

Politechnika Wrocławska

Wydział Matematyki

| | |
|--------------------------|---|
| Skład grupy: | Agata Sobczak 268873 Jakub Franczak 262271 Katarzyna Kudelko 268762 |
| Prowadząca laboratorium: | dr inż. Aleksandra Grzesiek |
| Prowadząca wykładu: | dr hab. Alicja Jokiel-Rokita |

Analiza Danych Ankietowych

Raport 2.

Lista 2.

Spis treści

| | | |
|----------|---|-----------|
| 1 | Zadanie 1. | 3 |
| 1.1 | Cel zadania | 3 |
| 1.2 | <i>Fisher.test</i> i <i>prop.test</i> | 3 |
| 1.3 | Wnioski | 3 |
| 2 | Zadanie 2. | 4 |
| 2.1 | Cel zadania | 4 |
| 2.2 | (a) Czy skuteczność leczenia jest niezależna od wielkości dawki? | 4 |
| 2.3 | (b) Czy skuteczność leczenia jest niezależna od rodzaju leku? | 4 |
| 2.4 | (c) Czy skuteczność leczenia jest niezależna od miejsca leczenia? | 5 |
| 3 | Zadanie 3. | 5 |
| 3.1 | Cel zadania | 5 |
| 3.2 | <i>chisq.test</i> | 6 |
| 3.3 | Wnioski | 7 |
| 4 | Zadanie 4. | 7 |
| 4.1 | Cel zadania | 7 |
| 4.2 | Wnioski | 8 |
| 5 | Zadanie 5. | 8 |
| 5.1 | Cel zadania | 8 |
| 5.2 | Reakcja a Dawka | 8 |
| 5.3 | Reakcja a Miejsce | 8 |
| 6 | Zadanie 6. | 9 |
| 6.1 | Cel zadania | 9 |
| 6.2 | Współzmienność | 9 |
| 6.3 | Analiza korespondencji | 10 |
| 6.3.1 | Kod | 11 |
| 6.3.2 | Wyniki | 13 |
| 6.4 | Wnioski | 15 |
| 7 | Zadanie 7. | 15 |
| 7.1 | Cel zadania | 15 |
| 7.2 | Wyniki | 15 |
| 7.3 | Wnioski | 18 |

1 Zadanie 1.

1.1 Cel zadania

Celem zadania jest zweryfikowanie czy na poziomie istotności 0.05 są podstawy do odrzucenia hipotezy o niezależności temperatury i wystąpienia uszkodzeń w promach kosmicznych.

Hipoteza zerowa- prawdopodobieństwo wystąpienia uszkodzenia promu kosmicznego jest takie same dla temperatury otoczenia dla co najwyżej $65^{\circ}F$ i powyżej $65^{\circ}F$.

Hipoteza alternatywna- zależność wystąpienia uszkodzenia promu kosmicznego dla temperatury otoczenia dla co najwyżej $65^{\circ}F$ i powyżej $65^{\circ}F$.

| | Brak uszkodzeń | Obecność uszkodzeń |
|---------------------|----------------|--------------------|
| Do 65 stopni F | 0 | 4 |
| Powyżej 65 stopni F | 17 | 3 |

Tabela 1: Tabela dwudzielcza dla zmiennych temperatura i obecność uszkodzeń

1.2 *Fisher.test* i *prop.test*

```
#fisher.test() i prop.test()

starty_wahadlowców <- matrix(c(0,4,17,3), nrow = 2, byrow = TRUE,
  dimnames = list(Temperatura = c("Do 65 stopni F", "Powyżej 65 stopni F"),
    uszkodzenia_pierścieni = c("Brak uszkodzeń", "obecność uszkodzeń")))

fisher.test(starty_wahadlowców, conf.level = 0.95)

prop.test(starty_wahadlowców, conf.level = 0.95)
```

| | <i>fisher.test</i> | <i>prop.test</i> |
|------------|--------------------|------------------|
| p- wartość | 0.003294 | 0.004927 |

Tabela 2: p - wartość testów *fisher.test* i *prop.test*

1.3 Wnioski

Tabela 2 pokazuje, że oba testy mają podobną p - wartość, która na poziomie istotności $\alpha = 0.05$ daje podstawy do odrzucenia hipotezy o niezależności temperatury otoczenia i wystąpienia uszkodzeń pierścieni na promie kosmicznym.

2 Zadanie 2.

2.1 Cel zadania

Celem zadania jest analiza danych zawartych w pliku *Reakcja.csv*, które obejmują informacje o 200 reakcjach na lek.

Informacje obejmują:

- *Reakcja* (0 - nie nastąpiła poprawa, 1 - nastąpiła poprawa),
- *Dawka* (w skali logarytmicznej),
- *Rodzaj* (0 - pierwsza firma farmaceutyczna, 1- druga firma farmaceutyczna)
- *Miejsce* (0 - dom, 1- szpital)

```
# wczytywanie danych
dane <- read.csv("Reakcja.csv", sep = ";")
dane
```

2.2 (a) Czy skuteczność leczenia jest niezależna od wielkości dawki?

| Reakcja | Dawka | | | | | SUMA |
|---------|-------|--------|--------|--------|--------|------|
| | -2 | -2.301 | -2.602 | -2.903 | -3.204 | |
| 0 | 21 | 25 | 32 | 32 | 37 | 147 |
| 1 | 19 | 15 | 8 | 8 | 3 | 53 |
| SUMA | 40 | 40 | 40 | 40 | 40 | 200 |

Tabela 3: Tabela wielodzielcza dla *Reakcja* i *Dawka*

```
skuteczność_dawka <- matrix(c(21,25,32,32,37,19,15,8,8,3), nrow = 2, byrow = TRUE,
  dimnames = list(Reakcja = c("Nie nastąpiła poprawa", "Nastąpiła poprawa"),
  Dawka = c("Dawka:-2", "Dawka:-2,301", "Dawka:-2,602", "Dawka:-2,903", "Dawka:-3,204")))
fisher.test(skuteczność_dawka)
```

p - wartość wynosi 0.0002993, co na poziomie istotności $\alpha = 0.05$ daje podstawy do odrzucenia hipotezy o niezależności skuteczności leczenia od wielkości dawki.

