



October 2011
Volume 49 Number 5
Article Number 5TOT7

[Return to Current Issue](#)

Simple Statistics for Correlating Survey Responses

Robert G. Hollingsworth

Research Entomologist
US Pacific Basin Agricultural Research Center
USDA-ARS
Hilo, Hawaii
robert.hollingsworth@ars.usda.gov

Topaz P. Collins

Research Experience for Undergraduates Intern
University of Hawaii at Hilo
Hilo, Hawaii
topaz@hawaii.edu

Virginia Easton Smith

Assistant Extension Agent
Kealahou, Hawaii
vsmith@hawaii.edu

Scot C. Nelson

Associate Specialist
Komohana Research and Extension Center
Hilo, Hawaii
scot.nelson@hawaii.edu

Abstract: The rank-sum test is a non-parametric hypothesis test that can be used to determine if there is a statistically significant association between categorical survey responses provided for two different survey questions. The use of this test is appropriate even when survey sample size is small. The rank-sum test is most useful when the goal is to determine whether two groups of respondents differ in their average response to a particular survey question for which response categories are logically ordered according to magnitude (e.g., Likert-scale responses).

Introduction

A large percentage of the manuscripts published in *JOE* present survey data. Surveys are frequently used to collect baseline data, document trends, or compare pre- and post-training scores, for example.

Typically, a respondent is asked to indicate his or her level of agreement to a particular statement according to the following Likert scale response variables:

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

When presenting survey results, some authors are content to simply report the percentage or number of responses within each category for each survey question asked. Such a presentation is descriptive in nature, and while these data can be very useful, the analysis does not go deep enough to provide additional perspective for a more meaningful interpretation of the relationship existing between and among response variables.

Others authors are interested in using the responses obtained to draw broader conclusions to explore correlations or other statistical relationships involving survey responses. For example, an Extension worker or researcher might want to know if the answer to a particular question is influenced by age, sex, geographic region in which the respondent resides, or some other demographic factor. Or, if the survey relates to a natural phenomenon, such as the severity of a particular pest problem, the author might want to know if pest severity ratings (analogous to Likert scale responses) are statistically related to specific agronomic practices or farm characteristics.

We conducted our own survey, of sorts, to estimate the percentage of surveys published in *JOE* that used descriptive statistics only. We searched *JOE*'s archives using the search term "survey" and randomly selected 25 surveys published between 2003 and 2009. Of these 25 surveys, nine presented survey responses using descriptive statistics only. In these papers there were no p-values, chi-square or ANOVA tests, t or F statistics, or any other kind of statistical test. Instead, these surveys tabulated responses separately for each question asked, reporting the number, percentage, or proportion of answers in each category.

One might think that low sample sizes might be a factor in the decision not to use statistical tests, due to a justifiable concern that the test used might be inappropriate given small sample sizes. However, in our sample of 25 *JOE* manuscripts, the average number of respondents per survey was 344.8 (SEM = 78) for manuscripts not using statistical tests, and 482.8 (SE = 129) for those that did use statistics. This difference was not statistically significant ($t = 0.75$, $p = 0.23$, $df = 23$, one-tailed t-test assuming equal variances).

Possibly a more common reason authors don't use statistics to evaluate survey responses is because they are unsure how to perform simple statistical analyses of categorical data. Because categorical data are not normally distributed, non-parametric methods are generally more appropriate for analyses than are parametric tests such as the t-test and the F-test, particularly when survey sample sizes are small.

Using Rank-Sum Tests on Likert-Scale Type Data

One very simple non-parametric test that can be used to determine whether survey responses to two different questions are statistically related is the rank-sum test. This test goes by several different names, including the

Mann-Whitney U test, the Wilcoxon-Mann-Whitney test, or the Wilcoxon Rank-Sum test. This test is one of the best known non-parametric tests and is usually included in statistical software packages.

To illustrate, let's assume we send out a survey, receive back 100 survey forms, and want to know if there is a statistical relationship between answers given to survey Question "A" and survey Question "B." Question A asks respondents whether they are "male" or "female." Question B asks a question requiring a Likert-scale response ("strongly disagree", "disagree," etc.). We want to determine whether the response to Question B was influenced by the gender of the respondents.

