МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МО ЭВМ

ОТЧЕТ

по лабораторной работе №5
по дисциплине «СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ»
Тема: Элементы регрессионного анализа. Выборочные прямые среднеквадратической регрессии. Корреляционное отношение.

Студент гр. 5381		Лянгузов А. А.
Преподаватель		Середа В. И.

Санкт-Петербург
2019

Цель работы.

Ознакомление с основными положениями метода наименьших квадратов (МНК), со статистическими свойствами МНК оценок, с понятием функции регрессии и роли МНК в регрессионном анализе, с корреляционным отношением, как мерой тесноты произвольной (в том числе и линейной) корреляционной связи.

Задание.

Для заданной двумерной (Х,У) выборки построить уравнения выборочных прямых среднеквадратической регрессии. Вычислить оценки остаточной дисперсии. Полученные функции регрессии отобразить графически (на одном графике).

Для заданной двумерной (Х,У) выборки построить уравнения выборочных кривых для параболической среднеквадратической регрессии. Вычислить оценки остаточной дисперсии. Полученные функции регрессии отобразить графически (на одном графике).

Полученные результаты содержательно проинтерпретировать.

Основные теоретические положения.

Регрессионный анализ — это статистический метод исследования влияния одной или нескольких независимых переменных $X_1, X_2, ..., X_p$ на зависимую переменную Y. Независимые переменные иначе называют регрессорами или предикторами, а зависимые переменные — критериальными.

Линейные функции выборочной среднеквадратической регрессии выглядят следующим образом:

$$y = \bar{y_B} + \bar{r_{xy}}\frac{S_y}{S_x}(x - \bar{x_B})x = \bar{x_B} + \bar{r_{xy}}\frac{S_x}{S_y}(y - \bar{y_B}),$$

где $\bar{x_B}$, $\bar{y_B}$ - статистические оценки математических ожиданий выборок X, Y соответственно;

$\bar{r_{xy}}$- статистическая оценка коэффициента корреляции;

S_x, S_y- статистические оценки среднеквадратических отклонений для выборок X, Y.

Чтобы рассчитать выборочное корреляционное отношение Y к X нужно рассчитать внутригрупповую и межгрупповую дисперсии. Внутригрупповая дисперсия вычисляется по формуле:

$$D_{внутригр\ ух} = \frac{1}{n}\sum_{i=1}^{k_2} n_{x_i} * D_{у\ гр\ i}$$

где n – объём выборки,

k_2 – количество интервалов интервального ряда X,

n_{x_i} – абсолютная частота для i-ого интервала интервального ряда X,

$D_{у\ групп\ i}$ – групповая дисперсия элементов выборки Y на i-ом интервале интервального ряда X.

Межгрупповая дисперсия вычисляется по формуле:

$$D_{межгр\ ух} = \frac{1}{n}\sum_{i=1}^{k_2} n_{x_i} * (\bar{y}_{гр\ i} - \bar{y}_B)^2$$

где n – объём выборки,

k_2 – количество интервалов интервального ряда X,

n_{x_i} – абсолютная частота для i-ого интервала интервального ряда X,

$\bar{y}_{гр\ i}$ i – групповое математическое ожидание элементов выборки Y на i-ом интервале интервального ряда X,

$\bar{y}_B$ – статистическая оценка математического ожидания Y.

Суммарная выборочная дисперсия рассчитывается по формуле:

$$D_{общ\ ух} = D_{внутригр\ ух} + D_{межгр\ ух}$$

Выборочное корреляционное отношение Y к X рассчитывается по следующей формуле:

$$\eta_{ух} = \sqrt{\frac{D_{межгр\ ух}}{D_{общ\ ух}}}$$

Для расчёта выборочного корреляционного отношения X к Y необходимо рассчитать те же величины по следующим формулам (меняем местами X и Y):

$$D_{внутригр\ xy} = \frac{1}{n}\sum_{i=1}^{k_1} n_{y_i} * D_{x\ гр\ i}$$

$$D_{межгр\ xy} = \frac{1}{n}\sum_{i=1}^{k_1} n_{y_i} * \left(\bar{x}_{гр\ i} - \bar{x}_B\right)^2$$

$$D_{общ\ xy} = D_{внутригр\ xy} + D_{межгр\ xy}$$

$$\eta_{xy} = \sqrt{\frac{D_{межгр\ xy}}{D_{общ\ xy}}}$$

В качестве коэффициентов квадратичной функции выборочной среднеквадратической регрессии $y = ax^2 + bx + c$ следует выбрать такие, чтобы значение функции $f(a, b, c) = \sum_{i=1}^{n} \left(y_i - ax_i^2 - bx_i - c\right)^2$ было минимально.

Взяв производные от данной функции и приравняв их нулю, получим:

$$\frac{\partial f}{\partial a} = -2\sum_{i=1}^{n} (y_i - ax_i^2 - bx_i - c)x_i^2$$

$$\frac{\partial f}{\partial b} = -2\sum_{i=1}^{n} \left(y_i - ax_i^2 - bx_i - c\right)x_i = 0$$

$$\frac{\partial f}{\partial c} = -2\sum_{i=1}^{n} \left(y_i - ax_i^2 - bx_i - c\right) = 0$$

Преобразовав данные равенства, получим следующую систему линейных уравнений.

$$\begin{cases} a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i^2 \\ a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i x_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i + cn = \sum_{i=1}^{n} y_i \end{cases}$$

Решив данную систему, найдём коэффициенты квадратичной функции выборочной среднеквадратической регрессии.

Обработка результатов эксперимента.

Графики линейной функции выборочной среднеквадратической регрессии $y(x)$ и x(y) представлены на рисунке 1.

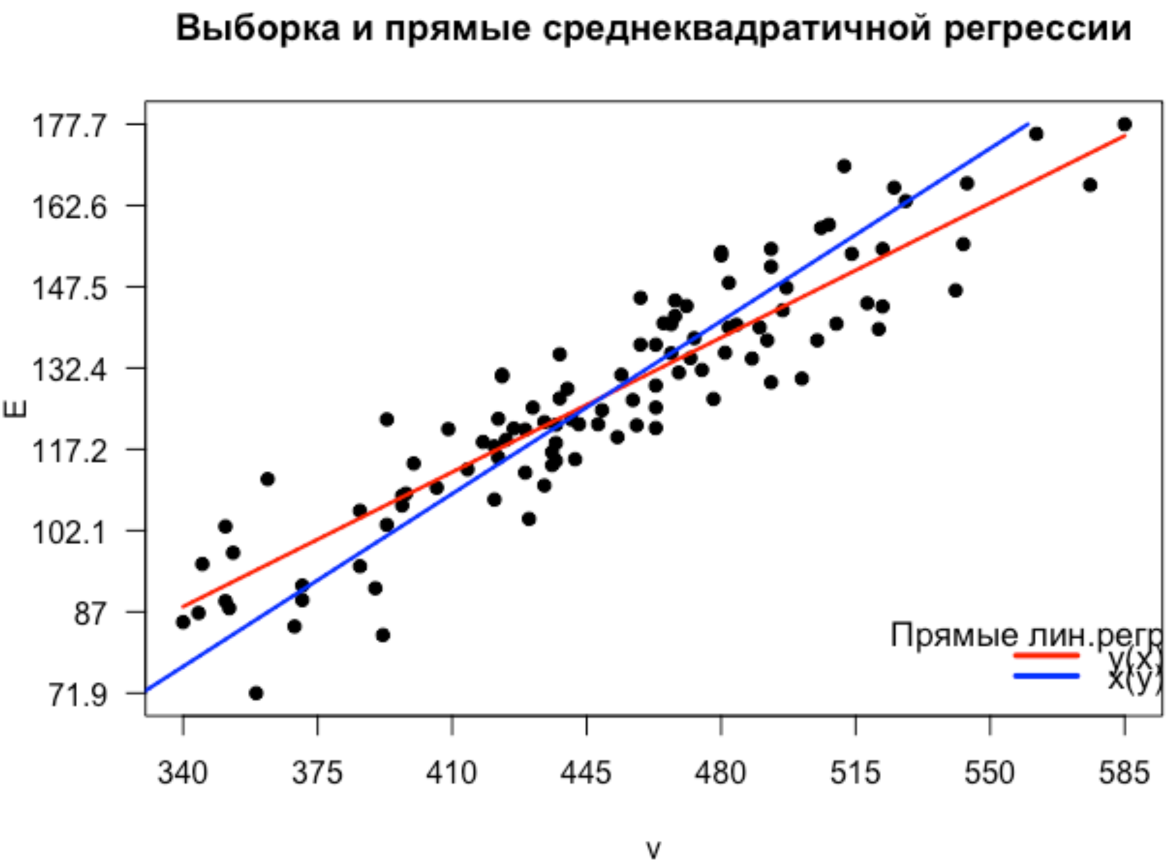

Рис. 1

Линейная функция среднеквадратической регрессии $y(x)$ была определена по формуле $y(x) = \bar{y}_B + \bar{r}_{xy}\frac{S_y}{S_x}(x - \bar{x}_B)$ В результате была получена следующая функция: y(x) = 0.3574181x + 127.2013

Линейная функция среднеквадратической регрессии *x(y)* была определена по формуле x(y) $= \bar{x}_B + \bar{r}_{xy}\frac{S_x}{S_y}(y - \bar{y}_B)$ В результате была получена следующая функция: x(y) = 2.179471y + 449.743

Найдем оценки остаточной дисперсии для полученных выборочных уравнений регрессии.

$D_{ост\,y} = 105.8849$

$D_{ост\,x} = 645.6670$

Вспомогательная таблица представлена в таблице 1.

Таблица 1.

$x_i y_i$	[71.9,87)	[87,102)	[102,117)	[117,132)	[132,147)	[147,163)	[163,178]	N	y_{gr}	$D_{y\,gr\,i}$
[340,375)	4	6	2	0	0	0	0	12	92.05238	107.87522
[375,410)	1	2	7	2	0	0	0	12	107.16667	145.94882
[410,445)	0	0	10	16	1	0	0	27	119.76190	67.68641
[445,480)	0	0	0	12	11	0	0	23	132.02857	57.00245
