Белорусский Государственный Университет

Информатики и Радиоэлектроники

Факультет компьютерных систем и сетей

Кафедра ЭВМ

Лабораторная работа №1

Тема «Корреляционный анализ»

Выполнил: Проверил:

Студент группы 7М2432 Марченко В.В.

Канаш В.Н.

Минск, 2017

Задание:

Входные данные: n объектов, каждый из которых характеризуется

двумя числовыми признаками: и .

Требуется исследовать степень взаимосвязи между двумя признаками некоторых объектов. Для каждого набора данных необходимо выполнить следующие задания:

1. Визуализировать данные на плоскости в виде точек с

координатами .

2. Статистически оценить коэффициент корреляции Пирсона между признаками *x* и *y*.

3. Проверить статистическую гипотезу о некоррелированности признаков *x* и *y* на уровне значимости 0,05.

Исходные данные:

1) значения объёма исследуемой выборки (n) – 10000;

2) значения вектора математических ожиданий (a) – (-1, 1);

3) корреляционных матриц (R) для моделируемой выборки из гауссовских случайных векторов –

Все описанные выше задания требуется выполнить для двух наборов данных.

1. Смоделированные независимые случайные векторы (X, Y), имеющие гауссовское распределение с заданным математическим ожиданием a и корреляционной матрицей R.

2. Реальные статистические данные из заданного набора (выдаются преподавателем).

Варианты реальных наборов данных №6. Wine

Название файла: 06-wine.txt

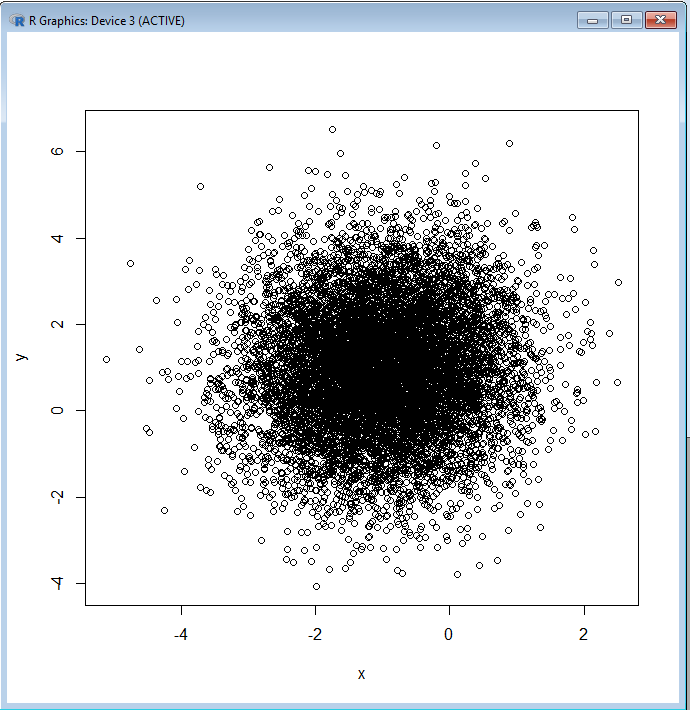
Ссылка: <http://archive.ics.uci.edu/ml/datasets/Wine>

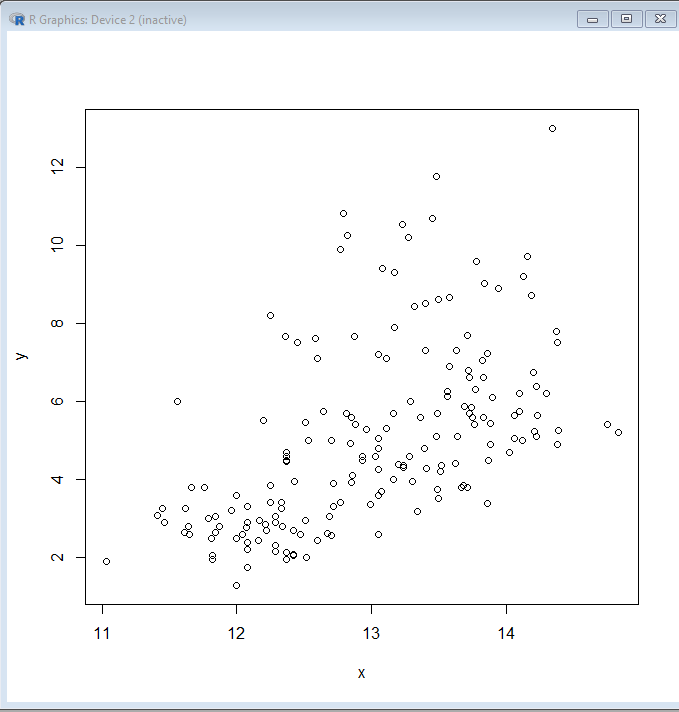
Первый признак: alcohol (столбец № 2)

