Analiza Danych Ankietowych
Sprawozdanie 2

Natalia Lach 262303, Alicja Myśliwiec 262275

Matematyka Stosowana
Wydział Matematyki Politechniki Wrocławskiej

2023

Spis treści

1. Zadanie 2

Celem zadania jest zweryfikowanie hipotezy zerowej, w której płeć i zajmowane stanowiska (kierownicze lub niekierownicze) są niezależne. Hipotezą alternatywną jest zależność tych dwóch zmiennych. W tym celu użyto funkcji *fisher.test*. Odpowiadającą jej hipotezą zerową jest ta, w której prawdopodobieństwo, że na stanowisku kierowniczym pracuje kobieta równe jest prawdopodobieństwu, że na stanowisku kierowniczym pracuje mężczyzna. Hipotezą alternatywną jest nierówność obu tych prawdopodobieństw. Do zweryfikowania tejże hipotezy użyto funkcji *prop.test*. Działano na parametrach otrzymanych z danych, które przedstawiono w poniższej tabeli 1.

P \ S	0	1
K	63	8
M	110	19

Tab. 1: Tablica dwudzielcza zmiennych S i P.

Używając wcześniej wymienionych funkcji, otrzymano następujące wyniki testów.

Metoda	*fisher.test*	*prop.test*
p-wartość	0.6659	0.6389

Tab. 2: p- wartości testów dotyczących zmiennych S i P.

Z tabeli 2 wynika, że oba testy zwróciły dość podobną p-wartość, która na poziomie istotności $\alpha = 0.05$ nie daje podstaw do odrzucenia hipotez zerowych na rzecz alternatywnych. Zatem nie ma podstaw do stwierdzenia, że rozważane zmienne nie są niezależne ani że ich rozkład nie jest taki sam.

2. Zadanie 3

Do wykonania każdego z poniższych podpunktów użyto funkcji *fisher.test* z odpowiednimi parametrami przedstawionymi w tabelach 3 i 4.

2.1. Podpunkt a.)

Hipoteza zerowa jest postaci - zajmowanie stanowiska kierowniczego nie zależy od wieku. Hipotezą alternatywną jest zależność tych dwóch zmiennych.

S \ Wiek	1	2	3	4
0	23	91	39	20
1	3	13	6	5

Tab. 3: Tablica dwudzielcza zmiennych Wiek i S.

Na podstawie danych z tabeli 3 otrzymano p-wartość $= 0.7823$.

Powyższy wynik nie daje podstaw do odrzucenia hipotezy zerowej na rzecz alternatywnej - nie ma podstaw do stwierdzenia, że zajmowanie stanowiska kierowniczego jest zależne od wieku.

2.2. Podpunkt b.)

Hipoteza zerowa jest postaci - zajmowanie stanowiska kierowniczego nie zależy od wykształcenia. Hipotezą alternatywną jest zależność tych dwóch zmiennych.

S \ Wyk	1	2	3
0	40	123	10
1	1	17	9

Tab. 4: Tablica dwudzielcza zmiennych Wyk i S.

Na podstawie danych z tabeli 4 otrzymano p-wartość $= 6.538 \cdot 10^{-5}$.

Powyższy wynik jest mniejszy od ustalonego poziomu ufności $\alpha = 0.05$, a więc istnieją podstawy do odrzucenia hipotezy zerowej na rzecz alternatywnej - przyjmujemy, że zajmowanie stanowiska kierowniczego jest zależne od wykształcenia.

3. Zadanie 4

Podobnie jak w zadaniu 3 zostanie użyta funkcja *fisher.test*. W każdym przypadku poziom istotności wynosi $\alpha = 0.05$. Tym razem jednak badana będzie niezależność zmiennej W1 względem innych zmiennych.

3.1. Podpunkt a.)

Hipoteza zerowa jest postaci - zadowolenie z wynagrodzenia nie zależy od zajmowanego stanowiska. Hipotezą alternatywną jest zależność tych dwóch zmiennych.

S \ W1	-2	-1	1	2
0	64	18	0	91
1	10	2	2	13

Tab. 5: Tablica dwudzielcza zmiennych W1 i S.

Na podstawie danych z tabeli 5 otrzymano p-wartość $= 0.0443$.

Na podstawie otrzymanej p-wartości mamy podstawy do odrzucenia hipotezy o niezależności zmiennych. Przyjmujemy więc, że poziom zadowolenia z wynagrodzenia zależy od zajmowanego stanowiska.

