ADPS 2025Z — Laboratorium 3

Konrad Jędrzejewski, Marek Rupniewski

Przykład 1 – test wartości średniej

Wygeneruj 30 liczb będących próbą z rozkładu N(1,4):

```
n = 30; mi = 1; sigma = 2
x = rnorm(n, mean = mi, sd = sigma)
```

Dla wygenerowanych danych zweryfikuj hipotezę zerową:

$$H_0 : \mu = \mu_0,$$
$$H_1 : \mu \neq \mu_0,$$

dla $\mu_0 = 1$, przy poziomie istotności $\alpha = 0.05$, przy założeniu, że wariancja rozkładu jest nieznana.

Nieznana wariancja:

```
mi_0 = 1
alfa = 0.05
T = abs(mean(x) - mi_0)*sqrt(n)/sd(x)
c = qt(1 - alfa/2, df = n-1)
p_val = 2*(1 - pt(T, df = n-1))
```

Wartość statystyki T = 0.3481.

Wartość krytyczna dla poziomu istotności $\alpha = 0.05$ wynosi c = 2.0452.

p-wartość = 0.7303.

Wykorzystanie funkcji t.test:

```
t.test(x, mu = mi_0, alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  x
## t = 0.34808, df = 29, p-value = 0.7303
## alternative hypothesis: true mean is not equal to 1
## 95 percent confidence interval:
##  0.298923 1.988656
## sample estimates:
## mean of x
##   1.14379
```

Powtórz testy dla innych wartości $\mu_0 = 0, 1.5, 3$.

Przykład 2 – test zgodności Pearsona

Kostka symetryczna

Zasymuluj 50 rzutów kostką i sprawdź, czy kostka jest symetryczna.

```
n = 50; x = sample(1:6, n, replace = T)
ni_i = as.data.frame(table(factor(x, levels = 1:6)))$Freq
p_i = rep(1/6, 6)
T = sum((ni_i - n*p_i)^2 / (n*p_i))
r = 6
alfa = 0.05
c = qchisq(1 - alfa, r - 1)
p_val = (1 - pchisq(T, r - 1))
```

Wartość statystyki T = 5.2.

Wartość krytyczna dla poziomu istotności $\alpha = 0.05$ wynosi c = 11.0705.

p-wartość = 0.392.

Wykorzystanie funkcji chisq.test:

```
chisq.test(ni_i, p = p_i)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  ni_i
## X-squared = 5.2, df = 5, p-value = 0.392
```

Kostka niesymetryczna

Powtórz testy w przypadku gdy zakładamy, że kostka nie jest symetryczna a prawdopodobieństwa wyrzucenia kolejnych liczb oczek są równe 1/9, 1/9, 1/9, 2/9, 2/9, 2/9.

```
n = 50; x = sample(1:6, n, replace = T)
ni_i = as.data.frame(table(factor(x, levels = 1:6)))$Freq
p_i = c(1/9, 1/9, 1/9, 2/9, 2/9, 2/9)
T = sum((ni_i - n*p_i)^2 / (n*p_i))
r = 6
alfa = 0.05
c = qchisq(1 - alfa, r - 1)
p_val = (1 - pchisq(T, r - 1))
```

Wartość statystyki T = 10.21.

Wartość krytyczna dla poziomu istotności $\alpha = 0.05$ wynosi c = 11.0705.

p-wartość = 0.0695.

Wykorzystanie funkcji chisq.test:

```
chisq.test(ni_i, p = p_i)
```

```
##
##  Chi-squared test for given probabilities
##
```

```
## data:  ni_i
## X-squared = 10.21, df = 5, p-value = 0.0695
```

