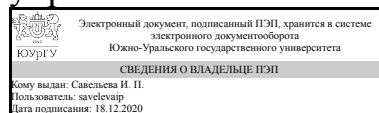


# ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:  
Директор института  
Высшая школа экономики и  
управления



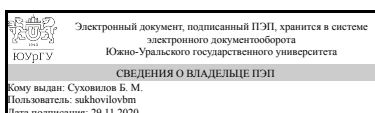
И. П. Савельева

## РАБОЧАЯ ПРОГРАММА

дисциплины Б.1.12 Анализ данных  
для направления 38.03.05 Бизнес-информатика  
уровень бакалавр тип программы Академический бакалавриат  
профиль подготовки  
форма обучения очная  
кафедра-разработчик Информационные технологии в экономике

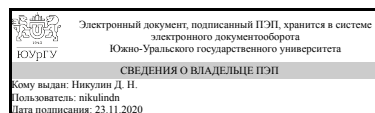
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 38.03.05 Бизнес-информатика, утверждённым приказом Минобрнауки от 11.08.2016 № 1002

Зав.кафедрой разработчика,  
Д.ТЕХН.Н., СИС



Б. М. Суховилов

Разработчик программы,  
старший преподаватель



Д. Н. Никулин

## 1. Цели и задачи дисциплины

Целью дисциплины "Анализ данных" является формирование у студентов знаний, умений и навыков в области математической статистики, компьютерного анализа данных и технологий извлечения знаний. Задачи дисциплины "Анализ данных": - изучение основных методов обработки структурированных данных; - приобретение опыта выявления закономерностей в массивах данных; - освоение технологий Data mining; - развитие навыков применения компьютерных программ для анализа данных.

## Краткое содержание дисциплины

Методы анализа количественных, порядковых и номинальных данных. Статистические показатели и графики, отражающие основные свойства наборов данных. Распространение результатов выборочного исследования на генеральную совокупность. Построение доверительных интервалов и проверка статистических гипотез. Непараметрические критерии для проверки статистической значимости. Сравнение зависимых и независимых выборок. Корреляционный и регрессионный анализ. Кластерный, факторный и дискриминантный анализ. Алгоритмы Data Mining. Языки Python и R.

## 2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине (ЗУНы)
ПК-17 способность использовать основные методы естественнонаучных дисциплин в профессиональной деятельности для теоретического и экспериментального исследования	Знать: классификацию основных типов данных; типологию методов анализа данных; основные алгоритмы Data Mining
	Уметь: использовать среду R для статистического анализа данных; обрабатывать количественные, порядковые и номинальные качественные данные; применять методы анализа одномерных и многомерных данных
	Владеть: методами построения репрезентативных выборок, оценки параметров генеральной совокупности; проверки статистических гипотез
ПК-18 способность использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования	Знать: свойства нормального и биномиального распределений; особенности исходных структур данных для применения одномерных и многомерных методов анализа
	Уметь: вычислять простые и взвешенные показатели центральной тенденции и изменчивости; выполнять проверку статистической значимости корреляционных связей и различий между выборками
	Владеть: методами проверки статистической значимости и многомерного анализа данных в среде R

## 3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Б.1.14 Информатика, Б.1.10 Теория вероятностей и математическая статистика	ДВ.1.07.02 OLAP технологии, ДВ.1.07.01 Хранилища данных, ДВ.1.08.02 Информационные системы поддержки принятия решений

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Дисциплина	Требования
Б.1.14 Информатика	Необходимо знать структуру и возможности программного обеспечения, позволяющего выполнять хранение, обработку и передачу информации; уметь использовать технологии работы с данными различных типов; владеть навыками обработки данных в табличных процессорах и представления результатов обработки в офисных пакетах
Б.1.10 Теория вероятностей и математическая статистика	Необходимо знать основы теории вероятностей и математической статистики; уметь пользоваться расчетными формулами, таблицами, графиками при решении статистических задач; владеть методами оценки доверительных интервалов и проверки гипотез

