

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Луганский государственный университет имени Владимира Даля»

Экономический факультет
Кафедра экономической кибернетики и прикладной статистики

УТВЕРЖДАЮ:
Декан экономического факультета
Тхор Е.С.
« 24 » _____ 2023 года



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

**«ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ОБРАБОТКИ
СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ»**

По направлению подготовки 38.04.05 Бизнес-информатика
Программа магистратуры: «Экономическая аналитика и бизнес-статистика»

Луганск - 2023


Лист согласования РПУД

Рабочая программа учебной дисциплины «Информационные технологии обработки статистической информации» по направлению подготовки 38.04.05 Бизнес-информатика. – 24 с.

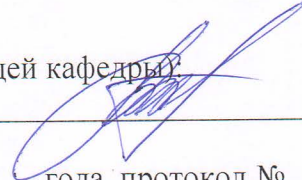
Рабочая программа учебной дисциплины «Информационные технологии обработки статистической информации» составлена с учетом Федерального государственного образовательного стандарта высшего образования по направлению подготовки 38.04.05 Бизнес-информатика Министерства науки и высшего образования Российской Федерации от 12 августа 2020 г. № 990.

СОСТАВИТЕЛЬ (СОСТАВИТЕЛИ):
к.э.н., доцент Спорняк С.А.

Рабочая программа дисциплины утверждена на заседании кафедры экономической кибернетики и прикладной статистики «18» 04 2023 г., протокол № 26

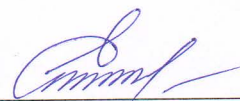
Заведующий кафедрой экономической кибернетики
и прикладной статистики  А.В. Велигура

Переутверждена: « » 20 г., протокол №

Согласована (для обеспечивающей кафедры):
Декан экономического факультета  Тхор Е.С.

Переутверждена: « » 20 года, протокол №

Рекомендована на заседании учебно-методической комиссии экономического факультета «21» апреле 2023 г., протокол № 4.

Председатель учебно-методической
комиссии экономического факультета  Е.Н. Шаповалова

Структура и содержание дисциплины

1. Цели и задачи дисциплины, ее место в учебном процессе

Целью дисциплины «Информационные технологии обработки статистической информации» является приобретение студентами необходимой квалификации для проведения статистического анализа различных экономических процессов и явлений с использованием программных средств статистического анализа.

Задачи:

изучение средств описательной статистики; изучение средств одномерного и двумерного дисперсионного анализа;

изучение методов снижения размерности, метода главных компонент; изучение средств множественного регрессионного статистического анализа и оценки результатов; изучение методов кластерного анализа; изучение способов глубокой обработки данных, “DataMining”;

изучение методов анализа рядов динамики и прогнозирования, применение авторегрессионных моделей скользящих средних для одномерного и множественного моделирования.

2. Место дисциплины в структуре ОПОП ВО

Дисциплина «Информационные технологии обработки статистической информации» относится к дисциплинам обязательной части.

Необходимыми условиями для освоения дисциплины являются: знания теории статистики, статистической методологии и методов исследования, прикладной статистики, бизнес-информатики и обработки статистических данных с помощью языка программирования R.

Содержание дисциплины является логическим продолжением содержания дисциплин «Бизнес-статистика», «Многомерный анализ и прогнозирование средствами языка программирования R», «Цифровые платформы и экосистемы современного бизнеса», «Информационные технологии в управлении проектами», и служит основой для выполнения научно-исследовательской работы, прохождения производственной и преддипломной практик, написания выпускной квалификационной работы.

3. Требования к результатам освоения содержания дисциплины

Код и наименование компетенции	Индикаторы достижений компетенции (по реализуемой дисциплине)	Перечень планируемых результатов
ОПК-3. Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности	ОПК-3.3. Выявляет возможности, создаваемые информационными и цифровыми технологиями	Знать: информационные и цифровые технологии для проведения статистического анализа;
		Уметь: использовать информационные и цифровые технологии для проведения статистического анализа;

использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта		Владеть: информационными и цифровыми технологиями для проведения статистического анализа.
ПК-4. Способен анализировать статистические данные в соответствии с базовой методологией и использованием программных средств статистического анализа	ПК-4.4. Способен проводить статистический анализ различных экономических процессов и явлений с использованием программных средств статистического анализа	Знать: методику проведения статистического анализа различных экономических явлений и процессов с использованием программных средств;
		Уметь: проводить статистический анализ с использованием программных средств;
		Владеть: программными средствами обработки статистических данных.

4. Структура и содержание дисциплины

4.1. Объем учебной дисциплины и виды учебной работы

Вид учебной работы	Очная форма	Заочная форма
Общая учебная нагрузка (всего)	144 (4 зач. ед)	144 (4 зач. ед)
Обязательная контактная работа (всего) в том числе:	56	28
Лекции	28	14
Семинарские занятия	-	-
Практические занятия	28	14
Лабораторные работы	-	-
Курсовая работа (курсовой проект)		
Другие формы и методы организации образовательного процесса (<i>расчетно-графические работы, индивидуальные задания и т.п.</i>)	-	-
Самостоятельная работа студента (всего)	88	116
Форма аттестации	зачет	зачет

4.2. Содержание разделов дисциплины

Тема 1. ВВЕДЕНИЕ В ДИСЦИПЛИНУ. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ

Цель и задачи дисциплины. Современные программные пакеты средств статистического анализа.

Тема 2. APACHE OPENOFFICE CALC: ОПИСАТЕЛЬНАЯ СТАТИСТИКА

Средства описательной статистики в OpenOffice Calc: возможности и ограничения. Анализ вариационного ряда. Результаты инструмента описательной статистики: дисперсия, мода, медиана, стандартная ошибка, уровень надежности, стандартное отклонение, асимметрия, эксцесс. Интерпретация полученных результатов.

Тема 3. APACHE OPENOFFICE CALC: ДИСПЕРСИОННЫЙ АНАЛИЗ, ОДНОМЕРНЫЙ И ДВУМЕРНЫЙ

Средства дисперсионного анализа в OpenOffice Calc. Проведение предварительного этапа, анализ исходной информации, оценка возможности проведения дисперсионного анализа: числовая и графическая, построение блочной диаграммы. Оценка результатов проверки гипотез. Двумерный дисперсионный анализ.

Тема 4. EViews: АВТОРЕГРЕССИОННОЕ МОДЕЛИРОВАНИЕ. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЕ

Авторегрессионное моделирование методом скользящих средних (ARMA, ARIMA) в Eviews. Проведение авторегрессионного моделирования динамики одномерных данных. Графический анализ исходных данных. Выбор параметров временного лага и сезонных колебаний. Проблема формальных критериев определения параметров модели. Многофакторное моделирование результативной переменной на основе авторегрессионных моделей.

Тема 5. ВВЕДЕНИЕ В ПРОГРАММИРОВАНИЕ НА ЯЗЫКЕ PYTHON

Введение в Python. Синтаксис языка. Базовые типы данных: числа, строки, списки, кортежи, словари, множества. Функции. Итераторы и генераторы. Классы и объекты. Декораторы. Ввод-вывод. Обработка исключений. IPython, Jupyter Notebook. Подключение библиотек, создание собственных модулей. Элементы функционального программирования (lambda, map, zip, reduce, filter). Чтение и запись данных в текстовом формате.

Тема 6. ОБРАБОТКА ДАННЫХ. МАССИВЫ И ВЕКТОРНЫЕ ВЫЧИСЛЕНИЯ. ПОСТРОЕНИЕ ГРАФИКОВ И ВИЗУАЛИЗАЦИЯ ДАННЫХ

Основы NumPy: многомерные массивы и векторные вычисления. Индексирование и вырезание. Универсальные функции: быстрые поэлементные операции над массивами. Обработка данных с применением массивов. Методы булевых массивов. Сортировка. Устранение дубликатов и другие теоретико-множественные операции. Файловый ввод-вывод массивов. Линейная алгебра. Генерация случайных чисел. Визуализация данных в

Python. Обзор библиотек: matplotlib, pandas, seaborn, plotly, bokeh. Базовые типы визуализаций: графики, столбчатые диаграммы, гистограммы, точечные диаграммы (scatter plots), ящики с усами. Комбинирование различных графических элементов. Построение интерактивных диаграмм с помощью plotly и bokeh.

