BOOK REVIEW

Multivariable analysis: a practical guide for clinicians—second edition

Mitchell H Katz
Cambridge University Press, The Edinburgh Building, Shaftesbury Road, Cambridge CB2 2RU, United Kingdom
£27.99 (US$45.00), pp 203, ISBN 0 521 54985 X (paperback), 0 521 84051 1 (hardback)

The term multivariable analysis is often mistaken for multivariate analysis in medical and health sciences research. Multivariate analysis refers to the analysis of multiple outcomes whereas multivariable analysis deals with only one outcome each time. That is, whether a regression analysis is multivariate depends on the number of outcomes, not the number of independent variables. Katz, as is obvious from the book title, focuses on multivariable, not multivariate, analysis.

Chapters 1 to 3 briefly distinguish between the possible different effects exerted by independent variables and introduce the different multivariable analyses arising from using different measurement scales for the outcome. Chapter 4 is a very short chapter, of only three pages, describing mainly ordinal and nominal independent variables that are handled as dummy variables in a multivariable analysis. This may not be of practical use to clinicians with experience of conducting a regression analysis. Instead, a discussion of the importance of assessing the sample size at each level of a categorical independent variable would be good, especially the effect of sample size on interpreting a surprisingly large effect which may just be due to the lack of a representative sample.

It is not uncommon to see analyses performed by a non-statistician that are not checked for adequacy despite the continual sensitivity of most journal statistical reviewers to the validity of the statistical method used. Knowing the assumptions behind an analysis is important when determining what needs to be checked to justify the method used. Katz focused on this issue in Chapter 5 (the longest chapter in the book), detailing the assumptions behind each multivariable analysis. Reading this chapter increases one's motivation to check for model adequacy. However, we are sceptical about the suggested automatic assumption of normality when the sample size is 100 or larger in an ordinary linear regression. Katz mentioned that this is derived from the Central Limit Theorem, but this theorem only tells us that the sample mean distributes approximately as normal when the sample size is sufficiently large. However in an ordinary linear regression, we assume normality on the residuals or the difference between observed and predicted outcome values. It is not clear how the theorem is applied. Nevertheless, it is not difficult to find examples with sample sizes greater than 100 that do not fulfill the normality assumption. Therefore, it remains desirable to check for normality, regardless of the sample size.

Chapter 7 is a useful chapter that discusses what and how independent variables should be considered in a multivariable analysis. When examining the effect of a risk factor, different adjustments of other factors may yield different or even confusing conclusions. The decision on what to adjust is often guided by an a priori theoretical or biological relationship among the different factors and the outcome. Having said that, establishing a complete relational structure may not be straightforward (indeed it is difficult most of the time). This may be due to different opinions in the literature, an attempt to consider an overwhelming number of factors at the same time, or from leaving the analysis to a statistician who has insufficient communication with the clinical investigator. Nevertheless, there should be at least an attempt to identify the intervening variables (those caused by the risk factor of interest but not vice versa) that, if mistakenly adjusted, may mask the actual association between the risk factor of interest and the outcome. Katz stressed the importance of this right at the start of the chapter. The rest of the chapter discusses the possible solutions when some independent variables are strongly associated and when there are missing values. However, this requires a more sophisticated knowledge of statistics before these can be performed in practice.

Katz then discusses how the different multivariable analyses are interpreted and checked for adequacy and validity, followed by descriptions of some more advanced topics including generalised estimating equations, the mixed effects model, conditional logistic regression, the counting processes approach and the classification and regression trees (CART) analysis for different situations and research purposes. Practical comparisons are made between different methods used for the same problem.

Overall, the book is easy and worth reading for clinicians. It is written in a conversational tone with informative examples, making it a lively and engaging book to read. The book teaches concepts and practical considerations rather than gives technical details. A knowledge of basic statistics, particularly regression analysis, will make reading it even more enjoyable.

DYT Fong, CF Lee
Department of Nursing Studies
Li Ka Shing Faculty of Medicine
The University of Hong Kong, Hong Kong