Raport 7

Aleksander Milach

20 May 2018

Zadanie 1

```
k05=qnorm(1-.05/2)
k10=qnorm(1-.1/2)
```

Wartości krytyczne dla Z wynoszą 1.959964 dla α=0,05 oraz 1.6448536 dla α=0,1.

Z obliczeń wykonanych na laboratorium szukane N w zależności od α wyraza sie wzorem $(\frac{\Phi^{-1}(\frac{1,95}{2})\sqrt{\alpha(1-\alpha)}}{0,1\alpha})^2$.

```
N05=((qnorm(1.95/2)*sqrt(.05*.95))/.1/.05)^2
N10=((qnorm(1.95/2)*sqrt(.1*.9))/.1/.1)^2
#n=50, n=100 skopiowac i zmienic na poczatku n=100
n=50
a=.1
v10=numeric(1000)
for(j in 1:1000)
{
  licz=0
  for(i in 1:N10)
  {
    p1=rexp(n)
    p2=rexp(n)
    z=abs(mean(p1)-mean(p2))*sqrt(n/2)
    if(z>k10){licz<-licz+1}
  }
  v10[j]<-licz/N10
}
mi501=mean(v10)
p501=length(v10[v10>1.1*a | v10<0.9*a])

n=100
a=.1
v11=numeric(1000)
for(j in 1:1000)
{
  licz=0
  for(i in 1:N10)
  {
    p1=rexp(n)
    p2=rexp(n)
    z=abs(mean(p1)-mean(p2))*sqrt(n/2)
    if(z>k10){licz<-licz+1}
  }
  v11[j]<-licz/N10
}
mi502=mean(v11)
p502=length(v11[v11>1.1*a | v11<0.9*a])
```

```
n=50
a=.05
v12=numeric(1000)
for(j in 1:1000)
{
  licz=0
  for(i in 1:N05)
  {
    p1=rexp(n)
    p2=rexp(n)
    z=abs(mean(p1)-mean(p2))*sqrt(n/2)
    if(z>k05){licz<-licz+1}
  }
  v12[j]<-licz/N05
}
mi503=mean(v12)
p503=length(v12[v12>1.1*a | v12<0.9*a])


n=100
a=.05
v13=numeric(1000)
for(j in 1:1000)
{
  licz=0
  for(i in 1:N05)
  {
    p1=rexp(n)
    p2=rexp(n)
    z=abs(mean(p1)-mean(p2))*sqrt(n/2)
    if(z>k05){licz<-licz+1}
  }
  v13[j]<-licz/N05
}
mi504=mean(v13)
p504=length(v13[v13>1.1*a | v13<0.9*a])

kable(matrix(c(mi501,p501,mi502,p502,mi503,p503,mi504,p504),ncol=4),col.names=c("n=50,a=.1",
                                                                                  "n=100,a=.1","n=50,a=.0
```

n=50,a=.1	n=100,a=.1	n=50,a=.05	n=100,a=.05
0.0997199	0.0997648	0.0504905	0.0501351
55.0000000	52.0000000	56.0000000	41.0000000

Dla obliczonych N05 i N10, tak jak chcieliśmy, błąd estymacji był większy niż 10%α dla około 50 symulacjach. Średnia wartość prawdopodobieństwa popełnienia błędu pierwszego rodzaju w naszych symulacjach jest bardzo bliska alfa.

Zadanie 2

```
ww1=c(.05,5,10)
ww2=c(.05,10,20)
ww3=c(.05,20,40)
ww4=c(.05,40,80)
ww5=c(.1,5,10)
ww6=c(.1,10,20)
ww7=c(.1,20,40)
ww8=c(.1,40,80)
N=matrix(numeric(16),ncol=2)
M=rbind(ww1,ww2,ww3,ww4,ww5,ww6,ww7,ww8)
for(e in 1:8)
{
  w1=numeric(100)
  for(j in 1:100)
  {
    licz=0
    for(i in 1:100)
    {
      B1=rnorm(M[e,][2],0,1)
      B2=rnorm(M[e,][3],0,5)
      if(abs(t.test(B1,B2)$statistic)>qt(1-(M[e,][1]/2),t.test(B1,B2)$parameter)){licz<-licz+1}
    }
    w1[j]<-licz/100
  }

  LPU1=mean(w1)-qt(1-M[e,][1]/2,99)*sd(w1)/sqrt(100)
  PPU1=mean(w1)+qt(1-M[e,][1]/2,99)*sd(w1)/sqrt(100)
  N[e,]=c(LPU1,PPU1)
}

O=matrix(numeric(16),ncol=2)
for(e in 1:8)
{
  w1=numeric(100)
  for(j in 1:100)
  {
    licz=0
    for(i in 1:100)
    {
      B1=rnorm(M[e,][2],0,1)
      B2=rnorm(M[e,][3],0,5)
      if(abs(t.test(B1,B2)$statistic)>qt(1-(M[e,][1]/2),(M[e,][2]+M[e,][3]-2))){licz<-licz+1}
    }
    w1[j]<-licz/100
  }

  LPU1=mean(w1)-qt(1-M[e,][1]/2,99)*sd(w1)/sqrt(100)
  PPU1=mean(w1)+qt(1-M[e,][1]/2,99)*sd(w1)/sqrt(100)
  O[e,]=c(LPU1,PPU1)
}
```

```r
P=matrix(numeric(16),ncol=2)
for(e in 1:8)
{
  w1=numeric(100)
  for(j in 1:100)
  {
    licz=0
    for(i in 1:100)
    {
      B1=rnorm(M[e,][2],0,1)
      B2=rnorm(M[e,][3],0,5)
      if(abs(t.test(B1,B2)$statistic)>qnorm(1-(M[e,][1]/2))){licz<-licz+1}
    }
    w1[j]<-licz/100
  }

  LPU1=mean(w1)-qt(1-M[e,][1]/2,99)*sd(w1)/sqrt(100)
  PPU1=mean(w1)+qt(1-M[e,][1]/2,99)*sd(w1)/sqrt(100)
  P[e,]=c(LPU1,PPU1)
}
Np=cbind(M,N)
Op=cbind(M,O)
Pp=cbind(M,P)

R=rbind(Np,Op,Pp)

kable(R,col.names=c("alfa","n1","n2","LPU","PPU"),format="markdown")
```

