**SPRAWOZDANIE PROJEKT ANALIZA STATYSTYCZNA**

# Spis treści

# WSTĘP:

Witaj użytkowniku! Poniższe sprawozdanie, które właśnie czytasz, służy jako manual do narzędzia przeprowadzającego podstawową analizę statystyczną dla danych ilościowych. W tym dokumencie zawarte są kompleksowe wyjaśnienia dotyczące metod statystycznych stosowanych przez program, interpretacja dołączanych plików/wyników, oraz schemat działania skryptu. Aby wygodniej czytało się dołączany dokument, sprawozdanie jest sprezentowane w formie od ogółu do szczegółu, czyli poszczególne komponenty będą dzielone wedle działań podejmowanych przez program. Miłego czytania 😊

# UWAGI OGÓLNE:

- Dane wprowadzane przez użytkownika powinny być w formacie .csv,
- Program na samym początku zapyta o pełną ścieżkę do folderu, w którym mają zapisywać się tworzone pliki,
- Program dodatkowo wyświetla pomiędzy poszczególnymi segmentami komunikaty, pełniące rolę wprowadzenia do danego komponentu. To czas dla ciebie na analizę danych i przygotowanych wykresów,
- Wymagane pakiety (Hmisc, dplyr, data.table, purr, rlang, ggpubr, ggplot2, car, dunn.test, FSA)
- Program jest przygotowany tak by móc go odpalić w trybie batchowym (ergo: przez terminal). Poniżej schemat:
    - W terminalu wpisz komendę: cd ścieżka_do_pliku_Rscript.exe (zazwyczaj C:\Program Files\R\R-4.1.3\bin)
    - Następnie wklej następującą komendę: Rscript.exe ścieżka_do_programu.R ścieżka_do_pliku.csv
      (np:Rscript.exe C:\Users\User\OneDrive\Dokumenty\projekt_rpis.R C:\Users\User\OneDrive\Dokumenty\przykladoweDane-Projekt.csv)

# PRZYGOTOWANIE DANYCH I ANALIZA WARTOŚCI ODSTAJĄCYCH

## TEORIA:

Dane statystyczne często nie są idealne. W wyniku ludzkiego błędu mogą pojawić się chociażby braki w badanych instancjach. Tudzież omyłkowo, badanie może zostać przeprowadzone w niepoprawny sposób przez co pojawią się wartości znacząco odbiegające od uśrednionej normy. O ile kwestia wartości odstających jest czysto subiektywna i ich ocena zależy od analityka i charakteru badania, tak wybrakowane wartości stanowią poważne zagrożenie dla analizy, ponieważ często uniemożliwiają przeprowadzenie niezbędnych operacji (takich jak testy zgodności). Dlatego takie wartości muszą zostać w jakiś sposób wyeliminowane zanim rozpocznie się prawdziwa analiza. Zazwyczaj w miejsce tych braków wstawia się medianę lub średnią, tak by zasymulować „realne” wyniki.

## SCHEMAT DZIAŁANIA ALGORYTMU:

Algorytm analizuje wprowadzone dane kolumna po kolumnie. Jeśli wykryje braki danych w danym miejscu, to informuje o tym użytkownika i prosi o potwierdzenie operacji zmiany. W przypadku zgody

na zamianę danych, algorytm ściąga średnią wartość z danej kolumny w obrębie danej grupy. Robi to dlatego, by w wypadku gdy pustych danych było więcej, w ramach jednej grupy, to wszystkie puste wartości w danej grupie i kolumnie, zostaną zamienione na to samo. Po przeprowadzeniu operacji zamiany, algorytm przystępuje do detekcji wartości odstających. Podobnie jak w przypadku mechanizmu wyciągania średniej, tak i tu analiza wartości odstających i ich miejsc występowania jest w obrębie danej grupy. Dodatkowo jest przygotowywany wykres o nazwie outliers.pdf który w formie graficznej reprezentuje wartości odstające w formie wykresu pudełkowego (boxplot) (patrz punkt WYKRES). Opisanie informacji zwrotnej zostanie zaprezentowane w podpunkcie FORMAT.

## FORMAT:

INFORMACJE DLA KOLUMNY: *Nazwa_badanej_kolumny*

---

Puste krotki: 0 → *Ilość pustych krotek (lub kratek) w danej kolumnie*

Wartosci odstajace (grupa*: grupa_w_tabeli* ): x //// Polozenie: y → *Wskazanie wartości odstających w danej grupie wraz z ich położeniem w danym miejscu w danej grupie*

