Analiza Danych Ankietowych
Sprawozdanie 3

Natalia Lach 262303, Alicja Myśliwiec 262275

Matematyka Stosowana
Wydział Matematyki Politechniki Wrocławskiej

2023

Spis treści

1. Zadanie 1

W niniejszym zadaniu należało zweryfikować hipotezę, że atmosfera w miejscu pracy w pierwszym badanym okresie (zmienna A1) i po roku od pierwszego badania (zmienna A2) odpowiada modelowi symetrii. Hipotezą alternatywną jest w takim przypadku to, że rozważane zmienne nie odpowiadają modelowi symetrii. W celu weryfikacji hipotez wykonano dwa testy, działając na danych pochodzących z tabeli 1.

A2 \ A1	-2	-1	0	1	2
-2	10	0	1	0	1
-1	2	15	1	0	1
0	1	1	32	1	0
1	1	1	6	96	1
2	0	0	0	3	26

Tab. 1: Tablica dwudzielcza zmiennych A1 i A2.

Analizowana tablica dwudzielcza, jest tabelą o wymiarach 5x5,a więc wykonać można uogólniony test Mcnemary oraz test ilorazu wiarogodności. Ustalony poziom istotności wynosi $\alpha = 0.05$. Wyniki przedstawiono w poniższej tabeli.

Test	Mcnemary	Ilorazu wiarogodności
p-wartość	NA	0.205975

Tab. 2: p-wartości testów dla zmiennych A1 i A2.

Jak wynika z tabeli 2, p-wartość uogólnionego testu Mcnemary jest niedostępna. W analizowanej tabeli dwudzielczej (tabela 1) występują zera na symetrycznie odpowiadających sobie miejscach. Konstrukcja rozważanego testu nie pozwala na analizę tak wyglądającego zestawu danych.

Inaczej jest w przypadku testu ilorazu wiarogodności. Test zwrócił p-wartość większą od ustalonego poziomu istotności, a więc nie mamy podstaw do odrzucenia hipotezy zerowej, czyli stwierdzić, że dane nie są realizacją modelu symetrii.

Biorąc pod uwagę wymiary analizowanej tabeli dwudzielczej, hipoteza o jednorodności rozkładów brzegowych nie jest równoważna z tą dotyczącą modelu symetrii. Istnieje jedynie jednostronne wynikanie. Skoro nie ma wystarczających dowodów, aby odrzucić hipotezę o symetrii w rozkładzie, to nie ma także podstaw, aby odrzucić hipotezę o jednorodności rozkładów brzegowych.

2. Zadanie 2

Tym razem do czynienia mamy z tablicą dwudzielczą o wymiarach 4x4 dotyczącą zadowolenia z wynagrodzenia w pierwszym badanym okresie (zmienna W1) i po roku od pierwszego badania (zmienna W2). Ponownie należy zweryfikować hipotezę, że obie zmienne odpowiadają modelowi symetrii na poziomie istotności $\alpha = 0.05$. I tym razem wykonano dwa testy, bazując na danych z tabeli dwudzielczej - tabeli 3.

W2 \ W1	-2	-1	1	2
-2	74	0	0	0
-1	0	19	0	0
1	0	1	1	0
2	0	0	1	104

Tab. 3: Tablica dwudzielcza zmiennych W1 i W2.

Biorąc pod uwagę wymiary analizowanej tabeli dwudzielczej, przeprowadzono dwa testy - uogólniony test Mcnemary oraz test ilorazu wiarogodności. Wyniki przedstawiono poniżej, w tabeli 4.

Test	Mcnemary	Ilorazu wiarogodności
p-wartość	NA	0.836800

Tab. 4: p-wartości testów dla zmiennych W1 i W2.

Ponownie jak w przypadku zadania w sekcji 2, test Mcnemary zwrócił nieokreśloną p-wartość. Jak wspomniano wcześniej, powodem jest występowanie zer na odpowiadających sobie miejscach w analizowanej tabeli kontyngencji.

P-wartość w przypadku testu ilorazu wiarogodności jest większa od ustalonego poziomu istotności, a więc nie ma podstaw do odrzucenia hipotezy zerowej dotyczącej modelu symetrii. Owa hipoteza nie jest jednak równoważna hipotezie o jednorodności rozkładów brzegowych. Skoro nie ma wystarczających dowodów, aby odrzucić hipotezę o symetrii w rozkładzie, to nie ma także podstaw, aby odrzucić hipotezę o jednorodności rozkładów brzegowych.

3. Zadanie 3

W niniejszym zadaniu należy zweryfikować hipotezę analogiczną do tych w poprzednich sekcjach. Tym razem jednak działano na zmodyfikowanych danych dotyczących zadowolenia z wynagrodzenia. Zmienne WM1 i WM2 odpowiadają bardziej ogólnemu opisowi opinii wydanych przez pracowników podczas odpowiednio pierwszego okresu badania i po roku od niego. Przykładowo odpowiedź (-1) oznacza ogólne niezadowolenie, czyli połączenie odpowiedzi (-2) i (-1) z oryginalnej zmiennej. Dane przedstawiono w tabeli 5. Poziom istotności pozostaje stały, równy $\alpha = 0.05$.

WM2 \ WM1	-1	1
-1	93	0
1	1	106

Tab. 5: Tablica dwudzielcza zmiennych WM1 i WM2.

Na podstawie tak skonstruowanych danych przeprowadzono test Mcnemary w dwóch wersjach - z poprawką dotyczącą ciągłości (*correct = TRUE*) i bez niej (*correct = FALSE*). Otrzymano następujące wyniki.

Test Mcnemary	*correct=TRUE*	*correct=FALSE*
p-wartość	1	0.3173

Tab. 6: p-wartości testów dla zmiennych WM1 i WM2.

Przy ustalonym poziomie istotności, p-wartości przedstawione w tabeli 6 nie dają podstaw do odrzucenia hipotezy zerowej. To znaczy, że nie mamy wystarczających dowodów, by stwierdzić, że dane nie są realizacjami z modelu symetrii. Zauważyć można znaczące różnice w otrzymanych wartościach z poprawką na ciągłość i bez niej. Jednak decyzja dotycząca hipotezy jest taka sama w obu testach.

Analizowana tablica dwudzielcza jest rozmiaru 2x2, zatem w tym przypadku hipoteza o modelu symetrii jest równoważna z tą o jednorodności rozkładów brzegowych. W takim razie nie mamy podstaw do odrzucenia żadnej z nich.

4. Zadanie 4

W tym zadaniu należało przeprowadzić symulacyjnie porównanie mocy testu Z oraz testu Z_0.

Zaczynając od implementacji owych testów, test Z:

```
test_z <- function(ftab){
    n <- sum(ftab)
    p1p <- rowSums(ftab)[1]/n
    pp1 <- colSums(ftab)[1]/n
    D <- p1p - pp1
    s2 <- (p1p*(1-p1p) + pp1*(1-pp1) -
            2*(ftab[1,1]*ftab[2,2] - ftab[1,2]*ftab[2,1])/n^2)/n
    z <- D/sqrt(s2)
    pval <- 2*(1-pnorm(abs(z)))
    }
```

Test Z_0:

```
test_z0  <- function(ftab){
    n <- sum(ftab)
    z0 <- (ftab[1,2]-ftab[2,1])/sqrt(ftab[1,2]+ftab[2,1])
    print(z0)
    pval <- 2*(1-pnorm(abs(z0)))
    }
```

Hipoteza zerowa jest w postaci

$$H_0 : p_{1+} = p_{+1},$$

natomiast alternatywna

$$H_1 : p_{1+} \neq p_{+1}.$$

W celu symulacji należy wygenerować odpowiedzi na pytania dla n respondentów.

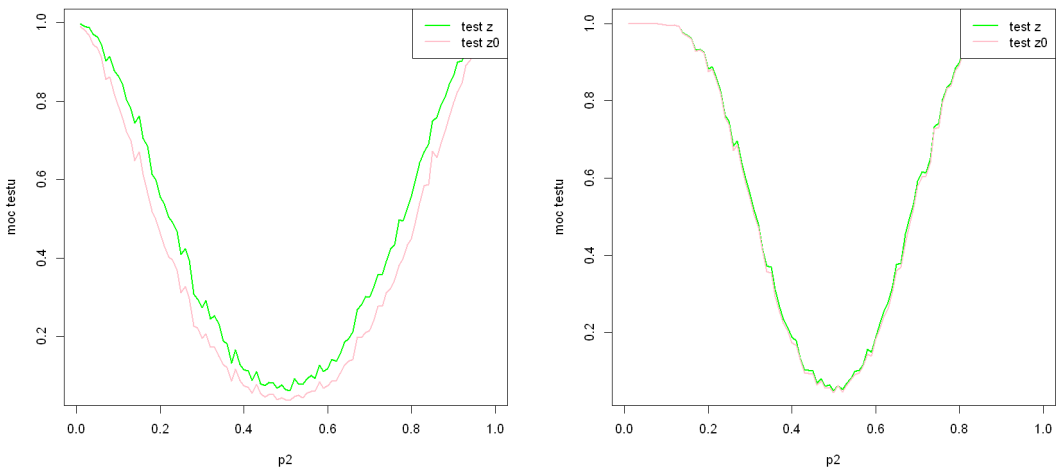
```
    testy <- function(p2,n, pval = 0.05){
        p1 <- 0.5
        odp_1 <- runif(n) ##generuje 1000 losowych wartosci miedzy 0 a 1
        odp_2 <- runif(n)
        o1 <- as.integer(odp_1<p1) ##patrze czy sa < prawdopodobienstwa sukcesu
        o2 <- as.integer(odp_2<p2)

        if (0 %in% o2 & 1 %in% o2) {
             tab <- table(o1, o2)
              }
        else if (0 %in% o2) {
            tab <- table(o1, o2)
            tab <- cbind(tab, 0)
```

```
        }
    else if (1 %in% o2) {
        tab <- table(o1, o2)
        tab <- cbind(0, tab)
      }
    else {
        tab <- matrix(0, nrow = 2, ncol = 2)
      }
    r1 <- test_z0(tab)<pval
    r2 <- test_z(tab)<pval
    data.frame(r1, r2)
}
```

W ten sposób otrzymujemy rezultaty testów Z oraz Z0 wygenerowanych przez nas danych. Poniżej kod służący do liczenia mocy testów oraz przykładową symulację.

