

Rachunek Prawdopodobieństwa i Statystyka

Projekt zaliczeniowy – Adam Łangowski 147896

Temat:

Analiza statystyczna z wykorzystaniem języka R dla danych medycznych, w której badane grupy są niezależne – badamy te same parametry u różnych grup chorych (rozdzielonych chociażby na podstawie stadium choroby) jak i dla grupy kontrolnej – zdrowej.

Przykładowe dane (analizowany przez skrypt plik z danymi (przykladoweDane-Projekt.csv) można rozszerzać o kolejne grupy i dane z nimi powiązane) :

	grupa	plec	wiek	hsCRP	ERY	PLT	HGB	HCT	MCHC	MON	LEU
1	CHOR1	k	36	2,711	4,19	201	13,2102	0,392	34,7149	0,48	11,86
2	CHOR1	m	39	4,69938	4,48	222	13,0491	0,38	35,3793	0,76	10,32
3	CHOR1	k	35	2,35354	3,59	278	10,1493	0,321	32,5556	1,06	13,6
4	CHOR1	m	29	2,27161	3,66	200	11,277	0,336	34,5488	0,63	10,11
5	CHOR1	m	29	4,46519	4,41	128	12,4047	0,363	35,2132		10,55
6	CHOR1	m	43	6,16269	3,68	176	11,4381	0,34	34,7149	0,83	9,28

Istotnym elementem jest nazwa grupy występująca koniecznie w pierwszej kolumnie. W kolejnych określamy płeć oraz wiek pacjenta. Pozostałe kolumny reprezentują wyniki morfologiczne – są danymi ilościowymi ciągłymi (mogą przyjmować dowolne wartości liczbowe z określanego przez badanie krwi przedziału).

Problem badawczy:

Określenie czy między scharakteryzowanymi grupami występują istotne statystycznie różnice; ocena zależności między konkretnymi parametrami.

Omówienie elementów przeprowadzonej analizy:

- **Weryfikacja i ocena danych wejściowych:**

- Przeanalizowane zostały 3 grupy – *Chorzy_1*, *Chorzy_2* oraz *Grupa Kontrolna*. Każda z nich zawiera po 25 rekordów – pacjentów.

	grupa	liczność_grupy	średnia_wieku
1	CHOR1	25	29.56
2	CHOR2	25	30.04
3	KONTROLA	25	32.32

- **Brakujące dane** (zidentyfikowane wartości *Not Available (NA)*):

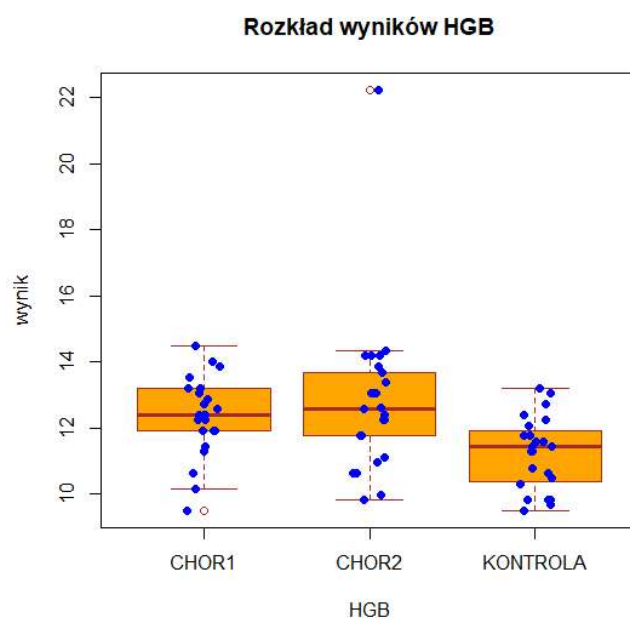
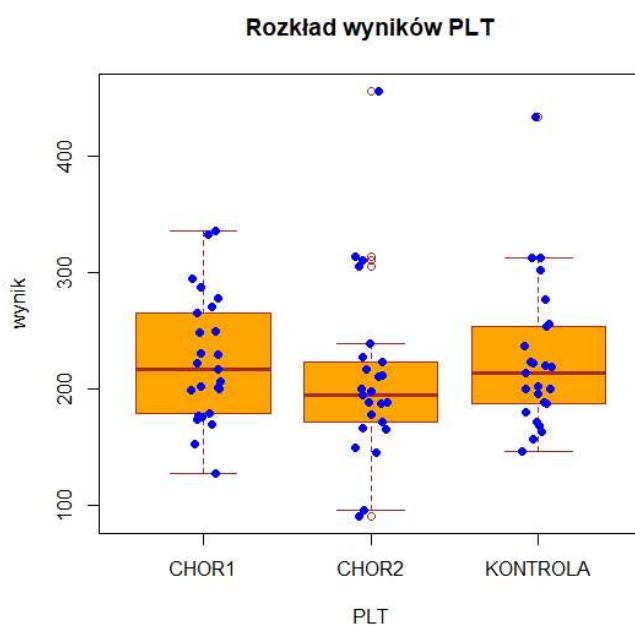
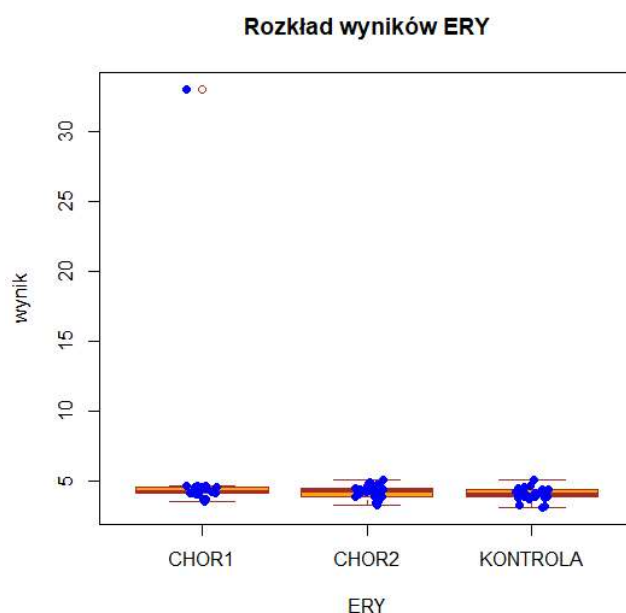
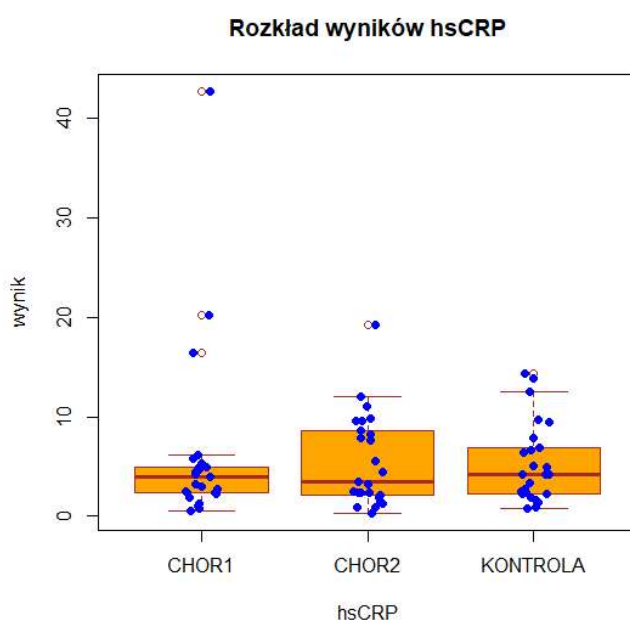
Dla grupy pierwszej (Chorzy_1):	Dla grupy drugiej (Chorzy_2):	Dla grupy trzeciej (kontrolnej):
Brak wyniku HGP dla indeksu: 13, Brak wyniku MON dla indeksu: 5	Dane są kompletne	Brak wyniku HGP dla indeksu: 18* *indeks '18' licząc dla grupy 3, a '68' dla całości danych.

