Chapter 6: Non-parametric models

Smart Alex's Solutions

Task 1

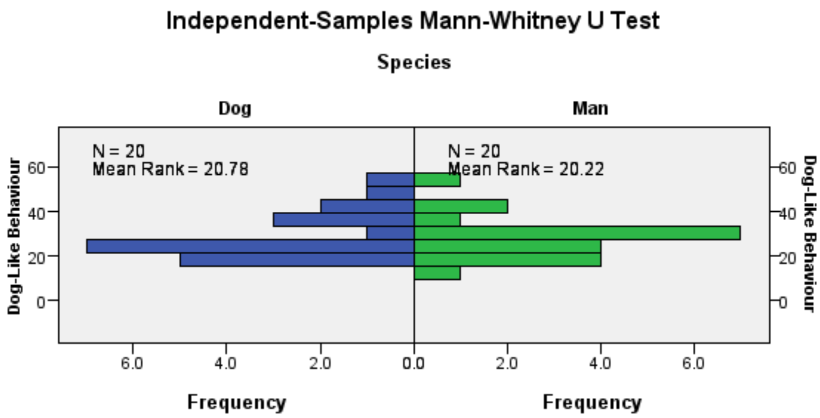
A psychologist was interested in the cross-species differences between men and dogs. She observed a group of dogs and a group of men in a naturalistic setting (20 of each). She classified several behaviours as being dog-like (urinating against trees and lamp posts, attempts to copulate with anything that moved, and attempts to lick their own genitals). For each man and dog she counted the number of dog-like behaviours displayed in a 24-hour period. It was hypothesized that dogs would display more dog-like behaviours than men. The data are in the file ***MenLikeDogs.sav****. Analyse them with a Mann–Whitney test.*

SPSS Output

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Dog-Like Behaviour is the same across categories of Species.	Independent-Samples Mann-Whitney U Test	.881	Retain the null hypothesis.

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



Total N	40
Mann-Whitney U	194.500
Wilcoxon W	404.500
Test Statistic	194.500
Standard Error	36.744
Standardized Test Statistic	-.150
Asymptotic Sig. (2-sided test)	.881

Calculating an effect size

The output tells us that z is –0.15 (standardized test statistic), and we had 20 men and 20 dogs so the total number of observations was 40. The effect size is, therefore:

$$r = \frac{-0.15}{\sqrt{40}} = -.02$$

This represents a tiny effect (it is close to zero), which tells us that there truly isn't much difference between dogs and men.

Writing and interpreting the result

We could report something like:

- ✓ Men (*Mdn* = 27) did not seem to differ from dogs (*Mdn* = 24) in the amount of dog-like behaviour they displayed, *U* = 194.5, *ns*.

Note that I've reported the median for each condition. Of course, we really ought to include the effect size as well. We could do two things. The first is to report the *z*-score associated with the test statistic. This value would enable the reader to determine both the exact significance of the test, and to calculate the effect size *r*:

- ✓ Men (*Mdn* = 27) and dogs (*Mdn* = 24) did not significantly differ in the extent to which they displayed dog-like behaviours, *U* = 194.5, *ns*, *z* = −0.15.

The alternative is to just report the effect size (because readers can convert back to the *z*-score if they need to for any reason). This approach is better because the effect size will probably be most useful to the reader.

- ✓ Men (*Mdn* = 27) and dogs (*Mdn* = 24) did not significantly differ in the extent to which they displayed dog-like behaviours, *U* = 194.5, *ns*, *r* = −.02.

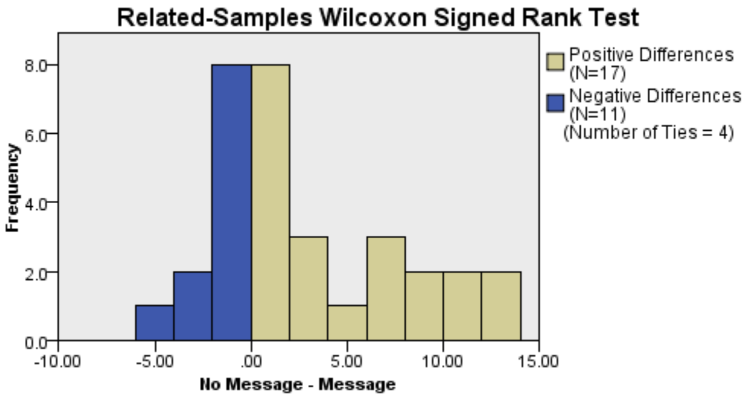
Task 2

There's been much speculation over the years about the influence of subliminal messages on records. To name a few cases, both Ozzy Osbourne and Judas Priest have been accused of putting backward masked messages on their albums that subliminally influence poor unsuspecting teenagers into doing things like blowing their heads off with shotguns. A psychologist was interested in whether backward masked messages really did have an effect. He took the master tapes of Britney Spears's 'Baby one more time' and created a second version that had the masked message 'deliver your soul to the dark lord' repeated in the chorus. He took this version, and the original, and played one version (randomly) to a group of 32 people. He took the same group six months later and played them whatever version they hadn't heard the time before. So each person heard both the original, and the version with the masked message, but at different points in time. The psychologist measured the number of goats that were sacrificed in the week after listening to each version. It was hypothesized that the backward message would lead to more goats being sacrificed. The data are in the file ***DarkLord.sav****. Analyse them with a Wilcoxon signed-rank test.*

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Message and No Message equals 0.	Related-Samples Wilcoxon Signed Rank Test	.036	Reject the null hypothesis.