	alfa	n1	n2	LPU	PPU
ww1	0.05	5	10	0.0442223	0.0523777
ww2	0.05	10	20	0.0460913	0.0549087
ww3	0.05	20	40	0.0490769	0.0569231
ww4	0.05	40	80	0.0432931	0.0513069
ww5	0.10	5	10	0.0940638	0.1039362
ww6	0.10	10	20	0.0983058	0.1094942
ww7	0.10	20	40	0.0934173	0.1031827
ww8	0.10	40	80	0.0975649	0.1080351
ww1	0.05	5	10	0.0479434	0.0572566
ww2	0.05	10	20	0.0475676	0.0564324
ww3	0.05	20	40	0.0449771	0.0522229
ww4	0.05	40	80	0.0438824	0.0525176
ww5	0.10	5	10	0.0999046	0.1104954
ww6	0.10	10	20	0.0968732	0.1061268
ww7	0.10	20	40	0.0990757	0.1081243
ww8	0.10	40	80	0.1009384	0.1110616
ww1	0.05	5	10	0.0715680	0.0832320
ww2	0.05	10	20	0.0592326	0.0685674
ww3	0.05	20	40	0.0515960	0.0622040
ww4	0.05	40	80	0.0457673	0.0536327
ww5	0.10	5	10	0.1272630	0.1391370
ww6	0.10	10	20	0.1112474	0.1217526
ww7	0.10	20	40	0.0983170	0.1092830
ww8	0.10	40	80	0.1000688	0.1111312

Biorąc kwantyl z rozkładu normalnego prawdopodobieństwo błędu pierwszego rodzaju jest wyższe, szczególnie dla małych prób. Wynika to ze znaczących różnic między kwantylami rozkładu studenta z małą ilośćią stopni swobody, a kwantylami rozkłądu normalnego.

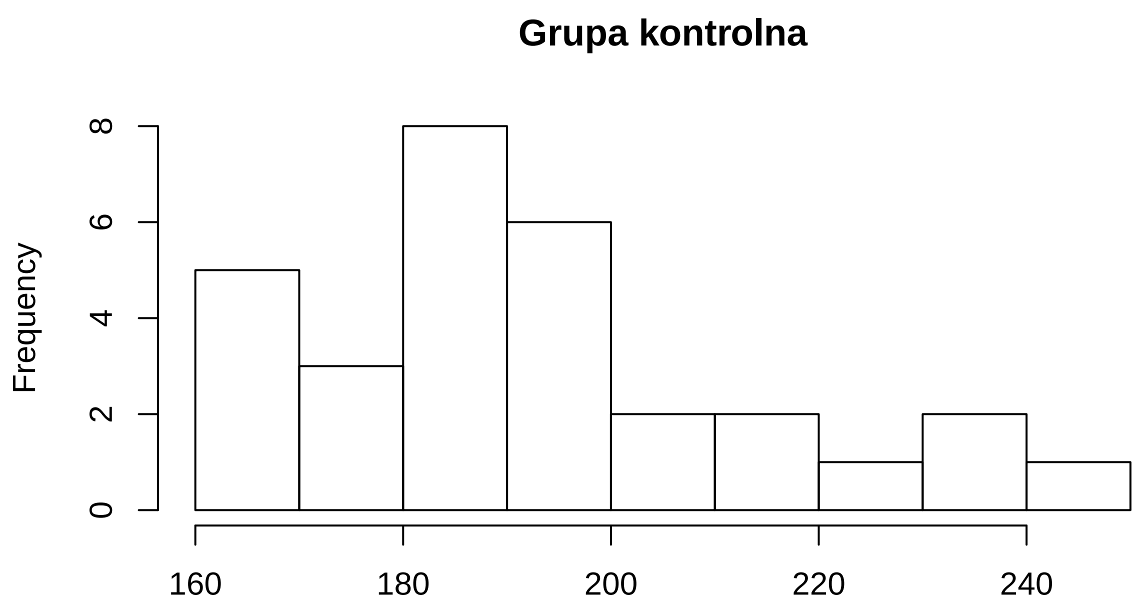
Zadanie 3

```
dane55=read.table("dane55.txt",fill=TRUE)
GK=dane55[[2]][29:58]
D2=dane55[[2]][1:28]
D4=dane55[[3]][1:28]
#D14=dane55[[4]][c(1,3,6,7,8,9,11,12,14,15,17,19,20,21,22,24,26,27,28)]
D14=c(156, 242, 256, 142, 216, 248, 168, 236, 200, 264, 264, 188, 182, 294, 214, 198, 256, 280, 204)
```

