

Laboratorium 5

1. G-Test

Rzucono 40 razy kostką do gry (wyniki w **Baza 5.1**). Na podstawie uzyskanych wyników przetestować czy kostka jest rzetelna.

```
library(readxl)

library(DescTools)

Baza<- read_excel("Baza 5.1.xlsx")

X<-Baza$"Wyniki"

n<-length(X)

N<-c()

for (i in 1:6)

{

  N[i]<-0

  for (j in 1:n)

  {

    if(X[j]==i){N[i]=N[i]+1}

  }

}

P<-c(1/6,1/6,1/6,1/6,1/6,1/6)

GTest(x=N,p=P)

G<-0

for (j in 1:6)

{

  if(N[j]!=0){G=G+2*N[j]*log(N[j]/(n*P[j]))}

}

p<-1-pchisq(G,5)

p

Obliczmy teraz „dokładną nie asymptotyczną” p-wartość metodą MC

GY<-c()

MC<-100000

for (k in 1:MC)

{

  Y <- ceiling(runif(n,0,6))

  NY<-c()

  for (i in 1:6)

  {
```

```

NY[i]<-0

for (j in 1:n)

{

  if(Y[j]==i){NY[i]<-NY[i]+1}

}

}

GY[k]<-0

for (j in 1:6)

{

  if(NY[j]!=0){GY[k]<-GY[k]+2*NY[j]*log(NY[j]/(n*P[j]))}

}

}

pMC<-0

for (k in 1:MC)

{

  if(GY[k]>=G){pMC=pMC+1}

}

pMC=pMC/MC

pMC

```

2. Klasyczny test Chi-kwadrat

Przeprowadzić test Chi-kwadrat rzetelności kostki dla danych z **Baza 5.1**.

```

library(readxl)

Baza<- read_excel("Baza 5.1.xlsx")

X<-Baza$"Wyniki"

n<-length(X)

N<-c()

for (i in 1:6)

{

  N[i]<-0

  for (j in 1:n)

  {

    if(X[j]==i){N[i]=N[i]+1}

  }

}

P<-c(1/6,1/6,1/6,1/6,1/6,1/6)

chisq.test(N,p=P)

```

Obliczmy teraz „dokładną nie asymptotyczną” p-wartość metodą MC

```
chisq.test(N,p=P,simulate.p.value = TRUE,B=100000)
```

Bez gotowca:

```
Q<-0
```

```
for (j in 1:6)
```

```
{
```

```
  Q=Q+(N[j]- (n*P[j]))^2/(n*P[j])
```

```
}
```

```
p<-1-pchisq(Q,5)
```

```
p
```

Zadanie do samodzielnego wykonania: Oblicz „dokładną nie asymptotyczną” p-wartość testu Chi-kwadrat metodą MC bez gotowca.

3. Test niezależności Chi-kwadrat

W latach sześćdziesiątych przeprowadzono badania wśród amerykańskiej młodzieży szkolnej mające wyjaśnić, czy istnieje związek pomiędzy poglądami politycznymi a paleniem marihuany (wyniki w **Baza 5.1**). Zweryfikować hipotezę o niezależności stosunku do palenia marihuany (1-nigdy, 2-okazjonalnie, 3-często) od poglądów politycznych (1-liberalne, 2-konserwatywne, 3-inne).

```
library(readxl)
```

```
Baza<- read_excel("Baza 5.2.xlsx")
```

```
X<-Baza$"Marichuana"
```

```
Y<-Baza$"Poglady"
```

```
table(X,Y)
```

```
chisq.test(table(X, Y))
```

Bez gotowca:

```
n<-length(X)
```

```
N<-table(X,Y)
```

```
m<-nrow(N)
```

```
k<-ncol(N)
```

```
Q<-0
```

```
for (i in 1:m)
```

```
{
```

```
  for (j in 1:k)
```

```
  {
```

```
    Q=Q+(N[i,j]-sum(N[i,])*sum(N[,j])/n)^2/(sum(N[i,])*sum(N[,j])/n)
```

```
  }
```

```
}
```

```
p<-1-pchisq(Q,k*m-(k+m-2)-1)
```

Q

p

4. Test zgodności Chi-kwadrat hipoteza złożona

Cechy dziedziczne są przekazywane za pomocą genów występujących parami. W najprostszym przypadku , każdy gen może mieć dwie formy (allele), które będziemy oznaczać odpowiednio literami A i a . Oznaczmy przez $P(A)=\theta$ nieznane prawdopodobieństwo występowania genu A w populacji. Z teorii Mendla losowego tworzenia się genotypów potomstwa otrzymujemy następujący rozkład prawdopodobieństwa genotypów populacji w stanie równowagi (równowaga Hardy’ego Weinberga)

	AA	Aa	aa
p	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

Zaobserwowano następujące licznosci genotypów

	n_{AA}	n_{Aa}	n_{aa}
Obserw	110	235	155

Czy na poziomie 0.05 są podstawy do odrzucenia hipotezy równowagi H-W?

Estymatorem największej wiarygodności θ jest $\hat{\theta} = \frac{2n_{AA}+n_{Aa}}{2n}$ (policzyć samodzielnie).

```
N<-c(110,235,155)
```

```
n<-sum(N)
```

```
T<-(2*N[1]+N[2])/(2*n)
```

```
P<-c(T^2,2*T*(1-T),(1-T)^2)
```

```
Q<-0
```

```
for (j in 1:3)
```

```
{
  Q=Q+(N[j]- (n*P[j]))^2/(n*P[j])
}
```

```
p<-1-pchisq(Q,1)
```

```
N/n
```

```
P
```

```
P
```