

# CHARAKTERYSTYKI LICZBOWE STRUKTURY ZBIOROWOŚCI (c.d.)

1. miary położenia - *wykład 2*
2. **miary zmienności** (dyspersji, rozproszenia)
3. miary asymetrii (skośności)
4. miary koncentracji

## MIARY ZMIENNOŚCI

Miary zmienności charakteryzują stopień zróżnicowania jednostek zbiorowości pod względem badanej cechy.

Miary zmienności dzielą się na miary klasyczne i pozycyjne.

1. miary klasyczne (wariancja, odchylenie standardowe, odchylenie przeciętne, współczynnik zmienności) oraz
2. miary pozycyjne (rozstęp, odchylenie ćwiartkowe, współczynnik zmienności).

# **Miary KLASYCZNE**

## **Wariancja, odchylenie standardowe, odchylenie przeciętne, współczynnik zmienności (klasyczny)**

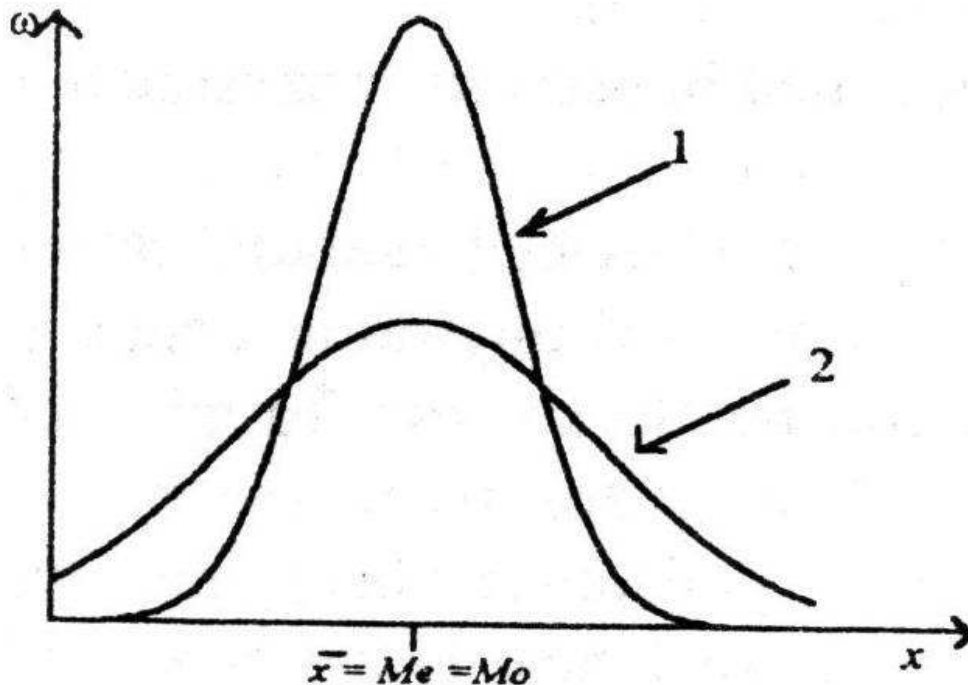
Wariancję ( $s^2$ ) definiuje się jako średnią arytmetyczną kwadratów odchyłeń wartości cechy od średniej arytmetycznej zbiorowości. Wariancja jest wielkością mianowaną w kwadracie miana badanej cechy i nie interpretujemy jej.

Odchylenie standardowe ( $s$ ) jest pierwiastkiem kwadratowym z wariancji. Jest ono wielkością mianowaną tak samo jak badana cecha. Odchylenie standardowe określa przeciętne zróżnicowanie badanej cechy od średniej arytmetycznej.

Odchylenie przeciętne ( $d$ ) jest średnią arytmetyczną bezwzględnych odchyłeń wartości cechy od jej średniej arytmetycznej. Jest ono wielkością mianowaną tak samo jak badana cecha. Odchylenie przeciętne interpretujemy podobnie jak odchylenie standardowe.

Współczynnik zmienności (klasyczny) ( $V_s$  lub  $V_d$ ) jest to iloraz odchylenia standardowego (lub przeciętnego) przez średnią arytmetyczną. Jest to wielkość niemianowana. Używamy go do porównań zmienności w dwu lub więcej zbiorowościach.

