

DR IAN HOCKING

MPSMD2RES: EXERCISES

Contents

0.1	<i>Versions</i>	13
0.2	<i>About this document</i>	13
1	<i>Measures of Central Tendency, Association and Difference</i>	17
1.1	<i>Overview</i>	17
1.2	<i>Objectives</i>	17
1.3	<i>Workshop</i>	18
1.4	<i>Measures of Central Tendency and Variance</i>	18
1.5	<i>Measures of Association</i>	20
1.6	<i>Difference Between Two Related Groups</i>	22
1.7	<i>Follow-up Analyses</i>	24
2	<i>Issues of Significance</i>	27
2.1	<i>Overview</i>	27
2.2	<i>Objectives</i>	27
2.3	<i>Workshop</i>	28
2.4	<i>Effect Size</i>	28
2.5	<i>Power</i>	30

2.6	<i>Confidence Interval</i>	32
2.7	<i>Supporting Materials</i>	33
2.8	<i>Power Table</i>	33
3	<i>One-Way ANOVA</i>	35
3.1	<i>Overview</i>	35
3.2	<i>Objectives</i>	35
3.3	<i>Workshop</i>	35
3.4	<i>Between-Groups ANOVA</i>	36
3.5	<i>Within-Groups ANOVA</i>	46
4	<i>Two-Way Between Groups ANOVA</i>	59
4.1	<i>Overview</i>	59
4.2	<i>Objectives</i>	59
4.3	<i>Workshop</i>	60
4.4	<i>Example 1: Traffic noise and Learning</i>	61
4.5	<i>Example 2: Alcohol and Music on Driving Ability</i>	70
5	<i>Two-Way Repeated and Mixed ANOVA</i>	81
5.1	<i>Overview</i>	81
5.2	<i>Objectives</i>	81
5.3	<i>Workshop</i>	81
5.4	<i>Repeated ANOVA</i>	82
5.5	<i>Mixed ANOVA</i>	97

6	<i>ANOVA Overview</i>	115
6.1	<i>Overview</i>	115
6.2	<i>Objectives</i>	115
6.3	<i>Workshop</i>	115
6.4	<i>Between-Groups ANOVA</i>	116
6.5	<i>Within Groups ANOVA</i>	124
6.6	<i>Effect Sizes</i>	128
7	<i>Advanced Nonparametric Statistics</i>	135
7.1	<i>Overview</i>	135
7.2	<i>Objectives</i>	135
7.3	<i>Workshop</i>	135
7.4	<i>Mann-Whitney Test</i>	136
7.5	<i>Wilcoxon's Matched Pairs Signed-Ranks Test</i>	142
7.6	<i>The Kruskal-Wallis One-way Analysis of Variance</i>	145
7.7	<i>The Friedman's Rank Test for k Correlated Samples</i>	152
8	<i>Correlation and Regression</i>	159
8.1	<i>Overview</i>	159
8.2	<i>Objectives</i>	159
8.3	<i>Workshop</i>	160
8.4	<i>Rescoring Reverse-Phrased Items</i>	160
8.5	<i>Calculating Reliability</i>	163
8.6	<i>Creating a Total Score.</i>	166
8.7	<i>Calculating Pearson's Correlation.</i>	168
8.8	<i>Regression</i>	170

9 *Multiple Regression (1)* 179

 9.1 *Overview* 179

 9.2 *Objectives* 179

 9.3 *Workshop* 179

10 *Multiple Regression (2)* 191

 10.1 *Overview* 191

 10.2 *Objectives* 191

 10.3 *Workshop* 191

11 *Colophon* 207

 11.1 *References* 207

List of Tables

1.2 Visuo-spatial scores for each female participant.	19
1.3 Verbal scores for each female participant.	20
1.4 Verbal scores for each male participant.	21
1.5 Verbal scores for each male participant after training.	23
2.2 Showing mean and standard deviation (SD) in creative problem solving score pre-intervention and post-intervention	28
4.2 The effects of alcohol and music on driving ability.	70
5.2 Before and after ratings for the statistics and criminal psychology courses.	
	83
6.2 Interpreting Effect Size for R	130
8.2 Cronbach's Alpha with associated Internal Consistency	164
9.2 Measures and descriptions for the American city health data	180

List of Figures

2.1 Power Appendix from Howell	34
3.1 SPSS boxplot dialogue	39
3.2 SPSS boxplot definition dialogue	39
3.3 Social activity level by therapy group	40
3.4 One-way ANOVA main dialogue	41
3.5 One-way ANOVA post hoc test dialogue	42
3.6 One-way ANOVA options dialogue	42
3.7 Output—Homogeneity of Variance	43
3.8 Output—ANOVA	44
3.9 Output—Post Hoc Comparisons	45
3.10 SPSS boxplot dialogue	49
3.11 SPSS boxplot definition dialogue	49
3.12 Number of errors by alcohol group	50
3.13 One-way within-groups ANOVA factor definitions	52
3.14 One-way within-groups ANOVA variables entered	53
3.15 Output—Multivariate Tests	53
3.16 Output—Mauchley’s Test	54
3.17 Output—Tests of Within-Groups Effects	55
4.1 SPSS compare means dialogue	62
4.2 Output—compare means output	63
4.3 SPSS Univariate ANOVA dialogue	64
4.4 SPSS Univariate ANOVA dialogue, Options	65
4.5 SPSS Univariate ANOVA dialogue, Profile Plots	66
4.6 Output—Homogeneity of Variance	67

4.7 Output—ANOVA	68
4.8 Output—Profile Plots	69
4.9 Output—Descriptives for error by alcohol group and music group	72
4.10 Output—Estimated Marginal Means	73
4.11 Output—Levene’s Test of Equality of Variances	75
4.12 Syntax—Univariate ANOVA, without SME analyses	76
4.13 Syntax—Univariate ANOVA, including SME analyses	77
4.14 Output—Univariate Tests, Simple Main Effects of Music within Alcohol	78
4.15 Output—Univariate Tests, Simple Main Effects of Alcohol within Music	78
5.1 SPSS Repeated Measures Define Factors Dialogue	86
5.2 SPSS Repeated Measures Main ANOVA Dialogue	87
5.3 SPSS Repeated Measures Main ANOVA Dialogue, completed	88
5.4 SPSS Repeated Measures Plots	88
5.5 SPSS Repeated Measures Options	89
5.6 Output—Mauchly’s Test	90
5.7 Output—Tests of Within-Subjects Effects	91
5.8 Output—Profile Plots	93
5.9 Therapy and Sociability Descriptives	99
5.10 SPSS Repeated Measures Define Factors Dialogue	100
5.11 SPSS Repeated Measures Factor Selection	100
5.12 Output—Plot (Therapy by Time)	102
5.13 Output—Plot (Time by Therapy)	102
5.14 Output—Mauchly’s Test of Sphericity	103
5.15 Output—Tests of Within-Subjects Effects	104
5.16 Output—Tests of Between-Subjects Effects	105
5.17 Repeated Measures ANOVA Dialogue, Pairwise Comparisons for Therapy Group	107
5.18 Output—Pairwise Comparisons for Levels within Therapy Group	108
5.19 Define Cases where Therapy Group = 1, No Therapy	111
5.20 Data View with only Therapy Group = 1, No Therapy	111
6.1 SPSS Explore	118

6.2 SPSS Explore—Plots	119
6.3 SPSS Output—Q-Q Plot for life satisfaction (now) when the educational level is ‘other’	120
6.4 SPSS Output—Tests of Normality	121
6.5 SPSS ANOVA Options	122
6.6 SPSS ANOVA Options, Post Hoc	122
6.7 SPSS Output—Robust Tests of Equality of Means	124
6.8 SPSS Output—One Way ANOVA	133
7.1 SPSS Two-Independent-Samples Tests	139
7.2 SPSS Two-Independent-Samples Tests, Exact Tests	140
7.3 SPSS Output—Mann-Whitney	141
7.4 SPSS Two-Related-Samples Tests	144
7.5 SPSS Output—Wilcoxon	146
7.6 SPSS Tests for Several Independent Samples	149
7.7 SPSS Output—Kruskal-Wallis	150
7.8 SPSS Tests for Several Related Samples	155
7.9 SPSS Output—Friedman	156
8.1 SPSS Recode into Different Variables, Incomplete	161
8.2 SPSS Recode into Different Variables, Complete	162
8.3 SPSS Recode into Different Variables: Old and New Values	163
8.4 SPSS Reliability Analysis	165
8.5 SPSS Reliability Analysis Statistics	165
8.6 SPSS Compute Variable	167
8.7 SPSS Bivariate Correlations	169
8.8 SPSS Linear Regression	171
8.9 SPSS Linear Regression, Statistics	171
8.10 SPSS Linear Regression, Plots	172
8.11 SPSS Output—Model Summary	173
8.12 SPSS Output—ANOVA Table	174
8.13 SPSS Output—Scatterplot	175
8.14 SPSS Output—Coefficients	176
9.1 SPSS Linear Regression Dialogue	182

9.2 SPSS Linear Regression Statistics	182
9.3 SPSS Output—Correlations, excerpt	184
9.4 SPSS Output—Model Summary	184
9.5 SPSS Output—ANOVA	185
9.6 SPSS Output—Coefficients Overall	187
9.7 SPSS Output—Coefficients, Bs and Betas	187
9.8 SPSS Output—Coefficients, Ts, CIs and Correlations	188
10.1 SPSS Multiple Regression dialogue	193
10.2 SPSS Linear Regression, Plots	194
10.3 SPSS Linear Regression, Statistics	195
10.4 SPSS Syntax for Multiple Regression	196
10.5 SPSS Syntax for Multiple Regression, Edited	197
10.6 SPSS Output—Model Summary	198
10.7 SPSS Output—ANOVA	198
10.8 SPSS Output—Coefficients, lefthand section	199
10.9 SPSS Output—Coefficients, righthand section	199
10.10 SPSS Output—Histogram of Frequency against Standardised Residuals	201
10.11 SPSS Output—Normal P-P Plot	202
10.12 SPSS Output—Scatterplot of Residuals against Predicted Values	202
10.13 SPSS Output—Residuals Statistics	203
10.14 SPSS Output—Casewise Diagnostics	203
10.15 SPSS Output—Outlier Statistics	204

0.1 Versions

This document is available in [standard book PDF](#). This document contains hyperlinks to [sections within it](#), [external webpages](#), and email addresses like ian.hocking@canterbury.ac.uk.

0.1.1 Licence

This work is licensed under a [Creative Commons](#) Attribution-NonCommercial-NoDerivatives 4.0 International License. The dark theme is based on [Solarised](#) by Ethan Schoonover.

0.2 About this document

This document is available in different formats for students who may have accessibility requirements. See [Versions](#). The system is still being piloted and I'd be interested in your [feedback](#).

0.2.1 Tasks and Your Research Journal

Use this booklet in conjunction with your own *Research Journal*, where you will record your workings, thoughts, and other comments related to the exercises. Your Research Journal can take any form, but a Word document might be best; you can copy and paste output from SPSS alongside your notes.

(If you're looking at a non-standard, accessible version of this document, some of the formatting below will be simplified.)

- When I ask you to complete a task, like calculate a mean, it will be formatted like this.

- This is what a Research Journal reminder looks like. I'll use these when asking you to make a note.

0.2.2 Other Aspects of this Booklet

- This formatting will be used to highlight something important.

Answer

Here I'll provide answers to questions. Note that this version of the document won't be available until after your workshop.

0.2.3 Mathematics and Statistics Help

If you're not confident in your algebra, which is important for dealing with equations, try this [Introduction to Algebra](#).

0.2.4 Answers

You'll be provided with a second version of this document, containing answers, a few days after your seminar. I'll include SPSS Syntax and possibly SPSS Data files to help you reproduce the correct answers quickly.

When you use menus and dialogue boxes within SPSS to do analyses, SPSS is actually building up a complex command in its native language, syntax, and then running this command. It is feasible for you to access these complex commands yourself. In any dialogue box, the *paste* button will produce the appropriate syntax to do a particular analysis. You can save this syntax as text and run it again at a later date to get the same output. If you want to repeat an analysis quickly, changing bits like variables or type of test, editing syntax is often the best way.

Paste the syntax into an *SPSS syntax window* using *File > New > Syntax*. Highlight the syntax and click the green arrow to make SPSS run the syntax, producing the appropriate output.

It would be a good idea to get used to SPSS Syntax, though I'm not expecting you to use it instead of the graphical, 'point and click'

interface.

1

Measures of Central Tendency, Association and Difference

1.1 Overview

Data files required None

Booklet Version 1.0

1.2 Objectives

This workshop is designed to get you back up to speed with some basics. It revises some important concepts from MPSMD1RES. If I were you, I would keep your notes from that module handy while you're learning new things on MPSMD2RES. Remember that you should still have access to the MPSMD1RES module Blackboard.

- In this booklet, the section **8 mean, median and standard deviation** will reacquaint you with measures of central tendency—which tell us something about the central point in a given set of data—as well as **standard deviation**, which tells us something about the dispersion of a set of data;
- You'll look at the r statistic in **9 simple correlation**;
- The section on **8 between-subjects differences** will remind

you how to test for differences between two groups of data that are not correlated;

- Finally, the section [8 within-subjects differences](#) will take you through testing for differences between two groups of data that *are* correlated.

1.3 Workshop

For these exercises, use [SPSS](#).

1.4 Measures of Central Tendency and Variance

Central tendency is all about describing the central or typical value of a set of data. It attempts to find a conceptual middle. Typically, we talk about the mean, mode and median.

By contrast, we also want a measure of dispersion. This tells us, in a value, something about the spread of data. One such value is standard deviation.

Calculating the (arithmetic) mean and median goes like this:

- Mean: Add all the values together and divide by the number of values
- Median: Order the values and take the middle one

For standard deviation (in its simplest form):

- Calculate the mean
- Subtract each data point from the mean, and square these values.
Call this *deviation*
- Take the mean of the *deviation*. This is called variance
- Take the square root of the *variance*. This is called standard deviation

Look at the following example. A group of 18 female participants takes part in a visuo-spatial memory task and produces the following scores:

Table 1.2: Visuo-spatial scores for each female participant.

<i>Participant number</i>	1	2	3	4	5	6	7	8	9
<i>Visual Score</i>	12	5	20	20	25	19	20	19	15
<i>Participant number</i>	10	11	12	13	14	15	16	17	18
<i>Visual Score</i>	19	20	6	25	20	20	5	8	19

In summarising these data, we want to get a sense of central tendency and dispersion of the sample.

1. Using SPSS, compute the mean, median and standard deviation scores.

1. Note these in your Research Journal

Answer

Mean is *16.50*

Median is *19*

Standard deviation is *6.47*

The SPSS Syntax for these  is as follows:

```
EXAMINE VARIABLES=visuoSpatialScore
/STATISTICS DESCRIPTIVES
/NOTOTAL.
```

1.5 Measures of Association

Lets look at an effect measure called correlation. The female participants from the memory task also completed a verbal comprehension test and produced the following range of scores:

Table 1.3: Verbal scores for each female participant.

<i>Participant number</i>	1	2	3	4	5	6	7	8	9
<i>Verbal Score</i>	68	82	45	45	32	52	45	52	65
<i>Participant number</i>	10	11	12	13	14	15	16	17	18
<i>Verbal Score</i>	55	45	78	33	46	47	81	66	52

2. Use SPSS to conduct an analysis assessing whether there is an association in performance on the two tasks (visuo-spatial and verbal).

2. Make a note of your answer using APA style.

Answer

There is a significant negative correlation between performance on the visual task and the verbal task, $r(18) = -.972, p < .001$.

The SPSS Syntax for these  data is as follows:

CORRELATIONS

```
/VARIABLES=verbalScore visuoSpatialScore
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
```

3. What conclusion(s) can you draw from your answer?

Answer

That females who are good at the visual task are poor at the verbal task.

1.5.1 *Difference Between Two Groups*

We've just looked at correlation, which, using the sample, tries to estimate what the population correlation might be.

Now we'll consider difference.

When comparing the means of two groups of data, you will always find that they are not the same. This means that you have a difference between your two samples. This difference is, in a sense, trivial, because you're very likely to find it. We are much more interested in when this mean difference is said to be significant. Why? Because this tells us there is probably a true difference between the two groups. In other words, while we always have a sample difference, we only sometimes have a population difference. The population difference is the one you should declare in your results. For many researchers, writing 'A difference was found between the two groups' is synonymous with 'A *population* difference was found between the two groups'.

Now, researchers thought that gender may be influencing performance on the verbal task so they had a separate group of 18 males complete the verbal comprehension test, producing the following scores:

Table 1.4: Verbal scores for each male participant.

<i>Participant number</i>	1	2	3	4	5	6	7	8	9
<i>Verbal Score</i>	62	71	38	41	29	35	28	51	58
<i>Participant number</i>	10	11	12	13	14	15	16	17	18
<i>Verbal Score</i>	53	29	49	23	38	28	65	54	27

3. Conduct a statistical test to ascertain whether there is a significant difference in performance of the females vs. the males on the verbal comprehension test.

4. Note your results in APA format.

Answer

Females scored significantly higher on the verbal comprehension test compared to the males (means of 54.9 and 43.2 respectively), $t(34) = 2.326, p < .05$.

The SPSS Syntax for these  is as follows:

```
T-TEST GROUPS=gender(1 2)
/MISSING=ANALYSIS
/VARIABLES=verbalScore
/CRITERIA=CI(.95).
```

1.6 Difference Between Two Related Groups

In the above example, the two groups had different people in each.

We'll now turn to another example where the same people were tested

on two occasions. Because the same people are now involved in both conditions, we can expect the data to be correlated, since each person will, to some extent, have stable individual characteristics.

The researchers, surprised at the poor performance of the males on the verbal task, took them away and gave them a week's training on verbal skill. Following a delay of one week they tested the men again and they produced the following scores:

Table 1.5: Verbal scores for each male participant after training.

<i>Participant</i>	1	2	3	4	5	6	7	8	9
<i>number</i>									
<i>Verbal Score</i>	68	75	59	58	51	62	59	62	59
<i>Participant</i>	10	11	12	13	14	15	16	17	18
<i>number</i>									
<i>Verbal Score</i>	55	49	52	65	52	45	68	59	45

4. Conduct a statistical test to ascertain whether there is any significant improvement in the verbal scores of the males.

5. Note your results in APA format.

Answer

Males scores on the verbal comprehension test improved significantly from time 1 to time 2 (means of 43.2 and 57.9 respectively), $t(18) = 5.438$, $p < .001$.

The SPSS Syntax for these  data is as follows:

```
T-TEST PAIRS=verbalScoreMenT1 WITH verbalScoreMenT2 (PAIRED)
```

```
/CRITERIA=CI(.9500)
/MISSING=ANALYSIS.
```

1.7 Follow-up Analyses

If you've managed the above and want to try something more advanced outside of the seminar, keep reading.

A1. What are *skewness* and *kurtosis*?

Answer

Skewness is a value representing (i) the degree of asymmetry and (ii) direction of asymmetry in a distribution. For example, a skewness of 0 is a symmetrical distribution. Negative skew pushes the tail of the distribution to the left. Positive skew pushes it towards the right.

Kurtosis is the degree of peakedness or flatness of a distribution.

A1. Calculate skewness and kurtosis for the data on the original 18 females undertaking the visuo-spatial memory task in the section

[8 measures of central tendency](#)

Answer

Skewness -0.8

Kurtosis -0.546

The SPSS syntax is:

```
EXAMINE VARIABLES=visuoSpatialScore
```

```
/STATISTICS DESCRIPTIVES
/NOTOTAL.
```

A2. What do the skewness and kurtosis tell us about our sample?

Answer

The thing to remember about measures of *skew* and *kurtosis* is that they relate to our sample. Even if our sample is non-normal, this doesn't mean that the population we are sampling from is non-normal.

Interpreting skew (following Bulmer, M. G. 1979. *Principles of Statistics*. Dover.):

- If skewness is less than -1 or greater than $+1$, the distribution is highly skewed
- If skewness is between -1 and -0.5 or between 0.5 and 1 , the distribution is moderately skewed
- If skewness is between -0.5 and 0.5 , the distribution is approximately symmetric

In this case, our skew is moderate.

Interpreting kurtosis:

- When kurtosis is around 3, the distribution is approximately normal
- When kurtosis is less than 3, it starts to get *platykurtic* (flat)
- When kurtosis is greater than 3, it begins to get *leptokurtic* (peaky)

Our sample, then, is reasonably flat.

2

Issues of Significance

2.1 Overview

Data files required None

Booklet Version 1.0

2.2 Objectives

This workshop is designed to build upon the basic concepts of central tendency, difference, and relationship that you revised last week.

This week, you will not need to use SPSS. Use pen and paper (or Word) and a calculator.

- In the section **Effect Size**, you'll compute a standard effect size by hand;
- Next, in **Power**, you'll perform a power calculation to determine the likelihood that the effect of an intervention will be successfully detected; you'll also work out how many participants you'll need for a given level of power;
- Lastly, in **Confidence Interval**, which you might not have time to do in the seminar, you will look at estimating the mean of a population (with a confidence of 95%)

2.3 Workshop

2.4 Effect Size

The standard effect size is an important, portable measure of difference or relationship.

A researcher creates three interventions that are supposed to help people become better problem solvers: intervention A, intervention B and intervention C. She trains three separate groups of 15 people on these interventions and then runs a pre- vs. post-intervention analysis on their creative problem solving scores. See Table 2.2.

Table 2.2: Showing mean and standard deviation (SD) in creative problem solving score pre-intervention and post-intervention

	Intervention A	..	Intervention B	..	Intervention C	..
	Mean	SD	Mean	SD	Mean	SD
Pre	56.6	7.5	62.8	6.5	72.0	6.9
Post	62.5	7.7	75.8	6.8	74.1	7.5

She is interested in calculating the effect size for each intervention.

To do this, she uses Equation 2.1 to find Cohen's D^1 :

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s} \quad (2.1)$$

Effect size can be thought of as the difference between the two means, but why do we divide by the standard deviation? The difference between the means is a number that won't be comparable to other sets of data because it is expressed in the units of the scale used for the experiment. When we divide it by standard deviation, a standard measure, we're effectively transforming to standard, or 'universal', units. This is not much different from converting a score like 22 out of

¹ Technically, because this is a *repeated* design—the same participants must be used the *pre* and *post* conditions—you should use a more complicated calculation for the standard deviation that accounts for the correlation between samples. However, the simpler calculation is still useful for learning about power.

43 (where 22 only makes sense as a score if you know that the maximum is 43) to a percentage: $(100 / 43) * 22 = 51\%$. Because everyone knows a percentage is a proportion of 100, percentages are portable and comparable. In this way, so is Cohen's D (and related standard measures of effect size).

1. Use Equation 2.1 to work out the effect size of each of the interventions (A, B and C). One thing to remember: you want the absolute difference between means irrespective of its sign, so make sure the difference itself is a positive number.

Answer

Intervention A

- $(56.6 - 62.5) / ((7.5 + 7.7)/2) = -5.9 / 7.6 = 5.9 / 7.6 = 0.77$

Intervention B

- $(62.8 - 75.8) / ((6.5 + 6.8)/2) = -13 / 6.65 = 13 / 6.65 = 1.95$

Intervention C

- $(72 - 74.1) / ((6.9 + 7.5)/2) = -2.1 / 7.2 = 2.1 / 7.2 = 0.29$

Having performed these calculations, you might have noticed something about the relationship between the mean 'pooled' standard deviation and the difference between the group means. Because this is a fraction, you divide the bottom by the top when converting it to a decimal. So, as the pooled standard deviation gets smaller (i.e. variability is reduced), and as the mean difference gets bigger, the effect size gets larger. The next time you look at a table of means and stan-

dard deviations, note the ratio between the pooled standard deviation and the mean differences; you can get a rough idea of what the effect size will be, and how seriously to take the result.

1. According to Cohen's (1988) guidelines, how would you interpret the size of these effects? (The guidelines are include in this week's lecture slides.)

Answer

The guidelines are:

Small = 0.2

Medium = 0.5

Large = 0.8

Intervention A

- 0.77: medium-to-large

Intervention B

- 1.95: very large

Intervention C

- 0.29: small

2.5 Power

The same researcher decides she would like to know the power level of the test conducted on intervention C. She will use Equation 2.2 to identify δ , which she'll then look up in a [table](#) to determine the power level:

$$\delta = \gamma\sqrt{N} \quad (2.2)$$

Note Equation 2.2 uses the term γ (*gamma*) for effect size, which is the equivalent of Cohen's D in this case.

2. Calculate the power for intervention C. Remember that you'll need to look up the value of δ in the [go table from Howell](#). Assume that your alpha level is 5% and your test is two-tailed.

Answer

$$\delta = 0.29 * \text{square root of } 15 = 0.29 * 3.87 = 1.12$$

Looking in Appendix Power in Howell (1992) you can see that a δ of 1.10 (we are rounding to the nearest value) with alpha of 0.05 (two-tailed) gives a power value of .20, which is a 20% chance of finding an effect. This is quite low power.

The researcher wants to run intervention C again. This time, she is interested in working out beforehand how many participants she would need for a power level of 80%—in other words, she wants an 80% likelihood that she will detect the effect of intervention C if it exists.

3. Look up a power level of .8 in the [go table from Howell](#) for a two-tailed test with an alpha of 0.05. What is the δ required?

Answer

$$\delta \text{ is } 2.8$$

4. Now that you have the δ , it can be plugged into Equation 2.2. However, what we really want to do is solve this equation for N , so that we can determine how many people to sample. We can use normal algebraic shifting to express Equation 2.2 as Equation 2.3:

$$N = \frac{\delta^2}{\gamma^2} \quad (2.3)$$

(Remember that γ is Cohen's D in this case.)

Use the above equation to determine how many participants (N) are needed.

Answer

$$N = (2.80 * 2.80) / (0.29 * 0.29) = 7.84 / 0.0841 = 93.22$$

Since we can't have .22 of a person, we can round this up to 94.

2.6 Confidence Interval

This is a bonus section. You might not have time to do it in the seminar, but it will be useful practise to do it afterwards.

The previous seminar, you looked at data involving verbal scores produced by male and female participants. We'll now return to these data.

5. Download these data

As before, run the T-test.

Answer

- Select *Analyse - Compare Means - Independent Samples T-test*
- Set your *Test Variable(s)* as *Verbal score*
- Set your *Grouping Variable* as *gender*
- Set *Define groups* as *User specified values*, with *Group 1* as *1* and *Group 2* as *2*

- Hit *OK*

2. You'll see that your output contains the lower and upper bound (i.e. the range, from the lower edge to the higher edge) of the confidence interval.

Note down the test results in APA style, including the confidence interval. (You'll see an example of how to do this in this week's lecture slides.)

Answer

$$t(34) = 2.32, p < .05, 95\% \text{ CI } [1.47, 21.86]$$

3. In plain English, explain what this confidence interval means.

Answer

This means, precisely, that if we were to repeat the T-test many times, the difference between the two means would be between 1.47 and 21.86 95% of the time.

In this booklet, you've worked through effect size, power, and the confidence interval.

2.7 Supporting Materials

2.8 Power Table

If the graphic of the power table is too small, [3 download a larger one](#).

Table D.5 Power as a Function of δ and Significance Level <i>(a) (Source: The entries in this table were computed by the author.)</i>	δ	Alpha for Two-Tailed Test			
		.10	.05	.02	.01
	1.00	0.26	0.17	0.09	0.06
	1.10	0.29	0.20	0.11	0.07
	1.20	0.33	0.22	0.13	0.08
	1.30	0.37	0.26	0.15	0.10
	1.40	0.40	0.29	0.18	0.12
	1.50	0.44	0.32	0.20	0.14
	1.60	0.48	0.36	0.23	0.17
	1.70	0.52	0.40	0.27	0.19
	1.80	0.56	0.44	0.30	0.22
	1.90	0.60	0.48	0.34	0.25
	2.00	0.64	0.52	0.37	0.28
	2.10	0.68	0.56	0.41	0.32
	2.20	0.71	0.60	0.45	0.35
	2.30	0.74	0.63	0.49	0.39
	2.40	0.78	0.67	0.53	0.43
	2.50	0.80	0.71	0.57	0.47
	2.60	0.83	0.74	0.61	0.51
	2.70	0.85	0.77	0.65	0.55
	2.80	0.88	0.80	0.68	0.59
	2.90	0.90	0.83	0.72	0.63
	3.00	0.91	0.85	0.75	0.66
	3.10	0.93	0.87	0.78	0.70
	3.20	0.94	0.89	0.81	0.73
	3.30	0.95	0.91	0.84	0.77
	3.40	0.96	0.93	0.86	0.80
	3.50	0.97	0.94	0.88	0.82
	3.60	0.98	0.95	0.90	0.85
	3.70	0.98	0.96	0.92	0.87
	3.80	0.98	0.97	0.93	0.89
	3.90	0.99	0.97	0.94	0.91
	4.00	0.99	0.98	0.95	0.92
	4.10	0.99	0.98	0.96	0.94
	4.20	...	0.99	0.97	0.95
	4.30	...	0.99	0.98	0.96
	4.40	...	0.99	0.98	0.97
	4.50	...	0.99	0.99	0.97
	4.60	0.99	0.98
	4.70	0.99	0.98
	4.80	0.99	0.99
	4.90	0.99
	5.00	0.99

Figure 2.1: Power Appendix from Howell

3

One-Way ANOVA

3.1 Overview

Data files required  [Activity after therapy](#),
 [Driving errors after alcohol](#)

Booklet Version 1.0

3.2 Objectives

This workshop introduces you to the Analysis of Variance (ANOVA), which is a statistical technique designed to help analyse designs with multiple groups.

- You'll begin by performing an ANOVA on group data where *different* people have contributed data to each group. This is known as a between-groups ANOVA analysis (sometimes a between-subjects analysis).
- You'll then carry out a second ANOVA using data where the *same* people have contributed data to each group. This is a within-groups ANOVA or a within-subjects ANOVA.

3.3 Workshop

Analysis of variance or ‘ANOVA’ is a statistical technique that allows us to look at the differences between mean scores when there are more

than two conditions. Basically, it compares the variance between the groups with the variance within the groups. The bigger the differences between the groups compared to within the groups, the more likely we are to get a significant result.

3.4 Between-Groups ANOVA

We'll begin with a one-way between-groups ANOVA in SPSS, and interpret the result.

- *One-way* means that the ANOVA has one ‘factor’, which translates as ‘basis on which we have created groups’. Typically, it is the thing we have manipulated.
- This is a *between* or *independent* ANOVA because the people in the groups are different. In other words, the group data are not correlated.

Answer

[GO Download the complete SPSS syntax](#)

[GO Download the complete SPSS output](#)

3.4.1 Our Scenario

A clinical psychologist was interested in the effect of ‘sunshine therapy’ on the activity levels of depressed patients. The therapy involved sitting in an empty room with artificial sunlight filtered through. She took a group of 300 depressed out-patients at a psychiatric hospital and split them into three sub-groups. She gave one sub-group a 3-hour session of sunshine therapy; the second sub-group received 3 hours of traditional Cognitive Behavioural Therapy (CBT); the third group received no therapy. After a series of sessions all the participants got together for an hour to socialise and support each other. During this session, they were observed and the amount of time during which each patient engaged in social activities was recorded.

3.4.2 Step One: Obtain the Data

1. Open  Activity after therapy.

Note that you have two columns. The first is ‘group’ and the second ‘activity’. ‘Group’ represents type of therapy (no therapy, CBT, and sunshine). ‘Activity’ represents social activity level following the treatment.

3.4.3 Step Two: Run the Descriptive Statistics

Before we carry out the ANOVA (or any other statistical test) it is always good practice to get an idea of what is going on with the data. So, calculate the scores for the three groups. Whereas before you’ve used the *Descriptives* function, we need to tell SPSS that we’d like means for one variable (‘activity’) as classified by the value of another (‘group’).

2. Go to *Analyze > Compare means > Means*. Put the dependent variable in ‘dependent’ box and the independent variable into the ‘independent’ box. Now click *OK*.

1. What are the mean activity ratings for the three groups?

Answer

- *No therapy:* 1
 - *CBT:* 2.8
 - *Sunshine therapy:* 2.1
-

SPSS Syntax:

```
MEANS TABLES=activity BY group
/CELLS=MEAN COUNT STDDEV.
```

3.4.4 Step Three: Get a Useful Picture of the Means

Box plots are very useful for getting a sense of what our grouped data *looks* like.

From [Wikipedia](#):

In descriptive statistics, a box plot or boxplot is a convenient way of graphically depicting groups of numerical data through their quartiles.

Box plots may also have lines extending vertically from the boxes (whiskers) indicating variability outside the upper and lower quartiles, hence the terms box-and-whisker plot and box-and-whisker diagram.

Boxplots can be constructed by going to *Graphs > Legacy Dialogues > Boxplot*. For this kind of plot, make sure *Simple* and *Summaries for groups of cases* are selected, like Figure 3.1.

When you've clicked *OK*, you can specify the dependent variable as *Variable* and the grouping variable as *Category Axis*, like Figure 3.2

3. Using the instructions above, create boxplots for the activity ratings within each group by going to *Graphs > Legacy > Boxplot*. Make sure that your output looks like Figure 3.3.

