

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное учреждение
высшего профессионального образования
"Казанский (Приволжский) федеральный университет"
Институт вычислительной математики и информационных технологий



УТВЕРЖДАЮ

Проректор
по образовательной деятельности КФУ
Проф. Таюрский Д.А.

"__" _____ 20__ г.

Программа дисциплины

Программирование на языке R Б1.В.ДВ.21

Направление подготовки: 01.03.02 - Прикладная математика и информатика

Профиль подготовки: Системное программирование

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Григорьева И.С.

Рецензент(ы):

Салимов Р.Ф.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Турилова Е. А.

Протокол заседания кафедры No ____ от " ____ " _____ 201__ г

Учебно-методическая комиссия Института вычислительной математики и информационных технологий:

Протокол заседания УМК No ____ от " ____ " _____ 201__ г

Регистрационный No

Казань
2017

Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) доцент, к.н. (доцент) Григорьева И.С. кафедра математической статистики отделение прикладной математики и информатики , Irina.Grigorieva@kpfu.ru

1. Цели освоения дисциплины

Цели освоения дисциплины "Программирование на языке R"

- ознакомить студентов с возможностями программирования в области статистики с использованием свободного ПО "Язык R";
- сформировать представление о возможностях языка R и сфере его применения;
- научить пользоваться специфическим векторным вариантом программирования;
- способствовать выработке навыков самостоятельного поиска информации (в том числе в на английском языке), использования справочных систем языка.

2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " Б1.В.ДВ.21 Дисциплины (модули)" основной образовательной программы 01.03.02 Прикладная математика и информатика и относится к дисциплинам по выбору. Осваивается на 3 курсе, 6 семестр.

Курс "Программирование на языке R" входит в вариативную часть подготовки бакалавра по направлению "Прикладная математика и информатика". Изучение дисциплины "Программирование на языке R" основывается на результатах изучения дисциплин, "Теория вероятности", "Математическая статистика", "Информатика".

Результаты подготовки по дисциплине "Программирование на языке R" могут быть использованы во всех дисциплинах прикладной направленности для моделирования изучаемых моделей.

Дисциплина изучается на 3 курсе.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

| Шифр компетенции | Расшифровка приобретаемой компетенции |
|---|--|
| ОПК-1 (профессиональные компетенции) | способность использовать базовые знания естественных наук, математики и информатики, основные факты, концепции, принципы теорий, связанных с прикладной математикой и информатикой |
| ОПК-2 (профессиональные компетенции) | способность приобретать новые научные и профессиональные знания, используя современные образовательные и информационные технологии |
| ПК-1 (профессиональные компетенции) | способность собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям |
| ПК-5 (профессиональные компетенции) | способность осуществлять целенаправленный поиск информации о новейших научных и технологических достижениях в информационно-телекоммуникационной сети ?Интернет? и в других источниках |
| ПК-7 (профессиональные компетенции) | способность к разработке и применению алгоритмических и программных решений в области системного и прикладного программного обеспечения |

В результате освоения дисциплины студент:

1. должен знать:

основные объекты, используемые в языке R и действия с ними;
способы ввода и вывода данных из/в файлы различных форматов;
набор задач, которые можно решать с помощью языка R;
способы обновления и расширения возможностей свободного ПО "R".

2. должен уметь:

применять векторный вариант объекто-ориентированного программирования;
решать с помощью R основные задачи статистической обработки данных;
подключать библиотеки программ для решения широкого класса статистических задач.

3. должен владеть:

постановки задач статистической обработки;
поиска в интернет обновлений и расширений языка, подключения их к работе.

4. должен демонстрировать способность и готовность:

использовать статистический язык R для решения статистических задач и задач обработки данных
совершенствовать навыки владения языком с использованием справочных и иных официальных материалов

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 3 зачетных(ые) единиц(ы) 108 часа(ов).

Форма промежуточного контроля дисциплины экзамен в 6 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

