

# Statystyka opisowa

Robert Pietrzykowski

email: [robert\\_pietrzykowski@sggw.pl](mailto:robert_pietrzykowski@sggw.pl)

[www.ekonometria.info](http://www.ekonometria.info)

**Jeśli mój sąsiad codziennie bije swoją żonę, ja zaś nie biję jej nigdy, to w świetle statystyki obaj bijemy je co drugi dzień.**

George Bernard Shaw

**Śmierć jednostki to tragedia - milion zabitych to tylko statystyka.**

Józef Stalin

---

*Statystyka jest bardziej sposobem myślenia lub wnioskowania niż pęczkiem recept na młócenie danych w celu odślonięcia odpowiedzi*

C. R. Rao

---

# Literatura

- Wasilewska E. (2009): Statystyka opisowa od podstaw, Wydawnictwo SGGW
- Piłatowska M. (2006): Repetytorium ze statystyki, PWN
- Aczel A. D. (2006): Statystyka w zarządzaniu, PWN
- Sobczyk M. (2005): Statystyka, PWN
- Parlińska M., Parliński J. (2011): Statystyczna analiza danych z Excelem, Wydawnictwo SGGW
- Kisielińska J., Skórnik-Pokorowska U. (2005): Podstawy statystyki z przykładami w Excelu, Wydawnictwo SGGW
  
- materiały wykładowe
- [www.ekonometria.info](http://www.ekonometria.info), [www.statystyka.info](http://www.statystyka.info)
- [www.ibuk.pl](http://www.ibuk.pl)

# ZADANIE 1

- Dane pochodzą z GUS. Obserwujemy liczbę bezrobotnych mężczyzn i kobiet w okresie od 2005 do 2010 roku. Zadanie 1 to porównanie liczby bezrobotnych ze względu na płeć. Zadanie 2 to określenie dynamiki zmian liczby bezrobotnych w latach.
- Zbiorowość statystyczna podlegając badaniu to bezrobotni w Polsce. Cechy badane to: liczba bezrobotnych mężczyzn i liczba bezrobotnych kobiet. Obserwowane cechy są ilościowe, skokowe.
- Narzędzia statystyczne: miary zróżnicowania (odchylenie standardowe), położenia (średnia i mediana), średnia geometryczna.

Tabela 1. Porównanie liczby bezrobotnych mężczyzn i kobiet na podstawie wybranych wskaźników w I kwartale roku 2006

Miary	Mężczyźni	Kobiety
średnia	92993,3	97791,4
Me	87883	96710
S	36797,3	39307,4

W roku 2006 średnio liczba bezrobotnych mężczyzn była mniejsza niż bezrobotnych kobiet. Typowy obszar zmienności dla mężczyzn wynosił (56 196; 129 791), a dla kobiet (58 484; 137 099).

# Zmiany liczby bezrobotnych w latach 2005 - 2010 na koniec I kwartału

Dla mężczyzn i kobiet obliczono tempo zmian w badanym okresie (2005 - 2010). Średnia geometryczna dla mężczyzn wyniosła 0,9345, a dla kobiet 0,9172. Oznacza to, że z roku na rok w pierwszym kwartale malała liczba bezrobotnych kobiet i mężczyzn odpowiednio o 8,3% i 6,6%.



# Anex

- Wzory wykorzystanych miar statystycznych
- Wykorzystane pozycje literatury
  - (Piłatowska M. (2006): Repetytorium ze statystyki, PWN, str. 12 def. typowy obszar zmienności)
  - (Kobus P. Statystyka opisowa wykład 2. Miary położenia i zmienności, <http://www.statystyka.info> dn. 29.09.2013, str. 1 wzór średniej)

## ZADANIE 2

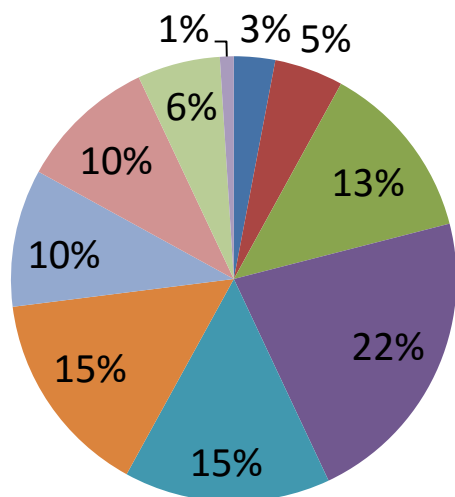
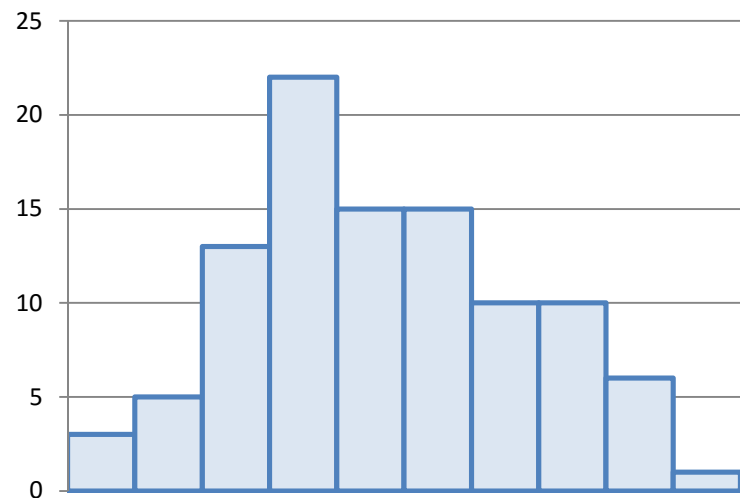
- Dane zebrano od studentów WNE z II roku Zarządzania. Obserwujemy wzrost mężczyzn i kobiet z WNE II roku Zarządzania niestacjonarnych. Zadanie to graficzna prezentacja badanego zjawiska.
- Zbiorowość statystyczna podlegając badaniu to studenci z WNE II roku Zarządzania niestacjonarnego. Cechy badane to: wzrost studentów [w cm]. Obserwowane cecha jest ilościowa, ciągła.
- Dane zestawiono w szeregu rozdzielczym.
- Narzędzia statystyczne: histogram, wykres kołowy, graficzna, krzywa częstości, graficzna prezentacja typowego obszaru zmienności.



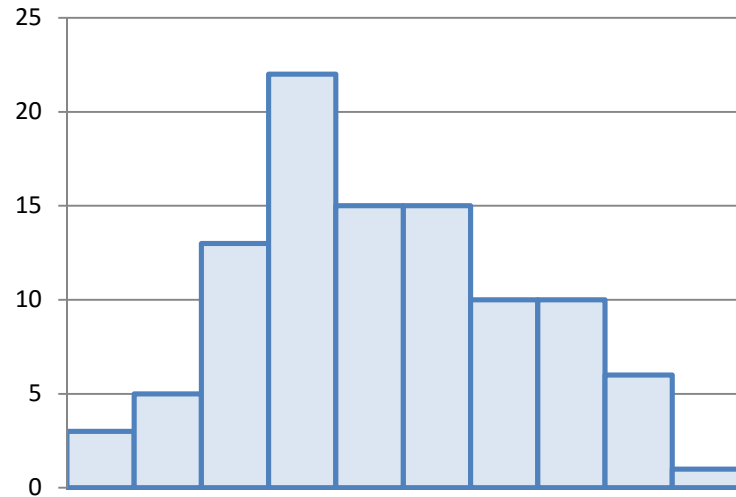
# Szereg szczegółowy

- 175 167 163 163 150 166 180 183 186 169  
190 177 170 180 170 167 191 181 171 165  
175 165 166 169 162 181 167 169 173 190  
165 191 157 164 162 159 180 173 171 184  
173 175 158 166 164 173 165 171 167 163  
165 196 172 162 157 164 168 186 171 158  
183 154 191 166 186 184 172 160 177 169  
183 160 152 175 177 162 170 185 161 182  
165 185 168 172 160 167 179 167 175 165  
190 173 179 178 177 162 170 180 162 176

