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 (de referência).
- No caso do modelos de regressão para dados binários, em que assume-se que $Y_i \overset{ind.}{\sim}$ binomial$(m_i, \mu_i), i = 1, 2, ..., n$, tem-se que $\mathcal{V}(Y_i) = m_i\mu_i(1 - \mu_i)$.
- No caso do Exemplo 8 (dos besouros), $n = 8$ e $m_i \in \{59, 60, 62, 56, 63, 59, 62, 60\}$.

Cont.

- 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 $D(\boldsymbol{y}; \widetilde{\boldsymbol{\mu}})$ é maior que o número de graus de liberdade (n-p), pode indicar presença 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 dentre 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).

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 $\mathcal{V}(Y_{ij}) = \mu_i(1-\mu_i)$, $i = 1, 2, ..., n$; $j = 1, 2, ..., m_i$. O número total de sucessos no i-ésimo grupo será definido por

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

- 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

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

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 $-\dfrac{1}{m_i - 1} < \delta < 1$.
- Logo, δ poderá assumir valores negativos apenas se houver $m_i's$ pequenos.
- Podemos notar que

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

Caso II

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

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

Cont.

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

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

(continua no próximo slide)

Cont.

- (Cont.)

$$
\begin{aligned}
&= m_i\left(\mu_i(1-\delta)(1-\mu_i)\right) + m_i^2\delta\mu_i(1-\mu_i) \\
&= m_i\mu_i(1-\mu_i)\left[1+\delta(m_i-1)\right],
\end{aligned}
$$

que coincidem com os resultado do Caso I, considerando $\delta > 0$.

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

$$
\begin{aligned}
f_y(y) &= \frac{\Gamma(m+1)}{\Gamma(y+1)\Gamma(m-y+1)} \frac{\Gamma(y+a)\Gamma(m-y+b)}{\Gamma(m+a+b)} \\
&\times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mathbb{1}_{\{0,1,2,3,\ldots,m\}}(y).
\end{aligned}
$$

Modelagem

- Temos que $\mathcal{E}(Y) = m\dfrac{a}{(a+b)}$ e $\mathcal{V}(Y) = \dfrac{mab(a+b+m)}{(a+b)^2(a+b+1)}$. (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, $a = \dfrac{\mu}{\sigma}$ e $b = \dfrac{1-\mu}{\sigma}$, então:

$$
\begin{aligned}
f_Y(y) &= \frac{\Gamma(m+1)}{\Gamma(y+1)\Gamma(m-y+1)} \frac{\Gamma\left(y+\frac{\mu}{\sigma}\right)\Gamma\left(\frac{1}{\sigma}\right)\Gamma\left(m+\frac{1-\mu}{\sigma}-y\right)}{\Gamma\left(m+\frac{1}{\sigma}\right)\Gamma\left(\frac{\mu}{\sigma}\right)\Gamma\left(\frac{1-\mu}{\sigma}\right)} \\
&\times \mathbb{1}_{\{0,1,2,\ldots,m\}}(y)
\end{aligned}
$$

Modelagem

- Nesse caso,

$$
\begin{aligned}
\mathcal{E}(Y) &= m\mu, \\
\mathcal{V}(Y) &= \frac{m\mu(1-\mu)(1+m\sigma)}{1+\sigma} = m\mu(1-\mu)\left[1+\frac{(m-1)\sigma}{1+\sigma}\right],
\end{aligned}
$$

em que σ é um parâmetro de dispersão (provar).

- O interesse agora reside em modelar μ_i, mantendo-se σ fixo ao longo das observações.

- Em termos práticos, à semelhança do modelo binomial, trabalha-se com a verossimilhança associada à $Y_i^* = \dfrac{Y_i}{m_i}$.

Modelo de regressão beta binomial

- Sejam $Y_i, i = 1, 2, ..., n$, $Y_i \overset{ind.}{\sim} \text{bb}(m_i, \mu_i, \sigma)$.
- Temos que $\mathcal{E}(Y_i^*) = \mu_i$ e $\mathcal{V}(Y_i^*) = \mu_i(1-\mu_i)\left[1 + \dfrac{(m_i - 1)\sigma}{1+\sigma}\right]$.
- $g(\mu_i) = \eta_i \rightarrow \mu = g^{-1}(\eta_i), \eta_i = \boldsymbol{X}_i'\boldsymbol{\beta}$ em que $g(.)$, em geral corresponde à $F^{-1}(.)$, 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 Aditive Models for Location Scale and Shape"). Existe um pacote no R homônimo (gamlss).

