

МИНОБРНАУКИ РОССИИ

**Федеральное государственное автономное образовательное
учреждение высшего образования "Пермский
государственный национальный исследовательский
университет"**

Кафедра информационных систем и математических методов в экономике

Авторы-составители: **Радионова Марина Владимировна
Васёва Галина Сергеевна
Ильин Иван Вадимович**

Рабочая программа дисциплины

АНАЛИЗ ДАННЫХ

Код УМК 101460

Утверждено
Протокол №8
от «17» апреля 2024 г.

Пермь, 2024

1. Наименование дисциплины

Анализ данных

2. Место дисциплины в структуре образовательной программы

Дисциплина входит в обязательную часть Блока « Б.1 » образовательной программы по направлениям подготовки (специальностям):

Направление подготовки: **01.03.02** Прикладная математика и информатика

направленность Анализ данных и искусственный интеллект в цифровой экономике

3. Планируемые результаты обучения по дисциплине

В результате освоения дисциплины **Анализ данных** у обучающегося должны быть сформированы следующие компетенции:

01.03.02 Прикладная математика и информатика (направленность : Анализ данных и искусственный интеллект в цифровой экономике)

УК.2 Способен определять круг задач в рамках поставленной цели и выбирать способы их решения, исходя из имеющихся ресурсов и ограничений

Индикаторы

УК.2.3 Обосновывает способ решения задачи с учетом имеющихся ресурсов и ограничений

ОПК.3 Способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач

Индикаторы

ОПК.3.3 Демонстрирует практический опыт решения прикладных задач с использованием систем программирования и специализированного программного обеспечения

4. Объем и содержание дисциплины

Направление подготовки	01.03.02 Прикладная математика и информатика (направленность: Анализ данных и искусственный интеллект в цифровой экономике)
форма обучения	очная
№№ триместров, выделенных для изучения дисциплины	5
Объем дисциплины (з.е.)	5
Объем дисциплины (ак.час.)	180
Контактная работа с преподавателем (ак.час.), в том числе:	84
Проведение лекционных занятий	28
Проведение практических занятий, семинаров	28
Проведение лабораторных работ, занятий по иностранному языку	28
Самостоятельная работа (ак.час.)	96
Формы текущего контроля	Защищаемое контрольное мероприятие (1) Итоговое контрольное мероприятие (1) Письменное контрольное мероприятие (2)
Формы промежуточной аттестации	Экзамен (5 триместр)

5. Аннотированное описание содержания разделов и тем дисциплины

Анализ данных

Тема 1. Предварительный анализ данных

Статистические методы обработки экспериментальных данных. Основные понятия и задачи математической статистики. Генеральная совокупность, выборка, результаты наблюдений, статистика, статистическая оценка, требования к оценкам. Классификация признаков по шкалам измерений. Описательная статистика: среднее значение, математическое ожидание, медиана, мода, дисперсия, среднее квадратичное отклонение, коэффициент вариации, показатель точности опыта, минимум, максимум, размах выборки, моменты распределения. Вариационная статистика: параметры классовых интервалов, группировка, функции эмпирического распределения. Общее понятие о моментах распределения. Начальные, центральные и условные моменты K -го порядка. Нормирование момента. Моменты распределения, используемые в качестве показателей асимметрии и эксцесса ряда. Генеральная и выборочная совокупность. Основные обобщающие характеристики генеральной выборочной совокупности.

Ранжирование. Методы первичной обработки данных. Основы визуализации: полигон, гистограмма, кумулята, огива, "ящик с усами".

Визуализация качественных данных

Тема 2. Доверительное оценивание и проверка статистических гипотез

Основы интервального оценивания и проверки статистических гипотез. Доверительные интервалы для параметров нормальной совокупности

Доверительный интервал для математического ожидания. Доверительный интервал для вероятности.

Доверительный интервал для доли совокупности.

Основные понятия проверки статистических гипотез. Нулевая и конкурирующая гипотезы. Критерии проверки.

Общая схема решения задачи проверки статистических гипотез. Параметрические и непараметрические гипотезы. Проверка случайности выборки из нормальной совокупности. Критерии обнаружения выбросов. Критерии проверки гипотез о параметрах нормального распределения. Критерии согласия.

Тема 3. Корреляционный анализ данных

Выборочный и теоретический коэффициенты корреляции. Парный коэффициент корреляции. Проверка гипотезы о значимости коэффициента корреляции. Доверительный интервал для парного коэффициента корреляции. Частный коэффициент корреляции. Проверка гипотезы о значимости частного коэффициента корреляции. Доверительный интервал для частного коэффициента корреляции.

Множественный коэффициент корреляции. Проверка гипотезы о значимости множественного коэффициента корреляции.

