

SPRAWOZDANIE 3

Analiza danych ankietowych

Aleksandra Grzeszczuk
album 255707

Jacek Wszola
album 255718

5 lutego 2022

Spis treści

1	Lista 1	1
1.1	Zadanie 1	1
1.1.1	Test McNemary’ego bez poprawki na ciągłość	1
1.1.2	Test McNemary’ego z poprawką na ciągłość	1
1.2	Test dokładny	2
1.3	Zadania 2 i 3	3
1.3.1	Symetria	3
1.3.2	Quasi-symetria	4
1.3.3	Quasi-niezależność	4
1.3.4	Brzegowa jednorodność	4
2	Lista 2	6
2.1	Zadanie 1	7
2.2	Zadanie 2	13
2.2.1	podpunkt a	14
2.2.2	podpunkt b	14
2.2.3	podpunkt c	14
2.3	Zadanie 3	16
2.3.1	podpunkt a	16
2.3.2	podpunkt b	16
2.3.3	podpunkt c	17
2.4	Zadanie 4	17
2.4.1	Testy	17
2.4.2	Kryterium AIC i BIC	20

1 Lista 1

1.1 Zadanie 1

Na podstawie danych dotyczących testowania dwóch leków, zawartych w poniższej tabeli, prze-testujemy hipotezę o jednakowej skuteczności obu leków (tzn. symetrii rozkładu). Skorzystamy z różnych wariantów testu McNemary'ego.

	N	P
N	1	5
P	2	4

Tabela 1: Reakcja na lek

1.1.1 Test McNemary'ego bez poprawki na ciągłość

W przypadku małej próby (2×2) mamy statystykę testową w postaci

$$Z = \frac{Y_{12} - Y_{21}}{\sqrt{Y_{12} + Y_{21}}}.$$

Ma ona w przybliżeniu rozkład normalny. Napišemy funkcję, która zwraca wartość poziomu krytycznego, po czym porównamy go z wartością, jaką zwraca funkcja `mcnemar.test()`.

```
mcnemar.bez.popr <- function(X){
  Z <- (X[1,2]-X[2,1])/sqrt(X[1,2]+X[2,1])
  p <- 1 - pchisq(Z^2, 1)
  return(p)
}

mcnemar.bez.popr(reakcja)

## [1] 0.2568393

mcnemar.test(reakcja, correct = FALSE)

##
## McNemar's Chi-squared test
##
## data:  reakcja
## McNemar's chi-squared = 1.2857, df = 1, p-value = 0.2568
```

W kosekwencji otrzymujemy te same wartości. Na mocy tego testu, przy p -value równym 0.2568, nie mamy podstaw do odrzucenia hipotezy o symetrii.

1.1.2 Test McNemary'ego z poprawką na ciągłość

W tej wersji testu statystyka testowa jest w postaci

$$Z = \frac{||Y_{12} - Y_{21}| - 1|}{\sqrt{Y_{12} + Y_{21}}}.$$

Postępujemy podobnie, jak w poprzednim przypadku.

```
mcnemar.popr <- function(X){
  Z <- (abs(X[1,2]-X[2,1])-1)^2/(X[1,2]+X[2,1])
  p <- 1 - pchisq(Z, 1)
  return(p)
}

mcnemar.popr(reakcja)

## [1] 0.4496918

mcnemar.test(reakcja, correct = TRUE)

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

Ponownie, otrzymujemy te same wartości i ponownie, przy p -value równym 0.4497, nie mamy podstaw do odrzucenia hipotezy o symetrii.

1.2 Test dokładny

W przypadku, gdy n jest małe, do weryfikacji hipotezy skorzystamy z testu warunkowego, przy warunku, że $Y_{12} + Y_{21} = n$ i przy założeniu, że rozkład warunkowy $Y_{12}|Y_{12} + Y_{21} = n$ to rozkład dwumianowy $\mathcal{B}(n, 1/2)$. W praktyce, wartość p -value liczymy następująco:

$$p = \begin{cases} 2 \cdot \sum_{k=0}^{Y_{12}} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} & \text{gdy } Y_{12} < \frac{n}{2} \\ 2 \cdot \sum_{k=Y_{12}}^n \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} & \text{gdy } Y_{12} > \frac{n}{2} \\ 1 & \text{gdy } Y_{12} = \frac{n}{2} \end{cases}$$

Tę funkcję zaimplementujemy poniżej:

```
exact <- function(X){
  n <- X[1,2] + X[2,1]
  if(X[1,2] < n/2){
    k <- seq(0, X[1,2], 1)
    Y <- choose(n, k)*(1/2)^n
    p <- 2*sum(Y)
  }
  if(X[1,2] > n/2){
    k <- seq(X[1,2], n, 1)
    Y <- choose(n, k)*(1/2)^n
    p <- 2*sum(Y)
  }
  else{
    p <- 1
  }
}
```

```

    return(p)
}

exact(reakcja)

## [1] 0.453125

```

P-wartość wychodzi zbliżona do tej z poprzedniego podpunktu, jest równa ok. 0.4531. Nie mamy więc podstaw do odrzucenia hipotezy o symetrii.

1.3 Zadania 2 i 3

Zajmiemy się teraz analizą wyników z dwóch kolokwiów. Zbadamy symetrię, quasi-symetrię, quasi-niezależność, a także sprawdzimy na poziomie istotności 0.05, czy studenci byli tak samo przygotowani do obu kolokwiów. Rozkład wyników z obu kolokwiów prezentuje poniższa tabela, gdzie w kolumnach mamy wyniki z kolokwium 1, zaś w wierszach – wyniki z kolokwium 2.

	2	3	3+	4	4+	5
2	4	3	1	0	0	0
3	6	3	2	2	0	0
3+	1	4	5	5	2	2
4	0	10	15	18	5	2
4+	1	2	5	3	2	2
5	0	1	3	4	3	2

Tabela 2: Wyniki z kolokwium

1.3.1 Symetria

W przypadku powyższego zestawu danych nie skorzystamy z testu Browkera, ponieważ przy wykonywaniu algorytmu w dwóch miejscach w mianowniku pojawi się zero, a test zwróci nieokreśloną wartość. Rzeczywiście,

```

mcnemar.test(wyniki)

##
## McNemar's Chi-squared test
##
## data: wyniki
## McNemar's chi-squared = NaN, df = 15, p-value = NA

```

Skorzystamy więc z testu ilorazu wiarygodności, zaimplementowanego w funkcji `glm()` z pakietu `gnm`. Odpowiednie p -value liczymy jako $1 - F(g^2)$, gdzie g^2 jest wartością statystyki testowej, zaś F jest dystrybuantą rozkładu χ^2 z $N(N-2)/2$ stopniami swobody. U nas $N = 6$, ponieważ rozważamy tabelę 6×6 .

