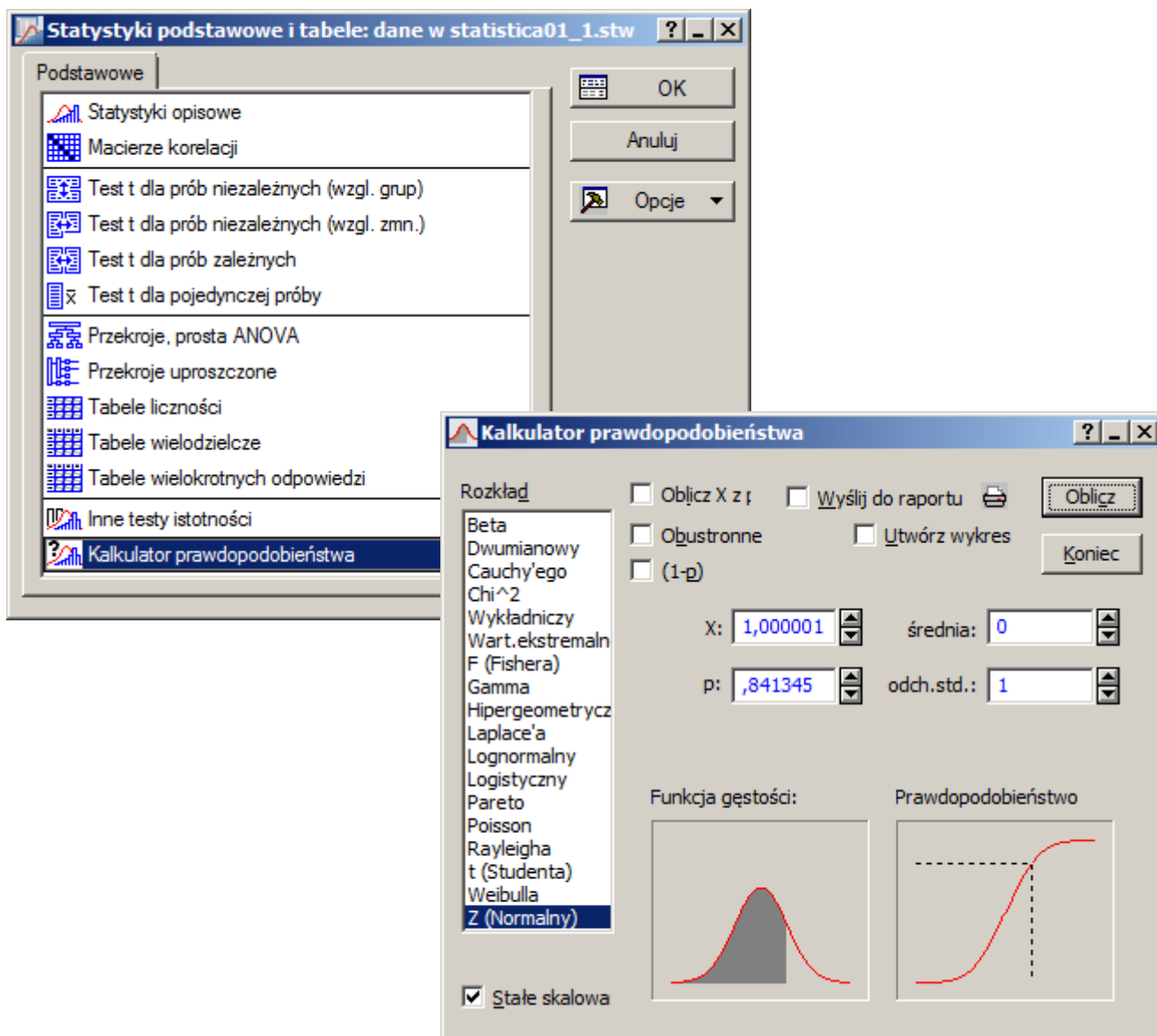


## 2. STATISTICA – FUNKCJE PODSTAWOWE

### 2.1. Kalkulator prawdopodobieństwa

Obliczanie prawdopodobieństw oraz poszukiwanie wartości odpowiadającej określonemu prawdopodobieństwu umożliwia okno **Kalkulator prawdopodobieństwa** (dostępne z menu głównego: **Statystyka/Statystyki podstawowe**).



W oknie kalkulatora użytkownik może wybrać:

