Stanford Psychology

Statistics and Programming Bootcamp!

Instructors: Shao Fang (Pam) Wang, Arielle Keller, & Leili Mortazavi

Emails: shaofang@stanford.edu, askeller@stanford.edu & leilimor@stanford.edu

Overview

In this two-day bootcamp, you will learn some basic coding skills, and practice these coding skills as we review some introductory statistical methods. You'll also learn some complementary skills to improve your computing workflow with Bash and Github. We'll also cover some of the many facets of open science and talk through some good coding practices. By the end of this bootcamp, you'll be well on your way to using these skills for your Psychology research!

Learning Goals

There are three main goals that we hope you'll achieve in this bootcamp:

1. Basic programming skills
   We will prioritize learning R, but will also provide some Python resources.
2. Ability to use programming skills to implement introductory statistical methods
   Including t-tests, correlations, linear/logistic regression, and test-set validation
3. Understanding of clear computing workflows
   Including basic skills with Bash and Github, and knowledge of open science and good coding practices.

Schedule

Day 1 (Sept. 10th)

9:00am: Welcome! Introductions.
9:30am: Basic R **OR** Basic Python (concurrent)
12pm: Lunch break
1pm: Introduction to Bash
2pm: Introduction to Github

Day 2 (Sept. 11th)

9:00am: Group activity
9:30am: Statistics with R / Python
12pm: Lunch break
1pm: Introduction to Open Science
2pm: Introduction to Good Coding Practices (Russ Poldrack)

A note about R vs. Python for this course

R has been specifically designed for statistical operations, whereas Python is a more general-purpose language and thus is powerful and very flexible. Though we recommend using R for statistics and data visualization, learning Python can be very useful for other components of your research. Therefore, for those already familiar and comfortable with R, we're using the same concepts and goals as an opportunity to introduce you to Python.

Modules (and table of contents)

Basic R

- Introduction to R syntax
    - Variables
    - Data types
    - Operators
    - Functions
    - Vectors
    - Indexing
- Data wrangling
    - Dataframes
    - Importing and exporting data
    - Inspecting data
    - Manipulating data with *tidyverse*
    - Tidy data
- Data visualization
    - Introduction to *ggplot*

Basic Python

- Introduction to Python

    - Variables (module 01-01)
    - Control structures (module 01-02)
- Data wrangling (module 01-03)
    - Dataframes
    - Importing and exporting data
    - Inspecting data
    - Manipulating data with *pandas*
- Data visualization (module 01-04)
    - Introduction to *seaborn*

Statistical concepts and implementation with R / Python

- Data-wrangling review + T-tests
- Correlation + linear regression
- Logistic regression + cross validation
- Practice with an example dataset (with R / Python)

Bash

- Introduction to Bash
- Creating, moving, and destroying directories
- Manipulating files
- Scripts
- Connecting to remote servers

Github

- Introduction to *git*
- Github repositories
- Cloning repositories
- Add/commit/push changes
- Rescuing older versions

Open Science

- What is “open science”?
- Pre-registration
- Reproducibility
- Sharing data and code

Good coding practices

Honor code

https://communitystandards.stanford.edu/

Additional Resources

https://docs.google.com/document/d/1V5OJSOnIERMRsjoNhabToG1nndMdmQwfbGcO8l-_T3g/edit

Contributors

Many people have contributed to developing and revising the tutorial material over the years:

Anna Khazenzon
Cayce Hook
Paul Thibodeau
Mike Frank
Benoit Monin
Ewart Thomas
Michael Waskom
Steph Gagnon
Dan Birman
Natalia Velez,
Kara Weisman
Andrew Lampinen
Joshua Morris
Yochai Shavit
Jackie Schwartz
Russ Poldrack