

Wprowadzenie do pakietu STATISTICA



Spis treści

Laboratorium I: Wprowadzenie do pakietu STATISTICA	1
Wiadomości ogólne.....	2
1. Okno podstawowe.	2
1.1 Paski narzędziowe.	2
2. Praca w arkuszu danych.	3
2.1 Poruszanie się po arkuszu:	4
2.2 Zaznaczanie obszarów:.....	4
2.3 Elementarne modyfikowanie arkusza danych	4
3. Wprowadzanie danych.....	5
4. Import i eksport plików	6
5. Operacje na zmiennych i przypadkach.....	7
6. Formuły i ich przeliczanie	8
7. Selekcja przypadków	11
8. Zapisywanie danych	12
Ćwiczenie	13

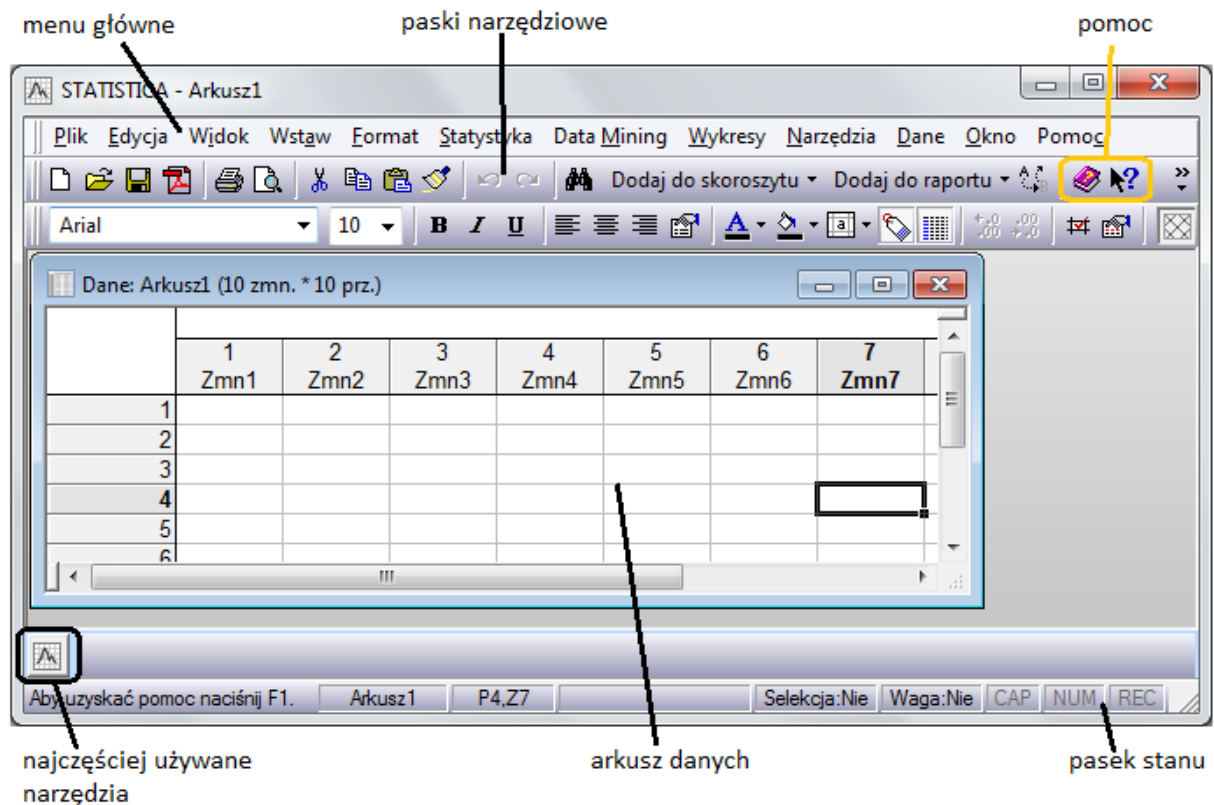
Wiadomości ogólne

Program STATISTICA uruchamia się poprzez kliknięcie w ikonkę:



1. Okno podstawowe.

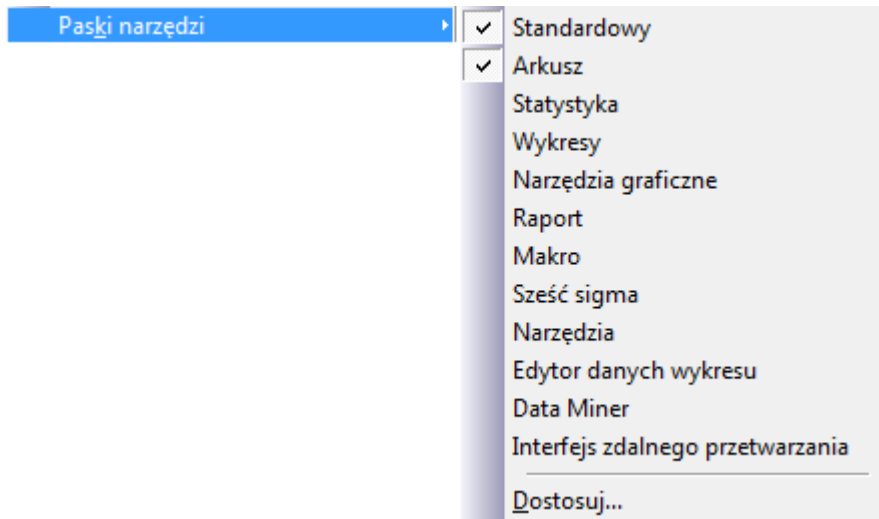
Po uruchomieniu programu STATISTICA wyświetla się okno podstawowe, które przedstawiono na Rys.1. Zawiera m.in. menu główne, paski narzędziowe, arkusz danych, pasek stanu.



