

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

Наглядный вероятностно-статистический анализ данных

Уровень высшего образования

Бакалавриат

Направление подготовки / специальность

09.03.04 - Программная инженерия

Направленность образовательной программы

Разработка программно-информационных систем

Форма обучения

очная

г. Нижний Новгород

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.01.01 Наглядный вероятностно-статистический 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-12: Способен осуществлять концептуальное, функциональное и логическое проектирование систем и подсистем малого и среднего масштаба и сложности.	ПК-12.1: Знает методы концептуального, функционального и логического проектирования ПК-12.2: Знает типовые архитектурные стили и паттерны проектирования ПК-12.3: Умеет применять архитектурные стили и паттерны проектирования при решении типовых задач ПК-12.4: Владеет навыками разработки проекта программной системы с учетом возможностей и ограничений	ПК-12.1: Умеет осуществлять автоматизированный сбор данных. Знает требования, необходимые для выполнения при подготовке данных к анализу. ПК-12.2: Знает способы проведения разведывательного анализа данных. Умеет выявлять зависимости в анализируемых данных ПК-12.3: Знает приемы и возможности визуализации данных. ПК-12.4: Умеет находить оценки параметров распределения данных. Владеет навыками проверки качества построенных оценок.	Задачи	Зачёт: Контрольные вопросы

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
--	-------

Общая трудоемкость, з.е.	3
Часов по учебному плану	108
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	24
- занятия семинарского типа (практические занятия / лабораторные работы)	12
- КСР	1
самостоятельная работа	71
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			
		Контактная работа (работа во взаимодействии с преподавателем), часы из них			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа (практические занятия/ лабора- торные работы), часы	Всего	
	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0
Роль теории вероятностей и математической статистики при анализе данных	4	2	0	2	2
Типы статистических данных (числовые, ординальные, номинальные). Генеральная совокупность, выборка. Репрезентативность выборки.	8	3	0	3	5
Одномерные и многомерные данные. Законы распределения.	15	3	2	5	10
Числовые характеристики одномерных случайных величин.	11	2	1	3	8
Числовые характеристики многомерной случайной величины.	11	2	1	3	8
Регрессия, подгонка прямой под облако точек.	15	3	2	5	10
Кластеризация данных, задача и методы решения.	13	3	2	5	8
Оценивание неизвестных параметров распределения.	15	3	2	5	10
Критерий согласия и их применение.	15	3	2	5	10
Аттестация	0				
КСР	1				1
Итого	108	24	12	37	71

Содержание разделов и тем дисциплины

Роль теории вероятностей и математической статистики при анализе данных, возможности языка Python и приложения Excel для анализа.

Типы статистических данных (числовые, ординальные, номинальные). Генеральная совокупность, выборка. Репрезентативность выборки. Способы обеспечения сопоставимости данных. Обзор

существующих открытых библиотек данных.

Одномерные и многомерные данные. Понятие одномерной случайной величины, дискретные и непрерывные одномерные случайные величины. Понятие многомерной случайной величины.

Эмпирическая плотность и эмпирическая функция распределения, примеры их построения.

Визуализация эмпирических распределений: построение гистограмм, графиков функций распределения.

Числовые характеристики одномерных случайных величин. Статистические числовые характеристики.

Анализ данных на основе статистических числовых характеристик: характеристик центрального положения, разброса. Оценка функции распределения на основе квантилей.

Числовые характеристики многомерной случайной величины. Ковариация, коэффициент корреляции.

Построение ковариационной матрицы. Корреляционный анализ. Построение диаграмм рассеивания.

Выявление зависимости между величинами на основе ковариационной матрицы.

Регрессия, подгонка прямой под облако точек. Простая линейная регрессия. Многомерная регрессия.

Полиномиальная регрессия. Построение простой и многомерной линейной регрессии.

Кластеризация данных, задача и методы решения (метод k -средних, иерархическая кластеризация).

Определение оптимального количества кластеров.

Оценивание неизвестных параметров распределения. Точечные оценки. Параметрическое и непараметрическое оценивание. Метод максимального правдоподобия и метод моментов.

Простые и сложные гипотезы. Ошибки первого и второго рода. Критерий согласия и их применение.

Обзор готовых средств для проверки гипотез.

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 "Теория вероятностей и математическая статистика ДО" (<https://e-learning.unn.ru/course/view.php?id=789>).

Иные учебно-методические материалы: Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и задачам для текущего контроля и промежуточной аттестации по итогам освоения дисциплины, приведенным в п. 5.

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 - Задачи) для оценки сформированности компетенции ПК-12:

Задача 1. Исследуется величина X - вес клетчатки (гр), который ежедневно потребляет человек. В результате интернет-опроса 30 респондентов была получена исходная выборка.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
100	80	240	230	180	160	150	210	250	230
x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
260	215	270	175	145	190	320	300	390	235
x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{30}
160	195	180	215	220	210	60	120	130	250

По выборке необходимо найти основные статистические числовые характеристики (статистические математическое ожидание, дисперсию, среднее квадратическое отклонение).

