Two sample t-test

Task 3

Gent Rexha

19th of March 2019

Introduction

The two sample t-test is also known as an independent t-test. An independent t-test is used to compare two population means.

Assumptions

- Requires independent groups for each sample.
- Normal distribution - there should roughly be a normal distribution in both the populations and in the samples.
- Approximately similar variance - in each of the samples.
- Datapoints - there should be the same number of data points on either sample. And then finally this works good with low numbers, but you generally want to be in the 20 to 30 number range.

Procedure for carrying out a paired t-test

To test the null hypothesis that the true mean difference is zero, the procedure is as follows:

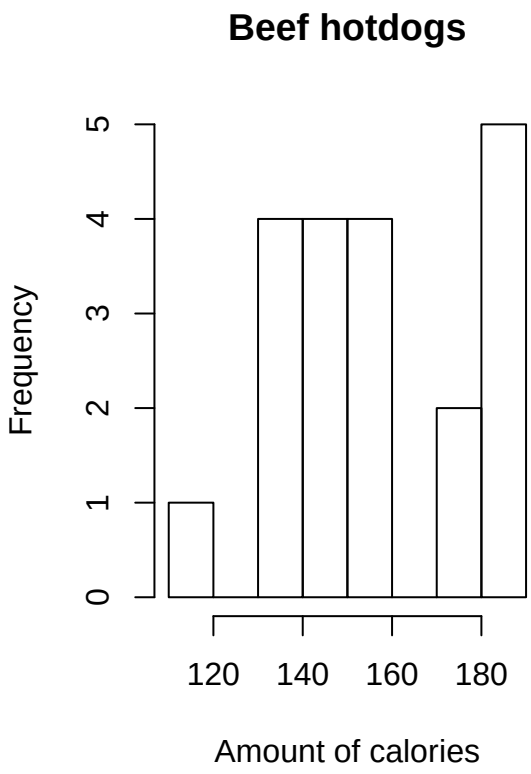
1. Calculate the difference between the two sample means, $\bar{x_1} - \bar{x_2}$.
2. Calculate the pooled standard deviation: $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$
3. Calculate the standard error of the difference between the means: $SE(\bar{x_1} - \bar{x_2}) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{N_2}}$
4. Calculate the T-statistic, which is given by $T = \frac{\bar{x_1}-\bar{x_2}}{SE(\bar{x_1}-\bar{x_2})}$. Under the null hypothesis, this statistic follows a t-distribution with $n_1 + n_2 - 2$ degrees of freedom.
5. Use tables of the t-distribution to compare your value for T to the $t_{n_1+n_2-2}$ distribution. This will give the p-value for the unpaired t-test.

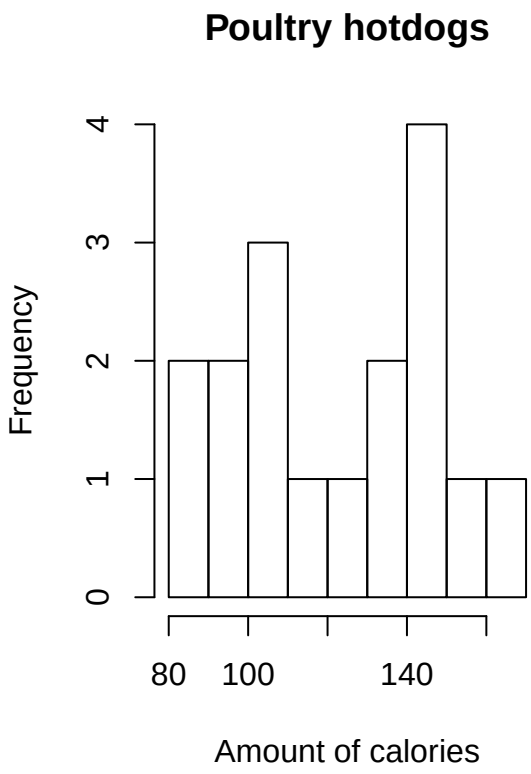
Example

(Data taken from Moore and McCabe - Introduction to the Practice of Statistics) A U.S. magazine, Consumer Reports, carried out a survey of the calorie and sodium content of a number of different brands of hotdog. There were three types of hotdog: beef, 'meat' (mainly pork and beef but can contain up to 15% poultry) and poultry. The results below are the calorie content of the different brands of beef and poultry hotdogs.

```
beef_hotdogs <- c(186, 181, 176, 149, 184, 190, 158, 139, 175, 148,
                  152, 111, 141, 153, 190, 157, 131, 149, 135, 132)
poultry_hotdogs <- c(129, 132, 102, 106, 94, 102, 87, 99,
                     170, 113, 135, 142, 86, 143, 152, 146, 144)
```

Before carrying out a t-test you should check whether the two samples are roughly normally distributed. This can be done by looking at histograms of the data. In this case there are no outliers and the data look reasonably close to a normal distribution; the t-test is therefore appropriate.





So, first we need to calculate the sample mean and standard deviation in each group, for this purpose R offers an already implemented function for this:

```
t.test(beef_hotdogs, poultry_hotdogs)
```

```
##
##  Welch Two Sample t-test
##
## data:  beef_hotdogs and poultry_hotdogs
## t = 4.3031, df = 32.394, p-value = 0.0001455
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  18.11306 50.64577
## sample estimates:
## mean of x mean of y
##  156.8500  122.4706
```

If we look this up in tables of the t-distribution with 35 degrees of freedom, we find $p < 0.001$. Therefore, there is strong evidence that the calorie content of poultry hotdogs is lower than the calorie content of beef hotdogs.

References

[1] "[No title]." [Online]. Available: http://www.statstutor.ac.uk/resources/uploaded/unpaired-t-test.pdf. [Accessed: 20-Mar-2019]