Тематический план дисциплины/модуля

| N | Раздел Дисциплины/ Модуля | Семестр | Неделя семестра | Виды и часы аудиторной работы, их трудоемкость (в часах) | | | Текущие формы контроля |
|----|---|---------|--------------------|---|-------------------------|------------------------|-----------------------------|
| | | | | Лекции | Практические занятия | Лабораторные работы | |
| 1. | Тема 1. Общее представление о задачах и особенностях среды R. Типы данных языка R. Объекты. Ввод данных с клавиатуры, с помощью таблицы, из файла. Организация обращения к элементам данных | 6 | 1-3 | 0 | 0 | 6 | Устный опрос |
| 2. | Тема 2. Ввод данных из разных источников. Вывод результатов расчета в файл. Сохранение и вызов программного кода | 6 | 4-5 | 0 | 0 | 4 | Письменное домашнее задание |
| 3. | Тема 3. Защита рефератов по основам языка R | 6 | 6-7 | 0 | 0 | 4 | Реферат |
| 4. | Тема 4. Простейшая графика (гистограммы, графики, точечные образы), вывод на экран и в файлы. Создание пользовательских функций и подпрограмм. | 6 | 7-8 | 0 | 0 | 4 | Устный опрос |
| 5. | Тема 5. Вычисление основных характеристик выборки. Вычисление результирующих величин для таблицы (по строкам и столбцам). Группа функций apply() | 6 | 9-10 | 0 | 0 | 4 | Контрольная работа |
| 6. | Тема 6. Визуализация данных. Использование метода главных компонент. Трехмерная графика. | 6 | 12-13 | 0 | 0 | 4 | Устный опрос |
| 7. | Тема 7. Применение полученных навыков: решение задачи по обработке данных | 6 | 11 | 0 | 0 | 2 | Творческое задание |

| N | Раздел Дисциплины/ Модуля | Семестр | Неделя семестра | Виды и часы аудиторной работы, их трудоемкость (в часах) | | | Текущие формы контроля |
|----|--|---------|-----------------|--|----------------------|---------------------|------------------------|
| | | | | Лекции | Практические занятия | Лабораторные работы | |
| 8. | Тема 8. Подключение внешних библиотек. Библиотека проверки нормальности Библиотеки преобразования данных | 6 | 15-16 | 0 | 0 | 4 | Устный опрос |
| 9. | Тема 9. Решение контрольных задач | 6 | 17-18 | 0 | 0 | 4 | Контрольная работа |
| | Тема . Итоговая форма контроля | 6 | | 0 | 0 | 0 | Экзамен |
| | Итого | | | 0 | 0 | 36 | |

4.2 Содержание дисциплины

Тема 1. Общее представление о задачах и особенностях среды R. Типы данных языка R. Объекты. Ввод данных с клавиатуры, с помощью таблицы, из файла. Организация обращения к элементам данных

лабораторная работа (6 часа(ов)):

Язык R как мощная современная среда статистической обработки данных. Принципы построения языка (интерпретируемый, язык командной строки, векторные вычисления). Использование справочных материалов и демонстрационных примеров. Встроенные данные. Векторы, их создание и использование. Массив (матрица) как особая форма вектора. Таблица (data.frame). Обращение к элементам/строкам/столбцам таблицы и матрицы. Применение основных вычислительных команд Объект список. Команда str() -- структура объектов Объект типа фактор (для категориальных данных). Упорядоченный фактор

Тема 2. Ввод данных из разных источников. Вывод результатов расчета в файл. Сохранение и вызов программного кода

лабораторная работа (4 часа(ов)):

Ввод данных с клавиатуры Ввод данных в таблицу (data.frame) из текстового файла или из файла типа .csv Вывод результатов на экран и в текстовый файл (команды print(), cat(), write()) и другие. Особенности вывода командами разных типов. Файл .Rhistory. Создание скриптов на его основе. Создание скриптов в текстовом файле и в редакторе для R. Настройка текущего каталога. Запуск скриптов командой source() и с помощью меню. Отладка скриптов.

Тема 3. Защита рефератов по основам языка R

лабораторная работа (4 часа(ов)):

Самостоятельное изучение некоторых разделов языка с последующим представлением их группе. Обсуждение результатов.

Тема 4. Простейшая графика (гистограммы, графики, точечные образы), вывод на экран и в файлы Создание пользовательских функций и подпрограмм.

лабораторная работа (4 часа(ов)):

Основные графические команды plot, lines, abline, points, text и другие. Графические параметры. Организация вывода в несколько окон. Вывод графической информации в файлы типа .pdf, .jpg, .png и т.п. Создание функций (подпрограмм) Синтаксис языка

Тема 5. Вычисление основных характеристик выборки. Вычисление результирующих величин для таблицы (по строкам и столбцам). Группа функций apply()

лабораторная работа (4 часа(ов)):

Вычисление среднего, дисперсии, медианы, квантилей, коэффициентов корреляции для объектов разного типа. Использование факторов (категориальных показателей) для группировки данных. Особенности применения команд к данным разного типа. Группа команд `apply()` - расширение возможностей для вычислений. Применение к строкам и столбцам. Формулы данных $x \sim y$. Простейшие графические представления данных: `boxplot`(ящик с усами), гистограмма.

