

Temat

Obliczenia statystyczne. Pakiet
STATISTICA

Zagadnienia

- 1.** Wprowadzenie: statystyczny opis danych.
- 2.** Zastosowanie pakietu statystycznego STATISTICA, arkusza kalkulacyjnego EXCEL lub CALC

Statystyczny opis danych

Dane liczbowe z pomiaru: X_1, X_2, \dots, X_n .

Pierwszym etapem analizy danych jest ich opis:

- rozkład wartości (w postaci tabel, wykresów)
- charakterystyka za pomocą parametrów (np. średnia arytmetyczna, mediana, odchylenie standardowe)

Przykład

Wiek chorych na pewną chorobę – **szereg statystyczny nieuporządkowany**.

30	80	92	65	90	19	21	81	76	24
80	38	31	38	38	20	92	28	50	55
73	47	30	62	41	48	69	61	44	57
49	56	43	56	45	27	45	70	89	44
52	56	61	49	53	66	81	56	34	88
76	49	51	45	48	48	60	57	43	59
41	51	85	69	17	36	76	59	52	43
77	54	51	47	48					

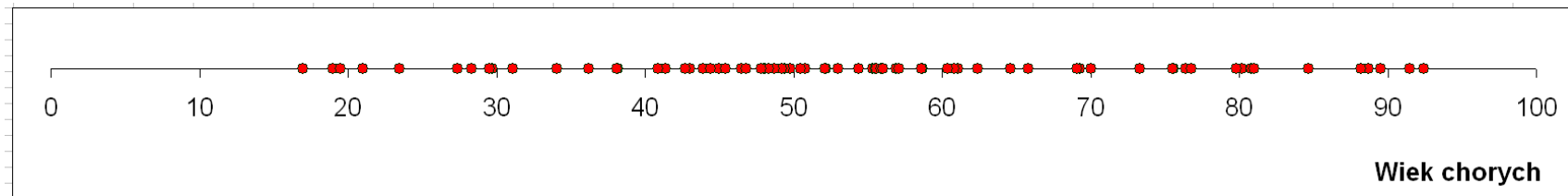
Przykład cd.

Wiek chorych na pewną chorobę – **szereg statystyczny uporządkowany** (posortowany rosnąco).

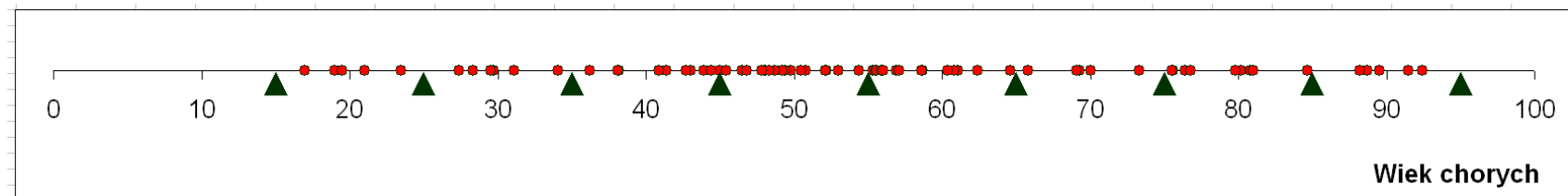
17	19	20	21	24	27	28	30	30	31
34	36	38	38	38	41	41	43	43	43
44	44	45	45	45	47	47	48	48	48
48	49	49	49	50	51	51	51	52	52
53	54	55	56	56	56	56	57	57	59
59	60	61	61	62	65	66	69	69	70
73	76	76	76	77	80	80	81	81	85
88	89	90	92	92					

Przykład cd.

Wiek chorych na pewną chorobę – szereg statystyczny uporządkowany (posortowany rosnąco).



Cały zakres wartości dzielimy na klasy,



zliczamy wartości w każdej klasie. Rezultat można przedstawić w tabeli lub na wykresie.

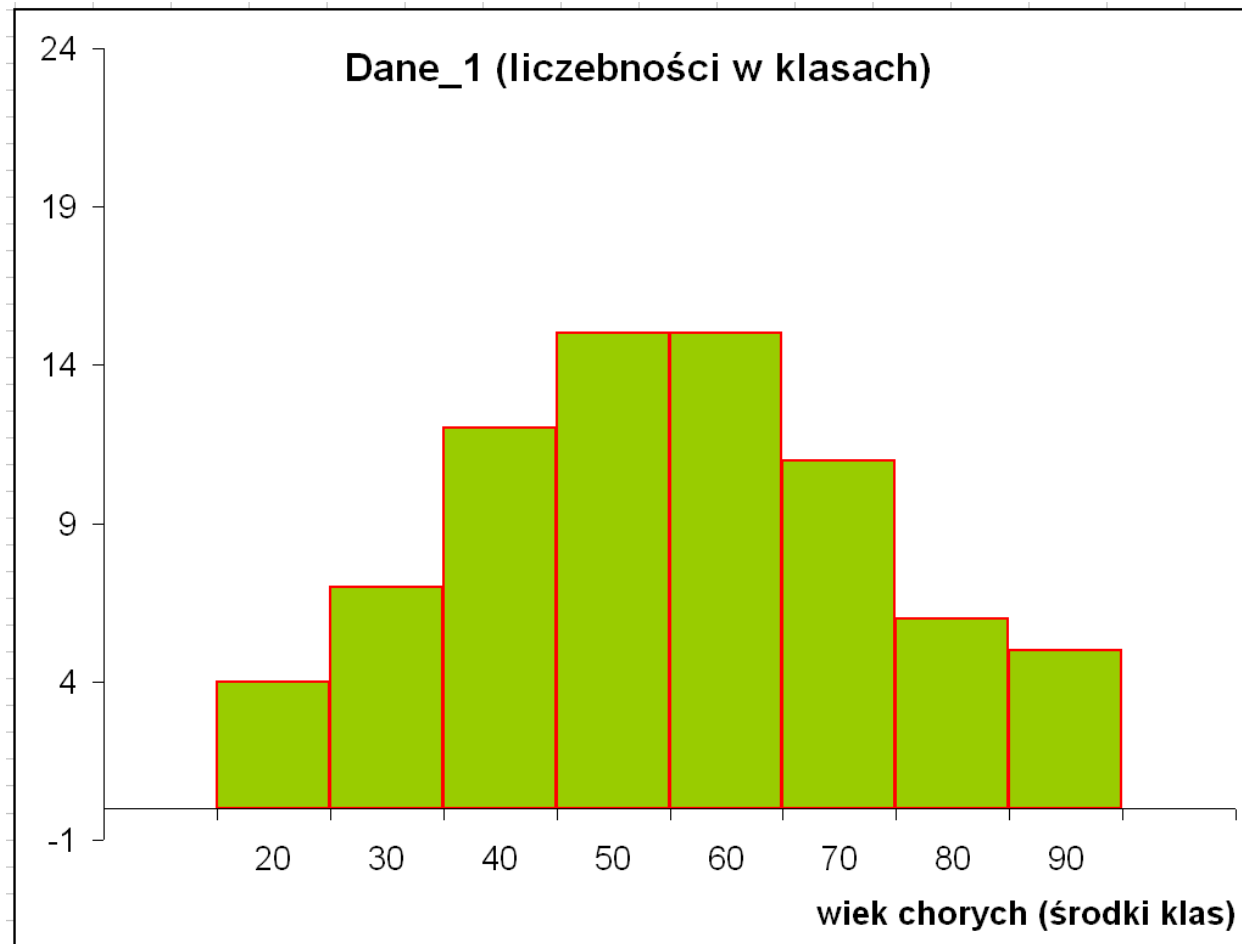
Empiryczny rozkład wartości

... przedstawiony w postaci **szeregu rozdzielczego**

Klasy wiekowe	Liczebność
<15; 25)	4
<25; 35)	7
<35; 45)	12
<45; 55)	15
<55; 65)	15
<65; 75)	11
<75; 85)	6
<85; 95>	5
Razem	75