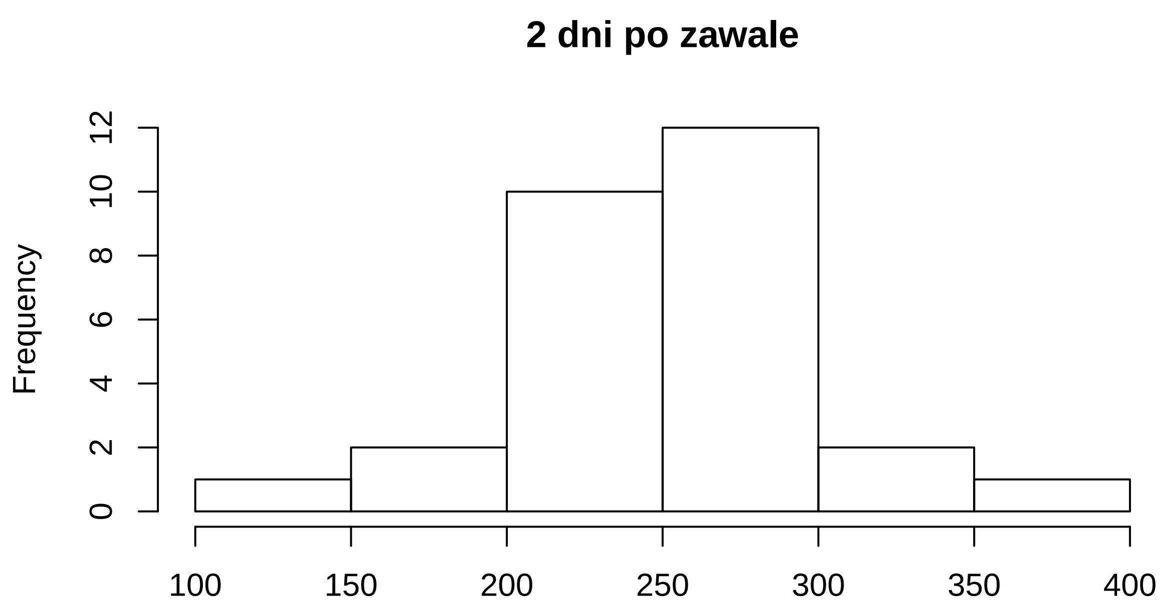
Po wyizolowaniu danych z pliku txt i umieszczeniu ich w wektorach, zadanie staje sie znacząco prostsze.

Podpunkt a

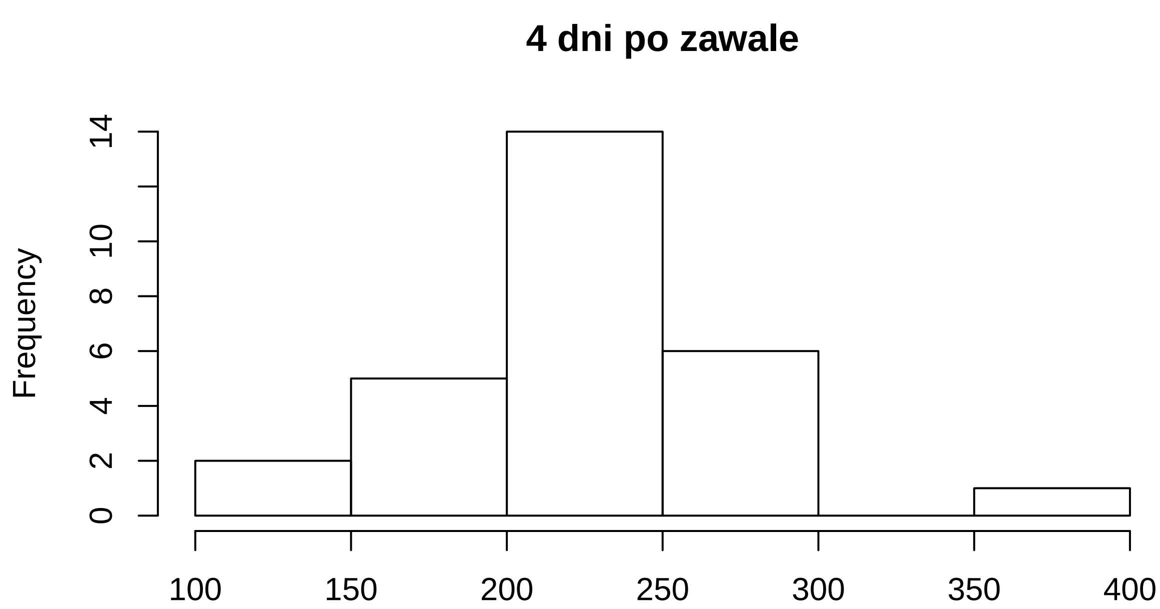
```
hist(GK,main="Grupa kontrolna",xlab="")
```



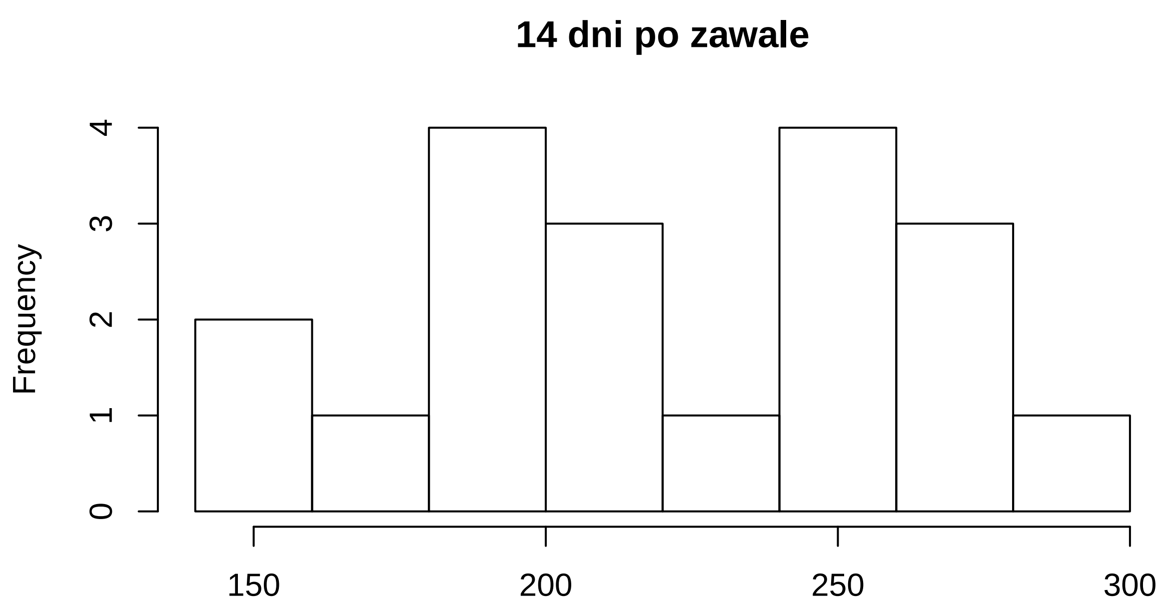
```r
hist(D2,main="2 dni po zawale",xlab="")
```



