

Гузенко А.М. Группа 7.2. Вариант 4

Лабораторная работа № 2

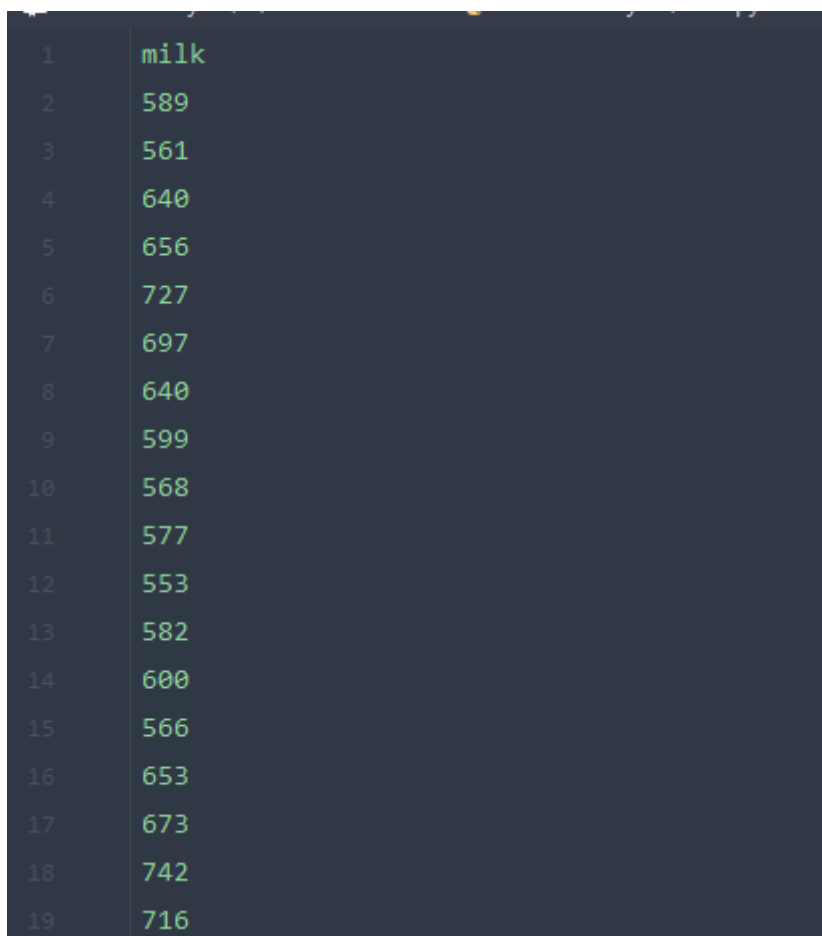
Линейная регрессия

“Количество молока в месяц”

### Цель

Провести линейный регрессионный анализ, спрогнозировать короткий временной промежуток.

### Описание данных



1	milk
2	589
3	561
4	640
5	656
6	727
7	697
8	640
9	599
10	568
11	577
12	553
13	582
14	600
15	566
16	653
17	673
18	742
19	716

Количество произведенного молока в месяц. У ряда есть тренд – с течением времени количество продукции увеличивается. Есть сезонность – чем ближе зима, тем продукции меньше, чем ближе лето, тем продукции становится больше, в мае количество продукции обычно максимальное. Данные не меняют свой характер. Выбросов в данных нет.

