МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКО ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. Н.И. ЛОБАЧЕВСКОГО»**

Дзержинский филиал

**Кафедра** Прикладная информатика

**ОТЧЕТ**

**по дисциплине «Интеллектуальный анализ данных»**

**к лабораторной работе № 2**

**Тема «R»**

Выполнил:

Студент гр. № 2721Б1ПИ

Погодин Дмитрий Евгеньевич

Проверил:

Ст.преподаватель, к.т.н.

\_\_\_\_\_\_\_\_\_/Нажимов А.В./ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дзержинск 2024

Оглавление

[**Лабораторная работа №2** 2](#_Toc135058871)

[**Вывод** 7](#_Toc135058872)

# **Лабораторная работа №2**

**Метод главных компонент**

**Цель**

Построить график, импортировать библиотеки для построения, применить метод главных компонент на нескольких примерах.

**Ход выполнения работы**

1. Добавим библиотеки cluster, ggplot2, factoextra. Для этого в пакетах найдём их по поиску и загрузим

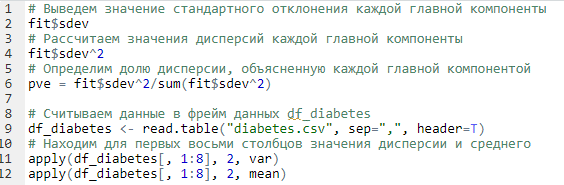


Рис.1. Код для переменных

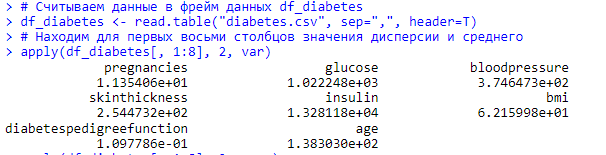


Рис. 2. Дисперсии для переменных набора данных df\_diabetes

По данным видно, что наибольшую дисперсию имеют переменные Insulin, Glucose и BloodPressure.

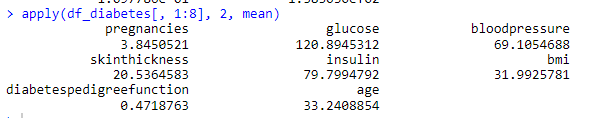


Рис. 3. Средние значения для переменных набора данных df\_diabetes

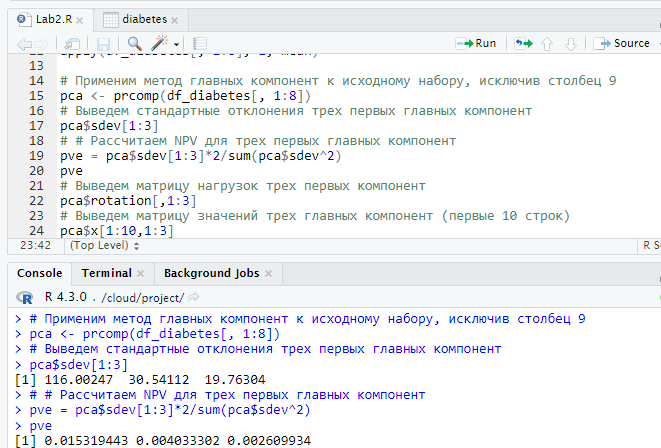


Рис.4. Значения стандартных отклонений трех главных компонент + расчёт NPV для трёх первых главных компонент

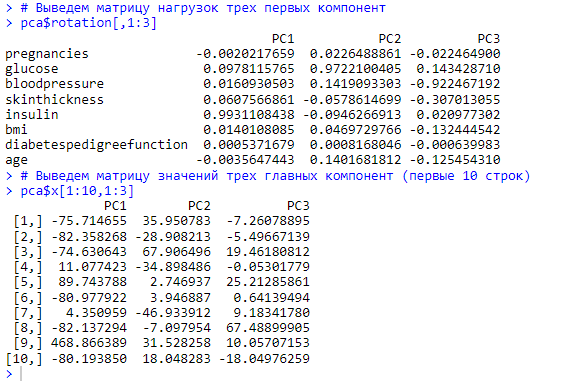


Рис.5. Матрица нагрузок первых трёх компонент + матрица значений трёх главных компонент

1. Стандартные отклонения главных компонент – убывающая последовательность, которая показывает, что каждая следующая компонента объясняет все меньшую долю диспресии наблюдений PVE:

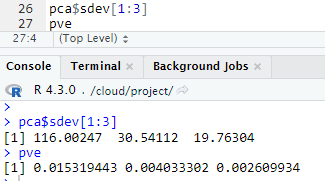


Рис.6. Средние значения для переменных набора данных df\_diabetes

1. Векторы нагрузок главных компонент показывают, как определяется каждая компонента через исходные переменные набора, а значения новых преобразованных переменных – главных компонент для всех наблюдений находится в матрице х результата

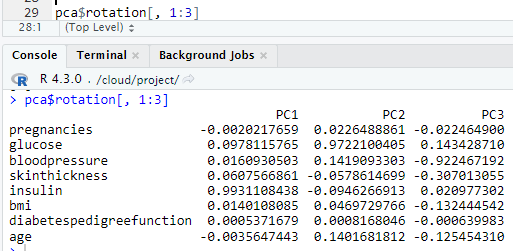


Рис.7. Векторы трёх главных компонент

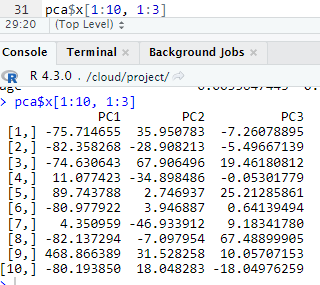


Рис.8. Первые 10 строк значений для первых трех главных компонент

1. Построить график «каменистой осыпи»

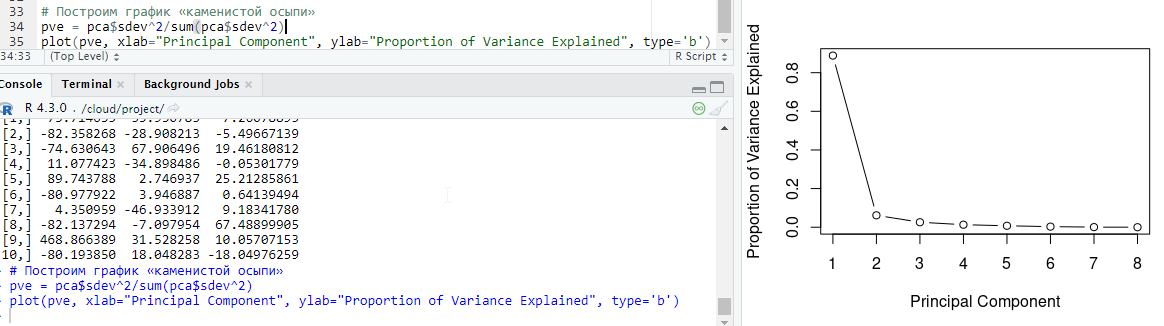


Рис.9. График «Каменистой осыпи»

Судя по графику, можно сказать, что главная компонента зависит от доли дисперсии. Кривая с самой высокой точки по y резко уходит вниз по х.

# **Вывод**

В ходе выполнения всех лабораторных работ мною были получены навыки технологии работы с кластерами, графиками. С помощью метода главных компонент мы смогли увидеть, как описывает данные каждая переменная. По данным видно, что наибольшую дисперсию имеют переменные Insulin, Glucose и BloodPressure.