Survey Statistics and Data Analytics MSc

List of contents

Economic research

- Econometrics
- Public Policy Analysis
- Public Policy Impact Assessment

Digital data analytics

- Social Media Analytics
- Natural Language Processing
- Understanding Digital Societies

Social research

- Social Network Analysis
- Applied Social Research

Business research

- Data Visualization
- Business Analytics

Final courses

- Internship
- Thesis Consultation

Free elective courses

- Fundamentals of Research Methodology

Mathematical Foundations

Fundamentals of Mathematics I

Aim

The purpose of the course is to provide an introduction to linear algebra and to show its connection with the most important classical mathematical statistical concepts and methods through application examples. Since the subject is included in the same semester as the Analysis and Probability Theory courses, it is partly based on and supplemented by them, discussing the knowledge in a less formal, more intuitive way. By completing the course, students will be able to understand the linear algebra involved in the most important statistical models, thereby gaining a deeper understanding and more confident use of the models in practice.

Content

This course uncovers the most important connections between basic linear algebra and analysis, and the fundamental notions and methods of mathematical statistics. Lab exercises are to be done in R.
Linear algebra, vectors, vector spaces, operations with vectors.
Matrices, operations with matrices. Linear transformations. Eigenvalue, eigenvector problem.
Matrix factorization, statistical applications.
Discrete Markov chain models: Markov properties, state space / phase space and its properties, population dynamics models.

Analysis Lecture

Aim

The aim of the course is to introduce students to the fundamentals of calculus that are necessary for their later studies. By completing the course, the student gets to know a significant part of the basic mathematical knowledge needed to master the theory of statistics used in market and social research. The student will be able to independently process the scientific literature discussing the theory of statistical models.

Content

Necessary concepts of mathematical real analysis which will be needed for later studies such as statistics and optimization. The main topics covered during the semester: Series, functions, continuity and limit. Derivative and its applications. Power series. Primitive function, Riemann integral. Multivariable functions, partial derivative, critical point, critical points subject to equation constraints. Integration in two dimensions.

Analysis Practice

Aim

The aim of the course is to illustrate concepts and theorems discussed in the lectures and to illustrate the role of these concepts in statistics and data analysis.

Content

Necessary concepts of mathematical real analysis which will be needed for later studies such as statistics and optimization. The main topics covered during the semester mirror that of the lecture: Series, functions, continuity and limit. Derivative and its applications. Power series. Primitive function, Riemann integral. Multivariable functions, partial derivative, critical point, critical points subject to equation constraints. Integration in two dimensions.

Probability Theory Lecture

Aim

Mastering the basic concepts of probability calculations, which play a key role in statistics, laying the foundation for the subject of later statistics, among others through the concepts of notable distributions, density function, expected value, and standard deviation. By completing the course, the student will be able to further develop her/his statistical knowledge and understand the essential elements of simpler probability models.

Content

The course introduces the basic concepts of probability and lay sthe foundations of statistics.

Axioms of probability, event space, events, relationship with relative frequency. Classical probability field and its applications: sampling with and without replacement.
Binomial-multinomial theorem, operations with events, sieve formula.
Conditional probability, Bayes theorem. Independence of events. Discrete random variables, expected value and standard deviation.
Notable discrete distributions.
Distribution function, known continuous distributions.
Expected value, standard deviation.
Joint distribution (joint distribution and functions of random variables, conditional distribution, conditional expected value (in the case of absolute continuous).
Covariance and correlation. Conditional expected value and forecast.
Markov, Chebyshev inequality, weak laws of large numbers.
Approximation of Poisson distribution with binomial, central limit distribution theorem.

Probability Theory Practice

Aim

The purpose of the course is to (1) illustrate and make more understandable the concepts and theorems discussed in the lecture through the solution of practical tasks, (2) to convey basic statistical analysis methods used in data analysis practice. By completing the course, the student will be able to further develop her/his statistical knowledge and understand the essential elements of simpler probability models.

Content

The aim of the course is to illustrate concepts and theorems discussed in the lectures and to discuss basic methods which are used in everyday data analysis practice.

Topics:

Axioms of probability, event space, events, relationship with relative frequency. Classical probability field and its applications: sampling with and without replacement.
Binomial-multinomial theorem, operations with events, sieve formula.
Conditional probability, Bayes theorem. Independence of events. Discrete random variables, expected value and standard deviation.
Notable discrete distributions.
Distribution function, known continuous distributions.
Expected value, standard deviation.
Joint distribution (joint distribution and functions of random variables, conditional distribution, conditional expected value (in the case of absolute continuous).
Covariance and correlation. Conditional expected value and forecast.
Markov, Chebyshev inequality, weak laws of large numbers.
Approximation of Poisson distribution with binomial, central limit distribution theorem.

Fundamentals of Mathematics II

Aim

One of the goals of the course is to deepen the knowledge of mathematical statistics by examining the procedures using software. The topics are closely related to the lecture material of Mathematical Statistics. We look at the simulations primarily in R, and the application examples of the methods in SPSS (possibly Stata). Another goal of the course is to discuss theoretical and methodological knowledge that was not covered in other courses and/or is necessary for a deeper understanding of several other courses. By completing the course, students will be able to understand the mathematics involved in the most important statistical models, thereby gaining a deeper understanding and more confident use of the models in practice.

Content

The course helps the students gain a deeper understanding of abstract notions of mathematical statistics by simulation methods. Numerical optimization for parameter estimation is also covered. Problems are solved in R.