Przykład 3 – test Kołmogorowa-Smirnowa

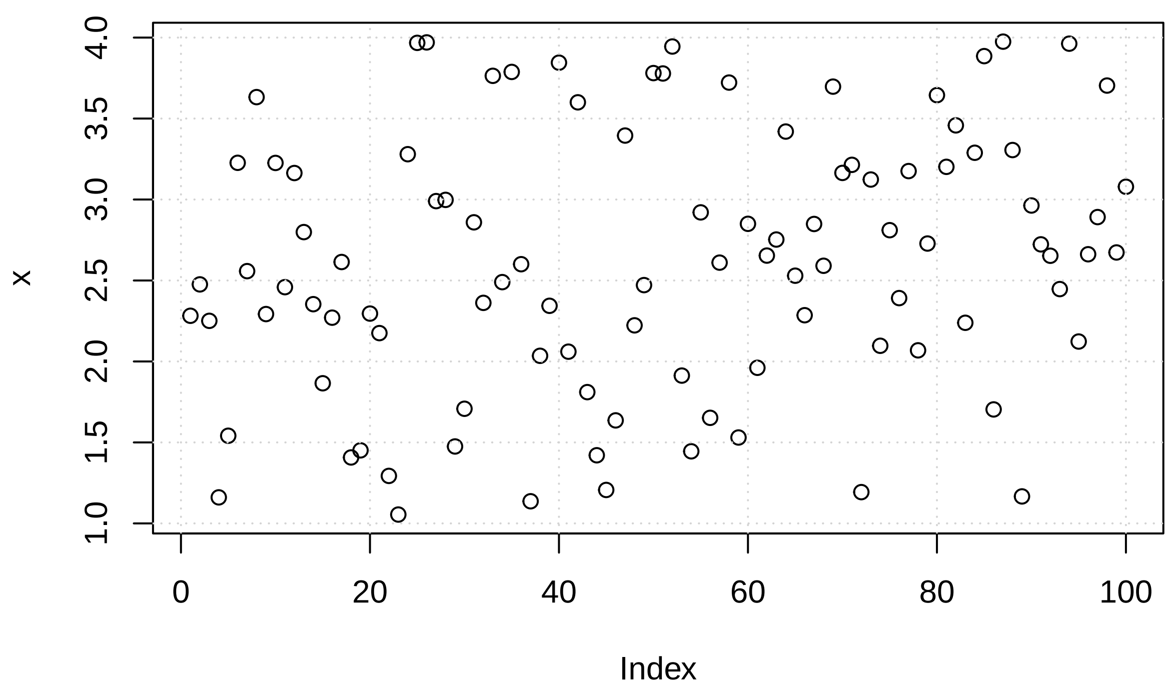
Dane o rozkładzie jednostajnym

Wygeneruj 100 liczb będących próbą z rozkładu jednostajnego w przedziale od 1 do 4:

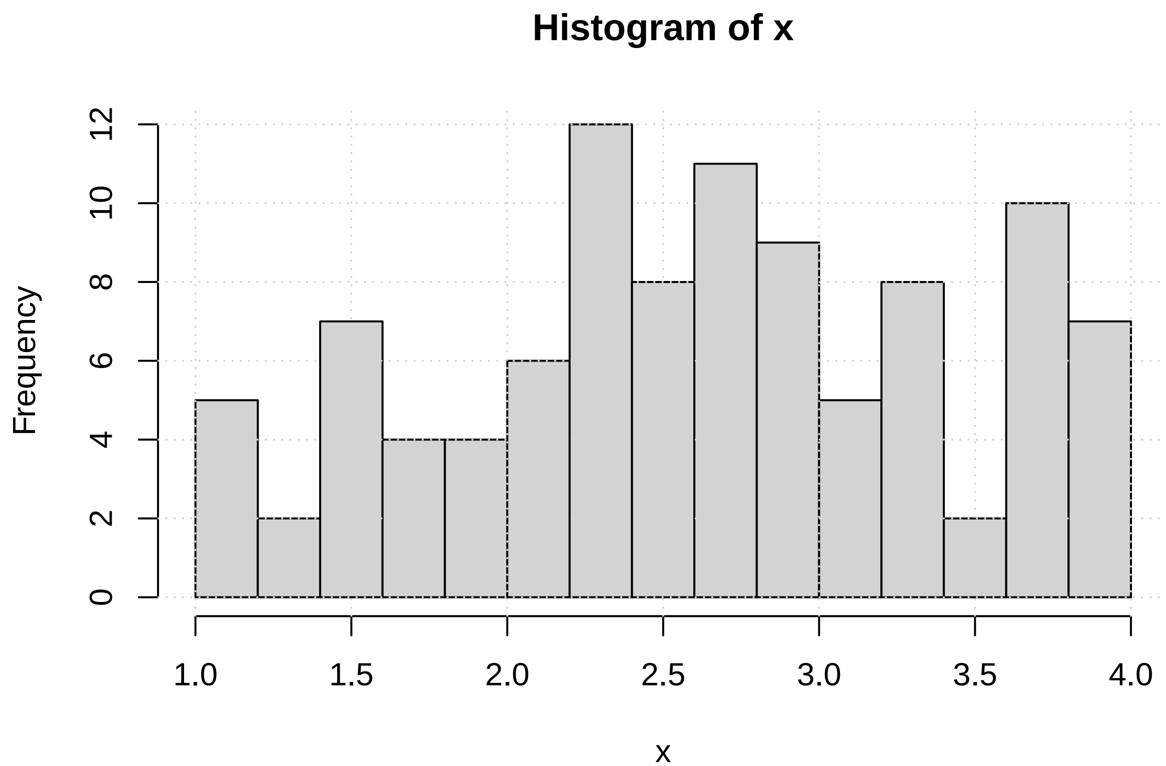
```
n = 100; a = 1; b = 4
x = runif(n, min = a, max = b)
```

