

Summary of Degree Programme 'Applied Statistics with Network Analysis'

Field of Studies

01.04.02 Applied Mathematics and Informatics

Approved by

The academic council meeting protocol № 03 from 26.02.2016

HSE University Educational Standard

[HSE University Educational Standard: Master's Degree \(from 2022\)\(signature\)](#) (PDF, 1.05 Mb)

Last Update

15.08.2023

Network Programme

No

Length of Studies, Mode of Studies, Credit Load

2 years [□]

Full-time, 120

Language of instruction

ENG

Instruction in English

Qualification upon graduation

Master

Double-degree Programme

No

Use of online learning

Прикладная статистика с методами сетевого анализа (2016): With online tools

Прикладная статистика с методами сетевого анализа (2023): [Online programme](#)

Tracks

2023/2024 ACADEMIC YEAR

Network Analysis

Type: General

Track Supervisor: [Batagelj, Vladimir](#)

Language of instruction: English

Use of online learning: With online tools

Qualification upon graduation: Магистр

2021/2022 ACADEMIC YEAR

Business-Analytics

Type: Applied

Track Supervisor: [Kuskova, Valentina](#)

Language of instruction: English

Use of online learning: With online tools

Qualification upon graduation: Магистр

Computational Social Sciences (PhD Track)

Type: Research

Track Supervisor: [Zaytsev, Dmitry](#)

Language of instruction: English

Use of online learning: With online tools

Qualification upon graduation: Магистр

Network Analysis

Type: General

Track Supervisor: [Batagelj, Vladimir](#)

Language of instruction: English

Use of online learning: With online tools

Qualification upon graduation: Магистр

2020/2021 ACADEMIC YEAR

Business-Analytics

Type: Applied

Track Supervisor: [Kuskova, Valentina](#)

Language of instruction: English

Use of online learning: With online tools

Qualification upon graduation: Магистр

Computational Social Sciences (PhD Track)

Type: Research

Track Supervisor: [Zaytsev, Dmitry](#)

Language of instruction: English

Use of online learning: With online tools

Qualification upon graduation: Магистр

Network Analysis

Type: General

Track Supervisor: [Batagelj, Vladimir](#)

Language of instruction: English

Use of online learning: With online tools

Qualification upon graduation: Магистр

Competitive Advantages

There is a shortage of specialists in applied statistics, especially in the area of social network analysis. At the same time, training in the field of statistics is carried out in different ways: the majority of educational programmes in this area belong to the field of economics and focus mainly on mathematical methods; in the field of sociology, the study of statistics is limited to the study of probability theory and introductory courses.

This programme is unique because it is the first programme in Russia to offer a comprehensive approach to data analysis in different areas. As part of the programme, students from different disciplines can come together to solve practical analytical problems. Those mathematically inclined gain an understanding of sociology and the object of research, while those with a background in the humanities will be able to build their skillset and gain a deeper understanding of statistical processes making up the data analysis that we teach. In addition, a special focus of the programme will be the analysis of social networks, a direction of data science that is becoming increasingly popular in foreign and Russian research practice.

Another important characteristic of the programme is its applied nature - students do not learn from abstract theoretical constructs, but rather from dealing with specific applied research questions. Students will be able to apply their knowledge by solving practical problems, working at the International Laboratory for Applied Network Analysis, Russian analytical centers and commercial companies.

Professional Activities and Competencies of Programme Graduates

The knowledge and skillset obtained by graduates of the programme will render them skilled practitioners, able to apply advanced complex data analysis techniques working in a range of organizations - both in commercial companies operating in various industries (banking, insurance, consulting, IT, medicine, pharmacy), and in

research organizations (sociology, marketing). The main competencies of the graduates of the program are:

General professional competencies:

— Is able to apply a systematic approach in setting objectives and choosing approaches to the solution, as well as to take into account conflicting goals and needs and demands.

— Is able to correctly use existing and introduce new concepts in the field of mathematics and informatics, integrate known facts, concepts, principles and theories related to applied mathematics and informatics.

— Is able to reasonably select and apply modern computer technologies to solve professional tasks, including operating systems, network technologies, programming languages, languages of data manipulation, digital libraries, application packages.

— Is able to communicate with specialists in the field of mathematical models and information technologies, as well as with experts from applied fields using various formal languages and notations.

— Is able to build mathematical models and use them in solving applied problems in accordance with the direction of training and specialization.

Professional competencies:

— Is able to organize research activities.

— Is able to create computer programs using models and algorithms of applied mathematics

— Is able to assess the correctness and reproducibility of applied mathematics and informatics methods

— Is able to maintain collective scientific communication, organize scientific events.

