

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Ректор Белорусского
государственного университета

А.Д.Король

15 июля 2024 г.

Регистрационный № 2135/м.



МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ

Учебная программа учреждения образования
по учебной дисциплине для специальности:

7-06-0533-05 Прикладная математика и информатика

Профилизации:

Компьютерный анализ данных

Интеллектуальные системы

2024 г.

Учебная программа составлена на основе ОСВО 7-06-0533-05-2023 и учебных планов № М53-5.3-43/уч. от 15.02.2023, № М53-5.3-79/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;

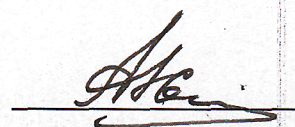
В.И. Малюгин, заведующий кафедрой математического моделирования и анализа данных факультета прикладной математики и информатики Белорусского государственного университета, доктор экономических наук, доцент.

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 «Многомерный статистический анализ» – 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 углубленным высшим образованием (магистра).

Учебная дисциплина относится к государственному компоненту и входит в модуль «Методы и компьютерные средства прикладной математики».

Основой для изучения дисциплины «Многомерный статистический анализ» является дисциплина общего высшего образования «Теория вероятностей и математическая статистика» (или «Высшая математика» с включением соответствующих разделов). Кроме того, лабораторный практикум предполагает дополнение дисциплиной «Программные средства анализа данных», преподаваемой параллельно. Дисциплина «Многомерный статистический анализ» способствует успешному освоению дисциплины «Математическое и компьютерное прогнозирование» (тот же модуль), а результаты ее изучения используются при прохождении практики и написании магистерских диссертаций.

Требования к компетенциям

Освоение учебной дисциплины «Многомерный статистический анализ» должно обеспечить формирование следующих компетенций:

универсальные компетенции (УК):

УК. Решать научно-исследовательские и инновационные задачи на основе применения информационно-коммуникационных технологий;

учебно-производственные компетенции (УПК):

УПК. На основе типовых моделей строить и исследовать сложные модели, адекватные решаемой прикладной задаче;

УПК. Решать прикладные задачи анализа многомерных данных с использованием свободно доступного современного программного обеспечения в области статистического анализа.

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 1 семестре. В соответствии с учебным планом всего на изучение учебной дисциплины «Многомерный статистический анализ» отведено для очной формы получения углубленного 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.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Основные понятия и определения в многомерном статистическом анализе

Тема 1.1. Особенности многомерного анализа.

Введение. Примеры задач, демонстрирующие принципиальные отличия от одномерного анализа данных.

Тема 1.2. Основные понятия.

Основные понятия и определения, содержательный смысл характеристик.

Раздел 2. Многомерное нормальное распределение и его свойства

Тема 2.1. Определения и основные свойства.

Определение многомерного нормального распределения вероятностей, его характеристики. Условные распределения вероятностей. Линейные преобразования гауссовских случайных векторов.

Тема 2.2. Функция регрессии.

Функция регрессии, ее оптимальные свойства. Частный и множественный коэффициенты корреляции.

Тема 2.3. Статистическая оценка параметров.

Оценка параметров многомерного нормального распределения. Распределение вероятностей Уишарта.

Раздел 3. Метод главных компонент

Тема 3.1. Построение главных компонент.

Метод главных компонент. Построение, свойства, геометрическая интерпретация.

Тема 3.2. Выборочные главные компоненты.

Главные компоненты для стандартизированных переменных. Выборочные главные компоненты. Построение доверительных интервалов.

Тема 3.3. Аппроксимация.

Графическое представление результатов. Аппроксимация с использованием главных компонент. Связь с ортогональной регрессией.

Раздел 4. Факторный анализ

Тема 4.1. Ортогональная модель.

Факторный анализ. Ортогональная модель. Анализ ковариационной структуры факторной модели. Неоднозначность факторных нагрузок.

Тема 4.2. Методы построения факторов.

Использование метода главных компонент в факторном анализе. Использование метода максимального правдоподобия в факторном анализе.

Тема 4.3. Вращение факторов.

Факторный анализ – вращение факторов.

Раздел 5. Дискриминантный и кластерный анализ

Тема 5.1. Решающие правила и их построение.

