M10Guided

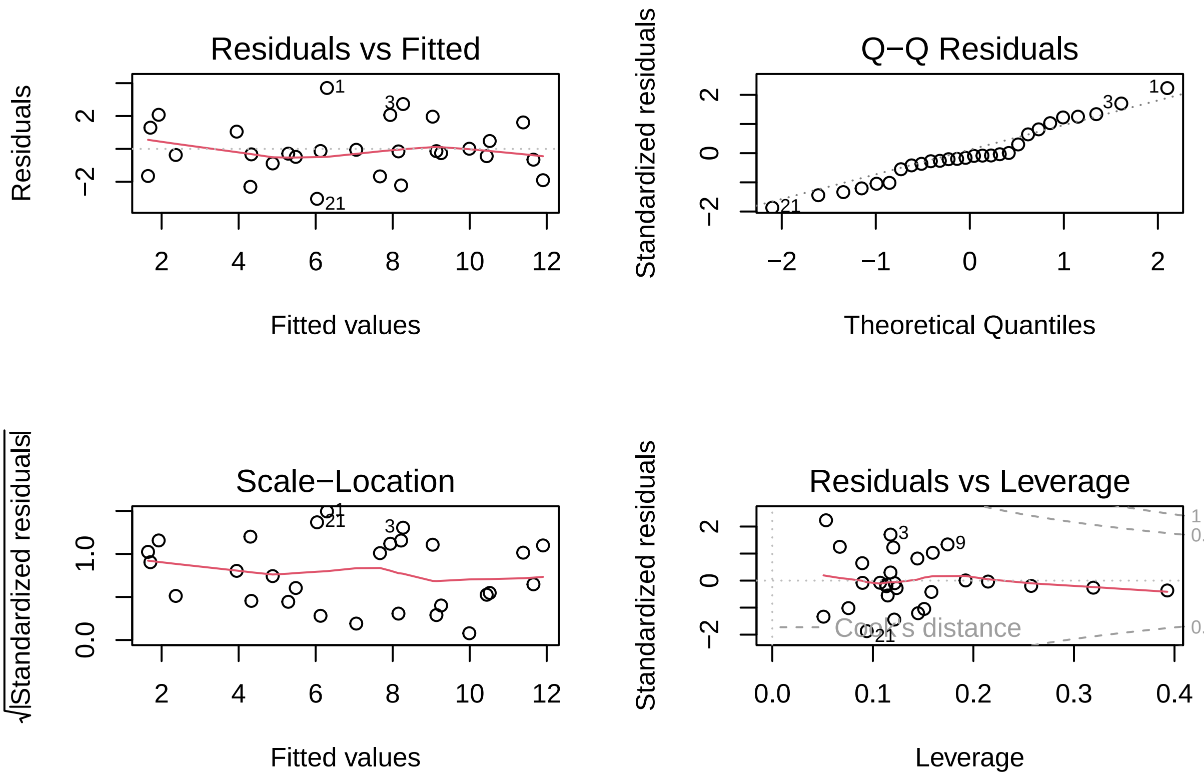
Alanna Hazlett

2024-04-07

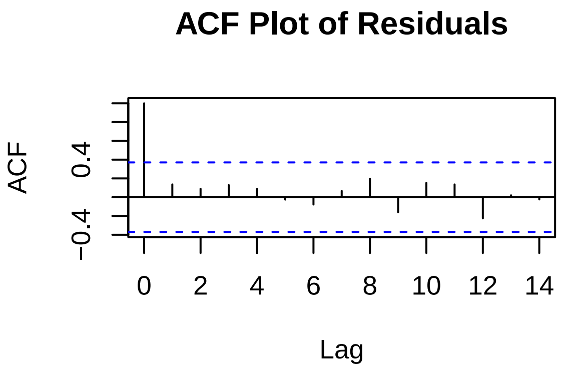
Problem 1

Create diagnostic plots for this regression. What are the plots telling us?

```r
result<-lm(y~x2+x7+x8,data=Data)
par(mfrow=c(2,2))
plot(result)
```




```r
acf(result$residuals, main="ACF Plot of Residuals")
```