- Introduction to statistical simulation (R, SPSS)
    - Recap of probability theory
    - Generating random samples
        - Inverses transform sampling
    - Solving problems with simulation
- Point and interval estimation
    - Maximum likelihood estimation
    - Parzen-Rosenblatt method
    - Studying the properties of estimates with simulation

- Numerical methods
    - Basics of numerical methods, theory of algorithms
        - Connected topics in analysis and linear algebra
    - Optimization
        - Gradient method
        - Newton-Raphson method
        - Gauss-Newton method
    - Estimating parameters of simple statistical models
    - Expectation-maximization algorithm
        - Fitting a mixture model with EM
- Introduction to Stata
- Hypothesis testing
    - Receiver Operating Characteristics
    - Nonparamteric tests
        - Kolmogorov-Smirnov test
        - Wilcoxon-test
        - Mann-Whitney u-test
        - Kruskal-Wallis-test

Mathematical Statistics Lecture

Aim

To learn the most commonly used methods of mathematical statistics, building on what has been learned in probability, and to master the mathematical foundations of previously learned estimation and hypothesis testing methods. By completing the course, students are able to understand the most important estimation and hypothesis testing procedures, select the appropriate one for the problem, and implement them independently in practice.

Content

The course provides the most commonly used methods of mathematical statistics, and the mathematical foundations of estimation and hypothesis testing methods.

Descriptive statistics, calculation of indices. Types of data. Means, basic statistics, boxplot.
Histogram, empirical distribution function, Glivenko-Cantelli theorem (basic theorem of statistics).
Ordered sample, estimation of density function, Parzen-Rosenblatt window method, kernel functions.
Estimation theory. Properties of estimates (unbiasedness, consistency). Comparison of estimates (efficiency).
Estimation methods: maximum likelihood estimation and asymptotic properties.
Method of moments.
Bayesian estimators.
Sufficient statistics, Rao-Blackwell theorem. Fisher information, Cramer-Rao inequality.
Confidence intervals.
Hypothesis testing, properties of tests, Neyman-Pearson lemma. Parametric tests, one- and two-sample z-, t-, F-tests. Welch test.
Non-parametric tests (chi-square tests: goodness of fit, estimated goodness of fit, independence test; test for positive correlation).
Kolmogorov-Smirnov test, sign test, Wilcoxon test. Test of normality.
Linear model, hypothesis testing with t-test and F-test.
Analysis of variance.
Time series analysis, autocorrelation coefficients and their estimation. Stationary processes.
Linear processes, estimation of coefficients.

Mathematical Statistics Practice

Aim

To learn the most commonly used methods of mathematical statistics, building on what has been learned in probability, and to master the mathematical foundations of previously learned estimation and hypothesis testing methods. The practice closely follows the lecture series and helps to deepen understanding by solving practical problems. By completing the course, students are able to understand the most important estimation and hypothesis testing procedures, select the appropriate one for the problem, and implement them independently in practice.

Content

The course provides the most commonly used methods of mathematical statistics, and the mathematical foundations of estimation and hypothesis testing methods.

Descriptive statistics, calculation of indices. Types of data. Means, basic statistics, boxplot.
Histogram, empirical distribution function, Glivenko-Cantelli theorem (basic theorem of statistics).
Ordered sample, estimation of density function, Parzen-Rosenblatt window method, kernel functions.
Estimation theory. Properties of estimates (unbiasedness, consistency). Comparison of estimates (efficiency).
Estimation methods: maximum likelihood estimation and asymptotic properties.
Method of moments.
Bayesian estimators.
Sufficient statistics, Rao-Blackwell theorem. Fisher information, Cramer-Rao inequality.
Confidence intervals.
Hypothesis testing, properties of tests, Neyman-Pearson lemma. Parametric tests, one- and two-sample z-, t-, F-tests. Welch test.
Non-parametric tests (chi-square tests: goodness of fit, estimated goodness of fit, independence test; test for positive correlation).
Kolmogorov-Smirnov test, sign test, Wilcoxon test. Test of normality.
Linear model, hypothesis testing with t-test and F-test.
Analysis of variance.
Time series analysis, autocorrelation coefficients and their estimation. Stationary processes.
Linear processes, estimation of coefficients.

Data collection and processing

Data Collection Methods and Sampling

Aim

The course focuses on the methodological issues of survey data collection and presents the related traditional and current theories with a special focus on sampling procedures. Through the course, students will be able to evaluate a survey data collection and assess the reliability and bias of the resulting estimate. The course provides a detailed description of the Total Survey Error framework and its potential for generalization to digital found data. The course also addresses current issues such as the future challenges of survey data collection in a changing social and technological environment and reflects on the impact of big data on surveys.

Content

Types of error, evaluation criteria — reliability, validity, bias and variance, sampling and non-sampling error, types of error according to Total Survey Error (TSE).
Data collection methods, response processes and questionnaire design.
Context effects, question phrasing, interviewers, evaluation/pretesting/post-stratification of questionnaires (and standard error).
Panels.

The effect of exit polls and opinion polls on public opinion (concepts; facts).
Comparability of opinion poll results.
Survey Data Processing

Survey Data Processing

Aim

The purpose of the course is to provide insight into the basics of data preparation and data processing, as well as data analysis, by going around the steps of data processing of survey-type data. The topics cover the stages from organizing the data database to data analysis to the final preparation stage. By completing the course, students learn basic database operations, are able to convert data into tidy data format, whether it is survey data or found data, are familiar with data missing mechanisms, are able to produce weight variables based on given aspects, carefully plan weighted analyses, know and know technique the most important imputation methods, abilities to design and implement simulations. Students learn the main principles of reproducible research management and apply them during all work processes of the course.

