Syllabus
Course Program

Mathematical statistics

Specialty
113 – Applied Mathematics

Institute
Institute of Computer Modeling, Applied Physics and Mathematics

Educational program
Computer and mathematical modeling

Department
Mathematical modeling and intelligent computing in engineering (161)

Level of education
Bachelor's degree

Course type
Special (professional), Compulsory

Semester
5

Language of instruction
English

Lecturers and course developers

Oleksiy Larin

Oleksiy.Larin@khpi.edu.ua
Doctor of Engineering Science, Professor,
Director of the Institute of Computer Modeling, Applied Physics and Mathematics. Specialist in computational mechanics, probabilistic modeling and reliability prediction. The main focus of his scientific works is on the development of models, methods, approaches and algorithms for computer modeling and statistical analysis of engineering systems, in particular with random parameters. Author of over 150 scientific and methodological works

Learn more about the teacher on the department's website

Ksenia Potopalska (assistant in laboratory practice)
Kseniia.Potopalska@khpi.edu.ua

Candidate of technical sciences, associate professor of the department of mathematical modeling and intellectual calculations in engineering, work experience - 7 years. Author of more than 40 scientific and methodical works works. Leading lecturer on disciplines : " Modeling and reverse engineering based on data ",
Modeling and reverse engineering based on of data " , teacher of practice

Learn more about the teacher on the department's website

General information

Summary

The statistics course is aimed at developing theoretical knowledge and practical skills in students to solve applied problems in the field of mathematical statistics. Students develop abstract thinking, analytical abilities and programming skills, using probabilistic statistical methods of data analysis, forecasting and making informed decisions. The course includes lectures, computer laboratory practicals and team

project work. The main topics include the study of descriptive statistics, the basics of statistical diagnostics and statistical data analysis. The team project allows students to apply the acquired knowledge in practice and demonstrate their skills in solving applied problems. The course takes into account modern information technologies and provides students with the opportunity to gain experience in using Python to solve problems of mathematical statistics. Final control - exam.

Course objectives and goals

The purpose of the course in mathematical statistics is to form in future specialists theoretical knowledge and practical skills in the application of probabilistic and statistical methods to solve applied problems in the field of modeling, data analysis, forecasting and decision-making, in particular with the use of modern information tools and when working in a team. The following goals are set:
1) Providing students with knowledge and ideas about the use of mathematical concepts and methods for statistical data analysis, developing forecasts and making statistically sound decisions in various fields.
2) Developing abstract thinking and analytical skills, allowing students to implement probabilistic and statistical methods in practice.
3) Providing practical experience in implementing statistical analysis algorithms and improving programming skills.
4) Improve the ability to independently formalize and solve problems using probabilistic and statistical methods, as well as to effectively collaborate in a team and interact with other specialists in the chosen field.

Format of classes

Lectures, laboratory classes, consultations. Final control in the form of an exam.

Competencies

GC06. Ability to abstract thinking, analysis and synthesis.
GC 13. Interpersonal skills.
SC01. Ability to use and adapt mathematical theories, methods and techniques to prove mathematical statements and theorems.
SC 02. Ability to perform tasks formulated in mathematical form.
SC 03. Ability to select and apply mathematical methods to solve applied problems, modeling, analysis, design, management, forecasting, decision-making.
SC 06. Ability to solve professional tasks using computer technology, computer networks and the Internet, in the environment of modern operating systems, using standard office applications.

Learning outcomes

LO2. To master the basic principles and methods of mathematical, complex and functional analysis, linear algebra and number theory, analytic geometry, theory of differential equations, in particular partial differential equations, probability theory, mathematical statistics and random processes, numerical methods.
LO3. To formalize problems formulated in the language of a specific subject area; to formulate their mathematical formulation and choose a rational solution method; to solve the obtained problems using analytical and numerical methods, to evaluate the accuracy and reliability of the obtained results.
LO5. To be able to develop and use in practice algorithms related to the approximation of functional dependencies, numerical differentiation and integration, solving systems of algebraic, differential and integral equations, solving boundary value problems, searching for optimal solutions.
LO6. Possess basic methods for developing discrete and continuous mathematical models of objects and processes, analytical research of these models for the existence and uniqueness of their solutions.
LO16. Demonstrate skills in interacting with other people, the ability to work in a team..

Student workload

The total volume of the course is 120 hours (4 ECTS credits): lectures - 32 hours, laboratory classes - 16 hours, self-study - 72 hours.

Course prerequisites

The main prerequisite for this course is the study of the discipline "Probability Theory".

Knowledge of the basics of mathematical analysis in the field of differential and integral calculus is also important.
The course includes computer laboratory practice, so programming skills are also necessary for its successful completion (initial level of Python proficiency is welcome but not required).

