Regression Models Class Project

Executive Summary

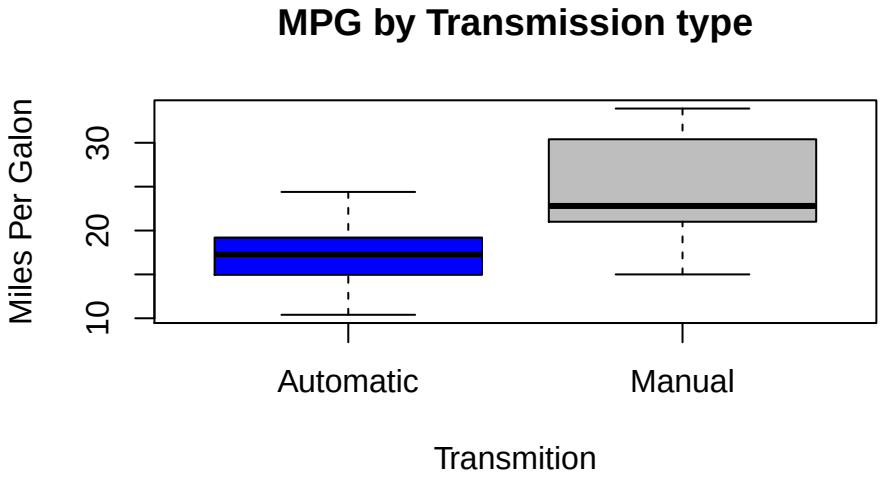
To help our subscribers select a car with the best gas milage, we are going to perform an analysys of MPG for Automatic vs Manual transmission. We will begin by looking at a data set for a collection of 1974 cars. During our analysis, it will be important to discuss how each feature set impacts MPG. Its even more important to compare like for like feature sets for Automatic and Manual Transmissionts

Exploratory Data Analysis

The data set consists of 32 observations with 11 variables or potential feature sets.

- **mpg**: Miles/(US) gallon
- **cyl**: Number of cylinders
- **disp**: Displacement (cu.in.)
- **hp**: Gross horsepower
- **drat**: Rear axle ratio
- **wt**: Weight (lb/1000)
- **qsec**: 1/4 mile time
- **vs**: V/S
- **am**: Transmission (0 = automatic, 1 = manual)
- **gear**: Number of forward gears
- **carb**: Number of carburetors

To better understand the data, lets look at a simple box plot to identify if there is a difference between the mean MPG for Automatic and Manual Transmissions. Data Analysis.pdf



This can further be verified by looking at the inference using a **T** test:

The mean MPG for the manual transmission cars is **24.3923**, or about **7.2449** higher MPG than the automatic cars. *See Figure 1 - Means Analysis*

Regression Analysis

Another method to analyze the data is to create a simple regression model model for MPG to Transmission types. However this model does not offer additional information compared to the T test. The **Intercept** of this model (17.1474) represents the mean MPG for Automatics. The **AM coefficient** (7.2449) represents the different MPG between the Automatic vs Manual Transmissions. These results are the same as the T-Test perfomred above.

The only conclusion we can make from this is that MPG is affected by other fatures of the data set.

By using the **COR** function in R we can determine how each field is related to the MPG.

From *Figure 2. (Correlations Analysis)*, we can see that wt-Weight, cyl-Number of cylinders, disp-Displacement (cu.in.), and hp-Gross horsepower and carb-Number of carburetors have a negative impact on MPG. While qsec-1/4 mile time, gear-Number of forward gears, vs(not sure what this is) and drat-Rear axle ratio have a positive impact on MPG.

The comparison of teh complex model adn teh simple model shows a very small ***p-value***(5.6466 $\times$ 10-6), which allows us to reject the **Null Hypothesis**. This indicates that the complex model may explain the explain the *MPG* better than the simple model. *See Figure 3 - Model COmparision*

Conclusion

By analysing the new multivariable model, we can see how MPG is affected by each variable. *See Figure 4 - Model Summary*

With a **R-squared** value of (0.8678), this model explains the variance in MPG when looking at all the variables in the data set. The order of impact to MPG is a follows.

1) The greatest impact on **MPG** is not the type of transmission in the car, but the weight of the car. For every 1000 lbs, **MPG** is decreased by avg of **-3.8661 MPG**
2) A Manual transmission will add an average of **2.8034** MPG.
3) Lastly the **drat**: Rear axle ratio and the **qsec**: 1/4 mile time have an impact on MPG of (0.8409) and (0.7951) respectively.

In summary, if a consumer wants to get the most **MPG** from there next car, they should first look at the lighter cars first then the Transmission type.

Apendix

Figure 1 - Means Analysis

```
print(means)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.767, df = 18.33, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.28  -3.21
## sample estimates:
## mean in group Automatic    mean in group Manual
##                   17.15                   24.39
```

Figure 2 - Correlations Analysis

```
sort(cor(mtcars_raw)[1,])
```

```
##      wt     cyl    disp      hp    carb    qsec    gear      am      vs
## -0.8677 -0.8522 -0.8476 -0.7762 -0.5509  0.4187  0.4803  0.5998  0.6640
##    drat     mpg
##  0.6812  1.0000
```

Figure 3 - Model Comparision

```
print(compare_models)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + cyl + disp + hp + carb + qsec + am + vs + drat
##   Res.Df RSS Df Sum of Sq    F  Pr(>F)
## 1     30 721
## 2     22 149  8       572 10.6 5.6e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4 - Model Summary

```
summary(complex_model)
```

```
##
## Call:
## lm(formula = mpg ~ wt + cyl + disp + hp + carb + qsec + am +
##     vs + drat, data = mtcars_raw)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.518 -1.434 -0.392  1.202  4.592
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.6418    16.7861    0.93    0.362
## wt           -3.8661     1.8285   -2.11    0.046 *
## cyl          -0.2732     0.9598   -0.28    0.779
## disp          0.0140     0.0175    0.80    0.433
## hp           -0.0206     0.0213   -0.97    0.343
## carb         -0.0451     0.7365   -0.06    0.952
## qsec          0.7951     0.7150    1.11    0.278
## am            2.8034     1.9166    1.46    0.158
## vs            0.3580     2.0636    0.17    0.864
## drat          0.8409     1.6006    0.53    0.605
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.6 on 22 degrees of freedom
## Multiple R-squared:  0.868,  Adjusted R-squared:  0.814
## F-statistic:   16 on 9 and 22 DF,  p-value: 9.89e-08
```