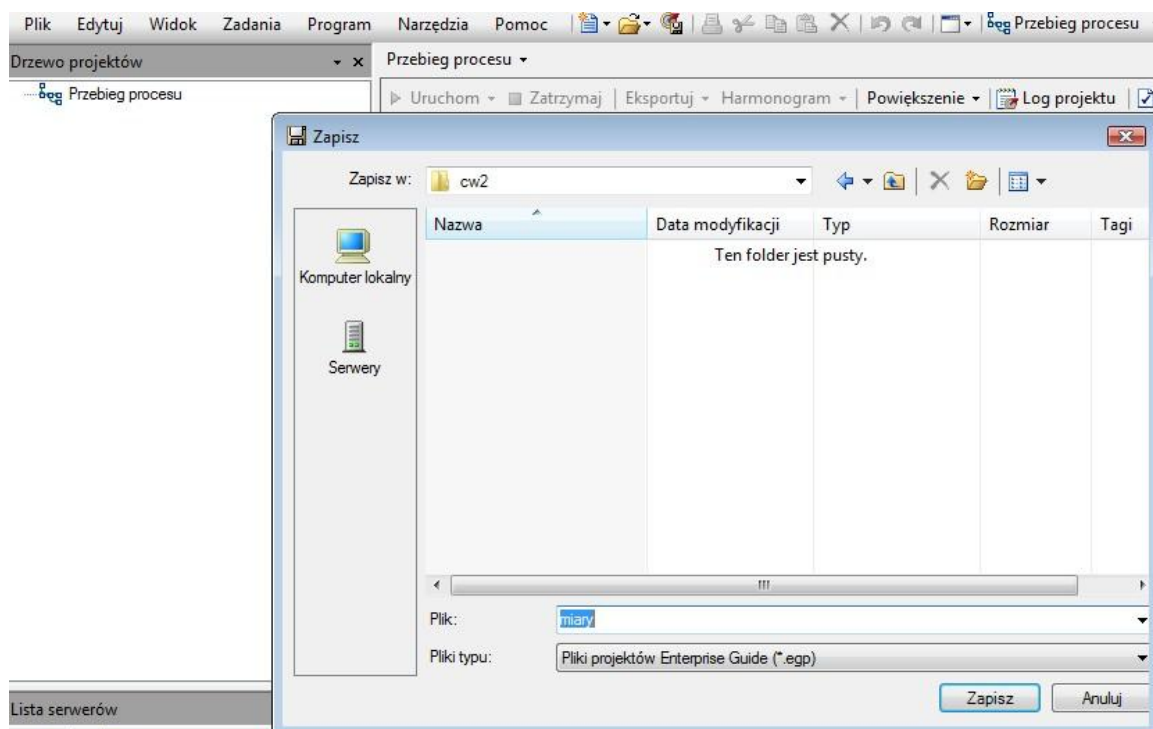


Ćwiczenie:

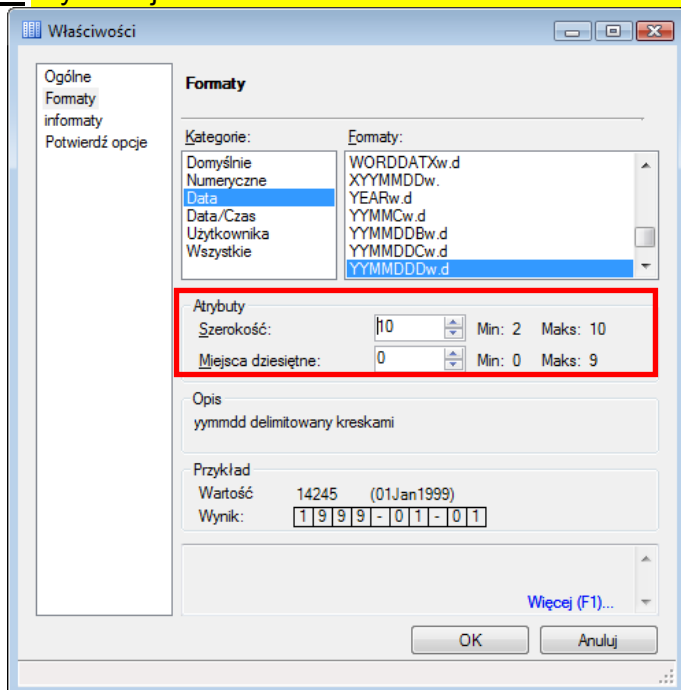
Wprowadzenie do obsługi programu statystycznego SAS Enterprise Guide.
Statystyka opisowa w SAS Enterprise Guide.

1. Załóż we własnym folderze podfolder o nazwie **cw2** i przekopiuj do niego plik **babulice100.xls** z **S:\bazy_XLS**.
2. Uruchom program Enterprise Guide
3. Rozpocznij pracę z nowym projektem (Plik → Nowy → Projekt). Zapisz jego zawartość w założonym folderze **cw2** pod nazwą **miary.egp**.

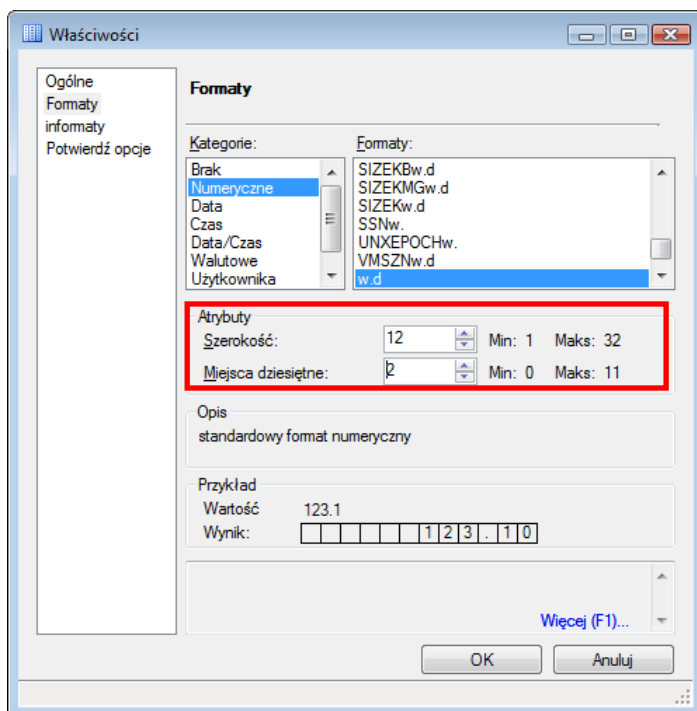


4. Zadeklaruj bibliotekę o nazwie **STAT**. Przypisz ją do lokalizacji własnego folderu **cw2** (Narzędzia → Przypisz bibliotekę projektu).
5. Importuj do EG arkusz roboczy **STAT** ze skoroszytu MS Excel o nazwie **babulice100.xls**. Importowany arkusz zapisz pod nazwą **babulice100** w bibliotece **STAT**.
6. Otwórz a następnie sformatuj kolumnę **Data** tak, aby data była zapisana w formacie YYYY-MM-DD. Następujące po kolumnie **Data** zmienne sformatuj do dwóch miejsc po przecinku. Zanim zaczniesz modyfikować zawartość kolumn zdejmij z nich ochronę: **Edycja → Chroń dane!**

