# Sprawozdanie z projektu

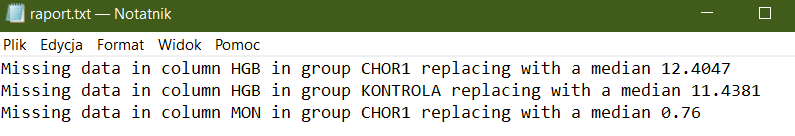
Weronika Koga

## Plik README

Porusza ważne kwestie m.in jak uruchomić program, aby uzyskać prawidłowe wyniki, jak przygotować dane, co jest potrzebne do wykonania analizy i zwracane jako rezultat działania programu.

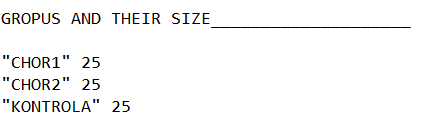
## Braki danych

Funkcja Replace\_blank\_with\_NA() odpowiada za zamianę ‘’ ‘’ występujących w pliku na NA.  
Następnie funkcja Remove\_NA() usuwa NA w kolumnach nienumerycznych oraz zastępuje NA medianą z danej grupy w kolumnach numerycznych. Wszystkie zmiany zapisywane są do pliku raport.txt. W przypadku kolumn numerycznych - w jakiej kolumnie i grupie wykryto brak w wartości oraz ile wynosi mediana, która jest wstawiana w to miejsce. W przypadku kolumn nienumerycznych – numer usuniętego wiersza.



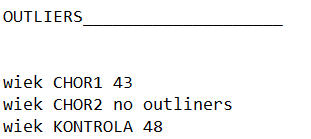
## Raport istniejących grup

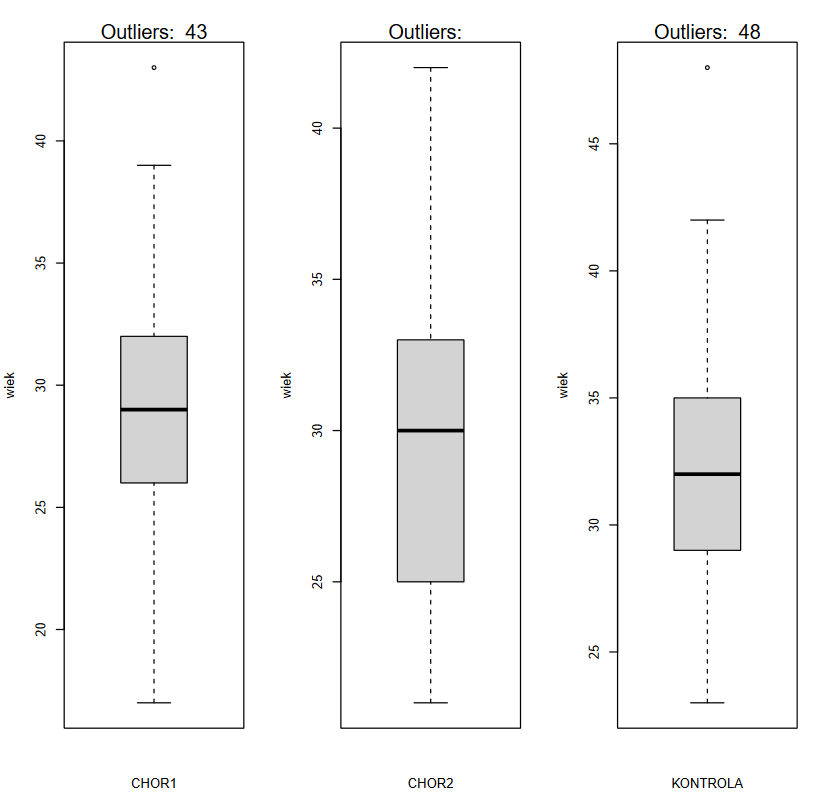
Funkcja count\_groups() odpowiada za wypisanie wszystkich istniejących w pliku csv grup, wraz z ilością wierszy do niej przypisanych, do pliku raport.txt.



## Wartości odstające

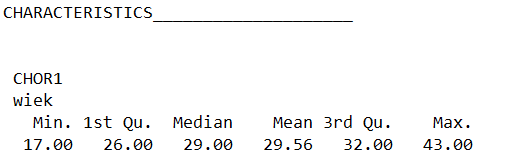
Funkcja Outliers\_detection() zapisuje do pliku raport.txt wartości odstające w każdej kolumnie numerycznej i dla każdej grupy. Tworzy również boxploty uwidaczniające wartości odstające lub ich brak zapisywany do pdf „Outliers”.   
Przykład :





## Charakterystyka badanych grup

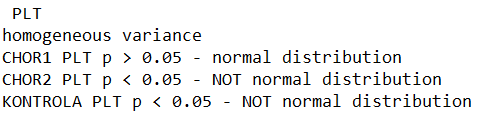
Funkcja Characteristics() dla każdej grupy w każdej kolumnie numerycznej przygotowuje raport minimalnej i maksymalnej wartości, średniej, mediany oraz 1 i 3 kwartylu.