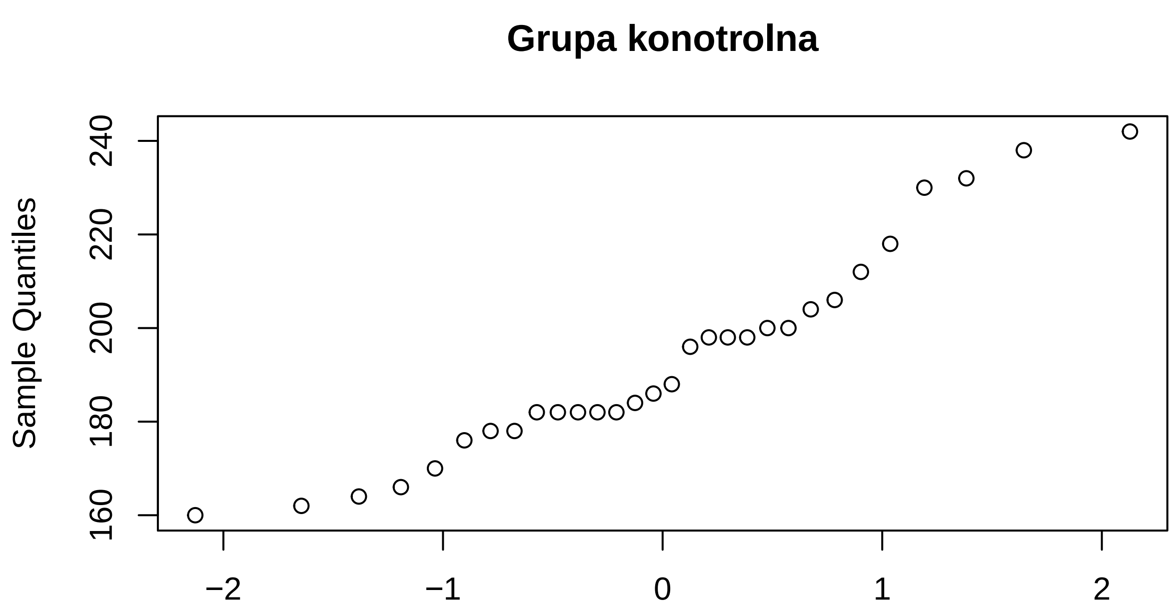

```r
hist(D4,main="4 dni po zawale",xlab="")
```



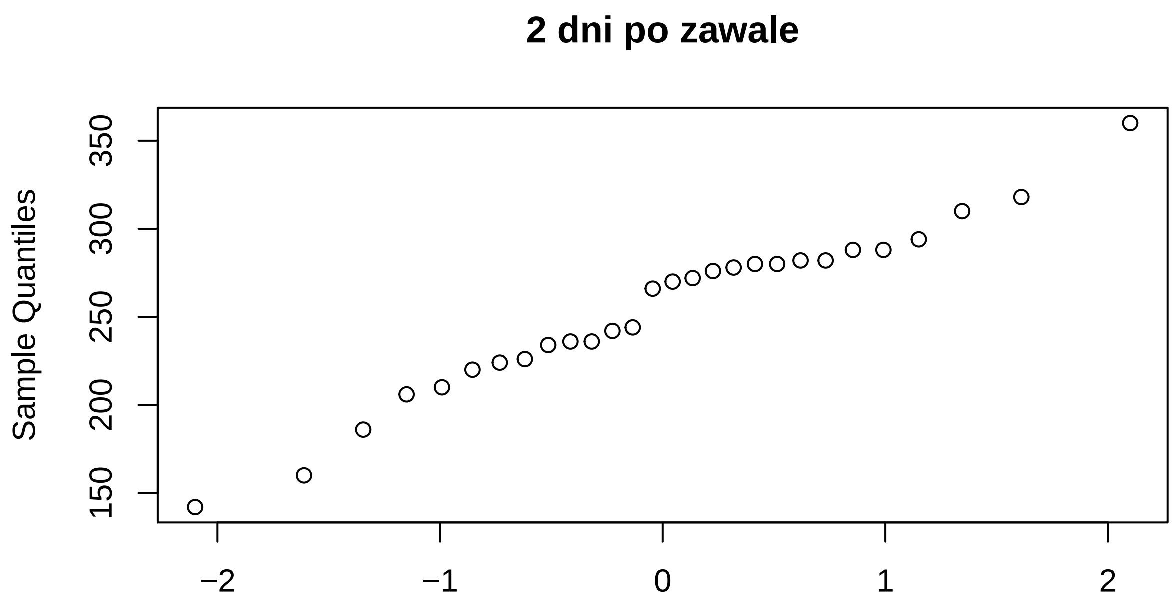
```r
hist(as.numeric(D14),main="14 dni po zawale",xlab="")
```



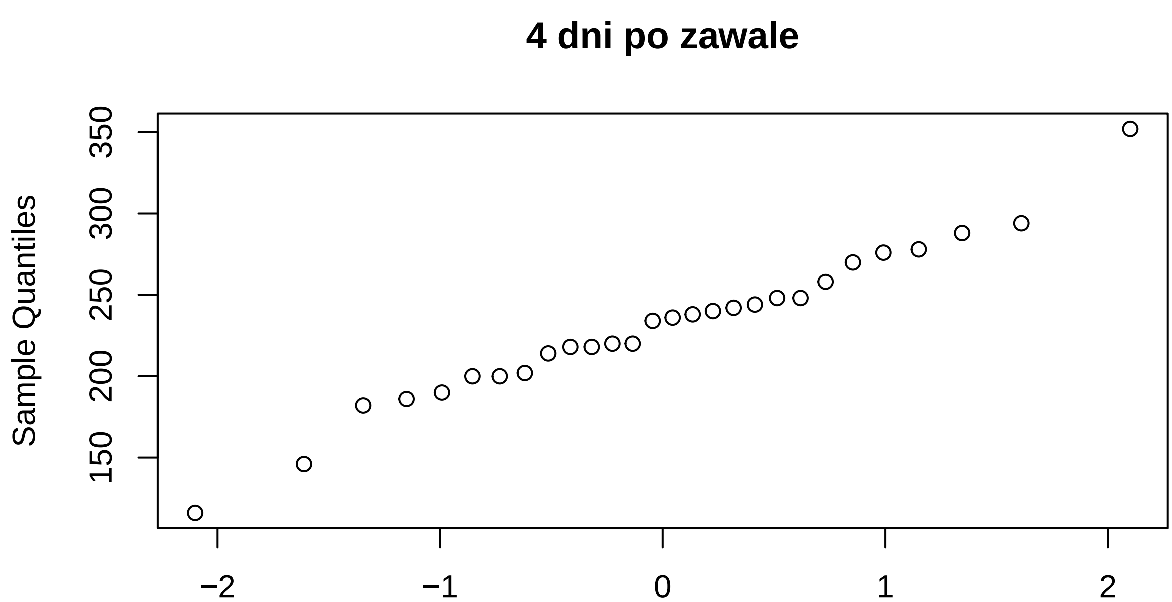

```r
qqnorm(GK,main="Grupa konotrolna",xlab="")
```