Obejrzyj wartosci próbek i ich histogram.

```
plot(x)
grid()
```




```
hist(x, breaks = 15)
grid()
```



Przeprowadź test Kołmogorowa-Smirnowa dla różnych założeń dotyczących krańców przedziałów:

```
ks.test(x, 'punif', min = a, max = b)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.12766, p-value = 0.07681
## alternative hypothesis: two-sided
```

```
ks.test(x, 'punif', min = a - 0.5, max = b - 0.5)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.29433, p-value = 5.977e-08
## alternative hypothesis: two-sided
```

```
ks.test(x, 'punif', min = a - 0.5, max = b + 0.5)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.16374, p-value = 0.009381
## alternative hypothesis: two-sided
```

Zweryfikuj hipotezę, że wygenerowane dane pochodzą z rozkładu normalnego o parametrach będących wartością średnią i wariancją z próby:

```
ks.test(x, 'pnorm', mean = mean(x), sd = sd(x))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.059188, p-value = 0.8748
## alternative hypothesis: two-sided
```

Sprawdź wynik po zwiększeniu liczby próbek do 500:

```
x = runif(500, min = a, max = b)
ks.test(x, 'pnorm', mean = mean(x), sd = sd(x))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.058236, p-value = 0.06731
## alternative hypothesis: two-sided
```

Dane o rozkładzie normalnym

Zweryfikuj czy dane z rozkładu normalnego dopasowują się do rozkładu jednostajnego:

```
n = 100; mi = 1; sigma = 2
x = rnorm(n, mean = mi, sd = sigma)
ks.test(x, 'punif', min = mean(x) - sd(x), max = mean(x) + sd(x))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.19, p-value = 0.001464
## alternative hypothesis: two-sided
```

```
ks.test(x, 'punif', min = mean(x) - 2*sd(x), max = mean(x) + 2*sd(x))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.11557, p-value = 0.1383
## alternative hypothesis: two-sided
```

```
ks.test(x, 'punif', min = mean(x) - 3*sd(x), max = mean(x) + 3*sd(x))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.21031, p-value = 0.0002879
## alternative hypothesis: two-sided
```

Przykład 4 – testy normalności

Dane o rozkładzie normalnym

Wygeneruj 200 liczb będących próbą losową z rozkładu N(1,4):

```
n = 200; mi = 1; sigma = 2
x = rnorm(n, mean = mi, sd = sigma)
```

Korzystając z testu Kołmogorowa-Smirnowa zweryfikuj hipotezę, że dane pochodzą z rozkładu N(1,4):

```
ks.test(x, 'pnorm', mean = 1, sd = 2)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.062922, p-value = 0.4069
## alternative hypothesis: two-sided
```

