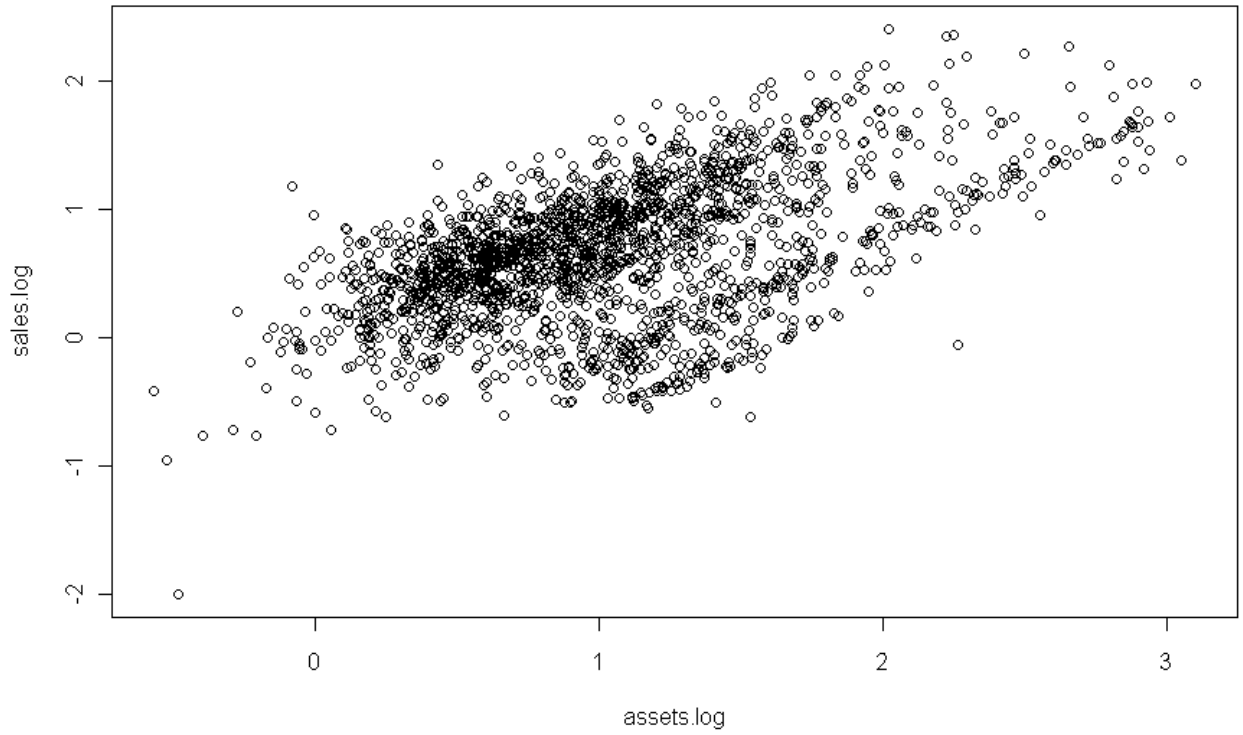
Projekt ***„Nowa oferta edukacyjna Uniwersytetu Wrocławskiego odpowiedzią na współczesne potrzeby rynku pracy i gospodarki opartej na wiedzy”***

Dane:

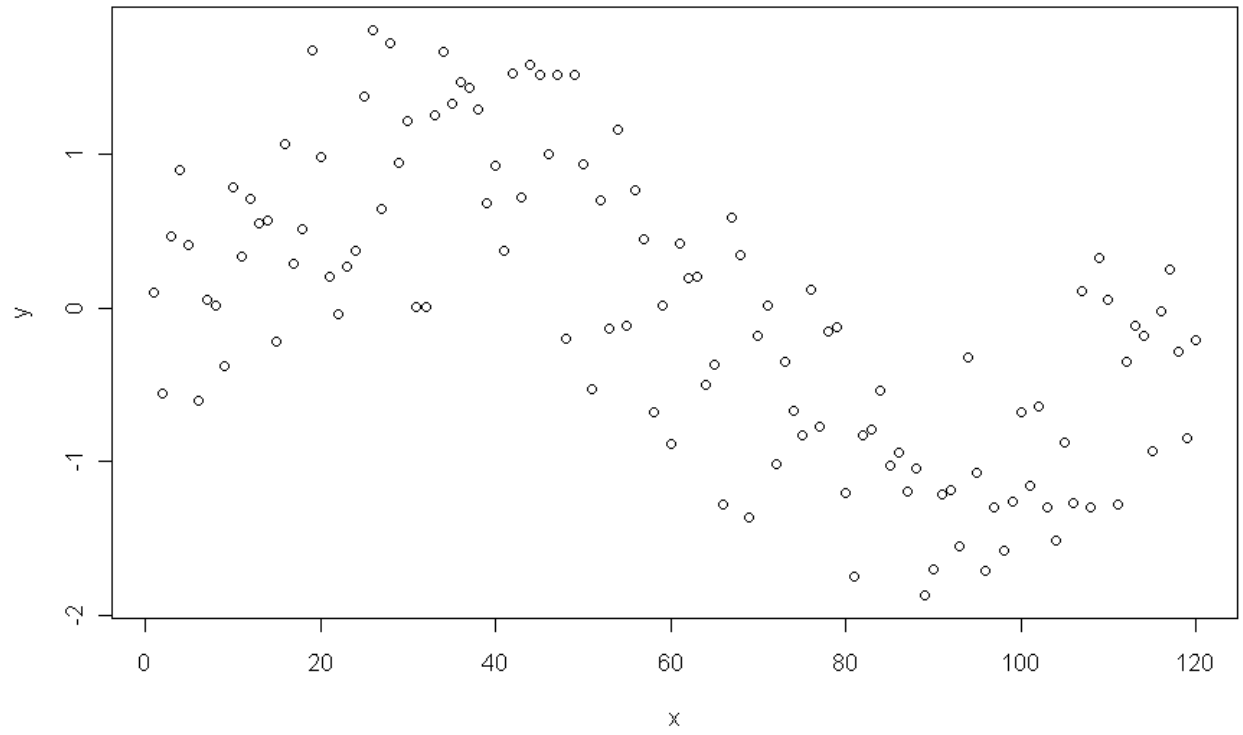
2000 największych spółek światowych z 2004 (*Forbes Magazine*)

```
data("Forbes2000",package="HSAUR2")
attach(Forbes2000)
```

