

Modelos de regressão para dados discretos (parte 5) (sub/superdispersão): dados binários

Prof. Caio Azevedo

Superdispersão (ou sobredispersão) e subdispersão

- Quando a variável resposta apresenta variância maior (sobre/super) ou menor (sub) do que aquela imposta pelo modelo probabilístico.
- No caso do modelos de regressão para dados binários, em que assume-se que $Y_i \stackrel{ind.}{\sim} \text{binomial}(m_i, \mu_i)$, $i = 1, 2, \dots, n$, tem-se que $\mathcal{V}(Y_i) = m_i \mu_i (1 - \mu_i)$. No caso do Exemplo 8 (dos besouros), $n = 8$ e $m_i \in \{59, 60, 62, , 56, , 63, 59, 62, 60\}$
- Portanto, espera-se que as variâncias amostrais não sejam distantes desse valor.
- Entretanto, do ponto de vista descritivo, muitas vezes não é simples verificar se se isso ocorre a menos que disponhamos das observações individuais que geraram as contagens binomiais.



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, 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 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).

Caso I

- Vamos supor inicialmente a existência de n grupos (contagens binomiais) de modo que para o i -ésimo grupo sejam observadas m_i repetições de uma variável aleatória $Y_{ij} \sim \text{Bernoulli}(\mu_i)$, assim, $\mathcal{E}(Y_{ij}) = \mu_i$ e $\mathcal{V}(Y_{ij}) = \mu_i(1 - \mu_i)$, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m_i$. O número total de sucessos no i -ésimo grupo será definido por

$$Y_i = Y_{i1} + \dots + Y_{im_i}.$$

- Supondo, adicionalmente, a existência de correlação entre as réplicas de Bernoulli (de sorte que $\text{Corre}(Y_{ij}, Y_{ik}) = \delta$, $\delta \in (-1, 1)$, $\forall j \neq k$, assim $\text{Cov} = \delta\mu_i(1 - \mu_i)$), então

Caso I

$$\begin{aligned}\mathcal{V}(Y_i) &= \sum_{j=1}^{m_i} \mathcal{V}(Y_{ij}) + \sum_{j=1}^{m_i} \sum_{k=1, k \neq j}^{m_i} \text{Cov}(Y_{ij}, Y_{ik}) \\ &= \sum_{j=1}^{m_i} \mu_i(1 - \mu_i) + \sum_{j=1}^{m_i} \sum_{k=1, k \neq j}^{m_i} \delta \mu_i(1 - \mu_i) \\ &= m_i \mu_i(1 - \mu_i) + m_i(m_i - 1) \delta \mu_i(1 - \mu_i) \\ &= \sigma_i^2 m_i \mu_i(1 - \mu_i)\end{aligned}$$

em que $\sigma_i^2 = 1 + (m_i - 1)\delta$.

Caso I

- Se for exigido que $\sigma_i^2 > 0$, então devemos ter $1 + (m_i - 1)\delta > 0$, o que implica que $\delta > -1/(m_i - 1)$ e, portanto, teremos $-\frac{1}{m_i - 1} < \delta < 1$.
- Logo, δ poderá assumir valores negativos apenas se houver m_i 's pequenos.
- Podemos notar que

$$\left\{ \begin{array}{l} \text{se } \sigma_i^2 > 1, \text{ tem-se superdispersão} \\ \text{caso contrário tem-se subdispersão} \end{array} \right.$$

Caso II

- Vamos supor agora que p_i representa a probabilidade de sucesso nas respostas do i -ésimo grupo tal que $\mathcal{E}(P_i) = \mu_i$ e $\mathcal{V}(P_i) = \delta\mu_i(1 - \mu_i)$, $\delta \geq 0$.
- A quantidade P_i , neste caso, é dita ser um efeito aleatório.
- Assumimos ainda que $Y_{ij}|p_i \sim \text{Bernoulli}(p_i)$, o que implica que $\mathcal{E}(Y_{ij}|p_i) = p_i$ e $\mathcal{V}(Y_{ij}|p_i) = p_i(1 - p_i)$. Portanto, temos que:

$$\begin{aligned}\mathcal{E}(Y_i) &= \mathcal{E}(\mathcal{E}(Y_i|P_i)) = \mathcal{E}\left(\mathcal{E}\left(\sum_{j=1}^{m_i} Y_{ij}|P_i\right)\right) = \mathcal{E}\left(\sum_{j=1}^{m_i} \mathcal{E}(Y_{ij}|P_i)\right) \\ &= m_i\mathcal{E}(P_i) = m_i\mu_i\end{aligned}$$

- Além disso, temos que (lembrando que $\mathcal{E}(P_i^2) = \mathcal{V}(P_i) + \mathcal{E}^2(P_i)$):

$$\begin{aligned}\mathcal{V}(Y_i) &= \mathcal{E}[\mathcal{V}(Y_i|P_i)] + \mathcal{V}(\mathcal{E}(Y_i|P_i)) \\ &= \mathcal{E}\left[\mathcal{V}\left(\sum_{j=1}^{m_i} Y_{ij}|P_i\right)\right] + \mathcal{V}\left[\mathcal{E}\left(\sum_{j=1}^{m_i} Y_{ij}|P_i\right)\right] \\ &= \mathcal{E}\left[\sum_{j=1}^{m_i} \mathcal{V}(Y_{ij}|P_i)\right] + \mathcal{V}\left[\sum_{j=1}^{m_i} \mathcal{E}(Y_{ij}|P_i)\right] \\ &= \mathcal{E}\left[\sum_{j=1}^{m_i} P_i(1 - P_i)\right] + \mathcal{V}\left[\sum_{j=1}^{m_i} P_i\right] \\ &= m_i (\mathcal{E}(P_i) - \mathcal{E}(P_i^2)) + m_i^2 \mathcal{V}(P_i) \\ &= m_i (\mu_i(1 - \delta)(1 - \mu_i)) + m_i^2 \delta \mu_i(1 - \mu_i) \\ &= m_i \mu_i(1 - \mu_i) [1 + \delta(m_i - 1)]\end{aligned}$$

Modelagem

- Devido à estrutura hierárquica apresentada anteriormente, uma candidato natural para o modelo probabilístico é a distribuição beta-binomial.
- Seja $Y|X = x \sim \text{binomial}(m, x)$ e $X \sim \text{beta}(a, b)$.
- Então $Y \sim \text{beta-binomial}(m, a, b) \equiv \text{bb}(m, a, b)$ (exercício), em que

$$f_y(y) = \frac{\Gamma(m+1)}{\Gamma(y+1)\Gamma(m-y+1)} \frac{\Gamma(y+a)\Gamma(m-y+b)}{\Gamma(m+a+b)} \\ \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mathbb{1}_{\{0,1,2,3,\dots,m\}}(y)$$

