Technical Report

IBM

The Friedman Rank Sum Test with Multiple Comparisons: A BASIC Program

James R. Lewis

February 15, 1993

TR54.733

Unclassified

Boca Raton, Florida

Unclassified

The Friedman Rank Sum Test with Multiple Comparisons: A BASIC Program

James R. Lewis

TR54.733

ABSTRACT

It is common for human factors engineers to collect data from non-normal distributions using single-factor within-subjects experimental designs. The appropriate analytical method for this type of data is the Friedman rank sum test. Although many computer statistical packages offer this analytical technique, none of them provide multiple comparisons based on the rank sums. This report documents a BASIC program that accepts either raw data or ranks as input and performs the Friedman test with multiple comparisons.

Introduction

The Friedman rank sum test (Friedman, 1937) is an appropriate method for analyzing data from single-factor within-subjects experimental designs, especially if the underlying distributions are not normal (Bradley, 1976; Hollander and Wolfe, 1973). Relative to its parametric analog, the F-test, the efficiency of the Friedman test is always greater than or equal to 0.576 and can be infinite. Hollander and Wolfe (1973) provided a method for a more sensitive test using a correction for tied ranks. They also provided a method for performing multiple comparisons based on Friedman rank sums.

Human factors engineers often collect data that is not likely to have an underlying normal distribution. For example, a human factors engineer could design an experiment to investigate the perceived legibility of different computer displays. Participants in the study would read different text samples on the displays, and would rank the displays along the dimension of perceived legibility. If a participant was unable to determine a difference among a subset of the displays, then those displays would split (take the average of) that subset of ranks for that participant (Bradley, 1976).

Although many statistical packages offer this analysis, no statistical package provides multiple comparisons based on Friedman rank sums. The purpose of this report is to document a BASIC program that analyzes within-subjects rank data with multiple comparisons.

The Program

This IBM[1] BASIC program evaluates either rank data or raw data (by conversion to ranks) in a two-way layout with the Friedman test, Hollander and Wolfe's (1973) correction for tied ranks, and multiple comparisons based on Friedman rank sums (Hollander and Wolfe, 1973). The input is a matrix of data with experimental conditions for columns and participants for rows. The program's output is a file that contains the rank totals, rank averages, Friedman's S statistic, the large sample approximation (using χ^2), a table that summarizes the number of ranks by conditions, and a table that lists the multiple comparisons. The program can accept up to 15 experimental conditions for up to 100 participants. The following examples illustrate the operation of the program.

Example 1: Rank Data as Input

Suppose a human factors engineer has conducted an experiment concerning the perceived legibility of three different displays. Eight participants have viewed and ranked the displays in order of preference, with the results shown in Table 1.

---

[1] IBM is a registered trademark of the International Business Machines Corporation.

First, use an ASCII editor to prepare a file that contains the data, with rows for participants and columns for experimental conditions (in this example, displays). Separate the ranks with commas, as shown in Figure 1. In this example, the human factors engineer saved the file as DISPLAY.DAT.

Table 1. Data for Hypothetical Display Preference Study: No Tied Ranks

Participant	Display 1	Display 2	Display 3
1	3	1	2
2	1	2	3
3	2	1	3
4	3	1	2
5	1	2	3
6	2	1	3
7	2	1	3
8	1	2	3

```
3,1,2
1,2,3
2,1,3
3,1,2
1,2,3
2,1,3
2,1,3
1,2,3
```

Figure 1. Sample Input File for Example 1 (DISPLAY.DAT)

Issue the command FRIEDMAN to start the program. The program requests the following information:

Input filename: The name of the file that contains the rank data. In this example, the input file name is DISPLAY.DAT.

Output filename: The name of the file that will contain the results of the analysis. In this example, the output file name is DISPLAY.OUT.

Number of levels of the independent variable: This is the same as the number of columns, which in this example is 3.

Number of participants: This is the same as the number of rows, which in this example is 8.

After you provide this information, the program processes the data; then it displays the output on the screen and writes it to the output file. Figure 2 shows the output file for this example.

In this example, the human factors engineer can reject the null hypothesis of no difference among the displays ($\chi^2(2)=7.75$, $p=.0208$). The multiple comparisons show that participants consistently ranked the second display more favorably than the third display. Neither the difference between rankings for the first and second displays nor that between the first and third displays was significant.

