Data Analysis Courses

2023 - 2024









Index

- **3** City map Ghent & Contact information
- 5 Micro-credentials, KMO-portefeuille & Doctoral Schools
- 6 Module 1 Inleiding tot SPSS
- 8 Module 2 Getting Started with R Software for Data Analysis
- 10 Module 3 Design and Analysis of Randomized Clinical Trials
- 12 Module 4 Drawing Conclusions from Data: an Introduction
- 14 Module 5 Getting Started with Python for Data Scientists
- 16 Module 6 Leverage your R Skills: Data Wrangling & Plotting with Tidyverse
- 18 Module 7 Dynamic Report Generation with R Markdown
- 20 Module 8 Exploiting Sources of Variation in your Data: the ANOVA Approach
- 22 Module 9 Getting Started with NVivo for Qualitative Data Analysis
- 24 Module 10 High Dimensional Data Analysis
- 26 Module 11 Explaining and Predicting Outcomes with Linear Regression
- 28 Module 12 Machine Learning with Python
- **30** Module 13 Structural equation modeling with lavaan





City Map Ghent

CAMPUSES GHENT

1 Het Pand

4

5

7

2 Campus Aula (Aula) **

3 Campus Ufo (Rectoraat, Ufo, De Brug) *

Campus Boekentoren (University Library) **

Campus Tweekerken (Therminal)

Campus Rommelaere

6 Campus Mercator *

Campus Dunant (GUSB) *

89 Campus Coupure *

Campus Ledeganck (GUM en Botanical Garden) **

Campus Schoonmeersen *

Campus Sterre (UGent second hand shop) *

Campus UZ Gent **

Campus Heymans **

Campus Proeftuin

Tech Lane Ghent Science Park – Campus Ardoyen *

RESTAURANTS AND CAFETERIAS

- Campus with resto
- ** Campus with cafetaria
- (17) HoGent Resto De Wijnaert

HoGent Resto Bijloke (18)

(19) HoGent Resto Vesalius

UNIVERSITY HALLS OF RESIDENCE

- Homes Groningen, Uppsala, Göttingen, Canterbury 20
- 21 Homes Fabiola, Vermeylen, Heymans
- 22 Home Astrid
- Home Bertha De Vriese 23
- 24 Home Boudewijn

PUBLIC AND CITY SERVICES

- 25 De Fietsambassade: bicycle rent / repair
- 26 Station Gent-Sint-Pieters
- ICC (International Convention Center) 27
- 28 Tourist office City of Ghent (Sint-Veerleplein 5)
- Administration Centre City of Ghent / (29
- Foreigners Registration Office (Woodrow Wilsonplein 1)
- De Krook / Municipal Library (Miriam Makebaplein 1) (30)

Contact



Krijgslaan 281, 9000 Ghent T 09 264 44 26 F 09 264 85 90 <u>ipvw.ices@UGent.be</u> www.ipvw-ices.UGent.be



Micro-credentials

Designed for the Life Long Learning audience, our course modules now extend to include select UGent courses and other structured learning options, collectively known as 'Microcredentials.' Micro-credentials are focused qualifications that validate specific skills and knowledge in a subject area. They offer a flexible and efficient way for individuals to showcase expertise and enhance professional development, accredited by Ghent University.

Some data analysis modules below are integrated into a micro-credential. If you are planning on registering for multiple modules, consider enrolling for the micro-credential instead.

Micro-credential: Module 3: Design and Analysis of Randomized Clinical Trials

Micro-credential: Data Analysis in R: Basics and Beyond

- Module 2 Getting Started with R Software for Data Analysis
- Module 6 Leverage your R Skills: Data Wrangling & Plotting with Tidyverse
- Module 7 Dynamic Report Generation with R Markdown and Quarto

Micro-credential: Applied Statistics: from Basics to Regression Modelling

- Module 4 Drawing Conclusions from Data: an Introduction
- Module 8 Exploiting Sources of Variation in your Data: the ANOVA Approach
- Module 10 Explaining and Predicting Outcomes with Linear Regression

KMO-portfolio

The Flemish Community recognizes the value of lifelong learning for the region's industry and economic development. Employers are granted financial support through the government's introduction of the KMO-portfolio. More about this stimulation initiative can be found on the website: www.kmo-portefeuille.be (in Dutch).

