

Подписано электронной подписью:
Вержицкий Данил Григорьевич
Должность: Директор КГПИ ФГБОУ ВО «КемГУ»
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Дата и время: 2024-02-21 00:00:00
Кузбасский государственный технический университет
федерального государственного бюджетного образовательного учреждения
высшего образования
«Кемеровский государственный университет»

Факультет информатики, математики и экономики

УТВЕРЖДАЮ
Декан А.В. Фомина
10 февраля 2022 г.

**Рабочая программа дисциплины
К.М.05.05 Многомерный анализ данных**

Направление подготовки
01.03.02 Прикладная математика и информатика

Направленность (профиль) подготовки
ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Программа бакалавриата

Квалификация выпускника
бакалавр

Форма обучения
Очная

Год набора 2022

Новокузнецк 2022

Оглавление

1 Цель дисциплины	3
1.1 Формируемые компетенции	3
1.2 Знания, умения, навыки (ЗУВ) по дисциплине	3
2 Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации.	4
3. Учебно-тематический план и содержание дисциплины.....	4
3.1 Учебно-тематический план	4
3.2. Содержание занятий по видам учебной работы.....	5
4 Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации.....	5
5 Материально-техническое, программное и учебно-методическое обеспечение дисциплины.	7
5.1 Учебная литература	7
5.2 Материально-техническое и программное обеспечение дисциплины.....	7
5.3 Современные профессиональные базы данных и информационные справочные системы.....	8
6 Иные сведения и (или) материалы.....	8
6.1. Примерные вопросы и задания / задачи для промежуточной аттестации	8

1 Цель дисциплины.

В результате освоения данной дисциплины у обучающегося должны быть сформированы компетенции основной профессиональной образовательной программы бакалавриата (далее - ОПОП): ОПК-2.

Содержание компетенций как планируемых результатов обучения по дисциплине см. таблицы 1 и 2.

1.1 Формируемые компетенции

Таблица 1 - Формируемые дисциплиной компетенции

Наименование вида компетенции (универсальная, общепрофессиональная, профессиональная)	Наименование категории (группы) компетенций	Код и название компетенции
<i>общепрофессиональная</i>	Теоретические и практические основы профессиональной деятельности	<i>ОПК-2 Способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач</i>

Индикаторы достижения компетенций

Таблица 2 – Индикаторы достижения компетенций, формируемые дисциплиной

Код и название компетенции	Индикаторы достижения компетенции по ОПОП	Дисциплины и практики, формирующие компетенцию ОПОП
ОПК-2 Способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач	2.1 Анализирует и адаптирует математические методы для решения прикладных задач 2.2 Разрабатывает алгоритмы на основе современных математических методов 2.3 Реализует алгоритмы с использованием современных систем программирования	К.М.05.03 Теория игр и исследование операций К.М.05.04 Методы оптимизации К.М.05.05 Многомерный анализ данных К.М.05.06 Математические модели и методы искусственного интеллекта К.М.09.02(У) Научно-исследовательская работа (получение первичных навыков научно-исследовательской работы) К.М.09.05(Н) Научно-исследовательская работа

1.2 Знания, умения, навыки (ЗУВ) по дисциплине

Таблица 3 – Знания, умения, навыки, формируемые дисциплиной

Код и название компетенции	Индикаторы достижения компетенции, закрепленные за дисциплиной	Знания, умения, навыки (ЗУВ), формируемые дисциплиной
ОПК-2 Способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации	2.1 Анализирует и адаптирует математические методы для решения прикладных задач	Знать: – Методы многомерного анализа данных для решения прикладных задач математической статистики в профессиональной области. Уметь: – Адаптировать математические модели многомерного анализа данных к конкретным профессиональным задачам.

Код и название компетенции	Индикаторы достижения компетенции, закреплённые за дисциплиной	Знания, умения, навыки (ЗУВ), формируемые дисциплиной
алгоритмов решения прикладных задач		– выбирать и применять методы многомерного анализа данных, необходимые для решения профессиональных задач Владеть: – Методами многомерного анализа данных для решения задач в профессиональной области.

2 Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации.

Таблица 4 – Объём и трудоёмкость дисциплины по видам учебных занятий

Общая трудоёмкость и виды учебной работы по дисциплине, проводимые в разных формах	Объём часов по формам обучения
	ОФО
1 Общая трудоёмкость дисциплины	108
2 Контактная работа обучающихся с преподавателем (по видам учебных занятий) (всего)	48
Аудиторная работа (всего):	48
в том числе:	
лекции	16
практические занятия, семинары	32
Внеаудиторная работа (всего):	
3 Самостоятельная работа обучающихся (всего)	60
4 Промежуточная аттестация обучающегося - зачет (5 семестр)	

3. Учебно-тематический план и содержание дисциплины.