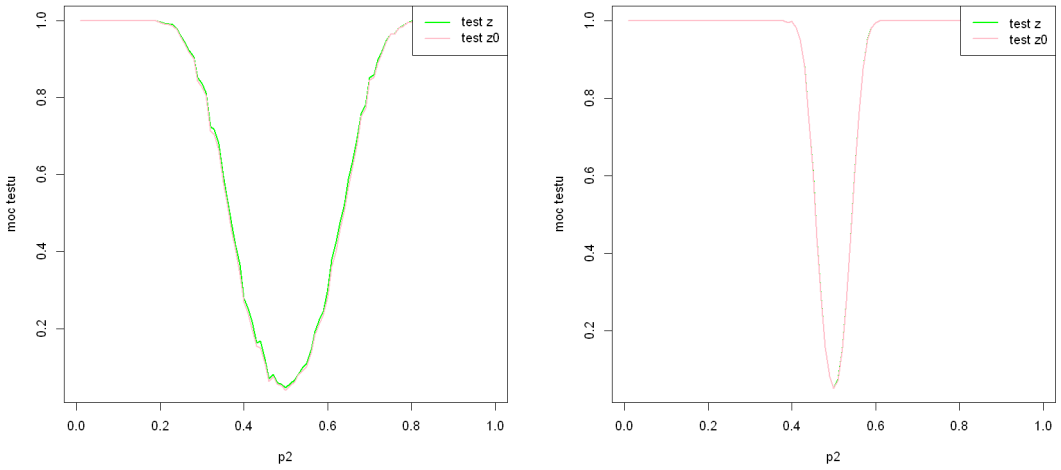
```
## funkcja liczaca moc
powers <- function(p2,n) {
    results <- do.call(rbind, replicate(1000, testy(p2, n),
                               simplify = FALSE))
    data.frame(mean(results$r2), mean(results$r1))
}

##przykladowe wykonanie dla n=20
a <- lapply(seq(0.01,0.99,0.01), function(p) powers(p, 20))

z <- c()
z_0 <- c()
for (i in 1:length(a)){
    z <- append(z, a[i][[1]][1,1])
    z_0 <- append(z_0, a[i][[1]][1,2])
}

data.frame(z,z_0)
```

Ostatecznie moc została wyznaczona dla $n \in [20, 50, 100, 1000]$, $p_1 = 0.5$ oraz $p_2 = 0.01, 0.02, \ldots 0.99$.



(a) $n = 20$ (b) $n = 50$

Rys. 1: Moce testów Z i Z_0 w zależności od wartości p_2



(a) $n = 100$ (b) $n = 1000$

Rys. 2: Moce testów Z i Z_0 w zależności od wartości p_2

Na rysunkach 1 i 2 można zauważyć, że oba testy są silniejsze w miarę zwiększania różnicy wartości p_2 od $p_2 = 0.5$. W miarę symulacji dla większego n widać, że ostatecznie oba wykresy się niemalże pokrywają, jednak przykładowo dla $n = 20$, test Z_0 wypada gorzej (moc testu ma mniejszą wartość).

5. Zadanie 5

W niniejszym zadaniu zakładamy, że

— zmienna 1 to zmienna S, czyli zajmowane stanowisko (kierownicze bądź nie),
— zmienna 2 to zmienna W1, czyli zadowolenie z wynagrodzenia w pierwszym badanym okresie (wartości w skali Stapela od -2 do 2),
— zmienna 3 to zmienna Wyk, czyli wykształcenie (wartości od 1 do 3).

Na podstawie wspomnianych zmiennych tworzone będą następujące modele log-liniowe. Zakładamy poziom ufności równy $\alpha = 0.05$. Testowana będzie hipoteza zerowa, że dany model został dopasowany poprawnie do posiadanych danych.

Wpierw jednak rozważane dane trzeba przygotować, to znaczy utworzyć odpowiednie tabele liczności. Przykładowy kod przedstawiono poniżej.

```
##zebranie odpowiednich zmiennych
col_s <- data$S
col_wyk <- data$Wyk
col_w <- data$W1

##stworzenie obiektu data_frame
df <- data.frame(col_s, col_w, col_wyk)
names(df) <- c('S','W1','Wyk')
df_tab <- as.data.frame(table(df))
```

5.1. Podpunkt a)

Model [1 3] zakłada niezależność zmiennych 1 i 3. Można go zapisać poniższym wzorem.

$$\ell_{ik} = \lambda + \lambda_i^{(1)} + \lambda_k^{(3)}$$

$$\forall i \in \{1,2\}, k \in \{1,2,3\}$$

Oznacza to, że zakładamy, że zajmowane stanowisko nie jest powiązane z wykształceniem. Dodatkowo zadowolenie z wynagrodzenia nie ma wpływu na model.

Odpowiednio przygotowane dane podano jako argument funkcji *glm* i dopasowano do nich rozważany model za pomocą poniższego kodu.

```
    mods <- glm(Freq ~ S + Wyk, ##[1 3]
                data = df_tab, family = poisson)

    summary(mods) ##wywolanie summary modelu

    deviance(mods) ##miara bledu

    1-pchisq(deviance(mods), df = df.residual(mods)) ##p-wartosc

    cbind(mods$data, fitted(mods)) ##tabela licznosci
}
```

Otrzymane *summary* oraz tabele liczności przedstawiono poniżej, na rysunku 3. Ponadto p-wartość wyniosła 0 a *deviance* ≈ 203.07.

```
Call:
glm(formula = Freq ~ S + Wyk, family = poisson, data = df_tab)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-7.7814  -2.4320  -0.7379   1.7898   5.8821

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.1823     0.1587  13.755  < 2e-16 ***
S1           -1.8575     0.2069  -8.977  < 2e-16 ***
Wyk2          1.2281     0.1776   6.916 4.65e-12 ***
Wyk3         -0.7691     0.2775  -2.772  0.00558 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 442.20  on 23  degrees of freedom
Residual deviance: 203.07  on 20  degrees of freedom
AIC: 267.82

Number of Fisher Scoring iterations: 5
```

(a) Podsumowanie funkcji glm.

	S	W1	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	-2	1	19	8.86625
2	1	-2	1	1	1.38375
3	0	-1	1	3	8.86625
4	1	-1	1	0	1.38375
5	0	1	1	0	8.86625
6	1	1	1	0	1.38375
7	0	2	1	18	8.86625
8	1	2	1	0	1.38375
9	0	-2	2	40	30.27500
10	1	-2	2	5	4.72500
11	0	-1	2	15	30.27500
12	1	-1	2	2	4.72500
13	0	1	2	0	30.27500
14	1	1	2	0	4.72500
15	0	2	2	68	30.27500
16	1	2	2	10	4.72500
17	0	-2	3	5	4.10875
18	1	-2	3	4	0.64125
19	0	-1	3	0	4.10875
20	1	-1	3	0	0.64125
21	0	1	3	0	4.10875
22	1	1	3	2	0.64125
23	0	2	3	5	4.10875
24	1	2	3	3	0.64125

(b) Tabela liczności.

Rys. 3: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

Jak widać na rysunku 3, liczności z modelu dość znacząco różnią się od tych z danych. Także otrzymana p-wartość daje podstawy do odrzucenia hipotezy zerowej, czyli do stwierdzenia, że faktycznie model został źle dopasowany.
Wniosek jest zatem taki, że musi istnieć pewna zależność pomiędzy zajmowanym stanowiskiem a

wykształceniem pracownika, a także zmienna W1 może mieć wpływ na odpowiedzi, czego rozważany przez nas model nie uwzględnia.

5.2. Podpunkt b)

Model [13] zakłada zależność zmiennych 1 i 3. Można go zapisać poniższym wzorem.

$$\ell_{ik} = \lambda + \lambda_i^{(1)} + \lambda_k^{(3)} + \lambda_{ik}^{(13)}$$

$$\forall i \in \{1,2\}, k \in \{1,2,3\}$$

Oznacza to, że zakładamy, że zajmowane stanowisko jest powiązane z wykształceniem.

