

## **Introduction to Statistics**

Mandatory MA level class

The class meeting times: Tue 17:20-19:00, Fri 11:00-12:40

by	Tamás Rudas	and	Levente (Levi) Littvay
<u>CV:</u>	<a href="http://statisztika.tatk.elte.hu/tanszeki_honlap/RT_CV.htm">http://statisztika.tatk.elte.hu/tanszeki_honlap/RT_CV.htm</a>		<a href="http://levente.littvay.hu/cv.pdf">http://levente.littvay.hu/cv.pdf</a>
<u>E-mail:</u>	<a href="mailto:rudas@tarki.hu">rudas@tarki.hu</a>		<a href="mailto:littvayl@ceu.hu">littvayl@ceu.hu</a>
<u>Phone:</u>	Make an appointment by email.		(mobile) +36 70 538-3683
<u>Office Hours:</u>	Make an appointment by email.		Tue 2-3:15/Wed 4-5 or by app

### **Summary**

Statistics is the science of data collection, analysis and interpretation, and familiarity with the appropriate statistical methods is a fundamental part of the methodological knowledge of political scientists. This class consists of two parts. During the first half of the term, the conceptual framework, the main applications and the most fundamental methods will be discussed. During the second half of the term, using computers and statistical software, the students will gain experience in applying some of the most important methods and will deepen their insight into statistical analysis and how it may be useful for political scientists.

### **FIRST HALF OF THE TERM**

#### **Topics covered**

- Official and inferential statistics
- Experiment, observational study
- Census, survey
- Descriptive statistics
- Probabilistic background
- Basic inferential statistics
- Behavior of experiments of the longrun (asymptotics)
- Estimation and tests of hypotheses
- Chi-squared test

#### **Learning goals**

To obtain a clear view of the applicability of statistics in political science problems, to understand the nature of statistical reasoning, to be able to perform simple statistical tests

#### **Textbook**

Freedman, Pisani, Purves: Statistics (3rd Ed), Norton, 1997

#### **Grading**

The first half of the term will be graded based on a test.

## **SECOND HALF OF THE TERM**

**Class Website:** <http://e-learning.ceu.hu>

You need to sign up for the on-line class. Once the class has started, you can log in with your CEU ID and password. The code to sign up for the class is: barack

Supporting material and assignments will be posted on line. Assignments will have to be submitted through the website. (You will need SPSS and Excel to do them.)

### **Evaluation**

Weekly assignments will be given in class. These are due the day before our next class at 11:00am and will be graded by the time of the class. (Grading your assignments before class helps me decide if we need to spend more time on a topic or if we can move on.) You can be up to two weeks late on one assignment without penalties. Use it wisely. Attendance (with timely arrival) is **required**. Late arrivals are counted as unexcused absences. Three unexcused absences will lead to an automatic failure of the class. If you will miss a class or come late for any reason, make sure I know about it **before** the class.

You will also have a research project in which you will have to apply the techniques learned in class. This project must include a description of your hypothesis (with a minimal review of literature), description of your data including possible problems it has, description of the method used including reasoning why you selected the method, discussion of possible alternatives, your model, the presentation of your results in appropriate format, interpretation and conclusion.

Your research topic has to be pre-approved.

### **Grading**

Assignments 70% (due 11am the day before class)

Project 30% (presented the last week of classes and due the last day of class)

### **Important Notice**

Complete academic honesty is expected of everyone. Failure to comply with this requirement will result in automatic failure in this course (and subsequently in the program) and additional disciplinary action on higher levels. This is an American university and American standards will be applied. For more information about these standards see: [http://en.wikipedia.org/wiki/Academic\\_dishonesty](http://en.wikipedia.org/wiki/Academic_dishonesty) (**READ VERY CAREFULLY!**)

All assignments are to be done individually. You can talk about how to do it but none of the actual work can be done in a group. Any evidence to the contrary will be investigated.

This mandatory class aims to show you how computers can make your life a lot easier when doing statistical analysis. The goal of the course is to teach you how to do the most

commonly used statistical analysis with the most commonly used statistical software. Software used in this class is Excel, SPSS and possibly R. This class will have two sections. Assignments will be made based on your statistical interest/knowledge. (One group will have to take an introductory statistics class before this one. The other group will start with this class and will have to take a multivariate statistics course after this one.)

Topics covered will depend on how fast we advance through the course. I expect the accelerated class to advance much faster. Below you find the topics we absolutely will cover and the topics we will try to cover. In the time remaining (if any) we will replicate the most commonly used techniques in R ([www.r-project.org](http://www.r-project.org)), an extremely flexible and modular freely available statistics package (that is not really user friendly).

No reader or text for this class. In-class presentations will be posted on the web. A number of SPSS books are available in the library

### **Topics We Will Definitely Cover**

- Intro to Applied Statistics (the debates)
- What is/are data and how to use it
- Intro to SPSS (GUI vs Scripts)
- Data and SPSS (Variables, Labels, Datasets, etc.)
- Descriptive Statistics
- Crosstabs and other presentation methods
- Chi-Square
- Reliability
- Correlations
- Regressions
- Regression Diagnostics
- Logistic Regression

### **Possible Additional Topics**

- T Test and ANOVA
- Factor analysis
- Cluster analysis
- Missing Data
- Overview of what's next

### **Replication of the most important topics in R**