3.2. Podpunkt b.)

Hipoteza zerowa jest postaci - zadowolenie z wynagrodzenia nie zależy od wykształcenia. Hipotezą alternatywną jest zależność tych dwóch zmiennych.

Wyk \ W1	-2	-1	1	2
1	20	3	0	18
2	45	17	0	78
3	9	0	2	8

Tab. 6: Tablica dwudzielcza zmiennych W1 i Wyk.

Na podstawie danych z tabeli 6 otrzymano p-wartość $= 0.01069$.

Ponownie otrzymana p-wartość jest mniejsza niż zadany poziom istotności, dlatego odrzucamy hipotezę zerową na rzecz alternatywnej, według której wynagrodzenie zależy od wykształcenia.

3.3. Podpunkt c.)

Hipoteza zerowa jest postaci - zadowolenie z wynagrodzenia nie zależy od płci. Hipotezą alternatywną jest zależność tych dwóch zmiennych.

P \ W1	-2	-1	1	2
K	25	10	1	35
M	49	10	1	69

Tab. 7: Tablica dwudzielcza zmiennych W1 i P.

Na podstawie danych z tabeli 7 otrzymano p-wartość $= 0.4758$.

Powyższy wynik jest duże większy niż ustalony poziom istostności α. Nie ma zatem podstaw do odrzucenia hipotezy zerowej, mówiącej o niezależności zadowolenia z wynagrodzenia i płci.

3.4. Podpunkt d.)

Hipoteza zerowa jest postaci - zadowolenie z wynagrodzenia nie zależy od wieku. Hipotezą alternatywną jest zależność tych dwóch zmiennych.

Wiek \ W1	-2	-1	1	2
1	9	1	0	16
2	42	9	1	52
3	12	6	0	27
4	11	4	1	9

Tab. 8: Tablica dwudzielcza zmiennych W1 i Wiek.

Na podstawie danych z tabeli 8 otrzymano p-wartość $= 0.31935$.

Po raz kolejny p-wartość przekracza 0.05, dlatego też nie ma podstaw do odrzucenia hipotezy zerowej na rzecz hipotezy alternatywnej.

4. Zadanie 6

Ponownie, jak w podpunkcie 3.1, bazowano na danych dotyczących zadowolenia z wynagrodzenia oraz zajmowanego stanowiska. Tablica dwudzielcza obu zmiennych przedstawiona jest poniżej w tabeli 9.

S \ W1	-2	-1	1	2
0	64	18	0	91
1	10	2	2	13

Tab. 9: Tablica dwudzielcza zmiennych W1 i S.

I tym razem badaną hipotezą zerową jest to, że zadowolenie z wynagrodzenia (w pierwszym badanym okresie) nie zależy od zajmowanego stanowiska. Hipoteza alternatywna dotyczy zależności obu zmiennych. Korzystając z testu chi-kwadrat Pearsona (*chisq.test*) oraz testu chi-kwadrat ilorazu wiarogodności (*assocstats*), wyznaczono p-wartości rozważanych testów.

Test	Chi-kwadrat Pearsona	Chi-kwadrat ilorazu wiarogodności
p-wartość	0.004397	0.0396896

Tab. 10: p- wartości testów dotyczących zmiennych W1 i S.

Na podstawie wyników z tabeli 10 można zauważyć dość dużą różnicę w otrzymanych wartościach poziomów krytycznych. Dla ustalonego dla tego zadania poziomu istotności $\alpha = 0.01$ przyjęte zostaną dwie różne hipotezy w zależności od zastosowanego testu. P-wartość testu Pearsona jest mniejsza od 0.01, zatem należy odrzucić hipotezę zerową na rzecz alternatywnej - zadowolenie z wynagrodzenia (w pierwszym badanym okresie) zależy od zajmowanego stanowiska. W przypadku testu chi-kwadrat ilorazu wiarogodności otrzymany wynik nie daje podstaw do odrzucenia hipotezy zerowej, gdyż p-wartość > 0.01.

$$\underline{p\text{-wartość} = 0.0443.}$$

Te same hipotezy były rozważane i weryfikowane w zadaniu 4a) w sekcji 3.1, lecz na innym poziomie istotności - to jest $\alpha = 0.05$. Po użyciu *fisher.test* odrzucono hipotezę zerową na rzecz alternatywnej, gdyż p-wartość wyniosła 0.0443. Gdyby jednak brać pod uwagę $\alpha = 0.01$, otrzymano by wniosek taki jak w przypadku testu chi-kwadrat ilorazu wiarogodności - nie istniałyby podstawy do stwierdzenia, że zadowolenie z wynagrodzenia zależy od zajmowanego stanowiska.

