

Statistical Data Analysis, Sampling, and Data Quality Control (204-1-3541)

Dr. David Lukatsky (Department of Chemistry)

lukatsky@bgu.ac.il

The aim of the course is to develop a fundamental understanding of statistical methods of modern data analysis and apply the obtained knowledge for a number of practical problems for experimental data analysis using Matlab programming language. During the course, the students will learn both theoretical material and Matlab programming.

The attendance is obligatory. The 100% of the final grade will be determined based on a final homework project (final research project) that students will perform at the end of the course. The duration of the final homework project will be from the end of the semester till the end of Moed bet. No additional exam will be given at the end of the course.

Introduction to probability theory:

1. Introduction to probability theory. Random processes. Discrete and continuous random processes.
2. The notion of probability distribution. Examples of probability distributions: Binomial distribution.
3. Mean. Standard deviation.
4. Poisson random process. Gaussian random process.
5. Central limit theorem.

Introduction to Matlab programming.

Statistical description of data:

1. Comparing two probability distributions.
2. Linear correlation.
3. Modeling of data. Fitting of data. Linear regression.
4. Introduction to Fourier transform. Discrete Fourier transform. Fast Fourier transform.
5. Introduction to Monte-Carlo method.

Examples of statistical data analysis and data quality control:

1. Statistical analysis of repetitive sequence patterns in genomic DNA.
2. Statistical analysis of DNA recognition by transcription factors.