

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

Кафедра математики и анализа данных
Факультет информационных технологий и анализа больших данных

Документ подписан усиленной неквалифицированной электронной подписью
Организация: Финансовый университет при Правительстве РФ
Утверждено: Проректор по учебной и методической работе Е.А. Каменева
Сертификат: eAZV7r97Mrmv4gqWfD9TqUxtRvsyUD+W
Дата: 03.10.2025 г.

П.Б. Лукьянов

Моделирование и анализ данных с помощью языка R

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки:

38.03.02 - Менеджмент,

Образовательная программа «Управление бизнесом»

Профиль:

«Отраслевой менеджмент»

Рекомендовано

*Факультет информационных технологий и анализа больших данных
(протокол № 05 от 18.02.2026 г.)*

Одобрено

*Кафедра математики и анализа данных
(протокол № 20 от 26.01.2026 г.)*

© Москва 2026

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	6
4.	Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	6
5.	Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	7
5.1.	Содержание дисциплины	7
5.2.	Учебно-тематический план	9
5.3.	Содержание семинаров	10
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	18
7.	Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине	20
8.	Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	29
9.	Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	29
10.	Методические указания для обучающихся по освоению дисциплины	30
11.	Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	30
12.	Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	31

1. Наименование дисциплины

«Моделирование и анализ данных с помощью языка R».

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

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПKN-10	Владение методами количественного и качественного анализа информации, а также навыками построения моделей, применяя для анализа, моделирования и поддержки принятия решений современные информационные технологии и программные средства, включая инструменты бизнес-аналитики, обработки и анализа данных	Использует методы получения информации, ее анализа для построения моделей и интерпретации результатов моделирования.	знать: основные форматы хранения данных и способы обращения к этим данным, знать методы статистической обработки и интерпретации результатов для оценки применимости различных моделей и результатов моделирования уметь: обращаться к внешним источникам данных с использованием различных функций языка R, уметь выполнять очистку, обработку и интерпретацию исходных данных, по результатам анализа уметь выбирать и строить соответствующую модель для выполнения моделирования и выявления связей между данными.
		Применяет приемы классификации и выбора подходящих измерительных шкал при описании организационных систем, происходящих в них процессов и явлений	знать: приемы классификации и оценивания различных наборов данных с целью использования их для достоверного описания процессов и явлений, знать меры и шкалы, используемые в анализе наборов данных различной природы уметь: определять структуру и метрики данных, формируемых в организационных системах, используя инструментальный язык R и функции классификации, уметь сопоставлять процессы и явления, происходящие в организационных системах, и особенности данных, формируемых в результате

		Использует навыки организации и проведения качественных и количественных исследований анализа информации, подготовки аналитических отчетов о состоянии и динамики развития рынков товаров и услуг.	функционирования таких систем. знать: основные требования, предъявляемые к проведению качественных и количественных исследований анализа информации в разрезе достоверности, полноты и репрезентативности при изучении рынков товаров и услуг, знать основные показатели, отражаемые в аналитических отчетах о состоянии и динамики развития рынков товаров и услуг. уметь: формулировать требования к качественным и количественным исследованиям, проводимым с целью оценки состояния и динамики развития рынков товаров и услуг в различных прикладных областях, уметь выполнять статистический анализ результатов исследований для выяснения тенденций и трендов развития рынков товаров и услуг.
ПКН-2	Способность применять математические методы для решения стандартных профессиональных задач, интерпретировать полученные математические результаты	Демонстрирует знания математических методов, применяемых в менеджменте	знать: методы выполнения расчетов, используемые при анализе экономической информации, знать основные понятия теории вероятностей и математической статистики, знать вероятностные и статистические методы анализа данных, используемые в менеджменте уметь: использовать инструменты описательной статистики и визуализации данных, реализованные на языке R, применять вероятностные и статистические методы обработки информации для решения профессиональных задач менеджера
		Применяет математические методы и модели для обоснования принятия управленческих решений	знать: методы получения, обработки, анализа и моделирования данных, используемые в менеджменте для формирования обоснованных управленческих решений. уметь: обрабатывать и оценивать реальные данные с использованием статистических функций языка R, в

			том числе анализировать структуру и качество данных, выполнять статистическое исследование данных и на основе полученных результатов выполнять моделирование данных с целью анализа управленческих решений и повышения их эффективности.
		Содержательно интерпретирует результаты, полученные при использовании математических моделей	знать: особенности и неявные допущения при построении различных математических моделей, знать условия применимости моделей, и как следствие, корректно интерпретировать результаты, полученные при использовании математических моделей. уметь: применять на практике результаты, полученные при использовании математических моделей в части выявления неочевидных зависимостей и связей между данными, описывающими различные бизнес-процессы и использования полученных знаний для повышения эффективности управленческих решений.

3. Место дисциплины в структуре образовательной программы

Дисциплина «Моделирование и анализ данных с помощью языка R» относится к «Модуль "IT Skills"».

