pearson 상관 계수

# example code

|  |
| --- |
| import pandas as pd  import numpy as np  from sklearn import datasets  data = datasets.load\_diabetes()  df = pd.DataFrame(data['data'], index=data['target'], columns=data['feature\_names'])  X = df.bmi.values  Y = df.index.values  # BMI(체질량지수)와 당뇨병수치(Target)의 산점도 그래프  import matplotlib.pyplot as plt  plt.scatter(X, Y, alpha=0.5)  plt.title('TARGET ~ BMI')  plt.xlabel('BMI')  plt.ylabel('TARGET')  plt.show()  cov = (np.sum(X\*Y) - len(X)\*np.mean(X)\*np.mean(Y)) / len(X)  cov  corr = cov / (np.std(X) \* np.std(Y))  corr  np.cov(X,Y)[0,1]  np.corrcoef(X,Y)[0,1]  import scipy.stats as stats  stats.pearsonr(X,Y)  #(상관계수, p-value) |

# testing result

|  |
| --- |
| 2.1480435755297007  0.5864501344746884  2.152914422639746  0.5864501344746886  (0.5864501344746887, 3.4660064451654114e-42) |