

Politechnika Lubelska
Wydział Podstaw Techniki

ANALIZA CZYNNIKÓW RYZYKA ZACHOROWANIA NA CHOROBE NIEDOKRWIENNĄ SERCA

PROJEKT Z ZAKRESU ANALIZY DANYCH

Autor:
MARCIN WOŹNIAK

Matematyka
III rok

1 Wstęp

W polskiej populacji niezmiennie od wielu lat choroby układu krążenia (ChUK) są główną przyczyną zgonów – W 2020 roku stanowiły ok. 37% wszystkich przyczyn zgonów. Wśród nich największy odsetek stanowi choroba niedokrwienna serca (ChNS, ang. Coronary Heart Disease – CHD) – w 2013 było to 23.0% zgonów z powodu ChUK w populacji ogólnej i 21.6% zgonów z powodu ChUK w populacji 65+. W 2017 roku na ChNS chorowało w Polsce 1.6 mln osób (4.2% populacji).

Choroba niedokrwienna serca to pojęcie obejmujące wszystkie stany niedokrwienia mięśnia sercowego. Klinicznie dzieli się ona na:

- dławicę piersiową stabilną
- ostre zespoły wieńcowe
- dławicę odmienną
- nieme niedokrwienie

Zidentyfikowano wiele czynników zwiększających ryzyko wystąpienia ChNS. Możemy podzielić je na niemodyfikowalne i modyfikowalne. Do tych pierwszych zaliczamy:

- wiek
- płeć
- wywiad rodzinny

Do czynników modyfikowalnych możemy zaliczyć między innymi:

- nadciśnienie tętnicze (NT)
- palenie tytoniu
- nadwagę i otyłość
- cukrzycę
- zaburzenia lipidowe

Celem projektu będzie zbadanie istnienia zależności pomiędzy niektórymi z wyżej wymienionych czynników a ryzykiem wystąpienia choroby niedokrwiennej serca.

2 Materiały i metody

Do analizy statystycznej został wykorzystany zbiór danych pochodzący ze strony internetowej kaggle.com, zawierający informacje pochodzące z badania „Framingham Heart Study”. Jest to badanie kohortowe prowadzone od 1948 roku na populacji miasteczka Framingham w USA. Należy zaznaczyć, że nie są to wszystkie dane pochodzące z tego badania, a jedynie nieznana ich część – najpewniej informacje dotyczące jednej z badanych kohort. Dane zawierają 4238 rekordów, z których dla każdego opisanego jest 16 cech.

Przygotowanie danych, analiza statystyczna i wizualizacja zostały wykonane z użyciem języka R i programu RStudio. Wykorzystano testy statystyczne z pakietu `stats`. Jako graniczną wartość p-value ustalono 0.05. Ilorazy szans obliczono z użyciem funkcji `epitools::oddsratio()`.

3 Opisanie zmiennych

- **male** – płeć (1 – mężczyzna, 0 – kobieta), zmienna nominalna
- **age** – wiek pacjenta, zmienna ilościowa
- **education** – wykształcenie pacjenta, w czterostopniowej skali, zmienna nominalna
- **currentSmoker** – czy pacjent obecnie pali (1 – tak, 0 – nie), zmienna nominalna
- **cigsPerDay** – przeciętna liczba wypalanych dziennie papierosów, zmienna ilościowa
- **BPMeds** – czy pacjent przyjmuje/przyjmował leki na nadciśnienie (1 – tak, 0 – nie), zmienna nominalna
- **prevalentStroke** – czy pacjent przeżył udar mózgu (1 – tak, 0 – nie), zmienna nominalna
- **prevalentHyp** – czy pacjent ma rozpoznane nadciśnienie tętnicze (1 – tak, 0 – nie), zmienna nominalna
- **diabetes** – czy pacjent ma rozpoznaną cukrzycę (1 – tak, 0 – nie), zmienna nominalna
- **totChol** – stężenie cholesterolu całkowitego we krwi, zmienna ilościowa
- **sysBP** – przeciętna wartość skurczowego ciśnienia tętniczego krwi, zmienna ilościowa
- **diaBP** – przeciętna wartość rozkurczowego ciśnienia tętniczego krwi, zmienna ilościowa
- **BMI** – Body Mass Index, zmienna ilościowa
- **heartRate** – przeciętna wartość częstości akcji serca, zmienna ilościowa
- **glucose** – stężenie glukozy we krwi – najprawdopodobniej jest to stężenie glukozy na czczo, zmienna ilościowa
- **TenYearCHD** – czy w ciągu 10 lat obserwacji u pacjenta rozpoznano ChNS (1 – tak, 0 – nie), zmienna nominalna

Zmienna **TenYearCHD** zostanie użyta do badania różnic w ryzyku wystąpienia ChNS.

4 Ustalenie hipotez badawczych

Hipotezy badawcze zostały ustalone na podstawie dostępnej literatury. Ich celem będzie zweryfikowanie, czy poznane czynniki ryzyka są istotne dla badanej przeze mnie populacji:

1. Wraz z wiekiem rośnie ryzyko wystąpienia ChNS.
2. U mężczyzn występuje większe ryzyko ChNS niż u kobiet.
3. U osób z rozpoznany nadciśnieniem istnieje większe ryzyko ChNS niż u osób bez nadciśnienia.
4. Wraz ze wzrostem skurczowego ciśnienia krwi rośnie ryzyko wystąpienia ChNS.
5. Wraz ze wzrostem rozkurczowego ciśnienia krwi rośnie ryzyko wystąpienia ChNS.
6. U osób z rozpoznaną cukrzycą istnieje większe ryzyko ChNS niż u osób bez cukrzycy.
7. Wraz ze wzrostem BMI rośnie ryzyko wystąpienia ChNS.
8. Wraz ze wzrostem stężenia cholesterolu całkowitego rośnie ryzyko wystąpienia ChNS.
9. Osoby palące papierosy mają większe ryzyko wystąpienia ChNS.

10. Wraz ze wzrostem liczby dziennie wypalanych papierosów rośnie ryzyko wystąpienia ChNS.

5 Przygotowanie danych

Przed przystąpieniem do analizy statystycznej należy wstępnie przygotować dostępne dane.

Tabela 1: Braki danych w analizowanym zbiorze

	liczba brakujących danych
male	0
age	0
education	105
currentSmoker	0
cigsPerDay	29
BPMeds	53
prevalentStroke	0
prevalentHyp	0
diabetes	0
totChol	50
sysBP	0
diaBP	0
BMI	19
heartRate	1
glucose	388
TenYearCHD	0

W analizowanym zbiorze brakuje łącznie 645 informacji, co stanowi ok. 0.95% wszystkich danych. Rekordów, w których brakuje danych jest 582, co stanowi 13.7% wszystkich rekordów. Najliczniejsze braki występują dla cech `glucose` oraz `education`.

