

Санкт-Петербургский Государственный Университет

Поляков Иван Михайлович

Отчёт по Лабораторной Работе № 10_3

Множественная 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

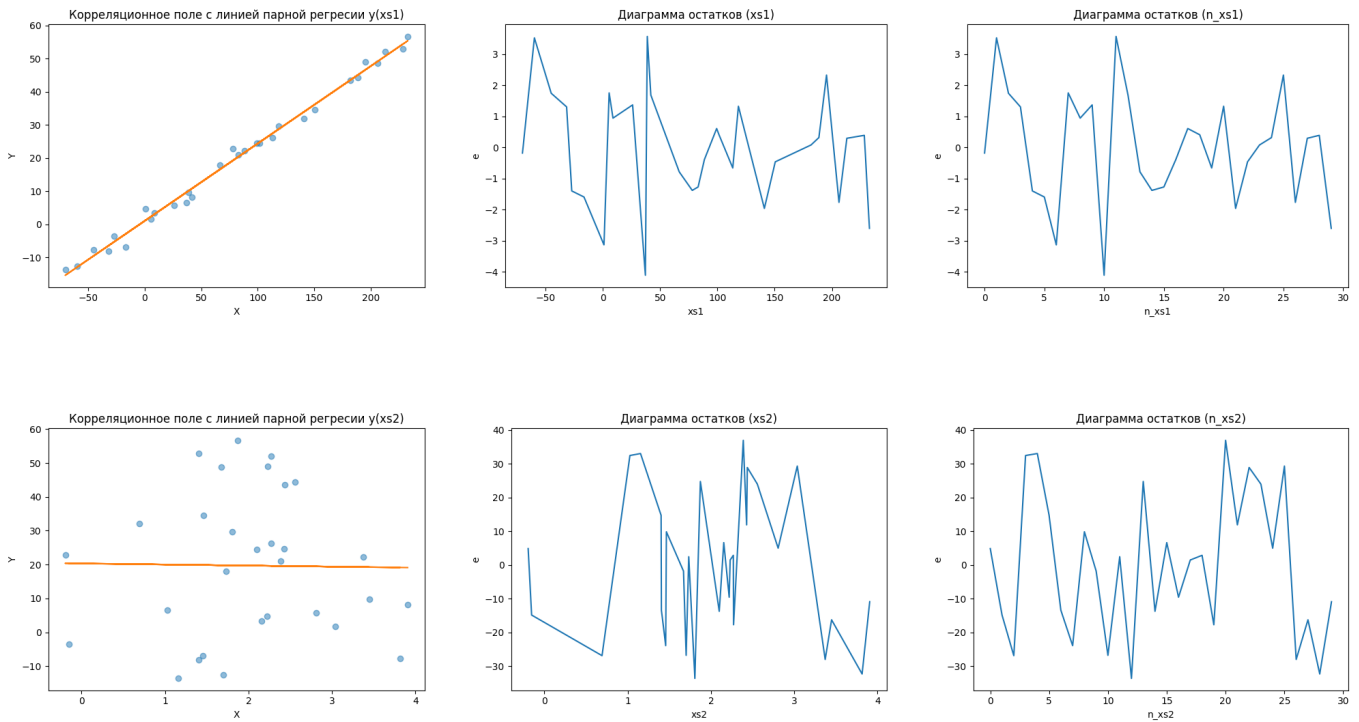
Все результаты исследований приведены как вывод из консоли в конце отчёта.

Изначально была сгенерирована выборка $\epsilon_1, \dots, \epsilon_n$, $n = 30$ из нормального распределения с $a = 0$ и $\sigma = 2$. Коэффициенты α , β_1 и β_2 были получены случайно из промежутка $[0, 1]$. Также был сгенерированы случайные числовые наборы x_{11}, \dots, x_{1n} , x_{21}, \dots, x_{2n} со случайными параметрами математического ожидания и стандартного отклонения. После чего была сформирована выборка наблюдений y_1, \dots, y_n следующим образом:

$$y_i = \alpha + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \epsilon_i, i = \overline{1, n} \quad (1)$$

Таким образом, исходные данные для последующего анализа являются значения показателя y и объясняющих переменных x_1 и x_2 .

Были построены парные линейные регрессии – зависимости результативного признака y от факторов, взятых по отдельности. Для каждой регрессии были построены поле корреляции с линией парной линейной регрессии и диаграммы остатков (остатки на фактор и остатки на номера наблюдений), а также вычислены коэффициенты уравнения линейной регрессии, коэффициент детерминации, коэффициент корреляции, величина средней ошибки аппроксимации.

