Correlations in R
Investigating Relationships

Lauren Kennedy

School of Psychology, University of Adelaide

2016-Version 1.1

1 Assumed knowledge

This guide specifically teaches you how to calculate a correlation and do a correlation test in R. Correlations are used to describe and test the relationship between two variables. To learn more about correlations, refer to the Descriptive Statistics chapter in Learning Statistics with R. This guide assumes that you have installed R and R Studio, have looked at the Getting Started Guide and downloaded your practical data. You should know how to use functions (see the guide titled "Fun with Functions"). We will use the *lsr* package in this guide. If you're not sure how to access the *lsr* package, see the Getting Started in R help guide.

2 The data

The data we are using for this guide is already pre loaded in R. You won't see it in your environment panel, but it's there. For this guide we are going to use a data frame called trees. It contains the height (measured in feet), Girth(measured in inches) and Volume(measured in cubic feet) of 31 felled black cherry trees. We're going to use this data set to calculate the correlation between height and girth of the trees, the correlation matrix of height, girth and volume, and conduct a test to see if the correlation between height and girth is significantly different from zero. If you type the following you will see the full data set.

```
View(trees)
```

3 Calculating Correlations

3.1 Calculating a single correlation

The first thing that we'd like to do is calculate the correlation between tree height and tree girth. To do this we are going to use the *correlate* function. This function takes two *vectors*, which are a collection of numbers that have an order. The two vectors we are going to use are the columns *Height* and *Girth*. To get those columns from the dataframe *trees*, we use the *$* operator. *trees$Height* and *trees$Girth* select the Height and Girth columns respectively. If we combine this with the *correlate* function we get the following code, which tells us the correlation is $.52$.

```
correlate ( trees$Height , trees$Girth )
```

3.2   Calculating multiple correlations at once

In the previous section we used the *correlate* function to calculate the correlation between Height and Girth. However, we might want to calculate the correlations between all of the variables in our data frame, or a correlation matrix. If not all of the columns are numeric (or made of numbers), only the the columns that are numeric will be shown.

Once you've checked this, you can give your whole data frame to the *correlate* function like below:

```
correlate ( trees )
```

This produces a *correlation matrix*, which is a table that shows the correlation between the variables listed in the rows and those in the columns. If you look at the second row (Girth) and the first column (Height) you'll see the correlation between girth and heigth, which we found before.

4   Testing Correlations

Whilst it's very useful to be able to calculate correlations, oftentimes you will need to do a test to see if that correlation is statistically different from zero. To do this, we would use a correlation test. We can still do this with the *cor.test* function.

Like before, the function takes two vectors. We'll use the same Girth and Height as before..

```
cor . test ( trees$Height , trees$Girth )
```

Which produces the output below.

```
  Pearson 's product−moment correlation

data :   trees$Height and trees$Girth
t = 3.2722 , df = 29 , p−value = 0.002758
alternative hypothesis : true correlation is not equal to 0
95 percent confidence interval :
 0.2021327 0.7378538
sample estimates :
       cor
0.5192801
```

As the p value is less than .05, this tells us that the correlation between the girth and height of the black cherry trees is significantly different from zero. Here's how we might right this up:

> Thirty one black cherry trees were included in the sample. The height (in feet) and girth (in inches) of the trees were measured. The relationship between the height and girth of the trees was significant, $r(29) = .52$, $p = .003$. As the correlation was positive, this indicates that an increase in girth of the trees is related to an increase in height.