## Ocena rozproszenia na podstawie obserwacji diagramów



Na rysunku pokazano dwa diagramy częstości (1) i (2). Dla uproszczenia miary położenia (średnia, mediana i modalna) są sobie równe i identyczne dla obu zbiorowości.

- Mniejsze rozproszenie wokół średniej występuje w zbiorowości (1).  
Diagram jest smuklejszy i wyższy.
- Większe rozproszenie wokół średniej występuje w zbiorowości (2).  
Diagram jest bardziej rozłożysty i niższy.

Odchylenie standardowe w zbiorowości (1) jest mniejsze niż w zbiorowości (2)

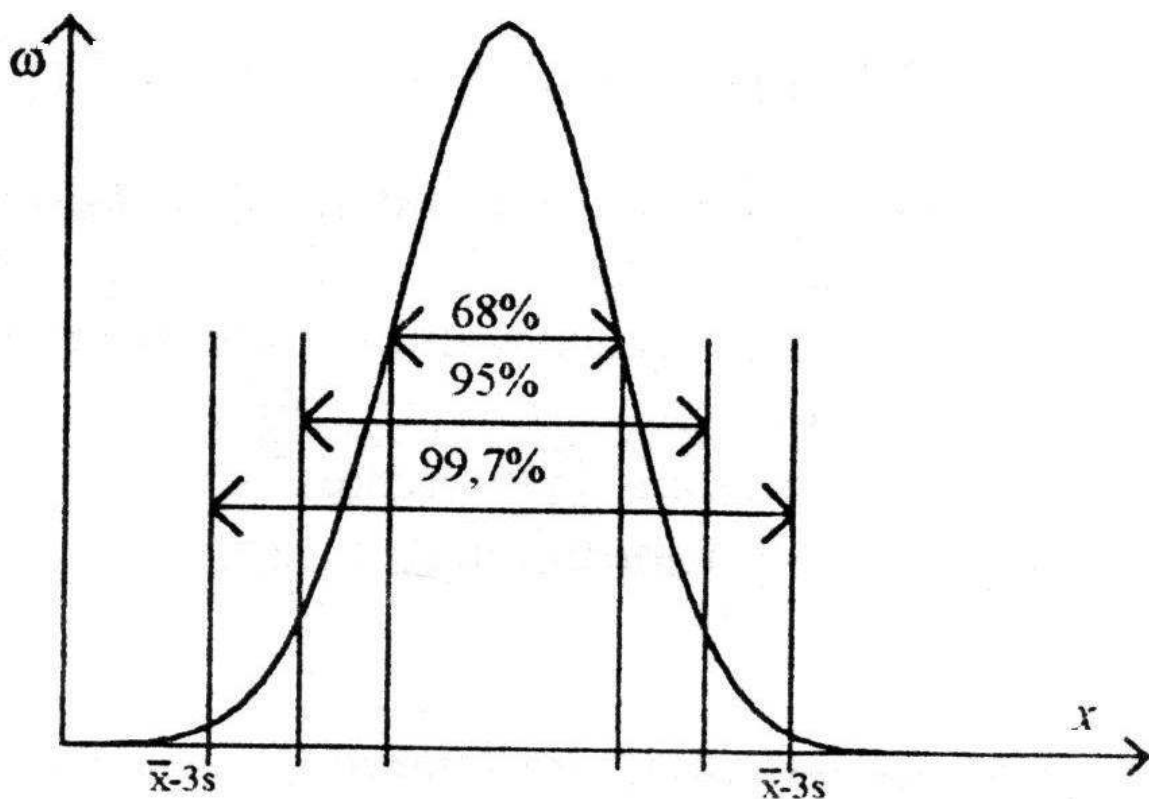
$$s_1 < s_2$$

## Przedział TYPOWYCH wartości cechy (miary klasyczne)

$$\bar{x} - s < x_{typ} < \bar{x} + s$$

Przedział taki ma tę własność, że około 70% jednostek badanej zbiorowości charakteryzuje się wartością cechy należącą do tego przedziału.

