DISCOVERING STATISTICS USING SPSS

THIRD EDITION

Non-parametric Tests

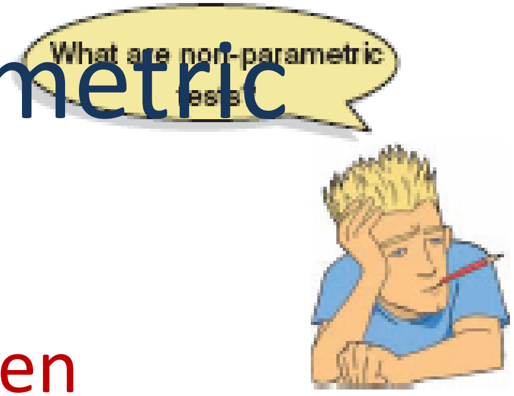
Lecture 12

ANDY FIELD

Aims

- When and why we use non-parametric tests?
  - Mann-Whitney test
    - Wilcoxon rank-sum test
  - Wilcoxon signed-rank test
  - Kruskal–Wallis test
  - Jonckheere–Terpstra test
  - Friedman's ANOVA
- Ranking data
- Interpretation of results
- Reporting results
- Calculating an Effect Size

When to use nonparametric tests

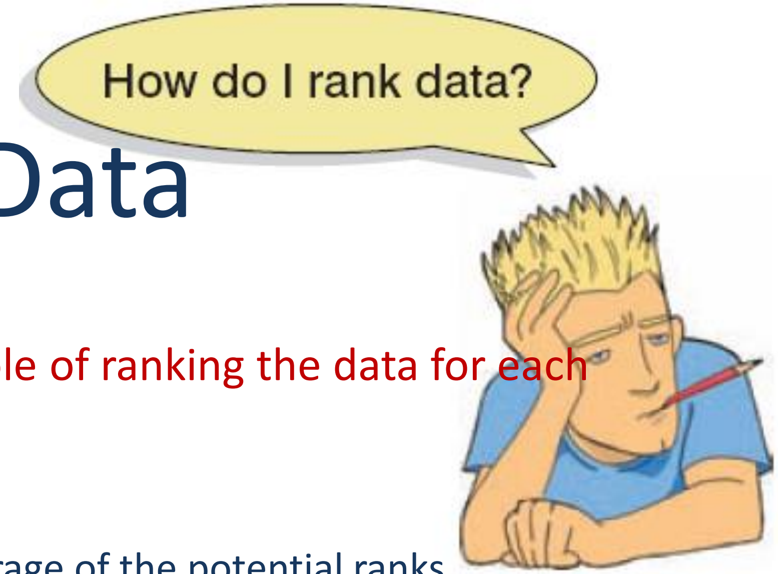


- Non-parametric tests are used when assumptions of parametric tests are not met.
- It is not always possible to correct for problems with the distribution of a data set
  - In these cases we have to use non-parametric tests.
  - They make fewer assumptions about the type of data on which they can be used.

The Wilcoxon rank-sum test and Mann–Whitney test

- These tests are the non-parametric equivalent of the independent *t-test.*

- Use either to test differences between two conditions in which different participants have been used.



Ranking Data

- The tests in this lecture work on the principle of ranking the data for each group:
  - Lowest score = a rank of 1,
  - Next highest score = a rank of 2, and so on.
  - Tied ranks are given the same rank: the average of the potential ranks.
- For an unequal group size
  - The test statistic (*Ws) = sum of ranks in the group that contains the least people.*
- *For an equal group size*
  - *Ws = the value of the smaller summed rank.*

- Add up the ranks for the two groups and take the lowest of these sums to be our test statistic.

- The analysis is carried out on the ranks rather than the actual data.

Theory

- A neurologist investigated the depressant effects of certain recreational drugs.
  - Tested 20 clubbers
  - 10 were given an ecstasy tablet to take on a Saturday night
  - 10 were allowed to drink only alcohol.
  - Levels of depression were measured using the Beck Depression Inventory (BDI) the day after and midweek.
- Rank the data *ignoring* the group to which a person belonged
  - A similar number of high and low ranks in each group suggests depression levels do not differ between the groups.
  - A greater number of high ranks in the ecstasy group than the alcohol group suggests the ecstasy group is more depressed than the alcohol group.

Ranking the Depression scores for Wednesday and Sunday

Wednesday Data

Score	3	5	6	6	7	8	9	10	17	24	27	28	29	30	32	35	35	35	36	39
Potential Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Actual Rank	1	2	3.5	3.5	5	6	7	8	9	10	11	12	13	14	15	17	17	17	19	20
Group	A	A	A	A	A	A	A	A	A	E	E	E	E	A	E	E	E	E	E	E

Sum of Ranks for Alcohol (A) = 59 Sum of Ranks for Ecstasy (E) = 151

Sunday Data

Score	13	13	14	15	15	15	16	16	16	16	17	18	18	18	19	19	20	20	27	35
Potential Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Actual Rank	1.5	1.5	3	5	5	5	8.5	8.5	8.5	8.5	11	13	13	13	15.5	15.5	17.5	17.5	19	20
Group	A	E	A	A	A	E	A	A	E	E	E	E	A	A	E	A	E	A	E	E

Sum of Ranks for Alcohol (A) = 90.5 Sum of Ranks for Ecstasy (E) = 119.5

FIGURE 15.3 Ranking the depression scores for Wednesday

Provisional analysis using SPSS

- First enter the data into SPSS
  - Because the data are collected using different participants in each group, we need to input the data using a coding variable.
    - **For example, 'Drug' with the codes; 1 = ecstasy group and 2 = alcohol group.**
    - **When you enter the data into SPSS remember to tell the computer that a code of 1 represents the group that was given ecstasy and a code of 2 represents the group that was restricted to alcohol.**
- There were no specific predictions about which drug would have the most effect so the analysis should be two-tailed.
- First, run some exploratory analyses on the data
  - Run these exploratory analyses for each group because we're going to be looking for group differences.

