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).
- Quando o modelo de regressão define grupos (covariáveis qualitativas, como no Exemplo 4) 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.

Cont.

- Do ponto de vista inferencial, quando o desvio $D(\boldsymbol{y};\widetilde{\boldsymbol{\mu}})$ 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, ausência de covariáveis relevantes 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 de superdispersão (Hinde e Demétrio (1998)

- Os erros-padrão estimados (a partir do modelo) podem ser (muito) subestimados consequentemente, 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

- Consideraremos uma abordagem que considera variáveis latentes.
- Suponha que para um conjunto fixo $\boldsymbol{x} = (x_1, ..., x_p)'$ 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,...,n) com $\boldsymbol{x}$ fixo, de modo que $E(Z) = \mu$.
- Assim $\mathcal{E}(Y) = \mathcal{E}(\mathcal{E}(Y|Z)) = \mathcal{E}(Z) = \mu$ e $\mathcal{V}(Y) = \mathcal{E}(\mathcal{V}(Y|Z)) + \mathcal{V}(\mathcal{E}(Y|Z)) = \mu + \mathcal{V}(Z)$.

Modelagem da super-dispersão

- 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 $\mathcal{V}(Y) = \mu + \mu^2/\phi$.
- Note que se $\phi \to \infty$ então $\mathcal{V}(Y) \to \mathcal{E}(Y) = \mu$ (à rigor teremos o modelo de Poisson).
- Além disso, $f(y|z) = \dfrac{e^{-z}z^y}{y!}\mathbb{1}_{\{0,1,2,\ldots\}}(y)$ 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

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

- 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 $Y_i, i = 1, 2, ..., n$, $Y_i \overset{ind.}{\sim} \text{BN}(\mu_i, \phi)$.
- Temos que $\mathcal{E}(Y_i) = \mu_i$ e $\mathcal{V}(Y_i) = \mu_i + \dfrac{\mu_i^2}{\phi}$.
- $g(\mu_i) = \eta_i \rightarrow \mu_i = g^{-1}(\eta_i)$, $\eta_i = \boldsymbol{X}_i\boldsymbol{\beta}$, $g(.)$ 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 $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi)$).

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

Estimação por MV

- Vetor escore para $\boldsymbol{\beta}$ (cada componente)

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

- 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

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

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

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

- Podemos notar que o sistema $\begin{cases} \boldsymbol{S}(\widetilde{\boldsymbol{\beta}}) = \mathbf{0}_{(p \times 1)} \\ S(\widetilde{\phi}) = 0 \end{cases}$ 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

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

- Informação de Fisher

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

Estimação por MV

- Matricialmente, temos $\boldsymbol{I}(\boldsymbol{\beta}, \boldsymbol{\beta}) = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}$.
- 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 $\boldsymbol{\beta}$ e ϕ são ortogonais, ou seja, $\boldsymbol{I}(\boldsymbol{\beta}, \phi) = \mathbf{0}_{(p \times 1)}$.

Estimação por MV

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

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

o algoritmo escore de Fisher, nesse caso, terá de ser conduzido, simultaneamente, para os dois parâmetros, pois $\boldsymbol{W}$ 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

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

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

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

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

$t = 0, 1, 2, ...,$ até que algum critério de convergência seja satisfeito. em que $\boldsymbol{y}^* = \boldsymbol{X}\beta + \boldsymbol{F}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$, $I_O(\phi,\phi) = \sum_{i=1}^{n}\left\{\Psi'(\phi + y_i) + \frac{(y_i - 2\mu_i - \phi)}{(\phi + \mu_i)^2}\right\} + n\phi^{-1}\{1 - \phi\Psi'(\phi)\}$ 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 $\Psi'(.)$ é a função trigama.