Content

Class 1: Introduction. Data types and survey data collection. (theory)
Class 2: Structured database from raw data I. (exercise)
Class 3: Structured database from raw data II. (exercise)
Class 4: Completing and expanding data (exercise), p-hacking
Class 5: Imputation and handling of missing values (theory and practice)
Class 6: Imputation and anonymization, synthetic data (theory and practice)
Class 7: Determining the standard deviation and standard error of estimates (theory & practice)
Class 8: Post-stratification I. (theory)
Class 9: Post-stratification II. (exercise)
Class 10: Preparation for the analysis & corrections for nonresponse and measurement error (theory & practice)
Class 11: Comparison of Big Data and classic data sources (theory)
Class 12: Simulation (theory & practice)

Foundations of business research

Market Research I

Aim

The aim of the Market Research courses is for the students to be informed about the market research framework after completion. The planning, preparation and analysis of independent research should be developed as a skill along the lines of practice-oriented tasks.

Content

The lectures touch on the most important customer types, methodologies, data collection techniques and customer thinking necessary for research planning.

During the first semester, the most important goal is to get to know and practice basic business thinking, company operation, problem formulation, research and planning. This is complemented by getting to know the legal framework of data collection, as well as practicing social statistics basics and bringing them up to skill level.

Topics: Social statistics from what, from whom, how many, target group calculation; Legislation, data protection; Company operation, information needs; Business and marketing metrics; Translation and formulation of business and research problems; Creating a research plan.

Market Research II

Aim

The purpose of the course is to familiarize the students with the key elements of Innovation methodologies, the purposes and methods of use of the analysis tools that arise during development through practical tasks. By the end of the course, students will be able to create an analysis/evaluation framework that can be used in any environment.

Content

The course focuses on the analytical tasks of the innovation process. A large proportion of analyst tasks are related to some kind of development, the development of a new product, product feature or some kind of change. This field has a separate set of rules, procedures, and vocabulary, which is important to know, because in most cases you have to work in a team, experts in other fields have to be understood, and supported with data and evaluations. If you don't yet know all the three-letter abbreviations or you don't know about each of them exactly, e.g. what BMC, VP, MVP, Waterfall, OKR, GTM, Lean, CP, Seed, UI/UX/SD, Agilis, SCRUM or even Idea- life cycle are then this is your place. Don't worry, these are just the entrance tickets to a World where they talk like this, the real value creation starts after this - if you know these buzz-words, you can assert yourself much more easily later.

Communication and Project Management

Aim

Development of students' verbal communicational skills, teaching about tools and problems of presentation and data visualization, making a Prezi presentation independently, data visualization in Python, getting to know git, displaying visualization on a webpage.

Knowledge, skills, abilities and attitude which can be attained during the course
- knowledge: about verbal communication, negotiation techniques, design of presentation (general and Prezi), programming data visualization
- skills: of efficient corporate and client communication, conflict management, efficient presentation of data and data visualization
- ability to give account of one's data analytical and research activities and their results on different forums — scientific community, customers, anyone from society in accordance with the knowledge

Content

Development of students' verbal communicational skills, tools and problems of presentation and the basics of project management.

l. Introduction: communication, information and statistics
2. Basics of verbal communication: active attention — communicational practices
3. Managing conflicts - communicational practices (in-person occasion — compulsory)
4. Techniques of negotiation - communicational practices
5. Visual thinking hand-drawing practices 6. Designing presentations and basics of presentation — correcting errors
7. Designing simple Prezi — step by step practice
8. Ideas for efficient Prezi — developing and handing in one's own Prezi
9. Basics of data visualization
10. Data visualization in Python: simple diagrams in the Altair package
11. Data visualization in Python: complex diagrams in the Altair package
12. Consultation about hand-ins
13. Show us what you made — independent presentation

Project Seminar

Aim

The aim is to enable students to critically assess the objectivity of empirical research. To develop an open, receptive, yet critical attitude towards all professional innovations, so that students support and apply those that meet the requirements of reliability and validity in their professional judgement.

The student will be able to translate a research question from an academic, governmental or business client into the language of statistics; to design a complete study to answer the question, to develop the data collection method and the sampling method, to interpret and communicate the results to the client.

Content

Project Seminar provides a systematic practical knowledge of planning, implementing and managing business projects, from start to finish by developing business plans. Students will prepare and develop practically applicable business plans, including market analysis and competition, identification of resources needed, etc.

Programming

Introduction to R

Aim

This course serves as an introduction to R programming, with its fundamentals applicable to almost any other current object-oriented programming language. The objective for students is to become familiar enough with R to utilize it as a tool for all statistical modeling and data analysis tasks in more advanced courses.

Upon completion of the course, students will be able to independently conduct statistical analyses, create simple simulations, and present results in R. They will also be equipped to deepen their knowledge using package documentation and online resources.

Content

Introduction to programming logic and R basics. Since this course runs concurrently with the Probability course and Linear Algebra and Analysis, it builds upon much of the material covered in those courses. However, many of those topics are discussed in a less formal and more intuitive manner.

Programming basics
Introduction
Operators, variable types, coding conventions
Cycles, applications, packages I
Cycles, functions, packages II
Complex data structures and their use
File operations, data sources
Depiction basics
Tidy-verse
Statistics in R
Descriptive statistics
Simulation of random processes
Data management, data cleaning
Hypothesis tests
Regression

Introduction to Python

Aim

The goal of the course is to introduce students to the basics of Python programming, techniques for collecting textual data with Python, and the preparation of data for analysis. Students who complete the course will be able to understand and modify program codes written by others, as well as be able to write their own code and select and apply appropriate Python packages in the field of data analysis and data visualization.

