

Ćwiczenie SOW

Statystyki opisowe i wizualizacja danych

Część teoretyczna

Materiał z wykładu dot. statystyk opisowych i wizualizacji danych.

Zadania do wykonania

Dokonaj opisu statystycznego i wizualizacji danych charakteryzujących samochody.

1. Wczytaj i zmodyfikuj zbiór danych:

```
load carbig

nr_gr = ?;
r_k = ?;

%modyfikacja danych
x = [Acceleration Displacement Horsepower Weight MPG];
rand('state',nr_gr*r_k);
mm = 1 + (rand(size(x))*2 - 1)/100;
x = x.*mm;
```

gdzie za nr_gr wstaw numer swojej sekcji a za r_k aktualny rok kalendarzowy.

Podejrzyj i opisz zmienne.

2. Wyznacz wartości minimalne i maksymalne oraz miary średnie (slajdy 22 i 23) dla każdej zmiennej numerycznej (kolumny macierzy x).
3. Wyznacz rozstęp, kwartyle, rozstęp międzykwartylowy, odchylenie ćwiartkowe i typowy obszar zmienności (slajd 24) dla każdej zmiennej numerycznej. Utwórz wykresy pudełkowe dla wszystkich zmiennych numerycznych (boxplot).
4. Wyznacz miary rozrzutu (slajd 26 i 27) dla każdej zmiennej numerycznej.
5. Utwórz histogramy dla każdej zmiennej numerycznej (histogram; slajd 32).
6. Wyznacz współczynniki asymetrii (skewness) oraz kurtozy (kurtosis) dla każdej zmiennej numerycznej (slajdy 30 i 31).
7. Utwórz wykres probplot dla każdej zmiennej numerycznej. Wyjaśnij co on obrazuje.
8. Utwórz i zinterpretuj wykres parallelcoords:

```
labels = {'Acceleration' 'Displacement' 'Horsepower' 'Weight' 'MPG'};
parallelcoords(x,'Standardize','on','Group',Cylinders,'Labels',labels);
```

9. Utwórz i zinterpretuj wykresy glyphplot:

```
for i=1:406
    ix(i) = strcmp(Mfg(i,:), 'toyota ');
end

figure
```

```

glyphplot(x(ix,:), 'glyph', 'star', 'obslabels', Model(ix,:), 'page', 'scroll')

figure
glyphplot(x(ix,:), 'glyph', 'face', 'obslabels', Model(ix,:), 'page', 'scroll')

```

10. Wyznacz współczynniki korelacji liniowej dla każdej pary zmiennych numerycznych (corrcoef; slajd 33).
11. Utwórz wykres obrazujący współzależności pomiędzy każdą parą zmiennych numerycznych (gplotmatrix; slajd 47).
12. Wyznacz dodatkową zmienną (Zuzycie) reprezentującą zużycie paliwa w litrach na 100 km (na podstawie zmiennej MPG). Przyjmij 1 mila = 1,6 km, 1 galon = 3,79 litra. Sporządź wykres zależności zużycia paliwa od MPG.
13. Utwórz wykresy zależności Zuzycie od Acceleration, Displacement, Horsepower, Weight. Używając polecenia Basic Fitting (opcja Tools na wykresie) pokaż wyniki aproksymacji tych zależności (wykresy, błędy aproksymacji, parametry modelu).

W nowszych wersjach Matlab'a użyj do tego celu narzędzia curveFitter.

14. Jakie jest średnie zużycie paliwa dla samochodów z 3, 4, 5, 6 i 8 cylindrami?

Co powinno znaleźć się w sprawozdaniu

- A) Cel ćwiczenia.
- B) Treść zadania.
- C) Raport z wykonania ćwiczenia. Dla każdego punktu: polecenia Matlab'a, wyniki (tabele, wykresy), komentarz.
- D) Wnioski końcowe.

Zadania dodatkowe dla ambitnych

1. Zaimplementuj ćwiczenie w innym środowisku, np. Python, R, C#, ...

Przykładowe zagadnienia i pytania zaliczeniowe

1. Statystyki opisowe i ich implementacje w Matlabie
2. Metody wizualizacji danych i ich implementacje w Matlabie.
3. Materiał ze sprawozdania.

Do przygotowania na następne zajęcia

1. Zapoznać się z instrukcją do kolejnego ćwiczenia.
2. Zapoznać się z częścią teoretyczną do kolejnego ćwiczenia.