

2. GRAFICZNA PREZENTACJA DANYCH

2.1. Histogram

Tabela wyników pomiarów

Najprostszym metodą gromadzenia danych jest tabela wyników pomiarów, w tabeli tej dane mogą być dowolnie zorganizowane (np. posortowane rosnąco).

	A	B	C	D	E	F	G	H	I	J	K
1		Lp.	Pomiary		Lp.	Pomiary					
2		(nieuporządkowane)			(uporządkowane rosnąco)						
3		1	21,6		1	19,5					
4		2	21,2		2	20					
5		3	20		3	20,9					
6		4	19,5		4	21,2					
7		5	20,9		5	21,6					
8		6	24		6	21,6					
9		7	22		7	22					
10		8	21,6		8	22					
11		9	22		9	22,6					
12		10	22,6		10	24					

Duże ilości danych mogą być prezentowane w postaci szeregu rozdzielczego (wartości dzielone na klasy (kategorie), dla każdej klasy podawana jest jej *liczebność* lub *częstość*).

<i>długość</i>	($-\infty$ 20]	(20 22]	(22 24]
<i>liczebność</i>	2	6	2
<i>częstość</i>	0.2	0.6	0.2

Przykład 1. Przedstaw wyniki widocznych powyżej pomiarów w postaci szeregu rozdzielczego przyjmując następujące klasy: ($-\infty$ 19], (19, 20], (20, 21], (21, 22], (22, 23], (23, $+\infty$).

	A	B	C	D	E	F	G	H
13	rozmiar próby	10						
14								
15		klasy	liczebności					
16		19	{=CZĘSTOŚĆ(C3:C12;B16:B20)}					
17		20						
18		21						
19		22						
20		23						
21		powyżej						



W MS Excel 2003 jest funkcja obliczająca częstość występowania wartości w określonym zakresie.

CZĘSTOŚĆ(liczby; klasy)

gdzie: **liczby** – obszar z danymi, **klasy** – pionowy zakres komórek zawierających granice przedziałów klasowych, zliczanie odbywa się w przedziałach prawostronnie domkniętych.

Funkcja **CZĘSTOŚĆ** zwraca tablicę obliczonych częstości – musi być więc wprowadzona jako formuła tablicowa (CTRL+SHIFT+ENTER). Tablica obliczanych przez funkcję wartości może zawierać o jedną komórkę więcej niż tablica wskazana w parametrze klasy – w takim przypadku ostatnia komórka zawiera ilość wartości większych od granicy ostatniego przedziału klasowego.

Histogram

Histogram to wykres słupkowy ilustrujący rozkład prawdopodobieństwa określonej cechy, jest konstruowany na podstawie danych szeregu rozdzielczego. Podstawę słupków stanowią przedziały klasowe szeregu, przyjmuje się że liczba tych przedziałów powinna być równa pierwiastkowi z liczby pomiarów, zwykle rysowane są histogramy o 6 do 12 przedziałach. Wysokość słupków jest ustalana dla określonej wartości lub przedziału na podstawie:

- *liczebności* (histogram liczebności),
- *częstości* (histogram częstości),
- *skumulowanych liczebności* (histogram liczebności skumulowanych),
- *skumulowanych częstości* (histogram częstości skumulowanych).

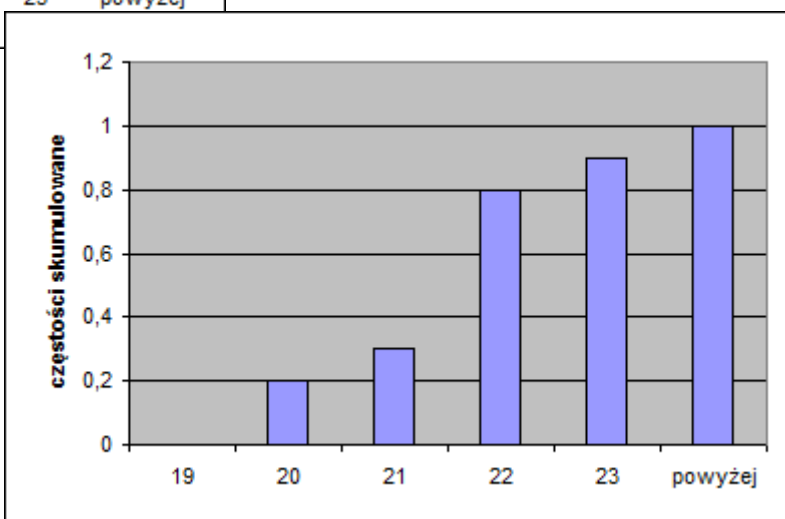
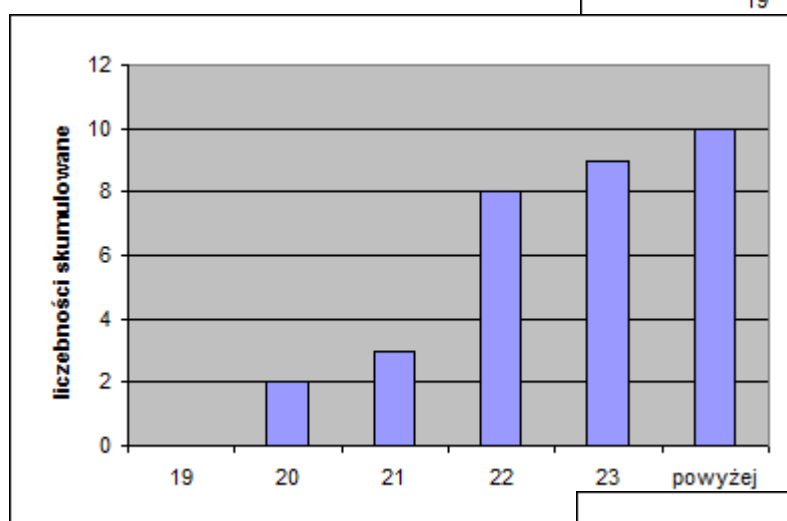
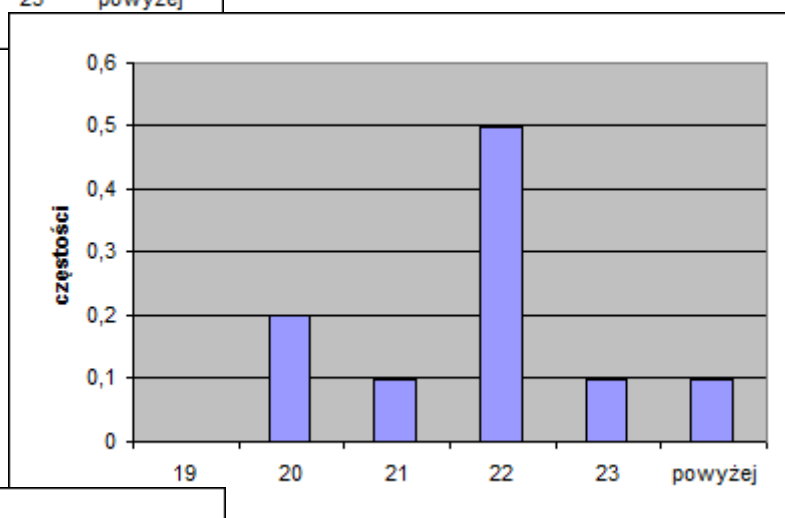
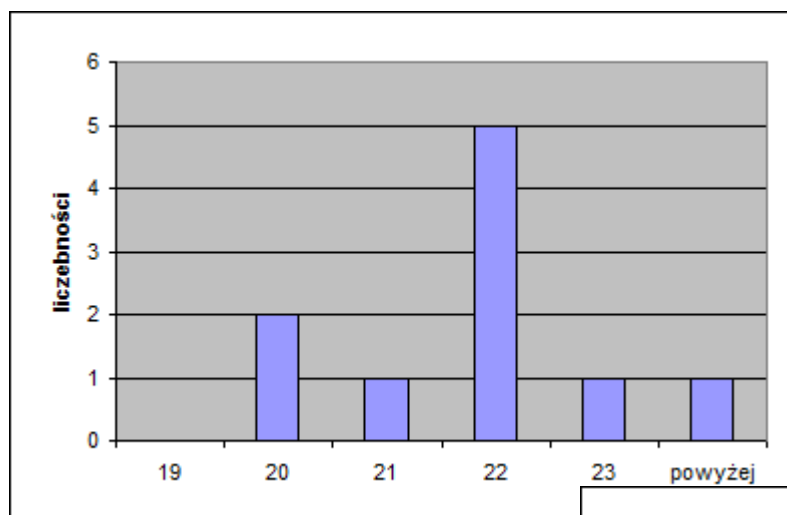
Przykład 2. Dla wyników pomiarów z przykładu 1. wyznacz liczebności i częstości skumulowane.

