

**Программа учебной дисциплины  
«Анализ данных в R»**

Утверждена  
Академическим советом ОП  
Протокол №\_\_\_\_\_ от \_\_\_\_\_.\_\_\_\_\_.20\_\_\_\_\_

Разработчик	Зарманбетов Ахмед Курманалиевич, Другой факультет (не ФКН)
Число кредитов	6
Контактная работа (час.)	64
Самостоятельная работа (час.)	164
Курс, Образовательная программа	4 (Б) курс, Социология
Формат изучения дисциплины	Без использования онлайн курса

**1. Цель, результаты освоения дисциплины и пререквизиты**

Цели:

1. Освоение такого инструмента как язык программирования R для анализа данных, начиная с импорта и подготовки данных до построения модели.

Планируемые результаты обучения (ПРО):

1. Приобрести необходимые навыки программирования: уметь работать с переменными, массивами и матрицами; уметь использовать циклы for и while, а также условный оператор if. Уметь писать собственные функции.
2. Научиться писать отчеты и доклады, совмещая текст, код и графики в одном документе.
3. Уметь преобразовывать и очищать данные.
4. Уметь подгружать данные из разных источников: интернет, файлы разных форматов.
5. Уметь визуализировать данные.
6. Уметь строить модели для задач регрессии и классификации.

**2. Содержание учебной дисциплины**

Тема (раздел дисциплины)	Объем в часах	Планируемые результаты обучения (ПРО), подлежащие контролю	Формы контроля
	лк		
	см		
	онл/сп		
Введение в R	2		

	4	<ul style="list-style-type: none"> <li>• Приобрести необходимы навыки программирования: уметь работать с переменными, массивами и матрицами; уметь использовать циклы for и while, а также условный оператор if. Уметь писать собственные функции.</li> </ul>	Д31, КР1.
	16		
Основы программирования	2	<ul style="list-style-type: none"> <li>• Приобрести необходимы навыки программирования: уметь работать с переменными, массивами и матрицами; уметь использовать циклы for и while, а также условный оператор if. Уметь писать собственные функции.</li> </ul>	Д31, КР1.
	6		
	16		
RMarkdown	0	<ul style="list-style-type: none"> <li>• Научиться писать отчеты и доклады, совмещая текст, код и графики в одном документе.</li> </ul>	Д32, Д33, Д34, Д35, КР1, Экз.
	4		
	10		
Данные	2	<ul style="list-style-type: none"> <li>• Уметь преобразовывать и очищать данные.</li> </ul>	Д32, КР1.
	4		
	10		
Импорт данных	0	<ul style="list-style-type: none"> <li>• Уметь подгружать данные из разных источников: интернет, файлы разных форматов.</li> </ul>	Д32, КР1.
	4		
	14		
Работа с данными	0	<ul style="list-style-type: none"> <li>• Уметь преобразовывать и очищать данные.</li> </ul>	Д32, КР1.
	4		
	16		
Работа с различными типами переменных	0	<ul style="list-style-type: none"> <li>• Уметь преобразовывать и очищать данные.</li> </ul>	Д32, КР1.
	4		
	10		
Визуализация данных	0	<ul style="list-style-type: none"> <li>• Уметь визуализировать данные.</li> </ul>	Д33, Экз.
	4		
	18		
Данные из интернета	2	<ul style="list-style-type: none"> <li>• Уметь подгружать данные из разных источников: интернет, файлы разных форматов.</li> </ul>	Д32, Д34, Д35, Экз.
	6		
	20		
Модель линейной регрессии	2	<ul style="list-style-type: none"> <li>• Уметь строить модели для задач регрессии и классификации.</li> </ul>	Д35, Экз.
	6		
	20		
Логистическая регрессия	2		Д35, Экз.
	6		

	14	<ul style="list-style-type: none"> <li>Уметь строить модели для задач регрессии и классификации.</li> </ul>	
<b>Часов по видам учебных занятий:</b>	12		
	52		
	164		
<b>Итого часов:</b>	228		

### *Содержание разделов дисциплины:*

#### 1. **Введение в R**

История языка программирования R и для чего он используется. Что такое IDE. Установка R и RStudio. Интерфейс RStudio. Работа с консолью и первый скрипт. Арифметические операции, переменные, массивы.

#### 2. **Основы программирования**

Циклы for и while, условный оператор if. Работа с матрицами. Функции. Рекурсия. Алгоритм сортировки пузырьком.

#### 3. **RMarkdown**

Оформление докладов, отчетов. Интегрирование кода, графиков и таблиц в отчеты. Связь с LaTeX.

#### 4. **Данные**

Как выглядит pipeline анализа данных. Типы переменных. Data Frame. List.

#### 5. **Импорт данных**

Различные источники импорта данных(базы данных, интернет, файлы). Использование пакетов: их установка и подгрузка. Импорт файлов формата .csv, .txt., .xlsx и других.

#### 6. **Работа с данными**

Очистка и преобразование данных. Работа с несколькими фреймами данных.

#### 7. **Работа с различными типами переменных**

Работа с переменными: дата/время, строковые переменные и категориальные(фактурные) переменные.

#### 8. **Визуализация данных**

Визуализация данных с помощью пакета ggplot2. Интерактивные графики с помощью пакета plotly.

