STATYSTYKA –wiadomości wstępne

- Co to jest statystyka
- Jak zbieramy i porządkujemy dane statystyczne
- Jak prezentujemy graficznie dane statystyczne
- Jakie są metody analizy statystycznej

STATYSTYKA –wiadomości wstępne

Termin **statystyka** współcześnie ma kilka znaczeń:

- zbiór danych liczbowych, przedstawiających kształtowanie się określonych zjawisk i procesów,
- wszelkie prace związane z gromadzeniem i opracowywaniem danych liczbowych,
- charakterystyki opisowe obliczane ze zbiorowości próbnych np. średnia arytmetyczna,
- dyscyplina naukowa mająca własne metody badawcze – nauka o ilościowych metodach badania prawidłowości występujących w zjawiskach masowych

STATYSTYKA –wiadomości wstępne

Statystyka jako nauka dzieli się na:

• **statystykę opisową (opis statystyczny),** która zajmuje się metodami gromadzenia, opracowania i prezentacji danych wraz z ich sumarycznym opisem, zajmuje się wstępnym opracowaniem próbki *bez posługiwania się rachunkiem prawdopodobieństwa.*

• **statystykę matematyczną (wnioskowanie statystyczne),** która zajmuje się metodami wnioskowania o całej zbiorowości na podstawie zbadania pewnej jej części, czyli próby, zajmuje się opisywaniem i analizą zjawisk masowych za pomocą *metod rachunku prawdopodobieństwa.*

STATYSTYKA –wiadomości wstępne

Metody statystyki:

- **opis statystyczny** - jest to liczbowy opis badanych zbiorowości:

· opis tabelkowy (szeregi, tablice)

· opis graficzny (wykresy)

· opis parametryczny (charakterystyki liczbowe, parametry, średnie, dominanta, mediana).

- **wnioskowanie statystyczne** wykorzystuje się, gdy badaniu statystycznemu nie jest poddawana cała zbiorowość (tzw. populacja), tylko część tej zbiorowości wybrana na drodze losowania (tzw. próba losowa). Wnioskowanie statystyczne polega na uogólnieniu wyników otrzymanych w badaniu próby losowej na całą populację, z której ta próba pochodzi.

STATYSTYKA –wiadomości wstępne

Zbiorowość statystyczna (populacja) to zbiór jednostek (osób, przedmiotów, zdarzeń) objętych badaniem statystycznym, które mają jedną lub kilka cech wspólnych oraz wiele cech różnicujących (zmiennych).

Zbiorowość statystyczna musi być jednoznacznie określona pod względem rzeczowym, przestrzennym oraz czasowym.

Jednostka statystyczna to najmniejszy element zbiorowości statystycznej objętej badaniem.

Zbiorowość (populacja) generalna - wszystkie elementy będące przedmiotem badania, co do których chcemy formułować wnioski ogólne.

Zbiorowość próbna (próba) - podzbiór populacji generalnej; wyniki badań próby są uogólniane na zbiorowość generalną. Próba musi być reprezentatywna.

Reprezentatywność zależy od: sposobu wyboru jednostek (celowy, losowy) oraz liczebności próby.

$n>30$ - duża próba, $n\leq 30$ - mała próba

STATYSTYKA –wiadomości wstępne

Cechy statystyczne (zmienne) to właściwości jednostek statystycznych tworzących badaną zbiorowość. Dzieli się je na cechy:

- jakościowe (niemierzalne) -podawane opisowo, np. płeć
- ilościowe (mierzalne) -podawane liczbowo, np. wiek
  - skokowe (dyskretne) – np. liczba kaw
  - ciągłe- np. wiek
  - quasi ciągłe, np. płace, zarobki, ceny

STATYSTYKA –wiadomości wstępne

Badanie statystyczne

Badanie statystyczne to zespół czynności zmierzających do uzyskania, za pomocą metod statystycznych, informacji charakteryzujących zbiorowość statystyczną objętą badaniem. Badanie statystyczne umożliwia wykrycie lub potwierdzenie istniejących prawidłowości statystycznych.

Etapy badania statystycznego:

- przygotowanie badania,
- obserwacja statystyczna,
- opracowanie i prezentacja materiału statystycznego,
- analiza statystyczna.

STATYSTYKA –wiadomości wstępne

Szeregi statystyczne - odpowiednio usystematyzowany i uporządkowany surowy materiał statystyczny.

Szeregi statystyczne dzielimy na szeregi:

- szczegółowe
- rozdzielcze (punktowe, przedziałowe)
- czasowe (momentów, okresów)

STATYSTYKA –wiadomości wstępne

Przykład 1. Szereg szczegółowy

Przedmiotem badania jest wadliwość produkcji na I zmianie w firmie Alfa. Liczba wyprodukowanych wyrobów wynosi 50 (*n*=50).

Cecha badana (X) oznacza liczbę usterek w wyrobie.

<u>Surowy materiał statystyczny</u> to ciąg 50 liczb oznaczający liczbę braków stwierdzonych w kolejnym wyrobie. Ma on postać:

3, 0, 1, 0, 0, 0, 0, 1, 0, 0, 4, 2, 0, 0, 1, 1, 0, 3, 0, 2, 0, 1, 0, 0, 1, 0, 2, 0, 0, 0, 3, 0, 0, 4, 0, 2, 0, 0, 1, 0,

0, 1, 3, 0, 2, 0, 0, 0, 0, 2

<u>**Szereg szczegółowy**</u> otrzymamy <u>sortując</u> te liczby <u>rosnąco</u> (najczęściej) lub <u>malejąco</u> (rzadziej). Liczby braków posortowane rosnąco dają następujący ciąg $\{x_i\}$ 50 liczb:

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2,

2, 2, 2, 2, 3, 3, 3, 3, 4, 4

STATYSTYKA –wiadomości wstępne

Szereg rozdzielczy punktowy

numer klasy	liczba braków	liczba wyrobów (liczebność)
i	x_i	n_i
1	0	30
2	1	8
3	2	6
4	3	4
5	4	2
razem	×	50

Szereg rozdzielczy przedziałowy

czas dojazdu do pracy $\langle x_{i}, x_{I+1})$	liczba pracowników
5 - 15	5
15 - 25	55
25 - 35	20
35 - 45	10
45 - 55	5
55 - 65	5
razem	100

STATYSTYKA –wiadomości wstępne

Przykład: Jak się tworzy szereg rozdzielczy przedziałowy?

Punkty	Liczebności
10,5	1
11,1	2
12,5	1
13,0	2
13,9	1
14,4	3
15,7	1
15,8	2
16,5	3
16,9	1
17,4	2
18,5	1

Liczbę przedziałów k liczymy wg wzoru:

1. $k = \sqrt{N} = \sqrt{20}$, k = 4,472135955
2. $k \leq 1 + 3{,}222\ logN$; k = 5,191918646
3. $k \leq 5logN$; k = 6,505149978

STATYSTYKA –wiadomości wstępne

- Wariant dla $k = 5$

$$x_{min} = 10{,}5, \qquad x_{max} = 18{,}5$$

$$i - rozpiętość\ przedziałów\ ,\ i = \frac{x_{max} - x_{min}}{k}$$

Tablica 4. Szereg rozdzielczy przedziałowy. Wariant $k = 5$

i	x0i	x1i	ni
1	10,5	12,1	3
2	12,1	13,7	3
3	13,7	15,3	4
4	15,3	16,9	7
5	16,9	18,5	3

STATYSTYKA –wiadomości wstępne

Przykład Szereg czasowy

Produkcja energii elektrycznej w Polsce w latach 1991-1994
(dane miesięczne w mld kWh)

Miesiące / Lata	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1991	14,0	12,9	12,8	11,3	10,6	9,2	9,0	9,0	9,3	11,2	12,1	13,3
1992	13,4	12,2	12,5	11,0	9,6	9,0	9,0	9,0	9,7	12,0	12,3	13,2
1993	13,4	12,2	12,8	10,7	9,3	8,9	8,7	9,1	9,8	11,5	13,0	13,5
1994	12,9	12,3	12,7	10,9	9,9	9,4	9,3	9,5	9,8	12,3	12,3	13,6

STATYSTYKA –wiadomości wstępne

Szereg czasowy momentów, to szereg zawierający informacje o poziomach badanego zjawiska w określonych momentach pewnego przedziału czasowego. Z kolei **szereg czasowy okresów** zawiera informacje o rozmairach zjawiska w ciągu kolejnych okresów danego przedziału czasowego.

Przykład 1. Szereg czasowy momentów:

data kalendarzowa	31 XII 2000	31 XII 2001	31 XII 2002	31 XII 2003	31 XII 2004	31 XII 2005	31 XII 2006
stan ludności Polski w tys.	38254, 0	38242, 2	38218, 5	38190, 6	38173, 8	38157, 1	38125, 5

Źródło: Roczniki Demograficzne.