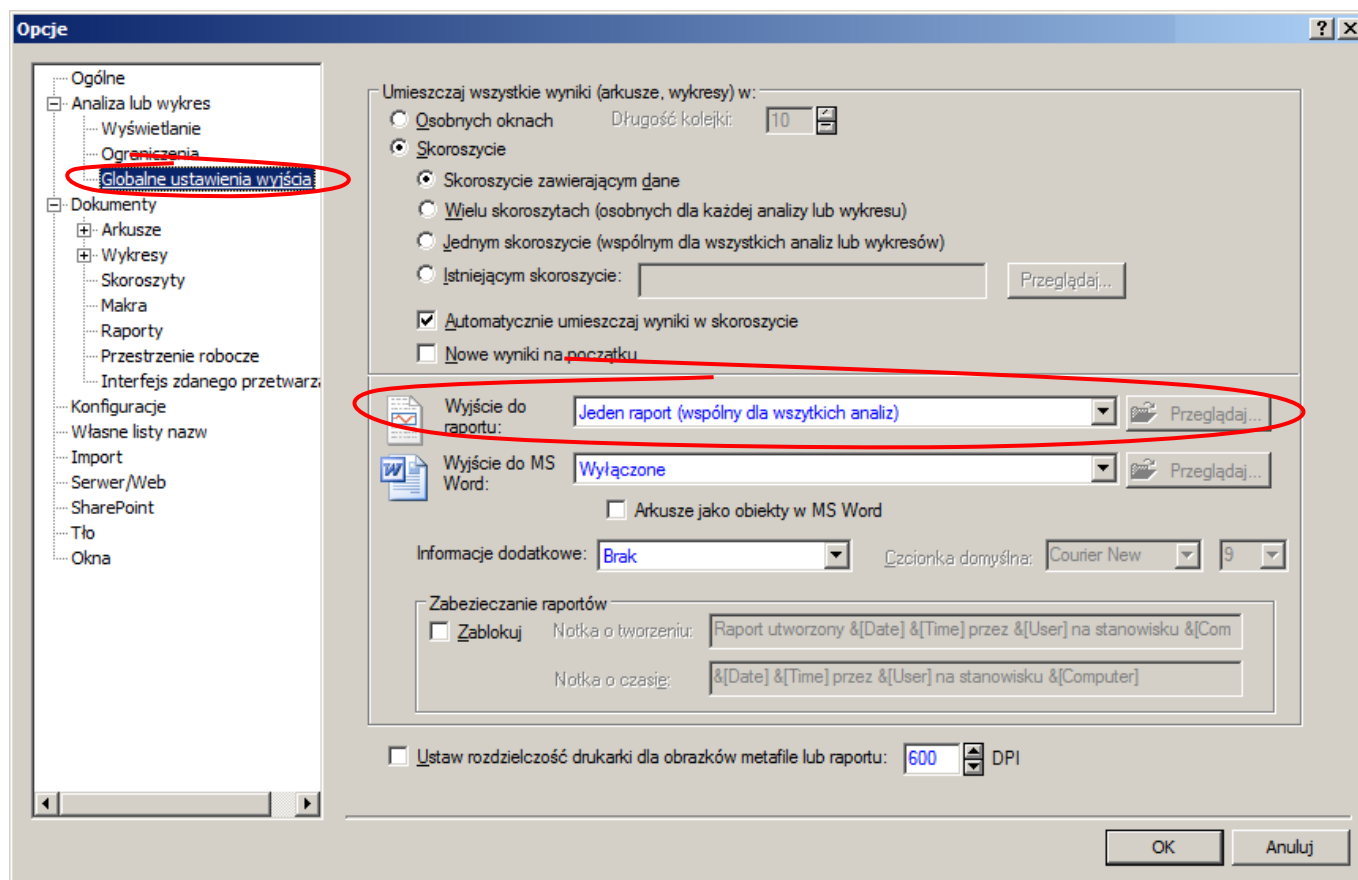
- **Rozkład** prawdopodobieństwa w oparciu o który będzie wykonywał obliczenia,
- **Oblicz X z prawdopodobieństwa** – jeżeli poszukuje wartości  $x$  odpowiadającej podanemu prawdopodobieństwu  $P(X \leq x)$ ,
- **Obustronne** – jeżeli oblicza prawdopodobieństwo  $P(|X| \leq x)$  lub poszukuje wartości  $x$  odpowiadającej podanemu prawdopodobieństwu  $P(|X| \leq x)$ , opcja dostępna jest tylko dla rozkładów symetrycznych,



- **(1-p)** – jeżeli obliczenia **Oblicz X z prawdopodobieństwa** mają odpowiadać prawdopodobieństwom  $P(X > x)$  lub  $P(|X| > x)$
- **X** – podawana lub poszukiwana wartość  $x$ ,
- **p** – podawana lub obliczana wartość prawdopodobieństwa,
- **parametry zależne od typu rozkładu**, np.: dla rozkładu normalnego: średnia odch.std., df, df1, df2 (liczba stopni swobody) dla rozkładu *t-Studenta*, rozkładu  $\chi^2$ , rozkładu F, itd.

W oknie **Kalkulatora** pomocniczo wyświetlane są funkcja gęstości wybranego rozkładu i wykres prawdopodobieństwa. Obliczenia wykonywane są po naciśnięciu przycisku **Oblicz**. Dodatkowo **Kalkulator** umożliwia generowanie wykresów przedstawiających funkcje gęstości i prawdopodobieństwa z zaznaczonymi obliczonymi wartościami (skala wykresów może być ustalana na podstawie parametrów rozkładu jeśli opcja **Stale skalowanie** zostanie odznaczona) jeśli w oknie **Kalkulatora** zostanie wybrana opcja **Utwórz wykres**. Obliczone wyniki mogą również zostać przesłane do raportu jeśli wybrana zostanie opcja **Wyślij do raportu**.

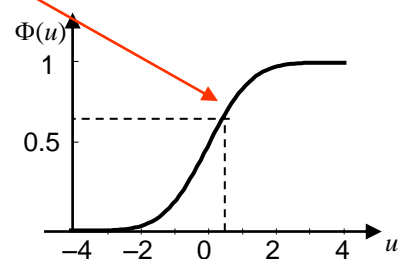
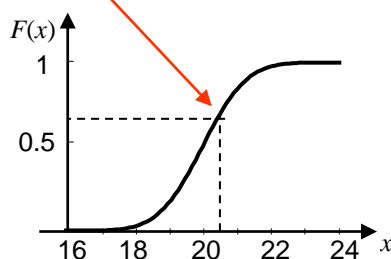
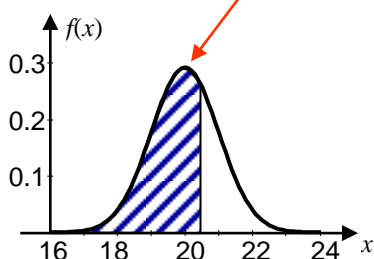
Na przedstawionym poniżej rysunku raporty zostały skonfigurowane w taki sposób, że wyniki wszystkich analiz trafiają do wspólnego raportu (okno **Opcje**, dostępne z menu głównego: **Narzędzia/Opcje**, opcja **Globalne ustawienia wyjścia**).



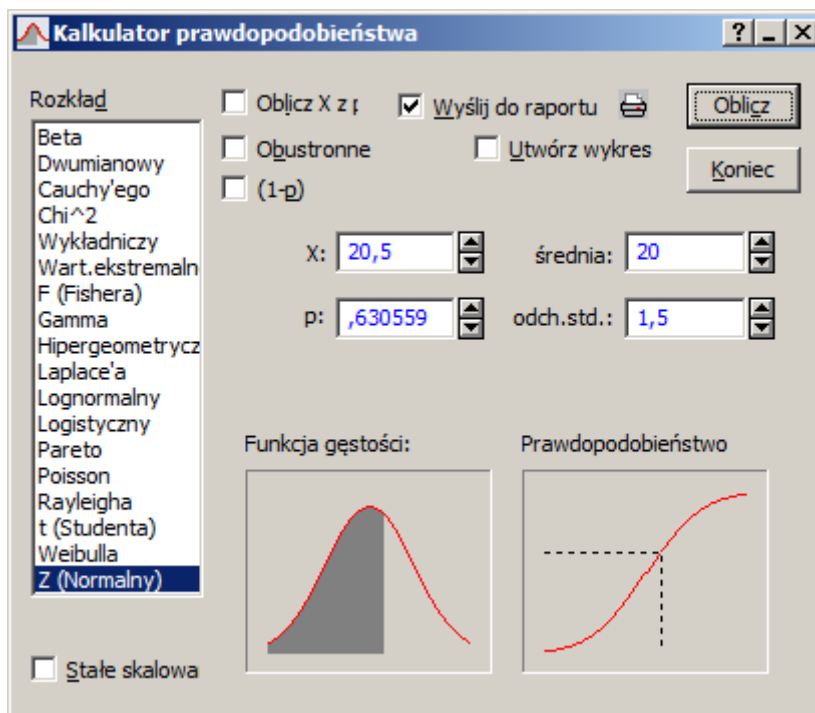
**Przykład 1.** Na podstawie pomiarów długości dużej partii detali wykonywanych na pewnym stanowisku stwierdzono, że rozkład długości jest rozkładem  $\mathcal{N}(20, 1.5)$ . Obliczyć prawdopodobieństwo, że długość losowo wybranego detalu:

- jest mniejsza lub równa 20.5,
- jest większa od 21.5,
- mieści się w przedziale  $(20.5, 21.5]$ ,
- co najmniej o 2 jednostki różni się od średniej,
- obliczyć odchylenie od średniej dla którego prawdopodobieństwo wystąpienia detali o długości przekraczającej wyznaczone odchylenie wyniesie 0.1.

$$a) \quad P(x \leq 20.5) = F_{\mathcal{N}(20,1.5)}(20.5) = \Phi\left(\frac{20.5 - 20}{1.5}\right) = \Phi(0.3333) = 0.6306$$

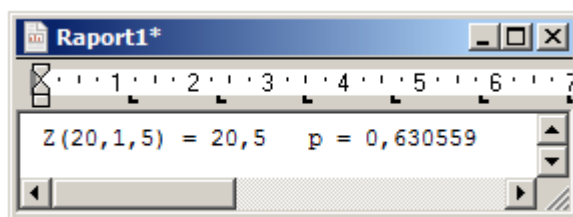


W oknie **Kalkulatora** wybrano rozkład normalny, wprowadzono jego parametry (średnia = 20, odchylenie standardowe = 1,5), następnie wprowadzono wartość X: 20,5. Po naciśnięciu **Oblicz** otrzymano wartość prawdopodobieństwa  $p = 0,630559$ .



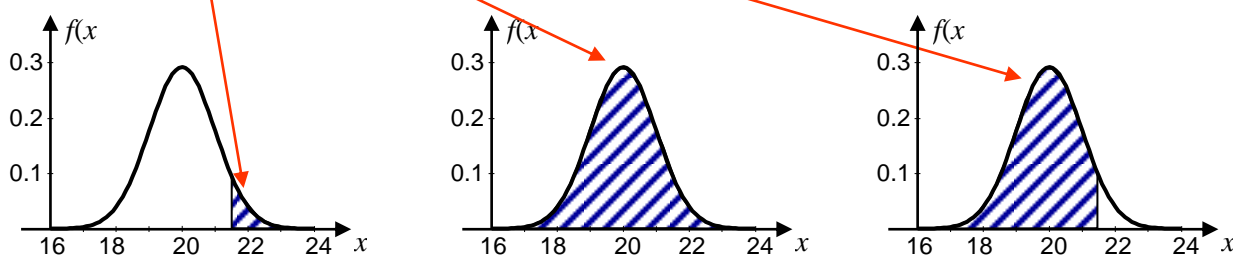
W **Kalkulatorze** zaznaczona została również opcja: **Wyślij do raportu**, co spowodowało utworzenie raportu o domyślnej nazwie *Raport1*. W raporcie pojawiła się informacja o rozwiązywanym zadaniu

(poszukiwanie prawdopodobieństwa odpowiadającego wartości 20,5 w rozkładzie normalnym o parametrach  $\mu = 20$  i  $\sigma = 1,5$ ).

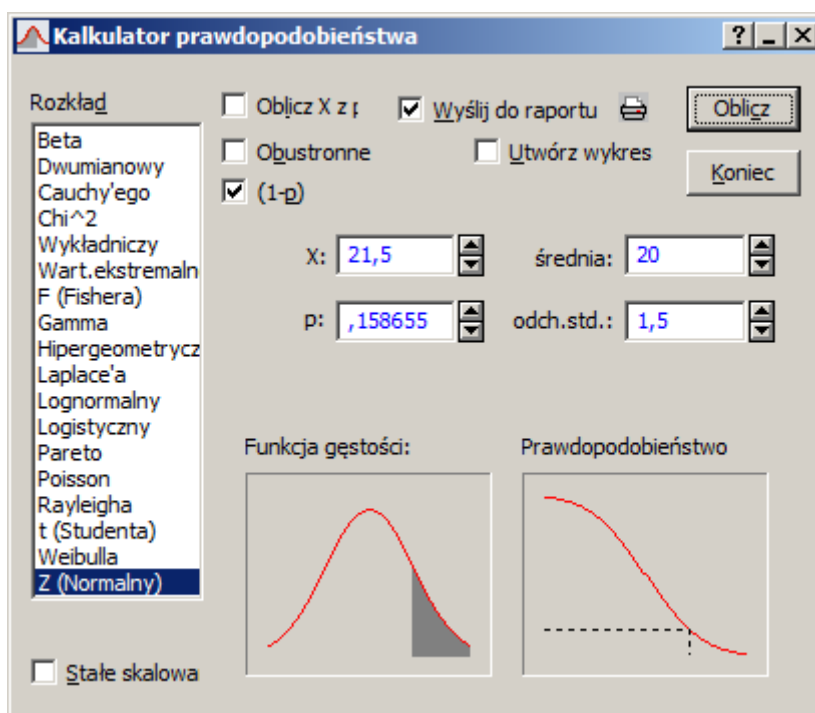


Po zakończeniu wszystkich obliczeń raport można zapisać w postaci osobnego pliku, można także dołączyć go do wybranego skoroszytu: opcja **Dodaj do skoroszytu** dostępna z menu głównego **Podstawowe**.

$$b) \quad P(x > 21.5) = 1 - P(x \leq 21.5) = 1 - F_{\mathcal{N}(20,1.5)}(21.5) = 1 - \Phi\left(\frac{21.5 - 20}{1.5}\right) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