Answer

SPSS syntax:

```
EXAMINE VARIABLES=activity BY group
/PLOT=BOXPLOT
/STATISTICS=NONE
/NOTOTAL.
```

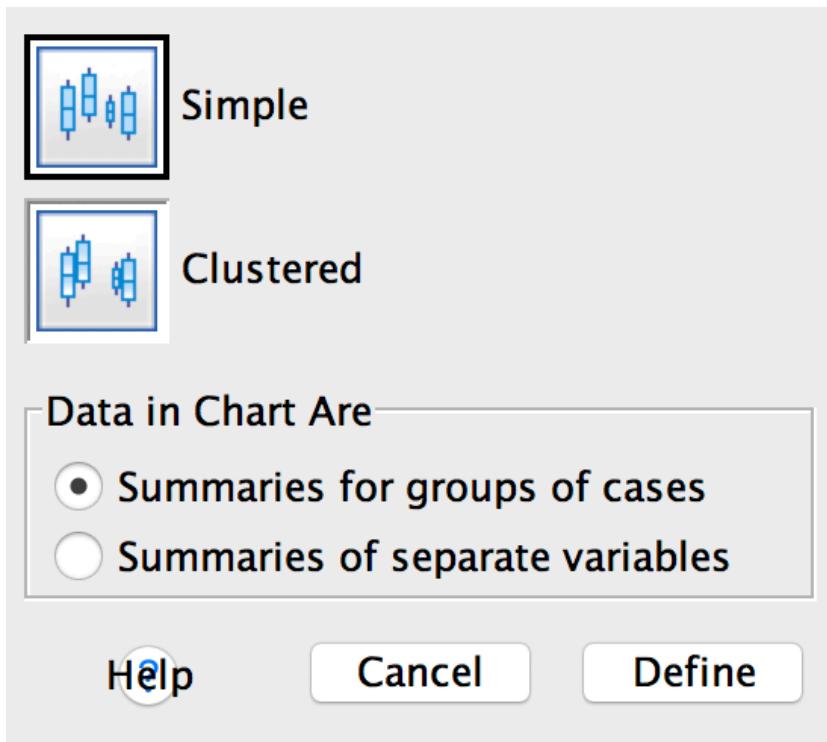


Figure 3.1: SPSS boxplot dialogue

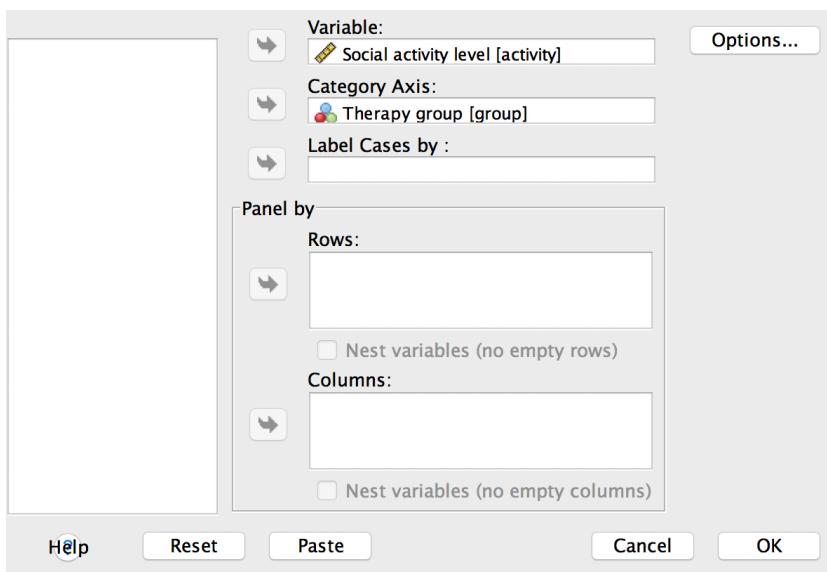


Figure 3.2: SPSS boxplot definition dialogue

Social activity level

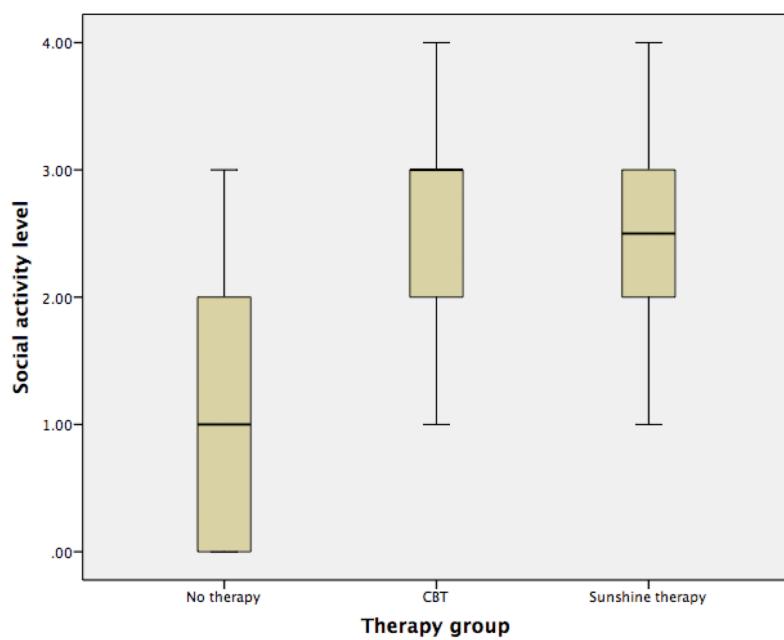


Figure 3.3: Social activity level by therapy group

3.4.5 Step Four: Set up the ANOVA

We can see that there are differences in activity between the groups. We're more interested in whether these differences are true. In other words, do the data better fit a null distribution (of no difference) or an alternative distribution (where the effect is true)? The Analysis of Variance can tell us this.

- Like any statistical test, the ANOVA requires certain conditions to be in place. We'll talk more about these later. Right now, I want you to experience producing the ANOVA; we'll return again to the important skill of knowing whether or not an ANOVA is appropriate. So, for the time being, we'll focus on checking just one assumption: *the homogeneity of variance*.
- Similarly, to keep things simple, we won't talk about *effect size* for now.

Before you run the ANOVA, read through the following to get an overall sense of how you do it.

- First, you'll go to *Analyze > Compare means > One-way ANOVA*, then put the dependent variable into the 'dependent' box and the independent variable into the 'factor' box (Figure 3.4)

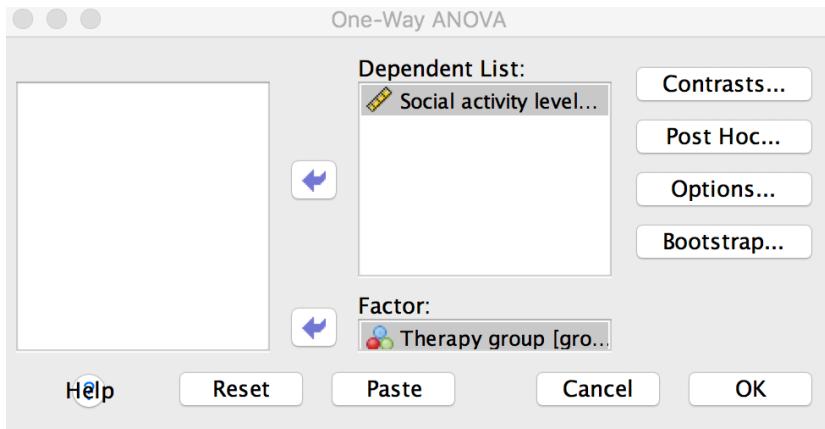


Figure 3.4: One-way ANOVA main dialogue

- To see exactly where the differences lie, if there are any, between the groups are you need to tell SPSS to carry out some *post hoc* tests. So you'll click on *post hoc* and select *Tukey* (Figure 3.5), then click *continue*.
- To ensure that we have satisfied the homogeneity of variance assumption, you'll click on *Options* and select *homogeneity of variance test* (Figure 3.6).
- Then you'll click on *continue* and *OK*.

4. Run the ANOVA according to the instructions above.

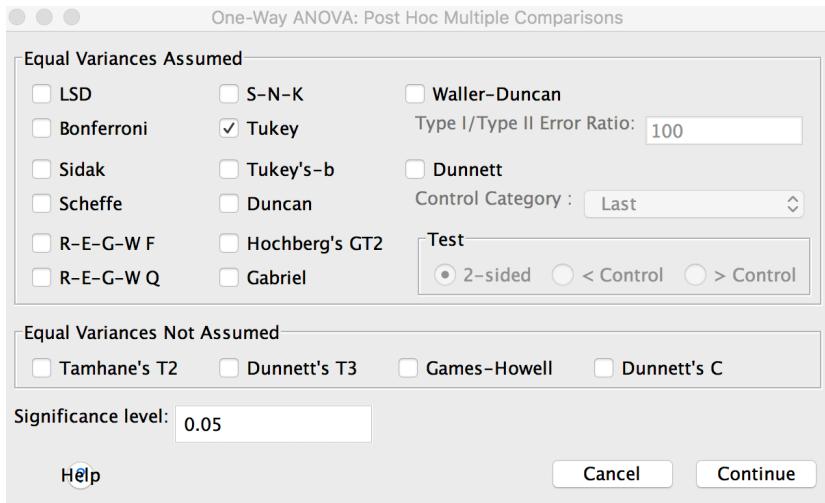


Figure 3.5: One-way ANOVA post hoc test dialogue

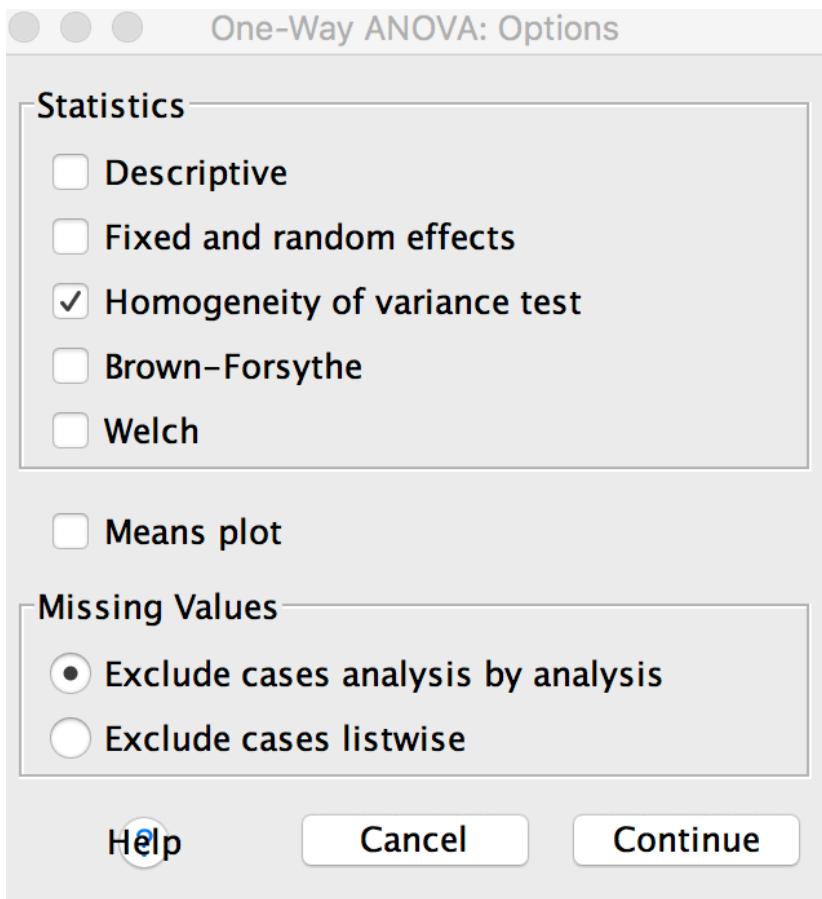


Figure 3.6: One-way ANOVA options dialogue

That's a lot of clicking. Remember that you can click *paste* and have SPSS produce the 'syntax' for running the ANOVA as you've just configured it. If you want to run it again with different options or variables, this is often the more efficient way of making changes.

Answer

For completeness, the SPSS syntax is:

```
ONEWAY activity BY group
/STATISTICS HOMOGENEITY
/MISSING ANALYSIS
/POSTHOC=TUKEY ALPHA(0.05).
```

3.4.6 Step Five: Interpret the Output

You should get the same output as I have below. Let's see first of all whether we have satisfied the assumption of homogeneity of variance. The first table in the output, Figure 3.7, compares the variance within each group against the others. If Levene's test is significant then, just like with T-tests, the assumption is violated and the result of the ANOVA may not be reliable. If it is non-significant, then the assumption is met and the result of the ANOVA should be reliable.

Test of Homogeneity of Variances

Social activity level

Levene Statistic	df1	df2	Sig.
.136	2	27	.873

Figure 3.7: Output—Homogeneity of Variance

2. Have we met the homogeneity of variance assumption? Explain why or why not.

Answer

We have indeed met it because Levene's test returns a *p* value of .873. This indicates that the variances do not differ significantly between groups.

The next table on the output, the 'ANOVA' table shows us whether the difference between the groups is statistically significant or not. Have a look at your own table (which should resemble Figure 3.8) and answer the next question in your Journal.

ANOVA

Social activity level

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	18.600	2	9.300	10.419	.000
Within Groups	24.100	27	.893		
Total	42.700	29			

Figure 3.8: Output—ANOVA

3. Is there a significant effect of the type of therapy a person receives on their subsequent activity levels? Write out your interpretation in plain English and report the ANOVA results in APA style.
- Be sure to include the F ratio itself, degrees of freedom for between and within groups, mean squared error, and the significance.

Answer

Yes, there is a significant effect of group, $F(2, 27) = 10.419$, $MSE = 0.89$, $p < .001$.

3.4.7 Follow-up Comparisons

The final table in the output, the ‘Multiple Comparisons’ table (Figure 3.9), tells us the result of the *post hoc* comparisons.

Post hoc comparisons can help us determine which of the groups differ from one another. This is important because the ANOVA tells us only that there is a difference *somewhere* between the groups (i.e. between at least two). Post hoc paired comparisons look more deeply into the groups.

Multiple Comparisons						
				95% Confidence Interval		
(I) Therapy group	(J) Therapy group	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
	No therapy	-1.80000*	.42251	.001	-2.8476	-.7524
CBT	CBT	-1.50000*	.42251	.004	-2.5476	-.4524
	Sunshine therapy	1.80000*	.42251	.001	.7524	2.8476
Sunshine therapy	No therapy	.30000	.42251	.760	-.7476	1.3476
	CBT	1.50000*	.42251	.004	.4524	2.5476
		-.30000	.42251	.760	-1.3476	.7476

*. The mean difference is significant at the 0.05 level.

Figure 3.9: Output—Post Hoc Comparisons

4. What conclusions can you draw from the *post hoc* tests?

Answer

- No therapy vs. CBT groups are significantly different at $p < .05$. Judging by the means, CBT leads to higher activity than no therapy.
- No therapy vs. Sunshine therapy groups are significantly different at $p < .05$. The means suggest, likewise, that Sunshine therapy leads to higher activity.
- CBT vs. Sunshine therapy are not, however, significantly different from one another. So our sample suggests these are equally effective in terms of their impact on social activity.

- Family-wise error is the problem of an increasing false positive rate as you perform multiple comparisons.
5. Why don't we need to make a family-wise error correction for the post-hoc tests?

Answer

Such tests are already conservative, i.e. have weaker statistical power, than planned comparisons, which means that they are corrected for the family-wise error rate.

3.5 Within-Groups ANOVA

We'll continue now with a one-way within-groups ANOVA using SPSS, and once again we will interpret the results.

- Remember, *one-way* means that the ANOVA has one 'factor', which translates as 'basis on which we have created groups'. Typically, it is the thing we have manipulated.
- This is a *within* or *repeated* ANOVA because the people in the groups are the same. In other words, the group data are not correlated. Because of this correlation, the ANOVA needs to account for the correlation by making an adjustment.

Answer

-  [Download the complete SPSS syntax](#)
-  [Download the complete SPSS output](#)

3.5.1 Our Scenario

Does alcohol affect driving ability? A researcher interested in answering this question tested 10 people on a driving test under 3 conditions (no alcohol, one pint, and two pints) and counted up the number of errors they made.

3.5.2 Step One: Obtain the Data

Open  [Driving errors after alcohol](#). Note that you have three columns: ‘control’, ‘onePint’, and ‘twoPints’.

- Why are the data now split across three columns? The answer lies with SPSS. It prefers to have people in rows. Earlier, in the between-groups example, each person was uniquely assigned to a group, so their score had a particular group code (no therapy, CBT, and sunshine). Now, in our within-groups example, each person has contributed to each group. In other words, they have been tested while sober, after one pint, and after two pints.

3.5.3 Step Two: Run the Descriptive Statistics

5. As before, we need to look at the means to get a sense of our data. Go to *Analyze > Descriptive Statistics > Descriptives...* and make sure all of the variables are transferred to the box marked *Variable(s)*. Click *OK*.

6. What are the mean number of driving errors made by participants in each of the three conditions?

Answer

- *No alcohol:* 1.40
 - *One pint:* 6.50
 - *Two pints:* 6.80
-

SPSS Syntax:

```
DESCRIPTIVES VARIABLES=control onePint twoPints
/STATISTICS=MEAN STDDEV MIN MAX.
```

3.5.4 Step Three: Get a Useful Picture of the Means

For the data on therapy, we created boxplots that used a grouping variable (see Figure 3.1). Now we'll simply create boxplots for each variable (i.e. each condition).

You'll start by going to *Graphs > Legacy Dialogues > Boxplots*.

Make sure that *Summaries of separate variables* are selected, like Figure 3.10.

When you've clicked *OK*, you can specify the variables of interest (all of them, in this case) in the *Boxes Represent* box. See Figure 3.11.

6. Create boxplots for the groups (no alcohol, one pint and two pints) by going to *Graphs > Legacy > Boxplot*. Make sure that it looks like Figure 3.12.

Answer

SPSS syntax:

```
EXAMINE VARIABLES=control onePint twoPints
/COMPARE VARIABLE
/PLOT=BOXPLOT
```

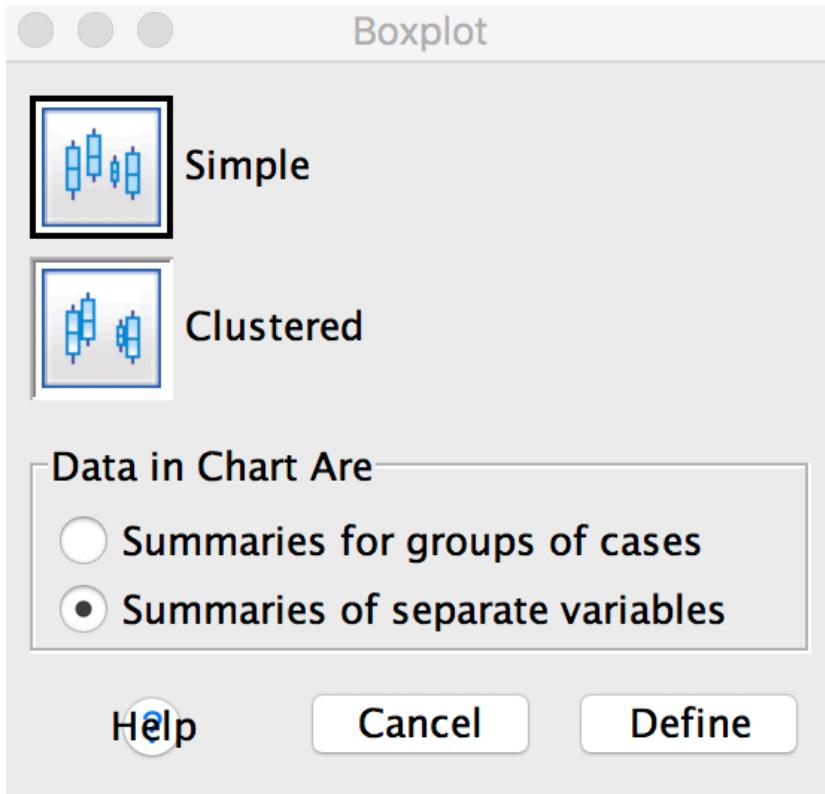


Figure 3.10: SPSS boxplot dialogue

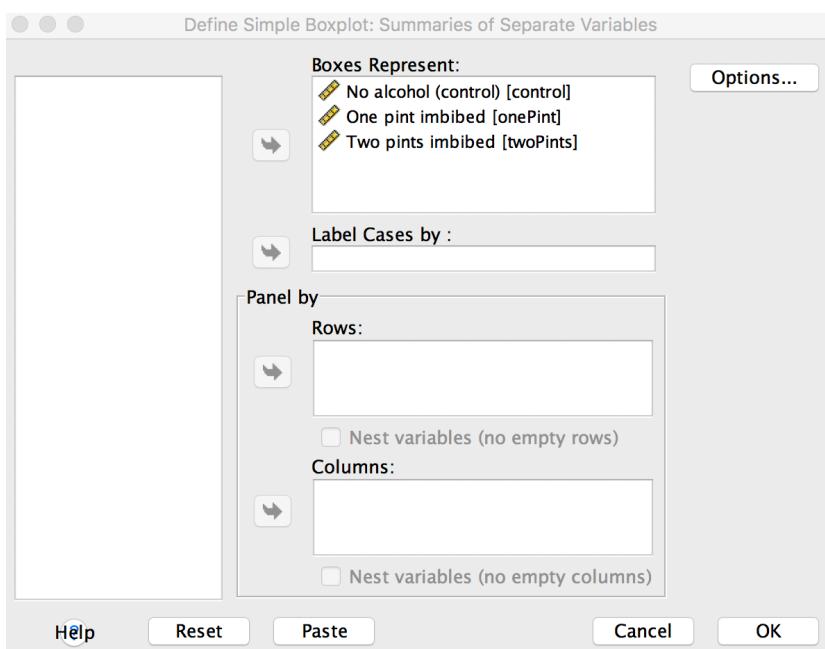


Figure 3.11: SPSS boxplot definition dialogue

```
/STATISTICS=NONE
/NOTOTAL
/MISSING=LISTWISE.
```

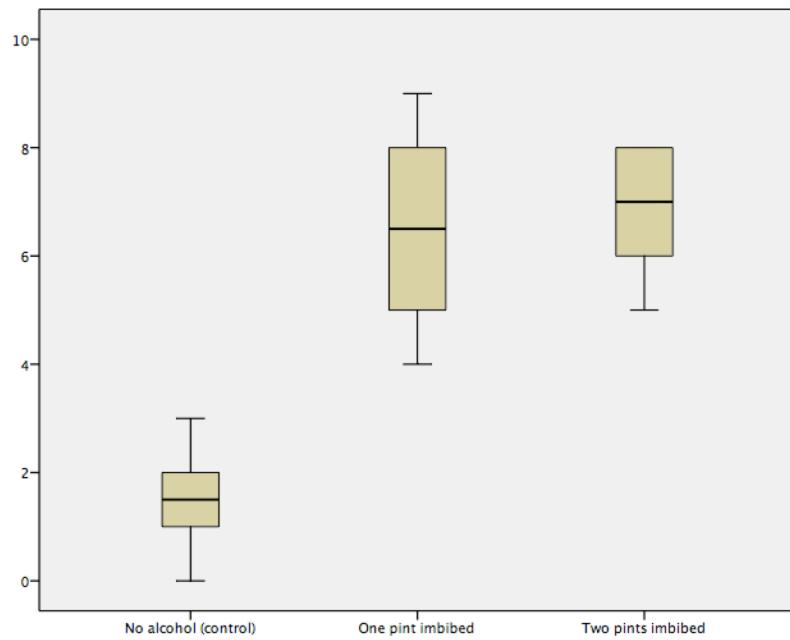


Figure 3.12: Number of errors by alcohol group

3.5.5 Step Four: Set up the ANOVA

As before, we can see that there are differences in activity between the groups, but these might superficial sample differences. We're more interested in whether these differences are true. In other words, do the data better fit a null distribution (of no difference) or an alternative distribution (where the effect is true)? The Analysis of Variance can tell us this.

- Again, we'll not concern ourselves with some of the prerequisites that need to be true before it is appropriate to run a repeated measures ANOVA, with the exception of one thing: the sphericity

assumption.

Before you run the ANOVA, read through the following to get an overall sense of how you do it.

- First, you'll go to *Analyze > General Linear Model > Repeated Measures.*
- A box like Figure 3.13 will appear where you have to tell SPSS what the repeated measures factor is and how many levels (conditions) it contains. In this study there is only one factor, 'drink', and there are three levels (no alcohol, one pint, and two pints). So enter the factor name (such as 'group' in the *Within-Groups Factor Name* box, and enter its levels (3), in the *Number of Levels* box. Then click *add*.

Click the arrow to pass each of the conditions over into the middle box so that it looks like Figure 3.14

Then you'll click *continue* and *OK*.

7. Run the ANOVA according to the instructions above.

Answer

SPSS syntax:

```
GLM control onePint twoPints
/WSFACTOR=group 3 Polynomial
/METHOD=SSTYPE(3)
/CRITERIA=ALPHA(.05)
/WSDESIGN=group.
```

3.5.6 Step Five: Interpret the Output

You should get the same output as I have here. Let's see what it tells us. We can ignore the *Multivariate Tests* table (Figure 3.15)

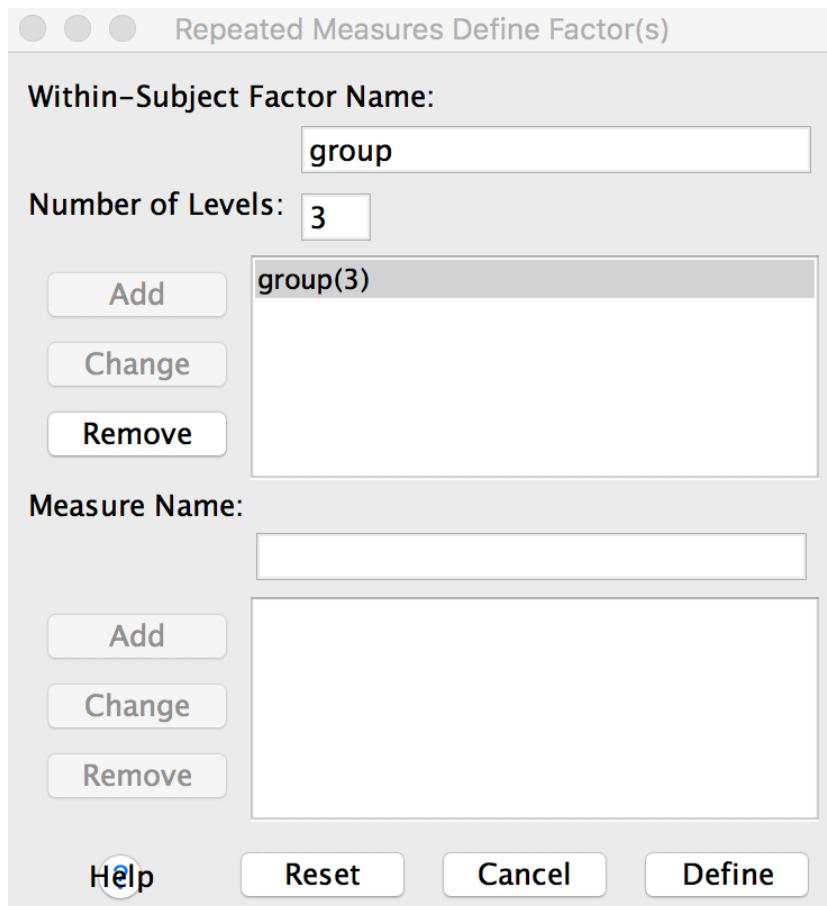


Figure 3.13: One-way within-groups ANOVA factor definitions

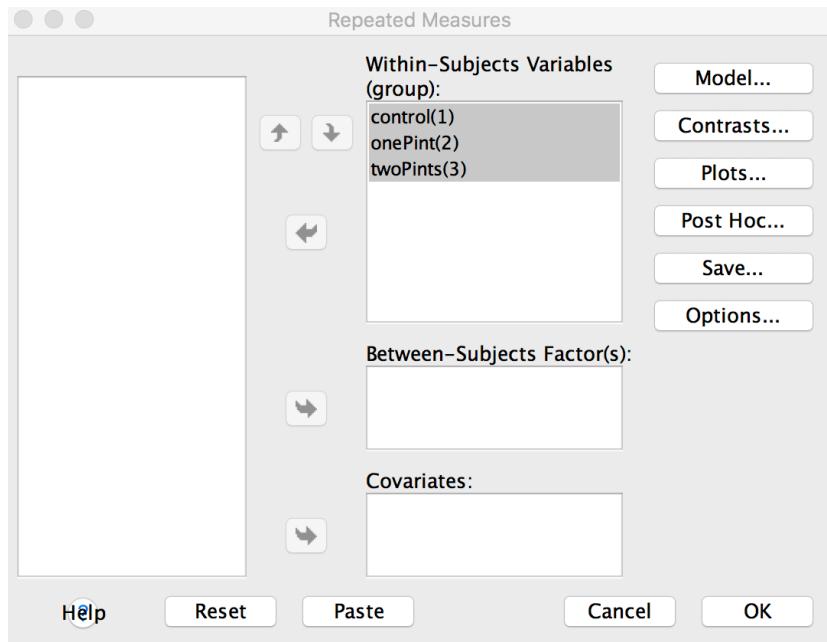


Figure 3.14: One-way within-groups ANOVA variables entered

Multivariate Tests ^a					
Effect	Value	F	Hypothesis df	Error df	Sig.
group Pillai's Trace	.980	196.187 ^b	2.000	8.000	.000
Wilks' Lambda	.020	196.187 ^b	2.000	8.000	.000
Hotelling's Trace	49.047	196.187 ^b	2.000	8.000	.000
Roy's Largest Root	49.047	196.187 ^b	2.000	8.000	.000

Figure 3.15: Output—Multivariate Tests

a. Design: Intercept
Within Subjects Design: group

b. Exact statistic

The table in Figure 3.16, however, is much more important. *Mauchly's test of sphericity* is the repeated measures equivalent of Levene's test for homogeneity of variance. That means we will check to see if we have violated the assumption of equal variances of the differences between levels. See Field (2009).

8. Check the significance. If it *is* significant, then we have violated the assumption and may not to rely on the result of the ANOVA without adjustment.

Mauchly's Test of Sphericity ^a							
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
group	.497	5.590	2	.061	.665	.737	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: group

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Figure 3.16: Output—Mauchley's Test

7. Have we met the sphericity assumption? Explain with regards to the statistic in the *Mauchly's Test of Sphericity* table.

Answer

Yes, we have met the assumption between Mauchly's test was not significant.

Depending on the result of Mauchly's test, we will decide which sub-row of data to look at on the next table, the *Tests of within-subjects effects* (Figure 3.17).

- If Mauchly's test was non-significant, we can assume sphericity, and we look at sub-row labelled 'Sphericity assumed'.
- If the test was significant, we cannot assume sphericity, and must look at one of the other rows. Each row is a corrected result based

on a slightly different formula. There are lots of complicated discussions about which one to use, but if you do not want to go into all the complications, Girden (1992) recommends that when the estimate of sphericity is greater than .75 the Huynh-Feldt correction should be used; otherwise, the Greenhouse-Geisser correction should be used. For more information, see Field (2009).

Tests of Within-Subjects Effects						
Measure: MEASURE_1						
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
group	Sphericity Assumed	184.200	2	92.100	109.546	.000
	Greenhouse-Geisser	184.200	1.331	138.408	109.546	.000
	Huynh-Feldt	184.200	1.475	124.921	109.546	.000
	Lower-bound	184.200	1.000	184.200	109.546	.000
Error(group)	Sphericity Assumed	15.133	18	.841		
	Greenhouse-Geisser	15.133	11.978	1.263		
	Huynh-Feldt	15.133	13.271	1.140		
	Lower-bound	15.133	9.000	1.681		

Figure 3.17: Output—Tests of Within-Groups Effects

8. Is there a significant main effect of alcohol on driving ability?

Write a sentence in this box to describe the finding and write out the result of the ANOVA in APA style.

Answer

The main effect of alcohol was significant, $F(2,18) = 109.55$, $MSE = .84$, $p < .001$. This means that the amount of errors participants made depended on how much alcohol they had drunk.

3.5.7 Follow-up Comparisons

Since the effect of alcohol on driving ability was significant, we need to do some follow-up analyses.

9. Do post-hoc tests on the data by carrying out three paired-samples T-tests comparing the three conditions (*Analyze > Compare Means > Paired Samples T-test*)
- T-test 1: no alcohol v one pint
 - T-test 2: no alcohol v two pints
 - T-test 3: one pint v two pints

Answer

SPSS syntax:

```
T-TEST PAIRS=control control onePint
  WITH onePint twoPints twoPints (PAIRED)
  /CRITERIA=CI(.9500)
  /MISSING=ANALYSIS.
```

9. What are the results of these T-tests? Report each in APA format and provide a verbal description of what the results mean.

Answer

- T-test 1: no alcohol v one pint: $t(9) = 9.70, p < .001$, 95% CI [-6.29, -3.91]
- T-test 2: no alcohol v two pints: $t(9) = 20.25, p < .001$, 95% CI [-6.00, -4.80]
- T-test 3: one pint v two pints: $t(9) = .76, p = .47$, 95% CI [-1.20, 0.60]

These results show that drinking one or two pints affected driving ability compared to the no alcohol condition. However,

there was no difference in performance between the one versus two pint conditions.

As you know, when carrying out a number of t-tests, we need to control our family-wise error rate.

10. Applying Bonferroni's correction, what would the new significance cut-off be? Does it change the results of the T-test?

Answer

$$.05 / 3 = .02$$

- The new cut-off doesn't change the interpretation of our results. Drinking one or two pints affected driving ability compared to no alcohol but the amount of alcohol consumed did not impact performance.

4

Two-Way Between Groups ANOVA

4.1 Overview

Data files required  [Example 1 SPSS data](#)

Booklet Version 1.0

4.2 Objectives

This workshop extends your knowledge of Analysis of Variance (ANOVA) to factorial designs. Remember that ANOVA uses the ratio between explained and unexplained variance to tell you whether the means are difference between levels (groups) of a factor (the thing you are manipulating, or your independent variable). Having looked at the simple case of one-way (i.e. one factor) ANOVA last week, in both independent and repeated forms, this week you'll look at the two-way ANOVA in its independent form.

- In the first example, you'll examine the effects of traffic noise and age on learning;
- In the second example, you'll examine the effects of music and alcohol on driving behaviour;
 - You'll also learn how to set up your SPSS data file.

Remember, this booklet will focus on between-groups analysis.

That is, for a given factor, there will be different people in each level. This means that the observations will be independent.

4.3 Workshop

Analysis of variance or ‘ANOVA’ is a statistical technique that allows us to examine the differences between mean scores when there are more than two conditions. Basically, it compares the variance *between* the conditions with the variance *within* the conditions. The bigger the differences between the conditions and the smaller the individual differences within a condition, the more likely we are to get a significant result. In this booklet we will carry out the analysis required for a design with two independent variables (i.e. two factors). Both of them are between-subjects (i.e. there are different people in each group or condition).

The ANOVA will tell us three things:

1. Is there a main effect of factor one on the dependent variable (DV)?
2. Is there a main effect of factor two on the DV?
3. Is there an interaction effect of the two factors on the DV?

Answer

All data and syntax for this booklet:

- [Example 1 SPSS data](#)
- [Example 1 SPSS syntax](#)
- [Example 2 SPSS data](#)
- [Example 2 SPSS syntax](#)

4.4 Example 1: Traffic noise and Learning

The conditions under which words are remembered are thought to affect the amount of subsequent recall. In particular, road traffic noise at the time of learning has been shown to affect subsequent recall and it seems to be most damaging for children. To look at this phenomenon, participants in the current experiment were randomly assigned to one of two noise conditions: loud (with road traffic) or quiet (no audible road traffic) and asked to read a 1000 word story. There were 20 children (aged 11 - 14) and 20 adults (aged 30 - 35). One week later they came back into the lab and were asked a series of 20 questions about the story. Recall (the dependent variable) was measured as the number of correct answers.