```

count <- c(4, 3, 1, 0, 0, 0,
           6, 3, 2, 2, 0, 0,
           1, 4, 5, 5, 2, 2,

```

```

      0, 10, 15, 18, 5, 2,
      1, 2, 5, 3, 2, 2,
      0, 1, 3, 4, 3, 2)
kol2 <- gl(6,6, labels = c("2", "3", "3+", "4", "4+", "5"))
kol1 <- gl(6,1, labels = c("2", "3", "3+", "4", "4+", "5"))
wyniki.dane <- data.frame(kol1, kol2, count)
symm <- glm(count ~ Symm(kol1, kol2), data = wyniki.dane, family = poisson)

1 - pchisq(symm$deviance, 6*(6-1)/2)

## [1] 0.1301103

```

Przy p -value równym 0.1301 nie mamy podstaw do odrzucenia hipotezy o symetrii.

1.3.2 Quasi-symetria

Ponownie skorzystamy z testu ilorazu wiarygodności, lecz teraz rozkład χ^2 będzie z $(N-1)(N-2)/2$ stopniami swobody.

```

quasi.symm <- glm(count ~ kol1 + kol2 + Symm(kol1, kol2),
                  data = wyniki.dane, family = poisson)
1 - pchisq(quasi.symm$deviance, (6-1)*(6-2)/2)

## [1] 0.970888

```

Ponieważ p -value jest w przybliżeniu równe 0.9708, ponownie nie mamy podstaw do odrzucenia hipotezy.

1.3.3 Quasi-niezależność

W tym przypadku weźmiemy rozkład χ^2 z $(N-1)^2 - N$ stopniami swobody.

```

quasi.indep <- glm(count ~ kol1 + kol2 + Diag(kol1, kol2),
                  data = wyniki.dane, family = poisson)
1 - pchisq(quasi.indep$deviance, (6-1)^2-6)

## [1] 0.005012503

```

Tutaj obserwujemy z kolei, że p -value jest stosunkowo małe i wynosi 0.005, więc hipotezę o quasi-niezależności odrzucimy z prawdopodobieństwem 1.

1.3.4 Brzegowa jednorodność

Chcąc sprawdzić, poziom przygotowania studentów do obu kolokwiów był równy, wystarczy badać brzegową jednorodność rozkładu. W przypadku, gdy zachodzi quasi-symetria, testowanie będzie sprowadzać się do testowania symetrii. Skorzystamy z testu ilorazu wiarygodności zaimplementowanego w funkcji `anova()`.

```
comparison <- anova(symm, quasi.symm)
comparison

## Analysis of Deviance Table
##
## Model 1: count ~ Symm(kol1, kol2)
## Model 2: count ~ kol1 + kol2 + Symm(kol1, kol2)
##   Resid. Df Resid. Dev Df Deviance
## 1         15    21.2146
## 2         10     3.3841  5    17.831

1 - pchisq(17.831, 5)

## [1] 0.003165839
```

Mała wartość p -value (0.0032) pokazuje, że hipotezę o jednorodności możemy odrzucić z prawdopodobieństwem 1. W szczególności, oznacza to nierówny poziom przygotowania studentów do obu kolokwiów. Nie możemy jednak wyciągać pochopnych wniosków. Zdarzenie, że jedno kolokwium było znacznie trudniejsze niż drugie, ma przecież dodatnie prawdopodobieństwo. Niestety, tej hipotezy nie zweryfikujemy już na podstawie dostarczonych danych.

2 Lista 2

Wszystkie poniższe zadania będziemy wykonywać w oparciu o dane w pliku `Reakcja3.csv`, które zawierają informacje o reakcji na lek (zmienna *Reakcja* na poziomie 0, gdy nie nastąpiła poprawa i na poziomie 1, gdy nastąpiła poprawa), dwóch firm farmaceutycznych (zmienna *Rodzaj*), pacjentów leczonych w domu (0) lub w szpitalu (1) - zmienna *Miejsce*.

Najpierw wczytamy dane, zamienamy nasze typy danych z jakościowych na ilościowe oraz, w celu ułatwienia dalszych działań, zamienimy nazwy zmiennych - nie będzie trzeba tworzyć byt dużych legend do wykresów.

```
dane <- mutate(dane, Reakcja = as.factor(Reakcja),
               Rodzaj = as.factor(Rodzaj),
               Miejsce = as.factor(Miejsce),
               Reakcja = fct_recode(Reakcja, "Nie ma poprawy" = "0",
                                     "Jest poprawa" = "1"),
               Miejsce = fct_recode(Miejsce, "Dom" = "0", "Szpital" = "1"))
```

Korzystając z funkcji `flat-table` tworzącą tabelę płaską zamieniamy nasze dane właśnie do takiej tabeli. Wyorzystując funkcję `as.data.frame(as.table)` przekształcimy nasze dane do nowej tabeli, w której zostaną podane częstości występowania danych zjawisk - `Freq`.

```
dane_nowe <- flat_table(dane)
addmargins(dane_nowe)
```

##			Sum
##	45	31	76
##	41	30	71
##	5	19	24
##	9	20	29
##	Sum	100	100
			200

```
prop.table(dane_nowe)
```

##			Miejsce	Dom	Szpital
##	Reakcja	Rodzaj			
##	Nie ma poprawy	0		0.225	0.155
##		1		0.205	0.150
##	Jest poprawa	0		0.025	0.095
##		1		0.045	0.100

```
dane.df <- as.data.frame(as.table(dane_nowe))
dane.df
```

##		Reakcja	Rodzaj	Miejsce	Freq
##	1	Nie ma poprawy	0	Dom	45
##	2	Jest poprawa	0	Dom	5
##	3	Nie ma poprawy	1	Dom	41
##	4	Jest poprawa	1	Dom	9