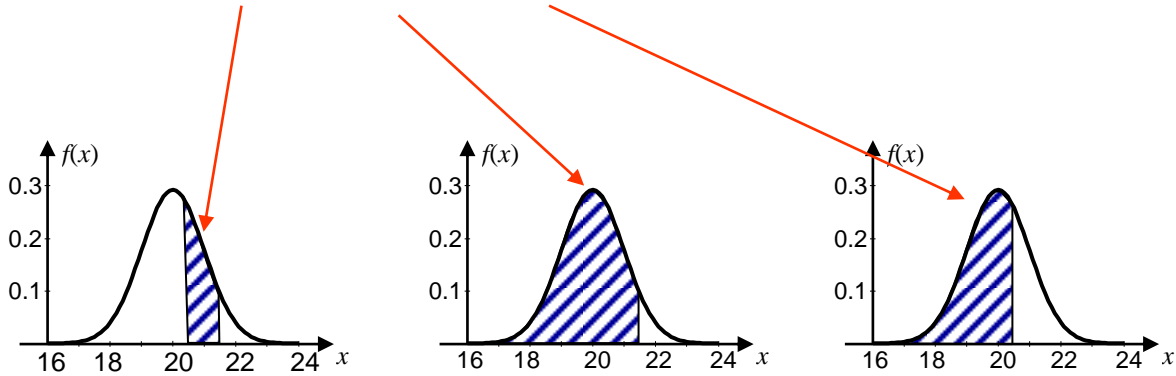


W oknie **Kalkulatora** wprowadzono tak jak poprzednio rozkład i jego parametry, następnie wprowadzono wartość X: 21,5 oraz zaznaczono opcję **(1-p)**. Po naciśnięciu **Oblicz** otrzymano wartość prawdopodobieństwa  $p = 0,158655$ .



c)

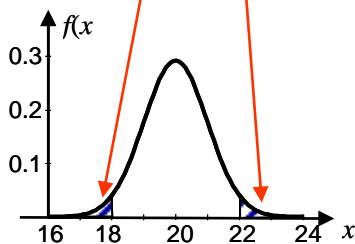
$$P(20.5 < x \leq 21.5) = P(x \leq 21.5) - P(x \leq 20.5) = F_{\mathcal{N}(20,1.5)}(21.5) - F_{\mathcal{N}(20,1.5)}(20.5) = \Phi(1) - \Phi(0.3333) = 0.2108$$



Ze względu na to, że wartości 20,5 i 21,5 nie są rozłożone symetrycznie względem średniej rozkładu nie można w tym przypadku skorzystać z opcji **Obustronne**. Odpowiedź można w tym przypadku otrzymać sumując prawdopodobieństwa obliczone w punktach a) i b).

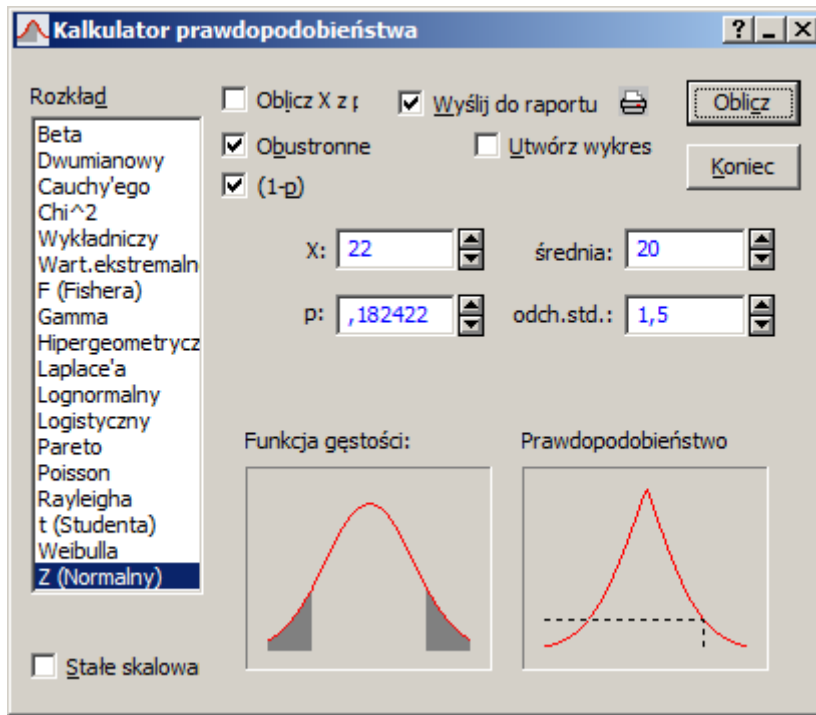
d)

$$P(|x - 20| \geq 2) = P(x \leq 18) + P(x \geq 22) = 2P(x \leq 18) = 2F_{\mathcal{N}(20,1.5)}(18) = 2\Phi\left(\frac{18-20}{1.5}\right) = 2\Phi(-1.3333) = 0.1824$$



Ze względu na to, że wartości 18 i 22 są rozłożone symetrycznie względem średniej rozkładu można w tym przypadku skorzystać z opcji **Obustronne**. Zadanie można rozwiązać wprowadzając wartość X: 18 lub X: 22 (z odznaczoną opcją **(1-p)** i zaznaczoną opcją **Obustronne**).



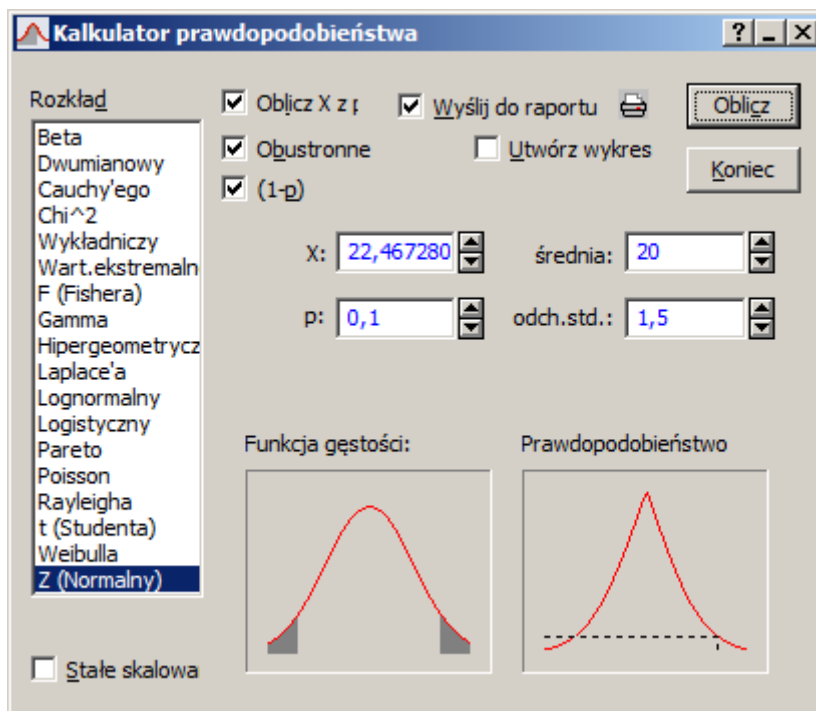


e)

