

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.
(подпись)
« 24 » 2023 года



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
«СОВРЕМЕННЫЕ ТЕХНОЛОГИИ ОБРАБОТКИ СТАТИСТИЧЕСКИХ
ДАНЫХ»

По направлению подготовки 38.03.05 Бизнес-информатика
Профиль: «Экономическая аналитика и бизнес-статистика»

Луганск – 2023

Лист согласования РПУД

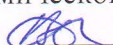
Рабочая программа учебной дисциплины «Современные технологии обработки статистических данных» по направлению подготовки 38.03.05 Бизнес-информатика. – 41 с.

Рабочая программа учебной дисциплины «Современные технологии обработки статистических данных» составлена с учетом Федерального государственного образовательного стандарта высшего образования по направлению подготовки 38.03.05 Бизнес-информатика, утвержденного приказом Министерства науки и высшего образования Российской Федерации от 29 июня 2020 года № 838.

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 №

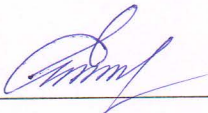
Согласована (для обеспечивающей кафедры):

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

Цель изучения дисциплины – приобретение студентами необходимой квалификации для проведения статистического анализа различных экономических процессов и явлений с использованием программных средств статистического анализа.

Задачи:

изучение методов статистики (в том числе и многомерных), современных технологий при выполнении конкретного экономико-статистического анализа данных на ПЭВМ;

овладение современными аналитическими технологиями для получения возможности выполнения многовариантных расчетов и формирования более обоснованных выводов и рекомендаций по выбору форм связей и зависимостей;

приобретение навыков использования современных вычислительных инструментов в процессе обработки реальных числовых массивов на ПЭВМ при выполнении конкретного экономико-статистического анализа данных, в том числе для ускорения выполнения аналитической части курсовых работ, при подготовке выпускных квалификационных работ.

2. Место дисциплины в структуре ОПОП ВО. Требования к результатам освоения содержания дисциплины

Дисциплина «Современные технологии обработки статистических данных» относится к части, формируемой участниками образовательных отношений.

Необходимыми условиями для освоения дисциплины являются: знания основных макроэкономических показателей, характеризующих развитие государства, теории вероятностей и математической статистики, умения осуществлять сбор экономических данных и расчеты экономических показателей, использовать прикладное программное обеспечение для обработки статистических данных.

Содержание дисциплины является логическим продолжением содержания дисциплин «Бизнес-информатика», «Бизнес-информатика 2», «Статистика», «Экономическая статистика», «Многомерные статистические методы», «Анализ данных средствами языка программирования R» и служит основой для прохождения преддипломной практики, и написания выпускной квалификационной работы.

3. Требования к результатам освоения содержания дисциплины

Код и наименование компетенции	Индикаторы достижений компетенции (по реализуемой дисциплине)	Перечень планируемых результатов
ПК-2. Способен осуществлять поддержку принятия управленческих	ПК-2.2. Способен осуществлять экономические расчеты с применением информационных	Знать: современные технические средства и информационные технологии, используемые в экономических

решений	технологий	расчетах.
		Уметь: осуществлять экономические расчеты, в том числе с использованием программных продуктов, формировать и оформлять аналитические отчеты – анализировать и прогнозировать хозяйственную деятельность организации и интерпретировать полученные результаты анализа.
		Владеть: методами инструментальных средств для обработки экономических данных.

4. Структура и содержание дисциплины

4.1. Объем учебной дисциплины и виды учебной работы

Вид учебной работы	Объем часов (зач. ед.)		
	Очная форма	Очно-заочная форма	Заочная форма
Общая учебная нагрузка (всего)	216 (6 зач. ед)	216 (6 зач. ед)	216 (6 зач. ед)
Обязательная контактная работа (всего) в том числе:	104	60	20
Лекции	52	30	10
Семинарские занятия	-	-	-
Практические занятия	52	30	10
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)	112	156	196
Форма аттестации (7 / 8 семестр)	зачет с оценкой / зачет	зачет с оценкой / зачет	зачет с оценкой / зачет

4.2. Содержание разделов дисциплины

Семестр 7

Тема 1. ВВЕДЕНИЕ В ДИСЦИПЛИНУ. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ

Цель и задачи дисциплины. Современные программные пакеты средств статистического анализа. Общие сведения о настройке «Пакет анализа» и статистические функции OpenOffice CALC.

Тема 2. OPENOFFICE CALC: ОПРЕДЕЛЕНИЕ ХАРАКТЕРА РАСПРЕДЕЛЕНИЯ И ФОРМИРОВАНИЕ ВЫБОРКИ. ОПИСАТЕЛЬНАЯ СТАТИСТИКА

Теоретические основы группировки и формирования выборки. Технология работы при построении гистограммы. Технология работы при формировании выборки.

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. OPENOFFICE CALC: ПРИНЯТИЕ СТАТИСТИЧЕСКИХ РЕШЕНИЙ

Проверка гипотез. Доверительные интервалы. Ковариация и корреляция. Проверка гипотезы о виде закона распределения. Исключение резко выделяющихся значений. Анализ одной выборки. Анализ двух выборок. Использование инструмента «Пакет анализа» для выявления различий между выборками.

Тема 4. OPENOFFICE CALC: КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

Средства корреляционно-регрессионного анализа в OpenOffice CALC. Проведение предварительного этапа, анализ исходной информации, оценка возможности проведения анализа. Построение корреляционной матрицы. Нахождение параметров уравнения регрессии и их оценка. Подбор и проверка значимости уравнения регрессии. Двумерный корреляционно-регрессионный анализ.

Поиск закономерностей для качественных данных. Расчет критерия Хи-квадрат в OpenOffice CALC

Тема 5. 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.

Семестр 8

Тема 6. R: МЕТОДЫ СНИЖЕНИЯ РАЗМЕРНОСТИ, МЕТОД ГЛАВНЫХ КОМПОНЕНТ, ФАКТОРНЫЙ АНАЛИЗ

Факторный анализ. Снижение размерности данных. Методы снижения размерности и факторного анализа в R. Объединение и обобщение.

Выделение главных компонент. Анализ главных компонент. Обзор результатов анализа. Факторный анализ как метод классификации. Подтверждающий факторный анализ (моделирование структурными соответствиями). Анализ соответствий (изучение структуры качественных переменных). Многомерный дисперсионный анализ.

Тема 7. R: АНАЛИЗ “DATA MINING”

Понятие «добычи данных». Основные направления метода добычи данных и проведение анализа средствами R. Хранилища данных. Оперативная аналитическая обработка (OLAP). Разведочный анализ данных (РАД) и методы добычи данных. Проверка гипотез в РАД. Вычислительные методы РАД. Графические методы. Проверка результатов РАД. «Нейронные сети».

