Impact of Transmission on Fuel Efficiency

Kumar Chandrakant

January 21, 2016

Executive Summary

The objective of this analysis is to establish if transmission type (automatic vs manual) has any statistical significance on the miles per gallon (mpg) of the motor vehicles. We will use the dataset “mtcars” for this analysis which is present in R by default.

Exploratory Analysis

We will begin by loading the “mtcars” dataset in the environment and perform some exploratory analysis.

The data frame consists of 32 rows of data with 11 variables.

```
dim(mtcars)
```

```
## [1] 32 11
```

A quick check of the data franme revealks that all the variables are of type num. We will convert the variable transmission (am) of type factor.

```
mtcars$am <- as.factor(mtcars$am)
```

As clearly evident from the boxplot in annexure the mean miles per gallon (mpg) for manual transmission is higher than that for automatic transmission. We will proceed to see if this observation can be corelated statistically as well.

Model Selection

To select a suitable model for the data we will perform the exercise of reducing the model whereby we will start with all predictors and in step wise manner remove the ones with highest p-value which indicates least statistical significance of the predictor in the model. Refer to annexure for details.

We begin as said with a model with all predictors.

```
fit <- lm(mpg ~ ., data = mtcars)
summary(fit)$r.squared
```

```
## [1] 0.8690158
```

After a series of steps removing less significant predictors we reach the model below.

```
fit <- lm(mpg ~ .-carb-gear-cyl-vs-drat-disp-hp, data = mtcars)
summary(fit)$r.squared
```

```
## [1] 0.8496636
```

As we can see that retaining just few predictors the final model is able to explain most of the variance as suggested by the Adjusted R-Squared value.

Diagnostic Analysis

Next we will analyse if the chosen model has any problem like heteroskedasticity using the residual plot. The plot of residuals as present in the annexure for the chosen model suggest no specific pattern (random dispersion around horizontal axis) and hence looks suitable to be used.

Statistical Inference

Let us now use the selected model to make the statistical inference using the coefficients of predictors and uncertainty involved.

- Interpretation of Coefficients

Let us examine the coefficients of the regressors as present in the annexure. The coefficient of the intercept suggests the mean mpg for automatic transmission "Intercept" and coefficients of the other regressors relative to this. Thus sum of coefficient of intercept and coefficient of regressor "am1"" gives the mean mpg for manual transmission.

- Confidence Interval of Slope (Uncertainty)

We will calculate the 95% confidence interval for the coefficient of the predictor "am1" to understand the variability involved.

```
confint(fit, 'am1', level=0.95)
```

```
##          2.5 %   97.5 %
## am1 0.04573031 5.825944
```

This shows that we are 95% confident that the slope will lie in the above range. Also as zero is not in this range we are 95% confident that manula transmission gives better miles per gallon compared to automatic transmission.

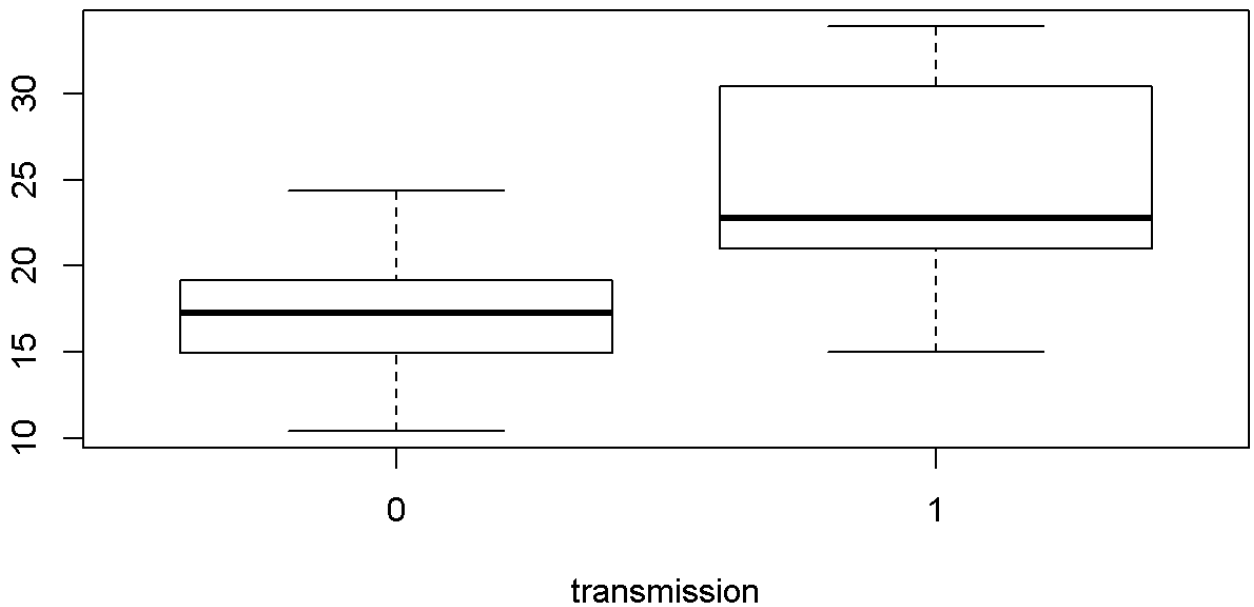
Conclusion

- Is an automatic or manual transmission better for MPG? ==> The mpg for automatic transmission is `9.6177805` while that for manual transmission is `12.5536177` .
- Quantify the MPG difference between automatic and manual transmissions. ==> The difference in mpg between automatic and manula transmission is `2.9358372` .

Annexure

BoxPlot of MPG vs. Transmission

```
boxplot(mpg ~ am, data = mtcars, xlab = "transmission")
```



Initial and Final Model as part of Selection

Initial Model with all predictors.

```
fit <- lm(mpg ~ ., data = mtcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am1          2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

Final model after a series of steps removing less significant predictors.

```
fit <- lm(mpg ~ .-carb-gear-cyl-vs-drat-disp-hp, data = mtcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ . - carb - gear - cyl - vs - drat - disp -
##     hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am1           2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Plot of Residuals for the Final Model

```
plot(fit, which=1)
```