Rys.1 Okno podstawowe.

1.1 Paski narzędziowe.

W menu **Widok** można wybrać wyświetlanie różnych pasków narzędziowych:



Rys.2 Wybór wyświetlanych pasków narzędzi.

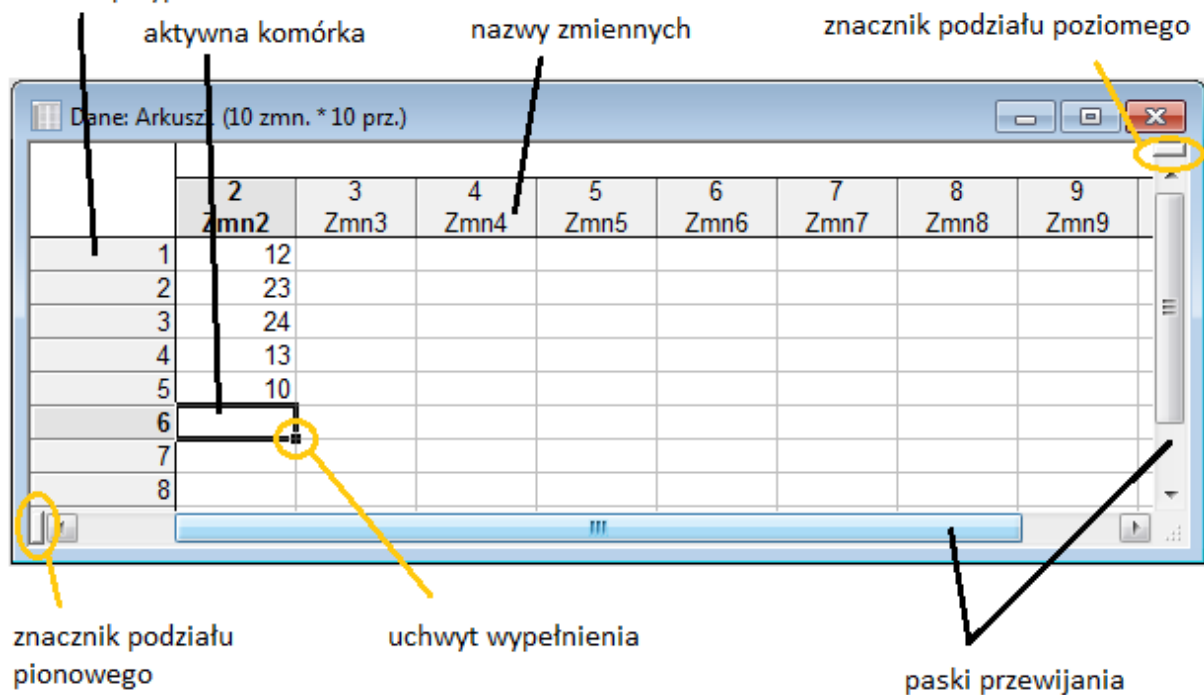
Korzystanie z pasków narzędziowych pozwala na szybkie wybieranie żądanej komendy bez konieczności odwoływania się do menu głównego. Wszystkie paski można dostosowywać do własnych potrzeb, klikając *Dostosuj...* w prawym rogu dowolnego paska:



2. Praca w arkuszu danych.

Aby utworzyć arkusz, należy kliknąć w menu *Plik/Nowy*. Arkusz danych przedstawiono poniżej:

numery przypadków



Rys. 3 Arkusz danych.

Dane w arkuszu zorganizowane są w **przypadki** (wiersze w arkuszu) oraz **zmienne** (kolumny). W pierwszej kolumnie mogą występować numery lub nazwy przypadków, pierwszy wiersz zawiera nazwy zmiennych.

2.1 Poruszanie się po arkuszu:

Poruszanie się po arkuszu może się odbywać za pomocą myszki, strzałek, a także innych przycisków klawiatury:

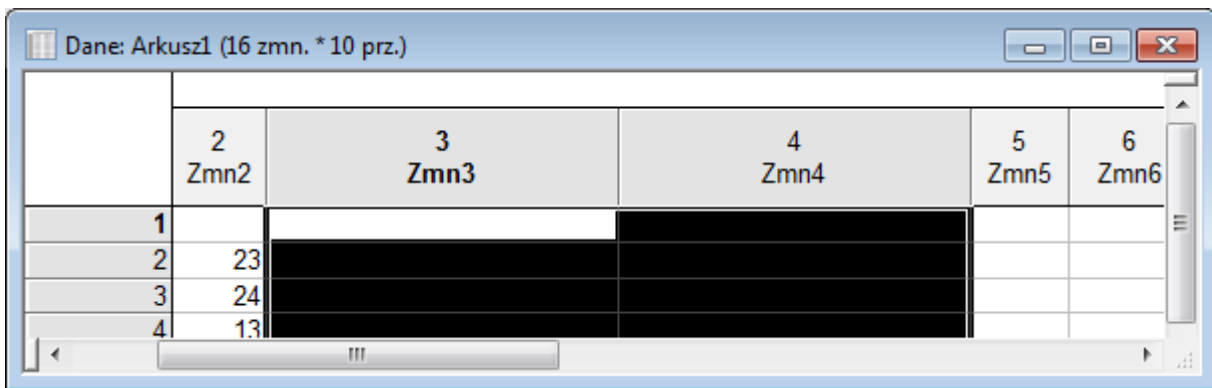
Home	Przemieszczenie do pierwszej zmiennej bieżącego przypadku
End	Przemieszczenie do ostatniej zmiennej bieżącego przypadku
Ctrl + ↑	Przemieszczenie do pierwszego przypadku bieżącej zmiennej
Ctrl + ↓	Przemieszczenie do ostatniego przypadku bieżącej zmiennej

