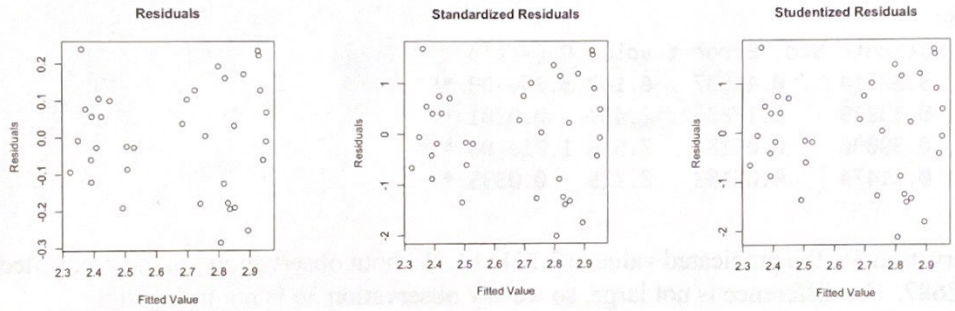
Yunlu Li
STAT 5120
Homework 8

Question 1

(a) *Share* = 3.18527 – 0.35269**price* + 0.39914**discount* + 0.11803**promo*

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.18527    0.36505   8.726  5.7e-10 ***
price       -0.35269    0.15738  -2.241   0.0321 *
discount     0.39914    0.05125   7.787  7.0e-09 ***
promo        0.11803    0.05149   2.292   0.0286 *
```

(b)
(i) (ii) (iii)

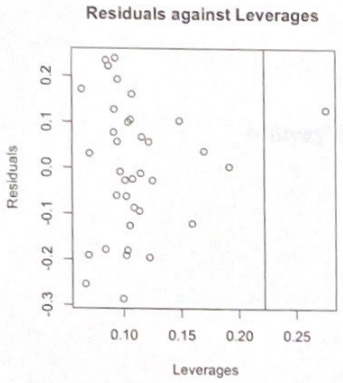


They all have similar patterns. The only difference is the range of plots. The standardized residuals and studentized residuals have different residuals from residuals is due to the standardizing nature of them.

(c) There is no outlying values in the response variable because no studentized residuals are greater than $t_{1-0.05/72,31}$.

```
> sort(student.res)
          8           3          23          32          30          29          28           5
-2.12242537 -1.81561462 -1.41335887 -1.36702714 -1.33878506 -1.28239802 -1.25645673 -0.88572450
         21          27          10          18          22          14           6          26
-0.87716804 -0.65828196 -0.60383799 -0.42918175 -0.41510328 -0.19006148 -0.18650762 -0.16783045
         19          17          24          33          20          15          35           7
-0.08079021 -0.05769732  0.02794362  0.21435971  0.27018938  0.39866815  0.40329435  0.47505038
          9           4           2          34          16          36          13          25
 0.53806408  0.70117575  0.74130342  0.74485268  0.89293085  1.00753169  1.14533207  1.19100514
         31           1          11          12
 1.37450335  1.59405716  1.68906204  1.73495234
> qt(1-0.05/72, 31)
[1] 3.51167
```

(d) Observation 36 has high leverages, which is greater than 2p/n. Here p = 4 and n = 36.




```
> lev[lev>2*4/36]
       36
0.2759335
```

(e) *Share* = 3.38614 - 0.43839**price* + 0.39096**discount* + 0.11474**promo*
The estimated regression equation does not differ greatly, so observation 36 is not influential.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.38614    0.41587   8.142 3.39e-09 ***
price       -0.43839    0.17886  -2.451   0.0201 *
discount     0.39096    0.05188   7.536 1.71e-08 ***
promo        0.11474    0.05158   2.225   0.0335 *
```

(f) With observation 36, the predicated value is 2.721614. Without observation 36, the predicated value is 2.672687. The difference is not large, so we say observation 36 is not influential.

(g) The observation 8 has DFFITS = -0.7062225, and its absolute value is greater than 2*sqrt(4/36), so observation 8 is influential based on DFFITS.
DFFITSi measures the dierence in predicted values for observation i, with and without observation i in the dataset.

(h) None of observations has Cook's distance larger than $F_{0.5,4,32}$, so no observation is influential based on Cook's distance.
Cook's distance examines the impact of a single observation on all the regression coefficients, with and without observation i in the dataset.

estion 2

(a). $t_6 = \frac{e_6}{\sqrt{MSE_{(6)}(1-h_{66})}} = \frac{120.82907}{\sqrt{22.6^2 \cdot (1-0.2396051)}} = 6.13$

(b) $h_{66} = 0.2396051$.

$\frac{2p}{n} = \frac{4}{19} = 0.2105263$.

$h_{66} > \frac{2p}{n}$ $\Rightarrow$ observation 6 has high leverage and thus is a outlier.

(c) $DFFITS_6 = t_6 \left(\frac{h_{66}}{1-h_{66}}\right)^{\frac{1}{2}} = 3.44$. Higher leverage makes DFFITs larger.

(d) $D_6 = \frac{e_6^2}{p \cdot MSE}\left(\frac{h_{66}}{(1-h_{66})^2}\right) = \frac{120.82907^2}{2 \cdot 40.13^2}\left(\frac{0.2396051}{(1-0.2396051)^2}\right) = 1.878$

(e). $F_{0.5,2,17} = 0.7221933$.

$D_6 > F_{0.5,2,17}$ $\Rightarrow$ it's influential.