Korzystając z testu Kołmogorowa-Smirnowa zweryfikuj hipotezę, że dane pochodzą z rozkładu N(0,9):

```
ks.test(x, 'pnorm', mean = 0, sd = 3)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.22959, p-value = 1.392e-09
## alternative hypothesis: two-sided
```

Korzystając z testu Kołmogorowa-Smirnowa zweryfikuj hipotezę, że dane pochodzą z rozkładu normalnego o parametrach wyestymowanych z próby:

```
ks.test(x, 'pnorm', mean = mean(x), sd = sd(x))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.03501, p-value = 0.967
## alternative hypothesis: two-sided
```

Za pomocą testu Shapiro-Wilka zweryfikuj hipotezę, że dane pochodzą z rozkładu normalnego:

```
shapiro.test(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.9944, p-value = 0.6611
```

Przeprowadź testy dot. normalności za pomocą innych testów:

```
library(moments)
anscombe.test(x)
```

```
##
##  Anscombe-Glynn kurtosis test
##
```

```
## data:  x
## kurt = 3.11341, z = 0.56588, p-value = 0.5715
## alternative hypothesis: kurtosis is not equal to 3
```

```r
agostino.test(x)
```

```
##
##  D'Agostino skewness test
##
## data:  x
## skew = 0.041122, z = 0.245523, p-value = 0.8061
## alternative hypothesis: data have a skewness
```

```r
jarque.test(x)
```

```
##
##  Jarque-Bera Normality Test
##
## data:  x
## JB = 0.16356, p-value = 0.9215
## alternative hypothesis: greater
```

Dane o rozkładzie jednostajnym

Wygeneruj 200 liczb będących próbą losową z rozkładu jednostajnego w przedziale od 1 do 4:

```r
n = 200; a = 1; b = 4
x = runif(n, min = a, max = b)
```

Zweryfikuj hipotezę, że dane pochodzą z rozkładu normalnego korzystając z testu Kołmogorowa-Smirnowa:

```r
ks.test(x, 'pnorm', mean = mean(x), sd = sd(x))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.072905, p-value = 0.2382
## alternative hypothesis: two-sided
```

Zweryfikuj hipotezę, że dane pochodzą z rozkładu normalnego korzystając z testu Shapiro-Wilka:

```r
shapiro.test(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.95038, p-value = 2.054e-06
```

Przeprowadź testy dot. normalności za pomocą innych testów:

```r
anscombe.test(x)
```

```
##
##  Anscombe-Glynn kurtosis test
##
## data:  x
## kurt = 1.8198, z = -9.0931, p-value < 2.2e-16
## alternative hypothesis: kurtosis is not equal to 3
```

```r
agostino.test(x)
```

```
##
##  D'Agostino skewness test
##
## data:  x
## skew = 0.11768, z = 0.70072, p-value = 0.4835
## alternative hypothesis: data have a skewness
```

```r
jarque.test(x)
```

```
##
##  Jarque-Bera Normality Test
##
## data:  x
## JB = 12.068, p-value = 0.002396
## alternative hypothesis: greater
```

Przykład 5 – porównywanie średnich

Test przy założeniu równości wariancji w obu próbach

Wygeneruj dwie próby z rozkładów normalnych N(0,1) oraz N(1,1) o licznościach 9 i 12:

```r
n_1 = 9; mi_1 = 0; sigma_1 = 1
n_2 = 12; mi_2 = 1; sigma_2 = 1
x_1 = rnorm(n_1, mean = mi_1, sd = sigma_1)
x_2 = rnorm(n_2, mean = mi_2, sd = sigma_2)
```

Przy poziomie istotności $\alpha = 0.05$ zweryfikuj hipotezę o równości średnich przy założeniu tej samej wariancji:

```r
mean_1 = mean(x_1); mean_2 = mean(x_2)
s2_1 = var(x_1); s2_2 = var(x_2)
s2 = ((n_1 -1)*s2_1 + (n_2 -1)*s2_2)/(n_1 + n_2 - 2)
alfa = 0.05; c = qt(1 - alfa/2, n_1 + n_2 - 2)
T = abs(mean_1 - mean_2) / ( sqrt(s2*( n_1^(-1) + n_2^(-1) ) ) )
p_val = 2*(1 - pt(T, n_1 + n_2 - 2))
```

Wartość statystyki T = 3.6452.

