

## Лабораторна робота №2 ІАД

Виконав студент групи МІТ-31

Тимохін Роман Миколайович

**Тема:** Використання NumPy, Pandas та Matplotlib для обробки даних та візуалізації результатів

**Мета:** Ознайомитися з основами роботи з бібліотеками NumPy, Pandas та Matplotlib для аналізу даних, проведення базових операцій з даними та візуалізації результатів.

### Завдання:

1. NumPy (створити набір даних (розмір за власним рішенням), провести CRUD операції над даними (скаляр, вектор, матриця)). Використати `arange`, `random`.
2. Pandas (створити `dataframe` різними способами, провести маніпуляції з даними (`head`, `describe`, `iloc`, `loc` ...))
3. Matplotlib (створити набір даних лінійної функції з урахуванням помилок вимірювань, побудувати графік функції та згенерованих даних, обчислити похибку з використанням метрик MAE, MSE). Записати отримані результати у csv файл (формат: X,Y,  $\hat{Y}$ , mAE, mSE)
4. Відповіді оформити .іруnb документом
5. Викласти у зазначений репозиторій GitHub в окремій папці з назвою Lab2 де має бути .іруnb файл та необхідні файли з даними

### Код

```
import numpy as np # Додаємо імпорт бібліотеки numpy
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import csv
```

```
# Створення DataFrame з випадковими даними
```

```
df = pd.DataFrame({
```

```
    'X': np.random.randint(1, 10, 10),
```

```
    'Y': np.random.randint(10, 20, 10)
```

```
})
```

```
# Огляд перших 5 рядків
```

```
print(f"Перші 5 рядків DataFrame:\n{df.head()}")
```

```
# Опис статистичних характеристик
```

```
print(f"Опис статистики:\n{df.describe()}")
```

```
# Доступ до даних за індексом
```

```
print(f"Дані за індексом 2 (iloc):\n{df.iloc[2]}")
```

```
# Доступ до даних за ім'ям стовпця
```

```
print(f"Дані по стовпцю 'Y' (loc):\n{df.loc[:, 'Y']}")
```

```
# Генерація даних
```

```
X = np.linspace(0, 10, 100)
```

```
true_Y = 2 * X + 1 # лінійна функція  $Y = 2X + 1$ 
```

```
# Додавання випадкових помилок вимірювань
```

```
noise = np.random.normal(0, 1, X.shape)
```

```
measured_Y = true_Y + noise
```

```
# Побудова графіка
```

```
plt.figure(figsize=(8, 6))
```

```
plt.plot(X, true_Y, label="Теоретична функція ( $Y = 2X + 1$ )", color='blue')
```

```
plt.scatter(X, measured_Y, label="Виміряні дані", color='red', alpha=0.5)
```

```
plt.xlabel('X')
```

```
plt.ylabel('Y')
```

```
plt.title('Графік лінійної функції з похибками')
```

```
plt.legend()
```

```
plt.show()
```

```
# Обчислення MAE та MSE
```

```
mae = np.mean(np.abs(measured_Y - true_Y))
```

```
mse = np.mean((measured_Y - true_Y) ** 2)
```

```
print(f"MAE (Mean Absolute Error): {mae}")
```

```
print(f"MSE (Mean Squared Error): {mse}")
```

```
# Запис результатів у CSV
```

```
with open('results.csv', mode='w', newline='') as file:
```

```
    writer = csv.writer(file)
```

```
    writer.writerow(['X', 'Y', 'Y_hat', 'MAE', 'MSE'])
```

```
    for i in range(len(X)):
```

```
        writer.writerow([X[i], true_Y[i], measured_Y[i], mae, mse])
```

## **Результати**

Перші 5 рядків DataFrame:

	X	Y
0	2	13
1	1	12
2	3	12
3	4	19
4	3	18

Опис статистики:

	X	Y
count	10.000000	10.000000
mean	3.200000	14.900000
std	2.440401	2.643651
min	1.000000	12.000000
25%	1.250000	12.250000
50%	3.000000	15.000000
75%	3.750000	16.750000
max	9.000000	19.000000

Дані за індексом 2 (iloc):

X	3
Y	12

Name: 2, dtype: int32

Дані по стовпцю 'Y' (loc):

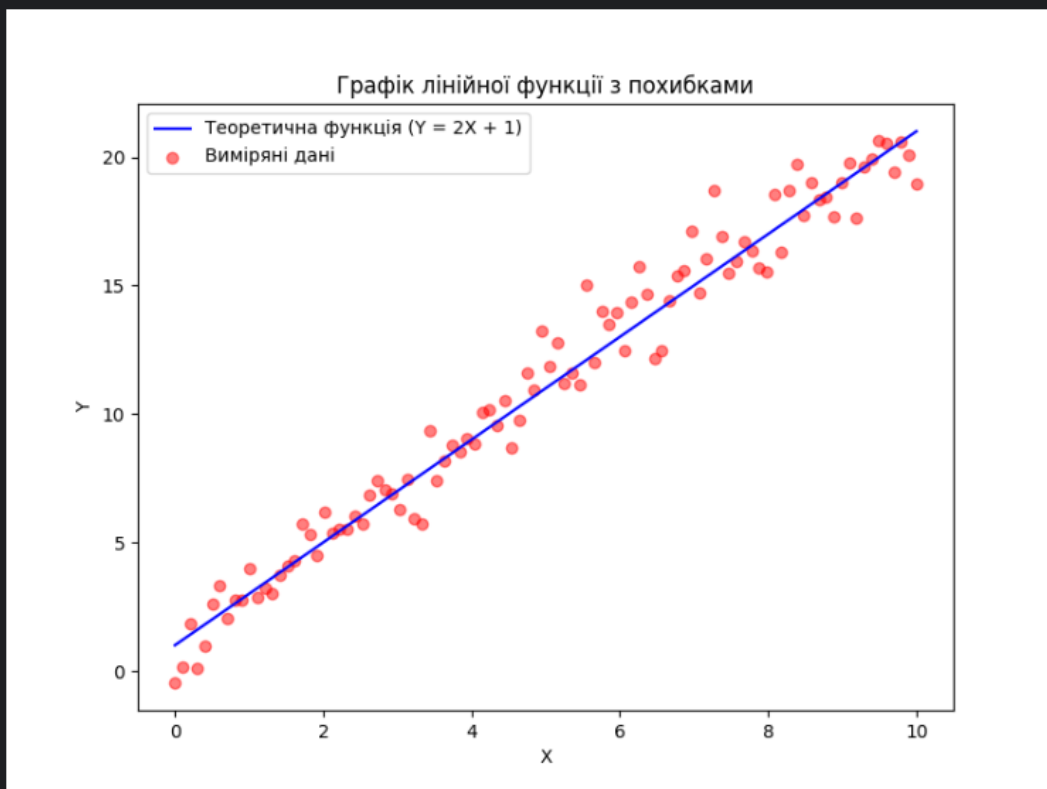
0	13
1	12
2	12
3	19
4	18
5	17
6	14
7	12
8	16
9	16

Name: Y, dtype: int32

MAE (Mean Absolute Error): 0.820107745526586

MSE (Mean Squared Error): 1.1602382293116742

Process finished with exit code 0



## Висновки:

В результаті виконання цього коду було не тільки здійснено основні операції з даними, а й обчислені важливі метрики похибки, що можуть бути корисні при аналізі моделі або результатів вимірювань. Крім того, запис результатів у CSV дозволяє ефективно зберігати та обмінюватися даними.

Загалом, цей підхід демонструє основи роботи з даними в Python за допомогою стандартних бібліотек, які є необхідними інструментами для аналізу та візуалізації даних у наукових та інженерних дослідженнях.