Второй признак: color-intensity (столбец № 11)

Результаты:

1. Изображения данных в виде точек на плоскости:





1. Статистические оценки коэффициентов корреляции Пирсона для каждого набора данных, сравнение статистической оценки коэффициента корреляции Пирсона с реальным коэффициентом корреляции Пирсона для смоделированных данных:

a) Данные из wine.csv:

data: x and y

t = 8.6542, df = 176, p-value = 3.056e-15

alternative hypothesis: true correlation is not equal to 0

5 percent confidence interval:

0.5430304 0.5496808

sample estimates:

cor

0.5463642

Число стьюдента для уровня значимости 0.05 и степеней свободы 200

равно 1.971

Т.к. |t| > 1,971, то гипотеза о некоррелированности отвергается.

b) Данные из модуляции по выборке:

data: x and y

t = 7.9895, df = 9998, p-value = 1.505e-15

alternative hypothesis: true correlation is not equal to 0

5 percent confidence interval:

0.07902570 0.08027207

sample estimates:

cor

0.07964892

Число Стьюдента для уровня значимости 0,05 и степеней свободы >100 равно 1,96.

Т.к. |t| > 1,96, то гипотеза о некоррелированности отвергается.

Листинг программы:

require(MASS)

analyse\_cor <- function(x, y) {

print(cor.test(x, y))

dev.new()

plot(x, y)

}

dat <- read.table("wine.csv", sep=",")

analyse\_cor(dat$V2, dat$V11)

n <- 10000

a <- c(-1, 1)

r <- cbind(c(1, 0.1), c(0.1, 2))

dat <- mvrnorm(n, a, r)

analyse\_cor(dat[,1], dat[,2])

Анализируемые данные:

|  |  |
| --- | --- |
| 14.23 | 5.64 |
| 13.2 | 4.38 |
| 13.16 | 5.68 |
| 14.37 | 7.8 |
| 13.24 | 4.32 |
| 14.2 | 6.75 |
| 14.39 | 5.25 |
| 14.06 | 5.05 |
| 14.83 | 5.2 |
| 13.86 | 7.22 |
| 14.1 | 5.75 |
| 14.12 | 5 |
| 13.75 | 5.6 |
| 14.75 | 5.4 |
| 14.38 | 7.5 |
| 13.63 | 7.3 |
| 14.3 | 6.2 |
| 13.83 | 6.6 |
| 14.19 | 8.7 |
| 13.64 | 5.1 |
| 14.06 | 5.65 |
| 12.93 | 4.5 |
| 13.71 | 3.8 |
| 12.85 | 3.93 |
| 13.5 | 3.52 |
| 13.05 | 3.58 |
| 13.39 | 4.8 |
| 13.3 | 3.95 |
| 13.87 | 4.5 |
| 14.02 | 4.7 |
| 13.73 | 5.7 |
| 13.58 | 6.9 |
| 13.68 | 3.84 |
| 13.76 | 5.4 |
| 13.51 | 4.2 |
| 13.48 | 5.1 |
| 13.28 | 4.6 |
| 13.05 | 4.25 |
| 13.07 | 3.7 |
| 14.22 | 5.1 |
| 13.56 | 6.13 |
| 13.41 | 4.28 |
| 13.88 | 5.43 |
| 13.24 | 4.36 |
| 13.05 | 5.04 |
| 14.21 | 5.24 |
| 14.38 | 4.9 |
| 13.9 | 6.1 |
| 14.1 | 6.2 |
| 13.94 | 8.90 |
| 13.05 | 7.2 |
| 13.83 | 5.6 |
| 13.82 | 7.05 |
| 13.77 | 6.3 |
| 13.74 | 5.85 |
| 13.56 | 6.25 |
| 14.22 | 6.38 |
| 13.29 | 6 |
| 13.72 | 6.8 |
| 12.37 | 1.95 |
| 12.33 | 3.27 |
| 12.64 | 5.75 |
| 13.67 | 3.8 |
| 12.37 | 4.45 |
| 12.17 | 2.95 |
| 12.37 | 4.6 |
| 13.11 | 5.3 |
| 12.37 | 4.68 |
| 13.34 | 3.17 |
| 12.21 | 2.85 |
| 12.29 | 3.05 |
| 13.86 | 3.38 |
| 13.49 | 3.74 |
| 12.99 | 3.35 |
| 11.96 | 3.21 |
| 11.66 | 3.8 |
| 13.03 | 4.6 |
| 11.84 | 2.65 |
| 12.33 | 3.4 |
| 12.7 | 2.57 |
| 12 | 2.5 |
| 12.72 | 3.9 |
| 12.08 | 2.2 |
| 13.05 | 4.8 |
| 11.84 | 3.05 |
| 12.67 | 2.62 |
| 12.16 | 2.45 |
| 11.65 | 2.6 |
| 11.64 | 2.8 |
| 12.08 | 1.74 |
| 12.08 | 2.4 |
| 12 | 3.6 |
| 12.69 | 3.05 |
| 12.29 | 2.15 |
| 11.62 | 3.25 |
| 12.47 | 2.6 |
| 11.81 | 2.5 |
| 12.29 | 2.9 |
| 12.37 | 4.5 |
| 12.29 | 2.3 |
| 12.08 | 3.3 |
| 12.6 | 2.45 |
| 12.34 | 2.8 |
| 11.82 | 2.06 |
| 12.51 | 2.94 |
| 12.42 | 2.7 |
| 12.25 | 3.4 |
| 12.72 | 3.3 |
| 12.22 | 2.7 |
| 11.61 | 2.65 |
| 11.46 | 2.9 |
| 12.52 | 2 |
| 11.76 | 3.8 |
| 11.41 | 3.08 |
| 12.08 | 2.9 |
| 11.03 | 1.9 |
| 11.82 | 1.95 |
| 12.42 | 2.06 |
| 12.77 | 3.4 |
| 12 | 1.28 |
| 11.45 | 3.25 |
| 11.56 | 6 |
| 12.42 | 2.08 |
| 13.05 | 2.6 |
| 11.87 | 2.8 |
| 12.07 | 2.76 |
| 12.43 | 3.94 |
| 11.79 | 3 |
| 12.37 | 2.12 |
| 12.04 | 2.6 |
| 12.86 | 4.1 |
| 12.88 | 5.4 |
| 12.81 | 5.7 |
| 12.7 | 5 |
| 12.51 | 5.45 |
| 12.6 | 7.1 |
| 12.25 | 3.85 |
| 12.53 | 5 |
| 13.49 | 5.7 |
| 12.84 | 4.92 |
| 12.93 | 4.6 |
| 13.36 | 5.6 |
| 13.52 | 4.35 |
| 13.62 | 4.4 |
| 12.25 | 8.21 |
| 13.16 | 4 |
| 13.88 | 4.9 |
| 12.87 | 7.65 |
| 13.32 | 8.42 |
| 13.08 | 9.40 |
| 13.5 | 8.60 |
| 12.79 | 10.8 |
| 13.11 | 7.1 |
| 13.23 | 10.52 |
| 12.58 | 7.6 |
| 13.17 | 7.9 |
| 13.84 | 9.01 |
| 12.45 | 7.5 |
| 14.34 | 13 |
| 13.48 | 11.75 |
| 12.36 | 7.65 |
| 13.69 | 5.88 |
| 12.85 | 5.58 |
| 12.96 | 5.28 |
| 13.78 | 9.58 |
| 13.73 | 6.62 |
| 13.45 | 10.68 |
| 12.82 | 10.26 |
| 13.58 | 8.66 |
| 13.4 | 8.5 |
| 12.2 | 5.5 |
| 12.77 | 9.899999 |
| 14.16 | 9.7 |
| 13.71 | 7.7 |
| 13.4 | 7.3 |
| 13.27 | 10.2 |
| 13.17 | 9.3 |
| 14.13 | 9.2 |