Odpowiednio przygotowane dane podano jako argument funkcji *glm* i dopasowano do nich rozważany model za pomocą poniższego kodu.

```
    mods <- glm(Freq ~ S + Wyk + S * Wyk, ##[1 3]
                data = df_tab, family = poisson)

    summary(mods) ##wywolanie summary modelu

    deviance(mods) ##miara bledu

    1-pchisq(deviance(mods), df = df.residual(mods)) ##p-wartosc

    cbind(mods$data, fitted(mods)) ##tabela licznosci
}
```

Otrzymane *summary* oraz tabele liczności przedstawiono poniżej, na rysunku 4. Ponadto *deviance* wyniosła ≈ 183.9797 a p-wartość jest równa 0.

```
Call:
glm(formula = Freq ~ S + Wyk + S * Wyk, family = poisson, data = df_tab)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-7.8422  -2.2361  -0.4385   1.3898   5.7820

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.3026     0.1581  14.563  < 2e-16 ***
S1           -3.6889     1.0124  -3.644 0.000269 ***
Wyk2          1.1233     0.1820   6.171 6.77e-10 ***
Wyk3         -1.3863     0.3535  -3.921 8.81e-05 ***
S1:Wyk2       1.7099     1.0449   1.636 0.101749
S1:Wyk3       3.5835     1.1117   3.223 0.001267 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 442.20  on 23  degrees of freedom
Residual deviance: 183.98  on 18  degrees of freedom
AIC: 252.73

Number of Fisher Scoring iterations: 5
```

(a) Podsumowanie funkcji glm.

	S	W1	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	-2	1	19	10.00
2	1	-2	1	1	0.25
3	0	-1	1	3	10.00
4	1	-1	1	0	0.25
5	0	1	1	0	10.00
6	1	1	1	0	0.25
7	0	2	1	18	10.00
8	1	2	1	0	0.25
9	0	-2	2	40	30.75
10	1	-2	2	5	4.25
11	0	-1	2	15	30.75
12	1	-1	2	2	4.25
13	0	1	2	0	30.75
14	1	1	2	0	4.25
15	0	2	2	68	30.75
16	1	2	2	10	4.25
17	0	-2	3	5	2.50
18	1	-2	3	4	2.25
19	0	-1	3	0	2.50
20	1	-1	3	0	2.25
21	0	1	3	0	2.50
22	1	1	3	2	2.25
23	0	2	3	5	2.50
24	1	2	3	3	2.25

(b) Tabela liczności.

Rys. 4: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

Jak wynika z tabeli z rysunku 4, oczekiwane liczności uzyskane z modelu różnią się od tych rzeczywistych. Także wartość *deviance* jest wysoka. Stwierdzamy, że model został źle dopasowany.

5.3. Podpunkt c)

Model [1 2 3] jest oparty na trzech zmiennych i zakłada ich niezależność względem siebie. Można go zapisać poniższym wzorem.

$$\ell_{ik} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)}$$

$$\forall i \in \{1,2\}, j \in \{1,2,3,4\}, k \in \{1,2,3\}$$

Oznacza to, że zakładamy, że ani zajmowane stanowisko ani zadowolenie z wynagrodzenia ani wykształcenie nie są w żaden sposób ze sobą powiązane.

Odpowiednio przygotowane dane podano jako argument funkcji *glm* i dopasowano do nich rozważany model za pomocą poniższego kodu.

```r
mods <- glm(Freq ~ S + W1 + Wyk,
            data = dfswyk_tab, family = poisson)

summary(mods) ##wywolanie summary modelu

deviance(mods) ##miara bledu
```

```r
1-pchisq(deviance(mods), df = df.residual(mods)) ##p-wartosc

cbind(mods$data, fitted(mods)) ##tabela licznosci
```

Otrzymane *summary* oraz tabele liczności przedstawiono poniżej, na rysunku 5. Ponadto p-wartość wyniosła około 0.00062 a *deviance* ≈ 42.2422.

```
Call:
glm(formula = Freq ~ S + W1 + Wyk, family = poisson, data = df_tab)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3992  -0.8209  -0.5129   0.2158   3.6711

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.5743     0.1835  14.026  < 2e-16 ***
S1           -1.8575     0.2069  -8.977  < 2e-16 ***
W1-1         -1.3083     0.2520  -5.191 2.09e-07 ***
W11          -3.6109     0.7166  -5.039 4.68e-07 ***
W12           0.3403     0.1521   2.238  0.02524 *
Wyk2          1.2281     0.1776   6.916 4.65e-12 ***
Wyk3         -0.7691     0.2775  -2.771  0.00558 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 442.195  on 23  degrees of freedom
Residual deviance:  42.242  on 17  degrees of freedom
AIC: 112.99

Number of Fisher Scoring iterations: 8
```

(a) Podsumowanie funkcji glm.

	S	W1	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	-2	1	19	13.12205
2	1	-2	1	1	2.04795
3	0	-1	1	3	3.54650
4	1	-1	1	0	0.55350
5	0	1	1	0	0.35465
6	1	1	1	0	0.05535
7	0	2	1	18	18.44180
8	1	2	1	0	2.87820
9	0	-2	2	40	44.80700
10	1	-2	2	5	6.99300
11	0	-1	2	15	12.11000
12	1	-1	2	2	1.89000
13	0	1	2	0	1.21100
14	1	1	2	0	0.18900
15	0	2	2	68	62.97200
16	1	2	2	10	9.82800
17	0	-2	3	5	6.08095
18	1	-2	3	4	0.94905
19	0	-1	3	0	1.64350
20	1	-1	3	0	0.25650
21	0	1	3	0	0.16435
22	1	1	3	2	0.02565
23	0	2	3	5	8.54620
24	1	2	3	3	1.33380

(b) Tabela liczności.

Rys. 5: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

Jak wynika z tabeli na rysunku 5, liczności oczekiwane i rzeczywiste nieco się od siebie różnią. Także wartość *deviance* jest wysoka, a p-wartość na ustalonym poziomie ufności daje podstawy do odrzucenia hipotezy zerowej. Model nie został dopasowany poprawnie.

5.4. Podpunkt d)

Model [12 3] zakłada za to, że wśród trzech zmiennych, zmienne 1 i 2 są od siebie zależne, jednak występuje brak zależności wobec zmiennej 3. Model można zapisać poniższym wzorem.

$$\ell_{ik} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \lambda_{ij}^{(12)}$$

$$\forall i \in \{1,2\}, j \in \{1,2,3,4\}, k \in \{1,2,3\}$$

Oznacza to, że zakładamy, że zajmowane stanowisko oraz zadowolenie z wynagrodzenia są od siebie zależne, jednak wykształcenie jest czynnikiem od nich niezależnym.

Odpowiednio przygotowane dane podano jako argument funkcji *glm* i dopasowano do nich rozważany model za pomocą poniższego kodu.

```r
mods <- glm(Freq ~ S + W1 + Wyk + S*W1,
            data = dfswyk_tab, family = poisson)

summary(mods) ##wywolanie summary modelu

deviance(mods) ##miara bledu

1-pchisq(deviance(mods), df = df.residual(mods)) ##p-wartosc

cbind(mods$data, fitted(mods)) ##tabela licznosci
```

Otrzymane *summary* oraz tabele liczności przedstawiono poniżej, na rysunku 6. Ponadto p-wartość wyniosła około 0.00212 a *deviance* ≈ 33.9138.

```
Call:
glm(formula = Freq ~ S + W1 + Wyk + S * W1, family = poisson,
    data = df_tab)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3087  -0.8377  -0.2620   0.4903   2.4074

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.57414    0.18712  13.756  < 2e-16 ***
S1           -1.85630    0.34004  -5.459 4.79e-08 ***
W1-1         -1.26851    0.26680  -4.755 1.99e-06 ***
W11         -21.09274 2883.98341  -0.007  0.99416
W12           0.35198    0.16314   2.158  0.03096 *
Wyk2          1.22807    0.17758   6.916 4.65e-12 ***
Wyk3         -0.76913    0.27753  -2.771  0.00558 **
S1:W1-1      -0.34093    0.81926  -0.416  0.67731
S1:W11       19.48330 2883.98351   0.007  0.99461
S1:W12       -0.08961    0.45115  -0.199  0.84255
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 442.195  on 23  degrees of freedom
Residual deviance:  33.914  on 14  degrees of freedom
AIC: 110.67

Number of Fisher Scoring iterations: 16
```

(a) Podsumowanie funkcji glm.

	S	W1	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	-2	1	19	1.312000e+01
2	1	-2	1	1	2.050000e+00
3	0	-1	1	3	3.690000e+00
4	1	-1	1	0	4.100000e-01
5	0	1	1	0	9.067214e-09
6	1	1	1	0	4.100000e-01
7	0	2	1	18	1.865500e+01
8	1	2	1	0	2.665000e+00
9	0	-2	2	40	4.480000e+01
10	1	-2	2	5	7.000000e+00
11	0	-1	2	15	1.260000e+01
12	1	-1	2	2	1.400000e+00
13	0	1	2	0	3.096122e-08
14	1	1	2	0	1.400000e+00
15	0	2	2	68	6.370000e+01
16	1	2	2	10	9.100000e+00
17	0	-2	3	5	6.080000e+00
18	1	-2	3	4	9.500000e-01
19	0	-1	3	0	1.710000e+00
20	1	-1	3	0	1.900000e-01
21	0	1	3	0	4.201880e-09
22	1	1	3	2	1.900000e-01
23	0	2	3	5	8.645000e+00
24	1	2	3	3	1.235000e+00

(b) Tabela liczności.

Rys. 6: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

Ponownie, tabela na rysunku 6 ukazuje niezbyt dobre dopasowanie modelu do danych. Liczności teoretyczne i rzeczywiste różnią się od siebie. *Deviance* w tym przypadku wynosi nieco mniej niż w poprzednim podpunkcie, jednak także i tym razem odrzucamy hipotezę zerową. Otrzymana p-wartość wskazuje na złe dopasowanie modelu do rozważanych danych.

