

Badanie statystyczne wpływu różnych czynników na poziom samopoczucia

Marcin Dziadosz, Alicja Hołowiecka, Matylda Jankowska

Spis treści

1	Wstęp	2
2	Opis danych	2
2.1	Zbieranie danych, opis badanej grupy	2
2.2	Import i oczyszczanie danych	2
3	Hipotezy badawcze oraz ich weryfikacja	5
3.1	Hipoteza I	5
3.2	Hipoteza II	6
3.3	Hipoteza III	8
3.4	Hipoteza IV	10
3.5	Hipoteza V	11
3.6	Hipoteza VI	12
3.7	Hipoteza VII	13
3.8	Hipoteza VIII	14
3.9	Hipoteza IX	16
3.10	Hipoteza X	17
4	Wnioski	17
5	Bibliografia	17

1 Wstęp

Celem naszego projektu jest analiza statystyczna wpływu różnych czynników na poziom samopoczucia. Zbadamy wpływ na odczuwanie poszczególnych emocji czynników z pięciu grup: dotyczących snu, diety, aktywności fizycznej, kontaktów towarzyskich oraz używania urządzeń elektronicznych. Sprawdzimy, czy średnie samopoczucie różni się pomiędzy płciami oraz pomiędzy różnymi grupami wiekowymi.

2 Opis danych

(Import, obróbka, wstępna analiza, testy na normalność)

2.1 Zbieranie danych, opis badanej grupy

Dane zebraliśmy na podstawie ankiety. Pytania dotyczące emocji pochodzą z artykułu “Skale do Pomiaru Nastroju i Sześciu Emocji” autorstwa Bogdana Wojciszke i Wiesława Baryły. Pozostałe pytania są naszego autorstwa. Ankietę stworzyliśmy za pomocą Formularzy Google i rozesłaliśmy respondentom drogą internetową.

Przebadana przez nas grupa składała się z 240 osób. Spośród nich 48.3% było kobietami (a zatem 51.7% było mężczyznami, bo nie daliśmy do wyboru opcji typu “nie wiem, jakiej płci jestem”). Większość osób (trochę ponad połowa) było w wieku co najwyżej 23 lat (w tym aż 30% w wieku 21 lat).

Pierwsze 10 pytań były to zdania dotyczące aktualnego samopoczucia. Należało ocenić w skali od 1 do 5, jak bardzo zgadzasz się z danym twierdzeniem. Po nich następowały 24 uczucia takie jak radość, złość, depresja itp. Należało ocenić w skali od 1 do 7 (od “nigdy” do “zawsze”) jak często odczuwasz daną emocję.

Potem pojawiały się pytania o czynniki, których wpływ na samopoczucie chcemy zbadać. Każdego z pięciu poniższych czynników dotyczyły trzy pytania:

- sen: liczba godzin przespanych w ciągu doby (do wyboru z 5 odpowiedzi), posiadanie wyznaczonych godzin snu (ocena w skali od 1 do 5), jakość snu (ocena w skali od 1 do 5)
- aktywność fizyczna: ilość godzin poświęcana tygodniowo na sport (do wyboru z 5 odpowiedzi), najczęściej wybierany środek transportu (do wyboru z 5 odpowiedzi, w tym możliwość podania własnej odpowiedzi przy opcji “inne”), kondycja fizyczna (ocena w skali od 1 do 5)
- używanie urządzeń elektronicznych: ilość godzin spędzana w ciągu dnia na korzystaniu z urządzeń elektronicznych w celu pracy/nauki (do wyboru z 5 odpowiedzi), ilość godzin spędzana w ciągu dnia na korzystaniu z urządzeń elektronicznych w celu rozrywki (do wyboru z 5 odpowiedzi), poczucie uzależnienia od urządzeń elektronicznych (ocena w skali od 1 do 5)
- dieta: przeważający składnik w diecie (do wyboru z 5 odpowiedzi), znaczenie jakości posiłków (do wyboru z 5 odpowiedzi), regularność posiłków (do wyboru z 5 odpowiedzi)
- kontakty towarzyskie: posiadanie bliskich osób, na których można polegać (do wyboru z 5 odpowiedzi), liczba bliższych znajomych (do wyboru z 5 odpowiedzi), znaczenie kontaktów z innymi ludźmi (ocena w skali od 1 do 5)

Jak widać, wszystkie te czynniki były oceniane w skali od 1 do 5, a czynniki dotyczące samopoczucia bądź w skali od 1 do 7, bądź w skali od 1 do 5. Zbadamy zarówno zależności pomiędzy pojedynczymi czynnikami, jak również pomiędzy zsumowanymi, na przykład Ogólna ocena samopoczucia.

2.2 Import i oczyszczanie danych

Do R zaimportowaliśmy dane z pliku `.xlsx`. Przy tworzeniu ankiety zastrzeżliśmy, że na każde pytanie trzeba odpowiedzieć, dlatego nie mieliśmy problemu z brakami danych. W kolumnie `wiek` mieliśmy kilka nieprawidłowo wprowadzonych danych, większość osób podawała wiek jako liczbę, ale niektórzy pisali np. “21 lat”. Poprawiliśmy to i zmieniliśmy typ kolumny na numeryczny. Musieliśmy też poprawić kilka odpowiedzi na

pytanie o jakość posiłków, ponieważ zmieniliśmy jedną z możliwych odpowiedzi w ankiecie już po rozpoczęciu zbierania danych.

Wszystkim kolumnom zmieniliśmy nazwy, tak aby pozbyć się polskich znaków i odstępów, oraz żeby ułatwić sobie pracę z danymi.

Typ danych w kolumnach, w których możliwych było 5 odpowiedzi na pytanie, zmieniliśmy na **factor**, i uszeregowaliśmy od odpowiedzi, która według nas w danym pytaniu była najgorsza, do najlepszej.

W pytaniach, gdzie należało oceniać czynnik w skali od 1 do 5 lub od 1 do 7, także zmieniliśmy typ na **factor**.

Dodaliśmy także kolumnę z Ogólną oceną samopoczucia, w której zsumowaliśmy odpowiedzi z pytań o emocje - pozytywne uczucia sumowaliśmy z plusem, a negatywne z minusem.

Te działania zostały wykonane w skrypcie `oczyszczanie_danych.R`, który załączymy do niniejszego raportu. Na koniec zapisaliśmy dane do pliku `dane.RData`.