Wartość krytyczna dla poziomu istotności $\alpha = 0.05$ wynosi c = 2.093.

p-wartość = 0.0017.

Wykorzystanie funkcji pakietu R:

```r
t.test(x_1, x_2, var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  x_1 and x_2
## t = -3.6452, df = 19, p-value = 0.001722
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.4291399 -0.6570865
```

```
## sample estimates:
##  mean of x  mean of y
## -0.3839679  1.1591453
```

Powtórz testy dla innych wartości μ_2, np. $\mu_2 = 0, 0.5, 2$.

Test bez zakładania równości wariancji

Przeprowadź analogiczne testy bez zakładania równości wariancji

```
s2 = s2_1/n_1 + s2_2/n_2
d = (s2^2)/(((s2_1/n_1)^2)/(n_1-1) + ((s2_2/n_2)^2)/((n_2-1)))
alfa = 0.05
c = qt(1 - alfa/2, d)
T = abs(mean_1 - mean_2)/sqrt(s2)
p_val = 2*(1 - pt(T, d))
```

Wartość statystyki T = 3.3162.

Wartość krytyczna dla poziomu istotności $\alpha = 0.05$ wynosi c = 2.2064.

p-wartość = 0.0071.

Wykorzystanie funkcji pakietu R:

```
t.test(x_1, x_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  x_1 and x_2
## t = -3.3162, df = 10.781, p-value = 0.007059
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.5698141 -0.5164124
## sample estimates:
##  mean of x  mean of y
## -0.3839679  1.1591453
```

Test Manna-Whitneya(-Wilcoxona)

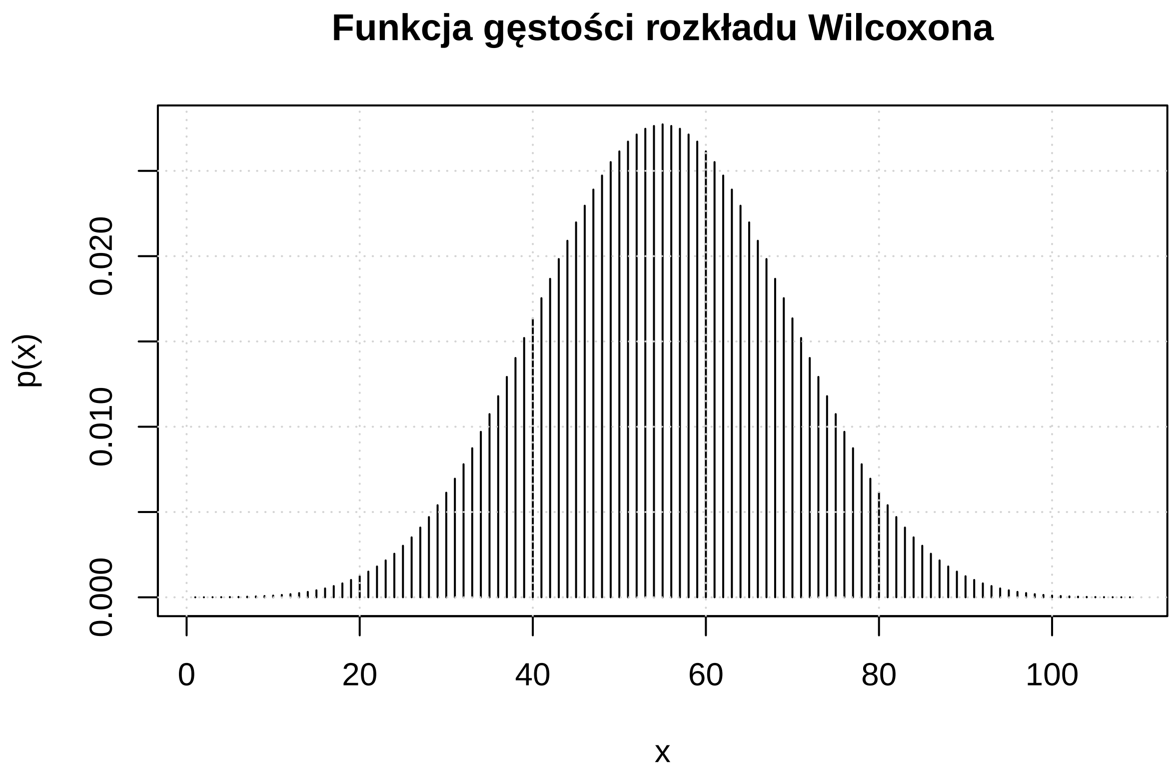
Przeprowadź analogiczne testy bez zakładania normalności rozkładów:

```
wilcox.test(x_1, x_2)
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  x_1 and x_2
## W = 12, p-value = 0.001803
## alternative hypothesis: true location shift is not equal to 0
```