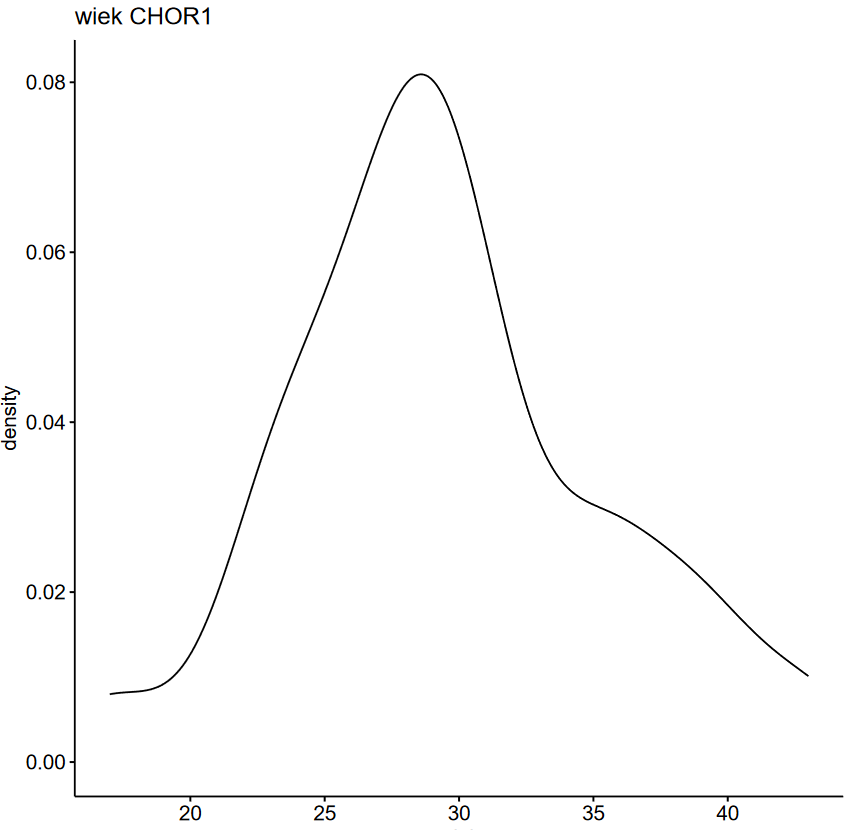


## Analiza porównawcza pomiędzy grupami

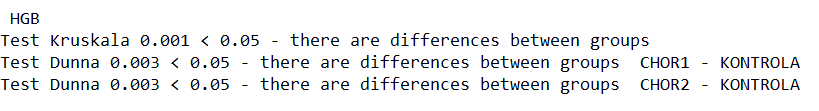
Dla każdego z testów przyjęłam wartość graniczną p.value=0.05

**Homogeniczność i rozkład normalny**   
Funkcja Homogenity\_of\_variance\_raport() sprawdza za pomocą testu Levene’a czy dla każdej kolumny numerycznej wariancja jest homogeniczna. Funkcja Normal\_distribution\_raport() przeprowadzając test Shapiro-Wilka sprawdza czy rozkład jest normalny. Wyniki tego sprawdzenia są raportowane do pliku raport.txt.   
Obie te funkcje wywoływane są w funkcji Density\_normal\_and\_homogenic\_info() która dodatkowo zwraca listę z wektorami :  
1.) nazw kolumn dla których istniała jakaś grupa która nie miała rozkładu normalnego  
2.) nazw kolumn z homogeniczną wariancją.   
Dodatkowo tworzony jest pdf „Density” z wykresami obrazującymi rozkłady w każdej grupie w kolumnie numerycznej.

Przykład:  




**Czy istnieją różnice pomiędzy grupami**

Funkcja Statistics\_test wywołuje funkcję Density\_normal\_and\_homogenic\_info() a następnie korzystając z tego co zwraca ta funkcja wywołuję funkcję Apply\_test().   
Apply\_test() korzystając z ilości grup oraz informacji które grupy dla jakich kolumn mają rozkład normalny i które kolumny mają wariancje homogeniczną, stosuje odpowiedni test statystyczny, którego nazwa, rezultat oraz to dla jakich danych był przeprowadzony, jest zapisywany do pliku raport.txt.  
Obecne testy:  
-Anova\_test (gdy grup jest więcej niż 2 a dane mają rozkład normalny i wariancję homogeniczną)  
-Tukey\_test (przeprowadzany automatycznie gdy test Anova wykaże p.value<0.05)  
-Kruskal\_test (gdy grup jest więcej niż 2 i warunek na Anova test nie jest spełniony)  
-Dunn\_test (przeprowadzany automatycznie gdy test Kruskala wykaże p.value<0.05)  
-T\_Student (gdy grupy są dwie a dane mają rozkład normalny i wariancję homogeniczną)  
-Welch (gdy grupy są dwie a dane mają rozkład normalny i niehomogeniczną wariancję)  
- Wilcoxon (gdy grupy są dwie a dane nie mają rozkładu normalnego)  
  
Przykład: 

## Analiza korelacji

Funkcja Correlation\_analysis() używa testu Spearmana i zapisuje do pliku raport.txt pomiędzy którymi kolumnami w obrębie jakiej grupy występuje korelacja oraz jaka jest jej siła i kierunek. Dodatkowo tworzony jest plik pdf „Correlations” z wykresami korelacji o ile takie istnieją. Regresja liniowa wskazuje kierunek korelacji a pole dookoła niej reprezentuje siłę korelacji pomiędzy danymi. Punkty dookoła reprezentują dane. Tworzony jest również plik pdf „Heatmaps” z heatmapą korelacji kolumn w obrębie każdej grupy (w funkcji Generate\_heatmap()).

Przykład:

