

Mateusz Misz 155628

Narzędzie skrypt.R służy do przeprowadzenia podstawowej analizy statystycznej dla danych numerycznych.

MANUAL:

Usage:

[1]*./R.exe CMD BATCH --vanilla --args input_file output_folder skrypt.R

[file.Rout]

[2]*./Rscrip.exe skrypt.R input_file output_folder

input_file - ścieżka do pliku .csv z danymi do analizy.

output_folder – folder w którym zapisywane będą wyniki

file.Rout – plik do przekierowane zostanie wyjście z narzędzia. Domyślnie tworzy plik o nazwieskryptu z rozszerzeniem .Rout

Wymagania:

- kolumna z przydziałem do grupy musi być pierwszą w tabeli

- wymagane są przynajmniej 3 grupy i 2 zmienne numeryczne

- kolumna z grupami nie może być kolumną numeryczną.

Stosowane testy:

test Saphiro-Wilka, test Levene'a, test ANOVA, test Tukeya, test Kruskala-Willisa, test Dunna

Wyjście:

Na wyjście zostają wypisywane komunikaty o wartościach odstających, uzupełnionych brakach danych, przeprowadzonych testach oraz zapisanych plikach(nie wszystkich)

Działanie:

Tworzy pliki ze statystykami ogólnymi, statystykami dla grup, wynikami analizy porównawczej(na podstawie testów opartych o średnie) i testów korelacji.

Struktura folderu wyjściowego:

output_folder:

- analiza_porownawcza_miedzy_grupami

- statystyki_ogolne

- statystyki_z_podzialem_na_grupy

- wyniki_testow_korelacji

OPIS DZIAŁANIA:

Narzędzie zaczyna od uzupełnienia brakujących danych numerycznych. W komórki z brakującymi danymi wpisuje wartość średniej danej zmiennej w danej grupie. Dane te zapisywane są bezpośrednio w folderze podanym jako wyjściowy.

Po uzupełnieniu braków narzędzie przechodzi do stworzenia ogólnych statystyk, które umieszcza w folderze „statystyki_ogolne” w postaci wykresów oraz plików .csv. Jeden plik .csv zawiera informacje o zakresie, średniej, medianie, IQR,

wariancji i odchyłaniu standardowym dla każdej z kolumn które zawierają dane numeryczne. Drugi plik zawiera informacje o danych odstających (dane te są również wypisywane na terminal/w pliku .Rout). Każdy z wykresów znajduje się w osobnym pliku nazwanym odpowiednio po nazwie kolumny w pliku wejściowym.

Następnie, narzędzie tworzy statystyki z podziałem na grupy. Robi to w folderze "statystyki_z _podziałem_na_grupy", analogicznie do tworzenia statystyk ogólnych, pomijając wypisywanie wartości odstających oraz stworzenie pliku z uzupełnionymi danymi wejściowymi.

Następnym etapem jest analiza porównawcza grup. Przeprowadzana jest na podstawie testów opierających się o średnie. Dla każdej zmiennej numerycznej sprawdza rozkład w grupach i jednorodność wariancji między grupami na podstawie p-value z testu Shapiro-Wilka oraz testu Levene'a. Jeśli zmienna posiada obie wspomniane cechy, przeprowadzany jest test Anova. Jeśli wartość p-value w tym teście wyniesie poniżej 0.05, różnice między grupami uważane są za istotne statystycznie i przeprowadzany jest test post hoc Tukeya, z którego uzyskiwane są dane o różnicach między grupami.

Dla zmiennych o rozkładzie innym niż normalny lub z niejednorodną wariancją w grupach, przeprowadzany jest test Kruskala-Wallisa. Jeśli p-value w tym teście wyniesie poniżej 0.05, różnice między grupami uważane są za istotne statystycznie i przeprowadzany jest test post hoc Dunna, z którego uzyskiwane są dane o różnicach między grupami.

Wykresy gęstości i boxploty dla każdej zmiennej z podziałem na grupy oraz pliki .txt z danymi o różnicach między grupami (po jednym pliku na każdą zmienną) znajdują się w folderze „analiza_porownawcza_miedzy_grupami”. Wyniki nieistotne statystycznie nie są nigdzie zapisywane.

Ostatnim etapem analizy jest analiza korelacji. Jest ona przeprowadzana dla każdej grupy i każdej pary zmiennych numerycznych, wykorzystując Test Pearsona (dla danych zgodnych z rozkładem normalnym) lub Test Spearmana (dla danych niezgodnych z rozkładem normalnym). Zgodność z rozkładem normalnym jest tutaj również badana przy pomocy testu Saphiro-Wilka. Dla obu testów, jeśli wartość p-value wyniesie poniżej 0.05, zapisywane są jego wyniki. Wyniki w postaci pliku .csv oraz wykresów (scatterplot) zapisywane są w folderze „wyniki_testow_korelacji”

ANALIZA WYNIKÓW:

Statystki ogólne i z podziałem na grupy:

Pierwsze dwa etapy analizy polegają na znalezieniu podstawowych parametrów.

