

## Sprawozdanie 3

### Lista 10

#### Zadanie 1

W tabeli 1 zawarte są wyniki (w skali pozytywny, negatywny) z pierwszego i drugiego kolokwium w pewnej grupie studentów. Przyjmując, że poziom trudności zadań na pierwszym i drugim kolokwium był taki sam, na podstawie tych danych, zweryfikować hipotezę, na poziomie istotności 0.05, że studenci byli tak samo przygotowani do obu kolokwiów.

Tablica 1: Dane do zadania 1.

Wynik z kolokwium 2	Wynik z kolokwium 1		Suma
	Negatywny	Pozytywny	
Negatywny	32	44	76
Pozytywny	22	38	60
Suma	54	82	136

```
mcnemar.test(data, correct = FALSE)

##
## McNemar's Chi-squared test
##
## data:  data
## McNemar's chi-squared = 7.3333, df = 1, p-value = 0.006769

zad1 <- function(data) {
  y12 <- data[1,2]
  y21 <- data[2,1]
  z0 <- ( y12 - y21 ) / ( sqrt( y12 + y21 ) )
  p <- 1 - pchisq(z0^2, df = 1)
  return(p)
}
zad1(data)

## [1] 0.006768741
```

Dla danych w Tabeli 1. przeprowadziliśmy test McNemary'ego i test  $Z_0$  i na poziomie istotności 0.05 odrzucamy hipotezę, że studenci byli tak samo przygotowani do obu kolokwiów. Z obu testów otrzymaliśmy takie same p-wartości równe 0.006768741.

## Zadanie 2

W tabeli 2 zawarte są dane dotyczące reakcji po godzinie od przyjęcia dwóch różnych leków przeciwbólowych (powiedzmy A i B) stosowanych w migrenie, zaaplikowanych grupie pacjentów w dwóch różnych atakach bólowych. Na podstawie tych danych, zweryfikować hipotezę, że leki te są jednakowo skuteczne korzystając z testu:

1. McNemary'ego z poprawką na ciągłość,
2. dokładnego (opisanego w sekcji 2.1.3 wykładu 9. do wydruku)

W drugim przypadku, najpierw napisać deklarację funkcji, której wartością będzie wartość poziomego krytycznego (p wartość) w tym warunkowym teście dokładnym.

Tablica 2: Dane do zadania 2.

Reakcja na lek B	Reakcja na lek A		Suma
	Negatywna	Pozytywna	
Negatywna	1	5	6
Pozytywna	2	4	6
Suma	3	9	12

1)

```

zad2 <- function(data, correct = TRUE) {
  y12 <- data[1,2]
  y21 <- data[2,1]
  if (correct) {
    z0 <- (abs( y12 - y21 ) - 1)/sqrt(y12 + y21)
  } else {
    z0 <- ( y12 - y21 ) / ( sqrt( y12 + y21 ) )
  }
  p <- 1 - pchisq(z0^2, df = 1)
  return(p)
}
mcnemar.test(data2,correct = TRUE)

##
## McNemar's Chi-squared test with continuity correction
##
## data:  data2
## McNemar's chi-squared = 0.57143, df = 1, p-value = 0.4497

```

```
zad2(data2)
## [1] 0.4496918
```

Na poziomie istotności 0.05, nie mamy podstaw do odrzucenia hipotezy o tym, że leki są jednakowo skuteczne. Wartość poziomu krytycznego zarówno w teście McNemary z poprawką na ciągłość i teście  $Z_0$  wynosi 0.4496918.

2)

```
zad2b <- function(data) {
  y12 <- data[1,2]
  y21 <- data[2,1]
  if (y12 < (y12+y21)/2) {
    p <- 2*sum(sapply(0:y12,
      function(k) choose(y12+y21,k) * (1/2)^k * (1/2)^(y12+y21-k)))
  } else if (y12 > (y12+y21)/2) {
    p <- 2*sum(sapply(y12:(y12+y21),
      function(k) choose(y12+y21,k) * (1/2)^k * (1/2)^(y12+y21-k)))
  } else {
    p <- 1
  }
  return(p)
}

zad2b(data2)
## [1] 0.453125

mcnemar.exact(data2)$p.value
## [1] 0.453125
```

W 2. podpunkcie przeprowadziliśmy warunkowy test dokładny, wynik porównaliśmy z wbudowanym testem McNemary z biblioteki **exact2x2**. Jak możemy zauważyć, p-wartości są sobie równe w tych dwóch metodach i wynoszą 0.453125, co dla poziomu istotności 0.05 nie daje nam podstaw do odrzucenia hipotezy o tym, że leki są jednakowo skuteczne.

