Lab 9

Kristen Fitzgerald

2024-07-09

Question 1

Using the cars2010 data set, run the regression with the following explanatory variables:

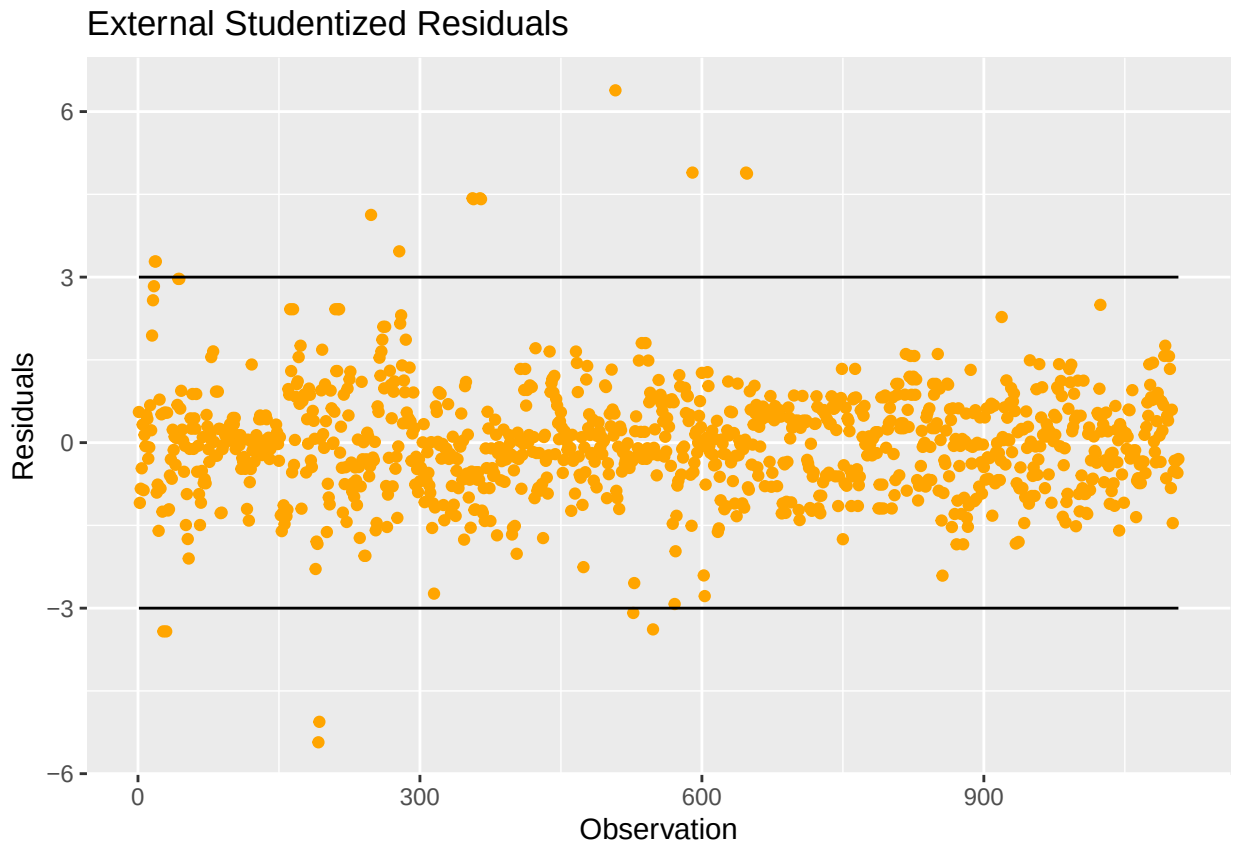
- `EngDispl`
- `Transmission`
- `AirAspirationMethod`
- `TransLockup`
- `TransCreeperGear`
- `DriveDesc`
- `IntakeValvePerCyl`
- `CarlineClassDesc`
- `VarValveLift`

```
cars.lm <- lm(FE ~ EngDispl + Transmission + AirAspirationMethod +
              TransLockup + TransCreeperGear + DriveDesc +
              IntakeValvePerCyl + CarlineClassDesc + VarValveLift,
              data = cars2010)

n.index = seq(1, nrow(cars2010))
cars_data <- cbind(cars2010, n.index)
```

(a) Use plots to identify potential influential observations based on the suggested cutoff values.