2.3 (b) Czy skuteczność leczenia jest niezależna od rodzaju leku?

| Reakcja | Rodzaj | | SUMA |
|---------|--------|-----|------|
| | 0 | 1 | |
| 0 | 76 | 71 | 147 |
| 1 | 24 | 29 | 53 |
| SUMA | 100 | 100 | 200 |

Tabela 4: Tabela wielodzielcza dla *Reakcja* i *Rodzaj*

```
skuteczność_rodzaj <- matrix(c(76,71,24,29), nrow = 2, byrow = TRUE,
                             dimnames = list(Reakcja = c("Nie nastąpiła poprawa", "Nastąpiła poprawa"),
                             Rodzaj = c("0", "1")))

fisher.test(skuteczność_rodzaj)
```

p -wartość wynosi 0.5218, co na poziomie istotności $\alpha = 0.05$ nie daje podstawy do odrzucenia hipotezy o niezależności skuteczności leczenia od rodzaju.

2.4 (c) Czy skuteczność leczenia jest niezależna od miejsca leczenia?

| Skuteczność | Miejsce | | SUMA |
|-------------|---------|-----|------|
| | 0 | 1 | |
| 0 | 86 | 61 | 147 |
| 1 | 14 | 39 | 53 |
| SUMA | 100 | 100 | 200 |

Tabela 5: Tabela wielodzielcza dla *Reakcja* i *Miejsce*

```
skuteczność_miejsce <- matrix(c(86,61,14,39), nrow = 2, byrow = TRUE,
                               dimnames = list(Reakcja = c("Nie nastąpiła poprawa", "Nastąpiła poprawa"),
                               Miejsce = c("0", "1")))

fisher.test(skuteczność_miejsce)
```

p -wartość wynosi $9.773e - 05$, co na poziomie istotności $\alpha = 0.05$ daje podstawy do odrzucenia hipotezy o niezależności skuteczności leczenia od miejsca.

3 Zadanie 3.

3.1 Cel zadania

Celem zadania jest zweryfikowanie hipotezy o niezależności stopnia zadowolenia z pracy i wynagrodzenia, korzystając z funkcji *chisq.test* na podstawie podanych danych.

| Tablica 1: Dane do zadania 3.i 6. | | | | | |
|-----------------------------------|-----------------------------|-----------|--------|-----------|------|
| Wynagrodzenie | Stopień zadowolenia z pracy | | | | Suma |
| | b. niezadow. | niezadow. | zadow. | b. zadow. | |
| poniżej 6000 | 32 | 44 | 60 | 70 | 206 |
| 6000-15000 | 22 | 38 | 104 | 125 | 289 |
| 15000-25000 | 13 | 48 | 61 | 113 | 235 |
| powyżej 25000 | 3 | 18 | 54 | 96 | 171 |
| Suma | 62 | 108 | 319 | 412 | 901 |

Rysunek 1: Tablica zawierająca dane o stopniu zadowolenia z pracy i wynagrodzenia

3.2 *chisq.test*

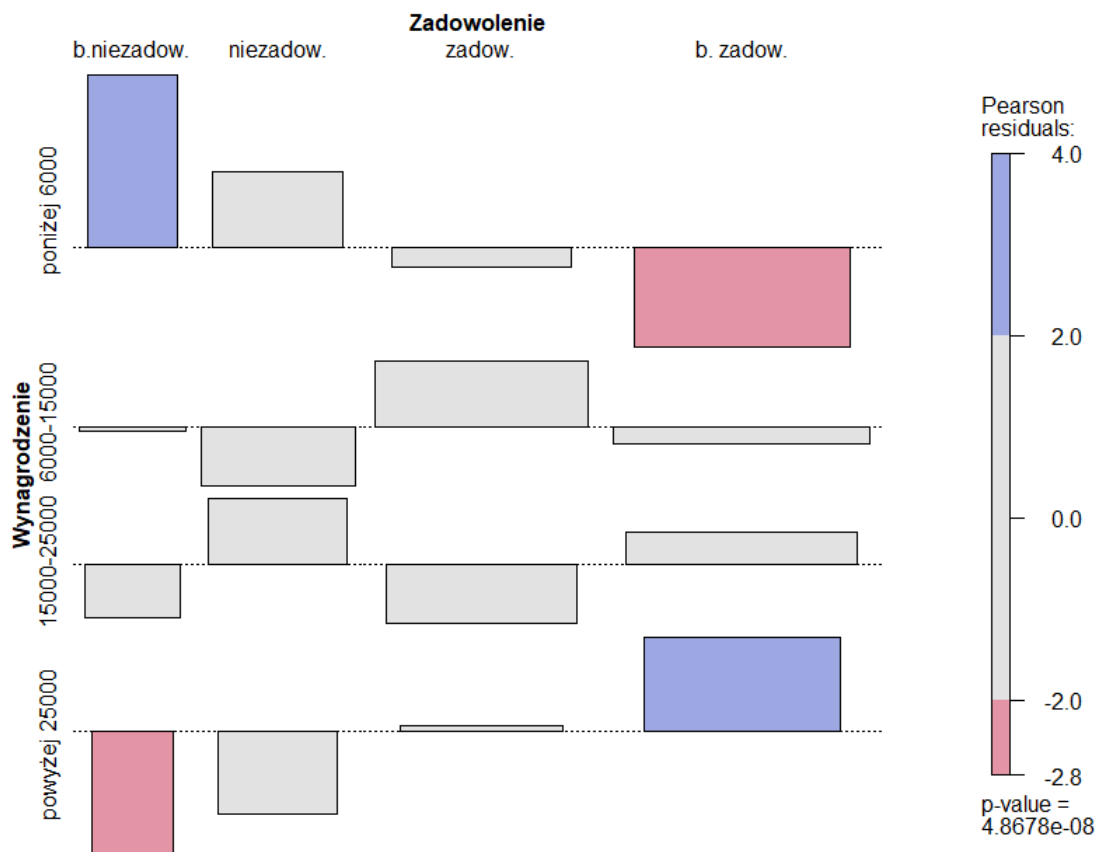
```
wynagrodzenie_zadowolenie <- matrix(c(32,44,60,70,22,38,104,125,13,48,61,113,3,18,54,96), nrow = 4, byrow = TRUE,
  dimnames = list(wynagrodzenie = c("poniżej 6000", "6000-15000", "15000-25000", "powyżej 25000"),
  zadowolenie = c("b.niezadow.", "niezadow.", "zadow.", "b. zadow.")))

chisq.test(wynagrodzenie_zadowolenie)
chisq.test(wynagrodzenie_zadowolenie, simulate.p.value = TRUE, B = 1000)
chisq.test(wynagrodzenie_zadowolenie, simulate.p.value = TRUE, B = 3000)
chisq.test(wynagrodzenie_zadowolenie, simulate.p.value = TRUE, B = 5000)
chisq.test(wynagrodzenie_zadowolenie, simulate.p.value = TRUE, B = 10000)

assoc(wynagrodzenie_zadowolenie, shade = TRUE)
```

