

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный университет
им. Н.И. Лобачевского»**

Институт информационных технологий, математики и механики

УТВЕРЖДЕНО

решением президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

Рабочая программа дисциплины

Методы моделирования и прикладной анализ данных

Уровень высшего образования

Бакалавриат

Направление подготовки / специальность

01.03.02 - Прикладная математика и информатика

Направленность образовательной программы

Прикладная математика и информатика (общий профиль)

Форма обучения

очная

г. Нижний Новгород

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.08.01 Методы моделирования и прикладной анализ данных относится к части, формируемой участниками образовательных отношений образовательной программы.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства	
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	Для текущего контроля успеваемости	Для промежуточной аттестации
ПК-4: Способен применять методы математического и компьютерного исследования при анализе задач на основе знаний фундаментальных математических и компьютерных наук	<p>ПК-4.1: Знает фундаментальные и теоретические основы, необходимые для исследования научных проблем</p> <p>ПК-4.2: Умеет самостоятельно применять полученные знания для анализа объекта исследования, определять цели и задачи исследования, а также выбирать корректный метод исследования научной проблемы</p> <p>ПК-4.3: Имеет практический опыт научно-исследовательской деятельности, а именно решения научных задач в соответствии с поставленной целью и выбранной методикой</p>	<p>ПК-4.1: Знает методы моделирования в сфере прикладного анализа данных: разведочный, корреляционный, дисперсионный анализ, снижение размерности методами кластерного анализа и методом главных компонент, основы факторного и регрессионного анализа. Знает понятийный аппарат, математические свойства (с доказательством) и способ применения. Знает типы и источники данных, способы представления данных в пакетах статистических программ. Знает основы статистического вывода для оценки качества обработки данных указанными методами.</p> <p>ПК-4.2: На примере модельных задач умеет формулировать цели исследования, применять методы прикладного анализа данных, анализировать протоколы обработки данных в профессиональных статистических пакетах, формулировать результаты. Имеет навыки оценки</p>	<p>Практическое задание</p> <p>Проект</p> <p>Собеседование</p> <p>Тест</p>	<p>Зачёт:</p> <p>Контрольные вопросы</p>

		<p>программных средств анализа данных, необходимые для развития таких программных средств и направлений их разработки.</p> <p>ПК-4.3: Владеет навыками самостоятельного выполнения проектов с использованием пакетов прикладных программ, включая выбор цели исследования, выбор метода моделирования, обоснование его применения, проверку качества решения, сравнительный анализ и интерпретацию результатов исследования.</p>		
--	--	--	--	--

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	2
Часов по учебному плану	72
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	16
- занятия семинарского типа (практические занятия / лабораторные работы)	16
- КСР	1
самостоятельная работа	39
Промежуточная аттестация	0 Зачёт

3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего (часы)	в том числе			Самостоятельная работа обучающегося, часы
		Контактная работа (работа во взаимодействии с преподавателем), часы из них			
		Занятия лекционного типа	Занятия семинарского типа (практические занятия/ лабора	Всего	

			торные работы), часы		
	о ф о	о ф о	о ф о	о ф о	о ф о
Тема 1. Введение в предмет. Примеры применения методов моделирования для анализа данных в бизнесе, промышленности и социальной сфере. Постановки управленческих задач. Источники и типы данных. Нормативно-правовые аспекты сбора данных. Интерпретация результатов и выводы для управленческих решений.	4	2		2	2
Тема 2. Методы разведочного, корреляционного и дисперсионного анализа.	10	2	2	4	6
Тема 3. Методы кластерного анализа. Теория и приложения.	10	2	2	4	6
Тема 4. Метод главных компонент. Теория и приложения.	14	4	4	8	6
Тема 5. Введение в регрессионный и факторный анализ.	14	4	4	8	6
Тема 6. Основы статистического вывода для методов п. 2 – п. 5.	7	2	2	4	3
Тема 7. Применение методов для анализа модельных задач.	12		2	2	10
Аттестация	0				
КСР	1			1	
Итого	72	16	16	33	39

Содержание разделов и тем дисциплины

Тема 1. Введение в предмет. Примеры применения методов моделирования для анализа данных в бизнесе, промышленности и социальной сфере. Постановки управленческих задач. Источники и типы данных. Нормативно-правовые аспекты сбора данных. Интерпретация результатов и выводы для управленческих решений.