5. Zadanie 7

Wykonując niniejsze zadanie, posłużono się następującym kodem.

```r
sym <- function(n, p, N=5000, pval=0.05) {
    fisher <- pearson <- iloraz <- c()
    for (i in 1:N) {
        a <- rmultinom(1, n, p)
        data <- matrix(a, nrow=2)
        fisher <- append(fisher, fisher.test(data)$p.val)
        pearson <- append(pearson, chisq.test(data)$p.val)
        iloraz <- append(iloraz, assocstats(data)[2]$chisq_tests[1,3])
    }
    c(sum(fisher < pval)/N, sum(pearson < pval)/N, sum(iloraz < pval)/N)
}
```

W zależności od podpunktu, jako argumenty funkcji podawano odpowiednie n oraz wektory prawdopodobieństwa p, które odpowiadały odpowiednim hipotezom. Parametry takie jak ilość powtórzeń symulacji ($N = 5000$) oraz poziom istotności ($pval = 0.05$) są takie same dla obu podpunktów.

5.1. Podpunkt a.)

W niniejszym podpunkcie należy oszacować rozmiar testu. To znaczy przeprowadzić symulację i znaleźć prawdopodobieństwo popełnienia błędu I rodzaju dla trzech wspominanych wcześniej testów - to jest testu Fishera, testu chi-kwadrat Pearsona oraz testu ilorazu wiarogodności. Aby to zrobić, należy sprawdzić ile razy odrzucono hipotezę zerową, kiedy była ona prawdziwa.

Y \ X	x_1	x_2	
y_1	0.05	0.45	0.5
y_2	0.05	0.45	0.5
	0.1	0.9	

Tab. 11: Rozkład łączny wraz z rozkładami brzegowymi X i Y.

Aby stwierdzić niezależność zmiennych X i Y, można posłużyć się twierdzeniem mówiącym, że zmienne X i Y są niezależne wtedy i tylko wtedy, gdy

$$p_{ij} = p_i \cdot p_j \quad \forall ij.$$

Podstawiając dane z tabeli 11 otrzymujemy

$$P(X = x_1, Y = y_1) = 0.05 = P(X = x_1)P(Y = y_1),$$
$$P(X = x_1, Y = y_2) = 0.05 = P(X = x_1)P(Y = y_2),$$
$$P(X = x_2, Y = y_1) = 0.45 = P(X = x_2)P(Y = y_1),$$
$$P(X = x_2, Y = y_2) = 0.45 = P(X = x_2)P(Y = y_2),$$

co oznacza, że zmienne są niezależne, co potwierdza słuszność hipotezy zerowej.

test \ n	50	100	1000
Fisher	0.0184	0.0258	0.0462
chi-kwadrat Pearsona	0.009	0.0164	0.0404
chi-kwadrat ilorazu wiarogodności	0.0828	0.0548	0.0522

Tab. 12: Rozmiary testów w zależności od długości próby.

Dla testu chi-kwadrat ilorazu wiarogodności rozmiar testu wyszedł większy niż zadany poziom $\alpha = 0.05$. Jest on też testem dla którego prawdopodobieństwo popełnienia błędu I rodzaju jest największe. Występuje to niezależnie od rozmiaru próby, jednak dla większych otzrymany wynik coraz bardziej zbliża sie do poziomu α, tak jak widać to w tabeli 12 przykładowo dla próby o rozmiarze 1000. Dla pozostałych dwóch testów ich rozmiar jest już dużo mniejszy, jednak podobnie do testu ilorazu wiarogodności - wraz z większym rozmiarem próby wynik jest bliższy zadanej wartości α.

5.2. Podpunkt b.)

W niniejszym podpunkcie należy oszacować moc testu. To znaczy przeprowadzić symulację i sprawdzić, ile razy odrzucono hipotezę zerową, kiedy była ona fałszywa.