2.2 Zaznaczanie obszarów:

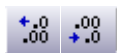
- dowolnie wybrany prostokątny obszar myszką (jak w programie Excel);
- cała kolumna (zmienna) – kliknięcie na nazwę zmiennej;
- cały wiersz (przypadek) – kliknięcie na numer/nazwę przypadku;
- cały arkusz – kliknięcie na polu będącym przecięciem nazw zmiennych i przypadków;
- za pomocą klawiatury: *Shift + strzałki*;

2.3 Elementarne modyfikowanie arkusza danych

Zmiana szerokości kolumn: przy podświetleniu kilku kolumn na raz, zmieniając szerokość jednej, w taki sam sposób zmienia się szerokość pozostałych podświetlonych.



Rys.4 Zmiana szerokości kolumn dwóch zmiennych na raz.



przyciski zwiększenia lub zmniejszenia miejsc dziesiętnych formatu wyświetlanej wartości bieżącej zmiennej;

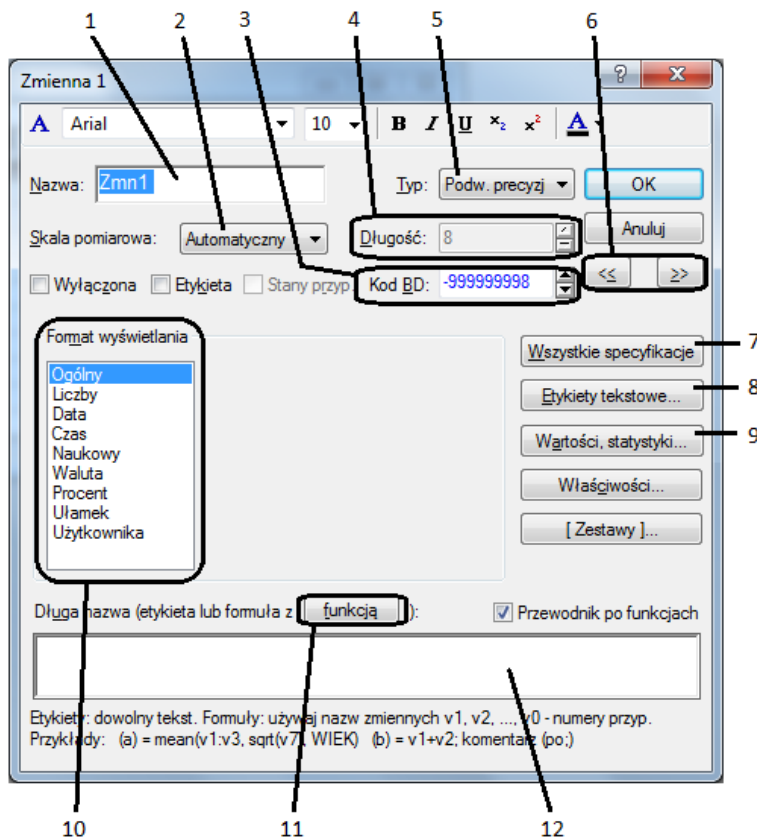
Różne części arkusza można oglądać w oknie na raz po podzieleniu go z wykorzystaniem znaczników poziomego i pionowego podziału (zaznaczone na Rys.3):

	10-pozycyjne badanie: wyniki dla			10-pozycyjne badanie: wyniki dla 40		
	1	2	3	8	9	10
	POZYCJA1	POZYCJA2	POZYCJA3	POZYCJA8	POZYCJA9	POZYCJA10
1	4	5		6	6	3
2	4	3		4	2	4
3	4	4		5	5	3
4	4	3		3	6	4
91	2	4		6	5	5
92	4	4		3	2	2
93	6	6		5	6	7
94	4	6		5	6	4
95	6	5		7	6	3

Rys.5 Arkusz podzielony za pomocą pionowego i poziomego znacznika podziału.