3.1 Учебно-тематический план

Таблица 5 - Учебно-тематический план очной формы обучения

№ недели п/п	Разделы и темы дисциплины по занятиям	Общая трудоёмкость (всего час.)	Трудоёмкость занятий (час.)		Формы текущего контроля и промежуточной аттестации успеваемости		
			ОФО				
			Аудиторн. занятия				СРС
			лекц.	практ.			
Семестр 5							
1	Задачи и методы многомерного статистического анализа	18	4	6	8	Лабораторная работа 1,2	
2	Факторный анализ	14	2	4	8	Лабораторная работа 3,4	
3	Кластерный анализ	24	4	4	16	Лабораторная работа 5,6	
4	Дискриминантный анализ	16	2	4	10	Лабораторная работа 7,8	
5	Метод канонических корреляций	12	2	4	6	Лабораторная работа 9	
6	Многомерное шкалирование	12	2	4	6	Лабораторная работа 10	
7	Дисперсионный и ранговый однофакторный анализ	12	2	4	6	Лабораторная работа 11	
	Промежуточная аттестация - зачет					зачет	
ИТОГО по семестру 5		108	16	32	60		

3.2. Содержание занятий по видам учебной работы

Таблица 6 – Содержание дисциплины

Содержание лекционного курса

№ п/п	Наименование раздела и темы дисциплины	Содержание
1	Задачи и методы многомерного статистического анализа	Методы многомерного статистического анализа. Многомерное признаковое пространство. Особенности обработки многомерных статистических данных
2	Факторный анализ	Метод главных компонент. Метод главных факторов. Вращение пространства общих факторов
3	Кластерный анализ	Иерархический кластерный Анализ. Метод k-средних. Метод поиска сгущений
4	Дискриминантный анализ	Расчет коэффициентов дискриминантной функции. Классификация при наличии двух и k-обучающих выборок
5	Метод канонических корреляций	Оценка значимости канонических корреляций. Экономическая интерпретация результатов канонического анализа
6	Многомерное шкалирование	Классическая модель многомерного шкалирования Торгерсона. Неметрические методы многомерного шкалирования
7	Дисперсионный и ранговый однофакторный анализ	Дисперсионный анализ в статистических исследованиях

Темы практических занятий

№ п/п	Наименование раздела и темы дисциплины	Содержание
1	Задачи и методы многомерного статистического анализа	Методы многомерного статистического анализа. Многомерное признаковое пространство. Особенности обработки многомерных статистических данных
2	Факторный анализ	Метод главных компонент. Метод главных факторов. Вращение пространства общих факторов
3	Кластерный анализ	Иерархический кластерный Анализ. Метод k-средних. Метод поиска сгущений
4	Дискриминантный анализ	Расчет коэффициентов дискриминантной функции. Классификация при наличии двух и k-обучающих выборок
5	Метод канонических корреляций	Оценка значимости канонических корреляций. Экономическая интерпретация результатов канонического анализа
6	Многомерное шкалирование	Классическая модель многомерного шкалирования Торгерсона. Неметрические методы многомерного шкалирования
7	Дисперсионный и ранговый однофакторный анализ	Дисперсионный анализ в статистических исследованиях

4 Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации.

Для положительной оценки по результатам освоения дисциплины обучающемуся

необходимо выполнить все установленные виды учебной работы. Оценка результатов работы обучающегося в баллах (по видам) приведена в таблице 7.

Таблица 7 - Балльно-рейтинговая оценка результатов учебной работы обучающихся по видам (БРС)

