

МИНОБРНАУКИ РОССИИ

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

Авторы-составители: **Анисимова Светлана Игоревна
Кнутова Наталия Сергеевна
Ренев Даниил Алексеевич
Сомхишвили Кристина Отариевна
Радионова Марина Владимировна
Исаева Екатерина Владимировна**

Рабочая программа дисциплины
СОВРЕМЕННЫЕ ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ
Код УМК 99832

Утверждено
Протокол №8
от «01» марта 2023 г.

Пермь, 2023

1. Наименование дисциплины

Современные технологии анализа данных

2. Место дисциплины в структуре образовательной программы

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

Направление подготовки: **06.03.01** Биология
направленность Экспериментальная биология

3. Планируемые результаты обучения по дисциплине

В результате освоения дисциплины **Современные технологии анализа данных** у обучающегося должны быть сформированы следующие компетенции:

06.03.01 Биология (направленность : Экспериментальная биология)

ОПК.2 Способен понимать принципы работы современных информационно-коммуникационных технологий и использовать их для решения профессиональных задач с учетом требований информационной безопасности

Индикаторы

ОПК.2.1 Демонстрирует базовые знания в области информационно-коммуникационных технологий

ОПК.2.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности

4. Объем и содержание дисциплины

Направление подготовки	06.03.01 Биология (направленность: Экспериментальная биология)
форма обучения	очная
№№ триместров, выделенных для изучения дисциплины	8
Объем дисциплины (з.е.)	3
Объем дисциплины (ак.час.)	108
Контактная работа с преподавателем (ак.час.), в том числе:	42
Проведение лекционных занятий	14
Проведение лабораторных работ, занятий по иностранному языку	28
Самостоятельная работа (ак.час.)	66
Формы текущего контроля	Итоговое контрольное мероприятие (1) Письменное контрольное мероприятие (2)
Формы промежуточной аттестации	Зачет (8 триместр)

5. Аннотированное описание содержания разделов и тем дисциплины

Современные технологии анализа данных

Обработка статистических данных уже давно применяется в самых разнообразных видах человеческой деятельности. Трудно назвать ту сферу, в которой она бы не использовалась. Но, пожалуй, ни в одной области знаний и практической деятельности обработка статистических данных не играет такой исключительно большой роли, как в экономике, имеющей дело с обработкой и анализом огромных массивов информации о социально-экономических явлениях и процессах. Всесторонний и глубокий анализ этой информации предполагает использование различных специальных методов, среди которых важное место занимают корреляционный и кластерный анализы обработки статистических данных.

Первичный статистический анализ данных

Основные понятия статистического анализа данных. Типы данных. Генеральная совокупность, выборочная совокупность. ЦПТ.

Методы центральной тенденции (мода, среднее арифметическое, медиана).

Меры разброса (размах, стандартное отклонение, дисперсия, межквартильный размах).

Моменты распределения, используемые в качестве показателей меры центральной тенденции.

Визуализация данных. Гистограммы с наложением плотности нормального распределения.

Визуальная проверка данных на выбросы (гистограммы и "ящик с усами").

Проверка статистических гипотез

Тестирование статистических гипотез. Основные понятия. Виды статистических гипотез. Ошибки I и II рода. Ошибка I рода. Уровень значимости и p-value. Основные принципы проверки статистических гипотез. Параметрические гипотезы. Критерии согласия. Критерии однородности.

Методы корреляционного анализа

Основные показатели исследования взаимосвязей. Коэффициенты корреляции

Пирсона и Спирмена. Таблицы сопряженности и их интерпретация.

Методы регрессионного анализа

Предпосылки линейной регрессии. Построение уравнения простой и множественной линейной регрессии.

Оценка качества построенной модели. Модели бинарного выбора.

Классификация: распознавание образов и типологизация

Общая постановка задачи классификации, классификации без обучения,

понятия кластерного анализа. Метрики расстояния и близости между объектами,

расстояния между кластерами. Основные типы задач и алгоритмов кластерного анализа.

Иерархические, параллельные и последовательные процедуры кластерного анализа.

Метод k-средних. Визуализация.

Итоговое контрольное мероприятие

Итоговое контрольное мероприятие в виде теста по всем темам дисциплины: первичный анализ данных, проверка статистических гипотез, корреляционный анализ и кластерный анализ данных.

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. Маккинли, Уэс Python и анализ данных / Уэс Маккинли ; перевод А. Слинкина. — 2-е изд. — Саратов : Профобразование, 2019. — 482 с. — ISBN 978-5-4488-0046-7. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/88752.html>

Дополнительная:

1. Пальмов, С. В. Интеллектуальный анализ данных : учебное пособие / С. В. Пальмов. — Самара : Поволжский государственный университет телекоммуникаций и информатики, 2017. — 127 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт].
<http://www.iprbookshop.ru/75376.html>

2. Болотова, Л. С. Системы поддержки принятия решений в 2 ч. Часть 2 : учебник и практикум для академического бакалавриата / Л. С. Болотова ; ответственный редактор В. Н. Волкова, Э. С. Болотов. — Москва : Издательство Юрайт, 2019. — 250 с. — (Бакалавр. Академический курс). — ISBN 978-5-9916-8251-0. — Текст : электронный // ЭБС Юрайт [сайт]. <https://www.urait.ru/bcode/437014>