[480,515)	0	0	0	2	10	8	1	21	145.67211	119.14189
[515,550)	0	0	0	0	4	2	3	9	153.34921	174.85656
[550,585]	0	0	0	0	0	0	3	3	170.14286	0.00000
N	5	8	19	32	26	10	7	107		
x_{gr}	364.5000	366.2500	407.2368	442.8125	485.3846	504.5000	542.5000			
$D_{x\,gr\,i}$	196.0000	229.6875	556.5097	607.7148	748.4098	196.0000	600.0000			

Рассчитаем внутригрупповую, межгрупповую, общую дисперсии, а также Выборочное корреляционное отношение.

Внутригрупповая дисперсия:

$$D_{внутригр\,xy} = \frac{1}{n}\sum_{i=1}^{k_1} n_{y_i} * D_{x\,гр\,i} = 546.3244$$

$$D_{внутригр\,yx} = \frac{1}{n}\sum_{i=1}^{k_2} n_{x_i} * D_{y\,гр\,i} = 95.88941$$

Межгрупповая дисперсия:

$$D_{межгр\,xy} = \frac{1}{n}\sum_{i=1}^{k_1} n_{y_i} * \left(\bar{x_{гр\,i}} - \bar{x}_B\right)^2 = 2347.71$$

$$D_{межгр\ yx} = \frac{1}{n}\sum_{i=1}^{k_2} n_{x_i}*(\bar{y_{гр\ i}} - \bar{y_B})^2 = 378.7121$$

Общая дисперсия:

$$D_{общ\ xy} = D_{внутригр\ xy} + D_{межгр\ xy} = 2894.034$$

$$D_{общ\ yx} = D_{внутригр\ yx} + D_{межгр\ yx} = 474.6015$$

Выборочное корреляционное отношение:

$$\eta_{xy} = \sqrt{\frac{D_{межгр\ xy}}{D_{общ\ xy}}} = 0.8112239$$

$$\eta_{yx} = \sqrt{\frac{D_{межгр\ yx}}{D_{общ\ yx}}} = 0.7979581$$

Для определения коэффициентов квадратичной функции выборочной среднеквадратической регрессии $y = ax^2 + bx + c$ была решена следующая система уравнений:

$$\begin{cases} a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i^2 \\ a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i x_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i + cn = \sum_{i=1}^{n} y_i \end{cases}$$

В результате были получены следующие значения коэффициентов:

$a = -0.0001149475 \qquad b = 0.5092418 \qquad c = -78.51749$

Полученное уравнение примет вид:

$$y = -0.0001149475*x^2 + 0.5092418*\text{x} - 78.51749$$

График квадратичной функции выборочной среднеквадратической регрессии y(x) представлен на рисунке 2.