Exploratory Analysis Output

Tests of Normality

	Type of Drug	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Beck Depression Inventory (Sunday)	Ecstasy	.276	10	.030	.811	10	.020
	Alcohol	.170	10	.200*	.959	10	.780
Beck Depression Inventory (Wednesday)	Ecstasy	.235	10	.126	.941	10	.566
	Alcohol	.305	10	.009	.753	10	.004

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
Beck Depression Inventory (Sunday)	Based on Mean	3.644	1	18	.072
	Based on Median	1.880	1	18	.187
	Based on Median and with adjusted df	1.880	1	10.076	.200
	Based on trimmed mean	2.845	1	18	.109
Beck Depression Inventory (Wednesday)	Based on Mean	.508	1	18	.485
	Based on Median	.091	1	18	.766
	Based on Median and with adjusted df	.091	1	11.888	.768
	Based on trimmed mean	.275	1	18	.606

Interpreting Output

- Testing for normality: Is the significance of the K–S and Shapiro–Wilk tests less than .05 (sig. different from normal) or greater than .05 (approximately normal).
  - For the Sunday data, the distributions for:
    - Ecstasy, $D(10) = 0.28, p < .05$, appears to be non-normal
    - Alcohol data, $D(10) = 0.17$, *ns,* were normal
  - *For the Wednesday data, the distributions for:*
    - Ecstasy were normal, $D(10) = 0.24$, *ns,*
    - Alcohol appeared to be significantly non-normal, $D(10) = 0.31$, $p < .01$. This finding would alert us to the fact that the sampling distribution might also be non-normal for the Sunday and Wednesday data and that a non-parametric test should be used.
- The second table in SPSS Output shows the results of Levene's test.
  - Sunday data, $F(1, 18) = 3.64$, *ns*
  - Wednesday, $F(1, 18) = 0.51$, *ns*
  - The variances are not significantly different, indicating that the assumption of homogeneity has been met.

Running the Analysis

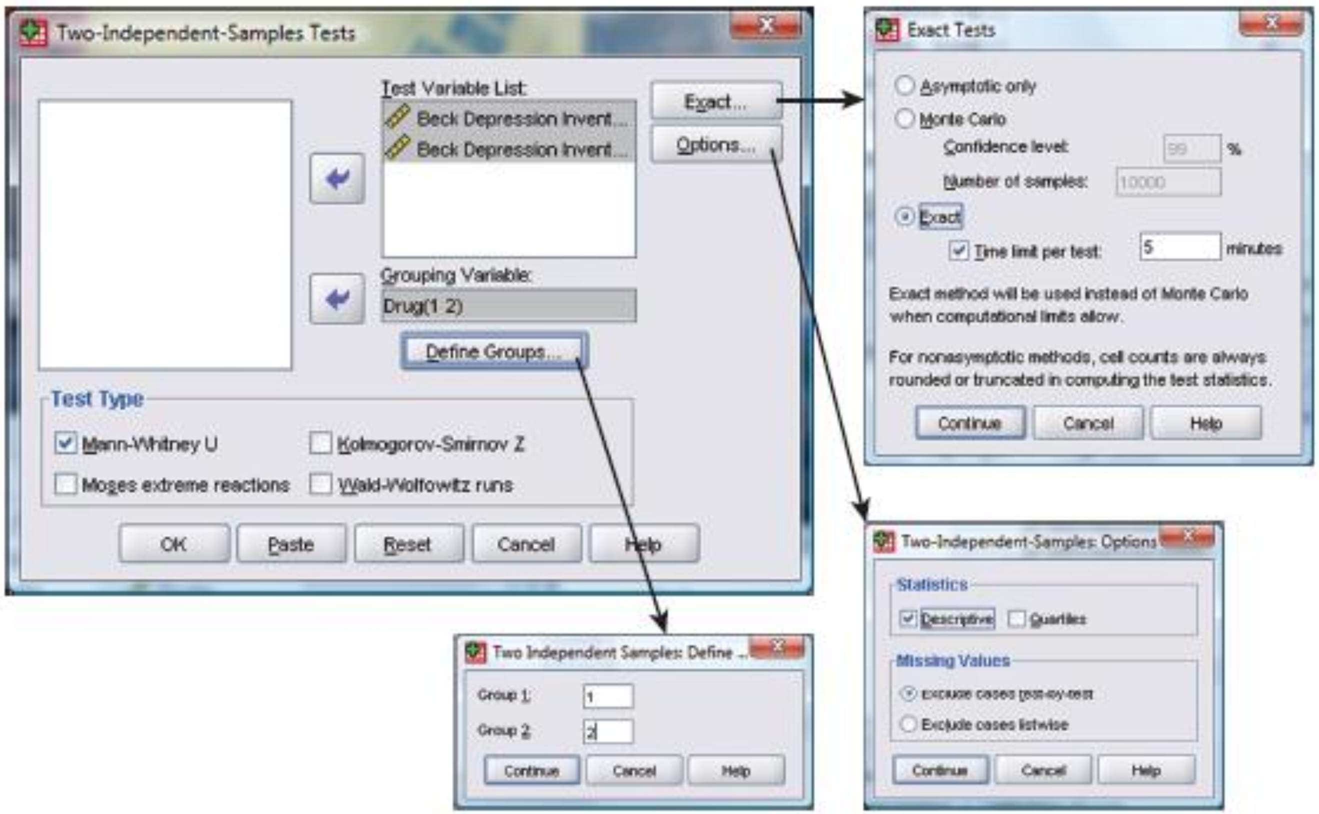


FIGURE 15.4 Dialog boxes for the Mann–Whitney test

Output from the Mann-Whitney test

- The first part of the output summarizes the data after they have been ranked.