3. Анализ данных качественных исследований : практикум / составители А. П. Истомина. — Ставрополь : Северо-Кавказский федеральный университет, 2016. — 94 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт].
<http://www.iprbookshop.ru/66014.html>

4. Протодяконов, А. В. Алгоритмы Data Science и их практическая реализация на Python : учебное пособие / А. В. Протодяконов, П. А. Пылов, В. Е. Садовников. — Москва, Вологда : Инфра-Инженерия, 2022. — 392 с. — ISBN 978-5-9729-1006-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/124000>

5. 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>

6. Болотова, Л. С. Системы поддержки принятия решений в 2 ч. Часть 1 : учебник и практикум для академического бакалавриата / Л. С. Болотова ; ответственный редактор В. Н. Волкова, Э. С. Болотов. — Москва : Издательство Юрайт, 2019. — 257 с. — (Бакалавр. Академический курс). — ISBN 978-5-9916-8250-3. — Текст : электронный // ЭБС Юрайт [сайт]. <https://www.urait.ru/bcode/436476>

7. Шнарева, Г. В. Анализ данных : учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева. — Симферополь : Университет экономики и управления, 2019. — 129 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт].
<http://www.iprbookshop.ru/89482.html>

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

<https://www.python.org/> Python

<https://sphere.vk.company/materials/video/> Введение в анализ данных

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

Образовательный процесс по дисциплине **Современные технологии анализа данных** 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;
- презентационные материалы (слайды по темам лекционных и практических занятий);
- офисный пакет приложений;
- IDE для Python версии 3 и выше;
- доступ к поисковым системам;
- доступ к бесплатному ПО для визуализации данных.

При освоении дисциплины использование специальных ПО не предусмотрено.

Самостоятельная работа студентов организуется и поддерживается учебными материалами онлайн-курса, размещенного на платформе online.psu.ru

Доступ к курсу по ссылке <https://online.psu.ru/course/view.php?id=385>

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

Материально-техническая база обеспечивается наличием:

1. Лекционные занятия - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.
2. Практические занятия - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.
3. Самостоятельная работа - аудитория для самостоятельной работы, оснащенная компьютерной техникой с возможностью подключения к сети <Интернет>, обеспеченная доступом в электронную информационно-образовательную среду университета. Помещения Научной библиотеки ПГНИУ.
4. Текущий контроль и промежуточная аттестация - аудитория, оснащенная презентационной техникой

(проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской

5. Индивидуальные и групповые консультации - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской или аудитория, оснащенная меловой (и) или маркерной доской.

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютерами 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
Современные технологии анализа данных**

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

ОПК.2

Способен понимать принципы работы современных информационно-коммуникационных технологий и использовать их для решения профессиональных задач с учетом требований информационной безопасности

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
ОПК.2.1 Демонстрирует базовые знания в области информационно-коммуникационных технологий	Знает основные понятия и современные технологии анализа данных. Умеет выбирать современные информационно-коммуникационные технологии для анализа данных. Владеет навыками первичного анализа данных с помощью современных информационно-коммуникационных технологий для анализа данных.	<p style="text-align: center;">Неудовлетворител</p> Не знает основные понятия и современные технологии анализа данных. Не умеет выбирать современные информационно-коммуникационные технологии для анализа данных. Не владеет навыками первичного анализа данных с помощью современных информационно-коммуникационных технологий для анализа данных. <p style="text-align: center;">Удовлетворительн</p> Не уверенно знает основные понятия и современные технологии анализа данных. Удовлетворительный уровень умения выбора современные информационно-коммуникационные технологии для анализа данных. Удовлетворительное владение навыками первичного анализа данных с помощью современных информационно-коммуникационных технологий для анализа данных. <p style="text-align: center;">Хорошо</p> В целом хорошее знание основных понятий и современных технологий анализа данных. В целом сформировано умение выбора современные информационно-коммуникационные технологии для анализа данных. В целом демонстрирует хорошее владение навыками первичного анализа данных с помощью современных информационно-коммуникационных технологий для анализа данных. <p style="text-align: center;">Отлично</p>