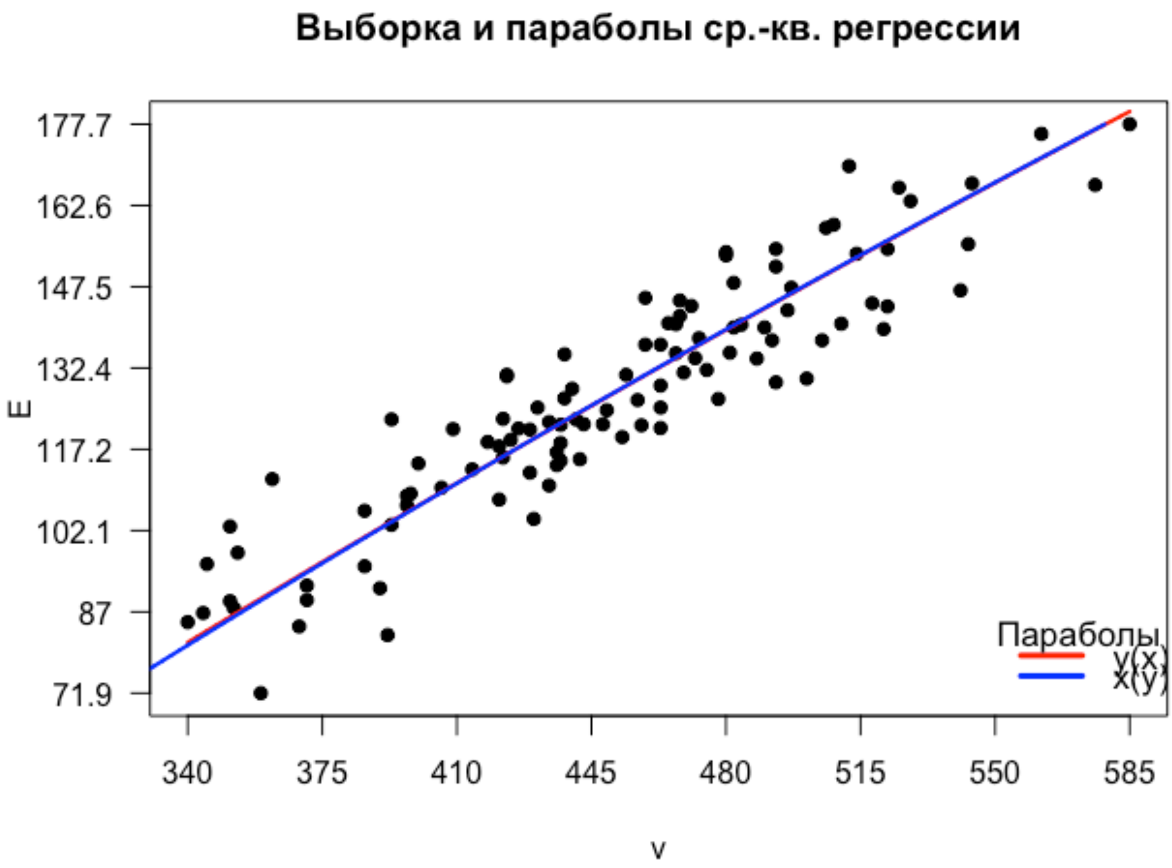
Выборка и параболы ср.-кв. регрессии
177.7
162.6
147.5
132.4
117.2
102.1
87
71.9
Ш
340
375
410
445
480
515
550
585
v
Параболы
y(x)
x(y)

Рис.2

Вывод.

Для заданной двумерной (Х,У) выборки построены уравнения выборочных прямых среднеквадратической регрессии. Вычислены оценки остаточной дисперсии

Полученные в результате выполнения работы выборочные корреляционные отношения характеризуют тесноту связи между X и Y. Функции среднеквадратической регрессии приближают зависимость между X и Y.

Литература.

1. Середа В.И. Курс лекций по статическим методам обработки экспериментальных данных. Лекция, 23.03.2019