Корреляционный анализ качественных данных. Исследование связи между номинальными переменными (таблица сопряженности признаков, критерий хи-квадрат, меры связи признаков: коэффициенты контингенции, ассоциации, среднеквадратической сопряженности, Пирсона, Крамера). Исследование связи между порядковыми переменными (ранговый коэффициент корреляции Спирмена, коэффициент согласованности Кендалла, коэффициент конкордации).

Проверка гипотез о значимости рангового коэффициента корреляции.

Тема 4. Методы классификации многомерных наблюдений

Классификация многомерных наблюдений без обучения (непараметрический случай), кластерный анализ. Общая постановка задачи автоматической классификации, классификации без обучения, понятия кластерного анализа. Метрики расстояния и близости между объектами, расстояния между кластерами. Функционалы качества разбиения. Основные типы задач и алгоритмов кластерного анализа.

Иерархические, параллельные и последовательные процедуры кластерного анализа. Метод к-средних. Зависимость выбора алгоритма классификации от цели статистического исследования. Классификация многомерных наблюдений при наличии обучающих выборок, дискриминантный анализ. Методы классификации с обучением, основные понятия дискриминантного анализа. Обучающие выборки. Линейный дискриминантный анализ при известных параметрах многомерного нормального закона распределения (случай двух классов и общий случай). Вероятность ошибочной классификации с помощью дискриминантной функции. Оценка качества дискриминантной функции и информативности отдельных признаков.

Контрольная точка 1

Контрольное мероприятие по темам "Предварительный анализ данных" и "Доверительное оценивание и проверка статистических гипотез"

Контрольная точка 2

Контрольное мероприятие по теме "Корреляционный анализ данных"

Контрольная точка 3

Контрольное мероприятие по всем разделам дисциплины

Итоговое контрольное мероприятие

Индивидуальный исследовательский проект по всем разделам дисциплины

6. Методические указания для обучающихся по освоению дисциплины

Освоение дисциплины требует систематического изучения всех тем в той последовательности, в какой они указаны в рабочей программе.

Основными видами учебной работы являются аудиторные занятия. Их цель - расширить базовые знания обучающихся по осваиваемой дисциплине и систему теоретических ориентиров для последующего более глубокого освоения программного материала в ходе самостоятельной работы. Обучающемуся важно помнить, что контактная работа с преподавателем эффективно помогает ему овладеть программным материалом благодаря расстановке необходимых акцентов и удержанию внимания интонационными модуляциями голоса, а также подключением аудио-визуального механизма восприятия информации.

Самостоятельная работа преследует следующие цели:

- закрепление и совершенствование теоретических знаний, полученных на лекционных занятиях;
- формирование навыков подготовки текстовой составляющей информации учебного и научного назначения для размещения в различных информационных системах;
- совершенствование навыков поиска научных публикаций и образовательных ресурсов, размещенных в сети Интернет;
- самоконтроль освоения программного материала.

Обучающемуся необходимо помнить, что результаты самостоятельной работы контролируются преподавателем во время проведения мероприятий текущего контроля и учитываются при промежуточной аттестации.

Обучающимся с ОВЗ и инвалидов предоставляется возможность выбора форм проведения мероприятий текущего контроля, альтернативных формам, предусмотренным рабочей программой дисциплины. Предусматривается возможность увеличения в пределах 1 академического часа времени, отводимого на выполнение контрольных мероприятий.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации.

При проведении текущего контроля применяются оценочные средства, обеспечивающие передачу информации, от обучающегося к преподавателю, с учетом психофизиологических особенностей здоровья обучающихся.

7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

При самостоятельной работе обучающимся следует использовать:

- конспекты лекций;
- литературу из перечня основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля);
- текст лекций на электронных носителях;
- ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины;
- лицензионное и свободно распространяемое программное обеспечение из перечня информационных технологий, используемых при осуществлении образовательного процесса по дисциплине;
- методические указания для обучающихся по освоению дисциплины.

8. Перечень основной и дополнительной учебной литературы

Основная:

1. Анализ данных : учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2019. — 490 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт].

<https://www.urait.ru/bcode/432178>

2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2019. — 174 с. — (Авторский учебник). — ISBN 978-5-9916-5009-0. — Текст : электронный // ЭБС Юрайт [сайт]. <https://www.urait.ru/bcode/432851>

3. Data Science и интеллектуальный анализ данных : учебное пособие / Д. М. Назаров, С. В. Бегичева, Д. Б. Ковтун, А. Д. Назаров. — Москва : Ай Пи Ар Медиа, 2023. — 304 с. — ISBN 978-5-4497-1931-7. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт].

