

Tworzenie podzbioru, analiza na podzbiorach. Grupowanie danych.

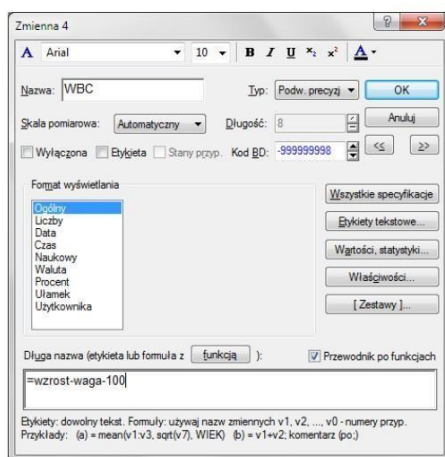
Zadanie 1.

1.1. Należy utworzyć nowy plik w programie Statistica oraz wprowadzić poniższe informacje.

Nazwisko i imię	Płeć	Wzrost	Waga
Góra Ewa	1	168	79
Nowak Jan	2	175	77
Rola Jerzy	2	182	69
Sala Anna	1	170	70
Róg Zofia	1	158	55
Bodo Emil	2	185	74
Dera Adam	2	178	96
Malak Ewa	1	171	57
Car Maria	1	182	94
Wilk Jan	2	178	72
Bendo Ola	1	162	63
Bik Kasia	1	171	82
Bora Piotr	2	180	81
Lis Monika	1	175	61
Mazur Anna	1	167	54

1.2. W kolejnym kroku należy wprowadzić kody dla zmiennej *Płeć*: 1 dla kobiet, 2 dla mężczyzn (należy dodać *etykiety tekstowe*).

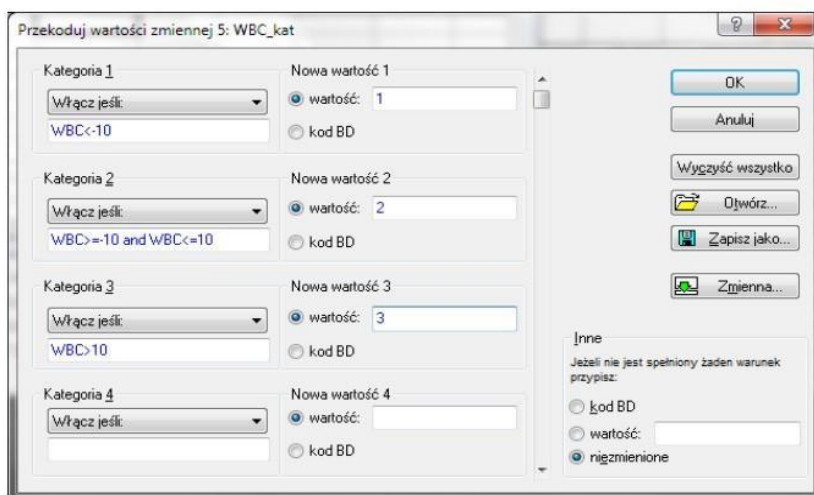
1.3. Należy dodać nową zmienną *WBC* (wskaźnik budowy ciała), którą obliczamy wg. formuły: **Wzrost-Waga-100** (formułę wpisujemy w polu *Długa nazwa*)



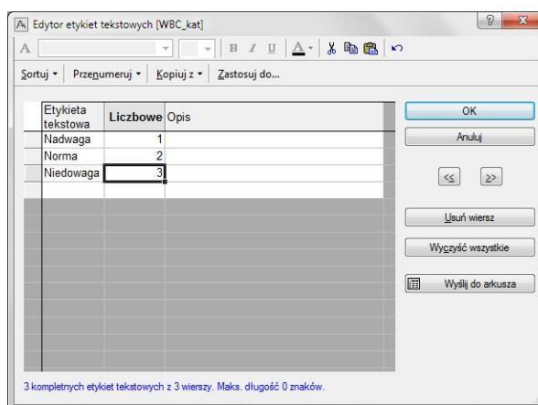
1.4. Należy dodać nową zmienną *WBC_kat*, w której umieszczone zostaną, określone w sposób umowny, kategorie proporcji wagowo-wzrostowych.

WBC	WBC_kat
poniżej -10	nadwaga
<-10;10>	norma
powyżej 10	niedowaga

W tym celu należy dla zmiennej *WBC_kat* wybrać *Dane -> Przekoduj*



Następnie należy nadać etykiety tekstowe jak poniżej



1.5. W tabelach liczości wykonać histogram dla zobrazowania udziału procentowego poszczególnych kategorii (wyciągnąć wnioski biorąc pod uwagę ogół badanych oraz rozróżniając płeć badanych).

Zadanie 2.

Należy otworzyć w programie Statistica plik **Zestawienie_transport.sta**.

Korzystając z pliku wyselekcjonować do badań tylko te osoby, **które mieszkają na stacji**.

Dane > Podzbiór > Zmienne (wszystkie), Przypadki (Włącz warunki selekcji – Określone przez: V9=2).

