



## MODELO MISTO PARA ANÁLISE DE DADOS DE HCV/HIV

Beatriz Regina Brum<sup>1</sup>, Omar Neves Pereira<sup>2</sup>, Claudinei Mesquita da Silva<sup>3</sup> e Isolde Previdelli<sup>4</sup>

<sup>1</sup>Programa de Pós-graduação em Bioestatística - UEM

<sup>2</sup>Programa de Pós-graduação em Bioestatística - UEM

<sup>3</sup>Programa de Pós-graduação em Ciências da Saúde - UEM

<sup>4</sup>Departamento de Estatística - UEM

### RESUMO

É comum nas áreas médicas o interesse na progressão de doenças, bem como no comportamento que biomarcadores naturais assumem em determinado período de tempo. A Síndrome da Imunodeficiência Humana Adquirida é um exemplo de enfermidade que necessita de observações ao longo do tempo. Cada indivíduo por sua vez, assume um comportamento particular, caracterizando seu próprio perfil. Modelos capazes de acomodar observações coletadas ao longo do tempo são caracterizados como dados longitudinais, vistas como ferramentas essenciais para explorar o comportamento e histórico dessa doença. Os modelos mistos assumem nesse contexto um papel fundamental na análise de medidas repetidas, sua grande flexibilidade garante com os efeitos aleatórios uma maior absorção da heterogeneidade contida entre e dentro de cada indivíduo. Sob essa perspectiva um modelo misto foi ajustado a um conjunto de dados de pacientes coinfectados pelos vírus HCV/ HIV. A análise de resíduos apresentou a adequabilidade satisfatória do modelo mediante os dados coletados.

**Palavras chave:** HIV, HCV, dados longitudinais e modelos mistos.

## 1 INTRODUÇÃO

Embora a infecção pelo vírus da Imunodeficiência Humana Adquirida venha tendo seu impacto reduzido na opinião pública, na medida em que, coquetéis antirretrovirais vem se mostrando eficazes no combate a proliferação do vírus HIV, novos vírus infecciosos vem se destacando no que diz respeito ao quadro clínico de pacientes com tal enfermidade. A preocupação com a comunidade de pacientes coinfectados por HCV/HIV, induz pesquisadores de diferentes áreas se interessarem por descrever o comportamento de biomarcadores longitudinais. A Estatística por sua vez, com seu caráter interdisciplinar fornece ferramentas capazes de descrever o comportamento médio de uma comunidade específica. Modelos como os de efeito aleatório são em geral, preferidos devido a sua grande flexibilidade, capaz de caracterizar a variabilidade de cada sujeito, bem como seu perfil médio. Esses modelos levam em consideração as particularidades individuais, possibilitando aos pesquisadores atuarem de forma interdisciplinar em um vasto campo de aplicações, auxiliando assim a tomada de decisões. A análise foi feita em um conjunto de dados referente a pacientes do Sudoeste do Paraná, coinfectados por HCV/HIV, o pesquisador tinha o interesse no perfil médio da contagem de CD4 ao longo do tempo para a amostra. A análise de resíduos confirmaram o bom ajuste do modelo.

## 2 METODOLOGIA

De acordo com Rizopoulos (2012), os modelos longitudinais apresentam uma riqueza de informações, que são extraídas de indivíduo a indivíduo, onde cada sujeito apresenta seu próprio perfil médio. Dessa forma é inevitável admitir que existe uma variabilidade inerente dos indivíduos. Ao utilizar um modelo de regressão simples, assume-se que as variáveis predictoras não contém erros de medição. Os modelos mistos são usados para suprir as lacunas de um modelo linear simples, pois leva em consideração os

erros de medição e atribui variabilidade as estimativas dos parâmetros fixos e aos efeitos aleatórios. A representação matricial de um modelo misto pode ser dada por

