

Raport końcowy

Analiza danych dotyczących choroby nerek

Przedmiot: Statystyka w medycynie

Prowadzący: dr hab. Agnieszka Goroncy

Ewa Kostrzewska

Wiktoria Musialska

studentki analizy danych

Narzędzie

Do przeprowadzenia analiz wykorzystano język programowania R w wersji 4.1.0 z następującymi bibliotekami: *readxl*, *moments*, *ggplot2*, *gridExtra*, *grid*, *dplyr*, *GGally*.

Informacje o analizach

Wszystkie analizy zostały przeprowadzone przyjmując z góry poziom istotności 0,05. Obserwacje odstające zostały zidentyfikowane na podstawie kryterium kwartylowego.

Zbiór danych

- Plik *micro-RNA 2020-2.xls* zawiera dane 26 pacjentów z chorobą nerek oraz 9 pacjentów będących w grupie kontrolnej.
- Dla ułatwienia przeprowadzenia analiz, dane zostały podzielone na dwa pliki *badana.xlsx* oraz *kontrola.xlsx* zawierający wyżej wspomnianych pacjentów w podziale na osoby chore i przypadki kontrolne.
- Dane grupy badanej składają się z 38 zmiennych: "nr", "nazwisko", "telefon", "numer próbki/numer pacjenta", "PChN etiologia", "wiek", "SBP office", "DBP office", "albumina", "białko całkowite", "wapń całkowity", "kreatynina", "MDRD", "CKD-EPI", "CKD stage", "glukoza", "fosforan", "UA", "mocznik", "CRP", "LDL", "cholesterol całkow.", "HDL", "TG", "fibrynogen", "Na", "K", "Cl", "DUB", "HGB", "PLT", "WBC", "wzrost", "waga", "Pas", "Biodra", "BSA", "BMI". Ze względu na poufność danych, zmienne "nazwisko" oraz "telefon" nie zawierały informacji.
- Dane grupy kontrolnej składają się z 38 zmiennych: "nr", "KONTROLA", "telefon", "...4", "...5", "wiek", "SBP office", "DBP office", "albumina", "białko całkowite", "wapń całkowity", "kreatynina", "MDRD", "CKD-EPI", "CKD stage", "glukoza", "fosforan", "UA", "mocznik", "CRP", "LDL", "cholesterol całkow.", "HDL", "TG", "fibrynogen", "Na", "K", "Cl", "DUB", "HGB", "PLT", "WBC", "wzrost", "waga", "Pas", "Biodra", "BSA", "BMI". Ze względu na poufność danych, zmienne "KONTROLA" oraz "telefon" nie zawierały informacji. Zmienne "...4" oraz "...5" nie miały nazwy w oryginalnym pliku z danymi, jednak wnioskując po ich umieszczeniu w grupie kontrolnej w porównaniu do grupy badanej, odpowiadałyby za "numer próbki/numer pacjenta" oraz "PChN etiologia".
- Dla ułatwienia analiz, nazwy zmiennych w grupie kontrolnej nazwano tak, jak w grupie badanej. Ze względu na to, że w grupie kontrolnej są osoby zdrowe, zmienna "PChN etiologia" nie ma wartości.
- Po wczytaniu danych do programu, okazało się, że zmienne *CRP* oraz *WBC* miały typ tekstowy, więc konieczne było skonwertowanie ich na dane liczbowe.

Podstawowe informacje o pacjentach

1. Wiek

	Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
Grupa badana	20,00	32,00	38,50	39,96	42,75	78,00
Grupa kontrolna	25,00	45,00	50,00	47,67	56,00	58,00

Tabela 1. Podstawowe statystyki wieku w grupie badanej oraz w grupie kontrolnej.

W grupie badanej są pacjenci w przedziale wiekowym od 20 do 78 lat, a ich średnia wieku wynosi 39,96 lat.

W grupie kontrolnej są pacjenci w przedziale wiekowym od 25 do 58 lat, a średnia wieku wynosi 47,67 lat.

Zarówno w grupie badanej jak i kontrolnej, zmienna *wiek* nie zawiera braków danych.

2. Wzrost

	Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
Grupa badana	0,0	170,0	176,0	168,6	180,0	195,0
Grupa kontrolna	152,0	165,0	174,0	171,4	175,0	190,0

Tabela 2. Podstawowe statystyki wzrostu w grupie badanej oraz w grupie kontrolnej.

Wzrost w grupie badanej wynosił od 0 do 195 cm. Wartość minimum równe 0 sugeruje, że u pewnego pacjenta/pacjentów nie zapisano wzrostu. Jest to pacjent o numerze (nr) 36. Drugim najmniejszym wzrostem jest 160 cm. Wzrost w grupie kontrolnej waha się między 152 a 190 cm, a jej średnia wynosi 171,4 cm.

3. Waga

	Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
Grupa badana	0,00	67,75	80,50	81,19	96,25	134,00
Grupa kontrolna	55	60	65	66	70	88

Tabela 3. Podstawowe statystyki wagi w grupie badanej oraz w grupie kontrolnej.

Waga w grupie badanej waha się między 0 a 134 kg. Wartość 0 wskazuje, że u pewnego pacjenta/pacjentów nie została wpisana waga. Jest to pacjent o numerze (nr) 36. W grupie kontrolnej, waga przyjmuje wartości od 55 do 88 kg, a jej średnia wynosi 66 kg.

4. BMI

W zmiennej *BMI* pojawił się brak danych u pacjenta o numerze (nr) 36. To ten sam pacjent, u którego zauważono zerową wartość wzrostu i wagi. Został on wykluczony w analizie tej zmiennej.

Zmienna *BMI* została skategoryzowana według następujących kryteriów:

- poniżej 16- wygłodzenie
- 16 – 16,99- wychudzenie
- 17 – 18,49- niedowaga
- 18,5- 24,99- waga prawidłowa
- 25- 29,99- nadwaga
- Powyżej 30- otyłość

Tak utworzona zmienna została nazwana jako *BMI_kat*.

waga prawidłowa	nadwaga	otyłość
11	7	7

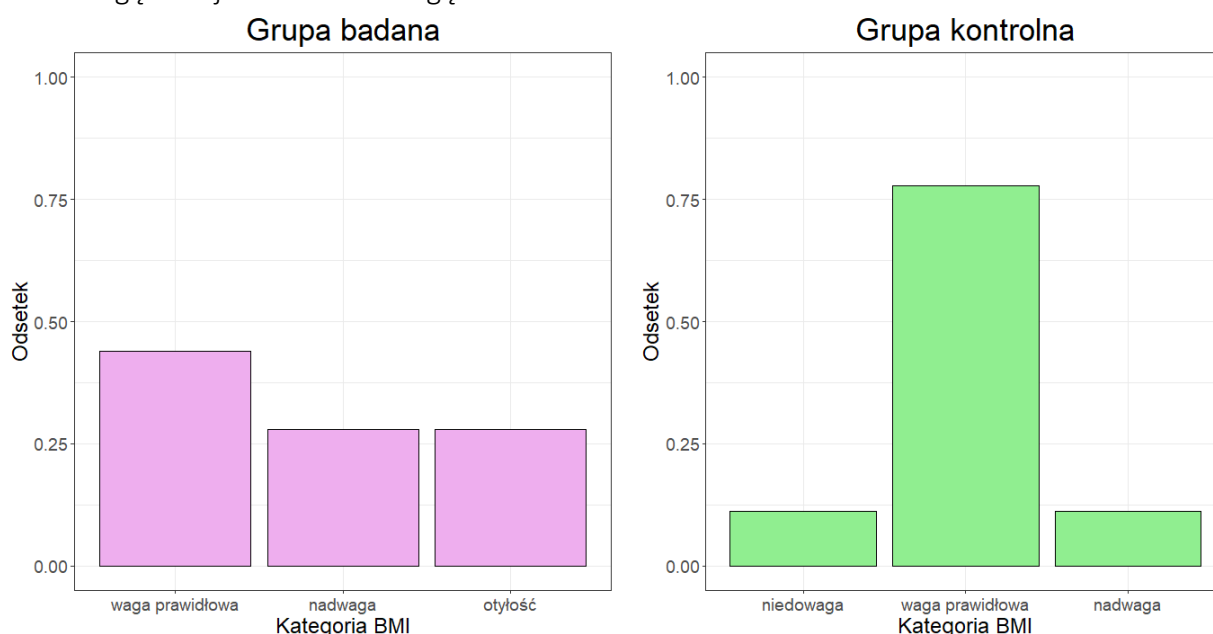
Tabela 4. Liczebności kategorii BMI w grupie badanych.

Liczebności w grupie nadwaga i otyłość są takie same i wynoszą 7, natomiast waga prawidłowa pojawiła się u 11 pacjentów. W grupie badanych nie ma osób z wygłodzeniem, wychudzeniem czy niedowagą.

niedowaga	waga prawidłowa	nadwaga
1	7	1

Tabela 5. Liczebności kategorii BMI w grupie kontrolnej.

