

**Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования
«ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»**

Структурное подразделение «Брикс кафедры»

УТВЕРЖДЕНА:
на заседании кафедры
Протокол №15 от 18 марта 2025 г.

Рабочая программа дисциплины

«АНАЛИЗ ДАННЫХ / DATA ANALYSIS»

Направление: 09.04.02 Информационные системы и технологии

Цифровые технологии, сети и большие данные / Information technologies, networks and big data

Квалификация: Магистр

Форма обучения: очная

Документ подписан простой электронной подписью Составитель программы: Аталян Алина Валерьевна Дата подписания: 17.06.2025

Документ подписан простой электронной подписью Утвердил: Киреенко Анна Павловна Дата подписания: 18.06.2025

Документ подписан простой электронной подписью Согласовал: Дорофеев Андрей Сергеевич Дата подписания: 18.06.2025
--

Год набора – 2025

Иркутск, 2025 г.

1 Перечень планируемых результатов обучения по дисциплине, соотнесённых с планируемыми результатами освоения образовательной программы

1.1 Дисциплина «Анализ данных / Data Analysis» обеспечивает формирование следующих компетенций с учётом индикаторов их достижения

Код, наименование компетенции	Код индикатора компетенции
ОПК-1 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	ОПК-1.4
ОПК-2 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	ОПК-2.4
ОПК-5 Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем	ОПК-5.4
ОПК-7 Способен разрабатывать и применять математические модели процессов и объектов при решении задач анализа и синтеза распределённых информационных систем и систем поддержки принятия решений	ОПК-7.2

1.2 В результате освоения дисциплины у обучающихся должны быть сформированы

Код индикатора	Содержание индикатора	Результат обучения
ОПК-1.4	Владеет навыками решения нестандартных профессиональных задач с применением современных методов анализа данных в профессиональной деятельности	Знать основные подходы к решению профессиональных задач методами анализа данных Уметь применять методы анализа данных для решения конкретных прикладных профессиональных задач Владеть навыками решения прикладных профессиональных задач методами анализа данных
ОПК-2.4	Использует технологии обработки, анализа данных для разработки оригинальных алгоритмов и программных средств, в процессе решения профессиональных задач	Знать основные алгоритмы анализа данных, методы их разработки и условия применения Уметь адаптировать известные алгоритмы анализа данных для решения прикладных задач Владеть методами использования алгоритмов анализа данных для решения прикладных задач
ОПК-5.4	Применяет технологии разработки информационных и	Знать основные принципы использования методов анализа

	автоматизированных систем для решения профессиональных задач, связанных с анализом данных	данных в профессиональных информационных системах Уметь применять современные информационные системы для решения задач анализа данных Владеть навыками анализа данных в профессиональных информационных системах
ОПК-7.2	Использует технологии обработки и анализа данных для разработки оригинальных алгоритмов и программных средств в процессе обработки информации	Знать основные алгоритмы анализа данных и методы их разработки Уметь Адаптировать известные алгоритмы анализа данных для решения прикладных задач Владеть методами использования алгоритмов анализа данных для решения прикладных задач

2 Место дисциплины в структуре ООП

Изучение дисциплины «Анализ данных / Data Analysis» базируется на результатах освоения следующих дисциплин/практик: «Прикладная математика в ИТ / Applied Mathematics in IT», «Визуализация данных / Data Visualization»

Дисциплина является предшествующей для дисциплин/практик: «Программирование искусственного интеллекта в приложениях / Artificial Intelligence Programming in Applications», «Производственная практика: научно-исследовательская работа / Scientific Research and Manufacturing Training»

3 Объем дисциплины

Объем дисциплины составляет – 3 ЗЕТ

Вид учебной работы	Трудоемкость в академических часах (Один академический час соответствует 45 минутам астрономического часа)	
	Всего	Семестр № 2
Общая трудоемкость дисциплины	108	108
Аудиторные занятия, в том числе:	28	28
лекции	14	14
лабораторные работы	14	14
практические/семинарские занятия	0	0
Самостоятельная работа (в т.ч. курсовое проектирование)	44	44
Трудоемкость промежуточной аттестации	36	36
Вид промежуточной аттестации (итогового контроля по дисциплине)	Экзамен	Экзамен