<https://www.iprbookshop.ru/127201.html>

Дополнительная:

1. Замятин, А. В. Интеллектуальный анализ данных : учебное пособие / А. В. Замятин. — Томск : Издательский Дом Томского государственного университета, 2020. — 194 с. — ISBN 978-5-94621-898-6. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт].

<https://www.iprbookshop.ru/116889.html>

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

https://praktikum.yandex.ru/data-scientist/?utm_source=google&utm_medium=cpc&utm_campaign=Google_Search_DS_Smart&utm_content=nt_g:pl_:cid_11035922147:gid_107384528359:kw_:tid_dsa-437398801577:crid_514817029584:adp_:d_c:dm_:lim_:lpm_1011993& Материалы по анализу данных

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Образовательный процесс по дисциплине **Анализ данных** предполагает использование следующего программного обеспечения и информационных справочных систем:

Образовательный процесс по дисциплине предполагает использование следующего программного обеспечения и информационных справочных систем:

- доступ в режиме on-line в Электронную библиотечную систему (ЭБС);
- доступ в электронную информационно-образовательную среду университета.

При освоении материала и выполнения заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ (student.psu.ru).

Перечень необходимого лицензионного и (или) свободно распространяемого программного обеспечения:

- офисный пакет приложений;
- приложение позволяющее просматривать и воспроизводить медиаконтент PDF-файлов «Adobe Acrobat Reader DC»;
- офисный пакет приложений «LibreOffice»;
- свободно распространяемое ПО Gretl (<http://gretl.sourceforge.net/ru.html>)
- свободно распространяемое ПО R (<https://www.r-studio.com/ru/>)
- специальное ПО не требуется.

При освоении материала и выполнения заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ (student.psu.ru).

При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

- система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).
- система LMS Moodle (<http://e-learn.psu.ru/>), которая поддерживает возможность использования текстовых материалов и презентаций, аудио- и видеоконтент, а так же тесты, проверяемые задания, задания для совместной работы.
- система тестирования Indigo (<https://indigotech.ru/>).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для лекционных занятий требуется аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для проведения практических занятий - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для лабораторных занятий – компьютерный класс, оснащенный персональными ЭВМ и

соответствующим программным обеспечением. Состав оборудования определен в Паспорте компьютерного класса

Для групповых (индивидуальных) консультаций - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для проведения текущего контроля - аудитория, оснащенная меловой (и) или маркерной доской.

Самостоятельная работа студентов: аудитория, оснащенная компьютерной техникой с возможностью подключения к сети «Интернет», с обеспеченным доступом в электронную информационно-образовательную среду университета, помещения Научной библиотеки ПГНИУ.

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютера с доступом к локальной и глобальной компьютерным сетям.

2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.

5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет Libreoffice.

Справочно-правовая система «КонсультантПлюс»

**Фонды оценочных средств для аттестации по дисциплине
Анализ данных**

**Планируемые результаты обучения по дисциплине для формирования компетенции.
Индикаторы и критерии их оценивания**

ОПК.3

Способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
ОПК.3.3 Демонстрирует практический опыт решения прикладных задач с использованием систем программирования и специализированного программного обеспечения	Демонстрирует практический опыт решения прикладных задач с использованием систем программирования и специализированного программного обеспечения Знать основные методы анализа взаимосвязи между экономическими процессами и явлениями, основные эконометрические методы анализа реальных социально-экономических явлений, уметь применять эти знания для анализа, верификации и интерпретации данных, владеть способностью применять методы и средства эконометрического анализа взаимосвязи между экономическими процессами и явлениями с помощью математических и статистических методов и моделей, в том числе с использованием систем программирования и специализированного программного обеспечения	Неудовлетворител Не знает основные понятия и методы анализа данных, основные методы анализа реальных социально-экономических явлений, Не умеет применять эти знания для анализа, верификации и интерпретации данных, Не владеет способностью применять методы и средства эконометрического анализа взаимосвязи между экономическими процессами и явлениями с помощью математических и статистических методов и моделей, в том числе с использованием систем программирования и специализированного программного обеспечения Удовлетворительн Не уверенно знает понятия и методы анализа данных, основные методы анализа реальных социально-экономических явлений, Удовлетворительный уровень умения применять эти знания для анализа, верификации и интерпретации данных, Удовлетворительно владеет способностью применять методы и средства эконометрического анализа взаимосвязи между экономическими процессами и явлениями с помощью математических и статистических методов и моделей, в том числе с использованием систем программирования и специализированного программного обеспечения Хорошо В целом хорошие знания основных понятий и методов анализа данных, основных