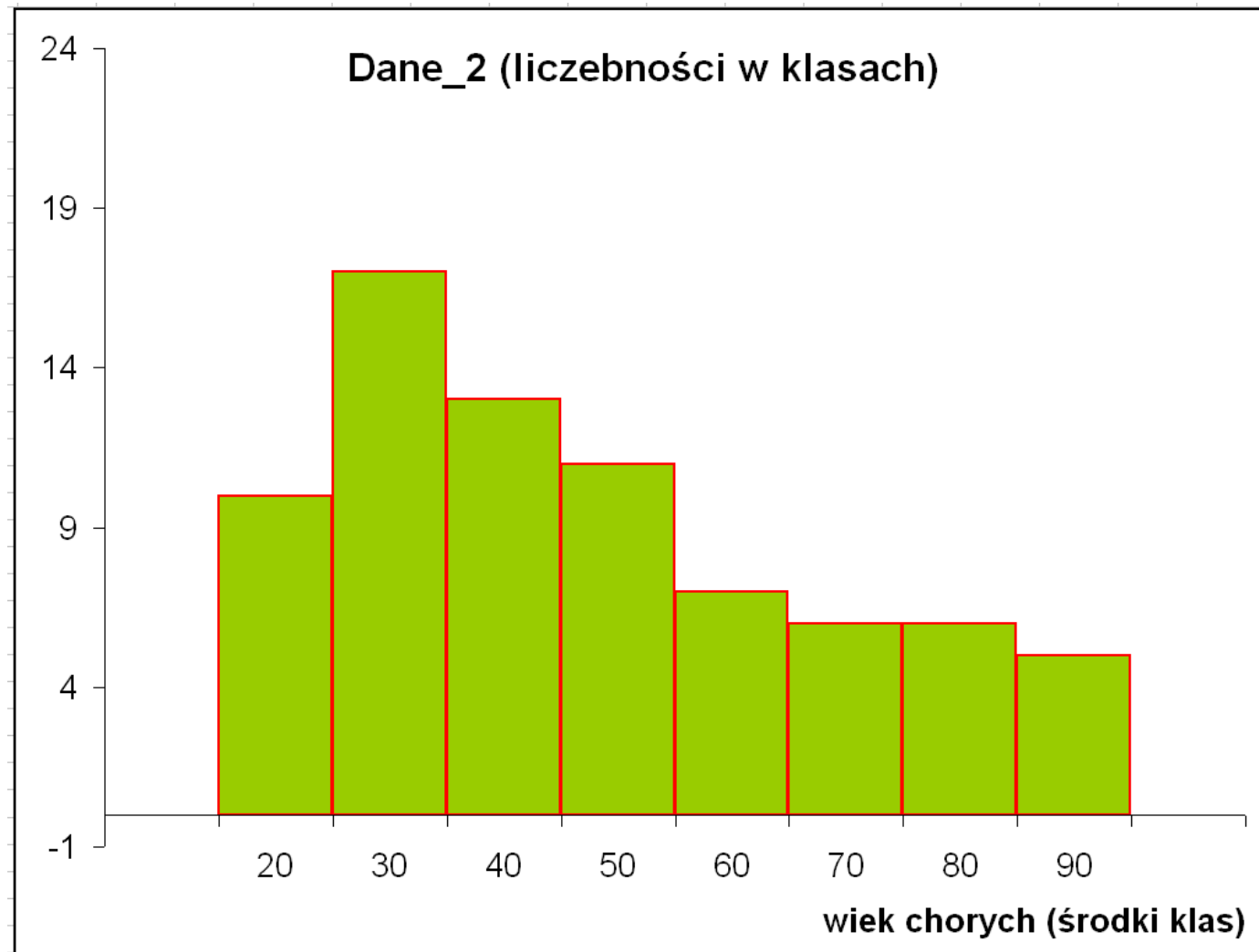
Empiryczny rozkład wartości

... przedstawiony w postaci **histogramu**



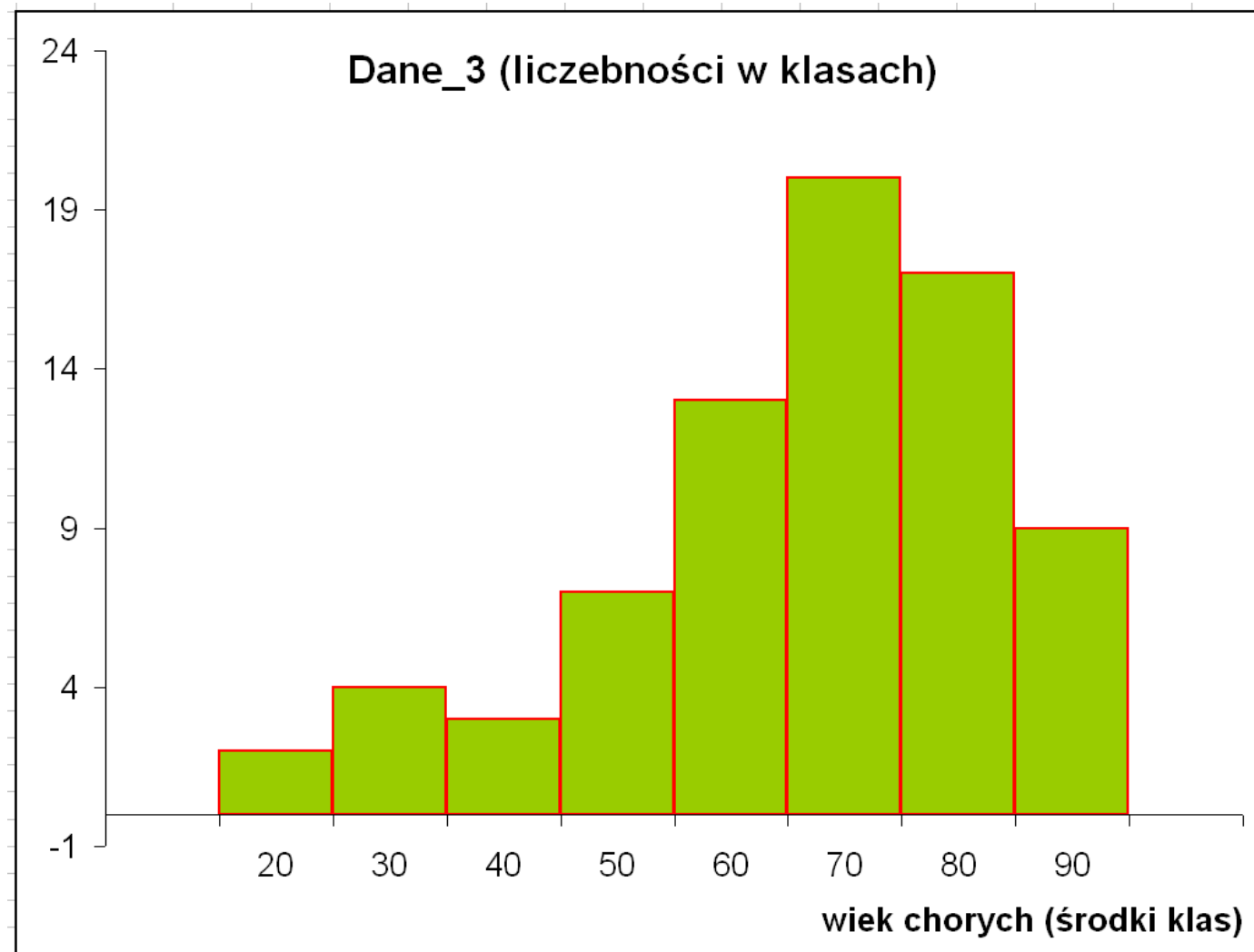
Rozkład symetryczny

Empiryczny rozkład wartości cd.



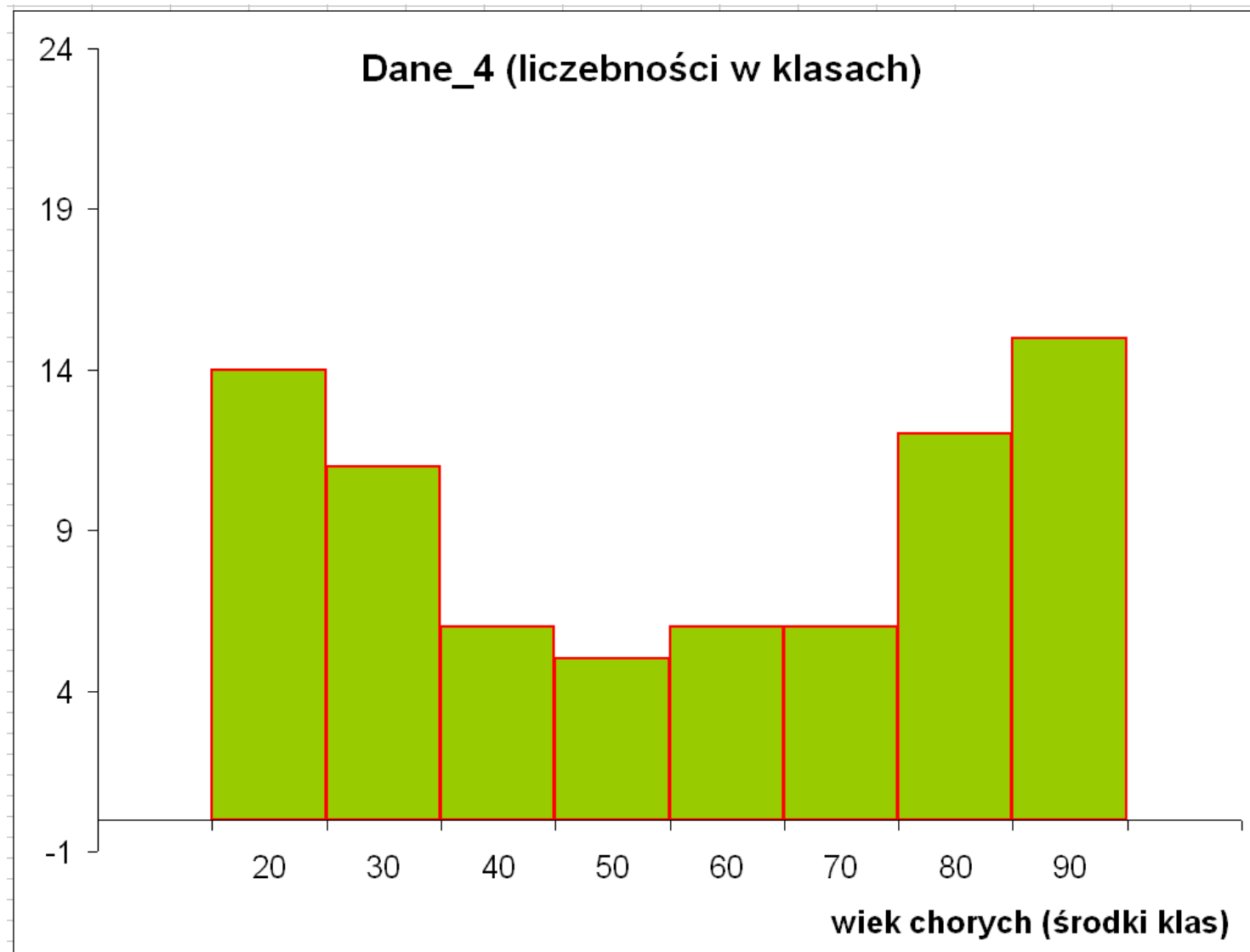
Rozkład asymetryczny
(prawoskośny, asymetria prawostronna)

Empiryczny rozkład wartości cd.



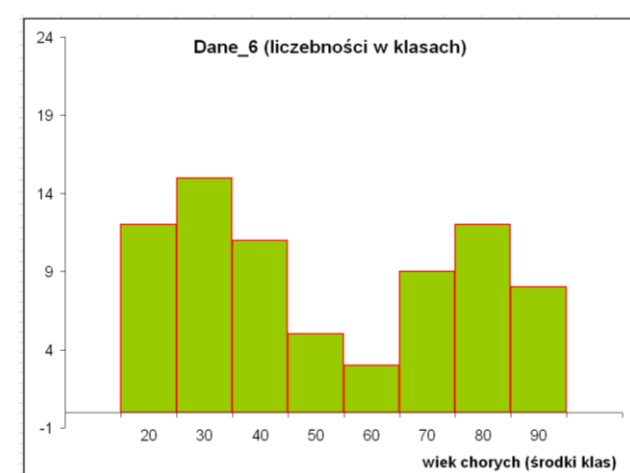
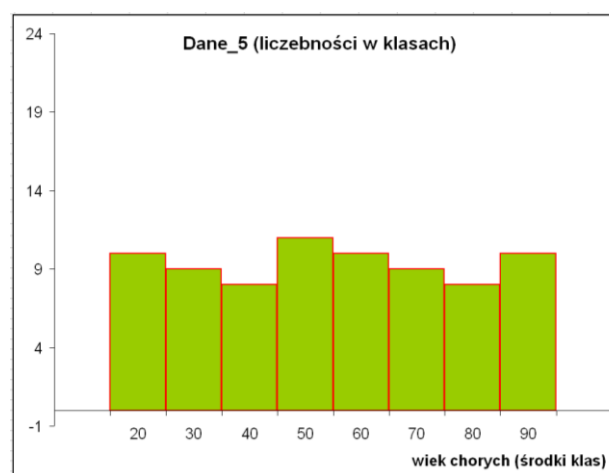
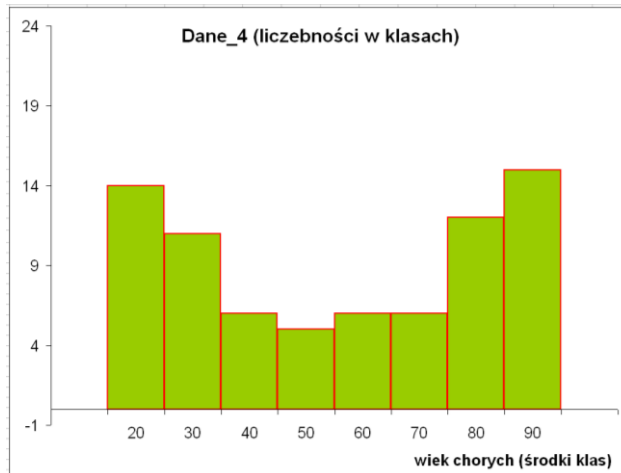
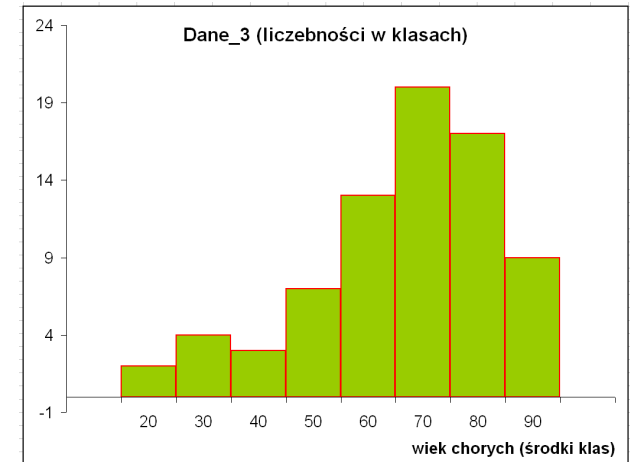
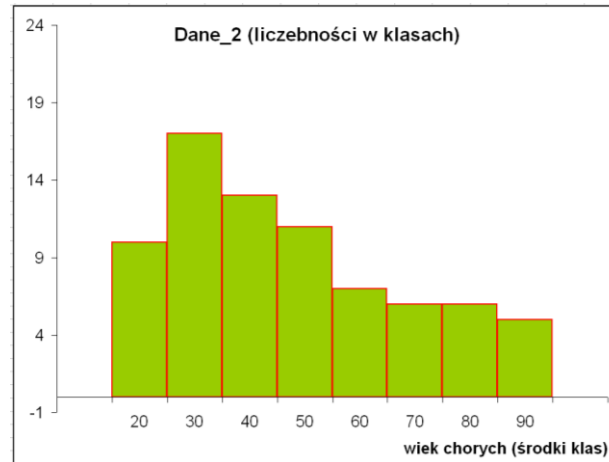
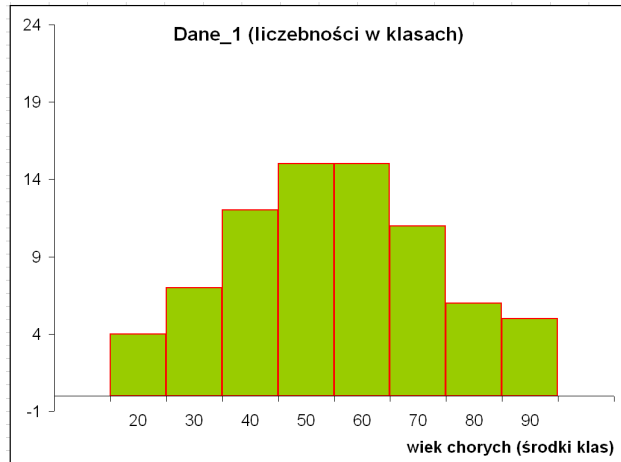
Rozkład asymetryczny
(lewoskośny, asymetria lewostronna)

