

```
import numpy as np
import pandas as pd
import plotly.express as px
import statsmodels as sm
import matplotlib.pyplot as plt
from scipy.stats import t
from scipy.stats import linregress
```

## Лабораторная работа №5

### Задание 1

{Фамилия Имя}, {Номер группы}, Вариант {Номер варианта}, ({Дата})

### Задача и данные

В таблице заданы частоты появлений значений двумерной дискретной случайной величины  $(X, Y)$ . При уровне значимости  $\alpha = 0,05$  найти коэффициент корреляции, проверить его значимость, найти линейные уравнения регрессии  $Y$  на  $X$  и  $X$  на  $Y$ . Построить корреляционное поле и на этом же графике изобразить обе прямые регрессии.

```
"78,14 79,72 85,06 87,19 89,29 77,45 84,27 92,56 88,51
85,03 95,49 80,31 81,32 71,28 95,20 87,01 81,93 81,68
77,15 64,65".replace(",", ".").replace(" ", ", ")
```

```
'78.14, 79.72, 85.06, 87.19, 89.29, 77.45, 84.27, 92.56, 88.51, 85.03,
95.49, 80.31, 81.32, 71.28, 95.20, 87.01, 81.93, 81.68, 77.15, 64.65'
```

```
data = pd.DataFrame({
    'X': [85.62, 93.21, 62.93, 68.74, 90.10, 79.74, 70.42, 90.25, 71.07,
65.37, 73.05, 93.39, 79.85, 84.11, 85.65, 69.31, 83.99, 63.85, 87.84,
78.24],
    'Y': [78.14, 79.72, 85.06, 87.19, 89.29, 77.45, 84.27, 92.56, 88.51,
85.03, 95.49, 80.31, 81.32, 71.28, 95.20, 87.01, 81.93, 81.68, 77.15,
64.65]
})
alpha = 0.05
```

```
print(f'a: {alpha}\nКоличество элементов в выборке x: {len(data["X"])}\nКоличество элементов в выборке y: {len(data["Y"])})')
```

a: 0.05

Количество элементов в выборке x: 20

Количество элементов в выборке y: 20

data.T

	0	1	2	3	4	5	6	7	8	9	10	11
X	85.62	93.21	62.93	68.74	90.10	79.74	70.42	90.25	71.07	65.37	73.05	93.3
Y	78.14	79.72	85.06	87.19	89.29	77.45	84.27	92.56	88.51	85.03	95.49	80.3

## Расчет средних значений

### Выборочное среднее для $X$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

```
x_mean = data['X'].mean()
print(f'Выборочное среднее для x: {x_mean:.2f}')
```

Выборочное среднее для x: 78.84

### Выборочное среднее для $Y$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i$$

```
y_mean = data['Y'].mean()
print(f'Выборочное среднее для y: {y_mean:.2f}')
```

Выборочное среднее для y: 83.16

### Выборочное среднее для $X$ и $Y$

$$\overline{XY} = \frac{1}{n} \sum_{i=1}^n (x_i y_i)$$

```
xy_mean = (data['X'] * data['Y']).mean()
print(f'Выборочное среднее для xy: {xy_mean:.2f}')
```

Выборочное среднее для xy: 6545.27

## Расчет дисперсий и стандартных отклонений

### Выборочная дисперсия для $X$

$$D_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2$$

```
x_var = data['X'].var(ddof=0)
print(f'Выборочная дисперсия для x: {x_var:.2f}')
```

Выборочная дисперсия для x: 96.09

### Стандартное отклонение для $X$

$$\sigma_x = \sqrt{D_x^2}$$

```
x_std = data['X'].std(ddof=0)
print(f'Стандартное отклонение для x: {x_std:.2f}')
```

Стандартное отклонение для x: 9.80

### Выборочная дисперсия для $Y$

$$D_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{Y})^2$$

```
y_var = data['Y'].var(ddof=0)
print(f'Выборочная дисперсия для y: {y_var:.2f}')
```

