

Белорусский государственный университет

УТВЕРЖДАЮ

Проректор по учебной работе и
образовательным инновациям

О.И.Чуприс



2019 г.

Регистрационный № УД-7206 /уч.

МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности:**

1-31 80 09 Прикладная математика и информатика

Профилизации:

Алгоритмы и системы обработки больших данных

Аналитическая логистика

Компьютерный анализ данных

Математическая кибернетика

2019 г.

Учебная программа составлена на ОСВО 1-31 80 09-2019 и учебного плана G31-072/уч., G31-073/уч. G31-075/уч., G31-074/уч. от 11.04.2019 г.

СОСТАВИТЕЛЬ:

А.Ю. Харин, заведующий кафедрой теории вероятностей и математической статистики Белорусского государственного университета, доктор физико-математических наук, доцент

РЕЦЕНЗЕНТЫ:

Б.А. Залесский, заведующий лабораторией обработки и распознавания изображений Объединенного института проблем информатики Национальной академии наук Беларуси, доктор физико-математических наук;

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

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой теории вероятностей и математической статистики Белорусского государственного университета

(протокол № 15 от 30 апреля 2019 г.);

Научно-методическим Советом Белорусского государственного университета (протокол № 5 от 28 июня 2019 г.)

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 II ступени 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 «Многомерный статистический анализ» относится к государственному компоненту и входит в модуль «Методы и программные средства прикладной математики».

Основой для изучения дисциплины «Многомерный статистический анализ» является дисциплина первой ступени «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», а результаты ее изучения используется при прохождении практики и написании магистерских диссертаций.

Требования к компетенциям

Освоение учебной дисциплины «Многомерный статистический анализ» должно обеспечить формирование следующих универсальных и учебно-производственных компетенций.

универсальные компетенции:

УК-2. Уметь формулировать решение на основе анализа сложных причинно-следственных связей.

учебно-производственные компетенции:

УПК-1. Уметь применять системный и сравнительный анализ для построения математических моделей повышенной сложности.

УПК-3. Обладать навыками решения прикладных задач анализа многомерных данных с использованием свободного доступного современного программного обеспечения в области статистического анализа.

В результате освоения учебной дисциплины студент магистратуры 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;

владеть:

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

Структура учебной дисциплины

Дисциплина изучается в первом семестре. Всего на изучение учебной дисциплины «Многомерный статистический анализ» 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 текущей аттестации по учебной дисциплине – экзамен.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Основные понятия и определения многомерного статистического анализа

Тема 1.1. Особенности многомерного анализа.

Введение. Примеры задач, показывающие принципиальные отличия от задач анализа скалярных 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 простейшие свойства.

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

Тема 2.2. Функция регрессии.

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

Тема 2.3. Статистическое оценивание параметров.

Статистическое оценивание параметров многомерного нормального распределения. Распределение 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 стандартизованных 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 кластер-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 (случай общей ковариационной матрицы). Дискриминантная функция Фишера.

Тема 5.2. Методы кластер-анализа.

Кластерный анализ. Матрица различий. Кластеризация методом K средних. Разделение вокруг медоидов (РАМ-алгоритм).

Тема 5.3. Графическая интерпретация.

Графическое представление результатов кластер-анализа.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Дневная форма получения образования

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УСР	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	2	3	4	5	6	7	8	9
1	Основные понятия и определения многомерного статистического анализа	4		2	4			
1.1	Особенности многомерного анализа	2			2			Собеседование
1.2	Основные понятия	2		2	2			Собеседование
2	Многомерное нормальное распределение и его свойства	4		2	4			
2.1	Определения и простейшие свойства	1			1			Собеседование
2.2	Функция регрессии	1			1			Собеседование
2.3	Статистическое оценивание параметров	2		2	2			Отчет по лаб. работе №1
3	Метод главных компонент	4		2	4			
3.1	Построение главных компонент	2		2	2			Собеседование
3.2	Выборочные главные компоненты	1			1			Собеседование
3.3	Аппроксимация	1			1			Отчет по лаб. работе №2
4.	Факторный анализ	4		2	4			
4.1	Ортогональная модель	2		2	2			Собеседование
4.2	Методы построения факторов	1			1			Отчет по лаб. работе №3
4.3	Вращение факторов	1			1			Командная игра для творческого решения заданий открытого типа
5	Дискриминантный и кластер-анализ	4		2	4			
5.1	Решающие правила и их построение	1			1			Собеседование
5.2	Методы кластер-анализа	2		2	2			Коллоквиум
5.3	Графическая интерпретация	1			1			Отчет по лаб. работе №4
	Всего: 50 часов	20		10	20			

