Regression Project.Rmd

Harold Trammel

September 7, 2015

Motor Trend Car Road Tests

Executive Summary

An analysis of the impact that transmission type has on gas mileage was performed on the *Motor Trend* car data. This data shows that automatic transmissions have negative impact on gasoline mileage when compared to manual transmissions. On average, cars in this data with automatic transmissions have 7.2449 fewer miles per gallon. This is a simplistic conclusion however as there were other variables such as car waight, horsepower, and engine displacemnt impacting this data. It is recommended that a decision to select a higher mileage car not be made simply on the transmission type.

Introduction

Motor Trend is interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They have provided a set of data from automobiles that they tested over a period of time. They are particularly interested in the following two questions:

- "Is an automatic or manual transmission better for MPG"
- "Quantify the MPG difference between automatic and manual transmissions"

A modern automobile is a complex machine and as such no one component is fully responsible for its performance and efficiency. Some of the carDF in the sample have been designed for performance while others were designed for performance or luxury. This inherent complexity must be kept in mind when looking a single complonent such as the type of transmission.

General Exploration

The data that *Motor Trend* provided was eleven different measured collected from 32 different automobiles. Three of the variables (Miles Per Gallon, Gross Horsepower, and 1/4 Mile Time) are dependent on different sets of the other variables. Some of the variables are factors even though their values are numeric. These are: Number of Cylinders, Vee vs Straight Cylinder Arrangement, Transmission Type, Number of Forward Gears, and Number of Carburators. Displacement, Rear Axle Ratio, and Weight are continous variables.

As part of the exploration phase, I generated correlation matrix, which is provided in Table 1. Analysis of this matrix identified Number of Cylinders, Gross Horsepower, Displacement, and Weight as having a negative correlation greater than 0.75. Rear Axle Ratio, Vee vs Straight Cylinder Arrangement, Transmission Type, and Number of Forward Gears had mild positive correlation with values between 0.48 and 0.68.

Focused Analysis

I compared the statistics related to milage recorded in cars with automatic transmissions to those with manual transmissions (see Table 2). The mean milage in automatic cars was 17.1474 $\pm$ 0.8796 (std. error of mean) with a standard deviation of 3.834. For manual transmission cars, the mean milage 24.3923 $\pm$ 1.7103 (std. error of mean) with a standard deviation of 6.1665.

Since there are only two types of tranmission, a linear relationship with mileage was obvious. I used a simple linear regression model to confirm my expectation and the resulting coeffients are shown in Table 3. The p-values were less that 0.05 but the variance seen in the basic statistics suggested more analysis was required. I prepared a plot of the residuals of this model (shown in Figure 1). This shows that the manual transmission has approximately a 20 mpg spread. I also prepared a residual QQ plot for this simple linear regression model and the residuals fell very close to the line. Based on this collection of results, it would seem that transmission type is a reasonable predictor of mileage; however, basic knowledge of automobile function warrants additional study.

Additional Analysis

I decided to look for possible factors that could be causing this. I decided to do nested model testing with other variables. I started with the Mileage to Transmission Type model and progressively added the following: Weight, Number of Carburators, Horsepower, Number of Cylindars, and Engine Displacement. (The ANOVA table is shown in Table ?????.) The addition of Weight had the strongest effect, which is understandable considering that lower weight cars had better gas mileage than heavier cars and that manual transmissions were more commonly seen in lower weight cars. This can bee seen in Figure 3.

Conclusion

Automatic transmissions have been maligned for causing lower gasoline mileage and this analysis supports that assertion. While transmission type plays a role in mileage, other factors are also involved.

Appendix

Table 1: Table 1 Correlation Matrix for All Variables

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	1.0000	-0.8522	-0.8476	-0.7762	0.6812	-0.8677	0.4187	0.6640	0.5998	0.4803	-0.5509
cyl	-0.8522	1.0000	0.9020	0.8324	-0.6999	0.7825	-0.5912	-0.8108	-0.5226	-0.4927	0.5270
disp	-0.8476	0.9020	1.0000	0.7909	-0.7102	0.8880	-0.4337	-0.7104	-0.5912	-0.5556	0.3950
hp	-0.7762	0.8324	0.7909	1.0000	-0.4488	0.6587	-0.7082	-0.7231	-0.2432	-0.1257	0.7498
drat	0.6812	-0.6999	-0.7102	-0.4488	1.0000	-0.7124	0.0912	0.4403	0.7127	0.6996	-0.0908
wt	-0.8677	0.7825	0.8880	0.6587	-0.7124	1.0000	-0.1747	-0.5549	-0.6925	-0.5833	0.4276
qsec	0.4187	-0.5912	-0.4337	-0.7082	0.0912	-0.1747	1.0000	0.7445	-0.2299	-0.2127	-0.6562
vs	0.6640	-0.8108	-0.7104	-0.7231	0.4403	-0.5549	0.7445	1.0000	0.1683	0.2060	-0.5696
am	0.5998	-0.5226	-0.5912	-0.2432	0.7127	-0.6925	-0.2299	0.1683	1.0000	0.7941	0.0575
gear	0.4803	-0.4927	-0.5556	-0.1257	0.6996	-0.5833	-0.2127	0.2060	0.7941	1.0000	0.2741
carb	-0.5509	0.5270	0.3950	0.7498	-0.0908	0.4276	-0.6562	-0.5696	0.0575	0.2741	1.0000