Тема 7. СПЕЦИАЛИЗИРОВАННЫЕ БИБЛИОТЕКИ PYTHON ДЛЯ АНАЛИЗА ДАННЫХ

Введение в анализ табличных данных в Python. Пакет pandas. Объекты Series (последовательность) и DataFrame (таблица). Чтение-запись данных в различных форматах. Запросы к таблицам: выборка строк/столбцов по заданным критериям. Переформатирование данных: очистка, преобразование, слияние, изменение формы. Фильтрация отсутствующих данных. Агрегирование данных и групповые операции. Основы работы с временными рядами.

Тема 8. СБОР ДАННЫХ PYTHON. ИЗВЛЕЧЕНИЕ ДАННЫХ С ВЕБ-СТРАНИЦ (WEB-SCRAPING). ИСПОЛЬЗОВАНИЕ ОТКРЫТЫХ ДАННЫХ ЧЕРЕЗ API

Сбор и подготовка данных. Извлечение данных с веб-страниц (web-scraping). Библиотека BeautifulSoup. Работа с динамическими сайтами с помощью Selenium. Массовый скреппинг с помощью scrapy. Работа со структурированными данными: JSON и XML. Разбор XML с помощью lxml.objectify. Взаимодействие с HTML и Web API.

4.3. Лекции

№ п/п	Название темы	Очная форма	Заочная форма
1.	Введение в дисциплину. Информационные технологии статистической обработки	2	1
2.	Apache OpenOffice Calc: Описательная статистика	2	1
3.	Apache OpenOffice Calc: Дисперсионный анализ, одномерный и двумерный	2	1
4.	Eviews: Авторегрессионное моделирование. Анализ временных рядов и прогнозирование	4	2
5.	Введение в программирование на языке Python	4	2
6.	Обработка данных. Массивы и векторные вычисления. Построение графиков и визуализация данных	4	2
7.	Специализированные библиотеки Python для анализа данных	6	3
8.	Сбор данных Python. Извлечение данных с веб-страниц (web-scraping). Использование открытых данных через API	4	2
Итого:		28	14

4.4. Практические (семинарские) занятия

№ п/п	Название темы	Очная форма	Заочная форма
1.	Введение в дисциплину. Информационные технологии статистической обработки	2	1
2.	Apache OpenOffice Calc: Описательная статистика	2	1
3.	Apache OpenOffice Calc: Дисперсионный анализ, одномерный и двумерный	2	1
4.	Eviews: Авторегрессионное моделирование. Анализ временных рядов и прогнозирование	4	2
5.	Введение в программирование на языке Python	4	2
6.	Обработка данных. Массивы и векторные вычисления. Построение графиков и визуализация данных	4	2
7.	Специализированные библиотеки Python для анализа данных	6	3
8.	Сбор данных Python. Извлечение данных с веб-страниц (web-scraping). Использование открытых данных через API	4	2
Итого:		28	14

4.5. Лабораторные работы (не предусмотрены)

4.6. Самостоятельная работа студентов

№ п/п	Название темы	Вид СРС	Очная форма	Заочная форма
1.	Введение в дисциплину. Информационные технологии статистической обработки	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	4	8
2.	Apache OpenOffice Calc: Описательная статистика	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	6	8
3.	Apache OpenOffice Calc: Дисперсионный анализ, одномерный и двумерный	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	6	8
4.	Eviews: Авторегрессионное моделирование. Анализ временных рядов и прогнозирование	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	6	10
5.	Введение в программирование на языке Python	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	16	18
6.	Обработка данных. Массивы и векторные вычисления.	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	14	20

	Построение графиков и визуализация данных			
7.	Специализированные библиотеки Python для анализа данных	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	18	20
8.	Сбор данных Python. Извлечение данных с веб-страниц (web-scraping). Использование открытых данных через API	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	14	20
	Зачет	Подготовка к зачету	4	4
Итого:			88	116

4.7. Курсовые работы/проекты не предполагаются учебным планом.

5. Образовательные технологии

Преподавание дисциплины ведется с применением следующих видов образовательных технологий: объяснительно-иллюстративного обучения (технология поддерживающего обучения, технология проведения учебной дискуссии), информационных технологий (презентационные материалы), развивающих и инновационных образовательных технологий.

Практические занятия проводятся с использованием развивающих, проблемных, проектных, информационных (использование электронных образовательных ресурсов (электронный конспект) образовательных технологий.

6. Формы контроля освоения дисциплины

Текущая аттестация студентов производится в дискретные временные интервалы лектором и преподавателем(ями), ведущими практические занятия по дисциплине в следующих формах:

- вопросы для обсуждения (в виде докладов и сообщений);
- тесты,
- индивидуальные задания;
- творческое задание.

Промежуточная аттестации по результатам освоения дисциплины проходит в форме зачета (включает в себя ответы на теоретические вопросы). Студенты, выполнившие 75% текущих и контрольных мероприятий на «отлично», а остальные 25 % на «хорошо», имеют право на получение итоговой оценки.

В экзаменационную ведомость и зачетную книжку выставляются оценки по шкале, приведенной в таблице.

Шкала оценивания (зачет с оценкой)	Характеристика знания предмета и ответов	Зачеты
отлично (5)	Студент глубоко и в полном объёме владеет программным материалом. Грамотно, исчерпывающе и логично его излагает в устной или письменной форме. При этом знает рекомендованную литературу, проявляет творческий подход в ответах на вопросы и правильно обосновывает принятые решения, хорошо владеет умениями и навыками при выполнении практических задач.	зачтено
хорошо (4)	Студент знает программный материал, грамотно и по сути излагает его в устной или письменной форме, допуская незначительные неточности в утверждениях, трактовках, определениях и категориях или незначительное количество ошибок. При этом владеет необходимыми умениями и навыками при выполнении практических задач.	
удовлетворительно (3)	Студент знает только основной программный материал, допускает неточности, недостаточно чёткие формулировки, непоследовательность в ответах, излагаемых в устной или письменной форме. При этом недостаточно владеет умениями и навыками при выполнении практических задач. Допускает до 30% ошибок в излагаемых ответах.	
неудовлетворительно (2)	Студент не знает значительной части программного материала. При этом допускает принципиальные ошибки в доказательствах, в трактовке понятий и категорий, проявляет низкую культуру знаний, не владеет основными умениями и навыками при выполнении практических задач. Студент отказывается от ответов на дополнительные вопросы.	не зачтено

7. Учебно-методическое и информационное обеспечение дисциплины:

1. Баранникова И.В., Теоретические основы автоматизированной обработки информации и управления: специальные функции MS Excel / И.В. Баранникова, Е.С. Могирева, О.Г. Харахан - М.: МИСиС, 2018. - 61 с. - ISBN -- Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: http://www.studentlibrary.ru/book/misis_0009.html.

2. Форман Дж., Много цифр: Анализ больших данных при помощи Excel / Джон Форман; Пер. с англ. А. Соколовой. - М.: Альпина Паблишер, 2016. - 461 с. - ISBN 978-5-9614-5032-3 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785961450323.html>.

3. Боровиков В.П., Популярное введение в современный анализ данных в системе STATISTICA: Учебное пособие для вузов / Боровиков В.П. - М.: Горячая линия - Телеком, 2013. - 288 с. - ISBN 978-5-9912-0326-5 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785991203265.html>.

4. Златопольский Д.М., Основы программирования на языке Python: учебник / Златопольский Д. М. - М.: ДМК Пресс, 2017. - 284 с. - ISBN 978-5-97060-552-3 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785970605523.html>.

5. Шелудько В.М., Основы программирования на языке высокого уровня Python: учебное пособие / Шелудько В. М. - Ростов н/Д: Изд-во ЮФУ, 2017. - 146 с. - ISBN 978-5-9275-2649-9 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785927526499.html>.

б) дополнительная литература:

1. Шелудько В.М., Язык программирования высокого уровня Python. Функции, структуры данных, дополнительные модули: учебное пособие / Шелудько В. М. - Ростов н/Д: Изд-во ЮФУ, 2017. - 107 с. - ISBN 978-5-9275-2648-2 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785927526482.html>.

2. Рутковская А.Э., Офисное программирование. Лабораторный практикум: учеб. пособие / А.Э. Рутковская - Минск: РИПО, 2017. - 146 с. - ISBN 978-985-503-705-8 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9789855037058.html>.

3. Биллиг В.А., Основы офисного программирования и документы Excel / Биллиг В.А. - М.: Национальный Открытый Университет "ИНТУИТ", 2016. - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: http://www.studentlibrary.ru/book/intuit_192.html.

4. Сдвижков О.А., Непараметрическая статистика в MS Excel и VBA / Сдвижков О.А. - М.: ДМК Пресс, 2014. - 172 с. - ISBN 978-5-94074-917-2 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785940749172.html>.

