

Eksploracji danych

Lista 2

1. Oblicz podstawowe charakterystyki statystyczne dla danych

−10, 9, 1, 2, 5, −2, 6, 2, 1, 0, 1, 4, 5, 6, 3, 7, 3, 2,
2, 3, 8, 5, 3, 4, 8, 0, 8, 0, 5, 1, 6, 4, 8, 13, 2, −13, 20.

2. Dla danych z poprzedniego zadania narysuj histogram.
3. Dla danych z zad. 1 narysuj wykres pudełkowy.
4. Dla danych z zad. 1 znajdź obserwacje odstające metodą trzech sigm i metodą półtora rozstępu ćwiartkowego (międzykwartyłowego).
5. Dla danych `mpg{ggplot2}` dla cechy `hwy` (highway miles per gallon) znajdź podstawowe charakterystyki statystyczne.
6. Dla danych `mpg{ggplot2}` dla cechy `hwy` narysuj histogram z różną liczbą przedziałów i wykres pudełkowy.
7. Dla danych `mpg{ggplot2}` dla cechy `hwy` znajdź wartości odstające metodą półtora rozstępu ćwiartkowego i trzech sigm. Znajdź wiersze odpowiadające obserwacjom odstającym.
8. Zmienna X opisuje staż pracy a Y wydajność. Na podstawie wykresu rozrzutu i korelacji określić zależność pomiędzy stażem a wydajnością.

X	1	5	10	8	9	1	2	4	5	6
Y	120	115	132	123	128	102	106	109	112	110

9. Dla danych `mtcars{datasets}` znajdź korelację i regresję liniową, zrób wykres rozrzutu dla zmiennych `mile/galon` (`mpg`, zmienna niezależna) i ciężar (`wt`, zmienna zależna). Umieść wszystko na jednym rysunku (skorzystaj też z funkcji `ggscatter{ggpubr}`).
10. Dla danych `bonds{expsmooth}` `visitors{expsmooth}` zrób wykres szeregu czasowego z opisem i wykresy rozrzutu z dla opóźnień 1,2,3,4 (funkcja `lag.plot{stats}`). Znajdź trend liniowy dla tych szeregów (metodą najmniejszych kwadratów) i podaj prognozę na podstawie trendu na następne 4 podokresy. Skorzystaj z funkcji `tslm{forecast}` i `forecast{forecast}`.