ELTE **MSc., PhD. Post Doc.**
Free choice

COURSE SHEET AND REQUIREMENTS
2021.08.23

1. Title: Multivariate Data Evaluation I. (Chemometrics 1.)

2.	Course code	Semester	requirement	credit	language	Course timing
	KÉM/327		2+2+0 f	6	English	>III-V

3. Responsible person: Károly Héberger, ELKH, TTK, Bp. Magyar Tudósok krt2
4. Lecturer:

Name:	Position:	Institute
Károly Héberger	Scientific advisor	ELKH Research Centre for Natural Sciences

5. The course is built on: Statistics I, chemometrics 1.
6. Recommended studies:
Probability theory, Analytical chemistry, Instrumental analysis
7. Aim of the subject:
To learn selected methods of multivariate data evaluation, (chemometric techniques); to apply them in practice, with special emphasis on avoiding the most frequent errors.
8. Short syllabus of the subject:
1. Introduction to chemometrics, basic ideas, terms, methods, techniques, examples. Elements of mathematical statistics (random variables, distribution function, expectation value, variance, hypothesis testing, examination of normality, etc.).
2. Robust statistics (non-parametric tests, asymmetric confidence bands, fuzzy sets & regression, generalized pair correlation method, GPCM)
3. The black box model and its usage (empirical, "statistical" modeling, choosing variables (factors). The correlation coefficient and its variants (COD = coefficient of determination), their abuse.
4. Multivariate techniques 1: Multiple linear regression, forward selection, backward elimination, best subset, (F test, t test), LFER, QSAR, examples.
5. Multivariate techniques 2: Pattern recognition 2. Principal Component Analysis, Factor Analysis. How and when to use them?
6. Multivariate techniques 3: Pattern recognition (supervised and unsupervised) 1. Hierarchical Cluster analysis, icon plots.
7. Partial least squares regression (PLSR), and its usage for classification. Principal component regression (PCR)
8. Selection of variables (factors) from (seemingly) equivalent variables. The Pair Correlation Method and its generalization. Cross-validation.
9. Linear discriminant analysis, wine, and olive oil authentication. Genetic algorithm.
10. Comparison of methods and models by consensus. Sum of ranking differences. Examples. Features of SRD ordering, methods of data fusion.
11. Curve fitting, ANOVA, regression (linear and non-linear case, model discrimination, checking the linearity, detection of trends etc., most frequent errors in regression and how to avoid them).
12. Consultation Practice, analysis of students' data sets.
13. Exam (written).

9. Method of teaching the subject: Lectures (2) Practice (2)
10. Requirements:
Following the verbal lectures, practicing (home), preparation of small lectures (5 slides, 10 minutes) based on selected literature sources.
11**. Literature** (Recommended reading)

- P.R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill Book Co., New York, 1969.
- N.R. Draper, H. Smith, Applied Regression Analysis, Wiley, New York, 1981 (2nd. ed.).
- **T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, New York, 2001., 2nd edition, 2009 February.**
  https://web.stanford.edu/~hastie/Papers/ESLII.pdf