

Министерство науки и высшего образования Российской Федерации

федеральное государственное бюджетное образовательное учреждение
высшего образования
РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ
УНИВЕРСИТЕТ

Кафедра Прикладной информатики

Рабочая программа дисциплины

Многомерный анализ данных

Основная профессиональная образовательная программа
высшего образования по направлению подготовки

38.03.05 «Бизнес-информатика»

Направленность (профиль):
Бизнес-аналитика

Уровень:
Бакалавриат

Форма обучения
Очная/заочная/очно-заочная

Согласовано
Руководитель ОПОП

Степанов С.Ю. Степанов С.Ю.

Председатель УМС
Палкин И.И. Палкин И.И.

Рекомендована решением
Учебно-методического совета РГГМУ
24 июня 2021 г., протокол № 9

Рассмотрена и утверждена на заседании кафедры
11 мая 2021 г., протокол № 6

Зав. кафедрой Истомин Е.П. Истомин Е.П.

Авторы-разработчики:

Истомин Е.П.
Степанов С.Ю.

Санкт-Петербург 2021

**Рассмотрено и рекомендовано к использованию в учебном процессе
на 2022/2023 учебный год без изменений***

**Протокол № 2 заседания кафедры Прикладной информатики от
17.03.2022г.**

1. Цель и задачи освоения дисциплины

Цель освоения дисциплины – формирование и расширение теоретических знаний и практических навыков применения методов и алгоритмов многомерного анализа данных для исследования социальных и экономических процессов и явлений

Задачи:

- освоение методов исследования взаимосвязей между показателями, характеризующими объекты в социально-экономических исследованиях
- освоение методов распознавания образов и типологизации объектов
- освоение современных пакетов прикладных программ, реализующих алгоритмы многомерного анализа данных
- приобретение навыков содержательной интерпретации результатов исследования

2. Место дисциплины в структуре основной профессиональной образовательной программы

Дисциплина «Многомерный анализ данных» относится к обязательной части блока «Дисциплины».

В соответствии с учебным планом, обучение проводится в 7,8 семестре для очной/очно-заочной формы обучения, на 5 году обучения для заочной формы обучения. Освоения дисциплины проводится на основе знаний, полученных во время изучения дисциплины “Линейная алгебра”, “Математический анализ” и “Анализ данных”.

3. Перечень планируемых результатов обучения

Процесс изучения дисциплины направлен на формирование компетенций:

ОПК-4.1

Общепрофессиональные компетенции

Код и наименование общепрофессиональной компетенции	Код и наименование индикатора достижения общепрофессиональной компетенции	Результаты обучения
ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.	<i>Знать: основные понятия и инструменты анализа данных, теории вероятностей</i> <i>Уметь: решать типовые математические задачи, используемые при принятии управленческих решений</i> <i>Владеть: программами Microsoft Office для работы с деловой информацией и основами Web-технологий, специализированными компьютерными программами</i>

4. Структура и содержание дисциплины

4.1. Объем дисциплины

Объем дисциплины составляет 7 зачетных единиц, 252 академических часа.

Таблица 4.

Объем дисциплины по видам учебных занятий в академических часах

Объем дисциплины	Всего часов		
	Очная форма обучения	Очно-заочная форма обучения	Заочная форма обучения
Объем дисциплины	252	252	252
Контактная работа обучающихся с преподавателем (по видам аудиторных учебных занятий) – всего:	98	64	28
в том числе:	-	-	-
лекции	32	26	10
занятия семинарского типа:			
практические занятия			
лабораторные занятия	56	38	18
Самостоятельная работа (далее – СРС) – всего:	154	188	224
в том числе:	-	-	-
курсовая работа			
контрольная работа			
Вид промежуточной аттестации	зачет, экзамен	зачет, экзамен	экзамен

4.2. Структура дисциплины

Таблица 5.

Структура дисциплины для очной формы обучения

№	Раздел / тема дисциплины	С е м е с т р	Виды учебной работы, в т.ч. самостоятельная работа студентов, час.			Формы текущего контроля успеваемости	Формируемые компетенции	Индикаторы достижения компетенций
			Ле кц ии	Пра ктич ески е заня тия	СР С			
1	Множественная линейная регрессия	7	14	28	66	Тест Лабораторная работа	ОПК-4. Способен использовать информацию, методы и программные	ОПК-4.1. Использует математические модели и методы

							средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	обработки данных.
2	Кластерный анализ	8	14	14	44	Контрольная работа Лабораторная работа	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
3	Факторный анализ	8	14	14	44	Контрольная работа Лабораторная работа	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
	ИТОГО	-	32	56	154	-	-	-

Таблица 6.