Najwięcej pacjentów w grupie kontrolnej ma wagę prawidłową i jest to 7 osób. Jedna osoba ma niedowagę oraz jedna ma nadwagę.



Wykresy 1. Rozkład odsetków osób według kategorii BMI w grupie badanej i kontrolnej.

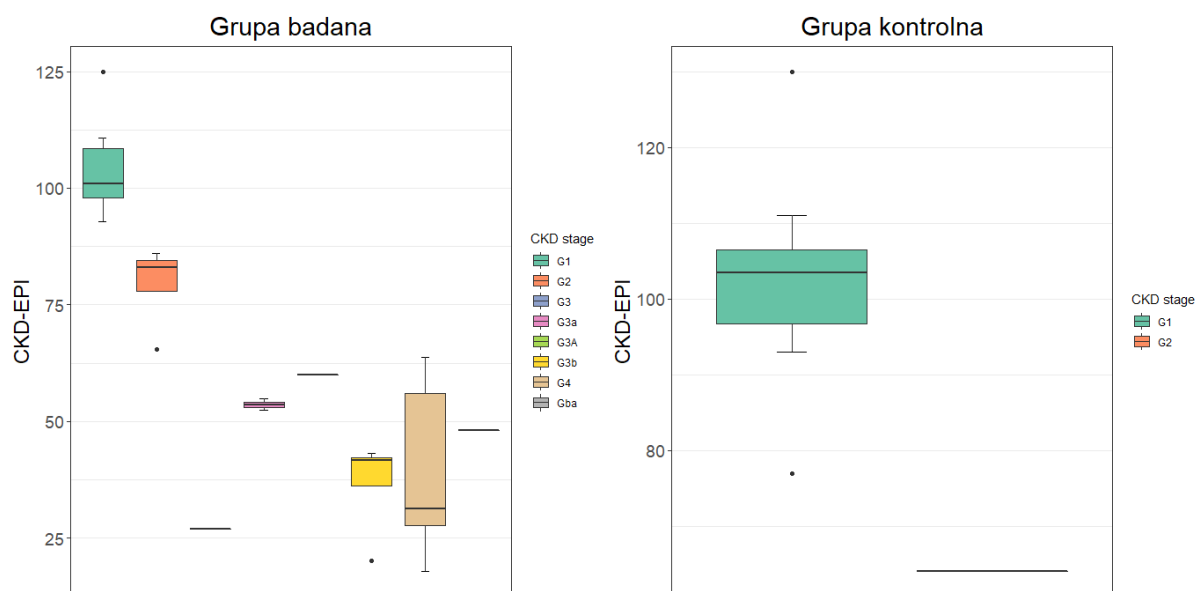
5. CKD stage

Zmienna *CKD stage* jest zmienną kategorialną, która informuje, jaki stadium choroby ma dany pacjent. Stadium zaawansowania choroby jest wyznaczany na podstawie szacowanego wskaźnika filtracji kłębuszkowej. Wskaźnik ten można wyznaczyć dwoma metodami, których wyniki były zapisane w zbiorze danych: *MDRD* oraz *CKD-EPI*. Na podstawie wyestymowanego wskaźnika, stadia choroby nerek pacjentów można przypisać w następujący sposób (na podstawie artykułu Łukasza Zdrojewskiego oraz Bolesława Rutkowskiego pt. „*MDRD czy CKD-EPI — rewolucja czy ewolucja?*”):

- Powyżej lub równe 90: G1
- 60 – 89: G2
- 45 – 59: G3A
- 30 – 44: G3B
- 15 – 29: G4
- Poniżej 15: G5

Oznacza to, że im niższa jest wartość *CKD-EPI* lub *MDRD*, tym wyższe jest stadium choroby. W analizowanych danych, zmienna *CKD stage* była utworzona w niepoprawny sposób, dlatego była wymagana zmiana przypisania stadium zaawansowania choroby do pacjentów. Nowego przypisania dokonano na podstawie zmiennej *CKD-EPI*.

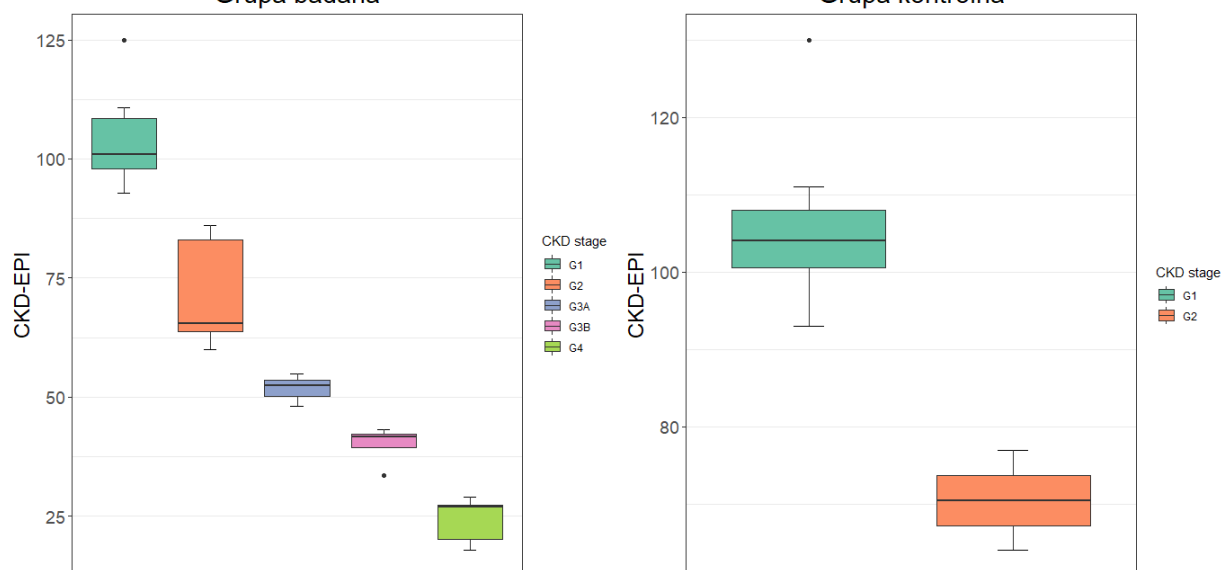
Przed poprawieniem wartości wg artykułu rozkład CKD według stadiów choroby wyglądał tak jak na wykresach 2.



Wykresy 2. Rozkład wskaźnika CKD-EPI w poszczególnych stadiach zaawansowania choroby przed naniesionymi poprawkami w grupie badanej i kontrolnej.

Widzimy, że wartości dla G2, G3, G3A, G3B oraz G4 są niezgodne z skalą zamieszczoną w artykule.

Po poprawieniu etykiet według skali zamieszczonej w artykule:



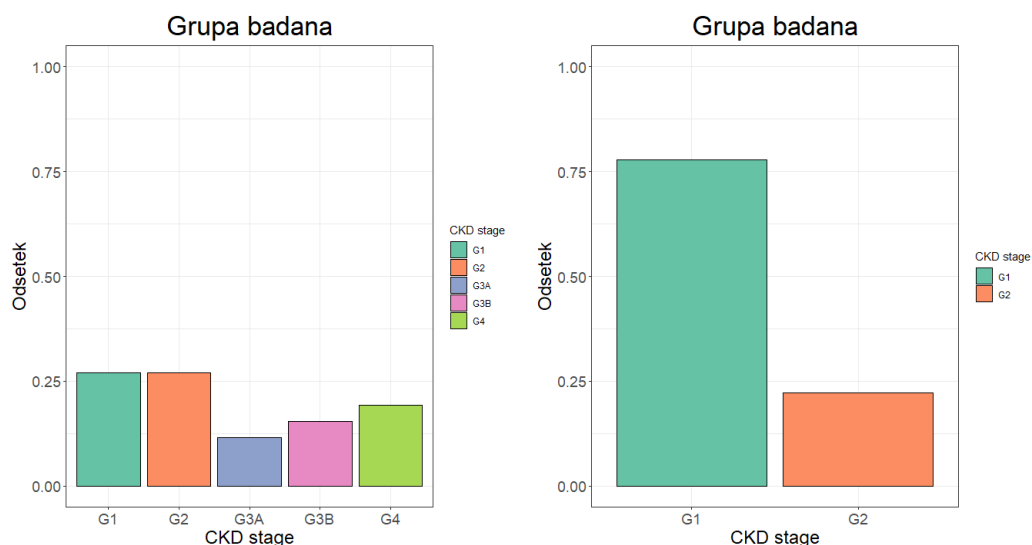
Wykresy 3. Rozkład wskaźnika CKD-EPI w poszczególnych stadiach zaawansowania choroby po naniesieniu poprawek.