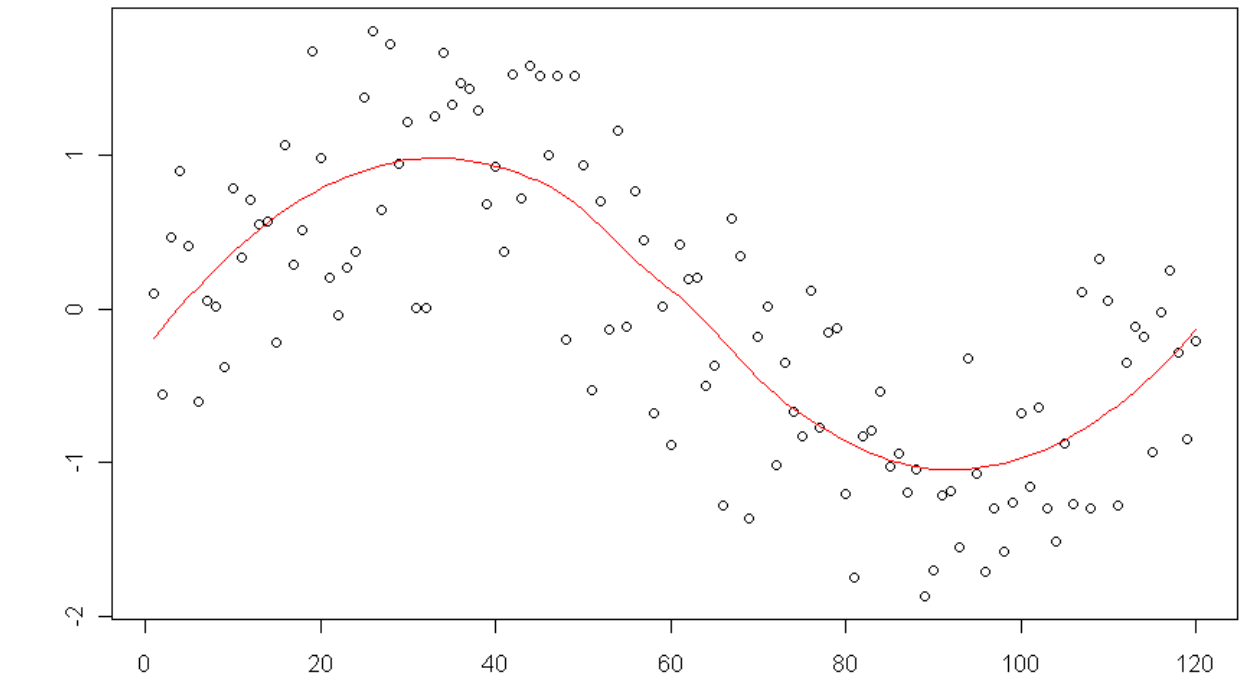
```
plot(log(assets),log(sales))
```



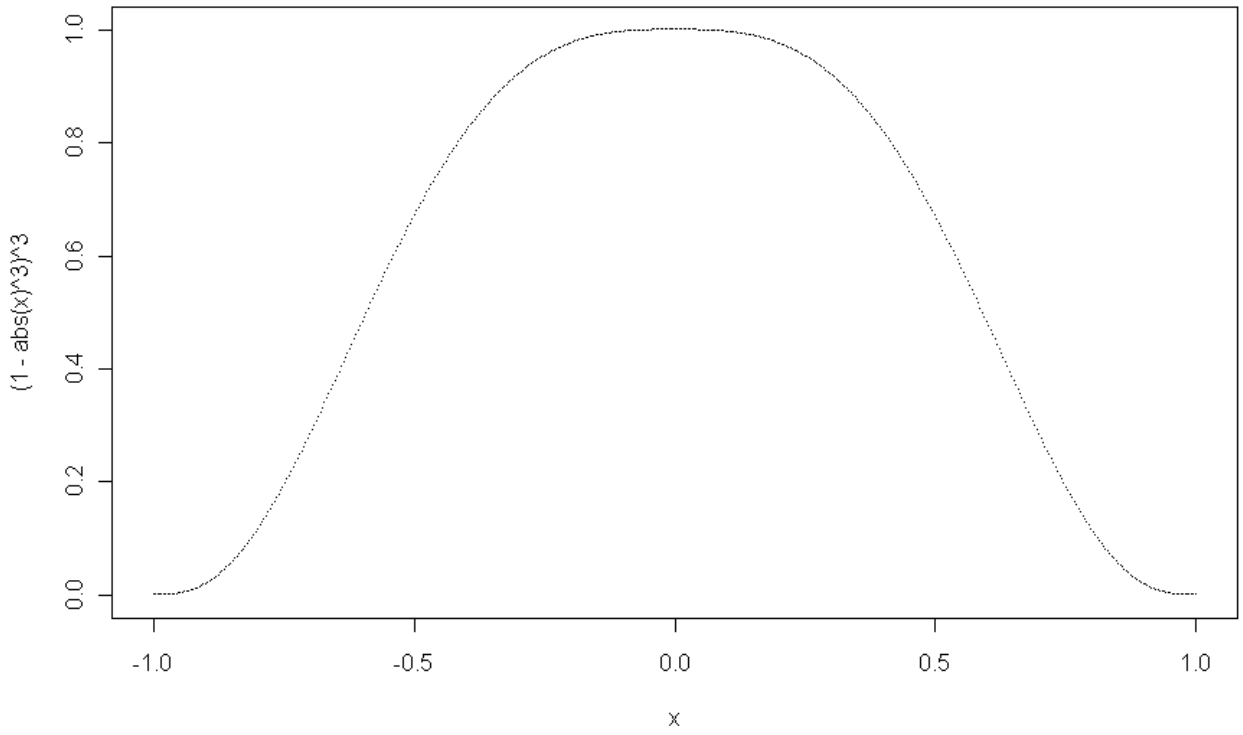
Loess



Po funkcji loess

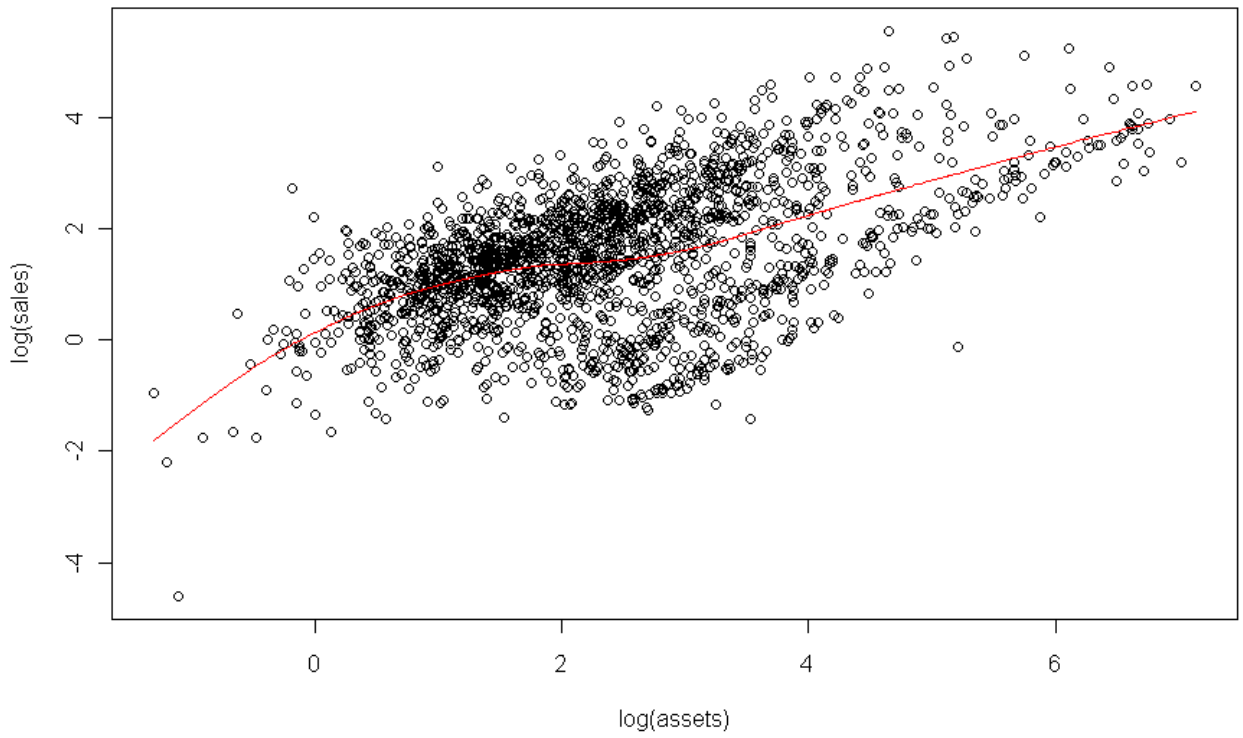


Funkcja wagowa




```
loplot <- function (x,y,xl="",yl="") {
  plot(x,y,xlab=xl,ylab=yl)
  loe <- loess(y ~ x)
  ly <- predict(loe,data.frame(x = sort(x)))
  lines(sort(x),ly,col="red")
}
```

