

**Федеральное государственное образовательное бюджетное учреждение
высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ
ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

**Кафедра математики и анализа данных
Факультета информационных технологий и анализа больших данных**

УТВЕРЖДАЮ
Проректор по учебной и
методической работе
_____ Е.А. Каменева
«26» декабря 2024 г.

Фридман М. Н., Цыганок И.И.

АНАЛИЗ ДАННЫХ

**Рабочая программа дисциплины
для студентов, обучающихся
по направлению подготовки 38.03.04 –
Государственное и муниципальное управление,
ОП «Государственное и муниципальное управление».**

*Рекомендовано Ученым советом
Факультета информационных технологий и анализа больших данных
(протокол №50 от 17 декабря 2024 г.)*

*Одобрено Советом кафедры математики и анализа данных
(протокол №4 от 18 ноября 2024 г.)*

Москва 2024

СОДЕРЖАНИЕ

1. Наименование дисциплины	3
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	3
3. Место дисциплины в структуре образовательной программы	5
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	5
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий.....	6
5.1. Содержание дисциплины.....	6
5.2. Учебно-тематический план.....	11
5.3. Содержание семинаров, практических занятий.....	13
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	14
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы.....	14
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю	16
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по данной дисциплине.....	25
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	43
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	45
10. Методические указания для обучающихся по освоению дисциплины	46
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.....	47
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	48

1. Наименование дисциплины

«Анализ данных».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Дисциплина «Анализ данных» обеспечивает формирование компетенций:

УК-4, УК-10, УК-15

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
УК-4	Способность использовать прикладное программное обеспечение при решении профессиональных задач	1.Использует основные методы и средства получения, представления, хранения и обработки данных.	Знать: источники актуальных данных о состоянии экономических субъектов Уметь: осуществлять поиск и импорт данных, необходимых для решения конкретных задач
		2.Демонстрирует владение профессиональными пакетами прикладных программ.	Знать: основные принципы программирования алгоритмов статистической обработки данных в табличных редакторах и R(RStudio) для анализа и прогноза внутри- и внешнеэкономических процессов Уметь: применять технические и инструментальные средства для решения прикладных задач управления и подготовки обоснований управленческих решений
		3.Выбирает необходимое прикладное программное обеспечение в зависимости от решаемой задачи.	Знать: тенденции развития новых программных продуктов и IT-сервисов, ориентированных на повышение эффективности сбора и обработки данных. Уметь: обосновывать выбор ПО в зависимости от решаемой задачи
		4. Использует прикладное программное обеспечение для решения конкретных прикладных задач.	Знать: основные принципы обработки, визуализации и анализа данных, вероятностные и статистические методы. Уметь: использовать инструменты описательной статистики и визуализации данных, вероятностные и статистические методы для решения

			организационно-управленческих задач, интерпретировать полученные результаты.
УК-10	Способность осуществлять поиск, критически анализировать, обобщать и систематизировать информацию, использовать системный подход для решения поставленных задач	1. Четко описывает состав и структуру требуемых данных и информации, грамотно реализует процессы их сбора, обработки и интерпретации.	<u>Знать:</u> основные принципы предварительной обработки, визуализации и анализа данных, вероятностные и статистические методы. <u>Уметь:</u> использовать инструменты предварительной обработки и аккуратизации данных
		2. Обосновывает сущность происходящего, выявляет закономерности, понимает природу вариабельности.	<u>Знать:</u> методы анализа тенденций развития экономических процессов <u>Уметь:</u> применять технические и инструментальные средства для оценки содержания экономической повестки дня
		3. Формулирует признак классификации, выделяет соответствующие ему группы однородных «объектов», идентифицирует общие свойства элементов этих групп, оценивает полноту результатов классификации, показывает прикладное назначение классификационных групп	<u>Знать:</u> современные методы многомерной классификации <u>Уметь:</u> применять классификационные методы, выбирая адекватные задаче математические и инструментальные методы, интерпретировать полученные результаты
		4. Грамотно, логично, аргументировано формирует собственные суждения и оценки. Отличает факты от мнений, интерпретаций, оценок и т. д. в рассуждениях других участников деятельности	<u>Знать:</u> основные методы теории вероятностей, математической и прикладной статистики, используемые для подготовки аналитических решений, экспертных заключений и рекомендаций. <u>Уметь:</u> использовать вероятностные и статистические методы для подготовки аналитических решений, 5 экспертных заключений и рекомендаций
		5. Аргументированно и логично представляет свою точку зрения посредством и на основе системного описания.	<u>Знать:</u> процедуры формулировки и проверки статистических гипотез <u>Уметь:</u> проверять гипотезы и делать обоснованные заключения на основе проверки

УК-15	Способность релевантно решать задачам использовать информационные ресурсы и информационно коммуникационные технологии для достижения целей, связанных с профессиональной деятельностью, обучением, участием в жизни общества и других сферах жизни	1. Самостоятельно выбирает и использует цифровые средства общения, осуществляет поиск и/или создание контента в соответствии с целью взаимодействия, в том числе для организации совместной деятельности.	<u>Знать:</u> современные информационные технологии анализа данных и особенности их применения для решения стандартных задач профессиональной деятельности; <u>Уметь:</u> осуществить выбор программных средств для анализа данных в исследованиях
		2. Владеет навыками организации взаимодействия и коммуникации с помощью информационных систем и/или цифровых сервисов и технологий.	<u>Знать:</u> тенденции развития технологии сбора данных в сети Internet <u>Уметь:</u> осуществлять поиск, сбор и первичную подготовку данных для прикладных исследований
		3. Осуществляет подбор и применение различных информационно коммуникационных средств для решения образовательных и профессиональных задач.	<u>Знать:</u> особенности современных информационных технологий представления результатов исследования <u>Уметь:</u> осуществить выбор информационной технологии для анализа данных и презентации результатов анализа

3. Место дисциплины в структуре образовательной программы

Дисциплина «Анализ данных» является обязательной дисциплиной Цикла математики и информатики по направлению подготовки 38.03.04 - Государственное и муниципальное управление. Дисциплина «Анализ данных» базируется на знаниях, полученных при изучении дисциплин «Математика», «Компьютерный практикум».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Институт онлайн-образования

Очно-заочная форма обучения

Таблица 1

Вид учебной работы по дисциплине	Всего (в з./е. и часах)	Семестр 3 (в часах)	Семестр 4 (в часах)
Общая трудоемкость дисциплины	6/216	108	108
Контактная работа - Аудиторные занятия	68	34	34
Лекции	20	8	12
Семинары, практические занятия	48	26	22
Самостоятельная работа	148	74	74
Вид текущего контроля	Домашнее творческое задание/Расчетно-аналитические работы	Домашнее творческое задание	Расчетно-аналитическая работа
Вид промежуточной аттестации	Зачет/Экзамен	Зачет	Экзамен

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Данные в экономике, их визуализация и предварительная обработка

1.1. Данные в экономике. Объекты, признаки и таблицы. Типы признаков в экономике: интервальные, порядковые, ранговые, дихотомические. Форматирование наборов данных как таблиц в табличном редакторе. Гистограммы в табличном редакторе. Условное форматирование в табличном редакторе. Графики и диаграммы рассеяния в табличном редакторе.

1.2. Инструменты описательной статистики в табличном редакторе. Измерение центра распределения. Измерение разброса данных. Описательная статистика в табличном редакторе и в R. Диаграммы размаха в R.

1.3. Предварительная обработка данных. Выбросы и их обработка в табличном редакторе. Пропущенные значения и их обработка в табличном редакторе.

Повторяющиеся строки и их обработка в табличном редакторе. Синтетические признаки.

Тема 2. Случайные события

2.1. Определение вероятности. Случайные события, их виды. Операции над событиями как операции над множествами. Комбинаторика. Классическая вероятностная схема. Схема геометрических вероятностей. Статистическая вероятность. Аксиомы теории вероятностей. Метод Монте-Карло. Моделирование случайных событий.

2.2. Основные теоремы теории вероятностей. Теорема сложения вероятностей. Условная вероятность. Независимость событий. Формула полной вероятности. Формула Байеса. Простейшие примеры применения теории вероятностей.

Тема 3. Случайные величины

3.1. Определение случайной величины. Понятие случайной величины. Функция распределения случайной величины. Свойства функции распределения. Индикатор события как простейшая случайная величина. Функция распределения индикатора события. Смеси распределений.

3.2. Дискретные случайные величины и их важнейшие числовые характеристики. Дискретная случайная величина. Ряд распределения и функция распределения дискретной случайной величины. Математическое ожидание, дисперсия и среднее квадратическое отклонение дискретной случайной величины.

3.3. Дискретные случайные величины, часто встречающиеся в практике. Последовательности испытаний. Биномиальная схема. Отрицательное биномиальное распределение (биномиальный закон распределения, геометрические законы распределения). Закон распределения Пуассона. Простейший поток событий. Гипергеометрический закон распределения.

