

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	3
2.3	Zbiorcza tabelka	3
3	Hipotezy badawcze oraz ich weryfikacja	7
3.1	Hipoteza I	7
3.2	Hipoteza II	8
3.3	Hipoteza III	10
3.4	Hipoteza IV	12
3.5	Hipoteza V	13
3.6	Hipoteza VI	14
3.7	Hipoteza VII	15
3.8	Hipoteza VIII	16
3.9	Hipoteza IX	18
3.10	Hipoteza X	19
4	Wnioski	19
5	Bibliografia	19

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`.

2.3 Zbiorcza tabela

Pytania dotyczące aktualnego nastroju były oceniane 5-cio stopniową skalą. Wartości 1-5 oznaczały odpowiednio: “1 - nie zgadzam się”, “2 - raczej się nie zgadzam”, “3 - trochę tak, trochę nie”, “4 - raczej się zgadzam”, “5 - zgadzam się”.

Cecha	Minimum	Mediana	Średnia	Maksimum	Czy rozkład jest normalny?
Jestem w złym humorze	1	2	1.95	5	NIE
Czuję się świetnie	1	3	3.3	5	NIE
Jestem w złym nastroju	1	2	2	5	NIE
Czuję się rozluźniony i spokojny	1	3	3.32	5	NIE
Czuję się szaro i beznadziejnie	1	2	1.95	5	NIE
Jestem w dobrym humorze	1	4	3.61	5	NIE
Jestem pogodny	1	4	3.52	5	NIE
Czuję się przygnębiony	1	2	2.12	5	NIE
Moje samopoczucie jest złe	1	2	2.04	5	NIE
Moje samopoczucie jest dobre	1	4	3.63	5	NIE

Przy ocenie częstości odczuwania poszczególnych emocji, do wyboru były odpowiedzi oceniane w skali 7-mio stopniowej. Oznaczały odpowiednio: “1 - nigdy”, “2 - bardzo rzadko”, “3 - rzadko”, “4 - czasami”, “5 - często”, “6 - bardzo często”, “7 - zawsze”.

Cecha	Minimum	Mediana	Średnia	Maksimum	Czy rozkład jest normalny?
Radość	1	5	4.8	7	NIE
Niepokój	1	4	3.61	7	NIE
Strach	1	3	2.8	7	NIE
Miłość	1	5	4.5	7	NIE
Gniew	1	4	3.51	7	NIE
Wesołość	2	5	4.98	7	NIE
Upokorzenie	1	2	2.15	7	NIE
Smutek	1	3	3.45	7	NIE
Wzburzenie	1	3	3.34	7	NIE
Lęk	1	3	2.85	7	NIE
Żal	1	3	2.88	7	NIE
Obawa	1	3	3.54	7	NIE
Złość	1	3	3.38	7	NIE
Załamanie	1	2	2.77	7	NIE
Szczęście	1	5	4.8	7	NIE
Wściekłość	1	3	2.73	7	NIE
Oddanie	1	4	4.13	7	NIE
Zadowolenie	1	5	4.77	7	NIE
Czułość	1	5	4.38	7	NIE
Wstyd	1	2	2.45	7	NIE
Poczucie winy	1	2	2.73	7	NIE
Przywiązanie	1	5	4.69	7	NIE
Depresja	1	2	2.35	7	NIE
Nieszczerść	1	2	2.4	7	NIE

Pytania dotyczące snu oceniane były w skali 5-cio stopniowej. Wartości w pytaniu dotyczącym:

- średniej liczby przesypianych godzin oznaczały: “1 - mniej niż 4”, “2 - 4-6h”, “3 - 6-8h”, “4 - 8-10h”, “5 - więcej niż 10h”;
- posiadania wyznaczonych pór snu określały: “1 - posiadanie regularnych pór snu” do “5 - posiadanie nieregularnych pór snu”;
- oceny jakości snu określały: “1 - bardzo zła jakość” do “5 - bardzo dobra jakość”.

Cecha	Minimum	Mediana	Średnia	Maksimum	Czy rozkład jest normalny?
Liczba godzin przespanych w ciągu doby	1	3	2.96	5	NIE
Posiadanie wyznaczonych pór snu	1	3	3.23	5	NIE
Jakość snu	1	3	3.44	5	NIE

Pytania dotyczące sportu oceniane były w skali 5-cio stopniowej. Wartości w pytaniu dotyczącym:

- liczby godzin poświęconych tygodniowo na sport oznaczały: “1 - nie uprawiam żadnego sportu”, “2 - mniej niż 1h”, “3 - 1-2,5h”, “4 - 2,5-4h”, “5 - więcej niż 4h”;

- najczęściej wybieranego środka transportu oznaczały: “1 - samochód”, “2 - komunikacja miejska”, “3 - rower”, “4 - chodzenie pieszo”, “5 - inny”;
- oceny własnej kondycji fizycznej określały: “1 - bardzo zła” do “5 - bardzo dobra”.

Cecha	Minimum	Mediana	Średnia	Maksimum	Czy rozkład jest normalny?
Godziny poświęcane tygodniowo na sport	1	3	2.98	5	NIE
Najczęściej wybierany środek transportu	1	3	3.27	5	NIE
Ocena kondycji fizycznej	1	3	3.22	5	NIE

Pytania dotyczące korzystania z urządzeń elektronicznych oceniane były w skali 5-cio stopniowej. Wartości w pytaniu dotyczącym:

- średniej liczby godzin poświęcanej w ciągu dnia na korzystanie z u.e. w celu pracy bądź nauki oznaczały: “1 - nie używam/rzadko używam komputera do pracy/nauki”, “2 - mniej niż 3”, “3 - 3-5h”, “4 - 5-8h”, “5 - powyżej 5h”;
- średniej liczby godzin poświęconej w ciągu dnia na korzystanie z u.e w celach rozrywkowych oznaczały: “1 - mniej niż 1h”, “2 - 1-2h”, “3 - 2-3h”, “4 - 3-4h”, “5 - więcej niż 4h”;
- częstości występowania poczucia uzależnienia od u.e. oznaczały: “1 - nigdy”, “2 - rzadko”, “3 - czasami”, “4 - często”, “5 - bardzo często”.

