



Course program and reading list

Semester 1 Year 2023

School: Efi Arazi School of Computer Science M.Sc.

Statistics and Data Analysis

Lecturer:

Mr. Ben Galili ben.galili@post.runi.ac.il

Prof. Zohar Yakhini zohar.yakhini@runi.ac.il

Tutors:

Mr. Saar Buchnik saar.buchnik@post.runi.ac.il

Ms. Shulamit Finley shulamit.finley@post.runi.ac.il

Teaching Assistant:

Mr. Ben Galili ben.galili@post.runi.ac.il

Ms. Shulamit Finley shulamit.finley@post.runi.ac.il

Course No.:	Course Type :	Weekly Hours :	Credit:
3620	Lecture	4	4

Course Requirements :	Group Code :	Language:
Final Exam	231362001	English

Prerequisites

Prerequisite:

52 - Calculus I

53 - Calculus II

54 - Linear Algebra I
55 - Linear Algebra II
56 - Discrete Mathematics
59 - Data Structures
69 - Logic And Set Theory
109 - Introduction to Probability
417 - Introduction To Computer Science
3141 - Machine Learning from Data

Students who took one of the courses listed below will not be allowed to register to the course Statistics and Data Analysis (3620):

3584 - Statistics and Data Analysis



Course Description

The course will cover statistics and data analysis from a computer age perspective. Data is becoming increasingly central, in scope and in scale, to science, engineering, economics and many more walks of life. Statistics and statistical interpretation is central to the successful and relevant use of data in all domains. Our course will generally follow the weekly outline below:

1. Python histograms and heatmaps
 2. Introduction to statistics and probability theory. Basic definitions and some basic distributions.
 3. More important distributions; Independence, convolutions; The coupon collector process
 4. Continuous probability distributions and CLT
 5. Gauss variants (logNormal, GMM)
 6. Parametric approximation and MLE; the EM algorithm
 7. Statistical inference, confidence intervals, p-values, hypothesis tests, comparing occurrence frequencies;
 8. Inference - continued. Non parametric tests - WRS
 9. Correlation measures; Correlations – parametric and non parametric. Statistical significance of correlations;
 10. Correlations, continued; The multinomial distribution; Markov CLT
 11. Multiple testing, Bonferroni and FDR corrections, analysis of gene expression data;
 12. Intro to information theory: entropy and information, KL and other distances between distributions (Wasserstein);
 13. Brief on survival analysis; Summary of course;
-



Course Goals

To present, understand and practice methodologies for statistical inference, modelling and data analysis, as relevant in developing and practicing data science.

To understand variants and to emphasize properties and assumptions used in statistical inference.

Methods will be implemented and applied in homework projects investigating example datasets.



Grading

4 HW assignments. Each will be 18/100 of the general grade.

An exam will be 28/100 of the general grade.

Must pass exam to pass the course.



Learning Outcomes

The students will be able to approach and perform analysis of data, to use statistically rigorous and efficient methods for this process, and to effectively present results and conclusions.

Example cases will be addressed and analyzed. Data will be taken from various science and engineering domains.



Lecturer Office Hours

By appointment



Tutor Office Hours

By appointment



Reading List

John Rice: Mathematical statistics and data analysis