

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

ОТЧЕТ
по лабораторной работе №5
по дисциплине «Статистические методы обработки экспериментальных
данных»
Тема: Элементы регрессионного анализа. Выборочные прямые средне-
квадратической регрессии. Корреляционные отношения.

Студент гр. 8383

Киреев К.А.

Студент гр. 8383

Муковский Д.В.

Преподаватель

Середа А.-В.И.

Санкт-Петербург

2022

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.

Основные теоретические положения

Метод наименьших квадратов (МНК) — метод, основанный на поиске минимума суммы квадратов отклонений значений некоторых функций от заданного множества значений. МНК является одним из основных методов регрессионного анализа и применяется для оценки параметров регрессионных моделей на основе выборочных данных.

$$M(X/y) = q_1(y)$$

$$M(Y/x) = q_2(x)$$

Пусть имеется двумерная случайная величина $\{X, Y\}$, где X и Y зависимые случайные величины. Функцию $g(x)$ называют линейной функцией среднеквадратической регрессии Y на X .

$$g(x) = m\left(\frac{Y}{x}\right) = m(Y) + r_{xy} \frac{\sigma_y}{\sigma_x} [x - m(X)]$$

В случае, когда известны только выборочные данные – двумерная выборка значений случайных величин X и Y , возможно построение только выборочных прямых среднеквадратической регрессии.

Уравнения выборочных прямых среднеквадратической регрессии:

$$\overline{y}_x = \overline{y}_B + \overline{r}_{xy} \frac{S_y}{S_x} (x - \overline{x}_B)$$

$$\overline{x}_y = \overline{x}_B + \overline{r}_{xy} \frac{S_x}{S_y} (y - \overline{y}_B)$$

Для оценки корреляционной зависимости между случайными величинами в общем, а не только линейной, может быть использовано так называемое корреляционное отношение.

Оценку общей дисперсии X можно представить, как сумму внутригрупповой и межгрупповой дисперсии:

$$D_{\text{общ}} = D_{\text{межгр}} + D_{\text{внгр}}$$

Внутригрупповая дисперсия вычисляется, как взвешенная по объемам групп средняя арифметическая групповых дисперсий.

Межгрупповая дисперсия вычисляется, как дисперсия условных (групповых) средних $\overline{x_{y_i}}$ относительно выборочной средней \overline{x}_y .

Выборочное корреляционное отношение Y к X определяется в соответствии с выражением:

$$\overline{\eta_{yx}} = \frac{\overline{\sigma_{y_x}}}{\overline{\sigma_y}} = \sqrt{\frac{D_{\text{межгр}}}{D_{\text{общ}}}},$$

где $\overline{\sigma_{y_x}} = \sqrt{D_{\text{межгр}}}$, $\overline{\sigma_y} = \sqrt{D_{\text{общ}}}$ – выборочные значения СКВО $\overline{y_x}$ и Y соответственно. Аналогично определяется выборочное корреляционное отношение X к Y .

Выборочное уравнение регрессии Y на X параболического вида:

$$\overline{y_x} = ax^2 + bx + c$$

Значения коэффициентов a, b и c определим с помощью МНК, что приводит к необходимости решать систему линейных уравнений третьего порядка:

$$\begin{cases} \left(\sum_{i=1}^m n_{x_i} x_i^4 \right) a + \left(\sum_{i=1}^m n_{x_i} x_i^3 \right) b + \left(\sum_{i=1}^m n_{x_i} x_i^2 \right) c = \sum_{i=1}^m n_{x_i} \overline{y_{x_i}} x_i^2 \\ \left(\sum_{i=1}^m n_{x_i} x_i^3 \right) a + \left(\sum_{i=1}^m n_{x_i} x_i^2 \right) b + \left(\sum_{i=1}^m n_{x_i} x_i \right) c = \sum_{i=1}^m n_{x_i} \overline{y_{x_i}} x_i \\ \left(\sum_{i=1}^m n_{x_i} x_i^2 \right) a + \left(\sum_{i=1}^m n_{x_i} x_i \right) b + Nc = \sum_{i=1}^m n_{x_i} \overline{y_{x_i}} \end{cases}$$

Постановка задачи

Для заданной двумерной выборки (X, Y) построить уравнения выборочных прямых среднеквадратической регрессии. Полученные линейные функции регрессии отобразить графически. Найти выборочное корреляционное отношение. Полученные результаты содержательно проинтерпретировать.

Выполнение работы

Двумерная выборка

Двумерная выборка показана на рис. 1.