```
## 5 Nie ma poprawy      0 Szpital  31
## 6 Jest poprawa        0 Szpital  19
## 7 Nie ma poprawy      1 Szpital  30
## 8 Jest poprawa        1 Szpital  20
```

2.1 Zadanie 1

Funkcja `glm` z biblioteki `stats` służy do dopasowania uogólnionych modeli liniowych, określonych przez podanie symbolicznego opisu predyktora liniowego i opisu rozkładu błędu. Wykorzystamy tą funkcję do interpretacji modeli log-liniowych.

Zadanie pierwsze dotyczy interpretacji modeli log-liniowych hierarchicznie uporządkowanych. Ustalmy, że zmienna *Reakcja* będzie zmienną W_1 , zmienna *Rodzaj* - W_2 oraz zmienna *Miejsce* - W_3 . Na podstawie powyższych danych mamy wyjaśnić znaczenie następujących modeli log-liniowych.

- [1 3] - oznacza, że zmienne W_1 oraz W_3 mają dowolne rozkłady oraz zmienne te są niezależne.

Czyli w naszym przypadku otrzymujemy, że zmienne *Reakcja* oraz *Miejsce* mają dowolne rozkłady, zaś zmienna *Rodzaj* ma rozkład równomierny.

```
model1 <- glm(Freq ~ Reakcja + Miejsce,
              data = dane.df, family = poisson)
summary(model1)$coefficients

##              Estimate Std. Error      z value      Pr(>|z|)
## (Intercept)    3.604138e+00  0.1086399  3.317508e+01 2.463575e-241
## ReakcjaJest poprawa -1.020141e+00  0.1602182 -6.367197e+00 1.925144e-10
## MiejsceSzpital   -5.879216e-11  0.1414206 -4.157257e-10 1.000000e+00
```

Wyznaczamy jeszcze p - *value* dla powyższego testu.

```
1 - pchisq(deviance(model1), df = df.residual(model1))

## [1] 0.003030787
```

Powyższa wartość jest mniejsza niż poziom istotności $\alpha = 0.5$, zatem uznajemy, że model nie jest dobry. To znaczy, że reakcja na lek zależy od tego, gdzie byliśmy leczeni.

Porównujemy licznosci z modelem:

```
cbind(model1$data, fitted(model1))

##      Reakcja Rodzaj Miejsce Freq fitted(model1)
## 1 Nie ma poprawy      0      Dom    45      36.75
## 2 Jest poprawa      0      Dom     5      13.25
```


## 3	Nie ma poprawy	1	Dom	41	36.75
## 4	Jest poprawa	1	Dom	9	13.25
## 5	Nie ma poprawy	0	Szpital	31	36.75
## 6	Jest poprawa	0	Szpital	19	13.25
## 7	Nie ma poprawy	1	Szpital	30	36.75
## 8	Jest poprawa	1	Szpital	20	13.25

W powyższej tabelce również widać, że wartości przyporządkowane przez model różnią się od tych właściwych, co potwierdza naszą hipotezę.

- [13] - oznacza, że zmienne W_1 oraz W_3 mają dowolne rozkłady oraz zmienne te nie są niezależne.

Czyli w naszym przypadku otrzymujemy, że zmienne *Reakcja* oraz *Miejsce* mają dowolne rozkłady oraz są zależne.

```
model2 <- glm(Freq ~ Reakcja + Miejsce +
               Reakcja*Miejsce,
               data = dane.df, family = poisson)
summary(model2)$coefficients
```

##		Estimate	Std. Error	z value
##	(Intercept)	3.7612001	0.1078328	34.879935
##	ReakcjaJest poprawa	-1.8152900	0.2881952	-6.298822
##	MiejsceSzpital	-0.3434734	0.1673958	-2.051864
##	ReakcjaJest poprawa:MiejsceSzpital	1.3679777	0.3536820	3.867819
##				Pr(> z)
##	(Intercept)	1.498074e-266		
##	ReakcjaJest poprawa	2.999169e-10		
##	MiejsceSzpital	4.018286e-02		
##	ReakcjaJest poprawa:MiejsceSzpital	1.098132e-04		

Wyznaczamy jeszcze p – value dla powyższego testu.

```
1 - pchisq(deviance(model2), df = df.residual(model2))

## [1] 0.8464359
```

Powyższa wartość jest większa niż poziom istotności $\alpha = 0.5$, zatem uznajemy, że model jest dobry. To znaczy, że reakcja na lek zależy od miejsca, w którym lek został podany - czy w domu, czy w szpitalu.

Porównujemy licznosci z modelem:

```
cbind(model2$data, fitted(model2))

##          Reakcja Rodzaj Miejsce Freq fitted(model2)
```

## 1	Nie ma poprawy	0	Dom	45	43.0
## 2	Jest poprawa	0	Dom	5	7.0
## 3	Nie ma poprawy	1	Dom	41	43.0
## 4	Jest poprawa	1	Dom	9	7.0
## 5	Nie ma poprawy	0	Szpital	31	30.5
## 6	Jest poprawa	0	Szpital	19	19.5
## 7	Nie ma poprawy	1	Szpital	30	30.5
## 8	Jest poprawa	1	Szpital	20	19.5

Widzimy, że w powyższej tabelce wartości dopasowane przez model nie różnią się zbytnio od tych właściwych, co oznacza, że nasz model jest poprawny.

- [1 2 3] - oznacza, że zmienne W_1 , W_2 oraz W_3 są wzajemnie niezależne. Czyli w naszym przypadku otrzymujemy, że zmienne *Reakcja*, *Rodzaj* oraz *Miejsce* są od siebie niezależne.

```
model3 <- glm(Freq ~ Reakcja + Rodzaj + Miejsce,
              data = dane.df, family = poisson)
summary(model3)$coefficients
```

##		Estimate	Std. Error	z value	Pr(> z)
##	(Intercept)	3.604138e+00	0.1296246	2.780443e+01	3.834820e-170
##	ReakcjaJest poprawa	-1.020141e+00	0.1602180	-6.367205e+00	1.925037e-10
##	Rodzaj1	-5.052647e-11	0.1414205	-3.572783e-10	1.000000e+00
##	MiejsceSzpital	-7.047677e-11	0.1414205	-4.983491e-10	1.000000e+00

Wyznaczamy jeszcze p – value dla powyższego testu.

