

Statystyczna analiza danych

Prezentacje tabelaryczne

Prezentacje graficzne

Miary położenia (średnie, tendencji centralnej)

Miary rozproszenia (zmienności)

Miary asymetrii

Miary zależności

Próba prosta (dane indywidualne)

$$X_1, X_2, \dots, X_n \quad (X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n})$$

Szereg rozdzielczy (dane skumulowane)

Przedział klasowy	Liczebność	Liczebność skumulowana
$x_0 - x_1$	n_1	$n_{(1)}$
$x_1 - x_2$	n_2	$n_{(2)}$
\vdots	\vdots	\vdots
$x_{k-1} - x_k$	n_k	$n_{(k)}$

Niech $0 \leq p \leq 1$

x_p : początek przedziału z obserwacją o numerze $p \cdot n$

n_p : liczebność przedziału z obserwacją o numerze $p \cdot n$

h_p : długość przedziału z obserwacją o numerze $p \cdot n$

$n_{(p)}$: liczebność skumulowana przedziału poprzedzającego przedział o początku x_p

$$\dot{x}_i = (x_{i-1} + x_i)/2$$

Mierniki położenia (próba prosta)

średnia

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k X_i$$

mediana

$$Me = \begin{cases} X_{(n+1)/2:n} & n \text{ nieparzyste} \\ (X_{n/2:n} + X_{n/2+1:n})/2 & n \text{ parzyste} \end{cases}$$

dolny kwartył

$$Q_1 = X_{[n/4]:n}$$

górnny kwartył

$$Q_3 = X_{[3n/4]:n}$$

dominanta (moda)

$D =$ najczęściej występująca wartość

minimum

$$Min = X_{1:n}$$

maksimum

$$Max = X_{n:n}$$

Mierniki położenia (szereg rozdzielczy) średnia

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k \dot{x}_i n_i$$

mediana

$$Me = x_{0.5} + \frac{h_{0.5}}{n_{0.5}} \left(\frac{n}{2} - n_{(0.5)} \right)$$

dolny kwartyl

$$Q_1 = x_{0.25} + \frac{h_{0.25}}{n_{0.25}} \left(\frac{n}{4} - n_{(0.25)} \right)$$

górnny kwartyl

$$Q_3 = x_{0.75} + \frac{h_{0.75}}{n_{0.75}} \left(\frac{3n}{4} - n_{(0.75)} \right)$$

dominanta (moda)

$$D = x_D + h_D \frac{n_D - n_{D-1}}{2n_D - n_{D+1} - n_{D-1}}$$

minimum

$$Min = x_0$$

maksimum

$$Max = x_k$$

Mierniki rozproszenia wariancja

$$S^2 = \begin{cases} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^2 \\ \frac{1}{n} \sum_{i=1}^k n_i (\dot{x}_i - \bar{x})^2 \end{cases}$$

odchylenie standardowe

$$S = \sqrt{S^2}$$

współczynnik zmienności

$$V = \frac{S}{\bar{x}} 100\%$$

rozstęp

$$R = Max - Min$$

odchylenie przeciętne

$$d = \begin{cases} \frac{1}{n} \sum_{i=1}^n |X_i - \bar{x}| \\ \frac{1}{n} \sum_{i=1}^k n_i |\dot{x}_i - \bar{x}| \end{cases}$$

odchylenie ćwiartkowe

$$Q = \frac{Q_3 - Q_1}{2}$$

Mierniki asymetrii

trzeci moment centralny

$$e_3 = \begin{cases} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^3 \\ \frac{1}{n} \sum_{i=1}^k n_i (\dot{x}_i - \bar{x})^3 \end{cases}$$