5.1 Cecha `education`

Ponieważ dane najpewniej nie pochodzą z ankiety, możemy przypuszczać, że parametry medyczne zostały uzupełnione przez badającego, natomiast informacja o poziomie edukacji pacjenta może być jedynie jego deklaracją - możemy sprawdzić, czy brak tej informacji nie jest zależny od płci i wieku pacjentów. Ponieważ rozkład wieku w zależności tego, czy została podana informacja o edukacji nie jest rozkładem normalnym (dla braku danych: $p = 0.0263$, dla pozostałych: $p = 6.55 \times 10^{-30}$, test Shapiro-Wilka), dlatego do porównania użyjemy testu Manna-Whitneya. Dla zbadania różnic w braku informacji o edukacji w zależności od płci użyjemy testu χ^2 .

Test Manna-Whitneya wskazuje na brak istotnych różnic w rozkładzie wieku w zależności od podania informacji o edukacji ($p = 0.1912$).

Test χ^2 sugeruje, że nie ma podstaw do odrzucenia hipotezy o niezależności płci i braku informacji o edukacji ($p = 0.1378$).

Cechę `education` uzupełnimy modą z użyciem funkcji `DMwR2::centralImputation()`.

5.2 Cechy `currentSmoker` i `cigsPerDay`

Cechy te są w oczywisty sposób ze sobą związane. Należy sprawdzić czy nie ma w tych kolumnach sprzecznych informacji tj. czy nie ma rekordów, dla których `currentSmoker = True` i `cigsPerDay = 0`, lub `currentSmoker = F` i `cigsPerDay != 0`:

```
nrow(filter(df, currentSmoker == TRUE & cigsPerDay == 0))
```

```
## [1] 0
```

```
nrow(filter(df, currentSmoker == FALSE & cigsPerDay != 0))
```

```
## [1] 0
```

Z powyższych obliczeń wynika, że w zbiorze nie występują zaprzeczające sobie informacje.

Następnie możemy ocenić braki danych w kolumnie `cigsPerDay` z uwzględnieniem kolumny `currentSmoker`:

```
nrow(filter(df, currentSmoker == FALSE & is.na(cigsPerDay)))
```

```
## [1] 0
```

```
nrow(filter(df, currentSmoker == TRUE & is.na(cigsPerDay)))
```

```
## [1] 29
```

W przypadku wszystkich 29 rekordów, w których brakuje zmiennej `cigsPerDay` wszystkie zmienne `currentSmoker` są równe 1. Możemy te dane uzupełnić medianą rekordów, dla których liczba dziennie wypalanych papierosów jest większa od 0 (dla wszystkich rekordów łącznie mediana wynosi 0 - uzupełnienie taką wartością doprowadziłoby do sprzeczności).

5.3 Pozostałe braki danych

Pozostałe dane zostaną uzupełnione medianą (zmienne ilościowe) lub modą (zmienne jakościowe) odpowiednich kolumn.

Tabela 2: Braki danych po edycji

	liczba brakujących danych
male	0
age	0
education	0
currentSmoker	0
cigsPerDay	0
BPMeds	0
prevalentStroke	0
prevalentHyp	0
diabetes	0
totChol	0
sysBP	0
diaBP	0
BMI	0

	liczba brakujących danych
heartRate	0
glucose	0
TenYearCHD	0

W badanym zbiorze nie mamy już brakujących informacji. Po tym wstępnym przetworzeniu danych możemy przejść do analizy statystycznej.

6 Podstawowe informacje o badanej grupie

W badanej grupie obecnych było 4238 pacjentów. Kobiety stanowiły 57.1% pacjentów a mężczyźni 42.9%. Średnia wieku pacjentów wynosiła 49.6 ± 8.6 lat (przedział wiekowy pacjentów: 32 – 70 lat).

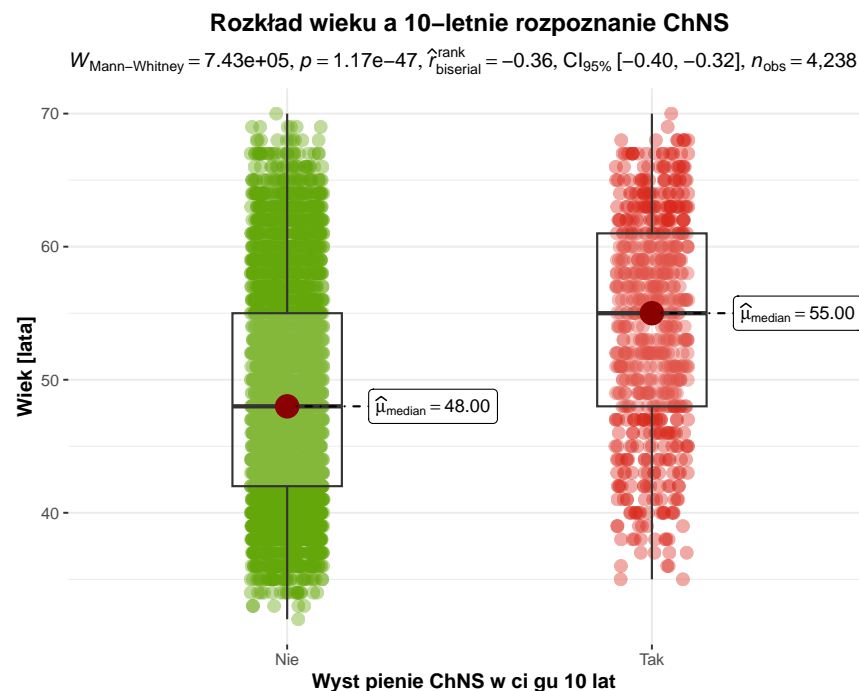
U 15.2% pacjentów w ciągu 10 lat rozpoznano chorobę niedokrwienną serca - zachorowało 18.9% wszystkich mężczyzn i 12.4% wszystkich kobiet.

