

ZARZĄDZANIE DANYMI W STATISTICA

1) Zastosowanie:

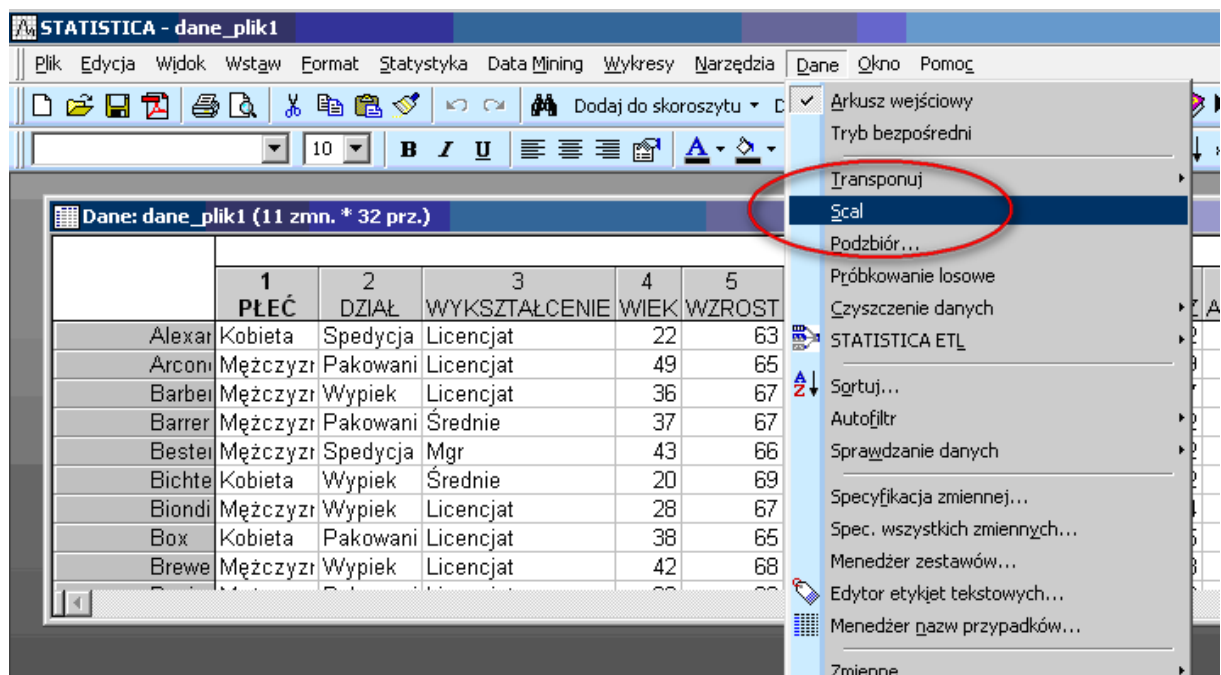
- ▶ STATISTICA umożliwia w zakresie zarządzania danymi m.in.:
 - scalanie plików
 - sprawdzanie danych
 - sortowanie danych
 - tworzenie podzbiorów danych
 - standaryzowanie danych
 - zastępowanie braków danych
 - tworzenie nowych plików danych
 - importowanie danych z innych plików i baz danych

2) Pobierz pliki:

http://www.metal.agh.edu.pl/~regulski/ed/04zaoczne/stat_pliki1/

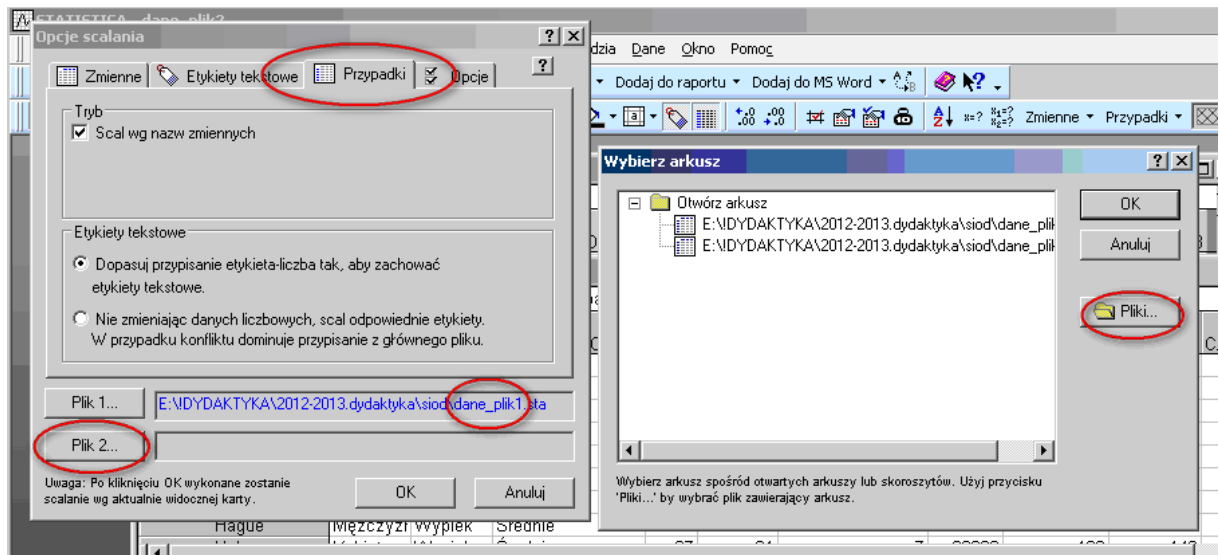
3) Scalanie plików:

- ▶ Pliki `dane_plik1.sta` oraz `dane_plik2.sta` zawierają informacje o różnych pracownikach Sooner Pretzel Company. Pliki te mają tę samą liczbę zmiennych, ale różne przypadki. Scal te pliki z pomocą STATISTICA.

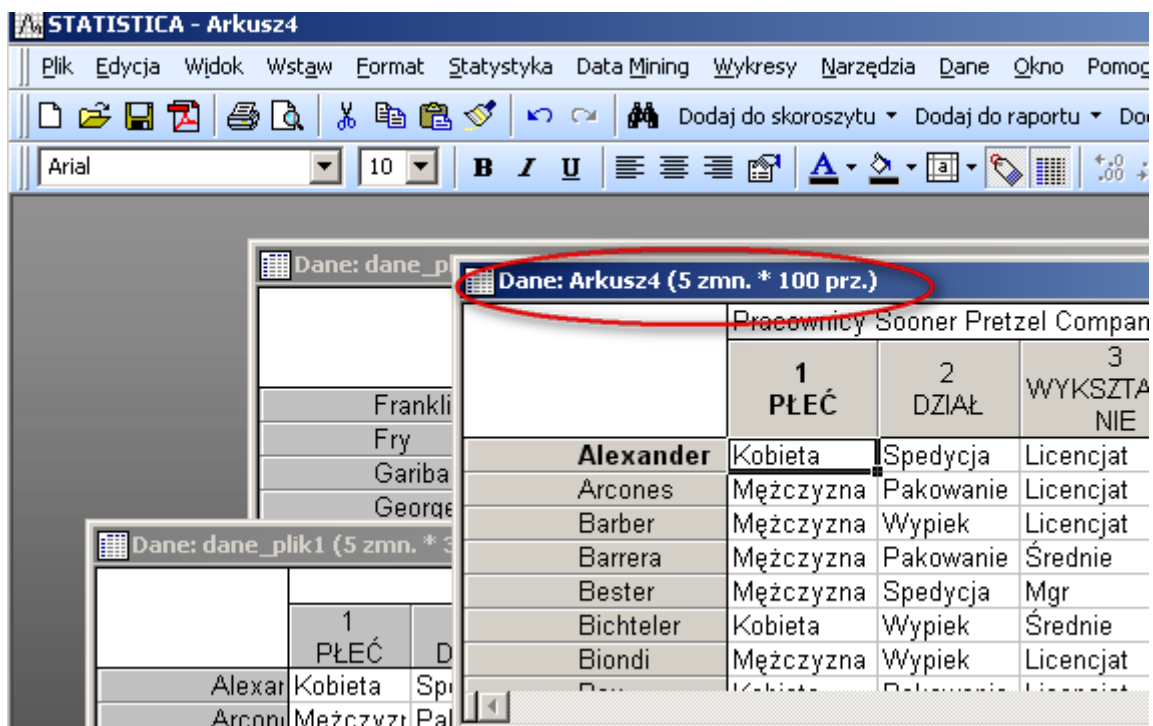


The screenshot shows the STATISTICA software interface. The main window displays a data table with columns labeled 1 through 5, corresponding to variables: PLEĆ, DZIAŁ, WYKSZTAŁCENIE, WIEK, and WZROST. The data rows include names like Alexar, Arconi, Barber, Barrer, Beste, Bichte, Biondi, Box, and Brewe. The 'Dane' menu is open, and the 'Scal' option is highlighted with a red circle. Other options in the menu include 'Arkusz wejściowy', 'Tryb bezpośredni', 'Transponuj', 'Podzbiór...', 'Próbkowanie losowe', 'Czyszczenie danych', 'STATISTICA ETL', 'Sortuj...', 'Autofiltr', 'Sprawdzanie danych', 'Specyfikacja zmiennej...', 'Spec. wszystkich zmiennych...', 'Menedżer zestawów...', 'Edytor etykiet tekstowych...', and 'Menedżer nazw przypadków...'. The status bar at the bottom indicates 'Dane: dane_plik1 (11 zm. * 32 prz.)'.

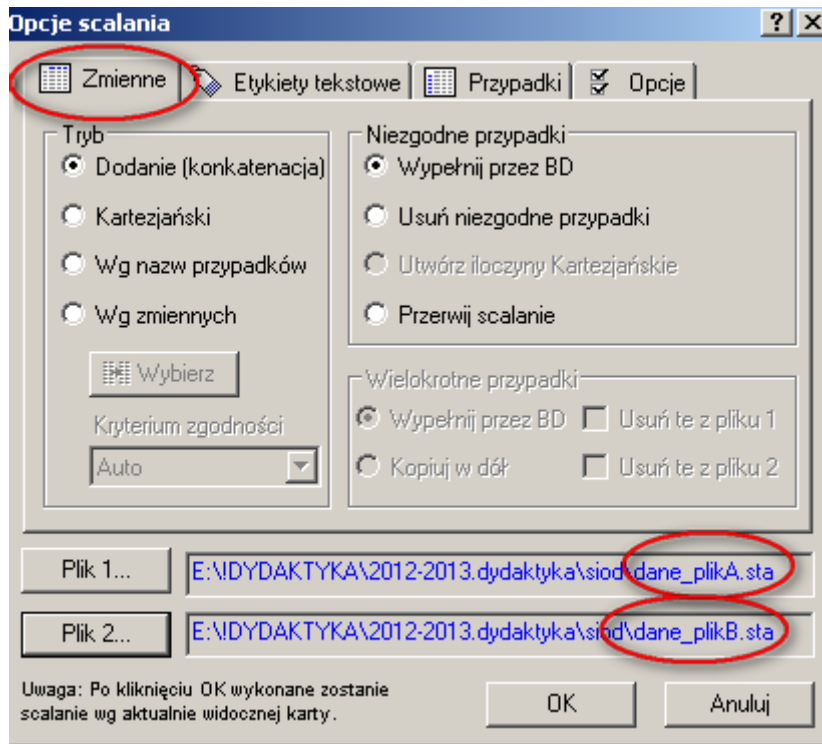
| | 1 | 2 | 3 | 4 | 5 |
|--------|----------|----------|---------------|------|--------|
| | PLEĆ | DZIAŁ | WYKSZTAŁCENIE | WIEK | WZROST |
| Alexar | Kobieta | Spedycja | Licencjat | 22 | 63 |
| Arconi | Mężczyzn | Pakowani | Licencjat | 49 | 65 |
| Barber | Mężczyzn | Wypiek | Licencjat | 36 | 67 |
| Barrer | Mężczyzn | Pakowani | Średnie | 37 | 67 |
| Beste | Mężczyzn | Spedycja | Mgr | 43 | 66 |
| Bichte | Kobieta | Wypiek | Średnie | 20 | 69 |
| Biondi | Mężczyzn | Wypiek | Licencjat | 28 | 67 |
| Box | Kobieta | Pakowani | Licencjat | 38 | 65 |
| Brewe | Mężczyzn | Wypiek | Licencjat | 42 | 68 |



- ▶ Jeśli wszystko pójdzie dobrze otrzymasz:



- ▶ Zapisz wynik w nowym pliku `dane_plikA.sta` (zapisz jako...)
- ▶ Następnie scal ze sobą pliki `dane_plikA.sta` oraz plik `dane_plikB.sta`. Pliki te zawierają różne zmienne, dlatego wybierz zakładkę ZMIENNE...



| Dane: Arkusz5 (11 zm. * 100 prz.) | | | |
|-----------------------------------|---|---|----------------------|
| | | | Pracownicy Sooner Pr |
| | 1 | 2 | |

▶ Otrzymasz:

| | | |
|--|---|---|
| | 1 | 2 |
|--|---|---|

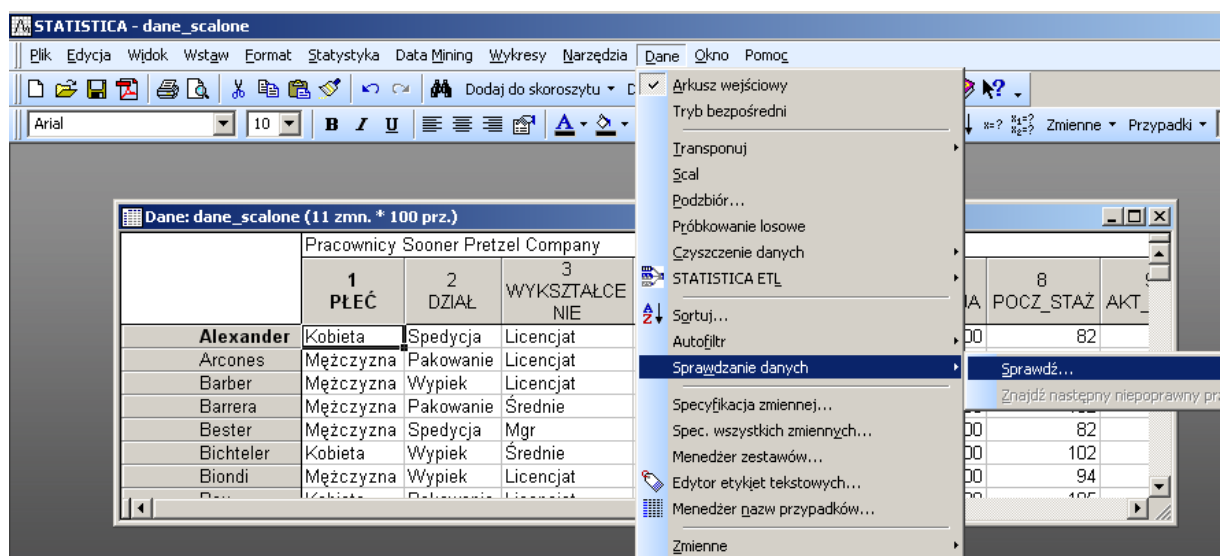
, zapisz plik jako dane_scalone.sta

▶ Pozamykaj wszystkie arkusze danych poza dane_scalone.sta.