## Выполнение работы

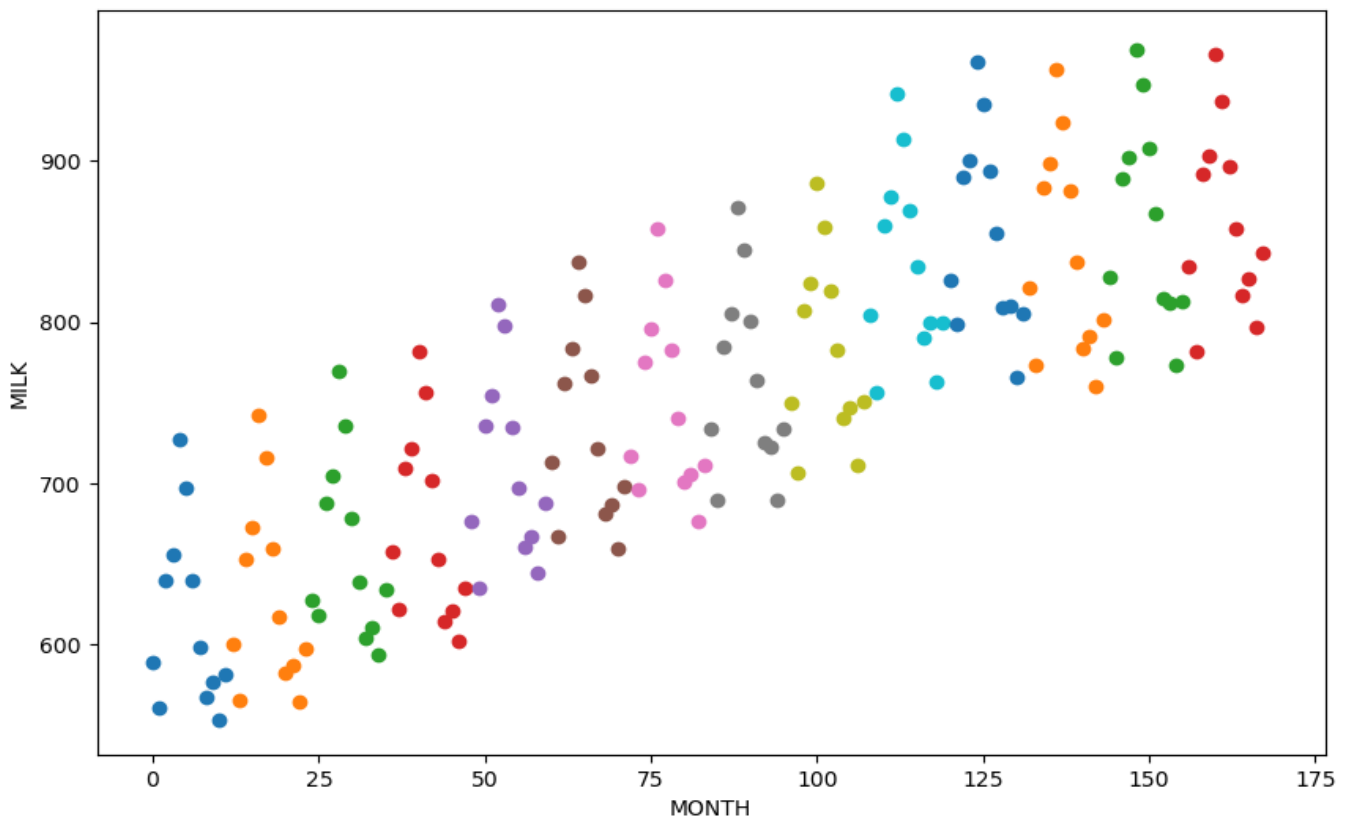
1. Импортируем нужные нам библиотеки, функции.

```
import pandas as pd
import os
import matplotlib.pyplot as plt
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import normalize
```

2. Считаем данные и выведем их на графике.

```
data = pd.read_csv(PATH + 'data\\milk.csv')
for i in range(0, 15):
    plt.scatter([j+(12*i) for j in range(0, len(data['milk'][i*12:12+12*i]))],
data['milk'][i*12:12+12*i])
plt.xlabel('MONTH')
plt.ylabel('MILK')
plt.show()
```