Inferência via MV

- Verossimilhança (em que $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma)'$):

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \frac{\prod_{i=1}^{n} \Gamma(m_i + 1)}{\prod_{i=1}^{n} [\Gamma(y_i + 1)\Gamma(m_i - y + 1)]} \\
&\times \frac{\left[\Gamma\left(\frac{1}{\sigma}\right)\right]^n \prod_{i=1}^{n} \left[\Gamma\left(y_i + \frac{\mu_i}{\sigma}\right)\Gamma\left(m_i + \frac{1-\mu_i}{\sigma} - y_i\right)\right]}{\prod_{i=1}^{n} \left[\Gamma\left(m_i + \frac{1}{\sigma}\right)\Gamma\left(\frac{\mu_i}{\sigma}\right)\Gamma\left(\frac{1-\mu_i}{\sigma}\right)\right]}.
\end{aligned}
$$

Inferência via MV

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

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

Inferência via MV

- Logverossimilhança

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

Inferência via MV

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

assim

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

Cont.

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

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

Cont.

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

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

Inferência via MV

- Vetor escore ($\boldsymbol{\beta}$)

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

- Função escore (σ)

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

Inferência via MV

- Hessiana ($\boldsymbol{\beta}$)

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

- Hessiana (σ)

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

Inferência via MV

- Hessiana $(\boldsymbol{\beta}', \sigma)'$

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

Inferência via MV

- Em que

$$
\begin{aligned}
g_1(\mu_i, \sigma) &= \ln \Gamma\left(y_i + \frac{\mu_i}{\sigma}\right); \\
g_2(\mu_i, \sigma) &= \ln \Gamma\left(m_i + \frac{1-\mu_i}{\sigma} - y_i\right); \\
g_3(\mu_i, \sigma) &= \ln \Gamma\left(m_i + \frac{1}{\sigma}\right); \\
g_4(\mu_i, \sigma) &= \ln \Gamma\left(\frac{\mu_i}{\sigma}\right); \\
g_5(\mu_i, \sigma) &= \ln \Gamma\left(\frac{1-\mu_i}{\sigma}\right).
\end{aligned}
$$

Cont.

- Temos que

$$
\begin{aligned}
\frac{\partial g_1(\mu_i,\sigma)}{\partial \beta_j} &= \frac{1}{\sigma}\Psi\left(y_i + \frac{\mu_i}{\sigma}\right)\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = \frac{1}{\sigma}\Psi\left(y_i + \frac{\mu_i}{\sigma}\right)\frac{\partial \mu_i}{\partial \eta_i}x_{ji}, \\
\frac{\partial g_2(\mu_i,\sigma)}{\partial \beta_j} &= -\frac{1}{\sigma}\Psi\left(m_i + \frac{1-\mu_i}{\sigma} - y_i\right)\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}, \\
&= -\frac{1}{\sigma}\Psi\left(m_i + \frac{1-\mu_i}{\sigma} - y_i\right)\frac{\partial \mu_i}{\partial \eta_i}x_{ji},
\end{aligned}
$$

Cont.

- Temos que

$$
\begin{aligned}
\frac{\partial g_3(\mu_i,\sigma)}{\partial \beta_j} &= 0,\\
\frac{\partial g_4(\mu_i,\sigma)}{\partial \beta_j} &= \frac{1}{\sigma}\Psi\left(\frac{\mu_i}{\sigma}\right)\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = \frac{1}{\sigma}\Psi\left(\frac{\mu_i}{\sigma}\right)\frac{\partial \mu_i}{\partial \eta_i}x_{ji},\\
\frac{\partial g_5(\mu_i,\sigma)}{\partial \beta_j} &= -\frac{1}{\sigma}\Psi\left(\frac{1-\mu_i}{\sigma}\right)\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = -\frac{1}{\sigma}\Psi\left(\frac{1-\mu_i}{\sigma}\right)\frac{\partial \mu_i}{\partial \eta_i}x_{ji}.
\end{aligned}
$$

Cont.

