Weizhen Xie
Department of Psychology
University of California, Riverside
Handout for PSYC213

Perspectives in Statistical Practice for Psychological Science

Analyzing data can be a blessing and a curse in graduate school. On the one hand, great discoveries and the joy of generating knowledge critically rely on appropriate statistical practice. On the other hand, the journey to making sense of the data can be uneasy: confusing equations may be haunting around, organizing messy statistical outputs can be laborious and tedious, and the worst, a full sheet of non-significant results can keep you up all night.

Indeed, when one is immersed in the jungle of data analysis, it is quite easy to get lost. Here, these confusions are not referred to as difficulties caused by fancy equations or complicated computations. Remember, we can always learn things that are invented by others. The jungle would get dangerous when one fails to find a right track and starts doing something unwise. In statistical practice, this can be the inability to follow appropriate statistical procedures or, even worse, adopting a procedure without questioning. In a statistical language, there may not be “right” or “wrong”; but there is surely a line between “appropriate” and “inappropriate.” However, this line is often blurry. How to make sure that we always stay on the appropriate side of the line? One way is to have a “big picture” view about statistical practice. Just like having a map may save one’s life in a jungle, having a bird’s eye view on statistical practice may help you make better decisions when it comes to data analysis.

This reading list is thus created to introduce different perspectives and their rebuttals about statistical practice for psychological science to help you formulate a bird’s eye view on the field. Note, similar principles and ideas may also be generalized to other scientific areas. The reading list starts with the introduction of issues and problems of everyday statistical practice, including reproducibility, false positive discovery, and publication bias, etc. With a knowledge about these issues and problems in their mind, readers should think about how to avoid these poisonous-snake-like obstacles in the jungle of data analysis. Next, the reading list introduces several approaches and ideas to mitigate the issues and problems, with a caveat that none are perfect. Therefore, readers are encouraged to consider both the promises and pitfalls of each approach, and decide wisely whether to adopt an approach in a specific case. Furthermore, it should be noted this is not an exhaustive list. New methods keep coming once a while. It is important for readers to keep track of emerging approaches and evaluate them critically.

All in all, the motivation of this reading list is to nurture sophisticated critical thinking about data analysis in psychological science, with two specific aims to 1) promote appropriate statistical practice and 2) to facilitate future methodological advance to resolve some issues in the jungle of data analysis.

1. Background: Everyday Statistical Practice as a Researcher

Perspectives

Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist*, *24*(2), 83–91.

Judd, C. M., McClelland, G. H., & Culhane, S. E. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology*, *46*(1), 433–465.

2. Reproducibility Issue

Perspectives

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4718.

Rebuttals

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science". *Science*, *351*(6277), 1037–1037.

Anderson, C. J., Bahník, Š., Barnett-Cowan, M., Bosco, F. A., Chandler, J., Chartier, C. R., et al. (2016). Response to Comment on "Estimating the reproducibility of psychological science". *Science*, *351*(6277), 1037–1037.

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(23), 6454–6459.

Drummond, C. (2009). Replicability is not reproducibility: nor is it good science. Retrieved from http://cogprints.org/7691/7/ICMLws09.pdf

3. False Positive Psychology: Excess Success, Researcher Degrees of Freedom, and Publication Bias

Perspectives

Francis, G., Tanzman, J., & Matthews, W. J. (2014). Excess success for psychology articles in the journal Science. *PLoS ONE*, *9*(12), e114255–15.

Francis, G. (2014). The frequency of excess success for articles in Psychological Science. *Psychonomic Bulletin & Review*, *21*(5), 1180–1187.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). p-Curve and Effect Size. *Perspectives on Psychological Science*, *9*(6), 666–681.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547.

Stahl, D., & Pickles, A. (2017). Fact or fiction: reducing the proportion and impact of false positives. *Psychological Medicine*, *2*, 1–10.

Rebuttals

Van Boxtel, J. J. A., & Koch, C. (2016). Reevaluating excess success in psychological science. *Psychonomic Bulletin & Review*, *23*(5), 1602–1606.

Ioannidis, J. P. A. (2013). Clarifications on the application and interpretation of the test for excess significance and its extensions. *Journal of Mathematical Psychology*, *57*(5), 184–187.

Morey, R. D. (2013). The consistency test does not–and cannot–deliver what is advertised: A comment on Francis (2013). *Journal of Mathematical Psychology*, *57*, 190–183.

Bruns, S. B., & Ioannidis, J. P. A. (2016). p-Curve and p-Hacking in Observational Research. *PLoS ONE*, *11*(2), e0149144–13.