Cecha	Minimum	Mediana	Średnia	Maksimum	Czy rozkład jest normalny?
Godziny spędzone dziennie na pracy na urządzeniach elektronicznych	1	3	3.14	5	NIE
Godziny spędzone dziennie na rozrywce na urządzeniach elektronicznych	1	3	2.81	5	NIE
Poczucie uzależnienia od urządzeń elektronicznych	1	3	3.27	5	NIE

Pytania dotyczące diety oceniane były w skali 5-cio stopniowej. Wartości w pytaniu dotyczącym:

- przeważającego składnika diety oznaczały: “1 - fast food”, “2 - żywność mocno przetworzona (mrożonki, gotowe dania)”, “3 - mięso i ciężkostrawne potrawy”, “4 - warzywa i owoce”, “5 - mam zbilansowaną dietę”;
- wagi przywiązywanej do jakości spożywanych posiłków oznaczały: “1 - wszystko mi jedno”, “2 - nie jest to moim priorytetem”, “3 - staram się jeść zdrowo, ale mi nie wychodzi”, “4 - jest to dla mnie dość ważne”, “5 - jestem tym, co jem!”;
- regularności spożywanych posiłków oznaczały: “1 - jem nieregularnie”, “2 - różnie z tym bywa”, “3 - staram się jeść regularnie”, “4 - zazwyczaj jem regularnie”, “5 - mam stałe pory posiłków”.

Cecha	Minimum	Mediana	Średnia	Maksimum	Czy rozkład jest normalny?
Przeważający składnik w diecie	1	5	3.92	5	NIE
Dbanie o jakość posiłków	1	3	3.21	5	NIE
Regularność posiłków	1	2	2.41	5	NIE

Pytania dotyczące kontaktów towarzyskich oceniane były w skali 5-cio stopniowej. Wartości w pytaniu dotyczącym:

- posiadania osoby, na której można polegać w trudnej sytuacji oznaczały: “1 - nie mam nikogo takiego”, “2 - z rodziną/znajomymi nie poruszam poważnych tematów”, “3 - nie czuję się komfortowo rozmawiając na trudne tematy z rodziną/znajomymi”, “4 - na ogół mam do kogo zwrócić się z problemem”, “5 - mam osoby, z którymi mogę porozmawiać o wszystkim”;
- liczby osób, z którymi utrzymuje się stały kontakt oznaczały: “1 - brak takiej osoby”, “2 - jedną osobę”, “3 - 2-3 osoby”, “4 - 4-5 osób”, “5 - 5 i więcej osób”;
- znaczenia, jakie mają dla ankietowanego kontakty z innymi ludźmi oznaczały “1 - brak znaczenia” do “5 - bardzo duże znaczenie”.

Cecha	Minimum	Mediana	Średnia	Maksimum	Czy rozkład jest normalny?
Posiadanie zaufanej osoby	1	5	4.35	5	NIE
Liczba osób, z którymi utrzymuje się bliski kontakt	1	4	3.9	5	NIE
Jakie znaczenie mają dla Ciebie kontakty z innymi ludźmi?	1	4	3.91	5	NIE
Ogólna ocena samopoczucia	-79	-0.5	-2.61	55	NIE

Ogólna ocena samopoczucia została uzyskana poprzez zsumowanie wartości uczuć pozytywnych ze znakiem dodatnim oraz negatywnych z ujemnym.

3 Hipotezy badawcze oraz ich weryfikacja

H_0 : nie istnieje zależność pomiędzy zmiennymi.

H_1 : istnieje zależność pomiędzy zmiennymi.

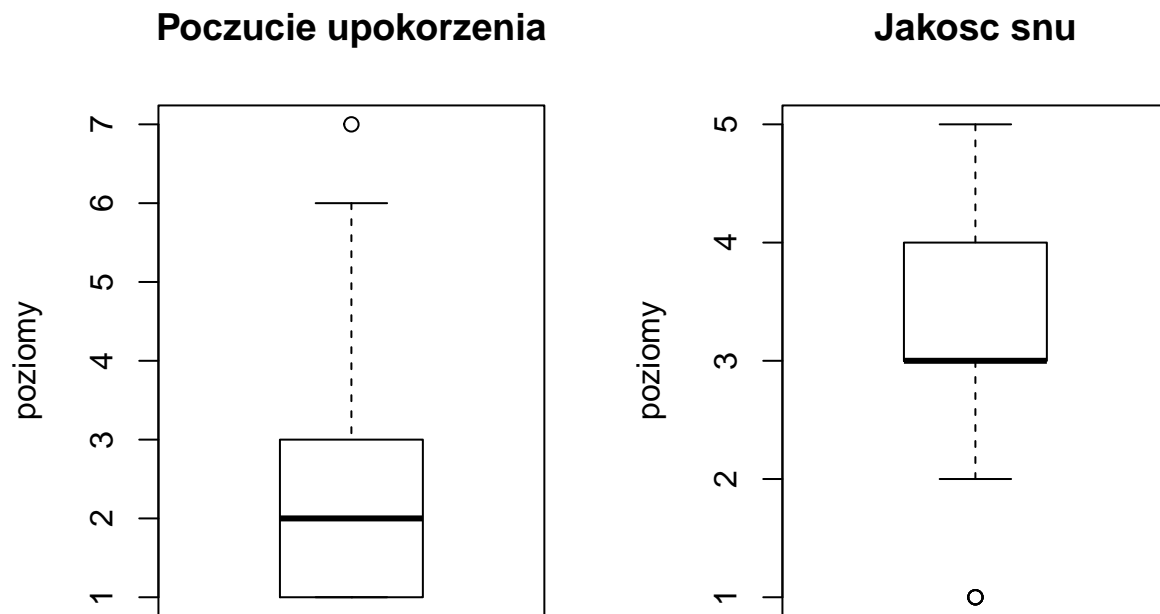
Hipotezy badane są na poziomie istotności $\alpha = 0.05$.

