Discussion 9

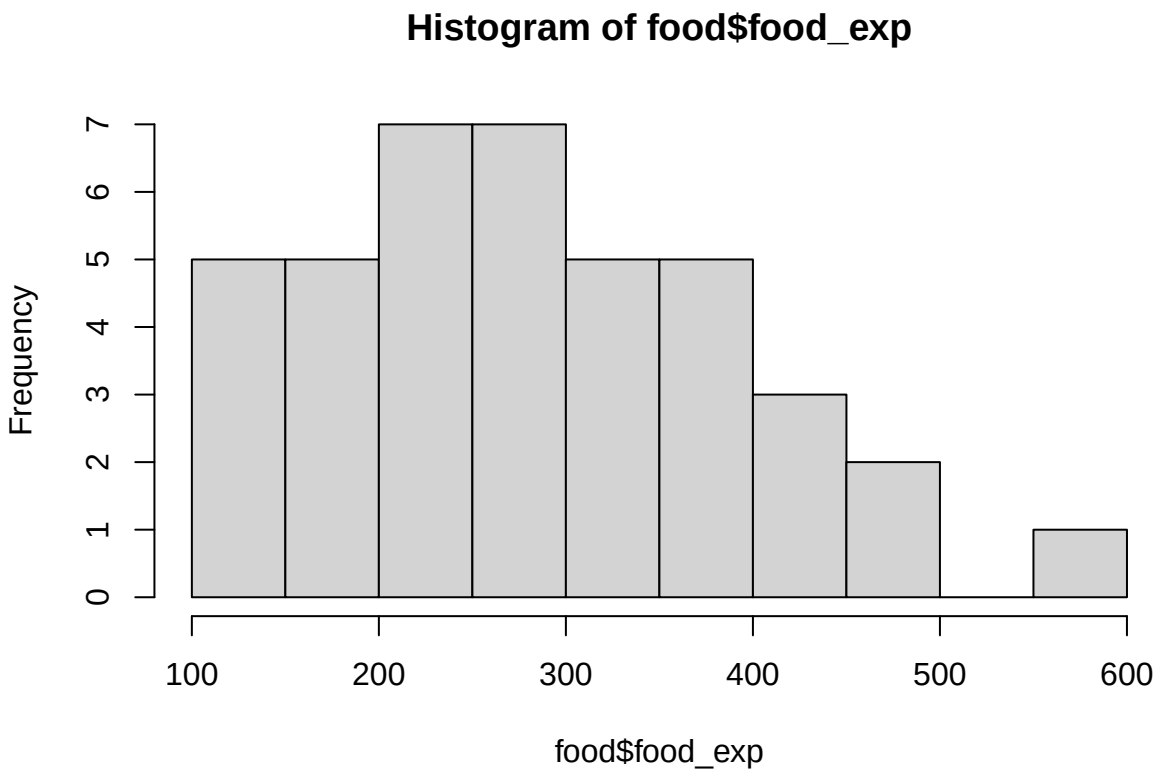
Hanyue Cao

03/28/2022

```
#install.packages("devtools")
#library(devtools)
#install_git("https://github.com/ccolonescu/PoEdata")
#data("food",package="PoEdata")
load(file="~/Boston/TF for 684/Disucssion 9/food.rda")
#load(file="YOUR OWN FILE PATH/food.rda")
head(food)
```

```
##   food_exp income
## 1   115.22   3.69
## 2   135.98   4.39
## 3   119.34   4.75
## 4   114.96   6.03
## 5   187.05  12.47
## 6   243.92  12.98
```