---

**Przykładowa interpretacja:**

INFORMACJE DLA KOLUMNY: wiek

---

Puste krotki: 0

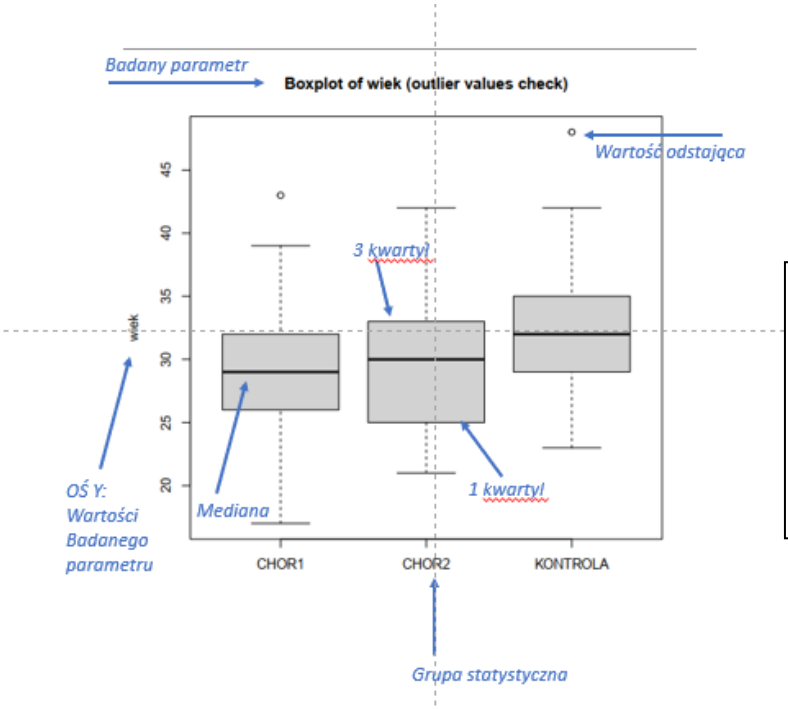
Wartosci odstajace (grupa: CHOR1 ): 43 //// Polozenie: 6 → *Wartość 43 będąca w wierszu 6 w ujęciu grupy CHOR1 odstaje od reszty wartości w kolumnie „wiek" w grupie CHOR1*

Wartosci odstajace (grupa: CHOR2 ): //// Polozenie: → *W grupie CHOR2 nie ma wartości odstających*

Wartosci odstajace (grupa: KONTROLA ): 48 //// Polozenie: 5

---

## WYKRES:




**NOTKA:** W przypadku braku wartości odstających, kropki nie występują. W celu wyjaśnienia pojęć takich jak mediana, 1/3 kwartyl sprawdź punkt TEORIA w części CHARAKTERYSTYKA GRUP

# CHARAKTERYSTYKA GRUP:

## TEORIA:

Po przygotowaniu danych do właściwej pracy, wypadałoby jeszcze przed właściwą analizą, sprawdzić czym właściwie one się charakteryzują. Do scharakteryzowania poszczególnych danych służą pewne matematyczne/statystyczne pojęcia. W programie użyte są następujące:

- **Minimum** → Najmniejsza wartość w danym zbiorze
- **Pierwszy kwartyl** → Jeśli posegregujemy wartości w zbiorze rosnąco, to pierwszy kwartyl to wartość stanowiąca 25% zbioru. Przykładowo – mając zbiór [1,2,3,4,5,6,7,8] pierwszym kwartylem byłaby liczba 2.
- **Mediana** → Wartość środkowa. Analogicznie do przykładu powyżej, medianą byłaby liczba ((4+5)/2) = 4.5
- **Średnia arytmetyczna** → Suma wszystkich wartości podzielona przez ilość wartości
- **Trzeci kwartyl** → To wartość stanowiąca 75% zbioru. W przykładzie z 1szego kwartylu, 3ecim kwartylem byłaby liczba 6.
- **Maksimum** → Największa wartość w zbiorze

Taka charakterystyka służy głównie jako informacja zwrotna dla analityka, którym grupom, parametrom warto się przypatrzeć przed wykonaniem dalszego testu

## SCHEMAT DZIAŁANIA ALGORYTMU:

Algorytm, w celu wykonania charakterystyki, dzieli dane użytkownika na „partie” zgodnie z analizowanymi grupami. Następnie za pomocą narzędzia map() z pakietu purrr wyświetla zbiorcze podsumowanie dla wszystkich partii w formie tabelarycznej. Dodatkowo zapisuje tabelę do pliku summary.txt. Jako że format w programie i format pliku stanowczo się różnią, zalecane jest przeczytanie sekcji FORMAT i PLIK

## FORMAT:

$*Badana_grupa*

*Badany_parametr*

Min. x → *Minimum*

1st Qu. X → *Pierwszy kwartyl*

Median. X → *Mediana*

Mean. X → *Średnia arytmetyczna*

3rd Qu. X → *Trzeci kwartyl*

Max. x → *Maksimum*

**Przykładowa interpretacja:**

```
$CHOR1
      wiek
Min.   :17.00
1st Qu.:26.00
Median :29.00
Mean   :29.56
3rd Qu.:32.00
Max.   :43.00
HGB
```