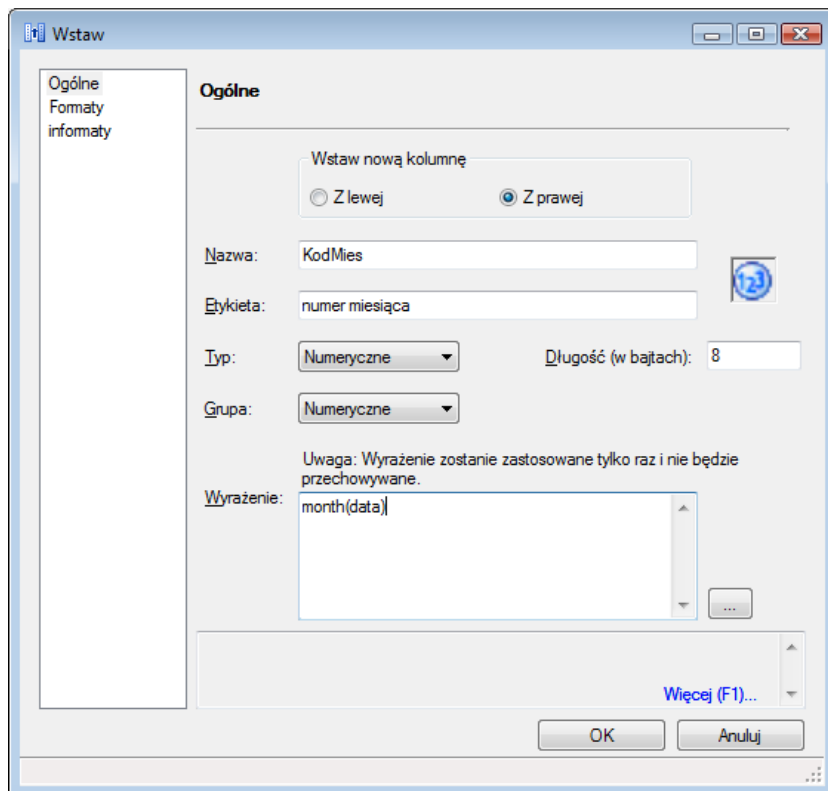
Data Wyszukaj format YYMMDDw.d i zmień szerokość na 10



SO2 D24 Wyszukaj format w.d i ustaw miejsca dziesiętne na 2.



7. Za zmienną **Data** wstaw nową kolumnę o nazwie **KodMies**, która będzie zawierała informację o miesiącu pomiaru wyrażoną w postaci liczby porządkowej. Wprowadź wyrażenie **month(data)**, jak na załączonym zrzucie ekranu.

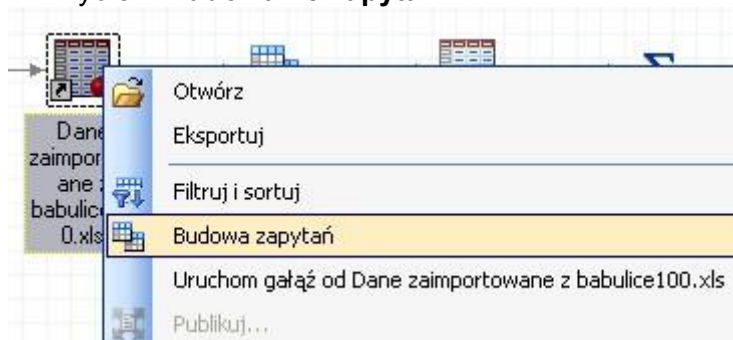


8. W kolejnym kroku utwórz kolumnę wyliczaną o nazwie **PoraRoku**, która zawierać będzie informację o porze roku, w jakiej prowadzono pomiary.

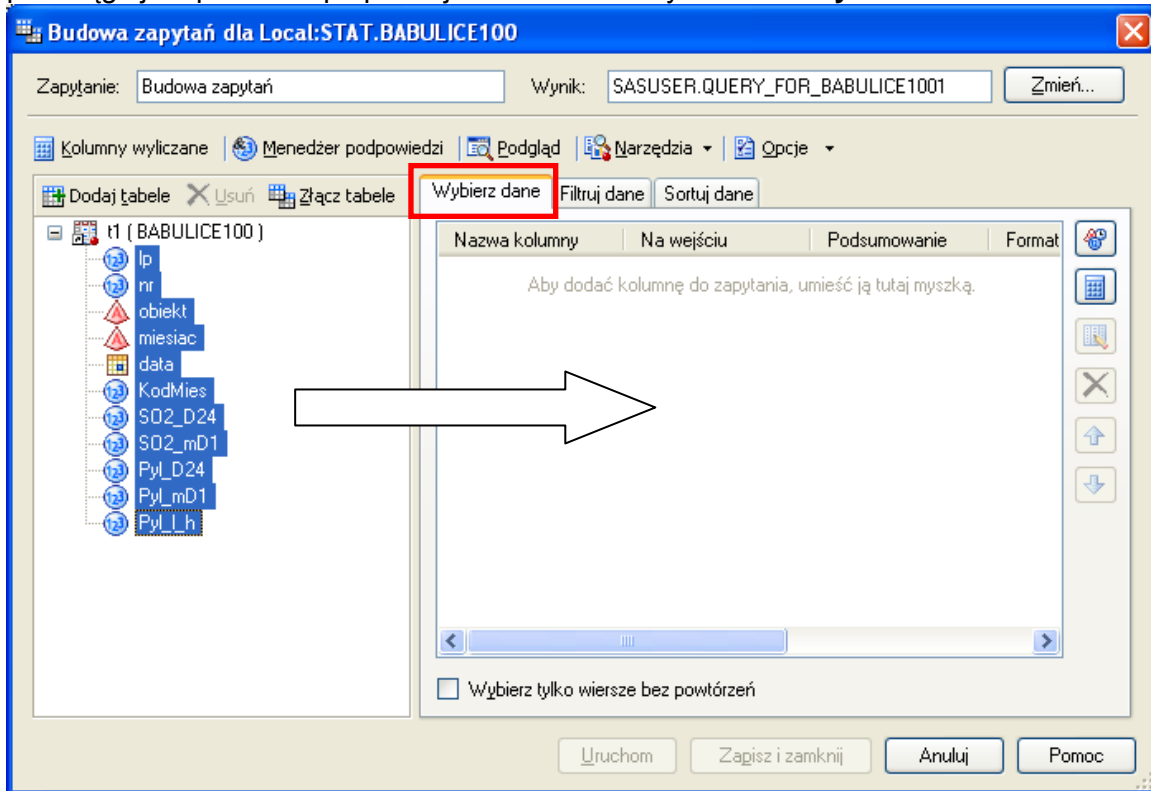
Tabela 1. Kodowanie pór roku

Pora roku	Miesiące (KodMies)
zima	12, 1, 2
wiosna	3, 4, 5
lato	6, 7, 8
jesień	9, 10, 11

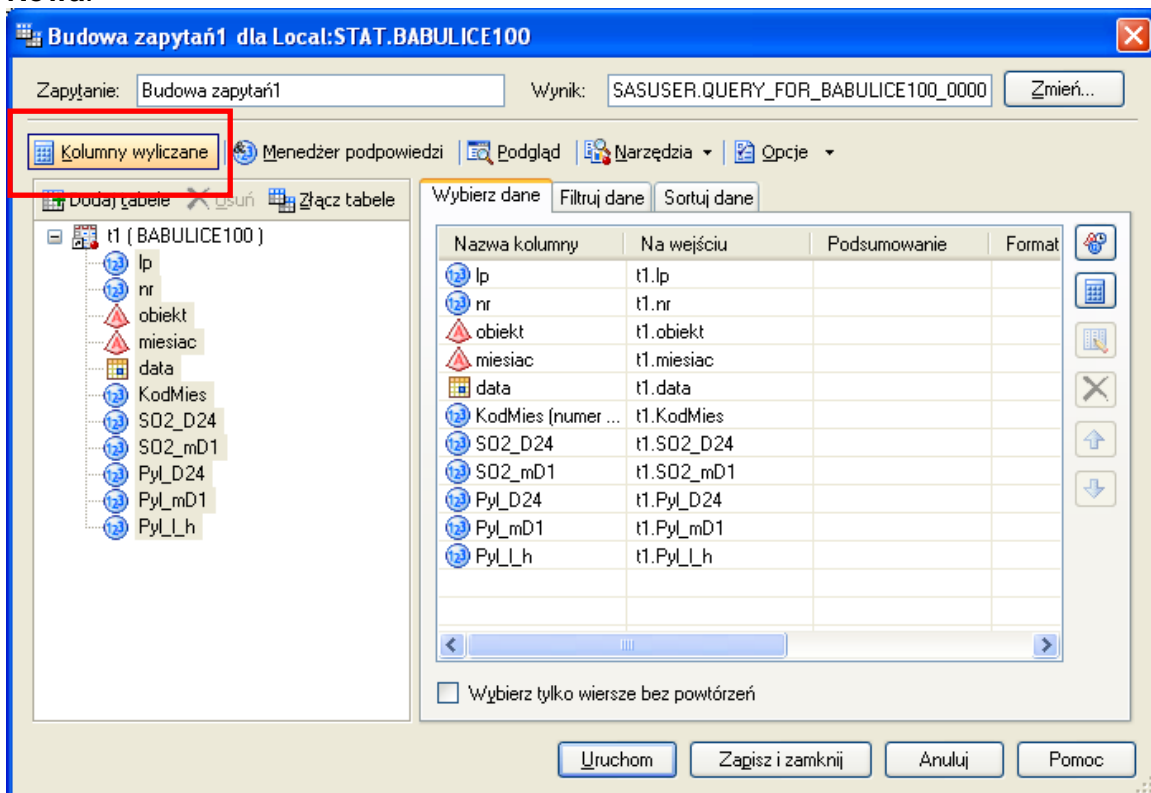
W tym celu zaznacz na diagramie projektu tabelę **babulice100**, a następnie z pop-menu wybierz **Budowanie zapytań!**



W kolejnym kroku zaznacz zmienne wchodzące w skład tabeli babulice100 a następnie przeciągnij na położone po prawej stronie okno zatytułowane **Wybierz dane**

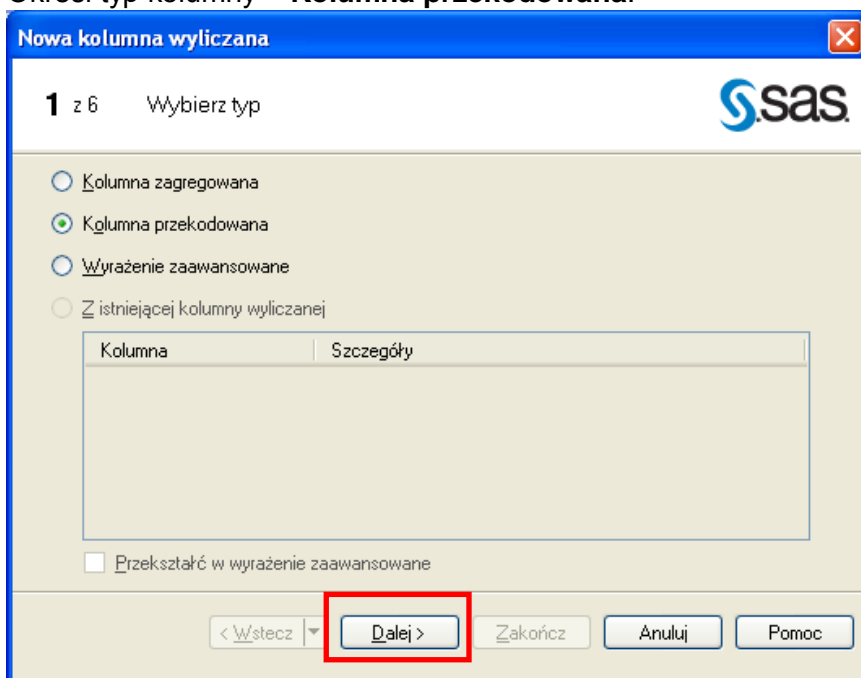


Widok po przeciągnięciu zmiennych. Kliknij na przycisk **Kolumny wyliczane**, a potem na **Nowa**.

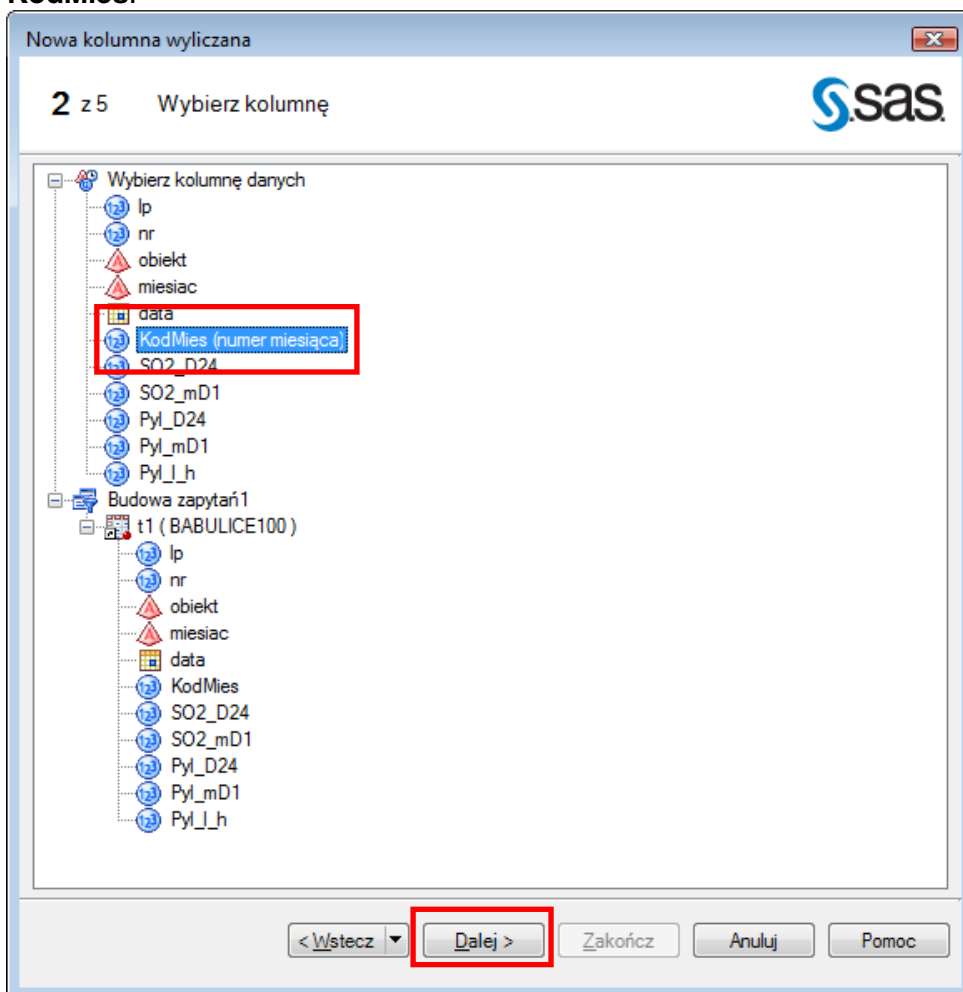




Określ typ kolumny – **Kolumna przekodowana**.



Wybierz kolumnę, na podstawie której zamierzasz utworzyć zmienną **PoraRoku**, tj. **KodMies**.



Na kolejnym etapie kliknij na **Dodaj...**

Nowa kolumna wyliczana

3 z 5 Podaj sposób zastępowania

Zastępowanie

Zastap	na
--------	----

Inne wartości

Pozostałe wartości zastap:

Wartością bieżącą

Brakiem danych

Podaj wartość:

...

Ujmij wartość w cudzysłów

Typ kolumny

Znakowy

Numeryczny

Dodaj... Edytuj... Usuń

< Wstecz Dalej > Zakończ Anuluj Pomoc

W okienku, które się pojawi kliknij na zakładkę **Warunek zastąpienia**. Wybierz za pomocą pola kombi operator „**Na liście**”. Według schematu podanego w tabeli 1 zakoduj pory roku.

Podaj sposób zastępowania

Zastap wartości Zastap przedział **Warunek zastąpienia**

Kolumna: t1.KodMies

Operator: **Na liście**

Generuj filtr wartości podpowiedzi (dotyczy tylko typów podpowiedzi)

Wartości: 12, 1, 2

Dodaj Usuń

t1.KodMies IN (12, 1, 2)

Ujmij wartości w cudzysłowy

Następującą wartością:

zima

Ujmij wartość w cudzysłów

OK Anuluj Pomoc

W celu wprowadzenia liczby 12 kliknij na **Dodaj**, następnie wpisz „12”. Ponów działanie, tj. kliknij na **Dodaj**, wpisz „1”, Kliknij na **Dodaj**, wpisz „2”. Wprowadź nazwę pory roku (**Ujmij wartość w cudzysłów**) i ostatecznie kliknij na **OK**.



