

# Raport nr 3

Natalia Iwańska 262270, Klaudia Janicka 262268

2023-07-02

## Zadanie 1

Tab. 1: Tablica dwudzielcza dla zmiennych A1 i A2.

	-2	-2	0	1	2	Sum
	10	2	1	1	0	14
	0	15	1	1	0	17
	1	1	32	6	0	40
	0	0	1	96	3	100
	1	1	0	1	26	29
Sum	12	19	35	105	29	200

## Test McNemary

Nie możemy skorzystać z testu McNemary, ponieważ w tablicy na odpowiadających sobie miejscach ( $Y_{ij}$  i  $Y_{ji}$ ) występują zera, co "psuje" nam statystykę testową (wynika to wprost z jej definicji).

## Test bazujący na ilorazie wiarygodności

Korzystając z testu bazującego na ilorazie wiarygodności na poziomie istotności  $\alpha = 0.05$  otrzymana p-wartość wyniosła 0.2059752. Zatem nie mamy podstaw do odrzucenia hipotezy o symetrii, która jest równoważna hipotezie o brzegowej jednorodności.

## Zadanie 2

Tab. 2: Tablica dwudzielcza dla zmiennych W1 i W2.

	-2	-1	1	2	Sum
	74	0	0	0	74
	0	19	1	0	20
	0	0	1	1	2
	0	0	0	104	104
Sum	74	19	2	105	200

## Test McNemary

Podobnie jak w poprzednim zadaniu nie możemy skorzystać z testu McNemary, ponieważ w tablicy na odpowiadających sobie miejscach ( $Y_{ij}$  i  $Y_{ji}$ ) występują zera, co “psuje” nam statystykę testową (wynika to wprost z jej definicji).

## Test bazujący na ilorazie wiarygodności

Korzystając z testu bazującego na ilorazie wiarygodności na poziomie istotności  $\alpha = 0.05$  otrzymana p-wartość wyniosła 0.8368001. Zatem nie mamy podstaw do odrzucenia hipotezy zerowej o symetrii, która jest równoważna hipotezie o brzegowej jednorodności.

## Zadanie 3

Tab. 3: Tablica dwudzielcza dla nowych zmiennych W1 i W2.

	-1	1
-1	93	1
1	0	106

```
#TEST Z
test_z <- function(tabela){
  n <- sum(rowSums(tabela))
  P <- tabela/n
  r <- rowSums(P)
  c <- colSums(P)
  D <- r[1] - c[1]
  sigma2_D <- (r[1]*(1-r[1])+c[1]*(1-c[1])-2*(P[1,1]*P[2,2]-P[1,2]*P[2,1]))/n
  Z <- D/sqrt(sigma2_D)
  p <- 2*(1 - pnorm(abs(Z)))
  return(p)
}

#TEST Z0
test_z0 <- function(tabela){
  n <- sum(rowSums(tabela))
  P <- tabela/n
  r <- rowSums(P)
  c <- colSums(P)
  D <- r[1] - c[1]
  sigma2_D0 <- (tabela[1,2]+tabela[2,1])/n^2
  Z_0 <- D/sqrt(sigma2_D0)
  p <- 2*(1 - pnorm(abs(Z_0)))
  return(p)
}
```

test	p-value
Test Z	0.3160976
Test Z0	0.3173105
McNemar test z poprawką na ciągłość	1.0000000
McNemar test bez poprawki	0.3173105

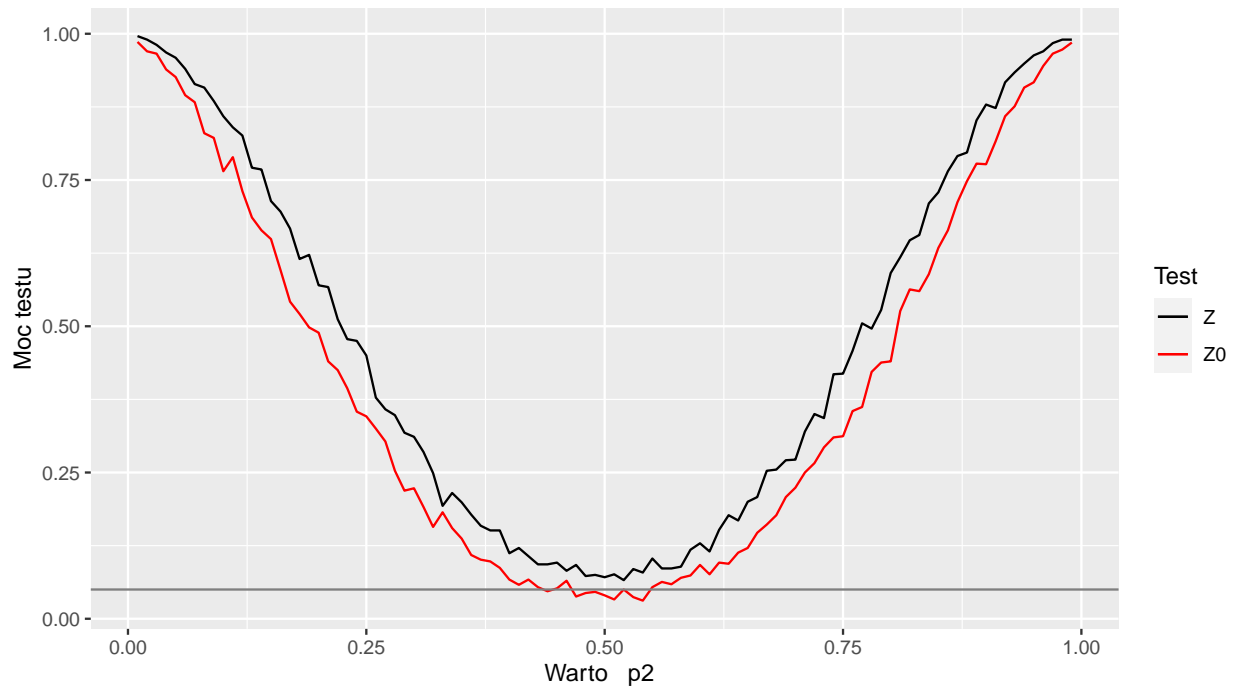
Na podstawie otrzymanych p-wartości testów przeprowadzonych na poziomie istotności  $\alpha = 0.05$  stwierdzamy, że nie mamy podstaw do odrzucenia hipotezy zerowej o symetrii, która jest równoważna hipotezie o brzegowej jednorodności.

