

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 в среде R Б1.В.ОД.2

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Григорьева И.С.

Рецензент(ы):

Халиуллин С.Г.

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
2016

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

Язык R - программная среда, предназначенная для обработки данных, широко используемая в научных и прикладных исследованиях. Владение этим языком необходимо каждому специалисту в области прикладной математики и информатики. Цель дисциплины - расширить навыки и умения во владении средой R, полученные на уровне бакалавриата

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

Данная учебная дисциплина включена в раздел " Б1.В.ОД.2 Дисциплины (модули)" основной образовательной программы 01.04.02 Прикладная математика и информатика и относится к обязательные дисциплины. Осваивается на 1 курсе, 1 семестр.

Курс языка R для магистров является продолжением аналогичного курса для бакалавров. Полученные навыки используются в курсах, связанных со статистическим и прочим анализом данных.

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-1 (общекультурные компетенции)	способность к абстрактному мышлению, анализу, синтезу
ОК-3 (общекультурные компетенции)	готовность к саморазвитию, самореализации, использованию творческого потенциала
ОПК-3 (профессиональные компетенции)	способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять своё научное мировоззрение
ОПК-4 (профессиональные компетенции)	способность использовать и применять углубленные знания в области прикладной математики и информатики
ПК-8 (профессиональные компетенции)	способность разрабатывать корпоративные стандарты и профили функциональной стандартизации приложений, систем, информационной инфраструктуры
ПК-9 (профессиональные компетенции)	способность к преподаванию математических дисциплин и информатики в образовательных организациях основного общего, среднего общего, среднего профессионального и высшего образования

В результате освоения дисциплины студент:

- Применять статистический язык R для анализа данных
- Самостоятельно следить за обновлением возможностей постоянно развивающейся среды R

- Создавать собственные программы и комплексы программ для прикладных задач

4. Структура и содержание дисциплины/ модуля

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

Форма промежуточного контроля дисциплины экзамен в 1 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. Основные объекты языка R (векторы, факторы, таблицы и списки). Векторные вычисления. Простейшие графические команды. Ввод данных из внешнего источника. Создание простых скриптов	1	1-3	0	0	6	домашнее задание
2.	Тема 2. Датчики случайных величин с разным распределением. Основные статистические задачи (вычисление параметров, доверительных интервалов, построение гистограмм, проверка гипотез) Создание банка программ для обработки табличных данных	1	4-6	0	0	6	контрольная точка

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
3.	Тема 3. Подключение библиотек. Методы проверки нормальности данных. Сравнительное исследование различных методов.	1	7-8	0	0	4	устный опрос
4.	Тема 4. Группа методов построения линейных моделей. Команда lm(). Построение линий регрессии	1	9-10	0	0	4	отчет
5.	Тема 5. Кластеризация данных. Методы k-means и иерархической кластеризации. Визуализация результатов кластеризации. 3d-объекты.	1	11-13	0	0	6	контрольная работа
6.	Тема 6. Применение полученных навыков для самостоятельного анализа данных	1	14	0	0	2	устный опрос
	Тема . Итоговая форма контроля	1		0	0	0	экзамен
	Итого			0	0	28	

4.2 Содержание дисциплины

Тема 1. Основные объекты языка R (векторы, факторы, таблицы и списки). Векторные вычисления. Простейшие графические команды. Ввод данных из внешнего источника. Создание простых скриптов

лабораторная работа (6 часа(ов)):

Повторение основных сведений о целях, особенностях, структуре и объектах статистической среды R. Вектор - базовый объект языка. Векторные вычисления - специфический метод организации работы с данными. Многомерные объекты (массивы, матрицы, таблицы), их особенности. Способы извлечения данных из многомерных объектов. Фактор - объект для обработки категориальных данных. Использование прямых команд для вычислений.

Использование команд группы apply(). Простейшие графические операторы (plot, lines, points, abline и т.п.) Построение графиков и скаттерплов. Первичная визуализация данных. Ввод данных из текстовых файлов и файлов типа .csv. Проверка результатов ввода, автоматический поиск ошибок. Создание скриптов, запуск их из консоли. Особенности синтаксиса команд на экране и в файле.