```
1 - pchisq(deviance(model3), df = df.residual(model3))

## [1] 0.001271524
```

Powyższa wartość jest mniejsza niż poziom istotności $\alpha = 0.5$, zatem uznajemy, że model nie jest dobry. Z tego wynika, że któreś ze zmiennych *Reakcja*, *Rodzaj* oraz *Miejsce* są od siebie zależne, czyli jedna z nich ma wpływ na drugą.

Porównujemy licznosci z modelem:

```
cbind(model3$data, fitted(model3))
```

##		Reakcja	Rodzaj	Miejsce	Freq	fitted(model3)
## 1	Nie ma poprawy	0	Dom	45		36.75
## 2	Jest poprawa	0	Dom	5		13.25
## 3	Nie ma poprawy	1	Dom	41		36.75
## 4	Jest poprawa	1	Dom	9		13.25
## 5	Nie ma poprawy	0	Szpital	31		36.75

## 6	Jest poprawa	0 Szpital	19	13.25
## 7	Nie ma poprawy	1 Szpital	30	36.75
## 8	Jest poprawa	1 Szpital	20	13.25

Wartości dopasowane przez model widocznie odbiegają od tych właściwych, stąd widzimy, że nasz model nie jest poprawny.

- [12 3] - oznacza, że zmienne W_1 , W_2 nie są niezależne, ale zmienna W_3 jest niezależna od zmiennej W_1 , W_2 .

Czyli w naszym przypadku otrzymujemy, że zmienne *Reakcja* oraz *Rodzaj* są zależne, ale zmienna *Miejsce* jest niezależna i od zmiennej *Reakcja* i od zmiennej *Rodzaj*. To znaczy, że reakcja na lek zależy od jego rodzaju oraz reakcja na lek i jego rodzaj nie zależy od miejsca podania leku.

```
model4 <- glm(Freq ~ Reakcja + Rodzaj + Miejsce +
               Reakcja*Rodzaj,
               data = dane.df, family = poisson)
summary(model4)$coefficients
```

##		Estimate	Std. Error	z value
##	(Intercept)	3.637586e+00	0.1347512	2.699483e+01
##	ReakcjaJest poprawa	-1.152680e+00	0.2341465	-4.922900e+00
##	Rodzaj1	-6.805346e-02	0.1650527	-4.123135e-01
##	MiejsceSzpital	1.225130e-15	0.1414214	8.662979e-15
##	ReakcjaJest poprawa:Rodzaj1	2.572955e-01	0.3215460	8.001825e-01
##		Pr(> z)		
##	(Intercept)	1.699747e-160		
##	ReakcjaJest poprawa	8.527114e-07		
##	Rodzaj1	6.801097e-01		
##	MiejsceSzpital	1.000000e+00		
##	ReakcjaJest poprawa:Rodzaj1	4.236051e-01		

Wyznaczamy jeszcze p – value dla powyższego testu.

```
1 - pchisq(deviance(model4), df = df.residual(model4))

## [1] 0.0006156983
```

Powyższa wartość jest mniejsza niż poziom istotności $\alpha = 0.5$, zatem uznajemy, że model nie jest dobry. To znaczy, że reakcja na lek nie zależy od jego rodzaju.

Porównujemy licznosci z modelem:

```
cbind(model4$data, fitted(model4))

##           Reakcja Rodzaj Miejsce Freq fitted(model4)
```

## 1	Nie ma poprawy	0	Dom	45	38.0
## 2	Jest poprawa	0	Dom	5	12.0
## 3	Nie ma poprawy	1	Dom	41	35.5
## 4	Jest poprawa	1	Dom	9	14.5
## 5	Nie ma poprawy	0	Szpital	31	38.0
## 6	Jest poprawa	0	Szpital	19	12.0
## 7	Nie ma poprawy	1	Szpital	30	35.5
## 8	Jest poprawa	1	Szpital	20	14.5

Wartości dopasowane przez model bardzo widocznie odbiegają od tych właściwych, dlatego stwierdzamy, że nasz model nie jest poprawny.

- [12 13] - oznacza, że przy ustalonej wartości zmiennej W_1 zmienne W_2 oraz W_3 są niezależne. Mówimy więc, że zmienne W_2 oraz W_3 są warunkowo niezależne. Czyli w naszym przypadku oznacza to, że zmienne *Rodzaj* oraz *Miejsce* są warunkowo niezależne. Zależą one od ustalonej wartości zmiennej *Reakcja*.

```
model5 <- glm(Freq ~ Reakcja + Rodzaj + Miejsce +
               Reakcja*Rodzaj + Reakcja*Miejsce,
               data = dane.df, family = poisson)
summary(model5)$coefficients
```

##	Estimate	Std. Error	z value
## (Intercept)	3.79464805	0.1341010	28.2969388
## ReakcjaJest poprawa	-1.94782880	0.3349782	-5.8147928
## Rodzaj1	-0.06805346	0.1650527	-0.4123135
## MiejsceSzpital	-0.34347343	0.1673958	-2.0518643
## ReakcjaJest poprawa:Rodzaj1	0.25729546	0.3215460	0.8001825
## ReakcjaJest poprawa:MiejsceSzpital	1.36797775	0.3536820	3.8678186

```
##
## Pr(>|z|)
## (Intercept) 3.769012e-176
## ReakcjaJest poprawa 6.070905e-09
## Rodzaj1 6.801097e-01
## MiejsceSzpital 4.018286e-02
## ReakcjaJest poprawa:Rodzaj1 4.236051e-01
## ReakcjaJest poprawa:MiejsceSzpital 1.098133e-04
```

Wyznaczamy jeszcze p – value dla powyższego testu.

```
1 - pchisq(deviance(model5), df = df.residual(model5))

## [1] 0.6891516
```

Powyższa wartość jest większa niż poziom istotności $\alpha = 0.5$, zatem uznajemy, że model jest dobry.

Porównujemy licznosci z modelem:

```
cbind(model5$data, fitted(model5))

##           Reakcja Rodzaj Miejsce Freq fitted(model5)
## 1 Nie ma poprawy      0      Dom   45    44.462585
## 2 Jest poprawa        0      Dom    5     6.339623
## 3 Nie ma poprawy      1      Dom   41    41.537415
## 4 Jest poprawa        1      Dom    9     7.660377
## 5 Nie ma poprawy      0 Szpital   31    31.537415
## 6 Jest poprawa        0 Szpital   19    17.660377
## 7 Nie ma poprawy      1 Szpital   30    29.462585
## 8 Jest poprawa        1 Szpital   20    21.339623
```

Wartości przyporządkowane przez model są bardzo zbliżone do tych właściwych co potwierdza naszą hipotezę, że model jest poprawnie dopasowany.

- [1 23] - oznacza, że zmienna W_1 jest niezależna od zmiennej W_2 oraz W_3 , ale zmienne W_2 i W_3 nie są niezależne.

Czyli w naszym przypadku oznacza to, że zmienne *Reakcja* jest niezależna od zmiennych *Rodzaj* i *Miejsce*, ale zmienne *Rodzaj* i *Miejsce* są od siebie zależne.