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon. \quad (1)$$

Para Searle et al.(2009), os efeitos aleatórios podem ser representados por  $\mathbf{b}_i$ , sendo paralela a  $\mathbf{X}\beta$ , onde  $\mathbf{X}$  é uma matriz de incidência,  $\mathbf{b}$  o vetor de efeitos aleatórios e  $\mathbf{Z}$  pode ser um subconjunto da matriz  $\mathbf{X}$ . A matriz de covariância de  $\mathbf{Y}$  é

$$\mathbf{V} = \mathbf{Z}Var(\mathbf{b})\mathbf{Z}' + Var(\epsilon). \quad (2)$$

onde  $G = Var(b)$ , representa a matriz de variabilidade dos efeitos aleatórios de modo que  $b \sim N(0, G)$  e  $R = Var(e)$  representa a variabilidade dos erros associados a resposta  $Y$  em que  $e \sim N(0, R)$ . As matrizes  $G$  e  $R$  são escritas da seguinte forma

$$\mathbf{G} = \begin{bmatrix} Var(b_{0i}) & cov(b_{0i}, b_{1i}) & \cdots & cov(b_{0i}, b_{qi}) \\ cov(b_{0i}, b_{1i}) & Var(b_{1i}) & \cdots & cov(b_{1i}, b_{qi}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(b_{0i}, b_{m_i i}) & cov(b_{1i}, b_{qi}) & \cdots & Var(b_{m_i i}) \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} Var(\epsilon_{0i}) & cov(\epsilon_{0i}, \epsilon_{1i}) & \cdots & cov(\epsilon_{0i}, \epsilon_{m_i i}) \\ cov(\epsilon_{0i}, \epsilon_{1i}) & Var(\epsilon_{1i}) & \cdots & cov(\epsilon_{1i}, \epsilon_{m_i i}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\epsilon_{0i}, \epsilon_{m_i i}) & cov(\epsilon_{1i}, \epsilon_{m_i i}) & \cdots & Var(\epsilon_{m_i i}) \end{bmatrix}$$

A estimação dos parâmetros é dada pelo método da máxima verossimilhança restrita ou residual escrita como

$$\ell_{re}(\beta, \tau) = -\frac{1}{2} \sum_{i=1}^N \log |\mathbf{V}_i| - \frac{1}{2} \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}) \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}) - \frac{1}{2} \log \left| \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right|. \quad (3)$$

Para a predição dos efeitos aleatórios a verossimilhança hierárquica proposta por Lee e Nelder (1996), é utilizada, sendo dada como segue

$$f(\mathbf{y}, \mathbf{b}; \beta, \tau) = f_{Y|b}(\mathbf{y}; \beta) f_b(\mathbf{b}; \tau) \quad (4)$$

$$= \frac{1}{\sqrt{2\pi^q} \det |\mathbf{R}|} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{b})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{b}) \right\}$$

$$- \frac{1}{\sqrt{2\pi^q} \det |\mathbf{G}|} \exp \left\{ -\frac{1}{2} \mathbf{b}^T \mathbf{G} \mathbf{b} \right\},$$

em que  $\tau$  representa o vetor de variâncias das estimativas dos efeitos aleatórios que compõe a matriz  $V$ .

### 3 RESULTADOS E DISCUSSÕES

O conjunto de dados corresponde a 95 pacientes coinfectados por HCV/HIV, dos quais 35 são do grupo de coinfeção e 61 do grupo controle.

Ao analisar o comportamento da variável resposta, que neste caso é a contagem de CD4 por  $mm^3$  de sangue, provenientes de contagem, foi preferido utilizar um modelo normal, de modo que a variável resposta foi transformada, aplicado a raiz quadrada na mesma. Os histogramas da Figura 1, representam a densidade da contagem de CD4 a direita, sugerindo que a distribuição dos dados é assimétrica e a direita a contagem do CD4 transformada.

Os gráficos de perfis (Figura 2) permitem verificar o comportamento da raiz quadrada da contagem do CD4 em indivíduos do grupo controle (somente HIV) e coinfectados (HCV/HIV).