Dane w pliku .csv z statystykami ogólnymi ułożone są wg schematu:

- w wierszach znajdują się wyniki dla poszczególnych zmiennych(nazwa zmiennej znajduje się w kolumnie „zmienna”

- nazwy kolumn określają jakiego parametru wartościami są dane wpisane w komórkach

W przypadku statystyk z podziałem na grupy plik.csv wygląda prawie identycznie. Posiada jedną dodatkową kolumnę określającą jakiej grupy dotyczy wiersz.

Wykresy boxplot (jedyne z tych dwóch etapów) odczytuje się następująco na pionowej osi znajdują się wartości zmiennej. Szary prostokąt wyznacza zakres w którym znajdują się 50% wszystkich wartości. Dolna krawędź prostokąta wyznacza pierwszy kwartył, a górna – 3. odległość między nimi to rozstęp międzykwartyłowy. Pozioma linia zaznaczona na prostokącie wyznacza medianę. Linie wychodzące z prostokąta wyznaczają większość z pozostałych wartości, a niepołączone z nimi punkty są wartościami odstającymi.

Analiza różnicująca:

Pliki .csv mogą być w dwóch formatach. Pierwszy z nich to wyniki z testu Dunna przeprowadzonego po teście kruskala-Wallis. Posiadają one kolumny „Comparison” z grupami, których porównania dotyczy wiersz, „Z” z wartością oznaczającą jak bardzo różnią się grupy, „P.unadj” z wartością p-value przed korekcją i „P.adj” z wartością p-value po korekcji z uwzględnieniem wielokrotnych powtórzeń. Aby różnica była uznana za istotną statystycznie – wartości p-value powinny wynosić poniżej 0.05.

Drugi format posiadają dane z wynikami z testu Tukeya wykonanym po ANOVA. Na początku pliku znajduje się informacja jakie prawdopodobieństwo ma przedział ufności oraz informacja o tym jakie argumenty zostały przekazane do testu(należy pomijać przy analizie). Poniżej znajduje się tabela z kolumnami zawierającymi kolejno: informacje dotyczącą porównania jakich grup dotyczą dane w wierszu, wartości różnicy między średnimi grup, dolnym i górnym ograniczeniu przedziału ufności 95% (zakres ten nie powinien obejmować zera aby różnica była istotna statystycznie) oraz wartość p-value po korekcji(powinna być mniejsza niż 0.05 aby różnica była uznana za istotną statystycznie).

W folderze dotyczącym analizy porównawczej znajdują się również boxploty dla każdej kolumny w której znaleziono istotne statystycznie różnice oraz wykresy gęstości(z podziałem na grupy) dla każdej zmiennej numerycznej, które umieszczone zostały w tym folderze ponieważ mogą służyć do graficznej oceny zgodności z rozkładem normalnym(jesli wykres przypomina krzywą gaussa, powinien zostać wykonany test Anova, w przeciwnym wypadku – test Kruskala-wallisa).

Znalezienie istotnej różnicy statystycznej może być podstawą do stwierdzenia że fakt rozróżniający grupy ma wpływ na dany parametr. Przykładowo, że konkretna choroba wpływa na obniżenie poziomu hemoglobiny we krwi

Analiza korelacji:

W folderze zawierającym wyniki testów korelacji znajduje się jeden plik .csv. Jego kolejne kolumny to:

Grupa – zawiera informacje o tym jakiej grupy dotyczą dane z danego wiersza
zmienna 1-|

|-wskazują których zmiennych dotyczą dane w wierszu
zmienna 2-|

Współczynnik korelacji – wartości wskazująca na to jak bardzo i w jaki sposób
zależne są od siebie zmienne (podpowiedź – spójrz na ostatnią kolumnę)

metoda – wskazuje na to jakim testem została zbadana korelacja (dla rozkładu
normalnego – pearson, dla innych -spearman)

p-value – wartość wskazująca na istotność statystyczną. Dane są istotne statystycznie
dla $p\text{-value} < 0.05$ i tylko takie dane znajdują się w pliku.

Ocena współczynnika korelacji – ocenia kierunek i siłę korelacji (dla korelacji
dodatniej, zmienne zmieniają się w tym samym kierunku, a dla ujemnej w
przeciwnych; im bliżej 1 lub -1 jest współczynnik korelacji, tym zmienne mocniej od
siebie zależą).

PRZYKŁADOWE DANE, WYNIKI I ANALIZA:

dane dla których zostanie pokazany przykładowy przebieg programu składają się z
kolumn: grupa;plec;wiek;hsCRP;ERY;PLT;HGB;HCT;MCHC;MON;LEU

są w nich 3 grupy i 9 zmiennych numerycznych

1. Uzupełnienie braków danych:

[1] "uzupełniono brakujące dane w rekordach: 13; 68"

na tym etapie program uzupełnił braki w danych i zapisał uzupełnione dane do
pliku.csv

2.statystyki ogólne i z podziałem na grupy:

[1] "statystyki ogólne:"

[1] "utworzono wykresy boxplot dla każdej zmiennej numerycznej"

zaraportowanie danych odstających:

[1] "dane odstające w kolumnie wiek :"

[1] 48

[1] "dane odstające w kolumnie hsCRP :"