- Temos que

$$
\begin{aligned}
\frac{\partial g_1(\mu_i,\sigma)}{\partial \sigma} &= -\frac{\mu_i}{\sigma^2}\Psi\left(y_i + \frac{\mu_i}{\sigma}\right),\\
\frac{\partial g_2(\mu_i,\sigma)}{\partial \sigma} &= -\frac{1-\mu_i}{\sigma^2}\Psi\left(m_i + \frac{1-\mu_i}{\sigma} - y_i\right),\\
\frac{\partial g_3(\mu_i,\sigma)}{\partial \sigma} &= -\frac{1}{\sigma^2}\Psi\left(m_i + \frac{1}{\sigma}\right),\\
\frac{\partial g_4(\mu_i,\sigma)}{\partial \sigma} &= -\frac{\mu_i}{\sigma^2}\Psi\left(\frac{\mu_i}{\sigma}\right),\\
\frac{\partial g_5(\mu_i,\sigma)}{\partial \sigma} &= -\frac{1-\mu_i}{\sigma^2}\Psi\left(\frac{1-\mu_i}{\sigma}\right);\ \frac{\partial \ln\left[\Gamma\left(\frac{1}{\sigma}\right)\right]}{\partial \sigma} = -\frac{1}{\sigma^2}\Psi\left(\frac{1}{\sigma}\right).
\end{aligned}
$$

Cont.

- Temos que

$$
\begin{aligned}
\frac{\partial^2 g_1(\mu_i,\sigma)}{\partial\beta_j\partial\beta_l} &= x_{ji}x_{li}\frac{1}{\sigma}\left[\Psi'\left(y_i+\frac{\mu_i}{\sigma}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\frac{1}{\sigma}+\Psi\left(y_i+\frac{\mu_i}{\sigma}\right)\frac{\partial^2\mu_i}{\partial\eta_i^2}\right],\\
\frac{\partial^2 g_2(\mu_i,\sigma)}{\partial\beta_j\partial\beta_l} &= -x_{ji}x_{li}\frac{1}{\sigma}\left[-\Psi'\left(m_i+\frac{1-\mu_i}{\sigma}-y_i\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\frac{1}{\sigma}\right.\\
&+ \left.\Psi\left(m_i+\frac{1-\mu_i}{\sigma}-y_i\right)\frac{\partial^2\mu_i}{\partial\eta_i^2}\right],\\
\frac{\partial^2 g_3(\mu_i,\sigma)}{\partial\beta_j\partial\beta_l} &= 0,
\end{aligned}
$$

Cont.

- Temos que

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

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

Cont.

- Temos que

$$
\begin{aligned}
\frac{\partial^2 g_1(\mu_i,\sigma)}{\partial\sigma^2} &= \mu_i\left[\frac{2}{\sigma^3}\Psi\left(y_i+\frac{\mu_i}{\sigma}\right)+\frac{\mu_i}{\sigma^4}\Psi'\left(y_i+\frac{\mu_i}{\sigma}\right)\right],\\
\frac{\partial^2 g_2(\mu_i,\sigma)}{\partial\sigma^2} &= (1-\mu_i)\left[\frac{2}{\sigma^3}\Psi\left(m_i+\frac{1-\mu_i}{\sigma}-y_i\right),\right.\\
&+ \left.\frac{1-\mu_i}{\sigma^4}\Psi'\left(m_i+\frac{1-\mu_i}{\sigma}-y_i\right)\right]\\
\frac{\partial^2 g_3(\mu_i,\sigma)}{\partial\sigma^2} &= \frac{2}{\sigma^3}\Psi\left(m_i+\frac{1}{\sigma}\right)+\frac{1}{\sigma^4}\Psi'\left(m_i+\frac{1}{\sigma}\right),
\end{aligned}
$$

Cont.

- Temos que

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

Cont.