| Ilość powtórzeń w teście Monte Carlo | 1000 | 3000 | 5000 | 10000 | | |
|--------------------------------------|----------|-----------|-------|-----------|--------------------|-----------|
| p- wartość | 0.000999 | 0.0003332 | 2e-04 | 9.999e-05 | p-wartość dokładna | 4.868e-08 |

Tabela 6: Tabela dla p -wartości w zależności od ilości powtórzeń Monte Carlo



Rysunek 2: Wykres asocjacji wskazujący odchylenia od określonego modelu niezależności

3.3 Wnioski

Dokładna p -wartość wynosi $4.8678e - 08$ co na poziomie istotności $\alpha = 0.05$ daje podstawy do odrzucenia hipotezy zerowej o niezależności zadowolenia w pracy i wynagrodzenia. Tabela 6 pokazuje, że w zależności od ilości powtórzeń w teście Monte Carlo p -wartość zbliża się do wartości dokładnej.

4 Zadanie 4.

4.1 Cel zadania

Celem zadania jest obliczenie wartości poziomu krytycznego (p -value) w teście niezależności opartym na ilorazie wiarygodności dla danych z tablicy 1 na liście zadań (tabela została już zaprezentowana przy okazji zadania 3)

```
# Definicja danych
ponizej_6000 <- c(32, 44, 60, 70)
miedzy_6000_i_15000 <- c(22, 38, 104, 125)
miedzy_15000_i_25000 <- c(13, 48, 61, 113)
powyzej_25000 <- c(3, 18, 54, 96)

# Tworzenie macierzy wynagrodzeń
macierz_wynagrodzen <- matrix(
  c(ponizej_6000,
    miedzy_6000_i_15000,
    miedzy_15000_i_25000,
    powyzej_25000),
  nrow = 4,
  byrow = TRUE
)
row.names(macierz_wynagrodzen) <- c("<6000", "[6000,15000]", "[15000,25000]", ">25000")
colnames(macierz_wynagrodzen) <- c("b. niezadow.", "niezadow.", "zadow.", "b. zadow.")

# Dodanie sum wierszy i kolumn
macierz_wynagrodzen <- cbind(macierz_wynagrodzen, rowSums(macierz_wynagrodzen))
macierz_wynagrodzen <- rbind(macierz_wynagrodzen, colSums(macierz_wynagrodzen))

# Funkcja obliczająca p-value
count_p_value <- function(data) {
  if (!is.matrix(data)) {
    stop("Podana dane nie są dwuwymiarową tablicą.")
  }

  results <- 1

  for (a in 1:4) {
    for (b in 1:4) {
      results <- results * (
        macierz_wynagrodzen[a, 5] * macierz_wynagrodzen[5, b] /
        (macierz_wynagrodzen[5, 5] * macierz_wynagrodzen[a, b])
      ) ^ macierz_wynagrodzen[a, b]
    }
  }

  g2 <- -2 * log(results)
  return(1 - pchisq(g2, df = 16))
}

# wywołanie funkcji
wynik <- count_p_value(macierz_wynagrodzen)

# wyświetlenie wyniku
print(wynik)
print(macierz_wynagrodzen)
```

Rysunek 3: Kod do wykonania obliczeń

4.2 Wnioski

Wyliczona wartość poziomu krytycznego wynosi $9.403163e-06$, więc jest to mniej niż 0.05, co skłania nas do odrzucenia hipotezy zerowej.

5 Zadanie 5.

5.1 Cel zadania

Celem zadania jest obliczenie miar współzmienności pomiędzy zmiennymi Reakcja (skuteczność leczenia) i Dawka (wielkość dawki), a także między zmiennymi Reakcja a Miejsce (miejsce leczenia) na podstawie danych zawartych w pliku Reakcja.csv. Następnie należy zinterpretować uzyskane wartości tych miar współzmienności.