5.5. Podpunkt e).

Model [12 13] zakłada, że wśród trzech zmiennych, zmienne 1 i 2, a także 1 i 3 są od siebie zależne. Przy ustalonej wartości zmiennej 1, zmienne 2 i 3 są od siebie niezależne. Wówczas mowa

o zmiennych 2 i 3 jako zmiennych warunkowo niezależnych. Taką relację można zapisać poniższym wzorem.

$$\ell_{ik} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \lambda_{ij}^{(12)} + \lambda_{ik}^{(13)}$$
$$\forall i \in \{1,2\}, j \in \{1,2,3,4\}, k \in \{1,2,3\}$$

Oznacza to, że między zajmowanym stanowiskiem i zadowoleniem z wynagrodzenia, a także między zajmowanym stanowiskiem a poziomem wykształcenia istnieje pewna zależność. Natomiast zadowolenie z wynagrodzenia i wykształcenie są zmiennymi warunkowo niezależnymi.

Odpowiednio przygotowane dane podano jako argument funkcji *glm* i dopasowano do nich rozważany model za pomocą poniższego kodu.

```
mods <- glm(Freq ~ S + W1 + Wyk + S*W1 + S*Wyk,
            data = dfswyk_tab, family = poisson)

summary(mods) ##wywolanie summary modelu

deviance(mods) ##miara bledu

1-pchisq(deviance(mods), df = df.residual(mods)) ##p-wartosc

cbind(mods$data, fitted(mods)) ##tabela licznosci
```

Otrzymane *summary* oraz tabele liczności przedstawiono poniżej, na rysunku 7. Ponadto p-wartość wyniosła około 0.2512 a *deviance* ≈ 14.8237.

```
Call:
glm(formula = Freq ~ S + W1 + Wyk + S * W1 + S * Wyk, family = poisson,
    data = df_tab)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.58698  -0.67881  -0.05727   0.60121   1.31445

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.69447    0.18667  14.435  < 2e-16 ***
S1           -3.68772    1.04776  -3.520 0.000432 ***
W1-1         -1.26851    0.26680  -4.755 1.99e-06 ***
W11         -21.93973 4404.68250  -0.005 0.996026
W12           0.35198    0.16314   2.158 0.030964 *
Wyk2          1.12330    0.18202   6.171 6.77e-10 ***
Wyk3         -1.38629    0.35355  -3.921 8.82e-05 ***
S1:W1-1      -0.34093    0.81926  -0.416 0.677306
S1:W11       20.33029 4404.68256   0.005 0.996317
S1:W12       -0.08961    0.45115  -0.199 0.842552
S1:Wyk2       1.70991    1.04497   1.636 0.101771
S1:Wyk3       3.58352    1.11181   3.223 0.001268 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 442.195  on 23  degrees of freedom
Residual deviance:  14.824  on 12  degrees of freedom
AIC: 95.576

Number of Fisher Scoring iterations: 17
```

(a) Podsumowanie funkcji glm.

	S	W1	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	-2	1	19	1.479769e+01
2	1	-2	1	1	3.703704e-01
3	0	-1	1	3	4.161850e+00
4	1	-1	1	0	7.407407e-02
5	0	1	1	0	4.384198e-09
6	1	1	1	0	7.407407e-02
7	0	2	1	18	2.104046e+01
8	1	2	1	0	4.814815e-01
9	0	-2	2	40	4.550289e+01
10	1	-2	2	5	6.296296e+00
11	0	-1	2	15	1.279769e+01
12	1	-1	2	2	1.259259e+00
13	0	1	2	0	1.348141e-08
14	1	1	2	0	1.259259e+00
15	0	2	2	68	6.469942e+01
16	1	2	2	10	8.185185e+00
17	0	-2	3	5	3.699422e+00
18	1	-2	3	4	3.333333e+00
19	0	-1	3	0	1.040462e+00
20	1	-1	3	0	6.666667e-01
21	0	1	3	0	1.096050e-09
22	1	1	3	2	6.666667e-01
23	0	2	3	5	5.260116e+00
24	1	2	3	3	4.333333e+00

(b) Tabela liczności.

Rys. 7: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

Tym razem otrzymana p-wartość jest większa od ustalonego poziomu ufności, a więc nie mamy podstaw do odrzucenia hipotezy zerowej o tym, że model został dobrze dobrany. Także wartość *deviance* jest najniższa jak dotąd, biorąc pod uwagę podpunkty dotyczące wszystkich trzech zmiennych. Liczności z tabeli 7 są do siebie zbliżone. Zatem nie mamy podstaw do stwierdzenia, że model mógł zostać źle dopasowany do analizowanych danych.

5.6. Podpunkt f)

Model [1 23] zakłada, że wśród trzech zmiennych, zmienne 2 i 3 są od siebie zależne, jednak występuje brak zależności wobec zmiennej 1. Można go zapisać poniższym wzorem.

$$\ell_{ik} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \lambda_{jk}^{(23)}$$

$$\forall i \in \{1,2\}, j \in \{1,2,3,4\}, k \in \{1,2,3\}$$

Oznacza to, że zakładamy, że zadowolenie z wynagrodzenia oraz wykształcenie są od siebie zależne, jednak zajmowane stanowisko jest czynnikiem od nich niezależnym.

Odpowiednio przygotowane dane podano jako argument funkcji *glm* i dopasowano do nich rozważany model za pomocą poniższego kodu.

```
mods <- glm(Freq ~ S + W1 + Wyk + W1*Wyk,
          data = dfswyk_tab, family = poisson)

summary(mods) ##wywolanie summary modelu

deviance(mods) ##miara bledu

1-pchisq(deviance(mods), df = df.residual(mods)) ##p-wartosc

cbind(mods$data, fitted(mods)) ##tabela licznosci
```

Otrzymane *summary* oraz tabele liczności przedstawiono poniżej, na rysunku 8. Ponadto p-wartość wyniosła około 0.01286 a *deviance* ≈ 23.970.

Ustalono poziom ufności na $\alpha = 0.05$, zatem otrzymana p-wartość sugeruje odrzucenie hipotezy zerowej. I w tym przypadku liczności przedstawione w tabeli na rysunku 8 są do siebie dość zbliżone, jednak nie na tyle, by stwierdzić, że model został dobrze dopasowany do danych. Wskazuje na to także wartość *deviance*.

```
Call:
glm(formula = Freq ~ S + W1 + Wyk + W1 * Wyk, family = poisson,
    data = df_tab)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2045  -0.5296  -0.0001   0.1900   2.1330

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.85071    0.22534  12.650  < 2e-16 ***
S1           -1.85745    0.20692  -8.977  < 2e-16 ***
W1-1         -1.89712    0.61914  -3.064  0.00218 **
W11         -21.28324 5674.58093  -0.004  0.99701
W12          -0.10536    0.32489  -0.324  0.74572
Wyk2          0.81093    0.26874   3.018  0.00255 **
Wyk3         -0.79851    0.40139  -1.989  0.04666 *
W1-1:Wyk2     0.92367    0.68145   1.355  0.17528
W11:Wyk2     -0.81093 8025.06932   0.000  0.99992
W12:Wyk2      0.65541    0.37496   1.748  0.08048 .
W1-1:Wyk3   -18.58761 5674.58099  -0.003  0.99739
W11:Wyk3     19.77916 5674.58098   0.003  0.99722
W12:Wyk3     -0.01242    0.58452  -0.021  0.98304
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 442.20  on 23  degrees of freedom
Residual deviance:  23.97  on 11  degrees of freedom
AIC: 106.72

Number of Fisher Scoring iterations: 17
```

(a) Podsumowanie funkcji glm.

	S	W1	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	-2	1	19	1.730000e+01
2	1	-2	1	1	2.700000e+00
3	0	-1	1	3	2.595000e+00
4	1	-1	1	0	4.050000e-01
5	0	1	1	0	9.882209e-09
6	1	1	1	0	1.542310e-09
7	0	2	1	18	1.557000e+01
8	1	2	1	0	2.430000e+00
9	0	-2	2	40	3.892500e+01
10	1	-2	2	5	6.075000e+00
11	0	-1	2	15	1.470500e+01
12	1	-1	2	2	2.295000e+00
13	0	1	2	0	9.882209e-09
14	1	1	2	0	1.542310e-09
15	0	2	2	68	6.747000e+01
16	1	2	2	10	1.053000e+01
17	0	-2	3	5	7.785000e+00
18	1	-2	3	4	1.215000e+00
19	0	-1	3	0	9.882209e-09
20	1	-1	3	0	1.542310e-09
21	0	1	3	0	1.730000e+00
22	1	1	3	2	2.700000e-01
23	0	2	3	5	6.920000e+00
24	1	2	3	3	1.080000e+00

(b) Tabela liczności.

Rys. 8: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

6. Zadanie 6

Podobnie jak w zadaniu 5 w sekcji 5, sprawdzane będzie dopasowanie modelów log-linowych do danych. W tym przypadku analizowane będą zmienne:

— zmienna 1 - zmienna S, czyli zajmowane stanowisko (kierownicze bądź nie),
— zmienna 2 - zmienna P, czyli płeć (litery M lub K),
— zmienna 3 - zmienna Wyk, czyli wykształcenie (wartości od 1 do 3).

Sposób przygotowywania danych, wykonywania funkcji *glm* oraz inne operacje i założenia pozostają takie same jak w zadaniu 5. Modele log-liniowe, które zostaną dopasowane do danych także się nie zmieniły, więc nie bedą ponownie opisywane.

