



## Comparando perfis longitudinais em uma mesma unidade experimental utilizando os modelos mistos

Omar C. N. Pereira<sup>1</sup>, Márcia L. A. dos Santos<sup>1</sup>, Tiago Sugiura<sup>1</sup>,  
Beatriz R. Brum<sup>1</sup>, Camila Rodrigo<sup>2</sup>,  
Tuane Krupek<sup>3</sup>, Sueli Ichisato<sup>2</sup>,  
Roberto Barbosa Bazotte<sup>3</sup>, Isolde Previdelli<sup>1</sup>

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

<sup>2</sup>UEM - Pós-graduação em Ciências Biológicas

<sup>3</sup>UEM - Pós-graduação em Enfermagem

### RESUMO

Por vezes, na área da saúde, uma variável resposta acompanhada ao longo do tempo em um mesmo indivíduo pode ter seu comportamento influenciado por uma intervenção realizada durante o período experimental. Nessa perspectiva, o presente estudo vem usar uma metodologia simples baseada no ajuste de modelos de efeitos mistos para a verificação de diferenças significativas no perfil de resposta perante uma intervenção. A notoriedade dessa proposta consiste do emprego de todas as observações repetidas da variável resposta para comparar o perfil longitudinal de um indivíduo nos momentos anterior e posterior a intervenção. O método foi motivado por dois conjuntos de dados reais, os quais necessitaram desta abordagem para responder ao objetivo da pesquisa. Modelos lineares de efeitos mistos foram ajustados no primeiro conjunto de dados, o qual refere-se a contagem de células CD4 em pacientes com HIV positivo que, durante 30 dias consecutivos, receberam uma suplementação alimentar com glutamina. No segundo problema, modelos não lineares de efeitos mistos foram ajustados às medições da massa corporal de recém nascidos prematuros, cuja dieta alimentar inicial baseou-se no leite materno, sendo posteriormente substituído por um suplemento alimentar. A metodologia utilizada foi capaz de detectar diferença na tendência de crescimento da contagem de células CD4 promovida pela suplementação com glutamina nos perfis dos pacientes observados. Além disso, forneceu evidências que sugerem a suplementação alimentar como alternativa ao leite materno na dieta de recém nascidos prematuros, por manter a tendência de crescimento da massa corporal.

**Palavras chave:** Modelo linear e não linear misto, perfil longitudinal, recém nascido prematuro, leite materno, HIV, glutamina.

## 1 INTRODUÇÃO

Estudos que observam repetidamente ao longo do tempo determinada variável ou característica de um indivíduo são frequentes na área da saúde. Para exemplificar, considere que determinada variável foi observada repetidamente ao longo do tempo num mesmo indivíduo e, em dado momento, ocorre uma intervenção. A natureza desta intervenção pode ser alimentar, educacional, social, cirúrgica, entre outras. Dessa forma, o propósito é verificar a eficácia da intervenção, isto é, deseja-se identificar possíveis mudanças nas observações anteriores e posteriores a intervenção, com o interesse de atribuir tal efeito a mesma.

A aplicação da metodologia utilizada é apresentada a partir de dois conjuntos de dados reais, um com doze unidades experimentais e outro com oito. Apesar de ambos poderem ser tratados como casos de pequenas amostras, as repetições em cada unidade experimental possibilita uma maior quantidade de valores observados para a análise e um aumento nos graus de liberdade dos modelos propostos. Além disso, desde que as correlações seja consideradas de forma adequada em um modelo de efeitos mistos, a

quantidade de unidades experimentais pode não limitar a análise (Pereira et al., 2017). Dessa forma, o primeiro conjunto de dados, ajustado por meio de um modelo linear de efeitos mistos, refere-se a dados de contagem de células CD4 em pacientes com HIV positivo que, durante 30 dias consecutivos, receberam uma suplementação alimentar com glutamina. O segundo conjunto de dados, ajustado por meio de um modelo não linear de efeitos mistos, diz respeito a massa corporal de recém nascidos prematuros internados numa maternidade, os quais receberam leite materno inicialmente e um suplemento alimentar comercial posteriormente. Em ambos os casos, o objetivo é verificar a eficácia das intervenções alimentares, ou seja, se as mudanças nas dietas resultaram no aumento do número de células CD4 nos pacientes HIV positivo e no aumento da taxa de variação da massa corporal das crianças prematuras, respectivamente.