Przykład 2. Szereg czasowy okresów:

lata	2000	2001	2002	2003	2004	2005	2006
urodzenia żywe w Polsce w tys.	378, 3	368, 2	353, 8	351, 1	356, 1	364, 4	374, 2

Źródło: Roczniki Demograficzne.

https://eks.stat.gov.pl/

STATYSTYKA –wiadomości wstępne

Prezentacja graficzna szeregów strukturalnych i czasowych

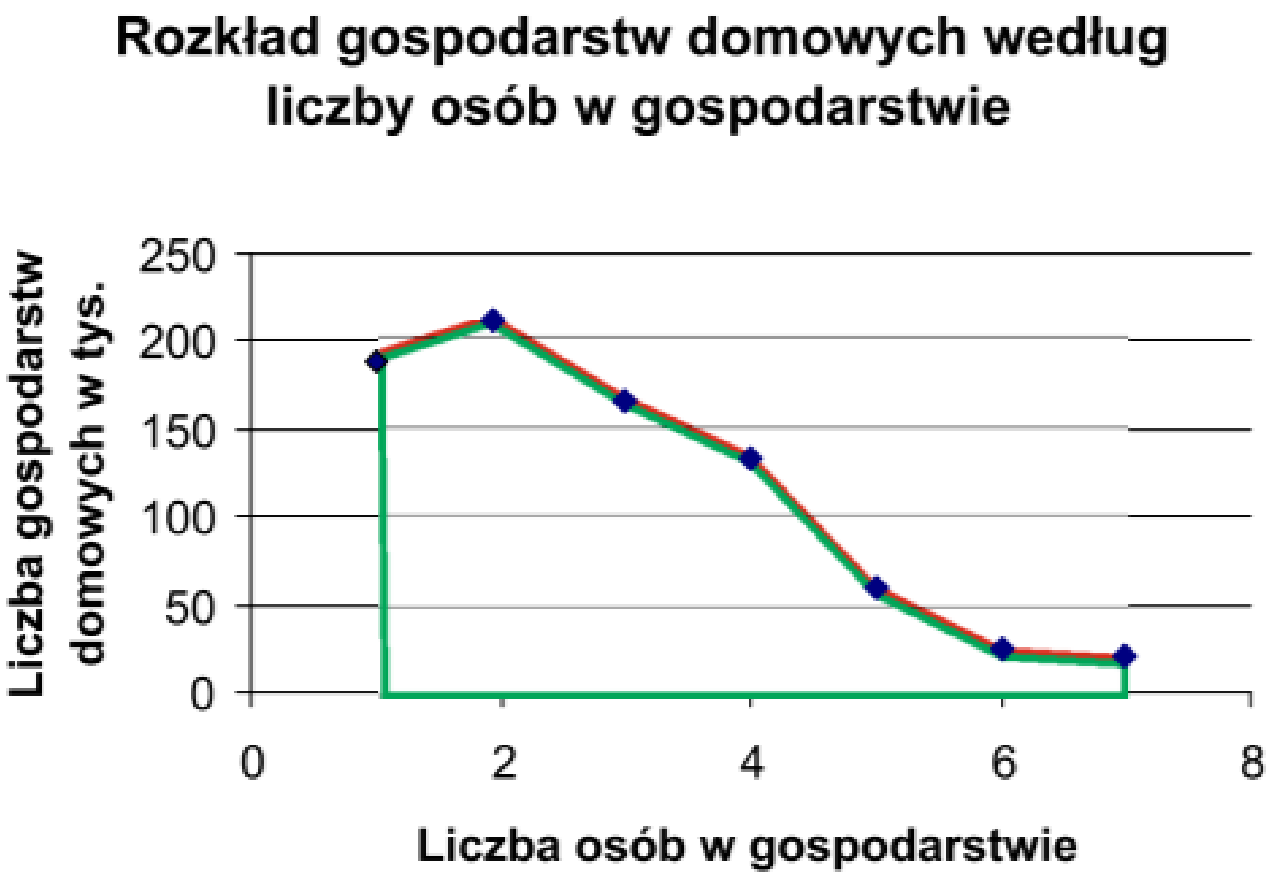
Ważnym narzędziem uwidaczniania prawidłowości występujących w zbiorowości statystycznej jest prezentacja graficzna materiału statystycznego w postaci wykresów.

Do najczęściej stosowanych graficznych form prezentacji zalicza się

wykresy: punktowe, liniowe i powierzchniowe.

- **Wykres punktowy** ma postać punktów, z których każdy prezentuje określoną liczbę jednostek zbiorowości, posiadających ten sam wariant cechy ilościowej. Jeżeli te punkty połączymy to otrzymamy **diagram.**
- Z kolei jeżeli diagram domkniemy liniami prostopadłymi do osi poziomej poprowadzonymi przez najniższą i najwyższą wartość cechy to otrzymamy **wielobok liczebności** (rys. 1).

STATYSTYKA –wiadomości wstępne



Rysunek 1. Wykres prezentujący szereg rozdzielczy punktowy

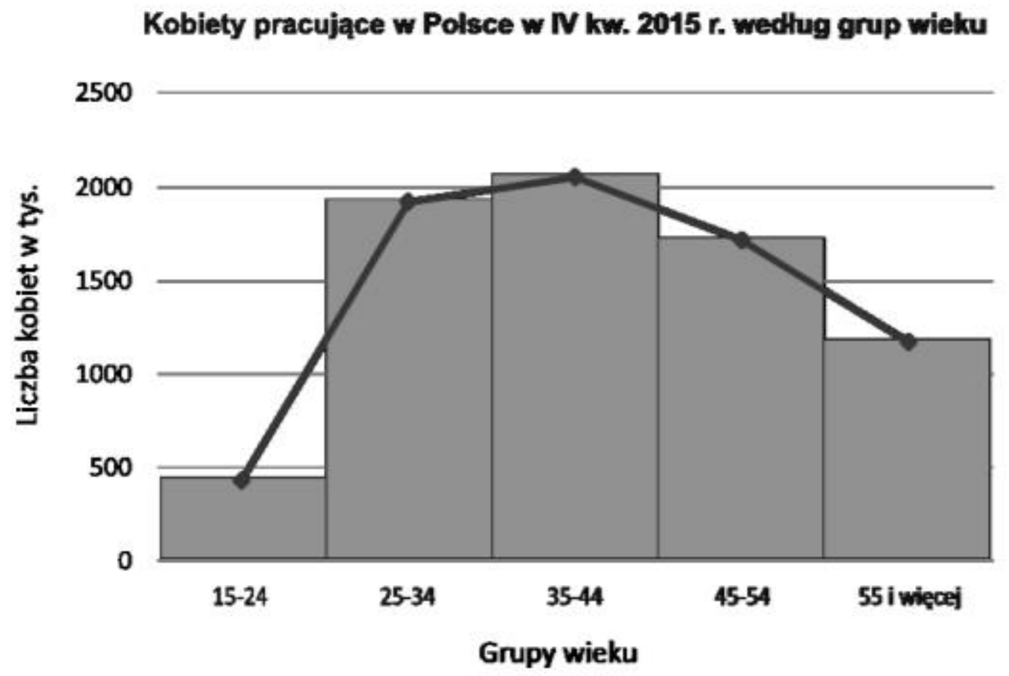
STATYSTYKA –wiadomości wstępne

Natomiast do prezentacji **szeregów rozdzielczych przedziałowych** zastosowanie ma wykres powierzchniowy zwany **histogramem**.

Stanowią go pola przystających do siebie prostokątów (słupków), przy czym jeden z boków każdego prostokąta ma długość odpowiadającą rozpiętości, a drugi liczebności przedziału klasowego.

Można również ten szereg przedstawić za pomocą wykresu liniowego zwanego **krzywą liczebności** (rys. 2).

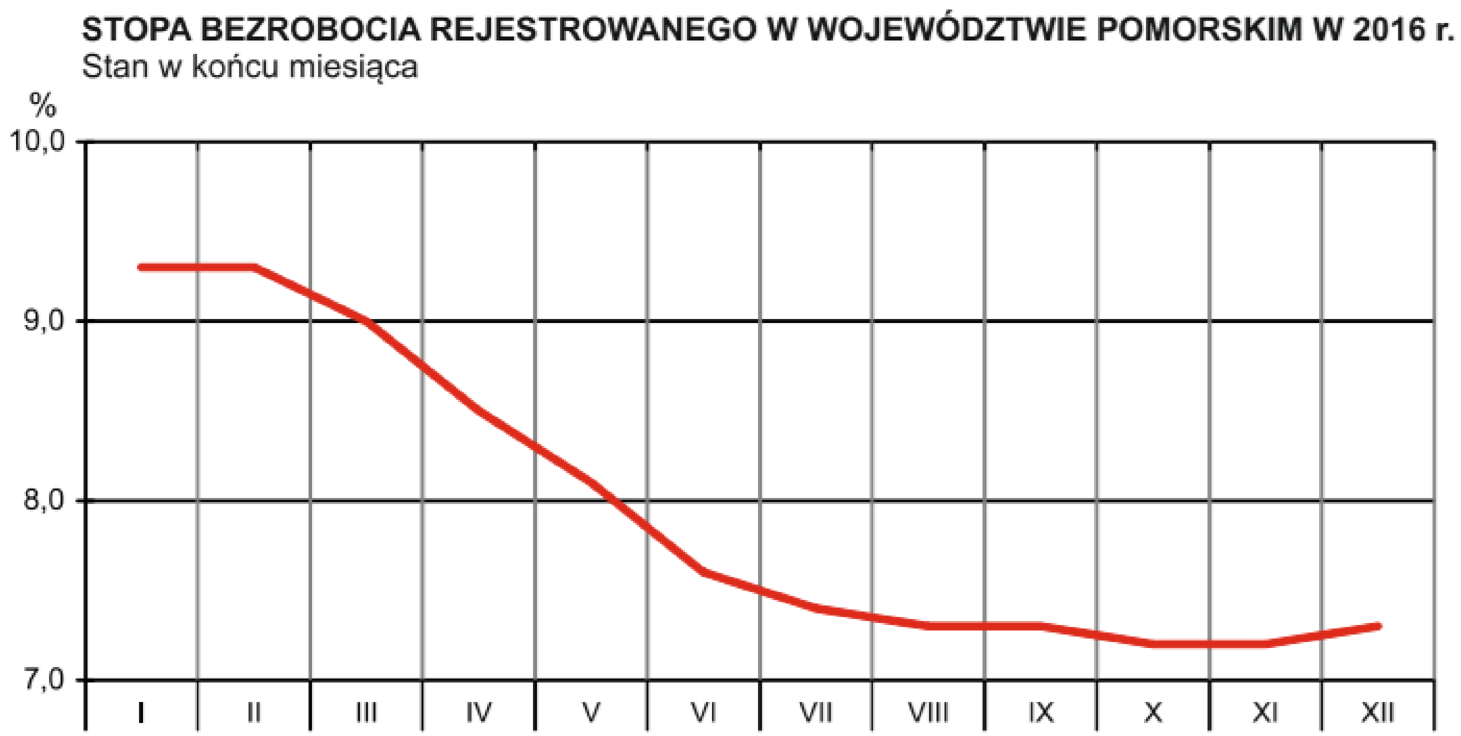
Krzywa powstaje z połączenia punktów, których współrzędnymi są środki przedziałów klasowych i liczebności poszczególnych przedziałów.