Example 2: Raw Data as Input

Lewis (1991, p. 1314) provided a hypothetical example for a rank-based method of comparing the usability of competing products. With this method, columns were products and rows were the average results for different tasks that users performed with the products. Table 2 shows the sample data for user-satisfaction ratings with the products. Because human factors engineers often collect user-satisfaction ratings with multipoint rating scales, it is likely that several ties will occur when this type of data is converted to ranks. In addition to illustrating raw-data input, this example also illustrates the treatment of tied ranks. The input file for this example (USERRAW.DAT) is in Figure 3.

Figure 4 shows the results of this analysis. In this example, three of five rows contained tied data (split ranks). The improvement in sensitivity from the test uncorrected for ties ($\chi^2(2)=4.9$, $p=.0863$) to the test with correction for ties ($\chi^2(2)=5.76$, $p=.0560$) is evident in this example. In this case, only the difference between the second and third products approached significance.

Rank averages and totals for file: DISPLAY.DAT

Level	Rank average	Rank total
1	1.88	15
2	1.38	11
3	2.75	22

Friedman's S = 62

CHI(2)= 7.75, p=0.0208 (No correction for ties.)

No ties in this data set.

==

Table of rank frequencies for file: DISPLAY.DAT

Conditions Rank	1	2	3
1.0	3	5	0
1.5	0	0	0
2.0	3	3	2
2.5	0	0	0
3.0	2	0	6
Rank Tot	15	11	22
Rank Ave	1.9	1.4	2.8

==

Multiple Comparisons Based on Friedman Rank Sums

Condition A	Condition B	Absolute Difference	Probability Level
1 : 15	2 : 11	4	>.20
1 : 15	3 : 22	7	<.20
2 : 11	3 : 22	11	<.025

Figure 2. Sample Output File for Example 1 (DISPLAY.OUT)

Table 2. Data for User Satisfaction Competitive Evaluation: Several Tied Ranks

Task	Product 1	Product 2	Product 3
1	3.5	2.0	2.0
2	2.2	2.0	3.0
3	1.3	1.3	2.0
4	1.0	1.0	1.7
5	.7	1.5	1.9

```
3.5,2.0,2.0
2.2,2.0,3.0
1.3,1.3,2.0
1.0,1.0,1.7
0.7,1.5,1.9
```

Figure 3. Sample Input File for Example 3 (USERRAW.DAT)

Rank averages and totals for file: USERRAW.DAT

Level	Rank average	Rank total
1	2.00	10
2	1.30	7
3	2.70	14

Friedman's S = 24.5

CHI(2)= 4.90, p=0.0863 (No correction for ties.)

CHI(2)= 5.76, p=0.0560 (With correction for ties.)

==

Table of rank frequencies for file: USERRAW.DAT

Conditions Rank	1	2	3
1.0	0	2	0
1.5	2	3	1
2.0	2	0	0
2.5	0	0	0
3.0	1	0	4
Rank Tot	10	7	14
Rank Ave	2.0	1.3	2.7

==

Multiple Comparisons Based on Friedman Rank Sums

Condition A	Condition B	Absolute Difference	Probability Level
1 : 10	2 : 6.5	3.5	>.20
1 : 10	3 : 13.5	3.5	>.20
2 : 6.5	3 : 13.5	7	<.10

Figure 4. Sample Output File for Example 3 (USERRAW.OUT)

Discussion

The Friedman test is appropriate for analyzing much of the data that human factors engineers collect with within-subjects designs. (For between-subjects designs, see Lewis, 1993). If the data is in rank form, the Friedman test is the only appropriate method. If the data is not already in rank form, the program will convert the raw data to ranks. The correction for tied ranks (a common occurrence in many situations) can substantially improve the sensitivity of the test, depending on the number of ties. In some cases, this correction could make the difference between rejecting or failing to reject the null hypothesis. Providing multiple comparisons allows the researcher to understand better the sources of the significance of the overall Friedman test.

How to Obtain the Program

(561) 615-4684

You can request a copy of the program by calling the author at tie-line ~~443-1066,~~ or by sending a note to JR~~LEWIS at BCR~~VM1.

jimlewis@us.ibm.com

References

Bradley, J. V. (1973). Probability; decision; statistics. Englewood Cliffs, NJ: Prentice-Hall.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, 32, 675-701.

Hollander, M. and Wolfe, D. A. (1973). Nonparametric statistical methods. New York, NY: John Wiley and Sons.

Lewis, J. R. (1991). A rank-based method for the usability comparison of competing products. In Proceedings of the Human Factors Society 35th Annual Meeting (pp. 1312-1316). Santa Monica, CA: Human Factors Society.

Lewis, J. R. (1993). The Kruskal-Wallis rank sum test with multiple comparisons: A BASIC program (Tech. Report 54.732). Boca Raton, FL: International Business Machines Corp.