

УТВЕРЖДАЮ:
Руководитель направления



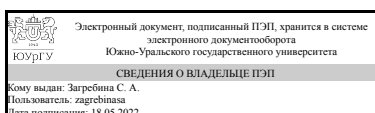
Л. Б. Соколинский

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.16.М1.02 Программирование для анализа данных
для направления 09.03.04 Программная инженерия
уровень Бакалавриат
форма обучения очная
кафедра-разработчик Математическое и компьютерное моделирование

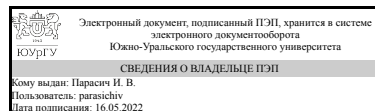
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 09.03.04 Программная инженерия, утверждённым приказом Минобрнауки от 19.09.2017 № 920

Зав.кафедрой разработчика,
д.физ.-мат.н., проф.



С. А. Загребина

Разработчик программы,
к.техн.н., доцент



И. В. Парасич

1. Цели и задачи дисциплины

Цель дисциплины: освоить методы и способы анализа данных с использованием языков программирования для решения различных научных и практических задач.
Задачи: познакомиться с различными видами анализа данных (описательная статистика, визуализация, корреляционный и регрессионный анализ, метод главных компонент, кластерный анализ) научиться на практике проводить анализ данных с использованием библиотек языков программирования Python и R;

Краткое содержание дисциплины

В данном курсе для анализа данных используются язык программирования Python и язык R - работа осуществляется в онлайн-среде Google Colab. Изучаются возможности использования библиотек и модулей Python для анализа данных (Numpy, Pandas, Scipy, Matplotlib, Scikit-learn, Statsmodels и др.). Рассматриваются функции и методы языка R для статистического анализа и визуализации данных. Большое внимание уделяется реализации оценки числовых характеристик, проверке статистических гипотез. Изучаются возможности проведения корреляционного и регрессионного анализа. Приводятся примеры использования метода главных компонент.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	Знает: инструментальные средства и информационные технологии анализа данных исходя из имеющихся ресурсов и ограничений
УК-6 Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	Умеет: адаптировать известные программные средства анализа данных в свою профессиональную область, с учётом возникающих ограничений по времени и ресурсам

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
1.Ф.16.М1.01 Анализ данных и технологии работы с данными	1.Ф.16.М1.03 Приложения и практика анализа данных

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

Дисциплина	Требования
1.Ф.16.М1.01 Анализ данных и технологии работы с данными	Знает: способы сбора, обработки и анализа данных для решения своих профессиональных

	задач с учётом имеющихся ресурсов и правовых норм Умеет: применять математические методы обработки данных для выбора и реализации оптимального способа решения профессиональных задач Имеет практический опыт: в использовании Excel Microsoft Office для обработки данных
--	--

4. Объём и виды учебной работы

Общая трудоёмкость дисциплины составляет 4 з.е., 144 ч., 72,25 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		4	
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)	0	0	
Лабораторные работы (ЛР)	32	32	
<i>Самостоятельная работа (СРС)</i>	71,75	71,75	
подготовка к проверочным заданиям и тестам, проводимым на лекциях	16	16	
подготовка к дифференцированному зачёту	23,75	23,75	
подготовка к лабораторным работам	32	32	
Консультации и промежуточная аттестация	8,25	8,25	
Вид контроля (зачет, диф.зачет, экзамен)	-	диф.зачет	

5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Знакомство с языком программирования Python.	4	2	0	2
2	Работа с библиотекой NumPy языка Python. Описательная статистика.	8	4	0	4
3	Работа с библиотекой Matplotlib. Построение графиков.	4	2	0	2
4	Корреляционный анализ и проверка статистических гипотез в Python (библиотека Scipy).	8	4	0	4
5	Работа с библиотекой Pandas языка Python. Анализ табличных данных.	18	8	0	10
6	Регрессионный анализ в Python (модуль Statsmodels и библиотека Scikit-learn).	12	6	0	6
7	Компонентный анализ в Python (метод PCA)	1	1	0	0
8	Основы кластерного анализа и задачи классификации.	1	1	0	0
9	Основы статистического анализа данных в R	8	4	0	4