Rysunek 2. Wykres prezentujący szereg rozdzielczy przedziałowy

STATYSTYKA –wiadomości wstępne

Do graficznej prezentacji szeregów czasowych wykorzystuje się wykresy liniowe, przy czym na osi OX zaznaczamy jednostki czasu, a na osi OY wielkość obserwowanego zjawiska (rys. 4).



Rysunek 4. Wykres prezentujący szereg czasowy

STATYSTYKA –wiadomości wstępne

Metody analizy statystycznej

W ramach statystyki opisowej można wyróżnić trzy podstawowe działy analizy:

- **analizę struktury zbiorowości**, która pozwala ustalić, jak są rozłożone poszczególne warianty cechy zmiennej wśród jednostek badanej zbiorowości statystycznej,
- **analizę współzależności zjawisk**, która zajmuje się badaniem powiązań między różnymi cechami zmiennymi charakteryzującymi zbiorowość statystyczną,
- **analizę dynamiki zjawisk**, której zadaniem jest określenie zmian zachodzących w kształtowaniu się cechy zmiennej w czasie.

METODY ANALIZY STATYSTYCZNEJ

Metody analizy statystycznej

- Analiza struktury zbiorowości
  - Wskaźnik struktury i natężenia
  - Miary położenia
  - Miary zmienności
  - Miary asymetrii
- Analiza współzależności zjawisk
  - Analiza korelacji
  - Analiza regresji
- Analiza dynamiki zjawisk
  - Miary indeksowe
  - Dekompozycja szeregu czasowego

STATYSTYKA –wiadomości wstępne

Wskaźnik struktury (częstość, liczebność względna, frakcja, odsetek) to stosunek liczby jednostek o danej wartości cechy zmiennej do łącznej liczebności zbiorowości

$$\omega_i = \frac{n_i}{N}$$

gdzie: n_i – liczebność cząstkowa określająca, ile jednostek zbiorowości przypada na daną wartość cechy zmiennej,N – liczebność zbiorowości.

Wskaźnik natężenia to stosunek liczby jednostek (wartości cechy) danej zbiorowości do liczby jednostek (wartości cechy) innej zbiorowości, które pozostają w przyczynowym lub logicznym związku $v_i = \frac{n_i}{m_i}$

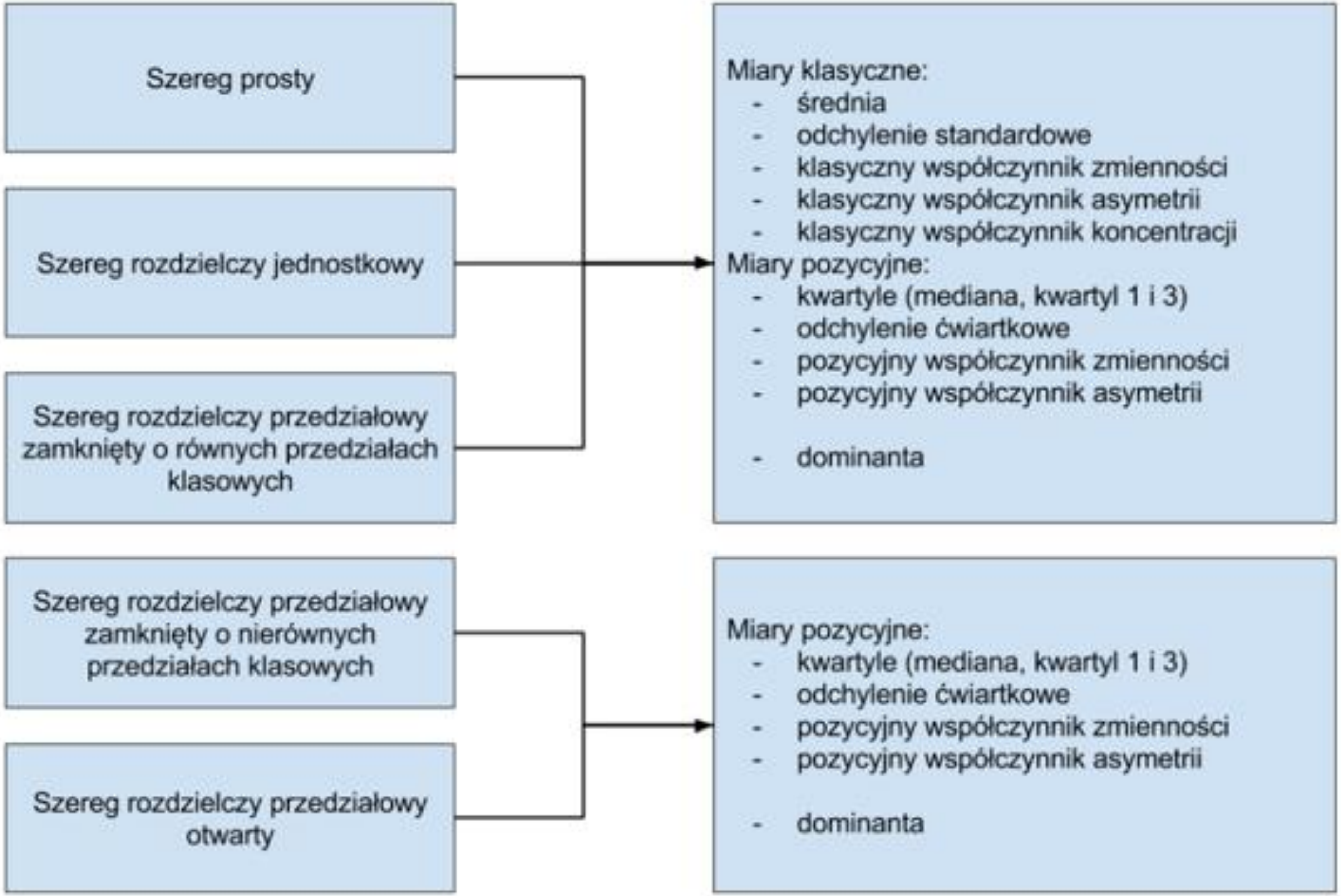
gdzie: n_i – liczba jednostek jednej zbiorowości, m_i – liczba jednostek drugiej zbiorowości.

STATYSTYKA –wiadomości wstępne

Parametry opisowe dzieli się na:

- **parametry klasyczne**, które liczone są na podstawie wartości cechy zmiennej wszystkich jednostek badanej zbiorowości,
- **parametry pozycyjne**, które wyznaczane są na podstawie wartości cechy zmiennej wybranych jednostek badanej zbiorowości zajmujących szczególną pozycję w szeregu statystycznym.