3.1 Hipoteza I

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

P-value dla testu **rho-Spearmana** wynosi $p = 0.054$, natomiast dla testu **tau-Kendalla** wynosi $p = 0.053$. 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

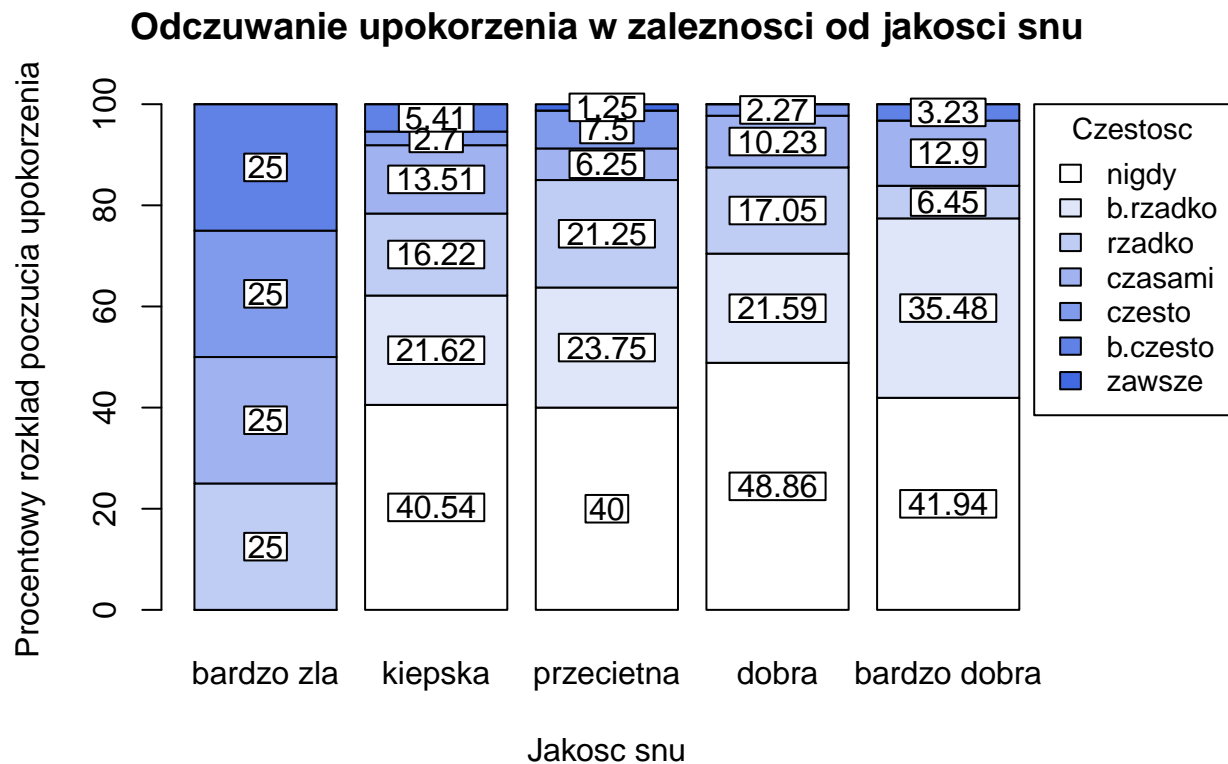


3.1.2 Tabela kontyngencji

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

Poniżej przedstawiony wykres zawiera wartości procentowe, nie liczebności dla poszczególnych odpowiedzi; podobnie jak w przypadku wykresów tego typu przedstawionych dla kolejnych hipotez.

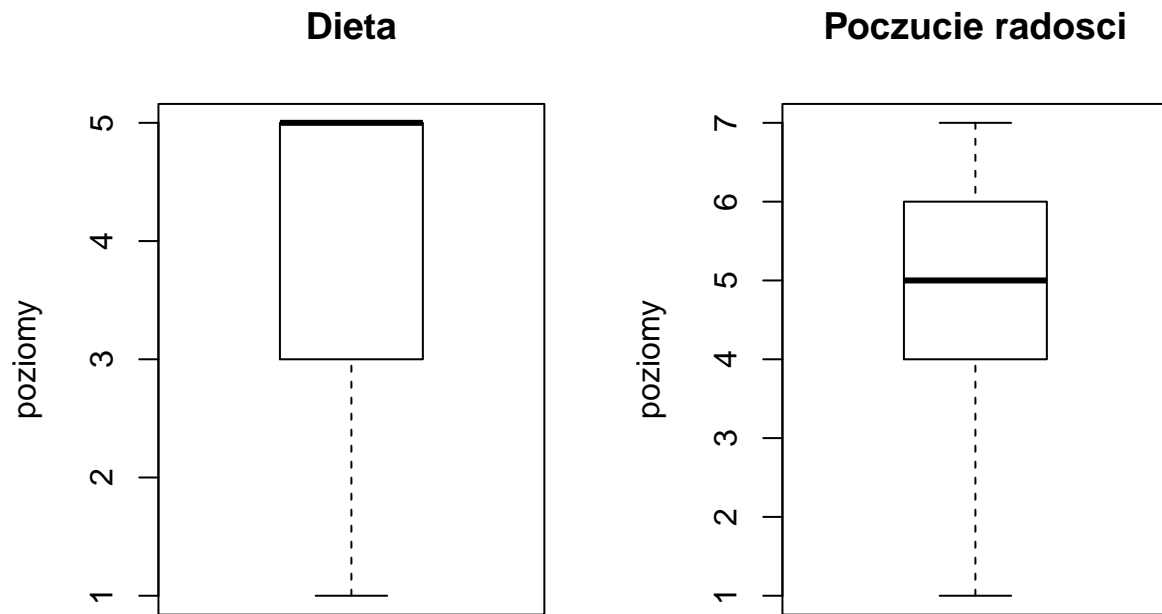


3.2 Hipoteza II

Badanie, czy istnieje zależność pomiędzy dietą a poczuciem radości.