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Язык программирования Python. Особенности и возможности языка. Краткий обзор основных библиотек для анализа данных. Обзор источников и площадок для работы, информации и изучения.	2
2	2	Библиотека NumPy (работа с массивами). Одномерные массивы NumPy : создание, основные операции, генерация случайных массивов, индексация, срезы, фильтрация, функции и методы. Двумерные массивы NumPy (матрицы) : создание, основные операции, генерация случайных массивов, индексация, срезы, фильтрация, функции и методы, матричные операции, элементы линейной алгебры. Описательная статистика в NumPy (функции и методы для вычисления статистических характеристик).	4
3	3	Библиотека Matplotlib. Графическое отображение данных (график, полигон, гистограмма, круговая диаграмма...).	2
4	4	Проверка статистических гипотез. Уровень значимости и р-значение. Парный корреляционный анализ. Диаграмма рассеяния. Линейная и ранговая корреляция. Проверка данных на нормальность (функции библиотеки Scipy) и выбор коэффициента корреляции. Проверка гипотезы о значимости коэффициента корреляции. Функции для вычисления коэффициентов корреляции и р- значений библиотеки Scipy.	4
5	5	Работа с библиотекой Pandas. Series (создание, атрибуты, индексация, операции, функции и методы). DataFrame (создание, атрибуты, индексация, срезы, фильтрация. преобразование таблицы, группировка, сортировка, графические возможности, статистические и др. функции, сводные таблицы...)	6
6	5	Монтирование Google Диска. Загрузка файлов различного формата в DataFrame (.xls, .csv, .txt...) с Google диска	2
7	6	Парная линейной регрессия и её реализация в модуле Statsmodels. Нахождение и анализ остатков. Коэффициент детерминации. Оценка качества уравнения. Проверка значимости уравнения и отдельных коэффициентов. Графическое представление линии регрессии. Построение парных диаграмм рассеяния для многомерной совокупности в библиотеке Seaborn. Построение уравнения множественной линейной регрессии в модуле Statsmodels	4
8	6	Построение уравнения линейной (парной и множественной) регрессии в библиотеке Scikit-learn. Разбиение выборки на обучающую и тестовую. Оценка качества уравнения регрессии. Прогноз с использованием регрессионного уравнения.	2
9	7	Снижение размерности данных. Метод главных компонент (PCA).	1
10	8	Понятие о кластерном анализе и проблемах классификации.	1
11	9	Язык программирования R. Особенности и возможности языка. Обзор источников и площадок для работы, информации и изучения. Программирование на R в Google Colab. Описательная статистика и визуализация данных в R.	2
12	9	Основы корреляционного и регрессионного анализа в R.	2

5.2. Практические занятия, семинары

Не предусмотрены

5.3. Лабораторные работы

№ занятия	№ раздела	Наименование или краткое содержание лабораторной работы	Кол-во часов
1	1	Работа с Google Диском. Основы работы в Google Colab. Арифметические операции в Python. (Лабораторные Работы. Часть 1. - 2 работы).	2
2	2	Одномерные массивы и матрицы библиотеки Numpy. (Лабораторные Работы. Часть 2. - 2 работы).	4
3	3	Визуализация данных в библиотеке Matplotlib - график, полигон, гистограмма, диаграмма рассеяния, круговая диаграмма. (Лабораторные Работы. Часть 3. - 1 работа).	2
4	4	Проверка данных на нормальность с помощью функций библиотеки Scipy (критерии Шапиро-Уилка, Хи-квадрат, Колмогорова-Смирнова). Проверка значимости коэффициентов корреляции. (Лабораторные Работы. Часть 4. - 3 работы).	4
5	5	Библиотека Pandas. Series. Dataframe. Основные возможности работы. Статистический анализ в Pandas. (Лабораторные Работы. Часть 5. - ЛР 5,1, 5.2).	6
6	5	Монтирование Google Диска. Загрузка датасета в DataFrame из текстового файла и таблицы Excel. (Лабораторные Работы. Часть 5 - ЛР 5.3).	4
7	6	Анализ связи между показателями. Построение диаграммы рассеяния. Выбор формы регрессионной кривой. Построение уравнения парной линейной регрессии в модуле Statsmodels. Коэффициент детерминации. Оценка качества уравнения. Проверка значимости уравнения и отдельных коэффициентов. Графическое представление линии регрессии. (Лабораторные Работы. Часть 6. - ЛР 6.1).	3
8	6	Построение уравнения множественной линейной регрессии в модуле Statsmodels. (Лабораторные Работы. Часть 6. - ЛР 6.2).	3
9	9	Язык программирования R. Запуск в Google Colab. Загрузка в программу файла excel с таблицей данных. Визуализация данных. Вычисление числовых характеристик (описательная статистика), корреляционный анализ и построение уравнений регрессии. (Лабораторные Работы. Часть 7. - ЛР 7.1).	4