### Zadanie 3

Przeprowadzić symulacje, w celu porównania mocy testu  $Z$  (opisanego w sekcji 2.1.1) i testu  $Z_0$  (opisanego w sekcji 2.1.2). Wyniki przedstawić w tabeli lub/i na wykresach i napisać odpowiednie wnioski.

```
n <- 10
mcs <- 1000
alpha <- 0.05
ps_1 <- numeric(mcs)
ps_2 <- numeric(mcs)
p2 <- seq(0.01,0.99,0.02)
p2_0 <- 0.5
```

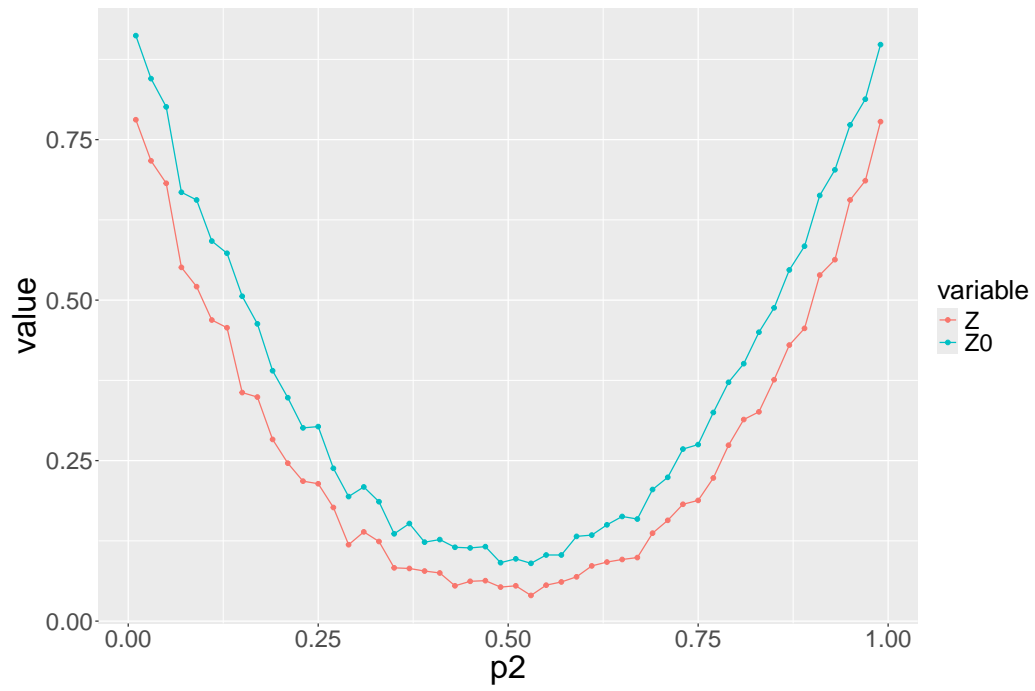
```
f_to_sup <- function(p) {
  p2 <- p
  for (mc in 1:mcs) {
    X <- rbinom(n,1, p = 0.5)
    Y <- rbinom(n, 1, p = p2)
    data <- matrix(0, nrow = 2, ncol = 2)
    for (i in 1:n) {
      x<-X[i]
      y<-Y[i]
      if (x==0 & y==0) {
        data[1,1] <- data[1,1] + 1
      } else if (x==0 & y==1){
        data[1,2] <- data[1,2] + 1
      }
      else if (x==1 & y==0){
        data[2,1] <- data[2,1] + 1
      } else{
        data[2,2] <- data[2,2] + 1
      }
    }
    y12 <- data[1,2]
    y21 <- data[2,1]
    p_data <- data/n
    p11 <- p_data[1,1]
    p12 <- p_data[1,2]
    p21 <- p_data[2,1]
    p22 <- p_data[2,2]

    p_.1 <- (data[1,1]+data[2,1])/n
    p_1. <- (data[1,1]+data[1,2])/n
    D <- p_1. - p_.1
    sig <- sqrt(
      ( p_1.*(1-p_1.) + p_.1*(1-p_.1) - 2*(p11*p22-p12*p21) )/n)
    z <- D/sig
```

```

    p_value1 <- 2*(1 - pnorm(abs(z)))
    z0 <- ( y12 - y21 ) / ( sqrt( y12 + y21 ) )
    p_value2 <- 2*(1 - pnorm(abs(z0)))
    ps_1[mc] <- p_value1 < 0.05
    ps_2[mc] <- p_value2 < 0.05
  }
  return(c( sum(ps_1, na.rm = TRUE)/mcs,
           sum(ps_2, na.rm = TRUE)/mcs))
}
f_to_sup_error_rm <- function(x){
  tryCatch(
    expr = {
      return(f_to_sup(x))
    },
    error = function(e){
    },
    warning = function(w){
      f_to_sup(x)
    },
    finally = {
    }
  )
}
sym <- sapply(p2, f_to_sup_error_rm)
df <- data.frame(p2 = p2, Z0 = sym[1,], Z = sym[2,] )
melt_Df <- melt(df, id=1, measure=c("Z", "Z0"))