Структура дисциплины для очно-заочной формы обучения

№	Раздел / тема дисциплины	С е м е с т р	Виды учебной работы, в т.ч. самостоятельная работа студентов, час.			Формы текущего контроля успеваемости	Формируемые компетенции	Индикаторы достижения компетенций
			Ле кц ии	Пра ктич ески е заня тия	СР С			
1	Множественная линейная регрессия	7	8	20	80	Тест Лабораторная работа	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
2	Кластерный анализ	8	8	8	54	Контрольная работа Лабораторная работа	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
3	Факторный анализ	8	10	10	54	Контрольная работа Лабораторная работа	ОПК-4. Способен использовать информацию, методы и	ОПК-4.1. Использует математические модели и методы

							программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	обработки данных.
	ИТОГО	-	26	38	188	-	-	-

Таблица 7.

Структура дисциплины для заочной формы обучения

№	Раздел / тема дисциплины	Год	Виды учебной работы, в т.ч. самостоятельная работа студентов, час.			Формы текущего контроля успеваемости	Формируемые компетенции	Индикаторы достижения компетенций
			Лекции	Практические занятия	СРС			
1	Множественная линейная регрессия	5	2	6	74	Тест Лабораторная работа	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
2	Кластерный анализ	5	2	6	74	Контрольная работа Лабораторная работа	ОПК-4. Способен использовать информацию, методы и программные средства ее	ОПК-4.1. Использует математические модели и методы обработки данных.

							сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	
3	Факторный анализ	5	4	6	76	Контрольная работа Лабораторная работа	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
	ИТОГО	-	10	18	224	-	-	-

4.3. Содержание разделов/тем дисциплины

Множественная линейная регрессия

Введение. Вычисление и оценивание параметров множественной линейной регрессии. Проблема мультиколлинеарности. Пошаговые методы построения оптимальной модели МЛР. Анализ остатков регрессионной модели.

Кластерный анализ

Теория распознавания образов. Общие понятия кластерного анализа. Постановка задачи кластерного анализа. Выбор исходной системы признаков. Снижение размерности пространства признаков. Выбор меры близости. Классификация алгоритмов кластеризации. Оценка качества классификации. Дискриминантный анализ

Факторный анализ

Метод главных компонент. Идеология факторного анализа. Дисперсия факторной модели. Основное уравнение факторного анализа. Геометрическая интерпретация факторного анализа. Фундаментальная теорема факторного анализа. Проблема общностей. Метод главных факторов. Проблема второго вращения.

4.4. Содержание занятий семинарского типа

Таблица 11.

Содержание лабораторных занятий для очной формы обучения

№ темы дисциплины	Тематика лабораторных занятий	Всего часов
Множественная линейная регрессия	“Модель множественной линейной регрессии методом исключения” “Модель множественной линейной регрессии методом включения” “Дискриминантный анализ”	28
Кластерный анализ	“Кластерный анализ”	14
Факторный анализ	“Анализ главных компонент” “Факторный анализ”	14

Таблица 12.

Содержание лабораторных занятий для очно-заочной формы обучения

№ темы дисциплины	Тематика лабораторных занятий	Всего часов
Множественная линейная регрессия	“Модель множественной линейной регрессии методом исключения” “Модель множественной линейной регрессии методом включения” “Дискриминантный анализ”	20
Кластерный анализ	“Кластерный анализ”	8
Факторный анализ	“Анализ главных компонент” “Факторный анализ”	10

Таблица 13.

Содержание лабораторных занятий для заочной формы обучения

№ темы дисциплины	Тематика лабораторных занятий	Всего часов
Множественная линейная регрессия	“Модель множественной линейной регрессии методом исключения” “Модель множественной линейной регрессии методом включения”	6
Кластерный анализ	“Кластерный анализ”	6
Факторный анализ	“Анализ главных компонент” “Факторный анализ”	6

5. Перечень учебно-методического обеспечения самостоятельной работы обучающихся по дисциплине

Moodle

6. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации по итогам освоения дисциплины

6.1. Текущий контроль

Типовые задания, методика выполнения и критерии оценивания текущего контроля по разделам дисциплины представлены в Фонде оценочных средств по данной дисциплине.

6.2. Промежуточная аттестация

Форма промежуточной аттестации по дисциплине – **зачет (7 семестр)/экзамен (8 семестр) для очной/очно-заочной формы обучения**
экзамен (5 год обучения) для заочной формы обучения

Форма проведения зачета: *тестирование*

Перечень вопросов для подготовки к зачету:

ОПК-4.1

1. Многомерный анализ данных, его виды
2. Множественная линейная регрессия
3. Вычисление и оценивание параметров множественной линейной регрессии.
4. Проблема мультиколлинеарности.
5. Пошаговые методы построения оптимальной модели МЛР.
6. Анализ остатков регрессионной модели.