Modelagem

- Temos que $\mathcal{E}(Y) = m \frac{a}{(a+b)}$ e $\mathcal{V}(Y) = \frac{mab(a+b+m)}{(a+b)^2(a+b+1)}$ (exercício)
- Podemos definir uma reparametrização da distribuição beta-binomial de tal forma que $\mathcal{E}(Y) = m\mu$ e que se tenha um parâmetro de dispersão (σ), ou seja, $a = \frac{\mu}{\sigma}$ e $b = \frac{1-\mu}{\sigma}$.

$$f(y) = \frac{\Gamma(m+1)}{\Gamma(y+1)\Gamma(m-y+1)} \frac{\Gamma\left(y + \frac{\mu}{\sigma}\right) \Gamma\left(\frac{1}{\sigma}\right) \Gamma\left(m + \frac{1-\mu}{\sigma} - y\right)}{\Gamma\left(m + \frac{1}{\sigma}\right) \Gamma\left(\frac{\mu}{\sigma}\right) \Gamma\left(\frac{1-\mu}{\sigma}\right)}$$

Modelagem

- Nesse caso, $\mathcal{E}(Y) = \mu$ e
$$\mathcal{V}(Y) = \frac{m\mu(1-\mu)(1+m\sigma)}{1+\sigma} = m\mu(1-\mu) \left[1 + \frac{(m-1)\sigma}{1+\sigma} \right],$$
 em que σ é um parâmetro de dispersão.
- O interesse agora reside em modelar μ_i , mantendo-se σ fixo ao longo das observações.
- Em termos práticos, à semelhança do modelo binomial, trabalhos com a verossimilhança associada à $Y_i^* = \frac{Y_i}{m_i}$.

Modelo de regressão beta binomial

- Sejam $Y_i, i = 1, 2, \dots, n$, $Y_i \stackrel{ind.}{\sim} \text{bb}(m_i, \mu_i, \sigma)$.
- Temos que $\mathcal{E}(Y_i) = \mu_i$ e $\mathcal{V}(Y_i) = m_i \mu_i (1 - \mu_i) \left[1 + \frac{(m_i - 1)\sigma}{1 + \sigma} \right]$.
- $g(\mu_i) = \eta_i \rightarrow \mu = g^{-1}(\eta_i)$, $\eta_i = \mathbf{X}_i \boldsymbol{\beta}$ em que $g(\cdot)$, em geral corresponde à $F^{-1}(\cdot)$, ou seja, à inversa da fda associada à uma vac com suporte nos reais.
- Este modelo de regressão pertence à classe de modelos chamada de GAMLSS (“Generalized Additive Models for Location Scale and Shape”). Existe um pacote no R homônimo (gamlss).

Inferência via MV

- Verossimilhança (em que $\theta = (\beta', \sigma)'$)

$$L(\theta) = \frac{\prod_{i=1}^n \Gamma(m_i + 1)}{\prod_{i=1}^n [\Gamma(y_i + 1)\Gamma(m_i - y_i + 1)]} \times \frac{\left[\Gamma\left(\frac{1}{\sigma}\right) \right]^n \prod_{i=1}^n \left[\Gamma\left(y_i + \frac{\mu_i}{\sigma}\right) \Gamma\left(m_i + \frac{1 - \mu_i}{\sigma} - y_i\right) \right]}{\prod_{i=1}^n \left[\Gamma\left(m_i + \frac{1}{\sigma}\right) \Gamma\left(\frac{\mu_i}{\sigma}\right) \Gamma\left(\frac{1 - \mu_i}{\sigma}\right) \right]}$$

Inferência via MV

- Verossimilhança (em que $\theta = (\beta', \phi)'$)

$$L(\theta) \propto \frac{\left[\Gamma\left(\frac{1}{\sigma}\right) \right]^n \prod_{i=1}^n \left[\Gamma\left(y_i + \frac{\mu_i}{\sigma}\right) \Gamma\left(m_i + \frac{1 - \mu_i}{\sigma} - y_i\right) \right]}{\prod_{i=1}^n \left[\Gamma\left(m_i + \frac{1}{\sigma}\right) \Gamma\left(\frac{\mu_i}{\sigma}\right) \Gamma\left(\frac{1 - \mu_i}{\sigma}\right) \right]}$$

Inferência via MV

■ Logverossimilhança

$$\begin{aligned}l(\theta) &= n \ln \left[\Gamma \left(\frac{1}{\sigma} \right) \right] + \sum_{i=1}^n \left\{ \ln \left[\Gamma \left(y_i + \frac{\mu_i}{\sigma} \right) \right] + \ln \left[\Gamma \left(m_i + \frac{1 - \mu_i}{\sigma} - y_i \right) \right] \right. \\ &\quad \left. - \ln \left[\Gamma \left(m_i + \frac{1}{\sigma} \right) \right] - \ln \left[\Gamma \left(\frac{\mu_i}{\sigma} \right) \right] - \ln \left[\Gamma \left(\frac{1 - \mu_i}{\sigma} \right) \right] \right\} + \text{const} \\ &= n \ln \left[\left(\frac{1}{\sigma} \right) \right] + \sum_{i=1}^n \left\{ g_{1i}(\mu_i, \sigma) + g_{2i}(\mu_i, \sigma) \right. \\ &\quad \left. - g_{3i}(\mu_i, \sigma) - g_{4i}(\mu_i, \sigma) - g_{5i}(\mu_i, \sigma) \right\} + \text{const}\end{aligned}$$

Inferência via MV

$$\mathbf{S}(\boldsymbol{\beta}) = \begin{bmatrix} \frac{\partial l(\boldsymbol{\beta}, \sigma)}{\partial \beta_1} \\ \frac{\partial l(\boldsymbol{\beta}, \sigma)}{\partial \beta_2} \\ \vdots \\ \frac{\partial l(\boldsymbol{\beta}, \sigma)}{\partial \beta_p} \end{bmatrix}; S(\sigma) = \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma}$$

assim

$$\mathbf{S}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{S}(\boldsymbol{\beta}) \\ S(\sigma) \end{bmatrix}$$

$$\mathbf{H}(\boldsymbol{\beta}, \sigma) = \begin{bmatrix} \mathbf{H}_{11}(\boldsymbol{\beta}, \sigma) & \mathbf{H}_{12}(\boldsymbol{\beta}, \sigma) \\ \mathbf{H}_{21}(\boldsymbol{\beta}, \sigma) & \mathbf{H}_{22}(\boldsymbol{\beta}, \sigma) \end{bmatrix}$$

