

# Raport nr 3

Natalia Iwańska 262270, Klaudia Janicka 262268

2023-07-01

## Zadanie 1

Table 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 weryfikowaną hipotezę o symetrii, która jest równoważna hipotezie o brzegowej jednorodności należy odrzucić.

## Zadanie 2

Table 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

```
#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_z(tabela)
```

```
## [1] NaN
```

test	p-value
Test Z	NaN
Test Z0	NaN
McNemar test z poprawką na ciągłość	NaN
McNemar test bez poprawki	NaN

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

```
mc <- 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))
}

```

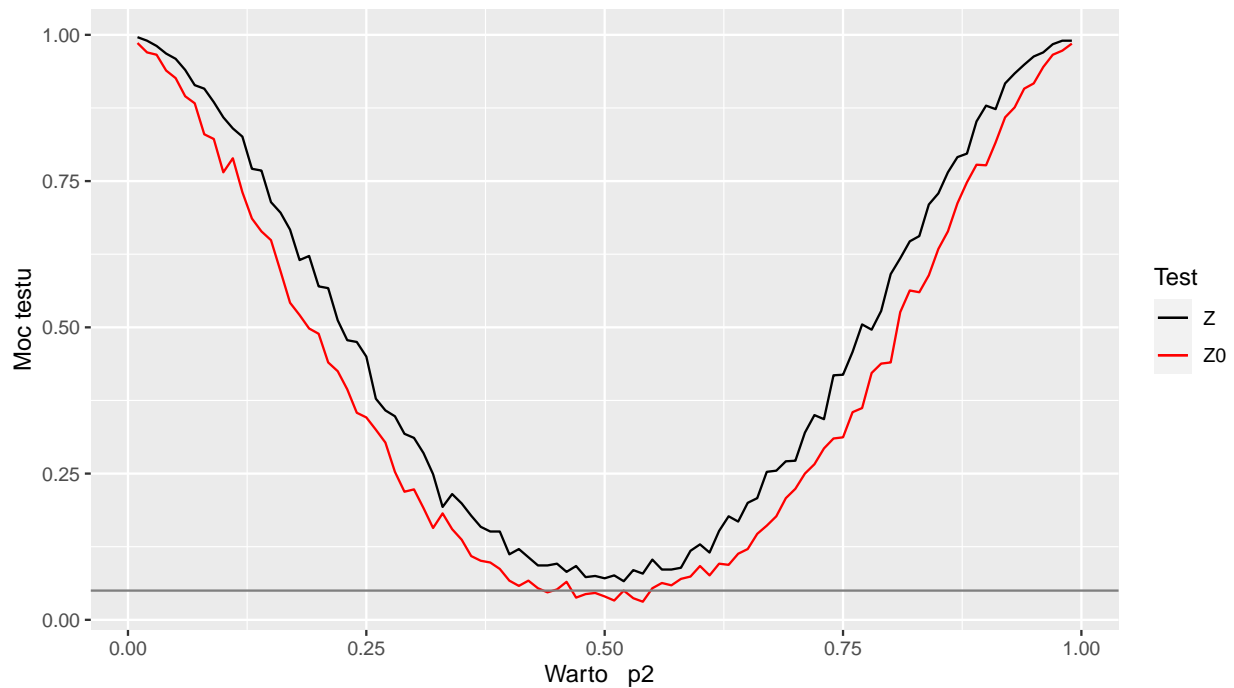


Figure 1: Wykres funkcji mocy testu  $Z$  i  $Z_0$  dla  $n = 20$

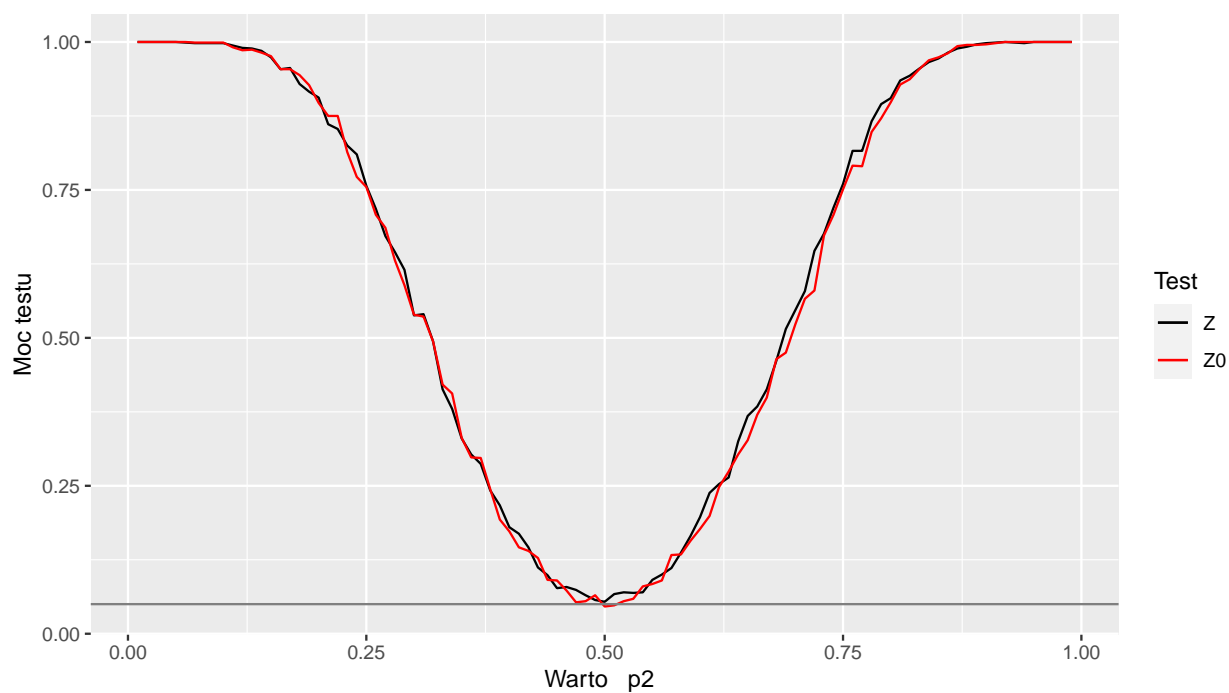


Figure 2: Wykres funkcji mocy testu  $Z$  i  $Z_0$  dla  $n = 50$

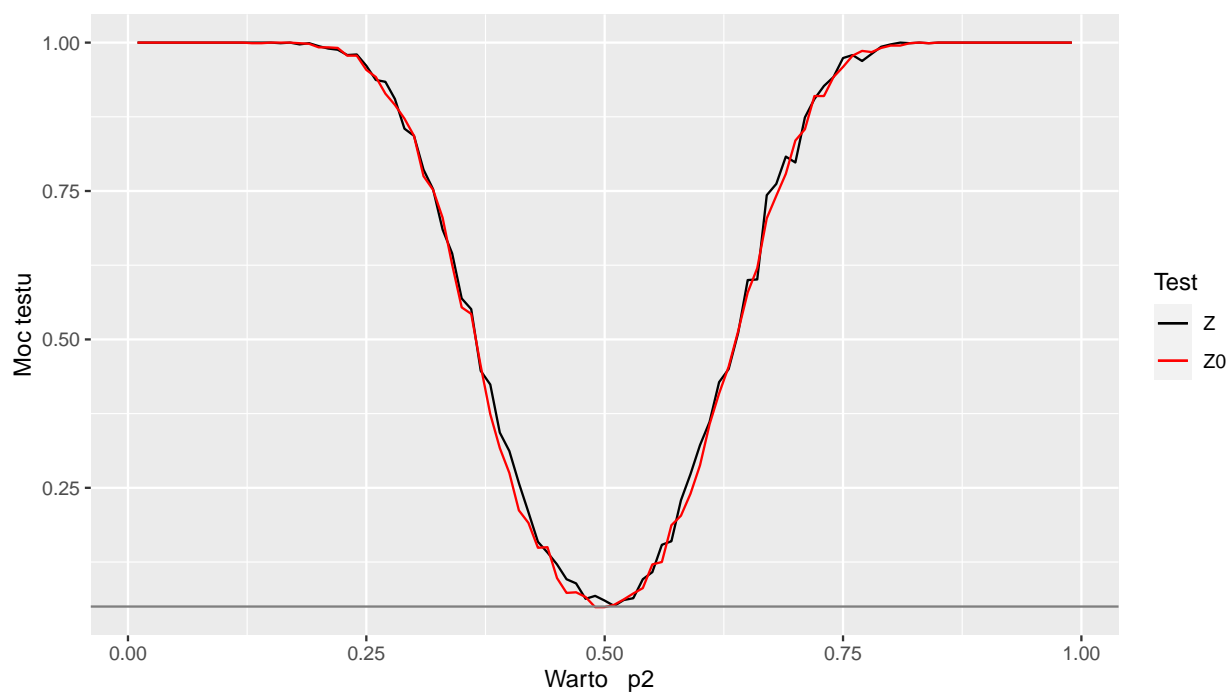
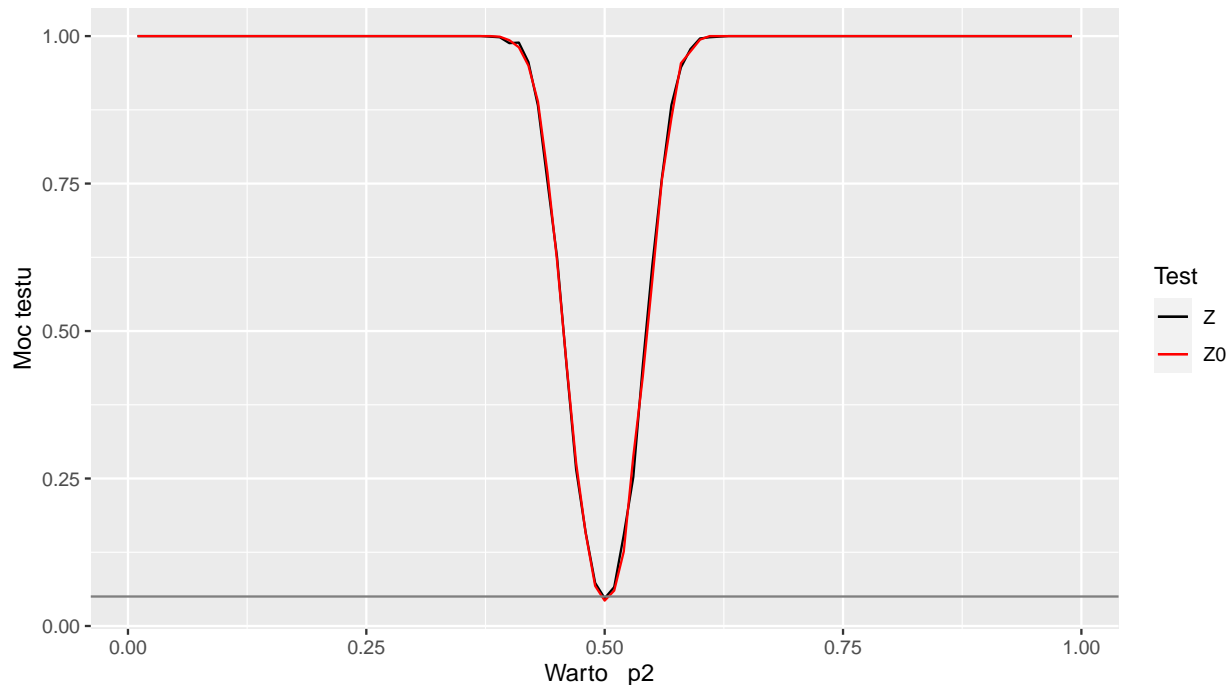