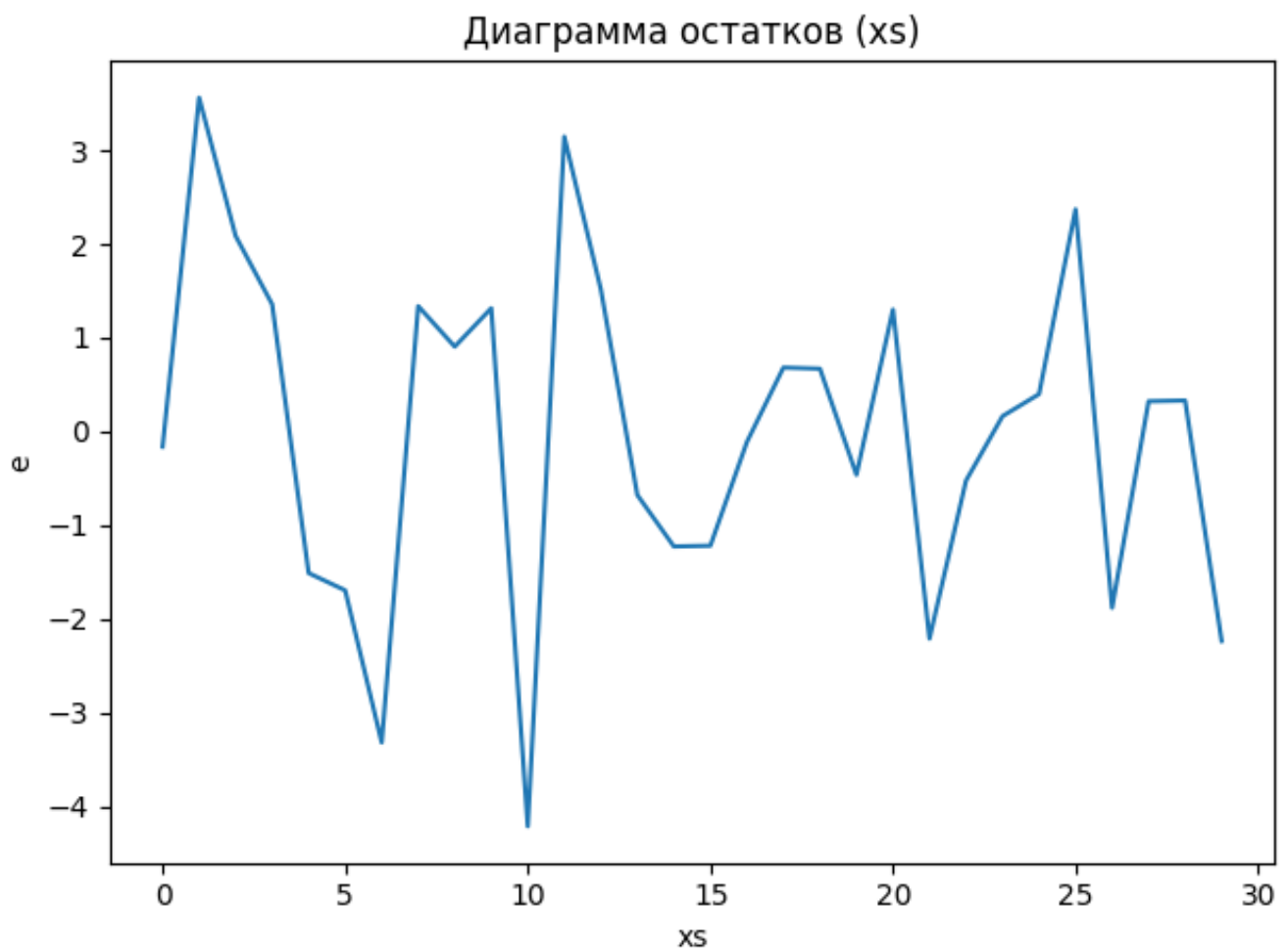


В соответствии с заданием были вычислены следующие параметры множественной регрессии:

- 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
- Величина средней ошибки аппроксимации
- Оценка для дисперсии остатков
- Множественный коэффициент детерминации

- Фактическое значение F критерия
- Частные коэффициенты корреляции в случае двух переменных
- Значения t -статистики для оценок a , b_1 и b_2 . Также были сделаны выводы о статистической значимости данных коэффициентов.
- Точечный прогноз для индивидуального значения
- Интервальный прогноз

Также была построена диаграмма остатков на номера наблюдений для множественной регрессии:



Все вычисленные результаты представлены ниже.