6.1. Podpunkt a)

Analizowany jest model [1 3]. Oznacza to, że zakładamy, że ani zajmowane stanowisko ani wykształcenie nie są w żaden sposób ze sobą powiązane. Płeć nie gra żadnej roli w analizowanym modelu.

Otrzymane *summary* oraz tabele liczności przedstawiono poniżej, na rysunku 9. Ponadto p-wartość wyniosła około $1.6342 \cdot 10^{-13}$ a *deviance* ≈ 77.392.

```
Call:
glm(formula = Freq ~ S + Wyk, family = poisson, data = df_tab)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.2644  -2.0923  -0.4669   1.2975   4.3522

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.8754     0.1586  18.125  < 2e-16 ***
S1           -1.8575     0.2069  -8.977  < 2e-16 ***
Wyk2          1.2281     0.1776   6.916 4.64e-12 ***
Wyk3         -0.7691     0.2775  -2.771  0.00558 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 316.518  on 11  degrees of freedom
Residual deviance:  77.392  on  8  degrees of freedom
AIC: 127.78

Number of Fisher Scoring iterations: 5
```

(a) Podsumowanie funkcji glm.

	S	P	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	K	1	1	17.7325
2	1	K	1	0	2.7675
3	0	M	1	39	17.7325
4	1	M	1	1	2.7675
5	0	K	2	54	60.5500
6	1	K	2	4	9.4500
7	0	M	2	69	60.5500
8	1	M	2	13	9.4500
9	0	K	3	8	8.2175
10	1	K	3	4	1.2825
11	0	M	3	2	8.2175
12	1	M	3	5	1.2825

(b) Tabela liczności.

Rys. 9: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

Jak wynika z tabeli z rysunku 9, liczności teoretyczne i rzeczywiste różnią się od siebie dość znacząco. Wskazuje na to także wartość *deviance*. Otrzymana p-wartość sugeruje odrzucenie hipotezy zerowej i przyjęcie, że model nie jest dobrze dopasowany.

6.2. Podpunkt b)

Analizowany jest model [13]. Oznacza to, że zakładamy, że zajmowane stanowisko oraz wykształcenie są ze sobą powiązane. Płeć nie gra żadnej roli w rozważanym modelu.

Otrzymane *summary* oraz tabele liczności przedstawiono poniżej, na rysunku 10. Ponadto p-wartość wyniosła około $9.95 \cdot 10^{-11}$ a *deviance* ≈ 58.3023.

```
Call:
glm(formula = Freq ~ S + Wyk + S * Wyk, family = poisson, data = df_tab)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.6576  -1.1320  -0.0044   1.0116   3.7538

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.9957     0.1581  18.947  < 2e-16 ***
S1           -3.6889     1.0124  -3.644 0.000269 ***
Wyk2          1.1233     0.1820   6.172 6.76e-10 ***
Wyk3         -1.3863     0.3536  -3.921 8.82e-05 ***
S1:Wyk2       1.7099     1.0450   1.636 0.101770
S1:Wyk3       3.5835     1.1118   3.223 0.001268 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 316.518  on 11  degrees of freedom
Residual deviance:  58.302  on  6  degrees of freedom
AIC: 112.69

Number of Fisher Scoring iterations: 5
```

(a) Podsumowanie funkcji glm.

	S	P	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	K	1	1	20.0
2	1	K	1	0	0.5
3	0	M	1	39	20.0
4	1	M	1	1	0.5
5	0	K	2	54	61.5
6	1	K	2	4	8.5
7	0	M	2	69	61.5
8	1	M	2	13	8.5
9	0	K	3	8	5.0
10	1	K	3	4	4.5
11	0	M	3	2	5.0
12	1	M	3	5	4.5

(b) Tabela liczności.

Rys. 10: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

Jak wynika z tabeli z rysunku 10, liczności teoretyczne i rzeczywiste różnią się od siebie dość znacząco. Wskazuje na to także wartość *deviance*. Otrzymana *p*-wartość sugeruje odrzucenie hipotezy zerowej i przyjęcie, że model nie jest dobrze dopasowany.

6.3. Podpunkt c)

Analizowany jest model [1 2 3]. Oznacza to, że zakładamy, że ani zajmowane stanowisko ani płeć ani wykształcenie nie są w żaden sposób ze sobą powiązane.

Otrzymane *summary* oraz tabele liczności przedstawiono poniżej, na rysunku 11. Ponadto *p*-wartość wyniosła około $1.298 \cdot 10^{-10}$ a *deviance* ≈ 60.328.

```
Call:
glm(formula = Freq ~ S + P + Wyk, family = poisson, data = df_tab)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.2561  -1.7038  -0.4112   1.7332   3.0601

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.5329     0.1851  13.685  < 2e-16 ***
S1           -1.8575     0.2069  -8.977  < 2e-16 ***
PM            0.5971     0.1478   4.041 5.32e-05 ***
Wyk2          1.2281     0.1776   6.916 4.65e-12 ***
Wyk3         -0.7691     0.2775  -2.771  0.00558 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 316.518  on 11  degrees of freedom
Residual deviance:  60.328  on  7  degrees of freedom
AIC: 112.72

Number of Fisher Scoring iterations: 5
```

(a) Podsumowanie funkcji glm.

	S	P	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	K	1	1	12.590075
2	1	K	1	0	1.964925
3	0	M	1	39	22.874925
4	1	M	1	1	3.570075
5	0	K	2	54	42.990500
6	1	K	2	4	6.709500
7	0	M	2	69	78.109500
8	1	M	2	13	12.190500
9	0	K	3	8	5.834425
10	1	K	3	4	0.910575
11	0	M	3	2	10.600575
12	1	M	3	5	1.654425

(b) Tabela liczności.

Rys. 11: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

Jak wynika z tabeli z rysunku 11, liczności teoretyczne i rzeczywiste różnią się od siebie dość znacząco. Wskazuje na to także wartość *deviance*. Otrzymana *p*-wartość sugeruje odrzucenie hipotezy zerowej i przyjęcie, że model nie jest dobrze dopasowany.

6.4. Podpunkt d)

Analizowany jest model [12 3]. Oznacza to, zakładamy, że zajmowane stanowisko oraz płeć są od siebie zależne, jednak wykształcenie jest czynnikiem od nich niezależnym.

Otrzymane *summary* oraz tabele liczności przedstawiono poniżej, na rysunku 12. Ponadto *p*-wartość wyniosła około $4.8343 \cdot 10^{-11}$ a *deviance* ≈ 59.8472.

```
Call:
glm(formula = Freq ~ S + P + Wyk + S * P, family = poisson, data = df_tab)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.3259  -1.7670  -0.3977   1.5670   3.1354

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.5584     0.1878  13.624  < 2e-16 ***
S1           -2.0637     0.3753  -5.498 3.83e-08 ***
PM            0.5573     0.1580   3.528 0.000419 ***
Wyk2          1.2281     0.1776   6.916 4.65e-12 ***
Wyk3         -0.7691     0.2775  -2.771 0.005582 **
S1:PM         0.3077     0.4501   0.684 0.494277
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 316.518  on 11  degrees of freedom
Residual deviance:  59.847  on  6  degrees of freedom
AIC: 114.24

Number of Fisher Scoring iterations: 5
```

(a) Podsumowanie funkcji glm.

	S	P	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	K	1	1	12.915
2	1	K	1	0	1.640
3	0	M	1	39	22.550
4	1	M	1	1	3.895
5	0	K	2	54	44.100
6	1	K	2	4	5.600
7	0	M	2	69	77.000
8	1	M	2	13	13.300
9	0	K	3	8	5.985
10	1	K	3	4	0.760
11	0	M	3	2	10.450
12	1	M	3	5	1.805

(b) Tabela liczności.

Rys. 12: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

Jak wynika z tabeli z rysunku 12, liczności teoretyczne i rzeczywiste różnią się od siebie. Wskazuje na to także dość wysoka wartość *deviance*. Otrzymana p-wartość ponownie sugeruje odrzucenie hipotezy zerowej i przyjęcie, że model nie jest dobrze dopasowany.

6.5. Podpunkt e)

Analizowany jest model [12 13]. Oznacza to, że między zajmowanym stanowiskiem i płcią, a także między zajmowanym stanowiskiem a poziomem wykształcenia istnieje pewna zależność. Natomiast płeć i wykształcenie są zmiennymi warunkowo niezależnymi.

```
Call:
glm(formula = Freq ~ S + P + Wyk + S * P + S * Wyk, family = poisson,
    data = df_tab)

Deviance Residuals:
      1       2       3       4       5       6       7       8
-4.6664  -0.7698   2.4923   0.3320   1.3323  -0.4795  -1.0627   0.2956
      9      10      11      12
 1.9686   0.7596  -2.0224  -0.5503

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.6787     0.1873  14.300  < 2e-16 ***
S1           -3.8951     1.0597  -3.676 0.000237 ***
PM            0.5573     0.1580   3.528 0.000419 ***
Wyk2          1.1233     0.1820   6.171 6.77e-10 ***
Wyk3         -1.3863     0.3536  -3.921 8.82e-05 ***
S1:PM         0.3077     0.4501   0.684 0.494285
S1:Wyk2       1.7099     1.0450   1.636 0.101771
S1:Wyk3       3.5835     1.1118   3.223 0.001268 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 316.518  on 11  degrees of freedom
Residual deviance:  40.757  on  4  degrees of freedom
AIC: 99.147

Number of Fisher Scoring iterations: 5
```

(a) Podsumowanie funkcji glm.