#### 4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 4 з.е., 144 ч.

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		5
Общая трудоёмкость дисциплины	144	144
<i>Аудиторные занятия:</i>	64	64
Лекции (Л)	32	32
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	32	32
Лабораторные работы (ЛР)	0	0
<i>Самостоятельная работа (СРС)</i>	80	80
Подготовка к экзамену по разделу "Общие сведения о языках Python и R"	20	20
Подготовка к экзамену по разделу "Первичные статистики и базовые статистические методы"	20	20
Подготовка к экзамену по разделу "Библиотеки Python и R для визуализации данных"	20	20
Подготовка к экзамену по разделу "Многомерные методы и Data Mining"	20	20
Вид итогового контроля (зачет, диф.зачет, экзамен)	-	экзамен

## 5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Общие сведения о языках Python и R	16	8	8	0
2	Первичные статистики и базовые статистические методы	16	8	8	0
3	Библиотеки Python и R для визуализации данных	16	8	8	0
4	Многомерные методы и Data Mining	16	8	8	0

### 5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Типы данных в библиотеках Python и R	4
2	1	Встроенные и пользовательские функции в Python и R. Математические, статистические и текстовые функции. Определение пользовательских функций	4
3	2	Меры центральной тенденции. Меры изменчивости. Распределение значений переменной	4
4	2	Статистические критерии проверки гипотез. Проверка уровня значимости. Построение доверительных интервалов.	4
5	3	Графические функции высокого уровня. Параметры диаграмм, осей, надписей, легенд и текстовых аннотаций	4
6	3	Столбчатые и круговые диаграммы. Гистограммы. Диаграммы размахов. Точечные диаграммы.	4
7	4	Многомерные регрессионные модели. Выбор оптимального набора предикторов линейной модели. Мультиколлинеарность. Нелинейные модели. Деревья регрессии. Сравнение построенных моделей и оценка информативности предикторов.	4
8	4	Модели классификации. Дискриминантный анализ. Деревья классификации. Логистическая регрессия. Метод k-ближайших соседей. Байесовский классификатор. Нелинейные классификаторы. Модель мультиномиального логита. Классификация на базе нейронных сетей.	4

### 5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1	1	Рабочее пространство и пользовательский интерфейс в Python и R. Управление вводом и выводом. Управление пакетами и библиотеками. Получение информации об объектах и функциях	4
2	1	Обработка данных и управляющие конструкции. Перекодировка переменных. Обработка пропущенных значений. Импорт данных. Объединение наборов данных. Извлечение данных	4
3	2	Проверка различий между выборками. Критерий Манна-Уитни, критерий Вилкоксона, критерий Стьюдента для независимых и зависимых выборок	4
4	2	Корреляционные связи. Коэффициенты корреляции Пирсона и Спирмена. Одномерный регрессионный анализ. Связи между категориальными	4

		переменными	
5	3	Графические функции низкого уровня. Добавление элементов к существующим диаграммам. Настройка параметров вывода диаграмм. Экспорт в файлы pdf	4
6	3	Коррелограммы. Пузырьковые диаграммы. Мозаичные диаграммы. Категоризованные графики (пакет lattice). Графическая система, основанная на графическом словаре (пакет ggplot2). Взаимодействие с диаграммами в интерактивном режиме.	4
7	4	Классификация в бинарных пространствах. Выделение ассоциативных правил. Анализ последовательностей событий. Моделирование порядковых переменных. Логит для порядковых переменных.	4
8	4	Кластерный анализ. Иерархическая кластеризация. Метод k-средних. Карты Кохонена. Многомерная ординация. Непараметрический дисперсионный анализ. Немеетрическое многомерное шкалирование.	4

### 5.3. Лабораторные работы

Не предусмотрены

### 5.4. Самостоятельная работа студента

Выполнение СРС		
Вид работы и содержание задания	Список литературы (с указанием разделов, глав, страниц)	Кол-во часов
Подготовка к экзамену по разделу "Общие сведения о языках Python и R"	ЭУМД, осн. лит. 2, с. 30-80	20
Подготовка к экзамену по разделу "Первичные статистики и базовые статистические методы"	ЭУМД, осн. лит. 2, с. 198-298	20
Подготовка к экзамену по разделу "Библиотеки Python и R для визуализации данных"	ЭУМД, осн. лит. 2, с. 81-112, 171-197	20
Подготовка к экзамену по разделу "Многомерные методы и Data Mining"	ЭУМД, доп. лит. 1, с. 88-320	20