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

Очная форма обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 6 (в часах)
Общая трудоёмкость дисциплины	3/108	108
Контактная работа- Аудиторные занятия	50	50
Лекции	8	8
Семинары, практические занятия	42	42
Самостоятельная работа	58	58
Вид текущего контроля	Проектная работа	Проектная работа
Вид промежуточной аттестации	Зачет	Зачет

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. Основы языка R

1. Установка R и RStudio
2. Скрипты, переменные, функции
3. Типы данных. Векторы и векторные операции
4. Основные конструкции языка
5. Функции Пользователя

Тема 2. Визуализация данных на языке R

1. Базовая графика, пакет graphics
2. Интерактивная графика, пакет plotly

Тема 3. Получение, обработка и анализ данных

1. Генерация тестовых синтетических данных
2. Получение и обработка реальных данных
3. Характеристики наборов данных
4. Очистка и обработка данных
5. Основные понятия статистики и первичный анализ данных
6. Разведочный анализ данных

Тема 4. Моделирование данных

1. Моделирование данных: математический аппарат, возможности и ограничения
2. Классическая линейная регрессия
3. Полиномиальная регрессия и сплайны

4. Обобщенные линейные модели

5.2. Учебно-тематический план

№ п/п	Наименование тем(разделов) дисциплины	Трудоемкость в часах				
		Всего	Контактная работа - Аудиторная работа			Самостоя тельная работа
			Общая, в т.ч.:	Лекции	Семинары, практическ ие занятия	
1	Основы языка R	32	12	2	10	20
2	Визуализация данных на языке R	20	10	2	8	10
3	Получение, обработка и анализ данных	28	14	2	12	14
4	Моделирование данных	28	14	2	12	14
	Итого	108	50	8	42	58

5.3. Содержание семинаров

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарах, практических занятиях	Формы проведения занятий
Основы языка R	1. Установка R и RStudio. Работа с RStudio в облаке; описание консольного интерфейса и инструментов RStudio. 2. Скрипты, переменные, функции. Создание и работа со скриптами в формате R Notebook. Первые программы. Запуск и сохранение скриптов. Формирование html-отчетов с результатами работы скрипта. 3. Типы данных. Векторы и векторные операции. Простейшие типы данных. Создание констант и переменных различных типов, явное и неявное преобразование типов в алгебраических и логических операциях. Векторы и функции работы с векторами, правила обработки векторов разного типа и разной длины в вычислительных операциях. 4. Основные конструкции языка. Обработка данных в циклах for(), while(), repeat(). Условия выхода из цикла. Обработка данных по условию. Ограниченность конструкции if - else. Обработка векторов по условию. Функции ifelse(), switch(). 5. Функции Пользователя. Правила создания и вызова функций. Параметры и тело функции. Передача и обработка параметров, формальные имена параметров. Параметры по умолчанию. Формирование возвращаемого значения. Ключевое слово return.	Программирование и отладка, обсуждение и решение задач в интерактивной форме, самостоятельное решение заданий, проверка выполнения, разбор ошибок.
Визуализация данных на языке R	1. Базовая графика. Функция plot(). Визуализация данных и отображение математических функций. Пакет graphics, основные и вспомогательные функции. Последовательность формирования изображения. Графические параметры для управления изображением, примеры использования. Точки и линии, формирующие график функции; графическое исследование функций с анализом точек экстремумов. 2. Интерактивная графика. Пакет plotly и функция plot_ly(). Конвейерные операторы > и %>% . Послойное формирование изображения. Интерактивный характер изображения, изменение масштаба. Форматирование графика. Настройки координатных осей, фона, границ, легенды, заголовков и подписей, аннотаций.	Программирование и отладка, обсуждение и решение задач в интерактивной форме, самостоятельное решение заданий, проверка выполнения, разбор ошибок.
Получение, обработка	1. Генерация и представление синтетических	Программирование и

и анализ данных	данных. Цель и задачи генерации синтетических данных. Математические функции для генерации данных. Контейнеры для хранения данных. Сохранение данных в файлы различных форматов. Определение параметров, задающих вид тестовых данных. Построение итоговой функции генерации, управляемой параметрами. Визуализация результатов. 2. Получение реальных данных. Хранилища данных, СУБД. Форматы хранения данных: табличный формат CSV, универсальный формат JSON. Способы обращения к данным разных форматов. Представление внешних данных в структурах R. Понятие датафрейма, запись данных в датафреймы. Обращение к столбцам, строкам и отдельным ячейкам датафрейма. Сохранение датафрейма в файл. 3. Характеристики наборов данных. Числовые и категориальные данные. Непрерывные и дискретные числовые данные. Упорядоченные и номинальные данные. Данные наблюдений и экспериментальные данные. Организация данных в структуры. Генеральная совокупность и выборка. Размер выборки. Репрезентативность выборки. Полная и неполная информация. Особенности генерации и визуализации данных разной природы. 4. Очистка и обработка данных. Подготовка данных для анализа, удаление дубликатов, исправление ошибок. Преобразование данных, приведение к единым форматам. Выявление, обработка и анализ выбросов. Отсутствующие данные. Анализ значений NaN, NA, Null, удаление, заполнение медианой, средним, модой, специальным значением. Некорректные типы данных, преобразование строк в даты, категории, числа. Итоговая оценка качества данных. 5. Основные понятия статистики и первичный анализ данных. Закономерности, взаимосвязи. Меры центральной тенденции. Среднее, медиана, мода. Вариация измерений, случайная вариация. Меры вариации: дисперсия, размах, стандартное отклонение. Описательная и индуктивная статистика. Гистограммы, диаграммы с усами, диаграммы рассеяния. Корреляция как мера линейной зависимости. Причинно-следственные связи и корреляции. Коэффициент корреляции Пирсона. Проверка статистических гипотез, ошибки первого и второго рода, уровень значимости, р-значение. Распределения нормальное, хи-квадрат,	отладка, обсуждение и решение задач в интерактивной форме, самостоятельное решение заданий, проверка выполнения, разбор ошибок.
-----------------	--	---

