1 Two independent sample t-test

1.1 Overview

The two independent sample t-test is used when we want to know whether there is a difference in the means of a quantitative variable on two populations that are independent (i.e groups that do not share any members or contain matched pairs). The hypothesis being tested is that the means are equal for the two populations.

Here is an example of when the test is used: A lecturer from the psychology department claims that women have on average a higher IQ than men. To test this claim one of her student's collects the IQ scores of 15 men and 15 women. A two-independent sample t-test is required to test the hypothesis that the means are the same.

The assumptions of the two independent sample t-test are:

- Random sampling from a defined population.

- Quantitative measurements.

- Measurements in population are normally distributed. However, there is a rule of thumb that for sample sizes (individual sample sizes; not total sample) larger than 30 the t-test can still be used. (This is due to the central limit theorem).

- The population variances are equal. A rule of thumb for this is that the largest standard error should not be larger than two times the smallest standard error. If you do have unequal variances then when using SPSS and R you need to state this, and it will carry out Welch's unequal variance two independent sample test.

If you do not have quantitative response data, then you are in the wrong part of the flowchart and it is advised that you start again and look at the definitions of ordinal, nominal and quantitative variables.

If you do not have a categorical predictor with two-independent levels you are advised to go through the flowchart again.

If your sample size is smaller than 30 and your measurements are far from normal then it would be better to use a Wilcoxon Mann-Whitney test.

1.2 Analysis plan

The descriptive statistics you should have carried out are: calculate the sample size, sample mean, variance and standard error for each sample. It may also be useful to calculate skewness and kurtosis to assess whether your data is normally distributed. You can use side by side box plots, one for the first sample and one for the second sample to visualise the data. You can do the same for histograms.

A good follow up test is to calculate a 95% confidence interval for the difference in means between the two populations.

Here are some examples of a two-independent sample t-test being carried out:
http://www.youtube.com/watch?v=vV0RtkI2jxU   (by hand)
http://www.youtube.com/watch?v=vtkW1il7kRc   (in SPSS)
http://ww2.coastal.edu/kingw/statistics/R-tutorials/independent-t.html   (in R)