

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

(факультет / институт / филиал)

УТВЕРЖДЕНО
президиумом Ученого совета ННГУ
протокол от
«14» декабря 2021 г. № 4

Рабочая программа дисциплины

Методы моделирования и прикладной анализ данных

(наименование дисциплины (модуля))

Уровень высшего образования

Бакалавриат

(бакалавриат / магистратура / специалитет)

Направление подготовки / специальность

02.03.02 Фундаментальная информатика и информационные технологии

(указывается код и наименование направления подготовки / специальности)

Направленность образовательной программы

Инженерия программного обеспечения

(указывается профиль / магистерская программа / специализация)

Форма обучения

очная

(очная / очно-заочная / заочная)

Нижний Новгород

2022 год

1. Место дисциплины в структуре ОПОП

№ варианта	Место дисциплины в учебном плане образовательной программы	Стандартный текст для автоматического заполнения в конструкторе РПД
2	Блок 1. Дисциплины (модули) Часть формируемая участниками образовательных отношений	Дисциплина <i>Б1.В.ДВ.05.02 Методы моделирования и прикладной анализ данных</i> относится к части ООП направления подготовки 02.03.02 Фундаментальная информатика и информационные технологии, формируемой участниками образовательных отношений.

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**	
ПК-3 Способен создавать и исследовать новые математические модели в естественных науках, промышленности и бизнесе, с учетом возможностей современных информационных технологий и программирования и компьютерной техники	ПК-3.1: Знает методы анализа и исследования математических моделей в области фундаментальной информатики и информационных технологий;	ЗНАТЬ 31 (ПК-3) <i>Постановки модельных задач дисциплины «Методы моделирования и прикладной анализ данных», понятийный аппарат и утверждения. Основные приемы и формулы, подходы к обоснованию методов, основы статистического вывода:</i> – источники данных; типы данных; нормативно-правовые аспекты их получения и обработки; – теоретические основы разведочного анализа и методов снижения размерности: кластерного анализа; метода главных компонент, факторного и регрессионного анализа; диагностику применения методов; – элементы статистического вывода; их обоснование; – практику применения методов в социально-экономических	<i>Собеседование, задача, тестирование</i>

		приложениях.	
	ПК-3.2: Умеет определять ключевые свойства и ограничения системы	<p>УМЕТЬ</p> <p><i>Формализовать постановку задачи и ставить цели исследования</i></p> <p><i>Применять методы анализа данных в экономических (социальных) приложениях; формулировать результаты анализа; предлагать их экономическую (социальную) интерпретацию;</i></p> <p><i>Обосновать возможность применения метода, его теоретические свойства</i></p> <p><i>Пользоваться навыками самостоятельного выполнения проектов с использованием современных пакетов статистических программ, включая выбор постановки задачи, выбор метода, анализ и интерпретацию результатов.</i></p> <p><i>Пользоваться навыками оценки программных средств (социально-экономические приложения), постановки задач развития программных средств, направлений их разработки</i></p>	<i>Собеседование, задача, тестирование</i>

3. Структура и содержание дисциплины

3.1. Трудоемкость дисциплины

	очная форма обучения
Общая трудоемкость	3 ЗЕТ
Часов по учебному плану	108
в том числе	
контактная работа:	33
- занятия лекционного типа	32
- занятия лабораторного типа	
- текущий контроль (КСР)	1
самостоятельная работа	75
Промежуточная аттестация – зачет	

3.2. Содержание дисциплины

Наименование и краткое содержание разделов и тем дисциплины	Всего (часы)	В том числе				
		Контактная работа (работа во взаимодействии с преподавателем), часы. Из них				Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа	Занятия лабораторного типа	Всего	
	Очная	Очная	Очная	Очная	Очная	Очная
Тема 1. Примеры моделирования в экономике и социальной сфере. Постановки управленческих задач. Источники и типы данных. Нормативно-правовые аспекты сбора данных. Интерпретация результатов и выводы для управленческих решений.	14	4			4	10
Тема 2. Методы разведочного, корреляционного и дисперсионного анализа.	14	4			4	10
Тема 3. Методы кластерного анализа. Теория и приложения.	14	4			4	10
Тема 4. Метод главных компонент. Теория и приложения.	23	8			8	15
Тема 5. Введение в регрессионный и факторный анализ.	14	4			4	10
Тема 6. Основы статистического вывода для методов п. 2 – п. 5.	14	4			4	10
Тема 7. Применение методов для анализа модельных задач	14	4			4	10
Текущий контроль (КСР)	1				1	
Промежуточная аттестация –зачет						
Итого	108	32			33	75

Текущий контроль успеваемости реализуется в формах опросов на занятиях лекционного типа. Промежуточная аттестация проходит в традиционной форме (зачет).

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

- решение задач и выполнение практических заданий средствами Excel
- выполнение практических заданий в профессиональных пакетах Statistica, SPSS
- выполнение проектов в профессиональных пакетах Statistica, SPSS
- подготовка презентаций и обсуждение результатов проектов.

