Laboratory Exercise Week 13

Brian Tipton

11/15/23

Directions:

- Write your R code inside the code chunks after each question.
- Write your answer comments after the `#` sign.
- To generate the word document output, click the button `Knit` and wait for the word document to appear.
- RStudio will prompt you (only once) to install the `knitr` package.
- Submit your completed laboratory exercise using Blackboard's Turnitin feature. Your Turnitin upload link is found on your Blackboard Course shell under the Laboratory folder.

For this exercise, you will need to use the packages `mosaic` and `dplyr` to find numerical and graphical summaries.

```
# install packages if necessary
if (!require(mosaic)) install.packages(`mosaic`)
if (!require(dplyr)) install.packages(`dplyr`)
# load the package in R
library(mosaic) # load the package mosaic to use its functions
library(dplyr) # load the package dplyr to use data management functions
```

1. Data from the US Federal Reserve Board (2002) on the percentage of disposable personal income required to meet consumer load payments and mortgage payments for selected years are found in the data below

```
debt <- read.csv("https://www.siue.edu/~jpailde/debt.csv")
debt
```

```
##    Consumer_Debt Household_Debt
## 1           7.88           6.22
## 2           7.91           6.14
## 3           7.65           5.95
## 4           7.61           5.83
## 5           7.48           5.83
## 6           7.49           5.85
## 7           7.37           5.81
## 8           6.57           5.79
## 9           6.24           5.73
## 10          6.09           5.95
## 11          6.32           6.09
## 12          6.97           6.28
## 13          7.38           6.08
## 14          7.52           5.79
## 15          7.84           5.81
```

i) Construct a scatterplot with a simple regression for this data set.

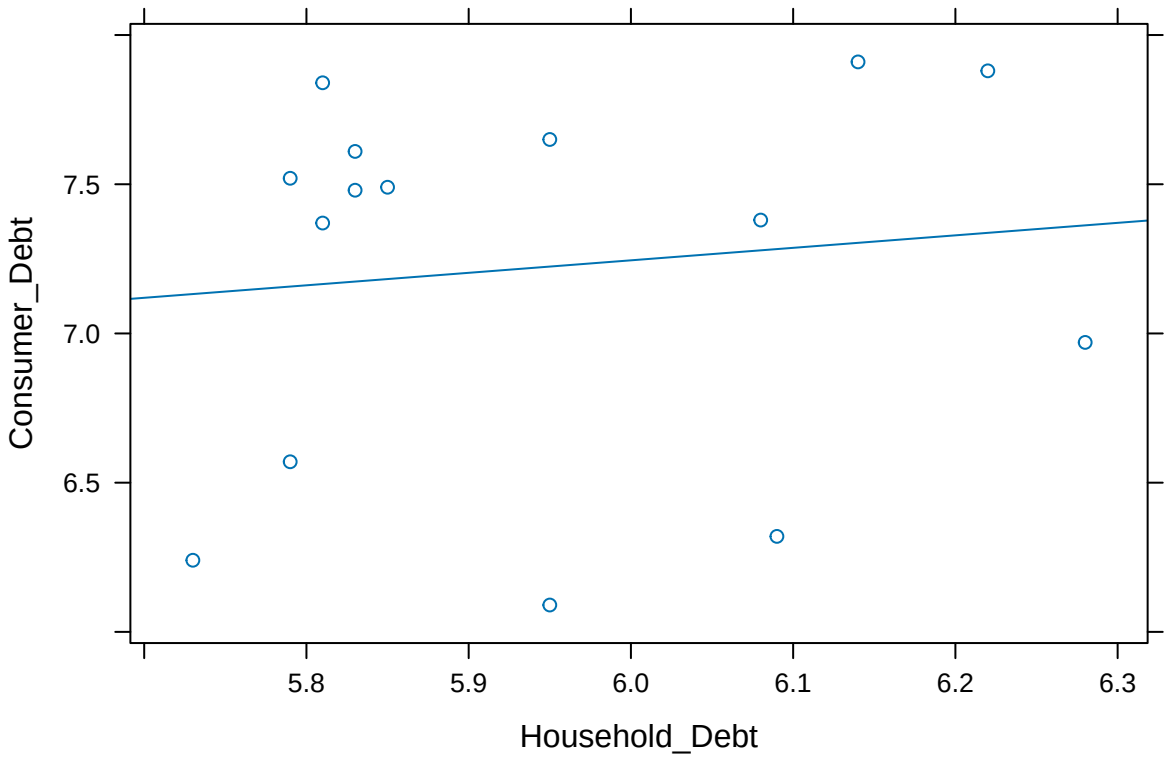
ii) Check the error model assumption visually by constructing a residual plot and QQplot of the residuals. Interpret what you see.

iii) Estimate the population regression slope by constructing 95% confidence interval. Give a brief interpretation of the esimated slope in the context of the problem.

iv) Perform a hypothesis test on the regression slope, use a 5% level of significance. Given an appropriate conclusion.