P-value dla testu rho-Spearmana wynosi $p = 0.084$, natomiast dla testu tau-Kendalla wynosi $p = 0.084$. Nie ma podstaw do odrzucenia H_0 . Nie istnieje istotna zależność pomiędzy dietą a poczuciem radości.

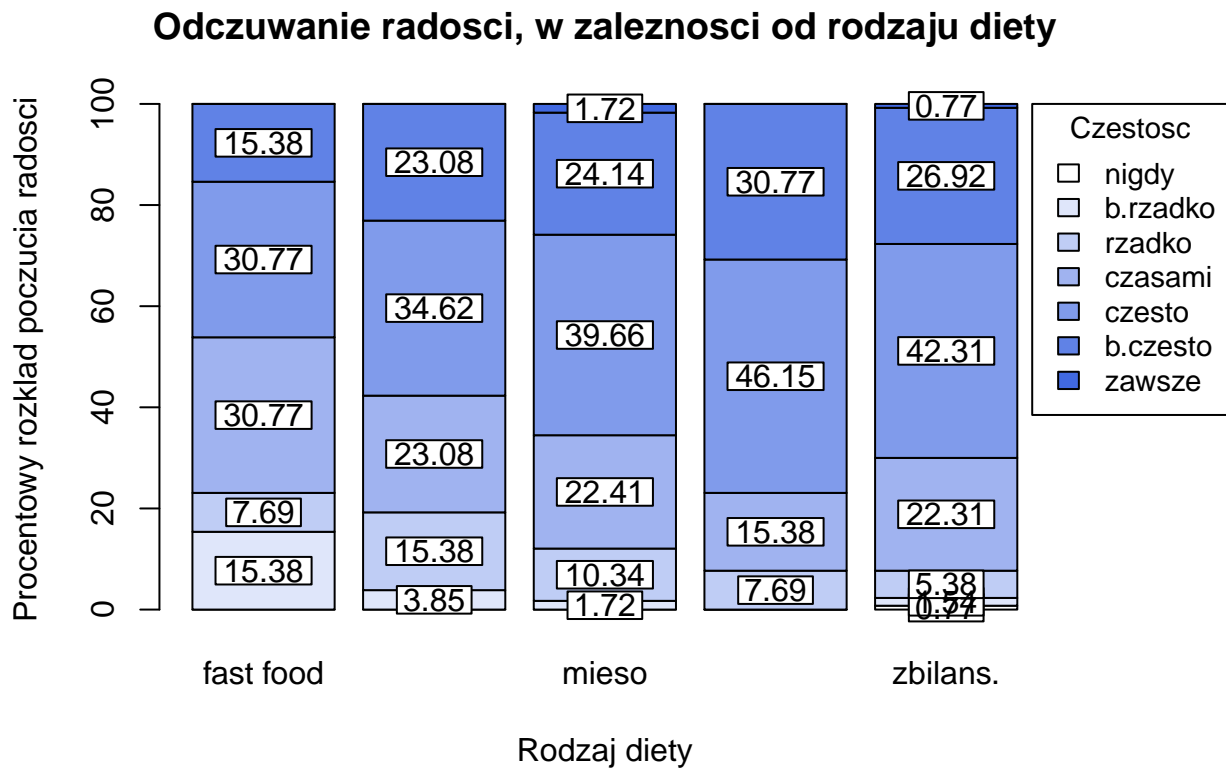
3.2.1 Wykresy boxplot



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.

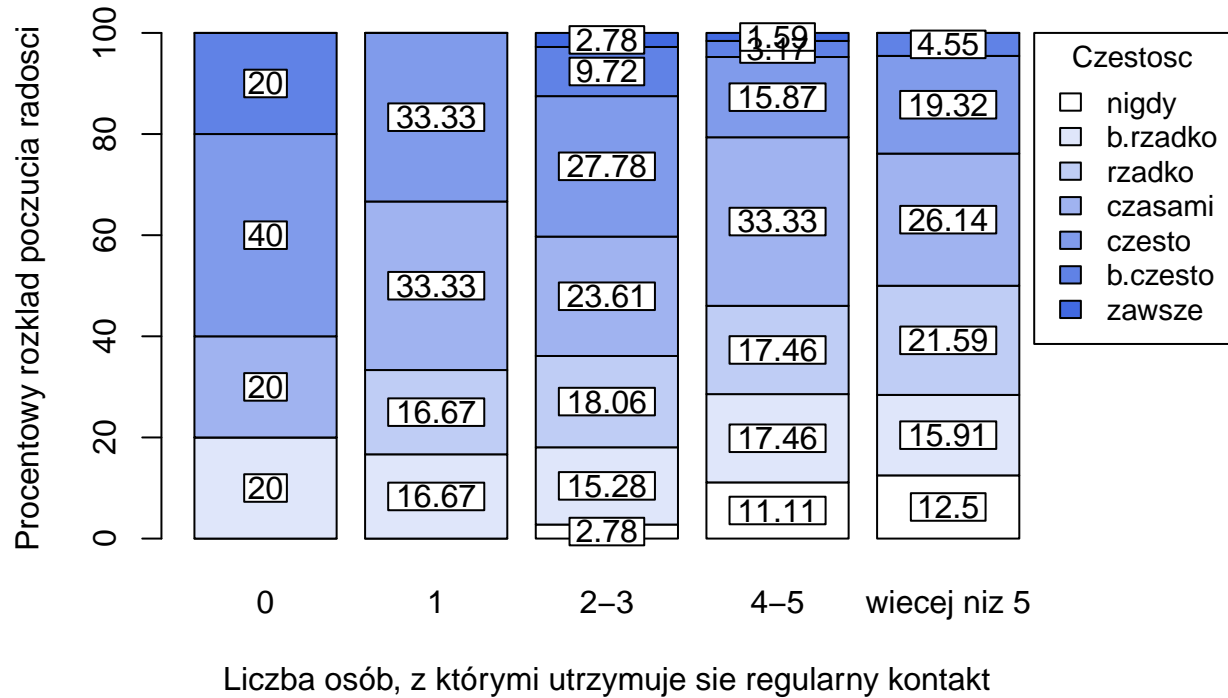
P-value dla obu testów wynosi około $p = 0.006$. Należy odrzucić hipotezę H_0 , na rzecz hipotezy alternatywnej H_1 . Istnieje zależność pomiędzy liczbą osób, z którymi utrzymuje się regularny kontakt a odczuwaniem niepokoju.

