

**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ /
BELARUSIAN STATE UNIVERSITY**

УТВЕРЖДАЮ / APPROVED

Ректор Белорусского

государственного университета/

Rector of Belarussian State University

И.Король /Andrei D.Karol

июль 2024 г. July 15, 2024

Регистрационный № 1855/м. / Registration №1855/m.



**ПРОГРАММНЫЕ СРЕДСТВА АНАЛИЗА ДАННЫХ/
DATA ANALYSIS SOFTWARE**

Учебная программа учреждения образования по учебной дисциплине для
специальности:

The program of the educational institution of the discipline for the speciality:

7-06-0533-05 Прикладная математика и информатика /

7-06-0533-05 Applied Mathematics and Computer Science

Профилизация / Profilization:

Компьютерный анализ данных / Computer Data Analysis

2024 г.

Учебная программа составлена на основе ОСВО 7-06-0533-05-2023 и учебного плана № М53а-53-115/уч. от 11.04.2023.

СОСТАВИТЕЛЬ:

П.А. Пашук, старший преподаватель кафедры теории вероятностей и математической статистики факультета прикладной математики и информатики Белорусского государственного университета, магистр прикладной математики и информационных технологий.

РЕЦЕНЗЕНТЫ:

М.С. Абрамович, заведующий НИИ статистического анализа и моделирования учреждения Белорусского государственного университета «НИИ прикладных проблем математики и информатики», кандидат физико-математических наук, доцент;

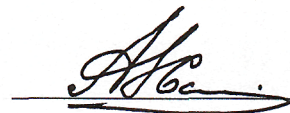
А.И. Кишкар, ведущий инженер-программист отдела информационных систем управления бизнес-приложений департамента производства ЗАО «Международный деловой альянс», магистр прикладной математики и информационных технологий.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой теории вероятностей и математической статистики БГУ
(протокол № 12 от 21.05.2024);

Научно-методическим советом БГУ
(протокол № 8 от 31.05.2024)

Заведующий кафедрой теории вероятностей
и математической статистики



А.Ю.Харин

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель учебной дисциплины «Data analysis software» – формирование необходимых навыков программной реализации методов и алгоритмов, а также их применения при анализе данных.

Задачи учебной дисциплины:

- 1) изучение базового синтаксиса и особенностей программной среды при анализе данных;
- 2) знакомство студентов с применением методов программной среды, а также их преимуществ и недостатками;
- 3) формирование практических навыков решения прикладных задач с использованием современного программного обеспечения.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием.

Учебная дисциплина относится к модулю «Специальные программные средства» государственного компонента.

Учебная программа составлена с учетом межпредметных **связей** и программ по дисциплинам: «Multivariate statistical analysis», «Visualization methods in data analysis by R», «Methods for statistical analysis of complex data», «Mathematical models and methods for computer vision».

Требования к компетенциям

Освоение учебной дисциплины «Data analysis software» должно обеспечить формирование следующих компетенций:

универсальные компетенции (UC):

UC. To perform foreign languages communications in academic, scientific and professional areas for realization of research and innovative activity;

учебно-производственные компетенции (UPC):

UPC. To apply knowledge of modern probability models for complex data analysis, to apply specific modern methods for complex data analysis.

В результате освоения учебной дисциплины студент должен:

знать:

- базовый синтаксис и основные стандартные модуля изучаемой программной среды для анализа данных;
- особенности программной среды при реализации методов и алгоритмов анализа данных;

уметь:

- реализовывать методы и алгоритмы анализа данных в программной среде;
- выбирать оптимальный способ реализации методов преобразования и анализа данных;

иметь навыки:

- работы с основными методами анализа данных в программной среде;
- компьютерной реализации основных методов;
- решения прикладных задач с использованием современного программного обеспечения;

Структура учебной дисциплины

Дисциплина изучается в 1 семестре. Всего на изучение учебной дисциплины «Data analysis software» отведено:

– для очной формы получения высшего образования – 90 часов, в том числе 40 аудиторных часов, из них: лекции – 20 часов, лабораторные занятия – 20 часов.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма промежуточной аттестации – экзамен.

EXPLANATORY NOTE

Aim and tasks of the discipline

Aim of the discipline «Data analysis software» – development of the necessary skills for software implementation of methods and algorithms, as well as their application in data analysis.

Tasks of the discipline:

- 1) learning the basic syntax and features of the software environment for data analysis;
- 2) familiarizing students with the use of software environment methods, as well as their advantages and disadvantages;
- 3) development of practical skills in solving applied problems using modern software.

Place of the academic discipline in the system of training a specialist with higher education.

The academic discipline is part of the module «Special software» of state component.

The curriculum is designed taking into account interdisciplinary connections and programs in disciplines: «Multivariate statistical analysis», «Visualization methods in data analysis by R», «Methods for statistical analysis of complex data», «Mathematical models and methods for computer vision».

