

## Лабораторная работа №7

**Цель работы:** научиться моделировать экономические процессы, строить прогнозы уровня экономических показателей.

### ЗАДАНИЕ К ЛАБОРАТОРНОЙ РАБОТЕ

По данным таблицы провести анализ динамического ряда соответствующего варианту показателя.

Для этого:

1. Рассчитать средний уровень ряда.
2. Рассчитать абсолютный прирост, темп роста и темп прироста для всех уровней и средние значения этих показателей.
3. Провести аналитическое выравнивание динамического ряда методом наименьших квадратов, т.е. функцию  $y = f(x)$ , заданную таблично, аппроксимировать многочленом первой степени  $y = P_1(x) = a_1 + a_2x$ ;
4. Рассчитать точность полученной динамической модели (уравнения) ряда, сделать прогноз на два года.
5. Определить колеблемость показателя и его устойчивость.
6. Показать на графике исходный и выровненный динамические ряды.
7. Средствами MS Excel построить точечную диаграмму по исходным табличным данным, соответствующим своему варианту. Добавить линии тренда: линейный, полиномиальный, экспоненциальный. Установить флаги «Показывать уравнение на диаграмме», «Поместить на диаграмму величину достоверности аппроксимации ( $R^2$ )». Для линейной линии тренда сравнить ее показатели с рассчитанными аналитическим путем коэффициентами. Сделать вывод о том, какая из линий тренда лучше аппроксимирует исходные данные.
8. Сделать выводы по работе.

### Ход выполнения работы.

1. Средний уровень ряда рассчитывается как простое среднее арифметическое показателей динамического ряда:

$$\bar{y} = \frac{\sum y_i}{n}$$

2. Для расчета абсолютного прироста, темпа роста и темпа прироста используют формулы:

Абсолютный прирост базовый:

$$\Delta y_{\delta i} = y_i - y_0$$

Абсолютный прирост цепной:

$$\Delta y_{\psi i} = y_i - y_{i-1}$$

темп прироста:      базовый -  $T_{\Pi \delta i} = \frac{\Delta y_{\delta i}}{y_0} 100\%$

цепной -  $T_{\Pi \psi i} = \frac{\Delta y_{\psi i}}{y_{i-1}} 100\%$

темп роста:      базовый -  $T_{p \delta i} = \frac{y_i}{y_0} 100\%$

цепной -  $T_{p \psi i} = \frac{y_i}{y_{i-1}} 100\%$

Средние значения этих показателей равны:

- 1) средний абсолютный прирост:

$$\overline{\Delta y} = \frac{y_n - y_0}{n - 1}$$

- 2) средний темп роста:

$$\bar{T}_p = \sqrt[n]{\frac{y_n}{y_0}} \cdot 100$$

3. Провести аналитическое выравнивание динамического ряда методом наименьших квадратов, т.е. функцию  $y = f(x)$ , заданную таблично, аппроксимировать многочленом первой степени  $\tilde{y} = P_1(x) = a_1 + a_2x$ , для чего

найдем коэффициенты  $a_1$  и  $a_2$  методом наименьших квадратов, решив систему уравнений:

$$\begin{cases} \sum_{i=1}^n y_i x_i - a_2 \sum_{i=1}^n x_i^2 - a_1 \sum_{i=1}^n x_i = 0 \\ \sum_{i=1}^n y_i - a_2 \sum_{i=1}^n x_i - a_1 n = 0 \end{cases}$$

Откуда  $a_1$  и  $a_2$  равны:

$$\begin{cases} \sum_{i=1}^n y_i x_i - a_2 \sum_{i=1}^n x_i^2 - a_1 \sum_{i=1}^n x_i = 0 \\ \sum_{i=1}^n y_i - a_2 \sum_{i=1}^n x_i - a_1 n = 0 \end{cases} \Rightarrow \begin{cases} a_1 = \frac{\sum_{i=1}^n y_i - a_2 \sum_{i=1}^n x_i}{n} \\ a_2 = \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i - n \sum_{i=1}^n y_i x_i}{(\sum_{i=1}^n x_i)^2 - n \sum_{i=1}^n x_i^2} \end{cases}$$

Вспомогательные данные – суммы  $\sum_{i=1}^n x_i, \sum_{i=1}^n y_i, \sum_{i=1}^n x_i^2, \sum_{i=1}^n y_i x_i$ , вычислить, используя табличный процессор MS Excel.