3.3.1 Tabela kontyngencji

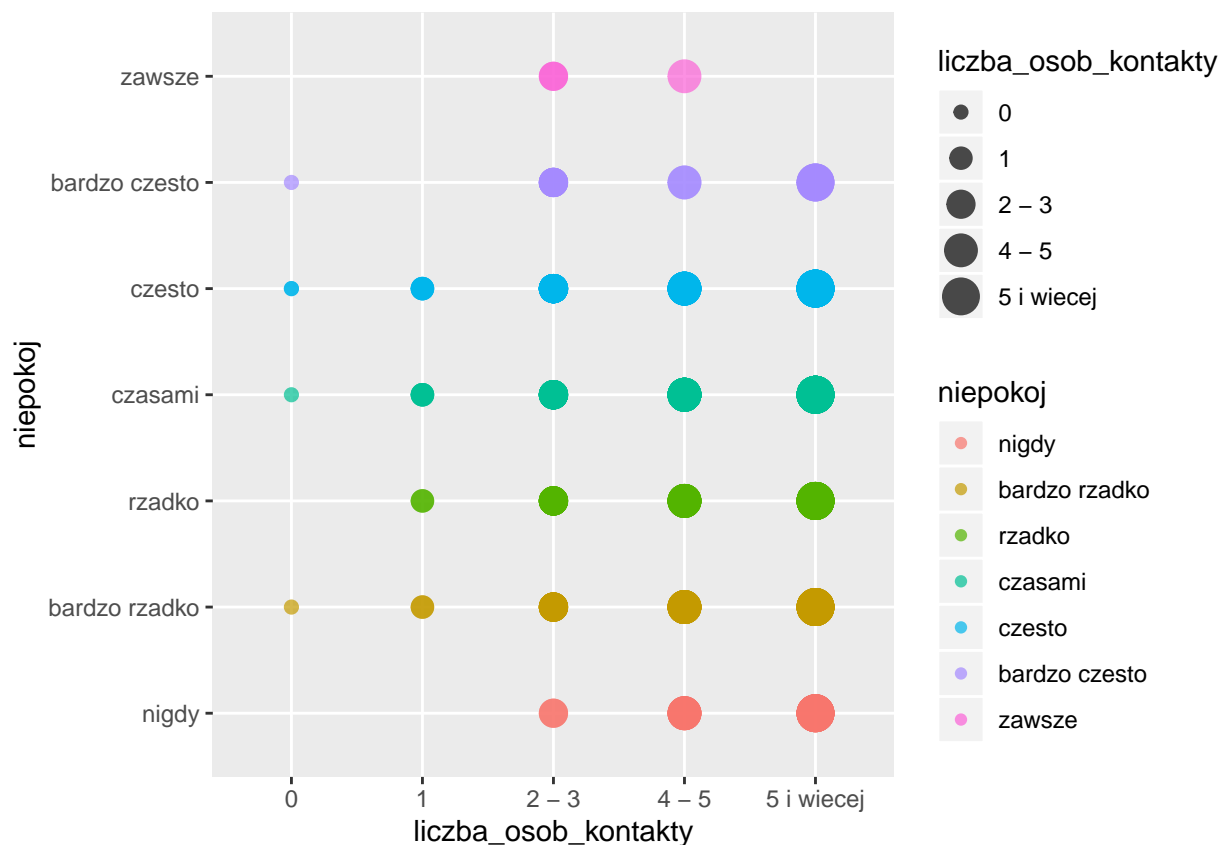
Niepokój	Liczba bliskich znajomych				
	0	1	2 - 3	4 - 5	5 i więcej
nigdy	0.0	0.0	2.8	11.1	12.5
bardzo rzadko	20.0	16.7	15.3	17.5	15.9
rzadko	0.0	16.7	18.1	17.5	21.6
czasami	20.0	33.3	23.6	33.3	26.1
często	40.0	33.3	27.8	15.9	19.3
bardzo często	20.0	0.0	9.7	3.2	4.5
zawsze	0.0	0.0	2.8	1.6	0.0

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

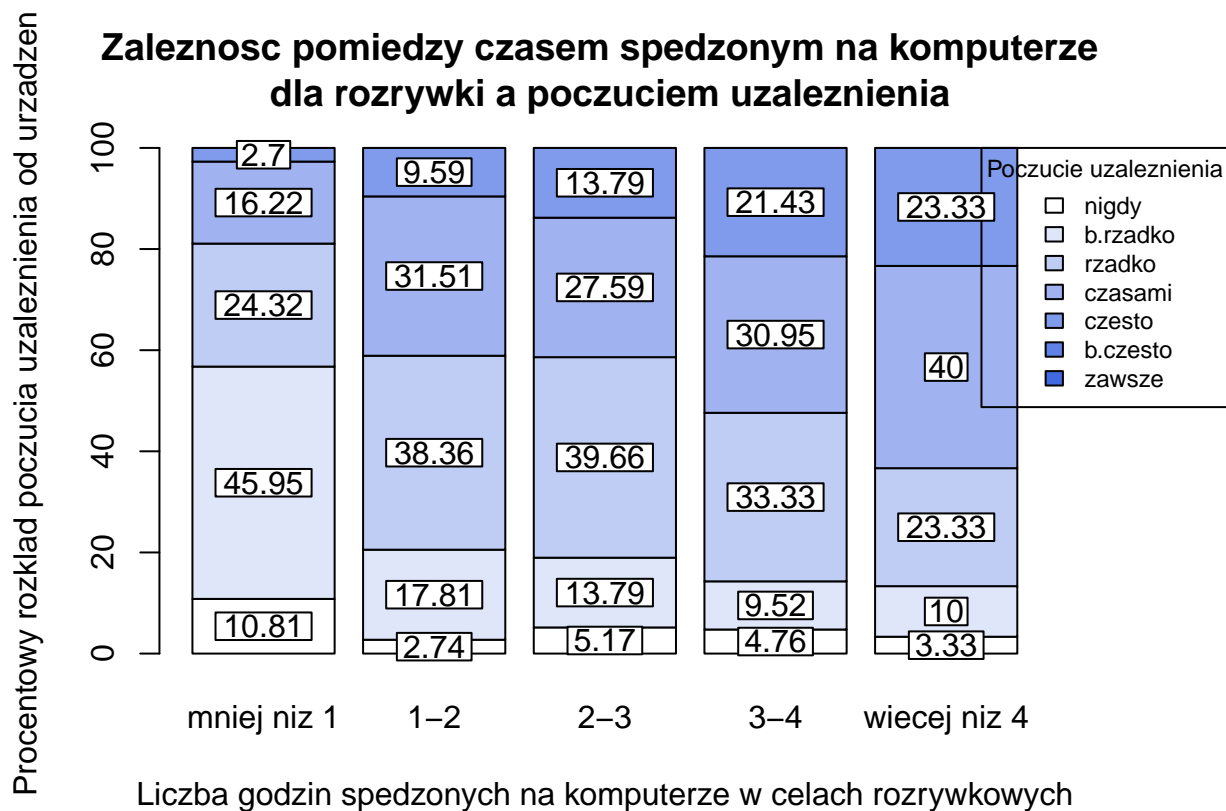
Badanie, czy istnieje zależność pomiędzy liczbą godzin przeznaczonych na korzystanie z urządzeń elektronicznych a poczuciem uzależnienia od urządzeń.