	KLASYCZNE	POZYCYJNE
MIARY POŁOŻENIA	- **średnia** $\bar{x}$ • arytmetyczna, • harmoniczna – gdy dane są w jednostkach względnych • geometryczna – badanie średniego poziomu zjawiska • chronologiczna – szeregi czasowe	-**mediana Me** -**kwantyle:** • kwartyle (4 części) • decyle (10 części) • centyle (percentyle) (100 części) -**dominanta** – MIARA DLA DANYCH JAKOŚCIOWYCH
MIARY ZMIENNOŚCI (ROZPROSZENIA, DYSPERSJI) – odchylenie od średniej	- **Odchylenie standardowe** S - **wariancja** S^2 - **współczynnik zmienności** V_S	-**Odchylenie ćwiartkowe Q** $Q = \frac{Q_3 - Q_1}{2}$ -**współczynnik zmienności** $V_S = \frac{Q}{Me}$ -**rozstęp** $\boldsymbol{R = x_{min} - x_{max}}$
MIARY ASYMETRII • Lewostronna $A < 0$ lub $D < \bar{X}$ • Prawostronna $A > 0$ lub $\bar{X} < D$ • $A \in < -1,1 >$	**Współczynnik asymetrii** $A_s = \frac{m_3}{S^3},$ $m_3 - moment\ centralny\ III\ rzędu$	**Współczynnik asymetrii** $AA_s = \frac{Q_3 + Q_1 - 2Me}{2Q}$
MIARY KURTOZY (koncentracji wokół średniej) K Bada skupinie wokół średniej- porównanie do normalnego • normalny to K=3 • Wykres spłaszczony (K<3) • Wykres smukły (K>3)	• Współczynnik kurtozy: $K = \frac{m_4}{S^4},$ $m_4 - moment\ centralny\ IV\ rzędu$	



Analiza struktury w zależności od typu szeregu

Średnia arytmetyczna

ŚREDNIA arytmetyczna

$$\bar{\boldsymbol{x}} = \frac{\boldsymbol{x_1+x_2+\cdots+x_n}}{\boldsymbol{n}} = \frac{\boldsymbol{\sum_{i=1}^{n} x_i}}{\boldsymbol{n}}$$ - szereg szczegółowy

$$\bar{\boldsymbol{x}} = \frac{\boldsymbol{x_1n_1+x_2n_2+\cdots+x_kn_k}}{\boldsymbol{n}} = \frac{\boldsymbol{\sum_{i=1}^{k} x_in_i}}{\boldsymbol{n}}$$ - szereg rozdzielczy punktowy

$$\bar{\boldsymbol{x}} = \frac{\boldsymbol{\dot{x}_1n_1+\dot{x}_2n_2+\cdots+\dot{x}_kn_k}}{\boldsymbol{n}} = \frac{\boldsymbol{\sum_{i=1}^{k} \dot{x}_in_i}}{\boldsymbol{n}}$$ - szereg rozdzielczy przedziałowy

gdzie $\dot{x}_i$ jest środkiem przedziału klasowego wyliczanym następująco:

$$\dot{x}_i = \frac{x_{0i}+x_{1i}}{2}$$

Średnia arytmetyczna

PRZYKŁAD

Obliczenia dla średniej

z czasem dojazdu w firmie ALFA przedstawia poniższa tabela.

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

numer klasy	czas dojazdu w ALFA	środek przedziału	liczba pracowników	obliczenia do średniej
i	$x_{0i} - x_{1i}$	$\dot{x}_i$	n_i	$\dot{x}_i n_i$
1	5 – 15	10	10	100
2	15 – 25	20	20	400
3	25 – 35	30	30	900
4	35 – 45	40	50	2000
5	45 – 55	50	80	4000
6	55 – 65	60	10	600
razem	×	×	200	8000

Średnia arytmetyczna

PRZYKŁAD

Obliczenia dla średniej

z czasem dojazdu w firmie ALFA przedstawia poniższa tabela.

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

$$\bar{x} = \frac{8000}{200} = 40$$

numer klasy	czas dojazdu w ALFA	środek przedziału	liczba pracowników	obliczenia do średniej
i	$x_{0i} - x_{1i}$	$\dot{x}_i$	n_i	$\dot{x}_i n_i$
1	5 – 15	10	10	100
2	15 – 25	20	20	400
3	25 – 35	30	30	900
4	35 – 45	40	50	2000
5	45 – 55	50	80	4000
6	55 – 65	60	10	600
razem	×	×	200	8000

Średnia arytmetyczna

Własności średniej arytmetycznej

1. Suma wartości cechy jest równa iloczynowi średniej arytmetycznej i liczebności populacji, tj.

$$n\bar{x} = \sum_{i=1}^{n} x_i \qquad n\bar{x} = \sum_{i=1}^{k} x_i n_i$$

2. Średnia arytmetyczna nie może być mniejsza od najmniejszej wartości cechy ani też większa od największej jej wartości

3. Suma odchyleń poszczególnych wartości cechy od średniej jest równa zero

4. Średnią arytmetyczną oblicza się w zasadzie **dla szeregów o zamkniętych klasach przedziałowych.** Można klasy sztucznie domknąć (i policzyć średnią) tylko wtedy, gdy odsetek jednostek w tych klasach jest niewielki (do 5%). Gdy ten odsetek jest duży należy stosować miary pozycyjne zamiast średniej.

Średnia arytmetyczna

5. Średnia arytmetyczna jest **czuła na skrajne wartości cechy.**

Są to wartości cechy dla jednostek nietypowych w badanej zbiorowości i przypadkowo (niepoprawnie) włączonych do badanej populacji.

Modalna-dominanta

Modalna **(Mo) zwana też** **dominantą** **(D)** jest to wartość cechy, która występuje

najczęściej w badanej zbiorowości.

$$\boldsymbol{M_o} = \boldsymbol{x_{0m}} + i_m \frac{\boldsymbol{n_m} - \boldsymbol{n_{m-1}}}{\boldsymbol{n_m} - \boldsymbol{n_{m-1}} + \boldsymbol{n_m} - \boldsymbol{n_{m+1}}}$$

m - numer klasy (przedziału) z modalną

x_{0m} - dolny kraniec przedziału modalnej

i_m - rozpiętość przedziału modalnej ($i_m = x_{1m} - x_{0m}$)

n_m - liczebność przedziału modalnej

n_{m-1} (n_{m+1}) - liczebność dla przedziałów sąsiadujących z przedziałem modalnej

Modalna-dominanta

PRZYKŁAD Wykorzystamy badanie czasu dojazdu w firmie Alfa

numer klasy	czas dojazdu w Alfa	liczba pracow-ników
i	$x_{0i} - x_{1i}$	n_i
1	5 – 15	10
2	15 – 25	20
3	25 – 35	30
4	35 – 45	50
5	**45** – 55	80
6	55 – 65	**10**
razem	×	200

Najpierw ustalamy przedział (klasę), gdzie jest dominanta

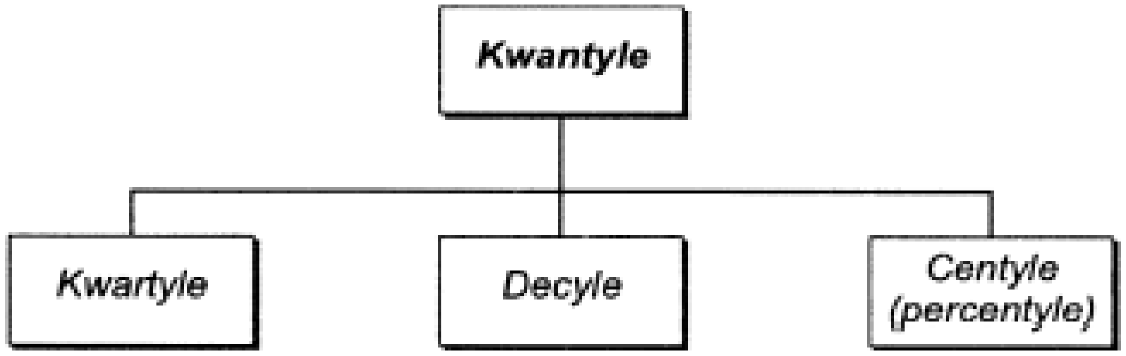
$$\boldsymbol{M_o} = \boldsymbol{x_{0m}} + i_m \frac{\boldsymbol{n_m} - \boldsymbol{n_{m-1}}}{\boldsymbol{n_m} - \boldsymbol{n_{m-1}} + \boldsymbol{n_m} - \boldsymbol{n_{m+1}}}$$

$$\boldsymbol{M_o} = \mathbf{45} + 10 \times \frac{80 - \mathbf{50}}{80 - \mathbf{50} + 80 - 10}$$

$$\boldsymbol{M_o} = 45 + 10 \times 30/100 = 45 + 3 = \mathbf{48}$$

WNIOSEK: najczęściej występującym czasem dojazdu wśród pracowników firmy Alfa jest 48 minut.

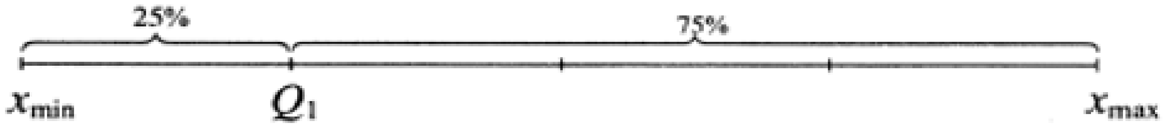
Kwantyle



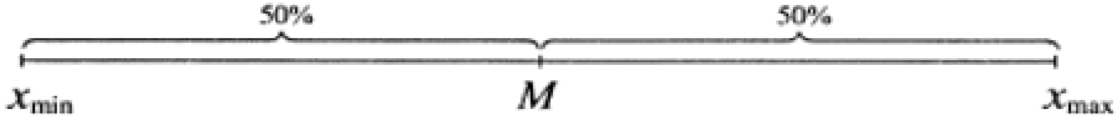
Kwartyle (Q) dzielą populację z dokładnością do ćwiartek, decyle (D) dzielą populację z dokładnością do dziesiątek, centyle (C) dzielą populację z dokładnością do pojedynczych procentów.

Kwantyle

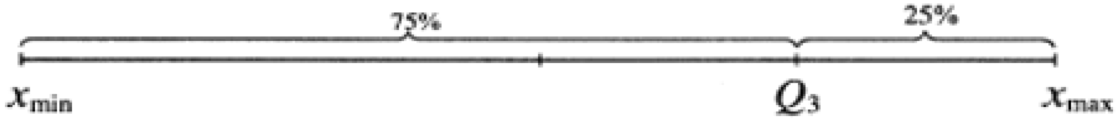
Pierwszy kwartyl (Q_1) odcina 25% elementów populacji o najmniejszych wartościach od pozostałych 75% jednostek:



Mediana (kwartyl drugi) symbolizowana jako M lub Q_2 dzieli populację na połowy:



Trzeci kwartyl (Q_3) oddziela 75% populacji od 25% jednostek o największych wartościach:



Kwantyle

Mediana (M_e) - wartość środkowa, inaczej: kwartyl 2 (Q_{II}).

Jest to taka wartość cechy X, która dzieli zbiorowość na dwie równe części, tj. połowa zbiorowości charakteryzuje się wartością cechy X mniejszą lub równą medianie, a druga połowa większą lub równą.

Mediana dla szeregu rozdzielczego

1. Ustalamy na początek tzw. **numer mediany** N_{Me}. Jest to połowa liczebności populacji: $N_{Me} = \frac{1}{2}n$

2. Kumulujemy liczebności (albo częstości).

3. Znajdujemy klasę, w której po raz pierwszy przekroczony został numer mediany.

4. Następnie wstawiamy do wzoru.

Kwartyle

Mediana dla szeregu przedziałowego

numer klasy	czas dojazdu w Alfa	liczba pracow-ników	skumul. liczebność
i	$x_{0i} - x_{1i}$	n_i	$n_{i\,sk}$
1	5 – 15	10	10
2	15 – 25	20	30
3	25 – 35	30	60
4	35 – 45	50	110
5	45 – 55	80	190
6	55 – 65	10	200
razem	×	200	×

$$M_e = x_{0Me} + i_{Me}\,\frac{N_{Me} - n_{ME-1\ sk}}{n_{Me}}$$

$$M_e = 35 + 10 \times \frac{100 - 60}{50} = 35 + 10 \times \frac{40}{50} = 35 + 8 = 43$$

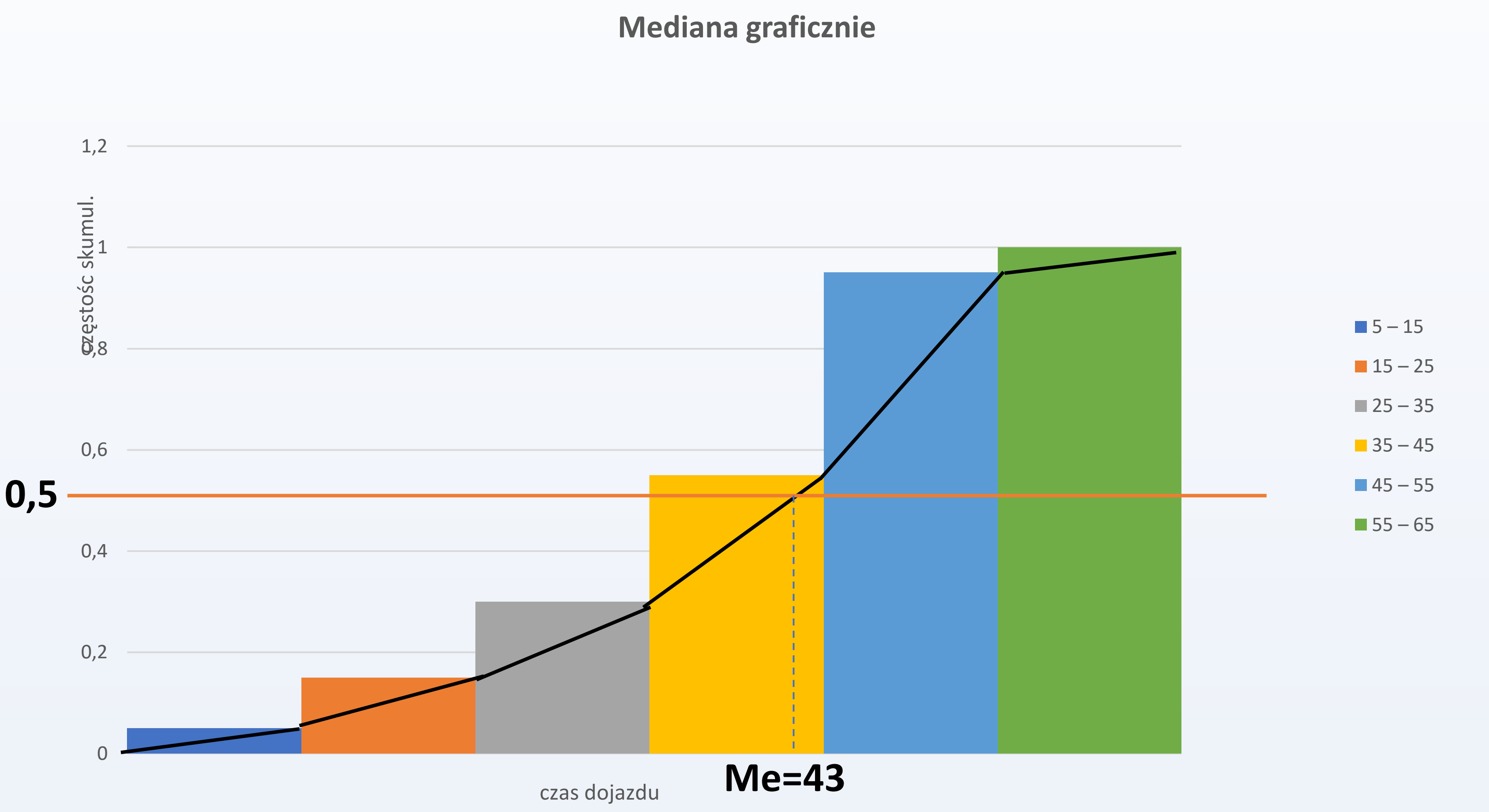
WNIOSEK: Połowa pracowników firmy Alfa dojeżdża do pracy w czasie nie dłuższym (≤) niż 43 minuty, a druga połowa w czasie nie krótszym (≥) niż 43 minuty

50

Kwantyle

Kwartyle graficznie

numer klasy	czas dojazdu w ZAUR	Częstość skumulowana
i	$x_{0i} - x_{1i}$	$\omega_{i\ skum}$
1	5 – 15	0,05
2	15 – 25	0,15
3	25 – 35	0,3
4	35 – 45	0,55
5	45 – 55	0,95
6	55 – 65	1,00
razem	×	



Kwantyle

Kwartyle dla szeregu przedziałowego

$$Q_1 = x_{Q_1} + i_{Q_1} \frac{\frac{N}{4} - n_{Q_1 - 1\ sk}}{n_{Q_1}}$$

$$Me = Q_2 = x_{Me} + i_{Me} \frac{\frac{N}{2} - n_{Me - 1\ sk}}{n_{Me}}$$

$$Q_3 = x_{Q_3} + i_{Q_3} \frac{\frac{3N}{4} - n_{Q_3 - 1\ sk}}{n_{Q_3}}$$

Kwantyle

PRZYKŁAD

numer klasy	czas dojazdu w ZAUR	liczba pracowników	skumul. liczebność
i	$x_{0i} - x_{1i}$	n_i	$n_{i\,sk}$
1	5 – 15	10	10
2	15 – 25	20	30
3	25 – 35	30	60
4	35 – 45	50	110
5	45 – 55	80	190
6	55 – 65	10	200
razem	×	200	×

$$\boldsymbol{Q_1 = x_{Q_1} + i_{Q_1}\frac{\frac{N}{4} - n_{Q_1 - 1\ sk}}{n_{Q_1}}}$$

Numer kwartyla : $N_{Q_1} = \frac{1}{4} \times 200 = 50$

$- \boldsymbol{kolumna}$ **$n_{i\,sk}$ aż przekroczymy po raz pierwszy 50**

Kwantyle

PRZYKŁAD

numer klasy	czas dojazdu w ZAUR	liczba pracowników	skumul. liczebność
i	$x_{0i} - x_{1i}$	n_i	$n_{i\,sk}$
1	5 – 15	10	10
2	15 – 25	20	30
3	25 – 35	30	60
4	35 – 45	50	110
5	45 – 55	80	190
6	55 – 65	10	200
razem	×	200	×

$$Q_1 = x_{Q_1} + i_{Q_1}\frac{\frac{N}{4} - n_{Q_1 - 1\ sk}}{n_{Q_1}}$$

$$Q_1 = 25 + 10 \times \frac{50 - 30}{30} = 25 + 10 \times \frac{20}{30} = 25 + 6{,}7 = 31{,}7$$

WNIOSEK: 25% firmy Alfa dojeżdża do pracy w czasie nie dłuższym (≤) niż 31,7 minut, a 75% a w czasie nie krótszym (≥) niż 31,7 minut.