Biorąc pod uwagę cechę *Czas dojazdu na uczelnię* (V1) ze zmiennej ilościowej stworzyć jakościową (krótki czas dojazdu, średni czas dojazdu, długi czas dojazdu). W tym celu należy wstawić nową zmienną po zmiennej V1 i nazwać ją **Czas dojazdu II** (*Zmienne -> Dodaj ->*

Nadać nazwę. Następnie wejść w Zmienne -> Przekoduj -> wstawić odpowiednie warunki)

Czas dojazdu na uczelnię	Czas dojazdu II	Warunki
poniżej 20 minut	krótki czas dojazdu	V1<20
od 20 do 40 minut	średni czas dojazdu	V1>=20 and V1<=40
powyżej 40 minut	długi czas dojazdu	V1>40

Dla nowo przekodowanej zmiennej sprawdzić etykiety tekstowe tak aby były w porządku rosnącym i w *Tabelach liczości* wykonać histogram dla zobrazowania udziału procentowego poszczególnych kategorii (wyciągnąć wnioski).

Zadanie 3.

Przeprowadzić samodzielnie taką samą analizę jak w zadaniu 2 dla **kobiet, które dojeżdżają**. Uwzględnić dwa warunki (płeć i miejsce zamieszkania podczas studiów) (*and, or*). Za pomocą statystyk opisowych scharakteryzować badaną zmienną. Tym razem czas dojazdu podzielić na dwie grupy (**krótki czas dojazdu** do 35 minut i **długi czas dojazdu** 35 minut i powyżej). Wyniki podziału przedstawić na wykresie kołowym.

Zadanie 4.

Na podstawie danych z pliku **Opinie o integracji z UE 2004** należy przedstawić rozkład wieku ankietowanych 300 osób w badaniu przeprowadzonym w roku 2004. Proszę wykorzystać *Tabele liczości* i przedstawić wynik w postaci tabeli a następnie histogramu. Czy przedstawione wyniki są czytelne, czy potrzebny jest taki poziom szczegółowości prezentacji danych?

W przypadku danych mających charakter liczbowy (a zwłaszcza, gdy cecha przyjmuje szeroki zakres takich wartości) sensowne wydaje się stworzenie **szeregu rozdzielczego przedziałowego**, co wymaga określenia początku pierwszego przedziału i ich rozpiętości. Przyjmijmy, iż chcemy pogrupować wiek w następujących przedziałach 15-24, 25-34, ..., itd.. W tym celu wykorzystujemy opcję *Krok* według wskazówek zamieszczonych obok tabeli (zakładka *Więcej* w oknie *Tabele liczości*).

Wiek	<i>n</i>	<i>n_{SK}</i>	%	% _{SK}
[15; 25)				
[25; 35)				
[35; 45)				
[45; 55)				
[55; 65)				
[65; 75)				
[75; 85)				

Krok:
 rozpocznij od: **od minimum**

n – liczba osób, *n_{SK}* – liczebność skumulowana, % – procent, %_{SK} – procent skumulowany

Na podstawie uzyskanych wyników proszę udzielić odpowiedzi na pytanie, ile ankietowanych osób (i jaki procent) miało mniej niż 45 lat

Proszę przedstawić wyniki w formie *Histogramu*, wykorzystując domyślny wykres wbudowany w okno analizy *Tabele liczości*.

Zadanie 4.

Celem analizy danych z pliku **Wynagrodzenia w powiatach 2002-2017** jest dokonanie opisu poziomu wynagrodzeń w Polsce w przekroju powiatów w roku 2015, 2016 i 2017. Oczywiście do analizy można wykorzystać procedury poznane na wcześniejszych zajęciach: sortowania, rangowania, wizualizacji w postaci wykresów słupkowych. Tym razem jednak nie będziemy analizować „imiennie” powiatów, lecz zbadamy ogólny rozkład wynagrodzeń, tworząc szereg przedziałowy o przedziałach różnej rozpiętości – wykorzystamy w tym celu możliwość wizualizacji danych w postaci histogramu. Oto kolejne etapy tworzenia odpowiedniego wykresu:

- Za pomocą poleceń **Wykresy / Wykresy 2W / Histogramy** wywołujemy okno analizy **Histogramy 2W** po czym wybieramy jako zmienne **2015, 2016 i 2017**;
- W zakładce **Więcej** wskazujemy **Rodzaj wykresu** jako **Wielokrotny** oraz wyłączamy opcję **Dopasuj**;
- Aby ustalić sposób grupowania płać, znajdujemy opcję **Granice**, po czym wprowadzamy następujące liczby **3250 3500 3750 4000 4500 5000** – w ten sposób dane zostaną pogrupowane w 7 przedziałów, przy czym skrajne przedziały są otwarte. Jak widać, przedziały nie muszą być równej długości – w przypadku płać, które wykazują się dość dużą asymetrią, wygodnie jest wybrać „krótsze” przedziały dla niższych przeciętnych wynagrodzeń.
- Zaznaczamy jeszcze opcję **Pokaż procenty** i **Odstępy między kolumnami**, po czym sporządzamy wykres.
- Wykres formatujemy, tak by był jak najbardziej zgodny ze wzorcem zamieszczonym poniżej.