# Szereg rozdzielczy



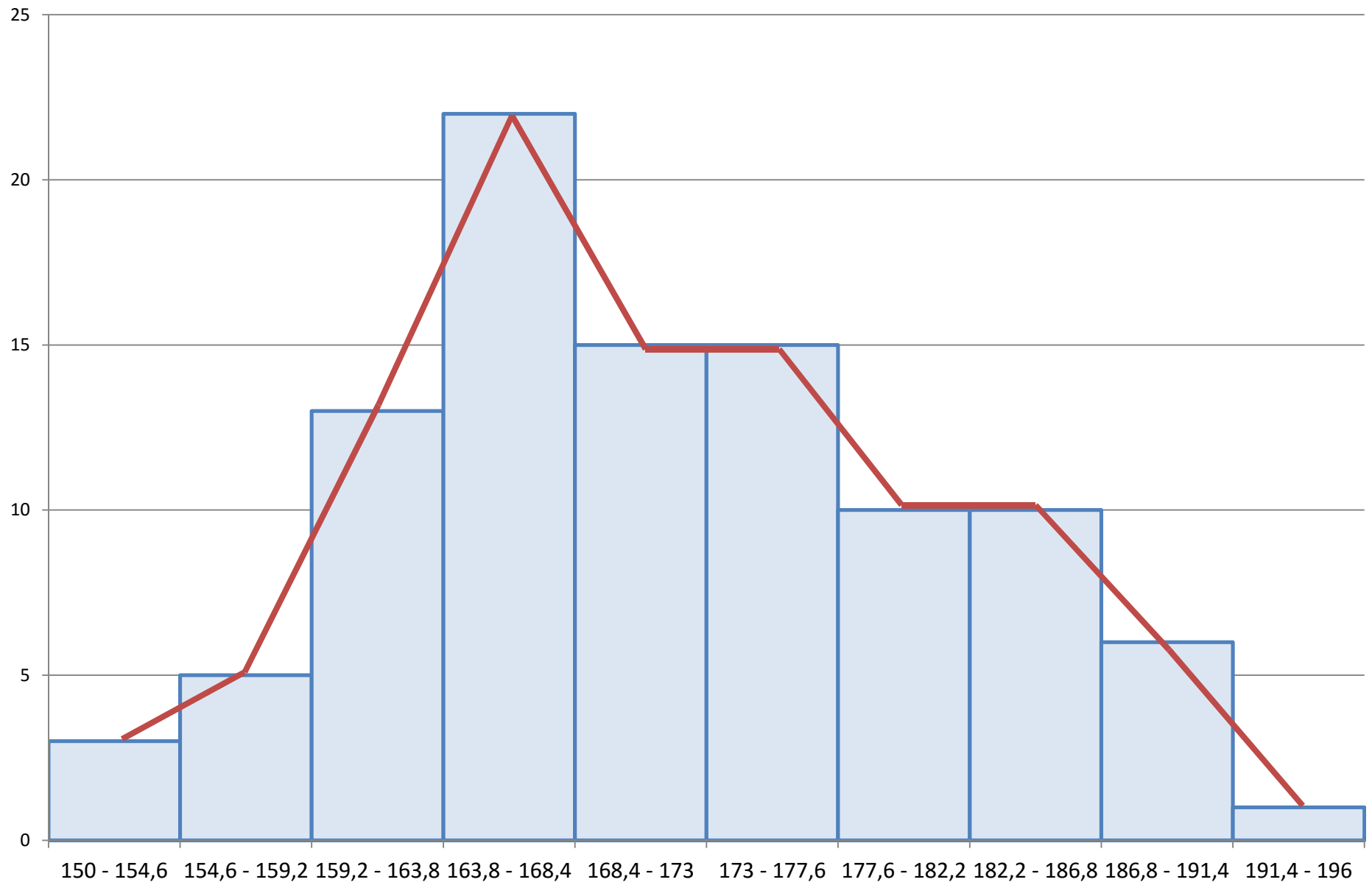
Wzrost studenta (przedział)	$X_d$	$X_g$	$n_i$	$w_i$
150 - 154,6	150	154,6	3	0,03
154,6 - 159,2	154,6	159,2	5	0,05
159,2 - 163,8	159,2	163,8	13	0,13
163,8 - 168,4	163,8	168,4	22	0,22
168,4 - 173	168,4	173	15	0,15
173 - 177,6	173	177,6	15	0,15
177,6 - 182,2	177,6	182,2	10	0,1
182,2 - 186,8	182,2	186,8	10	0,1
186,8 - 191,4	186,8	191,4	6	0,06
191,4 - 196	191,4	196	1	0,01
Razem			100	1



Rozkład empiryczny badanego zjawiska przedstawiono na histogramie. Najliczniejsza grupa 22 studentów to studenci o wzroście od 163,8 do 168,4. Stanowi to 22% całej badanej zbiorowości. Najmniej liczne grupy studentów to najniżsi studenci (3 osoby 150 – 154,6cm) i najwyżsi (1 student 191,4 do 196cm).

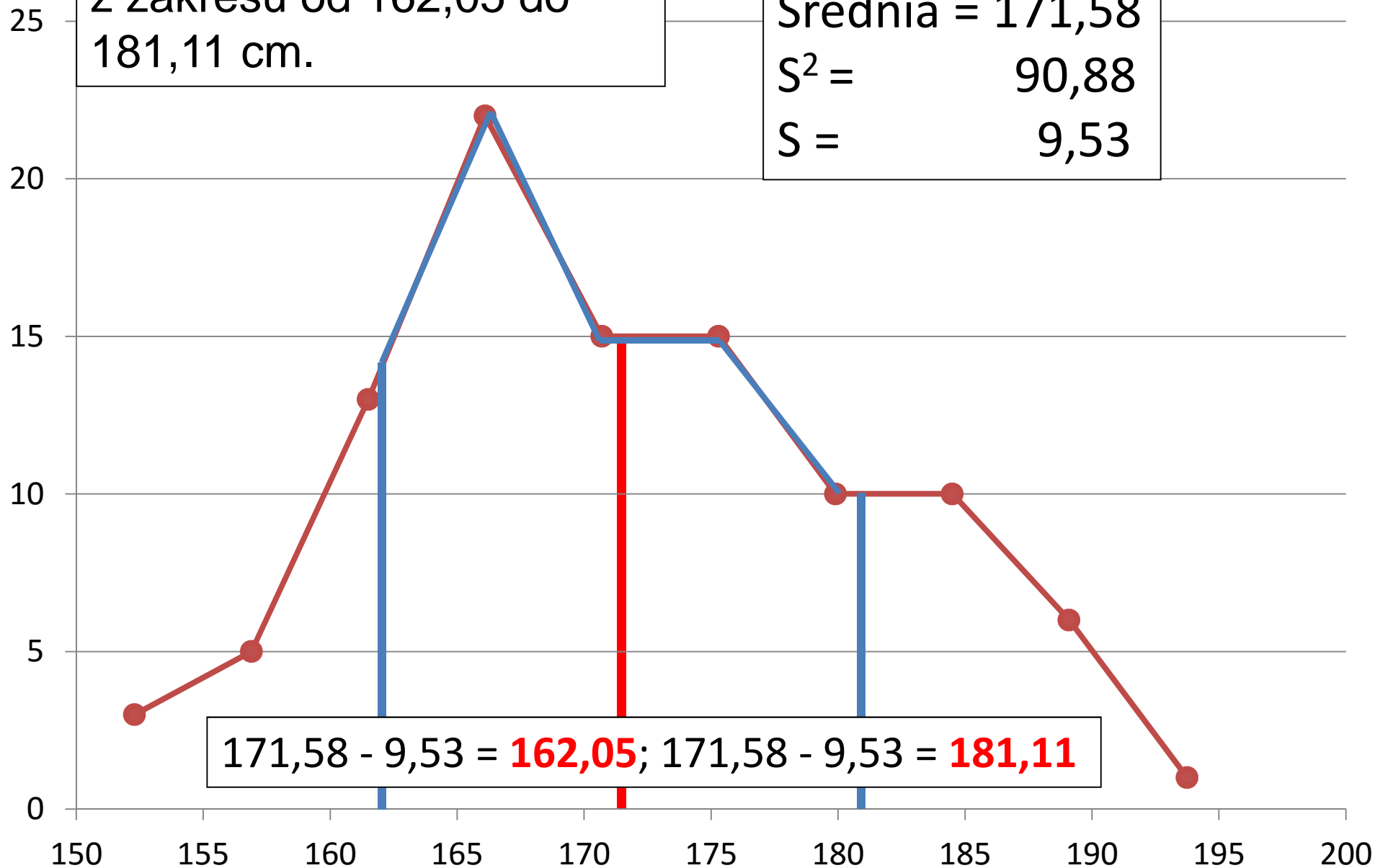
W oparciu o uzyskany wykres można mówić o prawostronnej asymetrii.

Wzrost studenta (przedział)	$X_d$	$X_g$	$n_i$	$w_i$
150 - 154,6	150	154,6	3	0,03
154,6 - 159,2	154,6	159,2	5	0,05
159,2 - 163,8	159,2	163,8	13	0,13
163,8 - 168,4	163,8	168,4	22	0,22
168,4 - 173	168,4	173	15	0,15
173 - 177,6	173	177,6	15	0,15
177,6 - 182,2	177,6	182,2	10	0,1
182,2 - 186,8	182,2	186,8	10	0,1
186,8 - 191,4	186,8	191,4	6	0,06
191,4 - 196	191,4	196	1	0,01
Razem			100	1



Wzrost 68% studentów jest z zakresu od 162,05 do 181,11 cm.

Średnia = 171,58  
 $S^2 = 90,88$   
 $S = 9,53$



$171,58 - 9,53 = 162,05$ ;  $171,58 + 9,53 = 181,11$

# **Statystyka** (*łc. status, stan rzeczy, państwo*)

— nauka zajmująca się ilościowymi technikami badania prawidłowości procesów masowych.

— wykazy informacji liczbowych.