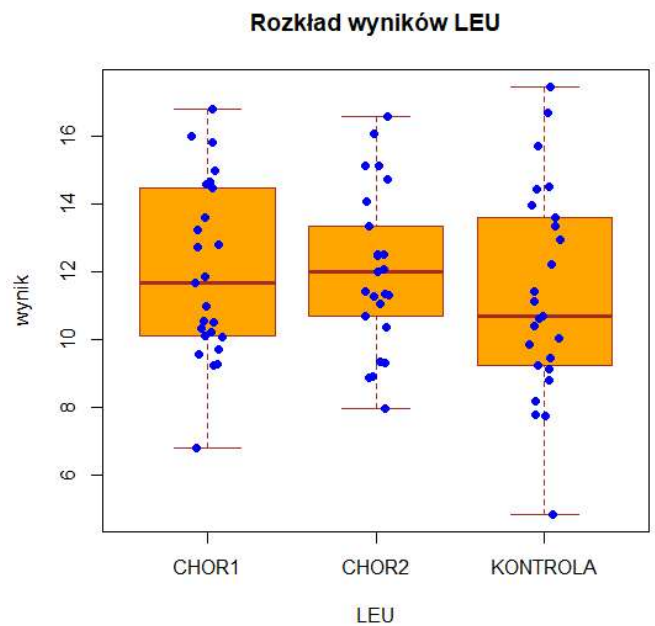
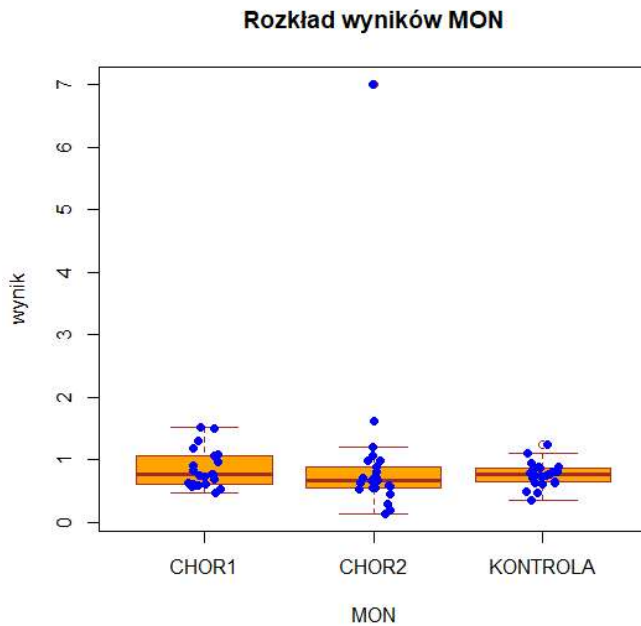
- W celu zachowania jak największej rzetelności wyników późniejszych analiz **wartości puste** zostały zastąpione przez **średnie danych parametrów dla grupy**.

Dla grupy pierwszej (Chorzy_1):	Dla grupy trzeciej (kontrolnej):
HGP indeks 13: NA -> 12.41141 MON indeks 5: NA -> 0.8579167	HGP indeks 18: NA -> 11.26357

- **Elementy statystyki opisowej:**

- **Identyfikacja wartości odstających od normy** – prezentacja przy pomocy wykresów typu boxplot wszystkich parametrów dla badanych grup:





Wykres pudełkowy w bardzo łatwy sposób pozwala ilustrować duże ilości danych, przejrzystość obrazując wartości odstające. Długość pudełka równa się rozstępowi ćwiartkowemu (czyli różnicy między kwartylem trzecim i pierwszym; pozioma kreska w pudełku to mediana. Długość wychodzących z pudełka pionowych linii, nazywanych ‘wąsami’ i wyliczana jest ona następująco

$$\text{Min} = Q1 - 1.5 \cdot \text{IQR} \quad \text{(dolna granica)}$$

$$\text{Max} = Q3 + 1.5 \cdot \text{IQR} \quad \text{(górna granica)}$$

Pojedyncze wartości paramentów w kilku przypadkach mocno odstają od normy, jednak z uwagi na wystarczająco dużą ilość rekordów w grupie, a także fakt, iż takie wyniki w większości mogą faktycznie mieć miejsce w rzeczywistości zdecydowałem się pozostawić wszystkie rekordy bez zmian.

➤ Poniższe tabele prezentują **wyniki analizy statystyki opisowej dla każdej z grup.**

Elementy analizy parametrów:

- Minimum – wartość minimalna w grupie
- Maximum – wartość maksymalna w grupie
- Średnia – średnia wartość w grupie
- Mediana – wartość środkowa zbioru wartości w grupie
- Odchylenie_std – odchylenie standardowe określa o ile wszystkie wartości badanego parametru różnią się od średniej arytmetycznej
- IQR – rozstęp międzykwartylowy - różnica między trzecim a pierwszym kwartylem; pomiędzy tymi kwartylami znajduje się z definicji 50% obserwacji, dlatego im większa szerokość rozstępu ćwiartkowego, tym większe zróżnicowanie cechy.
- Wariancja - średnia arytmetyczna kwadratów odchyłeń poszczególnych wartości cechy od wartości oczekiwanej
- Kurtoza - miara koncentracji (spłaszczenia), dla rozkładu normalnego kurtoza równa jest 0, dla rozkładu wysmukłego kurtoza przyjmuje wartości większe od 0, dla rozkładu spłaszczonego kurtoza przyjmuje wartości mniejsze od 0

Parametry:

1) hsCRP:

grupa	minimum	maximum	średnia	mediana	odchylenie_std	iqr	wariancja	kurtoza
1 CHOR1	0.467607	42.6499	6.103022	3.96646	8.824633	2.67086	77.87415	9.128154
2 CHOR2	0.335089	19.2124	5.536029	3.44546	4.645587	6.53121	21.58148	0.584128
3 KONTROLA	0.758440	14.3951	5.295149	4.22037	3.996580	4.54994	15.97265	-0.305736

2) ERY:

grupa	minimum	maximum	średnia	mediana	odchylenie_std	iqr	wariancja	kurtoza
1 CHOR1	3.53	33.00	5.3628	4.20	5.7671603	0.44	33.260138	18.097806
2 CHOR2	3.25	5.04	4.1976	4.27	0.4653697	0.58	0.216569	-0.846329
3 KONTROLA	3.09	5.05	4.0132	3.98	0.4571171	0.51	0.208956	-0.118212

3) PLT:

grupa	minimum	maximum	średnia	mediana	odchylenie_std	iqr	wariancja	kurtoza
1 CHOR1	128	336	225.28	217	54.22293	87	2940.127	-0.731565
2 CHOR2	91	456	209.12	195	75.22318	51	5658.527	2.535503
3 KONTROLA	147	434	225.88	214	63.81309	66	4072.110	2.160537

4) HGB:

grupa	minimum	maximum	średnia	mediana	odchylenie_std	iqr	wariancja	kurtoza
1 CHOR1	9.5049	14.4990	12.41141	12.4047	1.192466	1.2888	1.421975	-0.087965
2 CHOR2	9.8271	22.2318	12.80616	12.5658	2.373023	1.9332	5.631239	7.242734
3 KONTROLA	9.5049	13.2102	11.26357	11.4381	1.059385	1.2888	1.122296	-1.011412

5) HCT

grupa	minimum	maximum	średnia	mediana	odchylenie_std	iqr	wariancja	kurtoza
1 CHOR1	0.2800	0.405	0.363560	0.363	0.03081136	0.036	0.000949340	0.0663792
2 CHOR2	0.0423	0.412	0.345972	0.360	0.07303785	0.060	0.005334528	9.0837963
3 KONTROLA	0.2790	0.389	0.337560	0.339	0.03025872	0.033	0.000915590	-0.6761557

6) MCHC:

grupa	minimum	maximum	średnia	mediana	odchylenie_std	iqr	wariancja	kurtoza
1 CHOR1	32.5556	36.8742	35.12882	35.0471	0.8775039	0.88033	0.770013	1.124156
2 CHOR2	32.8878	38.8674	35.55204	35.5454	1.2906016	1.16270	1.665653	0.503720
3 KONTROLA	32.0573	36.0437	34.40263	34.5488	1.1197078	1.49490	1.253746	-0.773422

7) MON:

	grupa	minimum	maximum	średnia	mediana	odchylenie_std	iqr	wariancja	kurtoza
1	CHOR1	0.48	1.52	0.8579167	0.76	0.2929445	0.46	0.08581649	-0.36979
2	CHOR2	0.14	7.00	0.9528000	0.66	1.2978685	0.33	1.68446267	15.87148
3	KONTROLA	0.35	1.25	0.7604000	0.76	0.1875162	0.21	0.03516233	0.68300