P-value dla obu testów są nieistotnie różne od zera ($p < 0.001 **$). 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 rho-Spearmana i tau-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

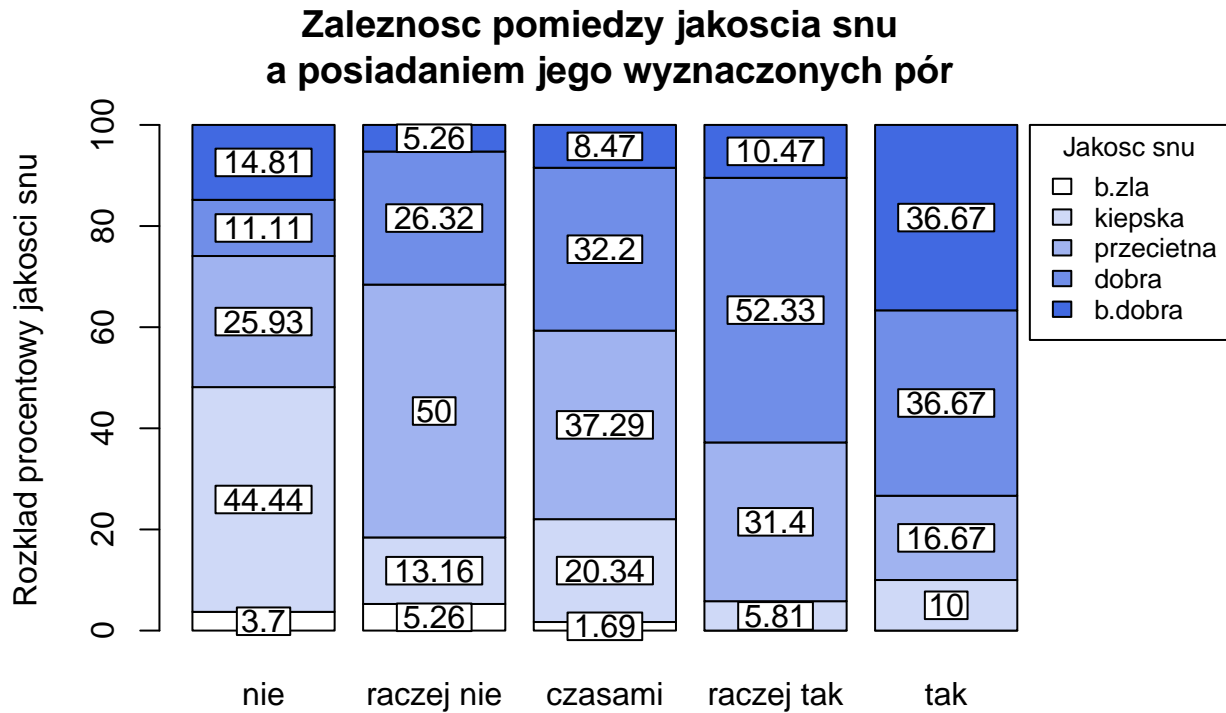
Badanie, czy istnieje zależność pomiędzy posiadaniem wyznaczonych pór snu a jego jakością.

P-value dla obu testów są nieistotnie różne od zera ($p < 0.001 **$). 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



Współczynniki korelacji dla testów ρ -Spearmana i τ -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