## Zadanie 4

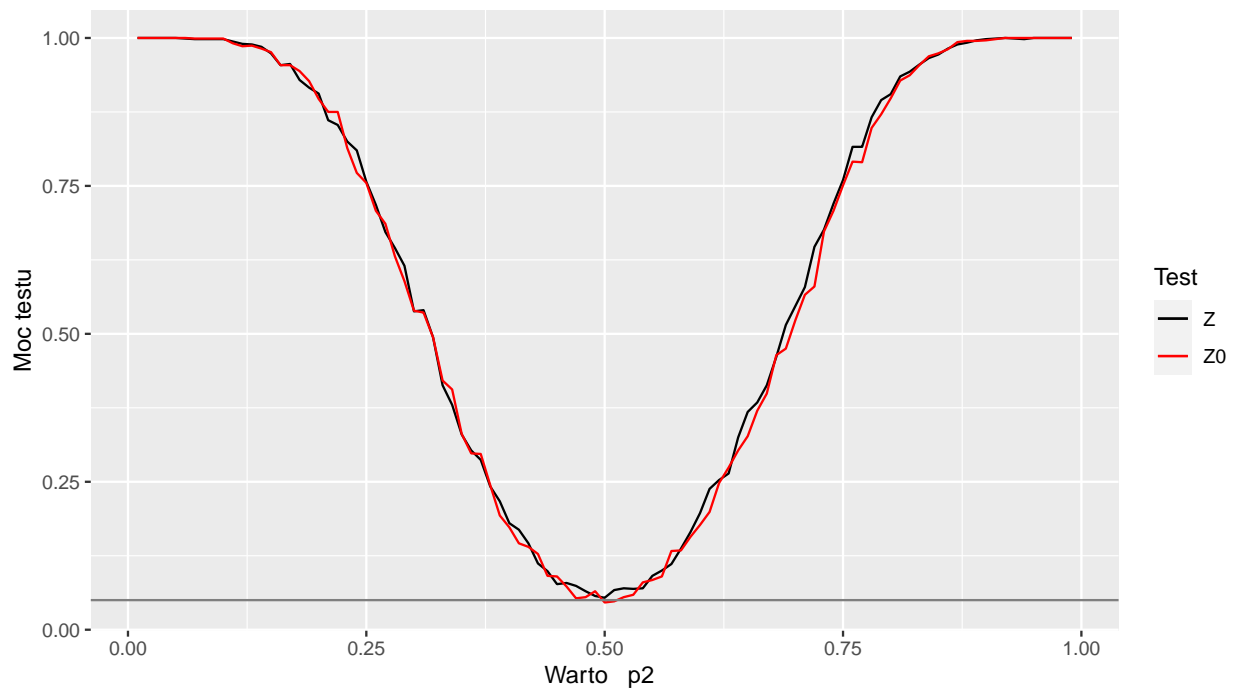
```

moc <- function(n, test){
  MC <- 1000
  p2 <- seq(0.01,0.99,0.01)
  p1 <- 0.5
  m <- length(p2)
  res <- rep(NA, m)
  for (i in 1:m){
    counter <- 0
    for (j in 1:MC){
      X <- factor(sample(c("1","0"), n, replace=TRUE, prob = c(p1,1-p1)), levels = 0:1)
      Y <- factor(sample(c("1","0"), n, replace=TRUE, prob = c(p2[i],1-p2[i])), levels=0:1)
      tab <- ftable(X,Y)
      if (test(tab) < 0.05){
        counter <- counter+1
      }
    }
    res[i] <- counter/MC
  }
  return(data.frame( 'prob' = p2, 'results' = res))
}

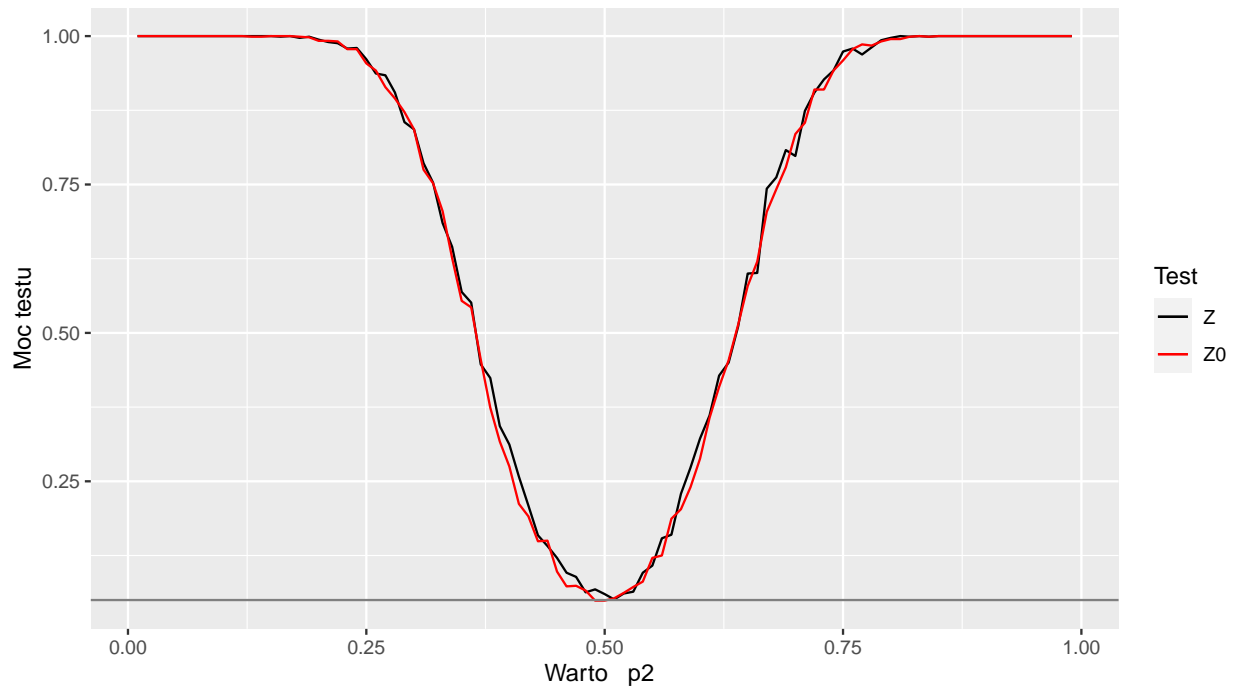
```



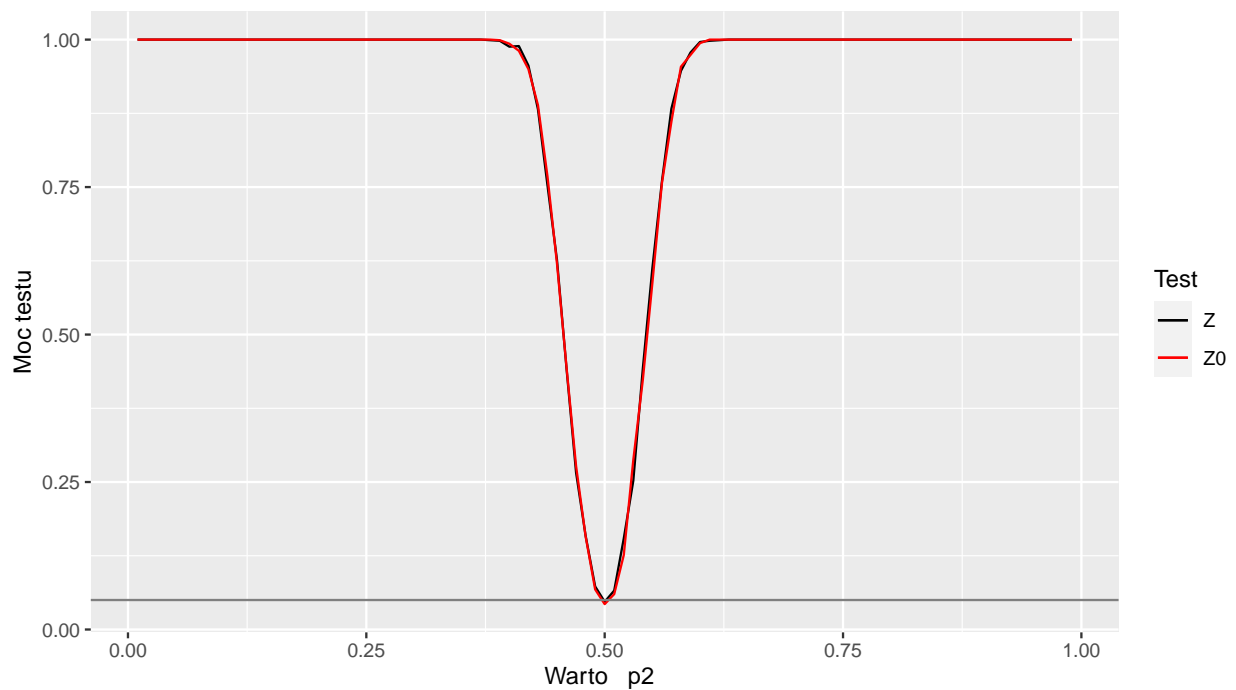
Wykres 1: Wykres funkcji mocy testu  $Z$  i  $Z_0$  dla  $n = 20$ .