Content

This course gives an introduction to Python programming using Anaconda, which contains Python, Jupyter Notebook and the most widespread packages and add-ons. After reviewing the general syntax, programming and data structures, the course focuses on data science applications, like database building using web-scraping and writing SQL queries, statistical modelling, and collaborative working with Git.

1. Introduction, basic operations
2. Conditions, try-except
3. Operations with strings, lists
4. Functions
5. For and while loops
(6. Consultation)
7. Dictionaries, tuples
8. Working with external files
9. Data collection with the BeautifulSoup package
10. Data collection with the BeautifulSoup package
11. Dataframes, pandas package
(12. Consultation)

Data Analysis Infrastructure (SQL, Git, other tools)

Aim

This course is designed to establish a robust data analysis infrastructure foundation, incorporating essential tools such as SQL, Git, and other indispensable resources. Participants will develop proficiency in SQL for data extraction and manipulation, and learn to utilize Git, a powerful version control system, for efficient collaborative data analysis project management. Throughout the course, students will also familiarize themselves with a range of additional tools and resources crucial for data analysis workflows. The goal is to empower students to confidently navigate data analysis tasks in both professional and research settings.

Structured for practicality, the course combines theoretical knowledge with hands-on application. Participants will gain proficiency in leveraging this infrastructure for streamlined and collaborative data analysis, equipping them well for the demands of data-driven industries.

Content

- Introduction to Data Analysis Infrastructure
- Git and Version Control
- SQL Fundamentals
- Practical Applications
- Optional topics: command line, LaTeX, docker, cloud computing platforms

Data analysis

Multivariate Statistics Lecture

Aim

This course covers the theory of some of the most prominent classical multivariate statistical methods. Students will familiarize themselves with the key concepts of multidimensional variables and the theory

behind linear models, and gain understanding of several other supervised and unsupervised, exploratory techniques.

Upon completion of the course, students will be able to critically assess the correct use of covered methods in scientific research and in the industry, as well as being able to propose a viable method for a specific problem based on a deep theoretical understanding of these.

Content

- multidimensional distributions,
- singular decomposition and eigen decomposition of matrices
- the multidimensional normal distribution (MVN),
- the Wishart-distribution,
- the Cochran-Fisher theorem,
- maximum likelihood estimation for the parameters of MVN,
- multidimensional z- and t-tests,
- the multidimensional linear model,
- multidimensional regression analysis and MANOVA,
- principal component analysis,
- factor analysis,
- discriminant analysis,
- multidimensional scaling.

Multivariate Statistics Practice

Aim

The course focuses on the practical application of the major classic multivariate statistical methods, building on the content of the related lecture. Students can learn about the practice of the methods discussed during the lectures, their practical advantages and disadvantages.

After completing the course, students are able to propose solutions to scientific, industrial, and market data analysis problems in an educated way, as well as to carry out the analysis and critically evaluate the results.

Content

The course follows closely the theme of the lecture with computer implementation of the statistical models discussed.

- multidimensional distributions,
- singular decomposition and eigen decomposition of matrices
- the multidimensional normal distribution (MVN),
- the Wishart-distribution,
- the Cochran-Fisher theorem,
- maximum likelihood estimation for the parameters of MVN,
- multidimensional z- and t-tests,
- the multidimensional linear model,
- multidimensional regression analysis and MANOVA,
- principal component analysis,
- factor analysis,
- discriminant analysis,
- multidimensional scaling.

Data Analysis I

Aim

Linear and logistic regression models, main types of generalized regression models, multilevel models and their practical applications. Fundamental problems in all areas of data analytics such as the assessment of model fit (nested and non-nested models), the inclusion of interactions/categorical predictors, merging and controlling, model building (for causal and predictive models), overfitting, cross-validation, bootstrap procedures in model building.

Students who have completed the course will be able to apply their statistical knowledge in practice, learn new methods of analysis and develop new tools and methods according to their practical needs. The ability to choose the appropriate statistical analysis technique for data analysis and to implement it using appropriate IT tools.

Content

linear regression — categorical predictors,
logistic regression,
general modeling problems,
the former in a unified framework: Generalized Linear Models,
Poisson regression,
multi-level models,
general concepts,
confounding,
control,
interaction,
model, prediction and causal models,
model building,
stepwise methods,
nested models,
overfitting,
validation, k-fold cross-validation,
bootstrap e.g. for variable selection.

Data Analysis II

Aim

The course covers the basic statistical methods of categorical data analysis and time series analysis through practical examples. Software implementation of the methods is also discussed.

The aim is to enable students to critically assess the objectivity of empirical research. To develop an open, receptive, yet critical attitude towards all professional innovations, so that students support and apply those that meet the requirements of reliability and validity in their professional judgement.

Content

Generalizations of independence;
Generalizations of odds ratios, conditional odds ratios;
Loglinear representation, loglinear model;
Basics of time series analysis;
Stationary processes;
Seasonality;
ARMA model.

Data Science Lecture

Aim

Data Science is high level course discussing the most important concepts in the theory of statistical learning and models used in machine learning and artificial intelligence. Statistical models developed in different fields are treated in a general theoretical framework to highlight the fundamental underlying problems of function approximation, pattern recognition, and generalization. The lecture series is accompanied by a computer lab practice to provide application examples for the different models.

Students who have completed the course will be familiar with the most important models in current data science and also understand the mathematical fundamentals which will allow them to adapt these models to specific problems and have the ability to thoroughly understand new methods and models developed in the future.