	Стьюдента, Фишера. 6. Разведочный анализ данных. Загрузка данных из внешних источников: форматы CSV, Excel, базы данных. Анализ структуры данных, определение типов данных. Проверка базовой статистики: среднее, медиана, стандартное отклонение, минимум, максимум, квартили, IQR. Визуализации: ящик с усами <code>boxplot()</code>, диаграммы рассеяния. Категориальные данные и таблицы частот, столбчатые и круговые диаграммы, графики плотности. Анализ взаимосвязей между переменными. Связь категориальных и числовых данных. Связь числовых данных, диаграмма рассеяния, линия тренда. Коэффициенты корреляции. Связь категориальных данных, сводная таблица, тепловые карты. Формулирование выводов и гипотез. Циклы визуализация - статистика - очистка данных - визуализация.	
Моделирование данных	1. Моделирование данных: математический аппарат, возможности и ограничения. Описание явлений и процессов с помощью алгебраических функций. Точность приближения и минимизация отклонения от исходных значений. Аппроксимация, интерполяция и экстраполяция. Критерии качества: максимальная ошибка, среднеквадратичная ошибка, равномерное приближение по Чебышеву. 2. Классическая линейная регрессия. Прогнозирование непрерывной величины, геометрическая интерпретация. Математическая формула модели, метод наименьших квадратов. Функция <code>lm()</code>. Диагностика и меры качества модели. Коэффициент детерминации, стандартная ошибка регрессии. F-статистика. Ограничения классической модели. 3. Полиномиальная регрессия и сплайны. Расширение линейной регрессии для сложных зависимостей. Выбор оптимальной степени полинома. Критерии выбора степени. Узлы и степень сплайна. Линейные, кубические, естественные, сглаживающие сплайны. Гибкость, гладкость, интерпретируемость. 4. Обобщенные линейные модели (Generalized Linear Models, GLM). Проблема моделирования данных, не имеющих нормального распределения. Биноминальное, пуассоновское, гамма, обратное гауссовское распределения. Биноминальные данные и логистическая регрессия. Счетные данные и пуассоновская регрессия. Положительные	Программирование и отладка, обсуждение и решение задач в интерактивной форме, самостоятельное решение заданий, проверка выполнения, разбор ошибок.

	непрерывные данные и гамма-регрессия. Диагностика и проверка моделей.	
--	--	--

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

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

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Основы языка R	1. Установка R и RStudio на домашний компьютер. Создание учетной записи на сайте posit. cloud для работы с RStudio в браузерной версии. Выполнение расчетов в консоли, анализ конструкций %% и %/%. Очистка консоли и Global Environment. 2. Установка пакетов для работы со скриптами в формате Rmd, настройка заголовков файлов R Notebook для автоматического создания отчетов в формате html. Создание, сохранение и загрузка скриптов в папках Пользователя, отправка и получение скриптов. Изучение возможностей функций print(), cat(), paste() для вывода результатов. 3. Типы данных. Векторы и векторные операции. Создание векторов с помощью функций c(), vector(), append(). Явное преобразование векторов различных типов с помощью функций as. integer(), as. character(), as. numeric(), as. logical(). Выяснение приоритетов выполнения арифметических операций с векторами разных типов. Неявное преобразование типов в логических выражениях. 4. Основные конструкции языка. Написание программ генерации паролей через использование генератора случайных чисел и заданных условий. Отличия в работе функции ifelse() от конструкции if - else. Выполнение расчетов с накопительной суммой при вычислении в циклах for(), while(), repeat(). 5. Функции Пользователя. Способы отладки функций Пользователя. Создание функций Пользователя для вычислений средних, минимальных и максимальных значений в векторе, переданном в качестве параметра.	Разбор материалов лекции и семинарских занятий по пройденной теме, решение задач домашнего задания, изучение рекомендованных к занятию литературных источников и электронных ресурсов, подготовка дополнительных вопросов по пройденной теме.
Визуализация данных на языке R	1. Базовая графика. Функция plot(). Построение графиков с использованием функций abline(), points(), lines(). Параметры управления изображением pch, cex, lwd, lty, type, col, xlim, ylim. Визуализация данных и отображение	Разбор материалов лекции и семинарских занятий по пройденной теме, решение задач

