11-5: FRIEDMAN RANK TEST FOR DIFFERENCES IN *C* MEDIANS

It sometimes happens that the data collected are only in rank form within each block or normality cannot be assumed in the randomized block design. In these situations, a nonparametric approach called the **Friedman rank test** can be utilized.

The Friedman rank test is primarily used to test whether c sample groups (i.e., the treatment levels) have been selected from populations having equal medians. That is, you test

$$H_0: M_{.1} = M_{.2} = \cdots = M_{.c}$$

against the alternative

$$H_1: \text{Not all } M_{.j} \text{ are equal (where } j = 1, 2, \ldots, c)$$

To develop the test you first replace the data by their ranks on a block-to-block basis. In each of the r independent blocks, the c observations are replaced by their corresponding ranks such that rank 1 is given to the smallest observation in the block and rank c to the largest. If any values in a block are tied, they are assigned the average of the ranks that they would otherwise have been given. Thus, R_{ij} is the rank (from 1 to c) associated with the jth group (where $j = 1, 2, \ldots, c$) in the ith block (where $i = 1, 2, \ldots, r$).

Under the null hypothesis of no differences in the c groups, each ranking within a block is equally likely. There are $c!$ possible ways of ranking within a particular block and $(c!)^r$ possible arrangements of ranks over all r independent blocks. If the null hypothesis is true, there will be no real differences among the average ranks for each group (taken over all r blocks).

From the above, the following test statistic F_R is obtained:

FRIEDMAN RANK TEST FOR DIFFERENCES IN *C* MEDIANS

$$F_R = \frac{12}{rc(c+1)} \sum_{j=1}^{c} R_{.j}^2 - 3r(c+1) \qquad \textbf{(11.28)}$$

where $R_{.j}^2$ is the square of the rank total for group j ($j = 1, 2, \ldots, c$)
r is the number of independent blocks
c is the number of groups or treatment levels

As the number of blocks in the experiment gets large (greater than 5), the test statistic F_R can be approximated by the chi-square distribution with $c - 1$ degrees of freedom. Thus, for any selected level of significance α, the decision rule is to reject the null hypothesis if the computed value of F_R is greater than χ_U^2, the upper-tail critical value for the chi-square distribution having $c - 1$ degrees of freedom as shown in Figure 11.40. That is,

$$\text{Reject } H_0 \text{ if } F_R > \chi_U^2;$$

$$\text{otherwise do not reject } H_0.$$

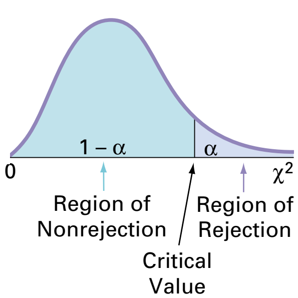

FIGURE 11.40
Determining the rejection region for the Friedman test

The critical values from the chi-square distribution are given in Table E.4.

To illustrate the Friedman rank test for differences in c medians, return to the fast-food-chain study from section 11.2. Recall that the customer service director for the chain designed a randomized block experiment in which 24 investigators are stratified into six blocks of four—based on food-service evaluation experience—and the four members of each block are randomly assigned to evaluate the service at one of the four restaurants owned by the chain.

The results of the experiment are displayed in Table 11.6 along with some summary computations. If the customer service director does not want to make the assumption that the service ratings were normally distributed for each restaurant, the nonparametric Friedman rank test for differences in the four population medians can be used.

The null hypothesis to be tested is that the median service ratings for the four restaurants are equal; the alternative is that at least one of the restaurants differs from the others.

$$H_0: M_{.1} = M_{.2} = M_{.3} = M_{.4}$$

$$H_1: \text{Not all the medians are equal}$$

Table 11.12 provides the 24 service ratings from Table 11.6 along with the ranks assigned within each block.

