

# SMD, Statistical Methods (Mathematical Theory with Data Science Applications)

Syllabus, Fall Semester 2024, reading course, BSM

- Lecturer: Prof. Marianna Bolla, DSc.
- Prerequisite: Undergraduate Calculus and Basic Probability
- **Course description:** Statistics teaches us how to behave in the face of uncertainties, according to the famous mathematician, Abraham Wald and the book ‘Statistics and Truth’ of C.R. Rao. Theoretically, we will learn strategies of treating chances in everyday life, where our inference is based on a randomly selected sample from a large population, and hence, we intensively use concepts of probability (laws of large numbers, Bayes rule). Parameter estimation and hypothesis testing (parametric and non-parametric inference) are introduced on a theoretical basis, but applications are intensively discussed and presented on real-life data. Methods of supervised and unsupervised learning are outlined for multivariate data sets; former include regression and discriminant analysis, while latter include factor and cluster analysis. The students are made capable of solving real-world problems by choosing the most convenient method or statistical test. Outputs of the BMDP (BioMeDical Program package), developed at Stanford in 1981, are also analyzed in the classes.
- **Topics covered:**
  1. Short overview of probability theory (sample spaces, random variables, notable distributions, Bayes rule, laws of large numbers, Central Limit Theorem).
  2. Basic concepts of estimation theory. Point estimation, unbiasedness. Confidence intervals.
  3. Inferences about a population, sampling statistics.
  4. Basic concepts of hypothesis testing, concept of a uniformly most powerful (UMP) test.
  5. Parametric inference, comparing two treatments (z, t, F tests).
  6. Nonparametric inference: Wilcoxon test.
  7. Analyzing categorized data (two-way classified tables), chi-square test.
  8. Introduction to linear models: regression analysis.
  9. Covariance, Pearson and Spearman correlation.
  10. Analyzing outputs of programs for medical and econometrical data.
- **References:**
  - Textbook: G. K. Bhattacharyya, R. A. Johnson: Statistical Concepts and Methods, Wiley, 1987, 1992.
  - Lecture notes to the topics that are not covered by the above textbook are on the lecturer’s homepage <https://www.math.bme.hu/marib/bsm>.
  - For those, who are more deeply interested: C. R. Rao: Statistics and Truth, World Scientific, 1997.
  - Handouts: tables of notable distributions and percentile values of basic test distributions (can be used in the midterm and final tests).
- **Grading:** 4 homework assignments (2x5=10 points for each), the midterm test (multiple choice, open-book), and the final exam (with theoretical questions and applications, only handout tables can be used) make up 40%, 20%, and 40% of the final grade, respectively. The final grade as the function of the total (maximum 100) points is the following. Below 45: F, 46-49: D, 50-56: C+, 57-63: B-, 64-70: B, 71-77: B+, 78-84: A-, 85-91: A, 92-100: A+.
- **Contact details** of the lecturer: Marianna Bolla, DSc, Full Prof. Budapest University of Technology and Economic, Institute of Mathematics, Tel: +3670346630. E-mail: [marib@math.bme.hu](mailto:marib@math.bme.hu). Homepage: <https://www.math.bme.hu/marib/bsm>.
- Office hours: by appointment.