Тема 8. R: МНОГОМЕРНАЯ КЛАССИФИКАЦИЯ, КЛАСТЕРНЫЙ АНАЛИЗ, ДИСКРИМИНАНТНЫЙ АНАЛИЗ

Основное назначение кластерного анализа. Возможности кластерного и дискриминантного анализа в R. Древоподобная классификация (объединение). Меры расстояния. Правила объединения. Метод К средних: общая логика и постановка задачи. Примеры анализа и интерпретации результатов. Дискриминантный анализ: основная цель и вычислительный подход. Пошаговый дискриминантный анализ. Дискриминация для двух групп. Дискриминация для нескольких групп. Проведение классификации дискриминантным методом.

Тема 9. 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.

4.3. Лекции

№ п/п	Название темы	Объем часов		
		Очная форма	Очно-заочная форма	Заочная форма
1	Введение в дисциплину. информационные технологии статистической обработки	4	2	1
2	OPENOFFICE CALC: Определение характера распределения и формирование выборки. Описательная статистика	6	4	2

3	OPENOFFICE CALC: Принятие статистических решений	6	2	1
4	OPENOFFICE CALC: Корреляционно-регрессионный анализ	6	4	1
5	OPENOFFICE CALC: Дисперсионный анализ, одномерный и двумерный	6	4	1
6	R: Методы снижения размерности, метод главных компонент, факторный анализ	6	4	1
7	R: Анализ “Data mining”	6	4	1
8	R: Многомерная классификация, кластерный анализ, дискриминантный анализ	6	4	1
9	EVIEWES: Авторегрессионное моделирование. анализ временных рядов и прогнозирование	6	4	1
Итого:		52	30	10

4.4. Практические (семинарские) занятия

№ п/п	Название темы	Объем часов		
		Очная форма	Очно-заочная форма	Заочная форма
1	Введение в дисциплину. информационные технологии статистической обработки	4	2	1
2	OPENOFFICE CALC: Определение характера распределения и формирование выборки. Описательная статистика	6	4	2
3	OPENOFFICE CALC: Принятие статистических решений	6	2	1
4	OPENOFFICE CALC: Корреляционно-регрессионный анализ	6	4	1
5	OPENOFFICE CALC: Дисперсионный анализ, одномерный и двумерный	6	4	1
6	R: Методы снижения размерности, метод главных компонент, факторный анализ	6	4	1
7	R: Анализ “Data mining”	6	4	1
8	R: Многомерная классификация, кластерный анализ, дискриминантный анализ	6	4	1
9	EVIEWES: Авторегрессионное моделирование. анализ временных рядов и прогнозирование	6	4	1
Итого:		52	30	10

4.5. Лабораторные работы

Учебным планом не предусмотрены.

4.6. Самостоятельная работа студентов

№ п/п	Название темы	Вид СРС	Объем часов		
			Очная форма	Очно-заочная форма	Заочная форма
1	Введение в дисциплину. информационные технологии статистической обработки	Подготовка к практическому занятию и написание отчета	8	14	18
2	OPENOFFICE CALC: Определение характера распределения и формирование выборки. Описательная статистика	Подготовка к практическому занятию и написание отчета	10	16	18
3	OPENOFFICE CALC: Принятие статистических решений	Подготовка к практическому занятию и написание отчета	10	14	16
4	OPENOFFICE CALC: Корреляционно-регрессионный анализ	Подготовка к практическому занятию и написание отчета	10	16	20
5	OPENOFFICE CALC: Дисперсионный анализ, одномерный и двумерный	Подготовка к практическому занятию и написание отчета	10	16	20
6	R: Методы снижения размерности, метод главных компонент, факторный анализ	Подготовка к практическому занятию и написание отчета	14	18	20
7	R: Анализ "Data mining"	Подготовка к практическому занятию и написание отчета.	14	18	20
8	R: Многомерная классификация, кластерный анализ, дискриминантный анализ	Подготовка к практическому занятию и написание отчета	14	18	28
9	EViews: Авторегрессионное моделирование. анализ временных рядов и прогнозирование	Подготовка к практическому занятию и написание отчета	14	18	28
	Контроль	Подготовка к зачету	8	8	8
			112	156	196

4.7. Курсовые работы/проекты.

Учебным планом не предусмотрено.

5. Образовательные технологии

Преподавание дисциплины ведется с применением следующих видов образовательных технологий:

Преподавание дисциплины ведется с применением следующих видов образовательных технологий:

Информационные технологии: использование электронных образовательных ресурсов (электронный конспект, размещенный на сайте кафедры ecpsdahl.ru) при подготовке к лекциям и лабораторным работам.

Мультимедийные технологии: презентации к лекциям.

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:

собеседование (устный или письменный опрос);

тестирование;

индивидуальные задания.

Фонды оценочных средств, включающие вопросы для собеседования, индивидуальные задания, тесты и методы контроля, позволяющие оценить результаты текущей и промежуточной аттестации обучающихся по данной дисциплине, помещаются в приложении к рабочей программе в соответствии с «Положением о фонде оценочных средств».

Форма аттестации по результатам освоения дисциплины проходит в форме устного экзамена (включает в себя ответ на теоретические вопросы и решение задач) либо в сочетании различных форм (компьютерного тестирования, решения задач и пр.). Студенты, выполнившие 75% текущих и контрольных мероприятий на «отлично», а остальные 25 % на «хорошо», имеют право на получение итоговой отличной оценки.

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

Шкала оценивания (экзамен)	Характеристика знания предмета и ответов	Зачеты
отлично (5)	Студент глубоко и в полном объеме 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:

а) основная литература:

1. Афонин П.Н., Статистический анализ с применением современных программных средств: учебное пособие / Афонин П.Н., Афонин Д.Н. - СПб.: ИЦ Интермедия, 2017. - 100 с. - ISBN 978-4383-0080-9 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN978438300809.html>

2. Баранникова И.В., Теоретические основы автоматизированной обработки информации и управления: специальные функции OpenOffice CALC / 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

3. 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>

4. 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>

5. 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>

6. 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:

7. 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>

8. 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>

9. 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

10. Сдвижков О.А., Непараметрическая статистика в OpenOffice CALC 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>

11. 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>

12. 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>

13. 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 «Современные технологии обработки статистических данных» для студентов направления подготовки 38.03.05 – Бизнес-информатика [Электронный ресурс] / сост. Н.А. Рязанцева. – Луганск: ЛНУ им. В. Даля, 2019. – 78 с.

2. Методические указания к самостоятельной работе по дисциплине «Современные технологии обработки статистических данных» для студентов

направления подготовки 38.03.05 – Бизнес-информатика [Электронный ресурс] / сост. Н.А. Рязанцева. – Луганск: ЛНУ им. В. Даля, 2019. – 25 с.

