

Санкт-Петербургский Государственный Университет

**Поляков Иван Михайлович**

**Отчёт по Лабораторной Работе № 10\_1**

***Парная линейная регрессия***

Направление 01.04.02: «Прикладная математика и информатика»  
Образовательная программа ВМ.5505.2021: «Математическое и информационное  
обеспечение экономической деятельности»

Преподаватель:  
доктор технических наук,  
профессор Буре Владимир Мансурович

Санкт-Петербург  
2022 г.

# 1 Ход работы

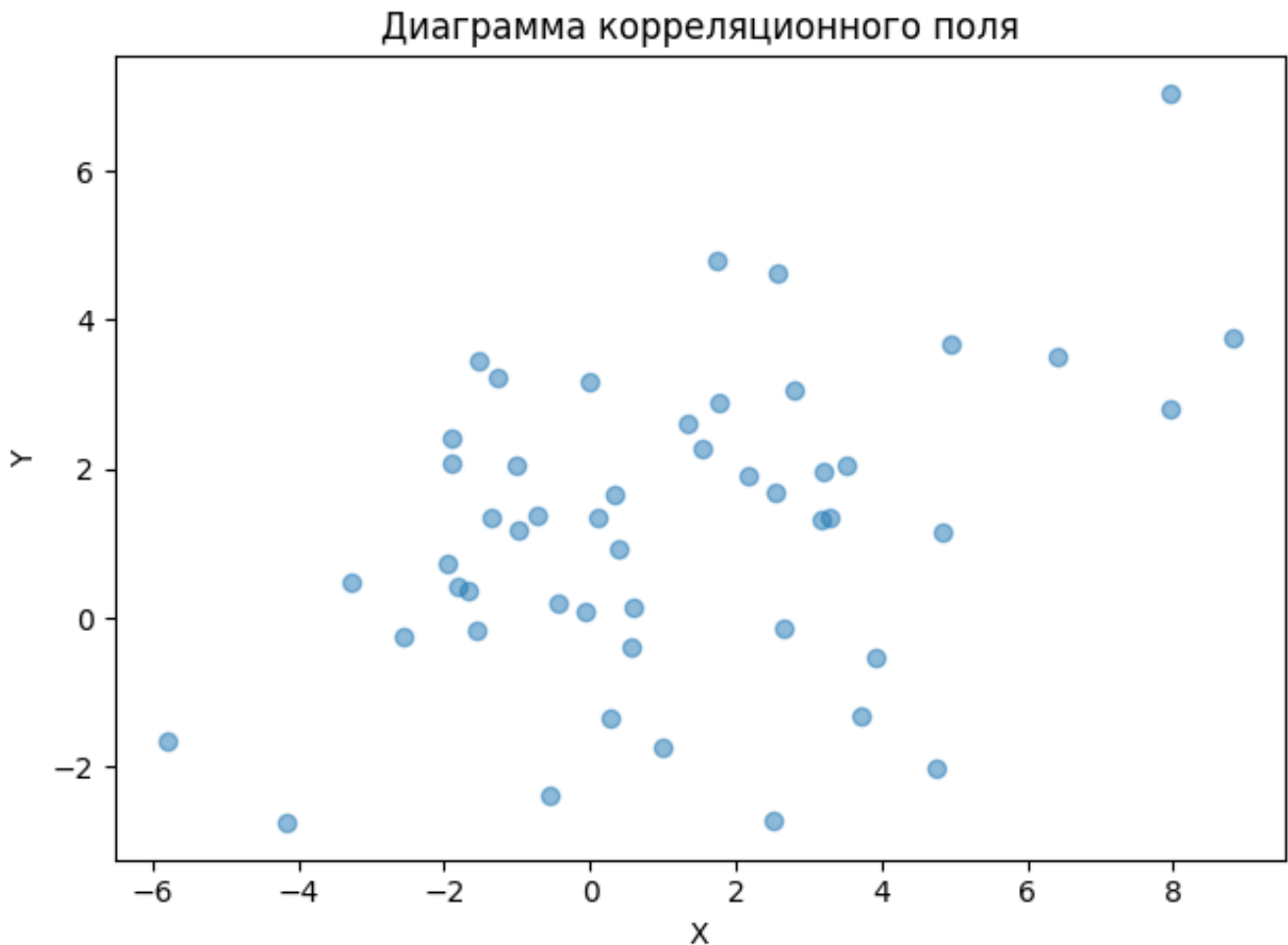
Все результаты исследований приведены как вывод из консоли в конце отчёта.

