

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Interpolação por splines aplicada a dados longitudinais

Breno Gabriel da Silva ¹Yana Miranda Borges ²Naiara Caroline Aparecido dos Santos³

Programa de Pós-graduação em Bioestatística, Universidade Estadual de Maringá - UEM.

Vanderly Janeiro ⁴Terezinha Aparecida Guedes ⁵

Professor do Departamento de Estatística, Universidade Estadual de Maringá - UEM e orientador no Programa de Pós-graduação em Bioestatística.

1 Introdução

Com o objetivo de investigar a possível relação entre uma ou mais variáveis, os modelos de regressão são uma das técnicas que podem ser empregadas. Esses modelos são ditos lineares quando a relação da variável resposta (variável dependente) às variáveis regressoras (variáveis independentes) é uma função linear de alguns parâmetros, caso contrário diz-se que o modelo é não-linear.

De acordo com [2], a regressão por *splines* também chamada *regressão segmentada* vem crescendo nas últimas décadas, sendo esta uma técnica de modelagem não linear bastante flexível para identificar a forma funcional da variável regressora com a resposta. Advoga [3], que *splines* são funções polinomiais segmentadas, unidas por pontos de corte denominados nós. A vantagem de se utilizar *splines* é a facilidade de convergência e manipulação, sendo úteis quando um polinômio de grau baixo não se ajusta aos dados e quando se trata de dados biológicos cuja natureza é oscilatória.

Neste trabalho, objetiva-se utilizar a *regressão segmentada* com o intuito de identificar a forma funcional da relação das variáveis regressoras (Tempo em semanas e Grupos de Tratamento) com a resposta (Peso), sendo esta metodologia aplicada a um conjunto de dados referentes a 57 camundongos suíços, infectados por *Trypanosoma cruzi* (agente etiológico da doença de chagas), machos com 56 dias de idade, provenientes do Biotério central da Universidade Estadual de Maringá - UEM. Para as análises estatísticas, utilizou-se o software R versão 3.5.0.

¹omatematico.breno@gmail.com²borges.yana@gmail.com³naicaroline2@gmail.com⁴vjaneiro@uem.br⁵taguedes@uem.br

2 Desenvolvimento e resultados preliminares

De acordo com [1], uma *spline* é definida como uma função de *regressão segmentada*, contínua por partes, em que é imposta uma restrição de que os pontos de cada segmento se unam nos pontos de corte denominados (nós). Assim, sejam n o grau da função *spline*, c a quantidade de nós, x_o o nó e $(x - x_o) = \begin{cases} (x - x_o), & x > x_o \\ 0, & x \leq x_o \end{cases}$, tem-se o modelo *spline*:

$$y = \sum_{j=1}^n \beta_{0j} x^j + \sum_{i=1}^c \beta_i (x - x_o)^n \quad (1)$$

No qual, para utilizá-lo, é necessário a presença do fenômeno oscilatório, que é averiguado por meio do modelo de efeitos. Uma vez ajustado tal modelo, detectou-se a presença de alterações no ganho de peso corporal após a 2^o semana, como segue a figura 1:

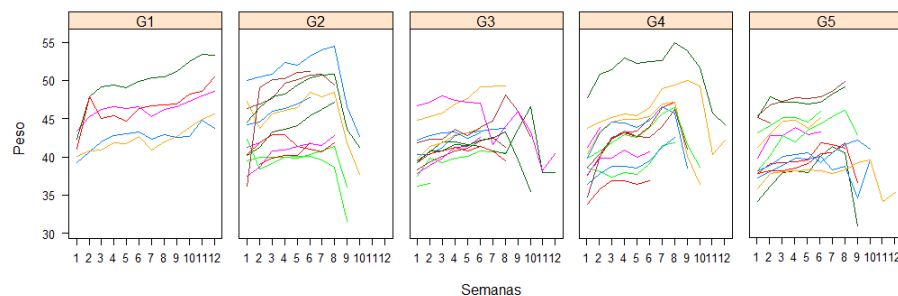


Figura 1: Evolução do peso no tempo

3 Conclusões

Portanto, sendo evidenciada à presença de alterações no ganho de peso, utilizou-se a regressão por *splines* para evidenciar o melhor tratamento que propicia o tempo de sobrevivência para os camundongos, pois com esta ferramenta é possível modelar fenômenos cuja natureza é oscilatória. Pretende-se ainda neste trabalho avaliar as suposições do modelo e estimar os parâmetros por meio do método dos mínimos quadrados, cujas soluções serão obtidas pelo processo iterativo de Gauss-Newton.

Referências

- [1] P. H. C. Eilers and B. D. Marx. Flexible Smoothing with B-Splines and Penalties. *Stat. Sci.* 11: 89-121, 1996.
- [2] S. Greenland. Dose-Response and Trend Analysis in Epidemiology: Alternatives to Categorical Analysis. *Epidemiology*, 6, 4, 356-365, 1995.
- [3] L.J. Keele. Semiparametric regression statistic for the science. John Wiley and Sons, 1^o edição, 2008.