```



Rysunek 1. Porównanie mocy testów

## Wnioski do Zadania 3

nie wiem co tu

## Lista 11

### Zadanie 1

W tabeli 3 (znajduje się poniżej) zawarte są wyniki (w skali 2, 3, +3, 4, +4, 5) z pierwszego i drugiego kolokwium w pewnej grupie studentów. Korzystając z odpowiedniego testu, na poziomie istotności  $\alpha = 0.05$ , zweryfikować hipotezę, że dane w tabeli 3 podlegają modelowi:

1. symetrii,
2. quasi-symetrii
3. quasi-niezależności.

Zwrócić uwagę na problem z zastosowaniem do analizowanych danych testu Bowkera

#### 0.0.1. Testowanie symetrii za pomocą testu ilorazu wiarygodności

```
symmetry <- glm(count ~ Symm(kol1, kol2), data=dane33,
                 family = poisson)
x = symmetry$deviance #wartość statystyki  $G^2$ 
r=15 #Liczba stopni swobody
1-pchisq(x,r) #p-wartość
## [1] 0.1004656
```

Do przetestowania symetrii dla danych w Tabeli 1. wykorzystaliśmy test ilorazu wiarygodności (IW), aby to zrobić musieliśmy przekształcić dane do postaci ramki danych. Następnie za pomocą funkcji **glm** z biblioteki **gnm** przeprowadziliśmy test symetrii. P-wartość owego testu wyznaczyliśmy za pomocą wzoru : **1-pchisq(x,r)**, gdzie **pchisq** jest dystrybuantą rozkładu  $\chi^2$ , **x** to wartość statystyki  $G^2$  z testu, a **r** to ilość stopni swobody.

Na poziomie istotności 0.05 nie mamy podstaw do odrzucenia hipotezy o symetrii. Wartość poziomu krytycznego wynosi 0.1004656.

#### 0.0.2. Testowanie quasi-symetrii

Hipotezę zerową, że dane podlegają modelowi quasi-symetrii również zweryfikujemy korzystając z funkcji **glm**.

```
quasi.symm <- glm(count ~ kol1+kol2 + Symm(kol1,kol2),
                  data=dane33, family =poisson)
x = quasi.symm$deviance #wartość statystyki  $G^2$ 
r=10 #Liczba stopni swobody
1-pchisq(x,r) #p-wartość
## [1] 0.9589187
```

Korzystając z testu ilorazu wiarygodności (IW), na poziomie istotności 0.05, nie ma podstaw do odrzucenia hipotezy o quasi-symetrii. Wartość poziomu krytycznego w teście wynosi 0.9589187.

### 0.0.3. Testowanie quasi-niezależności

Hipotezę zerową, że dane podlegają modelowi quasi niezależności również zweryfikujemy korzystając z funkcji **glm**.

```
quasi.indep <- glm(count ~ kol1 + kol2 + Diag(kol1, kol2),
                   data=dane33, family = poisson)
x=quasi.indep$deviance #wartość statystyki  $G^2$ 
r = 19 #Liczba stopni swobody
1-pchisq(x,r) #p-wartość
## [1] 0.00962481
```

Korzystając, z testu ilorazu wiarygodności (IW), na poziomie istotności 0.05, hipotezę o quasi-niezależności należy odrzucić. Wartość poziomu krytycznego w tym teście wynosi 0.00962481.