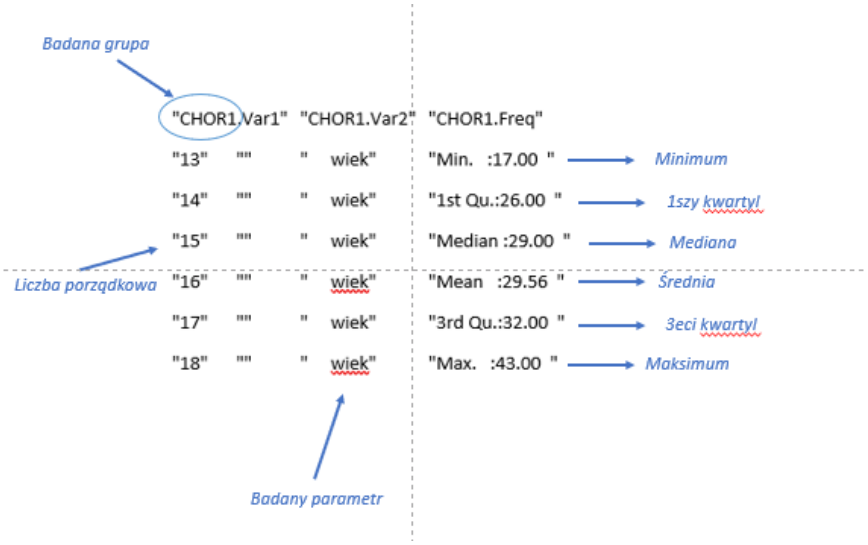
W obrębie grupy CHOR1, dla parametru wiek:

Wartość minimalna wynosi 17

Średnia arytmetyczna wynosi 29.56

Wartość maksymalna wynosi 46 *(najstarszy uczestnik badania miał 46 lat)*

PLIK:




# ANALIZA PORÓWNAWCZA

## TEORIA:

Każda statystyka ma na celu potwierdzenie lub obalenie jakiejś hipotezy przyjętej przed rozpoczęciem badania. W celu potwierdzenia/odrzucenia hipotezy stosuje się tak zwane testy statystyczne. To one pozwalają określić chociażby to, czy pomiędzy grupami i parametrami ilościowymi które badamy istnieją jakieś niezwykle ważne różnice, które mogą potwierdzać/dewaluować naszą hipotezę statystyczną. Jednakże dobranie odpowiedniego testu zależy od wielu czynników. (1) Między innymi najpierw musimy określić charakter grup. W przypadku grup zależnych, będziemy badali te same parametry ale w odstępie czasu. W przypadku grup niezależnych mówimy o badaniu tych samych parametrów, ale u różnych grup pacjentów. (2) Następnie określamy ile grup tak naprawdę mamy do zbadania. Różne testy są przystosowane do różnych ilości grup. (3) W kolejnym kroku przechodzimy do zebrania informacji o charakterze danych (patrz: TEST SHAPIRO-WILKA (ZGODNOŚĆ Z ROZKŁADEM NORMALNYM) i TEST LEVENE'A (HOMOGENICZNOŚĆ WARIANCJI)). (4) Po ocenieniu wszystkich wyżej opisanych czynności można przystąpić do wyboru testu. Należy podkreślić to, że w przypadku testów dotyczących więcej niż 2 grup, należy jeszcze przystąpić do przeprowadzenia analizy *post hoc,* która dokładnie wskazuje pomiędzy którymi grupami występują różnice. Poniżej grafika prezentująca wybór odpowiedniego testu:

Tablica 1: Wyboru testu statystycznego dla 2 i > 2 grup niezależnych.

| Porównanie grup niezależnych | | | |
|---|---|---|---|
| Ilość porównywanych grup | Zgodność z rozkładem normalnym | Jednorodność wariancji | Wybrany test |
| 2 | TAK | TAK | test t-Studenta (dla gr. niezależnych) |
| | | NIE | test Welcha |
| | NIE | - | test Wilcoxona (Manna-Whitneya) |
| >2 | TAK | TAK | test ANOVA (*post hoc* Tukeya) |
| | | NIE | test Kruskala-Wallisa (*post hoc* Dunna) |
| | NIE | - | |

Dalsza część segmentu TEORIA zostanie rozpisana w podpunktach odnoszących się do danych testów, w tym testów wymaganych by sprawdzić zgodność z rozkładem normalnym i jednorodność wariancji.

## TEST SHAPIRO-WILKA (ZGODNOŚĆ Z ROZKŁADEM NORMALNYM):

Test Shapiro-Wilka to test służący do oceny tego, czy zebrane dane są zgodne z rozkładem normalnym (ergo: zgodny z krzywą Gaussa). Istotnym parametrem tego testu jest parametr *p-value* świadczący o istotności statystycznej. Jeśli ten parametr ma wartość mniejszą niż 0.05, to znaczy że test osiągnął istotność statystyczną i dane nie posiadają rozkładu normalnego. Należy podkreślić jeszcze fakt że ten test prezentuje wartości *p-value* dla każdej grupy. W związku z tym, aby stwierdzić że dane w ujęciu ogólnym mają rozkład normalny, to każda grupa musi mieć wymagane *p-value.*

## TEST LEVENE'A (HOMOGENICZNOŚĆ WARIANCJI):

Test Levene'a to test którego zadaniem jest ocena tego, czy wariancja w podanym zbiorze danych jest równa w zakresie analizowanych grup. Podobnie jak w przypadku TESTU SHAPIRO-WILKA tu też mamy do czynienia z parametrem określającym istotność statystyczną testu – *p-value*. Jeśli *p-value* jest mniejsze niż wartość 0.05 to oznacza, że wariancje są niejednorodne (heterogeniczne) – ergo: Są różnice między wariancjami w porównywanych grupach. Homogeniczność wariancji pełni ważną rolę dla testu T-STUDENTA ponieważ  założenie o jednorodności wariancji pełni sporą rolę dla przeprowadzenia prawidłowo tego testu.

