

Labolatorium nr 4- zadanie domowe

Mateusz Kapusta

2022-03-24

Zadanie nr 1

Wpierw jak zwykle ładujemy dane i dodajemy kolumnę z województwami.

```
Zadluzenie.gmin <- read.delim("Zadluzenie gmin.csv", colClasses = c('factor', 'factor', 'factor', 'nume
Zadluzenie.gmin$Wojewodztwo<-sownik[sapply(Zadluzenie.gmin$Kod.Teryt,substr,start=0,stop=2)]
```

Przygotowujemy dwie ramki danych dla województwa Opolskiego i Dolnośląskiego.

```
dolne<-Zadluzenie.gmin[Zadluzenie.gmin$Wojewodztwo=="Dolnośląskie",]
opole<-Zadluzenie.gmin[Zadluzenie.gmin$Wojewodztwo=="Opolskie",]
```

Teraz pora policzyć odpowiednie wariancje i średnie

```
ndolne<-nrow(dolne)
nopoie<-nrow(opole)
dolmean<-mean(dolne$Zadłużenie.gmin)
opolmean<-mean(opole$Zadłużenie.gmin)
wardol<-var(dolne$Zadłużenie.gmin)
waropoie<-var(opole$Zadłużenie.gmin)
```

Teraz pora na policzenie wartości wartości statystyki Welscha.

```
wel<-(dolmean-opolmean)/sqrt(waropoie/nopoie+wardol/ndolne)
swo<-(waropoie/nopoie+wardol/ndolne)^2/(waropoie^2/(nopoie^2*(nopoie-1))+wardol^2/(ndolne^2*(ndolne-1)))
```

wartość naszej statystyki to 4.2447338 co biorąc pod uwagę liczbę stopni swobody 152.9490226 pozwala nam określić jaki jest zbiór krytyczny dla naszego przypadku.

```
a<-0.05
qt(a/2,df=swo)
```

```
## [1] -1.975596
```

widzimy więc, że wartość naszej statystyki wpada do naszego zbioru krytycznego i hipotezę o równym zadłużeniu należy odrzucić. Proawdopodobieństwo tego, że nasza statystyka przyjmuje co do modułu większe wartości niż wartość wel to

```
2*(1-pt(wel,swo))
```

```
## [1] 3.781472e-05
```

zastosowaliśmy tutaj test dwustronny. Porównując z funkcją dostępną w języku R

```
t.test(dolne$Zadłużenie.gmin, opole$Zadłużenie.gmin, alternative="two.sided", var.equal=FALSE)
```

```
##
## Welch Two Sample t-test
```

```
##
## data:  dolne$Zadłużenie.gmin and opole$Zadłużenie.gmin
## t = 4.2447, df = 152.95, p-value = 3.781e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  4.884523 13.389815
## sample estimates:
## mean of x mean of y
## 32.76731 23.63014
```

Widzimy więc, że otrzymujemy dokładnie takie same wartości jak ręcznie obliczone.

Zadanie nr 2

Macierz kontyngencji uzyskana na zajęciach to

```
kon<-matrix(c(11,62,17,20),nrow=2,byrow=TRUE)
```

Prawdopodobieństwo znalezienia takiego układu lub z mniejszą wartością liczby żuków znalezionych w składach drewna przed 2000 r to

```
phyper(11, 73, 37, 28,lower.tail = TRUE)
```

```
## [1] 0.0006345026
```

Wykonując teraz test Fishera zaimplementowany w języku R mamy

```
fisher.test(kon,alternative = "less")
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  kon
## p-value = 0.0006345
## alternative hypothesis: true odds ratio is less than 1
## 95 percent confidence interval:
##  0.0000000 0.4953036
## sample estimates:
## odds ratio
## 0.2122357
```

Widzimy więc, że otrzymane wyniki są takie same a hipotezę o niezależności należy odrzucić.