Code chunk

```
# start your code
#i)
xyplot(Consumer_Debt ~ Household_Debt,
       data = debt,
       type = c('p', 'r'))
```



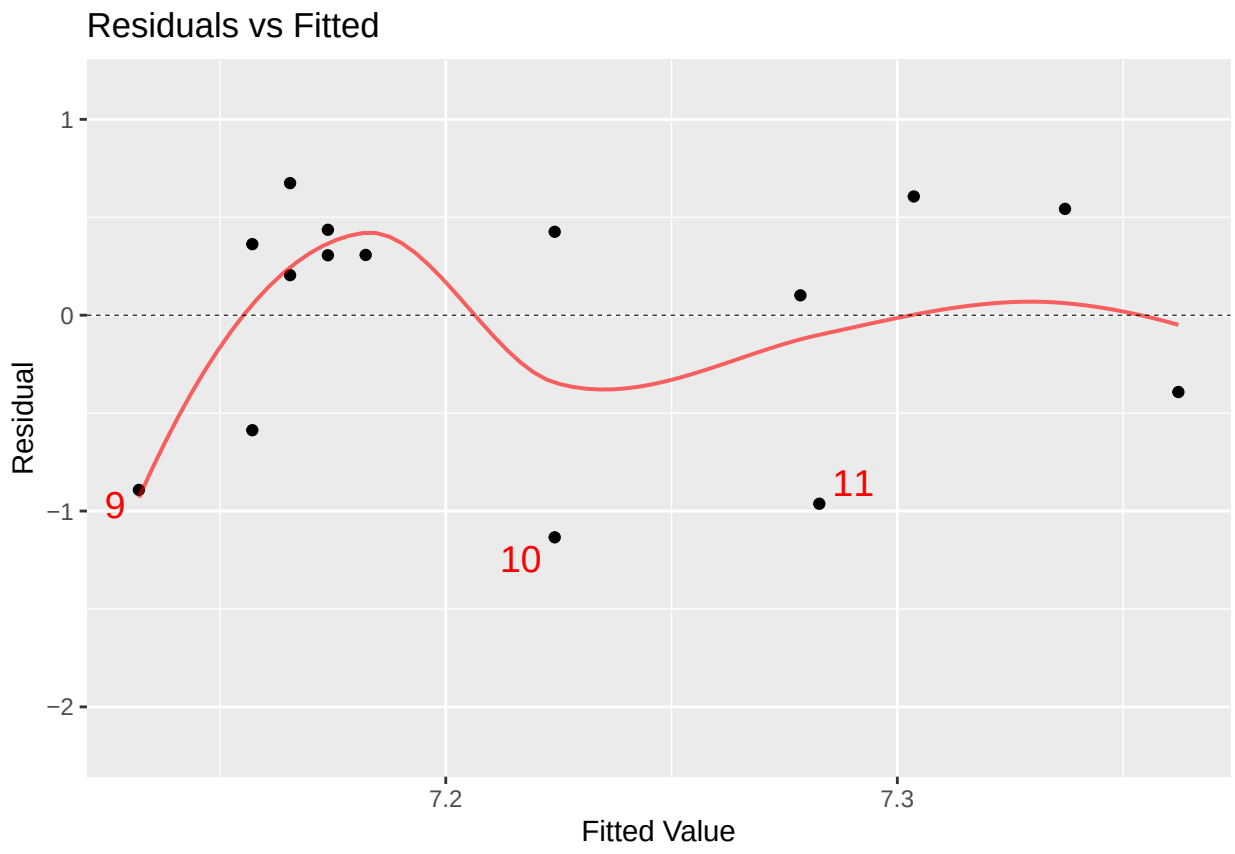

```
#ii)
model1 <- lm (Consumer_Debt ~ Household_Debt,
              data = debt)
print(model1)
```

```
##
## Call:
## lm(formula = Consumer_Debt ~ Household_Debt, data = debt)
```

