SHOULD WE STILL TEACH RANK-BASED DISTRIBUTION-FREE PROCEDURES?

E. Jacquelin Dietz
Meredith College, Raleigh NC, USA
dietzjac@meredith.edu

For decades, I have enjoyed teaching rank-based distribution-free inference procedures for two distinct reasons. First, I have believed these are useful data analysis methods that should be part of any applied statistician's repertoire of statistical methods. Second, I have found rank-based tests ideal for teaching hypothesis testing – many students report that they never really understood sampling distributions and p-values until they studied rank-based tests. Recently, many instructors have begun teaching inference in introductory courses using bootstrapping and randomization tests in place of traditional normal theory methods. New software has made it feasible to apply randomization methods to the original observations. Is there now less motivation to rank data? Can we teach the fundamental concepts of hypothesis testing just as well using randomization methods on the original observations? Are rank procedures still important methods of data analysis that we should be teaching to our students?

INTRODUCTION

As a graduate student at the University of Connecticut in the 1970's, I was strongly attracted to rank-based distribution-free inference procedures. My dissertation topic was "Bivariate Nonparametric Tests for the One-Sample Location Problem," and my first teaching experience was serving as a Teaching Assistant for an introductory statistics course that used Gottfried Noether's book *Introduction to Statistics: A Nonparametric Approach* (Noether, 1976). Shortly after arriving at North Carolina State University for my first faculty position, I developed a new course, Applied Nonparametric Statistics, which I taught 24 times during my 26 years there. When I moved to Meredith College ten years ago, I was surprised and pleased to learn that one of the three second-level statistics courses taught at Meredith was a course in Nonparametric Statistics; I have now taught that course five times.

I always begin my nonparametrics course by enumerating the advantages of rank-based distribution-free inference procedures. I characterize rank-based tests and confidence interval procedures as methods that require relatively mild assumptions about the underlying population distribution, and I explain that they are applicable in situations where classical normal-theory methods may not be. I tell my students that the procedures are often intuitively appealing and easy to compute. I describe the resistance of the procedures to outliers and tout their high power for heavy-tailed distributions.

Anyone who has taught a traditional introductory statistics course is likely to agree that sampling distributions and *p*-values are two of the most difficult topics for students. Even students who can correctly carry out the steps of a hypothesis test for means or proportions may be hard pressed to explain what the *p*-value of a test really means. Many times during my teaching career, students in my nonparametrics classes have reported that they never really understood sampling distributions and *p*-values until they studied rank-based tests. The reasoning involved in a rank-based nonparametric hypothesis test can be very intuitive for students compared to the reasoning required for a traditional normal or *t*-based test.

Gottfried Noether realized decades ago that "a nonparametric approach" to introductory statistics could have advantages over the traditional normal theory approach. The second edition of Noether's book, *Introduction to Statistics: A Nonparametric Approach* (Noether, 1976), quoted an introduction written by Herman Chernoff that appeared in the first edition of the book. Chernoff wrote: "What are the advantages of this approach? The basic ideas of inference appear at the beginning. The methods developed are easy to apply and require a minimal amount of computation. The methods are simple in principle; the common sense logic behind them is easy to perceive and to explain… Finally, there are the advantages of the robustness and of the wide applicability of the nonparametric methods discussed."

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

For all of these reasons, Nonparametric Statistics has always been one of my favorite courses to teach. I have enjoyed sharing useful data analysis methods while helping my students better understand the logic of hypothesis testing.

QUESTIONING MY ASSUMPTIONS

The Applied Nonparametric Statistics course at North Carolina State University was one of many applied statistics courses available to undergraduate and graduate students in statistics. I never questioned my belief that the course was an important addition to the curriculum. Only since coming to Meredith College have I started to question the role of the nonparametrics course taught there. Meredith College has a minor in statistics that consists of four statistics courses – Statistics I, Statistics II, a probability and mathematical statistics course, and the nonparametrics course – in addition to Calculus I and II. Given the variety of other applied statistics courses that we could potentially add to our curriculum, I have begun to question whether a course in nonparametric statistics remains a high priority.

