

## Ćwiczenie:

Badanie normalności rozkładu. Wyznaczanie przedziałów ufności.

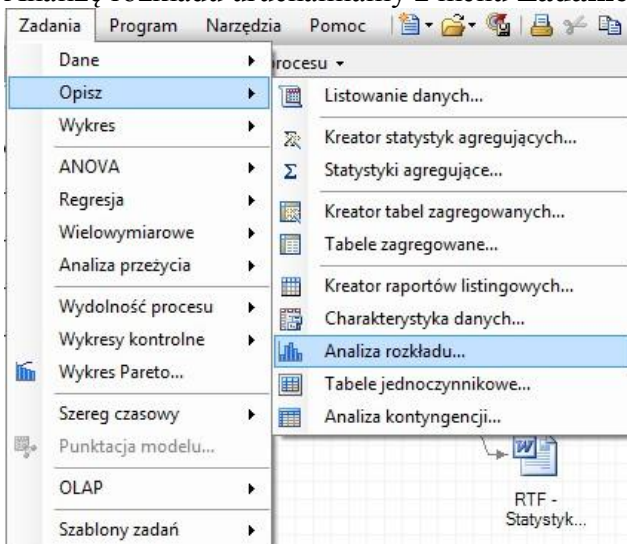
### Badanie normalności rozkładu

**Shapiro-Wilka:** jest on najbardziej zalecanym testem normalności rozkładu. Jednak wskazane jest, aby liczebność próby była poniżej 2000, gdyż w przeciwnym wypadku daje mylne wyniki. Test ten wymaga, aby cecha miała rozkład ciągły.

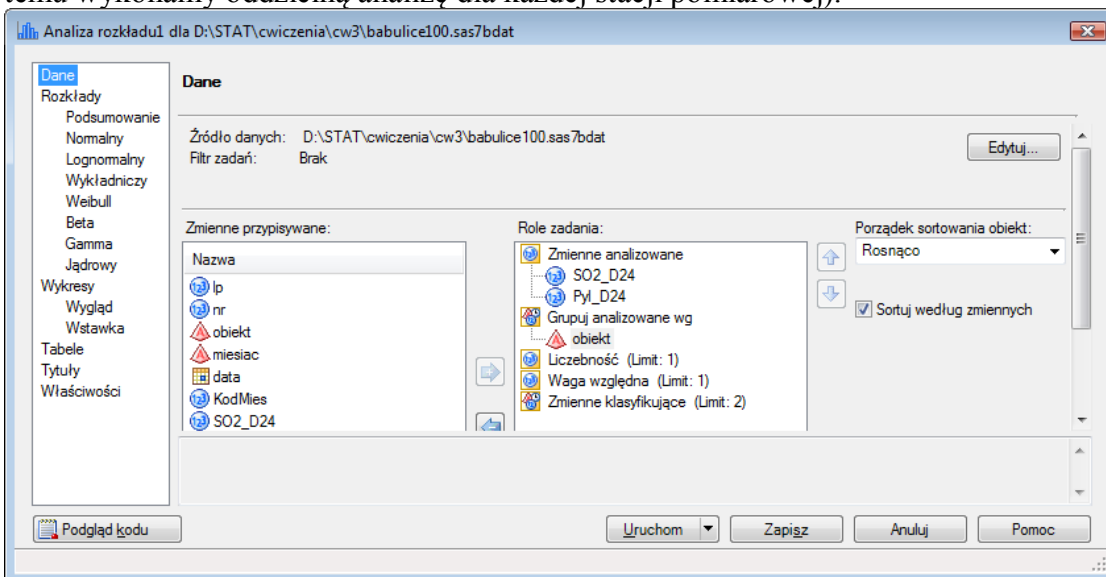
**Kolmogorowa-Smirnowa (K-S):** Wymaga znajomości średniej i odchylenie standardowego dla populacji. W sytuacji, gdy nie znamy wyżej wymienionych parametrów stosujemy test **K-S** z poprawką Lilieforsa. **Test K-S** wymaga, aby cecha była ciągła. Jest to test mocniejszy, aniżeli test  $X^2$ .

Przykład 1: Sprawdź czy zmienne: imisja SO<sub>2</sub> i pyłu posiadają rozkład zgodny z normalnym (babulice100)

Analizę rozkładu uruchamiamy z menu **Zadanie**.