Y \ X	x_1	x_2	
y_1	$\frac{1}{40}$	$\frac{19}{40}$	$\frac{20}{40}$
y_2	$\frac{3}{40}$	$\frac{17}{40}$	$\frac{20}{40}$
	$\frac{4}{40}$	$\frac{36}{40}$	

Tab. 13: Rozkład łączny wraz z rozkładami brzegowymi X i Y.

W tym przypadku możemy wykluczyć niezależność zmiennych X i Y. Posłuży do tego kontrprzykład, który nie spełnia warunków wyżej przedstawionego twierdzenia, mianowicie

$$P(X = x_1, Y = y_1) = 0.025 \neq 0.05 = P(X = x_1)P(Y = y_1).$$

Dzięki temu możemy stwierdzić, że hipoteza zerowa o niezależności zmiennych jest fałszywa. Dlatego też, przedstawiony wcześniej kod, tym razem posłuży nam do szacowania mocy testu.

Otrzymano następujące wyniki.

test \ n	50	100	1000
Fisher	0.1158	0.3128	0.9998
chi-kwadrat Pearsona	0.0706	0.2528	0.9998
chi-kwadrat ilorazu wiarogodności	0.2616	0.4094	0.9998

Tab. 14: Moce testów w zależności od długości próby.

Analizując otrzymane wyniki przedstawione w tabeli 14, można stwierdzić, że dla małej wielkości próby każdy z testów z dużym prawdopodobieństwem nie odrzuci fałszywej hipotezy zerowej. Nie pozwala to jednoznacznie stwierdzić, który z nich jest istotnie lepszy. Jednak warto zauważyć, że test ilorazu wiarogodności ma tendencję do osiągania wyższej mocy w porównaniu do testu Fishera i testu chi-kwadrat Pearsona. Pokrywa się to z analizą w podpunkcie 5.1, ponieważ mniejsze ryzyko popełnienia błędu I rodzaju zwiększa ryzyko popełnienia błedu II rodzaju, tj. maleje badana w naszym przypadku moc testu. Dla próby rozmiaru 1000, wszystkie trzy testy mają moc wynoszącą 0.998. Ta wartość jest bardzo wysoka i zbliżona do 1, co oznacza, że testy mają bardzo silną zdolność do wykrycia fałszywej hipotezy zerowej oraz odrzucenia jej.

6. Zadanie 8

W zależności od charakteru zmiennych, jako miary współzmienności zmiennych będą liczone współczynnik τ lub współczynnik γ. Pomogą w tym wbudowane funckje pakietu *DescTools* w języku R - odpowiednio *GoodmanKruskalTau* i *GoodmanKruskalGamma*.

6.1. podpunkt a.)

Miara współzmienności zmiennych - W1 (zadowolenie z wynagrodzenia w pierwszym badanym okresie) i S (zajmowane stanowisko), zostanie wyliczona na podstawie danych z tabeli 15.

S \ W1	-2	-1	1	2
0	64	18	0	91
1	10	2	2	13

Tab. 15: Tablica dwudzielcza zmiennych W1 i S.

Korzystając z podpunktu 3.1, wiemy że p-wartość $= 0.0443 < 0.05$, więc odrzucamy hipotezę o niezależności zmiennych - możemy wyliczyć odpowiednie miary. Tylko jedna ze zmiennych jest porządkowa (zmienna W1), dlatego liczymy współczynnik τ. Ponieważ współczynnik ten jest niesymetryczny, osobno został on policzony dla kolumn i wierszy (argument *direction*). Aby osiągnąć bardziej ogólny wskaźnik dla zmiennych, została wyznaczona średnia arytmetyczna obu wyników.

Współczynnik	τ_{column}	τ_{row}	$\frac{\tau_{\text{column}}+\tau_{\text{row}}}{2}$
Wartość	0.0016773	0.0655684	0.0336229

Tab. 16: Współczynniki zmienności zmiennych W1 i S.

Współczynnik τ osiąga wartości z przedziału $[0, 1]$, gdzie wartość 0 jest osiągana tylko i wyłącznie w przypadku, gdy badane zmienne są niezależne. Otrzymana średnia $\tau \approx 0.0336$, jest bliska zeru. Oznacza to słabą współzmienność zmiennych W1 i S. Zatem zadowolenie z wynagrodzenia i zajmowane stanowisko są od siebie bardzo słabo zależne.

6.2. podpunkt b.)

