Chapter 34
Non-parametric Tests for Three or More Samples (Friedman and Kruskal-Wallis)

1 General Purpose

The Friedman test is used for comparing three or more repeated measures that are not normally distributed, and is a non-parametric test, and extension of the Wilcoxon signed rank test (Chap. 6). The Kruskal-Wallis test compares multiple groups that are unpaired and not normally distributed, and is also a non-parametric test, and extension of the Mann–Whitney test (Chap. 7). This chapter assesses both methodologies, that are adequate for non-normal data, but can also be used with normal data.

2 Schematic Overview of Type of Data File

Outcome from treat 1	From treat 2	From treat 3
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.

treat = treatment modality

T.J. Cleophas, A.H. Zwinderman, *Clinical Data Analysis on a Pocket Calculator*,
DOI 10.1007/978-3-319-27104-0_34

3 Primary Scientific Question

How do we use the Friedman and Kruskal – Wallis tests for, respectively, testing repeated measures in one group, and single measures in multiple groups, if the assumption of normality is doubtful.

4 Friedman Test for Paired Observations

The underneath data are paired comparisons to test effect of 2 dosages of a sleeping drug versus placebo on hours of sleep

	Hours of sleep					
Patient	dose 1	dose 2	placebo	dose 1	dose 2	placebo
	(hours)			(ranks)		
1	6.1	6.8	5.2	2	3	1
2	7.0	7.0	7.9	1.5	1.5	3
3.	8.2	9.0	3.9	2	3	1
4.	7.6	7.8	4.7	2	3	1
5.	6.5	6.6	5.3	2	3	1
6.	8.4	8.0	5.4	3	2	1
7.	6.9	7.3	4.2	2	3	1
8.	6.7	7.0	6.1	2	3	1
9.	7.4	7.5	3.8	2	3	1
10.	5.8	5.8	6.3	1.5	1.5	3

The data are ranked for each patient in ascending order of hours of sleep. If the hours are equal, then an average ranknumber is given. Then, for each treatment the squared ranksum is calculated: for dose 1 it equals $(2+1.5+2+2+2+3+2+2+2+1.5)^2 = 400$, for dose 2 it is 676, for placebo it is 196. The following equation is used:

$$\text{chi-square} = \frac{12}{\text{nk}\ (\text{k}+1)}\left(\text{ranksum}_{\text{dose1}}{}^{2} + \text{ranksum}_{\text{dose2}}{}^{2} + \text{ranksum}_{\text{placebo}}{}^{2}\right) - 3\text{n}(\text{k}+1),$$

where n = the number of patients and k = the number of treatments.

The chi-square value as calculated is 7.2. The degrees of freedom is 3 – 1 = 2.

The underneath chi-square table has an upper row with areas under the curve, a left-end column with degrees of freedom, and a whole lot of chi-square values. It shows that, for 2 degrees of freedom, a chi-square > 5.991 is required to reject the null-hypothesis of no effect at $p < 0.05$.

Chi-squared distribution

	Two-tailed *P*-value			
df	0.10	0.05	0.01	0.001
1	2.706	3.841	6.635	10.827
2	4.605	5.991	9.210	13.815
3	6.251	7.851	11.345	16.266
4	7.779	9.488	13.277	18.466
5	9.236	11.070	15.086	20.515
6	10.645	12.592	16.812	22.457
7	12.017	14.067	18.475	24.321
8	13.362	15.507	20.090	26.124
9	14.684	16.919	21.666	27.877
10	15.987	18.307	23.209	29.588
11	17.275	19.675	24.725	31.264
12	18.549	21.026	26.217	32.909
13	19.812	22.362	27.688	34.527
14	21.064	23.685	29.141	36.124
15	22.307	24.996	30.578	37.698
16	23.542	26.296	32.000	39.252
17	24.769	27.587	33.409	40.791
18	25.989	28.869	34.805	42.312
19	27.204	30.144	36.191	43.819
20	28.412	31.410	37.566	45.314
21	29.615	32.671	38.932	46.796
22	30.813	33.924	40.289	48.268
23	32.007	35.172	41.638	49.728
24	33.196	36.415	42.980	51.179
25	34.382	37.652	44.314	52.619
26	35.536	38.885	45.642	54.051
27	36.741	40.113	46.963	55.475
28	37.916	41.337	48.278	56.892
29	39.087	42.557	49.588	58.301
30	40.256	43.773	50.892	59.702
40	51.805	55.758	63.691	73.403
50	63.167	67.505	76.154	86.660
60	74.397	79.082	88.379	99.608
70	85.527	90.531	100.43	112.32
80	96.578	101.88	112.33	124.84
90	107.57	113.15	124.12	137.21
100	118.50	124.34	135.81	149.45

