Alexander Martinez

Homework 3

6.3 While it is true we can never reduce R^2 by adding predictors, it does not make it a good thing. In fact, this is one of the big downside's of R^2 in MLR. You can include predictor variable's that do not effect the model at all and still increase R^2, which would make the regression a bad fit.

6.9 a. In R,
The stem plot for cases shipped (X1) the most concentrated # of cases shipped are between 240,000 and 320,000. Some go as high a 472,000, and as low as 210,000. There are a few scattered in between all those value sets

In R,
The stem for (X2) shows that the indirect costs of total labor hours as a % are concentrated between 6.1 and 8.3. It goes as low a 4.6 and as high as 9.6. There is a big gap between 4.6 and 6.1.

b. In R
(X1) - The seem to be close to one another during the early weeks, then they scatter and go in a positive direction in later weeks

★★ Accidentally did b. instead of c. C. is top of Next Page ★★

(X2) – It seems to be pretty consistent through all 52 weeks

(X3) – As expected the Holiday weeks ($X_{i3}=1$) are sporadic, but it is good to note that there is about a 20 week period of no holiday weeks.

6.10 a. In R

$\hat{Y} = 4149.887 + .0007870804 X_1 + -13.16602 X_2 + 623.5545 X_3$

b_1 = For every 1 unit of X_1 increased, the mean of total labor hours (Y) increases by .0007870804, holding all other variables constant

b_2 = For every 1 unit increase in X_2, the mean of the total labor hours (Y) decreases by -13.16602, holding all other variables constant

b_3 = For every 1 unit increase in X_3, the mean of the total labor hours (Y) increases by 623.5545, holding all other variables constant.

b. The boxplot shows the residuals are pretty evenly distributed on both sides of ∅

6.9 (skipped c., so I put it here)

C. a. The Scatterplot matrix shows a slight (+) trend w/ cases, a sligt (-) trend w/ X_2 and a really strong (+) trend w/ X_3

The correlation matrix shows a little Corr. w/ X_1, a tiny bit of corr w/ X_2 and a large corr. w/ X_3

6.10c. In R,

Every single residual plot seems to have no pattern, which is good. The one thing to note is that binary variable (X_3) seems to separate the data into 2, which is expected.

As for the NPP, the plot seems to check out. The points form nearly a linear line, which indicates the normal distribution is a good model for the set.

e. Done In R, (sorted by $\hat{Y}$) we tested whether there was significant departure from constant σ^2.

Pval = .3368 > .01

No sig. departure.

6.11

a. In R, test of regression Relation.

$H_0: \beta_1 = 0$ and $\beta_2 = 0$ and $\beta_3 = 0$
H_a: Not all β_1, β_2, and $\beta_3 = 0$

$F((.95); 3, 48) = 2.798061$

$F = 35.337 > 2.798061$

Conclude H_a, Total labor Hours are related to # of cases shipped, indirect costs of total labor hours as a %, and holiday weeks.

b. In R
$-.00003654 \leq \beta_1 \leq .001630748$

Do Not Reject Null, $\emptyset$ is included in the CI for β_1

$478.6896 \leq \beta_3 \leq 768.499$

Reject Null.

We cannot conclude w/ 95% confidence, adjusting for Bonferroni, that $\beta_1 \neq 0$.

We can conclude for β_3 that $\beta_3 \neq 0$ at a 95% conf. Level

In R

C. $R^2 = .6883$

68.83% of total variation in the Y_i's are explained by the X_i's in the model

6.12a. In R

$S = 3.7033$

$B = 2.6822$

$B < S$, so we use Bonferroni

The family of estmates can be found in R

b. Yes, by looking at the df, that would be within the scope of the model.

No, there is not one week where the percentage got to 9.9. The only time it got close was during a holiday week

6.13 In R

$S = 3.203274$

$B = 2.395323$

The family of prediction Intervals can be found in the R output.

