Министерство образования Республики Беларусь

Учреждение Образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Кафедра ИИТ

Лабораторная работа № 2

По дисциплине «Статистические основы индуктивного вывода»

По теме «Анализ зависимостей. Корреляционный и регрессионный анализ. Парная корреляция»

Выполнил: Ахроров М. Д.

Принял: Федосенко В. А.

Минск 2022

# Цель работы

Изучить основные понятия регрессионных моделей, ознакомиться с методикой построения регрессионных моделей.

# Ход работы

Данная работа состоит из двух этапов:

* Составление регрессии с помощью средств Microsoft Excel
* Составление регрессии с помощью средств Python

**Этап 1**

На основе имеющихся данных произвести регрессионный анализ.

|  |  |
| --- | --- |
| Электрическая активность сетчатки  (независимая переменная) | Проницаемость сосудов  сетчатки  (зависимая переменная) |
| 0 | 19,5 |
| 38,5 | 15 |
| 59 | 13,5 |
| 97,4 | 23,3 |
| 119,2 | 6,3 |
| 129,5 | 2,5 |
| 198,7 | 13 |
| 248,7 | 1,8 |
| 318 | 6,5 |
| 438,5 | 1,8 |

Рис 1. Имеющиеся данные.

0

5

10

15

20

25

0

50

100

150

200

250

300

350

400

450

500

Рис 2. График с имеющимися значениями

|  |  |  |
| --- | --- | --- |
|  | *Столбец 1* | *Столбец 2* |
| Столбец 1 | 1 |  |
| Столбец 2 | -0,677051424 | 1 |

Рис 3. Таблица корреляции.

|  |  |
| --- | --- |
| Множественный R | 0,677051424 |
| R-квадрат | 0,458398631 |
| Нормированный Rквадрат | 0,39069846 |
| Стандартная ошибка | 5,984632632 |
| Наблюдения | 10 |

Рис 4. Регрессионная статистика.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 𝑑𝑓 | 𝑆𝑆 | 𝑀𝑆 | 𝐹 | *Значимость* 𝐹 |
| Регрессия | 1 | 242,509378 | 242,509378 | 6,771011402 | 0,031513974 |
| Остаток | 8 | 286,526622 | 35,81582775 |  |  |
| Итого | 9 | 529,036 |  |  |  |

Рис 5. Дисперсионный анализ.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Коэффици енты* | *Стандар тная ошибка* | *t-*  *статис тика* | *P-*  *Значени*  *е* | *Нижние 95%* | *Верхние 95%* | *Нижние 98,0%* | *Верхние 98,0%* |
| Yпересеч ение | 16,565639 76 | 3,056569 184 | 5,41968 4217 | 0,00063 1037 | 9,51717 8588 | 23,6141 0094 | 7,71241 1075 | 25,4188 6845 |
| Переме  нная X 1 | -  0,0379098  01 | 0,014568 832 | -  2,60211  6716 | 0,03151 3974 | -  0,07150  5588 | -  0,00431  4014 | -  0,08010  7832 | 0,00428 823 |

Рис 6. Дисперсионный анализ.

Полученное уравнение регрессии: 16,5656397648245 − 0,0379098013039422 ∙ 𝑥

Корреляция: −0,677051424

|  |  |  |  |
| --- | --- | --- | --- |
| Электрическая активность сетчатки  (независимая переменная) | Проницаемость сосудов  сетчатки  (зависимая переменная) | Предсказание | Ошибка |
| 0 | 19,5 | 16,56563976 | 8,61046999 |
| 38,5 | 15 | 15,10611241 | 0,011259845 |
| 59 | 13,5 | 14,32896149 | 0,687177148 |
| 97,4 | 23,3 | 12,87322512 | 108,7176344 |
| 119,2 | 6,3 | 12,04679145 | 33,02561196 |
| 129,5 | 2,5 | 11,6563205 | 83,83820502 |
| 198,7 | 13 | 9,032962246 | 15,73738854 |
| 248,7 | 1,8 | 7,137472181 | 28,48860928 |
| 318 | 6,5 | 4,51032295 | 3,958814763 |
| 438,5 | 1,8 | -0,057808107 | 3,451450962 |
|  |  | Сумма ошибок | 286,526622 |

Рис 7. Результаты предсказания

y =

-

16,566

,0379x +

0

y =

-

0

,0379x +

16,566

-5

0

5

10

15

20

25

0

100

200

300

400

500

Эталон

Предсказание

Линейная (Эталон)

Линейная (Предсказание)

Рис 8. График с результатами.