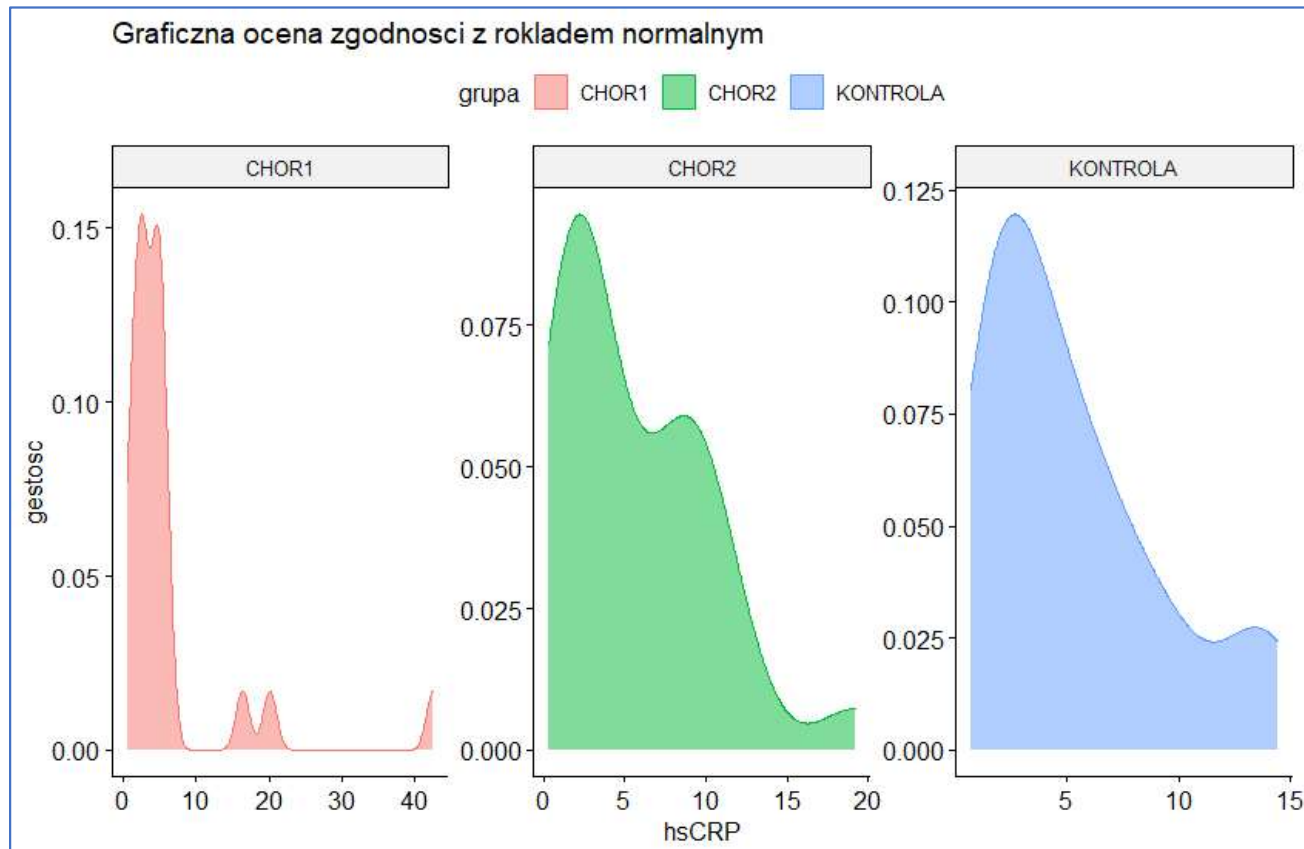
8) LEU:

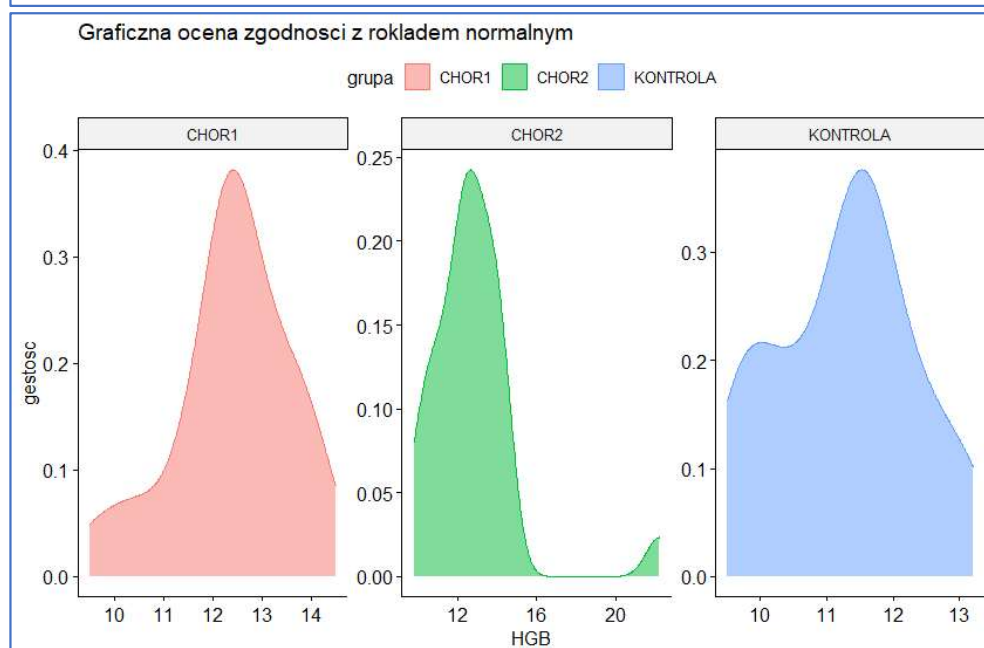
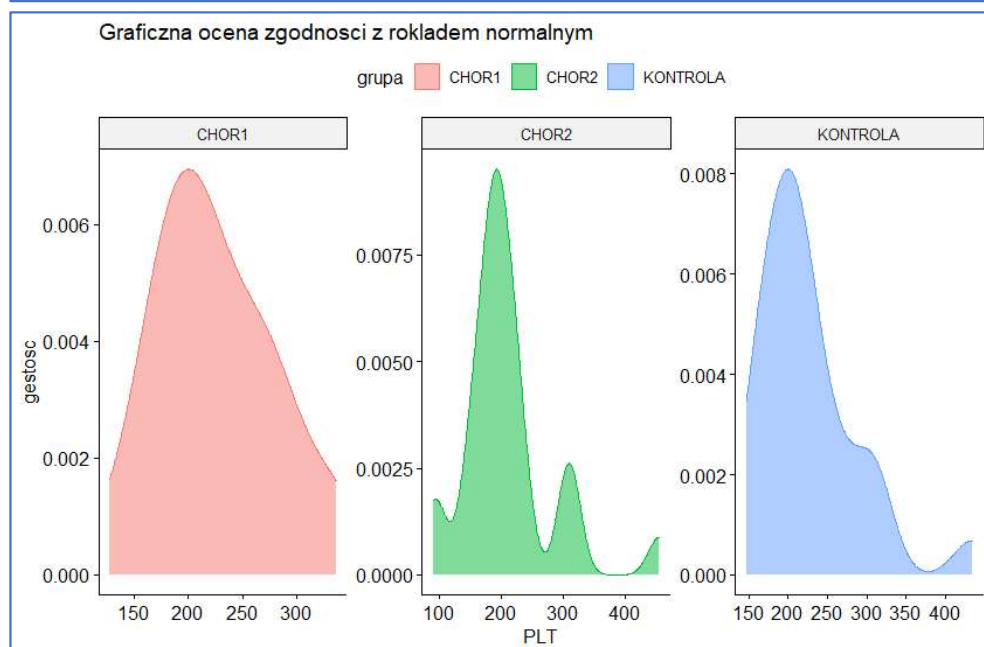
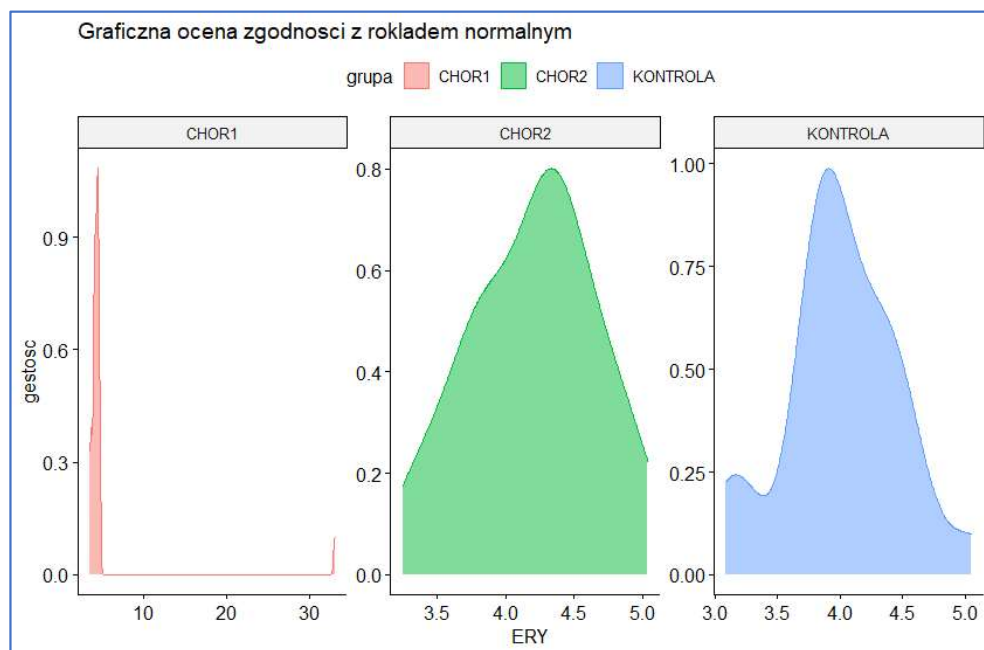
	grupa	minimum	maximum	średnia	mediana	odchylenie_std	iqr	wariancja	kurtoza
1	CHOR1	6.79	16.81	12.0220	11.66	2.582079	4.37	6.667133	-1.02893
2	CHOR2	7.95	16.59	12.0376	12.00	2.318206	2.64	5.374077	-0.85364
3	KONTROLA	4.83	17.46	11.3604	10.68	3.081249	4.37	9.494096	-0.68624