$$P(|x - 20| \geq odl) = 0.1$$

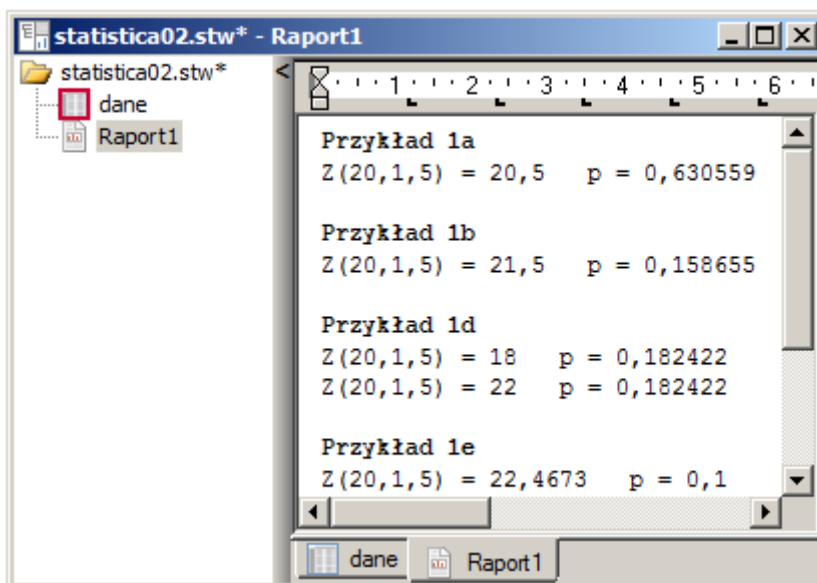
$$P(|x - 20| \geq odl) = 2P(x \leq 20 - odl) = 2F_{\mathcal{N}(20,1.5)}(20 - odl)$$

$$F_{\mathcal{N}(20,1.5)}(20 - odl) = 0.05 \longrightarrow 20 - odl = F_{\mathcal{N}(20,1.5)}^{-1}(0.05)$$

$$odl = 20 - F_{\mathcal{N}(20,1.5)}^{-1}(0.05) \longrightarrow odl = 2.4673$$


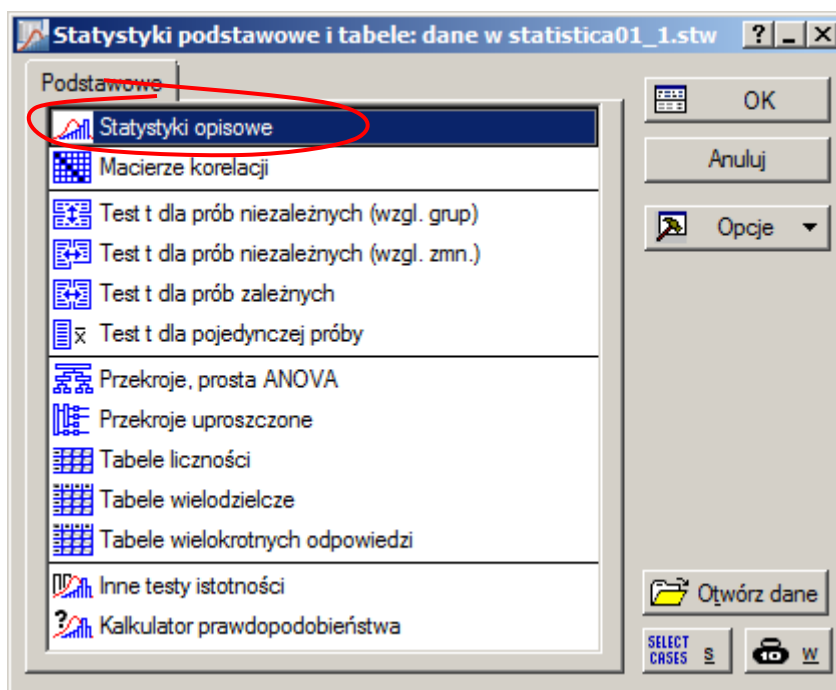
W tym przypadku poszukiwane są symetrycznie rozłożone wokół średniej wartości  $x$  takie, że prawdopodobieństwo wystąpienia wartości na zewnątrz wyznaczonego w ten sposób przedziału wynosi 0.1. Zadanie można rozwiązać wybierając opcje: **Oblicz X z prawdopodobieństwa, Obustronne** i **(1-p)** oraz wprowadzając wartość prawdopodobieństwa  $p$ : 0,1. Otrzymana wartość  $X$ : 22,467280 leży na prawo od średniej, wartość położoną na lewo można obliczyć wyznaczając odległość  $X$  od średniej i następnie odejmując ją od tej średniej.

Wyniki przeprowadzonych obliczeń były wysyłane do raportu, który został dołączony do bieżącego skoroszytu. Na poniższym rysunku przedstawiony został wygenerowany w ten sposób raport, który został uzupełniony dodatkowymi opisami.



