

# Análise de sensibilidade de respostas de contagem longitudinais usando R com aplicação a dados médicos

*Alejandra Tapia*

**Palabras clave:** Respostas de contagem longitudinais, Distribuição de Poisson, Aproximação de integrais, Influência local, Métodos de Monte Carlo e Metropolis-Hastings

## Resumo

Respostas de contagem longitudinais ocorrem frequentemente em diversas áreas da ciência. Estas respostas são comumente descritas pelo modelo Poisson com efeitos mistos. Uma questão relevante que deve ser considerada em toda modelagem estatística, uma vez que a estimação de parâmetros é realizada, corresponde a análise de sensibilidade. Apesar da popularidade deste modelo, ferramentas de análise de sensibilidade para avaliar seu ajuste ainda são escassas. Assim, este trabalho apresenta uma metodologia de análise de sensibilidade das respostas de contagem longitudinais do modelo Poisson com efeitos mistos usando R. Finalmente, uma aplicação a dados médicos mostra a utilidade da proposta.

## Metodologia

Respostas de contagem longitudinais ocorrem frequentemente em diversas áreas, por exemplo, medicina, biologia, psicologia, sociologia, humanidades, economia, agricultura. Estas respostas são comumente analisadas através do modelo Poisson com efeitos mistos (MPEM). Este modelo acomoda a estrutura de correlação existente entre as respostas por meio de efeitos aleatórios (não observáveis). A estimação de parâmetros do modelo é conduzida usualmente pelo método de máxima verossimilhança (ML). Porém, não é uma tarefa computacional fácil, pois a incorporação dos efeitos aleatórios leva à função de verossimilhança dos dados observados a incluir integrais, que não possuem solução analítica e podem ser de grande dimensão. Por consequência, aproximações dessas integrais por aproximação de Laplace (AL) e adaptativa quadratura de Gauss Hermite (AQGH) são necessárias para a obtenção destas estimativas. Neste trabalho, as estimativas de ML são obtidas usando a função `glmer` do pacote `lme4` de R. Note que esta função está disponível com AQGH quando o modelo possui um intercepto aleatório e com AL para dois ou mais efeitos aleatórios.

Uma questão relevante que deve ser considerada em toda modelagem estatística, uma vez realizada a estimação de parâmetros, e a análise de sensibilidade. Esta permite investigar/avaliar como as diferentes fontes de incerteza do modelo podem afetar as estimativas de ML e, conseqüentemente, as inferências e tomadas de decisão. No obstante, este tipo de ferramentas para o MPEM ainda são escassas, principalmente para suas respostas de contagem, que são uma fonte de incerteza relevante para o modelo. Neste contexto, este trabalho apresenta uma metodologia de análise de sensibilidade usando a técnica de influência local sob uma estratégia de perturbação indireta para as respostas de contagem de um MPEM. Esta metodologia é resumida em um algoritmo é implementada no R; para mais detalhes vide Tapia et al. (2019).

Como a influência local é uma técnica baseada na função de verossimilhança, seu uso para o MPEM apresenta o mesmo problema das integrais mencionado acima. No obstante, essa dificuldade pode ser evitada, tratando os efeitos aleatórios como dados faltantes, definindo assim a função de verossimilhança dos dados completos (observados e faltantes). Desta maneira, a técnica de influência local é desenvolvida com base na função de verossimilhança dos dados completos, compreendendo o uso do método de Monte Carlo e algoritmo Metropolis-Hastings para a aproximação de matrizes fundamentais no cálculo das curvaturas individuais que permitirão avaliar a influência de uma observação. Na prática, esta metodologia se resume no seguinte algoritmo: (i) realizar uma análise exploratória dos dados usando os pacotes `base` e `ggplot` para a obtenção de estatísticas e visualizações, respectivamente; (ii) baseado em (i), formular um MPEM e obter as estimativas de ML usando a