6.14 a. Using R

$3986.826 \leq \bar{Y}_{h(new)} \leq 4570.104$

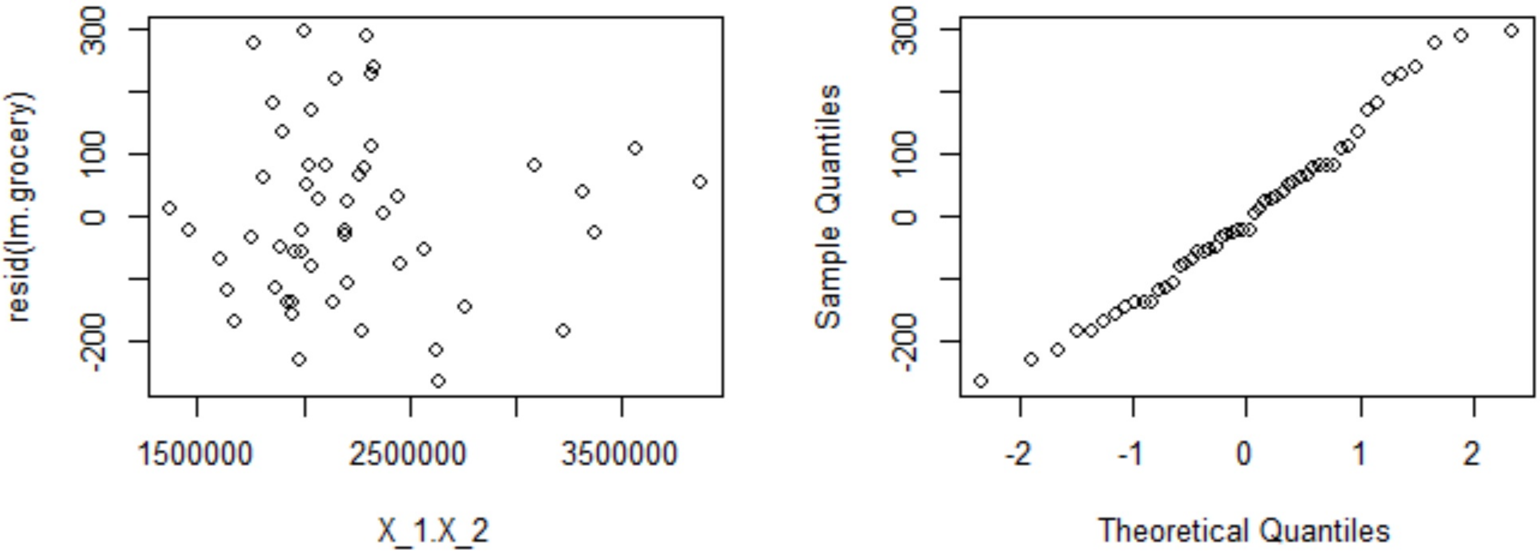
6.22 a. Yes

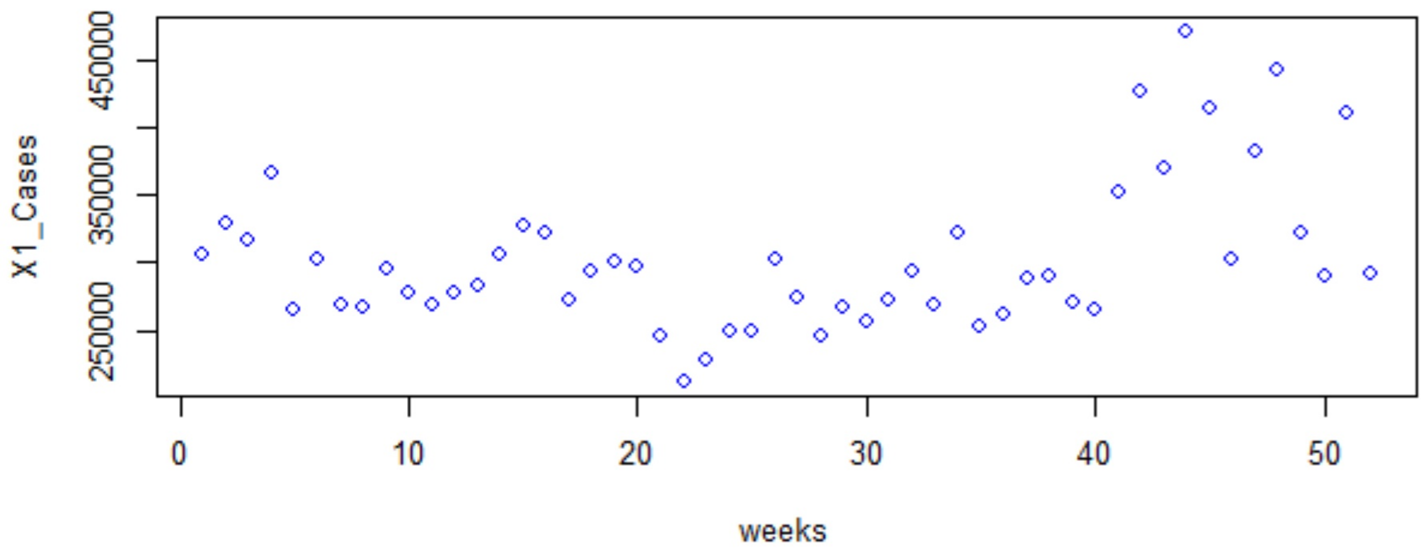
b. No, but you can take the log of both sides, get rid of error multiplying term

