■ RICK HESSE, Feature Editor, Graziadio Graduate School of Business and Management, Pepperdine University

Simplified Procedure for Implementing Nonparametric Tests in Excel

by Robert J. Pavur, University of North Texas, and Kellie B. Keeling, Virginia Tech

The shift to Excel as a primary computer package for statistics and quantitative methods courses has prompted business professors to search for new ways to make teaching with Excel relevant and fun. One statistical area clearly absent from Excel's data analysis tools is nonparametric statistics. We demonstrate how nonparametric procedures can be easily implemented by inputting ranks of the data into Excel's parametric procedures. The purpose here is not to start a philosophical debate on whether nonparametric statistics should be taught, but to present an approach for teaching nonparametric statistics in an undergraduate business statistics course.

Professors hoping to enlighten their students to the world of nonparametric statistics can encounter a number of difficulties. For example, the name "nonparametric statistics" seems esoteric. Perhaps calling it "statistical analysis of nonnormal data" or "statistical inference based on ranks" may sound more relevant to students. Students don't see any connection between the parametric and nonparametric procedures. To make matters worse, many nonparametric procedures have two versions. One version involves a simplified calculation for the case in which no ties or just a few ties exist, and the other version uses a calculation corrected for ties. Even Excel add-ins can vary in which version they use. MINITAB recognized that there is confusion over which version is used and presents both versions in its printouts. Excel does not provide a ranking function that assigns average ranks to tied observations, which can lead to a frustrating experience for instructors implementing even simple nonparametric tests in Excel. This problem of incorrect rankings has been noted in this column before (Hesse, 1998).

Teaching Students to Use Excel to Find the Average Rank for Tied Observations

When ties occur, the recommended procedure in nonparametric statistics is to assign the average of the ranks that would have been recorded if there were no ties. One deficiency in Excel's RANK function is that this function does not provide an option to use this "average" procedure. Statistics textbooks using Excel typically do not explain how to assign average ranks to tied observations and sometimes do not mention that Excel's **RANK** function does not use the "average" procedure.

Excel's **RANK** function uses the following format: **RANK** (***Number***, ***Ref***, ***Order***) where ***Number*** refers to the observation value to be ranked, ***Ref*** denotes a reference to the complete list of observations, and ***Order*** specifies the ranking scheme. A ***1*** or any nonzero number for ***Order*** indicates a ranking from lowest to highest (a rank of ***1*** being assigned to the smallest value) and a "***0***" indicates a ranking from highest to lowest (a rank of ***1*** being assigned to the largest value). Students should first explore these options on a simple set of data such as presented in cells **A2** through **A5** in Figure 1. For example, they discover something peculiar when asked to use Excel's **RANK** function to rank observations from smallest to largest. That is, they should notice that for tied observations, Excel assigns the smallest rank that would have been recorded if there were no ties.

In Figure 1, the formula **RANK**(**A2**,**A$2:A$5**,1) is typed in cell **B2** and dragged to cell **B5**. In a similar fashion, column **C** is calculated. This time the ranking function uses a 0 for the option of creating reverse ranks. We create column **D**

Robert J. Pavur *is a professor of business computer information systems at the University of North Texas. He holds a master and doctorate degree from Texas Tech University. He has papers published in such journals as the* Annals of Operations Research, IEEE Transactions on Reliability, European Journal of Operational Research, Journal of the Operational Research Society, *and* International Journal of Operations and Quantitative Management. *He is a coauthor of the textbook* Introduction to Business Statistics: A Computer Integrated, Data Analysis Approach *published by South-Western Publishers. He has held various officer positions in the Southwest Region of the Decision Sciences Institute and received Southwest Region's Distinguished Service Award in 2000.*

pavur@unt.edu

Kellie B. Keeling *is an assistant professor of business information technology at Virginia Tech. She received her Ph.D. in management science from the University of North Texas. Her research interests include data mining, acquisition of information technology, and the use of technology in the classroom. She has recently published in the* Communications of the ACM *and in* Multivariate Behavioral Research. *She recently received Virginia Tech's university award for excellence in teaching.*

kkeeling@vt.edu

	A	B	C	D	E	F
1	Data	RANK(...,1)	RANK(...,0)	N+1-RANK(...,0)	Average Rank	Simplified Average Rank
2	12	1	4	1	1	1
3	13	2	2	3	2.5	2.5
4	13	2	2	3	2.5	2.5
5	14	4	1	4	4	4