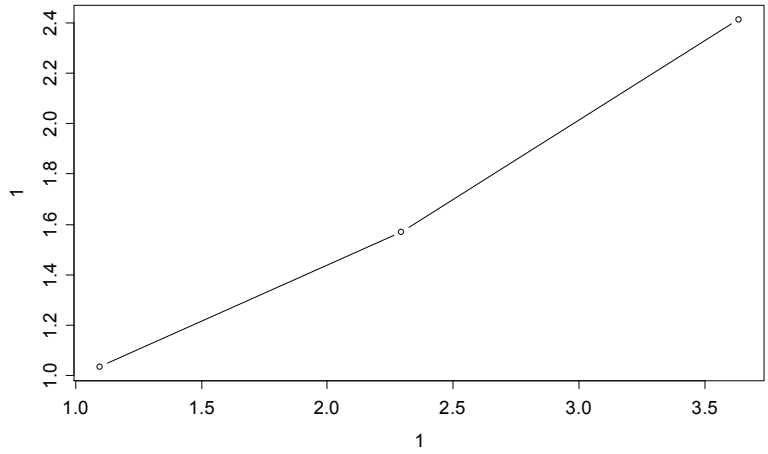
```
loplot(log(assets),log(sales),"log(assets)","log(sales)")
```



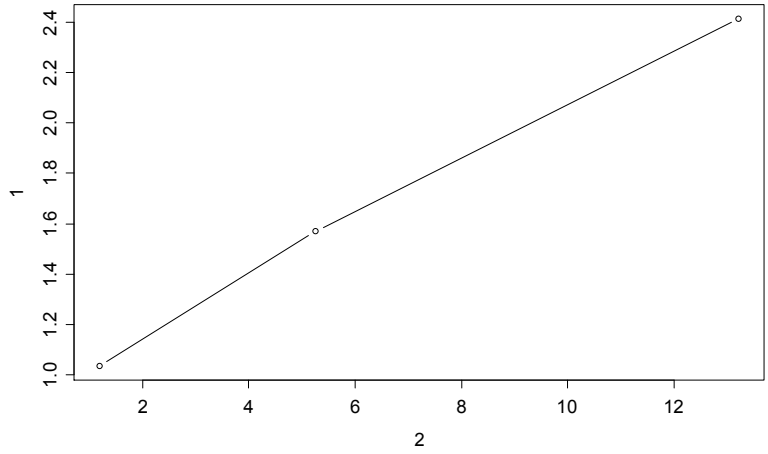
Strzałka Tukeya

strzalkaTukeya(log(assets),log(sales))

```
potęga x = 1  potęga y = 1
 błąd = -0.1703516  b1 = 0.4454196  b2 = 0.6283355

 potęga (x/y),koniec - ESC
```




```
Wybierz x,y,ESC: x
```

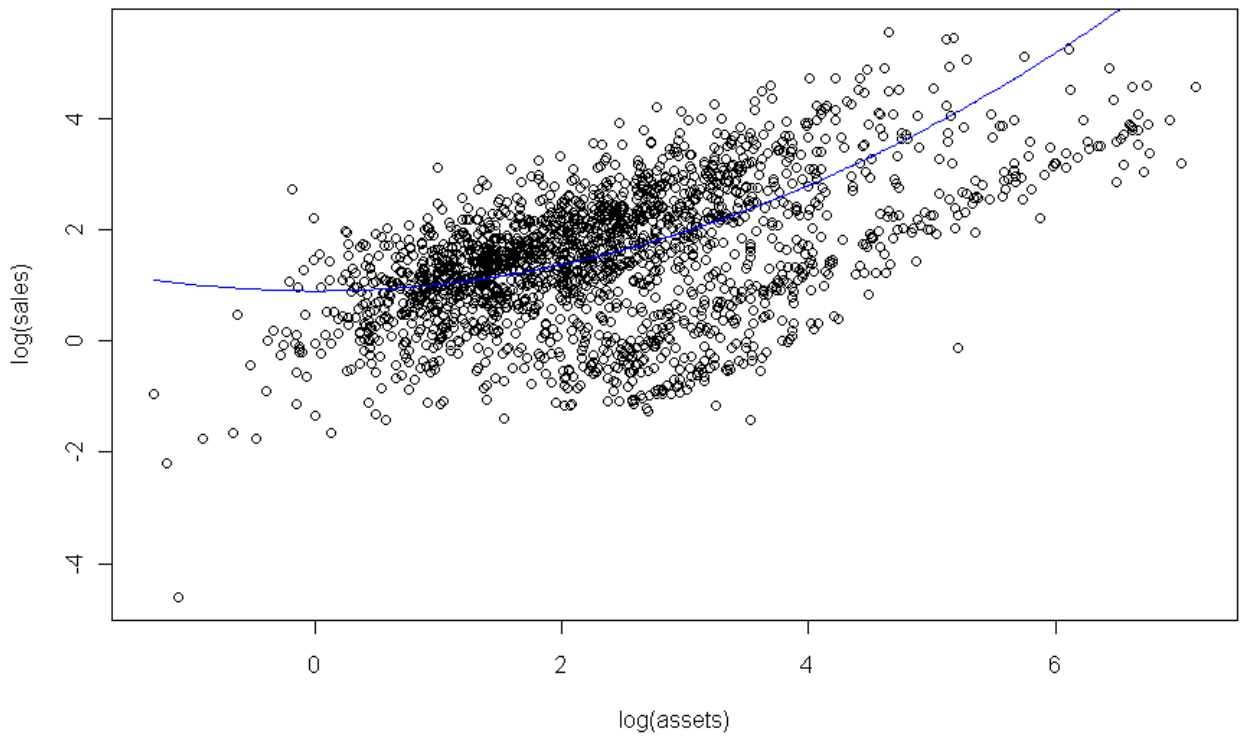



```
potęga x = 2  potęga y = 1
 błąd = 0.1071709  b1 = 0.1314185  b2 = 0.1059767

 potęga (x/y),koniec - ESC
Wybierz x,y,ESC:
```

strzalkaTukeya(log(assets),log(sales),prog=0.11)