```
model6 <- glm(Freq ~ Reakcja + Rodzaj + Miejsce +
               Rodzaj*Miejsce,
               data = dane.df, family = poisson)
summary(model6)$coefficients

##              Estimate Std. Error      z value      Pr(>|z|)
## (Intercept)  3.604138e+00  0.1476555  2.440910e+01 1.369160e-131
## ReakcjaJest poprawa -1.020141e+00  0.1602176 -6.367219e+00 1.924865e-10
## Rodzaj1         -4.376804e-10  0.1999979 -2.188424e-09 1.000000e+00
## MiejsceSzpital  -4.433278e-10  0.1999980 -2.216661e-09 1.000000e+00
## Rodzaj1:MiejsceSzpital 4.409790e-10  0.2828406 1.559108e-09 1.000000e+00
```

Wyznaczamy jeszcze p – value dla powyższego testu.

```
1 - pchisq(deviance(model6), df = df.residual(model6))

## [1] 0.0004539394
```

Powyższa wartość jest mniejsza niż poziom istotności $\alpha = 0.5$, zatem uznajemy, że model nie jest dobry. To znaczy, że rodzaj szczepionki nie zależy od miejsca, w którym została podana, oraz, że reakcja na lek zależy od rodzaju bądź miejsca.

Porównujemy licznosci z modelem:

```
cbind(model6$data, fitted(model6))
```

##		Reakcja	Rodzaj	Miejsce	Freq	fitted(model6)
## 1	Nie ma poprawy	0	Dom	45		36.75
## 2	Jest poprawa	0	Dom	5		13.25
## 3	Nie ma poprawy	1	Dom	41		36.75
## 4	Jest poprawa	1	Dom	9		13.25
## 5	Nie ma poprawy	0	Szpital	31		36.75
## 6	Jest poprawa	0	Szpital	19		13.25
## 7	Nie ma poprawy	1	Szpital	30		36.75
## 8	Jest poprawa	1	Szpital	20		13.25

Wartości przyporządkowane przez model znacznie się różnią od tych właściwych, stąd wiemy, że model [1 23] nie jest modelem dobrym.

2.2 Zadanie 2

W zadaniu drugim mamy przyjąć model log-liniowy [13], oznaczający, że zmienne W_1 oraz W_3 mają dowolne rozkłady oraz zmienne te nie są niezależne.

```
model_zad2 <- glm(Freq ~ Reakcja + Rodzaj + Miejsce +
  Reakcja*Miejsce,
  data = dane.df, family = poisson)
```

Wyznaczamy p-wartość.

```
1 - pchisq(deviance(model_zad2), df = df.residual(model_zad2))
```

```
## [1] 0.7085639
```

Oraz porównujemy licznosci z modelem:

```
cbind(model_zad2$data, fitted(model_zad2))
```

##		Reakcja	Rodzaj	Miejsce	Freq	fitted(model_zad2)
## 1	Nie ma poprawy	0	Dom	45		43.0
## 2	Jest poprawa	0	Dom	5		7.0
## 3	Nie ma poprawy	1	Dom	41		43.0
## 4	Jest poprawa	1	Dom	9		7.0
## 5	Nie ma poprawy	0	Szpital	31		30.5
## 6	Jest poprawa	0	Szpital	19		19.5
## 7	Nie ma poprawy	1	Szpital	30		30.5
## 8	Jest poprawa	1	Szpital	20		19.5

Widzimy, że model log-liniowy [13] jest dobrze dopasowany, ponieważ wartości dopasowane przez model są bardzo zbliżone do tych właściwych w modelu.

2.2.1 podpunkt a

Na podstawie danych `Reakcja3.csv` mamy oszacować prawdopodobieństwo pozytywnej reakcji ($Reakcja = 1$) pacjenta leczonego w domu ($Miejsce = 0$). Wykorzystamy w tym celu prawdopodobieństwo warunkowe - podzielimy licznosci pacjentów leczonych w domu, którzy pozytywnie zareagowali na leczenie przez licznosc pacjentów leczonych w domy.

```
zad2_dom <- cbind(model_zad2$data, fitted(model_zad2))

dane_dom_reakcja <- subset(zad2_dom, zad2_dom$Reakcja=="Jest poprawa" &
                           zad2_dom$Miejsce=="Dom")

dane_dom <- subset(zad2_dom, zad2_dom$Miejsce=="Dom")

sum(dane_dom_reakcja$fitted(model_zad2))/sum(dane_dom$fitted(model_zad2))

## [1] 0.14
```

Czyli otrzymaliśmy, że prawdopodobieństwo pozytywnej reakcji dla pacjenta leczonego w domu wynosi zaledwie 0,14.

2.2.2 podpunkt b

Na podstawie danych `Reakcja3.csv` mamy oszacować prawdopodobieństwo pozytywnej reakcji ($Reakcja = 1$) pacjenta leczonego w szpitalu ($Miejsce = 1$).

```
zad2_szpital <- cbind(model_zad2$data, fitted(model_zad2))

dane_szpital_reakcja <- subset(zad2_szpital, zad2_szpital$Reakcja=="Jest poprawa"
                              & zad2_szpital$Miejsce=="Szpital")

dane_szpital <- subset(zad2_szpital, zad2_szpital$Miejsce=="Szpital")

sum(dane_szpital_reakcja$fitted(model_zad2))/sum(dane_szpital$fitted(model_zad2))

## [1] 0.39
```

Prawdopodobieństwo poprawy pacjenta leczonego w szpitalu wynosi 0,39. Jest ono około 2,6 raza większa od prawdopodobieństwa poprawy pacjenta leczonego w domu. To znaczy, że wykwalifikowana opieka medyczna znacznie zwiększa naszą szanse na powrót do zdrowia. Jednakże i tak jest to małe prawdopodobieństwo poprawy.

2.2.3 podpunkt c

Zbadamy również, jakie byłyby oszacowania powyższych prawdopodobieństw przy założeniu modelu log-liniowego [12 13] oznaczającego, że zmienne W_2 oraz W_3 są warunkowo niezależne.