Cecha	Minimum	Mediana	Maksimum	Czy rozkład jest normalny?
Cecha	14	21	59	TAK
wiek	1	2	5	TAK
zly_humor	1	3	5	TAK
swietnie	1	2	5	TAK
zly_nastroj	1	3	5	TAK
roz_spok	1	2	5	TAK
szaro_bezn	1	4	5	TAK
dobry_humor	1	4	5	TAK
pogodny	1	2	5	TAK
przygnebiony	1	2	5	TAK
zle_samop	1	4	5	TAK
dobre_samop	1	5	7	TAK
radosc	1	4	7	TAK
niepokoj	1	3	7	TAK
strach	1	5	7	TAK
milosc	1	4	7	TAK
gniew	2	5	7	TAK
wesolosc	1	2	7	TAK
upokorzenie	1	3	7	TAK
smutek	1	3	7	TAK
wzburzenie	1	3	7	TAK
lek	1	3	7	TAK
zal	1	3	7	TAK
obawa	1	3	7	TAK
zlosc	1	2	7	TAK
zalamanie	1	5	7	TAK
szczescie	1	3	7	TAK
wscieklosc	1	4	7	TAK
oddanie	1	5	7	TAK
zadowolenie	1	5	7	TAK
czulosc	1	2	7	TAK
wstyd	1	2	7	TAK
pocz_winy	1	5	7	TAK
przywiazanie	1	2	7	TAK
depresja	1	2	7	TAK
nieszczescie	1	3	5	TAK
liczba_przespanych_godzin	1	3	5	TAK
wyzn_pory_snu	1	3	5	TAK
jakosc_snu	1	3	5	TAK
godziny_sportu	1	3	5	TAK
srodek_transportu	1	3	5	TAK
kondycja_fizyczna	1	3	5	TAK
godziny_pracy_urzadzenia	1	3	5	TAK
godziny_rozrywki_urzadzenia	1	3	5	TAK
poczucie_uzaleznienia_urzadzenia	1	5	5	TAK
dieta	1	3	5	TAK
jakosc_posilkow	1	2	5	TAK
regularnosc_posilkow	1	5	5	TAK
posiadanie_zaufanej_osoby	1	4	5	TAK
liczba_osob_kontakty	1	4	5	TAK
znaczenie_kontaktow	-2	0.9	2	TAK
suma_samop1sr	-3.04166666666667	-	1.54166666666667	NIE
suma_samop2sr	-79	0.333333333333333	55	TAK

3 Hipotezy badawcze oraz ich weryfikacja

H_0 : nie istnieje zależność pomiędzy zmiennymi.

H_1 : istnieje zależność pomiędzy zmiennymi.

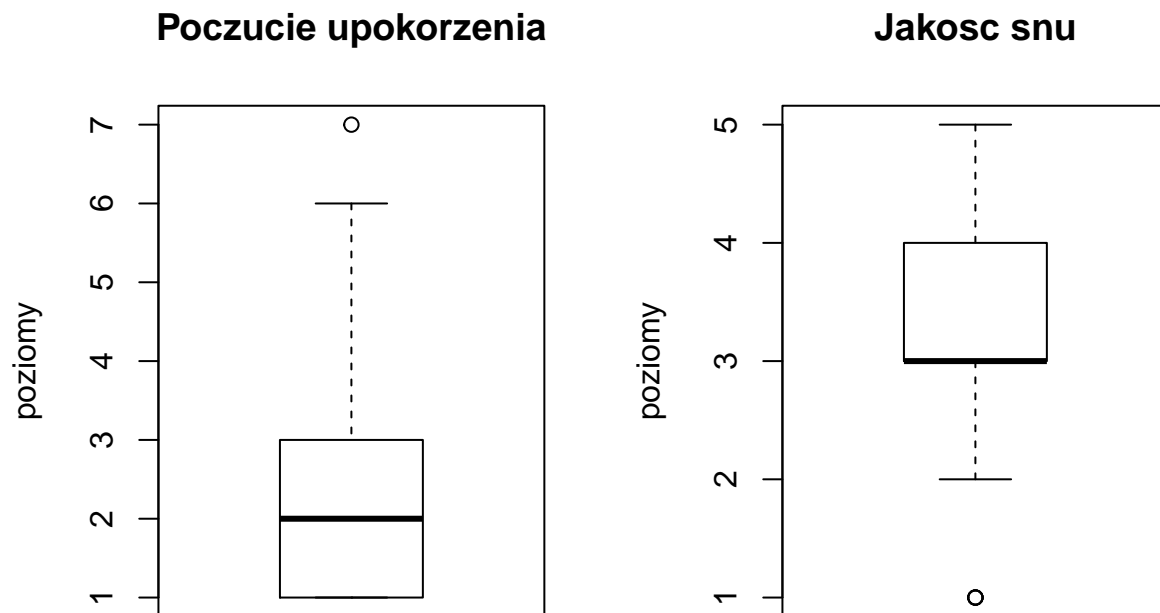
Hipotezy badane są na poziomie istotności $\alpha = 0.05$.

3.1 Hipoteza I

Badanie, czy istnieje zależność pomiędzy poczuciem upokorzenia a jakością snu.

P-value dla testu **rho-Spearmana** wynosi 0.0535368656, natomiast dla testu **tau-Kendalla** wynosi 0.0534696670. Nie ma podstaw do odrzucenia hipotezy H_0 , a więc nie istnieje istotna zależność pomiędzy poczuciem upokorzenia a jakością snu. Współczynniki korelacji dla testów **rho-Spearmana** oraz **tau-Kendalla** wynoszą odpowiednio -0.12478 oraz -0.10524, ale na poziomie istotności $\alpha = 0.05$ nie są istotne statystycznie.