4. Рассчитать точность полученной динамической модели (уравнения) ряда, сделать прогноз на два года. Для этого считаем среднеквадратическое отклонение:

$$\sigma_t = \sqrt{\frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{n-2}}$$

Интервал прогнозных значений равен:  $y_{n+1} = \tilde{y}_{n+1} \pm \sigma_t$

5. Определить колеблемость показателя и его устойчивость.

$$V_t = \frac{\sigma_t}{\bar{y}} \cdot 100\%$$

6. Показать на графике исходный и выровненный динамические ряды, рисунок 1.



Рисунок 1 – Линии исходных уровней динамического ряда и аппроксимирующей их

7. Чтобы добавить линии тренда на диаграмму, нужно щелкнуть правой кнопкой мыши по графику линии исходных данных и из контекстного меню выбрать пункт «добавить линию тренда». Установить флаги «Показывать уравнение на диаграмме» и «Поместить на диаграмму величину достоверности аппроксимации ( $R^2$ )», рисунок 2.

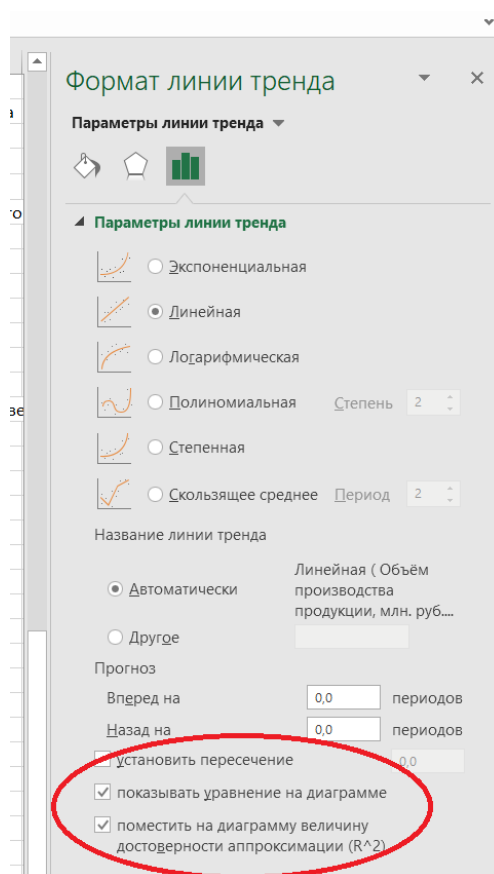


Рисунок 2 – Добавление линий тренда и их параметров.

Лучше аппроксимирует исходные данные та из линий тренда, у которой  $R^2$  ближе к 1, рисунок 3.

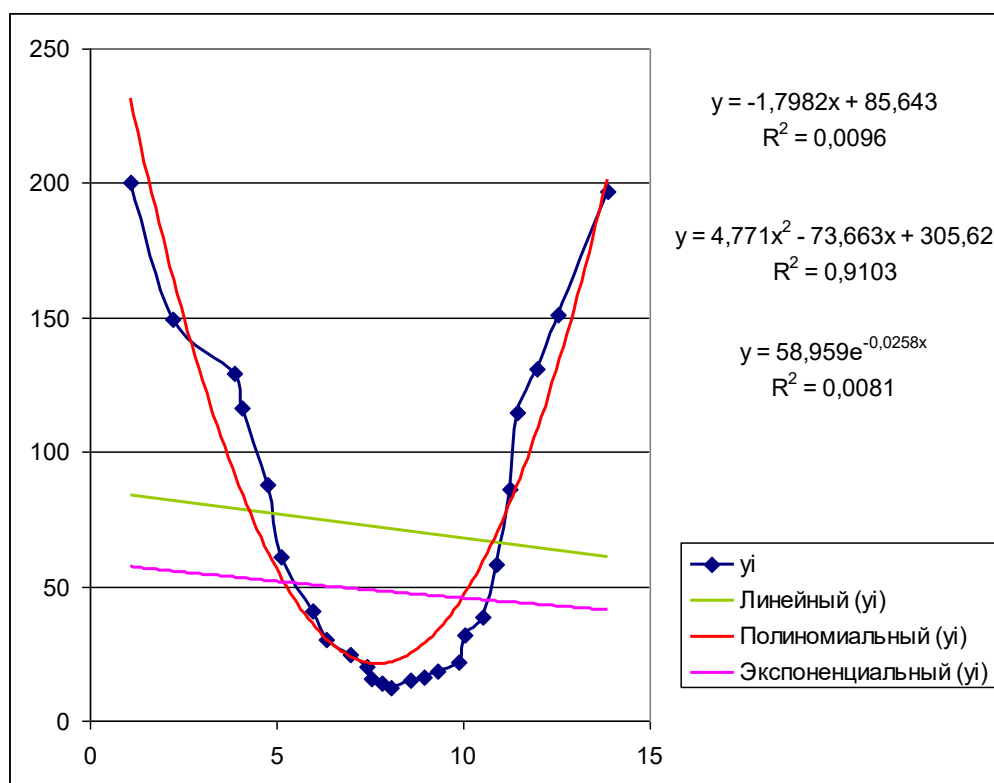


Рисунок 3. Различные линии тренда на диаграмме и их параметры.

