Foundational Statistics

Basics of the R statistical language

What is R?

- A computer programming language and environment
  - Interpreted language
- Designed for data visualization and analysis
- The language S (1976) -> adapted to R (1995)
  - Authors: R Development Core Team
  - Distributor: CRAN (Comprehensive R Archive Network)
  - GNU open-source

Why do so many people use R?

- Good general scripting tool for statistics and mathematics
- Powerful, flexible, and free!
- Runs on all computer platforms
- New packages coming out all the time
- Superb data management & graphics capabilities
- Reproducibility - can keep your scripts to see exactly what was done
- Can embed your R analyses in dynamic, polished files using Rmarkdown
- Write your own functions and packages
- Lots of online help available
- Can use an IDE such as Rstudio

Some Basics

Object R is an object oriented language and everything in R is an object. For example, a single number is an object, a variable is an object, output is an object, a data set is an object that is itself a collection of objects, etc.

Vector A collection of one or more *objects* of the same type (e.g. all numbers or all characters etc).

Function A set of instructions carried out on one or more objects. Functions are typically used to perform specific and common tasks that would otherwise require many instructions. For example, the *function* `mean()` is used to calculate the arithmetic mean of the values in a given *numeric vector*. Functions consist of a name followed by parentheses containing either a set of *parameters* (expressed as *arguments*) or left empty.

Parameter The kind of information that can be passed to a function. For example, the `mean()` *function* declairs a single required parameter (a valid object for which the mean is to be calculated is a compulsary) as well as a number of optional parameters that facilitate finer control over the function.

Argument The specific information passed to a function to determine how the function should perform its task. Arguments are expressions (in the form of `name=value`) given between the parentheses that follow the name of the function. For example, the `mean()` function requires at least one argument - either the name of an object that contains the values from which the mean is to be generated or a vector of values.

Arithmetic Mean

Description

Generic function for the (trimmed) arithmetic mean.

Usage

```
mean(x, ...)

## Default S3 method:
mean(x, trim = 0, na.rm = FALSE, ...)
```

Arguments

x	An R object. Currently there are methods for numeric/logical vectors and date, date-time and time interval objects. Complex vectors are allowed for trim = 0, only.
trim	the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed. Values of trim outside that range are taken as the nearest endpoint.
na.rm	a logical value indicating whether NA values should be stripped before the computation proceeds.
...	further arguments passed to or from other methods.

Some Basics

Operator Is a symbol that has a pre-defined meaning.

Operator	Description
`[ [[`	indexing
`::`	name space
`$`	component
`^`	exponentiation (evaluated right to left)
`- +`	sign (unary)
`:`	sequence
`%special%`	special operators (e.g. `%/%, %%`)
`* \`	multiplication, division
`+ -`	addition and subtraction
`< > <= >= == !=`	ordering and comparison
`!`	logical negation (not)
`& &&`	logical AND
`| ||`	logical OR
`~`	formula
`-> ->>`	assignment (left to right)
`=`	argument assignment (right to left)
`<- <<-`	assignment (right to left)
`?`	help

RStudio - Popular IDE