## Zadanie 2

W tabeli 3 zawarte są wyniki (w skali 2, 3, +3, 4, +4, 5) z pierwszego i drugiego kolokwium w pewnej grupie studentów. Przyjmując, że poziom trudności zadań na pierwszym i drugim kolokwium był taki sam, na podstawie tych danych, zweryfikować hipotezę, na poziomie istotności 0.05, że studenci byli tak samo przygotowani do obu kolokwiów.

Tablica 3: Dane do zadania 1 i 2.

Wyniki z kolokwium 2	Wyniki z kolokwium 1						Suma
	2	3	+3	4	+4	5	
2	5	2	1	0	0	0	8
3	6	3	2	2	0	0	13
+3	1	4	5	5	2	2	19
4	0	10	15	18	5	2	50
+4	1	2	5	3	2	2	15
5	0	1	3	4	3	2	13
Suma	13	22	31	32	12	8	118

```
comparison <- anova(symmetry, quasi.symm)
p <- 1-pchisq(comparison$Deviance[2], comparison$Df[2])
p
## [1] 0.002316914
```

Do weryfikacji hipotezy o tym, że studenci byli tak samo przygotowani do obu kolokwiów wykorzystamy funkcję **anova**, a dokładnie test **IW**, który



jest zaimplementowany w tą funkcję.

Korzystając z tego testu, na poziomie istotności 0.05, należy odrzucić hipotezę o symetrii, czyli że studenci byli tak samo przygotowani na oba kolokwia, pod warunkiem, że zachodzi quasi-symetria. Wartość poziomu krytycznego w tym teście wynosi .002316914.

## Lista 12,13 i 14

Wszystkie poniższe zadania należy wykonać w oparciu o dane w pliku Ankieta.csv, które zawierają, a wyniki ankietowania 40 losowo wybranych studentów PWr. Ankieta zawierała trzy pytania, które dotyczyły jakości snu (odpowiedź 1 oznaczała, że student sypia dobrze, 0, że źle), czy regularnie biega (1 – tak, 0 – nie) oraz czy posiada psa (1 – tak, 0 – nie).

### Zadanie 1

W przypadku powyższych danych, podać interpretację następujących modeli logliniowych:

■ [1 3],

$$l_{ij} = \lambda + \lambda_i^{(1)} + \lambda_j^{(3)}, \forall i \in \{1, \dots, R\} \text{ i } j \in \{1, \dots, C\},$$

Zmienne  $W_1$  i  $W_3$  mają dowolne rozkłady oraz zmienne te są niezależne.

■ [13],

$$l_{ij} = \lambda + \lambda_i^{(1)} + \lambda_j^{(3)} + \lambda_{ij}^{(13)}, \forall i \in \{1, \dots, R\} \text{ i } j \in \{1, \dots, C\},$$

Zmienne  $W_1$  i  $W_3$  mają dowolne rozkłady oraz zmienne te nie są niezależne.

■ [1 2 3],

$$l_{ijk} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)}, \forall i \in \{1, \dots, R\} \text{ i } j \in \{1, \dots, C\} \text{ i } k \in \{1, \dots, L\},$$

Zmienne  $W_1$  i  $W_2$  i  $W_3$  są wzajemnie niezależne.

■ [12 3],

$$l_{ijk} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \lambda_{ij}^{(12)},$$

$$\forall i \in \{1, \dots, R\} \text{ i } j \in \{1, \dots, C\} \text{ i } k \in \{1, \dots, L\},$$

Zmienna  $W_3$  jest niezależna od zmiennej  $W_1$  i  $W_2$ , ale zmienne  $W_1$  i  $W_2$  nie są niezależne.

■ [12 13],

$$l_{ijk} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \lambda_{ij}^{(12)} + \lambda_{ik}^{(13)},$$

$$\forall i \in \{1, \dots, R\} \text{ i } j \in \{1, \dots, C\} \text{ i } k \in \{1, \dots, L\},$$

Przy ustalonej wartości zmiennej  $W_1$ , zmienne  $W_2$  i  $W_3$  są niezależne.

Mówimy wówczas, że zmienne  $W_2$  i  $W_3$  są warunkowo niezależne.