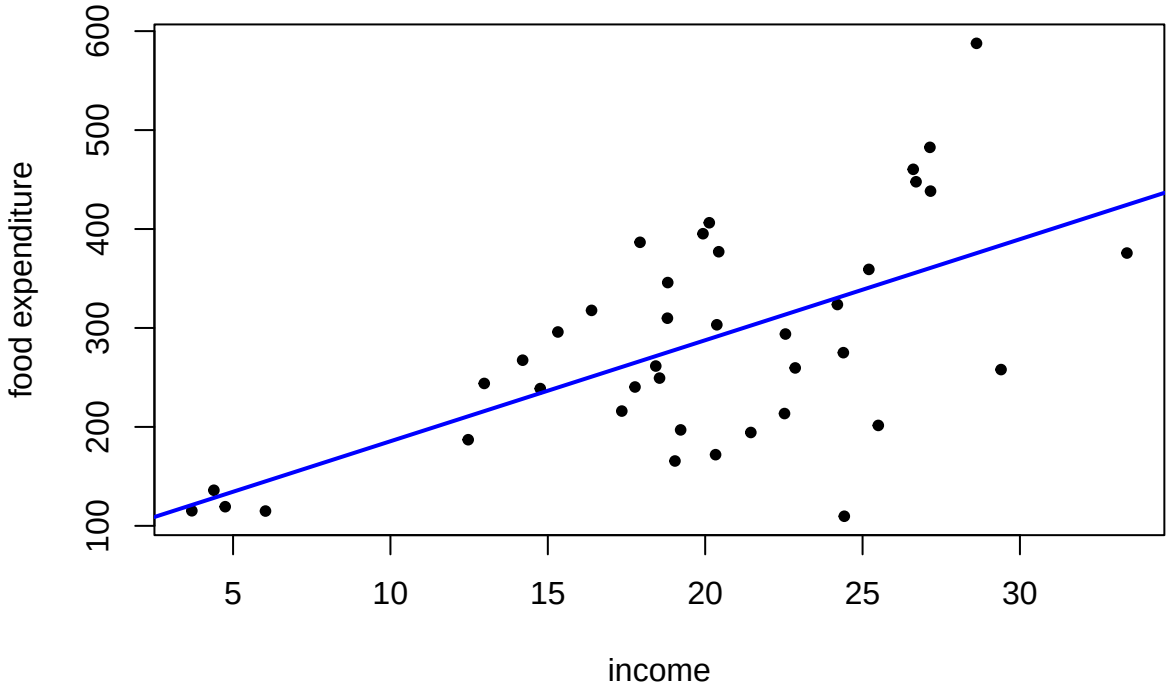
```
hist(food$food_exp)
```



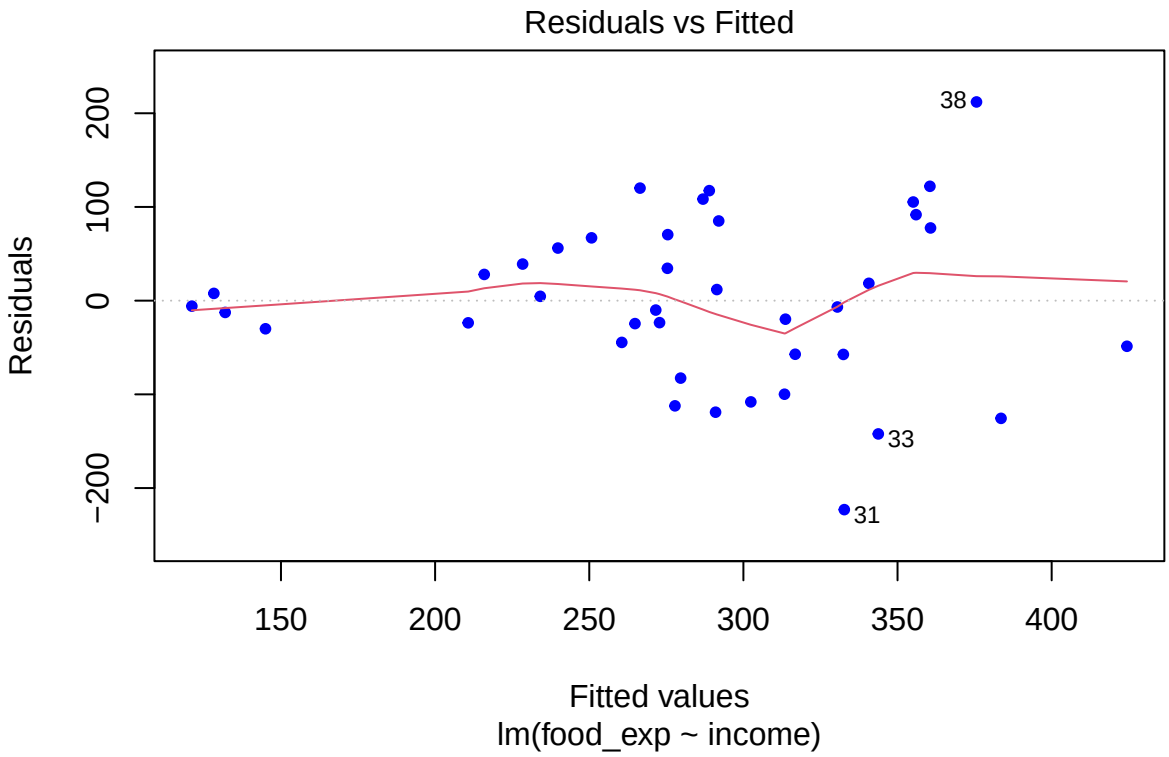

```
mod1 <- lm(food_exp~income, data=food)
summary(mod1)
```

```
##
## Call:
## lm(formula = food_exp ~ income, data = food)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -223.025  -50.816   -6.324   67.879  212.044
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   83.416     43.410   1.922   0.0622 .
## income        10.210      2.093   4.877 1.95e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89.52 on 38 degrees of freedom
## Multiple R-squared:  0.385,  Adjusted R-squared:  0.3688
## F-statistic: 23.79 on 1 and 38 DF,  p-value: 1.946e-05
```

```
plot(food$income,food$food_exp, type="p",
     xlab="income", ylab="food expenditure",pch=20)
abline(mod1,col="blue",lwd=2)
```




```
#par(mfrow=c(1,2))
#plot residual
plot(mod1, which=1, col="blue",pch=20)
```