## TEST T-STUDENTA (DLA GRUP NIEZALEŻNYCH):

Jest to najczęściej stosowany test do porównywania średnich z dwóch niezależnych od siebie grup. Pierwotna hipoteza tego testu zakłada, że między porównywanymi grupami/parametrami nie ma żadnych różnic. O tym czy te różnice są czy nie świadczy parametr  *p-value.* W przypadku gdy ten parametr jest mniejszy niż 0.05, oznacza to że możemy odrzucić hipotezę zerową i przyjąć że istnieją różnice między grupami. W tym wypadku na podstawie parametru *mean* jesteśmy w stanie określić jak istotne są to różnice.

## TEST WELCHA:

Najprościej rzecz ujmując jest to uogólnienie testu T-studenta w przypadku gdy nasze dane cechują się różnorodną wariancją. W celu lepszego zrozumienia testu przeczytaj TEST T-STUDENTA (DLA GRUP NIEZALEŻNYCH).

## TEST WILCOXONA:

Jest to nieparametryczna alternatywa dla testu T-studenta. Często stosowany do ponownej analizy danych, by sprawdzić czy nie wystąpiły różnice po przeprowadzeniu eksperymentu. W przeciwieństwie do testu T-studenta, w teście Wilcoxona porównuje się mediany.

## TEST ANOVA:

Test Anova jest to test stosowany w przypadku, gdy mamy więcej niż 2 grupy badawcze. W schemacie jednoczynnikowym ma ona za zadanie określić czy jedna zmienna niezależna wpływa na wynik jednej zmiennej zależnej. W najbardziej ogólnym ujęciu, test polega na porównaniu wariancji międzygrupowej (ujęcie ogólne) do wariancji wewnątrzgrupowej (ujęcie szczegółowe). Podobnie jak w każdym innym teście, parametr *p-value* informuje o występowaniu/niewystępowaniu różnic

między grupami. Jeśli test wykaże że istnieją różnice, najczęściej stosuje się test HSD Tukeya (TEST POST HOC)

### TEST KRUSKALA-WALLISA

Jest to nieparametryczny (niewymagający zgodności i homogeniczności wariancji) wariant testu ANOVA. Różni się on między innymi tym od ANOVY, że tu porównywane ze sobą są mediany parametrów. Podobnie jak w każdym innym teście, parametr *p-value* informuje o występowaniu/niewystępowaniu różnic między grupami. Jeśli test wykaże że istnieją różnice, najczęściej stosuje się test Dunna (TEST POST HOC)

### TEST POST HOC

Test stosowany zazwyczaj jako uzupełnienie innego testu dotyczącego większej ilości grup niż 2. Ma na celu dokładnie ukazać pomiędzy jakimi grupami występują różnice. Przedstawia to zazwyczaj za pomocą parametru *p-value* – jeśli w jakiejś parze grup *p-value* jest mniejsze niż 0.05 to pomiędzy tymi grupami konkretnie występuje różnica.

## SCHEMAT DZIAŁANIA ALGORYTMU:

Algorytm analizuje dostarczone dane kolumna po kolumnie. Na samym początku analizuje dane pod kątem zgodności z rozkładem normalnym i pod kątem jednorodności wariancji. Z racji zaoszczędzenia użytkownikowi zbędnych informacji, Program wyświetla tylko wartości P-value danych testów oraz werdykt (Zgodny z rozkładem / Homogeniczny). W międzyczasie program zbiera dane p-value dla danego parametru do odpowiedniego kontenera i przerabia je na wartości TRUE/FALSE (odpowiednik wyżej opisanego werdyktu). Ten kontener będzie służył do wyboru odpowiedniego testu. Należy wspomnieć jeszcze, że w ramach testu na rozkład normalny program tworzy plik **density.pdf** graficznie reprezentujący rozkład danych. Po zatwierdzeniu komunikatu rozpoczynającego właściwe testy statystyczne, kontener oraz liczba grup są przekazywane do zewnętrznych funkcji. Na podstawie tych danych algorytm decyduje o właściwym teście dla analizowanego parametru. Użytkownik dostaje zwrotną informacje zarówno o nazwie testu, wartości p-value, graficznej reprezentacji testu, oraz o werdykcie dotyczącym różnic między grupami (Są/Nie są). W przypadku gdy takie różnice występują, algorytm informuje o przeprowadzanym teście post-hoc, wyświetlając jego nazwę oraz graficzną reprezentacje. NALEŻY PODKREŚLIĆ ŻE WARUNKIEM KTÓRY DECYDUJE ZARÓWNO O NIEZDANIU TESTÓW NA ROZKŁAD NORMALNY, HOMOGENICZNOŚCI WARIANCJI, WYSTĘPOWANIU RÓŻNIC ORAZ POMIĘDZY KTÓRYMI KONKRETNIE GRUPAMI WYSTĘPUJĄ RÓŻNICE JEST TO, CZY PARAMETR P-VALUE JEST MNIEJSZY OD 0.05

## FORMAT (TESTY NA ROZKŁAD I HOMOGENICZNOŚĆ):

WYNIKI TESTU SHAPIRO/LEVENE DLA PARAMETRU: *jakiś_parametr* → *Analizowany Parametr*