3.4. Абсолютно непрерывные случайные величины и их важнейшие числовые характеристики. Абсолютно непрерывная случайная величина. Функция

распределения и функция плотности распределения абсолютно непрерывной случайной величины. Свойства функции плотности распределения. Математическое ожидание, дисперсия и среднее квадратическое отклонение абсолютно непрерывной случайной величины.

3.5. Абсолютно непрерывные случайные величины, часто встречающиеся в практике. Равномерный закон распределения. Экспоненциальный закон распределения. Нормальный закон распределения. Законы распределения, важные в математической статистике (законы распределения Стьюдента, χ^2 , Фишера — Снедекора). Реализация моделей абсолютно непрерывных случайных величин в пакете табличный редактор.

3.6. Моменты и критические границы случайной величины. Начальные и центральные моменты случайной величины. Асимметрия и эксцесс случайной величины. Квантили и процентные точки случайной величины. Вычисление квантилей и процентных точек в табличном редакторе. Медиана и мода случайной величины.

3.7. Системы случайных величин. Случайные векторы и условные законы распределения. Условный ряд распределения (для дискретных случайных величин), условная плотность распределения (для непрерывных случайных величин). Условное математическое ожидание. Формула полного математического ожидания. Формула полной дисперсии. Ковариация и коэффициент корреляции.

Тема 4. Предельные теоремы теории вероятностей

4.1. Закон больших чисел. Массовые случайные явления в экономике. Теорема Чебышёва и оценка математического ожидания. Теорема Бернулли и оценка вероятности. Обсуждение условий статистической устойчивости.

4.2. Центральная предельная теорема. Теорема Леви. Интегральная теорема Муавра — Лапласа. Место центральной предельной теоремы в изучении статистических закономерностей.

Тема 5. Выборочный метод математической статистики

5.1. Основы выборочного метода. Предмет и задачи математической статистики. Генеральная и выборочная совокупности. Случайная и конкретная выборки. Случайная повторная и случайная бесповторная выборки. Соотношение между предельной ошибкой среднего, уровнем значимости (риском) и объемом выборки. Использование этого соотношения в организации выборочных обследований.

5.2. Оценка плотности распределения и функции распределения. Вариационный ряд. Числовые характеристики выборки. Интервальный вариационный ряд. Полигон частот, кумулята. Оценка числовых характеристик генеральной случайной величины (выборочное среднее как оценка математического ожидания, относительная частота как оценка вероятности, выборочная дисперсия как оценка дисперсии).

5.3. Точечные оценки параметров. Понятие точечной оценки параметра генеральной совокупности. Свойства точечных оценок: состоятельность, несмещенность, эффективность. Выборочное среднее как состоятельная, несмещенная и эффективная оценка математического ожидания генеральной случайной величины. Смещенность выборочной дисперсии как оценки дисперсии генеральной случайной величины. Исправленная выборочная дисперсия как несмещенная и состоятельная оценка дисперсии генеральной случайной величины (для повторной и бесповторной выборок). Методы построения точечных оценок: метод моментов, метод максимального правдоподобия. Построение оценок параметров распределений случайных величин.

5.4. Интервальные оценки параметров. Понятие интервальной оценки параметра генеральной совокупности. Интервальные оценки вероятности, математического ожидания, дисперсии и коэффициента корреляции. Поправка на конечный объем генеральной совокупности (для повторной и бесповторной выборок). Асимптотический подход к интервальному оцениванию.

Тема 6. Проверка статистических гипотез

6.1. Статистические гипотезы. Понятие статистической гипотезы. Виды статистических гипотез: параметрические и непараметрические, простые и сложные. Критерий проверки гипотезы, критическое множество. Проверка гипотез с помощью интервальных оценок. Ошибки первого и второго рода. Мощность критерия. Наиболее мощный критерий. Достижимый уровень значимости (p-value).

6.2. Критерии случайности, независимости, однородности. Критерий Граббса (отсутствия аномальных значений). Таблицы сопряженности. Критерии однородности χ^2 и Смирнова. Критерий независимости χ^2 , критерий значимости коэффициента корреляции. Критерий инверсий.

6.3. Критерии согласия. Критерий согласия χ^2 Пирсона. Критерий χ^2 Пирсона для дискретных распределений. Критерии согласия Колмогорова и Смирнова.

6.4. Критерии согласия для сложных гипотез. Критерии для проверки нормальности (Лиллиефорса, Шапиро-Уилка и другие)

6.5. Параметрические критерии. Проверка гипотезы о равенстве математического ожидания теоретическому значению. Проверка гипотезы о равенстве двух математических ожиданий. Проверка гипотезы о равенстве дисперсии теоретическому значению. Проверка гипотезы о равенстве двух дисперсий. Проверка гипотезы о равенстве вероятности события теоретическому значению. Проверка гипотезы о равенстве двух вероятностей. Проверка гипотез для связанных выборок.

6.6. Основы непараметрической статистики. Распределение Вилкоксона и его критические границы. Критерий Вилкоксона-Манна-Уитни (непараметрический критерий сравнения математических ожиданий для независимых выборок). Коэффициент ранговой корреляции Спирмена. Коэффициент ранговой корреляции Кендалла. Коэффициент конкордации.

Тема 7. Введение в методы многомерной статистической классификации

7.1. Задачи регрессионного анализа. Парная регрессия. Спецификация регрессионной модели. Результативный признак и факторы. Существенность факторов. Выбор формы регрессионного уравнения. Оценка параметров уравнения парной регрессии методом наименьших квадратов. Модель множественной регрессии. Процедура отбора факторов. Теорема Гаусса-Маркова. Проблема мультиколлинеарности. Оценка параметров уравнения множественной регрессии средствами Ms Excel и R. Логистическая регрессия и задачи бинарной классификации.

7.2. Основы дискриминантного анализа. Предпосылки метода. Классификация с обучением. Обучающие выборки. Линейный дискриминантный анализ. Дискриминантная функция. Линейный дискриминантный анализ в R.

7.3. Основы кластерного анализа. Предпосылки метода. Иерархические кластер-процедуры. Принцип классификации без обучения. Меры расстояний между объектами. Метрики расстояния между кластерами. Метод К-средних. Реализация процедур кластерного анализа в R.

5.2. Учебно-тематический план

Таблица 2

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа			Самостоятельная работа	
			Общая, в т.ч.	Лекции	Семинары, практические занятия		
1	Данные в экономике, их визуализация и предварительная обработка		4	2	2	14	Вовлеченность в интерактивный процесс, выступления на вебинарах, домашние задания, участие в обсуждениях результатов в чате, расчетно-аналитическая работа №1 по темам 1-4

2	Случайные события		10	2	8	24	Вовлеченность в интерактивный процесс, выступления на вебинарах, домашние задания, участие в обсуждениях результатов в чате, расчетно-аналитическая работа №1 по темам 1-4
3	Случайные величины		15	3	12	24	Вовлеченность в интерактивный процесс, выступления на вебинарах, домашние задания, участие в обсуждениях результатов в чате, расчетно-аналитическая работа №1 по темам 1-4
4	Предельные теоремы теории вероятностей		5	1	4	12	Вовлеченность в интерактивный процесс, выступления на вебинарах, домашние задания, участие в обсуждениях результатов в чате, расчетно-аналитическая работа №1 по темам 1-4
	Итого 1 семестр	108	34	8	26	74	
5	Выборочный метод математической статистики		12	4	8	22	Вовлеченность в интерактивный процесс, выступления на вебинарах, домашние задания, участие в обсуждениях результатов в чате, расчетно-аналитическая работа №2 по темам 5-7
6	Проверка статистических гипотез		12	4	8	28	Вовлеченность в интерактивный процесс, выступления на вебинарах, домашние задания, участие в обсуждениях результатов в чате, расчетно-аналитическая работа №2 по темам 5-7
7	Введение в методы многомерной статистической классификации		10	4	6	24	Вовлеченность в интерактивный процесс, выступления на вебинарах, домашние задания, участие в обсуждениях результатов в чате, расчетно-аналитическая работа №2 по темам 5-7

	Итого 2 семестр	108	34	12	22	74	Вовлеченность в интер-активный процесс, выступления на вебинарах, домашние задания, участие в обсуждениях результатов в чате, расчетно-аналитическая работа №2 по темам 5-7
	В целом по дисциплине	216	68	20	48	148	Согласно учебному плану: расчетно-аналитическая работа/домашнее творческое задание
	Итого в %		31	29	71	69	

*объем контактной работы в очно-заочной/заочной формах обучения и индивидуальных учебных планах определяется соответствующими учебными планами. Темы, реализуемые в виде контактной работы, определяются преподавателем самостоятельно, исходя из уровня их сложности.