7 Weryfikacja hipotez

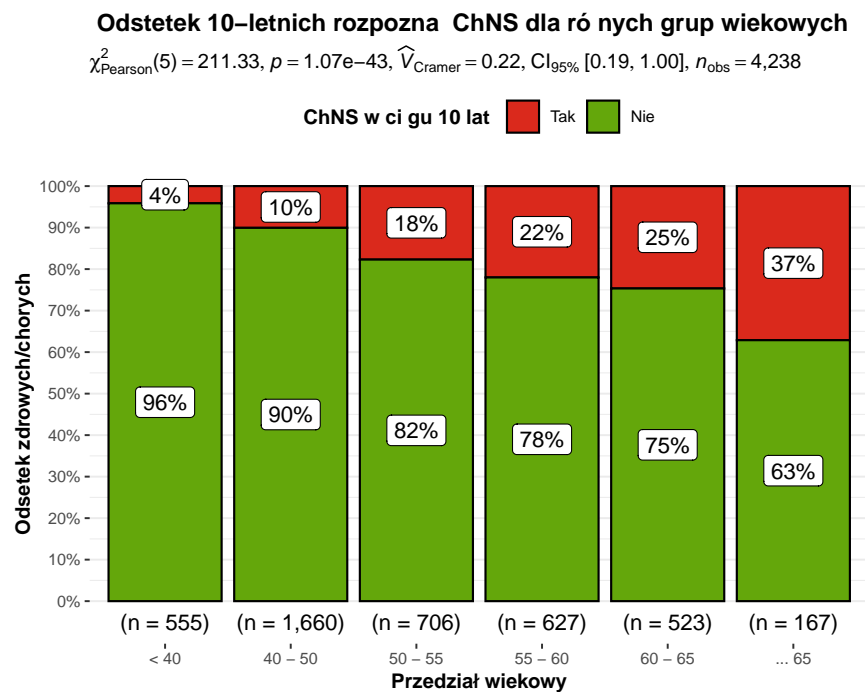
7.1 Wpływ wieku na ryzyko wystąpienia ChNS

Z powodu braku normalności rozkładu wieku dla obu wartości **TenYearCHD** (W teście Shapiro-Wilka w obu przypadkach $p < 0.05$) użyjemy testu Manna-Whitneya. Test sugeruje, że istnieją istotne różnice w rozkładzie wieku w zależności od wystąpienia w ciągu 10 lat ChNS ($p = 1.17 \times 10^{-47}$). Wiek ma istotny wpływ na wystąpienie ChNS.

Różnice obrazuje poniższy wykres:



Ryzyko wystąpienia ChNS możemy również przedstawić, dzieląc badanych na kilka przedziałów wiekowych.

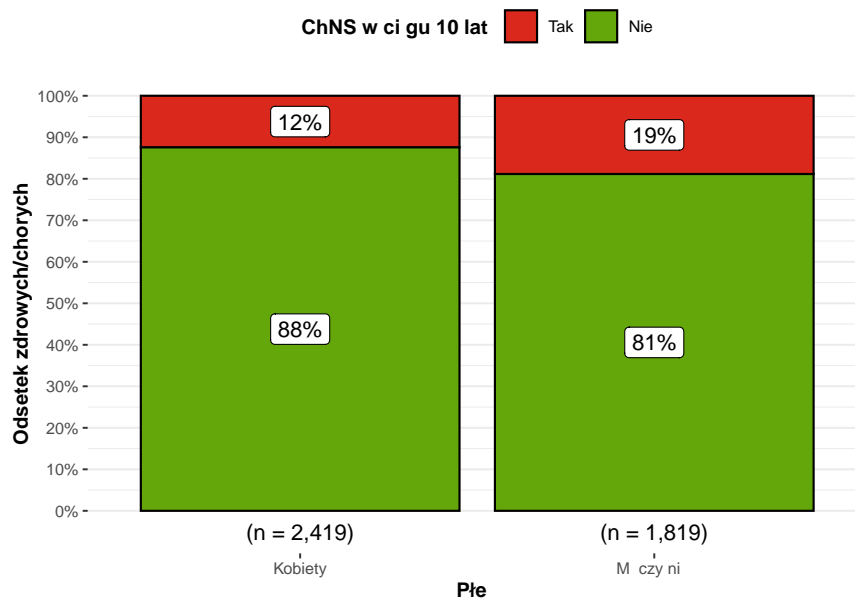


Dla tak utworzonych grup test χ^2 sugeruje, że należy odrzucić hipotezę o niezależności wieku od ryzyka wystąpienia ChNS.

7.2 Płeć a ryzyko wystąpienia ChNS

W badanym przypadku możemy użyć podwójnego testu istotności dla frakcji lub testu χ^2 . Test χ^2 wskazuje, że należy odrzucić hipotezę o niezależności płci i ryzyka wystąpienia ChNS. Ryzyko wystąpienia ChNS istotnie zależy od płci ($p = 1.11 \times 10^{-8}$). Również podwójny test istotności dla frakcji wskazuje, że należy odrzucić hipotezę o równości frakcji w populacji kobiet i mężczyzn. Odstetek wystąpienia ChNS u mężczyzn jest istotnie większy niż u kobiet ($p = 5.54 \times 10^{-9}$). Różnice obrazuje poniższy wykres.

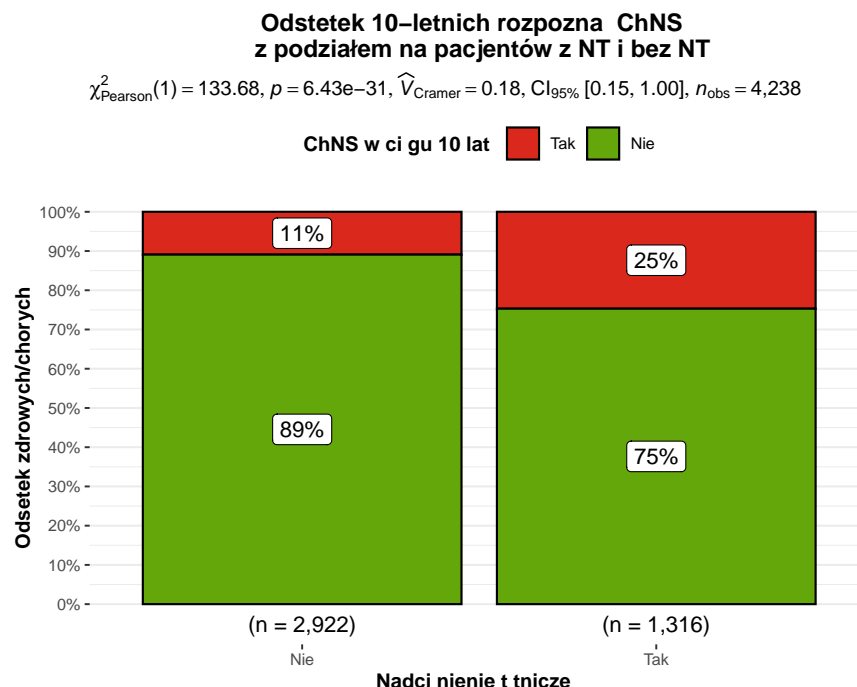
Odstetek 10-letnich rozpozna ChNS z podziałem na płe
 $\chi^2_{\text{Pearson}}(1) = 33.14, p = 8.58\text{e-}09, \hat{V}_{\text{Cramer}} = 0.09, \text{CI}_{95\%} [0.06, 1.00], n_{\text{obs}} = 4,238$



7.3 Nadciśnienie a ryzyko ChNS

Ponownie możemy użyć podwójnego testu istotności i testu χ^2 .