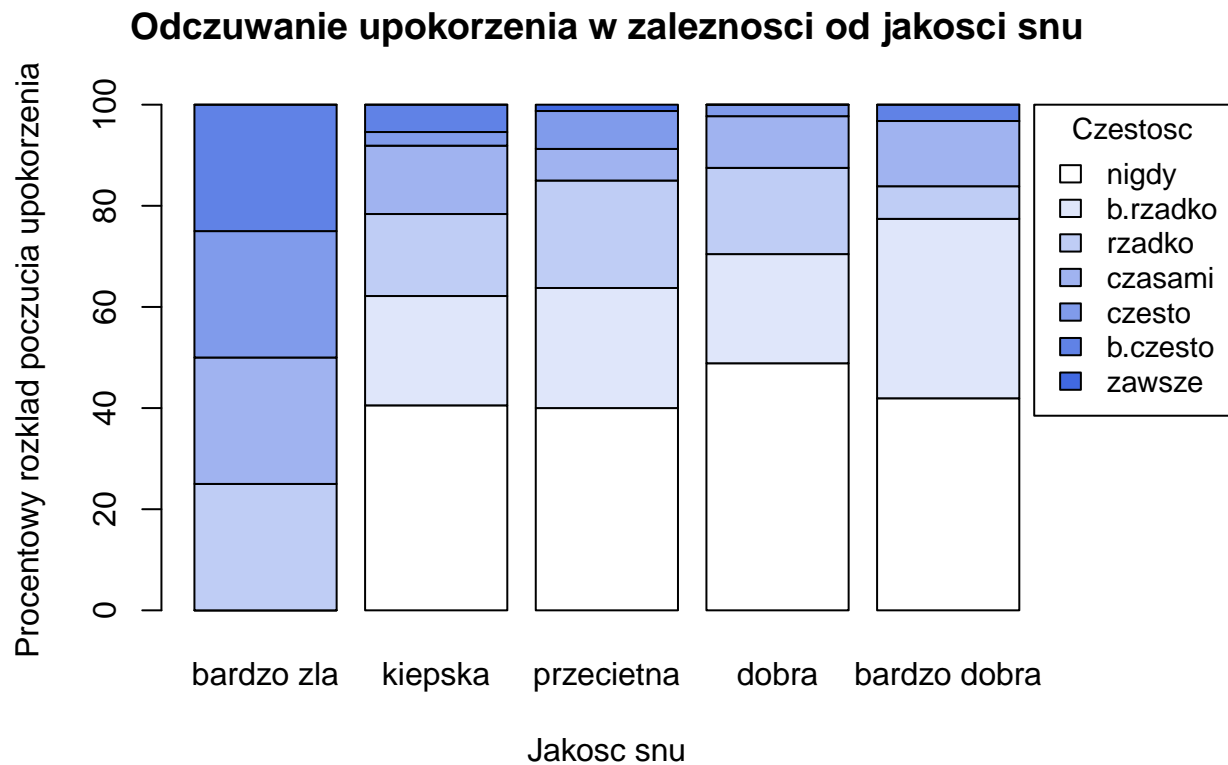
3.1.1 Wykresy boxplot



3.1.2 Tabela kontyngencji

Upokorzenie	Jakość snu				
	1	2	3	4	5
nigdy	0.0	40.5	40.0	48.9	41.9
bardzo rzadko	0.0	21.6	23.8	21.6	35.5
rzadko	25.0	16.2	21.2	17.0	6.5
czasami	25.0	13.5	6.2	10.2	12.9
często	25.0	2.7	7.5	2.3	0.0
bardzo często	25.0	5.4	0.0	0.0	3.2
zawsze	0.0	0.0	1.2	0.0	0.0

3.1.3 Wykresy słupkowe

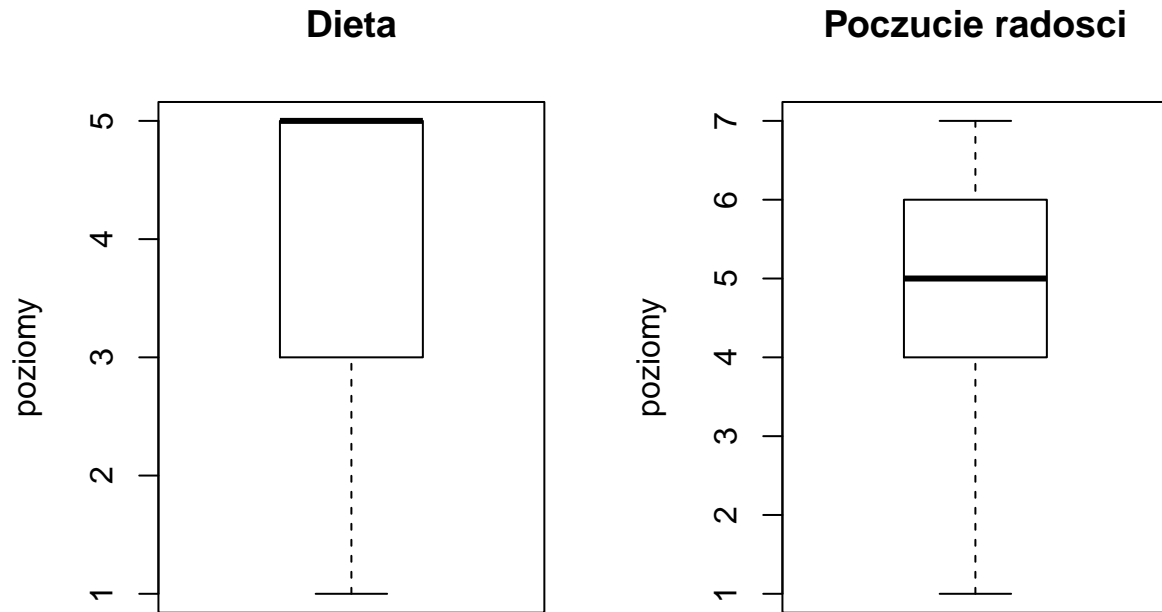


3.2 Hipoteza II

Badanie, czy istnieje zależność pomiędzy dietą a poczuciem radości.

P-value dla testu ρ -Spearmana wynosi 0.0839082132, natomiast dla testu τ -Kendalla wynosi 0.0839333604. Nie ma podstaw do odrzucenia H_0 . Nie istnieje istotna zależność pomiędzy dietą a poczuciem radości.

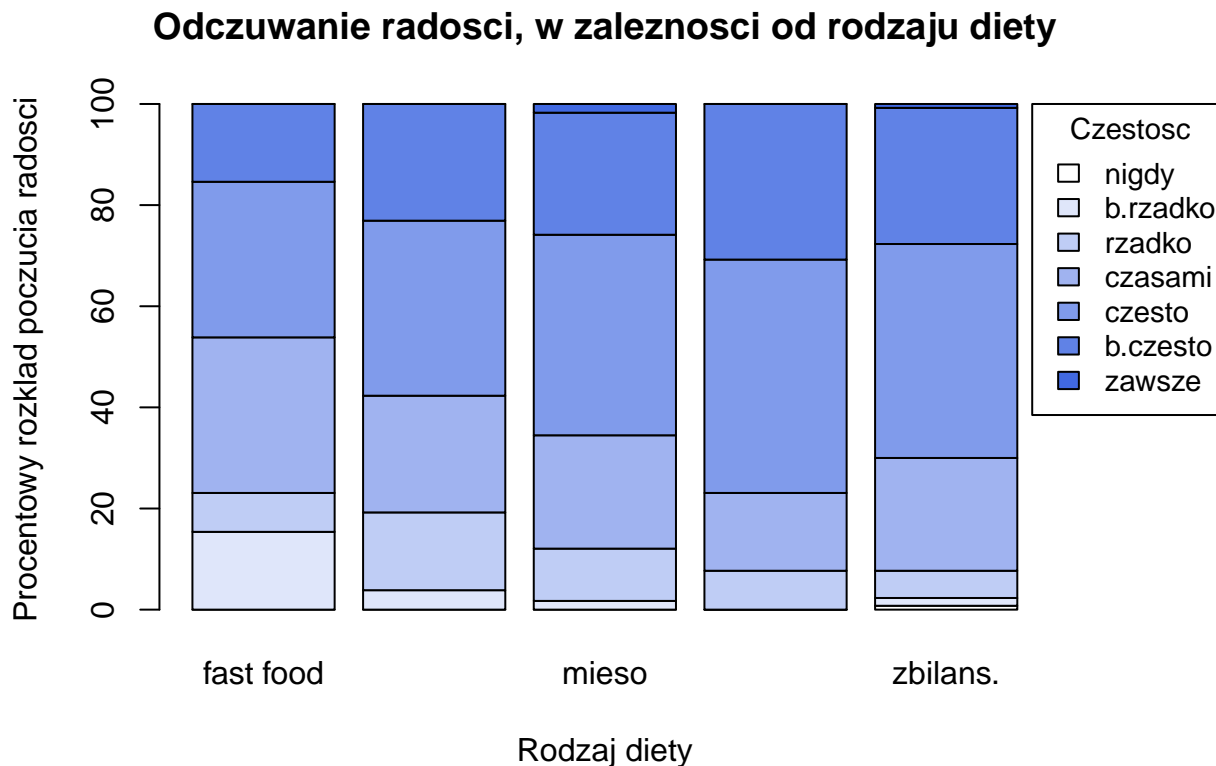
3.2.1 Wykresy boxplot



3.2.2 Tabela kontyngencji

Poczucie radości	Jakość diety				
	fast food	żywność mocno przetworzona (mrożonki, gotowe dania)	mięso i ciężkostrawne potrawy	warzywa i owoce	mam zbilansowaną dietę
nigdy	0.00	0.00	0.00	0.00	0.77
bardzo rzadko	15.38	3.85	1.72	0.00	1.54
rzadko	7.69	15.38	10.34	7.69	5.38
czasami	30.77	23.08	22.41	15.38	22.31
często	30.77	34.62	39.66	46.15	42.31
bardzo często	15.38	23.08	24.14	30.77	26.92
zawsze	0.00	0.00	1.72	0.00	0.77

3.2.3 Wykresy słupkowe



3.3 Hipoteza III

Badanie, czy istnieje zależność pomiędzy liczbą bliskich znajomych a odczuwaniem niepokoju.