На графике видно, что линия тренда совпадает с предсказаниями, составленными на основе полученного уравнения.

После тщательного изучения набора данных было принято решение о принятии значений (97,4; 23,3) и (129,5; 2,5) за выбросы.

0

5

10

15

20

25

0

50

100

150

200

250

300

350

400

450

500

Рис 9. Исходные данные и выбросы.

|  |  |  |
| --- | --- | --- |
|  | *Столбец 1* | *Столбец 2* |
| Столбец 1 | 1 |  |
| Столбец 2 | -0,828592694 | 1 |

Рис 10. Таблица корреляции.

|  |  |
| --- | --- |
| Множественный R | 0,828592694 |
| R-квадрат | 0,686565853 |
| Нормированный Rквадрат | 0,634326829 |
| Стандартная ошибка | 3,93276864 |
| Наблюдения | 8 |

Рис 11. Регрессионная статистика.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 𝑑𝑓 | 𝑆𝑆 | 𝑀𝑆 | 𝐹 | *Значимость* 𝐹 |
| Регрессия | 1 | 203,2749849 | 203,2749849 | 13,14277707 | 0,011027029 |
| Остаток | 6 | 92,80001506 | 15,46666918 |  |  |
| Итого | 7 | 296,075 |  |  |  |

Рис 12. Дисперсионный анализ.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Коэффици енты* | *Стандар тная ошибка* | *t-*  *статис тика* | *P-*  *Значени*  *е* | *Нижние 95%* | *Верхние 95%* | *Нижние 98,0%* | *Верхние 98,0%* |
| Yпересеч ение | 16,565639 76 | 3,056569 184 | 5,41968 4217 | 0,00063 1037 | 9,51717 8588 | 23,6141 0094 | 7,71241 1075 | 25,4188 6845 |
| Переме  нная X 1 | -  0,0379098  01 | 0,014568 832 | -  2,60211  6716 | 0,03151 3974 | -  0,07150  5588 | -  0,00431  4014 | -  0,08010  7832 | 0,00428 823 |

Рис 13. Дисперсионный анализ.

Полученное уравнение регрессии: 15,9720622449423 − 0,0354614233137677 ∙ 𝑥

Корреляция: −0,82

|  |  |  |  |
| --- | --- | --- | --- |
| Электрическая активность сетчатки  (независимая переменная) | Проницаемость сосудов  сетчатки  (зависимая переменная) | Предсказание | Ошибка |
| 0 | 19,5 | 15,97206224 | 12,4463448 |
| 38,5 | 15 | 14,60679745 | 0,154608247 |
| 59 | 13,5 | 13,87983827 | 0,144277111 |
| 97,4 | 23,3 | 12,51811961 | 116,2489447 |
| 119,2 | 6,3 | 11,74506059 | 29,64868478 |
| 129,5 | 2,5 | 11,37980793 | 78,8509888 |
| 198,7 | 13 | 8,925877432 | 16,5984747 |
| 248,7 | 1,8 | 7,152806267 | 28,65253493 |
| 318 | 6,5 | 4,695329631 | 3,25683514 |
| 438,5 | 1,8 | 0,422228122 | 1,898255348 |
|  |  | Сумма ошибок | 287,8999485 |

Рис 14. Результаты предсказания

y =

-

16,566

,0379x +

0

y =

-

,0379x +

0

16,566

y =

-

0

,0355x +

15,972

-5

0

5

10

15

20

25

0

100

200

300

400

500

Эталон

Старое предсказание

Новое предсказание

Линейная (Эталон)

Линейная (Старое

предсказание)

Линейная (Новое

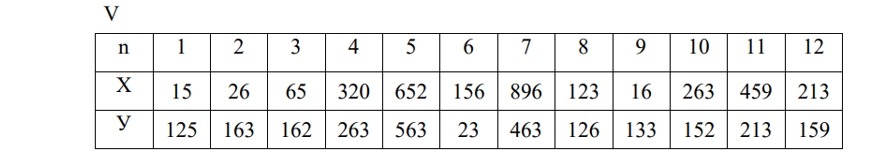
предсказание)

Рис 15. График с результатами.

После удаления из набора данных двух значений, принятых за выбросы корреляция между данными возросла на 0,15 (по модулю), среднеквадратическая ошибка предсказаний выросла 0,3%.