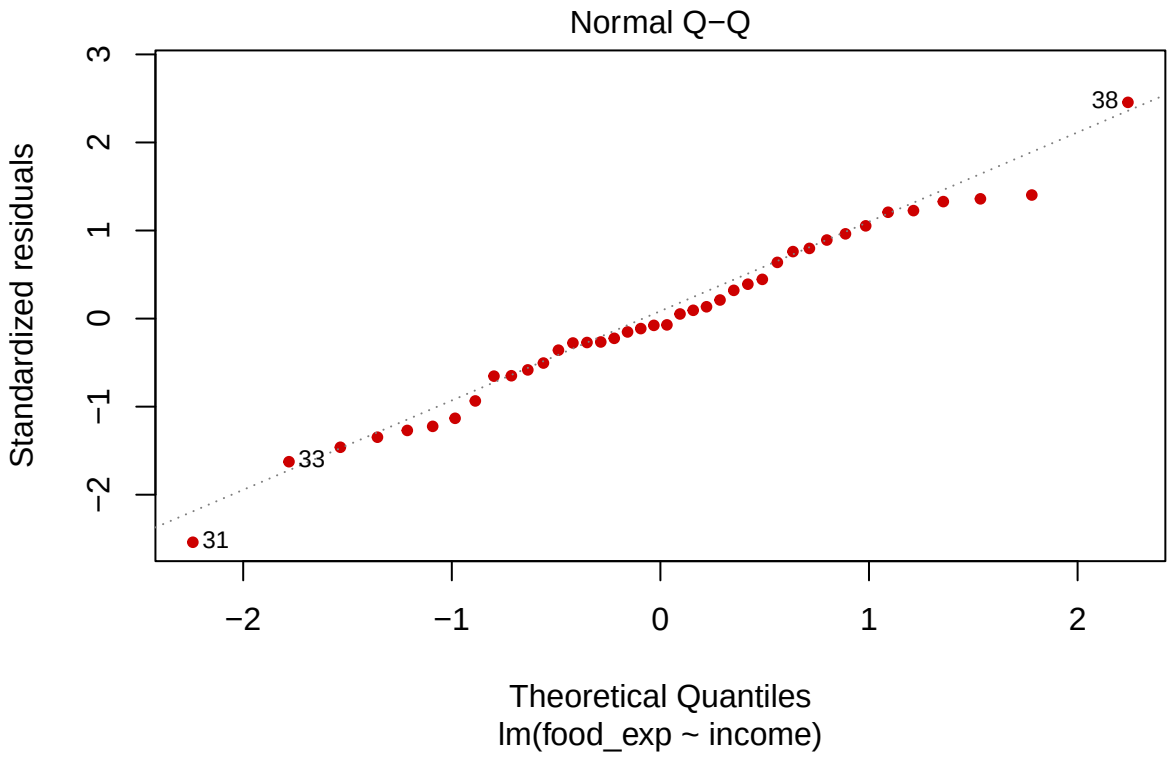

```
#QQ plot
plot(mod1,which=2, col="red3",pch=20)
library(lmtest)
```

```
## Warning: package 'lmtest' was built under R version 4.1.2
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```