Content

- Introduction to statistical learning theory
    - Function approximation and ill-defined problems
    - Empirical and structural risk minimization principle
    - Probabilistic bounds on estimates
    - Model complexity, capacity control, regularization
    - Constrained optimization
- Kernel methods
    - Hilbert spaces
    - Kernel functions
    - Reproducing Kernel Hilbert space
    - Regularization in Hilbert space
- Shrinkage methods
    - Ridge regression
    - LASSO, Least Angle Regression
- Spline models
    - Regression splines
    - Smoothing splines
    - Splines with regularization
- Tree based methods
    - Classification and regression trees
- Ensemble methods
    - Bagging
    - Random forest
    - Boosting, AdaBoost
- Support Vector Machines
    - Maximal margin and support vector classifier
    - Support vector machines
- Artificial neural networks
    - Rosenblatt's perceptron and precursor models
    - Back-propagation algorithm
    - Network architectures
    - Overview of current advancements (Deep Learning, Convolutional neural networks, Generative adversarial networks)

Data Science Practice

Aim

The aim of the course is to gain a deeper understanding of the models and algorithms discussed at the theoretical level in the data science lecture through application examples. The analyses are usually performed using the appropriate R and Python packages, but in some cases, we create our own implementation.

By completing the practice course, students will be able to implement the basic procedures learned and interpret the results.

Content

- Introduction to statistical learning theory
    - Function approximation and ill-defined problems
    - Empirical and structural risk minimization principle
    - Probabilistic bounds on estimates
    - Model complexity, capacity control, regularization
    - Constrained optimization
- Kernel methods
    - Kernel functions
    - Reproducing Kernel Hilbert space
    - Regularization in Hilbert space
- Shrinkage methods
    - Ridge regression
    - LASSO, Least Angle Regression
- Spline models
    - Regression splines
    - Smoothing splines
    - Splines with regularization
- Tree based methods
    - Classification and regression trees
- Ensemble methods
    - Bagging
    - Random forest
    - Boosting, AdaBoost
- Support Vector Machines
    - Maximal margin and support vector classifier
    - Support vector machines
- Artificial neural networks
    - Rosenblatt's perceptron and precursor models
    - Back-propagation algorithm
    - Network architectures
    - Overview of current advancements (Deep Learning, Convolutional neural networks, Generative adversarial networks)

Applied methods

Qualitative Methods

Aim

The purpose of the course is to provide an introduction to the qualitative research methods that are most often used in the field of market and public opinion research. During the semester, we review the interview method, focus groups, observations, netnography, and online/social listening methods and we

also talk about the applicability of visual and projective techniques, as well as the possibilities of combining different methods. The underlying motivation for using the tools presented during the semester is to learn as much as possible about the deeper reasons behind the actions (e.g. consumption, decisions) of the participants under investigation. At the same time, the course aims to convey theoretical knowledge about the individual techniques and, taking advantage of the opportunities provided by the small group, involves the students in trying out the individual methods, so that the students gain first-hand experience of the advantages and disadvantages of the techniques, the gains, pitfalls, and risks arising from their application. In addition to knowledge about methods, students also receive a general introduction to the various preparations necessary for the application of qualitative techniques. During the semester, if possible, employees of several market research companies will give a guest lecture, and we will also have a field visit. The aim of the semester is for the students to implement a mini research project, in which at least one technique learned during the semester is used as a test.

Content

1. Introduction, the epistemology of qualitative research methods
2. Individual and group interview
3. Focus group
4. Projective techniques
5. Netnography
6. Social listening
7. Visual techniques, photovoice
8. Ethical issues

Network Analysis

Aim

The course covers the basic methodological concepts of network research and the underlying mathematical notions of graph theory and introduces students to the practice of network research using R and Python. Students will be able to critically evaluate research using network methodology, both for small-scale networks where the focus is on individual nodes, and for large-scale networks, where graph structure and limiting properties are more important.

Knowledge: During the course, students will develop their knowledge by reviewing most prominent theories of network analysis literature, understanding the basic theoretical concepts, learning how to use methodological tools, and mastering them through practical examples.

Skill: The skills acquired during the course will focus on the interpretation and solution of a problem using network analysis tools; including the search for relevant data related to the problem, data collection, data processing, selection of applicable methods, analysis using related network analysis tools, data visualization, interpretation of results and drawing conclusions.

Attitude: During the course, students will learn to correctly assess the scientific problem to be analyzed, to select appropriate analytical tools, and to apply the available methods carefully and critically, taking into account and reflecting on possible methodological aspects and difficulties. In documenting and interpreting analytical results, they will also learn to follow best practices and quality assurance guidelines of the relevant national and international scientific community, as well as the principles of open science.

Autonomy and responsibility: The course will equip students with the knowledge and skills necessary to carry out independent empirical research using the methodological tools of network analysis, and with the scientific attitude necessary to discern the results of research in a responsible manner.

Content

- Graphs: vertices and edges, paths and circles, undirected, directed and weighted graphs.
- Representations of graphs.
- Local and global graph statistics.

- Graphs in statistics: tree graphs, directed acyclic graphs (DAG).
- Basic network models, generative models, exponential random graph model (ERGM).
- Propagation processes, message passing.
- Neural Networks, Graph Neural Networks (GCN, GAT).

Official Statistics and Data in Public Administration

Aim

The aim of the course is to present major European statistical data collections and their historical motives, as well as to review the main stages of statistical thinking. During the course, the domestic and European Union statistical system, the relevant legal regulations, and the main international official statistical data sources and data repositories will be presented. The course is prepared in collaboration with experts from ELTE and CSO (the national statistical office of Hungary), as well as the Hungarian Office of Education.

The student who completes the course will have a high level of commitment to the quality assurance guidelines set by national and international professional organisations. They are open and committed to all forms of cooperation in international professional relations. He/she will be committed to the widest possible publicity of research data and analyses, especially when the public interest so requires.

