Sociology Methods Camp Syllabus

Janet Xu, Xinyi Duan, and Matt Salganik

Fall 2017

Logistics

Dates of camp: Tuesday September 5th to Friday September 8th

Time structure of camp:

Breakfast: *9:00 AM*

Morning session, math: *9:30 - 11:30 AM*

Lunch session (qualitative methods): *12:00 - 1:30 PM*

Afternoon session, programming : *2:00 - 4:00 PM*

Location: 165 Wallace Hall

Office hours: 4:30-6:30 PM every day, 165 Wallace Hall (*Note*: in the past, students have found it helpful to use office hours as a time to just work on the homework together in one place)

Materials for over the summer and during the camp

- Gill, Jeff. *Essential Mathematics for Political and Social Research.* Cambridge University Press. 2006. ISBN: 978-0-521–68403-3. Can buy for <$50 on Amazon (please purchase over the summer for the pre-camp review materials. The summer assignment will adhere pretty closely to the structure of the textbook): Gill on Amazon

- DataCamp for learning R

- Bring computer to camp with R and Rstudio installed

Assignments during on-campus portion

Programming is learned best through practice rather than through passive listening, so each day we're going to have a short activity/assignment where you will apply the lessons learned in that morning's programming lecture. The assignment will be due at **9 AM the following day (right before next day's programming lecture) on Blackboard**.

For working on the assignment in and out of class, we'll randomly draw pairs at the end of each programming lecture and your pair will turn in one copy as a group. You'll have some time at the end of each programming lecture to work on the activity but may also need to spend some time on it out of class.

Because the focus will be on these worksheets, there are no required DataCamp modules– instead, we'll list the relevant DataCamp modules as optional references if you want more practice at another time. Likewise, the listed Gill readings can be used as references but aren't required before camp.

We'll be available in office hours and on Piazza to offer help on the assignment.

Schedule for on-campus portion of camp

Day 1

Math: Calculus and its Applications

- Review of summer assignment topics: chain rule, derivatives of logs/exponents

- Two tools useful for optimization: higher-order derivatives and partial derivatives

- Preview simple case of univariate optimization

- **Reading**: Gill Chapter 6 and Chapter 5 for those who are newer to calculus

- **Selected online resources** (these are all just suggestions, but ones we've found helpful!):

    - Vector derivatives and integrals: Vector differentiation video
    - Chain rule: Chain rule problems + solutions

Lunchtime qualitative speaker: Matthew Desmond

Computing: Review of Basics

- Summer review, indexing and manipulation of four main data structures: vectors, lists, matrices, and data.frames

- Three useful tools for data manipulation: logical statements, control flow, for loops

- Dplyr as a tool for data manipulation

- **Selected online resources**:

    - How to get help (you will thank the gods of peer-to-peer online communities many times for this site!): stackoverflow
    - Loops: MIT tutorial on writing loops in R

- **Relevant DataCamp modules**: Intermediate R

Day 2

Math: Matrix Algebra

- Vectors -> matrices

- Addition, subtraction, multiplication

- Linear independence

- Rank, inverse of a matrix

- Matrix applications: distance measures, dimension reduction, using matrices to solve systems of linear equations

- **Reading**: Gill Chapters 3-4

- **Selected online resources**:

– Online tool where you can practice one way of finding the rank of a matrix (transforming a matrix into reduced row echelon form): rref practice
– Good general guide to linear algebra: Intuitive linear algebra

Lunchtime qualitative speaker: Fred Wherry

Computing: Functions and the Apply family

- Functions
- Apply family

Day 3

Math: Basic Probability

- Counting, sets, methods of counting and sampling

- Conditional Probability and Baye's Rule

- Independence

- **Reading**: Gill chapter 7

- **Selected online resources**:

  – Some applications of bayes rule/conditional probability: Bayes examples
  – Penn state counting techniques (helpful discussion of permutations and combinations): PSU STAT 414

Lunchtime qualitative speaker: Patricia Fernandez-Kelly

Computing: Data Cleaning and more Advanced Manipulation

- Follow up from Day 2's with math lecture: matrix algebra in R
- Reading data from different file formats
- Merging data
- Reshaping data between long and wide format
- Basic string operations for renaming and recoding variables

Day 4

Math: Basic Optimization; Open-ended Q & A

- Numerical optimization: Newton-Raphson

- Maxima and minima: critical points, Hessians

- Application: Likelihood Functions

- (if there is time) Optimization beyond univariate case: multivariable function

- (if there is time) Optimization beyond univariate case: quadratic form notation

- Open-ended Question and Answer session about material covered this week

- **Reading**: Gill 6.4-6.9

- **Selected online resources**:

  - Khan academy series of vidoes/exercise on Taylor series approximations Khan Academy Taylor series
  - Short preview of how you'll use Taylor series approximations in statistics for calculating the standard error of transformed parameters UCLA ATS delta method

Lunchtime qualitative speaker: Bob Wuthnow

Computing: Plotting and Review of Day 1 material; Open-ended Q & A

- Preparing data to plot
- Plotting using ggplot2
- Open-ended Question and Answer session about material covered this week