W stadium G1 (w obu grupach) oraz G3B (w grupie badanej) występują obserwacje odstające.

	G1	G2	G3A	G3B	G4	G5
Grupa badana	7	7	3	4	5	-
Grupa kontrolna	7	2	-	-	-	-

Tabela 6. Liczebności stadiów zaawansowania choroby nerek w grupie badanej i kontrolnej.

W grupie badanej wystąpiły 4 stadia choroby, natomiast w grupie kontrolnej 2 stadia. Najniższe stadium G1 wystąpiło u 7 pacjentów, zarówno w grupie badanej, jak i kontrolnej. Stadium G2 choroby ma 7 pacjentów z grupy badanej oraz 2 pacjentów z grupy kontrolnej. W grupie badanej stadium G3A miało 3 osoby, stadium G3B 4 osoby, a stadium G4 5 osób.



Wykresy 4. Rozkład odsetków osób według stadia choroby w grupie badanej i kontrolnej.

WSTĘPNE HIPOTEZY – PORÓWNANIE WYBRANYCH ZMIENNYCH W GRUPACH

1. CKD-EPI

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
17,80	41,62	61,80	64,35	91,17	125,00

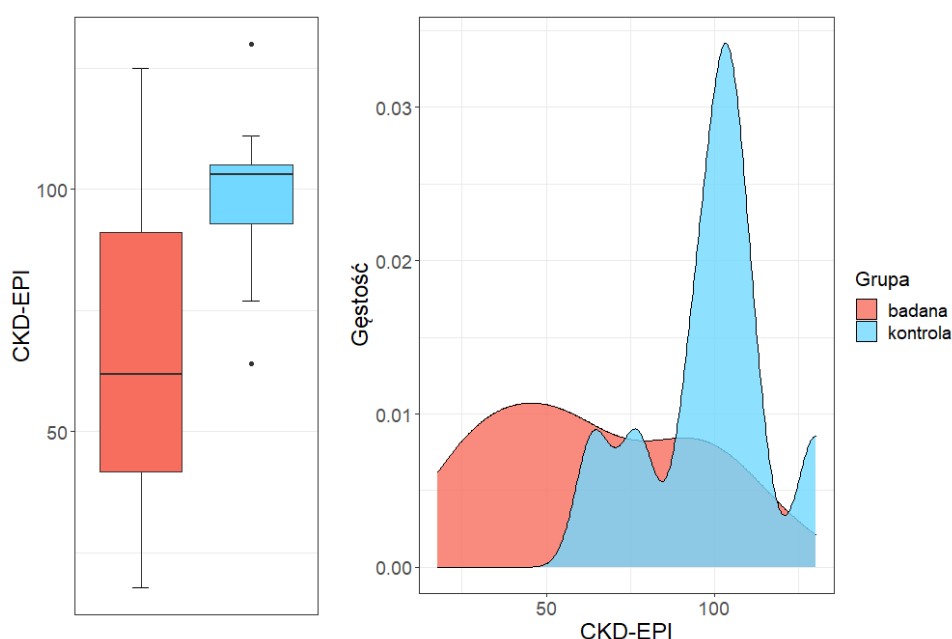
Tabela 7. Podstawowe statystyki CKD-EPI w grupie badanej.

CKD-EPI w grupie badanej przyjmuje wartości od 17,8 do 125, a jej średnia wynosi 64,35. Dane mogą być prawostronnie skośne, ponieważ średnia ma większą wartość niż mediana. Nie występują braki danych.

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
64,00	93,00	103,00	98,33	105,00	130,00

Tabela 8. Podstawowe statystyki CKD-EPI w grupie kontrolnej.

CKD-EPI w grupie kontrolnej przyjmuje wartości od 64 do 130 i jest to mniejsza rozbieżność niż dla grupy badanej. Jej średnia wynosi 98,33 i jest większa niż w grupie badanej. Dane mogą być lewostronnie skośne, ponieważ średnia jest mniejsza niż mediana. Nie występują braki danych.



Wykres 5. Rozkład CKD-EPI w grupie badanej i kontrolnej.

W grupie kontrolnej mamy dwie wartości odstające: 64 oraz 130, należące odpowiednio do pacjentów o numerze (nr) 10 i 15. Wykres gęstości dla grupy badanej obrazuje prawostronną skośność, natomiast wykres gęstości dla grupy kontrolnej jest bardziej symetryczny.

Badanie skośności rozkładu CKD-EPI w obu grupach przy użyciu testu D'Agostino:

H_0 : Rozkład zmiennej jest symetryczny (skośność jest równa 0).

H_1 : Rozkład zmiennej jest skośny (skośność jest różna od 0).

	Grupa badana	Grupa kontrolna
Skośność	0,22405	-0,30292
Statystyka testowa	0,55189	-0,52514
P-wartość	0,581	0,5995

Tabela 9. Wyniki testu skośności zmiennej CKD-EPI.

Rozkład w grupie badanej ma dodatnią skośność, ale ta skośność nie jest istotnie różna od 0 (p -wartość $> 0,05$), natomiast w grupie kontrolnej występuje ujemna skośność, która również nie jest istotna statystycznie. Rozkłady CKD-EPI w obu grupach są symetryczne.

Badanie normalności rozkładu CKD-EPI w obu grupach przy użyciu testu Shapiro-Wilka:

H_0 : Rozkład zmiennej jest normalny.

H_1 : Rozkład zmiennej nie jest normalny.

	Grupa badana	Grupa kontrolna
Statystyka testowa	0,9524	0,95438
P-wartość	0,2637	0,7381

Tabela 10. Wyniki testu normalności zmiennej CKD-EPI.

Zarówno w grupie badanej jak i w grupie kontrolnej p -wartość testu jest powyżej 0,05, co oznacza, że można założyć, że rozkłady CKD-EPI w tych grupach są zbliżone do rozkładu normalnego.

Ponieważ wartości odstające nie zaburzają normalności rozkładu CKD-EPI w grupie kontrolnej, użyto testu **t-Studenta dla prób niezależnych poprawką na ciągłość** z następującymi hipotezami:

H_0 : Wartości CKD-EPI są równe w grupie badanej i kontrolnej.

H_1 : Wartości CKD-EPI są różne w grupie badanej i kontrolnej.

Statystyka testowa $t = -3,8494$, p -wartość = 0,0008076 $< 0,05$

Wniosek: Odrzucamy H_0 na rzecz H_1 . Wartości CKD-EPI są różne w grupie badanej i kontrolnej.

2. CRP

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
0,240	2,380	3,415	3,897	4,825	13,080

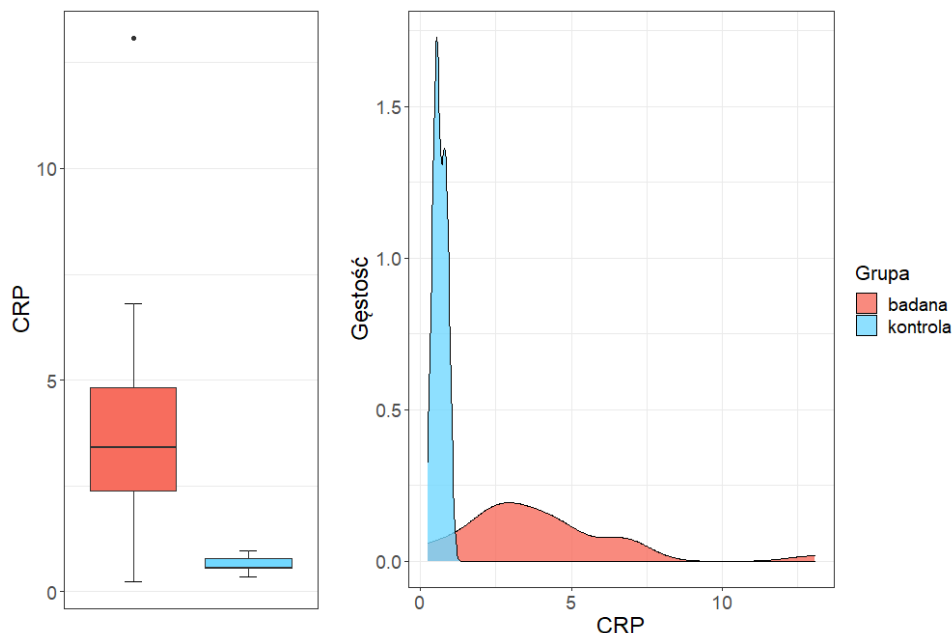
Tabela 11. Podstawowe statystyki CRP w grupie badanej.

CRP w grupie badanej przyjmuje wartości od 0,24 do 13,08, a jej średnia wynosi 3,897. Dane mogą być prawostronnie skośne, ponieważ średnia ma większą wartość niż mediana. Nie występują braki danych.