Funkcja gęstości rozkładu Wilcoxona

```
plot(dwilcox(0:108, length(x_1), length(x_2)), type = 'h',
     xlab = 'x', ylab = 'p(x)', main = 'Funkcja gęstości rozkładu Wilcoxona');
grid()
```



Przykład 6 – test niezależności

Przeprowadź testy niezależności dla danych zawartych w tabeli kontyngencji xx:

```
x_1 = c(16, 25, 11)
x_2 = c(13, 32, 15)
x_3 = c(31, 43, 26)
xx = cbind(x_1, x_2, x_3)
I = 3
J = 3
n_i = x_1 + x_2 + x_3
n_j = c(sum(x_1), sum(x_2), sum(x_3))
N = sum(n_j)
```

Obliczenie wartości statystyki decyzyjnej, progu i p-wartości:

```
T = 0
for (i in 1:I) {
   for (j in 1:J) {
      T = T + (N*xx[i,j] - n_i[i]*n_j[j])^2/(N*n_i[i]*n_j[j])
   }
}
alfa = 0.05
c = qchisq(1 - alfa, df = (I - 1)*(J - 1))
p_val = 1 - pchisq(T, df = (I - 1)*(J - 1))
```

Wartość statystyki T = 2.4985.

Wartość krytyczna dla poziomu istotności $\alpha = 0.05$ wynosi c = 9.4877.

p-wartość = 0.6449.

Wykorzystanie funkcji chisq.test:

```
chisq.test(xx)
```

```
##
##  Pearson's Chi-squared test
##
## data:  xx
## X-squared = 2.4985, df = 4, p-value = 0.6449
```

Powtórz testy niezależności gdy ostatni element zmiennej x_1 zmienimy z 11 na 41:

```
x_1 = c(16, 25, 41)
x_2 = c(13, 32, 15)
x_3 = c(31, 43, 26)
xx = cbind(x_1, x_2, x_3)
I = 3
J = 3
n_i = x_1 + x_2 + x_3
n_j = c(sum(x_1), sum(x_2), sum(x_3))
N = sum(n_j)
```

Obliczenie wartości statystyki decyzyjnej, progu i p-wartości:

```
T = 0
for (i in 1:I) {
   for (j in 1:J) {
      T = T + (N*xx[i,j] - n_i[i]*n_j[j])^2/(N*n_i[i]*n_j[j])
   }
}
alfa = 0.05
c = qchisq(1 - alfa, df = (I - 1)*(J - 1))
p_val = 1 - pchisq(T, df = (I - 1)*(J - 1))
```

Wartość statystyki T = 16.7223.

Wartość krytyczna dla poziomu istotności $\alpha = 0.05$ wynosi c = 9.4877.

p-wartość = 0.0022.

Wykorzystanie funkcji chisq.test:

```
chisq.test(xx)
```

```
##
##  Pearson's Chi-squared test
##
## data:  xx
## X-squared = 16.722, df = 4, p-value = 0.002188
```
