Coursera Regression Models Course Project

Sooihk Ro

5/16/2022

Objective

For this report the mtcars dataset will be analyzed to answer the following questions: 1) is an automatic or manual transmission better for MPG, and 2) Quantify the MPG difference between automatic and manual transmission.

Findings

The findings of the report indicate that manual transmission yields more miles per gallon than automatic transmission. On average a car with manual transmission gets .14 more mpg than automatic cars holding cyl, disp and wt variables constant.

Exploratory Analysis

```
library(ggplot2)
data(mtcars)
summary(mtcars)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am              gear            carb      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
##  Median :0.0000   Median :4.000   Median :2.000  
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812  
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```
head(mtcars)
```

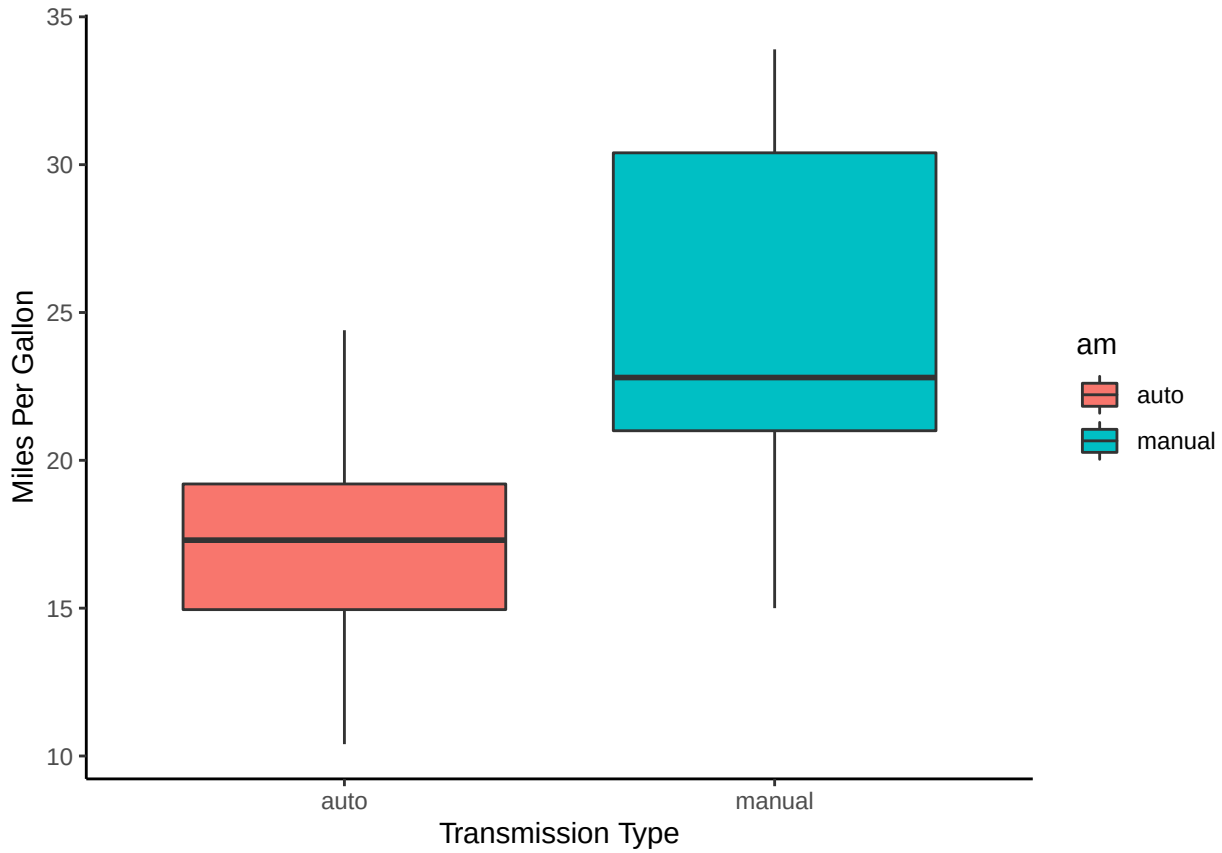
```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
#Transform variables into factors
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- as.factor(mtcars$am)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb)

#Replace mtcars$am variable from 0-1 to auto-manual
mtcars$am <- gsub("0", "auto", mtcars$am)
mtcars$am <- gsub("1", "manual", mtcars$am)
```

Figure 1. is a boxplot to visualize the effect between automatic and manual transmission on mpg. The figure displays that manual transmission has better mileage per gallon than automatic transmission.

```
boxFig <- ggplot(mtcars, aes(x=am, y = mpg, fill=am)) + geom_boxplot() + theme_classic() +
labs(x="Transmission Type", y="Miles Per Gallon")
boxFig
```



Statisical Inference

A t.test function is utilized to test our hypothesis that a manual transmission provides higher miles per gallon than a automatic transmission is statistically significant. The t.test function yields automatic transmission having on average 7 miles per gallon than manual transmission. We can conclude that the difference in mpg between automatic and manual transmission with a 95% confidence interval not containing a value 0 and a small p value.

```
t.test(mpg ~ am, mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means between group auto and group manual is not equal to
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
##   mean in group auto mean in group manual
##             17.14737             24.39231
```

Regression Model

We start with regressing mpg on the am variable alone. The summary of the model shows an average of 17.1 mpg for automatic and an average 7.2 mpg increase for manual. The R-squared value of .36 tells us that the variable am only explains 36% of the variance.

```
am_regression_model <- lm(mpg~am, mtcars)
summary(am_regression_model)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## ammanual       7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Additional variables will be included to the model to better explain the % variance. Analysis of variance (aov) function will be implemented to the data in order to obtain such variables.

