

Zad2_051012024

October 14, 2024

```
[1]: import pandas as pd
import numpy as np
```

```
[2]: df = pd.read_csv('IHME_GBD_2019_SMOKING_TOB_1990_2019_NUM_SMOKERS_Y2021M05D27.
↪CSV')
df
```

```
[2]:
```

	measure_name	location_id	location_name	sex_id	sex_name	\
0	Number of Smokers	1	Global	1	Male	
1	Number of Smokers	1	Global	2	Female	
2	Number of Smokers	1	Global	3	Both	
3	Number of Smokers	1	Global	1	Male	
4	Number of Smokers	1	Global	2	Female	
...	
20965	Number of Smokers	522	Sudan	2	Female	
20966	Number of Smokers	522	Sudan	3	Both	
20967	Number of Smokers	522	Sudan	1	Male	
20968	Number of Smokers	522	Sudan	2	Female	
20969	Number of Smokers	522	Sudan	3	Both	

	age_group_id	age_group_name	year_id	val	upper	\
0	29	15+ years	1990	8.031015e+08	8.096221e+08	
1	29	15+ years	1990	1.891488e+08	1.930929e+08	
2	29	15+ years	1990	9.922503e+08	1.000161e+09	
3	29	15+ years	1991	8.138972e+08	8.200339e+08	
4	29	15+ years	1991	1.905375e+08	1.944249e+08	
...	
20965	29	15+ years	2018	2.435999e+05	3.286166e+05	
20966	29	15+ years	2018	2.610672e+06	2.833943e+06	
20967	29	15+ years	2019	2.439150e+06	2.656579e+06	
20968	29	15+ years	2019	2.500800e+05	3.345384e+05	
20969	29	15+ years	2019	2.689230e+06	2.918332e+06	

	lower
0	7.959086e+08
1	1.855595e+08
2	9.847880e+08
3	8.069514e+08

```

4      1.869744e+08
...
20965  1.752508e+05
20966  2.409108e+06
20967  2.236450e+06
20968  1.816686e+05
20969  2.480656e+06

```

[20970 rows x 11 columns]

```
[3]: df.dropna()
```

```

[3]:      measure_name  location_id location_name  sex_id sex_name \
0      Number of Smokers           1      Global         1      Male
1      Number of Smokers           1      Global         2      Female
2      Number of Smokers           1      Global         3      Both
3      Number of Smokers           1      Global         1      Male
4      Number of Smokers           1      Global         2      Female
...
20965  Number of Smokers          522      Sudan         2      Female
20966  Number of Smokers          522      Sudan         3      Both
20967  Number of Smokers          522      Sudan         1      Male
20968  Number of Smokers          522      Sudan         2      Female
20969  Number of Smokers          522      Sudan         3      Both

      age_group_id  age_group_name  year_id      val      upper \
0              29      15+ years      1990  8.031015e+08  8.096221e+08
1              29      15+ years      1990  1.891488e+08  1.930929e+08
2              29      15+ years      1990  9.922503e+08  1.000161e+09
3              29      15+ years      1991  8.138972e+08  8.200339e+08
4              29      15+ years      1991  1.905375e+08  1.944249e+08
...
20965              29      15+ years      2018  2.435999e+05  3.286166e+05
20966              29      15+ years      2018  2.610672e+06  2.833943e+06
20967              29      15+ years      2019  2.439150e+06  2.656579e+06
20968              29      15+ years      2019  2.500800e+05  3.345384e+05
20969              29      15+ years      2019  2.689230e+06  2.918332e+06

      lower
0      7.959086e+08
1      1.855595e+08
2      9.847880e+08
3      8.069514e+08
4      1.869744e+08
...
20965  1.752508e+05
20966  2.409108e+06

```

```
20967  2.236450e+06
20968  1.816686e+05
20969  2.480656e+06
```

```
[20970 rows x 11 columns]
```

```
[4]: series1 = df['val']
      series1
```

```
[4]: 0      8.031015e+08
      1      1.891488e+08
      2      9.922503e+08
      3      8.138972e+08
      4      1.905375e+08
      ...
      20965  2.435999e+05
      20966  2.610672e+06
      20967  2.439150e+06
      20968  2.500800e+05
      20969  2.689230e+06
      Name: val, Length: 20970, dtype: float64
```

```
[5]: # obliczanie średniej
      np.mean(series1)
```

```
[5]: 12428071.383604305
```

```
[6]: #obliczanie mediany
      np.median(series1)
```

```
[6]: 577712.25205
```

```
[7]: #Odchylenie standardowe
      np.std(series1)
```

```
[7]: 64890362.21057887
```

```
[8]: #Wariancja
      np.var(series1)
```

```
[8]: 4210759107820122.0
```

```
[9]: series2 = df['upper']
```

```
series2
```

```
[9]: 0      8.096221e+08
      1      1.930929e+08
      2      1.000161e+09
      3      8.200339e+08
      4      1.944249e+08
      ...
      20965  3.286166e+05
      20966  2.833943e+06
      20967  2.656579e+06
      20968  3.345384e+05
      20969  2.918332e+06
      Name: upper, Length: 20970, dtype: float64
```

```
[10]: #Koleracja

correlation = np.corrcoef(series1,series2) [0,1]
print(correlation)

0.9999762487761583
```

```
[11]: #Kowariancja

covariance = np.cov(series1,series2) [0,1]
covariance
```

```
[11]: 4254193895188547.0
```

```
[ ]:
```