Исходные данные:

epsilons:

```
[ -0.64   3.304   1.238   0.5    -2.022  -2.022  -3.142   2.217   0.512   1.096
 -4.084   3.759   1.928  -1.597  -1.817  -1.805  -1.024   0.124  -0.343  -1.1
  0.568  -2.165  -1.094  -0.683  -0.221   1.58   -1.686   0.071   0.468  -3.361]
```

alpha: 0.722

beta1: 0.236

beta2: 0.256

xs1:

```
[101.38    0.928 -45.123 213.    228.252 150.504  37.084 -27.184 118.325]
```

```

66.652 -16.518 78.009 -70.23 188.6 25.937 113.387 38.738 83.127
88.623 5.444 232.727 140.98 206.148 181.715 99.329 195.259 -31.708
8.803 -59.826 41.867]
xs2:
[ 2.1 2.219 3.815 2.272 1.405 1.465 1.027 -0.154 1.807 1.733
1.457 -0.195 1.154 2.557 2.807 2.269 3.449 2.388 3.372 3.037
1.873 0.692 1.67 2.437 2.429 2.231 1.401 2.154 1.702 3.907]
ys = alpha + beta1 * xs1 + beta2 * xs2 + epsilon:
[ 24.545 4.813 -7.713 52.071 52.927 34.594 6.594 -3.516 29.622
17.991 -6.887 22.841 -13.629 44.29 5.745 26.257 9.722 21.075
22.157 1.685 56.693 32.005 48.706 43.547 24.565 48.954 -8.088
3.422 -12.494 8.242]

```

Парные линейные регрессии:

Коэффициенты:

ys/xs1 : 0.233, 1.07

ys/xs2: -0.309, 20.3

Коэффициент детерминации:

xs1: 0.993

xs2: 0.000202

Коэффициент корреляции:

xs1:

[[1. 0.996]

[0.996 1.]]

xs2:

[[1. -0.014]

[-0.014 1.]]

Средняя ошибка аппроксимации:

xs1: 0.154

xs1: 1.8

Множественная линейная регрессия

Коэффициенты:

ys/xs - b1, b2, a: [0.233 -0.192], [1.46]

Уравнение регрессии в развернутой форме:

$y = 1.46 + (0.233) * x1 + (-0.192) * x2$

Коэффициент детерминации:

0.993

Значения F-статистики

OLS Regression Results

```

=====
Dep. Variable:          0    R-squared:          0.993
Model:                OLS    Adj. R-squared:      0.992
Method:               Least Squares    F-statistic:      1869.
Date:                 Mon, 27 Jun 2022    Prob (F-statistic): 1.12e-29

```

```

Time:                18:48:51    Log-Likelihood:        -59.917
No. Observations:    30    AIC:                125.8
Df Residuals:        27    BIC:                130.0
Df Model:            2
Covariance Type:      nonrobust

```

```

=====
              coef      std err          t      P>|t|      [0.025      0.975]
-----
const          1.4595        0.850        1.718      0.097      -0.284        3.203
x1              0.2333        0.004       61.132      0.000        0.225        0.241
x2            -0.1917        0.354       -0.541      0.593      -0.918        0.535
=====
Omnibus:                0.406    Durbin-Watson:           2.069
Prob(Omnibus):           0.816    Jarque-Bera (JB):         0.328
Skew:                   -0.233    Prob(JB):                 0.849
Kurtosis:               2.786    Cond. No.                 316.
=====

```

Notes:

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

Средняя ошибка аппроксимации:
0.146

Оценка для дисперсии остатков
3.53

Частные коэффициенты корреляции
r_{yx1|x2}: 0.994
r_{yx2|x1}: -0.103

Стандартные ошибки коэффициентов регрессии
a: 0.85
b1: 0.00382
b2: 0.354

Коэффициент a статистически значим,
так как $t_a > t_{crit}$ ($1.72 > 1.71$)
Коэффициент b1 статистически значим,
так как $t_{b1} > t_{crit}$ ($61.1 > 1.71$)
Коэффициент b2 статистически не значим,
так как $t_{b2} \leq t_{crit}$ ($-0.541 \leq 1.71$)

Точечный прогноз

Значение: 56.2

Интервальный прогноз: [54.931 57.379]