

ГОУ ВПО РОССИЙСКО-АРМЯНСКИЙ (СЛАВЯНСКИЙ) УНИВЕРСИТЕТ

Составлен в соответствии с
государственными требованиями к
минимуму содержания и уровню
подготовки выпускников по
направлению “Менеджмент” и
Положением «Об УМКД РАУ».



УТВЕРЖДАЮ:
Директор института
Сандоян Э.М.

20 августа 2021 г.

Институт: Экономике и Бизнеса

**Кафедра: Математических методов и информационных
технологий в экономике и бизнесе**

Автор(ы): Маркосян Месроп Тигранович, преподаватель

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС

Дисциплина: Б1.В.05 Статистическое программирование в R

Для магистерских программ:

Магистерская программа: Бизнес-инжиниринг

Направление: Менеджмент (38.04.02)

Форма обучения: очная

ЕРЕВАН 2021

1. Аннотация

Язык программирования R является безусловным лидером среди свободно распространяемых систем статистического анализа. Ведущие университеты мира, аналитики крупнейших компаний и исследовательских центров регулярно используют эту программную среду при проведении научно-технических расчетов и создании крупных информационных проектов.

Широкое преподавание статистики на базе R и всемирная поддержка научным сообществом обусловили то, что приведение скриптов кода этого языка постепенно становится общепризнанным стандартом как в журнальных публикациях, так и при неформальном общении ученых всего мира.

Каждый день растет роль качественного статистического анализа данных. Результаты применяются в разных сферах и областях

- обнаружение аномалий, например, ненормальное поведение клиента, мошенничества;
- персонализированный маркетинг — электронные рассылки, ретаргетинг, системы рекомендаций;
- количественные прогнозы — показатели эффективности, качество рекламных кампаний и других мероприятий;
- скоринговые системы — обработка больших объемов данных, помощь в принятии решений, например, о предоставлении кредита;

Курс «Статистическое программирование в R» предназначен для повышения у обучающихся аналитических способностей для анализа данных с помощью R. Курс включает в себе знакомство с базовыми функциями и пакетами в R для обработки и анализа данных. Процесс анализа, в свою очередь, предполагает углубление в статистические вычисления и в регрессионный анализ. В ходе уроков студенты будут знакомиться с практическим применением статистики и анализа данных в финансах и экономике.

Содержание курса охватывает: основы программирования в R, обработка данных, описательная статистика и визуализация данных, проверка статистических гипотез, модели линейной регрессии, прогнозирование с помощью регрессионных моделей, построение скоринговой модели с помощью логистической регрессии.

Данный курс дает возможность использования R для статистики и анализа данных. А так же курс дает знания для практического применения статистики в разных сферах касающихся финансов и экономики.

Широкое преподавание статистики на базе R и всемерная поддержка научным сообществом обусловили то, что приведение скриптов кода этого языка постепенно становится общепризнанным стандартом как в журнальных публикациях, так и при неформальном общении ученых всего мира.

Общая трудоемкость дисциплины «Статистическое программирование в R» – 144 академических часов (4 академических кредита). Курс рассчитан на 36 часов практических занятий, а также на 108 часов самостоятельной работы студентов, которая будет заключаться в выполнении домашних заданий.

Проведение занятий предполагается в следующем формате: демонстрация материала, подготовленного преподавателем в виде R кода в RStudio. В ходе уроков студентам будут даны задачи для самостоятельной работы. Далее студентам после каждого занятия будут предоставляться задачи для домашнего задания.

Взаимосвязь с другими дисциплинами специальности: Дисциплина «Статистическое программирование в R» взаимосвязана с дисциплинами «Теория Вероятности и Статистика» и «Эконометрика».

Требования к исходным уровням знаний, умений и навыков студентов для прохождения дисциплины - для прохождения данной дисциплины студент должен:

- обладать базовыми знаниями и принципами в области статистики и теории вероятности
- владеть базовыми навыками работы с компьютером на уровне пользователя
- уметь пользоваться литературой и Интернетом для получения необходимой информации.

2. Учебная программа

2.1 Цели и задачи дисциплины

Цель дисциплины: сформировать у студентов системы знаний программирования с помощью R, умения сделать разные статистические вычисления, анализировать данные с помощью R и применять в практике.