One factor that has led me to re-evaluate my nonparametrics course in recent years is the increased interest of many statistics instructors in using bootstrap and randomization (or permutation) methods to teach inference in introductory courses. The theme of the 2011 U.S. Conference on Teaching Statistics was "The Next BIG Thing." A large proportion of conference attendees agreed that the next big thing in teaching statistics was the use of bootstrap and randomization methods to teach inference. Much of Chernoff's description about a nonparametric approach to teaching statistics could have been written more recently about the use of randomization methods to teach statistical inference: "The basic ideas of inference appear at the beginning… The methods are simple in principle; the common sense logic behind them is easy to perceive and to explain…"

When Noether's book was written in the 1970's, randomization tests on the original observations were too computationally intensive for use in introductory statistics courses. At that time, ranking the data made the computations required for randomization-based inference feasible. But new software and applets now exist that make it possible to introduce statistical inference to beginning students using bootstrap and randomization methods on the original observations. In addition, there are now several excellent textbooks that use a randomization approach to teach inference. These developments raise the question of whether we can teach the fundamental concepts of hypothesis testing just as well using randomization methods on the original observations as we can by ranking the data and performing nonparametric tests.

The availability of software for bootstrapping and randomization tests also raises the question of whether practitioners – both consulting statisticians and researchers in other disciplines – will continue to use rank-based inference methods for statistical analysis. Will data analysts who previously used rank-based methods switch to randomization methods on the original observations now that it is more feasible to do so?

A note on terminology: Some authors distinguish between "permutation tests," used in situations in which subjects are randomly sampled from one or more populations, and "randomization tests," used in situations where available subjects are randomly assigned to treatments. Because that distinction will not be crucial in this paper, I will use the terms interchangeably.

QUESTIONING OTHERS

As I pondered these issues (and prepared for my spring 2014 nonparametrics course), I decided to solicit opinions from others about whether they thought we should still teach courses on rank-based distribution-free inference procedures. On December 17, 2013, I sent the abstract for this paper to the Isolated Statisticians list, an e-mail discussion list read primarily by faculty who teach statistics courses at small colleges. I asked list members whether they thought there was now less need for courses on rank-based distribution-free methods, given that many of us teach introductory statistics using randomization methods. Of eight respondents, two saw no reason to teach the standard nonparametric procedures because it is now feasible to apply randomization procedures to the original observations. One stated, "Historically, one of the prime motivators for many nonparametric procedures was to provide an option for doing a test when conditions for standard procedures are not met. That's not so necessary if students know randomization procedures." Another respondent expressed the opinion that if "good software had been around in the 1940s then Wilcoxon, Whitney, et al. would not have developed their methods and no one would be using those methods today."

On the other hand, other respondents mentioned a variety of reasons why they thought there was still a role for a course on rank-based procedures. In the next section, I will enumerate several of their arguments for teaching these methods, along with others drawn from my own experience and reading.

BENEFITS OF TEACHING RANK-BASED INFERENCE METHODS

It is Useful for Students to be Exposed to a Variety of Data Analysis Methods

Students who have studied classical normal theory methods, randomization tests and bootstrapping, and rank-based nonparametric methods will have more tools to choose from when analyzing data.

Rank-based Procedures can be Applied to Data Measured on an Ordinal Scale

Many nonparametric procedures can be applied to count data or ordinal data that cannot be analyzed using normal theory or permutation tests. The sign test for the median can be carried out on data for which the only information available is an indicator of whether each observation is above or below the hypothesized median. A favorite example that I use in my nonparametrics class involves two-sample data on the stomach fullness of fish, measured using a "fullness index" with values 0.00, <0.25, 0.25, <0.50, 0.50, >0.50, 0.75, >0.75, 1.00. While these data can be ranked for a Wilcoxon rank sum test, it would not be possible to carry out a normal theory test or a permutation test based on means.

