# Uma aplicação de dados longitudinais para a distribuição odd log-logistica skew normal

Altemir da Silva Braga [1], Edwin M Marcos Ortega [2]

[1] Doutorando Estatística e Experimentação Agronômica / ESALQ.
[2] Professor Titular da Universidade de São Paulo/ESALQ.

## RESUMO

Fazer estudos envolvendo distribuições de probabilidades com aplicações a dados reais sempre é importante para avançar com as pesquisas na área da estatística. Neste trabalho destaca-se a distribuição com quatro parâmetros odd log-logística skew normal com modelos de covariância com efeitos aleatórios de intercepto. Para estimar os parâmetros do modelo foi utilizado o método da máxima verossimilhança. Para motivar a aplicação modelo foi realizado uma aplicação em dados proviniente de um delineamento experimental inteiramente casualizado com medida repetida no tempo. A nova proposta mostrou-se mais eficiente e obteve melhores resultados em relação ao modelo normal. Portanto, pode-se recomendar tal modelo para dados, principalmente, com evidencias de assimetria e bimodalidade.

**Palavras chave**: Log-logistica; Estimação verossimilhança; Skew; Bimodalidade.

# 1 INTRODUÇÃO

Em muitas áreas de pesquisas são comuns realizarem retiradas de amostras sobre a mesma unidade experimental, por exemplo, na avaliação do crescimento do peso de espécies animais ou o tempo que uma determinda espécie de fruta demora para amadurecer sobre o efeito de um tratamento, etc. Tais ensaios experimentais permitem que se avaliem as mudanças que ocorrem ao longo do tempo (DIGGLE, 1988) e (FERREIRA; MORAIS, 2013).

Tendo em vista que as medidas são realizadas nas mesmas unidades experimentais. Então, espera-se que ocorra correlação entre as medidas no tempo, assim como, a falta de homogeneidade de variâncias. Nesse sentido, é possível considerar que respostas para tempos mais próximos sejam mais correlacionadas do que aquelas com tempos mais distantes, tornando-se, uma característica comum a dados mensurados ao longo do tempo (LITTELL; HENRY; AMMERMAN, 1998).

Neste sentido, torna-se importante o uso de modelos lineares mistos, porque as amostras retiradas da mesma parcela não são independentes. Além disso, propõe-se um novo estudo utilizando uma distribuições de probabilidade odd log-logistica skew normal que é assimétrica e bimodal . Esse procedimento estatístico é extremante versátil e prático quando se tem a intensão de utilizar modelo de covariância por meio de curvas polinomiais de crescimento. Devido as possibilidades que se tem para se optar por uma estrutura de covariância que acomode a correlação, possivelmente, existente nas unidades experimentais.

Portanto, este estudo foi conduzido, com o objetivo de propor uma nova abordagem para analisar medidas repetidas no tempo por meio do modelo linear misto utilizando a distribuição odd log logística skew normal.

# 2 Especificação do modelo de regressão com efeito aleatório

Seja $Y_{ij}$ a j-ésima resposta associado ao i-ésimo grupo (indivíduo) para $i = 1, \ldots, N$ e $j = 1, \ldots, n_i$. Assumindo -se o modelo OLLSN com feito misto tem-se a seguinte estrutura hierárquica:

$$Y_{ij}|\mathbf{u}_i \overset{iid}{\sim} OLLSN(\mu_{ij}, \sigma^2, \lambda, \alpha) \text{ e } \mathbf{u}_i \overset{iid}{\sim} N(0, \sigma_m^2), \tag{1}$$

em que $\mu_{ij} = x_{ij}^T\boldsymbol{\beta} + z_{ij}^T\mathbf{u}_i$ e $x_{ij}^T = (x_{ij1}, \ldots, x_{ijp})^T$ é o vetor de variáveis explanatórias associados aos efeitos fixos. Nos estudos de Lobos (2010), foram apresentados modelos mistos com intercepto aleatório conforme o modelo dado por (HENDERSON, 1975).

$$Y_{ij} = x_{ij}^T\boldsymbol{\beta} + z_{ij}^T\mathbf{u}_i + \sigma\epsilon_{ij}, \tag{2}$$

em que $x_{ij}^T$ é uma matriz do planejamento experimental relativa a os parâmetros de efeitos fixos, de dimensões $n \times r$, e posto $r < n$; $\boldsymbol{\beta}$ é o vetor de parâmetros fixos, desconhecidos, de dimensão $r \times 1$; $z_{ij}^T$ é uma matriz de delineamento experimental relativa os efeitos aleatórios,$n \times q$, $\mathbf{u}_i$ é o vetor de efeitos aleatórios, desconhecidos, $q \times 1$, em que $\mathbf{u}_i \sim \mathrm{N}(0, \sigma_m^2)$; $\epsilon_{ij} \sim \mathrm{OLLSN}(0, \sigma^2, \lambda, \alpha)$ é o vetor de erros aleatórios não observáveis, $n \times 1$.