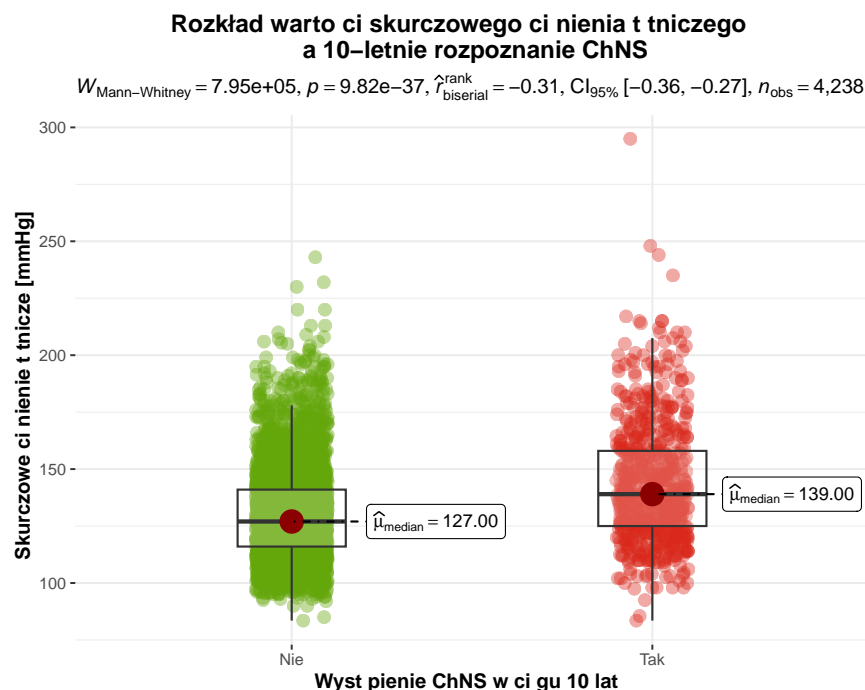
Test χ^2 wskazuje, że należy odrzucić hipotezę o niezależności obecności nadciśnienia i ryzyka wystąpienia ChNS. Ryzyko wystąpienia ChNS istotnie zależy od obecności nadciśnienia ($p = 1.11 \times 10^{-8}$). Również podwójny test istotności dla frakcji wskazuje, że należy odrzucić hipotezę o równości frakcji w populacjach osób z nadciśnieniem i bez. Odstetek wystąpienia ChNS u osób z nadciśnieniem jest istotnie większy niż u osób bez nadciśnienia ($p = 5.5 \times 10^{-31}$). Różnice obrazuje poniższy wykres.



7.4 Wpływ ciśnienia tętniczego krwi na ryzyko wystąpienia ChNS

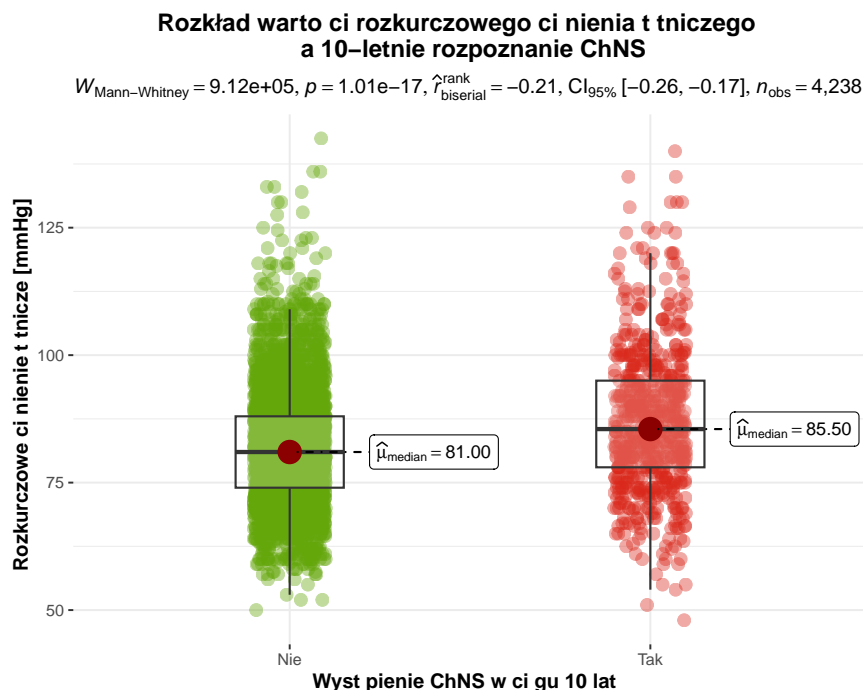
7.4.1 Wpływ skurczowego ciśnienia tętniczego krwi na ryzyko wystąpienia ChNS

Ponieważ rozkłady cechy **sysBP** w zależności od zmiennej **TenYearCHD** nie są normalne (dla obu wartości zmiennej w teście Shapiro-Wilka $p < 0.05$), zostanie użyty test Manna-Whitneya. Wykonany test wskazuje, że istnieją istotne różnice w rozkładzie skurczowego ciśnienia tętniczego w zależności od wystąpienia w ciągu 10 lat ChNS ($p = 9.82 \times 10^{-37}$). Skurczowe ciśnienie tętnicze krwi ma istotny wpływ na wystąpienie ChNS.



7.4.2 Wpływ rozkurczowego ciśnienia tętniczego krwi na ryzyko wystąpienia ChNS

Zachodzi podobna sytuacja jak w przypadku ciśnienia rozkurczowego (brak normalności rozkładów, $p < 0.05$). W związku z tym ponownie użyjemy testu Manna-Whitneya. Test sugeruje, że istnieją istotne różnice w rozkładzie rozkurczowego ciśnienia tętniczego w zależności od wystąpienia w ciągu 10 lat ChNS ($p = 1.01 \times 10^{-17}$). Rozkurczowe ciśnienie tętnicze krwi ma istotny wpływ na wystąpienie ChNS.