*Słownik Wyrazów Obcych Europa, Warszawa 2001*

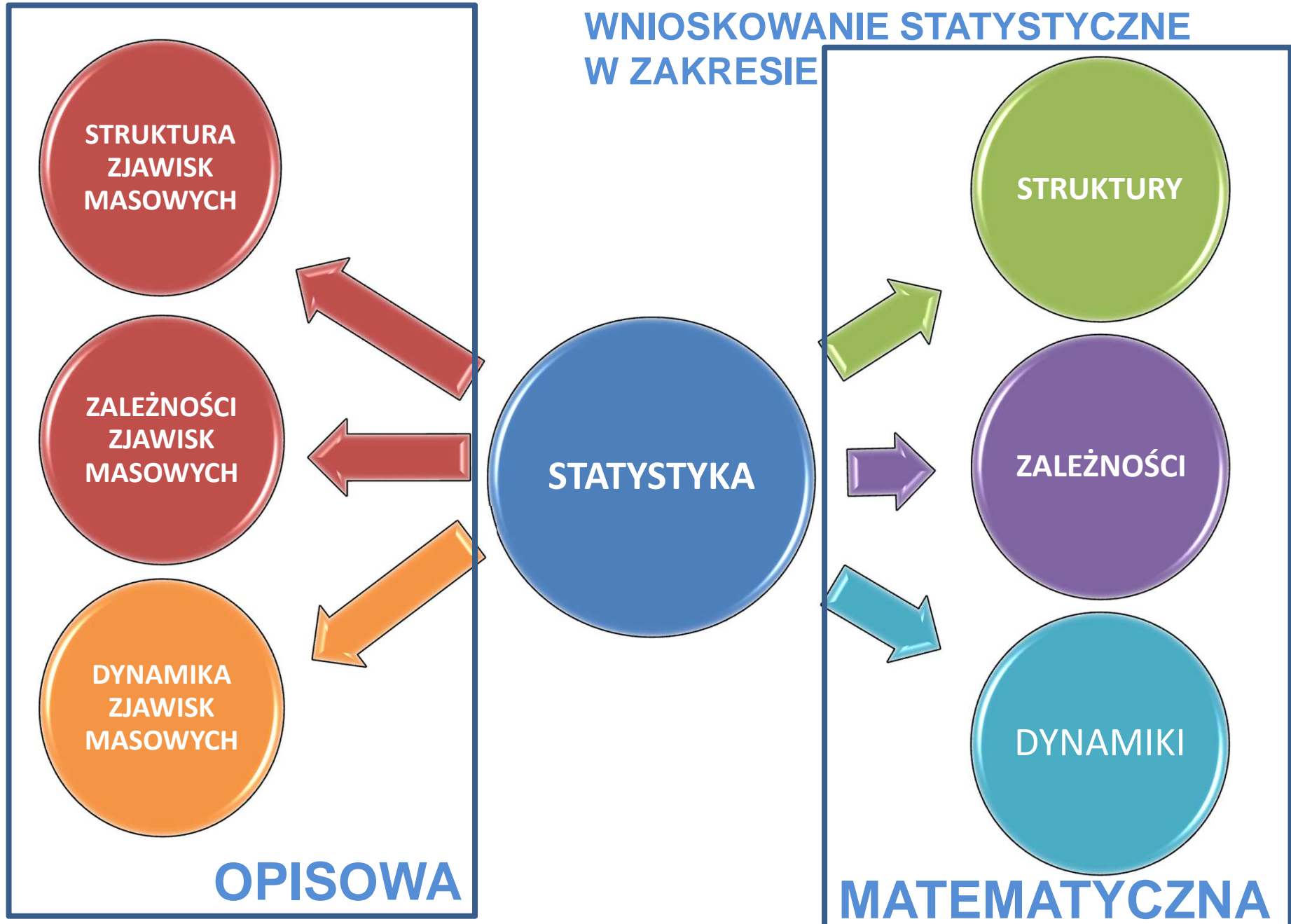
**STATYSTYKA MATEMATYCZNA:** dział matematyki stosowanej oparty na rachunku prawdopodobieństwa; zajmuje się badaniem zbiorów na podstawie znajomości własności ich części.

*Encyklopedia Popularna PWN, Warszawa 1982*

# STATYSTYKA OPISOWA

- **Statystyka opisowa** to dział statystyki zajmujący się metodami opisu danych statystycznych uzyskanych podczas badania statystycznego. Celem stosowania metod statystyki opisowej jest podsumowanie zbioru danych i wyciągnięcie pewnych podstawowych wniosków i uogólnień na temat zbioru.

# WNIOSKOWANIE STATYSTYCZNE W ZAKRESIE





# STATYSTYKA

nauka poświęcona metodom badania prawidłowości **ZJAWISK MASOWYCH** polega na systematyzowaniu obserwowanych cech ilościowych i jakościowych oraz przedstawianiu wyników w postaci zestawień tabelarycznych, wykresów, itp.; posługuje się **rachunkiem prawdopodobieństwa**. Jej celem jest poznanie występujących prawidłowości, ich ilościowe wyrażenie oraz wyodrębnienie w nich składnika systematycznego i przypadkowego.

Jakie ma dochody przeciętny Kowalski?

Jak wygląda przeciętny Kowalski?

Jak odżywia się przeciętny Kowalski?



**BADANIE PEŁNE**

**STATYSTYKA OPISOWA**

**POPULACJA**

**ZBIÓR OBIEKTÓW OBJĘTYCH  
BADANIEM STATYSTYCZNYM  
Z WYRÓŻNIONĄ CECHĄ WSPÓLNA**

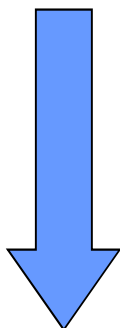
- **PŁEĆ**
- **WZROST**
- **WAGA**
- **LICZBA DZIECI**
- **ZAROBKI**
- **ODŻYWIANIE**
- **KOLOR OCZU**

