Guided Question Set 10 Solutions

1)

```r
Data<-read.table("nfl.txt", header=TRUE)
```

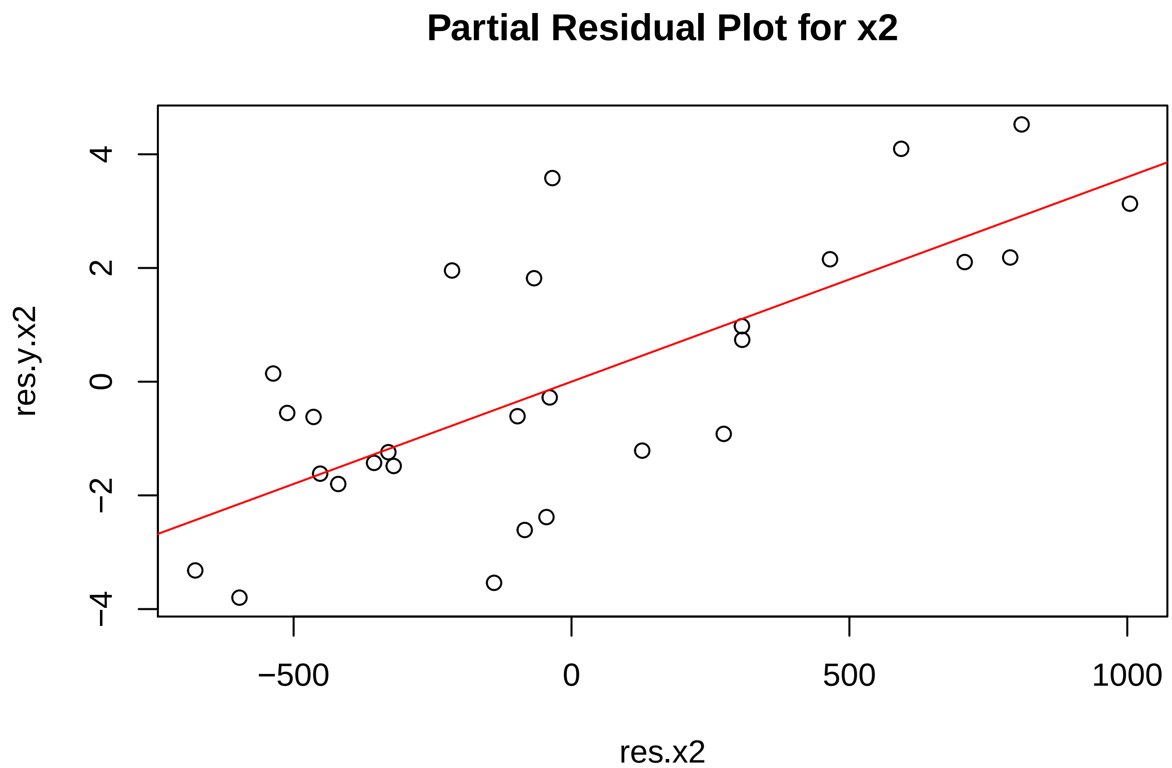
a)

The partial residual plot for x_2 is shown below

```r
result.y.x2<-lm(y~x7+x8, data=Data)
result.x2<-lm(x2~x7+x8, data=Data)

res.y.x2<-result.y.x2$residuals
res.x2<-result.x2$residuals

##partial residual plot for x2
plot(res.x2,res.y.x2,
     main="Partial Residual Plot for x2")
##overlay regression line
abline(lm(res.y.x2~res.x2), col="red")
```



The plots are evenly scattered across the regression line. The partial residual plot for x_2 informs us that a linear term for x_2 will be appropriate when x_7 and x_8 are already in the model, and that the estimated coefficient for x_2 would be positive in the MLR model with x_2, x_7, x_8 as predictors.

b)

```
summary(lm(res.y.x2~res.x2))
```

```
##
## Call:
## lm(formula = res.y.x2 ~ res.x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0370 -0.7129 -0.2043  1.1101  3.7049
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -6.714e-16  3.098e-01   0.000        1
## res.x2       3.598e-03  6.677e-04   5.388 1.21e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.639 on 26 degrees of freedom
## Multiple R-squared:  0.5276, Adjusted R-squared:  0.5094
## F-statistic: 29.04 on 1 and 26 DF,  p-value: 1.209e-05
```

The estimated slope is 0.003598 and the estimated intercept is 0.

c)

```r
result<-lm(y~x2+x7+x8, data=Data)
summary(result)
```

```
##
## Call:
## lm(formula = y ~ x2 + x7 + x8, data = Data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0370 -0.7129 -0.2043  1.1101  3.7049
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.808372   7.900859  -0.229 0.820899
## x2           0.003598   0.000695   5.177 2.66e-05 ***
## x7           0.193960   0.088233   2.198 0.037815 *
## x8          -0.004816   0.001277  -3.771 0.000938 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.706 on 24 degrees of freedom
## Multiple R-squared:  0.7863, Adjusted R-squared:  0.7596
## F-statistic: 29.44 on 3 and 24 DF,  p-value: 3.273e-08
```

The estimated slope for x_2 is 0.003598, which is the same as the estimated slope for the partial residual plot from the previous part.

