R Programming Training

An Introduction for Data Analysis and Graphics

UF INFORMATION TECHNOLOGY

Summer 2021
Authored by: Jose Lorenzo Silva-Lugo, Ph.D.

R Programming Training

An Introduction for Data Analysis and Graphics

Instructor

Dr. Jose Lorenzo Silva-Lugo, HUB 285, Phone: (352) 273-1579. Email: joselugo@ufl.edu.
Office Hours: TBA.

Textbooks

- Field, A, J. Miles, and Z. Field. 2012. Discovering Statistics using R. Sage Publications Ltd, London.
- Michael, W. Trosset. 2009. An Introduction to Statistical Inference and its Application with R. A Chapman & Hall Book, CRC Press. Boca Raton, Florida.
- Muenchen, R. A. 2011. R for SAS and SPSS Users. 2nd Edition. Springers, New York.
- William N. Venables, D. M. Smith, and the R Development Core Team. 2009. An Introduction to R. 2nd edition. Network Theory, Ltd.
- Hothorn, T. and B. S. Everitt. 2014. A handbook of statistical analysis using R. 3rd edition. CRC Press, Boca Raton, Florida.

Training Schedule

From June 1st (8:00 am) to July 30th (5:00 pm), 2021. No extension will be offered!
Location: online (self-paced)

Goal and Objectives

These training series are an introduction to the program language R specifically designed for faculty, staff, postdocs, graduate students, and teaching assistants. The purpose is to get familiar with the R commands in such a way that participants will feel self-prepared to continue by their own. After completing these trainings, participants will be able to:

- Install R, install and load packets and entering data
- Manage the data set and prepare it for analysis
- Explore the data set with descriptive statistics, normality test and graphics
- Perform elementary statistical analyses such as two or more mean comparison, one-way and two-way ANOVA, correlation, bivariate and multiple linear regression, proportion comparisons, and the logistic regression.

Training Series Description

This is an introductory R training for participants with basic knowledge in statistics at the graduate level. It is oriented toward the Natural and Social Science. Participants should be

familiar with all basic concepts and the statistical procedures presented in the objectives. Therefore, **participants should have taken at least a graduate course in statistics before enrolling** because these trainings summarize the most important statistical analyses included in three graduate courses in the Statistics Department: STA 6166 Statistical Methods in Research I, STA 6126 Statistics Methods in Social Research I and STA 6127 Statistic Methods in Social Research II. However, it is important to highlight that although these training series cover several topics in statistics, it is not a substitute of any of the courses mentioned above. R is just a statistical program language, and the training focuses on explaining how to use this language to carry out the statistical analysis.

The first training introduces data input with the programming language called 'S'. Then, it will describe how to format different types of data and how to summarize them, which they are essential steps before the actual analysis. The ability of R to simulate different types of random data allows you to carry out simulations in order to find out what type of distribution your data might have. During this process, graphic representation of the distribution of your data set is critical, and R is particularly very handy providing a variety of charts. Then, statistical inference comes into place by formulating hypotheses for testing means and medians. In this regard, the two-sample t-test and ANOVA are very useful and common statistical analyses. In addition, regression and correlation are other important components in any classical statistical analysis. After that, cross-tabulation will allow us to do another type of analysis: goodness of fit and independence. Finally, the logistic regression closes the training series with a sophisticated way to carry out the analysis. Therefore, these training series present you a free program language to carry out the most common statistical analyses.

Participants can register in any of the eight training sessions. However, if you are a beginner, it is strongly recommended to attend the first three trainings because the skill learned in these trainings will be used later in other trainings.

The teaching technique will follow the classical training session composes of lecturing and activities, in which students take the dynamic role of watching videos, asking questions, and working on hands-on activities. Lectures and activities will always be together to provide knowledge and to practice the skill. Class is 100% online and self-paced.

Training Content
1. **Introduction to R**:
   - What is R?
   - Installing R, R Studio and R Commander
   - Installing and loading packages
   - Loading data sets from a computer
   - Mathematical operations with objects
   - Working with vectors
   - Data structures

- Working with functions
- Working with data sets
- Working with graphics
- Getting help with R

2. **Data Preparation**:
   - Detecting anomalies
   - Working with missing values
   - Removing duplicate cases
   - Dealing with outliers
   - Aggregating cases
   - Normalizing variables
   - Merging files

3. **Exploratory data analysis**:
   - Performing exploratory data analysis for one or more variables
   - Creating charts for one variable
   - Transforming data
   - Computing a new variable:
     - SqRt, Log10/Ln, and 1/X
     - Mathematical Operations (+, -, *, /)
     - Raising to the power
     - Trigonometric transformations
   - Transforming scale data into an ordinal/binary one
   - Making decisions about the statistical analysis to follow based on the exploratory data analysis.

4. **Mean and Median Comparison**:
   - Making inferential statistics about the mean and the median
   - Appling the parametric t-test:
     - One-sample t-test
     - Two-independent samples t-test
     - Paired difference t-test
   - Applying non-parametric t-test:
     - Sign test
     - Mann-Whitney U Test
     - Wilcoxon Signed Rank Test
   - Build charts for two mean comparison
   - Carry out t-test involving violations of normality

5. **ANOVA, Correlation, and Bivariate Linear Regression**:
   - Making inferential statistics by using one-way ANOVA and two-way ANOVA
   - Executing the Kruskal-Wallis test (non-parametric of one-way ANOVA)

- Performing bivariate correlation parametric and non-parametric
- Carry out bivariate linear regression parametric and non-parametric

6. **Multiple Linear Regression**:
   - Making EDA and charts for more than 2 variables
   - Multiple linear regression and correlation
   - Model building and selection
   - Quantile regression

7. **Proportion Comparison:**
   - Two proportions comparison
   - Chi-squared goodness of fit tests
   - Chi-squared tests of independence

8. **Logistic Regression:**
   - Demonstrating the general procedure to carry out the logistic regression with scale, binary and categorical predictor variables:
     * Fit the model to the data
     * Estimate and interpret regression coefficients, the odds ratio, and confidence intervals
     * Estimate the probability of occurrence
   - Carrying out exercises with the combination of binary, categorical, and scale predictor variables.
   - Testing the linearity assumption

GENERAL NOTICE TO STUDENTS

Students with Disabilities

If you need classroom accommodation because of a disability, you must register with the Dean of Students Office (http://dso.ufl.edu/drc). This office will provide you several forms, and one of them must be turned in to the instructor. Since some of these accommodations require time to be in place, I will appreciate that the form is giving to me with two weeks in advanced.

Training Policies

- You are responsible for following the workflow during training and for studying all materials and resources posted in Canvas
- You are responsible for submitting all activities during the training sessions
- If you have questions, please contact me at joselugo@ufl.edu

Software Use

All faculty, staff, and students of the University are required to obey the laws and legal agreements regarding software use. It is Illegal to copy licensed and/or copy written materials. This is a third-degree felony under Florida law. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against

University policies and rules, disciplinary action will be taken as appropriate. The Office of Academic Technology and the members of the University of Florida community, pledge to hold our peers and ourselves to the highest standards of honesty and integrity.

UF Counseling Services

Resources are available on campus for students having personal problems or lacking clear career and academic goals that interfere with their academic performance. These resources include:

- U Matter We Care, 352-294-2273 | umatter@ufl.edu, help for students in distress
- Counseling and Wellness Center, 3190 Radio Road, 392-1575, personal, sexual assault, and career counseling
- Career Resources Center, Reitz Union, 392-1601, career development assistance and counseling