Ranks

	Type of Drug	N	Mean Rank	Sum of Ranks
Beck Depression Inventory (Sunday)	Ecstasy	10	11.95	119.50
	Alcohol	10	9.05	90.50
	Total	20		
Beck Depression Inventory (Wednesday)	Ecstasy	10	15.10	151.00
	Alcohol	10	5.90	59.00
	Total	20		

Output from the Mann-Whitney test

- The second table (below) provides the actual test statistics for the Mann–Whitney test, the Wilcoxon procedure and the corresponding *z-score*
- The significance value gives the two-tailed probability that a test statistic of at least that magnitude is a chance result, if the null hypothesis is true.

Test Statistics[b]

	Beck Depression Inventory (Sunday)	Beck Depression Inventory (Wednesday)
Mann-Whitney U	35.500	4.000
Wilcoxon W	90.500	59.000
Z	-1.105	-3.484
Asymp. Sig. (2-tailed)	.269	.000
Exact Sig. [2*(1-tailed Sig.)]	.280[a]	.000[a]

a. Not corrected for ties.

b. Grouping Variable: Type of Drug

Calculating an Effect Size

- The equation to convert a *z-score into the effect size estimate, r,* is as follows (from Rosenthal, 1991: 19):

$$r = \frac{Z}{\sqrt{N}}$$

  - *z* is the z-score that SPSS produces
  - *N* is the size of the study (i.e. the number of total observations)
  - We had 10 ecstasy users and 10 alcohol users and so the total number of observations was 20.

$$r_{\text{Sunday}} = \frac{-1.11}{\sqrt{20}} = -0.25$$

$$r_{\text{Wednesday}} = \frac{-3.48}{\sqrt{20}} -0.78$$

Reporting the Results

- For the Mann–Whitney test
  - Depression levels in ecstasy users (*Mdn* = 17.50) did not differ significantly from alcohol users (*Mdn* = 16.00) the day after the drugs were taken, $U = 35.50$, $z = -1.11$, *ns*, $r = -.25$. However, by Wednesday, ecstasy users (*Mdn* = 33.50) were significantly more depressed than alcohol users (*Mdn* = 7.50), $U = 4.00$, $z = -3.48$, $p < .001$, $r = -.78$.
- We could also choose to report Wilcoxon's test rather than Mann–Whitney's *U* statistic and this would be as follows:
  - Depression levels in ecstasy users (*Mdn* = 17.50) did not significantly differ from alcohol users (*Mdn* = 16.00) the day after the drugs were taken, $W_s = 90.50$, $z = -1.11$, *ns*, $r = -.25$. However, by Wednesday, ecstasy users (*Mdn* = 33.50) were significantly more depressed than alcohol users (*Mdn* = 7.50), $W_s = 59.00$, $z = -3.48$, $p < .001$, $r = -.78$.

Comparing two related conditions: the Wilcoxon signed-rank test

- Uses:
  - To compare two sets of scores, when these scores come from the same participants.
- Imagine the experimenter in the previous example was interested in the change in depression levels for each of the two drugs.
  - We still have to use a non-parametric test because the distributions of scores for both drugs were non-normal on one of the two days.

Ranking data in the Wilcoxon signed-rank test

BDI Sunday	BDI Wednesday	Difference	Sign	Rank	Positive Ranks	Negative Ranks
			Ecstasy			
15	28	13	+	2.5	2.5	
35	35	0	Exclude			
16	35	19	+	6	6	
18	24	6	+	1	1	
19	39	20	+	7	7	
17	32	15	+	4.5	4.5	
27	27	0	Exclude			
16	29	13	+	2.5	2.5	
13	36	23	+	8	8	
20	35	15	+	4.5	4.5	
				Total =	**36**	**0**
			Alcohol			
16	5	–11	–	9		9
15	6	–9	–	7		7
20	30	10	+	8	+8	
15	8	–7	–	3.5		3.5
16	9	–7	–	3.5		3.5
13	7	–6	–	2		2
14	6	–8	–	5.5		5.5
19	17	–2	–	1		1
18	3	–15	–	10		10
18	10	–8	–	5.5		5.5
				Total =	**8**	**47**