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



Total N	32
Test Statistic	294.500
Standard Error	43.699
Standardized Test Statistic	2.094
Asymptotic Sig. (2-sided test)	.036

Calculating an effect size

The output tells us that *z* is 2.094 (standardized test statistic), and we had 64 observations (although we only used 32 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:

$$r = \frac{2.094}{\sqrt{64}} = .26$$

This represents a medium effect (it is close to Cohen's benchmark of .3), which tells us that the effect of whether or a subliminal message was present was a substantive effect.

Writing and interpreting the result

We could report something like:

- ✓ The number of goats sacrificed after hearing the message (*Mdn* = 9) was significantly less than after hearing the normal version of the song (*Mdn* = 11), $T = 111.50$, $p < .05$.

As with the Mann–Whitney test, we should report either the *z*-score or the effect size. The effect size is most useful:

- ✓ The number of goats sacrificed after hearing the message (*Mdn* = 9) was significantly less than after hearing the normal version of the song (*Mdn* = 11), $T = 111.50$, $p < .05$, $r = .26$.

Task 3

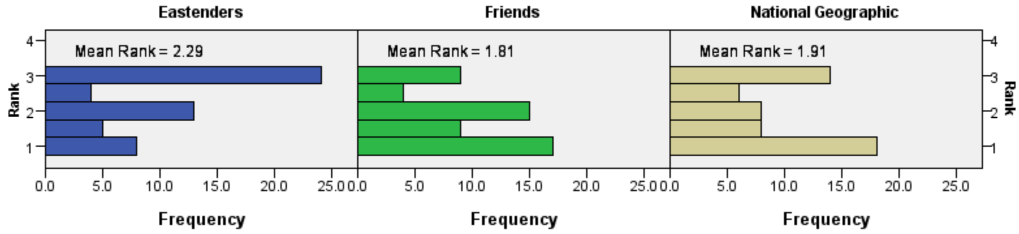
A psychologist was interested in the effects of television programmes on domestic life. She hypothesized that through 'learning by watching', certain programmes might actually encourage people to behave like the characters within them. This in turn could affect the viewer's own relationships (depending on whether the programme depicted harmonious or dysfunctional relationships). She took episodes of three popular TV shows and showed them to 54 couples, after which the couple were left alone in the room for an hour. The experimenter measured the number of times the couple argued. Each couple viewed all three of the TV programmes at different points in time (a week apart) and the order in which the programmes were viewed was counterbalanced over couples. The TV programmes selected were EastEnders *(which typically portrays the lives of extremely miserable, argumentative, London folk who like nothing more than to beat each other up, lie to each other, sleep with each other's wives and generally show no evidence of any consideration to their fellow humans),* Friends *(which portrays a group of unrealistically considerate and nice people who love each other oh so very much—but I love it anyway), and a National Geographic programme about whales (this was a control). The data are in the file* ***Eastenders.****sav. Access them and conduct Friedman's ANOVA on the data.*

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distributions of Eastenders, Friends and National Geographic are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	.023	Reject the null hypothesis.

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

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks



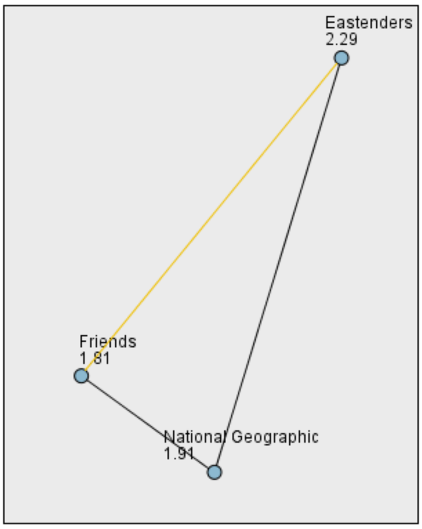
Total N	54
Test Statistic	7.586
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.023

The graph above shows the mean rank in each condition. These mean ranks are important later for interpreting any effects; they show that the ranks were highest after watching *EastEnders*.

The table below the graph shows the chi-square test statistic and its associated degrees of freedom (in this case we had three groups so the degrees of freedom are 3 – 1, or 2), and the significance. Therefore, we could conclude that the type of programme watched significantly affected the subsequent number of arguments (because the significance value is less than .05). However, this result doesn't tell us exactly where the differences lie. To see where the differences lie we can look at the pairwise comparisons that we requested.

Follow-up analysis

Pairwise Comparisons



Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Friends-National Geographic	-.102	.192	-.529	.597	1.000
Friends-Eastenders	.481	.192	2.502	.012	.037
National Geographic-Eastenders	.380	.192	1.973	.049	.146

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.

Looking at the diagram in the output of the pairwise comparisons above, we can see that the test comparing Friends to *EastEnders* is significant (as indicated by the yellow line); however, the other two comparisons were both non-significant (as indicated by the black lines). The table below the diagram confirms this and tells us the significance values of the three comparisons. The significance value of the comparison between *Friends* and *EastEnders* is .037, which is below our criterion of .05, therefore we can conclude that *EastEnders* led to significantly more arguments than *Friends*. The effect we got seems to mainly reflect the fact that *EastEnders* makes people argue more.

Calculating an effect size

For the first comparison (*Friends* vs. control) z is −.529, and because this is based on comparing two groups each containing 54 observations, we have 108 observations in total (remember that it isn't important that the observations come from the same people). The effect size is, therefore:

$$r_{Friends-\text{Control}} = \frac{-0.529}{\sqrt{108}} = -.05$$

This represents virtually no effect (it is close to zero). Therefore, *Friends* had very little effect in creating arguments compared to the control.