```
model_zadanie2c <- glm(Freq ~ Reakcja + Rodzaj + Miejsce +
                      Reakcja*Rodzaj + Reakcja*Miejsce,
                      data = dane.df, family = poisson)
```

Wyznaczamy p – *warto*.

```
1 - pchisq(deviance(model_zadanie2c), df = df.residual(model_zadanie2c))

## [1] 0.6891516
```

Oraz porównujemy licznosci z modelem:

```
cbind(model_zadanie2c$data, fitted(model_zadanie2c))

##          Reakcja Rodzaj Miejsce Freq fitted(model_zadanie2c)
## 1 Nie ma poprawy      0    Dom    45      44.462585
## 2 Jest poprawa        0    Dom     5       6.339623
## 3 Nie ma poprawy      1    Dom    41      41.537415
## 4 Jest poprawa        1    Dom     9       7.660377
## 5 Nie ma poprawy      0 Szpital  31      31.537415
## 6 Jest poprawa        0 Szpital  19      17.660377
## 7 Nie ma poprawy      1 Szpital  30      29.462585
## 8 Jest poprawa        1 Szpital  20      21.339623
```

Widzimy ponownie, że model [12 13] jest dobrze dopasowany, ponieważ wartości przez niego przyporządkowane nieznacznie odbiegają od wartości właściwych w modelu.

Korzystając z funkcji `subset` obliczamy prawdopodobieństwo poprawy pacjenta leczonego w domu.

```
zad2c_dom <- cbind(model_zadanie2c$data, fitted(model_zadanie2c))

dane_dom_reakcjac <- subset(zad2c_dom, zad2c_dom$Reakcja=="Jest poprawa" &
                             zad2c_dom$Miejsce=="Dom")

dane_domc <- subset(zad2c_dom, zad2c_dom$Miejsce=="Dom")

sum(dane_dom_reakcjac$fitted(model_zadanie2c))/sum(dane_domc$fitted(model_zadanie2c))

## [1] 0.14
```

Czyli prawdopodobieństwo poprawy stanu zdrowia pacjenta leczonego w domu wynosi 0,14. Tyle samo wynosiło to prawdopodobieństwo w modelu [13].

Obliczymy jeszcze prawdopodobieństwo poprawy stanu zdrowia pacjenta leczonego w szpitalu.

```
zad2_szpitalc <- cbind(model_zadanie2c$data, fitted(model_zadanie2c))

dane_szpital_reakcjac <- subset(zad2_szpitalc, zad2_szpitalc$Reakcja=="Jest poprawa" &
                                zad2_szpitalc$Miejsce=="Szpital")

dane_szpitalc <- subset(zad2_szpitalc, zad2_szpitalc$Miejsce=="Szpital")
```



```
sum(dane_szpital_reakcjac$'fitted(model_zadanie2c)')/sum(dane_szpitalc$'fitted(model_zadanie2c)')
## [1] 0.39
```

I ponownie widzimy, że wyszło nam takie same prawdopodobieństwo jak w modelu [13].

2.3 Zadanie 3

Na podstawie danych *Reakcja3.csv* zweryfikujemy 3 poniższe hipotezy.

2.3.1 podpunkt a

Zmienne losowe *Reakcja*, *Rodzaj* oraz *Miejsce* są wzajemnie niezależne - czyli mamy model log-liniowy postaci [1 2 3].

```
model_zadanie3a <- glm(Freq ~ Reakcja + Rodzaj + Miejsce,
                      data = dane.df, family = poisson)
```

Wyznaczamy p – *value* dla powyższego testu.

```
1 - pchisq(deviance(model_zadanie3a), df = df.residual(model_zadanie3a))
## [1] 0.001271524
```

Widzimy, że wartość poziomu krytycznego jest mniejsza niż poziom istotności $\alpha = 0.05$, zatem widzimy, że nasz model [1 2 3] oznaczający niezależność wszystkich zmiennych nie jest poprawny.

2.3.2 podpunkt b

Zmienna losowa *Reakcja* jest niezależna od pary zmiennych *Rodzaj* i *Miejsce* - czyli mamy model log-liniowy postaci [1 23].

```
model_zadanie3b <- glm(Freq ~ Reakcja + Rodzaj + Miejsce +
                      Rodzaj*Miejsce,
                      data = dane.df, family = poisson)
```

Wyznaczamy p – *value* dla powyższego testu.

```
1 - pchisq(deviance(model_zadanie3b), df = df.residual(model_zadanie3b))
## [1] 0.0004539394
```

Ponownie widzimy, że wartość poziomu krytycznego jest mniejsza niż poziom istotności $\alpha = 0.05$, zatem nasz model [1 23] oznaczający, że reakcja na lek jest nie zależy od rodzaju leku bądź miejsca jego podania jest niepoprawny.

2.3.3 punkt c

Zmienna losowa *Reakcja* jest niezależna od zmiennej *Miejsce*, przy ustalonej zmiennej *Rodzaj* - czyli mamy model log-liniowy [12 23].

```
model_zadanie3c <- glm(Freq ~ Reakcja + Rodzaj + Miejsce +
                      Reakcja*Rodzaj + Rodzaj*Miejsce,
                      data = dane.df, family = poisson)
```

Wyznaczamy p - value dla powyższego testu.

```
1 - pchisq(deviance(model_zadanie3c), df = df.residual(model_zadanie3c))
## [1] 0.0001759082
```

Wartość poziomu krytycznego jest mniejsza niż poziom istotności $\alpha = 0.05$, zatem przyjmujemy, że model log-liniowy [12 23] jest niepoprawny. Więc reakcja na lek nie jest niezależna od miejsca jego podania przy ustalonym rodzaju podanego leku.

Hipoteza zerowa jest postaci [1 2 3] oraz hipoteza alternatywna - [1 23].

```
1 - pchisq(deviance(model_zadanie3a)-deviance(model_zadanie3b),
          df = df.residual(model_zadanie3a)-df.residual(model_zadanie3b))
## [1] 1
```

p - value wyszła nam większa niż 0.05, czyli przyjmujemy hipotezę zerową [1 2 3] i odrzucamy hipotezę alternatywną [1 23].

Potwierzamy zadanie, tym razem za hipotezę zerową przyjmujemy model [1 23] a za alternatywną [12 23].