5.3. Содержание семинаров, практических занятий

Таблица 3

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Данные в экономике, их визуализация и предварительная обработка	Представление данных в виде форматированных таблиц. Гистограммы, диаграммы рассеяния, диаграммы размаха. Сводные таблицы. Визуализация качественных признаков. Импорт, предварительная обработка и визуализация данных в Excel. <i>Рекомендуемые источники:</i> [8.1,3,4], [9.2]	Решение задач онлайн, обсуждение в чате вебинара
Случайные события	Основы комбинаторики. Функции подсчета количества комбинаций в Excel и R. Классическая вероятностная схема. Схема геометрических вероятностей. Моделирование статистической вероятности. Вероятность суммы. Условные вероятности. Вероятность произведения. Полная вероятность. Формула Байеса. Последовательности испытаний. Схема Бернулли. <i>Рекомендуемые источники:</i> [8.1-4], [9.2]	Решение задач онлайн, обсуждение в чате вебинара
Случайные величины	Дискретные случайные величины и их важнейшие числовые характеристики. Дискретные случайные величины, встречающиеся в экономиче-	Решение задач онлайн, обсуждение в чате вебинара

	ской практике. Абсолютно непрерывные случайные величины и их важнейшие числовые характеристики. Абсолютно непрерывные величины, встречающиеся в практике экономики. Равномерный, показательный, нормальный законы распределения. Моменты случайных величин. Квантили и 100 α -процентные точки. Формулы корреляции и ковариации. Многомерная случайная величина. Условные распределения. <i>Рекомендуемые источники:</i> [8.1-4], [9.2]	
Предельные теоремы теории вероятностей	Применение неравенств Маркова и Чебышёва для решения практических задач <i>Рекомендуемые источники:</i> [8.1-6], [9.2]	Решение задач онлайн, обсуждение в чате вебинара
Выборочный метод математической статистики	Выборочные оценки функций плотности и распределения. Точечные оценки параметров. Описательная статистика средствами Ms Excel. Интервальные оценки параметров. <i>Рекомендуемые источники:</i> [8.2-6], [9.2]	Решение задач онлайн, обсуждение в чате вебинара
Проверка статистических гипотез	Формулировка статистически гипотез. Критерии проверки статистических гипотез. Параметрические гипотезы. Одно- и двухвыборочные тесты. Функции var.test(), t.test(), prop.test() R. Функции СТЬЮДЕНТ.ТЕСТ(), FТЕСТ() Ms Excel. Критерии согласия. Проверка значимости коэффициента корреляции Пирсона. Непараметрическая точечная оценка математического ожидания. Непараметрическая интервальная оценка математического ожидания. Критерий Вилкоксона (парный критерий знаковых рангов). Коэффициент ранговой корреляции Спирмена. <i>Рекомендуемые источники:</i> [8.2-6], [9.2]	Решение задач онлайн, обсуждение в чате вебинара
Введение в методы многомерной статистической классификации	Оценка параметров уравнения линейной регрессии в Ms Excel. Интерпретация результатов. Решение задач кластерного анализа. <i>Рекомендуемые источники:</i> [8.1,2,3,4], [9.2]	Решение задач онлайн, обсуждение в чате вебинара

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Таблица 4

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоя- тельное освоение	Формы внеаудиторной самостоятельной работы
1. Данные в эконо- мике, их визуализация и предварительная об- работка	Выбросы и их обработка в Microsoft Excel. Про- пущенные значения и их обработка в Microsoft Excel. Повторяющиеся строки и их обработка в Microsoft Excel. Синтетические признаки Импорт, предварительная обработка и визуали- зация данных в R	Работа с учебной литературой. Ре- шение типовых задач. Выполне- ние теста для са- моподготовки в ЭУКе на sam- pus.fa.ru. Выполнение рас- четно-аналитиче- ской работы/до- машнего творче- ского задания
2. Случайные события	Реализация сочетаний и размещений. Алгебра событий. Диаграммы Эйлера-Венна	Работа с учебной литературой. Ре- шение типовых задач. Выполне- ние тестов для са- моподготовки по теме в ЭУКе на портале sam- pus.fa.ru. Выпол- нение расчетно- аналитической работы/домаш- него творческого задания
3. Случайные величины	Полное матожидание и полная дисперсия. Многомерные непрерывные случайные вели- чины и условные распределения в экономике и управлении. Смеси распределений.	Работа с учебной литературой. Ре- шение типовых задач. Выполне- ние тестов для са- моподготовки по теме в ЭУКе на портале sam- pus.fa.ru. Выпол- нение расчетно- аналитической работы/домаш- него творческого задания
4. Предельные теоремы теории вероятностей	Метод Монте-Карло. Моделирование случайных величин в Excel и R.	Работа с учебной литературой. Ре- шение типовых

		задач. Выполнение тестов для самоподготовки по теме в ЭУКе на портале sam-pus.fa.ru. Вычисление характеристик случайных величин в процессе моделирования. Выполнение расчетно-аналитической работы/домашнего творческого задания
5. Выборочный метод математической статистики	Практические задачи оценки параметров генеральной совокупности. Описательная статистика средствами R	Работа с учебной литературой. Решение типовых задач. Выполнение тестов для самоподготовки по теме в ЭУКе на портале sam-pus.fa.ru. Выполнение дополнительных заданий на применение Excel и R. Выполнение расчетно-аналитической работы/домашнего творческого задания
6. Проверка статистических гипотез	Задачи непараметрической статистики в практике управления. Критерий Вилкоксона — Манна — Уитни (непараметрический критерий сравнения математических ожиданий для независимых выборок).	Работа с учебной литературой. Решение типовых задач. Выполнение тестов для самоподготовки по теме в ЭУКе на портале sam-pus.fa.ru. Выполнение дополнительных заданий на применение критериев: Уилкоксона, знаков, Манна- Уитни в Excel и R. Выполнение расчетно-

		аналитической работы/домашнего творческого задания
7. Введение в методы многомерной статистической классификации	Оценка параметров уравнения логистической регрессии. Бинарная классификация объектов средствами машинного обучения. Регрессионный анализ в R. Решение задач дискриминантного анализа	Работа с учебной литературой. Решение типовых задач. Выполнение тестов для самоподготовки по теме в ЭУКе на портале sam-rus.fa.ru. Выполнение дополнительных заданий регрессионного анализа в Excel и R. Выполнение расчетно-аналитической работы/домашнего творческого задания

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Пример домашнего творческого задания

Домашнее творческое задание в 1 семестре состоит в организации, проведении социологического опроса и обработки их результатов с использованием ИТ-инструментов (Excel, язык R, Python или другое средство программирования).

Для составления вопросов анкеты студенту-исследователю необходимо учесть следующие особенности:

- выбрать актуальную тему, чтобы заинтересовать респондентов участвовать в анкетировании;
- привлечь к участию не менее 70 респондентов;
- формулировка вопросов должна однозначно трактовать суть проблемы;
- использовать различные форматы ввода ответов;
- вопросы на ответы должны содержать как числовые ответы, так и категории.

Проведите свое небольшое социологическое исследование, реализовав следующий алгоритм действий:

1. Определитесь с тем, что хотите исследовать: объект исследования, представленный группой респондентов, и его исследуемое качество, свойство, отношение.
2. Составьте небольшую анкету (3–4 вопроса), ответы на которую помогут ответить на вопросы исследования.
3. Проверьте однозначность вопросов и корректность предлагаемых ответов.
4. Запрограммируйте свою анкету в Microsoft Forms (для этого в браузере необходимо запустить ссылку <https://forms.office.com/>) или Yandex Forms (<https://yandex.cloud/ru/docs/forms/new-form=>). Для создания опроса необходимо нажать кнопку «Создать форму».

В данной форме обучающийся формирует анкету с вопросами и выбирает соответствующий формат ответа.

5. Когда анкета готова полностью, студент отправляет ссылку на опрос респондентам, которые будут участвовать в социологическом исследовании.

(это могут быть друзья, студенты, коллеги по работе, случайные интернет-пользователи, люди с форумов или иных групп). Соберите необходимое количество ответов респондентов.

6. Для обработки полученных результатов опроса необходимо скачать все ответы в файл Excel, чтобы получить отчет. Визуализируйте результаты опроса с помощью Excel или R(RStudio). При этом для категориальных переменных должны быть построены гистограммы, а для числовых переменных – гистограммы и ящики с усами. Дайте интерпретацию результатов. Оформите и распечатайте результаты в MS Word с краткими выводами исследования.

