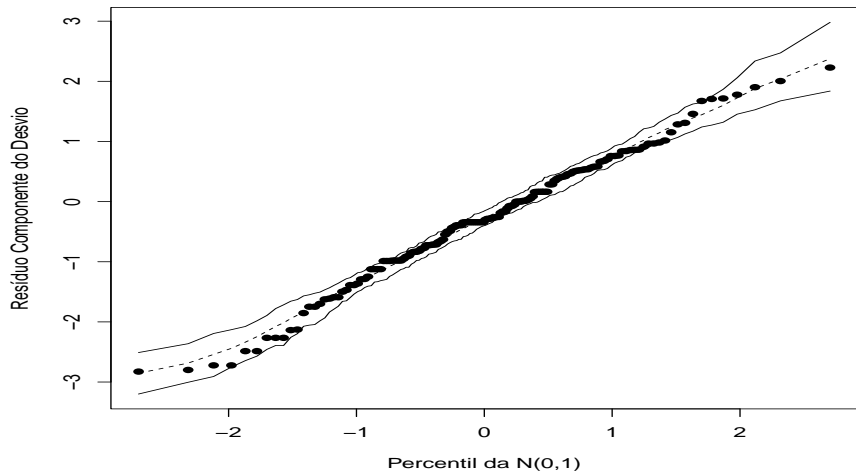
$$\ln \mu_{ik} = \alpha + \beta_i + \delta_k + (\beta\delta)_{ik}$$

$$\beta_1 = \delta_1 = (\beta\delta)_{1k} = (\beta\delta)_{i1} = 0, \forall i, k$$