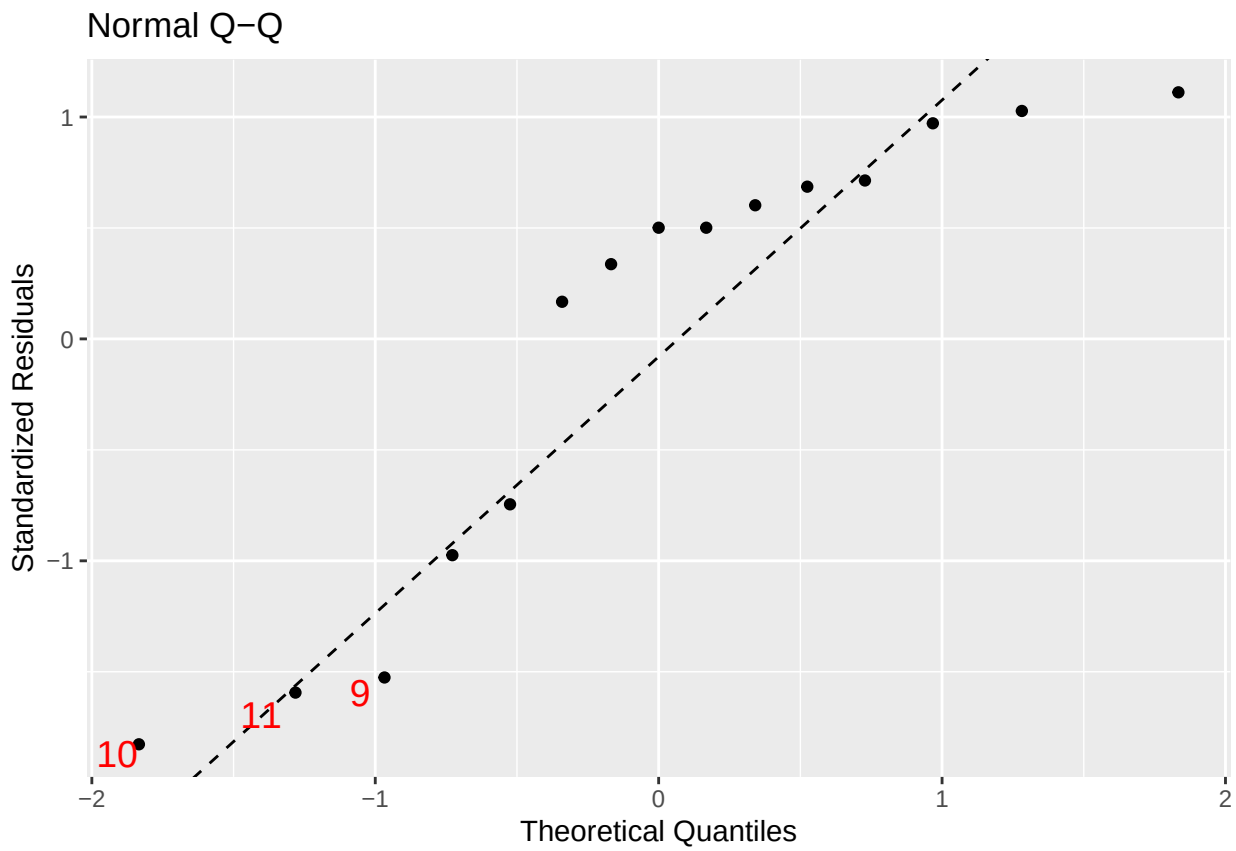
```
##
## Coefficients:
##    (Intercept)  Household_Debt
##         4.7338          0.4185
```

```
mplot(model1, which = 1)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
mplot(model1, which = 2)
```




```
# Interpretation: zero is still a plausible value from the interval test test above because zero is wit

# iii)
confint(model1)
```

```
##                    2.5 %    97.5 %
## (Intercept)    -7.834614 17.302150
## Household_Debt -1.695295  2.532389
```

```
# iv)
msummary(model1)
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.7338     5.8177   0.814    0.430
## Household_Debt   0.4185     0.9785   0.428    0.676
##
## Residual standard error: 0.6426 on 13 degrees of freedom
## Multiple R-squared:  0.01388,    Adjusted R-squared:  -0.06198
## F-statistic: 0.183 on 1 and 13 DF,  p-value: 0.6758
```

```
# Conclusion: We fail to reject the null hypothesis do to lack of evidence

# last R code line
```

2. The data below contains sale price, size, and land-to-building ratio for 10 large industrial properties

```
saleprice <- read.csv("https://www.siue.edu/~jpailde/saleprice.csv")
saleprice
```

```
##    Property price...M. Size..k.sq.ft. Ratio
## 1         1       10.6           2166   2.0
## 2         2        2.6            751   3.5
## 3         3       30.5           2422   3.6
## 4         4        1.8            224   4.7
## 5         5       20.0           3917   1.7
## 6         6        8.0           2866   2.3
## 7         7       10.0           1698   3.1
## 8         8        6.7           1046   4.8
## 9         9        5.8           1108   7.6
## 10       10        4.5            405  17.2
```

i) Construct a scatterpot for `sale price versus size` and `sale price versus land-to-building ratio`. Be sure to fit regression lines on the scatterplots.

ii) Use the `lm` function to estimated the equations of each regression model for `sale price versus size` and `sale price versus land-to-building ratio`.

iii) Check the error model assumption visually by constructing a residual plot and QQplot of the residuals for the two models.