função `glmer` do pacote `lme4`; (iii) baseado nas estimativas de ML, amostrar  $S = 10,000$  observações relacionada aos efeitos aleatórios através do algoritmo Metropolis-Hastings; (iv) com as observações amostradas em (iv), e depois de eliminar as primeiras  $M_0 = 1000$ , aproximar as matrizes fundamentais; (v) com as matrizes obtidas em (iv), calcular as curvaturas individuais usando o pacote `matrixcalc` e um ponto de corte adequado para identificar as observações influentes; (vi) realizar uma gráfica das curvaturas individuais versus o índice das observações, traçando o ponto de corte, para visualizar as observações identificadas como influentes; (vii) avaliar a magnitude do impacto exercido nas estimativas de ML pelas observações influentes ou um conjunto delas, calculando o erro percentual (PE) para todas as combinações de observações influentes e conjuntos de observações influentes (quando pertencem ao mesmo indivíduo).

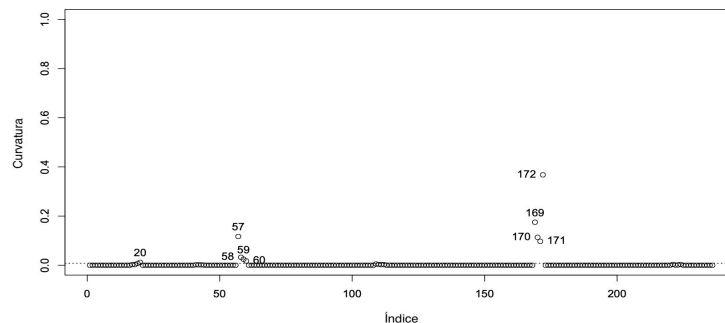
Por último, se deseja desenvolver um pacote de análise de sensibilidade para modelos mistos com respostas discretas que contenha tanto esta metodologia quanto outras.

## Aplicação

Um conjunto de dados de um ensaio clínico com 59 pacientes que sofrem de epilepsia foi apresentado por Thall e Vail (1990). O objetivo deste estudo foi analisar se uma nova droga reduz as crises epiléticas ou não. A resposta de interés é o número de crises epiléticas vivenciadas por cada paciente durante um período de duas semanas a cada quatro visitas à clínica. Depois de uma análise exploratoria dos dados, considera-se um modelo Poisson com efeitos mistos definido por

$$\log(\mu_{ij}) = \beta_0 + \beta_1 \text{Base}_{ij} + \beta_2 \text{Trt}_{ij} + \beta_3 \text{Base}_{ij} \text{Trt}_{ij} + \beta_4 \text{Age}_{ij} + \beta_5 \text{Visit}_{ij} + b_{i_0} + b_{i_1} \text{Visit}_{ij}, \quad (1)$$

para  $j = 1, \dots, 4$ ,  $i = 1, \dots, 59$ , e  $q = 236$ , em que “Base” é o número de crises de um pre-ensaio clínico; “Trt” é um indicador de tratamento o placebo; “Base Trt” é a interação entre “Base” e “Trt”; “Age” é a idade em anos; e “Visit” um indicador dos periodos. A Figura 1 mostra os pacientes indentificados como influentes. A eliminação de combinações de pacientes leva a variações (PE) importantes nas estimativas de  $\beta_0$ ,  $\beta_3$ ,  $\beta_4$  e  $\beta_5$ , sendo estas as mais sensíveis à estratégia de perturbação indireta adotada. No entanto, com um nível de significância do 5 % não foram detectadas alterações inferenciais. En conclusão, esta metodologia permite obter informações valiosas sobre estimativas de ML e respostas de contagem que precisam de escrutínio adicional neste modelo, permitindo obter uma melhor perspectiva sobre as consequências dos dados quando observações influentes são indentificadas.



**Figura 1:** Pacientes indentificados como influentes para os dados de epilepsia.

Tapia, A., Giampaoli, V., Diaz, M.P. e Leiva, V. 2019. "Sensitivity analysis of longitudinal count responses: a local influence approach and application to medical data". *Journal of Applied Statistics*, 46 (6): 1021–1042. <https://doi.org/10.1080/02664763.2018.1531978>

Alejandra Tapia  
Universidad Católica del Maule, Chile  
[alejandraandreatapiasilva@gmail.com](mailto:alejandraandreatapiasilva@gmail.com)