

МИНОБРНАУКИ РОССИИ

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

Кафедра информационных систем и математических методов в экономике

Авторы-составители: **Арбузов Вячеслав Олегович**

Рабочая программа дисциплины

АНАЛИЗ ДАННЫХ В R

Код УМК 95436

Утверждено
Протокол №9
от «06» июня 2022 г.

Пермь, 2022

1. Наименование дисциплины

Анализ данных в R

2. Место дисциплины в структуре образовательной программы

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

Направление подготовки: **01.03.05** Статистика
направленность Анализ больших данных

3. Планируемые результаты обучения по дисциплине

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

01.03.05 Статистика (направленность : Анализ больших данных)

ОПК.2 Способен формировать упорядоченные сводные массивы статистической информации и осуществлять расчет сводных и производных показателей в соответствии с утвержденными методиками, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ

Индикаторы

ОПК.2.3 Использует практический опыт решения стандартных задач математических и (или) естественных наук

4. Объем и содержание дисциплины

Направления подготовки	01.03.05 Статистика (направленность: Анализ больших данных)
форма обучения	очная
№№ триместров, выделенных для изучения дисциплины	11
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 практических занятий, семинаров	0
Проведение лабораторных работ, занятий по иностранному языку	28
Самостоятельная работа (ак.час.)	66
Формы текущего контроля	Входное тестирование (1) Защищаемое контрольное мероприятие (2) Итоговое контрольное мероприятие (1)
Формы промежуточной аттестации	Зачет (11 триместр)

5. Аннотированное описание содержания разделов и тем дисциплины

Входной контроль

Входной контроль призван проверить знания обучающихся по общим темам связанным с алгоритмизацией и программированием:
переменные, параметры, функции; циклы; классы объектов и их свойства; компиляция кода; алгоритмизация вычислений; вектора, массивы данных;

Тема 1. Введение в программирование на языке R

R — язык программирования для статистической обработки данных и работы с графикой, а также свободная программная среда вычислений с открытым исходным кодом в рамках проекта GNU. Язык создавался как аналогичный языку S, разработанному в Bell Labs, и является его альтернативной реализацией, хотя между языками есть существенные отличия, но в большинстве своём код на языке S работает в среде R.

Тема 2. Анализ данных, построение моделей и визуализация в R

R применяется везде, где нужна работа с данными. Это не только статистика в узком смысле слова, но и «первичный» анализ (графики, таблицы сопряжённости), и продвинутое математическое моделирование. R без особых проблем может использоваться и там, где сейчас принято использовать коммерческие программы анализа уровня MatLab/Octave. С другой стороны вполне естественно, что основная вычислительная мощь R лучше всего его проявляется при статистическом анализе: от вычисления средних величин до вейвлет-преобразований временных рядов.

Тема 3. Модели временных рядов, статистические тесты и панельные данные

В данной теме рассматриваются приемы работы с временными рядами, методы вызова статистических тестов и их интерпретация и базовые основы работы с временными рядами в R

Контрольное мероприятие по синтаксису языка R

Промежуточное контрольное мероприятие для контроля усвоения дисциплины обучающимся

Тема 4. Машинное обучение и построение ансамблевых моделей в R

Основы работы с машинным обучением, в частности изучаются детали и методы работы с библиотекой caret, использование встроенных функций, работа с документацией библиотеки, доступные модели и методы комбинирования

Тема 5. Создание аналитических веб-приложений. Пакет shiny

Основы работы с интерактивными приложениями в R, в частности изучаются детали и методы работы с библиотекой caret, использование встроенных функций, работа с документацией библиотеки, доступный функционал библиотеки и его изменение

Тема 6. Анализ данных с использованием библиотек Tidyverse

Основы работы с семейством библиотек Tidyverse в R, начиная с библиотеки для визуализации ggplot2 и заканчивая деталями и методами работы с библиотеками dplyr, tibble, stringr, использование встроенных функций, работа с документацией библиотек.