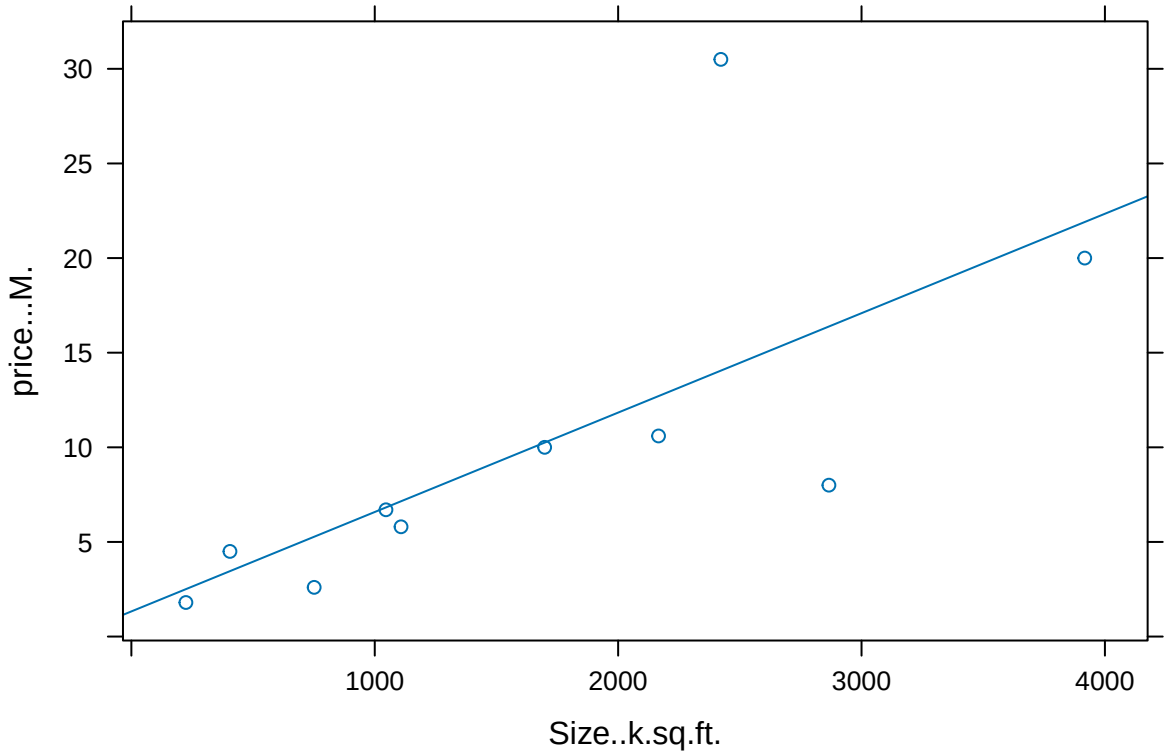
iv) Estimate the population regression slope of each model (line) by constructing 95% confidence interval. Give a brief interpretation of the esimated slope in the context of the problem.

v) Perform a hypothesis test on the regression slope of each model (line), use a 5% level of significance. Given an appropriate conclusion.