Учебная работа (виды)	Сумма баллов	Виды и результаты учебной работы	Оценка в аттестации	Баллы
Текущая учебная работа в семестре (Посещение занятий по расписанию и выполнение заданий)	80	Лабораторная работа №1 (отчет о выполнении контрольной работы)	3 балла (пороговое значение) 5 баллов (максимальное значение)	3-5
		Лабораторная работа №2 (отчет о выполнении контрольной работы)	3 балла (пороговое значение) 5 баллов (максимальное значение)	3-5
		Лабораторная работа №3 (отчет о выполнении контрольной работы)	3 балла (пороговое значение) 5 баллов (максимальное значение)	3-5
		Лабораторная работа №4 (отчет о выполнении контрольной работы)	3 балла (пороговое значение) 5 баллов (максимальное значение)	3-5
		Лабораторная работа №5 (отчет о выполнении контрольной работы)	4 балла (пороговое значение) 8 баллов (максимальное значение)	4-8
		Лабораторная работа №6 (отчет о выполнении контрольной работы)	4 балла (пороговое значение) 8 баллов (максимальное значение)	4-8
		Лабораторная работа №7 (отчет о выполнении контрольной работы)	5 балла (пороговое значение) 10 баллов (максимальное значение)	5-10
		Лабораторная работа №8 (отчет о выполнении контрольной работы)	5 балла (пороговое значение) 10 баллов (максимальное значение)	5-10
		Лабораторная работа №9 (отчет о выполнении контрольной работы)	4 балла (пороговое значение) 8 баллов (максимальное значение)	4-8
		Лабораторная работа №10 (отчет о выполнении контрольной работы)	4 балла (пороговое значение) 8 баллов (максимальное значение)	4-8
		Лабораторная работа №11 (отчет о выполнении контрольной работы)	3 балла (пороговое значение) 8 баллов (максимальное значение)	3-8
Итого по текущей работе в семестре				41-80
Промежуточная аттестация (зачет)	20	Теоретический вопрос 1	2 балла (пороговое значение) 5 баллов (максимальное значение)	2 - 5
		Теоретический вопрос 2	2 балла (пороговое значение) 5 баллов (максимальное значение)	2 - 5
		Решение задачи 1.	3 балла (пороговое значение) 5 баллов (максимальное значение)	3 - 5
		Решение задачи 2.	3 балла (пороговое значение) 5 баллов (максимальное значение)	3 - 5
Итого по промежуточной аттестации (зачет)				10-20
Суммарная оценка по дисциплине: Сумма баллов текущей и промежуточной аттестации				51 – 100 б.

В промежуточной аттестации оценка выставляется в ведомость в 100-балльной шкале и в буквенном эквиваленте (таблица 8)

Таблица 8 – Соотнесение 100-балльной шкалы и буквенного эквивалента оценки

Сумма набранных баллов	Уровни освоения дисциплины и компетенций	Экзамен		Зачет
		Оценка	Буквенный эквивалент	Буквенный эквивалент
86 - 100	Продвинутый	5	отлично	Зачтено
66 - 85	Повышенный	4	хорошо	
51 - 65	Пороговый	3	удовлетворительно	
0 - 50	Первый	2	неудовлетворительно	Не зачтено

5 Материально-техническое, программное и учебно-методическое обеспечение дисциплины.

5.1 Учебная литература

Основная учебная литература

1. Кулаичев, А.П. Методы и средства комплексного анализа данных [Электронный ресурс]: учебн. пособие / А.П. Кулаичев. – Электрон.текстовые дан. – Москва: ИНФРА-М, 2016. – Режим доступа: <https://new.znaniium.com/read?id=252560>
2. Козлов, А.Ю. Статистический анализ данных в MSExcel [Электронный ресурс]: учебн. пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. – Электрон.текстовые дан. – Москва: ИНФРА-М, 2012. – Режим доступа: <http://www.znaniium.com/bookread.php?book=238654>
3. Горлач, Б.А. Теория вероятностей и математическая статистика [Электронный ресурс]: Учебник / Б.А. Горлач – Электрон. текстовые дан. – Москва: Лань, 2013. – Режим доступа: <http://e.lanbook.com/reader/book/4864/>

Дополнительная учебная литература

1. Кибзун, А.И. Теория вероятностей и математическая статистика [Электронный ресурс]. Базовый курс с примерами и задачами / А.И. Кибзун, Е.Р. Горяинова, А.В. Наумов – Электрон. текстовые дан. – Москва: Лань, 2005. – Режим доступа: <http://e.lanbook.com/reader/book/2198/>
2. Хрущева, И.В. Теория вероятностей [Электронный ресурс]: Учебник / И.В. Хрущева – Электрон. текстовые дан. – Москва: Лань, 2009. – Режим доступа: <http://e.lanbook.com/reader/book/425/>
3. Туганбаев, А.А. Теория вероятностей и математическая статистика [Электронный ресурс]: Учебник / А.А. Туганбаев, В.Г. Крупин – Электрон. текстовые дан. – Москва: Лань, 2011. – Режим доступа: <http://e.lanbook.com/reader/book/652/>
4. Бородин, А.Н. Элементарный курс теории вероятностей и математической статистики [Электронный ресурс]: Учебник / А.Н. Бородин – Электрон. текстовые дан. – Москва: Лань, 2011. – Режим доступа: <http://e.lanbook.com/reader/book/2026/>

5.2 Материально-техническое и программное обеспечение дисциплины.