Дискриминантный анализ. Ожидаемая стоимость ошибки. Построение решающего правила. Минимизация полной вероятности ошибки. Классификация наблюдений из нормального распределения вероятностей (случай общей ковариационной матрицы). Дискриминантная функция Фишера.

Тема 5.2. Методы кластерного анализа.

Кластерный анализ. Матрица различий. Кластеризация методом К-средних. Разбиение по медоидам (РАМ-алгоритм). Иерархическая кластеризация.

Тема 5.3. Графическая интерпретация.

Графическое отображение результатов кластерного анализа.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Очная форма получения высшего образования с применением дистанционных образовательных технологий (ДОТ)

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УСР	Формы контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	Основные понятия и определения в многомерном статистическом анализе	4		2	4			
1.1	Особенности многомерного анализа	2			2			Устное собеседование
1.2	Основные понятия	2		2	2			Доклад на семинаре
2	Многомерное нормальное распределение и его свойства	4		2	4			
2.1	Определения и основные свойства	1			1			Устное собеседование
2.2	Функция регрессии	1			1			Отчет по лабораторной работе
2.3	Статистическая оценка параметров	2		2	2			Контрольная работа
3	Метод главных компонент	4		2	4			
3.1	Построение главных компонент	2		2	2			Доклад на семинаре
3.2	Выборочные главные компоненты	1			1			Отчет по лабораторной работе
3.3	Аппроксимация	1			1			Контрольная работа
4	Факторный анализ	4		2	4			
4.1	Ортогональная модель	2		2	2			Доклад на семинаре

4.2	Методы построения факторов	1			1			Контрольная работа
4.3	Вращение факторов	1			1			Отчет по лабораторной работе, командная игра для творческого решения заданий открытого типа
5	Дискриминантный и кластерный анализ	4		2	4			
5.1	Решающие правила и их построение	1			1			Устное собеседование
5.2	Методы кластерного анализа	2		2	2			Контрольная работа, доклад на семинаре
5.3	Графическая интерпретация	1			1			Отчет по лабораторной работе
ИТОГО		20		10	20			

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Основная литература

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>.

Дополнительная литература

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.

Перечень рекомендуемых средств диагностики и методика формирования итоговой отметки

Для диагностики компетенций используется следующий инструментарий:

1. *Устная форма*: собеседования, коллоквиум, командная игра для творческого решения заданий открытого типа;

2. *Устно-письменная форма*: отчеты по лабораторным работам с их устной защитой, доклады на семинарах;

3. *Письменная форма*: контрольные работы;

4. *Командная игра для творческого решения задач открытого типа*.

Формой промежуточной аттестации по дисциплине «Многомерный статистический анализ» учебным планом предусмотрен экзамен.

Для формирования итоговой отметки по учебной дисциплине используется модульно-рейтинговая система оценки знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая система предусматривает использование весовых коэффициентов для текущей и промежуточной аттестации студентов по учебной дисциплине.

Формирование итоговой отметки в ходе проведения контрольных мероприятий текущей аттестации (примерные весовые коэффициенты, определяющие вклад текущей аттестации в отметку при прохождении промежуточной аттестации):

- контрольные работы – 4 шт. по 15% каждая, в сумме 60%;
- доклады на семинаре – 25%;
- результаты устных собеседований – 15%.

Итоговая отметка по дисциплине рассчитывается на основе итоговой отметки текущей аттестации (рейтинговой системы оценки знаний) 40 % и экзаменационной отметки 60 %.

Примерная тематика лабораторных занятий

Занятие № 1. Особенности многомерного анализа.

Занятие № 2. Основные понятия в многомерном анализе.

Занятие № 3. Определения и основные свойства многомерного нормального распределения.

Занятие № 4. Статистическая оценка параметров многомерного нормального распределения.

Занятие № 5. Построение главных компонент.

Занятие № 6. Выборочные главные компоненты. Аппроксимация.

Занятие № 7. Ортогональная модель.

Занятие № 8. Методы построения факторов. Вращение факторов.

Занятие № 9. Решающие правила и их построение.

Занятие № 10. Методы кластерного анализа.

Примерная тематика семинарских занятий

Занятие № 1. Основные понятия в многомерном анализе.

Занятие № 2. Статистическая оценка параметров многомерного нормального распределения.

Занятие № 3. Построение главных компонент.