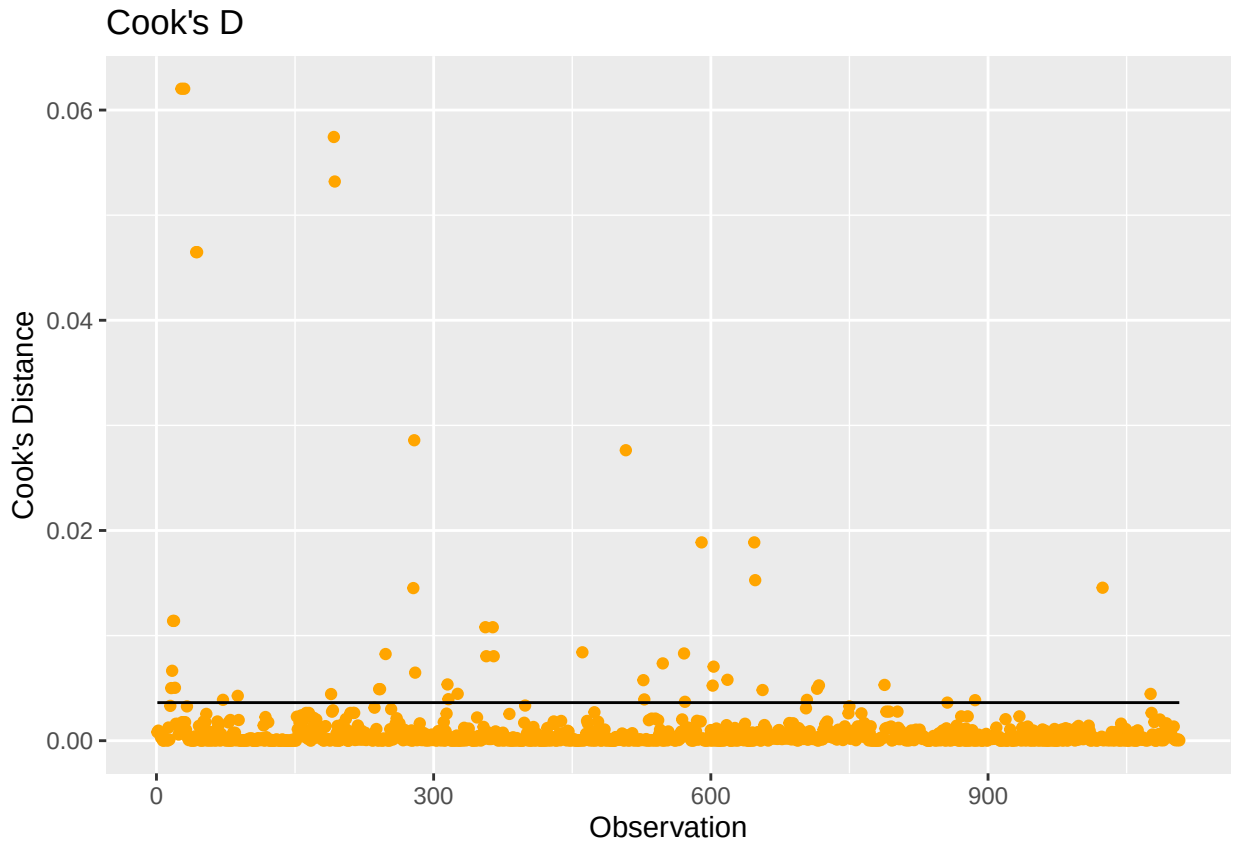
Studentized Residuals

```
ggplot(cars.lm,  aes( x = n.index, y = rstudent(cars.lm))) +
  geom_point(color="orange") +
  geom_line(y = -3) +
  geom_line(y = 3) +
  labs(title = "External Studentized Residuals", x = "Observation",
       y = "Residuals")
```