---

Shapiro p-values: x y z → *P-values z testu na zgodność z rozkładem normalnym*

Levene p-values : x → *P-value z testu na jednorodność wariancji*

WERDYKT: tekst → *Tekstowa reprezentacja wyników w formacie Zgodny/NIEZGODNY z rozkładem normalnym / Homogeniczny/NIE homogeniczny. W przypadku gdy parametr jest niezgodny z rozkładem normalnym, to werdykt dla homogeniczności nie wyświetli się (brak zgodności z rozkładem normalnym definiuje test z automatu patrz* **TEORIA** *tabela)*

---

**Przykładowa interpretacja:**

```
WYNIKI TESTU SHAPIRO|LEVENE DLA PARAMETRU: MCHC
-----------------------------------------------------------------
Shapro p-values: 0.227 0.283 0.274
Levene p-value: 0.2688996

WERDYKT: Zgodny z rozkladem normalnym \ Homogeniczny
-----------------------------------------------------------------
```

Wartości testu na rozkład normalny dla parametru MCHC wynoszą: 0.227, 0.283, 0.274

Wartość testu na homogeniczność wariancji dla parametru MCHC wynosi: 0.2688996

Na podstawie tego że wszystkie wartości p-value dla testu na rozkład normalny są większe niż 0.05, można stwierdzić że parametr jest Zgodny z rozkładem normalnym (WERDYKT)

Na podstawie tego że wartość p-value dla testu na homogeniczność wariancji jest większy niż 0.05, można stwierdzić że parametr jest Homogeniczny (WERDYKT)

## FORMAT (TEST STATYSTYCZNY):

### Ogólny wariant (Test bez różnic/różnice dla 2 grup):

ANALIZA DLA PARAMETRU: *jakiś_parametr → Analizowany parametr*

---

Wybrany test: *nazwa_testu → Dobrany test na podstawie ilości grup/testów Shapiro i Levene'a*

P-value: x → *Wartość p-value świadcząca o tym czy występują różnice czy nie występują*

*GRAFICZNA REPREZENTACJA TESTU → PATRZ KONKRETNE TEST PONIŻEJ*

WERDYKT *tekst → Tekstowa reprezentacja tego czy są różnice czy nie ma różnic*

---

**Przykładowa interpretacja będzie zaprezentowana w kontekście konkretnych testów.**

### Ogólny wariant (Test z różnicami dla > 2 grup):

ANALIZA DLA PARAMETRU: *jakiś_parametr → Analizowany parametr*

---

Wybrany test: *nazwa_testu → Dobrany test na podstawie ilości grup/testów Shapiro i Levene'a*

P-value: x → *Wartość p-value świadcząca o tym czy występują różnice czy nie występują*

*GRAFICZNA REPREZENTACJA TESTU → PATRZ KONKRETNIE DLA TESTÓW PONIŻEJ*

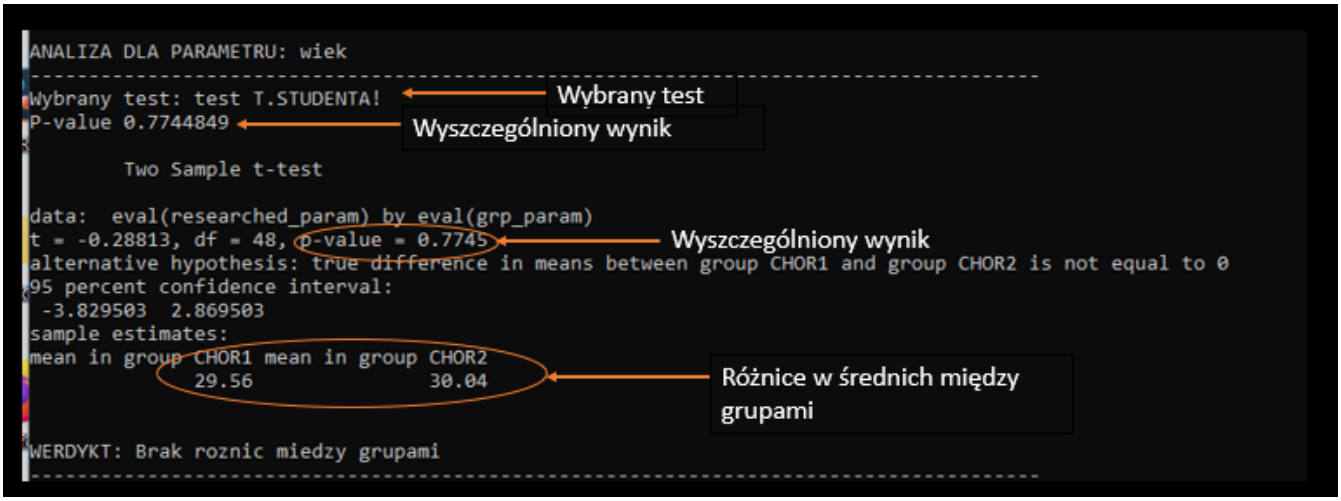
WERDYKT *tekst → Tekstowa reprezentacja tego czy są różnice czy nie ma różnic*

TEST: *NAZWA_TESTU* (post hoc aby sprawdzić pomiędzy konkretnie którymi grupami występują różnice) → *Nazwa testu post hoc dobranego aby zobaczyć, pomiędzy którymi grupami są różnice*