- Quantidades ω_i e f_i para algumas funções de ligação:

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

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

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

Inferência

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

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

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

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

Inferência

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

$$Q_t = \left(\boldsymbol{C}\widehat{\beta} - \boldsymbol{M}\right)' \left(\boldsymbol{C}\widehat{\mathcal{V}}(\widehat{\beta})\boldsymbol{C}'\right)^{-1} \left(\boldsymbol{C}\widehat{\beta} - \boldsymbol{M}\right).$$

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

Inferência

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

$$Q_t^* = \left(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{M}\right)' \left(\boldsymbol{C}\mathcal{V}(\widehat{\boldsymbol{\beta}})\boldsymbol{C}'\right)^{-1} \left(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{M}\right),$$

então sob H_0 e para n suficientemente grande, $Q_t^* \approx \chi^2_c$, em que c é o número de linhas de $\boldsymbol{C}$. Além disso, $\widehat{\boldsymbol{W}} \xrightarrow[n\to\infty]{P} \boldsymbol{W}$ o que implica que $\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}}) \xrightarrow[n\to\infty]{P} \mathcal{V}(\widehat{\boldsymbol{\beta}})$, pois cada componente de $\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}})$ é uma função contínua da respectiva componente de $\widehat{\boldsymbol{W}}$ e também por Slutsky pois $\widehat{\phi} \xrightarrow[n\to\infty]{P} \phi$.

Inferência

- Portanto, sob H_0 e para n suficientemente grande, pelos resultados anteriores $(Q_t^*, \widehat{\boldsymbol{W}})$, $Q_t \approx \chi^2_c$.
- Assim, rejeita-se H_0 se $p - valor \leq \alpha$, em que $p - valor \approx P(X \geq q_t | H_0)$, em que $X \sim \chi^2_c$
$q_t = \left(\boldsymbol{C}\widetilde{\boldsymbol{\beta}} - \boldsymbol{M}\right)' \left(\boldsymbol{C}\widetilde{\mathcal{V}}(\widehat{\boldsymbol{\beta}})\boldsymbol{C}'\right)^{-1} \left(\boldsymbol{C}\widetilde{\boldsymbol{\beta}} - \boldsymbol{M}\right)$.
- Sob H_1, temos que $Q_t \approx \chi^2_{(c,\delta)}$ (qui-quadrado não central com c graus de liberdade e parâmetro de assimetra δ), em que $\delta = (\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{M})' \left(\boldsymbol{C}\mathcal{V}(\widehat{\boldsymbol{\beta}})\boldsymbol{C}'\right)^{-1} (\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{M})$. Uma estimativa de δ é dada por $\widetilde{\delta} = \left(\boldsymbol{C}\widetilde{\boldsymbol{\beta}} - \boldsymbol{M}\right)' \left(\boldsymbol{C}\widetilde{\mathcal{V}}(\widehat{\boldsymbol{\beta}})\boldsymbol{C}'\right)^{-1} \left(\boldsymbol{C}\widetilde{\boldsymbol{\beta}} - \boldsymbol{M}\right)$.

Diagnóstico do modelo aqui

- A i-ésima componente do desvio do modelo para ϕ fixo é dada por $D^{*}(y_i,\widehat{\mu}_i) = 2\left\{\phi \ln\left[\frac{\widehat{\mu}_i+\phi}{y_i+\phi}\right] + y_i \ln\left[\frac{y_i(\widehat{\mu}_i+\phi)}{\widehat{\mu}_i(y_i+\phi)}\right]\right\}$, $\forall\, y_i \in \{1,2,...,\}$.

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

- Se $y_i = 0$ então

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

Diagnóstico do modelo

- Assim o desvio é dado por

$$
\begin{aligned}
D^*(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) &= \sum_{i=1}^{n} \left\{ \left[\phi \ln \left[\frac{\widehat{\mu}_i + \phi}{y_i + \phi} \right] + y_i \ln \left[\frac{y_i(\widehat{\mu}_i + \phi)}{\widehat{\mu}_i(y_i + \phi)} \right] \right] \mathbb{1}_{\{1,2,\ldots,\}}(y_i) \right. \\
&+ \left. \phi \ln \{ (\widehat{\mu}_i + \phi)/\phi \} \mathbb{1}_{\{0\}}(y_i) \right\}
\end{aligned}
$$

