University of Miskolc

Faculty of Materials Science and Engineering
Antal Kerpely Doctoral School of Materials Science & Technology

Data Analysis

Dr. Olivér Bánhidi

Data Analysis

Dr. Olivér Bánhidi

Lecturer

Dr. Olivér Bánhidi honorary professor, Institute for Chemistry, building A/2 etage, corridor B

mail: akmbo@uni-miskolc.hu, phone: +36-70-544-0171

Target group

The course is offered for all students of the Kerpely Doctoral School, especially in the field of research work that include chemometric prodecures to get information

Language

English or Hungarian.

Scope

The scope of the course is to get acquainted with the modern methods used for discovering the relations to data and data-heaps gathered during measurements and other experimental processes.

Learning outcomes

The students passing the course are expected to understand the basic chemometric concepts and techniques used to examine data and data structures.

Methodology

If the number of students is more than 2, normal lectures will take place using projector and the lecturer will answer the emerging questions of students. If there are 1 or 2 students, then the topics of the course are organised in 3 units. Each unit will be discussed during a separate consultation. After the third occasion there will be a fourth consultation, where the questions of the students will be discussed and answered.

Constituent topics

- Probability, probability variable, distribution, expectable value and variance.
- Specialities of the chemometric tests, the concept of significance level, Gaussian disrtribution, t-distribution, tests for examination of distributions.
- Tests to examine the expectable value (u- and t-tests). Confidence interval.
- Test for variances and standard deviation. ANOVA
- Correlation and independency, correlation and regression, linear and polynomial regression, multivariate correlation.
- Principal-componenst analysis (PCA) and principal component regression.
- Paramertric and non-parametric chemometric tests.
- The concept of outliers and staglers and tests to examine data for outliers and staglers
- The concept of correlation and the concept of being uncorrelated.
- The properties of the correlation coeffitient and test to examine and/or determine its value.

- The so-called rank-correlation test and its application to examination of trend vs time.

Test questions:
1. What is a probability variable?
2. What is the expectable value and variance as well as standard deviation?
3. What is the so-called significance-level?
4. What are the properties of the Gaussian distribution?
5. What tests are for examining Gaussian distribution?
6. What is the main difference between u—tests and t-tests?
7. What is the so-called confidence-interval and which filed is it applied to?
8. What is the so-called Welch-test and when is it applied?
9. What is the difference between the parametric and non-parametric statistical tests?
10. What is an outlier data-item?
11. What are principles of ANOVA test?
12. What is the relation to the correlation and independency of data?
13. What are the important properties of the correlation coeffitient?
14. What is the relation between the correlation and number of data-elements?
15. What is the principle of the rank correlation test?
16. When do we state trand (vs time) in a data-heap?
17. What is the difference between correlation and regression?
18. What are the requirements of the application of the least squares principle?
19. What fields are linear and polinomic regression applied to most frequently?
20. What is the so-called multivariate correlation?
21. What is the principal component analysis is based upon?

Recommended literature

1. Peter C. Meier, Richard E. Zünd.: Statistical Methods in Analytical Chemistry, John Wiley and Sons, Interscience publication, New York, Toronto, 2000
2. Brian Steele, John Chandler, Swarna Reddy : Algorithm for Data Science, Springer International Publishing, Switzerland 2016
3. José Unpingco: Python for Probability, Statistics and Machine Learning, Springer International Publishing, Switzerland, 2016

+ If need be, the student may get literature directly relevant to his/her research.

Completion, examination

During and oral exam, the students gets a topic and after some time to prepare he discusses the information on the topic with the lecturer.

Relevant topics for the complex examination
1. Probability variable, distributions, expectable value, variance and standard deviation, significance-level statistic estimation.
2. Tests for examination of expectable value and variance, the concept of confidence interval and its usage.
3. Principles of tests for outliers and staglers and analysis of variances, their usage and application.

4. The concept and testing of independency and correlation.
5. Correlation, regression, multivariate correlation, principal component analysis.