**JAKI WZROST MA PRZECIĘTNY KOWALSKI ?**  
**Jaki są zarobki przeciętnego Kowalskiego?**



**BADANIE WYRYWKOWE  
REPREZENTACYJNE**

**POPULACJA**

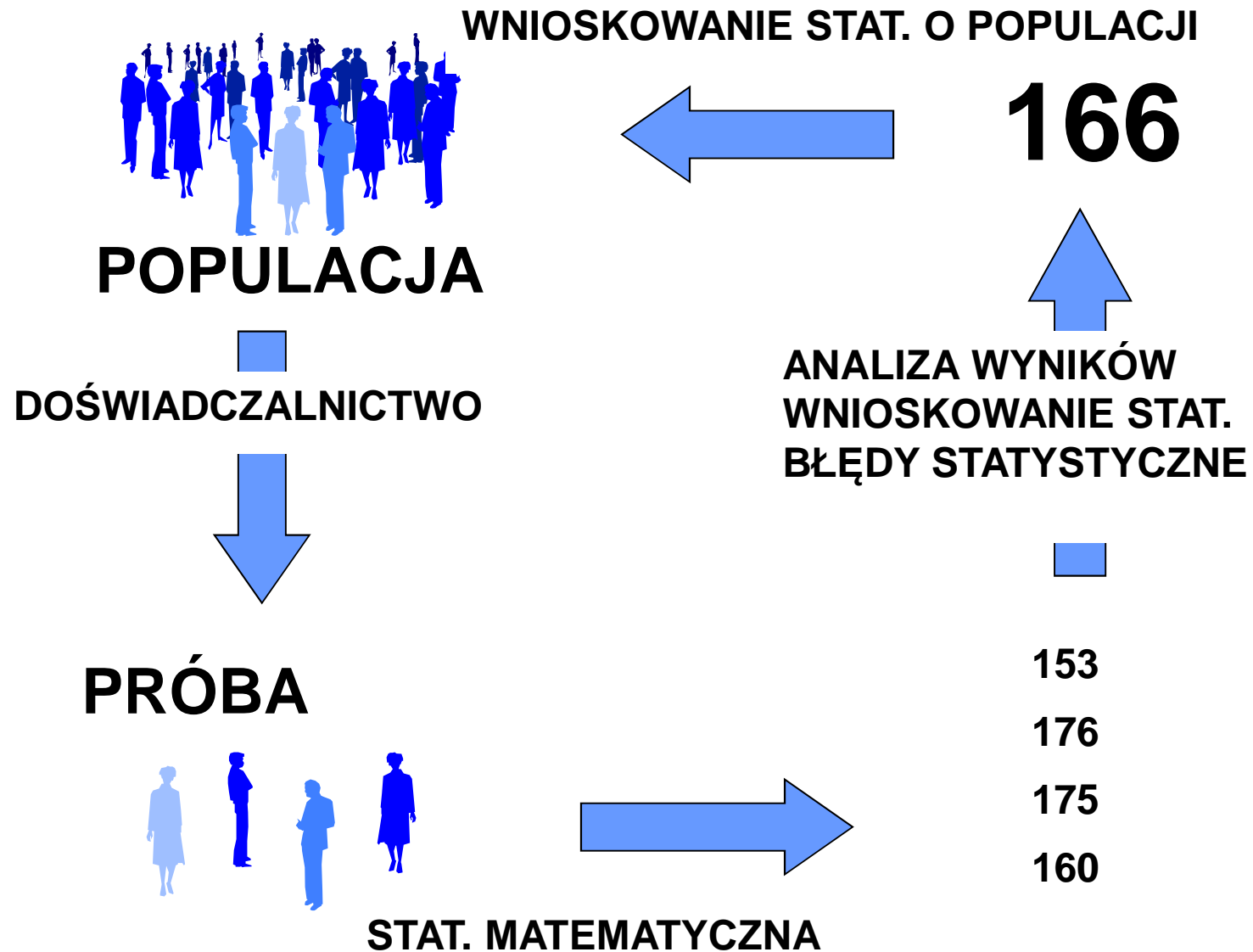


**PRÓBA**

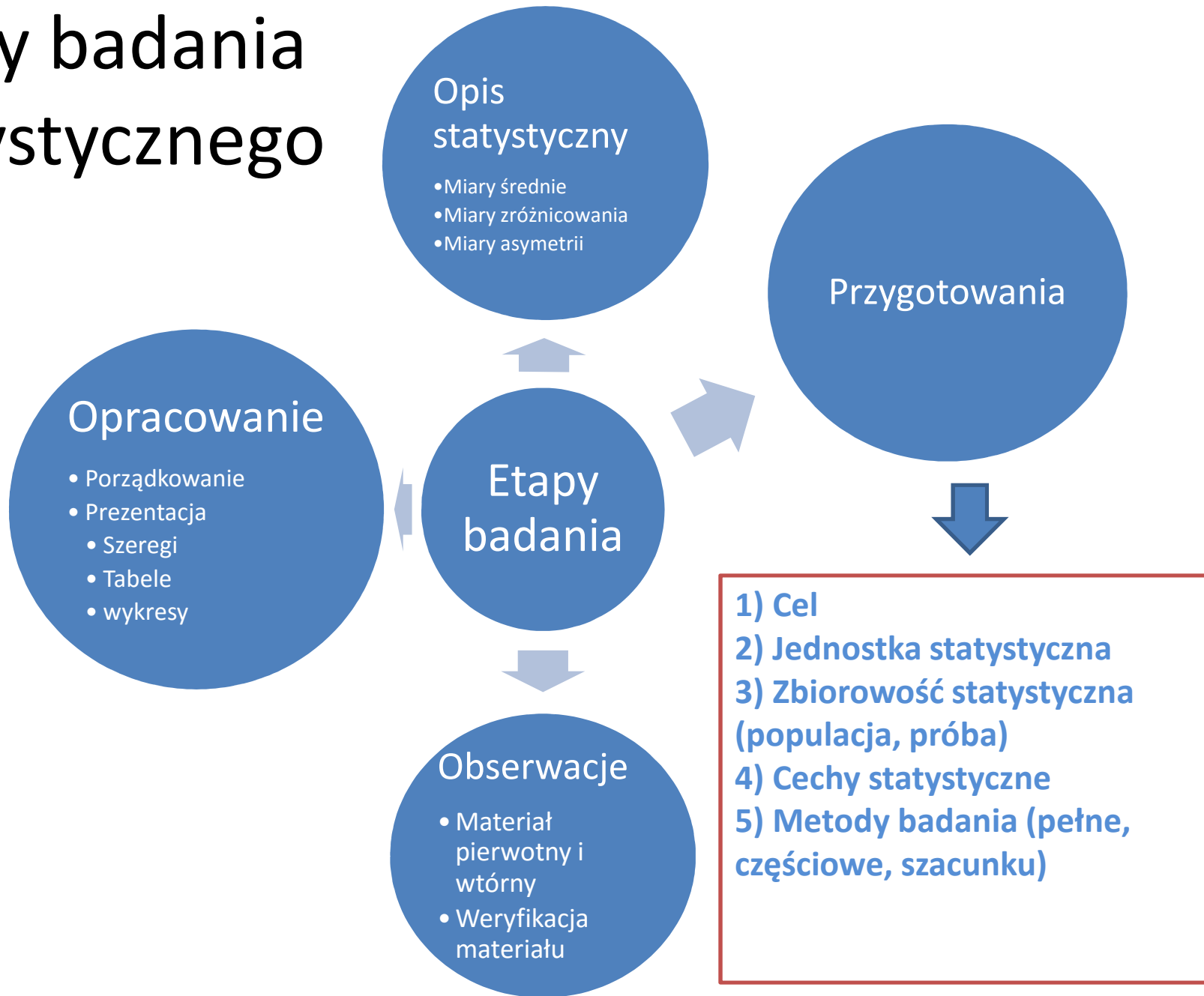


**WYBRANA CZĘŚĆ POPULACJI PODLEGAJĄCA BADANIU**

JAKI WZROST MA PRZECIĘTNY KOWALSKI ?  
Jaki są zarobki przeciętnego Kowalskiego?



# Etapy badania statystycznego



# Etapy badania statystycznego

- 1) Cel
- 2) Jednostka statystyczna
- 3) Zbiorowość statystyczna (populacja, próba)
- 4) Cechy statystyczne
- 5) Metody badania (pełne, częściowe, szacunku)

**Przygotowania**

**Opis statystyczny**

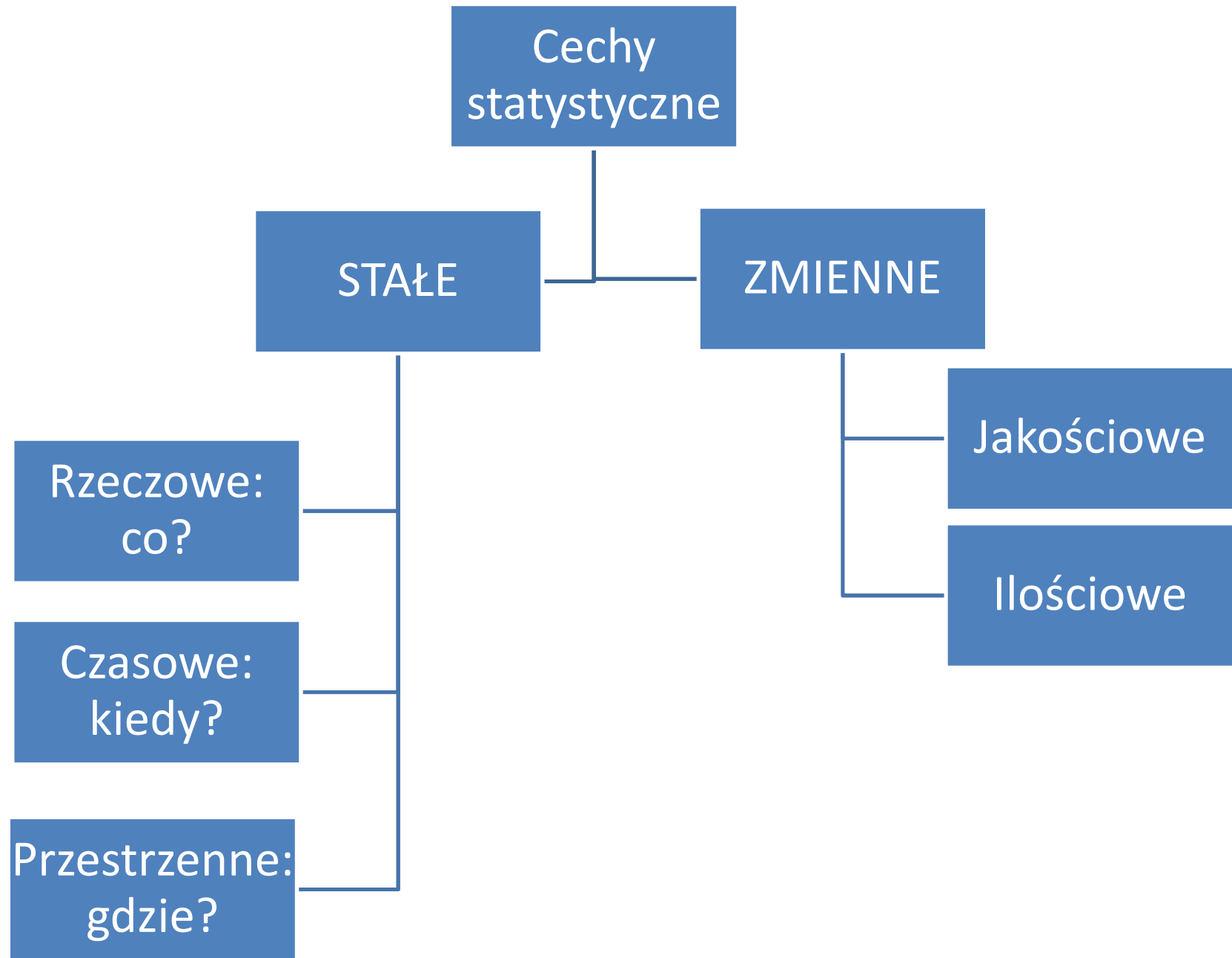
Miary średnie, Miary różnicowania, Miary asymetrii

