Non-parametric tests 2:

Non-parametric tests for comparing three or more groups or conditions:

(a) Kruskal-Wallis test:
Similar to the Mann-Whitney test, except that it enables you to compare *three or more* groups rather than just two.
***Different* subjects are used for each group.**

(b) Friedman's Test:
Similar to the Wilcoxon test, except that you can use it with *three or more* conditions.
Each subject does *all* of the experimental conditions.

One IV, with multiple levels:

Levels can differ

(a) qualitatively/categorically -
e.g. effects of managerial style (laissex-faire, authoritarian, egalitarian) on worker satisfaction.
effects of mood (happy, sad, neutral) on memory.
effects of location (Scotland, England or Wales) on happiness ratings.

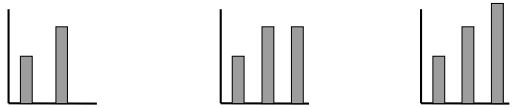
(b) quantitatively -
e.g. effects of age (20 vs 40 vs 60 year olds) on optimism ratings.
effects of study time (1, 5 or 10 minutes) before being tested on recall of faces.
effects of class size on 10 year-olds' literacy.
effects of temperature (60, 100 and 120 deg.) on mood.

Why have experiments with more than two levels of the IV?

(a) Increases generality of the conclusions:
e.g. comparing young (20) and old (70) subjects tells you nothing about the behaviour of intermediate age-groups.

(b) Economy:
Getting subjects is expensive - may as well get as much data as possible from them.

(c) Can look for trends:
What are the effects on performance of increasingly large doses of cannabis (e.g. 100mg, 200mg, 300mg)?



Kruskal-Wallis test, step-by-step:

Does it make any difference to students' comprehension of statistics whether the lectures are given in English, Serbo-Croat or Cantonese?

Group A: lectures in English;
Group B: lectures in Serbo-Croat;
Group C: lectures in Cantonese.

DV: student rating of lecturer's intelligibility on 100-point scale ("0" = "incomprehensible").

***Ratings* - so use a nonparametric test.**

English (raw score)	English (rank)	Serbo-Croat (raw score)	Serbo-Croat (rank)	Cantonese (raw score)	Cantonese (rank)
20	*3.5*	25	*7.5*	19	*1.5*
27	*9*	33	*10*	20	*3.5*
19	*1.5*	35	*11*	25	*7.5*
23	*6*	36	*12*	22	*5*

Step 1:
Rank the scores, ignoring which group they belong to.
Lowest score gets lowest rank.
Tied scores get the average of the ranks they would otherwise have obtained.

Step 2:
Find "Tc", the total of the ranks for each group.

Tc1 (the total for the English group) is 20.

Tc2 (for the Serbo-Croat group) is 40.5.

Tc3 (for the Cantonese group) is 17.5.

Step 3:

Find H.

$$H = \left[\frac{12}{N(N+1)} * \Sigma \frac{Tc^2}{n_c} \right] - 3 * (N+1)$$

N is the total number of subjects;
Tc is the rank total for each group;
nc is the number of subjects in each group.

$$H = \left[\frac{12}{N(N+1)} * \sum \frac{Tc^2}{n_c} \right] - 3 * (N+1)$$

$$\sum \frac{Tc^2}{n_c} = \frac{20^2}{4} + \frac{40.5^2}{4} + \frac{17.5^2}{4} = 586.62$$

$$H = \left[\left(\frac{12}{12*13} \right) * 586.62 \right] - (3*13) = 6.12$$

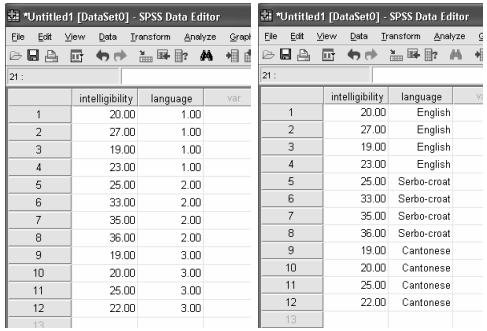
Step 4:
Degrees of freedom are the number of groups minus one. d.f. = 3 - 1 = 2.

