**Звіт до лабораторної роботи 6**

За даними таблиці, що використовувалася у попередніх завданнях, зробити три вектори даних, що відповідають стовпчикам I, J, K.  
  
Для цих вибірок попарно перевірити нульову гіпотезу про наявність статистичного зв`язку. Для цього:

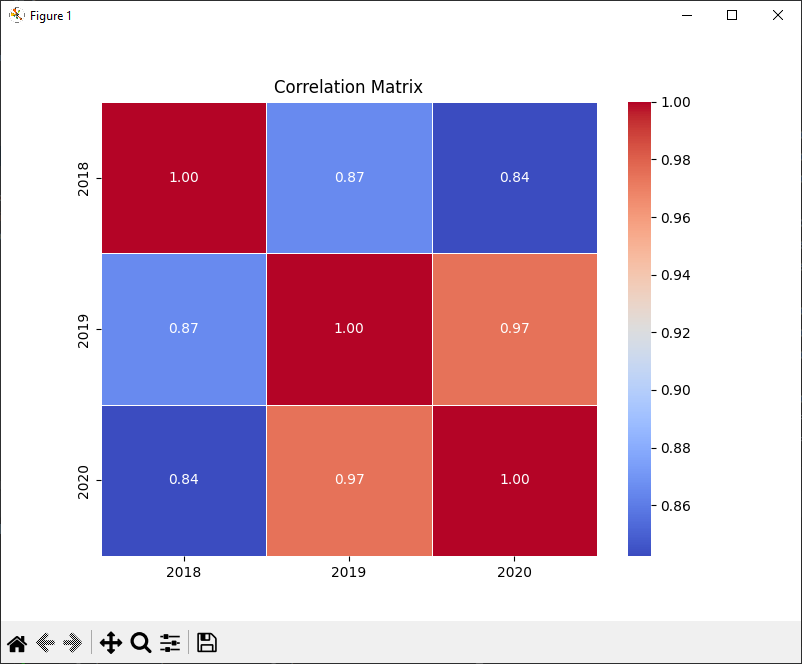
* побудувати кореляційне поле;
* розрахувати коефіцієнт кореляції Пірсона, коефіцієнти рангової кореляції Спірмена та Кендала, коефіцієнт детермінації.

Порівняти результати, отримані різними методами, та зробити висновки про значущість і лінійність зв'язку.

import pandas as pd  
import numpy as np  
import seaborn as sns  
import matplotlib.pyplot as plt  
import scipy.stats  
  
df = pd.read\_csv('output.csv')  
vector\_i = df['2018'].values  
vector\_j = df['2019'].values  
vector\_k = df['2020'].values

Зчитування потрібного файлу та створюємо три вектори відповідні нашим стовпчикам

correlation\_matrix = df[['2018', '2019', '2020']].corr()  
  
# Побудова кореляційної матриці  
plt.figure(figsize=(8, 6))  
sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=.5)  
plt.title('Correlation Matrix')  
plt.show()



Кореляційна матриця

Розрахунок всіх інших коефіцієнтів

cor\_coef\_I\_J = np.corrcoef(vector\_i, vector\_j)[0, 1]  
cor\_coef\_K\_J = np.corrcoef(vector\_k, vector\_j)[0, 1]  
cor\_coef\_I\_K = np.corrcoef(vector\_i, vector\_k)[0, 1]  
  
print(f"\nКоефіцієнт кореляції Пірсона між I J: {cor\_coef\_I\_J}")  
print(f"Коефіцієнт кореляції Пірсона між K J: {cor\_coef\_K\_J}")  
print(f"Коефіцієнт кореляції Пірсона між I K: {cor\_coef\_I\_K}")  
  
