



The Hebrew University of Jerusalem

Syllabus

Big Data in Economics - 57493

Last update 05-03-2023

HU Credits: 4

Degree/Cycle: 1st degree (Bachelor)

Responsible Department: Economics

Academic year: 0

Semester: 1st and/or 2nd Semester

Teaching Languages: Hebrew

Campus: Mt. Scopus

Course/Module Coordinator: Sarit Agami

Coordinator Email: sarit.agami@mail.huji.ac.il

Coordinator Office Hours:

Teaching Staff:

Dr. Sarit Agami

Course/Module description:

The course discusses methods for analyzing large data, ie data in which the number of observations is large and / or the number of variables is large. The aim of the course is to become familiar with the methods and apply them to real data collections. The application will be made using the R software.

Course/Module aims:

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

To analyze big data: describe and draw conclusions from them.

Attendance requirements(%):

90%

Teaching arrangement and method of instruction:

Course/Module Content:

1. Principal Component Analysis (PCA)
 - Population principal components, sample principal components, visualizing principal components.
 - The Kaiser-Meyer-Olkin Measure of Sampling Adequacy, Bartlett's test of sphericity
- Principal components regression (PCR)
- Robust PCA
- PCA on discrete data
2. Canonical Correlation Analysis (CCA)
 - Population canonical correlations, sample canonical correlations
 - Canonical correlations and transformed data
 - Canonical correlations and regression
3. Discriminant Analysis
 - Linear discriminant rules
 - Discrimination under Gaussian assumptions
 - PCA, discrimination and regression
4. Cluster Analysis
 - Hierarchical Agglomerative Clustering
 - k-Means Clustering

-
- *Principal Components and Cluster Analysis*
 - 5. *Factor Analysis*
 - *Population k-Factor Model, Sample k-Factor Model*
 - *Factor Loadings*
 - *Factor Scores and Regression*
 - 6. *Multidimensional Scaling*
 - 7. *Graphical models for high-dimensional data*
 - 8. *Ridge regression*
 - 9. *Lasso*

Required Reading:

None

Additional Reading Material:

Multivariate analysis
Mardia, Kantilal Varichand; London:Academic Press;1979

An introduction to applied multivariate analysis with R
Everitt, Brian;New York: Springer; 2011

Course/Module evaluation:

End of year written/oral examination 0 %
Presentation 10 %
Participation in Tutorials 10 %
Project work 0 %
Assignments 5 %
Reports 0 %
Research project 75 %
Quizzes 0 %
Other 0 %

Additional information:

During the course, homework will be given for the application of what is being studied. A score will be given for each exercise, with the weight of the exercises being 5% of the final score. Also, the students will be required to present a short presentation in class of an article that applies methods in big data, the weight of that presentation is 10% of the final grade. During the semester, each student will choose a topic for writing a seminar paper according to what interests him or from a

collection of topics that will be offered for selection, and individual meetings will be held with each student to work on the topic he has chosen.

** The lessons will be recorded, and the lesson recordings will be available to all students at any time.*