R for Social Statistics Workshop
Overview & Outline

Workshop Description

This one-day workshop will provide a friendly introduction to the statistical data analysis program, R. We will cover the basic structure and functions of R; data entry and management in R; and ways to describe, visualize, and analyze data in R. Please note that the course is intended for persons with some statistical background. Registrants should have a working knowledge of hypothesis testing, correlation, and regression, as these topics will be included in the workshop.

R (cran.r-project.org) is a free computing language and software environment that provides access to a variety of statistical and graphic techniques. Due to its open source structure and flexibility, R has been growing in popularity since its official release in 2000, making it one of the most widely used statistical software programs both inside and outside academia. Although many software programs and coding languages for data analysis exist, R is one of the few software environments directly oriented towards statistical analyses, which is continually expanding through the implementation of user-created packages. In addition, a large active community of programmers, statisticians, and researchers continue to strengthen the program and the resources available.

This workshop is made possible through the University of Alberta Population Research Laboratory (PRL) and the Department of Sociology.

Instructor Information

Instructor: Dr. Michelle Lee Maroto
Email: maroto@ualberta.ca
Phone: (780) · 492 · 0478
Office: 6-23 Tory Building
Website: michellemaroto.com

Dr. Maroto is an Assistant Professor of Sociology at the University of Alberta whose research interests include stratification, wealth inequality, social policy, and quantitative methodology. Michelle regularly uses R as the primary software for her research and she teaches students to use the program in both her introductory and more advanced statistics and data analysis courses.

Workshop Topics & Outline

Part 1: R Basics

Part 1 focuses on "getting to know R." We begin with an overview of the software for R and RStudio, along with the associated packages and libraries. We then take some time to dig into the R language with a focus on working with variables, vectors, matrices, and array. Finally, we discuss data entry and management with examples for reading data into R and subsequently subsetting, cleaning, and recoding missing data.

- 1A: R overview
  - R and RStudio
  - R packages and libraries
  - R as an overgrown calculator
- 1B: R language and programming
  - R language components
  - Variables and vectors
  - Matrices and arrays
  - Factors
  - Indexing and conditional selection
  - Data frames
  - Loops and sorting
  - The graphics system
- 1C: Data entry and management
  - Reading data into R
  - Working with data
  - Selecting and subsetting
  - Cleaning data
  - Missing data

Part 2: Describing Data in R

Part 2 focuses on the many options for describing data in R. In addition to reviewing functions for calculating measures of central tendency and dispersion, we also discuss how to apply survey weights to produce weighted estimates for complex survey data. The section then ends with a description of basic and more advanced plotting options in R.

- 2A: Descriptive statistics
  - Recoding variables
  - Frequency distributions and bivariate tables
  - Measures of central tendency and dispersion
    - Mean, median, and mode
    - Range, interquartile range, variance, and standard deviation
- 2B: Descriptive statistics for complex survey data
  - Using the survey( ) package
  - Frequency distributions and bivariate tables
  - Measures of central tendency and dispersion
- 2C: R graphics
  - Basic plots
    - Pie charts, bar charts, histograms, scatterplots
  - More advanced plotting options

Part 3: Inferential Statistics in R

Part 3 focuses on measures of association and inferential statistics in R. We begin with a discussion of hypothesis testing in different situations and ANOVA. We then review aspects of regression and correlation. This section concludes with an introduction to R libraries and functions for more advanced statistical methods.

- 3A: Hypothesis testing and ANOVA
    - Single and two sample means
    - Single and two sample proportions
    - ANOVA
    - Chi square
- 3B: Regression
    - Correlation
    - Linear regression basics
    - Regression diagnostics
    - Logistic regression
- 3C: More advanced methods
    - Generalized linear models
    - Multilevel models