Content

Introductory consultation: purpose of the lesson, circumstances, evaluation, presentation of the EMOS specialization.
Excerpts from the history of probability and statistics. Required reading: Fienberg, S. E. (1992). A brief history of statistics in three and one-half chapters: A review essay.
Survey research antecedents, domestic and international history.
Presentation of data sources, indicator systems, Eurostat. Mandatory literature: Towards a harmonized methodology for statistical indicators. Part I: Indicator typologies and terminologies. Eurostat 2014 edition.
The statistical system, the data generation model.
Professional statistics.
Modern data acquisition technologies. Experimental statistics, Technologies supporting data production.
Data management systems.
Disclosure of data, access practices
Statistical work of the Office of Education: DPR, matriculation exams, competence measurements, international examinations. Thesis, professional internship, and publication opportunity.
Preliminary exam: in the last week of the semester.

Social Science Research

Aim

The aim of the course is to familiarize the students of the survey statistics program with all the phenomena and mechanisms that can help in the systematic study of social relations, the sociological description of today's Hungarian society, and provide examples for the application of the methods learned in the program.

Students who have completed the course will be familiar with the basic context of the social sciences in relation to the structure of contemporary Hungarian society, the demographic situation, social inequalities and theories describing social change.

Content

During the semester, the course covers some aspects of the following topics: the most important population and demographic processes, the history of Hungarian structural research from the 1960s to the present day, the structure of today's Hungarian society, the most important characteristics and dimensions of the organization of social inequalities, the issue of social mobility, the material, cultural and the characteristics of the functioning of relational capital, the issue of lifestyle and status groups, the phenomenon of social detachment, and issues such as losers and winners after the regime change, the poor and impoverishment, unemployment, the consequences of marginalization in relation to deviance, and the ethnicization of poverty, the issues of cultural stratification and social identity.

The discussion of individual topics is embedded in an overview of the most important quantitative empirical research of the last two decades. This also offers an opportunity to review the most important research methodology issues, the dilemmas of operationalization and quantitative analysis, and the technique of writing studies in connection with a wide variety of topics.

Biomedical research

Biostatistics

Aim

The aim of the course is to prepare students for jobs in medicine or pharmaceuticals, and for students planning to enter other fields, to fertilize and broaden their perspectives with typical perspectives and solutions of biomedical research, and to enable them to find innovative solutions in the adaptation of methods for social research, business, economics.

Content

- basic epidemiological concepts, indicators, designs
- diagnostic tests, diagnostic function validation
- survival analysis, Kaplan-Meier analysis
- Cox regression
- comparison of heterogeneous populations: direct and indirect standardization
- Poisson regression
- simple nonparametric methods

Meta-analysis

Aim

The aim of the course is to present the process of meta-analysis to synthesize empirical information, from hypothesis generation to presentation of results, by familiarizing the students with the most common software used in the field. By completing the cruise, students will be able to evaluate research conducted using the meta-analysis, as well as prepare their own meta-analysis, competently selecting the appropriate methodology.

Content

Introduction — why we do meta-analyses, PICO research question framework.
Systematic literature research and available databases.
Qualitative evidence synthesis, selection and quality assessment protocols.
Quantitative evidence synthesis — meta-analysis.
Appropriate metrics, simple meta-analysis methods.
MetaXL — fixed and random effects meta-analysis.
MetaEssentials — subgroup analysis and moderator analysis.
Meta-regression in R.

Network meta-analysis.

Economic research

Econometrics

Aim

The aim of the course is to provide a comprehensive overview of basic econometric modelling methods, including appropriate question posing, common errors to avoid, main model frameworks, and checking application conditions. Cross-sectional methods, linear regression and its extensions, the resolution of linearity requirements, models with limited outcome variables, time series and panel econometrics.

Content

Week 1, Introduction. Cross-sectional models I. Multivariate linear regression. Parameter estimation, global and partial tests, model diagnostics, model selection.
Week 2, Cross-sectional models 2. Dummy variable, interaction, quadratic effect.
Week 3, Cross-sectional models 3. Multicollinearity, heteroskedasticity, autocorrelation.
Week 4, Cross-sectional models 4. Non-linear models.
Week 5, Cross-sectional models 5. Difference-in-differences, instrumental variables.
Week 6, Cross-sectional models 6. Models with limited outcome variables. Binary and multinomial logit, linear probability model, Tobit, Probit model.
Week 8, Time series analysis 1. Deterministic time series models. Trend, cycles, seasonality, residual autocorrelation.
Week 9, Time series analysis 2. Stochastic time series models. Autocorrelation, partial autocorrelation, correlogram, white noise, autoregressive models.
Week 11, Time series analysis 3. Stochastic time series models. Stationarity, unit root, integrated autoregressive models, random walk.
Week 12, Panel econometrics I. Random Effects, Fixed Effects, First Differentiating.
Week 13, Panel econometrics 2. Instrumental variables, non-linear models, difference-in-differences.

Public Policy Analysis

Aim

During the course, students become familiar with the public policy process as a whole: the methodological issues that arise, the concept and components of the public policy cycle, thus understanding the place and role of program evaluation in the broader context. The class requires a lot of independent (individual and group) work and ends with a test at the end of the semester.

Learning goals: The goal of the unit is for the participants to know and evaluate the tasks, professions, and challenges that usually arise during public policy processes, including typical pitfalls and mistakes. As an evaluator, they should be able to communicate productively with their clients, politicians, and civil servants, and they should also know a broader scope of project evaluation than what is to be learned in the second half of the course, and the aspects on the basis of which it is possible to decide what the adequate evaluation method is. In addition, learn about the tasks that you can get in other jobs related to public policy, for example as a junior decision maker.