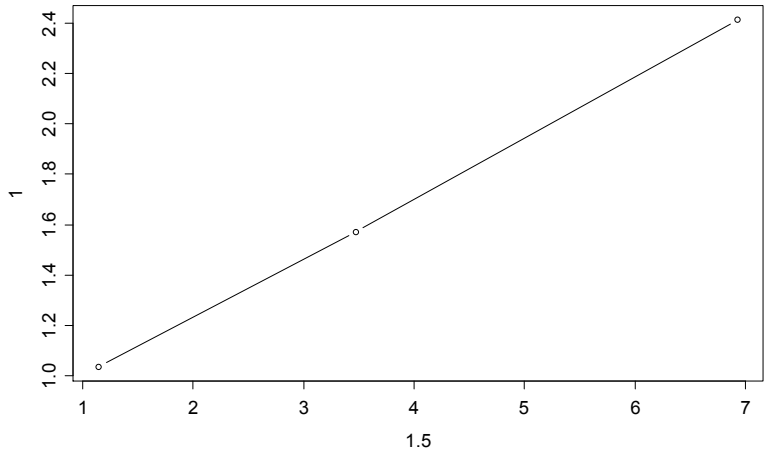
```
potęga x = 1  potęga y = 1
 błąd = -0.1703516  b1 = 0.4454196  b2 = 0.6283355

 potęga (x/y),koniec - ESC
Wybierz x,y,ESC: x

potęga x = 2  potęga y = 1
 błąd = 0.1071709  b1 = 0.1314185  b2 = 0.1059767

 PRZYBLIŻONE RÓWNANIE REGRESJI: a = 0.8943444 b = 0.1186976
```

```
plot(log(assets),log(sales))
lines(sort(log(assets)),0.8943444+0.1186976*sort(log(assets))^2,col="blue")
```



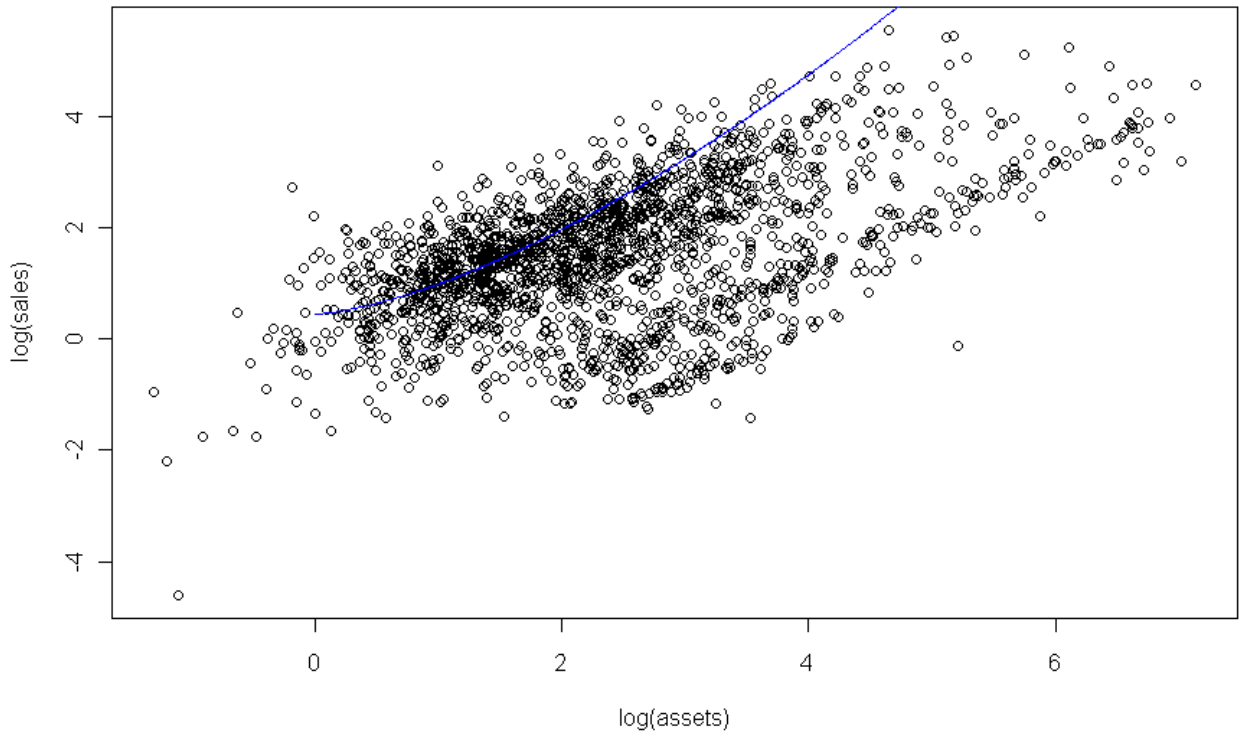
```
strzalkaTukeya(log(assets),log(sales),0.5)
```

```
potęga x = 1  potęga y = 1
 błąd = -0.1703516  b1 = 0.4454196  b2 = 0.6283355

 potęga (x/y),koniec - ESC
Wybierz x,y,ESC: x
```




```
potęga x = 1.5  potęga y = 1
 błąd = -0.03061239  b1 = 0.2293308  b2 = 0.2438149

 PRZYBLIŻONE RÓWNANIE REGRESJI: a = 0.7655625 b = 0.2365729
```

```
plot(log(assets),log(sales))
lines(sort(log(assets)),0.4487091+0.5368776*sort(log(assets))^1.5,col="blue")
```