Empiryczny rozkład wartości cd.



Rozkład typu U

Typy rozkładów empirycznych



Statystyczny opis danych

Dane liczbowe z pomiaru: X_1, X_2, \dots, X_n
w badaniu pełnym.

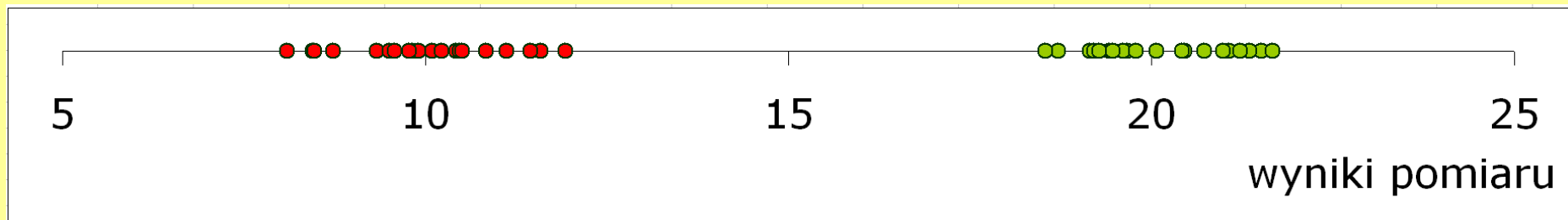
Pierwszym etapem analizy danych jest ich opis:

- rozkład wartości (w postaci tabeli, wykresu)
- charakterystyka za pomocą parametrów (np. średnia arytmetyczna, mediana, odchylenie standardowe)

Do czego potrzebujemy parametrów?

Parametry opisują własności całego zespołu danych liczbowych:

Rys. 1

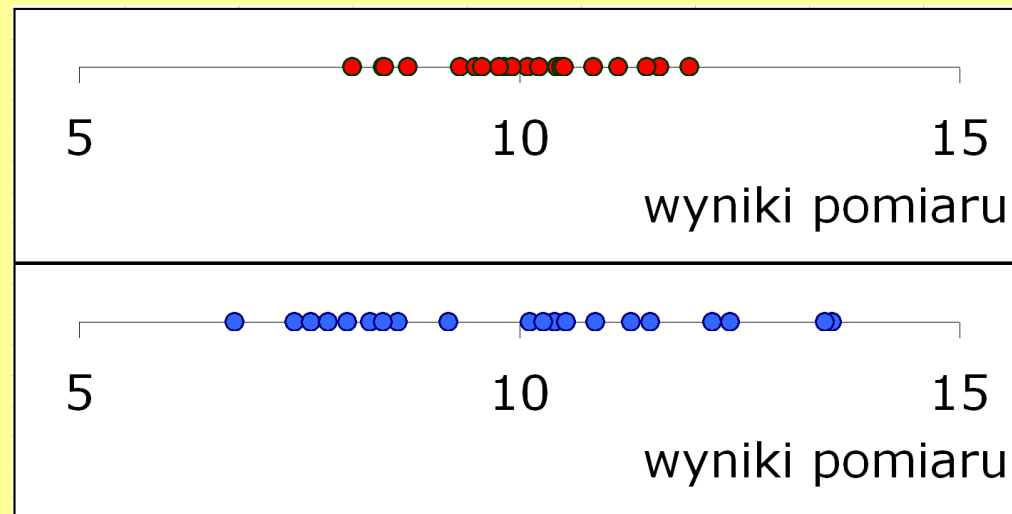


położenie (na osi liczbowej)

Do czego potrzebujemy parametrów?

Parametry opisują własności całego zespołu danych liczbowych:

Rys. 2

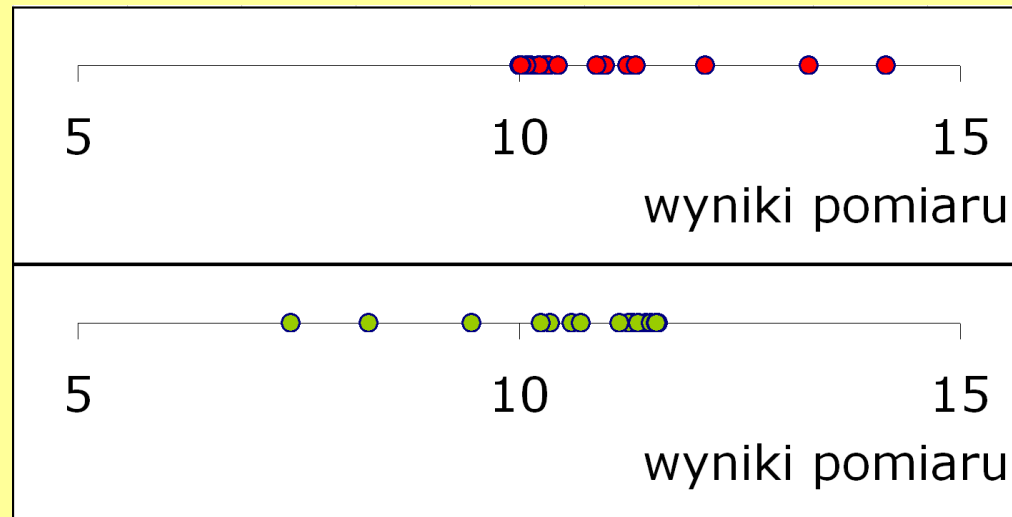


rozrzut (względem średniej)

Do czego potrzebujemy parametrów?

Parametry opisują własności całego zespołu danych liczbowych:

Rys. 3



asymetrię (względem średniej)

Parametry

Parametry opisują własności całego zespołu danych liczbowych:

- a. położenie (przeciętny poziom)
- b. rozrzut (rozproszenie, dyspersja) i zmienność
- c. asymetrię (skośność)
- d. spłaszczenie

Parametry klasyczne – obliczane na podstawie wszystkich wyników.

Parametry pozycyjne - wyznaczane na podstawie pozycji wyników w szeregu statystycznym lub częstości występowania.

Parametry położenia

Parametry położenia charakteryzują średni lub typowy poziom, wokół którego skupiają się wartości ze zbioru danych:

- wartość średnia (np. arytmetyczna, geometryczna, harmoniczna)
- wartość o ustalonej pozycji w rozkładzie (np. mediana i inne kwantyle)
- wartość najczęściej występująca (dominanta, inaczej: modalna, moda)

Średnia arytmetyczna

Dla danych x_1, x_2, \dots, x_n :

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

Przykład

W doświadczeniu medycznym badano czas snu pacjentów leczonych na pewną chorobę. Dla 12 pacjentów otrzymano wyniki (w min.):

435, 389, 533, 324, 561, 395, 416, 500, 499, 397, 356, 398.

Średni czas snu w badanej grupie pacjentów wynosi

$$\bar{x} = \frac{435 + 389 + \dots + 398}{12} = \frac{5203}{12} \approx 433,6 \text{ minuty}$$

Średnia harmoniczna

Średnia harmoniczna jest stosowana do danych wyrażonych w jednostkach względnych, np.: prędkość w km/h, zagęszczenie/zaludnienie w liczbie osób/km².

$$\bar{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Średnią harmoniczną stosujemy tylko do liczb dodatnich. Pozwala nadać większe znaczenie mniejszym wartościom w zbiorze danych.

Dominanta

Dominanta to wartość występująca najczęściej w zbiorze danych (modalna, moda).

Dla danych zestawionych w szeregu rozdzielczym wskazuje na nią najliczniejsza klasa, na histogramie – najwyższy szczyt.