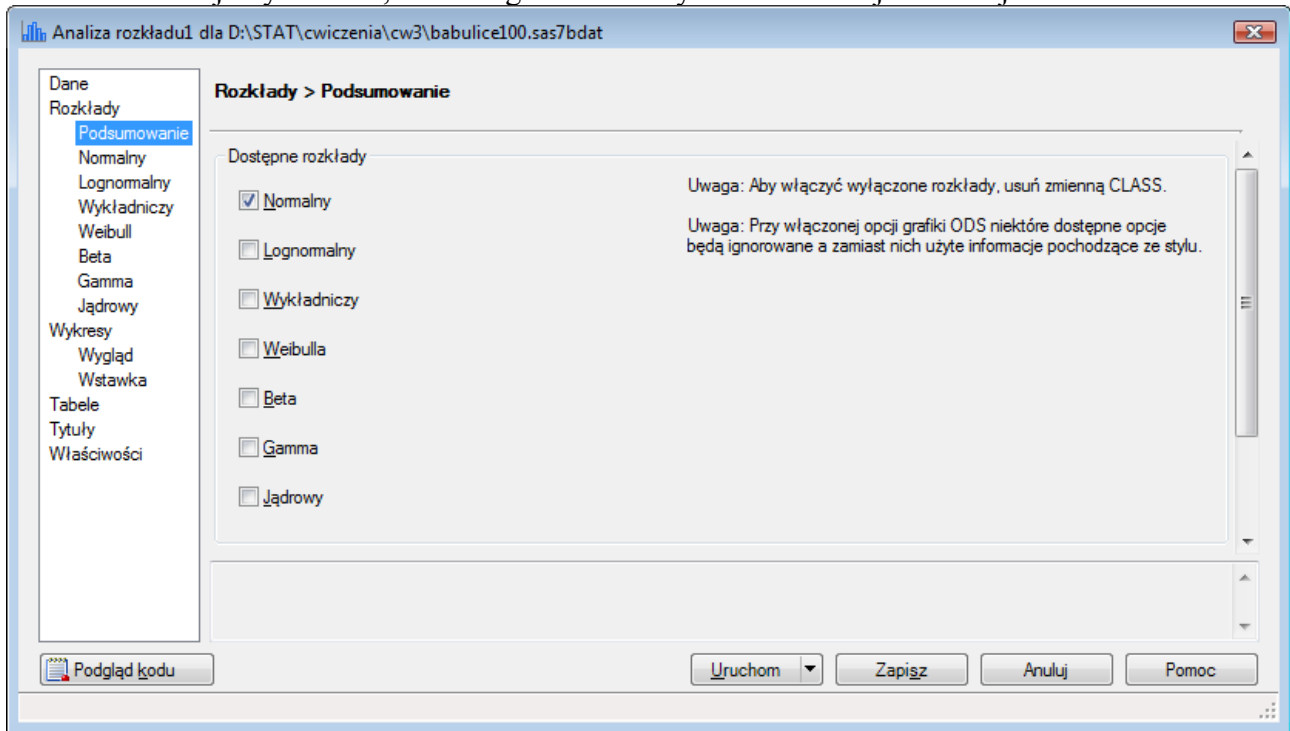


**Krok 1.** Ustalamy zmienne analizowane a także wprowadzamy zmienną grupującą **obiekt** (dzięki temu wykonamy oddzielną analizę dla każdej stacji pomiarowej).

