MATH-710

Data Analysis for Science and Engineering

Davison Anthony, Goldstein Darlene, Morgenthaler Stephan, Panaretos Victor

Cursus	Sem.	Type
Civil & Environmental Engineering		Obl.
Mathematics		Opt.

Language of teaching	English
Credits	4
Session	
Exam	Multiple
Workload	120h
Hours	**56**
Lecture	28
Practical work	28
Number of positions	

Remark

Postponed until further notice

Summary

An overview course intended for scientists and engineers who need to use statistical methods as part of their research, who have already attended a course at the second-year EPFL undergraduate level, and need revision and deepening of their knowledge at a more conceptual level.

Content

This four-credit course is intended for PhD students who need to use statistical ideas and data analysis as part of their research. It is assumed that they have already attended a first course in probability and statistics, at the level of an EPFL second-year course for engineers, and need a broader coverage at a more conceptual level. The course structure is akin to Diggle and Chetwynd (2011), but with different emphases and choices of material.
The course will consist of two classroom hours per week over one semester, plus assigned reading, plus exercises using the statistical package R. Students are expected to submit a problem (which might be a dataset) from their research before the course begins, so that the contents can be tailored to the problems proposed, and the course assessment will be based on a report and presentation in which ideas from the course are applied to the problem.

1. Introduction:
   - Introduction, aims of the course. Statistics and the scientific method. Examples.
   - Presentation of selected problems by students.

2. Looking at data:
   - Exploratory data analysis
   - Elements of graphical data analysis

3. Probability revision: Flipped classroom on basic probability (probability distribution, random variates, conditional distributions, limit theorems), based on assigned reading.
4. Probability models 1:
   - Multivariate distributions
   - Gaussian processes

5. Probability models 2:
   - Poisson process
   - Markov processes

6. Statistics revision: Flipped classroom on basic statistics (point and interval estimation, testing, likelihood), based on assigned reading.

7. Experimental design 1:
   - Basic ideas: randomisation, replication, blocking
   - Simple comparative experiments

8. Experimental design 2:
   - More complex designs
   - Clinical trials

9. Experimental design 3:
   - Significance, power. Multiple testing.
   - Observational vs designed studies

10. Statistical models 1:
    - Linear regression
    - Model-checking, Robust regression

11. Statistical models 2:
    - Model selection (AIC, BIC, cross-validation), Spline regression
    - Logistic regression, Log-linear models

12. Statistical models 3:
    - Generalised additive models, High-dimensional regression (lasso)
    - Mixed models

13. Statistical models 4: (Possible topics, to be determined by needs of participants)
    - Reliability
    - Time failure data

14. Statistical models 5: (Possible topics, to be determined by needs of participants)
    - Time series
    - Time series

References
Diggle, P. J. and Chetwynd, A. G. (2011) Statistics and Scientific Method. Oxford University Press.

Keywords

Data analysis; statistical methods, scientific method

Learning Prerequisites

Required courses

second-year course in probability/statistics for engineers and/or scientists, reasonable mathematical ability

Assessment methods

Project report/Oral presentation