Задача 2. Получены многомерные данные, представленные в таблице ниже:

Показатель	2013	2014	2015	2016	2017	2018
Численность безработных (по методологии МОТ), тыс. человек	139,8	75,2	75,2	76,3	75,2	73,1
Численность пенсионеров, тыс. человек	1006,3	1031,6	1040,9	1045,6	1050,2	1054,1
Среднедушевые денежные доходы населения в месяц, руб.	16477,3	27048,6	30003,5	30057,2	30325,7	31408,0

По представленным данным необходимо построить множественную линейную регрессию. В качестве зависимой переменной следует выбрать «Численность безработных», а в качестве независимых «Численность пенсионеров» и «Среднедушевые денежные доходы населения».

Критерии оценивания (оценочное средство - Задачи)

Оценка	Критерии оценивания
зачтено	Верное решение задачи
не зачтено	Ошибки в решении задачи или отсутствие решения

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

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено			зачтено			

(индикатор достижения)							
<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	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 несущественными 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 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 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 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-12

1. Перечислить способы представления выборочных значений
2. Перечислить типы шкал измерений и соответствующие типы статистических данных
3. Дать определение одномерных и многомерных данных.
4. Дать определение статистической (выборочной, эмпирической) функции распределения.
5. Дать определение статистической (выборочной, эмпирической) плотностью вероятностей и гистограммы
6. Определить, основные статистические числовые характеристик (выборочное среднее, дисперсию и стандарт)
7. Привести формулы для вычисления выборочный начального и центрального моментов k-го порядка, статистической медианы
8. Привести способы вычисления статистической ковариации и выборочного коэффициента корреляции, определить понятие несмещенной выборочной ковариации.
9. Перечислить способы выявления статистическая зависимости двух случайных величин
10. Дать определения понятию регрессии двух случайных величин
11. Определить простую линейную регрессию. Перечислить Ключевые различия между корреляцией и линейной регрессией.
12. Описать метод использования регрессии для прогнозирования. Определить, что такое подогнанные значения и остатки.

13. Дать определение множественной линейной регрессии.
14. Определить понятие нелинейной регрессии, указать, на какие классы она подразделяется.
15. Перечислить способы диагностики качества регрессионной модели
16. Дать определения понятия точечного оценивания неизвестных параметров распределения. Перечислить критерии качества статистических оценок
17. Описать Метод моментов
18. Описать Метод максимального правдоподобия
19. Описать метод интервального оценивания неизвестных параметров распределения. Определить понятие доверительного интервала
20. Описать методологию построения доверительных интервалов для математического ожидания
21. Описать методологию построения доверительных интервалов для неизвестной дисперсии
22. Описать методологию построения доверительных интервалов для неизвестной вероятности события
23. Дать определения понятию статистической гипотезы
24. Перечислить основные принципы построения критериев согласия
25. Описать метод проверки простых гипотез с помощью критерия согласия Колмогорова
26. Описать метод проверки простых гипотез о виде распределения с помощью критерия согласия хи-квадрат Пирсона

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	Полный ответ на контрольный вопрос
не зачтено	Неполный ответ на контрольный вопрос или отсутствие ответа

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

Основная литература:

1. Федоткин Михаил Андреевич. Основы прикладной теории вероятностей и статистики : учеб. для студентов вузов, обучающихся по специальности "Прикладная математика и информатика" и по направлению "Прикладная математика и информатика". - М. : Высшая школа, 2006. - 368 с. :

ил. - ISBN 5-06-005328-8 : 215.60., 183 экз.

2. Федоткин Михаил Андреевич. Модели в теории вероятностей : учебник. - М. : Физматлит : ННГУ, 2012. - 608 с. - (Библиотека Нижегородского государственного университета им. Н. И. Лобачевского). - ISBN 978-5-9221-1384-7 : 600.00., 200 экз.

3. Пройдакова Е. В. Задания для самоконтроля по теме «Многомерные случайные величины» : учебно-методическое пособие / Пройдакова Е. В., Бородина Т. С. - Нижний Новгород : ННГУ им. Н. И. Лобачевского, 2021. - 86 с. - Рекомендовано методической комиссией Института информационных технологий, математики и механики для студентов ННГУ, обучающихся по направлениям подготовки 09.03.04 «Программная инженерия», 01.03.02 «Прикладная математика и информатика», 09.03.03 «Прикладная информатика» и 02.03.02 «Фундаментальная информатика и информационные технологии». - Книга из коллекции ННГУ им. Н. И. Лобачевского - Математика., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=830094&idb=0>.

Дополнительная литература:

1. Белько Иван Васильевич. Высшая математика для экономистов : 3 семестр : Теория вероятностей и математическая статистика. - М. : Новое знание, 2002. - 144 с. - (Экспресс-курс). - Лит.: с. 140. - ISBN 5-94735-015-7 : 62-00., 2 экз.

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

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения.

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

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения, компьютерами.

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 09.03.04 - Программная инженерия.

Автор(ы): Пройдакова Екатерина Вадимовна, кандидат физико-математических наук, доцент.

Заведующий кафедрой: Зорин Андрей Владимирович, доктор физико-математических наук.

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