5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
подготовка к проверочным заданиям и тестам, проводимым на лекциях	1. Многомерные статистические методы : методические указания / составители И. Л. Макарова, А. М. Игнатенко. — Сочи : СГУ, 2018. — 64 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/147679 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей. 2. Герасимова, Е. А. Эконометрика: регрессионный анализ : учебно-методическое пособие / Е. А. Герасимова, М. Ю. Карышев. — Самара : СамГУПС, 2011. — 98 с. — Текст : электронный // Лань : электронно-библиотечная система.	4	16

	— URL: https://e.lanbook.com/book/130432 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей. Тексты лекций в электронном виде. Набор блокнотов Google Colab		
подготовка к дифференцированному зачёту	1. Многомерные статистические методы : методические указания / составители И. Л. Макарова, А. М. Игнатенко. — Сочи : СГУ, 2018. — 64 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/147679 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей. 2. Герасимова, Е. А. Эконометрика: регрессионный анализ : учебно-методическое пособие / Е. А. Герасимова, М. Ю. Карышев. — Самара : СамГУПС, 2011. — 98 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/130432 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей. Тексты лекций в электронном виде. Методические указания к лабораторным работам. Набор блокнотов Google Colab.	4	23,75
подготовка к лабораторным работам	1. Многомерные статистические методы : методические указания / составители И. Л. Макарова, А. М. Игнатенко. — Сочи : СГУ, 2018. — 64 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/147679 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей. 2. Герасимова, Е. А. Эконометрика: регрессионный анализ : учебно-методическое пособие / Е. А. Герасимова, М. Ю. Карышев. — Самара : СамГУПС, 2011. — 98 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/130432 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей. Тексты лекций в электронном виде. Методические указания к лабораторным работам. Набор блокнотов Google Colab.	4	32

6. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления баллов	Учитывается в ПА
1	4	Текущий контроль	Задание 1 "Среды выполнения"	2	2	Задание содержит 2 пункта, за правильное выполнение каждого пункта начисляется 1 балл	дифференцированный зачет
2	4	Текущий контроль	Задание 2 "Numpy1"	4	15	Задание содержит 15 подпунктов, за правильное выполнение каждого начисляется 1 балл	дифференцированный зачет
3	4	Текущий контроль	Задание 3 "Описательная статистика"	4	6	Задание содержит 6 пунктов, за правильное выполнение каждого пункта начисляется 1 балл	дифференцированный зачет
4	4	Текущий контроль	Задание 4 (тест) "Numpy1+Numpy2"	6	10	Задание (тест) содержит 10 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
5	4	Текущий контроль	Задание 5 "Графика"	2	5	Задание содержит 5 пунктов, за правильное выполнение каждого пункта начисляется 1 балл	дифференцированный зачет
6	4	Текущий контроль	Задание 6 "Проверка данных на нормальность"	4	4	Задание содержит 2 пункта, за правильное выполнение каждого пункта начисляется 2 балла	дифференцированный зачет
7	4	Текущий контроль	Задание 7 (тест) "Корреляция"	2	9	Задание (тест) содержит 9 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
8	4	Текущий контроль	Задание 8 "Коэффициент корреляции Пирсона"	4	5	Задание содержит 5 подпунктов, за правильное выполнение каждого начисляется 1 балл	дифференцированный зачет
9	4	Текущий контроль	Задание 9 "Series"	2	3	Задание содержит 3 пункта, за правильное выполнение каждого пункта начисляется 1 балл	дифференцированный зачет
10	4	Текущий контроль	Задание 10 "DataFrame"	4	5	Задание содержит 5 пунктов, за правильное выполнение каждого пункта начисляется 1 балл	дифференцированный зачет
11	4	Текущий	Задание 11	6	5	Задание содержит 5	дифференцированный