[1] 20.1548 16.4069 42.6499 19.2124  Wartości danych odstających dla zmiennej hsCRP

na tym etapie program utworzył pliki .csv dla zmiennych numerycznych, dla
zmiennych nienumerycznych, z wartościami odstającymi i zapisał wykresy
plik z danymi odstającymi:

dane odstajace.csv - LibreOffice Calc

Plik Edycja Widok Wstaw Format Style Arkusz Dane Narzędzia Okno Pomoc

Liberation Sans 10 pt G K P A

| 1 | A | B | C | D | E | F | G | H | I | J | K | L |
|----|-------|---------|---|---|---|---|---|---|---|---|---|---|
| 2 | wiek | 48 | | | | | | | | | | |
| 3 | hsCRP | 20.1548 | | | | | | | | | | |
| 4 | hsCRP | 16.4069 | | | | | | | | | | |
| 5 | hsCRP | 42.6499 | | | | | | | | | | |
| 6 | hsCRP | 19.2124 | | | | | | | | | | |
| 7 | ERY | | | | | | | | | | | |
| 8 | PLT | 456 | | | | | | | | | | |
| 9 | PLT | 434 | | | | | | | | | | |
| 10 | HGB | 22.2318 | | | | | | | | | | |
| 11 | HCT | 0.0423 | | | | | | | | | | |
| 12 | MCHC | 38.8674 | | | | | | | | | | |
| 13 | MCHC | 38.203 | | | | | | | | | | |
| 14 | MCHC | 32.0573 | | | | | | | | | | |
| 15 | MCHC | 32.2234 | | | | | | | | | | |
| 16 | MON | 1.5 | | | | | | | | | | |
| 17 | MON | 1.52 | | | | | | | | | | |
| 18 | MON | 1.31 | | | | | | | | | | |
| 19 | MON | 0.14 | | | | | | | | | | |
| 20 | MON | 1.61 | | | | | | | | | | |
| 21 | LEU | | | | | | | | | | | |
| 22 | | | | | | | | | | | | |
| 23 | | | | | | | | | | | | |

Arkusz 1 z 1 Domyślny Polski Średnia: Suma: 0 120%

Wyszukaj

plik ze statystykami dla zmiennych numerycznych:

statystyki_ogolne_zmiennych_numerycznych.csv - LibreOffice Calc

Plik Edycja Widok Wstaw Format Style Arkusz Dane Narzędzia Okno Pomoc

Liberation Sans 10 pt G K P A

| 1 | A | B | C | D | E | F | G | H |
|----|----------|----------|---------|-------------------|---------|--------------------|---------------------|------------------------|
| 2 | parametr | min | max | srednia | mediana | IQR | wariancja | odchylenie standardowe |
| 3 | wiek | 17 | 48 | 30.64 | 30 | | 834.1794594594595 | 5.84632016395437 |
| 4 | hsCRP | 0.335089 | 42.6499 | 5.64473342666667 | 4.18349 | 4.95114 | 37.5524363507157 | 6.12800427143419 |
| 5 | ERY | 3.09 | 5.05 | 4.12853333333333 | 4.14 | 0.6000000000000001 | 0.190782954954955 | 0.436787081946061 |
| 6 | PLT | 91 | 456 | 220.093333333333 | 202 | | 714170.5181981982 | 64.5795493805756 |
| 7 | HGB | 9.5049 | 22.2318 | 12.1603829 | 12.2436 | 1.7788125 | 3.0854052111723 | 1.75653215489279 |
| 8 | HCT | 0.0423 | 0.412 | 0.349030666666667 | 0.354 | 0.053 | 0.00245388945225225 | 0.0495367485030281 |
| 9 | MCHC | 32.0573 | 38.8674 | 35.0278324 | 35.0471 | 1.3288 | 1.42490097278606 | 1.19369215997512 |
| 10 | MON | 0.14 | 1.61 | 0.767866666666667 | 0.73 | 0.28 | 0.0761548468468469 | 0.275961676409691 |
| 11 | LEU | 4.83 | 17.46 | 11.8066666666667 | 11.4 | 3.84 | 7.08538738738739 | 2.66183909870364 |
| 12 | | | | | | | | |
| 13 | | | | | | | | |
| 14 | | | | | | | | |
| 15 | | | | | | | | |
| 16 | | | | | | | | |
| 17 | | | | | | | | |
| 18 | | | | | | | | |
| 19 | | | | | | | | |
| 20 | | | | | | | | |

