Белорусский Государственный Университет

Информатики и Радиоэлектроники

Факультет компьютерных систем и сетей

Кафедра ЭВМ

Лабораторная работа №2

Тема «Регрессионный анализ»

Выполнил: Проверил:

Студент группы 7М2432 Марченко В.В.

Пашковский А.А.

Минск, 2017

Задание:

Входные данные: n объектов, каждый из которых характеризуется

двумя числовыми признаками: и .

Требуется исследовать регрессионную зависимость признака y от признака х. Для каждого набора данных необходимо выполнить следующие задания:

1. Построить модель линейной регрессии *y = ax + b + ε*, оценив оптимальные параметры *a* и *b* из условия минимизации суммы квадратов отклонения для заданных значений признаков и .

2. Вычислить коэффициент детерминации для получившейся модели.

3. Визуализировать на одном графике точки и прямую y = ax + b.

Исходные данные:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Вариант | N | a | b | σ2 |
| 3 | 1000 | 2 | 0,1 | 0,1 |

Где N – это количество точек, a и b – коэффициенты в линейной функции y = ax + b + ε, а σ2 – дисперсия гауссовского белого шума ε. Сами значения x задаются в виде равномерной сетки на отрезке [0; 1].

Реальные статистические данные из заданного набора (выдаются преподавателем).

Название файла: 26-parkinsons.txt

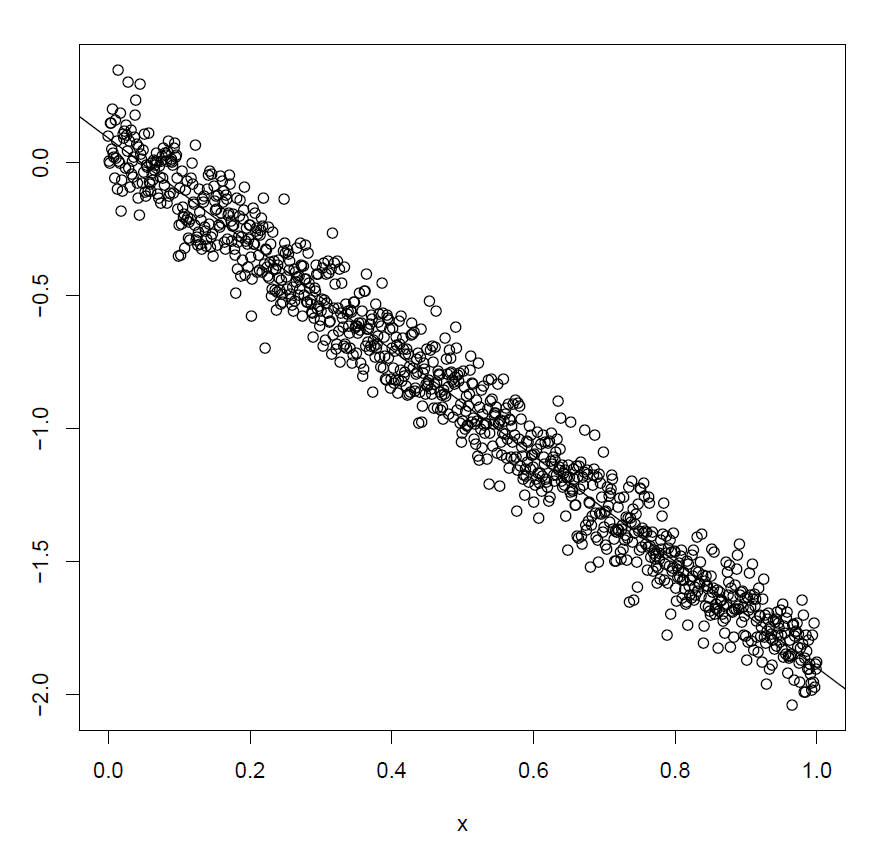
Ссылка: http://archive.ics.uci.edu/ml/datasets/Parkinsons

Предиктор: MDVP:Fhi(Hz) (столбец № 3)

Зависимая переменная: MDVP:Flo(Hz) (столбец № 4)

Результаты:

1. Смоделированные данные:



Call:

lm(formula = y ~ x)

Residuals:

Min 1Q Median 3Q Max

-0.35107 -0.06945 -0.00190 0.06919 0.29238

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.092459 0.006326 14.62 <2e-16 \*\*\*

x -1.990685 0.010954 -181.73 <2e-16 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

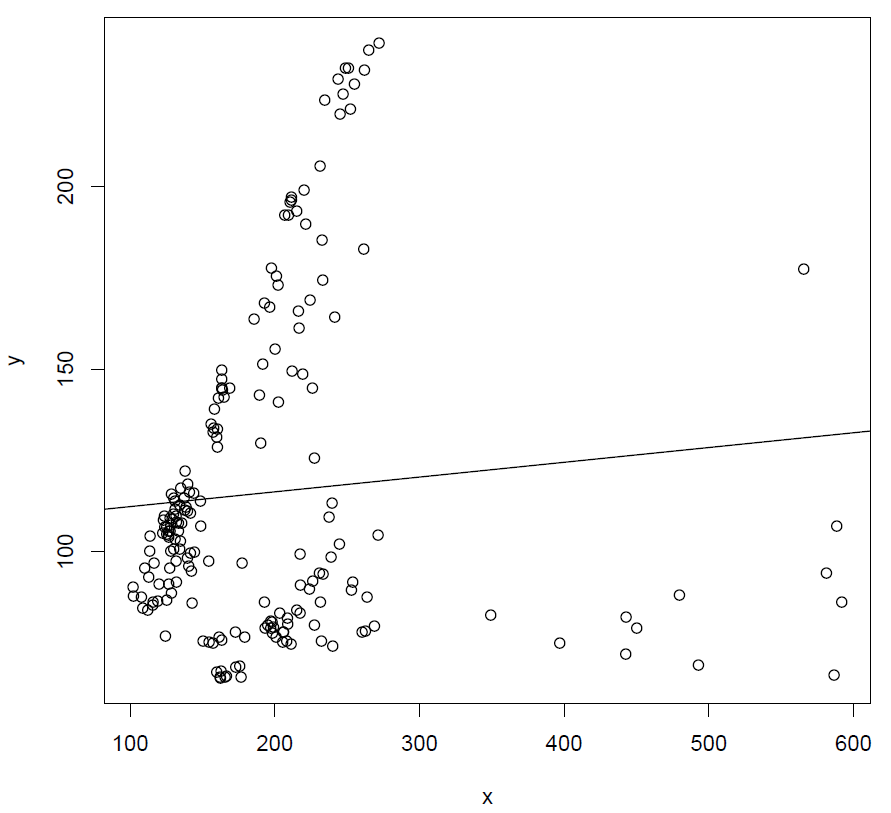
Residual standard error: 0.1001 on 998 degrees of freedom

Multiple R-squared: 0.9707, Adjusted R-squared: 0.9706

F-statistic: 3.303e+04 on 1 and 998 DF, p-value: < 2.2e-16

Коэффициент детерминации = 0.10

1. Реальные данные:



Call:

lm(formula = y ~ x)

Residuals:

Min 1Q Median 3Q Max

-65.91 -33.62 -10.29 24.37 119.81

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 108.35957 7.41039 14.623 <2e-16 \*\*\*

x 0.04041 0.03412 1.184 0.238

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 43.48 on 193 degrees of freedom

Multiple R-squared: 0.007217, Adjusted R-squared: 0.002073

F-statistic: 1.403 on 1 and 193 DF, p-value: 0.2377

Коэффициент детерминации = 0.007

Листинг программы:

