Introduction to R

Session 2 – Data subsetting

Statistical Consulting Centre

19 July, 2017

1. Installing an R package

R packages are collections of user-defined functions. The function `std.error`, for example, is contained in the `plotrix` package.

1. Let's look at what happens when we try to use a function before actually installing on our computer the package in which it is contained. E.g. Calculate the SEM of age using `std.error`.
2. Install the package `plotrix` while in your R session by following the instructions below:
   (a) Select `Packages` from the bottom right panel of your Rstudio interface.
   (b) Click on the `Install Packages` icon just below `Packages`.
   (c) Type `plotrix` in the blank space provided below "`Packages (separate multiple with space or comma):`"
   (d) Click on `No` if you are asked you to restart `R`.
   (e) Submit the code `library(plotrix)` to the `R` console to make the functions contained in `plotrix` available in the current `R` session.
3. Now, use `std.error` to calculate the standard error of the pH.
4. Try writing your own code to calculate the standard error of the pH. Hint: This only requires one line of code. Use online resources if you cannot remember how the SEM is calculated.

2. Write your own function

1. In Session 2 you were shown a simple function to calculate the standard error of the mean (SEM), i.e.

```
mystder <- function(x){
     mysd <- sd(x, na.rm = TRUE)
     n <- length(x)
     mysd/sqrt(n)
}
```

Type the above code into your `R` script and submit it to the `R` console.

2. Modify the function in 2.1 so that the output will have only 2 decimal places.
3. Calculate the SEM of `pH` using the function you created in 2.2.

2. Subsetting datasets

1. Print the following to the console:

- The pH of the first lake.

- The pH of the last lake.
- The pH values of the first and last lakes.
- All measurements made on the third lake.
- All pH values.

2. Calculate:

- The average pH of lakes with low Calcium concentrations.
- The average pH of lakes with low Calcium concentrations and Chlorophyll concentrations lower than 10.

4. Challenge

Modify the function given in 2., so that the function will return a 95% confidence interval (with 2 decimal places). Hint: A 95% confidence interval of a variable `x` is given by the mean of `x` $\pm$ 1.96 $\times$ SEM of `x`. You might find the `paste()` function useful.