Exercise 6

Arthur Junges Schmidt

08 July 2020

The data which can be found in a separate spreadsheet provides the highway gasoline mileage test results for 2005 model year vehicles from DaimlerChrysler. (1) Fit a multiple linear regression model to these data to estimate gasoline mileage that uses the following regressors: cid, rhp, etw, cmp, axle, n/v

```
Linear_Model <- lm(formula = mpg ~ cid + rhp + etw + cmp + axle +
                     `n/v`,
                   data = Original_Data);
summary(Linear_Model)
```

```
##
## Call:
## lm(formula = mpg ~ cid + rhp + etw + cmp + axle + `n/v`, data = Original_Data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0501 -0.8477  0.2360  1.0896  2.8193
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 49.9039998 19.6652426   2.538  0.02368 *
## cid         -0.0104474  0.0233788  -0.447  0.66180
## rhp         -0.0012042  0.0163061  -0.074  0.94217
## etw         -0.0032364  0.0009459  -3.421  0.00413 **
## cmp          0.2924277  1.7647364   0.166  0.87076
## axle        -3.8553646  1.3286464  -2.902  0.01160 *
## `n/v`        0.1897094  0.2729740   0.695  0.49845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.228 on 14 degrees of freedom
## Multiple R-squared:  0.8933, Adjusted R-squared:  0.8476
## F-statistic: 19.53 on 6 and 14 DF,  p-value: 4.664e-06
```

(2) Estimate σ^2 and the standard errors of the regression coefficients.

σ^2 values:

```
##  (Intercept)          cid          rhp          etw          cmp         axle
## 3.867218e+02 5.465676e-04 2.658901e-04 8.947518e-07 3.114294e+00 1.765301e+00
##         `n/v`
## 7.451482e-02
```

Standard errors:

```
##  (Intercept)          cid          rhp          etw          cmp         axle
## 1.966524e+01 2.337879e-02 1.630614e-02 9.459132e-04 1.764736e+00 1.328646e+00
##        `n/v`
## 2.729740e-01
```

(3) Test for significance of regression using $\alpha = 0.05$. What conclusions can you draw? Only the intercept and the 'etw' and 'axle' variables have a significance of 95% or higher. This leads to the conclusion that, with 95% certainty, only the axle ration and the equivalent test weight variables are significant to explain the mileage of the vehicles on the data set.

(4) Find the t-test statistic for each regressor. Using $\alpha = 0.05$, what conclusions can you draw? Does each regressor contribute to the model?

T-test statistic for each regressor:

```
## (Intercept)         cid         rhp         etw         cmp        axle
##  2.53767527 -0.44687602 -0.07385193 -3.42145672  0.16570617 -2.90172361
##       `n/v`
##  0.69497223
```

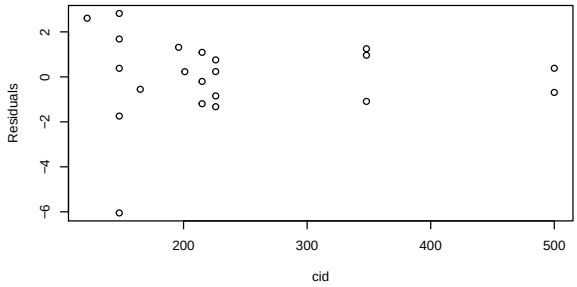
As on the question before, the only 2 variables that have a significant t-test are 'etw' and 'axle'. Not all the regressor contributes to the model since some of them add more error into it.

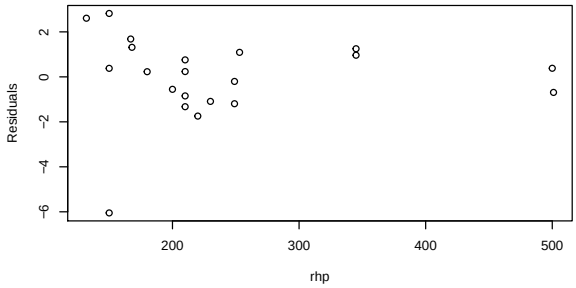
(5) Find 99% confidence intervals on the regression coefficients.

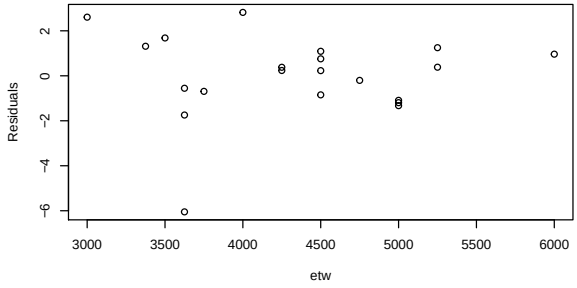
```
##                    0.5 %        99.5 %
## (Intercept) -8.636334830  1.084443e+02
## cid         -0.080042386  5.914755e-02
## rhp         -0.049745045  4.733657e-02
## etw         -0.006052236 -4.205662e-04
## cmp         -4.960914978  5.545770e+00
## axle        -7.810535975  9.980675e-02
## `n/v`       -0.622891389  1.002310e+00
```

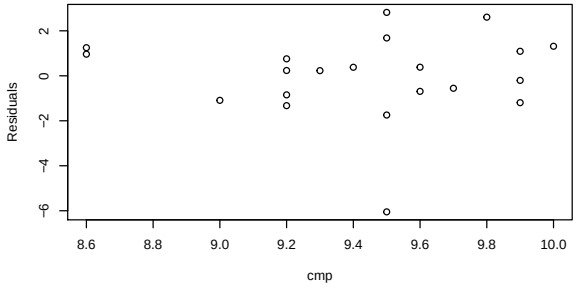
(6) Plot residuals versus $\hat{Y}$ and versus each regressor. Discuss these residual plots.

