

Desenvolvimento de ferramenta em linguagem R para avaliação de linearidade na validação de métodos analíticos

Development of a tool in R language to evaluate the linearity in the validation of analytical methods

Jonathan Silva^{1,2}, **Josiane Mônego**¹, **Raffael Nagel**¹, **Fernando Henrique Rosa**¹, **Nadir Hermes**¹

¹Research Product Centre Americas - Souza Cruz S/A; ²Universidade Federal do Rio Grande do Sul - UFRGS

E-mail: jonathan.silva@souzacruz.com.br

Resumo: A validação de métodos é uma etapa crucial para confirmar que o método analítico tem capacidade de desempenho consistente com sua aplicação. A linearidade é um dos parâmetros avaliados neste processo. No presente trabalho serão apresentados os principais conceitos de linearidade revisados na literatura, assim como sua implementação em uma ferramenta em linguagem R, a qual viabiliza a aplicação destes conceitos através de utilização de testes estatísticos robustos.

Palavras-chave: linearidade, validação de métodos, projeto R.

Abstract: Methods validation is an essential step to confirm that an analytical method has a performance capacity in line with its application. Linearity is one of the parameters observed in this process. In the current work we present some of linearity concepts we reviewed in the literature, as well as show an implementation in the statistical package R, which enables the user to evaluate linearity using the most robust statistical concepts.

Keywords: linearity, validation of method, R project.

1. INTRODUÇÃO

A confiabilidade dos resultados reportados pelos laboratórios, bem como a validade dos ensaios, é garantida através da validação dos métodos analíticos.

Conforme definido na norma ABNT NBR ISO/IEC 17025:2005, a validação é a confirmação por exame e fornecimento de

evidência objetiva de que os requisitos específicos para um determinado uso pretendido são atendidos.

Os parâmetros analíticos normalmente avaliados na validação do método são: seletividade, linearidade, precisão, exatidão, limite de detecção, limite de quantificação e robustez.

O parâmetro linearidade é frequentemente analisado através de avaliação visual e coeficiente de correlação r , considerando $r > 0,999$ um ajuste ideal dos dados à linha de regressão.

Para suportar a avaliação dos resultados obtidos na validação de métodos com maior embasamento estatístico e disponibilizando novas técnicas de abordagem, foi desenvolvida uma ferramenta de diagnóstico de linearidade utilizando a linguagem de programação R. Esta linguagem foi escolhida devido à flexibilidade e grande gama de métodos estatísticos disponíveis.

Este artigo tem como objetivo apresentar a ferramenta desenvolvida para avaliação completa do parâmetro de linearidade proveniente das validações realizadas na Área Analítica da Souza Cruz.

1. VALIDAÇÃO DE MÉTODOS

A realização da validação de métodos garante, por meios de estudos experimentais, que o método em questão atende às exigências das aplicações analíticas, assegurando a confiabilidade dos resultados gerados.

1.1. Linearidade

A análise de regressão é uma modelagem matemática que busca verificar a relação entre duas ou mais variáveis e avaliar como determinadas variáveis estão influenciando as demais.

O modelo de Regressão Linear Simples é caracterizado quando há o interesse na relação de apenas uma variável de entrada (ou de predição) com uma variável resposta.

Através da coleta de dados e do uso de métodos estatísticos é possível estabelecer suposições sobre o modelo, prever valores para a variável resposta e avaliar seu comportamento.

Linearidade é o parâmetro que trata da função de regressão obtida para os dados:

$$Y_i = \beta_0 + \beta_1 X_i \quad (1)$$

onde: Y_i : variável resposta, β_0 : intercepto, β_1 : slope, X_i : variável de predição.

Assim, a linearidade corresponde à capacidade do método em demonstrar que os resultados são diretamente proporcionais à concentração da substância em análise, dentro de um intervalo especificado.

Diferentes análises estatísticas devem ser realizadas nos resultados experimentais obtidos para avaliação da linearidade, tais como: equação que define a curva de resposta (linear, quadrática, etc), avaliação do comportamento da variância (homocedasticidade e heterocedasticidade), análise de resíduos, ausência de *lack of fit*.

2. FERRAMENTA DE DIAGNÓSTICO DE LINEARIDADE

Como a linguagem R é predominantemente textual, foi desenvolvida uma interface gráfica (GUI – *Graphical User Interface*), no qual o poder de processamento da linguagem é acessado através desta interface, tornando os cálculos e controles estatísticos acessíveis a todo perfil de usuários. Além disso, também foi desenvolvido um pacote para instalação da ferramenta.

Na figura 1 tem-se a interface do software desenvolvido. Nesta interface escolhe-se o arquivo de entrada (em formato Excel), as colunas com as variáveis, pontos a serem removidos da análise e as concentrações consideradas no estudo. Também pode-se escolher o tipo de modelo de regressão a ser ajustado. Na figura 2 mostra-se um dos gráficos obtidos no relatório de linearidade.