Тема 2. Датчики случайных величин с разным распределением. Основные статистические задачи (вычисление параметров, доверительных интервалов, построение гистограмм, проверка гипотез) Создание банка программ для обработки табличных данных

лабораторная работа (6 часа(ов)):

Создание модельных объектов с помощью датчиков случайных чисел, а также команд группы `as.*()` Вычисление основных параметров: среднее, дисперсия, стандартное отклонение, медиана. квантили, коэффициенты корреляции. Особенности применения соответствующих команд к данным разных типов. Доверительные интервалы для параметров. Проверка гипотез: команды группы `*.test()`. Структура входной и выходной информации. Создание собственных списков вывода. Создание подпрограммы для запуска произвольной команды проверки гипотезы и обработки результата. Вывод результатов в текстовый файл. Возможности языка R: использование имени команды в качестве фактического параметра процедуры. Использование текстовых строк для создания имен выводимых файлов.

Тема 3. Подключение библиотек. Методы проверки нормальности данных. Сравнительное исследование различных методов.

лабораторная работа (4 часа(ов)):

Библиотеки в R. Знакомство с набором библиотек по литературе и сайтам, в том числе официальному сайту R. Возможности по преобразованию данных. Библиотека проверки нормальности. Создание тестовых наборов данных для проверки. Создание скриптов, применяющих различные методы проверки к данным. Сравнение результатов такой проверки

Тема 4. Группа методов построения линейных моделей. Команда `lm()`. Построение линий регрессии

лабораторная работа (4 часа(ов)):

Команда `lm()` и создаваемый ею объект. Использование полученного объекта для поиска зависимостей в данных. Построение линий регрессии (линейной и нелинейной). Визуализация многомерных данных. Методы главных компонент.

Тема 5. Кластеризация данных. Методы k-means и иерархической кластеризации. Визуализация результатов кластеризации. 3d-объекты.

лабораторная работа (6 часа(ов)):

Изучение различных способов кластеризации данных. Создание тестовых (двумерных) наборов данных в различной структурой. Метод k-means разбиения на фиксированное число классов. Тестирование метода на заготовленных тестовых наборах. Влияние случайности и эвристичности метода. Метод иерархической кластеризации. Подготовка данных (вычисление матрицы расстояний командой `dist()`). Применение метода и визуализация полученного дерева. Выделение классов. Визуализация классов на скаттерплоте с использованием цвета и значков разного типа. Визуализация многомерных данных с помощью метода главных компонент.

Тема 6. Применение полученных навыков для самостоятельного анализа данных

лабораторная работа (2 часа(ов)):

Самостоятельное создание скриптов, их отладка и прогонка на тестовых данных.

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. Основные объекты языка R (векторы, факторы, таблицы и списки). Векторные вычисления. Простейшие графические команды. Ввод данных из внешнего источника. Создание простых					

скриптов

1	1-3	подготовка домашнего задания	10	домашнее задание
---	-----	------------------------------------	----	---------------------