	A	B	C	D	E	F	G
13	rozmiar próby	10					
14							
15	klasy	liczebności	częstości	liczebności skumulowane	częstości skumulowane		
16	19	0	=B16/\$B\$13	=C16	=SUMA(\$D\$16:D16)		
17	20	2	=B17/\$B\$13	=C17+E16	=SUMA(\$D\$16:D17)		
18	21	1	=B18/\$B\$13	=C18+E17	=SUMA(\$D\$16:D18)		
19	22	5	=B19/\$B\$13	=C19+E18	=SUMA(\$D\$16:D19)		
20	23	1	=B20/\$B\$13	=C20+E19	=SUMA(\$D\$16:D20)		
21	powyżej	1	=B21/\$B\$13	=C21+E20	=SUMA(\$D\$16:D21)		

Uwaga! Na czerwono zostały zaznaczone formuły wyliczające wartości skumulowane – formuły te pokazują dwa różne podejścia do wyznaczania tego typu wartości. Przy obliczaniu liczebności skumulowanych wykorzystano fakt, że w komórce powyżej znajduje się poprzednio obliczona wartość sumy. Przy obliczaniu częstości skumulowanych wykorzystana została funkcja suma z odpowiednio określonym zakresem – komórka początkowa obszaru zaadresowana jest bezwzględnie a komórka końcowa względnie. Obydwie formuły mogą zostać skopiowane do komórek znajdujących się poniżej wyliczając w ten sposób wartości skumulowane dla wszystkich przedziałów klasowych.

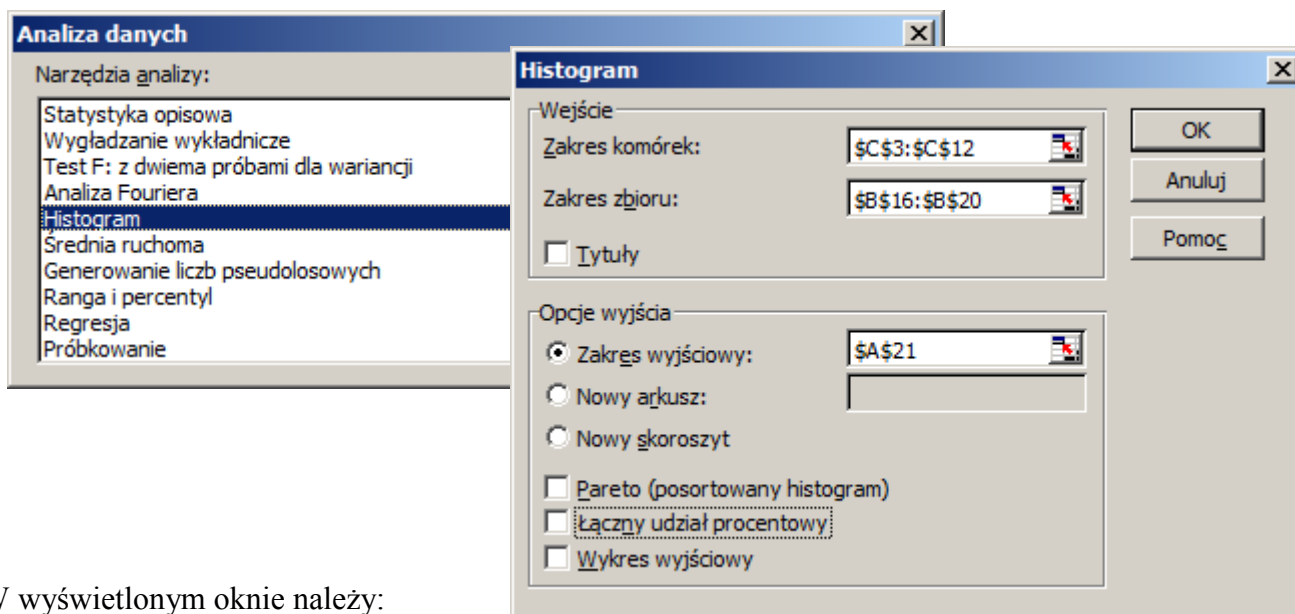
Po wyznaczeniu wartości wystarczy już tylko wygenerować odpowiedni **wykres kolumnowy**.





Dodatek Analysis ToolPak

Histogram można również narysować wykorzystując dostępną w Excelu funkcję „**Analiza danych**”. (funkcja jest dostępna z menu **Narzędzia|Analiza danych** o ile jest ona włączona – żeby włączyć funkcję należy wybrać opcję **Narzędzia|Dodatki** a następnie zaznaczyć pole wyboru: **Analysis ToolPak**).



W wyświetlonym oknie należy:

- wskazać obszar z danymi,
- wskazać obszar zawierający granice przedziałów klasowych,
- wskazać „opcje wyjścia” czyli:
 - miejsce w którym zwrócone zostaną obliczenia niezbędne do wykreślenia histogramu,
 - i opcjonalnie wybrać dodatkowe możliwości histogramu:
 - Pareto (posortowany histogram)
 - Łączny udział procentowy
 - Wykres wyjściowy.

Histogram w wersji podstawowej wyznacza liczebności we wskazanych przedziałach klasowych. Histogram Pareto dodatkowo sortuje otrzymane wyniki zaczynając od przedziałów klasowych o największej liczebności. Opcja „Łączny udział procentowy” uzupełnia obliczenia o skumulowane częstości (w %). Po wybraniu opcji „Wykres wyjściowy” generowany jest wykres kolumnowy przedstawiający uzyskane wyniki.

	A	B	C	D	E	F	G	H	I
21	Histogram			Histogram + Pareto (posortowany histogram) + łączny udział procentowy					
22	<i>Zbiór danych (koszyk)</i>	<i>Częstość</i>		<i>Zbiór danych (koszyk)</i>	<i>Częstość</i>	<i>Łączna wartość %</i>	<i>Zbiór danych (koszyk)</i>	<i>Częstość</i>	<i>Łączna wartość %</i>
23	19	0		19	0	0,00%	22	5	50,00%
24	20	2		20	2	20,00%	20	2	70,00%
25	21	1		21	1	30,00%	21	1	80,00%
26	22	5		22	5	80,00%	23	1	90,00%
27	23	1		23	1	90,00%	Więcej	1	100,00%
28	Więcej	1		Więcej	1	100,00%	19	0	100,00%



	A	B	C	D	E	F	G	H
29	Histogram + łączny udział procentowy				Histogram + Pareto (posortowany histogram)			
30	Zbiór danych (koszyk)	Częstość	Łączna wartość %		Zbiór danych (koszyk)	Częstość	Zbiór danych (koszyk)	Częstość
31	19	0	0,00%		19	0	22	5
32	20	2	20,00%		20	2	20	2
33	21	1	30,00%		21	1	21	1
34	22	5	80,00%		22	5	23	1
35	23	1	90,00%		23	1	Więcej	1
36	Więcej	1	100,00%		Więcej	1	19	0

Histogramy umożliwiają ocenę podobieństwa rozkładu empirycznego do rozkładu oczekiwanego. Aby ułatwić ocenę tego podobieństwa dobrze jest nanieść na wykresie, łącznie z histogramem, krzywą gęstości prawdopodobieństwa rozkładu oczekiwanego.

Uwaga! W MS Excel 2003 dołączenie krzywej gęstości do histogramu wymaga dodatkowych zabiegów: konieczne jest wyznaczenie danych niezbędnych do narysowania krzywej i utworzenie nowej serii danych z obliczonymi wartościami.

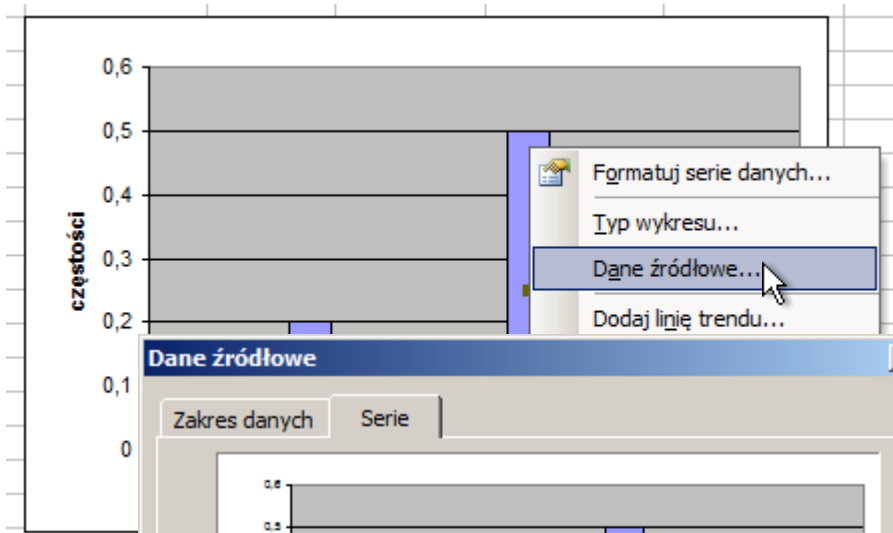
Założmy, że wyniki pomiarów mają rozkład normalny ze średnią i odchyleniem standardowym równym wartościom obliczonym z próby. W poniższym arkuszu obliczone zostały parametry oczekiwanego rozkładu normalnego i wartości funkcji gęstości w 20 punktach położonych w odległości $\pm 4\sigma$ od średniej.

