Index to R/RStudio Workbook

20160608:13.47

This is the index to a workbook which consists of a set of modules in medical statistics.

It assumes no knowledge of statistics or programing.

It is designed to guide the reader through all the steps required install a computing environment which will allow them run simple statistical analyses in R and to familiarize themselves with the R programming environment in order to then progress to more complex statistical programming.

Modules

Click in links below:

1. What is R?

The purpose of this module to give you some background information about R, how and why it was developed, and what it is designed to do

2. Installing R and RStudio

The purpose of this module to guide you through the process of installing R and RStudio. Everything that you need is available on line free of charge.

3. Introducing RStudio

This module is designed to provide you with:

- A basic understanding of RStudio as an Integrated Development Environment and its key features
- An understanding of the four panel layout of RStudio and the functions of each of the panels
- The ability to use simple arithmetic, algebraic and logical operators in R
- The ability to load in a CSV file from a web based address using the Import wizard
- The ability to set the Working Directory
- An understanding of .R and .Rdata file types
- The ability to load and run R scripts

4. Creating Vectors and working with them

This module is designed to enable you to:

- Create, write, name and save R scripts
- Understand when and how to use the colon operator, and 'combine' c() and 'sequence' seq() functions
- Understand what is a vector and how vectors can be used
- Know how to find 'help' using the various inbuilt R and RStudio mechanisms
- Understand 'command completion' and how to use it
- Have a basic understanding of how the data in a vector can be viewed

5. Data.frames, indexing and sub-setting

On completion of this module you will:

- Be able to write the contents of a data.frame out to a text file
- Be able to import a local text file into R to create a data.frame
- Be familiar with the functions names() and str() and will know how and when to use them
- Know how to rename the columns in a data.frame
- Be able to demonstrate an understanding of the row / column table structure of data frames and the essential differences between rows and columns
- Understand and be able to use $ notation for data.frame column names
- Understand and be able to use simple indexing as it applies to vectors and also single data.frame columns
- Understand and be able to use simple row / column indexing on data.frames
- Understand and be able to build more detailed row expressions to select subsets of population with particular characteristics
- Know how to add columns to a data.frame
- Know how to remove columns from a data.frame
- Know how to re-arrange the order of columns in a data-frame

6. Data.frames: Checking out the data

On completion of this module you will:

- Be able to demonstrate an understanding of the importance of visualising your data sample to check whether or not it is representative of the population(s) that you wish to study
- Be able to demonstrate an understanding of the importance of finding out what characteristics the variables in the data sample are supposed to represent and how they have been derived
- Know how to view the data in any data.frame
- Understand and be able to demonstrate that you know how to use the summary function
- Understand and be able to demonstrate that you know how to use the basic table function
- Understand the implications of NAs when using the table function and demonstrate the ability to handle these
- Be able to set up age bands on any data.frame that has age and gender columns
- Be able to produce age / sex profiles on any data.frame, or any subset of a data.frame, that has age and gender columns
- Have developed further expertise in using row / column indexing
- Know how to use the nrow() function to count rows in a data.frame and how to use that function to keep a tally of rows that have been selected or deselected
- Be able to demonstrate an understanding of what is meant by 'data cleaning'
- Know at least two different ways of handling extreme values and NAs when cleaning numeric data
- Understand and be able to use logical operators to combine expressions to clean data
- Understand and know how to use the is.na() function
- Know how to set up simple histograms using the hist() function and where to go to find further details to refine them