# Lista 7

*Mikołaj Słupiński*

*20 maja 2018*

## Zadanie 1

### a)

```
z.025 = qnorm(0.975)
z.05 = qnorm(0.95)
z.025
z.05
```

```
## [1] 1.959964
## [1] 1.644854
```

Wartość krytyczna dla $\alpha = 0.05$ wynosi około 1.96, natomiast dla $\alpha = 0.1$ wynosi około 1.64.

### b)

Liczba powtórzzeń eksperrymentu powinna wynosić $N = \lceil \alpha(1-\alpha)(\frac{Z_{1-\alpha/2}}{0.1*\bar{\alpha}})^2 \rceil$

```
alfa = 0.05
n = 50
alfa.t = 0.05
z = qnorm(1-alfa/2)
sigma = 1
N = ceiling((alfa*(1-alfa)*(qnorm(1-alfa/2)/(0.1*alfa.t))^2))
A = matrix(rnorm(2*n*N), 2*n, N)
m = 500
mean(sapply(1:m, function(i) abs(alfa-mean(apply(matrix(rnorm(2*n*N, sd = 1),2*n,N), 2, function(v) abs
alfa = 0.05
alfa.t = 0.1
z = qnorm(1-alfa/2)
N = ceiling((alfa*(1-alfa)*(qnorm(1-alfa/2)/(0.1*alfa.t))^2))
A = matrix(rnorm(2*n*N), 2*n, N)
mean(sapply(1:m, function(i) abs(alfa-mean(apply(matrix(rnorm(2*n*N, sd = 1),2*n,N), 2, function(v) abs
```

```
## [1] 0.956
## [1] 0.956
```

Jak widać otrzymujemy frakcje zbliżoną, do teoretycznego 0.95 co sugeruje, że nasze obliczenia są poprawne.

### c)

```
alfa = 0.05
n = 50
alfa.t = 0.05
z = qnorm(1-alfa/2)
```

```
sigma = 1
N = ceiling((alfa*(1-alfa)*(qnorm(1-alfa/2)/(0.1*alfa.t))^2))
A = matrix(rexp(2*n*N), 2*n, N)
m = 500
mean(sapply(1:m, function(i) abs(alfa-mean(apply(matrix(rexp(2*n*N),2*n,N), 2, function(v) abs(mean(v[1
alfa = 0.05
alfa.t = 0.1
z = qnorm(1-alfa/2)
N = ceiling((alfa*(1-alfa)*(qnorm(1-alfa/2)/(0.1*alfa.t))^2))
A = matrix(rexp(2*n*N), 2*n, N)
mean(sapply(1:m, function(i) abs(alfa-mean(apply(matrix(rexp(2*n*N),2*n,N), 2, function(v) abs(mean(v[1
```

```
## [1] 0.944
## [1] 0.928
```

W tym wypadku równiez otrzymujemy frakcję zbliżoną do 0.95, pomimo że rozkład nie jest normalny, tylko wykładniczy. Sugeruje to, że estymacja rozkładem normalnym sprawdza się również dla innych rozkładów.

## Zadanie 2

### a)

```
N=1000
n1 = 5
n2 = 10
sigma1 = 1
sigma2 = 5
alfa = 0.05
alfa.t = 0.05
A = matrix(c(rnorm(n1*N, sd = sigma1), rnorm(n2*N, sd = sigma2)), n1+n2, N, byrow = TRUE)
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,(var(x)/n1+var(y)/n2)^2/((var(x)/n1)^2/
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,n1+n2-2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qnorm(1-alfa/2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
alfa = 0.1
```

```
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,(var(x)/n1+var(y)/n2)^2/((var(x)/n1)^2/
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,n1+n2-2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qnorm(1-alfa/2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
```

```
## [1] 0.049
## [1] 0.03562061 0.06237939
## [1] 0.053
## [1] 0.03911452 0.06688548
## [1] 0.076
## [1] 0.05957555 0.09242445
## [1] 0.101
## [1] 0.0823238 0.1196762
## [1] 0.11
## [1] 0.09060725 0.12939275
## [1] 0.14
## [1] 0.1184939 0.1615061
```

**b)**

```
N=1000
n1 = 10
n2 = 20
sigma1 = 1
sigma2 = 5
alfa = 0.05
alfa.t = 0.05
A = matrix(c(rnorm(n1*N, sd = sigma1), rnorm(n2*N, sd = sigma2)), n1+n2, N, byrow = TRUE)
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,(var(x)/n1+var(y)/n2)^2/((var(x)/n1)^2/
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
```

```r
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,n1+n2-2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qnorm(1-alfa/2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
alfa = 0.1
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,(var(x)/n1+var(y)/n2)^2/((var(x)/n1)^2/
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,n1+n2-2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qnorm(1-alfa/2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
```

```
## [1] 0.049
## [1] 0.03562061 0.06237939
## [1] 0.052
## [1] 0.03823888 0.06576112
## [1] 0.067
## [1] 0.05150376 0.08249624
## [1] 0.1
## [1] 0.08140615 0.11859385
## [1] 0.107
## [1] 0.08784131 0.12615869
## [1] 0.114
## [1] 0.09430222 0.13369778
```

### c)

```
N=1000
n1 = 20
n2 = 40
sigma1 = 1
sigma2 = 5
alfa = 0.05
alfa.t = 0.05
A = matrix(c(rnorm(n1*N, sd = sigma1), rnorm(n2*N, sd = sigma2)), n1+n2, N, byrow = TRUE)
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,(var(x)/n1+var(y)/n2)^2/((var(x)/n1)^2/
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,n1+n2-2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qnorm(1-alfa/2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
alfa = 0.1
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,(var(x)/n1+var(y)/n2)^2/((var(x)/n1)^2/
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,n1+n2-2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qnorm(1-alfa/2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
```

```
## [1] 0.038
## [1] 0.02614975 0.04985025
## [1] 0.04
## [1] 0.02785455 0.05214545
## [1] 0.046
## [1] 0.03301622 0.05898378
## [1] 0.088
## [1] 0.07044152 0.10555848
## [1] 0.092
## [1] 0.07408632 0.10991368
## [1] 0.096
## [1] 0.07774138 0.11425862
```