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
0,3400	0,5500	0,5600	0,6489	0,7800	0,9700

Tabela 12. Podstawowe statystyki CRP w grupie kontrolnej.

CRP w grupie kontrolnej przyjmuje wartości od 0,34 do 0,97 i jest to mała rozbieżność niż w przypadku grupy badanej. Jej średnia wynosi 0,6489 i jest mniejsza niż w grupie badanej. Dane mogą być prawostronnie skośne, ponieważ średnia jest większa niż mediana. Nie występują braki danych.



Wykresy 6. Rozkład CRP w grupie badanej i kontrolnej.

Grupa badana zawiera obserwację odstającą, która wygląda na dość nietypową. Wynosi ona 13,1 i należy do pacjenta o numerze (nr) 26. Wykres gęstości dla grupy kontrolnej jest bliższy rozkładowi normalnemu. Wykres dla grupy badanej obrazuje prawostronną skośność.

Badanie skośności rozkładu CRP w obu grupach przy użyciu testu D'Agostino:

H_0 : Rozkład zmiennej jest symetryczny (skośność jest równa 0).

H_1 : Rozkład zmiennej jest skośny (skośność jest różna od 0).

	Grupa badana	Grupa kontrolna
Skośność	1,5460	0,11748
Statystyka testowa	3,1632	0,20410
P-wartość	0,001561	0,8383

Tabela 13. Wyniki testu skośności zmiennej CRP.

Rozkład w grupie badanej jest istotnie skośny (p -wartość $< 0,05$), jej skośność jest dodatnia, co oznacza, że rozkład w tej grupie jest prawostronnie skośny. W grupie kontrolnej występuje dodatnia skośność, ale skośność nie jest istotnie różna od 0, zatem rozkład w tej grupie jest symetryczny.

Badanie normalności rozkładu CRP w obu grupach przy użyciu testu Shapiro-Wilka:

H_0 : Rozkład zmiennej jest normalny.

H_1 : Rozkład zmiennej nie jest normalny.

	Grupa badana	Grupa kontrolna
Statystyka testowa	0,86983	0,95529
P-wartość	0,003536	0,7479

Tabela 14. Wyniki testu normalności zmiennej CRP.

W grupie badanej p-wartość jest mniejsza od 0,05, co oznacza, że rozkład CRP w tej grupie nie pochodzi z rozkładu normalnego, natomiast w grupie kontrolnej p-wartość jest większa od 0,05, więc można założyć, że rozkład CRP w grupie kontrolnej jest zbliżony do rozkładu normalnego.

Ze względu na to, że w grupie badanej jest duża wartość odstająca i nie jest spełnione założenie o normalności rozkładu, użyto testu nieparametrycznego **U Manna-Whitneya z poprawką na ciągłość** do przetestowania następujących hipotez:

H_0 : Wartości CRP są równe w grupie badanej i kontrolnej.

H_1 : Wartości CRP są różne w grupie badanej i kontrolnej.

Statystyka testowa $W = 214$, p-wartość = 0,0002685 < 0,05

Wniosek: Odrzucamy H_0 na rzecz H_1 . Wartości CRP są różne w grupie badanej i kontrolnej.

3. Fibrynogen

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
174,0	345,0	369,5	384,8	450,8	578,0

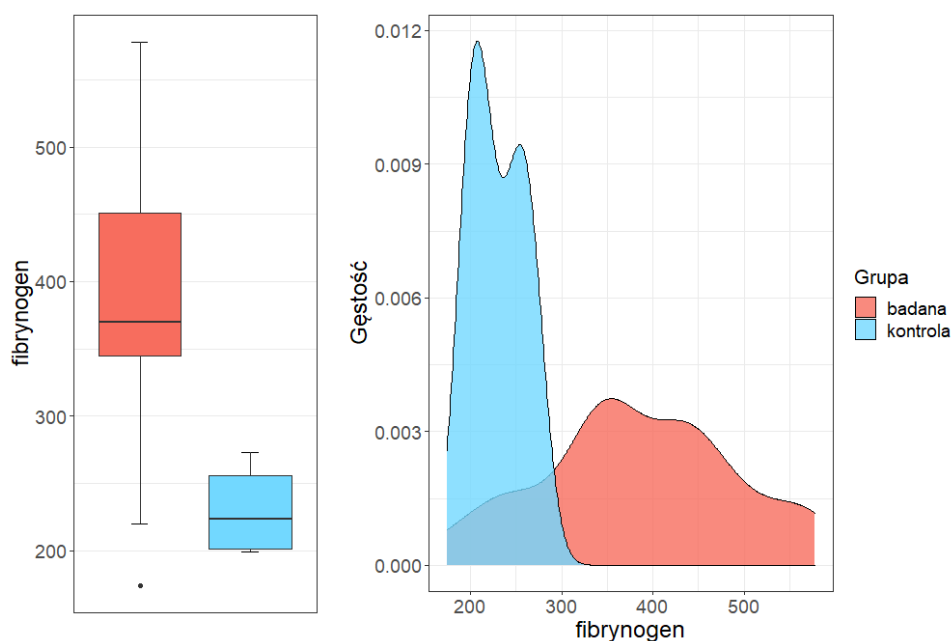
Tabela 15. Podstawowe statystyki fibrynogenu w grupie badanej.

Fibrynogen w grupie badanej przyjmuje wartości od 174 do 578, a jej średnia wynosi 384,8. Dane mogą być prawostronnie skośne, ponieważ średnia ma większą wartość niż mediana. Nie występują braki danych.

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
199,0	201,0	223,0	229,8	256,0	273,0

Tabela 16. Podstawowe statystyki fibrynogenu w grupie kontrolnej.

Fibrynogen w grupie kontrolnej przyjmuje wartości od 199 do 273 i jest to mała rozbieżność niż w przypadku grupy badanej. Jej średnia wynosi 229,8 i jest mniejsza niż w grupie badanej. Dane mogą być zbliżone do rozkładu normalnego, ponieważ mediana jest bliska średniej.



Wykresy 7. Rozkłady fibrynogenu w grupie badanej i kontrolnej.

W grupie badanej występuje wartość odstająca równa 174 i należy do pacjenta o numerze (nr) 27. Na wykresie gęstości widzimy, że rozkłady w obu grupach są zbliżone do rozkładu normalnego.

Badanie skośności rozkładu fibrynogenu w obu grupach przy użyciu testu D'Agostino:

H_0 : Rozkład zmiennej jest symetryczny (skośność jest równa 0).

H_1 : Rozkład zmiennej jest skośny (skośność jest różna od 0).

	Grupa badana	Grupa kontrolna
Skośność	-0,021140	0,23592
Statystyka testowa	-0,052385	0,40940
P-wartość	0,9582	0,6822

Tabela 17. Wyniki testu skośności zmiennej fibrynogen.

W grupie badanej skośność rozkładu danych jest ujemna i bliska zeru, natomiast w grupie kontrolnej jest dodatnia. Zarówno w grupie badanej jak i kontrolnej, zaobserwowane skośności nie są istotne statystycznie. Rozkłady w obu grupach są symetryczne.

Badanie normalności rozkładu fibrynogenu w obu grupach przy użyciu testu Shapiro-Wilka:

H_0 : Rozkład zmiennej jest normalny.

H_1 : Rozkład zmiennej nie jest normalny.

	Grupa badana	Grupa kontrolna
Statystyka testowa	0,97402	0,88745
P-wartość	0,7287	0,1878

Tabela 18. Wyniki testu normalności zmiennej fibrynogen.

Zarówno w grupie badanej jak i w grupie kontrolnej p-wartość testu jest powyżej 0,05, co oznacza, że można założyć, że rozkłady fibrynogenu w tych grupach są zbliżone do rozkładu normalnego.

Ponieważ wartość odstająca nie zaburza normalności rozkładu fibrynogenu w grupie badanej, użyto testu **t-Studenta dla prób niezależnych poprawką na ciągłość** z następującymi hipotezami:

H_0 : Wartości fibrynogenu są równe w grupie badanej i kontrolnej.

H_1 : Wartości fibrynogenu są różne w grupie badanej i kontrolnej.

Statystyka testowa $t = 6,7696$, **p-wartość** = $1,156 \cdot 10^{-7} < 0,05$

Wniosek: Odrzucamy H_0 na rzecz H_1 . Wartości fibrynogenu są różne w grupie badanej i kontrolnej.