c. No

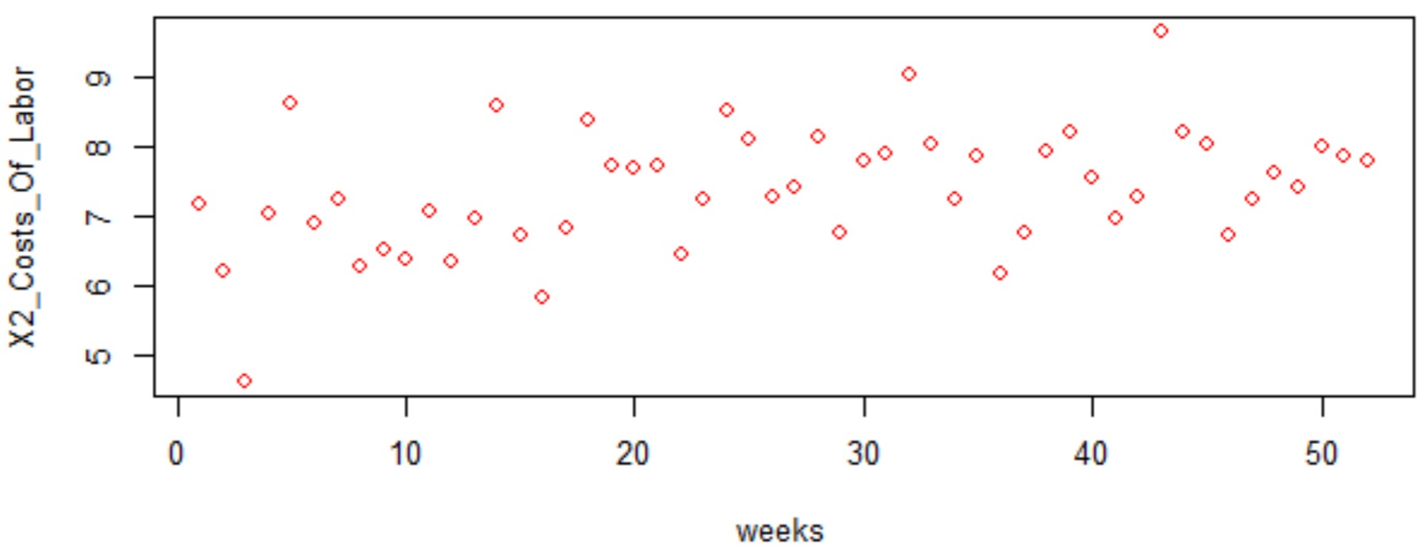
d. No, intercept is a function of β_1

e Yes.

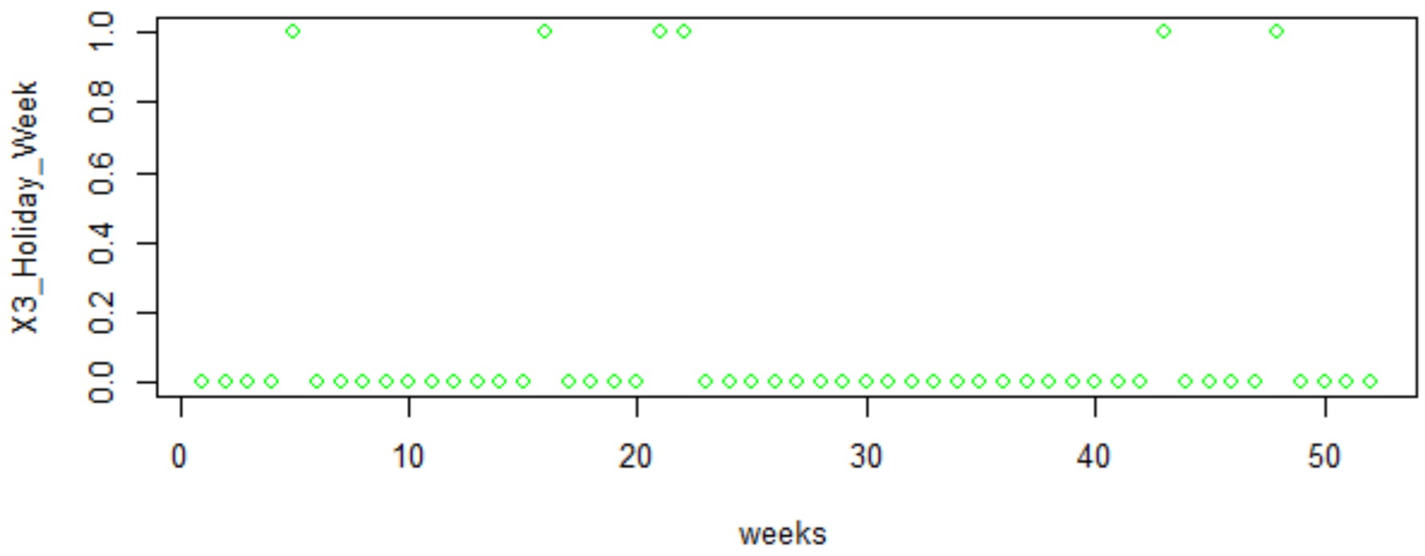

resid(lm.grocery)
300
100
0
-200
1500000
2500000
3500000
X_1.X_2
Sample Quantiles
300
100
0
-200
-2
-1
0
1
2
Theoretical Quantiles



