

Sprawozdanie 2

Lista 5,6 i 7

Zad 2

Założmy, że 200 losowo wybranych klientów (w różnym wieku) kilku (losowo wybranych) aptek zapytano, jaki lek przeciwbólowy zwykle stosują. Zebrane dane zawarte są w tablicy 1.

```
print(data)
```

##	Lek	Wiek_ankietowanych do lat 35	od 36 do 55	powyżej 55	suma
##	Lbuprom	35	0	0	35
##	Aap	22	22	0	44
##	Paracetamol	15	15	15	45
##	Ibuprofen	0	40	10	50
##	Panadol	18	3	5	26
##	Suma	90	80	30	200

a)

Prawdopodobieństwo, że losowo wybrana osoba z badanej populacji w przypadku bólu zażywa Apap jest mniejsze bądź równe $1/4$:

```
binom.test(44, 200, p=1/4, alternative = "g")$p.value
## [1] 0.8562401
prop.test(44, 200, p=1/4, alternative = "g", correct = T)$p.value
## [1] 0.8154462
prop.test(44, 200, p=1/4, alternative = "g", correct = F)$p.value
## [1] 0.8364066
```

Każda z powyższych p-wartości jest większa niż 0,05, zatem w zadanym z powyższych testów nie możemy odrzucić hipotezy, że prawdopodobieństwo, że losowo wybrana osoba z badanej populacji w przypadku bólu zażywa Apap jest mniejsze bądź równe $1/4$.

b)

Prawdopodobieństwo, że losowo wybrana osoba z badanej populacji w przypadku bólu zażywa Apap jest równe $1/2$:

```
binom.test(44, 200, p = 0.5, alternative = "t")$p.value
## [1] 6.837911e-16
prop.test(44, 200, p = 0.5, alternative = "t", correct = T)$p.value
## [1] 4.197514e-15
prop.test(44, 200, p = 0.5, alternative = "t", correct = F)$p.value
## [1] 2.382836e-15
```

Widzimy zatem, że p-wartości w każdym teście są mniejsze niż 0,05, zatem odrzucamy hipotezę, że prawdopodobieństwo, że losowo wybrana osoba z badanej populacji w przypadku bólu zażywa Apap jest równe $1/2$.

c)

Prawdopodobieństwo, że losowo wybrana osoba z badanej populacji w przypadku bólu zażywa Ibuprom jest większe bądź równe $1/5$:

```
binom.test(50, 200, p=1/5, alternative = "l")$p.value
## [1] 0.9655032
prop.test(50, 200, p=1/5, alternative = "l", correct = T)$p.value
## [1] 0.9534609
prop.test(50, 200, p=1/5, alternative = "l", correct = F)$p.value
## [1] 0.9614501
```

Każda z powyższych p-wartości jest większa niż 0,05, zatem w zadanym z powyższych testów nie możemy odrzucić hipotezy, że prawdopodobieństwo, że losowo wybrana osoba z badanej populacji w przypadku bólu zażywa Ibuprom jest większe bądź równe $1/5$.

d)

Powtórzyć punkt (a), (b) i (c), ale dla osoby z badanej populacji do lat 35:

```
binom.test(22, 90, p=1/4, alternative = "g")$p.value
## [1] 0.5885826
prop.test(22, 90, p=1/4, alternative = "g", correct = T)$p.value
## [1] 0.5
prop.test(22, 90, p=1/4, alternative = "g", correct = F)$p.value
```

```
## [1] 0.5484381
```

Tak samo jak w podpunkcie a) nie odrzucamy badanej hipotezy.

```
binom.test(22, 90, p = 0.5, alternative = "t")$p.value
## [1] 1.249258e-06
prop.test(22, 90, p = 0.5, alternative = "t", correct = T)$p.value
## [1] 2.101436e-06
prop.test(22, 90, p = 0.5, alternative = "t", correct = F)$p.value
## [1] 1.241945e-06
```

Tak samo jak w podpunkcie b) odrzucamy badaną hipotezy.

```
binom.test(0, 90, p=1/5, alternative = "l")$p.value
## [1] 1.897138e-09
prop.test(0, 90, p=1/5, alternative = "l", correct = T)$p.value
## [1] 1.997379e-06
prop.test(0, 90, p=1/5, alternative = "l", correct = F)$p.value
## [1] 1.050718e-06
```

W tym przypadku widzimy że p-wartości w kazym z wykonanych testów

Zad 3

Na podstawie danych w tablicy 1, korzystając z testu Fishera, na poziomie istotności $\alpha = 0.05$, zweryfikować hipotezę, że prawdopodobieństwo, że osoba do lat 35 zażywa Panadol jest równe prawdopodobieństwu, że osoba od 36 lat do 55 lat zażywa Panadol.

