WiFi

- Connect to UT-Open Network
- Open browser window
- Choose Unsponsored Guest
- Enter an email address
- Click Submit

Introduction to Statistics Using R

Presented by
Daniel Walker
Tennessee AFS Chapter Meeting
28 February 2017

Introduction

- R: the software and language
- Why should you use R?
- Download and Install R and RStudio
- Download datasets, code sheets
- Coding in R

What is R?

- R is a language and environment for statistical computing and graphics
- Contains a wide variety of analytical and graphical procedures, and is highly extensible
- Freely available software distributed under the Free Software Foundation General Public License
- Developed and distributed by the R Foundation, an international non-profit
- An implementation of the S statistical programming language developed in the 1970's

Why should you use R?

- It's entirely free, unlike SAS, SPSS, NCSS, etc.
- It 'ships' with a variety of functions in bundles called packages, with flexibility for user-defined functions and downloadable packages, unlike Excel
- It has sophisticated graphics capabilities in the default version, with even better graphics packages available – all with low-level control
- It is an environment, allowing users to complete all stages of an analysis from data reading and clean up to report generation

Why should you use R?

- Required coding encourages extensive documentation – saved code in documents = bread crumb trail of analysis = reproducibility
- The R language is object-oriented: save objects by name, access and manipulate objects with procedures (unlike logic-heavy procedural languages like C, C++, etc) = more intuitive, human-readable code
- Abundant references available at low or no cost – the chances are good that someone else has had a similar problem to you, and had their question resolved in a reputable internet forum

Downloading and Installing R and RStudio

- Nearly all R downloads you'll need will come from CRAN – the Comprehensive R Archive Network
- https://cran.r-project.org/mirrors.html (or Google 'R download')

- RStudio – an integrated development environment for R (an ancillary program that makes using R easier)
- https://www.rstudio.com/products/rstudio/download/ (or Google 'RStudio download')

Course Materials

- Go to http://jalfor12.wixsite.com/alfordlab/links
- Download each of the files
- Store all of the files in a single folder for this class

R documents

- .R filetype – an R specific plain text document
- Load into R, run code from it
- Key to generating documentation of analyses, preserving analyses for reproducibility
- RMarkdown – a markdown package for R, more features but some learning curve

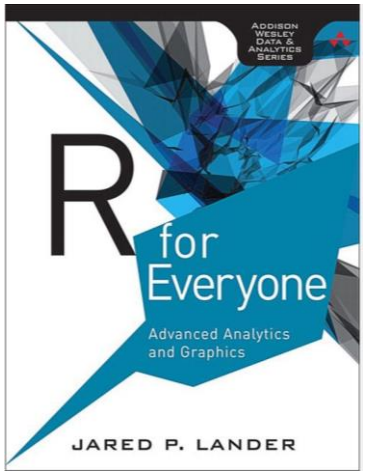
Switch to RStudio

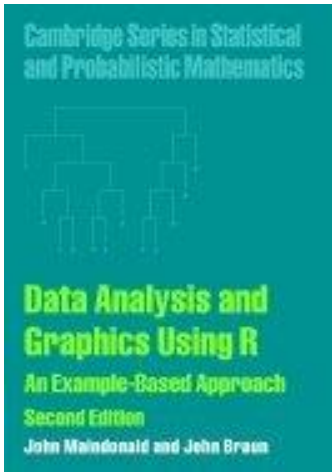
Additional Resources

- Google: " 'function/operation' and/or 'error message' r "
- www.r-tutor.com
- www.statmethods.net
- www.stats.stackexchange.com
- http://derekogle.com/fishR/courses

Additional Resources

- 'R for Everyone' first ed. - Lander 2013, $25 on Amazon. Second ed. releasing April, $40
- 'Data Analysis and Graphics Using R' - Maindonald and Braun 2003, hardcover, ~$100 on Amazon





Summary of Training

- R is free, flexible, and capable
- New processes on the cutting edge of statistical research are constantly deployed in packages
- The IDE and coding will let you develop your own unique data management and analysis pathways, streamlining your data handling

Final Quiz

1. Start with an empty environment, new R session
2. Set working directory
3. Load the channel catfish length-weight-age sheet from 'RforFishSci.xlsx' as a dataframe named 'catfish'
4. Load the crappie TL sheet from 'RforFishSci.xlsx' as a separate dataframe named 'crappie'
5. Load the spot length-at-age data from the 'spot.csv' file as a dataframe named 'spot'
6. Generate a length-frequency histogram from the crappie TL data using the default settings
7. Make a new length-frequency histogram by changing at least two plot settings

Final Quiz

8. Calculate the ratio of ideal to actual weights for the channel catfish
9. Classify whether each catfish's weight ratio is greater or less than 100
10. Identify the proportion of channel catfish who exceed a weight ratio of 100
11. Calculate the average relative weight for the catfish population
12. Export channel catfish data frame with intermediate steps to new csv file called 'catfish_new'
13. Solve an original VB function for the spot length at age data
14. Plot the spot length at age data and VB growth curve

Comments, Questions, Assistance

- dwalke44@vols.utk.edu – current address
- walker.dan890@gmail.com – permanent address