## 1.1 Descrição da metodologia

A relevância da metodologia utilizada neste estudo refere-se ao emprego do perfil de respostas na sua totalidade para compará-lo perante uma intervenção. Para isso, dois modelos de efeitos mistos são ajustados no perfil com a finalidade de modelar as observações repetidas anteriores e posteriores a intervenção e, em seguida, compará-las. Os modelos de efeitos mistos, introduzidos por Laird and Ware (1982), constituem uma importante ferramenta de regressão para dados correlacionados e são representados por

$$\mathbf{y}_i = g(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\beta}, \mathbf{b}_i) + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n, \quad (1)$$

em que  $\mathbf{y}_i = (y_{i1}, \dots, y_{ip_i})^T$  com dimensão  $(p_i \times 1)$  é o perfil de respostas do  $i$ -ésimo indivíduo com  $i = 1, \dots, n$ ,  $\boldsymbol{\beta}$  é um vetor com dimensão  $(t \times 1)$  de efeitos fixos,  $\mathbf{X}_i = (x_{i1}, \dots, x_{im})$  é uma matriz conhecida de especificação dos efeitos fixos com dimensão  $(p_i \times t)$ ,  $\mathbf{b}_i$  é um vetor com dimensão  $(q \times 1)$  de efeitos aleatórios,  $\mathbf{Z}_i$  é uma matriz, conhecida e de posto completo, de especificação dos efeitos aleatórios,  $g$  é uma função com derivadas contínuas até a segunda ordem e  $\boldsymbol{\epsilon}_i$  é um vetor de erros aleatórios com dimensão  $(p_i \times 1)$ . Segue também que  $\mathbf{b}_i \sim N_q(\mathbf{0}, \mathbf{G})$ , e  $\boldsymbol{\epsilon}_i \sim N_{p_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ , em que  $\mathbf{G}$  e  $\boldsymbol{\Sigma}_i$  são matrizes  $(q \times q)$  e  $(p_i \times p_i)$  de covariâncias, sendo matrizes simétricas definidas positivas. Além disso,  $\mathbf{b}_i$  e  $\boldsymbol{\epsilon}_i$  são variáveis aleatórias independentes.

A presença dos efeitos aleatórios no modelo permite que os efeitos fixos variem aleatoriamente de um indivíduo para outro, descrevendo o comportamento individual e incorporando a estrutura de covariância inerente a dados com medidas repetidas. Para melhor compreender, considere o caso particular em que  $g$  é uma função linear dos efeitos fixos  $\boldsymbol{\beta}$  e dos efeitos aleatórios  $\mathbf{b}_i$ , podendo ser escrita como  $g = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$ . Então, um modelo linear de efeitos mistos é obtido. Considere ainda que tal modelo declara o tempo como uma variável preditora e possui efeitos aleatórios no intercepto e na inclinação da função linear  $g$ , ou seja,

$$y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})x_{ij} + \epsilon_{ij}, \quad j = 1, \dots, p_i, \quad (2)$$

em que  $y_{ij}$  representa a resposta do  $i$ -ésimo indivíduo observada no  $j$ -ésimo tempo e  $\epsilon_{ij}$  é o erro aleatório associado a tal resposta. Além disto,  $\beta_0$  e  $\beta_1$  são os efeitos fixos e descrevem a resposta média compartilhada por todos os indivíduos. Assim, ao acrescentar o efeito aleatório  $b_{0i}$  no intercepto,  $\beta_0$ , deseja-se incorporar ao modelo a diferença entre o intercepto médio e o intercepto específico do  $i$ -ésimo indivíduo. Em geral, essa variabilidade entre indivíduos é proveniente de fatores não ponderados na análise, como fatores genéticos, por exemplo (Diggle et al., 1994; Molenberghs and Verbeke, 2001). Semelhantemente, ao acrescentar o efeito aleatório  $b_{1i}$  na inclinação,  $\beta_1$ , deseja-se capturar o desvio da inclinação particular do  $i$ -ésimo indivíduo com relação a inclinação média.

É importante atentar ao fato de que a inclinação  $\beta_1$  da função linear  $g$  é interpretada como a taxa de variação da variável observada  $y_{ij}$  com relação a variável preditora  $x_{ij}$ , isto é, expressa a alteração sofrida pela variável  $y_{ij}$  à medida que o tempo se modifica. Com o acréscimo do efeito aleatório,  $(\beta_1 + b_{1i})$  é interpretada como taxa de variação da resposta ao longo do tempo específica do  $i$ -ésimo indivíduo. Nessa perspectiva, considere que dois modelos lineares de efeitos mistos, análogos a expressão (2), são ajustados no perfil de respostas do indivíduo  $i$ , antes e depois da intervenção. Sendo assim, suponha que  $(\beta_{a_1} + b_{a_{1i}})$  e  $(\beta_{d_1} + b_{d_{1i}})$  indicam a taxas de variação das respostas anteriores e posteriores a intervenção, respectivamente. Portanto, verificar a influência de uma intervenção na variável observada  $y_{ij}$  equivale a verificar se há diferença estatística entre as taxas de variação dos modelos lineares de efeitos mistos ajustados no perfil do  $i$ -ésimo indivíduo antes e depois da intervenção.