Учебные занятия по дисциплине проводятся в учебных аудиториях КГПИ ФГБОУ ВО «КемГУ»:

404 Учебная аудитория для проведения: - занятий лекционного типа; - групповых и индивидуальных консультаций; - текущего контроля и промежуточной аттестации. Специализированная (учебная) мебель: доска меловая, кафедра, столы, стулья.	Учебный корпус №4. 654079,
--	-----------------------------------

Оборудование: <i>переносное</i> - ноутбук, экран, проектор. Используемое программное обеспечение: MSWindows (MicrosoftImaginePremium 3 year по лицензионному договору № 1212/KMP от 12.12.2018 г. до 12.12.2021 г.), LibreOffice (свободно распространяемое ПО), Яндекс.Браузер (отечественное свободно распространяемое ПО). Интернет с обеспечением доступа в ЭИОС.	Кемеровская область, г. Новокузнецк, пр-кт Metallургов, д. 19
502 Компьютерный класс. Учебная аудитория (мультимедийная) для проведения: - занятий лекционного типа; - занятий семинарского (практического) типа; - занятий лабораторного типа; - групповых и индивидуальных консультаций; - самостоятельной работы; - текущего контроля и промежуточной аттестации. Специализированная (учебная) мебель: доска меловая, столы компьютерные, стулья. Оборудование для презентации учебного материала: <i>стационарное</i> - компьютер, экран, проектор, наушники. Лабораторное оборудование: стационарное – компьютеры для обучающихся (16 шт.). Используемое программное обеспечение: MSWindows (MicrosoftImaginePremium 3 year по лицензионному договору № 1212/KMP от 12.12.2018 г. до 12.12.2021 г.), LibreOffice (свободно распространяемое ПО), Firefox 14 (свободно распространяемое ПО), Яндекс.Браузер (отечественное свободно распространяемое ПО), MicrosoftVisualStudio (MicrosoftImaginePremium 3 year по лицензионному договору № 1212/KMP от 12.12.2018 г. до 12.12.2021 г.), Среда статистических вычислений Rv.4.0.2 (свободно распространяемое ПО). Интернет с обеспечением доступа в ЭИОС.	Учебный корпус №4. 654079, Кемеровская область, г. Новокузнецк, пр-кт Metallургов, д. 19

5.3 Современные профессиональные базы данных и информационные справочные системы.

Перечень СПБД и ИСС по дисциплине

1. Информационная система «Единое окно доступа к информационным ресурсам» <http://window.edu.ru/catalog/>
2. Общероссийский математический портал (информационная система) - <http://www.mathnet.ru/>
3. Научная электронная библиотека eLIBRARY.RU – крупнейший российский информационный портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты - www.elibrary.ru
4. База данных Science Direct (более 1500 журналов издательства Elsevier, среди них издания по математике и информатике), режим доступа :<https://www.sciencedirect.com>

6 Иные сведения и (или) материалы.

6.1. Примерные вопросы и задания / задачи для промежуточной аттестации

1. Многомерная регрессионная модель.
2. Методы снижения размерности.
3. Компонентный анализ и основы факторного анализа.
4. Кластерный анализ.
5. Дискриминантный анализ.
6. Корреляционный анализ.
7. Регрессионный анализ.
8. Канонические корреляции.
9. Нелинейная регрессия. Примеры нелинейной регрессии.

10. Распределение генеральной совокупности.
11. Некоторые характеристики генеральной совокупности.
12. Параметры связи между признаками в генеральной совокупности.
13. Многомерная нормально распределенная генеральная совокупность.
14. Выборка из генеральной совокупности.
15. Точечные оценки параметров многомерной генеральной совокупности.
16. Доверительные области.
17. Гипотезы о параметрах многомерной нормально распределенной генеральной совокупности.
18. Основные понятия.
19. Двумерная и трехмерная модель.
20. Корреляционный анализ многомерной модели.
21. Проверка гипотезы о нормальном распределении генеральной совокупности.
22. Ранговая корреляция.
23. Простейшее линейное уравнение регрессии.
24. Множественное линейное уравнение регрессии.
25. Статистический подход в методе главных компонент.
26. Линейная модель метода главных компонент.
27. Квадратичные формы и главные компоненты.
28. Метод главных факторов и его алгоритм.
29. Проблема вращения.
30. Проблема оценки факторов и задачи классификации.
31. Классификация задач факторного анализа и метода главных компонент.
32. Классификация без обучения. Кластерный анализ.
33. Дискриминантный анализ.