For the second comparison (*Friends* vs. *EastEnders*) z is 2.502, and this was again based on 108 observations. The effect size is, therefore:

$$r_{Friends-EastEnders} = \frac{2.502}{\sqrt{108}} = .24$$

This represents a medium effect (it is close to Cohen's benchmark of .3), which tells us that the effect of *EastEnders* relative to *Friends* was a substantive effect: *EastEnders* produced substantially more arguments.

For the third comparison (*EastEnders* vs. Control) z is 1.973, and this was again based on 108 observations. The effect size is, therefore:

$$r_{\text{Control}-EastEnders} = \frac{1.973}{\sqrt{108}} = .19$$

This represents a small to medium effect. Therefore, *EastEnders* had very little effect in creating arguments compared to the control; however, it had more of an effect than *Friends*.

Writing and interpreting the result

For Friedman's ANOVA we need only report the test statistic (which we saw earlier is denoted by χ^2), its degrees of freedom and its significance. So, we could report something like:

- ✓ The number of arguments that couples had was significantly affected by the programme they had just watched, $\chi^2(2) = 7.59$, $p < .05$.

We need to report the follow up tests as well (including their effect sizes):

- ✓ The number of arguments that couples had was significantly affected by the programme they had just watched, $\chi^2(2) = 7.59$, $p < .05$. Pairwise comparisons with adjusted *p*-values showed that watching *EastEnders* significantly increased the number of arguments compared to watching *Friends* ($p = .012$, $r = .24$). However, there were no significant differences in number of arguments when watching *Friends* compared to the *control* programme ($p = 1.00$, $r = -.05$). Finally, *EastEnders* did not significantly increase the number of arguments compared to the control programme; however, there was a small to medium effect ($p = .146$, $r = .19$).

Task 4

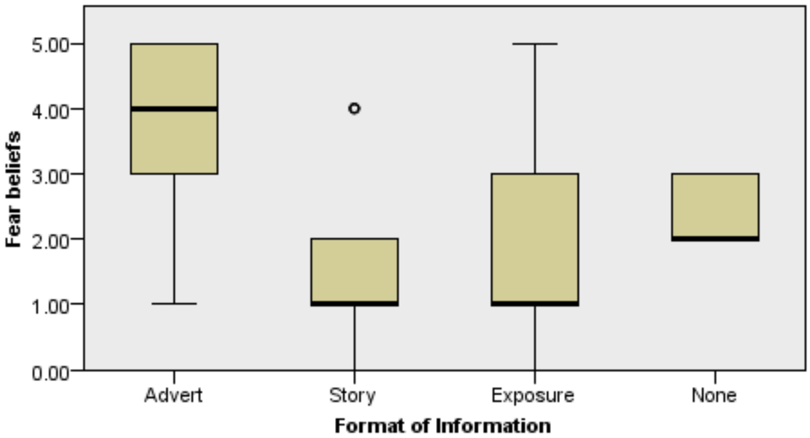
A researcher was interested in trying to prevent coulrophobia (fear of clowns) in children. She decided to do an experiment in which different groups of children (15 in each) were exposed to different forms of positive information about clowns. The first group watched some adverts for McDonald's in which their mascot Ronald McDonald is seen cavorting about with children and going on about how they should love their mums. A second group was told a story about a clown who helped some children when they got lost in a forest (although what on earth a clown was doing in a forest remains a mystery). A third group was entertained by a real clown, who came into the classroom and made balloon animals for the children. A final group acted as a control condition and had nothing done to them at all. The researcher took self-report ratings of how much the children liked clowns, resulting in a score for each child that could range from 0 (not scared of clowns at all) to 5 (very scared of clowns). The data are in the file ***coulrophobia.sav****, access them and conduct a Kruskal–Wallis test.*

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Fear beliefs is the same across categories of Format of Information.	Independent-Samples Kruskal-Wallis Test	.001	Reject the null hypothesis.

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

Independent-Samples Kruskal-Wallis Test

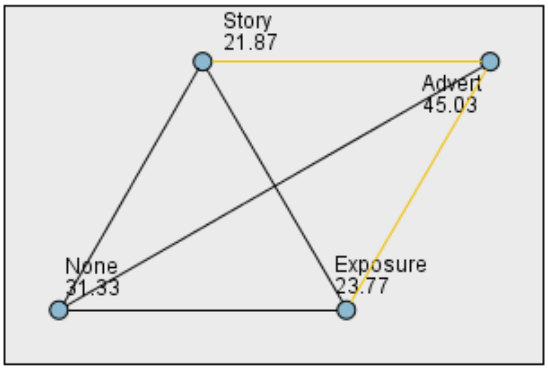


Total N	60
Test Statistic	17.058
Degrees of Freedom	3
Asymptotic Sig. (2-sided test)	.001

1. The test statistic is adjusted for ties.

This table shows this test statistic and its associated degrees of freedom (in this case we had four groups so the degrees of freedom are 4 – 1, or 3), and the significance (which is less than the critical value of .05). Therefore, we could conclude that the type of information presented to the children about clowns significantly affected their fear ratings of clowns. The boxplot in the output above gives us an indication of the direction of the effects, but to see where the significant differences lie we need to look at the follow-up analysis – in this case, the pairwise comparisons.

Pairwise Comparisons of Format of Information



Each node shows the sample average rank of Format of Information.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Story-Exposure	-1.900	6.237	-.305	.761	1.000
Story-None	-9.467	6.237	-1.518	.129	.774
Story-Advert	23.167	6.237	3.714	.000	.001
Exposure-None	-7.567	6.237	-1.213	.225	1.000
Exposure-Advert	21.267	6.237	3.410	.001	.004
None-Advert	13.700	6.237	2.197	.028	.168

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.

