

Wśród studentów jednej z krakowskich uczelni wykonano badania czasu dojazdu na uczelnię. Badania przeprowadzono wśród studentów dojeżdżających z Nowej Huty. O czasy dojazdu zapytano 100 osób, spośród których:

- 26 dojeżdża autobusem;
- 49 tramwajem;
- 20 na rowerze;
- 4 samochodem;
- 1 osoba dochodzi pieszo.

Czas dojazdu na uczelnię to cecha ilościowa, natomiast środek transportu – jakościowa. Cecha jakościowa dzieli cechę ilościową na kategorie, odpowiadające poszczególnym środkom transportu.

**Zad.1** Wśród 20 osób dojeżdżających na uczelnię na rowerze, 6 jest kobietami. Czasy dojazdów na rowerze dla tych kobiet wynoszą: 34, 36, 34, 36, 35, 37 minut. Wylicz średnią, wariancję, odchylenie, rozstęp dla tej próby i uzupełnij odpowiednio tabelę. Podobne parametry wyliczono dla grupy mężczyzn dojeżdżających na rowerach. Jakie wnioski można wysunąć porównując czasy dojazdu dla kobiet i mężczyzn?

Rowerzyści	$n$	$\bar{x}$	$s^2$	$s$	$R$
Mężczyźni	14	30,2	2,95	1,7	6
Kobiety					

**Zad.2** W tabeli przedstawiono czas dojazdu (w minutach) osób, które korzystają z tramwaju:

Czas w minutach	$\langle 20; 22 \rangle$	$\langle 22; 24 \rangle$	$\langle 24; 26 \rangle$	$\langle 26; 28 \rangle$	$\langle 28; 30 \rangle$	$\langle 30; 32 \rangle$	$\langle 32; 34 \rangle$
Liczba osób	2	8	12	12	7	5	4

Wylicz średnią, medianę, kwartyle, wariancję, odchylenie standardowe oraz rozstęp dla tej próby i uzupełnij poniższą tabelę. (warto wykorzystać atuty związane z pracą w grupie i podzielić się pracą związaną z liczeniem).

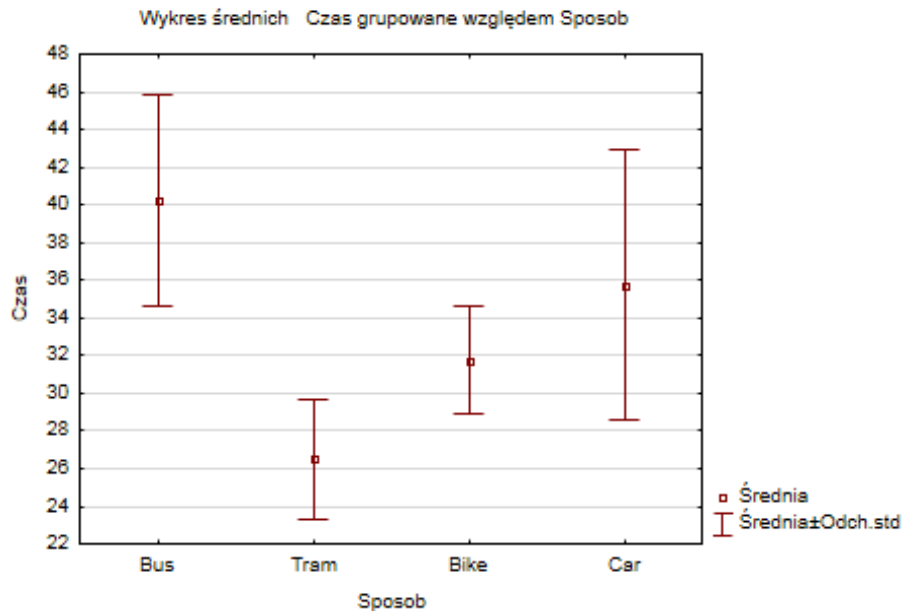
Podobne parametry wyliczono dla czasu dojazdu autobusem, samochodem i rowerem:

Sposób dojazdu	$n$	$\bar{x}$	$Me$	$Q_1$	$Q_3$	$s^2$	$s$	$R$
<b>Autobus</b>	26	40,2	38,5	36	43	31,94	5,7	22
<b>Tramwaj</b>								
<b>Samochód</b>	4	35,8	33,5	31	40,5	50,92	7,1	16
<b>Rower</b>	20	31,8	31	30,5	34	8,20	2,9	11
<b>Wszystkie łącznie</b>	100	32,4	30,5	26	36	118,88	10,9	95

**Zad.3** Na podstawie parametrów z powyższej tabeli, a także wykresu  $\bar{x} \pm s$  odpowiedz na pytania:

- a) Co można powiedzieć o czasie dojazdu na uczelnie różnymi środkami transportu porównując wartości średnie czasu dojazdu?
- b) A co porównując odchylenia standardowe?