## Reguła „3 sigma”





**Wariancja liczby braków:**

$$\begin{aligned}s^2 &= \frac{(x_1 - \bar{x})^2 + \dots + (x_{50} - \bar{x})^2}{50} = \\ &= \frac{(0 - 0,8)^2 + \dots + (4 - 0,8)^2}{50} = \frac{68}{50} = 1,36\end{aligned}$$

**Odchylenie standardowe:**

$$s = \sqrt{s^2} = \sqrt{1,36} \approx 1,17$$

**Odchylenie przeciętne:**

$$\begin{aligned}d &= \frac{|x_1 - \bar{x}| + \dots + |x_{50} - \bar{x}|}{50} = \\ &= \frac{|0 - 0,8| + \dots + |4 - 0,8|}{50} = \frac{48}{50} = 0,96\end{aligned}$$

**Współczynnik zmienności (klasyczny)**

$$V_s = \frac{1,17}{0,8} = 1,46 \quad \text{lub} \quad V_d = \frac{0,96}{0,8} = 1,2$$

## Dla szeregów rozdzielczych punktowych

### Wariancja

$$s^2 = \frac{(x_1 - \bar{x})^2 n_1 + \dots + (x_k - \bar{x})^2 n_k}{n} = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 n_i}{n}$$

### Odchylenie standardowe

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 n_1 + \dots + (x_k - \bar{x})^2 n_k}{n}} = \sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2 n_i}{n}} = \sqrt{s^2}$$

### Odchylenie przeciętne

$$d = \frac{|x_1 - \bar{x}| n_1 + \dots + |x_k - \bar{x}| n_k}{n} = \frac{\sum_{i=1}^k |x_i - \bar{x}| n_i}{n}$$

W przykładzie z liczbą braków obliczenia przedstawia poniższa tabela.

<i>numer klasy</i>	<i>liczba braków</i>	<i>liczba wyrobów</i>	<i>obliczenia dla wariancji</i>			<i>odchylenie przeciętne</i>
<i>i</i>	<i>x<sub>i</sub></i>	<i>n<sub>i</sub></i>	<i>x<sub>i</sub> - x̄</i>	<i>(x<sub>i</sub> - x̄)<sup>2</sup></i>	<i>(x<sub>i</sub> - x̄)<sup>2</sup> n<sub>i</sub></i>	<i> x<sub>i</sub> - x̄  n<sub>i</sub></i>
<b>1</b>	<b>0</b>	<b>30</b>				
<b>2</b>	<b>1</b>	<b>8</b>				
<b>3</b>	<b>2</b>	<b>6</b>	<b>1,2</b>	<b>1,44</b>	<b>8,64</b>	<b>7,2</b>
<b>4</b>	<b>3</b>	<b>4</b>	<b>2,2</b>	<b>4,84</b>	<b>19,36</b>	<b>8,8</b>
<b>5</b>	<b>4</b>	<b>2</b>	<b>3,2</b>	<b>10,24</b>	<b>20,48</b>	<b>6,4</b>
<b>razem</b>	<b>×</b>	<b>50</b>	<b>×</b>	<b>×</b>		

$$s^2 = \frac{68}{50} = 1,36 \quad s = \sqrt{1,36} \approx 1,17 \quad d = \frac{48}{50} = 0,96$$

### Współczynnik zmienności (klasyczny)

$$V_s = \frac{1,17}{0,8} = 1,46 \quad \text{lub} \quad V_d = \frac{0,96}{0,8} = 1,2$$

## Dla szeregów rozdzielczych przedziałowych

### Wariancja

$$s^2 = \frac{(\dot{x}_1 - \bar{x})^2 n_1 + \dots + (\dot{x}_k - \bar{x})^2 n_k}{n} = \frac{\sum_{i=1}^k (\dot{x}_i - \bar{x})^2 n_i}{n}$$

### Odchylenie standardowe

$$s = \sqrt{\frac{(\dot{x}_1 - \bar{x})^2 n_1 + \dots + (\dot{x}_k - \bar{x})^2 n_k}{n}} = \sqrt{\frac{\sum_{i=1}^k (\dot{x}_i - \bar{x})^2 n_i}{n}} = \sqrt{s^2}$$