	A	B	C
37	średnia	s	odstęp
38	=ŚREDNIA(C3:C12)	=ODCH.STANDARDOWE(C3:C12)	=8*B38/20
39	Lp.	x	f(x)
40	0	=\$A\$38-4*\$B\$38+A40*\$C\$38	=ROZKŁAD.NORMALNY(B40;\$A\$38;\$B\$38;FAŁSZ) * 1
41	1		
...	...		
60	20		

Uwaga! Po wyznaczeniu wartości funkcji gęstości konieczna jest jeszcze jej normalizacja umożliwiająca nałożenie krzywej na histogram. W przypadku histogramu częstości należy otrzymane wartości funkcji gęstości przemnożyć przez szerokość przedziału klasowego (w rozważanym przykładzie szerokość ta wynosi 1). W przypadku histogramu liczebności otrzymane wartości należy przemnożyć przez szerokość przedziału klasowego oraz przez liczebność próby.

Po przygotowaniu danych wystarczy tylko uzupełnić histogram częstości nakładając na niego krzywą. W tym celu w oknie z danymi źródłowymi wykresu należy dodać nową serię danych wskazując w polu „Wartości” komórki zawierające wartości funkcji gęstości.





Dane źródłowe

Zakres danych Serie

Serie

Serie1

Nazwa:

Wartości:

Dodaj **Usuń**

Etykiety osi kategorii (X):

Dane źródłowe

Zakres danych Serie

Serie

Serie1
Serie2

Nazwa:

Wartości:

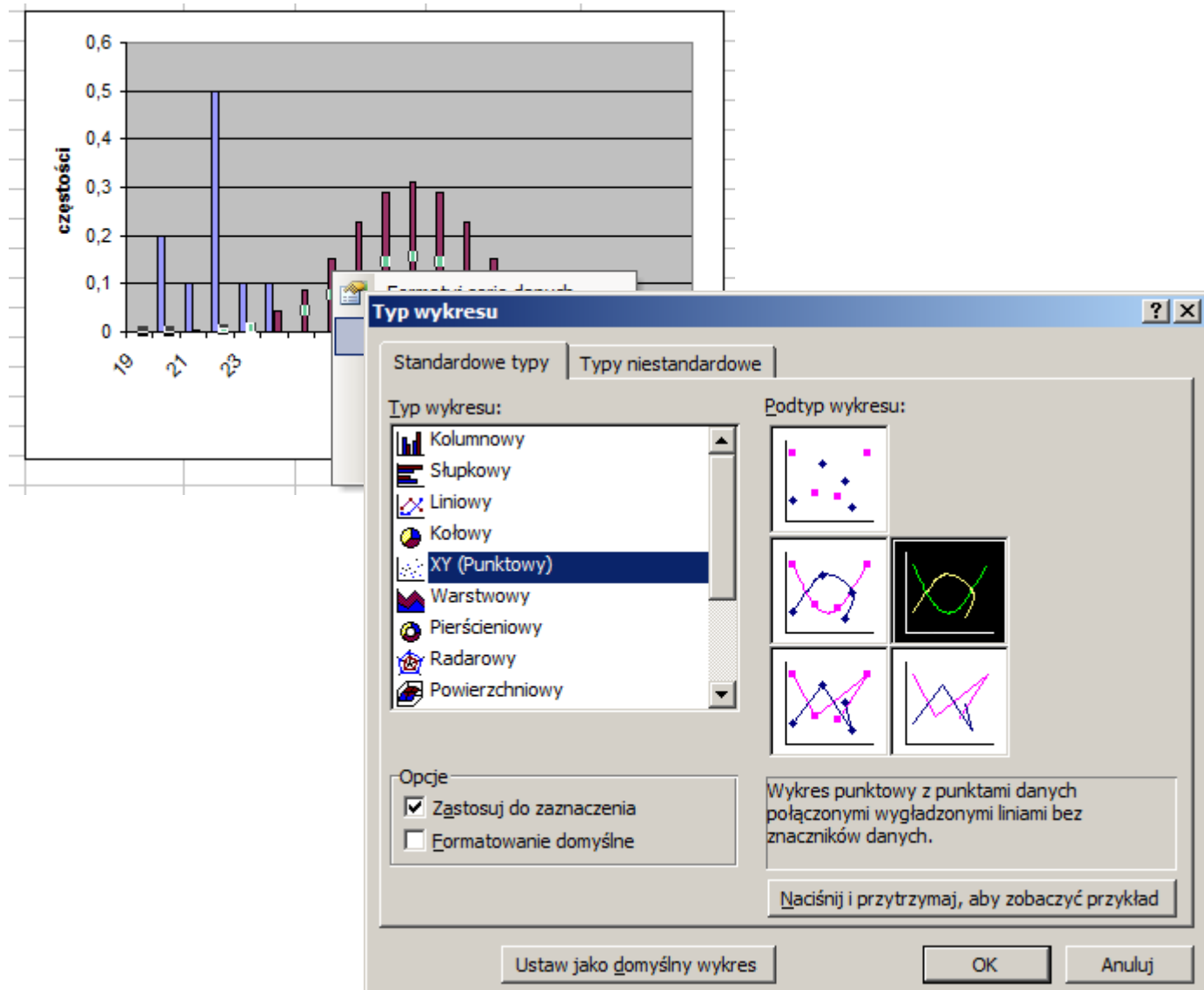
Dodaj **Usuń**

Etykiety osi kategorii (X):

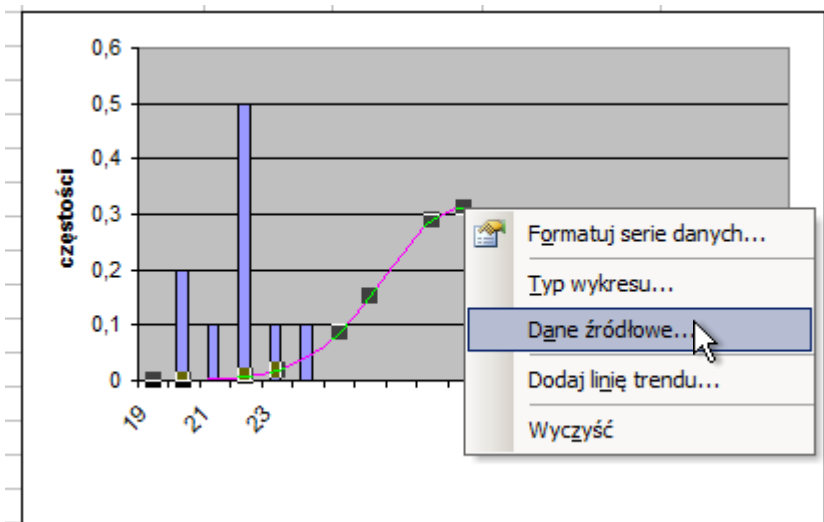
OK Anuluj

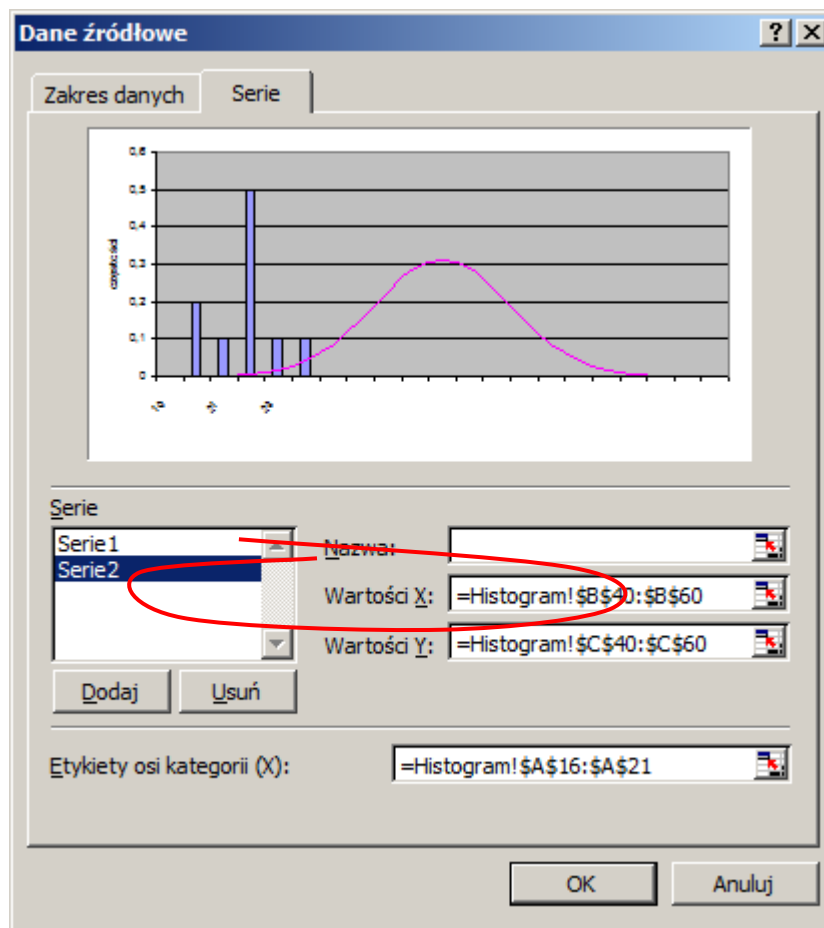


Teraz trzeba jeszcze tylko zmienić typ wykresu na „XY Punktowy”:

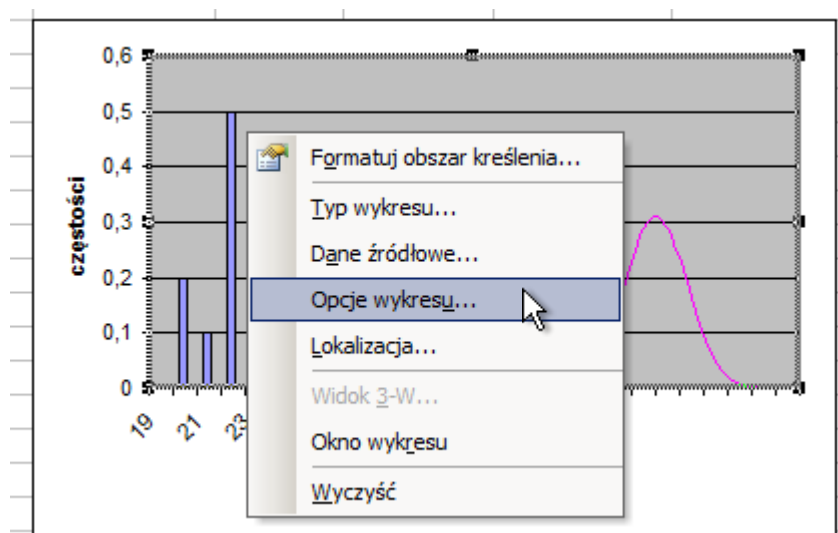


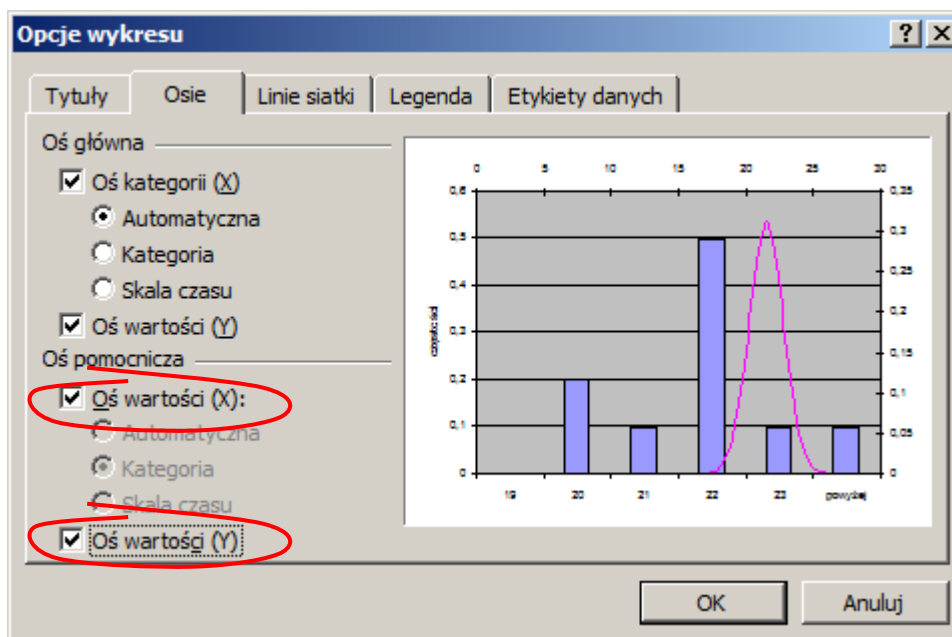
i wskazać komórki zawierające x -owe współrzędne punktów wykresu:



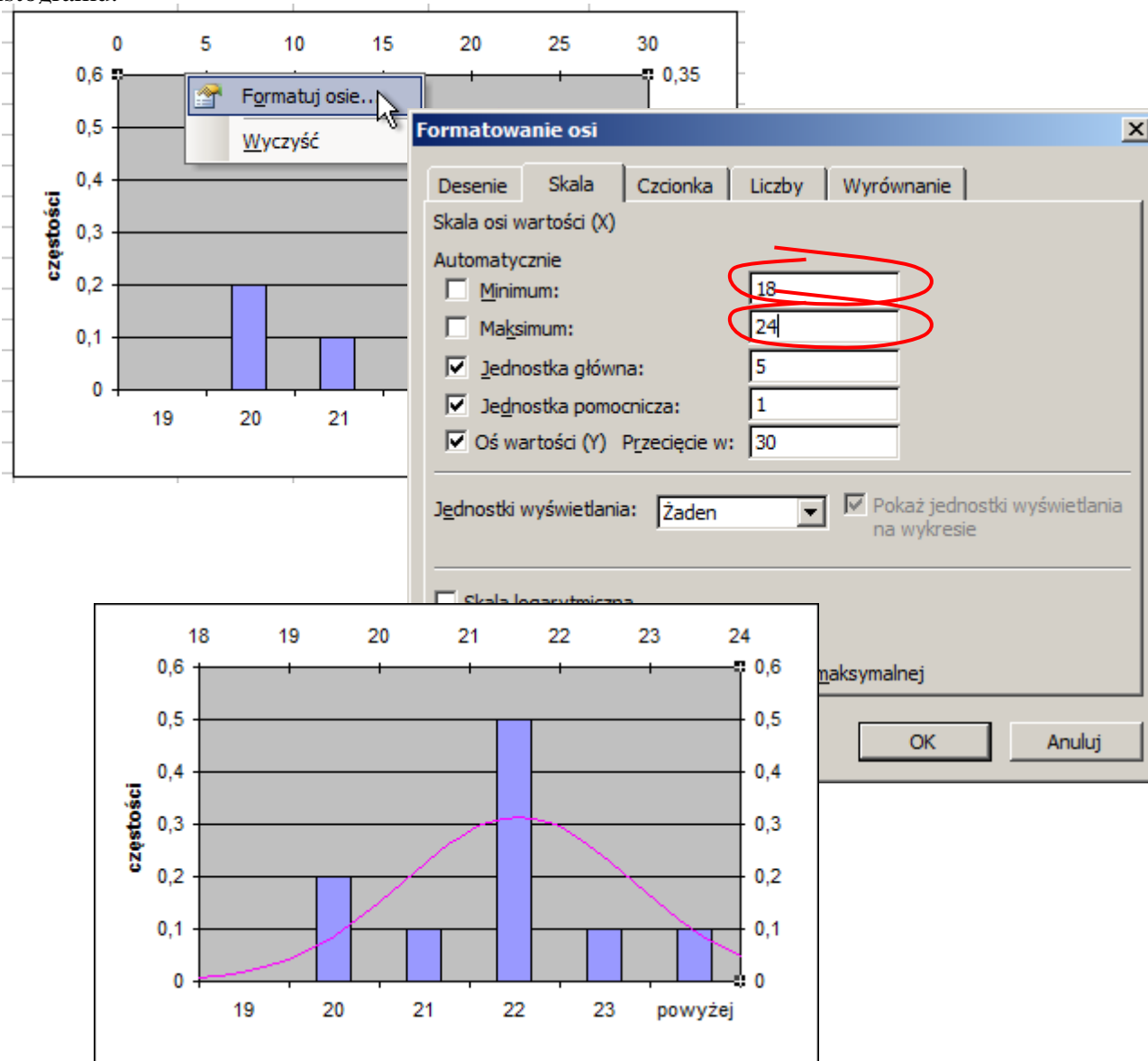


Wykres jest już prawie gotowy – wymaga jedynie wyrównania zakresów w odpowiadających sobie osiach obydwu wykresów. Najpierw należy pokazać osie pomocnicze, które są w tym przypadku osiami dla krzywej rozkładu:



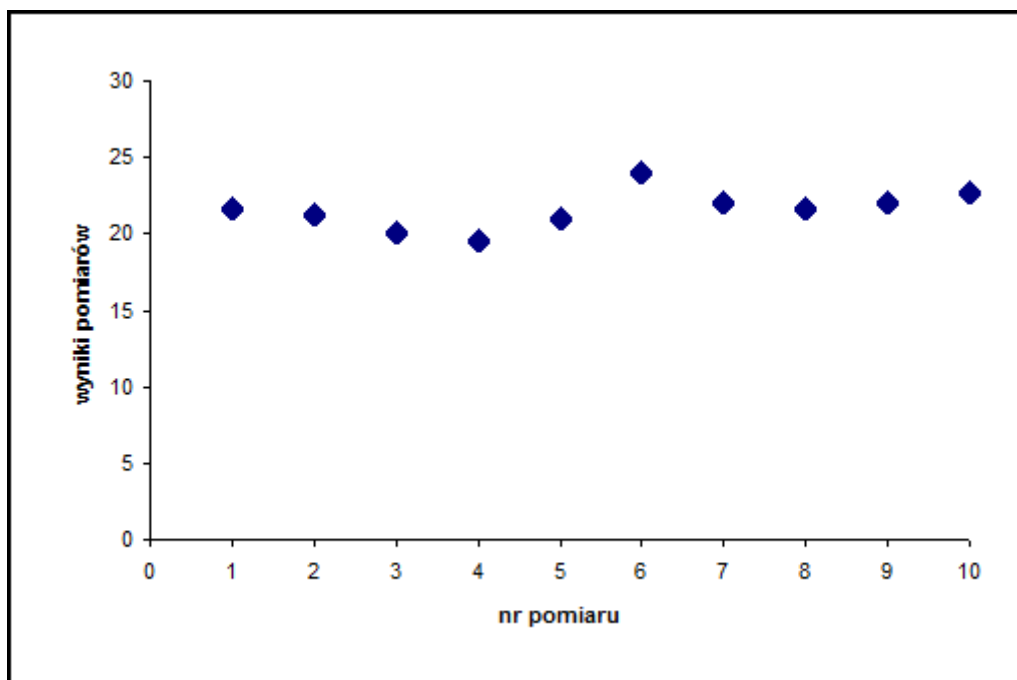


Na koniec należy ustalić taki sam zakres dla osi poziomej i pionowej dla wykresu krzywej jaki mają osie histogramu.



2.2. Wykres punktowy lub liniowy

Dziedziną wykresu jest zbiór wszystkich numerów pomiarów, przeciwdziedziną zbiór wyników pomiarów.



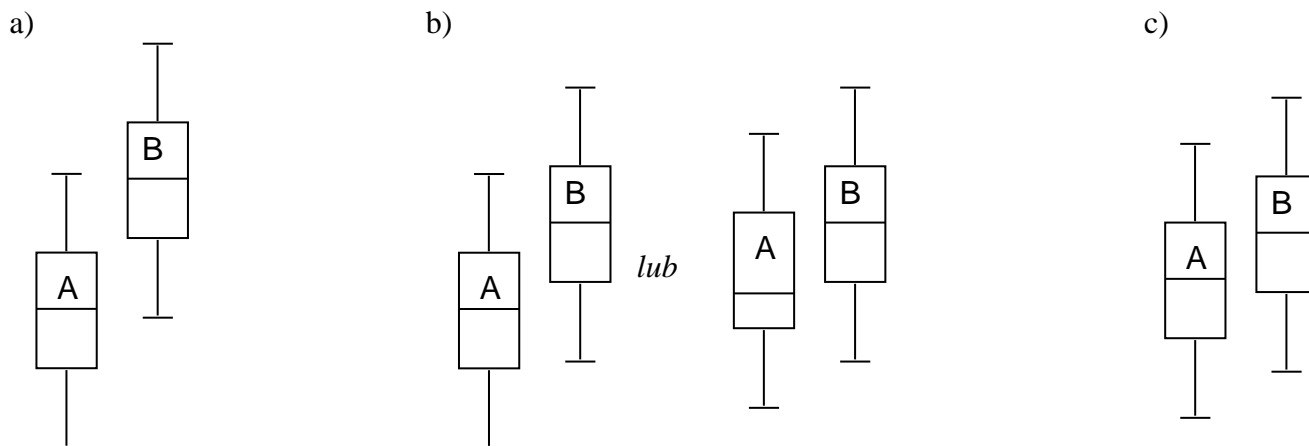
2.3. Wykres pudełkowy

Wykres pudełkowy – pozwala na ilustrację miar rozkładu jednowymiarowej zmiennej losowej. Jest rysowany w kilku odmianach:

<i>Ilustrowane miary</i>	<i>Wygląd</i>
min, max Q_1, Me, Q_3	<p>Wykres pudełkowy z etykietami: min, Q_1, Me, Q_3, max.</p>
\bar{x}, s	<p>Wykres pudełkowy z etykietami: $\bar{x} - s$, $\bar{x} - \frac{s}{\sqrt{n}}$, \bar{x}, $\bar{x} + \frac{s}{\sqrt{n}}$, $\bar{x} + s$.</p>

Wykresy pudełkowe są często rysowane dla wielu grup danych w celu ich porównania. Wyniki porównania można uogólnić na całe populacje jeśli rozmiary próbek na podstawie których narysowane były wykresy nie są mniejsze od 30.

Poniżej rozważone zostały typowe układy dwóch wykresów pudełkowych.



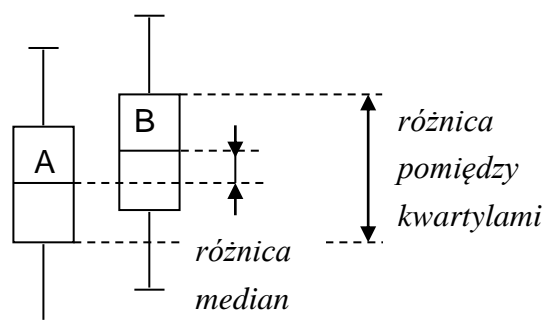
W przypadku:

- prostokąty wykresów nie nachodzą na siebie – w tej sytuacji uznaje się że ***B jest większe od A***,
- prostokąty nachodzą na siebie ale jedna z median jest poza prostokątem drugiego wykresu – w tej sytuacji wyciągany jest wniosek, że ***B jest prawdopodobnie większe od A***,
- prostokąty nachodzą na siebie, obydwie mediany są wewnątrz prostokątów sąsiedniego wykresu – w tej ***sytuacji nie można stwierdzić, że istnieje różnica pomiędzy A i B***.

W przypadku c) można wyznaczyć iloraz różnicy median do różnicy pomiędzy dalej położonym dolnym i górnym kwartylem. Jeśli wielkość tak wyznaczonego wskaźnika przekracza pewną wartość graniczną uznaje się że jest ***tendencja do różnicy pomiędzy A i B***.

Za wartość graniczną uznaje się wielkość:

- 0,33 dla próbek o rozmiarze 30,
- 0,2 dla próbek o rozmiarze 100,
- 0,1 dla próbek o rozmiarze 1000.



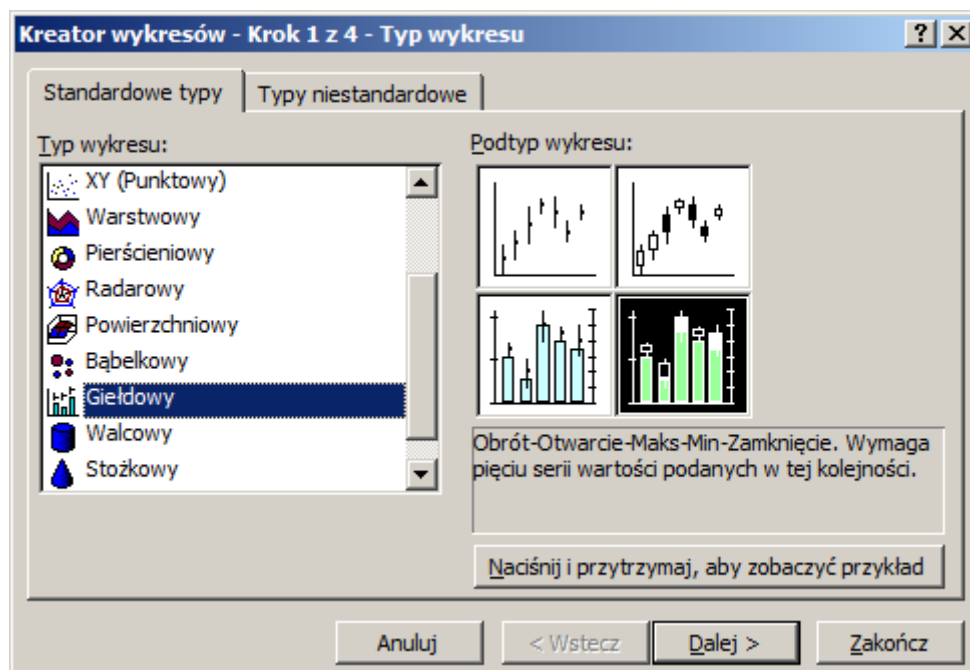
Uwaga! W MS Excel 2003 wykres pudełkowy jest niedostępny. Można go narysować na kilka różnych sposobów. Najprostsza metoda polega na wykorzystaniu wykresu giełdowego.