Можно перечислить следующие цели курса:

- овладение языком программирования для анализа данных.
- обновления технологических навыков студентов.
- знакомство с инструментами анализа данных и их применениями в экономике и в финансах
- развития аналитического мышления
- развивать у студентов умения правильно постановить задачи в практике которые можно решать с помощью анализа данных
- умения ясно и детально объяснять результаты статистического анализа

2.2 Требования к уровню освоения содержания дисциплины

По окончании курса студент должен:

- ✓ овладеть базовыми функциями и инструментами в языке R.
- ✓ знать пакеты и инструменты в R для статистики
- ✓ уметь создавать новые функции для задачи анализа данных
- ✓ уметь пользоваться библиотеками и пакетами в R
- ✓ уметь проводить разные статистические вычисления с помощью R
- ✓ знать различные методы для анализа данных
- ✓ знать применение статистических моделей в практике
- ✓ уметь объяснять результаты анализов данных

Обладать:

- ✓ (ОК-3) - готовностью к саморазвитию, самореализации, использованию творческого потенциала

- ✓ (ОПК-3) - способностью проводить самостоятельные исследования, обосновывать актуальность и практическую значимость избранной темы научного исследования
- ✓ (ПК-4) - способностью использовать количественные и качественные методы для проведения прикладных исследований и управления бизнес-процессами, готовить аналитические материалы по результатам их применения
- ✓ (ПК-5) - владением методами экономического и стратегического анализа поведения экономических агентов и рынков в глобальной среде

2.3 Трудоемкость дисциплины и виды учебной работы (в академических часах и зачетных единицах) (см. Таблица 1).

Таблица 1. Трудоемкость дисциплин и видов учебной работы

Виды учебной работы	Всего, в акад. часах	Распределение по семестрам
		2
1	2	3
1. Общая трудоемкость изучения дисциплины по семестрам, в т. ч.:	144	144
1.1. Аудиторные занятия, в т. ч.:	36	36
1.1.1. Лекции	0	0
1.1.2. Практические занятия, в т. ч.	36	36
1.1.2.1. Обсуждение прикладных проектов		
1.1.2.2. Кейсы		
1.1.2.3. Деловые игры, тренинги		
1.1.2.4. Контрольные работы		
1.1.2.5. Другое		
1.2. Самостоятельная работа	108	108
Итоговый контроль (Экзамен, Зачет, диф. зачет - указать)	Экзамен	Экзамен

2.4 Содержание дисциплины, разделы дисциплины с указанием видов занятий (лекции, семинарские и практические занятия, лабораторные работы) и их трудоёмкость в академических часах (см. Таблица 2).

Таблица 2. Тематический план и трудоемкость аудиторных занятий (модули, разделы дисциплины и виды занятий) по учебному плану

Разделы и темы дисциплины	Всего (ак. часов)	Практ. занятия (ак. часов)
---------------------------	----------------------	-------------------------------

Раздел 1. Базовые навыки программирования в R		
Тема 1.1. Векторы и Матрицы	2	2
Тема 1.2. Дата фреймы и Списки	2	2
Тема 1.3. Циклы	2	2
Тема 1.4. Функции	2	2
Тема 1.5. Обработка данных при помощи dplyr	2	2
Тема 1.6. Визуализация данных	4	4
Раздел 2. Статистика в R		
Тема 2.1. Описательная статистика	2	2
Тема 2.2. Доверительные интервалы	2	2
Тема 2.3. Проверка статистических гипотез	2	2
Тема 2.4. Линейная регрессия	4	4
Тема 2.5. Логистическая Регрессия	2	2
Раздел 3. Применение в практике		
Тема 3.1. Оценка рыночной стоимости квартир с помощью методов регрессионного анализа	2	2
Тема 3.2. Построение Скоринговых моделей с помощью логистической регрессии	2	2
Тема 3.3. Рекомендательные системы	4	4
ИТОГО	36	36

Введение

Раздел 1. Базовые навыки программирования в R

В этом разделе студенты будут знакомиться с базовыми навыками программирования в R.

Тема 1.1. Векторы и Матрицы

В этой теме входят знакомство с векторами и матрицами,

Тема 1.2. Дата фреймы и Списки

Эта тема включает работу с данными из типа дата фреймов и списков, манипуляцию данных и разные действия с этими видами данных

Тема 1.3. Циклы

Эта тема включает изучения циклов (if,else,for,while)

Тема 1.4. Функции

В этой теме входит знакомство с синтаксом в R для составления собственных функций и их применение.

Тема 1.5. Обработка данных при помощи dplyr

Эта тема включает работу с обработкой данных, манипуляции данных с помощью библиотеки в R под названием dplyr.

Тема 1.6. Визуализация данных

Эта тема включает и использование основных инструментов встроенных в R, и использование Пакета в R под названием ggplot2.

Раздел 2. Статистика в R

В этом разделе студенты будут использовать инструменты R для статистических вычислений, построения линейных моделей и других моделей.

Тема 2.1. Описательная статистика

Эта тема включает знакомство с основными показателями статистики и их применение в практике с помощью R

Тема 2.2. Доверительные интервалы

В этой теме входит вычисление доверительных интервалов с помощью R

Тема 2.3. Проверка статистических гипотез

Эта тема включает знакомство с основными статистическими тестами, которые применяются для проверки статистических гипотез.