Mediana

Mediana Me to wartość środkowa w zbiorze danych (parametr pozycyjny). Gdy liczba danych jest nieparzysta, medianę stanowi wartość środkowego elementu. W przypadku parzystej liczby pomiarów w próbie, medianę stanowi średnia wartość z dwóch sąsiadujących środkowych elementów uporządkowanego zbioru.

$$Me = \frac{x_{n/2} + x_{n/2+1}}{2} \quad \text{dla } n \text{ parzystych}$$

$$Me = x_{(n+1)/2} \quad \text{dla } n \text{ nieparzystych}$$

Przykłady

Przykład 1. Obliczanie mediany dla $n = 9$ danych:

35, 40, 36, 35, 39, 37, 38, 36, 38

Po uporządkowaniu:

35, 35, 36, 36, 37, 38, 38, 39, 40; $Me=37$

Przykład 2. Obliczanie mediany dla $n = 8$ danych:

35, 35, 36, 36, 37, 38, 38, 39;

$Me=(36+37)/2=36,5$

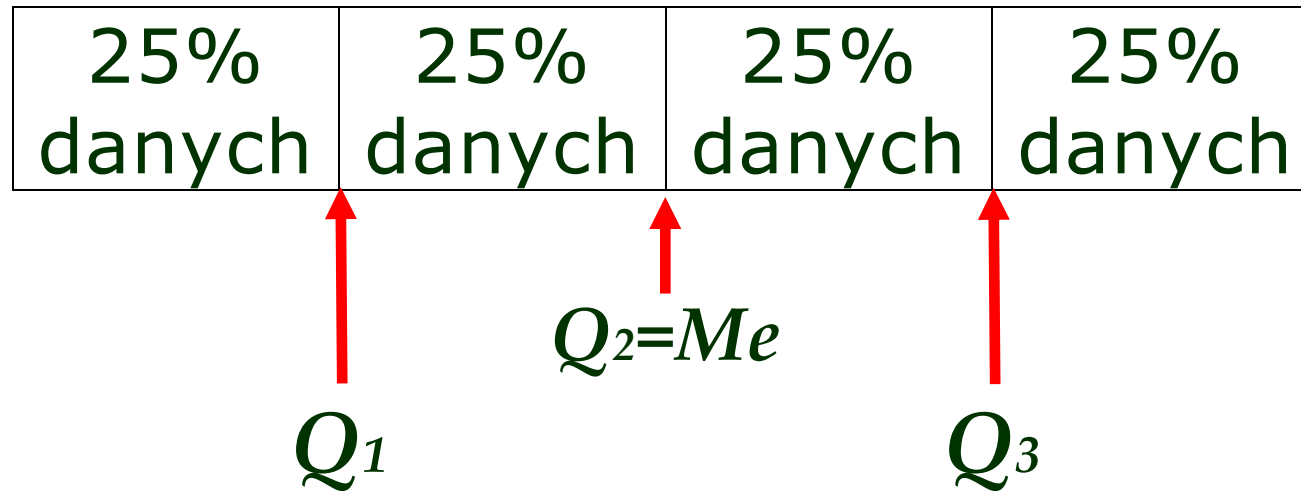
Kwantyle

Mediana jest jednym z kwantyli.

Kwantyle to wartości, które dzielą uporządkowany zbiór danych na części o jednakowej liczbie elementów.

Kwartyle dzielą zbiór danych na cztery części, **decyle** na dziesięć i **percentyle** (centyle) na sto części.

Kwartyle



Kwartyle

Pierwszy kwartył Q_1 dzieli uporządkowany zbiór danych w taki sposób, że 25% elementów zbioru ma wartości nie większe, a 75% nie mniejsze od tego elementu.

Drugi kwartył (mediana) $Q_2 = Me$ dzieli uporządkowany zbiór danych na dwie równe części. Wartości skrajne nie mają wpływu na jej wielkość.

Trzeci kwartył Q_3 dzieli uporządkowany zbiór danych w taki sposób, że 75% elementów zbioru ma wartości nie większe, a 25% nie mniejsze od tego elementu.

Parametry rozrzutu i zmienności

Parametry rozrzutu opisują zróżnicowanie w zbiorze danych:

- wariancja i odchylenie standardowe
- współczynnik zmienności
- rozstęp
- rozstęp kwartyłowy
- odchylenie ćwiartkowe

Wariancja

Wariancja s^2 pokazuje rozproszenie danych wokół średniej arytmetycznej.

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Najmniejsza wartość wariancji wynosi zero, im większe zróżnicowanie, tym większa wartość wariancji.

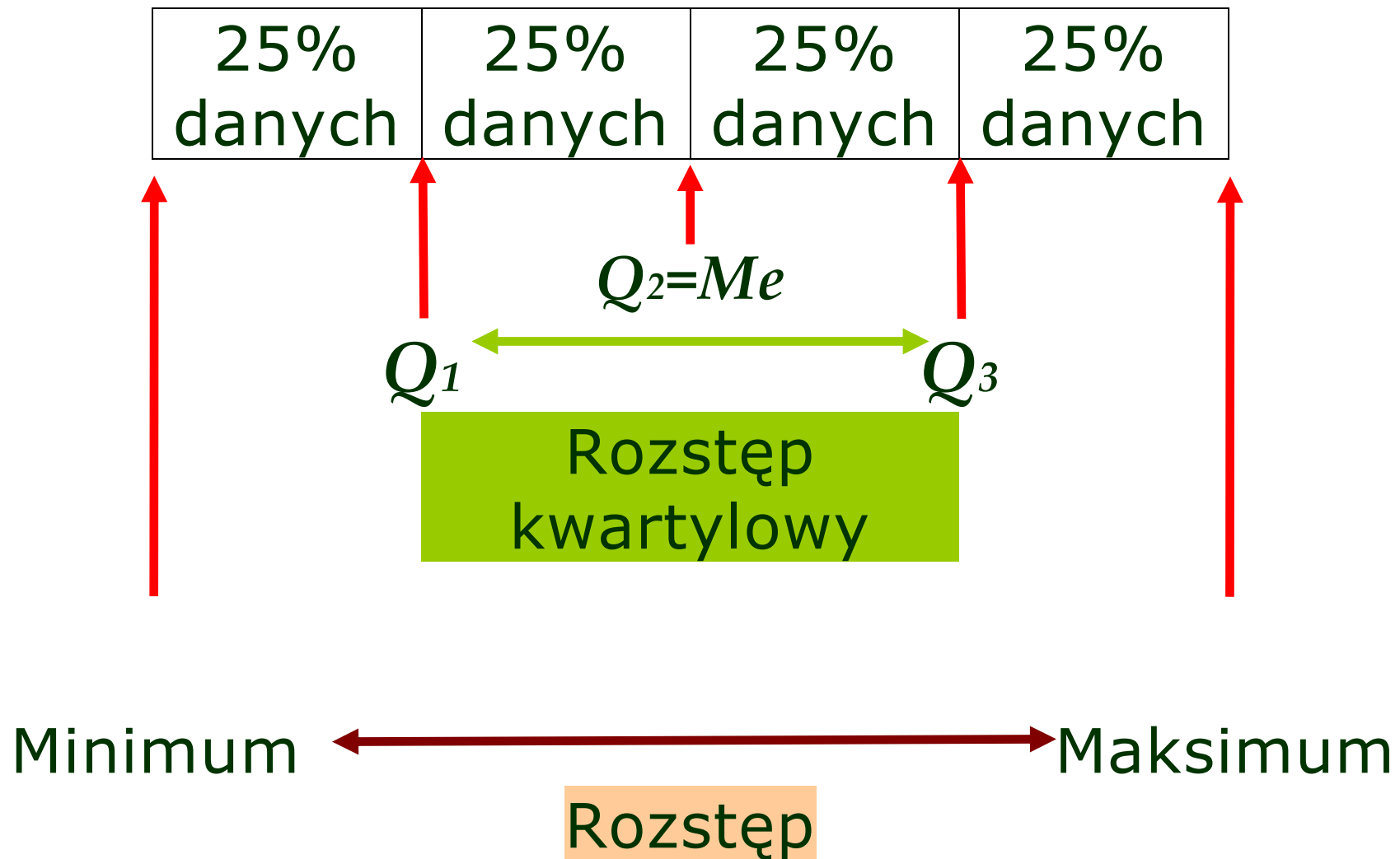
Odchylenie standardowe

Odchylenie standardowe s , SD (*standard deviation*)

$$SD = \sqrt{s^2}$$

Odchylenie standardowe pokazuje, o ile przeciętnie poszczególne wyniki różnią się od średniej, czyli pokazuje wielkość błędu pojedynczego pomiaru. Im mniejsza wartość odchylenia, tym obserwacje są bardziej skupione wokół średniej.

Parametry pozycyjne



Rozstęp

Rozstęp wskazuje na całkowity obszar zmienności badanej cechy. Jest obliczany jako różnica pomiędzy największą i najmniejszą wartością w zbiorze danych.

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

Określają go tylko dwie skrajne wartości, pozostałe nie mają żadnego wpływu na wielkość rozstępu.

Odchylenie ćwiartkowe

Rozstęp kwartylowy obliczamy jako różnicę pomiędzy trzecim i pierwszym kwartylem. Pomiedzy tymi kwartydami znajduje się 50% wszystkich obserwacji.

Odchylenie ćwiartkowe Q obliczamy jako połowę różnicy pomiędzy trzecim i pierwszym kwartylem.

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

Rozstęp kwartylowy i odchylenie ćwiartkowe są obliczane, gdy miarą położenia jest mediana. Wskazują one na rozrzut wyników wokół mediany.

Współczynnik zmienności

Współczynnik zmienności v , cv (*coefficient of variation*) oblicza się dzieląc odchylenie standardowe przez średnią arytmetyczną - jest parametrem bezwymiarowym:

$$v = \frac{SD}{\bar{x}} \quad \text{lub} \quad v = \frac{SD}{\bar{x}} \cdot 100\%$$

Im mniejszą przyjmuje wartość, tym mniejsza zmienność zbioru danych względem średniej.

Współczynnik zmienności - pozycyjny

Współczynnik zmienności pozycyjny V
oblicza się dzieląc odchylenie ćwiartkowe przez medianę:

$$V = \frac{Q}{Me} \quad \text{lub} \quad V = \frac{Q}{Me} \cdot 100\%$$

*Przykład

					średnia arytmetyczna	odch. standard	wsp. zmienności
Dane1	6	10	12	12	10,0	2,4	24%
Dane2	56	60	62	62	60,0	2,4	4%

Współczynnik zmienności umożliwia porównanie zmienności zbiorów danych różniących się znacznie wartością średniej lub zawierających pomiary wykonane w różnych jednostkach.

Współczynnik zmienności stosujemy do porównania, która cecha jest bardziej zróżnicowana relatywnie względem swojej średniej. Jeśli oceniamy zróżnicowanie względem mediany, to stosujemy pozycyjny współczynnik zmienności.

Parametry asymetrii (skośności)

Miarę asymetrii można oprzeć na spostrzeżeniu, że w szeregu statystycznym o rozkładzie symetrycznym średnia arytmetyczna, mediana oraz dominanta są równe

$$\bar{x} = Me = D$$

W szeregach asymetrycznych wartości tych parametrów odbiegają od siebie tym bardziej, im asymetria jest większa. **Wskaźnikiem asymetrii** może być różnica

$$\bar{x} - D$$

Gdy

$$\bar{x} - D > 0 \quad \textit{asymetria prawostronna}$$

$$\bar{x} - D < 0 \quad \textit{asymetria lewostronna}$$

Współczynnik asymetrii

Współczynnik asymetrii (jeden z wielu) jest liczbą niemianowaną, a jego znak mówi o kierunku asymetrii.

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

- $A_3=0$ rozkład symetryczny
- $A_3>0$ rozkład prawoskośny, asymetria prawostronna
- $A_3<0$ rozkład lewoskośny, asymetria lewostronna

Parametr spłaszczenia

Kurtoza (wskaźnik spłaszczenia, wskaźnik smukłości) wskazuje na koncentrację danych wokół średniej. Wartości kurtozy mniejsze od zera świadczą o rozkładzie spłaszczonym (platykurtycznym), zaś wartości większe od zera o rozkładzie wysmukłym (leptokurtycznym) w stosunku do rozkładu normalnego.

Podsumowanie

Parametry tak charakteryzują zbiór danych, że porównanie różnych zbiorów danych można sprowadzić do porównań parametrów.