```
1 - pchisq(deviance(model_zadanie3b)-deviance(model_zadanie3c),
          df = df.residual(model_zadanie3b)-df.residual(model_zadanie3c))
## [1] 0.4228065
```

p - value ponownie wychodzi większe niż poziom istotności $\alpha = 0.05$. Zatem ponownie przyjmujemy hipotezę zerową i odrzucamy hipotezę alternatywną.

2.4 Zadanie 4

Dokonamy teraz wyboru omawianych wyżej modeli log-liniowych różnymi sposobami.

2.4.1 Testy

Będziemy porównywać ze sobą poszczególne modele. Za początkowy model przyjmujemy model [1 2 3] i porównamy go z modelem [12 3]. Jeśli test wskaże, że odpowiednią hipotezę powinniśmy odrzucić, za nowy model przyjmujemy model [12 3] i porównamy go z kolejnym. Jeśli zaś nie będziemy mieli podstaw do odrzucenia, zostawimy model [1 2 3] i porównamy go z kolejnym.

Procedurę powtórzymy aż do ostatniego modelu. Skorzystamy z testu `anova()`. Na początku testujemy więc

$H_0 : [1 \ 2 \ 3]$ jest poprawny przeciwko $H_1 : [12 \ 3]$ jest poprawny.

```
anova(model3, model4)

## Analysis of Deviance Table
##
## Model 1: Freq ~ Reakcja + Rodzaj + Miejsce
## Model 2: Freq ~ Reakcja + Rodzaj + Miejsce + Reakcja * Rodzaj
##   Resid. Df Resid. Dev Df Deviance
## 1          4      17.934
## 2          3      17.291  1    0.6425

1-pchisq(deviance(model3)-deviance(model4),
         df = df.residual(model3)- df.residual(model4))

## [1] 0.4228065
```

p -value jest równe 0.4228, więc nie mamy podstaw do odrzucenia hipotezy H_0 . Zatem zostawiamy $[1 \ 2 \ 3]$. Porównujemy teraz z modelem $[1 \ 23]$.

```
anova(model3, model6)

## Analysis of Deviance Table
##
## Model 1: Freq ~ Reakcja + Rodzaj + Miejsce
## Model 2: Freq ~ Reakcja + Rodzaj + Miejsce + Rodzaj * Miejsce
##   Resid. Df Resid. Dev Df    Deviance
## 1          4      17.934
## 2          3      17.934  1 -1.4211e-14

1-pchisq(deviance(model3)-deviance(model6),
         df = df.residual(model3)- df.residual(model6))

## [1] 1

#p value duże, zostawiamy model [1 2 3]. porównamy z modelem [2 13]

model7 <- glm(Freq ~ Reakcja+Rodzaj+Miejsce + Miejsce*Reakcja,
              data = dane.df, family = poisson)
anova(model3, model7)

## Analysis of Deviance Table
##
## Model 1: Freq ~ Reakcja + Rodzaj + Miejsce
## Model 2: Freq ~ Reakcja + Rodzaj + Miejsce + Miejsce * Reakcja
##   Resid. Df Resid. Dev Df Deviance
## 1          4      17.9336
## 2          3      1.3871  1    16.547
```

```

1-pchisq(deviance(model3)-deviance(model7),
         df = df.residual(model3)- df.residual(model7))

## [1] 4.74713e-05

#p value małe, przyjmujemy model [2 13], który porównamy z modelem [12 13]

anova(model7, model5)

## Analysis of Deviance Table
##
## Model 1: Freq ~ Reakcja + Rodzaj + Miejsce + Miejsce * Reakcja
## Model 2: Freq ~ Reakcja + Rodzaj + Miejsce + Reakcja * Rodzaj + Reakcja *
##      Miejsce
##   Resid. Df Resid. Dev Df Deviance
## 1          3      1.38709
## 2          2      0.74459  1    0.6425

1-pchisq(deviance(model7)-deviance(model5),
         df = df.residual(model7)- df.residual(model5))

## [1] 0.4228065

#p value duże, zostawiamy model [2 13], który porównamy z modelem [23 13]

model8 <-glm(Freq ~ Reakcja+Rodzaj+Miejsce+Miejsce*Rodzaj+Miejsce*Reakcja,
             data = dane.df, family = poisson)
anova(model7, model8)

## Analysis of Deviance Table
##
## Model 1: Freq ~ Reakcja + Rodzaj + Miejsce + Miejsce * Reakcja
## Model 2: Freq ~ Reakcja + Rodzaj + Miejsce + Miejsce * Rodzaj + Miejsce *
##      Reakcja
##   Resid. Df Resid. Dev Df      Deviance
## 1          3      1.3871
## 2          2      1.3871  1 -3.9968e-15

1-pchisq(deviance(model7)-deviance(model8),
         df = df.residual(model7)- df.residual(model8))

## [1] 1

#p-value duże, zostawiamy model [2 13]

```

Metoda testów pozwoliła nam wybrać model [2 13] jako najlepszy model log-liniowy.

2.4.2 Kryterium AIC i BIC

By porównać modele metodami AIC oraz BIC, stworzymy 19 modeli log-liniowych (wszystkie możliwe kombinacje), po czym skorzystamy z funkcji AIC oraz BIC. Wyniki przedstawiamy w poniższej tabeli, z której następnie wyciągniemy odpowiednie wnioski.

	AIC	BIC
0	104.80	104.88
[1]	60.83	60.99
[2]	106.80	106.96
[3]	106.80	106.96
[1 : 2]	62.83	63.06
[2 : 3]	108.80	109.03
[1 : 3]	62.83	63.06
[12]	64.18	64.50
[23]	110.80	111.11
[13]	48.28	48.60
[1 : 2 : 3]	64.83	65.14
[1 : 23]	66.83	67.22
[2 : 13]	50.28	50.68
[3 : 12]	66.18	66.58
[12 : 23]	68.18	68.66
[13 : 23]	52.28	52.76
[12 : 13]	51.64	52.11
[12 : 13 : 23]	53.58	54.14
[123]	54.89	55.53

Tabela 3: Porównanie AIC i BIC

Ponieważ najlepszym modelem ze względu na kryteria AIC i BIC to te, których wartość jest najmniejsza, więc na podstawie tabeli możemy ustalić, że ze względu na oba kryteria najlepszy jest model [13]. Ten model porównamy więc z modelem [2 13], który uzyskaliśmy jako najlepszy poprzez metodę testów. Przetestujemy hipotezę:

$$H_0 : [13] \text{ jest poprawny} \quad \text{przeciwko} \quad H_1 : [2 \ 13] \text{ jest poprawny.}$$