Тема 2. Методы разведочного, корреляционного и дисперсионного анализа. Представление исходных данных. Описательная статистика признака и преобразования признаков. Анализ линейной зависимости признаков. Геометрический смысл парной корреляции. Геометрический смысл корреляции в пространстве признаков. Описательная статистика матрицы данных. Вычисление R и $V(X)$. Корреляции (ковариации) двух наборов признаков. Свойства собственных чисел и собственных векторов. Собственные пары матриц R и $V(X)$. Ортогональные матрицы и преобразования признаков. Дисперсионный анализ (примеры).

Тема 3. Методы кластерного анализа. Теория и приложения. Задача и инструменты кластерного анализа. Оценка качества кластерного решения. Иерархический кластерный анализ. Метод К-средних. Сопоставление методов. Иерархический кластерный анализ по методу Уорда. Критерии выбора кластерного решения.

Тема 4. Метод главных компонент. Теория и приложения. Извлечение главных компонент на основе корреляционной матрицы. Главные компоненты в матричной записи. Свойства главных компонент. Геометрический смысл главных компонент. Дисперсия исходных признаков и главных компонент. Примеры. Критерии снижения размерности. Корреляция исходных признаков и главных компонент. Воспроизведенные корреляции. Стандартизация главных компонент. Извлечение главных компонент на основе ковариационной матрицы $V(X)$. Специфика применения метода.

Тема 5. Введение в регрессионный и факторный анализ. Принцип наименьших квадратов. Обоснование МНК. Наклон и корреляция. Свойства остатков и оценочных значений. Правило разложения вариации. Коэффициент детерминации. Связь коэффициентов корреляции и детерминации. Анализ остатков. МНК в случае нескольких переменных. Векторная форма записи. Приведенный коэффициент детерминации. Геометрический смысл МНК и критерии качества решения. Формулировка факторной модели. Свойства матриц факторных нагрузок. Обоснование вращения факторов. Схема факторного анализа. Оценка факторных значений. Решение модельной задачи. Критерии качества решения.

Тема 6. Основы статистического вывода для методов п. 2 – п. 5. Анализ значимости корреляций, дисперсионный анализ значимости различий кластерных центров, анализ значимости модели регрессии, статистический тест на сферичность.

Тема 7. Применение методов для анализа модельных задач: разбор модельных задач для применения методов К-средних, главных компонент, иерархического кластерного анализа, оценки моделей простой и множественной регрессии, факторного анализа. Самостоятельный комбинированный анализ данных.

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

Для обеспечения самостоятельной работы обучающихся используются:

- электронный курс "Методы моделирования социально-экономических процессов: анализ данных" (<https://e-learning.unn.ru/course/view.php?id=814>).

Иные учебно-методические материалы: 1. Садовничий В.А. Большие данные в современном мире. (15 ноября 2017 года ректор МГУ имени М.В. Ломоносова академик В.А. Садовничий прочитал перед студентами, аспирантами и сотрудниками первую лекцию в рамках нового межфакультетского учебного курса). Видеозапись <http://media.msu.ru/?p=16379>

2. Балабанов А.С., Стронгина Н.Р. Анализ данных в экономических приложениях. Учебное пособие. Н. Новгород: Изд-во ННГУ, 2004. – 136 с.

3. Стронгина Н.Р., Марчева И.А. Регрессионный анализ в экономических приложениях: Учебное пособие. Н. Новгород: Изд-во ННГУ, 2002. – 214 с.

4. Балабанов А.С., Стронгина Н.Р. Анализ данных в экономических приложениях. Компьютерный практикум в SPSS. Н. Новгород: Изд-во ННГУ, 2003. – 100 с.

5. Балабанов А.С., Власова О.В., Стронгина Н.Р. Статистический вывод в экономических приложениях. Н. Новгород: Изд-во ННГУ, 2002. - 67 с.

6. Стронгина Н.Р. Элементы теории стабильного населения. Учебное пособие. Н. Новгород: Изд-во ННГУ, 2002. – 80 с.

7. Стронгина Н.Р. Методы моделирования социально-экономических процессов: регрессионный анализ. ЭУК, учебно-методический комплекс. Фонд электронных образовательных ресурсов ННГУ. Н. Новгород, 2014. Идентификационный номер 793Е.14.08. <https://e-learning.unn.ru/course/view.php?id=793>

5. Фонд оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю)

5.1 Типовые задания, необходимые для оценки результатов обучения при проведении текущего контроля успеваемости с указанием критериев их оценивания:

5.1.1 Типовые задания (оценочное средство - Практическое задание) для оценки сформированности компетенции ПК-4:

Практические задания по дисциплине включают:

1. Стартовый практикум. Первичный анализ данных на примере модельных задач, анализ структуры и динамики (рекомендуется EXCEL)
2. Решение модельных задач методами кластерного анализа (рекомендуется SPSS)