Przykład

W dwóch grupach chorych zmierzono skurczowe ciśnienie krwi. Otrzymano wyniki:

grupa I: 145, 125, 130, 155, 140, 150, 135

grupa II: 115, 150, 100, 180, 140, 165, 130

Obliczenia w arkuszu kalkulacyjnym

	A	B	C	D	E	F	G	H	I	J
1		grupa I	grupa II							
2		145	115							
3		125	150							
4		130	100							
5		155	180							
6		140	140							
7		150	165							
8		135	130							
9		średnia arytmetyczna	140,0	140,0						
10		odchylenie standardowe	10,0	25,8						
11		mediana	140,0	140,0						
12		odchylenie ćwiartkowe	7,5	17,5						

	A	B	C
1		grupa I	grupa II
2		145	115
3		125	150
4		130	100
5		155	180
6		140	140
7		150	165
8		135	130
9		średnia arytmetyczna	=ŚREDNIA(B2:B8)
10		odchylenie standardowe	=ODCH.STANDARD.POPUL(B2:B8)
11		mediana	=MEDIANA(B2:B8)
12		odchylenie ćwiartkowe	=(KWARTYL(B2:B8;3)-KWARTYL(B2:B8;1))/2

Przykład cd.

Jeśli położenie zbioru danych określamy używając średniej arytmetycznej (miara klasyczna), to do opisu rozrzutu także użyjemy miary klasycznej (odchylenie standardowe). Natomiast jeśli położenie rozkładu opisujemy używając miar pozycyjnych (mediana, dominanta), to rozrzut opiszemy za pomocą odchylenia ćwiartkowego (miara pozycyjna).

Obliczenia w pakiecie STATISTICA

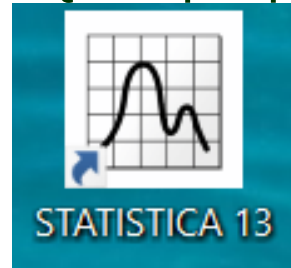


Przykład. (plik biegusy.xlsx) Zebrano dane o pomiarach 25 dorosłych i 38 młodych biegusów płaskodziobych. Należy przedstawić statystyki opisowe dla pomiaru skrzydła (kolumna SKRZ) oddzielnie dla obu grup wiekowych.

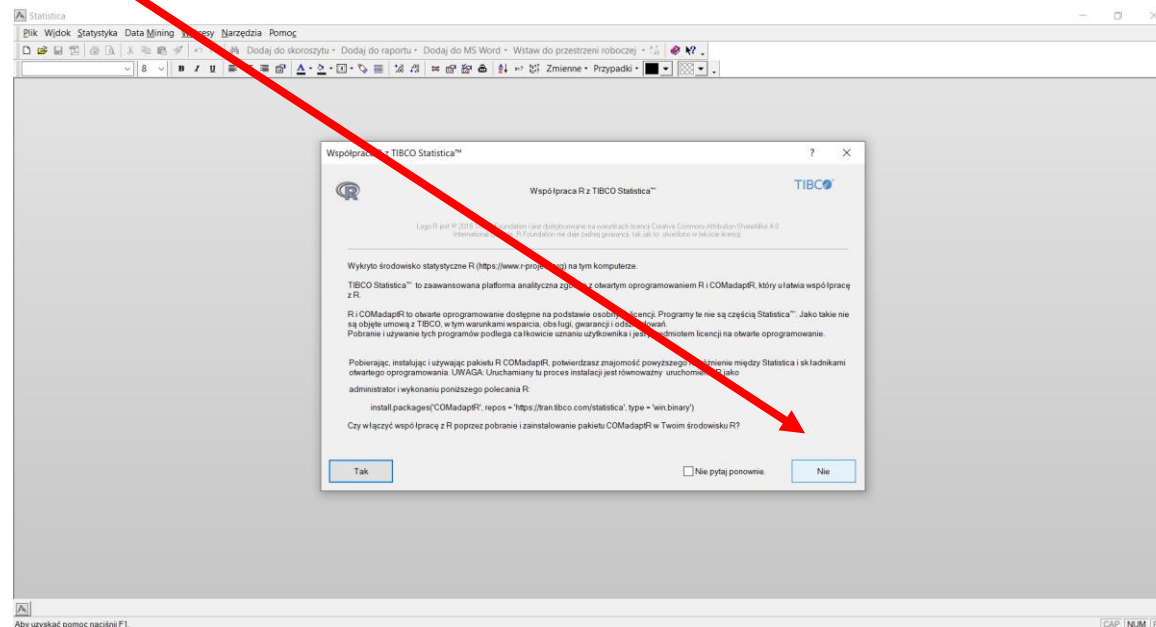
	A	B	C	D	E	F	G	H	I	J	K	
1	WIEK	DCG	DZ	SKOK	SKRZ	CIEZ						
2	1	50	28,7	21,5	107	30						
3	1	55,3	32,6	21,4	108	31						
4	1	53	31	22,1	108	32						
5	1	55,1	33,7	23	105	31						
6	1	54,2	33	109	115	33						
7	1	53,3	30,3	22,8	108	33						
8	1	51	28,5	22,8	107	42						
9	1	48,5	27,5	21,9	105	26						
10	1	54,3	32,6	21,2	108	33						
11	1	54,1	32,7	22	113	32						
12	1	53,8	32,2	21,2	110	44						
13	1	50,7	30	22,2	107	28						
14	1	52,2	31,7	22,7	111	40						

Obliczenia w pakiecie STATISTICA

Uruchamiamy pakiet (np. ikoną na pulpicie):

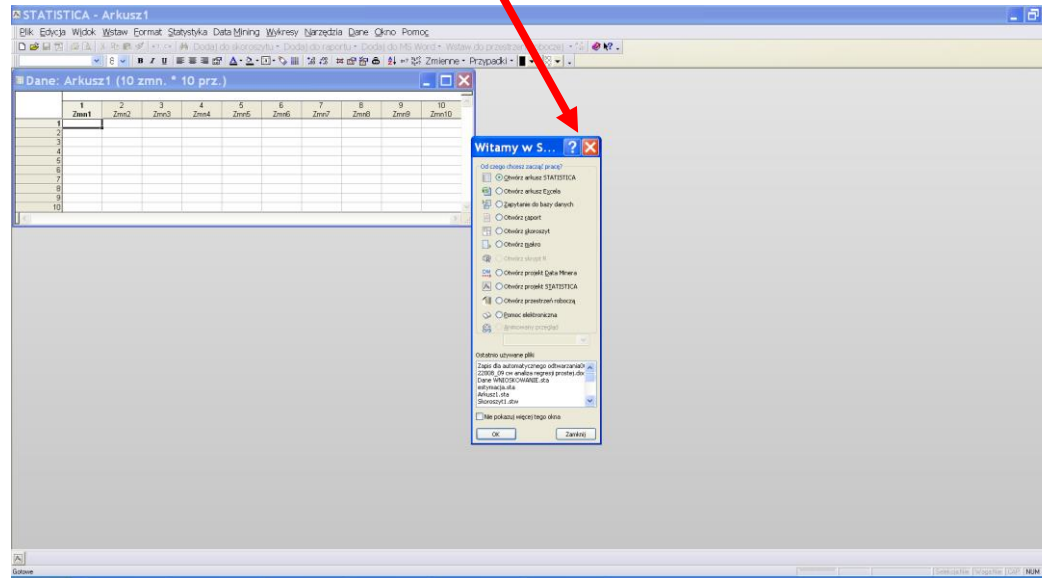


Jeśli nie zamierzamy współpracować ze środowiskiem R, zaznaczamy „Nie”



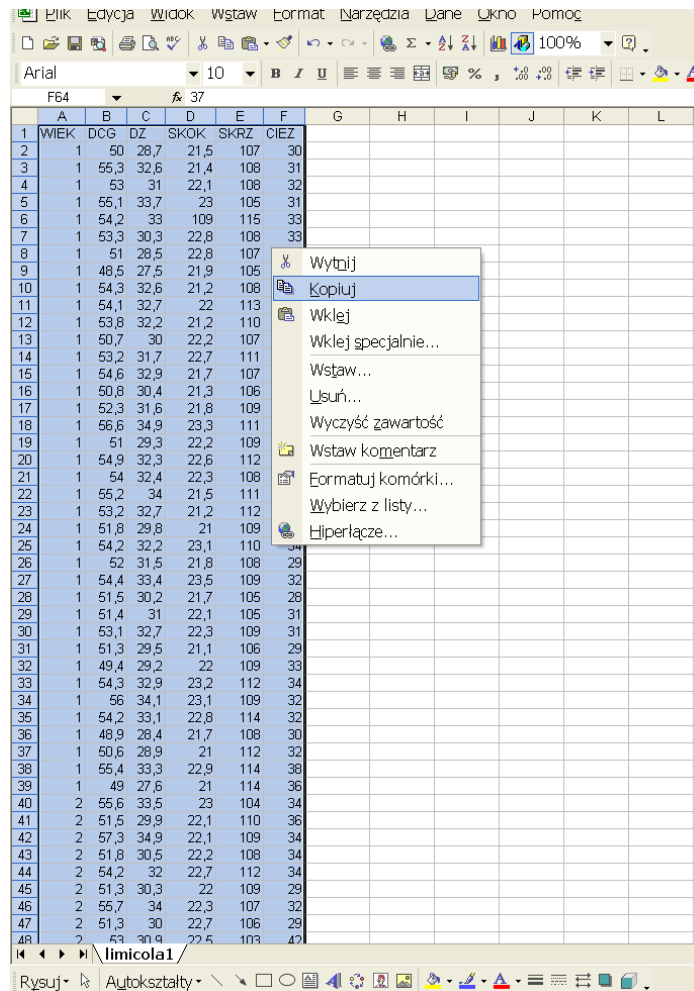
Obliczenia w pakiecie STATISTICA

Zamykamy okienko powitalne:



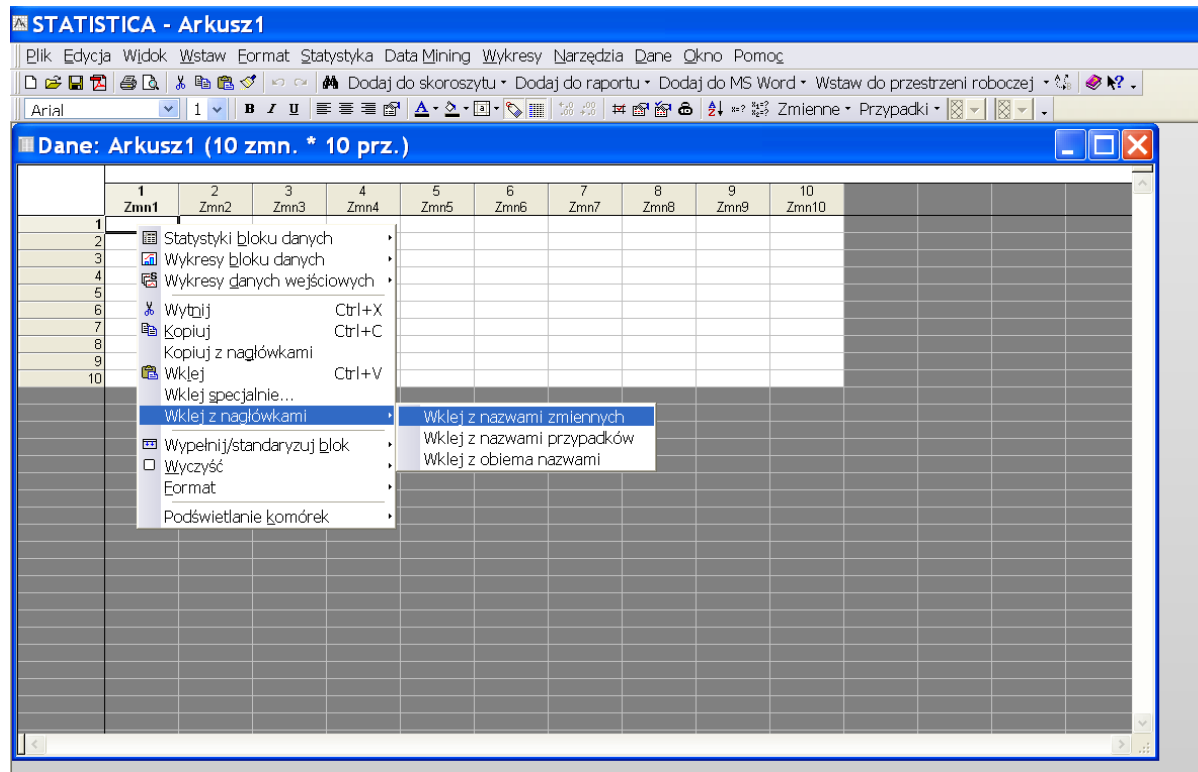
Obliczenia w pakiecie STATISTICA

Można otworzyć plik formatu xlsx (także innych) w pakiecie Statistica lub skopiować dane z pliku xlsx do pakietu. Zaznaczamy dane w arkuszu, wybieramy **Kopiuj** z menu podręcznego:



Obliczenia w pakiecie STATISTICA

Przechodzimy do pakietu, w pierwszej komórce pierwszej kolumny (zmiennej) klikamy prawym przyciskiem myszy, wybieramy **Wklej z nagłówkami/Wklej z nazwami zmiennych**



Obliczenia w pakiecie STATISTICA

plik Edycja Wzrost Wstaw Format Statystyka Data Planing Wykresy Narzędzia Dane Okno Pomoc

Dodaj do skoroszytu Dodaj do raportu Dodaj do MS Word Wstaw do przestrzeni roboczej

Arial 1 B I U

Dane: Arkusz4 (10 zm. * 63 prz.)

	1	2	3	4	5	6	7	8	9	10
	WIEK	DCG	DZ	SKOK	SKRZ	CIEZ	Zmn7	Zmn8	Zmn9	Zmn10
1	1	50	28,7	21,5	107	30				
2	1	55,3	32,6	21,4	108	31				
3	1	53	31	22,1	108	32				
4	1	55,1	33,7	23	105	31				
5	1	54,2	33	109	115	33				
6	1	53,3	30,3	22,8	108	33				
7	1	51	28,5	22,8	107	42				
8	1	48,5	27,5	21,9	105	26				
9	1	54,3	32,6	21,2	108	33				
10	1	54,1	32,7	22	113	32				
11	1	53,8	32,2	21,2	110	44				
12	1	50,7	30	22,2	107	28				
13	1	53,2	31,7	22,7	111	40				
14	1	54,6	32,9	21,7	107	37				
15	1	50,8	30,4	21,3	106	39				
16	1	52,3	31,6	21,8	109					
17	1	56,6	34,9	23,3	111	36				
18	1	51	29,3	22,2	109	44				
19	1	54,9	32,3	22,6	112	35				
20	1	54	32,4	22,3	108	45				
21	1	55,2	34	21,5	111	48				
22	1	53,2	32,7	21,2	112	46				
23	1	51,8	29,8	21	109	30				
24	1	54,2	32,2	23,1	110	34				
25	1	52	31,5	21,8	108	29				
26	1	54,4	33,4	23,5	109	32				
27	1	51,5	30,2	21,7	105	28				
28	1	51,4	31	22,1	105	31				
29	1	53,1	32,7	22,3	109	31				
30	1	51,3	29,5	21,1	106	29				
31	1	49,4	29,2	22	109	33				
32	1	54,3	32,9	23,2	112	34				
33	1	56	34,1	23,1	109	32				
34	1	54,2	33,1	22,8	114	32				
35	1	48,9	28,4	21,7	108	30				
36	1	50,6	28,9	21	112	32				
37	1	55,4	33,3	22,9	114	38				
38	1	49	27,6	21	114	36				
39	2	55,6	33,5	23	104	34				
40	2	51,5	29,9	22,1	110	36				
41	2	57,3	34,9	22,1	109	34				
42	2	51,8	30,5	22,2	108	34				
43	2	54,2	32	22,7	112	34				
44	2	51,3	30,3	22	109	29				
45	2	55,7	34	22,3	107	32				

Aby uzyskać pomoc naciśnij F1.

Obliczenia w pakiecie STATISTICA

Wybieramy z menu opcję **Statystyka/Statystyki podstawowe i tabele/Statystyki opisowe/OK**

The screenshot displays the STATISTICA software interface. The main menu is open, showing the path: **Statystyka / Statystyki podstawowe i tabele / Statystyki opisowe / OK**. The background shows a data table with columns labeled 'WIEK' and 'Zmn7' through 'Zmn10'. The data table contains the following values:

	WIEK	Zmn7	Zmn8	Zmn9	Zmn10
1	1	56,6	34,9	23,3	111
2	1	51	29,3	22,2	109
3	1	54,9	32,3	22,6	112
4	1	54	32,4	22,3	108
5	1	55,2	34	21,5	111
6	1	53,2	32,7	21,2	112
7	1	51,8	29,8	21	109
8	1	54,2	32,2	23,1	110
9	1	52	31,5	21,8	108
10	1	54,4	33,4	23,5	109
11	1	51,5	30,2	21,7	105
12	1	51,4	31	22,1	105
13	1	53,1	32,7	22,3	109
14	1	51,3	29,5	21,1	106
15	1	49,4	29,2	22	109
16	1	54,3	32,9	23,2	112
17	1	56	34,1	23,1	109
18	1	54,2	33,1	22,8	114
19	1	48,9	28,4	21,7	108
20	1	50,6	28,9	21	112
21	1	55,4	33,3	22,9	114
22	1	49	27,6	21	114
23	2	55,6	33,5	23	104
24	2	51,5	29,9	22,1	110
25	2	57,3	34,9	22,1	109
26	2	51,8	30,5	22,2	108
27	2	54,2	32	22,7	112
28	2	51,3	30,3	22	109
29	2	55,7	34	22,3	107

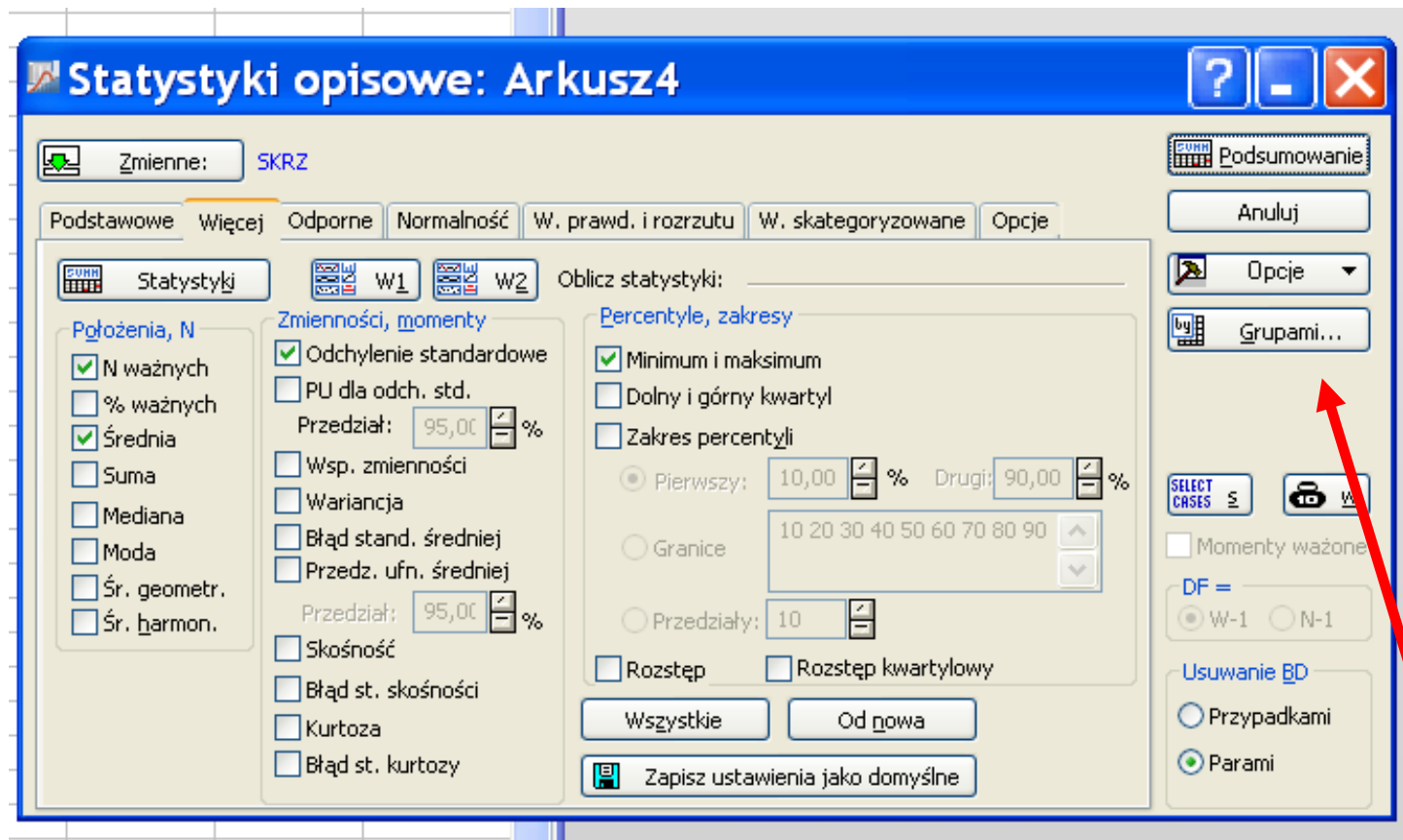
Obliczenia w pakiecie STATISTICA

W okienku **Statystyki opisowe** wybieramy przycisk **Zmienna** i zaznaczamy zmienną **SKRZ**, przyciskamy **OK**.

