



LITHUANIAN UNIVERSITY OF HEALTH SCIENCES

APPROVED
Kaunas Medical University
Senate
17 th of December in 2004
Resolution Nr. 3-11

RENEWED
24 th of October, 2024

**APPLICATION OF MATHEMATICAL STATISTICS AND ARTIFICIAL
INTELLIGENCE IN DATA SCIENCE
DOCTORAL STUDIES SUBJECT PROGRAM**

Subject program coordinators:

Department of Physics, Mathematics and Biophysics assoc. prof. dr. Renata Paukštaitienė

Departments participating in the subject program: Department of Physics, Mathematics and
Biophysics

Kaunas, 2024

Subject program data

| | |
|-----------------------|--|
| Fields of science | Medical and Health Sciences, Natural Sciences, Agricultural Sciences |
| Field of study (code) | For all fields |
| Title of the subject | Application of Mathematical Statistics and Artificial Intelligence in Data Science |
| Scope of the program | 160 hours. (6 ECTS) |
| Lectures | 40 h |
| Seminars | 40 h |
| Independent work | 80 h |

Subject program preparation group

| Eil. Nr. | Name, Surname | Responsibilities | Telephone (work) | e-mail address |
|----------|---------------------------------------|-----------------------------------|------------------|--|
| 1. | Assoc. prof. dr. Renata Paukstaitienė | Head of the department, professor | 327-370 (5758) | renata.paukstaitiene@ismuni.lt |
| 2. | prof. dr. Algimantas Kriščiukaitis | professor | 327-367 | algimantas.krisciukaitis@ismuni.lt |
| 3. | dr. Vita Špečkauskienė | associate professor | 327-367 | vita.speckauskiene@ismuni.lt |

DESCRIPTION OF THE SUBJECT PROGRAM

1. The need for the subject program: Nowadays, doctoral student of the field of medicine, health, natural or agricultural sciences must be able to apply the methods of mathematical statistics and artificial intelligence in data analysis using computer program packages.

2. Objectives of the subject program:

1. To provide doctoral students with knowledge and practical skills about the collection, systematization and presentation of the health data and the estimation of the necessary minimum size of the research sample according to the purpose and tasks of the planned scientific work.
2. To provide doctoral students with knowledge and practical skills in choosing and using the necessary methods of statistical data analysis, in formulating the conclusions of research and in presenting the results.
3. To provide doctoral students with knowledge and practical skills in choosing and using the necessary artificial intelligence methods in data analysis.
4. To provide doctoral students with practical skills in using statistical analysis computer software packages.

THEORETICAL PART - 40 hours

| No | Title and Short Description of Lecture | Duration | Teacher |
|--|---|----------|--|
| <i>Application of Mathematical Statistics</i> | | | |
| 1. | Introduction to data analysis Population and sample. Sampling methodology. Interval estimation of population characteristics by sample characteristics | 2 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| 2. | Testing of statistical hypotheses. Parametric tests An introduction to statistical hypothesis testing. Parametric tests for one sample and two samples. The minimum sample size. | 4 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| 3. | Analysis of variance Analysis of variance (ANOVA). Analysis of covariates (ANCOVA). Analysis of variance of blocked data. | 4 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| 4. | Testing of statistical hypotheses. Rank Tests Wilcoxon, Mann-Whitney, Kruskal Wallis and Friedman tests. | 2 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| 5. | Correlation and regression analysis of quantitative data Correlation coefficients. Linear regression | 2 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| 6. | Statistical hypothesis testing by analyzing qualitative data Chi-squared test and evaluation of baseline estimates, Cochran's Q test | 2 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| 7. | Correlation and regression analysis of qualitative data Coefficients of association. Binomial logistic regression, multinomial logistic regression and ordinal logistic regression models, their selection and application. | 4 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| 8. | Survival analysis. Survival tables. Kaplan-Meier estimates. Comparison of survival functions in groups. Regression models of survival. | 2 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| 9. | Analysis of questionnaire data. Reliability and validity analysis of the questionnaire. Factor analysis. | 2 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| 10. | Systematic literature review and meta-analysis | 2 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| <i>Application of Artificial Intelligence in Data Analysis</i> | | | |
| 11. | Basics of artificial intelligence. Machine and deep learning in health data analysis (open and big data analytics). Open and big data analysis. Concept and classification, overview of basic methods | 2 | dr. R. Petrolis prof. A. Kriščiukaitis dr. V. Špečkauskienė |
| 12. | Application of decision trees for the determination of diagnostically important risk factors | 3 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| 13. | Fundamentals of applying structural equation models in biomedical data analysis | 2 | dr. R. Paukštaitienė dr. I. Grabauskytė dr. V. Špečkauskienė |
| 14. | Cluster analysis. The purpose and stages of cluster analysis, its mathematical model. Measures of object similarity and classification of methods. | 2 | dr. R. Petrolis dr. A. Kriščiukaitis |