Выборочная дисперсия для y: 54.78

### Стандартное отклонение для $Y$

$$S_y = \sqrt{S_y^2}$$

```
s_std = data['Y'].std(ddof=0)
print(f'Стандартное отклонение для y: {s_std:.2f}')
```

Стандартное отклонение для y: 7.40

## Расчет коэффициента корреляции

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

где  $n$  - количество наблюдений,  $x_i$  и  $y_i$  - значения случайных величин,  $\bar{x}$  и  $\bar{y}$  - средние значения.

$$r_{XY} = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\sigma_x \cdot \sigma_y}$$

```
r_xy = (xy_mean - x_mean * y_mean) / (x_std * s_std)
print(f'Коэффициент корреляции: {r_xy}')
```

Коэффициент корреляции: -0.15058723424234788

## Расчет параметров $a$ и $b$

### Параметр $a$

$$a = \frac{S_y}{S_x} \cdot r_{XY}$$

```
a = s_std / x_std * r_xy
print(f'Параметр a: {a:.2f}')
```

Параметр a: -0.11

### Параметр $b$

$$b = \bar{Y} - a \cdot \bar{X}$$

```
b = y_mean - a * x_mean
print(f'Параметр b: {b:.2f}')
```

Параметр b: 92.13

## Проверка гипотезы о значимости коэффициента корреляции

$$H_0 : r_{XY} = 0$$

$$H_1 : r_{XY} \neq 0$$

### Расчетный критерий Стьюдента

$$t_{XY} = \frac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}$$

```
t_value = abs(r_xy) * np.sqrt(len(data['X']) - 2) / np.sqrt(1 - r_xy ** 2)
print(f'Значение критерия: {t_value:.2f}')
```

Значение критерия: 0.65

### Критическое значение критерия Стьюдента

$$t_{\alpha} = t_{1-\alpha, n-2}$$

```
t_alpha = t.ppf(1 - alpha / 2, len(data['X']) - 2)
print(f'Критическое значение критерия: {t_alpha:.2f}')
```

Критическое значение критерия: 2.10

## Вывод о значимости коэффициента корреляции

```
from IPython.display import Markdown, display

if abs(t_value) < t_alpha:
    display(Markdown('Коэффициент корреляции значимый, гипотеза $H_0$ принимается, т.к. $t_{\text{эмп}} < t_{\text{кр}}$.'))
else:
    display(Markdown('Коэффициент корреляции не значимый'))
```

Коэффициент корреляции значимый, гипотеза  $H_0$  принимается, т.к.  $t_{\text{эмп}} < t_{\text{кр}}$ .