Figure 3: Wykres funkcji mocy testu  $Z$  i  $Z_0$  dla  $n = 100$



#### \* 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 obciążoność 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  $Wyk$ .

- [1 3] 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, fitted(model_a))
```

```
##               model_a
## coefficients    numeric,4      8.86625
## residuals      numeric,24      1.38375
## fitted.values  numeric,24      8.86625
```

```
## effects          numeric,24      1.38375
## R                numeric,16      8.86625
## rank             4                1.38375
## qr               qr,5             8.86625
## family            family,12       1.38375
## linear.predictors numeric,24      30.275
## deviance          203.0698        4.725
## aic               267.822         30.275
## null.deviance     442.1953        4.725
## iter              5                30.275
## weights           numeric,24      4.725
## prior.weights     numeric,24      30.275
## df.residual       20                4.725
## df.null           23                4.10875
## y                 integer,24      0.64125
## converged         TRUE             4.10875
## boundary          FALSE           0.64125
## model             data.frame,3     4.10875
## call              expression      0.64125
## formula           Freq ~ S + Wyk  4.10875
## terms             Freq ~ S + Wyk  0.64125
## data              data.frame,4     8.86625
## offset            NULL            1.38375
## control            list,3          8.86625
## method            "glm.fit"       1.38375
## contrasts          list,2          8.86625
## xlevels            list,2          1.38375
```

P-wartość 0 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 „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(model_b))
```

```
##      S W1 Wyk Freq fitted(model_b)
## 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
```

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(model_c))
```

```
##      S W1 Wyk Freq fitted(model_c)
## 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
```

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(model_d))
```

```
##      S W1 Wyk Freq fitted(model_d)
## 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
```

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(model_e))
```

```
##      S W1 Wyk Freq fitted(model_e)
## 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
```

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(model_f))
```

```
##      S W1 Wyk Freq fitted(model_f)
## 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
```