B2: = **RANK(A2,$A$2:$A$5,1)** **C2:** =**RANK(A2,$A$2:$A$5,0)**

D2: = **COUNT($A$2:$A$5)+1-C2** **E2:** =**AVERAGE(B2,D2)**

F2: = 0.5*(**COUNT($A$2:$A$5)+1+RANK(A2,$A$2:$A$5,1)-RANK(A2,$A$2:$A$5,0)**)

Copy all formulas from Row 2 down through Row 5.

Figure 1: Ranking formulas.

which is essentially N + 1 (the total number of observations plus 1) minus the values in column **C**. At this point, students may recognize that in column **D** when ties are present, the value of the largest rank of the ranks that would be assigned to the tied values is used. From here, students can practically guess that to get the average rank for tied observations simply take the average of columns **B** and **D** in Figure 1.

To obtain the average ranks in one column all at once, it is easy to demonstrate by substituting into the formulas in column **E** that the long formula .5 * (4 + 1 + **RANK** (**A2**, **A$2:A$5**,1) – **RANK(A2**,**A$2:A$5**,0)) makes sense. Students now have two ways of properly assigning average ranks to tied observations. One way is to form columns as presented in columns **B** through **E** and another way is to use the long formula displayed in column **F**. The visualization benefit of using the spreadsheet allows students to discover why it is necessary to use this procedure to obtain an average rank.

Perhaps another approach is to first type into a cell of a column the formula **COUNT**(**A$2:A$5**) + 1 – (**RANK**(**A2**, **A$2:A$5**, 1) + **RANK**(**A2**, **A$2:A$5**, 0)) and then drag this formula to the other rows. The **COUNT**(**A$2:A$5**) function provides the number of observations (4 in this case). Students should recognize that this formula is equal to 0 when there is no tie. That is, the rank of an observation plus the reverse rank of an observation is always equal to the number of observations plus 1 when no ties are present. With a little guidance, students can visualize why the formula **RANK**(**A2**, **A$2:A$5**, 1) + .5*(**COUNT** (**A$2:A$5**) + 1 – (**RANK**(**A2**, **A$2:A$5**, 1) + **RANK**(**A2**, **A$2:A$5**, 0))) produces an average rank for tied observations. Here the part .5*(**COUNT**(**A$2:A$5**) + 1 – (**RANK** (**A2**, **A$2:A$5**, 1) + **RANK**(**A2**, **A$2:A$5**, 0)) is considered a correction term when data values are tied and is equal to zero otherwise. For the data value 13 in this example, the correction term is equal to .5 and its rank is 2 plus .5. These formulas allow students to easily calculate average ranks for tied observations without the need for add-ins that use a visual basic program to calculate the average ranks.

Showing Students How to Obtain Nonparametric Statistics by Using Parametric Statistics

Many nonparametric statistics have two versions, one simplified calculation for no ties and a more complicated one with a correction for ties. The formulas with a correction for ties are often not illustrated in undergraduate textbooks. However, they can be easily implemented using Conover and Iman's rank transformation procedure (Conover & Iman, 1981). An additional advantage to this procedure is that students look at the parametric and nonparametric statistics and see a connection instead of thinking of these test procedures as being isolated from one another. Undergraduate statistics textbooks using Excel tend to have add-ins for nonparametric procedures and do not explain the rank transform technique.

Nonparametric statistics such as the Wilcoxon-Mann-Whitney test, the Kruskal-Wallis test, the Wilcoxon signed-ranks test, the Friedman test, and Spearman's correlation coefficient can be obtained using the rank transform procedure. Students simply perform the following steps.

