

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

Анализ данных и имитационное моделирование в Python

Уровень высшего образования

Магистратура

Направление подготовки / специальность

01.04.02 - Прикладная математика и информатика

Направленность образовательной программы

Анализ данных в прикладных областях

Форма обучения

очная

г. Нижний Новгород

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.03.01 Анализ данных и имитационное моделирование в Python 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-11: Способен разрабатывать и анализировать концептуальные и теоретические модели решаемых задач производственно-технологической деятельности	<p>ПК-11.1: Знает методы разработки и анализа концептуальных и теоретических моделей решаемых производственно-технологических задач</p> <p>ПК-11.2: Умеет применять методы разработки и анализа концептуальных и теоретических моделей решаемых производственно-технологических задач</p> <p>ПК-11.3: Имеет навыки применения методов разработки и анализа концептуальных и теоретических моделей решаемых производственно-технологических задач</p>	<p>ПК-11.1: Знать математические основы концепции «Шесть сигм» при решении производственно-технологических задач</p> <p>ПК-11.2: Уметь проводить анализ данных при сравнении ключевых показателей производственно-технологической деятельности</p> <p>ПК-11.3: Иметь навыки разработки и применения программного обеспечения при анализе данных и постановке производственно-технологических задач с использованием концепции «Шесть сигм»</p>	Задания Собеседование	Зачёт: Контрольные вопросы
ПК-4: Способен разрабатывать и анализировать концептуальные и теоретические модели решаемых научных проблем и задач	<p>ПК-4.1: Знает методы разработки и анализа концептуальных и теоретических моделей решаемых научных проблем и задач</p> <p>ПК-4.2: Умеет применять методы разработки и анализа концептуальных и теоретических моделей</p>	<p>ПК-4.1: Знать: -базовые понятия теории вероятностей и математической статистики; -законы распределения вероятностей, наиболее распространенных на практике непрерывных</p>	Задания Собеседование	Зачёт: Контрольные вопросы

	<p>решаемых научных проблем и задач</p> <p>ПК-4.3: Имеет навыки применения методов разработки и анализа концептуальных и теоретических моделей решаемых научных проблем и задач</p>	<p>случайных величин; - постановку задачи квадратичного программирования; -ансамблевые методы в статистике и обучении машин</p> <p>ПК-4.2: Уметь проводить анализ данных и моделей согласование данных, строить модели на основе ансамблевых методов</p> <p>ПК-4.3: Иметь навыки разработки моделей согласования данных и применения ансамблевых методов.</p>		
--	---	---	--	--

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	16
- занятия семинарского типа (практические занятия / лабораторные работы)	16
- КСР	1
самостоятельная работа	75
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	Всего	

			занятия/лабораторные работы), часы		
	ОФ	ОФ	ОФ	ОФ	ОФ
Тема 1. Компьютерное моделирование случайных величин с различными законами распределения на языке Python	6	2	2	4	2
Тема 2. Анализ данных при сравнении ключевых показателей деятельности	8	2	2	4	4
Тема 3. Математические аспекты концепции «Шесть сигм»	32	4	4	8	24
Тема 4. Согласование данных и интеллектуальное использование данных процесса	28	4	4	8	20
Тема 5. Деревья классификации и регрессии. Ансамблевые методы в статистике и обучении машин	33	4	4	8	25
Аттестация	0				
КСР	1			1	
Итого	108	16	16	33	75

Содержание разделов и тем дисциплины

Тема 1. Компьютерное моделирование случайных величин с различными законами распределения на языке Python: равномерное распределение, нормальный закон распределения, логнормальное распределение, распределение Стюдента, распределение Хи-квадрат. Композиция законов распределения двух непрерывных случайных величин. Назначение модели, приемы моделирования. Библиотека SciPy. Визуализация данных, библиотека Matplotlib. Гистограммы. Имитационное моделирование. Метод статистических испытаний (Метод Монте-Карло), примеры его применения.

Тема 2. Анализ данных при сравнении ключевых показателей деятельности. Коллекции Series и DataFrame библиотеки Pandas. Генерация описательной статистики. Получение сведений о датафрейме. Экспорт и импорт данных. Отсутствующие данные. Агрегирование данных.

Квантили, квартили. Реализация в Python. Метод Джона Тьюки построения пяти числовой сводки, эксклюзивный подход. Инклюзивные квартили. Диаграмма «Ящик с усами», примеры использования при сравнении данных. Анализ данных и бенчмаркинг. Нормирование показателей. Упорядочивание данных, позиция в квартилях. Анализ динамики изменения данных. Практические примеры использования квартилей в научном и производственном анализе.