- Temos que

$$
\begin{aligned}
\frac{\partial g_1(\mu_i,\sigma)}{\partial\beta_j\partial\sigma} &= -\frac{1}{\sigma^2}\frac{\partial\mu_i}{\partial\eta_i}x_{ji}\left[\Psi\left(y_i+\frac{\mu_i}{\sigma}\right)+\frac{\mu_i}{\sigma}\Psi'\left(y_i+\frac{\mu_i}{\sigma}\right)\right],\\
\frac{\partial g_2(\mu_i,\sigma)}{\partial\beta_j\partial\sigma} &= \frac{1}{\sigma^2}\frac{\partial\mu_i}{\partial\eta_i}x_{ji}\left[\Psi\left(m_i+\frac{1-\mu_i}{\sigma}-y_i\right),\right.\\
&+ \left.(1-\mu_i)\Psi'\left(m_i+\frac{1-\mu_i}{\sigma}-y_i\right)\right]\\
\frac{\partial g_3(\mu_i,\sigma)}{\partial\beta_j\partial\sigma} &= 0,
\end{aligned}
$$

Cont.

- (cont.) Temos que

$$
\begin{aligned}
\frac{\partial g_4(\mu_i,\sigma)}{\partial \beta_j \partial \sigma} &= -\frac{\partial \mu_i}{\partial \eta_i} x_{ji} \frac{1}{\sigma^2}\left[\Psi\left(\frac{\mu_i}{\sigma}\right) + \mu_i \Psi'\left(\frac{\mu_i}{\sigma}\right)\frac{1}{\sigma}\right], \\
\frac{\partial g_5(\mu_i,\sigma)}{\partial \beta_j \partial \sigma} &= \frac{1}{\sigma^2}\frac{\partial \mu_i}{\partial \eta_i} x_{ji}\left[\Psi\left(\frac{1-\mu_i}{\sigma}\right) + \Psi'\left(\frac{1-\mu_i}{\sigma}\right)\frac{1-\mu_i}{\sigma}\right].
\end{aligned}
$$

Inferência via MV

- Podemos notar que o sistema $\begin{cases} \boldsymbol{S}(\widetilde{\boldsymbol{\beta}}) = \mathbf{0}_{(p\times 1)} \\ S(\widetilde{\sigma}) = 0 \end{cases}$ 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 $\boldsymbol{S}(\boldsymbol{\beta}, \sigma)$ e $H(\boldsymbol{\beta}, \sigma)$ e/ou escrever o código em R para implementar o algoritmo de Newton-Raphson.

Algoritmo de Newton-Raphson

- O algoritmo de Newton-Raphson (ANR) é definido como: Sejam $\boldsymbol{\beta}^{(0)}$ e $\sigma^{(0)}$ estimativas iniciais de $\boldsymbol{\beta}$ e σ (chutes iniciais), respectivamente, então faça

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

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

Cont.

- Uma vez que a obtenção da informação de Fisher (esperada) é complicada, podemos utilizar a informação observada $\boldsymbol{I}_O(\boldsymbol{\beta}, \sigma) = -\boldsymbol{H}(\boldsymbol{\beta}, \sigma)$, pois, sob certa condições de regularidade, temos que

$$|\boldsymbol{I}_O(\boldsymbol{\beta}, \sigma) - \boldsymbol{I}(\boldsymbol{\beta}, \sigma)| \xrightarrow[n \to \infty]{P} \mathbf{0}_{(p+1)\times(p+1))}$$

para construir os resultados inferenciais assintóticos.

Cont.

- Neste caso (sob as condições de regularidade supramencionadas), para n suficientemente grande, temos que

$$\widehat{\boldsymbol{\beta}} \approx N_p(\boldsymbol{\beta}, Cov(\boldsymbol{\beta})), \widehat{\sigma} \approx N(\sigma, \mathcal{V}(\sigma)),$$

em que:

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

- 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

- Os critérios e informação podem ser calculados calculadas de forma semelhante ao que fora feito para os MLG (aqui).
- Pesquisar sobre o comportamento assintótico da função desvio.
- Verificação da qualidade de ajuste do modelo: resíduo quantílico aleatorizado.
- Mais detalhes sobre aspectos inferenciais, veja aqui e aqui.
- Veja também aqui e aqui.

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).

Cont.

