

**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ /  
BELARUSIAN STATE UNIVERSITY**

**УТВЕРЖДАЮ / APPROVED**

Ректор Белорусского

государственного университета/

Rector of Belarusian State University

А.Д.Король /Andrei D.Karol

15.07.2024

Регистрационный №/ Registration № 2141/m.



**МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ/  
MULTIVARIATE STATISTICAL ANALYSIS**

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.

#### **СОСТАВИТЕЛИ:**

**А.Ю. Харин**, заведующий кафедрой теории вероятностей и математической статистики факультета прикладной математики и информатики Белорусского государственного университета, доктор физико-математических наук, профессор;

**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:**

**Б.А. Залесский**, заведующий лабораторией обработки и распознавания изображений ГНУ «Объединённый институт проблем информатики Национальной академии наук Беларуси», доктор физико-математических наук;

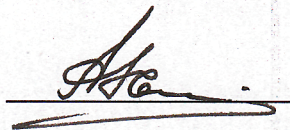
**В.И. Малюгин**, заведующий кафедрой математического моделирования и анализа данных факультета прикладной математики и информатики Белорусского государственного университета, доктор экономических наук, доцент.

#### **РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:**

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

Научно-методическим советом БГУ  
(протокол № 9 от 28.06.2024)

Заведующий кафедрой теории вероятностей  
и математической статистики



А.Ю.Харин

## ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

### Цели и задачи учебной дисциплины

**Цель** учебной дисциплины «Multivariate Statistical Analysis» – ознакомление студентов углубленной формы высшего образования (магистрантов) с основными вероятностными моделями, методами и алгоритмами статистического исследования данных, имеющих многомерную структуру.

При изложении учебной дисциплины важно уделить внимание компьютерной реализации основных методов изучаемой теории.

В рамках поставленной цели **задачи учебной дисциплины** состоят в следующем:

1. Изучение теоретических основ – математических моделей и методов статистического анализа данных многомерной структуры;

2. Формирование практических навыков решения прикладных задач анализа многомерных данных с использованием свободно доступного современного программного обеспечения в области статистического анализа.

**Место учебной дисциплины** в системе подготовки специалиста с высшим образованием (магистра).

Учебная дисциплина относится к государственному компоненту и входит в модуль «Methods and computer tools of applied mathematics».

Основой для изучения дисциплины «Multivariate Statistical Analysis» является дисциплина общего высшего образования «Probability Theory and Mathematical Statistics» (или «Higher Mathematics» с включением соответствующих разделов). Кроме того, лабораторный практикум предполагает дополнение дисциплиной «Data Analysis Software», преподаваемой параллельно. Дисциплина «Multivariate Statistical Analysis» способствует успешному освоению дисциплины «Mathematical and Computer Forecasting» (тот же модуль), а результаты ее изучения используются при прохождении практики и написании магистерских диссертаций.

### Требования к компетенциям

Освоение учебной дисциплины «Multivariate Statistical Analysis» должно обеспечить формирование следующих компетенций:

**универсальные** компетенции (UC):

UC. To solve scientific and innovative problems on the base of information and communication technologies basis;

**учебно-производственные** компетенции (UPC):

UPC. On the basis of typical models, to construct and analyze complex models that are adequate to the underlying applied problem;

UPC. To solve applied problems of multivariate data analysis by existing free software for statistical analysis.

В результате освоения учебной дисциплины студент должен:

**знать:**

– основные вероятностные модели, применяемые для статистического анализа многомерных данных;

- статистические методы анализа многомерных данных;
- правила применения методов статистического анализа многомерных данных, их свойства;

**уметь:**

- подбирать подходящую модель для решения конкретной задачи статистического анализа многомерных данных;
- исследовать потенциальную эффективность применения конкретного статистического метода для решения задачи анализа многомерных данных;

**иметь навык:**

- построения основных моделей статистического анализа многомерных данных;
- выбора и обоснования модели при решении конкретной задачи;
- компьютерной реализации основных методов решения задач.