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		Хорошо методов анализа реальных социально-экономических явлений, В целом сформировано умение применять эти знания для анализа, верификации и интерпретации данных, В целом хорошее владение навыками применения методов и средств эконометрического анализа взаимосвязи между экономическими процессами и явлениями с помощью математических и статистических методов и моделей, в том числе с использованием систем программирования и специализированного программного обеспечения Отлично Высокий уровень знания основных понятий и методик анализа данных, основных методов анализа реальных социально-экономических явлений, Сформировано уверенное умение применять эти знания для анализа, верификации и интерпретации данных, Высокий уровень владения навыками использования методов и средств эконометрического анализа взаимосвязи между экономическими процессами и явлениями с помощью математических и статистических методов и моделей, в том числе с использованием систем программирования и специализированного программного обеспечения

УК.2

Способен определять круг задач в рамках поставленной цели и выбирать способы их решения, исходя из имеющихся ресурсов и ограничений

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
УК.2.3 Обосновывает способ решения задачи с учетом имеющихся ресурсов и ограничений	Знает методы сбора, обработки и анализа научных данных; Умеет находить способы решения задачи с учетом имеющихся ресурсов и ограничений; Владеет навыками обоснования	Неудовлетворител Не знает методы сбора, обработки и анализа научных данных; Не умеет находить способы решения задачи с учетом имеющихся ресурсов и ограничений; Не владеет навыками обоснования способов

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
	способов решения задачи с учетом имеющихся ресурсов и ограничений.	Неудовлетворител решения задачи с учетом имеющихся ресурсов и ограничений. Удовлетворительн Слабо знает методы сбора, обработки и анализа научных данных; Допускает ошибки в ходе нахождения и обоснования способов решения задачи с учетом имеющихся ресурсов и ограничений; Демонстрирует неуверенное владение навыками обоснования способов решения задачи с учетом имеющихся ресурсов и ограничений. Хорошо Знает методы сбора, обработки и анализа научных данных; Допускает незначительные ошибки в процессе нахождения способов решения задачи с учетом имеющихся ресурсов и ограничений; Владеет навыками обоснования способов решения задачи с учетом имеющихся ресурсов и ограничений. Отлично Отлично знает методы сбора, обработки и анализа научных данных; Умеет находить способы решения задачи с учетом имеющихся ресурсов и ограничений; Демонстрирует уверенное владение навыками обоснования способов решения задачи с учетом имеющихся ресурсов и ограничений.

Оценочные средства текущего контроля и промежуточной аттестации

Схема доставки : Базовая

Вид мероприятия промежуточной аттестации : Экзамен

Способ проведения мероприятия промежуточной аттестации : Оценка по дисциплине в рамках промежуточной аттестации определяется на основе баллов, набранных обучающимся на контрольных мероприятиях, проводимых в течение учебного периода.

Максимальное количество баллов : 100

Конвертация баллов в отметки

«отлично» - от 81 до 100

«хорошо» - от 61 до 80

«удовлетворительно» - от 42 до 60

«неудовлетворительно» / «незачтено» менее 42 балла

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
УК.2.3 Обосновывает способ решения задачи с учетом имеющихся ресурсов и ограничений ОПК.3.3 Демонстрирует практический опыт решения прикладных задач с использованием систем программирования и специализированного программного обеспечения	Контрольная точка 1 Письменное контрольное мероприятие	Оценивается качество знаний и умений в заданиях по предварительному анализ данных
УК.2.3 Обосновывает способ решения задачи с учетом имеющихся ресурсов и ограничений ОПК.3.3 Демонстрирует практический опыт решения прикладных задач с использованием систем программирования и специализированного программного обеспечения	Контрольная точка 2 Письменное контрольное мероприятие	Оценивается качество знаний и умений в задачах на доверительное оценивание

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
УК.2.3 Обосновывает способ решения задачи с учетом имеющихся ресурсов и ограничений ОПК.3.3 Демонстрирует практический опыт решения прикладных задач с использованием систем программирования и специализированного программного обеспечения	Контрольная точка 3 Защищаемое контрольное мероприятие	Оценивается качество знаний и умений в задании на корреляц. анализ данных
УК.2.3 Обосновывает способ решения задачи с учетом имеющихся ресурсов и ограничений ОПК.3.3 Демонстрирует практический опыт решения прикладных задач с использованием систем программирования и специализированного программного обеспечения	Итоговое контрольное мероприятие Итоговое контрольное мероприятие	Оценивается качество знаний и умений в задачах на классификацию

Спецификация мероприятий текущего контроля

Контрольная точка 1

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **20**

Проходной балл: **8.5**

Показатели оценивания	Баллы
Верно решенное задание	1

Контрольная точка 2

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **20**

Проходной балл: **8.5**

Показатели оценивания	Баллы
Верно решенное задание	1

Контрольная точка 3

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **12.5**

Показатели оценивания	Баллы
Верно решенное задание	1

Итоговое контрольное мероприятие

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **12.5**

Показатели оценивания	Баллы
Верно решенное задание	1