г) Интернет-ресурсы:

1. Росстат [Электронный ресурс]. – Режим доступа:
2. Федеральная служба государственной статистики: Главная www.gks.ru
3. Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации [Электронный ресурс]. - Режим доступа: <https://digital.gov.ru/ru/>
4. Сервер раскрытия информации <https://www.e-disclosure.ru/>

Электронные библиотечные системы и ресурсы

1. Электронно-библиотечная система «Консультант студента» – <http://www.studentlibrary.ru/>
2. Электронно-библиотечная система «StudMed.ru» – <https://www.studmed.ru/>
3. Руководство по своду знаний по бизнес-анализу (BABOK 2.0.) (на рус.яз.). URL: <http://iiba.ru/chapter-1-introduction/>
4. Бизнес-информатика. Концепция ведения бизнеса нового поколения URL: http://donntu.org/sites/default/files/documents/prezentaciya_napravleniya_podgotovki.pdf

5. Системы дистанционного обучения кафедры экономической кибернетики и прикладной статистики Луганского национального университета имени Владимира Даля в среде Moodle. URL: <https://ecps.gnomio.com/>

6. Методическая работа в вузе: методические указания / сост . Н. П. Пучков. – Тамбов: Изд-во ГОУ ВПО ТГТУ, 2010. – 32 с. URL: <http://window.edu.ru/resource/184/73184/files/puchkov.pdf>

Информационный ресурс библиотеки образовательной организации

7. Научная библиотека имени А. Н. Коняева – <http://biblio.dahluniver.ru/>

8. Материально-техническое и программное обеспечение дисциплины

Для проведения лекционных занятий требуется аудитория на курс, оборудованная мультимедийным проектором с экраном.

Для проведения лабораторных занятий требуется компьютерный класс, подключенный к Интернет.

Программное обеспечение:

Функциональное назначение	Бесплатное программное обеспечение	Ссылки
Офисный пакет	OpenOffice 4.3.7	https://www.openoffice.org/
Операционная система	UBUNTU 19.04	https://ubuntu.com/
Браузер	Firefox Mozilla	http://www.mozilla.org/ru/firefox/fx
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
Редактор PDF	Adobe Acrobat Reader	https://get.adobe.com/ru/reader/
Аудиоплеер	VLC	http://www.videolan.org/vlc/
Язык программирования	Python	https://www.python.org/downloads/
Среда разработки	PyCharm Community	https://www.jetbrains.com/ru-ru/pycharm/download/

9. Оценочные средства по дисциплине

Паспорт фонда оценочных средств по учебной дисциплине " Современные технологии обработки статистических данных "

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	ПК-2	Способен осуществлять поддержку принятия управленческих решений	ПК-2.2. Способен осуществлять экономические расчеты с применением информационных технологий	Тема 1-12	7, 8

**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.	ПК-2	ПК-2.2	<p>Знать: современные технические средства и информационные технологии, используемые в экономических расчетах.</p> <p>Уметь: осуществлять экономические расчеты, в том числе с использованием программных продуктов, формировать и оформлять аналитические отчеты – анализировать и прогнозировать хозяйственную деятельность организации и интерпретировать полученные результаты анализа.</p> <p>Владеть: методами инструментальных средств для обработки экономических данных.</p>	Тема 1-12	Устный или письменный опрос, индивидуальные задания, тесты

**Фонды оценочных средств по дисциплине
«Современные технологии обработки статистических данных»
Вопросы для обсуждения на практических и семинарских занятиях
(устный или письменный опрос)**

1. Перечислите возможные практические задачи, которые целесообразно решить программными методами.
2. Какие показатели являются результатом вывода описательной статистики в OpenOffice CALC?
3. Какие основные ограничения имеются в 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 OpenOffice CALC
8. Как интерпретируются результатные показатели дисперсионного анализа в 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. Какие характеристики множественной регрессионной модели можно получить в R? Какие характеристики множественной регрессионной модели можно получить в R?
17. Назовите необходимые условия и этапы предварительного анализа для проведения множественного регрессионного моделирования.
18. Как проводить подгонку модели по результатам анализа в R?
19. Какие могут быть возможные затруднения? Как верно оценить значимость модели?
20. Какие средства оценки модели есть в наличии в R?
21. Сформулируйте сущность и необходимую информационную базу для применения метода «добычи данных».
22. Что представляют собой хранилища данных и какова их структура?
23. Что такое оперативная обработка данных (OLAP) и как она реализуется?
24. Что представляет собой разведочный анализ данных (РАД)?

25. Какими способами графического анализа данных располагает R? Как они применяются?
26. Что такое нейронные сети, какие результаты и интерпретацию они позволяют получить?
27. Сформулируйте основные направления и методы кластерного анализа.
28. Что представляет собой формальная постановка задачи и какие данные требуются для иерархической кластеризации?
29. Назовите необходимые исходные данные и постановку задачи для метода K средних.
30. Как проводится дискриминантный анализ в R?
31. Опишите в общих чертах сущность и задачи авторегрессионного моделирования, назовите условия его применения.
32. Как проводится выбор количества лаговых переменных? Опишите визуальный способ и назовите количественные критерии, сформулируйте возможные различия между ними.
33. Как проводится выбор количества сезонных переменных??
34. Назовите необходимые условия и этапы предварительного анализа для проведения множественного регрессионного моделирования.
35. Как проводить подгонку модели по результатам анализа в R?
36. Какие могут быть возможные затруднения? Как верно оценить значимость модели?
37. Какие средства оценки модели есть в наличии в R?
38. Сформулируйте сущность и необходимую информационную базу для применения метода «добычи данных».
39. Что представляют собой хранилища данных и какова их структура?
40. Что такое оперативная обработка данных (OLAP) и как она реализуется?
41. Что представляет собой разведочный анализ данных (РАД)?
42. Какими способами графического анализа данных располагает R? Как они применяются?
43. Что такое нейронные сети, какие результаты и интерпретацию они позволяют получить?
44. Сформулируйте основные направления и методы кластерного анализа.
45. Что представляет собой формальная постановка задачи и какие данные требуются для иерархической кластеризации?
46. Назовите необходимые исходные данные и постановку задачи для метода K средних.
47. Как проводится дискриминантный анализ в R?
48. Опишите в общих чертах сущность и задачи авторегрессионного моделирования, назовите условия его применения.
49. Как проводится выбор количества лаговых переменных? Опишите визуальный способ и назовите количественные критерии, сформулируйте возможные различия между ними.

50. Как проводится выбор количества сезонных переменных?

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

52. В чем заключается аналитическая работа экономистов, работающих в управлениях Госкомстата.

53. Укажите состав и название основных отделов, входящих в группу подразделений ВЦ, занимающихся проектированием машинной обработки статистической информации.

