**Лабораторный практикум по R**

Отчет должен быть подготовлен в течение 2 недель с помощью программы RStudio и технологии автоматической генерации отчетов RMarkdown. Отчет готовится с помощью специального шаблона, где нужно ввести необходимый код на языке R. Пример выполненного отчета прилагается. Рекомендуемый объем отчета до 10 с.

# Материалы по практикуму

### Ресурсы

1. В Интернете

ссылка <http://rpubs.com/AlexNovopoltsev> содержит следующие страницы

1) примеры использования языка R по разделам  
 (Раздел 1 – Раздел 5);

2) пример выполненных заданий «ВКИАД – пример выполненных заданий».

### Темы разделов по анализу данных в R

1. **Введение в R.**Основы работы со средой R, базовые конструкции языка и типы данных.
2. **Описательная статистика и графический анализ.**

Вычисление статистик, графический анализ и моделирование.

1. **Анализ статистических зависимостей.**

Корреляция, таблицы сопряженности, регрессионный анализ.

1. **Анализ неоднородных данных**

Анализ аномальных наблюдений и проверка гипотез о равенстве средних.

1. **Классификация неоднородных данных**

Кластерный и дискриминантный анализ, деревья решений.

# Подготовка отчета

1. Открыть программу RStudio.
2. Узнать путь к текущей рабочей директории (ввести команду getwd() в консоли) и скопировать туда два файла:
   1. файл шаблона “lab1.Rmd”:
   2. файл с данными “QC#.txt”, где # – номер варианта (1–14) согласно порядковому номеру в списке подгруппы (**описание данных приведено ниже**).
3. Изменить настройку кодировки в RStudio: выбрать кодировку UTF-8 для пункта “Defaulttextencoding” в Tools->GlobalOptions->General (для чтения кириллицы в шаблоне);
4. Открыть файл шаблона“vkiad.Rmd” из текущей рабочей директории.
5. Исправить
   1. ФИО, курс, группу, специальность;
   2. ФИО преподавателя, ведущего практику;
   3. вариант задания.
6. Заполнить файл “.Rmd”, при этом
   1. вставлять код R в пустые блоки кода:

```{r}

```.

* 1. за пределами блоков с кодом вставлять комментарии и выводы по полученным результатам.

**Замечания по заполнению отчета**

1) с момента ввода первой команды в файле “lab1.Rmd” и последующих команд желательно выделять каждую команду (можно просто поставить курсор на строку с командой) и выполнить их в порядке следования в интерактивном режиме с помощью команды “Ctrl+Enter”.

Последнее удобно для пошаговой проверки правильности проводимого анализа до получения всего отчета в целом. При этом можно видеть вновь созданные переменные на вкладке “Environment”.

2) текст заданий в итоговый отчет не включать, а вместо заданий написать краткие комментарии (названия статистик, графиков, критериев и переменных) и выводы.

3) правила форматирования отчета можно посмотреть, выбрав в выпадающем списке с пометкой «?» “MarkdownQuickReference” (на вкладке “Help” появится описание языка разметки RMarkdown).

В частности, для выделения текста жирным шрифтом используется синтаксис \*\*Текст\*\*, а для того, чтобы в итоговом документе, был переход на новую строку, соседние абзацы в “.Rmd” должны разделяться пустой строкой, или предыдущий абзац должен оканчиваться хотя двумя пробелами.

После выполнения всех заданий (заполнения кодом Rвсех пустых блоков) скомпилировать итоговый отчет в файл Word, для чего вверхней части редактора в выпадающем списке выбрать “KnitWord” для генерации отчета.

Распечатать файл Word с отчетом листа формата A4.

# Описание данных

Данные для занятий включают 14 типовых выборок многомерных разнотипных данных. Каждая выборка включает по 100 значений и 9 признаков (переменных). Данные представлены в текстовых файлах с расширением \*.txt(файлы qc1 – qc14 – для вариантов 1–4 соответственно).

## Содержательная интерпретация данных

Данные в каждой выборке представляют собой результаты оценки качества (QualityControl) партии 100 единиц продукции (услуг) некоторой компании по 9 признакам.

Используемые для оценки качества характеристики имеют следующую интерпретацию:

1. количественные признаки с ожидаемым нормальным законом распределения (помечены в табл. 1 символом «н»);
2. компания-поставщик продукции (№пост – номер (код) поставщика);
3. вид обнаруженного дефекта в единице продукции (№вд – номер (код) вида дефекта);
4. компания изготовитель продукции (№изг – номер (код) изготовителя);
5. число некритических дефектов в единице продукции (символ «п»);
6. уровень комплектации и исполнения (символ «б», максимальный уровень – код 0).