Post-hoc subgroups analyses (using Wilcoxon's tests) are required to find out exactly where the difference is situated, between group 1 and 2, between group 1 and 3, or between group 2 and 3 or between two or more groups. The subject of post-hoc testing will be further discussed in the Chaps. 18, 19, and 20.

5 Kruskal-Wallis Test for Unpaired Observations

The underneath data show three-samples of patients treated with placebo or 2 different NSAIDs (non steroidal anti-inflammatory drugs). The outcome variable is the fall in plasma globulin concentration (g/l). Group 1 patients are printed in italics, group 2 in normal standard letters, and group 3 in fat prints.

Globulin concentration (g/l)	ranknumber
−17	***1***
−16	***2***
−5	***3***
−3	***4***
−2	***5***
16	***6***
18	***7***
26	8
27	**9**
28	***10.5***
28	**10.5**
29	12
30	14
30	***14***
30	***14***
31	16
32	17
33	18
34	**19**
35	20
36	21
38	22.5
38	**22.5**
39	**24.5**
39	**24.5**
40	**26**
41	27
42	**28**
45	**29.5**
45	**29.5**

Three groups of patients with rheumatoid arthritis are treated with a placebo or one of two different NSAIDS. The fall in plasma globulin (g/l) is used to estimate the effect of treatments. First, we will give a ranknumber to every patient dependent on his/her magnitude of fall. If two or three patients have the same fall, they are given an average ranknumber. Then, we calculate the sum of the ranks for the three groups. For group 1 this amounts to $1 + 2 + 3 + 4 + 5 + 6 + 7 + 10.5 + 14 + 14 = 66.5$, for group 2 to 175.5, group 3 to 488.5. Then we use the equation:

$$\text{chi-square} = \frac{12}{30\,(30-1)}\left(\frac{\text{ranksum}_{\text{group1}}{}^2}{10} + \frac{\text{ranksum}_{\text{group2}}{}^2}{10} + \frac{\text{ranksum}_{\text{group3}}{}^2}{10}\right) - 3(30-1),$$

where the number 30 equals all values, 10 the patient number per group.

The chi-square-value equals 7744.3. This value is very large, indicating that the null-hypothesis of no difference in the data can be rejected. In this example the calculated chi-square value is much larger than the rejection chi-square for (3–1) degrees of freedom and, therefore, we conclude that there is a significant difference between the three treatments at $p < 0.0001$ (see the above chi-square table).

Post-hoc subgroup analyses (using Man-Whitney tests) are required to find out exactly where the difference is situated, between group 1 and 2, between group 1 and 3, or between group 2 and 3 or between two or more groups. The subject of post-hoc testing have been discussed in the Chaps. 18, 19, and 20.

6 Conclusion

For the analysis of efficacy data we test null-hypotheses. The t-test (Chaps. 6 and 7) is appropriate for two parallel-groups or two paired samples. Analysis of variance (ANOVA) (Chaps. 19 and 20) is appropriate for analyzing more than two groups / treatments. For data that do not follow a normal frequency distribution, non-parametric tests are available: for paired data the Wilcoxon signed rank (Chap. 6) or Friedman tests (current chapter), for unpaired data the Mann–Whitney test (Chap. 7) or Kruskal-Wallis tests (this chapter) are adequate.

7 Note

More background, theoretical and mathematical information of Friedman and Kruskal-Wallis tests are given in Statistics applied to clinical studies 5th edition, Chap. 2, Springer Heidelberg Germany, 2012, from the same authors.