## Friedman’s test

The Friedman test extends the previously mentioned Wilcoxon signed ranks test to more than two repeated values at more than two time points. Alternatively, to more than two matched groups where the individuals of each group are randomly assigned to a group.

The test examines the ranks at the different time points or matched pairs and tests whether the continuous underlying distribution of the variables is the same. It is the non-parametric equivalent of the repeated measures ANOVA.

### Examples from the Literature

The Friedman test is used by Kraemer et al in their paper: “Time-of-day variations of indicators of attention: performance, physiologic parameters, and self-assessment of sleepiness” (Biol Psychiatry 2000 Dec 1; 48(11):1069-80). The objective of this study was to analyse time-of-day variations of different indicators of attention and their interrelations. Time-of-day variations were tested non-parametrically with Friedman's test for repeated measurements.

Gustafson et al also use the Friedman test in their paper: “Effects of 4 hand-drying methods for removing bacteria from washed hands: a randomised trial” (Mayo Clin Proc 2000 Jul; 75(7): 705-8). The objective of this study was to evaluate the effects of 4 different drying methods to remove bacteria from washed hands. The Friedman test was used to show that there was no significant difference in the efficiency of the four methods at removing bacteria.

### Null Hypothesis

H0: There is no difference in median between the groups being tested.

HA: There is at least one difference in median between the groups.

This is a non-directional alternative hypothesis. If the alternative hypothesis is to be directional a more powerful way of analysing the data would be to carry out planned comparisons, with the appropriate correction for multiple testing. Alternatively, overall tests of significance followed by post hoc tests can be used.

### Assumptions

1. The data to be analysed are continuous and at least at the ordinal level of measurement.
2. The data from a randomly selected sample are either multiple observations from a single sample across more than two time periods or conditions. Otherwise, the data are blocks of matched subjects in which the subjects from a given block are each randomly assigned to one of the three or more conditions.
3. The subjects or blocks of subjects are independent; that is, the results within one block do not have an influence on the results within the other blocks.

### Method

1. Construct the null and alternative hypotheses.
2. Construct a two-way table with *N* (the number of subjects or matched sets of subjects) rows and *k* (the number of conditions or data collection periods) columns.
3. Rank each row from lowest to highest and sum ranks in each column.
4. If the null hypothesis is not true then the sum of the columns will vary from column to column. The Friedman test examines the extent to which these column sums vary from what is expected using the following formula:



where,

*Rj* = the sum of the ranks for column *j*

*N* = the number of subjects

*k* = the number of periods or conditions.

1. Look *Fr* up in tables of Friedman’s distribution.
2. Reject the null hypothesis in favour of the alternative hypothesis if the *Fr* value is greater than (or equal to) the value in the tables.

Note: If *N* and *k* are sufficiently large, then *Fr* can be compared to a *χ2* distribution on

*k-1* degrees of freedom.

### Example

The data for this example is taken from Rubin and Peter’s paper. The Friedman test will be used to study whether or not hydralazine would relieve high blood pressure in the lungs.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Person | Before | | 48 hrs after | | 6 months after | |
| Units | Rank | Units | Rank | Units | Rank |
| 1 | 22.2 | 3 | 5.4 | 1 | 10.6 | 2 |
| 2 | 17.0 | 3 | 6.3 | 2 | 6.2 | 1 |
| 3 | 14.1 | 3 | 8.5 | 1 | 9.3 | 2 |
| 4 | 17.0 | 3 | 10.7 | 1 | 12.3 | 2 |
| Rank sum | - | 12 | - | 5 | - | 7 |

**Table 7.1** Total pulmonary resistance before and after hydralazine

.

As 6.5 is the same as the value in the table there is sufficient evidence to reject the null hypothesis and conclude that at least one group is different from the others.

### Post hoc testing

#### Method 1

If the Friedman test shows that there is a difference in medians in the groups it is possible to carry out post hoc testing to see which groups there is actually a difference between. This is done by comparing average ranks in all the pairs or comparing to baseline. The null hypothesis that there is no difference in mean ranks between the pairs will be rejected if the absolute value of these differences is greater than a specified critical value. If the following condition holds, the null hypothesis will be rejected:



where,

 = the mean rank in period or condition *i*

= the mean rank in period or condition *j*

*Zα’*= the critical z value for α’

*α’ = α/[k(k-1)]*

*k* = the number of periods or conditions

*N* = the number of subjects

In the example above, the average ranks for the three time points are 3, 1.25 and 1.75. Since *k*= 3 and the critical value of the *z*-statistic is a *z* for which ′ = 0.05/3(2) = 0.0083. Looking this value up in the normal tables gives *z* = 2.39. The critical value



The absolute values of the three comparisons are:



The comparison between before treatment and 48 hours after treatment is the only one that is greater than the critical value of 1.68. Therefore, we can conclude that according to the post hoc approach, hydralazine only relieves high blood pressure in the lungs 48 hours after the treatment. This effect was not maintained 6 months after treatment.

#### Method 2

It is also possible to use the Wilcoxon ranked sign tests for post hoc testing. The procedure is carried out in the same way as described before. However, the Bonferroni correction must be applied to allow for multiple testing. That is the critical value of *α* becomes *α′=α*/*k*, where *k* is the number of tests to be carried out and α is the original significance level. The value of α′ is the one looked at in the table or that the output *p*-value is compared against.

### Presentation of the Results

The results of the Friedman test could be reported in the following way:

The results of the Friedman test indicate that there is a significant difference in median total pulmonary resistance across the three time periods. Therefore, we can conclude that hydralazine alters total pulmonary resistance (p=0.042).

Post hoc analyses with the adjustment of the two-tailed level to 0.0083 indicated that there were decreases in total pulmonary resistance from before treatment (Md=17.0) to 48 hours after treatment (Md= 7.4). No other significant differences were found.

**Note:** That post hoc testing was carried out here on a very small sample size as an illustration, in reality post hoc testing would not be carried out on such a small sample size.

### Advantages and Limitations

The Friedman test is very versatile and can be used with randomised block designs and multiple observations of a single sample. It is useful when the dependent variable is skewed.

There are some drawbacks however, it is possible for the medians not to change and there still to be significant differences between groups. Although it is often referred to as the Friedman two-way ANOVA by ranks, it is restricted to within group comparisons. It is not possible to test between group comparisons. This is a major disadvantage in clinical research as it is not possible to make experimental-control group comparisons. Each group can be analysed separately and compare their results. However, it is not possible to test a group and time interaction with independent groups.

### Summary

Friedman’s test, tests the null hypothesis that k related variables come from the same population. For each case, the *k* variables are ranked from 1 to *k*; the test statistic is based on these ranks.

After establishing a difference between one of the variables, post hoc testing can be carried out to decide which of the variables are actually different. An appropriate method for allowing for multiple testing must be carried out.