4. PLT

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
141,0	211,5	256,0	257,1	302,0	354,0

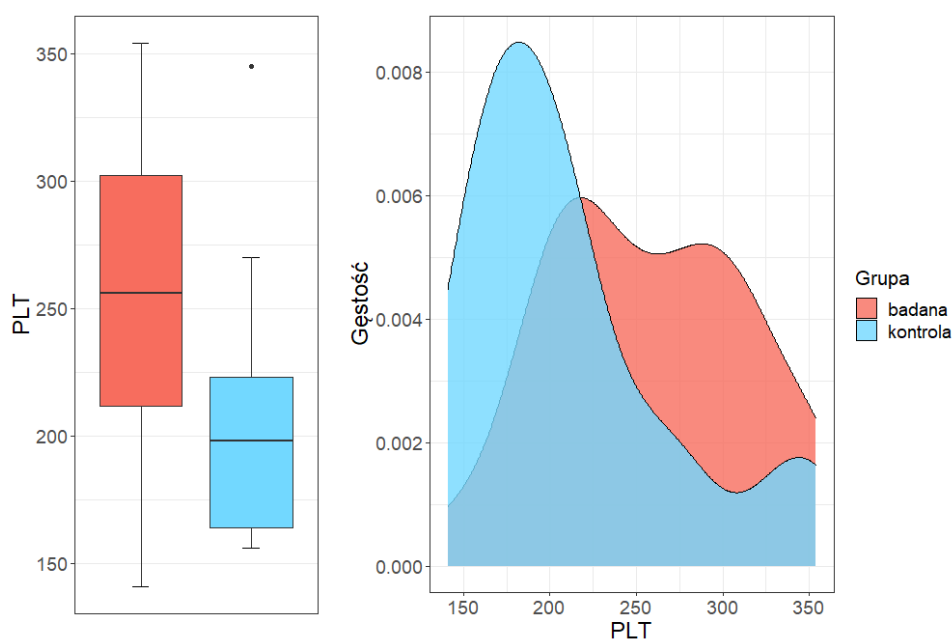
Tabela 19. Podstawowe statystyki PLT w grupie badanej.

PLT w grupie badanej przyjmuje wartości od 141 do 354, a jej średnia wynosi 257,1. Dane są zbliżone do rozkładu normalnego, ponieważ mediana jest bliska średniej. Nie występują braki danych.

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
156,0	164,0	198,0	211,9	223,0	345,0

Tabela 20. Podstawowe statystyki PLT w grupie kontrolnej.

PLT w grupie kontrolnej przyjmuje wartości od 156 do 345 i jest to mała rozbieżność niż w przypadku grupy badanej. Jej średnia wynosi 211,9 i jest mniejsza niż w grupie badanej. Dane mogą być prawostronnie skośne, gdyż mediana jest mniejsza od średniej. Nie występują braki danych.



Wykresy 8. Rozkład PLT w grupie badanej i kontrolnej.

W grupie kontrolnej występuje wartość odstająca, która jest równa 345 i należy do pacjenta o numerze (nr) 17. Na wykresach widać, że rozkład PLT w grupie badanej jest prawie symetryczny. Wykresy dla grupy kontrolnej sugerują prawostronną skośność.

Badanie skośności rozkładu PLT w obu grupach przy użyciu testu D'Agostino:

H_0 : Rozkład zmiennej jest symetryczny (skośność jest równa 0).

H_1 : Rozkład zmiennej jest skośny (skośność jest różna od 0).

	Grupa badana	Grupa kontrolna
Skośność	0,059167	1,2175
Statystyka testowa	0,146562	2,0375
P-wartość	0,8835	0,0416

Tabela 21. Wyniki testu skośności zmiennej PLT.

W grupie badanej skośność rozkładu danych nie jest istotnie różna od 0, natomiast w grupie kontrolnej skośność jest istotnie różna od 0. Jej skośność jest dodatnia, to oznacza, że rozkład PLT jest prawostronnie skośny. W grupie badanej rozkład tej zmiennej jest symetryczny.

Badanie normalności rozkładu PLT w obu grupach przy użyciu testu Shapiro-Wilka:

H_0 : Rozkład zmiennej jest normalny.

H_1 : Rozkład zmiennej nie jest normalny.

	Grupa badana	Grupa kontrolna
Statystyka testowa	0,9598	0,84809
P-wartość	0,3876	0,07099

Tabela 22. Wyniki testu normalności zmiennej PLT.

Zarówno w grupie badanej jak i w grupie kontrolnej p-wartość testu jest powyżej 0,05, co oznacza, że można założyć, że rozkłady PLT w tych grupach są zbliżone do rozkładu normalnego.

Ponieważ wartość odstająca nie zaburza normalności rozkładu PLT w grupie kontrolnej, użyto testu **t-Studenta dla prób niezależnych poprawką na ciągłość** z następującymi hipotezami:

H_0 : Wartości PLT są równe w grupie badanej i kontrolnej.

H_1 : Wartości PLT są różne w grupie badanej i kontrolnej.

Statystyka testowa $t = 1,9415$, **p-wartość** = 0,07403 > 0,05

Wniosek: Nie mamy podstaw do odrzucenia H_0 . Wartości PLT są równe w grupie badanej i kontrolnej.

5. WBC

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
3,890	6,410	7,535	7,791	8,568	12,680

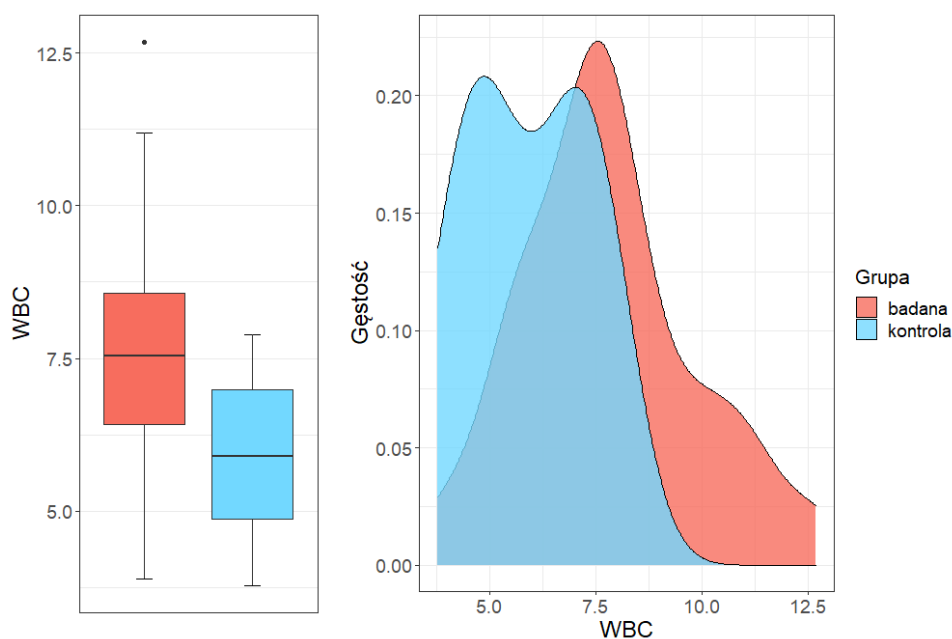
Tabela 23. Podstawowe statystyki WBC w grupie badanej.

WBC w grupie badanej przyjmuje wartości od 3,89 do 12,68, a jej średnia wynosi 7,791. Dane są zbliżone do rozkładu normalnego, ponieważ mediana jest bliska średniej. Nie występują braki danych.

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
3,780	4,870	5,890	5,907	6,980	7,890

Tabela 24. Podstawowe statystyki WBC w grupie kontrolnej.

WBC w grupie kontrolnej przyjmuje wartości od 3,78 do 7,89 i jest to mała rozbieżność niż w przypadku grupy badanej. Jej średnia wynosi 5,907 i jest mniejsza niż w grupie badanej. Dane są zbliżone do rozkładu normalnego, ponieważ mediana jest bliska średniej. Nie występują braki danych.



Wykresy 9. Rozkład WBC w grupie badanej i kontrolnej.

W grupie badanej występuje wartość odstająca, która jest równa 12,7 i należy do pacjenta o numerze (nr) 26. Na wykresach widać, że grupa badana może mieć rozkład prawostronnie skośny, natomiast rozkład w grupie kontrolnej jest bardziej symetryczny.

Badanie skośności rozkładu WBC w obu grupach przy użyciu testu D'Agostino:

H_0 : Rozkład zmiennej jest symetryczny (skośność jest równa 0).

H_1 : Rozkład zmiennej jest skośny (skośność jest różna od 0).

	Grupa badana	Grupa kontrolna
Skośność	0,46697	-0,012812
Statystyka testowa	1,12843	-0,022267
P-wartość	0,2591	0,9822

Tabela 25. Wyniki testu skośności zmiennej WBC.

W grupie badanej skośność rozkładu danych jest dodatnia, natomiast w grupie kontrolnej jest ujemnie bliska zeru, ale zaobserwowane skośności nie są istotnie różne od zera. To oznacza, że rozkład WBC w obu grupach są symetryczne.

Badanie normalności rozkładu WBC w obu grupach przy użyciu testu Shapiro-Wilka:

H_0 : Rozkład zmiennej jest normalny.