// Jeśli nie zostanie wyraźnie wydane takie polecenie, nie zamykaj otwartych analiz ani plików. Jest to również warunkiem otrzymania punktów z aktywności.

4) Sprawdzanie danych:

▶ Sprawdź poprawność danych:



Sprawdzanie danych [?] [X]

Przypadek będzie zaakceptowany jeśli spełnione są:

wszystkie warunki przynajmniej 1 warunek

Warunek 1

Niepoprawne jeśli: [v]
 PŁEĆ<>"Kobieta" AND PŁEĆ<> "

Warunek 2

Poprawne jeśli: [v]
 PENSJA > 0

Warunek 3

Poprawne jeśli: [v]
 []

Warunek 4

Poprawne jeśli: [v]
 []

Znajdź pierwszy

Podświetl niepoprawne

Anuluj

Wyczyść wszystko

Otwórz...

Zapisz jako...

Zakres

Od przypadku: [1] [v]

Do przypadku: [100] [v]

- ▶
- ▶ warunek 1: Niepoprawne jeśli PŁEĆ<>"Kobieta" AND PŁEĆ<> "Mężczyzna";
- ▶ warunek 2: Poprawne jeśli PENSJA>0

| | | | NIE | | | WU | | | | | | |
|----------|-----------|-----------|-----------|----|----|----|-------|-----|-----|----|----|--|
| Dunn | | Pakowanie | Licencjat | 39 | 65 | 12 | 33600 | 93 | 117 | 18 | 18 | |
| Einstein | Kobieta | Srednia | Srednia | 18 | 61 | 0 | 20600 | 95 | 95 | | 11 | |
| Williams | Kobieta | Wypiak | Licencjat | 41 | 66 | 1 | 21000 | 107 | 100 | 3 | 10 | |
| Wintore | Mężczyzna | Wybiak | Srednia | 27 | 67 | 9 | 31600 | 107 | 125 | 13 | 14 | |

- ▶ Warunki można zapisywać i korzystać z wcześniej przygotowanych. Otwórz warunki z pliku Warunki.ini

Sprawdzanie danych [?] [X]

Przypadek będzie zaakceptowany jeśli spełnione są:

wszystkie warunki przynajmniej 1 warunek

Warunek 1

Niepoprawne jeśli: [v]
 PŁEĆ<>"Kobieta" AND PŁEĆ<> "

Warunek 2

Poprawne jeśli: [v]
 PENSJA > 0

Warunek 3

Poprawne jeśli: [v]
 WIEK > 16

Warunek 4

Niepoprawne jeśli: [v]
 POCZ_STAZ < 0

Znajdź pierwszy

Podświetl niepoprawne

Anuluj

Wyczyść wszystko

Otwórz...

Zapisz jako...

Zakres

Od przypadku: [1] [v]

Do przypadku: [100] [v]

- ▶
- ▶ Usuń przypadek: Dunn

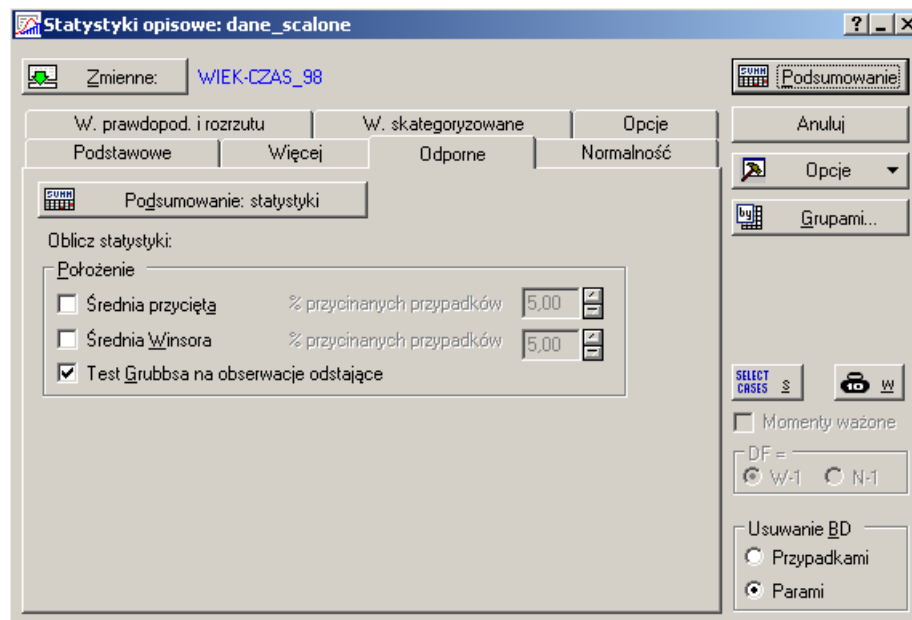
- ▶ W celu sprawdzenia, czy w zbiorze danych jest wartość odstająca ze względu na wartość średnią, w pierwszej kolejności musimy znaleźć kandydata na wartość odstającą, czyli wartość która jest najbardziej oddalona od ogólnej średniej. Test Grubbsa na obserwacje odstające może być użyty w celu wskazania jednej obserwacji odstającej podczas jednego przebiegu.
- ▶ Test ten polega na wyliczeniu jak daleko potencjalna obserwacja odstająca znajduje się od pozostałych wartości w zbiorze danych. Statystyka testu Grubbsa (G) liczona jest jako stosunek największego bezwzględnego odchylenia średniej z próby do odchylenia standardowego próby.
- ▶ Wartość bezwzględną różnicy pomiędzy wartością średnią a wartością kandydata należy podzielić przez odchylenie standardowe. Otrzymujemy w ten sposób wartość statystyki zwanej statystyką

Grubbsa:
$$G = \frac{|\bar{X} - X_{\max}|}{s}$$

- ▶ Pozostaje nam otrzymaną wartość G porównać z wartością krytyczną, którą może my odczytać z odpowiednich tablic. [WERYFIKACJA HIPOTEZ STATYSTYCZNYCH H_0 : dana obserwacja nie jest odstająca] Jeżeli G będzie zbyt wysokie, wtedy należy uznać kandydata za wartość odstającą.
- ▶ poziom istotności *ex post* czyli „poziom prawdopodobieństwa” lub **p-wartość** oznaczone jest literą **p**

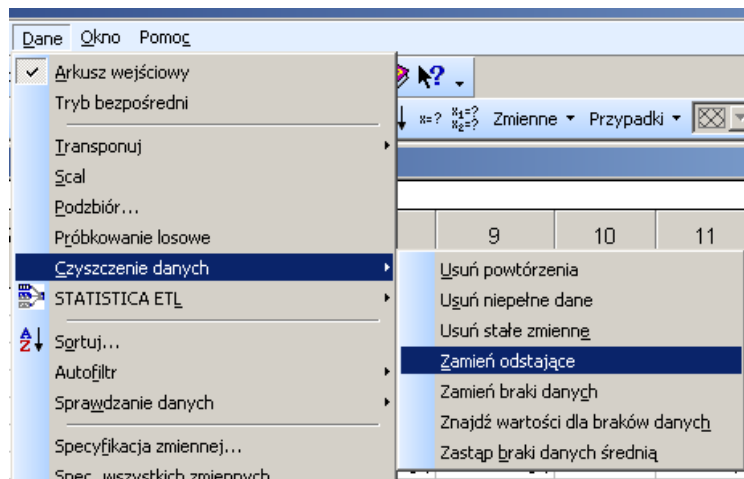
- ▶ WAŻNE: jeśli $p < \alpha$ oznacza to, że są one istotne statystycznie, czyli, że odrzucamy hipotezę zerową na danym poziomie istotności α
- ▶ w przypadku gdy $p > \alpha$, wynika, że nie ma podstaw do odrzucenia H_0 – czyli żadna obserwacja nie jest odstająca.

- ▶ Wykonaj test Grubbsa na obserwacje odstające.



| Zmienna | Statystyki opisowe (dane_scalone_zad) | | | | | | |
|----------------|---------------------------------------|----------|--------------------|-----------------|----------|----------|-----------|
| | N ważnych | Średnia | Grubbsa statystyka | poziom p | Minimum | Maksimum | Odch. std |
| WIEK | 99 | 30,75 | 3,257085 | 0,082017 | 18,000 | 54,00 | 7,139 |
| WZROST | 99 | 66,22 | 2,422542 | 1,000000 | 60,000 | 73,00 | 2,798 |
| STARSZEŃSTWO | 99 | 3,68 | 3,452211 | 0,037183 | 0,000 | 17,00 | 3,859 |
| PENSJA | 99 | 24171,73 | 4,448456 | 0,000267 | -999,000 | 42400,00 | 5658,307 |
| POCZ_STAŻ | 99 | 99,54 | 2,584263 | 0,861328 | 66,000 | 137,00 | 14,497 |
| AKT_STAŻ | 99 | 106,85 | 2,446891 | 1,000000 | 68,000 | 142,00 | 15,877 |
| CZAS_97 | 79 | 14,05 | 2,567859 | 0,701049 | 2,000 | 24,00 | 4,693 |
| CZAS_98 | 99 | 15,01 | 3,451222 | 0,037337 | 5,000 | 32,00 | 4,923 |

- ▶ Po tym jak wskazano obserwacje odstające, do badacza należy określenie czy obserwacje te przedstawiają szczególną własność danej zmiennej czy są wynikiem błędu albo innych anomalii, które nie powinny być modelowane. Przejdź ponownie do arkusza dane_scalone.sta
- ▶ W programie STATISTICA znajdują się opcje czyszczenia danych, które mogą być wykorzystane w celu zamiany obserwacji odstających. Dostęp do tej opcji możliwy jest przez wybór polecenia *Zamień odstające* z menu *Dane - Czyszczenie danych*.
- ▶ Zaznacz obserwacje odstające:



Zamiana wartości odstających i rzadkich

Węjście:
 Zmienne: PENSJA
 Przypadki: WSZYSTKIE
 K powtórzeń: 1 Powtarzaj dopóki wszystkie odstające nie zostaną zamienione Stosuj wagi

Sposób zamiany:

| Zmienna | Pomiar | Test | Parametr | Typ | Wartość | Oznaczenie |
|---------|-----------|---------------------|----------|------------------|---------|------------|
| PENSJA | Ilościowy | Grubbsa, dwustronny | 0,05 | Bez zastępowania | | Bez zmian |

Węjście:
 Zmienne: WSZYSTKIE
 Utwórz nowy arkusz Skopiuj formaty

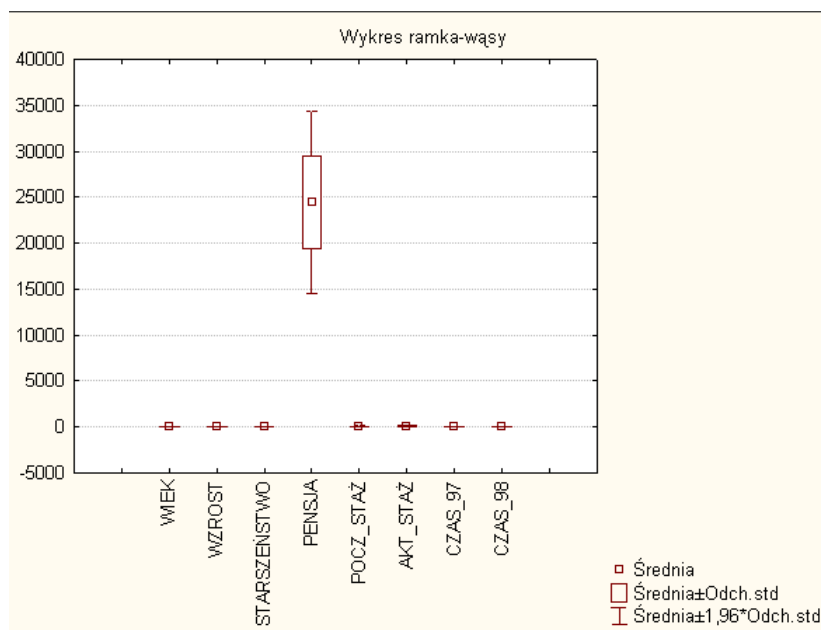
Kliknięcie OK skutkuje zamianą wartości nietypowych na nowe (na sposób określony w polu Węjście)