	математических функций. Использование функции <code>plot()</code> для анализа вектора. Графический анализ данных, имеющих двумерное представление. Способы изменения масштаба двумерного изображения. 2. Интерактивная графика. Пакет <code>plotly</code> и функция <code>plot_ly()</code>. Установка пакета с использованием функции <code>install.packages()</code>. Настройка параметров <code>plot_ly()</code> <code>type</code>, <code>mode</code>, <code>marker</code>, <code>line</code>. Функции <code>add_lines()</code>, <code>add_markers()</code>. График по умолчанию. Использование панели управления масштабом графика. Использование конвейерных операторов <code> </code> <code>></code> при формировании графика. Функция <code>layout()</code> и работа со слоями, раскраска слоев. Параметры форматирования графика. Задание шрифта, цвета, размера элементов графика <code>plot_ly()</code>. Использование тегов <code>html</code> для задания точных характеристик параметров функций.	домашнего задания, изучение рекомендованных к занятию литературных источников и электронных ресурсов, подготовка дополнительных вопросов по пройденной теме.
Получение, обработка и анализ данных	1. Генерация синтетических данных с использованием математических функций <code>sample()</code>, <code>runif()</code>, <code>rnorm()</code>, <code>rbinom()</code>, <code>rgamma()</code>, <code>sin()</code>, <code>cos()</code>, <code>abs()</code>, и др. Генерация тестовых данных с использованием функций Пользователя. Создание генератора данных, моделирующих периодическую, случайную и линейную составляющие. Визуализация данных с использованием различных графических библиотек. Управление свойствами набора данных через управляющие параметры функции генерации. Сохранение данных в файлы <code>csv</code>, <code>xlsx</code>. 2. Получение и обработка реальных данных. Получение данных с сайтов для поиска данных Google Dataset Search, Kaggle, UCI Machine Learning Repository, а также хранилищ данных по российским организациям https://data.rcsi.science/. Функции <code>read.csv()</code>, <code>read.table()</code>, <code>read_excel()</code>. Визуализация структуры полученных данных в виде датафреймов и графиков. Функции <code>head()</code>, <code>tail()</code>, <code>str()</code>, <code>summary()</code>, <code>View()</code>. Сохранение выбранных диапазонов данных в файлы для последующего анализа. Функции <code>write.table()</code>, <code>write.csv()</code>. 3. Характеристики наборов данных. Получение информации о данных, функции <code>dim()</code>, <code>nrow()</code>, <code>ncol()</code>, <code>names()</code>, <code>colnames()</code>, <code>rownames()</code>, <code>class()</code>. Загрузка реальных данных из разных прикладных областей, определение типов данных и анализ структуры. Подсчет количества строк и столбцов, необходимые преобразования типов. Визуализация данных разных типов. Особенности	Разбор материалов лекции и семинарских занятий по пройденной теме, решение задач домашнего задания, изучение рекомендованных к занятию литературных источников и электронных ресурсов, подготовка дополнительных вопросов по пройденной теме.

	представления числовых и категориальных данных, непрерывных и дискретных числовых данных. 4. Очистка и обработка данных. Технологии обработки данных в датафреймах. Выявление и анализ выбросов с помощью <code>boxplot()</code>. Обработка значений <code>NaN</code>, <code>NA</code>, <code>Null</code>. Функции <code>is.na()</code>, <code>na.omit()</code>, <code>complete.cases()</code>. Параметр <code>na.rm = TRUE</code>. Подсчет количества повторений, функции <code>duplicated()</code>, <code>table()</code>. Фильтрация данных, функция <code>filter()</code>. Удаление дубликатов с условиями. 5. Основные понятия статистики и первичный анализ данных. Меры центральной тенденции <code>mean()</code>, <code>median()</code>, <code>quantile()</code>. Меры вариации, функции <code>var()</code>, <code>sd()</code>, <code>IQR()</code>, <code>range()</code>, <code>min()</code>, <code>max()</code>. Гистограммы, диаграммы с усами, диаграммы рассеяния. Корреляция как мера линейной зависимости. Функции <code>cor()</code>, <code>cor.test()</code>. 6. Разведочный анализ данных. Задачи выявления структуры, закономерностей, аномалий и взаимосвязей. Загрузка данных из <code>csv</code>, <code>Excel</code>. Проверка базовой статистики: среднее, медиана, стандартное отклонение, минимум, максимум, квартили, <code>IQR</code>. Визуализации, функции <code>boxplot()</code>, <code>hist()</code>, <code>plot()</code>. Категориальные данные и таблицы частот, функции <code>barplot()</code>, <code>pie()</code>, <code>pie3d()</code>. Анализ взаимосвязей между переменными.	
Моделирование данных	1. Моделирование данных: Общие вопросы моделирования: математический аппарат, возможности и ограничения. Описание поведения данных с помощью алгебраических функций. Шум данных. Точность приближения и минимизация отклонения от исходных значений. Аппроксимация, интерполяция и экстраполяция. Построение математических функций по реальным и синтетическим данным. Регрессия как частный случай аппроксимации. 2. Классическая линейная регрессия. Уравнение линейной регрессии. Расчет коэффициента наклона через ковариацию и дисперсию. Функции <code>lm()</code>, <code>summary()</code>, <code>str()</code>. Требование нормальности распределения. Визуализация результатов, функции <code>plot()</code>, <code>plot_ly()</code>. Проверка допущений модели, функции <code>bptest()</code>, <code>shapiro.test()</code>, <code>residuals()</code>. 3. Полиномиальная регрессия и сплайны. Квадратичная и кубическая регрессии. Функции <code>lm()</code>, <code>poly()</code>, <code>predict()</code>. Визуализация <code>plot()</code>, <code>plot_ly()</code>. Расширение линейной регрессии для нелинейных зависимостей. Выбор оптимальной степени	Разбор материалов лекции и семинарских занятий по пройденной теме, решение задач домашнего задания, изучение рекомендованных к занятию литературных источников и электронных ресурсов, подготовка дополнительных вопросов по пройденной теме.