Doctoral Schools

As UGent PhD student you can incorporate some modules as 'specialist course' or 'transferable skills seminar: research & valorization' in your Doctoral Training Program (DTP). To get a refund of the registration fee from your Doctoral School (DS) please follow these strict rules and take the necessary action in time.

m 5

Inleiding tot SPSS

Deze cursus geeft je een stevige basis om je eigen onderzoeksvragen te vertalen naar een kwalitatief databestand en om zelf je SPSS vaardigheden verder te ontwikkelen in functie van je analysenoden. De introductie is ervaringsgericht opgevat. De deelnemers worden met een aantal problemen geconfronteerd waarna mogelijke oplossingen besproken en gedemonstreerd worden.

Sessie 1 = data en databestanden

Het verzamelen en opslaan van onderzoeksgegevens is vaak niet zo evident als het lijkt. Zeker niet als het de bedoeling is de gegevens later op een professionele manier te verwerken. In de eerste sessie krijg je een overzicht van de vele mogelijkheden van SPSS en leer je er zelf je weg in vinden. Na deze sessie kun je vlot data in SPSS invoeren en definiëren, data uit andere programma's importeren en gebruiken, en variabelen herschrijven en combineren. Sessie 2 = een steekproef beschrijven en verbanden bekijken

De kracht van SPSS zit in de syntax achter de interface. In de tweede sessie maak je kennis met deze syntax: waar vind je die en wat kan deze syntax voor jou betekenen? Je leert ook hoe je je steekproef grafisch en statistisch kunt beschrijven, en hoe je de resultaten van je beschrijfopdrachten moet interpreteren. Na deze sessie weet je hoe je SPSS-syntax kunt gebruiken, heb je zicht op de beschikbare beschrijvende statistieken, en kun je de gewenste statistieken genereren en interpreteren.

Sessie 3 = gemiddelden vergelijken

De derde sessie geeft je een voorsmaakje van wat statistiek écht voor je kan doen. Je wordt ondergedompeld in een bad van ttoetsen en ANOVA met de bedoeling de kennis, vaardigheden en inzichten uit de vorige sessies samen te brengen en verder uit te diepen.

DOELPUBLIEK

Deze introductie is gericht op alle personen die gegevens inzamelen en/of opslaan, met de bedoeling deze statistisch te analyseren en interpreteren.

LESMATERIAAL

Documentatie- en oefeningenbundel.

WANNEER?

- Maandag 2 oktober 2023: 17u 20u30
- Maandag 9 oktober 2023: 17u 20u30
- Maandag 16 oktober 2023: 17u 20u30

WAAR?

Faculteit Psychologische en Pedagogische Wetenschappen, Henri Dunantlaan 1, Gent, Auditorium 3

LESGEVER



Prof. Kris Erauw (UGent) is beleidsmedewerker aan de Faculteit Psychologie en Pedagogische Wetenschappen. Hij stond jarenlang mee in voor de begeleiding van studenten bij de vakken statistiek en methodologie en bij het schrijven van hun masterproef. Vandaag is hij onder andere actief in het ontwikkelen van online oplossingen voor dataverzameling en



dataverwerking.

PRIJZEN

Industrie	€ 540
Non-profit, overheid en personeel hoger onderwijs	€ 405
(Doctoraat)studenten	€ 325

Deelnameprijs omvat lesgeld, hand-outs, frisdranken, koffie en tussendoortje.

TAAL

De opleiding wordt volledig in het Nederlands gedoceerd.

INSCHRIJVING

<u>Beta-Academy</u>

Micro-credential: Data Analysis in R: Basics and Beyond

Module 2

Getting Started with R Software for Data Analysis

This course introduces the use of the R environment for the implementation of data management, data exploration, basic statistical analysis and automation of procedures. It starts with a description of the R GUI, the use of the command line and an overview of basic data structures. The application of standard procedures to import data or to export results to external files will be illustrated.

Creation of new variables, subsetting, merging and stacking of data sets will be covered in the data management section. Exploration of the data by histograms, box plots, scatter plots, summary numbers, correlation coefficients and cross-tabulations will be performed.

Simple statistical procedures that will be covered are:

- comparisons of observed group means (t-test, ANOVA and their non-parametric versions) and proportions
- test for independence in 2-way cross tables and linear regression (focusing on the Rimplementation of the statistical methods that are the subject of other modules of the statistics series)

Finally, installing new packages and automation of analysis procedures will also be discussed. Practical sessions and specific exercises will be provided to allow participants to practice their R skills in interaction with the teacher.



This course targets professionals and investigators from diverse areas with little to no R-programming experience who wish to start using R for their data manipulation, data exploration or statistical analysis.

COURSE PREREQUISITES

The course is open to all interested persons. Knowledge of basic statistical concepts and experience with other programming languages are considered advantages, but not required for learning the R language.