■ [1 23],

$$l_{ijk} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \lambda_{jk}^{(23)},$$

$$\forall i \in \{1, \dots, R\} \text{ i } j \in \{1, \dots, C\} \text{ i } k \in \{1, \dots, L\},$$

Zmienna  $W_1$  jest niezależna od zmiennej  $W_2$  i  $W_3$ , ale zmienne  $W_2$  i  $W_3$  nie są niezależne.

## Zadanie 2

Przyjmując model log-liniowy [12 3], na podstawie danych Ankieta.csv, oszacować prawdopodobieństwo:

1. dobrej jakości snu studenta, który regularnie biega,
2. tego, że student biega regularnie, gdy posiada psa.

Jakie byłyby oszacowania powyższych prawdopodobieństw przy założeniu modelu [12 23]?

```
dane <- read.csv("data/Ankieta.csv", sep = ";")
sen <- dane[,1]
bieganie <- dane[,2]
pies <- dane[,3]
ankieta <- table(sen, bieganie, pies)
ankieta.df <- as.data.frame(as.table(ankieta))
```

## Model [12 3]

```
mod1 <- glm(Freq ~ sen + bieganie + pies + sen*bieganie,
            data = ankieta.df, family = poisson)
diff <- cbind(mod1$data, fitted(mod1))
diff

##   sen bieganie pies Freq fitted(mod1)
## 1   0         0   0    6      3.400
## 2   1         0   0    5      4.250
## 3   0         1   0    1      1.275
## 4   1         1   0    5      8.075
## 5   0         0   1    2      4.600
## 6   1         0   1    5      5.750
## 7   0         1   1    2      1.725
## 8   1         1   1   14     10.925
```

## Pytanie 1.

```
La_mod1_fitt<-sum(
  diff[diff["sen"] == 1&diff["bieganie"] ==1,]["fitted(mod1)"])
Ma_mod1_fitt<-sum(diff[diff["bieganie"] == 1,]["fitted(mod1)"])
pa_mod1_fitt<-La_mod1_fitt/Ma_mod1_fitt
pa_mod1_fitt

## [1] 0.8636364

La_mod1_date <- sum(
  diff[diff["sen"] == 1&diff["bieganie"] == 1,]["Freq"])
Ma_mod1_date <- sum(diff[diff["bieganie"] == 1,]["Freq"])
pa_mod1_date <- La_mod1_date/Ma_mod1_date
pa_mod1_date
```

```
## [1] 0.8636364
```

## Pytanie 2.

```
Lb_mod1_fitt<-sum(
  diff[diff["bieganie"] ==1&diff["pies"] == 1 ,]["fitted(mod1)"])
Mb_mod1_fitt<-sum(diff[diff["pies"] ==1,]["fitted(mod1)"])
pb_mod1_fitt<-Lb_mod1_fitt/Mb_mod1_fitt
pb_mod1_fitt
## [1] 0.55

Lb_mod1_date <- sum(
  diff[diff["bieganie"] ==1&diff["pies"] == 1 ,]["Freq"])
Mb_mod1_date <- sum(diff[diff["pies"] ==1,]["Freq"])
pb_mod1_date <- Lb_mod1_date/Mb_mod1_date
pb_mod1_date
## [1] 0.6956522
```

## Model [12 23]

```
mod2 <- glm(Freq ~ sen + bieganie + pies + sen*bieganie + bieganie*pies,
  data = ankieta.df, family = poisson)
diff <- cbind(mod2$data, fitted(mod2))
diff
##      sen bieganie pies Freq fitted(mod2)
## 1     0         0    0    6   4.8888889
## 2     1         0    0    5   6.1111111
## 3     0         1    0    1   0.8181818
## 4     1         1    0    5   5.1818182
## 5     0         0    1    2   3.1111111
## 6     1         0    1    5   3.8888889
## 7     0         1    1    2   2.1818182
## 8     1         1    1   14  13.8181818
```

## Pytanie 1.

```
La_mod2_fitt<-sum(diff[diff["sen"] == 1&diff["bieganie"] ==1,]["fitted(mod2)"])
Ma_mod2_fitt<-sum(diff[diff["bieganie"] == 1,]["fitted(mod2)"])
pa_mod2_fitt<-La_mod2_fitt/Ma_mod2_fitt
pa_mod2_fitt
## [1] 0.8636364

La_mod2_date <- sum(diff[diff["sen"] == 1&diff["bieganie"] == 1,]["Freq"])
Ma_mod2_date <- sum(diff[diff["bieganie"] == 1,]["Freq"])
```

```
pa_mod2_date <- La_mod2_date/Ma_mod2_date
pa_mod2_date
## [1] 0.8636364
```