Изначально была сгенерирована выборка  $\epsilon_1, \dots, \epsilon_n$ ,  $n = 50$  из нормального распределения с  $a = 0$  и  $\sigma = 2$ . Коэффициенты  $\alpha$  и  $\beta$  были получены случайно из промежутка  $[0, 1]$ . Также был сгенерирован случайный числовой набор  $x_1, \dots, x_n$  со случайными параметрами математического ожидания и стандартного отклонения. После чего была сформирована выборка наблюдений  $y_1, \dots, y_n$  следующим образом:

$$y_i = \alpha + \beta \cdot x_i + \epsilon_i, i = \overline{1, n} \quad (1)$$

Таким образом, исходные данные для последующего анализа являются значения показателя  $y$  и объясняющей переменной  $x$ .

Была построена диаграмма зависимости показателя  $y$  от признака фактора  $x$ :



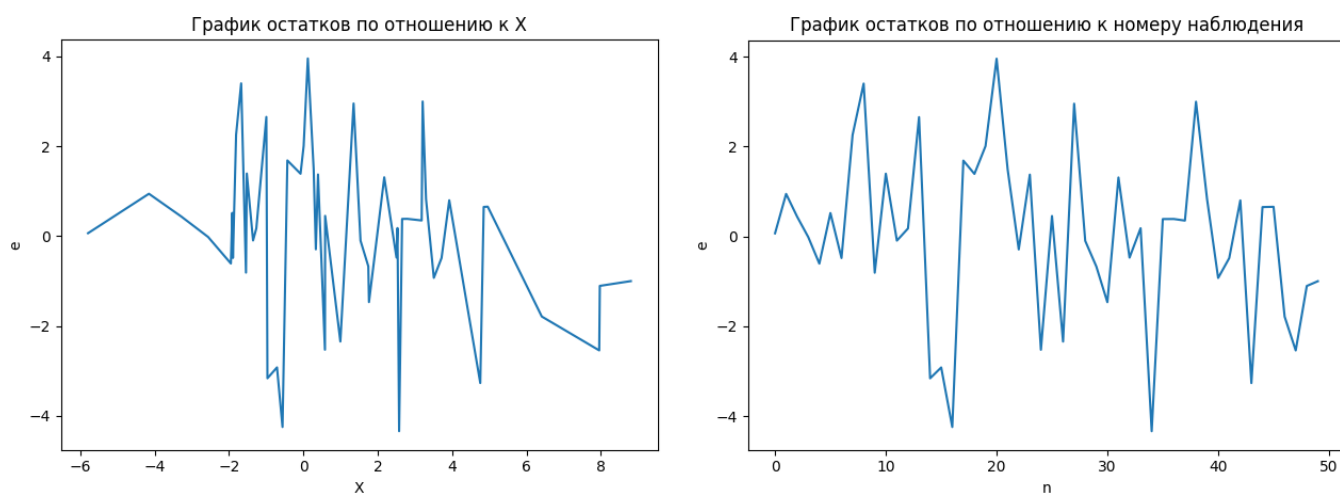
В соответствии с заданием были вычислены:

- Коэффициенты линейной регрессии
- Коэффициент корреляции
- Расчётные значения  $\hat{y}_i = a + b \cdot x_i$ ,  $i = \overline{1, n}$
- Отклонения  $e_i = y_i - \hat{y}_i$ ,  $i = \overline{1, n}$  истинных значений признака от расчётных
- Средняя квадратическая ошибка аппроксимации

