Car_Data-MiriamDalfin.R

mdalf

2024-12-19

```r
#Reading in the vehicle file
auto_info <- read.csv("auto-mpg(1).csv")
head(auto_info)
```

```
##   mpg cylinder displacement horsepower weight acceleration modelyear
origin
## 1  18        8          307        130   3504         12.0        70
1
## 2  15        8          350        165   3693         11.5        70
1
## 3  18        8          318        150   3436         11.0        70
1
## 4  16        8          304        150   3433         12.0        70
1
## 5  17        8          302        140   3449         10.5        70
1
## 6  15        8          429        198   4341         10.0        70
1
##                   car.name
## 1 chevrolet chevelle malibu
## 2         buick skylark 320
## 3        plymouth satellite
## 4             amc rebel sst
## 5               ford torino
## 6          ford galaxie 500
```

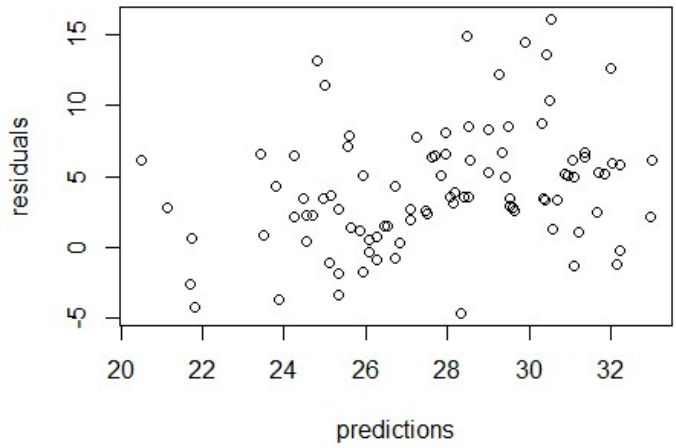
```r
#Turning horsepower into numeric data
auto_info$horsepower <- as.numeric(auto_info$horsepower)
```

```
## Warning: NAs introduced by coercion
```

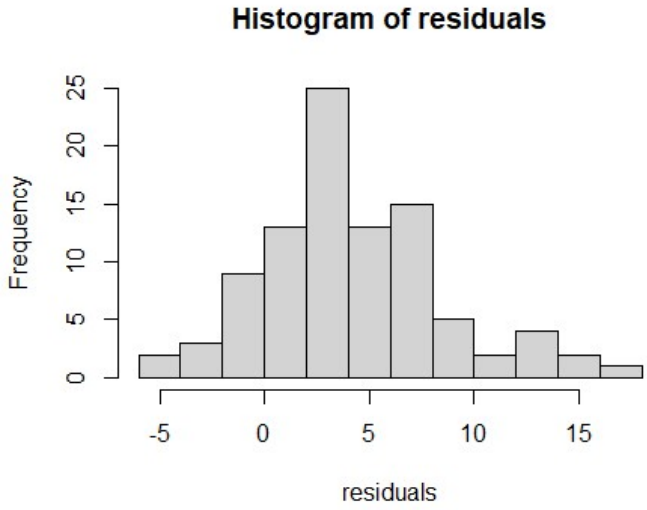
```r
#Splitting the data
train <- auto_info[1:300,]
test <- auto_info[301:398,]
#Creating the linear regression model
model <- lm(mpg ~
displacement+horsepower+weight+acceleration+modelyear+origin, data = train)

#Getting a summary of the model
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ displacement + horsepower + weight + acceleration +
```

```
##     modelyear + origin, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.0368 -1.6125  0.0888  1.5158 13.5305
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.1633922  4.9652103   1.040  0.29924
## displacement  0.0035638  0.0051570   0.691  0.49008
## horsepower   -0.0148579  0.0119785  -1.240  0.21583
## weight       -0.0054700  0.0005656  -9.672  < 2e-16 ***
## acceleration -0.0142687  0.0954542  -0.149  0.88128
## modelyear     0.4357457  0.0604479   7.209 4.89e-12 ***
## origin        0.9875643  0.3000279   3.292  0.00112 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.689 on 291 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.8216, Adjusted R-squared:  0.8179
## F-statistic: 223.4 on 6 and 291 DF,  p-value: < 2.2e-16

#Multiple R-squared:  0.7832,   Adjusted R-squared:  0.7802
# mpg = 40.3082911 - 0.0064554 * displacement - 0.0245072 * num_horsepower
# - 0.0046438 * weight - 0.0538346 * acceleration

#Check to see how well the test data fits the model
predictions <- predict(model, newdata = test)
summary(predictions)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   20.49   25.56   27.95   27.77   30.38   33.00       4

# Find the residuals of the predictions
residuals <- test$mpg - predictions

# Make a scatterplot
plot(predictions, residuals)
```




```
#Make a histogram
hist(residuals)
```




```
#This report shows three values that have a significant effect on the car's
mpg:
# model year - for every year later the car came out, on average a car will
have .43 more miles per gallon
# The origin also has a significant effect, with every increase in a unit of
origin, the car has about 1 extra mile per gallon
# Weight has a negative relationship with mpg so for every decrease of 100
lbs on average the car will go .54 extra miles per gallon

Something I changed in the dataset is that I changed the column model year to
be one word - modelyear
```