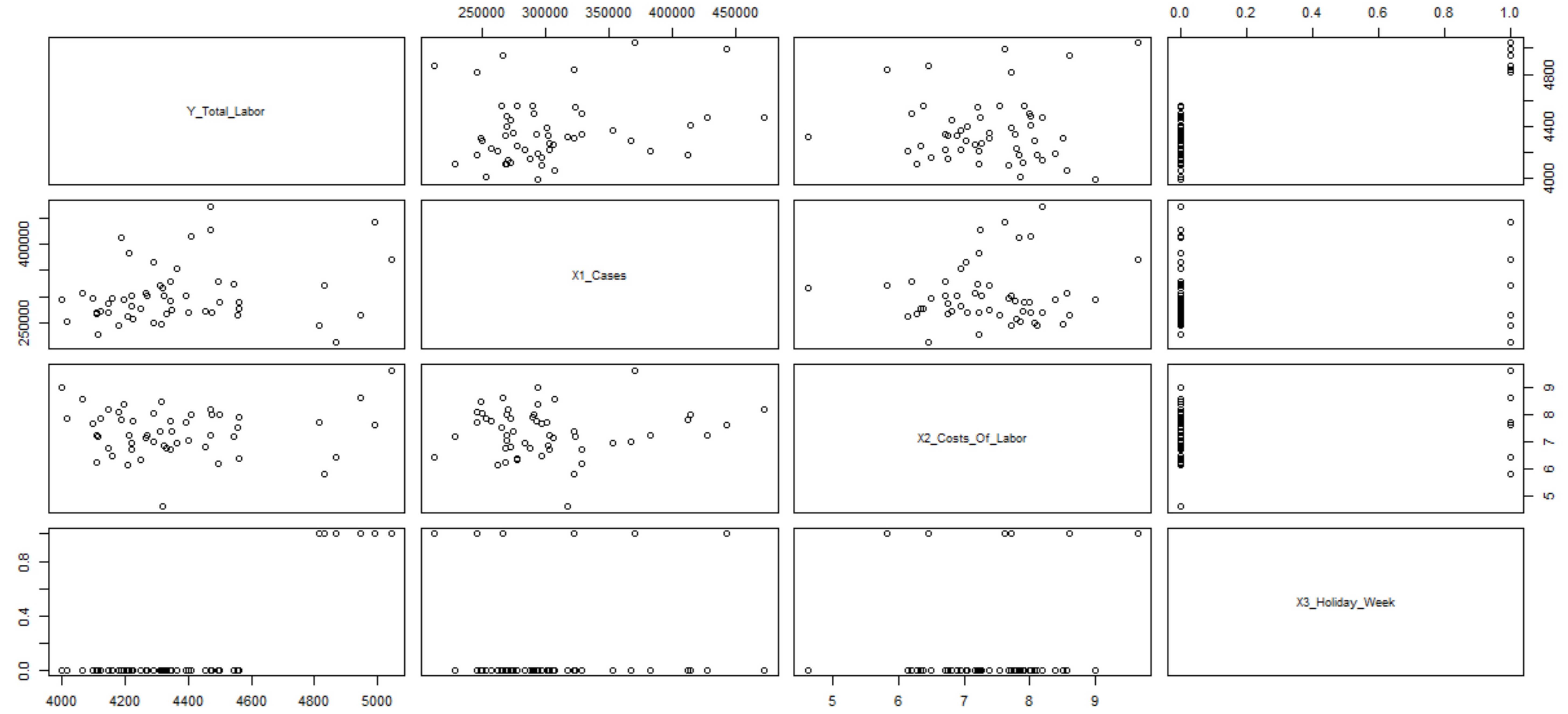
X1_Cases
250000
350000
450000
0
10
20
30
40
50
weeks