3.Wprowadzanie danych

Każda komórka arkusza może zawierać: liczbę, tekst, datę lub wartości formuł. Wpisywanie danych odbywa się tak samo jak w innych arkuszach kalkulacyjnych. Zmienne definiuje się dwukrotnie klikając myszką na nazwę zmiennej (domyślnie Zmn1, Zmn2.. lub Var1, Var2...). W wyniku dwukrotnego kliknięcia w nazwę zmiennej, otwiera się okno:



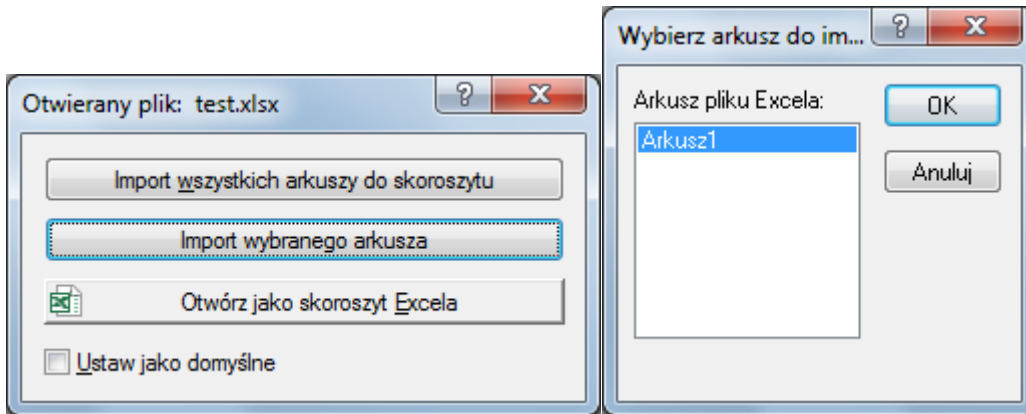
Rys.6 Okno specyfikacji zmiennych.

Przy pomocy oznaczonych pól można wykonać następujące czynności:

- 1) **Nazwa**: modyfikacja nazwy zmiennej;
- 2) **Skala pomiarowa**: pole wyboru skali pomiaru danej zmiennej: niezdefiniowana, automatyczna, ilościowa, jakościowa, porządkowa;
- 3) **Kod BD**: pole braku danych – gdy w danej komórce nie ma danych, pole przyjmuje domyślną wartość -999999998;
- 4) **Długość**: określenie długości danej zmiennej typu tekstowego;
- 5) **Typ**: określenie typu zmiennej: podwójna precyzja, całkowita, tekstowa, bajt;
- 6) **<<>>** : przejście do okna specyfikacji wcześniejszej lub kolejnej zmiennej;
- 7) **Wszystkie specyfikacje**: otwarcie okna specyfikacji wszystkich zmiennych - możliwość przeglądania i edycji wszystkich zmiennych, ich etykiet, formuł, połączeń;
- 8) **Etykiety tekstowe...**: wywołanie menedżera wartości tekstowych do edycji podwójnej notacji zmiennych. Każda wartość w programie STATISTICA może mieć 2 reprezentacje: liczbową np. 1 oraz tekstową np. Żeńska. Dzięki temu, łatwiej wprowadza się dane: przykładowo dla zmiennej płeć szybciej wpisuje się wartości 0 i 1. Następnie po zdefiniowaniu etykiety tekstowej „Żeńska” dla komórek zawierających „1”, wszystkie pola z „1” zostają zastąpione etykietą „Żeńska”;
- 9) **Wartości, statystyki...**: wywołanie okna z posortowaną listą wszystkich wartości zmiennej i z podstawowymi statystykami opisowymi dla tej zmiennej. Przycisk  umożliwia przekopiowanie wartości tych statystyk w dowolne miejsce, np. do sprawozdania;
- 10) **Format wyświetlania**: wybór formatu wyświetlania danych; W przypadku zmiennych określających datę: gdy w kolejnych zmiennych mamy dane określające dzień, miesiąc, rok, przy pomocy funkcji **Operacje na danych** dostępnej w menu **Zmienne** w Pasku zadań można utworzyć nową zmienną zawierającą datę utworzoną z tych 3 zmiennych. W odpowiednie pola okna **Operacje na danych** należy wpisać nazwy zmiennych źródłowych oraz zmiennej docelowej. Przy pomocy tego okna można również dokonać operacji odwrotnej.
- 11) **Funkcja**: wybór funkcji do tworzenia formuł;
- 12) **Długa nazwa**: nadawanie zmiennym dłuższych opisów, tworzenie formuł tworzących nowe zmienne z już istniejących.

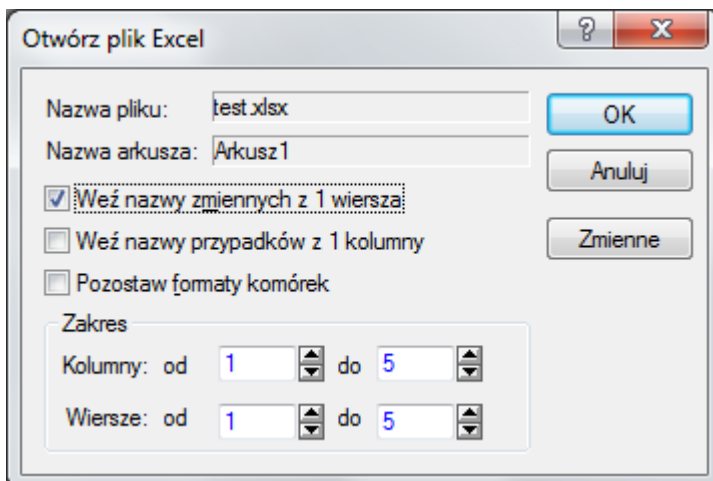
4. Import i eksport plików

W programie STATISTICA można wykorzystywać dane zapisane w różnych innych aplikacjach (arkusze kalkulacyjne, bazy danych). W celu wczytania danych z arkusza kalkulacyjnego utworzonego w programie Excel należy otworzyć odpowiedni plik poprzez **Plik/Otwórz**. W wyniku tego pokazuje się okienka przedstawione na Rys. 7.