Практические задания

1. Смоделировать выборку 80 значений случайной величины X , имеющей равномерное распределение в промежутке $[0, 4]$, проверив гипотезу о форме распределения.
2. Средний объем стока воды в реке фиксируется каждый месяц в течение двух лет. В таблице приведены наблюдаемые объемы стока в фут³/сек.

Год	Месяц											
	1	2	3	4	5	6	7	8	9	10	11	12
1	15	11	123	190	138	98	88	80	76	49	27	16
2	14	12	115	210	130	45	43	40	35	30	29	17

Используя критерии Вальда – Вольфовица и Вилкоксона при уровне значимости $\alpha = 0,05$ проверить гипотезу об отсутствии систематических изменений объема стока из года в год.

1. Произведено по 32 бросания пяти исследуемых монет и подсчитано количество k выпавших орлов $X = 0, 1, 2, 3, 4, 5$ в каждом испытании. В таблице приведены результаты.

X	0	1	2	3	4	5
k	10	1	5	1	5	10

Используя критерий Колмогорова – Смирнова, при уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что среди этих монет нет поддельных.

2. В таблице приведены ранги объектов X, Y .

X	1	2	3	4	5	6	7	8	9	10
Y	1	7	3	10	5	2	9	8	4	6

Найти выборочные коэффициенты ранговой корреляции Спирмена и Кендалла, а также проверить гипотезу об отсутствии ранговой корреляции при уровне значимости $\alpha = 0,05$.

1. Шесть пар участников соревнования по фигурному катанию на льду оценивались двумя членами жюри. В результате этой оценки у каждого члена жюри пара получала определенное место. Необходимо проверить с заданным уровнем значимости $\alpha = 0,05$ нулевую гипотезу $H_0 : W = 0$ об отсутствии согласованности оценок членов жюри. Результаты оценок приведены в таблице.

Член жюри	Номер пары					
	1	2	3	4	5	6
1	1	2	3	4	5	6
2	1	2	4	5	3	6

2. В таблице приведены данные об урожайности сельскохозяйственной культуры за 6 лет при разных технологиях обработки почвы. Выясним на уровне значимости $\alpha = 0,05$, зависит ли урожайность сельскохозяйственной культуры от технологии обработки почвы. Если зависит, то найти коэффициент детерминации технологии.

Номер технологии	Год наблюдения					
	1-й	2-й	3-й	4-й	5-й	6-й
1	140	141	140	141	142	145
2	150	149	150	147		
3	147	147	145	150	150	
4	144	147	142	146		

1. Смоделировать выборку 100 значений случайной величины X , распределенной нормально с параметрами $(20, 10)$, проверив гипотезу о форме распределения.

2. В стоматологическую поликлинику прибывает простейший поток больных с интенсивностью $\lambda = 5$ чел/час в обычное время дня, в часы «пик» она возрастает до $\lambda_+ = 12$ чел/час, а в часы «спада» достигает величины $\lambda_- = 3$ чел/час. Средняя продолжительность сеанса лечения составляет $\overline{t_{об}} = 0,5$ час/чел. Определить вероятность образования очереди $P_{оч}$ и среднюю длину очереди $\overline{l_{оч}}$ в течение дня, а также необходимое число стоматологов в часы «пик» n_+ и «спада» n_- , обеспечивающих такую же длину очереди и вероятность ее образования, как в обычное время дня.

В. 14

1. В стоматологическую поликлинику прибывает простейший поток больных с интенсивностью $\lambda = 5$ чел/час. Средняя продолжительность сеанса лечения составляет $\overline{t_{об}} = 0,5$ час/чел. Уровень суммарных потерь связан с простым средним числом свободных стоматологов $n_{св}$ и пребыванием среднего числа больных в очереди $L_{оч}$. Построить график зависимости $f(n) = n_{св} + L_{оч}$, где n – количество стоматологов, и определить оптимальное количество стоматологов, при котором уровень суммарных потерь минимален.

2. Фирма «Надежный компьютер», которая реализует компьютеры, стремится оптимизировать затраты, связанные с управлением запасов. Пусть суточный спрос является равномерно распределенной величиной, которая принимает значения $q = 1, \dots, 10$, а вероятности $p(q) = 0,1$ при всех значениях q . Рассматриваются затраты с такими же параметрами, как в примере 1: $\pi = 10$; $c = 3$; $K = 8$; $h = 2$. Найти методом статистических испытаний оценку суточных суммарных затрат для оптимальной стратегии пополнения

запасов и исследовать чувствительность этой оценки к параметрам стратегии.

Составитель (и): канд. технн. наук Вячкин Е.С.

(фамилия, инициалы и должность преподавателя (ей))