## Pytanie 2.

```
Lb_mod2_fitt<-sum(diff[diff["bieganie"] ==1&diff["pies"] == 1 ,]["fitted(mod2)"])
Mb_mod2_fitt<-sum(diff[diff["pies"] ==1,]["fitted(mod2)"])
pb_mod2_fitt<-Lb_mod2_fitt/Mb_mod2_fitt
pb_mod2_fitt
## [1] 0.6956522

Lb_mod2_date <- sum(diff[diff["bieganie"] ==1&diff["pies"] == 1 ,]["Freq"])
Mb_mod2_date <- sum(diff[diff["pies"] ==1,]["Freq"])
pb_mod2_date <- Lb_mod2_date/Mb_mod2_date
pb_mod2_date
## [1] 0.6956522
```

## Zadanie 3

Na podstawie danych Reakcja3.csv zweryfikować następujące hipotezy:

1. zmienne losowe Sen, Bieganie i Pies są wzajemnie niezależne,
2. zmienna losowa Pies jest niezależna od pary zmiennych Sen i Bieganie,
3. zmienna losowa Sen jest niezależna od zmiennej Pies, przy ustalonej zmiennej Bieganie.

## Pytanie 1.

```
mod1 <- glm(Freq ~ sen + bieganie + pies,
            data = ankieta.df, family = poisson)
mod2 <- glm(Freq ~ (sen + bieganie + pies)^2,
            data = ankieta.df, family = poisson)
mod3 <- glm(Freq ~ (sen + bieganie + pies)^3,
            data = ankieta.df, family = poisson)
```

```
test1 <- anova(mod1,mod2)
1-pchisq(test1$Deviance[2], df = test1$Df[2])
## [1] 0.01438801

test2 <- anova(mod1,mod3)
1-pchisq(test2$Deviance[2], df = test2$Df[2])
## [1] 0.02932791
```

## Pytanie 2.

```
mod1 <- glm(Freq ~ sen + bieganie + pies + sen*bieganie,  
            data = ankieta.df, family = poisson)  
mod2 <- glm(Freq ~ (sen + bieganie + pies)^2,  
            data = ankieta.df, family = poisson)  
mod3 <- glm(Freq ~ (sen + bieganie + pies)^3,  
            data = ankieta.df, family = poisson)
```

```
«>= test1 <- anova(mod1,mod2) 1-pchisq(test1Deviance[2],df = test1Df[2])  
test2 <- anova(mod1,mod3) 1-pchisq(test2Deviance[2],df = test2Df[2]) @
```

## Pytanie 3.

```
mod1 <- glm(  
  Freq ~ sen + bieganie + pies + sen*bieganie + pies*bieganie,  
  data = ankieta.df, family = poisson)  
mod2 <- glm(Freq ~ (sen + bieganie + pies)^2,  
            data = ankieta.df, family = poisson) #model niepełny  
mod3 <- glm(Freq ~ (sen + bieganie + pies)^3,  
            data = ankieta.df, family = poisson) #model pełny
```

```
mod2 <- glm(Freq ~ (sen + bieganie + pies)^2,  
            data = ankieta.df, family = poisson) #model niepełny  
mod3 <- glm(Freq ~ (sen + bieganie + pies)^3,  
            data = ankieta.df, family = poisson) #model pełny
```

## Zadanie 4

Na podstawie danych Ankieta.csv dokonać wyboru modelu w oparciu o:

1. testy,
2. kryterium AIC,
3. kryterium BIC.

W przypadku, gdy wybrane modele w punktach 1–3 są różne, dokonać ich porównania.