```
variance_analysis <- aov(mpg ~ ., mtcars)
summary(variance_analysis)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## cyl          2  824.8   412.4  51.377 1.94e-07 ***
## disp         1   57.6    57.6   7.181   0.0171 *
## hp           1   18.5    18.5   2.305   0.1497
## drat         1   11.9    11.9   1.484   0.2419
## wt           1   55.8    55.8   6.950   0.0187 *
## qsec         1    1.5     1.5   0.190   0.6692
## vs           1    0.3     0.3   0.038   0.8488
## am           1   16.6    16.6   2.064   0.1714
## gear         2    5.0     2.5   0.313   0.7361
## carb         5   13.6     2.7   0.339   0.8814
## Residuals   15  120.4     8.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

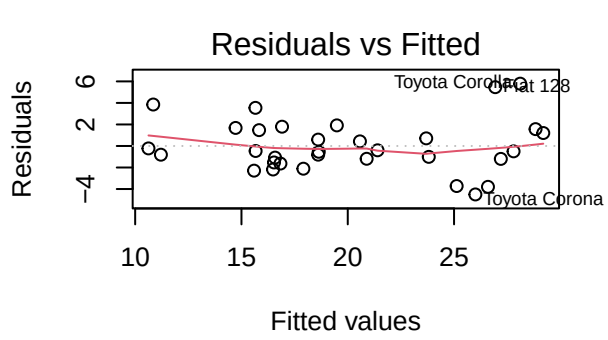
From the summary variables cyl, disp and wt have p values less than .05 which are significant.

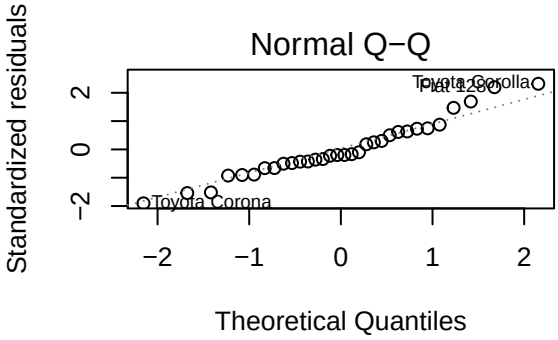
```
multi_regression <- lm(mpg ~ cyl + disp + wt + am, mtcars)
summary(multi_regression)
```

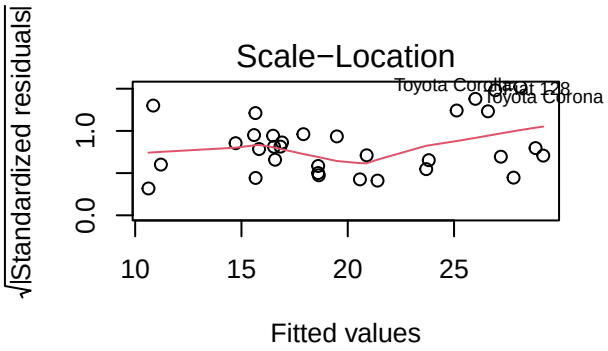
```
##
## Call:
## lm(formula = mpg ~ cyl + disp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5029 -1.2829 -0.4825  1.4954  5.7889
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.816067   2.914272  11.604 8.79e-12 ***
## cyl6        -4.304782   1.492355  -2.885  0.00777 **
## cyl8        -6.318406   2.647658  -2.386  0.02458 *
## disp         0.001632   0.013757   0.119  0.90647
## wt          -3.249176   1.249098  -2.601  0.01513 *
## ammanual     0.141212   1.326751   0.106  0.91605
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.652 on 26 degrees of freedom
## Multiple R-squared:  0.8376, Adjusted R-squared:  0.8064
## F-statistic: 26.82 on 5 and 26 DF,  p-value: 1.73e-09
```

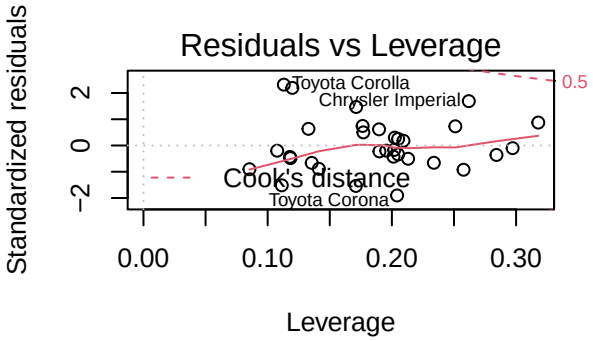
The multivariable regression model has an R-squared value of .84 which states the model explains 84% of the variance. The difference between automatic and manual transmmission is 0.14 miles per gallon. Diagnostic plotting shows the residuals are homoscedastic with the Residuals vs Fitted plot. The Normal Q-Q plot shows the distribution of the residuals is roughly normal.

```
par(mfrow=c(2,2))
plot(multi_regression)
```


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Toyota Corona


Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Fiat 128
Toyota Corona


Scale−Location
√|Standardized residuals|
Fitted values
Toyota Corolla
Fiat 128
Toyota Corona


Residuals vs Leverage
Standardized residuals
Leverage
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's distance
0.5