Критерии и шкала оценивания по оценочному средству собеседование (устный или письменный опрос)

Шкала оценивания (интервал баллов)	Критерий оценивания
5	собеседование (устный или письменный опрос) прошел на высоком уровне (студент в полном объеме осветил рассматриваемый вопрос, владеет профильным понятийным (категориальным) аппаратом и т.п.)
4	собеседование (устный или письменный опрос) прошел на среднем уровне (студент в целом осветил рассматриваемый вопрос, владеет профильным понятийным (категориальным) аппаратом и т.п.)
3	собеседование (устный или письменный опрос) на низком уровне (студент допустил существенные неточности, изложил материал с ошибками, не владеет в достаточной степени профильным категориальным аппаратом и т.п.)
2	собеседование (устный или письменный опрос) прошел 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 OpenOffice CALC. Рекомендованное время работы над документом – 2 часа 30 минут. Данные размещены в СДО кафедры ЭКИПС по адресу <https://ecpsdahl.ru/>

Таблица содержит данные по 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. При проведении статистического наблюдения за деятельностью предприятий корпорации получены выборочные данные по 32-м предприятиям, выпускающим однородную продукцию (выборка 10%-ная, механическая), о среднегодовой стоимости основных производственных фондов и о выпуске продукции за год.

Статистический анализ выборочной совокупности

1. Выявить наличие среди исходных данных резко выделяющихся значений признаков («выбросов» данных) с целью исключения из выборки аномальных единиц наблюдения.

2. Рассчитать обобщающие статистические показатели совокупности по изучаемым признакам: среднюю арифметическую, моду, медиану, размах вариации, дисперсию, средние отклонения – линейное (d) и квадратическое (σ), коэффициент вариации ($V\sigma$), структурный коэффициент асимметрии К.Пирсона ($A_{сп}$).

3. На основе рассчитанных показателей в предположении, что распределения единиц по обоим признакам близки к нормальному, оценить:

- а) степень колеблемости значений признаков в совокупности;
- б) степень однородности совокупности по изучаемым признакам;
- в) устойчивость индивидуальных значений признаков;
- г) количество попаданий индивидуальных значений признаков в диапазоны $(\bar{x} \pm \sigma)$, $(\bar{x} \pm 2\sigma)$, $(\bar{x} \pm 3\sigma)$.

4. Дать сравнительную характеристику распределений единиц совокупности по двум изучаемым признакам на основе анализа:

- а) вариации признаков;
- б) количественной однородности единиц;
- в) надежности (типичности) средних значений признаков;
- г) симметричности распределений в центральной части ряда.

5. Построить интервальный вариационный ряд и гистограмму распределения единиц совокупности по признаку Среднегодовая стоимость основных производственных фондов и установить характер (тип) этого распределения. Рассчитать моду M_o полученного интервального ряда и сравнить ее с показателем M_o несгруппированного ряда данных.

Задание 6. Дисперсионный анализ. Необходимо проанализировать производительность труда рабочих одного из цехов.

С этой целью была образована контрольная группа рабочих, имеющих стаж работы до 1 года, во вторую группу вошли рабочие со стажем от 1 до 3 лет и в третью группу объединили рабочих с большим стажем. Для чистоты эксперимента были сформированы пять групп, приблизительно равноценных по показателям. Результаты приведены в таблице исходных данных (<https://ecpsdahl.ru/>). Рассчитать эмпирическое корреляционное отношение и коэффициент детерминации. Сформулировать выводы. Учитывая специфику исходных данных, провести двухфакторный дисперсионный анализ с повторениями или без повторений в той же последовательности.

Задание 7. С использованием средств Вставка функций и Мастер диаграмм OpenOffice CALC проведите анализ по схеме Примера (<https://ecpsdahl.ru/>) следующих данных значений прибыли, млн. руб., для 75 предприятий:

24,8 33,4 29,3 32,7 30,4 25,9 30,0 30,2 26,7 32,3 29,6 34,0 30,7 29,4 32,0 30,7 28,0 31,5 29,6 33,5 28,0 26,4 38,4 30,5 28,0 29,9 34,2 36,0 30,7 28,0 28,0 28,6 26,1 30,7 33,1 37,3 32,3 30,0 34,1 25,5 25,5 31,5 31,8 36,4 31,8 25,5 33,1 25,5 28,2 31,6 33,8 31,9 32,2 28,0 26,4 30,2 31,2 33,4 32,2 34,4 30,5 32,3 30,8 30,0 37,0 24,2 33,5 35,3 33,7 30,6 31,8 35,4 29,5 32,8 27,4.

Проанализируйте влияние числа интервалов группирования на гистограмму для трех значений k : минимального, в полтора раза и вдвое больше минимального.

Проведите аппроксимацию и сглаживание построенных гистограмм. Сравнение полиномиальных трендов для трех различных значений числа интервалов группирования проведите для степени аппроксимирующего полинома 2.

Выполните процедуру генерации случайных чисел и визуализации данных с использованием средств Анализ данных и Мастер диаграмм OpenOffice CALC.

При помощи средства Анализ данных выполните процедуру описательной статистики по сгенерированным данным.

При помощи средства Анализ данных выполните процедуру построения гистограмм по сгенерированным в Задании 3 данным.

Задание 8. Построить линейную модель для двух наблюдаемых величин (например, объем реализованных подержанных автомобилей фирмой за указанное число недель) в соответствии с номером варианта (<https://ecpsdahl.ru/>).

Задание 9. Определить, влияет ли фактор образования на уровень зарплаты сотрудников в гостинице на основании следующих данных:

Образование	Зарплата сотрудника					
	Высшее	3200	3000	2600	2000	1900
Среднее спец.	2600	2000	2000	1900	1800	1700
Среднее	2000	2000	1900	1800	1700	1700

Задание 10. Приведены данные ежемесячной результативности (количество голов) футбольной команды в двух сезонах

Месяц	3	4	5	6	7	8	9	10	11
2020 г.	3	4	5	8	9	1		4	5
2021 г.	6	19	3	2	14	4	5	117	1

Определить, есть ли статистические различия в ежемесячной результативности команды в рассматриваемых сезонах?

Задание 11. Даны результаты бега на дистанцию 100 м в секундах в двух группах студентов. Студенты первой группы в течение года посещали факультативные занятия по физкультуре. Определить, достоверны ли отличия по результатам бега в этих группах.

Посещавшие факультатив	Не посещавшие
12,6	12,8

12,3	13,2
11,9	13,0
12,2	12,9
13,0	13,5
12,4	13,1

Задание 12. Исходные данные (Росстат) представляют собой удельные показатели ресурсов здравоохранения и социальных ресурсов для субъектов федерации РФ. Цель работы – исследовать взаимосвязь между заболеваемостью, инвалидностью и ресурсами здравоохранения, социальными ресурсами. Исследовать влияние на заболеваемость и инвалидность ресурсов здравоохранения. Найти наиболее значимые ресурсы здравоохранения и социальные ресурсы. Сформировать группы районов, обладающих близкими параметрами по различным показателям ресурсов здравоохранения и социальных ресурсов и по сумме всех показателей.