	полинома. 4. Обобщенные линейные модели. Функция <code>glm()</code>. Биноминальное, пуассоновское, гамма, обратное гауссовское распределения. Анализ распределений с помощью функций <code>gaussian()</code>, <code>binomial()</code>, <code>poisson()</code>, <code>rpois()</code>, <code>Gamma()</code>, <code>inverse. gaussian()</code>. Визуализация моделей.	
--	---	--

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

1. Типы данных в R. Создание переменных разных типов
2. Контейнеры для хранения и обработки данных
3. Использование функции `ifelse()` с векторами
4. Использование функции `switch()`
5. Использование функции `sample()` для получения случайных векторов
6. Использование функции `runif()` для получения случайных векторов
7. Создание пользовательских функций
8. Параметры функции `plot()` для отображения точек и линий
9. Параметры функции `plot_ly()` для отображения точек и линий
10. Параметры функции `layout()` для оформления и настройки графика
11. Назначение и параметры функции `legend()`
12. Основные и вспомогательные графические функции
13. Графические функции и их параметры
14. Способы задания области отображения при анализе наборов данных
15. Экспорт данных из файлов CSV и `xlsx`
16. Доступ к данным через адресацию в датафреймах
17. Импорт результатов расчета в файлы CSV и `xlsx`
18. Функции для генерации синтетических данных
19. Основы выборочного метода. Генеральная совокупность и выборка
20. Основные характеристики распределения и их выборочные аналоги
21. Построение графика функции с использованием пакета `plotly`
22. Размещение на графике отразить нескольких наборов данных

23. Задание диапазона для представления данных по осям OX, OY
24. Размещение на графике поясняющей метки
25. Отличие параметров функций `lines()`, `points()` от параметров функции `plot()`
26. Возможности функции `ifelse()` при работе с векторами
27. Особенности использования функции `switch()` при целочисленных и строковых значениях параметра EXPR
28. Задание последовательности чисел на заданном интервале с указанием количества элементов последовательности
29. Получение набора случайных действительных чисел в заданном диапазоне с помощью функции `sample()`
30. Получение набора случайных целых чисел в заданном диапазоне с помощью функции `runif()`
31. Статистическая гипотеза. Статистические критерии

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по 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

ПКН-10 Владение методами количественного и качественного анализа информации, а также навыками построения моделей, применяя для анализа, моделирования и поддержки принятия решений современные информационные технологии и программные средства, включая инструменты бизнес-аналитики, обработки и анализа данных

1) Использует методы получения информации, ее анализа для построения моделей и интерпретации результатов моделирования.

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: основные форматы хранения данных и способы обращения к этим данным, знать методы статистической обработки и интерпретации результатов для оценки применимости различных моделей и результатов моделирования

Уметь: обращаться к внешним источникам данных с использованием различных функций языка R, уметь выполнять очистку, обработку и интерпретацию исходных данных, по результатам анализа уметь выбирать и строить соответствующую модель для выполнения моделирования и выявления связей между данными.

Типовые контрольные задания

Для выполнения задания использовать язык R и среду разработки RStudio.

1. По ссылке, заданной преподавателем, скачать набор реальных данных в формате CSV или `xlsx`,
2. Сохранить набор данных в структуру `dataframe`,
3. Выполнить очистку данных, в том числе: обработать выбросы, заменив их на медианные значения, исключить повторы, преобразовать строковый тип данных в числовой, факторный или логический тип, исходя из фактической информации, содержащейся в полученном датафрейме,

4. Выполнить анализ структуры данных и рассчитать показатели центральной тенденции и разброса,
5. Оценить форму распределения данных, исходя из формы распределения обосновать применимость или неприменимость для модели линейной регрессии,
6. В случае применимости линейной регрессии построить модель, выполнить визуализацию модели.

2) Применяет приемы классификации и выбора подходящих измерительных шкал при описании организационных систем, происходящих в них процессов и явлений

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: приемы классификации и оценивания различных наборов данных с целью использования их для достоверного описания процессов и явлений, знать меры и шкалы, используемые в анализе наборов данных различной природы

Уметь: определять структуру и метрики данных, формируемых в организационных системах, используя инструментальный язык R и функции классификации, уметь сопоставлять процессы и явления, происходящие в организационных системах, и особенности данных, формируемых в результате функционирования таких систем.

Типовые контрольные задания

Для выполнения задания использовать язык R и среду разработки RStudio.