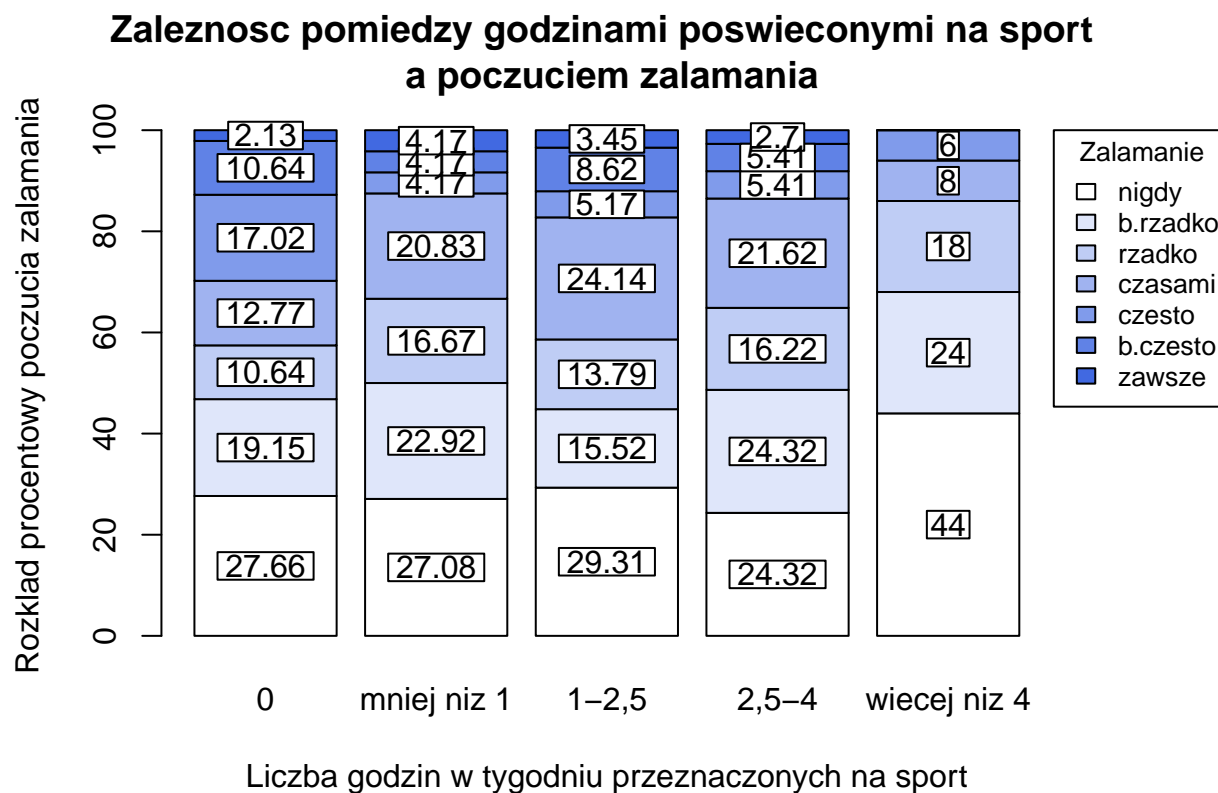
Badanie, czy istnieje zależność pomiędzy liczbą godzin przeznaczonych na uprawianie sportu a odczuwaniem załamania.

P-value dla testu $p = \rho$ -Spearmana wynosi 0.011, natomiast dla testu τ -Kendalla wynosi $p = 0.011$. 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

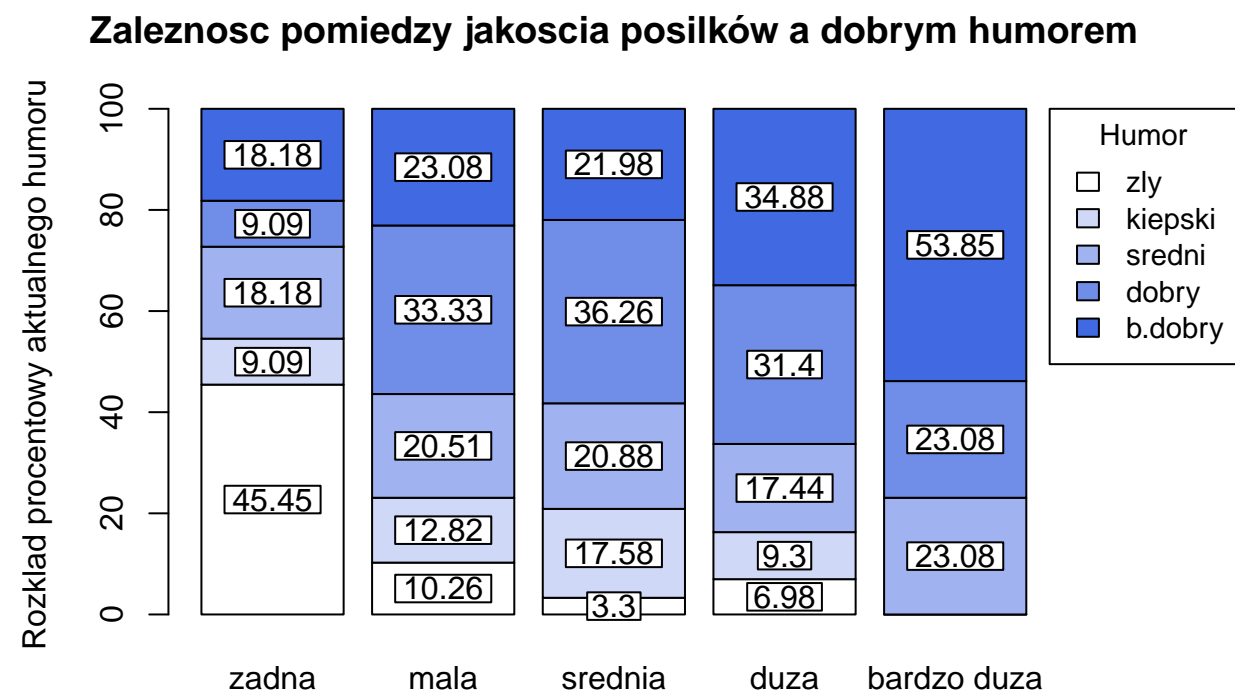
Badanie, czy istnieje zależność pomiędzy jakością posiłków a dobrym humorem.

P-value dla testu rho-Spearmana wynosi $p = 0.002$, natomiast dla testu tau-Kendalla wynosi $p = 0.001^{**}$. 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 rho-Spearmana i tau-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.

