

Undergraduate Program in International Relations

Introduction to Data Culture

Course objectives:

This course provides an introduction to modern data analysis methods through a series of seminars based on Excel. The aims of the course are to:

- provide an introduction to modern data science techniques;
- introduce main concepts of scientific data analysis;
- show best practices of working with data;
- train basic skills in MS Excel.

Learning outcomes:

At the end of the course, having completed the set readings and the activities, students will be able to:

- formulate and solve simple scientific analysis problems;
- search, download and process data;
- evaluate and compare the quality of applied data analysis methods;
- interpret the results using knowledge in the professional field;
- perform a data analysis in MS Excel.

Course description:

This course offers an introduction to the modern Data Science methods that are useful both for research and industrial careers. The main focus of the course is to teach students to find data on the Internet, process it and perform a simple data analysis. Students are trained to develop critical thinking and a scientific approach to problem solving.

The course starts from the basics of working with data. Students will be taught to perform a basic data analysis in MS Excel. In the first part of the course, students will learn how to sort and filter data, calculate various distribution characteristics (mean, median etc.) and create graphs and charts. A part of the course also concerns the main methods of data storage and their usage.

The second part concerns the main methods that lead to scientific results of the analyses in humanities starting from time series and linear regression analyses to the simplest predictive modelling. Students will learn to apply all these techniques in MS Excel.

Section 1: The basics of working with data.

Weeks 1-2:

Applied data science in the international relations, examples of applications, examples of erroneous applications. Data. Probability distributions, basic types of distributions. Main characteristics of the distributions: mean, mode, median. Correlations and causality.

Weeks 3-4:

Main methods of accumulation, storage and processing of data. Types of variables. Prospects for the development and application of data analysis. Processing polls, ratings. The simplest text analysis.

Section 2: Basic Data Analysis.

Weeks 5-6:

Time series. Basic methods for forecasting time series. Calculation of quality metrics of predictive models.

Week 7:

Linear regression analysis. Construction and visualization of regression lines. Obtaining predictions using a linear regression model. The concept of splitting data into train and test. Model quality evaluation.

Methods and Forms

The study methods and forms used in this course include teachers' explanations, practice work and homework. There is one weekly seminar. The seminar includes the explanation of the new material with the help of presentation slides and practical work in the computer class under the supervision of a teacher. Homework generally includes the computer-based assignment in Excel. Interim control work and final examination are conducted in Excel and include all cover topics by the time of assessment. In the beginning of each seminar students are asked to perform tests to check knowledge based on the materials of previous seminars. Each 100% correctly completed test provides the student with extra points to ongoing assessment.

Duration. Module 2, November and December 2018.

Course structure. Weekly seminars.

Final assessment. Homework, Classwork, Examination.

Module Grade:

Final assessment = $0.8 \times \text{Ongoing Assessment} + 0.2 \times \text{Examination Assessment}$,
where

Ongoing Assessment = $0.7 \times \text{Homework} + 0.3 \times \text{Control Work} + \text{Extra points}$

Instructors:

Dr. Elena Kantonistova (ekantonistova@hse.ru, elena.kantonistova@yandex.ru)

Ms. Marina Ananyeva (ananyeva.me@gmail.com).

Recommended Literature:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: Springer.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning*. New York: Springer series in statistics.

Support Material:

- MS Office Blog <https://blogs.office.com/en-us/?eu=true>
- Support documentation for Microsoft Excel: <https://support.office.com/en-US/Excel>

Syllabus for STATISTICS FOR INTERNATIONAL RELATIONS, PART 2 (Nov-Dec 2018)

Tutor: *Nadezhda Shilova*

description

Part 2 of the course will last 2 months, starting from November 2018, and will give ½ of the grade for the whole course of D.A. and Stats.

This particular part of the course is aimed to give further understanding of the statistical analysis process, based on the basic notions learned in the first semester.

The course will be read to the students of the 1st year, HSE and UoL International Relations program. No prerequisites are imposed rather than usual school maths. But successful completion of the course is crucial for understanding the further courses on economics, Data Culture, all kinds of analysis taught on the program.

Teaching Methods

The course program consists of:

- classes (labs),
- regular self-study. Self-study is supposed to take no less than 2/3 of the time needed to complete the course successfully.

Assessment and grade determination

Control takes the following forms:

- short review tests (10-15 min) in the beginning of labs;
- written exam at the end of the first module (60 min.).

The course grade will be determined from the following activities:

- short review tests (40%);
- lab work quality (30%);
- fall semester test (30%).

Main Reading

1. Newbold, P. Statistics for Business and Economics, 4th ed. University of Illinois, 1994, or
2. Newbold, P., Carlson, W., Thorne, B. Statistics for Business and Economics (Pearson Education, 2012).

Additional Reading

1. Lindley, D.V., Scott, W.F. New Cambridge Statistical Tables (Cambridge: Cambridge University Press, 1995)

Course outline.

1. Continuous random variables and probability distributions. The normal distribution.

Probability function. Cumulative probability function. Probability density function. Jointly distributed continuous random variables. Normal distribution and its properties.

2. Sampling and sampling distributions.

Random sample. Standard error. Sample proportion. Chi-square. Degrees of freedom.

3. Interval estimation.

Confidence intervals. Student's t distribution.

4. Hypothesis testing.

Hypothesis. Null hypothesis. Type I errors, type II error. Significance level. Measuring the power of the test. P-value.

5. Linear correlation and regression.

Correlation. Covariance. Least squares estimation. Dependent and independent variables. Standard assumptions for the linear regression model. Explanatory power of linear regression equation. Residuals. Coefficient of determination.