#### 9. **Данные из интернета**

Импорт данных из интернета. Принципы работы с API. Выкачивание данных из VK с помощью API и построение графа друзей. Скраппинг html страниц. Пакет RSelenium. Создание робота для выкачивания цен на авиабилеты.

#### 10. **Модель линейной регрессии**

Ликбез по теории вероятности и математической статистики. Что представляет собой линейная регрессия. Обучение линейной регрессии. Бинарные переменные. Оценки и свойства коэффициентов линейной регрессии. Доверительные интервала и t-тест.

## 11. Логистическая регрессия

Задача классификации. Оценка логистической регрессии (метод максимального правдоподобия).

### 3. Оценивание

- ДЗ1, Не блокирующее, Домашнее задание
- ДЗ2, Не блокирующее, Домашнее задание
- ДЗ3, Не блокирующее, Домашнее задание
- ДЗ4, Не блокирующее, Домашнее задание
- ДЗ5, Не блокирующее, Домашнее задание
- КР1, Не блокирующее, Контрольная работа
- Экз, Не блокирующее, Экзамен (письменный)

**Формула округления:** Стандартное арифметическое округление

**Вид формулы оценивания:** Линейная

**Формула оценивания:**

Окончательная оценка = Округление( $0.1 * ДЗ1 + 0.1 * ДЗ2 + 0.1 * ДЗ3 + 0.1 * ДЗ4 + 0.1 * ДЗ5 + 0.2 * КР1 + 0.3 * Экз$ )

Академическая этика

Преподаватель оставляет за собой право устроить устную защиту любой из форм контроля. Если при проверке работ (текущий и итоговый контроль) установлен факт нарушения академической этики, студент получает оценку «0» за данную работу. Работа студента, предоставившего свою работу для списывания, также аннулируется, к обоим студентам применяется дисциплинарное взыскание.

В случае нарушения правил проведения экзамена студент удаляется с экзамена с оценкой «0». К нарушениям правил проведения экзамена относятся: коммуникация с другими студентами во время выполнения работы, использование социальных сетей/телефона во время экзамена (с любой целью), списывание.

### 4. Примеры оценочных средств

Блокирующие элементы не предусмотрены.

Примеры задач:

1. Написать функцию, которая удваивает все четные или нечетные элементы (в зависимости от параметра) в массиве в  $n$  раз.
2. Импортировать .csv файл, отобрать необходимые наблюдения и признаки. Вычислить основные статистики признаков (среднее, медиана, дисперсия).

3. Визуализировать предоставленные данные: гистограмму распределения признаков, диаграмму рассеивания, временной ряд.
4. Выгрузить информацию о друзьях в VK и построить граф друзей.
5. Построить модель регрессии или классификации и проинтерпретировать результаты.

При отправлении работы после дедлайна, количество баллов по ней уменьшается на 50%.

При отсутствии на КР1 по уважительной причине, ее можно будет написать в другой день.

## 5. Ресурсы

### 5.1. Рекомендуемая основная литература

п/п	Наименование
1	<a href="#">Wickham, H. R for data science : import, tidy, transform, visualize, and model data, O'Reilly, 2017 г.</a>

### 5.2. Рекомендуемая дополнительная литература

п/п	Наименование
1	<a href="#">Кабаков Р. И. R в действии : анализ и визуализация данных в программе R, ДМК Пресс, 2014 г.</a>

### 5.3. Программное обеспечение

п/п	Наименование	Условия доступа/скачивания
1	Microsoft Windows 7 Professional RUS Microsoft Windows 8.1 Professional RUS Microsoft Windows 10	<i>Из внутренней сети университета (договор)</i>
2	Microsoft Office Professional Plus 2010	<i>Из внутренней сети университета (договор)</i>
3	R	Открытый продукт(open source). <a href="https://www.r-project.org">https://www.r-project.org</a>
4	RStudio	Открытый продукт(open source). <a href="https://www.rstudio.com">https://www.rstudio.com</a>

### 5.4. Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

п/п	Наименование	Условия доступа/скачивания
	<b><i>Профессиональные базы данных, информационно-справочные системы</i></b>	
1	Электронно-библиотечная система Юрайт	URL: <a href="https://biblio-online.ru/">https://biblio-online.ru/</a>
	<b><i>Интернет-ресурсы (электронные образовательные ресурсы)</i></b>	
1	Открытое образование	URL: <a href="https://openedu.ru/">https://openedu.ru/</a>

### 5.5. Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

- ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);
- мультимедийный проектор с дистанционным управлением.

Учебные аудитории для семинарских и самостоятельных занятий по дисциплине оснащены ПЭВМ, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.

Компьютерные классы оборудованы ПЭВМ с доступом в Интернет, операционными системами и программным обеспечением, необходимыми для освоения дисциплины. При необходимости допускается замена оборудования его виртуальными аналогами.

## **6. Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов**

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) а для инвалидов также в соответствии с индивидуальной программой реабилитации инвалида, могут предлагаться следующие варианты восприятия учебной информации с учетом их индивидуальных психофизических особенностей, в том числе с применением электронного обучения и дистанционных технологий:

6.1.1. *для лиц с нарушениями зрения:* в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); в печатной форме на языке Брайля; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации.

6.1.2. *для лиц с нарушениями слуха:* в печатной форме; в форме электронного документа; видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации.

6.1.3. *для лиц с нарушениями опорно-двигательного аппарата:* в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.