## 6. Инновационные образовательные технологии, используемые в учебном процессе

Инновационные формы учебных занятий	Вид работы (Л, ПЗ, ЛР)	Краткое описание	Кол-во ауд. часов
Обсуждение высказываний экспертов	Лекции	Просмотр видео-роликов (фрагменты выступлений экспертов-практиков на конференциях и вебинарах). Вопросы по просмотренному видео	20

## Собственные инновационные способы и методы, используемые в образовательном процессе

Не предусмотрены

Использование результатов научных исследований, проводимых университетом, в рамках данной дисциплины: нет

## 7. Фонд оценочных средств (ФОС) для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины

### 7.1. Паспорт фонда оценочных средств

Наименование разделов дисциплины	Контролируемая компетенция ЗУНы	Вид контроля (включая текущий)	№№ заданий
Все разделы	ПК-17 способность использовать основные методы естественнонаучных дисциплин в профессиональной деятельности для теоретического и экспериментального исследования	Текущий	Т1-Т16
Все разделы	ПК-18 способность использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования	Экзамен	Э1-Э16

### 7.2. Виды контроля, процедуры проведения, критерии оценивания

Вид контроля	Процедуры проведения и оценивания	Критерии оценивания
Текущий	Текущий контроль проводится в процессе проведения практических занятий и семинаров в форме устных вопросов	Отлично: если студент демонстрирует понимание более 86% учебного материала Хорошо: если студент демонстрирует понимание более 73% учебного материала Удовлетворительно: если студент демонстрирует понимание более 60% учебного материала Неудовлетворительно: если студент демонстрирует понимание менее 60% учебного материала
Экзамен	Экзамен проводится в устной форме по билетам, каждый из которых содержит 2 вопроса. Преподаватель оценивает долю усвоенного студентом учебного материала (отношение объема изложенного студентом учебного материала к общему объему вопросов билета).	Отлично: если студент демонстрирует понимание более 86% учебного материала Хорошо: если студент демонстрирует понимание более 73% учебного материала Удовлетворительно: если студент демонстрирует понимание более 60% учебного материала Неудовлетворительно: если студент демонстрирует понимание менее 60% учебного материала

### 7.3. Типовые контрольные задания

Вид	Типовые контрольные задания
-----	-----------------------------

контроля	
Текущий	<p>T1. Рабочее пространство и интерфейс среды R. Управление вводом и выводом.  T2. Управление пакетами и библиотеками R. Получение информации об объектам и функциях  T3. Обработка данных и управляющие конструкции. Перекодировка переменных. Обработка пропущенных значений.  T4. Импорт данных. Объединение наборов данных. Извлечение данных  T5. Проверка различий между выборками. Критерий Манна-Уитни, критерий Вилкоксона, критерий Стьюдента для независимых и зависимых выборок  T6. Корреляционные связи. Коэффициенты корреляции Пирсона и Спирмена.  T7. Одномерный регрессионный анализ.  T8. Связи между категориальными переменными  T9. Графические функции низкого уровня. Добавление элементов к существующим диаграммам.  T10. Настройка параметров вывода диаграмм. Экспорт в файлы pdf  T11. Коррелограммы. Пузырьковые диаграммы. Мозаичные диаграммы. Категоризованные графики (пакет lattice).  T12. Графическая система, основанная на графическом словаре (пакет ggplot2). Взаимодействие с диаграммами в интерактивном режиме.  T13. Классификация в бинарных пространствах. Выделение ассоциативных правил. Анализ последовательностей событий.  T14. Моделирование порядковых переменных. Логит для порядковых переменных.  T15. Кластерный анализ. Иерархическая кластеризация. Метод k-средних.  T16. Карты Кохонена. Многомерная ординация.  T17. Непараметрический дисперсионный анализ. Неметрическое многомерное шкалирование.  РП (Анализ Д) - вопросы для текущего контроля.doc</p>
Экзамен	<p>Э1. Типы данных в среде R.  Э2. Векторы, матрицы и таблицы.  Э3. Факторы и списки.  Э4. Встроенные функции среды R. Математические, статистические и текстовые функции.  Э5. Определение пользовательских функций в среде R.  Э6. Меры центральной тенденции и меры изменчивости.  Э7. Распределение значений переменной.  Э8. Статистические критерии проверки гипотез. Проверка уровня значимости.  Э9. Построение доверительных интервалов.  Э10. Графические функции высокого уровня. Параметры диаграмм, осей, надписей, легенд и текстовых аннотаций  Э11. Столбчатые и круговые диаграммы. Гистограммы. Диаграммы размахов. Точечные диаграммы.  Э12. Многомерные регрессионные модели. Выбор оптимального набора предикторов линейной модели. Мультиколлинеарность. Нелинейные модели.  Э13. Деревья регрессии. Сравнение построенных моделей и оценка информативности предикторов.  Э14. Модели классификации. Дискриминантный анализ.  Э15. Деревья классификации. Логистическая регрессия. Метод k-ближайших соседей.  Э16. Байесовский классификатор. Нелинейные классификаторы. Модель мультиномиального логита.  Э17. Классификация на базе нейронных сетей.  РП (Анализ Д) - вопросы к экзамену.doc</p>