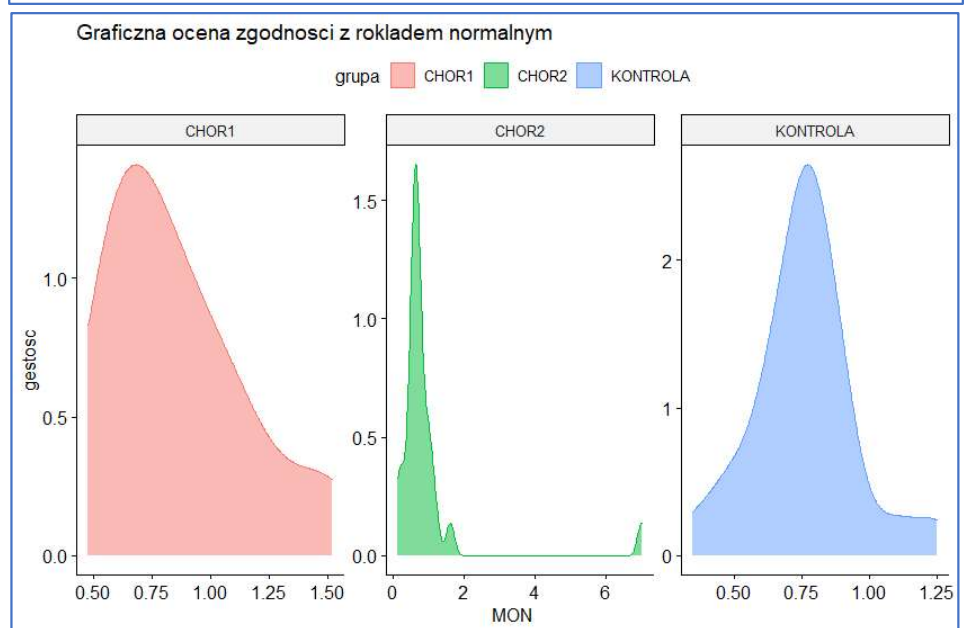
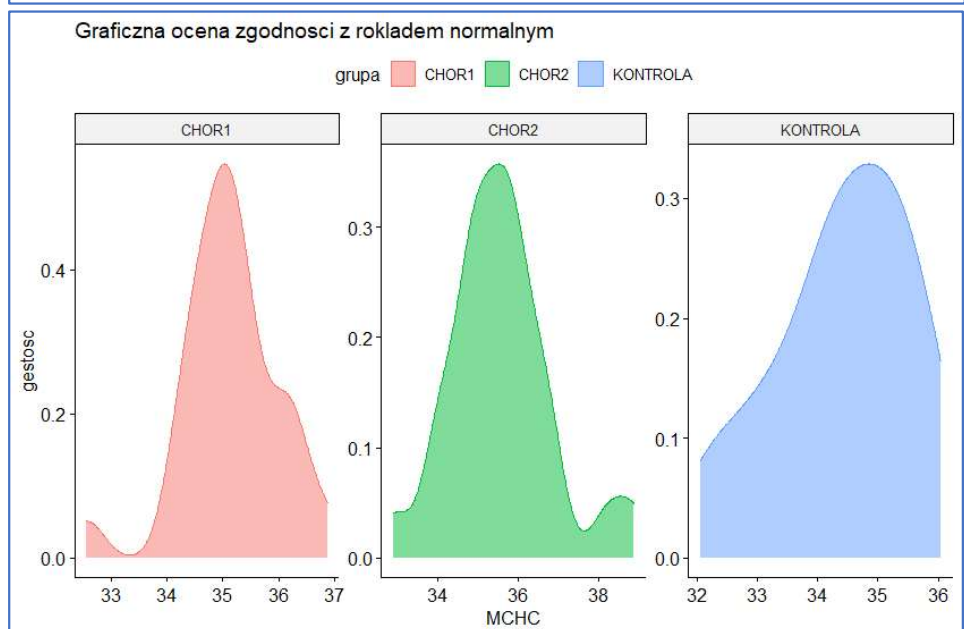
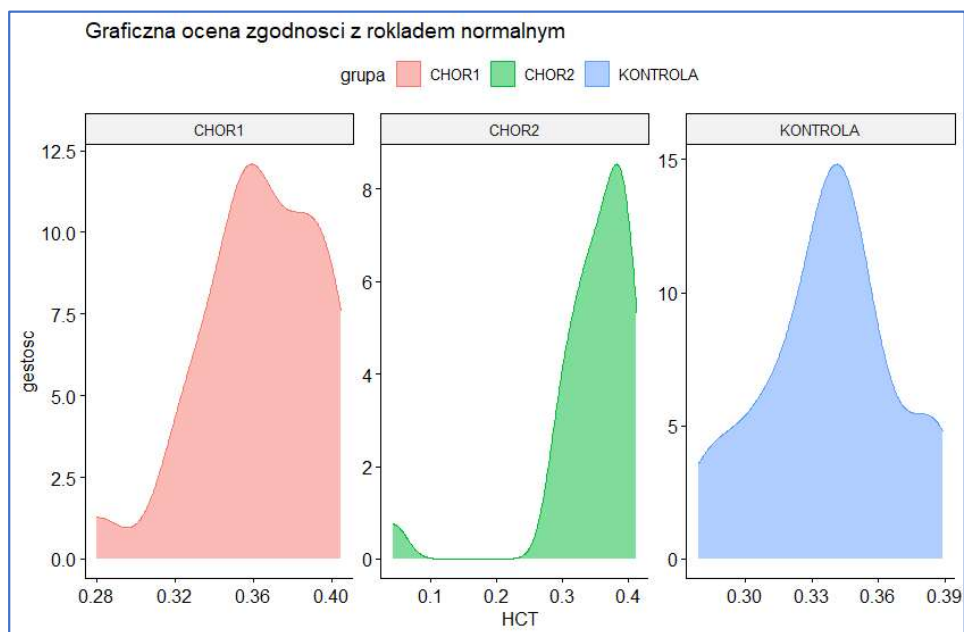
• Analiza porównawcza między grupami (określenie czy istnieją istotne statystycznie różnice)

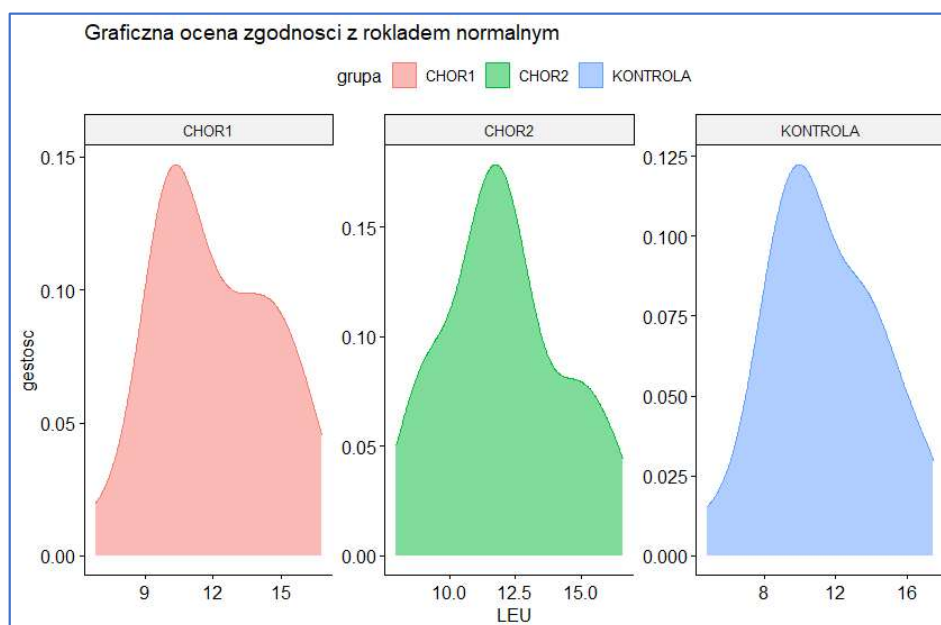
- Ocena zgodności danych z rozkładem normalnym dla każdego analizowanego parametru
Do testów zastosowano test Shapiro-Wilka.