Running the Analysis

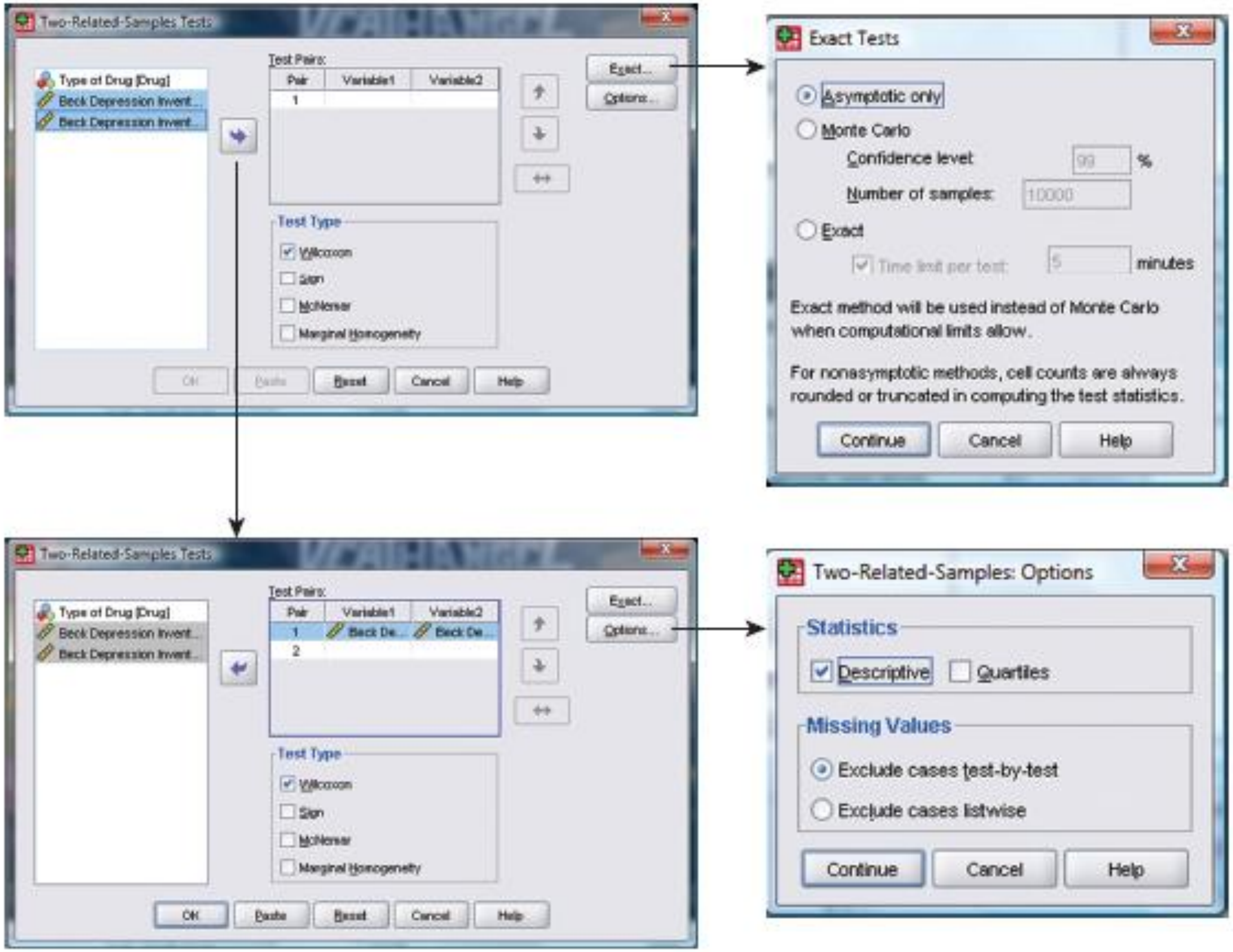


FIGURE 15.5 Dialog boxes for the Wilcoxon signed-rank test

Output for the Ecstasy Group

- If you have split the file, then the first set of results obtained will be for the ecstasy group

Ranks[d]

		N	Mean Rank	Sum of Ranks
Beck Depression Inventory (Wednesday) - Beck Depression Inventory (Sunday)	Negative Ranks	0[a]	.00	.00
	Positive Ranks	8[b]	4.50	36.00
	Ties	2[c]		
	Total	10		

a. Beck Depression Inventory (Wednesday) < Beck Depression Inventory (Sunday)
b. Beck Depression Inventory (Wednesday) > Beck Depression Inventory (Sunday)
c. Beck Depression Inventory (Wednesday) = Beck Depression Inventory (Sunday)
d. Type of Drug = Ecstasy

Test Statistics[b,c]

	Beck Depression Inventory (Wednesday) - Beck Depression Inventory (Sunday)
Z	-2.527[a]
Asymp. Sig. (2-tailed)	.012

a. Based on negative ranks.
b. Wilcoxon Signed Ranks Test
c. Type of Drug = Ecstasy

Output for the alcohol group

Ranks[d]

		N	Mean Rank	Sum of Ranks
Beck Depression Inventory (Wednesday) - Beck Depression Inventory (Sunday)	Negative Ranks	9[a]	5.22	47.00
	Positive Ranks	1[b]	8.00	8.00
	Ties	0[c]		
	Total	10		

a. Beck Depression Inventory (Wednesday) < Beck Depression Inventory (Sunday)

b. Beck Depression Inventory (Wednesday) > Beck Depression Inventory (Sunday)

c. Beck Depression Inventory (Wednesday) = Beck Depression Inventory (Sunday)

d. Type of Drug = Alcohol

Test Statistics[b,c]

	Beck Depression Inventory (Wednesday) - Beck Depression Inventory (Sunday)
Z	-1.990[a]
Asymp. Sig. (2-tailed)	.047

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

c. Type of Drug = Alcohol

Calculating an Effect size

- The effect size can be calculated in the same way as for the Mann–Whitney test.
- In this case SPSS Output tells us that for the ecstasy group *z* is –2.53, and for the alcohol group is −1.99.
- In both cases we had 20 observations
  - (although we only used 10 people and tested them twice, it is the number of observations, not the number of people, that is important here).
- *The effect size is therefore:*

$$r_{\text{Ecstasy}} = \frac{-2.53}{\sqrt{20}} = -0.57$$

$$r_{\text{Alcohol}} = \frac{-199}{\sqrt{20}} = -0.44$$

Reporting the results

- Reporting Test-Statistic:
  - For ecstasy users depression levels were significantly higher on Wednesday ($Mdn = 33.50$) than on Sunday ($Mdn = 17.50$), $T = 0$, $p < .05$, $r = -.57$. However, for alcohol users the opposite was true: depression levels were significantly lower on Wednesday ($Mdn = 7.50$) than on Sunday ($Mdn = 16.0$), $T = 8$, $p < .05$, $r = -.44$.
- Reporting the values of *z*:
  - For ecstasy users, depression levels were significantly higher on Wednesday ($Mdn = 33.50$) than on Sunday ($Mdn = 17.50$), $z = -2.53$, $p < .05$, $r = -.57$. However, for alcohol users the opposite was true: depression levels were significantly lower on Wednesday ($Mdn = 7.50$) than on Sunday ($Mdn = 16.0$), $z = -1.99$, $p < .05$, $r = -.44$.

Differences between several independent groups: the Kruskal–Wallis test

- The Kruskal–Wallis test (Kruskal & Wallis, 1952;) is the non-parametric counterpart of the one-way independent ANOVA .
  - If you have data that have violated an assumption then this test can be a useful way around the problem.
- The theory for the Kruskal–Wallis test is very similar to that of the Mann–Whitney (and Wilcoxon rank-sum) test,
  - Like the Mann–Whitney test, the Kruskal–Wallis test is based on ranked data.
  - The sum of ranks for each group is denoted by R_i (where i is used to denote the particular group).