Table 2: Table 2 Mileage Impact by Transmission Type

Statistic	Automatic	Manual
number of values (nbr.val)	19.0000	13.0000
number of null values (nbr.null)	0.0000	0.0000
number of missing values (nbr.na)	0.0000	0.0000
minimal value (min)	10.4000	15.0000

Statistic	Automatic	Manual
maximal value (max)	24.4000	33.9000
range (range that is max-min)	14.0000	18.9000
sum of all non-missing values (sum)	325.8000	317.1000
median (median)	17.3000	22.8000
mean (mean)	17.1474	24.3923
standard error on mean (SE.mean)	0.8796	1.7103
confidence interval of mean (CI.mean)	1.8479	3.7264
variance (var)	14.6993	38.0258
standard deviation (std.dev)	3.8340	6.1665
variation coefficient (coef.var)	0.2236	0.2528

Table 3: Table 3 Automatic vs Manual Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.147	1.125	15.248	0.0000
amManual	7.245	1.764	4.106	0.0003

Figure 1 Residuals of Transmission vs Mileage

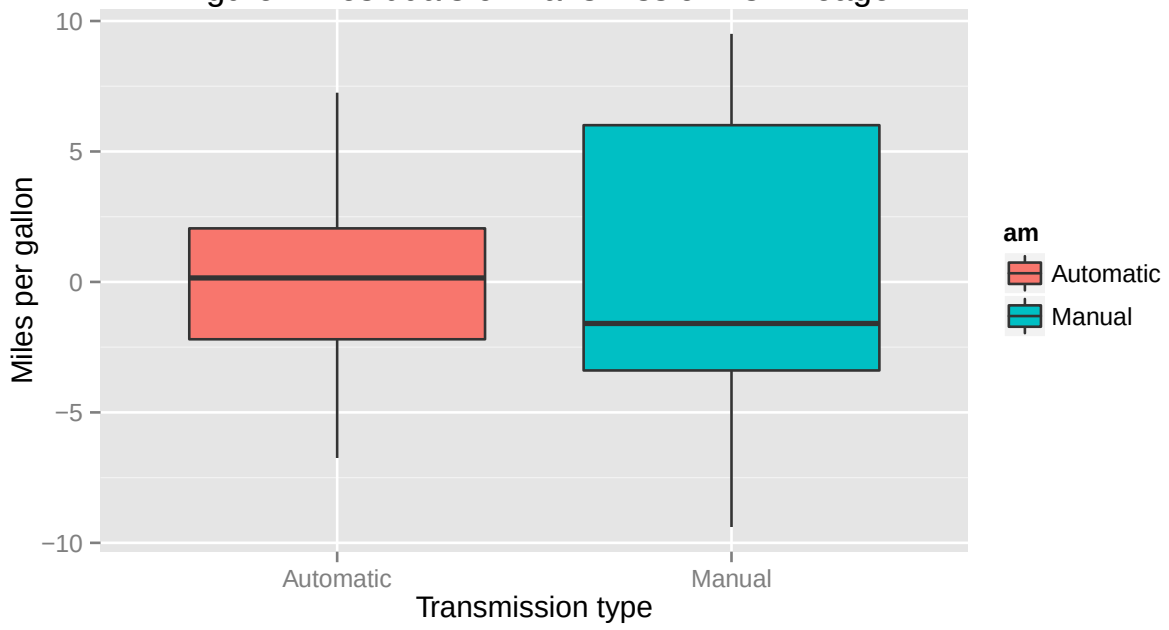