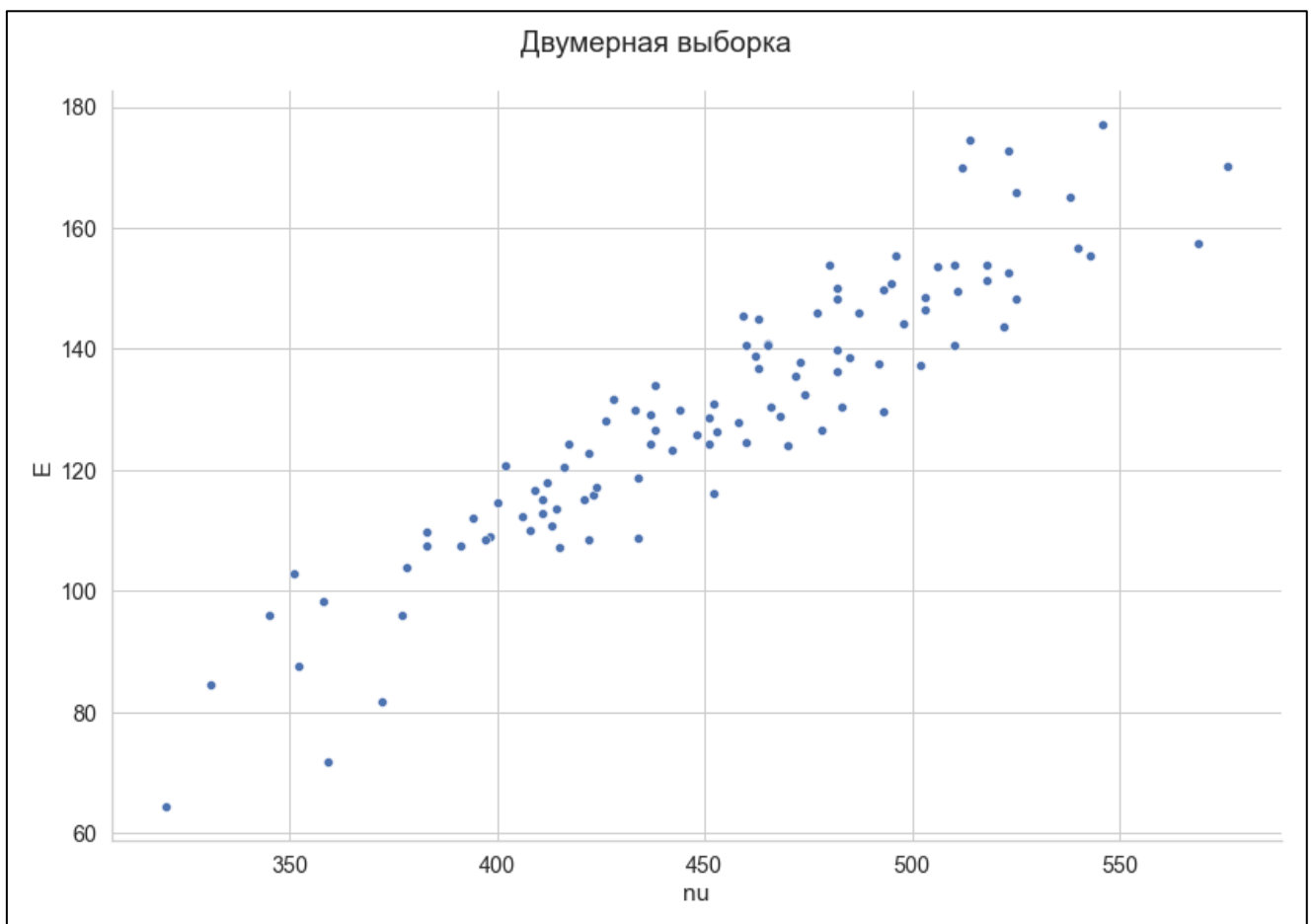


Рисунок 1 – Двумерная выборка

Уравнения средней квадратичной регрессии

Для заданной выборки построим уравнения средней квадратичной регрессии X на Y и Y на X и отобразим полученные прямые на множестве выборки.

Выборочная прямая средней квадратичной регрессии X на Y :

$$\bar{x}_y = \bar{x}_B + \bar{r}_{xy} \frac{S_x}{S_y} (y - \bar{y}_B)$$

$$x(y) = 453.71 + 0.8765 \frac{53.79}{22.06} (y - 129.98) = 2.1372 * y + 175.915$$

Выборочная прямая средней квадратичной регрессии Y на X :

$$\bar{y}_x = \bar{y}_B + \bar{r}_{xy} \frac{S_y}{S_x} (x - \bar{x}_B)$$

$$y(x) = 129.98 + 0.8765 \frac{22.06}{53.79} (x - 453.71) = 0.3595 * x - 33.1126$$

Двумерная выборка и выборочные прямые средней квадратичной регрессии представлены на рис. 2.

