6-5-2018

Robust Statistics

Bruce E. Blaine
St. John Fisher College, bblaine@sjfc.edu

Publication Information

Blaine, Bruce E. (2018). "Robust Statistics." *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* , 1434-1436.

Robust Statistics

Abstract
In lieu of an abstract, here is the entry's first paragraph:

Robust statistics are procedures that maintain nominal Type I error rates and statistical power in the presence of violations of the assumptions that underpin parametric inferential statistics. Since George Box coined the term in 1953, research on robust statistics has centered on the assumption of normality, although the violation of other parametric assumptions (e.g., homogeneity of variance) has their own implications for the accuracy of parametric procedures. This entry looks at the importance of robust statistics in educational and social science research and explains the robustness argument. It then describes robust descriptive statistics, their inferential extensions, and two common resampling procedures that are robust alternatives to classic parametric methods.

Disciplines
Statistics and Probability

Comments
This is an entry in: Frey, B. (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation* (Vols. 1-4). Thousand Oaks,, CA: SAGE Publications, Inc.: https://dx.doi.org/10.4135/9781506326139

Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology, 60*, 151–174.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*, 504–528.

Llorente, E., Warren, C. S., de Eulate, L. P., & Gleaves, D. H. (2013). A Spanish version of the sociocultural attitudes towards appearance questionnaire-3 (SATAQ-3): Translation and psychometric evaluation. *Journal of Clinical Psychology, 69*(3), 240–251. doi:10.1002/jclp.21944

Rodebaugh, T. L, Woods, C. M, Heimberg, R. G., Liebowitz, M. R., & Schneier, F. R. (2006). The factor structure and screening utility of the Social Interaction Anxiety Scale. *Psychological Assessment, 18*(2), 231–237.

Tay, L., & Drasgow, F. (2012). Theoretical, statistical, and substantive issues in the assessment of construct dimensionality: Accounting for the item response process. *Organizational Research Methods, 15*(3), 363–384.

VonSonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloSONE, 8*(7), e68967. doi:10.1371/journal.pone.0068967

Weijters, B., Baumgartner, H., & Schillewaet, N. (2013). Reversed item bias: An integrative model. *Psychological Method, 18*(3), 320–334. doi:10.1037/a0032121

Robust Statistics

Robust statistics are procedures that maintain nominal Type I error rates and statistical power in the presence of violations of the assumptions that underpin parametric inferential statistics. Since George Box coined the term in 1953, research on robust statistics has centered on the assumption of normality, although the violation of other parametric assumptions (e.g., homogeneity of variance) has their own implications for the accuracy of parametric procedures. This entry looks at the importance of robust statistics in educational and social science research and explains the robustness argument. It then describes robust descriptive statistics, their inferential extensions, and two common resampling procedures that are robust alternatives to classic parametric methods.

Robust statistics are important tools for educational and social science researchers because of three well-established findings. First, parametric methods (e.g., ANOVA, least squares regression) are the most commonly used procedures for significance testing in the social sciences; some estimates indicate that over 90% of published articles use a parametric significance test. Second, surveys of the educational and psychological literature show that nonnormally distributed data is the rule rather than the exception. Third, even modest departures from normality can substantially compromise both the Type I error rate and the power of parametric inferential procedures.

The robustness argument refers to the longstanding claim in the social sciences that parametric procedures such as the *t* test are "robust to violations" of the assumption of normality, meaning that the tests maintain accurate Type I error rates in the face of nonnormality. Originating in several research articles from the 1970s, the robustness argument has been repeated in introductory statistics textbooks, asserted by researchers in defense of their use of parametric methods, and over time become accepted as fact in the social science research community.

The near ubiquity of parametric procedures for significance testing in social science research speaks to the acceptance of the robustness argument. However, the research underlying the robustness argument has been criticized both for its methods and interpretation of results. Subsequent research has substantially, if not convincingly, established that beyond some very specific circumstances in which parametric procedures are in fact robust to violations of the assumption of normality, the robustness of *t* tests and other parametric procedures to violations of normality is the exception rather the rule.