1. По ссылке, заданной преподавателем, скачать набор реальных данных в формате CSV или xlsx,
2. Сохранить набор данных в структуру dataframe,
3. Выполнить очистку данных, в том числе: обработать выбросы, заменив их на медианные значения, исключить повторы, преобразовать строковый тип данных в числовой, факторный или логический тип, исходя из фактической информации, содержащейся в полученном датафрейме,
4. Выполнить классификацию набора данных: определить типы полученных данных, минимальное и максимальное значения для соответствующих числовых данных, определить размерности данных, выбрать шкалы для визуализации данных,
5. Выполнить визуализацию данных двумя способами, 1) используя пакет graphics и семейство функций plot() и др., и 2) пакет plotly и набор функций plot_ly() и др.
6. Обосновать выбор шкалы в каждом случае.

3) Использует навыки организации и проведения качественных и количественных исследований анализа информации, подготовки аналитических отчетов о состоянии и динамики развития рынков товаров и услуг.

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: основные требования, предъявляемые к проведению качественных и количественных исследований анализа информации в разрезе достоверности, полноты и репрезентативности при изучении рынков товаров и услуг, знать основные показатели, отражаемые в аналитических отчетах о состоянии и динамики развития рынков товаров и услуг.

Уметь: формулировать требования к качественным и количественным исследованиям, проводимым с целью оценки состояния и динамики развития рынков товаров и услуг в различных прикладных областях, уметь выполнять статистический анализ результатов исследований для выяснения тенденций и трендов развития рынков товаров и услуг.

Типовые контрольные задания

Для выполнения задания использовать язык R и среду разработки RStudio.

1. По ссылке, заданной преподавателем, скачать набор реальных данных в формате CSV или `xlsx`,
2. Сохранить набор данных в структуру `dataframe`,
3. Выполнить очистку данных, в том числе: обработать выбросы, заменив их на медианные значения, исключить повторы, преобразовать строковый тип данных в числовой, факторный или логический тип, исходя из фактической информации, содержащейся в полученном датафрейме,
4. Выполнить первичный статистический анализ данных, рассчитать показатели центральной тенденции и разброса, определить характер распределения данных,
5. Для полученных признаков выполнить визуализацию данных с использованием гистограмм, линейных графиков, ящиков с усами, диаграмм рассеяния и др.
6. Для каждому графику выполнить анализ и подготовить вывод, поясняющий и объясняющий поведение данных на графике,
7. Подготовить итоговый отчет, в котором на основании анализа признаков дать оценку качественным и количественным характеристикам датасета.

ПКН-2 Способность применять математические методы для решения стандартных профессиональных задач, интерпретировать полученные математические результаты

1) Демонстрирует знания математических методов, применяемых в менеджменте

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: методы выполнения расчетов, используемые при анализе экономической информации, знать основные понятия теории вероятностей и математической статистики, знать вероятностные и статистические методы анализа данных, используемые в менеджменте

Уметь: использовать инструменты описательной статистики и визуализации данных, реализованные на языке R, применять вероятностные и статистические методы обработки информации для решения профессиональных задач менеджера

Типовые контрольные задания

Для выполнения задания использовать язык R и среду разработки RStudio.

1. По ссылке, заданной преподавателем, скачать набор реальных данных в формате CSV или `xlsx`,
2. Выполнить обращение к датасету и скачивание альтернативным способом, для чего использовать функции доступа к данным языка R из разных пакетов, необходимые пакеты найти и установить в среду RStudio,
3. Сохранить набор данных в структуру `dataframe`,
4. Выполнить очистку данных, используя изученные методы обработки данных, в том числе: преобразование типов, анализ и обработку выбросов, нахождение и исключение дубликатов, замену пропущенных значений,
5. Используя различные статистические методы анализа данных, описать, оценить и охарактеризовать датасет с помощью основных показателей описательной статистики, выполнить расчет центральных тенденций и показателей вариации.
6. Используя методы визуализации данных, построить графики, отражающие характер датасета с помощью функций пакета `plotly`.

2) Применяет математические методы и модели для обоснования принятия управленческих решений

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: методы получения, обработки, анализа и моделирования данных, используемые в менеджменте для формирования обоснованных управленческих решений.

Уметь: обрабатывать и оценивать реальные данные с использованием статистических функций языка R, в том числе анализировать структуру и качество данных, выполнять статистическое исследование данных и на основе полученных результатов выполнять моделирование данных с целью анализа управленческих решений и повышения их эффективности.

Типовые контрольные задания

Для выполнения задания использовать язык R и среду разработки RStudio.

1. По ссылке, заданной преподавателем, скачать набор реальных данных в формате CSV или xlsx,
2. Сохранить набор данных в структуру dataframe,
3. Выполнить очистку данных, используя изученные методы обработки данных, в том числе: преобразование типов, анализ и обработку выбросов, нахождение и исключение дубликатов, замену пропущенных значений,
4. Используя различные статистические методы анализа данных, описать и охарактеризовать датасет с помощью основных показателей описательной статистики, выполнить расчет центральных тенденций и показателей вариации.
5. На основании показателей описательной статистики, выбрать допустимую модель для исследования различных связей между признаками,
6. Выполнить визуализацию модели с помощью функций пакета plotly,
7. На основании анализа показателей модели обосновать ее применимость или неприменимость для использования в подготовке управленческого решения.

