

SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

| | |
|---|---|
| Laboratorium Nr 1 Data 19.10.2024 Temat: Wprowadzenie do narzędzi i środowiska pracy w analizie danych Wariant: 5 | Mateusz Łysoń Informatyka II stopień, niestacjonarne, 1 semestr, gr. b |
|---|---|

Polecenie

Na podstawie swojego wariantu danych oblicz średnią, medianę, odchylenie standardowe oraz korelację.

Opis programu opracowanego

```
1 import pandas as pd
2
3 # Ścieżka do pliku CSV
4 csv_file_path = r"C:\Users\Sensorbtf\OneDrive\studia\Magisterka\Dane\IHME_GBD_2019_SMOKING_AGE_1990_2019_INIT_
5 df = pd.read_csv(csv_file_path)
6 df_cleaned = df.dropna(subset=['year_id', 'val'])
7
8 # Obliczamy średnie rozpowszechnienie palenia dla każdej płci
9 mean_prevalence_by_sex = df_cleaned.groupby('sex_name')['val'].mean().round(2)
10 print("\n===== Średnie rozpowszechnienie palenia w zależności od płci =====\n")
11 print(mean_prevalence_by_sex)
12
13 # 2. Obliczamy medianę rozpowszechnienia palenia (z użyciem funkcji round() z Pythona)
14 median_prevalence = round(df_cleaned['val'].median(), 2)
15 print("\n===== Mediana rozpowszechnienia palenia =====\n")
16 print(f"Mediana rozpowszechnienia palenia: {median_prevalence}")
17
18 # 3. Obliczamy odchylenie standardowe dla każdej płci
19 std_prevalence_by_sex = df_cleaned.groupby('sex_name')['val'].std().round(2)
20 print("\n===== Odchylenie standardowe rozpowszechnienia palenia w zależności od płci =====\n")
21 print(std_prevalence_by_sex)
22
23 # 4. Obliczamy kowariancję między 'sex_id' a 'val' (rozpowszechnieniem palenia)
24 covariance_sex = df_cleaned[['sex_id', 'val']].cov().round(4)
25 print("\n===== Kowariancja między płcią a rozpowszechnieniem palenia =====\n")
26 print(covariance_sex)
27
28 # 5. Obliczamy korelację między 'sex_id' a 'val' (rozpowszechnieniem palenia)
29 correlation_sex = df_cleaned[['sex_id', 'val']].corr().round(4)
30 print("\n===== Korelacja między płcią a rozpowszechnieniem palenia =====\n")
31 print(f"Korelacja między płcią a rozpowszechnieniem palenia: {correlation_sex}")
32
```

Console IJA X

val 0.4608 4.2638

===== Korelacja między płcią a rozpowszechnieniem palenia =====

Korelacja między płcią a rozpowszechnieniem palenia: 0.4458

runfile(C:/Users/Sensorbtf/OneDrive/studia/Magisterka/Dane/IHME_GBD_2019_SMOKING_AGE_1990_2019_INIT_2/SecondAndThird.py, wdir=C:/Users/Sensorbtf/OneDrive/studia/Magisterka/Dane/IHME_GBD_2019_SMOKING_AGE_1990_2019_INIT_2)

===== Średnie rozpowszechnienie palenia w zależności od płci =====

| sex_name | val |
|----------|-------|
| Female | 28.29 |
| Male | 16.45 |

Name: val, dtype: float64

===== Mediana rozpowszechnienia palenia =====

Mediana rozpowszechnienia palenia: 19.07

===== Odchylenie standardowe rozpowszechnienia palenia w zależności od płci =====

| sex_name | val |
|----------|------|
| Female | 2.34 |
| Male | 1.45 |

Name: val, dtype: float64

===== Kowariancja między płcią a rozpowszechnieniem palenia =====

| sex_id | val |
|--------|---------------|
| sex_id | 0.2506 0.4608 |
| val | 0.4608 4.2638 |

===== Korelacja między płcią a rozpowszechnieniem palenia =====

Korelacja między płcią a rozpowszechnieniem palenia: 0.4458

In [23]:

Wnioski

Z przeprowadzonej analizy rozpowszechnienia palenia wynika, że kobiety w grupie wiekowej 15-24 lata wykazują wyższe rozpowszechnienie palenia niż mężczyźni, a rozrzut wartości dla kobiet jest większy niż dla mężczyzn. Umiarkowana korelacja między płcią a rozpowszechnieniem palenia sugeruje, że płeć ma wpływ na wzorce palenia.

Github: <https://github.com/sensorbtf/Nauka-O-Danych/tree/main>