```
anova(m9, m12)

## Analysis of Deviance Table
##
## Model 1: Freq ~ Reakcja + Miejsce + Reakcja * Miejsce
## Model 2: Freq ~ Reakcja + Rodzaj + Miejsce + Reakcja * Miejsce
##   Resid. Df Resid. Dev Df   Deviance
## 1         4      1.3871
## 2         3      1.3871  1 2.2204e-16

1-pchisq(deviance(m9)-deviance(m12),
         df = df.residual(m9)- df.residual(m12))

## [1] 1
```

Duża wartość p -value sugeruje, że nie mamy podstaw do odrzucenia hipotezy o poprawności modeli [13], więc to on będzie najlepszym modelem log-liniowym.

Korzystając z biblioteki **MASS** i funkcji **stepAIC** wyznaczamy wartości AIC

```
MASS::stepAIC(m18)

## Start:  AIC=54.89
## Freq ~ Reakcja + Rodzaj + Miejsce + Reakcja * Rodzaj + Rodzaj *
##       Miejsce + Reakcja * Miejsce + Reakcja * Rodzaj * Miejsce
##
##               Df Deviance    AIC
## - Reakcja:Rodzaj:Miejsce  1  0.68843 53.581
## <none>                    0.00000 54.893
##
## Step:  AIC=53.58
## Freq ~ Reakcja + Rodzaj + Miejsce + Reakcja:Rodzaj + Rodzaj:Miejsce +
##       Reakcja:Miejsce
##
##               Df Deviance    AIC
## - Rodzaj:Miejsce    1   0.7446 51.637
## - Reakcja:Rodzaj    1   1.3871 52.280
## <none>              0.6884 53.581
## - Reakcja:Miejsce  1  17.2911 68.184
##
## Step:  AIC=51.64
## Freq ~ Reakcja + Rodzaj + Miejsce + Reakcja:Rodzaj + Reakcja:Miejsce
##
##               Df Deviance    AIC
## - Reakcja:Rodzaj    1   1.3871 50.280
## <none>              0.7446 51.637
## - Reakcja:Miejsce  1  17.2911 66.184
##
## Step:  AIC=50.28
## Freq ~ Reakcja + Rodzaj + Miejsce + Reakcja:Miejsce
##
##               Df Deviance    AIC
## - Rodzaj            1   1.3871 48.280
## <none>              1.3871 50.280
## - Reakcja:Miejsce  1  17.9336 64.826
##
## Step:  AIC=48.28
## Freq ~ Reakcja + Miejsce + Reakcja:Miejsce
##
##               Df Deviance    AIC
## <none>          1.3871 48.280
## - Reakcja:Miejsce  1  17.9336 62.826
##
## Call:  glm(formula = Freq ~ Reakcja + Miejsce + Reakcja:Miejsce, family = poisson,
```

```
##      data = dane.df)
##
## Coefficients:
##              (Intercept)              ReakcjaJest poprawa
##                  3.7612                  -1.8153
##      MiejsceSzpital  ReakcjaJest poprawa:MiejsceSzpital
##                  -0.3435                  1.3680
##
## Degrees of Freedom: 7 Total (i.e. Null);  4 Residual
## Null Deviance:      63.9
## Residual Deviance: 1.387  AIC: 48.28
```

A następnie korzystając z tej samej biblioteki oraz funkcji $\text{stepAIC}(m18, k = \log(8))$ wyznaczamy wartości BIC .

```
MASS::stepAIC(m18, k = log(8))

## Start:  AIC=55.53
## Freq ~ Reakcja + Rodzaj + Miejsce + Reakcja * Rodzaj + Rodzaj *
##      Miejsce + Reakcja * Miejsce + Reakcja * Rodzaj * Miejsce
##
##              Df Deviance    AIC
## - Reakcja:Rodzaj:Miejsce  1  0.68843 54.137
## <none>                    0.00000 55.528
##
## Step:  AIC=54.14
## Freq ~ Reakcja + Rodzaj + Miejsce + Reakcja:Rodzaj + Rodzaj:Miejsce +
##      Reakcja:Miejsce
##
##              Df Deviance    AIC
## - Rodzaj:Miejsce  1  0.7446 52.114
## - Reakcja:Rodzaj  1  1.3871 52.757
## <none>            0.6884 54.137
## - Reakcja:Miejsce  1 17.2911 68.661
##
## Step:  AIC=52.11
## Freq ~ Reakcja + Rodzaj + Miejsce + Reakcja:Rodzaj + Reakcja:Miejsce
##
##              Df Deviance    AIC
## - Reakcja:Rodzaj  1  1.3871 50.677
## <none>            0.7446 52.114
## - Reakcja:Miejsce  1 17.2911 66.581
##
## Step:  AIC=50.68
## Freq ~ Reakcja + Rodzaj + Miejsce + Reakcja:Miejsce
##
##              Df Deviance    AIC
## - Rodzaj  1  1.3871 48.598
```

```
## <none>                1.3871 50.677
## - Reakcja:Miejsce  1  17.9336 65.144
##
## Step:  AIC=48.6
## Freq ~ Reakcja + Miejsce + Reakcja:Miejsce
##
##                Df Deviance    AIC
## <none>                1.3871 48.598
## - Reakcja:Miejsce  1  17.9336 63.065
##
## Call:  glm(formula = Freq ~ Reakcja + Miejsce + Reakcja:Miejsce, family = poisson,
##           data = dane.df)
##
## Coefficients:
##                (Intercept)                ReakcjaJest poprawa
##                   3.7612                   -1.8153
##           MiejsceSzpital ReakcjaJest poprawa:MiejsceSzpital
##                   -0.3435                   1.3680
##
## Degrees of Freedom: 7 Total (i.e. Null);  4 Residual
## Null Deviance:      63.9
## Residual Deviance: 1.387  AIC: 48.28
```