Some helpful R commands

To read in data from an Excel csv file called *excel_data.csv* to R Studio and name it *mydata*, first use the drop down menus in R Studio **Session > Set Working Directory > Choose Directory** to indicate the location of *excel_data.csv* on your computer. The following code will then read the data in to R Studio:

```
mydata<-read.csv("excel_data.csv")
```

`attach(mydata)` — this adds the variable names

At the end of the analysis remember to use `detach(mydata)` to disassociate the variable names.

(a) Graphics

`hist(X,col="yellow",main="Histogram of X (units)")` — this produces a histogram of the variable X

`plot(X,Y,xlab="x-axis label",ylab="y-axis label", main="Scatterplot of Y on X",pch=21,bg="black")` — produces a scatter plot of Y on X with the required title, axis labels, and black dots

`pie(table(X), main="Title")` — this gives a simple pie chart of the categories in variable X with the specified title

`barplot(table(X), main="title", xlab="x-axis label",col="orange")` — this gives a bar chart of the categories in the variable X with the required title, axis labels and colour

`boxplot(X)` — produces a boxplot of the numerical variable X

(b) Descriptive Statistics

`mean(X)` — computes the mean of X

`sd(X)` — computes the standard deviation of X

`summary(X)` — computes the mean, median, minimum, maximum, and upper and lower quartiles

`table(X)` — computes the number of observations in each level of the categorical variable X

(c) Correlation and Regression

`cor.test(X,Y)` — computes the correlation between X and Y and performs a test of the null hypothesis of zero correlation

`lm(Y~X)` — fits a linear regression line to the data (lm command stands for linear model)

`abline(lm(Y~X))` — produces a scatterplot with the least squares linear regression line superimposed on the data

`summary(lm(Y~X))` — displays the coefficient of determination (r-squared)

`predict(lm(Y ~ X), newdata=data.frame(X=C),interval = "pred")` — computes the predicted value of Y when X=C along with a 95% prediction interval

(d) Hypothesis Testing

`t.test(X,Y)` — performs a two sample t-test between X and Y

`t.test(X,Y,paired=TRUE)` — performs a paired t-test between X and Y

`prop.test(x = c(a, b), n = c(n1, n2))` — performs a 2-sample test for equality of proportions with continuity correction