Aparentemente o grupo de coinfeção apresenta maior variabilidade na linha de base (tempo zero) e ao longo do tempo apresentam uma leve diferença de inclinação. Um modelo de efeitos mistos foi ajustado aos dados, assumindo que os pacientes apresentam  $\sqrt{CD4}$  distintas entre si e também comportam-se diferente ao longo do tempo. O modelo escolhido é escrito por

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) t_{ij} + \beta_2 g_j + \epsilon_{ij} \quad (5)$$

$$b_i \sim N(0, \sigma_b^2)$$

$$\epsilon_{ij} \sim N(0, \sigma^2).$$

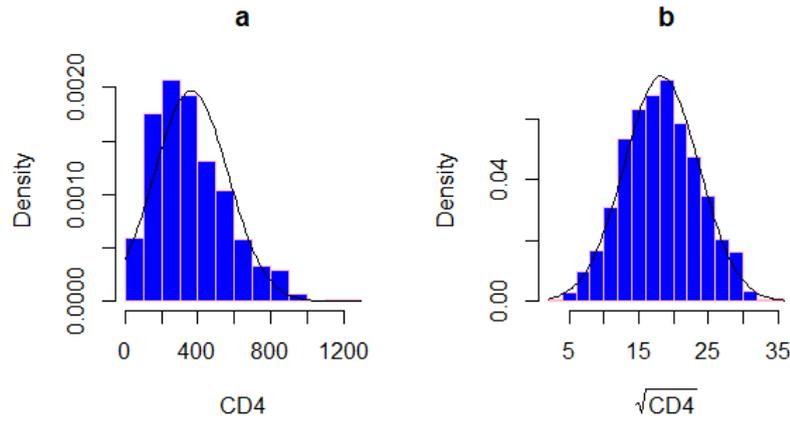


Figura 1: (a) Distribuição da variável CD4; (a) dados reais, (b) dados transformados por raiz quadrada.

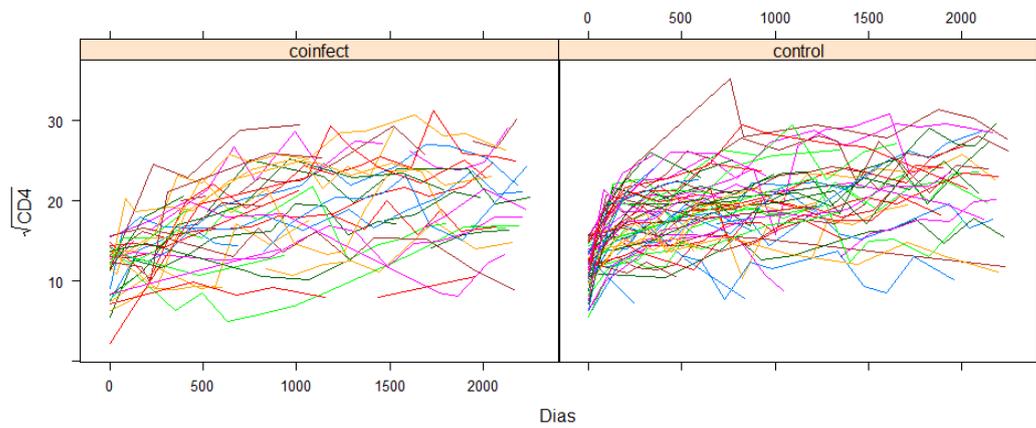


Figura 2: Perfis de pacientes dos grupos coinfeção e controle.

o efeito aleatório ocorre tanto no intercepto quanto na inclinação ( $b_{i0}, b_{i1}$ ),  $t_{ij}$  representa o tempo do  $i$ -ésimo indivíduo do  $j$ -ésimo grupo (controle e coinfeção) e  $g_i$  o  $i$ -ésimo grupo. A estimativa dos parâmetros assim como seu erro podem ser vistos na Tabela 1.