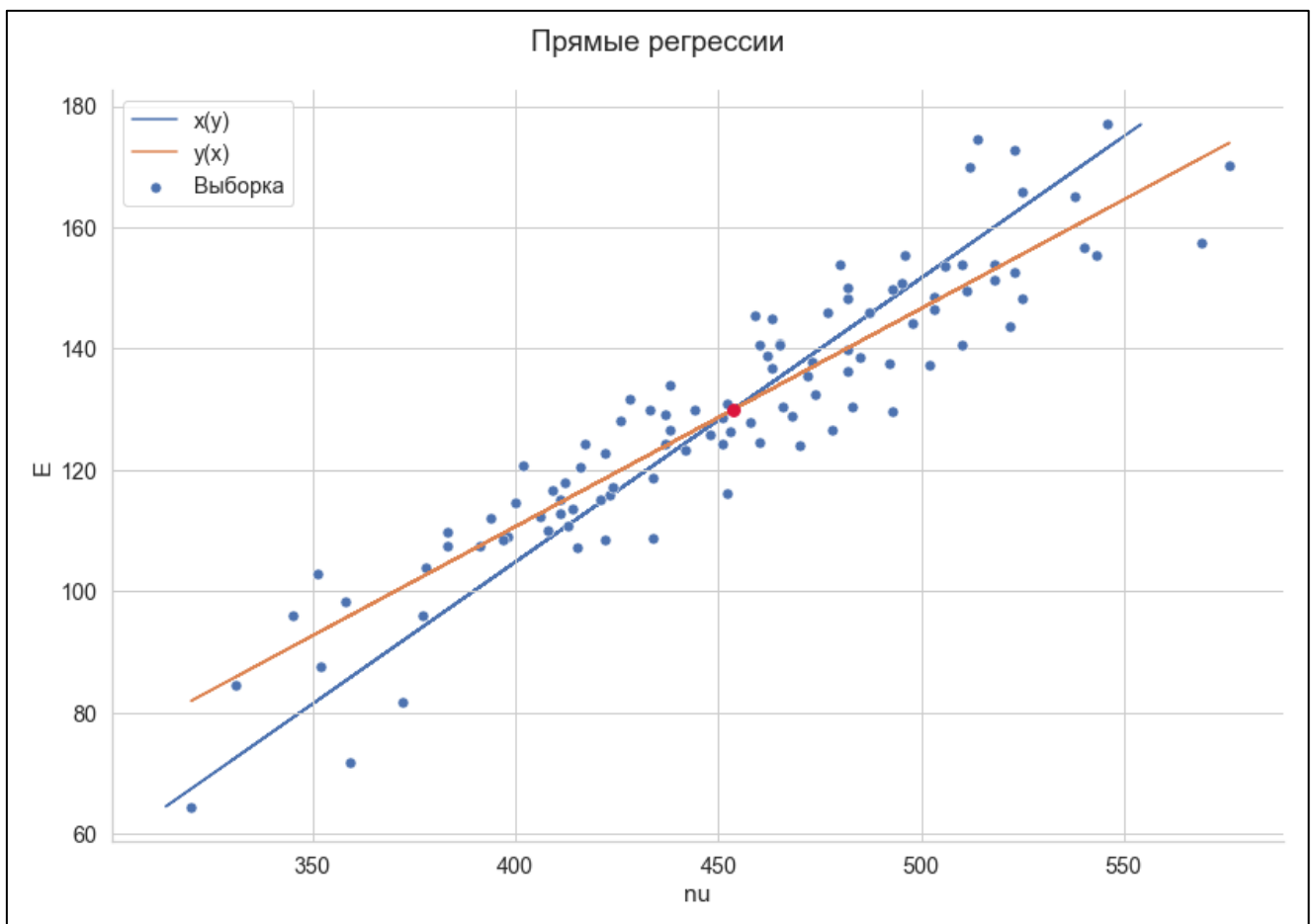


Рисунок 2 - Выборочные прямые средней квадратичной регрессии

Можно заметить, что пересечение выборочных прямых средней квадратичной регрессии находится в точке с координатами выборочного среднего для каждого из признаков.

Статистические оценки остаточной дисперсии для полученных выборочных прямых регрессии:

$$D_{\text{ост}_x} = S_x^2(1 - \overline{r_{xy}}^2) = 670.53$$

$$D_{\text{ост}_y} = S_y^2(1 - \overline{r_{xy}}^2) = 112.78$$

Выборочное корреляционное отношение

Корреляционная таблица для нахождения выборочного корреляционного отношения представлена в таблице 1. В данной таблице рассчитаны групповые выборочные средние и групповые выборочные дисперсии.

Таблица 1 - Корреляционная таблица

Y	X							n_{y_i}	\bar{x}_{y_i}	$D_{x_{y_i}}$
	338.5	375.5	412.5	449.5	486.5	523.5	559			
72.55	1	1	-	-	-	-	-	2	357	342.25
88.65	3	2	-	-	-	-	-	5	353.3	328.56
104.75	1	5	8	1	-	-	-	15	397.7	693.63
120.85	-	-	14	11	2	-	-	27	433.06	536.99
136.95	-	-	1	12	12	2	-	27	479.06	638.07
153.05	-	-	-	1	10	8	2	21	505.74	715.28
169.05	-	-	-	-	-	5	2	7	533.64	260.24
n_{x_j}	5	8	23	25	24	15	4	104	-	-
\bar{y}_{x_j}	88.65	96.7	115.95	129.22	142.32	156.24	161.05	-	-	-
$D_{y_{x_j}}$	103.68	129.6	77.42	106.69	99.86	108.7	64	-	-	-

Выборочное корреляционное отношение X к Y рассчитывается как отношение выборочных значений СКО $\overline{x_y}$ и X соответственно. Для этого были вычислены внутригрупповая, межгрупповая и общая дисперсии.

$$D_{\text{внгр}_{xy}} = \frac{1}{n} \sum_1^K D_{x_{y_i}} * n_{y_i} = 589.432$$

$$D_{\text{межгр}_{xy}} = \frac{1}{n} \sum_1^K (\bar{x}_{y_i} - \bar{x}_B)^2 * n_{y_i} = 2273.804$$

$$D_{\text{общ}_{xy}} = D_{\text{внгр}_{xy}} + D_{\text{межгр}_{xy}} = 2863.236$$

$$\overline{\eta_{xy}} = \sqrt{\frac{D_{\text{межгр}_{xy}}}{D_{\text{общ}_{xy}}}} = 0.8911$$

$$\overline{r_{xy}} = 0.8765$$

Неравенство $\overline{\eta_{xy}} \geq |\overline{r_{xy}}|$ выполняется.