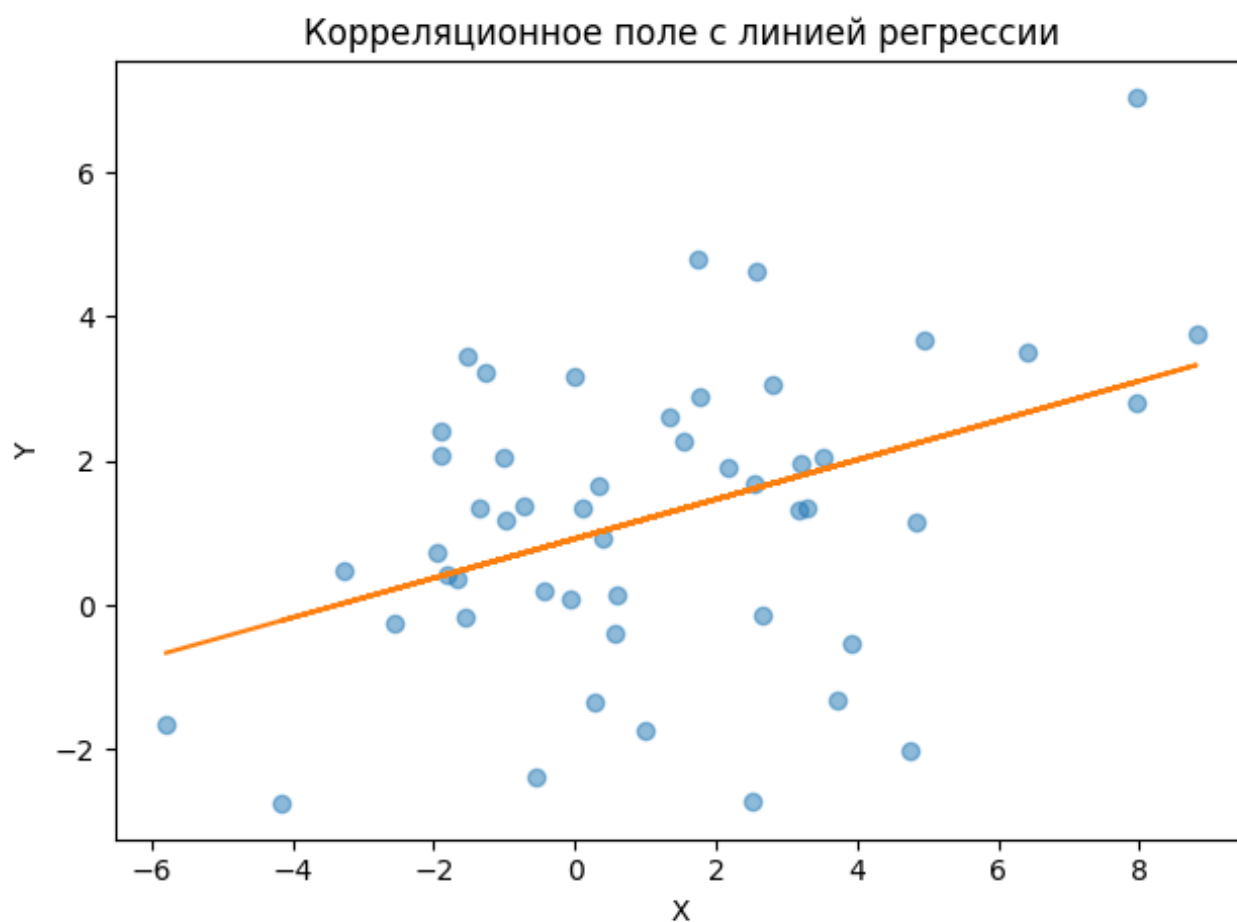
- Значения  $t$ -статистики для коэффициента корреляции и оценок  $a$  и  $b$ . Также были сделаны выводы о статистической значимости данных коэффициентов.
- Доверительные интервалы с уровнем значимости 5% для параметров  $\alpha$  и  $\beta$  линейной регрессии
- Стандартная ошибка  $m_y$  прогноза индивидуального значения
- Доверительный интервал полученного прогноза с уровнем значимости 5%

Также были построены следующие графики:

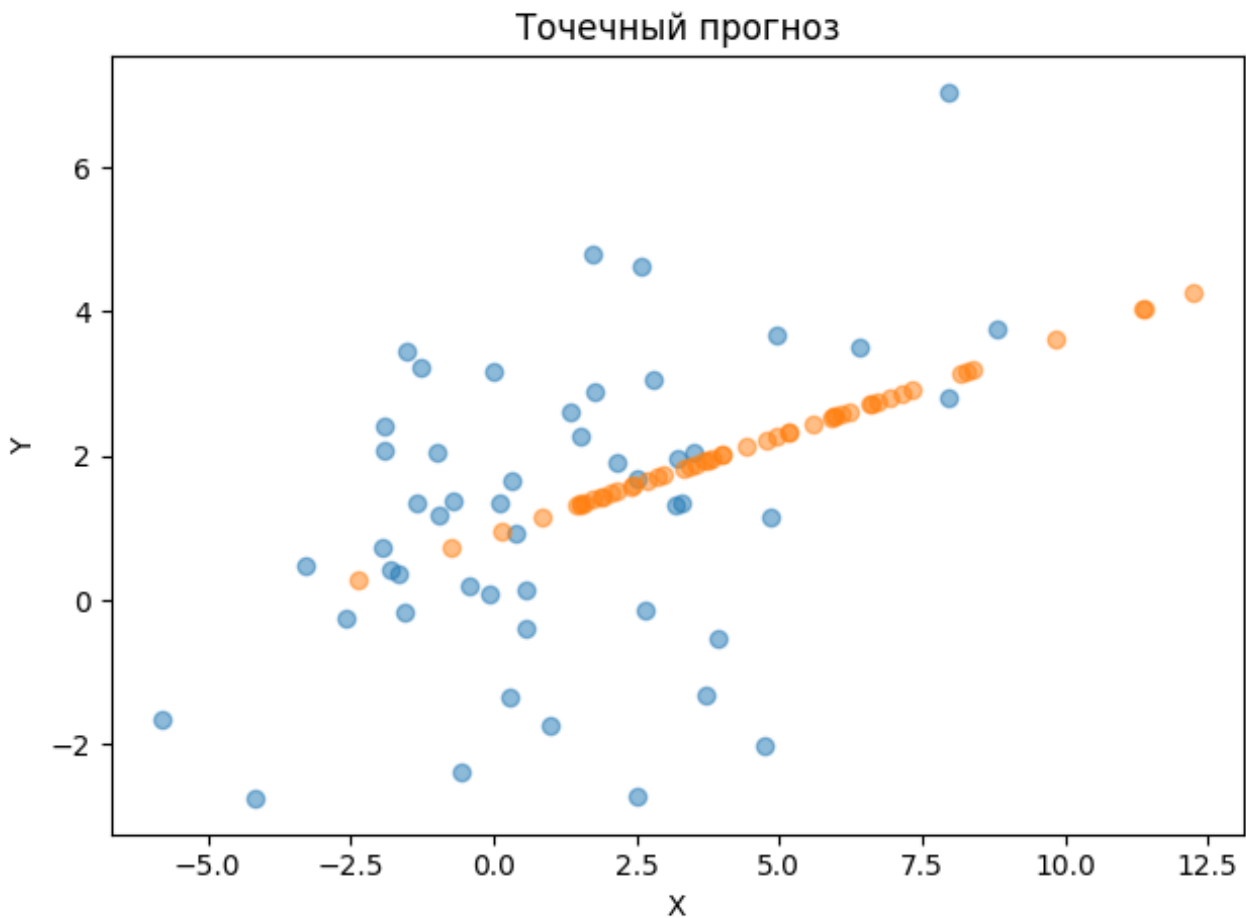
- График остатков  $e$  по отношению к фактору  $x$  и по отношению к номеру наблюдения



- Совместный график корреляционного поля и прямой линейной регрессии:



- Точечный прогноз  $\hat{y}_p = a + b \cdot x_p$ , где  $x_p$  на 200% больше, чем средний уровень  $\bar{x}$



Все вычисленные результаты представлены ниже.

