Biostatistics I: Introduction to R

Introduction

Eleni-Rosalina Andrinopoulou

Department of Biostatistics, Erasmus Medical Center

e.andrinopoulou@erasmusmc.nl

@erandrinopoulou

Erasmus MC
University Medical Center Rotterdam

Introduction to R

- **R** is a great tool to explore and investigate the data
- Several statistical methods can be performed with **R**
- It is important to understand the methods before applying them in **R**

How to use

R uses packages that perform specific tasks

- Install package only once
- Load package every time you open **R**

Introduction to R

- For this course: Rstudio (http://www.rstudio.org/)
  - free
  - works fine in Windows, MacOS and Linux
  - helpful with errors
  - alternative output options

Introduction to R

Basic functions

- `getwd()`, `setwd()`,
- `is.na()`, `is.finite()`, `is.null()`

Import/Export

- `read.csv()`, `write.csv()`
- `read.xlsx()`, `write.xlsx()`
- `read.table()`, `write.table()`

Save/Load

- `save()`, `saveRDS()`
- `load()`, `readRDS()`

Data Types/Structures

The simplest data types are:

- **numeric** : quantitative data
- **character** : qualitative data
- **integer** : whole numbers
- **logical** : TRUE or FALSE
- **factors** : qualitative data (levels)

Data Types/Structures

The most important data structures are:

- **Scalar** a single element
- **Vectors** have the same type of elements
- **Matrices** have the same type of elements with the same length
- **Arrays** have the same type of elements with the same length but can store the data in more than two dimensions
- **Data frames** have elements of different type with the same length
- **Lists** have elements of different type and length

Data Types/Structures

Data types

- `is.numeric()` / `as.numeric()`
- `is.character()` / `as.character()`
- `is.integer()` / `as.integer()`
- `is.logical` / `as.logical()`
- `is.factor()` / `as.factor()`
- `str()`, `mode()`

Data structures

- `c()`
- `matrix()`
- `array()`
- `data.frame()`
- `list()`

Other

- `ls()`, `objects()`

Indexing/Subsetting

- This can be done using square bracket (**[ ]**) notation and indices.
- Three basic types
  - position indexing
  - logical indexing
  - name indexing

Indexing/Subsetting

Vectors

- `[]`
- `[""]` - for categorical variables

Matrices

- `[,]`
- `[[]]`, `[]`

Arrays

- `[ , , ]`

Data frames

- `[,]`
- `[[]]`, `[]`
- `$`

Lists

- `[]`
- `[[]]`
- `$`

Data Transformation/Exploration/Visualization

Transformation

- `round()`
- `factor()`
- `order()`
- `reshape()`

Exploration

- `mean()`, `sd()`
- `median()`, `IQR()`
- `table()`

Visualization

- `plot()`, `legend()`
- `hist()`
- `barchart()`
- `boxplot()`
- `xyplot()`, `ggplot()`
- `par()`

Correlation

Pearson correlation

- magnitude of association
- linear association
- direction of the relationship

A relationship is linear when a change in one variable is associated with a proportional change in the other variable

Correlation

Spearman correlation

- direction of the relationship
- monotonic relationship

In a monotonic relationship, the variables tend to change together, but not always at a constant rate (as in the linear case)

Test hypothesis

- **parametric** (assumptions about the distribution) / **non-parametric** (distribution-free)
- **one sample** / **two samples** / .. / M **samples**
  - compare one group with a value
  - compare two groups paired / unpaired
  - compare M groups
- **one-sided (one-tailed)** / **two-sided (two-tailed)**

$$H_0 : \theta = \theta_0$$

$H_1 : \theta \neq \theta_0$ (two-sided)
$H_1 : \theta > \theta_0$ (one-sided)
$H_1 : \theta < \theta_0$ (one-sided)

Test hypothesis

- Choose a null hypothesis H_0 and an alternative hypothesis H_1
- Collect and visualize the data
- Choose and calculate the test statistic, which is a numerical summary of the data
- Determine the sampling distribution under the condition that the null-hypothesis holds
- Choose the type I error (significant level) α, usually α=0.05
- Determine the corresponding critical value(s)
- Compare the test statistic with critical value(s) and reject or not