Efekt naszego dotychczasowego działania.

Nowa kolumna wyliczana

3 z 5 Podaj sposób zastępowania

Zastępowanie

Zastęp	na
IN (12, 1, 2)	'zima'

Inne wartości

Pozostałe wartości zastęp:

Wartością bieżącą

Brakiem danych

Podaj wartość:

...

Ujmij wartość w cudzysłów

Typ kolumny

Znakowy

Numeryczny

< Wstecz Dalej > Zakończ Anuluj Pomoc

Kolejne pory roku zakoduj w analogiczny sposób. Zwróć uwagę, że kolumna zawiera dane tekstowe (znakowe).

Edytuj kolumnę wyliczaną

1 z 3 Podaj sposób zastępowania

Zastępowanie

Zastęp	na
IN (12, 1, 2)	'zima'
IN (3, 4, 5)	'wiosna'
IN (6, 7, 8)	'lato'
IN (9, 10, 11)	'jesień'

Inne wartości

Pozostałe wartości zastęp:

Wartością bieżącą

Brakiem danych

Podaj wartość:

...

Ujmij wartość w cudzysłów

Typ kolumny

Znakowy

Numeryczny

< Wstecz Dalej > Zakończ Anuluj Pomoc



Widok po w zakodowaniu wszystkich pór roku.

Zastęp	na
IN (12, 1, 2)	'zima'
IN (3, 4, 5)	'wiosna'
IN (6, 7, 8)	'lato'
IN (9, 10, 11)	'jesień'

Other options: Wartością bieżącą, Brakiem danych, Podaj wartość: [], Ujmij wartość w cudzysłowach.

Typ kolumny: Znakowy, Numeryczny.

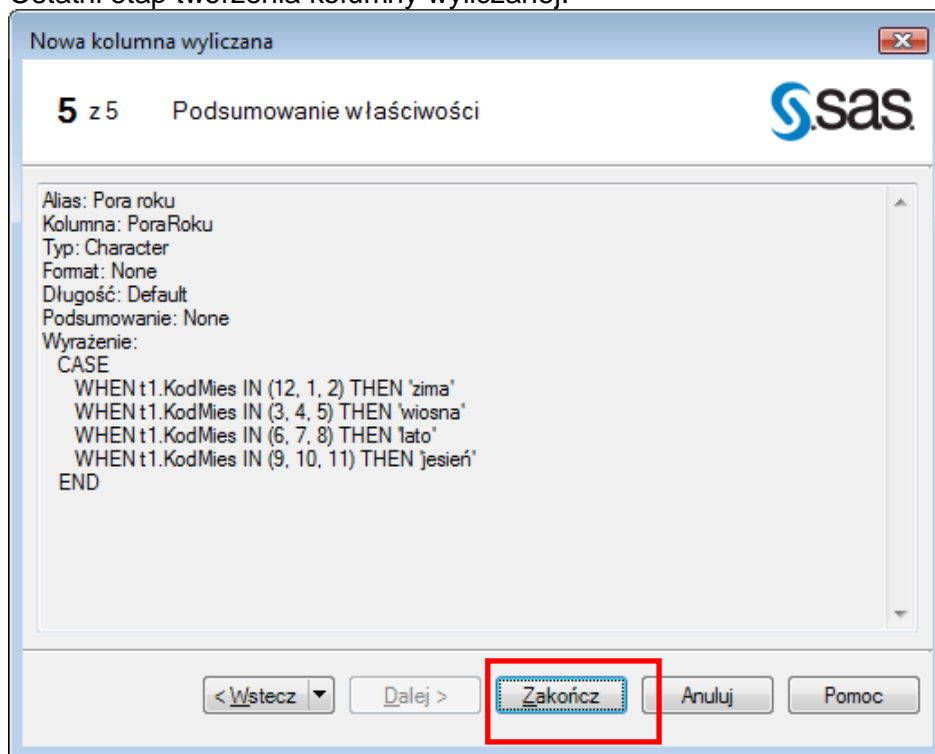
W kolejnym okienku kreatora wpisz nazwę i alias tworzonej kolumny. Zmieniamy format kolumny na **None** (Brak).

Kolumna: PoraRoku
Alias: Pora roku
Podsumowanie: NONE Długość (w bajtach):
Wyrażenie: CASE
WHEN t1.KodMies IN (12, 1, 2) THEN 'zima'
WHEN t1.KodMies IN (3, 4, 5) THEN 'wiosna'
Typ kolumny:
 Znakowy
 Numeryczny
Format: None

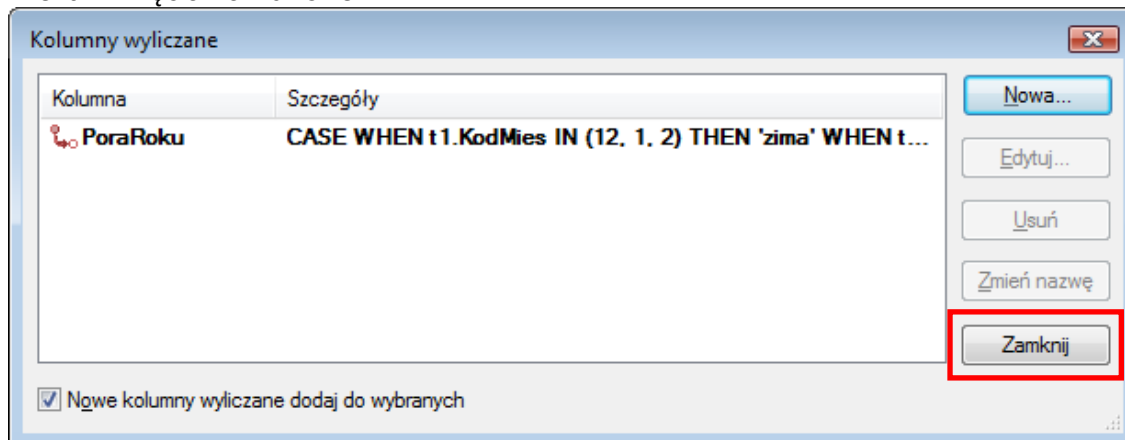
The 'Zmień...' button is highlighted with a red box.



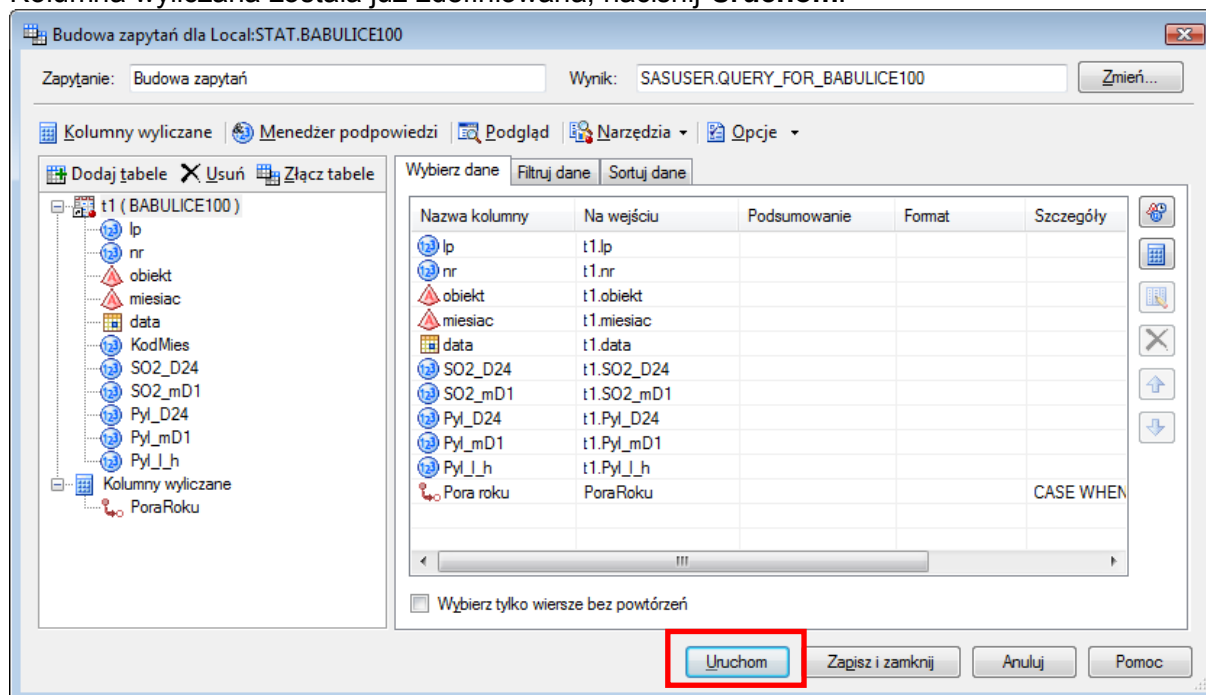
Ostatni etap tworzenia kolumny wyliczanej.