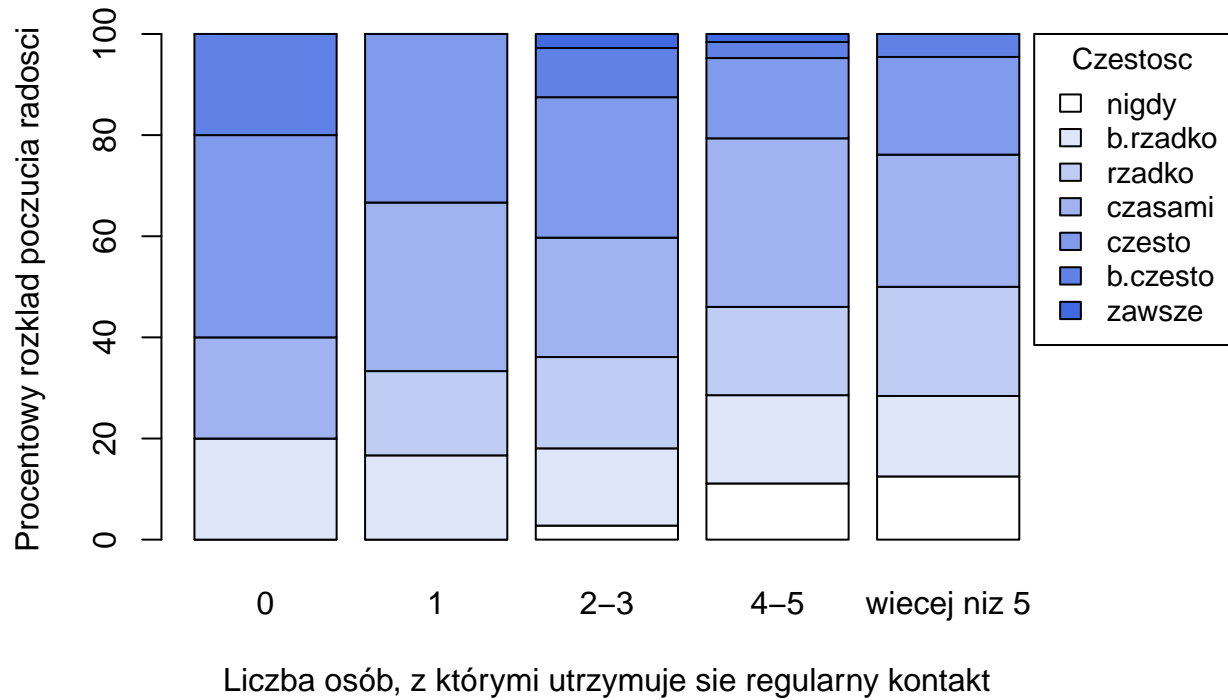
P-value dla testu ρ -Spearmana wynosi 0.0055186575, natomiast dla testu τ -Kendalla wynosi 0.0057809380. Należy odrzucić hipotezę H_0 , na rzecz hipotezy alternatywnej H_1 . Istnieje zależność pomiędzy liczbą osób, z którymi utrzymuje się regularny kontakt a odczuwaniem niepokoju.

3.3.1 Tabela kontyngencji

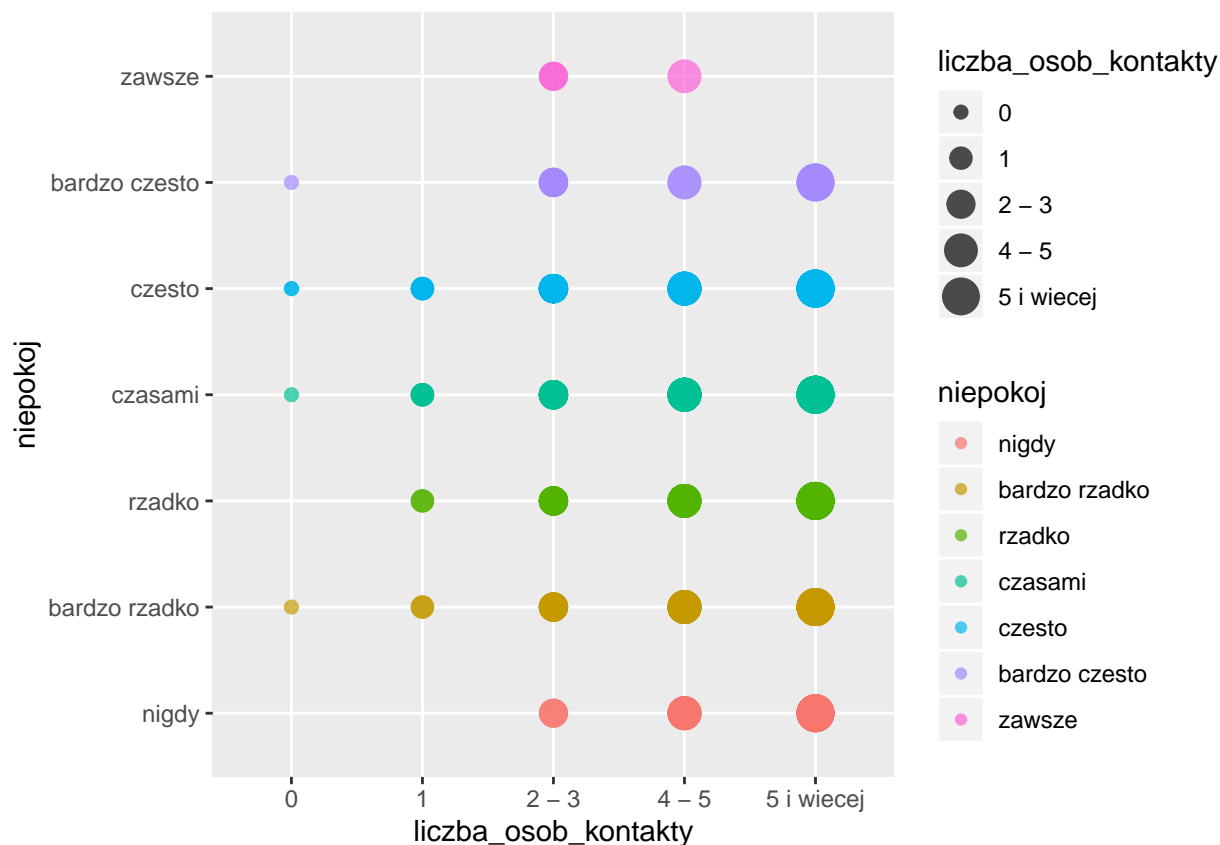
Niepokój	Liczba bliskich znajomych				
	0	1	2 - 3	4 - 5	5 i więcej
nigdy	0.0	0.0	2.8	11.1	12.5
bardzo rzadko	20.0	16.7	15.3	17.5	15.9
rzadko	0.0	16.7	18.1	17.5	21.6
czasami	20.0	33.3	23.6	33.3	26.1
często	40.0	33.3	27.8	15.9	19.3
bardzo często	20.0	0.0	9.7	3.2	4.5
zawsze	0.0	0.0	2.8	1.6	0.0

