**Министерство образования и науки Российской Федерации**

**Федеральное государственное бюджетное образовательное учреждение**

**высшего профессионального образования**

**«Московский государственный технический университет имени Н.Э. Баумана»**

**(МГТУ им. Н.Э.Баумана)**

**Мытищинский филиал**

**ФАКУЛЬТЕТ КОСМИЧЕСКИЙ**

**КАФЕДРА К-1 САУ**

**ЛАБОРАТОРНАЯ РАБОТА №8**

**ПО ДИСЦИПЛИНЕ**

**“МАтематическая статистика”**

**НА ТЕМУ:**

# **«**КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ**»**

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

**Студент К1-41Б**  **29.04.23 К. А. Тимофеев**

(Группа) (Подпись, дата) (ФИО)

**Руководитель**  **О. М. Полещук**

(Подпись, дата) (ФИО)

*2023 г.*

**Цель:**

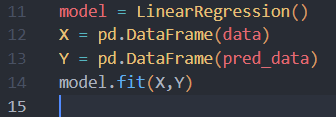
При помощи регрессионного анализа найти коэффициенты линейного уравнения регрессии .

**Решение:**

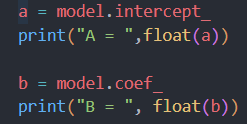
Решим задачу, написав программу для построения графика линейной регрессии на языке Python с использованием библиотек numpy, pandas, matplotlib, а также импортируем модуль Linear Regression для регрессионного анализа из библиотеки sklearn.linear\_model.



Имеется набор данных из исходных значений (xi и yi). Построим модель на основе полученных данных и обучим её:



Далее функциями intercept\_ и coef\_ найдем значения параметров a и b:

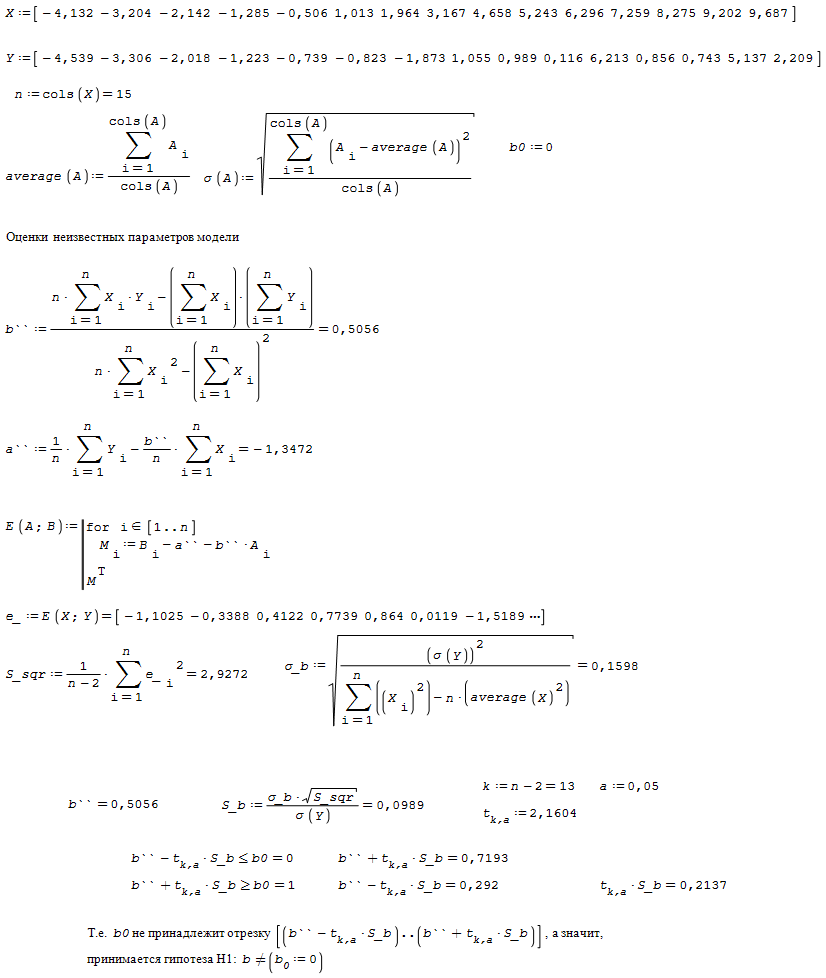


Сразу же можем получить информацию о том, насколько достоверна наша модель. Переменная R\_sqr хранит в себе значение R² - коэффициент детерминации, равный , где - это модельные значения величины y, а - среднее выборочное по исходным данным. При этом известно, что чем ближе коэффициент детерминации к 1, тем лучше построена модель и тем больше доля общей дисперсии объясняется уравнением.

 Модельные значения в свою очередь можно получить с помощью функции predict(), которая принимает значение исходных данных X.

Проверим гипотезу , конкурентом которой выступает .

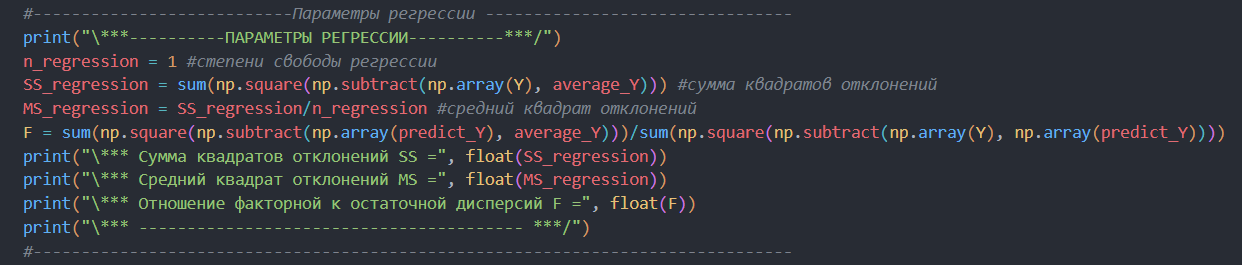
Тогда расчёт проверки основной гипотезы приведён ниже:



Взяв для построения интервала статистику , получим что, доверительный интервал для Y выглядит как:



Далее находим параметры для оценки регрессии:



Здесь ключевой является переменная F, которая также показывает успешность модели, а именно, чем больше F, тем больше факторная дисперсия, а соответственно лучше построена модель, остаточная дисперсия же указывает на

ошибки модели и в идеале должна стремиться к нулю.



