**統計方法 Statistical methods**

**FALL 2016**

**Homework 1, due: 2016/9/28**

1. 由於大部份的作業問題，會須要以 R 程式軟體來進行實作、分析，因此要求同學們的作業要以 R Markdown (http://rmarkdown.rstudio.com/)的格式來撰寫。請各位同學藉由 R markdown 的說明檔：http://rmarkdown.rstudio.com/lesson-1.html，或者網路上的開放課程與影片，來學習如何撰寫 R markdown 文件。

2. Download, install and load the swirl package:

```
install.packages("swirl")
library(swirl)
```

Go through the R Programming Basic Building Blocks tutorial and then use the skills you have just learned to answer the following questions.

   a. What version of R are you using (hint: make sure you download the latest version and then type version)?
   b. Create a numeric vector containing the numbers 2.23, 3.45, 1.87, 2.11, 7.33, 18.34, 19.23. What is the average of these numbers?
   c. Use a for loop to determine the value of $\sum_{i=1}^{25} i^2$
   d. The cars dataset is available in base R. You can type cars to see it. Use the class function to determine what type of object is cars.
   e. How many rows does the cars object have?
   f. What is the name of the second column of cars?
   g. The simplest way to extract the columns of a matrix or data.frame is using [. For example you can access the second column with cars[,2]. What is the average distance traveled in this dataset?
   h. Familiarize yourself with the which function. What row of cars has a distance of 85?

3. (Irizarry RA, Love MI (2015): Data Analysis for the Life Sciences, page 10 Exercises) Here we will test some of the basics of R data manipulation which you should know or should have learned.
   a. Start a project in RStudio.

b. Download the file `femaleMiceWeights.csv` from our course webpage

   http://ghuang.stat.nctu.edu.tw/course/statmethods16/files/data/femaleMiceWeights.csv

   into your created project.

c. Read in the file `femaleMiceWeights.csv` and report the exact name of the column containing the weights.

d. The `[` and `]` symbols can be used to extract specific rows and specific columns of the table. What is the entry in the 12th row and second column?

e. You should have learned how to use the `$` character to extract a column from a table and return it as a vector. Use `$` to extract the weight column and report the weight of the mouse in the 11th row.

f. The `length` function returns the number of elements in a vector. How many mice are included in our dataset?

g. To create a vector with the numbers 3 to 7, we can use `seq(3, 7)` or, because they are consecutive, `3:7`. View the data and determine what rows are associated with the high fat or `hf` diet. Then use the `mean` function to compute the average weight of these mice.

h. One of the functions we will be using often is `sample`. Read the help file for sample using `?sample`. Now take a random sample of size 1 from the numbers 13 to 24 and report back the weight of the mouse represented by that row. Make sure to type `set.seed(1)` to ensure that everybody gets the same answer.

4. (Irizarry RA, Love MI (2015): Data Analysis for the Life Sciences, page 13 Exercises) For these exercises, we will use a new dataset related to mammalian sleep. This data is described here: http://docs.ggplot2.org/0.9.3.1/msleep.html. We are going to read in this data, then test your knowledge of key `dplyr` functions select and filter.

   a. Start a project in RStudio.

   b. Download the CSV file `msleep_ggplot2.csv` from our course webpage

      http://ghuang.stat.nctu.edu.tw/course/statmethods16/files/data/femaleMiceWeights.csv

      into your created project.

   c. Read in the `msleep_ggplot2.csv` file with the function `read.csv` and use the

function `class` to determine what type of object is returned.

d. Now use the `filter` function to select only the primates. How many animals in the table are primates? Hint: the `nrow` function gives you the number of rows of a data frame or matrix.

e. What is the class of the object you obtain after subsetting the table to only include primates?

f. Now use the `select` function to extract the sleep (total) for the primates. What class is this object? Hint: use `%>%` to pipe the results of the `filter` function to `select`.

g. Now we want to calculate the average amount of sleep for primates (the average of the numbers computed above). One challenge is that the `mean` function requires a vector so, if we simply apply it to the output above, we get an error. Look at the help file for `unlist` and use it to compute the desired average.

h. For the last exercise, we could also use the dplyr `summarize` function. You can read the help file and repeat exercise g, this time using just `filter` and `summarize` to get the answer.