— Is able to organize the training of specialists in the field of applied mathematics in new methods and tools in accordance with the direction of training and specialization.

— Is able to analyze and reproduce the meaning of interdisciplinary texts using the language and apparatus of applied mathematics and informatics.

— Is able to create interdisciplinary texts using the language and apparatus of applied mathematics and informatics.

— Is able to formalize and present publicly the results of professional activity using information technologies.

— Is able to carry out a targeted multi-criteria search for information on the latest scientific and technological advances on the Internet and in other sources.

— Is able to create, describe and responsibly control the implementation of technological requirements and regulations in professional activities

— Is able to collect, clean, analyze and visualize large data

Universal competencies

— Is able to reflect (evaluate and process) the learned scientific methods and ways of activity.

— Is able to develop new theories, invent new ways and tools of professional activity.

— Is able to independently master new research methods, change the scientific and production profile of its activities

— Is able to improve and develop their intellectual and cultural level, build a track of professional development and career.

- Is able to make management decisions and ready to take responsibility for them
- Is able to analyze, verify, evaluate the completeness of information in the course of professional activities, if necessary, to fill in and synthesize missing information.
- Is able to organize and manage multilateral communication.
- Is able to conduct research activities in the international environment.

Programme Modules

Required Core Courses

This M.S. is based on the newly created graduate courses in statistical theory and methods taught by the faculty of the laboratory and renown Russian and international faculty. All candidates for this degree must take “Contemporary Data Analysis: Methodology of Interdisciplinary Research” and “Contemporary Decision Sciences Methods: an Integrated Perspective.” These two courses lay the foundation for the systems thinking that this program aims to develop. “Contemporary Data Analysis: Methodology and Methods of Interdisciplinary Research” is designed as a “gateway” to graduate work in statistics, where the mathematical concepts are bridged with applied concepts and research design, depending on the discipline. “Contemporary Decision Sciences Methods: an Integrated Perspective” provides a unified perspective that is aimed at developing improved decision-making process, where one needs to understand how decisions are made in practice and in what ways behavior differs from guidelines implied by normative theories of choice.

“Applied Linear Models” serves as a foundational course for all mathematical thinking in applied statistics and subsequent courses taken in the program. Statistical Consulting, an equivalent to which does not appear to be offered by competing programs, is designed to establish firm foundations for working with “someone else’s” data, extracting relevant information from it, and preparing easy-to-understand reports for accurate use by clients with no statistical knowledge. Foundational and advanced courses in network analysis are focused on developing the critical analytical skills necessary for working with network data – the emphasis of this program. Required courses will be offered every year, to be taken by incoming new students only.

Additional Requirements

Given the emphasis of the laboratory on network statistical analysis, the lab is offering an emphasis on network methods. Students interested in obtaining a specialization in network methods must take additional courses in networks.

Students not wishing to pursue the network component are welcome to choose from the remaining course offerings. All students must select a total of 15 courses from the offerings of the program, subject to the following restrictions:

1. Most elective courses will be offered every other year, to be taken by both Year 1 and Year 2 students. Only some elective courses, considered essential to the program (such as Network Analysis), will be also offered every year.
2. As needed, courses could be offered as short courses, taught by invited professors from internationally recognized research universities.
3. Some courses may not be offered in any given two-year period if there is not enough student interest or demand in opening the course. Exact number of students, required to open the course, will be determined by the program director, depending on the total number of students enrolled in the program. Course may not be offered even if there are enough interested students if such students do not meet prerequisite requirements for the course.
4. Regardless of the student enrollment numbers, a choice of courses will be offered to provide students with selection between more and less advanced courses from the standpoint of mathematics.

Programme Courses

1. Contemporary Data Analysis: Methodology of Interdisciplinary Research

- a. Prerequisites: One statistics course at the undergraduate level.
- b. Required course

This foundational course is designed to put together a unified research program for people from diverse disciplines. Its main purpose is to provide students with a firm foundation of research methodology, including topics in research design, theory building and testing with hypotheses generation, and advanced academic writing topics. This course is about conducting research, both in academia and in practice. Specifically, the students will focus on basic steps of the scientific inquiry, starting with the topic selection, and progressing through to literature review, hypotheses generation, choice of analysis method, and methods of propagating the research results to wide audiences (written and oral presentations). Whether they plan to work in the corporate world, or develop career in academia, they will be forced to generate knowledge and disseminate it to others, so there is no doubt that they will use the skills acquired in this course.