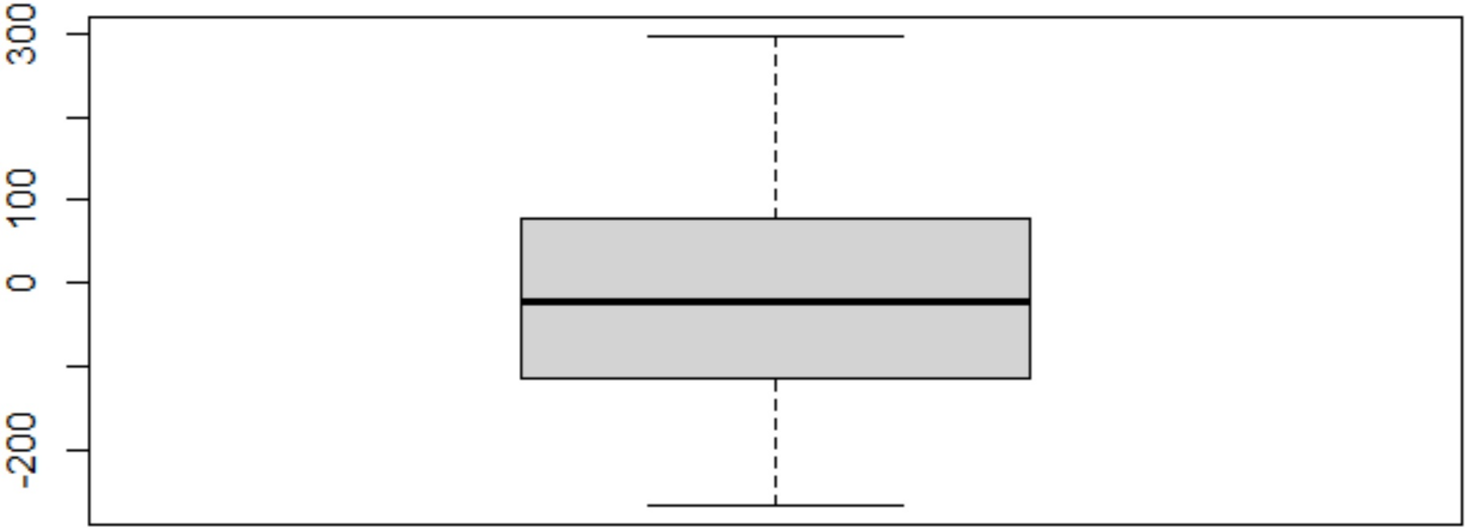


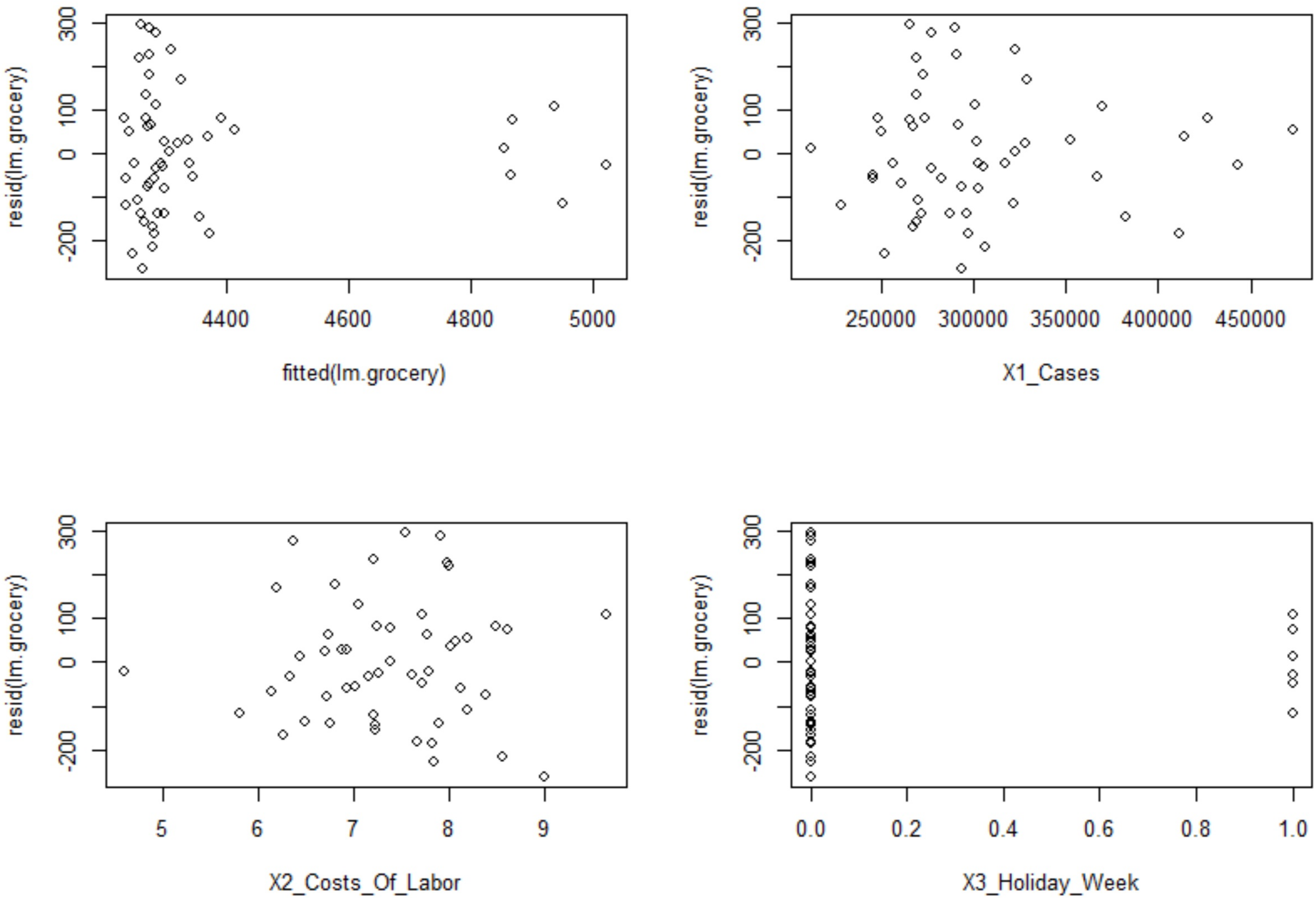
X2_Costs_Of_Labor
5
6
7
8
9
0
10
20
30
40
50
weeks



X3_Holiday_Week
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
weeks


Y_Total_Labor
X1_Cases
X2_Costs_Of_Labor
X3_Holiday_Week
250000
300000
350000
400000
450000
0.0
0.2
0.4
0.6
0.8
1.0
4000
4400
4800
5
6
7
8
9
4200
4600