The Distribution-free Property Allows Tabulation of the Distributions of Rank-based Statistics

Because distribution-free statistics have null distributions that do not vary from sample to sample, those distributions can be tabulated or computed relatively easily. In contrast, the null distribution of the test statistic in a permutation test is conditional on the sample and must be computed anew for each set of data. In a list of advantages of nonparametric procedures, Hollander, Wolfe, and Chicken (2014, p. 1) state, "Nonparametric procedures enable the user to obtain exact *P*-values for tests, exact coverage probabilities for confidence intervals, exact experimentwise error rates for multiple comparison procedures, and exact coverage probabilities for confidence bands without relying on assumptions that the underlying populations are normal." Although the development of new software has made it more feasible to carry out permutation tests on the original observations, many data analysts would likely lack the computing expertise required to address the variety of situations mentioned by Hollander, Wolfe, and Chicken.

Rank Tests can have High Power Relative to Normal Theory and Permutation Tests

Rank tests can have high power relative to competing tests for heavy-tailed data or data with outliers. For example, the asymptotic relative efficiency (a.r.e.) of the Wilcoxon rank sum test relative to the two-sample *t*-test can be arbitrarily large. The a.r.e. values for some heavy-tailed distributions are well-known to be 1.5 for the double exponential, 3.0 for the exponential, and ∞ for the Cauchy. For the normal distribution, for which the *t*-test is optimal, the asymptotic relative efficiency is 0.955, and Hodges and Lehmann (1956) showed that the a.r.e. is never less than 0.864.

While permutation tests on the original observations control the Type I error rate, under certain conditions they have the same asymptotic power as their normal theory counterparts. For example, Hoeffding (1952) showed that the permutation *t*-test (the permutation test based on the two-sample *t*-statistic) has the same asymptotic power as the ordinary normal theory *t*-test. As a result, the asymptotic relative efficiency values given above also apply to the Wilcoxon rank sum test relative to the permutation *t*-test. Keller-McNulty and Higgins (1987) conducted a simulation study to compare the small sample power of the Wilcoxon rank sum test to the power of permutation tests based on various test statistics. Their results showed that for heavy-tailed distributions, the Wilcoxon test tended to be more powerful than the permutation *t*-test (Higgins, 2004). Thus, for heavy-tailed distributions, rank-based tests tend to be more powerful than both normal theory tests and permutation tests.

Boos and Stefanski (2013) refer to the "Type II error robustness" of rank-based tests. They state that "for an appropriate data generation model, the permutation method can make any statistic Type I error robust (level α), but because rank tests are a function of the data only through the ranks, the influence of outliers is automatically limited."

Some Rank Tests do not have a Simple Normal Theory or Randomization Counterpart

Rank correlations are useful for detecting nonlinear relationships, and the Mann-Kendall Trend Test can be used to detect monotone, but nonlinear, trends.

The rank-based Jonckheere and Page tests are used to test equality of the medians for three or more populations, for independent and related samples, respectively. Both tests have an ordered alternative hypothesis that specifies an a priori ordering for the medians. There are not widely used normal theory counterparts for these tests.

The Kolmogorov-Smirnov two-sample test is a widely used rank-based test for detecting any kind of difference between two probability distributions.

Nonparametric Tests are Pedagogically Valuable for Discussing the Role of Assumptions in Inference

The distribution-free property of rank-based procedures is an interesting and important concept for students to understand. A comparison of normal-theory and distribution-free procedures clarifies the effect of the population distribution on the probability of Type I error. It is also instructive to understand that the null distribution of the test statistic in a permutation test is conditional on the observed data, while (for data without ties) the null distribution of a rank-based test statistic is unconditional (does not vary from sample to sample). In addition, a discussion of the efficiency of distribution-free procedures relative to their normal-theory counterparts clarifies the impact of the population distribution on the power of tests.

