Syllabus

Data Analysis 2

Department of Economics and Business
CENTRAL EUROPEAN UNIVERSITY

- **Instructor:** Arieda Muço
- **Credits:** 2 (4 ECTS)
- **Term:** Fall 2017-2018
- **Course level:** [MA/MSc]
- **Prerequisites:** Data Analysis 1

Course description

This introductory course focuses on classic statistics methods which we will discuss with applications. It focuses on summary statistics, probabilities, type of variables and their distributions. We will cover confidence intervals, standard errors and how to perform hypothesis testing.

To understand the patterns of the data we deal with, we will look at frequency tables, joint, marginal, and conditional probabilities.

We will also discuss the meaning and implication of key theories such as the Central Limit and Bayes' theorem.

Learning outcomes

Understand and interpret basic statistics and figures to summarize data

Understand and interpret probabilities and distributions

Being able to tell apart independent events, clustered observations

Formulate hypothesis and test them

Understand how biases or dealing carelessly with data can misrepresent findings

Reading list

Handouts (compulsory)

To deepen the understanding of the material you may also choose one of the books or a combination of books enlisted here below:

Introductory Statistics, by Ross

Statistics by Freedman, Pisani, Purves (in the CEU library)

Complete Business Statistics by Aczel (in the CEU library)

Statistics for Business and Economics by Anderson, Sweney, Williams (in the CEU library)

Statistics for Business and Economics by Newbold, Carlson, Thorne (in the CEU library)

Assessment

Quizzes 15%. Assignments 25%. Exam: 60%. Passing the course requires a passing score of over 50% on the written exam.

Course schedule and materials for each session

1	Basic statistics: Centrality and dispersion. Mean; median; Standard Deviation; Quantiles
2	Elementary probability theory. Bayesian foundations
3	Types of variables and their distributions
4	Data Collection: Access admin data; collect data from Web (scraping); design surveys.
5	Sampling. Representative samples; sample size. Consequences of selective coverage, item non-response.
6	Big data vs other types of data. Bias, selection in online data collection
7	Levels of generalization: similar datasets different situations. External validity.
8	Repeated samples. Bootstrap.
9	Standard errors and Robust standard errors. Analytical formulae using the Central Limit Theorem.
10	Confidence Interval as the prime tool to make inference.
11	Hypothesis testing: Null and alternative hypotheses. False negatives, false positives, types of error. Size and power. p-value. Short discussion of bayesian alternative.
12	Hypothesis testing: Testing multiple hypotheses and p-hacking