6 Effect of diet on early growth of chicks

In this section, we shall investigate a data set that is available in R. This data set is called `ChickWeight` and consists of a single data frame containing the weights of 50 chicks over 21 days (12 time points). The chicks have been split into four groups with each group fed a different diet.

Let's start by loading the data set and reading the documentation. We can also use the `head` function to print out the first few data points.

```
> data(ChickWeight)
> help(ChickWeight)
> head(ChickWeight)
  weight Time Chick Diet
1     42    0     1    1
2     51    2     1    1
3     59    4     1    1
4     64    6     1    1
5     76    8     1    1
6     93   10     1    1
```

You can see the four column structure of the data frame, together with the column names. We can also access the column names using the `names` function.

```
>names(ChickWeight)
 [1] "weight" "Time"   "Chick"  "Diet"
```

When looking at the data table, the columns "Chick" and "Diet" contain numerical data. However, although the entries clearly are integers, the data tables has been set up in such a way that these two columns are defined as so-called "factors". This means that R interprets them as categorical variables that define groups of values (such as all the datapoints for a given chick, or the measurements obtained for a given diet). Treating the entries as regular numerical values will sometimes lead to unexpected results. For example, while pulling out the data for chick number 5 works well (e.g., with `ChickWeight[ChickWeight$Chick==5,]`), subsetting for chicks 1-3 with `ChickWeight[ChickWeight$Chick<3,]` will not produce the expected result (try it out). Here, it would be necessary to combine several conditions such as in `ChickWeight[ChickWeight$Chick==1 | ChickWeight$Chick==2,]`.

It is also important to note that when reading in data from a file, R will interpret numbers as numerical data. This is fine for your measurement values, but can be problematic when you want to define groups of data, such as in the column "Chick" above. In this case, you would have to manually define Chick as a factor. You can avoid these problems by labelling chicks as "C1", "C2" etc when entering your data. Any value that contains characters is automatically interpreted as a factor when R reads in data tables. This note of caution goes slightly beyond the scope

of this tutorial, but is important to keep in mind once you start entering your own experimental data.

Exercise 6.1

1. Make a plot of the weight progression of Chick 1 over the 21 days and save it as a PNG file.
   Hint: Start by subsetting the data frame such that you have the data for Chick 1. For this chick, plot how its weight increases over the 21 days using the `plot` function. Plot time on the x axis and the weight on the y axis.

2. Compare the weight of Chick 1 on day 21 and when it was born. How many times heavier is it?

3. Which of the chicks has the largest weight at 21 days?
   Hint: Look at the documentation for `which.max`

4. Make histograms for the weights of the chicks on diets 1 and 2 after 21 days. Save them as PNG files.

5. Calculate the mean and standard deviations of the chicks on the four different diets after 21 days. On which diet do the chicks appear to be heavier at 21 days?