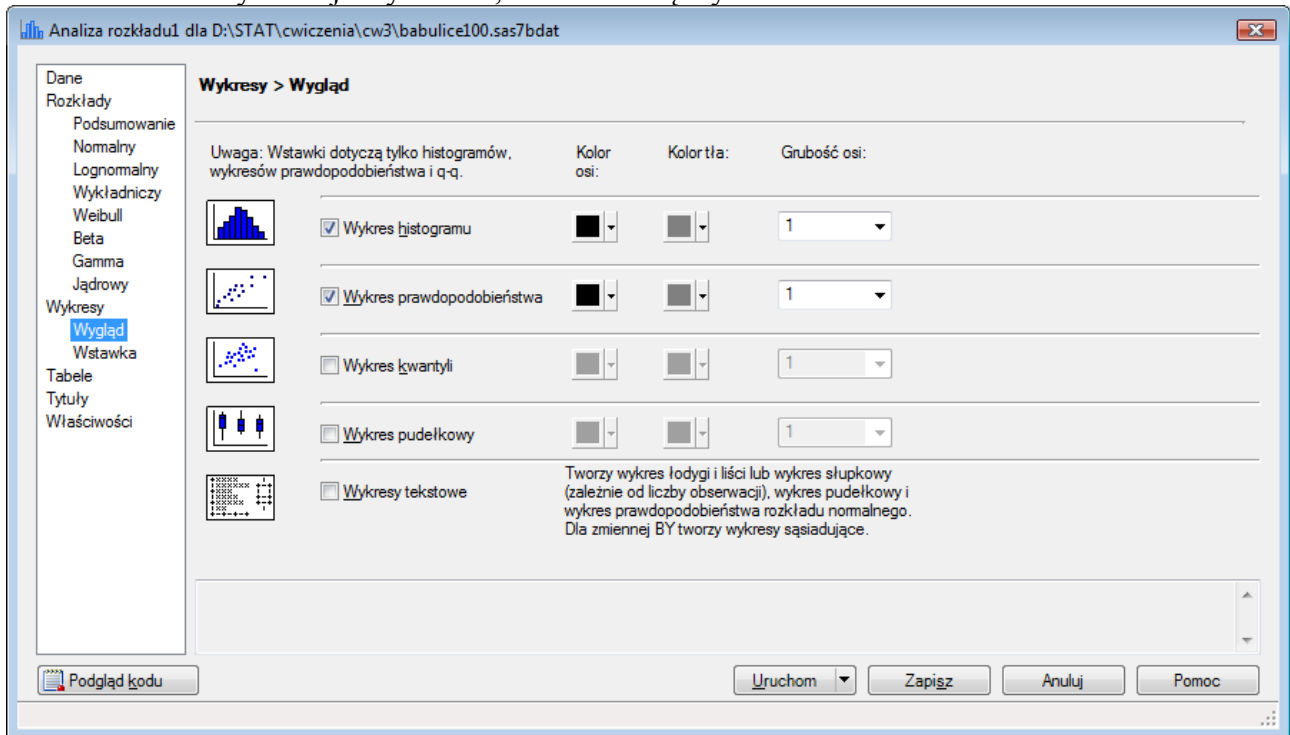




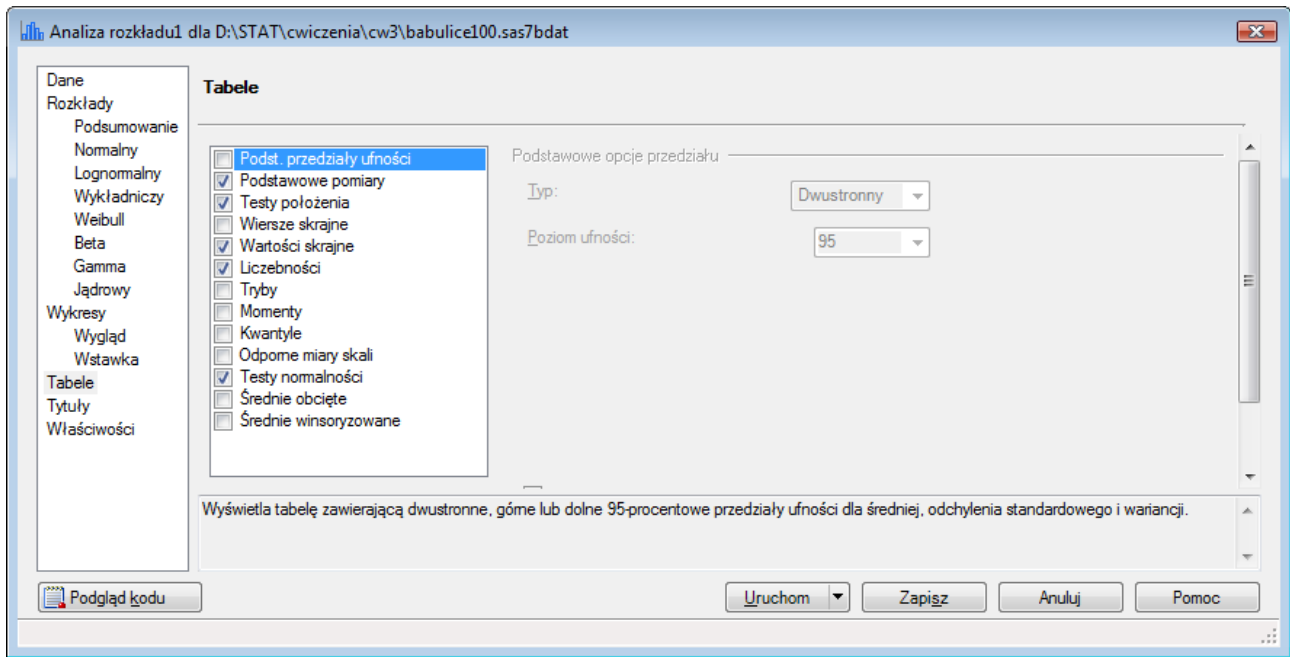
Krok 2. Wskazujemy rozkład, do którego odniesiemy rozkład naszej zmiennej.