3.3.2 Wykresy słupkowe

Odczuwanie niepokoju, w zależności od liczby bliskich znajomych



Współczynniki korelacji dla testów rho-Spearmana i tau-Kendalla wynoszą odpowiednio -0.1786209 oraz -0.1473437. Interpretując te współczynniki oraz dane z wykresów, można stwierdzić, że wraz ze wzrostem liczby osób, z którymi utrzymuje się regularny kontakt, spada częstość odczuwania niepokoju.



Ten bubble chart jest na razie bez sensu, wrzucam tutaj żebyśmy mieli jak to zrobić... Nie bardzo rozumiem jak to ma być kolorowane ani od czego wielkość ma zależeć (A.)

3.4 Hipoteza IV

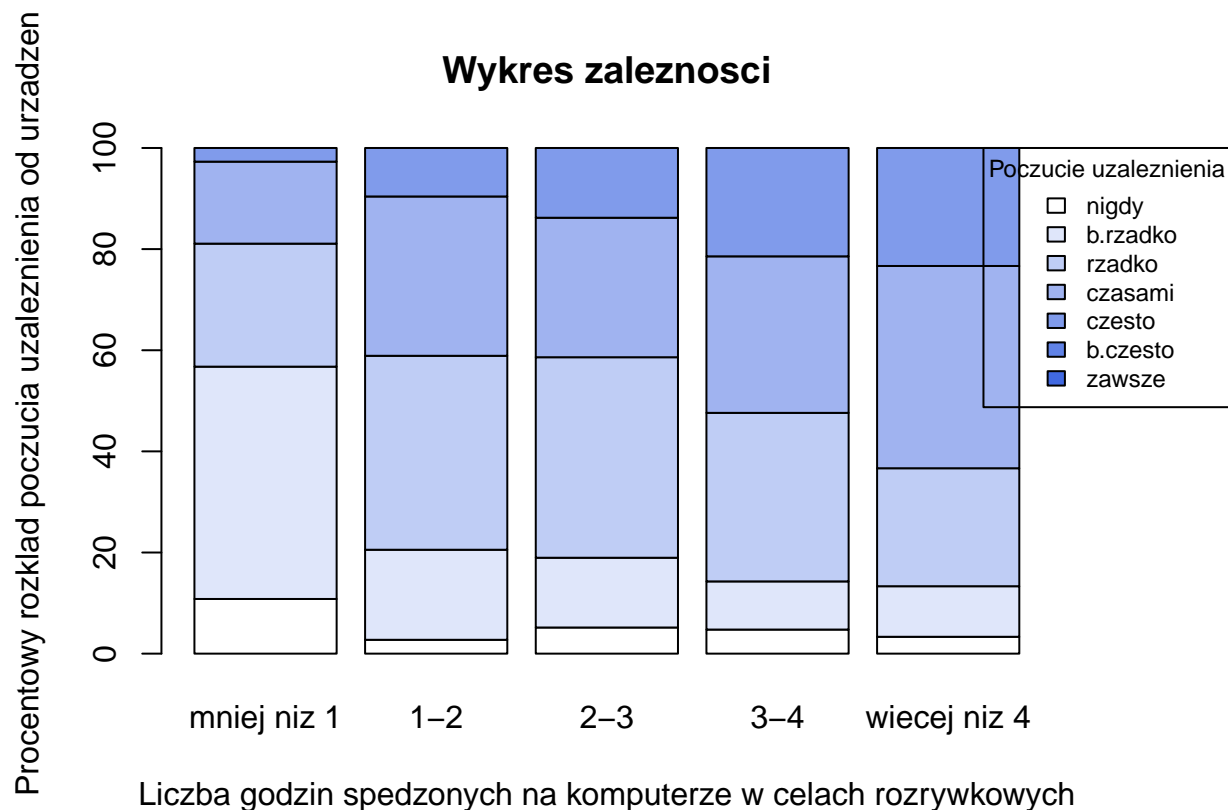
Badanie, czy istnieje zależność pomiędzy liczbą godzin przeznaczonych na korzystanie z urządzeń elektronicznych a poczuciem uzależnienia od urządzeń.

P-value dla testu **rho-Spearmana** wynosi 0.0000046081, natomiast dla testu **tau-Kendalla** wynosi 0.0000047417. Należy odrzucić hipotezę H_0 , na rzecz hipotezy alternatywnej H_1 . Istnieje zależność pomiędzy liczbą godzin spędzonych na używaniu urządzeń elektronicznych a poczuciem uzależnienia od nich.

3.4.1 Tabela kontyngencji

Poczucie uzależnienia	Godziny spędzone na używaniu urządzeń elektronicznych				
	mniej niż 1	1 - 2	2 - 3	3 - 4	więcej niż 4
nigdy	10.8	2.7	5.2	4.8	3.3
rzadko	45.9	17.8	13.8	9.5	10.0
czasami	24.3	38.4	39.7	33.3	23.3
często	16.2	31.5	27.6	31.0	40.0
bardzo często	2.7	9.6	13.8	21.4	23.3

3.4.2 Wykresy słupkowe



Współczynniki korelacji dla testów ρ -Spearmana i τ -Kendalla wynoszą odpowiednio 0.2908585 oraz 0.2426912. Wniosek: wraz ze wzrostem liczby godzin, spędzonych na korzystaniu dla rozrywki z urządzeń elektronicznych, wzrasta znacznie poczucie uzależnienia od tych urządzeń.

3.5 Hipoteza V

Badanie, czy istnieje zależność pomiędzy posiadaniem wyznaczonych pór snu a jego jakością.