```
shapiro.test(resid(mod1))
```

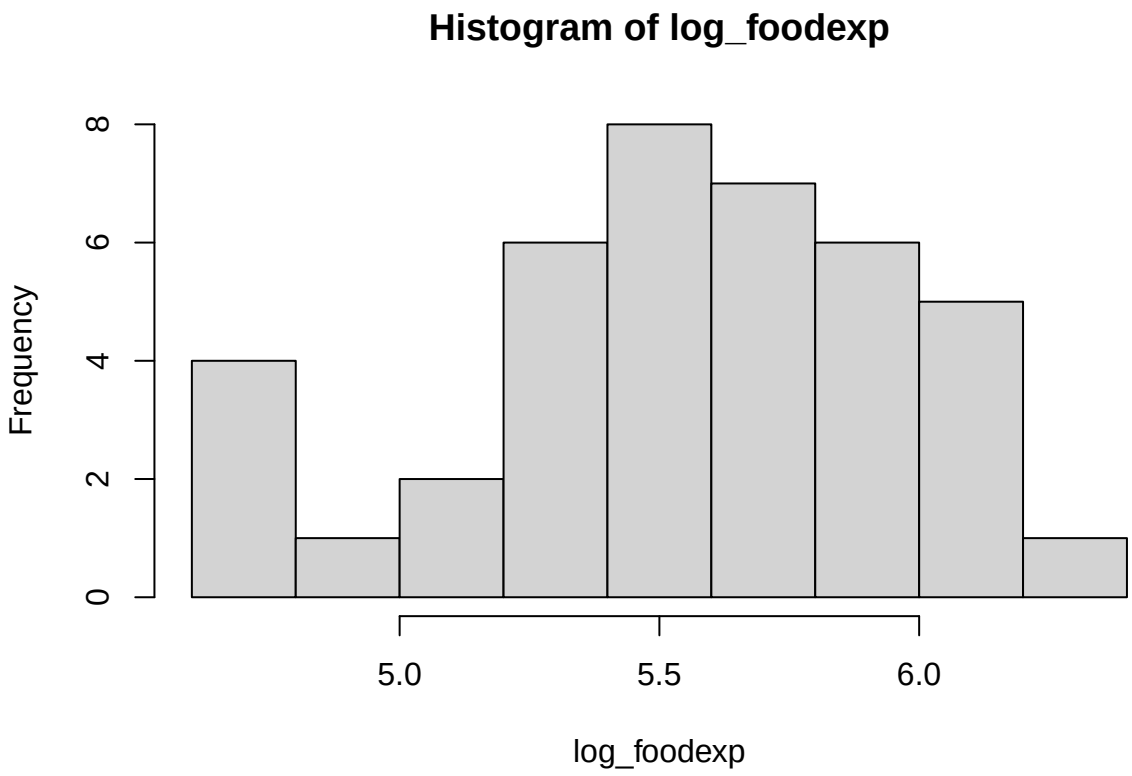
```
##
##  Shapiro-Wilk normality test
##
## data:  resid(mod1)
## W = 0.98838, p-value = 0.9493
```