Wykonano także graficzną ocenę zgodności z rozkładem normalnym (wygenerowane wykresy gęstości zapewniają wizualną ocenę tego, czy rozkład ma kształt dzwonu i ma zgodność z rozkładem normalnym):









Jeśli wartość $p\text{-value} > 0.05$ oznacza to, że rozkład danych nie różni się znacząco od rozkładu normalnego (możemy założyć normalność danych).

W analizowanych danych jedynie dla parametrów MCHC oraz LEU ma miejsce zgodność z rozkładem normalnym dla każdej z grup.

Przykład:

TESTY:

Dla parametru LEU:

Wartosci rozkladu normalnego:

0.3310257 0.6152727 0.8771294 $\leftarrow p\text{-value}$

- Ocena homogeniczności wariancji (inaczej jednorodność wariancji) – miara potrzebna w przypadku niektórych późniejszych testów statystycznych. Do testów zastosowano test Levene’a.

Jeśli wartość $p\text{-value} > 0.05$ oznacza to, że dane są zgodne z założeniem o jednorodności wariancji (możemy założyć homogeniczność danych).

Przykład (dla parametru LEU):

Występuje jednorodność wariancji

0.5965009 $\leftarrow p\text{-value}$

- Program na podstawie zidentyfikowanej na początku liczby grup niezależnych wybiera odpowiednie testy. Testy do porównywania grup pozwalają określić czy pomiędzy badanymi grupami występują istotne statystycznie różnice. Na wybór testu mają także wpływ: ocena zgodności danych z rozkładem normalnym oraz jednorodność wariancji, które zostały omówione powyżej.

Wszystkie możliwe wybory odpowiednich testów prezentuje poniższa tabela:

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 Wilcozona (Manna-Whitneya)
>2	TAK	TAK	test ANOVA (<i>post hoc</i> Tukeya)
		NIE	test Kruskala-Wallisa (<i>post hoc</i> Dunna)
	NIE	-	

➤ **Omówienie poszczególnych testów dla więcej niż dwóch grup:**

A. test Kruskala-Wallisa – wykonywany jeśli dane nie spełniają założenia o zgodności z rozkładem normalnym ($p\text{-value} < 0.05$) lub gdy są zgodne z rozkładem normalnym, ale nie spełniają założenia o jednorodności wariancji.

Gdy wartość $p\text{-value}$ jest mniejsza niż poziom istotności 0.05 możemy stwierdzić, że istnieją znaczące różnice między grupami.

Przykład:

```
[1] "Dla parametru hscRP"
[1] "wartosci rozkladu normalnego"
[1] 9.933661e-08 4.998515e-03 7.351605e-03
[1] "Grupy nie sa zgodne z rozkladem normalnym -> TEST KRUSKALA"
[1] "Test Kruskala:"

kruskal-wallis rank sum test

data: dane[, numer] by grupa
kruskal-wallis chi-squared = 0.25403, df = 2, p-value = 0.8807
[1] 0.8807212
```

```
[1] "Test KRUSKALA wykrył, że nie ma różnic między grupami"
```

B. Test Anova - wykonywany jeśli dane spełniają założenia o zgodności z rozkładem normalnym ($p\text{-value} > 0.05$) oraz spełniają założenia o jednorodności wariancji.

Gdy wartość $p\text{-value}$ jest mniejsza niż poziom istotności 0.05 możemy stwierdzić, że istnieją znaczące różnice między grupami.

Przykład:

```
[1] "Dla parametru LEU"
[1] "wartosci rozkladu normalnego"
[1] 0.3310257 0.6152727 0.8771294
[1] "Grupy sa zgodne z rozkladem normalnym"
[1] "wystepuje jednorodnosc wariancji -> TEST ANOVA"
[1] 0.5965009

[1] "Test Anova wykrył, że nie ma różnic między tymi grupami"
```

Testy post hoc - Jeśli występują istotne statystycznie różnice pomiędzy grupami, określamy pomiędzy którymi grupami występują te różnice i jak są znaczące. Wykorzystujemy do tego testy post hoc.

- **Testy post hoc dla więcej niż dwóch grup:**

a. Test Dunna – wykonywany jeśli test Kruskala-Wallisa wskazał na istnienie istotnie statystycznych różnic między grupami

```
[1] "Test KRUSKALA wykrył, że sa roznice miedzy grupami -> TEST DUNNA"
[1] "Test DUNNA:"
      Comparison      Z      P.unadj      P.adj
1  CHOR1 - CHOR2 -0.1201601 0.9043562952 0.904356295
2 CHOR1 - KONTROLA 3.2183431 0.0012893348 0.002578670
3 CHOR2 - KONTROLA 3.3385033 0.0008423104 0.002526931
[1] "Roznica miedzy CHOR1 - CHOR2 w parametrze ' Z ' jest istotna statycznie"
[1] "Roznica miedzy CHOR1 - KONTROLA w parametrze ' P.unadj ' jest istotna statycznie"
[1] "Roznica miedzy CHOR2 - KONTROLA w parametrze ' P.unadj ' jest istotna statycznie"
[1] "Roznica miedzy CHOR1 - KONTROLA w parametrze ' P.adj ' jest istotna statycznie"
[1] "Roznica miedzy CHOR2 - KONTROLA w parametrze ' P.adj ' jest istotna statycznie"
```

b. Test Tukeya - wykonywany jeśli test Anova wskazał na istnienie istotnie statystycznych różnic między grupami

```
[1] "Test ANOVA wykrył, że sa roznice miedzy tymi grupami -> TEST TUKEY"
[1] "Test TUKEY:"
      diff      lwr      upr      p adj
CHOR2-CHOR1 0.4232228 -0.3274109 1.17385653 0.372940393
KONTROLA-CHOR1 -0.7261892 -1.4768229 0.02444453 0.060043259
KONTROLA-CHOR2 -1.1494120 -1.9000457 -0.39877827 0.001352322
[1] "Sa istotne roznice statyczne w parametrze diff w wierszu 2"
[1] "Sa istotne roznice statyczne w parametrze diff w wierszu 3"
[1] "Sa istotne roznice statyczne w parametrze lwr w wierszu 1"
[1] "Sa istotne roznice statyczne w parametrze lwr w wierszu 2"
[1] "Sa istotne roznice statyczne w parametrze lwr w wierszu 3"
[1] "Sa istotne roznice statyczne w parametrze upr w wierszu 2"
[1] "Sa istotne roznice statyczne w parametrze upr w wierszu 3"
[1] "Sa istotne roznice statyczne w parametrze p adj w wierszu 3"
```

➤ **Zidentyfikowane różnice między grupami:**

1)

```
[1] "Dla parametru HGB"
[1] "wartosci rozkladu normalnego"
[1] 5.073768e-01 3.885087e-05 3.944054e-01
[1] "Grupy nie sa zgodne z rozkladem normalnym -> TEST KRUSKALA"
[1] "Test kruskala:"

      kruskal-wallis rank sum test

data: dane[, numer] by grupa
kruskal-wallis chi-squared = 14.345, df = 2, p-value = 0.0007673

[1] 0.0007673315

      kruskal-wallis rank sum test

data: dane[, numer] by grupa
kruskal-wallis chi-squared = 14.345, df = 2, p-value = 0.0007673

[1] "Test KRUSKALA wykrył, że sa roznice miedzy grupami -> TEST DUNNA"
[1] "Test DUNNA:"
      Comparison      Z      P.unadj      P.adj
1  CHOR1 - CHOR2 -0.1201601 0.9043562952 0.904356295
2 CHOR1 - KONTROLA 3.2183431 0.0012893348 0.002578670
3 CHOR2 - KONTROLA 3.3385033 0.0008423104 0.002526931
[1] "Roznica miedzy CHOR1 - CHOR2 w parametrze ' Z ' jest istotna statycznie"
[1] "Roznica miedzy CHOR1 - KONTROLA w parametrze ' P.unadj ' jest istotna statycznie"
[1] "Roznica miedzy CHOR2 - KONTROLA w parametrze ' P.unadj ' jest istotna statycznie"
[1] "Roznica miedzy CHOR1 - KONTROLA w parametrze ' P.adj ' jest istotna statycznie"
[1] "Roznica miedzy CHOR2 - KONTROLA w parametrze ' P.adj ' jest istotna statycznie"
```