H_1 : Rozkład zmiennej nie jest normalny.

	Grupa badana	Grupa kontrolna
Statystyka testowa	0,97294	0,93639
P-wartość	0,7005	0,5444

Tabela 26. Wyniki testu normalności zmiennej WBC.

Zarówno w grupie badanej jak i w grupie kontrolnej p-wartość testu jest powyżej 0,05, co oznacza, że można założyć, że rozkłady WBC w tych grupach są zbliżone do rozkładu normalnego.

Ponieważ wartość odstająca nie zaburza normalności rozkładu WBC w grupie badanej, użyto testu t-Studenta dla prób niezależnych poprawką na ciągłość z następującymi hipotezami:

H_0 : Wartości WBC są równe w grupie badanej i kontrolnej.

H_1 : Wartości WBC są różne w grupie badanej i kontrolnej.

Statystyka testowa $t = 2,987$, p-wartość = 0,007362 < 0,05

Wniosek: Odrzucamy H_0 na rzecz H_1 . Wartości WBC są różne w grupie badanej i kontrolnej.

6. Albumina

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
0,000	3,300	3,700	3,197	4,275	4,900

Tabela 27. Podstawowe statystyki albuminy w grupie badanej.

Najmniejsza wartość albuminy w grupie badanej jest równa 0, co może oznaczać, że u niektórych pacjentów nie zmierzono tego wskaźnika. Takich pacjentów jest pięciu o numerach (nr) 31, 35, 38, 43, 44. Po wykluczeniu tych pacjentów (ci pacjenci są wykluczeni w analizie tej zmiennej), podstawowe statystyki albuminy w grupie badanej wyglądają następująco:

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
2,900	3,600	3,900	3,959	4,400	4,900

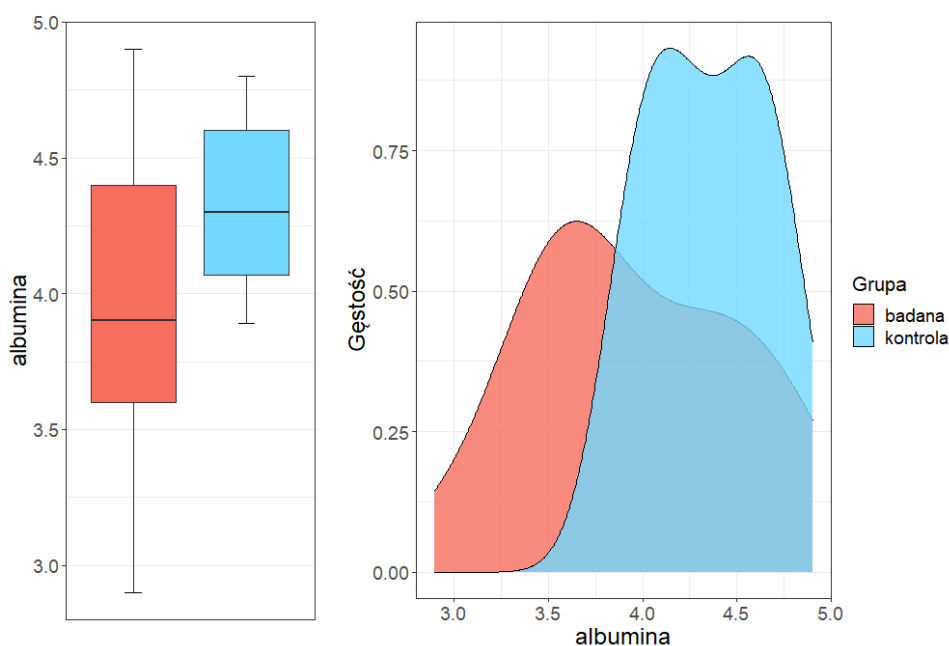
Tabela 28. Podstawowe statystyki albuminy w grupie badanej.

Albumina w grupie badanej przyjmuje wartości od 2,9 do 4,9, a jej średnia wynosi 3,959. Dane mogą być zbliżone do rozkładu normalnego, ponieważ mediana jest bliska średniej.

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
3,890	4,070	4,300	4,342	4,600	4,800

Tabela 29. Podstawowe statystyki albuminy w grupie kontrolnej.

Albumina w grupie kontrolnej przyjmuje wartości od 3,89 do 4,8 i jest to mniejsza rozbieżność niż w przypadku grupy badanej. Jej średnia wynosi 4,342 i jest większa niż w grupie badanej. Dane mogą być zbliżone do rozkładu normalnego, ponieważ mediana jest bliska średniej. Nie występują braki danych.



Wykresy 10. Rozkład albuminy w grupie badanej i kontrolnej po usunięciu wartości zerowych w grupie badanej.

W grupie badanej i w grupie kontrolnej nie występują wartości odstające. Wykresy gęstości dla obu grup przypominają rozkład normalny.

Badanie skośności rozkładu albuminy w obu grupach przy użyciu testu D'Agostino:

H_0 : Rozkład zmiennej jest symetryczny (skośność jest równa 0).

H_1 : Rozkład zmiennej jest skośny (skośność jest różna od 0).

	Grupa badana	Grupa kontrolna
Skośność	0,11251	0,0089256
Statystyka testowa	0,25710	0,0155124
P-wartość	0,7971	0,9876

Tabela 30. Wyniki testu skośności zmiennej albumina.

W grupie badanej i kontrolnej występują dodatnie skośności, ale te skośności nie są istotnie różne od zera. Rozkład albuminy w obu grupach jest symetryczny.

Badanie normalności rozkładu albuminy w obu grupach przy użyciu testu Shapiro-Wilka:

H_0 : Rozkład zmiennej jest normalny.

H_1 : Rozkład zmiennej nie jest normalny.

	Grupa badana	Grupa kontrolna
Statystyka testowa	0,96336	0,95319
P-wartość	0,5862	0,7251

Tabela 31. Wyniki testu normalności zmiennej albumina.

Zarówno w grupie badanej jak i w grupie kontrolnej p-wartość testu jest powyżej 0,05, co oznacza, że można założyć, że rozkład albuminy w tych grupach są zbliżone do rozkładu normalnego.

Ponieważ są spełnione założenia o normalności rozkładu, użyto testu t-Studenta dla prób niezależnych poprawką na ciągłość, z następującymi hipotezami:

H_0 : Wartości albuminy są równe w grupie badanej i kontrolnej.
 H_1 : Wartości albuminy są różne w grupie badanej i kontrolnej.

Statystyka testowa $t = -2,3616$, $p\text{-wartość} = 0,02615 < 0.05$

Wniosek: Odrzucamy H_0 na rzecz H_1 . Wartości albuminy są różne w grupie badanej i kontrolnej.

7. Kreatynina

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
0,5500	0,9325	1,3400	1,5304	1,9550	3,8000

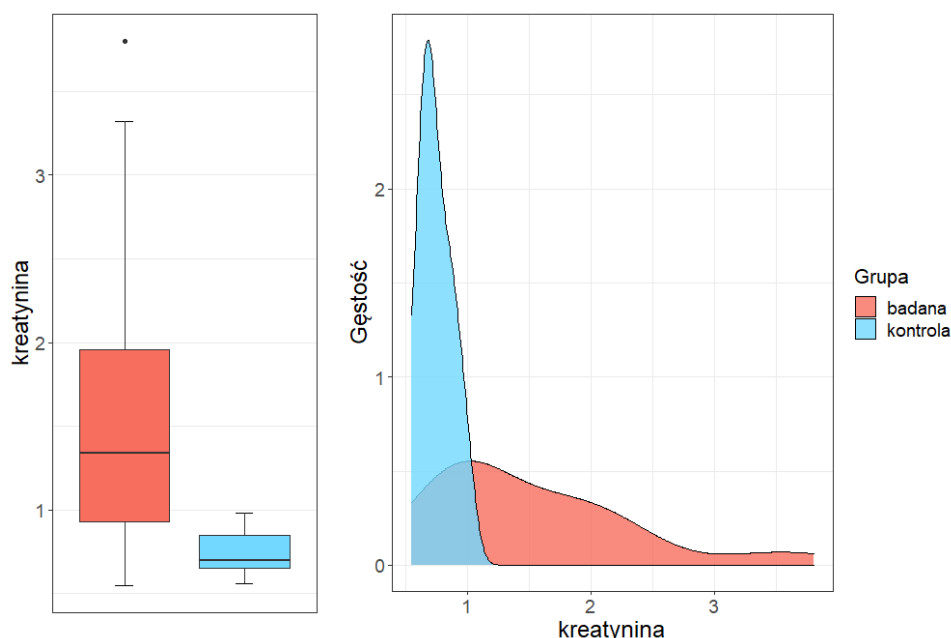
Tabela 32. Podstawowe statystyki kreatyniny w grupie badanej.