Wykres 2: Wykres funkcji mocy testu  $Z$  i  $Z_0$  dla  $n = 50$ .



Wykres 3: Wykres funkcji mocy testu  $Z$  i  $Z_0$  dla  $n = 100$ .



Wykres 4: Wykres funkcji mocy testu  $Z$  i  $Z_0$  dla  $n = 1000$ .

## Wnioski

Na wykresach powyżej przedstawiliśmy moce testów  $Z$  i  $Z_0$  dla  $n \in \{20, 50, 100, 1000\}$ , na podstawie symulacji Monte Carlo. Szarą linią na wykresach oznaczono poziom istotności  $\alpha = 0.05$ . Funkcja mocy testu

powinna przechodzić przez wartość poziomu istotności w punkcie  $p_2 = 0.5$ , ponieważ jest to miejsce, w którym oba prawdopodobieństwa są takie same, a wtedy hipoteza zerowa powinna być przyjmowana z prawdopodobieństwem  $1 - \alpha$ . Dla  $n = 20$  funkcja mocy dla testu  $Z$  jest lekko powyżej oczekiwanej wartości, ale wraz ze zwiększaniem się  $n$ , funkcja coraz bardziej zbliża się do pożądanej wartości. Można na tej podstawie wyciągnąć wniosek, że test  $Z$  jest testem asymptotycznie nieobciążonym. Dla testu  $Z_0$  sytuacja jest podobna, jednak wartość funkcji mocy dla najmniejszego rozważanego  $n$ , dla  $p_2 = 0.5$  jest trochę mniejsza niż założony poziom istotności  $\alpha$ , ale znów, ze wzrostem wartości  $n$ , zbliża się ona do poziomu istotności, więc podobnie jak przy teście  $Z$ , można wyciągnąć wniosek, że test  $Z_0$  jest testem asymptotycznie nieobciążonym. Natomiast obciążenie obydwu testów dla małych  $n$  nie jest duża. Dla zwiększających się wartości  $n$  widzimy, że wartości funkcji mocy są większe, dla  $p_2 \neq 0.5$ . Było to do przewidzenia, ponieważ wraz ze wzrastającą liczbą prób (ankietowanych), test powinien być częściej odrzucany dla  $p_1 \neq p_2$ , bo moc testu rośnie.

## Zadanie 5

Przyjmujemy za zmienną 1 zmienną  $S$  (zajmowane stanowisko), za zmienną 2 – zmienną  $W1$  (zadowolenie z wynagrodzenia w pierwszym badanym okresie) i za zmienną 3 – zmienną  $Wyk$ .

- [13] zmienne „S” i „Wyk” mają dowolne rozkłady oraz zmienne te są niezależne, a zmienna „W1” ma rozkład równomierny

```
model_a <- glm(Freq ~ S + Wyk,
               data = df, family = poisson)
p_a <- 1-pchisq(deviance(model_a), df = df.residual(model_a))
cbind(model_a$data, fitted = fitted(model_a)) %>%
  kable(caption = "Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami") %>%
  column_spec(1, border_left = TRUE) %>%
  column_spec(5, border_right = TRUE) %>%
  kable_styling(latex_options = "HOLD_position")
```

Tab. 4: Porównanie wyznaczonych licznosci na podstawie modeli z rzeczywistymi licznosciami danych

S	W1	Wyk	Freq	fitted
0	-2	1	19	8.86625
1	-2	1	1	1.38375
0	-1	1	3	8.86625
1	-1	1	0	1.38375
0	1	1	0	8.86625
1	1	1	0	1.38375
0	2	1	18	8.86625
1	2	1	0	1.38375
0	-2	2	40	30.27500
1	-2	2	5	4.72500
0	-1	2	15	30.27500
1	-1	2	2	4.72500
0	1	2	0	30.27500
1	1	2	0	4.72500
0	2	2	68	30.27500
1	2	2	10	4.72500
0	-2	3	5	4.10875
1	-2	3	4	0.64125
0	-1	3	0	4.10875
1	-1	3	0	0.64125
0	1	3	0	4.10875
1	1	3	2	0.64125
0	2	3	5	4.10875
1	2	3	3	0.64125

P-wartość 0 jest mniejsza niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Nasze dane nie pochodzą z modelu [13].

- [13] zmienne „S” i „Wyk” mają dowolne rozkłady oraz zmienne te nie są niezależne, a zmienna „W1” ma rozkład równomierny

```
model_b <- glm(Freq ~ S + Wyk + S*Wyk,
               data = df, family = poisson)
p_b <- 1-pchisq(deviance(model_b), df = df.residual(model_b))
cbind(model_b$data, fitted = fitted(model_b)) %>%
  kable(caption = "Porównanie wyznaczonych licznosci na podstawie modeli z rzeczywistymi licznosciami danych",
        column_spec(1, border_left = TRUE) %>%
        column_spec(5, border_right = TRUE) %>%
        kable_styling(latex_options = "HOLD_position"))
```

Tab. 5: Porównanie wyznaczonych licznosci na podstawie modeli z rzeczywistymi licznosciami danych

S	W1	Wyk	Freq	fitted
0	-2	1	19	10.00
1	-2	1	1	0.25
0	-1	1	3	10.00
1	-1	1	0	0.25
0	1	1	0	10.00
1	1	1	0	0.25
0	2	1	18	10.00
1	2	1	0	0.25
0	-2	2	40	30.75
1	-2	2	5	4.25
0	-1	2	15	30.75
1	-1	2	2	4.25
0	1	2	0	30.75
1	1	2	0	4.25
0	2	2	68	30.75
1	2	2	10	4.25
0	-2	3	5	2.50
1	-2	3	4	2.25
0	-1	3	0	2.50
1	-1	3	0	2.25
0	1	3	0	2.50
1	1	3	2	2.25
0	2	3	5	2.50
1	2	3	3	2.25

P-wartość 0 jest mniejsza niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Nasze dane nie pochodzą z modelu [13].

- [1 2 3] zmienne „S”, „W1” i „Wyk” są wzajemnie niezależne

```
model_c <- glm(Freq ~ S + Wyk + W1,
               data = df, family = poisson)
p_c <- 1-pchisq(deviance(model_c), df = df.residual(model_c))
cbind(model_c$data, fitted = fitted(model_c)) %>%
  kable(caption = "Porównanie wyznaczonych licznosci na podstawie modeli z rzeczywistymi licznosciami danych",
        column_spec(1, border_left = TRUE) %>%
        column_spec(5, border_right = TRUE) %>%
        kable_styling(latex_options = "HOLD_position"))
```



Tab. 6: Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danych

S	W1	Wyk	Freq	fitted
0	-2	1	19	13.12205
1	-2	1	1	2.04795
0	-1	1	3	3.54650
1	-1	1	0	0.55350
0	1	1	0	0.35465
1	1	1	0	0.05535
0	2	1	18	18.44180
1	2	1	0	2.87820
0	-2	2	40	44.80700
1	-2	2	5	6.99300
0	-1	2	15	12.11000
1	-1	2	2	1.89000
0	1	2	0	1.21100
1	1	2	0	0.18900
0	2	2	68	62.97200
1	2	2	10	9.82800
0	-2	3	5	6.08095
1	-2	3	4	0.94905
0	-1	3	0	1.64350
1	-1	3	0	0.25650
0	1	3	0	0.16435
1	1	3	2	0.02565
0	2	3	5	8.54620
1	2	3	3	1.33380

P-wartość  $6.18728 \times 10^{-4}$  jest mniejsza niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Nasze dane nie pochodzą z modelu [1 2 3].