2)

```
[1] "Dla parametru HCT"
[1] "wartosci rozkladu normalnego"
[1] 2.126689e-01 3.279665e-06 6.084101e-01
[1] "Grupy nie sa zgodne z rozkladem normalnym -> TEST KRUSKALA"
[1] "Test kruskala:"

kruskal-wallis rank sum test

data: dane[, numer] by grupa
kruskal-wallis chi-squared = 7.9308, df = 2, p-value = 0.01896

[1] 0.01896091

kruskal-wallis rank sum test

data: dane[, numer] by grupa
kruskal-wallis chi-squared = 7.9308, df = 2, p-value = 0.01896

[1] "Test KRUSKALA wykryl, ze sa roznice miedzy grupami -> TEST DUNNA"
[1] "Test DUNNA:"
      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
[1] "Roznica miedzy CHOR1 - KONTROLA w parametrze ' P.unadj ' jest istotna statycznie"
[1] "Roznica miedzy CHOR1 - KONTROLA w parametrze ' P.adj ' jest istotna statycznie"
```

3)

```
[1] "Dla parametru MCHC"
[1] "wartosci rozkladu normalnego"
[1] 0.2266247 0.2832569 0.2741520
[1] "Grupy sa zgodne z rozkladem normalnym"
[1] "wystepuje jednorodnosc wariacji -> TEST ANOVA"
[1] 0.00185981

[1] "Test ANOVA wykryl, ze sa roznice miedzy tymi grupami -> TEST TUKEY"
[1] "Test TUKEY:"
      diff      lwr      upr      p adj
CHOR2-CHOR1 0.4232228 -0.3274109 1.17385653 0.372940393
KONTROLA-CHOR1 -0.7261892 -1.4768229 0.02444453 0.060043259
KONTROLA-CHOR2 -1.1494120 -1.9000457 -0.39877827 0.001352322
[1] "Sa istotne roznice statyczne w parametrze diff w wierszu 2"
[1] "Sa istotne roznice statyczne w parametrze diff w wierszu 3"
[1] "Sa istotne roznice statyczne w parametrze lwr w wierszu 1"
[1] "Sa istotne roznice statyczne w parametrze lwr w wierszu 2"
[1] "Sa istotne roznice statyczne w parametrze lwr w wierszu 3"
[1] "Sa istotne roznice statyczne w parametrze upr w wierszu 2"
[1] "Sa istotne roznice statyczne w parametrze upr w wierszu 3"
[1] "Sa istotne roznice statyczne w parametrze p adj w wierszu 3"
```

Dla pozostałych parametrów nie stwierdzono różnic istotnych statystycznie w żadnych parametrach.

- **Do przetestowania programu, w sytuacji analizowania dwóch grup niezależnych, z pliku wejściowego usunięta została grupa KONTROLA. Analizie podlegają zatem teraz dwie grupy: Chorzy_1 oraz Chorzy_2.**

- **Omówienie poszczególnych testów dla dokładnie dwóch grup:**
 - a. Test Wilcoxona** (dokładniej test sumy rang Wilcoxona zwany również testem Manna–Whitneya) - wykonywany jeśli dane nie spełniają założenia o zgodności z rozkładem normalnym ($p\text{-value} < 0.05$).

Przykład:

```
[1] "TESTY:"
[1] "Dla parametru hsCRP"
[1] "wartosci rozkladu normalnego"
[1] 9.933661e-08 4.998515e-03
[1] "Grupy nie sa zgodne z rozkladem normalnym"
[1] "Test WILCOXONA"
```

```
      wilcoxon rank sum test with continuity correction

data:  dane[, numer] by grupa
W = 295, p-value = 0.7415
alternative hypothesis: true location shift is not equal to 0

[1] "Test WILCOXONA wykrył, że nie ma znaczych różnic między grupami"
```

- b. Test t-Studenta** – wykonywany jeśli dane spełniają założenie o zgodności z rozkładem normalnym ($p\text{-value} > 0.05$) oraz spełniają założenie o jednorodności wariancji ($p\text{-value} > 0.05$).

Przykład:

```
[1] "TESTY:"
[1] "Dla parametru MCHC"
[1] "wartosci rozkladu normalnego"
[1] 0.2266247 0.2832569
[1] "Grupy sa zgodne z rozkladem normalnym"
[1] "wystepuje jednorodnosc wariancji"
[1] "Test T.STUDENTA"
```

```
      Two sample t-test

data:  dane[, numer] by grupa
t = -1.3559, df = 48, p-value = 0.1815
alternative hypothesis: true difference in means between group CHOR1 and group CHOR2 is not equal to 0
95 percent confidence interval:
 -1.0508070  0.2043614
sample estimates:
mean in group CHOR1 mean in group CHOR2
      35.12882      35.55204

[1] 0.1814734
```

```
[1] "Test T-STUDENTA wykrył, że nie ma znaczących różnic między grupami"
```

- c. Test Welcha** – wykonywany jeśli dane spełniają założenie o zgodności z rozkładem normalnym ($p\text{-value} > 0.05$), ale nie spełniają założenia o jednorodności wariancji ($p\text{-value} < 0.05$).

Na podstawie grup Chorzy_1 oraz Chorzy_2 program nie był zmuszony do skorzystania z testu w żadnym przypadku.

We wszystkich testach jeśli wartość $p\text{-value}$ jest mniejsza niż poziom istotności 0.05 możemy stwierdzić, że istnieją znaczące różnice między grupami.

- **Analiza korelacji (określenie pomiędzy którymi parametrami w obrębie jakich grup występują istotne statystycznie korelacje oraz określenie siły i kierunku korelacji):**

- Do przeprowadzenia testu korelacji służy funkcja `cor.test()`, przyjmuje ona argument: 'pearson' lub 'spearman'.
- **Współczynnik korelacji liniowej Pearsona** – jest stosowaną miarą siły i kierunku zależności w przypadku parametrycznego testu korelacji (gdy są spełnione założenia o zgodności z rozkładem normalnym i założenia o homogeniczności wariancji).
- Nieparametrycznym odpowiednikiem współczynnika korelacji Pearsona jest **współczynnik korelacji rangowej Spearmana** – stosowany dla konkretnych parametrów gdy nie są spełnione założenia o zgodności z rozkładem normalnym i o jednorodność wariancji)

Podczas sprawdzania warunków o zgodności z rozkładem normalnym oraz jednorodności wariancji kolejne parametry zostały zapisane odpowiednio do listy `numery_parametryczne` lub `numery_nieparametryczne`.

Przy wyborze testu dla kolejnych dwóch parametrów sprawdzamy czy należą one do grupy **parametrycznych** bądź **nieparametrycznych** i na tego podstawie wybieramy odpowiednią metodę obliczania **współczynnika korelacji**.

- **Zidentyfikowane korelacje (dla danych z początkowymi trzema grupami):**

1.

```
"Grupa - KONTROLA | Parametry: hSCRp - LEU"
"wyniki:"
"P.value = 0.0350910111033194 < 0.05 - korelacja miedzy zmiennymi"
"Rodzaj korelacji:"
"r = 0.425384615384615 > 0 - korelacja Dodatnia"
"Interpretacja sily korelacji:"
"0.425384615384615 - korelacja dodatnia o srednim stezeniu"
```

2.

```
"Grupa - KONTROLA | Parametry: ERY - PLT"
"wyniki:"
"P.value = 0.0225324670621139 < 0.05 - korelacja miedzy zmiennymi"
"Rodzaj korelacji:"
"r = 0.454282972805216 > 0 - korelacja Dodatnia"
"Interpretacja sily korelacji:"
"0.454282972805216 - korelacja dodatnia o srednim stezeniu"
```