### **Структура учебной дисциплины**

Дисциплина изучается в первом семестре. В соответствии с учебным планом всего на изучение учебной дисциплины «Multivariate Statistical Analysis» отведено для очной формы получения высшего образования – 106 часов, в том числе 50 аудиторных часов, из них: лекции – 20 часов, лабораторные занятия – 20 часов, семинарские занятия – 10 часов.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма промежуточной аттестации – экзамен.

## EXPLANATORY NOTE

### **Aim and tasks of the discipline**

**Aim** of the discipline «Multivariate statistical analysis» – familiarization of advanced students of higher education (master's students) with the basic probabilistic models, methods and algorithms of statistical analysis of data with a multidimensional structure.

It is important to pay attention to the computer implementation of the main methods of the studied theory when presenting the academic discipline.

### **Tasks of the discipline:**

1) study of theoretical foundations - mathematical models and methods of statistical analysis of data of multidimensional structure;

2) formation of practical skills in solving applied problems in the analysis of multidimensional data using freely available modern software in the field of statistical analysis.

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

The academic discipline is part of the module « Methods and computer tools of applied mathematics » of state component.

The basis for studying the discipline "Multivariate Statistical Analysis" is the discipline of general higher education "Probability Theory and Mathematical Statistics" (or "Higher Mathematics" with the inclusion of relevant sections). In addition, the laboratory practical training involves supplementing the discipline "Data Analysis Software", taught in parallel. The discipline "Multivariate Statistical Analysis" contributes to the successful mastering of the discipline "Mathematical and Computer Forecasting" (the same module), and the results of its study are used during practical training and writing master's theses.

### **Requirements for competences**

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

***universal*** competences (UC):

UC. To solve scientific and innovative problems on the base of information and communication technologies basis;

***educational and production*** competences (UPC):

UPC. On the basis of typical models, to construct and analyze complex models that are adequate to the underlying applied problem;

UPC. To solve applied problems of multivariate data analysis by existing free software for statistical analysis.

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

**know:**

- basic probabilistic models used for statistical analysis of multidimensional data;
- statistical methods of analysis of multidimensional data;

- rules for applying methods of statistical analysis of multidimensional data, their properties;

**be able to:**

- constructing basic models of statistical analysis of multidimensional data;
- investigate the potential effectiveness of using a specific statistical method for solving a problem of multidimensional data analysis;

**have skills in:**

- working with basic data analysis methods in a software environment;
- selecting and justifying a model when solving a specific problem;
- computer implementation of basic methods for solving problems;

**Structure of the academic discipline**

The discipline is studied in the 1 semester. In total for the study of the discipline «Multivariate statistical analysis» is allocated for full-time higher education – 106 hours, including 50 in-class hours, of them: lectures – 20 часов, laboratory classes – 20 hours, seminar classes – 10 hours.

The labour intensity of the discipline is 3 credit units.

Form of certification – exam.

## **CONTENT OF THE STUDY MATERIAL**

### **Section 1. Basic concepts and definitions in multivariate statistical analysis**

#### **Topic 1.1. Multivariate analysis specific features.**

Introduction. Problem examples to show principal differences from the univariate data analysis.

#### **Topic 1.2. Main terms.**

Main terms and definitions, characteristics explained.

### **Section 2. Multivariate normal probability distribution and its properties**

#### **Topic 2.1. Definitions and fundamental properties.**

Multivariate normal probability distribution, its characteristics. Conditional probability distributions. Linear transforms of Gaussian random vectors.

#### **Topic 2.2. Regression function.**

Regression function, its optimal properties. Partial and multiple correlation coefficients.

#### **Topic 2.3. Statistical estimation of parameters.**

Multivariate normal probability distribution parameters estimation. Wishart probability distribution.

### **Section 3. Principal components analysis**

#### **Topic 3.1. Principal components construction.**

Principal components method. Construction, properties, geometric setting.

#### **Topic 3.2. Sample principal components.**

Principal components for z-scores. Sample principal components. Confidence interval derivation.

#### **Topic 3.3. Approximation.**

Results graphing. Approximation using principal components. Relation to the orthogonal regression.

### **Section 4. Factor analysis**

#### **Topic 4.1. Orthogonal model.**

Factor analysis. Orthogonal model. Covariance analysis. Multiple solution for factor loadings.