Исходные данные:

epsilons:

```
[ 0.245  1.137  0.521  0.225 -0.385  0.749 -0.314  2.472  3.59 -0.593
 1.625  0.145  0.343  2.888 -2.936 -2.717 -4.099  1.929  1.532  2.253
 4.052  1.677 -0.193  1.548 -2.367  0.716 -2.131  3.19  0.11 -0.431
-1.259  1.506 -0.22  0.345 -4.159  0.599  0.57  0.595  3.174  0.946
-0.72  -0.317  1.035 -3.106  0.862  0.883 -1.61  -2.263 -0.961 -0.697]
```

alpha: 0.704

beta: 0.288

xs:

```
[ 2.531e+00  1.536e+00  8.808e+00 -1.810e+00 -4.310e-01 -9.670e-01
 3.181e+00  7.000e-03  1.746e+00 -7.600e-02 -9.960e-01 -1.674e+00
 3.207e+00 -1.266e+00 -5.620e-01  9.940e-01  4.757e+00 -1.895e+00
 4.958e+00 -1.912e+00  7.957e+00  1.764e+00  7.973e+00  2.798e+00
 3.923e+00 -3.281e+00  2.760e-01 -1.522e+00  3.930e-01 -1.544e+00
 5.800e-01  1.350e+00 -2.563e+00  3.509e+00  2.503e+00  1.200e-01
 2.175e+00 -1.951e+00  2.575e+00  6.412e+00  5.850e-01  3.298e+00
-1.354e+00  3.716e+00  3.350e-01 -7.100e-01  2.656e+00 -4.155e+00
 4.850e+00 -5.792e+00]
```

ys = alpha + beta \* xs + epsilons:

```
[ 1.678  2.284  3.762  0.408  0.195  1.174  1.306  3.178  4.797  0.089
 2.042  0.367  1.971  3.227 -2.394 -1.727 -2.024  2.087  3.664  2.406]
```

```

7.048 2.889 2.807 3.058 -0.533 0.476 -1.347 3.456 0.927 -0.172
-0.388 2.599 -0.254 2.059 -2.734 1.337 1.9 0.737 4.62 3.496
0.153 1.337 1.349 -1.332 1.662 1.383 -0.141 -2.755 1.14 -1.661]

```

Найденное уравнение линейной регрессии:

$$y = 0.921 + (0.273) \cdot X$$

Коэффициент корреляции:

0.411

Коэффициент детерминации:

0.169

Средняя ошибка аппроксимации:

1.46

Результаты F-статистики:

#### OLS Regression Results

```

=====
Dep. Variable:          y      R-squared:          0.169
Model:                  OLS    Adj. R-squared:      0.152
Method:                 Least Squares    F-statistic:      9.781
Date:                   Mon, 27 Jun 2022    Prob (F-statistic): 0.00299
Time:                   16:35:57    Log-Likelihood:   -101.95
No. Observations:       50    AIC:              207.9
Df Residuals:           48    BIC:              211.7
Df Model:                1
Covariance Type:        nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	0.9205	0.286	3.216	0.002	0.345	1.496
x1	0.2732	0.087	3.127	0.003	0.098	0.449

```

=====
Omnibus:                0.979    Durbin-Watson:      1.806
Prob(Omnibus):           0.613    Jarque-Bera (JB):    0.741
Skew:                    -0.296    Prob(JB):            0.690
Kurtosis:                2.933    Cond. No.            3.54
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified

Значения t-статистик:

для коэффициента корреляции: 0.0541

для alpha: 3.22

для beta: 3.13

Коэффициент корреляции статистически не значим,

так как  $t_{corr} \leq t_{crit}$  ( $0.0541 \leq 1.68$ )

Коэффициент a статистически значим,

так как  $t_a > t_{crit}$  ( $3.22 > 1.68$ )

Коэффициент b статистически значим,

так как  $t_b > t_{crit}$  ( $3.13 > 1.68$ )

Доверительные интервалы:

для  $\alpha$ :  $[0.345 \ 1.496]$

для  $\beta$ :  $[0.098 \ 0.449]$

Значение точечного прогноза:

1.86

Ошибка точечного прогноза:

3.51

Доверительный интервал для прогноза:

$(-4.03; 7.74)$