7.4.3 Zestawienie wyników

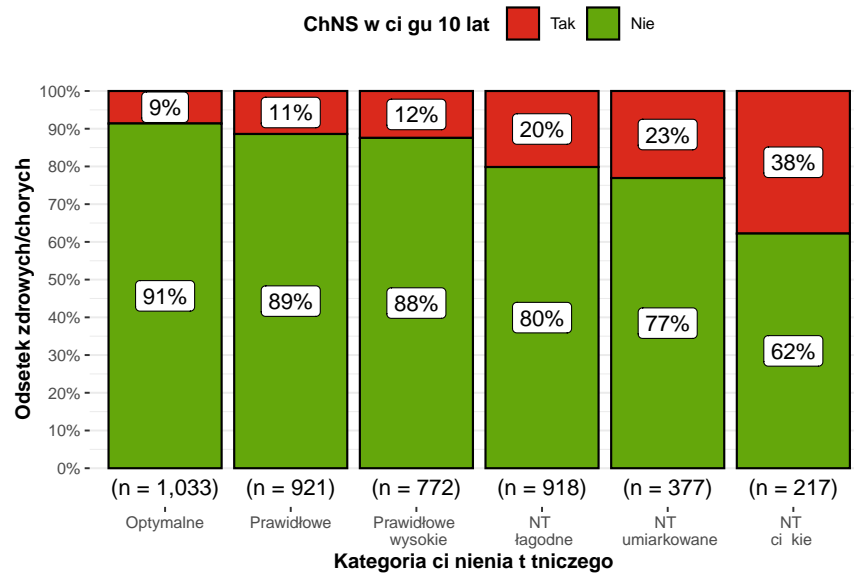
Dla lepszego zobrazowania różnic ryzyka zachorowania na ChNS możemy przedstawić dane o zachorowaniu w zależności od stopnia nasilenia nadciśnienia tętniczego. W tym celu podzielimy badaną grupę według następującego, obowiązującego obecnie podziału:

Kategoria	Ciśnienie tętnicze skurczowe (mm Hg)	Ciśnienie tętnicze rozkurczowe (mm Hg)
CIŚNIENIE TĘTNICZE		
Optymalne	<120	<80
Prawidłowe	120–129	80–84
Wysokie prawidłowe	130–139	85–89
NADCIŚNIENIE TĘTNICZE		
Stopień 1 (łagodne)	140–159	90–99
Stopień 2 (umiarkowane)	160–179	100–109
Stopień 3 (ciężkie)	≥180	≥110
Izolowane skurczowe	≥140	<90

Rysunek 1: klasyfikacja ciśnienia tętniczego

Odstetek 10-letnich rozpoznania ChNS dla różnych kategorii ciśnienia tętniczego

$\chi^2_{\text{Pearson}}(5) = 171.19$, $p = 4.07 \times 10^{-35}$, $\hat{V}_{\text{Cramer}} = 0.20$, $CI_{95\%} [0.17, 1.00]$, $n_{\text{obs}} = 4,238$

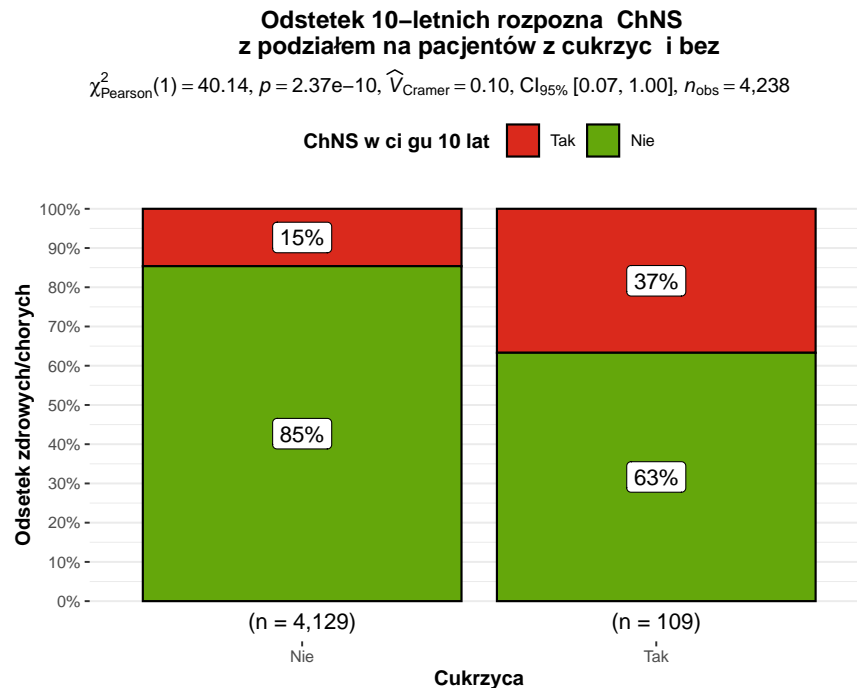


W przypadku tak ustalonych grup test χ^2 również sugeruje, że ryzyko zachorowania na ChNS istotnie zależy od kategorii ciśnienia tętniczego.

7.5 Cukrzyca a ryzyko ChNS

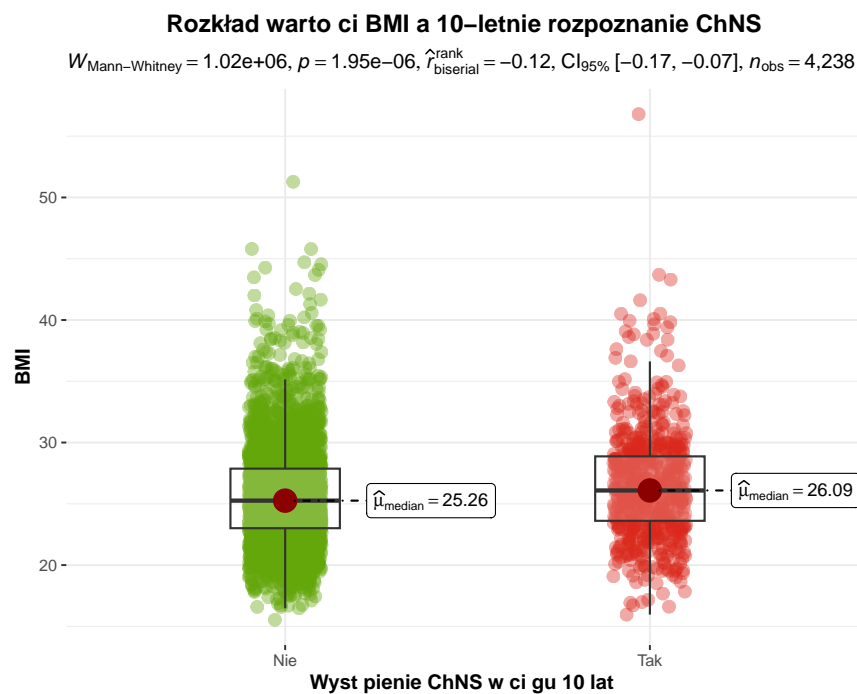
W badanym przypadku ponownie możemy użyć podwójnego testu istotności dla frakcji lub testu χ^2 .