### d)

```
N=1000
n1 = 40
n2 = 80
sigma1 = 1
sigma2 = 5
alfa = 0.05
alfa.t = 0.05
A = matrix(c(rnorm(n1*N, sd = sigma1), rnorm(n2*N, sd = sigma2)), n1+n2, N, byrow = TRUE)
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,(var(x)/n1+var(y)/n2)^2/((var(x)/n1)^2,
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,n1+n2-2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qnorm(1-alfa/2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
alfa.t = 0.1
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,(var(x)/n1+var(y)/n2)^2/((var(x)/n1)^2,
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
```

```r
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qt(1-alfa/2,n1+n2-2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
p = mean(apply(A, 2, function (v) {
  x = v[1:n1]
  y = v[(n1+1):(n1+n2)]
  abs(mean(x) - mean(y))/sqrt(var(x)/n1 + var(y)/n2)>qnorm(1-alfa/2)
}))
p
c(p - qnorm(1-alfa.t/2)*sqrt(p*(1-p)/N), p+qnorm(1-alfa.t/2) * sqrt(p*(1-p)/N))
```

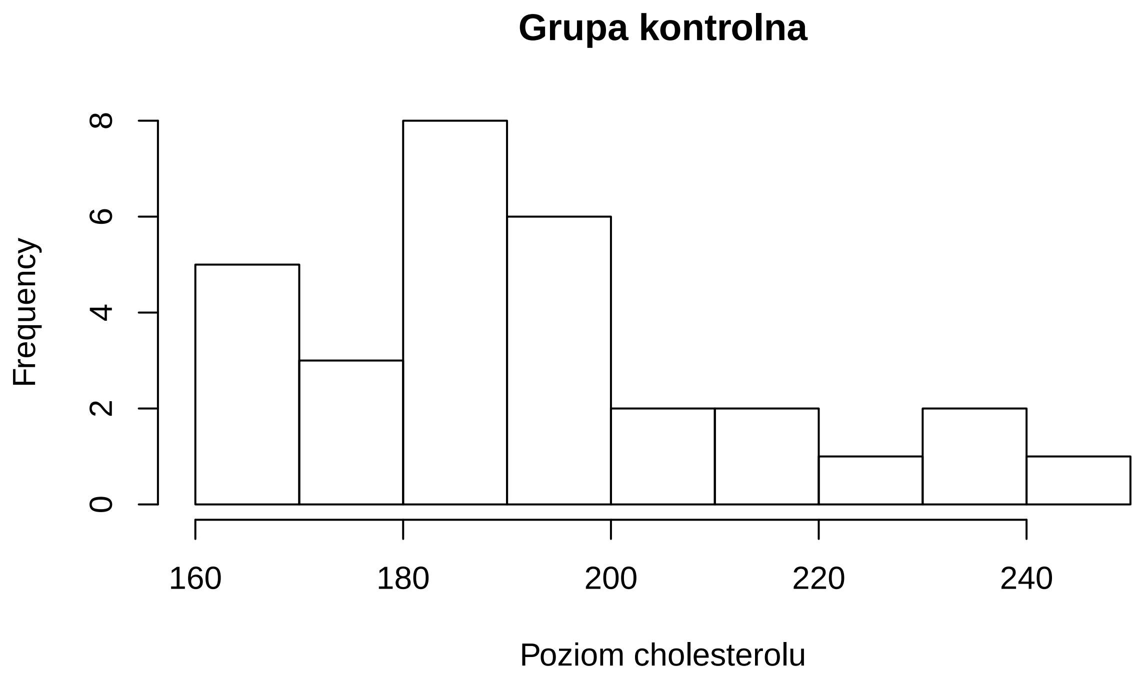
```
## [1] 0.05
## [1] 0.03649188 0.06350812
## [1] 0.052
## [1] 0.03823888 0.06576112
## [1] 0.054
## [1] 0.03999154 0.06800846
## [1] 0.05
## [1] 0.03866363 0.06133637
## [1] 0.052
## [1] 0.0404513 0.0635487
## [1] 0.054
## [1] 0.04224373 0.06575627
```

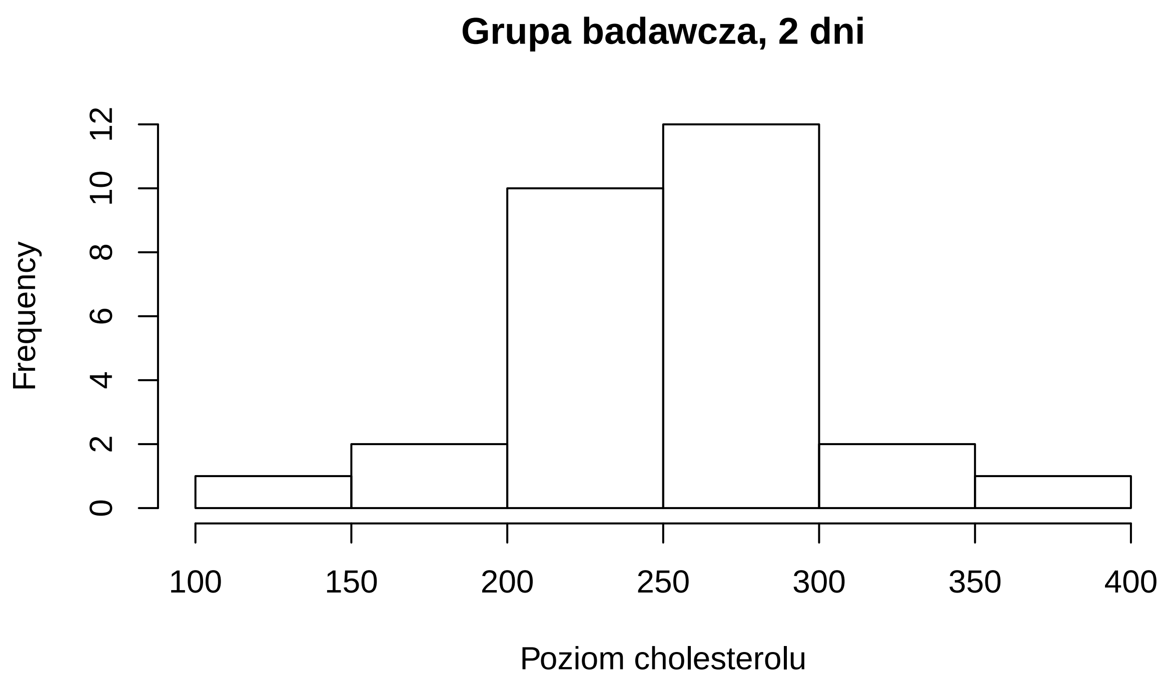
Ponownie otrzymaliśmy wyniki zbliżone do oczekiwanych.