5.2 Reakcja a Dawka

Jak wiadomo z punktu (2.2), p - wartość wynosi 0.0002993, co na poziomie istotności $\alpha = 0.05$ daje podstawy do odrzucenia hipotezy o niezależności skuteczności leczenia od wielkości dawki. Ponieważ nie są badane dwie zmienne porządkowe, użytą miarą współzmienności jest τ . Współczynnik ten został wyliczony za pomocą funkcji GoodmanKruskalTau z biblioteki DescTools.

```
tau_Reakcja_Dawka_C <- GoodmanKruskalTau(skutecznosć_dawka, direction="column")
print(paste("Współzmiennosć między Reakcją a Dawką (kolumna):", tau_Reakcja_Dawka_C))

tau_Reakcja_Dawka_R <- GoodmanKruskalTau(skutecznosć_dawka, direction="row")
print(paste("Współzmiennosć między Reakcją a Dawką (wiersz):", tau_Reakcja_Dawka_R))
```

Rysunek 4: Kod wykonujący obliczenia współczynnika współzmienności

```
[1] "Współzmiennosć między Reakcją a Dawką (kolumna): 0.0258631754588627"
[1] "Współzmiennosć między Reakcją a Dawką (wiersz): 0.103452701835451"
> |
```

Rysunek 5: Uzyskane wyniki

Ponieważ współczynnik τ jest niesymetryczny, (przyjmuje różne wartości w zależności od przyjętej zmiennej zależnej), policzono średnią arytmetyczną obu wyników.

| τ_C | τ_R | τ_{sr} |
|--------------------|-------------------|-----------------|
| 0.0258631754588627 | 0.103452701835451 | 0.0646579386425 |

Powyższy wynik jest dosyć mały, sugeruje to słabą zależność zmiennych.

5.3 Reakcja a Miejsce

Ponownie wracając do wcześniejszych punktów (2.4) p -wartość wynosi $9.773e - 05$, co na poziomie istotności $\alpha = 0.05$ daje podstawy do odrzucenia hipotezy o niezależności skuteczności leczenia od miejsca. Ponieważ badane są dwie zmienne nominalne, użytą miarą współzmienności ponownie jest τ . Kod jest analogiczny do przypadku (4.2).

| τ |
|--------------------|
| 0.0802207675523039 |

Tabela 7: Otrzymany współczynnik


```
[1] "Współzmiennność między Reakcją a Miejscem (kolumna): 0.0802207675523039"
[1] "Współzmiennność między Reakcją a Miejscem (wiersz): 0.0802207675523039"
> |
```

Rysunek 6: Uzyskane wyniki

Miara współzmienności jest symetryczna oraz równa 0.0802207675523039. Taki wynik oznacza niewielką współzmiennność zmiennych Reakcji i Miejsca. Można przyjąć, że reakcja na leczenie jest słabo zależna od miejsca leczenia.

6 Zadanie 6.

6.1 Cel zadania

Cel zadania to analiza współzmienności pomiędzy zmiennymi "Wynagrodzenie" i "Stopień zadowolenia z pracy" na podstawie danych zawartych w tablicy 1 (Rysunek 1). Następnie należy przeprowadzić analizę korespondencji, która obejmuje obliczenie wartości odpowiednich macierzy oraz współrzędnych punktów umożliwiających wizualizację danych na wykresach.

6.2 Współzmiennność

W tabeli umieszczone są zmienne porządkowe, zatem jako miarę współzmienności obliczono współczynnik γ . Został on wyliczony za pomocą funkcji GoodmanKruskalGamma z biblioteki DescTools. Skorzystano z tabeli wynagrodzenie_zadowolenie z punktu (3.2).

```
gamma_wynagrodzenie_zadowolenie_C <- GoodmanKruskalGamma(wynagrodzenie_zadowolenie, direction="column")
print(paste("Współzmiennność między Wynagrodzeniem a Zadowoleniem (kolumna):", gamma_wynagrodzenie_zadowolenie_C))

gamma_wynagrodzenie_zadowolenie_R <- GoodmanKruskalGamma(wynagrodzenie_zadowolenie, direction="row")
print(paste("Współzmiennność między Wynagrodzeniem a Zadowoleniem (wiersz):", gamma_wynagrodzenie_zadowolenie_R))
```

Rysunek 7: Kod potrzebny do wyliczenia współczynnika

```
[1] "Współzmiennność między wynagrodzeniem a Zadowoleniem (kolumna): 0.218101953188693"
[1] "Współzmiennność między wynagrodzeniem a Zadowoleniem (wiersz): 0.218101953188693"
> |
```

Rysunek 8: Uzyskane wyniki

Współzmiennność sprawdzono zarówno dla kolumn, jak i wierszów. Wyniki są symetryczne, $\gamma = 0.218101953188693$. Oczywiście $\gamma \in [-1, 1]$ oraz w tym przypadku jest to miara zależności dodatniej. Zmienne wykazują niewielką zależność.

6.3 Analiza korespondencji

Skorzystano z podejścia zaproponowanego przez Greenacre'a, polegającego na dekompozycji według wartości osobliwych macierzy.

$$A = D_r^{-1/2}(P - rc^T)D_c^{-1/2},$$

gdzie:

- A - macierz rezyduów standaryzowanych,
- P - macierz korespondencji,
- $a_{ij} = \frac{p_{ij} - r_i c_j}{\sqrt{r_i c_j}}$,
- $r_i = n_{i+}/n$ jest i -tą współrzędną wektora r ,
- $c_j = n_{+j}/n$ jest j -tą współrzędną wektora c .