2. Contemporary Decision Sciences Methods: an Integrated Perspective

- a. Prerequisites: intro to stats or consent of the instructor
- b. Required Course

This course is designed as an overview of a range of problems and applications to managerial decision making using scientific and analytical methodology. Topics include concepts and applications of the decision support system, including type of decisions, type of decision makers, modeling decisions, decisions within organizations, rule based expert systems, and simulation as a DSS application. This course also covers practical issues in DSS such as using Integer and Linear Programming as applications of modeling and solving choices and uncertainties of real world decision problems. Topics covered also include sensitivity analysis and an introduction to decision analysis. Problem recognition, model building, model analysis and managerial implications are the primary objectives with special emphasis on understanding the concepts and computer implementation and interpretation.

3. Machine Learning

- a. Prerequisites: intro to stats and linear algebra (or equivalent courses), or consent of instructor.
- b. Optional course

The main goal of statistical learning theory is to provide a framework for studying the problem of inference, that is of gaining knowledge, making predictions, making decisions or constructing models from a set of data. This is studied in a statistical framework, that is there are assumptions of statistical nature about the underlying phenomena (in the way the data is generated). Therefore, the course will cover the fundamental concepts and principles of data reduction and statistical inference, including the method of maximum likelihood, the method of least squares, and Bayesian inference. The course also introduces minimal sufficiency; exponential families; theory estimation, theory of optimal tests, and confidence intervals; robustness; and decision theory. Upon successfully completing this course, students will be able to:

- Discuss classical theory/methods for drawing statistical inference
- Discuss the statistical reasoning and theoretical justification behind two main streams in inference: hypothesis testing and estimation (point and interval)

4. Probability Theory

This course covers standard introductory probability theory topics such as probability spaces, discrete and continuous random variables, transformations, expectations, generating functions, conditional distributions, law of large numbers, central limit theorems, as well as advanced topics that are likely to be the most useful to someone planning to use research from the modern theory of stochastic processes in their daily work. The course has an applied component, with real-life applications examples of probability theory.

5. Nonparametric Theory and Data Analysis

- a. Prerequisites: Two statistics courses at the graduate level, or consent of instructor.

b. Optional course

The course is an introduction to statistics outside of the "classical" techniques. Over and above the material itself, the course is useful for reinforcement of and elaboration on concepts of testing and estimation seen in classical courses, and serves as a bridge to modern, computationally intensive branches of statistics like machine learning. Topics covered include statistical functionals, bootstrapping, empirical likelihood. Nonparametric density and curve estimation. Rank and permutation tests.

6. Bayesian Theory and Data Analysis

a. Prerequisites: Two statistics courses at the graduate level, or consent of instructor.

The course covers an introduction to the theory and practice of Bayesian inference. Topics covered include: Prior and posterior distributions, Bayes theorem, model formulation, Bayesian computation, model checking and sensitivity analysis. This is a general class on Bayesian methods. Some basic knowledge of probability distributions, calculus and linear algebra is assumed. We will examine Bayesian inference and prediction for simple parametric models, regression models, hierarchical models and mixture models that span a wide variety of applied data settings. In each of these areas, we will compare and contrast the Bayesian and classical viewpoints for data analysis. We will develop a wide range of methods for model implementation, including optimization algorithms and Markov chain Monte Carlo simulation techniques. We will also examine strategies for model evaluation and validation.

Course participants will have interest in applied data analysis as well as basic knowledge of principles for statistical inference and prediction. Participants should also have experience with basic probability topics, such as probability density functions, marginal and conditional probabilities, as well as transformation and simulation of random variables. We will be implementing our models using the statistical software package R, though prior experience with R is not required for the course.

7. Applied Linear Models

a. Prerequisites: intro to statistics and linear algebra (or equivalent courses), or consent of instructor.

An advanced course in applied statistics, Linear models will be used to treat a wide range of regression and analysis of variance methods. Topics include: matrix review; multiple, curvilinear, nonlinear, and stepwise regression; correlation; residual analysis; model building; use of the regression computer packages; use of indicator variables for analysis of variance and covariance models. The first part of the course will emphasize linear regression and the analysis of variance, including topics from the design of experiments and culminating in the general linear model. The second part – topics from experimental design.

In addition, most experimental situations measure several dependent variables. When studying the variables one at a time, one usually does not get a complete picture of the experimental results. For this reason, in recent years, multivariate methods, in which the variables are studied simultaneously have become increasingly popular. This course covers both the underlying theory required to understand the multivariate methods, as well as their applications in data analysis. Some of the methods/models covered in the course are principal component analysis, factor analysis, discriminant analysis, multivariate analysis of variance (MANOVA), PLS, cluster analysis and multivariate analysis of repeated measurements. The course includes computer labs where multivariate data analysis is performed using statistical software.