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

$$
\begin{aligned}
D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) &= \sum_{i=1}^{n} \left\{ \left[\widehat{\phi} \ln \left[\frac{\widehat{\mu}_i + \widehat{\phi}}{y_i + \widehat{\phi}} \right] + y_i \ln \left[\frac{y_i(\widehat{\mu}_i + \widehat{\phi})}{\widehat{\mu}_i(y_i + \widehat{\phi})} \right] \right] \mathbb{1}_{\{1,2,\ldots,\}}(y_i) \right. \\
&+ \left. \widehat{\phi} \ln \left\{ (\widehat{\mu}_i + \widehat{\phi})/\widehat{\phi} \right\} \mathbb{1}_{\{0\}}(y_i) \right\}
\end{aligned}
$$

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

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

em que

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

- e

  $\widehat{h}_{ii}$ é o i-ésimo elemento da diagonal principal da matriz,
  $\widehat{\boldsymbol{H}} = \widehat{\boldsymbol{W}}^{1/2}\boldsymbol{X}(\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{W}}^{1/2}$, ou seja,
  $\widehat{h}_{ii} = \frac{\widehat{\phi}\widehat{\mu}_i}{\widehat{\phi} + \widehat{\mu}_i}\boldsymbol{X}_i'\left(\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}_i$.
- 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.

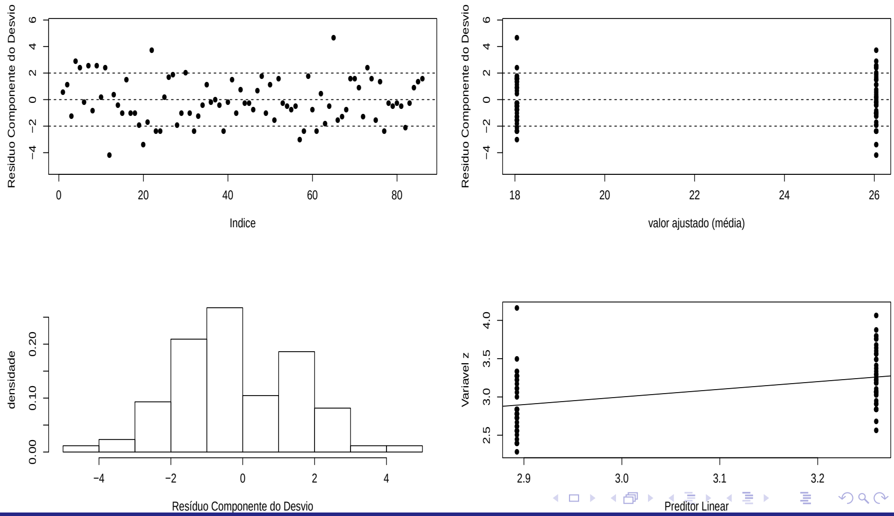
Voltando ao Exemplo 4: comparação do número de acidentes

Modelo

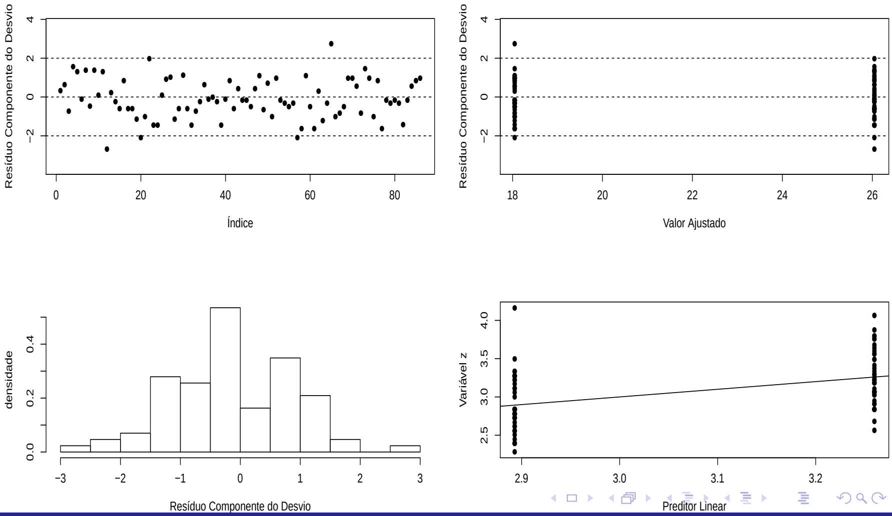
$$
\begin{aligned}
Y_{ij} &\overset{ind.}{\sim} \text{BN}(\mu_i, \phi), i = 1\,(\text{ano de } 1961), 2\,(\text{ano de } 1962), j = 1, ..., 43 \\
\ln \mu_i &= \mu + \alpha_i, \alpha_1 = 0
\end{aligned}
$$