OK Anuluj

| Imię | Płeć | Specjalizacja | Stano | Wiek | Wzrost | Staż | Wiek | Wzrost | Staż | Wiek | Wzrost |
|----------|-----------|---------------|-----------|------|--------|------|------|--------|------|------|--------|
| Williams | Kobieta | Wypiek | Licencjat | 33 | 65 | 4 | 86 | 94 | 9 | 13 | 13 |
| Witars | Mężczyzna | Wzrost | Średnie | 27 | 67 | 0 | 107 | 125 | 13 | 14 | 14 |

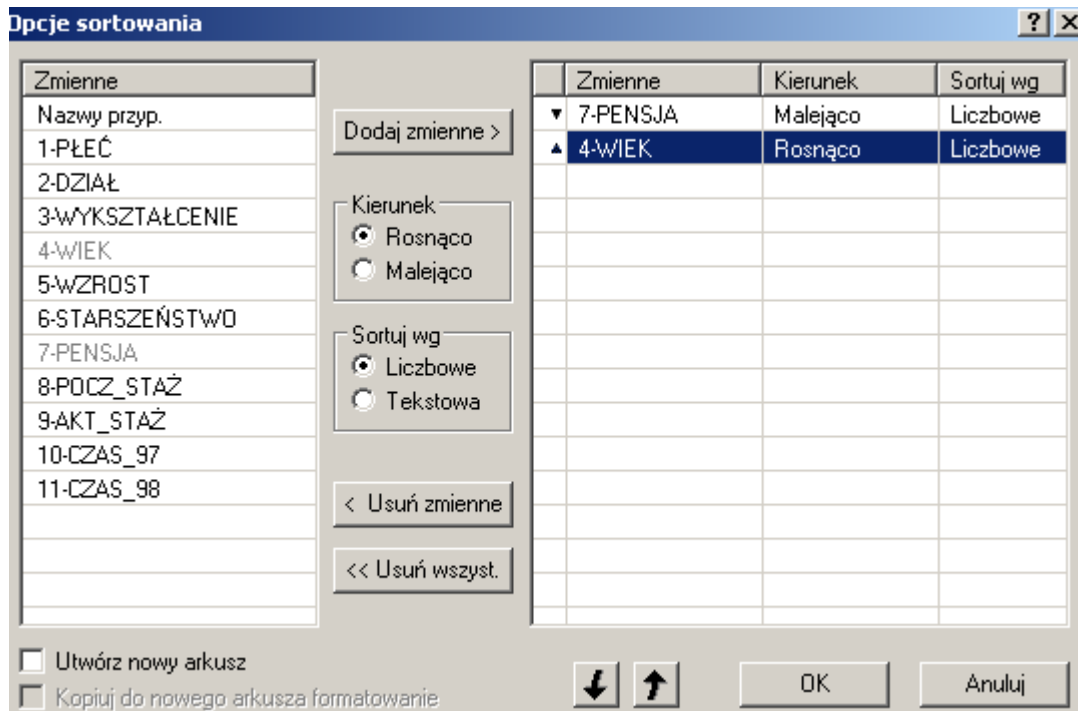
- ▶ Usuń przypadek Williams.
- ▶ Zapisz nowy arkusz danych jako dane_scalone_zad.sta

- ▶ Podejście graficzne. Popularnym sposobem wykrywania obserwacji odstających jest stworzenie wykresu ramkowego.



5) Sortowanie danych:

- Posortuj dane względem zmiennej PENSJA (malejąco) i WIEK (rosnąco)



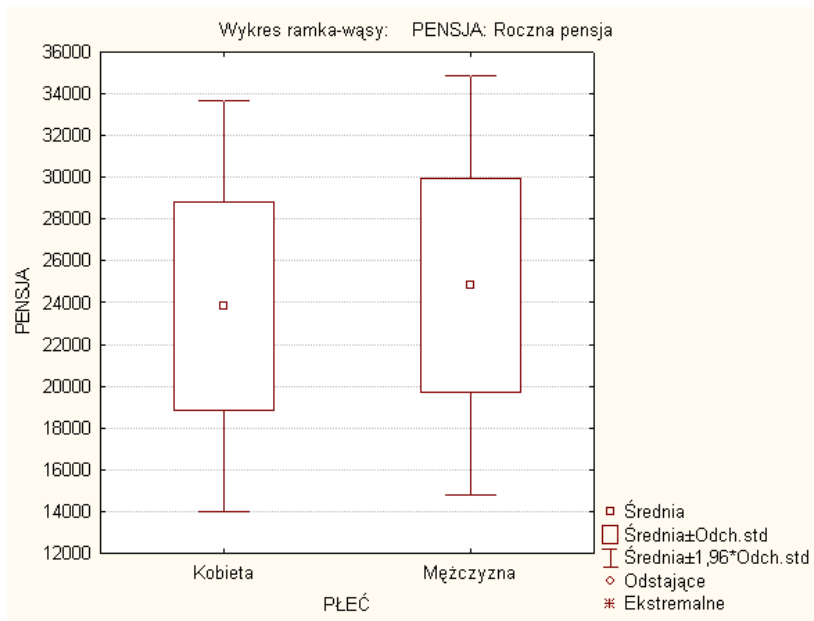
-

6) Tworzenie podzbiorów.

- Zrób zestawienie statystyk opisowych:

| Zmienna | N ważnych | Średnia | Minimum | Maksimum | Odch. std |
|--------------|-----------|----------|----------|----------|-----------|
| WIEK | 98 | 30,72 | 18,00 | 54,00 | 7,172 |
| WZROST | 98 | 66,23 | 60,00 | 73,00 | 2,809 |
| STARSZEŃSTWO | 98 | 3,67 | 0,00 | 17,00 | 3,879 |
| PENSJA | 98 | 24428,57 | 16000,00 | 42400,00 | 5074,273 |
| POCZ_STAŻ | 98 | 99,67 | 66,00 | 137,00 | 14,506 |
| AKT_STAŻ | 98 | 106,98 | 68,00 | 142,00 | 15,904 |
| CZAS_97 | 78 | 14,12 | 2,00 | 24,00 | 4,688 |
| CZAS_98 | 98 | 15,03 | 5,00 | 32,00 | 4,944 |

-



- ▶
- ▶ Utwórz losowy podzbiór. Wykonaj na zbiorze dane_scalone_zad.sta losowanie warstwowe. Weź 100% kobiet i 10% mężczyzn.

Twórz próbę losową [?] [X]

Zmienne: WSZYSTKIE

Przypadki: WSZYSTKIE

Proste losowanie | **Losowanie warstwowe** | Opcje [?]

Zmienna grupująca: PŁEĆ

| Warstwy | Przybliżony % |
|-----------|---------------|
| Kobieta | 100,000000 |
| Mężczyzna | 10,000000 |
| | |
| | |
| | |
| | |

Kody Równe prawdopodob. Przybliżony % = [] []
 Dokładna liczba lub procent przypadków

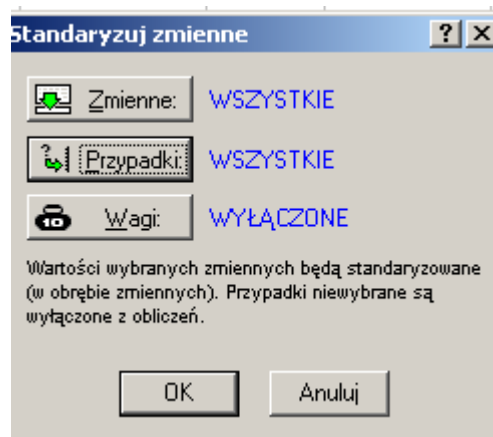
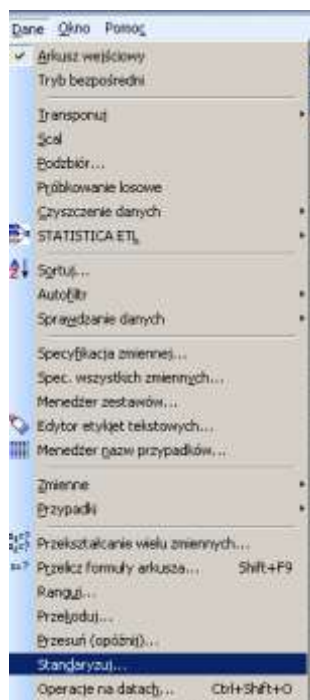
Aktywna karta określa rodzaj losowania wykonywanego po kliknięciu OK. [OK] [Anuluj]

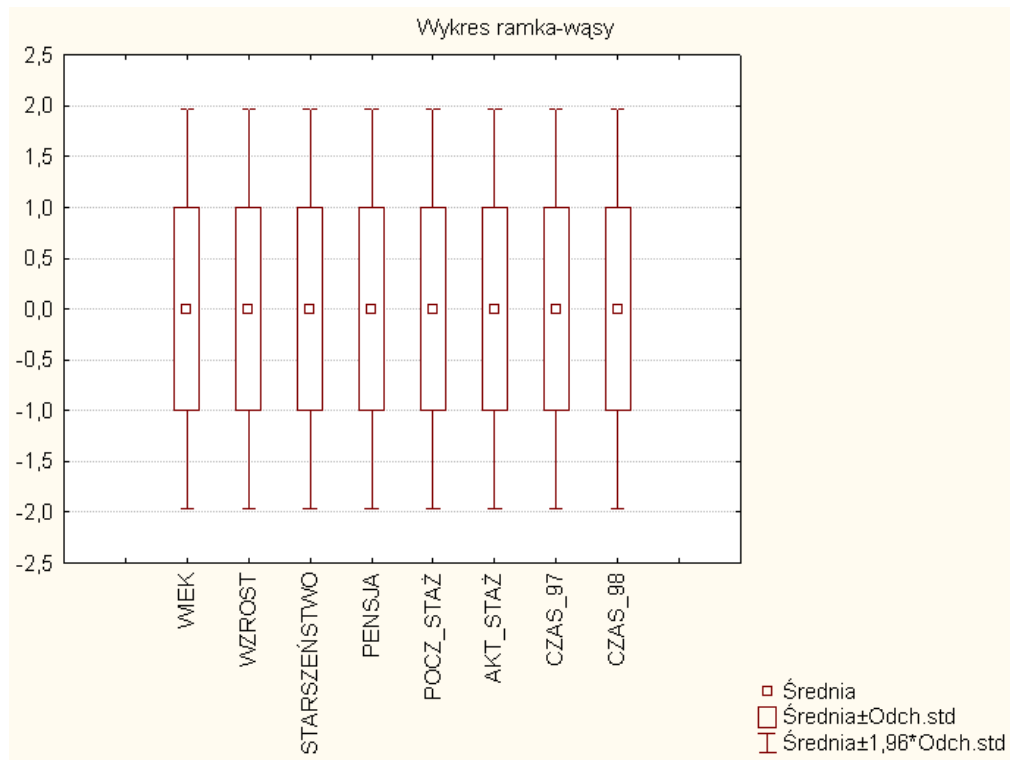
- ▶
- ▶ Ponownie policz statystyki:

| Zmienna | Statystyki opisowe (Arkusz7) | | | | |
|--------------|------------------------------|----------|----------|----------|----------|
| | N ważnych | Średnia | Minimum | Maksimum | Odch.std |
| WIEK | 43 | 29,23 | 18,00 | 54,00 | 7,118 |
| WZROST | 43 | 64,30 | 60,00 | 70,00 | 2,559 |
| STARSZEŃSTWO | 43 | 2,98 | 0,00 | 17,00 | 3,642 |
| PENSJA | 43 | 23579,07 | 18300,00 | 42400,00 | 4932,910 |
| POCZ_STAŻ | 43 | 98,33 | 75,00 | 126,00 | 11,435 |
| AKT_STAŻ | 43 | 104,23 | 78,00 | 141,00 | 14,123 |
| CZAS_97 | 32 | 14,44 | 7,00 | 22,00 | 4,428 |
| CZAS_98 | 43 | 15,84 | 5,00 | 32,00 | 5,481 |

7) Standaryzowanie zmiennych

- ▶ Na wykresie ramka-wąsy widać wyraźnie, że zmienna PENSJA znacznie odbiega od pozostałych skalą. Aby móc badać w przyszłości regresje i korelację, trzeba takie zmienne zestandaryzować, tzn. średnia będzie równa 0, a odchylenie standardowe równe będzie 1.
- ▶ W pliku będącym wylosowanym zbiorem zestandaryzuj wszystkie zmienne ilościowe.