# Розрахунок коефіцієнта рангової кореляції Спірмена  
cor\_coef\_spearman\_I\_J, \_ = scipy.stats.spearmanr(vector\_i, vector\_j)  
cor\_coef\_spearman\_I\_K, \_ = scipy.stats.spearmanr(vector\_i, vector\_k)  
cor\_coef\_spearman\_K\_J, \_ = scipy.stats.spearmanr(vector\_k, vector\_j)  
  
print(f"\nКоефіцієнт рангової кореляції Спірмена між I J: {cor\_coef\_spearman\_I\_J}")  
print(f"Коефіцієнт рангової кореляції Спірмена між I K: {cor\_coef\_spearman\_I\_K}")  
print(f"Коефіцієнт рангової кореляції Спірмена між K J: {cor\_coef\_spearman\_K\_J}")  
  
cor\_coef\_kendall\_I\_J, \_ = scipy.stats.kendalltau(vector\_i, vector\_j)  
cor\_coef\_kendall\_I\_K, \_ = scipy.stats.kendalltau(vector\_i, vector\_k)  
cor\_coef\_kendall\_K\_J, \_ = scipy.stats.kendalltau(vector\_k, vector\_j)  
  
print(f"\nКоефіцієнт рангової кореляції Кендала між I J: {cor\_coef\_kendall\_I\_J}")  
print(f"Коефіцієнт рангової кореляції Кендала між I K: {cor\_coef\_kendall\_I\_K}")  
print(f"Коефіцієнт рангової кореляції Кендала між K J: {cor\_coef\_kendall\_K\_J}")  
  
  
# Функція для розрахунку R-squared  
  
def calculate\_r\_squared(x, y):  
 mean\_x = np.mean(x)  
  
 ss\_x = np.sum((x - mean\_x) \*\* 2)  
 ss\_residual = np.sum((x - y) \*\* 2)  
  
 r\_squared = 1 - (ss\_residual / ss\_x)  
  
 return r\_squared  
  
  
# Розрахунок R-squared для всіх трьох пар  
r\_squared\_i\_j = calculate\_r\_squared(vector\_i, vector\_j)  
r\_squared\_i\_k = calculate\_r\_squared(vector\_i, vector\_k)  
r\_squared\_j\_k = calculate\_r\_squared(vector\_j, vector\_k)  
  
print(f"\nКоефіцієнт детермінації (R-squared) між i та j: {r\_squared\_i\_j}")  
print(f"Коефіцієнт детермінації (R-squared) між i та k: {r\_squared\_i\_k}")  
print(f"Коефіцієнт детермінації (R-squared) між j та k: {r\_squared\_j\_k}")

Вивід:

Коефіцієнт кореляції Пірсона між I J: 0.8654681068094782

Коефіцієнт кореляції Пірсона між K J: 0.974305424789148

Коефіцієнт кореляції Пірсона між I K: 0.8423420386672834

Коефіцієнт рангової кореляції Спірмена між I J: 0.8458618719825015

Коефіцієнт рангової кореляції Спірмена між I K: 0.7813300356691688

Коефіцієнт рангової кореляції Спірмена між K J: 0.9329912492379501

Коефіцієнт рангової кореляції Кендала між I J: 0.8405372166128019

Коефіцієнт рангової кореляції Кендала між I K: 0.7641132765835219

Коефіцієнт рангової кореляції Кендала між K J: 0.916549484408755

Коефіцієнт детермінації (R-squared) між i та j: 0.7183670934274322

Коефіцієнт детермінації (R-squared) між i та k: 0.6581684405902355

Коефіцієнт детермінації (R-squared) між j та k: 0.9460891720362395

З цього всього ми можемо зробити наступні висновки:

За коефіцієнтами Пірсона все має досить сильний лінійний зв’язок, але найбільш лінійно пов’язані K та J.

За Коефіцієнтом рангової кореляції Спірмена та Кендала у всіх пар присутній досить сильний монотонний зв’язок, але найбільш виражений він в тій самій парі K та J.

За коефіцієнтом детермінації отримаємо той самий результат, всюди досить сильна пояснювальна сила моделі, але в парі K та J вона досить близька до 1 та є найсильнішою.