Примеры заданий теста

1. При изучении интенсивности потока звонков в колл-центр в течение 15 дней получены следующие результаты:

112; 251; 300; 112; 93; 200; 130; 112; 120; 175; 200; 243; 287; 290; 302

Найти моду, медиану и среднее значение. Пояснить различие между медианой и средним значением. Построить диаграммы, визуализирующие полученные данные. (Тема 1)

2. Для каждого студента группы в таблице записаны: оценка по математике в первом семестре; размер обуви; любимый фильм; расстояние от места проживания до места учебы. Какие из этих признаков являются: а) количественными; б) качественными; в) дискретными; г) непрерывными; д) номинальными; е) порядковыми? Какие виды диаграмм можно применить к каждому из этих признаков? (Тема 1)

3. Из 250 студентов первого курса, сдававших в сессию три экзамена – по математике, философии и иностранному языку, отметку «удовлетворительно» получили: по математике 86 человек, по иностранному языку – 71, по философии – 50, по математике или иностранному языку – 130, по математике или философии – 112, по иностранному языку или философии – 94, по всем трем предметам – 18 человек. Сколько студентов сдали эти экзамены:

- а) без троек;
- б) с одной тройкой по математике;
- в) с одной тройкой. (Тема 2)

4. В десятиугольнике случайным образом выбираются две вершины. Чему равна вероятность того, что эти вершины являются соседними? (по теме 2)

5. Два теплохода должны подойти к одному и тому же причалу. Время прихода теплоходов независимо и равновозможно в течение данных суток. Определить вероятность того, что одному из теплоходов придется ожидать причала, если время стоянки первого теплохода – 1 час, а второго – 2 часа (Тема 2)

6. В некоторой страховой компании среди держателей полисов страхования автогражданской ответственности 30% относятся к группе высокого риска (у таких страхователей вероятность наступления страхового случая в течение года равна 0,2), а 70% – к группе низкого риска (у них вероятность наступления страхового случая в течение года равна 0,01). С какой вероятностью

случайно выбранный страхователь, у которого в прошлом году был страховой случай, относится к группе: а) высокого риска; б) низкого риска? (Тема 2)

7. Прибор, установленный на борту самолета, может работать в двух режимах: в условиях нормального полета и в условиях перегрузки при взлете и посадке. Нормальный режим полета составляет 80% времени полета, перегрузка – 20%. Вероятность выхода прибора из строя за время полета в нормальном режиме равна 0,01, в условиях перегрузки – 0,04. 1) Вычислить надежность прибора за время полета. 2) Известно, что прибор вышел из строя. Какова вероятность, что при этом полет происходил в нормальном режиме? (Тема 2)

8. Текущий тест по анализу данных состоит из трех разделов, для каждого из которых определен проходной балл. Вероятность того, что студент получит проходной балл по первому разделу теста равна 0,9, по второму – 0,8, по третьему – 0,7. Составить закон распределения случайной величины – числа разделов теста, по которым получен проходной балл и найти наиболее вероятное число таких разделов. Построить полигон распределения. Определить числовые характеристики случайной величины: математическое ожидание, дисперсию и среднее квадратическое отклонение. Найти функцию распределения и построить ее график. Найти вероятность того, что случайная величина примет значение: а) в промежутке $[1; 3]$; б) не менее чем 0,5; в) в промежутке $[1, 5; 3]$. (Тема 3)

9. Партия товара, хранящегося на складе, содержит 20 упаковок, среди которых 8 с просроченным сроком годности. Случайным образом выбирают 4 упаковки и отправляют заказчику. Составить закон распределения случайной величины – числа упаковок с непросроченным сроком годности. Построить полигон распределения. Определить числовые характеристики случайной величины: математическое ожидание, дисперсию и среднее квадратическое отклонение. Найти функцию распределения и построить ее график. Найти вероятность того, что случайная величина примет значение: а) в промежутке $[1; 5]$; б) менее, чем 3; в) в промежутке $[1; 3]$. (Тема 3)

10. Для случайных величин X и Y известно, что $E(X) = 2$,

$E(Y) = 3$, $Var(X) = 8$, $Var(Y) = 32$, $R(X, Y) = 0,25$. Найти $E(XY)$ и $Var(2X - 3Y + 5)$. (Тема3)

11. Величина X распределена по закону Пуассона с параметром $\lambda=2$. Тогда вероятность $P(0,5\sigma(X) \leq X - E(X) \leq 1 + \sigma(X))$ равна: (Тема3)

12. Тест состоит из 10 вопросов, на каждый из них предлагают 5 варианта ответа, из которых только один – правильный. Тест заполняется наугад, то есть ответы выбираются случайным образом. Пусть величина X – число правильных ответов на вопросы теста. Тогда дисперсия $Var(X)$ этой величины равна: (Тема3)

Примеры заданий расчетно-аналитической работы

1. В 10 измерениях длины стержня одним прибором получены следующие результаты: 5,95; 5,71; 5,93; 5,97; 5,97; 6,0; 5,92; 6,1; 5,92; 5,98. Найти среднее значение, выборочную и "исправленную" выборочную дисперсии показаний прибора и соответствующие средние квадратические отклонения, выборочную долю показаний, не превосходящих 5,95. (Тема 5)

2. В некоторой области по схеме собственно случайной бесповторной выборки было обследовано 100 предприятий малого бизнеса из 2500 с целью изучения объема привлечённых инвестиций. Получены следующие данные:

Величина выплаты (т.руб.)	Менее 600	600–700	700–800	800–900	900–1000	1000– 1100	Итого
Число выплат	14	21	25	20	6	4	100

Найти: а) вероятность того, что средняя выплата отличается от средней

выплаты в выборке не более чем на 15 т. руб.; б) границы, в которых с вероятностью 0,98 заключена доля предприятий, с объемом инвестиций от 600 до 900 т. руб; в) объем бесповторной выборки, при котором те же границы для среднего объема инвестиций (см. п. а)) можно гарантировать с вероятностью 0,95. (Тема 5)

3. При испытании 1000 приборов зарегистрировано 102 сбоя работы. Найти доверительный интервал, покрывающий неизвестную вероятность p сбоя работы прибора с надежностью 0,85. (Тема 5)

4. Глубина моря измеряется прибором, систематическая ошибка которого равна нулю, а случайные ошибки распределены нормально со среднеквадратичным отклонением $\sigma = 16$ м. Каково наименьшее число независимых измерений, при котором удастся определить глубину с ошибкой меньше 2 метров с надежностью не ниже 0.97? (Тема 5)

5. С целью определения средней суммы вкладов на 1 января текущего года в сберегательном банке, имеющем 2000 вкладчиков, по схеме собственно-случайной выборки с бесповторным отбором членов проведено обследование 200 лицевых счетов. Распределение вкладов по их величине (тыс. руб.) представлено в таблице:

612	442	498	284	667	563	709	388	518	717
218	600	605	131	547	517	448	818	732	842
501	385	238	682	400	498	305	610	463	618
537	453	546	723	190	608	607	620	117	705
562	212	520	414	316	408	405	355	457	569
367	429	254	568	413	572	423	755	154	588
594	473	340	335	566	402	401	502	756	558
792	565	474	526	502	408	674	828	483	465
596	670	502	601	452	523	741	261	327	556
541	496	141	274	394	555	409	511	644	560
549	763	739	455	475	287	522	743	535	630
494	562	488	562	656	559	540	592	591	348
498	495	457	644	379	877	398	272	363	597
231	539	667	583	369	492	559	662	239	532
574	568	621	663	223	714	649	476	619	428
494	567	536	359	502	511	389	621	573	305
520	561	634	609	563	359	343	702	489	136
725	495	507	627	775	489	419	430	598	511

661	593	386	643	182	366	611	464	665	427
389	779	761	644	607	536	706	694	462	354

Составить интервальный вариационный ряд. Записать эмпирическую функцию распределения и построить ее график. На одном чертеже изобразить гистограмму и полигон частот. Используя возможности Excel и R, вычислить выборочные числовые характеристики: среднее арифметическое, исправленную выборочную дисперсию, среднее квадратичное отклонение, коэффициент вариации, асимметрию, эксцесс, моду и медиану. (Тема 5)

6. По данным предыдущей задачи, заменить параметры генеральной совокупности соответственно их наилучшими выборочными числовыми характеристиками и используя χ^2 -критерий Пирсона, на уровне значимости $\alpha = 0,05$ проверить две гипотезы о том, что изучаемая случайная величина X – величина транспортных затрат распределена: а) по нормальному закону распределения; б) по равномерному закону распределения. Построить чертёж, на котором изображена гистограмма эмпирического распределения и соответствующие графики равномерного и нормального распределений. (Темы 5, 6)

7. По извлеченной из нормальной генеральной совокупности с известным стандартным отклонением $\sigma = 1,9$ случайной выборке объема $n = 396$ найдено выборочное среднее $\bar{X} = 39,2$. На уровне значимости $\alpha = 0,08$ проверить нулевую гипотезу $H_0: E(X) = 39,3$ при альтернативной гипотезе $H_1: E(X) < 41,9$ (Тема 6)

8. Ежедневные доходы (в тыс. руб.) двух аптек задаются нормальными независимыми случайными величинами X и Y с неизвестными генеральными дисперсиями. $\sigma_x^2 = 0,14$ и $\sigma_y^2 = 0,11$. По результатам 10 наблюдений величины X и 17 наблюдений величины Y были получены выборочные значения средних: $\bar{x} = 3,4$ и $\bar{y} = 5,6$ и выборочные дисперсии $\sigma_x^2 = 0,14$ и $\sigma_y^2 = 0,11$. На уровне значимости $\alpha = 0,01$ проверить гипотезу о равенстве средних ежедневных доходов аптек $H_0: E(X) = E(Y)$ против альтернативы $H_1: E(X) < E(Y)$. (Тема 6)