Выборочное корреляционное отношение Y к X :

$$D_{\text{внгр}_{yx}} = \frac{1}{n} \sum_1^K D_{y_{x_j}} * n_{x_j} = 98.91$$

$$D_{\text{межгр}_{yx}} = \frac{1}{n} \sum_1^K (\bar{y}_{x_j} - \bar{y}_B)^2 * n_{x_j} = 382.72$$

$$D_{\text{общ}_{yx}} = D_{\text{внгр}_{yx}} + D_{\text{межгр}_{yx}} = 481.63$$

$$\overline{\eta_{yx}} = \sqrt{\frac{D_{\text{межгр}_{yx}}}{D_{\text{общ}_{yx}}}} = 0.8914$$

$$\overline{r_{xy}} = 0.8765$$

Неравенство $\overline{\eta_{yx}} \geq |\overline{r_{xy}}|$ выполняется.

Корреляционные кривые

Для заданной выборки построим корреляционную кривую параболического вида $y = \beta_2 x^2 + \beta_1 x + \beta_0$. Выборочное уравнение регрессии Y на X :

$$\overline{y_x} = ax^2 + bx + c$$

Значения коэффициентов определим с помощью МНК, решив систему уравнений:

$$\begin{cases} \left(\sum_{i=1}^m n_{x_i} x_i^4 \right) a + \left(\sum_{i=1}^m n_{x_i} x_i^3 \right) b + \left(\sum_{i=1}^m n_{x_i} x_i^2 \right) c = \sum_{i=1}^m n_{x_i} \overline{y_{x_i}} x_i^2 \\ \left(\sum_{i=1}^m n_{x_i} x_i^3 \right) a + \left(\sum_{i=1}^m n_{x_i} x_i^2 \right) b + \left(\sum_{i=1}^m n_{x_i} x_i \right) c = \sum_{i=1}^m n_{x_i} \overline{y_{x_i}} x_i \\ \left(\sum_{i=1}^m n_{x_i} x_i^2 \right) a + \left(\sum_{i=1}^m n_{x_i} x_i \right) b + Nc = \sum_{i=1}^m n_{x_i} \overline{y_{x_i}} \end{cases}$$

Для вычисления сумм была построена таблица 2.

Таблица 2 – Таблица сумм МНК

x	n_x	$\overline{y_x}$	$n_x x$	$n_x x^2$	$n_x x^3$	$n_x x^4$	$n_x \overline{y_x}$	$n_x \overline{y_x} x$	$n_x \overline{y_x} x^2$
338.5	5.0	88.65	1692.5	572911.25	193930458.125	65645460075.3125	443.25	150040.125	50788582.3125
375.5	8.0	96.7	3004.0	1128002.0	423564751.0	159048564000.5	773.6	290486.8	109077793.4
412.5	23.0	115.95	9487.5	3913593.75	1614357421.875	665922436523.4375	2666.85	1100075.625	453781195.3125
449.5	25.0	129.22	11237.5	5051256.25	2270539684.375	1020607588126.5625	3230.5	1452109.75	652723332.625
486.5	24.0	142.32	11676.0	5680374.0	2763501951.0	1344443699161.5	3415.68	1661728.3199999998	808430827.68
523.5	15.0	156.24	7852.5	4110783.75	2151995293.125	1126569535950.9375	2343.6	1226874.6	642268853.1
559.0	4.0	161.05	2236.0	1249924.0	698707516.0	390577501444.0	644.2	360107.80000000005	201300260.20000002
Σ	104.0	-	47186.0	21706845.0	10116597075.5	4772814785282.25	13517.68	6241423.0199999999	2918370844.6299996

В результате решения системы были получены следующие значения коэффициентов:

$$a = -0.00033$$

$$b = 0.6563$$

$$c = -99.7989$$

Выборочное уравнение регрессии Y на X :

$$y(x) = -0.00033 * x^2 + 0.6563 * x - 99.7989$$

Корреляционная кривая параболического вида представлена на рис. 3.