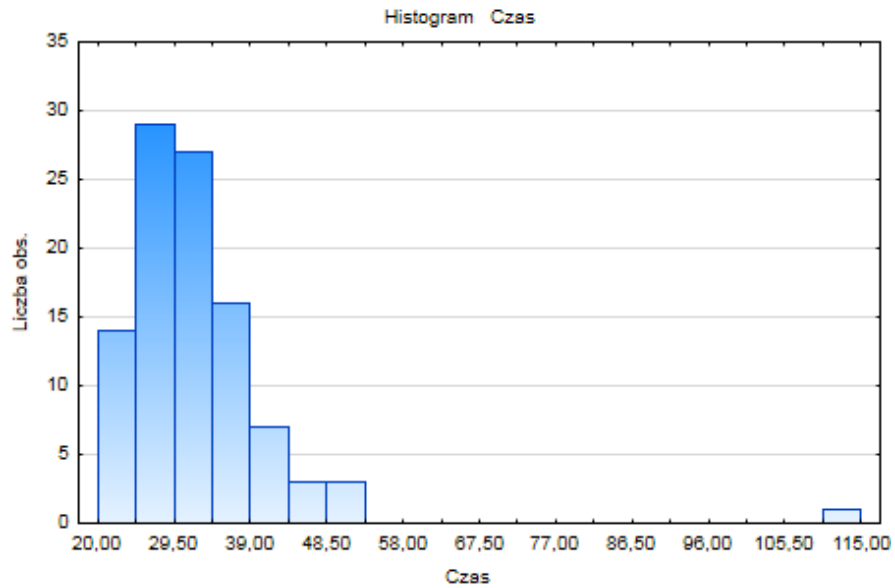
- c) Jakie podejście do wyliczenia parametrów statystyki opisowej daje więcej informacji: z podziałem na kategorie w zależności od środka transportu, czy analizowanie wszystkich osób łącznie?
- d) Wybierając który ze środków transportu jesteśmy w stanie najdokładniej określić godzinę przyjazdu na uczelnię?
- e) Jakie czynniki mogą mieć wpływ na takie wartości średnich i odchyłeń dla tych środków transportu?



**Zad.4** Wykres przedstawia histogram czasów dojazdu dla wszystkich osób, bez podziału na środki transportu. Wykorzystując dane z tabeli z zadania 2: zaznacz na nim wartość średnią, oraz wartości  $\bar{x} \pm s$ ,  $\bar{x} \pm 3s$ . Narysuj pod histogramem poziomy wykres pudełkowy, którego środek będzie równy wartości średniej, wielkość pudełka będzie ograniczona przez punkty  $\bar{x} \pm s$ , natomiast długość wąsów przez punkty  $\bar{x} \pm 3s$ . Obszar, który w ten sposób powstanie będzie ilustracją do często stosowanej w statystyce **reguły 3 sigma**, która definiuje **typowy obszar zmienności cechy**. Reguła mówi o tym, że gdy mamy do czynienia z rozkładem normalnym, to:

- 68% wartości cechy leży w przedziale  $\bar{x} \pm s$
- 95,5% wartości cechy leży w przedziale  $\bar{x} \pm 2s$
- 99,7% wartości cechy leży w przedziale  $\bar{x} \pm 3s$

Obserwacje, które wykraczają poza ten obszar mogą być rzeczywistymi obserwacjami odstającymi (nietypowymi) albo, gdy znacznie wykraczają, pomyłkami. Obserwacje odstające czasami się zdarzają, np. wzrost powyżej 2 metrów. Jednak gdy wzrost wynosi 2cm, to raczej mamy do czynienia z pomyłką w trakcie zbierania danych.



**Zad.5** Analizując czasy dojazdu na uczelnię w zależności od środka transportu, pieszego można potraktować jako obserwację nietypową. Po odrzuceniu z analizy osoby, która dochodzi na uczelnie pieszo, uzyskano następujące wartości parametrów:

<i>Sposób dojazdu</i>	<i>n</i>	$\bar{x}$	<i>Me</i>	$Q_1$	$Q_3$	$s^2$	<i>s</i>	<i>R</i>
<i>Wszystkie łącznie</i>	100	32,4	30,5	26	36	118,88	10,9	95
<i>Bez pieszego</i>	99	31,5	30	26	36	49,72	7,1	33

Jak zmieniły się te parametry po odrzuceniu pieszego z analizy?

Poniżej przedstawiono wskaźniki statystyki opisowej. Wskaźniki to miary liczbowe opisujące podstawowe własności rozkładu cechy. Przyjęto oznaczenia:  $n$  – liczebność próby;  $x_i$  – kolejne wartości cechy z szeregu szczegółowego.

### Miary położenia:

Średnia arytmetyczna:	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Mediana:	$Me = \begin{cases} x_{(n+1)/2}, & \text{gdyn} - \text{nieparzyste} \\ \frac{1}{2}(x_{n/2} + x_{1+n/2}), & \text{gdyn} - \text{parzyste} \end{cases}$
Pierwszy kwartył:	$Q_1 = x_{0,25(n-1)+1}$
Trzeci kwartył:	$Q_3 = x_{0,75(n-1)+1}$

### Miary zmienności:

Wariancja:	$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
Odchylenie standardowe:	$s = \sqrt{s^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
Odchylenie przeciętne:	$d = \frac{1}{n} \sum_{i=1}^n  x_i - \bar{x} $
Rozstęp:	$R = x_{max} - x_{min}$
Rozstęp międzykwartyłowy:	$IQR = Q_3 - Q_1$
Współczynnik zmienności:	$V_s = s/\bar{x}$