$$\mathbf{H}_{11}(\boldsymbol{\beta}, \sigma) = \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\beta}, \sigma)}{\partial \beta_1^2} & \frac{\partial^2 l(\boldsymbol{\beta}, \sigma)}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta}, \sigma)}{\partial \beta_1 \partial \beta_p} \\ \cdot & \frac{\partial^2 l(\boldsymbol{\beta}, \sigma)}{\partial \beta_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta}, \sigma)}{\partial \beta_2 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdot & \frac{\partial^2 l(\boldsymbol{\beta}, \sigma)}{\partial \beta_p^2} \end{bmatrix}$$

$$\mathbf{H}_{12}(\boldsymbol{\beta}, \sigma) = \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\beta}, \sigma)}{\partial \beta_1 \partial \sigma} \\ \frac{\partial^2 l(\boldsymbol{\beta}, \sigma)}{\partial \beta_2 \partial \sigma} \\ \vdots \\ \frac{\partial^2 l(\boldsymbol{\beta}, \sigma)}{\partial \beta_p \partial \sigma} \end{bmatrix}$$

$$H_{22}(\boldsymbol{\beta}, \sigma) = \frac{\partial^2 l(\boldsymbol{\beta}, \sigma)}{\partial \sigma^2}$$

Inferência via MV

■ Vetor escore β

$$S(\beta_j) = \sum_{i=1}^n \left\{ \frac{\partial g_{1i}(\mu_i, \sigma)}{\partial \beta_j} + \frac{\partial g_{2i}(\mu_i, \sigma)}{\partial \beta_j} - \frac{\partial g_{3i}(\mu_i, \sigma)}{\partial \beta_j} - \frac{\partial g_{4i}(\mu_i, \sigma)}{\partial \beta_j} - \frac{\partial g_{5i}(\mu_i, \sigma)}{\partial \beta_j} \right\}$$

■ Função escore σ

$$S(\sigma) = n \ln \frac{\partial \left[\Gamma \left(\frac{1}{\sigma} \right) \right]}{\partial \sigma} + \sum_{i=1}^n \left\{ \frac{\partial g_{1i}(\mu_i, \sigma)}{\partial \sigma} + \frac{\partial g_{2i}(\mu_i, \sigma)}{\partial \sigma} - \frac{\partial g_{3i}(\mu_i, \sigma)}{\partial \sigma} - \frac{\partial g_{4i}(\mu_i, \sigma)}{\partial \sigma} - \frac{\partial g_{5i}(\mu_i, \sigma)}{\partial \sigma} \right\}$$

Inferência via MV

■ Hessiana β

$$H(\beta_j, \beta_l) = \sum_{i=1}^n \left\{ \frac{\partial^2 g_{1i}(\mu_i, \sigma)}{\partial \beta_j \partial \beta_l} + \frac{\partial^2 g_{2i}(\mu_i, \sigma)}{\partial^2 \beta_j \partial \beta_l} - \frac{\partial^2 g_{3i}(\mu_i, \sigma)}{\partial \beta_j \partial \beta_l} - \frac{\partial^2 g_{4i}(\mu_i, \sigma)}{\partial \beta_j \partial \beta_l} - \frac{\partial^2 g_{5i}(\mu_i, \sigma)}{\partial \beta_j \partial \beta_l} \right\}$$

■ Hessiana σ

$$H(\sigma, \sigma) = n \ln \frac{\partial^2 \left[\Gamma \left(\frac{1}{\sigma} \right) \right]}{\partial \sigma^2} + \sum_{i=1}^n \left\{ \frac{\partial^2 g_{1i}(\mu_i, \sigma)}{\partial \sigma^2} + \frac{\partial^2 g_{2i}(\mu_i, \sigma)}{\partial \sigma^2} - \frac{\partial^2 g_{3i}(\mu_i, \sigma)}{\partial \sigma^2} - \frac{\partial^2 g_{4i}(\mu_i, \sigma)}{\partial \sigma^2} - \frac{\partial^2 g_{5i}(\mu_i, \sigma)}{\partial \sigma^2} \right\}$$

Inferência via MV

■ Hessiana β, σ

$$H(\beta_j, \sigma) = \sum_{i=1}^n \left\{ \frac{\partial^2 g_{1i}(\mu_i, \sigma)}{\partial \beta_j \partial \sigma} + \frac{\partial g_{2i}(\mu_i, \sigma)}{\partial^2 \beta_j \partial \sigma} - \frac{\partial^2 g_{3i}(\mu_i, \sigma)}{\partial \beta_j \partial \sigma} - \frac{\partial^2 g_{4i}(\mu_i, \sigma)}{\partial \beta_j \partial \sigma} - \frac{\partial^2 g_{5i}(\mu_i, \sigma)}{\partial \beta_j \partial \sigma} \right\}$$

Inferência via MV

- Em que

$$g_1(\mu_i, \sigma) = \ln \Gamma \left(y_i + \frac{\mu_i}{\sigma} \right);$$

$$g_2(\mu_i, \sigma) = \ln \Gamma \left(m_i + \frac{1 - \mu_i}{\sigma} - y_i \right);$$

$$g_3(\mu_i, \sigma) = \ln \Gamma \left(m_i + \frac{1}{\sigma} \right);$$

$$g_4(\mu_i, \sigma) = \ln \Gamma \left(\frac{\mu_i}{\sigma} \right);$$

$$g_5(\mu_i, \sigma) = \ln \Gamma \left(\frac{1 - \mu_i}{\sigma} \right)$$

■ Temos que

$$\frac{\partial g_1(\mu_i, \sigma)}{\partial \beta_j} = \frac{1}{\sigma} \Psi \left(y_i + \frac{\mu_i}{\sigma} \right) \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{1}{\sigma} \Psi \left(y_i + \frac{\mu_i}{\sigma} \right) \frac{\partial \mu_i}{\partial \eta_i} x_{ji}$$

$$\begin{aligned} \frac{\partial g_2(\mu_i, \sigma)}{\partial \beta_j} &= -\frac{1}{\sigma} \Psi \left(m_i + \frac{1 - \mu_i}{\sigma} - y_i \right) \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \\ &= -\frac{1}{\sigma} \Psi \left(m_i + \frac{1 - \mu_i}{\sigma} - y_i \right) \frac{\partial \mu_i}{\partial \eta_i} x_{ji} \end{aligned}$$

$$\frac{\partial g_3(\mu_i, \sigma)}{\partial \beta_j} = 0$$

$$\frac{\partial g_4(\mu_i, \sigma)}{\partial \beta_j} = \frac{1}{\sigma} \Psi \left(\frac{\mu_i}{\sigma} \right) \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{1}{\sigma} \Psi \left(\frac{\mu_i}{\sigma} \right) \frac{\partial \mu_i}{\partial \eta_i} x_{ji}$$

$$\frac{\partial g_5(\mu_i, \sigma)}{\partial \beta_j} = -\frac{1}{\sigma} \Psi \left(\frac{1 - \mu_i}{\sigma} \right) \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = -\frac{1}{\sigma} \Psi \left(\frac{1 - \mu_i}{\sigma} \right) \frac{\partial \mu_i}{\partial \eta_i} x_{ji}$$