Step 5:
H is statistically significant if it is *larger* than the critical value of Chi-Square for these d.f.
Here, H is 6.12. This is larger than 5.99, the critical value of Chi-Square for 2 d.f.

The three groups differ significantly; the language in which statistics is taught does make a difference to the lecturer's intelligibility (H(2) = 6.12, $p < .05$).

NB: the test merely tells you that the three groups *differ*; inspect group medians to decide *how* they differ.

Using SPSS for the Kruskal-Wallis test:



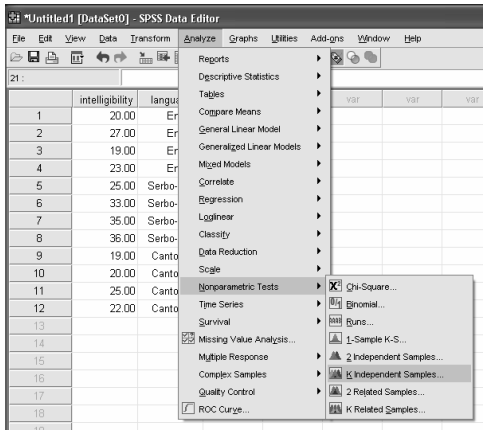
"1" for "English",
"2" for "Serbo-Croat",
"3" for "Cantonese".

Independent measures -

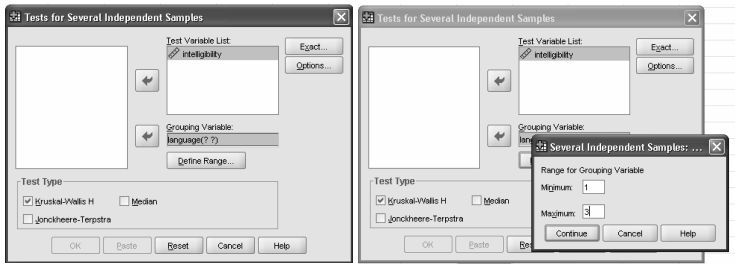
one column gives scores, another column identifies which group each score belongs to.

Using SPSS for the Kruskal-Wallis test:

Analyze > Nonparametric tests > k independent samples



Using SPSS for the Kruskal-Wallis test :



Output from SPSS for Kruskal-Wallis test :

Ranks

	language	N	Mean Rank
intelligibility	English	4	5.00
	Serbo-croat	4	10.13
	Cantonese	4	4.38
	Total	12	

Test Statistics[a,b]

	intelligibility
Chi-Square	6.190
df	2
Asymp. Sig.	.045

a. Kruskal Wallis Test

b. Grouping Variable: language

Friedman's test, step-by-step:

Effects on worker mood of different types of music:

Five workers. Each is tested three times, once under each of the following conditions:
condition 1: silence.
condition 2: "easy-listening" music.
condition 3: marching-band music.

DV: mood rating ("0" = unhappy, "100" = euphoric).
***Ratings* - so use a nonparametric test.**

NB: to avoid practice and fatigue effects, order of presentation of conditions is varied across subjects.

	Silence (raw score)	Silence (ranked score)	Easy (raw score)	Easy (ranked score)	Band (raw score)	Band (ranked score)
Wkr 1:	**4**	***1***	**5**	***2***	**6**	***3***
Wkr 2:	**2**	***1***	**7**	***2.5***	**7**	***2.5***
Wkr 3:	**6**	***1.5***	**6**	***1.5***	**8**	***3***
Wrkr 4:	**3**	***1***	**7**	***3***	**5**	***2***
Wrkr 5:	***3***	***1***	**8**	***2***	**9**	***3***

Step 1:
Rank *each subject's scores* individually.
Worker 1's scores are 4, 5, 6: these get ranks of 1, 2, 3.
Worker 4's scores are 3, 7, 5: these get ranks of 1, 3, 2 .