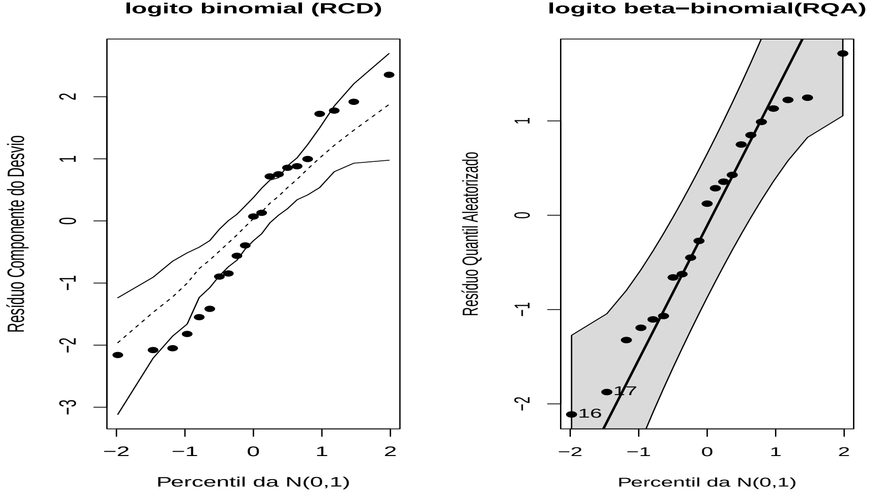
- As sementes foram cultivadas em diluições de $1/125$ a partir de dois tipos de extrato de raiz (pepino ou feijão), em um esquema fatorial 2×2 com replicações.
- Objetivo: avaliar o comportamento de cada espécie, sob cada um dos tipos de extrato, em termos da capacidade de germinação.

Modelo beta-binomial

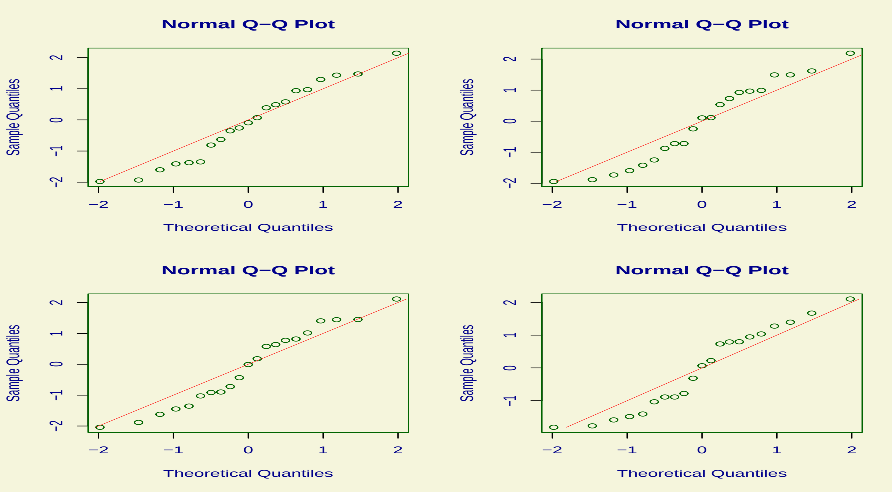
$$
\begin{aligned}
Y_{ijk} &\overset{ind.}{\sim} \mathrm{bb}(m_{ijk}, \mu_{ij}, \sigma) \\
\ln\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) &= \alpha + \beta_i + \gamma_j + (\beta\gamma)_{ij}, i = 1, 2, j = 1, 2, k = 1, 2, ..., n_{ij} \\
\beta_1 = \gamma_1 &= (\beta\gamma)_{1j} = (\beta\gamma)_{i1} = 0, \forall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

- n_{ij} : número total de sementes pertecentes a i-ésima espécie (1: O.aegyptiaca73, 2: O.aegyptiaca75) 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.

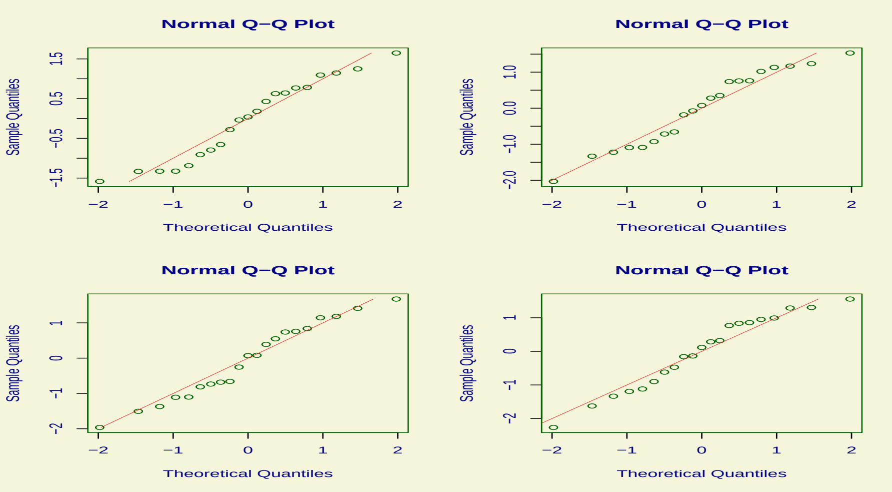
Gráficos de envelope: modelos binomial e beta-binomial