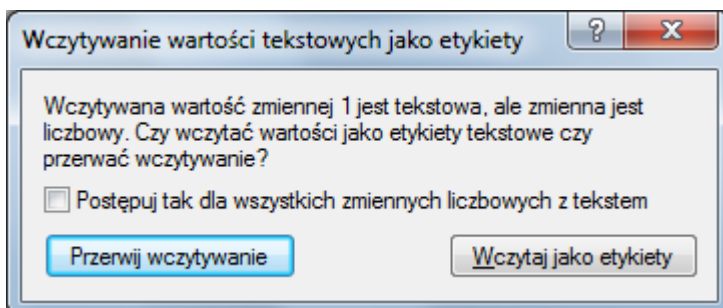


Rys. 7 Importowanie danych z arkusza Excela. Po wciśnięciu *Import wybranego arkusza* (lewe okno), należy wybrać odpowiedni Arkusz (prawe okno).

Należy także zaznaczyć, czy pierwszy wiersz i pierwsza kolumna są nazwami zmiennych i przypadków, a także zatwierdzić zakres wczytywanych danych (Rys.8). Jeśli wczytujemy cały arkusz, nie zmieniając zakresu, należy wcisnąć **OK**.



Rys.8 Jeden z etapów importowania do STATISTIKI danych z arkusza kalkulacyjnego Excel.



5. Operacje na zmiennych i przypadkach

Dla realizacji takich operacji wykorzystujemy przyciski **Zmienne ▾ Przypadki ▾** dostępne z paska narzędzi. Umożliwiają one wykonywanie operacji takich jak m.in.: dodawanie, usuwanie, przenoszenie, kopiowanie, tworzenie nazw, sortowanie. Ważniejsze operacje:

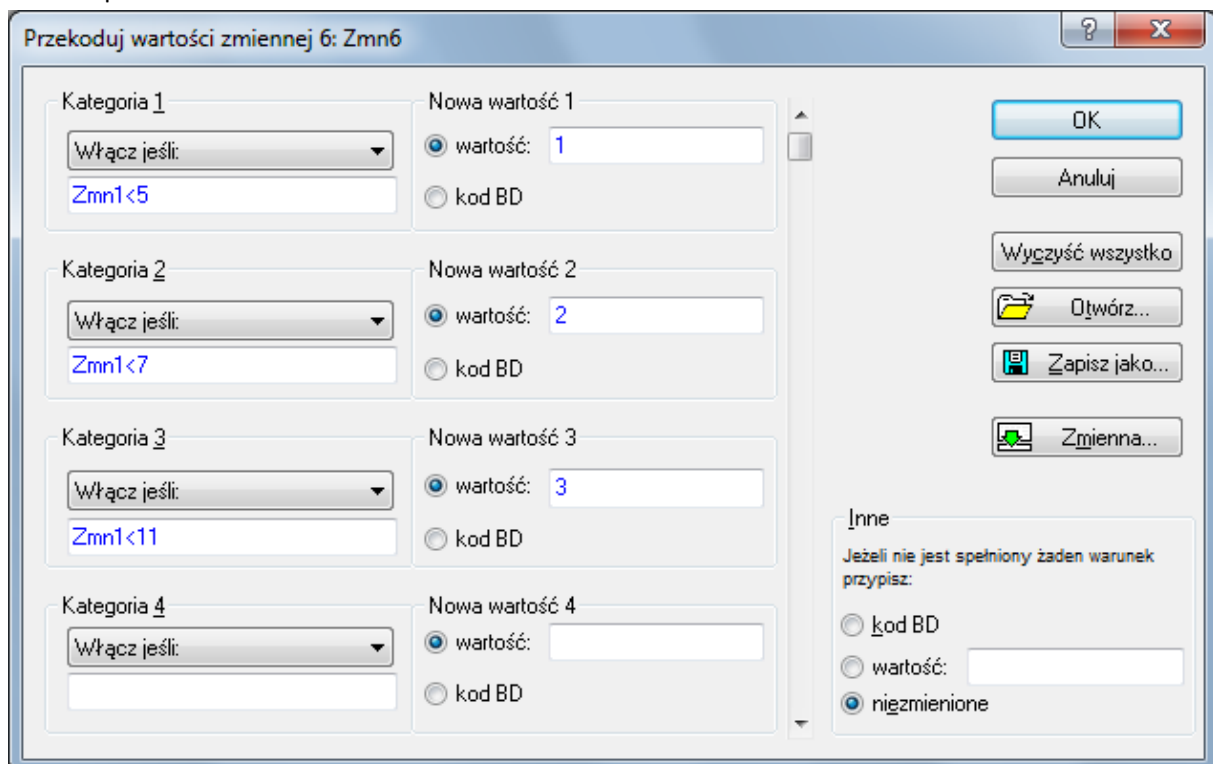
- **dodawanie nowej zmiennej:** wybierz **Zmienne/Dodaj...**Otwiera się okno, w którym można nadać nazwę nowej zmiennej, ustalić jej format wyświetlania, utworzyć ją korzystając z formuły (patrz rozdz. 6);
- **przesuwanie wybranej zmiennej wstecz („w górę”) lub wprzód („w dół”) w stosunku do innych zmiennych o wybrana liczbę przypadków : **Zmienna/Przesuń (opóźnij);****
- **przekodowanie aktualnych wartości zmiennej**(Rys. 9) na pewne nowe wartości: **Zmienna/Przekoduj;**

Poniżej przedstawiono sposób przekodowania w zależności od wartości zmiennej Zmn1, tak, aby nowopowstająca przekodowana zmienna przyjmowała wartości:

- 1 – dla przypadków, dla których wartość Zmn1 jest mniejsza od 5;
- 2 – dla przypadków, dla których wartość Zmn1 jest mniejsza od 7;
- 3 – dla przypadków, dla których wartość Zmn1 jest mniejsza od 11.