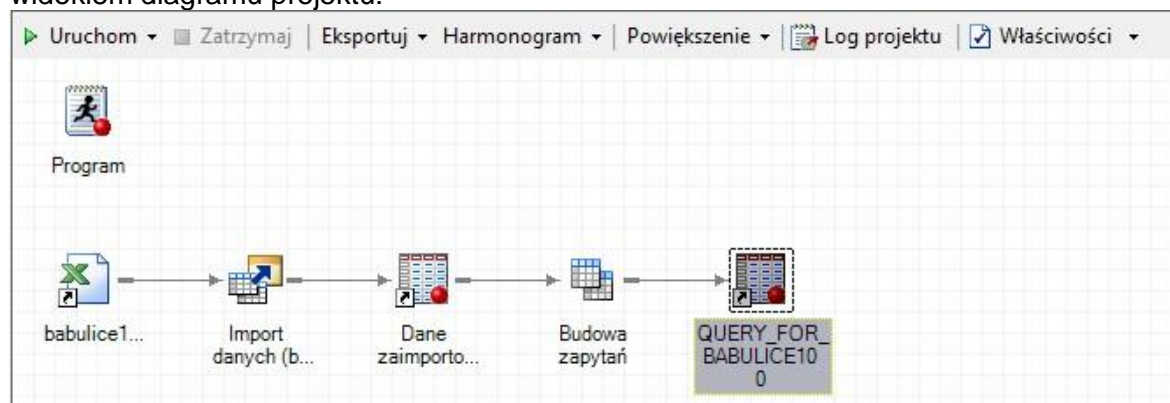
Efekt kliknięcia na **Zakończ**.



Kolumna wyliczana została już zdefiniowana, naciśnij **Uruchom**.



Zakończenie czynności związanych z tworzeniem kolumny wyliczanej skutkuje następującym widokiem diagramu projektu:



Fragment tabeli, która powstała na podstawie **babulice100** i zawierającej zmienną Pora roku.

Pyl_mD1	Pyl_I_h	Pora roku
96.8	999.0	zima
57.2	1,000.0	zima
27.9	998.0	zima
26.9	1,000.0	zima
18.0	1,000.0	zima
16.6	1,000.0	zima
47.5	1,000.0	zima
59.0	999.0	zima

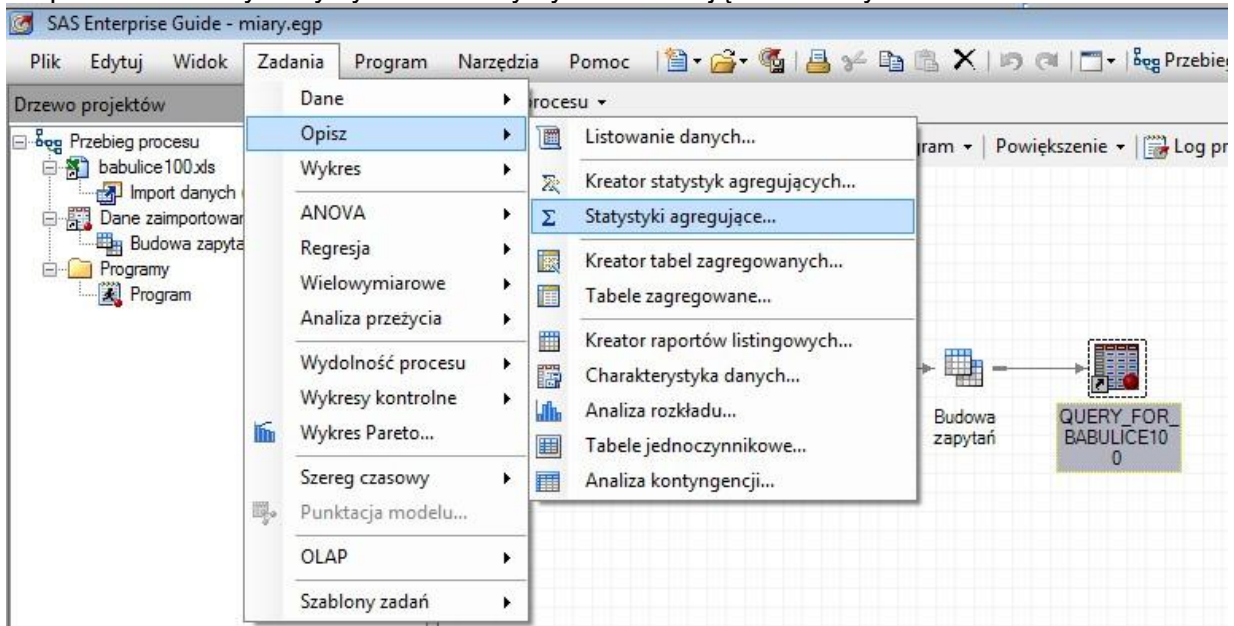


Zadania do realizacji:

Zadanie 1 (Przykład).

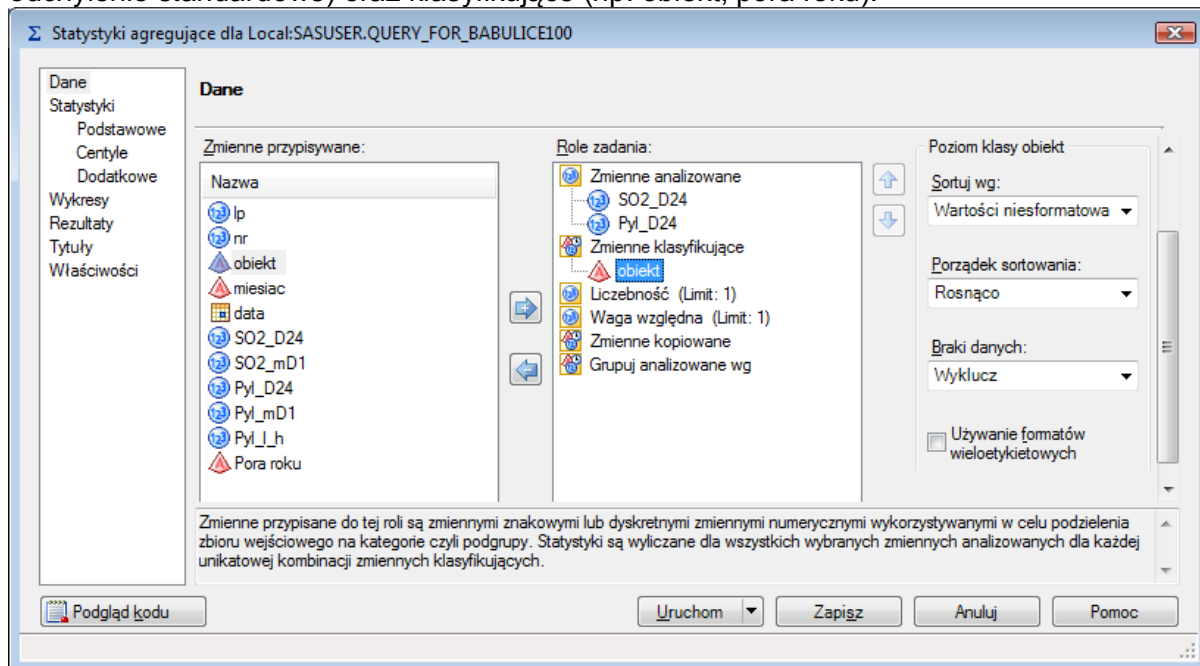
Oblicz podstawowe miary położenia i zmienności w zakresie wybranych wskaźników zanieczyszczenia powietrza. Uwzględnij podział bazy danych ze względu na obiekt i porę roku (tę część zadania wykonaj już samodzielnie). Uzyskane wyniki poddaj wnikliwej interpretacji słownej! Sporządź wykres słupkowy prezentujący średnie stężenie SO_2 w zależności od obiektu i pory roku.

