

Ivan Jacob Agaloos Pesigan

August 13, 2023

References

Eddelbuettel: Seamless R and C++ integration with Rcpp

Eddelbuettel-2013

Dirk Eddelbuettel. *Seamless R and C++ integration with Rcpp*. Springer New York, 2013. ISBN: 978-1-4614-6868-4. DOI: [10.1007/978-1-4614-6868-4](https://doi.org/10.1007/978-1-4614-6868-4).

Abstract: Illustrates a range of statistical computations in R using the Rcpp package. Provides a general introduction to extending R with C++ code. Features an appendix for R users new to the C++ programming language Rcpp packages are presented in the context of useful application case studies.

Koopman et al.: Pulling the Sobel test up by its bootstraps

Koopman-Howe-Hollenbeck-2014

Joel Koopman, Michael Howe, and John R. Hollenbeck. “Pulling the Sobel test up by its bootstraps”. In: Charles E. Lance and Robert J. Vandenberg. *More statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences*. Routledge/Taylor & Francis Group, 2014, pp. 224–243. ISBN: 9780203775851. DOI: [10.4324/9780203775851](https://doi.org/10.4324/9780203775851).

Abstract: In the domain of building and testing theory, mediation relationships are among the most important that can be proposed. Mediation helps to explicate our theoretical models (Leavitt, Mitchell, & Peterson, 2010) and addresses the fundamental question of why two constructs are related (Whetten, 1989). One of the better-known methods for testing mediation is commonly referred to as the “Sobel test,” named for the researcher who derived a standard error (Sobel, 1982) to test the significance of the indirect effect. Recently, a number of different research teams (e.g., Preacher & Hayes, 2004; Shrout & Bolger, 2002) have criticized the Sobel test because this standard error

requires an assumption of normality for the indirect effect sampling distribution. This distribution tends to be positively skewed (i.e., not normal), particularly in small samples, and so this assumption can be problematic (Preacher & Hayes, 2004; Stone & Sobel, 1990). As a result, the statistical power of the Sobel test may be lessened in these contexts (Preacher & Hayes 2004; Shrout & Bolger, 2002). In light of this concern, some scholars have advocated instead for the use of bootstrapping to test the significance of the indirect effect (e.g., Shrout & Bolger 2002). Bootstrapping requires no a priori assumption about the shape of the sampling distribution because this distribution is empirically estimated using a resampling procedure (Efron & Tibshirani, 1993). As a result, departures from normality are less troublesome when creating a confidence interval for the indirect effect. For this reason, bootstrapping is now widely believed to be inherently superior to the Sobel test when testing the significance of the indirect effect in organizational research. Our position is that this belief constitutes an urban legend. As with all statistical urban legends, there is an underlying kernel of truth to the belief that bootstrapping is superior to the Sobel test. However, as we discuss in this chapter, there are several reasons to be concerned with a broad belief in the superiority of bootstrapping. We begin with a brief overview of mediation testing focusing on the Sobel test and bootstrapping and then explain the underlying kernel of truth that has propelled bootstrapping to the forefront of mediation testing in organizational research. Subsequently, we discuss four areas of concern that cast doubt on the belief of the inherent superiority of bootstrapping. Finally, we conclude with recommendations concerning the future of mediation testing in organizational research.

Little et al.: Statistical analysis with missing data

Little-Rubin-2019

Roderick J. A. Little and Donald B. Rubin. *Statistical analysis with missing data*. 3rd ed. Wiley, Apr. 2019. ISBN: 9781119482260. DOI: [10.1002/9781119482260](https://doi.org/10.1002/9781119482260).