WHEN?

- 23rd of October 2023: 5.30 pm 9 pm
- 26th of October 2023: 5.30 pm 9 pm
- 30th of October2023: 5.30 pm 9 pm
- 6th of November 2023: 5.30 pm 9 pm

Where?

Faculty of Science, Campus Sterre, Krijgslaan 281, 9000 Ghent, Building S9, 3rd floor, Auditorium 3

TEACHER



Dr. Emmanuel Abatih is a post-doctoral research fellow at Ghent University

and he works as the FLAMES coordinator for UGent and as a statistical consultant for FIRE and DASS. He obtained a PhD in Epidemiology and Biostatistics in 2008 at the University of Copenhagen and has been teaching courses ranging from general statistics to more specialized areas like Machine Learning, Causal Inference and Structural Equations Modeling. He has experience with R, python, SPSS and STATA.

PRICES

Industry	€ 620
Non profit, government, higher education staff	€ 465
(Doctoral) students, unemployed	€280

LANGUAGE

This course will be taught in English

REGISTRATION

<u>Beta-Academy</u>



<u>______</u> 9

Micro-credential

Module 3

Design and Analysis of Randomized Clinical Trials

This course builds on the course on Principles of statistical data analysis, Analysis of continuous data and Categorical Data Analysis to teach knowledge and ideas to draw evidence from randomized studies with a focus on applications in medicine and public health. It will provide the student with an understanding of statistical principles, methodology and concepts involved in clinical research across the various phases. The student should acquire the skills to design a clinical trial, be aware and understand the limitations related to clinical trials, be aware of formal guidelines and be able to analyze the data according to appropriate statistical techniques.

Detailed content:

- Clinical Trials
- Protocol document and the role of the statistician therein
- Types of experimental designs (parallel, cross-over, factorial)
- Endpoints and estimands
- Sample Size/Power calculations
- Treatment allocation
- Data analysis and prognostic factors
- Concepts and methodology for (data-dependent) stopping and early termination of trials
- (interim analysis, adaptive designs and data monitoring committees)
- Meta-analysis
- Reporting of results, potential pitfalls and data quality



This course is aimed at anyone involved in setting up, running or evaluating clinical trials, looking for a better understanding of the underlying statistical principles that guide choices in design and analysis.

COURSE PREREQUISITES

The basics of statistical concepts, techniques and inference, knowledge of linear regression and categorical data analysis are considered known.

₩HEN?

- 25th of October 2023
- 8th, 22nd & 29th of November 2023
- 6th, 13rd & 20th of December 2023 from 8.30 am to 11.30 am

Where? (0)

Faculty of Science, Campus Sterre, Krijgslaan 281, 9000 Ghent, Building S12, lecture room 0.2

TEACHERS

- Prof. An Vandebosch
- Dries Reynders

PRICE

€ 890

LANGUAGE

This course will be taught in English



<u>UGent</u>





Micro-credential: Applied Statistics: from Basics to Regression Modelling

Module 4

<u></u> 12

Drawing Conclusions from Data: an Introduction

This course aims to provide insight into basic statistical concepts with emphasis on practical applications. Mathematical formulae will be kept to a minimum. The theory and the methods of analysis will be extensively illustrated with examples relating to a wide variety of different fields. To emphasize the practical approach in this course all classes will take place in a pc room.

The first session will be dedicated to getting to know the software package R. We start with concise graphical and numerical descriptions of data obtained from observational or experimental studies. The most common and frequently used probability distributions of discrete and continuous variables will be presented. Statistical inference draws conclusions about a population based on sampled data. Chance variations are taken into account such that a level of confidence is attached to these conclusions. We present the reasoning behind significance tests for the comparison of observed data with a hypothesis, the validity of which we want to assess. We apply this procedure to data obtained either from one or from two populations. The correct use of the t-test will be discussed. Nonparametric methods are considered as a possible alternative in case the requirements of the t-test are not met. We cover the basic concepts of hypothesis testing for categorical data, including the chisquare test. Quite often the relationship between two variables, where the outcome of one variable is seen as depending on the value of the other, is the focus of scientific interest. We will give an introduction to linear regression analysis, where a regression line based on observations obtained in a sample describes this relation. Hands-on exercises are worked out behind the PC using the R software.

This course will benefit professionals and investigators from diverse areas, research scientists, clinical research associates, investing in data handling and wishing to acquire insight into basic statistical methods or to refresh their knowledge and practice of statistics.

