MPSMD2RES Workshop 1: Measures of Central Tendency, Association and Difference

Dr Ian Hocking, Psychology Programme

Contents

1 Overview

Data files required	None
Booklet Version	1.0
Format	large text PDF

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 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 simple correlation;
- The section on between-subjects differences will remind you how to test for differences between two groups of data that are not correlated;
- Finally, the section within-subjects differences will take you through testing for differences between two groups of data that *are* correlated.

3 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.

3.1 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.

3.2 Other Aspects of this Booklet

- This formatting will be used to highlight something important.

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

3.3 Mathematics and Statistics Help

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

3.4 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.

4 Workshop

For these exercises, use SPSS.

5 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 2: Visuo-spatial scores for each female participant.

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

Visual Score	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

Mean is *16.50*

Median is *19*

Standard deviation is *6.47*

The SPSS Syntax for these data is as follows:

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

6 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 3: Verbal scores for each female participant.

Participant number	1	2	3	4	5	6	7	8	9
Verbal Score	68	82	45	45	32	52	45	52	65
Participant number	10	11	12	13	14	15	16	17	18
Verbal Score	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.

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?

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

6.1 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 4: Verbal scores for each male participant.

Participant number	1	2	3	4	5	6	7	8	9
Verbal Score	62	71	38	41	29	35	28	51	58
Participant number	10	11	12	13	14	15	16	17	18
Verbal Score	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.

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 data is as follows:

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

7 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 5: Verbal scores for each male participant after training.

Participant number	1	2	3	4	5	6	7	8	9
Verbal Score	68	75	59	58	51	62	59	62	59
Participant number	10	11	12	13	14	15	16	17	18
Verbal Score	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.

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.
```

8 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*?

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 measures of central tendency

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?

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.

In this booklet, you've revised mean, median and standard deviation; you've calculated a parametric statistic of association, r; you've calculated a parametric statistic of difference, t, for between-group designs and within-group designs. You've also reported these in APA style.

9 Versions

This document is available in standard PDF, simplified layout PDF, standard dark theme PDF, PDF with Open Dyslexic font, large text PDF, HTML (webpage) format and spoken format. This document contains hyperlinks to sections within it, external webpages, and email addresses like ian.hocking@canterbury.ac.uk.