Контрольное мероприятие по анализу данных, построению моделей и визуализации в R

Промежуточное контрольное мероприятие для контроля усвоения дисциплины обучающимся

Тема 7. Высокопроизводительные вычисления с использованием R

Основы работы с высокопроизводительными вычислениями, в частности изучаются методы параллельных расчетов, работы с большими данными, ускорения расчетов путем векторизации вычислений, использования специализированных ускоренных библиотек и встраивания C++ кода

(написания функций) в код R

Итоговое контрольное мероприятие

Итоговое контрольное мероприятие по всем темам 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 аудиторские 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. Федин, Ф. О. Анализ данных. Часть 2. Инструменты Data Mining : учебное пособие / Ф. О. Федин, Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 308 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/26445>

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>

Дополнительная:

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>

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

http://rusquant.ru/math_app_1/ Первое знакомство с R

<http://rusquant.ru/занятие-2-базовый-уровень-статистики/> Базовый уровень статистики

<http://rusquant.ru/занятие-7-временные-ряды-и-финансовые-д/> Временные ряды и финансовые данные с R

<http://rusquant.ru/машинное-обучение-с-r/> Машинное обучение с R

<http://rusquant.ru/занятие-8-веб-приложения-с-r/> Веб-приложения с R

<http://rusquant.ru/занятие-4-продвинутая-графика-с-ggplot2/> Графика ggplot2 и Tidyverse

<http://rusquant.ru/занятие-9-hpc-in-r/> Высокопроизводительные вычисления с использованием R

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

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

Образовательный процесс по дисциплине предполагает:

- презентационные материалы (слайды по темам лекционных и практических занятий);
- доступ в режиме on-line в Электронную библиотечную систему (ЭБС);
- доступ в электронную информационно-образовательную среду университета.

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

1. среда разработки языка программирования R (свободное ПО).
2. среда разработки Rstudio (свободное ПО).

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

1. Лекционные занятия - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.
2. Лабораторные занятия - Компьютерный класс, оснащенный персональными ЭВМ и соответствующим программным обеспечением. Состав оборудования определен в Паспорте компьютерного класса.
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 или аудитория, оснащенная 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 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 в R**

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

ОПК.2

Способен формировать упорядоченные сводные массивы статистической информации и осуществлять расчет сводных и производных показателей в соответствии с утвержденными методиками, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ

Индикатор	Планируемые результаты обучения	Критерии оценивания результатов обучения
<p>ОПК.2.3 Использует практический опыт решения стандартных задач математических и (или) естественных наук</p>	<p>Знать • основные типы данных и программные конструкции в R; • методы генерирования случайных величин с заданным законом распределения; • виды имитационных моделей; • о поддержке работы в R с регрессионными моделями и моделями временных рядов; • о поддержке в R работы с объектами нечисловой природы. Уметь • создавать функции и сценарии в R; • реализовывать в R детерминированные и стохастические имитационные модели с дискретным и непрерывным временем, а также гибридные модели; • планировать эксперимент; • обрабатывать результаты вычислений; • работать в R с регрессионными моделями — оценивать параметры, проверять статистические гипотезы, выполнять прогнозирование; • работать в R с моделями временных рядов; • работать с моделями, содержащими объекты нечисловой природы. Владеть • основными приемами работы в статистическом пакете R; • методами решения уравнений и</p>	<p align="center">Неудовлетворител</p> <p>Не умеет создавать функции и сценарии в языках R; Не владеет навыками обработки данных в R; Не умеет импортировать и экспортировать данные в R; Не умеет обрабатывать результаты вычислений; Не умеет визуализировать результаты вычислений; Не умеет работать в R с регрессионными моделями — оценивать параметры, проверять статистические гипотезы, выполнять прогнозирование; Не умеет работать в R с моделями временных рядов; Не умеет работать со стохастическими процессами и оценивать их параметры; Не умеет строить ансамблиевые модели и проводить их валидацию на обучающей и тестовой выборке;</p> <p align="center">Удовлетворительн</p> <p>Умеет создавать функции и сценарии в R; Владеет навыками обработки данных в R; Умеет импортировать и экспортировать данные в R; Не умеет обрабатывать результаты вычислений; Не умеет визуализировать результаты вычислений; Не умеет работать в R с регрессионными моделями — оценивать параметры, проверять статистические гипотезы, выполнять прогнозирование; Не умеет работать в R с моделями временных рядов; Не умеет работать со стохастическими процессами и оценивать их параметры; Не умеет строить ансамблиевые модели и проводить их валидацию на обучающей и тестовой выборке;</p>