■ Temos que

$$\frac{\partial g_1(\mu_i, \sigma)}{\partial \sigma} = -\frac{\mu_i}{\sigma^2} \Psi \left(y_i + \frac{\mu_i}{\sigma} \right)$$

$$\frac{\partial g_2(\mu_i, \sigma)}{\partial \sigma} = -\frac{1 - \mu_i}{\sigma^2} \Psi \left(m_i + \frac{1 - \mu_i}{\sigma} - y_i \right)$$

$$\frac{\partial g_3(\mu_i, \sigma)}{\partial \sigma} = -\frac{1}{\sigma^2} \Psi \left(m_i + \frac{1}{\sigma} \right)$$

$$\frac{\partial g_4(\mu_i, \sigma)}{\partial \sigma} = -\frac{\mu_i}{\sigma^2} \Psi \left(\frac{\mu_i}{\sigma} \right)$$

$$\frac{\partial g_5(\mu_i, \sigma)}{\partial \sigma} = -\frac{1 - \mu_i}{\sigma^2} \Psi \left(\frac{1 - \mu_i}{\sigma} \right)$$

$$\frac{\partial \ln \left[\Gamma \left(\frac{1}{\sigma} \right) \right]}{\partial \sigma} = -\frac{1}{\sigma^2} \Psi \left(\frac{1}{\sigma} \right)$$

■ Temos que

$$\frac{\partial^2 g_1(\mu_i, \sigma)}{\partial \beta_j \partial \beta_l} = x_{ji} x_{li} \frac{1}{\sigma} \left[\psi' \left(y_i + \frac{\mu_i}{\sigma} \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{\sigma} + \psi \left(y_i + \frac{\mu_i}{\sigma} \right) \frac{\partial^2 \mu_i}{\partial \eta_i^2} \right]$$

$$\begin{aligned} \frac{\partial^2 g_2(\mu_i, \sigma)}{\partial \beta_j \partial \beta_l} &= -x_{ji} x_{li} \frac{1}{\sigma} \left[-\psi' \left(m_i + \frac{1 - \mu_i}{\sigma} - y_i \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{\sigma} \right. \\ &\quad \left. + \psi \left(m_i + \frac{1 - \mu_i}{\sigma} - y_i \right) \frac{\partial^2 \mu_i}{\partial \eta_i^2} \right] \end{aligned}$$

$$\frac{\partial^2 g_3(\mu_i, \sigma)}{\partial \beta_j \partial \beta_l} = 0$$

$$\frac{\partial^2 g_4(\mu_i, \sigma)}{\partial \beta_j \partial \beta_l} = x_{ji} x_{li} \frac{1}{\sigma} \left[\psi' \left(\frac{\mu_i}{\sigma} \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{\sigma} + \psi \left(\frac{\mu_i}{\sigma} \right) \frac{\partial^2 \mu_i}{\partial \eta_i^2} \right]$$

$$\frac{\partial^2 g_5(\mu_i, \sigma)}{\partial \beta_j \partial \beta_l} = x_{ji} x_{li} \frac{1}{\sigma} \left[-\psi' \left(\frac{1 - \mu_i}{\sigma} \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{\sigma} + \psi \left(\frac{1 - \mu_i}{\sigma} \right) \frac{\partial^2 \mu_i}{\partial \eta_i^2} \right]$$

- Temos que

$$\frac{\partial^2 g_1(\mu_i, \sigma)}{\partial \sigma^2} = \mu_i \left[\frac{2}{\sigma^3} \Psi \left(y_i + \frac{\mu_i}{\sigma} \right) + \frac{\mu_i}{\sigma^4} \Psi' \left(y_i + \frac{\mu_i}{\sigma} \right) \right]$$

$$\begin{aligned} \frac{\partial^2 g_2(\mu_i, \sigma)}{\partial \sigma^2} &= (1 - \mu_i) \left[\frac{2}{\sigma^3} \Psi \left(m_i + \frac{1 - \mu_i}{\sigma} - y_i \right) \right. \\ &\quad \left. + \frac{1 - \mu_i}{\sigma^4} \Psi' \left(m_i + \frac{1 - \mu_i}{\sigma} - y_i \right) \right] \end{aligned}$$

$$\frac{\partial^2 g_3(\mu_i, \sigma)}{\partial \sigma^2} = \frac{2}{\sigma^3} \Psi \left(m_i + \frac{1}{\sigma} \right) + \frac{1}{\sigma^4} \Psi' \left(m_i + \frac{1}{\sigma} \right)$$

- Temos que

$$\begin{aligned} \frac{\partial^2 g_4(\mu_i, \sigma)}{\partial \sigma^2} &= \mu_i \left[\frac{2}{\sigma^3} \Psi \left(\frac{\mu_i}{\sigma} \right) + \frac{\mu_i}{\sigma^4} \Psi' \left(\frac{1 - \mu_i}{\sigma} \right) \right] \\ \frac{\partial^2 g_5(\mu_i, \sigma)}{\partial \sigma^2} &= (1 - \mu_i) \left[\frac{2}{\sigma^3} \Psi \left(\frac{1 - \mu_i}{\sigma} \right) + \frac{1 - \mu_i}{\sigma^4} \Psi' \left(\frac{1 - \mu_i}{\sigma} \right) \right] \\ \frac{\partial^2 \ln \left[\Gamma \left(\frac{1}{\sigma^2} \right) \right]}{\partial \sigma} &= \frac{2}{\sigma^3} \Psi \left(\frac{1}{\sigma} \right) + \frac{1}{\sigma^4} \Psi \left(\frac{1}{\sigma} \right) \end{aligned}$$