W wyniku przekodowania powstaje nowa zmienna zawierająca wartości 1, 2 lub 3.

Przekodowywanie można uzależnić od kilku zmiennych, wykorzystując operatory przedstawione w Tab.1.



Rys.9 Okno do przekodowywania wartości zmiennych. Alternatywnie zamiast nazwy zmiennej Zmn1 można w te same miejsca wpisać v1.

6. Formuły i ich przeliczanie

Formuły pozwalają na szybsze wykonanie niektórych typów obliczeń. Wprowadza się je w oknie specyfikacji zmiennej (Rys.6, pole 12)). Formuły mogą zawierać stałe, nazwy zmiennych, operatory i funkcje. Formułę zaczyna się od znaku „=”.

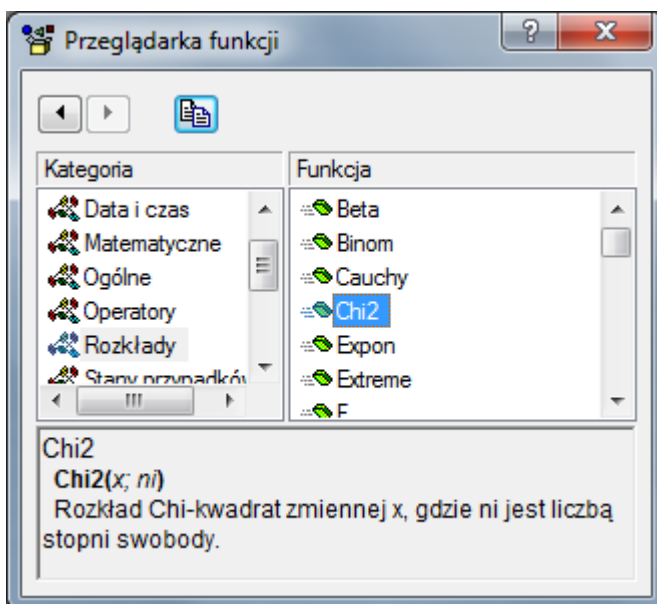
Tab. 1 Dostępne operatory.

Arytmetyczne	+	Dodawanie
	-	Odejmowanie
	*	Mnożenie
	/	Dzielenie
	^ (lub **)	Potęgowanie
Logiczne	()	Nawiasy
	AND (lub &)	Koniunkcja
	OR (lub)	Alternatywa
	NOT (lub ~)	Negacja
Relacyjne	=	Równe
	<	Mniejsze
	>	Większe
	<=	Mniejsze lub równe
	>=	Większe lub równe
	<>	Różne

W formułach do zmiennych można się odwoływać na 2 sposoby:

- przez nazwę (np.: = Zmn1+Zmn2; =LEK1 – LEK2),
- przez ich numery (od ang. *variable* – zmienna zostaje literka „v” np.: =v1 – v2).

Do formuły można dodać komentarz, dodaje się go po średniku (np.: = LEK1-LEK2; *różnica leku pierwszego i drugiego*). Poprzez v0 oznaczamy numer przypadku. Oprócz operatorów w formułach można wykorzystywać funkcje, dostępne w oknie specyfikacji zmiennej (Rys.6) pod przyciskiem **Funkcja** (ozn. na Rys.6 jako 11). W wyniku wciśnięcia przycisku pojawia się okno:



Rys.10 Przeglądarka funkcji.

Zaznaczenie danej kategorii i wybranie z niej konkretnej funkcji powoduje wyświetlenie się krótkiego opisu działania. Dwukrotne kliknięcie na funkcję skutkuje umieszczeniem jej w polu formuły w oknie specyfikacji zmiennej. Przykładowe funkcje przedstawiono w Tab.2 i Tab3, a przykłady formuł wraz z opisem działania w Tab.4.

Uwaga!

Przy tworzeniu formuł należy pamiętać o kolejności wykonywania działań.
 Gdy wykorzystana zmienna dla pewnego przypadku ma brakujące wartości, to w wyniku przeliczenia formuły, dla tego przypadku również wyświetlą się brakujące wartości.
 Formuły można wykorzystywać do przekodowywania zmiennych.

Tab.2 Zestawienie przykładowych funkcji matematycznych.

abs(x)	Wartość bezwzględna x
arcsin(x)	Arcus sinus x
arccos(x)	Arcus cosinus x
arctan(x)	Arcus tangens x
cos(x)	Cosinus x
exp(x)	e do potęgi x
log(x)	Logarytm naturalny z x
log2(x)	Logarytm o podstawie 2 z x
log10(x)	Logarytm o podstawie 10 z x
max(x,y)	Większa wartość z x i y
min(x,y)	Mniejsza wartość z x i y
rnd(x)	Liczba losowa z zakresu 0 do x
sign(x)	Signum x
sin(x)	Sinus x
sqrt(x)	Pierwiastek kwadratowy z x
tan(x)	Tangens x

Tab.3 Zestawienie przykładowych funkcji gęstości prawdopodobieństwa i dystrybuant.

