DATA605: Fundamentals of Computational Mathematics

Assignment 11

Donald Butler

05/07/2022

Regression Analysis

Using the cars dataset in R, build a linear model for stopping distance as a function of speed and replicate the analysis of your textbook Chapter 3 (visualization, quality evaluation of the model, and residual analysis).

Load cars dataset

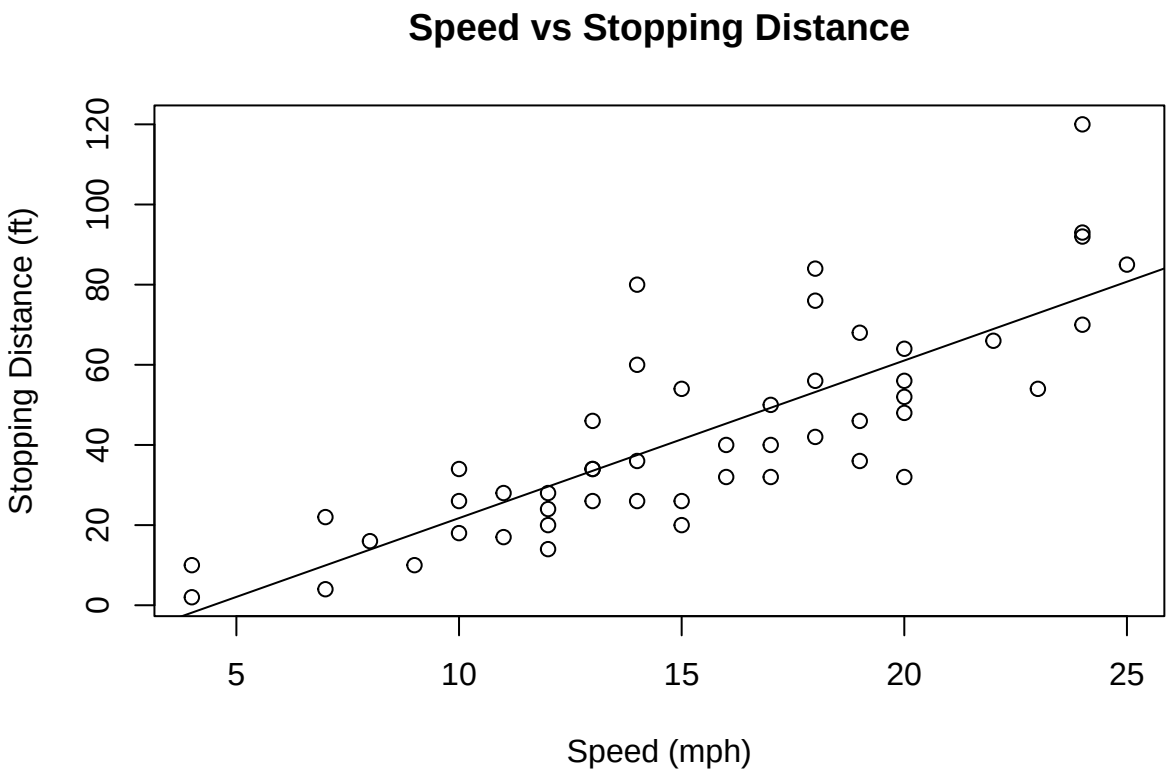
The cars dataset was collected in the 1920s and contains the speed in miles per hour and the stopping distance in feet.

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Construct Linear Model

```
cars.lm = lm(formula = dist ~ speed, data = cars)
plot(formula = dist ~ speed, data = cars, main = 'Speed vs Stopping Distance', xlab = 'Speed (mph)', yla
abline(cars.lm)
```