## 8. Учебно-методическое и информационное обеспечение дисциплины

## Печатная учебно-методическая документация

### а) основная литература:

Не предусмотрена

### б) дополнительная литература:

Не предусмотрена

### в) отечественные и зарубежные журналы по дисциплине, имеющиеся в библиотеке:

1. Машинное обучение и анализ данных
2. Моделирование и анализ данных
3. BPM World
4. Intelligent Data Analysis

### г) методические указания для студентов по освоению дисциплины:

1. Мокеев, В.В. Анализ данных: метод главных компонент / В.В. Мокеев. Издательский центр ЮУрГУ, 2015.–155с.
2. Никулин, Д.Н. Анализ данных в среде R: учебное пособие / Д.Н. Никулин; под ред. В.В. Мокеева. – Челябинск: Издательский центр ЮУрГУ, 2019. – 80 с.

*из них: учебно-методическое обеспечение самостоятельной работы студента:*

3. Мокеев, В.В. Анализ данных: метод главных компонент / В.В. Мокеев. Издательский центр ЮУрГУ, 2015.–155с.
4. Никулин, Д.Н. Анализ данных в среде R: учебное пособие / Д.Н. Никулин; под ред. В.В. Мокеева. – Челябинск: Издательский центр ЮУрГУ, 2019. – 80 с.

## Электронная учебно-методическая документация

№	Вид литературы	Наименование разработки	Наименование ресурса в электронной форме	Доступность (сеть Интернет / локальная сеть; авторизованный / свободный доступ)
1	Основная литература	С.Э. Мاستицкий, В.К. Шитиков. Статистический анализ и визуализация данных с помощью R. [Электронный ресурс]. Тольятти, 2014. - 401 с.	Электронно-библиотечная система издательства Лань	Интернет / Авторизованный
2	Основная литература	Роберт И. Кабаков. R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с.	Электронно-библиотечная система издательства Лань	Интернет / Авторизованный
3	Дополнительная литература	Шитиков В.К., Мاستицкий С.Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. [Электронный ресурс]. Тольятти, 2017. - 351 с.	Электронно-библиотечная система издательства Лань	Интернет / Авторизованный



4	Основная литература	Никулин, Д.Н. Анализ данных в среде R: учебное пособие / Д.Н. Никулин; под ред. В.В. Мокеева. – Челябинск: Издательский центр ЮУрГУ, 2019. – 80 с.	eLIBRARY.RU	Интернет / Свободный
5	Основная литература	Никулин, Д.Н. Анализ данных в среде R: учебное пособие / Д.Н. Никулин; под ред. В.В. Мокеева. – Челябинск: Издательский центр ЮУрГУ, 2019. – 80 с.	Электронный каталог ЮУрГУ	Интернет / Свободный

## 9. Информационные технологии, используемые при осуществлении образовательного процесса

Перечень используемого программного обеспечения:

1. Microsoft-Office(бессрочно)
2. -Python(бессрочно)

Перечень используемых информационных справочных систем:

Нет

## 10. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Лекции	143 (36)	Компьютер, проектор и динамики для просмотра видео-роликов
Практические занятия и семинары	115 (36)	Компьютерный класс с установленными программами Anaconda, R и Microsoft Office