Intro to R for Survey Researchers

Course Workbook

Welcome to Intro to R for Survey Researchers.

Been wanting to try out R? Here's your chance. During this 4-week class, students will use R to analyze survey data. We will cover descriptives, crosstabs and factor analysis. Click here for course agenda.

R is a hot skill. There's a reason many quant researchers are making an effort to gain this skill: they don't want to get leapfrogged. We make it easy by preparing demonstrations and exercises that are specific to the needs of survey research data analysis. And we even have a little fun along the way.

This training is guided by 3 key principles:

1. **Repetition aids retention**. Therefore, we will do exercises, so students have opportunities to repeat key points from the training materials.

2. **Lectures are boring**. Nobody wants to listen to an instructor for sixty minutes at a time. Therefore, at many points during the course, the instructor will ask questions of the students. Your active participation will optimize the learning experience—and make it more fun.

3. **Debate can be constructive**. If the instructor presents something that you just don't agree with, please say so. Differences of opinion can lead to conversations that everyone will learn from.

We hope you enjoy the course!

Course Outline

I. Day 1
   A. Data Types in R
   B. Basic Functions in R
   C. Manipulating Data in R
   D. Basic Descriptive Statistics
   E. Basic Data Visualizations

II. Day 2
   A. Live Demo of Homework
   B. Exploring Data Visually
   C. Exploring Underlying Assumptions

III. Day 3
   A. Live Demo of Homework
   B. Looking for Correlations in Data
   C. Looking for Regression in Data

IV. Day 4
   A. Live Demo of Homework
   B. Running t-Tests in R
   C. Running an Anova in R
   D. Running Crosstabs in R

Homework Day 1
Due on Class 1

1. Install R-Studio.

- **Recommended Readings:**
    - Take a look at this "R Studio Basics" tutorial:

        http://web.cs.ucla.edu/~gulzar/rstudio/basic-tutorial.html
    - Take a look at the different types of charts in the menu here:

        http://www.statmethods.net/graphs/bar.html

Homework Day 2
Due on Class 2

Prior to the next class upload your answers to the training portal.

1. Identify two variables in the "mock pizza data" csv:
    a. Var1 = Number of times eaten pizza in the past 7 days
    b. Var2 = Overall satisfaction with preferred pizza brand
2. Create a new dataframe based upon subsetting the two variables you identified.
3. Run descriptive statistics on the dataframe.
4. Run a basic pie charts on the scores of each variable.

- **Course Pre-Recorded Video**
    o How to use other packages for data visualization
- **Recommended Readings:**
    o Take a look at "Using the R Interface" here:
    http://www.statmethods.net/interface/index.html

Homework Day 3
Due on Class 3

Prior to the next class upload your answers to the training portal.

1. Using the mock pizza data, graph:
    a. A bar chart of the scores to "Number of times eaten pizza in the past 7 days"
    b. Scatterplots with a trendline
        i. Var1 = Number of times eaten pizza in the past 7 days
        ii. Var2 = Overall satisfaction with preferred pizza brand
    c. Scatterplots with trendline and using different colors for different categories
        i. Var1 = Number of times eaten pizza in the past 7 days
        ii. Var2 = Overall satisfaction with preferred pizza brand
        iii. Category = gender

- **Course Pre-Recorded Video**
    o An illustration of using the lm() function
- **Recommended Readings:**
    o Take a look at a website to help you with regressions here:
      http://www.statmethods.net/stats/regression.html

Homework Day 4
Due on Class 4

Prior to the next class upload your answers to the training portal.

1. Run correlations using three different functions
    a. Var1 = Number of times eaten pizza in the past 7 days
    b. Var2 = Overall satisfaction with preferred pizza brand
2. Run a regression using lm()
    a. IV = Number of times eaten pizza in the past 7 days
    b. IV = Importance of cheese quality
    c. DV = Overall satisfaction with preferred pizza brand

- **Course Pre-Recorded Video**
    o An illustration of using the crosstabs function
- **Recommended Readings:**
    o Take a look at a website to help you with ANOVA here:
       http://www.statmethods.net/stats/anova.html
    o Take a look at a website to help you with Generalized Linear Models (particularly Logistic Regression) here:
       https://www.statmethods.net/advstats/glm.html

Notes

Thank you for attending.

Want to earn hours for your IA PRC status?
Be sure to submit this course to your IA portal.

508.691.6004
Info@ResearchRockstar.com
www.ResearchRockstar.com
Twitter: @ResearchRocks
11 Main Street | Suite 3 | Southborough, MA 01772