TABLE 11.12
Converting data to ranks within blocks

FFCHAIN

RESTAURANTS

Blocks of Raters	*A* Rating	*A* Rank	*B* Rating	*B* Rank	*C* Rating	*C* Rank	*D* Rating	*D* Rank
1	70	**2.0**	61	**1.0**	82	**4.0**	74	**3.0**
2	77	**3.0**	75	**1.0**	88	**4.0**	76	**2.0**
3	76	**2.0**	67	**1.0**	90	**4.0**	80	**3.0**
4	80	**3.0**	63	**1.0**	96	**4.0**	76	**2.0**
5	84	**2.5**	66	**1.0**	92	**4.0**	84	**2.5**
6	78	**2.0**	68	**1.0**	98	**4.0**	86	**3.0**
Rank total		**14.5**		**6.0**		**24.0**		**15.5**

From Table 11.12 note the following rank totals for each group:

$$\text{Rank totals: } R_{.1} = 14.5 \quad R_{.2} = 6.0 \quad R_{.3} = 24.0 \quad R_{.4} = 15.5$$

Equation (11.29) provides a check on the rankings.

CHECKING THE RANKINGS

$$R_{.1} + R_{.2} + R_{.3} + R_{.4} = \frac{rc(c+1)}{2} \qquad \textbf{(11.29)}$$

For the fast-food-chain data

$$14.5 + 6 + 24 + 15.5 = \frac{(6)(4)(5)}{2}$$

$$60 = 60$$

Using Equation (11.28),

$$F_R = \frac{12}{rc(c+1)}\sum_{j=1}^{c} R_{.j}^2 - 3r(c+1)$$

$$= \left\{\frac{12}{(6)(4)(5)}[14.5^2 + 6.0^2 + 24.0^2 + 15.5^2]\right\} - (3)(6)(5)$$

$$= \left(\frac{12}{120}\right)(1{,}062.5) - 90 = 16.25$$

Since the computed F_R statistic is greater than 7.815, the upper-tail critical value χ^2_U under the chi-square distribution having $c - 1 = 3$ degrees of freedom (see Table E.4), the null hypothesis is rejected at the $\alpha = 0.05$ level. You conclude that there are significant differences (as perceived by the raters) with respect to the service rendered at the four restaurants. Using the Minitab output of Figure 11.41, observe that since the p-value $= 0.001 < 0.05$, the null hypothesis is rejected.

Note that these are the same conclusions that were made for these data using the randomized block F test in section 11.2.

Friedman Test

```
Friedman test for Rating by Restrat blocked by Raters

S = 16.25  DF = 3  P = 0.001
S = 16.53  DF = 3  P = 0.001 (adjusted for ties)

                       Est    Sum of
Restrat        N    Median     Ranks
A              6     76.87      14.5
B              6     66.50       6.0
C              6     90.38      24.0
D              6     79.25      15.5

Grand median  =     78.25
```

FIGURE 11.41
Minitab output of Friedman rank test for differences in c medians in fast-food-chain study.

Observe that for each restaurant branch, the sample size (N), the estimated median rating, and the rank total (Sum of Ranks) are displayed. An estimate of the overall (Grand) median based on all ratings is also shown. At the top of the Minitab printout appears the Friedman test statistic S (which is equivalent to the statistic F_R), the degrees of freedom (df), and the p-value. If there are ties in the rankings, as is the case in the fast-food-chain study, Minitab provides an adjustment to the test statistic S along with an adjusted p-value that can be used for interpreting results. Note that this adjustment has a minimal impact on these results.

Since you have rejected the null hypothesis and concluded that there is evidence of a significant difference among the restaurant branches with respect to the median ratings, the next step is a simultaneous comparison of all possible pairs of restaurant branches to determine which one or ones differ from the others. As a follow-up to the Friedman rank test, a *post hoc* multiple comparison procedure proposed by Nemenyi (see references 3, 4 and 9) can be used.

To use the Friedman rank test for differences in c medians you make the assumptions as listed in Exhibit 11.2:

◆ EXHIBIT 11.2 ◆

Assumptions of the Friedman Rank Test for Differences in *c* Medians

1. The r blocks are independent so that the measurements in one block have no influence on the measurements in any other block.
2. The underlying random variable of interest is *continuous* (to avoid ties).
3. The observed data constitute at least an ordinal scale of measurement within each of the r blocks.
4. There is no interaction between the r blocks and the c treatment levels.
5. The c populations have the same variability.
6. The c populations have the same shape.

Interestingly, the Friedman procedure still makes less stringent assumptions than does the randomized block F test. To use the Friedman procedure to test for differences in c medians, the measurements need only be ordinal within each of the blocks, and the common population distributions need only be continuous—their common shapes are irrelevant. In fact, if you ignore the last two assumptions, the Friedman rank test still could be used to test the null hypothesis of no differences in the c populations against the general alternative that at least one of the populations differs from at least one of the other populations in some characteristic—be it central tendency, variation, or shape.