3. Решение модельных задач методом главных компонент (рекомендуется EXCEL, SPSS)
4. Построение оценочного уравнения простой и множественной регрессии для модельных задач (рекомендуется EXCEL, SPSS)
5. Решение модельных задач методом факторного анализа (рекомендуется SPSS)
6. Практические задания теоретического характера

Содержание и порядок выполнения заданий, а также файлы данных представлены на ресурсе:

Стронгина Н.Р. Методы моделирования социально-экономических процессов: анализ данных. ЭУК, учебно-методический комплекс. Фонд электронных образовательных ресурсов ННГУ. Н. Новгород, 2014. Идентификационный номер 814E.14.08. - Доступ требует

авторизации, <https://e-learning.unn.ru/enrol/index.php?id=827>

(см. Стартовый практикум, Практикум 1, Практикум 2, Практикум 3 и Дополнительный практикум на ресурсе <https://e-learning.unn.ru/enrol/index.php?id=827>)

В ходе выполнения каждого практического задания

1. Опишите комплект данных модельной задачи.
2. Сформулируйте цель и задачу прикладного анализа данных.
3. Проведите первичный анализ данных.
4. Аргументируйте применение метода моделирования.
5. Опишите понятийный аппарат исследования, сформулируйте свойства метода.
6. Опираясь на математическую модель, объясните протокол обработки данных на компьютере.
7. Оцените качество построенной модели, сформулируйте результаты исследования.

Практические задания теоретического характера (примеры)

Сформулируйте определения и докажите утверждения:

1. Свойства коэффициента ковариации и свойства коэффициента корреляции Пирсона для исходных, центрированных, стандартизованных признаков.
2. Геометрические свойства суммарной дисперсии (кластерный анализ).
3. Геометрические свойства показателя качества кластерного решения D.
4. Правило разложения дисперсии на внутригрупповую и междугрупповую.
5. Утверждения о средних и дисперсиях главных компонент, о суммарной дисперсии главных компонент, об отсутствии корреляции главных компонент, свойствах весовых коэффициентов главных компонент.
6. Определение общностей, их смысл и свойства, определение доли суммарной объясненной дисперсии. Выбор количества главных компонент.
7. Факторная модель для стандартизованных признаков в канонической и векторной записи. Воспроизведенная матрица корреляций. Общности, характерности и доля объясненной дисперсии. Теорема о корреляции признаков и общих факторов.
8. Теорема о «диагональных» условиях нормировки. Принцип главных факторов.
9. Оценка факторных нагрузок методом главных факторов.
10. Оценка факторных значений.

Критерии оценивания (оценочное средство - Практическое задание)

Оценка	Критерии оценивания
зачтено	Задание выполнено верно в полном объеме, либо выполнено в полном объеме с незначительными ошибками
не зачтено	Задание не выполнено, выполнено в существенно неполном объеме или допущены существенные ошибки

5.1.2 Типовые задания (оценочное средство - Проект) для оценки сформированности компетенции ПК-4:

Проект: Самостоятельный прикладной анализ данных комбинированным методом
Тематика:

1. Самостоятельный анализ данных комбинированным методом первичного анализа, кластерного анализа (К-средних) и статистического вывода
2. Самостоятельный анализ данных комбинированным методом первичного анализа, иерархического кластерного анализа и статистического вывода
3. Самостоятельный анализ данных комбинированным методом первичного анализа, регрессионного анализа и статистического вывода
4. Самостоятельный анализ данных комбинированным методом первичного анализа, факторного анализа (метод главных факторов) и статистического вывода
5. Самостоятельный анализ данных комбинированным методом первичного анализа, метода главных компонент и статистического вывода

Индивидуальные и групповые задания проектов различаются комплектами (файлами) данных, тематикой и применением разных методов моделирования.

Комплекты (файлы) данных и методические указания представлены на ресурсе:

Стронгина Н.Р. Методы моделирования социально-экономических процессов: анализ данных. ЭУК, учебно-методический комплекс. Фонд электронных образовательных ресурсов ННГУ. Н. Новгород, 2014. Идентификационный номер 814Е.14.08. - Доступ требует авторизации, <https://e-learning.unn.ru/enrol/index.php?id=827>.

(см. Практикум 4 на ресурсе <https://e-learning.unn.ru/enrol/index.php?id=827>)

В ходе выполнения проекта:

1. Используйте данные в соответствии с индивидуальным заданием.
2. Определите цель исследования и массив исследуемых факторов.
3. Используйте методы первичного анализа, статистический вывод и методы снижения размерности.
4. Программные средства по выбору студента, рекомендуется статистический пакет.
5. В соответствии с методическими указаниями подготовьте отчет.

