DATA 605 - Homework 11

Manolis Manoli

```
library(ggplot2)
library(psych)
library(dplyr)
library(knitr)
library(tidyr)
library(GGally)
```

Cars - stopping distance vs speed

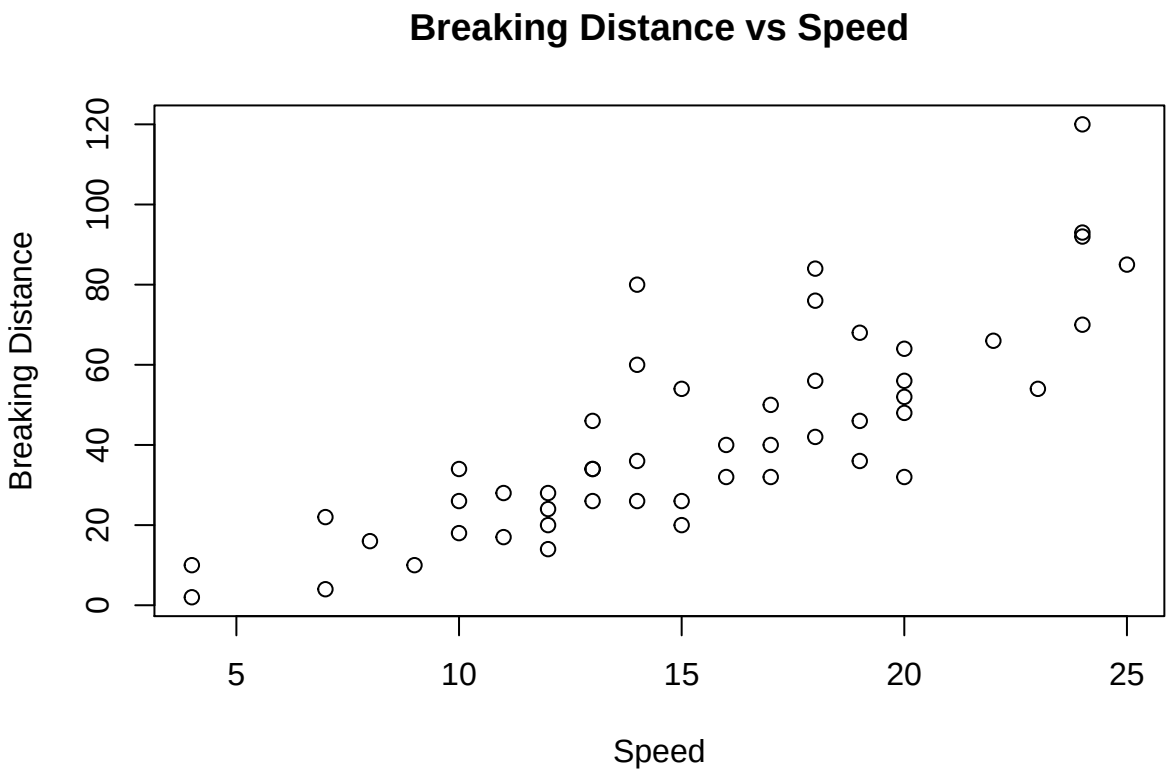
Using the "cars" dataset in R, build a linear model for stopping distance as a function of speed and replicate the analysis of your textbook chapter 3 (visualization, quality evaluation of the model, and residual analysis.)

Solution

```
data("cars")
```

Visualize the Data

```
plot(cars$speed,cars$dist, main="Breaking Distance vs Speed",
xlab="Speed", ylab="Breaking Distance")
```