Зачет оценивается по двухбалльной шкале: «зачтено»/ «незачтено».

Критерии оценивания:

«Зачтено» - студент ответил правильно на 70% вопросов теста

«Незачтено» - студент ответил правильно менее чем на 70% вопросов теста

Форма проведения экзамена: *письменно по билетам*

Перечень вопросов для подготовки к экзамену:

ОПК-4.1

1. Многомерный анализ данных, его виды
2. Множественная линейная регрессия
3. Вычисление и оценивание параметров множественной линейной регрессии.
4. Проблема мультиколлинеарности.
5. Пошаговые методы построения оптимальной модели МЛР.
6. Анализ остатков регрессионной модели.
7. Кластерный анализ
8. Измерение близости объектов
9. Характеристики близости объектов
10. Методы кластерного анализа
11. Иерархические алгоритмы
12. Процедуры эталонного типа
13. Факторный анализ
14. Цели и сущность факторного анализа
15. Линейная модель факторного анализа
16. Решение модели факторного анализа
17. Задача о числе факторов
18. Статистическая оценка факторных нагрузок и остаточных дисперсий
19. Дискриминантный анализ
20. Проблема классификации
21. Дискриминирующие функции

22. Классификация объектов с помощью функции расстояния

Экзамен оценивается по четырехбалльной шкале: «отлично» / «хорошо» / «удовлетворительно» / «неудовлетворительно».

Критерии оценивания:

«Отлично» - студент дал ответ на два вопроса, допускаются недочеты

«Хорошо» - студент дал ответ на два вопроса, допустил не грубые ошибки

«Удовлетворительно» - студент дал ответ на один из вопросов, или допустил грубые ошибки в одном из ответов

«Неудовлетворительно» - студент не дал ответ на вопросы или допустил грубые ошибки в обоих вопросах

7. Методические рекомендации для обучающихся по освоению дисциплины

Методические рекомендации ко всем видам аудиторных занятий, а также методические рекомендации по организации самостоятельной работы, в том числе по подготовке к текущему контролю и промежуточной аттестации представлены в Методических рекомендациях для обучающихся по освоению дисциплины «Многомерный анализ данных».

8. Учебно-методическое и информационное обеспечение дисциплины

8.1. Перечень основной и дополнительной учебной литературы

Основная литература

1. Александровская, Ю. П. Многомерный статистический анализ в экономике : учебное пособие / Ю. П. Александровская. — Казань : КНИТУ, 2017. — 96 с.

Дополнительная литература

1. М.С. Аракелов, К.П. Мавриди. Статистика. Учебно-методическое пособие. – СПб.: Изд. РГГМУ, 2012. – 96 с.

2. Кузьмин В. И., Гадзаов А. Ф. Методы анализа данных: Учебное пособие - 2020. 155с.

3. Вуколов В.И. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL.- М.: ФОРУМ-ИНФРА-М, 2004 - 462 с.

4. Александровская, Ю. П. Классификация многомерных данных в экономике: дискриминантный анализ : учебное пособие / Ю. П. Александровская. — Казань : КНИТУ, 2018. — 80 с.

5. Учебное пособие по дисциплине "Статистика" – СПб.: РГГМУ, 2019. – 72 с

6. Макарова Н.В., Трофимец В.Я. Статистика в Excel.- М.: Финансы и статистика, 2002.-365 с.

8.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

1. Савельев В. Статистика и котики- АСТ , Москва. 2018

8.3. Перечень программного обеспечения

1. Microsoft Office
2. Statistica
3. Past4
4. Radiant

8.4. Перечень информационных справочных систем

1. Elib RSHU
2. ЭБС “Лань”

8.5. Перечень профессиональных баз данных

1. Электронно-библиотечная система eLibrary;

9. Материально-техническое обеспечение дисциплины

Лекционные занятия

- Лекционная аудитория

Лабораторные занятия

- Аудитория, оснащенная персональными компьютерами или мультимедийным оборудованием (ауд.101,104,108,23 и кв.14 2-го корпуса РГГМУ)

Самостоятельная работа

- Читальный зал библиотеки, читальный зал Российской Национальной Библиотеки

10. Особенности освоения дисциплины для инвалидов и лиц с ограниченными возможностями здоровья

Обучение обучающихся с ограниченными возможностями здоровья при необходимости осуществляется на основе адаптированной рабочей программы с использованием специальных методов обучения и дидактических материалов, составленных с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся (обучающегося).

При определении формы проведения занятий с обучающимся-инвалидом учитываются рекомендации, содержащиеся в индивидуальной программе реабилитации инвалида, относительно рекомендованных условий и видов труда.

При необходимости для обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья создаются специальные рабочие места с учетом нарушенных функций и ограничений жизнедеятельности.