### Odchylenie przeciętne

$$d = \frac{|\dot{x}_1 - \bar{x}| n_1 + \dots + |\dot{x}_k - \bar{x}| n_k}{n} = \frac{\sum_{i=1}^k |\dot{x}_i - \bar{x}| n_i}{n}$$

### PRZYKŁAD 2 - czas dojazdu pracowników firmy ZAUR

numer klasy	czas dojazdu	środek klasy	liczba pracow.	obliczenia dla wariancji			odchylenie przeciętne
$i$	$x_{0i} - x_{1i}$	$\dot{x}_i$	$n_i$	$\dot{x}_i - \bar{x}$	$(\dot{x}_i - \bar{x})^2$	$(\dot{x}_i - \bar{x})^2 n_i$	$ \dot{x}_i - \bar{x}  n_i$
1	5 – 15	10	10				
2	15 – 25		20				
3	25 – 35	30	30	-10	100	3000	300
4	35 – 45	40	50	0	0	0	0
5	45 – 55	50	80	10	100	8000	800
6	55 – 65	60	10	20	400	4000	200
razem	×	×	200	×	×		

Jak pamiętamy:  $n=200$      $\bar{x} = 40$  [minut]

$$s^2 = \frac{32000}{200} = 160 \quad s = \sqrt{160} \approx 12,7 \quad d = \frac{2000}{200} = 10$$

### Współczynnik zmienności (klasyczny)

$$V_s = \frac{12,7}{40} = 0,32 \quad \text{lub} \quad V_d = \frac{10}{40} = 0,25$$



# Miary POZYCYJNE

## Rozstęp, odchylenie ćwiartkowe, współczynnik zmienności (pozycyjny)

**Rozstęp (  $R$  )** definiuje się jako różnicę pomiędzy największą i najmniejszą wartością cechy:

$$R = x_{\max} - x_{\min}$$

**Odchylenie ćwiartkowe (  $Q$  )** jest miarą rozproszenia wartości cechy od mediany. Definiuje się go jako połowę różnicy pomiędzy trzecim i pierwszym kwartylem:

$$Q = \frac{Q_{III} - Q_I}{2}$$

Odchylenie ćwiartkowe mierzy poziom zróżnicowania połowy jednostek populacji. Odrzucane są jednostki o wartościach badanej cechy poniżej pierwszego kwartyla (25%) oraz powyżej trzeciego kwartyla (25%).

**Współczynnik zmienności (pozycyjny) jest to iloraz odchylenia ćwiartkowego przez medianę. Jest to wielkość niemianowana. Używamy jej do porównań zmienności w dwu lub więcej zbiorowościach.**

$$V_Q = \frac{Q}{M_e}$$

### **Przedział TYPOWYCH wartości cechy (miary pozycyjne)**

**Definiujemy go podobnie jak w przypadku miar klasycznych (rolę średniej przejmuje tutaj mediana, a rolę odchylenia standardowego – odchylenie ćwiartkowe)**

$$M_e - Q < x_{typ} < M_e + Q$$

**Przedział ten będzie węższy od przedziału dla miar klasycznych.**



**PRZYKŁAD 4**

Weźmy dane z przykładu 7 (wykład 2):

10, 10, 10, 12, 12, 12, 12, 13, 13, 13,  
13, 13, 14, 14, 15, 15, 15

Rozstęp:  $R = x_{\max} - x_{\min} = 15 - 10 = 5$

Odchylenie ćwiartkowe:

$$Q_I = (x_4 + x_5)/2 = (12 + 12)/2 = 12$$

$$Q_{II} (M_e) = x_9 = 13$$

$$Q_{III} = (x_{13} + x_{14})/2 = (14 + 14)/2 = 14$$

$$Q = \frac{Q_{III} - Q_I}{2} = \frac{14 - 12}{2} = 1$$

Współczynnik zmienności (pozycyjny):

$$V_Q = \frac{Q}{M_e} = \frac{1}{13} \approx 0,077$$

Przedział typowych wartości cechy (pozycyjny):

$$M_e - Q < x_{typ} < M_e + Q$$

$$13 - 1 < x_{typ} < 13 + 1$$

$$12 < x_{typ} < 14$$

## Dla szeregów rozdzielczych punktowych

### PRZYKŁAD 5

Dane z przykładu 5 (wykład 2).