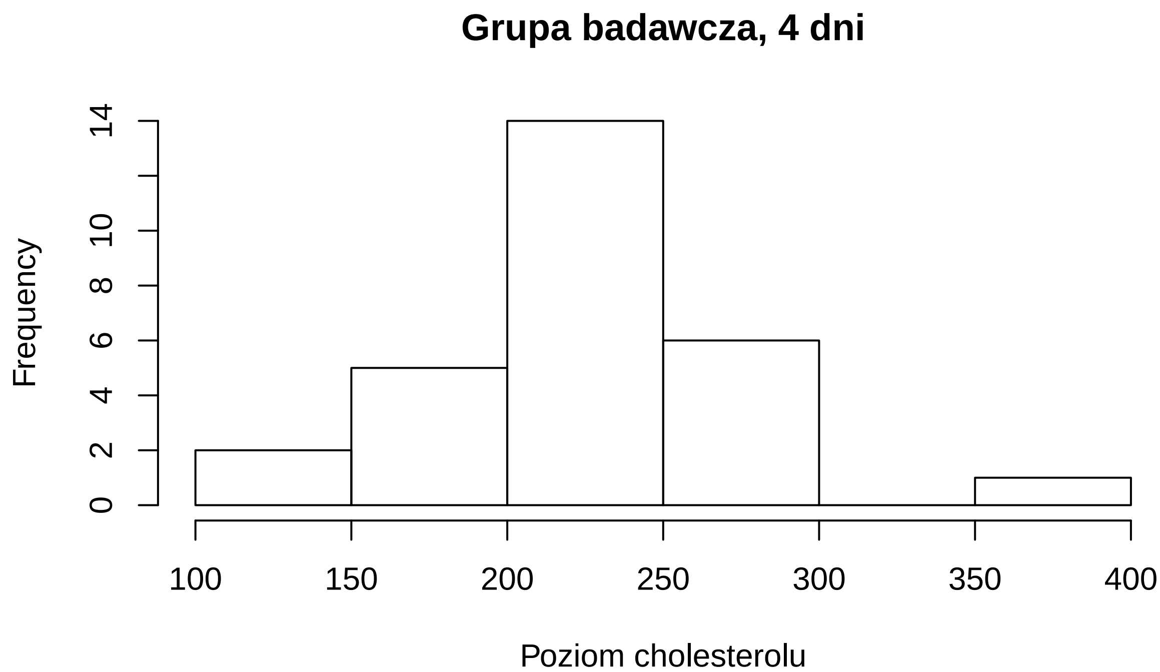
## Zadanie 3

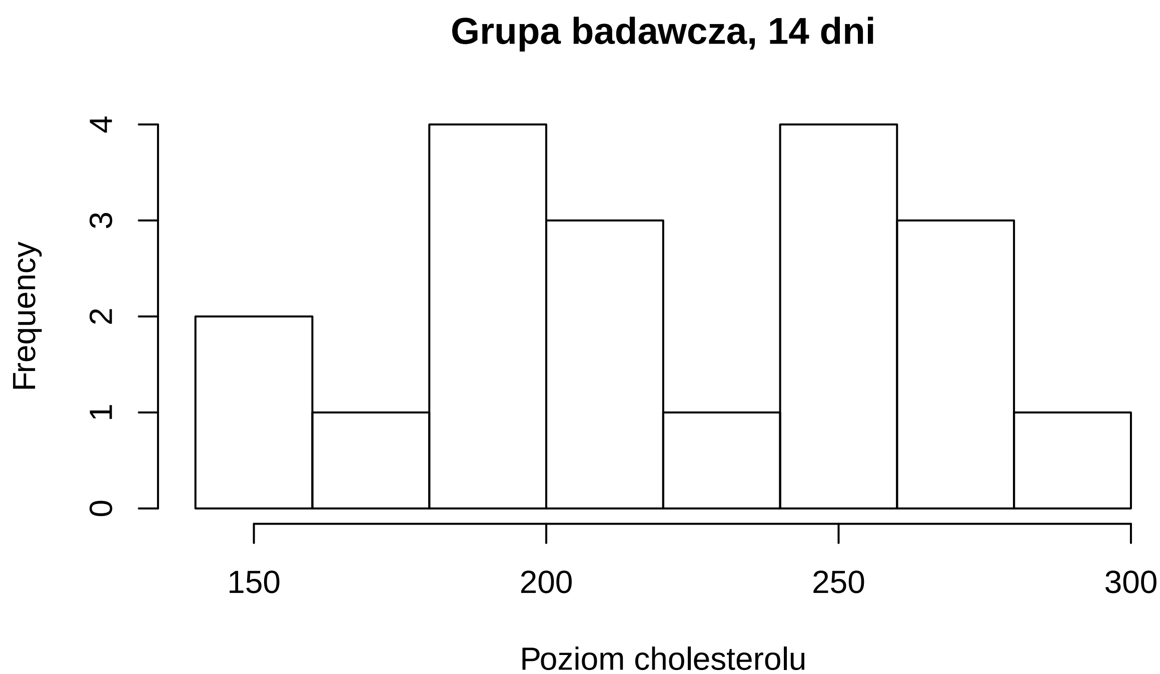
### a)