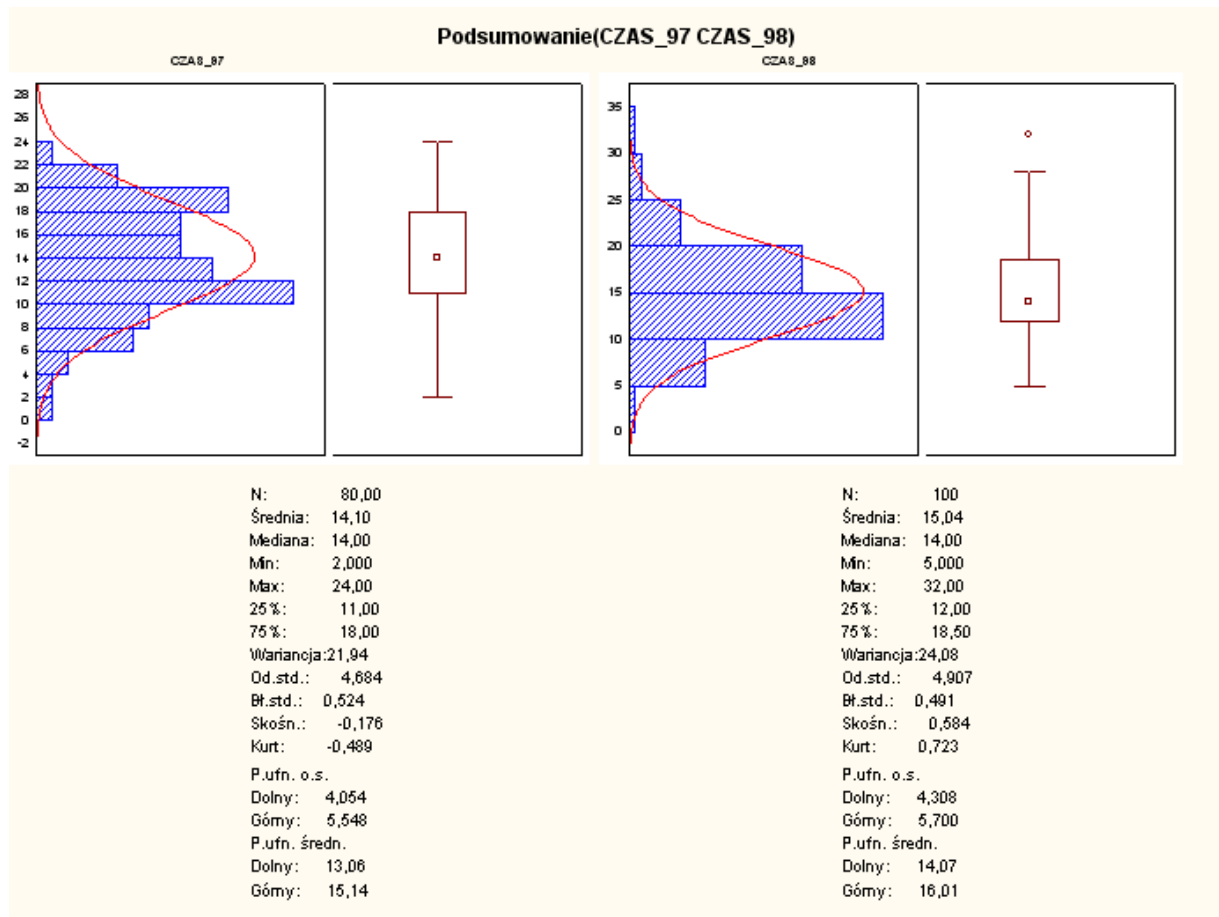




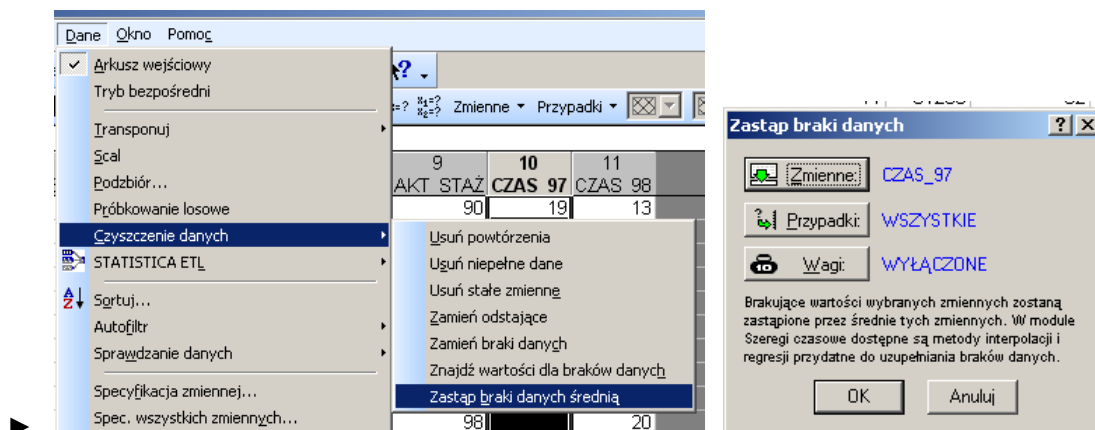
8) Zastępowanie braków danych

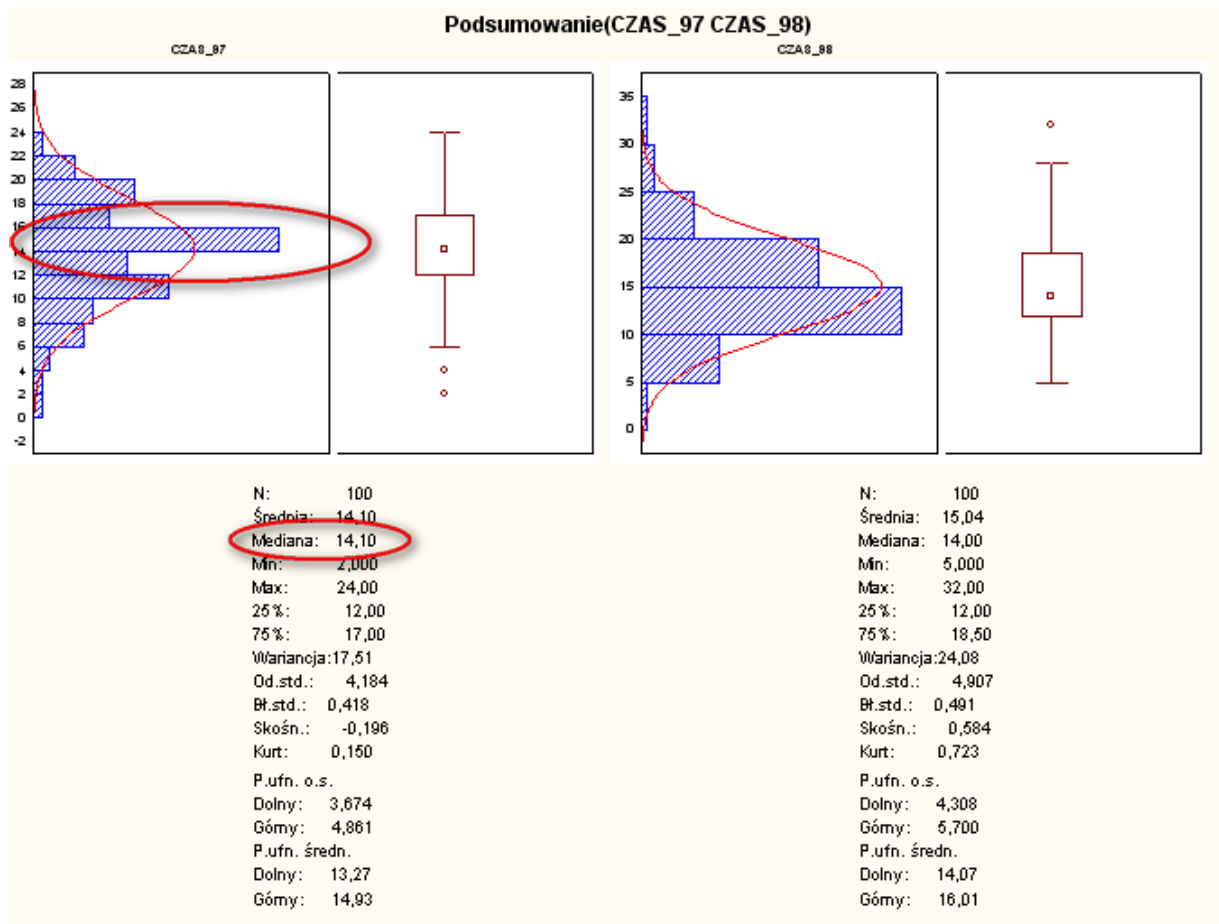
- ▶ Otwórz plik dane_scalone.sta.
- ▶ Zwróć uwagę, że zmienna CZAS_97 obejmuje jedynie 79 przypadków, podczas gdy pozostałe zmienne mają po 99.

| Zmienna | Statystyki opisowe (dane_scalone_zad) | | | | |
|--------------|---------------------------------------|----------|----------|----------|----------|
| | N ważnych | Średnia | Minimum | Maksimum | Odch.std |
| WIEK | 99 | 30,75 | 18,000 | 54,00 | 7,139 |
| WZROST | 99 | 66,22 | 60,000 | 73,00 | 2,798 |
| STARSZEŃSTWO | 99 | 3,68 | 0,000 | 17,00 | 3,859 |
| PENSJA | 99 | 24171,73 | -999,000 | 42400,00 | 5658,307 |
| POCZ_STAŻ | 99 | 99,54 | 66,000 | 137,00 | 14,497 |
| AKT_STAŻ | 99 | 106,85 | 68,000 | 142,00 | 15,877 |
| CZAS_97 | 79 | 14,05 | 2,000 | 24,00 | 4,693 |
| CZAS_98 | 99 | 15,01 | 5,000 | 32,00 | 4,923 |



- ▶
- ▶ STATISTICA umożliwi wiele wygodnych sposobów czyszczenia danych, wybierzmy tym razem zastępowanie braków danych średnią:





9) Import plików

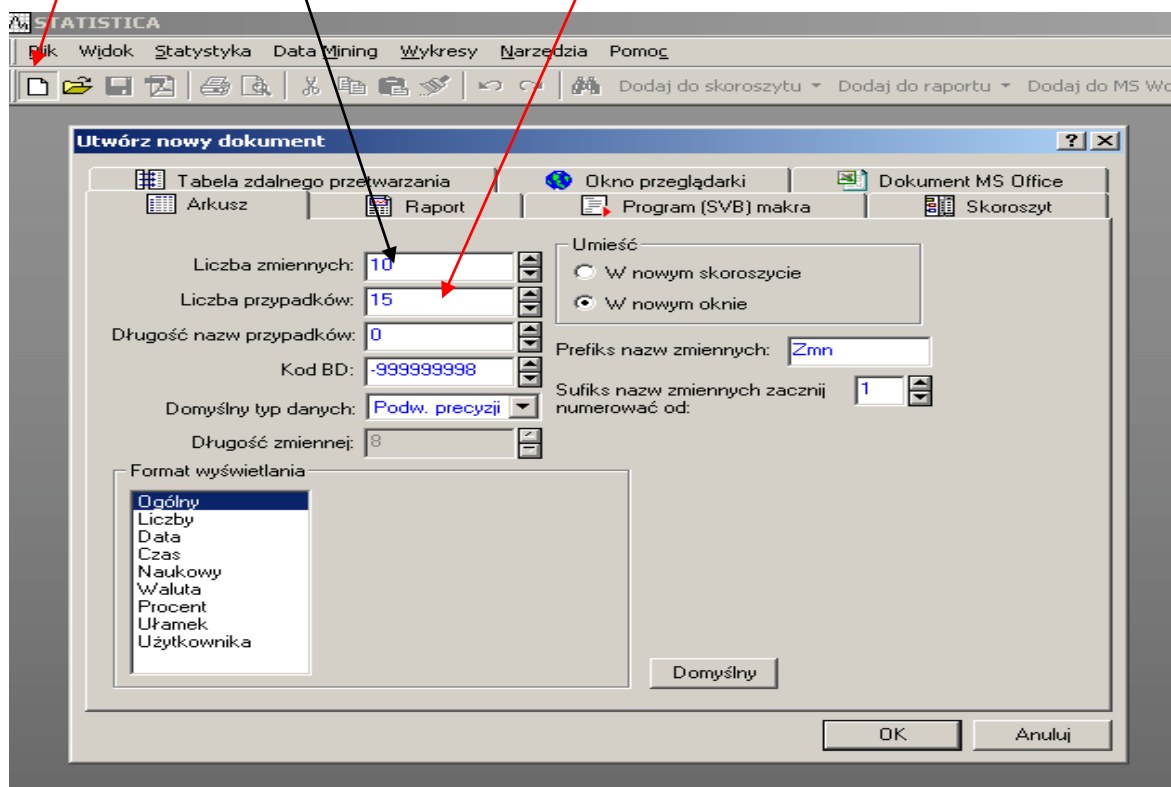
- ▶ STATISTICA umożliwia wygodny import plików.
- ▶ Wypróbujmy to na pliku Dane_siod.xls
- ▶ Zimportuj te dane do STATISTICA zachowując nazwy zmiennych i przypadków.
- ▶ Zadbaj o polskie znaki.
- ▶ Zwołaj mnie jak skończysz, aby otrzymać punkty.

