Project Multivariate and Hierarchical Data Discovering Associations

Geert Molenberghs, Geert Verbeke, Olivier Thas, Yannick Vandendijck, Steven Abrams, Johan Verbeeck and Marsha Nugroho

Project: Multivariate and Hierarchical Data

2023-2024

Overview of the course

See Syllabus on Blackboard

- 3 main topics in the classes
  - Multivariate methods $\longrightarrow$ Olivier Thas
  - Repeated measures $\longrightarrow$ Geert Verbeke
  - Clustered data $\longrightarrow$ Geert Molenberghs

- 2 projects in the course
  - Small individual assignment $\longrightarrow$ Yannick Vandendijck
  - **Discovering associations** on Multivariate, Repeated measures & Clustered data $\longrightarrow$ Yannick Vandendijck, Steven Abrams, Johan Verbeeck, Marsha Nugroho

Overview of the project

- Project learning from data
  - Apply methods introduced in Concept of Probability and Statistics, Linear Models to real-life data set
- This project:
  - Group project
  - Design your own experiment to address a scientific question (issues: what sample size is needed, what statistical design, what is a good primary endpoint, etc?)
  - Apply data analysis methods introduced in Hierarchical Data and Generalized Linear Models
  - Scientific reporting (writing, presentations, ethical aspects)

Organization of the project - lectures

1. Introduction to sample size calculations
   - (Video) Lectures by Y. Vandendijck
   - **Individual assignment** (submit in Blackboard **6/03/24** at 11h59pm CET)
2. Ethical aspects of consulting
   - (Video) Lectures by G. Molenberghs
3. Statistical research for pharmaceutical R&D
   - (Video) Lectures by Y. Vandendijck

Organization of the project - guided project

4. Training statistical consultancy
   - 1-day workshop (Y. Vandendijck, S. Abrams, J. Verbeeck & M. Nugroho)
     - 28/03/2024 (on-campus)
     - 25/03/2024, 26/03/2024 or 29/03/2024 (Distance Learning)
   - In groups of $4-6$ students
   - Design experiment, elicit relevant information from scientist
   - **Group assignment**: 1-page report at the end of the day with proposed approach (Design of experiment, Statistical Analysis Plan (SAP))

5. Statistical help desk (optional)
   - **Group appointment** to assist with sample size computations, 11/04/2024 (S. Abrams, J. Verbeeck) $\longrightarrow$ 15 minutes
   - Come up with final sample size for data set

Organization of the project - guided project

6. Presentation I
   - 18/04/2024
   - **Group assignment**: present proposed design + sample size computation (send presentation before 17/04 at 17h CET by email to marsha.nugroho@uhasselt.be)
   - Based on this work: final data set provided

7. Presentation II
   - 16/05/2024
   - **Group assignment**: presentation on exploratory data analysis (EDA) and (first) statistical analyses (send presentation before 15/05 at 17h CET by email to marsha.nugroho@uhasselt.be)

Organization of the project - guided project

11. Final written report
    - 27/05/2024 17h CET
    - **Group assignment**: written report that includes all relevant previous work
        - Max. 25 pages, including Appendix
        - Fixed template (LaTeX) on Blackboard
        - Only include relevant R/SAS code
        - Submit report in Blackboard

Summary of important deliverables

Due date	Type of assignment	Topic
6/03	Individual assignment, handwritten	Sample size
25/03 - 29/03	Group assignment, written	Protocol experiment, SAP
18/04	Group assignment, presentation	Design of experiment
16/05	Group assignment, presentation	EDA + statistical analysis
27/05	Group assignment, written	Final report