Looking at the diagram above, we can see that the test comparing the Story and Advert groups, and the test comparing the Exposure and the Advert groups were significant (yellow connecting lines). However, none of the other comparisons were significant (black connecting lines). The table below the diagram confirms this, and tells us the significance values of the three comparisons. The significance value of the comparison between Exposure and Advert is .004, which is below our criterion of .05. Therefore, we can conclude that hearing a story and exposure to a clown significantly decreased fear beliefs compared to watching the advert (I know the direction of the effects by looking at the boxplot above). There was no significant difference between hearing and exposure on children's fear beliefs. Finally, none of the interventions significantly decreased fear beliefs compared to the control condition.

Calculating an effect size

For the first comparison (Story vs. Exposure) *z* is –.305, and because this is based on comparing two groups each containing 15 observations, we have 30 observations in total. The effect size is, therefore:

$$r_{\text{Story–Exposure}} = \frac{-0.305}{\sqrt{30}} = -.06$$

This represents a very small effect, which tells us that the effect of a story relative to exposure was similar.

For the second comparison (story vs. control) *z* is –1.518, and this was again based on 30 observations. The effect size is, therefore:

$$r_{\text{Story–None}} = \frac{-1.518}{\sqrt{30}} = -.28$$

This represents a small to medium effect. Therefore, although non-significant, the effect of stories relative to the control was a fairly substantive effect.

For the next comparison (story vs. advert) *z* is 3.714, and this was again based on 30 observations. The effect size is, therefore:

$$r_{\text{Story–Advert}} = \frac{3.714}{\sqrt{30}} = .68$$

This represents a large effect. Therefore, the effect of a stories relative to adverts was a substantive effect.

For the next comparison (exposure vs. control) *z* is –1.213, and this was again based on 30 observations. The effect size is, therefore:

$$r_{\text{Exposure–None}} = \frac{-1.213}{\sqrt{30}} = -.22$$

This represents a small effect. Therefore, there was a small effect of exposure relative to the control.

For the next comparison (exposure vs. advert) *z* is 3.410, and this was again based on 30 observations. The effect size is, therefore:

$$r_{\text{Exposure–Advert}} = \frac{3.419}{\sqrt{30}} = .62$$

This represents a large effect. Therefore, the effect of a stories relative to adverts was a substantive effect.

For the final comparison (adverts vs. control) *z* is 2.197, and this was again based on 30 observations. The effect size is, therefore:

$$r_{\text{None–Advert}} = \frac{2.197}{\sqrt{30}} = .40$$

This represents a medium to large effect, Therefore, although non-significant, the effect of adverts relative to the control was a substantive effect.

Writing and interpreting the result

For the Kruskal–Wallis test, we need only report the test statistic (which we saw earlier is denoted by H), its degrees of freedom and its significance. So, we could report something like:

- ✓ Children's fear beliefs about clowns was significantly affected the format of information given to them, $H(3) = 17.06$, $p < .01$.

However, we need to report the follow-up tests as well (including their effect sizes):

- ✓ Children's fear beliefs about clowns was significantly affected the format of information given to them, $H(3) = 17.06$, $p < .01$. Pairwise comparisons with adjusted p-values showed that fear beliefs were significantly higher after the adverts compared to the story, $U = 23.167$, $r = .68$, and exposure, $U = 21.267$, $r = .62$. However, fear beliefs were not significantly different after the stories, $U = -9.467$, *ns*, $r = -.28$, exposure, $U = -7.557$, *ns*, $r = -.22$, or adverts, $U = 13.700$, *ns*, $r = .40$, relative to the control. Finally, fear beliefs were not significantly different after the stories relative to exposure, $U = -.305$, *ns*, $r = -.06$. We can conclude that clown information through adverts, stories and exposure did produce medium-size effects in reducing fear beliefs about clowns compared to the control, but not significantly so (future work with larger samples might be appropriate).

Task 5

Thinking back to Labcoat Leni's Real Research 3.1. Carry out an appropriate test to see whether there was a significant difference between offers made in people listening to Bon Scott compared to those listening to Brian Johnson. Compare your results to those reported by Oxoby (2008). The data are in ***Oxoby (2008) Offers.sav****.*

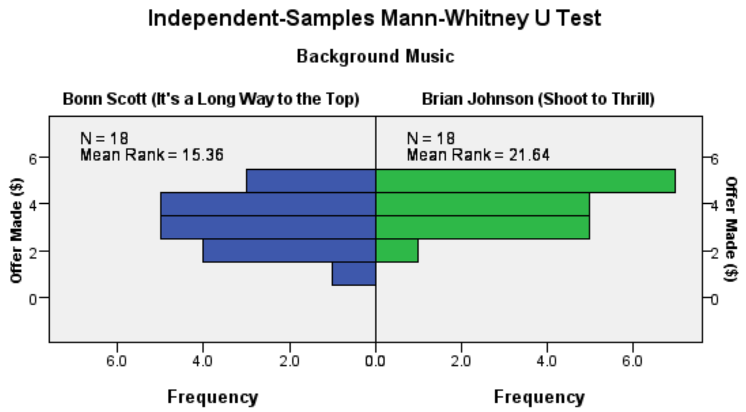
To answer this question we need to conduct a Mann–Whitney test because we want to compare scores in two independent samples: participants who listened to Bon Scott vs. those who listened to Brian Johnson.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Offer Made ($) is the same across categories of Background Music.	Independent-Samples Mann-Whitney U Test	.074[1]	Retain the null hypothesis.

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

[1] Exact significance is displayed for this test.