**Obserwacje**

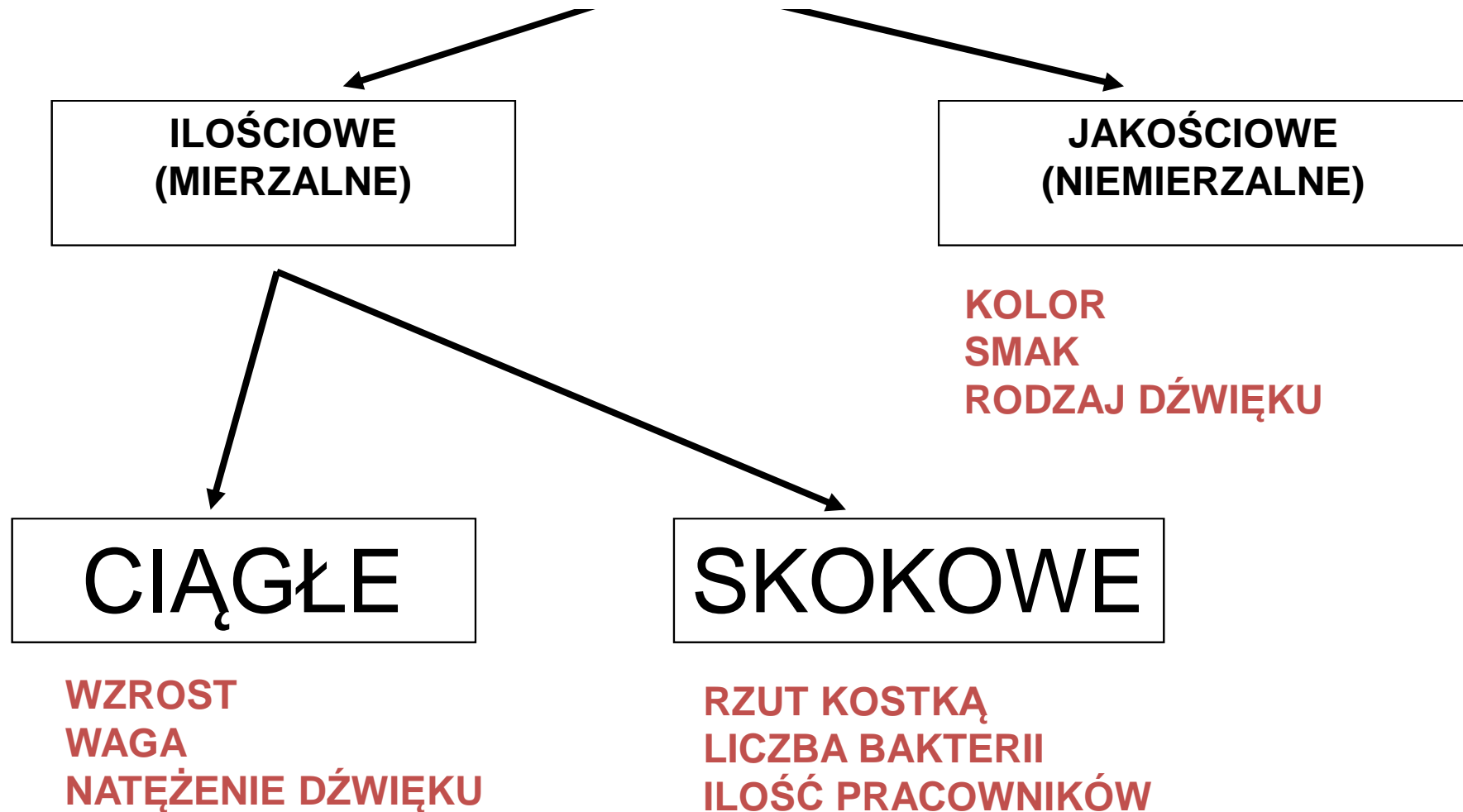
Materiał pierwotny i wtórny  
Weryfikacja materiału

**Opracowanie**

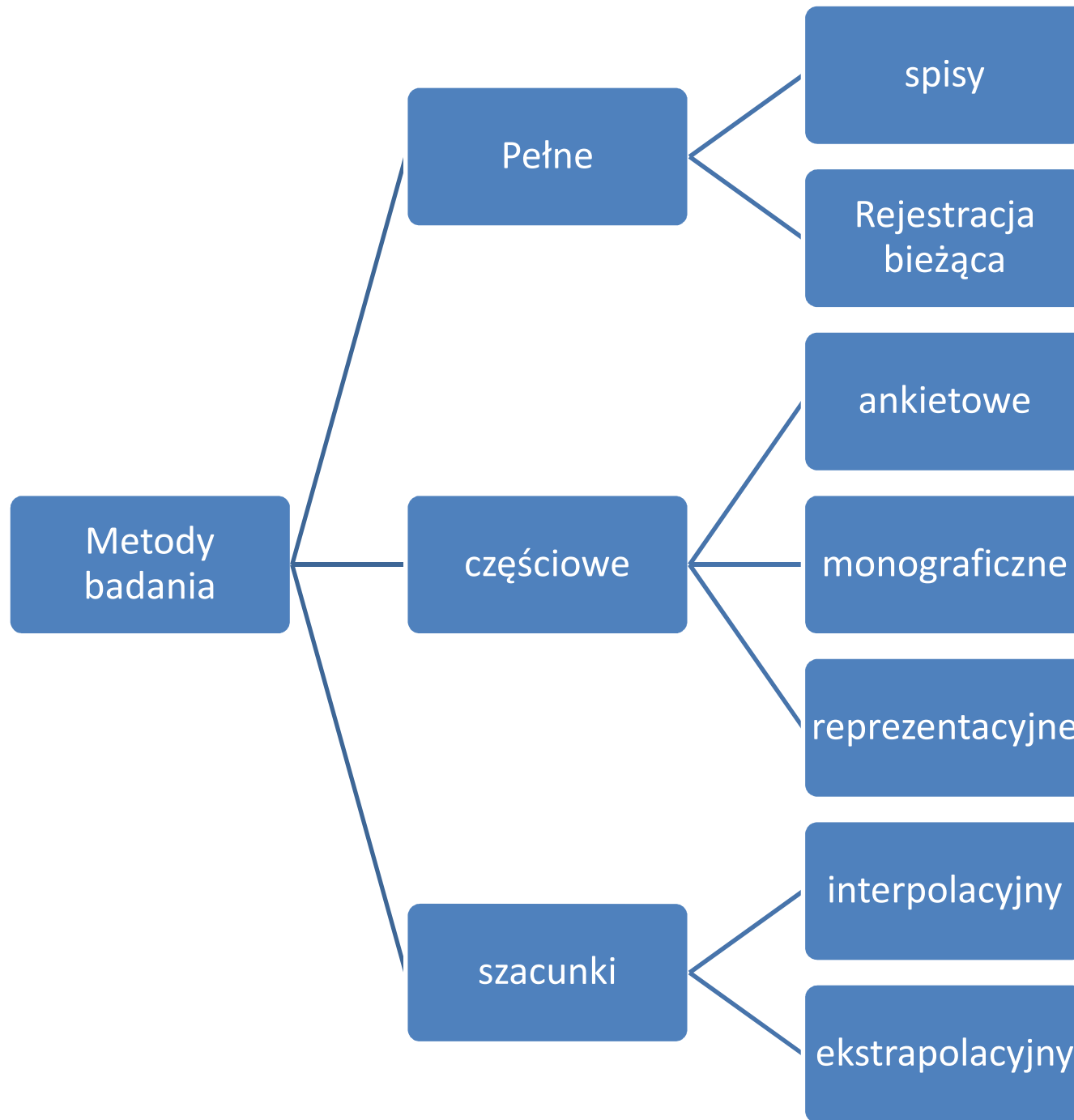
Porządkowanie, Prezentacja, Szeregi, Tabele, wykresy

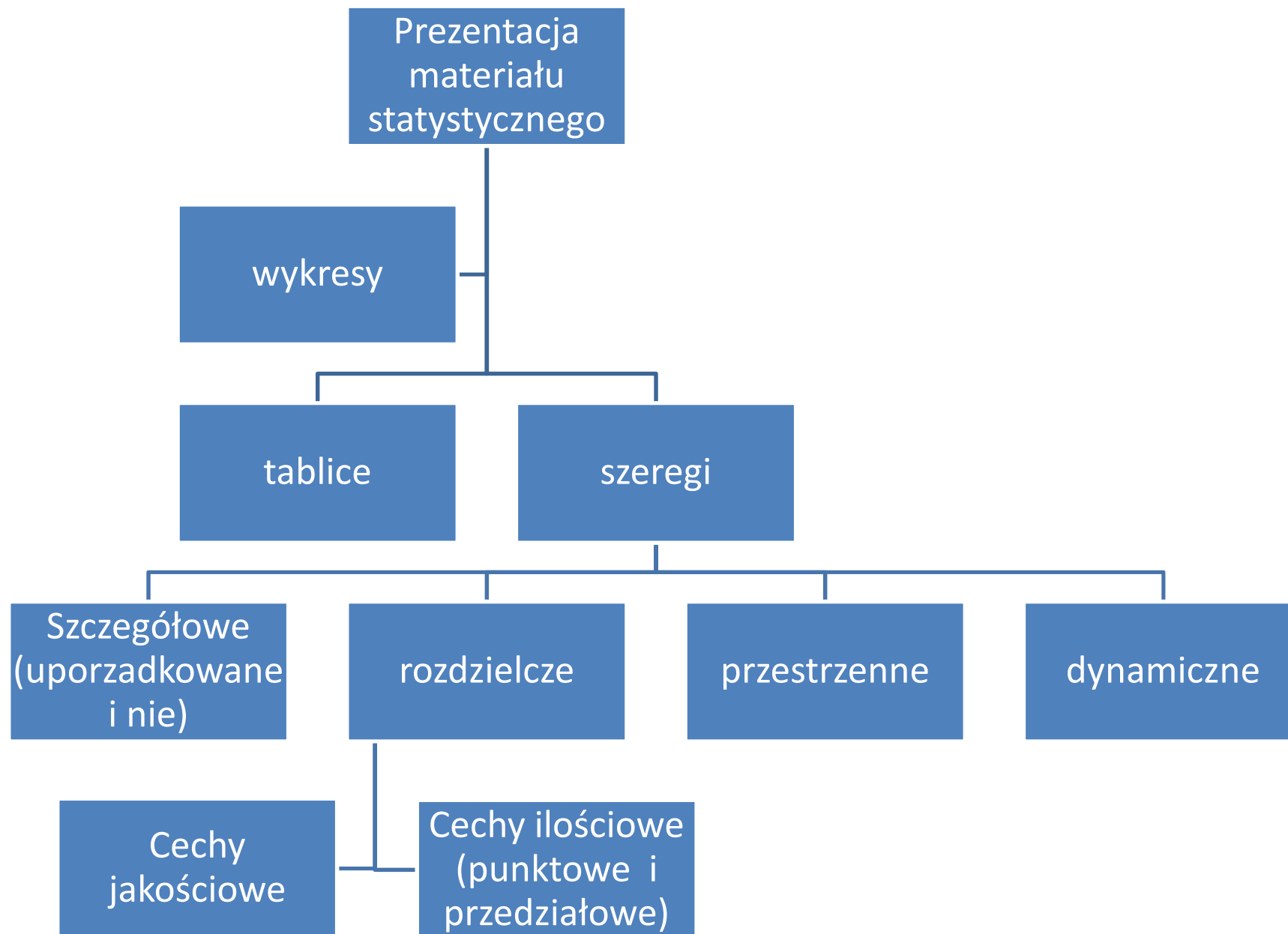


Funkcja o wartościach rzeczywistych określona na zbiorze zdarzeń elementarnych.

