Forecast Accuarcy

Ebraheem Amashah

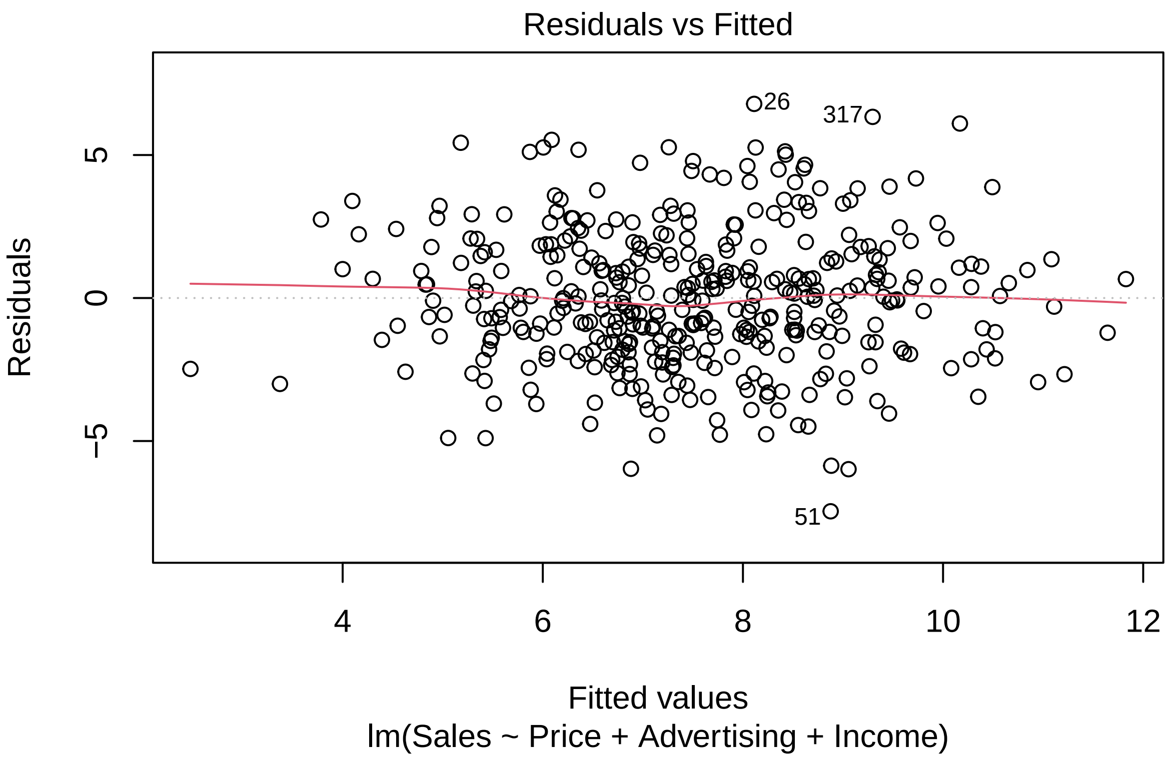
8/3/2021

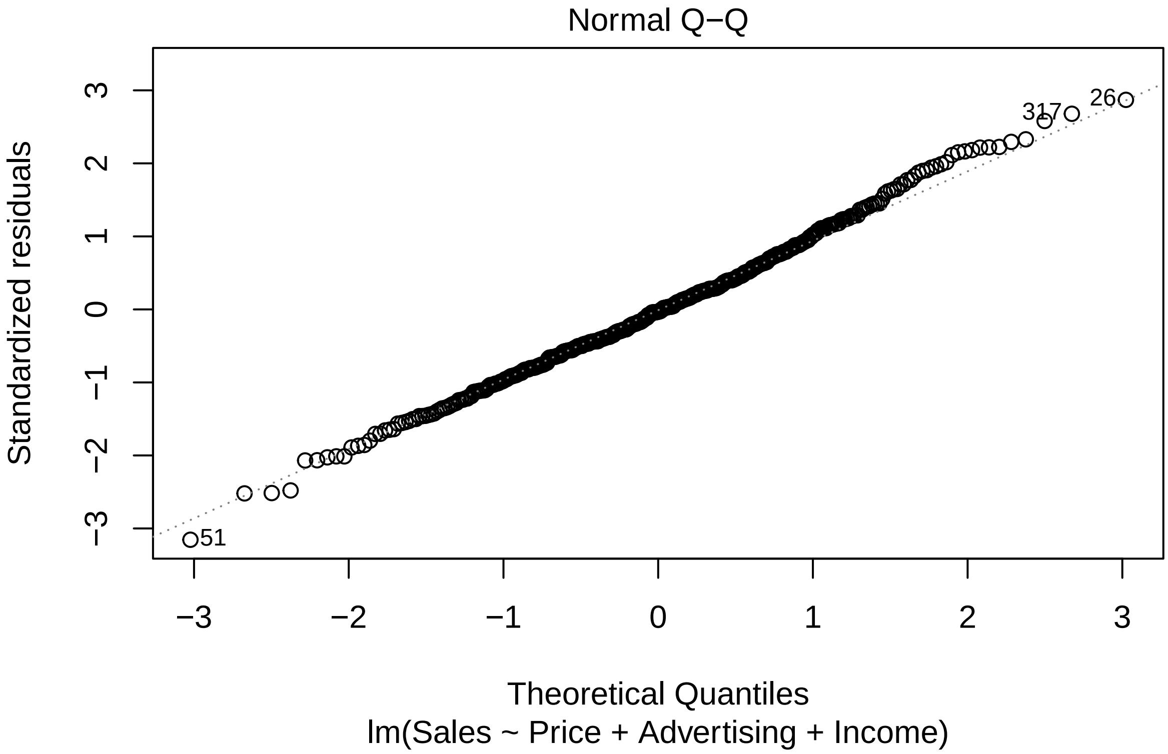
```
data(Carseats)

summary(Carseats)
```