Kruskal-Wallis Theory

- Once the sum of ranks has been calculated for each group, the test statistic, *H,* is calculated as:

$$H = \frac{12}{N(N-1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

  - R_i is the sum of ranks for each group,
  - N is the total sample size (in this case 80)
  - n_i is the sample size of a particular group (in this case we have equal sample sizes and they are all 20).

Example

- Does eating soya affect your sperm count?
- Variables
  - Outcome: sperm (millions)
  - IV: Number of soya meals per week
    - No Soya meals
    - 1 Soya meal
    - 4 soya meals
    - 7 soya meals
- Participants
  - 80 males (20 in each group)

Data for the Soya Example with Ranks

No Soya		1 Soya Meal		4 Soya Meals		7 Soya Meals	
Sperm (Millions)	Rank	Sperm (Millions)	Rank	Sperm (Millions)	Rank	Sperm (Millions)	Rank
0.35	4	0.33	3	0.40	6	0.31	1
0.58	9	0.36	5	0.60	10	0.32	2
0.88	17	0.63	11	0.96	19	0.56	7
0.92	18	0.64	12	1.20	21	0.57	8
1.22	22	0.77	14	1.31	24	0.71	13
1.51	30	1.53	32	1.35	27	0.81	15
1.52	31	1.62	34	1.68	35	0.87	16
1.57	33	1.71	36	1.83	37	1.18	20
2.43	41	1.94	38	2.10	40	1.25	23
2.79	46	2.48	42	2.93	48	1.33	25
3.40	55	2.71	44	2.96	49	1.34	26
4.52	59	4.12	57	3.00	50	1.49	28
4.72	60	5.65	61	3.09	52	1.50	29
6.90	65	6.76	64	3.36	54	2.09	39
7.58	68	7.08	66	4.34	58	2.70	43
7.78	69	7.26	67	5.81	62	2.75	45
9.62	72	7.92	70	5.94	63	2.83	47
10.05	73	8.04	71	10.16	74	3.07	51
10.32	75	12.10	77	10.98	76	3.28	53
21.08	80	18.47	79	18.21	78	4.11	56
Total (R_i)	**927**		**883**		**883**		**547**

Provisional Analysis

- Run some exploratory analyses on the data
  - We need to run these exploratory analyses for each group because we're going to be looking for group differences .

Tests of Normality

	Number of Soya Meals Per Week	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Sperm Count (Millions)	No Soya Meals	.181	20	.085	.805	20	.001
	1 Soya Meal Per Week	.207	20	.024	.826	20	.002
	4 Soyal Meals Per Week	.267	20	.001	.743	20	.000
	7 Soya Meals Per Week	.204	20	.028	.912	20	.071

a. Lilliefors Significance Correction

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
Sperm Count (Millions)	Based on Mean	5.117	3	76	.003
	Based on Median	2.860	3	76	.042
	Based on Median and with adjusted df	2.860	3	58.107	.045
	Based on trimmed mean	4.070	3	76	.010