#### **Topic 4.2. Factor deriving methods.**

Principal component method use for the factor analysis. Maximum likelihood method use for factor analysis.

#### **Topic 4.3. Factor rotation.**

Factor analysis – rotation of factors.

### **Section 5. Discriminant and cluster analysis**

#### **Topic 5.1. Decision rules and their construction.**

Discriminant analysis. Expected misclassification costs. Decision rule construction. Total misclassification probability minimization. Observations classification for the normal probability distribution (common covariance matrix case). Fisher discriminant function.

**Topic 5.2. Cluster analysis methods.**

Cluster analysis. Difference matrix.  $K$  means clusterization. Partition around medoids (PAM-algorithm).

**Topic 5.3. Graphical interpretation.**

Cluster analysis results graphing.

**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	<b>Basic concepts and definitions in multivariate statistical analysis</b>	4		2	4			
1.1	Multivariate analysis specific features	2			2			Random interview
1.2	Main terms	2		2	2			Seminar talk with a discussion
2	<b>Multivariate normal probability distribution and its properties</b>	4		2	4			
2.1	Definitions and fundamental properties	1			1			Random interview
2.2	Regression function	1			1			Written report on the in-class labs
2.3	Statistical estimation of parameters	2		2	2			Test No 1, seminar talk with a discussion
3	<b>Principal components analysis</b>	4		2	4			
3.1	Principal components construction	2		2	2			Seminar talk with a discussion
3.2	Sample principal components	1			1			Random interview
3.3	Approximation	1			1			Test No 2
4.	<b>Factor analysis</b>	4		2	4			
4.1	Orthogonal model	2		2	2			Seminar talk with a discussion
4.2	Factor deriving methods	1			1			Test No 3
4.3	Factor rotation	1			1			Team play for creative solution

								of open-type tasks
<b>5</b>	<b>Discriminant and cluster analysis</b>	<b>4</b>		<b>2</b>	<b>4</b>			
5.1	Decision rules and their construction	1			1			Random interview
5.2	Cluster analysis methods	2		2	2			Seminar talk with a discussion
5.3	Graphical interpretation	1			1			Test No 4
<b>TOTAL</b>		<b>20</b>		<b>10</b>	<b>20</b>			

## **INFORMATION AND METHODOLOGICAL PART**

### **List of basic literature**

1. Сошникова, Л.А. Многомерный статистический анализ. Практикум: учеб. пособие для студ. учреждений высш. образования по спец. "Статистика", "Бухгалтерский учет, анализ и аудит", "Финансы и кредит", "Национальная экономика", "Государственное управление и экономика" / Л. А. Сошникова, Е.Е. Шарилова. - Минск: БГЭУ, 2024. - 230 с.
2. Буре, В.М. Методы прикладной статистики в R и Excel: учебное пособие для студ. вузов, обуч. по направлениям "Прикладная математика и информатика", "Фундаментальная информатика и информационные технологии", а также другим математическим и естественно-научным направлениям и спец. в области техники и технологий / В. М. Буре, Е. М. Парилина, А. А. Седаков. - Изд. 5-е, стер. - Санкт-Петербург; Москва; Краснодар: Лань, 2024. - 148 с. - URL: <https://e.lanbook.com/book/319424>.
3. Концевая, Н.А. Многомерный статистический анализ в экономических задачах. Компьютерное моделирование в SPSS: учебное пособие для студентов высших учебных заведений, обучающихся по экономическим специальностям / Н. А. Концевая, И.В. Орлова. - Москва: Вузовский учебник: ИНФРА-М, 2024. - 309 с. - URL: <https://znanium.com/catalog/product/1850713>.