Content

Introduction + what is public policy? Community economics overview
Public policy in general: institutions, actors, instruments; the context of public policy
The public policy cycle: introduction, examples
The public policy cycle: Agenda and creation of public policy I
The public policy cycle: creation of public policy II
The public policy cycle: decision-making
The public policy cycle: implementation

The public policy cycle: evaluation
Alternative models of public policy-making

Public Policy Impact Assessment

Aim

The aim of the course is to introduce students to the methodology of public policy impact assessment. During the course, students will learn about the types, conditions, and possibilities of use of impact assessments, as well as the statistical methods used. In addition, during the course we will discuss the difference between causality and apparent correlation and the possibilities and limitations of exploring causal relationships. The aim of the course is for students to be able to critically interpret the results of an impact assessment, and to be able to plan and conduct an impact assessment independently, including the choice and implementation of appropriate statistical methods. Given that only a few databases are available that can be used to model real impact assessments, during the course we use other databases to illustrate and apply the discussed methods in practice.

Content

1. Introduction: Overview of the topic, purpose of impact assessments, review of applied methods, and necessary statistical foundations.

2. General aspects of impact assessments: Objectives of impact assessments, types of impact assessments, the Hungarian impact assessment system. Literature: Impact Assessment Manual Volume II, Chapter 1.

3. Planning impact assessments: Steps of impact assessment, determination of criteria, identification of intervention effects, examination of impact mechanisms, possibilities for measuring impacts, feasibility and risk analysis. Literature: Impact Assessment Manual Volume II, Chapter 2.

4. Potential data sources: Data required for preliminary and subsequent impact assessments, possible tools and practical aspects of data collection. Literature: Impact Assessment Manual Volume I, Chapter 4, and Impact Assessment Manual Volume II, Chapter 5.

5. Basic concepts of impact assessment: Correlation and causality, counterfactual state, target group, control group, intended and unintended effects. Literature: Impact Assessment Manual Volume I, Chapter I.

6. Preliminary impact assessments: Examples of preliminary impact assessments, forecasting with time-series analysis, forecasting with microsimulation. Literature: Impact Assessment Manual Volume I, Chapter 2.

7. Subsequent impact assessments I: Experiment as a tool for impact assessment, experimental design, conditions and limitations of method application, implementation. Example of experiment-based impact assessment. Literature: Impact Assessment Manual Volume I, Chapter 3.3; MHE Chapter 2.

8. Subsequent impact assessments II: Difference-in-differences method, conditions and limitations of method application, implementation. Application of control variables. Example of difference-in-differences method application. Literature: Impact Assessment Manual Volume I, Chapters 3.2 and 3.4; MHE Chapter 5.

9. Subsequent impact assessments III: Introduction of matching-based methods, logic of the method, conditions and limitations of application, implementation. Example of matching-based impact assessment. Literature: Impact Assessment Manual Volume I, Chapter 3.5; MHE Chapter 3.3.

10. Subsequent impact assessments IV: Additional methods: interrupted time series regression model, application of instrumental variables. Logic of the methods. Examples of impact assessments. Literature: Impact Assessment Manual Volume I, Chapter 3.7; MHE (4.) and Chapter 6.

Digital data analytics

Social Media Analytics

Aim

The aim of the course is to present comprehensively the directions of social media analysis, from targeted marketing use, through the development of data-driven business products, to the world of scientific publications. The course provides a comprehensive overview of the business analytics capabilities of the different social media platforms, the possibilities of data collection, definitions of metrics and their evaluation, and the specificities of the platforms.

The student who completes the course has the knowledge necessary for the practical implementation of business research in the field of research planning, analysis and evaluation. Able to choose a suitable statistical tool during data analysis, able to properly present and interpret the results in a business context. In terms of attitude, the student is committed to representing professional standards; is open, inclusive, but at the same time critical of all professional innovation efforts, supports and applies those that, according to its professional assessment, meet the requirements of reliability and validity.

Content

Digital Footprints in Social Sciences,
Analytic Tools,
Data Sources,
Evaluation of social media analytics metrics,
Characteristics of different social media platforms,
Domestic and international social media analytics case studies, and student case studies.

Natural Language Processing

Aim

This course introduces quantitative analysis of large text corpora using Natural Language Processing (NLP), progressing from simple word-based descriptive statistics to deep-learning language models. Theoretical concepts are predominantly introduced through practical tasks in Python, allowing students to gain hands-on experience with NLP.

Upon completion of this course, students will have a solid understanding of basic NLP concepts and algorithms. They will be equipped to comprehend NLP articles and effectively plan and execute less complex NLP projects.

Content

1. Introduction
2. Data collection
3. Preprocessing in theory and in practice I.
4. Preprocessing in theory and in practice II.
5. Supervised classification algorithms for NLP in Python
6. Unsupervised algorithms I - dimension reduction: SVD, MDS, t-SNE
7. Unsupervised algorithms II: - cluster analysis: distance measures, k-means, and hierarchical clustering, HDBSCAN
8. Topic modeling in theory and in practice.
9. Word embedding in theory and in practice I
10. Word embedding in theory and in practice II
11. N-gram language models
12. Transformers

Understanding Digital Societies

Aim

The aim of the course is for students to become familiar with the role and key methods of social media research in sociology. By completing this course, students will be able to place social media research within the context of sampling, and to formulate the essential questions necessary for evaluating research based on such data.

Content

The course introduces the sociology of digital research and social media. The course will go through the main aspects of the phenomenon (e.g. the impact of algorithms, the social impact of AI, self-representation in social media, etc.), together with the relevant theoretical approaches and current main research directions.