Wspomniane miary statystyczne obliczymy uruchamiając stosowny kreator:

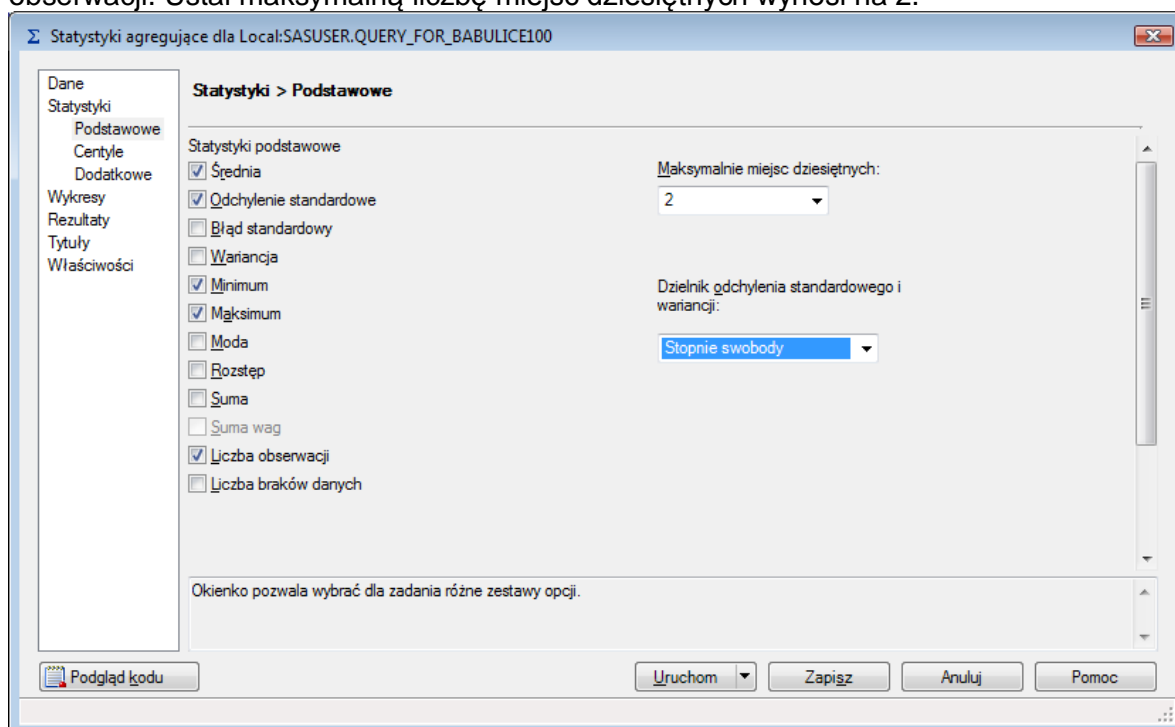




W pierwszym kroku wskaż zmienne analizowane (te dla których obliczamy średnią, odchylenie standardowe) oraz klasyfikujące (np. obiekt, pora roku).

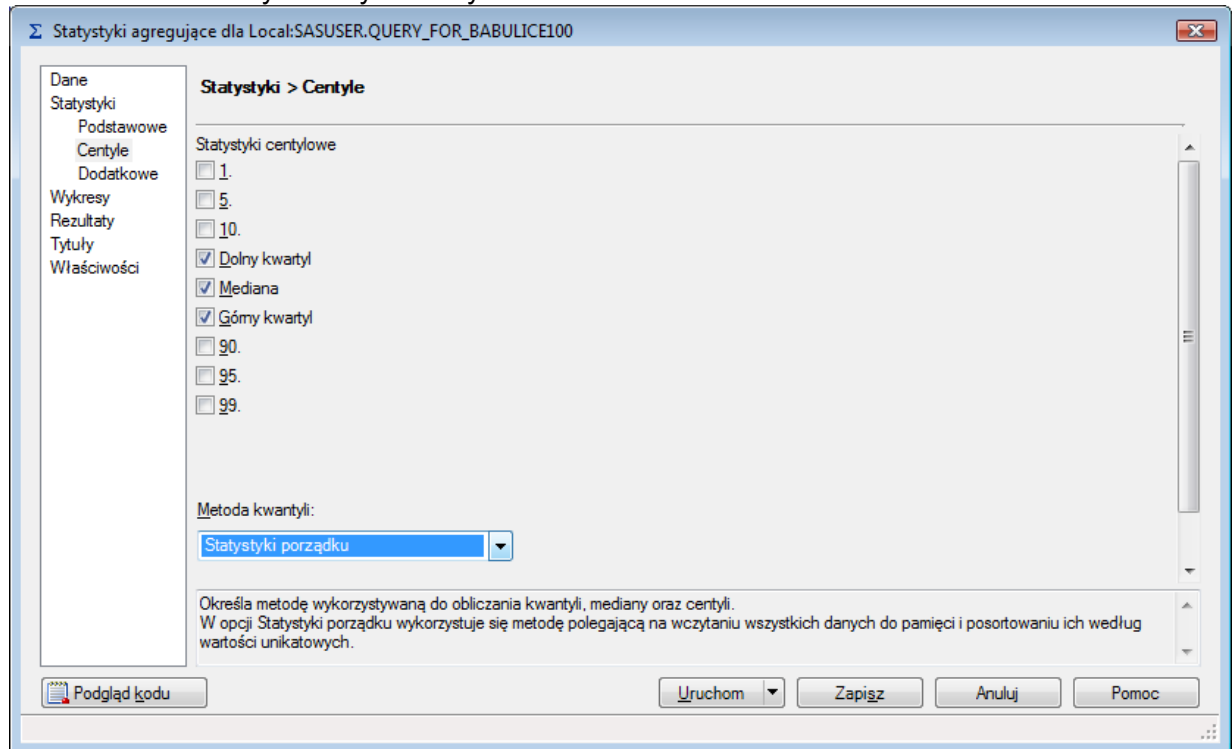


Zaznacza wybrane statystyki: średnia, odchylenie standardowe, minimum, maksimum, liczba obserwacji. Ustal maksymalną liczbę miejsc dziesiętnych wynosi na 2.