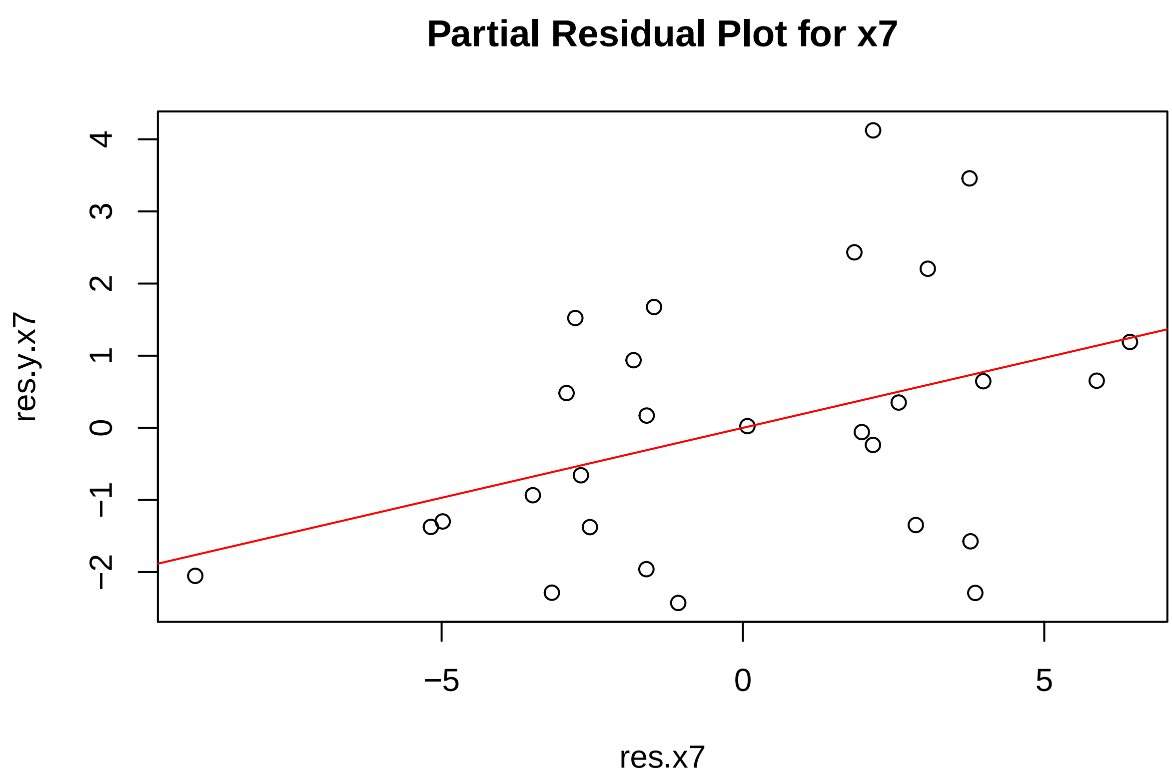
d)

The estimated slopes would be 0.1940 and -0.004816 respectively.

e)