**Этап 2. Вариант 5**

На основе имеющихся данных провести регрессионный анализ.



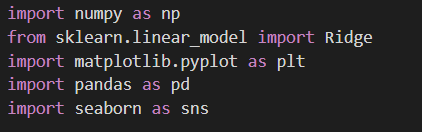


Рис. 16. Подключенные библиотеки (листинг 1).

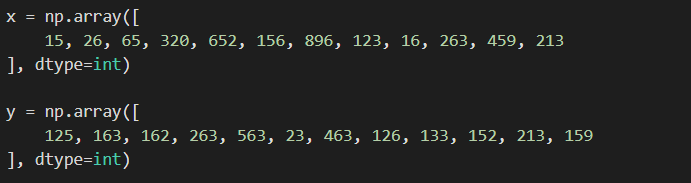


Рис. 17. Внесение данных в массив NumPy.

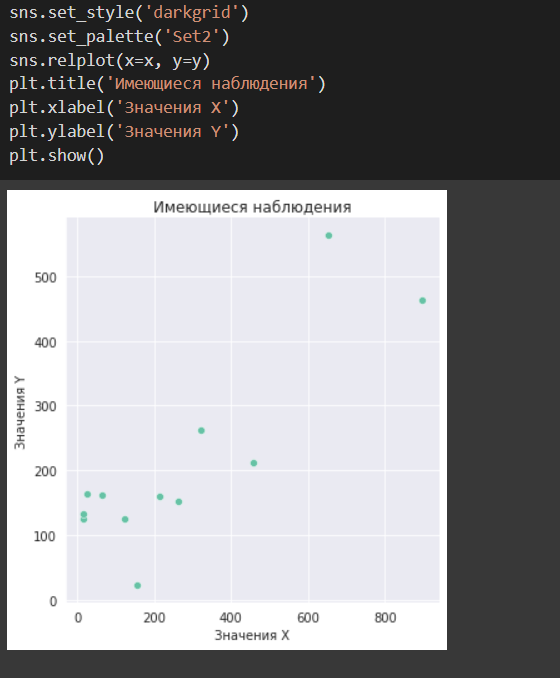


Рис. 18. Осмотр данных.

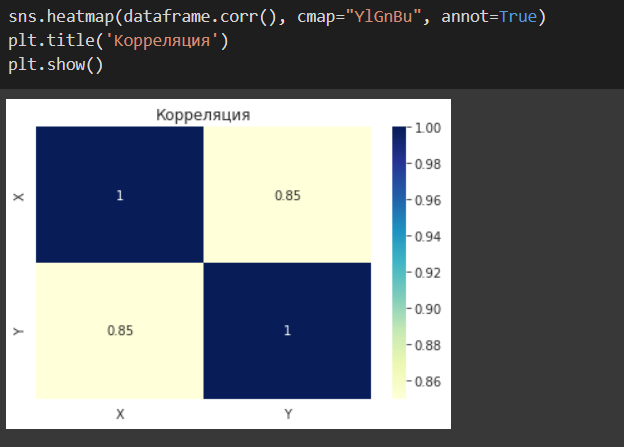


Рис. 19. Корреляция данных.