## 2.2. Statystyki opisowe

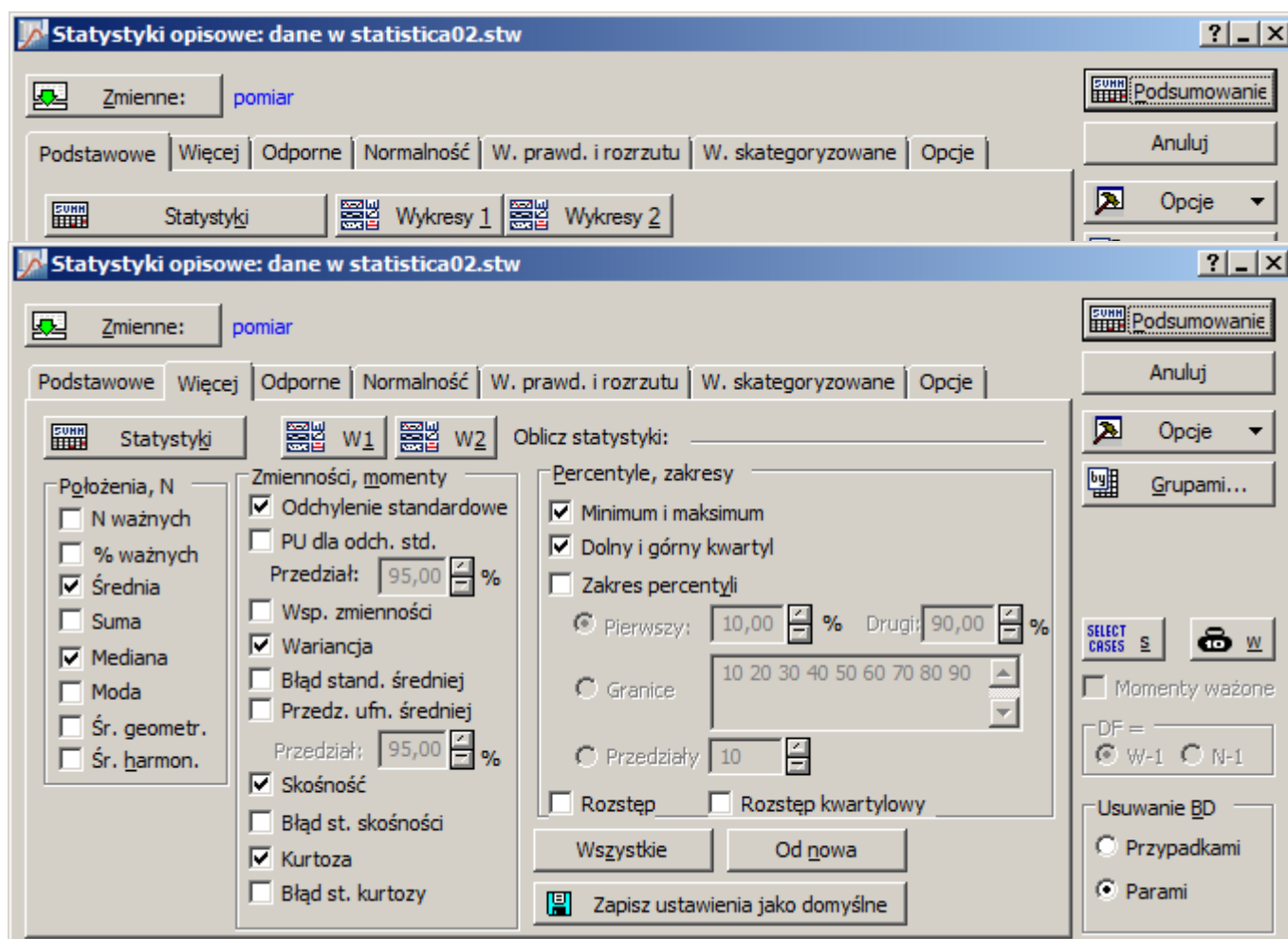
Miary położenia, rozproszenia i zniekształcenia mogą być obliczane z pomocą okna **Statystyki podstawowe** (dostępne z menu głównego: **Statystyka/Statystyki podstawowe**).



Przyciski **Statystyki** i **Podsumowanie** umożliwiają wykonanie obliczeń podstawowych statystyk dla zmiennej wskazanej przy pomocy przycisku **Zmienne** (na poniższym rysunku do analizy została wybrana zmienna **pomiar**).

O wyborze obliczanych miar decydują ustawienia na zakładce **Więcej**. Znaczenie większości miar jest oczywiste. W sposób skrótowy opisane zostały: liczba i procent ważnych obserwacji (N i % ważnych), przedział ufności dla odchylenia standardowego (PU dla odch. std.) z możliwością określania poziomu ufności (domyślnie 95%), przedział ufności dla średniej (Przedz. ufn. średniej) z możliwością określania poziomu ufności (domyślnie 95%).

Poniższy rysunek przedstawia przykładowe ustawienia zakładki **Więcej** dla zadania polegającego na wyznaczeniu: średniej, mediany, odchylenia standardowego, wariancji, minimum, maksimum, kwartyli dolnego i górnego, skośności i kurtozy.



Wyniki obliczeń odkładane są w postaci kolejnego folderu w bieżącym skoroszybie.

| Statystyki opisowe (dane w statistica02.stw) |          |          |          |          |                |            |
|--|----------|----------|----------|----------|----------------|------------|
| Zmienna                                      | Średnia  | Mediana  | Minimum  | Maksimum | Dolny Kwartył. | Górn Kwart |
| <b>pomiar</b>                                | 21,54000 | 21,60000 | 19,50000 | 24,00000 | 20,90000       | 22,00000   |

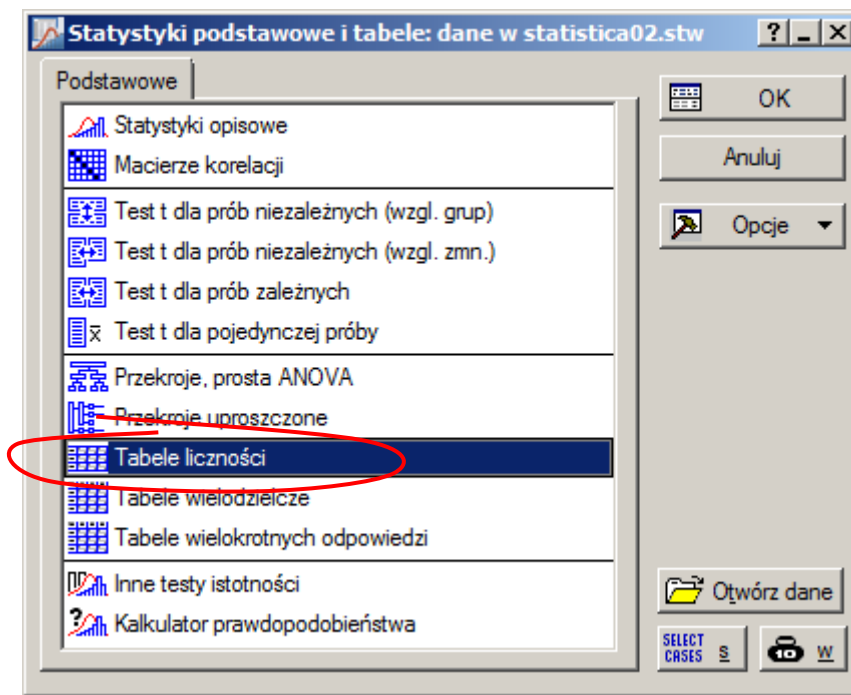




### 2.3. Tabele liczności

Tabele liczności pozwalają na wyznaczenie liczebności i częstości w przedziałach klasowych szeregu rozdzielczego. Tabele te dostępne z menu: **Statystyka/Statystyki podstawowe**.