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
5.	<p>Тема 2. Датчики случайных величин с разным распределением. Основные статистические задачи (вычисление параметров, доверительных интервалов, построение гистограмм, проверка гипотез)</p> <p>Работа в компьютерном классе</p> <p>Отчет перед одногруппниками о проделанной работе. Совместное обсуждение результатов</p>	1	4-6	подготовка к контрольной точке	8	контрольная точка
6.	<p>Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов</p>					
Тема 1.	<p>Тема 3. Подключение библиотек. Методы проверки нормальности данных. Сравнительное исследование различных методов</p> <p>Создание простых скриптов</p> <p>домашнее задание, примерные вопросы:</p> <p>Повторение сведений о языке R, полученных в ходе самостоятельного изучения</p> <p>Для магистрантов, не изучавших этот курс - самостоятельная подготовка</p>	1	7-8	подготовка к устному опросу	8	устный опрос
Тема 2.	<p>Тема 4. Группы статистических задач (вычисление параметров, доверительных интервалов, построение гистограмм, проверка гипотез) Создание банка программ для обработки табличных данных</p> <p>Команда lm(). Построение линейной регрессии</p> <p>Предоставление скриптов преподавателю для тестирования на наборах данных</p>	1	9-10	подготовка к отчету	8	отчет
Тема 3.	<p>Тема 5. Кластеризация данных. Методы k-means и иерархической кластеризации</p> <p>устный опрос, примерные вопросы:</p> <p>Проверка написанных студентами скриптов в индивидуальном порядке</p>	1	11-13	подготовка к контрольной работе	8	контрольная
Тема 4.	<p>Тема 6. Визуализация результатов кластеризации. 3d-объекты</p> <p>отчет, примерные вопросы:</p> <p>Отчет перед группой по результатам работы. Совместное обсуждение</p>	1	14	подготовка к устному опросу	2	устный опрос
Тема 5.	<p>Тема 6. Применение полученных навыков для самостоятельного анализа данных</p> <p>устный опрос, примерные вопросы:</p> <p>Индивидуальный опрос</p>					

Тема . Итоговая форма контроля

Примерные вопросы к экзамену:

Типы объектов языка R

ввод данных из внешних источников

Вывод данных в текстовые и графические файлы
Простейшая статистическая обработка данных
Проверка гипотез в языке R
Создание и использование подпрограмм
Визуализация данных. Дополнительные графические возможности
Использование линейных моделей. Регрессия
Кластеризация данных методом k-means
Иерархическая кластеризация

7.1. Основная литература:

1. Лагутин М.Б. Наглядная математическая статистика. - М.: БИНОМ. Лаборатория знаний, 2013. - 472 с.
<http://e.lanbook.com/view/book/56887/>
2. Свешников А.А. Прикладные методы теории вероятностей.- М.: Лань, 2012. - 480 с.
http://e.lanbook.com/books/element.php?pl1_id=3184
3. Алон Н., Спенсер Дж. Вероятностный метод. - М.: БИНОМ. Лаборатория знаний, 2015. - 323 с.
<http://e.lanbook.com/view/book/70705/>

7.2. Дополнительная литература:

1. А.Б.Шипунов, Е.М.Балдин, П.А.Волкова, А.И.Коробейников, С.А.Назарова, С.В.Петров, В.Г.Суфиянов. Наглядная статистика. Используем R! / Москва, ДМК-Пресс, 2012.- 298 с.
2. Теория вероятностей и математическая статистика: Учебник / Е.С. Кочетков, С.О. Смерчинская, В.В. Соколов. - 2-е изд., испр. и перераб. - М.: Форум: НИЦ ИНФРА-М, 2014. - 240 с.
<http://znanium.com/bookread.php?book=447828>
3. Бородин А.Н. Элементарный курс теории вероятностей и математической статистики. - СПб.: Лань, 2011. - 256 с.
http://e.lanbook.com/books/element.php?pl1_id=2026

7.3. Интернет-ресурсы:

CRAN - <http://cran.r-project.org/>
R-help - <https://stat.ethz.ch/pipermail/r-help>
Советы по использованию - http://zoonek2.free.fr/UNIX/48_R/all.html
Советы по использованию - <http://www.statmethods.net/index.html>
Советы по использованию R - <http://pj.freefaculty.org/R/Rtips.html>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Программирование в среде 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.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен студентам. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, УМК, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего профессионального образования (ФГОС ВПО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен студентам. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

Компьютеры, подключенные к интернету, на которых установлено программное обеспечение "Язык R"

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 01.04.02 "Прикладная математика и информатика" и магистерской программе Анализ данных и его приложения .

Автор(ы):

Григорьева И.С. _____

"__" _____ 201__ г.

Рецензент(ы):

Халиуллин С.Г. _____

"__" _____ 201__ г.