3.

```
"Grupa - CHOR1 | Parametry: ERY - HGB"
"Wyniki:"
"P.value = 0.000508823737600965 < 0.05 - Korelacja miedzy zmiennymi"
"Rodzaj korelacji:"
"r = 0.644304035226098 > 0 - Korelacja Dodatnia"
"Interpretacja sily korelacji:"
"0.644304035226098 - silna korelacja dodatnia"

"Grupa - CHOR2 | Parametry: ERY - HGB"
"Wyniki:"
"P.value = 5.08515255680143e-09 < 0.05 - Korelacja miedzy zmiennymi"
"Rodzaj korelacji:"
"r = 0.883104217853547 > 0 - Korelacja Dodatnia"
"Interpretacja sily korelacji:"
"0.883104217853547 - Bardzo silna korelacja dodatnia"

"Grupa - KONTROLA | Parametry: ERY - HGB"
"Wyniki:"
"P.value = 0.00805695649170014 < 0.05 - Korelacja miedzy zmiennymi"
"Rodzaj korelacji:"
"r = 0.517547858370556 > 0 - Korelacja Dodatnia"
"Interpretacja sily korelacji:"
"0.517547858370556 - silna korelacja dodatnia"
```

4.

```
"Grupa - CHOR1 | Parametry: ERY - HCT"
"Wyniki:"
"P.value = 0.0010689638173705 < 0.05 - Korelacja miedzy zmiennymi"
"Rodzaj korelacji:"
"r = 0.615014448352824 > 0 - Korelacja Dodatnia"
"Interpretacja sily korelacji:"
"0.615014448352824 - silna korelacja dodatnia"

"Grupa - CHOR2 | Parametry: ERY - HCT"
"Wyniki:"
"P.value = 1.49386813661715e-05 < 0.05 - Korelacja miedzy zmiennymi"
"Rodzaj korelacji:"
"r = 0.751445588857523 > 0 - Korelacja Dodatnia"
"Interpretacja sily korelacji:"
"0.751445588857523 - Bardzo silna korelacja dodatnia"

"Grupa - KONTROLA | Parametry: ERY - HCT"
"Wyniki:"
"P.value = 0.000191622089933282 < 0.05 - Korelacja miedzy zmiennymi"
"Rodzaj korelacji:"
"r = 0.678729560216607 > 0 - Korelacja Dodatnia"
"Interpretacja sily korelacji:"
"0.678729560216607 - silna korelacja dodatnia"
```

5.

```
"Grupa - KONTROLA | Parametry: ERY - MCHC"  
"wyniki:"  
"P.value = 0.00907403764088521 < 0.05 - Korelacja miedzy zmiennymi"  
"Rodzaj korelacji:"  
"r = -0.510798910888078 < 0 - Korelacja Ujemna"  
"Interpretacja sily korelacji:"  
"-0.510798910888078 - silna korelacja ujemna"
```

6.

```
"Grupa - CHOR2 | Parametry: PLT - MCHC"  
"wyniki:"  
"P.value = 0.00779894485526622 < 0.05 - korelacja miedzy zmiennymi"  
"Rodzaj korelacji:"  
"r = -0.519372468228172 < 0 - korelacja Ujemna"  
"Interpretacja sily korelacji:"  
"-0.519372468228172 - silna korelacja ujemna"
```

7.

```
"Grupa - CHOR1 | Parametry: HGB - HCT"  
"wyniki:"  
"P.value = 6.30675706805627e-12 < 0.05 - korelacja miedzy zmiennymi"  
"Rodzaj korelacji:"  
"r = 0.936175429355085 > 0 - Korelacja Dodatnia"  
"Interpretacja sily korelacji:"  
"0.936175429355085 - Bardzo silna korelacja dodatnia"  
  
"Grupa - CHOR2 | Parametry: HGB - HCT"  
"wyniki:"  
"P.value = 2.64876689268217e-05 < 0.05 - Korelacja miedzy zmiennymi"  
"Rodzaj korelacji:"  
"r = 0.736964573555573 > 0 - Korelacja Dodatnia"  
"Interpretacja sily korelacji:"  
"0.736964573555573 - Bardzo silna korelacja dodatnia"  
  
"Grupa - KONTROLA | Parametry: HGB - HCT"  
"wyniki:"  
"P.value = 1.2210574418466e-10 < 0.05 - Korelacja miedzy zmiennymi"  
"Rodzaj korelacji:"  
"r = 0.916716244015588 > 0 - Korelacja Dodatnia"  
"Interpretacja sily korelacji:"  
"0.916716244015588 - Bardzo silna korelacja dodatnia"
```

8.

```
"Grupa - CHOR1 | Parametry: HGB - MCHC"  
"wyniki:"  
"P.value = 0.00219641235166691 < 0.05 - Korelacja miedzy zmiennymi"  
"Rodzaj korelacji:"  
"r = 0.583574977286289 > 0 - Korelacja Dodatnia"  
"Interpretacja sily korelacji:"  
"0.583574977286289 - silna korelacja dodatnia"
```

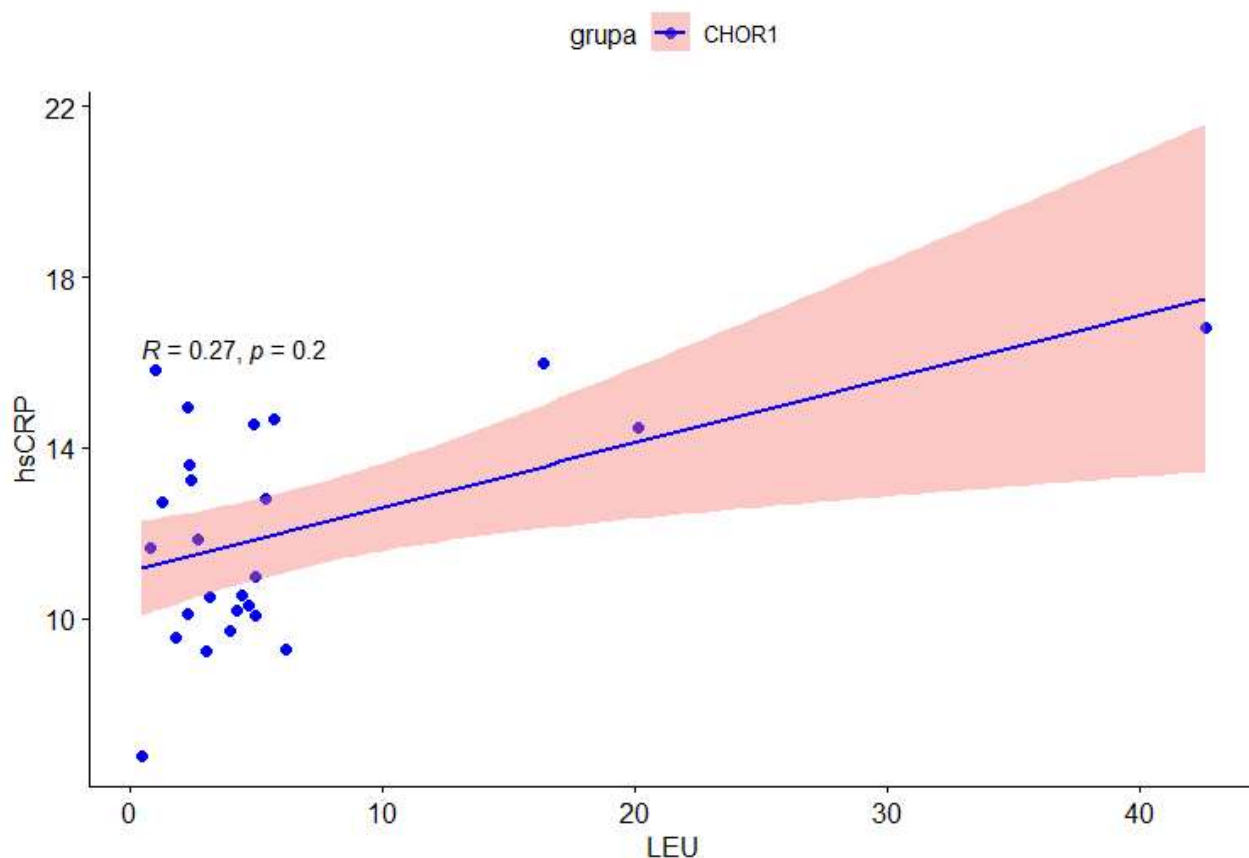
9.

```
"Grupa - CHOR1 | Parametry: HCT - MCHC"  
"wyniki:"  
"P.value = 0.0378334611214229 < 0.05 - korelacja między zmiennymi"  
"Rodzaj korelacji:"  
"r = 0.417519751807869 > 0 - korelacja Dodatnia"  
"Interpretacja siły korelacji:"  
"0.417519751807869 - korelacja dodatnia o srednim stezeniu"
```

- Dla zidentyfikowanych korelacji program tworzy również **wykres przedstawiający graficzne korelacje** między konkretnymi parametrami. Funkcja tworząca wykres posiada w argumencie dodatkowy element o dodaniu **modelu regresji liniowej** - regresja ta zakłada, że zależność między parametrami jest zależnością liniową. W efekcie dopasowana linia regresji reprezentuje **oszacowaną wartość oczekiwaną** zmiennej parametru pierwszego przy konkretnych wartościach parametru drugiego.