Total N	36
Mann-Whitney U	218.500
Wilcoxon W	389.500
Test Statistic	218.500
Standard Error	30.542
Standardized Test Statistic	1.850
Asymptotic Sig. (2-sided test)	.064
Exact Sig. (2-sided test)	.074

Calculating the Effect Size

$$r_{\text{Bon Scott–Brian Johnson}} = \frac{1.850}{\sqrt{36}} = .31$$

This represents a medium effect that when listening to Brian Johnson people proposed higher offers than when listening to Bon Scott, suggesting that they preferred Brian Johnson to Bon Scott. However, this effect was not significant, which shows how a fairly substantial effect size can still be non-significant in a small sample.

We could report something like:

- ✓ Offers made by people listening to Bon Scott (*Mdn* = 3.0) were not significantly different from offers by people listening to Brian Johnson (*Mdn* = 4.0), *U* = 218.50, *z* = 1.85, *p* = .074, *r* = .31.

I've reported the median for each condition because this statistic is more appropriate than the mean for non-parametric tests. You'll can get these values by running descriptive statistics, or you could report the mean ranks instead of the median. We could also choose to report Wilcoxon's test rather than the Mann–Whitney *U*-statistic and this would be as follows:

- ✓ Offers made by people listening to Bon Scott (*M* = 15.36) were not significantly different from offers by people listening to Brian Johnson (*M* = 21.64), W_s = 389.50, *z* = 1.85, *p* = .074, *r* = .31.

Task 6

Repeat the analysis above but for the minimum acceptable offer (Chapter 3, Task 3). Remember these data are in the file ***Oxoby (2008) MAO.sav****.*

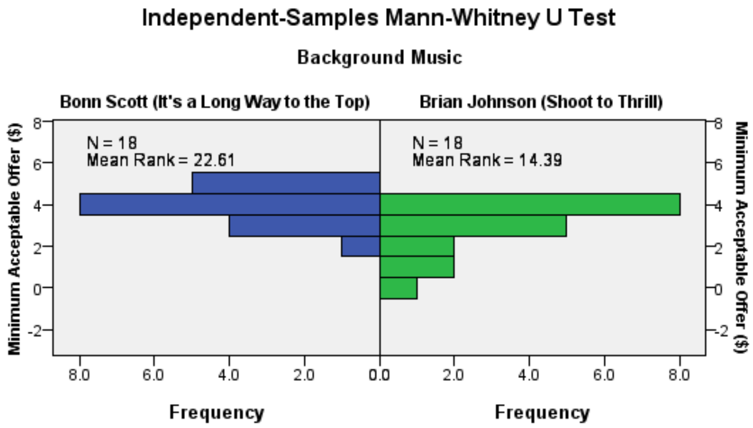
To answer this question we again need to conduct a Mann–Whitney test. This is because we are comparing two independent samples (those who listened to Brian Johnson and those who listened to Bon Scott).

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Minimum Acceptable Offer ($) is the same across categories of Background Music.	Independent-Samples Mann-Whitney U Test	.019[1]	Reject the null hypothesis.

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

[1]Exact significance is displayed for this test.



Total N	36
Mann-Whitney U	88.000
Wilcoxon W	259.000
Test Statistic	88.000
Standard Error	29.882
Standardized Test Statistic	-2.476
Asymptotic Sig. (2-sided test)	.013
Exact Sig. (2-sided test)	.019

Calculating the Effect Size

$$r_{\text{Bon Scott}-\text{Brian Johnson}} = \frac{-2.476}{\sqrt{36}} = -.41$$

This represents a medium effect. looking at the mean ranks in the output above, we can see that people accepted lower offers when listening to Brian Johnson than when listening to Bon Scott.

We could report something like:

- ✓ The minimum acceptable offer was significantly higher in people listening to Bon Scott (*Mdn* = 4.0) than in people listening to Brian Johnson (*Mdn* = 3.0), *U* = 88.00, *z* = –2.48, *p* = .019, *r* = –.41, suggesting that people preferred Brian Johnson to Bon Scott.

I've reported the median for each condition because this statistic is more appropriate than the mean for non-parametric tests. You'll can get these values by running descriptive statistics, or you could report the mean ranks instead of the median. We could also choose to report Wilcoxon's test rather than the Mann–Whitney *U*-statistic and this would be as follows:

- ✓ The minimum acceptable offer was significantly higher in people listening to Bon Scott (*M* = 22.61) than in people listening to Brian Johnson (*M* = 14.39), W_s = 259.00, *z* = –2.48, *p* = .019, *r* = –.41, suggesting that people preferred Brian Johnson to Bon Scott.

Task 7

*Using the data in **Shopping Exercise.sav** (Chapter 3, Task 4)., test whether men and women spent significantly different amounts of time shopping.*

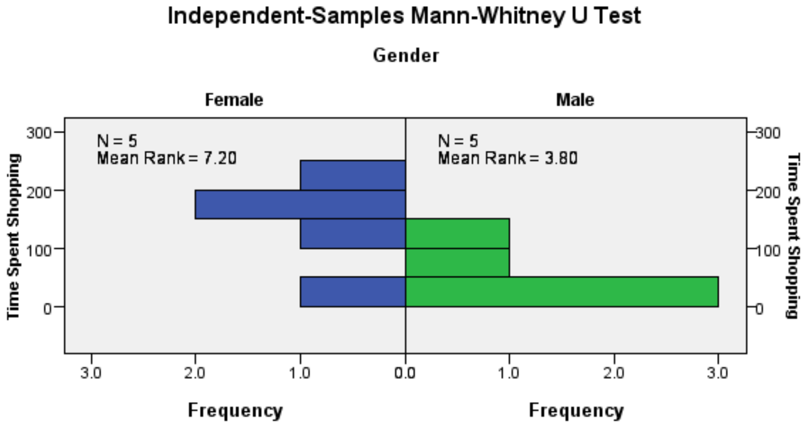
To answer this question we need to conduct a Mann–Whitney test because we are comparing two independent samples (men and women).