Abstract: An up-to-date, comprehensive treatment of a classic text on missing data in statistics. The topic of missing data has gained considerable attention in recent decades. This new edition by two acknowledged experts on the subject offers an up-to-date account of practical methodology

for handling missing data problems. Blending theory and application, authors Roderick Little and Donald Rubin review historical approaches to the subject and describe simple methods for multivariate analysis with missing values. They then provide a coherent theory for analysis of problems based on likelihoods derived from statistical models for the data and the missing data mechanism, and then they apply the theory to a wide range of important missing data problems. *Statistical Analysis with Missing Data*, Third Edition starts by introducing readers to the subject and approaches toward solving it. It looks at the patterns and mechanisms that create the missing data, as well as a taxonomy of missing data. It then goes on to examine missing data in experiments, before discussing complete-case and available-case analysis, including weighting methods. The new edition expands its coverage to include recent work on topics such as nonresponse in sample surveys, causal inference, diagnostic methods, and sensitivity analysis, among a host of other topics.

- An updated “classic” written by renowned authorities on the subject
- Features over 150 exercises (including many new ones)
- Covers recent work on important methods like multiple imputation, robust alternatives to weighting, and Bayesian methods
- Revises previous topics based on past student feedback and class experience
- Contains an updated and expanded bibliography

The authors were awarded The Karl Pearson Prize in 2017 by the International Statistical Institute, for a research contribution that has had profound influence on statistical theory, methodology or applications. Their work “has been no less than defining and transforming.” (ISI) *Statistical Analysis with Missing Data*, Third Edition is an ideal textbook for upper undergraduate and/or beginning graduate level students of the subject. It is also an excellent source of information for applied statisticians and practitioners in government and industry.

Library: QA276.

Pawitan: In all likelihood: Statistical modelling and inference using likelihood

Pawitan-2013

Yudi Pawitan. *In all likelihood: Statistical modelling and inference using likelihood*. Oxford University Press, Jan. 17, 2013. 544 pp. ISBN: 9780199671229.

Abstract: Based on a course in the theory of statistics this text concentrates on what can be achieved using the likelihood/Fisherian method of taking account of uncertainty when studying a statistical problem. It takes the concept of the likelihood as providing the best methods for unifying the demands of statistical modelling and the theory of inference. Every likelihood concept is illustrated by realistic examples, which are not compromised by computational problems. Examples range from a simple comparison of two accident rates, to complex studies that require generalised linear or semiparametric modelling. The emphasis is that the likelihood is not simply a device to produce an estimate, but an important tool for modelling. The book generally takes an informal approach, where most important results are established using heuristic arguments and motivated with realistic examples. With the currently available computing power, examples are not contrived to allow a closed analytical solution, and the book can concentrate on the statistical aspects of the data modelling. In addition to classical likelihood theory, the book covers many modern topics such as generalized linear models and mixed models, non parametric smoothing, robustness, the EM algorithm and empirical likelihood.

van Buuren: Flexible imputation of missing data

vanBuuren-2018

Stef van Buuren. *Flexible imputation of missing data*. 2nd ed. Chapman and Hall/CRC, July 2018. ISBN: 9780429492259. DOI: [10.1201/9780429492259](https://doi.org/10.1201/9780429492259).

Abstract: Missing data pose challenges to real-life data analysis. Simple ad-hoc fixes, like deletion or mean imputation, only work under highly restrictive conditions, which are often not met in practice. Multiple imputation replaces each missing value by multiple plausible values. The variability

between these replacements reflects our ignorance of the true (but missing) value. Each of the completed data set is then analyzed by standard methods, and the results are pooled to obtain unbiased estimates with correct confidence intervals. Multiple imputation is a general approach that also inspires novel solutions to old problems by reformulating the task at hand as a missing-data problem. This is the second edition of a popular book on multiple imputation, focused on explaining the application of methods through detailed worked examples using the MICE package as developed by the author. This new edition incorporates the recent developments in this fast-moving field. This class-tested book avoids mathematical and technical details as much as possible: formulas are accompanied by verbal statements that explain the formula in accessible terms. The book sharpens the reader's intuition on how to think about missing data, and provides all the tools needed to execute a well-grounded quantitative analysis in the presence of missing data.

Library: QA278.