При подготовке отчета:

1. Опишите комплект данных проекта, укажите целесообразность их фильтрации и (или) группировки.
2. Сформулируйте цели и задачи прикладного анализа данных.
3. Сформулируйте результаты первичного анализа данных.
4. Аргументируйте выбор методов моделирования.

5. Опишите понятийный аппарат методов, сформулируйте основные свойства методов.
6. Опираясь на математическую модель и протокол обработки данных, опишите этапы исследования.
7. Структурируйте полученные результаты, проведите сравнительный анализ, сформулируйте выводы.
8. Оцените качество построенной модели. 9. Укажите инструменты исследования.

Критерии оценивания (оценочное средство - Проект)

Оценка	Критерии оценивания
зачтено	Проект выполнен верно и в полном объеме, либо выполнен в полном объеме с незначительными ошибками
не зачтено	Проект не выполнен, выполнен в неполном объеме или допущены существенные ошибки

5.1.3 Типовые задания (оценочное средство - Собеседование) для оценки сформированности компетенции ПК-4:

1. Матрица данных, типы признаков (числовой, ординальный, номинальный), срезы данных (продольный, поперечный). Описательная статистика числового признака (среднее, дисперсия, стандарт). Центрирование и стандартизация данных. Мода, медиана, гистограмма, квартили, выброс.
2. Ковариация и корреляция. Понятия строгой, сильной, средней, слабой положительной и отрицательной корреляции и ее отсутствия. Свойства коэффициента ковариации и свойства коэффициента корреляции Пирсона для исходных, центрированных, стандартизованных признаков. Анализ диаграмм рассеивания. Ложная корреляция, выбросы, корреляция и нелинейная зависимость.
3. Статистика признаков. Вектор средних значений. Матрица ковариаций. Матрица корреляций. Матрица ковариаций двух наборов признаков (определения). Свойства матрицы ковариаций для центрированных и стандартизованных признаков. Способы записи и вычисления матрицы ковариаций и корреляций.
4. Статистика случайных векторов. Математическое ожидание случайного вектора. Матрицы ковариаций и корреляций случайного вектора. Матрица ковариаций двух случайных векторов (определения). Запись матрицы ковариаций и корреляций через математические ожидания. Свойства матрицы ковариаций стандартизованного случайного вектора.
5. Теорема о дисперсии линейной комбинации независимых случайных величин. Теорема о линейных свойствах ковариации случайных величин. Теорема о дисперсии линейной комбинации двух признаков. Теорема о линейных свойствах ковариации признаков (формулировки). Ортогональные матрицы и их свойства (формулировки).
6. Понятие суммарной дисперсии признаков (для матрицы данных). Геометрические свойства суммарной дисперсии.
7. Задача кластерного анализа. Таблица принадлежности объектов кластерам. Понятия центра кластера, расстояния от объекта до кластера, расстояния между кластерами, расстояния от объекта до «своего кластера». Понятие показателя качества кластерного решения D.
8. Геометрические свойства показателя качества кластерного решения D.
9. Правило разложения дисперсии на внутригрупповую и междугрупповую дисперсии. Статистический смысл показателя D и его свойства.
10. Алгоритм и основное свойство метода K-средних. Принцип построения цепочки кластерных решений и выбора решения.
11. Метод иерархического кластерного анализа. Способы определения сходства объектов, сходства объекта и группы объектов, сходство двух групп объектов (способы выбора расстояния, способы выбора

правила связи).