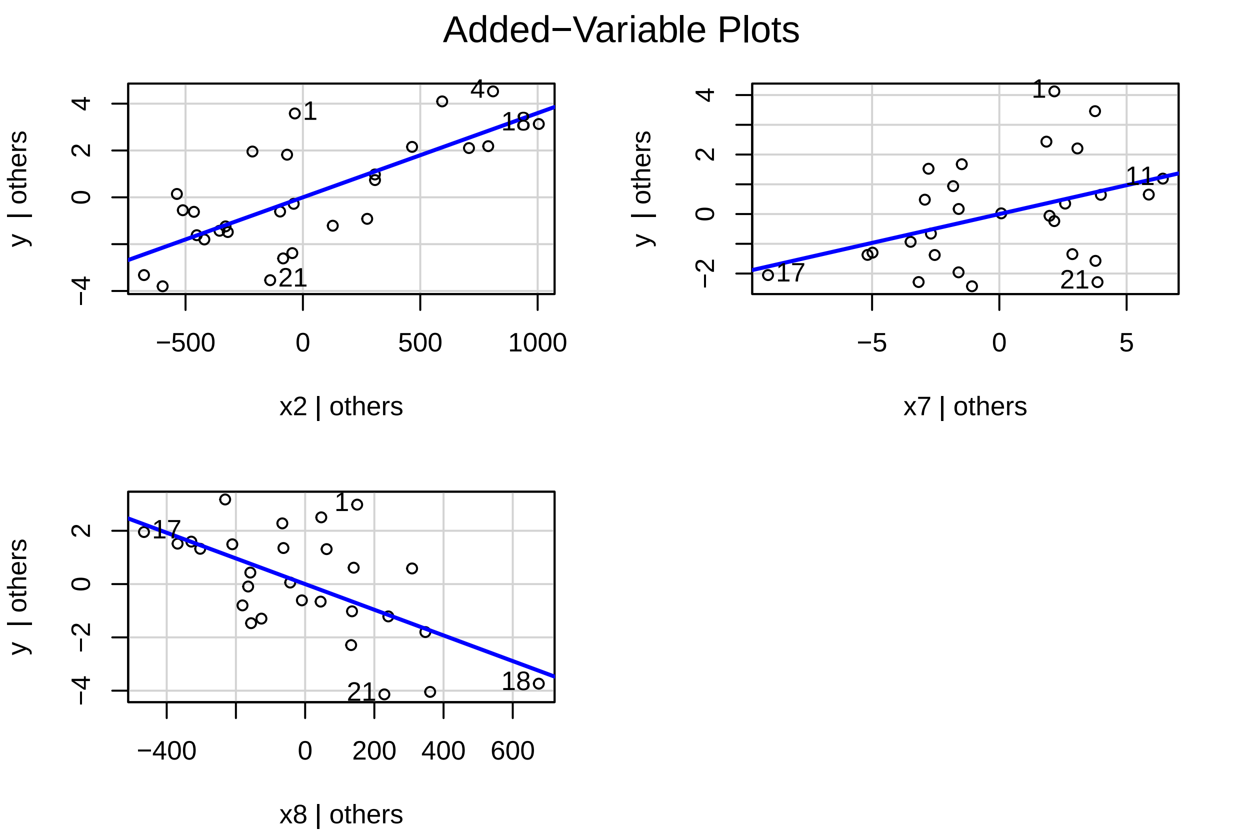


Assumption 1: Do the errors have mean of 0 for each value of the predictor -Yes
Assumption 2: Do the errors have constant variance for each value of the predictor -Yes
Assumption 3: Are the errors independent (acf plot) - Yes
Assumption 4: Are the errors normally distributed? -Yes
There is a linear relationship between the predictors x2 (Passing yards-Season), x7 (Percent rushing), x8 (Opponent's rushing yards) with our response variable (Games won).