Wykres giełdowy jest używany do ilustrowania wahań cen akcji. Wykres ten w zależności od liczby informacji może być konstruowany w oparciu o serie danych składające się z 3, 4 i 5 wartości.

Do konstrukcji pojedynczego „pudełka” potrzebnych jest 5 wartości (np. *min*, *max*, Q_1 , *Me*, Q_3) – wykres pudełkowy można zbudować wykorzystując do tego celu wykres giełdowy typu:

Obrót–Otwarcie–Maks–Min–Zamknięcie.





Aby zapewnić prawidłową interpretację danych parametry obliczone dla każdej serii danych należy ułożyć w kolejności:

- | | | |
|-------------|------------|-----------------------------|
| 1. Me | <i>lub</i> | 1. \bar{x} , |
| 2. Q_1 | | 2. $\bar{x} - s/\sqrt{n}$, |
| 3. maksimum | | 3. $\bar{x} + s$, |
| 4. minimum | | 4. $\bar{x} - s$, |
| 5. Q_3 . | | 5. $\bar{x} + s/\sqrt{n}$. |

Przykład 3.

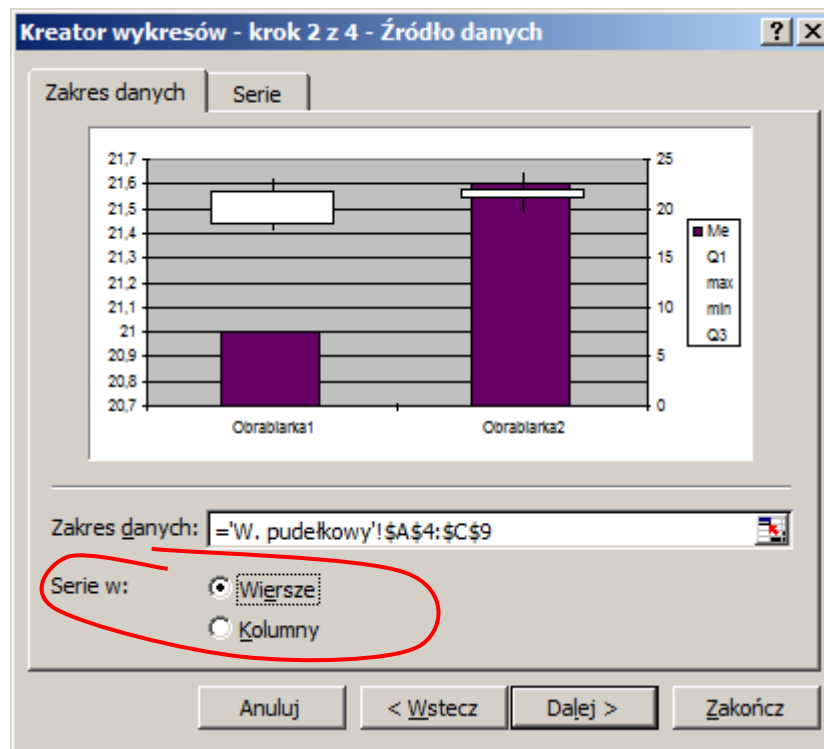
Zmierzono czasy wykonania pewnego detalu obrabianego na dwóch różnych obrabiarkach. Otrzymano następujące wyniki dla pierwszej obrabiarki: 18, 21, 22,4, 23, 21,3, 21,9, 17,6, 21, 17,8, 19,4 i dla drugiej obrabiarki: 19,5, 20, 20,9, 21,2, 21,6, 21,6, 22, 22, 22,6, 23,6. Porównać otrzymane wyniki na wykresie pudełkowym.

Dla potrzeb wykresu przygotowany został poniższy arkusz:

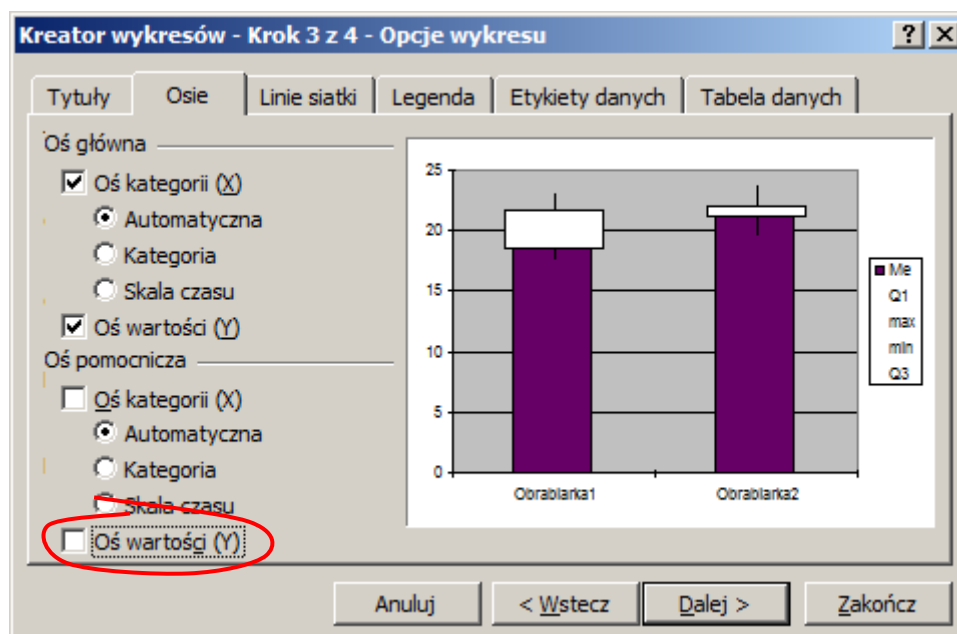
	A	B	C	D	E	F	G	H	I	J	K
1	Obr. 1	18	21	22,4	23	21,3	21,9	17,6	21	17,8	19,4
2	Obr. 2	19,5	20	20,9	21,2	21,6	21,6	22	22	22,6	23,6
3											
4		Obr.1	Obr. 2								
5	Me	21	=KWARTYL(B2:K2;2)								
6	Q1	18,35	=KWARTYL(B2:K2;1)								
7	max	23	=KWARTYL(B2:K2;4)								
8	min	17,6	=KWARTYL(B2:K2;0)								
9	Q3	21,75	=KWARTYL(B2:K2;3)								



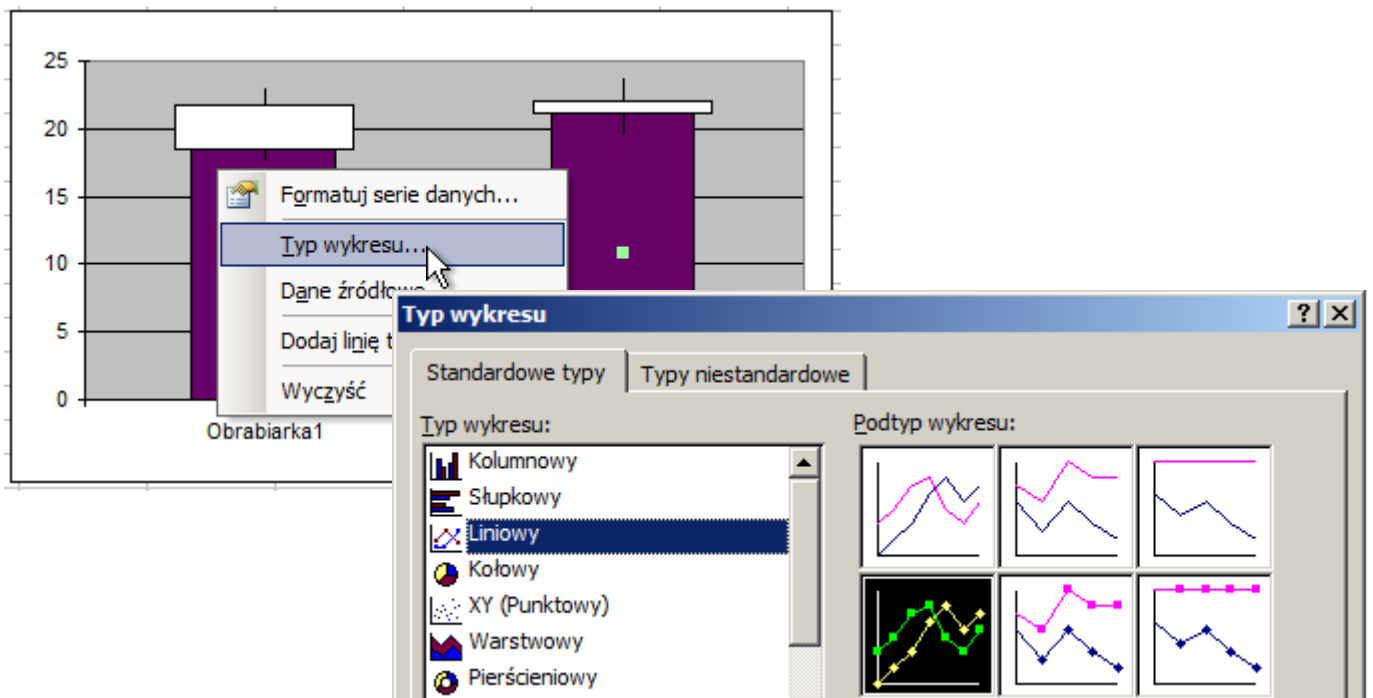
Następnie zaznaczono obszar z obliczonymi dla obydwu serii danych parametrami (A4:C9) i wybrano wykres giełdowy ustawiając opcję „Serie w” na „Wiersze”.