Занятие № 4. Ортогональная модель.

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используется **практико-ориентированный подход**, который предполагает:

- освоение содержания образования через решения практических задач;
- приобретение навыков эффективного выполнения разных видов профессиональной деятельности;
- ориентацию на генерирование идей, реализацию групповых студенческих проектов, развитие предпринимательской культуры;
- использованию процедур, способов оценивания, фиксирующих сформированность профессиональных компетенций.

При организации образовательного процесса используется **метод группового обучения**, который представляет собой форму организации учебно-познавательной деятельности обучающихся, предполагающую функционирование разных типов малых групп, работающих как над общими, так и специфическими учебными заданиями.

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы студентов по учебной дисциплине следует использовать современные информационные ресурсы: разместить на образовательном портале комплекс учебных и учебно-методических материалов (учебно-программные материалы, учебное издание для теоретического изучения дисциплины, методические указания к лабораторным занятиям, материалы текущего контроля и текущей аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего образования и учебно-программной документации, в т.ч. вопросы для подготовки к экзамену, задания, тесты, вопросы для самоконтроля, тематика докладов и др., список рекомендуемой литературы, информационных ресурсов и др.).

При изучении учебной дисциплины рекомендуется использовать следующие формы самостоятельной работы:

- поиск (подбор) и обзор литературы и электронных источников по индивидуально заданной тематике при подготовке к докладу;
- выполнение домашнего задания;
- решение задач и выполнение упражнений, выдаваемых на лабораторных занятиях;
- изучение материала, предназначенного для самостоятельного ознакомления;
- подготовка к экзамену.

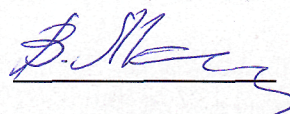
Примерный перечень вопросов к экзамену

1. Выборочные характеристики многомерных статистических данных и их интерпретация.
2. Многомерное нормальное распределение вероятностей, его функциональные и числовые характеристики.
3. Линейные преобразования гауссовских (нормальных) случайных векторов. Маргинальные распределения вероятностей.
4. Функция регрессии. Условная (частная) ковариационная матрица, частная дисперсия, частный коэффициент корреляции. Оптимальные свойства функции регрессии.
5. Множественный коэффициент корреляции и его свойства.
6. Статистическое оценивание параметров для многомерного нормального распределения вероятностей.
7. Метод главных компонент. Построение, свойства.
8. Метод главных компонент. Геометрическая интерпретация.
9. Главные компоненты для стандартизованных переменных.
10. Выборочные главные компоненты.
11. Построение доверительных интервалов в методе главных компонент.
12. Графическое представление результатов в методе главных компонент. Аппроксимация с использованием главных компонент.
13. Связь метода главных компонент с ортогональной регрессией.
14. Факторный анализ. Ортогональная модель.
15. Анализ ковариационной структуры факторной модели. Неоднозначность факторных нагрузок.
16. Использование метода главных компонент в факторном анализе.
17. Использование метода максимального правдоподобия в факторном анализе.
18. Факторный анализ – вращение факторов.
19. Дискриминантный анализ. Ожидаемая стоимость ошибочной классификации. Построение решающего правила.
20. Дискриминантный анализ – минимизация полной вероятности ошибки. Классификация наблюдений для нормального распределения (случай общей ковариационной матрицы).
21. Дискриминантная функция Фишера.
22. Кластерный анализ. Матрица различий. Кластеризация методом K-средних.
23. Разбиение вокруг медоидов (РАМ-алгоритм).
24. Графическое отображение результатов кластерного анализа.
25. Иерархическая кластеризация. Агломерационное гнездование, анализ разбиения.
26. Графическое отображение результатов иерархической кластеризации – дендрограмма, баннерная диаграмма.

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)
Математическое и компьютерное прогнозирование	Математического моделирования и анализа данных	Изменений не требуется	Протокол № 12 от 21.05.2024
Методы визуализации в анализе данных средствами языка R	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024
Программные средства анализа данных	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024
Математические модели и методы компьютерного зрения	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024

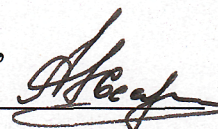
Заведующий кафедрой математического моделирования и анализа данных,
доктор экон. наук, доцент



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