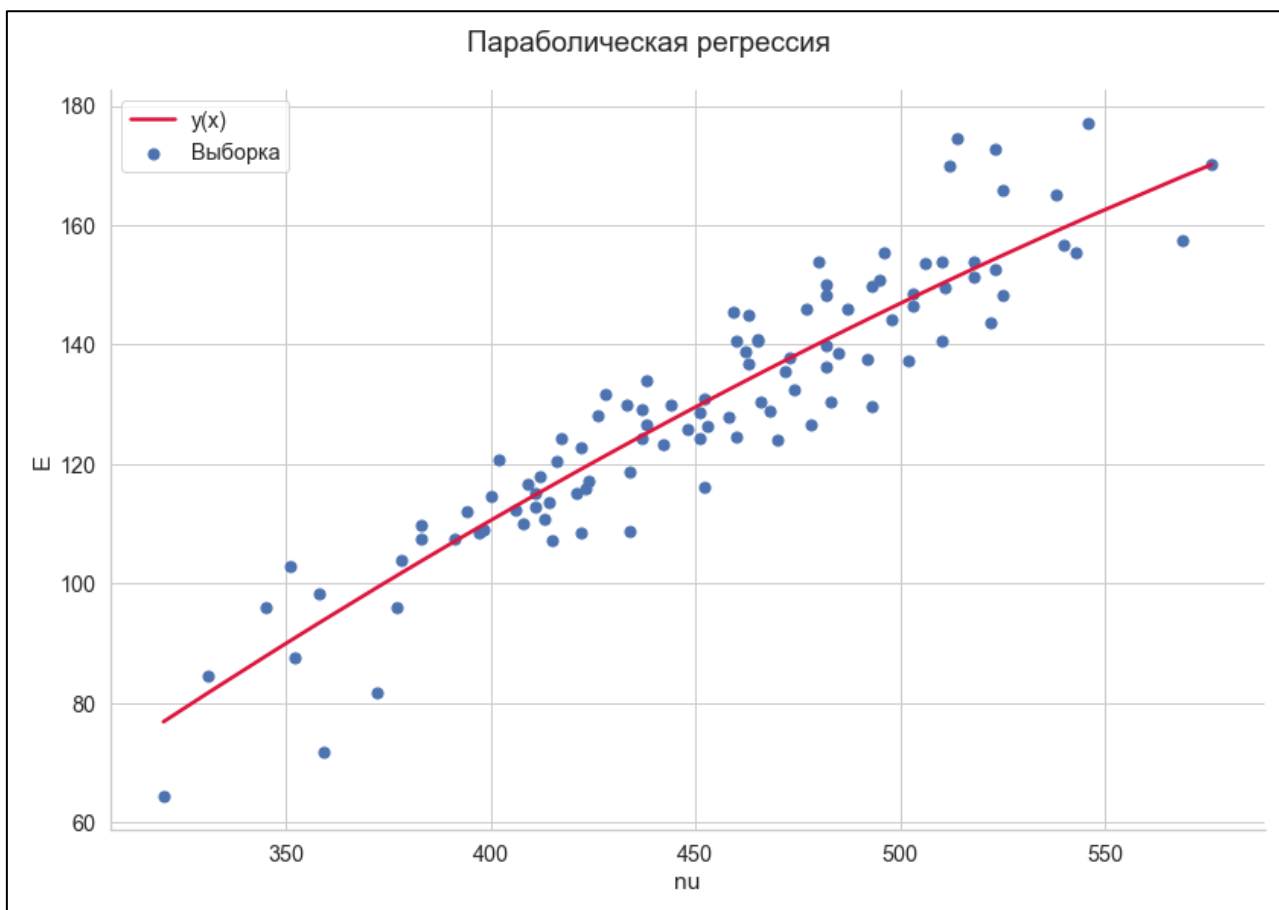


Рисунок 3 – Корреляционная кривая параболического вида

Для заданной выборки построим корреляционную кривую степенной функции $y = \beta_0 x^{\beta_1}$. Выборочное уравнение регрессии Y на X :

$$\overline{y_x} = a * x^b$$

Значения коэффициентов определим с помощью МНК, решив систему уравнений:

$$\begin{cases} \ln a + \left(\sum_{i=1}^m \ln x_i \right) b = \sum_{i=1}^m \ln \overline{y_{x_i}} \\ \left(\sum_{i=1}^m \ln x_i \right) \ln a + \left(\sum_{i=1}^m (\ln x_i)^2 \right) b = \sum_{i=1}^m \ln x_i * \ln \overline{y_{x_i}} \end{cases}$$

Для вычисления сумм была построена таблица 3.

Таблица 3 – Таблица сумм МНК

x	n_x	$\overline{y_x}$	$\ln x$	$(\ln x)^2$	$\ln y$	$\ln x * \ln y$
338.5	5.0	88.65	5.825	33.925	4.485	26.121
375.5	8.0	96.7	5.928	35.144	4.572	27.102
412.5	23.0	115.95	6.022	36.267	4.753	28.625
449.5	25.0	129.22	6.108	37.309	4.862	29.695
486.5	24.0	142.32	6.187	38.282	4.958	30.677
523.5	15.0	156.24	6.261	39.194	5.051	31.624
559.0	4.0	161.05	6.326	40.02	5.082	32.148
Σ	104.0	-	42.657	260.142	33.762	205.991

В результате решения системы были получены следующие значения коэффициентов:

$$\ln a = 0.002575 \Rightarrow a = 1.00258$$

$$b = 0.7914$$

Выборочное уравнение регрессии Y на X :

$$y(x) = 1.00258 * x^{0.7914}$$

Корреляционная кривая степенной функции представлена на рис. 4.