COURSE PREREQUISITES

The course is open to all. It is necessary to have an understanding of basic algebra (basic rules, solving equations, ...), exponents and square roots.

₩HEN?

- 7th, 14th, 21st & 28th of November 2023
- 5th & 12th of December 2023

from 5.30 pm - 9.30 pm

Where?

Faculty of Science, Campus Sterre, Krijgslaan 281, 9000 Ghent, Building S9, 3rd floor, Auditorium 3

TEACHER



Dr. Els Adriaens (Adriaens Consulting bvba) studied biology, obtained a PhD

in pharmaceutical sciences and a Master in Statistical Data Analysis at Ghent University.

She is consultant in statistical data analysis specialized in the field of the development and validation of alternatives to laboratory animals.

PRICES

Industry	€ 1320
Non profit, government, higher education staff	€990
(Doctoral) students, unemployed	€ 595

LANGUAGE

This course will be taught in English

REGISTRATION

<u>Beta-Academy</u>



<u></u> 13

Getting Started with Python for Data Scientists

Python started off as a general-purpose programming language, but in the last decade it has become a popular environment for data science. The reason is that the community of Python users have recently created useful add-on packages which are suitable for data manipulation, preparation, visualization and analysis. This practical course introduces both base Python and the most important packages in a hands-on way with many exercises.

The contents of the course are:

- Introduction: Python and the Anaconda distribution
- Data types: numbers, strings, lists, tuples, sets and dictionaries
- Automation: control flow and self-defined functions
- Importing data and exporting results
- Managing data with NumPy and pandas
- Graphs with matplotlib and seaborn
- Statistical analysis with statsmodels

The objective of the course is that you are capable of doing data management, visualization and analysis in Python on your own.

Python is an open-source programming language which you can freely <u>download</u> (i.e. the Anaconda distribution). Python version 3 or higher is recommended.



This course targets professionals and investigators from diverse areas with little to no Python-programming experience who wish to start using Python for their data manipulation, data exploration or statistical analysis.

COURSE PREREQUISITES

The course is open to all interested persons. Knowledge of basic statistical concepts and experience with other programming languages are considered advantages, but not required for learning the Python language.



• 4th, 7th, 11th, 14th & 18th of December 2023 from 5.30 pm - 9.30 pm

Where?

Faculty of Science, Campus Sterre, Krijgslaan 281, 9000 Ghent, Building S9, 3rd floor, Auditorium 3

LANGUAGE

This course will be taught in English

TEACHER

Dr. Koen Plevoets is a post-doctoral researcher at the Department of



PRICES

Translation, Interpreting and Communication of Ghent University. He obtained his PhD in linguistics in 2008 and he has specialized master's degrees in Artificial Intelligence and in Statistics. He has over 15 years of experience in categorical data analysis, multivariate statistics and text mining. His interests are visualizations of complex data, for which he uses the open-source programming languages R and Python.

Industry€ 1000Non profit, government, higher
education staff€ 750(Doctoral) students, unemployed€ 450

REGISTRATION

<u>Beta-Academy</u>



<u></u> 15

Micro-credential: Data Analysis in R: **Basics and Beyond**

Module 6

Leverage your R Skills: Data Wrangling & Plotting with Tidyverse

Tidyverse is a collection of R-packages used for data wrangling and visualization that share a common design philosophy. The goal of this course is to get you up to speed with the most upto-date and essential tidyverse tools for data exploration. After attending this course, you'll have the tools to tackle a wide variety of data wrangling and visualization challenges, using the best parts of R tidyverse.

This course covers the most essential tools from 3 main R tidyverse packages that are frequently used in general data analysis procedure.

Lectures with R code demonstrations are blended with hands-on exercises which allows you to try out the tools you've seen in the class under guides.

What you will learn:

- Data transforming and summarizing with dplyr: narrowing in on observations of interest, creating new variables that are functions of existing variables, and calculating a set of summary statistics (like counts or means)
- Data visualization with ggplot2: creating more informative graphs (e.g., scatter plot, bar plot, histogram, smoother/regression line, ...) in an elegant and efficient way. Arranging multiple plots on a grid
- Data ingest and tidying with tidyr: storing it in a consistent form that matches the semantics of the dataset with the way it is stored.
- Extra tools for programming: Merging and comparing two datasets based on various matching or filtering criterion. Other useful tools for R programming.



This course targets anyone who wants to use R for data processing and needs to produce professional looking graphs and/or summary statistics.

COURSE PREREQUISITES

The course is open to all interested persons. Basic R skills as provided in <u>Module 2</u> of this year's program are strongly advised.