Виды самостоятельной работы студентов

- ❖ Выполнение практических заданий (Практикум)
- ❖ Выполнение проектов (тематика в разделах Практикума)
- ❖ Подготовка к собеседованию (зачету)
- ❖ Тестирование

Образовательные материалы для самостоятельной работы студентов, практические задания для проведения текущего контроля

1. Балабанов А.С., Стронгина Н.Р. Анализ данных в экономических приложениях. Компьютерный практикум в SPSS. Н. Новгород: Изд-во ННГУ, 2002. – 100 с.
2. Стронгина Н.Р., Марчева И.А. Регрессионный анализ в экономических приложениях: Учебное пособие. Н. Новгород: Изд-во ННГУ, 2002. – 214 с.
3. 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
4. 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.

Контрольные вопросы и задания для проведения текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведены в п. 5.2.

5. Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю), включающий:

5.1. Описание шкал оценивания результатов обучения по дисциплине

Уровень сформированности компетенций (индикатора достижения компетенций)	Шкала оценивания сформированности 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	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 незначительных ошибок	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок.	Уровень знаний в объеме, превышающем программу подготовки.

	я от ответа						
<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.	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме.	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения, решены все основные задачи с отдельными несущественным недочетами, выполнены все задания в полном объеме.	Продemonстрированы все основные умения, решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие владения материалом. 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 некоторыми недочетами.	Продemonстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов.	Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов.	Продemonстрирован творческий подход к решению нестандартных задач.

Шкала оценки при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	Превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно»
	Отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично», при этом хотя бы одна компетенция сформирована на уровне «отлично»
	Очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо», при этом хотя бы одна компетенция сформирована на уровне «очень хорошо»
	Хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо», при этом хотя бы одна компетенция сформирована на уровне «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», ни одна из компетенций не сформирована на уровне «плохо»
	Плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

5.2. Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения

5.2.1 Контрольные вопросы

Вопрос	Код компетенции (согласно РПД)
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.	ПК-3
2. Ковариация и корреляция. Понятия строгой, сильной, средней, слабой положительной и отрицательной корреляции и ее отсутствия. <u>Свойства коэффициента ковариации и свойства коэффициента корреляции Пирсона для исходных, центрированных, стандартизованных.</u> Анализ диаграмм рассеивания. Ложная корреляция, выбросы, корреляция и нелинейная зависимость.	ПК-3
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.	ПК-3
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.	ПК-3
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).	ПК-3
6. Понятие суммарной дисперсии признаков (для матрицы данных). <u>Геометрические свойства суммарной дисперсии.</u>	ПК-3
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». <i>Понятие показателя качества кластерного решения D.</i>	ПК-3
8. <u>Геометрические свойства показателя качества кластерного решения D.</u>	ПК-3
9. <u>Правило разложения дисперсии на внутригрупповую и междугрупповую. Статистический смысл показателя D и его свойства.</u>	ПК-3

Вопрос	Код компетенции (согласно РПД)
10.Алгоритм и основное свойство метода К-средних. Принцип построения цепочки кластерных решений и <i>выбора решения</i> .	ПК-3
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).	ПК-3
12.Алгоритм метода иерархического кластерного анализа. Схема объединения, расстояние объединения, функция объединения. Критерий выбора кластерного решения.	ПК-3
13.Сравнение методов К-средних и иерархического метода кластеризации.	ПК-3
14.Условия, цели и схема применения метода главных компонент (МГК). Критерии качества снижения размерности (доля объясненной суммарной дисперсии, общности, <i>по лекциям</i>). Матрица весовых коэффициентов главных компонент (ВКГК), смысл ее коэффициентов. Выбор названий главных компонент.	ПК-3
15.Исходные данные для МГК, свойства матрицы корреляций (R), используемые при построении метода, <u>связь матриц R и Z</u> , определение главных компонент. Определение матрицы значений главных компонент. <u>Размерность, свойства и связь матриц C, Z, W</u> .	ПК-3
16.Свойства главных компонент: <u>утверждения о средних и дисперсиях ГК</u> .	ПК-3
17.Суммарная дисперсия набора признаков и ее геометрический смысл. <u>Утверждение о суммарной дисперсии ГК</u> , его геометрический смысл.	ПК-3
18. <u>Утверждение об отсутствии корреляции главных компонент</u> .	ПК-3
19. <u>Утверждения о свойствах матрицы ВКГК</u> (весовых коэффициентов главных компонент).	ПК-3
20.Определение общностей через коэффициенты матрицы ВКГК, <u>их смысл и свойства</u> (в т.ч. связь с определением из лекций).	ПК-3
21.Определение показателя δ через коэффициенты матрицы ВКГК, <u>его смысл и свойства</u> (в т.ч. связь с определением из лекций).	ПК-3
22.Выбор количества главных компонент при решении задач снижения размерности.	ПК-3
23.Модель факторного анализа в канонической и векторной записи. <u>Основное свойство факторной модели</u> . Воспроизведенная матрица ковариаций.	ПК-3
24.Общности, характерности и доля объясненной дисперсии. <u>Теорема о ковариациях исходных признаков и общих факторов</u> .	ПК-3