Krok 3. Określamy rodzaje wykresów, które zostaną wykonane.



Krok 4. Zaznaczamy statystyki, które będą obliczane (**Testy normalności**).



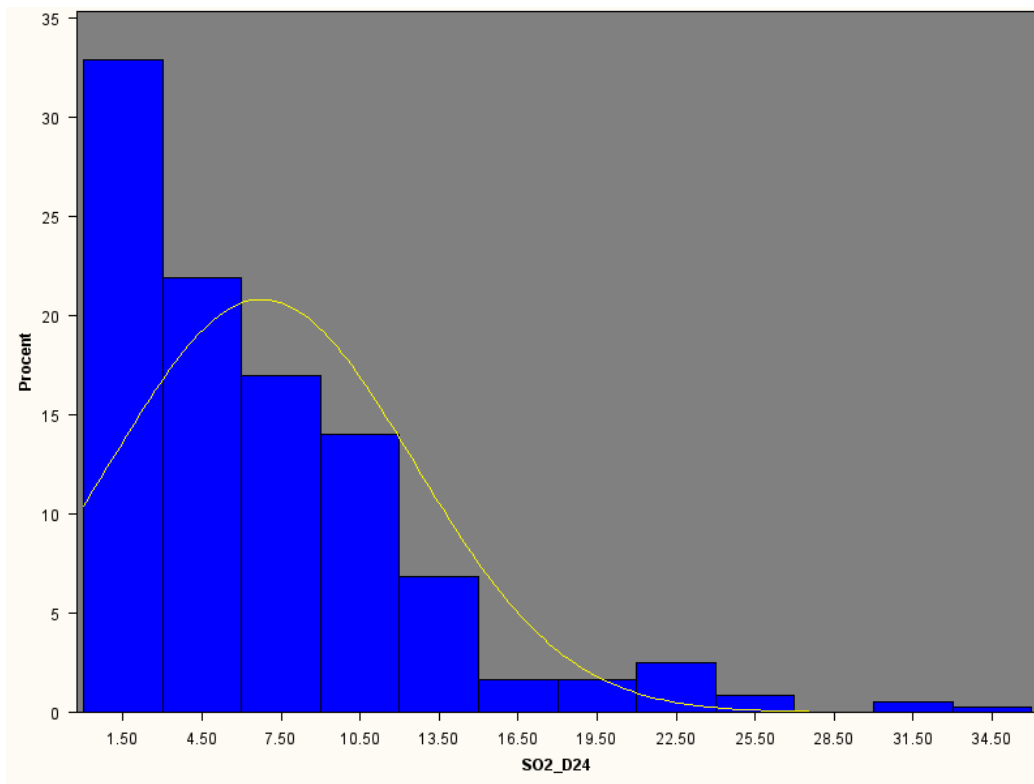
Klikamy na **Uruchom**

## WYNIKI KAMIONKI

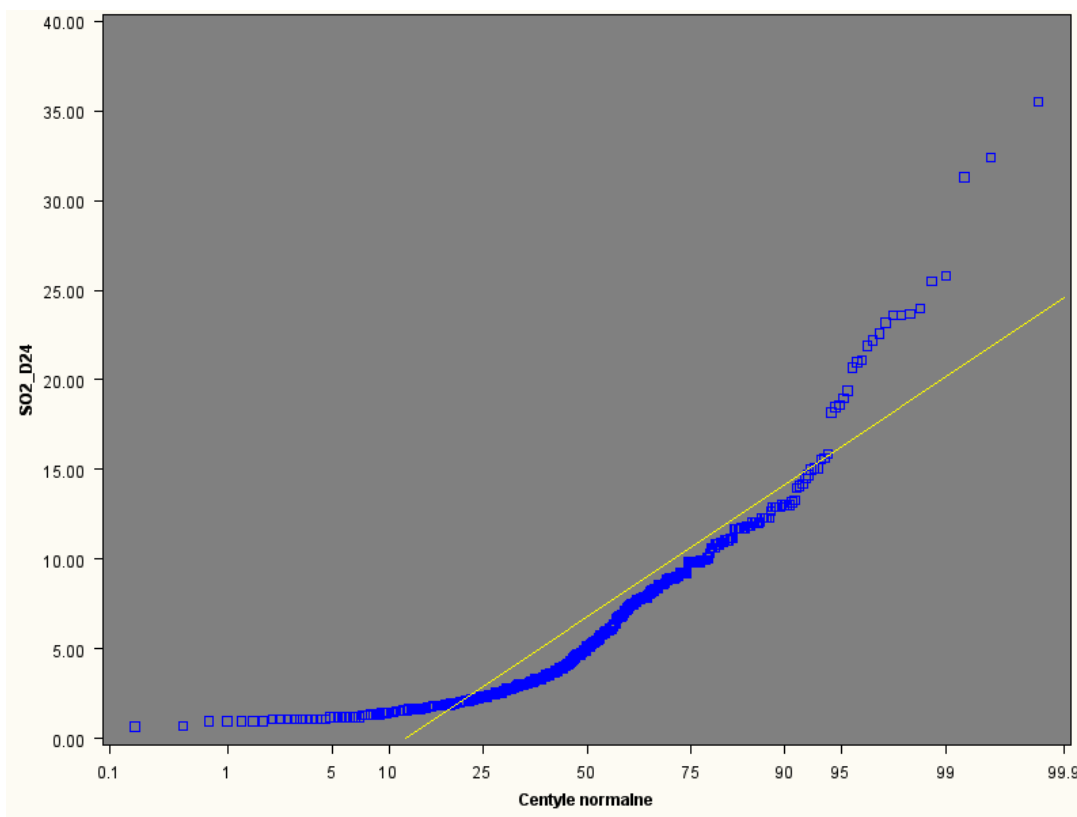
| Testy normalności         |                 |          |                          |         |
|---------------------------|-----------------|----------|--------------------------|---------|
| Test                      | Statystyka      |          | Wartość p                |         |
| <b>Shapiro-Wilk</b>       | <b>W</b>        | 0.837621 | <b>Pr. &lt; W</b>        | <0.0001 |
| <b>Kolmogorov-Smirnov</b> | <b>D</b>        | 0.152429 | <b>Pr. &gt; D</b>        | <0.0100 |
| <b>Cramer-von Mises</b>   | <b>W-kwadr.</b> | 2.197638 | <b>Pr. &gt; W-kwadr.</b> | <0.0050 |
| <b>Anderson-Darling</b>   | <b>A-kwadr.</b> | 14.31463 | <b>Pr. &gt; A-kwadr.</b> | <0.0050 |

W celu stwierdzenia czy badana cecha posiada rozkład zgodny z normalnym patrzymy na wartość „p”. Jeśli ta wartość jest równa lub mniejsza aniżeli 0,05, to mamy podstawę do odrzucenia hipotezy zerowej i stwierdzenia, iż badana cecha nie posiada rozkładu zgodnego z normalnego.

W powyższym przykładzie wszystkie testy normalności rozkładu dowodzą, iż istnieją podstawy do odrzucenia hipotezy zerowej. Oznacza to, że rozkład SO<sub>2</sub> i Pyłu jest niezgodny z rozkładem normalnym. Obydwie cechy posiadają rozkłady asymetryczne – dodatnie (prawostronne).