STATYSTYKA OPISOWA

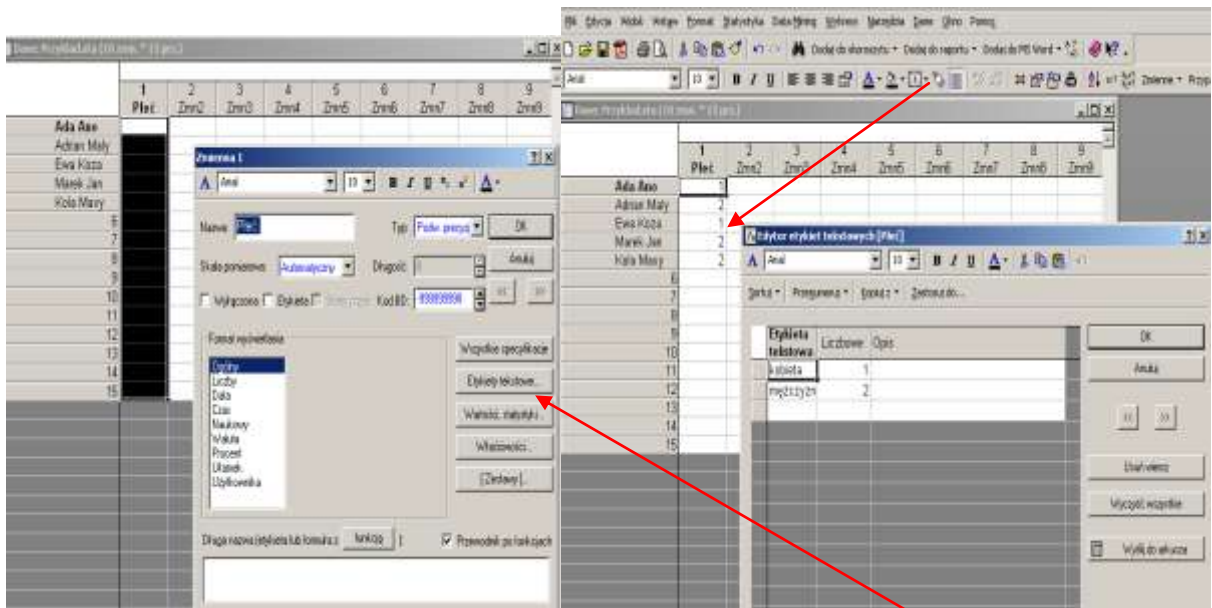
Tworzenie arkusza w pakiecie STATISTICA

Plik/Nowy

Ustawienia-liczba zmiennych, liczba przypadków



Plik/Zapisz: **Przyklad.sta**



Zmienną Płeć możemy oznaczać albo cyfrą albo zapisem tekstowym (etykieta tekstowa)

Jako kolejne wprowadzamy zmienne : zm2 **Wzrost** i zm3 **Waga**, a następnie zmienną zm4 jako **WBC**, jej wartości są obliczane z formuły = **Wzrost-Waga -100**, którą należy wpisać w polu **Długa nazwa**.

STATYSTYKA - Przykład.sta

Plik Edycja Widok Wstaw Format Statystyka Data Mining Wykresy Narzędzia Dane Okno Pomoc

Dane: Przykład.sta (10 zmn. * 15 prz.)

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-------------|----------|------|------|------|------|------|------|------|------|
| | Płeć | Zmn2 | Zmn3 | Zmn4 | Zmn5 | Zmn6 | Zmn7 | Zmn8 | Zmn9 |
| Ada Ano | kobieta | | | | | | | | |
| Adrian Maly | mężczyzn | | | | | | | | |
| Ewa Koza | kobieta | | | | | | | | |
| Marek Jan | mężczyzn | | | | | | | | |
| Kola Maxy | mężczyzn | | | | | | | | |

Zmienna 5

Nazwa: **Kody** Typ: **Podw. precyzj** OK

Skala pomiarowa: **Automatyczny** Długość: 8 Anuluj

Wyłączona Etykieta Stany przyp. Kod BD: **-999999998** << >>

Format wyświetlania

- Ogólny**
- Liczby
- Data
- Czas
- Naukowy
- Waluta
- Procent
- Ułamek
- Użytkownika

Długa nazwa (etykieta lub formuła z funkcją): Przewodnik po funkcjach

=1*(wbc<-10)+2*(wbc>=-10 AND wbc <=10)+3*(wbc>10)

Etykiety: dowolny tekst. Formuły: używaj nazw zmiennych v1, v2, ..., v0 - numery przyp. Przykłady: (a) = mean(v1:v3, sqrt(v7), WIEK) (b) = v1+v2; komentarz (po:)

Zmienna 5, o nazwie Kody jest obliczana wg formuły

$$= 1 * (\text{wbc} < -10) + 2 * (\text{wbc} \geq -10 \text{ AND } \text{wbc} \leq 10) + 3 * (\text{wbc} > 10)$$

Dla tej zmiennej wprowadzamy etykiety tekstowe

Włączyć Wartości statystyki dla sprawdzenia poszczególnych zmiennych.

Przełączyć dla zmiennych Płeć i Kody etykiety tekstowe i oznaczenia cyfrowe.

Statystyki Opisowe w pakiecie STATISTICA.

Empiryczne rozkłady zmiennych.

Realizacja przykładu: **Otwórz /Zarobki.sta**

Tworzymy rozkład empiryczny zmiennej jakościowej **Wykształcenie**

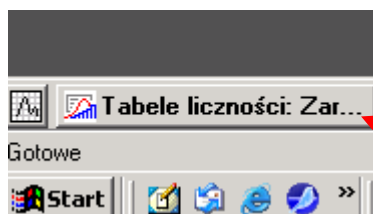
Statystyka /Statystyki podstawowe i tabele/Tabele licznosci

The screenshot shows the STATISTICA software interface. The main window displays a data table with the following columns: 1 Płeć, 2 Wykształcenie, 3 Wiek, 4 Staż, 5 Płaca. The data rows show 20 records, all with 'Kobieta' for gender and 'Średnie zawodowe' for education level. Overlaid on the table are two dialog boxes. The 'Tabele licznosci: Zarobki' dialog box is open, showing the 'Zmienne:' field set to 'Wykształcenie'. The 'Wybierz zmienną' dialog box is also open, showing a list of variables: 1 - Płeć, 2 - Wykształcenie, 3 - Wiek, 4 - Staż pracy, 5 - Praca brutto. Red arrows point from the text in the previous blocks to these dialog boxes.

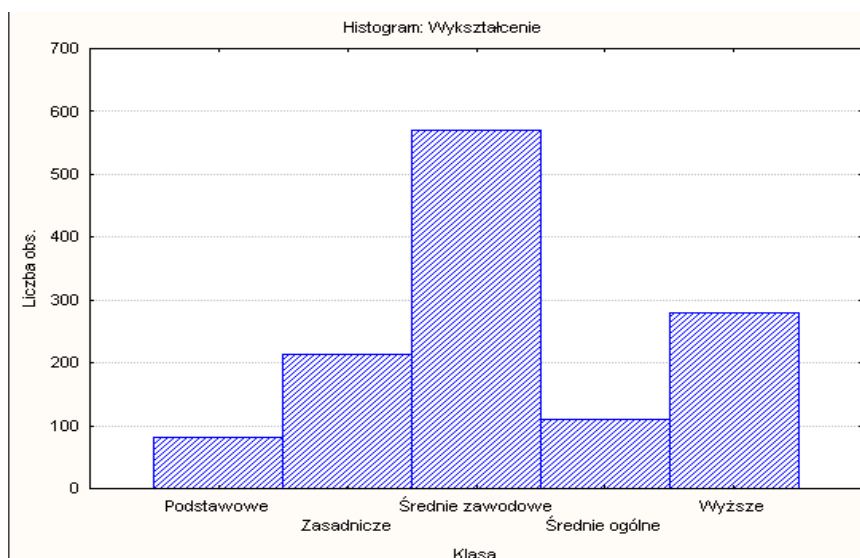
Uzyskamy tabele licznosci dla zmiennej Wykształcenie

Skoroszyt1* - Tabela licznosci: Wykształcenie (Zarobki)

| Klasa | Tabela licznosci: Wykształcenie (Zarobki) | | | |
|-------------------|---|---------------------|----------|----------------------|
| | Liczba | Skumulow. Liczba | Procent | Skumulow. Procent |
| Podstawowe | 81 | 81 | 6,45418 | 6,4542 |
| Zasadnicze | 214 | 295 | 17,05179 | 23,5060 |
| Średnie zawodowe | 571 | 866 | 45,49801 | 69,0040 |
| Średnie ogólne | 110 | 976 | 8,76494 | 77,7689 |
| Wyższe | 279 | 1255 | 22,23108 | 100,0000 |
| Braki | 0 | 1255 | 0,00000 | 100,0000 |

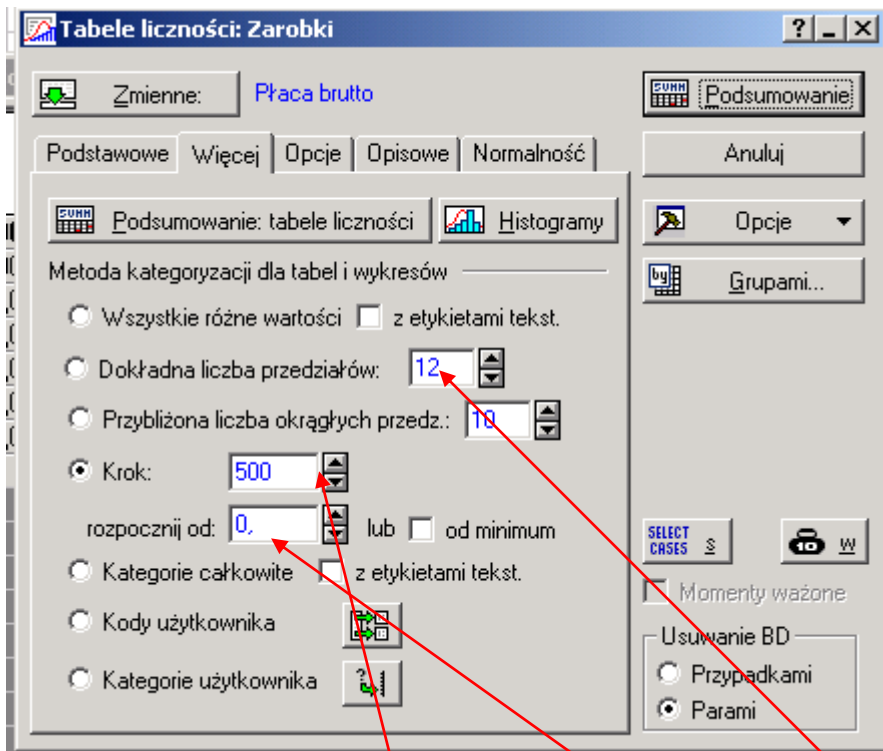


Klikając w ikonę, znajdującą się na pasku Analizy w dolnej części okna aplikacji, przywołujemy **Tabelę licznosci Zarobki** i po naciśnięciu **Histogramy** uzyskamy wykres

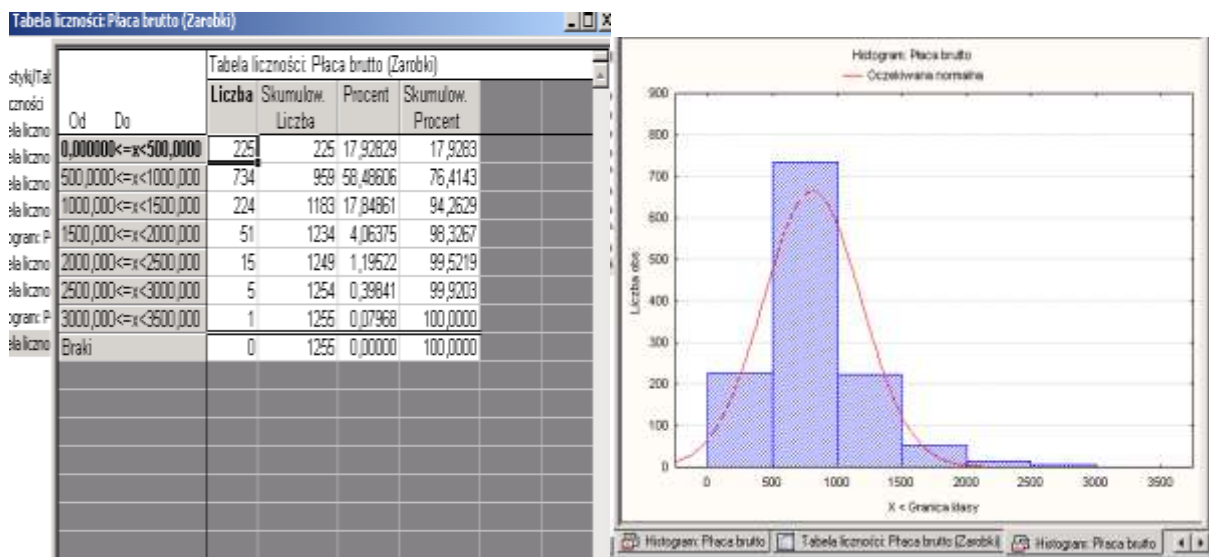


Wykształcenie podzielone jest według naturalnych kategorii. W przypadku gdy zmienna przyjmuje wartości liczbowe należy te kategorie (przedziały klasowe) stworzyć.

