

EXERCISE 8

Comparing the averages of two independent samples of data

Aim

The previous Exercises have used data from a questionnaire. The next few Exercises will be based on data from experiments designed to test experimental hypotheses. In real experiments, of course, a larger number of participants would have been used.

Before you start

Before proceeding with this Exercise, we suggest you read Chapter 6 carefully. In this Exercise, we shall be making an **independent samples *t* test** (Section 6.2.4). In Exercise 9, we shall be making a **paired-samples *t* test** (Section 6.2.5). Finally in Exercise 10, we shall be making **one-sample tests** (Section 6.5).

An investigation of the effects of a drug upon performance

The data we are going to explore in this Exercise might have been produced by the following project. A team of investigators has good reason to believe that a small dosage of a certain drug changes the speed with which people can make decisions. They decide to try to confirm this by carrying out an experiment in which the decision times of 14 people who have taken the drug are compared with those of a control group of 14 other people who have performed the task under a placebo condition. The experimenters expect that the average decision time of the experimental group will differ from that of the placebo group. The results are shown in Table 1.

DRUG GROUP				PLACEBO GROUP			
Case	Time	Case	Time	Case	Time	Case	Time
1	390	8	425	15	446	22	440
2	494	9	421	16	749	23	471
3	386	10	407	17	599	24	501
4	323	11	386	18	460	25	492
5	660	12	550	19	390	26	392
6	406	13	470	20	477	27	578
7	345	14	393	21	556	28	398

Opening SPSS

In the opening window of SPSS, select the **Type in data** radio button. If **Data View** appears first, click the tab **Variable View** to open **Variable View**.

Constructing the SPSS data set

Construct the data set as described in Section 3.1. In **Variable View**, the first variable, *Case*, will represent the participants. The second is the grouping variable (i.e. the type of treatment – drug or placebo). Call the grouping variable *Condition* and label the values: *1 = Drug; 2 = Placebo*. Label the variable *Experimental Condition*. The third variable, which can be named *Score*, contains all the participants' scores on the dependent variable. Notice that the *Score* variable includes the scores for **both** treatments. The grouping variable *Condition* is needed to enable the computer to identify the group to which a score belongs. Since there are no decimals in the data, ensure that the values in the **Decimals** column are all *0*.

Click the **Data View** tab and enter the data of Table 1 into **Data View** in the manner described in Section 3.1.2. When the data have been entered, save them to a file with a name such as **Drugs**.

Exploring the data

The first step is always to examine the data set to see whether there are any odd features.

We shall want a table of means and standard deviations, together with indicators of distribution shape such as histograms and boxplots. The statistics for the subgroups are most easily obtained with the **Means** procedure. (The plots are obtained with the **Explore** procedure.) Follow the instructions in Section 4.3.2, remembering that the dependent variable name is *Score* and the independent variable name is *Condition*.

- **Write down the values of the means and standard deviations. Do these statistics appear to support the scientific hypothesis?**

(Note that the **Means** procedure requires the presence of a grouping variable in the data set. Should the mean and standard deviation of a set of ungrouped data be required, use the **Descriptives** procedure.)

Graphical displays of the data

To draw the boxplots, proceed as described in Section 4.3.2. The dependent variable is *Score*, the factor is *Condition*, and the **Labels Cases by** is *Case*. (This choice labels any outliers or extreme scores in the boxplots with the number of the case, which is more useful than the default row number, especially if some cases have been deselected.) Remember to click the **Plots** radio button in the **Display** section of the **Explore** dialog box, thus turning off the **Both** radio button and ensuring that the descriptive statistics tables are omitted. Click **Plots**, deselect the **Stem-and-leaf** check box and select **Histogram** check box. Click **Continue** and finally **OK**.

The output in **SPSS Viewer** begins with the usual **Case Processing Summary** listing the number of valid cases in each group. Then it shows **Boxplots** and **Histograms** for the two groups. The boxplots show two outliers with the identifying numbers of the participants concerned (because you specified the participants' numbers in the **Label Cases by** box in the **Explore** dialog box).

When there is a marked discrepancy between the mean and median of a set of scores, the distribution is probably skewed or otherwise asymmetrical. Atypical scores, or **outliers** can also pull the value of the mean away from that of the median. Read Section 4.3.2 carefully for an explanation of SPSS's boxplot displays.

- **Identify any outliers by means of their identifiers, which are values in the variable case. State their values.**

Printing the output

If you want a hard copy of the output, follow the procedure described in Section 3.5. The precise details will depend upon your local set-up.

The independent samples *t* test

Run an independent samples *t* test on the full data set as described in Section 6.2.4 but do not remove any outliers at this stage.

Output for the independent samples *t* test

Guidance on how to interpret the output is given in Section 6.2.4. We suggest you study that section and try to answer the following questions.

- **On the basis of the Levene test p-value, which row of the *t* test results will you use?**
- **Write down the value of *t* and its tail probability. Is the p-value evidence against the null hypothesis? Remember that if the result is sufficiently unlikely (i.e. $p < 0.05$) under the null hypothesis, it is regarded as evidence against the null hypothesis and hence in favour of the experimental hypothesis.**
- **Write down your interpretation of the result of the test: has the *t* test confirmed the pattern shown by the means of the two groups?**
- **If the hypothesis had been one-tailed (e.g. that decision times of the experimental group will tend to be shorter than those of the control group), then the appropriate p-value would be obtained by dividing the two-tailed p-value by 2. What would be the one-tailed p-value in that case?**
- **Calculate the effect size using the formula given in Section 6.2.4. Is the effect size small, medium or large according to Cohen's table?**

A nonparametric equivalent of the independent samples *t* test: The Mann-Whitney U test

The running of the **Mann-Whitney** test on SPSS is described in Section 6.4.3. Run the procedure as described in that section.

Output for the Mann-Whitney test

The output gives the values of the statistics U and W (the W statistic belongs to a test by Wilcoxon which is the exact equivalent of the Mann-Whitney), followed by a standard normal deviate score *z* and a 2-tailed probability value corrected for ties. An exact 2-tailed probability value not corrected for ties concludes the table. If the p-value is less than 0.05, the null hypothesis can be rejected and the groups declared to differ significantly.

- **Write down the results of the Mann-Whitney test, including the value of U and its p-value. State whether the result is significant and whether the Mann-Whitney test confirms the result of the *t* test. In what circumstances might you expect the p-values of U and *t* to differ?**

Printing the output

To obtain a hard copy of the output, proceed as described in Section 3.5.

Re-running the tests after deselecting the two outliers

Deselect the two outliers using the **Select Cases...** procedure (see Section 3.3.1) by entering score <600 in the **Select Cases** dialog box. Then re-run the *t* test and the Mann-Whitney test.

- **Write down the new value of *t* and its p-value (assuming a two-tail test). Is the conclusion different from what it was with the complete data set?**
- **Calculate the effect size using the formula given near the end of Section 6.2.4. Is the effect size small, medium or large according to Cohen's table?**
- **Write down the results of the Mann-Whitney test (assuming a two-tail test), including the value of U and its p-value. Is the conclusion different from what it was with the complete data set?**
- **Write down your interpretation of the effects on each test of eliminating the outliers.**

Finishing the session

Close down SPSS and any other windows before logging out of the computer.