```
bptest(mod1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  mod1
## BP = 7.3844, df = 1, p-value = 0.006579
```

Fail to reject Shapiro test at $\alpha = 0.05$, so we conclude that the residuals follow a normal distribution.\ We reject Breusch-Pagan test at $\alpha = 0.05$, so we conclude that heteroscedasticity exists.

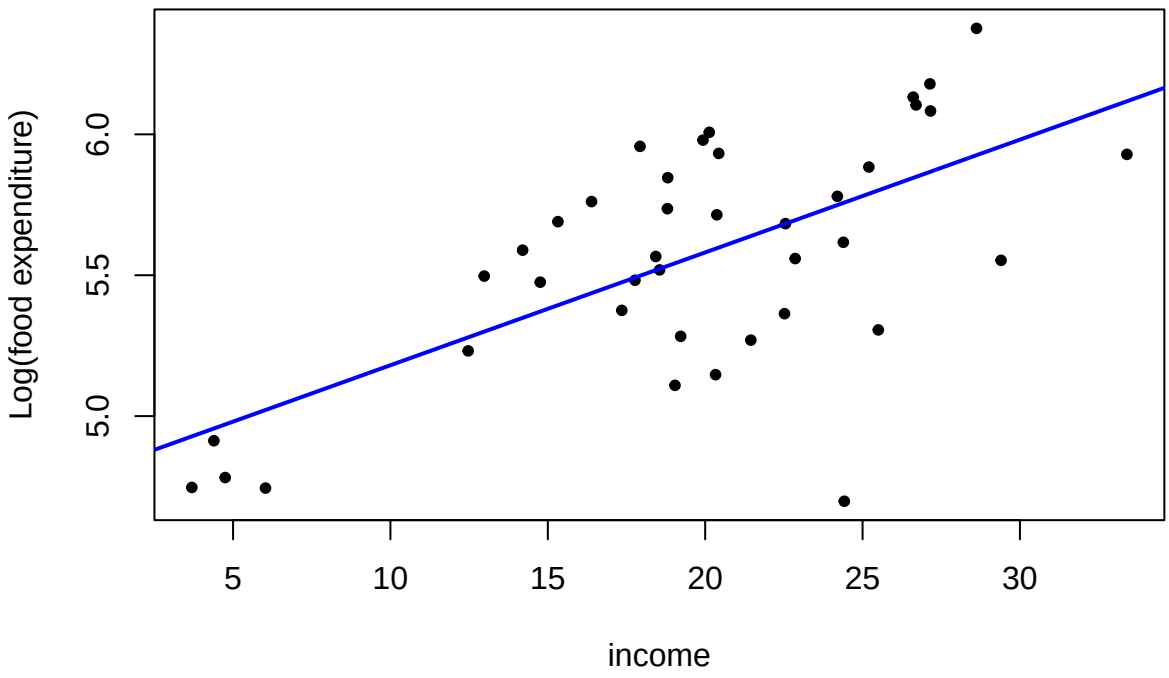
```
log_foodexp=log(food$food_exp)
hist(log_foodexp)
```



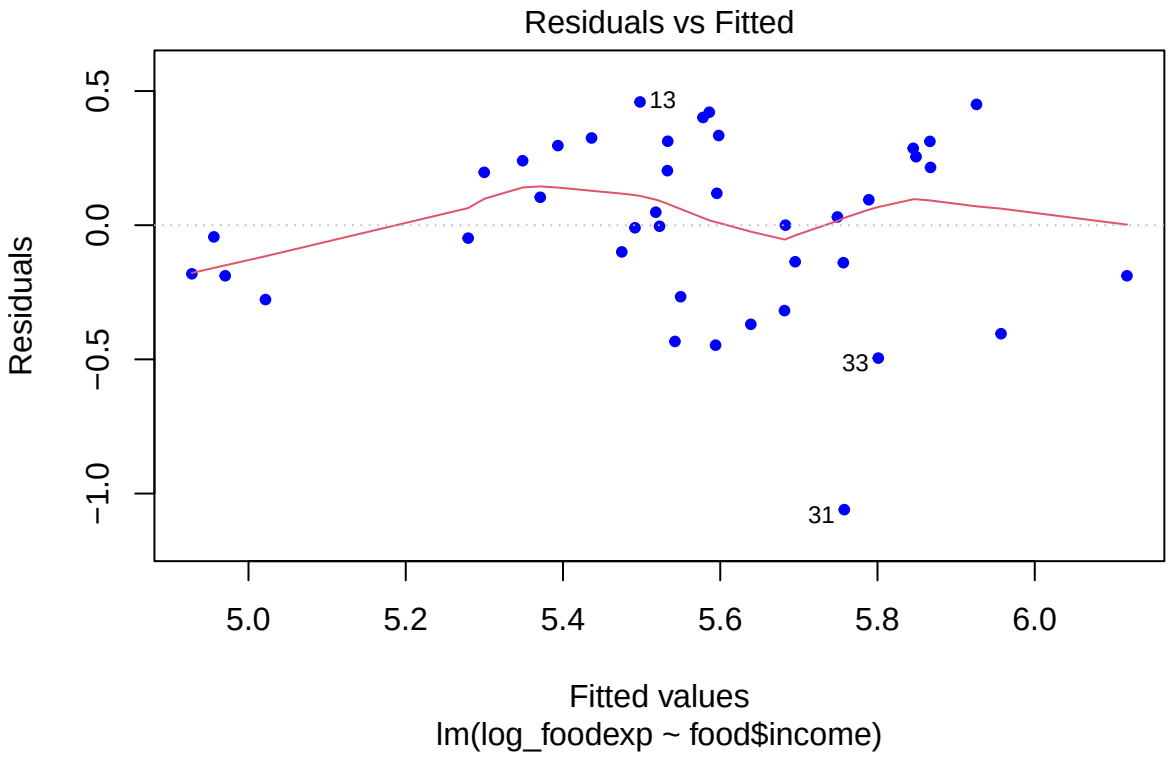