5. Лялин В.С., Статистика: теория и практика в Excel: учеб. пособие / В.С. Лялин, И.Г. Зверева, Н.Г. Никифорова. - М.: Финансы и статистика, 2010. - 448 с. - ISBN 978-5-279-03381-2 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785279033812.html>.

6. Александровская Ю.П., Многомерный статистический анализ в экономике: учебное пособие / Александровская Ю. П. - Казань: Издательство КНИТУ, 2017. - 96 с. - ISBN 978-5-7882-2191-5 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785788221915.html>.

7. Груздев А.В., Прогнозное моделирование в IBM SPSS Statistics и R: Метод деревьев решений / Груздев А. В. - М.: ДМК Пресс, 2016. - 278 с. - ISBN 978-5-97060-456-4 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785970604564.html>.

в) методические рекомендации

1. Методические указания к лабораторным работам по дисциплине «Информационные технологии обработки статистической информации» для студентов направления подготовки 38.04.05 – Бизнес-информатика [Электронный ресурс] / сост. С.А. Аст. – Луганск: ЛНУ им. В. Даля, 2019. – 69 с.

2. Методические указания к самостоятельной работе по дисциплине «Информационные технологии обработки статистической информации» для студентов направления подготовки 38.04.05 – Бизнес-информатика [Электронный ресурс] / сост. С.А. Аст. – Луганск: ЛНУ им. В. Даля, 2019. – 33 с.

г) Интернет-ресурсы:

1. Министерство образования и науки Российской Федерации [Электронный ресурс] / Режим доступа: <http://минобрнауки.рф/>

2. Федеральная служба по надзору в сфере образования и науки [Электронный ресурс] / Режим доступа: <http://obrnadzor.gov.ru/>

3. Министерство образования и науки Луганской Народной Республики [Электронный ресурс] / Режим доступа: <https://minobr.su>

4. Народный совет Луганской Народной Республики [Электронный ресурс] / Режим доступа: <https://nslnr.su>

5. Портал Федеральных государственных образовательных стандартов высшего образования [Электронный ресурс] / Режим доступа: <http://fgosvo.ru>

6. Федеральный портал «Российское образование» [Электронный ресурс] / Режим доступа: <http://www.edu.ru/>

7. Информационная система «Единое окно доступа к образовательным ресурсам» [Электронный ресурс] / Режим доступа: <http://window.edu.ru/>

8. Федеральный центр информационно-образовательных ресурсов [Электронный ресурс] / Режим доступа: <http://fcior.edu.ru/>

9. Электронно-библиотечная система «Консультант студента» [Электронный ресурс] / Режим доступа: <http://www.studentlibrary.ru/cgi-bin/mb4x>

10. Электронно-библиотечная система «StudMed.ru» [Электронный ресурс] / Режим доступа: <https://www.studmed.ru>

11. Научная библиотека имени А. Н. Коняева [Электронный ресурс] / Режим доступа: <http://biblio.dahluniver.ru/>

12. Сайте «Репетитор оценщика». URL: <http://dom-khv.ucoz.ru/>
Сайте «Репетитор оценщика». Практика. URL: <http://dom-khv.ucoz.ru/index/analitika/0-287>

8. Материально-техническое и программное обеспечение дисциплины

На лекциях и практических занятиях используется видеоаппаратура: проектор. Практические занятия проводятся в компьютерном классе. Презентации и материалы (тексты источников, вопросы и задания) в

электронном виде выкладываются в сетевые папки. Курс представлен в системе дистанционного обучения <https://ecps.gnomio.com/>

Программное обеспечение:

Функциональное назначение	Бесплатное программное обеспечение	Ссылки
Операционная система	UBUNTU 19.04	https://ubuntu.com/ https://ru.wikipedia.org/wiki/Ubuntu
Браузер	Firefox Mozilla	http://www.mozilla.org/ru/firefox/fx
Офисный пакет	Libre Office 6.3.1	https://www.libreoffice.org/ https://ru.wikipedia.org/wiki/LibreOffice
Файл-менеджер	Far Manager	http://www.farmanager.com/download.php
Архиватор	7Zip	http://www.7-zip.org/
Редактор PDF	PDFCreator	http://www.pdfforge.org/pdfcreator
Графический редактор	GIMP (GNU Image Manipulation Program)	http://www.gimp.org/ http://gimp.ru/viewpage.php?page_id=8 http://ru.wikipedia.org/wiki/GIMP
Аудиоплеер	VLC	http://www.videolan.org/vlc/

9. Оценочные средства по дисциплине

Паспорт оценочных средств по учебной дисциплине «Информационные технологии обработки статистической информации»

Перечень компетенций (элементов компетенций), формируемых в результате освоения учебной дисциплины (модуля) или практики

№ п/п	Код контролируемой компетенции	Формулировка контролируемой компетенции	Индикаторы достижений компетенции (по реализуемой дисциплине)	Контролируемые темы учебной дисциплины, практики	Этапы формирования (семестр изучения)
1	ОПК-3	Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта	ОПК-3.3	Темы 1-8	3
	ПК-4	Способен анализировать статистические данные в соответствии с базовой методологией и использованием программных средств статистического анализа	ПК-4.4	Темы 1-8	3

**Показатели и критерии оценивания компетенций, описание шкал
оценивания**

№ п/п	Код контролируемой компетенции	Индикаторы достижений компетенции (по реализуемой дисциплине)	Перечень планируемых результатов	Контролируемые темы учебной дисциплины	Наименование оценочного средства
1.	ОПК-3	ОПК-3.3. Выявляет возможности, создаваемые информационными и цифровыми технологиями	Знать: информационные и цифровые технологии для проведения статистического анализа; Уметь: использовать информационные и цифровые технологии для проведения статистического анализа; Владеть: информационными и цифровыми технологиями для проведения статистического анализа.	Темы 1-8	Вопросы для обсуждения (в виде докладов и сообщений); тесты, индивидуальные задания; творческое задание
2.	ПК-4	ПК-4.4. Способен проводить статистический анализ различных экономических процессов и явлений с использованием программных средств статистического анализа	Знать: методику проведения статистического анализа различных экономических явлений и процессов с использованием программных средств; Уметь: проводить статистический анализ с использованием программных средств; Владеть: программными средствами обработки статистических данных.	Темы 1-8	Вопросы для обсуждения (в виде докладов и сообщений); тесты, индивидуальные задания; творческое задание

Фонды оценочных средств по дисциплине «Информационные технологии обработки статистической информации»

Вопросы для обсуждения (в виде докладов и сообщений)

1. Перечислите возможные практические задачи, которые целесообразно решить программными методами.
2. Какие показатели являются результатом вывода описательной статистики в Apache OpenOffice Calc?
3. Какие основные ограничения наличествуют в Apache OpenOffice Calc для построения описательной статистики?
4. Как интерпретируется уровень надежности?
5. Перечислите действия, которые необходимо выполнить для построения теоретической кривой нормального распределения.
6. Какие условия должны выполняться для проведения дисперсионного анализа?
7. Перечислите действия и необходимые показатели для построения блочных диаграмм в Apache OpenOffice Calc
8. Как интерпретируются результатные показатели дисперсионного анализа в Apache OpenOffice Calc?
9. Как оценить уровень значимости результатов анализа?
10. Какие прикладные задачи позволяет решить метод снижения размерности?
11. В чем заключается предварительный этап выбор факторов для моделирования?
12. Что характеризуют «собственные значения» (Eigenvalues)?
13. Что является конечной целью анализа соответствий?
14. Сформулируйте постановку задачи и вывод результатов для подтверждающего факторного анализа.
15. Назовите основные этапы множественного регрессионного анализа.
16. Сформулируйте сущность и необходимую информационную базу для применения метода «добычи данных».
17. Что представляют собой хранилища данных и какова их структура?
18. Что такое оперативная обработка данных (OLAP) и как она реализуется?
19. Что представляет собой разведочный анализ данных (РАД)?
20. Что такое нейронные сети, какие результаты и интерпретацию они позволяют получить?
21. Сформулируйте основные направления и методы кластерного анализа.
22. Что представляет собой формальная постановка задачи и какие данные требуются для иерархической кластеризации?
23. Назовите необходимые исходные данные и постановку задачи для метода К средних.
24. Опишите в общих чертах сущность и задачи авторегрессионного моделирования, назовите условия его применения.