Miara współzmienności zmiennych - W1 (zadowolenie z wynagrodzenia w pierwszym badanym okresie) i Wyk (wykształcenie), zostanie wyliczona na podstawie danych z tabeli 17.

Wyk \ W1	-2	-1	1	2
1	20	3	0	18
2	45	17	0	78
3	9	0	2	8

Tab. 17: Tablica dwudzielcza zmiennych W1 i Wyk.

Wiemy, że p-wartość $= 0.01069 < 0.05$ (3.2), więc odrzucamy hipotezę o niezależności zmiennych. Obie zmienne są porządkowe, więc tym razem policzony zostanie współczynnik γ.

$$\gamma = 0.090843$$

Współczynnik γ w porównaniu do τ, przyjmuje już wartości z przedziału $[-1, 1]$. Ponieważ otrzymano wynik dodatni, uznajemy gammę za miarę zależności dodatniej. Wynik ten jednak jest dość niewielki, co sugeruje, że ogólnie osoby z wyższym wykształceniem mogą wykazywać nieznacznie wyższy poziom zadowolenia z wynagrodzenia w porównaniu do osób z niższym wykształceniem.

6.3. podpunkt c.)

Miara współzmienności zmiennych - S (zajmowane stanowisko) i Wyk (wykształcenie).

Wyk \ S	0	1
1	40	1
2	123	17
3	10	9

Tab. 18: Tablica dwudzielcza zmiennych S i Wyk.

Korzystając z testu Fishera otrzymano p-wartość $= 6.538 \cdot 10^{-05} < 0.05$, więc odrzucamy hipotezę o niezależności zmiennych. Tylko jedna ze zmiennych jest porządkowa, ponownie więc liczymy τ, korzystając z tabeli 18.

Współczynnik	τ_{column}	τ_{row}	$\frac{\tau_{\text{column}}+\tau_{\text{row}}}{2}$
Wartość	0.1158995	0.0305937	0.0732466

Tab. 19: Współczynniki zmienności zmiennych S i Wyk.

Przedstawiony średni wynik w tabeli 19 wyniósł ≈ 0.0735, co może sugerować słabą zależność zajmowania stanowisk kierowniczych od wykształcenia tych pracowników. Obie zmienne prawie w ogóle nie są współzmienne.

7. Zadanie 10

W zadaniu zostanie przeprowadzona analiza korespondencji zmiennych W1 i Wyk na podstawie podejścia polegającego na dekompozycji macierzy rezyduów standaryzowanych według jej wartości osobliwych. Wygląda ona następująco:

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

gdzie

— P - macierz korespodencji,
— r, c - wektory częstości brzegowych.

Poniżej przedstawiono macierz P - to znaczy tabelę częstotliwościową, wraz z rozkładami brzegowymi.

Wyk \ W1	-2	-1	1	2	r
1	0.100	0.015	0.000	0.090	0.205
2	0.225	0.085	0.000	0.390	0.700
3	0.045	0.000	0.010	0.040	0.095
c	0.370	0.100	0.010	0.520	

Tab. 20: Rozkład łączny wraz z rozkładami brzegowymi W1 i Wyk.

W tabeli 20 kolumna r odpowiada za sumę wartości w danym wierszu. Otrzymujemy tym sposobem wektor nazywany przeciętnym profilem wierszowym. Analogicznie wektor c - przeciętnym profilem kolumnowym.

Zależy nam na wyznaczeniu współrzędnych kategorii cech, tj. znaleźć macierze F oraz G, które dane są wzorami

$$F = D_r^{-1/2} U \Gamma, \tag{2}$$

$$F = D_c^{-1/2} V \Gamma. \tag{3}$$

gdzie U, Γ i V są elementami rozkładu macierzy A według wartości osobliwych.

$$A = U \Gamma V^T \tag{4}$$

Poniższy kod implementuje owe podejście i zwraca szukane macierze F oraz G, według podanych wyżej wzorów.

```
zad10 <- function(data, colname1="W1", colname2 = "Wyk") {
    p2 <- ftable(data, col.vars=colname1, row.vars=colname2)/nrow(data)
    r <- rowSums(p2)
    c <- colSums(p2)
    Dr <- diag(r)
    Dc <- diag(c)
    A <- sqrtm(solve(Dr)) %*% (p2-matrix(r) %*% c) %*% sqrtm(solve(Dc))
    URV <- svd(A)
    U <- URV$u
    V <- URV$v
    gamma <- diag(URV$d)
    F <- solve(sqrtm(Dr)) %*% U %*% gamma
    F <- F[,1:2]
    G <- solve(sqrtm(Dc)) %*% V %*% gamma
    G <- G[,1:2]
    return(list(F,G))
}
```