■ Temos que

$$\frac{\partial g_1(\mu_i, \sigma)}{\partial \beta_j \partial \sigma} = -\frac{1}{\sigma^2} \frac{\partial \mu_i}{\partial \eta_i} x_{ji} \left[\Psi \left(y_i + \frac{\mu_i}{\sigma} \right) + \frac{\mu_i}{\sigma} \Psi' \left(y_i + \frac{\mu_i}{\sigma} \right) \right]$$

$$\begin{aligned} \frac{\partial g_2(\mu_i, \sigma)}{\partial \beta_j \partial \sigma} &= \frac{1}{\sigma^2} \frac{\partial \mu_i}{\partial \eta_i} x_{ji} \left[\Psi \left(m_i + \frac{1 - \mu_i}{\sigma} - y_i \right) \right. \\ &\quad \left. + (1 - \mu_i) \Psi' \left(m_i + \frac{1 - \mu_i}{\sigma} - y_i \right) \right] \end{aligned}$$

$$\frac{\partial g_3(\mu_i, \sigma)}{\partial \beta_j \partial \sigma} = 0$$

$$\frac{\partial g_4(\mu_i, \sigma)}{\partial \beta_j \partial \sigma} = -\frac{\partial \mu_i}{\partial \eta_i} x_{ji} \frac{1}{\sigma^2} \left[\Psi \left(\frac{\mu_i}{\sigma} \right) + \mu_i \Psi' \left(\frac{\mu_i}{\sigma} \right) \frac{1}{\sigma} \right]$$

$$\frac{\partial g_5(\mu_i, \sigma)}{\partial \beta_j \partial \sigma} = \frac{1}{\sigma^2} \frac{\partial \mu_i}{\partial \eta_i} x_{ji} \left[\Psi \left(\frac{1 - \mu_i}{\sigma} \right) + \Psi' \left(\frac{1 - \mu_i}{\sigma} \right) \frac{1 - \mu_i}{\sigma} \right]$$

Inferência via MV

- Podemos notar que o sistema $\begin{cases} \mathbf{S}(\tilde{\beta}) = \mathbf{0}_{(p \times 1)} \\ S(\tilde{\sigma}) = 0 \end{cases}$ não apresenta solução analítica. Assim, algum método numérico deverá ser empregado e, nesse caso, utilizaremos o algoritmo de Newton-Raphson (o qual usa a matriz Hessiana ao invés da informação de Fisher).
- A obtenção da Informação de Fisher (esperada) é complicada.
- Exercício: escrever de forma matricial as quantidades $\mathbf{S}(\beta, \sigma)$ e $H(\beta, \sigma)$ e/ou escrever o código em R para implementar o algoritmo de Newton-Raphson.

- O algoritmo de Newton-Raphson (ANR) é definido como:

Sejam $\beta^{(0)}$ e $\sigma^{(0)}$ estimativas iniciais de β e σ (chutes iniciais), respectivamente, então faça

$$\begin{bmatrix} \beta^{(t+1)} \\ \sigma^{(t+1)} \end{bmatrix} = \begin{bmatrix} \beta^{(t)} \\ \sigma^{(t)} \end{bmatrix} + \begin{bmatrix} H_{11}(\beta^{(t)}, \sigma^{(t)}) & H_{12}(\beta^{(t)}, \sigma^{(t)}) \\ H_{21}(\beta^{(t)}, \sigma^{(t)}) & H_{22}(\beta^{(t)}, \sigma^{(t)}) \end{bmatrix}^{-1} \begin{bmatrix} S(\beta^{(t)}) \\ S(\sigma^{(t)}) \end{bmatrix}$$

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

- Uma vez que a obtenção da informação de Fisher (esperada) é complicada, podemos utilizar a informação observada $I_O(\beta, \sigma) = -\mathbf{H}(\beta, \sigma)$ para construir os resultados assintóticos.
- Neste caso, para n suficientemente grande, temos que $\hat{\beta} \approx N_p(\beta, \text{Cov}(\beta))$ e $\hat{\sigma} \approx N(\sigma, \mathcal{V}(\sigma))$, em que:

$$\begin{bmatrix} \text{Cov}(\hat{\beta}) & \text{Cov}(\hat{\beta}, \hat{\sigma}) \\ \text{Cov}(\hat{\sigma}, \hat{\beta}) & \mathcal{V}(\hat{\sigma}) \end{bmatrix} = - \begin{bmatrix} \mathbf{H}_{11}(\beta, \sigma) & \mathbf{H}_{12}(\beta, \sigma) \\ \mathbf{H}_{21}(\beta, \sigma) & H_{22}(\beta, \sigma) \end{bmatrix}^{-1}$$

- Os resultados apresentados anteriormente para o modelo binomial negativo (intervalos de confiança e testes de hipótese) continuam válidos com as devidas modificações.

Comparação de modelos e validação

- As estatísticas AIC e BIC podem ser calculadas de forma semelhante ao que fora feito para os MLG.
- Pesquisar sobre o comportamento assintótico da função desvio.
- Verificação da qualidade de ajuste do modelo: resíduos quantil aleatorizado.
- Veja mais em <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9876.2005.00510.x/full> e https://www.researchgate.net/publication/26538604-Generalised_additive_models_for_Location_Scale_and_Shape_GAMLSS_in_R.

Resíduo quantil aleatorizado (RQA)

- Sejam X e Y variáveis tais que F e G são, respectivamente, suas f.d.a's. Temos então que $U = F(X) \sim U(0, 1)$ (transformação integral de probabilidade) e $W = G^{-1}(U) \sim G$, ou seja W e Y possuem a mesma distribuição.
- Se $G \equiv \Phi(\cdot)$ (f.d.a da normal padrão) então $W \sim N(0, 1)$.
- Logo, se x for um valor simulado de X então $u = F(x)$ corresponderá à um valor simulado da distribuição $U(0,1)$ e, conseqüentemente $w = G(u)$ à um valor da $N(0,1)$.

Resíduo quantil aleatorizado (RQA)

- Esse resultado é muito útil para construir resíduos com distribuição (ainda que aproximada) $N(0,1)$, sob o ajuste adequado do modelo, quando a variável resposta é contínua.
- Contudo, em nosso caso, a variável resposta é discreta.
- Ocorre que se F for contínua para todo $q \in (0, 1)$, $\exists x, x = F^{-1}(q)$ o que não, necessariamente, ocorre quando F é discreta.
- No caso discreto (sendo $y_1 < y_2 < y_3, \dots$ o suporte de Y), temos que $P(Y = y_i) = F(y_i) - F(y_{i-1}) = P(F(y_{i-1}) \leq U \leq F(y_i))$, em que $U \sim U(0, 1)$.

Resíduo quantil aleatorizado (RQA)

- Ou seja, à cada valor de Y , digamos y_i , podemos associar um número uniforme no intervalo $[F(y_{i-1}), F(y_i)]$.
- Usualmente, para simular de uma vad , simula-se u , $U \sim U(0, 1)$ e $y = y_i$, se $u \in [F(y_{i-1}), F(y_i)]$.
- Por outro lado, se y_1, \dots, y_n forem valores (simulados) de Y , então $u_i \stackrel{\text{ind.}}{\sim} U([F(y_{i-1}), F(y_i)])$ e, conseqüentemente $u_i \stackrel{\text{ind.}}{\approx} U(0, 1)$.
Logo $z_i = \Phi^{-1}(u_i) \stackrel{i.i.d.}{\sim} N(0, 1)$.
- O problema é que para um conjunto de valores y_1, \dots, y_n temos um número virtualmente infinitos de conjunto de valores u_1, \dots, u_n .