25. Как проводится выбор количества лаговых переменных? Опишите визуальный способ и назовите количественные критерии, сформулируйте возможные различия между ними.

26. Как проводится выбор количества сезонных переменных?

27. Перечислите требования, предъявляемые к статистической информационной системе, как к сложной системе.

28. Файлы и исключения. Работа с внешними источниками данных.

29. Исключения, обработка исключений, вызов исключений (try-except-finally).

30. Утверждения (assert). Открытие, чтение, запись. (open, инструкция with).

31. Работа с текстовыми файлами, xml и csv - файлами.

32. Функциональное программирование. Лямбда-функции.

33. Использование функций map, filter, reduce, zip.

34. Генераторы, декораторы, рекурсия.

35. Модификация функций с помощью декораторов.

36. Итерируемые объекты. Использование генераторов (yield).

37. ООП в Python. Классы, объекты и экземпляры классов. Наследование.

38. Магические методы. Переопределение операторов. Методы классов.

39. Инкапсуляция. Условно частные и строго частные методы.

40. Регулярные выражения. Использование регулярных выражений.

Пакет re.

41. Наука о данных и Python. Библиотеки: NumPy, pandas, matplotlib, SciPy.

42. Основы NumPy: массивы и векторные вычисления.

43. Инструменты визуализации данных для Python.

44. Введение в API библиотеки matplotlib.

45. Библиотека pandas. Введение в структуры данных pandas.

46. Объекты Dataframe и Series.

47. Визуализация данных в pandas. Seaborn.

48. Агрегирование данных и групповые операции. 36. Сбор и подготовка данных в Python: извлечение данных с web-страниц (web-scraping). Библиотека BeautifulSoup.

49. Работа с динамическими сайтами с помощью Selenium.

50. Массовый скрепинг с помощью scrapy.

51. Работа со структурированными данными: JSON и XML.

52. Открытые API на примере Telegram API, VK API.

Критерии и шкала оценивания по оценочному средству «вопросы для обсуждения»

Шкала оценивания (интервал баллов)	Критерий оценивания
5	Доклад (сообщение) представлен(о) на высоком уровне (студент в полном объеме осветил рассматриваемую проблематику, привел аргументы в пользу своих суждений, владеет профильным понятийным (категориальным) аппаратом и т.п.)
4	Доклад (сообщение) представлен(о) на среднем уровне (студент в целом осветил рассматриваемую проблематику, привел аргументы в пользу своих суждений, допустив некоторые неточности и т.п.)
3	Доклад (сообщение) представлен(о) на низком уровне (студент допустил существенные неточности, изложил материал с ошибками, не владеет в достаточной степени профильным категориальным аппаратом и т.п.)
2	Доклад (сообщение) представлен(о) на неудовлетворительном уровне или не представлен (студент не готов, не выполнил задание и т.п.)

Индивидуальные задания

Типовые контрольные задания для проведения текущего контроля

Задание 1. Импортируйте набор данных о пассажирах “Титаника”. Проведите разведывательный анализ данных. Поставьте гипотезу. Проведите необходимые преобразования данных.

Задание 2. На основании поставленной гипотезы в предыдущем задании выберите модели для предсказания. Обоснуйте выбор. Постройте выбранные модели, выберите лучшую. Постройте ансамбль моделей. Улучшились ли предсказания? Оформите результаты в виде отчета.

Задание 3. Задание выполняется в табличном процессоре Excel. Рекомендованное время работы над документом – 2 часа 30 минут. Контрольное задание содержит 30 вариантов. Данные размещены в СДО кафедры ЭКиПС по адресу <https://ecpsdahl.ru/>

В таблице приведен фрагмент таблицы из книги Макконнел К., Брю С. Экономикс: Принципы, проблемы и политика. Т.2.- М.: Республика, 1992. — 400 Таблица содержит данные по 29 экономическим показателям за 10 лет. Выполните следующее:

1. Выберите один показатель для упражнения в соответствии с вашим вариантом.

2. Используя табличный процессор, попытайтесь найти истинную форму тренда показателя (а) за 6 лет, (б) за 9 лет. Какой из трендов недоступен в случае (а) и почему?

3. Оцените качества всех трендов, построенных в ходе поиска.

4. Сделайте прогноз: (а) на 7-й, (б) на 10-й год по найденному истинному тренду.

5. Сделайте точечный прогноз на один год вперед по простой статистической функции FORECAST (ПРЕДСКАЗ).

6. Сравните результаты учебных прогнозов по тренду и по FORECAST (ПРЕДСКАЗ) между собой, а также с известными фактическими данными.

8. Определите степень отклонения вашего прогноза от фактического значения (в процентах).

9. Сделайте выводы о прогнозе с формальной точки зрения и с точки зрения здравого смысла.

10. Оформите все ваши поисковые расчеты, аргументы и выводы.

11. Если студент желает удвоить количество оценочных рейтинговых баллов, то необходимо подобрать актуальные современные данные (с обязательными ссылками на источник). Обработать данные с помощью изученной технологии. Оформить результаты, дать заключительную справку о возможном прогнозе выбранного показателя на конкретный период.

Задание 4. Визуализация значений заданных переменных с использованием статистических графиков. Для файла с исходными данными (СДО <https://ecpsdahl.ru/>) построить гистограммы (команда Graphs/Histograms) для двух переменных на одном графике в зависимости от номера вашего варианта. Использовать опцию Multiple (несколько графиков на одной сетке) во вкладке Quick.

Для тех же переменных построить столбчатую диаграмму (Graphs/2D Graphs/Bar Columns Plots).

Для переменной VarN построить круговую диаграмму (Pie Chart-Counts). Обратит внимание, как строится график Pie Chart при изменении переменной Categories.

Для переменных VarN, VarN+1, VarN+2 построить 3D график. Обратит внимание, что трехмерный график можно разворачивать на любой угол в подменю «свойства графика/все свойства». Настроить графики, подписав переменные и оси.

Задание 5. Решение задачи в Eviews. По предложенным вам экспериментальным данным, представляющим собою макроэкономические показатели или показатели финансовой (денежно-кредитной) системы некоторой страны, т.е. случайной выборке объема n – построить математическую модель зависимости случайной величины Y от случайных величин X_1 и X_2 . Построение и оценку качества экономико-математической (эконометрической) модели вести в следующей последовательности:

1. Построить корреляционную матрицу для случайных величин и оценить статистическую значимость корреляции между ними.

2. Исходя из наличия между эндогенной переменной и экзогенными переменными, линейной зависимости, оценить параметры регрессионной модели по методу наименьших квадратов. Вычислите вектора регрессионных значений эндогенной переменной и случайных отклонений.

3. Найдите средние квадратические ошибки коэффициентов регрессии. Используя критерий Стьюдента проверьте статистическую значимость параметров модели. Здесь и далее принять уровень значимости 0,05 (т. е. надежность 95%).

4. Вычислите эмпирический коэффициент детерминации и скорректированный коэффициент детерминации. Проверьте, используя критерий Фишера, адекватность линейной модели.

5. Установите наличие (отсутствие) автокорреляции случайных отклонений модели. Используйте для этого метод графического анализа, статистику Дарбина-Уотсона и критерий Бреуша-Годфри.

6. Установите наличие (отсутствие) гетероскедастичности случайных отклонений модели. Используйте для этого графический анализ, тест Вайта и тест Парка для вариантов с добавочным индексом А (графический метод, тест Глейзера и тест Бреуша-Пагана для вариантов с добавочным индексом В).

7. Обобщите результаты оценивания параметров модели и результаты проверки модели на адекватность.

Задание 6. Используя данные (<https://ecpsdahl.ru/>), сформулируйте и проверьте гипотезу о наличии на исследуемом временном интервале точки разрыва (имеется сдвиг свободного члена или коэффициента наклона). В случае, если предварительный графический анализ не подтверждает наличия разрыва на временном интервале, примите, что точка разрыва находится посередине.

Задание 7. Введите в эконометрическую модель, построенную в задании 17 сезонные фиктивные переменные и с помощью соответствующей модели, исследуйте наличие или отсутствие сезонных колебаний.

Задание 8. Проанализировать данные фондового рынка, чтобы получить представление о том, как НММ работает с данными последовательных или временных рядов. Реализовать пример на Python.

Задание 9. Обработать и разрезать данные временных рядов с помощью Pandas. Обратите внимание, что используются данные по месячным колебаниям в Арктике, которые можно загрузить с сайта month.ao.index.b50.current.ascii и которые мы можем преобразовать в текстовый формат.