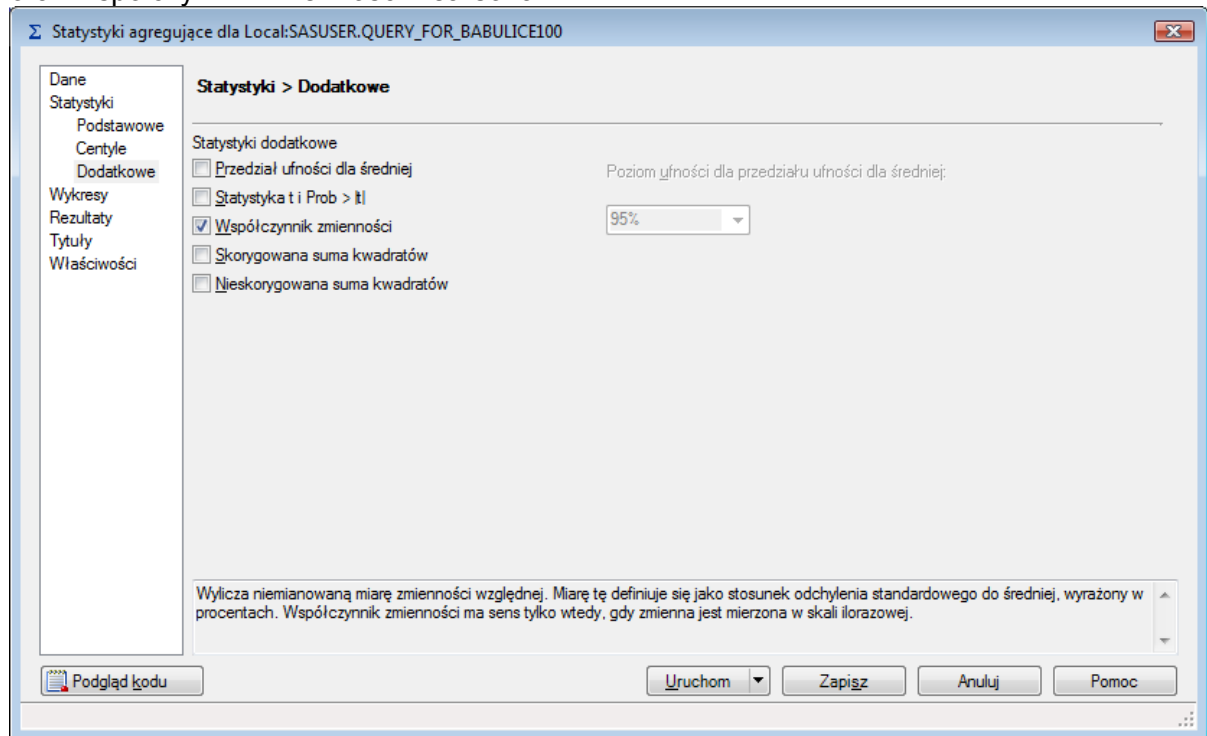




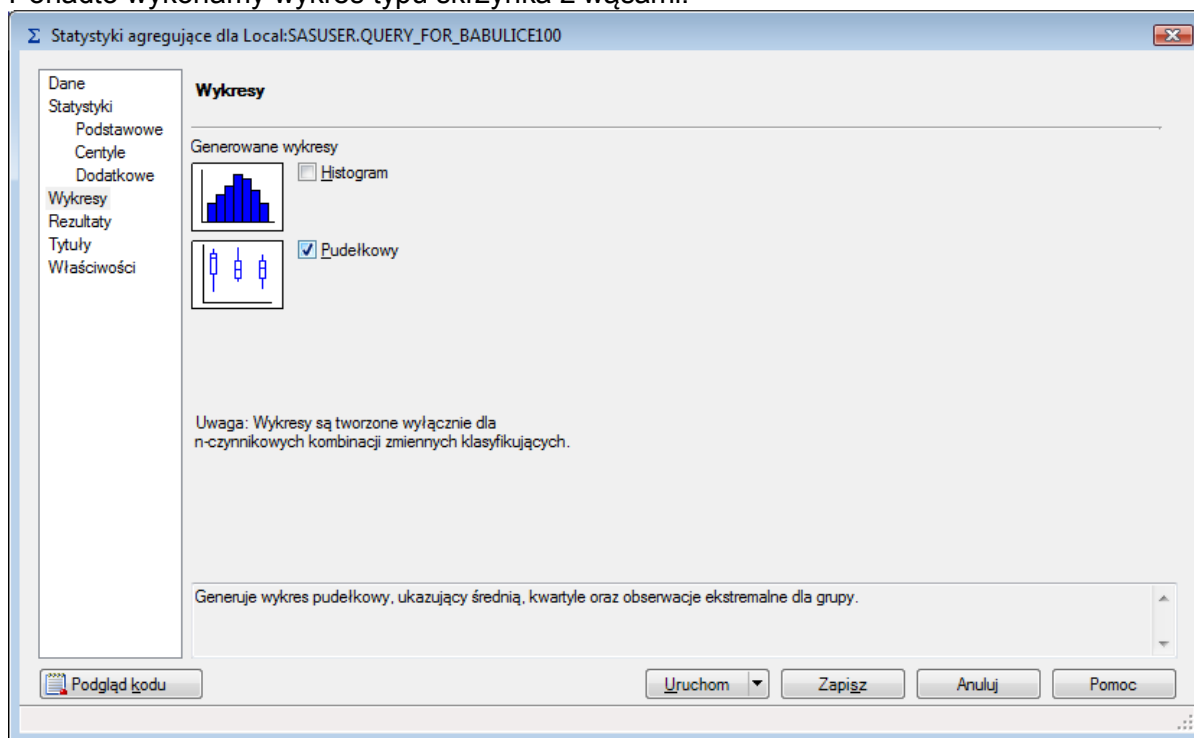
Dodatkowo możemy obliczyć kwartyle...



oraz współczynnik zmienności Pearsona.



Ponadto wykonamy wykres typu skrzynka z wąsami.



Wyniki

Statystyki agregujące

Rezultaty

Procedura MEANS

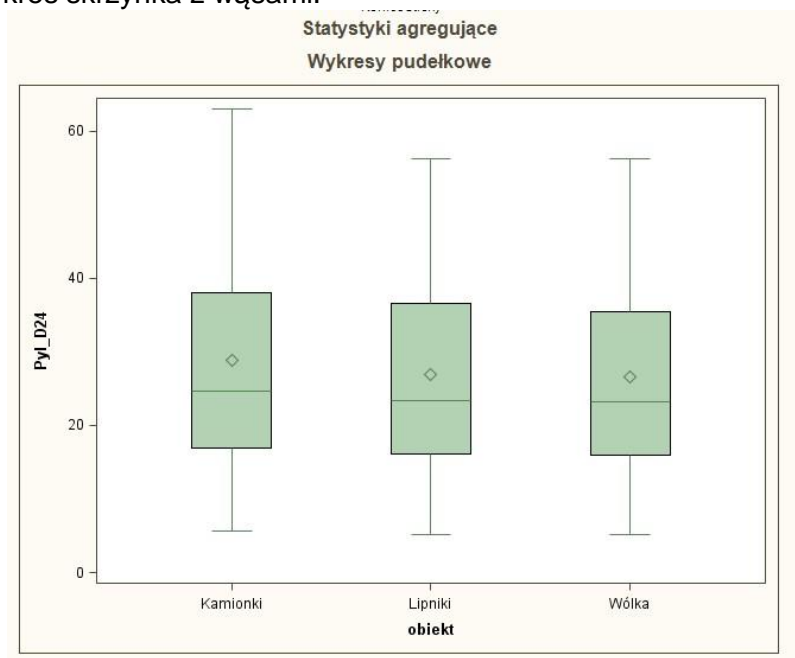
obiekt	N	Zmienna	Srednia	Odch. std.	Minimum	Maksimum	N	Dolny kwartyl	Mediana	Górny kwartyl	Wsp. zmienności
Kamionki	365	SO2_D24	6.78	5.75	0.67	35.50	365	2.34	5.17	9.86	84.87
		Pył_D24	28.81	14.35	5.71	63.02	365	16.91	24.64	38.07	49.82
Lipniki	365	SO2_D24	5.18	3.35	0.60	13.97	365	2.09	4.62	8.00	64.72
		Pył_D24	26.94	13.06	5.10	56.27	365	16.06	23.40	36.60	48.48
Wólka	365	SO2_D24	7.98	5.75	1.87	36.70	365	3.54	6.37	11.06	72.11
		Pył_D24	26.61	12.95	5.10	56.27	365	15.95	23.21	35.40	48.68

Stwierdzono znaczne zróżnicowanie imisji dwutlenku siarki rejestrowanego przez stacje pomiarowe. Najwyższe, przeciętne, dobowe stężenie SO₂ (7,98 µg/m³) zarejestrowano w stacji pomiarowej Wólka. Z kolei najniższy poziom tego związku zanotowano w miejscowości Lipniki (5,18 µg/m³).

Biorąc pod uwagę stację pomiarową Kamionki możemy wnioskować, że spośród 365 wykonanych pomiarów, 25% nie przekroczyło w zakresie stężenia SO₂ poziomu 2,34 µg/m³. W 50% przypadków przeciętne stężenie nie przekroczyło poziomu 5,17 µg/m³. Natomiast w 25% przypadków przeciętne, dobowe stężenie SO₂ było wyższe niż 9,86 µg/m³. Minimalne, przeciętne, dobowe stężenie SO₂ wyniosło 0,67 µg/m³, zaś maksymalne – 35,50 µg/m³.