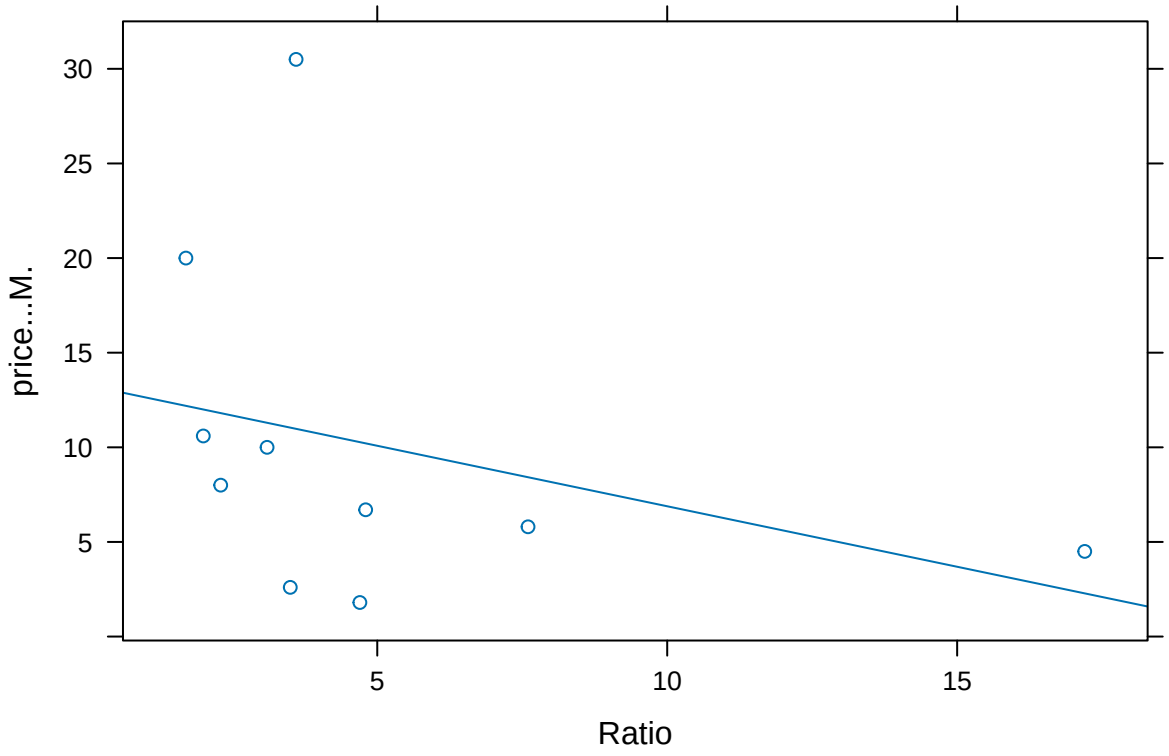
Code chunk

```
# start your code

# i)
xyplot(price...M. ~ Size..k.sq.ft.,
       data = saleprice,
       type = c('p', 'r'))
```




```
xyplot(price...M. ~ Ratio,
       data = saleprice,
       type = c('p', 'r'))
```




```
# ii)
model1 <- lm(price...M. ~ Size..k.sq.ft.,
             data = saleprice)
model1
```

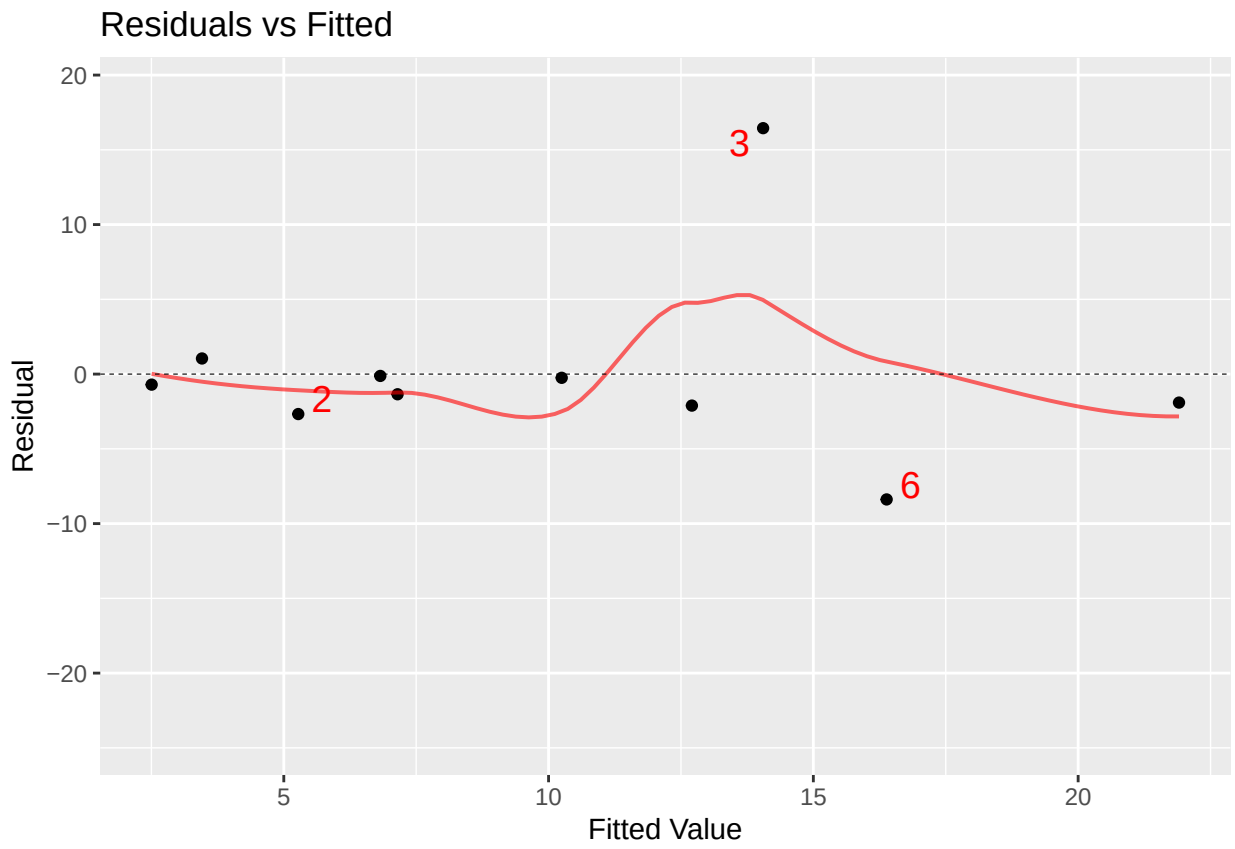
```
##
## Call:
## lm(formula = price...M. ~ Size..k.sq.ft., data = saleprice)
##
## Coefficients:
##    (Intercept)  Size..k.sq.ft.
##       1.328142        0.005253
```

```
model2 <- lm(price...M. ~ Ratio,
             data = saleprice)
model2
```