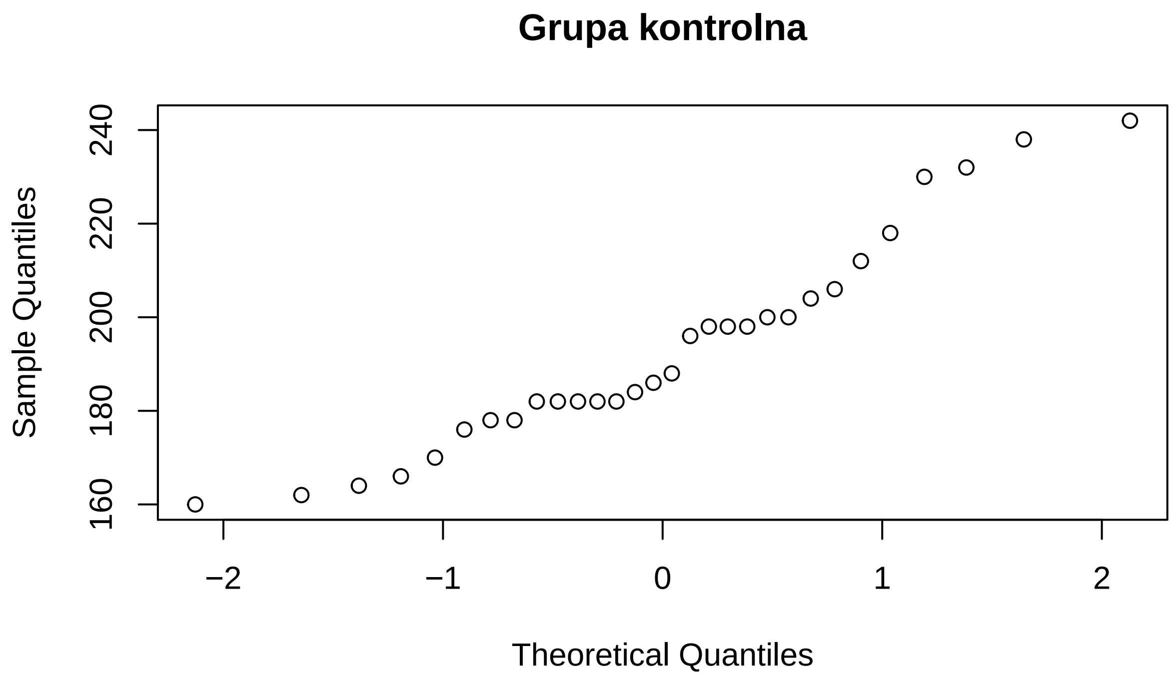
```r
library(stringr)
tab = read.table("http://www.math.uni.wroc.pl/~mbogdan/Podstawy/Dane/dane5.txt", skip = 5, fill = TRUE,
tab$V1 = as.integer(str_trim(tab$V1))
tab$V2 = as.integer(str_trim(tab$V2))
tab$V3 = as.integer(str_trim(tab$V3))
tab$V4 = as.integer(str_trim(tab$V4))
hist(tab[tab$V1 == 2, 2], main = "Grupa kontrolna", xlab = "Poziom cholesterolu")
hist(tab[tab$V1 == 1, 2], main = "Grupa badawcza, 2 dni", xlab = "Poziom cholesterolu")
hist(tab[tab$V1 == 1, 3], main = "Grupa badawcza, 4 dni", xlab = "Poziom cholesterolu")
hist(tab[tab$V1 == 1, 4], main = "Grupa badawcza, 14 dni", xlab = "Poziom cholesterolu")
qqnorm(tab[tab$V1 == 2, 2], main = "Grupa kontrolna")
qqnorm(tab[tab$V1 == 1, 2], main = "Grupa badawcza, 2 dni")
qqnorm(tab[tab$V1 == 1, 3], main = "Grupa badawcza, 4 dni")
qqnorm(tab[tab$V1 == 1, 4], main = "Grupa badawcza, 14 dni")
```