- [123] zmienna “Wyk” jest niezależna od zmiennych “S” i “W1”, ale zmienne “S” i “W1” nie są niezależne

```
model_d <- glm(Freq ~ S + Wyk + W1 + S*W1,
               data = df, family = poisson)
p_d <- 1-pchisq(deviance(model_d), df = df.residual(model_d))
cbind(model_d$data, fitted = fitted(model_d)) %>%
  kable(caption = "Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danych",
        column_spec(1, border_left = TRUE) %>%
        column_spec(5, border_right = TRUE) %>%
        kable_styling(latex_options = "HOLD_position"))
```

Tab. 7: Porównanie wyznaczonych licznosci na podstawie modeli z rzeczywistymi licznosciami danych

S	W1	Wyk	Freq	fitted
0	-2	1	19	13.120
1	-2	1	1	2.050
0	-1	1	3	3.690
1	-1	1	0	0.410
0	1	1	0	0.000
1	1	1	0	0.410
0	2	1	18	18.655
1	2	1	0	2.665
0	-2	2	40	44.800
1	-2	2	5	7.000
0	-1	2	15	12.600
1	-1	2	2	1.400
0	1	2	0	0.000
1	1	2	0	1.400
0	2	2	68	63.700
1	2	2	10	9.100
0	-2	3	5	6.080
1	-2	3	4	0.950
0	-1	3	0	1.710
1	-1	3	0	0.190
0	1	3	0	0.000
1	1	3	2	0.190
0	2	3	5	8.645
1	2	3	3	1.235

P-wartość 0.0021231 jest mniejsza niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Nasze dane nie pochodzą z modelu [12 3].

- [12 13] przy ustalonej wartości zmiennej “S”, zmienne “W1” i “Wyk” są niezależne

```
model_e <- glm(Freq ~ S + Wyk + W1 + S*W1 + S*Wyk,
               data = df, family = poisson)
p_e <- 1-pchisq(deviance(model_e), df = df.residual(model_e))
cbind(model_e$data, fitted = fitted(model_e)) %>%
  kable(caption = "Porównanie wyznaczonych licznosci na podstawie modeli z rzeczywistymi licznosciami danych",
        column_spec(1, border_left = TRUE) %>%
        column_spec(5, border_right = TRUE) %>%
        kable_styling(latex_options = "HOLD_position"))
```

Tab. 8: Porównanie wyznaczonych licznosci na podstawie modeli z rzeczywistymi licznosciami danych

S	W1	Wyk	Freq	fitted
0	-2	1	19	14.7976879
1	-2	1	1	0.3703704
0	-1	1	3	4.1618497
1	-1	1	0	0.0740741
0	1	1	0	0.0000000
1	1	1	0	0.0740741
0	2	1	18	21.0404624
1	2	1	0	0.4814815
0	-2	2	40	45.5028902
1	-2	2	5	6.2962963
0	-1	2	15	12.7976879
1	-1	2	2	1.2592593
0	1	2	0	0.0000000
1	1	2	0	1.2592593
0	2	2	68	64.6994220
1	2	2	10	8.1851852
0	-2	3	5	3.6994220
1	-2	3	4	3.3333333
0	-1	3	0	1.0404624
1	-1	3	0	0.6666667
0	1	3	0	0.0000000
1	1	3	2	0.6666667
0	2	3	5	5.2601156
1	2	3	3	4.3333333

P-wartość 0.251222 jest większa niż założony poziom istotności  $\alpha$ , więc nie mamy podstaw do odrzucenia hipotezy zerowej, czyli zakładamy, że nasze dane pochodzą z modelu [12 13].

- [1 23] zmienna “S” jest niezależna od zmiennych “Wyk” i “W1”, ale zmienne “Wyk” i “W1” nie są niezależne

```
model_f <- glm(Freq ~ S + Wyk + W1 + W1*Wyk,
               data = df, family = poisson)
p_f <- 1-pchisq(deviance(model_f), df = df.residual(model_f))
cbind(model_f$data, fitted = fitted(model_f)) %>%
  kable(caption = "Porównanie wyznaczonych licznosci na podstawie modeli z rzeczywistymi licznosciami danych",
        column_spec(1, border_left = TRUE) %>%
        column_spec(5, border_right = TRUE) %>%
        kable_styling(latex_options = "HOLD_position"))
```

Tab. 9: Porównanie wyznaczonych licznosci na podstawie modeli z rzeczywistymi licznosciami danych

S	W1	Wyk	Freq	fitted
0	-2	1	19	17.300
1	-2	1	1	2.700
0	-1	1	3	2.595
1	-1	1	0	0.405
0	1	1	0	0.000
1	1	1	0	0.000
0	2	1	18	15.570
1	2	1	0	2.430
0	-2	2	40	38.925
1	-2	2	5	6.075
0	-1	2	15	14.705
1	-1	2	2	2.295
0	1	2	0	0.000
1	1	2	0	0.000
0	2	2	68	67.470
1	2	2	10	10.530
0	-2	3	5	7.785
1	-2	3	4	1.215
0	-1	3	0	0.000
1	-1	3	0	0.000
0	1	3	0	1.730
1	1	3	2	0.270
0	2	3	5	6.920
1	2	3	3	1.080

P-wartość 0.01286 jest mniejsza niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Nasze dane nie pochodzą z modelu [1 23].

Tab. 10: P-wartości testów statystycznych

	Modele					
	[1 3]	[13]	[1 2 3]	[12 3]	[12 13]	[1 23]
P-wartość	0	0	$6.18728 \times 10^{-4}$	0.0021231	0.251222	0.01286

## Zadanie 6

Przyjmujemy za zmienną 1 zmienną  $S$  (zajmowane stanowisko), za zmienną 2 – zmienną  $P$  (płeć) i za zmienną 3 – zmienną  $Wyk$  (wykształcenie).

- [1 3] zmienne “S” i “Wyk” mają dowolne rozkłady oraz zmienne te są niezależne, a zmienna “P” ma rozkład równomierny

```
model_a <- glm(Freq ~ S + Wyk,
               data = df, family = poisson)
p_a <- 1-pchisq(deviance(model_a), df = df.residual(model_a))
cbind(model_a$data, fitted = fitted(model_a)) %>%
```

```
kable(caption = "Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danyh",
      column_spec(1, border_left = TRUE) %>%
      column_spec(5, border_right = TRUE) %>%
      kable_styling(latex_options = "HOLD_position"))
```

Tab. 11: Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danyh

S	P	Wyk	Freq	fitted
0	K	1	1	17.7325
1	K	1	0	2.7675
0	M	1	39	17.7325
1	M	1	1	2.7675
0	K	2	54	60.5500
1	K	2	4	9.4500
0	M	2	69	60.5500
1	M	2	13	9.4500
0	K	3	8	8.2175
1	K	3	4	1.2825
0	M	3	2	8.2175
1	M	3	5	1.2825

P-wartość  $1.6342483 \times 10^{-13}$  jest mniejsza niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Nasze dane nie pochodzą z modelu [1 3].