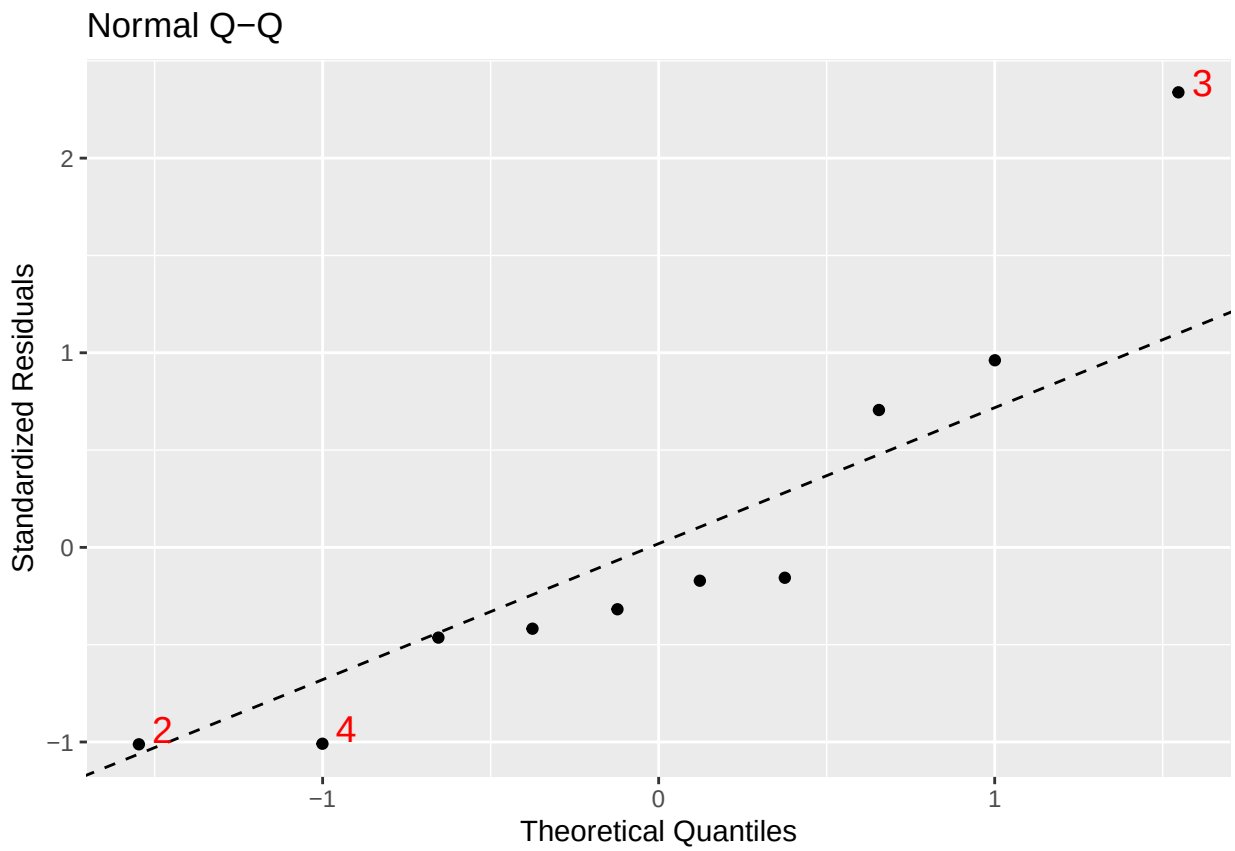
```
##
## Call:
## lm(formula = price...M. ~ Ratio, data = saleprice)
##
## Coefficients:
## (Intercept)        Ratio
##     