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Time Spent Shopping is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	.095[1]	Retain the null hypothesis.

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

[1] Exact significance is displayed for this test.



Total N	10
Mann-Whitney U	21.000
Wilcoxon W	36.000
Test Statistic	21.000
Standard Error	4.787
Standardized Test Statistic	1.776
Asymptotic Sig. (2-sided test)	.076
Exact Sig. (2-sided test)	.095

Calculating the Effect Size

$$r_{\text{men}-\text{women}} = \frac{1.776}{\sqrt{10}} = .56$$

This represents a large effect, which highlights how large effects can be non-significant in small samples. looking at the mean ranks in the output above, we can see that women spent more time shopping than men.

We could report something like:

- ✓ Men (*Mdn* = 37.0) and women (*Mdn* = 160.0) did not significantly differ in the length of time they spent shopping, $U = 21.00$, $z = 1.78$, $p = .095$, $r = .56$.

I've reported the median for each condition because this statistic is more appropriate than the mean for non-parametric tests. You'll can get these values by running descriptive statistics, or you could report the mean ranks instead of the median. We could also choose to report Wilcoxon's test rather than the Mann–Whitney *U*-statistic and this would be as follows:

- ✓ Men (*M* = 3.8) and women (*M* = 7.2) did not significantly differ in the length of time they spent shopping, $W_s = 36.00$, $z = 1.78$, $p = .095$, $r = .56$.

Task 8

Using the same data as in Task 7 above, test whether men and women walked significantly different distances while shopping.

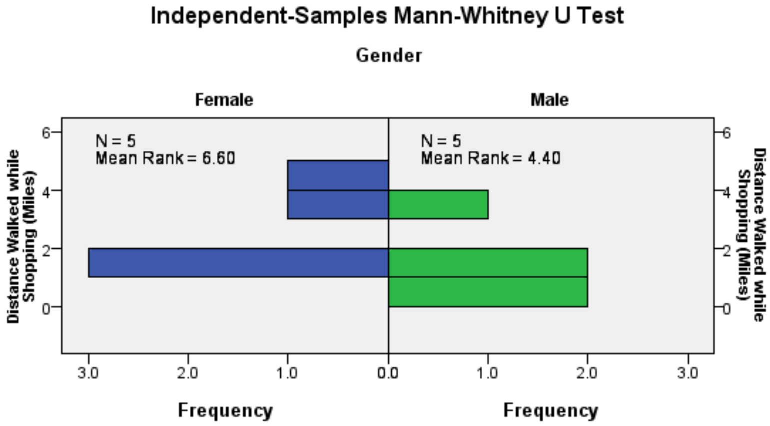
Again, we need to conduct a Mann–Whitney test because – yes, you guessed it – we are once again comparing two independent samples (men and women).

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Distance Walked while Shopping (Miles) is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	.310[1]	Retain the null hypothesis.

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

[1]Exact significance is displayed for this test.



Total N	10
Mann-Whitney U	18.000
Wilcoxon W	33.000
Test Statistic	18.000
Standard Error	4.787
Standardized Test Statistic	1.149
Asymptotic Sig. (2-sided test)	.251
Exact Sig. (2-sided test)	.310

Calculate the Effect Size

$$r_{\text{men}-\text{women}} = \frac{1.149}{\sqrt{10}} = .36$$

This represents a medium effect, which highlights how substantial effects can be non-significant in small samples. Looking at the mean ranks in the output above, we can see that women travelled greater distances while shopping than men (but not significantly so).

We could report something like:

- ✓ Men (*Mdn* = 1.36) and women (*Mdn* = 1.96) did not significantly differ in the distance walked while shopping, *U* = 18.00, *z* = 1.15, *p* = .310, *r* = .36.

I've reported the median for each condition because this statistic is more appropriate than the mean for non-parametric tests. You'll can get these values by running descriptive statistics, or you could report the mean ranks instead of the median. We could also choose to report Wilcoxon's test rather than the Mann–Whitney *U*-statistic and this would be as follows:

- ✓ Men (*M* = 4.4) and women (*M* = 6.6) did not significantly differ in the distance walked while shopping, W_s = 33.00, *z* = 1.15, *p* = .310, *r* = .36.

We can conclude that differences in men and women did produce substantial effects in distance walked and time spent shopping, but not significantly so (future work with larger samples might be appropriate).

Task 9

*Using the data in **Goat or Dog.sav** (Chapter 3, Task 5), test whether people married to goats and dogs differed significantly in their life satisfaction.*

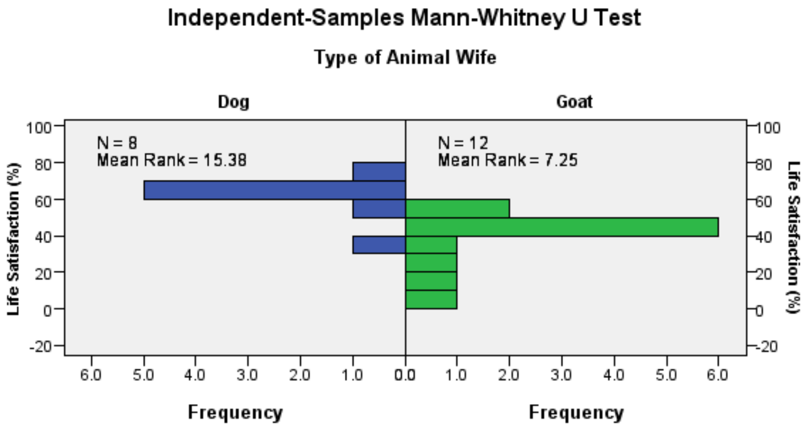
To answer this question we need to run a Mann–Whitney test. The reason for choosing this test is that we are comparing two independent groups (men could be married to a goat or a dog, not both – that would be weird).