Poniżej przedstawiono P - macierz korespondencji:

| | b.niezadow. | niezadow. | zadow. | b. zadow. | r |
|---------------|-------------|-----------|---------|-----------|---------|
| poniżej 6000 | 0.03552 | 0.04883 | 0.06659 | 0.07769 | 0.22863 |
| 6000-15000 | 0.02442 | 0.04218 | 0.11543 | 0.13873 | 0.32076 |
| 15000-25000 | 0.01443 | 0.05327 | 0.06770 | 0.12542 | 0.26082 |
| powyżej 25000 | 0.00333 | 0.01998 | 0.05993 | 0.10655 | 0.18979 |
| c | 0.07770 | 0.16426 | 0.30965 | 0.44839 | 1 |

Tabela 8: Tabela częstotliwości wraz z rozkładami brzegowymi

gdzie:

kolumna r - wektor przeciętnego profilu wierszowego, a wiersz c - wektor przeciętnego profilu kolumnowego.

Następnie obliczono macierze będące współrzędnymi kategorii cech, odpowiednio dla wierszy oraz kolumn:

$$F = D_r^{-1/2}U\Gamma, \quad G = D_c^{-1/2}V\Gamma$$

```

[[1]]
      [,1]      [,2]      [,3]      [,4]
[1,] 0.133203867 0.05819913 -0.015805310 -0.07753757
[2,] -0.003183739 -0.04579816 0.051097476 -0.01341809
[3,] -0.040991673 0.05039544 -0.045963056 0.02475694
[4,] -0.094008538 -0.06341766 0.004801831 0.07352502

[[2]]
      [,1]      [,2]      [,3]      [,4]
[1,] -0.35660100 -0.003711145 0.06630346 1.547229e-09
[2,] 0.03666254 -0.206508071 -0.08067414 1.832609e-09
[3,] 0.04611606 0.247727011 -0.05895610 1.652551e-09
[4,] 0.28967393 -0.017870033 0.10121872 1.409675e-09

[[3]]
      [,1]      [,2]      [,3]      [,4]
[1,] -0.35768330 -0.07965395 0.08081084 -9.019241e-10
[2,] -0.17935236 0.21906919 -0.07458282 -1.311450e-09
[3,] 0.03160571 -0.21050349 -0.08017514 -1.800624e-09
[4,] 0.23117646 0.07549559 0.07813112 -2.166764e-09

```

Rysunek 9: Macierze kolejno: A, F, G

Wyliczono także inercję całkowitą i porównano ją z wbudowaną funkcją.

6.3.1 Kod

```
dane <- matrix(c(32, 44, 60, 70,
                22, 38, 104, 125,
                13, 48, 61, 113,
                3, 18, 54, 96), nrow = 4, byrow = TRUE)

rownames(dane) <- c("poniżej 6000", "6000-15000", "15000-25000", "powyżej 25000")
colnames(dane) <- c("b.niezadow.", "niezadow.", "zadow.", "b. zadow.")
```

Rysunek 10: Inicjacja macierzy

```
analiza_korespondencji <- function(dane, rysuj = TRUE){
  P <- dane / sum(dane)
  suma_r <- rowSums(P)
  suma_k <- colSums(P)
  D_r <- diag(suma_r)
  D_k <- diag(suma_k)
  A <- solve(sqrt(D_r)) %*% (P - suma_r %*% t(suma_k)) %*% solve(sqrt(D_k))
  U <- svd(A)$u
  V <- svd(A)$v

  F <- U %*% diag(sqrt(svd(A)$d))
  G <- V %*% diag(sqrt(svd(A)$d))

  if (rysuj){
    x_F <- F[, 1]
    y_F <- F[, 2]
    x_G <- G[, 1]
    y_G <- G[, 2]
    plot(x_F, y_F,
         xlim = c(min(c(x_F, x_G)) - 0.1, max(c(x_F, x_G)) + 0.1),
         ylim = c(min(c(y_F, y_G)) - 0.1, max(c(y_F, y_G)) + 0.1),
         col = "blue",
         xlab = "Dimension 1", ylab = "Dimension 2",
         main = "Analiza Korespondencji metoda własna")

    points(x_G, y_G, col = "red", pch = 17)
    abline(v = 0, h = 0, col = c("black", "black"))
    text(F[, 1], F[, 2], labels = rownames(dane))
    text(G[, 1], G[, 2], labels = colnames(dane))
  }

  return(list(F, G))
}

analiza_korespondencji(dane, rysuj = TRUE)
```