| No | Title and Short Description of Lecture | Duration | Teacher |
|-----|---|----------|---|
| 15. | Discriminant analysis The purpose and stages of discriminant analysis, its model. Selection of the classification rule, its reliability. | 2 | dr. R. Petrolis prof. A. Kriščiukaitis |
| 16. | Application of artificial neural networks and other classifiers in data analysis | 3 | dr. R. Petrolis prof. A. Kriščiukaitis |

PRACTICAL PART – 40 hours.

Practical classes and seminars are held in computer classrooms. During them, doctoral students use computer statistical programs to perform statistical analysis of the given data by applying the theoretical material presented in the lectures.

| No | Title and Short Description of Lecture | Duration | Teacher |
|--|--|----------|---|
| <i>Application of Mathematical Statistics</i> | | | |
| 1. | Introduction to data analysis | 2 | dr. R. Paukštaitienė prof. A. Kriščiukaitis dr. I. Grabauskytė dr. V. Špečkauskienė dr. R. Petrolis |
| 2. | Testing of statistical hypotheses. Parametric tests | 4 | |
| 3. | Analysis of variance | 4 | |
| 4. | Testing of statistical hypotheses. Rank Tests | 2 | |
| 5. | Correlation and regression analysis of quantitative data | 2 | |
| 6. | Statistical hypothesis testing by analyzing qualitative data | 2 | |
| 7. | Correlation and regression analysis of qualitative data | 4 | |
| 8. | Survival analysis. | 2 | |
| 9. | Analysis of questionnaire data. | 2 | |
| 10. | Systematic literature review and meta-analysis | 2 | |
| 11. | Practical tests | 4 | |
| <i>Application of Artificial Intelligence in Data Analysis</i> | | | |
| 12. | Basics of artificial intelligence. | 2 | dr. R. Petrolis dr. V. Špečkauskienė dr. I. Grabauskytė dr. A. Kriščiukaitis dr. R. Paukštaitienė |
| 13. | Machine and deep learning in health data analysis (open and big data analytics). | 2 | |
| 14. | Application of decision trees for the determination of diagnostically important risk factors | 2 | |
| 15. | Fundamentals of applying structural equation models in biomedical data analysis | 2 | |
| 16. | Cluster analysis and discriminant analysis | 2 | |
| 17. | Application of artificial neural networks and other classifiers in data analysis | | |
| 18. | Test | 2 | |

Independent work - 80 hours.

During independent work, doctoral students study special literature, critically analyze scientific articles, use information databases, independently solve additional tasks obtained during practice seminars. Throughout this work, doctoral students are advised by professors.

EVALUATION OF STUDIES

In the course of mathematical statistics and its applications cumulative assessment of knowledge is applied. The practical skills of doctoral students are evaluated during the control works. Practical data analysis tasks are solved by using computer software. Acquired theoretical knowledge is evaluated during the exam at the end of the course. The exam task consists of 3 questions related to the material of the studied subject.

The cumulative grade of the course consists of: First control work 30 % + second control work 40 % + 30% third test.

RECOMMENDED LITERATURE

| No. | Title | Authors | Years of publication |
|-----|--|--|------------------------|
| 1. | Basics of Biostatistics | R. Šimoliūnienė, J. Tomkevičiūtė, Ž. Jokšienė et.al. | 2016, Kaunas, |
| 2. | Artificial Intelligence By Example | Rothman, Denis | 2018, Packt Publishing |
| 3. | Design and analysis of Clinical Trials | Shein-Chung Chow, Jen-Pei Liu | 2014, Wiley |

LIST OF TEACHERS:

- Professors or senior researchers teaching in the subject program:
 prof. dr. Algimantas Kriščiukaitis
 prof. dr. Renata Paukštaitienė
- Other teachers of the subject program - mathematicians:
 assoc. professor dr. Ingrida Grabauskytė
 assoc. professor Robertas Petrolis
 assoc. professor Vita Špečkauskienė