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

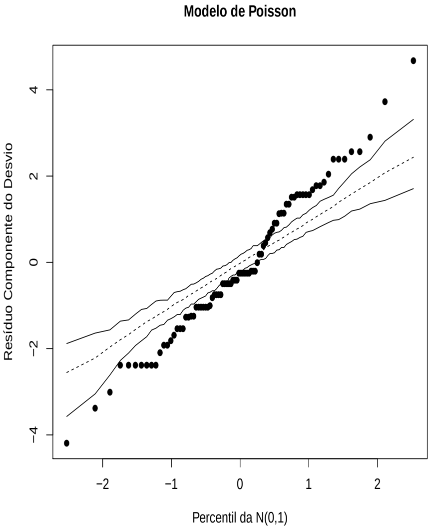
Gráficos de diagnóstico: Poisson

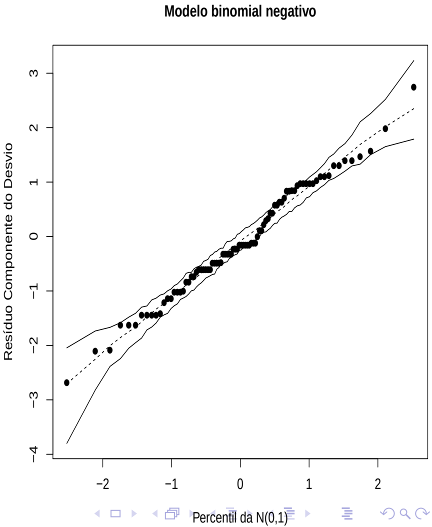


Gráficos de diagnóstico: binomial negativo

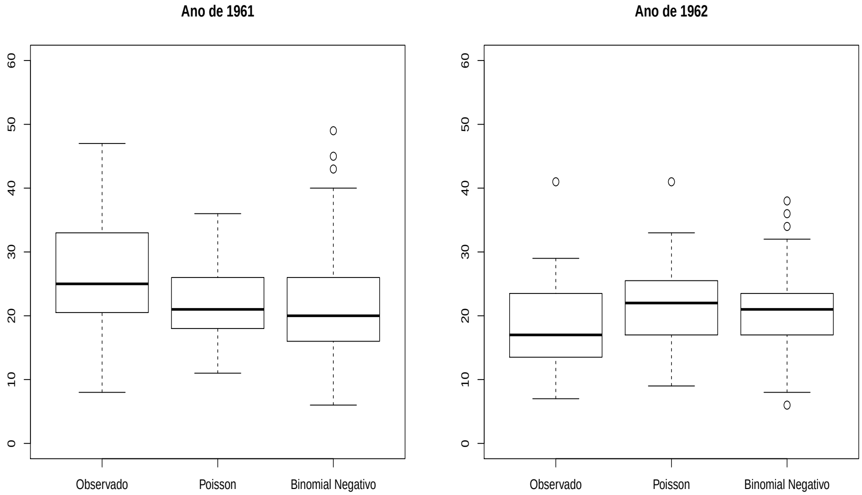


Gráficos de envelopes: Poisson e binomial negativo





Distribuições preditas e observadas (boxplot)



Estatísticas de ajuste e comparação de modelos

Modelo	AIC	BIC	Desvio	p-valor desvio
Poisson	656,94	661,85	235,17	$<$0,0001
Binomial negativo	595,37	602,73	86,66	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]	-	-