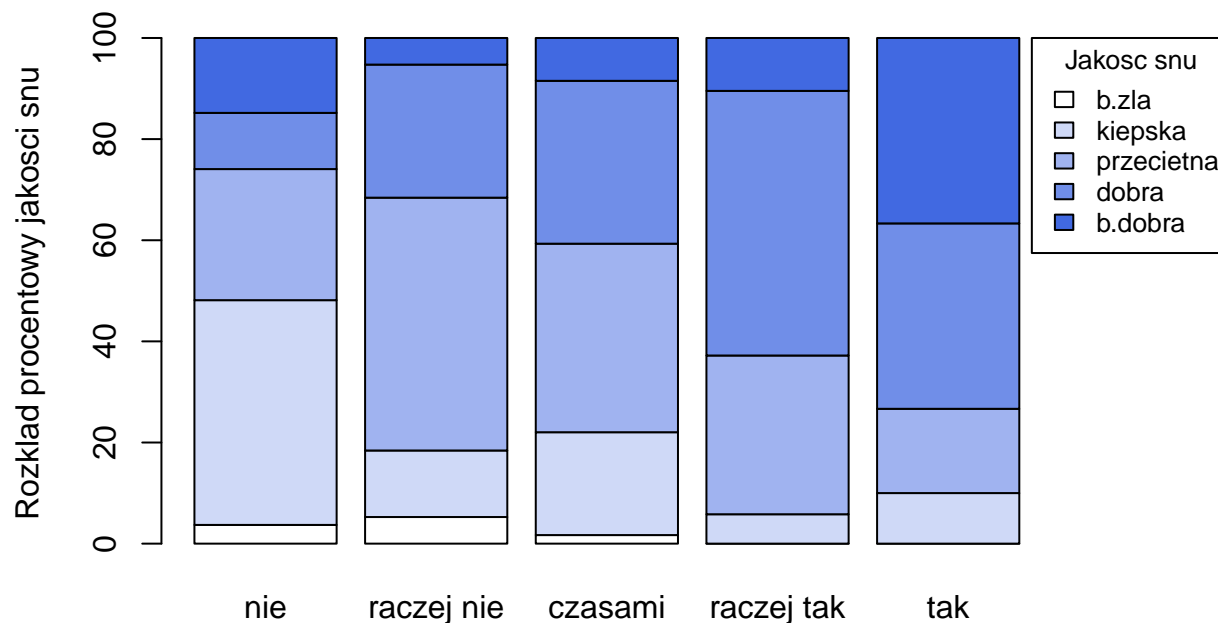
P-value dla testu ρ -Spearmana wynosi 0.0000000172, natomiast dla testu τ -Kendalla wynosi 0.0000000134. Należy odrzucić hipotezę H_0 , na rzecz hipotezy alternatywnej H_1 . Istnieje zależność pomiędzy posiadaniem wyznaczonych pór snu a jego jakością.

3.5.1 Tabela kontyngencji

Jakość snu	Posiadanie wyznaczonych pór snu				
	1	2	3	4	5
1	3.7	5.3	1.7	0.0	0.0
2	44.4	13.2	20.3	5.8	10.0
3	25.9	50.0	37.3	31.4	16.7
4	11.1	26.3	32.2	52.3	36.7
5	14.8	5.3	8.5	10.5	36.7

3.5.2 Wykresy słupkowe

ależność między jakością snu a posiadaniem jego wyznaczonych pór



Współczynniki korelacji dla testów **rho-Spearmana** i **tau-Kendalla** są dość duże i wynoszą odpowiednio 0.3539432 oraz 0.3064142. Im bardziej regularne są pory snu, tym lepsza jest jego jakość.

3.6 Hipoteza VI

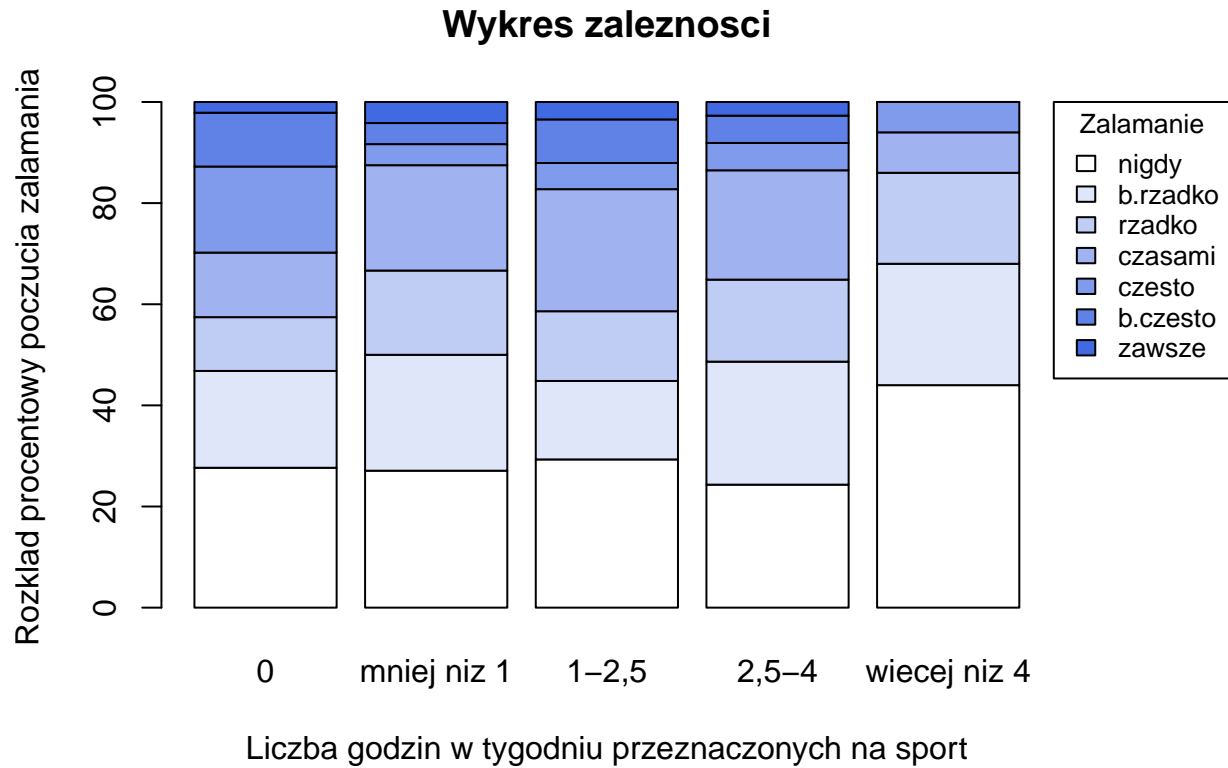
Badanie, czy istnieje zależność pomiędzy liczbą godzin przeznaczonych na uprawianie sportu a odczuwaniem załamania.

P-value dla testu **rho-Spearmana** wynosi 0.0114094634, natomiast dla testu **tau-Kendalla** wynosi 0.0107243185. Należy odrzucić hipotezę H_0 , na rzecz hipotezy alternatywnej H_1 . Istnieje zależność pomiędzy liczbą godzin przeznaczonych na uprawianie sportu a odczuwaniem załamania.

3.6.1 Tabela kontyngencji

Załamanie	Liczba godzin przeznaczonych na sport				
	nie uprawiam żadnego sportu	mniej niż 1	1 - 2,5	2,5 - 4	więcej niż 4
nigdy	27.7	27.1	29.3	24.3	44.0
bardzo rzadko	19.1	22.9	15.5	24.3	24.0
rzadko	10.6	16.7	13.8	16.2	18.0
czasami	12.8	20.8	24.1	21.6	8.0
często	17.0	4.2	5.2	5.4	6.0
bardzo często	10.6	4.2	8.6	5.4	0.0
zawsze	2.1	4.2	3.4	2.7	0.0

3.6.2 Wykresy słupkowe



Współczynniki korelacji dla testów rho-Spearmana i tau-Kendalla wynoszą odpowiednio -0.1630613 oraz -0.1318066. Im więcej godzin poświęca się tygodniowo na sport, tym rzadziej odczuwa się załamanie.