The partial residual plot for x_7 is shown below

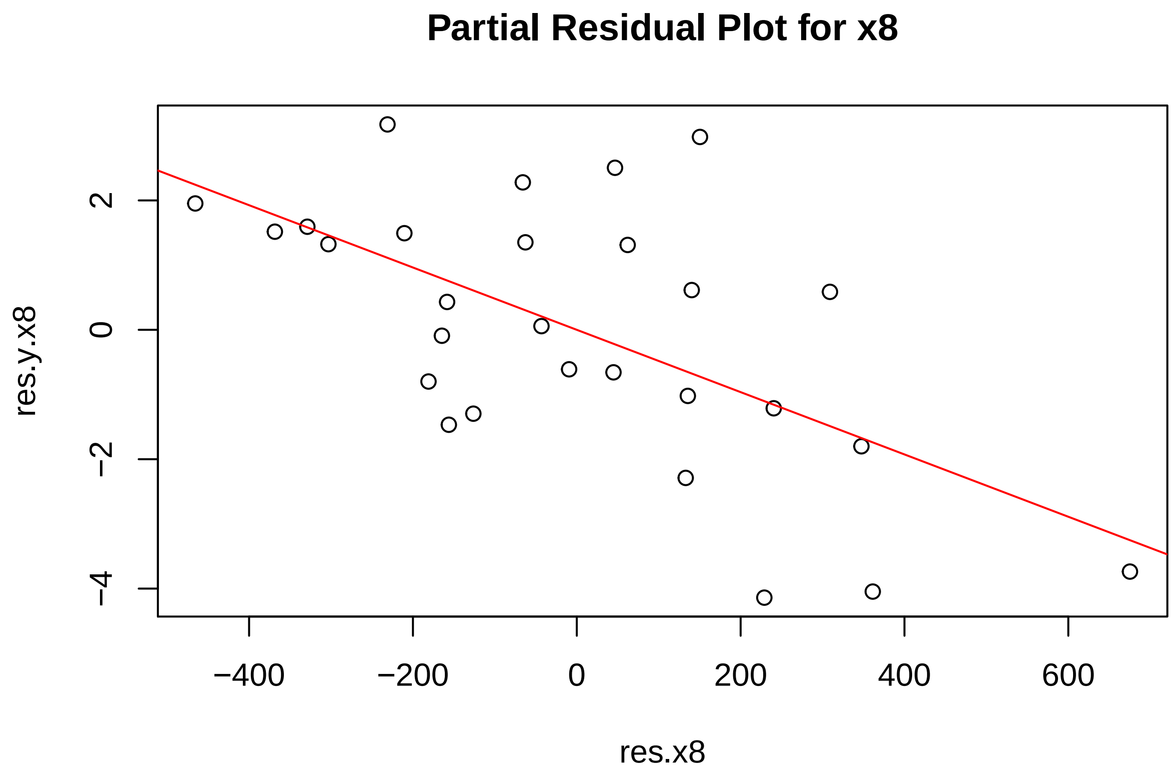
```r
result.y.x7<-lm(y~x2+x8, data=Data)
result.x7<-lm(x7~x2+x8, data=Data)
res.y.x7<-result.y.x7$residuals
res.x7<-result.x7$residuals
plot(res.x7,res.y.x7, main="Partial Residual Plot for x7")
abline(lm(res.y.x7~res.x7), col="red")
```



Since the plots are evenly scattered across the regression line, x_7 should be added as a linear term.

The partial residual plot for x_8 is shown below

```r
result.y.x8<-lm(y~x2+x7, data=Data)
result.x8<-lm(x8~x2+x7, data=Data)
res.y.x8<-result.y.x8$residuals
res.x8<-result.x8$residuals
plot(res.x8,res.y.x8, main="Partial Residual Plot for x8")
abline(lm(res.y.x8~res.x8), col="red")
```



Since the plots are evenly scattered across the regression line, x_8 should be added as a linear term.

2)

```
##critical value using Bonferroni procedure
n<-dim(Data)[1]
p<-4
crit<-qt(1-0.05/(2*n), n-1-p)
##externally studentized residuals
ext.student.res<-rstudent(result)
##identify
ext.student.res[abs(ext.student.res)>crit]
```

```
## named numeric(0)
```

No outliers based on externally studentized residuals.

3)

```r
##leverages
lev<-lm.influence(result)$hat
##identify
lev[lev>2*p/n]
```

```
##        18        27
## 0.3928394 0.3192801
```

Two teams that have high leverage, teams 18 and 27.

4)

```r
DFFITS<-dffits(result)
DFFITS[abs(DFFITS)>2*sqrt(p/n)]
```

```
## named numeric(0)
```

There are no teams that are influential in terms of $DFFITS_i$.

```r
DFBETAS<-dfbetas(result)
abs(DFBETAS)>2/sqrt(n)
```

```
##    (Intercept)    x2    x7    x8
## 1        FALSE FALSE FALSE FALSE
## 2        FALSE FALSE FALSE FALSE
## 3        FALSE FALSE FALSE FALSE
## 4        FALSE FALSE FALSE FALSE
## 5        FALSE FALSE FALSE FALSE
## 6        FALSE FALSE FALSE FALSE
## 7        FALSE FALSE FALSE FALSE
## 8        FALSE FALSE FALSE FALSE
## 9        FALSE FALSE FALSE FALSE
## 10       FALSE FALSE FALSE  TRUE
## 11       FALSE FALSE FALSE FALSE
## 12       FALSE FALSE FALSE FALSE
## 13       FALSE FALSE FALSE FALSE
## 14       FALSE FALSE FALSE FALSE
```

```
## 15       FALSE FALSE FALSE FALSE
## 16       FALSE FALSE FALSE FALSE
## 17       FALSE FALSE FALSE FALSE
## 18       FALSE FALSE FALSE FALSE
## 19       FALSE FALSE FALSE FALSE
## 20       FALSE FALSE FALSE FALSE
## 21       FALSE FALSE  TRUE FALSE
## 22       FALSE FALSE FALSE FALSE
## 23       FALSE FALSE FALSE FALSE
## 24       FALSE FALSE FALSE FALSE
## 25       FALSE FALSE FALSE FALSE
## 26       FALSE FALSE FALSE FALSE
## 27       FALSE FALSE FALSE FALSE
## 28       FALSE FALSE FALSE FALSE
```

Team 21 is influential in terms of β_3 and team 10 is influential in terms of β_4.

```
COOKS<-cooks.distance(result)
COOKS[COOKS>qf(0.5,p,n-p)]
```

```
## named numeric(0)
```

There are no teams that are influential in terms of Cook's distance.