Figure 2 QQ Plot of Transmission to Mileage LRM

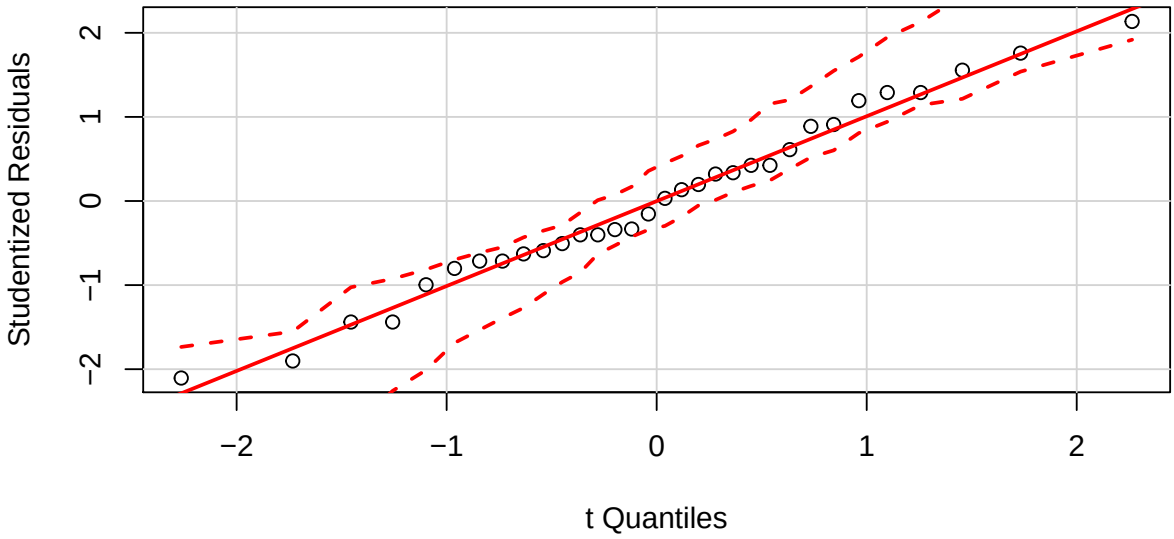

Table 4: ANOVA Table of Nested Models

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + carb
## Model 4: mpg ~ am + wt + carb + hp
## Model 5: mpg ~ am + wt + carb + hp + cyl
## Model 6: mpg ~ am + wt + carb + hp + cyl + disp
##   Res.Df RSS Df Sum of Sq     F  Pr(>F)
## 1     30 721
## 2     29 278  1       443 64.28 1.1e-07 ***
## 3     24 214  5        64  1.86   0.146
## 4     23 165  1        49  7.18   0.014 *
## 5     21 145  2        19  1.41   0.269
## 6     20 138  1         8  1.12   0.303
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## geom_smooth: colour = black
## stat_smooth: method = lm
## position_identity: (width = NULL, height = NULL)
```

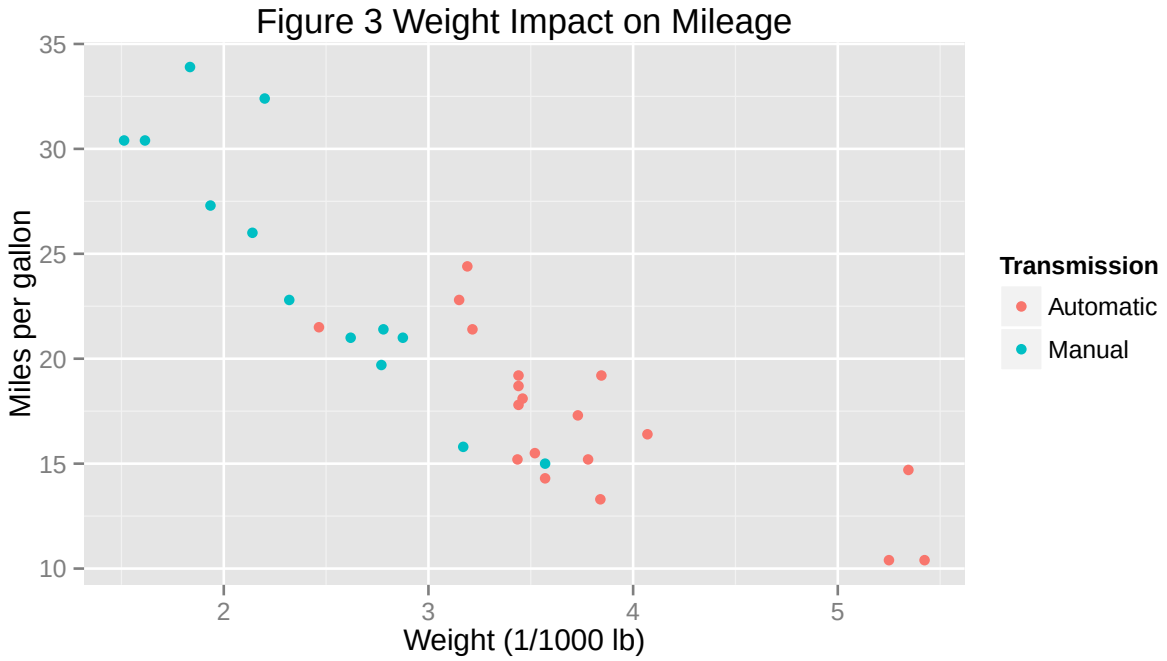
Figure 3 Weight Impact on Mileage
Miles per gallon
35
30
25
20
15
10
2
3
4
5
Weight (1/1000 lb)
Transmission
Automatic
Manual