RQA

- Cálculo do RQA para o modelo beta-binomial.
- Considera-se as observações (ordenadas) y_1, \dots, y_n
- Com as estimativas de β e σ simula-se (m vezes) valores de $u_i \sim U([F(y_{(i-1)}), F(y_i)])$, $F(\cdot)$ é a fda da distribuição beta-binomial, gerando-se $u_{i1} < \dots < u_{im}$, $i = 1, 2, \dots, n$. Se $y_1 = 0$ então $F(y_0) = 0$.
- Usualmente ordena-se os valores para cada i , ou seja, $u_{i(1)}, \dots, u_{i(m)}$ e considera-se a mediana (digamos $u_{i(0.5)}$) dos valores simulados.
- Analisa-se a distribuição de $\Phi^{-1}(u_{1(0.5)}), \dots, \Phi^{-1}(u_{n(0.5)})$. Sob o bom ajuste do modelo espera-se que esses resíduos tenham, aproximadamente, distribuição $N(0,1)$.

Ilustração resíduo quantil

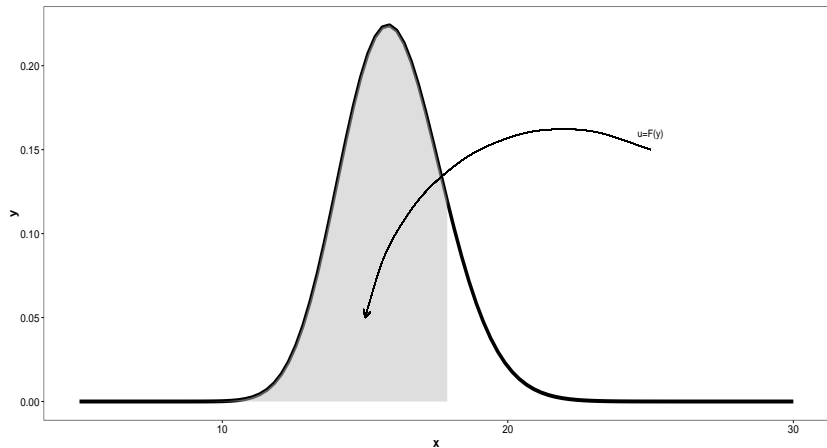


Ilustração resíduo quantil

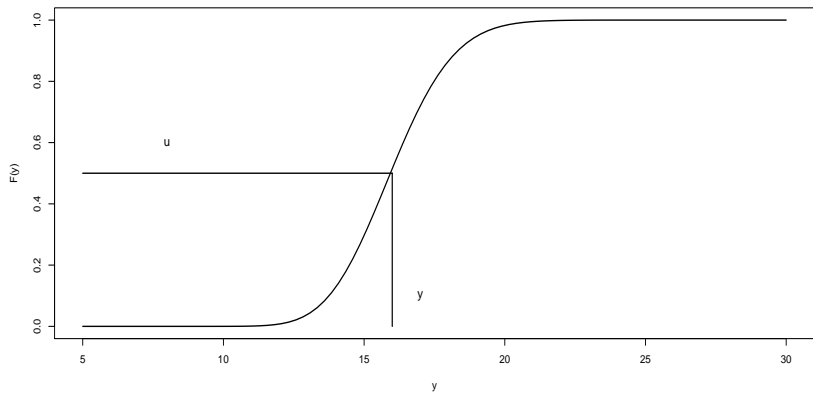


Ilustração resíduo quantil

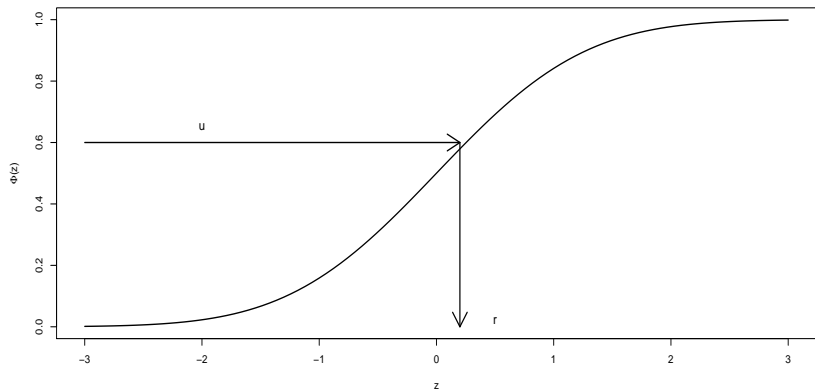


Ilustração resíduo quantil aleatorizado

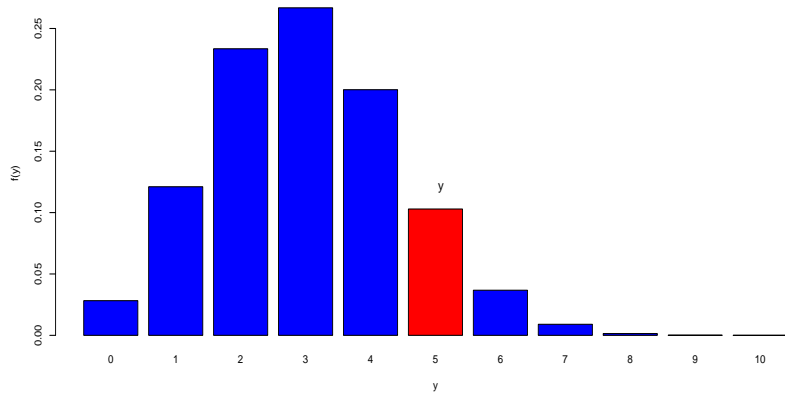


Ilustração resíduo quantil aleatorizado

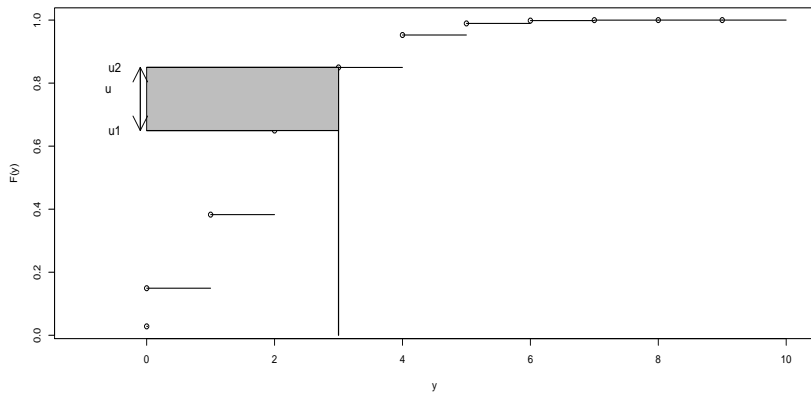
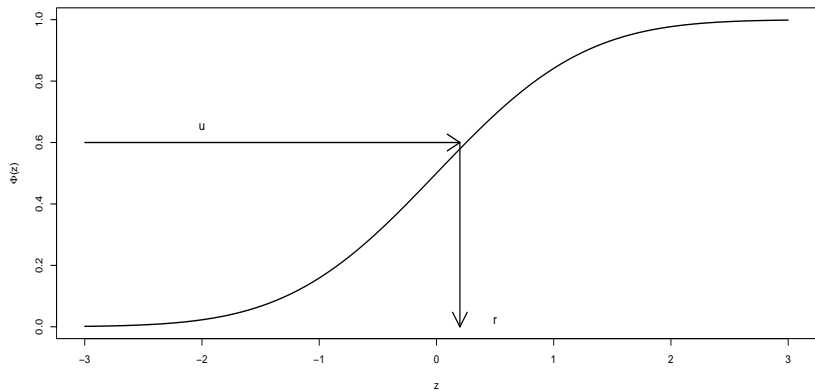


Ilustração resíduo quantil aleatorizado



Exemplo 10: germinação de sementes de Orobanche

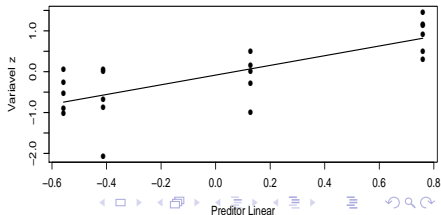
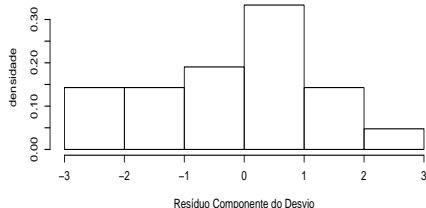
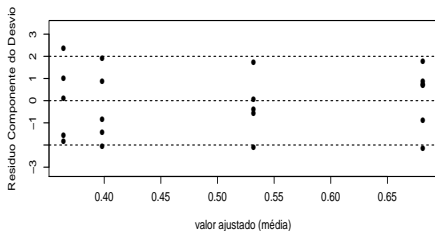
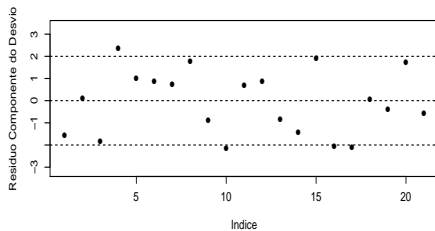
- Orobanche: tipo de organismo vegetal (planta).
