Book Review

Reviewed by Zhongxian Wang* and
Sylvain Jaume

Email: wangz@mail.montclair.edu
Email: sjaume@saintpeters.edu
*Corresponding author

Statistics Done Wrong: The Woefully Complete Guide (1st edition)
by: Alex Reinhart
Published March 16, 2015
by No Starch Press
245 8th street, San Francisco, CA 94103, USA, 176pp
ISBN: 978-1-59327-620-1

Modern society depends on scientific discovery and technology breakthroughs (Carriquiry, 2015). Statistical analysis is an extremely powerful tool because almost every scientific discipline depends on statistical analysis of experimental data. Statistics plays a predominant role in almost all natural and social sciences. A well-founded understanding of basic statistics is critical for everyone in science to produce statistically sound research.

However, statistics analysis could become an entangling snare since there are so many pitfalls in statistical analysis. As Reinhart pointed out, you would be surprised how many scientists are doing it wrong. The first step toward statistics done right is *Statistics Done Wrong*.

In Chapter 1, the author provides an introduction to statistical significance. When a difference between two events is larger than could easily be produced by luck and/or random variation, statisticians could claim that the difference is statistically significant. The power of p values shows a measure of surprise if your assumption is wrong, but not how right you are or how important a difference is. The author emphasises that statistical significance does not mean your result has any *practical* significance whereas statistical insignificance does not tell you much.

Statistical power and underpowered statistics are presented in Chapter 2. The author suggests that determining the appropriate sample size is a critical starting point for a research. Most likely, colossal amounts of data would be needed to consistently detect a small difference. We should look skeptically on the results of clearly underpowered studies because they may be hyperbolic due to truth inflation. We need to compute confidence intervals when comparing groups of different sizes.

In Chapter 3, ‘Pseudoreplication: choose your data wisely’, the author sees the substance of the word *random* in a randomised controlled trial. We should ensure that our statistical analysis really answers our research question. In order to account for a strong dependence between our measurements, it is better for us to use statistical methods such as hierarchical models and clustered standard errors. Also, we could design experiments to eliminate hidden sources of correlation between variables.

The p value and the base rate fallacy are explored in Chapter 4. P values are hard to interpret even they can be easily found in a software report. Theoretically, we can reject a null hypothesis if $p < 0.05$, as any statistics textbooks claim. However, it does not mean that our result is 95% accurate. We should use prior estimates of the base rate to calculate the probability that a given result is a false positive. In the case of testing multiple hypotheses or looking for correlations between many variables, we should use a procedure, such as Bonferroni or Benjamini-Hochberg, etc., to control for the excess of false positive.

Chapter 5 demonstrates bad judges of significance. There two sides for the same coin with using too many statistical significance test: either misleading or finding an unexpected difference. We should compare groups directly using appropriate statistical tests, rather than judge the significance of a difference by eyes. We need to adjust for making multiple comparisons when comparing many groups. Overlapping confidence intervals do not mean two values are not significantly different at all.

Double-dipping in the data is the focusing point in Chapter 6. An exploratory data analysis involves making numerous plots, trying a few plans, and following any emerging leads. In the meantime, separate data should be used to avoid biased results. Stopping rules should be arranged in advance carefully. Otherwise, false positives and truth inflation could disguise the truth. That is, exploratory findings should be considered tentative until confirmed.

Chapter 7 checks continuity errors. According to the author, in case of need to split continuous variables into groups for some reason, do not choose the groups to maximise our statistical significance. Defining the split in advance, using the same split as in previous similar research, or using outside standards are possible options. That is, we should not arbitrarily split continuous variables into discreet groups.

Model abuse had been noticed in Chapter 8 via regression. There are many more versions of regression in the real world. The relationship among variables could be nonlinear, rather than linear. The dependent variable could be categorical. Simpson's Paradox could arise. Therefore, an obvious trend can be eliminated or even reversed by splitting the data into natural groups. We can use random assignment to solve this dilemma. We need to avoid biased stepwise regression when possible. Other selection techniques (such as the lasso), or no variable selection, may be more appropriate.

Chapter 9 discovers researcher freedom by using good variables. The author believes that there are many decisions for a researcher to make, such as what do I measure? Which variables do I adjust for? Which cases do I exclude? How do I define groups? What about missing data? How much data should I collect, etc.? Researcher should plan data analysis, accounting for multiple comparisons and including any possible effects, before collecting data. We should not constantly run the model and force the data produce our expected results.

Everybody makes mistakes, as Chapter 10 displays. The author suggests that researchers should test all analysis programs against known input and ensure the results make sense. When writing programs and scripts to analyse imported data, researchers should follow the 'Ten simple rules for reproducible computational research'. Researchers can use a reproducible research tool like Sweave to automatically include data from their analysis in their paper.

Chapter 11 handles hiding the data. Researchers should make all data available when possible, through specialised databases such as GenBank and PDB or through generic data repositories such as Dryad and Figshare. Any deviations from the trial protocol should be documented and discussed in their published paper. Their software source code, Excel workbooks, or analysis scripts used to analyse their data should also be published.

Chapter 12 investigates what can be done to clean the mess. In line with the author, improving statistical education is a starting and decisive factor. Misconceptions are everywhere because much of basic statistics is not intuitive. Academic journals should pressure authors to more rigorous standards. Using article-level metrics could be one of solutions. Especially online-only journals can easily measure the number of views of an article, the number of citation it has received in other articles. Outsourcing helps a lot. Otherwise, a researcher needs a solid foundation in statistics to play the game lonely.

There is progress in checking the adequacy of a regression model recently. For instance, Zhang et al. (2015) propose a residual-based empirical process test statistic marked by proper functions of the regressors. They study tools for checking the validity of a parametric regression model, when both response and predictors are unobserved and distorted in a multiplicative fashion by an observed confounding variable.

Statistics Done Wrong is an indispensable guide that can help researchers to avoid statistical missteps in modern science. We really enjoyed reading this book due to its rigorous analysis, valuable tips, convincing examples, and humorous statements.

You would be surprised to see so many exaggerated results, awkward faults and lapses in recent research (Schwartzberg et al., 2015). This abridged and authoritative guide will show you how to shun these pitfalls that might skew public's perceptions of important issues. You will learn procedures to follow, precautions to take, and helpful software.

References

Carriquiry, A.L. (2015) 'Measurement error models', in Wright, J.D. (Ed.): *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed., pp.850–855, Elsevier Ltd., Oxford, DOI: 10.1016/B978-0-08-097086-8.42145-8.

Schwartzberg, D., Ivanovic, S., Patel, S. and Burjonrappa, S.C. (2015) 'We thought we would be perfect: medication errors before and after the initiation of computerized physician order entry', *Journal of Surgical Research*, in press, Uncorrected Proof, Available online 30 April 2015.

Zhang, J., Li, G. and Feng, Z. (2015) 'Checking the adequacy for a distortion errors-in-variables parametric regression model', *Computational Statistics & Data Analysis*, March, Vol. 83, pp.52–64.