Rozkład	Funkcja gęstości	Dystrybuanta
dwumianowy	binom(x,p,n)	ibinom(x,p,n)
Cauchy'ego	cauchy(x,h,q)	icauchy(x,h,q)
chi-kwadrat <i>i</i> F(x,n,w)	chi2(x,n)	ichi2(x,n)
F	F(x,n,w)	iF(x,n,w)
normalny	normal(x,m,s)	inormal(x,m,s)
Poissona	poisson(x,l)	ipoisson(x,l)
t Studenta	student(x,df)	istudent(x,df)

Tab.4 Przykładowe formuły wraz z opisem działania.

Formuła	Opis działania
=(miara1+miara2+miara3)/3	Średnia arytmetyczna 3 zmiennych
=(v4<=0)*1+(v4>0)*2	Przekodowanie czwartej zmiennej: wartości mniejsze lub równe 0 przekodowuje na 1, a pozostałe na 2;

$=\text{normal}(\text{rnd}(1),0,1)$	Generuje liczby losowe o standaryzowanym rozkładzie normalnym;
$= (((v1=K) \text{ OR } (v1=M))<>0)*1$	Zwraca 1, gdy zmienna losowa v1 przyjmuje wartość K lub M. Dla pozostałych 0.
$= \text{sign}(\text{student}(v3,9) - 0,5)$	Zwraca +1 jeśli wartości gęstości prawdopodobieństwa rozkładu t Studenta są większe od 0,5, -1 gdy są mniejsze od 0,5, a 0 gdy są równe 0,5. Rozkład t Studenta ma 9 stopni swobody.
$= ((v5 \leq 10) \text{ AND } (v1 > \text{Nowak}))*1$	Zwraca wartość 1 dla osób o nazwisku Nowak, dla których wartość zmiennej v5 jest mniejsza lub równa 10, a dla pozostałych przypadków 0.

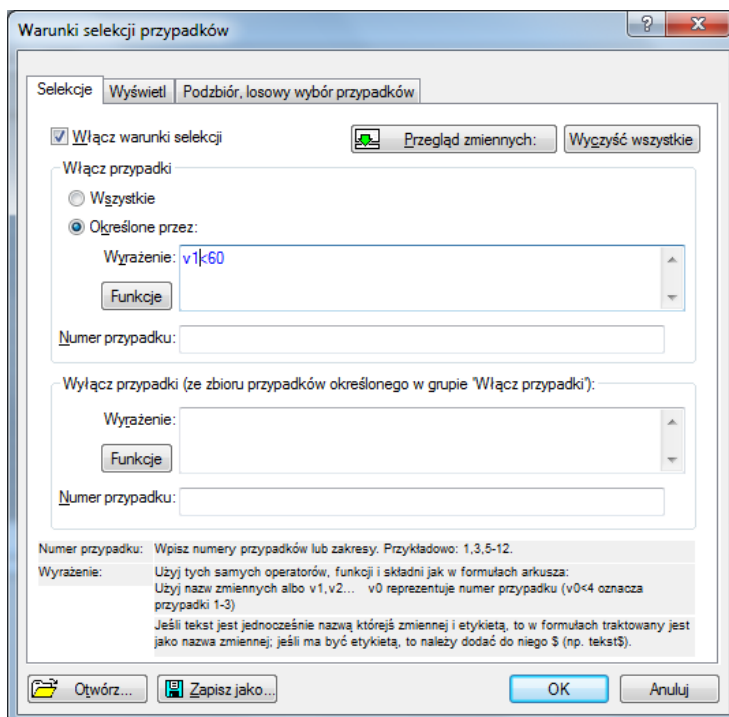
7. Selekcja przypadków

Na dole okna aplikacji STATISTICA znajduje się pasek stanu, na którym wyświetlane są różne komunikaty:



Rys.11 Pasek stanu.

Na Rys.11 widać informację **Selekcja: Nie**, co oznacza, że wszystkie dane są brane pod uwagę w trakcie analizy statystycznej. Dwukrotne kliknięcie pola **Selekcja** wywołuje okno z informacjami na temat obowiązującej selekcji danych do analizy, umożliwi użytkownikowi wybór części danych, które mają zostać poddane analizie. Warunki selekcji mogą zostać pobrane z pliku, lub zapisane w postaci formuły w polu **Wyrażenie**:

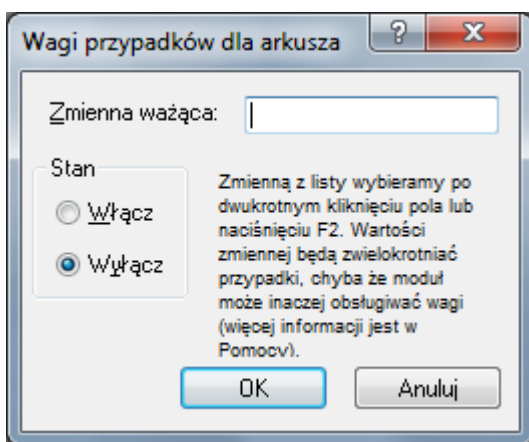


Rys.12 Warunki selekcji przypadków.