Running the Analysis

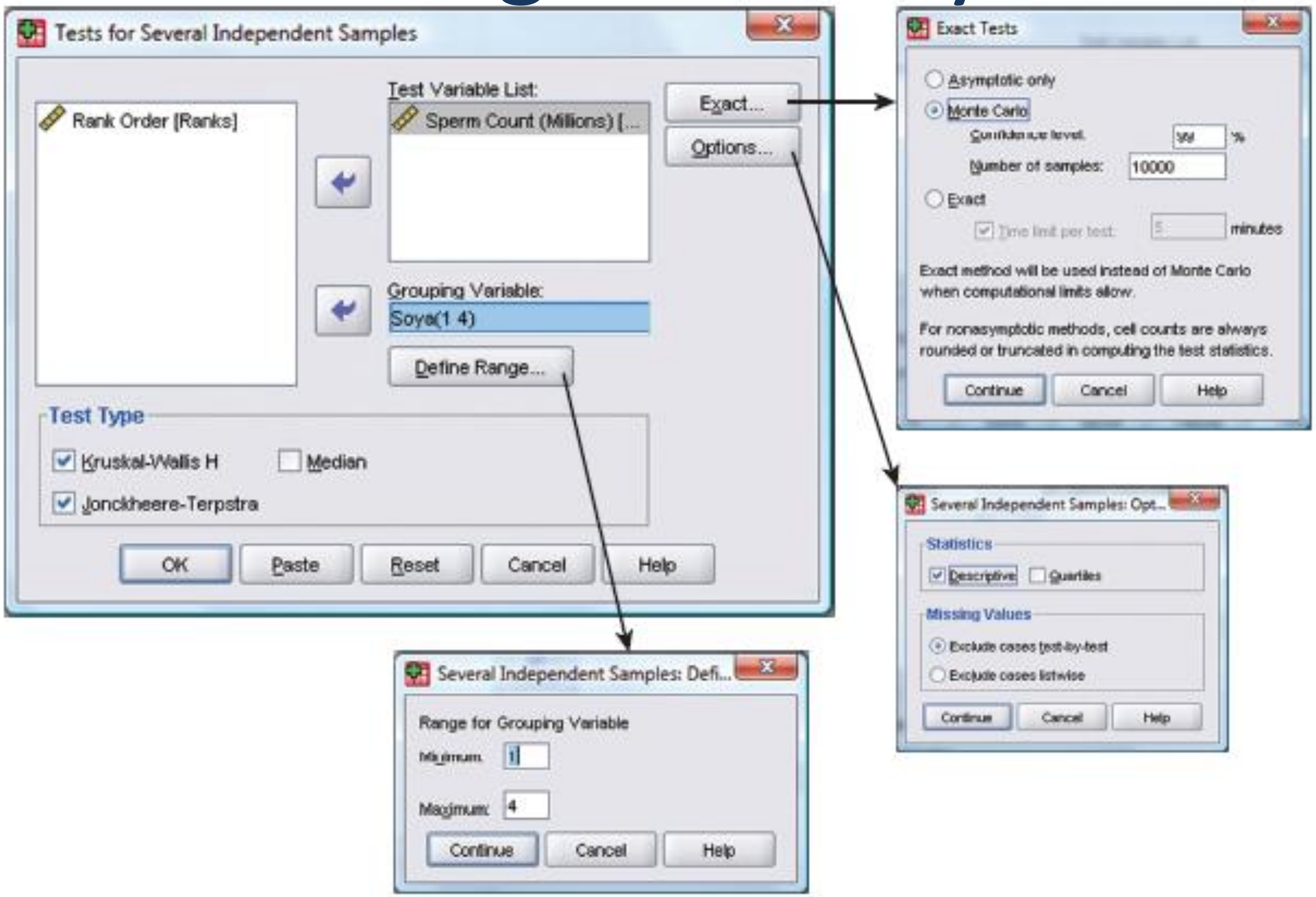


FIGURE 15.7 Dialog boxes for the Kruskal–Wallis test

Output from the Kruskal–Wallis test

Ranks

	Number of Soya Meals	N	Mean Rank
Sperm Count (Millions)	No Soya Meals	20	46.35
	1 Soya Meal Per Week	20	44.15
	4 Soyal Meals Per Week	20	44.15
	7 Soya Meals Per Week	20	27.35
	Total	80	

Test Statistics[b,c]

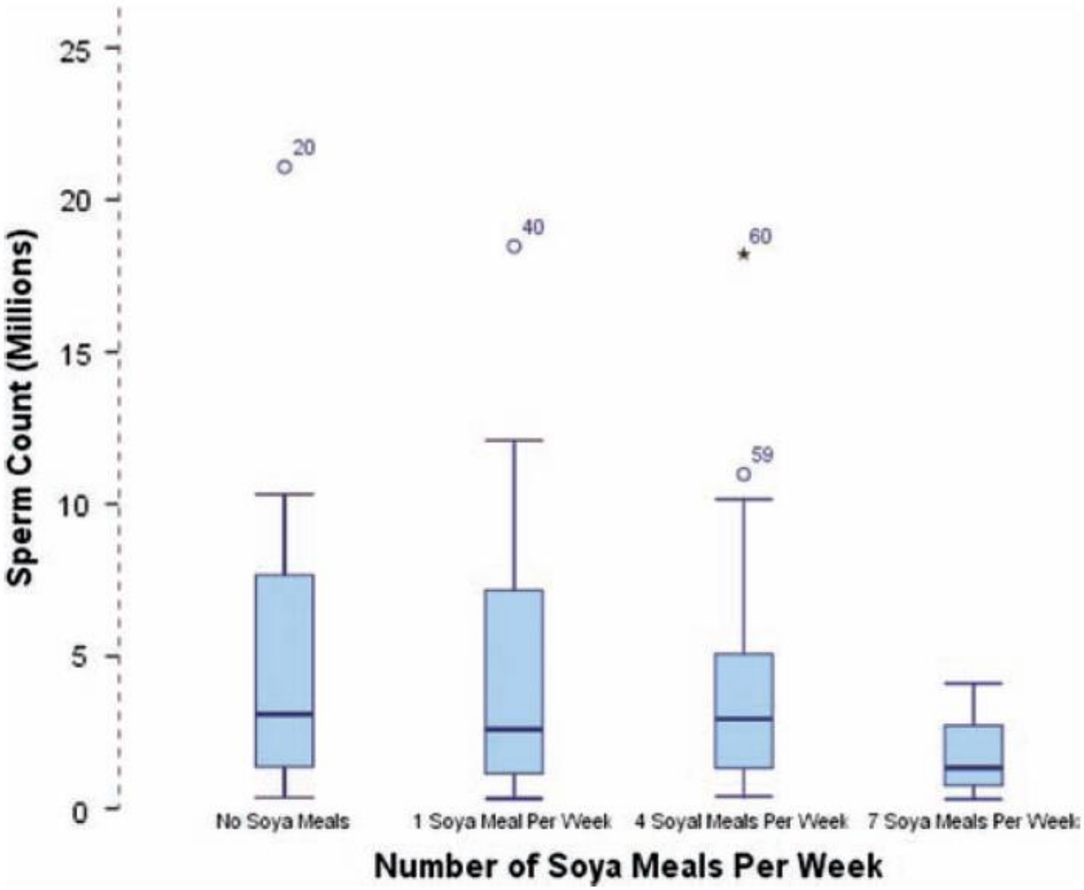
			Sperm Count (Millions)
Chi-Square			8.659
df			3
Asymp. Sig.			.034
Monte Carlo Sig.	Sig.		.031[a]
	99% Confidence Interval	Lower Bound	.027
		Upper Bound	.036

a. Based on 10000 sampled tables with starting seed 2000000.

b. Kruskal Wallis Test

c. Grouping Variable: Number of Soya Meals Per Week

Boxplot for the sperm counts of individuals eating different numbers of soya meals per week



Post hoc tests for the Kruskal–Wallis test

- One way to do a non-parametric *post hoc* procedure is to use Mann–Whitney tests.
  - However, lots of Mann–Whitney tests will inflate the Type I error rate.
    - Bonferroni correction
      - Instead of using .05 as the critical value for significance for each test, you use a critical value of .05 divided by the number of tests you've conducted.
    - It is very restrictive. Therefore, it's a good idea to be selective about the comparisons you make.
    - In this example, we have a control group which had no soya meals. As such, a nice succinct set of comparisons would be to compare each group against the control:
- Comparisons:
  - Test 1: one soya meal per week compared to no soya meals
  - Test 2: four soya meals per week compared to no soya meals
  - Test 3: seven soya meals per week compared to no soya meals
- Bonferroni correction:
  - Rather than use .05 as our critical level of significance, we'd use .05/3 = .0167.