9. При проверке на уровне значимости $\alpha = 0,05$ гипотезы о том, что курение существенно влияет на формирование повышенного артериального давления получены следующие результаты:

	Имеют повышенное артериальное давление	Имеют артериальное давление в пределах нормы
Курят	40	30
Не курят	32	48

Необходимо:

1) выбрать критерий согласия для проверки этой гипотезы; 2) в случае применения критерия χ^2 найти наблюдаемое значение статистики для критерия χ^2 ; 3) в случае применения критерия χ^2 найти критическое значение для формирования правосторонней критической области и сделать вывод. (Тема 6)

10. Распределение 100 средних фермерских хозяйств по числу наемных рабочих X (чел.) и их средней месячной заработной плате на одного человека Y (тыс. руб.) представлено в таблице:

$X \backslash Y$	10-15	15-20	20-25	25-30	30-35	Итого
20-25			6	8	4	18
25-30		2	10	2	2	16
30-35	2	6	8	2		18
35-40	4	12	10	2		28
40-45	10	6	4			20
Итого	16	26	38	14	6	100

Необходимо:

1) Вычислить групповые средние x_i и y_j , построить эмпирические линии регрессии;
2) Предполагая, что между переменными X и Y существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

б) вычислить коэффициент корреляции; на уровне значимости $\alpha = 0,05$ оценить его значимость и сделать вывод о тесноте и направлении связи между переменными X и Y ;

в) используя соответствующее уравнение регрессии, оценить среднюю месячную заработную плату одного рабочего в хозяйстве, в котором работают 7 наемных рабочих. (Тема 7)

11. При исследовании корреляционной зависимости между количеством заявок на участие в ежемесячном телевизионном конкурсе $Y(y_i)$ и затратами в месяц на рекламу конкурса $X(x_i)$ получены следующие данные:

x_i	36	40	42	44	41	49	56	51	54	55
y_i	100	120	115	120	117	122	120	118	121	124

Обозначив уравнение регрессии Y на X как $y = \alpha_{yx} + \beta_{yx}x$, а уравнение регрессии X на Y как $x = \alpha_{xy} + \beta_{xy}y$, найти: 1) произведение коэффициентов β_{yx} и β_{xy} ; 2) значения коэффициентов α_{yx} и α_{xy} ; 3) выборочный коэффициент линейной корреляции. (Тема 7)

Критерии бальной оценки различных форм текущего контроля успеваемости

Критерии бальной оценки различных форм текущего контроля успеваемости содержится в соответствующих методических рекомендациях кафедры математики и анализа данных.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по данной дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. «Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

**Типовые контрольные задания или иные материалы, необходимые для
оценки индикаторов достижения компетенций, умений и знаний**

Таблица 5

Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
Способность использовать прикладное программное обеспечение при решении профессиональных задач (УК-4)	1.Использует основные методы и средства получения, представления, хранения и обработки данных.	Знать: источники актуальных данных о состоянии экономических субъектов; Уметь: осуществлять поиск и импорт данных, необходимых для решения конкретных задач	Используя сайт mfd.ru или другие доступные сайты, получите данные о котировках акций трех российских компаний и проведите первичную обработку этих данных с помощью табличного редактора.
	2.Демонстрирует владение профессиональными пакетами прикладных программ.	Знать: основные принципы программирования алгоритмов статистической обработки данных в табличных редакторах и R(RStudio) для анализа и прогноза внутри- и внешнеэкономических процессов Уметь: применять технические и инструментальные средства для решения прикладных задач управления и подготовки обоснований управленческих решений	Используя инструменты Описательная статистика и Регрессия в надстройке Анализ данных в Excel, найти уравнения прямых регрессии для данных вашего варианта расчетно-аналитической работы и дать экономическую интерпретацию полученных уравнений
	3.Выбирает необходимое прикладное программное обеспечение в зависимости от решаемой задачи.	Знать: тенденции развития новых программных продуктов и IT-сервисов, ориентированных на повышение эффективности сбора и обработки данных. Уметь: обосновывать выбор ПО в зависимости от решаемой задачи	Для сгенерированного массива данных постройте гистограммы и диаграммы размаха (для каждого из признаков), диаграммы рассеяния (для каждой пары признаков). Используйте в каждом случае одно или несколько программных средств, сделайте вывод об их сравнительных функциональных возможностях на основе

			проведенной визуализации
	4. Использует прикладное программное обеспечение для решения конкретных прикладных задач.	<p>Знать: основные принципы обработки, визуализации и анализа данных, вероятностные и статистические методы.</p> <p>Уметь: использовать инструменты описательной статистики и визуализации данных, вероятностные и статистические методы для решения организационно-управленческих задач, интерпретировать полученные результаты.</p>	Для данных, предложенных в вашем варианте задания, дайте точечные оценки средних значений, стандартных отклонений, эксцесса и коэффициента асимметрии. Интерпретируйте полученные результаты.
Способность осуществлять поиск, критически анализировать, обобщать и систематизировать информацию, использовать системный подход для решения поставленных задач (УК-10)	1. Четко описывает состав и структуру требуемых данных и информации, грамотно реализует процессы их сбора, обработки и интерпретации	<p>Знать: основные принципы предварительной обработки, визуализации и анализа данных, вероятностные и статистические методы.</p> <p>Уметь: использовать инструменты предварительной обработки и аккуратизации данных</p>	Для данных, полученных с сайта mfd.ru или других доступных сайтов, используя инструменты в надстройке Анализ данных в Excel, найти точечные оценки средних значений, стандартных отклонений, эксцесса и коэффициента асимметрии. Интерпретируйте полученные результаты.
	2. Обосновывает сущность происходящего, выявляет закономерности, понимает природу вариативности	<p>Знать: методы анализа тенденций развития экономических процессов</p> <p>Уметь: применять технические и инструментальные средства для оценки содержания экономической повестки дня</p>	По данным двумерной выборки, содержащей сведения об уровне среднемесячного дохода респондентов и доле расходов на образование в общей сумме расходов, оцените параметры линейного уравнения парной регрессии, в которой доля расходов – результативный признак. Как изменится доля расходов на образование при увеличении среднемесячного дохода на 1000 рублей?
	3. Формулирует признак классификации, выделяет соответствующие ему группы	<p>Знать: современные методы многомерной классификации</p> <p>Уметь: применять классификационные методы,</p>	Даны 2 группы многомерных объектов (предприятия, регионы и т.п.) Выбрав и обосновав наиболее подходящий

	<p>однородных «объектов», идентифицирует общие свойства элементов этих групп, оценивает полноту результатов классификации, показывает прикладное назначение классификационных групп.</p>	<p>выбирая адекватные задаче математические и инструментальные методы, интерпретировать полученные результаты</p>	<p>метод классификации, определите, к какой из двух групп относится объект X</p>
	<p>4. Грамотно, логично, аргументировано формирует собственные суждения и оценки. Отличает факты от мнений, интерпретаций, оценок и т. д. в рассуждениях других участников деятельности.</p>	<p>Знать: основные методы теории вероятностей, математической и прикладной статистики, используемые для подготовки аналитических решений, экспертных заключений и рекомендаций. Уметь: использовать вероятностные и статистические методы для подготовки аналитических решений, 5 экспертных заключений и рекомендаций</p>	<p>Дана трехмерная выборка показателей, характеризующих профессиональные качества выбранной группы работников отрасли. По данным выборки вычислите коэффициенты корреляции Пирсона для каждой пары признаков, проверьте гипотезы о незначимости каждого из коэффициентов. Сделайте вывод о наличии/отсутствии связи между ними.</p>
	<p>5. Аргументированно и логично представляет свою точку зрения посредством и на основе системного описания</p>	<p>Знать: процедуры формулировки и проверки статистических гипотез Уметь: проверять гипотезы и делать обоснованные заключения на основе проверки</p>	<p>По данным выборки, содержащей ответы на вопросы социологического исследования, оцените доверительные интервалы для истинной доли конкретного ответа (указывается в зависимости от выборки), на уровне значимости 0,05 проверьте, используя критерий согласия Пирсона, гипотезу о равномерном распределении ответов респондентов.</p>

Способность релевантно решаемым задачам использовать информационные ресурсы и информационнокоммуникационные технологии для достижения целей, связанных с профессиональной деятельностью, обучением, участием в жизни общества и других сферах жизни (УК-15)	1.Самостоятельно выбирает и использует цифровые средства общения, осуществляет поиск и/или создание контента в соответствии с целью взаимодействия, в том числе для организации совместной деятельности	Знать: современные информационные технологии анализа данных и особенности их применения для решения стандартных задач профессиональной деятельности; Уметь: осуществить выбор программных средств для анализа данных в исследованиях	Согласно исследованию, доля работников данной отрасли меняет место работы чаще, чем каждые 3 года, составляет в среднем 0,5%. Определите вероятность того, что из 1000 работников данной отрасли число меняющих место работы чаще, чем каждые 3 года, не превысит 10 человек. Расчеты проведите при помощи соответствующих функций Excel, табличного редактора, или R
	2.Владеет навыками организации взаимодействия и коммуникации с помощью информационных систем и/или цифровых сервисов и технологий	Знать: тенденции развития технологии сбора данных в сети Internet Уметь: осуществлять поиск, сбор и первичную подготовку данных для прикладных исследований	Вероятности сбоя по техническим причинам в работе мессенджеров А, В и С составляет соответственно 0,0005, 0,005 и 0,001. Определите вероятность того, что сообщение будет доставлено, считая равновероятными предположения о том, каким именно из трех перечисленных мессенджеров воспользуется клиент
	3.Осуществляет подбор и применение различных информационнокоммуникационных средств для решения образовательных и профессиональных задач.	Знать: особенности современных информационных технологий представления результатов исследования; Уметь: осуществить выбор информационной технологии для анализа данных и презентации результатов анализа	При исследовании корреляционной зависимости между уровнем текучести кадров на фирме $Y(y_i)$ и объемом ипотечных кредитов работников $X(x_i)$ использовать имеющиеся данные, которые следует проанализировать с помощью инструментов Excel и R и построить графики линейной регрессии.