4 Структура и содержание дисциплины

4.1 Сводные данные по содержанию дисциплины

Семестр № 2

№ п/п	Наименование раздела и темы дисциплины	Виды контактной работы						СРС		Форма текущего контроля
		Лекции		ЛР		ПЗ(СЕМ)		№	Кол. Час.	
		№	Кол. Час.	№	Кол. Час.	№	Кол. Час.			
1	2	3	4	5	6	7	8	9	10	11
1	Введение в дисциплину	1	2							Устный опрос
2	Подготовка данных	2	2	1	2					Отчет по лабораторной работе
3	Методы статистического анализа данных	3, 4	4	2, 3, 4	10			1	10	Отчет по лабораторной работе
4	Расширенные методы анализа данных	5, 6	4	5	2			1	10	Отчет по лабораторной работе
5	Приложения методов и примеры использования. Этические аспекты анализа данных.	7	2					2	24	Устный опрос
	Промежуточная аттестация								36	Экзамен
	Всего		14		14				80	

4.2 Краткое содержание разделов и тем занятий

Семестр № 2

№	Тема	Краткое содержание
1	Введение в дисциплину	Что такое анализ данных? Определение и основные понятия. Роль анализа данных в бизнесе, науке и технологиях. Жизненный цикл данных (сбор, очистка, анализ, визуализация, интерпретация). Типы данных. Структурированные и неструктурированные данные. Качественные и количественные данные. Временные ряды, категориальные данные, текстовые данные. Инструменты анализа данных. Обзор популярных инструментов анализа данных.
2	Подготовка данных	Отсутствующие данные. Определение и работа со значениями NA. Методы импутации. Выбросы. Определение выбросов. Обнаружение выбросов. Приложения и методы для работы с выбросами. Преобразование данных. Изменение формы и агрегирование данных. Объединение наборов данных
3	Методы статистического анализа данных	Качественный и количественный анализ. Распределение вероятностей. Описательная статистика. Выводная статистика (статистические тесты). Разведочный анализ данных (РАД).

		<p>Корреляционный анализ. Основные характеристики корреляции.</p> <p>Коэффициент корреляции Пирсона. Коэффициент ранговой корреляции Спирмена. Коэффициент корреляции Кендалла.</p> <p>Проверка гипотез и доверительные интервалы.</p> <p>Другие виды корреляции. Факторы, влияющие на корреляционный анализ.</p> <p>Регрессионный анализ. Типы регрессионных моделей. Примеры применения.</p>
4	Расширенные методы анализа данных	<p>Кластерный анализ. Проблемы кластерного анализа. Основные концепции или «дорожная карта» для кластеризации. Типы методов кластеризации. Иерархический агломеративный кластерный анализ.</p> <p>Критерии определения количества кластеров.</p> <p>Характеристика методов разбиения. Метод кластеризации k-means. Практические советы по k-means. Анализ главных компонент (РСА). Рабочий процесс РСА. Стандартизация переменных.</p> <p>Уменьшение размерности. Масштабирование и интерпретация. Введение в анализ временных рядов. Классификация временных рядов.</p> <p>Стационарные и нестационарные временные ряды. Основные методы анализа стационарных временных рядов.</p>
5	Приложения методов и примеры использования. Этические аспекты анализа данных.	<p>Примеры некоторых реальных приложений аналитики данных: здравоохранение, финансы, розничная торговля, маркетинг, транспорт и логистика, образование, производство, телекоммуникации. Этика в анализе данных.</p> <p>Основные этические проблемы. Трехмерная этическая структура. Эволюция этических структур анализа данных.</p>

4.3 Перечень лабораторных работ

Семестр № 2

№	Наименование лабораторной работы	Кол-во академических часов
1	Практикум по предобработке данных в R: импутационные методы и обнаружение аномалий в данных	2
2	Оценка параметров статистических выборок методом максимального правдоподобия.	2
3	Тестирование статистических гипотез в R	4
4	Исследование зависимости между признаками. Корреляционный и регрессионный анализ. Перестановочный тест оценки параметров.	4
5	Подготовка временных рядов к анализу: обработка пропусков, аномалий и приведение к	2