Arkusz 1 z 1 Domyślny Polski Średnia: Suma: 0 140%

Wyszukaj

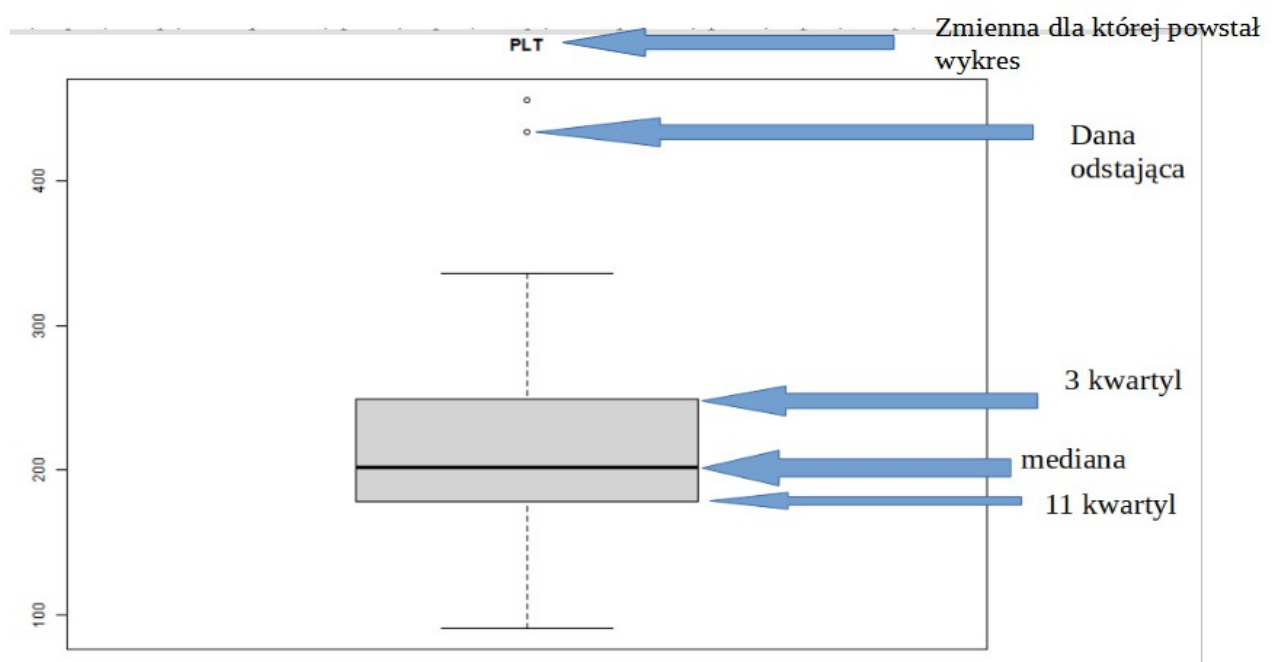
plik z danymi dla zmiennych nienumerycznych:

```
statystyki_ogolne_zmiennych_nienumerycznych.txt — Notatnik
Plik Edycja Format Widok Pomoc
zmienna: grupa

CHOR1    CHOR2    KONTROLA
25       25       25
zmienna: plec

k  m
40 35
```

przykładowy wykres:



[1] "utworzono pliki ze statystykami z podziałem na grupy"

na tym etapie utworzono wykresy oraz plik .csv

plik .csv

statystyki_grup.csv - LibreOffice Calc

Plik Edycja Widok Wstaw Format Style Arkusz Dane Narzędzia Okno Pomoc

LibreOffice Calc

10 pt G K P A B C D E F G H I J K L M

| 1 | A | B | C | D | E | F | G | H | I | J | K | L | M |
|----|----------|----------|----------|---------|------------|---------|-------------------|---------------------|------------------------|---|---|---|---|
| 2 | grupa | parametr | min | max | srednia | mediana | IQR | wariancja | odchylenie_standardowe | | | | |
| 3 | CHOR1 | wiek | 17 | 43 | 29.56 | 29 | | 6.34.59 | 5.88132638101305 | | | | |
| 4 | CHOR1 | hsCRP | 0.487607 | 42.6499 | 6.10302208 | 3.96646 | 2.67086 | 77.8741540567661 | 8.82463336670516 | | | | |
| 5 | CHOR1 | ERY | 3.3 | 4.62 | 4.1748 | 4.19 | 0.460000000000001 | 0.141667666666667 | 0.376387654774525 | | | | |
| 6 | CHOR1 | PLT | 128 | 336 | 225.28 | 217 | | 87.2940.12666666667 | 54.2229348769196 | | | | |
| 7 | CHOR1 | HGB | 9.5049 | 14.499 | 12.4114125 | 12.4047 | 1.2888 | 1.42197457359375 | 1.19246575363561 | | | | |
| 8 | CHOR1 | HCT | 0.28 | 0.405 | 0.36356 | 0.363 | 0.036 | 0.00094934 | 0.0308113615408343 | | | | |
| 9 | CHOR1 | MCHC | 32.5556 | 36.8742 | 35.1288212 | 35.0471 | 0.880330000000001 | 0.770013011819335 | 0.877503852880051 | | | | |
| 10 | CHOR1 | MON | 0.47 | 1.52 | 0.8424 | 0.76 | 0.46 | 0.0918356666666667 | 0.303044001205546 | | | | |
| 11 | CHOR1 | LEU | 6.79 | 16.81 | 12.022 | 11.66 | 4.37 | 6.66713333333333 | 2.58207926550161 | | | | |
| 12 | CHOR2 | wiek | 21 | 42 | 30.04 | 30 | | 8.34.79 | 5.89830484122345 | | | | |
| 13 | CHOR2 | hsCRP | 0.335089 | 19.2124 | 5.5360292 | 3.44546 | 6.53121 | 21.5814799641269 | 4.64558714955676 | | | | |
| 14 | CHOR2 | ERY | 3.25 | 5.04 | 4.1976 | 4.27 | 0.58 | 0.216569 | 0.465369745471276 | | | | |
| 15 | CHOR2 | PLT | 91 | 456 | 209.12 | 195 | | 51.5658.52666666667 | 75.2231790518499 | | | | |
| 16 | CHOR2 | HGB | 9.8271 | 22.2318 | 12.8061612 | 12.5658 | 1.9332 | 5.631239264896 | 2.3730232331113 | | | | |
| 17 | CHOR2 | HCT | 0.0423 | 0.412 | 0.345972 | 0.36 | 0.06 | 0.0053345279333333 | 0.0730378527431724 | | | | |
| 18 | CHOR2 | MCHC | 32.8878 | 38.8674 | 35.552044 | 35.5454 | 1.1627 | 1.66565257173334 | 1.29060163169482 | | | | |
| 19 | CHOR2 | MON | 0.14 | 1.61 | 0.7008 | 0.66 | 0.25 | 0.097282666666667 | 0.311901693914391 | | | | |
| 20 | CHOR2 | LEU | 7.95 | 16.59 | 12.0376 | | 122.64 | 5.37407733333333 | 2.31820562792288 | | | | |
| 21 | KONTROLA | wiek | 23 | 48 | 32.32 | 32 | | 6.31476666666667 | 5.61040699652589 | | | | |
| 22 | KONTROLA | hsCRP | 0.75844 | 14.3951 | 5.295149 | 4.22037 | 4.54994 | 15.9726547583583 | 3.9965803830723 | | | | |
| 23 | KONTROLA | ERY | 3.09 | 5.05 | 4.0132 | 3.98 | 0.51 | 0.208956 | 0.457117052843142 | | | | |
| 24 | KONTROLA | PLT | 147 | 434 | 225.88 | 214 | | 66.4072.11 | 63.8130864321731 | | | | |
| 25 | KONTROLA | HGB | 9.5049 | 13.2102 | 11.263575 | 11.4381 | 1.2888 | 1.122296101875 | 1.05938477517614 | | | | |
| 26 | KONTROLA | HCT | 0.279 | 0.389 | 0.33756 | 0.339 | 0.033 | 0.00091559 | 0.030258717520793 | | | | |
| 27 | KONTROLA | MCHC | 32.0573 | 36.0437 | 34.402632 | 34.5488 | 1.4949 | 1.25374566643334 | 1.11970784869685 | | | | |
| 28 | KONTROLA | MON | 0.35 | 1.25 | 0.7604 | 0.76 | 0.21 | 0.0351623333333333 | 0.187516221520522 | | | | |
| 29 | KONTROLA | LEU | 4.83 | 17.46 | 11.3604 | 10.68 | 4.37 | 9.49409566666667 | 3.08124904327233 | | | | |