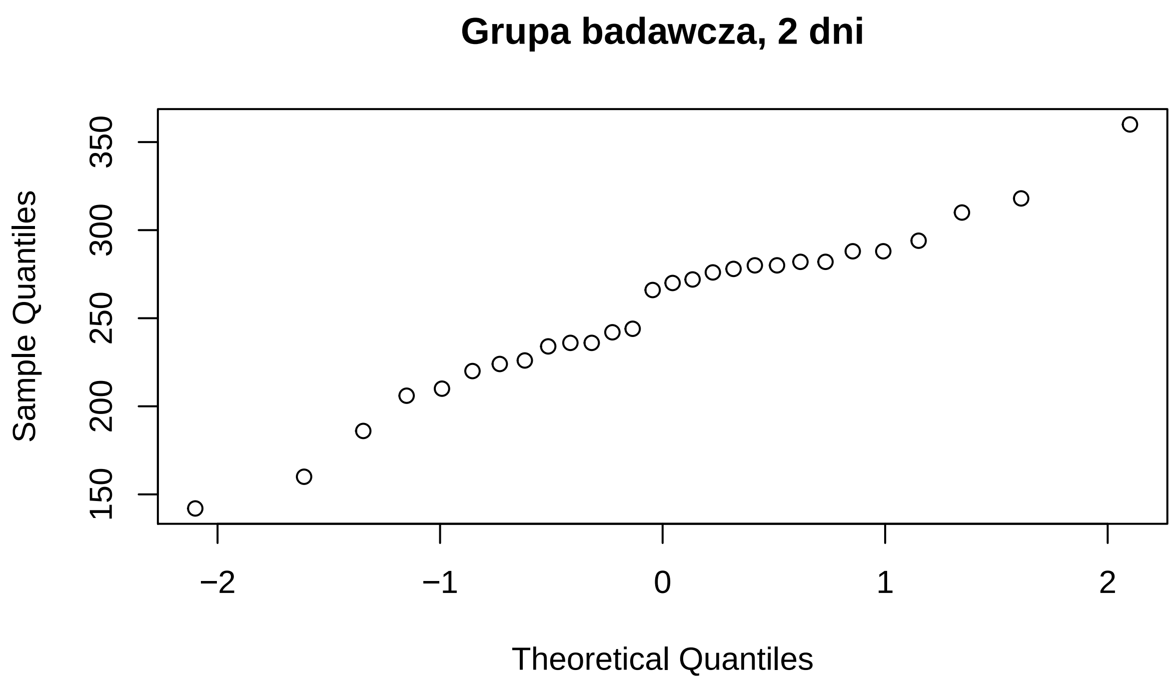

Grupa kontrolna
Frequency
0
2
4
6
8
160
180
200
220
240
Poziom cholesterolu

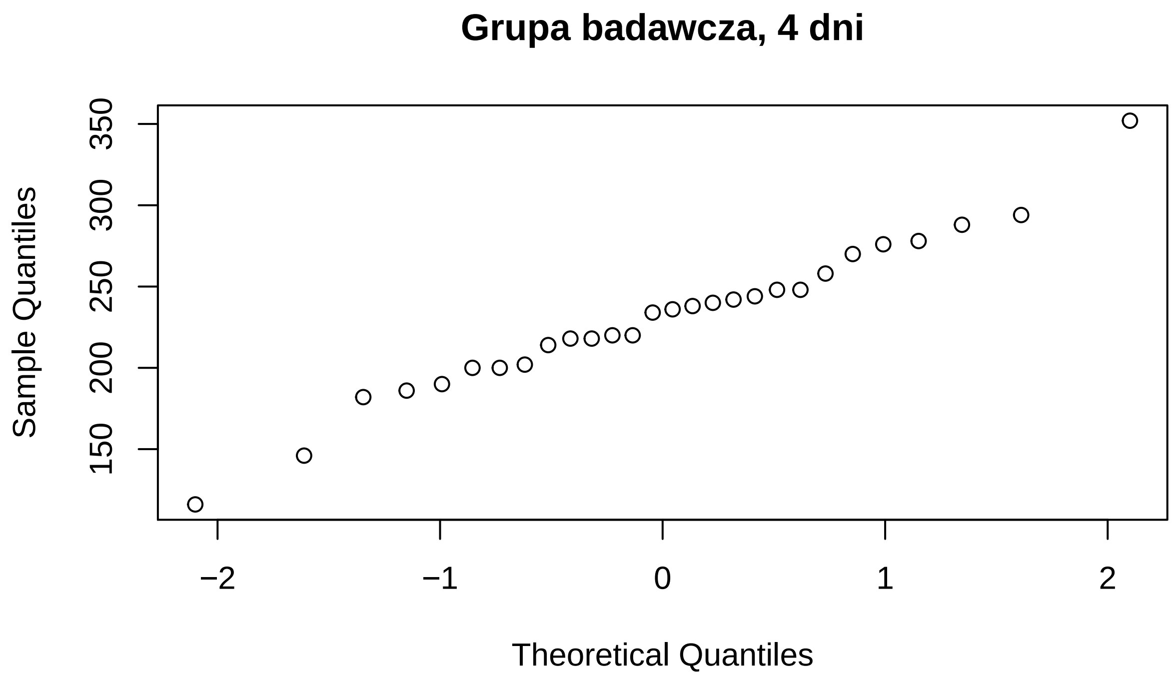

Grupa badawcza, 2 dni
Frequency
0
2
4
6
8
10
12
100
150
200
250
300
350
400
Poziom cholesterolu