Figure 1

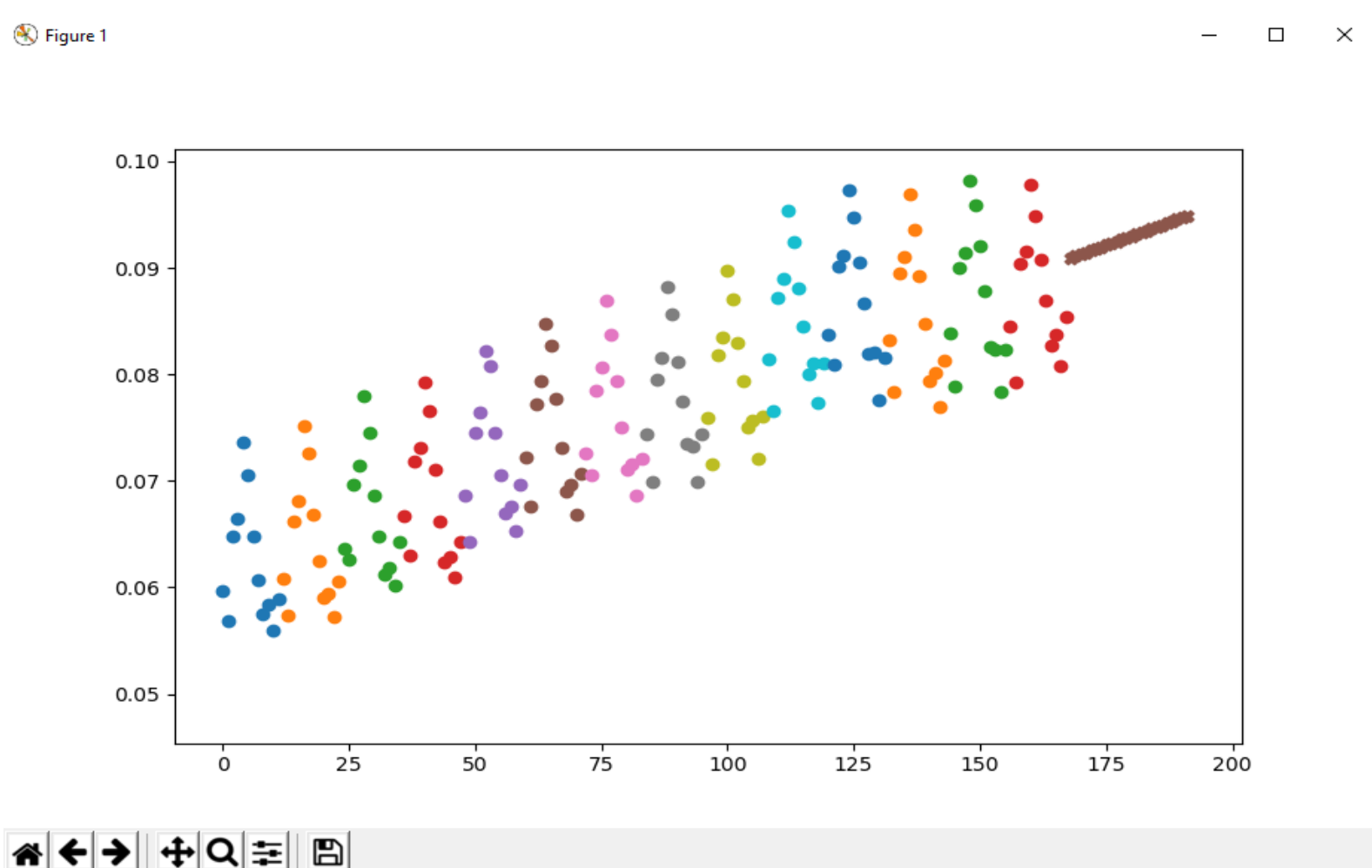


3. Подготовим данные, создадим модель на этих данные и сделаем прогноз.

```
x = np.array([i for i in range(0, len(data['milk']))]).reshape((-1, 1))
norm_data = normalize(np.array(data['milk']).reshape((1, -1)))
lr = LinearRegression().fit(x, norm_data[0])
x_predict = np.array([i for i in range(len(data['milk']),
len(data['milk'])+24)]).reshape((-1, 1))
y_predict = lr.predict(x_predict)
```

4. Выведем подготовленные данные вместе с прогнозом, на следующие два года, на графике.

```
for i in range(0, 15):
    plt.scatter([j+(12*i) for j in range(0, len(norm_data[0][i*12:12+12*i])],
norm_data[0][i*12:12+12*i])
plt.scatter(x_predict, y_predict, marker='X')
plt.show()
```



## **Вывод**

При помощи линейной регрессии мы смогли спрогнозировать рост, тренд, молочной продукции на следующие два года, но мы не смогли спрогнозировать сезонность, хотя это можно сделать установив зависимость между месяцами года и наложить ее на спрогнозированные данные.