Dessa forma, a verificação da existência de diferença estatística entre as inclinações pode ser realizada utilizando o teste t de student pareado com  $n - 1$  graus de liberdade, sendo a hipótese a ser testada  $H_0$  e a hipótese alternativa  $H_1$  enunciadas como,

$$\begin{aligned} H_0 &: \beta'_{a_1} = \beta'_{d_1} \\ H_1 &: \beta'_{a_1} \neq \beta'_{d_1} \end{aligned}$$

em que  $\beta'_{a_1}$  e  $\beta'_{d_1}$  são os parâmetros das inclinações das retas antes e depois da intervenção.

Não rejeitar  $H_0$  significa que as taxas de variação anterior e posterior a intervenção não são estatisticamente diferentes, dado que o comportamento da variável observada  $y_{ij}$  é inalterado pela intervenção. Melhor dizendo, ao longo do tempo, o comportamento da variável observada seguiu o percurso esperado na ausência de intervenções. Em contrapartida, rejeitar  $H_0$  significa que as taxas de variação anterior e posterior a intervenção são estatisticamente diferentes, ou seja, o comportamento esperado da variável  $y_{ij}$  ao longo do tempo  $x_{ij}$  é influenciado pela presença da intervenção.

A metodologia proposta continua válida quando  $g$  é uma função não linear dos efeitos fixos  $\beta$  e dos efeitos aleatórios  $b_i$ , dando origem a um modelo não linear de efeitos mistos.

## 1.2 Contagem de células CD4

Um total de 12 pacientes foi selecionado de um programa especial de atendimento à pessoas com doenças sexualmente transmissíveis/AIDS da cidade de Maringá, Paraná, Brasil, entre os anos de 2014 e 2015. Esses pacientes tiveram a contagem de células CD4 registradas em visitas ao local de atendimento. Entre as visitas, no período de 30 dias, os pacientes receberam diariamente uma suplementação alimentar com glutamina e nos momentos inicial e final deste período, a contagem de células CD4 foi registrada. Assim, cada paciente possui um perfil de respostas composto pelas contagens de células CD4, sendo que a penúltima contagem em cada perfil representa o momento do início da suplementação com glutamina e a última contagem o final dessa intervenção. Diante do exposto, quer-se responder a seguinte questão: A suplementação com glutamina promoveu o aumento de células CD4 em pacientes HIV positivo?

Dessa forma, dois modelos lineares de efeitos mistos, com intercepto e inclinação aleatórios, foram ajustados para as contagens de células CD4 antes e depois da suplementação com glutamina. Em ambos os modelos, considerou-se o tempo como uma variável preditora e como *base line* o momento inicial das visitas dos pacientes ao local de atendimento. Deste modo, o primeiro modelo é expresso por,

$$y_{a_{ij}} = (\beta_{a_0} + b_{a_{0i}}) + (\beta_{a_1} + b_{a_{1i}})x_{a_{ij}} + \epsilon_{a_{ij}}, \quad (3)$$

em que  $y_{a_{ij}}$  representa a resposta do  $i$ -ésimo paciente no  $j$ -ésimo tempo e  $(\beta_{a_0} + b_{a_{0i}})$  e  $(\beta_{a_1} + b_{a_{1i}})$  denotam o intercepto e a inclinação para o  $i$ -ésimo paciente, respectivamente. Por sua vez, o segundo modelo é escrito como,

$$y_{d_{ij}} = (\beta_{d_0} + b_{d_{0i}}) + (\beta_{d_1} + b_{d_{1i}})x_{d_{ij}} + \epsilon_{d_{ij}}, \quad (4)$$

em que  $y_{d_{ij}}$  representa a resposta do  $i$ -ésimo paciente no  $j$ -ésimo tempo e  $(\beta_{d_0} + b_{d_{0i}})$  e  $(\beta_{d_1} + b_{d_{1i}})$  denotam o intercepto e a inclinação para o  $i$ -ésimo paciente depois da intervenção, respectivamente.