# 3 Estimação por máxima verossimilhança

Seja $y_{11}, \ldots, y_{IJ}$ uma amostra de tamanho n da distribuição OLLSN. Então, considerando o logarítmo da função verossimilhança da distribuição marginal, supondo o modelo normal para o efeito aleatório e o vetor de parâmetros $\boldsymbol{\theta} = (m, \boldsymbol{\tau}^T, \sigma, \lambda, \alpha)^T$, em que $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_I)^T$. Tem-se que o logaritmo da função de verossimilhança da expressão pode ser reescrita da eguinte forma:

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{i=1}^{I}\sum_{j=1}^{J}\log\prod_{\mathrm{i=1}}^{\mathrm{n_j}}\left\{\frac{2\alpha\phi(z_{ij})\Phi(z_{ij}\lambda)\Phi_{SN}^{\alpha-1}(z_{ij},\lambda)\left[1-\Phi(z_{ij},\lambda)\right]^{\alpha-1}}{\sigma\left\{\Phi_{SN}^{\alpha-1}(z_{ij},\lambda)+\left[1-\Phi(z_{ij},\lambda)\right]^{\alpha-1}\right\}^2}\right\} \\ &\quad \times\sum_{\nu=1}^{m}\frac{w_k^+}{\sqrt{\pi}}\exp\left\{-\frac{1}{2}\left(\frac{s_k^+}{\sigma_m}\right)^2\right\}, \end{aligned} \tag{3}$$

em que $z_{ij} = (y_{ij} - m - \tau_i)/\sigma$.

Sob certas condições de regularidade o vetor de parâmetros $\boldsymbol{\theta}$, em seu espaço paramétrico, tem distribuição assintótica $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ normal multivariada $N_{I+4}(0, K(\boldsymbol{\theta})^{-1})$, em que $K(\boldsymbol{\theta})$ é a matrix de informação esperada. A matriz de covariâncias assintóticas $K(\boldsymbol{\theta})^{-1}$ de $\widehat{\boldsymbol{\theta}}$ pode ser aproximada pela inversa de $(I+4)\times(I+4)$ da matriz de informação observada $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$, isto é, pode-se inverter a matrix de informação observada da função dos parâmetros para obter uma aproximação da matriz de covariâncias. Assim, com as EMVs podem fornecer regiões de confiança utilizando a normalidade assintótica. Então, as inferências assintóticas para os vetor de parâmetros $\boldsymbol{\theta}$ podem ser reralizadas utilizando aproximação normal $N_{I+4}(0,-\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1})$ para $\widehat{\boldsymbol{\theta}}$ e os erros- padrão para as EMVs podem ser obtidas pela raiz quadrada dos elementos da diagonal principal da inversa da matrix de informação observada e podem ser formuladas hipóteses a serem testadas. A distribuição normal assintótica $N_{I+4}(0,-\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1})$ pode ser utilizada para construir regiões aproximadas de confiança para o vetor de parâmetros $\boldsymbol{\theta}$. Neste caso, os intervalos de confiança assintóticos $100(1-\gamma)\%$ para cada componente do vetor de parâmetros $\theta_r$ é dado por:

$$IC_r = \left(\hat{\theta}_r - z_{\gamma/2}\sqrt{-\widehat{\ddot{L}}^{r,r}}, \hat{\theta}_r + z_{\gamma/2}\sqrt{-\widehat{\ddot{L}}^{r,r}}\right) \tag{4}$$

em que $-\widehat{\ddot{L}}^{r,r}$ denota o $r$-ésimo elemento da diagonal da inversa da matriz de informação observada $-\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}})^{-1}$ e $z_{\gamma/2}$ é o quantil $1-\gamma/2$ da distribuição normal padrão.

# 4 Resultados

Este experimento foi desenvolvido no Laboratório de Fisiologia e Bioquímica no Departamento de Ciências Biológicas da ESALQ, em Piracicaba. Este experimento teve como objetivo avaliar o uso de antioxidantes no lichias " Bengal ". O experimento foi utilizado o delineamento experimental inteiramente casualizado, com fatorial $5\times 6$ (tratamentos $\times$ por dias). Foram utilizadas quatro repetições para os tratamentos.

