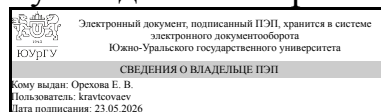


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Руководитель направления



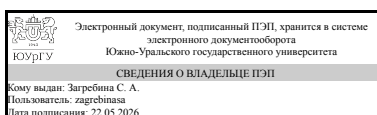
Е. В. Орехова

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.14.М2.02 Программирование для анализа данных
для направления 45.03.03 Фундаментальная и прикладная лингвистика
уровень Бакалавриат
форма обучения очная
кафедра-разработчик Математическое и компьютерное моделирование

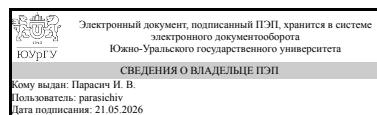
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 45.03.03 Фундаментальная и прикладная лингвистика, утверждённым приказом Минобрнауки от 24.04.2018 № 323

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 языка программирования Python для решения различных научных и практических задач. Задачи: -познакомиться с различными видами анализа данных (описательная статистика, визуализация, корреляционный и регрессионный анализ, проверка статистических гипотез) -научиться на практике проводить комплексное исследование датасетов с использованием библиотек языка программирования Python.

Краткое содержание дисциплины

В данном курсе для анализа данных используются язык программирования Python - работа осуществляется в онлайн-среде Google Colab. Изучаются возможности использования библиотек и модулей Python для статистического анализа и визуализации данных (Numpy, Pandas, Scipy, Matplotlib, Scikit-learn, Statsmodels и др.). Большое внимание уделяется оценке числовых характеристик показателей, изучению взаимосвязей данных, измеренных в разных шкалах, построению регрессионных моделей, проверке их качества и адекватности, рассматриваются вопросы прогнозирования.

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.14.М2.01 Анализ данных и технологии работы с данными	1.Ф.14.М2.03 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.14.М2.01 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 Имеет практический опыт:
--	---

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 4 з.е., 144 ч., 72,5 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>	71,5	71,5
подготовка к лабораторным работам	32	32
подготовка к дифференцированному зачёту	23,5	23,5
подготовка к проверочным заданиям и тестам, проводимым на лекциях	16	16
Консультации и промежуточная аттестация	8,5	8,5
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
0	Элементы математической статистики	2	0	0	2
1	Библиотека Numpy (одномерные массивы и матрицы).	16	8	0	8
2	Библиотека Pandas. Обработка и анализ табличных данных. Визуализация.	16	8	0	8
3	Корреляционный анализ и проверка статистических гипотез в Python (библиотека Scipy)	16	8	0	8
4	Регрессионный анализ в Python	14	8	0	6

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

2,3	1	Библиотека Numpy (работа с массивами). Одномерные массивы Numpy : создание, основные операции, генерация случайных массивов, индексация, срезы, фильтрация, функции и методы. Описательная статистика в Numpy (функции и методы для вычисления статистических характеристик).	4
4	1	Двумерные массивы Numpy. Матричные операции в linalg	2
5,6,7	2	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
8	2	Монтирование Google Диска. Загрузка файлов различного формата в DataFrame (.xls, .csv, .txt....) с Google диска	2
9,10	3	Проверка статистических гипотез. Уровень значимости и р-значение. Проверка на нормальность (функции библиотеки Scipy). Понятие корреляции. Парный корреляционный анализ. Диаграмма рассеяния. Линейная и ранговая корреляция. Проверка гипотезы о значимости коэффициента корреляции. Функции для вычисления коэффициентов корреляции и р- значений библиотеки Scipy.	4
11	3	Связь дихотомических, категориальных данных. Коэффициенты корреляции точечно-бисериальный, рангово-бисериальный, "фи", Крамера. Выбор коэффициента корреляции.	2
12	3	Многомерная корреляция. Парные, частные, множественные коэффициенты корреляции, Проверка значимости.	2
13,14	4	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
15	4	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
16	4	Регрессионные модели в Scikit-Learn. Полиномиальная регрессия. Регрессия Лассо и Риджа. Использование фиктивных переменных в регрессионном анализе. Проверка адекватности модели (условия Гаусса-Маркова). Мультиколлинеарность и способы борьбы с ней. Нелинейные регрессионные модели	2

5.2. Практические занятия, семинары

Не предусмотрены

5.3. Лабораторные работы

№ занятия	№ раздела	Наименование или краткое содержание лабораторной работы	Кол-во часов
1	0	Основы математической статистики. Гистограммы и описательная статистика. Пакет анализа (Excel) (Лабораторная работа 0)	2
2,3,4	1	Основы работы в Google Colab (Python). Одномерные массивы и матрицы библиотеки Numpy. (Лабораторные Работы. Часть 1 - 3 работы).	6
5	1	Двумерные массивы Numpy. Функции linalg (Лабораторные Работы. Часть 1	2

		- 1 работа).	
6,7	2	Библиотека Pandas. Series. Dataframe. Основные возможности работы. Статистический анализ и визуализация в Pandas. (Лабораторные Работы. Часть 2 - 2 работы).	4
8,9	2	Монтирование Google Диска. Загрузка датасета в DataFrame из текстового файла и таблицы Excel. Визуализация в Python (Лабораторные Работы. Часть 2 - 2 работы).	4
10,11	3	Проверка данных на нормальность с помощью функций библиотеки Scipy (критерии Шапиро-Уилка, Хи-квадрат, Колмогорова-Смирнова). Коэффициенты корреляции Спирмена, Кендалла, Пирсона. Выбор, вычисление и проверка значимости коэффициентов корреляции. Построение диаграмм рассеяния. (Лабораторные работы. Часть 3 - 3 работы).	4
12,13	3	Коэффициенты корреляции бисериальные, ФИ, Крамера. (Лабораторные работы. Часть 3 - 2 работы).	4
14	4	Корреляционно-регрессионный анализ в ПА (Лабораторные Работы. Часть 4 - 1 работа).	2
15	4	Построение уравнения парной и множественной линейной регрессии в Statsmodels (матричные формулы) и модуле Statsmodels. Коэффициент детерминации. Оценка качества модели, значимости всего уравнения и отдельных коэффициентов (Лабораторные Работы. Часть 4 - 1 работа).	2
16	4	Регрессионные анализ в библиотеке Scikit-Learn. Обучающая и тестовая выборка. Линейная и полиномиальная модели. Оценка качества и выбор оптимальной модели. Прогнозирование. (Лабораторные Работы. Часть 4 - 1 работа).	2

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 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
подготовка к дифференцированному зачёту	1. Многомерные статистические методы : методические указания / составители И.	4	23,5

	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.		
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. — 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	4	16

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
0	4	Текущий контроль	Лабораторные работы. Часть 0 "МС" (1 работа)	1	3	ЛР содержит 3 пункта, за правильное выполнение каждого пункта	дифференцированный зачет

						начисляется 1 балл	
1	4	Текущий контроль	Лабораторные работы. Часть 1 "Numpy" (4 работы)	2	4	За правильно выполненную работу начисляется 1 балл	дифференцированный зачет
2	4	Текущий контроль	Проверочное задание 1 "Python"	1	4	Задание содержит 4 пункта, за правильное выполнение каждого начисляется 1 балл	дифференцированный зачет
3	4	Текущий контроль	Проверочное задание 2 "Numpy 1"	1	5	Задание (тест) содержит 5 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
4	4	Текущий контроль	Проверочное задание 3 "Numpy2"	1	4	Задание (тест) содержит 4 вопроса, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
5	4	Текущий контроль	Лабораторные работы. Часть 2 "Pandas и графика" (4 работы)	3	4	За каждую правильно выполненную работу начисляется 1 балл	дифференцированный зачет
7	4	Текущий контроль	Лабораторные работы. Часть 3 "Гипотезы. Корреляция" (5 работ)	2	5	За каждую правильно выполненную работу начисляется 1 балл	дифференцированный зачет
8	4	Текущий контроль	Проверочное задание 5 "Измерительные шкалы"	1	8	Задание (тест) содержит 8 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
9	4	Текущий контроль	Проверочное задание 6 "Понятие корреляции"	1	5	Задание (тест) содержит 5 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
10	4	Текущий контроль	Проверочное задание 7 "Многомерная корреляция"	1	5	Задание (тест) содержит 5 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
11	4	Текущий контроль	Лабораторные работы. Часть 4 "Регрессия" (3 работы)	1	3	За правильно выполненную работу начисляется 1 балл	дифференцированный зачет
12	4	Текущий контроль	Проверочное задание 8 "Множественная регрессия"	1	9	Задание (тест) содержит 9 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
13	4	Промежуточная аттестация	ФОС "Задание на зачёте"	-	20	Задание (тест) содержит 20 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
дифференцированный зачет	На дифференцированном зачете происходит оценивание знаний, умений и приобретенного опыта обучающихся по дисциплине " Программирование для анализа данных" на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля. При недостаточной и/или не устраивающей студента величине рейтинга ему может быть предложено выполнить задание на зачете. В результате складывается совокупный рейтинг студента, который дифференцируется в оценку и проставляется в ведомость, зачетную книжку студента. Отлично: Величина рейтинга обучающегося по дисциплине 85...100 % Хорошо: Величина рейтинга обучающегося дисциплине 75...84 % Удовлетворительно: Величина рейтинга обучающегося по дисциплине 60...74 % Неудовлетворительно: Величина рейтинга обучающегося по дисциплине 0...59 %. Оценка по дисциплине вносится в «Приложение к диплому бакалавра».	В соответствии с пп. 2.5, 2.6 Положения

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ												
		0	1	2	3	4	5	7	8	9	10	11	12	13
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.

б) дополнительная литература:

1. 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.
2. Математическая статистика Учеб. для втузов 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.
3. Дубров А. М. Многомерные статистические методы: Для экономистов и менеджеров : Учеб. для экон. специальностей вузов / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. - М. : Финансы и статистика, 2003. - 349, [1] с.

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

Не предусмотрены

г) методические указания для студентов по освоению дисциплины:

1. Ссылки на блокноты Google Colab
2. Датасеты для анализа
3. Тексты лекций
4. Методические указания к лабораторным работам (Части 0-7)

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Датасеты для анализа

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Основная литература	ЭБС издательства Лань	Титов, А. Н. Обработка данных в Python. Основы работы с библиотекой Pandas : учебно-методическое пособие / А. Н. Титов, Р. Ф. Тагиева. — Казань : КНИТУ, 2022. — 116 с. — ISBN 978-5-7882-3164-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/331013 (дата обращения: 21.05.2026). — Режим доступа: для авториз. пользователей.
2	Основная литература	ЭБС издательства Лань	Маккинни, У. Python и анализ данных. Первичная обработка данных с применением pandas, NumPy и Jupiter : справочник / У. Маккинни ; перевод с английского А. А. Слинкина. — 3-е изд. — Москва : ДМК Пресс, 2023. — 536 с. — ISBN 978-5-93700-174-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/348086 (дата обращения: 21.05.2026). — Режим доступа: для авториз. пользователей.
3	Основная литература	ЭБС издательства Лань	Круценюк, К. Ю. Корреляционно-регрессионный анализ в эконометрических моделях : учебное пособие / К. Ю. Круценюк. — Норильск : ЗГУ им. Н.М. Федоровского, 2018. — 108 с. — ISBN 978-5-89009-698-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/155915 (дата обращения: 21.05.2026). — Режим доступа: для авториз. пользователей.

Перечень используемого программного обеспечения:

1. Microsoft-Windows(бессрочно)
2. Microsoft-Office(бессрочно)
3. ФГАОУ ВО "ЮУрГУ (НИУ)"-Портал "Электронный ЮУрГУ" (<https://edu.susu.ru>)(бессрочно)
4. -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
Лекции	405 (1)	ПК с проектором для преподавателя+ ПК для студентов
Практические занятия и семинары	405 (1)	Компьютерный класс с выходом в интернет
Контроль самостоятельной работы	405 (1)	ПК с проектором для преподавателя+ ПК для студентов
Дифференцированный зачет	405 (1)	ПК с проектором для преподавателя+ ПК для студентов