	стационарности. Подбор оптимальной модели (ARMA/ARIMA/SARIMA)	
--	---	--

4.4 Перечень практических занятий

Практических занятий не предусмотрено

4.5 Самостоятельная работа

Семестр № 2

№	Вид СРС	Кол-во академических часов
1	Оформление отчетов по лабораторным и практическим работам	20
2	Подготовка к экзамену	24

В ходе проведения занятий по дисциплине используются следующие интерактивные методы обучения: Дискуссия

5 Перечень учебно-методического обеспечения дисциплины

5.1 Методические указания для обучающихся по освоению дисциплины

5.1.1 Методические указания для обучающихся по лабораторным работам:

Аталян А.В. Data Analysis/Анализ данных: электронный курс / А.В. Аталян
<https://el.istu.edu/course/view.php?id=10588> (дата обращения 17.06.25) Режим доступа: для зарегистрированных пользователей.

5.1.2 Методические указания для обучающихся по самостоятельной работе:

Аталян А.В. Data Analysis/Анализ данных: электронный курс / А.В. Аталян
<https://el.istu.edu/course/view.php?id=10588> (дата обращения 17.06.25) Режим доступа: для зарегистрированных пользователей.

6 Фонд оценочных средств для контроля текущей успеваемости и проведения промежуточной аттестации по дисциплине

6.1 Оценочные средства для проведения текущего контроля

6.1.1 семестр 2 | Устный опрос

Описание процедуры.

Групповой разбор вопросов длительностью по 15–20 минут на каждую тему (в зависимости от сложности вопросов).

Возможны дополнительные практические задания (например, устное объяснение шагов решения задачи).

Примеры тем для опроса:

Основы анализа данных, типы данных (количественные, категориальные, временные ряды), методы описательной статистики (среднее, медиана, дисперсия, корреляция), визуализация данных (графики, гистограммы, box-plot), предобработка данных, обработка пропущенных значений и выбросов, нормализация и стандартизация данных,

статистические методы и проверка гипотез, P-value, доверительные интервалы, статистические критерии (например, Стьюдента, хи-квадрат и т.д.).

Критерии оценивания.

- 5 (отлично) – полные, аргументированные ответы с примерами, понимание взаимосвязей между методами;
- 4 (хорошо) – правильные ответы с незначительными неточностями;
- 3 (удовлетворительно) – поверхностное понимание, затруднения в объяснении;
- 2 (неудовлетворительно) – отсутствие ответа на основные вопросы.

6.1.2 семестр 2 | Отчет по лабораторной работе

Описание процедуры.

Выполнение каждой лабораторной работы состоит из двух частей: практической части в ходе которой необходимо выполнить индивидуальное задание и подготовка и защита отчета по лабораторной работе. Реализация решения индивидуального задания осуществляется средствами языка программирования R. В ходе выполнения лабораторной работы студент должен применить для выполнения задания теоретические знания, полученные в ходе изучения курса, а также приобрести практические навыки использования процедур и функций языка R. Подготовка отчета по лабораторной работе необходима для закрепления и переосмысления полученных знаний и навыков. Для успешной сдачи лабораторной работы необходимо продемонстрировать работу программы, реализующую индивидуальное задание, устранить выявленные преподавателем недостатки, подготовить и защитить отчет. Требования к оформлению программного кода, а также содержанию отчета указаны в соответствующем разделе методических указаний для каждой лабораторной работы.

Критерии оценивания.

В процессе защиты отчета по лабораторной работе, студенту задаются контрольные вопросы теоретического и практического характера, соответствующие теме работы. Для успешной защиты отчета студенту необходимо дать краткое изложение основных результатов полученных в ходе выполнения лабораторной работы, устно ответить на теоретические вопросы по теме лабораторной работы, а также продемонстрировать умение ориентироваться в написанном программном коде. Успешная защита отчета является необходимым условием для выставления оценки «зачтено» по соответствующей лабораторной работе.