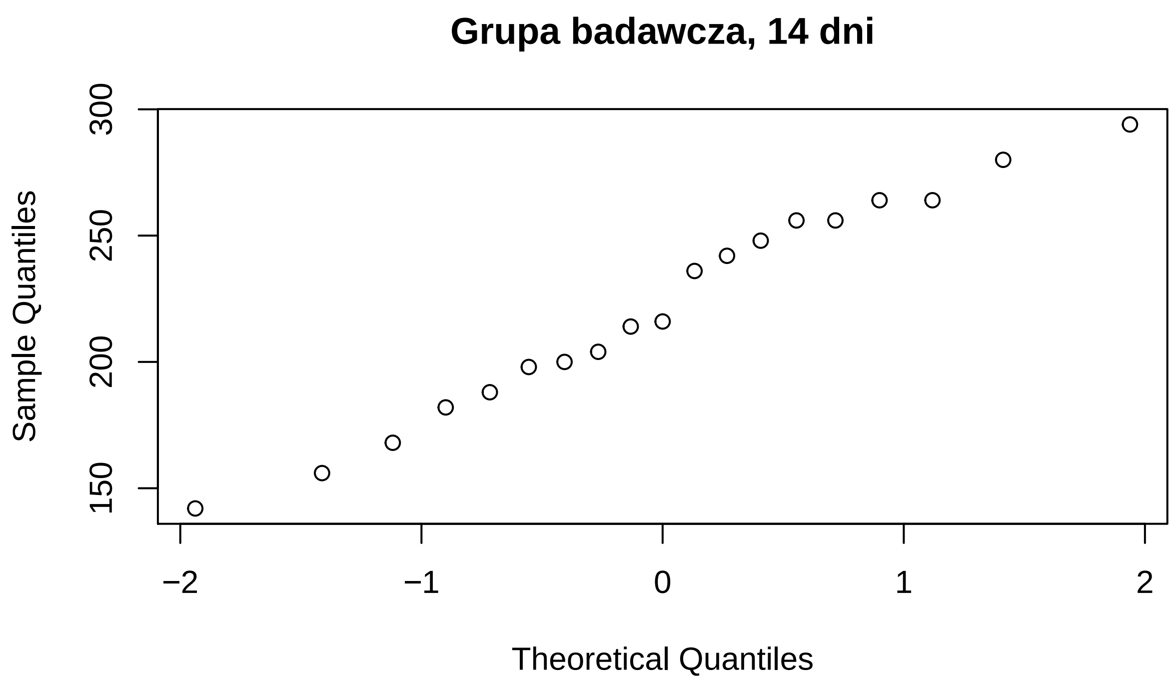

Grupa badawcza, 4 dni
Frequency
0
2
4
6
8
10
14
100
150
200
250
300
350
400
Poziom cholesterolu


Grupa badawcza, 14 dni
Frequency
0
1
2
3
4
150
200
250
300
Poziom cholesterolu


Grupa kontrolna
Sample Quantiles
240
220
200
180
160
−2
−1
0
1
2
Theoretical Quantiles


Grupa badawcza, 2 dni
Sample Quantiles
350
300
250
200
150
−2
−1
0
1
2
Theoretical Quantiles

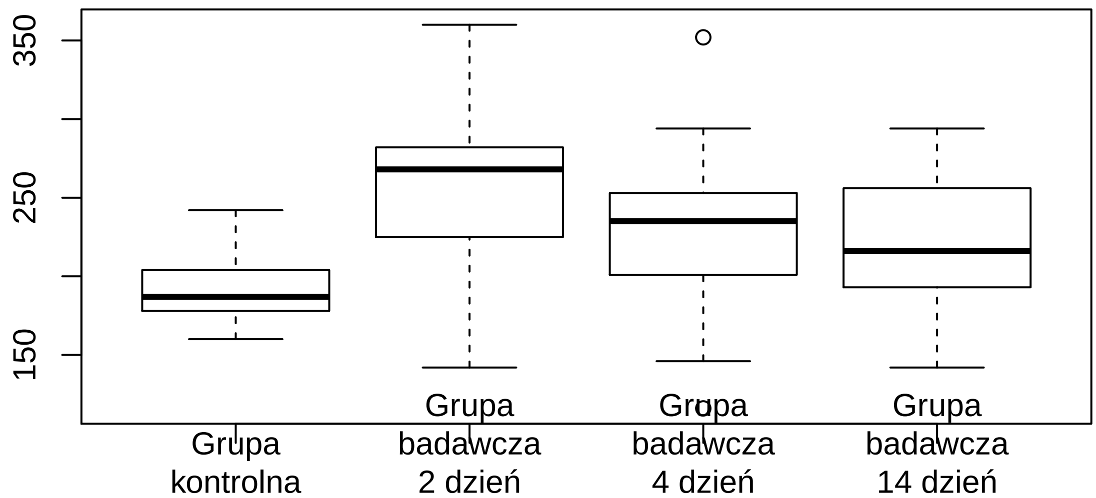



Możemy założyć, że poziom cholestorolu w grupue badawczej 2 dni po zawale ma rozkład normalny, podobne założenie możemy zrobić o grupie kontrolnej.

**b)**

```
boxplot(tab[tab$V1 == 2, 2], tab[tab$V1 == 1, 2], tab[tab$V1 == 1, 3], tab[tab$V1 == 1, 4], names = c("(
```




Widzmy tutaj, że cholesterol wydaje się niższy u grupy kontrolnej, niż u pacjentów po zawale.

```
X = tab[tab$V1 == 2, 2]
Y = tab[tab$V1 == 1, 2]
n1 = length(X)
n2 = length(Y)
X_avg = mean(X)
Y_avg = mean(Y)
s1_sq = var(X)
s2_sq = var(Y)
se1 = sqrt(s1_sq)/sqrt(n1)
se2 = sqrt(s1_sq)/sqrt(n2)
alfa = 0.05
t = (X_avg - Y_avg)/sqrt(s1_sq/n1 + s2_sq/n2)
df = (se1^2 + se2^2)^2/(se1^4/(n1-1) + se2^4/(n2-1))
ta = qt(1 - alfa, df)
t < -ta
```