12. Алгоритм метода иерархического кластерного анализа. Схема объединения, расстояние объединения, функция объединения. Критерий выбора кластерного решения.
13. Сравнение методов К-средних и иерархического метода кластеризации.
14. Условия, цели и схема применения метода главных компонент (МГК). Критерии качества снижения размерности (доля объясненной суммарной дисперсии, общности, по лекциям). Матрица весовых коэффициентов главных компонент (ВКГК), смысл ее коэффициентов. Выбор названий главных компонент.
15. Исходные данные для МГК, свойства матрицы корреляций (R), используемые при построении метода, связь матриц R и Z , определение главных компонент. Определение матрицы значений главных компонент. Размерность, свойства и связь матриц C , Z , W .
16. Свойства главных компонент: утверждения о средних и дисперсиях ГК.
17. Суммарная дисперсия набора признаков и ее геометрический смысл. Утверждение о суммарной дисперсии ГК, его геометрический смысл.
18. Утверждение об отсутствии корреляции главных компонент. ПК-3
19. Утверждения о свойствах матрицы ВКГК (весовых коэффициентов главных компонент).
20. Определение общностей через коэффициенты матрицы ВКГК, их смысл и свойства.
21. Определение показателя δ через коэффициенты матрицы ВКГК, его смысл и свойства.
22. Выбор количества главных компонент при решении задач снижения размерности.
23. Модель факторного анализа в канонической и векторной записи. Основное свойство факторной модели. Воспроизведенная матрица ковариаций.
24. Общности, характерности и доля объясненной дисперсии. Теорема о ковариациях исходных признаков и общих факторов.
25. Ортогональные матрицы. Теорема о вращения факторов. Инварианты при вращении. Цели вращения факторов.
26. Факторный анализ: теорема о линейном преобразовании признаков. Модели
27. Факторная модель для стандартизованных признаков в канонической и векторной записи. Основное свойство и его обоснование. Воспроизведенная матрица корреляций.
28. Общности, характерности и доля объясненной дисперсии (определения, свойства и их доказательства).
29. Теорема о корреляции признаков и общих факторов. Диаграммы факторных нагрузок. Принципы вращения факторов.
30. Модель для стандартизованных признаков. Задачи и схема исследования. Особенности применения модели. Критерии качества решения.
31. Геометрический смысл корреляции признаков в пространстве объектов и пространстве признаков. Геометрический смысл факторного анализа.
32. Проблемы поиска факторных нагрузок. Теорема о «диагональных» условиях нормировки.
33. Принцип главных факторов (теорема). Формулировки теорем о ранге матричного произведения и спектральном разложении.
34. Обоснование ограничений на число общих факторов.
35. Оценка факторных нагрузок методом главных факторов: постановка задачи, алгоритм и критерии остановки. Оценка числа общих факторов и начальная оценка общностей.
36. Оценка факторных значений: матрица данных как запись реализаций случайных векторов, задача на отыскание факторных значений, регрессионный способ оценки. Свойства оценочных матриц F , $V(F)$ и A . Факторные значения и вращение факторов.
37. Многомерная статистическая модель «Территория. Год. Показатели». Анализ структуры и динамики.
38. Модели поколения и населения. Показатели таблиц дожития.
39. Статистический анализ таблиц дожития.

Критерии оценивания (оценочное средство - Собеседование)

Оценка	Критерии оценивания
зачтено	Ответы на вопросы даны верно или имеют незначительные неточности
не зачтено	Ответов нет или допущены грубые ошибки

5.1.4 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ПК-4:

Вопросы теста о процедуре применении метода (пример)

Вопрос 1

Тип вопроса: ввод значения

Формулировка вопроса: Файл данных Практикум_1.sav. Формат выдачи: SPSS.

Модельная задача для метода К-средних. (формат «целое число»)

Ответ 1: число исследуемых признаков...

Верное значение: 4

Ответ 2: число объектов ...

Верное значение: 108

Ответ 3: число объектов с пропусками данных ...

Верное значение: 0

Ответ 4: строим кластеры в пространстве размерности ...

Верное значение: 4

Вопрос 2

Тип вопроса: одиночный выбор

Формулировка вопроса: Файл данных Практикум_1.sav. Формат выдачи: SPSS.

Модельная задача для метода К-средних.

Кластеризация будет проведена...

Варианты ответов:

* для исходных признаков

* для исходных стандартизованных признаков (+)

Вопрос 3

Тип вопроса: ввод значения

Формулировка вопроса: Файл данных Практикум_1.sav. Формат выдачи: SPSS.

Модельная задача для метода К-средних.

Общий центр в исходных координатах (формат **.*)

Ответ 1: LE_m...

Верное значение: 62.7

Ответ 2: LE_f...

Верное значение: 67.8

Ответ 3: КМС...

Верное значение: 46.9

Ответ 4: КСР...

Верное значение: 3.7

Вопрос 4

Тип вопроса: ввод значения

Формулировка вопроса: Файл данных Практикум_1.sav. Формат выдачи: SPSS.

Модельная задача для метода К-средних.

Общий центр в стандартизованных координатах (формат **.*)

Ответ 1: LE_m... Верное значение: 0.0

Ответ 2: LE_f... Верное значение: 0.0

Ответ 3: КМС... Верное значение: 0.0

Ответ 4: КСР... Верное значение: 0.0

Вопрос 5

Тип вопроса: ввод значения

Формулировка вопроса: Файл данных Практикум_1.sav. Формат выдачи: SPSS.

Модельная задача для метода К-средних. (формат *.***)

Ответ: исследуемая суммарная дисперсия ...

Верное значение: 4.000

Вопрос 6

Тип вопроса: ввод значения

Формулировка вопроса:

Файл данных Практикум_1.sav. Формат выдачи: SPSS.