Przywołajmy **tabelę licznosci**, wybieramy zmienną **płaca brutto** i zakładkę **więcej**

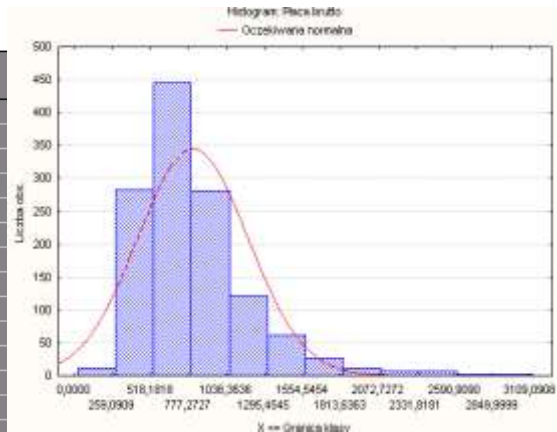


Tu jest możliwość zdecydowania czy wybieramy **liczbę przedziałów**, czy też szerokość przedziału klasowego czyli **krok**, a także od którego punktu chcemy rozpocząć kwalifikację. Przykładowy rozkład, gdzie punktem początkowym jest zero, a szerokość przedziału klasowego 500, pokazano na następnym rysunku.



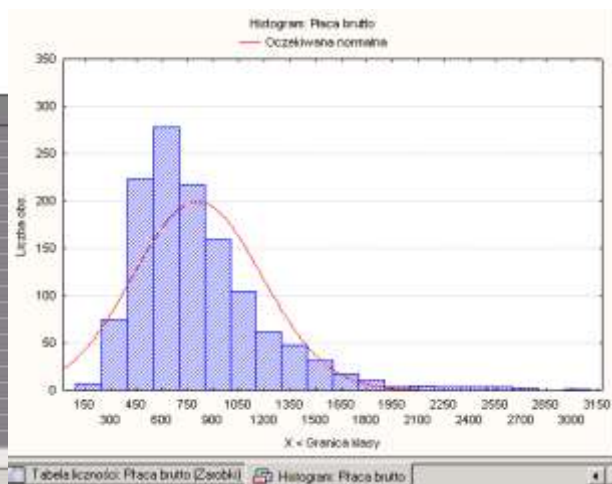
Można też postąpić inaczej, zadać liczbę klas np. 12, program oblicza granice klas

| | | Tabela licznosci: Płaca brutto (Zarobki) | | | |
|----------|--------------|--|------------------|----------|-------------------|
| Od | Do | Liczba | Skumulow. Liczba | Procent | Skumulow. Procent |
| 20,45455 | <x<=279,5455 | 11 | 11 | 0,87649 | 0,8765 |
| 279,5455 | <x<=538,6364 | 262 | 293 | 22,47012 | 23,3466 |
| 538,6364 | <x<=797,7273 | 445 | 738 | 35,45817 | 58,8048 |
| 797,7273 | <x<=1056,818 | 281 | 1019 | 22,39044 | 81,1952 |
| 1056,818 | <x<=1315,909 | 121 | 1140 | 9,64143 | 90,8367 |
| 1315,909 | <x<=1575,000 | 61 | 1201 | 4,86056 | 95,6972 |
| 1575,000 | <x<=1834,091 | 26 | 1227 | 2,07171 | 97,7689 |
| 1834,091 | <x<=2093,182 | 12 | 1239 | 0,96618 | 98,7251 |
| 2093,182 | <x<=2352,273 | 7 | 1246 | 0,56777 | 99,2829 |
| 2352,273 | <x<=2611,364 | 6 | 1252 | 0,47809 | 99,7610 |
| 2611,364 | <x<=2870,455 | 2 | 1254 | 0,15936 | 99,9203 |
| 2870,455 | <x<=3129,545 | 1 | 1255 | 0,07968 | 100,0000 |
| Braki | | 0 | 1255 | 0,00000 | 100,0000 |

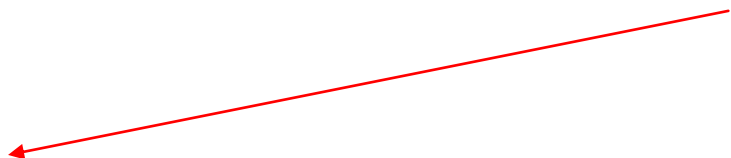


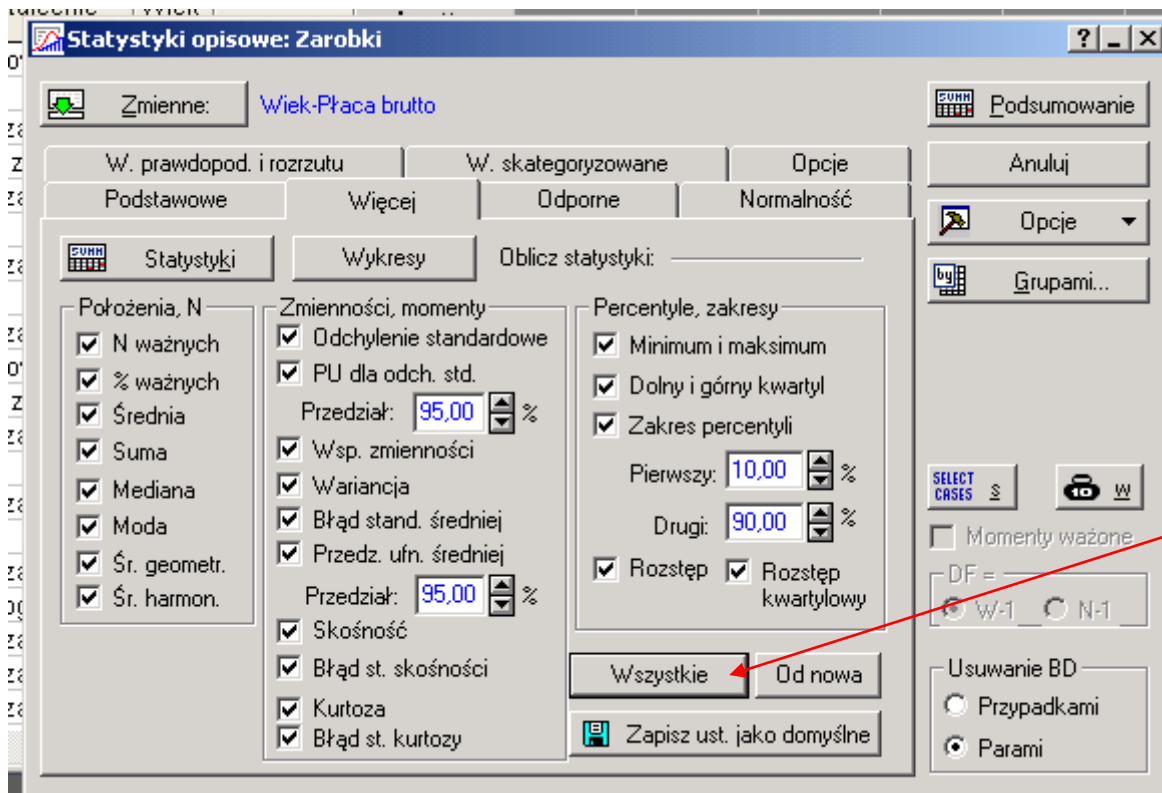
Następne ćwiczenie polega na kolejnej zmianie obrazu rozkładu empirycznego, tak by szerokość przedziału wynosiła 150, i rozpoczynamy od wartości 100

| | | Tabela licznosci: Płaca brutto (Zarobki) | | | |
|-----------|---------------|--|------------------|----------|-------------------|
| Od | Do | Liczba | Skumulow. Liczba | Procent | Skumulow. Procent |
| 100,0000 | <x<=250,0000 | 7 | 7 | 0,55777 | 0,5578 |
| 250,0000 | <x<=400,0000 | 74 | 81 | 5,89641 | 6,4542 |
| 400,0000 | <x<=550,0000 | 223 | 304 | 17,76892 | 24,2231 |
| 550,0000 | <x<=700,0000 | 278 | 582 | 22,15139 | 46,3745 |
| 700,0000 | <x<=850,0000 | 218 | 800 | 17,37052 | 63,7450 |
| 850,0000 | <x<=1000,0000 | 159 | 959 | 12,66932 | 76,4143 |
| 1000,0000 | <x<=1150,0000 | 105 | 1064 | 8,36653 | 84,7809 |
| 1150,0000 | <x<=1300,0000 | 62 | 1126 | 4,94024 | 89,7211 |
| 1300,0000 | <x<=1450,0000 | 48 | 1174 | 3,82470 | 93,5458 |
| 1450,0000 | <x<=1600,0000 | 31 | 1205 | 2,47012 | 96,0159 |
| 1600,0000 | <x<=1750,0000 | 19 | 1223 | 1,43426 | 97,4502 |
| 1750,0000 | <x<=1900,0000 | 10 | 1233 | 0,79681 | 98,2470 |
| 1900,0000 | <x<=2050,0000 | 4 | 1237 | 0,31873 | 98,5657 |
| 2050,0000 | <x<=2200,0000 | 5 | 1242 | 0,39841 | 98,9641 |
| 2200,0000 | <x<=2350,0000 | 4 | 1246 | 0,31873 | 99,2829 |
| 2350,0000 | <x<=2500,0000 | 3 | 1249 | 0,23904 | 99,5219 |
| 2500,0000 | <x<=2650,0000 | 3 | 1252 | 0,23904 | 99,7610 |
| 2650,0000 | <x<=2800,0000 | 2 | 1254 | 0,15936 | 99,9203 |
| 2800,0000 | <x<=2950,0000 | 0 | 1254 | 0,00000 | 99,9203 |
| 2950,0000 | <x<=3100,0000 | 1 | 1255 | 0,07968 | 100,0000 |



Statystyki opisowe: na zakładce **Więcej** wybieramy trzy zmienne **Wiek, Staż pracy i Płaca brutto**