Możemy znaleźć te same wartości używając wbudowanych funkcji języka R.

```
dane <- as.data.frame.matrix(ftable(data, col.vars="W1", row.vars="Wyk"))
res.ca <- CA(dane, graph = FALSE)
F <- res.ca$row$coord
G <- res.ca$col$coord
```

Posłużyły one jedynie do weryfikacji wyników powstałych po implementacji własnej funkcji. Otrzymano tym sposobem szukane macierze A, F oraz G.

$$A = \begin{bmatrix} 0.08768783 & -0.03841367 & -0.04527693 & -0.05084280 \\ -0.06680809 & 0.05669467 & -0.08366600 & 0.04309458 \\ 0.05253800 & -0.09746794 & 0.29362077 & -0.04229260 \end{bmatrix}$$

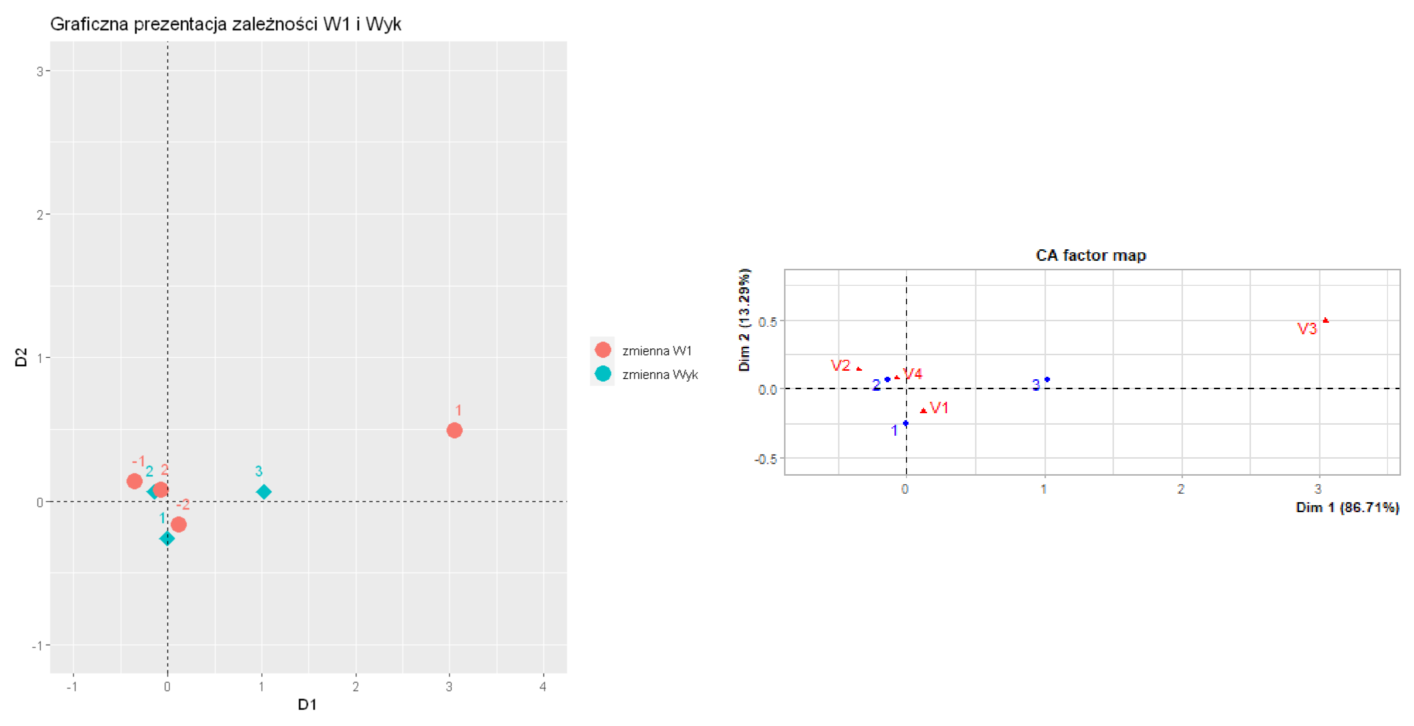
$$F = \begin{bmatrix} -0.003060932 & -0.25943499 \\ -0.138246087 & 0.06714975 \\ 1.025260549 & 0.06504575 \end{bmatrix} \qquad G = \begin{bmatrix} 0.11826617 & -0.1622275 \\ -0.35056182 & 0.1378599 \\ 3.04672664 & 0.4937333 \\ -0.07532609 & 0.0794247 \end{bmatrix}$$