Requirements for competences

Mastering of the academic discipline «Data analysis software» should provide the formation of the following universal and advanced professional competences:

universal competences (UC):

UC. To perform foreign languages communications in academic, scientific and professional areas for realization of research and innovative activity;

educational and production competences (UPC):

UPC. To apply knowledge of modern probability models for complex data analysis, to apply specific modern methods for complex data analysis.

As a result of mastering the academic discipline, the student is expected to:

know:

- basic syntax and main standard modules of the studied software environment for data analysis;
- features of the software environment when implementing methods and algorithms for data analysis;

be able to:

- implement methods and algorithms for data analysis in a software environment;
- choose the optimal way to implement data transformation and analysis methods;

have skills in:

- working with basic data analysis methods in a software environment;
- computer implementation of basic methods;
- solving applied data analysis problems using modern software;

Structure of the academic discipline

The discipline is studied in the 1 semester. In total for the study of the discipline «Data analysis software» is allocated:

– for full-time higher education – 90 hours, including 40 in-class hours, of them: lectures – 20 часов, laboratory classes – 20 hours.

The labour intensity of the discipline is 3 credit units.

Form of certification – exam.

CONTENT OF THE STUDY MATERIAL

Topic 1. Introduction to programming language.

Basic information about the programming language. Versions of the programming language. Main development environments used. General description of the development process. Setting up the working environment.

Topic 2. Basic data types.

Basic data types. Vector and matrix structures for storing data. Lists and data tables. Accessing elements of a list or data table.

Topic 3. Data processing.

Vector calculations. The principle of cyclic repetition. Basic operations. Operations on matrices. Sorting. Working with missing values.

Topic 4. Conditional selection operators and loop operators.

Conditional selection operator. Conditional multiple selection operator. For and while loop operators.

Topic 5. Functions.

Creating new user-defined functions. Scope of variables. Named arguments, default values, arbitrary number of arguments. Function as an argument to a function. A function that returns a function. Recursion.

Topic 6. Organization of data input/output.

Data editor. Access to external data. Reading data from the keyboard. Reading data from an external file, web page. Saving data to a file.

Topic 7. Basic data analysis libraries.

Calculating basic descriptive statistics. Testing hypotheses about whether a sample belongs to a given distribution. Plotting simple graphs. Plotting classic graphs (histograms, box plots, scatter plots, etc.). Fine-tuning the attributes and properties of graphs. Generating discrete and continuous random variables. Monte Carlo method. Clustering and discriminant analysis. Using machine learning algorithms.

Topic 8. Preparing reports.

Features of preparing a report in the LaTeX environment containing scripts and language inserts. Working with documents in different encodings.

TEACHING AND METHODOLOGICAL MAP OF THE DISCIPLINE
Full-time form of higher education with the use of distance learning technologies (DLT)

Title of section, topic	Title of section, topic	In-class hours					Independent work	Form of control
		Lectures	Practical classes	Seminar classes	Laboratory classes	Other		
1	Introduction to programming language	2			2			Oral test
2	Basic data types	2			2			Individual tasks
3	Data processing	2			2			Test, report on the performance of laboratory task
4	Conditional selection operators and loop operators	2			2			Control work
5	Functions	2			2			Individual tasks, report on the performance of laboratory task
6	Organization of data input/output	2			2			Test
7	Basic data analysis libraries	6			6			Individual tasks, report on the performance of laboratory task
8	Preparing reports	2			2			Report on the performance of laboratory task
TOTAL		20			20			

INFORMATION AND METHODOLOGICAL PART

List of basic literature

1. Васильев, Ю. Python для Data Science / Ю. Васильев; [пер. с англ. А. Алимova]. – Санкт-Петербург; Москва; Минск: Питер, 2023. – 270 с. – URL: <https://ibooks.ru/bookshelf/390133/reading>.
2. Брюс, П. Практическая статистика для специалистов Data Science: 50+ важнейших понятий с использованием R и Python / П. Брюс, Э. Брюс, П. Гедек; [пер. с англ. А. Логунова]. – 2-е изд., перераб. и доп. – Санкт-Петербург: БХВ-Петербург, 2021. – 346 с. – URL: <https://ibooks.ru/reading.php?short=1&productid=380029>.
3. Ланц, Б. Машинное обучение на R: экспертные техники для прогностического анализа / Б. Ланц; [пер. с англ. Е. Сандицкой]. – Санкт-Петербург [и др.]: Питер, 2020. – 462 с. – URL: <https://ibooks.ru/bookshelf/367984>.
4. Яворски, М. Python. Лучшие практики и инструменты / М. Яворски, Т. Зиаде; [пер. с англ. Е. Матвеев]. – 3-е изд. – Санкт-Петербург; Москва; Минск: Питер, 2021. – 558 с. – URL: <https://ibooks.ru/bookshelf/376831>.
5. Доусон, М. Програмируемо на Python / М. Доусон; [пер. с англ. В. Порицкий]. – Санкт-Петербург; Москва; Минск: Питер, 2023. – 414 с.