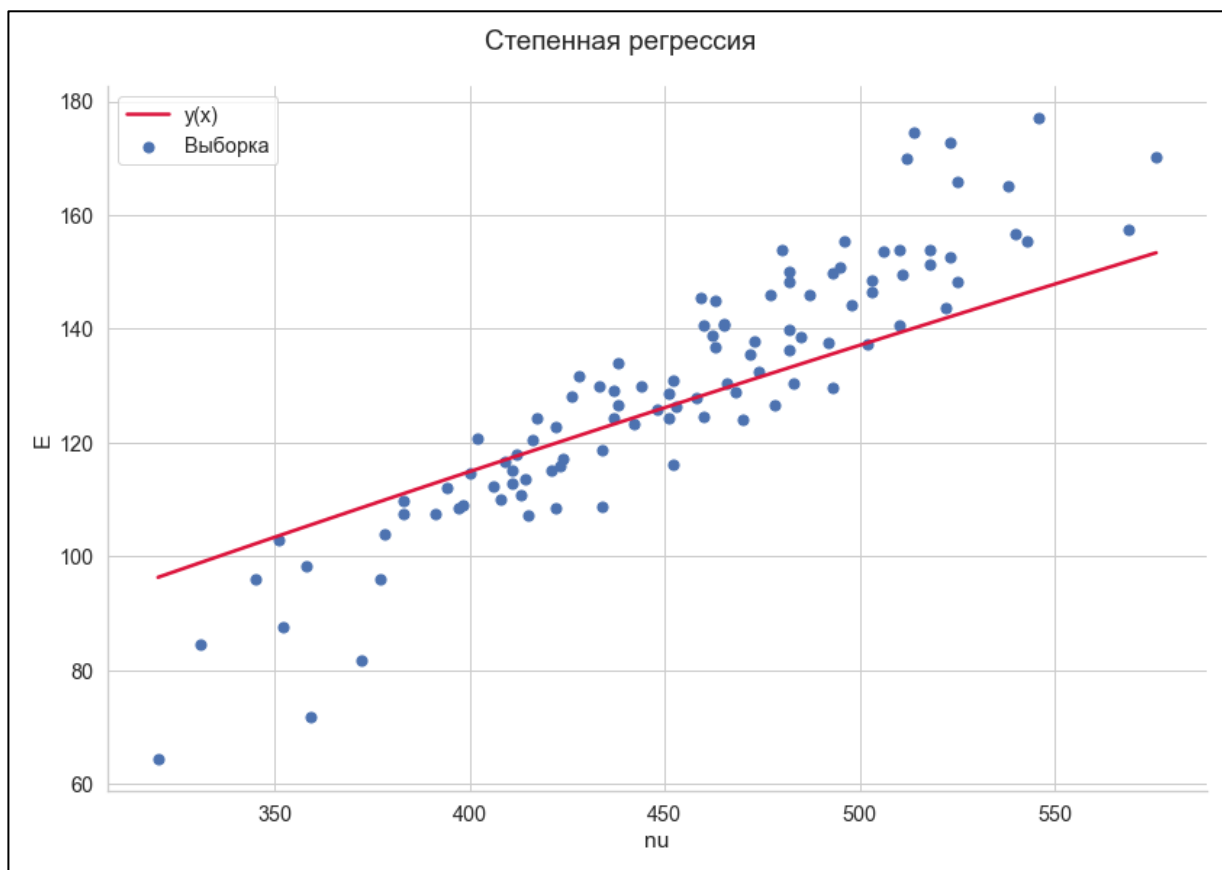


Рисунок 4 – Корреляционная кривая степенной функции

Выводы

Для заданной выборки были получены выборочные прямые средней квадратичной регрессии X на Y и Y на X .

$$x(y) = 2.1372 * y + 175.915$$

$$y(x) = 0.3595 * x - 33.1126$$

Прямые были построены на множестве выборки.

Найдены выборочные корреляционные отношения $\overline{\eta_{xy}} = 0.8911$ и $\overline{\eta_{yx}} = 0.8194$. Определено, что выполняются неравенства $\overline{\eta_{xy}} \geq |\overline{r_{xy}}|$ и $\overline{\eta_{yx}} \geq |\overline{r_{yx}}|$. На основе полученных значений выборочного корреляционного отношения можно предположить, что X и Y связаны корреляционной зависимостью, но не линейной корреляционной зависимостью и не функциональной зависимостью. Характер корреляционной зависимости не определен.

Были построены корреляционные кривые параболического и степенного вида. Визуально можно сделать вывод о том, что корреляционная зависимость признаков может быть выражена параболической функцией, но в меньшей мере степенной.