Figura 1. Interface da ferramenta de diagnóstico de linearidade.

A saída do programa é obtida em HTML, a qual contém o relatório de diagnósticos de linearidade. Este relatório apresentará todas as informações necessárias para crítica do modelo, testes e gráficos.

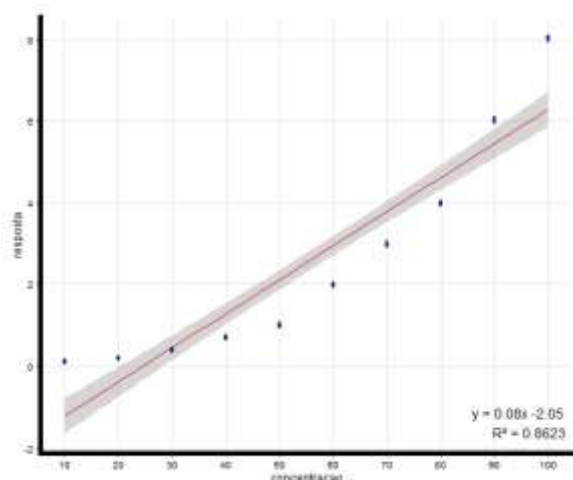


Figura 2. Gráfico de dispersão obtido pela ferramenta de diagnóstico de linearidade.

2.1. Análises estatísticas

Uma série de análises estatísticas é realizada a partir dos dados de entrada, incluindo modelos de regressão de diversas formas (linear, não linear, 8º Congresso Brasileiro de Metrologia, Bento Gonçalves/RS, 2015

linear com pesos, não linear com pesos), teste de ajuste e falta de ajuste (*Lack of Fit*), testes de homocedasticidade, testes de *outliers*, resíduos, entre outros.

Além disso, uma série de representações gráficas também está disponível, incluindo gráfico de dispersão, gráfico dos resíduos vs concentração, gráfico do módulo dos resíduos vs concentração, gráfico dos resíduos em sequência, *boxplot* dos resíduos, gráfico de probabilidade normal, *Cook's Distance* e resíduos vs influência.

Assim, a tomada de decisão é baseada em todos os testes realizados, e não somente através de um único resultado.

3. EXEMPLO DE USO

O uso da ferramenta será exemplificado, seguido da análise dos dados apresentados na figura 2.

Pode ser observado que a reta ajustada, incluindo a banda de confiança em cinza, não parece ser adequada aos dados. Assim, o gráfico de resíduos vs concentração (figura 3) deve ser avaliado para auxiliar na interpretação.

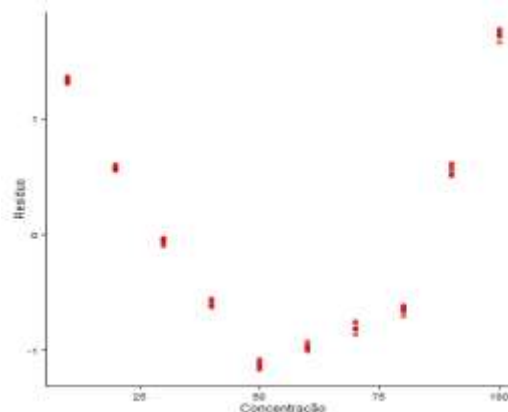


Figura 3. Gráfico de resíduos vs concentração obtido pela ferramenta de diagnóstico de linearidade.

Nota-se que este gráfico confirma que o ajuste linear não parece ser adequado, sendo esta percepção confirmada baseada nas demais saídas da ferramenta, como ilustrado na figura 4, onde

observa-se que o P-valor do teste de Falta de Ajuste é de 0%.

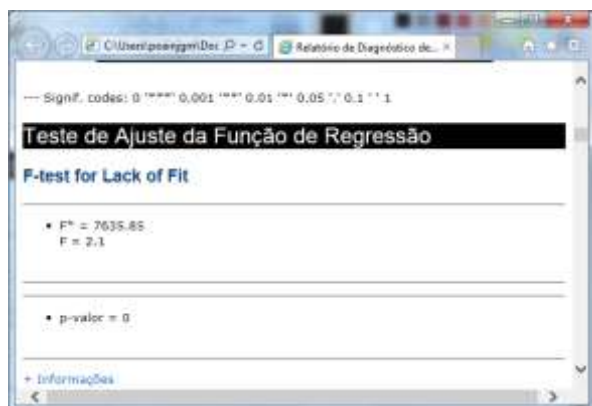


Figura 4. Saída parcial do relatório em HTML, mostrando o teste de ajuste de *Lack of Fit* (F-Test).

Neste caso, foi necessário retornar para a tela inicial da ferramenta, e escolher um ajuste de modelo quadrático (figuras 5 e 6).