Na Tabela 1 se apresenta o resumo dos efeitos das estimativas na contagem da raiz do CD4 para os efeitos fixos, na qual é possível observar que o intercepto ( $\beta_0$ ), o tempo ( $\beta_1$ ) e o grupo ( $\beta_3$ ) a 5% de significância. O tempo tem efeito positivo, evidenciando uma aumento de 0,45% na taxa da variável resposta ( $\sqrt{CD4}$ ).

Tabela 1: Coeficientes de regressão estimados (efeitos fixos) e erros padrão

Variavel	Estimativa	EP	DF	t	p-valor
$\beta_0$	13,6655	0,5522	828	24,7456	0,0000
$\beta_1$	0,0045	0,0003	828	11,7909	0,0000
$\beta_2$	1,4013	0,6558	828	2,136707	0,0329

Por meio das matrizes  $\mathbf{G}$  é possível verificar a variabilidade na linha de base em 8,3505, na taxa de inclinação em 0,0029, com  $\rho = -0.051$  e para a matriz  $\mathbf{R}$  a variância dos erros de medição foi 2,7988.

$$\mathbf{G} = \begin{bmatrix} 8,3505 & -0,0004 \\ -0,0004 & 0,0029 \end{bmatrix}$$

Os gráficos de resíduos (Figura 3 e Figura 4) evidenciam o bom ajuste do modelo.

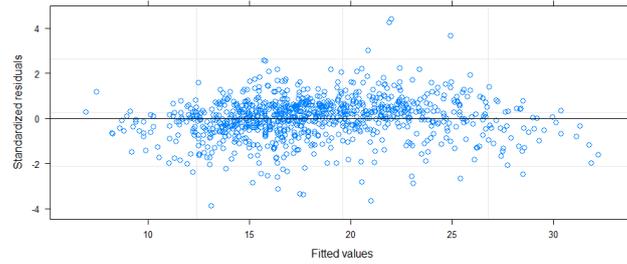


Figura 3: Gráfico de resíduos versus preditos.

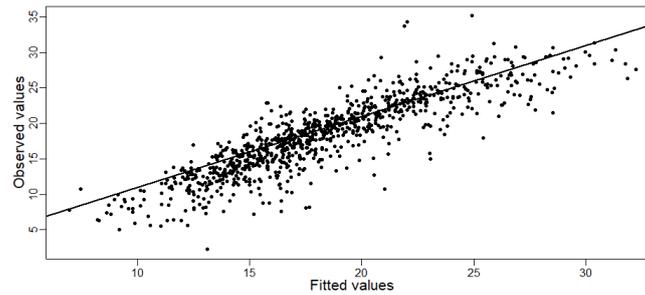


Figura 4: Gráfico de resíduos, valores preditos versus valores estimados.

## 4 CONCLUSÃO

O efeito do intercepto, tempo e do grupo foram ambos significativos. Também foi possível observar que na linha de base há uma grande variabilidade entre os indivíduos, por outro lado ao longo do tempo essa variância é pequena. Por meio da análise de resíduos verificou-se que o modelo ajustado representou bem o comportamento da contagem da raiz quadrada do CD4 em função do tempo.

## 5 AGRADECIMENTOS

PBE, DCS, Fundação Araucária e Caixa Econômica Federal.

## Referências

- [1] DIGGLE, P. J.; HEAGERTY, P.; LIANG, K-Y; ZEGER, S. L. Analysis of longitudinal data (2nd edition). Clarendon Press: Oxford, 2002.
- [2] LEE, Y.; NELDER, J. A. Hierarchical generalized linear models. Journal of the Royal Statistical Society. Series B (Methodological), JSTOR, p. 619–678, 1996.
- [3] RIZOPOULOS, D. Joint models for longitudinal and time-to-event data: With applications in R. [S.l.]: CRC Press, 2012.
- [4] SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. Variance components. [S.l.]: John Wiley the Sons, 2009. v. 391. 138–142 p.