

R

a complete and free data analysis package

- R is free software -- download from <http://cran.r-project.org/>
- R runs on Windows, Linux and Mac systems
- R can exchange data with Excel (via 'csv' [comma-separated variable] format)
- R is actively developed and maintained through open-source activities -- meaning that if you don't like what it does, you can change it yourself

Documentation

- Installed on your computer with R
- On the Web
- Books:
 - Different orientations (statistics, graphics, regression, etc.)
 - Listed on <http://www.r-project.org/>

Applications

- Medical studies (statistics)
- Social sciences (statistics)
- Engineering (mathematics)
- Data modelling in any discipline

Features

- Complex graphical displays
- Programming language
- Symbolic algebra (differentiation)
- Special function evaluation (bye-bye to Abramowitz and Stegun, Bronshtein and Semendyayev, etc.)
- ...and simple calculations

Data types

- Scalars (1, 2, 3, 1.732, $2+3i$)
- Vectors (collections of scalars)
- Matrices (m by n arrays of scalars)
- Functions (for programming)
- Expressions (for symbolic algebra and typesetting)
- Tables (associated vectors of data)

Example demo

- Expressions
- Vectors and operations with vectors
- Typesetting with expressions
- Plotting functions
- Contouring
- (any requests ... ?)

Tables

- Basic data importing structure
- Data located on your computer or on the web (loaded by URL) -- easy sharing!
- Blank- or comma-separated
- Columns of data associated by row
 - Each column can have a header/name
- Data may be numerical or nominal

Table example

- Table on computer
- Table over the Web

Statistics

- Functions operate on vectors (columns in a table)
- `sum(...)`, `mean(...)`, `median(...)`, `sd(...)`, ...
- Significance testing functions (t-test, chi-squared, Q)

Data modelling

- Linear and non-linear models handled
- Easy formula-related syntax
- Straightforward analysis of tabulated data

Linear model

- Linear model formed from constant parameter (to be estimated) that multiplies a known quantity (basis function)
- Example: $y(x) = a + b x$
 - two constant parameters (a and b)
 - known quantities are 1 and x
 - rewrite as $y(x) = a \times 1 + b \times x$
- $\text{lm}(y \sim 1 + x)$ -- straight line fit of y to x

Linear model fit example

- Data from table of time vs voltage
- Read data table and attach to access columns
- Fit using $\text{lm}(MI \sim I + \text{time})$
- Value and significance of each fit parameter calculated

Linear model fit example

- Test significance of one- versus two-parameter model: constant vs sloped line
- Use F-test to compare chi-squared values:
$$F = (X_1^2/DF_1) / (X_2^2/DF_2)$$
- Compare F against critical values
 - Get F comparison function from online R source
 - Conclude that sloped line is better fit

Programming example

- Earthquake location
- Uses station names and locations, P- and S-wave arrival times, and seismic wave speeds to calculate expected arrival time given guessed position
- Iterates to improve location

Another programming example

- Solves sudoku puzzle given initial tableau
- Have a puzzle on hand?