Kwantyle

numer klasy	czas dojazdu W Alfa	liczba pracow-ników	skumul. liczebność
i	$x_{0i} - x_{1i}$	n_i	$n_{i\,sk}$
1	5 – 15	10	10
2	15 – 25	20	30
3	25 – 35	30	60
4	35 – 45	50	110
5	45 – 55	80	190
6	55 – 65	10	200
razem	×	200	×

$$\boldsymbol{Q_3 = x_{Q_3} + i_{Q_3} \frac{\frac{3N}{4} - n_{Q_3 - 1\ sk}}{n_{Q_3}}}$$

Numer kwartyla: $N_{Q_3} = \frac{3}{4} \times 200 = 150$

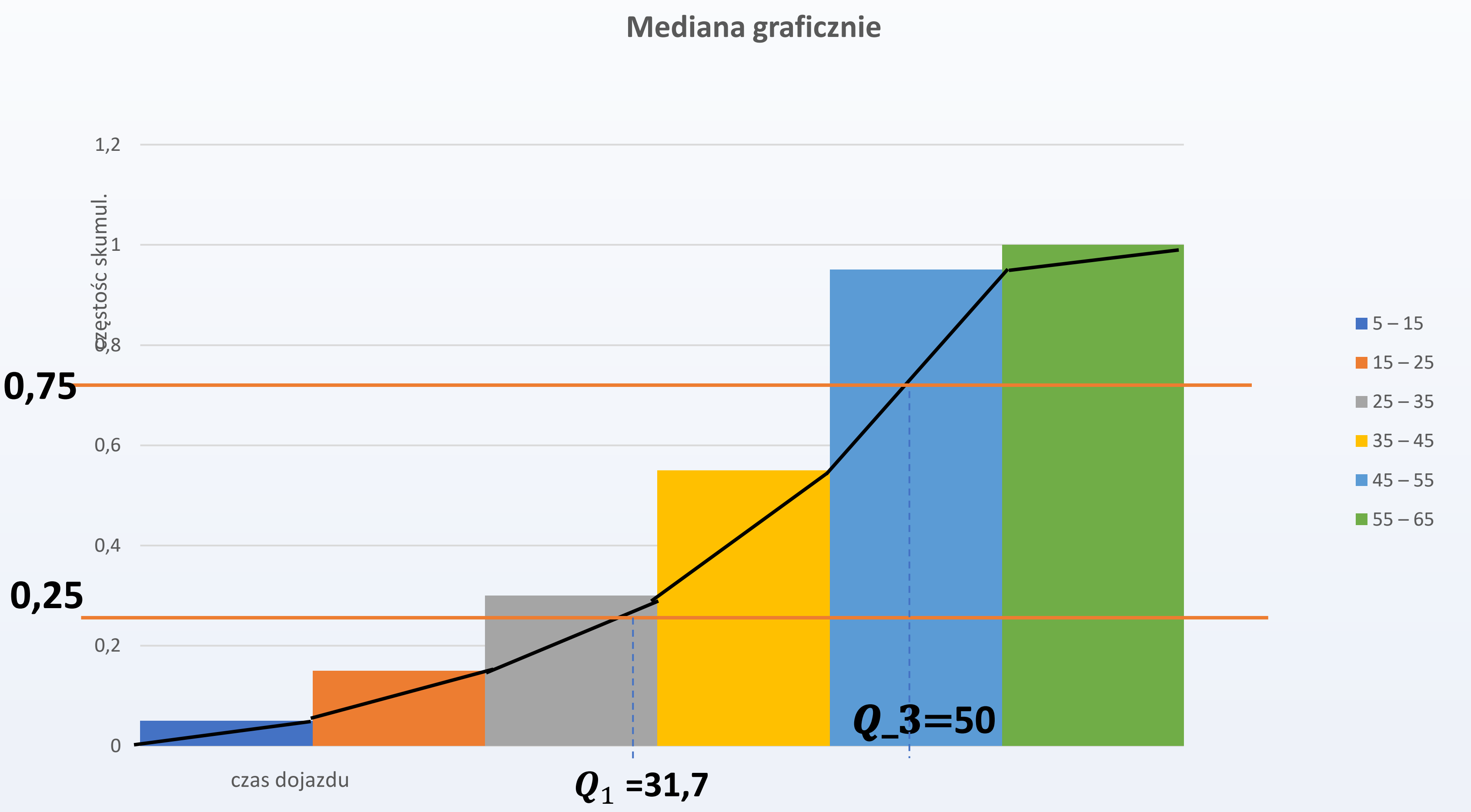
$$\boldsymbol{Q_3} = 45 + 10 \times \frac{150 - 110}{80} = $$
$$= 45 + 10 \times \frac{40}{80} = $$
$$= 45 + 5 = 50$$

WNIOSEK: 75% firmy Alfa dojeżdża do pracy w czasie nie dłuższym (≤) niż 50 minut, a 25% a w czasie nie krótszym (≥) niż 50 minut.

Kwantyle

Kwartyle graficznie

numer klasy	czas dojazdu w ZAUR	Częstość skumulowana
i	$x_{0i} - x_{1i}$	$\omega_{i\ skum}$
1	5 – 15	0,05
2	15 – 25	0,15
3	25 – 35	0,3
4	35 – 45	0,55
5	45 – 55	0,95
6	55 – 65	1,00
razem	×	



2. 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)**

2. miary pozycyjne (rozstęp, odchylenie ćwiartkowe, współczynnik zmienności).

2. Miary zmienności – klasyczne

Wariancja, odchylenie standardowe, współczynnik zmienności (klasyczny)

Wariancję definiuje się jako średnią arytmetyczną kwadratów odchyleń 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.

Współczynnik zmienności (klasyczny) (Vs) jest to iloraz odchylenia standardowego (lub przeciętnego) przez średnia arytmetyczną. Jest to wielkość niemianowana.

2. Miary zmienności –klasyczne

Wariancja

$$s^2 = \frac{(x_1-\bar{x})^2+\dots+(x_n-\bar{x})^2}{n} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$$ - dla szeregu szczegółowego

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

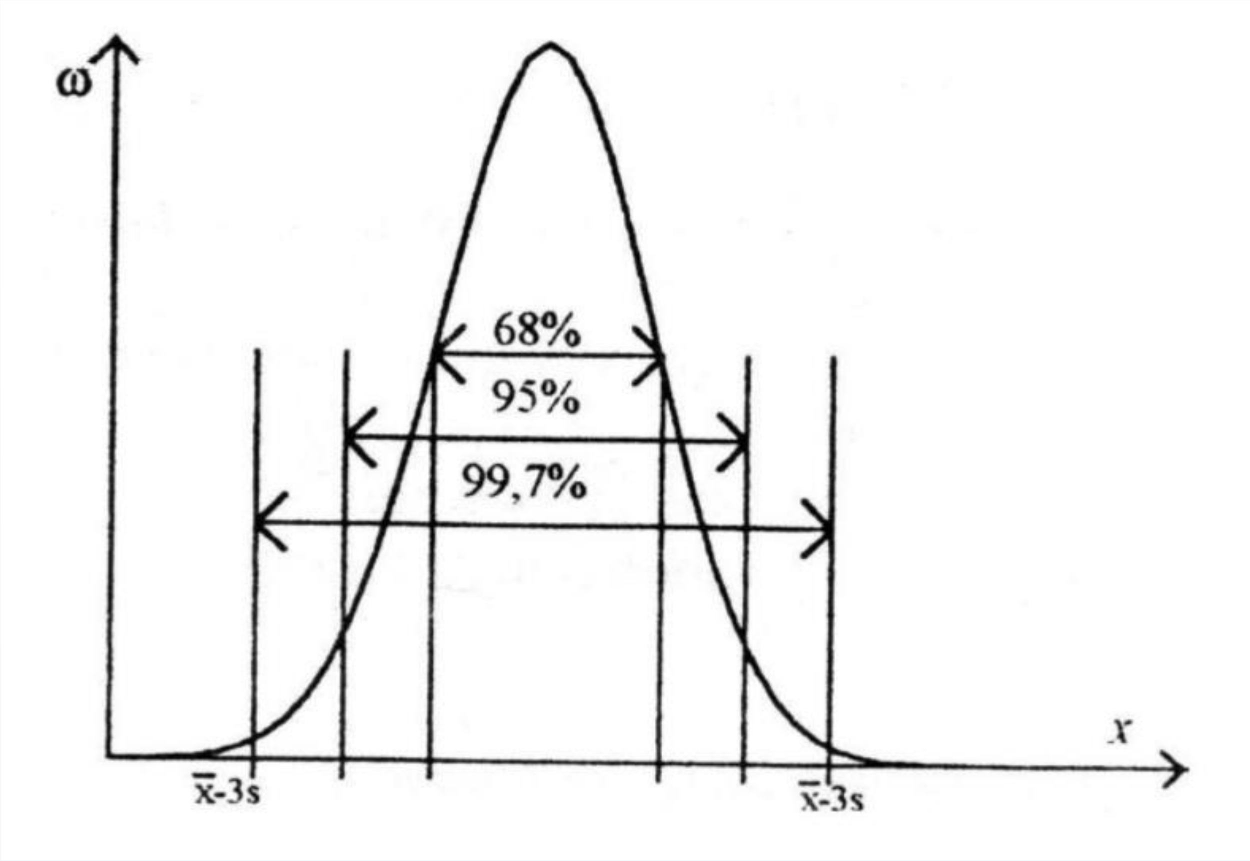
$$s^2 = \frac{(\dot{x}_1-\bar{x})^2 n_1+\dots+(\dot{x}_k-\bar{x})^2 n_i}{n} = \frac{\sum_{i=1}^{k}(\dot{x}_i-\bar{x})^2 n_i}{n}$$ - dla szeregu rozdzielczego