ПРИЛОЖЕНИЕ А

ИСХОДНЫЙ КОД

```
#!/usr/bin/env python
```

```
# coding: utf-8
```

```
# In[1]:
```

```
import numpy as np
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
from IPython.core.interactiveshell import InteractiveShell
```

```
InteractiveShell.ast_node_interactivity = "all"
```

```
# pd.set_option('display.max_columns', None)
```

```
# pd.set_option('display.max_rows', None)
```

```
# ## Выборка
```

```
# In[2]:
```

```
df = pd.read_csv('c:/Users/gandh/dev/unv/smoed/me/data/main_data.csv')
```

```
X = df['nu']
```

```
Y = df['E']
```

```
int_rowX = pd.read_csv('c:/Users/gandh/dev/unv/smoed/me/data/inter-  
val.csv')
```

```
int_rowY = pd.read_csv('c:/Users/gandh/dev/unv/smoed/me/data/inter-  
val2.csv')
```

```
kor = pd.read_csv('c:/Users/gandh/dev/unv/smoed/me/data/kor.csv')
```

```
# In[3]:
```

```
sns.set_theme(style="whitegrid", palette='deep', context='notebook',  
font_scale=1.3)
```

```
ax = sns.relplot(data=df, x='nu', y='E', kind='scatter', height=8.27, aspect=11.7/8.27)
ax.set_axis_labels('nu', 'E')
ax.fig.suptitle('Двумерная выборка')
plt.tight_layout()
plt.savefig('pics/1.png')
```

```
# ## Прямые регрессии
```

```
# In[4]:
```

```
N = 104
xv, yv = 453.71, 129.98
sx, sy = 53.79, 22.06
r = 0.8765
```

```
# ### Прямая x на y
```

```
# In[5]:
```

```
regr_xy = lambda y: xv + r*(sx/sy)*(y-yv)
```

```
# In[6]:
```

```
ost_var_xy = (sx**2)*(1-r**2)
```

```
# ### Прямая y на x
```

```
# In[7]:
```

```
regr_yx = lambda x: yv + r*(sy/sx)*(x-xv)
```

```
# In[8]:
```

```
ost_var_yx = (sy**2)*(1-r**2)
```

```
# ### График
```

```
# In[9]:
```

```
# Регрессия Y на X
```

```
# ax = sns.lmplot(data=df, x='nu', y='E', height=8.27, aspect=11.7/8.27)
```

```
# In[10]:
```

```
ax = sns.relplot(data=df, x='nu', y='E', kind='scatter', height=8.27,  
                 aspect=11.7/8.27, s=50, label='Выборка')  
plt.plot(regr_xy(df['E']), df['E'], label='x(y)', zorder=0)  
plt.plot(df['nu'], regr_yx(df['nu']), label='y(x)', zorder=1)  
plt.scatter(xv, yv, s=60, c='crimson', zorder=2)  
ax.set_axis_labels('nu', 'E')  
ax.fig.suptitle('Прямые регрессии')  
plt.legend()  
plt.tight_layout()  
plt.savefig('pics/2.png')
```

```
# In[11]:
```

```
ost_var_xy
```

```
ost_var_yx
```

```
# ## Выборочное корреляционное отношение
```

```
# ### Таблица
```

```
# In[12]:
```

```

kor.loc[1:7,'Xi'] = [np.sum(kor.iloc[i,1:8]) for i in range(1,8)]
kor.iloc[8,1:8] = [np.sum(kor.iloc[1:8,i]) for i in range(1,8)]
kor.iloc[8,8] = 104

# ##### Средний x для данного y (условный выборочный x)

# In[13]:

kor.loc[1:7,'yX']
=[(np.dot(kor.iloc[0,1:8],kor.iloc[i,1:8])/kor.loc[i,'Xi']).round(2) for
i in range(1,8)]

# ##### Средний y для данного x (условный выборочный y)

# In[14]:

kor.iloc[9,1:8]
=[(np.dot(kor.iloc[1:8,0],kor.iloc[1:8,i])/kor.iloc[8,i]).round(2) for i
in range(1,8)]

# ##### Групповая выборочная дисперсия X

# In[15]:

kor['D_grX'] = np.NaN
for i in range(1,8):
    x0_arg_kv = kor.iloc[0,1:8]**2
    dt = np.dot(x0_arg_kv,kor.iloc[i,1:8])/kor.loc[i,'Xi']
    dt -= kor.loc[i,'yX']**2
    kor.loc[i,'D_grX'] =(dt).round(2)

# ##### Групповая выборочная дисперсия Y

# In[16]:

```

```

kor = kor.append(pd.Series(dtype='float64'), ignore_index=True)
for i in range(1,8):
    y0_arg_kv = kor.iloc[1:8,0]**2
    dt2 = np.dot(y0_arg_kv,kor.iloc[1:8,i])/kor.iloc[8,i]
    dt2 -= kor.iloc[9,i]**2
    kor.iloc[10,i] =(dt2).round(2)

```

```

# In[17]:

```

```

kor

```

```

# ### Дисперсии X к Y

```

```

# ##### Внутригрупповая дисперсия X к Y

```

```

# In[18]:

```

```

D_vngr_xy = np.dot(kor.loc[1:7,'Xi'],kor.loc[1:7,'D_grX'])/kor.iloc[8,8]
D_vngr_xy.round(4)

```

```

# ##### Межгрупповая дисперсия X к Y

```

```

# In[19]:

```

```

kv_mezh_xy = (kor.loc[1:7,'yX']-xv)**2
D_mezh_xy = np.dot(kor.loc[1:7,'Xi'],kv_mezh_xy)/kor.iloc[8,8]
D_mezh_xy.round(4)

```

```

# ##### Общая дисперсия X к Y

```

```

# In[20]:

```

```

D_obsh_xy = D_vngr_xy + D_mezh_xy
D_obsh_xy.round(4)