<i>numer klasy</i>	<i>czas obróbki [minuta]</i>	<i>liczba pracow- ników</i>	<i>liczebność skumulowana</i>
<i>i</i>	<i>x<sub>i</sub></i>	<i>n<sub>i</sub></i>	<i>n<sub>i sk</sub></i>
<b>1</b>	<b>10</b>	<b>10</b>	<b>10</b>
<b>2</b>	<b>11</b>	<b>30</b>	
<b>3</b>	<b>12</b>	<b>80</b>	
<b>4</b>	<b>13</b>	<b>50</b>	
<b>5</b>	<b>14</b>	<b>20</b>	
<b>6</b>	<b>15</b>	<b>10</b>	<b>200</b>
<b>razem</b>	<b>×</b>	<b>200</b>	<b>×</b>

Rozstęp:  $R = x_{\max} - x_{\min} = 15 - 10 = 5$

Odchylenie ćwiartkowe:

$$Q_I = x_3 = 12$$

$$Q_{II} (M_e) = x_3 = 12$$

$$Q_{III} = x_4 = 13$$

$$Q = \frac{Q_{III} - Q_I}{2} = \frac{13 - 12}{2} = 0,5$$

Współczynnik zmienności (pozycyjny):

$$V_Q = \frac{Q}{M_e} = \frac{0,5}{12} \approx 0,042$$

Przedział typowych wartości cechy (pozycyjny):

$$12 - 0,5 < x_{typ} < 12 + 0,5$$

$$11,5 < x_{typ} < 12,5$$

## Dla szeregów rozdzielczych przedziałowych

**PRZYKŁAD 6** Dane z przykładu 10 (wykład 2).

<i>numer klasy</i>	<i>czas dojazdu w ZAUR</i>	<i>liczba pracow- ników</i>	<i>skumul. liczebność</i>
<i>i</i>	$x_{0i} - x_{1i}$	$n_i$	$n_{i\ sk}$
<b>1</b>	<b>5 – 15</b>	<b>10</b>	<b>10</b>
<b>2</b>	<b>15 – 25</b>	<b>20</b>	
<b>3</b>	<b>25 – 35</b>	<b>30</b>	
<b>4</b>	<b>35 – 45</b>	<b>50</b>	
<b>5</b>	<b>45 – 55</b>	<b>80</b>	
<b>6</b>	<b>55 – 65</b>	<b>10</b>	<b>200</b>
<b><i>razem</i></b>	<b>×</b>	<b>200</b>	<b>×</b>

Rozstęp:  $R = x_{\max} - x_{\min} = 65 - 5 = 60$

Odchylenie ćwiartkowe:

$$Q_I \approx 31,7$$

$$Q_{II} (M_e) = 43$$

$$Q_{III} = 50$$

$$Q = \frac{Q_{III} - Q_I}{2} = \frac{50 - 31,7}{2} \approx 9,2$$

Współczynnik zmienności (pozycyjny):

$$V_Q = \frac{Q}{M_e} = \frac{9,2}{43} \approx 0,213$$

Przedział typowych wartości cechy (pozycyjny):

$$43 - 9,2 < x_{typ} < 43 + 9,2$$

$$33,8 < x_{typ} < 52,2$$

**Przykład 7 (*praca domowa*)****Płace (stawka godzinowa) w firmach A, B i C**

<i>klasa</i>	Stawka [zł/godz.]		liczba pracowników ( $n_i$ )		
$i$	$x_{0i}$	$x_{1i}$	firma A	firma B	firma C
1	2	4	15	15	20
2	4	6	30	105	50
3	6	8	60	75	50
4	8	10	30	75	70
5	10	12	15	30	10
×	<i>razem</i>				

W ramach ćwiczenia wyznacz w każdej z firm:

1. średnią
2. wariancję
3. odchylenie standardowe
  
4. medianę
5. modalną
  
6. na wspólnym wykresie narysuj diagramy częstości stawki w firmach A, B i C

*Uzyskany materiał będzie podstawą dla kolejnego wykładu.*