Output of the test statistics from the Mann–Whitney tests

No Soya vs. 1 Meal per week:

Test Statistics[b]

	Sperm Count (Millions)
Mann-Whitney U	191.000
Wilcoxon W	401.000
Z	-.243
Asymp. Sig. (2-tailed)	.808
Exact Sig. [2*(1-tailed Sig.)]	.820[a]

a. Not corrected for ties.

b. Grouping Variable: Number of Soya Meals Per Week

No Soya vs. 4 Meals per week:

Test Statistics[b]

	Sperm Count (Millions)
Mann-Whitney U	188.000
Wilcoxon W	398.000
Z	-.325
Asymp. Sig. (2-tailed)	.745
Exact Sig. [2*(1-tailed Sig.)]	.758[a]

a. Not corrected for ties.

b. Grouping Variable: Number of Soya Meals Per Week

No Soya vs. 7 Meals per week:

Test Statistics[b]

	Sperm Count (Millions)
Mann-Whitney U	104.000
Wilcoxon W	314.000
Z	-2.597
Asymp. Sig. (2-tailed)	.009
Exact Sig. [2*(1-tailed Sig.)]	.009[a]

a. Not corrected for ties.

b. Grouping Variable: Number of Soya Meals Per Week

Testing for trends: the Jonckheere–Terpstra test

- This statistic tests for an ordered pattern to the medians of the groups you're comparing.
- Essentially it does the same thing as the Kruskal–Wallis test but it incorporates information about whether the order of the groups is meaningful.
  - Use this test when you expect the groups you're comparing to produce a meaningful order of medians.
  - In the current example we expect that the more soya a person eats, the more their sperm count will go down.

Output for Jonckheere-Terpstra Test

Jonckheere-Terpstra Test[b]

			Sperm Count (Millions)
Number of Levels in Number of Soya Meals Per Week			4
N			80
Observed J-T Statistic			912.000
Mean J-T Statistic			1200.000
Std. Deviation of J-T Statistic			116.333
Std. J-T Statistic			-2.476
Asymp. Sig. (2-tailed)			.013
Monte Carlo Sig. (2-tailed)	Sig.		.012[a]
	99% Confidence Interval	Lower Bound	.009
		Upper Bound	.015
Monte Carlo Sig. (1-tailed)	Sig.		.006[a]
	99% Confidence Interval	Lower Bound	.004
		Upper Bound	.008

a. Based on 10000 sampled tables with starting seed 2000000.

b. Grouping Variable: Number of Soya Meals Per Week

Calculating Effect Sizes for the Mann-Whitney tests

- For the first comparison
  - (no soya vs. 1 meal) $z = -0.243$, the effect size is therefore:

$$r_{\text{NoSoya}-1\text{Meal}} = \frac{-0.243}{\sqrt{40}} = -.04$$

- For the second comparison
  - (no soya vs. 4 meals) $z = -0.325$, the effect size is therefore:

$$r_{\text{NoSoya}-1\text{Meal}} = \frac{-0.325}{\sqrt{40}} = -.05$$

- For the third comparison
  - (no soya vs. 7 meals) $z = -2.597$, the effect size is therefore:

$$r_{\text{NoSoya}-1\text{Meal}} = \frac{-2.597}{\sqrt{40}} = -.41$$

Writing the Results

- For the Kruskal–Wallis test, we need only report the test statistic (*H*), its degrees of freedom and its significance:
  - Sperm counts were significantly affected by eating soya meals, $H(3) = 8.66$, $p < .05$.
- However, we need to report the follow-up tests as well (including their effect sizes):
  - Sperm counts were significantly affected by eating soya meals, $H(3) = 8.66$, $p < .05$. Mann–Whitney tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a .0167 level of significance.
  - It appeared that sperm counts were not significantly different when one soya meal ($U = 191$, $r = -.04$) or four soya meals ($U = 188$, $r = -.05$) were eaten per week compared to none.
  - However, when seven soya meals were eaten per week sperm counts were significantly lower than when no soya was eaten ($U = 104$, $r = -.41$).
  - We can conclude that if soya is eaten every day it significantly reduces sperm counts compared to eating none; however, eating soya less than every day has no significant effect on sperm counts.
- Or, we might want to report our trend:
  - All effects are reported at $p < .05$. Sperm counts were significantly affected by eating soya meals ($H(3) = 8.66$). Jonckheere's test revealed a significant trend in the data: as more soya was eaten, the median sperm count decreased, $J = 912$, $z = -2.48$, $r = -.28$.

Differences between several related groups: Friedman's ANOVA

- Used for testing differences between conditions when:
  - There are more than two conditions
  - The same participants have been used in all conditions (each case contributes several scores to the data).
- If you have violated some assumption of parametric tests then this test can be a useful way around the problem.

Theory of Friedman’s ANOVA

- The theory for Friedman’s ANOVA is much the same as the other tests: it is based on ranked data.
- Once the sum of ranks has been calculated for each group, the test statistic, F_r, is calculated as:

$$F_r = \left[\frac{12}{Nk(k+1)}\sum_{i=1}^{k} R_i^2\right] - 3N(k+1)$$

Example

- Does the ‘andikins’ dert work?
- Variables
  - Outcome: weight (Kg)
  - IV: Time since beginning the diet
    - Baseline
    - 1 Month
    - 2 Months
- Participants
  - 10 women