```



```
# ##### Выборочное корреляционное отношение X к Y
```

```
# In[21]:
```

```
eta_xy = np.sqrt(D_mezh_xy/D_obsh_xy)
```

```
eta_xy.round(4)
```

```
r
```

```
# ### Дисперсии Y к X
```

```
# ##### Внутригрупповая дисперсия Y к X
```

```
# In[22]:
```

```
D_vngr_yx = np.dot(kor.iloc[8,1:8],kor.iloc[10,1:8])/kor.iloc[8,8]
```

```
D_vngr_yx
```

```
# ##### Межгрупповая дисперсия Y к X
```

```
# In[23]:
```

```
kv_mezh_yx = (kor.iloc[9,1:8]-yv)**2
```

```
D_mezh_yx = np.dot(kor.iloc[8,1:8],kv_mezh_yx)/kor.iloc[8,8]
```

```
D_mezh_yx.round(4)
```

```
# ##### Общая дисперсия Y к X
```

```
# In[24]:
```

```
D_obsh_yx = D_vngr_yx + D_mezh_yx
```

```
D_obsh_yx.round(4)
```

```
# ##### Выборочное корреляционное отношение Y к X
```

```

# In[25]:

eta_yx = np.sqrt(D_mezh_yx/D_obsh_yx)
eta_yx.round(4)
r

# ## Корреляционные кривые

# In[26]:

kor

# ### Параболическая регрессия Y на X

# In[27]:

df_prbl_x = pd.DataFrame({'x': kor.iloc[0,1:8], 'n': kor.iloc[8,1:8], 'y':
kor.iloc[9,1:8]})

# In[28]:

for i in range(1,5):
    df_prbl_x[f'nx{i}'] = df_prbl_x['n']*(df_prbl_x['x']**i)
df_prbl_x['ny'] = df_prbl_x['n']*df_prbl_x['y']
df_prbl_x['nyx1'] = df_prbl_x['nx1']*df_prbl_x['y']
df_prbl_x['nyx2'] = df_prbl_x['nx2']*df_prbl_x['y']

df_prbl_xf = df_prbl_x.append(df_prbl_x.sum(), ignore_index=True)
df_prbl_xf.iloc[-1,[0,2]] = np.NaN
df_prbl_xf.to_csv('data/parabolxy.csv', index=False)
df_prbl_xf

```

```

M1 = np.array([[df_prbl_xf.loc[7, 'nx4'], df_prbl_xf.loc[7, 'nx3'], df_prbl_xf.loc[7, 'nx2']],

[ df_prbl_xf.loc[7, 'nx3'], df_prbl_xf.loc[7, 'nx2'], df_prbl_xf.loc[7, 'nx1']]

,

[ df_prbl_xf.loc[7, 'nx2'], df_prbl_xf.loc[7, 'nx1'], df_prbl_xf.loc[7, 'n']]])
v1 = np.array([df_prbl_xf.loc[7, 'nyx2'], df_prbl_xf.loc[7, 'nyx1'], df_prbl_xf.loc[7, 'ny']])
a, b, c = np.linalg.solve(M1, v1)
parab_regr = lambda x: a*x*x+b*x+c
a, b, c

ax = sns.relplot(data=df, x='nu', y=parab_regr(df['nu']), kind='line',
linewidth=2.5,
height=8.27, aspect=11.7/8.27, label='y(x)',
color='crimson')
plt.scatter(df['nu'], df['E'], s=50, label='Выборка')
ax.set_axis_labels('nu', 'E')
ax.fig.suptitle('Параболическая регрессия')
plt.legend()
plt.tight_layout()
plt.savefig('pics/3.png')

# ### Степенная регрессия Y на X

df_step_x = pd.DataFrame({'x': kor.iloc[0,1:8], 'n': kor.iloc[8,1:8], 'y':
kor.iloc[9,1:8]})

df_step_x['log_x'] = np.log(df_step_x['x'])
df_step_x['log_x2'] = np.log(df_step_x['x'])**2
df_step_x['log_y'] = np.log(df_step_x['y'])

```

```

df_step_x['log_x_log_y'] = df_step_x['log_x']*df_step_x['log_y']

df_step_xf = df_step_x.append(df_step_x.sum(), ignore_index=True)
df_step_xf.iloc[-1,[0,2]] = np.NaN
df_step_xf.round(3).to_csv('data/stepxy.csv', index=False)
df_step_xf
M1 = np.array([[df_step_xf.loc[7,'n'],df_step_xf.loc[7,'log_x']],
               [df_step_xf.loc[7,'log_x'],df_step_xf.loc[7,'log_x2']]])
v1 = np.array([df_step_xf.loc[7,'log_y'],df_step_xf.loc[7,'log_x_log_y']])
a2, b2 = np.linalg.solve(M1, v1)
step_regr = lambda x: np.exp(a2)*(x**b2)
np.exp(a2), b2, a2
dfst = df.copy()
dfst['1'] = parab_regr(dfst['nu'])
dfst['2'] = step_regr(dfst['nu'])
dfstm = dfst.melt(id_vars='nu', value_vars=['1','2'])
dfstm
ax = sns.relplot(data=dfstm, x='nu', y='value', hue='variable',
kind='line', linewidth=2.5,
                height=8.27, aspect=11.7/8.27)
plt.scatter(df['nu'], df['E'], s=50, label='Выборка')
ax.set_axis_labels('nu', 'E')
plt.legend()
ax = sns.relplot(data=df, x='nu', y=step_regr(df['nu']), kind='line', linewidth=2.5,
                height=8.27, aspect=11.7/8.27, label='y(x)',
color='crimson')
plt.scatter(df['nu'], df['E'], s=50, label='Выборка')
ax.set_axis_labels('nu', 'E')
ax.fig.suptitle('Степенная регрессия')
plt.legend()
plt.tight_layout()
plt.savefig('pics/4.png')

```