Вопрос	Код компетенции (согласно РПД)
25.Ортогональные матрицы. <u>Теорема о вращения факторов</u> . Инварианты при вращении. Цели вращения факторов.	ПК-3
26.Факторный анализ: <u>теорема о линейном преобразовании признаков</u> . Модели для стандартизованных признаков.	ПК-3
27.Факторная модель для стандартизованных признаков в канонической и векторной записи. <u>Основное свойство и его обоснование</u> . Воспроизведенная матрица корреляций.	ПК-3
28. <u>Общности, характерности и доля объясненной дисперсии</u> (определения, свойства и их доказательства).	ПК-3
29. <u>Теорема о корреляции признаков и общих факторов</u> . Диаграммы факторных нагрузок. Принципы вращения факторов.	ПК-3
30.Модель для стандартизованных признаков. Задачи и схема исследования. Особенности применения модели. <i>Критерии качества решения</i> .	ПК-3
31.Геометрический смысл корреляции признаков в пространстве объектов и пространстве признаков. <i>Геометрический смысл факторного анализа</i> .	ПК-3
32.Проблемы поиска факторных нагрузок. <u>Теорема о «диагональных» условиях нормировки</u> .	ПК-3
33. <u>Принцип главных факторов (теорема)</u> . Формулировки теорем о ранге матричного произведения и спектральном разложении.	ПК-3
34.Обоснование ограничений на число общих факторов.	ПК-3
35. <u>Оценка факторных нагрузок методом главных факторов</u> : постановка задачи, алгоритм и критерии останковки. Оценка числа общих факторов и начальная оценка общностей.	ПК-3
36. <u>Оценка факторных значений</u> : матрица данных как запись реализаций случайных векторов, задача на отыскание факторных значений, регрессионный способ оценки. <u>Свойства матриц</u> \hat{F} , $V(\hat{F})$ и \hat{A} . Факторные значения и вращение факторов.	ПК-3

5.2.2. Типовые тестовые задания для оценки сформированности компетенции ПК-3

Вопрос 1

Тип вопроса: ввод значения

Формулировка вопроса:

Файл данных *демография_95.sav*. Компьютерный пакет SPSS.

Презентация «Метод К-средних».

(формат «целое число»)

Ответ 1: число исследуемых признаков

Верное значение: 4

Ответ 1: 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 *демография_95.sav*. Компьютерный пакет SPSS.

Презентация «Метод К-средних».

Кластеризация будет проведена

Варианты ответов:

- для исходных признаков
- для исходных стандартизованных признаков

5.2.3. Типовые задачи для оценки сформированности компетенции ПК-3

Основные понятия и свойства. Разведочный анализ

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, квантили, 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. Анализ диаграмм рассеивания. Ложная корреляция, выбросы, корреляция и нелинейная зависимость.

Кластерный анализ

3. Понятие суммарной дисперсии признаков (для матрицы данных). Геометрические свойства суммарной дисперсии.
4. 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.*

❖ **Практическое задание (Практикум) для оценивания результатов обучения в виде умений.**

Стартовый практикум. Первичный анализ данных, анализ структуры и динамики (в EXEL)

Практикум 1. Первичный анализ данных. Метод К-средних (в SPSS®)

Практикум 2. Метод главных компонент (в SPSS®)

Практикум 3. Факторный анализ (в SPSS®)

Дополнительный практикум. Применение метода главных компонент для анализа финансовых данных (в SPSS®)

❖ Типовые задания для оценивания результатов обучения в виде умений

Сформулируйте определения и докажите утверждения:

Основные понятия и свойства. Разведочный анализ

1. Свойства коэффициента ковариации и свойства коэффициента корреляции Пирсона для исходных, центрированных, стандартизованных.

Кластерный анализ

2. Геометрические свойства суммарной дисперсии.
3. Геометрические свойства показателя качества кластерного решения D.
4. Правило разложения дисперсии на внутригрупповую и междугрупповую.

6. Учебно-методическое и информационное обеспечение дисциплины

Основная литература

1. 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.
2. 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
3. Российский статистический ежегодник. Росстат (ежегодное официальное издание). http://www.gks.ru/bgd/regl/b16_13/Main.htm, сайт Федеральной службы государственной статистики

Дополнительная литература

4. Стронгина Н. Р., Марчева И. А. - Регрессионный анализ в экономических приложениях: учеб. пособие. - Н. Новгород: Изд-во Нижегород. гос. ун-та, 2003. - 214 с. (5 экз.)

7. Материально-техническое обеспечение дисциплины

Помещения представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой, оснащенные оборудованием и техническими средствами обучения: операционная система Windows (лицензия), Microsoft Visual Studio (лицензия), SPSS или Statistica (лицензия), Microsoft Office (лицензия).

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 02.03.02
Фундаментальная информатика и информационные технологии.

Автор: к.ф.-м.н., доцент Стронгина Н.Р.

Рецензент (ы) _____

Заведующий кафедрой _____ Калинин А.В.