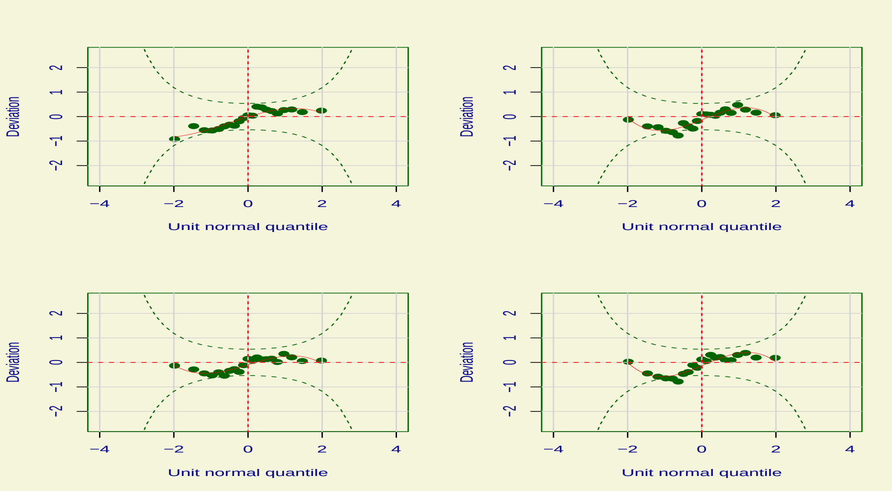
QQ plots RQA: modelos binomial



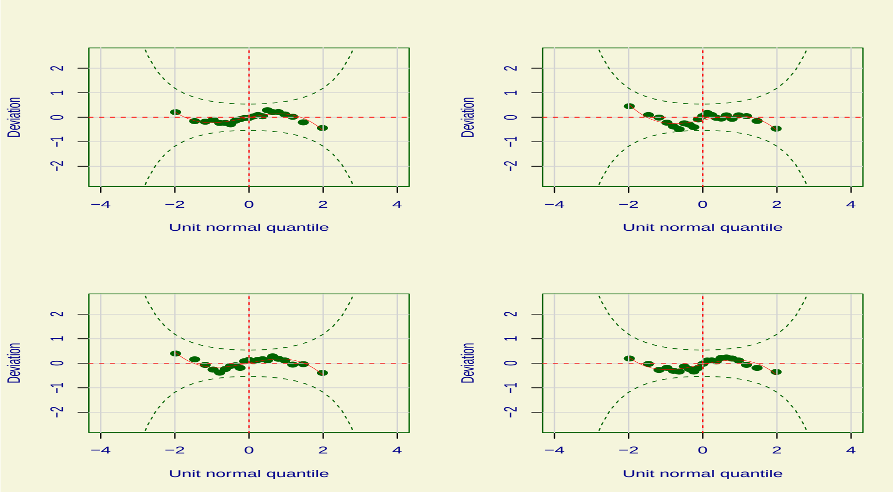
QQ plots RQA: modelo beta-binomial



Worm plots RQA: modelo binomial



Worm plots RQA: modelo beta binomial



Comparação entre os modelos

Modelo	AIC	BIC	AICc	SABIC	HQCIC	CAIC	DABM
bin.	117,874	122,052	120,374	109,704	118,781	126,052	0,106
bb	117,534	122,756	121,534	107,321	118,667	127,756	0,105