- Os dados foram obtidos a partir de um estudo sobre germinação de duas espécies de sementes de Orobanche (*O. aegyptiaca* 75 e *O. aegyptiaca* 73), veja (Hinde e Demétrio (1998)). Correspondem à quantidade de sementes analisadas (m) e germinadas (y) e tem-se o interesse na resposta y/m (conjunto de sementes).
- As sementes foram cultivadas em diluições de $1/125$ a partir de dois tipos de extrato de raiz (pepino ou feijão), em um esquema fatorial 2×2 com replicações.
- Objetivo: avaliar o comportamento de cada espécie, sob cada um dos tipos de extrato, em termos da capacidade de germinação.

Modelo beta-binomial

$$Y_{ijk} \stackrel{ind.}{\sim} \text{bb}(m_{ijk}, \mu_{ij}, \sigma)$$
$$\ln\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \alpha + \beta_i + \gamma_j + (\beta\gamma)_{ij}, i = 1, 2, j = 1, 2, k = 1, 2, \dots, n_{ij}$$
$$\beta_1 = \gamma_1 = (\beta\gamma)_{1j} = (\beta\gamma)_{i1} = 0, \forall i, j.$$

- n_{ij} : número total de sementes pertencentes a i -ésima espécie (1: *O.aegyptiaca*₇₃, 2: *O.aegyptiaca*₇₅) e tratados com o j -ésimo tipo de extrato (1: feijão, 2: pepino), $n_{11} = 5, n_{12} = 5, n_{21} = 5, n_{22} = 6$.
- Y_{ijk} : é o total de sementes germinadas na k -ésimo réplica, pertencentes à i -ésima espécie, e tratadas com o j -ésimo tipo de extrato vegetal.

Gráficos de diagnóstico: modelo logístico binomial



Gráficos de diagnóstico: modelo logístico beta-binomial

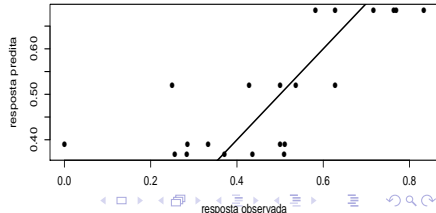
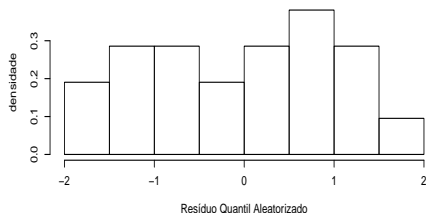
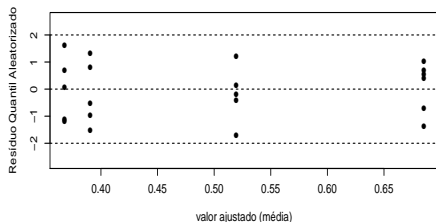
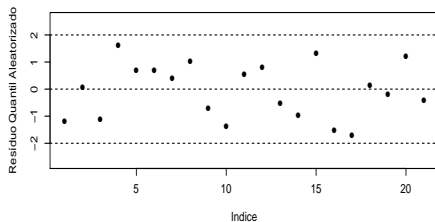
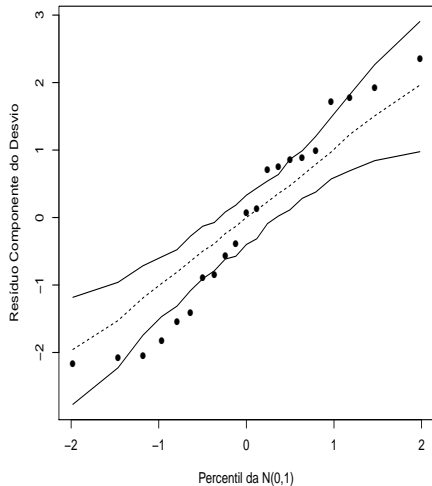
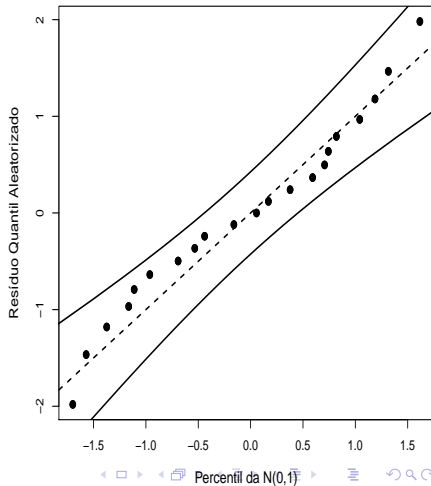