The screenshot displays the STATISTICA software interface. The main window shows a data table with columns labeled WIEK, DCG, DZ, SKOK, SKRZ, and CIEZ. Overlaid on this is the 'Statystyki opisowe: Arkusz4' dialog box. Within this dialog, the 'Zmienna' button is highlighted with a red arrow. A sub-dialog box titled 'Wybierz zmienną' is also open, showing a list of variables where '5 - SKRZ' is selected. Another red arrow points to the 'OK' button in this sub-dialog. The background data table is as follows:

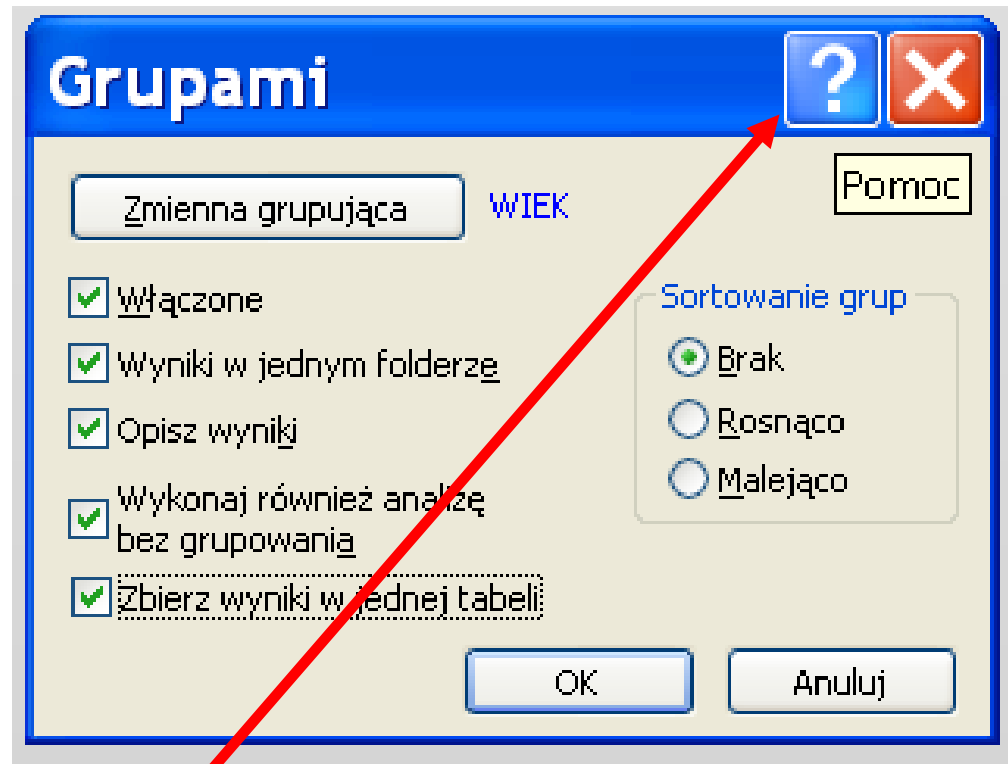
	1	2	3	4	5	6	7	8	9	10
	WIEK	DCG	DZ	SKOK	SKRZ	CIEZ	Zmn7	Zmn8	Zmn9	Zmn10
1	1	50	28,7	21,5	107	30				
2	1	55,3	32,6	21,4	108	31				
3	1	53	31	22,1	108	32				
4	1	55,1	33,7	23	105	31				
5	1	54,2	33	109	115	33				
6	1	53,3	30,3	22,8	108	33				
7	1	51	28,5	22,8	107	42				
8	1	48,5	27,5	21,9	105	26				
9	1	54,3	32,6	21,2	108	33				
10	1	54,1	32,7	22	113	32				
11	1	53,8	32,2	21,2	110	44				
12	1	50,7	30	22,2	107	28				
13	1	53,2	31,7	22,7	111	40				
14	1	54,6	32,9	21,7	107	37				
15	1	50,8	30,4	21,3	106	39				
16	1	52,3	31,6	21,8	109					
17	1	56,6	34,9	23,3	111	36				
18	1	51	29,3	22,2	109	44				
19	1	54,9	32,3	22,6	112	35				
20	1	54	32,4	22,3	108	45				
21	1	55,2	34	21,5	111	48				
22	1	53,2	32,7	21,2	112	46				
23	1	51,8	29,8	21	109	30				
24	1	54,2	32,2	23,1	110	34				
25	1	52	31,5	21,8	108	29				
26	1	54,4	33,4	23,5	109	32				
27	1	51,5	30,2	21,7	105	28				
28	1	51,4	31	22,1	105	31				
29	1	53,1	32,7	22,3	109	31				
30	1	51,3	29,5	21,1	106	29				
31	1	49,4	29,2	22	109	33				
32	1	54,3	32,9	23,2	112	34				
33	1	56	34,1	23,1	109	32				
34	1	54,2	33,1	22,8	114	32				
35	1	48,9	28,4	21,7	108	30				
36	1	50,6	28,9	21	112	32				
37	1	55,4	33,3	22,9	114	38				
38	1	49	27,6	21	114	36				
39	2	55,6	33,5	23	104	34				
40	2	51,5	29,9	22,1	110	36				
41	2	57,3	34,9	22,1	109	34				
42	2	51,8	30,5	22,2	108	34				
43	2	54,2	32	22,7	112	34				
44	2	51,3	30,3	22	109	29				
45	2	55,7	34	22,3	107	32				

Obliczenia w pakiecie STATISTICA



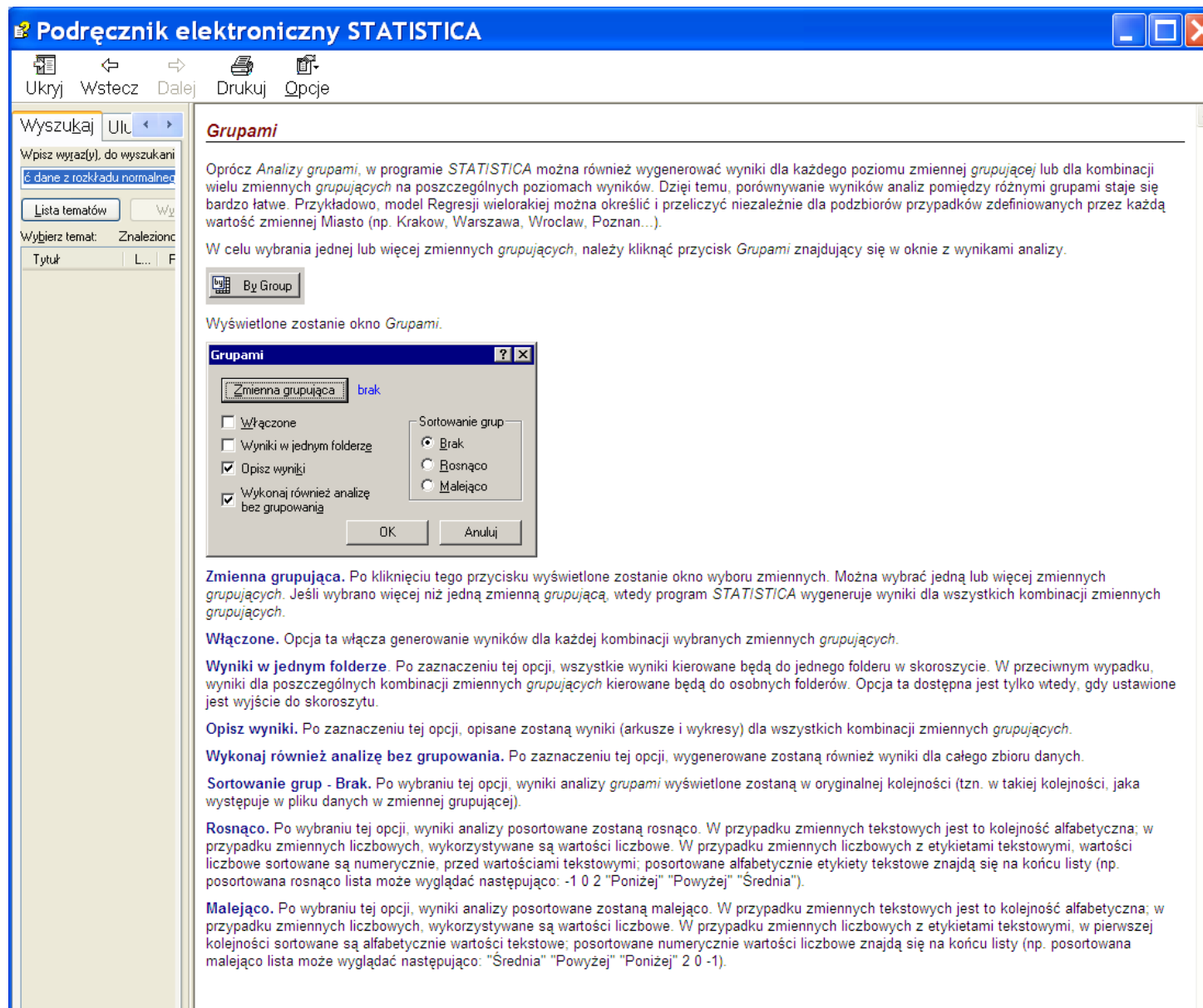
W oknie Statystyki opisowe, po prawej stronie przycisk **Grupami...** umożliwia zdefiniowanie podziału na grupy. Jako zmienną grupującą zaznaczamy WIEK. Zaznaczamy opcję **Wyniki w jednym folderze**.

Obliczenia w pakiecie STATISTICA



W razie niejasności, można kliknąć pytajnik, który otwiera pomoc ekranową.

Obliczenia w pakiecie STATISTICA



Grupami

Oprócz *Analizy grupami*, w programie *STATISTICA* można również wygenerować wyniki dla każdego poziomu zmiennej *grupującej* lub dla kombinacji wielu zmiennych *grupujących* na poszczególnych poziomach wyników. Dzięki temu, porównywanie wyników analiz pomiędzy różnymi grupami staje się bardzo łatwe. Przykładowo, model Regresji wielorakiej można określić i przeliczyć niezależnie dla podzbiorów przypadków zdefiniowanych przez każdą wartość zmiennej *Miasto* (np. Krakow, Warszawa, Wrocław, Poznan...).

W celu wybrania jednej lub więcej zmiennych *grupujących*, należy kliknąć przycisk *Grupami* znajdujący się w oknie z wynikami analizy.

Wyświetlone zostanie okno *Grupami*.

Zmienna grupująca brak

Włączone

Wyniki w jednym folderze

Opisz wyniki

Wykonaj również analizę bez grupowania

Sortowanie grup

Brak

Rosnąco

Malejąco

OK Anuluj

Zmienna grupująca. Po kliknięciu tego przycisku wyświetlone zostanie okno wyboru zmiennych *grupujących*. Jeśli wybrano więcej niż jedną zmienną *grupującą*, wtedy program *STATISTICA* wygeneruje wyniki dla wszystkich kombinacji zmiennych *grupujących*.

Włączone. Opcja ta włącza generowanie wyników dla każdej kombinacji wybranych zmiennych *grupujących*.

Wyniki w jednym folderze. Po zaznaczeniu tej opcji, wszystkie wyniki kierowane będą do jednego folderu w skoroszytcie. W przeciwnym wypadku, wyniki dla poszczególnych kombinacji zmiennych *grupujących* kierowane będą do osobnych folderów. Opcja ta dostępna jest tylko wtedy, gdy ustawione jest wyjście do skoroszytu.

Opisz wyniki. Po zaznaczeniu tej opcji, opisane zostaną wyniki (arkusze i wykresy) dla wszystkich kombinacji zmiennych *grupujących*.

Wykonaj również analizę bez grupowania. Po zaznaczeniu tej opcji, wygenerowane zostaną również wyniki dla całego zbioru danych.

Sortowanie grup - Brak. Po wybraniu tej opcji, wyniki analizy *grupami* wyświetlone zostaną w oryginalnej kolejności (tzn. w takiej kolejności, jaka występuje w pliku danych w zmiennej *grupującej*).