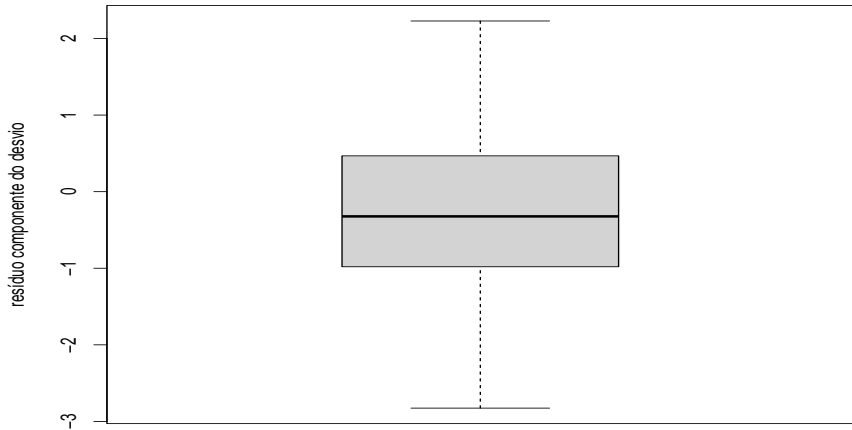
Gráficos de diagnóstico: modelo BN3



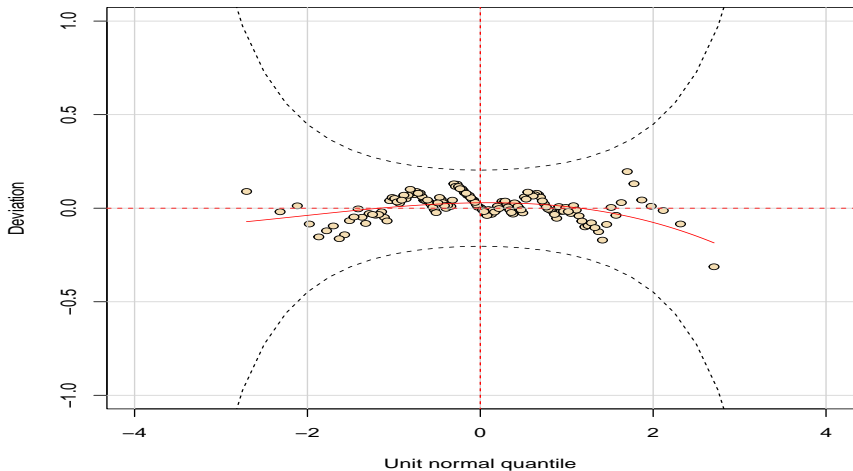
Gráficos de envelope: modelo BN3



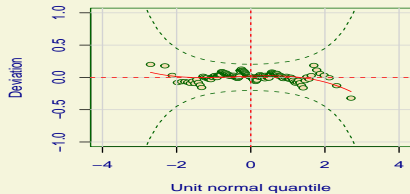
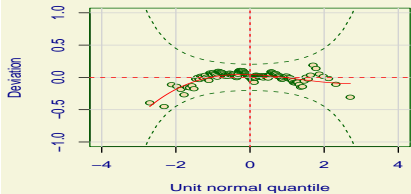
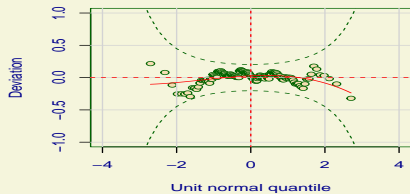
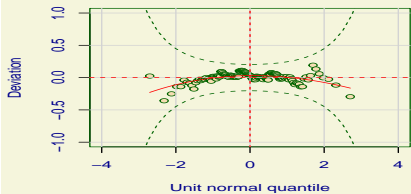
Boxplot do RCD modelo BN3



Wormplot do RQA: modelo BN3



Wormplots do RQA: modelo BN3



Estimativas modelo BN3

Par.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
α	2,63	0,25	[2,14 ; 3,12]	10,54	< 0,0000
β_2 (não-aborígene)	0,13	0,35	[-0,55 ; 0,81]	0,38	0,7044
δ_2 (1º ano)	0,18	0,32	[-0,45 ; 0,80]	0,56	0,5766
δ_3 (2º ano)	0,83	0,32	[0,20 ; 1,45]	2,61	0,0092
δ_4 (3º ano)	0,37	0,33	[-0,28 ; 1,02]	1,11	0,2665
$(\beta\delta)_{22}$	-0,99	0,44	[-1,85 ; -0,13]	-2,26	0,0204
$(\beta\delta)_{23}$	-1,24	0,45	[-2,11 ; -0,36]	-2,78	0,0055
$(\beta\delta)_{24}$	-0,18	0,46	[-1,08 ; 0,73]	-0,38	0,7038
ϕ	1,36	0,17	[1,02 ; 1,70]	-	-

Estimativas modelo BN3

- O teste $\mathbf{C}\boldsymbol{\beta} = \mathbf{M}$ para testar $H_0 : \beta_2 = \delta_2 = \delta_4 = (\boldsymbol{\beta}\boldsymbol{\delta})_{24} = 0$ vs H_1 : há pelo menos uma diferença, indicou que H_0 não pode ser rejeitada $q_t = 1,66(0,7985)$.
- Portanto, ajustaremos um modelo sem esses parâmetros, doravante modelo BN4.

Exemplo 12: Modelo BN4

$$Y_{ikr} \stackrel{ind.}{\sim} \text{BN}(\mu_{ik}, \phi),$$

$i = 1, 2$ (etnia: 1 - aborígene, 2 - não aborígene),

$k = 1, 2, 3, 4$, (ano escolar: 1 - 8ª série, 2 - 1º ano, 3 - 2º ano,
4 - 3ºano)

$r = 1, 2, \dots, n_{ik}$ (unidades amostrais)

$$\ln \mu_{ik} = \alpha + \delta_k + (\beta\delta)_{ik}$$

$$\beta_1 = \delta_1 = \delta_2 = \delta_4 = (\beta\delta)_{1k} = (\beta\delta)_{i1} = (\beta\delta)_{24} = 0, \forall i, k$$

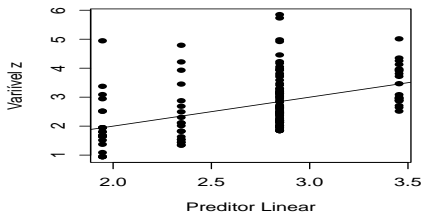
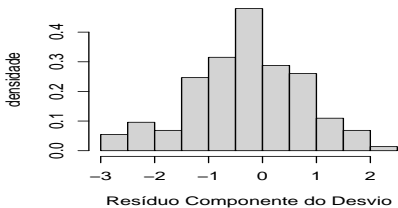
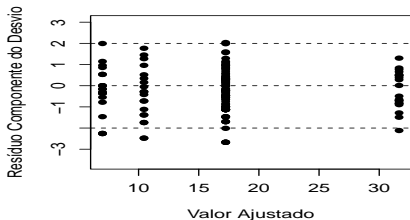
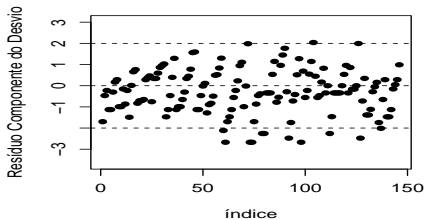
$$\beta = (\alpha, \delta_2, (\beta\delta)_{22}, (\beta\delta_{23}))'.$$