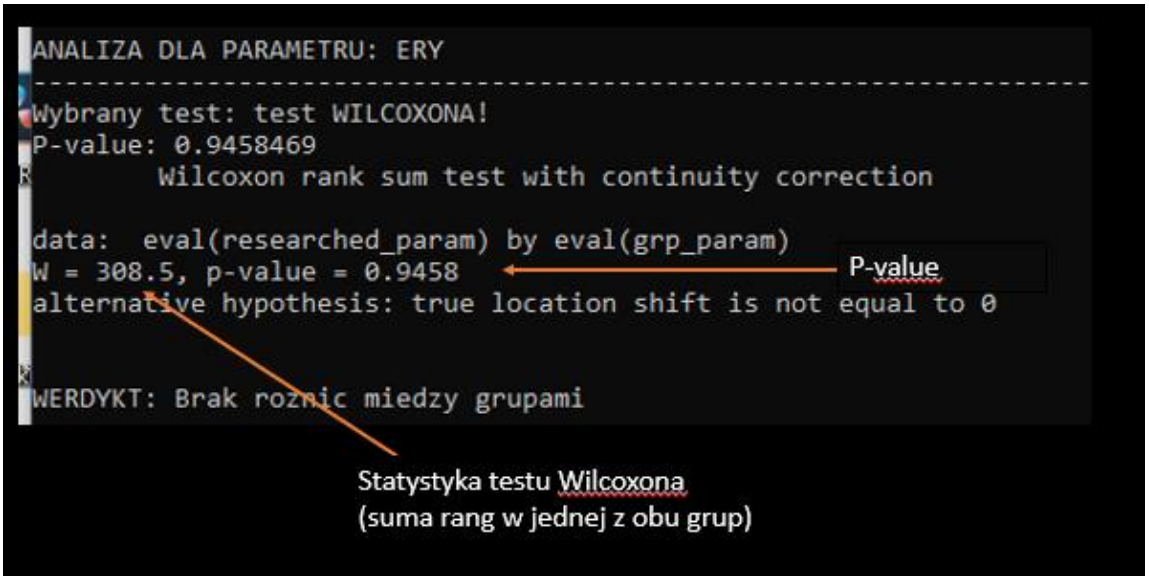
*GRAFICZNA REPREZENTACJA TESTU POST HOC → PATRZ KONRETNIE DLA TESTÓW PONIŻEJ*

**Przykładowa interpretacja będzie zaprezentowana w kontekście konkretnych testów.**

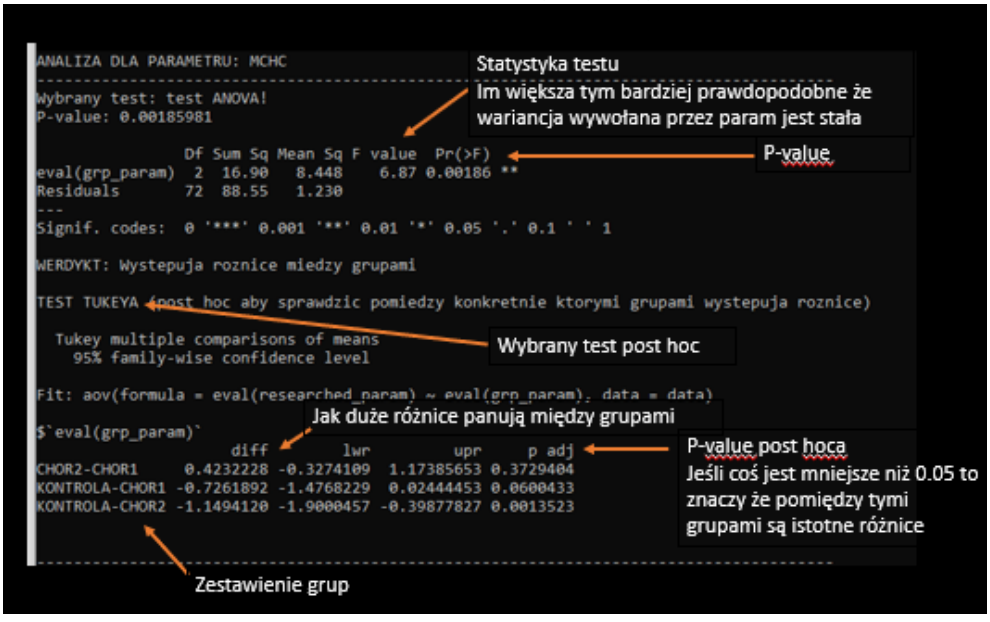
*PRZYKŁADOWA INTERPRETACJA TEST T-STUDENTA*