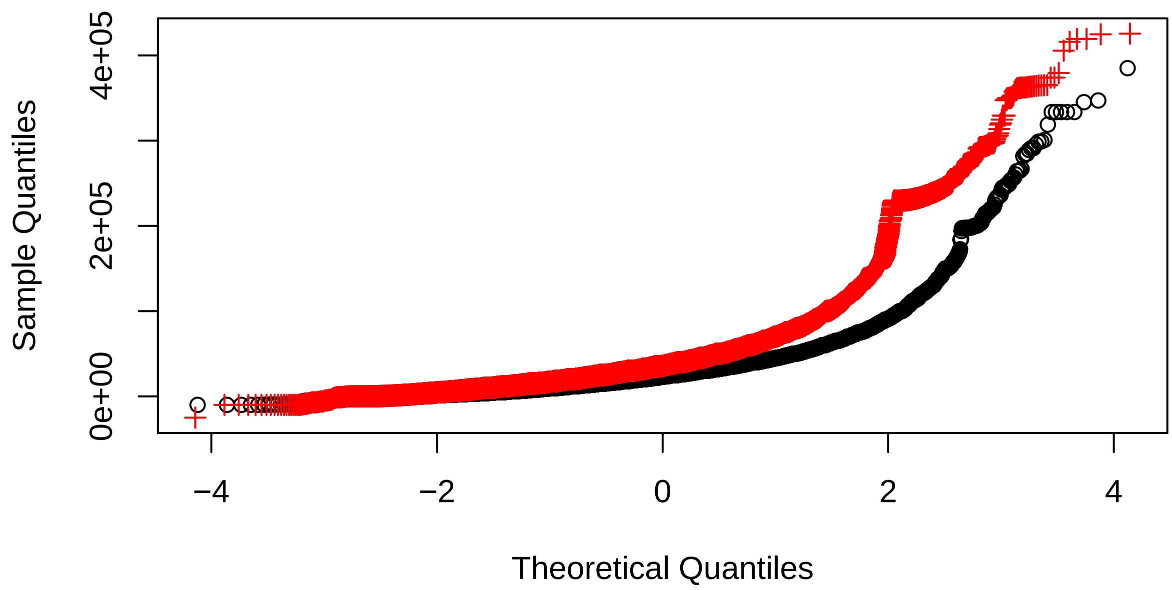
```
## [1] TRUE
```

Testujemy tutaj, czy średnia poziomu cholesterolu u grupy kontrolnej jest istotnie niższa, niż u grupy badawczej 2 dni po zawale. Test wykonano na poziomie istotności $\alpha = 0.05$. Przy zadanym poziomie istotności, odrzucamy hipotezę o równości średnich, na rzecz alternatywy, czyli średni poziom cholesterolu u grupy kontolnej jest niższy

### Zadanie 4

#### a)

```
tab = read.table("http://www.math.uni.wroc.pl/~mbogdan/Podstawy/Dane/individuals.dat")
q1 <- qqnorm(tab[tab[4] == 2, 5], plot.it = FALSE)
q2 <- qqnorm(tab[tab[4] == 1, 5], plot.it = FALSE)
plot(range(q1$x, q2$x), range(q1$y, q2$y), type = "n", ylab = "Sample Quantiles", xlab = "Theoretical Qu
points(q1)
points(q2, col = "red", pch = 3)
```

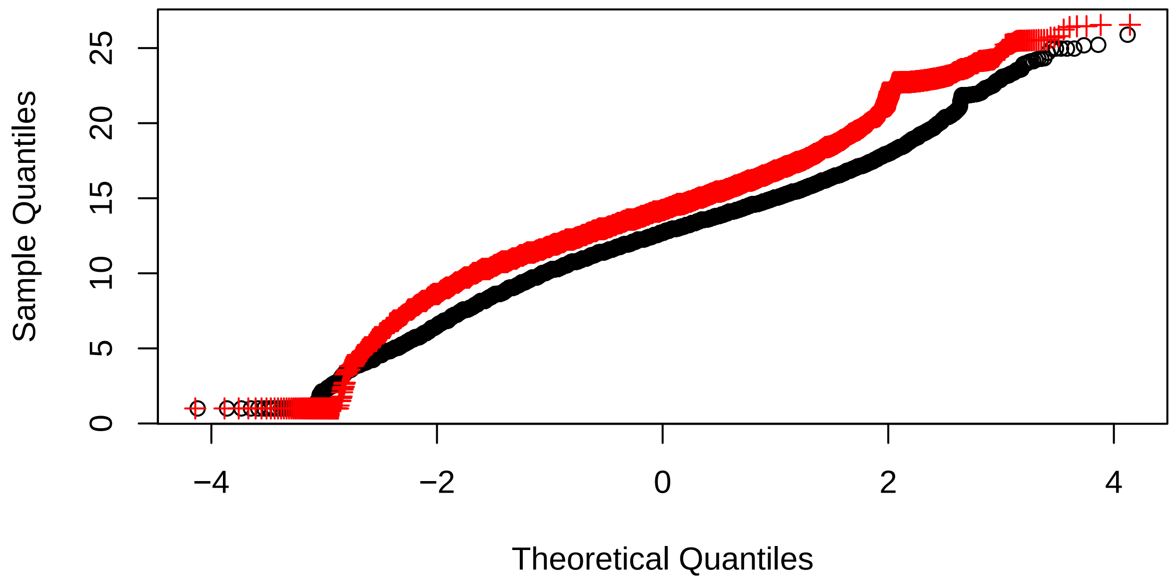



Na pierwszy rzut oka widać, że rozkłady nie są normalne.

#### b)

```
transincome = tab[tab[5] >= 0, 5]^0.253
transincome.genre = tab[tab[5] >= 0, 4]
transincome.sector = tab[tab[5] >= 0, 6]
transincome.education = tab[tab[5] >= 0, 3]
q1 <- qqnorm(transincome[transincome.genre == 2], plot.it = FALSE)
q2 <- qqnorm(transincome[transincome.genre == 1], plot.it = FALSE)

plot(range(q1$x, q2$x), range(q1$y, q2$y), type = "n", ylab = "Sample Quantiles", xlab = "Theoretical Qu
points(q1)
points(q2, col = "red", pch = 3)
```




Rozkłady zbliżyły się do rozkładu normalnego.