3) Содержательно интерпретирует результаты, полученные при использовании математических моделей

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: особенности и неявные допущения при построении различных математических моделей, знать условия применимости моделей, и как следствие, корректно интерпретировать результаты, полученные при использовании математических моделей.

Уметь: применять на практике результаты, полученные при использовании математических моделей в части выявления неочевидных зависимостей и связей между данными, описывающими различные бизнес-процессы и использования полученных знаний для повышения эффективности управленческих решений.

Типовые контрольные задания

Для выполнения задания использовать язык R и среду разработки RStudio.

1. По ссылке, заданной преподавателем, скачать набор реальных данных в формате CSV или xlsx из некоторой прикладной области,
2. Сохранить набор данных в структуру dataframe,

3. Выполнить очистку данных, используя преобразование типов, анализ и обработку выбросов, нахождение и исключение дубликатов, замену пропущенных значений, выполнить оценку качества данных по шкале от 0 до 10, где 10 характеризует данные, не требующие очистки,
4. Используя функции статистического анализа данных, оценить и охарактеризовать датасет с помощью основных показателей описательной статистики, выполнить расчет центральных тенденций и показателей вариации. интерпретировать результаты с точки зрения полноты и нормальности распределения признаков.
5. Построить математическую модель, основанную на найденном распределении, оценить качество модели,
6. Выполнить визуализацию модели с помощью функций пакета `plotly`,
7. На основании анализа показателей модели выполнить содержательную интерпретацию результатов с точки зрения применимости модели для обоснованного принятия управленческого решения.

Примеры практико-ориентированных заданий

Задание 1. Дана выборка из нормального распределения с неизвестным средним и дисперсией: 2.160718; 3.151646; 3.100161; 2.253975; 2.699964; 2.540674; 3.671576; 4.089237; 2.271442; 3.651401; 2.947188; 1.622963; 2.470867; 3.322804; 3.101575; 1.290382; 3.406223; 3.458536; 2.786844; 2.504702

Построить центральные доверительные интервалы для среднего и дисперсии 95% уровня надёжности.

Задание 2. По заданной двумерной выборке сгруппировать данные по значениям атрибутивного признака, построить описательную статистику для каждой группы. Рассчитать выборочное среднее, стандартное отклонение, медиану, верхнюю и нижнюю границу нормы, количество выбросов. Построить для каждой группы гистограмму и ящик с усами.

Задание 3. По данным многомерной выборки, содержащей сведения о поле, возрасте, месте проживания и уровне доходов жителей региона N (приводятся данные) 1) проверить гипотезу о равенстве распределения мужчин и женщин по возрасту, используя критерий однородности Колмогорова-Смирнова 2) проверить гипотезу о равенстве математических ожиданий доходов респондентов, проживающих в городской и в сельской местностях

Задание 4. Преобразовать данные, полученные из файла в формате csv, в фрейм данных. Определить, сколько пропущенных значений (N/A) содержит признак, размещенный в столбцах фрейма. Определить объем многомерной выборки после удаления пропусков.

Задание 5. Дана трехмерная выборка экономических показателей предприятия. По данным выборки вычислить коэффициенты корреляции Пирсона для каждой пары признаков, проверить гипотезы о незначимости каждого из коэффициентов

Задание 6. По результатам опроса ответы респондентов на определенный вопрос анкеты представлены в виде выборки (задается выборка). Используя функции R, очистить выборку от пропусков, обозначенных как "NA", и определить объем исходной выборки, количество пропущенных данных "NA" в исходной выборке; определить объем очищенной от "NA" выборки; определить количество различных вариантов ответов респондентов, встречающиеся в очищенной выборке; определить количество респондентов, которые дали ответ (приводится вариант ответа).

Задание 7. Имеются данные о расстоянии x (км) и времени y (мин) доставки товара курьерами, данные представлены парами (x, y) :

(1.5 15), (2.0 18), (2.7 22), (3.2 25), (4.0 30), (0.8 12), (3.5 28), (2.2 19), (5.1 37), (4.8 33).

Построить диаграмму рассеяния, где по оси X расстояние, а по оси Y время доставки. Визуально оценить наличие или отсутствие линейной зависимости. Оценить знак корреляции. Найти уравнение линейной регрессии вида $y = a + b \cdot x$, где y - время доставки, x - расстояние до клиента. Рассчитать средние значения x и y , значения коэффициентов a , b . Объяснить смысл a , b . Используя модель определить, сколько минут займет доставка на расстояние 6 км, 9 км, 15 км? Сравнить значения, предсказанные моделью с реальными данными. Где ошибка модели максимальна и почему? Можно ли утверждать, что доставка на расстояние 3 км не превысит 25 минут?