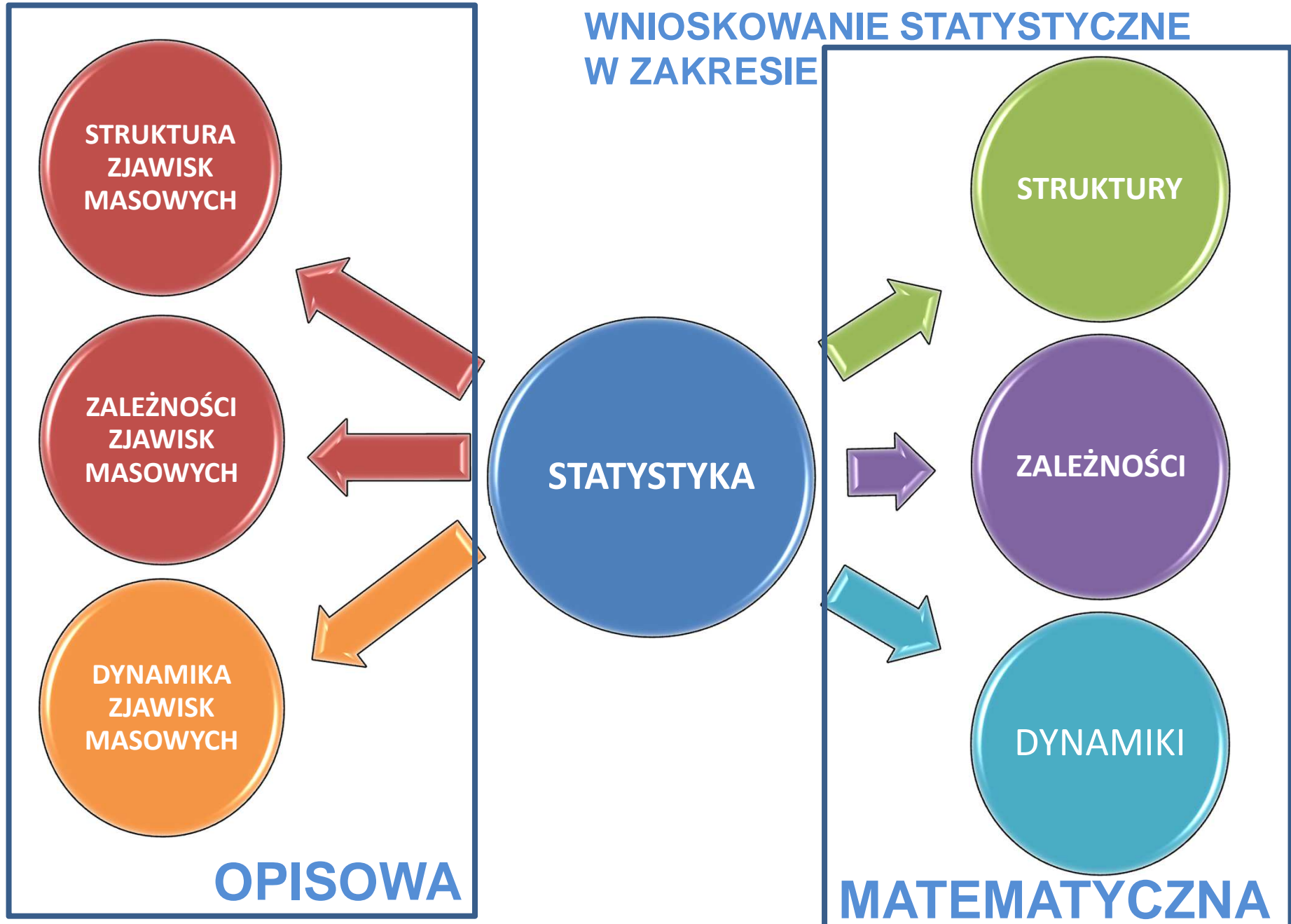








# WNIOSKOWANIE STATYSTYCZNE W ZAKRESIE

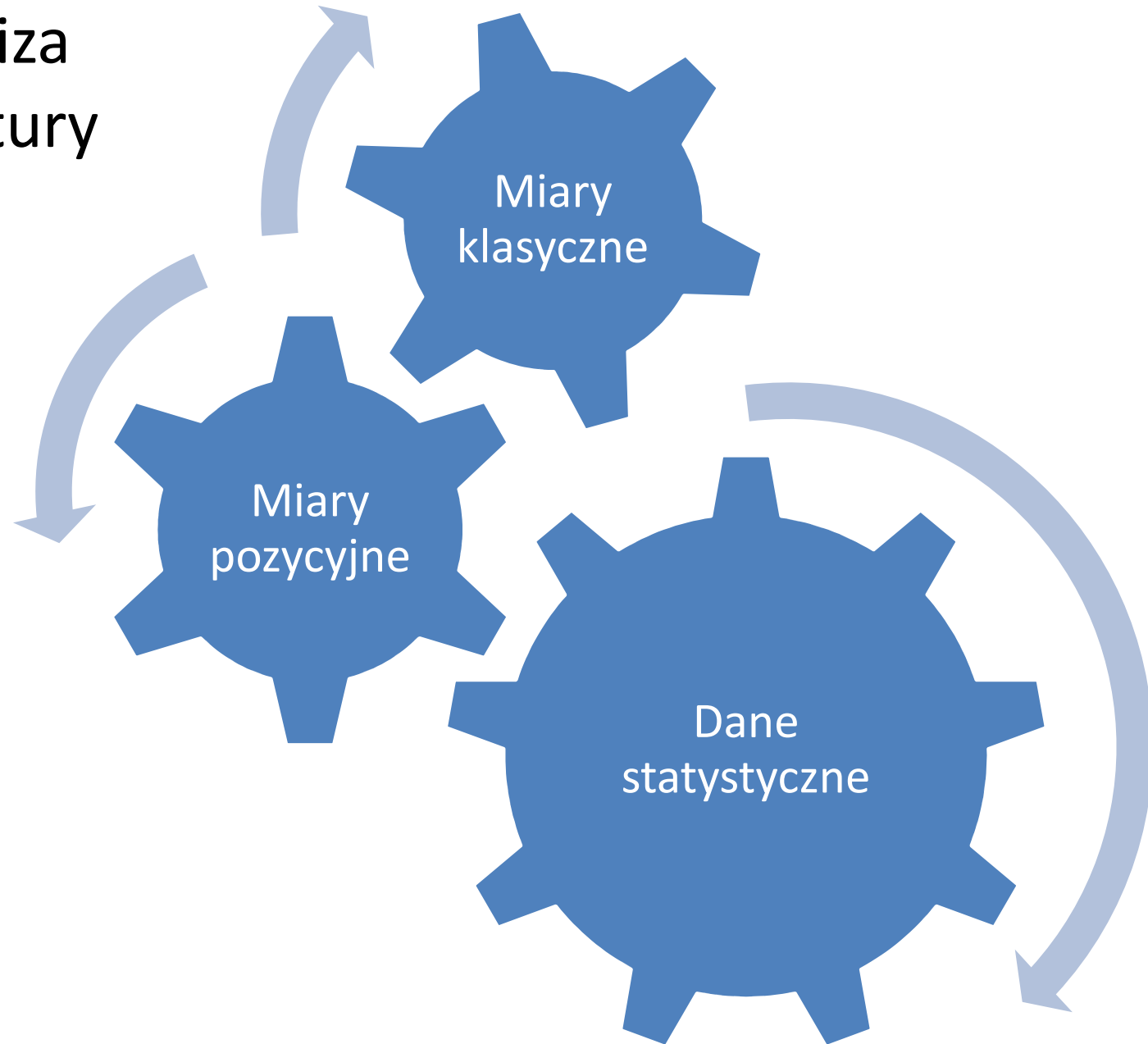


- Szereg szczegółowy
- Szereg rozdzielczy

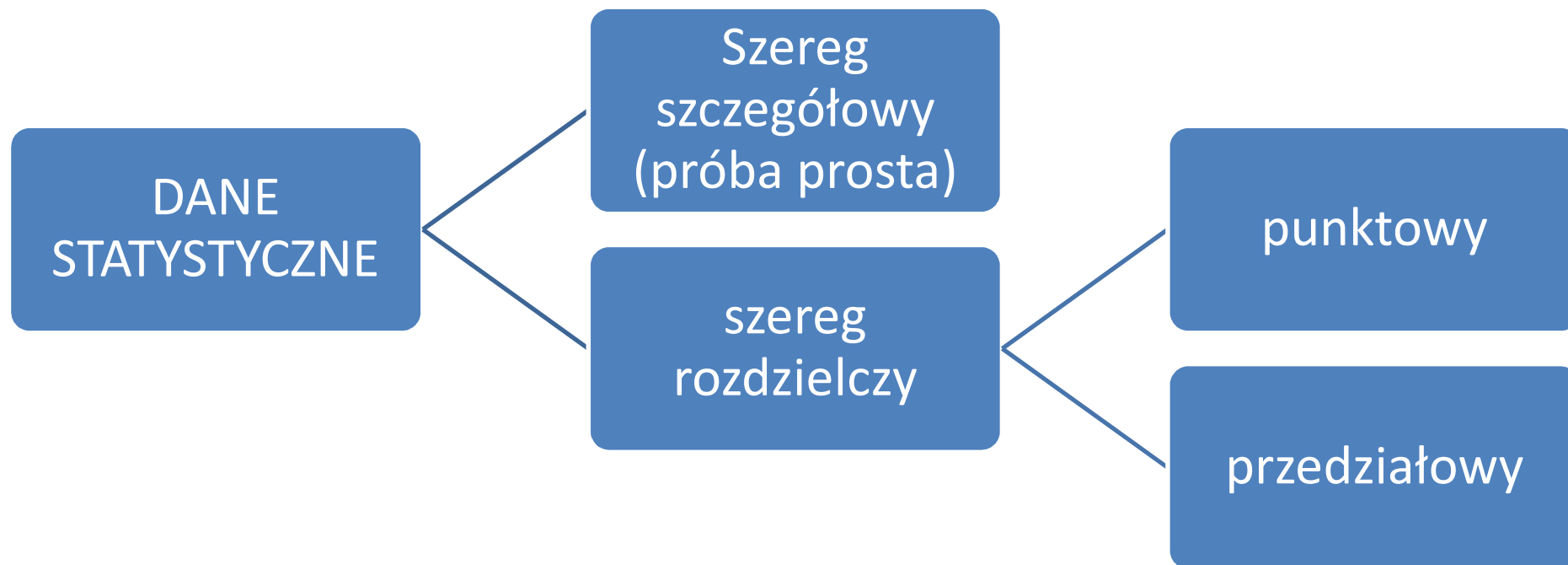
- Położenia
- Zmienności
- Koncentracji
- Skupienia



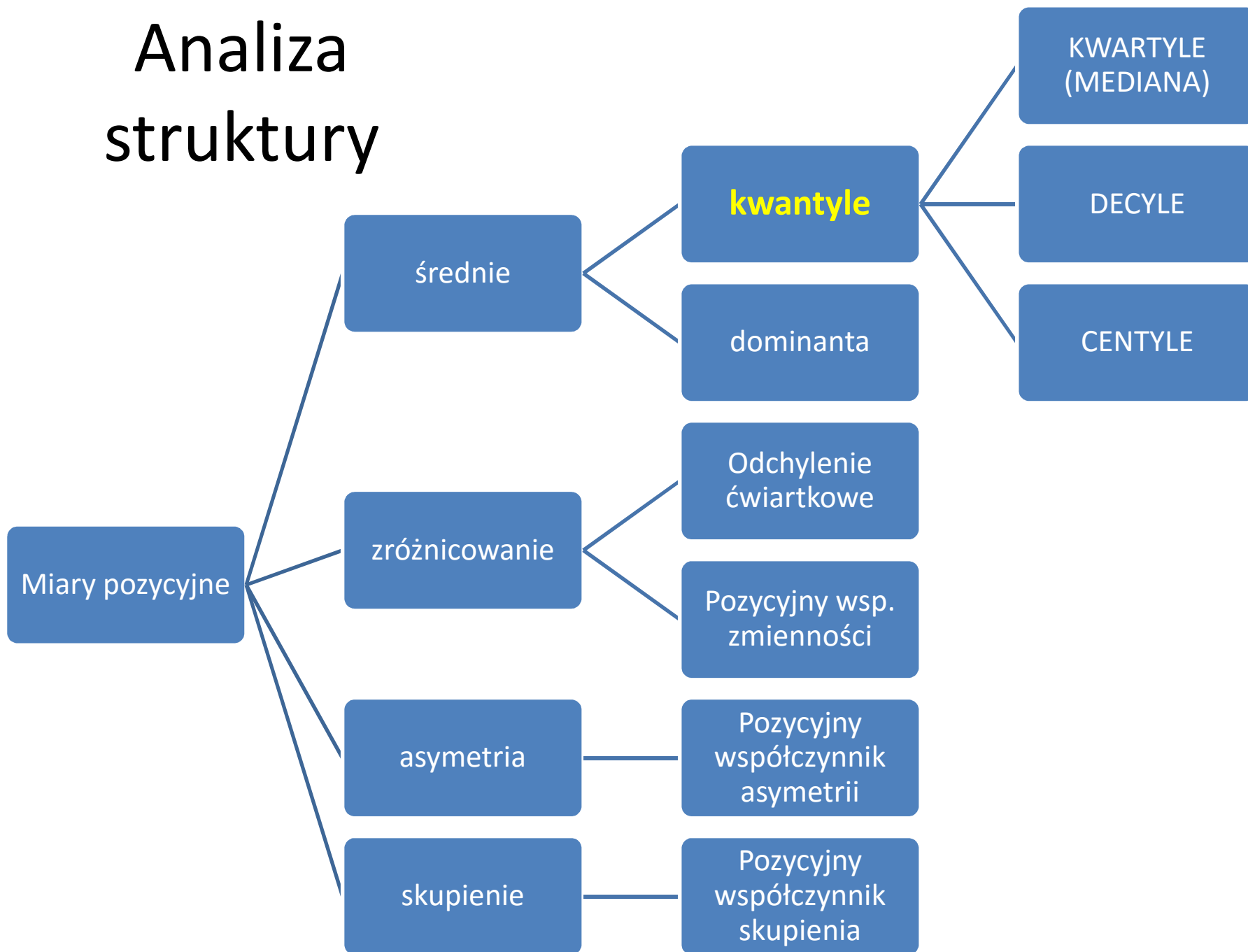