- $\mathbf{Y}_{ij}$: teor de ácido ascórbico;
- **Tratamentos**: **Trat1** - sem tratamentos (água destilada); **Trat2** - imersão em sulução 4-hexylresorcinol (300 mg $L^{-1}$); **Trat3** - imersão em sulução ácido ascórbico (300 mg $L^{-1}$); **Trat4** - imersão em sulução ácido citrico (300 mg $L^{-1}$); **Treat5** - imersão em sulução de ácido cítrico + ácido ascórbico (300 mg $L^{-1}$);
- **T: Fator longitudinal:** (0, 3, 6, 9, 12 e 15 dias).

Na Tabela 1 estão as estimativas dos parâmetros de escala ($\sigma$, $\sigma_m$) dos parâmetros de forma ($\nu$, $\tau$). Para estudar o efeito linear entre a variável resposta Teor de ácido ascórbico (Y) em função dos tratamentos e do fator longitudinal (T) considerou-se o modelo de covariância (5) OLLSN dado por:

$$Y_{tij} = \underbrace{\beta_1}_{efeito\ \text{intercepto}} + \underbrace{\beta_2 T}_{efeito\ angular} + \underbrace{\beta_{1t}}_{efeito\ \text{aleatorio}} + \sigma z_{ij} \tag{5}$$

Tabela 1: Estimativas de máxima verossimilhança para os parâmetros de escala e de forma e o critério da informação para o conjunto de dados.

| Parâmetros | Estimativas | Parâmetros | Estimativas | Preditor linear |
|---|---|---|---|---|
| $\beta_1$ | -0.6321 | $\beta_2$ | 34.5697 | 4.364x$10^{-08}$ |
| $\beta_1$ | -0.6321 | $\beta_2$ | 34.5697 | -5.223x$10^{-08}$ |
| $\beta_1$ | -0.6321 | $\beta_2$ | 34.5697 | 1.120x$10^{-08}$ |
| $\beta_1$ | -0.6321 | $\beta_2$ | 34.5697 | -3.411x$10^{-08}$ |
| $\beta_1$ | -0.6321 | $\beta_2$ | 34.5697 | 2.531x$10^{-08}$ |
| $\beta_1$ | -0.6321 | $\beta_2$ | 34.5697 | 6.177x$10^{-09}$ |
| Parâmetros | Estimativas | EP | Z-valor | P-valor |
| $\sigma^2$ | 9.99x$10^{-01}$ | 9.99x$10^{-01}$ | - | - |
| $\sigma_{\beta_{1t}}$ | 3.77x$10^{-09}$ | 6.14x$10^{-05}$ | - | - |
| $\lambda$ | 0.8293 | 0.1772 | 4.679 | 0.0000 |
| $\alpha$ | -0.0915 | 0.09020 | -1.015 | 0.3120 |

Portanto, conforme a Tabela 1 pode-se afirmar que para cada dia estudado o teor de ácido ascóbico teve um decréscimo de 0.6321. Como oe efeitos aleatórios foram praticamente zero, então o preditor linear não interferiu no efeito de intercepto.

# 5 Considerações finais

Neste estudo foi proposto modelo de covariância com efeitos aleatórios para distribuição odd log logística skew normal. A nova distribuição é versátil e analiticamente tratável, além disso, o juste poder ser realizado utilizando o método da máxima verossimilhança por meio de rotinas computacionais implementadas no software estatístico R.

# Referências


[1] DIGGLE, P. J.An approach to the analysis of repeated measurements. Biometrics, JSTOR, p. 959-971, 1988.

[2] FERREIRA, W. L.; MORAIS, A. R. de. Análise da influência do café no ganho de peso de animais (ratos) por meio de modelo linear misto. Rev. Bras. Biom, v. 31, n. 4, p. 485-500, 2013.

[3] HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. Biometrics, JSTOR, p. 423-447, 1975.

[4] LITTELL, R. C.; HENRY, P.R.; AMMERMAN, C.-B. Statistical analysis of repeated measures data using sas procedures. Journal of animal science, v. 76, n. 4, p. 1216-1231, 1998.

[5] LOBOS, C. M. V.. Modelos log-Birnbaum-Saunders mistos. Tese (Doutorado) - Universidade de São Paulo, 2010.