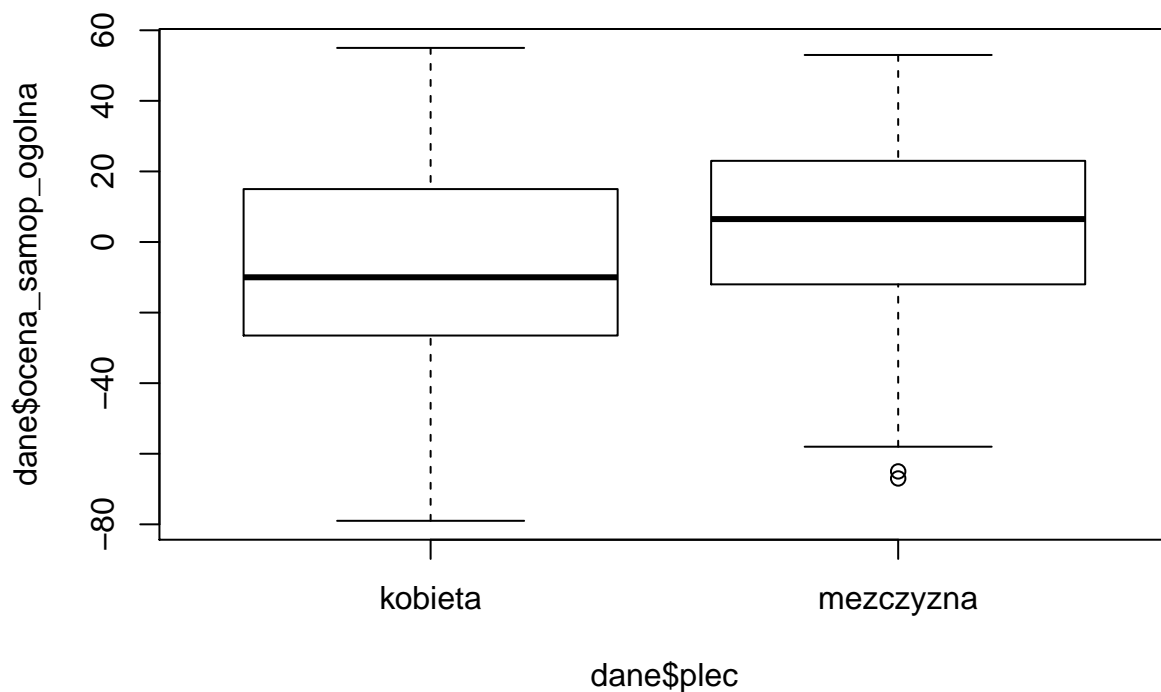
3.8 Hipoteza VIII

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 $p = 0.007$, zatem ta cecha nie ma rozkładu normalnego. Stosujemy test Manna - Whitney'a - Wilcoxona.

P-value z testu Manna - Whitney'a - Wilcoxona wynosi $p = 0.001$ **. 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 ($p = 0.532$), zatem stosujemy test T dla średnich.

P-value z testu T wynosi 0.002, zatem odrzucamy hipotezę o równości średnich. Średnia ogólna ocena samopoczucia różni się istotnie w grupach kobiet i mężczyzn.

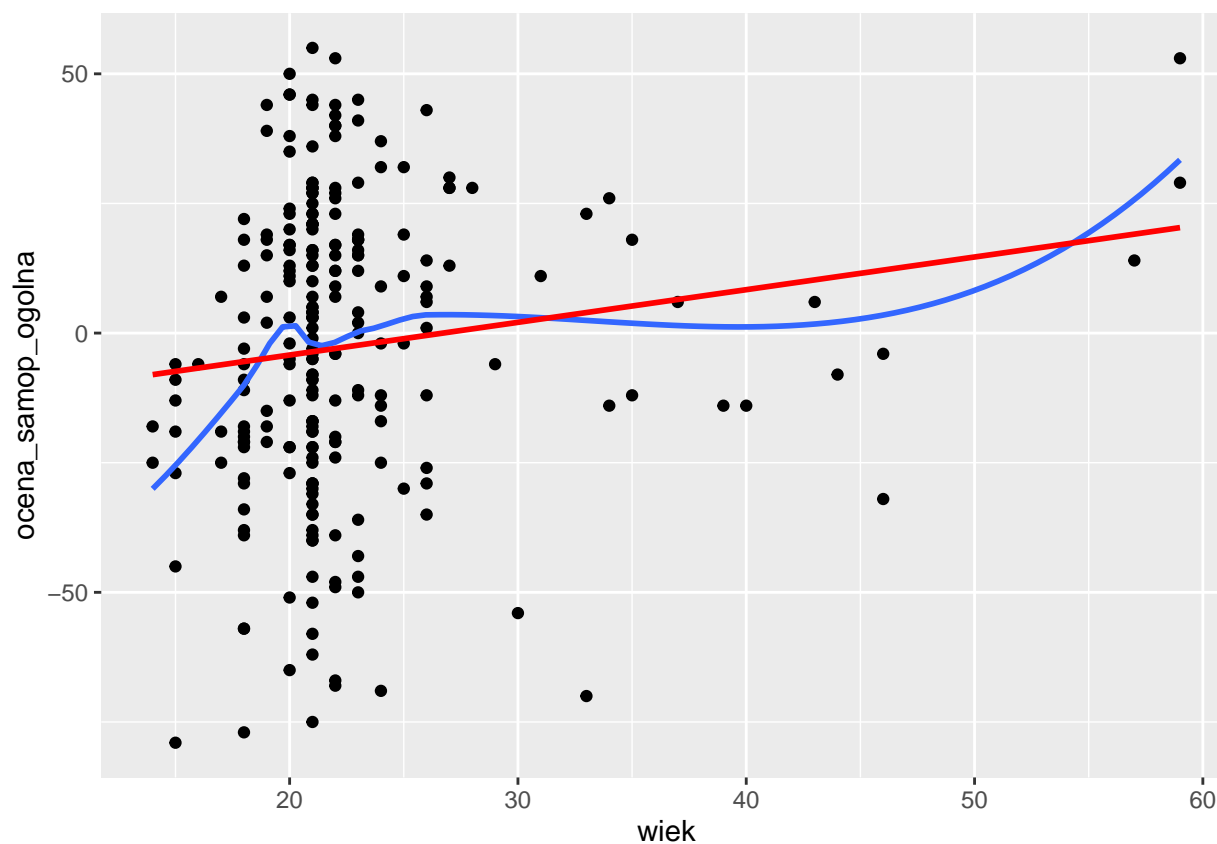
3.9 Hipoteza IX

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.