Gráfico de envelope: modelos binomial e beta-binomial

Modelo logito binomial



Modelo logito beta-binomial



Comparação com outros modelos (funções de ligação)

Função de ligação	AIC	BIC	DABM	Desvio	p-valor (desvio)
binomial	117,87	122,05	0,11	33,28	0,0104
beta-binomial	117,53	122,75	0,10	-	-

Estimativas dos parâmetros: modelos bin. e bb

Modelo	Parâm.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
binomial	α	-0,41	0,18	[-0,77 ; -0,05]	-2,24	0,0252
	β_2	-0,15	0,22	[-0,58 ; 0,29]	-0,65	0,5132
	γ_2	0,54	0,25	[0,05 ; 1,03]	2,16	0,0306
	$(\beta\gamma)_{22}$	0,78	0,31	[0,18 ; 1,38]	2,54	0,0111
beta-binomial	α	-0,44	0,22	[-0,87 -0,02]	-2,05	0,0564
	β_2	-0,10	0,27	[-0,63 ; 0,44]	-0,36	0,7243
	γ_2	0,52	0,30	[-0,06 ; 1,10]	1,76	0,0961
	$(\beta\gamma)_{22}$	0,80	0,38	[0,06 ; 1,54]	2,11	0,0497
	σ	0,013	0,189	-	-	-

Modelo 2

$$Y_{ijk} \stackrel{ind.}{\sim} \text{bb}(m_{ijk}, \mu_{ij}, \sigma)$$
$$\ln\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \alpha + \gamma_j + \delta_{ij}, i = 1, 2, j = 1, 2, k = 1, 2, \dots, n_{ij}$$
$$\gamma_1 = \delta_{1j} = \delta_{i1} = 0, \forall i, j.$$

- n_{ij} : número total de sementes pertencentes a i -ésima espécie (1: *O.aegyptiaca*₇₃, 2: *O.aegyptiaca*₇₅) e tratados com o j -ésimo tipo de extrato (1: feijão, 2: pepino), $n_{11} = 5, n_{12} = 5, n_{21} = 5, n_{22} = 6$.
- Y_{ijk} : é o total de sementes germinadas na k -ésimo réplica, pertencentes à i -ésima espécie, e tratadas com o j -ésimo tipo de extrato vegetal.

Modelo 2

- Médias induzidas pelo modelo 2

$$\mu_{11} = \frac{e^{\alpha}}{1 + e^{\alpha}} \quad (73, \text{ feijão})$$

$$\mu_{12} = \frac{e^{\alpha + \gamma_2}}{1 + e^{\alpha + \gamma_2}} \quad (73, \text{ pepino})$$

$$\mu_{21} = \frac{e^{\alpha}}{1 + e^{\alpha}} \quad (75, \text{ feijão})$$

$$\mu_{22} = \frac{e^{\alpha + \gamma_2 + \delta_{22}}}{1 + e^{\alpha + \gamma_2 + \delta_{22}}} \quad (75, \text{ pepino})$$

Gráficos de diagnóstico: modelo 2

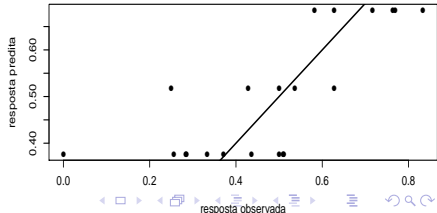
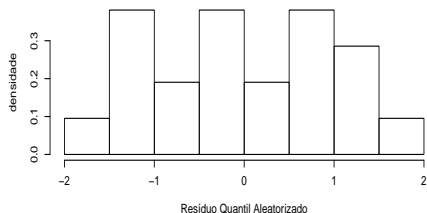
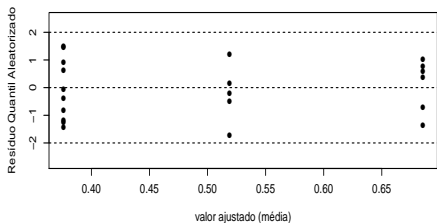
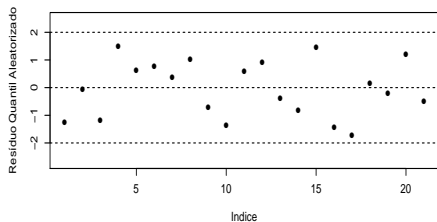
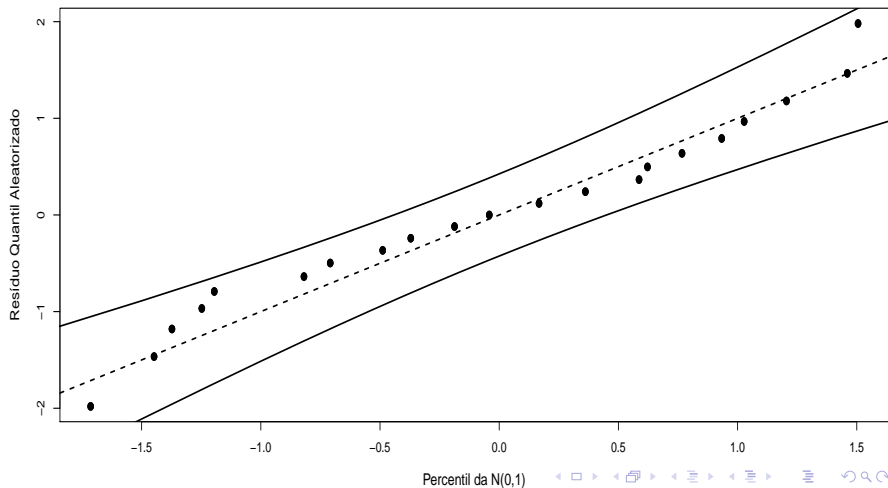


Gráfico de envelope: modelos 2



Estimativas dos parâmetros: modelo 2

Parâm.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
α	-0,51	0,13	[-0,77 ; -0,25]	-3,84	0,0012
γ_2	0,58	0,24	[0,11 ; 1,06]	2,40	0,0274
δ_{22}	0,70	0,26	[0,18 ; 1,22]	2,66	0,0161
σ	0,013	0,378	-	-	-

Médias observadas e previstas

