



The Hebrew University of Jerusalem

Syllabus

MODERN STATISTICAL DATA ANALYSIS - 52311

Last update 24-01-2019

HU Credits: 4

Degree/Cycle: 1st degree (Bachelor)

Responsible Department: Statistics

Academic year: 0

Semester: 2nd Semester

Teaching Languages: Hebrew

Campus: E. Safra

Course/Module Coordinator: Or Zuk

Coordinator Email: or.zuk@mail.huji.ac.il

Coordinator Office Hours: By appointment

Teaching Staff:

Dr. Or Zuk
Mr. omer ronon

Course/Module description:

The course will introduce modern statistical methods, and concentrate on high-dimensional and large-scale datasets. We will discuss the novel computational and statistical challenges arising from such datasets. Emphasis will be given on practical methods and computational efficiency.

During the course we will use and implement modern statistical procedures and apply them to simulated and real-life datasets from different domains.

Course/Module aims:

The goal of the course is to introduce the student to modern methods and tools in statistics.

Learning outcomes - On successful completion of this module, students should be able to:

to understand a few modern statistical methods, implement them in a standard programming language efficiently, and apply them to empirical datasets in order to solve a concrete scientific problem

Attendance requirements(%):

none

Teaching arrangement and method of instruction: Lectures and practice sessions

Course/Module Content:

Tentative list:

0. Data Pre-processing: normalization and transformation, missing data, censoring, imputation, visualization
1. Hypothesis Testing: permutation tests, power calculations, multiple hypothesis testing (Bonferroni, FDR)
2. Regression: multivariate linear regression, variable selection and sparsity: lasso, lars, elastic-net.
3. Classification: logistic regression, random forest, neural networks

-
4. *Model Selection and Averaging: AIC, BIC, cross-validation, bagging, SURE*
 5. *Dimensionality Reduction: linear methods (SVD, PCA) and non-linear methods (manifold learning, kernel PCA, Isomap, LLE)*
 6. *Clustering, k-means, EM-algorithm*

Required Reading:

None

Additional Reading Material:

The Elements of Statistical Learning – Data mining, inference and prediction
(Tibshirani, Hastie and Friedman)

<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>

Large Scale Inference, Bradley Efron

http://statweb.stanford.edu/~ckirby/brad/LSI/monograph_CUP.pdf

Advanced Data Analysis from an Elementary Point of View, Cosma Rohilla Shalizi

<http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/>

Course/Module evaluation:

End of year written/oral examination 0 %

Presentation 0 %

Participation in Tutorials 0 %

Project work 40 %

Assignments 60 %

Reports 0 %

Research project 0 %

Quizzes 0 %

Other 0 %

Additional information:

(will be updated)