3.7 Hipoteza VII

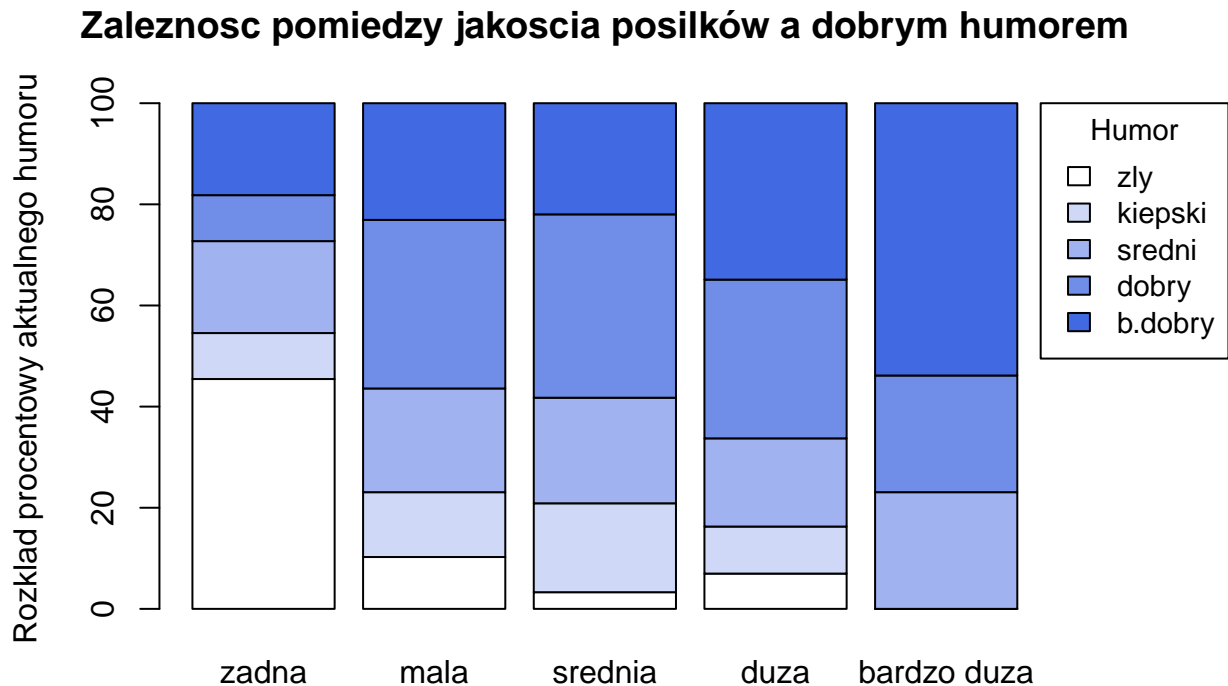
Badanie, czy istnieje zależność pomiędzy jakością posiłków a dobrym humorem.

P-value dla testu rho-Spearmana wynosi 0.0015012366, natomiast dla testu tau-Kendalla wynosi 0.0013687508. Należy odrzucić hipotezę H_0 , na rzecz hipotezy alternatywnej H_1 . Istnieje zależność pomiędzy jakością posiłków a dobrym humorem.

3.7.1 Tabela kontyngencji

Dobry humor	Jakość posiłków				
	wszystko mi jedno	nie jest to moim priorytetem	staram się jeść zdrowo, ale mi nie wychodzi	jest to dla mnie dość ważne	jestem tym, co jem!
nie zgadzam się	45.5	10.3	3.3	7.0	0.0
raczej się nie zgadzam	9.1	12.8	17.6	9.3	0.0
trochę tak, trochę nie	18.2	20.5	20.9	17.4	23.1
raczej się zgadzam	9.1	33.3	36.3	31.4	23.1
zgadzam się	18.2	23.1	22.0	34.9	53.8

3.7.2 Wykresy słupkowe



Zarówno współczynniki korelacji dla testów ρ -Spearmana i τ -Kendalla wynoszące odpowiednio 0.2038184 oraz 0.1738616, jak i wykresy słupkowe jasno wskazują na istotną zależność pomiędzy badanymi zmiennymi. Im większe znaczenie ma jakość spożywanych posiłków, tym lepszy humor się odczuwa.

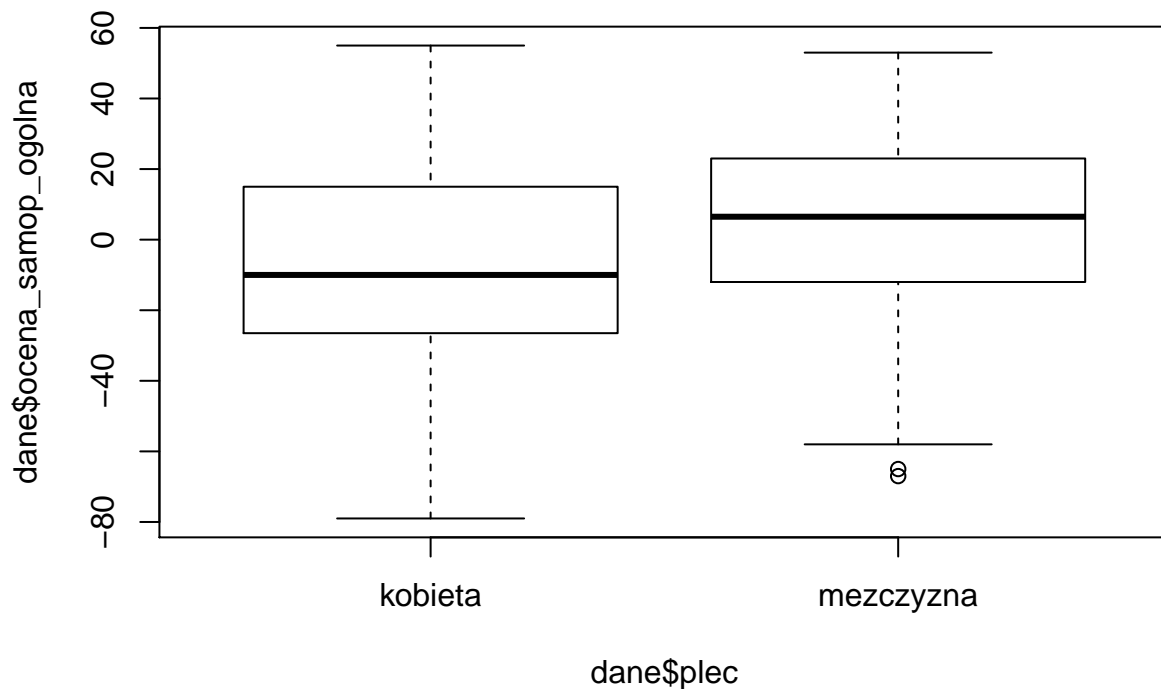
3.8 Hipoteza VIII

Badanie zależności między płcią a ogólną oceną samopoczucia.