Задание 10. Написать функцию `any_news_about_harry(url)`, принимающую на вход адрес веб-страницы `url`, загружающую эту веб-страницу и проверяющую, встречается ли в ней слово Harry (с большой буквы). Функция должна возвращать `True`, если встречается, и `False` в противном случае. Также функция должна возвращать `False`, если страницу не удалось открыть (например, была получена ошибка *404 Not Found*.)

Задание 11. Написать функцию `median(x1, x2, ...)`, принимающую на вход несколько чисел и возвращающую их медиану (для чётного числа элементов возьмите среднее арифметическое между двумя серединными элементами). Пользоваться библиотечными функциями нельзя!

Задание 12. Создать список, заполнить список случайными целыми числами в указанном диапазоне и отсортировать список по возрастанию и убыванию.

Задание 13. Отсортировать заданный массив методом "сортировка пузырьком".

Задание 14. Выполнить обработку элементов прямоугольной матрицы А, имеющей N строк и М столбцов. Нормировать элементы каждой строки по отношению к наибольшему элементу этой строки.

Задание 15. Выполнить обработку элементов прямоугольной матрицы А, имеющей N строк и М столбцов. Нормировать элементы каждого столбца матрицы по отношению к наибольшему элементу этого столбца.

Задание 16. Известны данные о количестве мальчиков и девочек в нескольких классах. Отсортировать названия классов по возрастанию процента мальчиков, определить количество классов, к которых мальчиков больше, чем девочек и вывести названия этих классов отдельно.

Задание 17. Решить задачу, связанную с оценкой экономической деятельности группы предприятий на основе известных данных: название предприятий; плановый объем розничного товарооборота; фактический объем розничного товарооборота. Требуется определить:

- а) процент выполнения плана каждым предприятием;
- б) количество предприятий, недовыполнивших план на 10% и более;
- в) наименьший плановый товарооборот;
- г) упорядочить предприятия по убыванию планового товарооборота.

Критерии и шкала оценивания по оценочному средству «Индивидуальные задания»

Шкала оценивания (интервал баллов)	Критерий оценивания
5	Индивидуальное задание выполнено на высоком уровне (правильные ответы даны на 90-100% вопросов/задач)
4	Индивидуальное задание выполнено на среднем уровне (правильные ответы даны на 75-89% вопросов/задач)
3	Индивидуальное задание выполнено на низком уровне (правильные ответы даны на 50-74% вопросов/задач)
2	Индивидуальное задание выполнено на неудовлетворительном уровне (правильные ответы даны менее чем на 50%)

Творческое задание

Задание 1. Соберите данные и исследуйте описательную статистику по одному из следующих статистических наблюдений, или самостоятельно выбранной области: 1) обследование работников промышленных предприятий; 2) перепись оборудования на промышленных предприятиях; 3) обследование семей рабочих и служащих; 4) обследование строительных организаций; 5) обследование торговых предприятий; 6) изучение спроса на некоторые товары; 7) изучение общественного мнения по отдельным вопросам. По указанному наблюдению определите цель и задачи наблюдения; объекты и единицу наблюдения; основные признаки, подлежащие регистрации; вид, форму и способ наблюдения.

Задание 2. Сформулируйте прикладную задачу бизнес-аналитики (например, классификация предприятий промышленной сферы, выявление

нетривиальных закономерностей), для решения которой возможно применение программных средств статистического анализа, и найдите данные по этой задаче.

Задание 3. Используя данные о заработной плате работников по двум предприятиям, проведите дисперсионный анализ средствами Apache OpenOffice Calc и сделайте выводы о различиях заработной платы. Насколько можно доверять полученным результатам? Сформулируйте ограничения результатов анализа.

Задание 4. Используя статистический ежегодник «Россия в цифрах», выберите экономический показатель и влияющие на него факторы, проведите моделирование с использованием метода главных компонент.

Используя статистический ежегодник «Россия в цифрах» Проведите анализ соответствий по данным о среднем времени поиска работы / занятости, или другим данным.

Задание 5. Проведите множественный регрессионный анализ на основе данных о ВВП России. Проведите множественный регрессионный анализ фондового индекса (ММВБ, РТС, или любого другого) по историческим данным.

Задание 6. Выявите закономерности и изучите структуру распределения данных о потребительской корзине товаров. Изучите состав и структуру распределения клиентской базы торгового предприятия.

Задание 7. По данным о результатах торгов фондовой биржи ММВБ или РТС проведите кластеризацию эмитентов, входящих в состав фондового индекса, иерархическим методом, сделайте выводы. С помощью метода К средних проведите классификацию товаров (например, ноутбуков) по параметрам их количественных потребительских свойств. На основе базы данных клиентов определите основные параметры, влияющие на спрос, используя дискриминантный анализ.

Задание 8. По данным о результатах торгов на фондовом или валютном рынке проведите прогнозирование курса ценной бумаги или валюты с помощью моделей БоксаДженкинса. По данным макроэкономической статистики (сборник «Россия в цифрах» или официальные данные), проведите парное или многофакторное авторегрессионное моделирование и прогноз объема ВВП.

Задание 9. Соберите данные из социальной сети или с новостного сайта для последующего текстового анализа. Проведите предварительную обработку данных. Поставьте задачу классификации. Проведите необходимые преобразования данных. Постройте модель. Оцените, насколько хорошо модель справляется с поставленной задачей. Опишите результат, назовите возможные причины плохой работы модели (в случае, если результаты неудовлетворительные). Используйте для реализации язык программирования Python.

Критерии и шкала оценивания по оценочному средству творческое задание

Шкала оценивания (интервал баллов) ²	Критерий оценивания
5	Выставляется, если обучающийся/группа обучающихся выразили свое мнение по сформулированной проблеме, аргументировали его, точно определив ее содержание и составляющие. Приведены данные отечественной и зарубежной литературы, статистические сведения, информация нормативно-правового характера. Продемонстрировано знание и владение навыком самостоятельной исследовательской работы по теме исследования. Фактических ошибок, связанных с пониманием проблемы, нет.
4	Работа обучающегося/группы обучающихся характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Для аргументации приводятся данные отечественных и зарубежных авторов. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет.
3	Проведен достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимание базовых основ и теоретического обоснования выбранной темы. Не достаточно привлечены основные источники по рассматриваемой теме. Допущены ошибки в смысле или содержании проблемы.
2	Работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая составляющая темы. Допущено большое количество ошибок смыслового содержания раскрываемой проблемы. Работа не представлена.

Тесты

Задания закрытого типа

1. Для проверки адекватности математической модели объекта исследования используют критерий:

- a) Кохрена.
- b) Стьюдента.
- c) Фишера.
- d) Пирсона.

2. В ситуации, если необходимо образовать заранее известное число групп предприятий для проведения классификации, применяется метод кластерного анализа:

- a) иерархический;
- б) метод К средних;
- в) ассоциативный;
- г) дискриминантный.

3. При отсутствии априорных представлений о характере и природе связей между переменными, для выявления и анализа взаимосвязи применяется:

- a) дисперсионный анализ;
- б) кластерный анализ;
- в) множественный регрессионный анализ;

г) метод «добычи данных».

4. В состав описательной статистики не входит такая характеристика:

- а) среднее;
- б) стандартная ошибка;
- в) медиана;
- г) мода;
- д) коэффициент корреляции.

5. При моделировании рядов данных на основе авторегрессионных моделей Бокса-Дженкинса другими переменными, необходимо, чтобы выполнялось условие:

- а) стационарности ряда;
- б) равенстве дисперсий;
- в) равенстве периода временного лага;

6. Целью факторного анализа методом главных компонент может являться:

- а) сокращение числа переменных;
- б) определение структуры взаимосвязей;
- в) классификация данных;
- г) всё вышеперечисленное;
- д) только а) и б)

7. Для воспроизведения нетривиальных сложных зависимостей используется:

- а) авторегрессионное моделирование;
- б) «нейронные сети»;
- в) кластерный анализ;
- г) метод главных компонент.