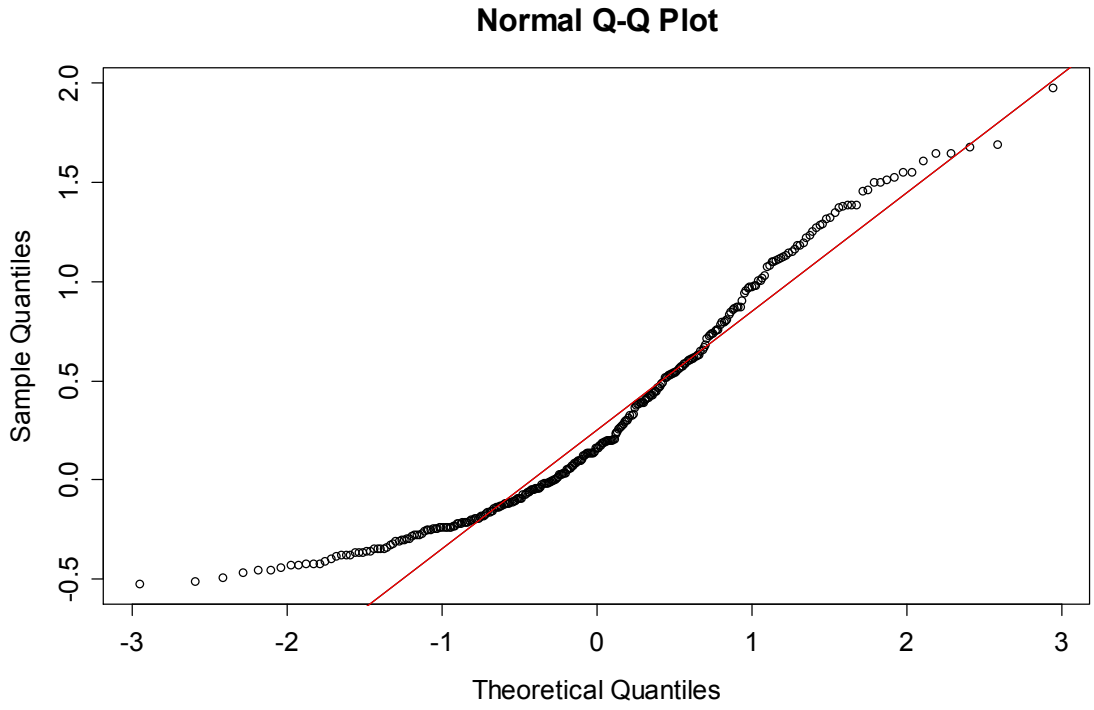


MODELE
ZMIENNA CIĄGŁA~CZYNNIK

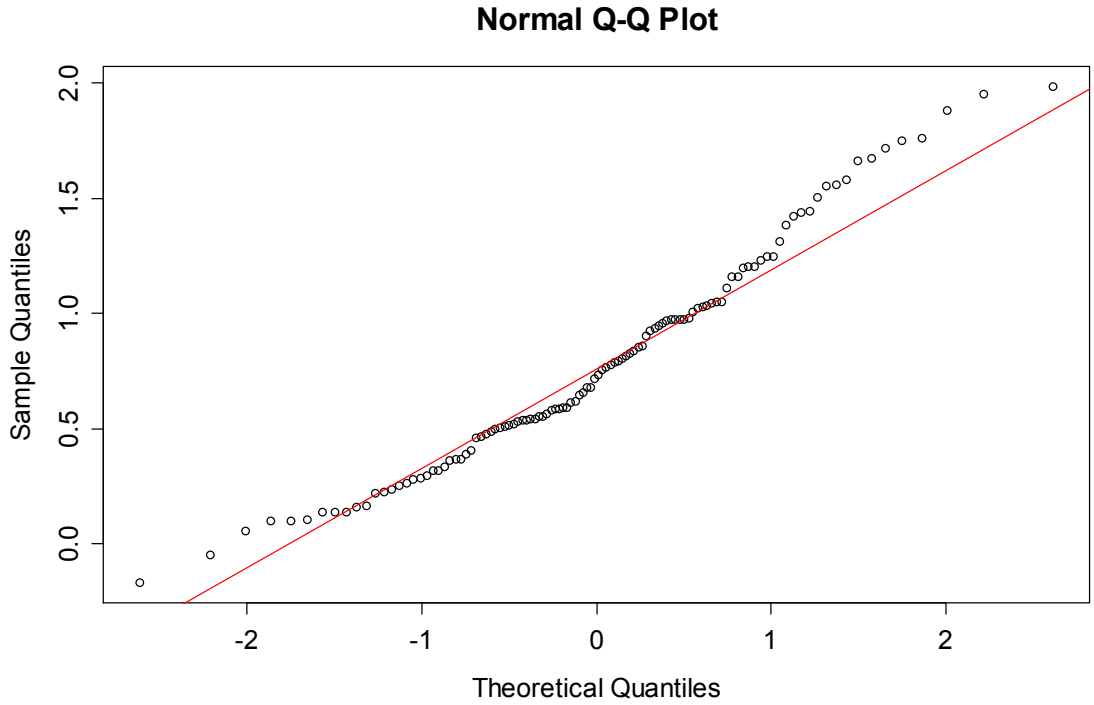
```
lsales.BIM <- log10(sales[category %in% c("Banking","Insurance","Media")])
catBIM <- as.factor(category[category %in% c("Banking","Insurance","Media")])
sBIM <- data.frame(lsales.BIM,catBIM)
```

```
str(sBIM)
'data.frame':  486 obs. of  2 variables:
 $ lsales.BIM: num  1.98 1.88 1.69 1.65 1.75 ...
 $ catBIM    : Factor w/ 27 levels "Aerospace & defense",..: 2 16 2 2 16 2 2 2 2
2 .
```

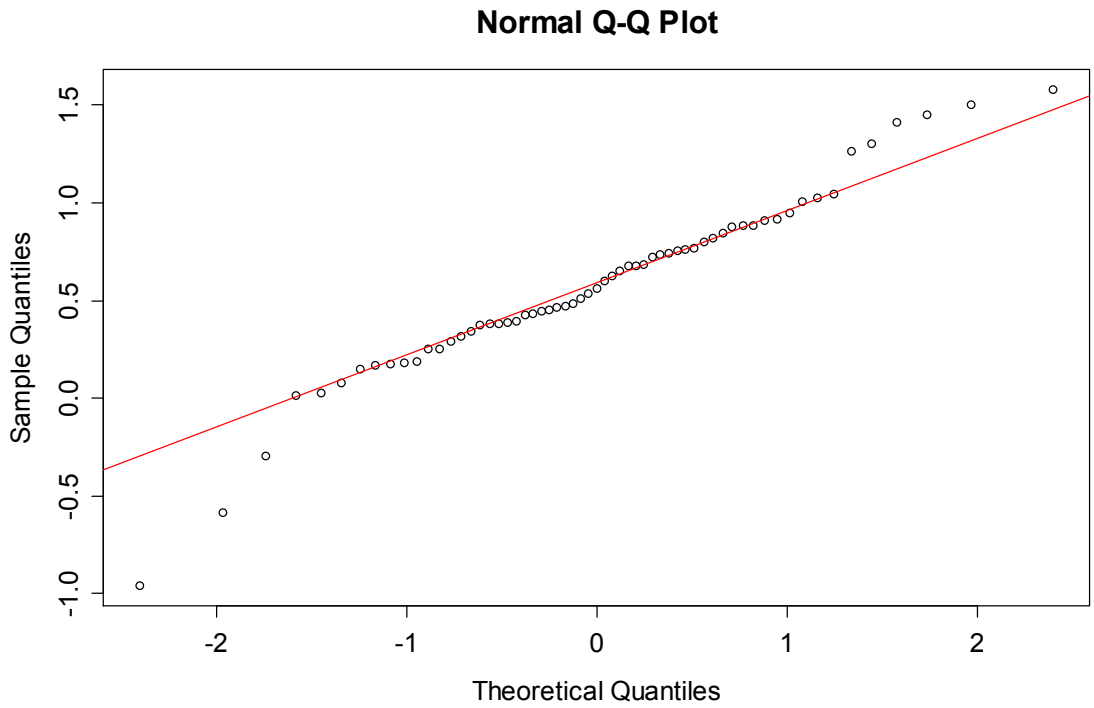
```
catBIM[1:10]
[1] Banking   Insurance Banking   Banking   Insurance Banking   Banking
Banking
 [9] Banking   Banking
qqnorm(lsales.BIM[catBIM=="Banking"])
qqline(lsales.BIM[catBIM=="Banking"],col=2)
```




```
qqnorm(lsales.BIM[catBIM=="Insurance"])
qqline(lsales.BIM[catBIM=="Insurance"],col=2)
```




```
qqnorm(lsales.BIM[catBIM=="Media"])
qqline(lsales.BIM[catBIM=="Media"],col=2)
```




```
cat("Banking",mean(lsales.BIM[catBIM=="Banking"]),"\n",
   "Insurance",mean(lsales.BIM[catBIM=="Insurance"]),"\n",
   "Media",mean(lsales.BIM[catBIM=="Media"]))
```

```
Banking 0.3050799
 Insurance 0.7846592
 Media 0.5770945
```

```
cat("Banking",sd(lsales.BIM[catBIM=="Banking"]),"\n",
   "Insurance",sd(lsales.BIM[catBIM=="Insurance"]),"\n",
   "Media",sd(lsales.BIM[catBIM=="Media"]))
```

```
Banking 0.562247
 Insurance 0.4816577
 Media 0.4625564
```

```
library("lawstat")
levene.test(lsales.BIM,catBIM,location="mean")
```

```
        classical Levene's test based on the absolute deviations from the mean (
        none not applied because the location is not set to median )

data:  lsales.BIM
Test Statistic = 5.8394, p-value = 0.00312
```

```
sBIM1 <- subset(sBIM,lsales.BIM>0)
```

```
colnames(sBIM1)<- c("lsales.BIM1","catBIM1")
```

```
cat("Banking",mean(lsales.BIM1[catBIM1=="Banking"]),"\n",
   "Insurance",mean(lsales.BIM1[catBIM1=="Insurance"]),"\n",
   "Media",mean(lsales.BIM1[catBIM1=="Media"]))
```

```
Banking 0.641726
 Insurance 0.8009084
 Media 0.6386005
```

```
cat("Banking",sd(lsales.BIM1[catBIM1=="Banking"]),"\n",
   "Insurance",sd(lsales.BIM1[catBIM1=="Insurance"]),"\n",
   "Media",sd(lsales.BIM1[catBIM1=="Media"]))
```

```
Banking 0.4693891
 Insurance 0.4703896
 Media 0.3782024
```

```
levene.test(lsales.BIM1,catBIM1,location="mean")
```