300
100
0
-200



resid(lm.grocery)
300
100
0
-200
4400
4600
4800
5000
fitted(lm.grocery)
250000
300000
350000
400000
450000
X1_Cases
5
6
7
8
9
X2_Costs_Of_Labor
0.0
0.2
0.4
0.6
0.8
1.0
X3_Holiday_Week

```
> grocery <- read.delim(file = "https://www.math.arizona.edu/~piegorsch/571A/Data/C
hapter06/CH06PR09.txt",
+                       header = F, sep = "")
> head(grocery)
    V1     V2   V3 V4
1 4264 305657 7.17  0
2 4496 328476 6.20  0
3 4317 317164 4.61  0
4 4292 366745 7.02  0
5 4945 265518 8.61  1
6 4325 301995 6.88  0
> X1_Cases <- grocery$V2
> X2_Costs_Of_Labor <- grocery$V3
> X3_Holiday_Week <- grocery$V4
> Y_Total_Labor <- grocery$V1
>
> stem(X1_Cases,scale = 4)

  The decimal point is 4 digit(s) to the right of the |

  21 | 2
  22 | 8
  23 |
  24 | 668
  25 | 027
  26 | 15688999
  27 | 022477
  28 | 38
  29 | 0023467
  30 | 123367
  31 | 7
  32 | 22388
  33 |
  34 |
  35 | 2
  36 | 7
  37 | 0
  38 | 3
```

```
  39 |
  40 |
  41 | 24
  42 | 7
  43 |
  44 | 3
  45 |
  46 |
  47 | 2

> stem(X2_Costs_Of_Labor, scale = 8)

  The decimal point is 1 digit(s) to the left of the |

  46 | 1
  47 |
  48 |
  49 |
  50 |
  51 |
  52 |
  53 |
  54 |
  55 |
  56 |
  57 |
  58 | 2
  59 |
  60 |
  61 | 4
  62 | 07
  63 | 47
  64 | 59
  65 |
  66 |
  67 | 1256
  68 | 28
  69 | 44
```

```
  61 | 4
  62 | 07
  63 | 47
  64 | 59
  65 |
  66 |
  67 | 1256
  68 | 28
  69 | 44
  70 | 25
  71 | 7
  72 | 123356
  73 | 99
  74 |
  75 | 5
  76 | 17
  77 | 2279
  78 | 359
  79 | 29
  80 | 128
  81 | 29
  82 | 0
  83 | 8
  84 |
  85 | 06
  86 | 1
  87 |
  88 |
  89 |
  90 | 1
  91 |
  92 |
  93 |
  94 |
  95 |
  96 | 5

>
```

```
> #b.
> par(mfrow = c(2,2))
> weeks = 1:52
> plot(X1_Cases ~ weeks, col = "blue")
>
> plot(X2_Costs_Of_Labor ~ weeks, col = "red")
>
> plot(X3_Holiday_Week ~ weeks, col= "green")
>
> #c.
> ##scatter plot matrix
> grocery.df <- data.frame(Y_Total_Labor ,X1_Cases , X2_Costs_Of_Labor, X3_Holiday_
Week)
> grocery.df
   Y_Total_Labor X1_Cases X2_Costs_Of_Labor X3_Holiday_Week
1           4264   305657              7.17               0
2           4496   328476              6.20               0
3           4317   317164              4.61               0
4           4292   366745              7.02               0
5           4945   265518              8.61               1
6           4325   301995              6.88               0
7           4110   269334              7.23               0
8           4111   267631              6.27               0
9           4161   296350              6.49               0
10          4560   277223              6.37               0
11          4401   269189              7.05               0
12          4251   277133              6.34               0
13          4222   282892              6.94               0
14          4063   306639              8.56               0
15          4343   328405              6.71               0
16          4833   321773              5.82               1
17          4453   272319              6.82               0
18          4195   293880              8.38               0
19          4394   300867              7.72               0
20          4099   296872              7.67               0
21          4816   245674              7.72               1
22          4867   211944              6.45               1
23          4114   227996              7.22               0
```

```
23          4114   227996                   7.22                    0
24          4314   248328                   8.50                    0
25          4289   249894                   8.08                    0
26          4269   302660                   7.26                    0
27          4347   273848                   7.39                    0
28          4178   245743                   8.12                    0
29          4333   267673                   6.75                    0
30          4226   256506                   7.79                    0
31          4121   271854                   7.89                    0
32          3998   293225                   9.01                    0
33          4475   269121                   8.01                    0
34          4545   322812                   7.21                    0
35          4016   252225                   7.85                    0
36          4207   261365                   6.14                    0
37          4148   287645                   6.76                    0
38          4562   289666                   7.92                    0
39          4146   270051                   8.19                    0
40          4555   265239                   7.55                    0
41          4365   352466                   6.94                    0
42          4471   426908                   7.25                    0
43          5045   369989                   9.65                    1
44          4469   472476                   8.20                    0
45          4408   414102                   8.02                    0