Модельная задача для метода К-средних.

Показатели качества решений (в процентах)... (формат ***.*)

Ответ 1: 2 кластера ...

Верное значение: 79.4%

Ответ 2: 3 кластера ...

Верное значение: 88.0%

Ответ 3: 4 кластера ...

Верное значение: 91.1%

Вопрос 7

Тип вопроса: одиночный выбор

Формулировка вопроса:

Файл данных Практикум_1.sav. Формат выдачи: SPSS.

Модельная задача для метода К-средних.

Для решения на 3 кластера при значимости 0.01 кластерные центры

Варианты ответов:

* незначимо отличаются по всем координатам

* по некоторым координатам значимо не отличаются

* по каждой координате отличаются значимо (+)

* нет значимых отличий по всем координатам

Вопрос 8

Тип вопроса: ввод значения

Формулировка вопроса: Файл данных Практикум_1.sav. Формат выдачи: SPSS.

Модельная задача для метода К-средних.

Для решения на 3 кластера: стандартизованные координаты кластерного центра самого многочисленного кластера ... (формат *.***)

Ответ 1: LE_m...

Верное значение: 0.762

Ответ 2: LE_f...

Верное значение: 0.812

Ответ 3: KMC...

Верное значение: -0.808

Ответ 4: KCP...

Верное значение: -0.901

Вопрос 9

Тип вопроса: классификация

Формулировка вопроса: Файл данных Практикум_1.sav. Формат выдачи: SPSS.

Модельная задача для метода К-средних.

Для решения на 3 кластера: координата кластерного центра самого многочисленного кластера отличается от координаты центра совокупности...

Ответ 1: LE_m, Классифицирующий термин – с превосходством

Ответ 2: LE_f, Классифицирующий термин – с превосходством

Ответ 3: KMC, Классифицирующий термин – с недостатком

Ответ 4: KCP, Классифицирующий термин – с недостатком

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
зачтено	Вопросы теста отвечены правильно, работа над ошибками проведена
не зачтено	Тест не отвечен, либо ответы частично правильны, работа над ошибками не проведена

5.2. Описание шкал оценивания результатов обучения по дисциплине при промежуточной аттестации

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатор достижения компетенций)	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено			зачтено			
<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущено много негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Ошибок нет.	Уровень знаний в объеме, превышающем программу подготовки.

<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие базовых навыков. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторым и недочетами	Продемонстрированы базовые навыки при решении стандартных задач с некоторым и недочетами	Продемонстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрирован творческий подход к решению нестандартных задач

Шкала оценивания при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно», продемонстрированы знания, умения, владения по соответствующим компетенциям на уровне выше предусмотренного программой
	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично».
	очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо».
	удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно».
	плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

5.3 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения на промежуточной аттестации с указанием критериев их оценивания:

5.3.1 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ПК-4

Контрольные вопросы по итогам выполнения практического задания теоретического характера

1. Приведите постановку задачи, сформулируйте метод моделирования, приведите определения и доказательство утверждений в соответствии с заданием.
2. Укажите место и роль доказанных утверждений в системе теоретического обоснования метода
3. Укажите место и роль доказанных утверждений в сфере применения метода.

Контрольные вопросы по итогам выполнения других практических заданий (практикум)

1. Опишите исходные данные практического задания (файл данных).
2. Сформулируйте цель и задачу исследования.
3. Проведите первичный анализ данных.
4. Аргументируйте применение предложенного в задании метода.
5. Опишите понятийный аппарат проведенного исследования, сформулируйте его основные свойства.
6. Комментируйте компьютерную диагностику построения решения.
7. Проверьте корректность применения метода (если нужно, используйте статистический вывод), и определите качество решения.
8. Сформулируйте результаты исследования.
9. Приведите характеристику использованных программных средств.

Контрольные вопросы по итогам выполнения проекта

1. Опишите исходные данные проекта, целесообразность их фильтрации и (или) группировки.
2. Опишите первоначальные и окончательные цели и задачи исследования (их может быть несколько).
3. Сформулируйте результаты первичного анализа данных.
4. Аргументируйте выбор методов для дальнейшего исследования. Опишите их понятийный аппарат.
5. Опираясь на компьютерную диагностику, опишите этапы применения методов.
6. Проверьте корректность применения методов (если нужно, используйте статистический вывод), определите качество предложенных решений.
7. Структурируйте полученные результаты, сформулируйте выводы.
8. Проведите сравнительный анализ использованных подходов и методов.
9. Приведите комплексную характеристику использованных программных средств.