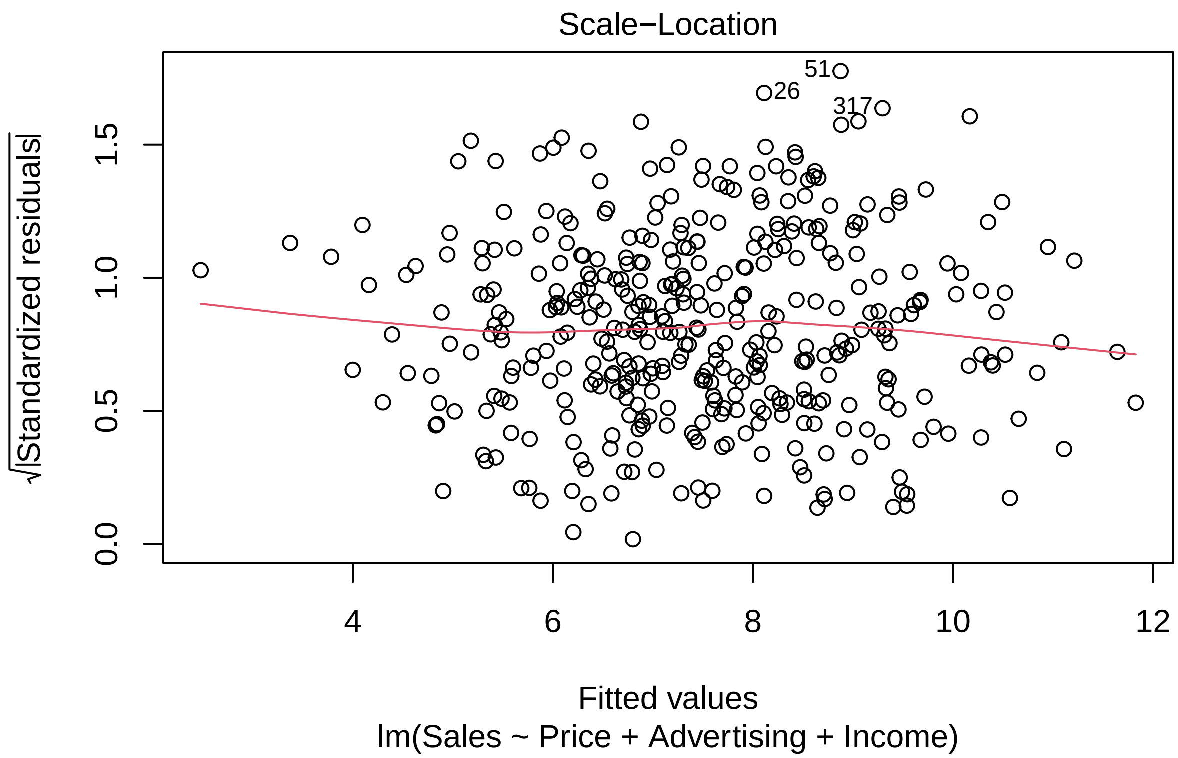
```
##      Sales          CompPrice       Income        Advertising
##  Min.   : 0.000   Min.   : 77   Min.   : 21.00   Min.   : 0.000
##  1st Qu.: 5.390   1st Qu.:115   1st Qu.: 42.75   1st Qu.: 0.000
##  Median : 7.490   Median :125   Median : 69.00   Median : 5.000
##  Mean   : 7.496   Mean   :125   Mean   : 68.66   Mean   : 6.635
##  3rd Qu.: 9.320   3rd Qu.:135   3rd Qu.: 91.00   3rd Qu.:12.000
##  Max.   :16.270   Max.   :175   Max.   :120.00   Max.   :29.000
##    Population        Price        ShelveLoc        Age          Education
##  Min.   : 10.0   Min.   : 24.0   Bad   : 96   Min.   :25.00   Min.   :10.0
##  1st Qu.:139.0   1st Qu.:100.0   Good  : 85   1st Qu.:39.75   1st Qu.:12.0
##  Median :272.0   Median :117.0   Medium:219   Median :54.50   Median :14.0
##  Mean   :264.8   Mean   :115.8                Mean   :53.32   Mean   :13.9
##  3rd Qu.:398.5   3rd Qu.:131.0                3rd Qu.:66.00   3rd Qu.:16.0
##  Max.   :509.0   Max.   :191.0                Max.   :80.00   Max.   :18.0
##  Urban       US
##  No :118   No :142
##  Yes:282   Yes:258
##
##
##
##
```

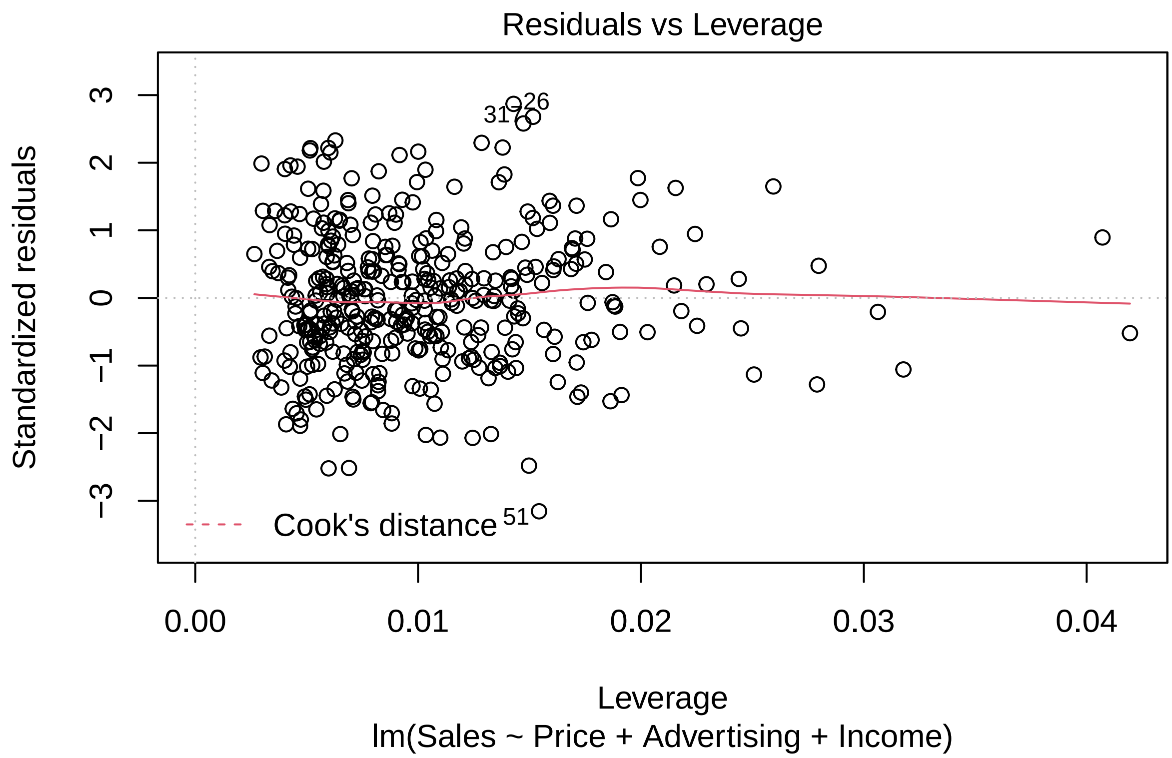
A. Model W/ Price, Advertising, and Income

```
lm.fit=lm(Sales~Price+Advertising+Income, data=Carseats)
plot(lm.fit)
```