	S	P	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	K	1	1	14.5664740
2	1	K	1	0	0.2962963
3	0	M	1	39	25.4335260
4	1	M	1	1	0.7037037
5	0	K	2	54	44.7919075
6	1	K	2	4	5.0370370
7	0	M	2	69	78.2080925
8	1	M	2	13	11.9629630
9	0	K	3	8	3.6416185
10	1	K	3	4	2.6666667
11	0	M	3	2	6.3583815
12	1	M	3	5	6.3333333

(b) Tabela liczności.

Rys. 13: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

Otrzymane *summary* oraz tabele liczności przedstawiono na rysunku 13. Ponadto p-wartość wyniosła około $3.0177 \cdot 10^{-8}$ a *deviance* ≈ 40.7571.

Także tym razem p-wartość jest mniejsza od ustalonego poziomu ufności, a więc odrzucamy hipotezę zerową i stwierdzamy, że model został źle dopasowany.

6.6. Podpunkt f)

Analizowany jest model [1 23]. Oznacza to, że zakładamy, że zadowolenie z wynagrodzenia oraz wykształcenie są od siebie zależne, jednak zajmowane stanowisko jest czynnikiem od nich niezależnym

Otrzymane *summary* oraz tabele liczności przedstawiono poniżej, na rysunku 14. Ponadto p-wartość wyniosła około 0.000174 a *deviance* ≈ 24.4908.

```
Call:
glm(formula = Freq ~ S + P + Wyk + P * Wyk, family = poisson,
    data = df_tab)

Deviance Residuals:
      1       2       3       4       5       6       7       8
 0.1416  -0.5196   0.7329  -2.3296   0.5341  -1.5122  -0.2302   0.5643
      9      10      11      12
-0.7701   1.5719  -1.9181   2.9240

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.1450     1.0004  -0.145 0.884735
S1           -1.8575     0.2069  -8.977  < 2e-16 ***
PM            3.6889     1.0124   3.644 0.000269 ***
Wyk2          4.0604     1.0086   4.026 5.68e-05 ***
Wyk3          2.4849     1.0408   2.387 0.016967 *
PM:Wyk2      -3.3426     1.0269  -3.255 0.001133 **
PM:Wyk3      -4.2279     1.1186  -3.780 0.000157 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 316.518  on 11  degrees of freedom
Residual deviance:  24.491  on  5  degrees of freedom
AIC: 80.881

Number of Fisher Scoring iterations: 6
```

(a) Podsumowanie funkcji glm.

	S	P	Wyk	Freq	fitted(mods)
	<fct>	<fct>	<fct>	<int>	<dbl>
1	0	K	1	1	0.865
2	1	K	1	0	0.135
3	0	M	1	39	34.600
4	1	M	1	1	5.400
5	0	K	2	54	50.170
6	1	K	2	4	7.830
7	0	M	2	69	70.930
8	1	M	2	13	11.070
9	0	K	3	8	10.380
10	1	K	3	4	1.620
11	0	M	3	2	6.055
12	1	M	3	5	0.945

(b) Tabela liczności.

Rys. 14: Podsumowanie wykonanej funkcji wraz z tabelą liczności modelu.

Jak wynika z tabeli z rysunku 14, liczności teoretyczne i rzeczywiste różnią się od siebie. Wskazuje na to także wartość *deviance*. Otrzymana p-wartość ponownie sugeruje odrzucenie hipotezy zerowej i przyjęcie, że model nie jest dobrze dopasowany.

7. Zadanie 7

W niniejszym zadaniu skupiamy się na zmiennych z zadania 5 z sekcji 5. Rozważane będą dwa modele log-liniowe - [13 23] oraz [123]. W celu dopasowania modelów skorzystano z następującego kodu.

```
col_s <- data$S
col_wyk <- data$Wyk
col_w <- data$W1
```

```
df <- data.frame(col_s, col_w, col_wyk)
names(df) <- c('S','W1','Wyk')
df_tab <- as.data.frame(table(df))

mod_1 <- glm(Freq ~ S + W1 + Wyk + S*Wyk + W1*Wyk, ##[13 23]
          data = df_tab, family = poisson)

mod2 <- glm(freq7b ~ (S + W1 + Wyk)^2 + S*W1*Wyk, #[123]
          data = df_tab, family = poisson)

fit1 <- cbind(mod_1$data, fitted(mod_1))
fit2 <- cbind(mod_2$data, fitted(mod_2))
```

Później na podstawie $fit1$ oraz $fit2$ określane są prawdopodobieństwa.

Za dane przyjmujemy zmienne:

— 1 - zmienna S - stanowisko,
— 3 - zmienna Wyk - wykształcenie,
— 2 - zmienna W1 - zadowolenie z wynagrodzenia w pierwszym badanym okresie.

7.1. Podpunkt a)

W tym podpunkcie należy oszacować prawdopodobieństwo, że osoba pracująca na stanowisku kierowniczym jest zdecydowanie zadowolona ze swojego wynagrodzenia.

Model / Wartości	[13 23]	[123]
Teoretyczne	0.48148	0.48148
Estymowane	0.50740	0.48148

Tab. 7: Teoretyczne i estymowane wartości prawdopodobieństw.

Analizując tabelę 7 można zauważyć, że wartości teoretyczne i estymowane prawdopodobieństw dla modelu [123] są niemal identyczne (wartości podane w tabeli są przybliżeniami otrzymanych wyników, jednak różnice występują na dalekim miejscu po przecinku, więc są one w tym przypadku nierozróżnialne). Jednak dla modelu [13 23] owe prawdopodobieństwa różnią się. Może to świadczyć o niepoprawnym dopasowaniu modelu do danych. Otrzymane wyniki sugerują, że prawdopodobieństwo, że osoba pracująca na stanowisku kierowniczym jest zdecydowanie zadowolona ze swojego wynagrodzenia wynosi około 0.48.

7.2. Podpunkt b)

W tym podpunkcie należy oszacować prawdopodobieństwo, że osoba z wykształceniem zawodowym pracuje na stanowisku kierowniczym.

Model / Wartości	[13 23]	[123]
Teoretyczne	0.02439	0.02439
Estymowane	0.02439	0.02439

Tab. 8: Teoretyczne i estymowane wartości prawdopodobieństw.

Analizując tabelę 8, zauważyć można, że i tym razem wartości teoretyczne i estymowane w obu modelach wyniosły w przybliżeniu dokładnie tyle samo. Zatem prawdopodobieństwo, że osoba z wykształceniem zawodowym pracuje na stanowisku kierowniczym jest równe około 0.02.

7.3. Podpunkt c)

W tym podpunkcie należy oszacować prawdopodobieństwo, że osoba z wykształceniem wyższym nie pracuje na stanowisku kierowniczym.

Wartości \ Model	[13 23]	[123]
Teoretyczne	0.526316	0.526316
Estymowane	0.526316	0.526316

Tab. 9: Teoretyczne i estymowane wartości prawdopodobieństw.

Analizując tabelę 9, ponownie zauważyć można, że wartości teoretyczne i estymowane w obu modelach wyniosły w przybliżeniu dokładnie tyle samo. Zatem prawdopodobieństwo, że soba z wykształceniem wyższym nie pracuje na stanowisku kierowniczym jest równe około 0.52.

8. Zadanie 8

Procedura wykonania zadania jest identyczna jak w zadaniu 7 (sekcja 7). Jedynie zostały zmienione analizowane dane. Przyjmujemy zmienne:

— 1 - zmienna S - stanowisko,
— 2 - zmienna P - płeć,
— 3 - zmienna Wyk - wykształcenie.

8.1. Podpunkt a)

W tabeli 10 podano teoretyczne oraz estymowane prawdopodobieństwa, że osoba pracująca na stanowisku kierowniczym jest kobietą.

Wartości \ Model	[13 23]	[123]
Teoretyczne	0.296	0.296
Estymowane	0.472	0.296

Tab. 10: Teoretyczne i estymowane wartości prawdopodobieństw.

Tym razem estymowana wartość prawdopodobieństwa jest niezgodna z jej rzeczywistą wartością w przypadku modelu [13 23]. Może to świadczyć o złym dopasowaniu modelu do danych. Jednak w przypadku modelu [123], obie wartości są w przybliżeniu takie same. Zatem można stwierdzić, że szukane prawdopodobieństwo wynosi około 0.3.

8.2. Podpunkt b)

W tabeli 11 podano teoretyczne oraz estymowane prawdopodobieństwa, że osoba z wykształceniem zawodowym pracuje na stanowisku kierowniczym.

Wartości \ Model	[13 23]	[123]
Teoretyczne	0.02439	0.02439
Estymowane	0.02439	0.02439

Tab. 11: Teoretyczne i estymowane wartości prawdopodobieństw.

Otrzymane wartości wynoszą w przybliżeniu tyle samo w każdym przypadku, a więc można stwierdzić, że szukane prawdopodobieństwo wynosi około 0.02.
Jak się okazuje, jest to wartość identyczna do tej uzyskanej w zadaniu 7b) z sekcji 7.2.

8.3. Podpunkt c)

W tabeli 12 podano teoretyczne oraz estymowane prawdopodobieństwa, że osoba z wykształceniem wyższym jest mężczyzną.

Wartości \ Model	[13 23]	[123]
Teoretyczne	0.368	0.3684
Estymowane	0.3684	0.3684

Tab. 12: Teoretyczne i estymowane wartości prawdopodobieństw.

Otrzymane wartości wynoszą w przybliżeniu tyle samo w każdym przypadku, a więc można stwierdzić, że ów prawdopodobieństwo jest równe 0.37.