Контрольные вопросы по итогам изучения теоретического материала

1. Матрица данных, типы признаков (числовой, ординальный, номинальный), срезы данных (продольный, поперечный). Описательная статистика числового признака (среднее, дисперсия, стандарт). Центрирование и стандартизация данных. Мода, медиана, гистограмма, квартили, выброс.
2. Ковариация и корреляция. Понятия строгой, сильной, средней, слабой положительной и отрицательной корреляции и ее отсутствия. Свойства коэффициента ковариации и свойства коэффициента корреляции Пирсона для исходных, центрированных, стандартизованных признаков. Анализ диаграмм рассеивания. Ложная корреляция, выбросы, корреляция и нелинейная зависимость.
3. Статистика признаков. Вектор средних значений. Матрица ковариаций. Матрица корреляций. Матрица ковариаций двух наборов признаков (определения). Свойства матрицы ковариаций для

центрированных и стандартизованных признаков. Способы записи и вычисления матрицы ковариаций и корреляций.

4. Статистика случайных векторов. Математическое ожидание случайного вектора. Матрицы ковариаций и корреляций случайного вектора. Матрица ковариаций двух случайных векторов (определения). Запись матрицы ковариаций и корреляций через математические ожидания. Свойства матрицы ковариаций стандартизованного случайного вектора.
5. Теорема о дисперсии линейной комбинации независимых случайных величин. Теорема о линейных свойствах ковариации случайных величин. Теорема о дисперсии линейной комбинации двух признаков. Теорема о линейных свойствах ковариации признаков (формулировки). Ортогональные матрицы и их свойства (формулировки).
6. Понятие суммарной дисперсии признаков (для матрицы данных). Геометрические свойства суммарной дисперсии.
7. Задача кластерного анализа. Таблица принадлежности объектов кластерам. Понятия центра кластера, расстояния от объекта до кластера, расстояния между кластерами, расстояния от объекта до «своего кластера». Понятие показателя качества кластерного решения D .
8. Геометрические свойства показателя качества кластерного решения D .
9. Правило разложения дисперсии на внутригрупповую и междугрупповую дисперсии. Статистический смысл показателя D и его свойства.
10. Алгоритм и основное свойство метода K -средних. Принцип построения цепочки кластерных решений и выбора решения.
11. Метод иерархического кластерного анализа. Способы определения сходства объектов, сходства объекта и группы объектов, сходство двух групп объектов (способы выбора расстояния, способы выбора правила связи).
12. Алгоритм метода иерархического кластерного анализа. Схема объединения, расстояние объединения, функция объединения. Критерий выбора кластерного решения.
13. Сравнение методов K -средних и иерархического метода кластеризации.
14. Условия, цели и схема применения метода главных компонент (МГК). Критерии качества снижения размерности (доля объясненной суммарной дисперсии, общности, по лекциям). Матрица весовых коэффициентов главных компонент (ВКГК), смысл ее коэффициентов. Выбор названий главных компонент.
15. Исходные данные для МГК, свойства матрицы корреляций (R), используемые при построении метода, связь матриц R и Z , определение главных компонент. Определение матрицы значений главных компонент. Размерность, свойства и связь матриц C , Z , W .
16. Свойства главных компонент: утверждения о средних и дисперсиях ГК.
17. Суммарная дисперсия набора признаков и ее геометрический смысл. Утверждение о суммарной дисперсии ГК, его геометрический смысл.
18. Утверждение об отсутствии корреляции главных компонент. ПК-3
19. Утверждения о свойствах матрицы ВКГК (весовых коэффициентов главных компонент).
20. Определение общностей через коэффициенты матрицы ВКГК, их смысл и свойства.
21. Определение показателя δ через коэффициенты матрицы ВКГК, его смысл и свойства.
22. Выбор количества главных компонент при решении задач снижения размерности.
23. Модель факторного анализа в канонической и векторной записи. Основное свойство факторной модели. Воспроизведенная матрица ковариаций.
24. Общности, характерности и доля объясненной дисперсии. Теорема о ковариациях исходных признаков и общих факторов.
25. Ортогональные матрицы. Теорема о вращения факторов. Инварианты при вращении. Цели вращения факторов.
26. Факторный анализ: теорема о линейном преобразовании признаков.
27. Факторная модель для стандартизованных признаков в канонической и векторной записи. Основное свойство и его обоснование. Воспроизведенная матрица корреляций.

