



O método de Bland e Altman para análise da qualidade do ajuste de um modelo de regressão não linear

João Pedro Serenini¹, Isolde Terezinha Santos Previdelli² e Aline Edlaine de Medeiros³

¹Programa de Pós-Graduação em Bioestatística PBE-UEM

²Programa de Pós-Graduação em Bioestatística PBE-UEM

³Programa de Pós-Graduação em Bioestatística PBE-UEM

RESUMO

Este trabalho visa empregar o Método de Bland e Altman para a validação de modelo de regressão não linear. Neste sentido, descreveu-se um modelo de regressão não linear descrito da literatura, para modelar dados da cinética do oxigênio durante um teste ergométrico, e aplicou-se o método de Bland e Altman para validar tal modelo. Além disso, realizou-se uma breve revisão da literatura fornecendo aos mestrandos uma familiarização com esta metodologia.

Palavras chave: Regressão Não Linear, Bland e Altman, Ajuste do Modelo.

1 INTRODUÇÃO

Em modelagem estatística um dos procedimentos essenciais é a validação do modelo, principalmente em modelos de regressão não linear, que empregam técnicas articulares no processo de ajuste do modelo, e que dependem de certas hipóteses laboratoriais observadas pelo pesquisador. Assim, neste trabalho descreve-se o emprego adequado do método de Bland e Altman (1986)[3] para analisar a concordância dos dados observados com os valores preditos de um modelo de regressão não linear.

Mais especificamente, o método de Bland e Altman (1986)[3] foi utilizado para validar o modelo proposto por Baty et.al.(2015) [2] em um estudo do comportamento do oxigênio durante o teste ergométrico (6MWT), onde apresenta-se uma análise gráfica e das estatística envolvidas neste método.

2 METODOLOGIA

No estudo de modelagem estatística busca-se empregar modelos parcimoniosos, e de simples manipulação, por exemplo, o modelo de regressão linear (MRL). Mas em algumas áreas, como, a Biologia, a Química e Engenharias, frequentemente surgem problemas que não podem ser explicados pelo MRL, e se faz necessário o emprego de técnicas que valorizem o conhecimento prévio do pesquisador, e satisfaça as inclinações científicas da teoria em questão, neste contexto, é muito oportuno o emprego de modelos de regressão não linear.

Baty et. al.(2015)[2] desenvolveram um estudo sobre a cinética do oxigênio durante o teste ergométrico (6MWT), que destaca-se ao descrever as várias etapas do processo de modelagem em regressão não linear no pacote *nlstools*[1], do software estatístico *R*[4]. Ao analisar os gráficos dos resíduos e resíduos padronizados, além das técnicas já empregadas pelos autores, sugere-se observar a qualidade do ajuste do modelo pelo método de Bland e Altman.

Em 1983 Bland e Altman [3] publicaram um artigo baseado em Eksborg (1981) [5] que apresentava uma metodologia de análise para comparar as medidas obtidas por medidores de fluxo máximo, tal metodologia ficou bastante conhecida no meio estatística como método de Bland e Altman. A princípio a proposta para a utilização de tal método era baseada na dificuldade que se encontrava em identificar, por exemplo, se novos aparelhos, como o de pressão arterial, realmente estavam apresentando resultados reais. Porém com o passar do tempo tal método passou a ser utilizado em outras tarefas, como na utilização de validação de modelos.

Com a finalidade de se testar a concordância entre duas medias a correta aproximação estatística não é óbvia, vários estudos trazem que o valor de correlação entre duas medidas é um indicador de concordância, porém hoje se sabe que uma alta correlação não necessariamente implica em uma alta concordância, pois o coeficiente de correlação mede a força de relação entre duas variáveis, não a concordância entre elas, além disso uma mudança de escala em uma das variáveis não altera a correlação, porém pode afetar de forma significativa a concordância entre elas.

O método gráfico de Bland e Altman tem como objetivo analisar a concordância entre duas medidas quantitativas através da diferença entre elas e da construção de limites de concordância. Através deste gráfico podemos visualizar o viés, ou seja, o quanto as diferenças entre as duas medidas se afastam do valor 0, ou seja, a medida do erro, além de outliers e certas tendências das variáveis.

Por ser um método gráfico o sistema de gráficos de Bland e Altman não nos diz se a concordância é suficiente ou adequada para se usar um método ou outro. Ele simplesmente quantifica o viés e a amplitude da concordância, dentro da qual 95% das diferenças entre uma medição e a outra estão incluídas. Uma vez que permite avaliar a real diferença entre os valores ajustados e os observados, o método informa o quanto o modelo proposto é bom em prever as medidas, sendo de grande importância na validação de modelos.

