Analiza pliku: „insurance.xlsx”

Jest to przykładowy zbiór danych na temat klientów firmy ubezpieczeniowej.

Charakteryza zbioru danych:

- Mamy 7 zmiennych: Wiek, Płeć, BMI, Liczba Dzieci, Status Palacza, Region Zamieszkania i Wysokość opłaty za ubezpieczenie.

- Zmienne ilościowe: Wiek ,BMI, Liczba Dzieci, Wysokość opłaty za ubezpieczenie.

- Zmienne jakościowe: Płeć, Status Palacza, Region Zamieszkania.

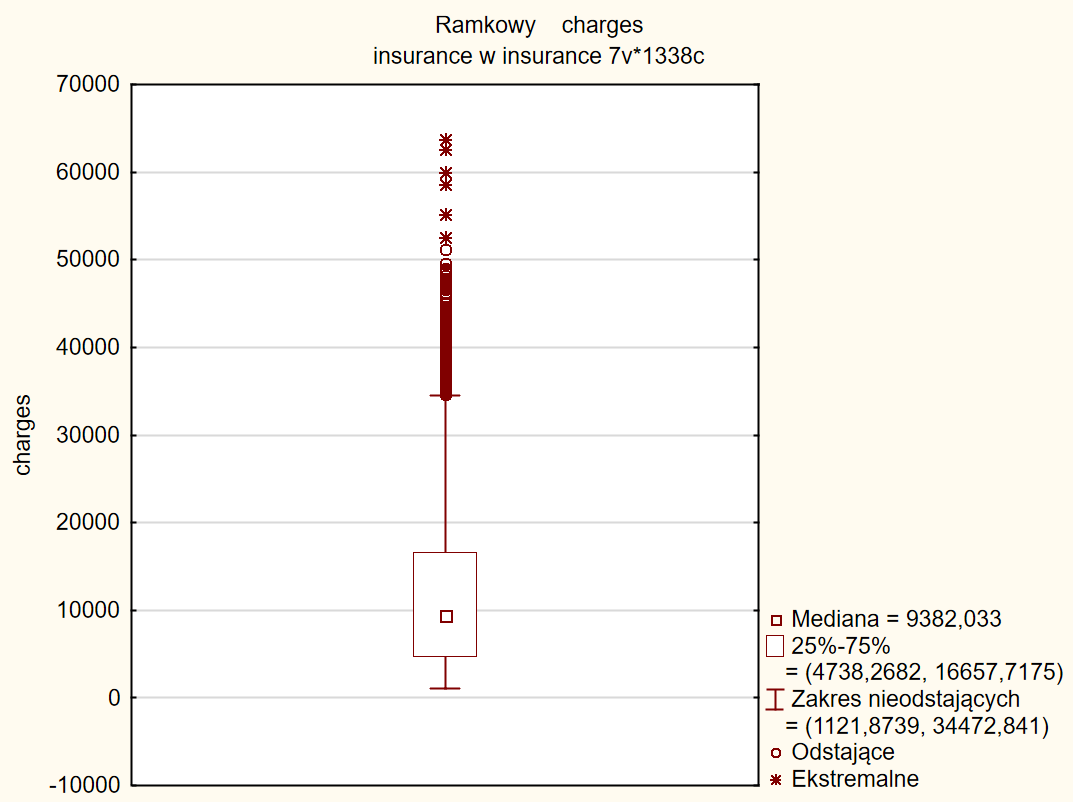
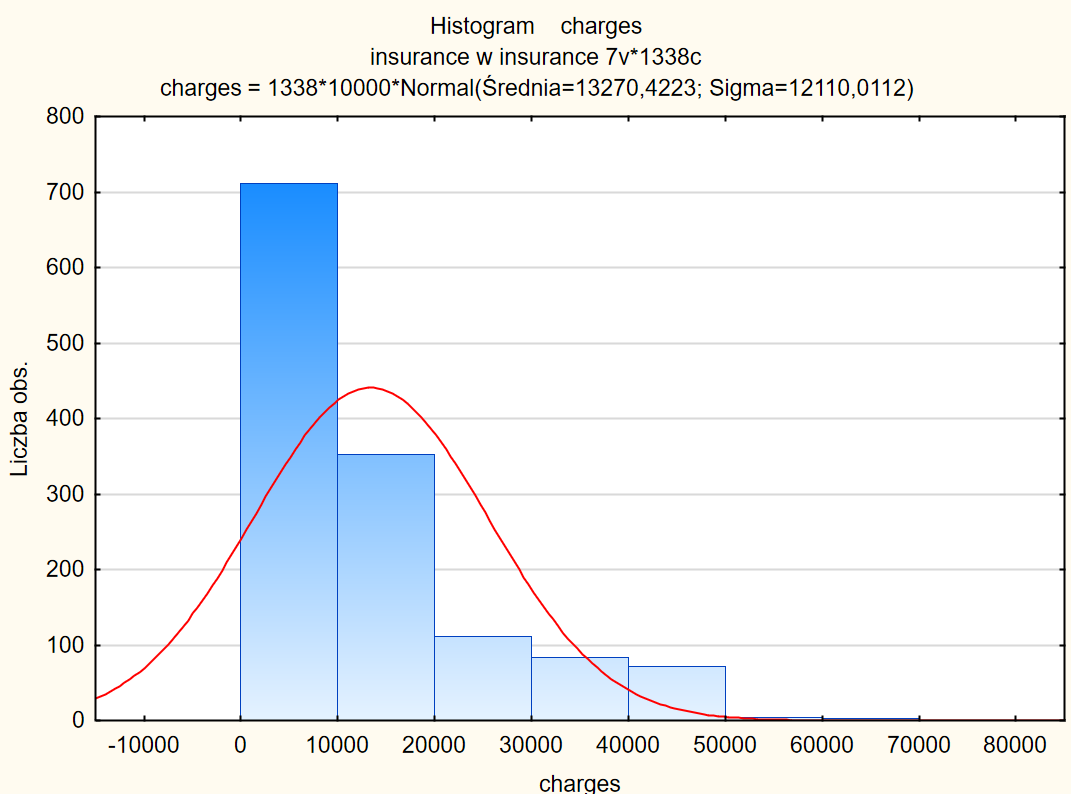
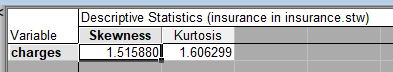
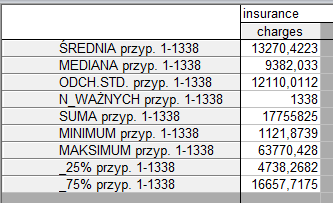
- Skala Nominalna: Płeć, Status Palacza, Region Zamieszkania