Model Evaluation

```
summary(cars.lm)
```

```
##
## Call:
## lm(formula = dist ~ speed, data = cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601   0.0123 *
## speed         3.9324     0.4155   9.464 1.49e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.38 on 48 degrees of freedom
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

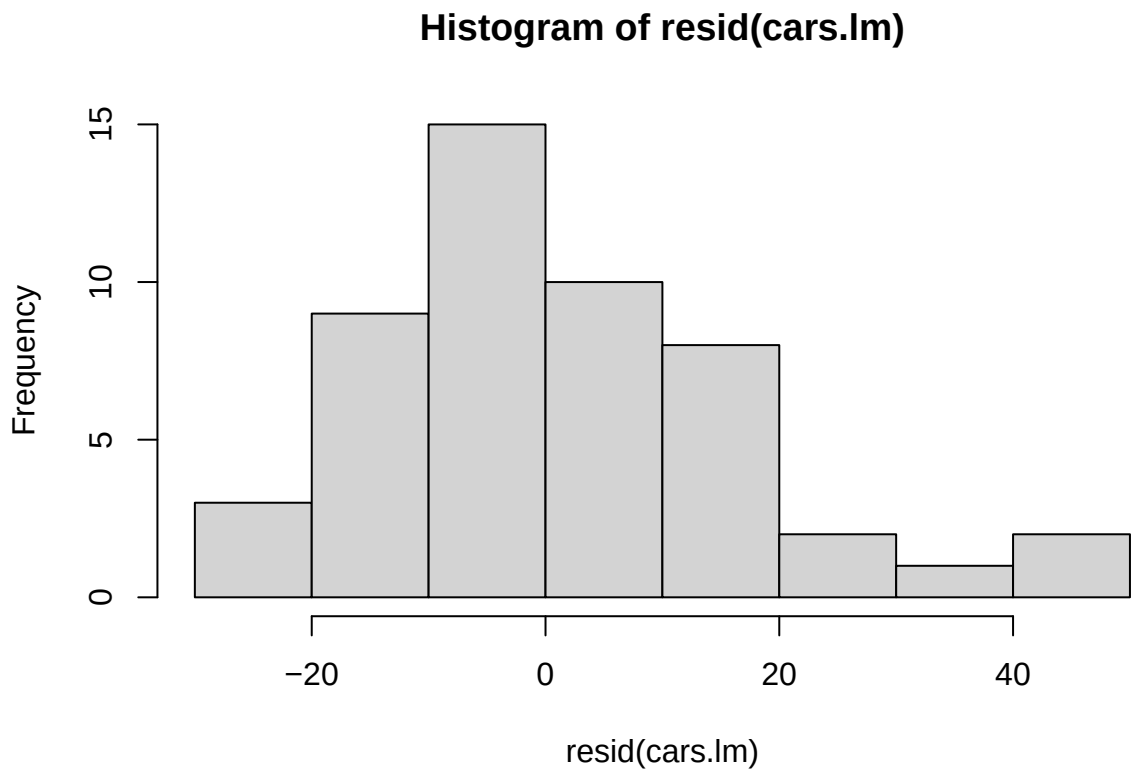
The p-value of 1.5e-12 is less than .05 which indicates that speed is a significant variable. The R^2 value

indicates that variations in speed account for 65% of the variation in stopping distance. The inter-quartile range is consistent with the residual standard error which supports that the residuals are normally distributed.

Residual Analysis

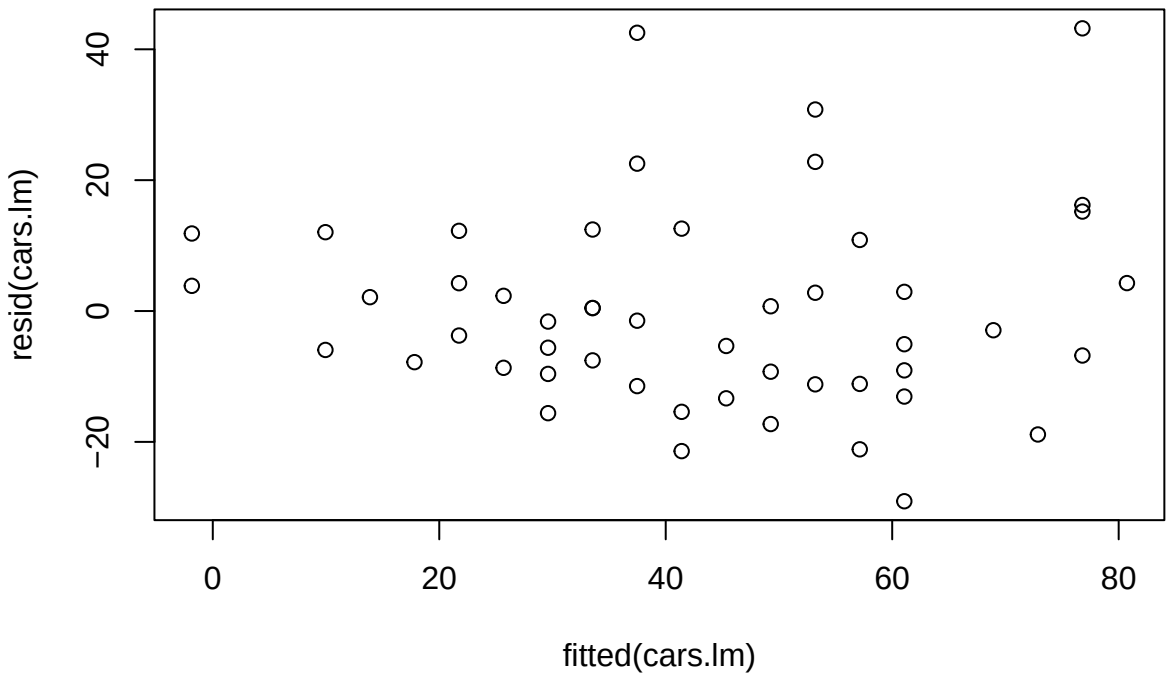
Viewing a histogram of the residuals shows that they are nearly normal centered at 0.