28. Общности, характерности и доля объясненной дисперсии (определения, свойства и их доказательства).
29. Теорема о корреляции признаков и общих факторов. Диаграммы факторных нагрузок. Принципы вращения факторов.
30. Модель для стандартизованных признаков. Задачи и схема исследования. Особенности применения модели. Критерии качества решения.
31. Геометрический смысл корреляции признаков в пространстве объектов и пространстве признаков. Геометрический смысл факторного анализа.
32. Проблемы поиска факторных нагрузок. Теорема о «диагональных» условиях нормировки.
33. Принцип главных факторов (теорема). Формулировки теорем о ранге матричного произведения и спектральном разложении.
34. Обоснование ограничений на число общих факторов.
35. Оценка факторных нагрузок методом главных факторов: постановка задачи, алгоритм и критерии остановки. Оценка числа общих факторов и начальная оценка общностей.
36. Оценка факторных значений: матрица данных как запись реализаций случайных векторов, задача на отыскание факторных значений, регрессионный способ оценки. Свойства оценочных матриц F , $V(F)$ и A . Факторные значения и вращение факторов.
37. Многомерная статистическая модель «Территория. Год. Показатели». Анализ структуры и динамики.
38. Модели поколения и населения. Показатели таблиц дожития.
39. Статистический анализ таблиц дожития.

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	Контрольные вопросы отвечены верно и в полном объеме, могут быть допущены незначительные неточности
не зачтено	Ответов на контрольные вопросы нет, ответы даны не в полном объеме или допущены существенные ошибки

6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература:

1. Гнеденко Борис Владимирович. Курс теории вероятностей : [учеб. для мех.-мат. специальностей ун-тов]. - 6-е изд., перераб. и доп. - М. : Наука, 1988. - 466, [1] с. : ил. - ISBN 5-02-013761-8 (в пер.) : 1.20., 483 экз.
2. Анализ данных : учебник / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. - Москва : Юрайт, 2023. - 490 с. - (Высшее образование). - ISBN 978-5-534-00616-2. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=847079&idb=0>.

Дополнительная литература:

1. Дубров Абрам Моисеевич. Многомерные статистические методы : учебник. - М. : Финансы и статистика, 1998. - 352 с. - 44.70., 2 экз.
2. Миркин Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. - Москва :

Юрайт, 2023. - 174 с. - (Высшее образование). - ISBN 978-5-9916-5009-0. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=843081&idb=0>.

3. Айвазян Сергей Артемьевич. Прикладная статистика. Основы эконометрики : учеб. для вузов : в 2 т. Т. 2. Основы эконометрики. - 2-е изд., испр. - М. : ЮНИТИ-ДАНА, 2001. - 432 с. - (Takis). - ISBN 5-238-00305-6 : 336.00., 12 экз.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

Операционная система Windows (лицензия), SPSS или Statistica (лицензия), Microsoft Office (лицензия), текстовый редактор Word, табличный редактор EXCEL, расширение "Анализ данных" в пакете EXCEL, базовый комплект и комплект "Анализ данных" профессиональных статистических пакетов Statistica или SPSS,

опционально программные средства анализа данных на базе Python, C++

Интернет-ресурсы:

1. Стронгина Н.Р. Методы моделирования социально-экономических процессов: анализ данных. ЭУК, учебно-методический комплекс. Фонд электронных образовательных ресурсов ННГУ. Н. Новгород, 2014. Идентификационный номер 814E.14.08. <https://e-learning.unn.ru/course/view.php?id=814> - Требуется регистрация.

2. Стронгина Н.Р. Методы моделирования социально-экономических процессов: регрессионный анализ. ЭУК, учебно-методический комплекс. Фонд электронных образовательных ресурсов ННГУ. Н. Новгород, 2014. Идентификационный номер 793E.14.08. <https://e-learning.unn.ru/course/view.php?id=793> - Требуется регистрация.

3. Российский статистический ежегодник. Росстат (ежегодное официальное издание). Сайт Федеральной службы государственной статистики. <https://rosstat.gov.ru/folder/210/document/12994>, http://www.gks.ru/bgd/regl/b16_13/Main.htm

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения, компьютерами, специализированным оборудованием: Учебная аудитория для проведения лекций, оснащенная партами, стульями, обязательно учебной доской. Оборудование для показа презентаций (персональный компьютер, проектор, экран). Терминал-класс с лицензионным программным обеспечением.

Учебная и научная литература, представленная в библиотечном фонде ННГУ, в электронных библиотеках и на кафедре ДУМЧА ИИТММ.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 01.03.02 - Прикладная математика и информатика.

Автор(ы): Стронгина Наталья Романовна, кандидат физико-математических наук, доцент.

Заведующий кафедрой: Калинин Алексей Вячеславович, доктор физико-математических наук.

Программа одобрена на заседании методической комиссии от 13.12.2023, протокол № 3.