statystyki grup

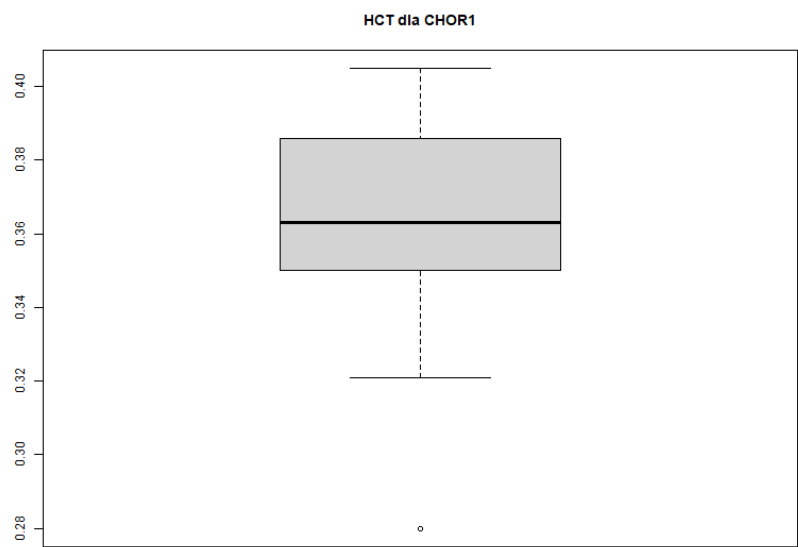
Arkusz 1 z 1

Domyslny Polski

Srednia : Suma: 0

18:34 31.05.2024

przykładowy wykres:



3. Analiza porównawcza:

[1] "analiza porownawcza:"

[1] "sprawdzenie zgodności z rozkładem normalnym i jednorodności wariancji:"

wyswietlanie wyników sprawdzenia rozkladu i jednorodności wariancji:

dla wiek p-value = $0.3897302 > 0.05$ - można założyć zgodność z rozkładem normalnym

[1] "ocena jednorodności wariancji dla: wiek"

jest jednorodność wariancji

dla wiek p-value = $0.3897302 > 0.05$ - można założyć zgodność z rozkładem normalnym

[1] "ocena jednorodności wariancji dla: wiek"

jest jednorodność wariancji

dla wiek p-value = $0.3897302 > 0.05$ - można założyć zgodność z rozkładem normalnym

