TEACHING STATISTICAL INFERENCE AT THE UNIVERSITY LEVEL: THE ITALIAN EXPERIENCE

Franca Agnoli
Department of Developmental Psychology and Socialization (DPSS)
University of Padova, Padova, Italy
Franca.Agnoli@unipd.it

The reproducibility crisis in psychology has been attributed to a faulty application of the NHST paradigm. Some researchers have advocated and adopted other methods, which calls for changes in university-level statistics courses. Instructors should de-emphasize NHST and teach new approaches such as effect size estimation and Bayesian statistics. A multi-phase investigation was conducted of statistics courses for psychology students at Italian Universities to determine how these courses are evolving and the problems these changes pose for students. Phase 1: an analysis of the course contents and required texts listed on university websites. Phase 2: a survey of the time devoted to traditional and newer analysis methods in these courses. Phase 3: interviews with instructors regarding the topics that are proving most difficult for students.

INTRODUCTION

Increasing awareness of methodological and statistical problems in psychological research has led to discussions of a crisis of confidence in the published research literature (e.g., Pashler & Wagenmakers, 2012). Researchers have found evidence of widespread methodological and statistical biases in the literature. Among the most shocking findings is Ioannidis' (2005) observation that most research findings in the biomedical field are false because of these biases. A large part of the very high rate of false positive findings in the social science literature can be explained as the result of flexibility in data collection, analysis, and reporting practices (Simmons, Nelson, & Simonsohn, 2011; Hubbard, 2016). John, Loewenstein, and Prelec (2012) surveyed academic psychologists in the US regarding their use of some questionable research practices and found that nearly all reported having used at least one such practice. I found similar results in a survey of Italian psychologists (Agnoli, Wicherts, Veldkamp, Albiero, & Cubelli, 2017), with 88% reporting that they had used at least one of the practices included in the survey. Both the US and Italian researchers who reported using a questionable research practice generally considered their use of the practice to be defensible and not an error.

Faced with these issues in the published psychological literature, researchers have begun proposing, testing, and analyzing various ways to avoid or reduce these problems in the future. Solutions to these systemic problems will require a greater emphasis on the quality of research instead of its quantity (see Cubelli & Della Sala, 2015; Sarewitz, 2016). Recent papers have suggested steps that would introduce greater transparency in the research process with respect to data (Wicherts, Bakker, & Molenaar, 2011), collaboration (Veldkamp, Nuijten, Dominguez-Alvarez, van Assen, & Wicherts, 2014), research materials (Nosek, Spies, & Motyl, 2012), and analyses (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). Pre-registration of research hypotheses and detailed analyses plans (e.g., Chambers, 2013) and increasing the statistical power of studies (Bakker, Hartgerink, Wicherts, & van der Maas, 2016) should help lower QRP use and increase reproducibility and replicability of research findings in psychology.

As Ziliak and McCloskey (2008) observed, changes in education are an essential element of statistical reform because students have been learning false beliefs from university professors in a persistent cycle of misinformation. They wrote (p. 250), "The textbooks are wrong. The teaching is wrong. The seminar you just attended is wrong. The most prestigious journal in your scientific field is wrong." Education is a key component of any serious plan to correct the existing problems (Kline, 2013). This includes educating current researchers about the steps they should take to ensure their research is conducted properly and can be trusted. It also includes educating the next generation of researchers – the undergraduate and graduate students at our universities.

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org [© 2018 ISI/IASE]

EVOLUTION OF STATISTICAL TRAINING FOR PSYCHOLOGY STUDENTS IN ITALY

All Italian universities are expected to include a course in statistics as part of their three-year undergraduate programs in psychology, but each instructor is at liberty to decide what content to cover and what textbook to use in that course. At the University of Padova, we have been struggling with how to revise our undergraduate statistics courses, what to include, and how to most effectively teach it. I decided to investigate how universities throughout Italy have addressed this, while also implementing some tentative innovations in our own courses.

I devised a multi-phase plan for investigating this. The first phase was to analyze the course contents and required texts listed on the websites of undergraduate statistics courses for psychologists in Italian universities. The second phase was a survey of the time devoted to traditional and newer analysis methods in these courses. The third phase was interviews with instructors about the topics that are proving most difficult for students. As of this writing, only the first phase has been completed, and the results are presented below.

THE STATISTICS CURRICULUM FOR PSYCHOLOGY IN ITALY

I extracted the content of 87 courses from the websites of 18 Italian universities. Most of these 18 universities require that psychology students complete an introductory statistics course, whereas others require only a course in research methodology. The 87 courses extracted in this investigation included both introductory statistics and research methodology courses, as well as more advanced courses taught as part of a two-year masters program.

