Motor Trend Analysis: Does Transmission affect Mileage(mpg) of a Car

S Zaki

4/22/2018

Executive Summary

In this analysis we will use the mtcars data set and fit different models to determine if the type of transmission (which is a binomial variable) the predictor, has an effect on a cars mileage or miles per gallon(outcome) and to quantify the difference between the mean mpg for automatic vs manual transmission cars and fit models to the data.

Exploring the data set

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

```
##                    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
```

Diffrence in mean between Automatic and Manual transmission

The diffrence in mean mpg between automatic and manual transmission is 7.2449393

Fitting Linear Regression Model

```r
fit1 <- lm(mpg ~ am,mtcars)
e <- resid(fit1)
confint(fit1)
summary(fit1)$r.squared
```

Hypothesis Testing: Null hypothesis H0: beta1 = 0; Ha:beta1 != 0. The p value for the slope estimate is 2.850207e-04 which is highly significant and therefore we reject the null hypothesis (that there is no linear relationship between the transmission) predictor and mpg(outcome) for the alternative hypothesis (that is there is a linear relationship between transmission and mpg). According to the slope coefficient we can expect a 7.24 increase in mpg for cars that are group 1-manual transmission when compared to cars that belong to group 0-automatic transmission. R squared value shows that this model explains only ~36% of variability. **Confidence Interval:** The confidence interval for the slope is [3.64151, 10.84837]. We can say with 95% confidence that changing from automatic to manual transmission will result in a 3.64151 to 10.84837 increase in miles per gallon. **Residual Inference:** The residual plot shows the residuals stacking up at the two ends as our predictor is binary. The plot shows a specific pattern, which may be an indication of a poor model fit.

Maybe adding more variables to our model will improve the R squared value and give a better fit.

Fitting Nested Models

```
anova(fit1,fit2,fit3,fit4,fit5,fit6,fit7,fit8,fit9)
```

From the above we find that fit2 and fit5 are of interest

Calculating R Squared fot fit1,fit2 and fit5

```
fit1 <- lm(mpg ~ am,mtcars)
fit2 <- lm(mpg ~ am + cyl,mtcars)
fit5 <- lm(mpg ~ am + cyl + disp + hp + wt,mtcars)
##Comparing the coeffcients of the fits
summary(fit1)$r.squared ## original fit with mpg outcome and trans as predictor
```

```
## [1] 0.3597989
```

```
summary(fit2)$r.squared
```

```
## [1] 0.7590135
```

```
summary(fit5)$r.squared
```

```
## [1] 0.8551394
```

We see from the R squared value that fit2 explains ~76% and fit5 ~85% only of total variability that is explained by the linear relationship with the predictor transmission. However we also know that adding more regressors can lead to variance inflation. **Checking for Varaiance Inflation**

```
##        am       cyl      disp        hp        wt
##  2.553064  7.209456 10.401420  4.501859  6.079452
```

Disp is inflating the variance a lot since it is correlated with other variables of interest. So, we fit a model with all variables from fit5 but without disp and see how that affects vif and r squared values.

Fittiing final model and checking VIF and R Squared

```
##       am      cyl       hp       wt
## 2.546159 5.333685 4.310029 3.988305
```

```
## [1] 0.8490314
```

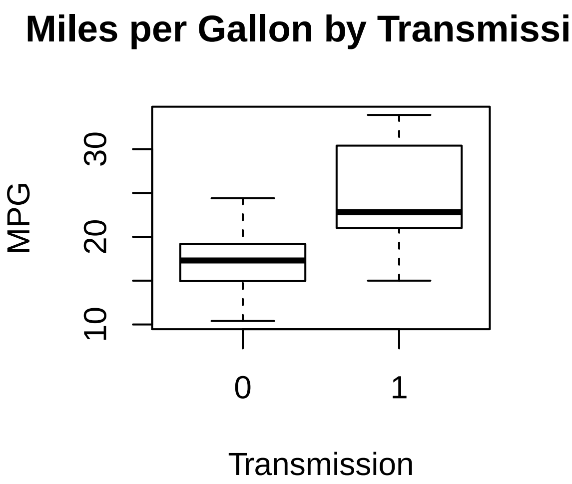
Removing disp reduces the variance inflation in our model and the r squared value is still close to ~ 85%. Therefore we can conclude that this might be the best fit for our data. We see that basic linear regression model shows that the manual transmission is significantly better than the automatic transmission and that the difference between the mean mpg for two types is 7.2449393mpg. We can improve over this model by adding other regressors to improve R squared while keeping variance inflation low and come up with the most suitable model.