[1] "ocena jednorodności wariancji dla: wiek"

jest jednorodność wariancji

komunikat o przeprowadzonym teście oraz jego wyniku

[1] "przeprowadzono test ANOVA"

brak istotnych statystycznie różnic między grupami

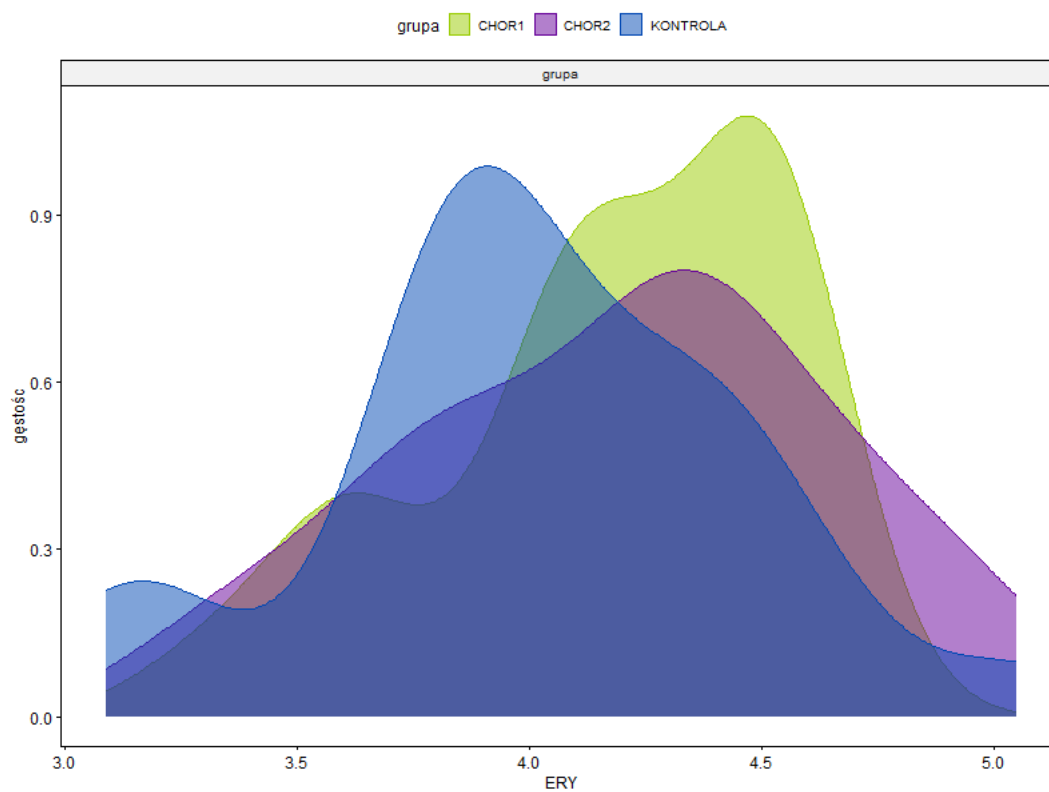
dla każdej zmiennej z istotnymi statystycznie różnicami tworzony jest wykres gęstości z podziałem na grupy oraz plik .csv z wynikami

plik .csv:

| | Comparison | Z | P.unadj | P.adj |
|---|------------------|-----------|-------------|------------|
| 1 | CHOR1 - CHOR2 | 0.7885475 | 0.430376497 | 0.43037650 |
| 2 | CHOR1 - KONTROLA | 2.7355785 | 0.006227073 | 0.01868122 |
| 3 | CHOR2 - KONTROLA | 1.9470310 | 0.051531020 | 0.10306204 |

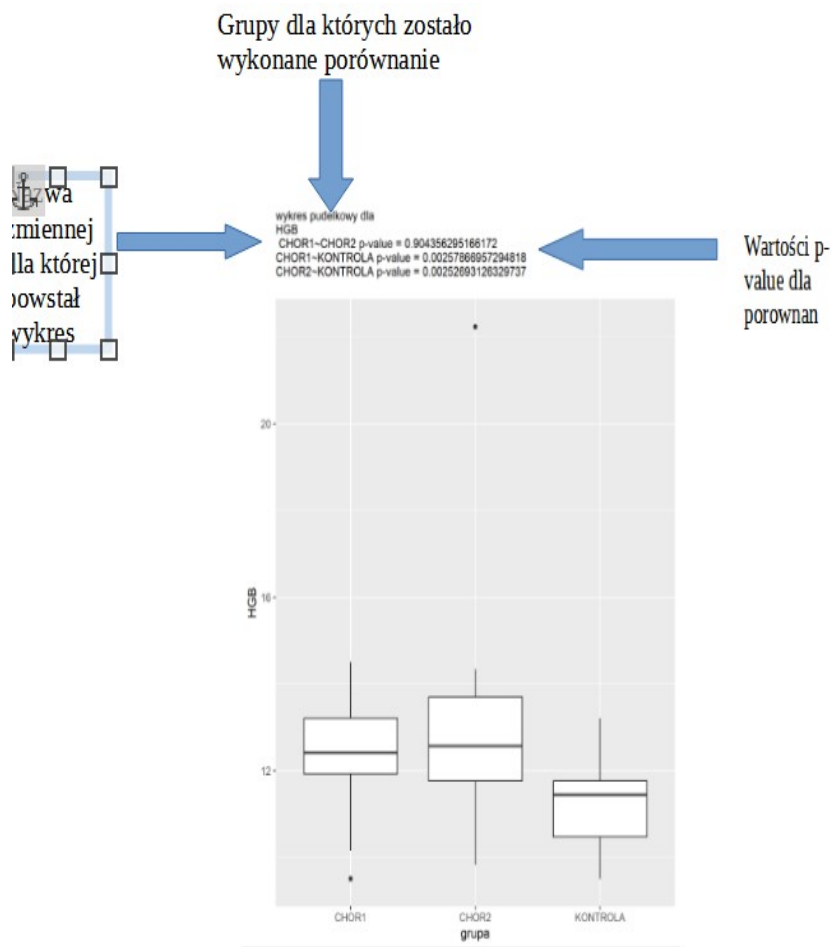


wykres:



krawędź obszaru na wykresie jest funkcją gęstości dla każdej z grup. Wyznacza ona prawdopodobieństwo wystąpienia danej wartości.

Boxplot:



4. Analiza zależności między parametrami

[1] "analiza zaleznosci miedzy zmiennymi"

komunikaty o rodzaju testu jaki przeprowadzono oraz o jego wyniku

[1] "w grupie CHOR1 dla parametrow wiek i hsCRP przeprowadzono test korelacji spearmana"

[1] "brak korelacji"

[1] "w grupie CHOR1 dla parametrow ERY i HGB przeprowadzono test korelacji spearmana"

[1] "istnieje korelacja; wspolczynnik = 0.87565198324265"

Saving 7 x 7 in image

-komunikat o zapisaniu wykresu

po ostatnim teście tworzony jest plik .csv z wynikami:

zaleznosci_miedzy_parametrami.csv - LibreOffice Calc

Plik Edycja Widok Wstaw Format Style Arkusz Dane Narzędzia Okno Pomoc

Liberation Sans 10 pt G K P A

| A | B | C | D | E | F | G | H | I | J | K |
|----|----------|----------|----------|------------------------|----------|----------------------|--|---|---|---|
| 1 | grupa | zmienna1 | zmienna2 | wspolczynnik_korelacji | metoda | p-value | ocena_wspolczynnika_korelacji | | | |
| 2 | CHOR1 | ERY | HGB | 0.87565198324265 | spearman | 9.9650337132721e-09 | bardzo_silna_korelacja_dodatnia | | | |
| 3 | CHOR1 | ERY | HCT | 0.846005790457171 | spearman | 1.00068880981556e-07 | bardzo_silna_korelacja_dodatnia | | | |
| 4 | CHOR1 | ERY | MCHC | 0.437501172147629 | spearman | 0.0287352537035762 | korelacja_dodatnia_o_srednim_natężeniu | | | |
| 5 | CHOR1 | HGB | MCHC | 0.953342396800469 | pearson | 1.87042804617836e-13 | bardzo_silna_korelacja_dodatnia | | | |
| 6 | CHOR1 | HGB | MCHC | 0.637305862411678 | pearson | 0.000611749123444521 | silna_korelacja_dodatnia | | | |
| 7 | CHOR1 | HCT | MCHC | 0.426816687940546 | pearson | 0.0333516912200337 | korelacja_dodatnia_o_srednim_natężeniu | | | |
| 8 | CHOR2 | ERY | HGB | 0.883104217853547 | spearman | 5.08515255680143e-09 | bardzo_silna_korelacja_dodatnia | | | |
| 9 | CHOR2 | ERY | HCT | 0.751445588857523 | spearman | 1.49386813661715e-05 | bardzo_silna_korelacja_dodatnia | | | |
| 10 | CHOR2 | PLT | MCHC | -0.519372468228172 | spearman | 0.00779894485526622 | silna_korelacja_ujemna | | | |
| 11 | CHOR2 | HGB | HCT | 0.736964573555573 | spearman | 2.64876689268217e-05 | bardzo_silna_korelacja_dodatnia | | | |
| 12 | CHOR2 | MCHC | MON | -0.46611253312818 | pearson | 0.0188501842843055 | korelacja_ujemna_o_srednim_natężeniu | | | |
| 13 | KONTROLA | wiek | ERY | 0.4609922821345 | pearson | 0.0203786404708158 | korelacja_dodatnia_o_srednim_natężeniu | | | |
| 14 | KONTROLA | wiek | LEU | -0.496139853126458 | pearson | 0.0116545940192987 | korelacja_ujemna_o_srednim_natężeniu | | | |
| 15 | KONTROLA | hsCRP | LEU | 0.425384615384615 | spearman | 0.0350910111033194 | korelacja_dodatnia_o_srednim_natężeniu | | | |
| 16 | KONTROLA | ERY | PLT | 0.454282972805216 | spearman | 0.0225324670621139 | korelacja_dodatnia_o_srednim_natężeniu | | | |
| 17 | KONTROLA | ERY | HGB | 0.587117362396295 | pearson | 0.00203256176231683 | silna_korelacja_dodatnia | | | |
| 18 | KONTROLA | ERY | HCT | 0.775705406532287 | pearson | 5.22360102250932e-06 | bardzo_silna_korelacja_dodatnia | | | |
| 19 | KONTROLA | ERY | MCHC | -0.494071763114614 | pearson | 0.0120632187287476 | korelacja_ujemna_o_srednim_natężeniu | | | |
| 20 | KONTROLA | HGB | HCT | 0.941801266281483 | pearson | 2.24424482992436e-12 | bardzo_silna_korelacja_dodatnia | | | |
| 21 | KONTROLA | MON | LEU | 0.519823280076833 | pearson | 0.0077362714043327 | silna_korelacja_dodatnia | | | |
| 22 | | | | | | | | | | |
| 23 | | | | | | | | | | |
| 24 | | | | | | | | | | |
| 25 | | | | | | | | | | |
| 26 | | | | | | | | | | |
| 27 | | | | | | | | | | |
| 28 | | | | | | | | | | |