3 APLICAÇÕES

Baty et. al.(2015)[2] estudando o comportamento da cinética do oxigênio aplicaram um modelo com quatro parâmetros, dado por

$$VO_2(t) = \begin{cases} VO_{2rest} & \text{se } t \leq \lambda, \\ VO_{2rest} + (VO_{2peak} - VO_{2rest})(1 - e^{-(t-\lambda)/\mu}) & \text{se } t > \lambda \end{cases} \quad (1)$$

em que, VO_{2rest} , VO_{2peak} , μ e λ representam, respectivamente, o nível de oxigênio restante após a etapa de repouso, o oxigênio máximo absorvido durante o teste, a taxa de variação, que caracteriza a inclinação da curva a medida em que o tempo t aumenta, com $\mu > 0$, e λ representa o período de repouso controlado pelos pesquisadores.

Neste experimento, adotou-se $\lambda = 5,883$, e utilizando a função $nls()$ os autores obtiveram as estimativas $VO_{2rest} = 356,759$, $VO_{2peak} = 1630,884$ e $\mu = 1,186$, gerando o modelo descrito em 2.

$$VO_2(t) = \begin{cases} 356,759 & \text{se } t \leq 5,883, \\ 356,759 + (1630,884 - 356,759)(1 - e^{-(t-5,883)/1,186}) & \text{se } t > 5,883 \end{cases} \quad (2)$$

Para avaliar a qualidade deste ajuste empregou-se o método de Bland e Altman que forneceu as seguintes constatações

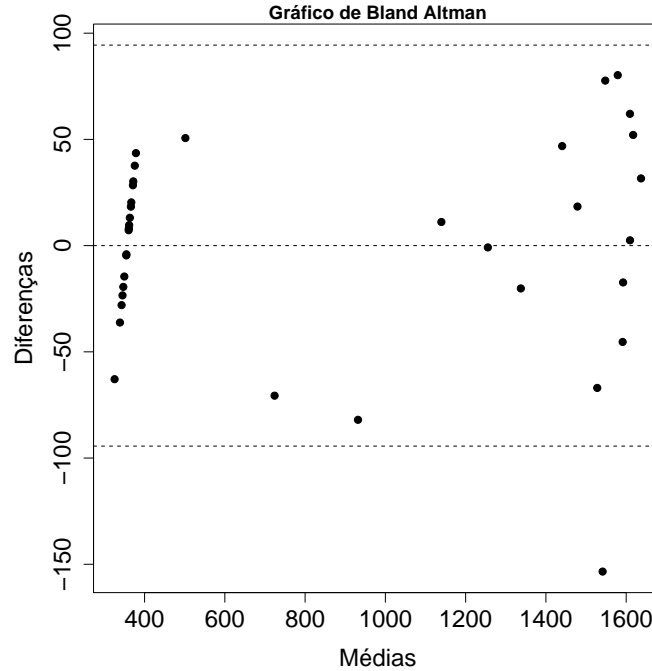


Figura 1: Diferenças *vs* médias dos dados observados e valores ajustados pelo modelo para o comportamento da cinética do oxigênio

Para esses dados a média das diferenças é $\mu_d = 1,298892e^{-07}$ com desvio-padrão igual a $s = 48,15$. As duas linhas que aparecem no gráfico são conhecidos como

limites de concordância que são os pontos dados por $\mu_d - 1,96s$ e $\mu_d + 1,96s$, a linha central é exatamente μ_d . Para que haja uma boa concordância entre as medidas esperamos que a linha central esteja o mais próximo possível do valor 0, e que pelo menos 95% das observações estejam contidas entre os limites de concordância, o que ocorre em nesse caso, indicando portanto um bom ajuste do modelo.

Referências

- [1] BATY, F., RITZ, C., CHARLES, S., BRUTSCHE, M., FLANDROIS, J.-P., AND DELIGNETTE-MULLER, M.-L. A toolbox for nonlinear regression in R: The package nlstools. *Journal of Statistical Software* 66, 5 (2015), 1–21.
- [2] BATY, F., RITZ, C., CHARLES, S., BRUTSCHE, M., FLANDROIS, J.-P., DELIGNETTE-MULLER, M.-L., ET AL. A toolbox for nonlinear regression in r: the package nlstools. *J Stat Softw* 66 (2015), 1–21.
- [3] BLAND, J. M., AND ALTMAN, D. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet* 327, 8476 (1986), 307–310.
- [4] R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.
- [5] S, E. Evaluation of method-comparison data. *Clin Chem* (1981), 27–1311.