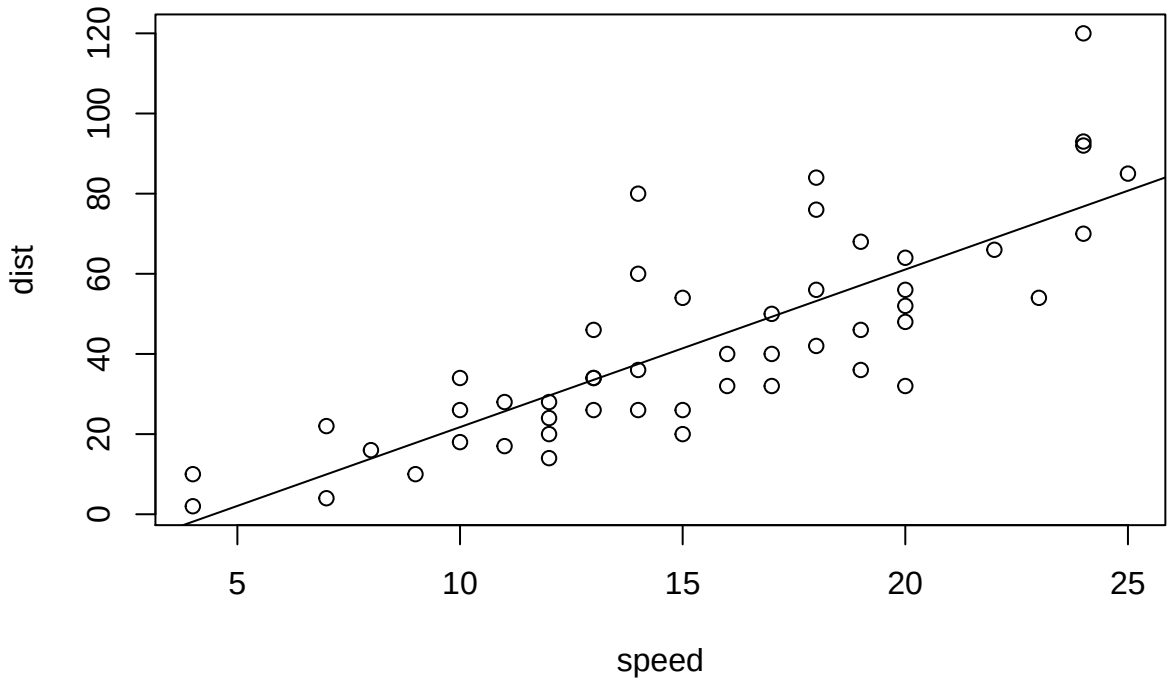
As expected, there is a general tendency of increasing distance with increasing speed.

The Linear Model Function

```
attach(cars)
cars.lm <- lm(dist ~ speed)
cars.lm
```

```
##
## Call:
## lm(formula = dist ~ speed)
##
## Coefficients:
## (Intercept)        speed
##     -17.579        3.932
```

```
plot(speed,dist)
abline(cars.lm)
```



