

Podstawy statystyki praktycznej

laboratorium – lista 1

1. Zbiór danych `grades.txt` zawiera dane 89 uczniów siódmej klasy pewnej szkoły w USA. Dla każdego ucznia mamy informację o średniej ocen (ocenom F, D-, D, D+, C-, C, C+, B-, B, B+, A-, A, odpowiadają liczby od 0 do 11), wynik standardowego testu IQ, płeć oraz punktacja na teście psychologicznym (Piers-Harris Children's Self-Concept Scale).
 - (a) Narysuj histogramy wszystkich zmiennych ilościowych. Opisz kształt, środek i rozrzut rozkładów. Podaj wartości odpowiednich statystyk (minimum, maximum, mediana, kwartyle, średnia, odchylenie standardowe, wariancja, współczynnik zmienności).
 - (b) Powtórz punkt (a) osobno dla chłopców, a osobno dla dziewcząt. Porównaj wyniki.
2. Dla wybranej zmiennej ilościowej ze zbioru `grades.txt` narysuj histogramy dla kilku różnych liczb klas (w tym bardzo małej liczby klas i bardzo dużej liczby klas). Skomentuj jaki efekt otrzymujemy wybierając zbyt mało klas, a jaki efekt, gdy wybieramy zbyt dużo klas.
3. Zbiór danych `income.dat` zawiera wyniki ankiet przeprowadzonych przez Bureau of Labor Statistics (USA) na próbie 55 899 osób. Dla każdej osoby podany jest wiek (w latach), wykształcenie (1 = podstawowe, 2 = niepełne średnie, 3 = średnie, 4 = niepełne wyższe, 5 = wyższe (licencjat), 6 = wyższe (magisterium)), płeć (1 = mężczyzna, 2 = kobieta), roczne zarobki (w dolarach) oraz sektor zatrudnienia (5 = sektor prywatny, 6 = sektor publiczny, 7 = samozatrudnienie).
 - (a) Narysuj histogram zarobków. Opisz kształt, środek i rozrzut. Podaj wartości odpowiednich statystyk (minimum, maximum, mediana, kwartyle, średnia, odchylenie standardowe, wariancja, współczynnik zmienności).
 - (b) Powtórz punkt (a) osobno dla mężczyzn, a osobno dla kobiet. Porównaj wyniki.
 - (c) W punkcie (a) wyznacz obserwacje odstające.