Задание 13. 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 14. Исходные данные (<https://ecpsdahl.ru/>) представляют собой поквартальную динамику компонент расходов в реальном ВВП Великобритании, 1955-2013 гг.). Все расчеты сделать в Eviews.

1. Построить описательную статистику и графики компонентов ВВП, гистограмму ошибок основного макроэкономического тождества.

2. Провести тест на причинно-следственную связь между какой-либо парой переменных (тест Гранжера).

3. Составить какое-либо макроэкономическое уравнение для одной из компонентов расходов ВВП, провести его оценивание методом наименьших квадратов и провести:

а) тестирование статистической гипотезы о структурной изменчивости коэффициента регрессии при какой-либо независимой переменной (тест Чоу)

б) тестирование на наличие автокорреляции остатков в оцененной регрессии (тест Дарбина-Уотсона, LM-test).

4. Сделать безусловный поквартальный прогноз объясняемой переменной на 2002 г. в уравнении регрессии п.3.

5. Составить какую-либо систему одновременных макроэкономических уравнений, выбрать инструментальные переменные и провести оценивание методами TSLS (2SLS) и 3SLS.

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%)

Тесты

Задания закрытого типа

1. Комплекс программных средств, обеспечивающий ввод статистических данных, их обработку и представление результатов обработки в наглядном виде называют:

- информационно-вычислительным комплексом;
- программным комплексом;
- программным статистическим комплексом;
- статистическим комплексом.

2. Программные статистические комплексы применяются:
- а) только в технике;
 - б) только в экономике;
 - в) только в организации производства;
 - г) в различных сферах деятельности.
3. Отечественным программным статистическим комплексом является:
- а) MS Excel;
 - б) Mathematica;
 - в) Mathcad;
 - г) SPSS.
4. К зарубежным программным статистическим комплексам не относится:
- а) MS Excel;
 - б) Mathematica;
 - в) Mathcad;
 - г) SPSS.
5. К программным статистическим комплексам не относится:
- а) MS Excel;
 - б) Mathematica;
 - в) Mathcad;
 - г) Power point.
6. К табличным программным статистическим комплексам относится:
- а) MS Excel;
 - б) Mathematica;
 - в) Mathcad;
 - г) Statistica.
7. Наиболее распространенным и широко используемым программным статистическим комплексом является:
- а) MS Excel;
 - б) Mathematica;
 - в) Mathcad;
 - г) R.
8. Структура программных статистических комплексов в общем случае включает:
- а) две составляющих;
 - б) три составляющих;
 - в) четыре составляющих;
 - г) пять составляющих.
9. Алгоритмическое (теоретическое) обеспечение программных статистических комплексов основывается:
- а) на логике и теории чисел;
 - б) теории множеств;
 - в) математической физике;

d) теории вероятностей и математической статистике.

10. В состав Microsoft Office входит программный статистический комплекс:

- a) MS Excel;
- b) Mathematica;
- c) Mathcad;
- d) Statistica.

11. Для подключения к программе OpenOffice CALC пакета статистического анализа данных пользователю требуется выполнить следующее действие ...

a) произвести установку модуля «пакет анализа» из режима «надстройки» программы;

б) удалить программу OpenOffice CALC и произвести ее повторную стандартную установку;

в) перезагрузить компьютер.

12. В OpenOffice CALC реализована схема _____ отбора.

- a) случайного повторного;
- б) случайного безповторного;
- в) серийного;
- г) комбинированного.

13. Какие методы анализа в модуле «Анализ данных» доступны:

- a) Anova;
- б) T-тест;
- в) корреляция;
- г) регрессия;
- д) описательная статистика.

14. Для вычисления среднего арифметического значения применяется функция ...

- a) МЕДИАНА(...)
- б) СРГАРМ(...)
- в) СРЗНАЧ(...)
- г) СРГЕОМ(...)

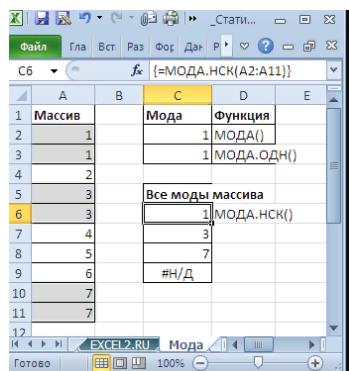
15. Для подключения к программе OpenOffice CALC пакета статистического анализа данных пользователю требуется выполнить следующее действие ...

a) произвести установку модуля «Пакет анализа» из режима «Надстройки» программы;

б) удалить программу OpenOffice CALC и произвести ее повторную стандартную установку;

в) перезагрузить компьютер.

16. На рисунке показано вычисление моды, какая команда позволяет выводит все значения моды?



- а) МОДА();
- б) МОДА.ОДН();
- в) МОДА.НСК();
- г) МОДА. ЛСК().

17. В число инструментов Пакета анализа MS Excel не входит:

- а) описательная статистика;
- б) двухфакторный дисперсионный анализ;
- с) регрессия;
- д) дисперсия.

18. Для создания таблиц основных статистических характеристик для одного или нескольких множеств входных значений используется инструмент Пакета анализа MS Excel:

- а) описательная статистика;
- б) двухфакторный дисперсионный анализ;
- с) регрессия;
- д) дисперсия.

19. Изучение влияния одного или нескольких факторов на рассматриваемый признак является задачей:

- а) дисперсионного анализа;
- б) корреляционного анализа;
- с) регрессионного анализа;
- д) математического анализа.

20. Получить стандартное отклонение и оценку дисперсии выборки позволяет инструмент Пакета анализа MS Excel:

- а) дисперсия;;
- б) двухфакторный дисперсионный анализ;
- с) регрессия;
- д) описательная статистика.

21. При проведении дисперсионного анализа с использованием Пакета анализа MS Excel вероятность возникновения ошибки по умолчанию принимается равной:

- а) 0,01;
- б) 0,03;
- с) 0,05;
- д) 0,10.

22. Определить, носят ли расхождения между группами данных случайный характер или вызваны конкретными обстоятельствами, позволяет проведение:

- а) дисперсионного анализа;

- b) корреляционного анализа;
- c) регрессионного анализа;
- d) математического анализа.

23. Получить математическую модель объекта исследования позволяет проведение:

- a) дисперсионного анализа;
- b) корреляционного анализа;
- c) регрессионного анализа;
- d) математического анализа.

24. Для установления влияния одной величины на другую по экспериментальным данным используется инструмент:

- a) описательная статистика;
- b) двухфакторный дисперсионный анализ;
- c) регрессия;
- d) дисперсия.

25. Определить коэффициенты математической модели объекта исследования позволяет применение инструмента:

- a) описательная статистика;
- b) двухфакторный дисперсионный анализ;
- c) дисперсия;
- d) регрессия.

