khai báo các thư viện

import math

import statistics

import numpy as np

import scipy.stats

import pandas as pd

tạo dữ liệu

x = [8.0, 1, 2.5, 4, 28.0]

x\_with\_nan = [8.0, 1, 2.5, math.nan, 4, 28.0]

x

* [8.0, 1, 2.5, 4, 28.0]

Tạo x với nan

x\_with\_nan

* [8.0, 1, 2.5, nan, 4, 28.0]

tạo np.ndarray và pd.Series tương ứng với x và x\_with\_nan

y, y\_with\_nan = np.array(x), np.array(x\_with\_nan)

z, z\_with\_nan = pd.Series(x), pd.Series(x\_with\_nan)

y

* array([ 8. , 1. , 2.5, 4. , 28. ])

tạo y, z với nan

y\_with\_nan

* array([ 8. , 1. , 2.5, nan, 4. , 28. ])

z\_with\_nan

0 8.0

1 1.0

2 2.5

3 NaN

4 4.0

5 28.0

* dtype: float64

tính giá trị trung bình

mean\_ = sum(x) / len(x)

mean\_

* 8.7

sử dụng hàm statistics.mean(x) để thống kê

mean\_ = statistics.mean(x)

mean\_

chỉ chạy trên python 3.8

mean\_ = statistics.fmean(x)

mean\_

thống kê dữ liệu không phải số thì trả về giá trị không phải số(nan). kết quả này giống với kết quả của hàm sum

mean\_ = statistics.mean(x\_with\_nan)

mean\_

* Nan

Hàm np.mean trả về giá trị trung bình

mean\_ = np.mean(y)

mean\_

* 8.7

mean\_ = y.mean()

mean\_

* 8.7