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Life Satisfaction (%) is the same across categories of Type of Animal Wife.	Independent-Samples Mann-Whitney U Test	.002[1]	Reject the null hypothesis.

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

[1] Exact significance is displayed for this test.



Total N	20
Mann-Whitney U	87.000
Wilcoxon W	123.000
Test Statistic	87.000
Standard Error	12.952
Standardized Test Statistic	3.011
Asymptotic Sig. (2-sided test)	.003
Exact Sig. (2-sided test)	.002

Calculate the Effect Size

$$r_{\text{goats}-\text{dogs}}\frac{3.011}{\sqrt{20}} = .67$$

This represents a very large effect. Looking at the mean ranks in the output above, we can see that men who were married to dogs had a higher life satisfaction than those married to goats – well, they do say that dogs are man's best friend.

We could report something like:

- ✓ Men who were married to dogs (*Mdn* = 63) had significantly higher levels of life satisfaction than men who were married to goats (*Mdn* = 44), *U* = 87.00, *z* = 3.01, *p* = .002, *r* = .67.

I've reported the median for each condition because this statistic is more appropriate than the mean for non-parametric tests. You'll can get these values by running descriptive statistics, or you could report the mean ranks instead of the median. We could also choose to report Wilcoxon's test rather than the Mann–Whitney *U*-statistic and this would be as follows:

- ✓ Men who were married to dogs (*M* = 15.38) had significantly higher levels of life satisfaction than men who were married to goats (*M* = 7.25), W_s = 123.00, *z* = 3.01, *p* = .002, *r* = .67.

Task 10

*Use the **SPSSExam.sav** (Chapter 5, Task 2) data to test whether students at Sussex and Duncetown universities differed significantly in their SPSS exam scores, their numeracy, their computer literacy, and the number of lectures attended.*

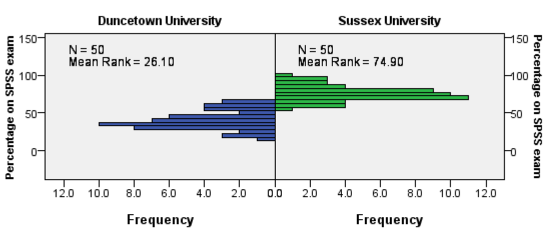
To answer this question we need to run a Mann–Whitney test. The reason for choosing this test is that we are comparing two unrelated groups (students who attended Sussex University and students who attended Duncetown University).

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Percentage on SPSS exam is the same across categories of University.	Independent-Samples Mann-Whitney U Test	.000	Reject the null hypothesis.
2	The distribution of Computer literacy is the same across categories of University.	Independent-Samples Mann-Whitney U Test	.327	Retain the null hypothesis.
3	The distribution of Percentage of lectures attended is the same across categories of University.	Independent-Samples Mann-Whitney U Test	.152	Retain the null hypothesis.
4	The distribution of Numeracy is the same across categories of University.	Independent-Samples Mann-Whitney U Test	.019	Reject the null hypothesis.

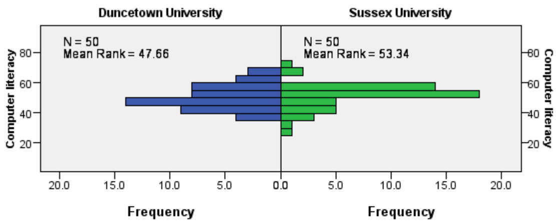
Asymptotic significances are displayed. The significance level is .05.

Independent-Samples Mann-Whitney U Test

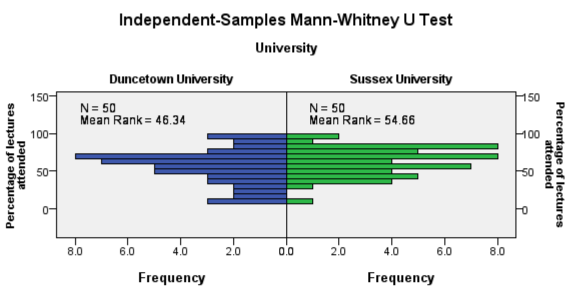


Total N	100
Mann-Whitney U	2,470.000
Wilcoxon W	3,745.000
Test Statistic	2,470.000
Standard Error	145.025
Standardized Test Statistic	8.412
Asymptotic Sig. (2-sided test)	.000

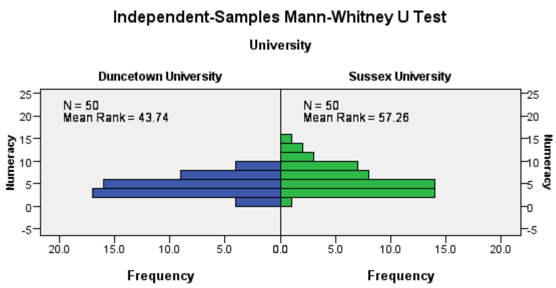
Independent-Samples Mann-Whitney U Test



Total N	100
Mann-Whitney U	1,392.000
Wilcoxon W	2,667.000
Test Statistic	1,392.000
Standard Error	144.827
Standardized Test Statistic	.980
Asymptotic Sig. (2-sided test)	.327



Total N	100
Mann-Whitney U	1,458.000
Wilcoxon W	2,733.000
Test Statistic	1,458.000
Standard Error	145.046
Standardized Test Statistic	1.434
Asymptotic Sig. (2-sided test)	.152