Conclusion:

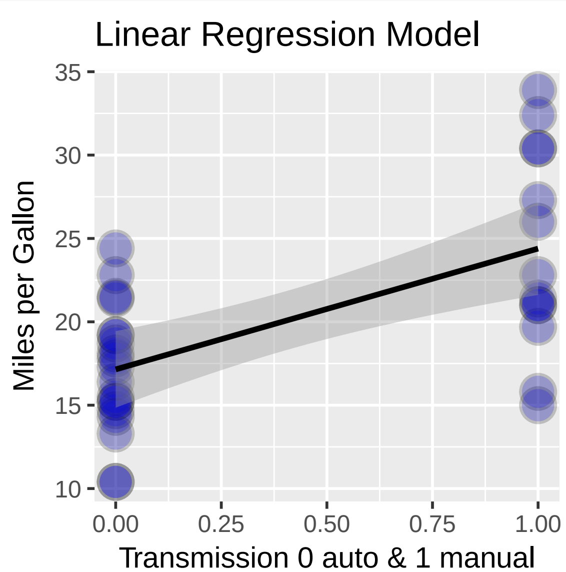
From the nested models we found that the model which takes into account cyl, hp and wt along with mpg and am has the best fit for our data. Overall we can conclude that for this particular analysis where transmission type is the predictor and miles per gallon is the outcome 1) Manual transmission cars are better for MPG than automatic. 2) The difference in mean MPG between the manual and automatic transmission cars is 7.2449393.

Appendix:

```
boxplot(mpg~am,data=mtcars,main="Miles per Gallon by Transmission",
        xlab="Transmission",ylab="MPG")
```

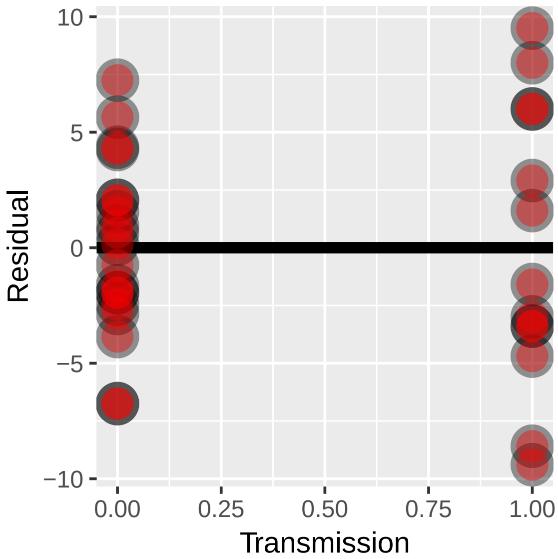



```
par(mfcol=c(1,2))
##Linear Regression Plot using ggplot lm method
g <- ggplot(mtcars,aes(x=am,y=mpg))
g <- g + labs(title="Linear Regression Model")
g <- g + xlab("Transmission 0 auto & 1 manual")
g <- g + ylab("Miles per Gallon")
g <- g + geom_point(size=6,color="black",alpha=0.2)
g <- g + geom_point(size=5,color="blue",alpha=0.2)
g <- g + geom_smooth(method=lm,color="black")
g
```




```
##Residual plot
h <- ggplot(mtcars,aes(x = am, y = e))
h <- h + geom_hline(yintercept = 0, size = 2)
h <- h + geom_point(size = 7,colour = "black", alpha = 0.4)
h <- h + geom_point(size = 5,colour = "red", alpha = 0.4)
```

```r
h <- h + xlab("Transmission") + ylab("Residual")
h
```



Diagnostics of Linear Model with mpg and transmission only

```r
par(mfcol=c(2,2))
plot(fit1)
```