## Wszystkie 19 modeli

```
#123  
mod123 <- glm(Freq ~ sen + bieganie + pies +  
              sen*bieganie +  
              sen*pies +  
              bieganie*pies +  
              sen*bieganie*pies,  
              data = ankieta.df, family = poisson)
```

```

#12 13 23
mod12_13_23 <- glm(Freq ~ sen + bieganie + pies +
  sen*bieganie +
  sen*pies +
  bieganie*pies,
  data = ankieta.df, family = poisson)

#12 13
mod12_13 <- glm(Freq ~ sen + bieganie + pies +
  sen*bieganie +
  sen*pies,
  data = ankieta.df, family = poisson)

#12 23
mod12_23 <- glm(Freq ~ sen + bieganie + pies +
  sen*bieganie +
  bieganie*pies,
  data = ankieta.df, family = poisson)

#13 23
mod13_23 <- glm(Freq ~ sen + bieganie + pies +
  sen*pies +
  bieganie*pies,
  data = ankieta.df, family = poisson)

#12 3
mod12_3 <- glm(Freq ~ sen + bieganie + pies +
  sen*bieganie,
  data = ankieta.df, family = poisson)

#13 2
mod13_2 <- glm(Freq ~ sen + bieganie + pies +
  sen*pies,
  data = ankieta.df, family = poisson)

#23 1
mod23_1 <- glm(Freq ~ sen + bieganie + pies +
  bieganie*pies,
  data = ankieta.df, family = poisson)

#12
mod12 <- glm(Freq ~ sen + bieganie +
  sen*bieganie,
  data = ankieta.df, family = poisson)

#13
mod13 <- glm(Freq ~ sen + pies +
  sen*pies,
  data = ankieta.df, family = poisson)

```

```

#23
mod23 <- glm(Freq ~ bieganie + pies +
              bieganie*pies,
              data = ankieta.df, family = poisson)

#1 2 3
mod1_2_3 <- glm(Freq ~ sen + bieganie + pies,
                 data = ankieta.df, family = poisson)

#1 2
mod1_2 <- glm(Freq ~ sen + bieganie ,
              data = ankieta.df, family = poisson)

#1 3
mod1_3 <- glm(Freq ~ sen + pies,
              data = ankieta.df, family = poisson)

#2 3
mod2_3 <- glm(Freq ~ bieganie + pies,
              data = ankieta.df, family = poisson)

#1
mod1 <- glm(Freq ~ sen,
            data = ankieta.df, family = poisson)

#2
mod2 <- glm(Freq ~ bieganie,
            data = ankieta.df, family = poisson)

#3
mod3 <- glm(Freq ~ pies,
            data = ankieta.df, family = poisson)

#staly
mod0 <- glm(Freq ~ 1,
            data = ankieta.df, family = poisson)

```

## Pytanie 1.

```

test1 <- anova(mod1_2_3, mod13_2)
1-pchisq(test1$Deviance[2], df = test1$Df[2]) # wybieramy mod1_2_3
## [1] 0.09634701

test2 <- anova(mod1_2_3, mod12_3)
1-pchisq(test2$Deviance[2], df = test2$Df[2]) # wybieramy mod12_3
## [1] 0.02850318

test3 <- anova(mod12_3, mod12_13)
1-pchisq(test3$Deviance[2], df = test3$Df[2]) # wybieramy mod12_3

```

```
## [1] 0.09634701
test4 <- anova(mod13_2, mod13_23)
1-pchisq(test4$Deviance[2], df = test4$Df[2]) # wybieramy mod12_3
## [1] 0.02999623
```

## Pytanie 2.

```
AIC(mod123)
## [1] 41.81545
AIC(mod12_13_23)
## [1] 40.02544
AIC(mod12_13)
## [1] 41.01874
AIC(mod12_23)
## [1] 39.07422
AIC(mod13_23)
## [1] 41.1066
AIC(mod12_3)
## [1] 41.78373
AIC(mod13_2)
## [1] 43.81611
AIC(mod23_1)
## [1] 41.8716
AIC(mod12)
## [1] 40.68714
AIC(mod13)
## [1] 42.21678
AIC(mod23)
## [1] 48.26987
AIC(mod1_2_3)
## [1] 44.5811
AIC(mod1_2)
## [1] 43.48451
```



```

AIC(mod1_3)
## [1] 42.98177
AIC(mod2_3)
## [1] 50.97938
AIC(mod1)
## [1] 41.88518
AIC(mod2)
## [1] 49.88278
AIC(mod3)
## [1] 49.38005
AIC(mod0)
## [1] 48.28345

