Undergraduate Program in International Relations

Data Analysis in R

Course objectives:

To provide a comprehensive overview of R and cover the essential exploratory techniques for summarizing data and creating model simulations in R.

Learning Outcomes:
- Mastering R language fundamentals and basic syntax
- Becoming familiar with major R data structures
- Applying basic techniques of data analysis to real-world data sets

Course outline

1. Introduction to Data
Data basics. Role of quantitative data analysis in decision making process. Common pitfalls in drawing conclusions from data. Introduction to RStudio.
2. Fundamentals of R Syntax
Basic data types in R. Assigning a variable. Creating, naming and selecting elements from vectors. Basic operations with vectors, matrices, data frames, tibbles and lists. R base graphics.
3. Summarizing Data
Measures of location: mean, median, mode. Measures of spread: standard deviation, interquartile range, range. Percentiles. Robust statistics. Data transformation.
4. Getting and Cleaning Data
Importing data to R. Various data sources: text files, web, APIs. Raw and processed data. Working with dates. ZOO and XTS data formats. The principles of tidy data.
5. Probability Distributions
Probability mass functions. Probability density functions. Conditional Probability. Expected values. Data variance. Standard error of the mean. Frequencies and the mode. Empirical distribution function.
6. Linear Regression
Univariate linear regression. Interpreting coefficients. Residuals. Linear regression for prediction.
7. Statistical Inference and Hypothesis Testing
Confidence intervals. T tests. P values. Calculating power of tests. Bootstrapping procedure.

Duration: Fall 2017 (Module 1)

Course Materials:

- Instructor's Handouts
- An Introduction to Statistical Learning: with Applications in R
  http://www-bcf.usc.edu/~gareth/ISL/index.html
- Quick-R:
  http://www.statmethods.net

Course Structure: The course revolves around the most essential structural elements of R, which will be illustrated through case studies.

Forms of Final Assessment: home assignments+group project

Module Grade: 50% - home assignments, 50% - group project

(96-100% - 10, 90-95% - 9, 80-89% - 8, 75-79% - 7, 65-74% - 6, 55-64% - 5, 45-54% - 4, 35-44% - 3, 25-34% - 2, 0-24% - 1)

Instructors: Mikhail Vladimirovich Kamrotov (kamrotov@gmail.com) and Nikolai Igorevich Korzhenevsky (nick@rbk.ru)

Office hours: by appointment

Classroom policies:

- Hand-in assignments policy: All home assignments should be submitted electronically via instructor's email on the due date. No deadline extensions are possible.
- Cheating policy: In case of any kind of plagiarism (with the detected source), the assignment is evaluated as zero without the chance to make up for it. In case of two written assignments with the similarity index of 50% and higher from two students, both get a zero for the assignment.