Примерные вопросы для подготовки к зачету

1. Ошибки первого и второго рода; уровень значимости и мощность критерия; p -значение
2. Особенности нормального распределения, распределения хи-квадрат, распределения Стьюдента

3. Задание переменной в R. Типы данных в R, особенности работы с ними
4. Создание функции Пользователя на языке R. Задание параметров по умолчанию для функции
5. Функции импорта данных из Excel в R
6. Функции импорта данных из R в Excel
7. Построение в R графика функции, гистограммы, ящика с усами, диаграммы рассеяния
8. Критерий проверки гипотезы о значении среднего для нормального распределения в случае известной и неизвестной дисперсии
9. Критерий проверки гипотезы о значении дисперсии для нормального распределения в случае известного и неизвестного среднего
10. Критерий проверки гипотезы о равенстве средних для двух нормальных распределений в зависимости от их дисперсии
11. Критерий проверки гипотезы о равенстве дисперсий двух нормальных распределений
12. Смысл коэффициента корреляции. Проверка гипотезы о некоррелированности двух величин
13. Проверка гипотезы о распределении, которому соответствует выборка, имеет заданный вид, с помощью критерия хи-квадрат
14. Проверка гипотезы о независимости признаков с помощью критерия хи-квадрат
15. Проверка гипотезы об однородности выборок с помощью критерия хи-квадрат
16. Тесты на нормальность распределения. Применение тестов в R
17. Проверка статистических гипотез, ошибки первого и второго рода, уровень значимости, р-значение
18. Моделирование статистических распределений в синтетических данных
19. Построение описательной статистики, и её показатели
20. Коэффициент корреляции как мера линейной связи случайных величин.
21. Выборочный коэффициент корреляции
22. Проверка гипотез о виде распределения. Критерий согласия Пирсона (хи-квадрат)
23. Проверка гипотезы о независимости признаков с помощью критерия Пирсона

24. Проверка гипотезы об однородности выборок
25. Графическая проверка нормальности распределения. График квантиль квантиль
26. Условия применимости классической линейной регрессионной модели. Анализ распределения
27. Свойства классической линейной регрессионной модели, построение и диагностика
28. Ограничения линейной регрессионной модели
29. Условия применимости полиномиальной регрессионной модели. Анализ распределения
30. Свойства полиномиальной регрессионной модели, построение и диагностика
31. Ограничения полиномиальной регрессионной модели
32. Условия применимости обобщенных линейных моделей. Анализ распределений
33. Свойства обобщенных линейных моделей, построение и диагностика
34. Ограничения обобщенных линейных моделей
35. Линейные, кубические, естественные, сглаживающие сплайны
36. Биномиальные данные и логистическая регрессия
37. Счетные данные и пуассоновская регрессия
38. Положительные непрерывные данные и гамма-регрессия

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

Основная литература:

1. Лукьянов, П. Б. Графический анализ на языке R [Электронный ресурс] : учебно-методическое пособие / Лукьянов П. Б. Москва : Прометей, 2022 48 с. Книга из коллекции Прометей - Экономика и менеджмент <https://e.lanbook.com/book/264971> ISBN 978-5-00172-311-0. [БИК ID: RU-LAN-BOOK-264971]

Дополнительная литература:

1. Апалькова, Тамара Геннадьевна Математическая статистика. Практикум : Учебное пособие / Финансовый университет при Правительстве Российской Федерации 1 Москва : ООО "Научно-издательский центр ИНФРА-М", 2025 254 с. (Высшее образование (Финансовый университет)) ВО - Магистратура <https://znanium.ru/catalog/document?id=464951> ISBN 978-5-16-017913-1 ISBN 978-5-16-110927-4 (электр. издание) . [БИК ID: RU\infra-m\znanium\bibl\2207576]

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

1. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
2. Образовательная платформа "ЮРАЙТ" <https://urait.ru/>
3. Электронно-библиотечная система издательства "Лань" <https://e.lanbook.com/>
4. Информационно-образовательный портал Финуниверситета: <https://org.fa.ru>
5. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf>)
6. Электронно-библиотечная система BOOK.RU <http://www.book.ru>

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 бакалавриата и магистратуры в Финансовом университете, утвержденные приказом Финансового университета от 11.05.2021 г. № 1040 (см. сайт Финансового Университета: на главной странице раздел «Наш университет»; далее «Единая правовая база Финансового университета»), использовать методические рекомендации Кафедры математики и анализа данных Факультета информационных технологий и анализа больших данных.

Методические указания для обучающихся по освоению дисциплины (комплекс рекомендаций и разъяснений, позволяющий студенту оптимальным образом организовать процесс изучения учебного материала дисциплины) представлены в Учебно-организационном комплексе для дисциплин Кафедры математики и анализа данных Факультета информационных технологий и анализа больших данных, размещенном на странице Кафедры математики и анализа данных сайта Финансового университета.

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

Комплект лицензионного программного обеспечения:

1. Язык R
2. Среда разработки RStudio

Современные профессиональные базы данных и информационные справочные системы:

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

Сертифицированные программные и аппаратные средства защиты информации:

1. не предусмотрены

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

1. Компьютерный класс для проведения учебных занятий, предусмотренных программой, в том числе групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенный оборудованием и техническими средствами обучения: мебель аудиторная (столы, стулья, доска аудиторная), персональные компьютеры, набор демонстрационного оборудования (проектор, экран)