współczynnik asymetrii

$$A = \frac{e_3}{S^3}$$

pozycyjny współczynnik asymetrii

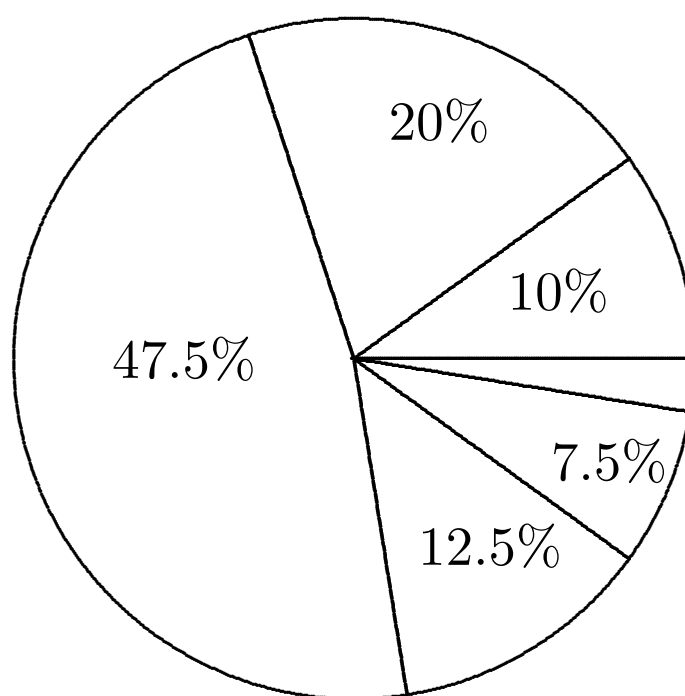
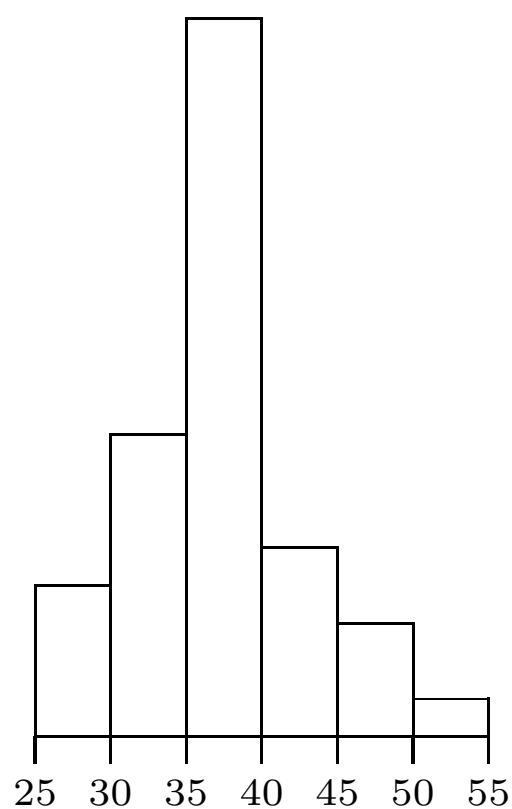
$$A_1 = \frac{Q_3 - 2Me + Q_1}{2Q}$$

współczynnik skośności

$$A_3 = \frac{\bar{x} - D}{S}$$

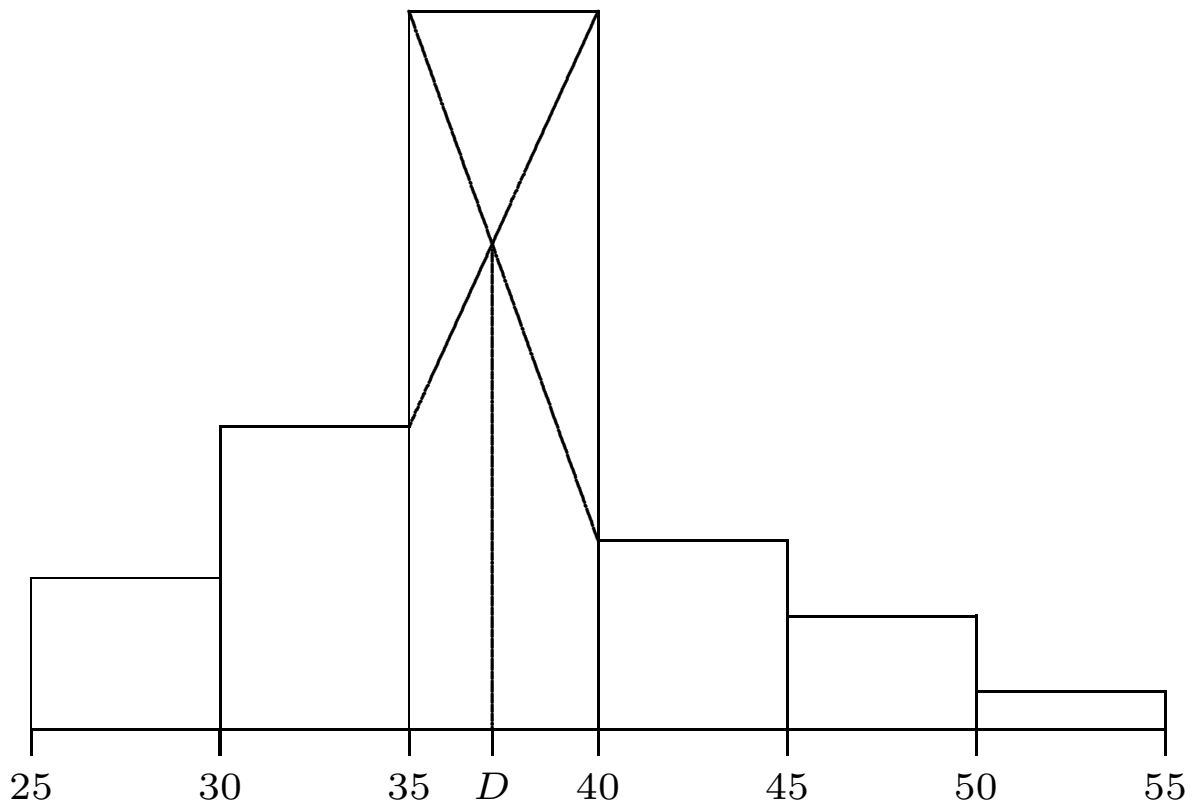
Przykład. Badano przebieg opon samochodowych wycofanych z eksploatacji.