Nonparametric Statistics is an Active Area of Research that Interests Many Statisticians

The *Journal of Nonparametric Statistics* is a publication of the American Statistical Association. It has been published since 1991; there are four issues a year. Its scope includes "rank and other robust and distribution-free procedures." *Statistical Science* published a special issue on Nonparametric Statistics in November 2004 (Volume 19, Number 4). The Section on Nonparametric Statistics of the American Statistical Association was founded in 1999 and maintains a busy schedule of conferences, paper sessions at the Joint Statistical Meetings, and other activities. The Section presents Student Paper Awards and Best Paper Awards for papers appearing in the *Journal of Nonparametric Statistics*. The American Statistical Association established the Gottfried E. Noether Awards in 1999 as a tribute to Gottfried Noether. There are two Noether Awards presented each year, a Senior Scholar Award given to a distinguished senior researcher/teacher in nonparametric statistics and a Young Scholar Award given to an accomplished young researcher. A quick search of the online program for the 2013 Joint Statistical Meetings found 57 papers with the word "rank" either in the title or as a keyword.

Nonparametric Tests are Widely used for Data Analysis

To gather some quick information about the frequency with which certain rank-based nonparametric tests are used for data analysis in the scientific literature, on January 25, 2014, I searched a few widely read journals for references to "Wilcoxon" and "Kruskal-Wallis" between 2000 and 2014. A search of the three journals published by the American Association for the Advancement of Science (*Science*, *Science Signaling*, and *Science Translational Medicine*) found 313 articles published since 2000 that included the word "Wilcoxon" and 158 that included "Kruskal-Wallis." A similar search of *Nature* found 339 articles that included the word "Wilcoxon" and 146 that included "Kruskal-Wallis"; corresponding counts for *Genetics* were 157 for "Wilcoxon" and 88 for "Kruskal-Wallis." Finally, 152 articles in *The Journal of Nutrition* mentioned "Wilcoxon" and 254 mentioned "Kruskal-Wallis." While this is admittedly an idiosyncratic selection of journals and search terms, it is clear that articles abound in the recent scientific literature that report results of rank-based distribution-free tests.

CONCLUSION

So are rank-based inference procedures still important methods of data analysis that we should teach to our students? After reflecting on the many advantages and desirable features of these methods, I think the answer is yes. I certainly applaud the movement toward teaching inference in introductory courses using bootstrapping and randomization tests, and I teach a course myself that uses that approach. Recent advances in software will likely result in increased use of randomization procedures by practitioners. The nonparametric statistics course of the future will likely include bootstrap and randomization procedures and nonparametric density estimation and regression, in addition to rank-based procedures. But the simplicity, applicability, ease of computation, high power, robustness to outliers, and widely-perceived usefulness of distribution-free rank tests should assure them a continued place in our courses and in the toolboxes of applied statisticians.

REFERENCES

Boos, D. D., & Stefanski, L. A. (2013). *Essential statistical inference: Theory and methods*. New York: Springer.

Higgins, J. J. (2004). *Introduction to modern nonparametric statistics*. Pacific Grove, CA: Brooks/Cole – Thomson Learning.

Hodges, J. L., & Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the *t*-test. *The Annals of Mathematical Statistics*, *27*(2), 324-335.

Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *The Annals of Mathematical Statistics*, *23*(2), 169-192.

Hollander, M., Wolfe, D. A., & Chicken, E. (2014). *Nonparametric statistical methods* (3rd edition). New York: Wiley.

Keller-McNulty, S., & Higgins, J. J. (1987). Effect of tail weight and outliers on power and type-I error of robust permutation tests for location. *Communications in Statistics: Simulation and Computation*, *16*(1), 17-36.

Noether, G. (1976). *Introduction to statistics: A nonparametric approach* (2nd edition). Boston: Houghton Mifflin Company.