

November 2020

## **Introduction to R**

*Instructor: Avi Bar Massada*

### **Course objectives**

1. To learn the fundamentals of R syntax
2. To understand the concept and practices of data preprocessing
3. To learn how to conduct common and advanced statistical methods
4. To learn to generate both basic and high-level graphics suitable for scientific publications

Ideally, after completing the course students should feel comfortable enough in their skills to begin utilizing R as their main statistical package. To successfully implement the knowledge gained during the course, students should begin using R in their own projects as soon as possible, or else it is likely that they will forget the basic syntax of R, which is the key hurdle towards effectively incorporating R in one's own research.

### **Background requirements**

Students should have taken an advanced-level statistical course (graduate level). The course assumes knowledge and understanding of basic statistical tests, linear models, and multivariate analyses.

### **Student evaluation**

Weekly assignments (100%).

### **Course structure**

The course will be divided into units, roughly representing the standard way of utilizing R in research. Within each unit, the class will feature a formal presentation followed by hands-on practice by the students. While most practices will be based on data provided by the instructor, it is possible to practice using data related to one's own research, after the original tasks are completed.

## **Program**

### *Unit 1: Intro and syntax*

What is R? GUI types (Rgui, RStudio), ways of getting help, working with packages

R syntax

Object types (vectors, matrices, arrays, data frames, model objects, lists)

Working with scripts

Writing functions

A general framework for using R during research

### *Unit 2: Data management and preprocessing*

Importing/exporting data (from/to text files, excel files)

Data processing (clearing, subsetting, slicing, reshaping, conversions, shuffling)

Generating new data using different statistical distributions

### *Unit 3: Descriptive statistics and basic plotting*

Summary statistics

Basic plotting (using R-base)

### *Unit 4: Statistical inference*

Common statistical tests (anova, Z, t,  $\chi^2$ , nonparametric tests)

Linear models (simple and multivariate)

Generalized linear models (logistic regression)

Mixed models (linear and generalized)

### *Unit 5: Multivariate analyses*

Distance matrices

Ordinations (PCA, NMDS, CCA)

Significance tests for multivariate data (Mantel, ANOSIM, ADONIS, PERMANOVA)

Plotting multivariate data

### *Unit 6: Advanced plotting*

Complex plots using R-base; 3D plotting, Using ggplot2 (optional, TBD)