Residuals vs Fitted
Residuals
Fitted values
lm(Sales ~ Price + Advertising + Income)
5
0
−5
4
6
8
10
12
26
317
51


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Sales ~ Price + Advertising + Income)
−3
−2
−1
0
1
2
3
51
317
26


Scale−Location
51
26
317
√|Standardized residuals|
0.0
0.5
1.0
1.5
4
6
8
10
12
Fitted values
lm(Sales ~ Price + Advertising + Income)




```
lm.fit
```

```
##
## Call:
## lm(formula = Sales ~ Price + Advertising + Income, data = Carseats)
##
## Coefficients:
## (Intercept)        Price  Advertising       Income
##    12.17270     -0.05384      0.12024      0.01107
```

B.Coefficients Interpretation

$$\beta_1$$

Price: Negative relationship,each 1 unit increase in price yeilds a 0.05 *decrease* on Sales

$$\beta_2$$

Advertising: Positive Relationship, each 1 unite increase yeilds a 0.12 unit *increase* in sales

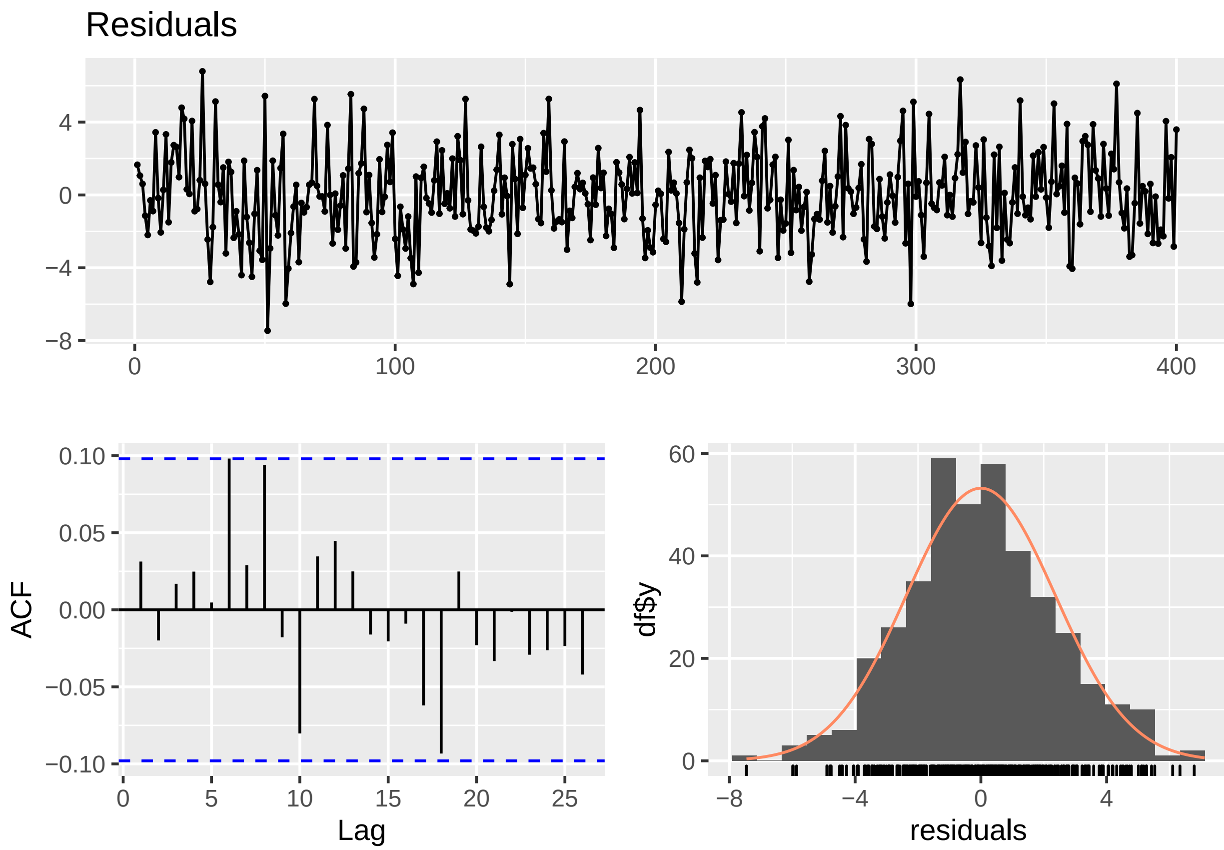
$$\beta_3$$

Income: Positive relationship, each 1 unit increase in income, yeilds 0.01 *increase* in Sales

C.Model Equation

$$Sales = \beta_0 + \beta_1 Price + \beta_2 Advertising + \beta_3 Income + \epsilon$$

```
checkresiduals(lm.fit)
```




```
##
##  Breusch-Godfrey test for serial correlation of order up to 10
##
## data:  Residuals
## LM test = 11.86, df = 10, p-value = 0.2945
```

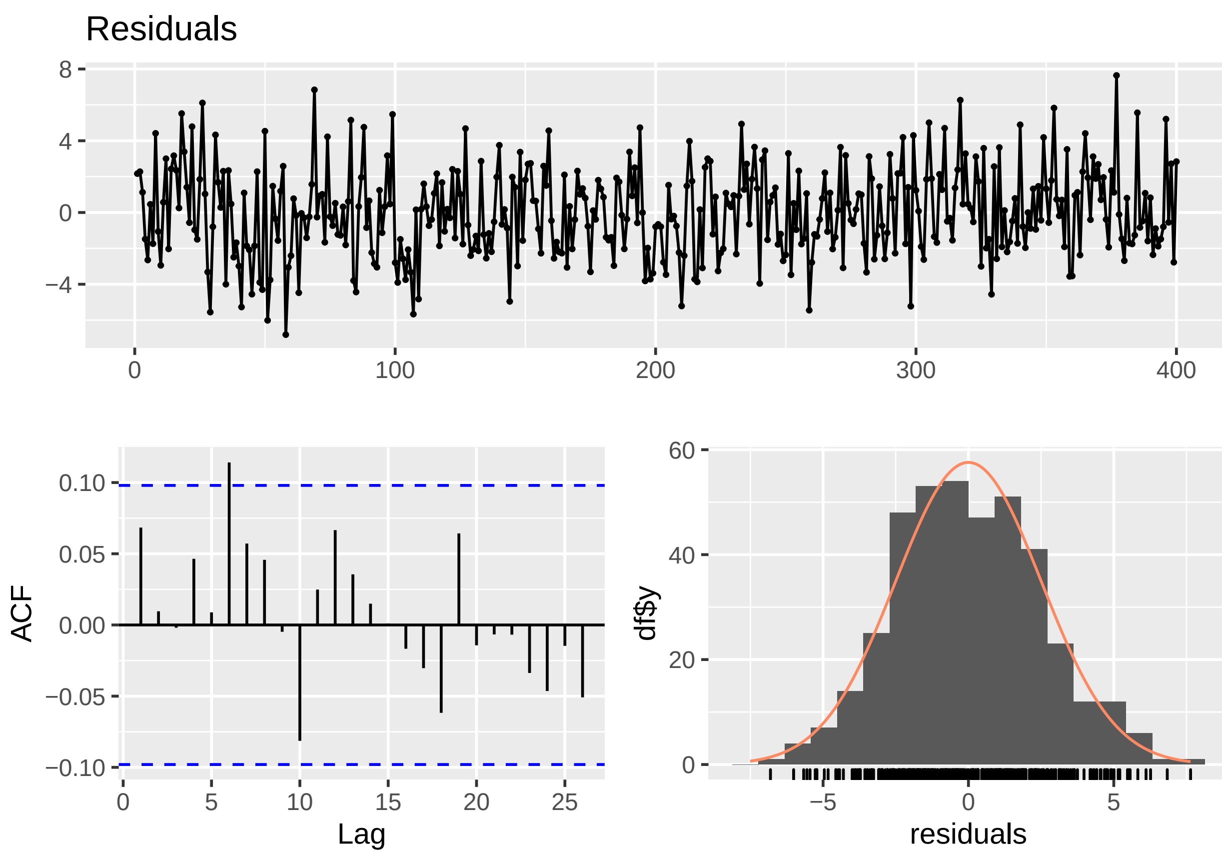
since p value is >0.05, we do not have neough evidence to reject null hypothesis for any predictor.