Total N	100
Mann-Whitney U	1,588.000
Wilcoxon W	2,863.000
Test Statistic	1,588.000
Standard Error	143.847
Standardized Test Statistic	2.350
Asymptotic Sig. (2-sided test)	.019

Calculate the Effect Sizes

SPSS exam:

$$r_{\text{Sussex−Duncetown}} = \frac{8.412}{\sqrt{100}} = .84$$

Computer literacy:

$$r_{\text{Sussex-Duncetown}} = \frac{.980}{\sqrt{100}} = .10$$

Percentage of lectures attended:

$$r_{\text{Sussex-Duncetown}} = \frac{1.434}{\sqrt{100}} = .14$$

Numeracy:

$$r_{\text{Sussex-Duncetown}} = \frac{2.35}{\sqrt{100}} = .24$$

We could report something like:

- ✓ Students from the Sussex University (*Mdn* = 75) scored significantly higher on their SPSS exam than students from Duncetown University (*Mdn* = 38), *U* = 2,470.00, *z* = 8.41, *p* = .00, *r* = .84. Sussex students (*Mdn* = 5) were also significantly more numerate than those at Duncetown University (*Mdn* = 4), *U* = 1,588.00, *z* = 2.35, *p* = .019, *r* = .24. However, Sussex students (*Mdn* = 54), were not significantly more computer literate than Duncetown students (*Mdn* = 49), *U* = 1,392.00, *z* = .980, *p* = .327, *r* = .10, nor did Sussex students (*Mdn* = 65.75) attend significantly more lectures than Duncetown students (*Mdn* = 60.50), *U* = 1,458.00, *z* = 1.43, *p* = .152, *r* = .14. Sussex students are just more intelligent, naturally. ☺

Task 11

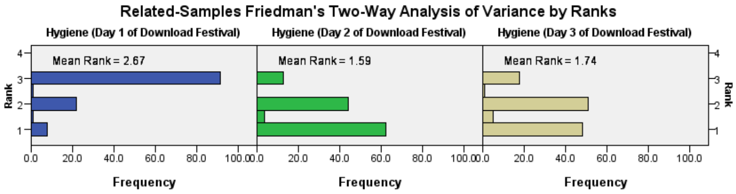
Use the ***DownloadFestival.sav*** *data from Chapter 5 to test whether hygiene levels changed significantly over the three days of the festival.*

To answer this question we need to conduct a Friedman's ANOVA. This is because we want to compare more than two (day 1, day 2 and day 3) related samples (the same participants were used across the three days of the festival).

Hypothesis Test Summary

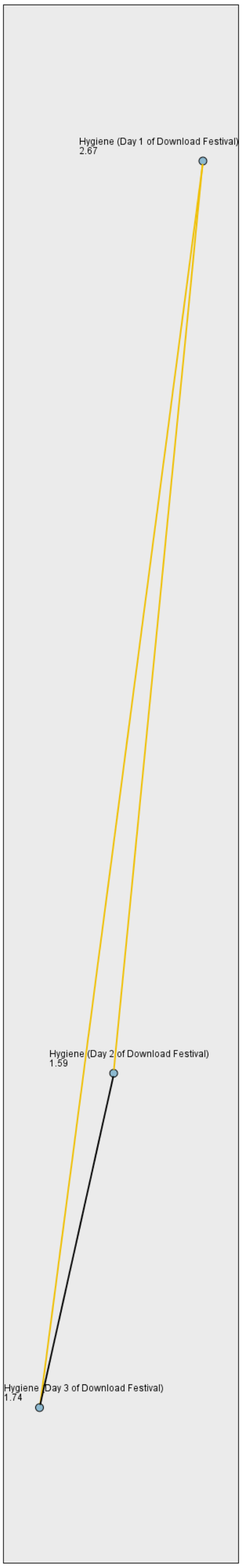
	Null Hypothesis	Test	Sig.	Decision
1	The distributions of Hygiene (Day 1 of Download Festival), Hygiene (Day 2 of Download Festival) and Hygiene (Day 3 of Download Festival) are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	.000	Reject the null hypothesis.

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



Total N	123
Test Statistic	86.535
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.000

Pairwise Comparisons



Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Hygiene (Day 2 of Download Festival)-Hygiene (Day 3 of Download Festival)	-.154	.128	-1.211	.226	.677
Hygiene (Day 2 of Download Festival)-Hygiene (Day 1 of Download Festival)	1.089	.128	8.544	.000	.000
Hygiene (Day 3 of Download Festival)-Hygiene (Day 1 of Download Festival)	.935	.128	7.332	.000	.000

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.

Calculating the Effect Sizes

$$r_{\text{Day 1}-\text{Day 2}} = \frac{8.544}{\sqrt{246}} = .54$$

$$r_{\text{Day 1}-\text{Day 3}} = \frac{7.332}{\sqrt{246}} = .47$$

$$r_{\text{Day 2}-\text{Day 3}} = \frac{-1.211}{\sqrt{246}} = -.08$$

For Friedman's ANOVA we need only report the test statistic, which is denoted by χ^2_F, its degrees of freedom and its significance. We can also report the follow-up tests with their effect sizes. So, we could report something like:

- ✓ The hygiene levels significantly decreased over the three days of the music festival, $\chi^2(2)$= 86.54, *p* = .000. However, pairwise comparisons with adjusted *p*-values revealed that while hygiene scores significantly decreased between days 1 and 2, (*p* = .000, *r* = .54), and days 1 and 3, (*p* = .000, *r* = .47), they did not significantly decrease between days 2 and 3 (*p* = .677, *r* = –.08).