Ponadto wykonano wykresy ilustrujące związek pomiędzy W1 oraz Wyk.

```
val <- zad10(data)
F <- data.frame(val[1])
G <- data.frame(val[2])

##funckje wlasne
ggplot() +
geom_point(aes(F[,1], F[,2], color="zmienna Wyk"), shape=18, size=5) +
geom_point(aes(G[,1], G[,2], color="zmienna W1"), shape=16, size=5) +
theme(legend.title=element_blank()) + xlim(-1,4) + ylim(-1,3) +
labs(x="D1", y="D2") + geom_hline(yintercept=0, linetype="dashed") +
geom_vline(xintercept = 0, linetype="dashed") +
geom_text(data = data.frame(F), aes(x, y, color="zmienna Wyk"),
    label=c(1,2,3), nudge_x=-0.05, nudge_y=0.15 , check_overlap = T) +
geom_text(data = data.frame(G), aes(x, y, color="zmienna W1"),
    label=c(-2,-1,1,2), nudge_x=0.05, nudge_y=0.15 , check_overlap = T) +
ggtitle("Graficzna prezentacja zaleznosci W1 i Wyk") +


##funkcje R do rysowania wykresu
dane <- as.data.frame.matrix(ftable(data, col.vars="W1", row.vars="Wyk"))
CA(dane, graph = TRUE)
```



(a) Własna funkcja.

(b) Wbudowana funkcja R.

Rys. 1: Graficzny związek pomiędzy W1 oraz Wyk

Otrzymany wykres przy pomocy własnej funkcji jest zbliżony do tego wykonanego funkcją wbudowaną. Może to wizualnie potwierdzić poprawność jej działania i otrzymanych wcześniej wyników. Najbardziej od innych danych odbiega wartość odpowiadająca zwykłemu zadowoleniu z zarobków wartość 1 w skali Stapela). Następną odstającą wartością jest grupa z wyższym wykształceniem. Może to być spowodowane tym, że ta właśnie grupa najczęściej odpowiadała na pytanie o zadowolenie z wynagrodzenia w sposób pozytywny (odpowiedź *zgadzam się*). Pozostałe punkty są raczej skupione na jednym obszarze. Oznacza to podobieństwo grup niezadowolonych, a także zdecydowanie zadowolonych ze swojego wynagrodzenia. Zbliżone są do siebie także grupy z niskim i średnim wykształceniem. Ich odpowiedzi dotyczące zarobków były zatem dość podobne.

Obliczona została również inercja, czyli miara rozproszenia profili.

```
##Kody zwracajace inercje rozwazanych zmiennych
inercja_calk <- sum(URV$d^2) ##inercja calkowita
d1 <- URV$d[1]^2/inercja_calk ## % inercji w pierwszym wymiarze
d2 <- URV$d[2]^2/inercja_calk ## % inercji w drugim wymiarze
```

Inercja całkowita	0.1306
Procent zmienności w pierwszym wymiarze	0.8671
Procent zmienności w drugim wymiarze	0.1329

Tab. 21: Wartości inercji zmiennych W1 i Wyk.

Inercja całkowita wynosi 0.1306, co oznacza raczej małe rozproszenie profili. Można to było zauważyć na wykresie 1, gdyż większość z punktów była skupiona wokół przeciętnych profili i punktu (0,0). Ponadto można porównać otrzymane procenty zmienności w obu wymiarach z wynikami na rysunku 1 dla funkcji wbudowanej języka R. Dla pierwszego wymiaru wynik wynosi w obu przypadkach 86.71%. Jest to wysoka wartość, dzięki której możemy zauważyć, że pierwszy wymiar (wymiar D1) aż w prawie 90% tłumaczy nam rozważane zmienne i ich rozproszenie. Mimo wszystko drugi wymiar także wnosi część informacji (ponad 13%), więc w zależności od celu analizy, można podjąć decyzję o redukcji wymiarów do tylko i wyłącznie tego pierwszego.