KSooklall_Homework11

Kenan Sooklall

4/8/2021

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.)

```
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
```

```
lmodel = lm(dist ~ speed, data=cars)
summary(lmodel)
```

```
##
## 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
```
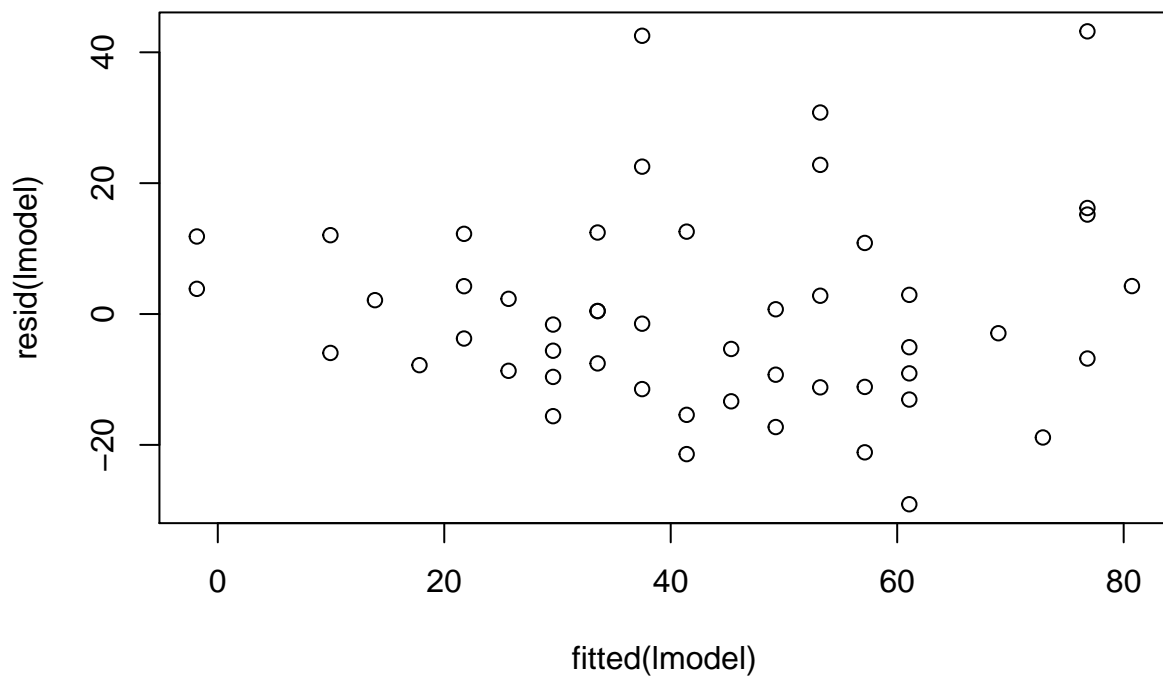
For every unit of speed (mph) that a car travels the distance the car will stop when the breaks is slammed is 4ft. The r^2 of 0.65 implies that 65% of the variance on the distance the car traveled is explained by the speed.

```r
cars %>% ggplot(aes(speed, dist)) + geom_point() + stat_smooth(method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```
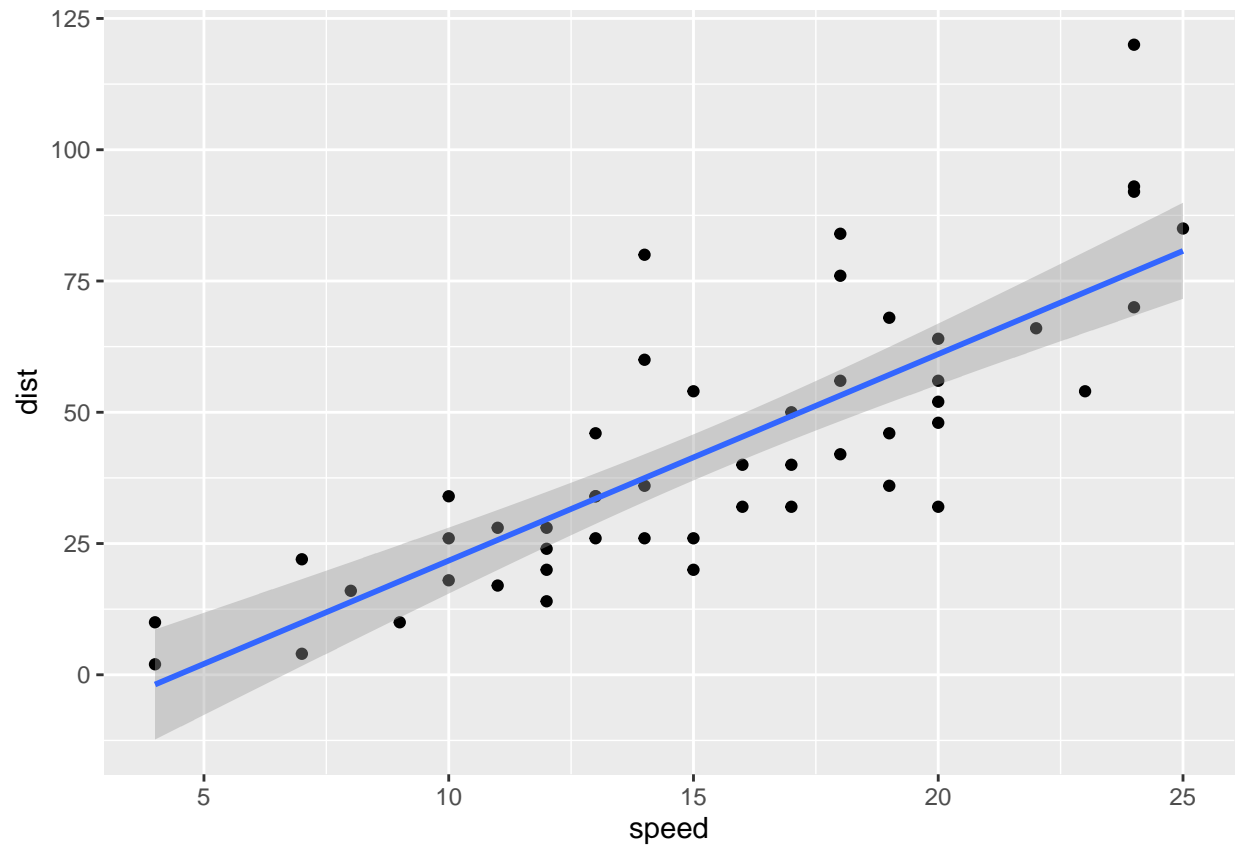


```r
plot(fitted(lmodel),resid(lmodel))
```

The residuals are some what scattered above and below zero. Overall, this plot tells us that using the regression model is an ok predictor between speed and distance.