4.4.1 Step One: Load the data into SPSS

The full data set is available for download here: [Example 1 SPSS data](#)

1. Open the data file and familiarise yourself with its contents. Use the paragraph above you help you.

1. What is the design of this study?

Answer

This is a 2 x 2 between-groups factorial design. Factor 1 is Age (child; adult) and factor 2 is Noise Condition (loud; quiet).

4.4.2 Step Two: Run descriptive statistics

Before we carry out the ANOVA, we should get an idea of what is going on with the data.

2. Calculate the mean scores for the four conditions by going to *Analyze > Compare means > Means*. As shown in Figure 4.1, put the DV (recall) into the *Dependent List* and ‘ageGroup’ into the *Independent List*. Now click on *Next* and ‘ageGroup’ should disappear. You are now specifying the second layer of your comparison (note where it says *Layer 2 of 2*). Click on ‘noise’ and put this into the *Independent List* and click *OK*.
- Your means per condition should be the same as mine in Figure 4.2.

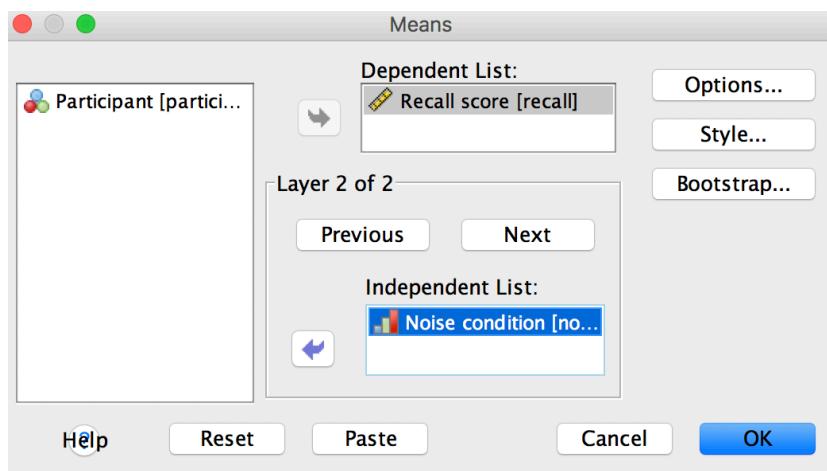


Figure 4.1: SPSS compare means dialogue

2. What do the descriptives tell you?

Answer

In our sample:

- Children had lower recall than adults.
- Quiet noise conditions resulted in better recall than louder conditions.

Report				
Recall score				
Age group	Noise condition	Mean	N	Std. Deviation
Child	Loud	6.1000	10	1.37032
	Quiet	8.8000	10	2.85968
	Total	7.4500	20	2.58488
Adult	Loud	11.4000	10	1.71270
	Quiet	13.6000	10	3.37310
	Total	12.5000	20	2.83772
Total	Loud	8.7500	20	3.10983
	Quiet	11.2000	20	3.91488
	Total	9.9750	40	3.70369

Figure 4.2: Output—compare means output

- The difference between quiet and load conditions in children was greater than for adults, so noise impaired children's recall more than adults'.
-

SPSS Syntax:

```
MEANS TABLES=recall BY ageGroup BY noise
/CELLS=MEAN COUNT STDDEV.
```

4.4.3 Step Three: Set up the ANOVA

By looking at the descriptive statistics, we have some inkling of group differences. Now let's see whether those differences are significant or not.

Read through the instructions below before doing them yourself.

- Remember that all the SPSS statistical analyses are under the *Analyze* menu, so go there, select *General Linear Model*, then *Univariate*.

2. As I've done in Figure 4.3, put the dependent variable into the *Dependent Variable* box and the independent variables into the *Fixed Factor(s)* box.

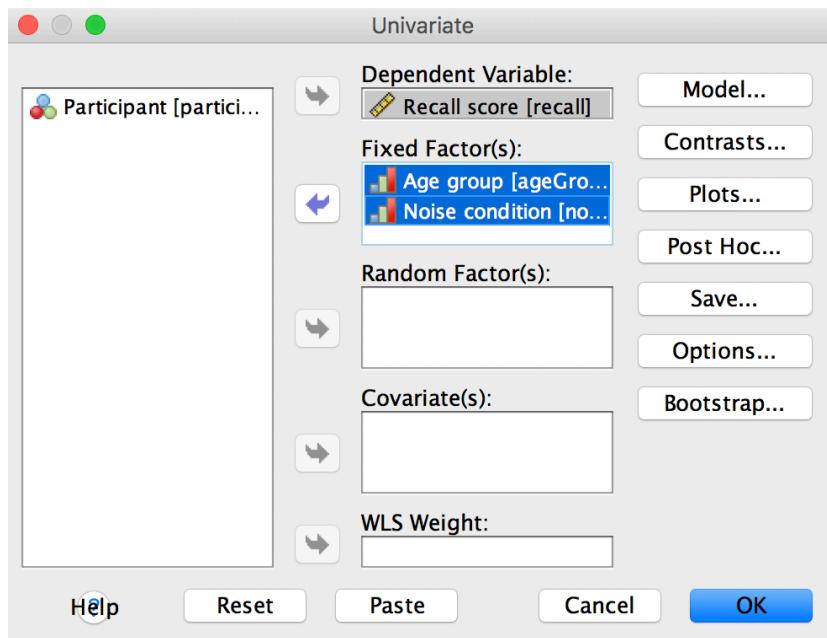


Figure 4.3: SPSS Univariate ANOVA dialogue

3. Click on *Options* and make sure you have selected the factors and interactions that are included in Figure 4.4.
4. Now click *Continue*. We know that graphs aid the understanding of patterns in our data, so click *Plots* and ask SPSS to produce a graph with 'ageGroup' on the horizontal axis and 'noise' as separate lines, just as you see in Figure 4.5.
5. Click on *Add* and *Continue*. Then click *OK*.

By the end of the above process, you will have set up and run your first two-way between-subjects ANOVA.

3. Go ahead and run the ANOVA described above.

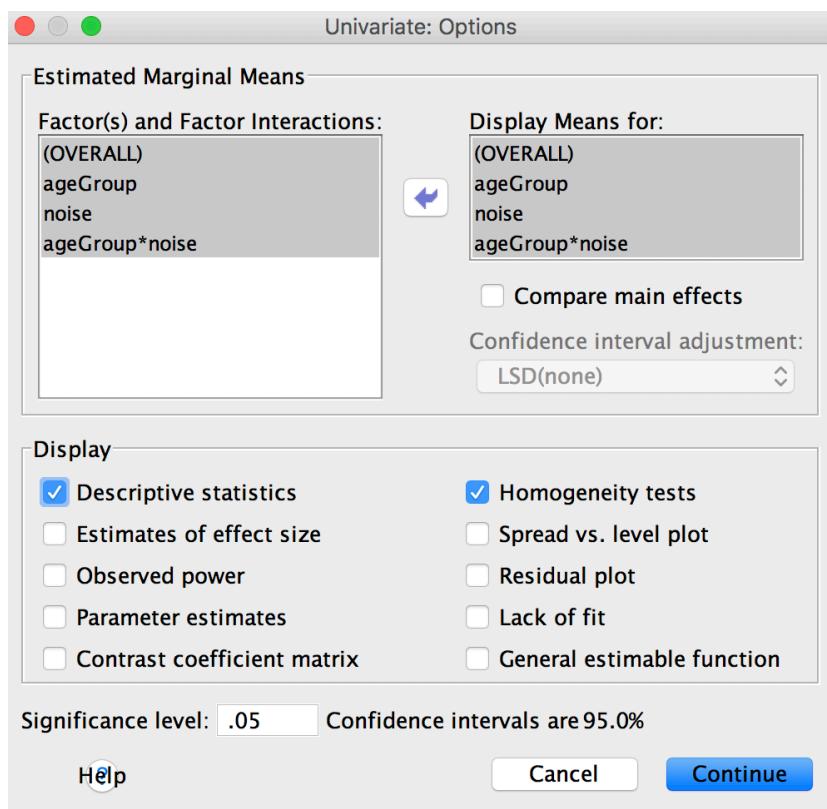


Figure 4.4: SPSS Univariate ANOVA dialogue, Options

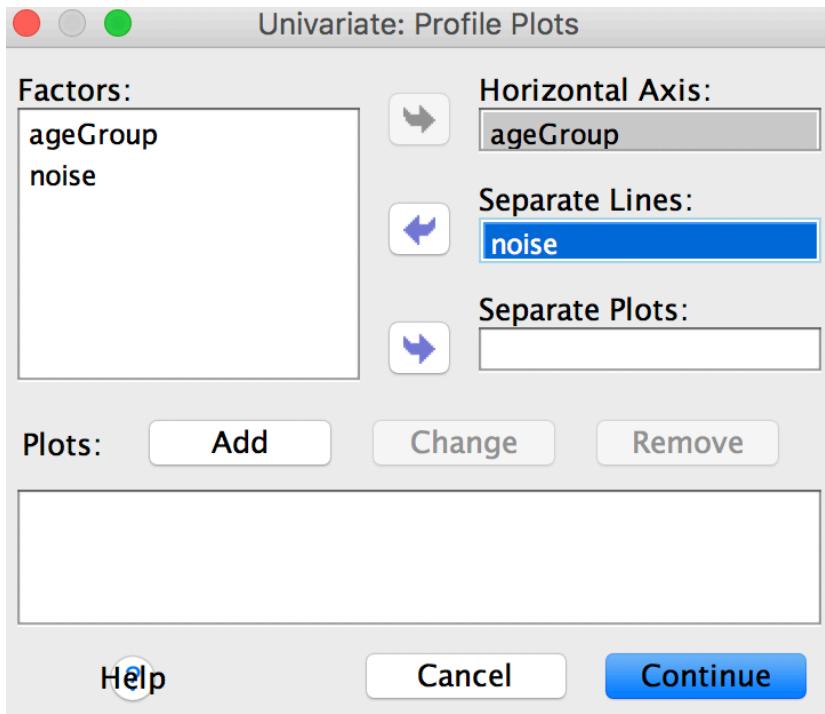


Figure 4.5: SPSS Univariate ANOVA dialogue, Profile Plots

Answer

SPSS syntax for the ANOVA:

```
* Run the first ANOVA on recall with age and noise
* as between participants factors.

UNIANOVA recall BY ageGroup noise
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/PLOT=PROFILE(ageGroup*noise)
/EMMEANS=TABLES(OVERALL)
/EMMEANS=TABLES(ageGroup)
/EMMEANS=TABLES(noise)
/EMMEANS=TABLES(ageGroup*noise)
/PRINT=HOMOGENEITY DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/DESIGN=ageGroup noise ageGroup*noise.
```

4.4.4 Step Four: Interpret the output

We are now ready to answer the important question. Are those differences significant or not?

The first box tells us how many people we have under each condition—nothing new here. The second box gives us the descriptive statistics that we should already have looked at in [8 Step Two](#). The first new bit of information comes with Levene's Test of the assumption of homogeneity of variance. Remember that we want the variances to be equal (not different), so this needs to be non-significant.

3. Look at the homogeneity of variance statistic in Figure 4.6. Has the homogeneity of variance assumption been met? Include the statistics to back that up.

Answer

Yes, Levene's test is nonsignificant, $F(3,26) = 1.93$, $p = .142$.

Levene's Test of Equality of Error Variances^a

Dependent Variable: Recall score

F	df1	df2	Sig.
1.934	3	36	.142

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

- a. Design: Intercept + ageGroup + noise
+ ageGroup * noise

Figure 4.6: Output—Homogeneity of Variance

Take a look at the ANOVA table (*Tests of Between-Subjects Effects*), which should resemble Figure 4.7.

Tests of Between-Subjects Effects					
Dependent Variable: Recall score					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	315.675 ^a	3	105.225	17.274	.000
Intercept	3980.025	1	3980.025	653.356	.000
ageGroup	255.025	1	255.025	41.865	.000
noise	60.025	1	60.025	9.854	.003
ageGroup * noise	.625	1	.625	.103	.751
Error	219.300	36	6.092		
Total	4515.000	40			
Corrected Total	534.975	39			

a. R Squared = .590 (Adjusted R Squared = .556)

Figure 4.7: Output—ANOVA

3. Use Figure 4.7 to answer the following questions, providing the statistics to back up your answers.
- 3.1 Is there a main effect of factor one on the DV?
- 3.2 Is there a main effect of factor two on the DV?
- 3.3 Is there an interaction effect of the two factors on the DV?

Answer

- 3.1. Yes, there is a main effect of noise on recall: $F(1,36) = 9.85$, $MSE = 6.09$, $p = .003$.
- 3.2. Yes, there is a main effect of age on recall: $F(1,36) = 41.87$, $MSE = 6.09$, $p < 0.01$.
- 3.3. There was no interaction: $F(1,36) = .10$, $MSE = 6.09$, $p < .75$.

Providing charts in APA style is an important skill. SPSS charts need some work to make them conform. Figure 4.8 is a spruced up version of the plot that SPSS original produced. Editing charts is achieved by double-clicking on them in the output window. This exposes a toolbox with which can change visual properties.

4. See if you can get your plot to look like Figure 4.8.

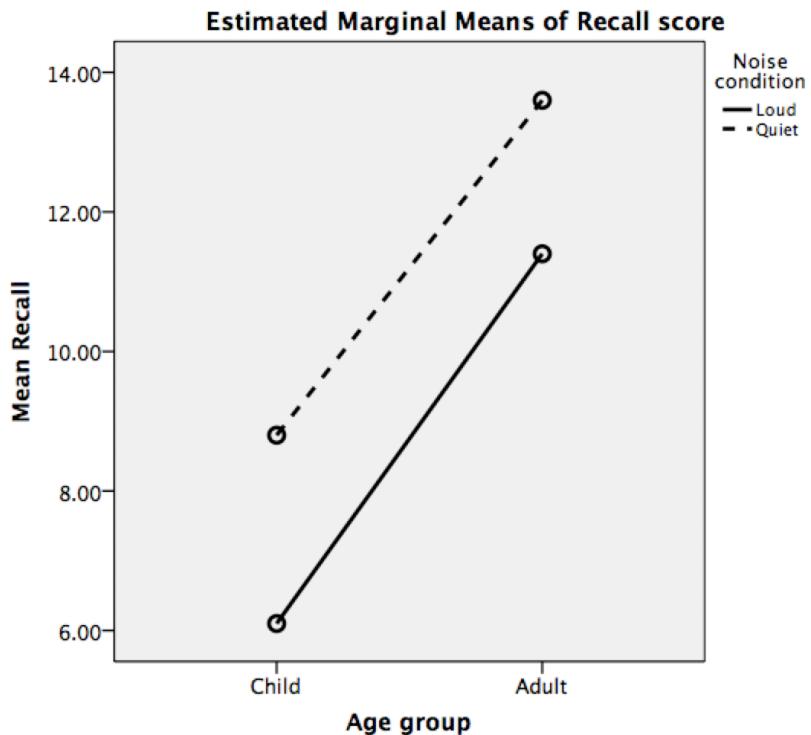


Figure 4.8: Output—Profile Plots

4. What do the significant effects reported in Figure 4.7, together with the plot in Figure 4.8, tell us?

Answer

The effect of age was significant, and adult participants had better recall than younger ones. The effect of noise was significant, and quiet conditions led to better recall than loud conditions. The interaction was not significant, so the effect of noise was the same for both adults and children (the differences observed in the descriptive statistics failed to reach significance).

4.5 Example 2: Alcohol and Music on Driving Ability

In the lecture, I described an experiment investigating the effects of alcohol and music on driving behaviour. Below you will find the raw data. Participants were allocated to one of four groups, depending on whether or not they had some alcohol before getting in the car and whether or not they had music whilst driving. Alcohol and Music are the two factors, both between-subjects. The DV is driving ability, measured as the number of errors committed whilst driving.

Table 4.2: The effects of alcohol and music on driving ability.

<i>participant</i>	<i>music</i>	<i>alcohol</i>	<i>errors</i>
1	1	1	8
2	1	1	10
3	1	1	7
4	1	2	0
5	1	2	2
6	1	2	1
7	2	1	5
8	2	1	4
9	2	1	3
10	2	2	2
11	2	2	0
12	2	2	1

5. Type these data into SPSS, which will give you more practice with setting up data files correctly. Make sure:
- You enter the above data correctly in the *Data* view;

- In the *Variable* view, give each variable an appropriate label and name (I've called my variables *participant*, *music*, *alcohol*, *errors*—use the same for now, so that you don't have any problems running the syntax below);
- Also in the *Variable* view, use the *Values* column to assign values to your grouping numbers (without these, your output will be more difficult to interpret);
- Finally, in the *Variable* view, set the *Measure* view correctly for each variable (without this, SPSS may not perform the correct analyses).

Answer

- My values for *music* are 1 = *Music*, 2 = *No music*; values for *alcohol* are 1 = *Alcohol*, 2 = *No alcohol*
- *participant* is *nominal*, *music* is *ordinal*, *alcohol* is *ordinal*, *errors* is *scale*

5. Report the design of the study

Answer

This is a 2 x 2 between-groups factorial design. Factor one is music (no music; music) and factor two is alcohol (no alcohol; alcohol).

6. As usual, get an overview of your data by looking at the means. Go to *Analyse > Compare Means* and compute them for each group. You should get something like Figure 4.9.

Report				
Errors				
Music condition	Alcohol condition	Mean	N	Std. Deviation
Music	Alcohol	8.3333	3	1.52753
	No alcohol	1.0000	3	1.00000
	Total	4.6667	6	4.17931
No music	Alcohol	4.0000	3	1.00000
	No alcohol	1.0000	3	1.00000
	Total	2.5000	6	1.87083
Total	Alcohol	6.1667	6	2.63944
	No alcohol	1.0000	6	.89443
	Total	3.5833	12	3.28795

Figure 4.9: Output—Descriptives for error by alcohol group and music group

6. Describe what you see in the means.

Answer

Participants produced more errors when they heard music (mean 4.67) versus not hearing music (2.50). With alcohol, their errors increased markedly (from 1.00 to 6.2).

7. Now set up the ANOVA as you did in Example 1's [Step Three](#).

Remember:

- to ask that SPSS produce a table of descriptives as well marginal means for all our effects;
- to test the homogeneity of variance;
- to produce a plot depicting the results (ask for alcohol as separate lines and music on the horizontal axis, so that you get the same plot as mine).

Answer

SPSS Syntax:

```
UNIANOVA errors BY music alcohol
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/PLOT=PROFILE(music*alcohol)
/EMMEANS=TABLES(OVERALL)
/EMMEANS=TABLES(music)
/EMMEANS=TABLES(alcohol)
/EMMEANS=TABLES(music*alcohol)
/PRINT=HOMOGENEITY DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/DESIGN=music alcohol music*alcohol.
```

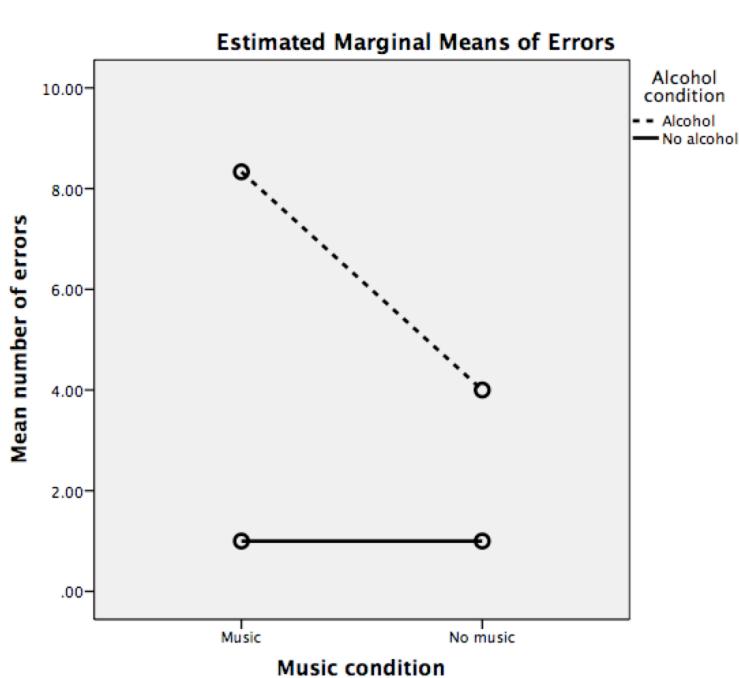


Figure 4.10: Output—Estimated Marginal Means

8. Before you check anything else, have a look at your plot to see the familiar data pattern. Edit it to make it look like Figure 4.10.

7. What does the graph tell us about our results? On the basis of the graph, would you expect a main effect of music? A main effect of alcohol? An interaction? Explain your answer to each of those questions.

Answer

- Yes, you might expect a main effect of music. When collapsed across alcohol condition, the means will be different.
- You might expect a main effect of alcohol because, again, when collapsing across the other factor (music), the means will be different.
- Since the lines are not parallel, we might expect an interaction.

Now let's have a look at the results and find out what the statistics are telling us.

7. Have we met the homogeneity of variance assumption? Report the statistics from Figure 4.11 to back up your answer.

Answer

Yes, Levene's test is nonsignificant: $F(3,8) = 0.40, p = .76$. A non significant result means that the variances are likely to be equal.

Levene's Test of Equality of Error Variances^a

Dependent Variable: Errors

F	df1	df2	Sig.
.400	3	8	.757

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

- a. Design: Intercept + music + alcohol + music * alcohol

Figure 4.11: Output—Levene's Test of Equality of Variances

8. Report all main and interaction effects. In plain English, what does each effect mean in terms of group differences? Which of these effects need post hoc tests and why?

Answer

We have:

- A significant main effect of music on driving ability, $F(1, 8) = 10.56$, $MSE = 1.33$, $p = .012$. Driving ability differed depending on music group (with music, performance was worse).
- A significant main effect of alcohol on driving ability, $F(1, 8) = 60.06$, $MSE = 1.33$, $p < .001$. Driving ability differed depending on alcohol group (with alcohol, performance was worse).
- A significant interaction of music and alcohol on driving ability, $F(1, 8) = 10.56$, $MSE = 1.33$, $p = .012$

We should carry out simple main effects analyses for the interaction. No post hoc testing is required to follow up the significant main effects because both factors have only two con-

ditions; because of this, there's nothing more we can learn from post hoc pairwise comparisons.

The (spoiler alert) significant interaction effect needs to be followed up with simple main effects analyses. Remember that in order to carry out SME analyses, we need to access the SPSS Syntax Editor.

9. Go to *Analyse > General Linear Model > Univariate*.

- If you haven't exited SPSS, you'll find that the settings for your last ANOVA will still be present in the dialogue box. In this case, simply click *paste* and the syntax for the ANOVA will appear in a new syntax window. (The window might be somewhere in the background.)
- If you have exited SPSS, you'll have to re-enter the settings for your last ANOVA before clicking *paste*.

10. Ensure your syntax looks like Figure 4.12.

```
UNIANOVA errors BY music alcohol
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/PLOT=PROFILE(music*alcohol)
/EMMEANS=TABLES(OVERALL)
/EMMEANS=TABLES(music)
/EMMEANS=TABLES(alcohol)
/EMMEANS=TABLES(music*alcohol)
/PRINT=HOMOGENEITY DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/DESIGN=music alcohol music*alcohol.
```

Figure 4.12: Syntax—Univariate ANOVA, without SME analyses

- Syntax is a programming language, and these need attention to detail. Even a misplaced comma will stop a program running correctly.
- In SPSS Syntax, variables are black and all other text colour-coded. If you find that you've typed a command but SPSS has not coloured-coded it, then this suggests that SPSS has not recognised the command; have you spelled it correctly?

In Figure 4.12, look at the line reading:

```
/EMMEANS=TABLES(music*alcohol)
```

We're going to extend this line so that it asks SPSS to calculate the simple main effects *within music* of *alcohol*¹, while making a Bonferroni adjustment to control for familywise error:

```
/EMMEANS=TABLES(music*alcohol) compare (music) ADJ(BONFERRONI)
```

Additionally, beneath it, we will include the following, which produces the simple main effects *within alcohol* of *music*² (while making the same familywise error adjustment):

```
/EMMEANS=TABLES(music*alcohol) compare (alcohol) ADJ(BONFERRONI)
```

Your syntax should now look like Figure 4.13.

```
UNIANOVA errors BY music alcohol
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/PLOT=PROFILE(music*alcohol)
/EMMEANS=TABLES(OVERALL)
/EMMEANS=TABLES(music)
/EMMEANS=TABLES(alcohol)
/EMMEANS=TABLES(music*alcohol) compare (music) ADJ(BONFERRONI)
/EMMEANS=TABLES(music*alcohol) compare (alcohol) ADJ(BONFERRONI)
/PRINT=HOMOGENEITY DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/DESIGN=music alcohol music*alcohol.
```

¹ These SMEs will tell us (i) whether listening to music affected driving ability in the group that had alcohol and (ii) whether listening to music affected driving in the group that had no alcohol.

² These SMEs will tell us (i) whether drinking alcohol affected driving ability in the group that had music in the car and (ii) whether alcohol affected driving in the group that did have music.

Figure 4.13: Syntax—Univariate ANOVA, including SME analyses

10. Run the syntax.

- Make sure the syntax window has focus and the syntax you wish to run is currently under cursor selection.
- Then click on the menu bar *Run* followed by *Selection*.
- Check that your two new *Univariate Tests* tables look like Figure 4.14 and Figure 4.15. If they don't, something might be wrong with your syntax, so go back and see if you can spot your error.

Univariate Tests					
Dependent Variable: Errors					
	Sum of Squares	df	Mean Square	F	Sig.
Alcohol condition	28.167	1	28.167	21.125	.002
	10.667	8	1.333		
No alcohol	.000	1	.000	.000	1.000
	10.667	8	1.333		

Each F tests the simple effects of Music condition within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

Figure 4.14: Output—Univariate Tests, Simple Main Effects of Music within Alcohol

Univariate Tests					
Dependent Variable: Errors					
	Sum of Squares	df	Mean Square	F	Sig.
Music condition	80.667	1	80.667	60.500	.000
	10.667	8	1.333		
No music	13.500	1	13.500	10.125	.013
	10.667	8	1.333		

Each F tests the simple effects of Alcohol condition within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

Figure 4.15: Output—Univariate Tests, Simple Main Effects of Alcohol within Music

9. Were the simple main effects of music within alcohol significant (Figure 4.14)? Report the statistics to back up your answer.

Answer

The SME of music was significant within the alcohol condition, $F(1,8) = 21.13$, $MSE = 1.33$, $p = .002$, but it wasn't significant within the no alcohol condition, $F(1,8) = .001$, $MSE = 1.33$, $p > .99$.

10. Were the simple main effects of alcohol within music significant (Figure 4.15)? Report the statistics to back up your answer.

Answer

The SME of Alcohol was significant within the no music condition, $F(1,8) = 60.50$, $MSE = 1.33$, $p < .001$, and it was also significant within the music condition, $F(1,8) = 10.13$, $MSE = 1.33$, $p = .013$.

11. Based on the results of the ANOVA and post hoc analyses, what can we conclude?

Answer

- Both music and alcohol had an effect on driving ability independently (which we have determined by observing the main effect of each factor). The presence of music made their performance worse, as did the presence of alcohol.

There was a significant interaction. Follow up analyses indicated:

- When participants had alcohol, music made their performance worse. When participants had no alcohol, music didn't have an effect on their performance. (This is concluded from the SME of music within alcohol.)

- Participants were made worse by alcohol no matter whether they heard music or not. (This is concluded from the SME of alcohol within music.)

5

Two-Way Repeated and Mixed ANOVA

5.1 Overview

Data files required  **Therapy**

Booklet Version 1.0

5.2 Objectives

This workshop extends your knowledge of Analysis of Variance (ANOVA) to repeated and mixed factorial designs. In a *repeated* design, all factors are repeated. In a *mixed* design, there is at least one *repeated* factor and one *independent* factor.

5.3 Workshop

Analysis of variance or ‘ANOVA’ is a statistical technique that allows us to examine the differences between mean scores when there are more than two conditions. In this booklet, we will carry out the analysis required for a design with two independent variables (i.e. two factors). In the  **first case**, we will look at an analysis where both factors are *repeated*. In the  **second**, we will look at a *mixed* analysis where one factor is *repeated* and one is *independent*.

In both cases, the ANOVA will tell us three things:

1. Is there a main effect of factor one on the dependent variable (DV)?
2. Is there a main effect of factor two on the DV?
3. Is there an interaction effect of the two factors on the DV?

There's a lot to get through in this week's booklet, but it's important material. Be prepared to continue with the exercises outside of the seminar.

Answer

All data and syntax for this booklet:

- [Repeated ANOVA SPSS data](#)
- [Repeated ANOVA SPSS syntax](#)
- [Mixed ANOVA SPSS data](#)
- [Mixed ANOVA SPSS syntax](#)

5.4 Repeated ANOVA

We'll start by revisiting the example I showed you in the lecture.

Students from the University of Statisticsburg were asked to rate the lectures that they received in relation to how much they enjoyed them. Each student surveyed was taking a course in statistics and a course in criminal psychology and were given the survey at the beginning of the course and again ten weeks later at the end. The researcher wanted to know three things:

1. Did people rate one course more positively than the other?
2. Did people rate the course differently from the start to the end of the course?

3. Was the difference in ratings from the start to the end of the course the same for both courses?

Ratings are provided per student in the table below. Responses could range from 0 (hate it) to 10 (love it).

Table 5.2: Before and after ratings for the statistics and criminal psychology courses.

Statistics Start	Criminal Start	Statistics End	Criminal End
0	5	4	5
2	7	5	7
1	8	3	6
2	4	2	5
1	5	5	5
0	6	7	6
1	7	5	6
1	5	4	5
1	8	1	7
2	4	5	4

5.4.1 Step One: Load the data into SPSS

- Set up a new SPSS data file and enter the data in the table above.
- Call the variables *stats_start*, *criminal_start*, *stats_end* and *criminal_end*
 - Give them appropriate labels
 - Make sure SPSS knows the data type associated with each (which

should be *scale*)

5.4.2 Step Two: Run descriptive statistics

Before we carry out the ANOVA it is a good idea to get an idea of what is going on with the data.

2. Calculate the mean scores for the four conditions by going to *Analyze > Descriptive statistics > Descriptives*. Put all four conditions into the right hand box and click *OK*.

1. Describe what the means tell you.

Answer

These means represent the average ratings for Statistics and for Criminal Psychology at the beginning and end of the courses. They show that Statistics had much lower ratings than Criminal Psychology at the beginning of the term but the difference was much smaller by the end of the term. Whereas Statistics increased, Criminal Psychology decreased slightly.

SPSS Syntax:

```
DESCRIPTIVES VARIABLES=stats_start criminal_start  
stats_end criminal_end  
/STATISTICS=MEAN STDDEV MIN MAX.
```

5.4.3 Step Three: Set up the ANOVA

By looking at the descriptive statistics you have some idea of what students thought about the courses and whether their opinions became more or less positive by the end of the course. Now let's see whether those differences are significant or not.

Read through the instructions below before doing them yourself.

1. Remember that all the SPSS statistical analyses are under the *Analyze* menu, so go there, select *General Linear Model*, then *Repeated Measures*. A window will appear where you can define the repeated measures factors (independent variables). As we have two repeated measures factors, we will define two.
2. First type 'time' into the box next to Within-Subject Factor Name. This variable has two levels (the start and the end of term) so type 2 into the box next to *Number of Levels*. Then click on *Add* and it will show up in the main box. Next, define the second repeated measures variable: 'course'. This also has two levels (statistics and criminal psychology). Once this has been defined and added you should have something that looks like Figure 5.1.
3. Now click on *Define* and a new box will appear, which should look like Figure 5.2. This is where we tell the SPSS which columns of data to look at.

- Pay close attention to the numbers in Figure 5.2's *Within-Subjects Variables* (*time, course*) box. Each line in this box refers to the column in the *Data View* where SPSS can find the data for a particular factor combination. (You can think of these combinations as the cells in the overall table when you put ANOVA into tabular form, with marginal means.) For instance, _?_(1,1) is SPSS's way of requesting the data that matches level 1 of the first factor and

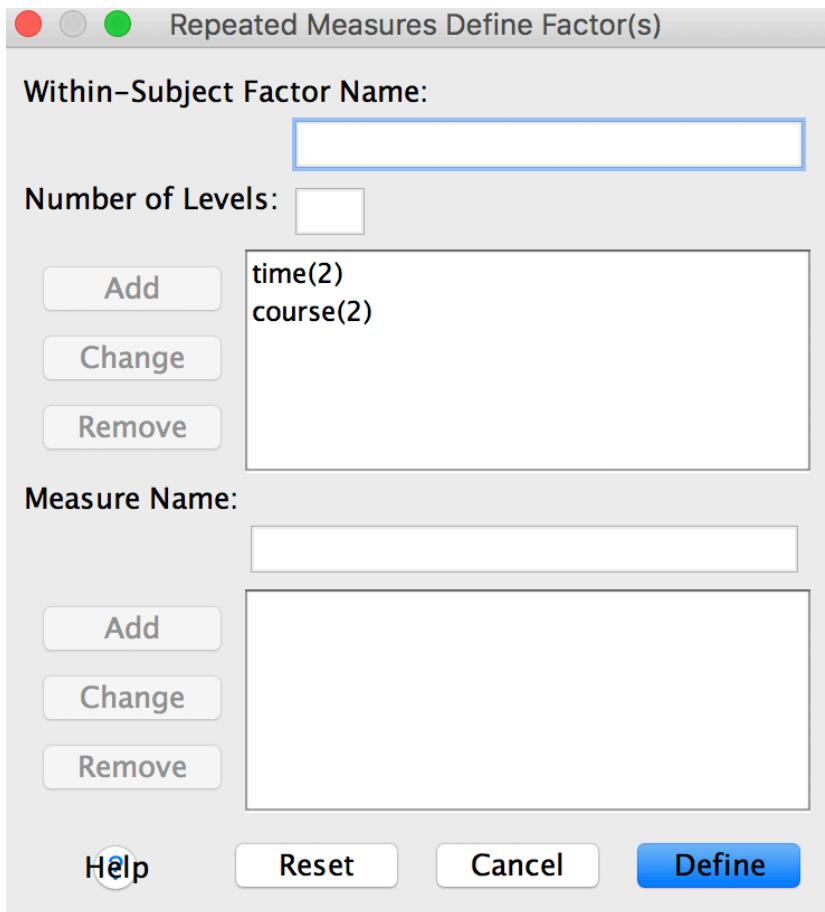


Figure 5.1: SPSS Repeated Measures Define Factors Dialogue

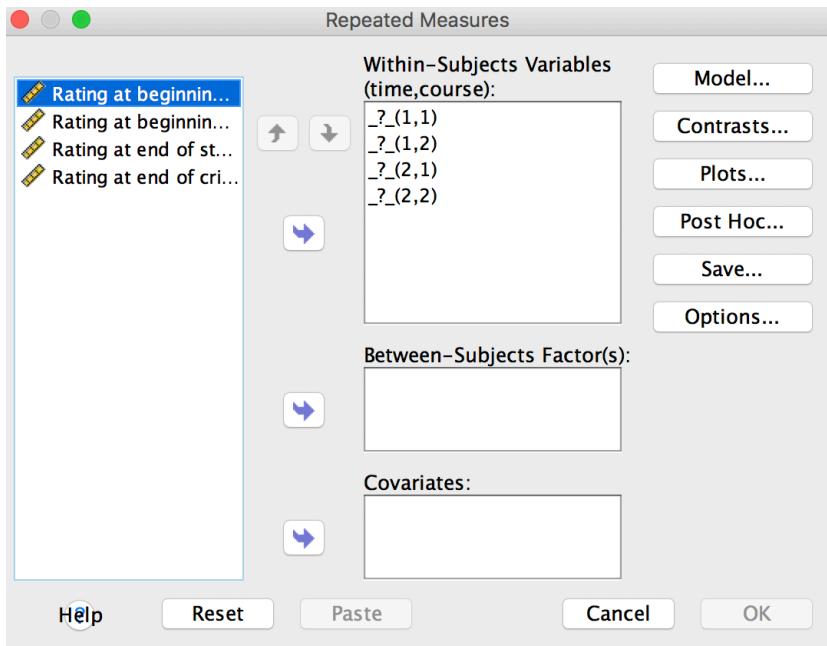


Figure 5.2: SPSS Repeated Measures Main ANOVA Dialogue

level 1 of the second factor. Likewise, $?(1,2)$ requests level 1 of the first factor and level 2 of the second. Bearing in mind that factor one is Time (start vs. end) and factor two is Course (statistics vs. criminal psychology):

- stats_start is (1,1)
- criminal_start is (1,2)
- stats_end (2,1)
- criminal_end (2,2)

4. Drag the variables across to the *Within-Subjects Variables (time, course)* box. It should look like Figure 5.3.
5. Make sure you also ask SPSS to create a graph of the results by clicking on *Plots*. Put ‘Time’ along the x-axis (category axis) and enter ‘Course’ as separate lines. See Figure 5.4. Now click *Add* then *Continue*.