Индикатор	Планируемые результаты обучения	Критерии оценивания результатов обучения
	<p>задач математического программирования; • основными методами статистической обработки результатов вычислительного эксперимента.</p>	<p>Хорошо Умеет создавать функции и сценарии в R; Владеет навыками обработки данных в R; Умеет импортировать и экспортировать данные в R; Умеет обрабатывать результаты вычислений; Умеет визуализировать результаты вычислений; Умеет работать в R с регрессионными моделями — оценивать параметры, проверять статистические гипотезы, выполнять прогнозирование; Не умеет работать в R с моделями временных рядов; Не умеет работать со стохастическими процессами и оценивать их параметры; Не умеет строить ансамблиевые модели и проводить их валидацию на обучающей и тестовой выборке;</p> <p>Отлично Умеет создавать функции и сценарии в R; Владеет навыками обработки данных в R; Умеет импортировать и экспортировать данные в R; Умеет обрабатывать результаты вычислений; Умеет визуализировать результаты вычислений; Умеет работать в R с регрессионными моделями — оценивать параметры, проверять статистические гипотезы, выполнять прогнозирование; Умеет работать в R с моделями временных рядов; Умеет работать со стохастическими процессами и оценивать их параметры; Умеет строить ансамблиевые модели и проводить их валидацию на обучающей и тестовой выборке;</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 50 до 60

«неудовлетворительно» / «незачтено» менее 50 балла

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
Входной контроль	Входной контроль Входное тестирование	Входное тестирование представляет из себя тест состоящий из 20 вопросов и задач по общим темам связанным с алгоритмизацией и программированием: - переменные, параметры, функции;- циклы;- классы объектов и их свойства;- компиляция кода;- алгоритмизация вычислений;- вектора, массивы данных;
ОПК.2.3 Использует практический опыт решения стандартных задач математических и (или) естественных наук	Контрольное мероприятие по синтаксису языка R Защищаемое контрольное мероприятие	Практическое самостоятельное задание связанное с пониманием синтаксиса языка R и решением практического задания по работе с данными с использованием языка R

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
<p>ОПК.2.3 Использует практический опыт решения стандартных задач математических и (или) естественных наук</p>	<p>Контрольное мероприятие по анализу данных, построению моделей и визуализации в R Защищаемое контрольное мероприятие</p>	<p>Практическое самостоятельное задание связанное с анализом данных, построением моделей и визуализацией в R. Всего предлагается 14 заданий, каждое задание оценивается по 1 баллу. Итоговое количество баллов подсчитывается и выставляется итоговая оценка за контрольное мероприятие. Примерный список заданий для самостоятельного выполнения: 1. Построить гистограмму на основе случайного распределения Стьюдента 2. Выбрать 5 столбец, 3 строку из набора данных mtcars 3. Удалить 7 столбец из набора данных mtcars 4. Удалить 1 строку из набора данных mtcars 5. Экспортировать данные mtcars в txt файл, а затем импортировать обратно. Отсортировать автомобили по мощности в порядке возрастания 7. Выбрать все автомобили, которые имеют 6 цилиндров из набора данных mtcars 8. Вычислить средние значения показателей по столбцам из набора данных mtcars 9. Вычислить среднюю мощность автомобиля в зависимости от количества цилиндров mtcars 10. Построить диаграмму размаха зависимости средней мощности автомобиля от количества цилиндров mtcars 11. На основе данных mtcars построить линейную регрессию зависимости расхода топлива от прочих факторов 12. Кластеризовать все автомобили из набора данных mtcars на основе метода k-средних 13. Построить дерево решений зависимости мощности от прочих факторов 14. Визуализировать одновременно любые 6 измерений данных mtcars (используя оси XY, цвет,</p>