# Analiza struktury



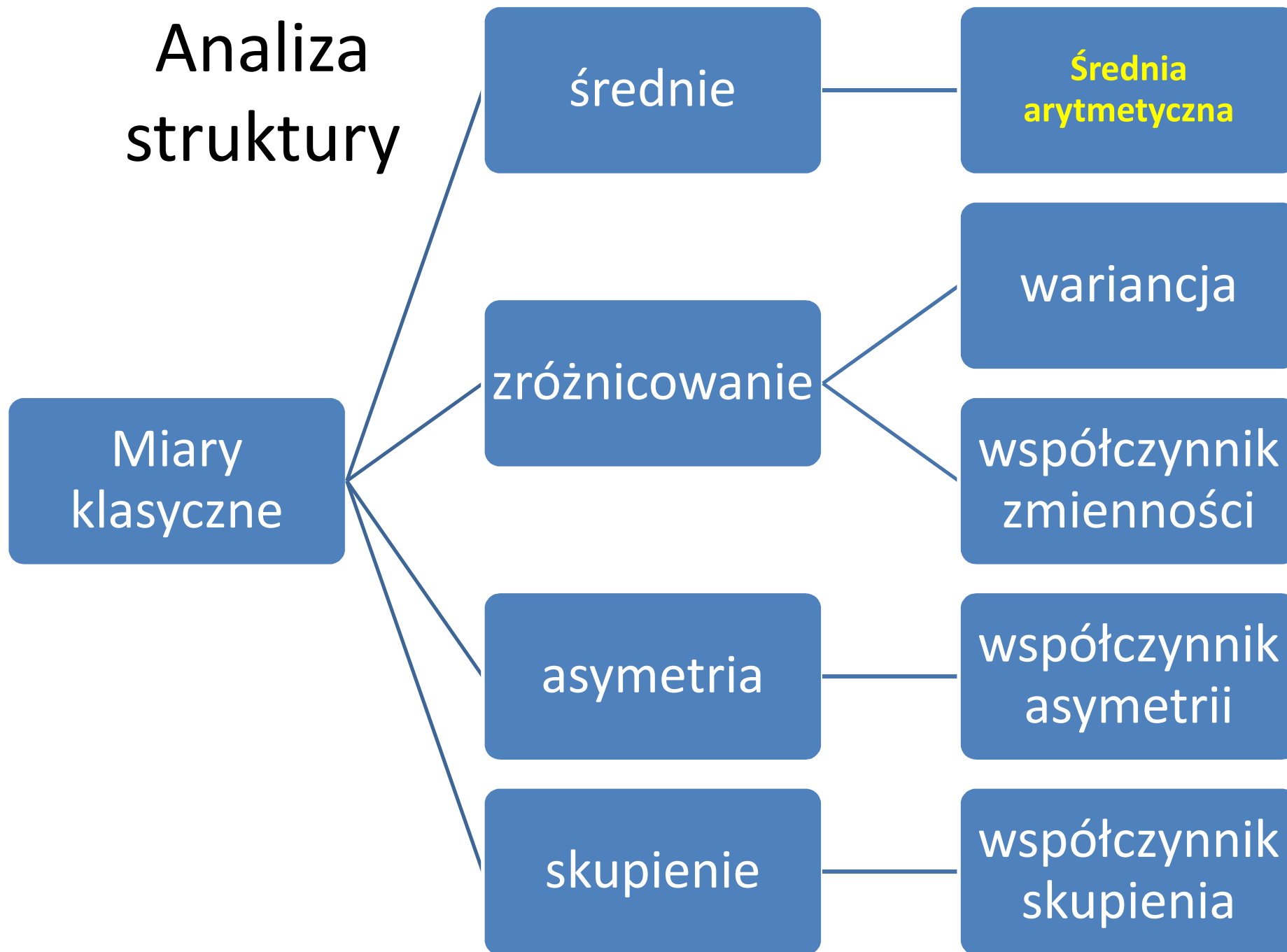
# Analiza struktury



# Analiza struktury



# Analiza struktury





# KLASYCZNE miary położenia

- Średnia arytmetyczna

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{N} = \frac{\sum_{i=1}^n x_i}{N}$$