External Studentized Residuals
Residuals
6
3
0
−3
−6
0
300
600
900
Observation

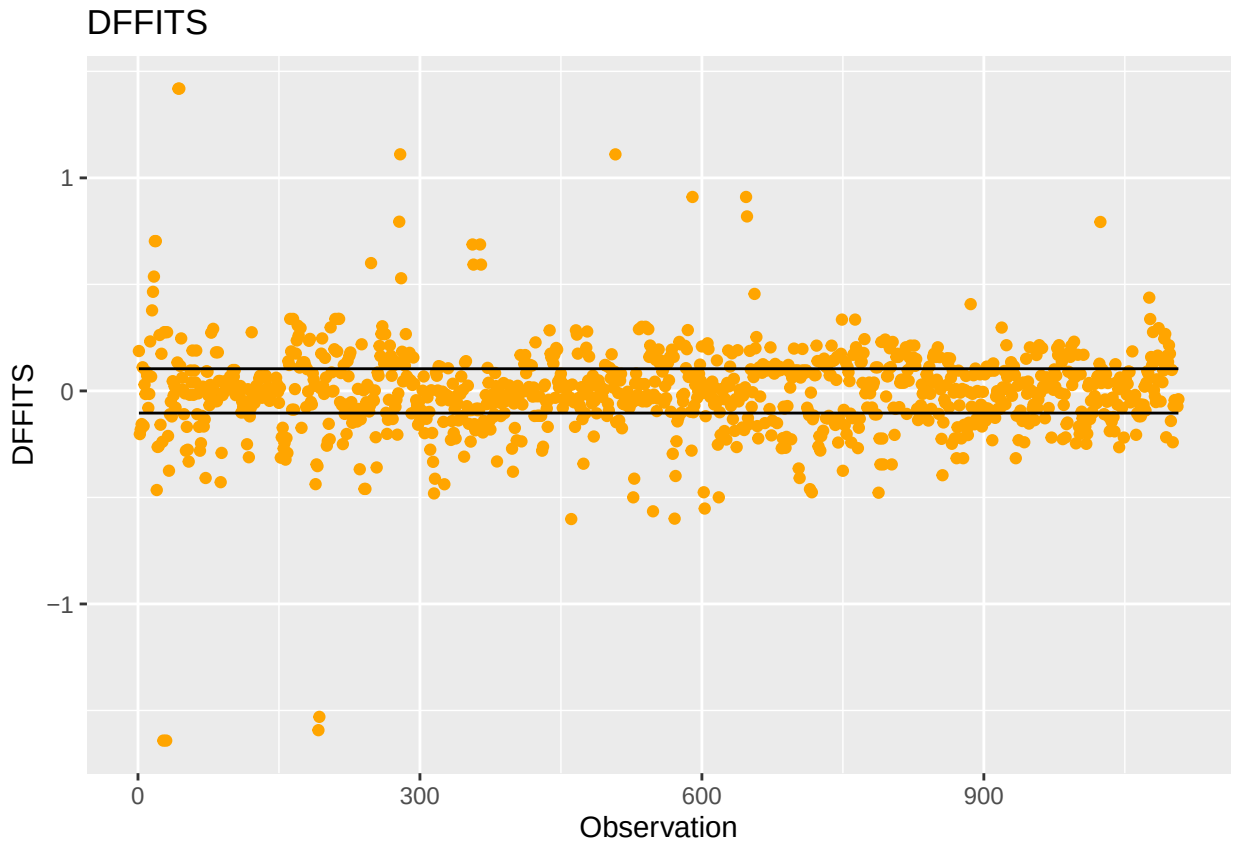
Cook's Distance

```
D.cut = 4 / (nrow(cars_data) - 3)

ggplot(cars.lm, aes( x = n.index, y = cooks.distance(cars.lm))) +
  geom_point(color = "orange") +
  geom_line(y = D.cut) +
  labs(title = "Cook's D", x = "Observation", y = "Cook's Distance")
```