### **List of additional literature**

1. Johnson R.A. Applied multivariate statistical analysis / R.A. Johnson, D.W. Wichern. – New York: Pearson. – 2017. – 800 p.
2. Everitt B. An introduction to applied multivariate analysis with R / B. Everitt, T. Hothorn. – New York: Springer. – 2018. – 274 p.
3. Hastie T. The elements of statistical learning / T. Hastie, R. Tibshirani, J. Friedman. – New York : Springer. – 2008. – 758 p.
4. Харин, А. Статистическое моделирование / А. Харин, К. Кроукс, Ш. ван Аэльст, П. Фильцмозер, М. Хьюберт – Минск: БГУ, 2017 – 151 с.
5. Харин Ю.С. Математические и компьютерные основы статистического моделирования и анализа данных / Ю.С. Харин, В.И. Малюгин, М.С. Абрамович. – Минск : БГУ. – 2008. – 455 с.
6. Харин Ю.С. Эконометрическое моделирование / Ю.С. Харин, В.И. Малюгин, А.Ю. Харин. – Минск : БГУ. – 2004. – 313 с.

### **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: test; report on the performance of laboratory task, individual tasks, random interview.

The form of interim certification in the discipline “Multivariate statistical analysis” 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 test control work – 60 %;
- reports at the seminar – 25 %;
- random interview – 15 %.

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**

Seminar No 1. Regression analysis.

Seminar No 2. Principal component analysis.

Seminar No 3. Factor analysis.

Seminar No 4. Discriminant analysis.

Seminar No 5. Cluster analysis.

### **Approximate list of laboratory classes**

Class No 1. Multivariate normal probability distribution.

Class No 2. Properties of the multivariate normal probability distribution.

Class No 3. Regression analysis.

Class No 4. Principal components method.

Class No 5. Factor analysis. Orthogonal model.

Class No 6. Defining number of factors in the model.

Class No 7. Factor rotation.

Class No 8. Discriminant analysis.

Class No 9. Cluster analysis.

Class No 10. Graphing of cluster analysis results.

### **Approximate list of test topics**

Test No 1. Multivariate normal probability distribution.

Test No 2. Principal component analysis.

Test No 3. Factor analysis.

Test No 4. Discriminant and cluster analysis.

### **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.

When organizing the educational process, the group learning method is used, which is a form of organizing the educational and cognitive activities of students, which involves the functioning of different types of small groups working on both general and specific educational tasks.

### **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**

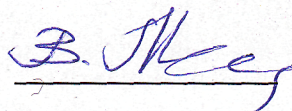
1. Sample characteristics (typical statistics), their interpretation.
2. Multivariate normal probability distribution, its characteristics – functional and numerical.
3. Linear transformations of Gaussian (normal) random vectors. Marginal probability distributions.
4. Regression function. Conditional (partial) covariance matrix; partial variance; partial correlation coefficient. Optimal properties of the regression function.

5. Multiple correlation coefficient and its properties.
  6. Statistical estimation of parameters for multivariate normal probability distribution.
  7. Principal components analysis. Construction, properties.
  8. Principal components analysis. Geometric interpretation.
  9. Principal components for z-scores.
  10. Sample principal components.
  11. Confidence intervals construction in the principal components analysis.
  12. Graphical representation of results in the principal components analysis.
- Approximation using principal components.
13. Relation of the principal components analysis to the orthogonal regression.
  14. Factor analysis. Orthogonal model.
  15. Covariance structure analysis for the factor model. Variability of factor loadings.
  16. Use of principal components analysis in the factor analysis.
  17. Use of maximal likelihood method in the factor analysis.
  18. Factor analysis – factors rotation.
  19. Discriminant analysis. Expected misclassification cost. Decision rule construction.
  20. Discriminant analysis – total misclassification probability minimization.
- Observations classification for the normal distribution (case of common covariance matrix).
21. Fisher's discriminant function.
  22. Cluster analysis. Matrix of differences.  $K$  means clusterization.
  23. Partitioning around medoids (PAM-algorithm).
  24. Graphing of cluster analysis results.
  25. Hierarchical clusterization. Agglomerative nesting, partitioning analysis.
  26. Graphing the hierarchical clusterization results – dendrogram, banner plot.

## 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)
Mathematical and computer forecasting	Математического моделирования и анализа данных	Изменений не требуется	Протокол № 12 от 21.05.2024
Data analysis software	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
Mathematical models and methods for computer vision	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024

Заведующий кафедрой математического моделирования и анализа данных,  
доктор экон. наук, доцент

 В.И. Малюгин

21 мая 2024 г.

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

\_\_\_\_\_