- [13] zmienne “S” i “Wyk” mają dowolne rozkłady oraz zmienne te nie są niezależne, a zmienna “P” ma rozkład równomierny

```
model_b <- glm(Freq ~ S + Wyk + S*Wyk,
               data = df, family = poisson)
p_b <- 1-pchisq(deviance(model_b), df = df.residual(model_b))
cbind(model_b$data, fitted = fitted(model_b)) %>%
  kable(caption = "Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danyh",
        column_spec(1, border_left = TRUE) %>%
        column_spec(5, border_right = TRUE) %>%
        kable_styling(latex_options = "HOLD_position"))
```

Tab. 12: Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danyh

S	P	Wyk	Freq	fitted
0	K	1	1	20.0
1	K	1	0	0.5
0	M	1	39	20.0
1	M	1	1	0.5
0	K	2	54	61.5
1	K	2	4	8.5
0	M	2	69	61.5
1	M	2	13	8.5
0	K	3	8	5.0
1	K	3	4	4.5
0	M	3	2	5.0
1	M	3	5	4.5

P-wartość  $9.9507624 \times 10^{-11}$  jest mniejsza niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Nasze dane nie pochodzą z modelu [13].

- [1 2 3] zmienne “S”, “P” i “Wyk” są wzajemnie niezależne

```
model_c <- glm(Freq ~ S + Wyk + P,
               data = df, family = poisson)
p_c <- 1-pchisq(deviance(model_c), df = df.residual(model_c))
cbind(model_c$data, fitted = fitted(model_c)) %>%
  kable(caption = "Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danych",
        column_spec(1, border_left = TRUE) %>%
        column_spec(5, border_right = TRUE) %>%
        kable_styling(latex_options = "HOLD_position"))
```

Tab. 13: Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danych

S	P	Wyk	Freq	fitted
0	K	1	1	12.590075
1	K	1	0	1.964925
0	M	1	39	22.874925
1	M	1	1	3.570075
0	K	2	54	42.990500
1	K	2	4	6.709500
0	M	2	69	78.109500
1	M	2	13	12.190500
0	K	3	8	5.834425
1	K	3	4	0.910575
0	M	3	2	10.600575
1	M	3	5	1.654425

P-wartość  $1.2979651 \times 10^{-10}$  jest mniejsza niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Nasze dane nie pochodzą z modelu [1 2 3].

- [12 3] zmienna “Wyk” jest niezależna od zmiennych “S” i “P”, ale zmienne “S” i “P” nie są niezależne

```
model_d <- glm(Freq ~ S + Wyk + P + S*P,
               data = df, family = poisson)
p_d <- 1-pchisq(deviance(model_d), df = df.residual(model_d))
cbind(model_d$data, fitted = fitted(model_d)) %>%
  kable(caption = "Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danych",
        column_spec(1, border_left = TRUE) %>%
        column_spec(5, border_right = TRUE) %>%
        kable_styling(latex_options = "HOLD_position"))
```

Tab. 14: Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danych

S	P	Wyk	Freq	fitted
0	K	1	1	12.915
1	K	1	0	1.640
0	M	1	39	22.550
1	M	1	1	3.895
0	K	2	54	44.100
1	K	2	4	5.600
0	M	2	69	77.000
1	M	2	13	13.300
0	K	3	8	5.985
1	K	3	4	0.760
0	M	3	2	10.450
1	M	3	5	1.805

P-wartość  $4.8342885 \times 10^{-11}$  jest mniejsza niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Nasze dane nie pochodzą z modelu [12 3].

- [12 13] przy ustalonej wartości zmiennej “S”, zmienne “P” i “Wyk” są niezależne

```
model_e <- glm(Freq ~ S + Wyk + P + S*P + S*Wyk,
               data = df, family = poisson)
p_e <- 1-pchisq(deviance(model_e), df = df.residual(model_e))
cbind(model_e$data, fitted = fitted(model_e)) %>%
  kable(caption = "Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danych",
        column_spec(1, border_left = TRUE) %>%
        column_spec(5, border_right = TRUE) %>%
        kable_styling(latex_options = "HOLD_position"))
```

Tab. 15: Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danych

S	P	Wyk	Freq	fitted
0	K	1	1	14.5664740
1	K	1	0	0.2962963
0	M	1	39	25.4335260
1	M	1	1	0.7037037
0	K	2	54	44.7919075
1	K	2	4	5.0370370
0	M	2	69	78.2080925
1	M	2	13	11.9629630
0	K	3	8	3.6416185
1	K	3	4	2.6666667
0	M	3	2	6.3583815
1	M	3	5	6.3333333

P-wartość  $3.0176519 \times 10^{-8}$  jest mniejsza niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Nasze dane nie pochodzą z modelu [12 13].

- [123] zmienna “S” jest niezależna od zmiennych “Wyk” i “P”, ale zmienne “Wyk” i “P” nie są niezależne

```

model_f <- glm(Freq ~ S + Wyk + P + P*Wyk,
               data = df, family = poisson)
p_f <- 1-pchisq(deviance(model_f), df = df.residual(model_f))
cbind(model_f$data, fitted = fitted(model_f)) %>%
  kable(caption = "Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami",
        column_spec(1, border_left = TRUE) %>%
        column_spec(5, border_right = TRUE) %>%
        kable_styling(latex_options = "HOLD_position"))

```

Tab. 16: Porównanie wyznaczonych licznosci na podstawie modelow z rzeczywistymi licznosciami danych

S	P	Wyk	Freq	fitted
0	K	1	1	0.865
1	K	1	0	0.135
0	M	1	39	34.600
1	M	1	1	5.400
0	K	2	54	50.170
1	K	2	4	7.830
0	M	2	69	70.930
1	M	2	13	11.070
0	K	3	8	10.380
1	K	3	4	1.620
0	M	3	2	6.055
1	M	3	5	0.945

P-wartość  $1.7468073 \times 10^{-4}$  jest mniejsza niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Nasze dane nie pochodzą z modelu [1 23].

Tab. 17: P-wartości testów statystycznych

Model	p-wartość
[1 3]	$1.6342483 \times 10^{-13}$
[13]	$9.9507624 \times 10^{-11}$
[1 2 3]	$1.2979651 \times 10^{-10}$
[12 3]	$4.8342885 \times 10^{-11}$
[12 13]	$3.0176519 \times 10^{-8}$
[1 23]	$1.7468073 \times 10^{-4}$

## Zadanie 7

Do zmiennych  $S$ ,  $W1$  i  $Wyk$  przyjmujemy model log-liniowy [13 23] oraz [123] i na tej podstawie obliczymy prawdopodobieństwa.

- Prawdopodobieństwo, że osoba pracująca na stanowisku kierowniczym jest zdecydowanie zadowolona ze swojego wynagrodzenia.

Przykładowy kod obliczający to prawdopodobieństwo dla modelu [13 23]:



```
sum(result1$`fitted(model1)`[result1$S == 1 & result1$W1 == 2])/
(sum(result1$`fitted(model1)`[result1$S == 1]))
```

```
## [1] 0.5074047
```

Tab. 18: Oszacowane prawdopodobieństwa

	Prawdopodobieństwo
Wartość empiryczna	0.4814815
Model [13 23]	0.5074047
Model [123]	0.4814815

Na podstawie wyników przedstawionych w tabeli możemy zauważyć, że do naszych danych lepiej niż model [13 23] dopasował się model [123] - prawdopodobieństwo obliczone z wykorzystaniem tego modelu jest identyczne jak prawdopodobieństwo empiryczne.

