

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Писарев Сергей Станиславович
Должность: Ректор
Дата подписания: 08.12.2029 15:07:46
Уникальный программный ключ:
b9d7463b91f434da3d4dc1afa9a0cf32d3c58650

Негосударственное образовательное учреждение высшего образования
«Школа управления СКОЛКОВО»

Утверждено
ректор С.С. Писарев
ШКОЛА УПРАВЛЕНИЯ
СКОЛКОВО
“19” августа 2024 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Введение в анализ данных

Направление подготовки	38.03.02 Менеджмент
Квалификация выпускника	Бакалавр
Образовательная программа	Управление и предпринимательство
Форма обучения	Очная
Рабочая программа дисциплины разработана	

Трудоемкость		Контактная работа		Самостоятельная работа	Форма контроля	Семестр/кв артиль
з.е.	часы	лекции	семинарские занятия			
2	72	16	16	40	Экзамен	2/3

1. АННОТАЦИЯ ДИСЦИПЛИНЫ

Данный курс знакомит студентов с фундаментальными принципами и практиками анализа данных. Он охватывает полный жизненный цикл анализа данных: от сбора и обработки до анализа, визуализации и интерпретации. Студенты познакомятся с ключевыми понятиями описательной статистики и базовыми методами анализа данных. Курс включает практическую работу с современными аналитическими инструментами, что позволяет студентам развить навыки, необходимые для решения реальных задач с использованием подходов, основанных на данных.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В случае успешного освоения курса студенты будут:

знать

- ключевые понятия, определения и задачи анализа данных;
- типы данных и их характеристики;
- основные принципы математической статистики, используемые в анализе данных;
- основные методы анализа данных;
- обзор современных программных инструментов для анализа данных и области их применения;
- ключевые этические аспекты и вопросы безопасности при работе с данными.

уметь

- собирать данные из различных источников, включая порталы открытых данных;
- выполнять первичную обработку данных: очищать, преобразовывать и подготавливать данные для анализа;
- проводить описательный статистический анализ выборок данных;
- применять базовые методы анализа данных для решения практических задач;
- выбирать подходящие статистические методы и инструменты для анализа конкретных типов данных и задач;
- визуализировать данные и результаты анализа с использованием современных программных инструментов для эффективного представления;
- интерпретировать полученные результаты и формулировать выводы на основе анализа данных.

владеть

- навыками использования программного обеспечения для анализа данных для статистических расчетов и визуализации;
- навыками применения статистических методов для анализа объемов данных;
- навыками подготовки отчетов и презентаций по результатам анализа данных;
- навыками критической оценки достоверности и адекватности полученных моделей и результатов анализа.

Дисциплина направлена на развитие следующих компетенций и их индикаторов:

Код компетенции	Формулировка компетенции и/или ее индикатора (ов)
ОПК-5.	Способен использовать при решении профессиональных задач современные информационные технологии и программные средства, включая управление крупными массивами данных и их интеллектуальный анализ
ОПК-5-1.	Использует цифровые технологии, включая информационные системы и

	базы данных, системы искусственного интеллекта и системы интеллектуального анализа и обработки данных для решения профессиональных задач
ОПК-5-2.	Оценивает возможности и целесообразность использования цифровых технологий в деятельности организации, использует современные цифровые технологии и программные продукты для решения профессиональных задач
ОПК-6.	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности
ОПК-6-1.	Знает процессы, методы поиска, сбора, хранения, обработки, представления, распространения информации и способы осуществления таких процессов и методов; современные инструментальные среды, программно-технические платформы и программные средства
ОПК-6-2.	Рационально выбирает современные информационные технологии для решения задач профессиональной деятельности

3. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Название раздела/темы	Всего часов	Трудоемкость (час.) по видам учебных занятий			
		Контактная работа			Самостоятельная работа
		Всего	Лекции	Семинары	
Тема 1. Введение в анализ данных	11	5	3	2	6
Тема 2. Основы технологий хранения и обработки данных	12	6	4	2	6
Тема 3. Программные средства для анализа данных	13	5	3	2	8
Тема 4. Статистические методы анализа данных	11	5	3	2	6
Тема 5. Процессы сбора и подготовки данных	11	5	3	2	6
Тема 6. Практикум по анализу данных	14	6	0	6	8
Итого	72	32	16	16	40