Rysunek 11: Funkcja wykonująca algorytm analizy korespondencji

```

#wbudowana
analiza_korespondencji_wbud <- ca(dane)
summary(analiza_korespondencji_wbud)

#wykresy
plot(analiza_korespondencji_wbud, col.row = "red", col.col = "blue",
     main = "Analiza Korespondencji metoda wbudowana")

#Inercja
inercja <- function(table){
  P <- table/sum(table)
  r <- apply(P, 1, sum)
  c <- apply(P, 2, sum)

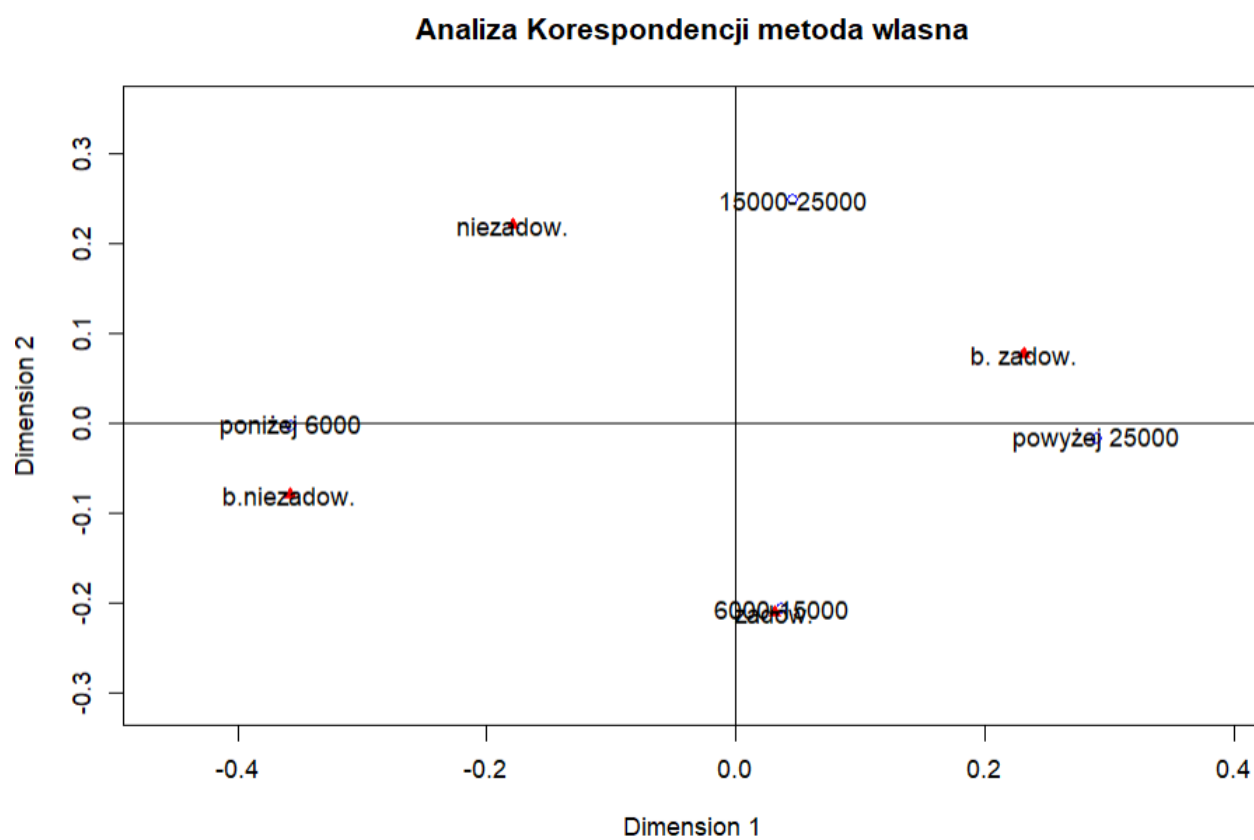
  D_r <- diag(r)
  D_c <- diag(c)
  A <- solve(sqrt(D_r)) %*% (P - r %*% t(c)) %*% solve(sqrt(D_c))

  Gamma <- svd(A)$d
  lambda <- sum(Gamma^2)
  return(lambda)
}
cat("Funkcji własna:", inercja(dane))
cat("Funkcji wbudowana:", sum(analiza_korespondencji_wbud$colinertia))

```

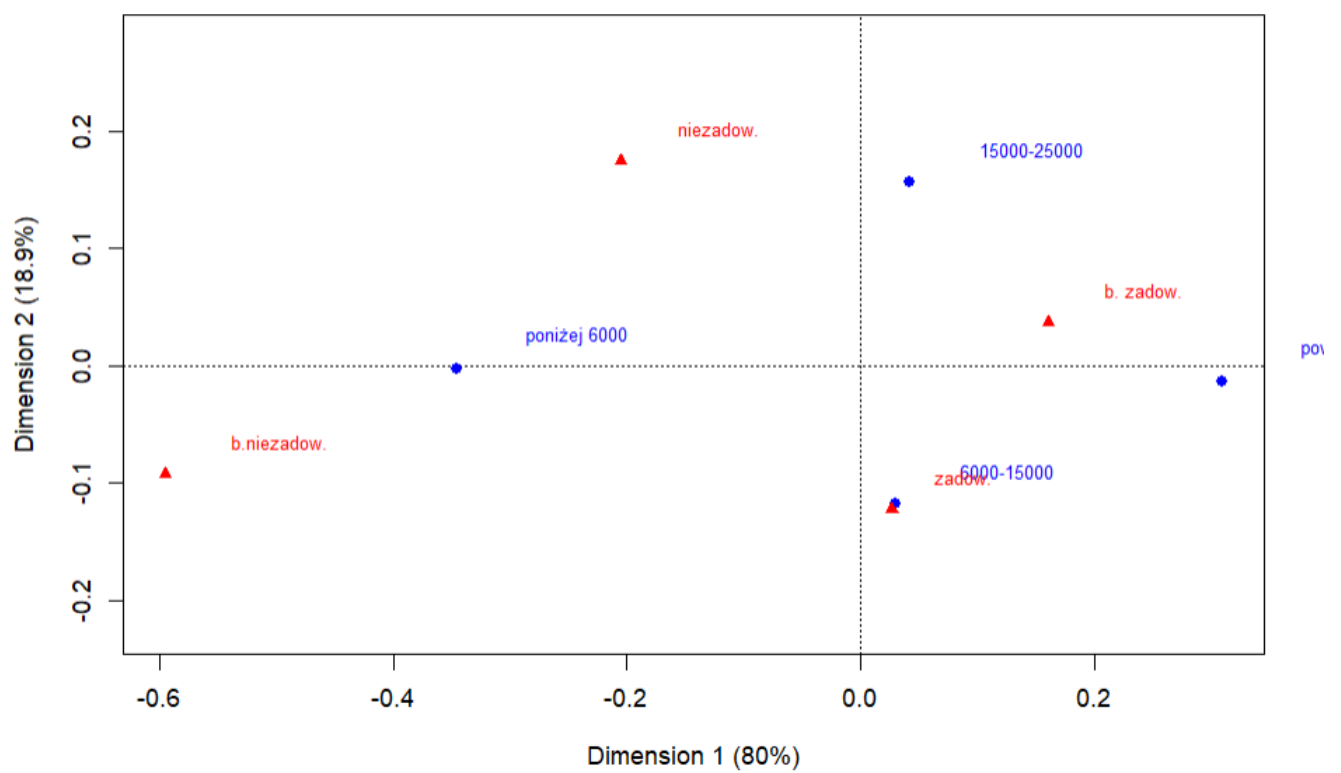
Rysunek 12: Kod realizujący pozostałe obliczenia