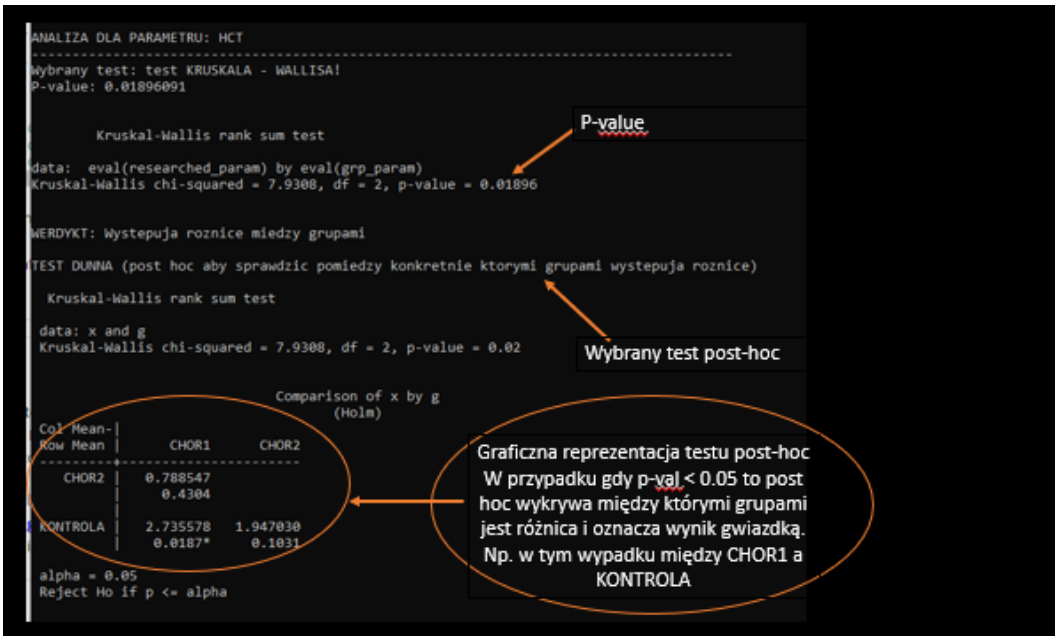

*PRZYKŁADOWA INTERPRETACJA TEST WILCOXONA*



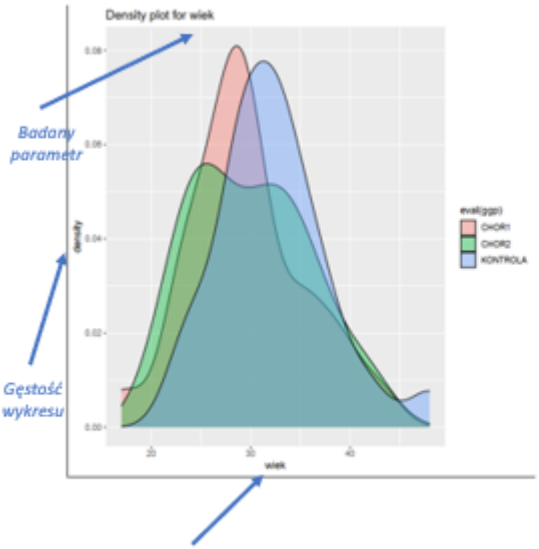

*PRZYKLADOWA INTERPRETACJA TEST ANOVA (razem z post hoc)*

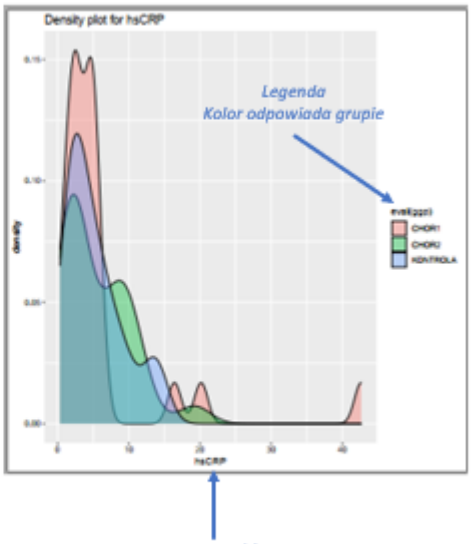


*PRZYKŁADOWA INTERPRETACJA TEST KRUSKALA (razem z post hoc)*




## WYKRES (Rozkład normalny)



# BADANIE KORELACJI

## TEORIA:

Podczas analizy statystycznej możemy wyszczególnić pewne, wcześniej niebadane związki pomiędzy poszczególnymi zmiennymi. Możemy przykładowo zauważyć, że gdy jedna zmienna maleje to druga gwałtownie rośnie. Albo gdy jedna zmienna maleje, druga delikatnie maleje. Takie związki nazwyamy korelacją. Korelacja statystyczna cechuje się dwoma poszczególnymi, opisującymi ją parametrami: Siła i kierunek. Kierunek korelacji określa to czy zmienna rośnie/maleje (ergo: porusza się w górę/w dół). Siła korelacji z kolei określa jak mocno zmienna maleje/rośnie (Silnie, średnio, słabo). Do analizy korelacji służy specjalny test korelacji. Cechują go dwa parametry: p-value i współczynnik korelacji. Jeśli p-value jest mniejsze od wartości 0.05, to oznacza że pomiędzy elementami istnieje korelacja. W tym wypadku przechodzimy do analizy siły i kierunku. Poniżej zaprezentowane są grafiki podsumowujące określenie w.w parametrów.

- $r > 0$ korelacja dodatnia – gdy zmienna X rośnie to Y także rośnie,
- $r = 0$ brak korelacji – gdy zmienna X rośnie to Y czasem rośnie a czasem maleje,
- $r < 0$ korelacja ujemna – gdy zmienna X rośnie to Y maleje.