Rosnąco. Po wybraniu tej opcji, wyniki analizy posortowane zostaną rosnąco. W przypadku zmiennych tekstowych jest to kolejność alfabetyczna; w przypadku zmiennych liczbowych, wykorzystywane są wartości liczbowe. W przypadku zmiennych liczbowych z etykietami tekstowymi, wartości liczbowe sortowane są numerycznie, przed wartościami tekstowymi; posortowane alfabetycznie etykiety tekstowe znajdują się na końcu listy (np. posortowana rosnąco lista może wyglądać następująco: -1 0 2 "Ponizej" "Powyzej" "Średnia").

Malejąco. Po wybraniu tej opcji, wyniki analizy posortowane zostaną malejąco. W przypadku zmiennych tekstowych jest to kolejność alfabetyczna; w przypadku zmiennych liczbowych, wykorzystywane są wartości liczbowe. W przypadku zmiennych liczbowych z etykietami tekstowymi, w pierwszej kolejności sortowane są alfabetycznie wartości tekstowe; posortowane numerycznie wartości liczbowe znajdują się na końcu listy (np. posortowana malejąco lista może wyglądać następująco: "Średnia" "Powyzej" "Ponizej" 2 0 -1).

Obliczenia w pakiecie STATISTICA

W zakładce **Więcej** zaznaczamy parametry, które mają być obliczone. Po kliknięciu przycisku **Podsumowanie** wyniki są zestawione w skoroszytcie.

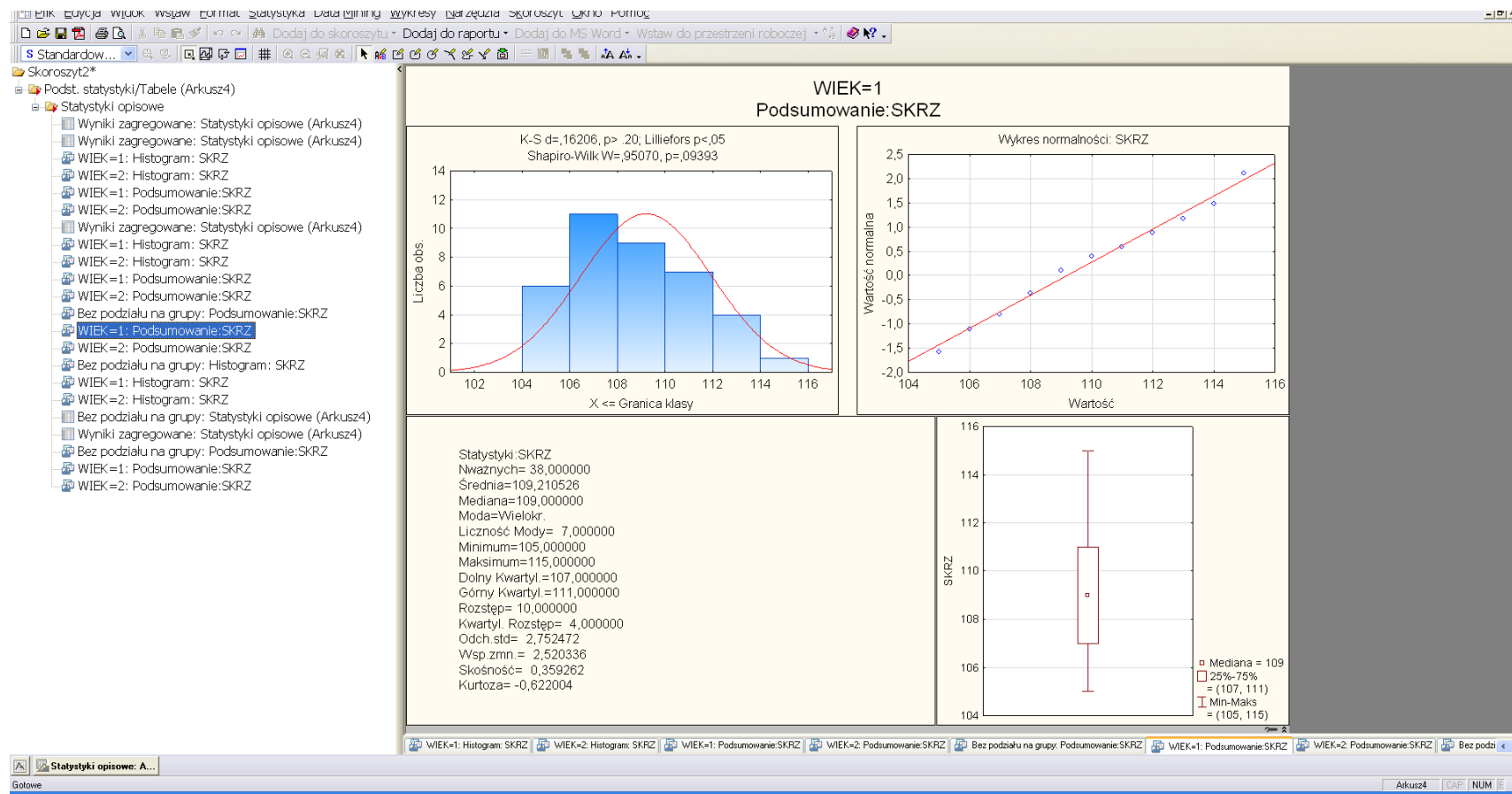
The screenshot displays the STATISTICA software interface. The main window, titled "Skoroszyt2* - Wyniki zagregowane: Statystyki opisowe (Arkus4)", shows a summary report for two variables (SKRZ). The report includes a table with the following data:

Zmienna	WIEK	Nważnych	Średnia	Mediana	Liczność Mody	Minimum	Maksimum	Dolny Kwartyl	Górnym Kwartyl	Rozstęp	Kwartyl. Rozstęp	Odch. std	Wsp. zmn.	Skośność	Kurtoza
SKRZ	1	38	109,2105	109,0000	7	105,0000	115,0000	107,0000	111,0000	10,00000	4,000000	2,752472	2,520336	0,359262	-0,622004
SKRZ	2	25	106,5600	106,0000	5	102,0000	115,0000	105,0000	108,0000	13,00000	3,000000	3,026751	2,842296	1,016750	1,178666

The bottom toolbar contains a button labeled "Statystyki opisowe: A..." which is circled in red. The status bar at the bottom indicates "Gotowe" and "Arkus4".

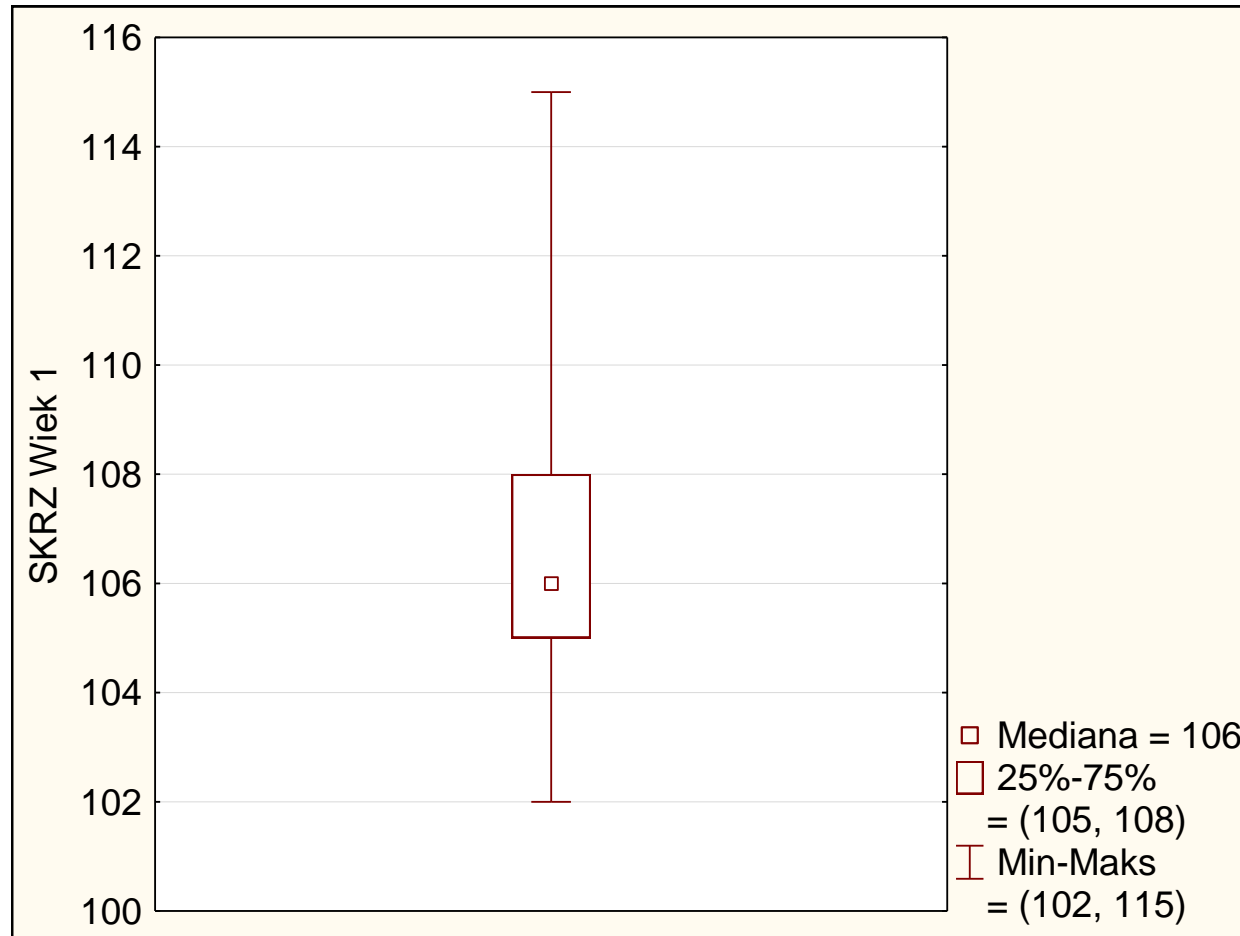
Obliczenia w pakiecie STATISTICA

Można obejrzeć histogramy rozkładu analizowanej zmiennej po kliknięciu na przycisk **Wykresy** na karcie **Więcej** lub **Histogramy** na karcie **Normalność**. W tym drugim przypadku możemy ustalić liczbę przedziałów.



Obliczenia w pakiecie STATISTICA

Wykres ramka-wąsy, pudełko z wąsami (box-and-whisker plot)



Dostęp do pakietu STATISTICA

Opis dostępu do pakietu Statistica (i innych) znajduje się na stronie Centrum Informatycznego. Można dostać się tam wchodząc na stronę uczelni **sggw.pl**, przewijając ją do samego dołu:



lub kontaktując się z (sekretariatem?) osobiście:

Kontakt

Centrum Informatyczne SGGW w Warszawie
ul. Nowoursynowska 166, 02-787 Warszawa
budynek nr 1

Sekretariat (pokój nr 19):
e-mail: ci@sggw.pl
tel.: +48 22 59 355 40
fax.: +48 22 59 355 41

Pomoc Informatyczna (pokój nr 1):
[Elektroniczny System Obsługi Zgłoszeń](#)
e-mail: pomoc_ci@sggw.pl
tel.: 22 59 355 50

Zamówienia publiczne IT (pokój nr 11):
tel.: 22 59 355 62