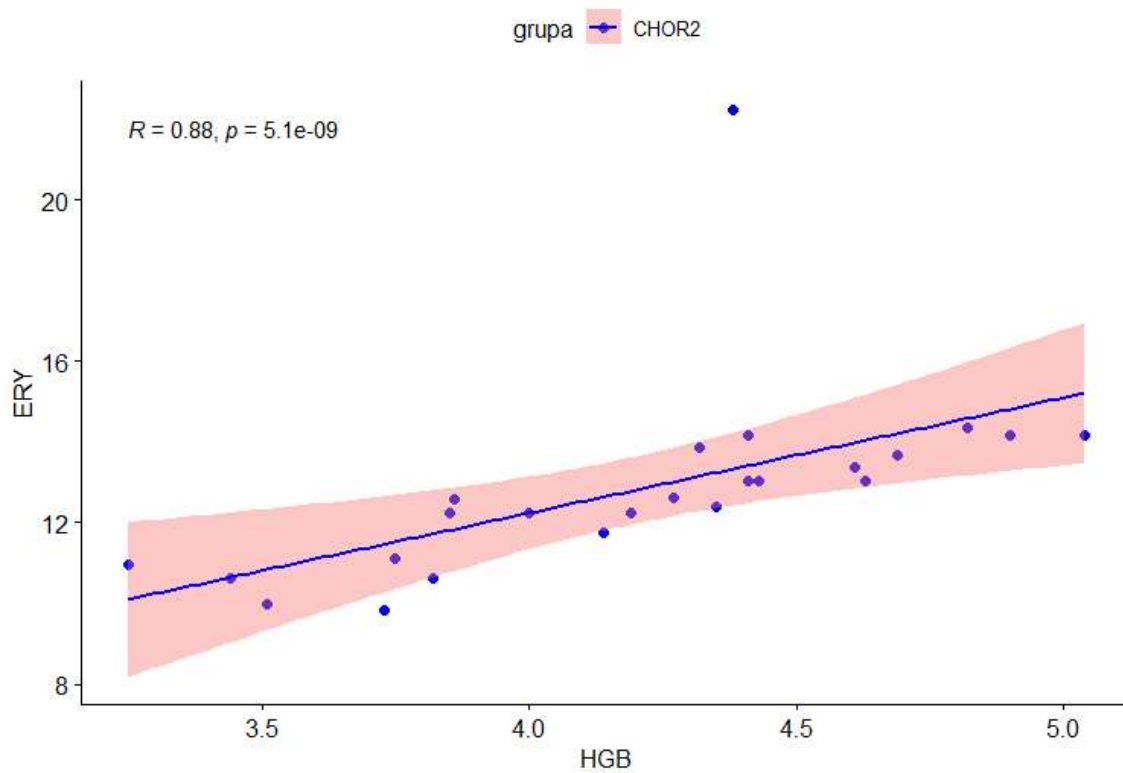
Wybrane wykresy:

Wykres dla metody spearman dla parametrow hsCRP - LEU



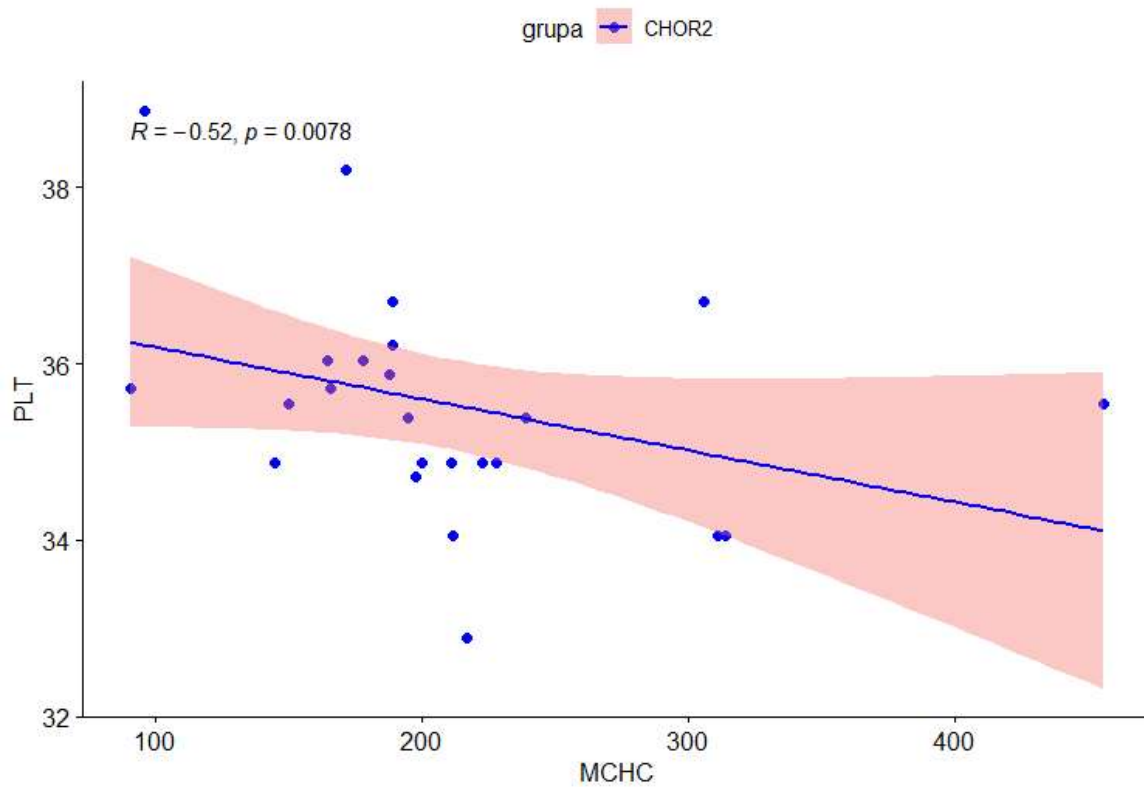
- przykład korelacji dodatniej o średnim stężeniu

Wykres dla metody spearman dla parametrow ERY - HGB



- przykład bardzo silnej korelacji dodatniej

Wykres dla metody spearman dla parametrow PLT - MCHC



- przykład silnej korelacji ujemnej

- **Uruchomienie programu z trybu wsadowego (BATCH MODE):**

- 1) Poleceniem 'cd *wybrany_Dysk* : ' wybieramy dysk z pobranym projektem
- 2) Poleceniem 'cd *sciezka_do_pliku*' ustalamy dokładny adres, w którym znajduje się projekt i plik *Projekt_skrypt.R*

```
C:\Users\adam1>cd D:  
D:\Studia\2 rok\2 semestr\PP\Statystyka\Projekt\Projekt_zaliczeniowy_Adam_tangowski
```

- 3) Będąc już w odpowiednim katalogu – w którym znajdują się pliki *Projekt_skrypt.R* oraz *przykladoweDane-Projekt.csv* (to plik z naszymi danymi medycznymi) wpisujemy polecenie:

```
"C:\Program Files\R\R-4.2.0\bin\R.exe" CMD BATCH --vanilla "--args  
przykladoweDane-Projekt.csv" Projekt_skrypt.R
```


Należy zwrócić uwagę na wersję R – 4.2.0.

Plik *przykladoweDane-Projekt.csv* jest argumentem przekazywanym do programu.

```
D:\Studia\2 rok\2 semestr\PP\Statystyka\Projekt\Projekt_zaliczeniowy_Adam_tangowski  
>"C:\Program Files\R\R-4.2.0\bin\R.exe" CMD BATCH --vanilla "--args przykladoweDane-  
Projekt.csv" Projekt_skrypt.R
```

W efekcie poprawnego zadziałania programu w folderze ze skrypcem pojawiają się pliki:

- plik *Projekt_skrypt.Rout* (który możemy odczytać dowolnym edytorem tekstowym) – przechowuje on wszystkie wyniki działania programu oraz
- plik *Rplots.pdf* prezentujący wszystkie wykresy wykonane na potrzeby analizy danych

 Projekt_skrypt.Rout	03.06.2022 09:55	Plik ROUT	59 KB
 Rplots.pdf	03.06.2022 09:55	Microsoft Edge P...	165 KB

**Konieczne biblioteki zainstalowane w R:*

library(dplyr), library(Hmisc), library(ggplot2), library(dplyr), library(e1071),
library(ggpubr), library(car), library(FSA)