Przebieg $x_{i-1} - x_i$	Liczba n_i	$n_{(i)}$	Odsetek
25 – 30	20	20	10.00%
30 – 35	40	60	20.00%
35 – 40	95	155	47.50%
40 – 45	25	180	12.50%
45 – 50	15	195	7.50%
50 – 55	5	200	2.50%



Średni przebieg dwustu opon

$$\bar{x} = \frac{27.5 \cdot 20 + 32.5 \cdot 40 + \dots + 52.5 \cdot 5}{200} = 37.25$$

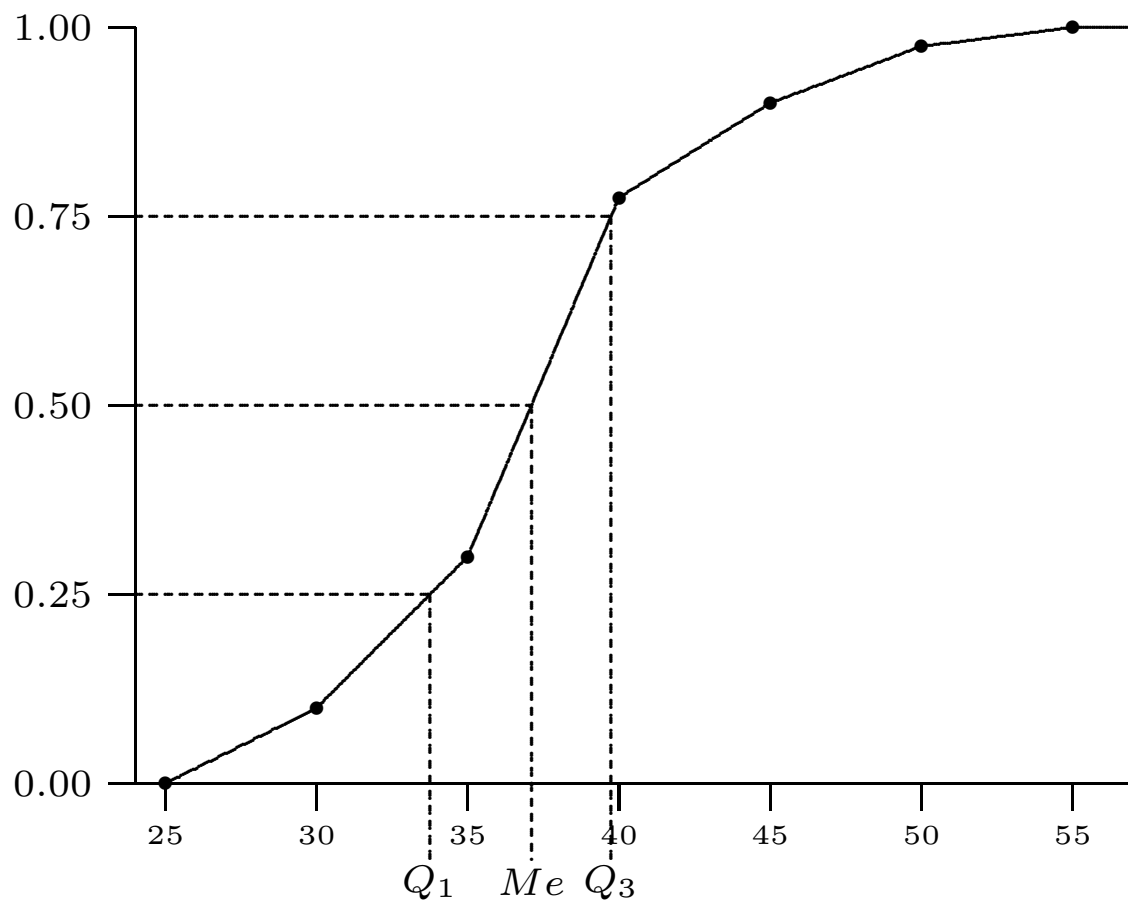


Dominanta przebiegu dwustu opon

$$x_D = 35 \quad h_D = 5$$

$$n_D = 95 \quad n_{D-1} = 40 \quad n_{D+1} = 25$$

$$D = 35 + 5 \cdot \frac{95 - 40}{2 \cdot 95 - 40 - 25} = 37.2$$



Mediana przebiegu dwustu opon

$$\frac{n}{2} = 100 \quad x_{0.5} = 35$$

$$h_{0.5} = 5 \quad n_{0.5} = 95 \quad n_{(0.5)} = 60$$

$$Me = 35 + \frac{5}{95} \cdot (100 - 60) = 37.11$$

Dolny kwartył przebiegu dwustu opon

$$\frac{n}{4} = 50 \quad x_{0.25} = 30$$

$$h_{0.25} = 5 \quad n_{0.25} = 40 \quad n_{(0.25)} = 20$$

$$Q_1 = 30 + \frac{5}{40} \cdot (50 - 20) = 33.75$$

Górny kwartył przebiegu dwustu opon

$$\frac{3n}{4} = 150 \quad x_{0.75} = 35$$

$$h_{0.75} = 5 \quad n_{0.75} = 95 \quad n_{(0.75)} = 60$$

$$Q_3 = 35 + \frac{5}{95} (150 - 60) = 39.74$$

Minimalny przebieg dwustu opon

$$Min = 25$$

Maksymalny przebieg dwustu opon

$$Max = 55$$

Wariancja

$$S^2 = \frac{1}{200} (20(27.5 - 37.25)^2 + 40(32.5 - 37.25)^2 + \dots + 5(52.5 - 37.25)^2) = 31.18$$

Odchylenie standardowe

$$S = \sqrt{S^2} = 5.58$$

Współczynnik zmienności

$$V = \frac{5.58}{37.25} \cdot 100\% = 14.99\%$$

Rozstęp

$$R = 55 - 25 = 30$$

Odchylenie przeciętne

$$d = \frac{1}{200} (20|27.5 - 37.25| + 40|32.5 - 37.25| + \dots + 5|52.5 - 37.25|) = 3.85$$

Odchylenie ćwiartkowe