Тема 3. Математические аспекты концепции «Шесть сигм». Стандартное нормальное распределение. Уровень сигм и качество. Компьютерное моделирование в бизнес-процессах. Распределение Парето. Диаграмма Парето в концепции «Шесть сигм». Метод DMAIC (define, measure, analyze, improve, control — определение, измерение, анализ, совершенствование, контроль). Корреляционный и регрессионный анализ при реализации проектов «Шесть сигм». Парадоксы регрессии. Квартет Энскомба. График нормальной вероятности (Normal probability plot). Библиотека Seaborn. Практические примеры использования концепции «Шесть сигм» в ведущих Компаниях мира.

Тема 4. Согласование данных и интеллектуальное использование данных процесса. Анализ данных, поступающих из систем измерения. Базовые понятия технологии согласования данных. Необходимое число систем измерения в модели баланса. Согласование данных материального баланса. Математическая постановка задачи согласования данных. Квадратичное программирование. Имитационные модели согласования данных. Статистический глобальный тест определения грубых ошибок измерения (The Global test). Статистический тест измерений (The Measurement Test). Качество баланса (Balance Quality). Порядок действий при анализе данных измерений. Современные IT-системы согласования данных. Интеграция систем различного уровня управления в IT-системах (DCS, PHD, MES, ERP).

Тема 5. Деревья классификации и регрессии. Ансамблевые методы в статистике и обучении машин. Представление в виде набора логических правил деревьев классификации и регрессии. Представление в виде дерева. Узлы, родители и потомки, конечные узлы. Пороговые значения. Меры чистоты узла (impurity measures): критерий Джини, энтропия, ошибки классификации. Правила остановки обучения дерева. Информативность переменных. Мудрость толпы. Бэггинг. Ключевые параметры модели. Out-of-bag error. Метод случайного леса. Бустинг.

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.

Основная литература:

1. Федоткин М.А. Основы прикладной теории вероятностей и статистики. — М.: Высшая школа. 2006. - 368 с. (90 экз.)
2. Федоткин М.А. Модели в теории вероятностей. — М.: ФИЗМАТЛИТ, 2012. – 608с. (196 экз.)
3. Федоткин М.А. Лекции по анализу случайных явлений. — М.: ФИЗМАТЛИТ, 2016. – 404 с. (166 экз.)
4. Зорин А. В., Зорин В. А., Федоткин М. А. Моделирование случайных величин и проверка гипотез о виде распределения: Учебно-методическое пособие. — Нижний Новгород: Нижегородский госуниверситет, 2017. — 19 с. Фонд электронных образовательных ресурсов ННГУ, электронный адрес <http://www.unn.ru/books/resources.html> (регистр. номер 1640.17.06)

Дополнительная литература:

1. Miller Curtis. Training Systems Using Python Statistical Modeling: Explore Popular Techniques for Modeling Your Data in Python. - Packt Publishing, 2019. - ISBN 9781838820640. - ISBN 9781838823733, электронный адрес: <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=854298&idb=0> (1 online resource)
2. Майкл Л. Дж. Бережливое производство + шесть сигм» в сфере услуг. — М.: Альпина Бизнес Букс, 2005. — 402 с. (1 экз.)
3. Molin Stefanie, Jee Ken. Hands-On Data Analysis with Pandas : A Python Data Science Handbook for Data Collection, Wrangling, Analysis, and Visualization. - Packt Publishing, 2021. - 1 online resource. - ISBN 9781800565913. - ISBN 9781800563452, электронный адрес: <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=854289&idb=0> (1 online resource)

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 - Задания) для оценки сформированности компетенции ПК-11:

Задача 1. Проиллюстрировать работу алгоритма согласования данных на примере с одним узлом. На вход поступило 2 потока, измеренная масса за балансовый период составила $y_1 = 150$ т и $y_2 = 40$ т. Измеренное значение на выходе $y_3 = 200$ т. Для согласованных значений x_i выполняется закон сохранения масс. Измерения равноточные. Вычислить согласованные значения с использованием Python. Провести проверку на наличие грубых ошибок измерения с использованием глобального теста

Задача 2. Построить график нормальной вероятности (Normal probability plot), используя библиотеки Python, для выборки из нормального распределения и для выборки из равномерного распределения.

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

Задача 1. Используя Python для нормального распределения с параметрами математическое ожидание $\mu = 1,5$, среднее квадратическое отклонение $\sigma = 1$ построить график плотности распределения и вывести на печать таблицу для различного уровня сигм.

