SPSS

stats*t*utor

community project

encouraging academics to share statistics support resources

All stcp resources are released under a Creative Commons licence

stcp-marshall-FriedmanS

The following resources are associated:
The SPSS dataset 'Video', Repeated measures in ANOVA resource.

Friedman test in SPSS
(Non-parametric equivalent to repeated measures ANOVA)

Dependent variable: Continuous (scale) but not normally distributed or ordinal

Independent variable: Categorical (Time/Condition)

Common Applications: Used when several measurements of the same dependent variable are taken at different time points or under different conditions for each subject and the assumptions of repeated measures ANOVA have not been met. It can also be used to compare ranked outcomes.

Data: The dataset 'Video' contains some results from a study comparing videos made to aid understanding of a particular medical condition. Participants watched three videos (A, B, C) and one product demonstration (D) and were asked several Likert style questions about each. These were summed to give an overall score for each e.g. TotalAGen below is the total score of the ordinal questions for video A.

TOTALA_GEN	TOTALB_DOC	TOTALC_OLD	TOTALD_DEMO		RANKA_GEN	RANKB_DOC	RANKC_OLD	RANKD_DEMO
25	23	13	22		4	3	1	2
20	17	14	23		2	1	1	4
25	22	*Friedman ranks each participants responses*			2	1	3
24	22	23	25		1	2	4

The Friedman test ranks each person's score from lowest to highest (as if participants had been asked to rank the methods from least favourite to favourite) and bases the test on the sum of ranks for each column. For example, person 1 gave C the lowest Total score of 13 and A the highest so SPSS would rank these as 1 and 4 respectively. As the raw data is ranked to carry out the test, the Friedman test can also be used for data which is already ranked e.g. the ranked example columns RANKA - RANKD. There should be one column per condition/ time point being compared containing the score or rank for that condition. If the Friedman test is significant, post hoc tests to locate which pairs are different are needed.

Research question: Which method is best for understanding the medical condition?

The dependent variable is the overall score and this is recorded in a separate variable for each method. Histograms showed that the residuals obtained from a repeated measures ANOVA were

skewed so the assumptions for carrying out a repeated measures ANOVA are not met and the Friedman test is more appropriate.

Summary Statistics: As we are carrying out a non-parametric test, use medians to compare the scores for the different methods.

Median

Overall score (video A)	Overall score (video B)	Overall score (video C)	Overall score (demo D)
24	23	14	22

Video C has a much lower median than the others. A box-plot is also useful for assessing differences.

Steps in SPSS

There are two methods in SPSS when carrying out a Friedman test.

<u>Legacy Dialogs</u>: *Analyze → Nonparametric Tests → Legacy Dialogs → K Related Samples*
<u>New:</u> *Analyze → Nonparametric Tests → Related Samples*

The legacy method is easier to use but the new method has options for more tests and runs the Dunn-Bonferroni post hoc tests automatically to look for differences between pairs if the main test is significant. If you use the legacy dialogs method, you will need to run Wilcoxon tests for each pair of variables and then make a Bonferroni adjustment (multiply p-values from the Wilcoxon tests by the number of Wilcoxon tests being carried out).

The new method will only run the tests if all the variables to be entered are classified as scale in SPSS (have a ruler next to them). Even though the Friedman test is suitable for testing differences between ordinal variables, they need to be classified as scale to run. Before running the analysis, check that the four variables of interest are classified as scale in variable View.

To use the new method, go to *Analyse → Nonparametric tests → Related samples.* There are three tabs in the dialogue box. Click on the **Fields** tab and move the four variables of interest from the *Fields* to the *Test Fields* box.