		контроль	"Парная линейная регрессия"			пунктов, за правильное выполнение каждого пункта начисляется 1 балл	зачет
12	4	Текущий контроль	Лабораторные работы. Часть 1 "Основы Python" (2 работы)	6	2	За каждую правильно выполненную работу начисляется 1 балл	дифференцированный зачет
13	4	Текущий контроль	Лабораторные работы. Часть 2 "Numpy" (2 работы)	9	2	За каждую правильно выполненную работу начисляется 1 балл	дифференцированный зачет
14	4	Текущий контроль	Лабораторные работы. Часть 3 "Matplotlib" (1 работа)	6	1	За правильно выполненную работу начисляется 1 балл	дифференцированный зачет
15	4	Текущий контроль	Лабораторные работы. Часть 4 "Scipy" (3 работы)	9	3	За каждую правильно выполненную работу начисляется 1 балл	дифференцированный зачет
16	4	Текущий контроль	Лабораторные работы. Часть 5 "Pandas" (3 работ)	12	3	За каждую правильно выполненную работу начисляется 1 балл	дифференцированный зачет
17	4	Текущий контроль	Лабораторные работы. Часть 6 "Регрессия" (2 работы)	9	2	За каждую правильно выполненную работу начисляется 1 балл	дифференцированный зачет
18	4	Текущий контроль	Лабораторные работы. Часть 7 "Язык R" (1 работа)	6	2	ЛР содержит 6 заданий, за правильное выполнение любых 3 начисляется 1 балл	дифференцированный зачет
19	4	Промежуточная аттестация	Задание на зачёте	-	13	Задание содержит 7 пунктов, за правильное выполнение которых начисляется, соответственно, 2,1,2,2,1,2,3 балла - максимум 13 баллов	дифференцированный зачет

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
дифференцированный зачет	<p>При оценивании результатов учебной деятельности обучающегося по дисциплине используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (Положение о БРС утверждено приказом ректора от 24.05.2019 г. № 179, в редакции приказа ректора от 10.03.2022 г. No 25-13/09).</p> <p>Оценка за дисциплину формируется на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля. Отлично: Величина рейтинга обучающегося по дисциплине 85...100 %.</p> <p>Хорошо: Величина рейтинга обучающегося по дисциплине 75...84 %. Удовлетворительно: Величина</p>	В соответствии с пп. 2.5, 2.6 Положения

	рейтинга обучающегося по дисциплине 60...74 %. Неудовлетворительно: Величина рейтинга обучающегося по дисциплине 0...59 %. Если студент не согласен с оценкой, полученной по результатам текущего контроля, студент проходит мероприятие промежуточной аттестации, на котором ему предлагается выполнить Задание на зачёте (выполняется на ПК на языке Python, содержит 7 пунктов). На выполнение задания дается 60 минут. В этом случае оценка за дисциплину рассчитывается на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и промежуточной аттестации. Фиксация результатов учебной деятельности по дисциплине проводится в день экзамена при личном присутствии студента.	
--	---	--

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ																		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
УК-2	Знает: инструментальные средства и информационные технологии анализа данных исходя из имеющихся ресурсов и ограничений	+		++				++++		+			+	+			+	+		+
УК-6	Умеет: адаптировать известные программные средства анализа данных в свою профессиональную область, с учётом возникающих ограничений по времени и ресурсам		+			++				+			+	+			+			+

Типовые контрольные задания по каждому мероприятию находятся в приложениях.