6.2 Оценочные средства для проведения промежуточной аттестации

6.2.1 Критерии и средства (методы) оценивания индикаторов достижения компетенции в рамках промежуточной аттестации

Индикатор достижения компетенции	Критерии оценивания	Средства (методы) оценивания промежуточной аттестации
ОПК-1.4	Применяет навыки решения прикладных профессиональных задач методами анализа данных	Устное собеседование и/или практические

		задания и/или тест
ОПК-2.4	Демонстрирует знание основных алгоритмов анализа данных, сферы и условия их применения. Демонстрирует способность применять эти алгоритмы для решения прикладных профессиональны	Решение практических задач и устное собеседование.
ОПК-5.4	Демонстрирует знание основных алгоритмов анализа данных, сферы и условия их применения. Демонстрирует способность применять эти алгоритмы для решения прикладных профессиональны	Решение практических задач и устное собеседование
ОПК-7.2	Демонстрирует знание основных алгоритмов анализа данных, сферы и условия их применения. Демонстрирует способность применять эти алгоритмы для решения прикладных профессиональны	Решение практических задач и устное собеседование

6.2.2 Типовые оценочные средства промежуточной аттестации

6.2.2.1 Семестр 2, Типовые оценочные средства для проведения экзамена по дисциплине

6.2.2.1.1 Описание процедуры

До экзамена допускаются студенты, которые выполнили все, предусмотренные промежуточные практические (лабораторные) работы.

Экзамен состоит из теоретической части в виде теста из 40 вопросов.

Каждый вопрос имеет вес 2,5 балла, максимальное количество - 100 баллов.

Перечень вопросов для теоретической части экзамена:

1. Что такое анализ данных?
2. Какой механизм пропуска данных называется случайным пропуском?
3. Какой механизм пропуска данных называется систематическим пропуском?
4. Какие методы обработки пропущенных данных обеспечивают несмещённые оценки при условии MCAR?
5. Какие методы импутации сохраняют несмещённость оценок, если данные пропущены по механизму MAR?
6. Пожалуйста, отметьте существующие типы выбросов
7. Какие методы дают эффективную оценку для обнаружения выбросов?
8. В контексте обработки данных изменение формы данных это... преобразование структуры набора данных без изменения его содержимого, преобразование структуры набора данных с изменением его содержимого.
9. Для чего используется изменение формы данных (Reshaping Data)?
10. Что включает в себя процедура изменение формы данных (Reshaping Data)?
11. Что такое melting data?
12. Что такое Качественный анализ данных?
13. Перечислите категории, которые относятся к качественному анализу данных.
14. Перечислите преимущества качественного анализа данных

15. Что такое количественный анализ данных?
16. Какие из следующих распределений имеют плотность вероятности относительно меры Лебега?
17. Какая мера центральной тенденции наиболее подходит для асимметричных распределений непрерывных переменных?
18. Какова роль p -значений и доверительных интервалов в выводной статистике?
19. Какова основная цель разведочного анализа данных (РАД)?
20. Какое из следующих утверждений верно относительно мультиколлинеарности:
 - a) Мультиколлинеарность возникает, когда независимые переменные сильно коррелируют, что затрудняет оценку индивидуального вклада каждого предиктора.
 - b) Мультиколлинеарность всегда приводит к смещенным оценкам коэффициентов в регрессионных моделях
 - c) Удаление одной из коррелированных переменных — единственный способ решения проблемы мультиколлинеарности
 - d) Мультиколлинеарность не влияет на предсказательную силу модели, пока прогнозы делаются в пределах диапазона исходных данных?
21. Что такое корреляция?
22. Что означает аббревиатура RPMCC?
23. О чем свидетельствует коэффициент корреляции Пирсона (r), равный $-0,9$?
24. Когда следует использовать ранговую корреляцию Спирмена вместо Пирсона?
25. Что может искусственно завысить коэффициент корреляции?
26. Если 95% доверительный интервал для коэффициента корреляции включает 0, о чем это говорит?
27. Что представляет собой значение R^2 в линейной регрессии?
28. Какая регрессионная модель подходит для бинарной зависимой переменной?
29. Когда возникает мультиколлинеарность в регрессии?
30. На что указывает значение $p > 0,05$ в тесте Дики-Фуллера?
31. Выберите преимущества качественного анализа: данные и анализ «обоснованы», данные богаты и подробны, допускается неоднозначность и противоречия, существует перспектива альтернативных объяснений, данные могут быть менее репрезентативными, интерпретация связана с «я» исследователя, существует вероятность деконтекстуализации смысла, существует опасность чрезмерного упрощения объяснения.
32. Преобразование Бокса-Кокса в основном используется для: Устранения сезонности, Стабилизации дисперсии и придания данным более нормального вида, Преобразования данных в стационарный ряд, Устранения автокорреляции.
33. Если данные временного ряда повторяются каждые 12 месяцев, это называется: Тренд, Сезонность, Случайность, Автокорреляция.
34. Общие этапы многомерного регрессионного моделирования ...
35. Что оценивает тест Квятковского-Филлипса-Шмидта-Шина (KPSS) в сравнении с тестом Дики-Фуллера (ADF)?
36. Какой статистический тест используется для формальной проверки стационарности?
37. Что отличает «периодичность» от «сезонности»?
38. Для временного ряда с трендом и сезонностью, какая модель является наиболее подходящей?
39. Какой компонент временного ряда описывает долгосрочное увеличение или уменьшение данных?
40. Что наилучшим образом описывает данные временных рядов?