8. Основные задачи статистической обработки данных, которые наиболее важны по своим практическим применениям:

- а) выбор способов сбора и группировки статистических сведений
- б) определение законов распределения случайной величины по статистическим данным
- в) проверка правдоподобия гипотез
- г) нахождение неизвестных параметров распределения
- получение статистической информации

9. Дискретные случайные величины среди приведенных величин

- а) «Количество мальчиков, рожденных за тот или иной год»
- б) «Число очков, выпадающих при бросании игральной кости»
- в) «Количество секунд в одном часе»
- г) «Значение функции на множестве действительных чисел»

10. Мода в математической статистике:

- а) значение признака (варианты), которое чаще всего встречается в данной совокупности
- б) значение признака (варианты), которое находится в середине ранжированного ряда
- в) среднее арифметическое значение признака (варианты)

11. Что будет напечатано в результате выполнения кода:

```
list = [1, 2, 3, 4]
del list[2:3]
print(list)
```

- а) [1, 2, 3, 4]
- б) [1, 2, 4]
- в) [1, 4]
- г) [1, 2]
- д) возникнет ошибка

12. Каким будет результат выполнения следующего кода:

```
value = 0
value = value > 0 ? 0 : 1
print(value)
```

- а) 0
- б) 1
- в) возникнет ошибка

13. Что будет напечатано в результате выполнения следующего кода:

```
1: qwe = 'qwe'
2: qwe[0] = 'a'
3: print(qwe)
```

- а) qwe
- б) awe
- в) возникнет ошибка в строке 2
- г) возникнет ошибка в строке 3

14. Каким будет результат:

```
value = 0
def func(arg=1):
    arg = arg + 1
    return arg
print(func() + func(value))
```

- а) 1
- б) 2
- в) 3
- г) 4
- д) возникнет синтаксическая ошибка

15. Что будет напечатано в результате выполнения кода:

```
list = [1, 2, 3, 4]
del list[2:3]
print(list)
```

- а) [1, 2, 3, 4]
- б) [1, 2, 4]
- в) [1, 4]
- г) [1, 2]
- д) возникнет ошибка

16. Каким будет результат выполнения следующего фрагмента кода:


```
for i in range(3):
    print(i, end=' ')
```

- а) 3
- б) 0 1 2
- в) 1 2 3
- г) 0 1 2 3
- д) возникнет ошибка

17. Что напечатает следующий код:

```
value = 0
if value > 0:
    print('>0')
elif:
    print('<=0')
```

- а) 0
- б) код ничего не напечатает
- в) возникнет синтаксическая ошибка

18. Имея следующую последовательность действий, чему равна переменная L2 ?:

```
>>> L1 = [2, 3, 4]
>>> L2 = L1
>>> L1[0] = 24
>>> L1
[24, 3, 4]
>>> L2
```

- а) [2, 3, 4]
- б) [24, 3, 4]
- в) [2, 3, 24]
- г) [2, 4, 3]

19. Какая из функций вернет итерируемый объект?

- а) len()
- б) xrange()
- в) range()
- г) ord()

20. Как можно вызвать метод func у следующего класса (выберите все подходящие варианты):

```
class myClass:
    def func(self):
        print('hello')
myClass.func()
```

- а) obj = myClass() obj.func()
- б) obj = myClass() myClass.func(obj)
- в) obj = myClass() obj.func
- г) ни один из перечисленных

21. Имея следующую последовательность действий, чему равна переменная L2 ?:

```
>>> L1 = [2, 3, 4]
>>> L2 = L1
>>> L1[0] = 24
>>> L1
[24, 3, 4]
>>> L2
```

- а) [2, 3, 4]
- б) [24, 3, 4]
- в) [2, 3, 24]
- г) [2, 4, 3]

22. Что напечатает следующий код:

```
for i in range(3):
    if i < 1:
        print(i)
    else:
        print(i)
        break
```

- а) 0
- б) 0 1
- в) 0 1 2
- г) ничего не напечатает
- д) возникнет ошибка

23. Каким будет результат выполнения данной программы?

```
list = []

for i in range(100):
    list.append(lambda x, i = i : x+i)

list[42](3)
```

- а) 42
- б) 45
- в) 102
- г) ошибка интерпретатора

24. Укажите модули входящие в стандартный набор Python:

- а) math
- б) smath
- в) numpy
- г) PIL

25. Библиотека statsmodels – это

а) модуль для Python, который дает легкий доступ к различным статистическим инструментам для проведения статистических тестов и изучения данных.

б) модуль для SPSS, который дает легкий доступ к различным статистическим инструментам для проведения статистических тестов и изучения данных.

в) модуль для подготовки набора данных к запуску регрессий, ANOVAs и других тестов, но в целом набор данных готов к работе;

г) нет правильного ответа.

26. Нормальное распределение является наиболее важным распределением вероятностей в статистике, потому что

а) многие процессы в природе следуют нормальному распределению.

б) некоторые из примеров - возраст, рост, вес и артериальное давление человека.

в) линейная регрессия предполагает, что ошибки или остатки подчиняются нормальному распределению.

г) некоторые алгоритмы машинного обучения, такие как линейный дискриминантный анализ и квадратичный дискриминантный анализ, основаны на предположении нормального распределения.

27. Метод `transform`, который похож на `apply`, но налагает больше ограничений на то, какие функции можно использовать:

а) функция может возвращать скалярное значение, которое будет уложено на группу;

б) функция может возвращать объект такой же формы, как входная группа;

в) функция не должна модифицировать свой вход.

г) все ответы верные.

28. Перекрестная таблица (`cross-tabulation`, или для краткости `crosstab`) – частный случай сводной таблицы, в которой представлены

а) частоты групп;

б) частости групп;

в) середины интервалов;

г) квантили.

29. Перечислите несколько видов моделей, имеющих в `statsmodels`:

а) линейные модели, обобщенные линейные модели и робастные линейные модели;

б) линейные модели со смешанными эффектами;

в) методы дисперсионного анализа (ANOVA);

г) временные ряды и модели в пространстве состояний;

д) обобщенные методы моментов.

30. Под функцией "`groupby`" понимают процесс, включающий один или более шагов:

а) разделение данных на группы по каким-либо критериям;

б) применение функции к каждой группе независимо друг от друга;

в) объединение результатов в единую структуру данных;

г) все шаги верные.

31. Самый простой способ сформировать список:

а) перечисление в квадратных скобках через запятую;

б) перечисление в квадратных скобках через точку;

в) деление элементов;

г) умножение элементов в круглых скобках.

Задания открытого типа

1. Какие показатели являются результатом вывода описательной статистики в Apache OpenOffice Calc?

2. Какие основные ограничения наличествуют в Apache OpenOffice Calc для построения описательной статистики? Как интерпретируется уровень надежности?

3. Перечислите действия, которые необходимо выполнить для построения теоретической кривой нормального распределения.

4. Какие условия должны выполняться для проведения дисперсионного анализа?

5. Перечислите действия и необходимые показатели для построения блочных диаграмм в Apache OpenOffice Calc

6. Как интерпретируются результатные показатели дисперсионного анализа в Apache OpenOffice Calc? \

7. В состав описательной статистики входят такие характеристики: среднее; стандартная ошибка; медиана; мода; стандартное отклонение; дисперсия выборки; эксцесс; асимметричность; интервал; минимум ; максимум; сумма; счет.

(Верно / неверно)

8. Какие установки нужно выполнить в поле Transform variable (series) prior to analysis (Преобразовать переменную (ряд) перед проведением анализа) прежде, чем строить прогноз значений ряда?

9. Как проверяется качество модели временного ряда, полученной с помощью процедуры ARIMA?

10. Файл в формате CSV (comma-separated values - значения, разделенные запятыми) – текстовая структура данных, предназначенная для переноса табличной информации между приложениями (электронными таблицами, СУБД, адресными книгами и т.п.).

(Верно / неверно)

11. Методы plot.bar() и plot.barh() строят соответственно вертикальную и горизонтальную столбчатые диаграммы. В этом случае индекс Series или DataFrame будет использоваться для нанесения рисок на ось x (bar) или y (bar)

(Верно / неверно)

12. Функция plot с параметром kind='kde' строит график плотности, применяя стандартный метод комбинирования нормальных распределений. Для построения таких графиков необходима библиотека SciPy.

Seaborn еще упрощает построение гистограмм и графиков плотности благодаря методу histplot, который может строить одновременно гистограмму и непрерывную оценку плотности.

(Верно / неверно)

13. Функция regress (воспользовавшись эконометрической библиотекой statsmodels), которая вычисляет регрессию методом обыкновенных наименьших квадратов для каждого блока данных.

(Верно / неверно)

14. Какие типы данных для представления даты и времени, а также средства, относящиеся к календарю имеются в стандартной библиотеке Python.

(datetime, time и calendar)

15. Как и для других объектов Series, арифметические операции над временными рядами с различными индексами автоматически приводят к выравниванию дат.

