SPSS Instructions for two-sample/independent samples t-test

To carry out a two-sample t-test, you need to:

- Explore the data visually using a boxplot
- Carry out the two-sample t-test
- Check that the data is Normally distributed

The data file for this example is *upgrade5.sav*. The file records the height, weight, alcohol and cigarette consumption of a sample of men aged either 16 or 18years.

Open SPSS and the data file

On the Brookes PC network, select Start > All programs > Statistics > SPSS, choosing version 15.0.

If necessary, when SPSS has loaded, click Cancel to avoid a tutorial session.

To open the data file, *upgrade5.sav*, select File > Open and then navigate to the folder where you have downloaded and saved the file for this example, and open the file.

The data file contains ten variables, in this example we will use:

Height   Height in m
Age   Age in years, either 16 or 18.

We will use a two-sample t-test to determine whether mean height of 18 year old men differs from that of 16 year olds.The two-sample t-test is appropriate because:

1) The samples are independent (The data was obtained from two groups of men. If the same individuals had been measured at different ages, we would have related samples.)

2) The measurement being compared, height, is measured on a continuous scale.

## Explore the data visually using a boxplot

- Select **Graphs > Legacy dialogs> Boxplot > Simple > Summaries for groups of cases > Define.**
- Select *height* and click on the ▸ symbol next to the box labelled **Variable**.
- Select *age* and click on the ▸ symbol next to the box labelled **Category Axis**.
- Click on **OK**.

Carry out a two-sample t-test
- Select **Analyse > Compare Means > Independent Samples T-Test.**
- Select *height* and click on the ▸ symbol next to the box labelled **Test Variable.**
- Select *age* and click on the ▸ symbol next to the box labelled **Grouping Variable.**
- Click on **Define Groups** and enter *16* and *18* in the boxes.
- Select **Continue > OK.**

SPSS generates two tables of output. The first table summarises the scores in each group, the second contains the test output:

Your output should look something like this. Check that it is the same:
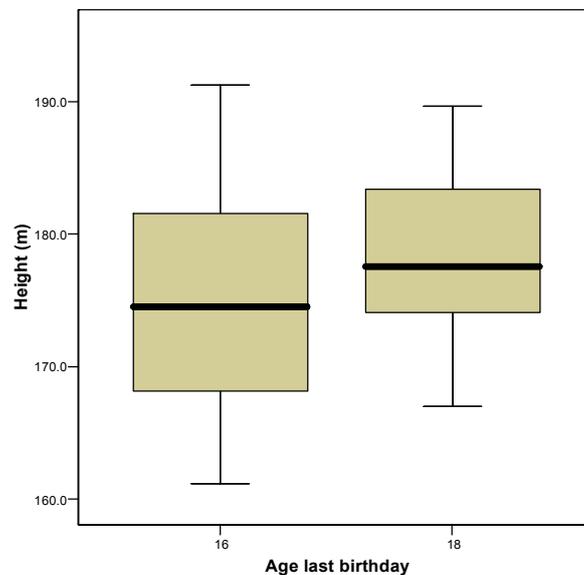
Age last birthday

T-Test

Group Statistics

	Age last birthday	N	Mean	Std. Deviation	Std. Error Mean
Height (m)	16	66	174.648	7.5472	.9290
	18	46	178.393	5.6677	.8357

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Height (m)	Equal variances assumed	4.837	.030	-2.850	110	.005	-3.7450	1.3139	-6.3489	-1.1411
	Equal variances not assumed			-2.997	109.341	.003	-3.7450	1.2495	-6.2215	-1.2685

Can you use the output to answer the following questions?

First look at the boxplot

1   How did the heights differ between the 16 year olds and the 18 year olds measured in the sample?

2   Does the symmetry and the variation within groups shown in the boxplot support the assumptions underlying a two-sample t-test?

3   What are the hypotheses tested by the two-sample t-test?

4   Why is a one sided test suitable here?

5   Which line of t-test output should be used and why?

Answers

1   The median height was somewhat higher for the 18 year olds. Neither sample contained outliers.

2   Variation in height appears reasonably similar – which tends to support the equal variance assumption. Both distributions appear reasonably symmetric. This, with the lack of outliers tends to support the assumption that height is Normally distributed for men in age group.

3   Null: mean height is the same for men aged 16 and 18
Alternative: mean height is higher for men aged 18 than men aged 16

4   A one-sided test might be used if the objective is to determine whether 16 year old men are still growing.

5   The first line, showing the