18th, 19th & 21st of December 2023
from 1 pm to 4 pm.

Where?

Faculty of Science, Campus Sterre, Krijgslaan 281, 9000 Ghent, Building S9, 3rd floor, room 3.1 & Auditorium 3

TEACHER



Dr. Limin Liu is a postdoc researcher at the Center for Statistics at Ghent University. She studied social work and sociology in Beijing and Berlin where she has accumulated several years of research experience in the field of sociology of eduction, social mobility and stratification. Upon her empirical research experience, she achieved a Master in Statistical Data Analysis at Ghent University. Since 2019, she has joined the team of statisticians and focused on both qualitative and quantitative research methods.

PRICES

Industry	€ 495
Non profit, government, higher education staff	€ 370
(Doctoral) students, unemployed	€ 225

LANGUAGE

This course will be taught in English

REGISTRATION<u>Beta-Academy</u>

<u>m</u> 17

Micro-credential: Data Analysis in R: Basics and Beyond

Module 7

Dynamic Report Generation with R Markdown and Quarto

R offers many first class features for statistics and data science. One of these, is certainly Rmarkdown, that allows seamless integration of analysis (code) and text. This greatly improves reproducibility, reduces copy-paste- and others errors and enhances possibilities for automation.

R markdown offers three main types of output: pdf, html and docx. The first session introduces the basic framework, the output-specific possibilities, the bookdown-extension and the easy and rewarding move to its recent less platform dependent successor Quarto.

The second session explores some general approaches for automation (using self-built templates for report-sections or complete reports) and presents Officedown. The latter is less flexible than Rmarkdown, but offers more options for docx-output.



This module targets anyone who wants to produce professional reports using R.

COURSE PREREQUISITES

Basic knowledge of R as provided in <u>Module 2</u> of this year's program is required. Knowledge of tidyverse as provided in <u>Module 6</u> is helpful.



 22nd of December 2023 from 9 am to 12 pm and from 1 pm to 4 pm

Where?

Faculty of Science, Campus Sterre, Krijgslaan 281, 9000 Ghent, Building S9, 3rd floor, room 3.3 Konrad Zuse

TEACHER



Dries Reynders studied Physics and Statistical Data Analysis at Ghent University.

He is an experienced teacher and is, in that role, well trained in explaining the link between mathematics and the reality it describes. Currently, he works as statistical consultant for the Stat-Gent consortium.

PRICES

Industry	€ 330
Non profit, government, higher education staff	€ 250
(Doctoral) students, unemployed	€ 150

LANGUAGE

This course will be taught in English

REGISTRATION

<u>Beta-Academy</u>





Micro-credential: Applied Statistics: from Basics to Regression Modelling

Module 8

Exploiting Sources of Variation in your Data: the ANOVA Approach

To emphasize the practical approach in this course all classes will take place in a pc room. Analysis of variance (ANOVA) is a statistical tool used in the comparison of means of a random variable over populations that differ in one or more characteristics (factors), e.g. treatment, age, sex, subject, etc.

First, we cover one-way ANOVA, where only one factor is of concern. Depending on the type of the factor, the conclusions pertain to just those factor levels included in the study (fixed factor model), or to a population of factor levels of which we observed a sample (random effects model).

In two-way and multi-way ANOVA where populations differ in more than one characteristic, the effects of factors are studied simultaneously. This yields information about the main effects of each of the factors as well as about any special joint effects (factorial design). We also consider nested designs, where each level of a second (mostly random) factor occurs in conjunction with only one level of the first factor. One special challenge in multi-way ANOVA lies in verifying the assumptions that must be satisfied.

In this course we will focus on correct execution of data analysis and understanding its results. We pay attention to expressing these conclusions in a correct and understandable way. The different methods will be extensively illustrated with examples from scientific studies in a variety of fields. Exercises are worked out behind PC using the R software.

<u></u>20

This course targets professionals and investigators from diverse areas, who need to use statistical methods in the collection and handling of data in their research, in particular for assessing the effect of e.g. different treatments.

COURSE PREREQUISITES

Participants are expected to have an active knowledge of the basic principles underlying statistical strategies, at a level equivalent to <u>Module 4</u> of this year's program. Some R skills are advised consistent with the course content of <u>Module 2</u> of this year's program.



- 9th, 16th, 23rd & 30th of January 2024
- 6th of February 2024

from 5.30 pm - 9.30 pm

Where?