DFFITS

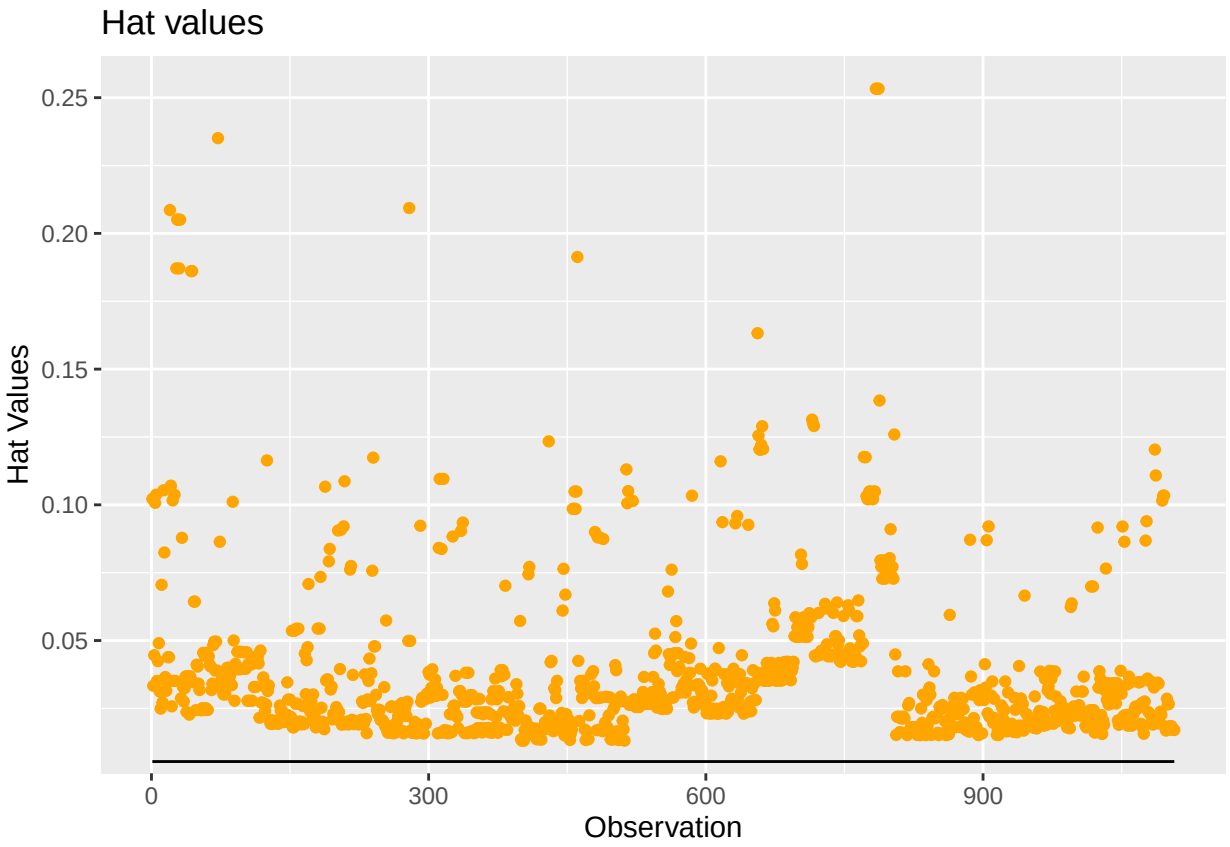
```
df.cut = 2*(sqrt( 3 / nrow(cars_data)))

ggplot(cars.lm, aes(x = n.index, y = dffits(cars.lm))) +
  geom_point(color="orange") +
  geom_line(y = df.cut) +
  geom_line(y = -df.cut) +
  labs(title = "DFFITS", x="Observation", y="DFFITS")
```




```
db.cut = 2 / sqrt(nrow(cars_data))
```

Hat Values

```
hat.cut = 2*(3) / nrow(cars_data)

ggplot(cars.lm, aes(x = n.index, y = hatvalues(cars.lm))) +
  geom_point(color = "orange") +
  geom_line(y = hat.cut) +
  labs(title = "Hat values", x="Observation", y="Hat Values")
```



b. Are there any observations with a dffits larger than 1 AND studentized residuals larger than 3 in magnitude? If so, list the observations.

Solution: Yes, observation 1596 has a dffits larger than 1 AND studentized residuals larger than 3 in magnitude.

```
Anomalous <- (rstudent(cars.lm) > 3) & (dffits(cars.lm) > 1)
Anomalous[Anomalous == TRUE]
```

```
## 1596
## TRUE
```