The robustness argument invokes the central limit theorem, which provides for normal sampling distributions of the mean (given adequate sample size) even when the parent population is not normally distributed. However, the central limit theorem says nothing about the distribution of *t*, from which probabilities are derived for *t* tests of null hypotheses and *t* quantiles derived for constructing confidence intervals (CIs). Simulation studies show that under conditions of nonnormality, inferences based on the *t* distribution are inaccurate (i.e., nominal Type I error rates are not maintained) and can be very inaccurate even with modest departures from normality in the parent population. Combined with the commonality of nonnormally distributed data mentioned earlier, the influence of the robustness argument on statistical practices has broad implications for research literatures in education, psychology, and beyond.

The most pernicious departures from normality, from the standpoint of undermining parametric significance tests, are those that take the form of heavy tailed distributions. Heavy tailed (also called contaminated normal) distributions are common in educational research where a target population is contaminated with cases from subpopulations that have different means and variances, or from the presence of outliers, or both. Worse, heavy tailed distributions appear to be normal by visual inspection, and their nonnormality often goes undetected by tests of normality. Robust descriptive statistics are, by definition, resistant to the influence of outliers, and inferential procedures that use robust descriptive statistics inherit the same resistant quality.

Common robust descriptive statistics include the trimmed mean and variance, Winsorized mean and variance, and M-estimators. As a group, these statistics moderate the influence of outliers or heavy tails on estimates of location and variability and are much preferable to data transformations as methods to deal with outliers or restore normality. Trimming removes a set percentage of cases in the upper and lower tails and calculates the mean or variance of the remaining cases. The median, which is widely appreciated as being resistant to the influence of outliers, is a 50% trimmed mean and therefore a robust estimator of location.

Winsorizing involves the systematic recoding of cases in the tails of a skewed or heavy tailed distribution, with the mean and variance calculated from the recoded data. Like Winsorizing, M-estimators also reassign values to observations in the tails of a distribution but do so based on one of several estimating functions. When these robust descriptive statistics are used in parametric inferential procedures, such as when a *t* test is calculated with trimmed means and variances, those procedures in turn become more robust. Robust descriptive and inferential statistics can be generated in most modern statistical software packages.

The robust procedures just described rely on theoretical probability distributions (e.g., *t*) to approximate the underlying distribution and generate probabilities for inference but do so with robust estimators of mean and variance. This category of robust inferential procedures is therefore still *parametric*. In contrast, other robust methods create empirical probability distributions from sample data and use those distributions for inference and estimation.

Certain robust procedures are freed from parametric assumptions, such as the assumption of normality, because the underlying probability distribution is directly estimated from sample data rather than approximated by a mathematical distribution. Two common examples are the bootstrapped CI and the permutation test for a mean difference. A bootstrapped 95% CI for estimating μ is produced via a resampled distribution of thousands of sample means. From that distribution, the 2.5% and 97.5% quantiles become the lower and upper limits, respectively, of the CI. A permutation test for a mean difference also starts with sample data, creating a probability distribution of mean differences from thousands of independent shufflings of scores into two random samples, each generating a mean difference, from which a *p* value for the observed mean difference can be retrieved. Resampled, robust alternatives exist for most parametric inferential procedures and are also part of most statistical software packages.

B. Evan Blaine

See also Random Assignment; Winsorizing

Further Readings

Erceg-Hurn, D., & Mirosevich, V. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist, 63*(7), 591–601.

Huber, P. (1972). The 1972 Wald Lecture robust statistics: A review. *Annals of Mathematical Statistics, 43(4)*, 1041–1067.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156–166.

Wilcox, R. (2009). Robust data analysis. In R. Milsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 387–403). Thousand Oaks, CA: SAGE.