Тема 6. Визуализация данных. Использование метода главных компонент. Трехмерная графика.

лабораторная работа (4 часа(ов)):

Дальнейшее изучение возможностей графических команд. Параметры, отвечающие за шрифты, цвета, значки и оформление графического поля. Создание нескольких графиков в одном окне вывода. Визуализация многомерных данных. Метод главных компонент. Возможности изображения трехмерных объектов.

Тема 7. Применение полученных навыков: решение задачи по обработке данных

лабораторная работа (2 часа(ов)):

Задача: построить выборку из случайной величины, равной сумме n равномерных величин. Проверить гипотезу нормальности этой величины (принцип больших чисел). Подготовить наборы данных разного размера и с разным законом распределения. Протестировать программу на этих данных.

Тема 8. Подключение внешних библиотек. Библиотека проверки нормальности. Библиотека преобразования данных

лабораторная работа (4 часа(ов)):

Расширение возможностей языка за счет подключения библиотек.

Тема 9. Решение контрольных задач

лабораторная работа (4 часа(ов)):

Создание собственных комплексов программ для статистической обработки данных и тестирование их на примерах.

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

| N | Раздел Дисциплины | Семестр | Неделя семестра | Виды самостоятельной работы студентов | Трудоемкость (в часах) | Формы контроля самостоятельной работы |
|----|---|---------|-----------------|---------------------------------------|------------------------|---------------------------------------|
| 1. | Тема 1. Общее представление о задачах и особенностях среды R. Типы данных языка R. Объекты. Ввод данных с клавиатуры, с помощью таблицы, из файла. Организация обращения к элементам данных | 6 | 1-3 | подготовка к устному опросу | 4 | устный опрос |
| 2. | Тема 2. Ввод данных из разных источников. Вывод результатов расчета в файл. Сохранение и вызов программного кода | 6 | 4-5 | подготовка домашнего задания | 3 | письменное домашнее задание |

| N | Раздел Дисциплины | Семестр | Неделя семестра | Виды самостоятельной работы студентов | Трудоемкость (в часах) | Формы контроля самостоятельной работы |
|----|--|---------|-----------------|---------------------------------------|------------------------|---------------------------------------|
| 3. | Тема 3. Защита рефератов по основам языка R | 6 | 6-7 | подготовка к реферату | 3 | реферат |
| 4. | Тема 4. Простейшая графика (гистограммы, графики, точечные образы), вывод на экран и в файлы Создание пользовательских функций и подпрограмм. | 6 | 7-8 | подготовка к устному опросу | 3 | устный опрос |
| 5. | Тема 5. Вычисление основных характеристик выборки. Вычисление результирующих величин для таблицы (по строкам и столбцам). Группа функций apply() | 6 | 9-10 | подготовка к контрольной работе | 3 | контрольная работа |
| 6. | Тема 6. Визуализация данных. Использование метода главных компонент. Трехмерная графика. | 6 | 12-13 | подготовка к устному опросу | 3 | устный опрос |
| 7. | Тема 7. Применение полученных навыков: решение задачи по обработке данных | 6 | 11 | подготовка к творческому заданию | 2 | творческое задание |
| 8. | Тема 8. Подключение внешних библиотек. Библиотека проверки нормальности Библиотеки преобразования данных | 6 | 15-16 | подготовка к устному опросу | 3 | устный опрос |
| 9. | Тема 9. Решение контрольных задач | 6 | 17-18 | подготовка к контрольной работе | 3 | контрольная работа |
| | Итого | | | | 27 | |

5. Образовательные технологии, включая интерактивные формы обучения

Обучение происходит в форме практических занятий в компьютерном классе.

На практическом занятии:

- преподавателем ставится задача, либо по самому языку, либо по его возможным приложениям;
- студенты ищут решение задачи с помощью справочных материалов языка или с помощью сведений из интернета;

- крупная задача, подводящая итог раздела, предлагается для самостоятельного решения вне аудиторного занятия;
- происходит защита решенного задания перед преподавателем;
- преподаватель проверяет решение, указывает на ошибки и при необходимости предлагает свое решение;
- некоторые из решенных задач обсуждаются совместно группой студентов.