On the other hand, to use the F test the level of measurement must be higher than an ordinal scale and you must assume that the c samples are coming from underlying normal populations having equal variances. Both the F test and the Friedman test assume that there is no *interacting effect* between the treatments and the blocks. That is, in the fast-food-chain study you need to assume that any differences between the treatments (the restaurants) are consistent across the entire set of blocks of raters.

For situations involving randomized block designs, when the more stringent assumptions of the F test hold, you should select it over the Friedman test because it will be slightly more powerful in its ability to detect significant treatment effects. However, if the more stringent assumptions cannot be met, the Friedman rank test likely will be more powerful than the F test and you should choose this procedure.

PROBLEMS FOR SECTION 11.5

Learning the Basics

11.65 If the Friedman rank test is used at the 0.10 level of significance when testing for the equality of the medians in six populations, what is the upper-tail critical value χ^2_U from the chi-square distribution?

11.66 From problem 11.65:

a. State the decision rule for testing the null hypothesis that all six groups have equal population medians.

b. What is your statistical decision if the computed value of the test statistic F_R is 11.56?

Applying the Concepts

11.67 A taste-testing experiment has been designed so that four brands of Colombian coffee are to be rated by nine experts. To avoid any carryover effects, the tasting sequence for the four brews is randomly determined for each of the nine expert tasters until a rating on a 7-point scale (1 = extremely unpleasing, 7 = extremely pleasing) is given for each of the following four characteristics: taste, aroma, richness, and acidity. The following table displays the summated ratings—accumulated over all four characteristics. **COFFEE**

BRAND				
Expert	*A*	*B*	*C*	*D*
C.C.	24	26	25	22
S.E.	27	27	26	24
E.G.	19	22	20	16
B.L.	24	27	25	23
C.M.	22	25	22	21
C.N.	26	27	24	24
G.N.	27	26	22	23
R.M.	25	27	24	21
P.V.	22	23	20	19

a. At the 0.05 level of significance, use the Friedman rank test to determine whether there is evidence of a difference in the summated ratings of the four brands of Colombian coffee. What do you conclude?

b. Are there any differences in the results of (a) from those of problem 11.23? Discuss.

11.68 The dean of a well-known business school wants to study the student-faculty evaluation process at his campus since it is used in reappointment, promotion, and tenure decisions. In particular, he is interested in determining the type of educational setting most conducive to higher faculty evaluations from students—MBA courses, advanced undergraduate courses, or required undergraduate courses. Since the faculty's semester workload at this institution is three courses, the dean takes a random sample of 10 faculty from his school that have been assigned one course in each of the three types of educational settings and retrieved their end-of-semester evaluation forms. The following results are mean ratings on a 5-point scale (1 = very poor, 5 = outstanding) to the question: "Compared to other teachers you have had, how would you rate this individual's teaching ability?" Each of the ratings is from classes containing 25–30 students. **RATING**

TYPE OF CLASS			
Faculty Member	*MBA Course*	*Advanced Undergrad*	*Required Undergrad*
L.M.	4.12	4.06	3.38
N.R.	4.87	4.72	4.60
A.C.	3.46	3.49	2.39
J.K.	3.87	3.61	3.23
J.B.	4.04	3.83	3.55
D.B.	2.90	3.23	3.52
W.F.	4.16	4.07	3.68
R.S.	4.19	3.76	3.83
M.L.	4.75	4.39	4.22
V.P.	4.29	4.34	3.67

a. At the 0.05 level of significance, use the Friedman rank test to determine whether there is evidence of a difference in the median ratings based on type of class. What can you conclude?
b. Are there any differences in the results in (a) from those of problem 11.24? Discuss.