Примерные вопросы для подготовки к зачету:

1. Данные в экономике. Типы признаков в экономике и управлении: интервальные, порядковые, ранговые, дихотомические.
2. Измерение центра распределения. Измерение разброса данных.
3. Выбросы и их обработка. Пропущенные значения и их обработка. Повторяющиеся строки и их обработка. Синтетические признаки.
4. Основы комбинаторики. Правила суммы и произведения. Перестановки, размещения и сочетания без повторений. Перестановки, размещения и сочетания с повторениями.
5. Определение вероятности. Случайные события, их виды. Операции над событиями как операции над множествами. Классическая вероятностная схема.
6. Схема геометрических вероятностей.
7. Статистическая вероятность.
8. Аксиоматическое построение теории вероятностей.
9. Теорема сложения вероятностей.
10. Условные вероятности. Независимость событий. Формула полной вероятности. Формула Байеса.
11. Последовательности испытаний. Биномиальная схема. Формула Бернулли. Последовательности испытаний в экономике и управлении.
12. Определение случайной величины. Понятие случайной величины.
13. Функция распределения случайной величины. Свойства функции распределения.
14. Индикатор события как простейшая случайная величина. Функция распределения индикатора события.

15. Дискретные случайные величины и их важнейшие числовые характеристики. Дискретная случайная величина. Ряд распределения и функция распределения дискретной случайной величины.

16. Математическое ожидание, дисперсия и среднее квадратическое отклонение дискретной случайной величины.

17. Биномиальный закон распределения.

18. Геометрический закон распределения.

19. Закон распределения Пуассона.

20. Простейший поток событий.

21. Гипергеометрический закон распределения.

22. Сравнение случайных величин: отношение предпочтения, ожидаемая полезность, оптимальность по Парето.

23. Абсолютно непрерывные случайные величины и их важнейшие числовые характеристики. Абсолютно непрерывная случайная величина. Функция распределения и функция плотности распределения абсолютно непрерывной случайной величины. Свойства функции плотности распределения.

24. Математическое ожидание, дисперсия и среднее квадратическое отклонение абсолютно непрерывной случайной величины.

25. Равномерный закон распределения.

26. Показательный закон распределения.

27. Нормальный закон распределения.

28. Логарифмически нормальный закон распределения и ценообразование финансовых инструментов.

29. Закон распределения Парето и задачи налогообложения.

30. Законы распределения, важные в математической статистике (законы распределения Стьюдента, χ^2 , Фишера — Снедекора).

31. Смеси распределений.

32. Начальные и центральные моменты случайной величины. Асимметрия и эксцесс случайной величины.

33. Квантили и процентные точки случайной величины.

34. Ценность под риском.

35. Медиана и мода случайной величины.

36. Случайные векторы и условные законы распределения. Условный ряд распределения (для дискретных случайных величин), условная плотность распределения (для непрерывных случайных величин).

37. Условное математическое ожидание. Формула полного математического ожидания. Формула полной дисперсии.

38. Ковариация и коэффициент корреляции.

39. Портфель финансовых инструментов

40. Функции случайных величин. Функции одной случайной величины. Функции нескольких случайных величин. Формула композиции. Композиция равномерных случайных величин.

41. Закон больших чисел. Массовые случайные явления в экономике. Теорема Чебышёва и оценка математического ожидания. Теорема Бернулли и оценка вероятности. Обсуждение условий статистической устойчивости.

42. Центральная предельная теорема. Теорема Леви. Интегральная теорема Муавра — Лапласа. Место центральной предельной теоремы в изучении статистических закономерностей в экономике, финансах и управлении.

43. Математические основы теории страхования.

44. Метод Монте-Карло. Моделирование случайных величин.

Примерные вопросы для подготовки к экзамену:

1. Основы выборочного метода. Генеральная и выборочная совокупности. Случайная и конкретная выборки. Случайная повторная и случайная бесповторная выборка.

2. Предмет и задачи математической статистики.

3. Соотношение между предельной ошибкой выборки, уровнем значимости (риском) и объемом выборки. Использование этого соотношения в организации выборочных обследований.

4. Оценка плотности распределения и функции распределения. Вариационный ряд. Выборочная случайная величина (статистический ряд распределения). Интервальный вариационный ряд. Полигон частот, кумулята.

5. Оценка числовых характеристик генеральной случайной величины с помощью выборочной случайной величины. Выборочное среднее как оценка математического ожидания. Относительная частота как оценка вероятности. Выборочная дисперсия как оценка дисперсии. 6. Точечные оценки параметров. Понятие точечной оценки параметра генеральной совокупности. Свойства точечных оценок: состоятельность, несмещенность, эффективность.

7. Выборочное среднее как состоятельная, несмещенная и эффективная оценка математического ожидания генеральной случайной величины.

8. Смещенность выборочной дисперсии как оценки дисперсии генеральной случайной величины. Исправленная выборочная дисперсия как несмещенная и состоятельная оценка дисперсии генеральной случайной величины.

9. Методы построения точечных оценок: метод моментов, метод максимального правдоподобия. Примеры построения оценок параметров распределений случайных величин, применяемых в экономике и управлении.

10. Интервальные оценки параметров. Понятие интервальной оценки параметра генеральной совокупности. Точные интервальные оценки вероятности, математического ожидания, дисперсии и коэффициента корреляции. Поправка на конечный объем генеральной совокупности. Асимптотический подход к интервальному оцениванию.

11. Статистические гипотезы. Понятие статистической гипотезы. Виды статистических гипотез: параметрические и непараметрические, простые и сложные.

12. Критерий проверки гипотезы, критическое множество. Проверка гипотез с помощью интервальных оценок. Ошибки первого и второго родов. Мощность критерия. Наиболее мощный критерий.

13. Проверка гипотезы о равенстве математического ожидания теоретическому значению. Проверка гипотезы о равенстве двух математических ожиданий.

14. Проверка гипотезы о равенстве дисперсии теоретическому значению. Проверка гипотезы о равенстве двух дисперсий.

15. Проверка гипотезы о равенстве вероятности события теоретическому значению. Проверка гипотезы о равенстве двух вероятностей.

16. Проверка гипотез о значимости коэффициента корреляции.

17. Критерии согласия. Критерий согласия χ^2 Пирсона. Критерий χ^2 Пирсона при неизвестных параметрах распределения.

18. Критерий χ^2 для проверки однородности данных.

19. Задачи машинного обучения. Обучение с учителем и обучение без учителя. Классы задач машинного обучения: регрессия, классификация, кластерный анализ, поиск аномалий. Примеры задач машинного обучения в экономике, управлении и финансах.

20. Линейная регрессия. Постановка задачи регрессионного анализа. Парная линейная регрессия.

21. Множественная линейная регрессия. Точечный и интервальный прогноз по модели регрессии. Примеры задач регрессии в экономике. Понятие о гетероскедастичности и автокорреляции.

22. Классификация с обучением. Постановка задачи классификации с обучением. Логистическая регрессия. Понятие о деревьях решений. Кредитный скоринг.

23. Кластерный анализ и поиск аномалий. Постановка задачи кластерного анализа. Метод К-средних. Сегментирование потребителей. Понятие о методах машинного обучения в задачах поиска аномалий.

Зачет и экзамен проходят путем тестирования, в процессе которого студенты решают задачи в LMS Moodle. Варианты заданий генерируются автоматически в системе.

Пример экзаменационного билета

Пример билета приводится в аутентичной транскрипции системы Moodle постранично.

Пример билета для зачета (3 семестр)

Задача 1. (5 баллов)

Из предложенного ниже перечня необходимо выбрать только пары порядковых признаков:

прибыль за год (млн. руб.) и рейтинг в данной отрасли

рейтинг в данной отрасли и уровень экологичности производства (высокий, средний, низкий)

уровень экологичности производства (высокий, средний, низкий) и вложения в модернизацию оборудования (млн. руб.)