Test χ^2 wskazuje, że należy odrzucić hipotezę o niezależności obecności cukrzycy i ryzyka wystąpienia ChNS. Ryzyko wystąpienia ChNS istotnie zależy od obecności cukrzycy ($p = 5.64 \times 10^{-10}$). Również podwójny test istotności dla frakcji wskazuje, że należy odrzucić hipotezę o równości frakcji w populacji cukrzyków i osób bez cukrzycy. Odstetek wystąpienia ChNS u cukrzyków jest istotnie większy niż u osób bez cukrzycy ($p = 2.82 \times 10^{-10}$). Różnice obrazuje poniższy wykres.



7.6 Wpływ BMI na ryzyko ChNS

Rozkład BMI nie jest rozkładem normalnym (test Shapiro-Wilka: $p < 0.05$). W związku z tym ponownie użyjemy testu Manna-Whitneya. Test sugeruje, że istnieją istotne różnice w rozkładzie BMI w zależności od wystąpienia w ciągu 10 lat ChNS ($p = 1.95 \times 10^{-6}$). BMI ma istotny wpływ na wystąpienie ChNS.



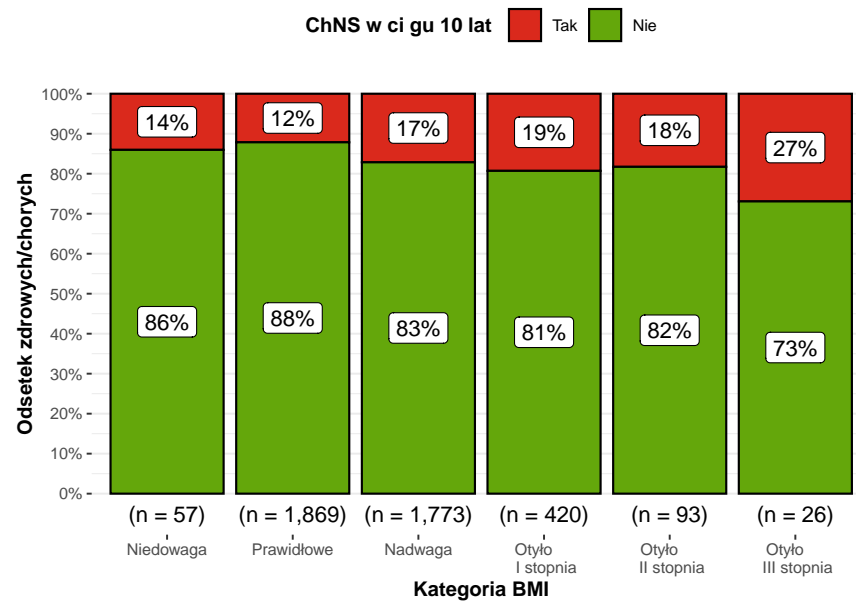
Dla lepszego zobrazowania różnic podzielimy badaną grupę według obecnie obowiązujących kategorii:

Kategoria	BMI (kg/m ²)
Niedowaga	<18,5
Prawidłowe	18,5–24,9
Nadwaga	25–29,9
Otyłość I stopnia	30–34,9
Otyłość II stopnia	35–39,9
Otyłość III stopnia	≥40

Rysunek 2: klasyfikacja BMI

Odstetek 10-letnich rozpoznania ChNS dla różnych kategorii BMI

$\chi^2_{\text{Pearson}}(5) = 27.70$, $p = 4.17 \times 10^{-5}$, $\hat{V}_{\text{Cramer}} = 0.07$, $CI_{95\%} [0.03, 1.00]$, $n_{\text{obs}} = 4,238$

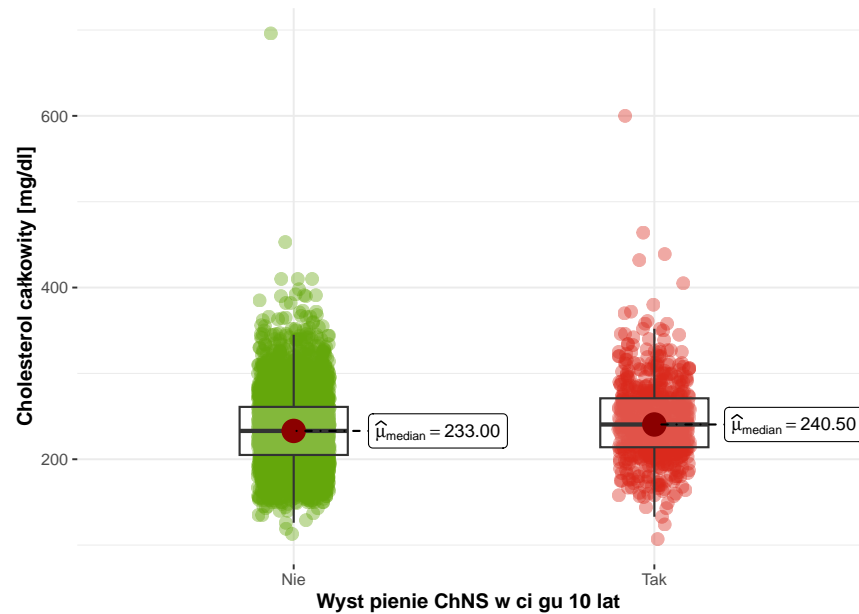


7.7 Stężenie cholesterolu całkowitego a ryzyko ChNS

Rozkład całkowitego stężenia cholesterolu nie jest rozkładem normalnym (w obu grupach w teście Shapiro-Wilka $p < 0.05$). W związku z tym tu również użyjemy testu Manna-Whitneya. Test sugeruje, że istnieją istotne różnice w rozkładzie całkowitego stężenia cholesterolu we krwi w zależności od wystąpienia w ciągu 10 lat ChNS ($p = 2.67 \times 10^{-7}$). Całkowite stężenie cholesterolu we krwi ma istotny wpływ na wystąpienie ChNS.