Odchylenie standardowe $s = \sqrt{s^2}$

Współczynnik zmienności (klasyczny) $V_s = \frac{s}{\bar{x}}$

2. Miary zmienności -klasyczne

Reguła „3 sigma”



numer klasy	czas dojazdu	środek klasy	liczba pracow.	obliczenia dla wariancji		
i	$x_{0i} - x_{1i}$	$\dot{x}_i$	n_i	$\dot{x}_i - \bar{x}$	$(\dot{x}_i - \bar{x})^2$	$(\dot{x}_i - \bar{x})^2 n_i$
1	5 – 15	10	10	-30	900	9000
2	15 – 25	20	20	-20	400	8000
3	25 – 35	30	30	-10	100	3000
4	35 – 45	40	50	0	0	0
5	45 – 55	50	80	10	100	8000
6	55 – 65	60	10	20	400	4000
razem	×	×	200	×	×	**32000**

$$\bar{x} = 40$$

$$s^2 = \frac{\sum_{i=1}^{k} (\dot{x}_i - \bar{x})^2 n_i}{n}$$

numer klasy	czas dojazdu	środek klasy	liczba pracow.	obliczenia dla wariancji		
i	$x_{0i} - x_{1i}$	$\dot{x}_i$	n_i	$\dot{x}_i - \bar{x}$	$(\dot{x}_i - \bar{x})^2$	$(\dot{x}_i - \bar{x})^2 n_i$
1	5 – 15	10	10	-30	900	9000
2	15 – 25	20	20	-20	400	8000
3	25 – 35	30	30	-10	100	3000
4	35 – 45	40	50	0	0	0
5	45 – 55	50	80	10	100	8000
6	55 – 65	60	10	20	400	4000
razem	×	×	200	×	×	**32000**

$$s^2 = \frac{\sum_{i=1}^{k}(\dot{x}_i - \bar{x})^2 n_i}{n}$$

$$\bar{x} = 40$$

$$s^2 = \frac{32000}{200} = 160$$

$$s = \sqrt{160} \approx 12{,}7$$

Współczynnik zmienności (klasyczny)

$$V_s = \frac{12{,}7}{40} = 0{,}32$$

WNIOSEK: Średnio o 12,7 min różni się czas dojazdu każdego pracownika od średniej arytmetycznej 40 min.

2. Miary zmienności – miary pozycyjne

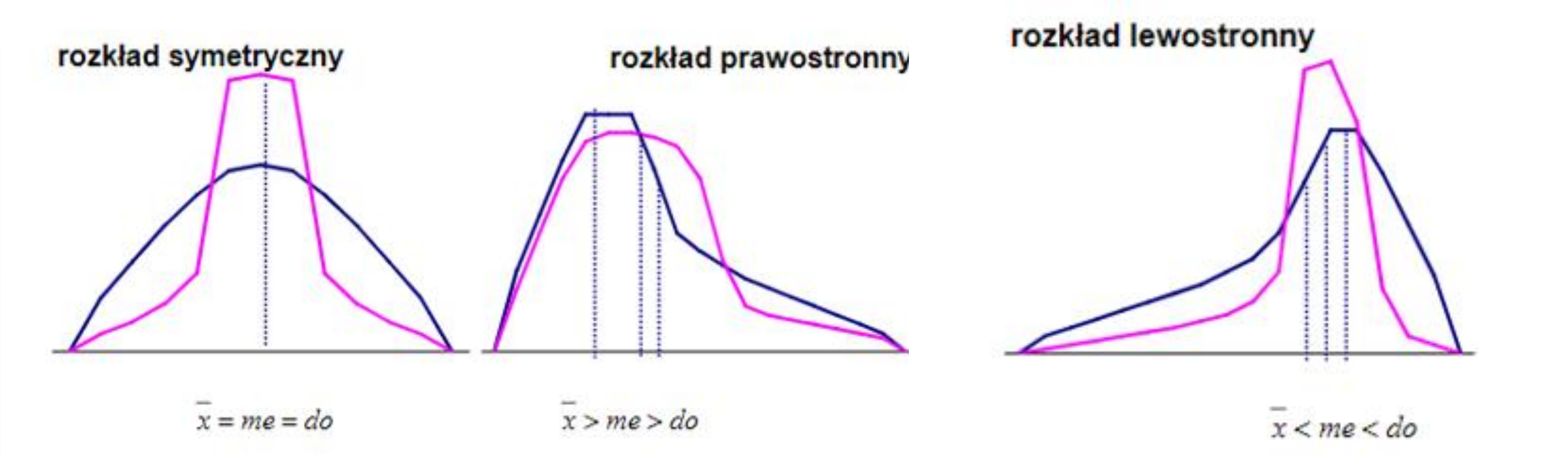
Rozstęp (*R*) definiujemy go jako różnicę pomiędzy największą i najmniejszą wartością cechy: $R = x_{max} - x_{min}$

Odchylenie ćwiatkowe (*Q*) jest miarą rozproszenia wartości cechy od mediany. Definiuje się go jako połowę różnicy pomiędzy trzecim i pierwszym kwartylem: $Q = \frac{Q_3 - Q_1}{2}$

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

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

3. Miary asymetrii



3. Miary asymetrii

Miary asymetrii charakteryzują rodzaj i stopień odstępstwa od symetrii rozkładu badanej cechy.

Miary asymetrii dzielą się na miary klasyczne i pozycyjne.

1. miary klasyczne – współczynnik skośności klasyczny A_s

2. miary pozycyjne – współczynnik skośności pozycyjny A_Q

3. Miary asymetrii

Współczynnik asymetrii $A_s = \frac{m_3}{s^3}$,

gdzie m_3**- moment centralny 3 -go rzędu**

$$m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3$$

- **szereg szczegółowy**

$$m_3 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^3 n_i$$

- **szereg rozdzielczy punktowy**

$$m_3 = \frac{1}{n}\sum_{i=1}^{k}(\dot{x}_i - \bar{x})^3 n_i$$

- **szereg rozdzielczy przedziałowy**

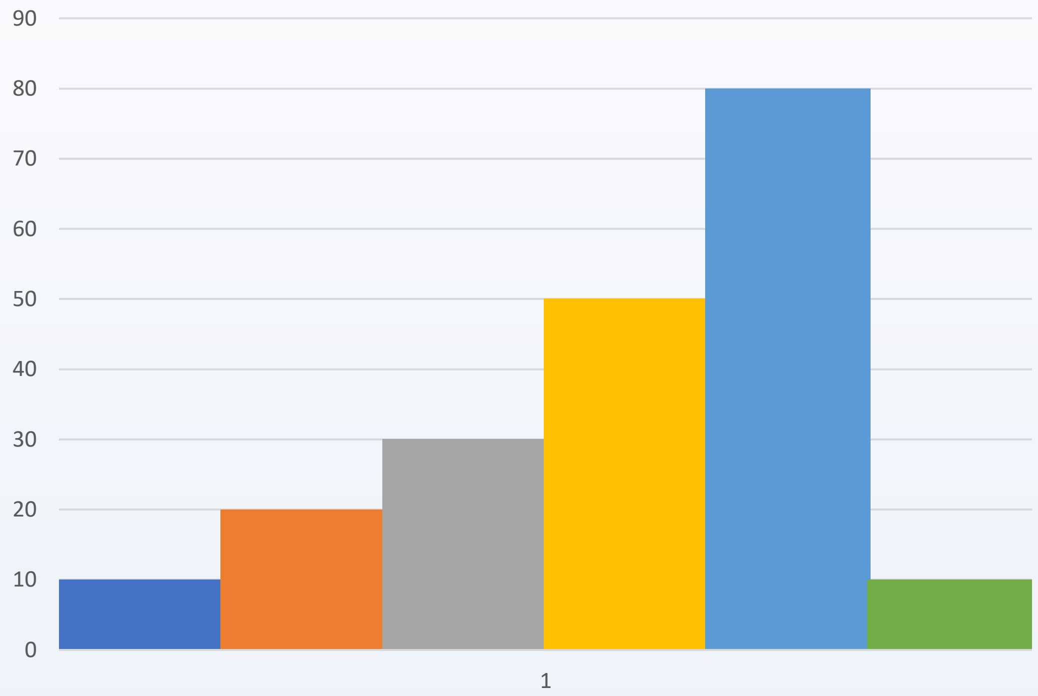
3. Miary asymetrii

numer klasy	czas dojazdu	środek klasy	liczba pracow.	obliczenia dla wariancji			obliczenia dla skośności	
1	5 – 15	10	10	-30	900	9000	-27000	-270000
2	15 – 25	20	20	-20	400	8000	-8000	-160000
3	25 – 35	30	30	-10	100	3000	-1000	-30000
4	35 – 45	40	50	0	0	0	0	0
5	45 – 55	50	80	10	100	8000	1000	80000
6	55 – 65	60	10	20	400	4000	8000	80000
razem			200			32000		-300000