26. Проверка размерности переменных проводится в статистическом комплексе:

- a) Mathcad;
- b) MS Excel;
- c) R;
- d) Mathematica.

27. Математические формулы в обычном их виде записываются при использовании программного статистического комплекса:

- a) Mathematica;
- b) MS Excel;
- c) R;
- d) Mathcad.

28. Динамическая обработка данных в числовом и аналитическом (формульном) виде обеспечивается при использовании программного статистического комплекса:

- a) Mathematica;
- b) MS Excel;
- c) Mathcad;
- d) R.

29. В полном факторном эксперименте минимальное число опытов для k факторов равняется:

- a) $2k$;

- b) $k/2$;
- c) 2^k ;
- d) 2^{k-1} .

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

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

31. 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.

32. 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».

33. В состав описательной статистики не входит такая характеристика:

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

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

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

35. Целью факторного анализа методом главных компонент может являться:

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

36. Для воспроизведения нетривиальных сложных зависимостей используется:

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

37. 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

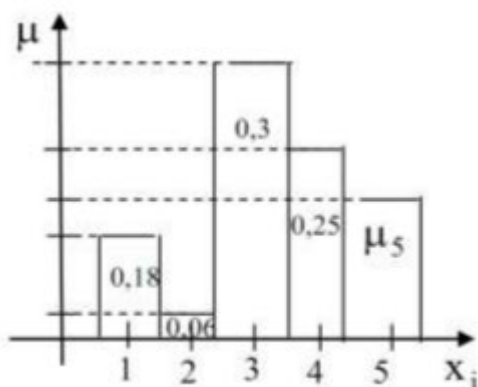
38. Дискретные случайные величины среди приведенных величин

- 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»

39. Мода в математической статистике:

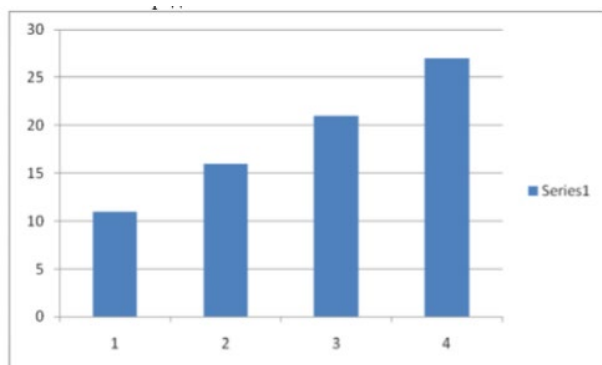
- а) значение признака (варианты), которое чаще всего встречается в данной совокупности
- б) значение признака (варианты), которое находится в середине ранжированного ряда
- в) среднее арифметическое значение признака (варианты)

40. Выборка задана в виде гистограммы относительных частот. Тогда относительная частота μ_5 равна



- а) 0,21
- б) 0,23
- в) 0,22
- г) 0,2

41. Гистограмма частот построена для выборки, представленной в виде группированного статистического ряда



а)

Границы интервалов	10-15	15-20	20-25	25-30
Частота	4	7	3	1

б)

Границы интервалов	0-12,5	12,5-17,5	17,5-22,5	22,5-27,5
Частота	20	35	15	5

в)

Границы интервалов	0-11,5	11,5-16,5	16,5-21,5	21,5-27
Частота	20	35	15	5

г)

Номер интервала	1	2	3	4
Частота	11	16	21	27

42. Для вычисления стандартного отклонения в R предусмотрена функция

- а) `sd()`;
- б) `mean()`;
- в) `cor()`;
- г) `hist()`.

43. В R реализовано несколько разных техник, отвечающих на вопрос о том, что данные имеют нормальное распределение:

- а) тест Шапиро-Уилкса;
- б) тест Колмогорова-Смирнова;
- в) графический способ
- г) все ответы верные.

44. Интересный способ представления корреляционной матрицы предоставляется пакетом `ellipse`. В этом случае значения коэффициентов корреляции рисуются в виде эллипсов. Чем ближе значение коэффициента корреляции к +1 или 1 — тем более вытянутым становится эллипс. Наклон эллипса отражает знак. Для получения изображения необходимо вызвать функцию:

- а) `plotcorr`;
- б) `symnum()`;
- в) `cor()`;
- г) `chisq.test()`.

45. Для того чтобы сравнивать регрессии, используется:

- а) анализ ковариаций (ANCOVA);
- б) TRUE;
- в) `symnum(cor(longley))`;

г) однофакторный дисперсионный анализ (ANOVA).

46. С помощью данного кода вычисляется:

```
> x.norm.q <- quantile(x.norm, probs=seq(0,1,0.1))
> round(x.norm.q, 2)
      0%   10%   20%   30%   40%   50%   60%   70%   80%
90%  100%
-4.12 -1.51 -0.14  0.59  1.52  2.15  2.70  3.89  4.51
 5.22  8.15
```

- а) квантили выборки с шагом 10% (по 10 элементов в интервале);
- б) выбираем интервалы;
- в) вычисляем (по выборке) теоретические вероятности для каждого интервала;
- г) выбираем интервалы.

47. Для построения линейной модели регрессии используется функция:

- а) `lm(formula=f)`;
- б) `cor(longley)`;
- в) `boxplot()`;
- г) `anova(object)`.

48. Что произойдёт, если не указать название файла при вызове функции `svg()` или `png()`?

- а) график не будет сохранён в файл, но будет показан в правой нижней части окна
- б) график будет сохранён в файл `Rplot001` с соответствующим расширением, а при каждом следующем вызове функции этот файл будет перезаписан
- в) произойдёт ошибка, и ничего не будет сохранено
- г) график будет сохранён в файл `Rplot001` с соответствующим расширением, а при каждом следующем вызове функции будут сохраняться файлы `Rplot002`, `Rplot003` и так далее

49. Отличается ли результат работы этих двух команд?

```
ggplot(iris, aes(Species, Petal.Width, fill = Species)) + geom_boxplot()
ggplot(iris, aes(Species, Petal.Width)) + geom_boxplot(aes(fill = Species))
```

- а) да
- б) нет

50. Вы нашли коэффициент корреляции между ростом человека в см и весом в килограммах. Пусть он составляет 0.7. Изменится этот коэффициент, если рост перевести в метры?