9. Zadanie 9

W tym zadaniu zostaną zweryfikowane następujące hipotezy na poziomie istotności $\alpha = 0.05$

9.1. Podpunkt a)

- zmienne losowe S, W1 i Wyk są wzajemnie niezależne.
H_0 o danych z modelu [1 2 3] będzie testowana przeciwko hipotezie alternatywnej: dane pochodza z modelu [123]

```
## Wszelkie tablice beda tworzone w nastepujacy sposob
col_s <- data$S
col_wyk <- data$Wyk
col_w1 <- data$W1

##tabela S, W1, Wyk
df <- data.frame(col_s, col_w1, col_wyk)
names(df) <- c('S','W1','Wyk')
df_tab <- as.data.frame(table(df))
```

```
mods <- glm(Freq ~ W1+S+Wyk, data=df_tab, family=poisson)
p_val <- 1-pchisq(deviance(mods), df = df.residual(mods))
```

Otrzymano $p\_val = 0.0006187$, która jest mniejszą wartością od zadanego poziomu istotności . Odrzucamy więc hipotezę zerową na rzecz alternatywnej.

Następnie rozpatrywana hipoteza zerowa (model [1 2 3]) będzie testowana przeciwko hipotezie alternatywnej o tym, że dane pochodzą z nadmodelu [12 32].

```
mods2 <- glm(Freq ~ S+W1+Wyk + S*Wyk + Wyk*W1, data=df_tab, family=poisson)
test <- anova(mods, mods2)
p_val <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Korzystając z funkcji *anova* otrzymano $p\_val = 9.870781e-06$, co ponownie jest znacznie mniejsze od zadanego poziomu istotności. Odrzucamy więc w obu przypadkach hipotezę zerową i wnioskujemy, iż podane zmienne nie są niezależne.

9.2. Podpunkt b)

- zmienna losowa W1 jest niezależna od pary zmiennych S i Wyk (model [2 13]).
H_0 o danych z modelu [2 13] będzie testowana przeciwko hipotezie alternatywnej: dane pochodza z modelu [123]

```
mods <- glm(Freq ~ S+W1+Wyk + W1 + S*Wyk, data=df_tab, family=poisson)
p_val <- 1-pchisq(deviance(mods), df = df.residual(mods))
```

Otrzymano $p\_val = 0.080963$, co jest większe od zadanego $\alpha = 0.05$. Dlatego też nie ma podstaw do odrzucenia hipotezy zerowej.

Następnie rozpatrywana hipoteza zerowa będzie testowana przeciwko hipotezie alternatywnej o tym, że dane pochodzą z nadmodelu [12 32] .

```
mods2 <- glm(Freq ~ S+W1+Wyk + S*W1 + Wyk*W1, data=df_tab, family=poisson)
test <- anova(mods, mods2)
p_val <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymano $p\_val = 0.377725$. W tym przypadku wartość ta jest mniejsza od zadanego α, co pozwala nam odrzucić hipotezę zerową o tym, że zmienna losowa W1 jest niezależna od pary zmiennych S i Wyk.

9.3. Podpunkt c/d)

- zmienna losowa W1 jest niezależna od zmiennej losowej S, przy ustalonej wartości zmiennej Wyk (model [13 23]). H_0 o danych z modelu [13 23] będzie testowana przeciwko hipotezie alternatywnej: dane pochodza z modelu [123].

```
mods <- glm(Freq ~ S+W1+Wyk + S*Wyk + W1*Wyk, data=df_tab, family=poisson)
p_val <- 1-pchisq(deviance(mods), df = df.residual(mods))
```

Otrzymano $p\_val = 0.84464$. Wartość ta jest większa od zadanego $\alpha = 0.05$, dlatego też nie ma podstaw do odrzucenia hipotezy zerowej.

Następnie rozpatrywana hipoteza zerowa będzie testowana przeciwko hipotezie alternatywnej o tym, że dane pochodzą z nadmodelu [12 13 23].

```
mods2 <- glm(Freq ~ S+W1+Wyk + S*Wyk + S*W1 + W1*Wyk, data=df_tab,
    family=poisson)
test <- anova(mods, mods2)
p_val <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymano $p\_val = 0.34998$. W tym przypadku wartość ta jest mniejsza od zadanego α, co pozwala nam odrzucić hipotezę zerową o tym, że zmienna losowa W1 jest niezależna od pary zmiennych S i Wyk.

9.4. Podpunkt e)

- zmienna losowa S jest niezależna od zmiennej P, przy ustalonej wartości zmiennej Wyk (model [13 32]).
H_0 o danych z modelu [13 32] będzie testowana przeciwko hipotezie alternatywnej: dane pochodza z modelu [123]

```
col_s <- data$S
col_wyk <- data$Wyk
col_p <- data$P

df <- data.frame(col_s, col_p, col_wyk)
names(df) <- c('S','P','Wyk')
df_tab <- as.data.frame(table(df))
```

```
mods <- glm(Freq ~ S+P+Wyk + S*Wyk + P*Wyk, data=df_tab, family=poisson)
p_val <- 1-pchisq(deviance(mods), df = df.residual(mods))
```

Otrzymano $p\_val = 0.1446957$. Wartość ta jest mniejsza od zadanego α, dlatego też odrzucamy hipotezę zerową o niezależności zmiennej losowa S od zmiennej P, przy ustalonej wartości zmiennej Wyk, na rzecz alternatywnej.

Następnie rozpatrywana hipoteza zerowa będzie testowana przeciwko hipotezie alternatywnej o tym, że dane pochodzą z nadmodelu [12 13 23].

```
mods2 <- glm(Freq ~ S+P+Wyk+ S*Wyk + S*P + P*Wyk, data=df_tab,
    family=poisson)
test <- anova(mods, mods2)
p_val <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymano $p\_val = 0.024487$. W tym przypadku wartość ta jest ponownie mniejsza od zadanego α, co pozwala nam odrzucić hipotezę zerową.

10. Zadanie 10

W niniejszym zadaniu analizowane będą modele log-liniowe dla następujących zmiennych.

— Zmienna 1 to zmienna A1, czyli atmosfera w miejscu pracy w pierwszym badanym okresie (wartości w skali Likerta)
— Zmienna 2 to zmienna W1, czyli zadowolenie z wynagrodzenia w pierwszym badanym okresie (wartości w skali Stapela)
— Zmienna 3 to zmienna P, czyli płeć

Ustalony poziom ufności wynosi $\alpha = 0.05$.

10.1. Podejście względem kryterium informacyjnego.

Aby porównać wartości kryteriów informacyjnych AIC i BIC dla rozważanych zmiennych, wykorzystano następujący kod.

```
## stworzenie tabeli z danymi
data10 <- data[c('A1','W1','P')]
table10 <- ftable(data10, row.vars = c('A1','W1'))
table10 <- as.data.frame(as.table(table10))
```

```
freq10 <- table10$Freq
##stworzenie przykladowego modelu (jednego z 19 mozliwych)
model10_1 <- glm(freq10 ~ (A1 + W1 + P), ##[1 2 3]
             data = table10,
             family = poisson)
##policzenie kryteriow
aic <- AIC(model10_1)
bic <- BIC(model10_1)
```

Powtórzono tę procedurę dla każdego z modeli i wyniki przedstawiono w tabeli na poniższym rysunku 15.

MODEL	AIC	BIC
<chr>	<dbl>	<dbl>
[1 2 3]	314.2426	329.4425
[12 3]	123.5923	159.0588
[1 23]	318.0580	338.3245
[13 2]	320.6063	342.5617
[12 13]	129.9560	172.1780
[12 23]	127.4077	167.9408
[23 13]	324.4217	324.4217
[12 23 13]	133.5509	180.8396
[123]	150.1856	217.7408
[1]	484.1343	492.5787
[2]	427.4253	434.1808
[3]	567.1889	570.5666
[13]	475.4339	492.3227
[12]	138.6564	172.4340
[23]	416.1766	429.6877
[1 2]	329.3067	342.8177
[1 3]	469.0702	479.2035
[2 3]	412.3613	420.8056
[ ]	582.2529	583.9418

Rys. 15: Wartości kryteriów informacyjnych.

Z przedstawionej tabeli wynika, że najmniejsze wartości obu kryteriów osiągnięte zostały dla modelu [12 3]. Oznacza to, że ów model najlepiej oddaje zależności pomiędzy analizowanymi danymi. Atmosfera w pracy oraz zadowolenie z wynagrodzenia są od siebie zależne, podczas gdy płeć nie jest czynnikiem ściśle z nimi powiązanym. Otrzymane wyniki wydają się być zgodne z rzeczywistością.