Т а б л и ц а 1. **Описание вариантов тестовых данных**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Вариант** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **qc1** | н | б | п | №изг | №пост | н | н | н | №вд |
| **qc2** | н | б | п | №изг | №пост | н | н | н | №вд |
| **qc3** | н | б | п | №изг | №пост | н | н | н | №вд |
| **qc4** | н | б | п | №изг | №пост | н | н | н | №вд |
| **qc5** | н | б | п | №изг | №пост | н | н | н | №вд |
| **qc6** | н | б | п | №изг | №пост | н | н | н | №вд |
| **qc7** | б | п | №изг | №пост | н | н | н | н | №вд |
| **qc8** | б | п | №изг | №пост | н | н | н | н | №вд |
| **qc9** | н | н | б | п | №изг | №пост | н | н | №вд |
| **qc10** | н | н | б | п | №изг | №пост | н | н | №вд |
| **qc11** | н | н | б | п | №изг | №пост | н | н | №вд |
| **qc12** | н | н | б | п | №изг | №пост | н | н | №вд |
| **qc13** | н | н | б | п | №изг | №пост | н | н | №вд |
| **qc14** | н | н | б | п | №изг | №пост | н | н | №вд |

**Замечания:**

1) у разных вариантов при подготовке отчета переменные должны быть переименованы в том порядке, который указан в таблице;

# Предварительная настройкаR

Настройка средыRна своем компьютере (дополнительно можно посмотреть в **Разделе 1.Введение в R**).

1. **УстановкаR и RStudio:**
   1. <http://cran.rstudio.com>

<http://www.rstudio.com/products/RStudio/#Desk>

**Замечание.** По ссылкам **a**можно получить новейшие версии программ.

1. **Настройка авторизации для скачивания библиотек**

**(**при наличии прокси-сервера, если нет прокси-сервера – пропустить этот пункт)

*Для компьютеров БГУ*:

* 1. открыть RStudio;
  2. снятьопцию “Use Internet Explorer library/proxy for HTTP” в Tools->Global Options->Packages;
  3. перезапустить RStudio;
  4. набрать в консоли две команды:

Sys.setenv(http\_proxy="http://proxy.bsu:3128/")

Sys.setenv(http\_proxy\_user="ask")

**Замечание.** Пункт **d** требуется выполнять после каждого перезапуска RStudio или рабочей сессии R.

1. **Установка библиотек для RMarkdown**

Происходит при первом создании в RStudioфайла с расширением “.Rmd” (NewFile -> R Markdown)при условии выполнения предыдущих пунктов (пункт два не требуется в случае отсутствия прокси-сервера).При этом появится окно с запросом на авторизацию, куда надо ввести имя пользователя и пароль в сети БГУ (для перемещения курсора между полями использовать клавишу Tab).

1. **Установка дополнительных библиотек**

Для выполнения заданий требуется установка только одной дополнительной библиотеки с кодовым названием “e1071”, для чего необходимо ввести в консоли RStudioследующую команду:

install.packages(“e1071”)

1. **Настройка кодировки для поддержки кириллицы**

ПрименитькодировкуUTF-8 дляопции “Defaulttextencoding” вTools->GlobalOptions->General.

# Литература

## Анализ данных вR

1. ШипуновА.Б., БалдинЕ.М. АнализданныхсR.
2. Verzani, J. Simple R: Using R for Introductory Statistics / The CSI Math Department, 2001.
3. Yakir, B. Introduction to Statistical Thinking / The Hebrew University, 2011.

## R tutorials

1. <http://ww2.coastal.edu/kingw/statistics/R-tutorials/index.html>
2. <http://www.math.csi.cuny.edu/Statistics/R/simpleR/>

<http://www.statmethods.net>

## Теория по анализу данных и моделированию

1. Харин Ю.С., Малюгин В.И., Абрамович М.С. Математические и компьютерные основы статистического моделирования анализа данных. Минск: БГУ, 2008.
2. Имитационное и статистическое моделирование : практикум для студентов мат. и экон. спец. / В. И. Лобач, В. П. Кирлица, В. И. Малюгин, С. Н. Сталевская — Минск : БГУ, 2004. — 189 с. — Библиогр.: с. 188-189.