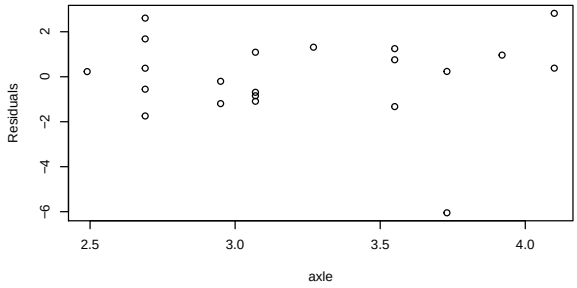
```
par(mfrow = c(5,2))
plot(x = Linear_Model$model$cid, y = resid(Linear_Model), xlab = "cid", ylab = "Residuals")
plot(x = Linear_Model$model$rhp, y = resid(Linear_Model), xlab = "rhp", ylab = "Residuals")
plot(x = Linear_Model$model$etw, y = resid(Linear_Model), xlab = "etw", ylab = "Residuals")
plot(x = Linear_Model$model$cmp, y = resid(Linear_Model), xlab = "cmp", ylab = "Residuals")
plot(x = Linear_Model$model$axle, y = resid(Linear_Model), xlab = "axle", ylab = "Residuals")
plot(x = Linear_Model$model$`n/v`, y = resid(Linear_Model), xlab = "n/v", ylab = "Residuals")
plot(Linear_Model)
```

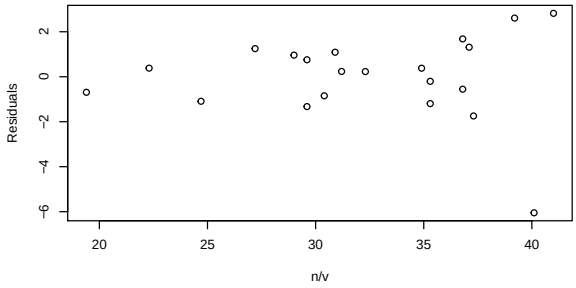

Residuals
cid

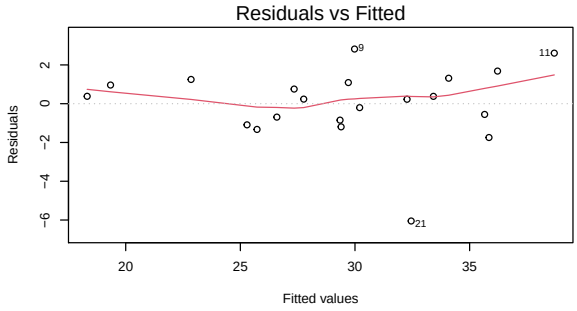

Residuals
rhp

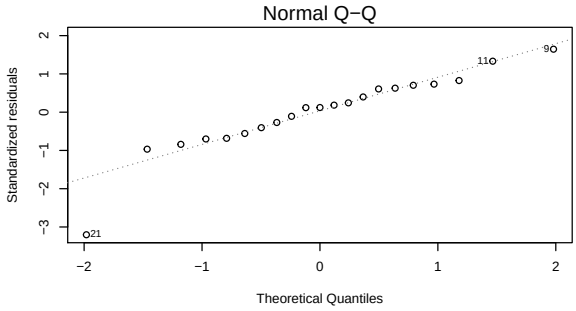

Residuals
etw

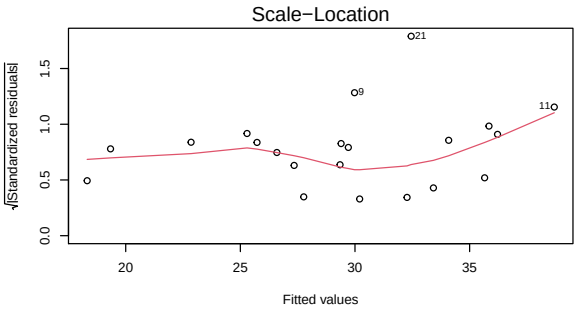

Residuals
cmp

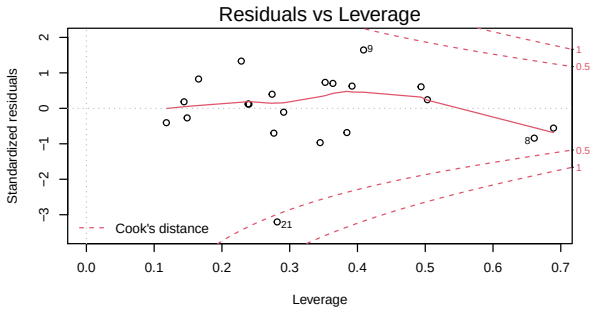

Residuals
axle


Residuals
n/v


Residuals vs Fitted
Residuals
Fitted values


Normal Q−Q
Standardized residuals
Theoretical Quantiles


Scale−Location
√|Standardized residuals|
Fitted values


Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage