

Statistics Workshop

Date: Monday 15 June – Wednesday 17 June

Online workshop over Microsoft Teams (Links for access see email)

For Aachen participants: Meeting location in Spiegelsaal or Conference Room Arbeitsmedizin (Level 4, Elevator D1 left, Room 4)

Speakers:

Prof. Dr. Georg Heinze, Center for Medical Statistics, Medical University of Vienna, Austria (Main organizer)

Dr. Paul Perco, Medical University of Innsbruck, Austria

Prof. Dr. Florian Frommlet, Center for Medical Statistics, Medical University of Vienna, Austria

Jeppe Christensen (H2020-ITN CaReSyAn)

Giulia Bagarolo (H2020-ITN CaReSyAn)

Max number of participants: 15

| Timeslot | Topic | Speaker |
|--------------------|---|--|
| Monday, 15 June | | |
| 14-14.10 | Welcome & Introduction | Georg Heinze, Heidi Noels & Joachim Jankowski |
| 14.10-14.55 | Introduction to Data Science tasks | Georg Heinze |
| 14.55-15.40 | Estimating population quantities: an experiment | Jeppe Christensen |
| 15.40-16.10 | Coffee break | |
| 16.10-16.55 | Describing data and initial data analysis | Giulia Bagarolo |
| 16.55-17.40 | Introduction to R and R-Markdown to create reproducible research reports | Jeppe Christensen |
| | | |
| Tuesday, 16 June | | |
| 9-10.30 | Bioinformatics and Systems Biology (incl. discussion) <i>OPEN SESSION</i> | Paul Perco, Medical University of Innsbruck, Austria |
| 10.30-11.00 | Coffee break | |
| 11.00-12.30 | T-tests and ANOVA | Georg Heinze |
| 12.30-14.00 | Lunch break | |
| 14-15.30 | P-values: the only thing that matters? (incl. discussion) | Georg Heinze & Jeppe Christensen |
| 15.30-16.00 | Coffee break | |
| 16-17.30 | Introduction to regression analysis | Georg Heinze |
| | | |
| Wednesday, 17 June | | |
| 9.00-10.30 | Reproducibility of animal experiments <i>OPEN SESSION</i> | Florian Frommlet |
| 10.30-11 | Coffee Break | |
| 11-11.45 | Multivariable regression | Georg Heinze |
| 11.45-12.30 | An outlook to high-dimensional regression <i>Conferenceroom Arbeitsmedizin</i> | Georg Heinze |

Summary of Invited Speaker Talks

Dr. Paul Perco (Medical University of Innsbruck, Austria)

Network-based disease and drug modeling supporting biomarker and drug target identification

High-throughput technologies such as microarrays, mass spectrometry, or next-generation sequencing enable the parallel measurement of thousands of molecular features on the level of the genome, transcriptome, proteome, and metabolome.

Initial analyses of these large datasets focused on statistical procedures for identifying maximally dysregulated molecular features between disease and healthy subjects under study followed by functional interpretation. Analyses procedures in biomedical research more and more focus on the integrated analysis of data from multiple sources. Molecular profiling data are joined with clinical and histological data, using on top information from literature mining approaches or publicly available biological databases. Making further use of protein interaction data, a new discipline in biomedical informatics termed network biology originated. Aims of these integration approaches are among others (i) an improved disease classification making use of molecular data, (ii) the elucidation of molecular disease processes, (iii) the identification of biomarkers and drug targets, (iv) the repositioning of approved drugs, or (v) the stratification of patients in clinical trials to increase drug efficacy.

In this talk a brief introduction to graph theory will be given followed by the introduction of different biological networks such as co-expression networks, protein-protein interaction networks, or literature co-annotation networks. Bioinformatics data integration concepts will be discussed analysing Omics data in the context of biological networks. Application examples in the context of kidney disease, cardiovascular disease but also in the area of oncology will be presented.

Prof. Dr. Florian Frommlet (Medical University of Vienna, Austria)

Reproducibility in Animal Trials

In recent years there has been an ongoing discussion about the fact that too many results from published preclinical studies turn out to be false positives. This culminated in the proclamation that there is a reproducibility crisis. But what exactly is meant by reproducibility? Which factors contribute to the problem that many published results cannot be reproduced? We will have a look at these questions and also discuss some potential remedies which might improve the situation for the future.