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Johnson R.A. Applied multivariate statistical analysis / R.A. Johnson, D.W. Wichern. – New York : Pearson. – 2017. – 800 p.
2. Харин Ю.С. Теория вероятностей, математическая и прикладная статистика / Ю.С. Харин, Н.М. Зуев, Е.Е. Жук. – Минск : БГУ. – 2011. – 463 с.
3. 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.

Перечень дополнительной литературы

1. Харин Ю.С. Эконометрическое моделирование / Ю.С. Харин, В.И. Малюгин, А.Ю. Харин. – Минск : БГУ. – 2004. – 313 с.
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. Лобач В.И. Имитационное и статистическое моделирование / В.И. Лобач, В.П. Кирлица, В.И. Малюгин, С.Н. Сталевская. – Минск : БГУ. – 2004. – 189 с.

Перечень рекомендуемых средств диагностики и методика формирования итоговой оценки

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.

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

При формировании итоговой оценки используется рейтинговая оценка знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний и текущей аттестации студентов по дисциплине.

Примерные весовые коэффициенты, определяющие вклад текущего контроля знаний и текущей аттестации в рейтинговую оценку:

- работа на лабораторных занятиях и семинарах – 50 %;
- коллоквиум – 50 %.

Рейтинговая оценка по дисциплине рассчитывается на основе оценки текущей успеваемости и экзаменационной оценки с учетом их весовых коэффициентов. Оценка по текущей успеваемости составляет 30 %, экзаменационная оценка – 70 %.

Примерная тематика лабораторных работ

Лабораторная работа № 1. Применение статистического оценивания параметров многомерного нормального распределения для анализа реальных наборов данных.

Лабораторная работа № 2. Компьютерная реализация метода главных компонент при решении задач анализа реальных наборов данных.

Лабораторная работа № 3. Компьютерная реализация методов построения факторов и проведение вычислительных экспериментов в задачах факторного анализа, содержательная интерпретация полученных результатов.

Лабораторная работа № 4. Компьютерная реализация методов дискриминантного и кластерного анализа, решение задач классификации реальных наборов данных.

Примерная тематика семинаров

Семинар № 1. Основные понятия многомерного статистического анализа, свойства типовых статистик.

Семинар № 2. Многомерное нормальное распределение и его свойства.

Семинар № 3. Применение метода главных компонент.

Семинар № 4. Применение факторного анализа.

Семинар № 5. Применение дискриминантного и кластер-анализа.

Описание инновационных подходов и методов к преподаванию учебной дисциплины (эвристический, проективный, практико-ориентированный)

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.).

Примерный перечень вопросов к экзамену

1. Выборочные характеристики многомерных статистических данных, их интерпретация.
2. Многомерное нормальное распределение, 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 многомерного нормального распределения.
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 ошибки. Построение решающего правила.
20. Дискриминантный анализ – минимизация полной вероятности ошибки. Классификация наблюдений из нормального распределения (случай общей ковариационной матрицы).
21. Дискриминантная функция Фишера.
22. Кластерный анализ. Матрица различий. Кластеризация методом K средних.
23. Разделение вокруг медоидов (РАМ-алгоритм).
24. Графическое представление результатов кластер-анализа.
25. Иерархическая кластеризация. Агломерационное гнездование, отделяющий анализ.
26. Графическое представление результатов иерархической кластеризации – дендрограмма, график баннеров.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

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) ¹
Методы статистического анализа сложных данных	Теории вероятностей и математической статистики	Нет	Оставить содержание учебной дисциплины без изменения, протокол № 15 от 30 апреля 2019 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 _____ 201_ г.)
(название кафедры)

Заведующий кафедрой

(ученая степень, ученое звание) (подпись) (И.О.Фамилия)

УТВЕРЖДАЮ
Декан факультета

(ученая степень, ученое звание) (подпись) (И.О.Фамилия)