Medina, J., & Cason, S. (2017). No evidential value in samples of transcranial direct current stimulation (tDCS) studies of cognition and working memory in healthy populations. *Cortex*, *94*, 131–141.

4. Cumulating Research Evidence: Meta-Analysis

Perspectives

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, *52*, 59–82. http://doi.org/10.1146/annurev.psych.52.1.59

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to Meta-Analysis. Hoboken, NJ: John Wiley & Sons.

Rubin, D. B. (1992). Meta-analysis: Literature synthesis or effect-size surface estimation? *Journal of Educational and Behavioral Statistics*, *17*(4), 363–374. http://doi.org/10.3102/10769986017004363

Rebuttals

Bailar, J. C. (1997). The promise and problems of meta-analysis. *New England Journal of Medicine*, *337*(8), 559–561.

Boden, W. E. (1992). Meta-analysis in clinical trials reporting: has a tool become a weapon? *The American Journal of Cardiology*, *69*(6), 681–686.

LeLorier, J., Grégoire, G., Benhaddad, A., Lapierre, J., & Derderian, F. (1997). Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *New England Journal of Medicine*, *337*(8), 536–542.

Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: why validity issues in meta-analysis will not go away. *Clinical Psychology Review*, *17*(8), 881–901.

Thompson, S. G., & Pocock, S. J. (1991). Can meta-analyses be trusted? *The Lancet*, *338*(8775), 1127–1130.

5. Re-defining Significance

Perspectives

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., et al. (2017). Redefine statistical significance. *Nature Human Behaviour*, 1–0.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement on p-Values: Context, Process, and Purpose. *American Statistician*, *70*(2), 129–133.

Rebuttals

Crane, H. (2017). Why “Redefining Statistical Significance” Will Not Improve Reproducibility and Could Make the Replication Crisis Worse. Retrieved from https://arxiv.org/abs/1711.07801.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*(12), 997–1003.

6. Advocating Open Science Practice

Perspectives

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., et al. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, du, N. P., et al. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 1–9.

Rebuttals

Drew, T. W., & Mueller-Doblies, U. U. (2017). Dual use issues in research - A subject of increasing concern? *Vaccine*, *35*(44), 5990–5994.

7. Another Way Out: Bayesian Hierarchical Modeling

Perspectives

Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, *12*(3), 403–408.

Rouder, J. N., Morey, R. D., & Pratte, M. S. (1965). Bayesian Hierarchical Models. In W. Batchelder, H. Colonius, E. Dzhafarov, & J. I. Myung (Eds.), *New Handbook of Mathematical Psychology, Voume 1: Measurement and Methodology* (pp. 1–82). London, UK: Cambridge University Press.

Gelman, A., & Shalizi, C. R. (2012). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 8–38.

Zyphur, M. J., & Oswald, F. L. (2015). Bayesian estimation and inference: A user's guide. *Journal of Management*, *41*(2), 390–420. http://doi.org/10.1177/0149206313501200

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*(3), 299–312.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189.

Rebuttals

Efron, B. (1986). Why Isn't Everyone a Bayesian? *American Statistician*, *40*(1), 1–5.

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science*, *6*(3), 274–290.

8. Reinstating Statistical Inference as a Form of Model Comparison

Perspectives

Platt, J. R. (1964). Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*, *146*(3642), 347–353.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). Contrasts and effect sizes in behavioral research: A correlational approach. New York: Cambridge University Press.

McElreath, R. (2016). Statistical Rethinking: A Bayesian Course with Examples in R and Stan. Boca Raton, Florida: CRC Press. (Read Chapter 1).

Wagenmakers, E. J., & Morey, R. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, *25*(3), 169–176.

Gelman, A., & Shalizi, C. R. (2012). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 8–38.

Rebuttals

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*(10), 421–425. http://doi.org/10.1016/S1364-6613(02)01964-2

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367.

Roberts, S., & Pashler, H. (2002). Reply to Rodgers and Rowe (2002). *Psychological Review*, *109*(3), 605–607. http://doi.org/10.1037//0033-295X.109.3.605

Rodgers, J. L., & Rowe, D. C. (2002a). Postscript: Theory development should not end (but always begins) with good empirical fits: Response to Roberts and Pashler's (2002) reply. *Psychological Review*, *109*(3), 603–604. http://doi.org/10.1037/h0087895

Rodgers, J. L., & Rowe, D. C. (2002b). Theory development should begin (but not end) with good empirical fits: a comment on Roberts and Pashler (2000). *Psychological Review*, *109*(3), 599–604. http://doi.org/10.1037//0033-295X.109.3.599

Veksler, V. D., Myers, C. W., & Gluck, K. A. (2015). Model flexibility analysis. *Psychological Review*, *122*(4), 755–769.