10.2. Podejście krokowe.

W podejściu krokowym postępujemy według następującej procedury.

```
##przygotowanie danych
data10 <- data[c('A1','W1','P')]
table10 <- ftable(data10, row.vars = c('A1','W1'))
table10 <- as.data.frame(as.table(table10))
freq10 <- table10$Freq

##inicjalizacja tabel na przyszlosc
table10_1 <- data.frame('M' = c(),'p-value' = c(), 'AIC(M)' = c(),
                        'BIC(M)' = c())
table10_2 <- data.frame('M' = c(),'p-value' = c(), 'AIC(M)' = c(),
                        'BIC(M)' = c())
```

```
table10_3 <- data.frame('M' = c(),'p-value' = c(), 'AIC(M)' = c(),
                        'BIC(M)' = c())

##utworzenie modeli
model10_1 <- glm(freq10 ~ (A1 + W1 + P), ##[1 2 3]
                 data = table10,
                 family = poisson)
model10_2 <- glm(freq10 ~ (A1 + W1 + P + A1*W1), ##[12 3]
                 data = table10,
                 family = poisson)
model10_3 <- glm(freq10 ~ (A1 + W1 + P + W1*P), ##[1 23]
                 data = table10,
                 family = poisson)
model10_4 <- glm(freq10 ~ (A1 + W1 + P + A1*P), ##[13 2]
                 data = table10,
                 family = poisson)
model10_5 <- glm(freq10 ~ (A1 + W1 + P + A1*W1 + A1*P), ##[12 13]
                 data = table10,
                 family = poisson)
model10_6 <- glm(freq10 ~ (A1 + W1 + P + A1*W1 + W1*P), ##[12 23]
                 data = table10,
                 family = poisson)
model10_7 <- glm(freq10 ~ (A1 + W1 + P + A1*P + W1*P), ##[13 23]
                 data = table10,
                 family = poisson)
model10_8 <- glm(freq10 ~ (A1 + W1 + P + A1*W1 + W1*P + A1*P), ##[12 23 13]
                 data = table10,
                 family = poisson)
model10_9 <- glm(freq10 ~ ((A1 + W1 + P)^2+ A1*W1*P), ##[123]
                 data = table10,
                 family = poisson)

##funkcja do testowania czy model1 jest lepszy od modelu2
test <- function(model1, model2){

    if (model1$aic != model2$aic){
      test <- anova(model1,model2)
      test_g <- 1 - pchisq(test$Deviance[2],df = test$Df[2])
      test_BIC <- BIC(model2)
      test_AIC <- AIC(model2)
      c(test_g,test_AIC,test_BIC)

    } else {
      test_BIC <- BIC(model2)
      test_AIC <- AIC(model2)
      c(NaN,test_AIC,test_BIC)
    }
}
##hierarchie modeli
models1 = list(model10_1, model10_2, model10_3, model10_4)
```

Pierwszym krokiem jest przeprowadzenia testów oraz wyliczenia kryteriów dla modelu podstawowego [1 2 3] względem modeli [12 3], [1 23], [2 13].

```
##pierwszy poziom modeli
```

```
for (model in models1){
  tests <- test(model10_1,model)
  print(tests)
  table10_1 <- rbind(table10_1,tests)
  colnames(table10_1) <- c('p-value', 'AIC(Mr)','BIC(Mr)')
}
Mr <- c('M_1 [1 2 3]','M_2 [12 3]','M_3 [1 23]', 'M_4 [13 2]')
table10_1 <- cbind(Mr, table10_1)
```

Otrzymano w ten sposób tabelę z podsumowaniem uzyskanych wyników.

Mr	p-value	AIC(Mr)	BIC(Mr)
<chr>	<dbl>	<dbl>	<dbl>
M_1 [1 2 3]	NaN	314.2426	329.4425
M_2 [12 3]	0.0000000	123.5923	159.0588
M_3 [1 23]	0.5349822	318.0580	338.3245
M_4 [13 2]	0.8022519	320.6063	342.5617

Rys. 16: Tabela kroku nr. 1.

Z tabeli na rysunku 16 wynika, że hipotezę zerową testu ANOVA odrzucono tylko w przypadku modelu drugiego, to jest modelu [12 3]. Także wartości kryteriów są najniższe dla tego właśnie przypadku. Wskazuje to na fakt, że ten model jest najlepszym spośród tych analizowanych i będzie on porównywany dalej, z kolejnymi nadmodelami.

```
##drugi poziom
models2 = list(model10_2, model10_5, model10_6)
for (model in models2){
  tests <- test(model10_2,model)
  table10_2 <- rbind(table10_2,tests)
  colnames(table10_2) <- c('p-value' , 'AIC(Mr)' ,'BIC(Mr)' )
}
Mr <- c('M_2 [12 3]','M_5 [12 13]', 'M_6 [12 23]')
table10_2 <- cbind(Mr, table10_2)
```

Mr	p-value	AIC(Mr)	BIC(Mr)
<chr>	<dbl>	<dbl>	<dbl>
M_2 [12 3]	NaN	123.5923	159.0588
M_5 [12 13]	0.8022519	129.9560	172.1780
M_6 [12 23]	0.5349822	127.4077	167.9408

Rys. 17: Tabela kroku nr. 2.

Analizując tabelę na rysunku 17, można zauważyć, że każda z otrzymanych p-wartości jest wyższa od ustalonego poziomu ufności, a więc nie mamy podstaw do odrzucenia hipotezy zerowej i do stwierdzenia, że któryś z rozważanych modeli jest lepiej dopasowany do danych niż model [12 3]. Także wartości kryteriów są wyższe niż te dla modelu podstawowego.

Procedura krokowa na tym etapie powinna zostać zakończona i wybrany powinien zostać model [12 3]. Jest to dokładnie ten sam model, który został uznany za najodpowiedniejszy także w poprzednim podejściu, podczas wyboru modelu z najmniejszą wartością kryteriów informacyjnych. Interpretacja pozostaje taka sama.

11. Zadanie 11

W niniejszym zadaniu powtórzono procedurę z zadania 10 z sekcji 10. Zmieniły się jedynie analizowane dane.

— Zmienna 1 to zmienna D, czyli dział, w którym pracownik jest zatrudniony.
— Zmienna 2 to zmienna A1, czyli atmosfera w miejscu pracy w pierwszym badanym okresie (wartości w skali Likerta).
— Zmienna 3 to zmienna P, czyli płeć.

11.1. Podejście względem kryterium informacyjnego.

Otrzymano następujące wyniki.

MODEL	AIC	BIC
<chr>	<dbl>	<dbl>
[1 2 3]	209.3334	224.5333
[12 3]	217.8385	253.3050
[1 23]	215.6971	237.6525
[13 2]	158.9377	179.2043
[12 13]	167.4428	207.9759
[12 23]	224.2022	266.4242
[23 13]	165.3014	165.3014
[12 23 13]	173.7756	221.0643
[123]	184.5309	252.0861
[1]	322.5161	329.2716
[2]	277.1697	285.6141
[3]	360.2243	363.6020
[13]	257.0563	270.5674
[12]	232.9026	266.6802
[23]	268.4693	285.3581
[1 2]	224.3975	237.9085
[1 3]	307.4520	315.8964
[2 3]	262.1057	272.2389
[ ]	375.2884	376.9772

Rys. 18: Wartości kryteriów informacyjnych.

Analizując tabelę przedstawioną na rysunku 18, widzimy, że najmniejsze wartości kryteriów informacyjnych zostały osiągniete dla modelu [13 2]. Otrzymane wyniki sugerują, że płeć oraz dział, w którym dany pracownik pracuje są od siebie zależne, jednak to co uważa o atmosferze w miejscu pracy jest już czynnikiem od nich niezależnym. Otrzymane wyniki wydają się być zgodne z rzeczywistością.

11.2. Podejście krokowe.

Także w tym podejściu korzystać będziemy z procedur wytłumaczonych w poprzednim zadaniu, w sekcji 10.2. Nastąpiła jedynie zmiana danych.

```
data11 <- data[c('D','A1','P')]
table11 <- ftable(data11, row.vars = c('D','A1'))
table11 <- as.data.frame(as.table(table11))
freq10 <- table11$Freq
```

Porównując modele [12 3], [1 23], [2 13] względem modelu podstawowego [1 2 3], otrzymano następujące wyniki.

Mr	p-value	AIC(Mr)	BIC(Mr)
<chr>	<dbl>	<dbl>	<dbl>
M_1 [1 2 3]	NaN	209.3334	224.5333
M_2 [12 3]	2.154821e-01	217.8385	253.3050
M_3 [1 23]	8.022519e-01	215.6971	237.6525
M_4 [13 2]	3.458678e-12	158.9377	179.2043

Rys. 19: Tabela kroku nr. 1.

Jak wynika z tabeli na rysunku 19, jedyny model dla którego p-wartość testu ANOVA wyniosła mniej niż ustalony poziom ufności, to model [13 2]. Daje to podstawy do odrzucenia hipotezy zerowej i stwierdzenia, że jest on bardziej odpowiedni wobec posiadanych danych niż model podstawowy. Jednocześnie, uzyskał on najmniejsze wartości kryteriów informacyjnych AIC i BIC. Zatem to model [13 2] będzie dalej porównywany i testowany.

Mr	p-value	AIC(Mr)	BIC(Mr)
<chr>	<dbl>	<dbl>	<dbl>
M_4 [13 2]	NaN	158.9377	179.2043
M_5 [12 13]	0.2154821	167.4428	207.9759
M_7 [13 23]	0.8022519	165.3014	192.3235

Rys. 20: Tabela kroku nr. 2.

Tabela na rysunku 20 przedstawia drugi krok w stosowanej procedurze. Otrzymane p-wartości są większe od ustalonego poziomu ufności, a więc brak nam podstaw do odrzucenia hipotezy zerowej i stwierdzenia, że któryś z rozważanych modeli jest lepszy od tego podstawowego. Potwierdzają to także wysokie wartości kryteriów informacyjnych.

Kończymy zatem procedurę i dochodzimy do wniosku, że wybrany zostać powinien model [13 2]. To on najlepiej oddaje charakter danych. Także i w tym przypadku, wybrany model powiela się z tym, wybranym metodą uwzględniającą minimum z kryteriów informacyjnych. Interpretacja wyników pozostaje taka sama.