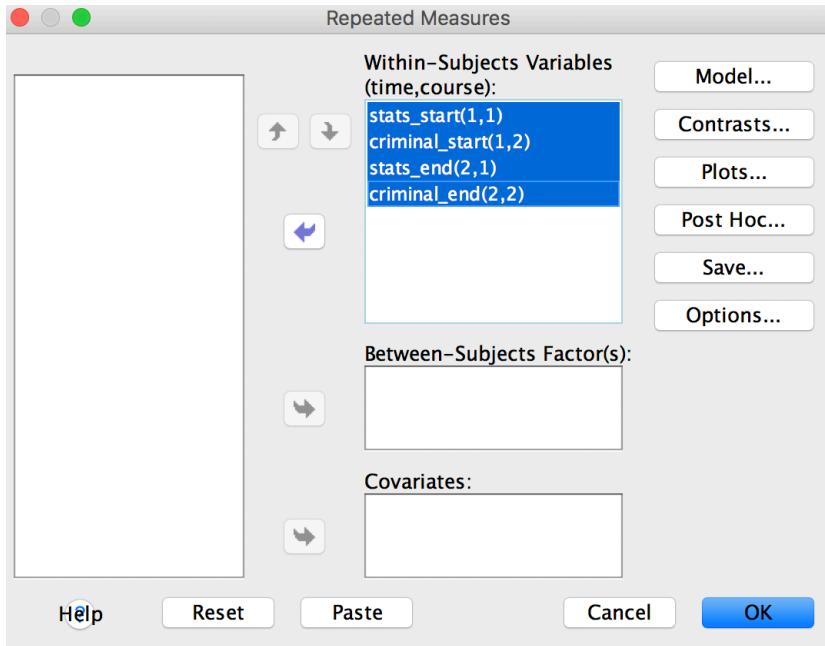


Figure 5.3: SPSS Repeated Measures Main ANOVA Dialogue, completed

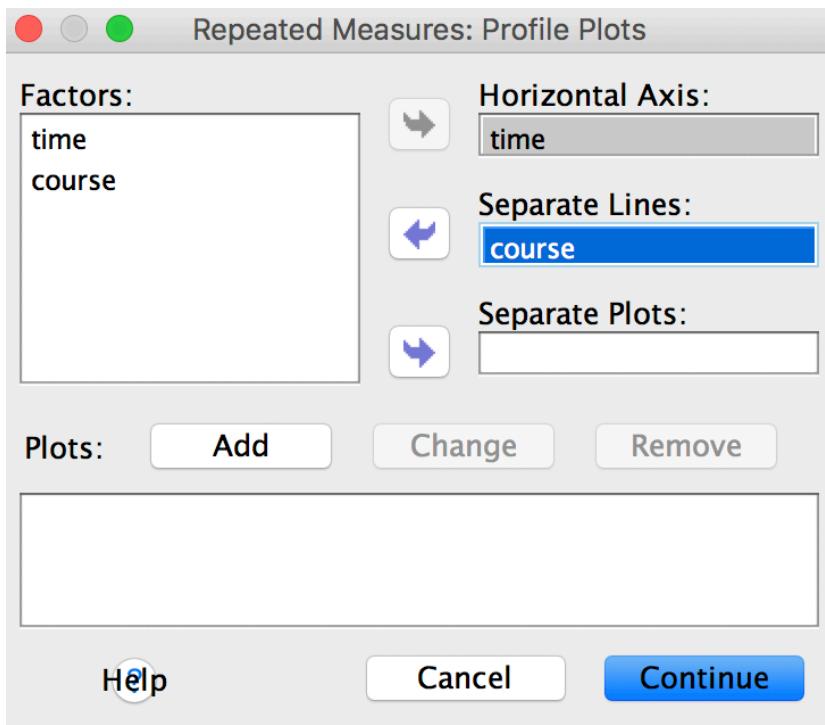


Figure 5.4: SPSS Repeated Measures Plots

6. Lastly, click on *Options* and move everything in the left-hand box into the right-hand box; also tick *Descriptives* and *Homogeneity Tests* (see Figure 5.5). Click on *Continue* and *OK*.

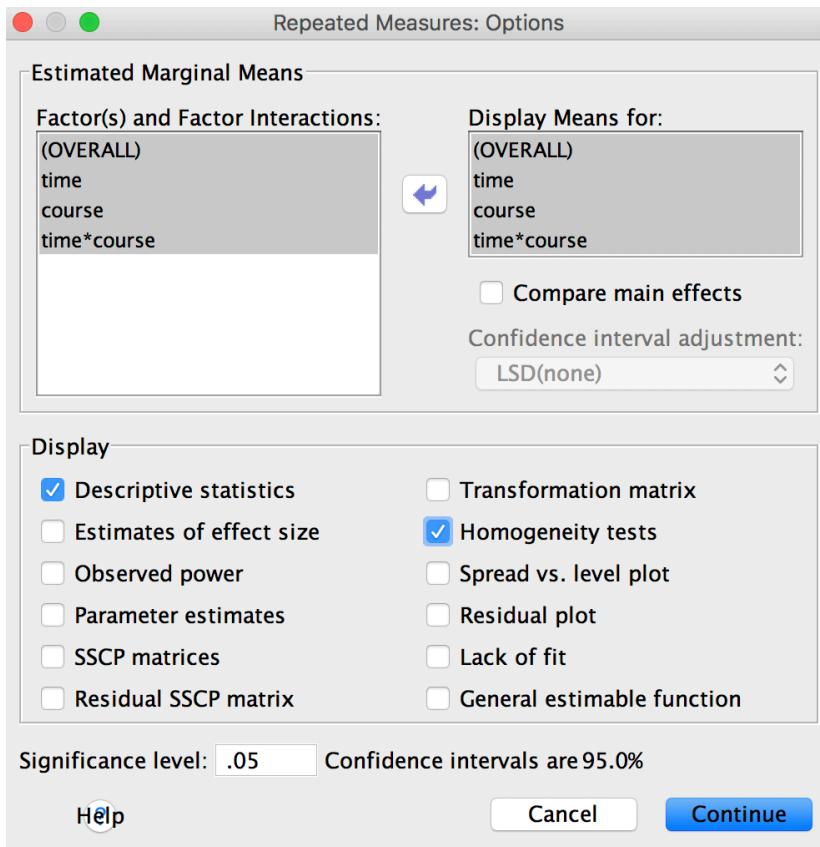


Figure 5.5: SPSS Repeated Measures Options

Now we're ready to run the ANOVA.

2. Go ahead and run the ANOVA as described above.

Answer

SPSS syntax for the ANOVA:

```
* Repeated measures ANOVA (Time by Course).
GLM stats_start criminal_start stats_end criminal_end
/WSFACCTOR=time 2 Polynomial course 2 Polynomial
```

```

/METHOD=SSTYPE(3)
/PLOT=PROFILE(time*course)
/EMMEANS=TABLES(OVERALL)
/EMMEANS=TABLES(time)
/EMMEANS=TABLES(course)
/EMMEANS=TABLES(time*course)
/PRINT=DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/WSDESIGN=time course time*course.

```

5.4.4 Step Four: Interpret the Output

We are now ready to answer the important question. Are those differences significant or not?

The table that we should look at first is *Mauchly's Test of Sphericity*. You should have one that looks like Figure 5.6.

Mauchly's Test of Sphericity ^a						
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b	
					Greenhouse-Geisser	Huynh-Feldt
time	1.000	.000	0	.	1.000	1.000
course	1.000	.000	0	.	1.000	1.000
time * course	1.000	.000	0	.	1.000	1.000

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: time + course + time * course

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Figure 5.6: Output—Mauchly's Test

- What do you make of the Mauchly's Test results?

Answer

- Mauchly's test cannot be completed without at least one factor with more than two levels. Because both our factors have two levels each, this table will remain blank.

- When a factor has two levels, we don't actually need to worry about sphericity (the property being assessed by Mauchly's test) because the data happen to have perfect sphericity in this case. To find out more, take a look at Field (2009).

Next, let's look at the *Tests of Within-Subjects Effects*, Figure 5.7.

Tests of Within-Subjects Effects					
Measure:	MEASURE_1	Type III Sum of Squares	df	Mean Square	F
time	Sphericity Assumed	18.225	1	18.225	14.232
	Greenhouse-Geisser	18.225	1.000	18.225	14.232
	Huynh-Feldt	18.225	1.000	18.225	14.232
	Lower-bound	18.225	1.000	18.225	.004
Error(time)	Sphericity Assumed	11.525	9	1.281	
	Greenhouse-Geisser	11.525	9.000	1.281	
	Huynh-Feldt	11.525	9.000	1.281	
	Lower-bound	11.525	9.000	1.281	
course	Sphericity Assumed	99.225	1	99.225	36.413
	Greenhouse-Geisser	99.225	1.000	99.225	36.413
	Huynh-Feldt	99.225	1.000	99.225	.000
	Lower-bound	99.225	1.000	99.225	.000
Error(course)	Sphericity Assumed	24.525	9	2.725	
	Greenhouse-Geisser	24.525	9.000	2.725	
	Huynh-Feldt	24.525	9.000	2.725	
	Lower-bound	24.525	9.000	2.725	
time * course	Sphericity Assumed	27.225	1	27.225	23.280
	Greenhouse-Geisser	27.225	1.000	27.225	.001
	Huynh-Feldt	27.225	1.000	27.225	.001
	Lower-bound	27.225	1.000	27.225	.001
Error (time*course)	Sphericity Assumed	10.525	9	1.169	
	Greenhouse-Geisser	10.525	9.000	1.169	
	Huynh-Feldt	10.525	9.000	1.169	
	Lower-bound	10.525	9.000	1.169	

Figure 5.7: Output—Tests of Within-Subjects Effects

- Answer the following questions in plain English and back them up with the appropriate statistics written in APA format. Remember, you will also need to look at the descriptive statistics to interpret

the findings fully.

4. Did people rate one course more positively than the other? Is there a need for further analysis?
5. Did people rate the courses differently from the start to the end of the course? And is there a need for further analysis?
6. Was the difference in ratings from the start to the end of the courses the same for both courses? Is there a need for further analysis?

Answer

4. Yes, the main effect of Course was significant: $F(1,9) = 36.41$, $MSE = 2.73$, $p < .001$. The means tell us that Criminal Psychology was rated more positively than Statistics. Because this factor has only two groups, there is no scope for further analysis.
5. Yes, there was a significant main effect of Time: $F(1,9) = 14.23$, $MSE = 1.28$, $p = .004$. The descriptive statistics tell us that ratings, averaged across both courses, were higher at the end of the courses than at the beginning. Since there are only two time points, there is no need for further analysis.
6. The difference in ratings was not the same for both courses, since the interaction between Course and Time was significant: $F(1,9) = 23.28$, $MSE = 1.17$, $p = .001$. We need to explore this with further analysis.

3. As you did last time, make the interaction graph look like the one in Figure 5.8 by using the SPSS graph editor.

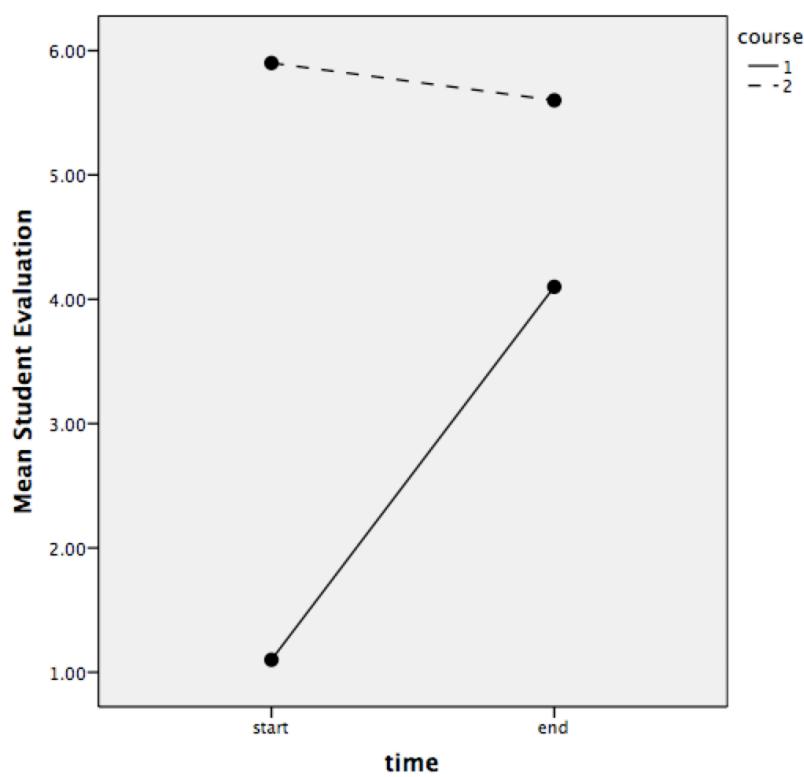


Figure 5.8: Output—Profile Plots

7. Generally, and with regards to a potential interaction, what does the graph show?

Answer

The lines are not parallel, which might indicate the interaction is significant. The pattern of cell means would indicate that the ratings for Statistics changed more than the ratings for Criminal between the start and end.

Since the interaction was significant, we'll need to carry out Simple Main Effects analyses. To do this, you'll need to carry out one-way repeated measures ANOVAs for each factor within the levels of the other factor. Carry out these analyses using the instructions in the lecture slides and answer the questions below. Remember that this analysis is very similar to one you've just done with two factors, but now you'll have one factor (for instance, Course) constrained to one level of the other factor (for instance, Time 1).

8. Were the Simple Main Effects of Course within each level of Time significant? Report the statistics to back up your response and explain what these results mean in plain English.

Answer

- The SME of Course was significant at the Start: $F(1,9) = 70.05$, $MSE = 1.64$, $p < .001$. The means show that Criminal Psychology was rated more positively than Statistics.
- The SME of Course was not significant at the End: $F(1,9) = 5.00$, $MSE = 2.25$, $p = .052$, which means that by the end of term, students rated Statistics and Criminal Psychology equally. (Alternatively, this can be treated as a marginally

significant difference, showing that Criminal was still liked better than Statistics, although the gap is much narrower than at the start of term.)

SPSS syntax for the above ANOVAs:

```
* SME of Course for Time1 (beginning).
GLM stats_start criminal_start
/WSFACTOR=Course_time1 2 Polynomial
/METHOD=SSTYPE(3)
/PLOT=PROFILE(Course_time1)
/EMMEANS=TABLES(OVERALL)
/EMMEANS=TABLES(Course_time1)
/PRINT=DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/WSDESIGN=Course_time1.

* SME of Course for Time2 (end).
GLM stats_end criminal_end
/WSFACTOR=Course_time2 2 Polynomial
/METHOD=SSTYPE(3)
/PLOT=PROFILE(Course_time2)
/EMMEANS=TABLES(OVERALL)
/EMMEANS=TABLES(Course_time2)
/PRINT=DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/WSDESIGN=Course_time2.
```

9. Were the Simple Main Effects of Time within each level of Course significant? Report the statistics to back up your response and

explain what these results mean in plain English.

Answer

- The SME of Time was significant for Statistics: $F(1,9) = 21.32$, $MSE = 2.11$, $p = .001$. The means show that ratings increased for Statistics from the start to the end of the course.
 - However, the SME of Time was not significant within Criminal Psychology: $F(1,9) = 1.33$, $MSE = 0.34$, $p = .279$. This means there was no change in ratings between the start and the end of the course for Criminal Psychology.
-

SPSS syntax for the above ANOVAs:

```
* SME of Time for Course1 (statistics).
GLM stats_start stats_end
/WSFACTOR=Time_Course1 2 Polynomial
/METHOD=SSTYPE(3)
/PLOT=PROFILE(Time_Course1)
/EMMEANS=TABLES(Time_Course1)
/EMMEANS=TABLES(OVERALL)
/PRINT=DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/WSDESIGN=Time_Course1.

* SME of Time for Course2 (criminal psychology).
GLM criminal_start criminal_end
/WSFACTOR=Time_Course2 2 Polynomial
/METHOD=SSTYPE(3)
/PLOT=PROFILE(Time_Course2)
/EMMEANS=TABLES(Time_Course2)
```

```
/EMMEANS=TABLES(OVERALL)  
/PRINT=DESCRIPTIVE  
/CRITERIA=ALPHA(.05)  
/WSDESIGN=Time_Course2.
```

5.5 Mixed ANOVA

So far you have looked at ANOVAs where all independent variables have been of one design or another (either repeated measures/within-groups or independent/between-groups). Now we will work through the analysis required when you have two factors (two-way), one of which has participants contributing to all conditions (repeated measures) while the other has different participants in each group (independent measures). We call this design *mixed*.

As always, the ANOVA will tell us three things:

1. Is there a main effect of IV1 on the DV?
2. Is there a main effect of IV2 on the DV?
3. Is there an interaction effect of the two IVs on the DV?

In the following analysis, we'll build upon an example you've looked at before.

A researcher put 30 depressed patients into one of three therapy groups (CBT, Sunshine therapy, no therapy). She found that having some therapy had a positive effect on sociability levels but it didn't matter whether that therapy was CBT or sunshine. Three months later the researcher was interested in finding out whether the effects of therapy had lasted so she got the group of patients together again and recorded, just like the first time, the amount of time each patient engaged in social activities.

5.5.1 Step One: Load the data into SPSS

4. Download the [data file](#).

5.5.2 Step Two: Run descriptive statistics

As usual, let's get an overview of what's going on with the data.

5. Calculate mean scores for the three groups. Do this by going to *Analyze > Compare Means > Means*. Put 'activity 1' and 'activity 2' in the *Dependent* box and 'group' into the *Independent* box. Now click *OK*.

Answer

SPSS syntax:

```
MEANS TABLES=activity1 activity2 BY group
/CELLS=MEAN COUNT STDDEV.
```

You should get Figure 5.9.

10. On the basis of the descriptives, do you think the effect of therapy varies over time?

Answer

Although the overall amount of sociability between time 1 and time 2 hasn't changed, there are differences depending on the group. Activity in the group receiving sunshine therapy has decreased from 2.50 to 1.10 hours, in the CBT group sociability increased from 2.80 to 4 hours, and in the no therapy group it remained more or less stable.

Report		Sociability level at time 1	Sociability level at time 2
Therapy Group			
No therapy	Mean	1.0000	1.200
	N	10	10
	Std. Deviation	1.05409	1.1353
CBT	Mean	2.8000	4.000
	N	10	10
	Std. Deviation	.91894	1.0541
Sunshine therapy	Mean	2.5000	1.100
	N	10	10
	Std. Deviation	.84984	.8756
Total	Mean	2.1000	2.100
	N	30	30
	Std. Deviation	1.21343	1.6887

Figure 5.9: Therapy and Sociability Descriptives

5.5.3 Step Three: Set up the ANOVA

We'll now see whether the differences in the sample can be said to be differences in the population. To do this, we'll use the same *Repeated Measures* SPSS menu option as before, because this dialogue box allows us to specify not only within-groups factors but between-groups factors as well. (In other words, don't look for an SPSS procedure specifically labelled 'mixed'.)

Read through the instructions below before doing them yourself.

1. Go to *Analyze > General Linear Model > Repeated Measures*. A window will appear where you can define the repeated measures factor (independent variable) you have. As we have only one repeated measures factor (Time: immediately after therapy and three months later), we will just define this one. If you can't remember how to define your repeated measures factor, have a look at [§ the repeated measures example](#).
2. Your next window, where you define factors, should look like Figure 5.10.

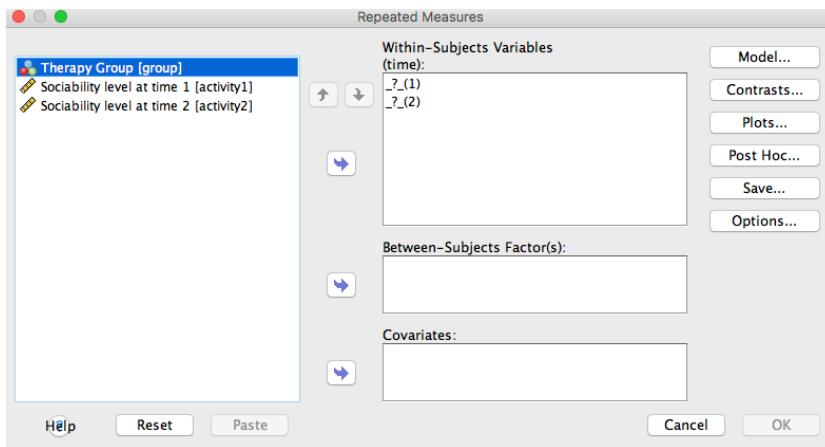


Figure 5.10: SPSS Repeated Measures Define Factors Dialogue

3. Put the two repeated measures conditions into the *Within-Subjects Variable* box and the independent measures factor into the *Between-Subjects Factor(s)* box. This should look like Figure 5.11.

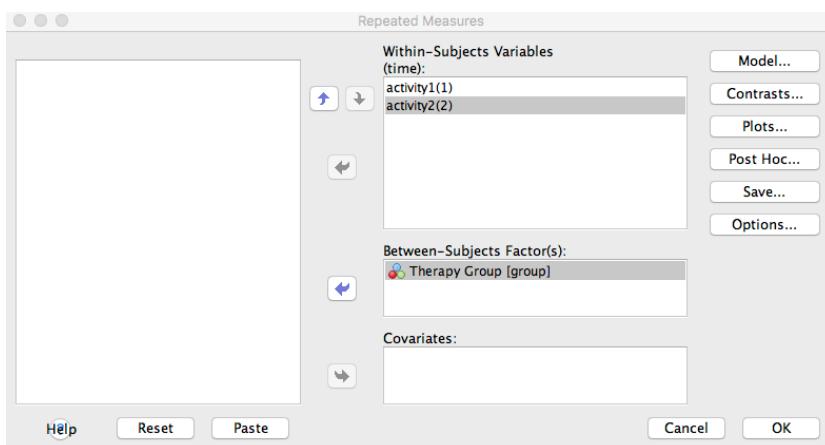


Figure 5.11: SPSS Repeated Measures Factor Selection

4. As you've done several times by now, click on *Options* ask for means for every effect. This will give us more detailed descriptive statistics. Click on *Continue*.
5. Ask SPSS for two graphs: one with 'Therapy Group' on the horizontal axis and 'Time' as separate lines; and one with 'Time' on the horizontal axis and 'Therapy Group' as separate lines. Click *Continue* and *OK*. You've now set up a basic two-way mixed ANOVA.

5. Carry out the ANOVA according to the method above.

Answer

SPSS syntax for the ANOVA:

```
* Mixed (Therapy Group by Time) ANOVA.  
GLM activity1 activity2 BY group  
/WSFACTOR=time 2 Polynomial  
/METHOD=SSTYPE(3)  
/PLOT=PROFILE(group*time time*group)  
/EMMEANS=TABLES(OVERALL)  
/EMMEANS=TABLES(group)  
/EMMEANS=TABLES(time)  
/EMMEANS=TABLES(group*time)  
/PRINT=DESCRIPTIVE  
/CRITERIA=ALPHA(.05)  
/WSDESIGN=time  
/DESIGN=group.
```

5.5.4 Step Four: Interpret the Output

Before we look at the basic statistics, let's look at the graphs produced by SPSS, Figure 5.12 and Figure 5.13. I've suggested you produce two of them because each represents a way of looking at the data. It could well be that you find one arrangement more intuitive than another; be aware, however, that they depict the same information.

10. What can we infer from the graph in terms of main and interaction effects? Explain how you arrived at your answer.

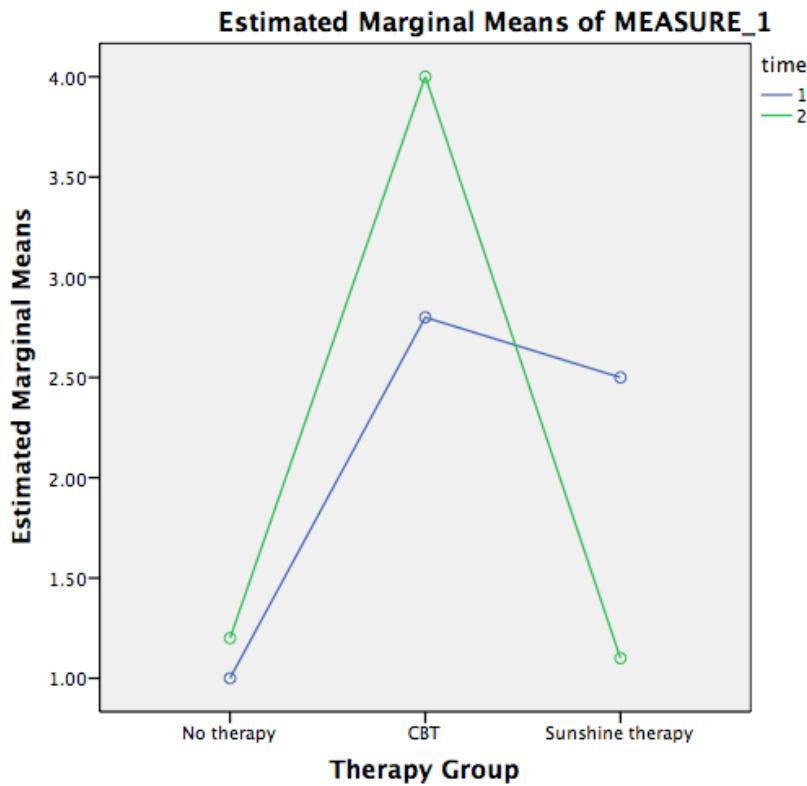


Figure 5.12: Output—Plot (Therapy by Time)

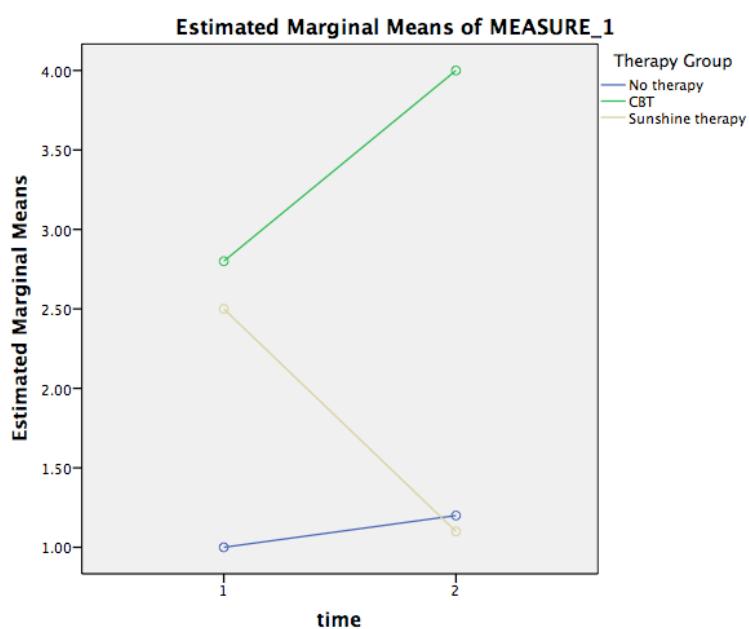


Figure 5.13: Output—Plot (Time by Therapy)

Answer

- Since the lines are not parallel, we might expect an interaction.
- The effect of Time would not appear to be significant (averaging data points across each line, there wouldn't be much of a difference between the two marginal means).
- There seems to be an effect of Therapy, as the marginal means for each group (average of Time 1 & Time 2 means for each Therapy group), are different.

Let's see if your inferences are backed up by the ANOVA. Again, there are a number of tables we don't need to look at to understand our findings. The first table that we should look at is *Mauchly's Test of Sphericity*. You should have one that looks like Figure 5.14.

Mauchly's Test of Sphericity ^a						
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b	
					Greenhouse-Geisser	Huynh-Feldt
time	1.000	.000	0	.	1.000	1.000

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + group
Within Subjects Design: time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Figure 5.14: Output—Mauchly's Test of Sphericity

11. What assumption does the above table relate to? Why are there no significance values?

Answer

The table relates to the Sphericity assumption. There are no significance values because it only applies when the within-subjects factors have more than two levels, which is not the case here.

Next, let's look at the *Tests of Within-Subjects Effects*, Figure 5.15. This will tell us about the main effect of our repeated measures factor (Time) and the interaction between Time and Therapy Group.

Tests of Within-Subjects Effects						
Measure: MEASURE_1		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	.000	1	.000	.000	1.000
	Greenhouse-Geisser	.000	1.000	.000	.000	1.000
	Huynh-Feldt	.000	1.000	.000	.000	1.000
	Lower-bound	.000	1.000	.000	.000	1.000
time * group	Sphericity Assumed	17.200	2	8.600	21.500	.000
	Greenhouse-Geisser	17.200	2.000	8.600	21.500	.000
	Huynh-Feldt	17.200	2.000	8.600	21.500	.000
	Lower-bound	17.200	2.000	8.600	21.500	.000
Error(time)	Sphericity Assumed	10.800	27	.400		
	Greenhouse-Geisser	10.800	27.000	.400		
	Huynh-Feldt	10.800	27.000	.400		
	Lower-bound	10.800	27.000	.400		

Figure 5.15: Output—Tests of Within-Subjects Effects

- Answer the following questions in plain English and back them up with the appropriate statistics written in APA format. Remember, you will also need to look at the descriptive statistics to interpret the findings fully.
12. Was there an effect of time on sociability? Is there a need for further analysis?
13. Was there an interaction between time and therapy on sociability? Is there a need for further analysis?

Answer

- No, there was no effect of time on sociability – this means that participants did not engage in more social activity in time 1 than in time 2 or vice versa: $F(1,27) < 0.001$, $MSE = 0.40$, $p > .99$. Since the effect is not significant, there is no need for further analysis. And, even if it were significant,

there is no need to track down where the difference might lie because the factor has only two levels.

- Yes, the interaction between time and therapy was significant, $F(2,27) = 21.50$, $MSE = 0.40$, $p < .001$. This means that there was a different amount of change in levels of activity between time 1 and time 2, depending on the therapy group. Because the interaction is significant, we need Simple Effects analysis to help us determine where the differences across time in Figure 5.13 actually differ.

The last thing to find out is whether there was an effect of therapy on sociability generally, regardless of how long after therapy the measure was taken. To find this out come down to the table titled *Tests of Between-Subjects Effects*, shown in Figure 5.16.

Tests of Between-Subjects Effects					
Measure:	MEASURE_1	Transformed Variable:	Average		
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	264.600	1	264.600	170.914	.000
group	55.600	2	27.800	17.957	.000
Error	41.800	27	1.548		

Figure 5.16: Output—Tests of Between-Subjects Effects

14. Was there a main effect of therapy on sociability? Is there a need for further analysis?

Answer

Yes, there was a difference in sociability levels depending on the group: $F(2,27) = 17.96$, $MSE = 1.55$, $p < .001$. We need to follow this up with post-hoc comparisons because we have three groups, so we don't know which mean differs from which.

In this example, we have a main effect of Therapy Group and an interaction between Therapy and Time. We need, therefore, to carry out *post hoc analyses*.

5.5.5 Investigating the Main Effect of Therapy Group

Therapy is a three-level factor (CBT, sunshine and no therapy). Our main effect tells us that there is a significant difference between at least one pair of levels, or groups (i.e. CBT v. sunshine, CBT v. no therapy, or sunshine v. therapy). Because we're looking for paired differences, we need to do *pairwise comparisons*.

6. Run the pairwise comparisons on Therapy Group to investigate the main effect.
 - Go back into *General Linear Model > Repeated Measures* and click *Define*. (The ANOVA should still be set up if you haven't exited SPSS.)
 - Go to *Options* and ask for the marginal means of 'Group' only. (You can remove the other effects and the 'overall' mean by highlighting them and clicking on the arrow between the two boxes).
 - Click on *compare main effects* and selected *Bonferroni* from the drop-down list immediately below. (This will offer pair-wise comparisons for the means of the levels of 'Group', controlling for family-wise error rate.) Click on *Continue*. Your dialogue box should look like Figure 5.17. Click *OK*.