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

Тема 1. Общее представление о задачах и особенностях среды R. Типы данных языка R. Объекты. Ввод данных с клавиатуры, с помощью таблицы, из файла. Организация обращения к элементам данных

устный опрос , примерные вопросы:

Простейшие формы объектов. Обращение к элементам объектов и заполнение их данными
Различие операторов [, [[, \$ Использование логических значений в качестве индексов Чем отличаются объекты "вектор" и "матрица"? "вектор" и "фактор"? " матрица" и "таблица"? Как их преобразовать их друг в друга? Что такое объект "список"? Какие примеры списков вы знаете

Тема 2. Ввод данных из разных источников. Вывод результатов расчета в файл. Сохранение и вызов программного кода

письменное домашнее задание , примерные вопросы:

Создать свой файл данных для последующей обработки. Организовать его ввод в среду R.
Создать скрипты по каждой выполненной задаче. В скриптах организовать вывод результатов в файлы

Тема 3. Защита рефератов по основам языка R

реферат , примерные темы:

1 Вывод графической информации (графиков, гистограмм). Команды plot, lines, points и т.п. 2 Вывод данных на экран и в файлы. Команды print, write и т.п. Управление расположением данных в файле, кавычками и др. 3 Вывод графической информации в файл. Команды png, jpeg, и т.п. Переключение канала вывода (device). Организация страницы вывода 4 Тип данных "дата". Функции от дат. 5 Обработка персональных данных (объекты класса person). Список авторов. 6 Действия с матрицами (как с алгебраическими объектами). 7 Обработка пропущенных данных. 8 Упорядочение векторов и таблиц.

Тема 4. Простейшая графика (гистограммы, графики, точечные образы), вывод на экран и в файлы Создание пользовательских функций и подпрограмм.

устный опрос , примерные вопросы:

Как организован вывод графики на экран? Чем отличаются команды plot и lines, points,...
Применение команды plot к данным разных типов

Тема 5. Вычисление основных характеристик выборки. Вычисление результирующих величин для таблицы (по строкам и столбцам). Группа функций apply()

контрольная работа , примерные вопросы:

Повторение материалов мат. статистики: выборка, параметры выборки, оценка параметров, доверительные интервалы

Тема 6. Визуализация данных. Использование метода главных компонент. Трехмерная графика.

устный опрос , примерные вопросы:

Повторение материалов математической статистики (факторный анализ, метод главных компонент)

Тема 7. Применение полученных навыков: решение задачи по обработке данных

творческое задание , примерные вопросы:

Повторение материалов теории вероятностей: закон больших чисел, центральная предельная теорема. Повторение материалов мат. статистики: нормальное распределение, проверка гипотезы нормальности

Тема 8. Подключение внешних библиотек. Библиотека проверки нормальности Библиотеки преобразования данных

устный опрос , примерные вопросы:

Изучение справочной системы R и поддерживающих сайтов

Тема 9. Решение контрольных задач

контрольная работа , примерные вопросы:

Повторение пройденного. Создание собственной библиотеки скриптов

Тема . Итоговая форма контроля

Примерные вопросы к экзамену:

Примерные контрольные вопросы для зачета

1. Какой объект для представления данных является основным в R? Чем от него отличаются остальные объекты?
2. Какие основные возможности для графического представления данных существуют в языке R?
3. Что такое векторные вычисления? Как заменить ими классические циклы?
4. Как происходит ввод и вывод различных объектов: исходных данных, результатов вычислений, графики, программных скриптов?
5. Какие типы случайных величин реализованы в языке R? Какие действия с ними можно производить?
6. Что такое метод apply и связанные с ним методы? Для чего они применяются?
7. Что такое библиотека? Как ее найти и подключить?

Задания для контрольной работы по статистическому языку R

- 1) Изучите список встроенных наборов данных командой `data()`. Определите структуру и тип данных объекта `cars`. Постройте график этой таблицы. Найдите уравнения регрессии и постройте их на том же графике.
- 2) Изучите список встроенных наборов данных командой `data()`. Определите структуру и тип данных объекта `ChickWeight`. Постройте график по двум первым столбцам так, чтобы цвет точек соответствовал значению в четвертом столбце. В качестве значка выберите звездочку.
- 3) Создайте две случайных величины (распределенные нормально и логнормально). Постройте их гистограммы на одном рисунке.
- 4) Исследуйте встроенный объект `ChickWeight`. Постройте средние значения для первых двух столбцов отдельно для каждого типа диеты одной командой. Сравните действие команд `tapply()` и `by()`. При необходимости преобразуйте тип данных (например, в `data.frame`).
- 5) Создайте в текстовом файле таблицу с пропущенными данными. Запрограммируйте замену пропущенных данных на среднее остальных данных по столбцам.
- 6) Постройте две выборки с помощью датчика равномерно распределенных чисел с разными параметрами. Проверьте гипотезу о сходстве их распределений с помощью критерия Вилкоксона (используйте автоматизированный тест).
- 7) Выберите какой-нибудь набор встроенных данных, содержащий как численные значения, так и пометки (тип данных "фактор"). Проведите исследование зависимости распределения численной величины от значения фактора: а) по критерию Фишера б) по критерию Вилкоксона. Сравните результаты.