Пример задания:

пример вопроса из теоретического теста с вариантами ответа:

Какое из следующих утверждений верно относительно мультиколлинеарности:

- a) Мультиколлинеарность возникает, когда независимые переменные сильно коррелируют, что затрудняет оценку индивидуального вклада каждого предиктора.
- b) Мультиколлинеарность всегда приводит к смещенным оценкам коэффициентов в регрессионных моделях
- c) Удаление одной из коррелированных переменных — единственный способ решения проблемы мультиколлинеарности
- d) Мультиколлинеарность не влияет на предсказательную силу модели, пока прогнозы делаются в пределах диапазона исходных данных?

Правильные варианты ответа a, d.

6.2.2.1.2 Критерии оценивания

Отлично	Хорошо	Удовлетворительно	Неудовлетворительно
Для оценки "отлично" нужно набрать 80 и более баллов.	Для оценки "хорошо" нужно набрать от 70 до 79 баллов.	Для оценки "удовлетворительно" нужно набрать от 60 до 69 баллов.	Оценка "неудовлетворительно" выставляется, если набрано менее 60 баллов.

7 Основная учебная литература

1. Тюрин Ю. Н. Анализ данных на компьютере : учебное пособие по направлениям "Математика" , "Математика . Прикладная математика" / Ю. Н. Тюрин , А. А. Макаров, 2011. - 366,[1].
2. Сапрыкин О. Н. Интеллектуальный анализ данных : учебное пособие / О. Н. Сапрыкин, 2020. - 80.
3. Золкин, А. Л. Математическое моделирование и анализ данных : учебное пособие для вузов / А. Л. Золкин, М. В. Сартаков. — Санкт-Петербург : Лань, 2025. — 128 с. — ISBN 978-5-507-51616-2 Add to Citavi project by ISBN. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/455660> (дата обращения: 03.06.2025). — Режим доступа: для авто

8 Дополнительная учебная литература и справочная

1. Миркин Б. Г. Введение в анализ данных : учебник и практикум для вузов / Б. Г. Миркин, 2022. - 174.
2. Маркова С. В. Анализ данных на языке R (с практикумом) : учебник / С. В. Маркова, 2024. - 224.
3. Шумвей Р. Х., Стоффер Д. С. Анализ временных рядов и его приложения : пер. с англ. — 3-е изд. — Нью-Йорк : Springer, 2000. — Т. 3. — 549 с.

9 Ресурсы сети Интернет

1. <http://library.istu.edu/>
2. <https://e.lanbook.com/>
3. <https://posit.cloud/>

10 Профессиональные базы данных

1. <http://new.fips.ru/>
2. <http://www1.fips.ru/>

11 Перечень информационных технологий, лицензионных и свободно распространяемых специализированных программных средств, информационных справочных систем

1. Свободно распространяемое программное обеспечение <https://www.r-project.org/>

12 Материально-техническое обеспечение дисциплины

1. Проектор Epson EB-W04LCD.WXGA 1280*800.3000:1.2800 ANSI Lumens
2. Компьютер Asustek P8H6-M/Intel Core i5
2400/4Gb/HDD2TB/DVD-RW/ATX550W/LCD22/ИБП1