вложения в модернизацию оборудования (млн. руб.) и прибыль за год (млн. руб.)

прибыль за год (млн. руб.) и присутствие иностранного капитала (да - нет)

Задача 2. (5 баллов)

Двадцать компьютеров в бухгалтерии некоторой фирмы имели следующее количество сбоев за последний месяц: 2, 2, 3, 4, 4, 2, 3, 5, 4, 4, 3, 3, 4, 4, 4, 5, 3, 4, 5, 5. Найти моду, медиану и среднее значение этого ряда наблюдений, в ответе записать меньшее из них.

Задача 3. (5 баллов)

Исследовался признак X -количество обращений в день с жалобами на выполнение коммунальных услуг по городу в течение месяца, по результатам было получено среднее значение $\bar{x} = 13$ и дисперсия $S_x^2 = 4$. Какое преобразование признака X приводит к тому, что у нового признака будет нулевое среднее и стандартное отклонение, равное единице:

mini-max -преобразование $\frac{X-x_{min}}{x_{max}-x_{min}}$

логарифмирование

возведение в квадрат

Z-преобразование $\frac{X-4}{13}$

Z-преобразование $\frac{X-13}{2}$

Задача 4. (5 баллов)

Изучается интенсивность потока звонков в регистратуру поликлиники по выборке: 130; 140; 120; 175; 315; 300; 330; 302; 290; 112; 251; 220; 105; 95; 100. Выбрать все верные ответы:

среди перечисленных значений нет выбросов

значение 95 принадлежит межквартильному размаху IQR

значение 100 не принадлежит межквартильному размаху IQR

значение 290 является выбросом

значение 290 принадлежит межквартильному размаху IQR

Задача 5. (5 баллов)

Опыт состоит в том, что на отрезок длиной 40 см наугад вбрасывают точку. Определить вероятность того, что она окажется далее 5 см от середины отрезка:

Задача 6. (5 баллов)

Изучается возвратность кредитов банку. Известно, что банк разделил своих клиентов на две группы – имевших нарушения и имеющих хорошую кредитную историю – в отношении 1:9. Вероятность нарушения сроков погашения кредита для клиентов первой группы равна 0,4, а для клиентов второй группы – 0,001. Случайным образом выбранный клиент нарушил срок погашения кредита. Вероятность того, что он принадлежит второй группе, равна (в ответе оставить 3 знака после десятичной запятой):

Задача 7. (5 баллов)

Опыт состоит в том, что сотрудник социологической службы обзванивает население микрорайона до тех пор, пока не встретит абонента, согласного ответить на вопросы. Вероятность того, что абонент согласен отвечать, равна $p = 0,2$. Если случайная величина X – количество осуществлённых попыток сотрудника до первого согласия, тогда дисперсия величины X равна

Задача 8. (5 баллов)

Исследуется непрерывная случайная величина с заданной функцией плотности вероятности

$$f(x) = \begin{cases} 0, & x \leq -4 \\ \frac{1}{8} \cdot (x + 4), & -4 < x \leq 0 \\ 0, & x > 0 \end{cases}$$

тогда математическое ожидание $E(X)$ этой величины будет равно:

Задача 9. (5 баллов)

Исследуется непрерывная величина X , распределённая по нормальному закону:

$X \sim N(\mu = 3, \sigma^2 = 9)$. Тогда вероятность $P(|X - E(X)| < 2\sigma(X))$ равна:

Задача 10. (5 баллов)

Исследуется двумерная случайная величина (X, Y) , закон распределения которой имеет вид:

	$X = 0$	$X = 15$
$Y = -15$	0	0,3
$Y = 0$	0,1	0,3
$Y = 10$	0,2	0,1

Вычислить условную вероятность $P(X \neq 0 | Y = 0)$:

Задача 11. (5 баллов)

Анализируется невозвратность выданных банком кредитов. По опытным данным полученная оценка вероятности того, что клиент банка не возвращает кредит составила 0,08 для каждого клиента. Оценить вероятность того, из 1000 клиентов, взявших кредит, от 50 до 110 его не возвратят.

Задача 12. (5 баллов)

Анализируется доходность пакетов акций на рынке ценных бумаг.

Опытным путём получена оценка вероятности того, что пакет даёт доход владельцу, равная 0,38 (для каждого пакета). Оценить вероятность $P(A)$ события A : «Из 1800 пакетов акций дадут доход менее 700 пакетов».

Пример экзаменационного билета (4 семестр)**Задача 1. (5 баллов)**

При изучении признака X из генеральной совокупности с известной дисперсией, равной 25, были извлечены две повторные выборки объёмов 200 и 400 соответственно. Сравнить полученные в результате средние ошибки и выбрать меньшую из них.

Задача 2. (5 баллов)

Исследуется возраст покупателей современной бытовой техники по выборке:

Значения X(лет)	22	24	28	35	37	41	45
Количество покупателей(чел.)	7	14	19	44	54	37	25

Вычислить несмещённую оценку генеральной дисперсии признака
(в ответе оставить два десятичных знака после десятичной запятой).

Задача 3. (5 баллов)

По результатам повторной выборки объёма $n = 200$ получены выборочное среднее значение признака $\bar{X} = 94$ и его выборочная дисперсия $\sigma^2 = 67$. Тогда верхняя граница доверительного интервала, который с вероятностью 0,9 накрывает генеральное среднее, равна (в ответе оставить два знака после десятичной запятой):

Задача 4. (5 баллов)

Установить соответствие между принятым решением при статистической проверке и видом ошибки:

L1: Отвергнута изначально верная основная гипотеза

L2: Принята изначально неверная основная гипотеза

L3: Отвергнута изначально неверная основная гипотеза

R1: допущена ошибка первого рода

R2: допущена ошибка второго рода

R3: нет ошибки

R4: допущена ошибка и первого, и второго рода

Задача 5. (5 баллов)

Из генеральной совокупности извлечена выборка:

x_i	2-4	4-6	6-8	8-10	10-12	12-14	14-16	16-18	18-20	20-22
n_i	19	21	15	26	22	16	16	22	18	25

Определить критическое значение для формирования правосторонней критической области при проверке гипотезы на соответствие распределения признака в генеральной совокупности равномерному закону с применением критерия Пирсона на уровне значимости 0,05.

Задача 6. (5 баллов)

Анализируется эффективность новой технологии, для чего отобраны две группы рабочих: в первой группе численностью $n = 60$ человек, где применялась новая технология, выборочная средняя выработка составила $\bar{X} = 82$ (изделий), во второй группе численностью $m = 70$ человек, где технология осталась прежней, выборочная средняя составила $\bar{Y} = 77$ (изделий). Генеральные дисперсии выработки известны и равны $D(X) = 100$ и $D(Y) = 74$ соответственно.

На уровне значимости $\alpha = 0,01$ выяснить, влияет ли новая технология на среднюю производительность, считая генеральные совокупности нормально распределёнными (Использовать гипотезы $H_0: \bar{X}_0 = \bar{Y}_0$ при альтернативной $H_1: \bar{X}_0 > \bar{Y}_0$, где \bar{X}_0 – генеральное среднее выработки при применении новой технологии, \bar{Y}_0 – генеральное среднее выработки по прежней технологии). Тогда наблюдаемое значение статистики будет равно:

Задача 7. (5 баллов)

По извлеченной из нормальной генеральной совокупности случайной выборке объема $n = 250$ найдена исправленная выборочная дисперсия $s^2 = 2,17$. Требуется на уровне значимости $\alpha = 0,02$ проверить нулевую гипотезу $H_0: \sigma^2 = 2,21$ при альтернативной гипотезе $H_1: \sigma^2 < 2,21$. Тогда критическая область определяется как (выбрать правильный ответ):

$(-\infty; -205,34) \cup (303,84; +\infty)$

$(-\infty; -361,72) \cup (361,72; +\infty)$

$(303,84; +\infty)$

$(401,44; 401,44)$

$(-\infty; 205,34)$
 $(201,44; 210,44)$

Задача 8. (5 баллов)

В исследовании изучалась связь между заболеваемостью гриппом детей школьного возраста и приёмом витаминов этими детьми, получены данные:

не принимал витамины и не заболел- 5 школьников

не принимал витамины и заболел- 40 школьников

принимал витамины и не заболел-155 школьников

принимал витамины и заболел- 5 школьников

На уровне значимости $\alpha = 0,05$ с применением критерия χ^2 проверяется гипотеза о том, что приём школьниками витаминов связан с заболеваемостью гриппом; найти наблюдаемое значение критерия, в ответе оставить один знак после десятичной запятой.

Задача 9. (5 баллов)

Анализируется эффективность двух методов обучения студентов, для чего проведено тестирование по 30-бальной шкале в группе X и группе Y. Были получены следующие результаты:

X: 7, 15, 14, 11, 13, 10, 18,

Y: 10, 8, 16, 10, 19, 7, 15, 14, 29, 20, 15

С помощью критерия Манна-Уитни проверяется гипотеза о существовании различий в результатах при использовании разных методов обучения для уровня статистической значимости $\alpha = 0,05$. Вычислить наблюдаемое значение критерия.