Answer

SPSS syntax for the above analysis:

```
GLM activity1 activity2 BY group
```

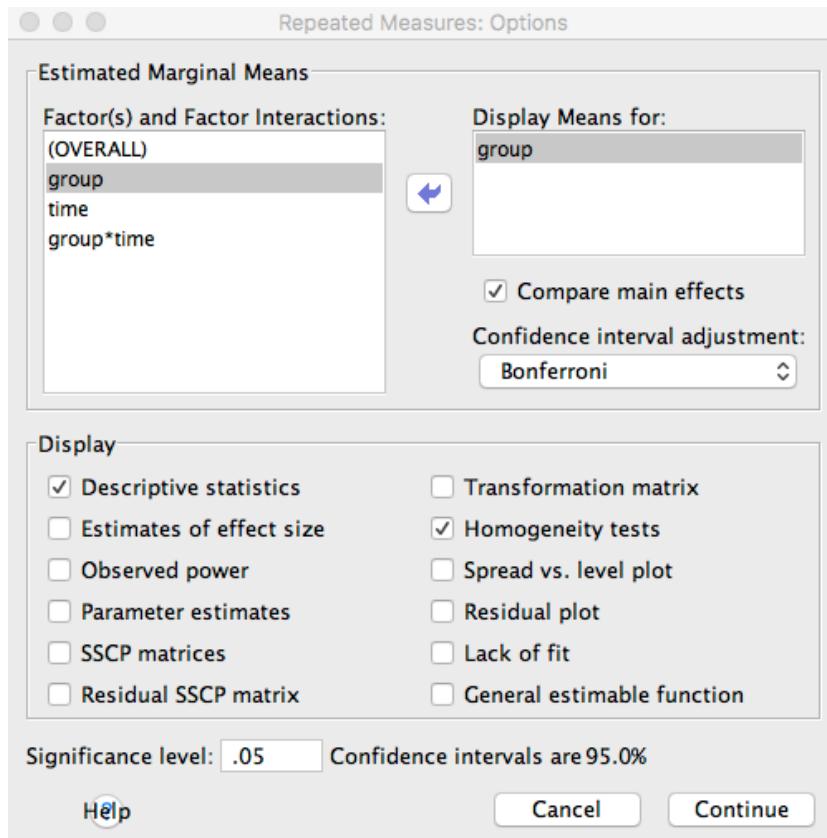


Figure 5.17: Repeated Measures ANOVA Dialogue, Pairwise Comparisons for Therapy Group

```

/WSFACTOR=time 2 Polynomial
/METHOD=SSTYPE(3)
/PLOT=PROFILE(group*time time*group)
/EMMEANS=TABLES(group) COMPARE ADJ(BONFERRONI)
/PRINT=DESCRIPTIVE HOMOGENEITY
/CRITERIA=ALPHA(.05)
/WSDESIGN=time
/DESIGN=group.

```

Your *Pairwise Comparisons* table should look like Figure 5.18.

Pairwise Comparisons						
(I) Therapy Group	(J) Therapy Group	Mean Difference (I-J)	Std. Error	95% Confidence Interval for Difference ^b		
				Sig. ^b	Lower Bound	Upper Bound
No therapy	CBT	-2.300*	.393	.000	-3.304	-1.296
	Sunshine therapy	-.700	.393	.259	-1.704	.304
CBT	No therapy	2.300*	.393	.000	1.296	3.304
	Sunshine therapy	1.600*	.393	.001	.596	2.604
Sunshine therapy	No therapy	.700	.393	.259	-.304	1.704
	CBT	-1.600*	.393	.001	-2.604	-.596

Based on estimated marginal means

*. The mean difference is significant at the

b. Adjustment for multiple comparisons: Bonferroni.

Figure 5.18: Output—Pairwise Comparisons for Levels within Therapy Group

15. What were the results of the post hoc tests on the main effect of Therapy? Report the statistics to back up your answer.

Answer

CBT differed from both no therapy ($p < .01$) and sunshine therapy conditions ($p < .01$). Looking at the means, we can see that CBT resulted in higher sociability than the other two. CBT is, therefore, driving the main effect that we see.

5.5.6 Investigating the Interaction of Therapy Group and Time

In our mixed analysis, we have an independent factor (Therapy Group) and a repeated factor (Time). For various reasons that we don't need to go into— unless you're interested; cf. Howell (1997)), it is quite straightforward to examine the interaction in terms of the repeated factor, but more difficult to examine it in terms of the independent factor (and we do not expect you to do this in your assignment). For this reason, we'll focus on looking at the interaction from the perspective of the repeated factor.

This means we'll be asking 'What is the effect of Time within each Therapy Group?' In other words, we're looking at how the repeated factor explains variability within each level of the independent factor. In other words, again, we're going to run three separate one-way repeated measures ANOVAs, presented below as the analyses followed by the question each analysis answers:

1. Time (for No Therapy): Did the No Therapy Group change over time?
2. Time (for CBT): Did the CBT group change over time?
3. Time (for Sunshine): Did the Sunshine group change over time?

At this point, it might be useful to touch upon the analysis we *won't* be doing, i.e. 'What is the effect of Therapy Group within Time?' This analysis, from the perspective of the independent/between-groups factor, would have allowed us examine:

1. Therapy Group (for Time 1): Did Therapy Groups differ at Time 1?
2. Therapy Group (for Time 2): Did Therapy Groups differ at Time 2?

In this case, such an analysis would be theoretically less interesting for us, since we're really interested in the effects of therapy *across* time,

which best answered by the three one-way repeated ANOVAs above (Time (for No Therapy) etc.).

- Again, if you find yourself needing to do Simple Main Effects from the perspective of the between-groups factor, [see this webpage](#), or Howell (1997).

Let's now do the Simple Main Effects analysis of Therapy Group from the perspective of Time. At the moment, our SPSS data file has all cases (participants) selected. We want to focus on each level of Therapy Group and perform an analysis on each. So, because Therapy Group is a between-groups factor, marking group membership on a case-by-case basis, we can simply use *Select Cases* to identify, or filter, the participants we're interested in.

- Read through the instructions for selecting cases and running each ANOVA before performing them.

- Start by going to *Data > Select Cases*.
- We're now going to describe our filter to SPSS. Select *If condition is satisfied*.
- Define the Therapy Group to be selected. Start with *If Group = 1*. Your dialogue box should look like Figure 5.19.
- Verify that the selection is working as you expect. Check that the *Data View* shows the appropriate cases to be *selected* (row number is untouched) and *unselected* (row number is struck out). When the filter is 'Therapy Group = 1', you should see something like Figure 5.20.
- Now that the scope of the data has been restricted to those participants in the No Therapy group, we can run a one-way repeated

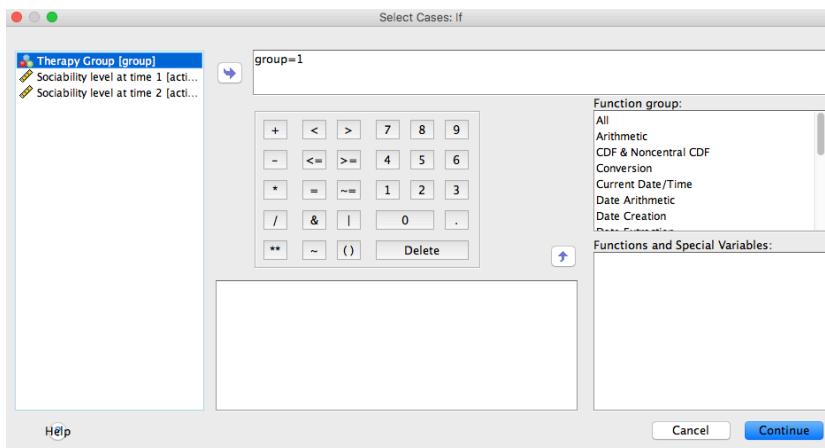


Figure 5.19: Define Cases where Therapy Group = 1, No Therapy

The screenshot shows the SPSS Data View window titled 'Untitled2 [DataSet2] - IBM SPSS Statistics Data Editor'. The top menu bar includes file, edit, view, analyze, transform, format, and help options. The toolbar below has icons for file operations, data entry, and analysis. The data table has a header row with columns: '1 : group' (containing '1.00'), 'group', 'activity1', 'activity2', 'filter_\$', 'var', 'var', 'var', and 'var'. The 'group' column contains values 1.00 through 16.00. The 'activity1' and 'activity2' columns show varying numerical values. The 'filter_\$' column contains mostly 1.00, except for row 11 which is 0.00. The 'var' columns are all empty. The data rows are numbered 1 through 16.

1 : group	group	activity1	activity2	filter_\$	var	var	var	var
1	1.00	2.00	2.0	1				
2	1.00	3.00	3.0	1				
3	1.00	1.00	3.0	1				
4	1.00	.00	1.0	1				
5	1.00	.00	.0	1				
6	1.00	.00	.0	1				
7	1.00	1.00	.0	1				
8	1.00	1.00	1.0	1				
9	1.00	2.00	1.0	1				
10	1.00	.00	1.0	1				
11	2.00	1.00	2.0	0				
12	2.00	3.00	3.0	0				
13	2.00	4.00	5.0	0				
14	2.00	3.00	4.0	0				
15	2.00	2.00	3.0	0				
16	2.00	3.00	4.0	0				
		---	---	-				

Figure 5.20: Data View with only Therapy Group = 1, No Therapy

measures ANOVA (the factor being Time) and our F ratio will tell us whether there is an effect of time within the No Therapy group.

7. Using the *Select Cases* procedure above, and referring to an earlier workshop booklet on one-way repeated measure ANOVA, go ahead and run one-way repeated ANOVAs of Time for each of the three levels of Therapy Group.
- Tip: If you paste the syntax from the *Select Cases* command, as well as the ANOVA, you may find changing the group variable and running the syntax much faster than doing everything manually via the menu system.

Answer

SPSS syntax for the above analysis:

```
* Examining Simple Main Effects of Time
* within levels of Therapy Group.

* Start by defining our filter to select
* case where Group = 1 (no therapy).
USE ALL.

COMPUTE filter_$(group=1).
VARIABLE LABELS filter_$ 'group=1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$(f1.0).
FILTER BY filter_$. 
EXECUTE.

* Repeated measures ANOVA, one way,
* examining effect of Time within
* Therapy Group = 1.
GLM activity1 activity2
```

```
/WSFACTOR=time 2 Polynomial  
/METHOD=SSTYPE(3)  
/PRINT=DESCRIPTIVE  
/CRITERIA=ALPHA(.05)  
/WSDESIGN=time.
```

* Select cases for Therapy Group = 2.

```
USE ALL.  
  
COMPUTE filter_$(group=2).  
VARIABLE LABELS filter_$ 'group=2 (FILTER)'.  
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.  
FORMATS filter_$ (f1.0).  
FILTER BY filter_$.  
EXECUTE.
```

* Repeated measures ANOVA, one way,

* examining effect of Time within

* Therapy Group = 2.

```
GLM activity1 activity2  
/WSFACTOR=time 2 Polynomial  
/METHOD=SSTYPE(3)  
/PRINT=DESCRIPTIVE  
/CRITERIA=ALPHA(.05)  
/WSDESIGN=time.
```

* Select cases for Therapy Group = 3.

```
USE ALL.  
  
COMPUTE filter_$(group=3).
```

```

VARIABLE LABELS filter_$ 'group=3 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

* Repeated measures ANOVA, one way,
* examining effect of Time within
* Therapy Group = 3.
GLM activity1 activity2
/WSFACTOR=time 2 Polynomial
/METHOD=SSTYPE(3)
/PRINT=DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/WSDESIGN=time.

```

16. Report what the above analysis tells you.

Answer

- Sociability levels did not change for those in the No Therapy group: $F(1,9) = .47$, $MSE = 0.42$, $p = .51$.
- There was a significant change in sociability for those in the CBT group, $F(1,9) = 17.05$, $MSE = 0.42$, $p = .003$, and in the sunshine group, $F(1,9) = 27.56$, $MSE = 0.36$, $p = .001$. Looking at the means in the *Descriptive Statistics* table, we can see that sociability went up for the CBT group and decreased in the sunshine group.

6

ANOVA Overview

6.1 Overview

<i>Data files required</i>	Wellbeing data
	Wellbeing data, trimmed

<i>Booklet Version</i>	1.0
------------------------	-----

6.2 Objectives

This workshop will reinforce some of the assumptions of ANOVA and what to do when they are broken. It will also introduce measures of effect size for the ANOVA. We'll begin with the ANOVAs (between groups and within) and then consider the effect size statistics R^2 , r and ω^2 .

6.3 Workshop

This workshop represents just about as far as we'll go with ANOVA.

Answer

SPSS data files and syntax:

Wellbeing data
Wellbeing syntax
Wellbeing data, trimmed

8 Wellbeing syntax, trimmed

6.4 Between-Groups ANOVA

For this example, we'll use the 8 Wellbeing data set. This study captured questionnaire data on life satisfaction and some other demographic variables. It is not an experiment—in which groups are created and manipulated—but uses existing groups in the manner of a correlational design.

1. Spend a few moments looking at the data. Switch between the *Variable View* and the *Data View* to get a sense of it. Pay particular attention to the labels given each to variable and what its values are.

We're going to run an analysis looking at differences in 'lifenow' ("Overall, how satisfied are you with your life nowadays?") by 'high-education' ("What is the highest level of qualification?"; this can be broken into four levels, as you can see from the *Variable View*).

2. How would you describe this analysis?

Answer

- This is a one-way between groups ANOVA looking at the effect of Education Level (degree, below degree, other, none) on life satisfaction.

3. As usual, we should start by looking at the means. Using *Analyze > Compare Means > Means* investigate the overall pattern of means for life satisfaction broken down by Education.

Answer

- The picture in the means (6.78, 7.34, 7.00, and 5.76) is unclear, though it could well be that those in the ‘degree’ group and those in the ‘none’ group have a slightly lower life satisfaction.
-

SPSS syntax for the above:

```
MEANS TABLES=lifenow BY higheducation
/CELLS=MEAN COUNT STDDEV.
```

Recall that between groups ANOVA has three main assumptions:

1. Independence of cases
 2. Normality
 3. Equality (or “homogeneity”) of variances
- We can test for (2) by trying to ensure that the dependent variable is normally distributed within the different levels of the factor. There are many ways of assessing normality but I’d like to focus on Q-Q plots and the Shapiro-Wilk Test.
 - We can test for (3) by looking at variance differences. For this, we’ll examine the statistic produced by Levene’s Test.

6.4.1 Assessing Normality

Read through the instructions below, which describe how you can obtain some measures of normality, before you actually follow the instructions.

1. Begin by heading to *Analyse > Descriptive Statistics > Explore*.

2. Enter the appropriate variables in the *Dependent List* and the *Factor List* (Figure 6.1).

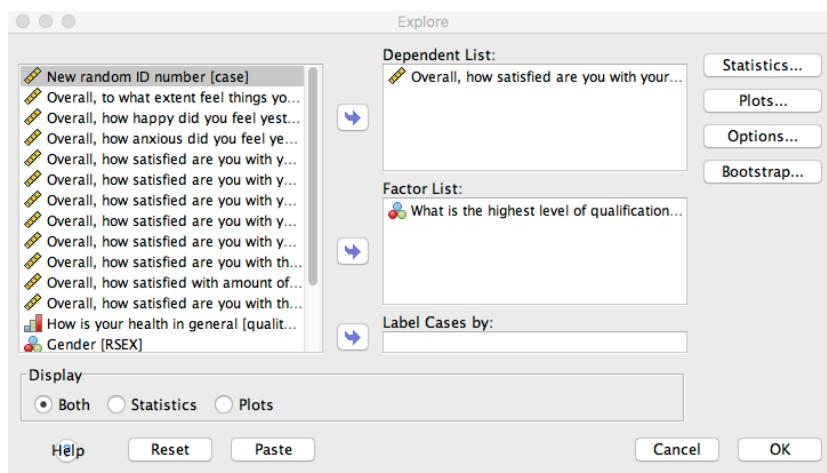


Figure 6.1: SPSS Explore

3. Go to *Plots* and set the options described in Figure 6.2.
4. Click *OK* and brace yourself for a raft of output.

1. Go ahead and run the *Explore* command as described above.

Answer

SPSS syntax for the above:

```
EXAMINE VARIABLES=lifenow BY higheducation
/PLOT BOXPLOT STEMLEAF HISTOGRAM NPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```

There are two things to look at:

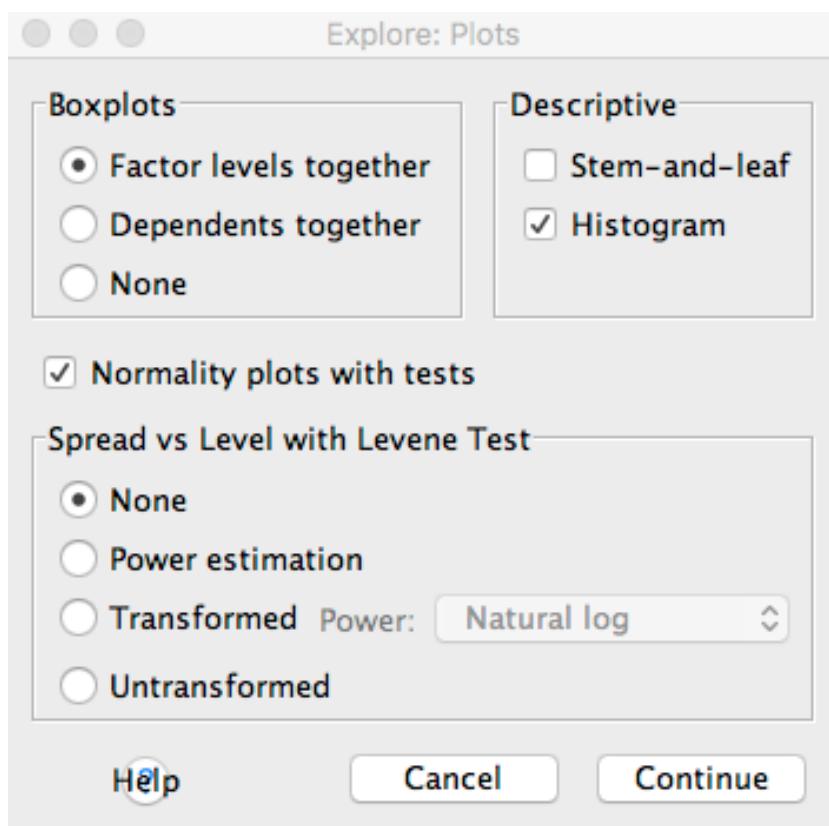


Figure 6.2: SPSS Explore—Plots

1. Q-Q Plots
2. The Shapiro-Wilk statistic

Let's talk about the Q-Q plots first. When data are normally distributed, data points should fall along the diagonal. Figure 6.3 shows the Q-Q plot for life satisfaction when the educational level is 'other'. There's certainly deviation from the diagonal line, and in this case I'd say we need to be a bit careful about whether we can assume normality.

Normal Q-Q Plot of Overall, how satisfied are you with your life nowadays?

for higheducation= Other

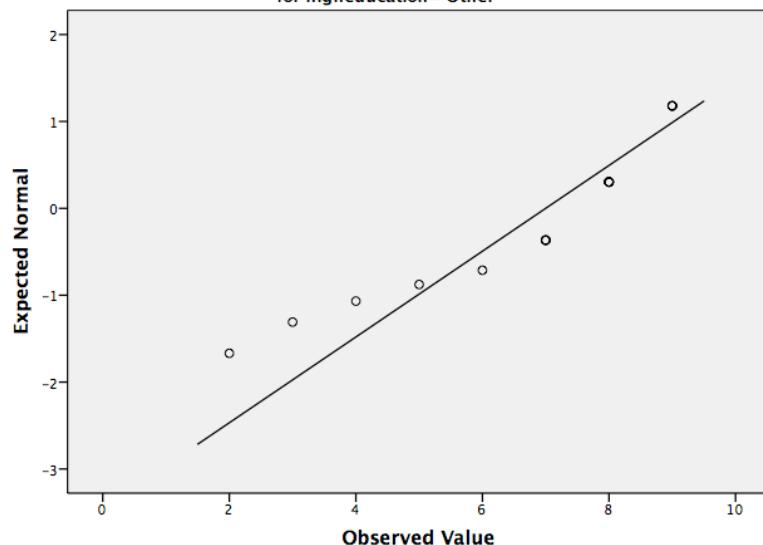


Figure 6.3: SPSS Output—Q-Q Plot for life satisfaction (now) when the educational level is 'other'

4. Look at the other Q-Q plots. What do you make of them?

Answer

Certainly, none is quite so out-of-line as Figure 6.3. But there are grounds for being cautious about their normality.

Now we'll turn to the Shapiro-Wilk statistic. The null-hypothesis of this test states that the population is normally distributed. Thus,

when we see a p value that is lower than our criterion value of 5%, we can call this a deviation from normality. To view the statistic for each factor level, look at Figure 6.4.

Tests of Normality						
What is the highest level of qualification?	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Overall, how satisfied are you with your life nowadays?	Degree or equivalent	.197	54	.000	.913	54
	Below degree level	.196	128	.000	.923	128
	Other	.250	20	.002	.829	20
	None (no qualifications)	.158	41	.012	.937	41

a. Lilliefors Significance Correction

Figure 6.4: SPSS Output—Tests of Normality

5. Interpret this table in plain English. What do you see here?

Answer

The Shapiro-Wilk statistics are all significant, so we may well have problems with our ANOVA achieving the normality.

The last assumption, equality of variances, will be tested by the Levene statistic, which we can compute when we run the ANOVA itself.

Let's prepare to run a one-way ANOVA to test the effect we've been talking about. Before you run it, make sure you select the options shown in Figure 6.5.

Under *Post hoc* select the comparisons Bonferroni, Tukey, Scheffe and Games-Howell, as you see in Figure 6.6.

2. Run a one-way ANOVA of Education Level on Life Satisfaction (now). Refer to the earlier seminar booklet if you need to.

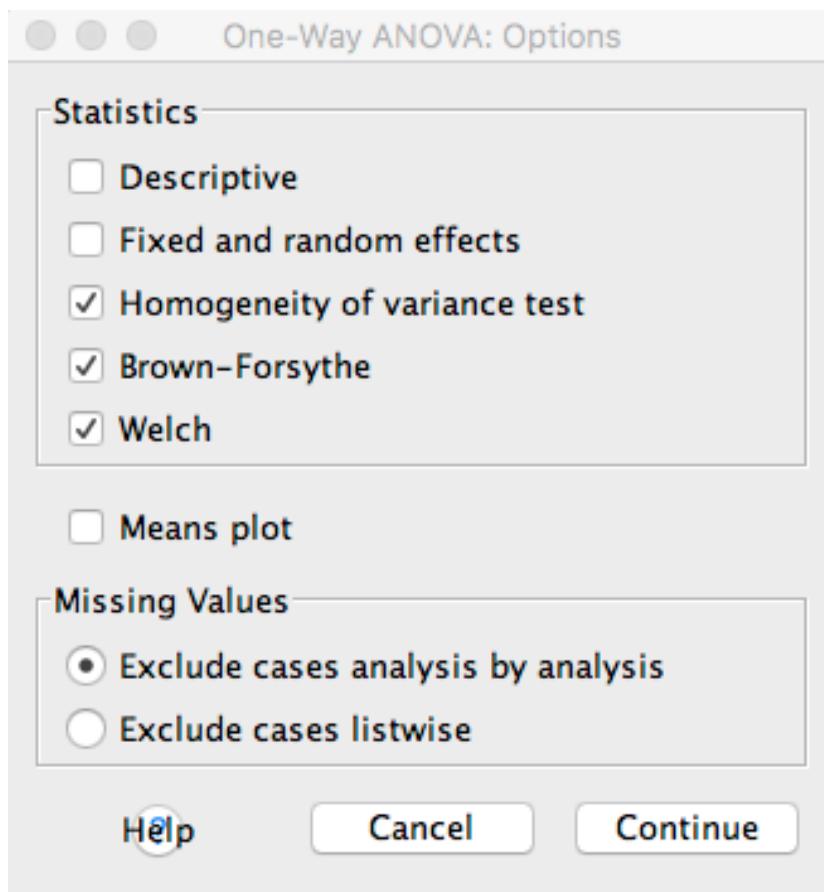


Figure 6.5: SPSS ANOVA Options

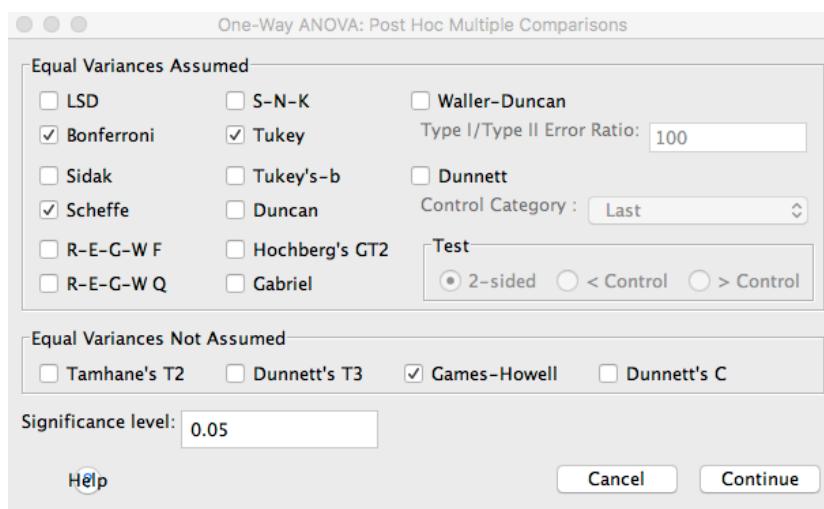


Figure 6.6: SPSS ANOVA Options, Post Hoc

Answer

SPSS syntax for the above ANOVA:

```
ONEWAY lifenow BY higheducation
/STATISTICS DESCRIPTIVES HOMOGENEITY BROWNFORSYTHE
WELCH
/PLOT MEANS
/MISSING ANALYSIS.
/POSTHOC=TUKEY SCHEFFE BONFERRONI GH ALPHA(0.05).
```

- What do you make of Levene's test for the equality of variances?

Answer

- In the table *Test of Homogeneity of Variances*, Levene's statistic(3,239) = 7.94, $p < .001$.
- Because the null hypothesis for this test is for the equality of variances, we can say that the ANOVA has failed this assumption.

Given that (spoiler alert) the ANOVA has failed the test of equality of variances (probably because of the uneven cell sizes), we will need to look for an adjusted statistic; the original F is no longer reliable. The one we'll look for in this case is the Welch corrected F , which can be found in Figure 6.7 (under some circumstances, which you can read more about in Field, 2009, you might wish to report the Brown-Forsythe).

We report Welch's F like this (fictitious example):

Welch's $F(3,30.17) = 10.95, p < .001$

Robust Tests of Equality of Means

Overall, how satisfied are you with your life nowadays?

	Statistic ^a	df1	df2	Sig.
Welch	4.303	3	63.994	.008
Brown-Forsythe	5.495	3	105.797	.002

a. Asymptotically F distributed.

Figure 6.7: SPSS Output—Robust Tests of Equality of Means

7. Report Welch's F for the ANOVA you've just performed.

Answer

Welch's $F(3,63.99) = 4.30, p < .01$. We're seeing a main effect (adjusted for the violation of homogeneity of variance) of Education.

When Levene's test is significant, we'll need to look at a corrected pairwise comparison, which in this case is the Games-Howell post hoc test.

8. Where do the differences lie within education level?

Answer

The Games-Howell statistic tells us the difference lies between 'none' (mean of 5.76; the lowest) and 'below degree level' (mean of 7.34; the highest).

6.5 Within Groups ANOVA

Download the SPSS data  Wellbeing data, trimmed.

The repeated measures ANOVA has two assumptions:

1. Normality
2. Sphericity

3. Familiarise yourself with the new data set. Notice that we now have fewer cases and that some variables are measured at three different times (T1, T2 and T3).

9. Use *Explore*, as you did  [above](#), to produce a descriptives report on the answer to the question ‘Overall, how satisfied are you with your work situation?’ at time 1, time 2 and time 3 (work_T1, work_T2 and work_T3), as well as the answer to the question ‘Overall, how satisfied are you with your financial situation?’ (financial_T1, financial_T2 and financial_T3). What do you think of the normality? Use the Q-Q plots and Shapiro-Wilk Test statistics.
- Don’t put anything in the factor list of the *Explore* box; that’s more useful when you have a between-groups factor. Simply drop all four variables into the *Dependent List* box.

Answer

- In terms of means, the highest mean is the Time 2 measure of Work satisfaction ($M=5.89$) while the lowest is the one reported by participants for financial satisfaction at Time 3 ($M=5.04$). (Remember too that standard deviation provides a measure of dispersion, and this can be informative.)
- In terms of the Q-Q plots and Shapiro-Wilk Test statistics, the data are not normally distributed. There’s a good

chance that our ANOVA will be unreliable under these circumstances.

SPSS syntax for the above:

```
EXAMINE VARIABLES=work_T1 financial_T1 work_T2  
financial_T2 work_T3 financial_T3  
/PLOT BOXPLOT HISTOGRAM NPLOT  
/COMPARE GROUPS  
/STATISTICS DESCRIPTIVES  
/CINTERVAL 95  
/MISSING PAIRWISE  
/NOTOTAL.
```

4. Use what you've learned in an earlier booklet to run a two-way repeated measures ANOVA to test for differences in Work and Financial satisfaction at Time 1, Time 2 and Time 3. Thus, you will have two factors: Wellbeing (Work or Financial) and Time (1, 2 and 3).
 - Use the *Plots* menu to create a Time by Wellbeing plot and another one for the reversed order, Wellbeing by Time. This will give you two plots so you can see the effects from different perspectives.
 - In the *Options* menu put the following into *Display means for*: overall, Time, Wellbeing and Time*Wellbeing. This will give you a comprehensive set of means.
 - If you want to do follow-up analyses, use the *compare main effects* with a Bonferroni adjustment. Check the *Descriptive Statistics* box.

Answer

SPSS syntax for the above:

```
GLM work_T1 work_T2 work_T3 financial_T1
financial_T2 financial_T3
/WSFACTOR=Wellbeing 2 Polynomial Time 3 Polynomial
/METHOD=SSTYPE(3)
/PLOT=PROFILE(Wellbeing*Time Time*Wellbeing)
/EMMEANS=TABLES(OVERALL)
/EMMEANS=TABLES(Wellbeing) COMPARE ADJ(BONFERRONI)
/EMMEANS=TABLES(Time) COMPARE ADJ(BONFERRONI)
/EMMEANS=TABLES(Wellbeing*Time)
/PRINT=DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/WSDESIGN=Wellbeing Time Wellbeing*Time.
```

When Mauchly's Test shows non-sphericity, we need to report an adjusted F ratio such as the Greenhouse-Geisser.

10. What do you make of Mauchly's Test?

Answer

Given that this highly significant for Time ($p < 0.001$), we need to apply the Greenhouse-Geisser correction to both the F ratio and its associated Mean Square Error.

11. Making use of adjusted statistics where necessary, report all main and interaction effects for the ANOVA.

Answer

- Wellbeing: $F(1,111) = 9.97$, $MSE = 6.39$, $p < 0.01$. Work satisfaction was higher than financial satisfaction.
- Time: $F(2,222) = 0.46$, $MSE = 15.756$, $p = .955$. There was no change attributable to time overall.
- Wellbeing * Time: $F(2,222) = 4.91$, $MSE = 5.06$, $p < 0.05$. There was interaction between Wellbeing and Time, which could be further investigated using Simple Main Effects.

6.6 Effect Sizes

In the second seminar booklet, you learned about effect size. This is a simple way of quantifying differences. It's particularly handy—once you've noted which effect size is being used—for getting a quick sense of how large an experimental effect actually is.

The two measures of effect size I'd like to concentrate on are r , which is the correlation statistic that you're used to, and ω^2 , or *omega squared*.

It would be nice if SPSS produced these for us. Unfortunately, it doesn't.

6.6.1 r

In the first seminar booklet, you looked at the statistic r . This is a measure of association that runs from -1 to +1. If you take a correlation of .5, for instance, you can claim a strong correlation. We might see such an association when we collect data on temperature and road rage incidents: a correlation of .5 would tell us that, as temperate increases, so too does road rage.

The r statistic can also tell something else about variance. When we take its square, we derive a measure of the variability overlap—or *shared variance*—between the two variables in the correlation. For

instance, keeping to the example of $r = .5$, look what happens when we square it:

$$r = .5$$

$$R^2 = .25$$

This value, R^2 ¹, gives us a measure of how much variability in the first variable is shared with—or accounted by—the second variable. So we can now say that there is a 25% shared variability between temperature and road rage. In other words, 25% of the variability in road rage can be accounted for by temperature, and vice versa. Because R^2 is the square of r , which can run between -1 and +1, R^2 will always be a positive number between 0 and 1. You can always multiply R^2 by 100 to get a percentage score of overlap.

Now let's apply this to the ANOVA, starting with a one-way between subjects ANOVA.

Think about effect size for ANOVA as the amount of variability explained by our factor versus the total amount of variability that we have. In other words, we can consider it as the amount of variability explained by our factor versus the amount of variability left over. Thus, it is the sum of squares for the model, SS_M , versus the total sum of squares, SS_T .

A better way of expressing this is as a fraction representing the proportion of all the variability in the data that the factor can explain, and this is shown in Equation 6.1.

$$R^2 = \frac{SS_M}{SS_T} \quad (6.1)$$

This value is also known as η^2 , or *eta squared*, as shown in Equation 6.2.

$$R^2 = \eta^2 \quad (6.2)$$

We can either stick with R^2 and interpret our findings in terms of

¹ R^2 can be written in uppercase or as r^2 . However, I'll stick to R^2 because you'll most commonly see it in this form.

shared variance, or we can simply take the square of R^2 to find r .

Let's do an example—which happens to be the one you'll find in Field (2009, p.389). Here, we're looking at a one-way ANOVA with a factor whose model sum of squares is 20.13 and whose total sum of squares is 43.73.

$$R^2 = \frac{SS_M}{SS_T}$$

$$= \frac{20.13}{43.73}$$

$$= .46$$

This gives us a shared variance of .46. In other words, the factor accounts for 46% of the total variability in the study.

Now let's convert this into r , which will give us a more familiar statistic.

$$r = \sqrt{.46}$$

$$= .68$$

We now have an r statistic of .68, and from here we can apply the benchmark of Cohen (1992, as cited by Field, 2009) in Table 6.2.

Table 6.2: Interpreting Effect Size for R

r	Effect size	Variance accounted for
.1	Small	1%
.3	Medium	9%
.5	Large	25%

Thus, our value of $r = .68$ would be a large effect size.

Remember that effect size interpretations need to be made in the context of the literature, not just the rule of thumb above. For instance, in a clinical context, one might need to reach a particular effect size in a study comparing the effectiveness of a new drug against a control group in order for that drug to be viewed as clinically effective.

12. Assume that you have produced an ANOVA whose main effect SS_M is 4.6 and SS_T is 30.1.

- Report the R^2 .
- Report the r .
- Interpret the r according to Table 6.2

Answer

- $R^2 = .46$
- $r = \sqrt{R^2} = .68$
- According to Table 6.2, this is a very large effect.

You might have noticed above that our measure of variance was the *total* squares, not the *mean* squares. The mean squares are, as we've seen earlier in the module, the sum of squares divided by *degrees of freedom*. This allows the F ratio to be a better fit for the population by adjusting it slightly (according to how many groups and participants we have). Without this adjustment, the F ratio tells us more about our sample and less about the population. As psychologists, we're more interested in the population than the sample.

The same is true, then, for r . When it is based on unadjusted sums of squares, it's a poorer fit for the population. The adjusted statistic

that you often see for this task is ω^2 , or *omega squared*. Calculating ω^2 is similar to calculating r .

The equation for ω^2 is below..

$$\omega^2 = \frac{SS_M - (df_M)MS_E}{SS_T + MS_E} \quad (6.3)$$

Let's talk about what's happening in Equation 6.3. The top part of the fraction is still the variability explained by the model. The bottom part is still the total variability. But, now, the top part is adjusted so that as the error for the model, MS_E , gets larger, the top part of the fraction gets smaller; this means that, the larger the error, the more likely the overall effect is to get smaller. Secondly, the bottom part will become greater as the model error, MS_E , increases, which in turn reduces the overall effect size.

The other thing to plug into Equation 6.3 is the degrees of freedom for the effect, which you can get from SPSS output².

Let's go through an example. Again, we'll use the same one as Field (2009), p. 389.

$$\omega^2 = \frac{20.13 - (2)1.97}{43.73 + 1.97}$$

$$= \frac{16.19}{45.70}$$

$$= .35$$

So our ω^2 statistic, an adjusted effect size, is (at .35) a little smaller than our R^2 (which was .46), but we can be sure it represents the population more accurately.

Remember that ω^2 , in its squared form, is a measure of shared variance telling you how much of a factor's variance overlaps with the total variability. In the form ω —having taken the square root—interpretation is just like r .

² For a main effect, this will be the number of levels in the factor minus one.

13. Assume that you have run a one-way between groups ANOVA and see the output in Figure 6.8. This shows a significant main effect of Therapy Group on social activity level.

- Report ω^2
- Remembering that ω is equivalent to r , interpret ω according to Table 6.2.

ANOVA

Social activity level	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	18.600	2	9.300	10.419	.000
Within Groups	24.100	27	.893		
Total	42.700	29			

Figure 6.8: SPSS Output—One Way ANOVA

Answer

- $\omega^2 = \frac{18.6 - (2)0.893}{42.7 + 0.893} = 0.39$
- $\omega = \sqrt{\omega^2}$
- $\omega = .62$
- According to Figure 6.8, this is a very large effect size, so we can call the effect of Therapy Group a very large one.

As Field (2009) points out, most of the time a statistic like ω^2 isn't very informative because it's telling us about the overall effect for a factor, which can often have several levels. It's somewhat more useful to examine the effect size of two-group comparisons. Knowing that the effect size refers to two groups makes it easier to interpret. The statistic we'll use for this is, once again, r .

Let's say that we've just run a t-test on two levels in a factor (as a follow-up to a main effect) and found t statistic of 2.474. Our degrees of freedom can be found in the SPSS output. In this case, $df = 12$.

Have a look at Equation 6.4.

$$r = \sqrt{\frac{t^2}{t^2 + df}} \quad (6.4)$$

Using this, we can run through the figures to determine r .

$$\begin{aligned} r &= \sqrt{\frac{2.474^2}{2.474^2 + 12}} \\ &= \sqrt{\frac{6.12}{18.12}} \\ &= .58 \end{aligned}$$

Now that we have the value for r , we can look it up using Cohen's (1992, as cited by Field, 2009) criteria in Table 6.2 and see that it is a large effect size.

14. Assume that you have found a significant difference between two conditions, where $t = 1.82$ and $df = 20$.

- Report the r .
- Interpret the r according to Table 6.2

Answer

- $r = \sqrt{\frac{1.82^2}{1.82^2+20}} = 0.38$
- This is a medium effect size.

Use effect sizes where you can and report them alongside the statistic whose effect size you are estimating.

7

Advanced Nonparametric Statistics

7.1 Overview

<i>Data files required</i>	Mann-Whitney data
	Wilcoxon data
	Kruskal-Wallis data
	Friedman data

<i>Booklet Version</i>	1.0
------------------------	-----

7.2 Objectives

This workshop will introduce you to nonparametric tests of difference for two independent samples, two paired samples, multiple independent samples, and multiple related samples.

7.3 Workshop

In this workshop, we'll look at the [Mann-Whitney](#), [Wilcoxon](#), [Kruskal-Wallis](#) and [Friedman](#) tests, which are all nonparametric tests of group differences.

The *parametric* statistical tests you've looked at so far on MPSMD2RES, like the t-test, make assumptions about your population. For instance, if you measure reaction times in two conditions and compare them using a t-test, the t-test assumes that the reaction time population¹

¹ The population represents all the possible reaction times that you might have sampled.

has characteristics such as a normal distribution.

Nonparametric statistical tests—sometimes called *distribution-free*—rely less on assumptions about the population. Because a test is distribution-free, that makes it very useful under certain circumstances, such as when your sample² is taken from a population that may not have the restrictive characteristics required by parametric tests. Now, nonparametric tests do make *some* assumptions; but these are far more liberal than those of parametric tests. Partly as a result of this, they have lower *power*—i.e., they are less likely to detect an effect if it exists. Nonparametric tests do, however, have an advantage. They are less sensitive to outliers because their calculations are based on ranks (where last is last, no matter how far from the mean), and thus work as tests of *median* differences.

² Remember that your sample should look like your population, as long as the sample is a fair one.

Answer

SPSS data files and syntax:

- [Mann-Whitney data](#)
- [Mann-Whitney syntax](#)
- [Wilcoxon data](#)
- [Wilcoxon syntax](#)
- [Kruskal-Wallis data](#)
- [Kruskal-Wallis syntax](#)
- [Friedman data](#)
- [Friedman syntax](#)

7.4 Mann-Whitney Test

Use this test when you wish to test for the differences of *two conditions* of *independent* data. This is the nonparametric equivalent of the independent samples t-test.

Assumptions of this test are:

- Observations from the two groups are independent
- The observations are *ordinal* (i.e. they can be put in a meaningful order; or, in other words, they can be ranked)

The null hypothesis (H_0) for this test is:

The overall difference between the two groups is zero

As usual, our p will tell us the probability of seeing our observed data under the assumption that H_0 is true.

- Our alternative hypothesis (H_1) is that ranked observations in each group are different. Whereas a t-test, being parametric, tells us about the likelihood of finding *mean* differences, the Mann-Whitney test tells us more about the *median* difference; in other words, these are both tests of central tendency but use different measures of central tendency.

In the following example, a researcher was interested in the effect of distraction on a measure of verbal working memory called reading span (Daneman & Carpenter, 1980). She had participants attempt this test—providing a *reading span score*—under either noisy or silent conditions. A higher reading span score indicate higher reading span.

1. Download and familiarise yourself with the  data

2. Using *Analyse > Descriptive Statistics > Explore*, get descriptive statistics for the dependent variable broken down by the independent variable. I'd suggest requesting Q-Q plots, histograms, and tests of normality.

Answer

SPSS Syntax for the above:

* Descriptive statistics.

```
DATASET ACTIVATE DataSet1.
EXAMINE VARIABLES=readingSpan BY group
/PLOT BOXPLOT HISTOGRAM NPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```

It's worth taking a moment to look at these descriptive statistics.

The median span for noise and silence are 6 and 7 respectively, which suggests that reading span was higher under the silence condition—an expected finding. However, we need to be careful about concluding anything from the means because we need an inferential statistic to tell us whether we can trust the difference. Which test do we choose?

Because we have small, unequal cell sizes, a parametric test might not be appropriate. A look at the Q-Q plots is also worrying. For noise, the Q-Q plot shows a poor fit between observed values and values that we would predict on the basis of the normal distribution.

Similarly, while the reading span scores in the silence condition show a non-significant Shapiro-Wilk W statistic, this is not the case for noise: $W(5) = .68, p < .001$.

Taken together, this suggests that a more appropriate test is the Mann-Whitney.

Let's get ready to run the test. Look over the instructions below before you do so.

1. Navigate to *Analyse > Nonparametric > Legacy Dialogues > 2 Independent samples*

2. Set ‘readingSpan’ as your test variable, as you see in Figure 7.1.
3. Set ‘group’ as your grouping variable. Define its groups as ‘1’ and ‘2’.
4. Make sure that ‘Mann-Whitney U’ is selected.
5. Because we want the *exact* significance level, we will request it by clicking *Exact...* on this dialogue box and then selecting *Exact* (with a *Time limit per test* of five minutes, if you like). See Figure 7.2.

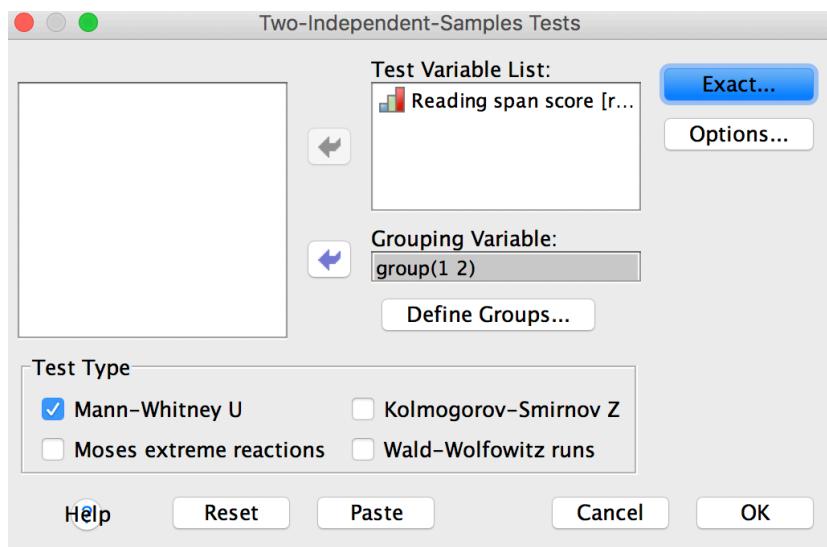


Figure 7.1: SPSS Two-Independent-Samples Tests

When reporting the significance (p) value for nonparametric tests, SPSS often provides different methods. These are:

1. The Asymptotic Method
2. The Exact Method
 - The first case—the *asymptotic* method—is most accurate with large samples and doesn’t use up too much computer processing time.
 - The second—*exact*—is more accurate when we have fewer data points or when the data are poorly distributed.
 - Field (2009) recommends going for the *exact* significance where

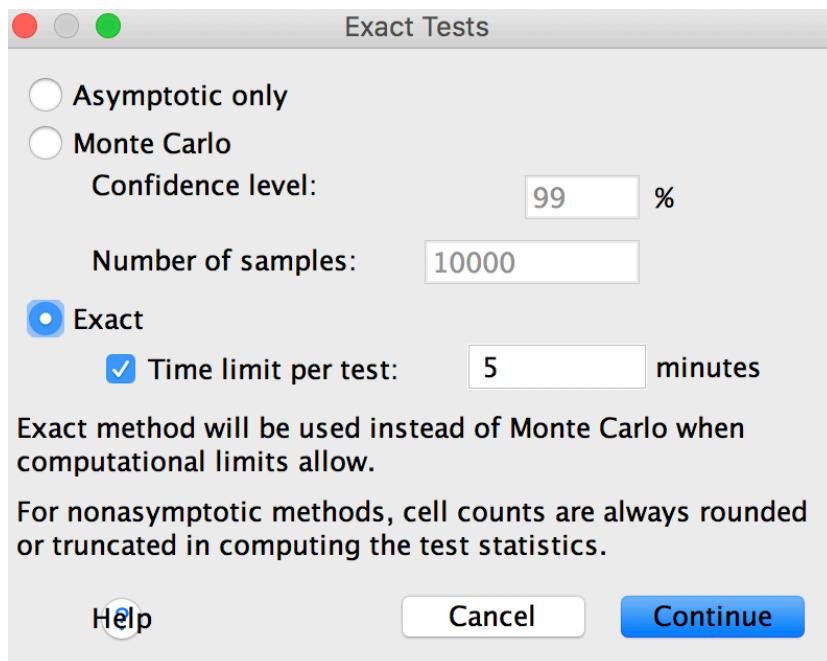


Figure 7.2: SPSS Two-Independent-Samples Tests, Exact Tests

possible, since we clearly would like precision(pp. 546-547). Because the *exact* method is labour-intensive for the computer, Field suggests a *Monte Carlo* option when the sample is large and a straightforward *exact* test when the sample is small. Most of our samples in psychology are small.

- Lastly, if don't see the *Exact Tests* table in your version of SPSS, it could be that you don't have the *Exact tests* module installed.

3. Run the Mann-Whitney U test.

Answer

SPSS Syntax:

NPAR TESTS

```
/M-W= readingSpan BY group(1 2)
```

```
/MISSING ANALYSIS
/METHOD=EXACT TIMER(5).
```

Your output should look like Figure 7.3. In the *Ranks* you will find the mean ranks of both groups (remember that the Mann-Whitney is comparing ranks, not means), and the sum of ranks. It's good practice to verify that the number of participants is correct.

The test statistic— U —is shown in the lower table of Figure 7.3. In this case, it's 6. Note that SPSS is also showing the Wilcoxon W (which can be converted to a z score).

→ NPar Tests

Figure 7.3: SPSS Output—Mann-Whitney

Mann-Whitney Test

Ranks

	Noise group	N	Mean Rank	Sum of Ranks
Reading span score	Noise	5	4.20	21.00
	Silence	7	8.14	57.00
	Total	12		

Test Statistics^a

	Reading span score
Mann-Whitney U	6.000
Wilcoxon W	21.000
Z	-1.944
Asymp. Sig. (2-tailed)	.052
Exact Sig. [2*(1-tailed Sig.)]	.073 ^b
Exact Sig. (2-tailed)	.052
Exact Sig. (1-tailed)	.032
Point Probability	.019

a. Grouping Variable: Noise group

b. Not corrected for ties.

As an aside, remember that a z score, or standard score, is a value on the standard normal distribution. It tells us how far a value is

above the mean. (The mean of the standard normal distribution is 0.) The higher the absolute value of a z score, the further from the mean, and the lower the chances are of finding it at random. For a one-tailed test, scores $+/- 1.64$ will occur less than 5% of the time (and the one-tailed significance for our z in Figure 7.3 is indeed $p = .032 < 0.05$). You can read more about z scores in Field (2009).

If we assume that the researcher had a preconceived idea about noise being detrimental to reading span performance, we should attend to the one-tailed exact significance. We might report the Mann-Whitney in the following manner (note the use of the median instead of the mean, as the Mann-Whitney is telling us about median differences):

Descriptive statistics showed that participants in the noise condition had lower reading span scores (median 6) than those in the silence condition (median 7). The Mann-Whitney U was found to be 6 ($z = -1.94$), $p < 0.05$.

7.5 Wilcoxon's Matched Pairs Signed-Ranks Test

Use this test when you wish to test for the differences of *two conditions of related* data. This is the nonparametric equivalent of the related samples t-test.

The assumptions of this test are:

- Data are paired
- The observations are *ordinal* (i.e. they can be put in a meaningful order; or, in other words, they can be ranked)

The null hypothesis (H_0) for this test is:

The overall difference between the two groups is zero

As usual, our p will tell us the probability of seeing our observed data under the assumption that H_0 is true.

- Our alternative hypothesis (H_1) is that there is a difference between the groups. Whereas a related t-test, being parametric, tells us about the likelihood of finding *mean* differences between two related groups, the Wilcoxon test tells us more about the *median* difference; in other words, these are both tests of central tendency but use different measures of central tendency. The Wilcoxon looks at the ranks of the differences rather than the absolute difference.

In our next example, we'll consider a design where the researcher wished to measure the susceptibility of reading span to training, so she measured reading span twice: once before training and once after.

4. Download and familiarise yourself with the  [data](#)
5. Obtain descriptive statistics for the dependent variable broken down by the independent variable.
1. What do you make of the normality in the data and the median differences?

Answer

- Before training median = 3, after training median = 6 (so perhaps in the direction we might expect)
- The histogram (suggesting a bimodal distribution) and Shapiro-Wilks ($W [12] = .84, p < 0.05$) for the after training group both suggest a lack of normality.

SPSS Syntax for the above:

```

DATASET ACTIVATE DataSet2.

EXAMINE VARIABLES=time1 time2
/PLOT BOXPLOT HISTOGRAM NPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.

```

We'll examine these group differences using the Wilcoxon test. Look over the instructions below before you carry them out.

1. Navigate to *Analyse > Nonparametric > Legacy Dialogues > 2 Related Samples*
2. Set 'time1' and 'time2' as your only test pair. Note that in Figure 7.4 the variables are shown by their longer labels.
3. Make sure that 'Wilcoxon' is selected.

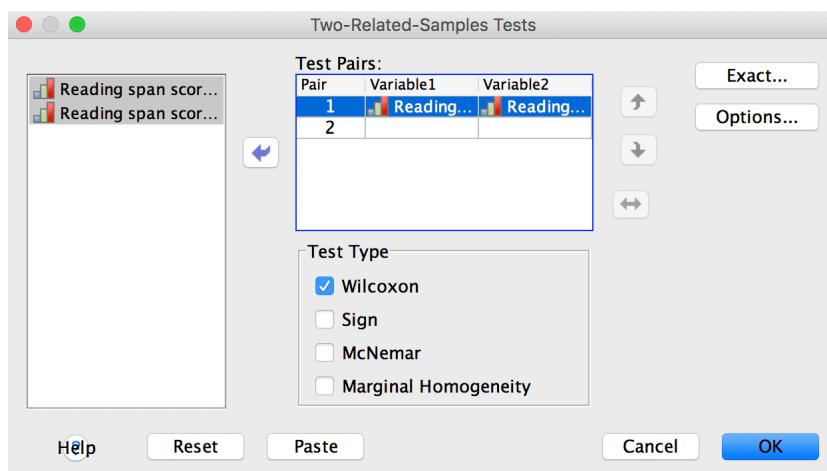


Figure 7.4: SPSS Two-Related-Samples Tests

4. As before, because we want the *exact* significance level, we will request it by clicking *Exact...* on this dialogue box and then selecting *Exact*. This is identical to the Mann-Whitney *Eaxct...* box shown in Figure 7.2.

6. Go ahead and run the Wilcoxon.

Answer

SPSS syntax for the above:

```
NPAR TESTS
/WILCOXON=time1 WITH time2 (PAIRED)
/MISSING ANALYSIS
/METHOD=EXACT TIMER(5).
```

Your output should look like Figure 7.5.

2. Assume the that the researcher was testing a one-tailed hypothesis.

Using the medians as your descriptive statistics, how might you report the findings? The key statistic for the Wilcoxon is z .

Answer

Descriptive statistics showed that participants improved their reading span scores after training, with a before training median of 3 and an after training median of 6. A Wilcoxon signed rank test indicated that this difference was reliable, $z = -2.59$, $p < 0.05$

7.6 The Kruskal-Wallis One-way Analysis of Variance

Use this test when you wish to test for the differences between *multiple conditions* of *independent* data. This is an extension of the non-parametric Mann-Whitney test from two groups to multiple groups. You can also think of it as a nonparametric version of the parametric one-way between-groups ANOVA.

Figure 7.5: SPSS Output—Wilcoxon

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
Reading span score after training –	Negative Ranks	1 ^a	4.00	4.00
Reading span score before training	Positive Ranks	10 ^b	6.20	62.00
	Ties	1 ^c		
	Total	12		

a. Reading span score after training < Reading span score before training

b. Reading span score after training > Reading span score before training

c. Reading span score after training = Reading span score before training

Test Statistics^a

	Reading span score after training – Reading span score before training
Z	-2.587 ^b
Asymp. Sig. (2-tailed)	.010
Exact Sig. (2-tailed)	.007
Exact Sig. (1-tailed)	.003
Point Probability	.001

a. Wilcoxon Signed Ranks Test

The assumptions of this test are:

- Data are independent
- The observations are *ordinal* (i.e. they can be put in a meaningful order; or, in other words, they can be ranked)

The null hypothesis (H_0) for this test is:

The overall difference between the groups is zero

As usual, our p will tell us the probability of seeing our observed data under the assumption that H_0 is true.

- Our alternative hypothesis (H_1) is that there is a difference between the groups. Whereas a one-way between-groups ANOVA, being parametric, tells us about the likelihood of finding *mean* differences between two independent groups, the Kruskal-Wallis test tells us more about the *median* difference; in other words, these are both tests of central tendency but use different measures of central tendency. The Kruskal-Wallis looks at the ranks of the differences rather than the absolute difference.

In this example, a researcher was interested in the effect of training on a measure called *reading fluency*, but rather than look at the effect within participants, she created three conditions—low, medium and high training—and placed different participants in each.

7. Download and familiarise yourself with the  data

3. Report and comment on the normality of the data and the median group differences.

Answer

- Medians for training group (low, medium and high) are: 12.5, 4, and 9.5
 - The Shapiro-Wilk statistics are nonsignificant for low and medium groups, and marginally nonsignificant for the high group: $W(4) = .76$, $p = .051$
 - The histograms for low and high training look decidedly non-normal; the Q-Q plots tell a similar story.
-

SPSS syntax for the above:

```
EXAMINE VARIABLES=readingFluency BY training
/PLOT BOXPLOT HISTOGRAM NPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```

We'll examine these group differences using the Kruskal-Wallis test. Look over the instructions below before you carry them out.

1. Navigate to *Analyse > Nonparametric > Legacy Dialogues > K Independent Samples*
2. Set the *Test variable list* as ‘readingFluency’ and *Grouping variable* as ‘training’, as you see in Figure 7.6. Because our training groups are numbered 3, 4, and 5 (these are arbitrary), use 3-5 after clicking *Define Range*.
3. Make sure ‘Kruskal-Wallis H’ is selected.
4. As before, request an *Exact p* value.

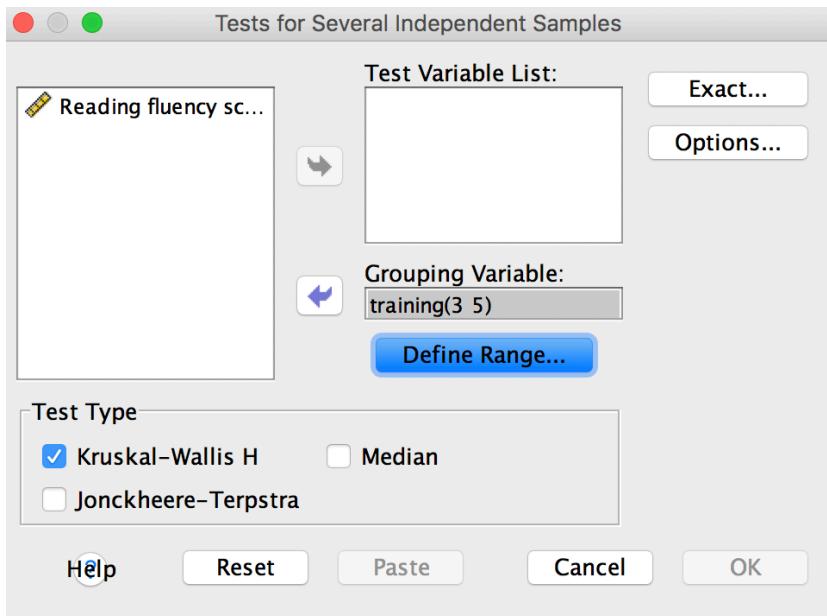


Figure 7.6: SPSS Tests for Several Independent Samples

8. Run the Kruskal-Wallis.

Answer

SPSS syntax for the above:

```
NPAR TESTS
/K-W=readingFluency BY training(3 5)
/MISSING ANALYSIS
/METHOD=EXACT TIMER(5).
```

Your output should look like Figure 7.7.

4. Using the medians as your descriptive statistics, how might you report the findings? The key statistic for the Kruskal-Wallis is the Chi-Square, written χ^2 .

Kruskal-Wallis Test

Figure 7.7: SPSS Output—Kruskal-Wallis

Ranks

	Training group	N	Mean Rank
Reading fluency score	low	4	13.50
	medium	7	4.21
	high	4	9.13
	Total	15	

Test Statistics^{a,b}

	Reading fluency score
Chi-Square	11.442
df	2
Asymp. Sig.	.003
Exact Sig.	.000
Point Probability	.000

a. Kruskal Wallis Test

b. Grouping Variable: Training group

Answer

Descriptive statistics showed that participants differed in their reading fluency according to level of training (with medians of 12.5, 4 and 9.5 for low, medium and high training conditions respectively). A Kruskal-Wallis one-way analysis of variance indicated that this difference was reliable, $\chi^2 (2) = 11.44, p < 0.001$.

5. Investigate the source of the differences between conditions using two Mann-Whitney U tests. Examine low v. high and medium v. high. Don't forget that these are post hoc tests and require a correction for familywise error. I'd suggest the Bonferroni correction.

We won't exhaust all the possible comparisons because we want to keep our Bonferroni correction to a minimum—remember that, as the number of post hoc comparisons increase, the corrected alpha level gets harder and harder to beat.

Answer

Our Bonferroni correction gives us a new alpha level of $(5\% / 2) = 2.5\%$. Note that we won't exhaust all the possible comparisons because we want to keep our Bonferroni correction to a minimum. Remember that, as comparison increase, the corrected alpha level gets harder and harder to beat.

- Curiously enough, those in the low training group (median 12.5) scored more than the medium and high training groups (medians of 4 and 9.5). Low-medium Mann-Whitney U was 0, $z = -2.67, p < 0.025$. Low-high Mann-Whitney was 0, $z = 2.32, p < 0.025$.
- (Note that a Mann-Whitney statistic of 0 occurs when all the values in one sample are higher than another.)

SPSS syntax for the above:

- * Mann-Whitney 'low' versus 'medium' (Group 3 v 4).
- * Asking for asymptotic and exact significance.

NPAR TESTS

```
/M-W= readingFluency BY training(3 4)
```

```
/MISSING ANALYSIS
```

```
/METHOD=EXACT TIMER(5).
```

- * Mann-Whitney 'low' versus 'high' (Group 3 v 5).

- * Asking for asymptotic and exact significance.

NPAR TESTS

```
/M-W= readingFluency BY training(3 5)
```

```
/MISSING ANALYSIS
```

```
/METHOD=EXACT TIMER(5).
```

7.7 The Friedman's Rank Test for k Correlated Samples

Use this test when you wish to test for the differences of *multiple conditions of related* data. This is an extension of the Wilcoxon test and represents the nonparametric equivalent of the one-way repeated-measures ANOVA.

The assumptions of this test are:

- Data are related
- The observations are *ordinal* (i.e. they can be put in a meaningful order; or, in other words, they can be ranked)

The null hypothesis (H_0) for this test is:

The overall difference between the groups is zero

As usual, our p will tell us the probability of seeing our observed data under the assumption that H_0 is true.

- Our alternative hypothesis (H_1) is that there is a difference between the groups. Whereas a one-way repeated-measures ANOVA, being parametric, tells us about the likelihood of finding *mean* differences between two related groups, the Friedman test tells us more about the *median* difference; in other words, these are both tests of central tendency but use different measures of central tendency. The Friedman looks at the ranks of the differences rather than the absolute difference.

In our final example, a researcher was interested in whether a visual stimulus could lead to changes in appetite. She showed participants a movie—‘The Hunger Games’—and asked them to rate their hunger on three occasions: at the beginning, middle and end of the film.

9. Download and familiarise yourself with the  data

6. Report and comment on the normality of the data and the median group differences.

Answer

- Medians for hunger score at the beginning, middle and end of the film are 1.5, 2 and 3 respectively.
- The Shapiro-Wilk statistics are significant for the beginning and end hunger scores, though not the middle. Beginning: $W(10) = .66, p < 0.001$; end: $W(10) = .76, p < 0.01$.
- The histograms and Q-Q plots for the beginning seem non-normal; they look reasonably normal for the middle; for the

end, they're also looking problematic.

SPSS syntax:

```
* Descriptive statistics (via Explore option).
EXAMINE VARIABLES=beginning middle end
/PLOT BOXPLOT HISTOGRAM NPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```

We'll examine these group differences using the Friedman test.

Look over the instructions below before you carry them out.

1. Navigate to *Analyse > Nonparametric > Legacy Dialogues > K Related Samples*
2. Set the *Test variables* as our beginning, middle and end hunger measures, as you see in Figure 7.8.
3. Again, request an *Exact p* value.

10. Run the Friedman.

Answer

SPSS syntax for the above:

```
NPAR TESTS
/FRIEDMAN=beginning middle end
/MISSING LISTWISE
/METHOD=EXACT TIMER(5).
```

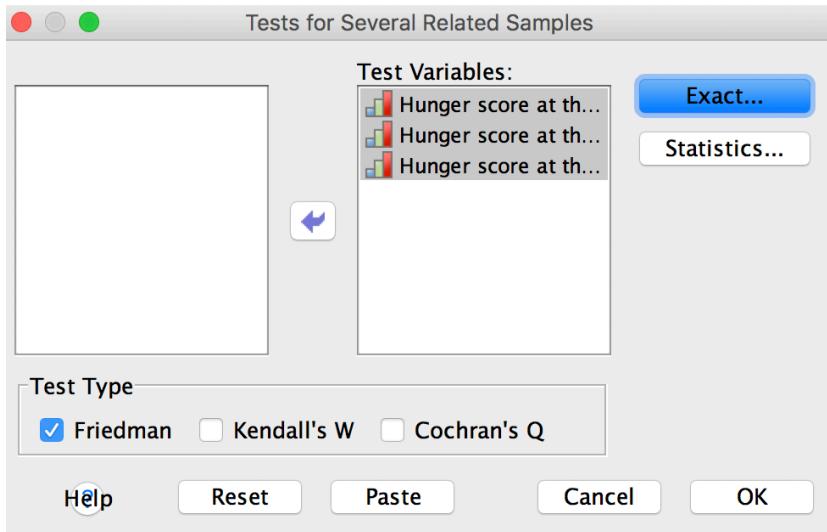


Figure 7.8: SPSS Tests for Several Related Samples

Your output should look like Figure 7.9.

7. Using the medians as your descriptive statistics, how might you report the findings? The key statistic for the Friedman is Chi-Square (χ^2).

Answer

Descriptive statistics showed that participants differed in their hunger scores throughout the film (with medians of 1.5, 2 and 3 for beginning, middle and end respectively). A Friedman one-way analysis of variance indicated a main effect of time, $\chi^2 (2) = 12.25, p < 0.05$.

8. Investigate the source of the differences between conditions using two Wilcoxon matched pair tests examining (i) the beginning and the end hunger scores, and (ii) beginning and middle hunger scores. Don't forget to correct for familywise error.

Friedman Test

Figure 7.9: SPSS Output—Friedman

Ranks

	Mean Rank
Hunger score at the beginning of the film	1.30
Hunger score at the middle of the film	2.00
Hunger score at the end of the film	2.70

Test Statistics^a

N	10
Chi-Square	12.250
df	2
Asymp. Sig.	.002
Exact Sig.	.001
Point Probability	.000

a. Friedman Test

Answer

- Our Bonferroni correction gives us a new alpha level of $(5\% / 2) = 2.5\%$.

Wilcoxon tests indicate a difference between the beginning and the end, but not between the beginning and the middle:

- For beginning vs. end: $z = -2.70, p < .025$
 - For beginning vs. middle: $z = -2.00, p > .025$
-

SPSS syntax for the above:

```
* Follow-up analyses (testing difference at
* beginning versus end)
* Wilcoxon.
```

NPAR TESTS

```
/WILCOXON=beginning WITH end (PAIRED)
/MISSING ANALYSIS
/METHOD=EXACT TIMER(5).
```

```
* Follow-up analyses (testing difference at
* beginning versus middle)
* Wilcoxon.
```

NPAR TESTS

```
/WILCOXON=beginning WITH middle (PAIRED)
/MISSING ANALYSIS
/METHOD=EXACT TIMER(5).
```


8

Correlation and Regression

8.1 Overview

Data files required  [Conscientiousness data](#)

Booklet Version 1.0

8.2 Objectives

This workshop will cover:

- transforming variables (reverse scoring variables that were reverse-phrased in their original forms);

- reliability (to ensure that multiple variables do indeed test the same construct);

- correlation;

- and regression.

8.3 Workshop

Answer

All SPSS data files and syntax:

[Go](#) [Conscientiousness data](#)

[Go](#) [Conscientiousness syntax](#)

In psychology, we often use a set of items (or statements) to measure some attitude or personality type. Participants are asked to respond to each item on a fixed scale, like a Likert scale, in a way that best reflects the way that they are. If you have used a scale to measure some attitude or personality type then it is essential that you know what those scores mean and that you find out how reliable they are.

We'll illustrate this by looking at conscientiousness (Costa & McCrae, 1985).

- (1) Download the [Go](#) [Conscientiousness data](#) and inspect it.

Each item has its own column and each respondent's answers are on one row. What we need is one score that represents how conscientious each respondent is. To do this we will need to re-score any reverse scored items, calculate reliability, and then add up all the individual item scores.

8.4 Rescoring Reverse-Phrased Items

Some items are reverse phrased. That is, a high score on these items would indicate less conscientiousness, while a high score on the other items would suggest more conscientiousness. The first thing that we must do then is to reverse the scoring for these reverse-phrased items. The reversed phrased items were items 1, 4, 5, 8, 9, 12, 13, 14 and 15.

- (2) Go to *Transform > Recode – Into different variables*, which should look like Figure 8.1 (note that I've asked SPSS to display variable names instead of labels).
- Select 'cons' items 1, 4, 5, 8, 9, 12, 13, 14, 15 from the left hand box and arrow them into the right hand box.

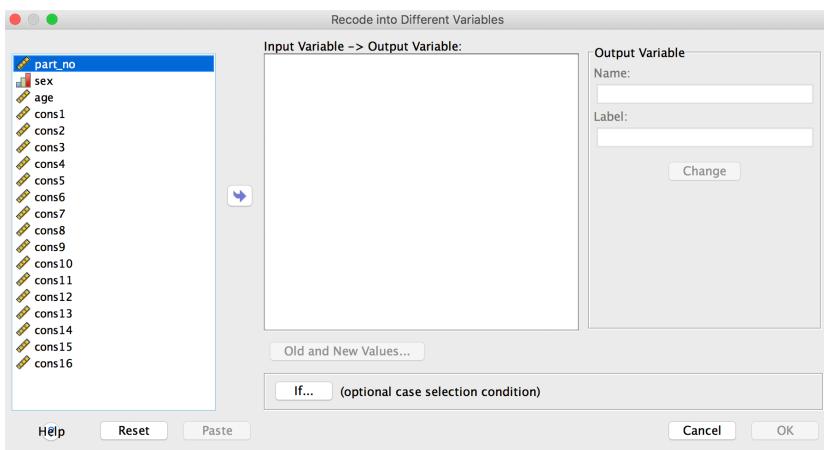


Figure 8.1: SPSS Recode into Different Variables, Incomplete

The recoding dialogue box will allow us to create a new variable for each of these items where the score for each individual is reversed. Now you need to tell SPSS what you will call these new variables.

