# The Hebrew University of Jerusalem

## Syllabus

## MODERN STATISTICAL DATA ANALYSIS - 52311

Last update 24-01-2019

<u>HU Credits:</u>  4

<u>Degree/Cycle:</u> 1st degree (Bachelor)

<u>Responsible Department:</u> Statistics

<u>Academic year:</u> 0

<u>Semester:</u> 2nd Semester

<u>Teaching Languages:</u> Hebrew

<u>Campus:</u> E. Safra

<u>Course/Module Coordinator:</u> Or Zuk

<u>Coordinator Email:</u> [or.zuk@mail.huji.ac.il](mailto:or.zuk@mail.huji.ac.il)

<u>Coordinator Office Hours:</u> By appointment

<u>Teaching Staff:</u>

*Dr. Or Zuk*
*Mr. omer ronen*

*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 Rohillla 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)