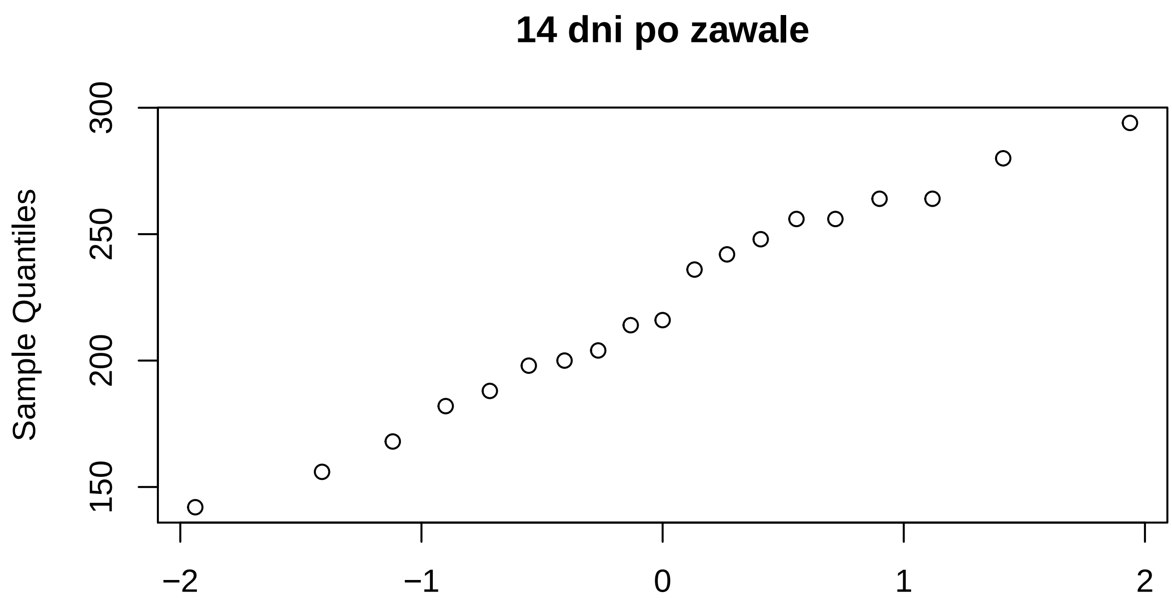
```r
qqnorm(D2,main="2 dni po zawale",xlab="")
```




```r
qqnorm(D4,main="4 dni po zawale",xlab="")
```

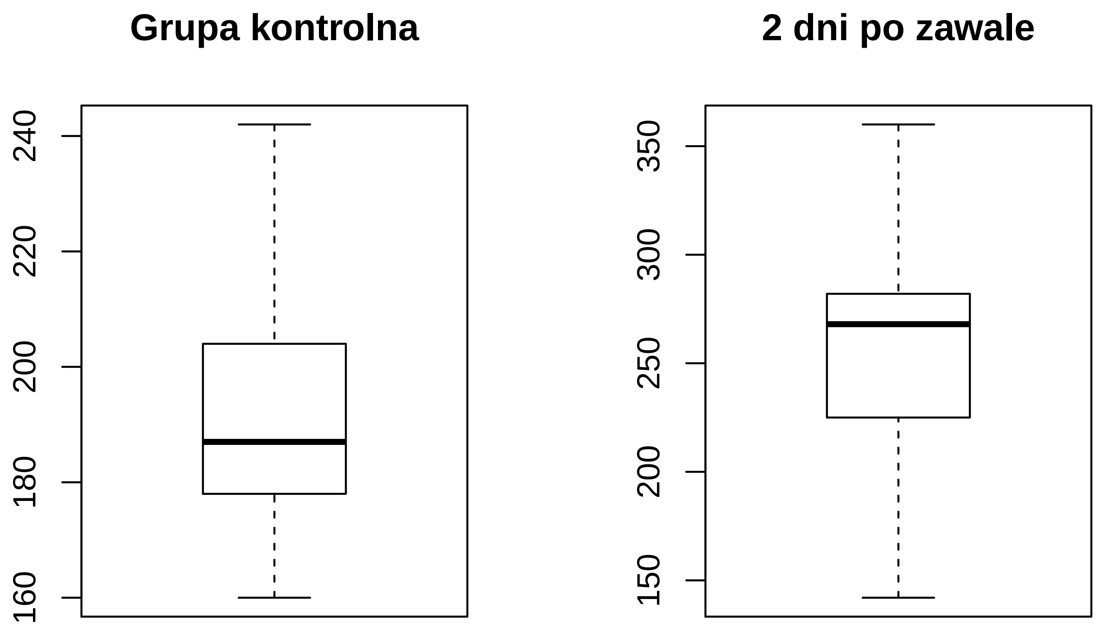


```r
qqnorm(D14,main="14 dni po zawale",xlab="")
```