Médias a partir do modelo

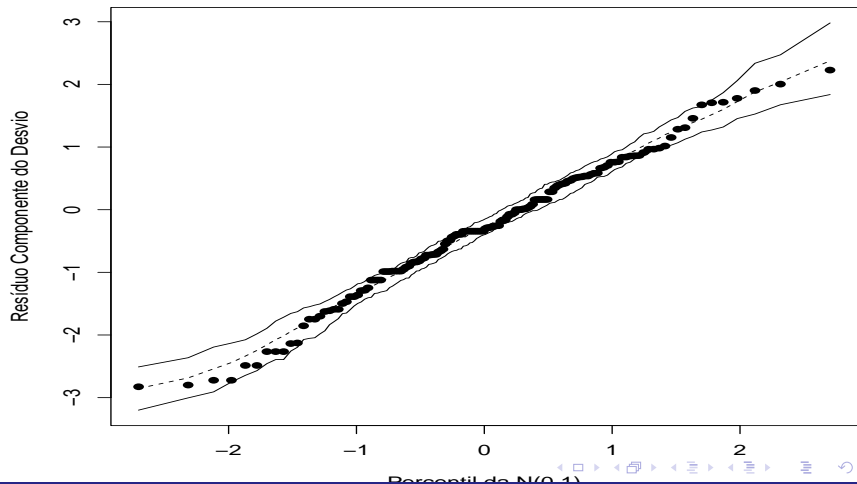
média	grupo (descrição)
$\mu_{11} = \mu_{12} = \mu_{14} = \mu_{21} = \mu_{24} = e^{\alpha}$	todos os alunos da 8ª série e 3º ano (independentemente dos outros fatores) e aborígenes do 1º ano - (grupo 1)
$\mu_{13} = e^{\alpha + \delta_3}$	aborígenes do 2º ano - (grupo 2)
$\mu_{22} = e^{\alpha + (\beta\delta)_{22}}$	não-aborígenes do 1º ano (grupo 3)
$\mu_{23} = e^{\alpha + \delta_3 + (\beta\delta)_{23}}$	não-aborígenes do 2º ano (grupo 4)

independentemente dos outros fatores (sexo e desempenho)