Kreatynina w grupie badanej przyjmuje wartości od 0,55 do 3,8, a jej średnia wynosi 1,5304. Dane mogą być prawostronnie skośne, ponieważ średnia ma większą wartość niż mediana. Nie występują braki danych.

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
0,5600	0,6500	0,7000	0,7433	0,8500	0,9800

Tabela 33. Podstawowe statystyki kreatyniny w grupie kontrolnej.

Kreatynina w grupie kontrolnej przyjmuje wartości od 0,56 do 0,98 i jest to mniejsza rozbieżność niż w przypadku grupy badanej. Jej średnia wynosi 0,7433 i jest mniejsza niż w grupie badanej. Dane mogą być prawostronnie skośne, ponieważ średnia jest większa niż mediana. Nie występują braki danych.



Wykresy 11. Rozkład kreatyniny w grupie badanej i kontrolnej.

W grupie badanej występuje wartość odstająca. Jest ona równa 3,8 i należy do pacjenta o numerze (nr) 37. Z wykresów można zauważyć, że rozkład kreatyniny w grupie kontrolnej jest zbliżony do rozkładu normalnego. W przypadku grupy badanej, rozkład jest prawostronnie skośny.

Badanie skośności rozkładu kreatyniny w obu grupach przy użyciu testu D'Agostino:*H₀: Rozkład zmiennej jest symetryczny (skośność jest równa 0).**H₁: Rozkład zmiennej jest skośny (skośność jest różna od 0).*

	Grupa badana	Grupa kontrolna
Skośność	1,1924	0,46311
Statystyka testowa	2,5937	0,80016
P-wartość	0,009495	0,4236

Tabela 34. Wyniki testu skośności zmiennej kreatyniny.

W grupie badanej występuje skośność istotnie różna od zera, natomiast w grupie kontrolnej skośność nie jest istotnie różna od zera. Skośność w grupie badanej jest dodatnia, co oznacza, że rozkład kreatyniny jest prawostronnie skośny. Rozkład zmiennej kreatyniny w grupie kontrolnej jest symetryczny.

Badanie normalności rozkładu kreatyniny w obu grupach przy użyciu testu Shapiro-Wilka:*H₀: Rozkład zmiennej jest normalny.**H₁: Rozkład zmiennej nie jest normalny.*

	Grupa badana	Grupa kontrolna
Statystyka testowa	0,88522	0,95473
P-wartość	0,007436	0,7419

Tabela 35. Wyniki testu normalności zmiennej kreatyniny.

W grupie badanej p-wartość jest mniejsza od 0,05, co oznacza, że rozkład kreatyniny w tej grupie nie pochodzi z rozkładu normalnego, natomiast w grupie kontrolnej p-wartość jest większa od 0,05, więc można założyć, że rozkład kreatyniny w grupie kontrolnej jest zbliżony do rozkładu normalnego.

Ponieważ nie można założyć, że rozkład kreatyniny w grupie badanej jest zbliżony do rozkładu normalnego ze względu na małą liczebność próbki oraz występującą wartość odstającą, użyto testu nieparametrycznego **U Manna-Whitneya z poprawką na ciągłość** do przetestowania następujących hipotez:

*H₀: Wartości kreatyniny są równe w grupie badanej i kontrolnej.**H₁: Wartości kreatyniny są różne w grupie badanej i kontrolnej.*Statystyka testowa **W** = 208, **p-wartość** = 0,0006351 < 0,05

Po przeprowadzeniu testu w R otrzymano ostrzeżenie, że nie można obliczyć dokładnej p-wartości przy użyciu testu za pomocą funkcji *wilcox.test* dla prób niezależnych z powodu występowania remisów (ties) w danych.

Wniosek: Odrzucamy H₀ na rzecz H₁. Wartości kreatyniny są różne w grupie badanej i kontrolnej.

8. UA

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
0,000	5,710	7,100	6,782	8,502	11,600

Tabela 36. Podstawowe statystyki UA w grupie badanej.

Najmniejsza wartość UA w grupie badanej jest równa 0, co może oznaczać, że u niektórych pacjentów nie zmierzono tego wskaźnika. Takich pacjentów jest dwóch o numerach (nr) 35, 39. Po wykluczeniu tych pacjentów (także w dalszej analizie tej zmiennej), podstawowe statystyki UA w grupie badanej wyglądają następująco:

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
3,350	6,400	7,250	7,348	8,625	11,600

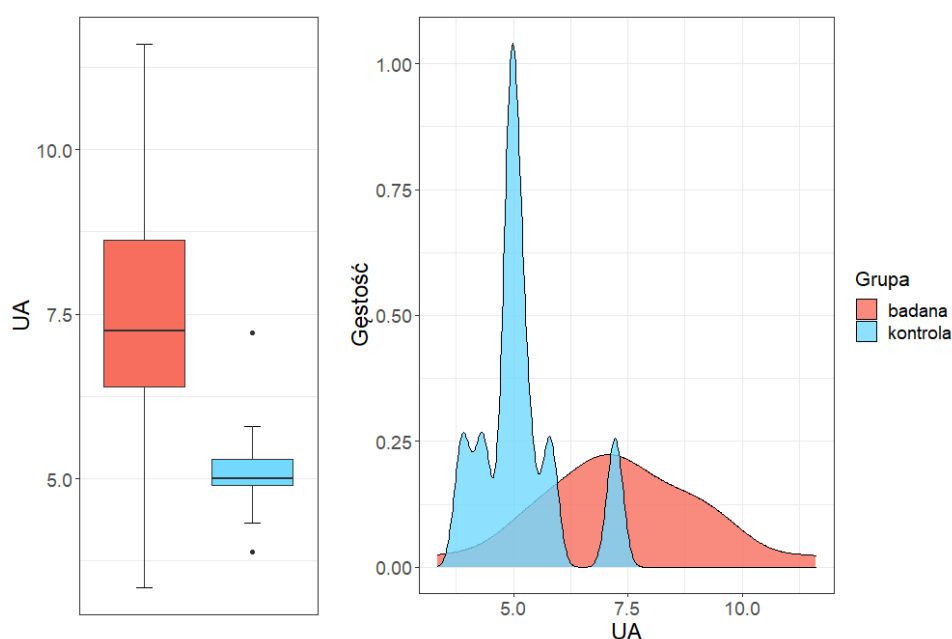
Tabela 37. Podstawowe statystyki UA w grupie badanej.

UA w grupie badanej przyjmuje wartości od 3,35 do 11,6, a jej średnia wynosi 7,348. Dane mogą być prawostronnie skośne, ponieważ mediana jest mniejsza od średniej.

Minimum	Pierwszy kwantyl	Mediana	Średnia	Trzeci kwantyl	Maksimum
3,890	4,900	5,000	5,158	5,300	7,220

Tabela 38. Podstawowe statystyki UA w grupie kontrolnej.

UA w grupie kontrolnej przyjmuje wartości od 3,89 do 7,22 i jest to mniejsza rozbieżność niż w przypadku grupy badanej. Jej średnia wynosi 5,158 i jest mniejsza niż w grupie badanej. Dane mogą być prawostronnie skośne, ponieważ średnia jest większa niż mediana. Nie występują braki danych.



Wykresy 12. Rozkład UA w grupie badanej i kontrolnej po usunięciu wartości odstających z grupy badanej.

W grupie kontrolnej występują dwie wartości odstające: 3,89 oraz 7,22, które należą do pacjentów o numerze (nr) 10 oraz 9. Rozkład UA w grupie badanej jest zbliżony do rozkładu normalnego, natomiast rozkład w grupie kontrolnej przypomina rozkład multimodalny.

Badanie skośności rozkładu UA w obu grupach przy użyciu testu D'Agostino:

H_0 : Rozkład zmiennej jest symetryczny (skośność jest równa 0).

H_1 : Rozkład zmiennej jest skośny (skośność jest różna od 0).

	Grupa badana	Grupa kontrolna
Skośność	0,13258	1,0098
Statystyka testowa	0,31811	1,7087
P-wartość	0,7504	0,0875

Tabela 39. Wyniki testu skośności zmiennej UA.

W grupie badanej i kontrolnej występują dodatnie skośności, ale te skośności nie są istotnie różne od 0. Rozkład UA w obu grupach jest symetryczny.

Badanie normalności rozkładu UA w obu grupach przy użyciu testu Shapiro-Wilka:

H_0 : Rozkład zmiennej jest normalny.

H_1 : Rozkład zmiennej nie jest normalny.

	Grupa badana	Grupa kontrolna
Statystyka testowa	0,98751	0,88986
P-wartość	0,9867	0,1989

Tabela 40. Wyniki testu normalności zmiennej UA.

Ponieważ wartości odstające nie zaburzają normalności rozkładu UA w grupie kontrolnej, użyto testu **t-Studenta dla prób niezależnych poprawką na ciągłość** z następującymi hipotezami:

H_0 : Wartości UA są równe w grupie badanej i kontrolnej.