Rozkład stężenia cholesterolu całkowitego a 10-letnie rozpoznanie ChNS

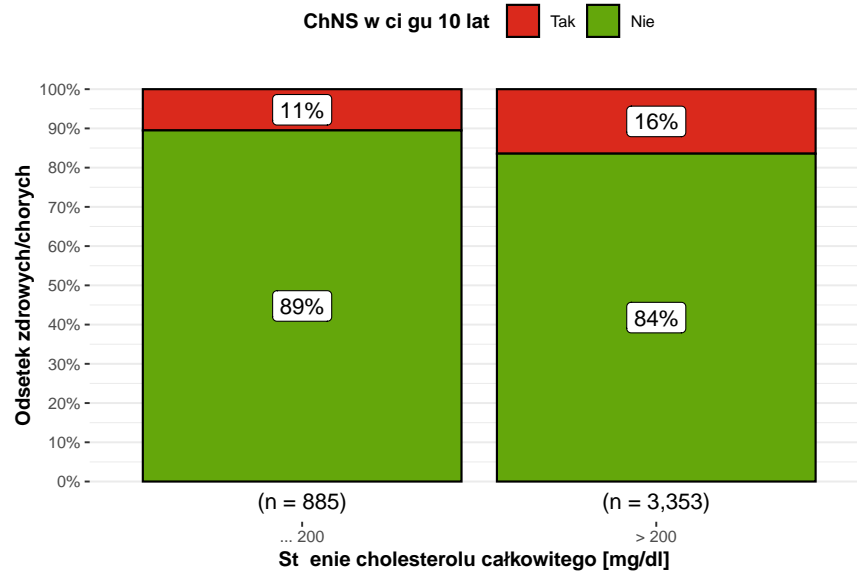
$W_{\text{Mann-Whitney}} = 1.01\text{e}+06$, $p = 2.67\text{e}-07$, $\hat{\rho}_{\text{rank biserial}} = -0.13$, $CI_{95\%} [-0.17, -0.08]$, $n_{\text{obs}} = 4,238$



Dane można również przedstawić w podziale na dwie grupy: o wartości stężenia cholesterolu całkowitego poniżej i powyżej górnej granicy normy (200 mg/dl).

Odsetek 10-letnich rozpoznania ChNS z podziałem na granicę normy stężenia cholesterolu całkowitego

$\chi^2_{\text{Pearson}}(1) = 19.07$, $p = 1.26\text{e}-05$, $\hat{V}_{\text{Cramer}} = 0.07$, $CI_{95\%} [0.04, 1.00]$, $n_{\text{obs}} = 4,238$

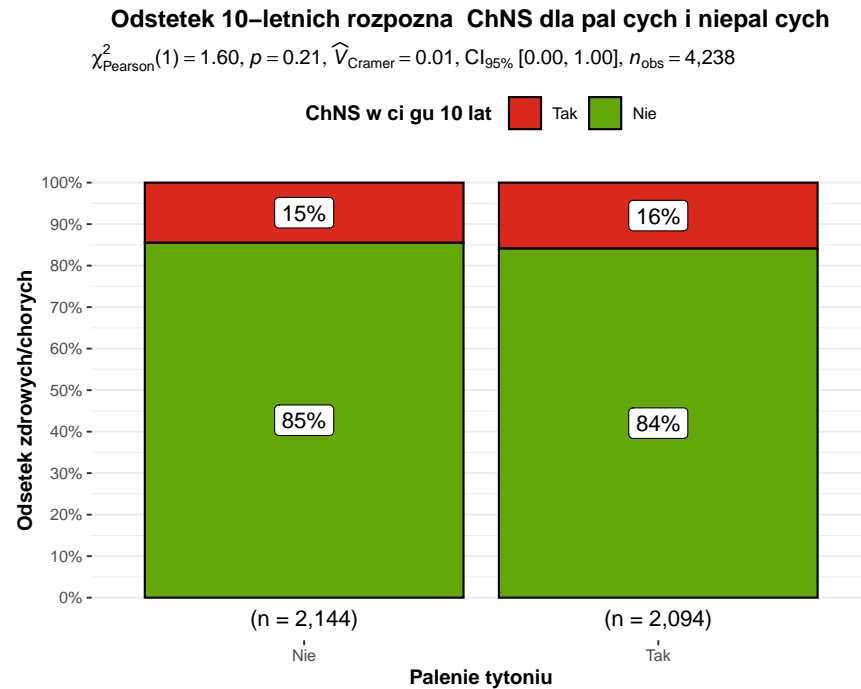


7.8 Palenie tytoniu a ryzyko ChNS

Tutaj również możemy użyć podwójnego testu istotności dla frakcji lub testu χ^2 .

Test χ^2 wskazuje, że nie ma podstaw do odrzucenia hipotezy o niezależności palenia tytoniu od ryzyka

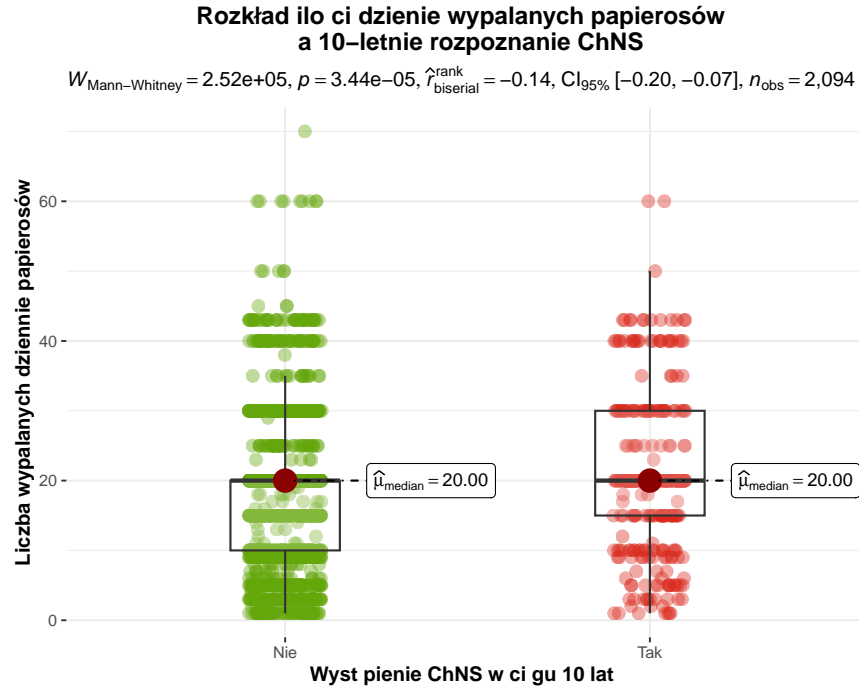
wystąpienia ChNS ($p = 0.221$). Również podwójny test istotności dla frakcji wskazuje, że nie należy odrzucać hipotezy o równości odsetka ChNS u palaczy i osób niepalących ($p = 0.111$). Wyniki możemy zobrazować na wykresie.