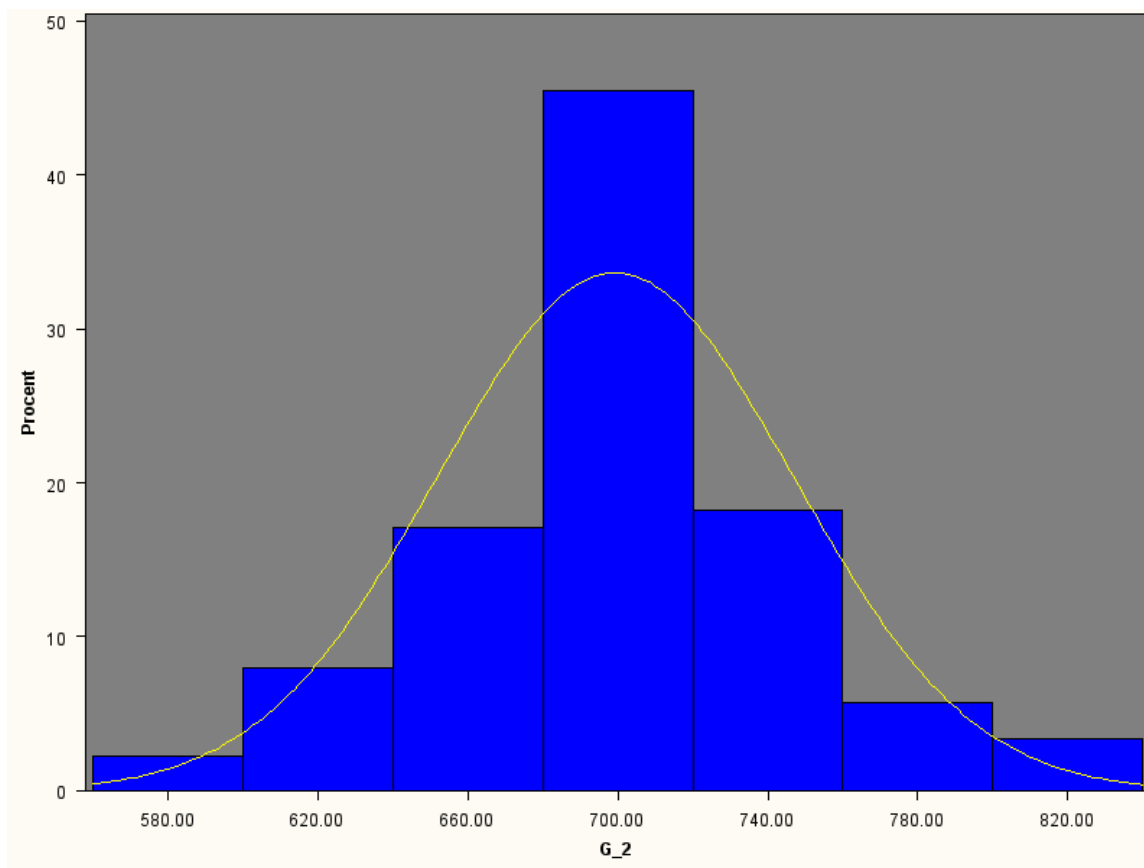


Rycina 1. Histogram



Rycina 2. Wykres normalności kwantyl-kwantyl

Niżej umieszczony histogram został wykonany na podstawie zmiennej G\_2, tj. grubości rogówki oka ludzkiego – rozkład tej zmiennej jest zgodny z normalny.

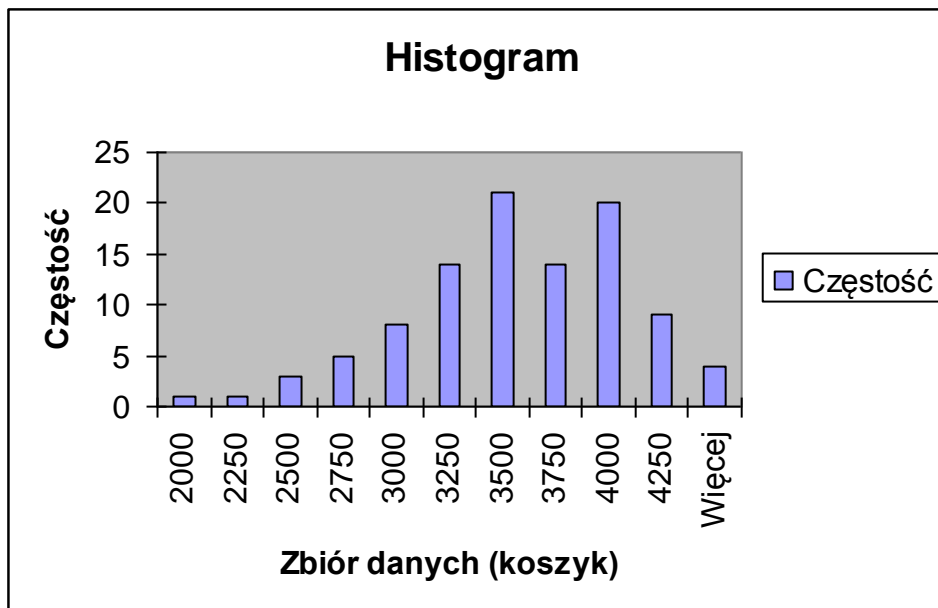


Rycina 3. Histogram – grubość rogówki.