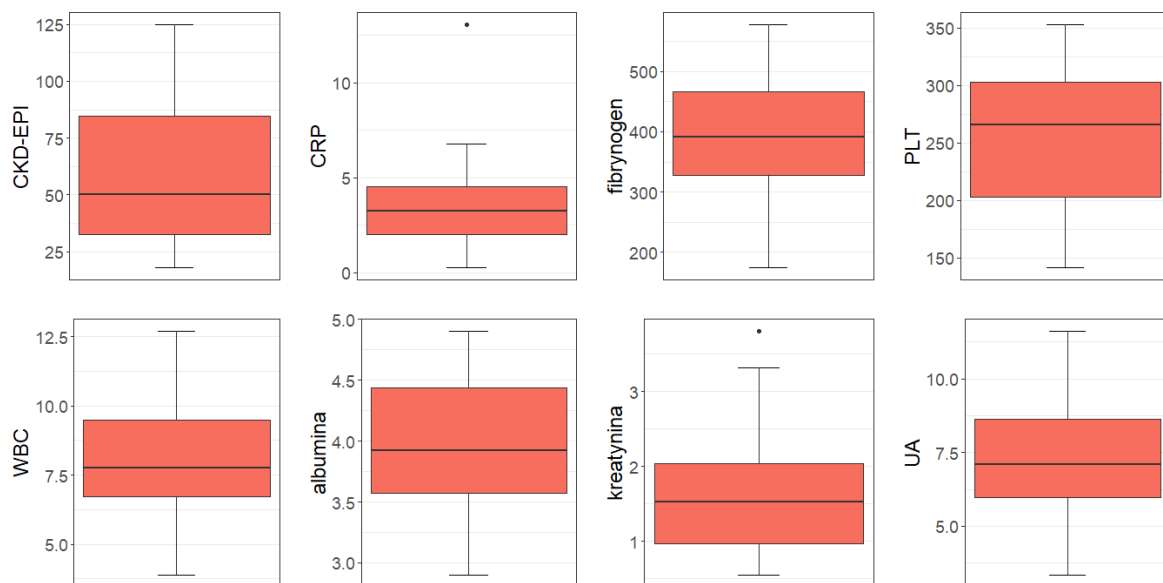
H_1 : Wartości UA są różne w grupie badanej i kontrolnej.

Statystyka testowa $t = 4,5343$, **p-wartość** = 0,0001057 < 0,05

Wniosek: Odrzucamy H_0 na rzecz H_1 . Wartości UA są różne w grupie badanej i kontrolnej.

ANALIZA ZALEŻNOŚCI MIĘDZY BADANYMI ZMIENNYMI W GRUPIE BADANEJ

W grupie badanej w zmiennych albumina oraz UA występują zerowe wartości, które mogą oznaczać, że nie zostały zapisane wartości tych parametrów u niektórych pacjentów. Dlatego na potrzeby przeprowadzenia analizy zależności pomiędzy niektórymi zmiennymi, wydzielono podzbiór pacjentów, którzy mają zapisane wszędzie niezerowe wartości interesujących nas zmiennych. W tym podziorze danych wykluczono 6 pacjentów o numerach (nr) 31, 35, 38, 39, 43, 44. Nowy zbiór danych zawiera informacje o 20 pacjentach, takie jak CKD-EPI, CRP, fibrynogen, PLT, WBC, albumina, kreatynina oraz UA.



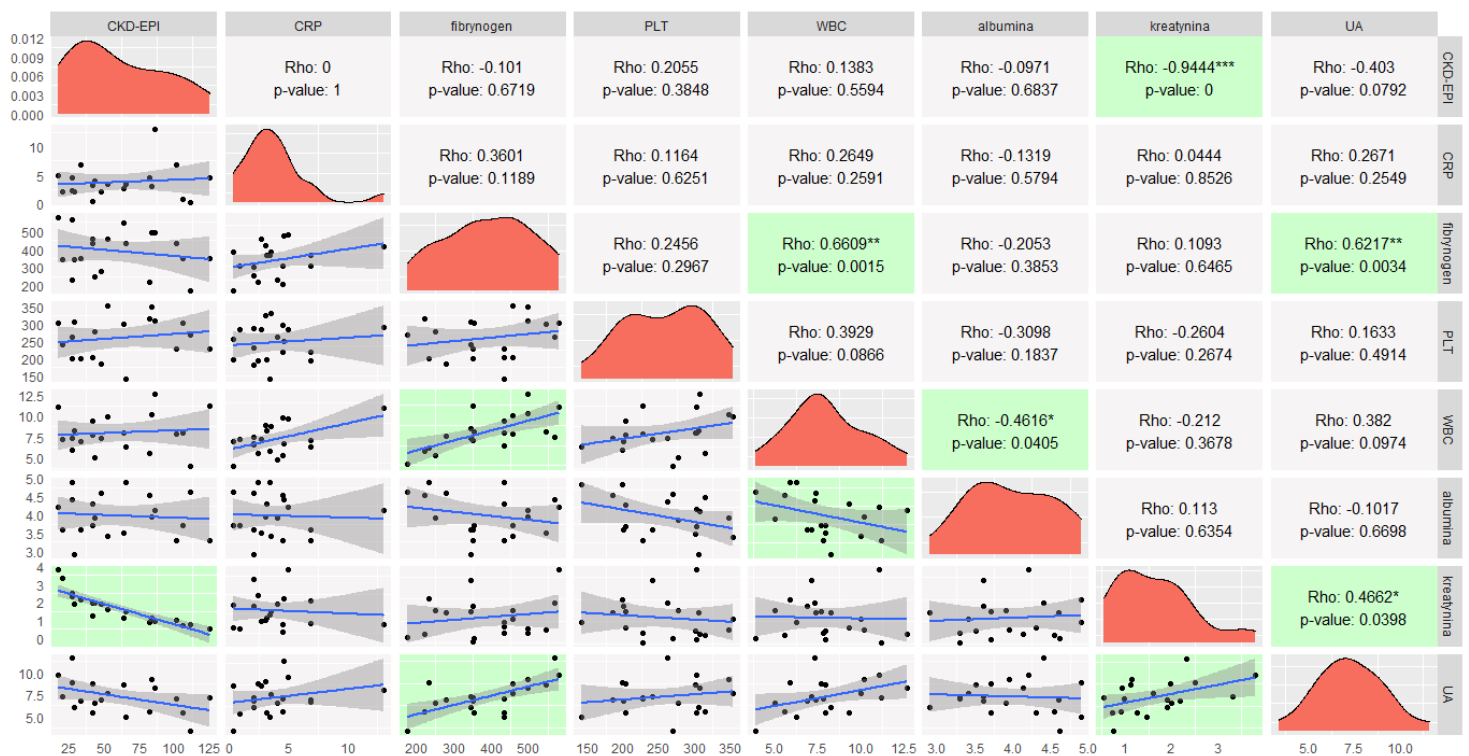
Wykresy 13. Rozkłady CKD-EPI, CRP, fibrynogenu, PLT, WBC, albuminy, kreatyniny oraz UA w grupie badanej po usunięciu pacjentów, którzy mieli w zmiennej albumina lub UA zerowe wartości.

W nowym zbiorze danych grupy badanej występują obserwacje odstające w zmiennej CRP oraz kreatynina, które mogą wpłynąć na wyznaczenie korelacji, dlatego do szukania zależności pomiędzy 8 zmiennymi posłuży **test korelacji Spearmana** (wyniki testu zostały zaokrąglone do czterech miejsc po przecinku) z następującymi hipotezami:

H_0 : Korelacja pomiędzy dwoma zmiennymi jest równa zero (nie ma liniowej zależności pomiędzy zmiennymi) w grupie badanej.

H_1 : Korelacja pomiędzy dwoma zmiennymi jest różna od zera (zmienne są liniowo zależne) w grupie badanej.

Po przeprowadzeniu testu w R otrzymano ostrzeżenie, że nie można obliczyć dokładnej p-wartości przy użyciu testu za pomocą funkcji `cor.test` metodą Spearmana z powodu występowania remisów (ties) w danych.



Wykresy 14. Zależności między zmiennymi.

Między pięcioma parami zmiennych występują korelacje istotnie różne od zera:

- dodatnie korelacje (im wyższy był pomiar jednej zmiennej, tym wyższy był pomiar drugiej zmiennej):
 - 1) WBC i fibrynogen: 0,6609;
 - 2) UA i fibrynogen: 0,6217;
 - 3) UA i kreatynina: 0,4662;
- ujemne korelacje (im niższy był pomiar jednej zmiennej, tym wyższy był pomiar drugiej zmiennej):
 - 1) albumina i WBC: -0,4616;
 - 2) kreatynina i CKD-EPI: -0,403.