7.9 Liczba wypalanych papierosów a ryzyko ChNS

W badanym przypadku uwzględniono jedynie osoby deklarujące się jako palące.

W celu zbadania omawianej zależności użyjemy testu Manna-Whitney'a - rozkłady cechy `cigsPerDay` względem cechy `TenYearCHD` nie są rozkładami normalnymi (w obu grupach w teście Shapiro-Wilka $p < 0.05$). Test wskazuje, że należy odrzucić hipotezę o braku różnic w rozkładzie ilości dziennie wypalanych papierosów w grupach w których wystąpiła i nie wystąpiła ChNS ($p = 0.00662$). Różnice w rozkładach obrazuje poniższy wykres.



8 Wnioski

W badaniu wykazano, że ryzyko jej wystąpienia istotnie zwiększają:

- Wiek
- Płeć męska – OR = 1.63 ($\text{CI}_{95\%}$: 1.38 – 1.94)
- Nadciśnienie tętnicze – OR = 2.68 ($\text{CI}_{95\%}$: 2.25 – 3.17)
- Cukrzyca – OR = 3.39 ($\text{CI}_{95\%}$: 2.25 – 5.03)
- Otyłość - dla danych podzielonych na dwie grupy wartością BMI równą 30: OR = 1.42 ($\text{CI}_{95\%}$: 1.12 – 1.78)
- Wysokie stężenie cholesterolu - wyniki są istotne statystycznie, jednak w badanej grupie bezwzględne różnice w rozkładach stężenia cholesterolu są niewielkie, wynik jest niejednoznaczny. Różnice są lepiej widoczne dla danych podzielonych górną granicą normy dla stężenia cholesterolu całkowitego (200 mg/dl) - OR = 1.67 ($\text{CI}_{95\%}$: 1.33 – 2.12). Warto zauważyć, że jedynie niecała 1/4 (885 osób) badanej grupy ma stężenie cholesterolu będące poniżej tej granicy.

Nie wykazano istotnych różnic w ryzyku wystąpienia ChNS między osobami palącymi i niepalącymi. Z kolei liczba dziennie wypalanych papierosów przez palaczy miała istotny wpływ na ryzyko wystąpienia ChNS. Z tego powodu nie można wykluczyć negatywnego wpływu palenia tytoniu na wystąpienie ChNS.

Należy zaznaczyć, że wyniki nie powinny być uogólniane na populację polską - badania pochodzą z USA, a literatura wyróżnia populacje o różnym stopniu ryzyka, badana populacja niekoniecznie musi mieć tego samego stopnia ryzyka, co populacja polska. Wspomnieć należy również o jakości użytych danych – nie są między innymi znane dokładne informacje o zmiennych **sysBP** i **diaBP** – czy są to przeciętne wartości ciśnień,

czy też wartości ciśnień z pojedynczych pomiarów w gabinecie lekarskim. Podobnie z cechami `heartRate` i `glucose` (nie analizowanymi w tym badaniu).

Po przeanalizowaniu i wizualizacji powyższych danych można dojść do wniosku, że ocena istotności statystycznej jest przede wszystkim działaniem pomocniczym, a nie decyzyjnym. Wiele zależy również od sposobu przedstawienia danych – wpływ badanych czynników był lepiej widoczny dla danych pogrupowanych przedstawionych w formie wykresów słupkowych niż dla danych niezagregowanych w formie wykresów ramka-wąsy.

9 Bibliografia

1. <https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset>
2. „Umieralność w 2020 roku. Zgony według przyczyn – dane wstępne” GUS, 2021, <https://stat.gov.pl/obszary-tematyczne/ludnosc/statystyka-przyczyn-zgonow/umieralnosc-i-zgony-wedlug-przyczyn-w-2020-roku,10,1.html>
3. „Zachorowalność i umieralność na choroby układu krążenia a sytuacja demograficzna polski” Redakcja naukowa: Zbigniew Strzelecki, Janusz Szymborski, Rozdział 5, https://bip.stat.gov.pl/files/gfx/bip/pl/zamowieniapubliczne/426/248/1/81_gp_rrl_2015_monografia_kardiologiczna.pdf
4. „NFZ o zdrowiu Choroba niedokrwienna serca” , Warszawa, kwiecień 2020, Centrala Narodowego Funduszu Zdrowia, Departament Analiz i Innowacji
5. „Kardiologia” pod redakcją Tadeusza Mandeckiego, Warszawa, 2005
6. „Eksploracja danych”, Dariusz Majerek, Katedra Matematyki Stosowanej, Wydział Podstaw Techniki, Politechnika Lubelska 2020-03-11, <https://dax44.github.io/datamining/>
7. „Nadciśnienie tętnicze”, <https://www.mp.pl/interna/chapter/B16.II.2.20>
8. <https://www.mp.pl/nadwaga-i-otylosc/wytyczne/246952,-nadwaga-i-otylosc-u-doroslych>

Spis treści

- 1 Wstęp
- 2 Materiały i metody
- 3 Opisanie zmiennych
- 4 Ustalenie hipotez badawczych
- 5 Przygotowanie danych
 - 5.1 Cecha `education`
 - 5.2 Cechy `currentSmoker` i `cigsPerDay`
 - 5.3 Pozostałe braki danych
- 6 Podstawowe informacje o badanej grupie

7 Weryfikacja hipotez

- 7.1 Wpływ wieku na ryzyko wystąpienia ChNS
- 7.2 Płeć a ryzyko wystąpienia ChNS
- 7.3 Nadciśnienie a ryzyko ChNS
- 7.4 Wpływ ciśnienia tętniczego krwi na ryzyko wystąpienia ChNS
 - 7.4.1 Wpływ skurczowego ciśnienia tętniczego krwi na ryzyko wystąpienia ChNS
 - 7.4.2 Wpływ rozkurczowego ciśnienia tętniczego krwi na ryzyko wystąpienia ChNS
 - 7.4.3 Zestawienie wyników
- 7.5 Cukrzyca a ryzyko ChNS
- 7.6 Wpływ BMI na ryzyko ChNS
- 7.7 Stężenie cholesterolu całkowitego a ryzyko ChNS
- 7.8 Palenie tytoniu a ryzyko ChNS
- 7.9 Liczba wypalanych papierosów a ryzyko ChNS

8 Wnioski

9 Bibliografia