Estimativas (gamlss) 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	0,0389
	β_2	-0,15	0,22	[-0,58 ; 0,29]	-0,65	0,5219
	γ_2	0,54	0,25	[0,05 ; 1,03]	2,16	0,0452
	$(\beta\gamma)_{22}$	0,78	0,31	[0,18 ; 1,38]	2,54	0,0212
beta-binomial	α	-0,44	0,22	[-0,87 -0,02]	-2,05	0,0587
	β_2	-0,10	0,27	[-0,63 ; 0,44]	-0,36	0,7266
	γ_2	0,52	0,30	[-0,06 ; 1,10]	1,76	0,0978
	$(\beta\gamma)_{22}$	0,80	0,38	[0,06 ; 1,54]	2,11	0,0510
	σ	0,013	0,012	[0,00;0,035]*	-	-

(*) truncado à esquerda do zero

Observações

- O modelo bb se ajustou bem aos dados e melhor do que o modelo binomial. Assim, escolheremos o modelo bb.
- Usualmente, retirar-se-ia o parâmetro que representa a interação ($(\alpha\beta)_{22}$). Contudo, dado que o p-valor associado ao teste de nulidade para β_2 é maior do que para $(\alpha\beta)_{22}$, vamos ajusta um modelo sem (β_2).
- Com efeito (modelo binomial):

Modelo	AIC	BIC	AICc	SABIC	HQCIC	CAIC
$\beta_2 = 0$	115,66	119,84	118,16	107,49	116,57	123,84
$(\beta\gamma)_{22} = 0$	119,66	123,84	122,16	111,49	120,57	127,84

Modelo 2

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

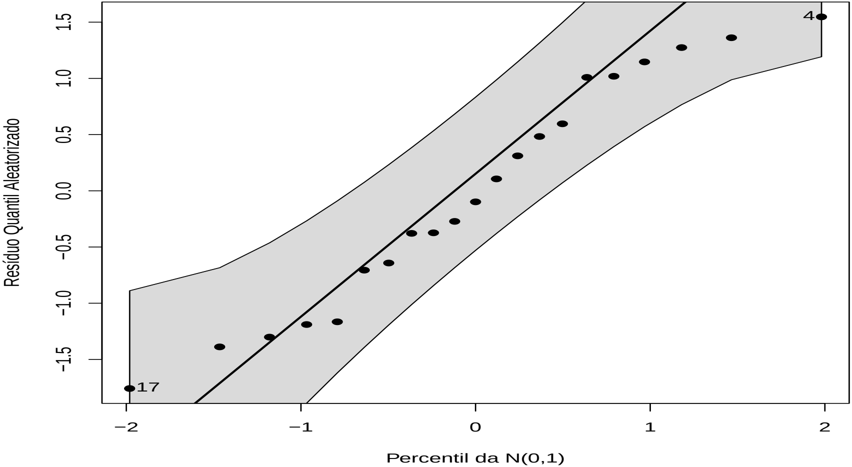
- n_{ij} : número total de sementes pertecentes a i-ésima espécie (1: O.aegyptiaca73, 2: O.aegyptiaca75) e tratados com o j-ésimo tipo de extrato (1: feijão,2:pepino), $n_{11} = 5, n_{12} = 5, n_{21} = 5, n_{22} = 6$.
- Y_{ijk} : é o total de sementes germinadas na k-ésimo réplica, pertencentes à i-ésima espécie, e tratadas com o j-ésimo tipo de extrato vegetal.