List of additional literature

1. Kabacoff, R.I. R in action. Third edition. Data analysis and graphics with R and Tidyverse / R.I. Kabacoff. – Maning Shelter Island, 2022 – 595 p.
2. Everitt B. An introduction to applied multivariate analysis with R / B. Everitt, T. Hothorn. – New York: Springer. – 2018. – 274 p
3. Мэтлофф, Н. Искусство программирования на R. Погружение в большие данные / Н. Мэтлофф. – Санкт-Петербург [и др.]: Питер, 2019. – 413 с.: ил.
4. Мастицкий, С.Э. Статистический анализ и визуализация данных с помощью R / Мастицкий С. Э., Шитиков В. К. – Москва: ДМК Пресс, 2015. – 495 с.: ил.
5. Хобсон, Л. Обработка естественного языка в действии / Л. Хобсон, Х. Ханнес, Х. Коул; [пер. с англ. И. Пальти, С. Черников]. – Санкт-Петербург [и др.]: Питер, 2020. – 575 с.: ил.
6. Кабаков, Р.И. R в действии / Р.И. Кабаков. – 3-е изд. – Москва: ДМК Пресс, 2023. – 768 с.
7. Бейдер, Д. Знакомство с Python / Д. Бейдер и др.; [пер. с англ. Е. Матвеев]. – Санкт-Петербург; Москва; Минск: Питер, 2023. – 507 с.: ил.
8. Вандер Плас, Дж. Python для сложных задач: наука о данных и машинное обучение / Дж. Вандер Плас; [пер. с англ. И. Пальти]. – Санкт-Петербург; Москва; Минск: Питер, 2023. – 573 с.: ил.

List of recommended diagnostic tools and methodology for final mark formation

The object of diagnostics of students' competences is the knowledge and skills acquired as a result of studying the academic discipline. Identification of students' learning achievements is carried out by means of current and interim certification.

The following means of current certification can be used to diagnose competences: control work; report on the performance of laboratory task, individual tasks, oral test.

The form of interim certification in the discipline "Data Analysis Software" in accordance with the curriculum is exam.

A rating system of the student knowledge is used for the final mark formation, which makes it possible to trace and evaluate the dynamics within the process of achieving learning objectives. The rating system stipulates the use of weighting coefficients for current and interim certification of students in the academic discipline.

The final mark formation in the course of control measures for current certification (approximate weighting coefficients determining the contribution of current certification to the mark for passing interim certification) includes:

- performance of control work – 25 %;
- performance of test – 25 %;
- report on the performance of laboratory task – 50 %.

The final mark for the discipline is calculated on the basis of the mark of current certification (rating system of knowledge) — 60 % and exam mark — 40 %.

Approximate list of laboratory classes

Class № 1. Introduction to programming language.

Class № 2. Basic data types.

Class № 3. Data processing.

Class № 4. Conditional selection operators and loop operators.

Class № 5. Functions.

Class № 6. Organization of data input/output.

Class № 7. Calculating descriptive statistics. Testing simple hypotheses.

Class № 8. Modeling of random variables. Monte Carlo method.

Class № 9. Clustering and discriminant analysis.

Class № 10. Preparing reports.

Description of innovative approaches and methods for teaching the discipline

When organizing the educational process, a practice-based approach is used, which entails the following:

- mastering the educational content through solving practical tasks;
- acquiring skills for effective performance in various types of professional activities;

- orientation towards idea generation, implementation of students' group projects, development of business culture;
- use of evaluation procedures, assessment methods, indicating the formation of professional competences.

Methodological recommendations for the organization of independent work

Independent work for the purpose of studying the material of the academic discipline involves working with recommended educational literature and Internet resources. Theoretical information is consolidated by completing laboratory assignments, during which one should be guided by the methodological developments posted in the electronic library of the university and on the educational portal. Additional assignments (tests, assignments for independent completion) may also be offered for self-assessment and deeper assimilation of the material received.

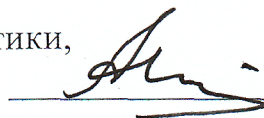
Sample list of questions for the exam

1. Basic data types and operations with them.
2. Data structures and their features.
3. Logical expressions.
4. Indexing and accessing array elements.
5. Conditional constructions.
6. Cycles.
7. Basic syntax for creating functions.
8. Creating functions with an infinite number of arguments.
9. Creating functions with default arguments.
10. Examples of creating functions that return a function.
11. Examples of creating a function that uses functions as an argument.
12. Global and local environments. Superassignment.
13. Connecting and using additional packages for data analysis.
14. Importing and exporting data.
15. Graphical representation of data.
16. Modeling random variables.
17. Testing statistical hypotheses.
18. Correlation analysis.
19. Building simple linear regression models.
20. Nonlinear regression models.
21. Analysis of residuals in regression models.
22. Creating reports.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Multivariate statistical analysis	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024
Visualization methods in data analysis by R	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024
Methods for statistical analysis of complex data	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024
Mathematical models and methods for computer vision	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024

Заведующий кафедрой теории
вероятностей и математической статистики,
доктор физ.-мат. наук, профессор



А.Ю. Харин

21 мая 2024 г.

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 202_ г.)

Заведующий кафедрой

УТВЕРЖДАЮ
Декан факультета