3.8.1 Średnia ocena samopoczucia w grupie kobiet i mężczyzn

	Płeć	
	kobieta	mężczyzna
Ogólna ocena samopoczucia	-8.0	3.2

3.8.2 Wykres boxplot



3.8.3 Badanie zależności

P-value z testu Shapiro-Wilka dla cechy Ogólna ocena samopoczucia wynosi 0.00671, zatem ta cecha nie ma rozkładu normalnego. Stosujemy test Manna - Whitney'a - Wilcoxona.

P-value z testu Manna - Whitney'a - Wilcoxona wynosi 0.00150. Istnieje zależność pomiędzy płcią a ogólną oceną samopoczucia.

3.8.4 Badanie równości średnich

Test F wykazuje, że wariancje Ogólnej oceny samopoczucia są równe w grupach kobiet i mężczyzn (p-value 0.53174), zatem stosujemy test T dla średnich.

P-value z testu T wynosi 0.00236, zatem odrzucamy hipotezę o równości średnich. Średnia ogólna ocena samopoczucia różni się istotnie w grupach kobiet i mężczyzn.

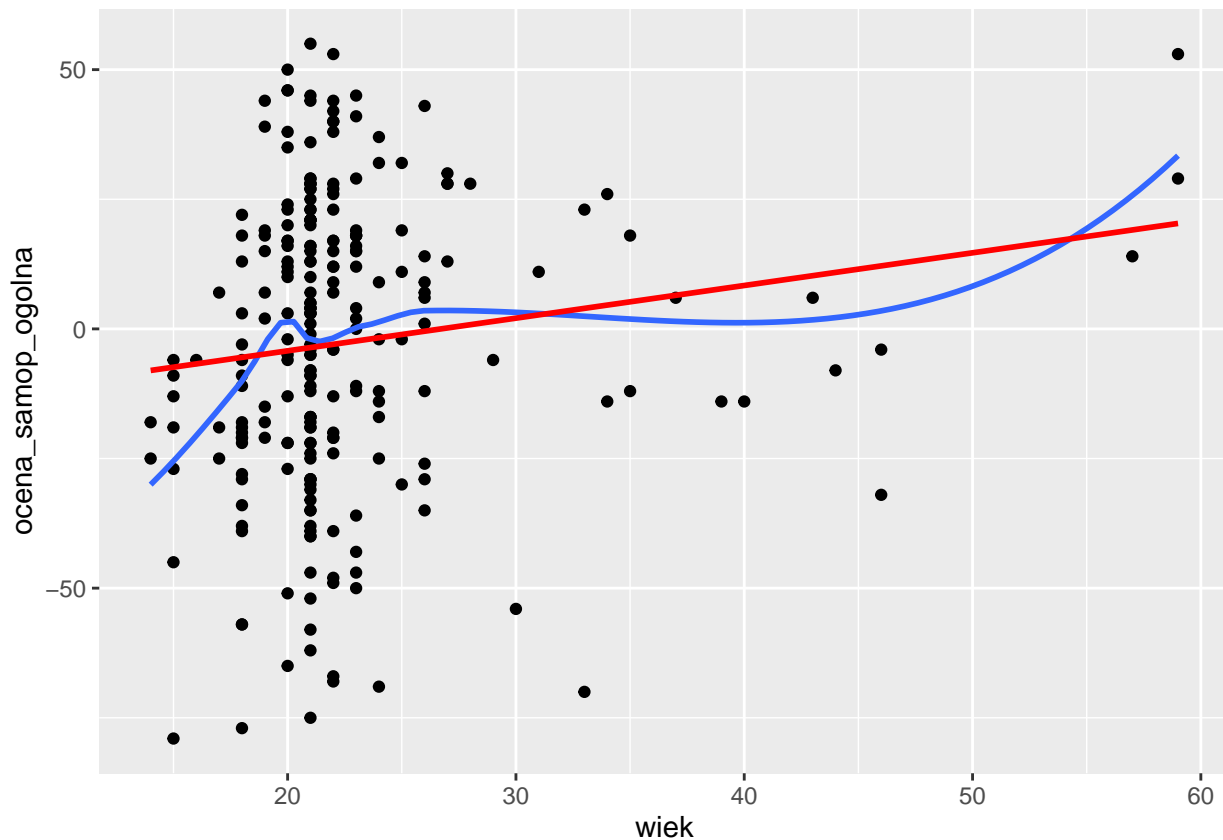
3.9 Hipoteza IX

Badanie zależności między wiekiem a ogólną oceną samopoczucia.

Tablica 1:

<i>Dependent variable:</i>	
ocena_samop_ogolna	
wiek	0.630** (0.292)
Constant	-16.832** (6.847)
Observations	240
R ²	0.019
Adjusted R ²	0.015
Residual Std. Error	28.413 (df = 238)
F Statistic	4.649** (df = 1; 238)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Istnieje istotna zależność pomiędzy ogólną oceną samopoczucia a wiekiem.



Na niebiesko regresja nieparametryczna, na czerwono liniowa.

Na podstawie wykresów diagnostycznych oraz odpowiednich testów mamy, że model spełnia założenia Gaussa

- Markova, a jedynym problemem jest brak normalności rozkładu reszt (choć na wykresach diagnostycznych wygląda trochę na heterogeniczność wariancji).

3.10 Hipoteza X

Zamiast badać związek pomiędzy wiekiem a średnim samopoczuciem, możemy też sprawdzić, czy średnie samopoczucie różni się w kilku grupach wiekowych (ANOVA?), np. poniżej 18, 18-25, powyżej 25.

4 Wnioski

Wykazaliśmy następujące związki:

- wraz ze wzrostem liczby bliskich znajomych, spada częstość odczuwania niepokoju
- im więcej godzin spędzamy na korzystaniu z urządzeń elektronicznych dla rozrywki, tym bardziej czujemy się od nich uzależnieni (co wydaje się oczywistym związkiem, i według testów statystycznych jest prawie pewne, ponieważ p-value było bardzo małe dla tej hipotezy, praktycznie zero)
- z podobną pewnością jak powyższą hipotezę, wykazaliśmy, że im bardziej regularne pory snu, tym lepsza jest jego jakość
- im więcej godzin tygodniowo poświęca się na sport, tym rzadziej odczuwa się załamanie
- im większą wagę ktoś przywiązuje do jakości swoich posiłków, tym lepszy ma humor
- zależność między wiekiem a samopoczuciem można opisać modelem liniowym

Wykazaliśmy także, że nie zachodzą związki:

- pomiędzy jakością snu a poczuciem upokorzenia - chociaż tutaj wynik był na krawędzi, ocierał się o wniosek, że im gorsza jakość snu, tym częściej odczuwamy upokorzenie
- pomiędzy dietą a poczuciem radości

Porównując wyniki w różnych podgrupach, odkryliśmy, że:

- średnia ocena ogólna samopoczucia różniła się między grupą kobiet a mężczyzn, w dodatku była znacznie niższa (-8) w grupie kobiet niż mężczyzn (3.2)

5 Bibliografia

(Bogdan Wojciszke 2005) (Wątroba 2009)

Bogdan Wojciszke, Wiesław Baryła. 2005. *Skale do Pomiaru Nastroju i Sześciu Emocji*.

Wątroba, Janusz. 2009. *Wspomaganie Statystycznej Analizy Wyników Badań Empirycznych w STATISTICA 9*. StatSoft Polska Sp. z o. o.