Задача 10. (5 баллов)

Два специалиста А и В дегустировали 6 сортов продукции, в результате были получены следующие оценки (по 10-бальной шкале):

А: 3, 10, 5, 4, 2, 1,

В: 5, 9, 4, 3, 1, 8.

Вычислить коэффициент ранговой корреляции Кендалла (в ответе оставить один знак после десятичной запятой)

Задача 11. (5 баллов)

При анализе влияния изменения стоимости тарифа страхования X на выручку страховой компании Y были получены следующие результаты:

1) ковариация 49

2) дисперсии $D(X) = 100$ и $D(Y) = 25$

Вычислить коэффициент детерминации (в ответе оставить два знака после десятичной запятой).

Задача 12. (5 баллов)

При анализе корреляционной зависимости между сроком службы медицинского оборудования $Y(y_i)$ и затрат на его профилактическое обслуживание $X(x_i)$ получены дисперсии $D(X) = 169$ и $D(Y) = 100$, а также коэффициент в уравнении регрессии $\beta_{yx} = 0,44$. Установить соответствие между числовыми коэффициентами и их значениями:

L1: коэффициент линейной корреляции

L2: коэффициент детерминации

L3: ковариация

R1: 0,57

R2: 0,33

R3: 74,36

R4: -0,57

Приказ от 01.10.2024 №2187/о «Об утверждении Положения о проведении текущего контроля успеваемости и промежуточной аттестации студентов, обучающихся по образовательным программам высшего образования в Финансовом университете»

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

а) основная:

1. Соловьев, В. И. Анализ данных в экономике: теория вероятностей, прикладная статистика, обработка и визуализация данных в Microsoft Excel: учебник / В. И. Соловьев ; Финуниверситет. – Москва : Кнорус, 2019. - 498 с. - Текст : непосредственный. - То же. - 2025. - ЭБС BOOK.ru. — URL: <https://book.ru/book/955517> (дата обращения : 18.11.2024). — Текст : электронный.

2. Кремер, Н. Ш. Теория вероятностей и математическая статистика: учебник и практикум для вузов / Н. Ш. Кремер. — 5-е изд., перераб. и доп. — Москва : Юрайт, 2024. — 538 с. — ЭБС Юрайт. — URL: <https://urait.ru/bcode/541918> (дата обращения : 18.11.2024). - Текст : электронный.

б) дополнительная:

3. Браилов, А. В. Практикум для самостоятельной работы студентов по дисциплине "Теория вероятностей и математическая статистика" / А. В. Браилов, С. А. Зададаев, П. Е. Рябов. — Москва : Финуниверситет, 2014. — ЭБ Финуниверситета. — URL: <http://elib.fa.ru/rbook/praktikum.pdf/view> (дата обращения : 18.11.2024). — Текст : электронный.

4. Сборник задач по курсу "Математика в экономике". В 3 частях. Часть 3. Теория вероятностей: учебное пособие / А. В. Браилов, А. С. Солодовников ; под редакцией В. А. Бабайцева, В. Б. Гисина. – Москва : Финансы и статистика, 2013, 2017. - 125 с. – Текст : непосредственный.

5. Расчетно-аналитическая работа по математической статистике: практикум / Т. Г. Апалькова, В. И. Глебов, С. Я. Криволапов, К. Г. Левченко. – Москва : Финуниверситет, Департамент математики, 2021. – 126 с. - org.fa.ru. - URL: https://org.fa.ru/app/umm/tree?login=yes&_url=%2Fumm%2Ftree&documentId=acb641b2-6161-44a8-a063-7ee66eddd6e8 (дата обращения : 18.11.2024). – Текст : электронный.

6. Анализ данных. Часть 1: учебное пособие / А. В. Потемкин, М. Н. Фридман, И. И. Цыганок, И. М. Эйсымонт. – Москва : Финуниверситет, Департамент анализа данных, принятия решений и финансовых технологий, 2019. - 106 с. -

org.fa.ru. - URL: <https://org.fa.ru/app/umm/tree> (дата обращения : 18.11.2024). – Текст : электронный.

7. Зададаев, С. А. Применение R (RStudio) в математической статистике: учебное пособие / С. А. Зададаев, М. В. Солдаткина. – Москва : Финуниверситет, Департамент анализа данных, принятия решений и финансовых технологий, 2019. – 29 с. — org.fa.ru. - URL: https://org.fa.ru/app/umm/tree?login=yes&_url=%2Fumm (дата обращения : 18.11.2024). — Текст : электронный.

9.Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Информационно-образовательный портал Финансового университета при Правительстве Российской Федерации <http://org.fa.ru/>

2. Система дистанционного обучения Финансового Университета URL: <https://campus.fa.ru>

3. Личный кабинет обучающегося <https://org.fa.ru>

4. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>

5. Электронно-библиотечная система BOOK.RU <http://www.book.ru>

6. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>

7. Электронно-библиотечная система Znanium <http://www.znanium.com>

8. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>

9. Деловая онлайн-библиотека Alpina Digital <https://finunivers.alpinadigital.ru/>

10. Анализ данных <https://stepik.org/course/129/promo?search=2818469770>

10. Методические указания для обучающихся по освоению дисциплины

Студентам при подготовке следует использовать нормативные документы Финансового университета, Методические рекомендации по планированию и организации внеаудиторной самостоятельной работы студентов по образовательным программам ба-

калавриата и магистратуры в Финансовом университете, утвержденные приказом Финуниверситета от 11.05.2021 г. № 1040 (см. сайт Финансового Университета: на главной странице раздел «Наш университет»; далее «Единая правовая база Финуниверситета»), использовать методические рекомендации кафедры.

Самостоятельная работа студентов проходит аудиторно в виде вебинаров на платформе МТС- Линк (ранее — Webinar Group), <https://mts-link.ru/>, и внеаудиторно, в специально организованном для каждой группы на образовательной платформе MOODL электронном учебном комплексе ЭУК. Наполнение ЭУКа состоит из видеолекций, тестов для самоподготовки, текущих тестов, заданий домашнего творческого задания, вариантов расчетно-аналитической работы и видеозаписей прошедших вебинаров. Для организации самостоятельной работы служит учебно-тематический план изучения дисциплины. В этом плане указана тематика лекций, практических занятий, вопросы и задания для самостоятельного изучения. Домашние задания, тесты для самоподготовки и текущие задания следует выполнять регулярно при подготовке к практическим занятиям. При подготовке к вебинару целесообразно предварительно ознакомиться с содержанием лекции по соответствующей теме и выделить наиболее трудные вопросы. Во время вебинаров студенту следует участвовать в решениях задач, отвечать на вопросы преподавателя, самому задавать вопросы, быть вовлеченным в интерактивный процесс. После занятий следует провести работу с записью вебинара, разобраться в непонятых моментах, еще раз повторить решения задач, выполнить домашнее задание. Практические занятия (вебинары) структурно состоят из следующих компонент: 1) краткое повторение предыдущего материала и ответы на вопросы студентов, возникшие в процессе самостоятельной работы; 2) рассмотрение теоретических вопросов, связанных с текущим практическим занятием; 3) разбор методов выполнения практических заданий и решения задач; 4) корректировка заданий для самостоятельной работы студентов; 5) интерактивная форма – совместное решение со студентами задач, аналогичных разобранным.

Подготовка к выполнению домашнего творческого задания осуществляется при помощи пособий [8.7,8,10,11]

Подготовка к выполнению расчетно-аналитической работы осуществляется при помощи пособий [8.7,8,9]. Примерные задания приведены в п 6.2. настоящей программы. Для подготовки к процедурам промежуточной аттестации (зачет и экзамен) на странице курса «Анализ данных» на сайте campus.fa.ru размещены тесты для самоподготовки для самостоятельного выполнения. Тесты охватывают все вопросы, выносимые на контроль, разбиты по темам и позволяют, благодаря значительному количеству вариантов, закреплять наиболее сложные для каждого конкретного студента навыки.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

11.1. Комплект лицензионного программного обеспечения:

2. Операционные системы ASTRA LINUX или MS Windows 10. 32
2. Браузеры Microsoft Edge, Google Chrome, Firefox.
3. Пакет MS Office 2016 и выше.
4. Антивирус Kaspersky.

11.2. Современные профессиональные базы данных и информационные справочные системы:

1. Информационно-правовая система «Консультант Плюс».
2. Информационно-правовая система «Гарант». URL:
3. Электронная энциклопедия: URL: <http://ru.wikipedia.org/wiki/Wiki>
4. Система комплексного раскрытия информации «СКРИН». URL: <http://www.skrin.ru>

11.3. Сертифицированные программные и аппаратные средства защиты информации – не предусмотрено

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для проведения вебинаров необходим компьютер с постоянным подключением к сети Интернет и возможностью доступа к ресурсам и технологиям, указанным в параграфах 9 и 11.