Используя функцию ЛИНЕЙН вычислить числовые характеристики зависимости  $y$  от  $x$ . Сравнить свои вычисления по методу наименьших квадратов с результатами, полученными при помощи функции ЛИНЕЙН.

8. Сделать выводы по работе.

## Варианты заданий.

Индивидуальный вариант выбирается по остатку от деления номера студента в списке журнала группы на 8. Если этот остаток равен 0, выбираем 8-й вариант.

№ варианта	1	2	3	4	5	6	7	8
Экономи- ческий показа- тель ( $y_i$ )  Годы ( $x_i$ )	Объём производ- ства продук- ции, млн. руб.	Выручка от продаж единицы продук- ции, тыс. руб.	Средний объём затрат на топливо, р / день	Уровень безработ- ицы, чел.	Средняя заработ- ная плата руб/мес	Объём производ- ства продук- ции, млн. руб.	Выручка от продаж единицы продук- ции, тыс. руб.	Средний объём затрат на топливо, р / день
2010	1316	1377	1266	18708	18696	18717	18674	11350
2011	1400	1383	1365	18731	18554	18715	18718	11296
2012	1609	1501	1590	18469	18577	18458	18540	11483
2013	1723	1625	1658	18387	18332	18481	18429	11526
2014	1809	1756	1734	18285	18331	18281	18387	11619
2015	1909	1931	1882	18209	18221	18250	18154	11901
2016	1957	2120	1930	18147	17966	17990	18156	11875
2017	2112	2162	2120	17902	17860	18017	17887	12054
2018	2319	2195	2288	17884	17764	17890	17907	12152
2019	2384	2471	2297	17649	17785	17619	17807	12331
2020	2543	2593	2429	17533	17676	17670	17532	12496

## Лабораторная работа № 7

### «Построение прогнозов с применением машинного обучения»

**Цель работы:** научиться моделировать экономические процессы, строить прогнозы уровня экономических показателей с применением элементов машинного обучения.

### ЗАДАНИЕ К ЛАБОРАТОРНОЙ РАБОТЕ

- 1) Прочитать данные из файла dataset\_ML.csv (рисунок 1)

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from sklearn.linear_model import LinearRegression
```

```
df = pd.read_csv('dataset_ML.csv', sep = ';')
df
```

	X	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8
0	2010	1316	1377	1266	18708	18696	18717	18674	11350
1	2011	1400	1383	1365	18731	18554	18715	18718	11296
2	2012	1609	1501	1590	18469	18577	18458	18540	11483
3	2013	1723	1625	1658	18387	18332	18481	18429	11526
4	2014	1809	1756	1734	18285	18331	18281	18387	11619
5	2015	1909	1931	1882	18209	18221	18250	18154	11901
6	2016	1957	2120	1930	18147	17966	17990	18156	11875

Рисунок 1. Чтение данных из файла

- 2) Выбрать из файла вектор-столбец признаков X и значения целевой функции Y в соответствии со своим вариантом.
- 3) Преобразовать вектор признаков X в двумерный массив, состоящий из n строк и одного столбца с помощью метода reshape библиотеки numpy (рисунок 2)

```
x = np.array(df['X']).reshape((-1, 1))
y = np.array(df['Y1'])
```

x

```
array([[2010],
       [2011],
       [2012],
       [2013],
       [2014],
       [2015],
       [2016],
       [2017],
       [2018],
       [2019],
       [2020]], dtype=int64)
```

y

```
array([1316, 1400, 1609, 1723, 1809, 1909, 1957, 2112, 2319, 2384, 2543],
      dtype=int64)
```

Рисунок 2. Вывод исходных данных

- 4) Обучить модель, вывести расчетные значения параметров модели, получить прогноз на ближайшие 3 года (рисунок 3).

```
model_reg = LinearRegression().fit(x, y) # обучаем модель
model_reg.score(x, y) # коэффициент детерминации
```

```
0.9890269675777921
```

```
model_reg.coef_ # угловой коэффициент
```

```
array([119.33636364])
```

```
model_reg.intercept_ # отрезок (свободный член)
```

```
-238546.31818181823
```

```
x1 = np.array([2021, 2022]).reshape((-1, 1)) # период упреждения
y_pred = model_reg.predict(x1) # получаем прогноз
print('Прогноз:', y_pred, sep='\n')
```

```
Прогноз:
[2632.47272727 2751.80909091]
```

Рисунок 3. Обучение модели, вывод параметров модели и получение прогноза.

- 5) Получить прогноз на ближайшие 3 года с помощью рассчитанных параметров модели (рисунок 4).