Тема 2.4. Линейная регрессия

В этой теме студенты будут использовать инструменты R для построения линейных регрессионных моделей.

Тема 2.5. Логистическая регрессия

В этой теме будут использовать инструменты R для построения логистических регрессионных моделей.

Раздел 3. Применение в практике

В этом разделе студенты будут знакомиться с применением статистики и анализа данных в финансах и экономике.

Тема 3.1. Оценка рыночной стоимости квартир с помощью методов регрессионного анализа

Эта тема включает практическое применение линейной регрессии для оценки рыночной стоимости квартир.

Тема 3.2. Построение Скоринговых моделей с помощью логистической регрессии

Эта тема включает практическое применение логистической регрессии для построения скоринговых моделей.

Тема 3.3. Рекомендательные системы

Эта тема включает знакомство с алгоритмами для рекомендательных систем и их практическое применение.

2.5 Материально-техническое обеспечение дисциплины

Материально-техническое обеспечение дисциплины представлено следующими компонентами:

- 1) Компьютерные аудитории, с установленными программами R и RStudio или же компьютеры с доступом в интернет и необходимым количеством памяти для установки программного обеспечения;
- 2) Аудитория с презентационным оборудованием;
- 3) Google-аккаунты;
- 4) Слайды, представляемые в течение практических занятий, составленные авторами;
- 5) Учебники и учебные пособия, находящиеся в свободном доступе и необходимом количестве в библиотеке РАУ;
- 6) Учебники и учебные пособия в электронном виде, предоставляемые студентам в начале курса.

2.6 Модульная структура дисциплины с распределением весов по формам контролей (см. таблица 3).

Таблица 3. Распределение весов по видам контролей

Формы контролей	Веса форм текущих контролей в результирующих оценках текущих контролей		Веса оценок промежуточных контролей и результирующих оценок текущих контролей в итоговых оценках промежуточных контролей		Веса итоговых оценок промежуточных контролей в результирующей оценке промежуточных контролей	
	M1	M2	M1	M2		
Вид учебной работы/контроля						
Контрольная работа (работа над индивидуальным кейсом)			1	1		

Весы результирующих оценок текущих контролей в итоговых оценках промежуточных контролей						
Вес итоговой оценки 1-го промежуточного контроля в результирующей оценке промежуточных контролей					0,5	
Вес итоговой оценки 2-го промежуточного контроля в результирующей оценке промежуточных контролей					0,5	
Вес результирующей оценки промежуточных контролей в результирующей оценке итогового контроля						0,5
Экзамен/зачет (оценка итогового контроля)						Экзамен 0,5
	$\Sigma = 1$					

3 и 4. Теоретический и практический блоки

Учебно-методическое обеспечение дисциплины

Рекомендуемая литература:

1. Наглядная статистика. Используем R! - А. Б. Шипунов, Е. М. Балдин, П. А. Волкова, А. И. Коробейников, С. А. Назарова, С. В. Петров, В. Г. Суфиянов
2. R в действии. Анализ и визуализация данных в программе R. – Роберт И. Кабаков.
3. Искусство программирования на R. Погружение в большие данные - Мэтлофф Норман
4. R Cookbook – Пол Титор
5. The Book of R: A First Course in Programming and Statistics - Тилман Мю Девис

Дополнительная литература:

1. Эконометрика, учебник - Елисеева И.И., Курышева С.В., Костеева
2. Временные ряды Финансы и статистика - Кендэл М.
3. Как правильно пользоваться статистикой. - Кимбл Г.

5. Блок ОДС и КИМ

Образцы заданий промежуточных и итоговых контролей

Итоговый контроль

Вариант билета для экзамена состоит из четырех задач. Задачи должны решаться на компьютерах используя язык программирования R

Пример варианта

- 1) Для случайно сгенерируемой выборки 20 чисел из нормального распределения с параметрами $\mu=0$ и $\sigma=1$, вычислить медиану, среднее значение и стандартное отклонение
- 2) Создать функцию, которая будет возвращать то же число что задано если задано положительное число и 0 когда задано отрицательное число.
- 3) Используя данные AirPassengers проверять гипотезу, что среднее значение количества авиапассажиров в месяц равно 250.000 , объяснить результаты теста.
- 4) В данных mtcars
 - а) используя функции из пакета dplyr выбрать те данные где параметр carb > 2 и disp > 160
 - б) построить корреляционную матрицу и визуализировать результаты
 - в) сделать линейную регрессию mpg зависимо от cyl. Напечатать R квадрат регрессионной модели и значимость объясняющего переменного cyl и объяснить результаты регрессионной модели.