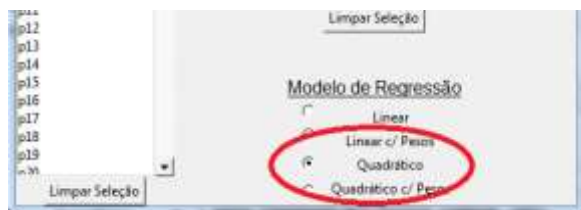


Figura 5. Escolha do modelo quadrático.

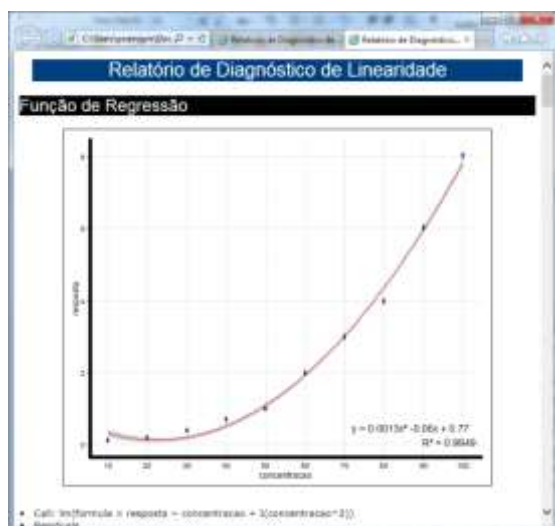


Figura 6. Saída da função de regressão com o modelo quadrático ajustado.

O modelo quadrático parece melhor ajustado conforme observado no gráfico da figura 6. Entretanto, é necessário avaliar, em conjunto, as demais saídas do relatório, para a tomada de decisão.

A ideia é que o processo de avaliação de linearidade seja interativo e iterativo, permitindo que o especialista do laboratório faça vários modelos preliminares e avalie os resultados baseado nas saídas da ferramenta.

Por fim, no final do relatório, há uma explicação detalhada de cada conceito estatístico utilizado e como interpretá-los corretamente (figura 7).

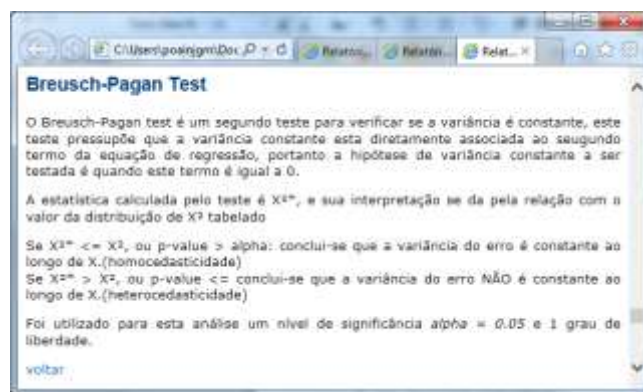


Figura 7. Explicação do teste de Breusch-Pagan apresentada no final do relatório obtido pela ferramenta de diagnóstico de linearidade.

4. CONCLUSÕES

Um dos principais benefícios do desenvolvimento da ferramenta foi possibilitar a aplicação de novos recursos para a avaliação do parâmetro linearidade, utilizando técnicas estatísticas mais completas e robustas.

Além disso, o desenvolvimento da ferramenta baseada na aplicação R, possibilitou o emprego de cálculos e funções estatísticas sofisticadas não disponíveis em outros programas tradicionalmente utilizados pelo laboratório, como o Excel.

Como resultado final, os nossos laboratórios passam a ter uma ferramenta muito mais completa e robusta para avaliação de linearidade, viabilizando a entrega de resultados analíticos mais confiáveis.

5. REFERÊNCIAS

THOMPSON, Michael; STEPHEN, Ellison R. and WOOD, Roger. Harmonized Guidelines for Single-Laboratory Validation of Methods of Analysis. Pure Appl. Chem., Vol. 74, No. 5, 2002.

MILLER, James N. and MILLER, Jane C. Statistics and Chemometrics for Analytical Chemistry. 6th ed., Pearson, 2010.

FUNK, Verner; DAMMANN, Vera and DONNEVERT, Gerhild. Quality Assurance in Analytical Chemistry. 2nd ed., WILEY-VCH Verlag GmbH & Co. KGaA, 2007.

JCGM 200:2008 - International Vocabulary of Metrology — Basic and general concepts and associated terms (VIM).

KUTNER, Michael H. *et al.* Applied Linear Statistical Models. 5th ed., Mc Graw Hill Irwin series Operations and decision sciences, 2005.

RIBAMI, Marcelo *et al.* 2004 Validação em métodos cromatográficos e eletroforéticos *Química Nova* **27** 771-780.

LAWRENCE, Michael F. and VERZANI, John. Programing Graphical User Interfaces in R - The R Series - CRC Press

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>