## Линейные уравнения регрессии

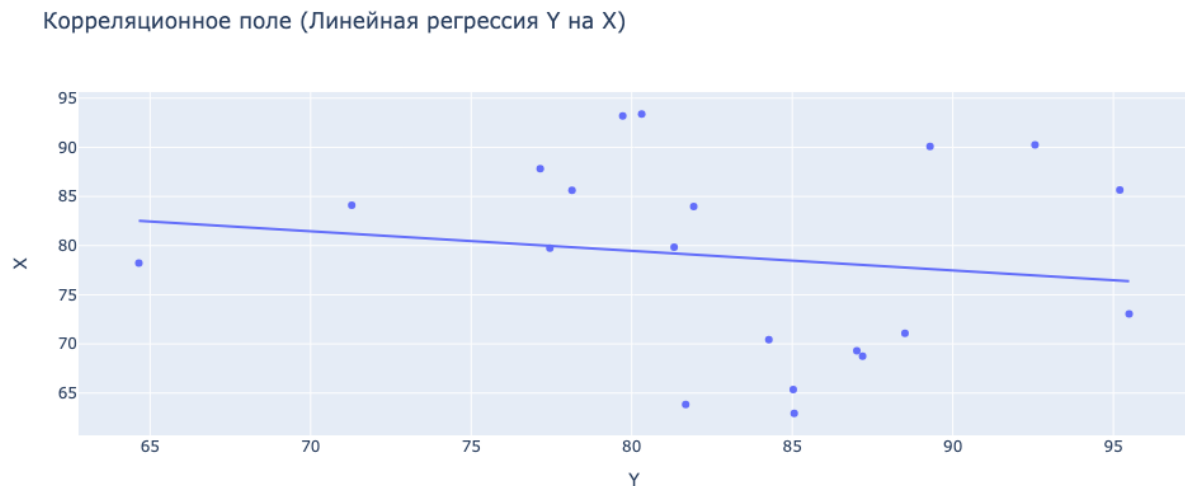
## Линейное уравнение регрессии Y на X

$$Y = aX + b$$

```
slope_yx, intercept_yx, r_value_yx, p_value_yx, std_err_yx =  
linregress(data['Y'], data['X'])  
print("Линейное уравнение регрессии X на Y: X = {:.2f}Y +  
{:.2f}".format(slope_yx, intercept_yx))
```

Линейное уравнение регрессии X на Y:  $X = -0.20Y + 95.42$

```
fig = px.scatter(data, x='Y', y='X', trendline='ols',  
title='Корреляционное поле (Линейная регрессия Y на X)')  
fig.show()
```



## Линейное уравнение регрессии X на Y

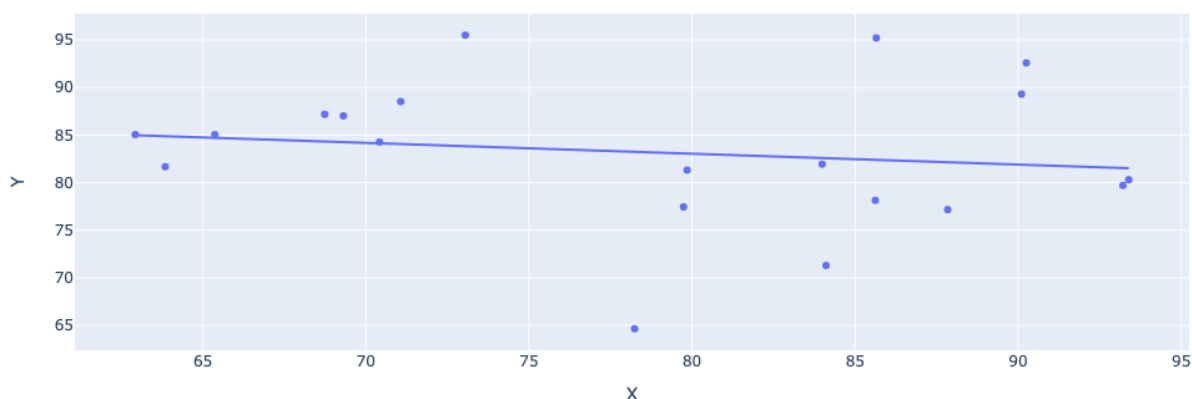
$$X = aY + b$$

```
slope_xy, intercept_xy, r_value_xy, p_value_xy, std_err_xy =  
linregress(data['X'], data['Y'])  
print("Линейное уравнение регрессии Y на X: Y = {:.2f}X +  
{:.2f}".format(slope_xy, intercept_xy))
```

Линейное уравнение регрессии Y на X:  $Y = -0.11X + 92.13$

```
px.scatter(data, x='X', y='Y', trendline='ols', title='Корреляционное поле  
(Линейная регрессия Y на X)')
```

Корреляционное поле (Линейная регрессия Y на X)



## Линейная регрессия Y на X и X на Y на одном графике

```
fig = px.scatter(data, x='X', y='Y', trendline='ols',
title='Корреляционное поле (Линейная регрессия Y на X и X на Y)',
color_discrete_sequence=['red'])
fig.add_trace(px.scatter(data, x='Y', y='X', trendline='ols',
color_discrete_sequence=['blue']).data[1])

fig.update_layout(showlegend=False)
fig.show()
```

Корреляционное поле (Линейная регрессия Y на X и X на Y)