- Prawdopodobieństwo, że osoba z wykształceniem zawodowym pracuje na stanowisku kierowniczym.

Tab. 19: Oszacowane prawdopodobieństwa

	Prawdopodobieństwo
Wartość empiryczna	0.0243902
Model [13 23]	0.0243902
Model [123]	0.0243902

Prawdopodobieństwa obliczone z wykorzystaniem obu modeli są identyczne jak prawdopodobieństwo empiryczne. Okazuje się, że prawdopodobieństwo, że osoba z wykształceniem zawodowym pracuje na na stanowisku kierowniczym jest niezwykle małe.

- Prawdopodobieństwo, że osoba z wykształceniem wyższym nie pracuje na stanowisku kierowniczym.

Tab. 20: Oszacowane prawdopodobieństwa

	Prawdopodobieństwo
Wartość empiryczna	0.5263158
Model [13 23]	0.5263158
Model [123]	0.5263158

Znów prawdopodobieństwa obliczone z wykorzystaniem modeli są identyczne jak prawdopodobieństwo empiryczne.

## Zadanie 8

Do zmiennych  $S$ ,  $P$  i  $Wyk$  przyjmijmy model log-liniowy [13 23] i na tej podstawie obliczymy prawdopodobieństwo.

- Prawdopodobieństwo, że osoba pracująca na stanowisku kierowniczym jest kobietą.

Przykładowe wywołanie:

```
m <- sum(result$`fitted(model)`[result$S == 1 & result$P == 'K'])/
  (sum(result$`fitted(model)`[result$S == 1]))
emp <- sum(result$Freq[result$S == 1 & result$P == 'K'])/
  (sum(result$Freq[result$S == 1]))
```

Tab. 21: Oszacowane prawdopodobieństwa

	Prawdopodobieństwo
Wartość empiryczna	0.2962963
Model [13 23]	0.4722762

Prawdopodobieństwo uzyskane z wykorzystaniem modelu [13 23] dość mocno odbiegają od otrzymanego wyniku empirycznego.

- Prawdopodobieństwo, że osoba z wykształceniem zawodowym pracuje na stanowisku kierowniczym.

Tab. 22: Oszacowane prawdopodobieństwa

	Prawdopodobieństwo
Wartość empiryczna	0.0243902
Model [13 23]	0.0243902

Prawdopodobieństwo otrzymane za pomocą modelu jest identyczne jak to otrzymane z wykorzystaniem modelu.

Tab. 23: Porównanie prawdopodobieństw tego samego problemu uzyskanych w zadaniu 7. i 8.

	Prawdopodobieństwo
Zadanie 7	0.0243902
Zadanie 8	0.0243902

Prawdopodobieństwa tego samego problemu z zadania 7. i 8. otrzymane za pomocą różnych modeli log-liniowych są identyczne.

- Prawdopodobieństwo, że osoba z wykształceniem wyższym jest mężczyzną.

Tab. 24: Oszacowane prawdopodobieństwa

	Prawdopodobieństwo
Wartość empiryczna	0.3684211
Model [13 23]	0.3684211

Prawdopodobieństwo otrzymane za pomocą modelu jest identyczne jak to otrzymane z wykorzystaniem modelu.

## Zadanie 9

W tym zadaniu będziemy testować hipotezy zerowe przeciwko dwóm pewnym modelom, w których jeden jest pełny (zawierają wszystkie interakcje) a drugi jest nadmodelem modelu z hipotezy zerowej, ale nie jest modelem pełnym. Hipotezy testujemy na poziomie istotności  $\alpha = 0.05$ .

- $H_0$ : Dane pochodzą z modelu  $[1\ 2\ 3]$  przeciwko
- $H_1$ : Dane pochodzą z modelu pełnego  $[123]$
- $H_1$ : Dane pochodzą z modelu  $[12\ 3]$

Przykładowy kod, w którym dopasowujemy modele do naszych danych:

```
m_0 <- glm(Freq ~ S+W1+Wyk, data = df_W1, family = poisson)
m_1 <- glm(Freq ~ S+W1+Wyk+S*W1+Wyk*W1+S*Wyk+S*W1*Wyk, data = df_W1, family = poisson)
m_2 <- glm(Freq ~ S+W1+Wyk+S*W1, data = df_W1, family = poisson)
```

Prawdopodobieństwo dla pierwszej z hipotez alternatywnych  $6.18728 \times 10^{-4}$  oraz dla drugiej 0.0396896 są mniejsze niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową.

- $H_0$ : Dane pochodzą z modelu  $[2\ 13]$  przeciwko
- $H_1$ : Dane pochodzą z modelu  $[123]$
- $H_1$ : Dane pochodzą z modelu  $[13\ 23]$

```
m_0 <- glm(Freq ~ S+W1+Wyk+ S*Wyk, data = df_W1, family = poisson)
m_1 <- glm(Freq ~ S+W1+Wyk+S*W1+Wyk*W1+S*Wyk+S*W1*Wyk, data = df_W1, family = poisson)
m_2 <- glm(Freq ~ S+W1+Wyk+Wyk*W1+S*Wyk, data = df_W1, family = poisson)

test <- anova(m_0, m_1)
p_1 <- 1-pchisq(test$Deviance[2], df = test$Df[2])

test <- anova(m_0, m_2)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Prawdopodobieństwo dla pierwszej z hipotez alternatywnych 0.0809632 jest większe niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej. Natomiast dla drugiej 0.0055869 jest mniejszy niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową.

- $H_0$ : Dane pochodzą z modelu  $[13\ 23]$  przeciwko
- $H_1$ : Dane pochodzą z modelu  $[123]$
- $H_1$ : Dane pochodzą z modelu  $[12\ 13\ 23]$

```
m_0 <- glm(Freq ~ S+W1+Wyk+ S*Wyk+W1*Wyk, data = df_W1, family = poisson)
m_1 <- glm(Freq ~ S+W1+Wyk+S*W1+Wyk*W1+S*Wyk+S*W1*Wyk, data = df_W1, family = poisson)
m_2 <- glm(Freq ~ S+W1+Wyk+S*W1+Wyk*W1+S*Wyk+S*W1, data = df_W1, family = poisson)

test <- anova(m_0, m_1)
p_1 <- 1-pchisq(test$Deviance[2], df = test$Df[2])

test <- anova(m_0, m_2)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Prawdopodobieństwo dla pierwszej z hipotez alternatywnych 0.8446445 oraz dla drugiej 0.349985 jest większe niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