Introductory statistics is taught either in the first or second year of the undergraduate three-year program in psychology. The contents of the typical introductory course include the following:

- Measurement in psychology (measurement scales and transformations)
- Descriptive statistics
- Probability theory
- Random variables (discrete and continuous)
- Probability distributions (binomial, normal, Chi-square, and Student *t*)
- Statistical inference: Null Hypothesis Significance Testing paradigm
- Inferential tests (two-sample tests with known and unknown variance)
- Contingency tables
- Linear correlation and regression

Note that this course outline could readily describe a course offered in an undergraduate psychology program taught in the United States 10, 20, or even 30 years ago. I looked for evidence that instructors were integrating statistics reform in their courses. The evidence I was seeking included topics such as the role of statistical power, confidence intervals, effect size estimation, importance of replication, the controversy regarding the Null Hypothesis Significance Testing (NHST) paradigm, and Bayesian estimation. The evidence was strikingly absent. A few courses explicitly mentioned effect sizes (5 courses) and a few mentioned confidence intervals, but none mentioned both effect sizes and confidence intervals. It should be noted that confidence intervals might have been included as part of the section on inferential tests, with all the problems that those tests entail. There is a wealth of research showing that researchers misunderstand confidence intervals and standard error bars (Belia, Fidler, Williams, & Cumming, 2005). None of the introductory course descriptions included anything about the role of statistical power, the importance of replication, the controversy regarding NHST, or Bayesian estimation. Only one course description distinguished between statistical significance and the strength of an effect. None of the course descriptions mentioned anything about misinterpretations of *p* values or related statistical errors (Nuijten, Hartgerink, van Assen, Epskamp, & Wicherts, 2016).

The most notable widely adopted innovation among the 87 Italian statistics courses is teaching students to use R to perform analyses. This is a potentially important step because it gives the students access to modern modeling methods and also may be used to calculate confidence intervals, effect size estimates, and Bayesian statistics if students learn that these calculations are necessary and important.

Some of the two-year masters programs in Italy require additional training in statistics and/or psychological testing. At my own university, we offer two advanced courses, one that I teach focuses on the controversies regarding the NHST paradigm and emphasizes use of effect sizes for different experimental designs. A colleague offers another course that contrasts the frequentist and Bayesian approaches. Outside my university, advanced courses emphasize classical inferential statistics, including topics such as analysis of variance, multivariate analysis, and nonparametric methods. Only a few course descriptions included any mention of effect sizes and just two mentioned Bayesian approaches, one course aimed at clinical diagnosis and another intended for forensic applications.

The textbooks used in both the undergraduate and graduate courses are generally Italian translations of English-language statistics books published many years ago. Some books are written in Italian for professors to use in their own classes, and these books closely follow the organizational structure of older English-language statistics books.

CONCLUSION

This analysis of statistics courses in Italian psychology programs was initiated to gain some insights about how these courses are evolving in response to the need for statistical reform. The results were disappointing and suggest that there has been little or no evolution in the courses. It should be noted, however, that this first phase of analysis was based entirely on course descriptions posted on the university websites, and some instructors may have added material or introduced innovations that are not yet reflected on their websites.

Some instructors may also be reluctant to change their courses without greater agreement and clearer guidance from the research community. Although some journals and some editors are strongly advocating statistical reform, papers that rely entirely on null hypothesis significance testing are still published, and evaluations of research productivity are based primarily on number of publications, not the quality of publications. University instructors may also be discouraged from revising their course content because textbooks that advocate statistical reform (e.g., Cumming, 2012) are not yet available in Italian.

REFERENCES

Agnoli, F., Wicherts, J. M., Veldkamp, C. L., Albiero, P., & Cubelli, R. (2017). Questionable Research Practices among Italian Research Psychologists. *PLoS One, 12*(3), e0172792.

Bakker, M., Hartgerink, C. H., Wicherts, J. M., & van der Maas, H. L. (2016). Researchers' Intuitions about Power in Psychological Research. *Psychological Science, 27*(8), 1069-1077.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers Misunderstand Confidence Intervals and Standard Error Bars. *Psychological Methods, 10*(4), 389.

Chambers, C. D. (2013). Registered Reports: A New Publishing Initiative at Cortex. *Cortex, 49*(3), 609-610.

Cubelli, R., & Della Sala, S. (2015). Write Less, Write Well. *Cortex, 73*, A1-2.

Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York: Routledge.

Hubbard, R. (2016). *Corrupt Research: The Case for Reconceptualizing Empirical Management and Social Science.* Los Angeles: Sage.

Ioannidis, J. P. (2005). Why Most Published Research Findings Are False. *PLoS Medicine, 2*(8), e124.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling. *Psychological Science, 23*(5), 524-532.

Kline, R. B. (2013). *Beyond Significance Testing: Statistics Reform in the Behavioral Sciences* (2nd edition). Washington, DC: American Psychological Association.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring Incentives and Practices to Promote Truth over Publishability. *Perspectives on Psychological Science, 7*(6), 615-631.

Nuijten, M. B., Hartgerink, C. H., van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016). The Prevalence of Statistical Reporting Errors in Psychology (1985–2013). *Behavior Research Methods, 48*(4), 1205-1226.

Pashler, H., & Wagenmakers, E-J. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science. *Perspectives on Psychological Science, 7*(6), 528-530.

Sarewitz, D. (2016). The Pressure to Publish Pushes Down Quality. *Nature, 533*(7602), 147.

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.

Veldkamp, C. L., Nuijten, M. B., Dominguez-Alvarez, L., van Assen, M. A., & Wicherts, J. M. (2014). Statistical Reporting Errors and Collaboration on Statistical Analyses in Psychological Science. *PloS One*, *9*(12), e114876.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, *7*(6), 632-638.

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. *PloS One*, *6*(11), e26828.

Ziliak, S., & McCloskey, D. N. (2008). *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor: University of Michigan Press.