**c)**

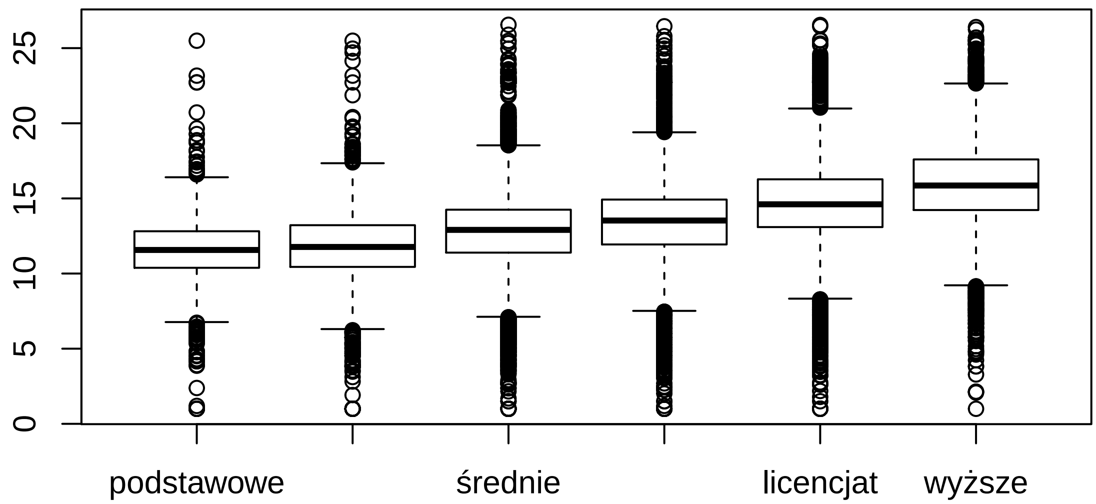
```
X = transincome[transincome.genre == 2]
Y = transincome[transincome.genre == 1]
n1 = length(X)
n2 = length(Y)
X_avg = mean(X)
Y_avg = mean(Y)
s1_sq = var(X)
s2_sq = var(Y)
se1 = sqrt(s1_sq)/sqrt(n1)
se2 = sqrt(s1_sq)/sqrt(n2)
alfa = 0.05
t = (X_avg - Y_avg)/sqrt(s1_sq/n1 + s2_sq/n2)
df = (se1^2 + se2^2)^2/(se1^4/(n1-1) + se2^4/(n2-1))
ta = qt(1 - alfa, df)
t < -ta
```

```
## [1] TRUE
```

W teście t-studenta na poziomie istotności $\alpha = 0.05$ otrzymaliśmy, że średnia zarobków kobiet jest istotnie niższa od średniej zarobków mężczyzn.

**d)**

```
boxplot(transincome[transincome.education == 1], transincome[transincome.education == 2], transincome[t
```




Patrząc na wykres można nabrać wrażenia, że osoby z wykształceniem wyższym zarabiają istotnie więcej od osób z licencjatem.

**e)**

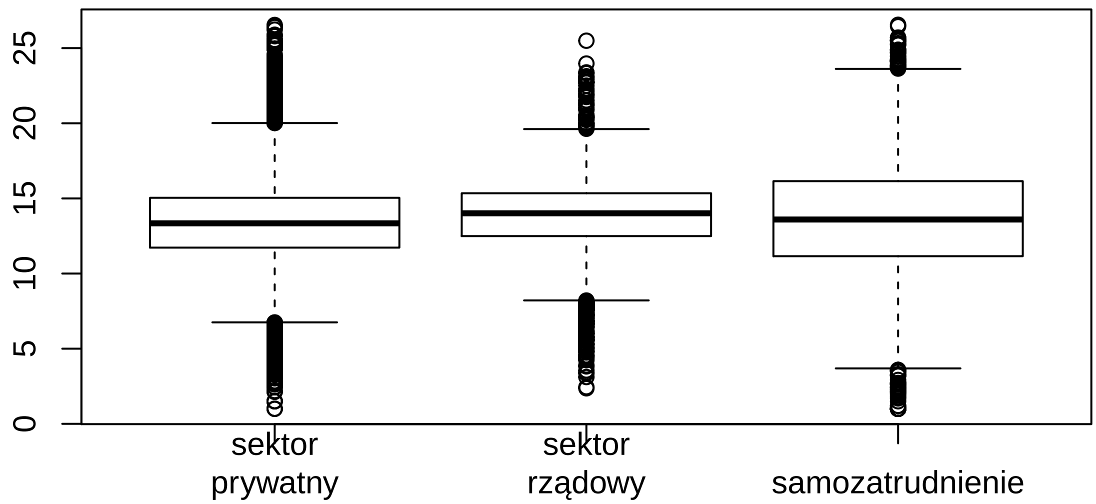
```
X = transincome[transincome.education == 6]
Y = transincome[transincome.education == 5]
n1 = length(X)
n2 = length(Y)
X_avg = mean(X)
Y_avg = mean(Y)
s1_sq = var(X)
s2_sq = var(Y)
se1 = sqrt(s1_sq)/sqrt(n1)
se2 = sqrt(s1_sq)/sqrt(n2)
alfa = 0.05
t = (X_avg - Y_avg)/sqrt(s1_sq/n1 + s2_sq/n2)
df = (se1^2 + se2^2)^2/(se1^4/(n1-1) + se2^4/(n2-1))
ta = qt(1 - alfa, df)
t > ta
```

```
## [1] TRUE
```

Również w tym wypadku test t-studenta wykonany na poziomie istotności $\alpha = 0.05$ wskazuje na to, że średnia zarobków osób z licencjatem jest istotnie niższa od średniej zarobków osób z wykształceniem wyższym.

**f)**

```r
boxplot(transincome[transincome.sector == 5], transincome[transincome.sector == 6], transincome[transinc
```




Patrząc na wykres a się wrażenie, ze zarobki w sektorze rządowym i sektorze prywatnym zbyt mocno od siebie nie odbiegają.

**g)**

```r
X = transincome[transincome.sector == 5]
Y = transincome[transincome.sector == 6]
n1 = length(X)
n2 = length(Y)
X_avg = mean(X)
Y_avg = mean(Y)
s1_sq = var(X)
s2_sq = var(Y)
se1 = sqrt(s1_sq)/sqrt(n1)
se2 = sqrt(s1_sq)/sqrt(n2)
alfa = 0.05
t = abs(X_avg - Y_avg)/sqrt(s1_sq/n1 + s2_sq/n2)
df = (se1^2 + se2^2)^2/(se1^4/(n1-1) + se2^4/(n2-1))
ta = qt(1 - alfa/2, df)
t < -ta | t > ta
```

```
## [1] TRUE
```

I w tym wypadku test t-studenta przeprowadzony na poziomie istotności $\alpha = 0.05$ wskazuje na to, że różnica zarobków jest istotna.