TABLE 15.5 Data for the diet example with ranks

	Weight				Weight		
	Start	Month 1	Month 2		Start (Ranks)	Month 1 (Ranks)	Month 2 (Ranks)
Person 1	63.75	65.38	81.34		1	2	3
Person 2	62.98	66.24	69.31		1	2	3
Person 3	65.98	67.70	77.89		1	2	3
Person 4	107.27	102.72	91.33		3	2	1
Person 5	66.58	69.45	72.87		1	2	3
Person 6	120.46	119.96	114.26		3	2	1
Person 7	62.01	66.09	68.01		1	2	3
Person 8	71.87	73.62	55.43		2	3	1
Person 9	83.01	75.81	71.63		3	2	1
Person 10	76.62	67.66	68.60		3	1	2
				R_i	19	20	21

Provisional analysis

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Weight at Start (kg)	.228	10	.149	.784	10	.009
Weight after 1 month (kg)	.335	10	.002	.685	10	.001
Weight after 2 months (kg)	.203	10	.200*	.877	10	.121

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Running the Analysis

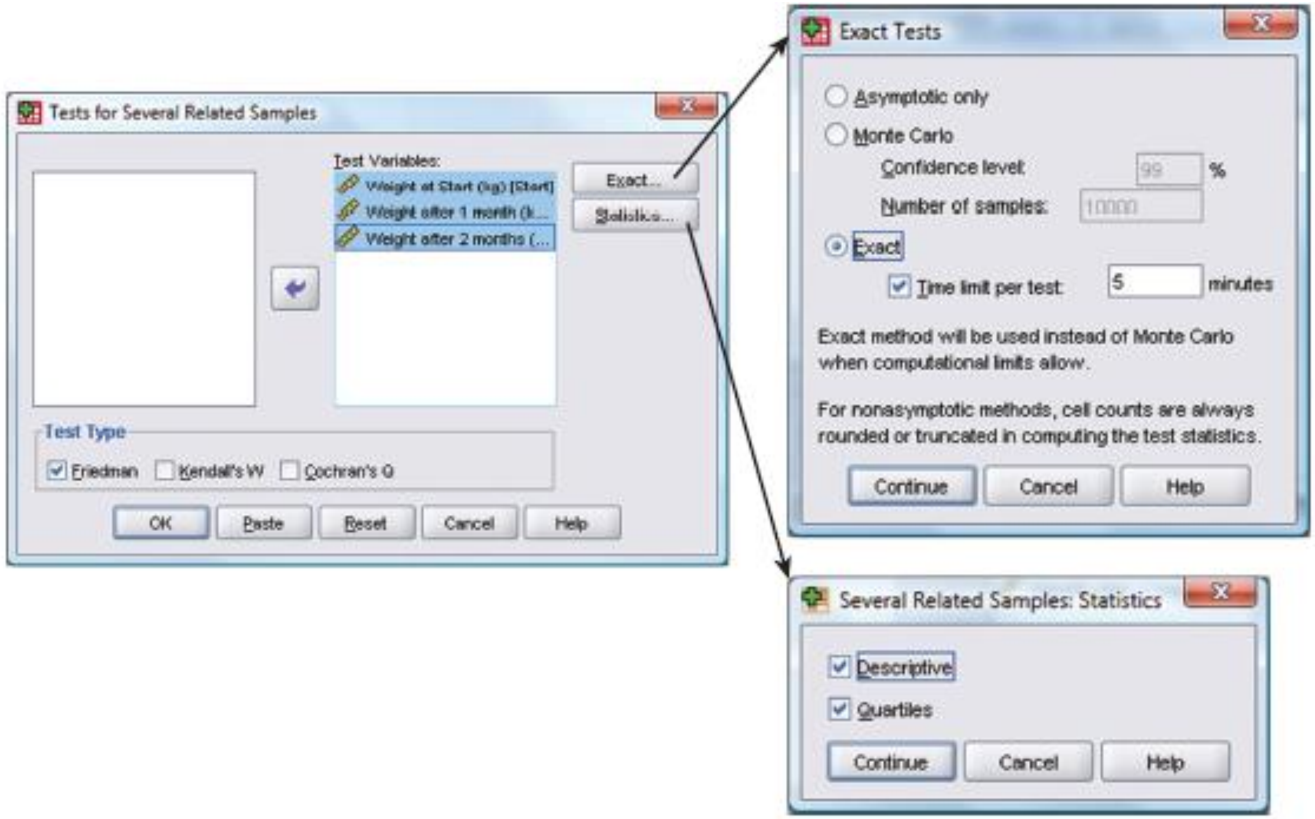


FIGURE 15.9 Dialog boxes for Friedman's ANOVA

Output from Friedman's ANOVA

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Weight at Start (kg)	10	78.0543	20.23008	62.01	120.46	63.5549	69.2288	89.0709
Weight after 1 month (kg)	10	77.4635	18.61502	65.38	119.96	66.2065	68.5728	82.5385
Weight after 2 months (kg)	10	77.0668	16.10612	55.43	114.26	68.4525	72.2493	83.8365

Ranks

	Mean Rank
Weight at Start (kg)	1.90
Weight after 1 month (kg)	2.00
Weight after 2 months (kg)	2.10

Test Statistics[a]

N	10.000
Chi-Square	.200
df	2.000
Asymp. Sig.	.905
Exact Sig.	.974
Point Probability	.143

a. Friedman Test

- There is no need to do any post hoc tests for this example because the main ANOVA was non-significant.

Writing and interpreting the results

- For Friedman's ANOVA we need only report the test statistic (χ^2), its degrees of freedom and its significance:
  - The weight of participants did not significantly change over the two months of the diet, $\chi^2(2) = 0.20$, $p > .05$

To sum up …

- When data violate the assumptions of parametric tests we can sometimes find a nonparametric equivalent
  - Usually based on analysing the ranked data
- Mann-Whitney/Wilcoxon rank-sum Test
  - Compares two independent groups of scores
- Wilcoxon signed rank Test
  - Compares two dependent groups of scores
- Kruskal-Wallis Test
  - Compares > 2 independent groups of scores
- Friedman's Test
  - Compares > 2 dependent groups of scores