13.2787      -0.6393
```

```
# iii)
mplot(model1, which = 1)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

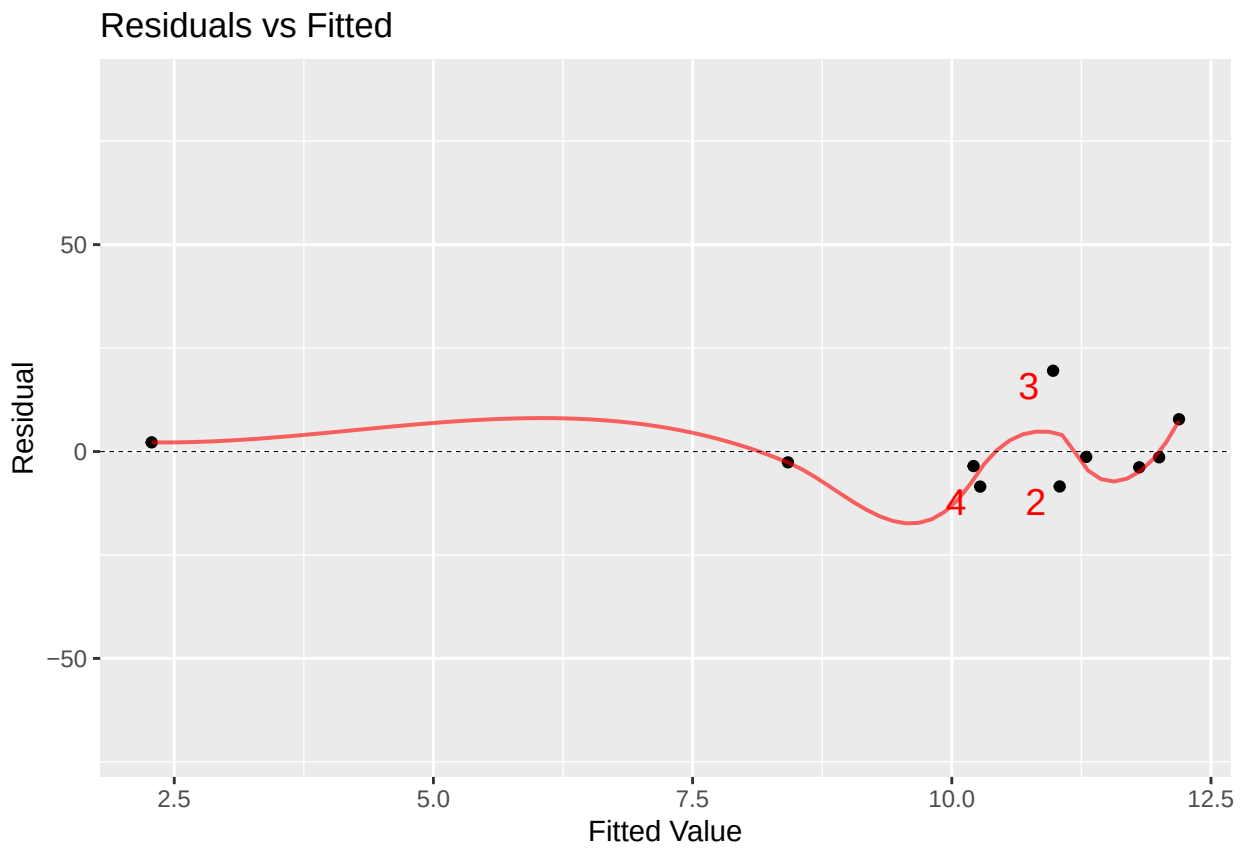