Формы самостоятельной работы:

- 1) Студент готовит рефераты по предложенным темам.
- 2) Практические задания, не решенные в аудитории, решаются самостоятельно.
- 3) Все выполненные задания подготавливаются к защите перед преподавателем.

Темы практических заданий

- 1) Создание выборки из случайной величины. Нахождение ее основных характеристик.
 - 2) Ввод данных из файла. Объект "таблица". Вычисление итоговых значений по строкам и столбцам.
 - 3) Графическое представление данных. Гистограммы.
 - 4) Проверка гипотезы больших чисел.
 - 5) Решение систем неравенств графическим способом.
 - 6) Проверка нормальности выборки различными способами при помощи библиотеки "nortest"
- Темы рефератов:

- 1) Основные понятия и структуры языка R.
- 2) Библиотеки языка. Автоматизированное решение статистических задач.

7.1. Основная литература:

1. Боровков А.А. Математическая статистика.- СПб.: Лань, 2010. - 704 с.
ЭБС "Лань": http://e.lanbook.com/books/element.php?pl1_id=3810
2. Плохотников К.Э. Метод и искусство математического моделирования. - М.: ФЛИНТА, 2012. - 520 с.
ЭБС "Знаниум": <http://znanium.com/bookread.php?book=456334>
3. Алон Н., Спенсер Дж. Вероятностный метод. - М.: БИНОМ. Лаборатория знаний, 2015. - 323 с.
ЭБС "Лань": <http://e.lanbook.com/view/book/70705/>
4. Лагутин М.Б. Наглядная математическая статистика. - М.: БИНОМ. Лаборатория знаний, 2013. - 472 с.
ЭБС "Лань": <http://e.lanbook.com/view/book/56887/>
5. Свешников А.А. Прикладные методы теории вероятностей.- СПб.: Лань, 2012. - 480 с.
ЭБС "Лань": http://e.lanbook.com/books/element.php?pl1_id=3184

7.2. Дополнительная литература:

1. Бородин А.Н. Элементарный курс теории вероятностей и математической статистики. - СПб.: Лань, 2011. - 256с.
ЭБС "Лань": http://e.lanbook.com/books/element.php?pl1_id=2026
2. Сулейманов Р.Р. Компьютерное моделирование математических задач. - М.: БИНОМ. Лаборатория знаний, 2012. - 381 с.
ЭБС "Лань": <http://e.lanbook.com/view/book/4421/>
3. Королев В.Ю., Бенинг В.Е., Шоргин С.Я. Математические основы теории риска. - М.: Физматлит, 2011.- 620 с.
ЭБС "Лань": <http://e.lanbook.com/view/book/2742/>
4. Партыка Т.Л., Попов И.И. Математические методы. - М.: ИНФРА-М, 2007. - 464 с.
ЭБС "Знаниум": <http://znanium.com/bookread.php?book=137102>

7.3. Интернет-ресурсы:

Д.Мертц, Б.Хантинг. Статистическое программирование на R. Часть 1. Покупаемся в изобилии статистических возможностей. - <http://www.ibm.com/developerworks/ru/library/l-r1/index.html>

Д.Мертц, Б.Хантинг. Статистическое программирование на R. Часть 2. Функциональное программирование и анализ данных. -

<http://www.ibm.com/developerworks/ru/library/l-r3/index.html>.

Д.Мертц, Б.Хантинг. Статистическое программирование на R. Часть 3. Повторное использование кода и объектное программирование. -

<http://www.ibm.com/developerworks/ru/library/l-r3/index.html>.

Зеркало для скачивания обновлений - CRAN <http://cran.r-project.org/>

Поиск в материалах по R - R <http://finzi.psych.upenn.edu/nmz.html>

1. Сайт проекта - <http://www.r-project.org/>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Программирование на языке R" предполагает использование следующего материально-технического обеспечения:

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Все компьютеры подключены к корпоративной компьютерной сети КФУ и находятся в едином домене.

Компьютерный класс с установленным программным обеспечением для языка R

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 01.03.02 "Прикладная математика и информатика" и профилю подготовки Системное программирование .

Автор(ы):

Григорьева И.С. _____

"__" _____ 201__ г.

Рецензент(ы):

Салимов Р.Ф. _____

"__" _____ 201__ г.