*1 Analiza Kierunku Korelacji*

- $-1 < r < -0.7$ bardzo silna korelacja ujemna
- $-0.7 < r < -0.5$ silna korelacja ujemna
- $-0.5 < r < -0.3$ korelacja ujemna o średnim natężeniu
- $-0.3 < r < -0.2$ słaba korelacja ujemna
- $-0.2 < r < 0.2$ brak korelacji
- $0.2 < r < 0.3$ słaba korelacja dodatnia

- $0.3 < r < 0.5$ korelacja dodatnia o średnim natężeniu
- $0.5 < r < 0.7$ silna korelacja dodatnia
- $0.7 < r < 1$ bardzo silna korelacja dodatnia

*2 Analiza siły korelacji*

## SCHEMAT DZIAŁANIA ALGORYTMU:

Algorytm na początku segmentu, dzieli sobie dostarczone dane na segmenty wedle badanych grup. Następnie analizuje każdy segment osobno *(Pierwsza pętla)* wyosobniając kolumny numeryczne aż do przedostatniej *(Druga pętla)* , jednocześnie pobierając kolumnę znajdującą się „dalej" niż kolumna badana w *pierwszej pętli (Trzecia pętla).* Takie rozwiązuje gwarantuje, że zautomatyzujemy proces pobierania parametrów potrzebnych do analizy korelacji unikając powtarzalnych doświadczeń. Ostatecznie mając dwie kolumny do badania, program analizuje korelacje między nimi (patrz RYSUNEK 1 i RYSUNEK 2) i zwraca użytkownikowi informacje zwrotną. Dodatkowo program zapisuje wykresy wszystkich korelacji w pliku **corelation.pdf** (Patrz WYKRES)

## FORMAT:

DANE DLA GRUPY: *nazwa_grupy*

PARA: *parametr_1 – parametr_2 → Badana korelacja w obrębie grupy*

Korelacja: *informacja_zwrotna → Czy istnieje korelacja (na podstawie p-value). Jeśli korelacja nie istnieje (BRAK) to poniższe wiadomości się nie wyświetlą.*

Współczynnik korelacji: *x → Współczynnik korelacji*

Kierunek: *informacja_zwrotna → Informacja o kierunku korelacji*

Sila: *informacja_zwrotna → Informacja o sile korelacji*

**Przykładowa interpretacja będzie zaprezentowana w kontekście konkretnych testów.**

DANE DLA GRUPY: KONTROLA

PARA: wiek -- ERY

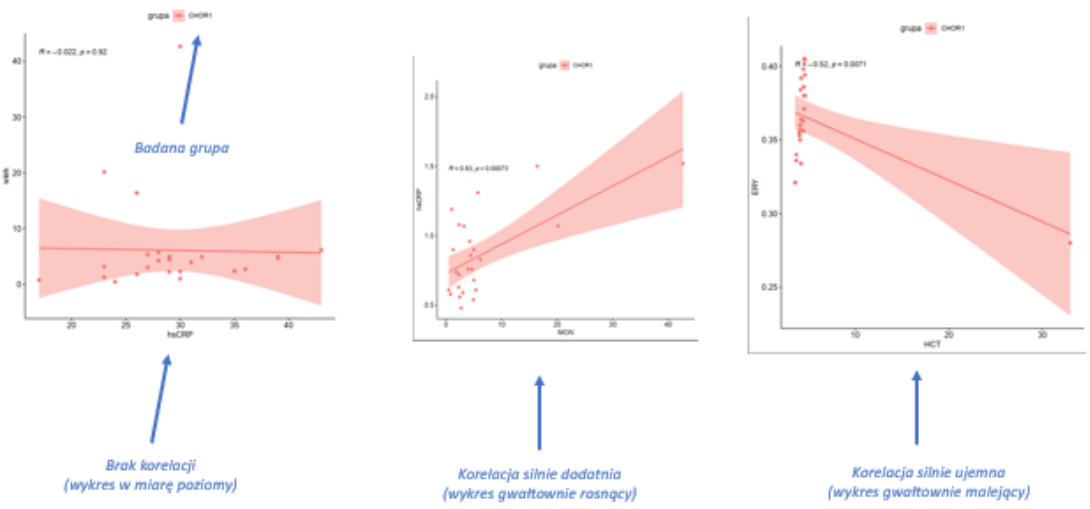
Korelacja: Istnieje! → *P-value z testu korelacji było mniejsze niż 0.05*

Współczynnik korelacji: 0.4609923

Kierunek: Korelacja dodatnia (gdy X rośnie to Y też rośnie) → *Współczynnik korelacji > 0*

Sila: Korelacja dodatnia o srednim natężeniu → *0.3 < Współczynnik korelacji < 0.5*

## WYKRES



NOTKA: Różnica w sile (między słabą  a średnią a silną korelacją ) objawia się poprzez zakrzywienie linii. W wypadku zaprezentowanym na przykładzie, słaba korelacja w porównaniu z silną będzie miała czerwoną linie bardziej skierowaną do dołu (dodatnia), albo skierowaną do góry (ujemna)

# KONIEC

Dziękujemy za wybranie najlepszego programu do analizy statystycznej 😎

*Program i sprawozdanie wykonał:* ***Jakub Łozowski***

*Numer indeksu:* ***147901***

***Bioinformatyka semestr IV***

***Wydział Informatyki i Telekomunikacji PP***