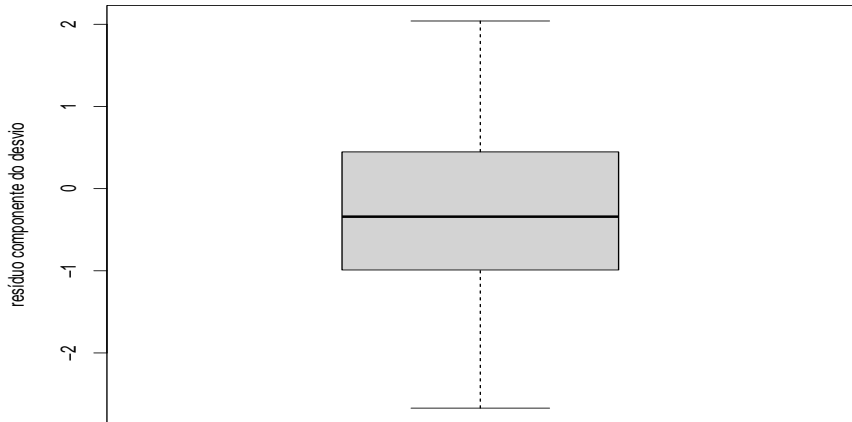
Gráficos de diagnóstico: modelo BN4



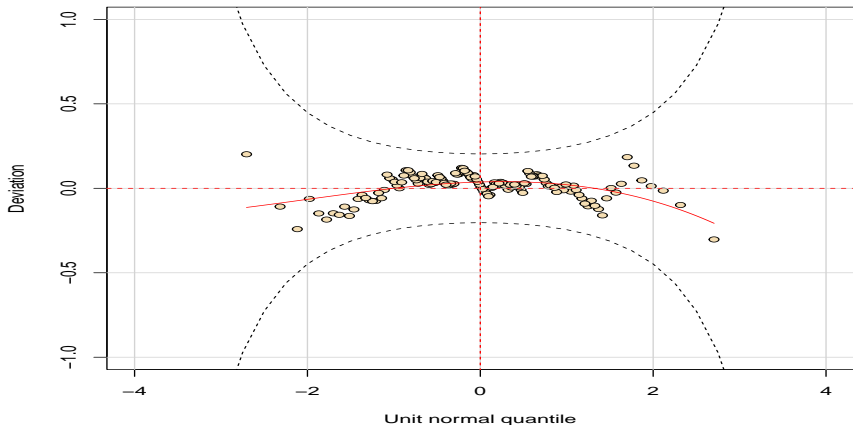
Gráficos de envelope: modelo BN4



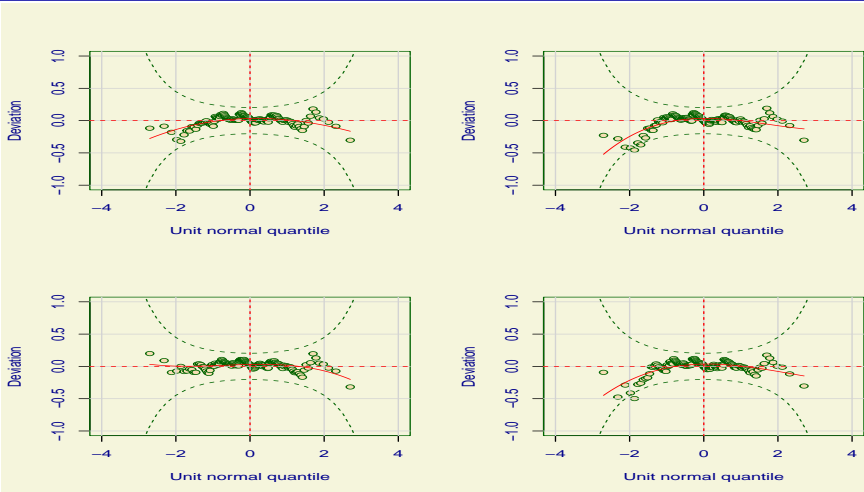
Boxplot do RCD modelo BN4



Wormplot do RQA: modelo BN4



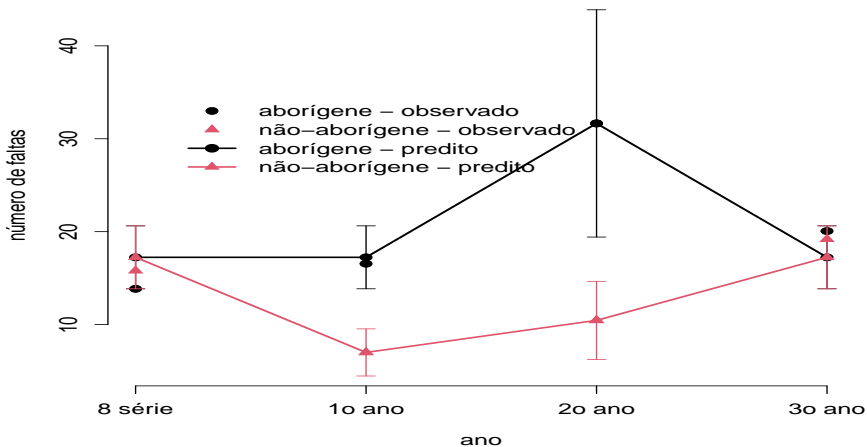
Wormplots do RQA: modelo BN4



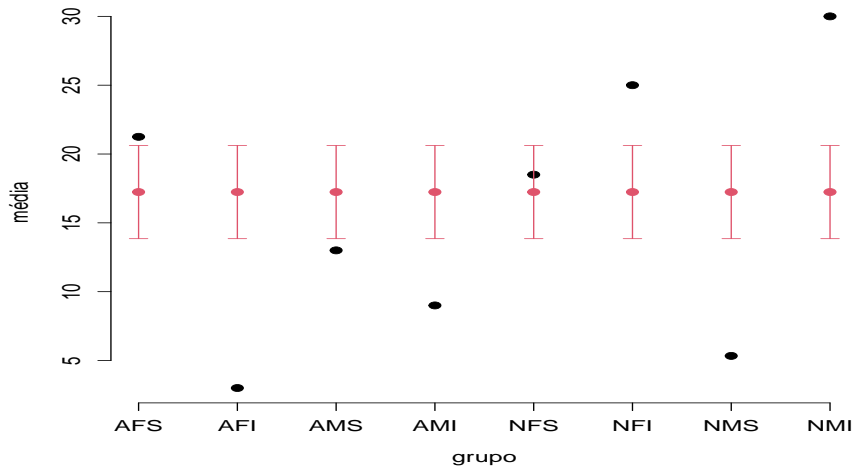
Estimativa do modelo BN4

Par.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
α	2,85	0,10	[2,65 ; 3,04]	28,40	< 0,0001
δ_3	0,61	0,22	[0,17 ; 1,04]	2,75	0,0060
$(\beta\delta)_{22}$	-0,90	0,21	[-1,31 ; -0,49]	-4,28	< 0,0001
$(\beta\delta)_{23}$	-1,11	0,28	[-1,67 ; -0,55]	-3,89	< 0,0001
ϕ	1,36	0,17	[1,02;1,70]	-	-

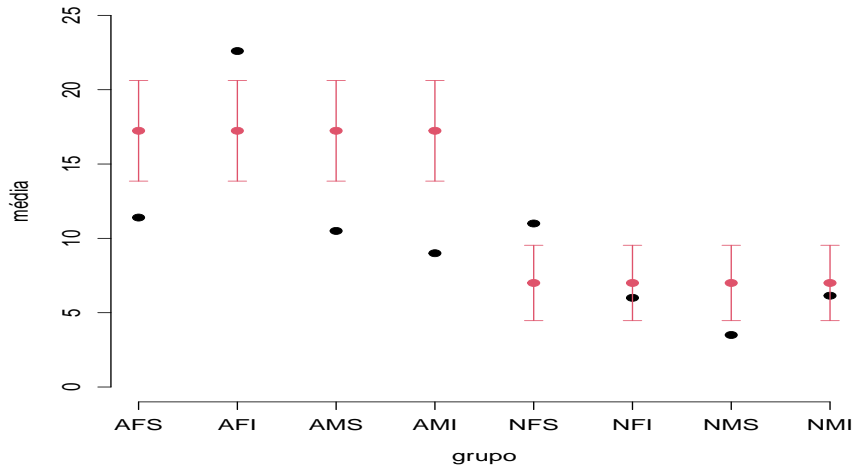
Médias observadas e previstas pelo modelo BN4



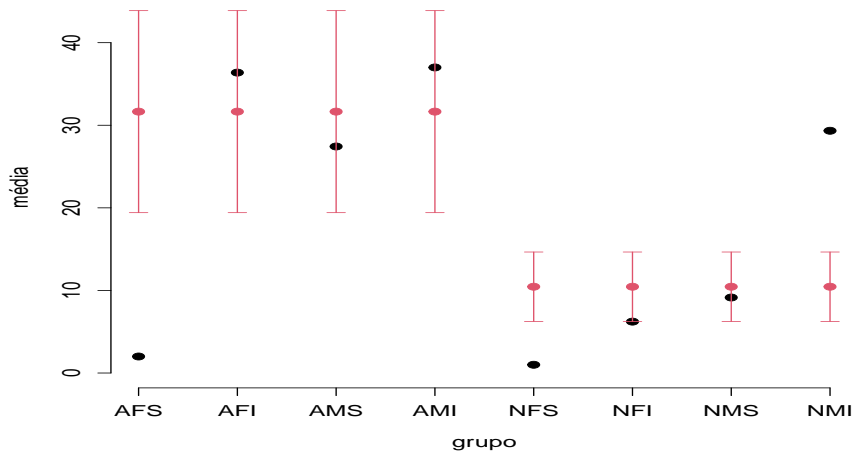
Médias observadas e previstas pelo modelo BN4 (8ª série)



