

Introduction to R

Session 1 – Introduction

Statistical Consulting Centre

19 July, 2017

1. Using R as a calculator

- Find the values of:
 - $1 + 4$
 - $2^3 + \frac{4}{\sqrt{34}}$
 - $\log 30$
 - $\log_{10} 30$
 - $|-2|$ (Hint: $|x|$ denotes the *absolute value* of x . Search on Google if you're unsure.)
- Now open Rstudio, open a R script clicking **File** → **New** → **R script**.
- Save this script by clicking **File** → **Save As...**
- Select a directory/location and save the script. Note: the saved script should have **.r** as extension. For example, if you call your file **exercise one**, then you should save it as **exercise one.r**
- Copy and paste the code you typed (*not the output, not the > symbol, just the code you typed*) at the console for into the R script opened in Rstudio.
- Submit your entire script at once to the R Console by highlighting all codes and pressing **Ctrl + R**.
- From now on, type all of your code in your R script and submit it to the R Console using **Ctrl + R**.

2. Reading data into R

- lake.csv** contains data on mercury contamination in 53 different lakes in Florida. The variable names and what has been measured are presented below.
 - ID**: ID number of the lake
 - Lake**: Name of the lake
 - pH**: pH value
 - Calcium**: concentration of Calcium
 - Chlorophyll**: concentration of Chlorophyll (mg/L)
- Read the data into R, saving it in object named **lake.df**.
- Use **dim()** and **head()** to look at some of the properties of the dataset you have just read into R. *Always* perform this important step to check that your dataset is as it should be.
- Calculate the mean and standard deviation of both **pH** and **Chlorophyll**.
- Check out what **summary()** does by running **summary(lake.df\$pH)**.
- Check the frequency of **Calcium** concentration
- Turn the frequency table in 2.6 into proportion, keep only 2 decimal place