```
data = pd.DataFrame(x1,y_pred)
data
```

```
      0
2632.472727 2021
2751.809091 2022
```

```
y_pred = model_reg.intercept_ + model_reg.coef_ * x1
print('Прогноз:', y_pred, sep='\n')
```

```
Прогноз:
[[2632.472727]
 [2751.809091]]
```

Рисунок 4. Расчет прогнозных значений показателя с помощью параметров модели.

- 6) Построить графики исходных данных, найденной модели и прогнозных значений (рисунки 5, 6).

```
plt.scatter(df['x'],df['y1']);
```

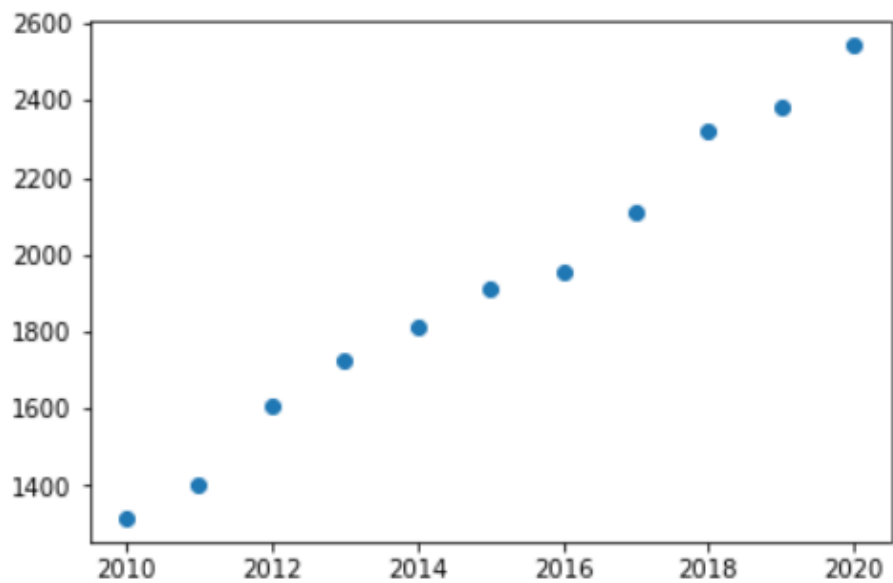


Рисунок 5. График исходных данных

```

y_pred = model_reg.intercept_ + model_reg.coef_ * x

plt.plot(df['X'],df['Y1'],label = 'Y1')
plt.plot(x, y_pred,label = 'y_pred')

plt.legend();

```

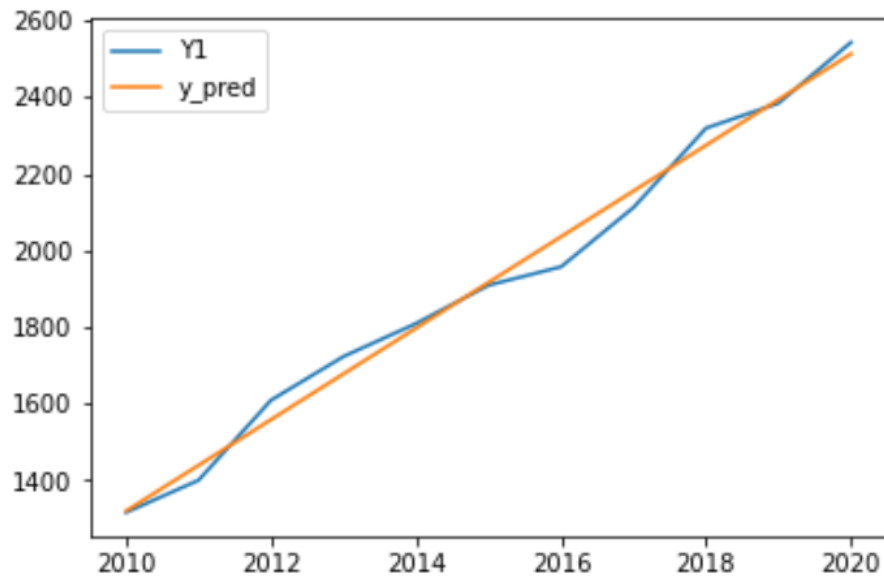


Рисунок 6. Линии, соответствующие исходным данным и найденной модели.

- 7) Сравнить полученные результаты с результатами первой части лабораторной работы, сделать выводы.