

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
ИНСТИТУТ ПО МАТЕМАТИКА И ИНФОРМАТИКА

сигнатура:				
4.5	TPMS	S	06	v1
професионално направление	код на докт. програма	вид курс (базов/спец.)	номер	версия
попълва се административно след приемане от НС на ИМИ				

Утвърдил:

(чл.-кор. П. Бойваленов, Директор на ИМИ-БАН)

Учебна програма
за специализиран докторантски курс

Област на висше образование:	4. Природни науки, математика и информатика
професионално направление:	4.5. Математика
докторска програма:	Теория на вероятностите и математическа статистика
тема:	Статистически анализ на експериментални данни
лектор:	доц. д-р Асен Чорбаджиев
данни за връзка с лектора (тел., имейл)	atchorbadjieff@math.bas.bg 0892-59070
хорариум:	20 часа лекции и 20 часа практически упражнения
кредити съгл. кредитната система на ЦО на БАН:	20

1. Анотация

Съвременните научни изследвания силно зависят от наличните експериментални данни. За тази цел от особена значимост е правилното събиране и обработка на все повече качествени експериментални данни. Липсата на познания и умения при работата с данни често могат да подложат на изпитание и провалят всеки съвременен научен експеримент. Затова целта на този курс ще бъде насочена за подготовка на докторанти физика, химия, биология и науки за земята, икономика и социология за работа с данни, тяхната точната класификация, обработка и основните методи за анализ. Курсът съдържа обяснителни лекции и практически компютърни упражнения на R или Python.

2. Необходими предварителни знания

Опит при работа с Ексел и програмиране с R и Python е предимство.

3. Компетентности, придобити в резултат на обучението

Първична и базова обработка на експериментални данни. Разработка на скриптове за ежедневна обработка на данни.

4. Тематично съдържание

Тема	брой часове лекции	брой часове практически упражнения
1. Увод в статистиката - Типове данни, вероятност и вероятностни разпределения. Основни понятия. - Оценки и доверителни интервали. - Параметрични и непараметрични тестове. - Линейна регресия.	10	10
2. Методи за напреднали - Подготовка на експеримента и качество на данни. Откриване и премахване на Outliers. - Времеви редове. - GLM регресия. - Дисперсионен анализ. Смаляване на размерности. - Пространствени данни. - Cross-validation. Bootstrapping. Автоматизирани статистически пресмятания.	10	10

5. Конспект

1. Типове данни и основни понятия в статистиката
2. Оценяване на параметри. Точкови оценки и доверителни интервали
3. Параметрични и непараметрични тестове: същност и приложение
4. Линейна регресия: изграждане, оценка и интерпретация на модела
5. Подготовка на данни за анализ. Качество на данните
6. Откриване и обработка на отклоняващи се стойности (outliers)
7. Времеви редове: анализ и моделиране
8. Генерализирани линейни модели (GLM): същност и приложение
9. Методи за намаляване на размерността на данни
10. Анализ на пространствени данни
11. Оценка на модели чрез cross-validation
12. Bootstrapping методи в статистическия анализ

6. Препоръчана литература:

1. Edward W. Frees; Regression Modeling with Actuarial and Financial Applications. (1993)
2. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel; Time Series Analysis, (1970)
3. Ton J. Cleophas, A. H. Zwingerman, et.al.; Statistics Applied To Clinical Trials, (2009)
4. Robert I. Kabacoff; R in Action; (2011)
5. Efron, Bradley, and Trevor Hastie; Computer age statistical inference, student edition: algorithms, evidence, and data science. Vol. 6. Cambridge University Press, (2021).

7. Ресурсно осигуряване на обучението:

Курсът е планиран за ползване на отворени за ползване програмни продукти.
Данните за компютърните практикуми ще бъдат осигурени от преподавателя.

8. Критерии за оценка

От докторанта се изисква разработка на проект с тематика близка до работната тематика по докторантурата на студента. Изпитът се състои от представяне на разработената курсова работа. Изпитваният трябва да защити своето решение пред жури и да предаде валидни програмни кодове, разработени в рамките на курсовата работа.

Крайната оценка е от 2 до 6 (с точност до 0.5).

Тя се формира на базата на следното съответствие:

Отличен (6 или 5.50)	Отлично владее материала. Изложението е изчерпателно, последователно, компетентно, логично и хармонично. Правилно обосновава предлаганите решения, знае как да обобщава и излага материала без да прави грешки. Умее да изпълнява практически задачи.
Мн. добър (5 или 4.50)	Познава материала. Излага го правилно без да допуска съществени неточности. Притежава умения за изпълнение на практически задачи с леки неточности.
Добър (4 или 3.50)	Владее голяма част от материала, но допуска неточности при изложението и отговорите на въпросите. Затруднява се при прилагането на материала в практически ситуации.
Среден (3)	Владее само част от материала, но се затруднява в отделните детайли. Допуска неточности във формулировките и нарушава последователността при представянето. Има затруднения при изпълнение на практически задачи.
Слаб (2)	Не познава значителна част от материала, допуска съществени грешки и с големи трудности изпълнява практически задачи.

Учебната програма е обсъдена и одобрена на заседание на секция „Изследване на опрещите, вероятности и статистика“ на 04.02.2025 г.

