**Statystyczna Analiza Danych Medycznych – Projekt**

Kamil Hulewicz 248889, Aleksander Pawłowski 250995

1. **Statystyka opisowa**

Dane użyte do wykonania statystyki opisowej pochodzą z bazy danych KEEL (<https://sci2s.ugr.es/keel/dataset.php?cod=21>). Dane pochodzą od pacjentek (kobiet) w wieku co najmniej 21 lat, pochodzącymi z plemienia Indian Pima.

Etykieta klasy reprezentuje, czy dana osoba ma cukrzycę (testowany\_pozytywny) lub czy dana osoba nie ma cukrzycy (testowany\_negatywny).

Informacje o atrybutach:

1. Preg - Liczba, mówiąca razy kobieta była w ciąży.

2. Plas - Stężenie glukozy w osoczu w doustnym, 2 godzinnym teście tolerancji glukozy.

3. Pres - Rozkurczowe ciśnienie krwi (mm Hg)

4. Skin - grubość fałdu skórnego mięśnia trójgłowego (mm)

5. Insu - 2-godzinna insulina w surowicy (mu U/ml)

6. Mass - wskaźnik masy ciała ( / )

7. Pedi - Funkcja rodowodu cukrzycy

8. Age - Wiek (w latach)

Wyniki poszczególnych miar statystyki opisowej przedstawiono w tabeli 1.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Preg | Plas | Pres | Skin | Insu | Mass | Pedi | Age |
| Średnia | Zbiór | 3,85 | 120,89 | 69,11 | 20,54 | 79,80 | 31,99 | 0,47 | 33,24 |
| tested\_positive | 4,87 | 141,26 | 70,82 | 22,16 | 100,34 | 35,14 | 0,55 | 37,07 |
| tested\_negative | 3,30 | 109,98 | 68,18 | 19,66 | 68,79 | 30,30 | 0,43 | 31,19 |
| Mediana | Zbiór | 3,00 | 117,00 | 72,00 | 23,00 | 30,50 | 32,00 | 0,37 | 29,00 |
| tested\_positive | 4,00 | 140,00 | 74,00 | 27,00 | 0,00 | 34,25 | 0,45 | 36,00 |
| tested\_negative | 2,00 | 107,00 | 70,00 | 21,00 | 39,00 | 30,05 | 0,34 | 27,00 |
| Moda | Zbiór | 1,00 | 100,00 | 70,00 | 0,00 | 0,00 | 32,00 | 0,26 | 22,00 |
| tested\_positive | 0,00 | 125,00 | 70,00 | 0,00 | 0,00 | 32,90 | 0,25 | 25,00 |
| tested\_negative | 1,00 | 99,00 | 74,00 | 0,00 | 0,00 | 0,00 | 0,21 | 22,00 |
| Rozstęp | Zbiór | 17,00 | 199,00 | 122,00 | 99,00 | 846,00 | 67,10 | 2,34 | 60,00 |
| tested\_positive | 17,00 | 199,00 | 114,00 | 99,00 | 846,00 | 67,10 | 2,33 | 49,00 |
| tested\_negative | 13,00 | 197,00 | 122,00 | 60,00 | 744,00 | 57,30 | 2,25 | 60,00 |
| Odchylenie standardowe | Zbiór | 3,37 | 31,95 | 19,34 | 15,94 | 115,17 | 7,88 | 0,33 | 11,75 |
| tested\_positive | 3,73 | 31,88 | 21,45 | 17,65 | 138,43 | 7,25 | 0,37 | 10,95 |
| tested\_negative | 3,01 | 26,12 | 18,05 | 14,88 | 98,77 | 7,68 | 0,30 | 11,66 |
| Wariancja | Zbiór | 11,34 | 1020,92 | 374,16 | 254,14 | 13263,89 | 62,08 | 0,11 | 138,12 |
| tested\_positive | 13,94 | 1016,33 | 460,17 | 311,41 | 19162,90 | 52,55 | 0,14 | 119,85 |
| tested\_negative | 9,09 | 682,00 | 325,62 | 221,27 | 9754,80 | 59,02 | 0,09 | 135,86 |
| Współczynnik zmienności | Zbiór | 0,88 | 0,26 | 0,28 | 0,78 | 1,44 | 0,25 | 0,70 | 0,35 |
| tested\_positive | 0,77 | 0,23 | 0,30 | 0,80 | 1,38 | 0,21 | 0,68 | 0,30 |
| tested\_negative | 0,91 | 0,24 | 0,26 | 0,76 | 1,44 | 0,25 | 0,70 | 0,37 |
| Kwartyl Q1 | Zbiór | 1,00 | 99,00 | 62,00 | 0,00 | 0,00 | 27,30 | 0,24 | 24,00 |
| tested\_positive | 1,75 | 119,00 | 66,00 | 0,00 | 0,00 | 30,80 | 0,26 | 28,00 |
| tested\_negative | 1,00 | 93,00 | 62,00 | 0,00 | 0,00 | 25,40 | 0,23 | 23,00 |
| Kwartyl Q2 | Zbiór | 3,00 | 117,00 | 72,00 | 23,00 | 30,50 | 32,00 | 0,37 | 29,00 |
| tested\_positive | 4,00 | 140,00 | 74,00 | 27,00 | 0,00 | 34,25 | 0,45 | 36,00 |
| tested\_negative | 2,00 | 107,00 | 70,00 | 21,00 | 39,00 | 30,05 | 0,34 | 27,00 |
| Kwartyl Q3 | Zbiór | 6,00 | 140,25 | 80,00 | 32,00 | 127,25 | 36,60 | 0,63 | 41,00 |
| tested\_positive | 8,00 | 167,00 | 82,00 | 36,00 | 167,25 | 38,78 | 0,73 | 44,00 |
| tested\_negative | 5,00 | 125,00 | 78,00 | 31,00 | 105,00 | 35,30 | 0,56 | 37,00 |
| Skośność | Zbiór | 0,90 | 0,17 | -1,84 | 0,11 | 2,27 | -0,43 | 1,92 | 1,13 |
| tested\_positive | 0,50 | -0,50 | -1,94 | 0,12 | 1,84 | 0,00 | 1,72 | 0,58 |
| tested\_negative | 1,11 | 0,17 | -1,81 | 0,03 | 2,50 | -0,67 | 2,01 | 1,57 |
| Kurtoza | Zbiór | 0,16 | 0,64 | 5,18 | -0,52 | 7,21 | 3,29 | 5,59 | 0,64 |
| tested\_positive | -0,44 | 1,43 | 4,70 | -0,21 | 4,36 | 4,76 | 4,56 | -0,35 |
| tested\_negative | 0,69 | 1,90 | 5,69 | -1,04 | 9,46 | 3,06 | 6,12 | 1,96 |

Tabela 1. Poszczególne miary statystyki opisowej.

1. **Testy nieparametryczne**

Dla Pima

Obraz zawierający tekst

Opis wygenerowany automatycznie

Dla obu klas test Shapiro odrzuca hipteze – rozkłady różne od normalnego.

Test Manna Whitneya U p<0.05, ale p>0.01. Wnioski?