Features of the course, teaching and learning methods, and technologies

The teaching and learning methods of the course include traditional lectures for theoretical foundations, as well as a computer laboratory workshop to consolidate practical skills, in which students perform statistical analysis and diagnostics tasks using the Python language in the PyCharm environment. Laboratory tasks are performed mainly as part of independent work, while laboratory classes provide consulting and defense of completed work.
As part of the course, students also complete a mandatory calculation task, which is submitted in the form of a team project. The completion of project work and its defense is a mandatory element of the academic discipline, the results of which are included in the formation of the final grade..

Program of the course

Topics of the lectures

Topic 1. Basic concepts of mathematical statistics.
Basic definitions of mathematical statistics. Formulation of descriptive statistics problems. Analytical statistics: diagnostics, forecasting, decision-making.
Topic 2. Statistical hypotheses.
General concepts regarding 0 and alternative hypotheses. Errors of the first and second types. Making statistical decisions: significance level and p-value.
Topic 3. The problem of statistical diagnostics.
The problem of making statistical decisions based on the application of the generalized Bayes formula is considered.
Topic 4. Statistical characteristics of the sample and their point estimates.
The concept of general values and sample estimates for numerical characteristics of sample distributions. Requirements for point estimates. Examples of meeting the requirements for estimates. Determining the power, unmixedness and efficiency of statistical estimates.
Topic 5. Point estimates for the mathematical expectation and variance.
Statistical estimates for the mathematical expectation and variance. Theoretical estimates for their power and unbiasedness
Topic 6. Confidence estimates and interval estimation.
Principles of constructing confidence intervals. Distribution quantiles and quartiles. One-sided and two-sided estimates.
Topic 7. Interval estimation of the mathematical expectation and variance.
Construction of confidence intervals for the mathematical expectation. Student's test. Pearson's chi-square test.
Topic 8. Comparison of samples.
Problem statement. Comparison of 2 variances (Fisher's F-test). Comparison of mathematical expectations. Comparison of several samples (definition of concepts regarding the Bartlett and Cochrane test).
Topic 9. Statistical distributions of the sample.
Distribution selection Algorithm for constructing a histogram. Principles of distribution selection. Method of moments and use of the principle of maximum likelihood.
Topic 10. Nonparametric statistical hypotheses.
Checking the correctness of the choice of distribution. Kolmogorov and Pearson consistency criteria.
Topic 11. Elements of variance analysis.
The concept of variance analysis. One-way variance analysis. Features of two- and multifactor variance analysis.
Topic 12. Linear regression analysis.
The concept of regression analysis. Theoretical foundations of the least squares method, formulated in relation to the principle of maximum likelihood.

Topic 13. Generalized linear regression analysis.
Generalization of the method of linear regression analysis for the application of different systems of orthonormal basis functions.
Topic 14. Statistical analysis of the regression model.
Interval estimates for regression parameters, confidence intervals and error estimates.
Topic 15. Elements of correlation analysis.
The concept of stochastic relationship and correlation. Point estimates of the sample correlation coefficient. Correlation matrix.
Topic 16. Elements of statistical decision making theory.
Definition of statistical risk and its use for decision making. Empirical risk minimization approach and minimax approach.

Topics of the workshops

Topic 1. Introduction to the Python language and the PyCharm environment.
Construction of approximations. Data visualization.
Topic 2. Solving the diagnostic problem using the Bayesian method.
Elementary writing of program code for the problem that was analyzed in the practical lesson is proposed (actually "manual" implementation of formulas)
Topic 3. Solving the medical diagnostic problem using the Bayesian method.
For a given dataset of data on symptoms of patients with multiple diagnoses, a diagnostic decision support system should be implemented programmatically
Topic 3a. Creating a graphical interface for a system that performs automated medical diagnostics (optional work).
For preliminary laboratory work, a user-oriented software solution should be created regarding the interaction interface and visualization of results.
Topic 4. Selection of a distribution for the sample. Nonparametric hypothesis testing.
For a given dataset (sample), it is necessary to calculate point estimates, construct a histogram, select a distribution, and check the consistency criteria.
Topic 5. Regression analysis.
For a given dataset (sample), construct a simple linear regression and obtain confidence intervals for it ("manual implementation" is assumed). Using libraries, construct polynomial, trigonometric, and multivariate regression.
Topic 6. Correlation analysis.
For a given dataset, construct and visualize a correlation matrix.
Topic 7. Statistical decision making.
For an existing dataset, determine the cutoff values for making a decision regarding a diagnostic parameter in assessing a viral disease in patients.
Defense of team projects.