6.3.2 Wyniki



Rysunek 13: Wykres analizy korespondencji z użyciem metody własnej

Analiza Korespondencji metoda wbudowana



Rysunek 14: Wykres analizy korespondencji z użyciem metody wbudowanej

| | |
|--------------------------------------|------------|
| Inercja z funkcji własnej: | 0.05752481 |
| Inercja z funkcji wbudowanej: | 0.05752481 |

Tabela 9: Porównanie wyników

6.4 Wnioski

Otrzymany wykres za pomocą własnej funkcji jest zbliżony do tego stworzonego przy użyciu funkcji wbudowanej, co potwierdza poprawność jej działania oraz zgodność otrzymanych wcześniej wyników z wynikami otrzymanymi poprzez wbudowane funkcje z bibliotek do analizy korespondencji. Można również zauważyć konkretne zależności:

- Osoby zarabiające powyżej 25000 są w znacznej większości zadowolone z zarobków,
- rozkład zadowolenia wśród osób poniżej 6000 jest stosunkowo równomierny wyłączając opcję b.zadowolony,
- osoby w przedziale 6000 – 15000 są w zdecydowanej większości zadowolone i b.zadowolone

Inercja całkowita wynosi 0.05752481, co oznacza małe rozproszenie profili. Można potwierdzić to patrząc na rysunek wykresu analizy korespondencji, gdzie większość punktów jest skupiona niemalże kółkiem wokół punktu (0, 0).