- а) да
- б) нет

Задания открытого типа

1. Окно выбора требуемого инструмента надстройки Пакет анализа открывается командой (Сервис – Анализ данных)
2. Если статистическая функция является формулой массива, то ввод формулы необходимо завершить нажатием: (на комбинацию клавиш Ctrl+Shift+Enter)
3. Инструменты надстройки Пакет анализа возвращают результаты в виде: (массива констант)
4. При изменении исходных данных, являющихся аргументами встроенных статистических функций OpenOffice CALC, результат пересчитывается: (в автоматическом режиме)
5. Автоматический пересчет результата при изменении исходных данных отсутствует при использовании: (инструментов Пакета анализа)
6. Если статистическая функция является формулой массива, то ввод формулы необходимо завершить нажатием: (на комбинацию клавиш Ctrl+Shift+Enter)
7. Запишите уравнение регрессии на основании данных регрессионного анализа:

Регрессионная статистика								
Множественный R		0,75						
R-квадрат		0,56						
Нормированный R-квадрат		0,51						
Стандартная ошибка		0,77						
Наблюдения		20						
Дисперсионный анализ								
	df	SS	MS	F	Значимость F			
Регрессия	2	12,92	6,46	10,95	0,001			
Остаток	17	10,03	0,59					
Итого	19	22,95						
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	1,91	0,92	2,09	0,05	-0,02	3,84	-0,02	3,84
x1, %	0,04	0,03	1,64	0,12	-0,01	0,10	-0,01	0,10
x2, %	0,08	0,03	2,48	0,02	0,01	0,14	0,01	0,14

$$(y = 1,91 + 0,04x_1 + 0,08x_2)$$

8. 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)
9. Как выводятся численные результаты статистического анализа в системе R?
10. Данная группа позволяет вычислить характеристики многих стандартных вероятностных распределений: нормального, хи-квадрат, Стьюдента, F-распределения и др.
11. Параметр row.vars позволяет указать номера переменных в наборе данных, которые следует объединить в один-единственный фактор, градации

которого и будут индексировать строки таблицы сопряженности. Параметр `col.vars` проделывает то же самое, но для столбцов таблицы.

(Верно / неверно)

12. Как определить коэффициенты уравнения регрессии, используя пакет R?

13. В R пропущенные данные принято обозначать двумя большими буквами латинского алфавита «NA».

(Верно / неверно)

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

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

16. В уже созданной переменной `my_numbers` сохранен вектор из 20 целых чисел. Ваша задача создать новую переменную `my_numbers_2`, в которой будет сохранен 2, 5, 7, 9, 12, 16 и 20 элемент вектора `my_numbers`.

17. Создайте новую числовую переменную `new_var` в данных `mtcars`, которая содержит единицы в строчках, если в машине не меньше четырёх карбюраторов (переменная "carb") или больше шести цилиндров (переменная "cyl"). В строчках, в которых условие не выполняется, должны стоять нули.

18. В уже существующей переменной `my_vector` сохранен вектор из 50 чисел. Решите задачу используя конструкцию:

```
if () {  
} else {  
}
```

Если среднее значение вектора `my_vector` больше 20, в переменную `result` сохраните "My mean is great", если среднее значение `my_vector` меньше или равно 20 то в переменную `result` сохраните строку "My mean is not so great".

19. Рассчитайте среднее значение времени разгона (`qsec`) для автомобилей, число цилиндров (`cyl`) у которых не равняется 3 и показатель количества миль на галлон топлива (`mpg`) больше 20.

Получившийся результат (среднее значение) сохраните в переменную `result`.

20. Нужно построить scatterplot с помощью `ggplot` из `ggplot2`, по оси x которого будет `mpg`, по оси y - `disp`, а цветом отобразить переменную (`hp`).

Полученный график нужно сохранить в переменную `plot1`. Таким образом в ответе должен быть скрипт:

```
plot1 <- ggplot(data, aes())+  
  geom_****()
```

На сопоставление

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

1. Показатели положения: а) Эти показатели определяют, насколько кучно

- | | |
|--|--|
| <p>2. Показатели разброса:</p> <p>3. Показатели асимметрии:</p> <p>4. Показатели, описывающие закон распределения:</p> | <p>основная масса данных группируется около центра.</p> <p>б) характеризуют симметрию распределения данных около своего центра. К ним можно отнести коэффициент асимметрии, положение медианы относительно среднего и т. п.</p> <p>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).</p> <p>г) дают представление о законе распределения данных. Сюда относятся таблицы частот, таблицы частостей, полигоны, кумуляты, гистограммы.</p> |
|--|--|

Ответ 1-В; 2-А; 3-Б; 4-Г.

2. Соотнесите:

- | | |
|--|---|
| <p>1. Модуль Дисперсионный анализ (AN OVA/MAN OVA):</p> <p>2. Модуль Факторный анализ (Factor analysis):</p> <p>3. Модуль Непараметрическая статистика и подгонка распределения (Nonparametrics/Distribution):</p> | <p>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.</p> <p>б) дает возможность оценить степень воздействия различных факторов на измеряемые данные и выделить среди них наиболее значимые (существенные).</p> <p>в) дает возможность сравнить распределение наблюдаемых величин с большим количеством различных теоретических распределений.</p> |
|--|---|

Ответ 1-Б; 2-А; 3-В .

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

- | | |
|---|---|
| <p>1.</p> <p>2.</p> <p>3.</p> <p>4.</p> <p>5.</p> <p>6.</p> | <p>а) далее необходимо открыть файл данных и выбрать переменные для анализа из открытого файла;</p> <p>б) после выбора из переключателя модулей открывается стартовая панель выбранного статистического модуля;</p> <p>в) затем выбираются метод анализа и конкретная вычислительная процедура с соответствующими параметрами расчета из меню в стартовой панели модуля;</p> <p>г) далее запускается вычислительная процедура. Если</p> |
|---|---|

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;

д) используя графические возможности и специальные таблицы вывода с вычисленными разнообразными статистиками, можно осуществить всесторонний просмотр и анализ результатов;

е) выбирается следующий шаг анализа.

Ответ 1-Б; 2-А; 3-В; 4-Г; 5-Д; 6-Е.

4. Самые распространенные методы преобразований с указаниями, как их делать в R (мы предполагаем, что ваши данные находятся в векторе data) (Соотнесите):

- | | |
|------------------------------|---|
| 1) $\log(\text{data} + 1)$: | а) Логарифмическое. Если распределение скошено вправо, то может дать нормальное распределение. Может также делать более линейными зависимости между переменными и уравнивать дисперсии. «Бойтсся» нулей в данных, поэтому рекомендуется прибавлять единицу. |
| 2) $\sqrt{\text{data}}$: | б) Квадратное. Если распределение скошено влево, может дать нормальное распределение. Линеаризует зависимости и выравнивает дисперсии. |
| 3) data^2 : | в) Квадратного корня. Похоже по действию на логарифмическое. «Бойтсся» отрицательных значений. Обратное: $1/(\text{data} + 1)$. Эффективно для стабилизации дисперсии. «Бойтсся» нулей. |
| 4) $\log(p/(1-p))$: | г) Логит. Чаще всего применяется к пропорциям. Линеаризует так называемую сигмовидную кривую. Кроме логитпреобразования, для пропорций часто используют и арксинус преобразование, $\text{asin}(\sqrt{p})$ |

Ответ 1-А; 2-В; 3-Б; 4-Г.