- $H_0$ : Dane pochodzą z modelu [13 23] przeciwko
- $H_1$ : Dane pochodzą z modelu [123]
- $H_1$ : Dane pochodzą z modelu [12 13 23]

```
m_0 <- glm(Freq ~ S+P+Wyk+S*Wyk+P*Wyk, data = df_P, family = poisson)
m_1 <- glm(Freq ~ S+P+Wyk+S*P+Wyk*P+S*Wyk+S*P*Wyk, data = df_P, family = poisson)
m_2 <- glm(Freq ~ S+P+Wyk+S*P+Wyk*P+S*Wyk+S*P, data = df_P, family = poisson)

test <- anova(m_0, m_1)
p_1 <- 1-pchisq(test$Deviance[2], df = test$Df[2])

test <- anova(m_0, m_2)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Prawdopodobieństwo dla drugiej z hipotez alternatywnych 0.0244872 jest mniejszy niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową. Natomiast dla pierwszej 0.1446957 jest mniejszy, zatem nie mamy podstaw do odrzucenia hipotezy zerowej.

## Zadanie 10

### Testy

Wybór modelu w oparciu o testy został wykonany w ten sposób, że najpierw wzięliśmy model, w którym nie występują żadne interakcje [1 2 3], wykonaliśmy test ilorazu wiarygodności, gdzie hipotezą alternatywną  $H_1$  był nadmodel modelu z hipotezy alternatywnej. Jeśli uzyskana p-wartość była większa niż założony poziom istotności  $\alpha = 0.05$ , to przyjmowaliśmy że model z hipotezy zerowej  $H_0$  lepiej opisuje nasze dane, w przeciwnym wypadku, dla następnych testów, model z hipotezy alternatywnej stawał się modelem hipotezy zerowej. Postępując w ten sposób doszliśmy do modelu, który według testów najlepiej opisuje dane.

- Model [1 2 3] przeciwko [12 3]

```
df <- as.data.frame(ftable(personel$A1,personel$W1,personel$P))
names(df) <- c('A1', 'W1', 'P', 'Freq')

m_0 <- glm(Freq ~ A1+W1+P, data = df, family = poisson)
m_1 <- glm(Freq ~ A1+W1+W1+A1*W1, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 0 jest mniejsza niż założony poziom istotności, więc odrzucamy hipotezę zerową. Model [12 3] staje się modelem wyjściowym.

- Model [12 3] przeciwko [12 13]

```
m_0 <- glm(Freq ~ A1+W1+P+A1*W1, data = df, family = poisson)
m_1 <- glm(Freq ~ A1+W1+P+A1*W1+A1*P, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 0.8022519 jest większa niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

- Model [12 3] przeciwko [12 23]

```
m_0 <- glm(Freq ~ A1+W1+P+A1*W1, data = df, family = poisson)
m_1 <- glm(Freq ~ A1+W1+P+A1*W1+W1*P, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 0.5349822 jest większa niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

- Model [12 3] przeciwko [12 13 23]

```
m_0 <- glm(Freq ~ A1+W1+P+A1*W1, data = df, family = poisson)
m_1 <- glm(Freq ~ A1+W1+P+A1*W1+W1*P+P*A1, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 0.775001 jest większa niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

- Model [12 3] przeciwko [123]

```
m_0 <- glm(Freq ~ A1+W1+P+A1*W1, data = df, family = poisson)
m_1 <- glm(Freq ~ A1+W1+P+A1*W1+W1*P+P*A1+A1*W1*P, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 0.9093978 jest większa niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

Z przeprowadzonych testów wynika, że model [12 3] najlepiej opisuje nasze dane.

## Kryterium AIC

```
models_AIC <- c(AIC(model_), AIC(model_1), AIC(model_2), AIC(model_3),
  AIC(model_1_2), AIC(model_1_3), AIC(model_2_3), AIC(model_12),
  AIC(model_13), AIC(model_23), AIC(model_1_2_3), AIC(model_12_3),
  AIC(model_13_2), AIC(model_1_23), AIC(model_12_23), AIC(model_13_23),
  AIC(model_12_13), AIC(model_12_23_13), AIC(model_123))
```

Tab. 25: Wartości obliczonych kryteriów AIC dla poszczególnych modeli

Model	Wartość kryterium
[ ]	582.2529387
[1]	484.1343098
[2]	427.4253204
[3]	567.1888702
[1 2]	329.3066915
[1 3]	469.0702413
[2 3]	412.3612519
[12]	138.6563749
[13]	475.4339317
[23]	416.1766225
[1 2 3]	314.242623
[12 3]	123.5923063
[13 2]	320.6063134
[1 23]	318.0579936
[12 23]	127.4076769
[13 23]	324.421684
[12 13]	129.9559967
[12 13 23]	133.5509426
[123]	150.1855778

Najmniejsza wartość kryterium informacyjnego wynosi 123.5923063, a otrzymano ją dla modelu [12 3], czyli tego samego który otrzymaliśmy z wykorzystaniem testów.

## Kryterium BIC

```
models_BIC <- c(BIC(model_), BIC(model_1), BIC(model_2), BIC(model_3),
  BIC(model_1_2), BIC(model_1_3), BIC(model_2_3), BIC(model_12),
  BIC(model_13), BIC(model_23), BIC(model_1_2_3), BIC(model_12_3),
  BIC(model_13_2), BIC(model_1_23), BIC(model_12_23), BIC(model_13_23),
  BIC(model_12_13), BIC(model_12_23_13), BIC(model_123))
```

Tab. 26: Wartości obliczonych kryteriów BIC dla poszczególnych modeli

Model	Wartość kryterium
[ ]	583.9418182
[1]	492.578707
[2]	434.1808382
[3]	570.5666291
[1 2]	342.8177271
[1 3]	479.203518
[2 3]	420.8056492
[12]	172.4339639
[13]	492.3227263
[23]	429.6876582
[1 2 3]	329.4425381
[12 3]	159.0587748
[13 2]	342.5617463
[1 23]	338.324547
[12 23]	167.9407838
[13 23]	351.4437553
[12 13]	172.1779831
[12 13 23]	180.8395673
[123]	217.740756

Najmniejsza wartość kryterium informacyjnego wynosi 159.0587748, a otrzymano ją dla modelu [12 3], czyli tego samego który otrzymaliśmy z wykorzystaniem testów i kryterium AIC.

## Metoda krokowa

```
step(model_123)
```

```
## Start:  AIC=150.19
## Freq ~ A1 + W1 + P + A1 * W1 + A1 * P + W1 * P + A1 * W1 * P
##
##           Df Deviance    AIC
## - A1:W1:P 12    7.3654 133.55
## <none>      0.0000 150.19
##
## Step:  AIC=133.55
## Freq ~ A1 + W1 + P + A1:W1 + A1:P + W1:P
##
##           Df Deviance    AIC
## - A1:P     4     9.222 127.41
## - W1:P     3     9.770 129.96
## <none>      7.365 133.55
## - A1:W1 12   222.236 324.42
##
## Step:  AIC=127.41
## Freq ~ A1 + W1 + P + A1:W1 + W1:P
##
##           Df Deviance    AIC
```

```
## - W1:P    3    11.407 123.59
## <none>          9.222 127.41
## - A1:W1 12   223.872 318.06
##
## Step:  AIC=123.59
## Freq ~ A1 + W1 + P + A1:W1
##
##           Df Deviance    AIC
## <none>          11.407 123.59
## - P           1    28.471 138.66
## - A1:W1 12    226.057 314.24

##
## Call:  glm(formula = Freq ~ A1 + W1 + P + A1:W1, family = poisson, data = df)
##
## Coefficients:
## (Intercept)      A1-1      A10      A11      A12      W1-1
##      1.5293      0.2076      1.0986     -1.4663     -1.4663    -22.1342
##           W11      W12      PM    A1-1:W1-1    A10:W1-1    A11:W1-1
##     -2.5649    -22.1342      0.5971     -0.2076     -1.0986     23.8688
##    A12:W1-1    A1-1:W11    A10:W11     A11:W11     A12:W11    A1-1:W12
##    22.1342    -19.7769    -1.0986    -18.1029    -18.1029     19.3616
##    A10:W12    A11:W12    A12:W12
##    -1.0986     25.4176     24.1711
##
## Degrees of Freedom: 39 Total (i.e. Null);  19 Residual
## Null Deviance:      510.1
## Residual Deviance: 11.41    AIC: 123.6
```

Z wykorzystaniem metody krokowej otrzymaliśmy ten sam model, co wyżej.