Faculty of Science, Campus Sterre, Krijgslaan 281, 9000 Ghent, Building S9, 3rd floor, Auditorium 3

LANGUAGE

This course will be taught in English

TEACHER



Dr. Els Adriaens (Adriaens Consulting bvba) studied biology, obtained a PhD in pharmaceutical sciences and a Master in Statistical Data Analysis at Ghent University.

She is consultant in statistical data analysis specialized in the field of the development and validation of alternatives to laboratory animals.

PRICES

Industry	€ 1100
Non profit, government, higher education staff	€ 825
(Doctoral) students, unemployed	€ 495



Beta-Academy



<u>m</u> 21

Getting Started with NVivo for Qualitative Data Analysis

NVivo is a widely used computer assisted qualitative data analysis software package which provides a potentially useful tool for the management and analysis of qualitative research data. This course is intended as a basic introduction to using NVivo for qualitative data analysis. Whether you are completely new to NVivo or have some previous experience with it, you will find this course both useful and enjoyable. This course blends lectures with hands-on exercises which allows you to try out the tools you've seen in the class under guidance.

What you will learn:

At the end of this course you will master the core functionalities to apply the latest version of NVivo (1.0) to your project, including:

- Import Creating a research project and importing different data formats such as Word documents, PDFs, webpages, audio, video and images into NVivo; classifying data files and managing their classifications
- Organize Organizing codes, code text and create codes; apply coding stripes and highlights; use cases with classification and attributes; make annotations and memos, create sets and links to files
- Explore Exploring lexical queries, word frequency and text search; apply code and matrix queries; illustrate with visualizations such as mind maps, concept maps, and coding matrix charts; coordinate team work by applying coding comparison



Young researchers and data analysts who are new to qualitative research and curious about NVivo.

COURSE PREREQUISITES

There are no course prerequisites for this course. Anyone can join.

SOFTWARE

It is advised to bring your own laptop to class. If you don't have access to NVivo through your employer you can download the NVivo 14-day free trial for Windows and Mac via this <u>link</u>.



• 1st of February 2024 from 9 am to 4 pm

Where?

Faculty of Science, Campus Sterre, Krijgslaan281, 9000 Ghent, Building S9, 3rd floor, room3.3 Konrad Zuse

LANGUAGE

This course will be taught in English

TEACHER

Dr. Limin Liu is a postdoc researcher at the Center for Statistics at Ghent



University. She studied social work and sociology in Beijing and Berlin where she has accumulated several years of research experience in the field of sociology of eduction, social mobility and stratification. Upon her empirical research experience, she achieved a Master in Statistical Data Analysis at Ghent University. Since 2019, she has joined the team of statisticians and focused on both qualitative and quantitative research methods.

PRICES

Industry	€ 330
Non profit, government, higher education staff	€ 250
(Doctoral) students, unemployed	€ 150

REGISTRATION





High Dimensional Data Analysis

Modern high throughput technologies easily generate data on thousands of variables; e.g. health care data, genomics, chemometrics, environmental monitoring, web logs, movie ratings, etc. Conventional statistical methods are no longer suited for effectively analysing such high-dimensional data. Multivariate statistical methods may be used, but for often the dimensionality of the data set is much larger than the number of (biological) samples. Modern advances in statistical data analyses allow for the appropriate analysis of such data. Methods for the analysis of high dimensional data rely heavily on multivariate statistical methods. Therefore a large part of the course content is devoted to multivariate methods, but with a focus on high dimensional settings and issues. Multivariate statistical analysis covers many methods. In this course a selection of techniques is covered based on our experience that they are frequently used in industry and research institutes. The course is taught using case studies with applications from different fields (analytical chemistry, ecology, biotechnology, genomic, etc.).

1. Dimension reduction: Singular Value Decomposition (SVD), Principal Component Analysis (PCA), Multidimensional Scaling (MDS) and biplots for dimension-reduced data visualisation

2. Sparse SVD and sparse PCA

- 3. Prediction with high dimensional predictors: principal component regression; ridge, lasso and elastic net penalised regression methods
- 4. Classification (prediction of class membership): (penalised) logistic regression and linear discriminant analysis
- 5. Evaluation of prediction models: sensitivity, specificity, ROC curves, mean squared error, cross validation

6. Clustering

7. Large scale hypotheses testing: FDR, FDR control methods, empirical Bayes (local) FDR control



This course targets professionals and investigators from all areas that are highdimensional. Course prerequisites are ready at hand knowledge of basic statistics: data exploration and descriptive statistical statistics, modeling, and inference: linear models, confidence intervals, t-tests, F-tests, anova, chisquared test, such as covered in M2: Introductory Statistics with R (or SPSS), M5: Analysis of Variance with R (or SPSS) and M12: Applied Linear Regression) of this years' course program.



