

Практическая работа №1: Основы работы с языком R

Цель работы

Изучить приёмы работы со средой RStudio, научиться создавать программы на языке R, формировать и обрабатывать с их помощью статистические данные, а также пользоваться контекстной помощью.

Основные теоретические положения

R – язык программирования для статистической обработки данных и работы с графикой, а также свободная программная среда вычислений с открытым исходным кодом в рамках проекта GNU. R широко используется как статистическое программное обеспечение для анализа данных и фактически стал стандартом для статистических программ. R поддерживает широкий спектр статистических и численных методов и обладает хорошей расширяемостью с помощью пакетов. Пакеты представляют собой библиотеки для работы специфических функций или специальных областей применения. В базовую поставку R включен основной набор пакетов, а всего по состоянию на 2017 год доступно более 11778 пакетов. Ещё одна особенность R – возможность создания качественной графики, которая может включать математические символы.

RStudio – свободная среда разработки программного обеспечения с открытым исходным кодом для языка программирования R, который предназначен для статистической обработки данных и работы с графикой. RStudio написана на языке программирования C++ и использует фреймворк Qt для графического интерфейса пользователя.

Постановка задачи

Используя средства языка R провести анализ таблицы данных «Социологический опрос», получить необходимые результаты, сделать выводы.

Порядок выполнения работы

1. Создать таблицу данных «Социологический опрос» из 6 случаев и 4 переменных с помощью конструкции `data.frame()`. Имена переменных задать следующие: `name` (ФИО), `sex` (Пол), `age` (Возраст), `education` (Образование), `income` (Доход).
2. Переменные Пол и Образование представить как факторные. Проверить, что эти переменные действительно факторные любым удобным способом с соответствующим сообщением в консоли.
3. Добавить в таблицу новую переменную `income_20` (Доход_20), значения которой равны Доход + премия в размере 20% от Дохода.

4. Рассчитать среднее значение Дохода и вывести его на экран.
5. Используя инструменты R построить столбчатую диаграмму для Возраста и Дохода.
6. Добавить в таблицу новую переменную `income_S` (Доход_C), значения которой равны Доход + Доход_20.
7. Добавить новую переменную `income_M` (Доход_M). Необходимо найти максимальный доход и для каждого респондента определить, сколько процентов от максимального дохода составляет его доход.
8. Рассчитать среднее значение дохода в зависимости от пола.
9. Удалить переменную `income_20`.
10. Составить таблицу 1000×3 , содержащую в качестве первой переменной значения $x_i \in [0, 2\pi]$ (x_i изменяется с постоянным шагом), в качестве второй – значения $\sin x_i$ и в качестве третьей – значения $\sin x_i + X$, где $X \sim \mathcal{N}(0, 0.005)$.
11. Создать новый dataframe, в котором будут сохранены только первые 100 строк и последние 50 строк таблицы из п. 10.
12. Используя инструменты R вывести первые 5 строчек датафрейма из п. 11.
13. Используя инструменты R вывести последние 5 строчек датафрейма из п. 11.
14. Нарисовать 2 синусоиды, используя данные из п. 10.

Таблица "Социологический опрос"

ФИО	Пол	Возраст	Образование	Доход
John Davis	Male	21	Secondary	125
Angela Williams	Female	23	Higher	345
Janice Markhammer	Female	40	-	50
Bullwinkle Moose	Male	35	Secondary	399
Mary Rayburn	Female	18	Secondary	136
Greg Knox	Male	20	Higher	120

Содержание отчёта

From:

<http://se.moevm.info/> - **se.moevm.info**

Permanent link:

http://se.moevm.info/doku.php/courses:data_analysis_and_interpretation:task1?rev=1605609906

Last update: **2022/12/10 09:08**