Tabela. Testy normalności przeprowadzone w odniesieniu do grubości rogówki

| Testy normalności  |            |          |                |
|--------------------|------------|----------|----------------|
| Test               | Statystyka |          | Wartość p      |
| Shapiro-Wilk       | W          | 0.975714 | Pr. < W        |
|                    |            |          | 0.0970         |
| Kolmogorov-Smirnov | D          | 0.084883 | Pr. > D        |
|                    |            |          | 0.1181         |
| Cramer-von Mises   | W-kwadr.   | 0.148993 | Pr. > W-kwadr. |
|                    |            |          | 0.0241         |
| Anderson-Darling   | A-kwadr.   | 0.828348 | Pr. > A-kwadr. |
|                    |            |          | 0.0327         |

Przykład 2 (MS EXCEL). Wykonaj wykresy typu HISTOGRAM, na którym zaprezentujesz rozkład masy i długości ciała noworodków. W tym celu posłuż się modulem **Analiza danych**.

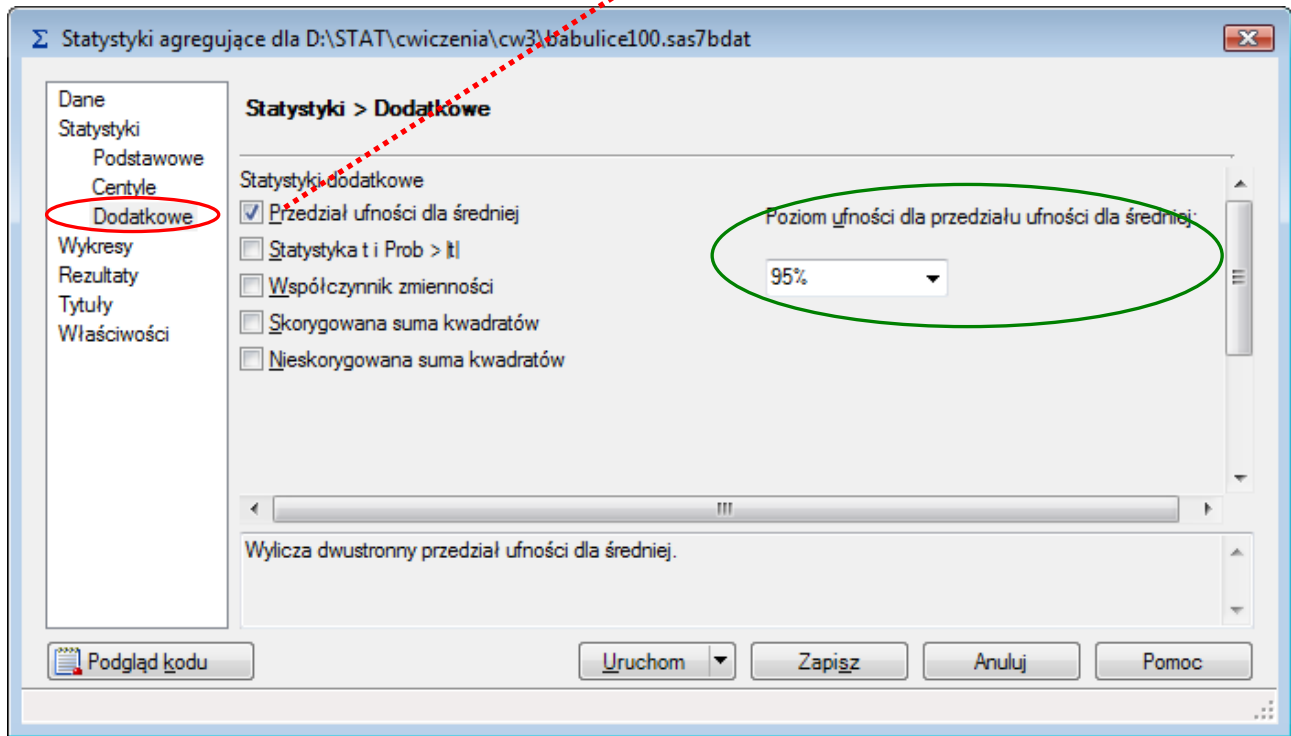


- Zadanie 1.** Sprawdź czy długość, szerokość i masa jaj srok posiadają rozkład zgodny z normalnym (**jajaSROK.xls**).
- Zadanie 2.** Jaki rozkład posiadają liczebności drobnoustrojów (**bakterie100.xls**)? Czy rozkład uległ zmianie po wykonaniu transformacji logarymicznych?
- Zadanie 3.** Czy grubość rogówki posiada rozkład zgodny z normalnym?
- Zadanie 4.** Czy wymiary studentów posiadają rozkład zgodny z normalnym (**stopyX.xls**)?
- Zadanie 5.** Czy masa ciała jagniąt ma rozkład zgodny z normalnym (**MasaCiała.xls**)?
- Zadanie 6.** Czy wydajność mleka posiada rozkład zgodny z normalnym (**krowy.xls**)?

