Statistics with R

Introduction

Zhuanghua Shi (Strongway)

16 April 2018

Challenges in Research

- Psychological research requires
  - theory development
  - Hypothesis testing
- Empirical data analyses play key role in bridging the gap
  - Statistics knowledge
  - Statistics software (SPSS, Matlab, Excel, R)

Why R?

Why R?

- Open source
- Large community
- ...

Most important features

R and RStudio constitute a powerful tool for reproducible research (RR)

Reproducible Research (RR)

- Research process is not linear, rather often redo some steps.
  - rerun experiments
  - reanalyze data

An ideal RR process should be able to adapt to flexible research process:

- easy to expand experiments and data analyses
- easy to maintain codes and text
- easy to replot figures and to output statistics

Solution: Git, R, and Markdown

- git for version controls (e.g., multiple experiments, minor variations)
- R for statistics
- RMarkdown for writing

RStudio integrates all of them seaminglessly. This course we will focus on *R for Statistics*.

Aims of the course

- After this course, you should be able to
  - Exploratory data analysis (EDA)
  - Import/Export various types of data
  - Basic statistics (ANOVA etc.)
  - Linear Regression
  - Multiple and logistic regression
- Lecturer and Tutors
  - Strongway (me)
  - Jan Nasenmann
  - Artyom Zinchenko

Schedule

dates	contents
2018-04-16	Introduction
2018-04-23	Data Visualiation
2018-04-30	Practice 1
2018-05-07	— (Online assignment)
2018-05-14	Data Wrangling
2018-05-21	— (Online assignment)
2018-05-28	Hypothesis Testing
2018-06-04	Practice 2
2018-06-11	Linear Regression
2018-06-18	Practice 3
2018-06-25	Multiple and Logistic Regression
2018-07-02	Practice 4

Online learning materials and sources

- The online book An Introduction to Statistical and Data Sciences via R

This book will be our reference book of the course.

- Datacamp

We have a learning group in datacamp - Statistics with R. Please join in this group via the email I sent. Datacamp has huge learning materials, providing not only R but also Python and others. Please use this opportunity to learn data science.

Start with R and RStudio

- R, RStudio, and Packages
  - Please read the chapter 2
- Get familiar with RStudio Inferface

Using R project

- R projects
  - make it straightforward to manage your files
  - are platform independent
  - have a root folder
  - Version control
  - .Rproj extension and one hiden folder
  - Store your history .Rhistory

Using R projects make you easy to run your codes across platforms and collaborate with colleagues.

Data Import/Export

- We encounter many types of data formats
  - csv, excel, matlab, spss, web table, Sql
- R {utils} provides basic import methods
  - read.csv
  - read.delim
  - read.table
- data.table package provides powerful function
  - fread()
- readr package (fast)
  - read_csv
  - read_delim
  - read_tsv

import text file basic functions

read.table(filename, header=TRUE, sep=',',