# POZYCYJNE miary położenia

- Dominanta (moda)
- Kwantyle (kwartyle, decyle, centyle, percentyle)

# KWANTYL RZĘDU ALPHA...

$K_\alpha$  – kwantyl rzędu  $\alpha$ , ( $\alpha \in \langle 0, 1 \rangle$ )  
wartość dzieląca zbiorowość na dwie części, takie że  
w pierwszej części wszystkie wartości są nie większe  
niż  $K_\alpha$  podczas gdy w drugiej części wszystkie są nie  
mniejsze niż  $K_\alpha$ . Proporcje liczebności:  $\alpha \cdot 100\%$  do  
 $(1 - \alpha) \cdot 100\%$

# KLASYCZNE miary zmienności

- Wariancja

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}$$

- Odchylenie standardowe

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}}$$

# KLASYCZNE miary zmienności

- Odchylenie przeciętne

$$d = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{N}$$

- Współczynnik zmienności

$$V = \frac{s}{\bar{x}}$$

# POZYCYJNE miary zmienności

- Rozstęp

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

- Odchylenie kwartyłowe

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

- Współczynnik zmienności

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

- 1) Cel
- 2) Jednostka statystyczna
- 3) Zbiorowość statystyczna (populacja, próba)
- 4) Cechy statystyczne
- 5) Metody badania (pełne, częściowe, szacunku)

1. Zbadanie jak kształtuje się powierzchnia mieszkań w pewnym mieście
2. Mieszkanie
3. Mieszkania
4. Powierzchnia mieszkania [m<sup>2</sup>]
5. Pełne/Częściowe

# Pomiar, skale pomiarowe

- SKALA NOMINALNA – wartości na tej skali nie mają oczywistego uporządkowania (np. płeć), relacją umożliwiającą porównanie jest równoważność (równość)
- SKALA PORZĄDKOWA – wartości mają jasno określony porządek (np. wykształcenie), oprócz równoważności możliwe są relacje porządkujące
- SKALA INTERWAŁOWA (PRZEDZIAŁOWA) – różnice między wartościami mają sensowną interpretację, ale ich iloraz nie ma sensu (np. data urodzenia, temperatura w stopniach Celsjusza)
- SKALA ILORAZOWA (STOSUNKOWA) – nie tylko różnice, ale także ilorazy wielkości mają interpretację (np. masa – coś może być dwa razy cięższe), wielkości w skali ilorazowej można dodawać, odejmować i dzielić przez siebie

CECHA BADANA: powierzchnia mieszkania w m<sup>2</sup>

	1	2	3	4	5	6	7	8	9	10
0	35	35	35	35	35	35	35	35	35	35
10	35	45	45	45	45	45	45	45	45	45
20	45	45	45	45	45	45	45	45	45	45
30	45	45	45	45	55	55	55	55	55	55
40	55	55	55	55	55	55	55	55	55	55
50	55	55	55	55	55	55	55	55	55	55
60	55	55	55	55	55	55	55	65	65	65
70	65	65	65	65	65	65	65	65	65	75
80	75	75	75	75	75	85	85	85	85	85
90	85	85	85	95	95	95	105	105	115	115

przedział	środek	liczba mieszkań	częstości
30 – 40	35	11	0.11
40 – 50	45	23	0.23
50 – 60	55	33	0.33
60 – 70	65	12	0.12
70 – 80	75	6	0.06
80 – 90	85	8	0.08
90 – 100	95	3	0.03
100 – 110	105	2	0.02
110 – 120	115	2	0.02
razem		100	1

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



CECHA BADANA: powierzchnia mieszkania w m<sup>2</sup>

	1	2	3	4	5	6	7	8	9	10
0	35	35	35	35	35	35	35	35	35	35
10	35	45	45	45	45	45	45	45	45	45
20	45	45	45	45	45	45	45	45	45	45
30	45	45	45	45	55	55	55	55	55	55
40	55	55	55	55	55	55	55	55	55	55
50	55	55	55	55	55	55	55	55	55	55
60	55	55	55	55	55	55	55	65	65	65
70	65	65	65	65	65	65	65	65	65	75
80	75	75	75	75	75	85	85	85	85	85
90	85	85	85	95	95	95	105	105	115	115

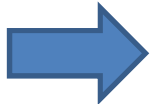
przedział	środek	liczba mieszkań	częstości
30 – 40	35	11	0.11
40 – 50	45	23	0.23
50 – 60	55	33	0.33
60 – 70	65	12	0.12
70 – 80	75	6	0.06
80 – 90	85	8	0.08
90 – 100	95	3	0.03
100 – 110	105	2	0.02
110 – 120	115	2	0.02
razem		100	1

$$K_{\alpha} = x_{\alpha} + h_{\alpha} \frac{\alpha \cdot N - n_{(\alpha)}}{n_{\alpha}}$$

CECHA BADANA: powierzchnia mieszkania w m<sup>2</sup>

$$K_{\alpha} = 50 + 10 \cdot \frac{0,5 \cdot 100 - 34}{33} (\alpha)$$

przedział	środek	liczba mieszkań	częstości
30 – 40	35	11	
40 – 50	45	23	
50 – 60	55	33	0.33
	65	12	0.12
	75	6	0.06
80 – 90	85	8	0.08
90 – 100	95	3	0.03
100 – 110	105	2	0.02
110 – 120	115	2	0.02
razem		100	1


  
 $11 + 23 = 34$ 
  
 $60 - 50 = 10$

Wzrost studenta (przedział)	X <sub>d</sub>	X <sub>g</sub>	n <sub>i</sub>	n <sub>(i)</sub>	zawartość liczb w przedziale n	środek	w <sub>i</sub>	w <sub>(i)</sub>
150 - 154,6	150	154,6	3	3	1, 2, 3	152,30	0,03	0,03
154,6 - 159,2	154,6	159,2	5	8	4, 5, 6, 7, 8	156,90	0,05	0,08
159,2 - 163,8	159,2	163,8	13	21	9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21	161,50	0,13	0,21
163,8 - 168,4	163,8	168,4	22	43	22, 23, 24, <b>25</b> , 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43	166,10	0,22	0,43
168,4 - 173	168,4	173	15	58	44, 45, 47, 48, 49, <b>50</b> , 51, 52, 53, 54, 55, 56, 57, 58	170,70	0,15	0,58
173 - 177,6	173	177,6	15	73	59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73	175,30	0,15	0,73
177,6 - 182,2	177,6	182,2	10	83	74, <b>75</b> , 76, 77, 78, 79, 80, 81, 82, 83	179,90	0,10	0,83
182,2 - 186,8	182,2	186,8	10	93	84, 85, 86, 87, 88, 89, 90, 91, 92, 93	184,50	0,10	0,93
186,8 - 191,4	186,8	191,4	6	99	94, 95, 96, 97, 98, 99	189,10	0,06	0,99
191,4 - 196	191,4	196	1	100	100	193,75	0,01	1,00
		Razem	100				1,00	

# Szereg szczegółowy

- 175 167 163 163 **150** 166 180 183 186 169  
190 177 170 180 170 167 191 181 171 165  
175 165 166 169 162 181 167 169 173 190  
165 191 157 164 162 159 180 173 171 184  
173 175 158 166 164 173 165 171 167 163  
165 196 172 162 157 164 168 186 171 158  
183 **154** 191 166 186 184 172 160 177 169  
183 160 **152** 175 177 162 170 185 161 182  
165 185 168 172 160 167 179 167 175 165  
190 173 179 178 177 162 170 180 162 176

- **150** 160 163 165 167 **170** 173 177 181 186
- 152** 161 164 166 168 171 173 177 181 186
- 154** 162 164 166 168 171 173 177 182 186
- 157 162 164 166 169 171 175 178 183 190
- 157 162 165 166 169 171 175 179 183 190
- 158 162 165 167 169 172 175 179 183 190
- 158 162 165 167 169 172 175 180 184 191
- 159 162 165 167 170 172 175 180 184 191
- 160 163 165 167 170 173 176 180 185 191
- 160 163 165 167 **170** 173 177 180 185 196

# PYTANIA

1. Dane są następujące obserwacje: 1, 2, 3, 4, 5. Oblicz średnią, medianę, dominantę i odchylenie standardowe.
2. Podano dane w postaci szeregu rozdzielczego.

Przedziały	Liczebności
10 – 20	1
20 – 30	8
30 – 40	1

Oblicz: średnią, wariancję i medianę.

3. W jakich przypadkach miary obliczane na podstawie szeregów rozdzielczych będą różnić od obliczanych z próby prostej.

# PYTANIA CD...

1. Jakie są własności poszczególnych miar tendencji centralnej?
2. Badano zarobki w dwóch zakładach pracy A i B. Uzyskano następujące wyniki: średnie w zakładzie A = 1800, B = 2000 i mediany w zakładzie A = 2000 i B = 1800. W którym zakładzie większość pracowników ma lepsze zarobki?
3. Wymienić pozycyjne miary położenia.
4. Wymienić klasyczne miary zmienności.
5. Wymienić pozycyjne miary zmienności.
6. W jakich przypadkach nie należy stosować współczynnika zmienności?
7. W pewnej zbiorowości wyznaczono średnia wartość badanej zmiennej i odchylenie standardowe, uzyskując odpowiednio: 100 i 9. Podaj zakres typowej zmienności.
8. Dla pewnego rocznika studentów średni wynik ze statystyki wynosi 3.57 oraz mediana 3.28. Czy większość studentów ma ocenę ze statystyki większą od średniej czy nie? Odpowiedz uzasadnić.