SPSS will automatically apply a Friedman test if there are three or more variables and a Wilcoxon Signed rank test if there are only two. After selecting the variables to be included, click **Run** to run the analysis to get the following output.
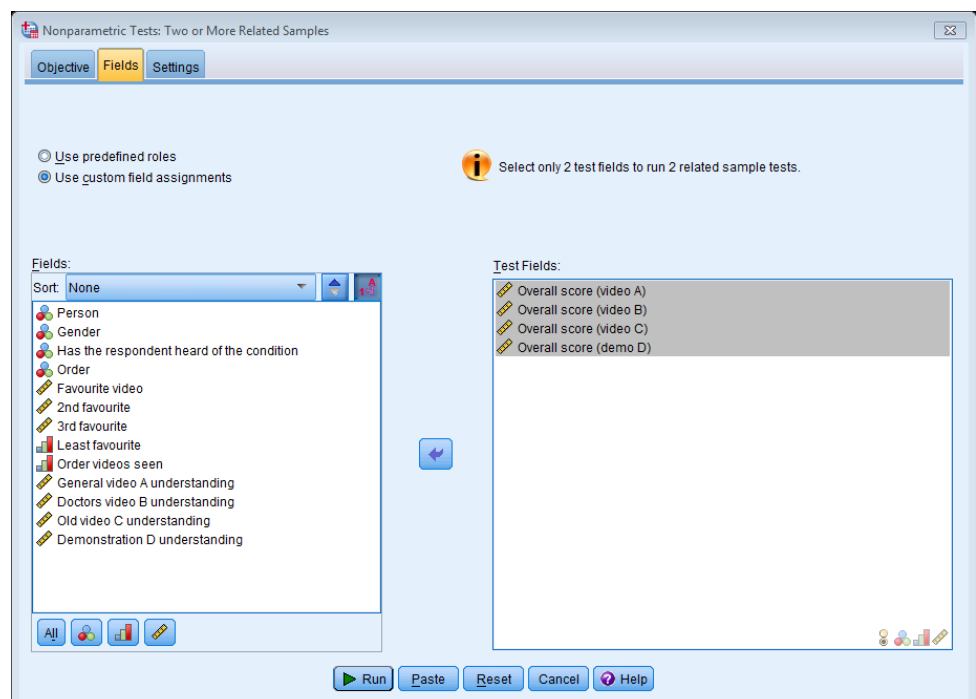
Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distributions of Overall score (video A), Overall score (video B), Overall score (video C) and Overall score (demo D) are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	.000	Reject the null hypothesis.

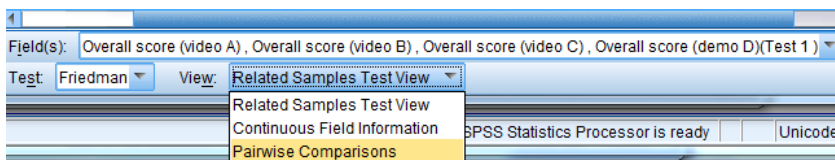Asymptotic significances are displayed. The significance level is .05.

As p < 0.001, there is a significant difference between at least two methods.

Double click on this table in the output window to open the **Model Viewer** window with more output.

Total N	20
Test Statistic	41.372
Degrees of Freedom	3
Asymptotic Sig. (2-sided test)	.000

Here the test statistic (41.372) and degrees of freedom (3) are reported for the Friedman test.

To find out which pairs are different, click on the View box at the bottom of the right hand side of the **Model Viewer** and choose '*Pairwise Comparisons*'. Note: If the Friedman test is not significant, pairwise comparisons will not be reported.

Field(s): Overall score (video A) , Overall score (video B) , Overall score (video C) , Overall score (demo D)(Test 1)

Test: Friedman View: Related Samples Test View

Related Samples Test View
Continuous Field Information   SPSS Statistics Processor is ready   Unicode:
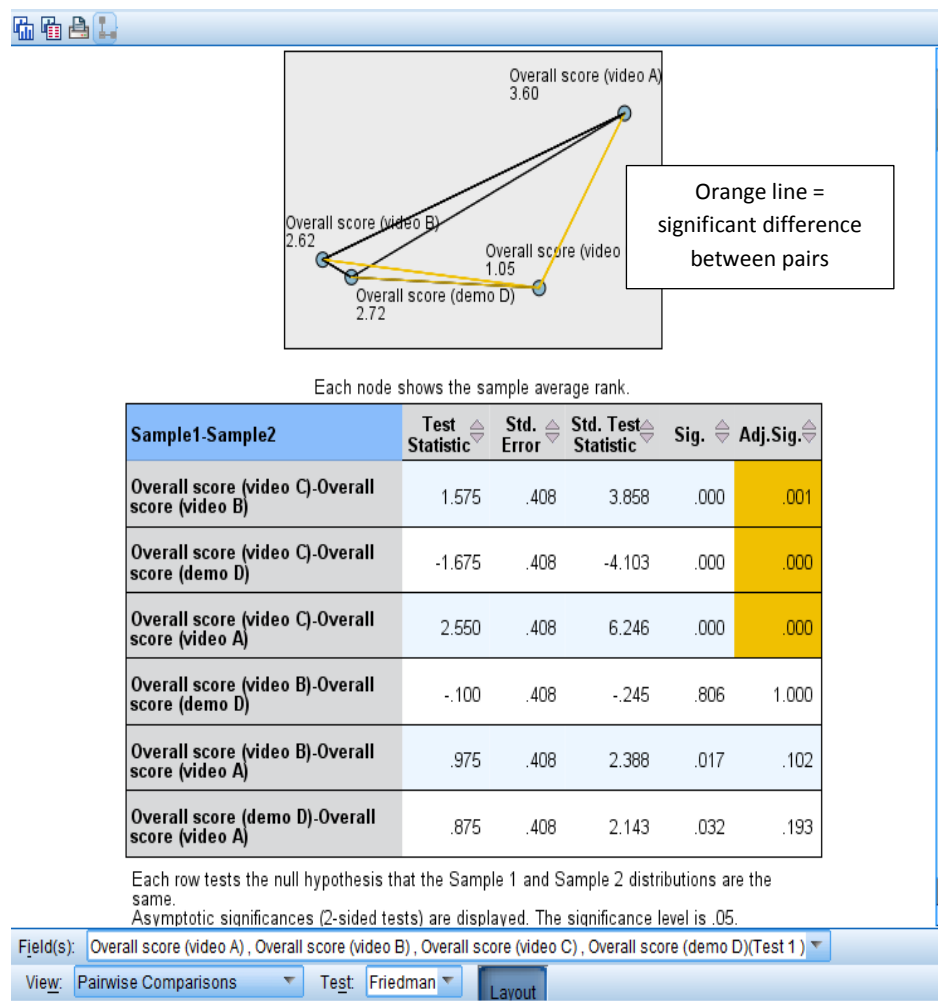Pairwise Comparisons