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

	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

Table 3: P-wartości testów statystycznych

## Zadanie 6

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

Liczności danych	Liczności modeli					
	[1 3]	[13]	[1 2 3]	[12 3]	[12 13]	[1 23]
19	8.86625	10	13.12205	13.12	14.7976879	17.3
1	1.38375	0.25	2.04795	2.05	0.3703704	2.7
3	8.86625	10	3.5465	3.69	4.1618497	2.595
0	1.38375	0.25	0.5535	0.41	0.0740741	0.405
0	8.86625	10	0.35465	$9.067214 \times 10^{-9}$	$4.3841983 \times 10^{-9}$	$9.8822092 \times 10^{-9}$
0	1.38375	0.25	0.05535	0.41	0.0740741	$1.5423101 \times 10^{-9}$

Table 4: Porównanie wyznaczonych liczności na podstawie modeli z rzeczywistymi licznościami danych

- [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, fitted(model_a))
```

```
##               model_a
## coefficients    numeric,4      17.7325
## residuals      numeric,12      2.7675
## fitted.values   numeric,12      17.7325
## effects        numeric,12      2.7675
## R              numeric,16      60.55
## rank           4              9.45
## qr             qr,5          60.55
## family         family,12      9.45
## linear.predictors numeric,12    8.2175
## deviance       77.39248       1.2825
## aic            127.7821       8.2175
## null.deviance  316.5179       1.2825
## iter          5              17.7325
## weights       numeric,12      2.7675
## prior.weights numeric,12      17.7325
## df.residual    8              2.7675
## df.null       11              60.55
## y             integer,12      9.45
## converged     TRUE            60.55
## boundary     FALSE           9.45
## model        data.frame,3     8.2175
## call         expression      1.2825
## formula      Freq ~ S + Wyk  8.2175
## terms        Freq ~ S + Wyk  1.2825
## data         data.frame,4     17.7325
## offset       NULL            2.7675
## control      list,3          17.7325
## method       "glm.fit"       2.7675
## contrasts     list,2          60.55
## xlevels      list,2          9.45
```

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(model_b))
```

```
##      S P Wyk Freq fitted(model_b)
## 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
```

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(model_c))
```

```
##      S P Wyk Freq fitted(model_c)
## 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
```

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(model_d))
```

```
##      S P Wyk Freq fitted(model_d)
## 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

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(model_e))
```

##		S	P	Wyk	Freq	fitted(model_e)
##	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

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].

- [12 3] 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(model_f))
```

##		S	P	Wyk	Freq	fitted(model_f)
##	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

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].

	Modele					
	[1 3]	[13]	[1 2 3]	[12 3]	[12 13]	
P-wartość	$1.6342483 \times 10^{-13}$	$9.9507624 \times 10^{-11}$	$1.2979651 \times 10^{-10}$	$4.8342885 \times 10^{-11}$	$3.0176519 \times 10^{-8}$	1.746

Table 5: P-wartości testów statystycznych

## 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] 0.5074047

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

Table 6: Oszacowane prawdopodobieństwa

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.

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

Table 7: Oszacowane prawdopodobieństwa

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.

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

Table 8: Oszacowane prawdopodobieństwa

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

## Zadanie 8

Do zmiennych  $S$ ,  $P$  i  $Wyk$  przyjmiemy 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]))
```

	Prawdopodobieństwo
Wartość empiryczna	0.4722762
Model [13 23]	0.2962963

Table 9: Oszacowane prawdopodobieństwa

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

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

	Prawdopodobieństwo
Wartość empiryczna	0.0243902
Model [13 23]	0.0243902

Table 10: Oszacowane prawdopodobieństwa

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

	Prawdopodobieństwo
Zadanie 7	0.0243902
Zadanie 8	0.0243902

Table 11: Porównanie prawdopodobieństw tego samego problemu uzyskanych w zadaniu 7. i 8.

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.

	Prawdopodobieństwo
Wartość empiryczna	0.3684211
Model [13 23]	0.3684211

Table 12: Oszacowane prawdopodobieństwa

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]

```
##      S W1 Wyk Freq
## 1  0 -2   1   19
## 2  1 -2   1    1
## 3  0 -1   1    3
## 4  1 -1   1    0
## 5  0  1   1    0
## 6  1  1   1    0
## 7  0  2   1   18
## 8  1  2   1    0
## 9  0 -2   2   40
## 10 1 -2   2    5
## 11 0 -1   2   15
## 12 1 -1   2    2
## 13 0  1   2    0
## 14 1  1   2    0
## 15 0  2   2   68
## 16 1  2   2   10
## 17 0 -2   3    5
## 18 1 -2   3    4
## 19 0 -1   3    0
## 20 1 -1   3    0
## 21 0  1   3    0
## 22 1  1   3    2
## 23 0  2   3    5
## 24 1  2   3    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 [12 13]

```
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+S*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 oraz dla drugiej 0.0396896 jest mniejszy niż założony poziom istotności  $\alpha$ , więc odrzucamy hipotezę zerową.