```

Minimum dla modelu [12,23]. Podobnie możemy użyć funkcji *step*:

```

step(mod123)
## Start:  AIC=41.82
## Freq ~ sen + bieganie + pies + sen * bieganie + sen * pies +
##      bieganie * pies + sen * bieganie * pies
##
##              Df Deviance    AIC
## - sen:bieganie:pies  1  0.20999 40.025
## <none>                0.00000 41.815
##
## Step:  AIC=40.03
## Freq ~ sen + bieganie + pies + sen:bieganie + sen:pies + bieganie:pies
##
##              Df Deviance    AIC
## - sen:pies      1  1.2588 39.074
## <none>           0.2100 40.025
## - bieganie:pies  1  3.2033 41.019
## - sen:bieganie   1  3.2912 41.107
##
## Step:  AIC=39.07
## Freq ~ sen + bieganie + pies + sen:bieganie + bieganie:pies
##
##              Df Deviance    AIC
## <none>           1.2588 39.074
## - bieganie:pies  1  5.9683 41.784
## - sen:bieganie   1  6.0561 41.872
##
## Call:  glm(formula = Freq ~ sen + bieganie + pies + sen:bieganie + bieganie:pies

```

```
##      family = poisson, data = ankieta.df)
##
## Coefficients:
##      (Intercept)          sen1      bieganie1      pies1      sen1:bie
##      1.5870          0.2231      -1.7876      -0.4520
## bieganie1:pies1
##      1.4328
##
## Degrees of Freedom: 7 Total (i.e. Null);  2 Residual
## Null Deviance:      20.47
## Residual Deviance: 1.259  AIC: 39.07
```

Otrzymujemy minimum wartości kryterium AIC dla modelu [12,23]

### Pytanie 3.

```
BIC(mod123)
## [1] 42.45098
BIC(mod12_13_23)
## [1] 40.58153
BIC(mod12_13)
## [1] 41.49539
BIC(mod12_23)
## [1] 39.55087
BIC(mod13_23)
## [1] 41.58325
BIC(mod12_3)
## [1] 42.18094
BIC(mod13_2)
## [1] 44.21332
BIC(mod23_1)
## [1] 42.2688
BIC(mod12)
## [1] 41.0049
BIC(mod13)
## [1] 42.53454
BIC(mod23)
```

```
## [1] 48.58763
BIC(mod1_2_3)
## [1] 44.89887
BIC(mod1_2)
## [1] 43.72283
BIC(mod1_3)
## [1] 43.2201
BIC(mod2_3)
## [1] 51.2177
BIC(mod1)
## [1] 42.04406
BIC(mod2)
## [1] 50.04166
BIC(mod3)
## [1] 49.53893
BIC(mod0)
## [1] 48.36289
```

Minimum dla modelu [12,23]. Podobnie możemy użyć funkcji *step* z `criterion = "BIC"` :

```
step(mod123, criterion = "BIC")
## Start:  AIC=41.82
## Freq ~ sen + bieganie + pies + sen * bieganie + sen * pies +
##      bieganie * pies + sen * bieganie * pies
##
##              Df Deviance    AIC
## - sen:bieganie:pies  1  0.20999 40.025
## <none>                0.00000 41.815
##
## Step:  AIC=40.03
## Freq ~ sen + bieganie + pies + sen:bieganie + sen:pies + bieganie:pies
##
##              Df Deviance    AIC
## - sen:pies      1  1.2588 39.074
## <none>           0.2100 40.025
## - bieganie:pies  1  3.2033 41.019
## - sen:bieganie   1  3.2912 41.107
##
```

```
## Step: AIC=39.07
## Freq ~ sen + bieganie + pies + sen:bieganie + bieganie:pies
##
##              Df Deviance    AIC
## <none>              1.2588 39.074
## - bieganie:pies    1    5.9683 41.784
## - sen:bieganie     1    6.0561 41.872
##
## Call: glm(formula = Freq ~ sen + bieganie + pies + sen:bieganie + bieganie:pies,
##            family = poisson, data = ankieta.df)
##
## Coefficients:
##      (Intercept)              sen1      bieganie1      pies1  sen1:bieganie1
##           1.5870           0.2231          -1.7876          -0.4520
## bieganie1:pies1
##           1.4328
##
## Degrees of Freedom: 7 Total (i.e. Null);  2 Residual
## Null Deviance:      20.47
## Residual Deviance: 1.259  AIC: 39.07
```

Otrzymujemy minimum wartości kryterium AIC dla modelu [12,23]  
 Czyli każda z powyżej przedstawionych metod prowadzi nas do tego samego wyniku.