

Introduction to R

Session 2 – Subsetting data

Statistical Consulting Centre

consulting@stat.auckland.ac.nz
The Department of Statistics
The University of Auckland

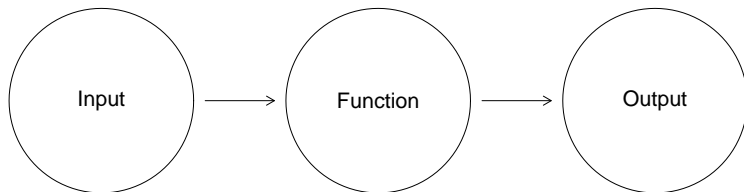
19 July, 2017



SCIENCE
DEPARTMENT OF STATISTICS

Functions

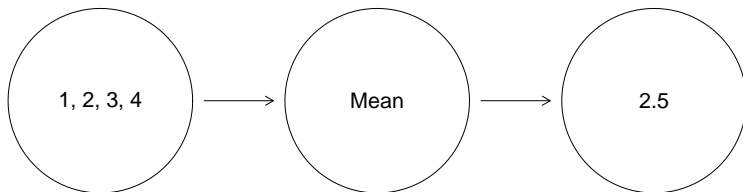
A function is a relationship between a set of inputs (arguments) and a set of outputs. E.g., the function is fed some information on which it operates, the results of which are the output.



This is an essential building block for the R package.

Functions

We have seen many functions, e.g. `log`, `mean`, `table`, `with`, etc.



Working with functions

- Functions can be user-defined, i.e., you can write your own.
- Output is the last line of the function. You can use `return()` to specify the output.
- Here is a function calculates the standard error of the mean (SEM).

```
mystder <- function(x) {  
  mysd <- sd(x, na.rm = TRUE)  # Calc std. deviation  
  n <- length(x)  # Calc sample size  
  mysd/sqrt(n)  # Definition of SEM  
}  
mystder(Growth.df$biomass)
```

```
## [1] 0.3651815
```

- A set of user-defined functions can be bundled together into an R package.

Getting data into R

- Base R includes only functions which read data sets saved in simple file formats, e.g. csv, txt, tab delimited, etc.
- What if your data was saved in another format, e.g. STATA, SPSS, or SAS spreadsheets?
- The haven package for R contains functions that may help <https://cran.r-project.org/web/packages/haven/index.html>

```
> library(haven)
> stata <- read_dta("data.dta")
> spss <- read_sav("data.sav")
> sas <- read_sas("data.sas7bdat")
> sasxport <- read_xpt("data.xpt")
```

However, it is always the easiest and safest to read data into R from a csv file.

Packages

- Currently, the CRAN package repository features 10,098 available packages (4 Jul. 2017). There are about 13,169 CRAN, BioConductor and Github packages in total.
- To install packages from the R GUI, click on Packages → Install Package(s) ... → New Zealand (or whatever region you are located) → Package name
- Or, you can type `install.packages(package name)`, e.g. `install.packages("haven")`.
- After the installation, use `library("package name")` to load it into R.

Note: Installation is performed only once; however, it must be loaded (i.e. use the command `library("package name")`) in every R session.