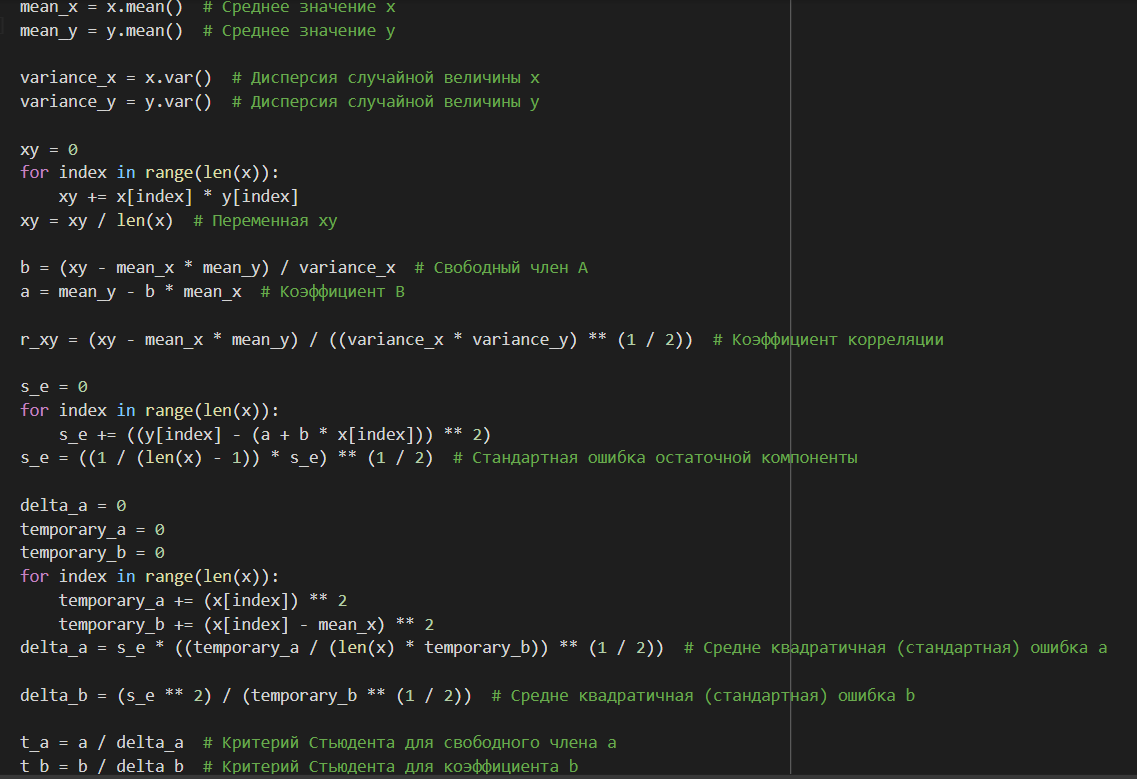


Рис. 20. Параметры для построения регрессии.

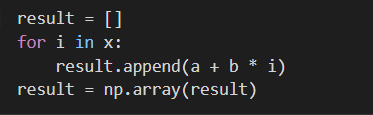


Рис. 21. Таблица с результатами предсказания значений 𝑌 на основе полученной формулы.

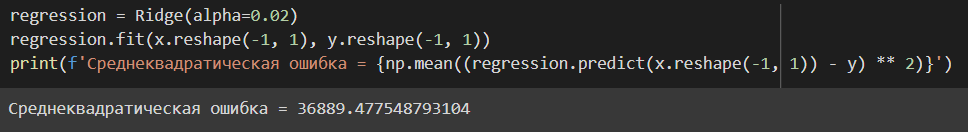


Рис. 22. Регрессионный анализ с помощью средств библиотеки sklearn.

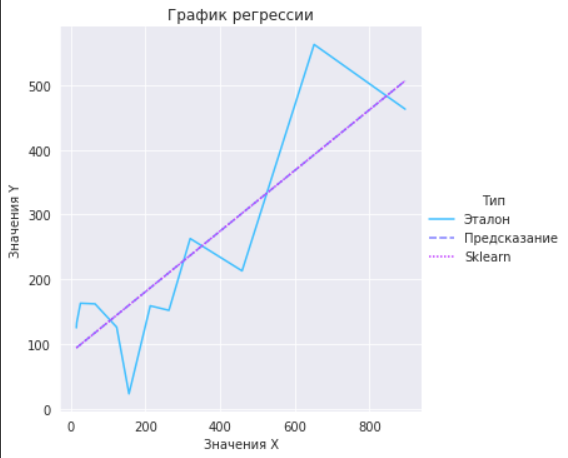


Рис. 23. График полученных значений («Предсказание» и «Sklearn» объединились).

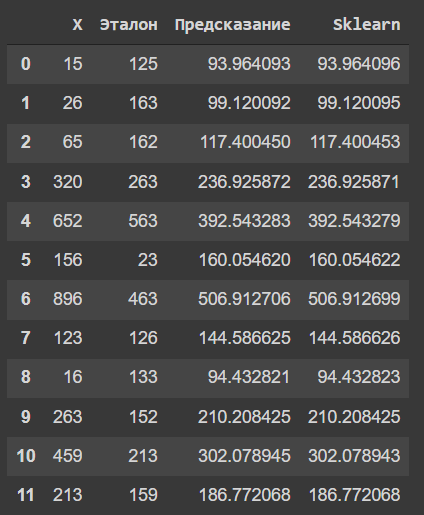


Рис. 24. Результаты в табличном виде.

**3. Выводы**

В ходе данной работы были получены навыки проектирования, решения и реализации регрессионных моделей средствами Excel и Python. Изучил основные понятия регрессионных моделей, ознакомился с методикой построения регрессионных моделей.