Czy na podstawie uzyskanego wyniku można (na zadanym poziomie istotności) odrzucić hipotezę o niezależności wyboru leku Panadol w leczeniu bólu od wieku, przy uwzględnieniu tylko dwóch grup wiekowych - do lat 35 i od 36 do 55 lat?

```
data2 <- matrix(c(18,72,3,77),nrow = 2, ncol = 2)
dimnames(data2) <- list(
  Wiek_ankietowanych = c("Panadol", "Inny"),
  Lek = c("do lat 35", "od 36 do 55"))
print(data2)

##              Lek
## Wiek_ankietowanych do lat 35 od 36 do 55
##           Panadol      18          3
##           Inny       72         77

fisher.test(data2)$p.value
```

```
## [1] 0.001788538
```

Jak widzimy p-wartość jest mniejsza niż 0.05, zatem odrzucamy hipotezę o niezależności rozkładów. Zatem badane rozkłady warunkowe nie są jednorodne.

Zad 4

Korzystając z funkcji `chisq.test` w pakiecie R, na poziomie istotności 0.05, zweryfikować hipotezę o niezależności stopnia zadowolenia z pracy i wynagrodzenia na podstawie danych w tablicy 2. Zwrócić uwagę na stosowaną w tej funkcji poprawkę.

data3						
##	Zadowolenia					
##	Wynagrodzenie	b. niezadow.	niezadow.	zadow.	b. zadow.	Suma
##	do 6000	20	24	80	82	206
##	6000–15000	22	38	104	125	289
##	15000–25000	13	28	81	113	235
##	powyżej 25000	7	18	54	92	171
##	Suma	62	108	5	26	901

```
chisq.test(data3[1:4,1:4])$p.value  
## [1] 0.2139542
```

Jak możemy zobaczyć p-wartość jest większa niż 0.05, zatem nie mamy podstaw aby odrzucić hipotezę o niezależności rozkładów warunkowych.

Zad 5

Napisać deklarację funkcji, która dla danych w tablicy dwudzielczej oblicza wartość poziomu krytycznego (p-value) w asymptotycznym teście niezależności opartym na ilorazie wiarygodności. Korzystając z napisanej funkcji, obliczyć tę wartość dla danych z zadania 4.

Lista 8 i 9

Zad 1

Na podstawie danych z przykładu 1 na wykładzie 7, obliczyć wartości odpowiednich miar współzmienności zmiennych Segregacja (odpowiedź na pytanie dotyczące segregacji śmieci) i Wiek oraz Segregacja i Miejsce zamieszkania. W przypadku zmiennej Wiek, wartości miar obliczyć przy “wyjściowych” kategoriach wiekowych, jak również przy połączonych kategoriach wiekowych (jak w przykładzie). Podać interpretację tych wartości. Następnie przeprowadzić analizę korespondencji, tzn. obliczyć wartości odpowiednich macierzy, współrzędnych punktów oraz utworzyć odpowiednie wykresy.

Zad 2

Założmy, że 200 klientów (w różnym wieku) kilku aptek zapytano, jaki lek przeciwbólowy zwykle stosują. Zebrane dane zawarte są w tablicy 2. Na podstawie tych danych, obliczyć odpowiednie miary współzmienności oraz przeprowadzić analizę korespondencji, tzn. obliczyć wartości odpowiednich macierzy, współrzędnych punktów oraz utworzyć odpowiednie wykresy.

Zad 3

Na podstawie danych zawartych w tablicy 1, obliczyć (odpowiednią) miarę współzmienności zmiennych Wynagrodzenie i Stopień zadowolenia z pracy (dane te są trochę inne niż te rozpatrywane na poprzedniej liście). Następnie, przeprowadzić analizę korespondencji, tzn. obliczyć wartości odpowiednich macierzy, współrzędnych punktów oraz utworzyć odpowiednie wykresy.