Topics of the laboratory classes

In this course the laboratory classes are given as a programming (coding) assessment in Python language within a topics that provided in the workshops section of his syllabus.

Self-study

Study of lecture material. Performing laboratory work, including additional individual tasks and questions for them. Performing team project work. Requirements for the form of project work and the procedure for its implementation: 1) Project work is performed by a team of 2–4 students. 2) The work begins in the 12th week and its implementation period is 3 weeks. 3) Reporting form: public defense with a presentation of the results of the project and a demonstration of the software tools used to achieve these results. 4) Project work is certified by a commission consisting of the lecturer of the discipline and teachers who conduct practical/laboratory work. 5) Students are free to choose the software and hardware tools to perform this work. The main requirements for project work in essence: The work involves the students performing statistical data analysis. The team must select a data array from a certain subject area with a volume of about 1000 values or more. The subject area (topic) of the project can be chosen by the student team arbitrarily using open data sources (this can be data on ecology and its

changes, data on sports competitions, open data on financial and market issues, medical statistics, infrastructure, security, etc.). The above list is very conditional and is not intended to limit the choice of students, but the topic must be agreed with the lecturer before the start of the project. Based on the results of the project, the team can receive up to 20 points for the final certification. The project is evaluated as a whole, that is, the entire team, but the assessment of the answers to the questions is personal for each student member of the team.
For independent study, students are recommended to take non-formal education courses on the Coursera platform:
Inferential Statistical Analysis with Python (University of Michigan)
Fitting Statistical Models to Data with Python (University of Michigan)
https://www.coursera.org/learn/inferential-statistical-analysis-python?specialization=statistics-with-python

Course materials and recommended reading

[1] A. V. Skorokhod Probability theory and mathematical statistics K.: TViMS, 1995
[2] Senyo P. S. Probability theory and mathematical statistics: textbook. – K.: Znannia, 2007, 556p.
[3] Turchyn V.M. Probability theory and mathematical statistics. Basic concepts, examples, problems. – D.: IMApres, 2014.
[4] Dragomyretska Kh.T. et al. Probability theory and mathematical statistics. – Lviv: Lviv Polytechnic, 2012, 389 p.
[5] Shchogolev S. A. Fundamentals of probability theory and mathematical statistics: teaching and methodical manual. – Odesa: "Odessa National University named after I. I. Mechnikova», 2015. – 206 p.
[6] Methodical instructions for practical classes in the course "Probability theory, probabilistic processes and mathematical statistics" for students of the specialty "Information technologies of design" / compiled by M. V. Matyushenko, G. V. Fedchenko, I. B. Shelikhova. — Kharkiv: NTU "KhPI", 2015. — 35 p.
[7] Methodical instructions for completing individual tasks of a computer workshop on Python in the course "Mathematical statistics" for students of technical disciplines, part 1 / compiled by O. O. Larin, O. I. Sukhanova – Kharkiv: NTU "KhPI". – 2020, 48 p.
[8] Kartashov M. V. Probability, processes, statistics. K.: Kyiv University, 2008. – 494 p.

Assessment and grading

Criteria for assessment of student performance, and the final score structure

Theoretical part: 20 points
in the form of an oral exam: 2 questions in the exam paper

Practical part of the course:60 points
8 laboratory works, which, depending on the complexity, are rated from 5 to 10 points
The score is the recommendation of the teacher conducting the laboratory work.
Performing laboratory work is admission to the exam.
Team project work: 20 points
Completeness of the problem disclosure 3..5 points
Mathematical complexity 3..5
Program and algorithmic solutions 0..2
Quality of presentation 1..3
Answers to the commission's questions 0..5 (individually for each team member)
Completing the recommended informal education courses (see the Independent Work section) allows the student, after an interview, to receive up to 20 points based on the results of individual laboratory work and/or a project..

Grading scale

Total points	National	ECTS
90–100	Excellent	A
82–89	Good	B
75–81	Good	C
64–74	Satisfactory	D
60–63	Satisfactory	E
35–59	Unsatisfactory (requires additional learning)	FX
1–34	Unsatisfactory (requires repetition of the course)	F

Norms of academic integrity and course policy

The student must adhere to the Code of Ethics of Academic Relations and Integrity of NTU "KhPI": to demonstrate discipline, good manners, kindness, honesty, and responsibility. Conflict situations should be openly discussed in academic groups with a lecturer, and if it is impossible to resolve the conflict, they should be brought to the attention of the Institute's management.
Regulatory and legal documents related to the implementation of the principles of academic integrity at NTU "KhPI" are available on the website: http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/

Approval

Approved by	Date, signature	Head of the department Oleksii VODKA
	Date, signature	Head of the educational program Gennady LVOV