

ZADANIA

Zadanie 0. W badaniu opinii publicznej często pojawiają się więcej niż dwie opinie. Dla przykład, założmy, że przeprowadzono ankietę zarejestrowanych wyborców, aby zobaczyć, który kandydat prawdopodobnie zostanie wybrany w nadchodzących wyborach. Dla uproszczenia zakładamy że są dwaj kandydaci, A i B. Potencjalny wyborca może wybrać jednogo z nich lub może być niezdecydowany. Jeśli ankietowanych jest 100 osób, a wyniki wynoszą 35 dla A, 40 dla B i 25 niezdecydowanych, czy jest różnica poparciem dla kandydata A oraz B jest znacząca? Założmy na przykład, że chcemy wiedzieć, czy rozkład poparcia dla kandydatów w całej populacji głosujących jest zgodny z rozkładem, w którym 35% wyborców popiera kandydata A, 35% kandydata B oraz 30% wyborców jest niezdecydowanych?

Możemy to sprawdzić wykonując następujący skrypt w programie R

```
> y <- c(35, 40, 25)
> p <- c(35, 35, 30) # zakładane procentowe poparcie
> p <- p/sum(p) # zakładany rozkład
> n <- sum(y)
> chi2 <- sum( (y - n*p)^2 / (n*p) )
> chi2
```

```
[1] 1.547619
```

```
> pchisq(chi2, df=3 - 1, lower.tail=FALSE)
```

```
[1] 0.4612526
```

Możemy to samo wykonać korzystając z gotowej funkcji **chisq.test**, programu R.

```
> y <- c(35, 40, 25)
> p <- c(35, 35, 30)
> p <- p/sum(p)
> chisq.test(y,p=p)
```

Chi-squared test for given probabilities

```
data: y
X-squared = 1.5476, df = 2, p-value = 0.4613
```

Wracając do wyjściowego zagadnienia. Chcemy sprawdzić hipotezę zerową postaci

$$H_0 : p_1 = p_2,$$

przeciwko hipotezie alternatywnej

$$H_1 : p_1 \neq p_2.$$

Hipotezę tą można przetestować za pomocą statystyki χ^2 , ale musimy najpierw określić, co mamy na myśli pisząc $p_1 = p_2$, ponieważ w H_0 nie jest to w pełni określone. Wartości p_1 i p_2 zależą od wartości p_3 , która nie jest określona. Zakładając, że $p_1 = p_2$ możemy,

na podstawie danych dokonać estymowacji wartości p_1 , $i = 1, 2, 3$. Aby oszacować $p_1 = p_2$, używamy obu wartości danych odnoszących się do p_1 i p_2 i przyjmujemy, że

$$\hat{p}_1 = \hat{p}_2 = \frac{y_1 + y_2}{2n}$$

oraz

$$\hat{p}_3 = \frac{y_3}{n} = 1 - \hat{p}_1 - \hat{p}_2.$$

Statystyka χ^2 ma wtedy postać

$$\chi^2 = \sum_{i=1}^3 \frac{(y_i - n\hat{p}_i)^2}{n\hat{p}_i}.$$

Ponieważ estymowaliśmy jedną wartość $p = p_1 = p_2$, więc statystyka testowa ma 1 stopień swobody. Korzystając z programu R możemy to sprawdzić wykonując następujący skrypt:

```
> y <- c(35, 40, 25)
> n <- sum(y)
> phat1 <- phat2 <- sum(y[1:2])/(2*n)
> phat3 <- 1 - phat1 - phat2
> phat <- c(phat1, phat2, phat3)
> obs <- sum((y - n*phat)^2/(n*phat))
> obs
```

```
[1] 0.3333333
```

```
> pchisq(obs, df =1 , lower.tail=FALSE)
```

```
[1] 0.5637029
```

Krzysztof Topolski