11.69 The manager of a nationally known real estate agency has just completed a training session on appraisals for three newly hired agents. To evaluate the effectiveness of his training, the manager wishes to determine whether there is any difference in the appraised values placed on houses by these three different individuals. A sample of 12 houses is selected by the manager, and each agent is assigned the task of placing an appraised value (in thousands of dollars) on the 12 houses. The results are summarized as follows. **REAPPR3**

HOUSE	AGENT 1	AGENT 2	AGENT 3
1	181.0	182.0	183.5
2	179.9	180.0	182.4
3	163.0	161.5	164.1
4	218.0	215.0	217.3
5	213.0	216.5	218.4
6	175.0	175.0	216.1
7	217.9	219.5	220.1
8	151.0	150.0	152.4
9	164.9	165.5	166.1
10	192.5	195.0	197.0
11	225.0	222.7	226.4
12	177.5	178.0	179.7

a. At the 0.05 level of significance, use the Friedman rank test to determine whether there is evidence of a difference in the median appraised value for the three agents. What can you conclude?
b. Are there any differences in the results of (a) from those of problem 11.25? Discuss.

11.70 Philips Semiconductors is a leading European manufacturer of integrated circuits. Integrated circuits are produced on silicon wafers, which are ground to target thickness early in the production process. The wafers are positioned in various locations on a grinder and kept in place through vacuum decompression. One of the goals of process improvement is to reduce the variability in the thickness of the wafers in different positions and in different batches. Data were collected from a sample of 30 batches. In each batch the thickness of the wafers on positions 1 and 2 (outer circle), 18 and 19 (middle circle), and 28 (inner circle) was measured. The results are given in the **CIRCUITS** file.

Source: K. C. B. Roes, and R. J. M. M. Does, " Shewhart-Type charts in nonstandard situations," *Technometrics*, 37, 1995, 15–24.

a. At the 0.05 level of significance, use the Friedman rank test to determine whether there is evidence of a difference in the median thickness of the wafers for the five positions. What can you conclude?
b. Are there any differences in the results of (a) from those of problem 11.26? Discuss.

11.71 The data in the **CONCRETE2** file represent the compressive strength in thousands of pounds per square inch (psi) of 40 samples of concrete taken 2, 7, and 28 days after pouring.

Source: O. Carrillo-Gamboa and R. F. Gunst, " Measurement-error-model collinearities," *Technometrics*, 34, 1992, 454–464.

a. At the 0.05 level of significance, use the Friedman rank test to determine whether there is evidence of a difference in the median compressive strength after 2, 7, and 28 days. What can you conclude?
b. Are there any differences in the results of (a) from those of problem 11.27? Discuss.

Using Minitab for the Friedman Test

To illustrate the use of Minitab for the Friedman test, open the **FFCHAIN.MTW** worksheet. Select **Stat** | **Nonparametrics** | **Friedman**. In the Friedman dialog box (see Figure 11.42), enter **C3** or **'Rating'** in the Response: edit box, **C2** or **'Restrat'** in the Treatment: edit box, and **C1** or **'Raters'** in the Blocks: edit box. Click the **OK** button.

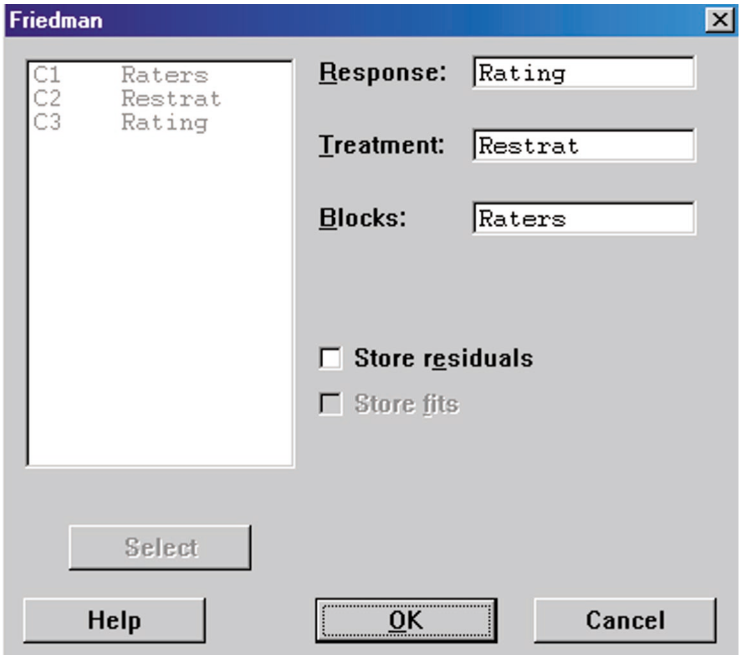

FIGURE 11.42
Minitab Friedman dialog box