6st, 8th, 13th, 15th, 20th & 22nd of February
2024 from 5.30 pm to 9,30 pm

Where?

Faculty of Science, Campus Sterre, Krijgslaan 281, 9000 Ghent, Building S1 Room 3.2 & S9 Auditorium 3.

TEACHER



Prof. Lieven Clement is an Associate Professor of Statistical Genomics at Ghent University. He specializes in developing statistical methods and open source tools for omics data analysis, focusing on (single cell) transcriptomics and proteomics. He

played a key role in establishing the Master of Science in Bioinformatics at Ghent University and has extensive experience teaching statistics, statistical genomics, and high-dimensional data analysis to life science and statistical data analysis students. He also conducts short courses on statistics and proteomics data analysis in renowned bioinformatics programs across Europe, including the Wellcome Trust Advanced Courses and Gulbenkian Institute Training Programme.

PRICES

Industry	€ 1320
Non profit, government, higher education staff	€990
(Doctoral) students, unemployed	€ 595

LANGUAGE

This course will be taught in English

REGISTRATION

<u>Beta-Academy</u>



Micro-credential: Applied Statistics: from Basics to Regression Modelling

Module 11

Explaining and Predicting Outcomes with Linear Regression

Linear regression addresses how a continuous dependent variable is associated by one or more predictors of any type. The fact that many practical problems deal with continuous outcomes (e.g. income, blood pressure, temperature, affect) makes linear regression a popular tool, and most of us will be familiar with the concept of drawing a line through a cloud of data points.

The first two sessions of this module introduce the conceptual framework of this method using the simple case of a single predictor. Formulas and technicalities are kept to a minimum and the main focus is on interpretation of results and assessing model validity. This includes confidence statements on the predictor effect (hypothesis tests and confidence intervals), using the regression model to predict future results and verification of model assumptions. In session 3 and 4 we allow for more than one predictor leading to the multiple linear regression model. We focus on either explanation or prediction. How to come to a parsimonious model starting from a large number of predictors will be discussed in detail. In these complex linear models special attention will be given to interpreting individual predictor effects, as they critically depend on other terms in the model and underlying relations between predictors (confounding).

In the last session a more elaborate data analysis is discussed. We touch on problems where linear regression is not appropriate and replaced by related approaches such as generalized linear models and mixed models.

Different features will be illustrated with case examples from the instructors practical experience, and participants are encouraged to bring examples from their own work. Hands-on exercises are worked out behind the PC using the R software.



This course targets professionals and investigators from all areas who are involved in prediction problems or need to model the relationship between a dependent variable and one or more explanatory variables.

COURSE PREREQUISITES

Participants are expected to have an active knowledge of the basic principles underlying statistical strategies, at a level equivalent to the <u>Module 4</u> of this program.



- 29th of February 2024
- 7th, 14th, 21st & 28th of March 2024 from 5.30 pm to 9.30 pm.

Where?

Faculty of Science, Campus Sterre, Krijgslaan 281, 9000 Ghent, Building S9, 3rd floor, Auditorium 3





Dries Reynders studied Physics and Statistical Data Analysis at Ghent University.

He is an experienced teacher and is, in that role, well trained in explaining the link between mathematics and the reality it describes. Currently, he works as statistical consultant for the Stat-Gent consortium.

PRICES

Industry	€ 1100	
Non profit, government, higher education staff	€ 825	
(Doctoral) students, unemployed	€ 495	

LANGUAGE

This course will be taught in English

REGISTRATIONBeta-Academy



<u>m</u> 28

Machine Learning with Python

Many modern digital applications increasingly rely on artificial intelligence and machine learning as a means to derive predictive strength from high-dimensional data sets. Compared to traditional statistics, the absence of a focus on scientific hypotheses, and the need for easily leveraging detailed signals in the data require a different set of models, tools, and analytical reflexes.

This course aims to bring participants to the level where they can independently tackle the modeling part of machine learning projects. This means that the most common types of projects will be addressed - regression-type with continuous outcomes, classification with categorical outcomes, and clustering. For each of these, the practical use of a set of standard methods will be shown, like Random Forests, Gradient Boosting Machines, Support Vector Machines, k-Nearest-Neighbors, K-means,... Furthermore, throughout the course, concepts will be highlighted that are of concern in every machine learning project, like the curse of dimensionality, model capacity, overfitting and regularization, and practical strategies will be offered to deal with them, introducing techniques such as the Lasso and ridge regression, cross-validation, bagging and boosting. Instructions will also be given on a selection of specific techniques that are often of interest, such as modern visualization of high-dimensional data with algorithms like UMAP, model calibration, outlier detection using isolation forests, explanability of black-box models,... Finally, the last lecture will introduce the idea of deep learning as a powerful tool for data analysis, discussing when and how to practically use it, and when to shy away from it.