**Szereg rozdzielczy** – wartości z próby dzielone są na *klasy* (kategorie), dla każdej klasy podawana jest jej *liczebność* lub *częstość* (stosunek liczebności klasy do liczebności całej próby).



Sposób korzystania z tabel zostanie wyjaśniony na poniższym przykładzie.

#### Przykład 2.

Załóżmy, że arkusz *dane* zawiera wynik pomiaru pewnej cechy dziesięciu wybranych detali.

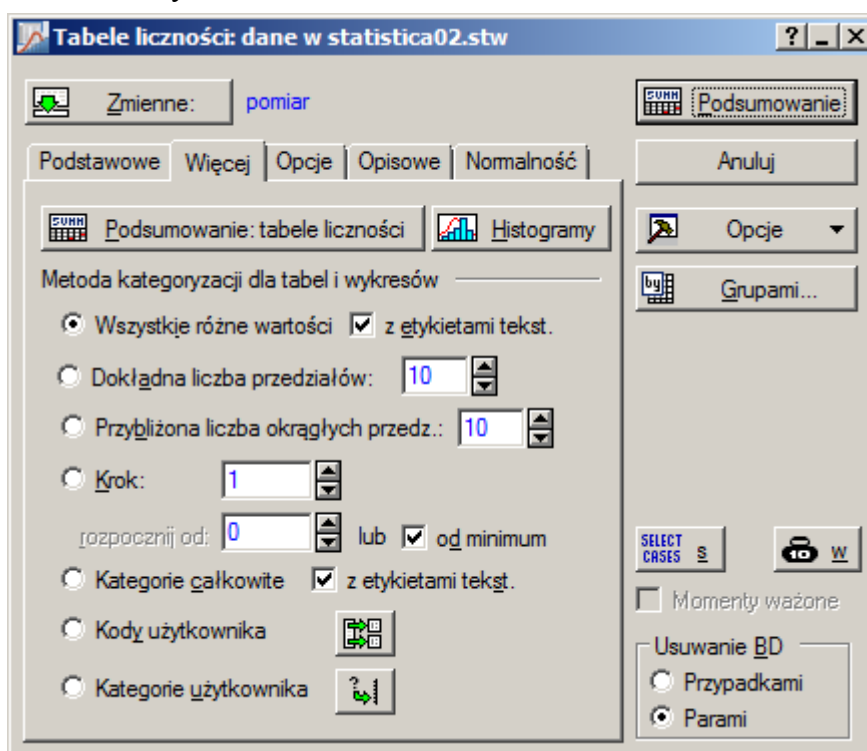
|    | 1 pomiar |
|----|----------|
| 1  | 21,6     |
| 2  | 21,2     |
| 3  | 20       |
| 4  | 19,5     |
| 5  | 20,9     |
| 6  | 24       |
| 7  | 22       |
| 8  | 21,6     |
| 9  | 22       |
| 10 | 22,6     |

Wyniki pomiarów należy przedstawić w postaci szeregu rozdzielczego przyjmując następujące klasy:  $(-\infty, 19]$ ,  $(19, 20]$ ,  $(20, 21]$ ,  $(21, 22]$ ,  $(22, 23]$ ,  $(23, +\infty)$ .

Wybór opcji **Tabele liczości** powoduje wyświetlenie przedstawionego poniżej okna. Dla potrzeb analizy należy wybrać zmienną zawierającą zliczane wartości – w przykładzie zmienną tą jest zmienna **pomiar**.



Przyciski **Podsumowanie: tabele liczości** i **Podsumowanie** umożliwiają wykonanie obliczeń (liczebności, liczebności skumulowanych, częstości, częstości skumulowanych, itd.) na podstawie ustawień na zakładkach **Więcej** oraz **Opcje**). Zakładka **Więcej** pozwala na ustalenie sposobu wyznaczenia przedziałów klasowych.

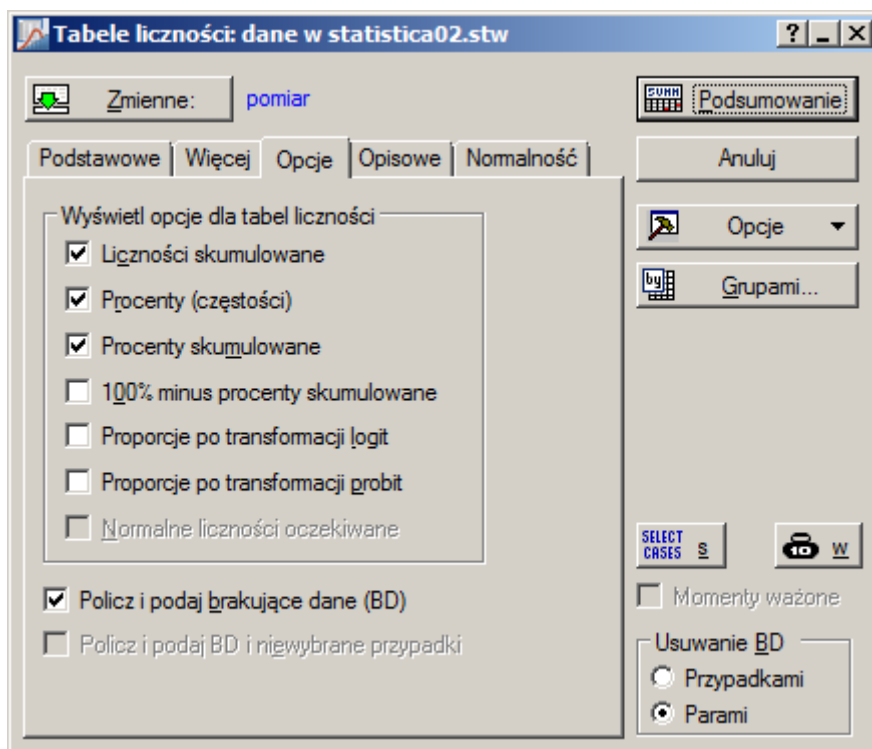


Dostępne są następujące metody kategoryzacji:

- **Wszystkie różne wartości** – zbudowanie tylu przedziałów klasowych ile jest różnych wartości analizowanej zmiennej,
- **Dokładna liczba przedziałów** – zbudowanie wskazanej liczby przedziałów, przedziały te wypełnią równomiernie przedział pomiędzy minimalną a maksymalną wartością zmiennej,
- **Przybliżona liczba okrągłych przedziałów** – przedziały zostaną wyznaczone w taki sposób, żeby ich granice były zaokrąglone a ich liczba nie przekraczała wskazanej,
- **Krok** – granice wyznaczone od wartości minimalnej lub wskazanej z zadanyim krokiem,
- **Kategorie całkowite** – obliczenia przeprowadzane jedynie dla całkowitych wartości analizowanej zmiennej,
- **Kody użytkownika** – obliczenia przeprowadzane jedynie dla wskazanych przez użytkownika wartości analizowanej zmiennej,
- **Kategorie użytkownika** – obliczenia przeprowadzane w oparciu o warunki logiczne wskazane przez użytkownika.