- (3) With 'cons1' highlighted in the middle box type 'recons1' into the right hand box called *Output Variable*. In the *Label* box, type:
reversed cons1
- Now click on *Change* and this new name should replace the question mark in the middle box next to 'cons1'.
 - Do the same for the rest of the variables that need recoding so that it looks like Figure 8.2.

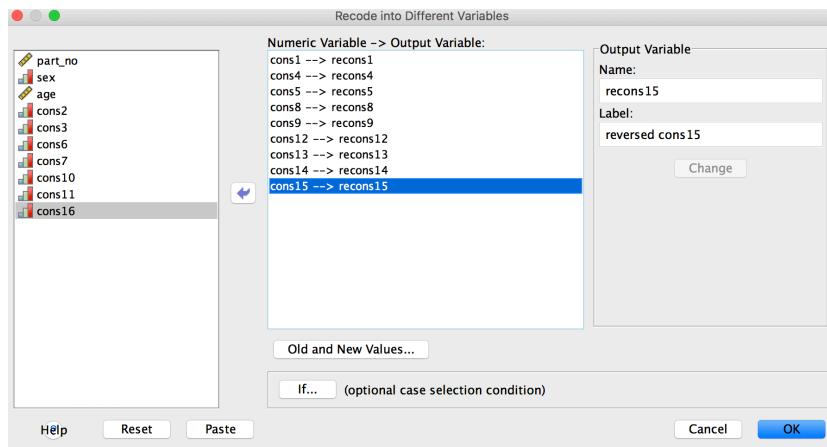


Figure 8.2: SPSS Recode into Different Variables, Complete

- (4) Next, click on *Old and New Values*. This will open a new dialogue box for you to tell SPSS what the old scores will be converted to. Tell SPSS that a score of 5 will become a score of 1, 4 will become 2, 3 will stay the same, 2 will become 4, and 1 will become 5.
- Remember, *all five* values must be reversed. This is shown in Figure 8.3. Remember to click on *Add* after you tell SPSS what each old value must be converted to. Now SPSS knows to reverse score each of these items for each participant.

- (5) Click on *Continue* and *OK*.

Answer

SPSS syntax for the above:

```

RECODE cons1 cons4 cons5 cons8 cons9 cons12 cons13
cons14 cons15 (1=5) (2=4) (3=3) (4=2) (5=1) INTO
recons1 recons4 recons5 recons8 recons9
recons12 recons13 recons14 recons15.

VARIABLE LABELS  recons1 'reversed cons1' /recons4

```

```
'reversed cons4' /recons5 'reversed cons5'
/recons8 'reversed cons8' /recons9 'reversed cons9'
/recons12 'reversed cons12' /recons13
'reversed cons13' /recons14 'reversed cons14'
/recons15 'reversed cons15'.

EXECUTE.
```

Look at the *data view* and verify that you have nine new columns at the end of the spreadsheet. All the values on these variables should be the reverse of the original score.

When we create a total score we will use these transformed variables (reversed negative items) and the original positive items so that high scores will mean higher conscientiousness.

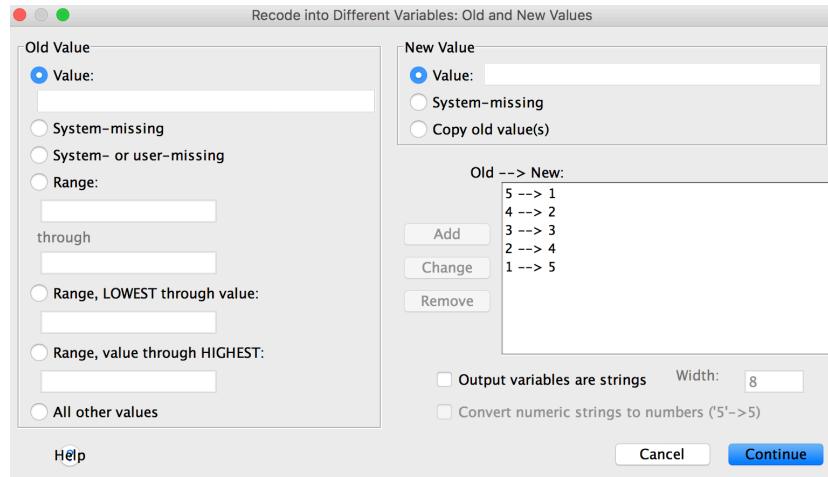


Figure 8.3: SPSS Recode into Different Variables: Old and New Values

8.5 Calculating Reliability

To look at how consistently all of the items measure conscientiousness, we need to calculate Cronbach's Alpha (written α), which is a measure of the degree of relationship between variables. Our rationale is this: If variables measure the same theoretical construct, we expect positive correlations among them, and this is basis for Cronbach's Alpha. The

more that items in a scale measure a construct, the greater reliability the scale has.

Table 8.2: Cronbach's Alpha with associated Internal Consistency

Cronbach's Alpha	Internal Consistency
$\alpha \geq 0.9$	Excellent
$0.9 > \alpha \geq 0.8$	Good
$0.8 > \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	Poor
$0.5 > \alpha$	Unacceptable

Let's start.

- (6) Go to *Analyze > Scale > Reliability analysis*. Put all the items that make up the conscientiousness scale (i.e. the recoded items recons1, recons4, recons5, recons8, recons9, recons12, recons13, recons14, recons15, plus the original items for the ones that were not reversed) into the right hand box, as you see in Figure 8.4. Make sure that where it says *Model, Alpha* is selected (this should be the default).
- Click on *Statistics* and make sure the options are selected as you see in Figure 8.5.

- (7) Click *continue* and *OK*.

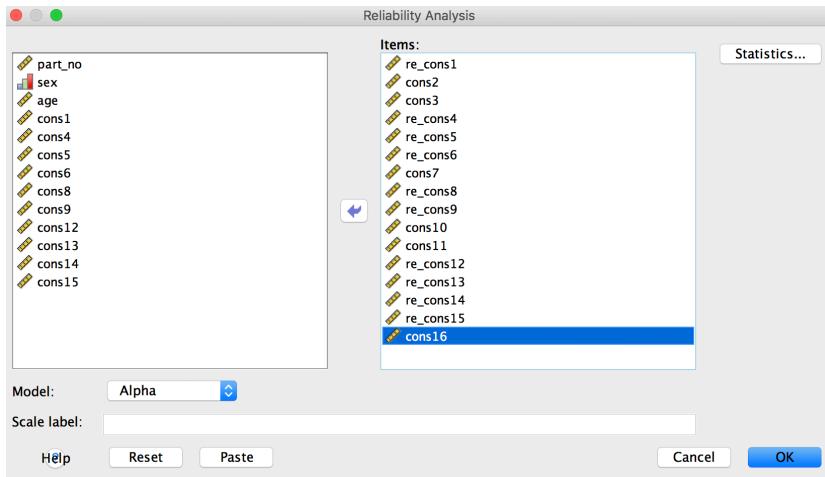


Figure 8.4: SPSS Reliability Analysis

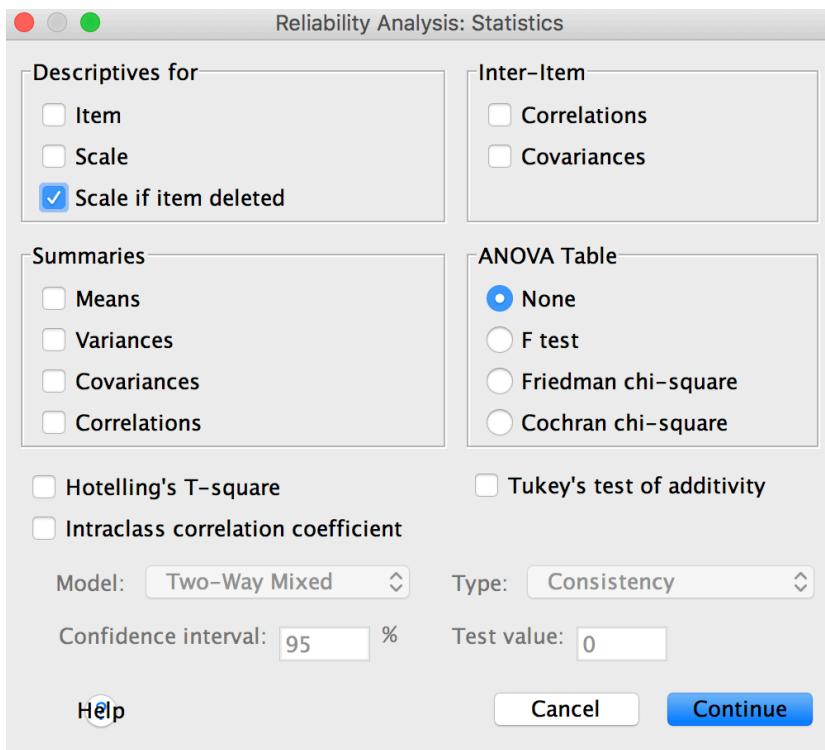


Figure 8.5: SPSS Reliability Analysis Statistics

Answer

SPSS syntax for the above:

```
* Reliability.

RELIABILITY

/VARIABLES=recons1 cons2 cons3 recons4
recons5 cons6 cons7 recons8 recons9 cons10 cons11
recons12 recons13 recons14 recons15 cons16
/SCALE('ALL VARIABLES') ALL
/MODEL=ALPHA
/SUMMARY=TOTAL.
```

- (8) What is the Cronbach's Alpha estimate for this scale? Is the scale reliable according to Table 8.2?

Answer

Cronbach's Alpha is .846 (to be reported in APA style as .85).

Yes, it is reliable because it's above .70.

- (9) Are there any items whose elimination from the scale would make it more reliable?

Answer

Eliminating item 1 (reversed) would bring it up slightly to .851 but the minimal increase in alpha doesn't make it worthwhile.
(‘Reversed cons13’ would improve it fractionally also to .847.)

8.6 Creating a Total Score.

Now let's create a total score on the scale for each individual.

(10) Go to *Transform > Compute*. In *target variable* type the name of the new variable that you are going to create.

- As we are creating a variable that represents people's total score, call it 'total'.
- In the *Numeric Expression* box we need to tell SPSS how to calculate the total. So arrow across 'recons1' (as opposed to the original 'cons1', as it was reversed scored) first of all, then click '+'; then arrow over 'cons2', then click '+'. Keep doing this until all the items forming the final scale are in, but remember to use the reversed items, not the original, for all the items we recoded earlier. See Figure 8.6.

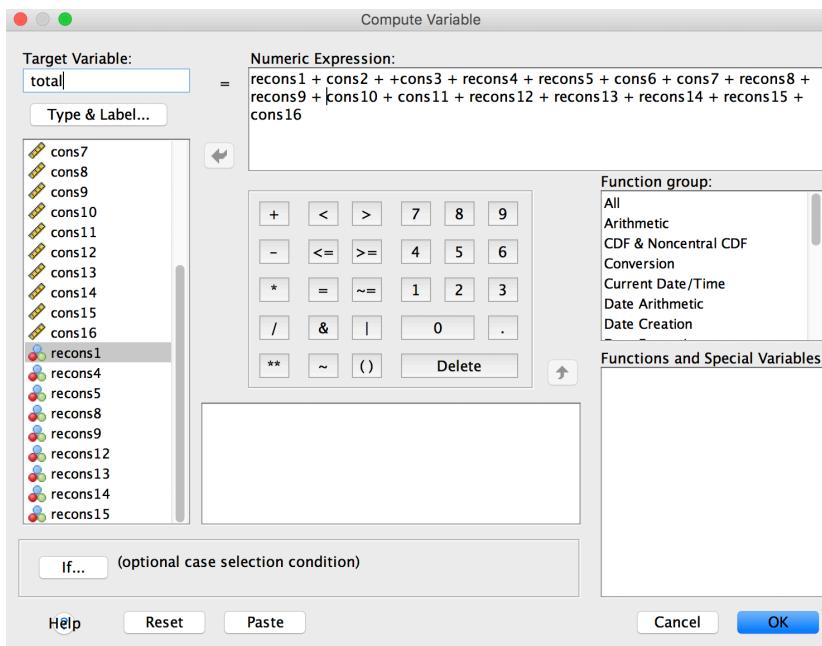


Figure 8.6: SPSS Compute Variable

Answer

SPSS syntax for the above:

```
COMPUTE total=recons1 + cons2 + +cons3 + recons4 +
recons5 + cons6 + cons7 + recons8 + recons9 +
cons10 + cons11 + recons12 + recons13 + recons14
+ recons15 + cons16.

EXECUTE.
```

Check that the ‘total’ mean is 53.87.

(11) Who is the most conscientious participant?

Answer

204 (total = 77)

8.7 Calculating Pearson’s Correlation.

Now let’s see if conscientiousness is related to a person’s age.

(12) Go into *Analyze > Correlate* and select *Bivariate* from the menu.

- Move the two variables you want to correlate across to the right-hand window. In this case we select ‘Age’ and our total conscientiousness score ‘Total’. As you see in Figure 8.7, make sure that *Pearson* is ticked (this should be the default) and, because we aren’t sure about the direction of our prediction, stick to a two-tailed test.
- Click *OK*.

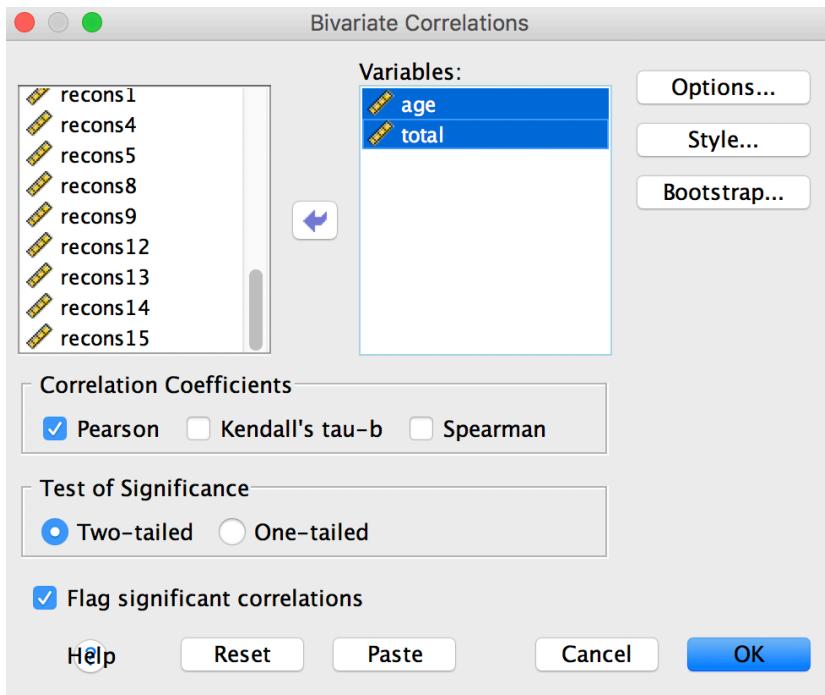


Figure 8.7: SPSS Bivariate Correlations

(13) Was age related to conscientiousness? Report Pearson's correlation between the two variables.

(14) What proportion of the variance in conscientiousness was accounted for by its relationship with age? (Look back at last week's booklet if you need a reminder of how to calculate shared variance.)

Answer

- Age was not related to conscientiousness, $r(239) = -.04$, $p > .05$.
- The proportion of variance in conscientiousness accounted for by its relationship with age was .0016, or 0.16%. You obtain this value by taking the square of r (-.04).

8.8 Regression

Let's say that we're interested in predicting the relationship between two variables, so that if we collect further data from more participants, we can predict the value of one variable given the value of another for each new participant. In the following example, we'll look at *one* of the conscientiousness measures, 'cons13', which captures a rating of the statement 'I'm not conscientious about cleaning'. Because this is a reverse-phrased item, we'll need to use its reverse-scored version ('recons13') so that positive values of the variable represents a positive value of the construct—which we can call 'cleaning conscientiousness'.

Imagine that the researcher believes 'cleaning conscientiousness' to be related to age. They wish to test this relationship with a regression so that they can use the age of participants, in another sample, to predict their 'cleaning conscientiousness'.

(15) Go to *Analyze > Regression > Linear...* and, in the dialogue box that appears (Figure 8.8), enter the variables 'age' and 'recons13'.

- Remember that thing you are trying to predict should be the *dependent* variable and the thing you are trying to predict *from* should be an *independent* variable. (This is not quite how the terms dependent and independent are used in experimental designs.)

(16) Click *Statistics* and ensure that *Estimates, Confidence Intervals, Model Fit* and *Descriptives* are selected (Figure 8.9).

(17) It's usually a good idea to visually inspect any relationship we find. To do this, click *Plots* and have it resemble Figure 8.10.

- Note that what we predict goes, by convention, on the Y-axis. Our predictor will go on the X-axis (here, this is the Z or standardised

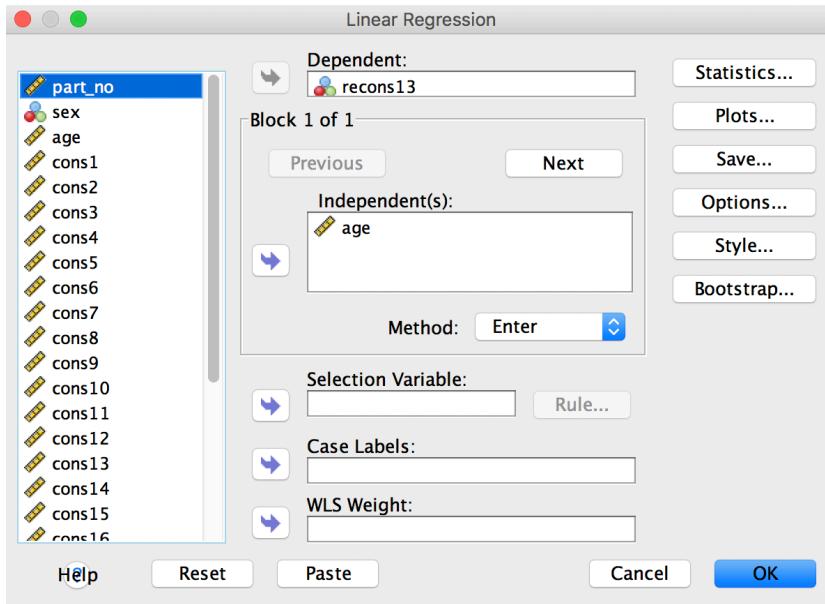


Figure 8.8: SPSS Linear Regression

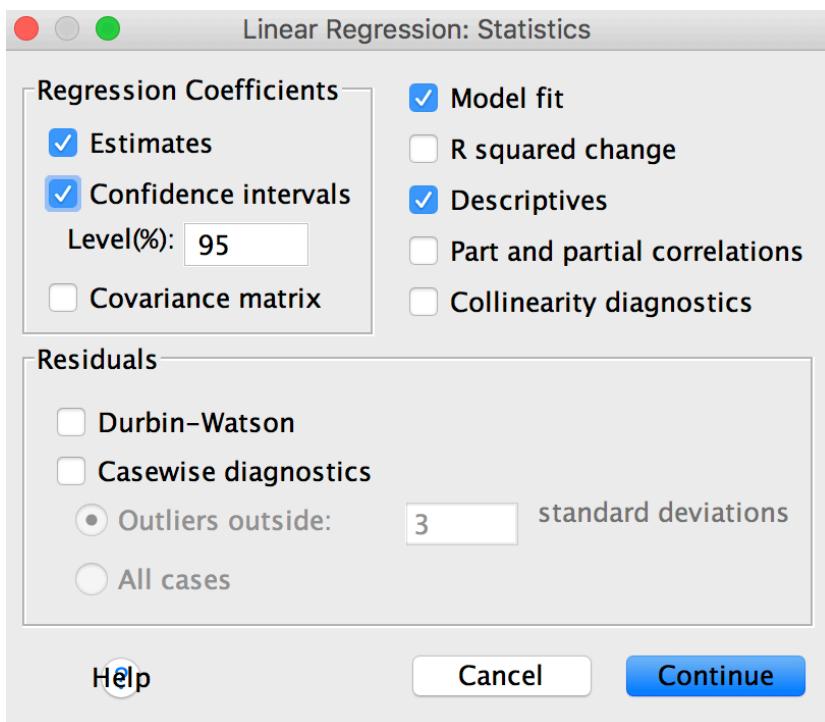


Figure 8.9: SPSS Linear Regression, Statistics

score of the predictor).

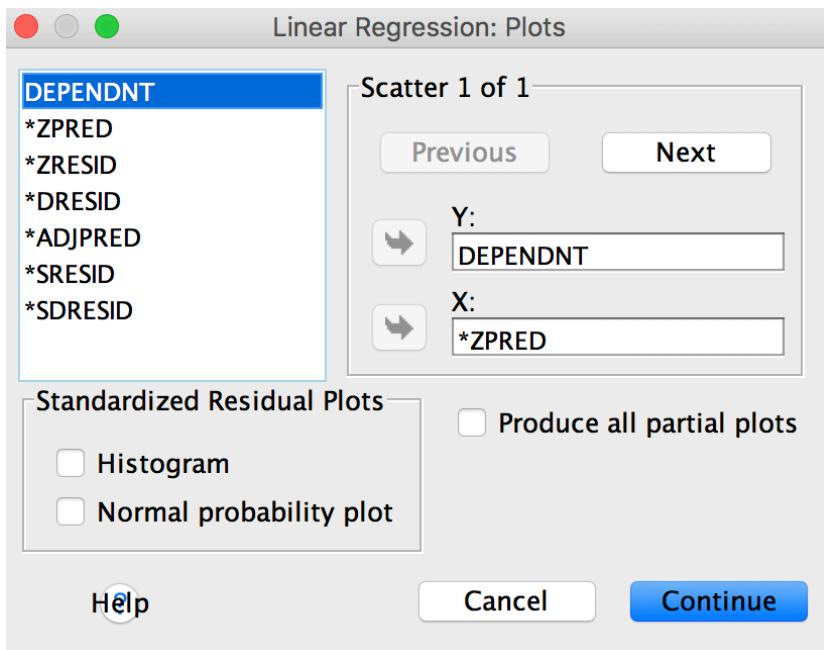


Figure 8.10: SPSS Linear Regression, Plots

- (18) Return to the main regression dialogue box and click *OK* to run the regression.

Answer

SPSS syntax for the above:

```
REGRESSION
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT recons13
```

```
/METHOD=ENTER age
/SCATTERPLOT=(recons13 ,*ZPRED) .
```

Let's see what SPSS has produced. The Model Summary (Table 8.11) provides us with the *Model* (we can run lots of these; here there is only one), R , R^2 , R_{adj}^2 , as well the standard error of the estimate.

The *standard error of the estimate* tells us how well the model fits the data in terms of the mean value of how far observations fall from the regression line, expressed as units of the dependent measure. In Table 8.11, this is 1.37, which means that, on average, cleaning conscientiousness scores are 1.37 conscientiousness-scale units away from the best-fitting regression line.

Model Summary^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.147 ^a	.021	.017	1.37372

a. Predictors: (Constant), age

b. Dependent Variable: reversed cons13

Figure 8.11: SPSS Output—Model Summary

- (19) Looking at the Model Summary Table 8.11, and using what you know from the lecture, what is the strength and direction of the regression?
- What is the shared variance between 'age' and 'cleaning conscientiousness'?

Answer

- Because this is simple regression (the case being more complex in multiple regression), R represents the Pearson product moment correlation—i.e. the standard parametric correlation coefficient—and is .15. This is a weak correlation.
- As ever, when we square the correlation coefficient, we see the shared variance between the two variables. $R_{adj}^2 = .02$, $R^2 = .02$. This means that ‘age’ is predicting only 2% of ‘cleaning conscientiousness’—not a lot.

Now let’s turn to the ANOVA (Table 8.12). Remember that the F ratio represents the mean squares of the model (which is the improvement of the regression line over a simpler model of the mean of Y) divided by the residual mean squares (which is how poorly the model fits the data). If our model—i.e. our regression line—is no great improvement over a simpler model of the mean of Y, and/or our model is a poor fit for the observations, F will be low. So what we’re testing with this ANOVA is a sense of the regression line’s explanatory power. The null hypothesis is that the regression line has no explanatory power, and the line has to have good explanatory power before we accept the model as reliable.

ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	9.813	1	9.813	5.200	.023 ^b
Residual	447.242	237	1.887		
Total	457.054	238			

a. Dependent Variable: reversed cons13

b. Predictors: (Constant), age

Figure 8.12: SPSS Output—ANOVA Table

- (20) Looking at the ANOVA, is the model significant overall? Report the F ratio in APA style.

(21) In plain English, what is this table telling you?

Answer

- $F(1, 237) = 5.20$, $MSE=1.89$ $p < .05$.
- It tells us that the model (our sloped line) fits the relationship well, and it would be more sensible to trust its predictive ability.

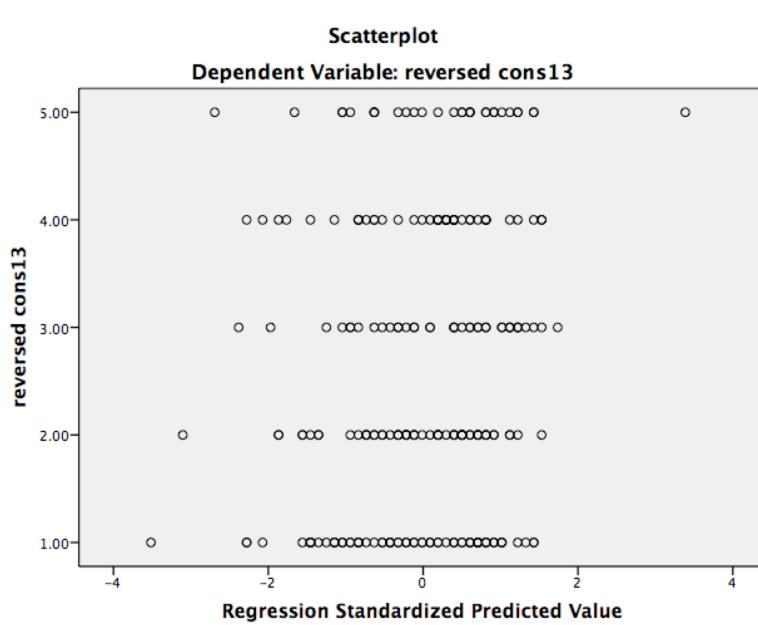


Figure 8.13: SPSS Output—
Scatterplot

(22) Does the scatterplot make you cautious about interpreting the regression line?

Answer

- With the bottom left and top right cases removed, we might see the regression lose its fit. This means the regression model is somewhat fragile.

Table 8.14 shows the Table of Coefficients. From this table, we can construct our raw and standardised regression equations.

Model	Coefficients ^a					
	B	Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B
						Lower Bound
1	(Constant) 3.390	.341		9.948	.000	2.719
	age -.021	.009	-.147	-2.280	.023	4.062 -.039 -.003

a. Dependent Variable: reversed cons13

Figure 8.14: SPSS Output—Coefficients

- (23) What is the unstandardised coefficient, its confidence intervals, and its p value?
- (24) What comments do you have on this coefficient?

Answer

- B is $-.02$, 95% CI $[-.04, .0]$, and this is significant: $t = 9.95$, $p < 0.05$.
- The coefficient is very small, as we saw earlier from an inspection of the regression relationship expressed as R and R^2 . The Confidence Intervals show that, if we were to sample again, the coefficient would vary by up to ten times, and is quite close to zero—not a particularly consistent effect. So while the t statistic is significant, ‘age’ is not a particularly convincing predictor in this model.

(25) Report the unstandardised and standardised regression models.

Answer

- Unstandardised:

$$(CleaningConscientiousness_{pred}) = 3.390 - .021(Age)$$

- Standardised:

$$(zCleaningConscientiousness_{pred}) = -.147(zAge)$$

(26) Report the full regression results as you would in an APA Results section.

Answer

- A simple linear regression was calculated to predict a participant's 'cleaning conscientiousness' based on their age. A significant regression equation was found, $F(1,237) = 5.20$, $MSE = 1.89$, $p < .05$., with an R^2 of .02. Participants' predicted 'cleaning conscientiousness' is equal to $3.390 - .021(Age)$ when age is measured in years (95% CI for B [-.04, .0]). Participants' mean 'cleaning conscientiousness' decreased by .021 for each year of age.

(27) Use the *unstandardised* regression model to predict cleaning conscientiousness for two new participants, one of whom is aged 24 and the other 78.

Answer

1. Age = 24

- $CC = 3.39 - (.021 * 24)$
 - $= 2.87$
-

2. Age = 78

- $CC = 3.39 - (.021 * 78)$
 - $= 1.75$
-

- Thus our model, should we believe it, is telling us a precise value for cleaning conscientiousness given age.

9

Multiple Regression (1)

9.1 Overview

Data files required [g Health](#)

Booklet Version 1.0

9.2 Objectives

This workshop booklet will take you through a single example of multiple regression. You will learn how to set up and run the regression, interpret each aspect of the output, and make an overall summary of the model.

9.3 Workshop

Answer

All SPSS data files and syntax:

[g Health data](#)

[g Health syntax](#)

In this workshop, we're going to run two multiple regression analyses on the same set of data, which contain health information about selected American cities. In the file, each city is a row. The variables

provide information about certain health statistics, which you can see in Table 9.2.

Table 9.2: Measures and descriptions for the American city health data

<i>Measure</i>	<i>Description</i>
deathRate	Death rate per 10000 residents
doctorAvailability	Doctor availability per 100,000 residents
hospitalAvailability	Hospital availability per 100,000 residents
annualIncome	Annual per capita income in thousands of dollars
populationDensity	Population density people per square mile

The scenario:

Imagine that you are a US government researcher tasked with predicting the availability of doctors based on the other health data you have (i.e. the other variables in Table 9.2). You're interested in developing a model that can be applied to other cities where the doctor availability is unknown, but where the other variables are known.

We'll now run a multiple regression in which we try to predict the availability of a doctor from the other variables. Thus, the *outcome* variable (or dependent variable) will be availability of a doctor. The *predictor* variables (or independent variables) will be the remaining variables.

Read through the instructions below before carrying out your multiple regression.

On this occasion, we will not examine whether the data fit the assumptions of multiple regression. This should keep things simpler for your first attempt. Assumptions, however, are very important, and we'll be looking at them again in the next booklet.

9.3.1 Step One: Set up the Linear Regression

- (28) Download [3 Health data](#) and familiarise yourself with the data by running *Descriptives*.

Answer

We have a range of variables here with different characteristics.

Per capita income seems quite low (range 7.2k to 13k), but bear in mind these data are historical.

SPSS syntax for the above:

```
DESCRIPTIVES VARIABLES=deathRate doctorAvailability
hospitalAvailability annualIncome
populationDensity
/STATISTICS=MEAN STDDEV MIN MAX.
```

- (29) Go to *Analyze > Regression > Linear...* and enter ‘doctorAvailability’ as your dependent variable (i.e. the one you wish to predict) and all the other variables into the *Independent(s)* box, as you see in Figure 9.1.

- (30) Select the correct statistics. We want *Estimates*, *Confidence intervals*, the *Model fit*, *Descriptives*, and *Part and partial correlations*. See Figure 9.2.

In this simple case, that’s as much as we need to tell SPSS.

- (31) Go ahead and run the multiple regression using the steps above.

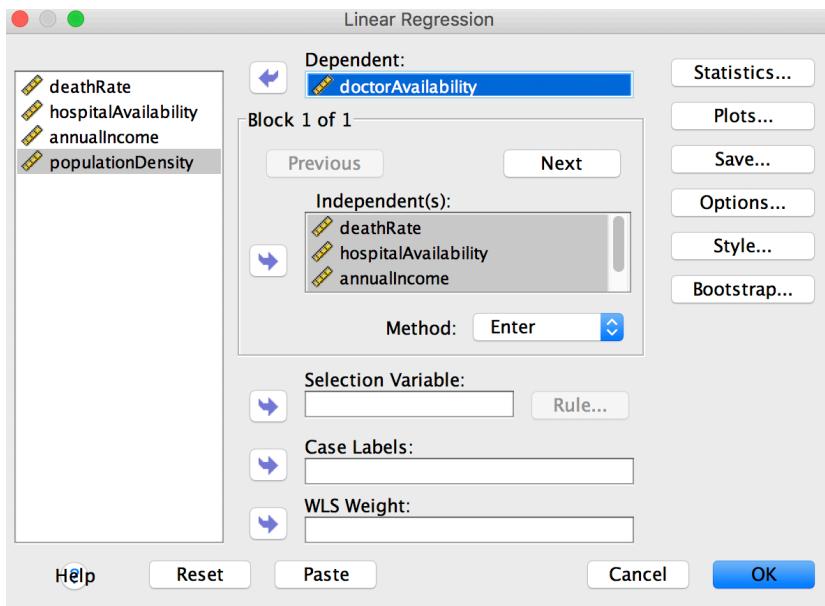


Figure 9.1: SPSS Linear Regression Dialogue

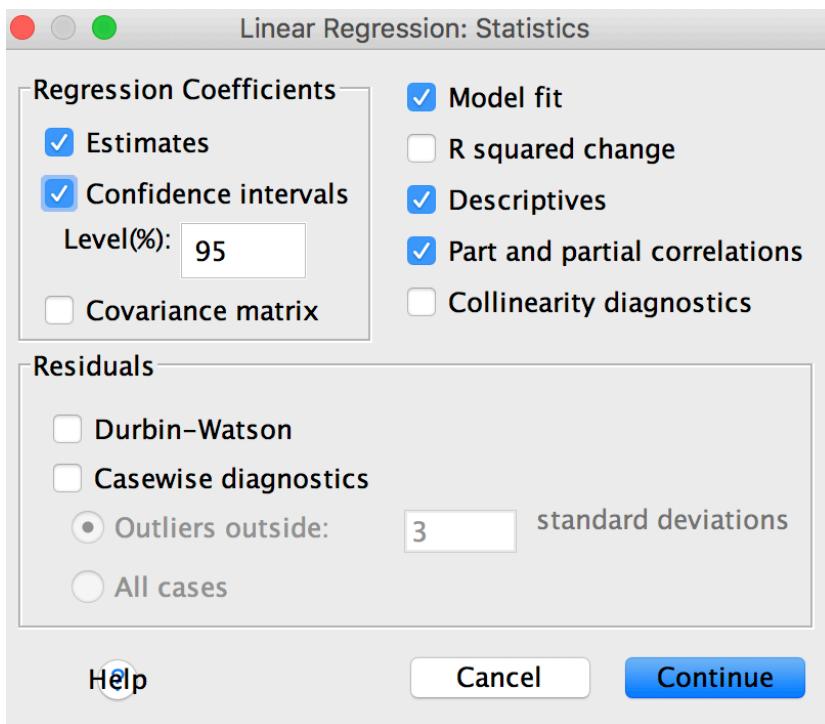


Figure 9.2: SPSS Linear Regression Statistics

Answer

SPSS syntax:

```
* Regression.

REGRESSION

/DESCRIPTIVES MEAN STDDEV CORR SIG N

/MISSING LISTWISE

/STATISTICS COEFF OUTS CI(95) R ANOVA ZPP

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT doctorAvailability

/METHOD=ENTER deathRate hospitalAvailability
annualIncome populationDensity.
```

9.3.2 Step 2: Interpret the Output

SPSS should now have produced:

- *Descriptive statistics*, which should be the same as your original descriptive statistics. Use this table to check that your data have been entered into the regression correctly.
- *Correlations*, which show you the correlation¹ between the variables. As we shall see, correlations among your predictor variables can be problematic for the regression model.
- *Variables Entered/Removed*, which tells you the variables that have been used. It is possible to remove variables. If you've requested this, it will be reported here.
- *Model Summary*, which tells about the fit of the regression model.
- *ANOVA*, which gives you a sense of whether your model is better than a random one.
- *Coefficients*, from which you can build a complete picture of your regression equation.