	Silence (raw score)	Silence (ranked score)	Easy (raw score)	Easy (ranked score)	Band (raw score)	Band (ranked score)
Wkr 1:	4	*1*	5	*2*	6	*3*
Wkr 2:	2	*1*	7	*2.5*	7	*2.5*
Wkr 3:	6	*1.5*	6	*1.5*	8	*3*
Wrkr 4:	3	*1*	7	*3*	5	*2*
Wrkr 5:	*3*	*1*	8	*2*	9	*3*

Step 2:
Find the rank total for each condition, using the ranks from all subjects within that condition.

Rank total for "Silence" condition: 1+1+1.5+1+1 = 5.5.
Rank total for "Easy Listening" condition = 11.
Rank total for "Marching Band" condition = 13.5.

Step 3:
Work out χr^2

$$\chi r^2 = \left[\left(\frac{12}{N * C * (C+1)}\right) * \Sigma Tc^2\right] - 3 * N * (C+1)$$

C is the number of conditions.
N is the number of subjects.
ΣTc^2 is the sum of the squared rank totals for each condition.

$$\chi r^2 = \left[\left(\frac{12}{N * C * (C+1)}\right) * \Sigma Tc^2\right] - 3 * N * (C+1)$$

To get ΣTc^2 :

(a) square each rank total:
5.5^2 = 30.25. 11^2 = 121. 13.5^2 = 182.25.

(b) Add together these squared totals.
30.25 + 121 + 182.25 = 333.5.

In our example,

$$\chi r^2 = \left[\left(\frac{12}{N * C * (C+1)}\right) * \Sigma Tc^2\right] - 3 * N * (C+1)$$

$$\chi r^2 = \left[\left(\frac{12}{5 * 3 * 4}\right) * 333.5\right] - 3 * 5 * 4 = 6.7$$

$\chi r^2 = 6.7$

Step 4:
Degrees of freedom = number of conditions minus one.
df = 3 - 1 = 2.

Step 5:
Assessing the statistical significance of χr^2 depends on the number of subjects and the number of groups.

(a) Less than 9 subjects:
Use a special table of critical values (as on my website).

(b) 9 or more subjects:
Use a Chi-Square table (on my website).
Compare your obtained χr^2 value to the critical value of Chi-Square for your d.f.
If your obtained χr^2 is *bigger* than the critical Chi-Square value, your conditions are significantly different.

The test only tells you that *some kind* of difference exists; look at the median score for each condition to see where the difference comes from.

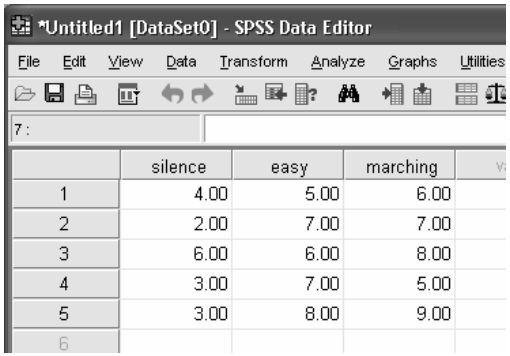
We have 5 subjects and 3 conditions, so use Friedman table for small sample sizes:

3 groups:

N=4		N=5		N=6		N=7		N=8		N=9	
Chi-R Sq.	*p*	*Chi-R Sq.*	*p*	*Chi-R Sq.*	*p*	*Chi-R Sq.*	*p*	*Chi-R Sq.*	*p*	*Chi-R Sq.*	*p*
6.5	0.042	6.4	0.039	6.33	0.052	6	0.052	6.25	0.047	6.222	0.048
8	0.0048	7.6	0.024	7	0.029	7.143	0.027	6.75	0.038	6.889	0.031
		8.4	0.0085	8.33	0.012	7.714	0.021	7	0.03	8	0.019
		10	0.00077	9	0.0081	8	0.016	7.75	0.018	8.222	0.016
				9.33	0.0055	8.857	0.0084	9	0.0099	8.667	0.01
				10.33	0.0017	10.286	0.0036	9.25	0.008	9.556	0.006
				12	0.00013	10.571	0.0027	9.75	0.0048	10.667	0.0035
						11.143	0.0012	10.75	0.0024	10.889	0.0029
						12.286	0.00032	12	0.0011	11.556	0.0013
						14	0.000021	12.25	0.00086	12.667	0.00066
								13	0.00026	13.556	0.00035

Obtained χr^2 is 6.7.
For N = 5, a χr^2 value of 6.4 would occur by chance with a probabilty of 0.024.
Our obtained value is *bigger* than 6.4.
Conclusion: the conditions are significantly different.
Music *does* affect worker mood.

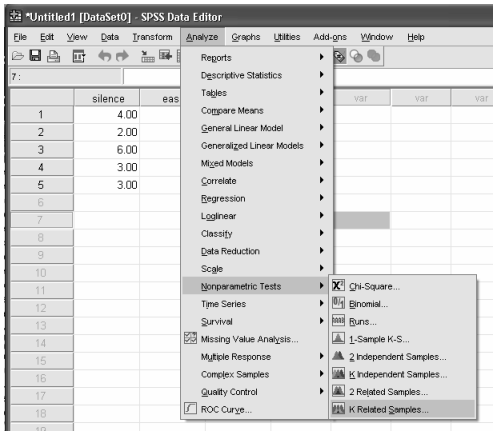
Using SPSS to perform Friedmans test:



***Repeated measures -* each row is one participant's data.**

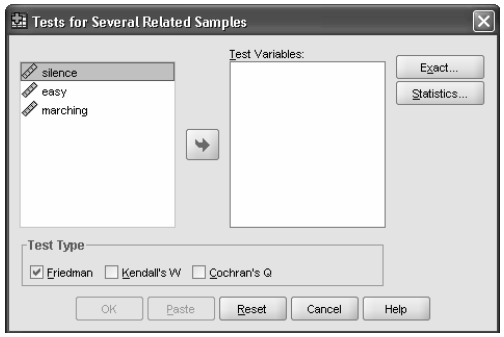
Using SPSS to perform Friedmans test:

Analyze > Nonparametric Tests > k related samples



Using SPSS to perform Friedmans test:

Analyze > Nonparametric Tests > k related samples



Using SPSS to perform Friedmans test:

Analyze > Nonparametric Tests > k related samples

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
silence	5	3.6000	1.51658	2.00	6.00
easy	5	6.6000	1.14018	5.00	8.00
marching	5	7.0000	1.58114	5.00	9.00

Ranks

	Mean Rank
silence	1.10
easy	2.20
marching	2.70

Test Statistics[a]

N	5
Chi-Square	7.444
df	2
Asymp. Sig.	.024

a. Friedman Test

NB: slightly different value from 6.7 worked out by hand

Which nonparametric test?

Differences in fear ratings for 3, 5 and 7-year olds in response to sinister noises from under their bed.	**Independent measures, three groups:** **Kruskal-Wallis.**
Effects of cheese, brussel sprouts, wine and curry on vividness of a person's dreams on four separate nights.	**Repeated measures, four groups:** **Friedman's.**
Number of people spearing their eardrums after enforced listening to Britney Spears, Beyonce, Robbie Williams and Boyzone.	**Frequency data:** **Chi-Square.**
Pedestrians rate the aggressiveness of owners of different types of car. Group A rate Micra owners; group B rate 4x4 owners; group C rate Subaru owners; group D rate Mondeo owners).	**Independent measures, four groups:** **Kruskal-Wallis.**