zaleznosci_miedzy_parametrami

Arkusz 1 z 1

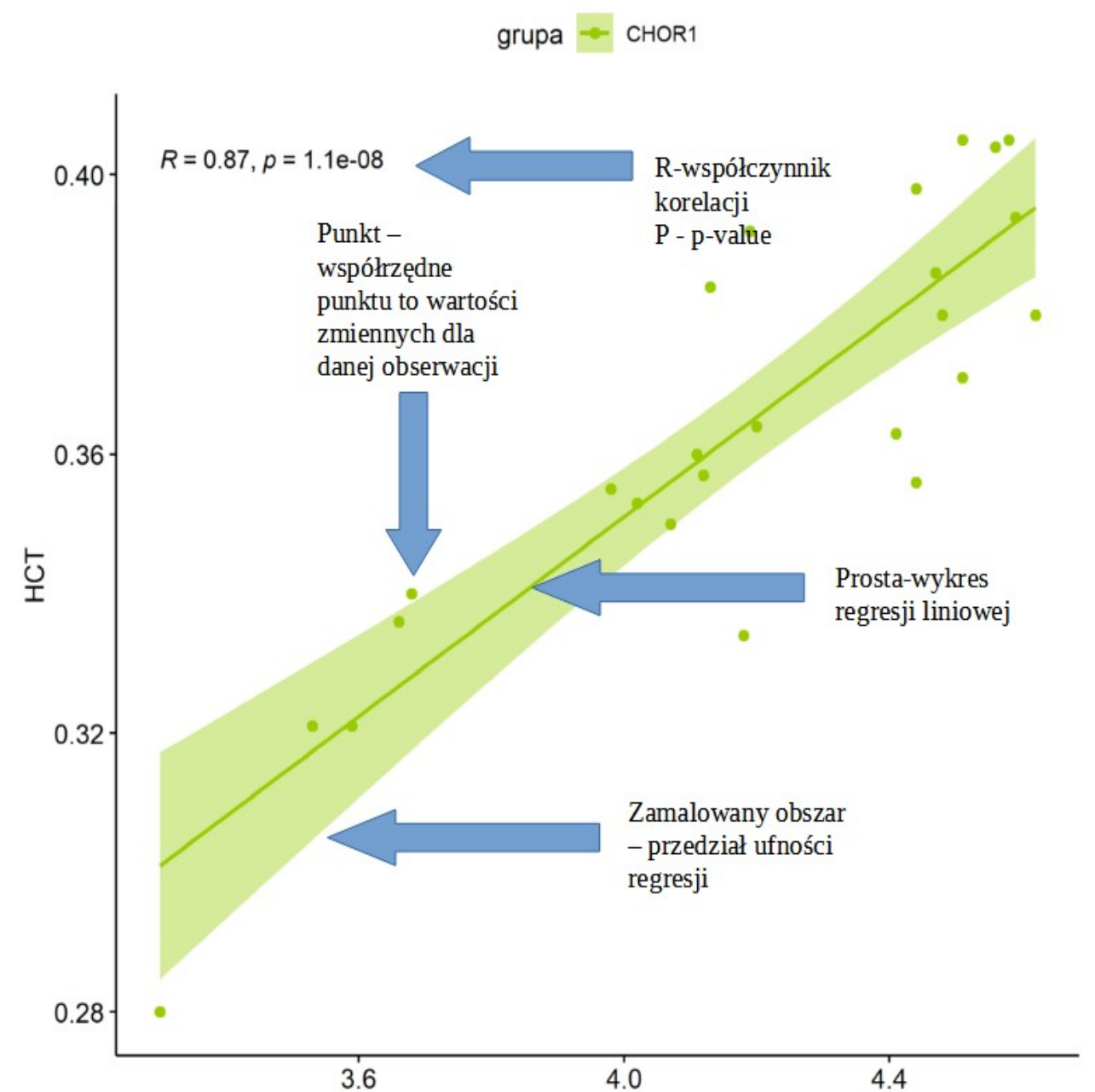
Domyślny Polski Średnia: Suma: 0 19:26 31.05.2024

w pliku znajdują się tylko dane istotne statystycznie(p-value dla wszystkich jest poniżej 0.05)

zasada oceniania 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

przykładowy wykres scatterplot:



przedział ufności dla regresji – jest to obszar na wykresie w którym może znajdować się prosta prawdziwej rzeczywistej regresji liniowej, zależy od odległości w jakiej może znajdować się rzeczywisty współczynnik korelacji względem wyliczonego. Im mniejsza pewność że współczynnik obliczony jest równy rzeczywistemu, tym większy obszar na wykresie.