Problem 2

Generate partial regression plots for each of the predictors. Interpret what these plots are telling us.

```
car::avPlots(result)
```



UL: Linear pattern present, positive linear association. Slope of 250 = estimated coefficient of x2 in the model.
UR: Linear pattern present, positive linear association. Slope of 0.2 = estimated coefficient of x7 in the

model.
LL: Linear pattern present, negative linear association. Slope of -0.005 = estimated coefficient of x8 in the model.

Problem 3

Using externally studentized residuals, do we have any outliers? What teams are these?

```r
sort(abs(rstandard(result)))
```

```
##           5          11          19          13          26          14
## 0.006124483 0.036468456 0.078998342 0.083851688 0.094055761 0.160668820
##          17          20          27          24           8          18
## 0.196937383 0.206464327 0.262130195 0.276544687 0.299328499 0.365011749
##           6          23          16          22          25           4
## 0.418876221 0.551056514 0.644990078 0.817274105 1.018586104 1.029767789
##          28           7           2          12          15           9
## 1.048746774 1.206836995 1.225616368 1.251090093 1.335367350 1.338032316
##          10           3          21           1
## 1.441760607 1.702625305 1.869940122 2.231851618
```

Guideline externally studentized residuals greater than 2 are flagged as outliers
Team 1 is a potential outlier.

Problem 4

Do we have any high leverage data points for this multiple linear regression? What teams are these?

```r
sort(lm.influence(result)$hat)
```

```
##         15          1         12         25         16         13         21
## 0.05091949 0.05342996 0.06711700 0.07572840 0.08946483 0.08972171 0.09396060
##         19         14         20         23          8          3          2
## 0.10721396 0.11315466 0.11352891 0.11475999 0.11752620 0.11758271 0.12033048
##         10         26         24         22          7         28          6
## 0.12129504 0.12173698 0.12364134 0.14431020 0.14497494 0.15105325 0.15824616
##          4          9          5         11         17         27         18
## 0.15962224 0.17431803 0.19222108 0.21455958 0.25746278 0.31928012 0.39283935
```

```r
hii<-lm.influence(result)$hat
n<-nrow(Data)
p<-4
hii[hii>2*p/n]
```

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

*Guideline for significant leverage is hii > 2p/n = 2(4)/28 =
My results show that there are no high leverage data points in this data set.

Problem 5

Use DFFITSi, DFBETASj,i, and Cook's distance to check for influential observations. What teams are influential?
(DFFITS)

```
p<-4
n<-nrow(Data)
DFFITS<-dffits(result)
DFFITS[abs(DFFITS)>2*sqrt(p/n)]
```

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

According to DFFITS we have no influential observations.
(DFBETAS)

```
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
```

```
DFBETAS[10,]
```

```
## (Intercept)          x2          x7          x8
##  0.32281766  0.05400214 -0.30760184 -0.42602002
```

```r
DFBETAS[21,]
```

```
## (Intercept)          x2          x7          x8 
##   0.3528626   0.1181012  -0.4148213  -0.3565277
```

Guideline of significance = 2/sqrt(n) = 0.3780
Team 10 coefficient for x8 = -0.4260
Team 21 coefficient for x7 = -0.4148
(Cook's Distance)

```r
COOKS<-cooks.distance(result)
sort(COOKS)
```

```
##            5           11           13           19           26           14 
## 2.231451e-06 9.082584e-05 1.732555e-04 1.873613e-04 3.065552e-04 8.234321e-04 
##           20           24            8           17           27            6 
## 1.364809e-03 2.697444e-03 2.983108e-03 3.361960e-03 8.057078e-03 8.246307e-03 
##           23           16           25           18           15           12 
## 9.841510e-03 1.021884e-02 2.125175e-02 2.155088e-02 2.391787e-02 2.815286e-02 
##           22           28            4            2            7            1 
## 2.816153e-02 4.892501e-02 5.035440e-02 5.136949e-02 6.173783e-02 7.029149e-02 
##           10           21            9            3 
## 7.173421e-02 9.065549e-02 9.449367e-02 9.657121e-02
```

```r
#COOKS[COOKS>1]
```

According to cook's distance we have no influential observations (teams).