Post hoc tests

The diagram shows the mean rank for each method and an orange line joins significantly different pairs.
SPSS carries out Dunn's pairwise post hoc tests. The first test statistic, is simply the difference between the mean ranks from the Friedman test for the two groups. However, it has to be converted to a standardised test statistic in order to calculate the p-value (Sig.). Then a Bonferroni correction for multiple testing is applied (the p-values in the 'Sig' column are multiplied by the number of tests being carried out to give the 'Adj. Sig' p-values).

Example: The difference in the mean ranks for C and B is 1.575, the standardised test statistic 3.858 and the p-value to be reported is p=0.001 (Adj. Sig).

Overall score (video A)
3.60

Overall score (video B)
2.62

Overall score (video C)
1.05

Overall score (demo D)
2.72

Orange line = significant difference between pairs

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Overall score (video C)-Overall score (video B)	1.575	.408	3.858	.000	.001
Overall score (video C)-Overall score (demo D)	-1.675	.408	-4.103	.000	.000
Overall score (video C)-Overall score (video A)	2.550	.408	6.246	.000	.000
Overall score (video B)-Overall score (demo D)	-.100	.408	-.245	.806	1.000
Overall score (video B)-Overall score (video A)	.975	.408	2.388	.017	.102
Overall score (demo D)-Overall score (video A)	.875	.408	2.143	.032	.193

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Field(s): Overall score (video A) , Overall score (video B) , Overall score (video C) , Overall score (demo D)(Test 1)

View: Pairwise Comparisons   Test: Friedman   Layout

Alternative post hoc tests

The automatic Dunn-Bonferroni takes into account the fact that there are multiple groups and the mean ranks from the Friedman. However, some people prefer using pairwise Wilcoxon tests so that the number of other measurements is not taken into account and the result slightly less conservative. To do this, go to *Analyze → NonparametricTests → Related samples* and follow the steps in the 'Wilcoxon in SPSS' resource for every pair of variables. Then adjust the resulting p-values by multiplying by the number of pairs you are testing (Bonferroni correction).

Reporting the results

A Friedman test was carried out to compare the total understanding scores for the four methods. There was found to be a significant difference between the methods,

$\chi^2(3) = 41.372, \ p < 0.001$ Dunn-Bonferroni post hoc tests were carried out and there were significant differences between the Old video C and the Doctors video B (p = 0.001), the demonstration D (p <0.001) and video A (p<0.001) after Bonferroni adjustments. There were no significant differences between any other methods.

Effect sizes

Effect sizes cannot be calculated directly for a Friedman test but they can be carried out on pairwise comparisons with a Z statistic using the formula $\frac{z}{\sqrt{n}}$ where n is the number of pairs.

Alternatively, Kendall's W (Coefficient of concordance) is a test which looks at agreement between subjects and gives a value which ranges between 0 and 1. A Kendall's W of one indicates that all subjects ranked the four methods in the same way and therefore they were in complete agreement.

To carry out this test, repeat the analysis through *Analyze → Nonparametric Tests → Related Samples* and select 'Customize tests' in the Settings tab and then 'Kendall's coefficient of concordance (k samples).
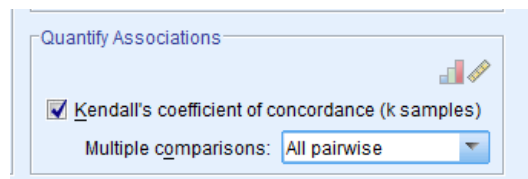
Again, you will need to double click on the table in the output window to get further results. The

Total N	20
Kendall's W	.690
Test Statistic	41.372
Degrees of Freedom	3
Asymptotic Sig. (2-sided test)	.000

Kendall's W is 0.69 which indicates a large effect size as well as good agreement between subjects on the preferable ordering of the methods.

Note: Kendalls uses the Cohen's interpretation guidelines of 0.1 (small effect), 0.3 (moderate effect) and above 0.5 as a strong effect.