Задача 2. Напишите программу, которая позволяет построить графики функции и плотности распределения вероятностей для распределения Парето с различными значениями параметров.

Критерии оценивания (оценочное средство - Задания)

Оценка	Критерии оценивания
зачтено	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно»
не зачтено	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно»

5.1.3 Типовые задания (оценочное средство - Собеседование) для оценки сформированности компетенции ПК-11:

1. Что такое инклюзивные квартили?
2. Что представляет собой график нормальной вероятности (Normal probability plot) в Python?
3. Какой подход Вы предложите к анализу производственных проблем с использованием концепции «Шесть сигм»?

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

1. Стандартное нормальное распределение и правило трех сигм.
2. Квантили, квартили. Сколько всего квартилей?
3. Математическая постановка задачи согласования данных.

Критерии оценивания (оценочное средство - Собеседование)

Оценка	Критерии оценивания
зачтено	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно»
не зачтено	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно»

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

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатора 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	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	При решении стандартных задач не продемонстрир	Имеется минимальный набор навыков	Продemonстрированы базовые навыки при	Продemonстрированы базовые навыки при	Продemonстрированы навыки при	Продemonстрирован творческий подход к

	оценить наличие навыков вследствие отказа обучающегося от ответа	ованы базовые навыки. Имели место грубые ошибки	для решения стандартных задач с некоторым и недочетами	решении стандартных задач с некоторым и недочетами	решении стандартных задач без ошибок и недочетов	решении нестандартных задач без ошибок и недочетов	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-11

1. Коллекции Series и DataFrame библиотеки Pandas. Генерация описательной статистики. Получение сведений о датафрейме. Экспорт и импорт данных. Отсутствующие данные. Агрегирование данных.

2. Порядок действий при анализе данных измерений.

3. Квантили, квартили. Реализация в Python. Диаграмма «Ящик с усами», примеры использования при сравнении данных.

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

1. Компьютерное моделирование случайных величин с различными законами распределения на языке Python: нормальный закон распределения, логнормальное распределение, распределение Стьюдента, распределение Хи-квадрат.
2. Математическая постановка задачи согласования данных.
3. Тесты определения грубых ошибок измерения.

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно»
не зачтено	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно»

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. Федоткин Михаил Андреевич. Лекции по анализу случайных явлений : учеб. пособие для студентов вузов, обучающихся по направлениям ВПО 010400 "Приклад. математика и информатика" и 010300 "Фундам. информатика и информ. технологии" / ННГУ. - М. : Физматлит, 2016. - 464 с. - ISBN 978-5-9221-1679-4 : 599.50., 250 экз.
4. Зорин Андрей Владимирович. Моделирование случайных величин и проверка гипотез о виде распределения : учебно-методическое пособие / А. В. Зорин, В. А. Зорин, М. А. Федоткин ; ННГУ им. Н. И. Лобачевского. - Нижний Новгород : Изд-во ННГУ, 2017. - 20 с. - Текст : электронный., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=823611&idb=0>.

Дополнительная литература:

1. Curtis Miller. Training Systems Using Python Statistical Modeling : Explore Popular Techniques for Modeling Your Data in Python. - Packt Publishing, 2019. - 1 online resource. - ISBN 9781838820640. - ISBN 9781838823733. - Текст : электронный., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=854298&idb=0>.
2. Джордж Майкл. Бережливое производство + шесть сигм в сфере услуг = Lean Six Sigma For Service : Как скорость бережливого производства и качество шести сигм помогают

совершенствованию бизнеса / пер. с англ. [Т. Гутман ; науч. ред. С. Турко ; ред. С. Огарева]. - М. : Манн, Иванов и Фербер, 2011. - 496 с. - (Библиотека Сбербанка). - ISBN 978-5-91657-250-6 (рус.) : 300.00., 1 экз.

3. Stefanie Molin. Hands-On Data Analysis with Pandas : A Python Data Science Handbook for Data Collection, Wrangling, Analysis, and Visualization. - Packt Publishing, 2021. - 1 online resource. - ISBN 9781800565913. - ISBN 9781800563452. - Текст : электронный., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=854289&idb=0>.

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

Общероссийский математический портал <http://www.mathnet.ru>

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 01.04.02 - Прикладная математика и информатика.

Автор(ы): Кувыкин Вячеслав Иванович, доктор физико-математических наук, старший научный сотрудник.

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

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