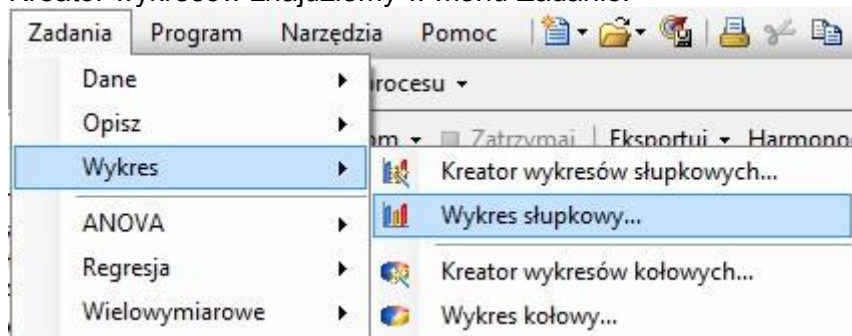
Analizując obliczone miary rozproszenia można stwierdzić, że stężenie charakteryzowało się znacząco zmiennością. Wartość odchylenia standardowego wahała się w przedziale od 3,35 do 5,75 µg/m³. Wysokie wartości przyjmował współczynnik zmienności Pearsona: 64,72-84,87%. Największe zróżnicowanie pomiarów zanotowano w stacji pomiarowej Kamionki, najniższe w okolicach miejscowości Lipniki.

Wykres skrzynka z wąsami.



Tworzenie wykres słupkowego

Kreator wykresów znajdziemy w menu Zadanie.





Ustal zmienne konieczne do sporządzenia wykresu.

Wykres słupkowy dla Local:SASUSER.QUERY_FOR_BABULICE100

Dane

Źródło danych: Local:SASUSER.QUERY_FOR_BABULICE100
Filtr zadań: Brak

Kolumny przypisywane:

Nazwa
lp
nr
objekt
miesiac
data
SO2_D24
SO2_mD1
Pyl_D24
Pyl_mD1
Pyl_J_h
Pora roku

Role zadania:

- Wykreślana kolumna (Limit: 1)
- objekt
- Grupuj słupki wg (Limit: 1)
- Pora roku
- Suma (Limit: 1)
- SO2_D24
- Grupuj wykresy wg

Podgląd kodu Uruchom Zapisz

Zmień rodzaj statystyki wyznaczającej słupek (Średnia).

Wykres słupkowy dla Local:SASUSER.QUERY_FOR_BABULICE100

Wygląd > Zaawansowane

Statystyka wyliczająca słupki:

Średnia

Słupki błędów są właściwe tylko wtedy, gdy statystyka wykresu to średnia lub procent, a wykres nie jest grupowanym wykresem słupkowym trójwymiarowym, ani nie należy do typu mającego przypisaną kolumnę

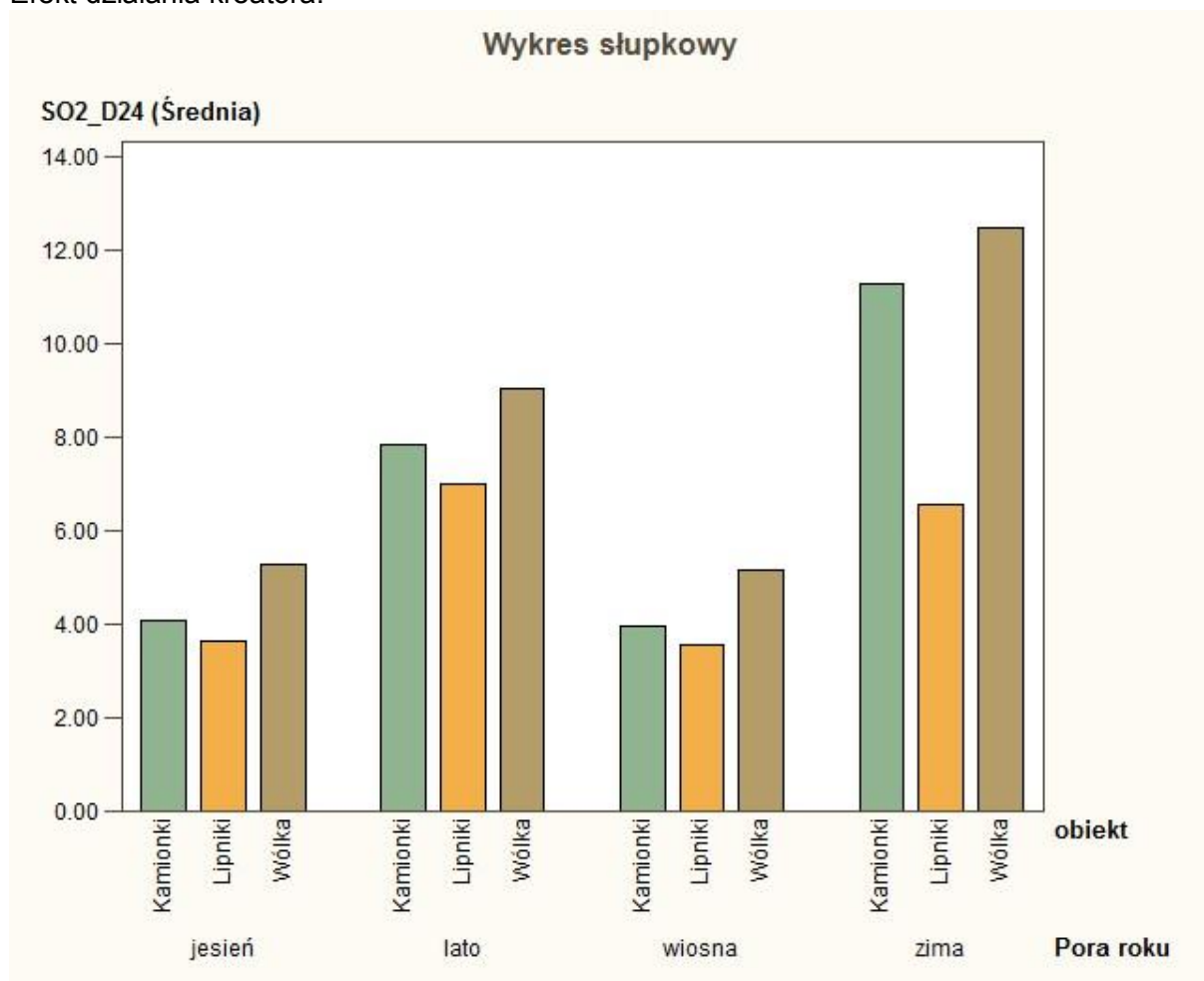
Wyświetl słupki błędów

Typ: CLM:

Oba 95

Kolor słupka błędów:

Efekt działania kreatora:



Zadanie 2. Wykonaj analogiczną analizę w zakresie biologicznego zanieczyszczenia ścieków o różnym pochodzeniu! Wcześniej jednak importuj arkusz roboczy **Dane** skoroszytu Excela o nazwie **bakterie.xls**. Wykonaj transformację logarymiczną zmiennych zawierających liczbę kolonii drobnoustrojów. W wyniku transformacji zostaną dodane nowe kolumny.

Zadanie 3. Importuj do EG zawartość skoroszytu Excela **dziecko100.xls (arkusz DANE)**. Oblicz znane Ci miary położenia klasyczne (średnia arytmetyczna) i pozycyjne (kwartyle, dominanta) oraz miary zmienności (odchylenie standardowe, współczynnik zmienności) w odniesieniu do zmiennej Apgar, masa i długość

Zadanie 4. Importuj do EG zawartość skoroszytu Excela **kret.xls**. Oblicz znane Ci miary położenia klasyczne (średnia arytmetyczna) i pozycyjne (kwartyle, dominanta) oraz miary zmienności (odchylenie standardowe, współczynnik zmienności) w odniesieniu do zmiennych masa i wymiary ciała.