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ОПК.2.3 Использует практический опыт решения стандартных задач математических и (или) естественных наук	Итоговое контрольное мероприятие Итоговое контрольное мероприятие	размер, тип и прозрачность) Основы синтаксиса языка программирования R, основные структуры языка. Особенности применения языка R, специализированных пакетов в нем для автоматизации обработки и анализа экономических данных, сложных экономических расчетов Решение нестандартных задач связанные с вычислениями, анализ ошибок возникающих в коде, самостоятельное нахождение пути решения устранения этих ошибок. Использование продвинутых методов решения задач обработки и анализа экономических данных, позволяющих минимизировать время исполнения кода.

Спецификация мероприятий текущего контроля

Входной контроль

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

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

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **0**

Проходной балл: **0**

Показатели оценивания	Баллы
От шестнадцати до двадцати правильных ответов – 100 баллов	100
От одиннадцати до пятнадцати правильных ответов – 70 баллов	70
От восьми до десяти правильных ответов – 50 баллов	50
От одного до семи правильных ответов – 20 баллов	20

Контрольное мероприятие по синтаксису языка R

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

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

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

Проходной балл: **15**

Показатели оценивания	Баллы
Знает основы синтаксиса языка программирования. Знает особенности применения специализированных пакетов и умеет решать, автоматизировать расчеты сложных задач.	20

Способен решать нестандартные задачи связанные с вычислениями, проводить анализ ошибок возникающих в коде и самостоятельно находить пути решения устранения этих ошибок. Умеет пользоваться продвинутыми методами решения задач, позволяющих минимизировать время исполнения кода.	
Знает основы синтаксиса языка программирования. Знает особенности применения специализированных пакетов и умеет решать и автоматизировать расчеты сложных задач.	15
Знает основы синтаксиса языка программирования и способен решать простейшие задачи в указанной теме	10
Не знает основ синтаксиса языка программирования и не способен решать простейшие задачи в указанной теме	0

Контрольное мероприятие по анализу данных, построению моделей и визуализации в R

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

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

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

Проходной балл: **15**

Показатели оценивания	Баллы
Знает основы синтаксиса языка программирования. Знает особенности применения специализированных пакетов и умеет решать, автоматизировать расчеты сложных задач. Способен решать нестандартные задачи связанные с вычислениями, проводить анализ ошибок возникающих в коде и самостоятельно находить пути решения устранения этих ошибок. Умеет пользоваться продвинутыми методами решения задач, позволяющих минимизировать время исполнения кода.	20
Знает основы синтаксиса языка программирования. Знает особенности применения специализированных пакетов и умеет решать и автоматизировать расчеты сложных задач.	15
Знает основы синтаксиса языка программирования и способен решать простейшие задачи в указанной теме	10
Не знает основ синтаксиса языка программирования и не способен решать простейшие задачи в указанной теме	0

Итоговое контрольное мероприятие

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

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

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

Проходной балл: **20**

Показатели оценивания	Баллы
Знает основы синтаксиса языка программирования. Знает особенности применения специализированных пакетов и умеет решать, автоматизировать расчеты сложных задач. Способен решать нестандартные задачи связанные с вычислениями, проводить анализ ошибок возникающих в коде и самостоятельно находить пути решения устранения этих ошибок. Умеет пользоваться продвинутыми методами решения задач, позволяющих	60

минимизировать время исполнения кода.	
Знает основы синтаксиса языка программирования. Знает особенности применения специализированных пакетов и умеет решать и автоматизировать расчеты сложных задач.	40
Знает основы синтаксиса языка программирования и способен решать простейшие задачи в указанной теме	20
Не знает основ синтаксиса языка программирования и не способен решать простейшие задачи в указанной теме	0