## Zadanie 11

### Testy

Analogicznie jak w zadaniu 10.

- Model [1 2 3] przeciwko [12 3]

```
m_0 <- glm(Freq ~ D+A1+P, data = df, family = poisson)
m_1 <- glm(Freq ~ D+A1+P+D*A1, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 0.2154821 jest większa niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

- Model [1 2 3] przeciwko [1 23]



```
m_0 <- glm(Freq ~ D+A1+P, data = df, family = poisson)
m_1 <- glm(Freq ~ D+A1+P+P*A1, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 0.8022519 jest większa niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

- Model [1 2 3] przeciwko [13 2]

```
m_0 <- glm(Freq ~ D+A1+P, data = df, family = poisson)
m_1 <- glm(Freq ~ D+A1+P+D*P, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość  $3.4586778 \times 10^{-12}$  jest mniejsza niż założony poziom istotności, więc odrzucamy hipotezę zerową. Model [13 2] staje się modelem wyjściowym.

- Model [13 2] przeciwko [12 13]

```
m_0 <- glm(Freq ~ D+A1+P+D*P, data = df, family = poisson)
m_1 <- glm(Freq ~ D+A1+P+D*A1+D*P, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 0.2154821 jest większa niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

- Model [13 2] przeciwko [13 23]

```
m_0 <- glm(Freq ~ D+A1+P+D*P, data = df, family = poisson)
m_1 <- glm(Freq ~ D+A1+P+D*P+A1*P, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 0.8022519 jest większa niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

- Model [13 2] przeciwko [12 23]

```
m_0 <- glm(Freq ~ D+A1+P+D*P, data = df, family = poisson)
m_1 <- glm(Freq ~ D+A1+P+D*A1+A1*P, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 1 jest większa niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

- Model [13 2] przeciwko [12 13 23]

```
m_0 <- glm(Freq ~ D+A1+P+D*P, data = df, family = poisson)
m_1 <- glm(Freq ~ D+A1+P+D*A1+A1*P+D*P, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 0.3751853 jest większa niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

- Model [13 2] przeciwko [123]

```
m_0 <- glm(Freq ~ D+A1+P+D*P, data = df, family = poisson)
m_1 <- glm(Freq ~ D+A1+P+D*A1+A1*P+D*P+A1*P*D, data = df, family = poisson)
test <- anova(m_0, m_1)
p_2 <- 1-pchisq(test$Deviance[2], df = test$Df[2])
```

Otrzymana p-wartość 0.3440413 jest większa niż założony poziom istotności, więc nie mamy podstaw do odrzucenia hipotezy zerowej.

Z przeprowadzonych testów wynika, że model [13 2] najlepiej opisuje nasze dane.

## Kryterium AIC

```
models_AIC <- c(AIC(model_), AIC(model_1), AIC(model_2), AIC(model_3),
  AIC(model_1_2), AIC(model_1_3), AIC(model_2_3), AIC(model_12),
  AIC(model_13), AIC(model_23), AIC(model_1_2_3), AIC(model_12_3),
  AIC(model_13_2), AIC(model_1_23), AIC(model_12_23), AIC(model_13_23),
  AIC(model_12_13), AIC(model_12_23_13), AIC(model_123))
```

Tab. 27: Wartości obliczonych kryteriów AIC dla poszczególnych modeli

Model	Wartość kryterium
[]	375.2883547
[1]	322.5160804
[2]	277.1697258
[3]	360.2242863
[1 2]	224.3974514
[1 3]	307.4520119
[2 3]	262.1056573
[12]	232.9025829
[13]	257.0563372
[23]	268.4693478
[1 2 3]	209.3333829
[12 3]	217.8385145
[13 2]	158.9377083
[1 23]	215.6970734
[12 23]	224.2022049
[13 23]	165.3013987
[12 13]	167.4428395
[12 13 23]	173.7756383
[123]	184.5309303

Najmniejsza wartość kryterium informacyjnego wynosi 158.9377083, a otrzymano ją dla modelu [13 2], czyli tego samego który otrzymaliśmy z wykorzystaniem testów.

## Kryterium BIC

```
models_BIC <- c(BIC(model_), BIC(model_1), BIC(model_2), BIC(model_3),
  BIC(model_1_2), BIC(model_1_3), BIC(model_2_3), BIC(model_12),
  BIC(model_13), BIC(model_23), BIC(model_1_2_3), BIC(model_12_3),
  BIC(model_13_2), BIC(model_1_23), BIC(model_12_23), BIC(model_13_23),
  BIC(model_12_13), BIC(model_12_23_13), BIC(model_123))
```

Tab. 28: Wartości obliczonych kryteriów BIC dla poszczególnych modeli

Model	Wartość kryterium
[]	376.9772342
[1]	329.2715982
[2]	285.6141231
[3]	363.6020452
[1 2]	237.908487
[1 3]	315.8964092
[2 3]	272.2389341
[12]	266.680172
[13]	270.5673729
[23]	285.3581423
[1 2 3]	224.533298
[12 3]	253.304983
[13 2]	179.2042617
[1 23]	237.6525063
[12 23]	266.4241913
[13 23]	192.32347
[12 13]	207.9759464
[12 13 23]	221.064263
[123]	252.0861085

Najmniejsza wartość kryterium informacyjnego wynosi 179.2042617, a otrzymano ją dla modelu [13 2], czyli tego samego który otrzymaliśmy z wykorzystaniem testów i kryterium AIC.

## Metoda krokowa

```
step(model_123)
```

```
## Start:  AIC=184.53
## Freq ~ D + A1 + P + D * A1 + D * P + A1 * P + D * A1 * P
##
##           Df Deviance    AIC
## - D:A1:P 12   13.245 173.78
## <none>      0.000 184.53
##
```

```

## Step: AIC=173.78
## Freq ~ D + A1 + P + D:A1 + D:P + A1:P
##
##      Df Deviance   AIC
## - D:A1 12   28.770 165.30
## - A1:P  4   14.912 167.44
## <none>      13.245 173.78
## - D:P   3   69.671 224.20
##
## Step: AIC=165.3
## Freq ~ D + A1 + P + D:P + A1:P
##
##      Df Deviance   AIC
## - A1:P  4   30.407 158.94
## <none>      28.770 165.30
## - D:P   3   85.166 215.70
##
## Step: AIC=158.94
## Freq ~ D + A1 + P + D:P
##
##      Df Deviance   AIC
## <none>      30.407 158.94
## - D:P   3   86.802 209.33
## - A1    4  136.525 257.06

##
## Call: glm(formula = Freq ~ D + A1 + P + D:P, family = poisson, data = df)
##
## Coefficients:
## (Intercept)      DP      DS      DZ      A1-1      A10
##    0.47623    -0.13976    0.04256   -1.74920    0.19416    1.04982
##      A11      A12      PM      DP:PM      DS:PM      DZ:PM
##    1.96611    0.72824   -2.03688    3.39786    1.90335    3.94642
##
## Degrees of Freedom: 39 Total (i.e. Null);  28 Residual
## Null Deviance:      268.8
## Residual Deviance: 30.41   AIC: 158.9

```

Z wykorzystaniem metody krokowej otrzymaliśmy ten sam model, co z wykorzystaniem metod wyżej.