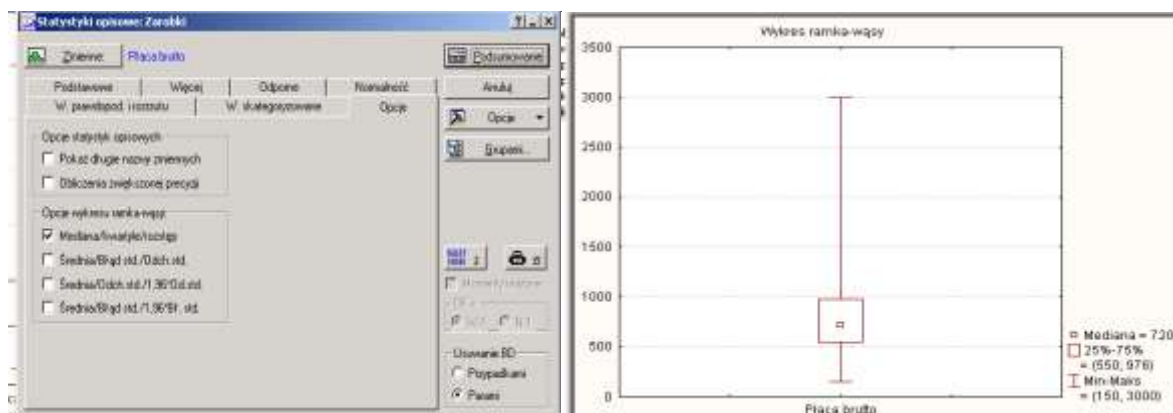




Klikamy przycisk **Wszystkie** a następnie **Podsumowanie** i otrzymamy

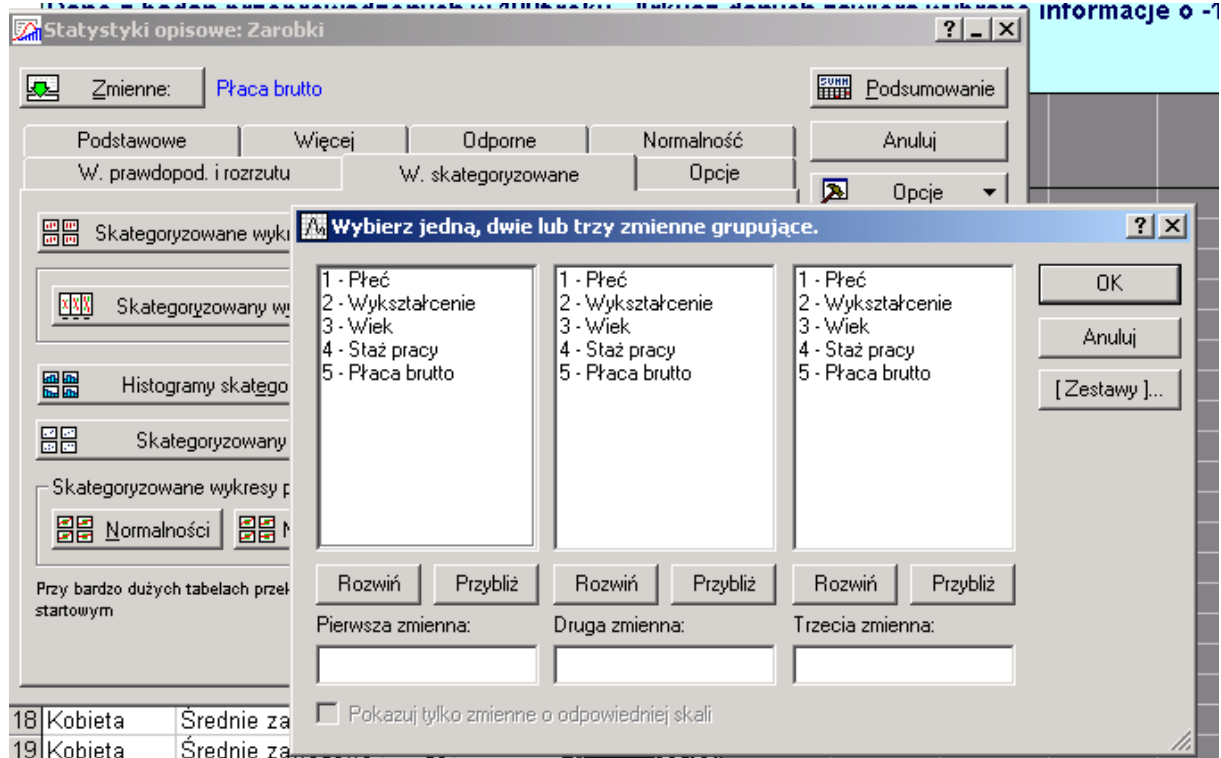
| Statystyki opisowe (Zarobki) | | N ważnych | % Ważnych | Średnia | Ufność -95,000% | Ufność 95,000 | Średnia przycięta 5,0000% | Średnia Winsora 5,0000% | Grubbsa statystyka | poziom p | Geometr. Średnia | Harmon. Średnia | Mediana |
|------------------------------|--|-----------|-----------|----------|-----------------|---------------|---------------------------|-------------------------|--------------------|----------|------------------|-----------------|----------|
| Zmienna | | | | | | | | | | | | | |
| Wiek | | 1246 | 99,2829 | 39,0658 | 38,5351 | 39,5965 | 38,9893 | 38,9904 | 3,134974 | 1,000000 | 37,8351 | 36,5512 | 40,0000 |
| Staż pracy | | 1227 | 97,7689 | 10,6406 | 10,1148 | 11,1664 | 9,9882 | 10,4866 | 3,233980 | 1,000000 | | | 8,0000 |
| Płaca brutto | | 1255 | 100,0000 | 808,6739 | 787,8162 | 829,5316 | 775,4726 | 791,5893 | 5,818169 | 0,000012 | 736,5090 | 673,6504 | 720,0000 |

W oknie **Statystyki opisowe** jest dostęp do większej liczby opcji związanych z opisową analizą danych. Wybieram Płace brutto i z zakładki **Opcje**, w polu **Opcje wykresu ramka wąsy** usuwamy pole **Średnia/odch....** a zaznaczymy **Mediana/kwartyle /rozstęp**, przechodzimy na zakładkę **Podstawowe** i wybieramy **Wykres ramka wąsy**

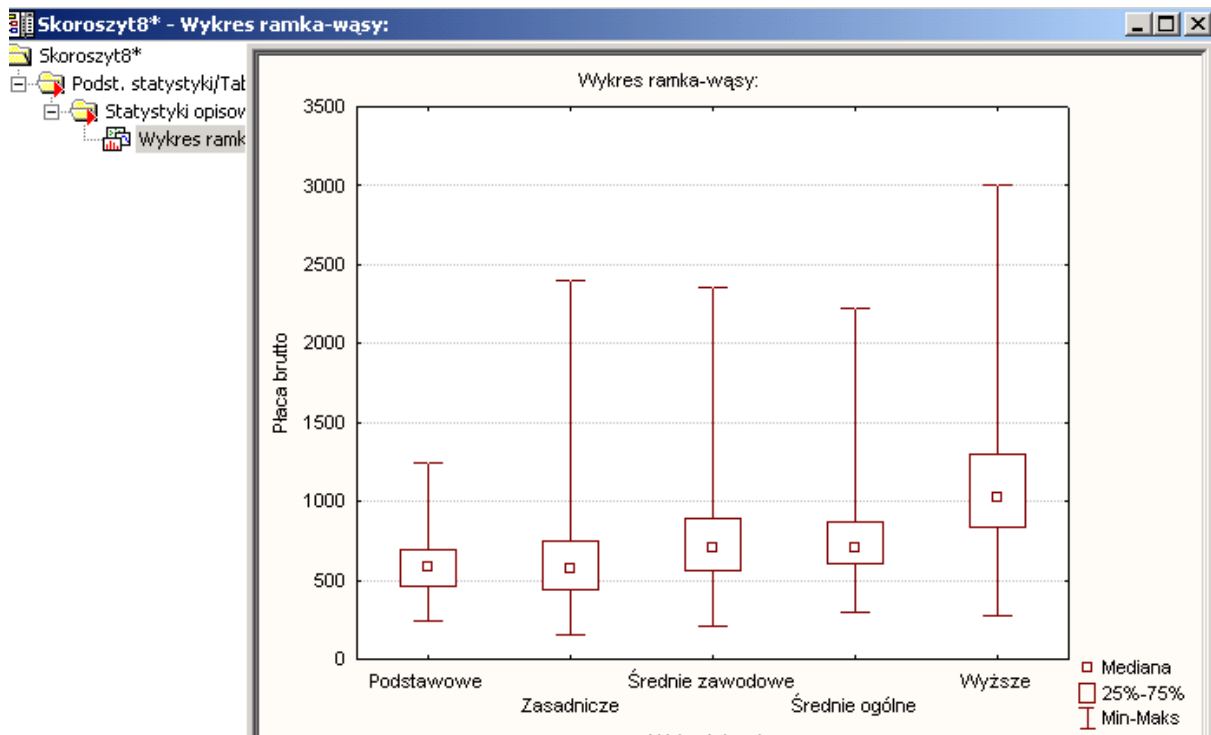


Przeanalizujemy jak kształtują się zarobki w poszczególnych kategoriach wykształcenia.

Wybieramy **Statystyki opisowe/ W.skategoryzowane/skategoryzowan, e wyk. Ramka-wąsy**, wybieramy zmienną grupującą **Wykształcenie**,

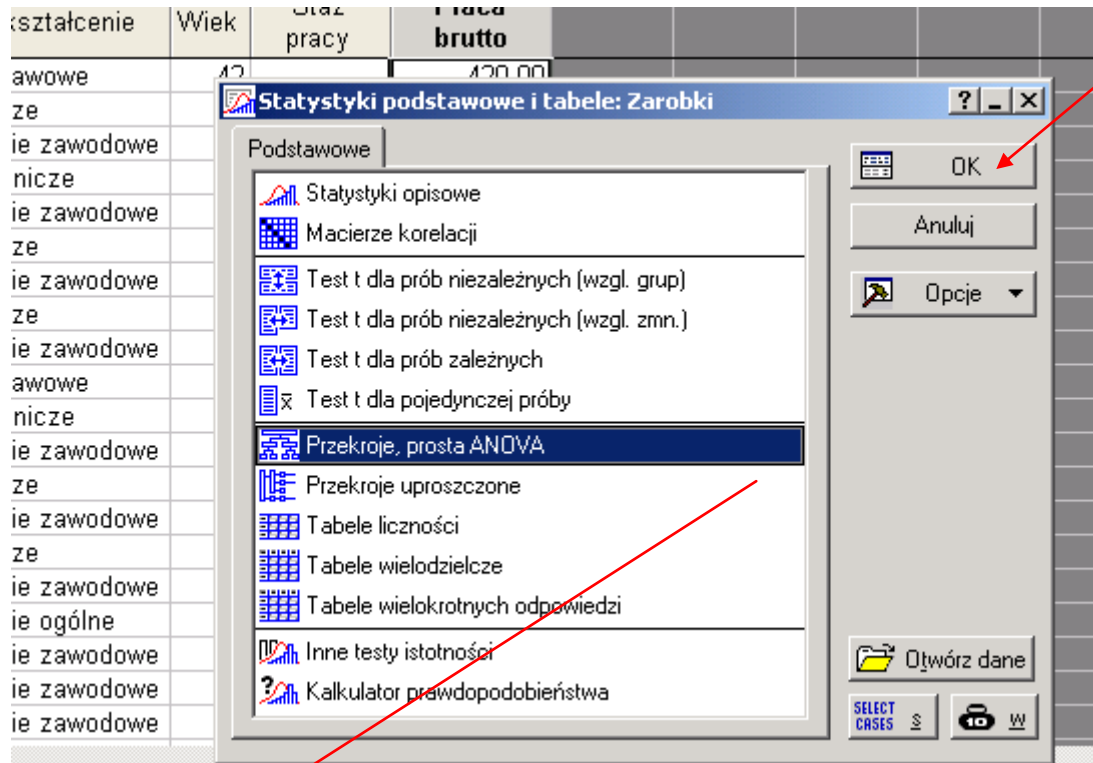


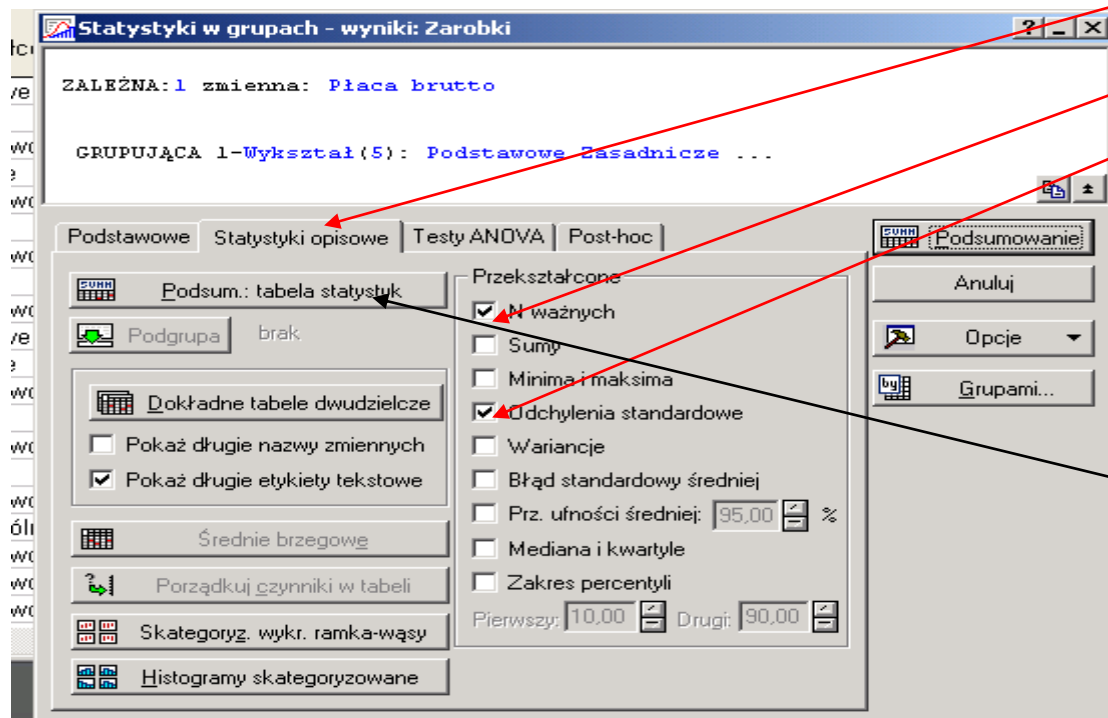
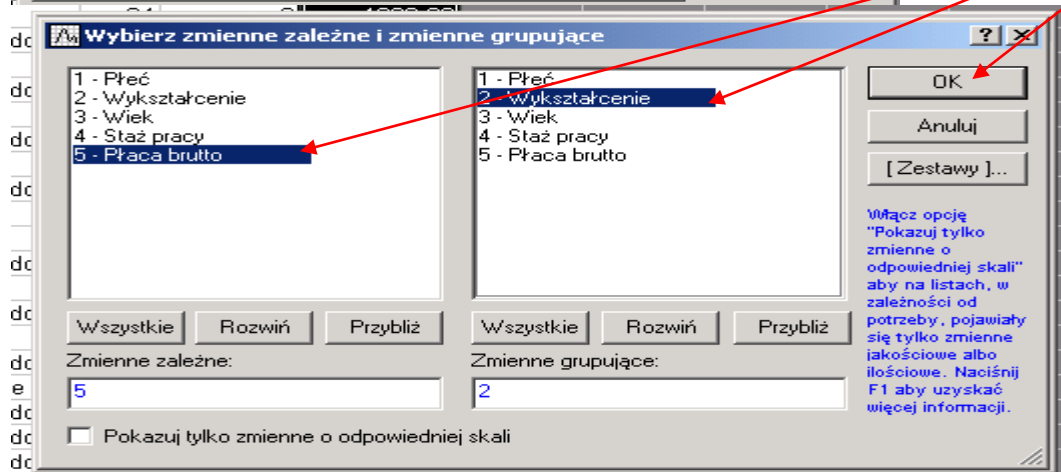
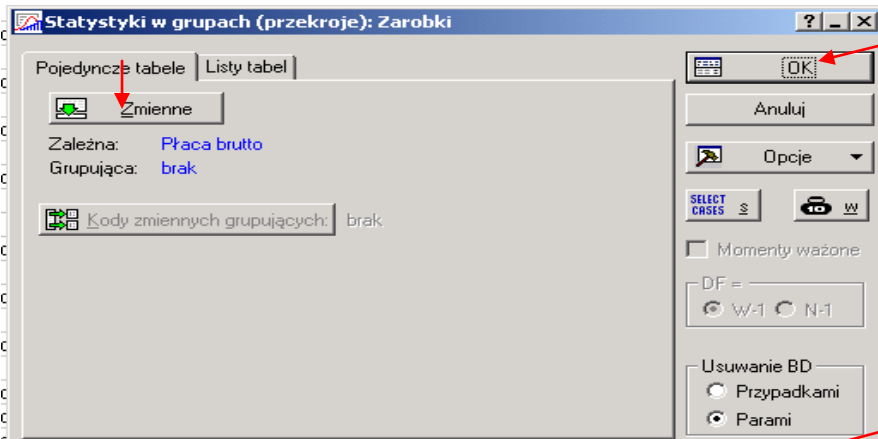
Otrzymujemy następujący wykres:



ANALIZA PRZEKROJOWA

Wybierzmy Statystyka / Statystyki podstawowe i tabele/przekroje, prosta ANOVA



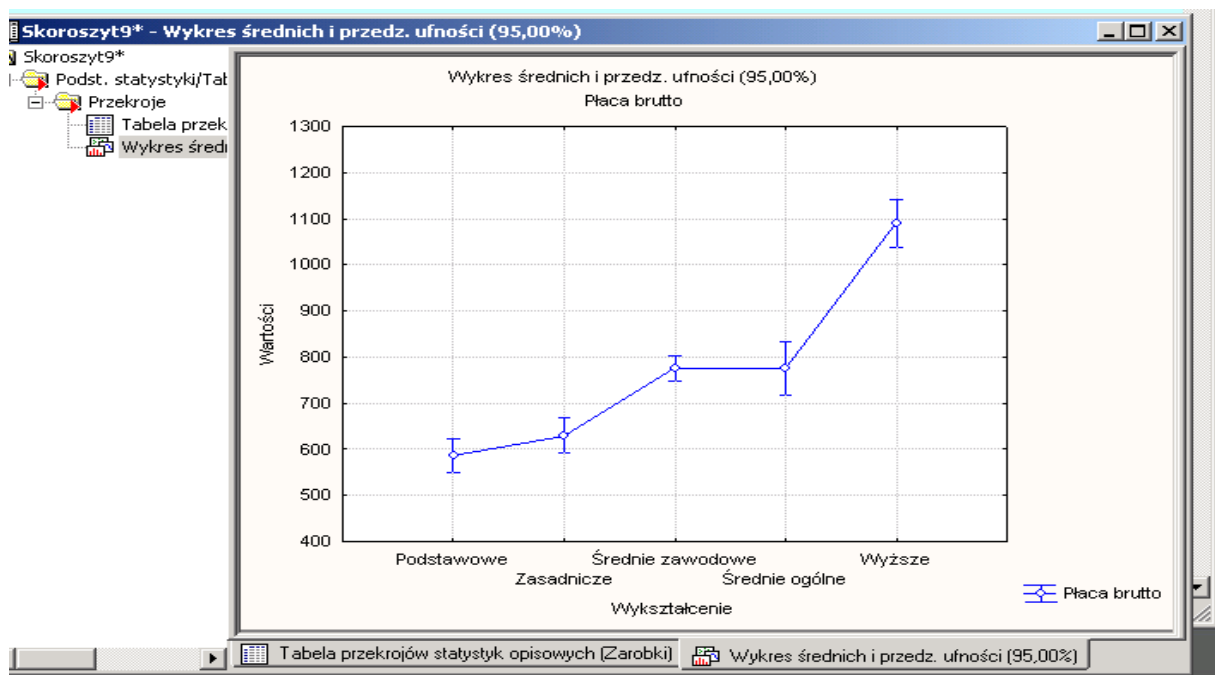
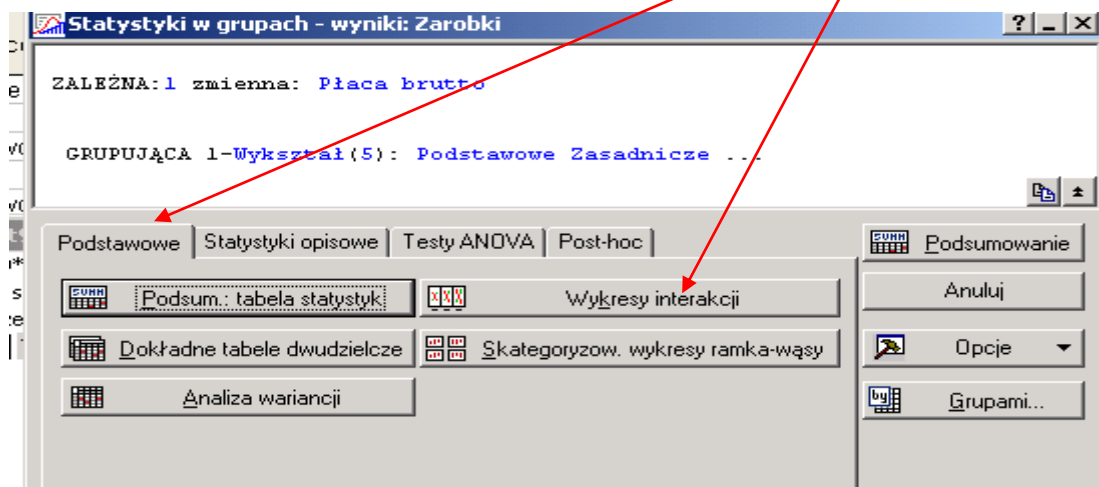


Skoroszyt9* - Tabela przekrojów statystyk opisowych (Zarobki)

Tabela przekrojów statystyk opisowych (Zarobki)
N=1255 (Zmienne zależne nie zawierają BD)

| Wykształcenie | Płaca brutto Średnie | Płaca brutto ważnych | Płaca brutto Odch.std |
|------------------|----------------------|----------------------|-----------------------|
| Podstawowe | 586,490 | 81 | 168,3500 |
| Zasadnicze | 629,555 | 214 | 278,9317 |
| Średnie zawodowe | 776,350 | 571 | 328,9721 |
| Średnie ogólne | 775,483 | 110 | 307,3914 |
| Wyższe | 1089,807 | 279 | 438,3830 |
| Ogół grp. | 808,674 | 1255 | 376,6350 |

Graficznie przedstawimy średnie: Ctrl+R:

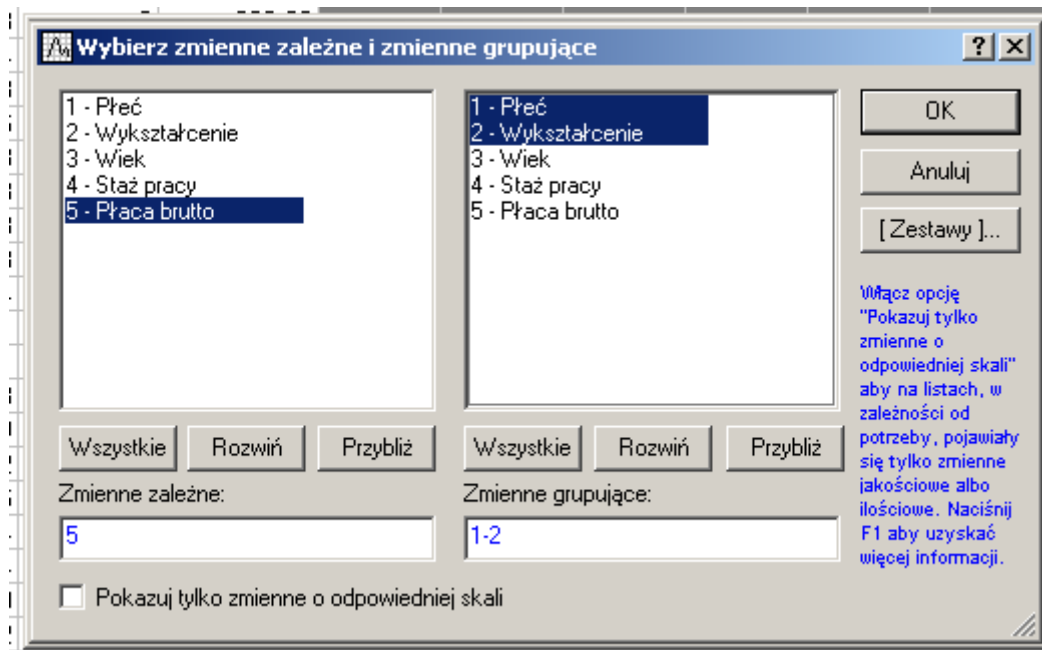


Otrzymany wykres można przekształcać według potrzeb zmieniając napisy, prawym przyciskiem myszki klikamy w obrębie wykresu, otwieramy okno właściwości wykresu i wybieramy przycisk np. tytuły wykresu..

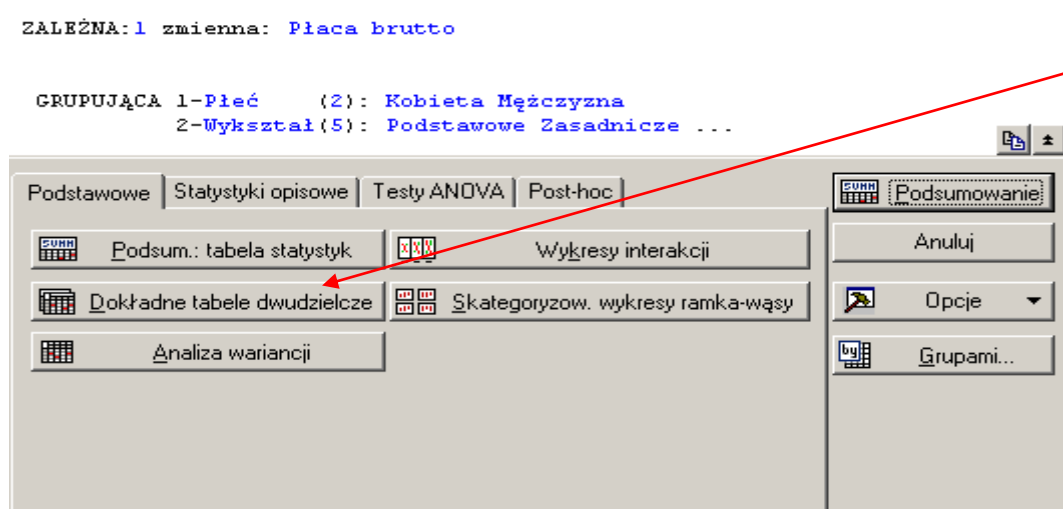
Analiza podzbiorów danych skategoryzowanych według wykształcenia i płci

CTRL+R

Statystyka/Statystyki podstawowe, tabele// przekroje proste ANOVA /Statystyki w grupach -



OK./ Podsum: tabela statystyk



otrzymujemy

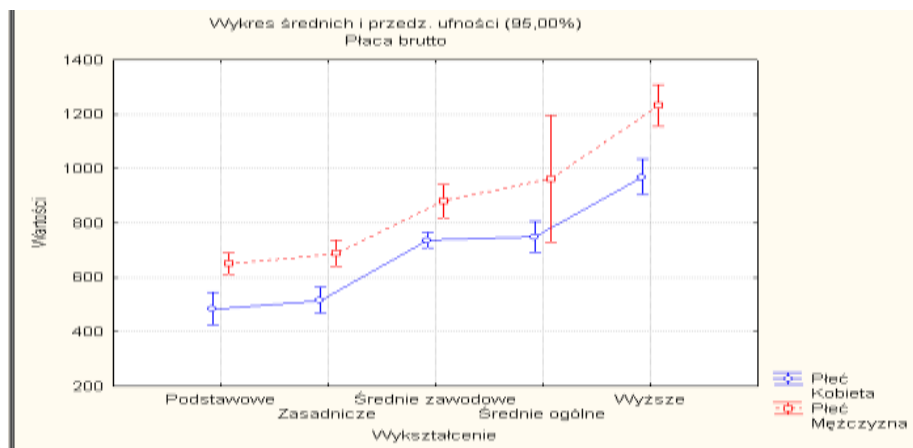
Skoroszyt1* - Tabela przekrojów statystyk opisowych (Zarobki.sta)

- Skoroszyt1*
- Podst. statystyki/Tabele (Za
- Przekroje
- Tabela przekrojów st

Tabela przekrojów statystyk opisowych (Zarobki.sta)
N=1255 (Zmienne zależne nie zawierają BD)

| Płeć | Wykształcenie | Płaca brutto Średnie | Płaca brutto ważnych | Płaca brutto Odch. std |
|-----------|------------------|----------------------|----------------------|------------------------|
| Kobieta | Podstawowe | 483,533 | 31 | 166,5363 |
| Kobieta | Zasadnicze | 515,506 | 72 | 210,3443 |
| Kobieta | Średnie zawodowe | 737,109 | 415 | 297,4444 |
| Kobieta | Średnie ogólne | 748,276 | 96 | 282,8746 |
| Kobieta | Wyższe | 969,692 | 151 | 405,2983 |
| Mężczyzna | Podstawowe | 650,323 | 50 | 135,7496 |
| Mężczyzna | Zasadnicze | 687,383 | 142 | 291,9642 |
| Mężczyzna | Średnie zawodowe | 880,741 | 156 | 382,8584 |
| Mężczyzna | Średnie ogólne | 962,051 | 14 | 405,6590 |
| Mężczyzna | Wyższe | 1231,505 | 128 | 435,0523 |
| Ogół grup | | 808,674 | 1255 | 376,6350 |

Przywołujemy z paska analizy Statystyki w grupach i na zakładce podstawowe wybieram przycisk wykresy interakcji, zaznaczam w pierwszym okienku wykształcenie, w drugim płeć i



otrzymuję wykres