46          4219   302507                   6.72                    0
47          4211   382686                   7.23                    0
48          4993   442782                   7.61                    1
49          4309   322303                   7.39                    0
50          4499   290455                   7.99                    0
51          4186   411750                   7.83                    0
52          4342   292087                   7.77                    0
> pairs(grocery.df)
> cor(grocery.df)
                  Y_Total_Labor   X1_Cases X2_Costs_Of_Labor X3_Holiday_Week
Y_Total_Labor         1.0000000 0.20766494        0.06002960      0.81057940
X1_Cases              0.2076649 1.00000000        0.08489639      0.04565698
X2_Costs_Of_Labor     0.0600296 0.08489639        1.00000000      0.11337076
X3_Holiday_Week       0.8105794 0.04565698        0.11337076      1.00000000
> ########################################################################
```

```
> ##6.10
> #a.
>
> lm.grocery <- lm(Y_Total_Labor ~ X1_Cases + X2_Costs_Of_Labor + X3_Holiday_Week)
> summary(lm.grocery)

Call:
lm(formula = Y_Total_Labor ~ X1_Cases + X2_Costs_Of_Labor + X3_Holiday_Week)

Residuals:
    Min      1Q  Median      3Q     Max
-264.05 -110.73  -22.52   79.29  295.75

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        4.150e+03  1.956e+02  21.220  < 2e-16 ***
X1_Cases           7.871e-04  3.646e-04   2.159   0.0359 *
X2_Costs_Of_Labor -1.317e+01  2.309e+01  -0.570   0.5712
X3_Holiday_Week    6.236e+02  6.264e+01   9.954 2.94e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 143.3 on 48 degrees of freedom
Multiple R-squared:  0.6883,    Adjusted R-squared:  0.6689
F-statistic: 35.34 on 3 and 48 DF,  p-value: 3.316e-12

> coef(lm.grocery)
      (Intercept)          X1_Cases X2_Costs_Of_Labor   X3_Holiday_Week
     4.149887e+03      7.870804e-04     -1.316602e+01      6.235545e+02
> #b.
> residuals <- resid(lm.grocery)
> sum(residuals)
[1] -3.113065e-13
> boxplot(resid(lm.grocery))
>
> #c.
> par(mfrow= c(2,2))
```

```
> plot(resid(lm.grocery) ~ fitted(lm.grocery))
>
> plot(resid(lm.grocery) ~ X1_Cases)
>
> plot(resid(lm.grocery) ~ X2_Costs_Of_Labor)
>
> plot(resid(lm.grocery) ~ X3_Holiday_Week)
>
> X_1.X_2 <- X1_Cases*X2_Costs_Of_Labor
>
> plot(resid(lm.grocery) ~ X_1.X_2)
>
>
> #NPP
>
> qqnorm(resid(lm.grocery), main = "")
>
> #e.
>
> library("lawstat")
> ei <- resid(lm.grocery)
> require(lawstat)
> BF.htest <- levene.test(ei[order(Y_Total_Labor)], group = c(rep(1,26),rep(2,26)),
 location = "median")
> BF.htest

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

data:  ei[order(Y_Total_Labor)]
Test Statistic = 0.94049, p-value = 0.3368


>
> ##################################################################
> ##6.11
> #a.
> anova(lm(Y_Total_Labor ~ 1), lm.grocery)
Analysis of Variance Table
```

```
Model 1: Y_Total_Labor ~ 1
Model 2: Y_Total_Labor ~ X1_Cases + X2_Costs_Of_Labor + X3_Holiday_Week
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     51 3162136
2     48  985530  3   2176606 35.337 3.316e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> qf( .95, 3,48)
[1] 2.798061
>
> #b.
> g = length(coef(lm.grocery))-2
> g
[1] 2
>
> alpha = .05
> confint(lm.grocery, level = 1 - (alpha/g))
                         1.25 %      98.75 %
(Intercept)         3.697369e+03 4.602406e+03
X1_Cases           -5.646080e-05 1.630622e-03
X2_Costs_Of_Labor -6.659796e+01 4.026592e+01
X3_Holiday_Week     4.786096e+02 7.684993e+02
>
> B <- qt(1-(.05/4), 48)
>
> #c.
> summary(lm.grocery)

Call:
lm(formula = Y_Total_Labor ~ X1_Cases + X2_Costs_Of_Labor + X3_Holiday_Week)

Residuals:
    Min      1Q  Median      3Q     Max
-264.05 -110.73  -22.52   79.29  295.75

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        4.150e+03  1.956e+02  21.220  < 2e-16 ***
```

```
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         4.150e+03  1.956e+02  21.220  < 2e-16 ***
X1_Cases            7.871e-04  3.646e-04   2.159   0.0359 *
X2_Costs_Of_Labor -1.317e+01  2.309e+01  -0.570   0.5712
X3_Holiday_Week     6.236e+02  6.264e+01   9.954 2.94e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 143.3 on 48 degrees of freedom
Multiple R-squared:  0.6883,    Adjusted R-squared:  0.6689
F-statistic: 35.34 on 3 and 48 DF,  p-value: 3.316e-12


>
>
> ##########################################################################
> ##6.12
> #a.
> g1 = 5
> alpha1= .05
> Spoint = sqrt(g1*qf(1-alpha1,g1, lm.grocery$df.residual))
> Spoint
[1] 3.470241
>
> Bpoint = qt(1 - (.5*(alpha1/g1)), lm.grocery$df.residual)
> Bpoint
[1] 2.682204
>
> newdata.df = data.frame(X1_Cases = c(302000,245000,280000,350000,295000),
+                         X2_Costs_Of_Labor= c(7.2,7.4,6.9,7,6.7),
+                         X3_Holiday_Week = c(0,0,0,0,1))
>
> predict.lm(lm.grocery, newdata=newdata.df,se.fit=T, interval='confidence',level =
 1- (alpha1/5))
$fit
       fit      lwr      upr
1 4292.790 4235.507 4350.073
2 4245.293 4165.626 4324.960