$$m_3 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^3 n_i}{n} = \frac{-300\,000}{200} = -1500$$

$$A_s = \frac{-1500}{12{,}7^3} = -0{,}76$$ - asymetria lewostronna

3. Miary asymetrii



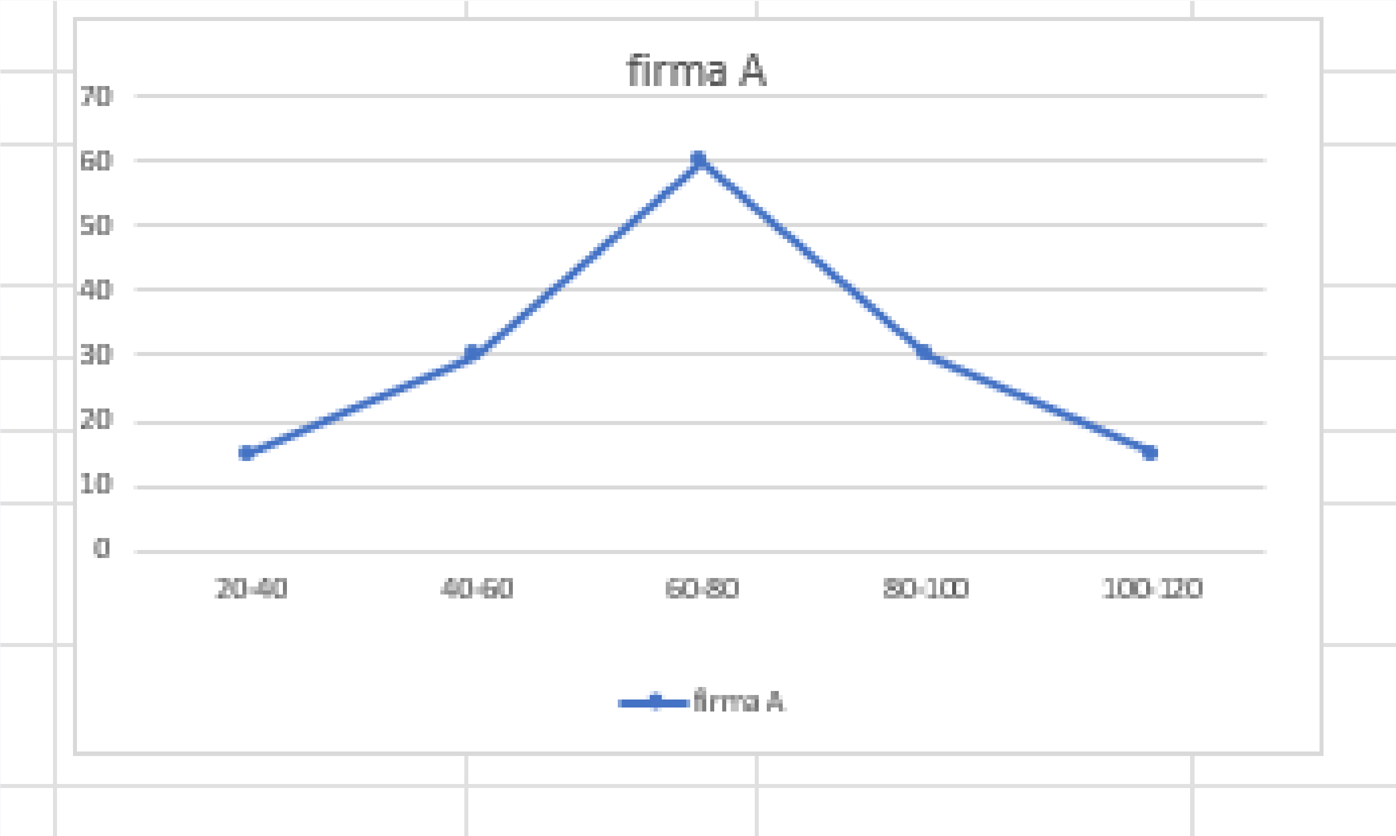
3. Miary asymetrii

Przedział	firma A	śr przedziału xi	xi razy ni	xi-xśr	(xi-xśr) ^2	(xi-x śr)^3	(xi-x śr)^2 * ni	(xi-x śr)^3 * ni
20-40	15	30	450	-40	1600	-64000	24000	-960000
40-60	30	50	1500	-20	400	-8000	12000	-240000
60-80	60	70	4200	0	0	0	0	0
80-100	30	90	2700	20	400	8000	12000	240000
100-120	15	110	1650	40	1600	64000	24000	960000
	150		**10500**				**72000**	**0**

$$s^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{n} = \frac{72000}{150} = 480, \qquad s = \sqrt{480} = 21{,}9$$

$$m_3 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^3 n_i}{n} = \frac{0}{150} = 0, \qquad A_s = \frac{0}{21{,}9^3} = 0$$

3. Miary asymetrii



3. Miary spłaszczenia

Współczynnik kurtozy $K = \frac{m_4}{s^4}$,

gdzie m_4**- moment centralny 4 -go rzędu**

Licznik powyższego ułamka (m_4**) wyliczamy odmiennie dla każdego sposobu pogrupowania materiału statystycznego. I tak:**

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$$

- szereg szczegółowy

$$m_4 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^4 n_i$$

- szereg rozdzielczy punktowy

$$m_4 = \frac{1}{n}\sum_{i=1}^{k}(\dot{x}_i - \bar{x})^4 n_i$$

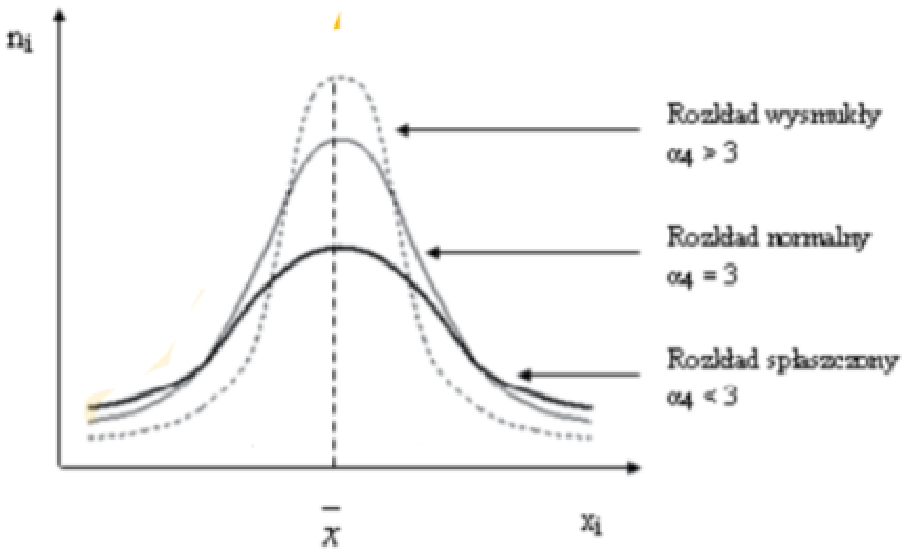
- szereg rozdzielczy przedziałowy

3. Miary spłaszczenia

Kurtoza w rozkładzie normalnym jest zawsze równa trzy (*K*=3).

W praktyce policzoną kurtozę porównujemy z kurtozą rozkładu normalnego. I tak jeżeli:

- K>3 - rozkład badanej cechy jest wyższy i smuklejszy od rozkładu normalnego
- K<3 - odwrotnie; niższy i bardziej rozłożysty



4. Miary spłaszczenia

numer klasy	czas dojazdu	środe klasy	liczba praco	obliczenia dla kurtozy	
i	$x_{0i} - x_{1i}$	$\dot{x}_i$	n_i	$(\dot{x}_i - \bar{x})^4$	$(\dot{x}_i - \bar{x})^4 n_i$
1	5 – 15	10	10	810000	8100000
2	15 – 25	20	20	160000	3200000
3	25 – 35	30	30	10000	300000
4	35 – 45	40	50	0	0
5	45 – 55	50	80	10000	800000
6	55 – 65	60	10	160000	1600000
razem			200		14000000

$$m_4 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^4 n_i}{n} = \frac{14000\,000}{200} = 70000$$

$$K = \frac{70\,000}{12{,}7^4} = 2{,}65$$

wykres spłaszczony

4. Miary spłaszczenia

Przedział	firma A	śr przedziału xi	(xi-x śr)^4	(xi-x śr)^4 * ni
20-40	15	30	2 560 000	38400000
40-60	30	50	16 000	480000
60-80	60	70	0	0
80-100	30	90	16 000	480000
100-120	15	110	2 560 000	38400000
	150		5 152 000	77760000

$$m_4 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^4 n_i}{n} = \frac{77760\,000}{150} = 518\,400$$

$$K = \frac{518400}{21{,}9^4} = 2{,}25$$

wykres spłaszczony