(Верно / неверно)

16. Под передискретизацией понимается процесс изменения частоты временного ряда

(Верно / неверно)

17. Есть два способа задать коэффициент затухания, самый популярный – использовать промежуток (span). В pandas имеется оператор ewm (exponentially weighted moving – экспоненциально взвешенный сдвиг), который работает совместно с rolling и expanding.

(Верно / неверно)

18. Линейные модели в statsmodels имеют два основных интерфейса: на основе массивов и на основе формул.

(Верно / неверно)

19. Библиотека Matplotlib обеспечивает объектно-ориентированный интерфейс для встраивания графиков в приложения, используя инструменты GUI общего назначения, такие как WxPython, Qt, или GTK+. Существует также процедурный ruIab-интерфейс напоминающий MATLAB. Библиотека научных расчетов SciPy также использует matplotlib для работы с графиками.

(Верно / неверно)

20. Кто является разработчиком Python?

(Гвидо ван Россум)

21. Значение p вычисляется с использованием статистики теста, возвращаемой функцией kstest. Если $pvalue > \alpha$, мы не можем отклонить нулевую гипотезу, в противном случае принимаем альтернативную гипотезу

(Верно / неверно)

22. Как выполняется считывание данных в Python?

23. Какие проблемы при импорте данных могут возникнуть, какие способы их решения могут быть использованы?

24. Есть ли в наборе данных количественная дискретная переменная?

25. Есть ли в данных категориальные переменные? Как преобразовать их в числовые векторы?

26. Есть ли в данных количественные переменные? Какой у них тип?

27. Как построить срез данных по условию? Как провести сортировку набора данных по одному из столбцов?

28. С помощью параметров описательной статистики опишите одну из количественных переменных и одну из категориальных переменных.

29. Есть ли в количественной переменной выбросы? Какое значение является выбросом и почему?

30. По какой категориальной переменной можно группировать данные? Какие группировки представляют интерес и почему?

31. Как рассчитать параметры описательной статистики для переменных набора с использованием методов библиотеки pandas языка Python.

32. Как провести тестирование гипотез с использованием методов на языке Python.

33. Какие гипотезы нужно протестировать для ответа на исследовательский вопрос? Альтернативная гипотеза, сформулированная вами, является односторонней или двусторонней?

34. Какой статистический критерий необходимо использовать при тестировании гипотез? _____

35. Что позволяет определить уровень значимости при тестировании гипотез? _____

36. Какие методы на языке Python использовались для тестирования гипотез? _____

37. Какие визуализации построены для количественных переменных? Какие методы библиотек на языке Python и с какими входными параметрами при этом использовались? Какие особенности распределения данных они позволяют наблюдать?

38. Какие визуализации построены для категориальных переменных? Какие методы библиотек на языке Python и с какими входными параметрами при этом использовались? Какие особенности распределения данных они позволяют наблюдать?

39. Какие визуализации построены для пар переменных? Какие методы библиотек на языке Python и с какими входными параметрами при этом использовались? Какие особенности распределения данных они позволяют наблюдать?

На сопоставление

1. Показатели описательной статистики можно разбить на несколько групп (Соотнесите):

1. Показатели положения:

2. Показатели разброса:

3. Показатели асимметрии:

4. Показатели, описывающие закон распределения:

а) Эти показатели определяют, насколько кучно основная масса данных группируется около центра.

б) характеризуют симметрию распределения данных около своего центра. К ним можно отнести коэффициент асимметрии, положение медианы относительно среднего и т. п.

в) описывают положение данных на

числовой оси. Примеры таких показателей - минимальный и максимальный элементы выборки (первый и последний члены вариационного ряда), верхний и нижний квартили (ограничивают зону, в которую попадают 50% центральных элементов выборки).

г) дают представление о законе распределения данных. Сюда относятся таблицы частот, таблицы частостей, полигоны, кумуляты, гистограммы.

Ответ 1-В; 2-А; 3-Б; 4-Г.

2. Соотнесите:

- | | |
|--|---|
| 1. Модуль Дисперсионный анализ (ANOVA/MANOVA): | а) позволяет проводить факторный анализ, основная цель которого заключается в том, чтобы выделить скрытые общие факторы, т.е. воздействующие на все параметры объекта, а не на какой-то один параметр или группу. |
| 2. Модуль Факторный анализ (Factor analysis): | б) дает возможность оценить степень воздействия различных факторов на измеряемые данные и выделить среди них наиболее значимые (существенные). |
| 3. Модуль Непараметрическая статистика и подгонка распределения (Nonparametrics/Distribution): | в) дает возможность сравнить распределение наблюдаемых величин с большим количеством различных теоретических распределений. |

Ответ 1-Б; 2-А; 3-В.

3. Структура диалога пользователя в каждом статистическом модуле имеет общие черты. Установите порядок:

- | | |
|----|--|
| 1. | а) далее необходимо открыть файл данных и выбрать переменные для анализа из открытого файла; |
| 2. | б) после выбора из переключателя модулей открывается стартовая панель выбранного статистического модуля; |
| 3. | в) затем выбираются метод анализа и конкретная вычислительная процедура с соответствующими параметрами расчета из меню в стартовой панели модуля; |
| 4. | г) далее запускается вычислительная процедура. Если процедура итерационная, то система дает возможность на каждом шаге просмотреть результаты в появившемся на экране окне и при необходимости добавить число итераций для увеличения точности оценок; |
| 5. | д) используя графические возможности и специальные таблицы вывода с вычисленными разнообразными статистиками, можно осуществить всесторонний просмотр и анализ результатов; |
| 6. | е) выбирается следующий шаг анализа. |

Ответ 1-Б; 2-А; 3-В; 4-г; 5-Д; 6-Е.

4. Распределите статистические тесты, для проверки распределения на нормальность, соответственно коду Python для каждого теста.

1. Тест Шапиро-Уилка
2. К-квадрат тест
- Д'Агостино
3. Тест нормальности хи-квадрат

а)

```
from scipy.stats import normaltest

data = np.random.normal(loc = 20, scale = 5, size=150)
stat, p = normaltest(data)
print('stat=%.3f, p=%.3f\n' % (stat, p))
if p > 0.05:
    print('Probably Gaussian')
else:
    print('Probably not Gaussian')
```

б)

```
from scipy.stats import shapiro

data = np.random.normal(loc = 20, scale = 5, size=150)
stat, p = shapiro(data)
print('stat=%.3f, p=%.3f\n' % (stat, p))
if p > 0.05:
    print('Probably Gaussian')
else:
    print('Probably not Gaussian')
```

в)

```
from scipy.stats import chisquare

data = np.random.normal(loc = 20, scale = 5, size=150)
statistic, pvalue = chisquare(data)
print('statistic=%.3f, p=%.3f\n' % (statistic, pvalue))
if pvalue > 0.05:
    print('Probably Gaussian')
else:
    print('Probably not Gaussian')
```

Ответ 1-Б; 2-А; 3-В

5. Даны параметры метода Series.plot (соотнесите аргументы с их описанием)

- | | |
|------------|---|
| 1. Label | а) Объект подграфика matplotlib, внутри которого строим график. Если |
| 2. Ax | параметр не задан, то используется активный подграфик |
| 3. Style | б) Метка для пояснительной надписи на графике |
| 4. Kind | в) Строка стиля, например 'ko--', которая передается matplotlib |
| 5. figsize | г) Может принимать значения "area", "bar", "barh", "density", "hist", "kde", "line", "pie"; по умолчанию "line" |
| | д) Размер создаваемого объекта рисунка |

Ответ 1-Б; 2-А; 3-В; 4-Г; 5- Д.

6. Даны аргументы метода resample (соотнесите аргументы с их описанием)

- | | |
|---------------|---|
| 1. rule | а) Строка или объект DateOffset, задающий новую частоту, например 'M', '5min' или Second (15) |
| 2. axis | б) Ось передискретизации, по умолчанию 0 |
| 3. closed | fill_method Способ интерполяции при повышающей передискретизации, например |
| 4. label | "ffill" или "bfill". По умолчанию интерполяция не производится |
| 5. kind | в) При понижающей передискретизации определяет, какой конец интервала |
| 6. convention | должен включаться: "right" (правый) или "left" (левый) |
| | г) При понижающей передискретизации определяет, |

следует ли помечать агрегированный результат меткой правого или левого конца интервала.

Например, пятиминутный интервал от 9:30 до 9:35 можно пометить меткой 9:30 или 9:35. По умолчанию "right" (т. е. 9:35 в этом примере)

д) Агрегировать в периоды ("period") или временные метки ("timestamp"); по умолчанию определяется видом индекса, связанного с данным временным рядом

е) При передискретизации периодов соглашение ("start" или "end") о преобразовании периода низкой частоты в период высокой частоты.