Zakładka **Opcje** umożliwia ustalenie dodatkowych wartości (oprócz liczebności) obliczanych dla każdej kategorii. Dodatkowo mogą być obliczane np.:

- **Liczebności skumulowane,**
- **Procenty (częstości)** – częstości wyrażone w procentach,
- **Procenty skumulowane,**
- **100% minus procenty skumulowane.**



Powyższe zadanie można wykonać na kilka sposobów. Wykorzystując opcję **Przybliżona liczba okrągłych przedziałów**:

The dialog box 'Tabele licznosci: dane w statistica02.stw' shows the 'Podsumowanie: tabele licznosci' tab. Under 'Metoda kategoryzacji dla tabel i wykresow', the option 'Przybliżona liczba okrągłych przedz.: 10' is selected and circled in red. The 'okragle przedzialy' window displays the following table:

| Od    | Do               | Tabela licznosci |
|-------|------------------|------------------|
| 19    | $\leq x \leq 20$ | 2                |
| 20    | $\leq x \leq 21$ | 1                |
| 21    | $\leq x \leq 22$ | 5                |
| 22    | $\leq x \leq 23$ | 1                |
| 23    | $\leq x \leq 24$ | 1                |
| 24    | $\leq x \leq 25$ | 0                |
| 25    | $\leq x \leq 26$ | 0                |
| Braki |                  | 0                |

Wykorzystując opcję **Krok** należy wskazać wartość początkową: 18 i krok: 1:

The dialog box 'Tabele licznosci: dane w statistica02.stw' shows the 'Podsumowanie: tabele licznosci' tab. Under 'Metoda kategoryzacji dla tabel i wykresow', the option 'Krok: 1' is selected and circled in red. The 'ustalony krok' window displays the following table:

| Od    | Do            | Tabela licznosci |
|-------|---------------|------------------|
| 18    | $\leq x < 19$ | 0                |
| 19    | $\leq x < 20$ | 1                |
| 20    | $\leq x < 21$ | 2                |
| 21    | $\leq x < 22$ | 3                |
| 22    | $\leq x < 23$ | 3                |
| 23    | $\leq x < 24$ | 0                |
| 24    | $\leq x < 25$ | 1                |
| Braki |               | 0                |



Pełną kontrolę nad sposobem ustalania przedziałów daje opcja: **Kategorie użytkownika**, która w tym przypadku wymagałaby ustalenia warunków dla poszczególnych przedziałów w postaci:

Określ kategorie

Kategoria 2  
Włącz jeśli:  Otwórz/Zapisz

Kategoria 3  
Włącz jeśli:  Otwórz/Zapisz

Kategoria 4  
Włącz jeśli:  Otwórz/Zapisz

Kategoria 5  
Włącz jeśli:  Otwórz/Zapisz

Przebieg zmiennych  
Otwórz...  
Zapisz...

Określ kategorie wpisując kryteria podziału na grupy (kategoryzacji). Program sprawdza kryteria jedno po drugim i przypisuje przypadek do pierwszej grupy, dla której spełnia on kryteria.  
Każdy przypadek jest tylko raz klasyfikowany, nawet jeżeli pasuje do następnym kategorii nie będzie ponownie zliczany.

statistica02.stw\* - kategorie u.

| Klasa                          | Tabela liczn |
|--------------------------------|--------------|
|                                | Liczba       |
| Włącz pomiar<=19               | 0            |
| Włącz pomiar>19 and pomiar<=20 | 2            |
| Włącz pomiar>20 and pomiar<=21 | 1            |
| Włącz pomiar>21 and pomiar<=22 | 5            |
| Włącz pomiar>22 and pomiar<=23 | 1            |
| Włącz pomiar>23                | 1            |
| Nie określone                  | 0            |

okragłe przedziały   ustalony krok   kategorie u.

W przykładzie do zdefiniowania wszystkich kategorii poza pierwszą i ostatnią wykorzystany został warunek złożony z dwóch warunków logicznych połączonych operatorem „and” reprezentującym logiczną koniunkcję. Taka konstrukcja warunków nie jest jednak konieczna ponieważ program sprawdza warunki w kolejności ich wprowadzenia i przypadek jest zaliczany do pierwszej kategorii dla której zachodzi opisujący ją warunek. Dodatkowo nie jest również konieczna konstrukcja ostatniej kategorii – przypadki nie zaliczone do żadnej z kategorii zostaną zliczone jako „Nie określone”. Uproszczona wersja definicji kategorii użytkownika została pokazana na kolejnym rysunku.

The image shows two windows from the STATISTICA software. The top window is titled "Określ kategorie" and contains four categories for defining data groups. Each category has a "Włącz jeśli:" dropdown menu and an "Otwórz/Zapisz" button. The criteria entered are:

- Kategoria 1: pomiar <= 19
- Kategoria 2: pomiar <= 20
- Kategoria 3: pomiar <= 21
- Kategoria 4: pomiar <= 22

Buttons on the right include "OK", "Anuluj", "Przegląd zmiennych", "Otwórz...", and "Zapisz...". A text box at the bottom right of this window reads: "Określ kategorie wpisując kryteria podziału na grupy (kategoryzacji). Program".

The bottom window is titled "statistica02.stw\* - kategorie u." and displays a frequency table. The left pane shows a file explorer with the "kategorie u." file selected. The right pane shows the following table:

| Klasa              | Tabela liczn |
|--------------------|--------------|
| Włącz pomiar <= 19 | 0            |
| Włącz pomiar <= 20 | 2            |
| Włącz pomiar <= 21 | 1            |
| Włącz pomiar <= 22 | 5            |
| Włącz pomiar <= 23 | 1            |
| Nie określone      | 1            |