W **Wyrażeniu** można wykorzystywać takie same operatory, jak przy formułach. Okno **Warunki selekcji przypadków** zawiera trzy zakładki. Po kliknięciu OK w zakładce **Selekcje** do analiz statystycznych brane są tylko przypadki, które spełniają opisane tu warunki, równocześnie zostają podświetlone w arkuszu danych. W zakładce **Podzbiór, losowy wybór przypadków** można utworzyć osobny arkusz zawierający jedynie wyselekcjonowane dane (klikając przycisk **Podzbiór** i zatwierdzając wybrane przypadki). Utworzone warunki można zapisać i wykorzystywać w innych zbiorach danych (pliki *.sel). Przykład warunku selekcji (v0 – numer przypadku):

v0<100 AND PŁEĆ = 'K' AND WIEK>35 – oznacza, że analiza obejmie tylko te z pierwszych 100 kobiet, które są w wieku poniżej 35 lat.

Na pasku stanu z Rys.11 warte uwagi jest jeszcze jedno pole: **Waga**. Dwukrotne kliknięcie tego pola powoduje wywołanie okna **Definiowania wagi**:



Rys.13 Definiowanie zmiennej ważącej.

Dwukrotne kliknięcie w puste pole **Zmienna ważąca:** powoduje wyświetlenie listy zmiennych, z których należy wybrać zmienną ważącą, czyli taką, której wartości mają się stać wagami dla poszczególnych przypadków. Okno definiowania zmiennej ważącej może zostać wywołane również poprzez naciśnięcie przycisku  z paska zadań, albo z klawiatury klawisza F7.

8. Zapisywanie danych

Po zakończeniu pracy z danymi, plik można zapisać również w formacie m.in. programu Excel, a także jako *.pdf, *.txt. Rozszerzenie arkusza danych zapisanego w formacie STATISTIKI to *.sta. Gdy oprócz arkusza danych zostały wykonane wykresy, całość projektu zostaje zapisywana jako Skoroszyt z rozszerzeniem *.stw.

Ćwiczenie

- Otwórz w STATISTICE plik z danymi zapisany w formacie programu Excel: dane1.xls. Plik zawiera dane dotyczące stężenia w krakowskim powietrzu dwutlenku azotu NO₂ oraz pyłu zawieszonego PM₁₀ w µg/m³ od początku roku 2011 do lutego 2013. Dane znajdują się w Arkuszu1. Nazwy zmiennych zawiera pierwszy wiersz. Stężenia są przedstawione jako średnie dla kolejnych miesięcy. Pomiary stężeń zostały przeprowadzone w stacji badawczej Kraków Kurdwanów, dane pochodzą ze strony: <http://213.17.128.227/iseo/>
- Połącz dane na temat daty znajdujące się w zmiennych Miesiąc i Rok. Połączoną informację przedstaw w zmiennej Data w formacie Mar-2013.

Wskazówki:

- Zmienna **Miesiąc**: w edytorze **Etykiety tekstowych** (patrz Rys.6) zmień etykiety tekstowe wpisując odpowiednie nazwy miesięcy.
- Wykorzystaj **Operacje na danych** dostępne z paska narzędzi **Zmienne/Operacje na danych**. Zwróć uwagę na format wyświetlania daty.

- Uzupełnij zmienne NO₂_proc i PM₁₀_proc. Zmienne te powinny zawierać informacje w % na temat tego, jaki procent względem normy stanowi stężenie poszczególnych zanieczyszczeń dla każdego miesiąca. Normy dla stężeń tych zanieczyszczeń wynoszą:

NO₂ 40 µg/m³
 PM₁₀ 40 µg/m³

Wskazówka: wykorzystaj formuły.

- Wyznacz podstawowe statystyki dla zmiennych NO₂_proc i PM₁₀_proc. Wklej je do notatnika.

Wskazówka: przeczytaj dokładnie opis do Rys.6.

- Utwórz nowe zmienne określające stan powietrza ze względu na stężenia poszczególnych zanieczyszczeń: NO₂_STAN oraz PM₁₀_STAN. W zależności od wartości zmiennych NO₂_proc i PM₁₀_proc zmienne stanu powietrza będą przyjmowały następujące wartości:

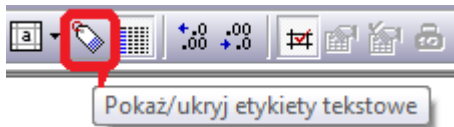
< 50% normy	50
50% - 75% normy	75
75-100% normy	100
>100% normy	200

Wskazówki: wykorzystaj przekodowanie.

- W zależności od wartości liczbowej, dodaj do zmiennych STAN odpowiednie etykiety tekstowe, wyświetl je:

50	Dobry
75	Dostateczny
100	Zły
200	BardzoZły

Wskazówka: do wyświetlenia etykiet wykorzystaj odpowiedni przycisk z paska narzędzi:



7. Wykorzystując pole Selekcja z paska stanu, podświetl te przypadki, dla których stan powietrza ze względu na obydwa zanieczyszczenia jest zły lub bardzo zły. Ile miesięcy od początku 2011 roku obejmuje taki stan?

Do zaliczenia arkusz danych z:

- odpowiednio sformatowaną datą,
- uzupełnionymi zmiennymi NO2_proc i PM10_proc,
- dodanymi i uzupełnionymi zmiennymi NO2_STAN oraz PM10_STAN,
- dodaną i uzupełnioną zmienną STAN z wyświetlonymi etykietami tekstowymi,
- podświetlonymi przypadkami określonymi w punkcie 7.