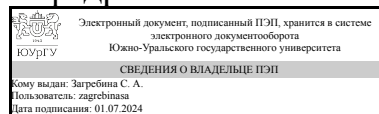


УТВЕРЖДАЮ:
Заведующий выпускающей
кафедрой



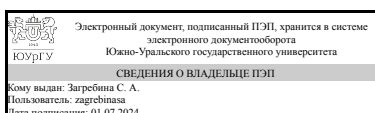
С. А. Загребина

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.М0.08 Многомерный анализ данных
для направления 01.04.05 Статистика
уровень Магистратура
магистерская программа Статистическое и компьютерное моделирование
форма обучения очно-заочная
кафедра-разработчик Математическое и компьютерное моделирование

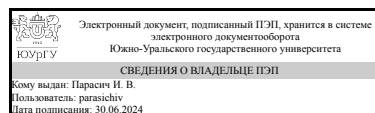
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 01.04.05 Статистика, утверждённым приказом Минобрнауки от 14.08.2020 № 1030

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, парный и множественный корреляционный 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, Linalg, Pandas, Scipy, Matplotlib, Statsmodels и др.). Большое внимание уделяется оценке числовых характеристик показателей, изучению взаимосвязей данных, измеренных в разных шкалах, проверке значимости корреляций, построению различных диаграмм и графиков

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-1 Способен активно участвовать в проведении экспериментальных статистических расчетов по оригинальным методикам и критически оценивать их результаты	Знает: методы проведения экспериментальных статистических расчетов по оригинальным методикам и критически оценивать их результаты Имеет практический опыт: проведения анализа и теоретического обобщения научных данных в соответствии с задачами исследования

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Нет	Статистические методы и модели прогнозирования, Пакеты прикладных статистических программ, Производственная практика (преддипломная) (5 семестр)

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

Нет

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 4 з.е., 144 ч., 57,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		2	
Общая трудоёмкость дисциплины	144	144	
<i>Аудиторные занятия:</i>	48	48	
Лекции (Л)	16	16	
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	0	0	
Лабораторные работы (ЛР)	32	32	
<i>Самостоятельная работа (СРС)</i>	86,5	86,5	
подготовка к проверочным заданиям и тестам, проводимым на лекциях	8	8	
Выполнение расчётов и оформление пояснительной записки в рамках Курсовой работы	45	45	
подготовка к дифференцированному зачёту	23,5	23,5	
подготовка к лабораторным работам	10	10	
Консультации и промежуточная аттестация	9,5	9,5	
Вид контроля (зачет, диф.зачет, экзамен)	-	диф.зачет, КР	

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). Визуализация в Matplotlib	12	4	0	8
2	Библиотека Pandas. Обработка и анализ табличных данных.	16	6	0	10
3	Парный корреляционный анализ и проверка статистических гипотез в Python (библиотека Scipy)	12	4	0	8
4	Многомерный корреляционный анализ	6	2	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. Библиотека Numpy (работа с массивами). Одномерные массивы Numpy : создание, основные операции, генерация случайных массивов, индексация, срезы, фильтрация, функции и методы.	2
2	1	Двумерные массивы Numpy (матрицы) : создание, основные операции, генерация случайных массивов, индексация, срезы, фильтрация, функции и методы, матричные операции, элементы линейной алгебры. Описательная	2

		статистика в NumPy (функции и методы для вычисления статистических характеристик). Библиотека Matplotlib. Графическое отображение данных (график, полигон, гистограмма, круговая диаграмма...).	
3,4	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...)	4
5	2	Монтирование Google Диска. Загрузка файлов различного формата в DataFrame (.xls, .csv, .txt...) с Google диска	2
6	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.	2
7	3	Связь дихотомических, категориальных данных. Коэффициенты корреляции точечно-бисериальный, рангово-бисериальный, "фи", Крамера. Выбор коэффициента корреляции.	2
8	4	Многомерная корреляция. Парные, частные, множественные коэффициенты корреляции, Проверка значимости.	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. Визуализация данных в библиотеке Matplotlib - график, полигон, гистограмма, диаграмма рассеяния, круговая диаграмма. (Лабораторные Работы. Часть 1 - ЛР 1.1, 1.2, 1.3 - 3 работы).	6
5	1	Визуализация данных в библиотеке Matplotlib - график, полигон, гистограмма, диаграмма рассеяния, круговая диаграмма. (Лабораторные Работы. Часть 1 - ЛР 1.4 - 1 работа).	2
6	2	Библиотека Pandas. Series. (Лабораторные Работы. Часть 2 - ЛР 2.1).	2
7,8,9	2	Библиотека Pandas. Dataframe. Основные возможности работы. Статистический анализ и визуализация в Pandas. (Лабораторные Работы. Часть 2 - ЛР 2.2).	6
10	2	Монтирование Google Диска. Загрузка датасета в DataFrame из текстового файла и таблицы Excel. (Лабораторные Работы. Часть 2 - ЛР 2.3).	2
11	3	Проверка данных на нормальность с помощью функций библиотеки Scipy (критерии Шапиро-Уилка, Хи-квадрат, Колмогорова-Смирнова). (Лабораторные работы. Часть 3 - ЛР 3.1).	2
12,13,14	3	Парная корреляция. Коэффициенты корреляции Спирмена, Кендалла, Пирсона, бисериальные. ФИ, Крамера. Выбор, вычисление и проверка значимости коэффициентов корреляции. Построение диаграмм рассеяния. (Лабораторные работы. Часть 3 - ЛР 3.2, ЛР 3.3, ЛР 3.4).	6

15, 16	4	Вычисление множественных коэффициентов корреляции и проверка их значимости (Лабораторные Работы. Часть 4 - ЛР 4.1).	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. — 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	2	8
Выполнение расчётов и оформление пояснительной записки в рамках Курсовой работы	1. 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. 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	45
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	2	23,5

	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 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.	2	10

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	2	Текущий контроль	Проверочное задание 0 "Основы МС"	1	12	Задание содержит 12 пунктов, за правильное выполнение каждого пункта начисляется 1 балл	дифференцированный зачет
2	2	Текущий контроль	Проверочное задание 1 "Python"	1	4	Задание содержит 4 пункта, за правильное выполнение каждого	дифференцированный зачет

						начисляется 1 балл	
3	2	Текущий контроль	Проверочное задание 2 "Numpy 1"	1	5	Задание (тест) содержит 5 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
4	2	Текущий контроль	Проверочное задание 3 "Series"	1	5	Задание содержит 5 пунктов, за правильное выполнение каждого пункта начисляется 1 балл	дифференцированный зачет
5	2	Текущий контроль	Проверочное задание 4 "DataFrame"	1	6	Задание содержит 6 пунктов, за правильное выполнение каждого пункта начисляется 1 балл	дифференцированный зачет
6	2	Текущий контроль	Проверочное задание 5 "Понятие корреляции"	1	5	Задание (тест) содержит 5 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
7	2	Текущий контроль	Проверочное задание 6 "Измерительные шкалы"	1	8	Задание (тест) содержит 8 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
8	2	Текущий контроль	Проверочное задание 7 "Коэффициент корреляции Спирмена"	1	5	Задание содержит 5 пунктов, за правильное выполнение каждого начисляется 1 балл	дифференцированный зачет
9	2	Текущий контроль	Проверочное задание 8 "Коэффициент корреляции ФИ"	1	5	Задание содержит 5 пунктов, за правильное выполнение каждого пункта начисляется 1 балл	дифференцированный зачет
10	2	Текущий контроль	Проверочное задание 9 "Многомерная корреляция"	1	5	Задание (тест) содержит 5 вопросов, за каждый правильный ответ начисляется 1 балл	дифференцированный зачет
13	2	Текущий контроль	Лабораторные работы. Часть 0 "МС" (1 работа)	1	3	ЛР содержит 3 пункта, за правильное выполнение каждого пункта начисляется 1 балл	дифференцированный зачет
14	2	Текущий контроль	Лабораторные работы. Часть 1 "Numpy" (4 работы)	2	4	За правильно выполненную работу начисляется 1 балл	дифференцированный зачет
15	2	Текущий контроль	Лабораторные работы. Часть 2 "Pandas" (3 работы)	2	3	За каждую правильно выполненную работу начисляется 1 балл	дифференцированный зачет
16	2	Текущий контроль	Лабораторные работы. Часть 3 "Парная корреляция" (4 работы)	2	4	За каждую правильно выполненную работу начисляется 1 балл	дифференцированный зачет
17	2	Текущий контроль	Лабораторные работы. Часть 4 "Множественная	1	1	За каждую правильно выполненную работу начисляется 1 балл	дифференцированный зачет

			корреляция" (1 работа)				
21	2	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	<p>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».</p>	В соответствии с пп. 2.5, 2.6 Положения
курсовые работы	<p>Магистры выполняют статистический анализ выбранного датасета с использованием библиотек языка программирования Python. Результаты расчётов обсуждаются с преподавателем, после чего оформляется пояснительная записка. В EDU в задание Курсовые работы прикрепляется : 1. текст записки (в формате pdf или word), содержащий в том числе ссылку на блокнот Colab 2. файл с расчётами на python (pdf) 3. файл с датасетом (xlsx, csv, ...). Защита проводится с демонстрацией результатов расчётов в присутствии комиссии и магистров. Результат оценивается по четырём бальной системе: Оценка "отлично" выставляется за полностью и качественно выполненное задание, содержащее демонстрацию всех возможностей и python при проведении статистических расчётов по теме КР, грамотно и логично оформленную пояснительную записку, уверенные ответы на вопросы при защите. Оценка "хорошо" выставляется за полностью выполненное задание, но содержащее погрешности в расчётах и/или не все возможности при расчётах, также частично допускаются неполные ответы на вопросы комиссии при защите. Оценка "удовлетворительно" ставится за выполненное задание, содержащее погрешности в расчётах и выводах, при частичных нарушениях логики оформления пояснительной записки, а также не совсем уверенном ответе на вопросы при защите.</p>	В соответствии с п. 2.7 Положения

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ																			
		1	2	3	4	5	6	7	8	9	10	13	14	15	16	17	21				
ПК-1	Знает: методы проведения экспериментальных статистических расчетов по оригинальным методикам и критически оценивать их результаты	+	+	+	+	+	+	+	+	+	+			+	+	+	+				
ПК-1	Имеет практический опыт: проведения анализа и теоретического обобщения научных данных в соответствии с задачами исследования	+	+	+	+				+	+		+	+	+	+	+	+				

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. Заляпин, В. И. Математическая статистика [Текст] учеб. пособие В. И. Заляпин, Е. В. Харитоновна ; Юж.-Урал. гос. ун-т, Каф. Мат. анализ ; ЮУрГУ. - Челябинск: Издательство ЮУрГУ, 2008. - 146 с.
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.

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 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 0-4)
2. Тексты лекций (Темы 1-4)
3. Датасеты для анализа
4. Ссылки на блокноты Google Colab (Темы 1-4)
5. Методические указания к КР

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Датасеты для анализа
2. Методические указания к КР

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	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.
5	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.
6	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. 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
Практические занятия и семинары	405 (1)	Компьютерный класс с выходом в интернет
Лекции	405 (1)	ПК с проектором для преподавателя+ ПК для студентов
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