This course targets professionals and investigators from all areas that are involved in predictive modeling based on large and/or high-dimensional databases.

COURSE PREREQUISITES

Participants are expected to be familiar with basic statistical modeling (as for instance taught in <u>Module 4</u> of this program), and to have had a first experience programming in Python (as for instance taught in <u>Module 5</u> of this program).



- 15th, 22nd & 29th of April 2024
- 6th, 13th & 27th of May 2024
- 3rd of June 2024

from 5.30 pm to 9 pm.

Where?

Faculty of Science, Campus Sterre, Krijgslaan 281, 9000 Ghent, Building S9, 3rd floor, Auditorium 3

TEACHER

Dr. Bart Van Rompaye is the Head of Advanced Analytics and Machine Learning



at KPMG, leading a diverse team of data scientists. Previously, he served as a Lead Data Scientist at KBC Group for 6 years, where he developed innovative products like Matti, Indigo (Czech Republic), and Belgium's first AI-assisted investment fund. Bart earned his PhD in survival analysis from Ghent University and held postdoctoral positions at Ghent University and Umea University, Sweden. He has also taught numerous courses for the Master in Statistical Data Analysis, the Institute for Continuing Education in Science, and FLAMES,.

PRICES

Industry	€ 1470	
Non profit, government, higher education staff	€ 1105	
(Doctoral) students, unemployed	€ 600	

LANGUAGE

This course will be taught in English

REGISTRATION

<u>Beta-Academy</u>



<u></u> 30

Structural equation modeling with lavaan

Structural equation modeling (SEM) is a general statistical modeling technique to study the relationships among observed and latent variables. It spans a wide range of multivariate methods including path analysis, mediation analysis, confirmatory factor analysis, growth curve modeling, and many more. Many applications of SEM can be found in the social, economic, behavioral and health sciences, but the technology is increasingly used in disciplines like biology, neuroscience and operation research. SEM is often used to test theories or hypotheses that can be represented by a path diagram. In a path diagram, observed variables are depicted by boxes, while latent variables (hypothetical constructs measured by multiple indicators) are depicted by circles. Hypothesized (possibly causal) effects among these variables are represented by single-headed arrows. If you had ever found yourself drawing a path diagram in order to get a better overview of the complex interrelations among some key variables in your data, this course is for you.

The first day of the course provides an introduction to the theory and application of structural equation modeling, and illustrates how to use the open-source R package 'lavaan' (see https://lavaan.org) to conduct an SEM analysis. On the second day, we discuss several special topics that are often needed by applied users (handling missing data, nonnormal data, categorical data, longitudinal data, multilevel data, etc.). With the exception of a short practical session at the end of the first day, the two days are mostly lectures, to maximize the amount of information that we can teach. However, do-it-yourself practicals (with written feedback and solutions) will be made available and illustrate all the topics that are covered in this course.

This course is for anyone interested in testing theories or models involving observed and latent variables. It caters to both beginners with no prior SEM experience and experienced users seeking to refresh or update their understanding of structural equation modeling.

COURSE PREREQUISITES

Solid understanding of regression analysis and basic statistics (hypothesis testing, p-values, etc.). Some knowledge of (exploratory) factor analysis (or PCA) is recommended, but not required. Because lavaan is an R package, some experience with R (reading in a dataset, fitting a regression model) is recommended, but not required.

₩HEN?

• 20th & 21st of May 2024

from 9am to 5pm

Where?

Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, Ghent, Auditorium 3

LANGUAGE

This course will be taught in English

TEACHER







the Department of Data Analysis, Faculty of Psychology and Educational Sciences, Ghent University. His main research interest is structural equation modeling. He is the author of lavaan, an R package for structural equation modeling.

Jasper Bogaert is a PhD researcher and teaching assistant at Ghent University at the Department of Data Analysis at the Faculty of Psychological and Education Sciences. His current research focuses on Structural Equation Modeling and alternative estimation approaches for estimating parameters from the structural part of the model, frequently using the lavaan package for simulation studies and teaching purposes.

PRICES

Industry	€ 750
Non profit, government, higher education staff	€ 565
(Doctoral) students, unemployed	€ 340

REGISTRATION

Beta-Academy



<u></u> 31