здесь

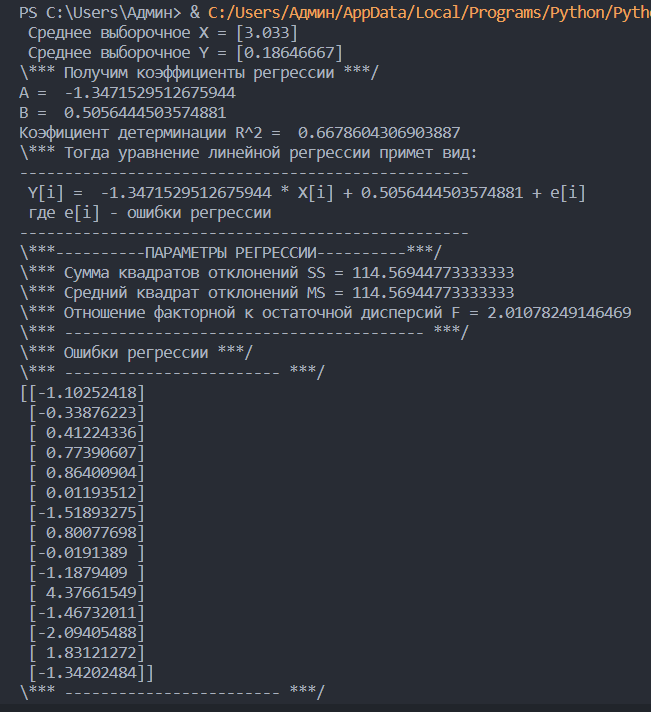
**df** - степени свободы статистики

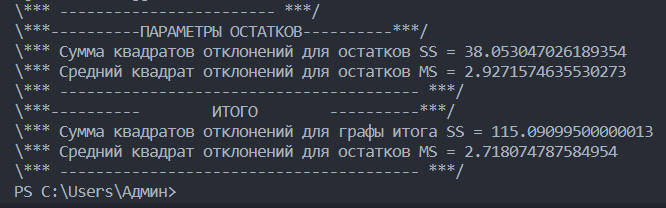
**SS** - сумма квадратов отклонений

**MS** - средний квадрат отклонений

**Значимость F** - квантиль t-критерия Стьюдента , находят из таблицы значений критерия Стьюдента.

Выведем необходимые значения и график регрессионной модели:





*Получили что:*

Коэффициент детерминации , что говорит о неплохом уровне модели, однако это не предел ожиданий.

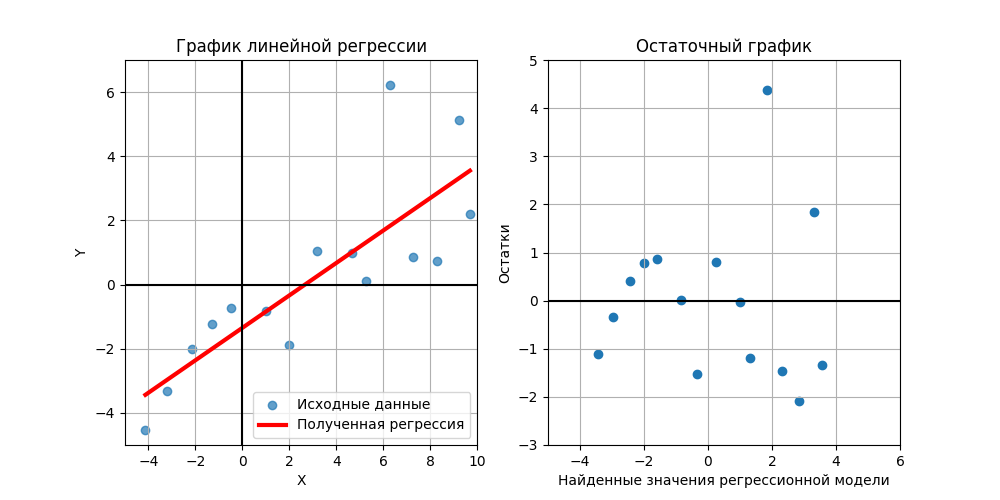
Линейное уравнение регрессии получилось в виде  с коэффициентами  и .

Зная коэффициенты, можем построить доверительный интервал для Y:





Также нашли отношение факторной к остаточной дисперсий . Однако это значение значительно меньше ожидаемого, что также не говорит о хорошей работе регрессионной модели.

Наконец-то рассмотрим полученные графики.

На графике линейной регрессии наблюдаем одновременно точки, координатами которых являются исходные xi и yi, и прямую образованную модельными значениями, найденными в ходе работы. Именно красную прямую описывает найденное уравнение линейной регрессии, иначе говоря, именно эта прямая из всех других возможных по мнению нашей модели имеет наименьшую ошибку или остаток для относительно всех точек, облако которых мы построили с помощью исходной информации.

На графике остатков (остаточный график) мы можем отчетливее пронаблюдать отклонение от прямой . Сумма этих отклонений стремится к нулю.

**Вывод:**

Произвели рассчет с помощью метода регрессионного анализа, нашли коэффициенты линейного уравнения регрессии, и построили его график.

Поскольку коэффициент детерминации оказался мал, можем сделать вывод о том, что исходные данные не ложатся на линейную регрессию, а значит имеет смысл использовать нелинейный регрессионный анализ.