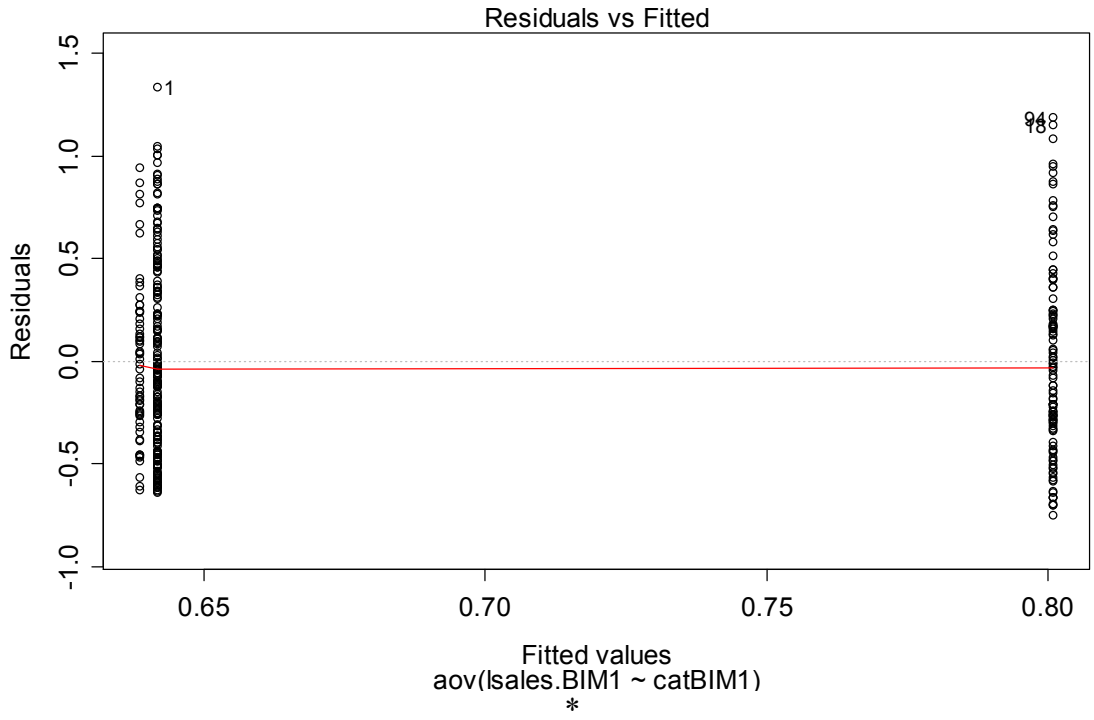
```
        classical Levene's test based on the absolute deviations from the mean (
        none not applied because the location is not set to median )

data:  lsales.BIM1
Test Statistic = 2.919, p-value = 0.05529
```

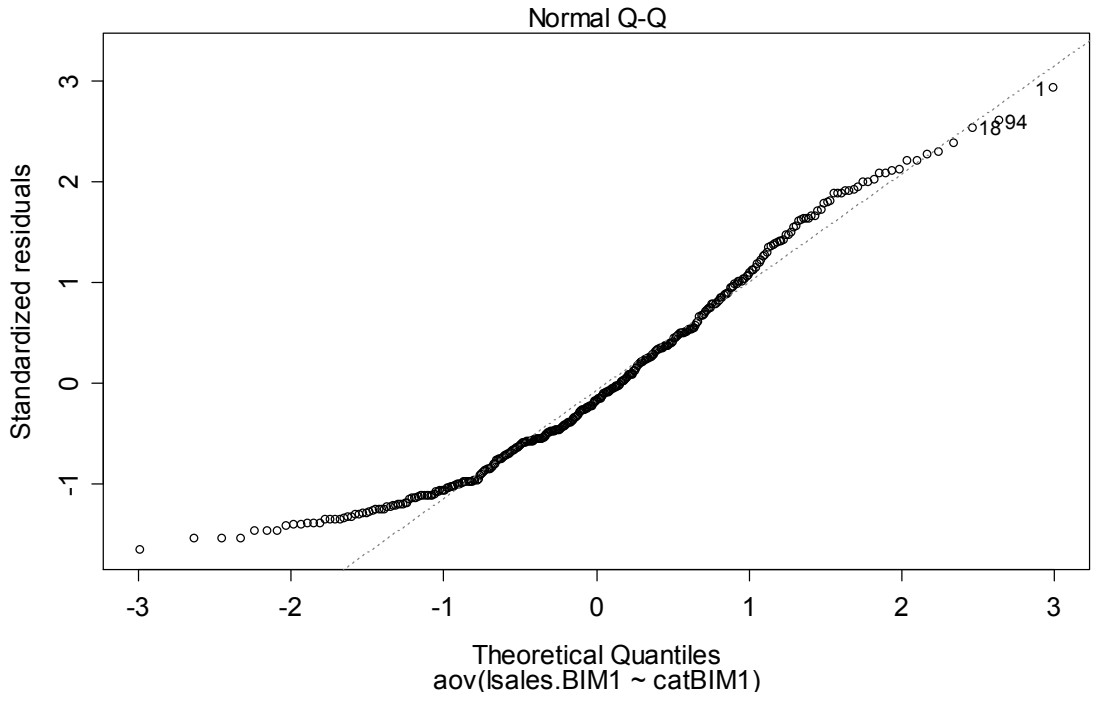
```
(anova(lm(lsales.BIM1~catBIM1)))
```

```
Analysis of Variance Table

Response: lsales.BIM1
           Df Sum Sq Mean Sq F value   Pr(>F)
catBIM1     2  1.949 0.97454  4.6807 0.009853 **
Residuals 355 73.913 0.20821
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
sal.aov <- aov(lsales.BIM1~catBIM1)
plot(sal.aov)
```

seria wykresów diagnostycznych:



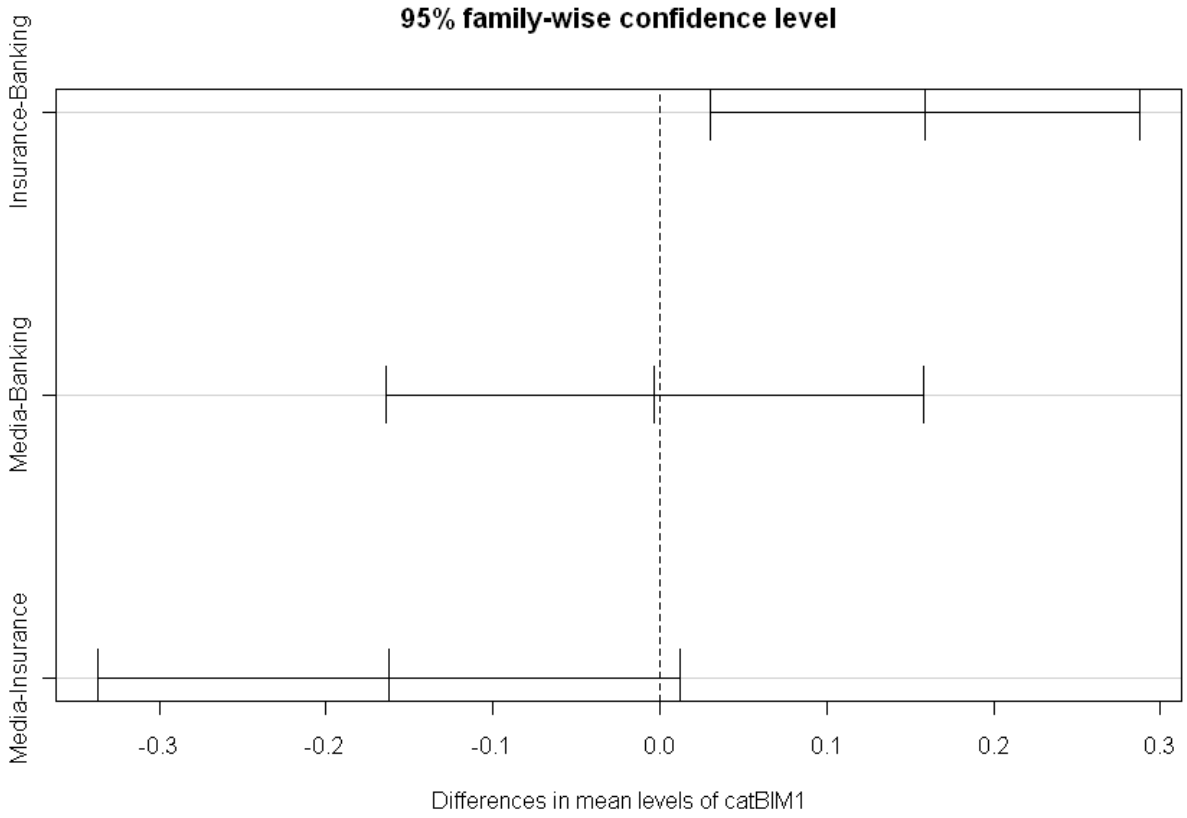


*

```
TukeyHSD(sal.aov)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = lsales.BIM1 ~ catBIM1)

$catBIM1
                           diff         lwr        upr     p adj
Insurance-Banking  0.159182463  0.03051665 0.28784828 0.0106456
Media-Banking     -0.003125459 -0.16423110 0.15798018 0.9988513
Media-Insurance   -0.162307922 -0.33657686 0.01196102 0.0739079
```

```
plot(TukeyHSD(sal.aov))
```



*

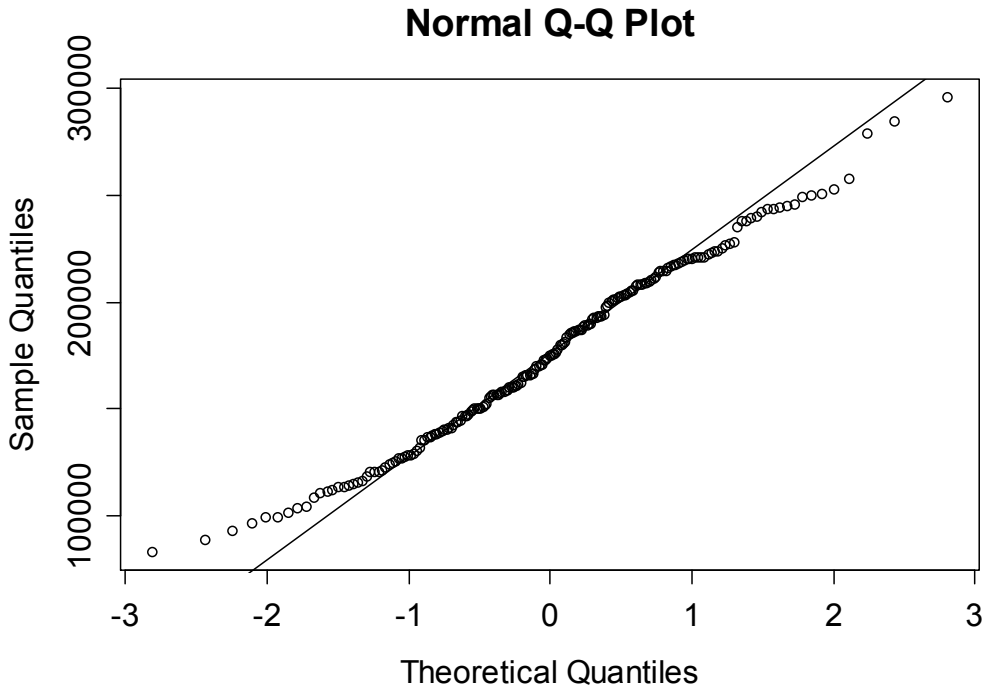
```
> mieszkania <- read.table("http://www.biecek.pl/R/dane/daneMieszkania.csv",
+                          header=T,sep=";")
> summary(mieszkania)
      cena            pokoi       powierzchnia         dzielnica
typ.budynku
 Min.   : 83280   Min.   :1.00   Min.   :17.00   Biskupin   :65   kamienica :61
 1st Qu.:143304   1st Qu.:2.00   1st Qu.:31.15   Krzyki     :79   niski blok:63
 Median :174935   Median :3.00   Median :43.70   Srodmiescie:56   wiezowiec :76
 Mean   :175934   Mean   :2.55   Mean   :46.20
 3rd Qu.:208741   3rd Qu.:3.00   3rd Qu.:61.40
 Max.   :295762   Max.   :4.00   Max.   :87.70
```

```
> str(mieszkania)
'data.frame':  200 obs. of  5 variables:
 $ cena        : num  226298 128121 118333 202064 130303 ...
 $ pokoi       : int  4 1 1 4 1 2 2 1 3 1 ...
 $ powierzchnia: num  76.4 19.4 21.5 65.2 21.4 35.8 41.1 18.1 54.6 18.9 ...
 $ dzielnica   : Factor w/ 3 levels "Biskupin","Krzyki",..: 2 1 3 3 2 1 1 2 3 2
...
 $ typ.budynku : Factor w/ 3 levels "kamienica","niski blok",..: 3 3 3 3 3 3 1 1
3 3 ...
```

*

```
> shapiro.test(cena)

        Shapiro-Wilk normality test

data:  cena
W = 0.9879, p-value = 0.0873
```

*



*

```
> levene.test(cena,dzielnica,"median",bootstrap=T)

        bootstrap modified robust Brown-Forsythe Levene-type test based on the
        absolute deviations from the median

data:  cena
Test Statistic = 0.6934, p-value = 0.461
```

