

Elements of Statistics and Probability

Ioannis Ntzoufras, Professor, AUEB, ntzoufras@aueb.gr

Overview

The course is a sort and brief introduction to the basics of probability, statistics and data analysis. The aim is to remind to all graduates the basic notions of statistics in combination with a short introduction in R.

Key Outcomes

By completing the course the students will be able to

- Know the basic notions of statistics and probability
- Be familiar (in basic level) with R
- Employ statistical methods and models.
- Solve statistical problems using data analysis

Requirements and Prerequisites

This course acts as a review of basic knowledge and notions in statistics. We assume that all students are familiar with a basic undergraduate course in Statistics and Probability but this is not required.

Bibliography

1. Basic Statistics and probability:

- a. Diez, D., Barr, C., & Cetinkaya-Rundel, M. (2012). *OpenIntro statistics* (Second Edition). Free Open Book; available at <http://www.openintro.org/stat/textbook.php>

2. Regression:

- a. Faraway, J. (2002). *Practical regression and ANOVA using R*; available at <http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>
- b. Fox J. & Weisberg H.S. (2011). *An R Companion to Applied Regression*. 2nd edition. SAGE Publications Inc.

3. R related books

- a. Crawley M.J. (2014). *Statistics: An Introduction Using R*. 2nd Edition. Wiley-Blackwell.
- b. Ντζούφρας, Ι., Καρλής, Δ., 2015. Εισαγωγή στον προγραμματισμό και στη στατιστική ανάλυση με R. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/2601>
- c. Φωκιανός Κ & Χαραλάμπους Χ. (2010). *Εισαγωγή στην R – Πρόχειρες σημειώσεις*. Πανεπιστημιακές σημειώσεις. 2^η έκδοση. Τμήμα Μαθηματικών & Στατιστικής. Πανεπιστήμιο Κύπρου, url: <http://cran.r-project.org/doc/contrib/mainfokianoscharalambous.pdf>.
- d. Φουσκάκης Δ. (2013). *Ανάλυση Δεδομένων με Χρήση της R*. Εκδόσεις Τσότρας. Αθήνα. (Κωδικός Βιβλίου στον Εύδοξο: 33134029).

Grading

With a written examination at the end of the course.

Course Syllabus

The course comprises of 5 sessions of three one-hour lectures which will be spread out in the first two weeks of the first semester.

Session 1: Motivation, Basic Probability and Descriptive Measures

1. Motivation – Statistics and the Data Science in the 21st century

The first lecture aims to motivate the students and explain the usefulness and the importance of Statistics and Data Science in the modern world. Examples from several fields are presented.

2. Basic Notions of Statistics and Probability

Basic definitions (population, sample, basic sampling methods, types of variables, data). Basic Probability Concepts, Conditional Probability, Statistical independence, marginal probability, Bayes' Theorem, Random variables (discrete and continuous), probability function, probability density function, cumulative probability, expectation and variance, Well known distribution functions (Bernoulli, Binomial, Poisson, Uniform, Normal etc.)

3. Descriptive Statistics

Definition, interpretation and usage of descriptive measures: (a) of central location: mean, median, mode, geometric mean (b) relative location (c) of dispersion: variance, standard deviance, range, IQR and (d) of shape: kurtosis and skewness.

Session 2: Introduction to R and Descriptive Statistics

A basic introduction in R will be given.

Session 3: Estimation and Hypothesis testing

Going from a sample to the population. Estimators. Properties. How we construct good estimators. The likelihood function, Maximum likelihood estimator, Point estimate, Confidence intervals. How we construct hypothesis tests and why. What is the null hypothesis. Rejection area and values. P-value. t-tests, ANOVA, chi-square tests

Session 4: Regression analysis

Basic principles of regression analysis. Estimation. Example with interpretation. Assumptions. Revising the model.

Session 5: Closure and Examination