Podpunkt b

```r
par(mfrow=c(1,2))
boxplot(GK,main="Grupa kontrolna")
boxplot(D2,main="2 dni po zawale")
```




```
par(mfrow=c(1,1))
```

Podpunkt c

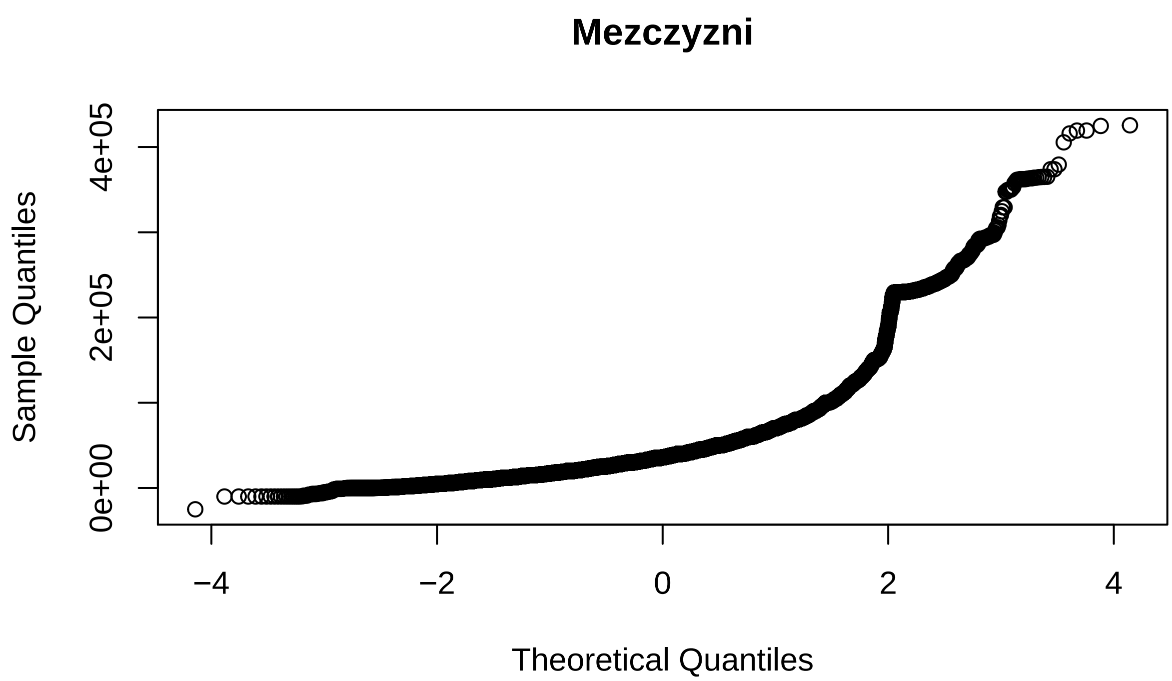
```
z3c2=t.test(D2,GK,alternative = "g")$statistic
```

Wartość 6.1452451 statystyki testowej jest większa od jakiejkoliwiek sensownej wartości krytycznej, zatem średnie ciśnienie o osób dwa dni po zawale jest większe niż u osób z grupy kontrolnej.