8. Categorical Data Analysis

a. Prerequisites: Two statistics courses at the graduate level, or consent of instructor.

The analysis of cross-classified categorical data. Loglinear models; regression models in which the response variable is binary, ordinal, nominal, or discrete. Logit, probit, multinomial logit models; logistic and Poisson regression.

This class focuses on the basic regression models for categorical dependent variables. While advances in

software have made it simple to estimate these models, post-estimation interpretation is difficult due to the nonlinearities of the models. The class begins by considering the general objectives for interpreting the results of any regression type model and then considers why achieving these objectives is more difficult with nonlinear models. Basic concepts and notation are introduced through a review of the linear regression model. Within this familiar context, the method of maximum likelihood estimation is presented. These ideas are used to develop the logit and probit models for binary outcomes. A variety of practical methods for interpreting nonlinear models are presented. The models and methods of interpretation for binary outcomes are extended to ordinal outcomes using the ordinal logit and probit models. The multinomial logit model for nominal outcomes is then discussed. Finally, a series of models for count data, including Poisson regression, negative binomial regression, and zero modified models are presented. A major component of the course is using Stata to estimate and interpret the models and particularly the special commands for post-estimation interpretation. The course assumes familiarity with the linear regression model. Familiarity with Stata is not assumed.

9. Multilevel Models

Prerequisites: Two statistics courses at the graduate level, or consent of instructor.

This course is designed to provide students with a training experience in the concept and application of multilevel statistical modeling. You will be motivated to think about correlated and dependent data structures that arise due to sampling design and/or are inherent in the population (such as pupils nested within schools; patients nested within clinics; individuals nested within neighborhoods and so on). The substantive purpose of this course is to enable quantitative assessments on the role of contexts (e.g., schools, clinics, neighborhoods) in predicting individual outcomes. This will be accomplished by developing a range of multilevel models along with a detailed discussion of the statistical properties and the interpretation of each model. Empirical presentations and homework assignments will focus on multilevel analysis using MLwiN – a specialized software to handle models with complex data structures.

Topics covered include: Introduction to the general multilevel model with an emphasis on applications. Discussion of hierarchical linear models, and generalizations to nonlinear models. How such models are conceptualized, parameters estimated and interpreted. Model fit via software. Major emphasis throughout the course will be on how to choose an appropriate model and computational techniques.

10. Structural Equation Modeling

a. Prerequisites: Two statistics courses at the graduate level, or consent of instructor.

Path analysis. Introduction to multivariate multiple regression, confirmatory factor analysis, and latent variables. Structural equation models with and without latent variables. Mean-structure and multi-group analysis.

This course is designed for students and faculty who would like to acquire a significant familiarity with statistical techniques known collectively as "structural equation modeling," "causal modeling," or "analysis of covariance structures." As learning in this course demands basic understanding of statistical principles and techniques such as regression and factor analysis, the course will start with an overview of basic applied statistics and linear algebra, and will progress to more complex models in a sequential manner. The goals of the course are: To ensure that students understand topics and principles of applied statistical techniques; to provide students with an understanding of the basic principles of latent variable structural equation modeling and lay the foundation for future learning in the area; to explore the advantages and disadvantages of latent variable structural equation modeling, and how it relates to other methods of analysis; to develop student familiarity, through hands-on experience, with the major structural equation modeling programs, so that they can use them and interpret their output; to develop and/or foster critical reviewing skills of published empirical research using structural equation modeling.

11. Time Series Analysis

a. Prerequisites: Two statistics courses at the graduate level, or consent of instructor.

Techniques for analyzing data collected at different points in time. Probability models, forecasting methods, analysis in both time and frequency domains, linear systems, state-space models, intervention analysis, transfer function models and the Kalman Filter. Topics also include: Stationary processes, autocorrelations, partial autocorrelations, autoregressive, moving average, and ARMA processes, spectral density of stationary processes, periodograms and estimation of spectral density.

Students are assumed to understand basics of statistical inference, regression analysis, and scalar and matrix algebra. Some topics that will be covered include ARIMA models, intervention analysis, regression analysis of time series, cointegration, error correction models, vector autoregression, pooled time series, and time varying parameter models.

12. Data Mining

a. Prerequisites: Two statistics courses at the graduate level, or consent of instructor.

Covers topics in data mining, including visualization techniques, elements of machine learning theory, classification and regression trees, Generalized Linear Models, Spline approach, and other related topics.
Literature

13. Exploratory Data Analysis

a. Prerequisites: Two statistics courses at the graduate level, or consent of instructor.