7 Zadanie 7.

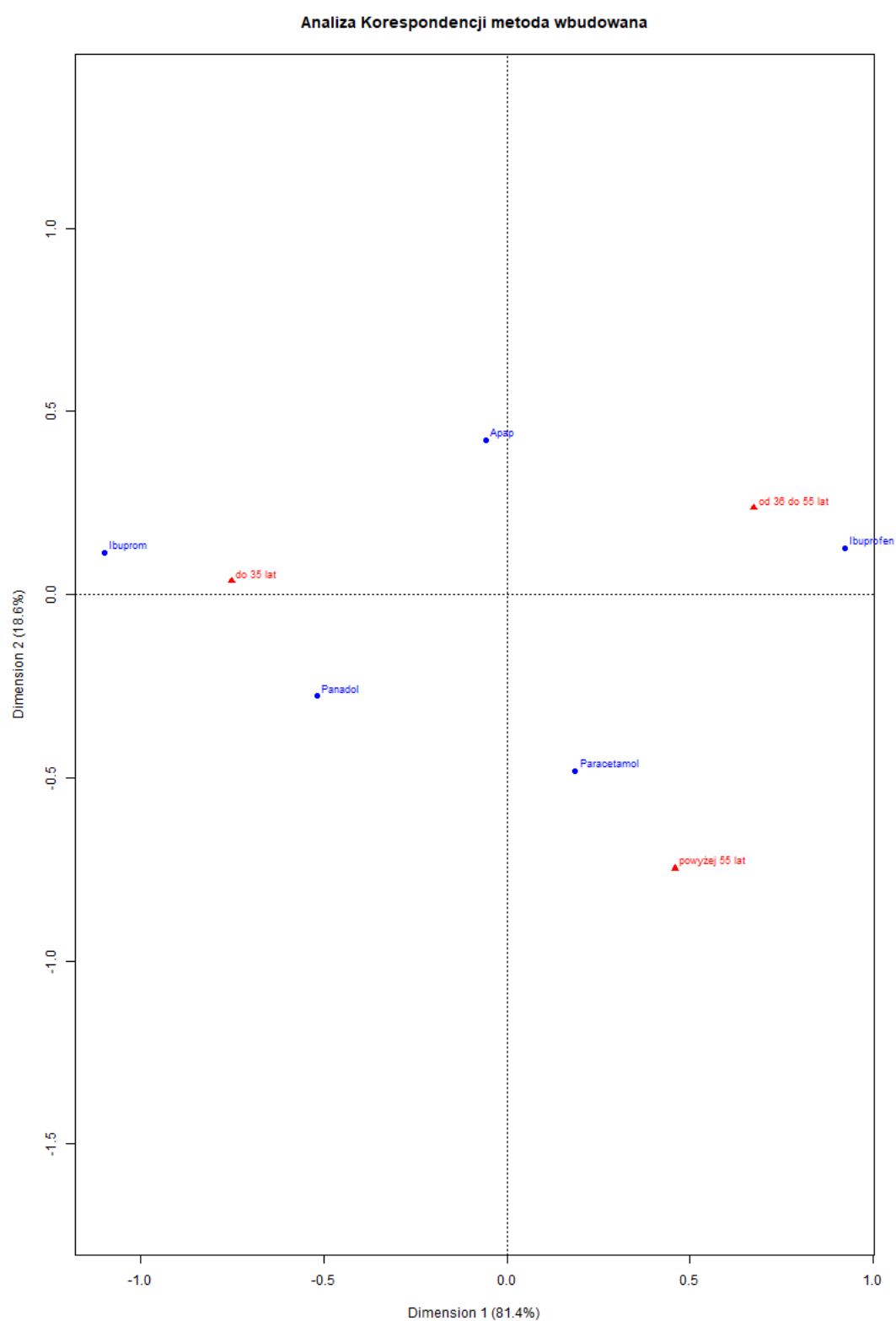
7.1 Cel zadania

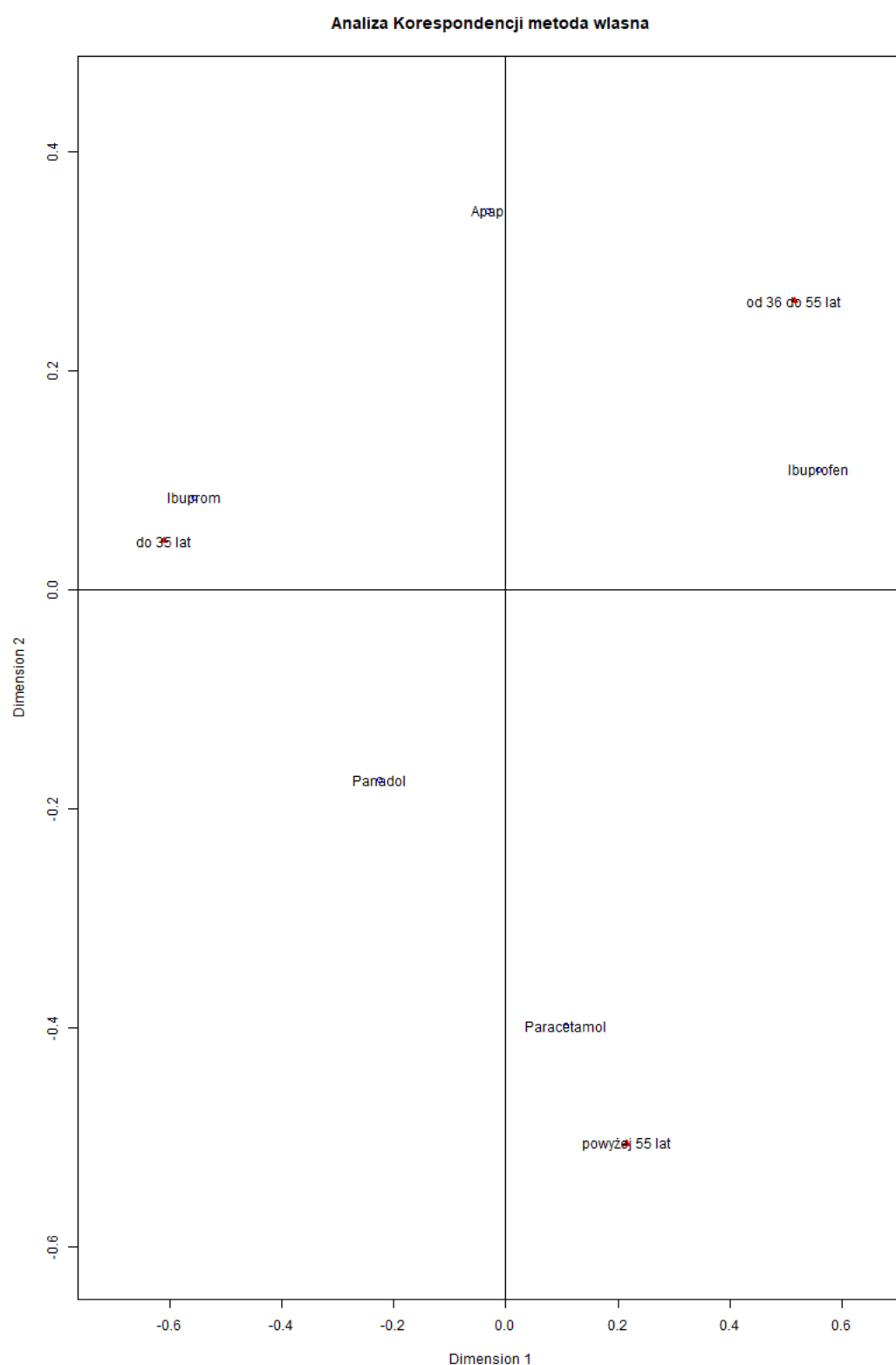
Celem zadania jest obliczenie odpowiednich miar współzmienności oraz przeprowadzenie analizy korespondencji, tzn. obliczenie wartości odpowiednich macierzy, współrzędnych punktów oraz utworzenie odpowiednich wykresów.

W zadaniu skorzystano z analogicznego kodu jak w zadaniu 6 (sekcja 6), a miara współzmienności została wyliczona za pomocą funkcji GoodmanKruskalTau z pakietu DescTools.

7.2 Wyniki

| | |
|--------------------------------------|-----------|
| Inercja z funkcji własnej: | 0.5748397 |
| Inercja z funkcji wbudowanej: | 0.574839 |
| Miara współzmienności: | 0.3477173 |





7.3 Wnioski

W przypadku wartości miary współzmienności $= 0.3477173$ można stwierdzić, że istnieje pewna zależność między badanymi zmiennymi, ale nie jest to związek bardzo silny.

Jeśli natomiast chodzi o wartość inercji wynoszącej w przybliżeniu 0.57, to jest ona stosunkowo wysoka. Można to interpretować jako to, że istnieje pewna zmienność w danych, co może sugerować istnienie pewnych zależności między analizowanymi zmiennymi w tablicy korespondencji.

Z wykresów analizy korespondencji można wyciągnąć takie wnioski:

- Ibuprom jest najczęściej wybierany przez osoby poniżej 35 roku życia,
- Apap jest stosunkowo równomiernie wybierany we wszystkich grupach wiekowych
- Paracetamol jest częściej wybierany przez osoby w wieku powyżej 55 lat,
- Ibuprofen jest preferowany głównie przez osoby między 36 a 55 rokiem życia,
- Panadol jest stosunkowo równo rozłożony w różnych grupach wiekowych.