Social research

Social Network Analysis

Aim

The course aims to provide practical knowledge of network analysis with a social science focus using Python. The analysis of networks now pervades modern society, from online social networks to business and political organizations: an interdisciplinary field that aims to explain the complex phenomena in which relationships between actors can be defined.

Students will grasp the fundamentals of social network analysis, providing both theoretical insights and practical skills to delve into key contemporary network science concepts and theories. Through readings and projects, students will discern and comprehend the nuances in defining and utilizing these concepts, alongside exploring their theoretical and empirical applications. Additionally, students will gain proficiency in analyzing social networks using the appropriate software.

Content

The course will delve into various social network principles, including social capital, homophily, preferential attachment, propinquity, contagion, and more, demonstrating their relevance to social science theory and research, as well as for business applications.

- analyzing social networks by processing data and visualizing the network,
- understanding the mechanisms that create networks,
- dynamic networks,
- examine the analysis of social networks through case studies.

Applied Social Research

Aim

The aim of the course is to demonstrate the practical application of the modelling tools learned. Through sociological research as case studies, students are confronted with practical problems and apply their solutions to their own data and research questions.

Upon completion of the course, the student will be able to design an empirical research project in the social sciences. They are able to choose the appropriate statistical analysis technique for the research question, to implement it using appropriate IT tools and to interpret the results.

A number of guest lectures take place throughout the semester. These guest lectures are included in the timetable so that students:

- Become familiar with the wide range of research projects currently underway in Hungary;
- To learn about the variety of research designs, approaches and methodologies currently being used;
- Understand the complexity of the work involved in conducting social science research (including methodological challenges);
- get to know the wider research community.

Guest lectures will be given by professionals from the public and private sectors who have a track record of applied research in one or more areas. These lectures will vary from year to year and may cover topics such as health-related issues, ageing, ethnic minorities, poverty/socio-economic disadvantage, mental health, sociology of education, political science or social aspects of sustainability.

Content

- Contemporary sociological research issues, offline and digital space,
- data quality, opportunities and limitations,
- sociological theory and empirical evidence.

Business research

Data Visualization

Aim

The purpose of this practice course is to introduce data visualization techniques and their potential, and to learn how to form expectations related to data representation and to gain experience in goal-oriented visualization of data. During the classes the students get to know simpler and more complex data visualization techniques based on the basic knowledge acquired during their previous methodological studies. During the course, great emphasis is placed on practical orientation. In addition to the primary goal, it is also important to strengthen group work and foster a critical attitude.

Content

Introduction to Tableau, practice
Data visualization in Tableau
Getting to know the interface; Data connection, calling up databases
Basic commands in Tableau
Getting to know the database: descriptive statistics based on own data
Editing diagrams, formatting figures. More complex visualizations: the dual axis
Editing diagrams, formatting figures
Complex visualizations, visualizations on own data
Posters
Maps
Dashboard and story in Tableau
Practice
Editing a scientific poster using a template in Power Point
Static vs. interactive figures
Simple Maps in Tableau
Dashboard editing using a module grid; Interactions
Story from dashboard
Creating a dashboard and story from your own data

Business Analytics

Aim

The main objective is to understand how a business analysis company works, what to look for, what are the conditions, what is the market, what are the players and how to characterize the latest trends. We invite speakers from a wide range of fields. We will seek to introduce new areas, companies, and topics, i.e., topics that are not covered or less emphasized in other courses. By completing the course, the students will be able to see through the operation of a business analysis company, so they will be able to effectively participate in the operation of such a company.

Content

During the semester, we welcome invited speakers from the widest possible range of data science applications to present and analyze a business problem and prepare students for the situations they will encounter in their future workplace. Students will learn about business planning and running a research company.

Final courses

Internship

Aim

The aim of the internship is to give students an insight into the practical application of the acquired data analysis knowledge, to be able to apply it in new contexts and to adapt new tools and methods according to the needs of the practice, to work with others, to be able to communicate professionally.

Content

The traineeship is a 240-hour, two-month work in a company or public institution working in the field of data collection, data analytics, statistical-based business decision making or other relevant statistical activities.

Thesis Consultation

Aim

The purpose of the thesis consultation is for the consultant to assist the students in preparing their thesis within a formal framework. During the regular consultation sessions, the student reports on the progress of the thesis and can ask his consultant for guidance. This same course provides an opportunity for the initial development of the concept of the thesis and preparation for the thesis presentation.

By completing the course, the student will be able to interpret and use English language literature in the field of statistics, will be able to manage and analyze large databases and interpret the results of the analyses. Furthermore, in the thesis, the student strives to enforce the quality assurance guidelines formulated by domestic and international professional organizations.

Content

The content of the course varies depending on the specific thesis topic. It is typically carried out in the form of personal and/or online consultation.

Free elective courses

Fundamentals of Research Methodology

Aim

The course aims to enable students who come to the MSc program in Survey Statistics and Data Analytics without full credit transfer or from non-social science disciplines to master the basics of social research.

Content

The course consists of the collaborative examination of the following topics:

Topic 1: Introduction to social science research. Theory and empiricism, differences between natural and social sciences, qualitative and quantitative, exploratory and confirmatory approaches in the social sciences.

Topic 2: Causality in social science research.

Topic 3: Conceptualization, operationalization, and the quality of measurement.

Topic 4: Types of social science research.

Topic 5: Research design, topic selection, research question, and hypothesis.

Topic 6: Quantitative measurement in social science research. The Lazarsfeld paradigm. Statistics in social science research.

Topic 7: „Big Data" and social science