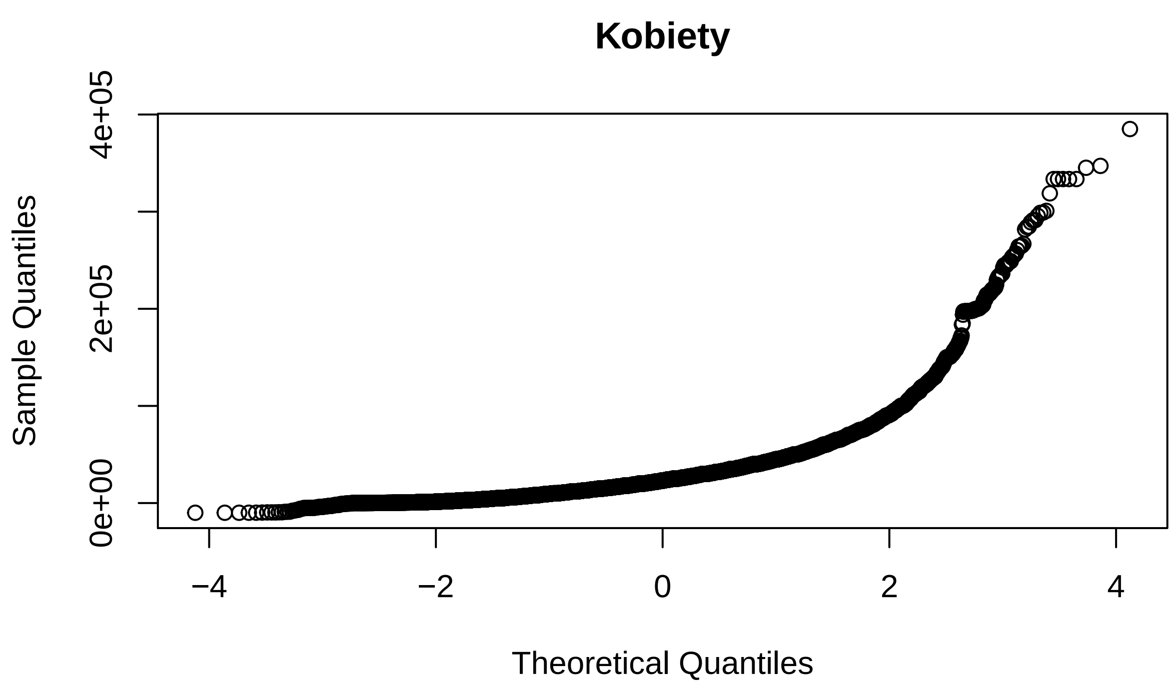
Zadanie 4

Podpunkt a

```
qqnorm(t[[5]][t[4]=="1"],main="Mężczyżni")
```



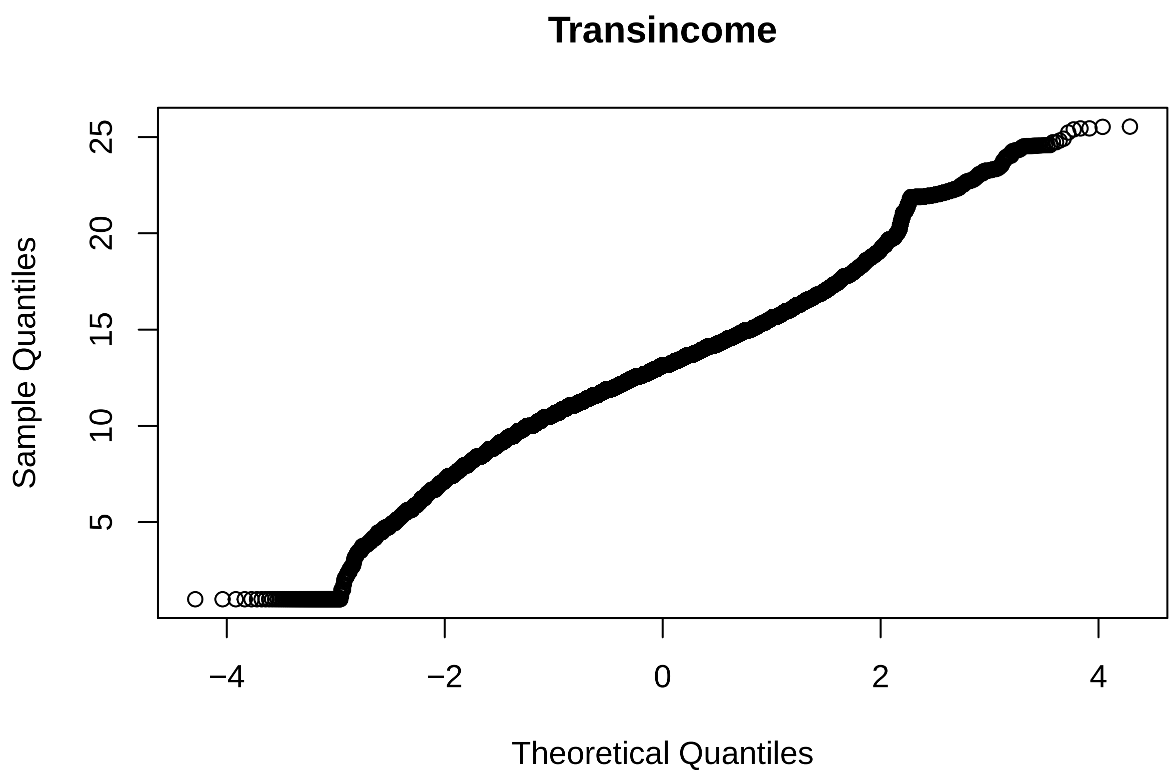

```r
qqnorm(t[[5]][t[4]=="2"],main="Kobiety")
```



Oba histogramy są skośne w prawo, u kobiet ilość obserwacji dla rosnących zarobków spada szybciej niż u mężczyzn.

Podpunkt b

```
transincome=(t[[5]]^.25)
qqnorm(transincome,main="Transincome")
```



Jeśli odrzucilibyśmy skrajnie małe wartości dochodu, to wykres QQ zmiennej transincome byłby prostą, zatem histogram tej zmiennej przypomina gęstość rozkłądu normalnego.

Podpunkt c

```
tiM=transincome[t[4]=="1"]
tiF=transincome[t[4]=="2"]

z4c1=t.test(tiM,tiF,alternative = "g")$statistic
```

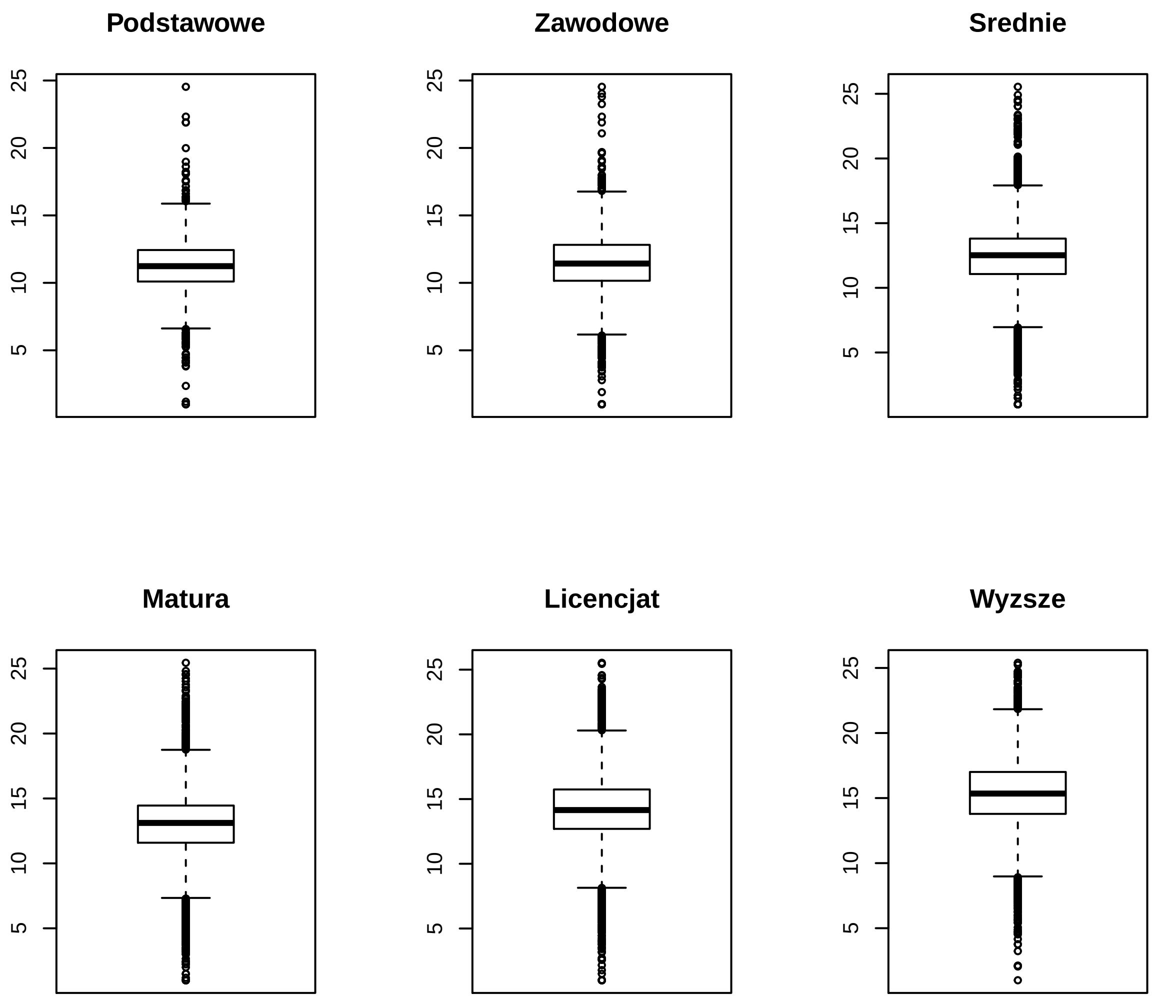
Wartość statystyki testowej wynosi 72.5759446 i jest znacznie większa od jakiejkolwiek sensownej wartośći krytycznej, zatem średnia zmiennej transincome dla mężczyzn jest większa niż u kobiet

