Review of guidance papers on regression modeling in statistical series of medical journals

Christine Wallisch[1,2]*, Paul Bach[1,3], Lorena Hafermann[1], Nadja Klein[3], Willi Sauerbrei[4], Ewout W. Steyerberg[5], Georg Heinze[2], Geraldine Rauch[1]*, on behalf of topic group 2 of the STRATOS initiative[¶]

PLOS ONE; https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0262918

Abstract

[…] published series of statistical tutorials and (shorter) papers mainly addressing medical researchers. The aim of this review was to assess the current level of knowledge with regard to regression modeling contained in such statistical papers.
We identified 23 series including 57 topic-relevant articles. Within each article, two independent raters analyzed the content by investigating 44 predefined aspects on regression modeling.

Software

Detailed code examples were provided in two articles only [16, 58]. In the article of Curran- Everett [58], the R script file was provided as appendix and in the article of Obuchowski [16], code chunks were included throughout the text directly showing how to derive the reported results.

Series and articles recommended to read

Depending on the aim of the planned study, as well as the focus and knowledge level of the reader, different series and articles might be recommended. The series in *Circulation* comprised three papers about multiple linear and logistic regression [24–26], which provide basics and describe many essential aspects of univariable and multivariable regression modeling. For more advanced researchers, we recommend the article of *Nuñez* et al. in *Revista Española de Cardiologia* [22], which gives a quick overview of aspects and existing methods including functional forms and variable selection. The *Nature Methods* series published short articles focusing on few, specific aspects of regression modeling [34–42]. This series might be of interest if one likes to spent more time on learning about regression modeling. If someone is especially interested in prediction models, we recommend a concise publication in the *European Heart Journal* [31], which provides details on model development and validation for predictive purposes. For the same topic we can also recommend the paper by Grant et al. [21]. We consider all series and articles recommended in this paragraph as suitable reading for medical research- ers but this does not imply that we agree to all explanations, statements and aspects discussed.

16. Obuchowski NA. Multivariate statistical methods. Am J Roentgenol. 2005; 185(2):299–309. PMID: 16037496

21. Grant SW, Collins GS, Nashef SAM. Statistical Primer: developing and validating a risk prediction model. Eur J Cardio-Thorac. 2018;54(2):203–8. pmid:29741602

22. Nuñez E, Steyerberg EW, Nuñez J. Regression modeling strategies. Rev Esp Cardiol. 2011;64(6):501–7. pmid:21531065

31. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925. pmid:24898551

24. Slinker BK, Glantz SA. Multiple linear regression—Accounting for multiple simultaneous determinants of a continuous dependent variable. Circulation. 2008;117(13):1732–7. pmid:18378626

25. LaValley MP. Logistic regression. Circulation. 2008;117(18):2395–9. pmid:18458181

26. Crawford SL. Correlation and regression. Circulation. 2006;114(19):2083–8. pmid:17088476

31. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925. pmid:24898551

34. Altman N, Krzywinski M. Simple linear regression. Nat Methods. 2015;12(11):999–1000. pmid:26824102

35. Altman N, Krzywinski M. Association, correlation and causation. Nat Methods. 2015;12(10):899–900. pmid:26688882

36. Altman N, Krzywinski M. Analyzing outliers: influential or nuisance? Nat Methods. 2016;13(4):281–2. pmid:27482566

37. Altman N, Krzywinski M. Regression diagnostics. Nature Methods. 2016;13(5):385–6. https://www.nature.com/articles/nmeth.3854

38. Krzywinski M, Altman N. Multiple linear regression. Nat Methods. 2015;12(12):1103–4. pmid:26962577

39. Krzywinski M, Altman N. Classification and regression trees. Nature Methods. 2017;14(8):755–6. https://www.nature.com/articles/nmeth.4370

40. Lever J, Krzywinski M, Altman N. Regularization. Nature Methods. 2016;13(10):803–4. https://www.nature.com/articles/nmeth.4014

41. Lever J, Krzywinski M, Altman N. Model selection and overfitting. Nature Methods. 2016;13(9):703–4. https://www.nature.com/articles/nmeth.3968

42. Lever J, Krzywinski M, Altman N. Logistic regression. Nature Methods. 2016;13(7):541–2. https://www.nature.com/articles/nmeth.3904

58. Curran-Everett D. Explorations in statistics: regression. Adv Physiol Educ. 2011; 35(4):347–52. PMID: 22139769