Numerical and graphical techniques for summarizing and displaying data. Exploration versus confirmation. Connections with conventional statistical analysis and data mining. Applications to large data sets.

14. Machine Learning

a. Prerequisites: Two statistics courses at the graduate level, or consent of instructor.

This course will take a modern, data-analytic approach to the multiple regression model. Our coverage of the material will emphasize the ways that graphical tools can augment traditional methods for describing how the conditional distribution of a dependent variable changes along with the values of one or more independent variables. The course will examine the basic nature and assumptions of the linear regression model, diagnostic tools for detecting violations of the regression assumptions, and strategies for dealing with situations in which the basic assumptions are violated..

15. Methods of Statistical Consulting

a. Prerequisites: Consent of instructor.

Development of effective consulting skills, including the conduct of consulting sessions, collaborative problem-solving, using professional resources, and preparing verbal and written reports. Real-life clients could be obtained from companies in Moscow; to them, service will be provided for free.

16. Network Analysis

a. Prerequisites: none

An introduction to various concepts, methods, and applications of social network analysis drawn from the social and behavioral sciences. The primary focus of these methods is the analysis of relational data measured on groups of social actors. Topics to be discussed include a basic introduction to network analysis, graphs and matrices, basic network measures and visualization, reciprocity and transitivity, dyadic and triadic analysis, centrality, egocentric networks, two-mode networks (affiliations, bibliographic/scientometric analysis), cohesive subgroups, equivalences and blockmodeling, hubs & authorities, cores & peripheries, clustering and graph partitioning, large scale structure of networks, statistical modeling in network (ergm/p*/RSiena) and network dynamics, and change in networks.

17. Advanced Topics in Network Analysis

a. Prerequisites: introduction to network analysis or consent of the instructor

The conventional categorization of data analytic methods into descriptive and inferential statistics can be fruitfully applied to network analysis. Descriptive methods of network analysis are important for illuminating structural features of a given network, but they cannot be used to build and/or test theories about the generation of networks. Inferential methods of network analysis can be used to test hypotheses about the generation and evolution of a network, derive measures of uncertainty for network indices, and find probabilistic models that accurately describe the overall features of a network.

18. Network Analysis: Statistical Approaches and modeling

a. Prerequisites: introduction to network analysis or consent of the instructor

Advanced statistical methods for analyzing social network data, focusing on testing hypotheses about network structure (e.g. reciprocity, transitivity, and closure), the formation of ties based on attributes (e.g. homophily), and network effects on individual attributes (social influence or contagion models). Statistical models (blockmodeling, diffusion, etc.)

19. Network Analysis: Application in R

a. Prerequisites: none

The focus of the course will be how to develop questions about social networks and appropriately test them using the R statistical programming language. Because it is critically important for researchers to be able to analyze the data, and standardized packages hardly ever offer the required set of analytic methods, we are faced with having to write our own code for analysis of specific datasets. Minimal programming skills are desirable, though not required.

20. Introduction to Statistics

a. Prerequisites: introduction to network analysis or consent of the instructor

This course is an introductory course in network analysis, designed to familiarize graduate students with the general concepts and basic techniques of network analysis in sociological re-search, gain general knowledge of major theoretical concepts and methodological techniques used in social network analysis, and get some hands-on experience of collecting, analyzing, and mapping network data with SNA software. In addition, this course will provide ample opportunities to include network concepts in students' master theses work.

21. Programming in R and Python

a. Prerequisites: introduction to network analysis or consent of the instructor

Students who have never programmed are afraid that it is difficult. This course is designed to introduce them to the basics of programming languages such as R and Python. This course will discuss the difference between these languages, the strengths of each of them. Students will learn the basics of programming and working with these languages.

Options for Students with Disabilities

This degree programme of HSE University is adapted for students with special educational needs (SEN) and disabilities. Special assistive technology and teaching aids are used for collective and individual learning of students with SEN and disabilities. The specific adaptive features of the programme are listed in each subject's full syllabus and are available to students through the online Learning Management System.

Programme Documentation

All documents of the degree programme are stored electronically on this website. Curricula, calendar plans, and syllabi are developed and approved electronically in corporate information systems. Their current versions are automatically published on the website of the degree programme. Up-to-date teaching and learning guides, assessment tools, and other relevant documents are stored on the website of the degree programme in accordance with the local regulatory acts of HSE University.



I hereby confirm that the degree programme documents posted on this website are fully up-to-date.

Vice Rector Sergey Yu. Roshchin

[Summary of Degree Programme 'Applied Statistics with Network Analysis'](#)