$$Q = \frac{39.74 - 33.75}{2} = 2.99$$

Trzeci moment centralny

$$e_3 = \frac{1}{200} (20(27.5 - 37.25)^3 + 40(32.5 - 37.25)^3 + \dots \\ + 5(52.5 - 37.25)^3) = 73.406$$

Współczynnik asymetrii

$$A = \frac{73.406}{5.58^3} = 0.059$$

Pozycyjny współczynnik asymetrii

$$A_1 = \frac{39.74 - 2 \cdot 37.11 + 33.75}{2 \cdot 2.99} = -0.121$$

Współczynnik skośności

$$A_3 = \frac{37.25 - 37.2}{5.58} = 0.004$$

Koncentracja Lorentza

Przedział	Liczebność	Częstość
$x_0 - x_1$	n_1	$w_1 = n_1/n$
$x_1 - x_2$	n_2	$w_2 = n_2/n$
\vdots	\vdots	\vdots
$x_{k-1} - x_k$	n_k	$w_k = n_k/n$
Razem	n	1

Środek	t_i	z_i	$z_{(i)}$
\dot{x}_1	$t_1 = n_1 \dot{x}_1$	$z_1 = t_1/t$	$z_{(1)}$
\dot{x}_2	$t_2 = n_2 \dot{x}_2$	$z_2 = t_2/t$	$z_{(2)}$
\vdots	\vdots	\vdots	
\dot{x}_k	$t_k = n_k \dot{x}_k$	$z_k = t_k/t$	$z_{(k)}$
	t	1	

$$z_{(i)} = z_1 + \cdots + z_{i-1}$$

Współczynnik koncentracji Lorentza

$$K = 1 - \sum_{i=1}^k [z_{(i)} + z_{(i-1)}] w_i$$

Przykład. Wyznaczyć i porównać koncentrację utargów w dwóch sieciach sklepów

Sieć 1		Sieć 2	
utargi	sklepy	utargi	sklepy
2 – 4	4	2 – 4	45
4 – 6	10	4 – 6	5
6 – 8	3	6 – 8	0
8 – 10	2	8 – 10	5
10 – 12	1	10 – 12	45

Sieć 1

n_i	w_i	środek	t_i	z_i	$z_{(i)}$
4	0.20	3	12	0.11	0.11
10	0.50	5	50	0.45	0.55
3	0.15	7	21	0.19	0.74
2	0.10	9	18	0.16	0.90
1	0.05	11	11	0.10	1.00
20	1.00		112	1.00	

Współczynnik koncentracji: $K = 0.194643$

Sieć 2

n_i	w_i	środek	t_i	z_i	$z_{(i)}$
45	0.45	3	135	0.19	0.19
5	0.05	5	25	0.04	0.23
0	0.00	7	0	0.00	0.23
5	0.05	9	45	0.06	0.29
45	0.45	11	495	0.71	1.00
100	1.00		700	1.00	

Współczynnik koncentracji: $K = 0.284286$