D.Model W/ Only Price and Income

```
lm.fit2=lm(Sales~Price+Income, data=Carseats)
lm.fit2
```

```
##
## Call:
## lm(formula = Sales ~ Price + Income, data = Carseats)
##
## Coefficients:
## (Intercept)        Price       Income
##    12.66155     -0.05221      0.01283
```

```
checkresiduals(lm.fit2)
```




```
##
##  Breusch-Godfrey test for serial correlation of order up to 10
##
## data:  Residuals
## LM test = 12.913, df = 10, p-value = 0.2286
```

E.How well do the 2 models fit the Data? *First Model*

```
str(summary(lm.fit))
```

```
## List of 11
##  $ call         : language lm(formula = Sales ~ Price + Advertising + Income, data = Carseats)
##  $ terms        :Classes 'terms', 'formula'  language Sales ~ Price + Advertising + Income
##   .. ..- attr(*, "variables")= language list(Sales, Price, Advertising, Income)
##   .. ..- attr(*, "factors")= int [1:4, 1:3] 0 1 0 0 0 0 1 0 0 0 ...
##   .. .. ..- attr(*, "dimnames")=List of 2
##   .. .. .. ..$ : chr [1:4] "Sales" "Price" "Advertising" "Income"
##   .. .. .. ..$ : chr [1:3] "Price" "Advertising" "Income"
##   .. ..- attr(*, "term.labels")= chr [1:3] "Price" "Advertising" "Income"
##   .. ..- attr(*, "order")= int [1:3] 1 1 1
##   .. ..- attr(*, "intercept")= int 1
##   .. ..- attr(*, "response")= int 1
```

```
##   .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
##   .. ..- attr(*, "predvars")= language list(Sales, Price, Advertising, Income)
##   .. ..- attr(*, "dataClasses")= Named chr [1:4] "numeric" "numeric" "numeric" "numeric"
##   .. .. ..- attr(*, "names")= chr [1:4] "Sales" "Price" "Advertising" "Income"
##  $ residuals    : Named num [1:400] 1.657 1.061 0.604 -1.138 -2.201 ...
##   ..- attr(*, "names")= chr [1:400] "1" "2" "3" "4" ...
##  $ coefficients : num [1:4, 1:4] 12.1727 -0.0538 0.1202 0.0111 0.6827 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:4] "(Intercept)" "Price" "Advertising" "Income"
##   .. ..$ : chr [1:4] "Estimate" "Std. Error" "t value" "Pr(>|t|)"
##  $ aliased      : Named logi [1:4] FALSE FALSE FALSE FALSE
##   ..- attr(*, "names")= chr [1:4] "(Intercept)" "Price" "Advertising" "Income"
##  $ sigma        : num 2.38
##  $ df           : int [1:3] 4 396 4
##  $ r.squared    : num 0.294
##  $ adj.r.squared: num 0.288
##  $ fstatistic   : Named num [1:3] 54.9 3 396
##   ..- attr(*, "names")= chr [1:3] "value" "numdf" "dendf"
##  $ cov.unscaled : num [1:4, 1:4] 0.082136 -0.000531 -0.000232 -0.000242 -0.000531 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:4] "(Intercept)" "Price" "Advertising" "Income"
##   .. ..$ : chr [1:4] "(Intercept)" "Price" "Advertising" "Income"
##  - attr(*, "class")= chr "summary.lm"
```

