Statistics with R

Introduction

Zhuanghua Shi (Strongway)

8 May 2019

- 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
2019-05-08	Introduction
2019-05-15	Data Visualization
2019-05-22	Practice 1
2019-05-29	Data Wrangling
2019-06-05	Practice 2
2019-06-12	Hypothesis Testing
2019-06-19	Bayes Analysis
2019-06-26	Practice 3
2019-07-03	Linear Regression
2019-07-10	Generalized Linear models and Logistic Regression
2019-07-17	Practice 4
2019-07-24	-

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

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=',',