```
mod2 <- lm(log_foodexp~food$income)
summary(mod2)
```

```
##
## Call:
## lm(formula = log_foodexp ~ food$income)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.05993 -0.18838  0.01555  0.26305  0.45933
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.780239   0.158959  30.072  < 2e-16 ***
## food$income 0.040030   0.007665   5.222 6.62e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3278 on 38 degrees of freedom
## Multiple R-squared:  0.4178, Adjusted R-squared:  0.4025
## F-statistic: 27.27 on 1 and 38 DF,  p-value: 6.616e-06
```

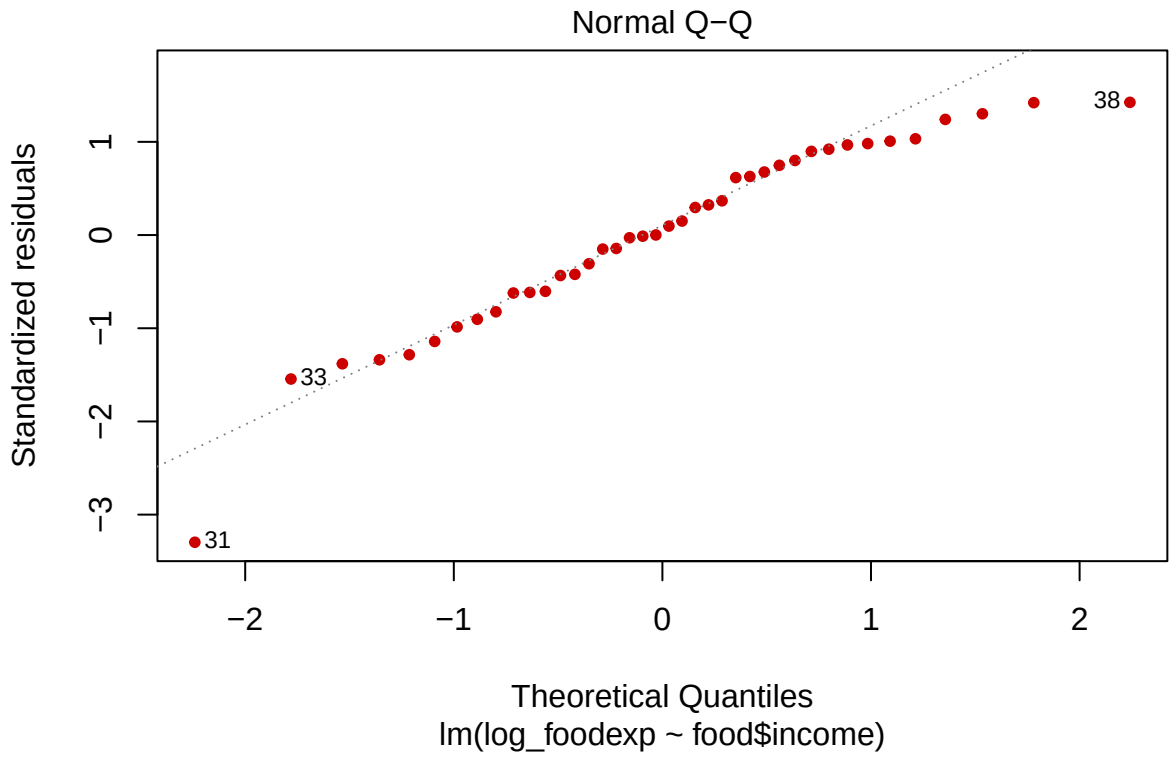
```
plot(food$income,log_foodexp, type="p",
     xlab="income", ylab="Log(food expenditure)",pch=20)
abline(mod2,col="blue",lwd=2)
```




```
#plot residual
plot(mod2, which=1, col="blue",pch=20)
```




```
#QQ plot
plot(mod2,which=2, col="red3",pch=20)
```




```
shapiro.test(resid(mod2))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(mod2)
## W = 0.93951, p-value = 0.03329
```

```
bptest(mod2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  mod2
## BP = 1.711, df = 1, p-value = 0.1909
```

Reject Shapiro-Wilk normality test and fail to reject Breusch-Pagan at $\alpha = 0.05$. We conclude that normality assumption are violated and homoscedasticity assumption are satisfied.