*

Spełnione są założenia ANOVA

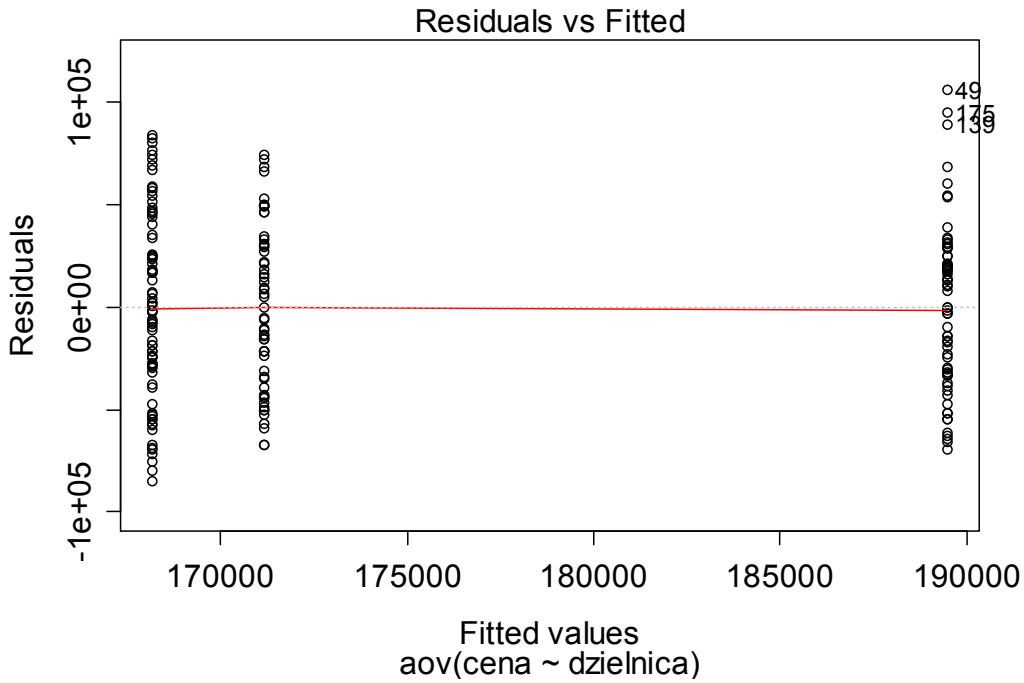
```
> a1 <- anova(lm(cena~dzielnica))
> (a1 <- anova(lm(cena~dzielnica)))
Analysis of Variance Table

Response: cena
            Df     Sum Sq    Mean Sq F value   Pr(>F)
dzielnica    2 1.7995e+10 8997691613  5.0456 0.007294 **
Residuals  197 3.5130e+11 1783263361
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

*

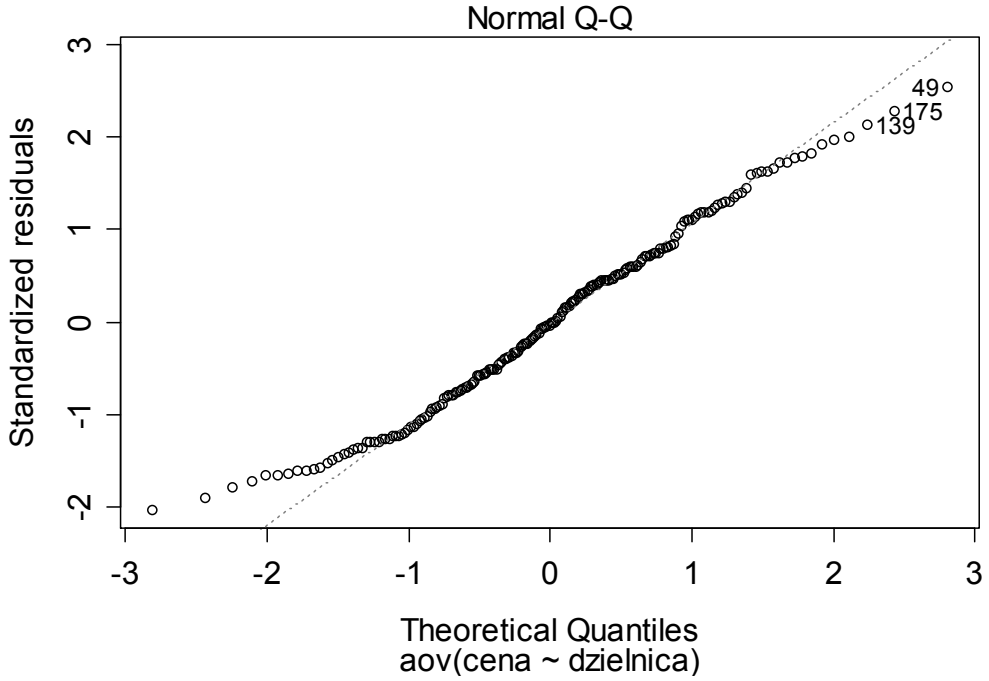
```
> tapply(cena,dzielnica,mean)
   Biskupin      Krzyki Srodmiescie
   189494.0    168173.0    171143.5
```

*

```
a2 <- aov(cena~dzielnica)
plot(a2)
```



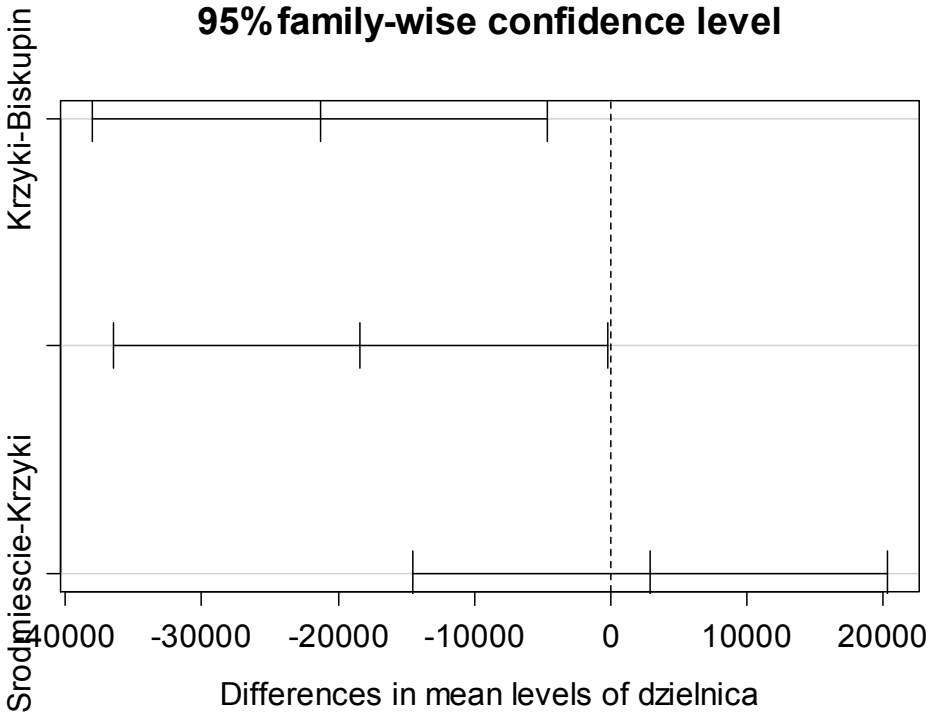
*



*

```
> TukeyHSD(a2,"dzielnica")
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = cena ~ dzielnica)

$dzielnica
                           diff       lwr         upr     p adj
Krzyki-Biskupin      -21321.019 -38021.10 -4620.9333 0.0081457
Srodmiescie-Biskupin -18350.541 -36532.88  -168.2053 0.0473579
Srodmiescie-Krzyki     2970.478 -14450.28 20391.2340 0.9145465
```

Grupy jednorodne {K,S} {B}



```
# kontrasty np
# czy cena na Biskupinie różni się od średniej na Krzykach i Śródmiesciu?
# czy ceny na Krzykach i Śródmiesciu są różne?
```

```
kontrasty=cbind(c(2,-1,-1),c(0,1,-1))
model=lm(cena~dzielnica,contrasts=list(dzielnica=kontrasty))
> summary(model)$coef
             Estimate Std. Error    t value      Pr(>|t|)
(Intercept) 176270.141   3015.732 58.4501978 2.070588e-126
dzielnica1    6611.927   2135.391  3.0963539  2.244773e-03
dzielnica2   -1485.239   3688.392 -0.4026793  6.876208e-01
*
```