Podpunkt d

```
tiW1=transincome[t[3]=="1"]
tiW2=transincome[t[3]=="2"]
```

```
tiW3=transincome[t[3]=="3"]
tiW4=transincome[t[3]=="4"]
tiW5=transincome[t[3]=="5"]
tiW6=transincome[t[3]=="6"]

par(mfrow=c(2,3))
boxplot(tiW1,main="Podstawowe")
boxplot(tiW2,main="Zawodowe")
boxplot(tiW3,main="Średnie")
boxplot(tiW4,main="Matura")
boxplot(tiW5,main="Licencjat")
boxplot(tiW6,main="Wyższe")
```




```
par(mfrow=c(1,1))
```

Zarobki rosną wraz z wykształceniem, największa różnica zauważalna jest między kolejnymi wykresami jest między licencjatem a wykształceniem wyższym, oraz maturą a licencjatem.

Podpunkt e

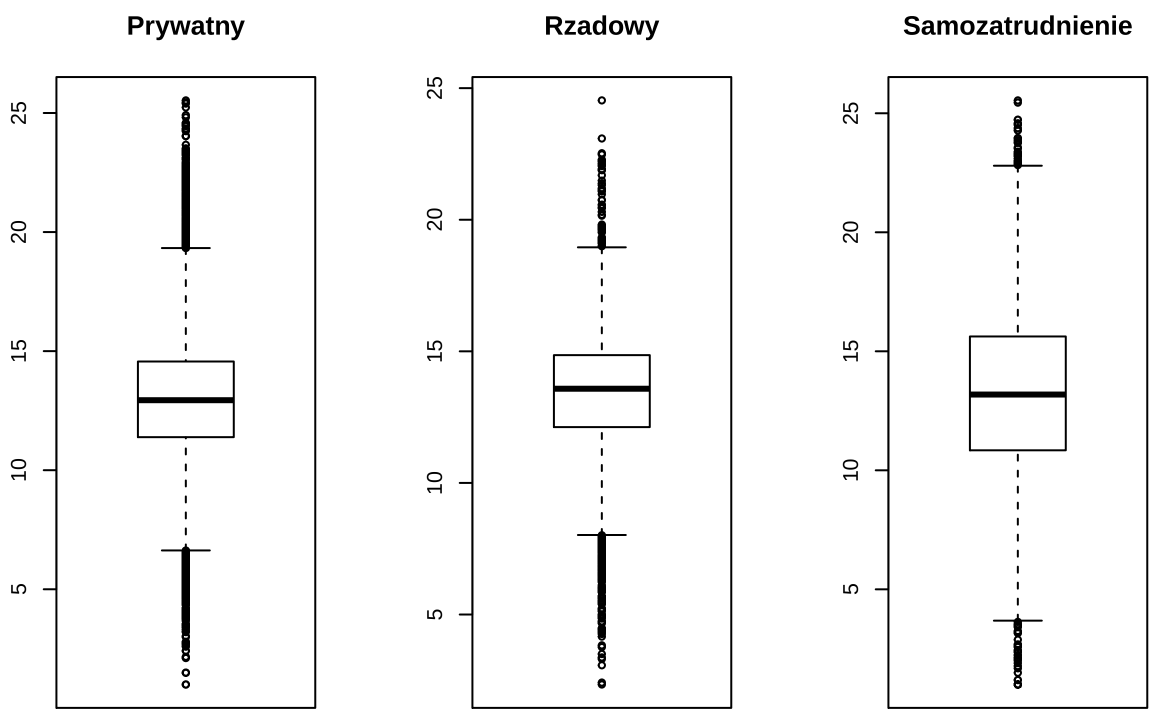
```
z4e1=t.test(tiW6,tiW5,alternative = "g")$statistic
```

Wartość statystyki testowej wynosi 27.0792726,zatem średnia zmiennej transincome dla osób z wykształceniem wyższym jest istotnie wyższa od średniej osób z licencjatem.

Podpunkt f

```
tiS5=transincome[t[6]=="5"]
tiS6=transincome[t[6]=="6"]
tiS7=transincome[t[6]=="7"]

par(mfrow=c(1,3))
boxplot(tiS5,main="Prywatny")
boxplot(tiS6,main="Rządowy")
boxplot(tiS7,main="Samozatrudnienie")
```



```
par(mfrow=c(1,1))
```

Podpunkt g

```
z4g1=t.test(tiS6,tiS5)$statistic
```

Wartość statystyki testowej wynosi 12.7539617,zatem średnie dla osób pracujących w sektorze rządowym i prywatnym są istotnie różne.