7. Учебно-методическое и информационное обеспечение дисциплины

Печатная учебно-методическая документация

а) основная литература:

1. Кремер, Н. Ш. Эконометрика [Текст] учебник для вузов по специальностям экономики и упр. Н. Ш. Кремер, Б. А. Путко ; под ред. Н. Ш. Кремера. - 3-е изд., перераб. и доп. - М.: ЮНИТИ-ДАНА, 2013. - 328 с.
2. Гмурман, В. Е. Теория вероятностей и математическая статистика [Текст] учеб. пособие для вузов В. Е. Гмурман. - 12-е изд. - М.: Юрайт, 2014. - 478, [1] с. ил.

б) дополнительная литература:

1. Математическая статистика Учеб. для вузов В. Б. Горяинов, И. В. Павлов, Г. М. Цветкова, О. И. Тескин; Под ред.: В. С. Зарубина, А. П. Крищенко. - М.: Издательство МГТУ им. Н. Э. Баумана, 2001. - 423 с. ил.

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

Не предусмотрены

г) методические указания для студентов по освоению дисциплины:

1. Методические указания к лабораторным работам (Части 1-7)

2. Многомерный статистический анализ
3. Тексты лекций
4. Датасеты для анализа
5. Ссылки на блокноты Google Colab

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Многомерный статистический анализ

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Дополнительная литература	Электронно-библиотечная система издательства Лань	Чураков, Е. П. Введение в многомерные статистические методы : учебное пособие / Е. П. Чураков. — Санкт-Петербург : Лань, 2016. — 148 с. — ISBN 978-5-8114-2149-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/87598 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей.
2	Основная литература	Электронно-библиотечная система издательства Лань	Программные системы статистического анализа. Обнаружение закономерностей в данных с использованием системы R и языка Python : учебное пособие / В. М. Волкова, М. А. Семёнова, Е. С. Четвертакова, С. С. Вожов. — Новосибирск : НГТУ, 2017. — 74 с. — ISBN 978-5-7782-3183-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/118287 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей.
3	Основная литература	Электронно-библиотечная система издательства Лань	Копырин, А. С. Программирование на Python : учебное пособие / А. С. Копырин, Т. Л. Салова. — Сочи : СГУ, 2018. — 48 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/147665 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей.
4	Дополнительная литература	Электронно-библиотечная система издательства Лань	Герасимова, Е. А. Эконометрика: регрессионный анализ : учебно-методическое пособие / Е. А. Герасимова, М. Ю. Карышев. — Самара : СамГУПС, 2011. — 98 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/130432 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей.
5	Основная литература	Электронно-библиотечная система издательства Лань	Круценюк, К. Ю. Корреляционно-регрессионный анализ в эконометрических моделях : учебное пособие / К. Ю. Круценюк. — Норильск : НГИИ, 2018. — 108 с. — ISBN 978-5-89009-698-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/155915 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей.
6	Дополнительная литература	Электронно-библиотечная система издательства Лань	Шкаберина, Г. Ш. Программирование. Основы языка Python : учебное пособие / Г. Ш. Шкаберина, Н. Л. Резова. — Красноярск : СибГУ им. академика М. Ф. Решетнёва, 2018. — 92 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL:

			https://e.lanbook.com/book/147450 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей.
7	Основная литература	Электронно-библиотечная система издательства Лань	Многомерные статистические методы : методические указания / составители И. Л. Макарова, А. М. Игнатенко. — Сочи : СГУ, 2018. — 64 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/147679 (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей.

Перечень используемого программного обеспечения:

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

Перечень используемых профессиональных баз данных и информационных справочных систем:

Нет

8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Практические занятия и семинары	405 (1)	Компьютерный класс с выходом в интернет
Лекции	405 (1)	ПК с проектором для преподавателя+ ПК для студентов