

## The Department of Mathematics

2017–18–B term

**Course Name** Statistical methods for big data

**Course Number** 201.1.9131

**Course web page**

<https://www.math.bgu.ac.il/en/teaching/spring2018/courses/demography>

**Lecturer** Dr. Luba Sapir, <lsapir@post.bgu.ac.il>, Office -109

**Office Hours** <https://www.math.bgu.ac.il/en/teaching/hours>

### Abstract

### Requirements and grading<sup>1</sup>

### Course topics

- .1 Review of probability: a. Basic notions. b. Random variables, Transformation of random variables, Independence. c. Expectation, Variance, Covariance. Conditional Expectation.
- .2 Probability inequalities: Mean estimation, Hoeffding's inequality.
- .3 Convergence of random variables: a. Types of convergence. b. The law of large numbers. c. The central limit theorem.
- .4 Statistical inference: a. Introduction. b. Parametric and non-parametric models. c. Point estimation, confidence interval and hypothesis testing.
- .5 Parametric point estimation: a. Methods for finding estimators: method of moments; maximum likelihood; other methods. b. Properties of point estimators: bias; mean square error; consistency c. Properties of maximum likelihood estimators. d. Computing of maximum likelihood estimate

---

<sup>1</sup>Information may change during the first two weeks of the term. Please consult the webpage for updates



- .6 Parametric interval estimation a. Introduction. b. Pivotal Quantity. c. Sampling from the normal distribution: confidence interval for mean, variance. d. Large-sample confidence intervals.
- .7 Hypothesis testing concepts: parametric vs. nonparametric a. Introduction and main definitions. b. Sampling from the Normal distribution. c. p-values. d. Chi-square distribution and tests. e. Goodness-of-fit tests. f. Tests of independence. g. Empirical cumulative distribution function. Kolmogorov-Smirnov Goodness-of fit test.
- .8 Regression. a. Simple linear regression. b. Least Squares and Maximum Likelihood. c. Properties of least Squares estimators. d. Prediction.
- .9 Handling noisy data, outliers.