```
hist(resid(cars.lm))
```



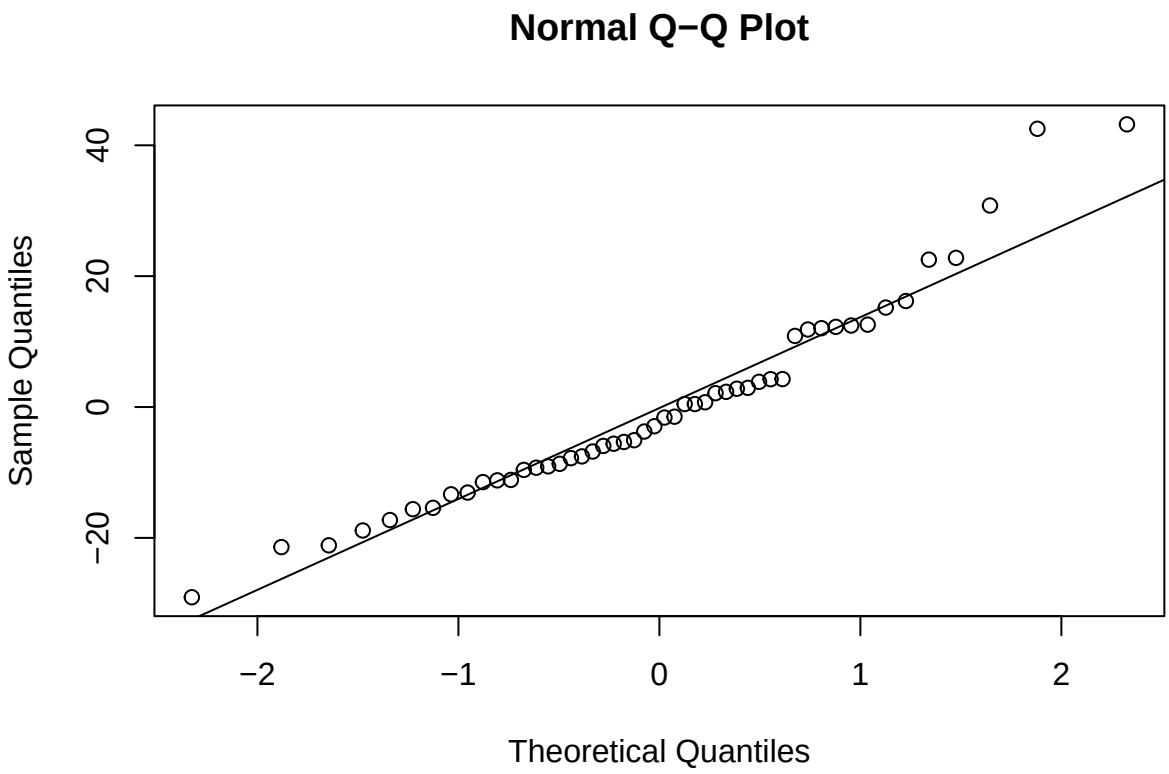
The residuals appear to have constant variability.

```
plot(fitted(cars.lm),resid(cars.lm))
```



The Q-Q plot of residuals shows that they are distributed nearly normal. There are a few outliers at higher speeds, but not too concerning.

```
qqnorm(resid(cars.lm))
qqline(resid(cars.lm))
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
40
20
0
−20
−2
−1
0
1
2