Ответ 1-А; 2-Б; 3-В; 4-Г; 5- Д; 6-Е.

7. Распределите статистические тесты, для проверки распределения на нормальность, соответственно коду Python для каждого теста.

1. resample() –
2. pd.concat –

а) объединяет объекты Pandas вдоль определенной оси с дополнительным набором конфигураций на других осях. Также может добавить слой иерархической индексации на

ось конкатенации, что может быть полезно, если метки одинаковы (или дублируются) на пройденной числовой оси. Аргументы:

– axis : {0, 1, ...}, default 0 =None; ось для объединения;

– dtype; указание типа данных для данных или столбцов.

б) стандартный метод для преобразования и передискретизации регулярных временных рядов данных. Аргументы:

– how: string; нисходящая или повторная передискретизация, по умолчанию 'средняя' для нисходящей.

Ответ 1-Б; 2-А.

8. Распределите важные функции среднего значения и меры центрального расположения.

1. mean(): а) - Эта функция возвращает среднее значение данных, переданных в ее аргументах.
2. mode(): б) - Эта функция возвращает число с максимальным числом вхождений.
3. median(): в) - Эта функция используется для вычисления медианы, т.е. среднего элемента данных.
4. median_low(): г) - Эта функция возвращает медиану данных в случае нечетного числа элементов, но в случае четного числа элементов возвращает меньший из двух средних элементов.
5. median_high(): д) - Эта функция возвращает медиану данных в случае нечетного числа элементов, но в случае четного числа элементов возвращает больший из двух средних элементов.
6. median_grouped() : е) - Эта функция используется для вычисления групповой медианы, т.е. 50-го перцентиля данных.

Ответ 1-А; 2-Б; 3-В; 4-Г; 5- Д.

9. Распределите важные функции статистики меры распространения

- | | |
|------------------|---|
| 1. дисперсия (): | а) - Эта функция вычисляет дисперсию, т.е. меру отклонения данных, чем больше значение дисперсии, тем больше разброс значений данных. В этой функции вычисляется выборочная дисперсия, предполагающая, что данные относятся к части генеральной совокупности. |
| 2. pvariance(): | б) - Эта функция вычисляет дисперсию всей совокупности. Данные интерпретируются как относящиеся ко всей совокупности. |
| 3. stdev(): | 3. - Эта функция возвращает стандартное отклонение (квадратный корень из выборочной дисперсии) данных. |
| 4. pstdev(): | 4. - Эта функция возвращает стандартное отклонение совокупности (квадратный корень из дисперсии совокупности) данных. |

Ответ 1-А; 2-Б; 3-В; 4-Г.

10. Распределите важные функции статистики меры распространения

- | | |
|--|---|
| 1. Функция <code>scipy.stats.zmap</code> | а) (оценки, сравнение, <code>axis=0</code> , <code>ddof=0</code>) вычисляет относительную Z-оценку входных данных. Оценки, стандартизированные до нулевого среднего значения и единичной дисперсии, где среднее значение и дисперсия вычисляются из массива сравнения. |
| 2. <code>stats.invgamma.rvs()</code> | б) метода мы можем сгенерировать случайную величину из инвертированной гамма-обобщенной функции нормального распределения |
| 3. <code>stats.halfgennorm.logcdf()</code> | в) метода мы можем получить логарифмическое значение кумулятивной функции распределения |
| 4. <code>stats.halfgennorm.stats()</code> | г) метода мы можем получить значение обратной функции выживания, которая является обратной ($1 - cdf$) |
| 5. <code>scipy.stats.foldnorm()</code> | д) метода мы можем получить значение среднего ('m'), дисперсии ('v'), перекоса ('s') и / или эксцесса ('k'). |
| | е) представляет собой свернутую нормальную непрерывную случайную величину, которая определяется со стандартным форматом и некоторыми параметрами формы для завершения ее спецификации. |

Ответ 1-А; 2-Б; 3-В; 4-Г.

Критерии и шкала оценивания по оценочному средству тесты

Шкала оценивания (интервал баллов)	Критерий оценивания
5	Тесты выполнены на высоком уровне (правильные ответы даны на 90-100% тестов)
4	Тесты выполнены на среднем уровне (правильные ответы даны на 75-89% тестов)
3	Тесты выполнены на низком уровне (правильные ответы даны на 50-74% тестов)
2	Тесты выполнены на неудовлетворительном уровне (правильные ответы даны менее чем на 50% тестов)

Оценочные средства для промежуточной аттестации

Вопросы для подготовки к зачету

«Информационные технологии обработки статистической информации»:

1. Современные программные пакеты средств статистического анализа.
2. Средства описательной статистики в Apache OpenOffice Calc: возможности и ограничения.
3. Анализ вариационного ряда в Apache OpenOffice Calc.
4. Инструменты описательной статистики в Apache OpenOffice Calc: дисперсия, мода, медиана, стандартная ошибка.
5. Анализ формы распределения в Apache OpenOffice Calc.
6. Уровень надежности в Apache OpenOffice Calc. Интерпретация полученных результатов.
7. Средства дисперсионного анализа в Apache OpenOffice Calc.
8. Оценка возможности проведения дисперсионного анализа: числовая и графическая. построение блочной диаграммы.
9. Оценка результатов проверки гипотез дисперсионного анализа, F-критерий.
10. Проведение двумерного дисперсионного анализа в Apache OpenOffice Calc.
11. Факторный анализ. Снижение размерности данных.
12. Авторегрессионное моделирование методом скользящих средних (ARMA, ARIMA) в Eviews.
13. Проведение авторегрессионного моделирования динамики одномерных данных.
14. Выбор параметров временного лага и сезонных колебаний.
15. Проблема формальных критериев определения параметров модели.
16. Многофакторное моделирование результативной переменной на основе авторегрессионных моделей.

17. Язык Python и особенности его стиля программирования. Интерактивный режим Python. Ipython. Jupyter Notebook.

18. Синтаксис и управляющие конструкции языка Python. Переменные, значения и их типы. Типы данных в Python.

19. Встроенные операции и функции. Основные алгоритмические конструкции.

20. Условный оператор. Множественное ветвление.

21. Циклы и счетчики.

22. Определение функций. Параметры и аргументы. Вызовы функций. Оператор возврата. Конструкции *args, **kwargs.

23. Списки, кортежи и словари.

24. Операторы общие для всех типов последовательностей.

25. Специальные операторы и функции для работы со списками. Срезы.

26. Работа со словарями. Методы словарей.

27. Случайные числа. random, randrange, choice.

28. Функции обработки строк. join, replace, split.

29. Стандартная библиотека и pip. Модули и пакеты в Python. Основные стандартные модули.

30. Импортное модулей. Создание собственных модулей и их импортное. Специализированные модули и приложения.

Шкала оценивания (экзамен)	Характеристика знания предмета и ответов
отлично (5)	Студент глубоко и в полном объёме владеет программным материалом. Грамотно, исчерпывающе и логично его излагает в устной или письменной форме. При этом знает рекомендованную литературу, проявляет творческий подход в ответах на вопросы и правильно обосновывает принятые решения, хорошо владеет умениями и навыками при выполнении практических задач.
хорошо (4)	Студент знает программный материал, грамотно и по сути излагает его в устной или письменной форме, допуская незначительные неточности в утверждениях, трактовках, определениях и категориях или незначительное количество ошибок. При этом владеет необходимыми умениями и навыками при выполнении практических задач.
удовлетворительно (3)	Студент знает только основной программный материал, допускает неточности, недостаточно чёткие формулировки, непоследовательность в ответах, излагаемых в устной или письменной форме. При этом недостаточно владеет умениями и навыками при выполнении практических задач. Допускает до 30% ошибок в излагаемых ответах.
неудовлетворительно (2)	Студент не знает значительной части программного материала. При этом допускает принципиальные ошибки в доказательствах, в трактовке понятий и категорий, проявляет низкую культуру знаний, не владеет основными умениями и навыками при выполнении практических задач. Студент отказывается от ответов на дополнительные вопросы.

Лист изменений и дополнений

№ п/п	Виды дополнений и изменений	Дата и номер протокола заседания кафедры (кафедр), на котором были рассмотрены и одобрены изменения и дополнения	Подпись (с расшифровкой) заведующего кафедрой (заведующих кафедрами)