```
mplot(model2, which = 2)
```

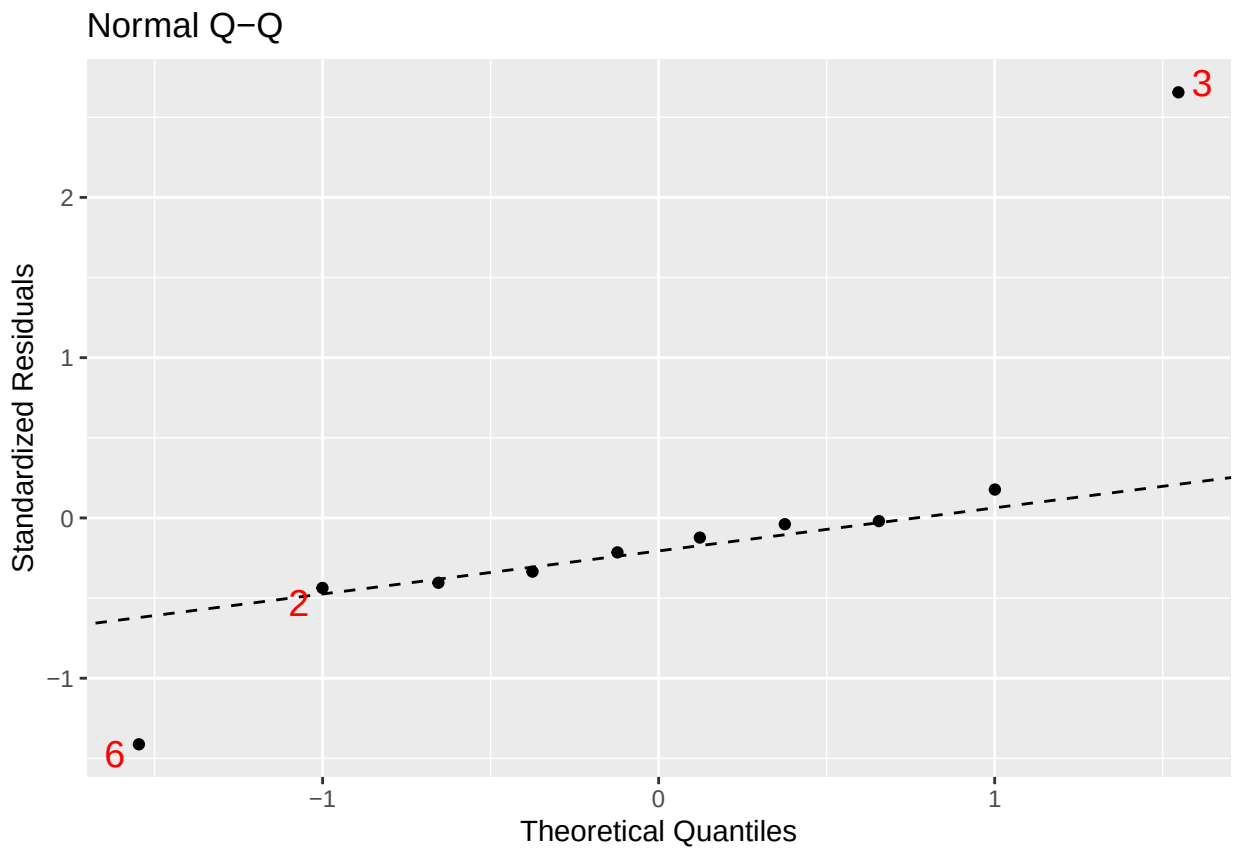



```
mplot(model2, which = 1)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
mplot(model1, which = 2)
```




```
# iv)
confint(model1)
```

```
##                        2.5 %       97.5 %
## (Intercept)    -7.4165326479 10.072817539
## Size..k.sq.ft.  0.0008861725  0.009620191
```

```
# Interpretation: Zero is out of the bounds for model1 so we reject
confint(model2)
```

```
##                 2.5 %     97.5 %
## (Intercept)  3.405707 23.1516090
## Ratio       -2.118144  0.8394672
```

```
# Interpretation: Zero (mu) is in the bound so we fail to reject

# v)
msummary(model1)
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.328142   3.792133   0.350   0.7352
## Size..k.sq.ft. 0.005253   0.001894   2.774   0.0241 *
##
## Residual standard error: 6.704 on 8 degrees of freedom
## Multiple R-squared:  0.4903, Adjusted R-squared:  0.4266
## F-statistic: 7.695 on 1 and 8 DF,  p-value: 0.02415
```

```r
# Interpretation: p-value is less than alpha so we reject the null hypothesis

msummary(model2)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.2787     4.2814   3.101   0.0146 *
## Ratio        -0.6393     0.6413  -0.997   0.3480
##
## Residual standard error: 8.856 on 8 degrees of freedom
## Multiple R-squared:  0.1105, Adjusted R-squared:  -0.000674
## F-statistic: 0.9939 on 1 and 8 DF,  p-value: 0.348
```

```r
# Interpretation: Since p-value is greater than alpha we fail to reject the null hypothesis

# Summary the size seems to have an effect on the price but the ratio does not.

# last R code line
```