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		<p>Отлично</p> <p>Высокий уровень знания основных понятий и современных технологий анализа данных. Сформировано уверенное умение осознанного выбора современных информационно-коммуникационных технологии для анализа данных. Высокий уровень владения навыками первичного анализа данных с помощью современных информационно-коммуникационных технологий для анализа данных.</p>
<p>ОПК.2.2</p> <p>Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности</p>	<p>Знает и понимает сущность и значение информации и связанных с ней современных технологий в развитии современной науки и общества. Умеет применять ранее полученные знания в решении практических задач. Владеет навыками поиска и самостоятельного освоения новых информационно-коммуникационных технологий. Знает основные принципы, стадии и подходы к решению исследовательских задач с применением современных информационно-коммуникационных технологий. Умеет применять современные инструменты и технологии для решения прикладных задач. Владеет навыками самостоятельного подбора и эффективного использования современных технологий на практике.</p>	<p>Неудовлетворител</p> <p>Не знает и понимает сущность и значение информации и связанных с ней современных технологий в развитии современной науки и общества. Не умеет применять ранее полученные знания в решении практических задач. Не владеет навыками поиска и самостоятельного освоения новых информационно-коммуникационных технологий.</p> <p>Удовлетворительн</p> <p>Не уверенно знает и понимает сущность и значение информации и связанных с ней современных технологий в развитии современной науки и общества. Умеет на удовлетворительном уровне применять ранее полученные знания в решении практических задач. На удовлетворительном уровне владеет навыками поиска и самостоятельного освоения новых информационно-коммуникационных технологий.</p> <p>Хорошо</p> <p>В целом хорошее знание и понимание сущности и значения информации и связанных с ней современных технологий в развитии современной науки и общества. В целом сформировано умение применять ранее полученные знания в решении практических задач. В целом демонстрирует хорошее владение навыками поиска и самостоятельного освоения новых информационно-коммуникационных</p>

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		<p>Хорошо технологий.</p> <p>Отлично Высокий уровень знания и понимания сущности и значения информации и связанных с ней современных технологий в развитии современной науки и общества. Сформировано уверенное умение применения ранее полученных знаний в решении практических задач. Высокий уровень владения навыками поиска и самостоятельного освоения новых информационно-коммуникационных технологий.</p>

Оценочные средства текущего контроля и промежуточной аттестации

Схема доставки : Базовая

Вид мероприятия промежуточной аттестации : Зачет

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 43 до 60

«неудовлетворительно» / «незачтено» менее 43 балла

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ОПК.2.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности ОПК.2.1 Демонстрирует базовые знания в области информационно-коммуникационных технологий	Методы корреляционного анализа Письменное контрольное мероприятие	Знает основные этапы статистического исследования и особенности подготовки данных для статистического исследования, типы и виды данных, шкалы измерения. Умеет вычислять основные описательные статистики. Владеет методами расчета дисперсии, правило сложения дисперсией и его использование в анализе связи. Знает основные принципы проверки статистических гипотез и общую схему решения задачи проверки статистических гипотез. Умеет применять проверку гипотез для реальных задач. Знает и решает задачи на параметрические гипотезы, критерии согласия. Знает основные понятия корреляционного анализа. Умеет рассчитывать и применять корреляционный анализ для различных наблюдений. Владеет навыками параметрического и непараметрического анализа информации. Умеет выбирать основные библиотеки Python для решения поставленных задач.

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
<p>ОПК.2.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности</p> <p>ОПК.2.1 Демонстрирует базовые знания в области информационно-коммуникационных технологий</p>	<p>Классификация: распознавание образов и типологизация</p> <p>Письменное контрольное мероприятие</p>	<p>Знает основные этапы регрессионного анализа данных и прогнозирования, методы кластеризации данных. Умеет применять методы регрессионного и кластерного анализа, интерпретировать полученные результаты с точки зрения сущности явлений. Владеет навыками логического анализа информации. Знает основные библиотеки Python, умеет выбирать и применять библиотеки и функции Python для решения прикладных задач. Владеет навыками самостоятельного подбора и эффективного использования современных технологий на практике.</p>
<p>ОПК.2.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности</p> <p>ОПК.2.1 Демонстрирует базовые знания в области информационно-коммуникационных технологий</p>	<p>Итоговое контрольное мероприятие</p> <p>Итоговое контрольное мероприятие</p>	<p>Знает основные понятия методов статистического анализа данных, первичного анализа информации, корреляционного анализа, методов снижения размерности признаков и кластерного анализа. Умеет применять статистические методы и модели на практике, собирать и проводить статистическую обработку информации с целью выявления основных характеристик совокупности. Умеет осуществлять оценку взаимосвязей различных показателей с помощью статистических методов, интерпретировать полученные результаты по оценке взаимосвязей с точки зрения сущности явлений. Владеет навыками логического анализа информации. Знает основные библиотеки Python, умеет выбирать и применять библиотеки и функции Python для решения прикладных задач. Владеет навыками самостоятельного подбора и эффективного использования современных технологий на практике.</p>

Спецификация мероприятий текущего контроля

Методы корреляционного анализа

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

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

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

Проходной балл: **13**

Показатели оценивания	Баллы
Верно решенное задание (максимальный балл)	30
Верно решенное задание (проходной балл)	13
Верно решенное задание (минимальный балл за 1 задание)	3

Классификация: распознавание образов и типологизация

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

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

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

Проходной балл: **13**

Показатели оценивания	Баллы
Верно решенное задание (максимальный балл)	30
Верно решенное задание (проходной балл)	13
Верно решенное задание (за 1 задание)	3

Итоговое контрольное мероприятие

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

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

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **40**

Проходной балл: **17**

Показатели оценивания	Баллы
Верно решенное задание (максимальный балл)	40
Верно решенное задание (проходной балл)	17
Верно решенное задание (за 1 задание)	3