5. Соотнесите список некоторых основных пакетов R, содержащих стандартные статистические тесты:

- | | |
|------------|--|
| 1. actest | а) – методы, используемые в “Разведочном анализе данных”; |
| 2. eda | б) – классические тесты (Фишера, “Стьюдента”, Пирсона, Бартлетта, Колмогорова- Смирнова...); |
| 3. lqs | в) – регрессия и оценка ковариации; |
| 4. modreg | г) – многомерный анализ; |
| 5. mva | д) – современные методы построения регрессионных моделей: сглаживание и локальные регрессии; |
| 6. nls | е) – нелинейные модели регрессии; |
| 7. splines | ж) – исследования временных рядов; |
| 8. stepfun | з) – эмпирические функции распределения; |
| 9. ts | |

и) – сплайны.

Ответ 1-Б; 2-А; 3-В; 4-Д; 5-Е; 6-Е; 7-И; 8-Ж; 9-З.

6. Дано Описание функции `t.test` (`x`, `y = NULL`, `alternative = c("two.sided", "less", "greater")`, `var.equal = FALSE`, `conf.level = 0.95`, `paired = FALSE,...`)
Соотнесите параметры с их значениями:

- | | |
|-----------------------------|--|
| 1. <code>y</code> | а) – Числовой вектор значений. |
| 2. <code>x</code> | б) – Числовой вектор значений (используется для парного теста, см. ниже). |
| 3. <code>alternative</code> | в) – Признак парного теста: проверяется гипотеза для <code>x</code> - <code>y</code> , поэтому вектор <code>y</code> должен присутствовать и соответствовать по длине вектору <code>x</code> . |
| 4. <code>paired</code> | г) – Символьная строка, определяющая альтернативную гипотезу, должна быть одна из "two.sided" – (по умолчанию)-двусторонняя критическая область, "greater" – правосторонняя критическая область или "less"– левосторонняя критическая область. |
| 5. <code>var.equal</code> | д) – Логическая переменная, указывающая на равенство дисперсий |
| 6. <code>conf.level</code> | е) – Доверительная вероятность. |

Ответ 1-Б; 2-А; 3-В; 4-Г; 5-Д; 6-Е.

7. Сопоставьте названия функций и описание того, что они делают:

- | | |
|---------------------------|--|
| 1. <code>replicate</code> | а) Применение заданной функции к всевозможным комбинациям двух аргументов. |
| 2. <code>mapply</code> | б) Запуск одного и того же выражения некоторое количество раз. |
| 3. <code>outer</code> | в) Применение заданной функции к списку аргументов |
| | г) Применение заданной функции к произвольному количеству векторов либо списков аргументов, в итерационном порядке |

Ответ 1-Б; 2-А; 3-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)

Эссе

1. Назначение и общая характеристика ППП "ФОРМА".
2. Содержание основных настроечных массивов в ППП "ФОРМА".
3. Назначение и общая характеристика АРМ "ПЕРМСТАТ".
4. Организация решения задач с помощью ППП "ОЛИМП".
5. Организация решения задач с помощью ППП "МЕЗОЗАВР".
6. Организация решения задач с помощью ППП "SPSS".
7. Экономическая эффективность внедрения информационных технологий.
8. Организационные формы внутримашинной технологии обработки данных.
9. Организационные основы автоматизации обработки статистической информации.
10. Технология решения задач экономического анализа и информационно-справочного обслуживания.
11. Технология интегрированной обработки статистической информации.
12. Технология обработки материалов переписей и обследований.
13. Организационные основы автоматизации обработки статистической информации на предприятиях и в организациях.
14. "Организация решения регламентных статистических задач с использованием АРМ "Пермстат".
15. Организация решения задач с помощью ППП "TATGRAPHICSPLUS".
16. Организация решения задач с помощью ППП " S-PLUS ".

Критерии и шкала оценивания по оценочному средству «эссе»

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

Оценочные средства для промежуточной аттестации

Вопросы для подготовки к зачету (с оценкой)

«Современные технологии обработки статистических данных»:

1. Современные программные пакеты средств статистического анализа.
2. Средства описательной статистики в OpenOffice CALC: возможности и ограничения.
3. Анализ вариационного ряда в OpenOffice CALC.
4. Инструменты описательной статистики в OpenOffice CALC: дисперсия, мода, медиана, стандартная ошибка.
5. Анализ формы распределения в OpenOffice CALC.
6. Уровень надежности в OpenOffice CALC. Интерпретация полученных результатов.
7. Средства дисперсионного анализа в 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 OpenOffice CALC.
11. Факторный анализ. Снижение размерности данных.
12. Методы снижения размерности и факторного анализа в R. Объединение и обобщение.

13. Выделение главных компонент по результатам анализа в R. Анализ главных компонент. Обзор результатов анализа.
14. Факторный анализ как метод классификации.
15. Подтверждающий факторный анализ (моделирование структурными соответствиями).
16. Анализ соответствий (изучение структуры качественных переменных).
17. Многомерный дисперсионный анализ.
18. Проведение основных этапов множественного регрессионного анализа в R. Корреляционная матрица. Выявление мультиколлинеарности.
19. Выбор числа переменных. Подгонка полиномиальных моделей. Определение степени и формы модели.
20. Анализ регрессионных остатков, выбросы. Критерии значимости. Критерий Дарбина-Уотсона.
21. Понятие «добычи данных». Основные направления метода добычи данных и проведение анализа средствами R.
22. Хранилища данных.
23. Оперативная аналитическая обработка (OLAP).
24. Разведочный анализ данных (РАД) и методы добычи данных.
25. Проверка гипотез в РАД.
26. Вычислительные методы РАД.
27. Графические методы. Проверка результатов РАД.
28. «Нейронные сети».
29. Основное назначение кластерного анализа. Возможности кластерного и дискриминантного анализа в R.
30. Древовидная классификация (объединение). Меры расстояния. Правила объединения.
31. Метод K средних: общая логика и постановка задачи. Примеры анализа и интерпретации результатов.
32. Дискриминантный анализ: основная цель и вычислительный подход. Пошаговый дискриминантный анализ.
33. Дискриминация для двух групп.
34. Дискриминация для нескольких групп. Проведение классификации дискриминантным методом.
35. Авторегрессионное моделирование методом скользящих средних (ARMA, ARIMA) в Eviews.
36. Проведение авторегрессионного моделирования динамики одномерных данных.
37. Выбор параметров временного лага и сезонных колебаний.
38. Проблема формальных критериев определения параметров модели.
39. Многофакторное моделирование результативной переменной на основе авторегрессионных моделей.

Критерии и шкала оценивания к промежуточной аттестации «зачет с оценкой»

Шкала оценивания (экзамен)	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.	
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.	зачтено
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
-------------------------	---	------------

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)