Evaluating the Quality of the Model

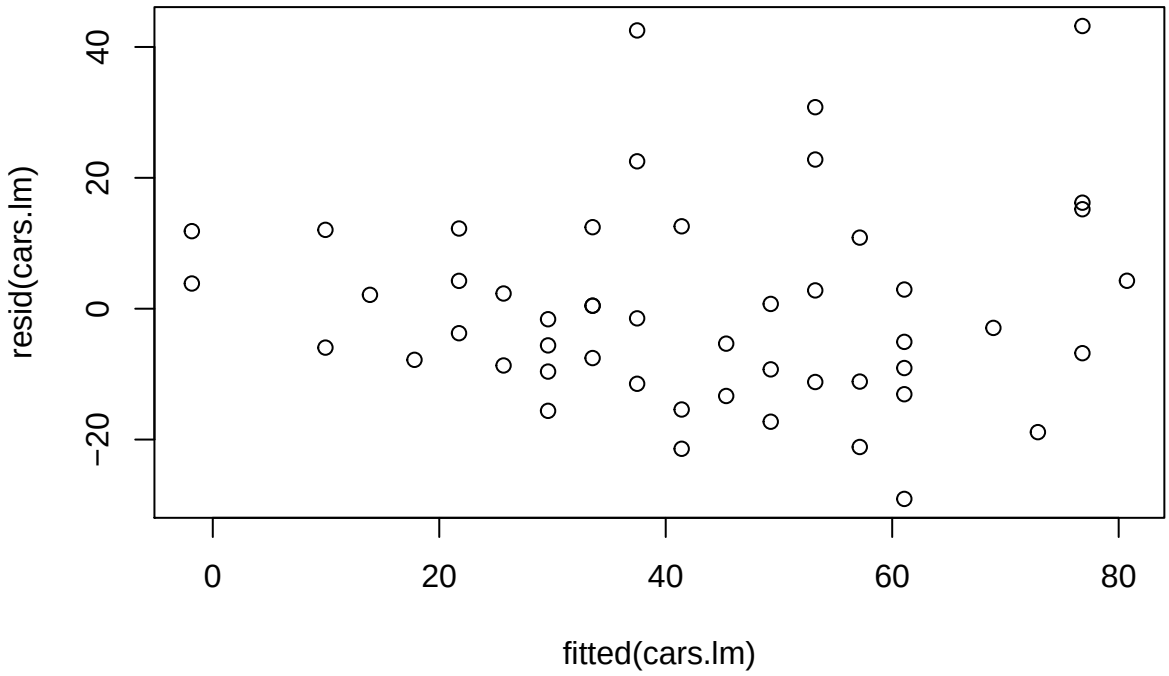
```
summary(cars.lm)
```

```
##
## Call:
## lm(formula = dist ~ speed)
##
## 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 is very small and R-squared of 65% which in comination would indicate that it's a good fit

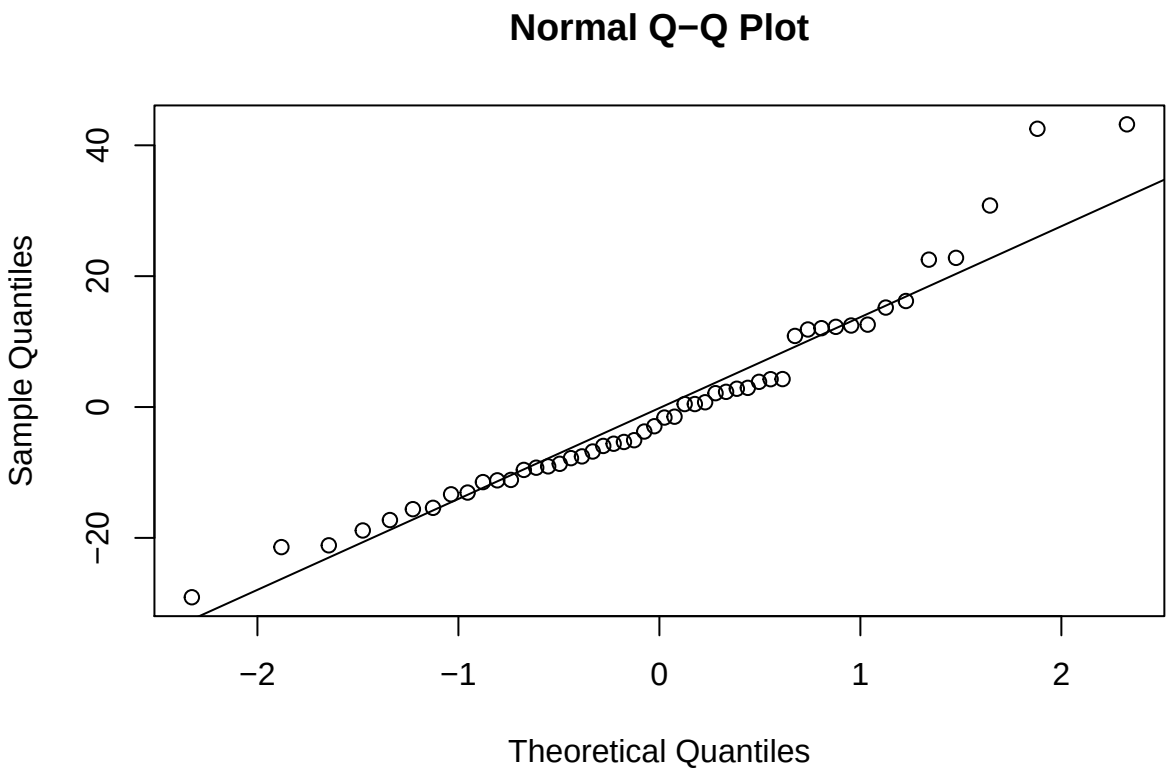
Residual Analysis

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



It seems like there is a little bit of a U shape in the residual. The errors in the middle of the plot seem to be mostly below zero.

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



We can see that the tails are diverging from the line slightly which might indicate slightly fatter tails than that of a normal distribution.

Overall, I would have assumed a quadratic dependency between speed and distance - however I actualyl do not see this very clearly in the plot we have used so far.

https://github.com/chilleundso/Data605_CompMath/tree/master/Homework11