3 4279.424 4213.859 4344.989
```

```
4 4333.203 4255.609 4410.798
5 4917.418 4749.781 5085.055

$se.fit
       1        2        3        4        5 
21.35667 29.70206 24.44444 28.92934 62.49980 

$df
[1] 48

$residual.scale
[1] 143.2895

> 
> 
> #b.
> 
> grocery.df
   Y_Total_Labor X1_Cases X2_Costs_Of_Labor X3_Holiday_Week
1           4264   305657              7.17               0
2           4496   328476              6.20               0
3           4317   317164              4.61               0
4           4292   366745              7.02               0
5           4945   265518              8.61               1
6           4325   301995              6.88               0
7           4110   269334              7.23               0
8           4111   267631              6.27               0
9           4161   296350              6.49               0
10          4560   277223              6.37               0
11          4401   269189              7.05               0
12          4251   277133              6.34               0
13          4222   282892              6.94               0
14          4063   306639              8.56               0
15          4343   328405              6.71               0
16          4833   321773              5.82               1
17          4453   272319              6.82               0
18          4195   293880              8.38               0
19          4394   300867              7.72               0
```

```
19          4394   300867                7.72                 0
20          4099   296872                7.67                 0
21          4816   245674                7.72                 1
22          4867   211944                6.45                 1
23          4114   227996                7.22                 0
24          4314   248328                8.50                 0
25          4289   249894                8.08                 0
26          4269   302660                7.26                 0
27          4347   273848                7.39                 0
28          4178   245743                8.12                 0
29          4333   267673                6.75                 0
30          4226   256506                7.79                 0
31          4121   271854                7.89                 0
32          3998   293225                9.01                 0
33          4475   269121                8.01                 0
34          4545   322812                7.21                 0
35          4016   252225                7.85                 0
36          4207   261365                6.14                 0
37          4148   287645                6.76                 0
38          4562   289666                7.92                 0
39          4146   270051                8.19                 0
40          4555   265239                7.55                 0
41          4365   352466                6.94                 0
42          4471   426908                7.25                 0
43          5045   369989                9.65                 1
44          4469   472476                8.20                 0
45          4408   414102                8.02                 0
46          4219   302507                6.72                 0
47          4211   382686                7.23                 0
48          4993   442782                7.61                 1
49          4309   322303                7.39                 0
50          4499   290455                7.99                 0
51          4186   411750                7.83                 0
52          4342   292087                7.77                 0
>
> ##################################################################
> ##6.13
> Spoint1 = sqrt(g2*qf(1-alpha2,g2, lm.grocery$df.residual))
```

```
32          4342   292087                7.77                0
>
> ####################################################################
> ##6.13
> Spoint1 = sqrt(g2*qf(1-alpha2,g2, lm.grocery$df.residual))
> Spoint1
[1] 3.203274
>
> Bpoint1 = qt(1 - (.5*(alpha2/g2)), lm.grocery$df.residual)
> Bpoint1
[1] 2.595323
>
>
> newdata.df2 = data.frame(X1_Cases = c(230000,250000,280000,340000),
+                          X2_Costs_Of_Labor = c(7.5,7.3,7.1,6.9),
+                          X3_Holiday_Week = c(rep(0,4)))
> g2=4
> alpha2 = .05
> predict.lm(lm.grocery, newdata=newdata.df2, interval='prediction',
+            level = 1 - (alpha2/g2))
       fit      lwr      upr
1 4232.171 3849.911 4614.430
2 4250.545 3871.478 4629.613
3 4276.791 3900.122 4653.460
4 4326.649 3947.913 4705.385
> ############################################################################
> ##6.14
> #a.
> alpha3 = .05
> g3 = 3
> newdata.df3 = data.frame(X1_Cases = c(282000),
+                           X2_Costs_Of_Labor = c(7.1),
+                           X3_Holiday_Week = c(0))
>
> predict.lm(lm.grocery, newdata=newdata.df3, interval='prediction')
       fit      lwr      upr
1 4278.365 3986.626 4570.104
>
```