Modelo 2

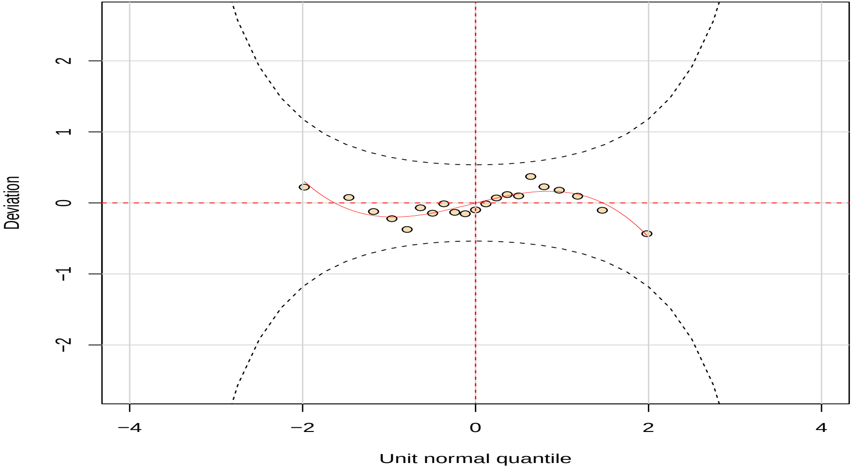
- Médias induzidas pelo modelo 2

$$
\begin{aligned}
\mu_{11} &= \frac{e^{\alpha}}{1+e^{\alpha}} \text{ (73, feijão)} \\
\mu_{12} &= \frac{e^{\alpha+\gamma_2}}{1+e^{\alpha+\gamma_2}} \text{ (73, pepino)} \\
\mu_{21} &= \frac{e^{\alpha}}{1+e^{\alpha}} \text{ (75, feijão)} \\
\mu_{22} &= \frac{e^{\alpha+\gamma_2+\delta_{22}}}{1+e^{\alpha+\gamma_2+\delta_{22}}} \text{ (75, pepino)}
\end{aligned}
$$

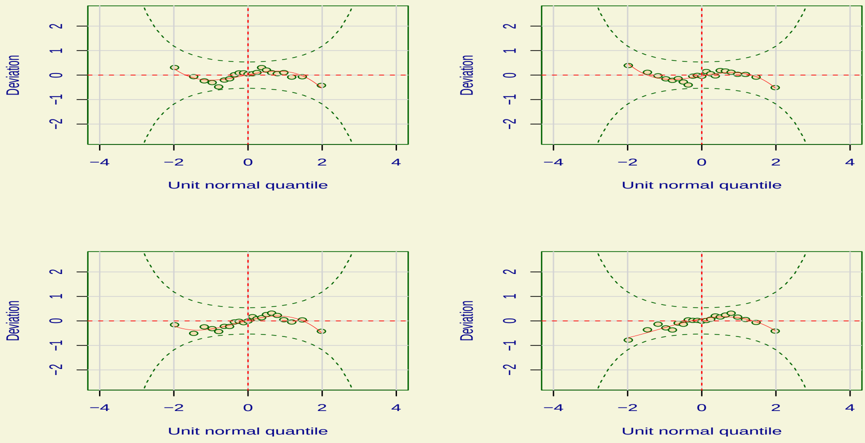
Gráfico de envelope: modelo 2



Worm plot: modelo 2



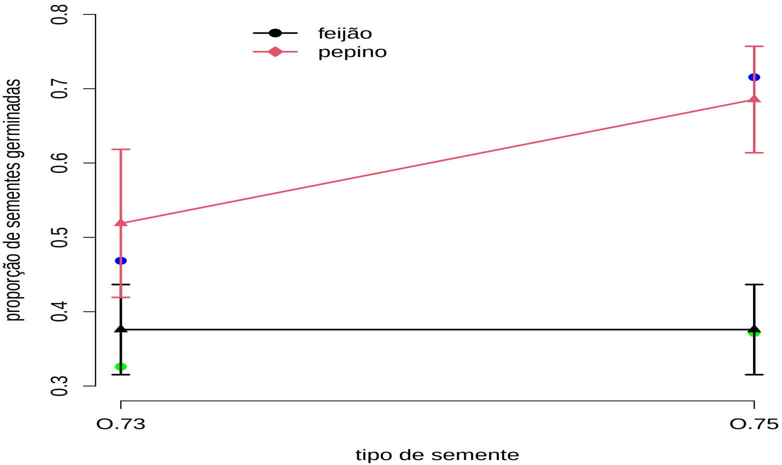
Worm plots: modelo 2



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 preditas



Médias preditas

Grupo	Est.	EP	IC(95%)
Feijão-O73	0,38	0,03	[0,32;0,44]
Feijão-O73	0,38	0,03	[0,32;0,44]
Pepino-O73	0,52	0,05	[0,42;0,62]
Pepino-O75	0,69	0,04	[0,61;0,76]

Melhor desempenho: Sementes do tipo O75 para as quais se usou extrato de pepino.

Comentários

- As conclusões entre o modelo beta binomial (bem ajustado) e binomial (mal ajustado) foram diferentes.
- Enquanto que o modelo binomial indicou presença de interação e efeito dos dois fatores (as quatro proporções médias são diferentes entre si), o modelo bb também identificou interação, mas efeito de tipo de semente somente para o extrato "pepino".
- Pesquisar sobre análise de influência para modelos de regressão beta binomiais.