1. Replace the original data by its ranks. Average ranks are used in case of ties.
2. The parametric counterpart of the nonparametric procedure is calculated on the ranks. This test statistic is referred to as the rank transformed parametric statistic.
3. The value of the nonparametric test statistic is calculated from the formula in Conover and Iman (1981) showing the relationship between the rank transformed statistic and the nonparametric procedure.

Next, we demonstrate this procedure with the nonparametric one-way ANOVA.

Obtaining the Kruskal Wallis Statistic by Using the Parametric Procedure: An Example

We provide an example for the calculation of the Kruskal Wallis test. Consider the data presented in Table 1. In this type of problem, students would be asked to stack the data in Excel and find the average ranks. We will adopt the notation KW and F_R to represent the Kruskal Wallis test statistic and the parametric F statistic calculated using the ranks of the data, respectively. The following relationship between these two statistics is found in Conover and Iman (1981) with N representing the total number of observations and k representing the number of groups:

$$KW = \frac{F_R(N-1)}{F_R + (N-k)/(k-1))}.$$

From this formula, students should realize that the value of the Kruskal Wallis test statistic depends on the value of the parametric statistic using the ranks. In fact, if F_R is large, the KW statistic will tend to be large. Thus, these statistics are not isolated concepts. This relationship may help business students to transition from parametric statistics to nonparametric statistics without getting bogged down in new abstract-looking formulas for the nonparametric statistics.

Following the steps of the rank transform procedure, we use the F statistic from Excel's single factor ANOVA procedure on the ranks of the data instead of the data itself. Using the data in Table 1, we obtain an F_R statistic of 4.388. Now the KW can be readily calculated as follows:

$$KW = \frac{4.388(15-1)}{4.388 + (15-3)/(3-1)} = 5.91.$$

The data in Table 1 were admittedly selected with lots of ties to illustrate differences in test procedures. In Table 2, we illustrate that the uncorrected and corrected KW statistic can be quite different. Furthermore, if students were to use Excel's ranks on either of these statistics, we see that the difference is even greater between the two versions of the KW test than if the average rank procedure was used. In fact, the KW statistic on Excel's ranks can be negative. This result should readily illustrate to students that use of these ranks can lead to strange results.

Group 1		Group 2		Group 3	
Data	Ranks	Data	Ranks	Data	Ranks
5	13.0	2	6.5	2	6.5
5	13.0	2	6.5	5	13.0
2	6.5	2	6.5	2	6.5
2	6.5	1	1.5	2	6.5
5	13.0	1	1.5	5	13.0

Table 1: Data for Kruskal Wallis test.

Test Procedure	Test Statistic	Significance Level
KW uncorrected for ties using average ranks	48.1	.090
KW corrected for ties using average ranks	5.91	.052
KW uncorrected for ties using Excel's ranks	–21.97	Student should realize test isn't valid since it is negative.
KW corrected for ties using Excel's ranks	21.89	.0000015

Table 2: Results of statistical test procedures.

Concluding Remarks

Since spreadsheets have become ubiquitous in the business world, practitioners should be aware of Excel's statistical capabilities. The techniques illustrated here allow students to readily assign average ranks to tied observations and to obtain nonparametric statistics without the use of add-ins. In addition, students may not be aware of the strange results that are possible if Excel's ranks are in fact used in nonparametric procedures. Add-ins do make the implementation of nonparametric statistics easier. This paper is not a case against add-ins, only an illustration of an alternate approach allowing an easy transition from parametric procedures to nonparametric techniques. This approach allows students to see a connection between the parametric and nonparametric procedures. While this bridge between nonparametric and parametric procedures has been around for a long time, it seems strange that its use now becomes handy when using a computer package in which nonparametric procedures are not available. ■

Editor's Note: Portions of this paper were presented at the 2000 Southeast Decision Sciences Conference.

References

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, *35*(3), 124-128.

Hesse, R. (1998). "Normal probability plots," *Decision Line*, *29*(1), 17-19.

Dr. Rick Hesse
Graziadio Graduate School of Business and Management
Pepperdine University
Malibu, CA 90265
rickhesse@aol.com