Тема 1. Введение в анализ данных

Основные понятия и терминология анализа данных. Жизненный цикл анализа данных. Data Mining. Типы данных и их характеристики. Основные задачи и цели анализа данных. Роль анализа данных в современном бизнесе. Этические аспекты и безопасность данных

Тема 2. Основы технологий хранения и обработки данных

Понятие больших данных и их характеристики (3V-модель). Технологии хранения данных: базы данных и СУБД. Реляционные и нереляционные базы данных. Системы распределенного хранения данных. Основные модели данных и их особенности. Принципы организации хранилищ данных. Подготовка исходных данных для анализа.

Тема 3. Программные средства для анализа данных

Обзор современных инструментов анализа данных. Коммерческое ПО: Statistica, SPSS, SAS. Бесплатные инструменты: R, Python, KNIME. Электронные таблицы (Excel) в анализе данных. Сравнительный анализ программных средств. Критерии выбора инструментов для различных задач.

Тема 4. Статистические методы анализа данных

Основные понятия математической статистики. Основы описательной статистики. Проверка статистических гипотез. Корреляционный и регрессионный анализ. Дисперсионный анализ. Кластерный и факторный анализ. Временные ряды и прогнозирование.

Тема 5. Процессы сбора и подготовки данных

Методы сбора данных из различных источников. Работа с открытыми данными и API. Оценка качества и достоверности данных. Техники очистки и трансформации данных. Работа с пропущенными значениями и выбросами. Создание признаков и преобразование переменных.

Тема 6. Практикум по анализу данных

Представление исходных данных в RStudio (векторы, массивы, матрицы, списки, таблицы). Статистическая обработка данных в MS Excel и RStudio: вычисление описательной статистики, графическое представление данных. Группировка данных, выявление значимых зависимостей и трендов посредством анализа доступной информации, выявление взаимосвязей между различными типами данных. Применение различных методов извлечения, поиска и группировки данных, позволяющих выявить системные структуры данных и вывести правила для принятия решений и прогнозирования их последствий (регрессия, дисперсионный анализ, кластерный, дискриминантный, факторный анализы). Возможности графического представления информации в RStudio: графические функции для отображения одномерных и многомерных данных, графический вывод с использованием графических параметров.

4. ОЦЕНОЧНЫЕ СРЕДСТВА И ПРИМЕРЫ ЗАДАНИЙ ДЛЯ ОЦЕНКИ РЕЗУЛЬТАТОВ ОСВОЕНИЯ ДИСЦИПЛИНЫ

4.1 Текущий контроль

Оценка за курс складывается из следующих видов заданий текущего контроля, каждый из которых обладает своим весом в общей системе:

Компоненты	Процент в итоговой оценке
Устные опросы	20%
Практические задания	60%
Итоговый тест	20%

На курсе используется 10 балльная система оценивания. За каждое задание студент получает от 1 до 10 баллов. Итоговый балл за каждый вид заданий рассчитывается как среднее арифметическое всех полученных баллов за все задания в рамках одного вида (O1, O2, O3). Невыполненное в срок задание оценивается в 0 баллов.

Общая оценка за курс (O) рассчитывается как:

$$O = O1 \times 0,2 + O2 \times 0,6 + O3 \times 0,2$$

Если по результатам текущего контроля студент получил положительную оценку (не ниже «удовлетворительно»), оценка за промежуточную аттестацию выставляется автоматически.

Устные опросы

Опрос проводится после изучения соответствующей темы. Преподаватель задает студентам вопросы устно во время лекций. Цель – обеспечить постоянное взаимодействие с материалом и оценить понимание ключевых теоретических концепций из лекций и материалов для чтения.

Практические задания

Практические задания с использованием программного обеспечения (RStudio, Excel) для обработки, анализа и визуализации данных. Портфолио из 6 заданий, сдаваемых в виде отчетов или файлов скриптов. Каждый отчет должен включать использованный код/команды, вывод (таблицы, графики) и краткую письменную интерпретацию результатов. Цель – развить и оценить практические навыки работы с данными, их анализа и интерпретации с использованием актуальных инструментов.