## Wyznaczanie przedziałów ufności

Przykład 1. Wyznacz 95% przedział ufności w odniesieniu do imisji SO<sub>2</sub> i pyłu.

1. Otwórz istniejącą tabelę babulice100.
2. Kliknij na Zadania → Opisz → Statystyki agregujące.
3. Wskaż zmienne do analizy: SO<sub>2</sub>\_D24 oraz Pyl\_D24.
4. W statystykach **Dodatkowe** zaznacz pole wyboru Przedział ufności dla średniej.
5. Kliknij na **Uruchom**.



## WYNIKI

| Zmienna | Średnia    | Odch. std. | Minimum   | Maksimum   | N    | Dolna gr. prz. ufn. 95% | Górna gr. prz. ufn. 95% |
|---------|------------|------------|-----------|------------|------|-------------------------|-------------------------|
| SO2_D24 | 6.6455872  | 5.2040437  | 0.6000000 | 36.7000000 | 1095 | 6.3370108               | 6.9541636               |
| Pyl_D24 | 27.4527158 | 13.4926157 | 5.1000000 | 63.0168000 | 1095 | 26.6526643              | 28.2527672              |

**Odp.:** Wyznaczony przedział ufności pozwala wnioskować z prawdopodobieństwem 95%, że średnia imisja dwutlenku siarki w otoczeniu zakładów przemysłowych o profilu zbliżonym do badanego powinna zawierać się w przedziale 6,34-6,95 µg / m<sup>3</sup>. Należy oczekiwać, że średni poziom pyłu winien zawierać się w przedziale 26,65-28,25 µg / m<sup>3</sup>.



### Zadanie 1 (MS EXCEL).

Przykładowe rozwiązanie dotyczące wyznaczania przedziałów ufności dla średniej arytmetycznej w MS EXCEL znajduje się w pliku kret.xls, arkusz roboczy „ufnosc” S:\OS\_STAT\bazy\_XLS\

1. Otwórz plik roztocze.xls (wcześnie przekopij plik do własnego folderu) zawierający dane dotyczące liczebności roztoczy w pobliżu zakładów produkcyjnych.
2. Skopiuj do kolejnego wolnego arkusza wiersz zawierający nazwy cech.
3. Zmień nazwę tegoż arkusza roboczego na „ufnosc”.
4. Sporządź tabelę zawierającą wybrane miary położenia i zmienności w odniesieniu do liczby roztoczy w próbach z okolic Białowieży (średnia arytmetyczna, odchylenie standardowe).
5. Oblicz lewą i prawą stronę ufności (95%) dla wszystkich rodzajów mechowców i roztoczy.

$$\bar{x} - \left(t_{\alpha} \cdot \frac{S_x}{\sqrt{n}}\right) < \Theta < \bar{x} + \left(t_{\alpha} \cdot \frac{S_x}{\sqrt{n}}\right),$$

gdzie:  $\bar{x}$  – średnia arytmetyczna,  $S_x$  – odchylenie standardowe,  $n$  – liczba próbek.

Wartości krytyczne rozkładu t-Studenta można otrzymać w wyniku zastosowania funkcji: =ROZKŁAD.T.ODW( $\alpha$ ;  $\nu$ ).  $\alpha$  – poziom istotności,  $\nu$  – liczba stopni swobody.

**Zadanie 2 (EG).** Importuj do EG plik **winniczek.xls**, który znajduje się w S:\OS\_STAT\BAZY\_XLS. Skonstruuj przedziały ufności (95 i 99%) dla masy ciała ślimaków, szerokości i wysokości muszli.

**Zadanie 3 (EG).** Skonstruuj przedziały ufności (95 i 99%) dla grubości rogówki (oczy.xls).

**Zadanie 4 (EG).** Skonstruuj przedziały ufności (95 i 99%) dla cech mleczności (krowy.xls).