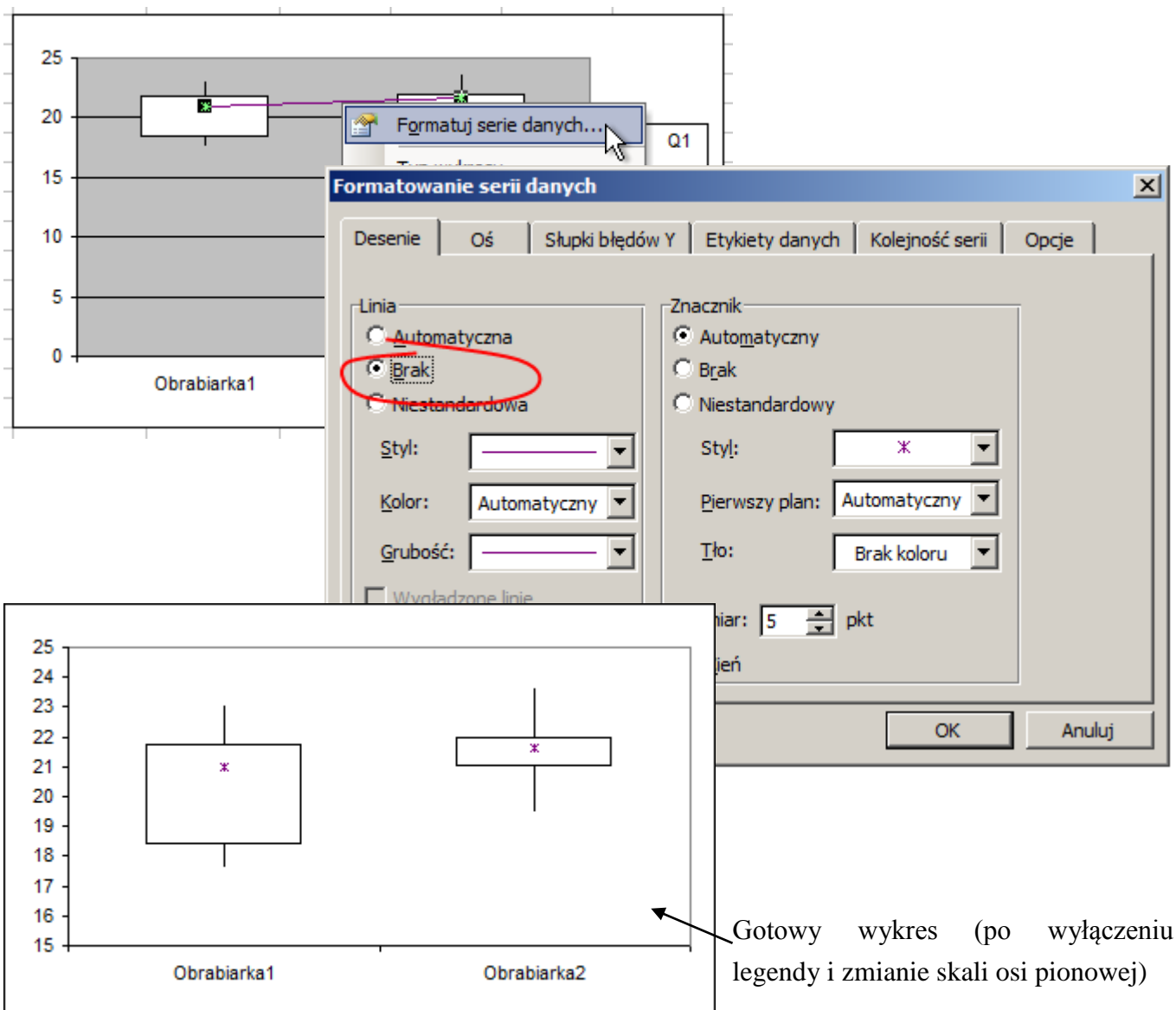
W kolejnym kroku **wyłączona** została, zbędna w tym przypadku **pionowa oś pomocnicza**:



Na koniec, zmieniono jeszcze typ wykresu dla wartości zawierających medianę na **wykres liniowy ze znacznikami danych** wyświetlanymi przy każdej wartości (w tym celu np. z menu podręcznego wybrano opcję „Typ wykresu”).



Teraz pozostaje jeszcze tylko wyłączyć linie łączące mediany (opcja „Formatuj serie danych”).



Gotowy wykres (po wyłączeniu legendy i zmianie skali osi pionowej)



2.4. Wykres kwanty–kwanty (wykres K-K lub Q-Q)

Przykład 4.

Na wykresach Q-Q porównaj:

- rozkład empiryczny długości detalu z przykładu 1. z rozkładem o parametrach wyznaczonych na podstawie danych z tego przykładu,
- rozkład empiryczny długości detalu z przykładu 1. z rozkładem o normalnym standaryzowanym.

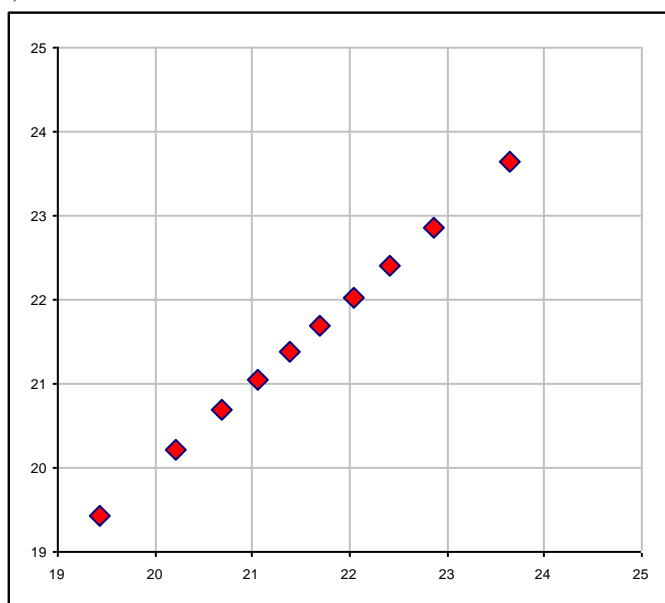
Uwaga! Rzędy kwantyli wyznaczyć wykorzystując metodę Clevelanda.