Итоговый тест

Комплексный письменный тест, охватывающий все материалы курса, включая теоретические вопросы и практические задачи.

4.2 Промежуточная аттестация

Студентам, набравшим достаточные для удовлетворительной оценки баллы за текущий контроль, оценка за дисциплину выставляется равной оценке за текущий контроль (См. п. 4.1)

Студентам, получившим неудовлетворительную оценку по результатам текущего контроля, необходимо по согласованию с преподавателем сдать один или несколько компонентов текущего контроля. Преподаватель вправе предложить студентам выполнить задание, не повторяющее задание текущего контроля, но проверяющее аналогичные знания, умения и навыки.

4.3 Примеры заданий

Примеры вопросов для устных опросов

1. Из каких ключевых этапов состоит процесс информационного поиска?
2. Что понимается под термином «Data Mining»?
3. В чем заключается основная суть концепции «большие данные»?
4. Какие технологии используются для хранения больших массивов данных?
5. Приведите примеры коммерческого и свободного программного обеспечения для анализа данных.
6. Назовите известные вам модели данных.
7. Какие операции входят в этап первичной обработки данных?
8. Какие критерии применяются для оценки статистической значимости различий между выборками?
9. В чем состоит главная задача корреляционного анализа?

10. Какую основную цель преследует применение регрессионного анализа?
11. Какие виды регрессионных моделей вам известны?
12. Какими способами можно оценить адекватность построенной регрессионной модели?
13. В чем состоит основная идея кластерного анализа?

Примеры практических заданий

1. Сбор данных из различных источников в Интернете. Сохранение данных в программе MS Excel. Первичный анализ данных и визуализация данных в Excel.
2. Выборка и преобразование исходных данных в RStudio. Удаление пропущенных значений.
3. Сравнение выборок. t-критерий Стьюдента, критерий хи-квадрат Пирсона, критерий Колмогорова-Смирнова.
4. Корреляционный анализ. Вычисление коэффициентов корреляции Пирсона, Спирмена, Кендалла.
5. Регрессионный анализ (линейная зависимость). Построение линейной модели. Проверка адекватности построенной модели.
6. Регрессионный анализ (нелинейная зависимость). Определение типа зависимости. Построение модели. Проверка адекватности построенной модели.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1 Литература

1. Анализ данных : учебник для вузов / под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2025. — 448 с. — (Высшее образование). — ISBN 978-5-534-19964-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/560311>
2. Миркин, Б. Г. Базовые методы анализа данных : учебник и практикум для вузов / Б. Г. Миркин. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 297 с. — (Высшее образование). — ISBN 978-5-534-19709-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/560414>
3. Митина, О. А. Бизнес-аналитика. Введение в обработку и анализ данных : учебник для вузов / О. А. Митина. — Москва : Издательство Юрайт, 2025. — 172 с. — (Высшее образование). — ISBN 978-5-534-21811-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/582199>

5.2 Электронные образовательные ресурсы

Материалы дисциплины размещены в LMS: <https://l.skolkovo.ru/login/index.php>

6. ЛИЦЕНЗИОННОЕ И СВОБОДНО РАСПРОСТРАНЯЕМОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Операционная система Simple Linux, браузер Yandex браузер, антивирусное ПО Calmantivirus;

Свободно распространяемое ПО, в том числе отечественного производства:

Офисный пакет Libre Office, Okular PDF Reader, 7-Zip Архиватор, GIMP Редактирования фотографий, Inkscape Векторная графика, Blender 3D графика, Kdenlive Видеоредактор, Audacity Аудиоредактор, VLC Медиаплеер, Thunderbird Почтовый клиент, Flameshot Создание скриншотов

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебная аудитория для проведения занятий лекционного типа, оснащенная мультимедийным оборудованием, учебной мебелью, доской или со стенами с маркерным покрытием.

Учебная аудитория для проведения занятий семинарского типа, оснащенная мультимедийным оборудованием, учебной мебелью, доской или со стенами с маркерным покрытием.

Аудитория (коворкинг) для самостоятельной работы, оснащенная учебной мебелью, ноутбуками.

Материально-техническое обеспечение аудиторий представлено на официальном сайте <https://bbask.ru/sveden/objects/>.