First Model (Price, Advertising and Income), is only able to fit %29.4 of the data.

Second Model

```
str(summary(lm.fit2))
```

```
## List of 11
##  $ call         : language lm(formula = Sales ~ Price + Income, data = Carseats)
##  $ terms        :Classes 'terms', 'formula'  language Sales ~ Price + Income
##   .. ..- attr(*, "variables")= language list(Sales, Price, Income)
##   .. ..- attr(*, "factors")= int [1:3, 1:2] 0 1 0 0 0 1
##   .. .. ..- attr(*, "dimnames")=List of 2
##   .. .. .. ..$ : chr [1:3] "Sales" "Price" "Income"
##   .. .. .. ..$ : chr [1:2] "Price" "Income"
##   .. ..- attr(*, "term.labels")= chr [1:2] "Price" "Income"
##   .. ..- attr(*, "order")= int [1:2] 1 1
##   .. ..- attr(*, "intercept")= int 1
##   .. ..- attr(*, "response")= int 1
##   .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
##   .. ..- attr(*, "predvars")= language list(Sales, Price, Income)
##   .. ..- attr(*, "dataClasses")= Named chr [1:3] "numeric" "numeric" "numeric"
##   .. .. ..- attr(*, "names")= chr [1:3] "Sales" "Price" "Income"
##  $ residuals    : Named num [1:400] 2.17 2.28 1.13 -1.48 -2.65 ...
##   ..- attr(*, "names")= chr [1:400] "1" "2" "3" "4" ...
##  $ coefficients : num [1:3, 1:4] 12.66155 -0.05221 0.01283 0.71519 0.00532 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3] "(Intercept)" "Price" "Income"
##   .. ..$ : chr [1:4] "Estimate" "Std. Error" "t value" "Pr(>|t|)"
##  $ aliased      : Named logi [1:3] FALSE FALSE FALSE
##   ..- attr(*, "names")= chr [1:3] "(Intercept)" "Price" "Income"
```

```
## $ sigma        : num 2.51
## $ df           : int [1:3] 3 397 3
## $ r.squared    : num 0.214
## $ adj.r.squared: num 0.21
## $ fstatistic   : Named num [1:3] 54.1 2 397
##  ..- attr(*, "names")= chr [1:3] "value" "numdf" "dendf"
## $ cov.unscaled : num [1:3, 1:3] 8.12e-02 -5.34e-04 -2.45e-04 -5.34e-04 4.49e-06 ...
##  ..- attr(*, "dimnames")=List of 2
##  .. ..$ : chr [1:3] "(Intercept)" "Price" "Income"
##  .. ..$ : chr [1:3] "(Intercept)" "Price" "Income"
## - attr(*, "class")= chr "summary.lm"
```

Second Model(Price and Income only), is only able to fit %21.4 of the data. also comparing the AIC and BIC of both models we can see the first model is better at fitting the data;

```
AIC(lm.fit)
```

```
## [1] 1835.565
```

```
AIC(lm.fit2)
```

```
## [1] 1876.34
```

```
BIC(lm.fit)
```

```
## [1] 1855.522
```

```
BIC(lm.fit2)
```

```
## [1] 1892.305
```

Keeping the Advertising predictor does improve the accuracy of our model.