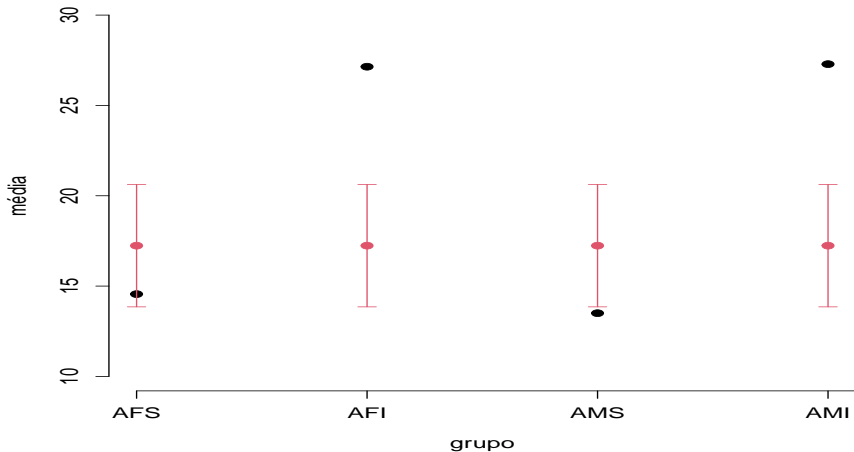
Médias observadas e previstas pelo modelo BN4 (1º ano)



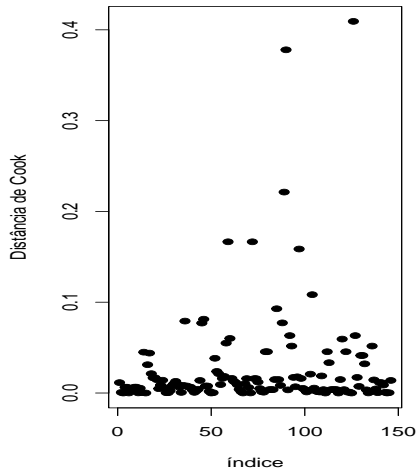
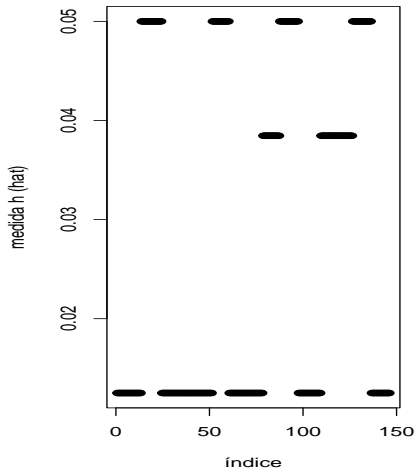
Médias observadas e previstas pelo modelo BN4 (2º ano)



Médias observadas e previstas pelo modelo BN4 (3º ano)



Análise de Influência



Análise de sensibilidade

Parâmetro (α)

Obs.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
1	2,85	0,10	[2,65;3,04]	28,40	< 0,0001
2	2,85	0,10	[2,65;3,04]	28,62	< 0,0001
3	2,85	0,10	[2,65;3,04]	28,70	< 0,0001
4	2,85	0,10	[2,65;3,04]	28,93	< 0,0001

Análise de sensibilidade

Parâmetro (δ_3)

Obs.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
1	0,61	0,22	[0,17;1,04]	2,75	0,0060
2	0,61	0,22	[0,18;1,04]	2,77	0,0056
3	0,61	0,22	[0,18;1,04]	2,78	0,0055
4	0,61	0,22	[0,18;1,03]	2,80	0,0051

Análise de sensibilidade

Parâmetro $((\beta\delta)_{22})$

Obs.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
1	-0,90	0,21	[-1,31;-0,49]	-4,28	< 0,0001
2	-0,90	0,21	[-1,31;-0,49]	-4,32	< 0,0001
3	-1,03	0,21	[-1,45;-0,61]	-4,82	< 0,0001
4	-1,03	0,21	[-1,44;-0,61]	-4,86	< 0,0001

Análise de sensibilidade

Parâmetro $((\beta\delta)_{23})$

Obs.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
1	-1,11	0,28	[-1,67;-0,55]	-3,89	0,0001
2	-1,25	0,29	[-1,81;-0,68]	-4,33	< 0,0001
3	-1,11	0,28	[-1,66;-0,56]	-3,93	0,0001
4	-1,25	0,28	[-1,80;-0,69]	-4,38	< 0,0001

Comentários (modelo BN4)

- O modelo BN4 se ajustou bem aos dados, do ponto de vista residual.
- A retirada de nenhum dos pontos destacados na análise de influência levou à mudança significativa nos resultados,
- As perdições do número médio de faltas poderia melhorar com a seleção de outras variáveis (presentes no modelo) e/ou de parâmetros (ou ainda a consideração de outras covariáveis).