¹ Pearson product moment correlation.

Let's go through each of these. We'll skip *Descriptives*.

- (32) What do you notice about the *Correlations* matrix, excerpted in Figure 9.3?

- Don't be confused by the table. The row labelled *Pearson Correlation* shows the correlations. The second row, labelled *Sig. (1-tailed)*, shows the *p* value of each corresponding *r* from the row above.

Correlations						
	Doctor availability per 100,000 residents	Death rate per 10000 residents	Hospital availability per 100,000 residents	Annual per capita income in thousands of dollars	Population density people per square mile	
Pearson Correlation	Doctor availability per 100,000 residents	.1000	.116	.296	.433	-.020
	Death rate per 10000 residents	.116	1.000	.111	-.172	-.278
	Hospital availability per 100,000 residents	.296	.111	1.000	.028	.187
	Annual per capita income in thousands of dollars	.433	-.172	.028	1.000	.129
	Population density people per square mile	-.020	-.278	.187	.129	1.000

Figure 9.3: SPSS Output—Correlations, excerpt

Answer

Many of the correlations are not significant, but the availability of doctors is correlated with annual income ($r = .43, p < .05$) and with hospital availability ($r = .27, p < .05$).

Look at the *Model Summary*, Figure 9.4.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.550 ^a	.302	.244	32.94571

- a. Predictors: (Constant), Population density people per square mile , Annual per capita income in thousands of dollars, Hospital availability per 100,000 residents, Death rate per 10000 residents

Figure 9.4: SPSS Output—Model Summary

- (33) Report *Multiple R*, R^2 , and R_{adj}^2 .
- (34) Include a plain English description of what each measure is telling you.

Answer

- $R = .55$. This shows that the overall relationship between the dependent variable (doctor availability) and the independent variables is moderate.
- $R^2 = .30$. The independent variables account for 30% of the variability in doctor availability data (unadjusted for the number of independent variables—that is, predictors—and based solely on our sample characteristics).
- $R_{adj}^2 = .24$. The independent variables account for 30% of the variability in doctor availability data (*adjusted* for the number of independent variables—that is, predictors—and based on an estimate of the population characteristics).

Now look at the *ANOVA* table, Figure 9.5.

ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	22540.385	4	5635.096	5.192	.001 ^b
Residual	52100.143	48	1085.420		
Total	74640.528	52			

a. Dependent Variable: Doctor availability per 100,000 residents

b. Predictors: (Constant), Population density people per square mile , Annual per capita income in thousands of dollars, Hospital availability per 100,000 residents, Death rate per 10000 residents

Figure 9.5: SPSS Output—ANOVA

- (35) Report F in the standard way.
- (36) In plain English, what is this F ratio telling you?

Answer

- $F(4, 48) = 5.19$, $MSE = 1085.42$, $p < .05$.
- This tells us that our regression model (a multidimensional plane, in this case, because each new variable adds a new dimension) fits the variability in the dependent variable (doctor availability) better than a ‘null’ model where we use the mean of the dependent variable. On this basis, because the F ratio is significant, we can consider our overall model to be reliable.

We'll now look at the table of *Coefficients*, Figure 9.6. Because its text might be too small, I've broken it down into two figures, Figure 9.7 and Figure 9.8. These show:

- *Unstandardized coefficients*, which tell us the strength and direction of the relationship of each predictor with the dependent variable.
- *Standardised coefficients*, which tell us the strength and direction of each predictor-dependent variable relationship *on a standard scale* (the Z distribution). Furthermore, because these values are standardised, they can be meaningfully compared (whereas the unstandardized coefficients cannot, because they are expressed in units of the dependent variable).
- *t (with associated significance)*, which tells us whether slope of the unstandardised coefficient (above) is different from 0. If it's not significant, the predictor is no better than chance at predicting variability in the dependent variable.
- *95% confidence interval for B*, which gives us the lower and upper bounds of a range within which each unstandardised coefficient should fall on 95% of occasions if were to keep sampling forever.
- *Correlations*, which tell us (1) zero order, (2) partial and (3) part correlations. Remember that (1) is a simple correlation between the dependent variable and a given independent variable; (2) is the

correlation between the DV and an IV when the variance explained by all other IVs has been removed from both the IV and DV; (3) is (in shared variance terms) the amount of variance in the DV uniquely explained by the IV.

Model	Coefficients ^a								
	B	Unstandardized Coefficients	Std. Error	Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	Correlations	
							Lower Bound	Upper Bound	Zero-order
1	(Constant)	-77.118	54.135		-1.425	.161	-185.964	31.728	
	Death rate per 10000 residents	3.129	2.935	.137	1.066	.292	-2.773	9.031	.116 .152 .129
	Hospital availability per 100,000 residents	.033	.014	.285	2.289	.026	.004	.061	.296 .314 .276
	Annual per capita income in thousands of dollars	16.248	4.330	.461	3.752	.000	7.542	24.955	.433 .476 .452
	Population density people per square mile	-.076	.104	-.094	-.731	.469	-.285	.133	-.020 -.105 -.088

a. Dependent Variable: Doctor availability per 100,000 residents

Figure 9.6: SPSS Output—Coefficients Overall

Model	Unstandardized Coefficients			Standardized Coefficients
	B	Std. Error	Beta	
1	(Constant)	-77.118	54.135	
	Death rate per 10000 residents	3.129	2.935	.137
	Hospital availability per 100,000 residents	.033	.014	.285
	Annual per capita income in thousands of dollars	16.248	4.330	.461
	Population density people per square mile	-.076	.104	-.094

Figure 9.7: SPSS Output—Coefficients, Bs and Betas

(37) Which of the predictors are reliable?

Answer

Only hospital availability ($t = 2.29, p < .05$) and annual income ($t = 3.75, p < .05$) are significant predictors.

t	Sig.	95.0% Confidence Interval for B		Correlations		
		Lower Bound	Upper Bound	Zero-order	Partial	Part
-1.425	.161	-185.964	31.728			
1.066	.292	-2.773	9.031	.116	.152	.129
2.289	.026	.004	.061	.296	.314	.276
3.752	.000	7.542	24.955	.433	.476	.452
-.731	.469	-.285	.133	-.020	-.105	-.088

Figure 9.8: SPSS Output—Coefficients, Ts, CIs and Correlations

(38) State the unstandardised regression equation.

Answer

$$\begin{aligned} \text{doctor availability} &= -77.12 + 3.13(\text{death rate}) + \\ &0.03(\text{hospital availability}) + 16.25(\text{annual income}) - \\ &0.08(\text{population}) \end{aligned}$$

(39) State the standardised regression equation.

Answer

$$\begin{aligned} Z\text{doctor availability} &= 0.14(Z\text{death rate}) + \\ &0.29(Z\text{hospital availability}) + 0.46(Z\text{annual income}) - \\ &0.09(Z\text{population}) \end{aligned}$$

(40) What do you notice about difference in ranked standardised versus unstandardised coefficients?

Answer

- In terms of unstandardised coefficients, the order (greatest first) is annual income, death rate, hospital availability and population density. When these are standardised, the order changes to annual income, hospital availability, death rate, and population density.
- So death rate and hospital availability have swapped. While death rate might appear to have a stronger relationship with doctor availability than hospital availability (from the perspective of the unstandardised coefficients, which are not fully comparable), the standardised model suggests that hospital availability has the stronger relationship.

(41) Which variable uniquely explains the most variability in doctor availability, and why?

Answer

Annual per capita income. Because its *part* correlation is .452. We can square this value to get a measure of unique variability, which is 20.43%.

(42) Summarise the findings.

Answer

The association between the dependent and predictor variables was moderate (Multiple R = .55). Overall, the model was significant, $F(4,48) = 5.19$, $MSE = 1085.42$, $p < .05$. Together, death rate, hospital availability, annual income and population density accounted for 30.2% of variation in doctor availability

(adjusted R-square = 24.4%). Only hospital availability ($t = 2.29, p < .05$) and annual income ($t = 3.75, p < .05$) were significant predictors. The unstandardised regression coefficient for hospital availability was .03 (95% CI [.00, .06]), while the unstandardised coefficient for annual income was 16.25 (95% CI [7.54, 24.96]). The standardised coefficients were .29 and .46 respectively. Annual income explained more unique variance in doctor availability (20.43%) than did hospital availability (7.61%).

10

Multiple Regression (2)

10.1 Overview

Data files required  **Worry**

Booklet Version 1.0

10.2 Objectives

This workshop booklet will take you through a second multiple regression, including the testing of assumptions.

10.3 Workshop

Answer

All SPSS data files and syntax:

 **Worry data**

 **Worry syntax**

Here is the scenario:

The researcher is interested in finding out the extent to which worry and conscientiousness predict a person's time management behaviour, the theory being that people who are highly conscientious and worry more will be more likely to engage in time-management behaviours.

This datafile comes from 100 respondents who filled out a questionnaire that contained measures of Worry (16 items), Conscientiousness (16 items), and Time Management Behaviour (34 items). Notice that there are columns to represent where an item has been reverse-scored and that at the very end of the datafile you have three ‘total’ columns that have each respondents’ total scores in them.

10.3.1 Step One: Data overview

- (43) Download the  [Worry data](#) and familiarise yourself with it by running *Descriptives*.

Answer

There’s much we could look at here. In terms of the three total variables, we have conscientiousness (mean = 46.16, standard deviation = 10.00), time management behaviour (mean = 103.79, standard deviation = 18.44) and worry (mean = 50.20, 11.89).

SPSS syntax for the above:

```
DESCRIPTIVES VARIABLES=consTotal worryTotal tmbTotal
/STATISTICS=MEAN STDDEV MIN MAX.
```

10.3.2 Step Two: Choose your method for entering predictors into the model

For this project you will use *Forced Entry*, in which all the predictors are entered at the same time. This method tells us how much variance in the outcome can be explained by the whole model (all predictors together) and how much variance in the outcome can be *uniquely* explained by each individual predictor. The alternative would be to

enter, or remove, them one at a time based on some kind of criterion (Field, 2009).

10.3.3 Step Three: Set up the multiple regression

Before you go ahead and run the regression, read through the next steps so that you can get a sense of them.

(44) Go to *Analyze > Regression > Linear* and enter the variable you wish to predict as the *Dependent*, and the variables you are using to predict it as the *Independents*. Your dialogue box should look like Figure 10.1.

- Make sure that the *Method:* box reads *Enter*.

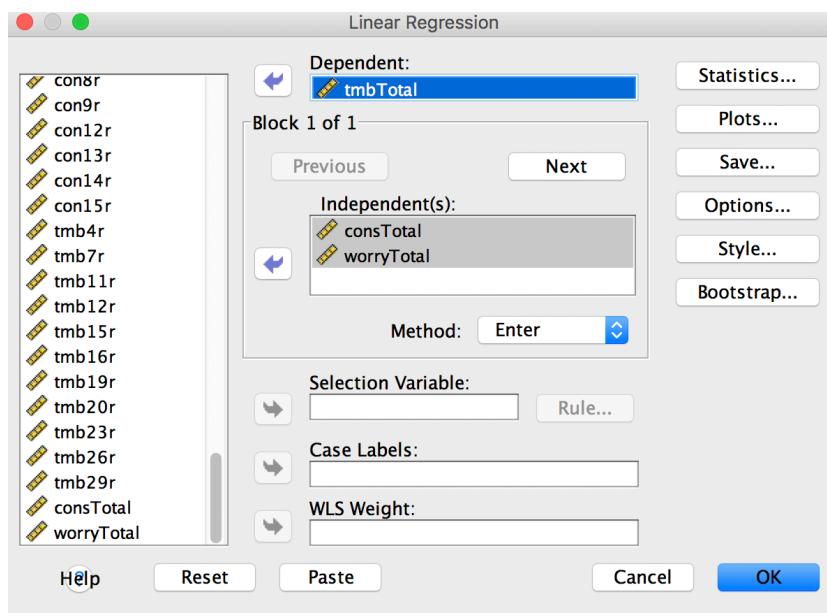


Figure 10.1: SPSS Multiple Regression dialogue

10.3.4 Step Four: Checking assumptions of multiple regression

Last time, you carried out the multiple regression without checking for assumptions. But, like any statistical test, multiple regression has certain prerequisites. The analysis will become unreliable if these are

not observed.

(45) Test for *normality*, *homoscedasticity* and *linearity* by clicking on *Plots*.

- In *Standardized Residual Plots*, select both *Histogram* and *Normal Probability Plot*.
- In the *Scatter 1 of 1* option, select **ZRESID* from the left hand box and move it into the *Y:* box; select **ZPRED* and move it into the *X:* box. This will produce a scatterplot of the standardised residuals (**ZRESID*) against the standardised predicted outcome (**ZPRED*).
- Your settings should look like the Figure 10.2.

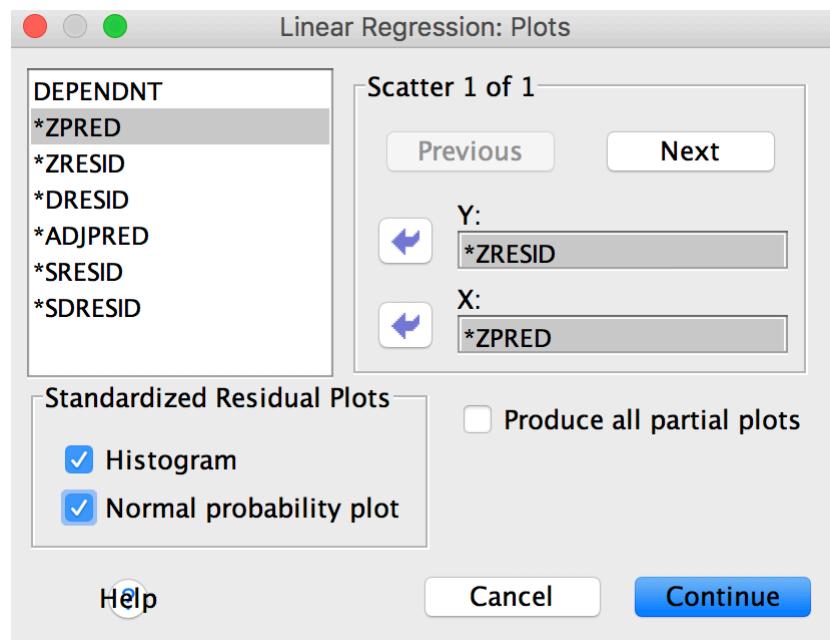


Figure 10.2: SPSS Linear Regression, Plots

(46) Test for the assumption of *independent errors* by producing the Durbin-Watson statistic. You'll find this in the *Statistics* box,

Figure 10.3.

- Include the other checkboxes you see in Figure 10.3: *Estimates*, *Model fit*, *Descriptives*, *Part and partial correlations*, and *Collinearity diagnostics*.

- (47) Test for outliers by looking at *casewise diagnostics*. This is also something that can be selected from the *Statistics* box, Figure 10.3.

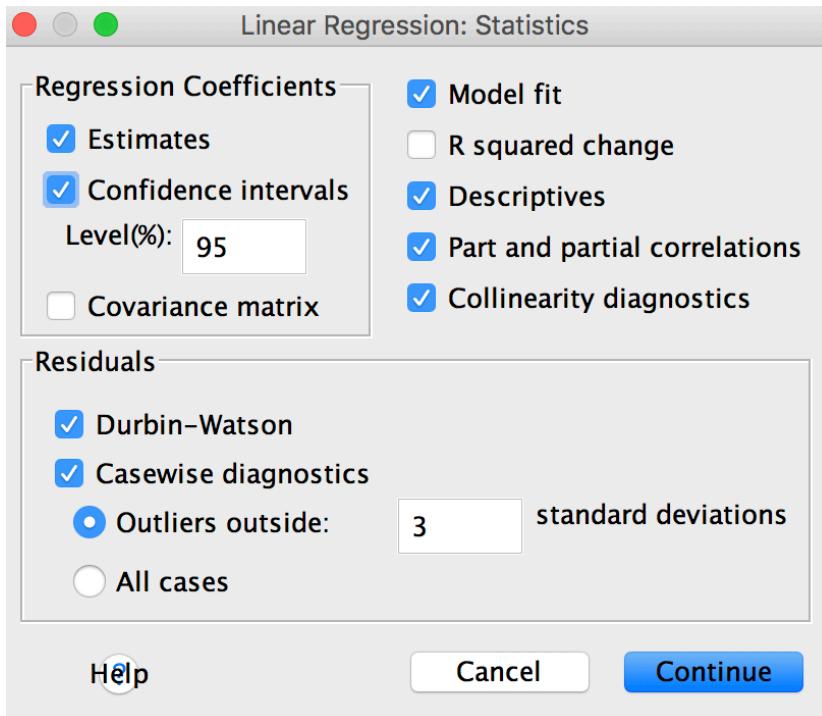


Figure 10.3: SPSS Linear Regression, Statistics

An *outlier* is an extreme case of an individual who responded in a very different way compared with the rest of the sample. As multiple regression is sensitive to outliers, we may want to remove them and re-run the analysis without them. Extreme outliers can be easily

identified using a statistic called Cook's distance.

- (48) To ask SPSS to compute Cook's distance, return to the main Linear Regression dialogue Figure 10.1. There, click *Paste* to generate the syntax that will run the multiple regression. It should look like Figure 10.4.

```

11 * Regression.
12 REGRESSION
13 /DESCRIPTIVES MEAN STDDEV CORR SIG N
14 /MISSING LISTWISE
15 /STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL ZPP
16 /CRITERIA=PIN(.05) POUT(.10)
17 /NOORIGIN
18 /DEPENDENT tmbTotal
19 /METHOD=ENTER consTotal worryTotal
20 /SCATTERPLOT=(*ZRESID ,*ZPRED)
21 /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
22 /CASEWISE PLOT(ZRESID) OUTLIERS(3).
23

```

Figure 10.4: SPSS Syntax for Multiple Regression

- (49) Identify the line that reads:

/RESIDUALS DURBIN HIST(ZRESID) NORMPROB(ZRESID)

- (50) Type *OUTLIERS(COOK)* after '*NORMPROB(ZRESID)*' so that the line now looks like this:

/RESIDUALS DURBIN HIST(ZRESID) NORMPROB(ZRESID)
OUTLIERS(COOK)

- The Syntax window should now look like Figure 10.5.

You have now asked SPSS for Cook's distance. This is a reliable way of identifying cases that could unduly influence your model.

```

11 * Regression.
12 REGRESSION
13 /DESCRIPTIVES MEAN STDDEV CORR SIG N
14 /MISSING LISTWISE
15 /STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL ZPP
16 /CRITERIA=PIN(.05) POUT(.10)
17 /NOORIGIN
18 /DEPENDENT tmbTotal
19 /METHOD=ENTER consTotal worryTotal
20 /SCATTERPLOT=(*ZRESID ,*ZPRED)
21 /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID) OUTLIERS(COOK)
22 /CASEWISE PLOT(ZRESID) OUTLIERS(3).
23

```

Figure 10.5: SPSS Syntax for Multiple Regression, Edited

Answer

SPSS syntax for the above:

```

* Regression.

REGRESSION

/DESCRIPTIVES MEAN STDDEV CORR SIG N

/MISSING LISTWISE

/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL ZPP

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT tmbTotal

/METHOD=ENTER consTotal worryTotal

/SCATTERPLOT=(*ZRESID ,*ZPRED)

/RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)

OUTLIERS(COOK)

/CASEWISE PLOT(ZRESID) OUTLIERS(3).

```

- (51) Go ahead and run the multiple regression. Do this by highlighting the syntax in the Syntax Window and clicking *Run selection* (if this isn't visible as part of the window itself, go to the menu *Run > Selection*).

10.3.5 Step Five: Interpret the output

You've already interpreted a multiple regression in last week's session.

Remember that the key things to look at are:

- The R^2 (and R^2_{adj}), which tells you how much variance in the outcome can be explained by the model. Find this in the *Model Summary*, Figure 10.6.
- The F produced by the ANOVA, which tells you whether or not the model is reliable. Find this in the *ANOVA* table, Figure 10.7.

Model Summary^b					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.495 ^a	.245	.229	16.19186	2.190

a. Predictors: (Constant), worryTotal, consTotal

b. Dependent Variable: tmbTotal

Figure 10.6: SPSS Output—Model Summary

ANOVA^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	8239.476	2	4119.738	15.714	.000 ^b
Residual	25431.114	97	262.176		
Total	33670.590	99			

a. Dependent Variable: tmbTotal

b. Predictors: (Constant), worryTotal, consTotal

Figure 10.7: SPSS Output—ANOVA

- The B and *beta* coefficients, which provide information regarding the strength and direction of the relationship between each predictor and the outcome. See Figures Figure 10.8 and Figure 10.9. Use these to write the unstandardised or raw (Equation 10.1) and standardised regression (Equation 10.2) equations:

$$(DV_{pred}) = b_0 + (b_1 IV_1) + (b_2 IV_2) \quad (10.1)$$

$$(ZDV_{pred}) = (\beta_1 ZIV_1) + (\beta_2 ZIV_2) \quad (10.2)$$

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	157.472	10.510		14.983	.000
consTotal	-.865	.163	-.469	-5.312	.000
worryTotal	-.274	.137	-.177	-1.999	.048

Figure 10.8: SPSS Output—Coefficients, lefthand section

95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
136.612	178.332					
-1.188	-.542	-.462	-.475	-.469	.998	1.002
-.546	-.002	-.158	-.199	-.176	.998	1.002

Figure 10.9: SPSS Output—Coefficients, righthand section

- The *t* values, which tell us the predictors that predicted a significant amount of variance in the outcome variable.
- The *correlations*, which you can use to calculate the contributions of each predictor to the outcome. For instance, how much variance in the outcome did each individual predictor independently explain? Remember, because r^2 is shared variance (measured from 0 to 1), you need to square these values and multiply by 100 to get percentages of variance explained by each predictor.

(52) Was the model significant? Report the appropriate statistics in APA style.

Answer

- Yes, the model was significant. $F(2,97) = 15.71$, $MSE = 262.13$, $p < .001$

(53) How much variance in the outcome variable did the model explain? Use the unadjusted and adjusted measures.

Answer

- $R^2 = .245 = 24.5\%$
- $R_{adj}^2 = .229 = 22.9\%$

(54) What are the unstandardised and standardised multiple regression equations?

Answer

Unstandardised:

$$(timeManagement_{pred}) = 157.47 - 0.87(conscientiousness) - 0.27(worry)$$

Standardised:

$$(ZtimeManagement_{pred}) = -0.47(Zconscientiousness) - 0.18(Zworry)$$

(55) Were the individual predictors significant? How much unique variance did each of them explain? (Report appropriate statistics to justify your answer.)

Answer

- Both predictors were significant individually, since both t-values (5.31, 2.00) were significant ($p < .001$, and $p = .05$).
- Conscientious predicted 22.09% ($(-0.47^2) * 100$) unique variance in time management, whereas worry explained 3% ($(-0.18^2) * 100$).

10.3.6 Assumptions

You've performed the main analysis. Now let's look at the assumptions.

- Tolerance allows you assess problems of collinearity between your predictors. You find the statistics on the right hand side of the *Coefficients* table. See Figure 10.9. Remember that tolerances below .10 are problematic.
- You can assess normality, linearity and homoscedasticity using the histograms (Figure 10.10), P-P plot (Figure 10.11) and scatterplot (Figure 10.12).

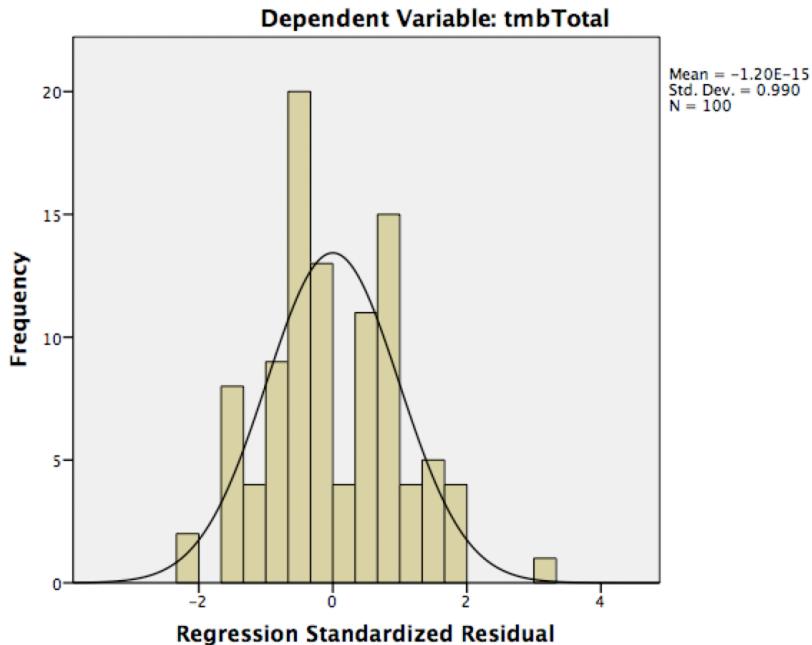


Figure 10.10: SPSS Output—Histogram of Frequency against Standardised Residuals

- Residual statistics (Figure 10.13) show us the range of the predicted values on the outcome measure. More importantly, it shows us the range of the residuals. Everyone in your sample has a residual that represents the difference between their actual score on the outcome measure and the predicted score based on the regression

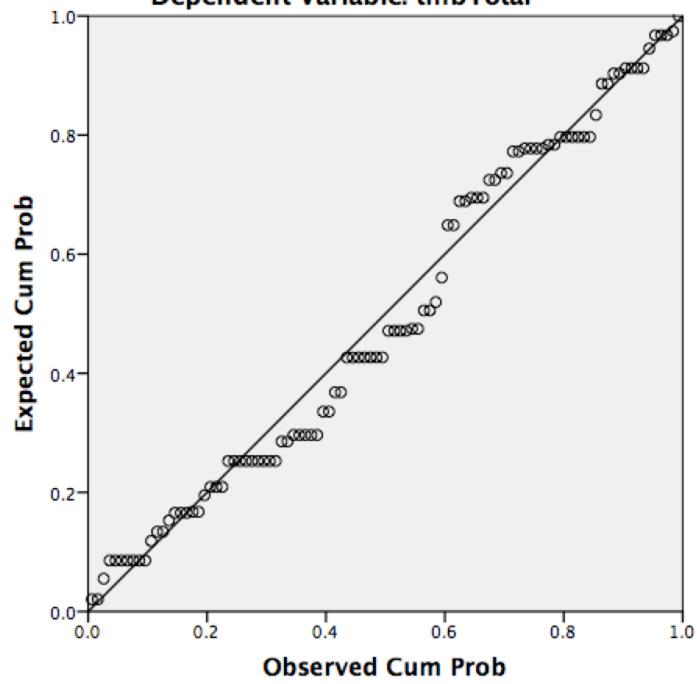
Normal P-P Plot of Regression Standardized Residual**Dependent Variable: tmbTotal**

Figure 10.11: SPSS Output—Normal P-P Plot

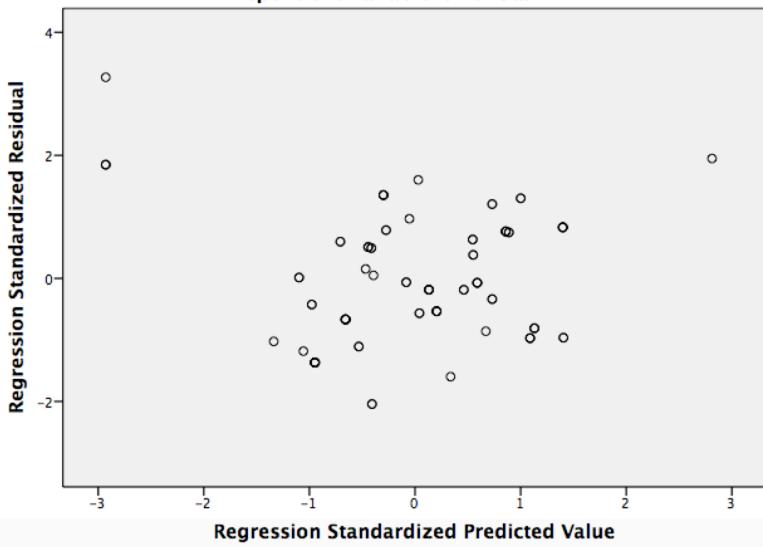
Scatterplot**Dependent Variable: tmbTotal**

Figure 10.12: SPSS Output—Scatterplot of Residuals against Predicted Values

model. The larger these values, the worse your model is in predicting the outcome.

Residuals Statistics^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	77.0746	129.4461	103.7900	9.12289	100
Std. Predicted Value	-2.928	2.812	.000	1.000	100
Standard Error of Predicted Value	1.814	5.621	2.684	.818	100
Adjusted Predicted Value	69.8219	126.2924	103.6118	9.59756	100
Residual	-33.07341	52.92538	.00000	16.02747	100
Std. Residual	-2.043	3.269	.000	.990	100
Stud. Residual	-2.068	3.485	.005	1.016	100
Deleted Residual	-33.90773	60.17815	.17818	16.90208	100
Stud. Deleted Residual	-2.104	3.707	.008	1.029	100
Mahal. Distance	.253	10.942	1.980	2.204	100
Cook's Distance	.000	.555	.019	.063	100
Centered Leverage Value	.003	.111	.020	.022	100

a. Dependent Variable: tmbTotal

Figure 10.13: SPSS Output—Residuals Statistics

What we are really interested in, however, is if the model is bad in predicting anyone. In other words, does anyone have a very large residual? We've already requested *casewise diagnostics* (Figure 10.3), so if anyone does have a very large residual, they should appear in a table like Figure 10.14. The Std. Residual in Figure 10.13 tells us the range (minimum and maximum) of the standard deviations of the residuals (i.e. the range of the standardized residuals). If the range falls within +/- 3 standard deviations, then we don't have any outliers. (Note that the maximum for Std. Residual is 3.269, beyond the accepted range, indicating the presence of some large residuals.)

We can see this in Figure 10.14, below, which identifies the outlying case (30).

Casewise Diagnostics^a				
Case Number	Std. Residual	tmbTotal	Predicted Value	Residual
30	3.269	130.00	77.0746	52.92538

a. Dependent Variable: tmbTotal

Figure 10.14: SPSS Output—Casewise Diagnostics

But just because there is a large residual (error) associated with a case, that doesn't mean the case is influencing our regression unduly.

To determine how much influence it has, we need to look at Cook's distance, which you'll see in Figure 10.15.

Outlier Statistics^a				
	Case Number	Statistic	Sig. F	
Cook's Distance	1	.555	.646	
	2	.177	.911	
	3	.177	.911	
	4	.177	.911	
	5	.139	.936	
	6	.037	.991	
	7	.036	.991	
	8	.036	.991	
	9	.026	.994	
	10	.026	.994	

a. Dependent Variable: tmbTotal

Figure 10.15: SPSS Output—Outlier Statistics

This table lists the top 10 cases, or, in other words, the 10 cases with the highest Cook's Distance. If any of the cases listed are unduly influencing your results, their Cook's distance will be greater than 1.00. If this is the case, you should consider carefully what to do (the case might need to be removed). Note the Cook's distance for case 30.

(56) Bearing in mind the above tables, do the data contain outliers?

Do you need to worry about them?

- Report the appropriate statistics to justify your answer.

Answer

Yes, case 30 was an outlier because its standardised residual was greater than 3 standard deviations (3.27) from the mean residual (0, since this is a standard distribution). However, Cook's distance for this case was .56, which is less than the

recommend cut-off of 1, so we need not be unduly concerned with case 30.

(57) Again, bearing in mind the above tables, do the data meet the assumptions of multiple regression? If not, why not?

- Report the appropriate statistics to justify your answer.

Answer

- Independence was met (Durbin-Watson was just over 2). See Figure 10.6.
- Normality: Some deviation from normality, as shown in the Histogram (Figure 10.10) and the Normal P-P plot (Figure 10.11).
- Linearity and homoscedasticity: The scatterplot between the standardised predicted outcome and standardised residuals show that there is no problem with linearity or homoscedasticity (Figure 10.12).
- Multicollinearity was not an issue since the Tolerance statistic is above .10 (Figure 10.9).

(58) Summarise the results of this regression.

Answer

The model was significant ($F(2,97) = 15.71$, $MSE = 262.13$, $p < .001$) and it predicted 25% [23% if you prefer to use the adjusted R^2] of the variance in Time Management Behaviour (TBM). Both predictors were significant (both p values $< .05$)

but conscientiousness was the most important predictor and it uniquely explained 22% of the variance in TBM, whereas worry only explained 3%. [If you wish, you can include the standardised and unstandardised models you provided in a previous answer.]

Colophon

11.1 References

- Costa, P. T., & McCrae, R. R. (1985). *The NEO personality inventory: Manual, form S and form R*. Psychological Assessment Resources.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage Publications.
- Howell, D. C. (1997). *Statistical Methods for Psychology (Fourth Edition)*. California, USA: Wadsworth.