Ръководител секция:

(доц. Цветелин Заевски)

Разгледана от Директорския съвет на ИМИ-БАН на 13.02.2025 (протокол № 7).

Приета от Научния съвет на ИМИ-БАН на 07.03.2025 (протокол № 3).

**BULGARIAN ACADEMY OF SCIENCES
INSTITUTE OF MATHEMATICS AND INFORMATICS**

Signature:				
4.5	TPMS	S	06	v1
Professional Field	PhD Programme Code	Course Type	Number	Version
<i>To be filled in after the acceptance by the Scientific Council of IMI</i>				

Approved:
(Prof. DSc P. Boyvalenkov, Director of IMI-BAS)

Curriculum of a Specialized PhD Course

Higher Education Area:	4. Natural Sciences, Mathematics and Informatics
Professional Field:	4.5 Mathematics
PhD Programme:	Theory of Probabilities and Mathematical Statistics
Theme:	Statistical Analysis of Experimental Data
Lecturer:	Assoc. Prof. Assen Tchorbadjieff
Contact Details of the Lecturer (phone, email):	atchorbadjieff@math.bas.bg +359-892-59070
Hours:	20 hours of lectures and 20 hours of practical exercises
Credits According to the Credit System of the Training Centre of BAS:	20

1. Annotation

Modern scientific research heavily relies on the availability of experimental data. For this reason, proper collection and processing of increasingly high-quality experimental data is of particular importance. A lack of knowledge and skills in working with data can often challenge and even compromise any modern scientific experiment.

Therefore, the goal of this course is to prepare doctoral students in physics, chemistry, biology, earth sciences, economics, and sociology for working with data, including its accurate classification, processing, and fundamental methods of analysis. The course consists of explanatory lectures and practical computer exercises using **R** or **Python**.

2. Prerequisites

Working experience with Excel and programming with R or Python is advantage.

3. Expected Learning Outcomes

Primary and basic processing of experimental data. Development of scripts for daily data processing.

4. Topical Outline of Content

Topic	Hours Lectures	Hours Exercises
1. Introduction to Statistics - Types of data, probability, and probability distributions. Basic concepts. - Estimates and confidence intervals. - Parametric and non-parametric tests. - Linear regression.	10	10
2. Advanced Methods - Experiment design and data quality. Detection and removal of outliers. - Time series analysis. - GLM regression. - Analysis of variance (ANOVA) and dimensionality reduction. - Spatial data analysis. - Cross-validation, bootstrapping, and automated statistical computations.	10	10

5. Questionary

1. Types of data and basic concepts in statistics
2. Parameter estimation. Point estimates and confidence intervals
3. Parametric and non-parametric tests: essence and application
4. Linear regression: model building, estimation, and interpretation
5. Data preparation. Data quality
6. Detection and handling of outliers
7. Time series: analysis and modeling
8. Generalized linear models (GLM): essence and application
9. Methods for dimensionality reduction
10. Spatial data analysis
11. Model evaluation through cross-validation
12. Bootstrapping methods in statistical analysis

6. References

1. Edward W. Frees; Regression Modeling with Actuarial and Financial Applications. (1993)
2. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel; Time Series Analysis, (1970)
3. Ton J. Cleophas, A. H. Zwingerman, et.al.; Statistics Applied To Clinical Trials, (2009)
4. Robert I. Kabacoff; R in Action; (2011)
5. Efron, Bradley, and Trevor Hastie; Computer age statistical inference, student edition: algorithms, evidence, and data science. Vol. 6. Cambridge University Press, (2021).

7. Resource provision of training

The course is designed for the use of open-source software. The data for the computer practicums will be provided by the instructor.

8. Evaluation criteria

The PhD student have to develop a project related to the student's doctoral research topic. The exam consists of presenting a developed coursework project. The examinee must defend their solution before a jury and submit valid program codes developed for the coursework project.

It is formed based on the following correspondence:

Excellent (6 or 5.50)	Excellent command of the material. Comprehensive, consistent, competent, logical and harmonious presentation. Proper justification of the proposed solutions, good summary and presentation of the material without making mistakes. Good necessary skills to perform practical tasks.
Very good (5 or 4.50)	Satisfactory command of the material. Correct explanation without significant inaccuracies. Proper application of the theoretical principles and appropriate performance of practical tasks.
Good (4 or 3.50)	Good command of the material, but with inaccuracies in the presentation and in the answers to questions. There are some ambiguities in attempts to apply the material in practical situations.
Average (3)	Limited command of the material, difficulties in the individual details. Inaccuracies in the wording and inconsistency in the presentation of the material. Difficulties in the performing of practical tasks.
Weak (Failing grade) (2)	A significant part of the material is not known, serious mistakes are made, and the practical tasks are performed with great difficulty.

The curriculum was discussed and approved at a meeting of the Department “Operational Research, Probability and Statistics”, held on 04.02.2025

Head of Department:

(Assoc. Prof. Ts. Zaeovski)

Approved by the Board of Directors of IMI-BAS on 13.02.2025 (Minutes No. 7)

Accepted by the Scientific Council of IMI-BAS on 07.03.2025 (Minutes No. 3)