	A	B	C	D	E	F	G
1	Lp	Pomiary (uporządkowane rosnąco)	prawdopodo bieństwo	kwantyli teoretyczny	kwanty teoretyczny rozkład N(0,1)		
2							
3	1	19,5	= (A3-0,5)/10	=ROZKŁAD.NORMALNY.ODW(C3;\$D\$13;\$D\$14)			
4	2	20	0,15	20,2170411	=ROZKŁAD.NORMALNY.S.ODW(C4)		
5	3	20,9	0,25	20,6790453	-0,67448975		
6	4	21,2	0,35	21,0481564	-0,385320466		
7	5	21,6	0,45	21,3795991	-0,125661347		
8	6	21,6	0,55	21,7004009	0,125661347		
9	7	22	0,65	22,0318436	0,385320466		
10	8	22	0,75	22,4009547	0,67448975		
11	9	22,6	0,85	22,8629589	1,036433389		
12	10	24	0,95	23,639579	1,644853627		
13			średnia	=ŚREDNIA(B3:B12)			
14			odchylenie	=ODCH.STANDARDOWE(B3:B12)			

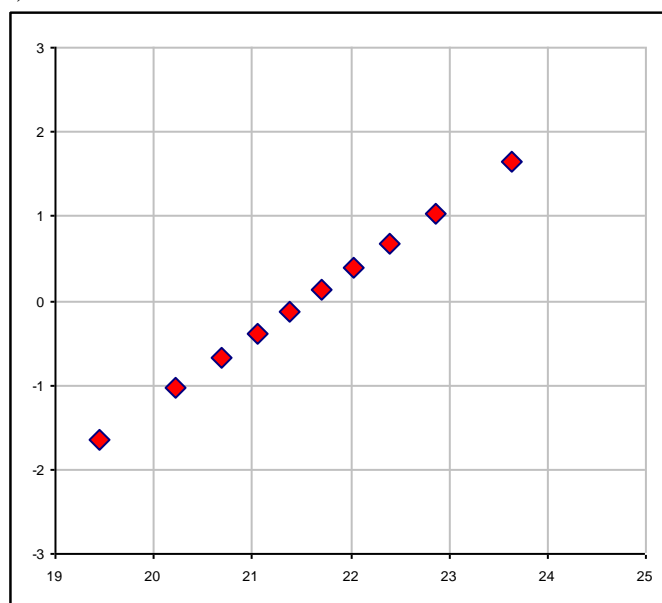
Można przyjąć że posortowane rosnąco dane pomiarowe odpowiadają kwantylom rzędu $p_i = i - 0.5/n$.

W kolumnie D wyznaczone zostały odwrotności dystrybuant rozkładu normalnego o średniej i odchyleniu standardowym wyznaczonym na podstawie danych z próby, w kolumnie E odwrotności dystrybuant rozkładu standaryzowanego. Wartości te odpowiadają kwantylom teoretycznym obydwu teoretycznych rozkładów normalnych o rzędach odpowiadającym rzędom kwantyli z próby.

a)

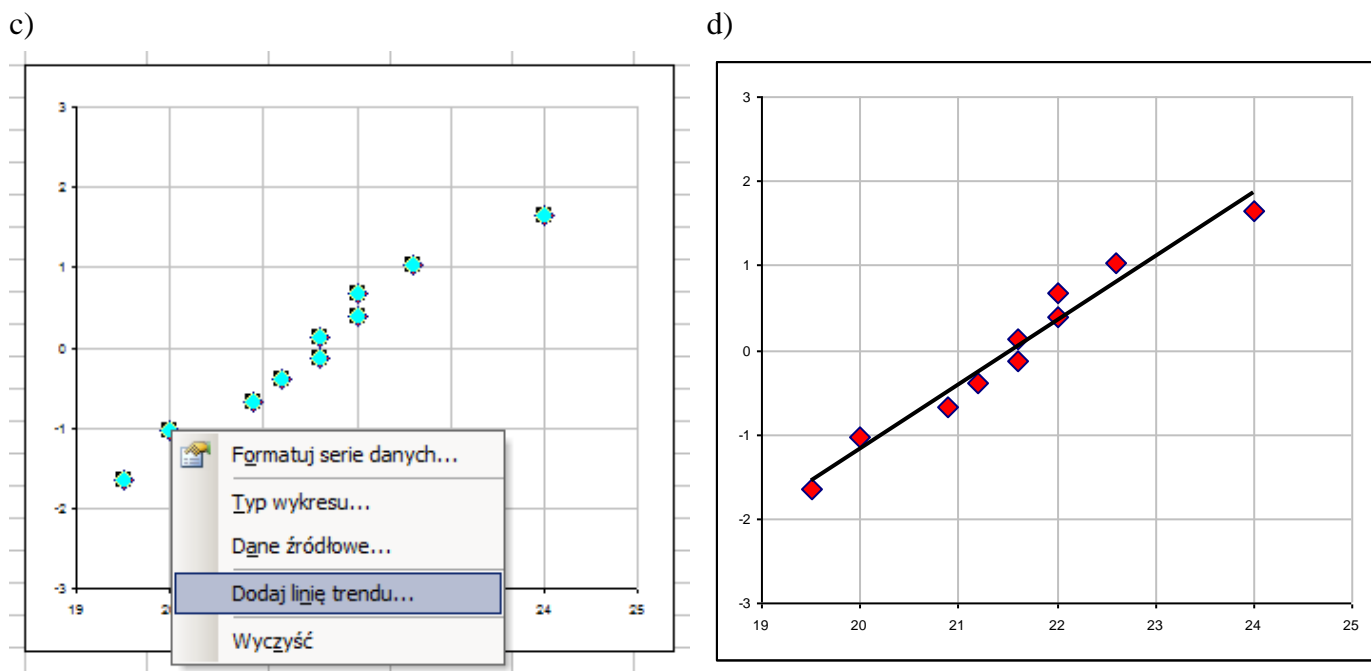


b)



Wykres $Q-Q$ można sporządzić w Excelu wykorzystując do tego celu **Wykres XY (Punktowy)**. Dla obydwu rysowanych wykresów kwantyle z próby odkładane są na osi X a kwantyle teoretyczne na osi Y. Porównanie kwantyli z próby z kwantylami teoretycznymi rozkładu normalnego o parametrach (średniej i odchyleniu standardowym) wyznaczonych na podstawie próby zostało przedstawione na rys. a, porównanie z rozkładem standaryzowanym widoczne jest na rys. b. Punkty na obydwu wykresach układają się wzdłuż linii prostej – można więc wnioskować, że rozkład danych jest rozkładem normalnym.

Wykresy kwanty-kwantyl często są uzupełniane liniowym trendem ułatwiającym sprawdzenie zgodności rozkładów. Sposób dodawania trendu został pokazany na poniższych rysunkach.



rys. c) dodawanie linii trendu d) wykres z linią trendu

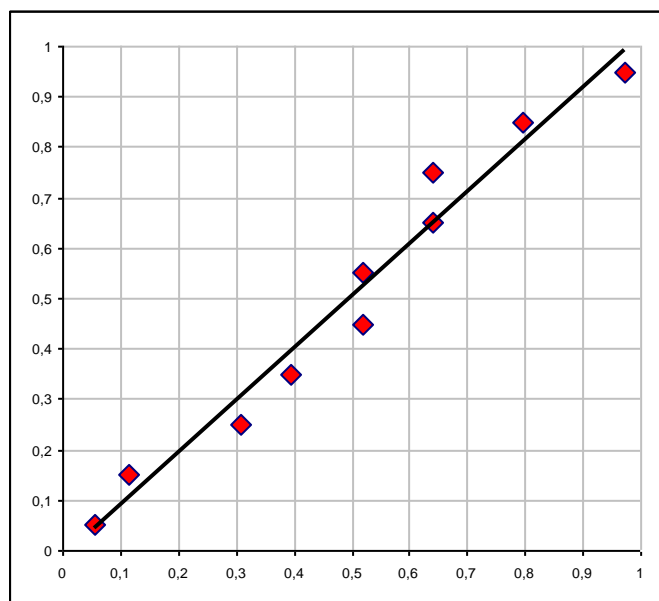
2.5. Wykres prawdopodobieństwo-prawdopodobieństwo (wykres P–P)

Przykład 5.

Porównaj z pomocą wykresu P–P rozkład pomiarów długości detalu z rozkładem normalnym o średniej i odchyleniu standardowym wyznaczonym na podstawie danych z próby.

	A	B	C	D	E	F	G
1	Lp	Pomiary (uporządkowane rosnąco)	prawdopodo bieństwo z próby	prawdopodo bieństwo teoretyczne			
2							
3	1	19,5	=ROZKŁAD.NORMALNY(B3;\$C\$13;\$C\$14;PRAWDA)				
4	2	20	0,113819	=(A4-0,5)/10			
5	3	20,9	0,308049	0,25			
6	4	21,2	0,39498	0,35			
7	5	21,6	0,518745	0,45			
8	6	21,6	0,518745	0,55			
9	7	22	0,640716	0,65			
10	8	22	0,640716	0,75			
11	9	22,6	0,796851	0,85			
12	10	24	0,973024	0,95			
13		średnia	=ŚREDNIA(B3:B12)				
14		odchylenie	=ODCH.STANDARDOWE(B3:B12)				

Po uporządkowaniu danych przy pomocy dystrybuanty rozkładu normalnego wyznaczone zostały wartości prawdopodobieństw skumulowanych dla danych z próby (po obliczeniu parametrów rozkładu: średniej i odchylenia standardowego). Następnie obliczono skumulowane prawdopodobieństwa teoretyczne $p_i = i - 0,5/n$ kolejnych pomiarów z próby i wykreślony został wykres punktowy – na osi poziomej odłożone zostały skumulowane prawdopodobieństwa z próby a na osi pionowej prawdopodobieństwa teoretyczne. Uzyskany w ten sposób wykres P–P uzupełniony o linię trendu został przedstawiony na poniższym rysunku.



rys. Wykres P–P z naniesioną linią trendu