

Statistics

Week 3 Recitation

ESD, SUTD

Term 5, 2017

Introduction to R

What is R?

- R is a language and environment for statistical computing and graphics

Why use R?

- Free - available to everyone
- Widely used (~ 2 million users worldwide) - active community
- Open-sourced - new features and packages are being developed

Introduction to R

What is RStudio?

- RStudio is an integrated development environment (IDE) for R.

Why use RStudio?

- User friendly, making developing R programs more easily and efficiently

You can use R without RStudio, but can't use RStudio without R.

RStudio Working Environment

Key concepts:

- Console - where commands are interpreted and executed
- R script - where commands are documented
- Working directory - a directory in your computer that you can save files to

To do:

- Create a new R script
- Set working directory
- Save R script to the working directory

Variables and Functions

Key concepts:

- Variable - where values are stored
- Function - takes in values/variables as input, and returns output after calculation (similar to functions in Excel)

To do:

- Create a few variables, assign values to them
- Calculate the square roots for the positive variables; and absolute values for the negative variables
- Assign the results to new variables

Example code

```
a <- 5  
b = -3.34  
aSquareRoot = sqrt(a)  
bAbs = abs(b)
```

Vector and Data Frame

Key concepts:

- Vector - a series of numbers/characters stored as the same object
- Data Frame - multiple vectors stored as one object

To do:

- Create two vectors
- Combine them into one data frame

Example code

```
pillar = c("ASD","EPD","ESD","ISTD")  
population = c(83,119,43,100)  
cohort2015 = data.frame(pillar,population)
```

Basic Data Analysis

Excercise: speed of light data

To do:

- Import data from the csv file *w3 light R.csv*
- Calculate the mean and sample standard deviation of the light data

Command outline

```
# loading csv file
dataLight = read.csv("w3 light R.csv")

# basic statistical analysis
mean(x)
sd(x)
```

Plots

Excercise: speed of light data (continued)

To do:

- Draw boxplot, histogram, and Q-Q plot using the light data
- Get rid of the outliers and redo the analysis and plots

Command outline

```
# plotting
boxplot(x, main, ylab)
hist(x, xlim, breaks)
qqnorm(x)
qqline(x)

# subset data
subset(dataLight, TimeD > 0)
```


Excercise: triple jump data

To do:

- Import data from the csv file *w3 triple jump R.csv*
- Plot the distances over the years

Command

```
plot(x)
```

Summary

After today's recitation, you should be able to work with R to...

- Manually create variables
- Import data from csv files
- Run basic statistical analysis and calculations
- Draw boxplot, histogram and Q-Q plot

A few helpful websites:

- Official Page: www.r-project.org
- Quick R: www.statmethods.net
- R Resources: www.ats.ucla.edu/stat/r/