Considerando que a transformação raiz quadrada da variável contagem de células CD4 a torna com distribuição aproximadamente normal (Brown and Prescott, 2014; Rizopoulos, 2012), a verificação da presença de diferença estatística entre as inclinações é realizada pelo teste t de student pareado, de maneira que a hipótese a ser testada  $H_0$  e a hipótese alternativa  $H_1$  são enunciadas como,

$$\begin{aligned} H_0 &: \beta'_{a_1} = \beta'_{d_1} \\ H_1 &: \beta'_{a_1} < \beta'_{d_1}. \end{aligned}$$

Realizando o teste em nível de significância de 0.05, o valor do nível descritivo obtido foi inferior a 0.001. Assim, tem-se evidências de que a suplementação com glutamina influenciou na tendência de crescimento da variável observada, promovendo um aumento na contagem de células CD4 maior do que o esperado ao longo do tempo supondo a ausência de intervenções.

## 1.3 Massa corporal de crianças prematuras

A evolução da massa corporal de recém-nascidos nos primeiros dias de vida, pode ser usada como um indicador da adequação da amamentação (Gartner et al., 2005). Em vista disso, medições da massa corporal de oito recém-nascidos prematuros foram registradas ao longo do tempo, a fim de comprovar a eficácia de uma intervenção alimentar. Durante os primeiros dias de vida, todas as crianças foram alimentadas com leite materno, *LM*, e, após alguns dias, passaram a ser alimentadas unicamente com um suplemento alimentar comercial, *FM*. O momento da mudança na alimentação foi peculiar a cada

criança, porém todas receberam os dois alimentos na mesma sequência. Além disso, como recém-nascidos perdem, em média,  $\sim 5 - 7\%$  da massa corporal nos primeiros 2-3 dias de vida (Bertini et al., 2015), a menor massa ao longo do período de avaliação foi tomada como medida inicial do perfil de cada criança. Assim, o perfil de respostas do *i-ésimo* recém nascidos é composto por várias medições da massa corporal antes e depois da intervenção alimentar, sendo o valor inicial correspondente a menor medição. Nessas circunstâncias, deseja-se responder a seguinte questão: a dieta com o suplemento alimentar comercial promoveu uma evolução na massa corporal das crianças prematuras equivalente ou superior ao leite materno?

É essencial ponderar que a taxa de variação da massa corporal  $y_i$  de cada criança ao longo do tempo  $x$ ,  $dy_i/dx$ , é proporcional a massa corporal da criança num dado instante  $x_{ij}$ . Quer dizer, quanto maior a massa corporal da criança, maior será a variação da mesma. Assim, resolvendo a equação diferencial  $dy_i/dx = ky_i$ , sendo  $k$  uma constante de proporcionalidade e considerando que o tempo inicial é aquele no qual a criança atingiu a menor massa,  $\beta_{0i}$ , obtem-se que  $y_i(x) = \beta_{0i} \exp(\beta_{1i}x)$ , sendo  $\beta_{0i}$  e  $\beta_{1i}$  parâmetros a serem estimados. Portanto, uma relação não linear baseada no mecanismo de geração das respostas é estabelecida entre a taxa de variação da massa corporal de recém nascidos e o tempo.

Realizando o teste t de student pareado em nível de significância de 0.05, o valor do nível descritivo obtido foi  $\sim 0.7$ . Portanto, tem-se evidências de que a suplementação alimentar comercial não influenciou na tendência de crescimento da taxa de variação da massa dos recém nascidos prematuros, impossibilitando um aumento da massa maior do que o esperado ao longo do tempo perante a dieta do leite materno.

## Referências

- Bertini, G., Breschi, R., and Dani, C. (2015). Physiological weight loss chart helps to identify high-risk infants who need breastfeeding support. *Acta Paediatrica*, 104(10):1024–1027.
- Brown, H. and Prescott, R. (2014). *Applied mixed models in medicine*. John Wiley & Sons.
- Diggle, P., Liang, K., and Zeger, S. (1994). Analysis of longitudinal data.
- Gartner, L. M., Morton, J., Lawrence, R. A., Naylor, A. J., O'Hare, D., Schanler, R. J., and Eidelman, A. I. (2005). Breastfeeding and the use of human milk. *Pediatrics*, 115(2):496–506.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.
- Molenberghs, G. and Verbeke, G. (2001). A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling*, 1(4):235–269.
- Pereira, O. C. N., da Costa Pereira, P. V., Sybuia, M. F., Barili, E., Santana, R., and Previdelli, I. (2017). A non-linear mixed-effects model to describe the effect of acarbose intake on postprandial glycaemia in a single rat. *Acta Scientiarum. Health Sciences*, 39(1):1.
- Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC Press.