To carry out the test, we first assign values to the Likert responses. We could assign a value of "1" for "strongly disagree," "2" for "disagree," up to a value of "5" for "strongly agree." Then, if we are carrying out the procedure by hand, we must pool together the values of both samples (males and females), listing them in order of increasing magnitude. Ranks are assigned to each of the 100 responses, from 1 to 100. Tied observations are given the mean rank. For example, imagine there were six "strongly disagree" responses: 5 from males and 1 from a female. These responses would be listed first, because they had the lowest magnitude (a value of "1"). The ranking assigned to each of these tied values would be 3.5 (the mid-point between 1 and 6).

The sum of the ranks for females and males is computed separately, and a formula is used to adjust each for the number of observations in each sample. Then the smaller of these values (called a "U" statistic) is compared to tabular values to determine the significance level. The Mann-Whitney U test is easy to carry out by hand for small datasets ($\hat{=}$ \approx 20 observations per sample) if ties are rare or non-existent. Otherwise, a normal theory approximation must be used to compute the probability value of the U statistic (Sheskin 2000, p. 296), and it becomes more practical in such cases to use a statistical package for analysis.

The hypothesis being tested by the Mann-Whitney U test is, "Do two independent samples represent two populations with different median values?" (Sheskin, 2000, p. 289). In our case, the hypothesis can be re-phrased as "Does the median rank of Likert scores (scores that vary from 1 to 5) differ between males and females? Once the concept behind using the Mann-Whitney U test is grasped, it becomes an easy matter to use this procedure to test for correlations in survey responses associated with any pair of survey questions, provided that the responses are categorical or can be sorted into categories.

An important point is that the two samples (males and females in the above example) can be constructed using any logical method to split survey respondents into two groups, and the categories (corresponding to different Likert responses in the example) can be constructed using any response variable for which it is possible to logically assign ordinal values. Let's suppose that the two survey questions of interest (Questions A & B) each request Likert responses. Before we carry out the rank-sum procedure, we first need to use the response data for one of the questions to create two new categories or groups. For example, one group could be constructed by selecting those who selected "strongly disagree" or "disagree" for Question A (call this "Group 1"), while the other group could be those chose "strongly agree" or "agree" ("Group 2"). Then, using the method shown above for the male/female example, we could determine if these two groups differed significantly in their median response to survey Question B.

In some cases, it is more convenient to divide the respondents into three or more groups before carrying out a rank-sum test on the response variable. For this situation, the Kruskal-Wallis test is the appropriate test to apply. It uses essentially the same procedures as the Mann-Whitney U test, except that rank-sums are computed for three or more groups instead of for two. As you might expect, *JOE* authors are already using these techniques to analyze Likert scale responses (e.g., Nichols, 2004; O'Neill & Xiao, 2006), although details of the statistical methods employed are seldom provided.

References

Nichols, A. (2004). The effect of tenure and promotion policy on evaluation and research in Extension. *Journal of Extension* [On-line], 42(2) Article 2RIB1. Available at: <http://www.joe.org/joe/2004april/rb1.php>

O'Neill, B., & Xiao, J. J. (2006). Financial fitness quiz findings: Strengths, weaknesses, and disconnects. *Journal of Extension* [On-line], 44(1) Article 1RIB5. Available at: <http://www.joe.org/joe/2006february/rb5.php>

Sheskin, D. (2000). *Handbook of parametric and nonparametric statistical procedures*. Second Edition. Boca Raton: Chapman & Hall/CRC.

Copyright © by *Extension Journal, Inc.* ISSN 1077-5315. Articles appearing in the Journal become the property of the Journal. Single copies of articles may be reproduced in electronic or print form for use in educational or training activities. Inclusion of articles in other publications, electronic sources, or systematic large-scale distribution may be done only with prior electronic or written permission of the *Journal Editorial Office*, joe-ed@joe.org.

If you have difficulties viewing or printing this page, please contact [JOE Technical Support](#).