- Skala Porządkowa: Brak

- Skala Interwałowa: Brak

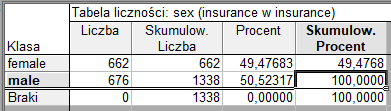
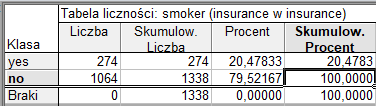
- Skala ilorazowa: Wiek, Bmi, Liczba Dzieci, Wysokość opłaty za ubezpieczenie.

Zmienna „Charges”( Wysokość opłaty za ubezpieczenie):



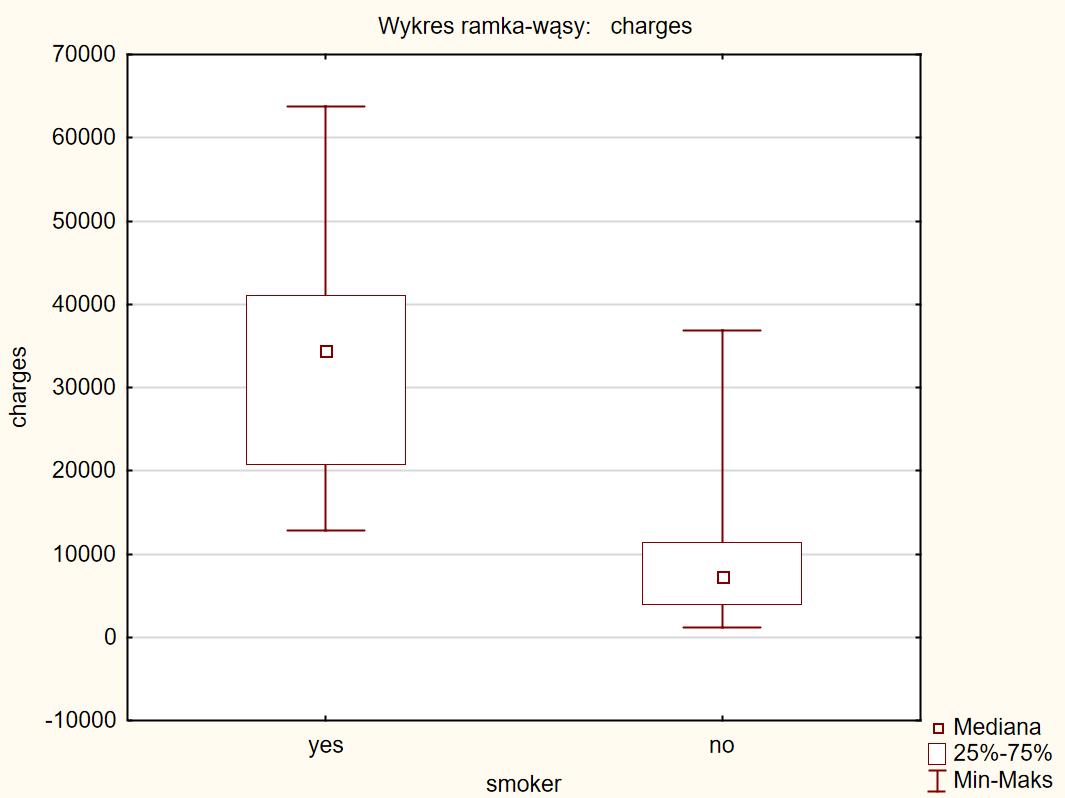
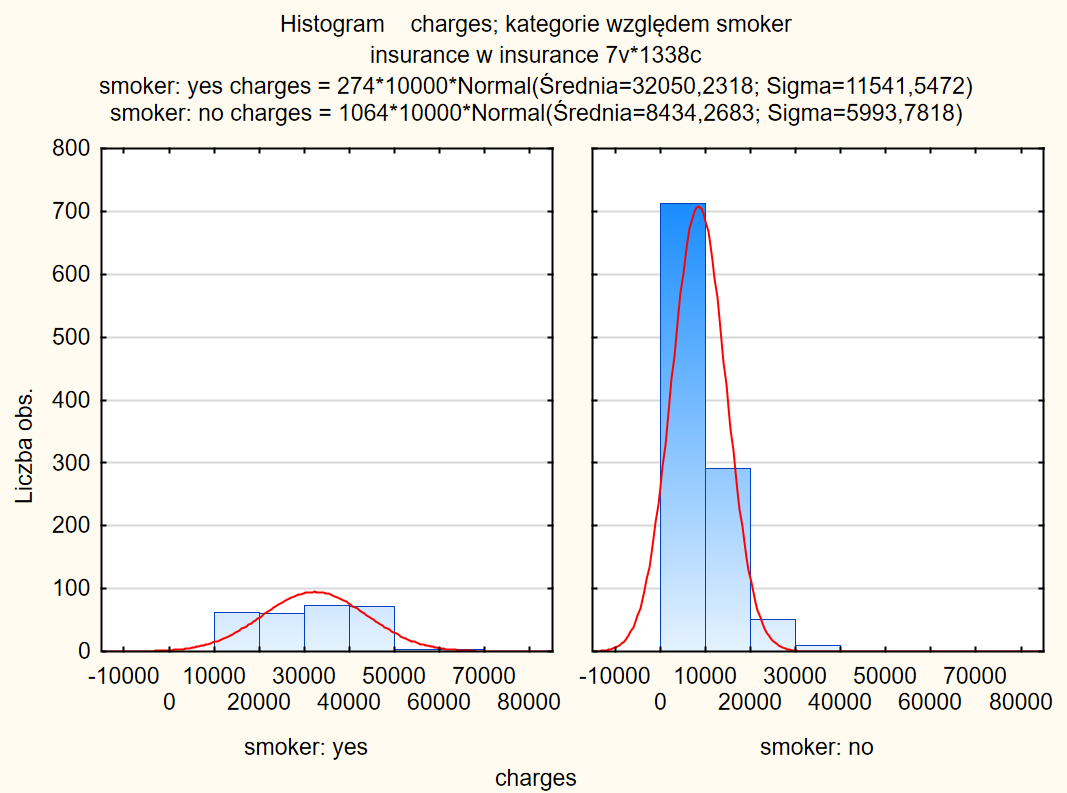
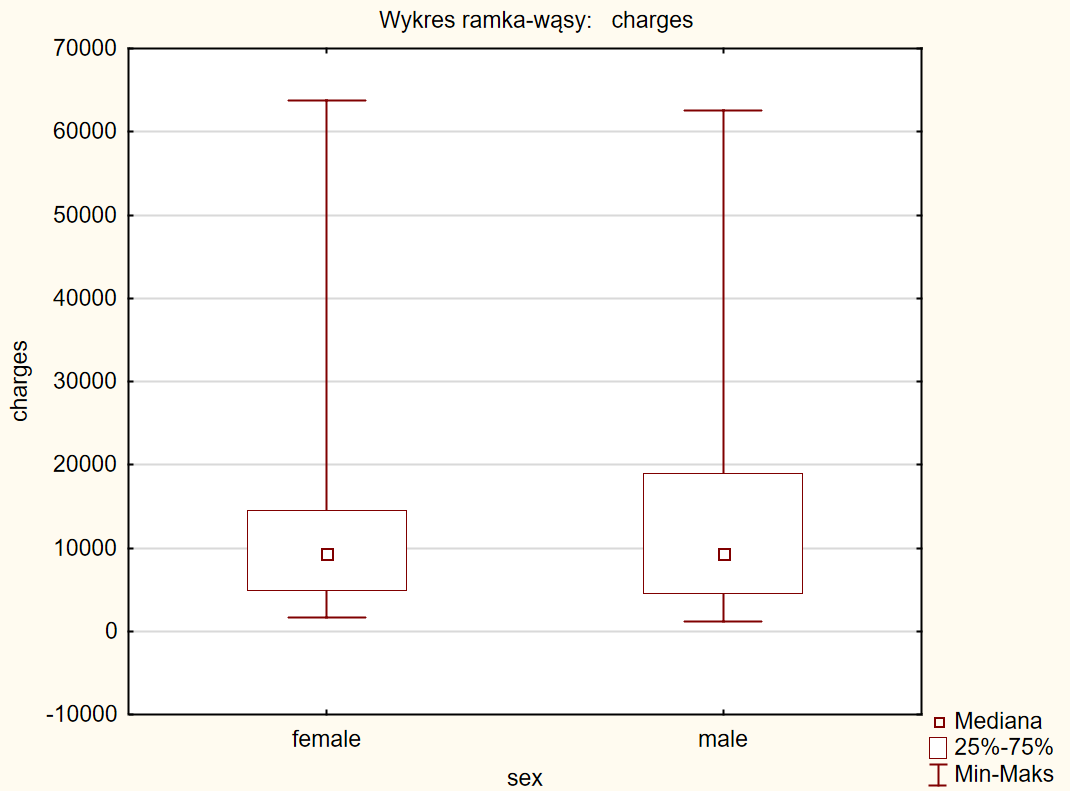
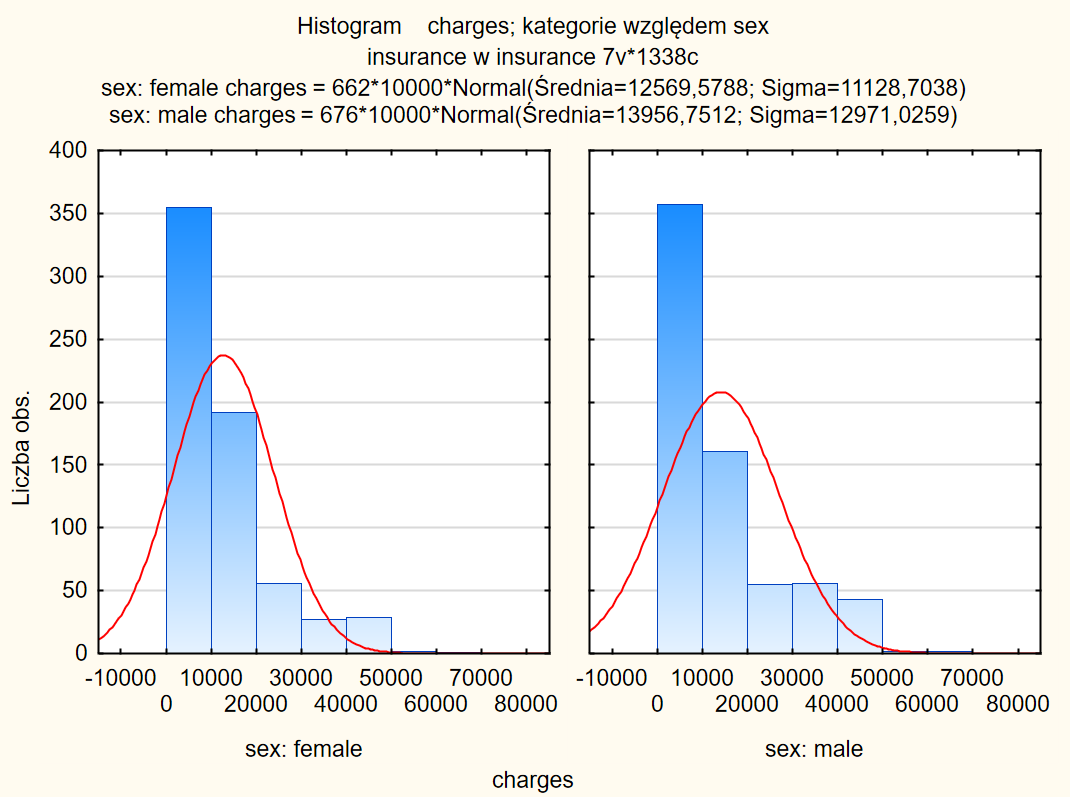
Z tego wynika że średnia wynosi 13270 a mediana 9382 pomimo większości opłat w przedziale 0-20000 na co wpływ mają odstające i ekstremalne wartości z największą wynoszącą 63770. Skośność jest prawostronna a kurtoza jest poniżej 3 a więc dane są skupione bardziej w pobliżu średniej i mają mniejsze wartości niż w przypadku rozkładu normalnego . Odchylenie standardowe wynosi 12110. Zarówno długość wąsów jak i ilość wartości odstających i ekstremalnych jest większa w górnej części wykresu.

4.Tabele liczności dla płci i statusu palacza:



Palacze stanowią zaledwie 20% próby, natomiast różnice między płciami są minimalne. Znaczne różnice są jedynie między ilością palących a niepalących, co prawdopodobnie pokrywałoby się z ogółem populacji biorąc pod uwagę historyczny spadek palenia.

5.



W przypadku podziału na płci nie występują znaczące różnice, aczkolwiek składki mężczyzn mają większe pudełko. Natomiast w przypadku różnic między palaczami a niepalącymi są one już znaczące. Poza ogólnie wyższymi stawkami pudełko jest znacznie większe dla palaczy, oraz położone wyżej osi y(opłata za ubezpieczenie).