Determination if Transmission Type statistically impacts MPG

Bill Martersteck

September 06, 2016

Executive Summary

This report explores the relationship between the vehicle transmission type (standard versus automatic) and the miles per gallon (mpg) outcome of the vehicles in the `mtcars` data set. Specifically, the following questions are answered:

1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions?

The analysis of the data shows that the vehicle transmission type is one of the significant variables defining the miles per gallon of the vehicles in this data set.

Summary of the `mtcars` Dataset

The `mtcars` data set defines 11 different automobile characteristics of 32 different cars. The variables for each car include:

- mpg - Miles Per Gallon
- cyl - Number of Engine Cylinders (4, 6, or 8)
- disp - Total Cylinder Displacement (cubic inches)
- hp - Gross Horsepower
- drat - Rear axle ratio
- wt - Weight (1000 lbs)
- qsec - 1/4 mile time (in seconds)
- vs - whether the cylinders are oriented in a "V" or a straight configuraion (0 = V, 1 = Straight)
- am - Transmission Type (0 = automatic, 1 = manual)
- gear - Number of forward gears
- carb - Number of carburetors

Inference Analysis

Figure 1 (in the appendix) shows the mpg of the various vehicles plotted by transmission type. The mpg for automatic transmission vehicles have mean of 17.15 while manual transmission vehicles have mean of 24.39.

Performimg a `t.test(mpg ~ am, data = mtcars)$p.value` results in a P-Value of 0.0013736. This value is below .025 telling us to reject the null hypothesis that the mean mpg for the automatic and the manual transmission vehicles are the same.

Lets look further at the regression model to determine how the transmission type impacts mileage.

Regression Analysis

The single variable model looking at mpg solely against transmission type gives a significant P-Value of 0.00029 but an adjusted R-Squared value of 0.3384589 indicating the model explains a very small percentage of the residual error. Given this, we must add additional variables to the model in an attempt to explain more of the residual errors.

To determine the other significant variables, model `mpg` as the outcome and the other 10 variables as the regressors. This model results in the following:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	12.3034	18.7179	0.66	0.5181
cyl	-0.1114	1.0450	-0.11	0.9161
disp	0.0133	0.0179	0.75	0.4635
hp	-0.0215	0.0218	-0.99	0.3350
drat	0.7871	1.6354	0.48	0.6353
wt	-3.7153	1.8944	-1.96	0.0633
qsec	0.8210	0.7308	1.12	0.2739
vs	0.3178	2.1045	0.15	0.8814
am	2.5202	2.0567	1.23	0.2340
gear	0.6554	1.4933	0.44	0.6652
carb	-0.1994	0.8288	-0.24	0.8122

This model gives an Adjusted R-Squared value of 0.8066423 which explains a significant amount of the error, but none of the variables are significant (their P-Values are all above .05), and so the variable with the largest P-Value is removed from the model (as it is least significant), and the model is rerun with the remaining variables. This process continues until all the remaining variables in the model are significant (P-Value < .05).

The variables drop out in the order: `cyl`, `vs`, `carb`, `gear`, `drat`, `disp`, and `hp`, leaving `wt`, `qsec`, and `am` being the significant variables driving mpg. This model produces the following:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.6178	6.9596	1.38	0.1779
wt	-3.9165	0.7112	-5.51	0.0000
qsec	1.2259	0.2887	4.25	0.0002
am	2.9358	1.4109	2.08	0.0467

This model has an Adjusted R-Squared value of 0.8335561 which is even better than the full model.

Diagnostics

Figure 2, in the appendix, has various residual diagnostic plots. The fitted values plot shows independance given there does not appear to be any pattern or significant outliers. The Normal Q-Q plot shows all the data on a line indicating nomality of the residuals. The randomly distributed Scale-Location plot shows constant variance. The Residuals vs. Leverage plot shows outliers. From these plots, we can say the model can be used to show the relationship of transmission type to mpg.

Results

The analysis of this data shows that a manual transmission vehicle will get better gas mileage over an automatic transmission vehicle, and, holding the other variables constant, the manual transmission vehicles ($am = 1$) see a 2.9358 mile per gallon improvement in gas mileage over the automatic transmission vehicles ($am = 0$).

Appendix

```
boxplot(mpg ~ am, data = mtcars, xlab="Transmission (0 = Automatic, 1 = Manual)", ylab="MPG")
```

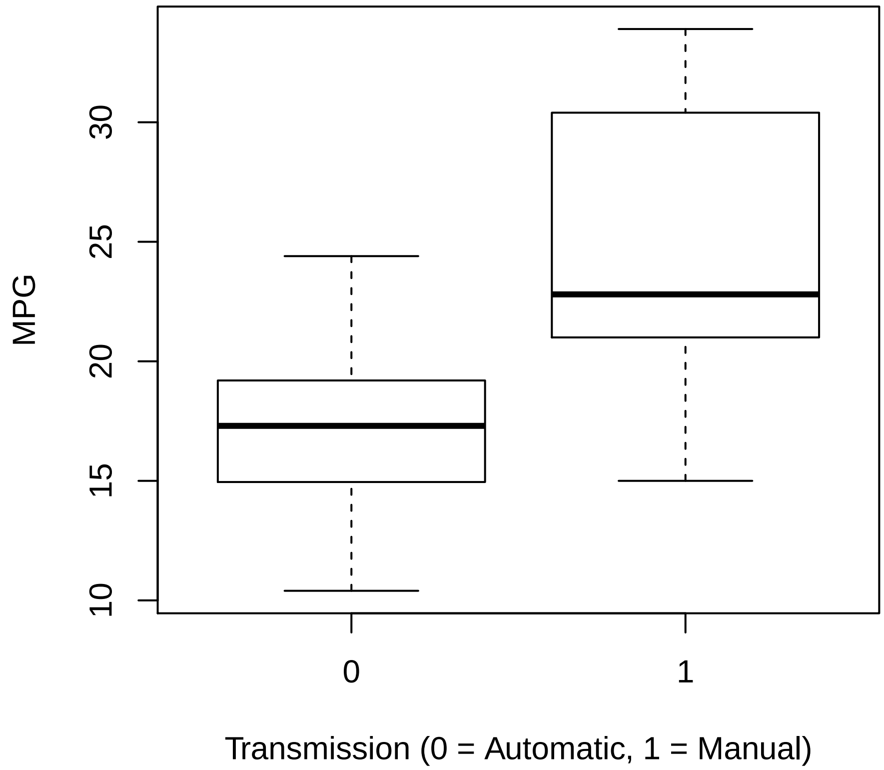

Figure 1: MPG divided by Transmission Type

```r
par(mfrow = c(2, 2))
fit <- lm(mpg ~ am + wt + qsec, data=mtcars)
plot(fit)
```

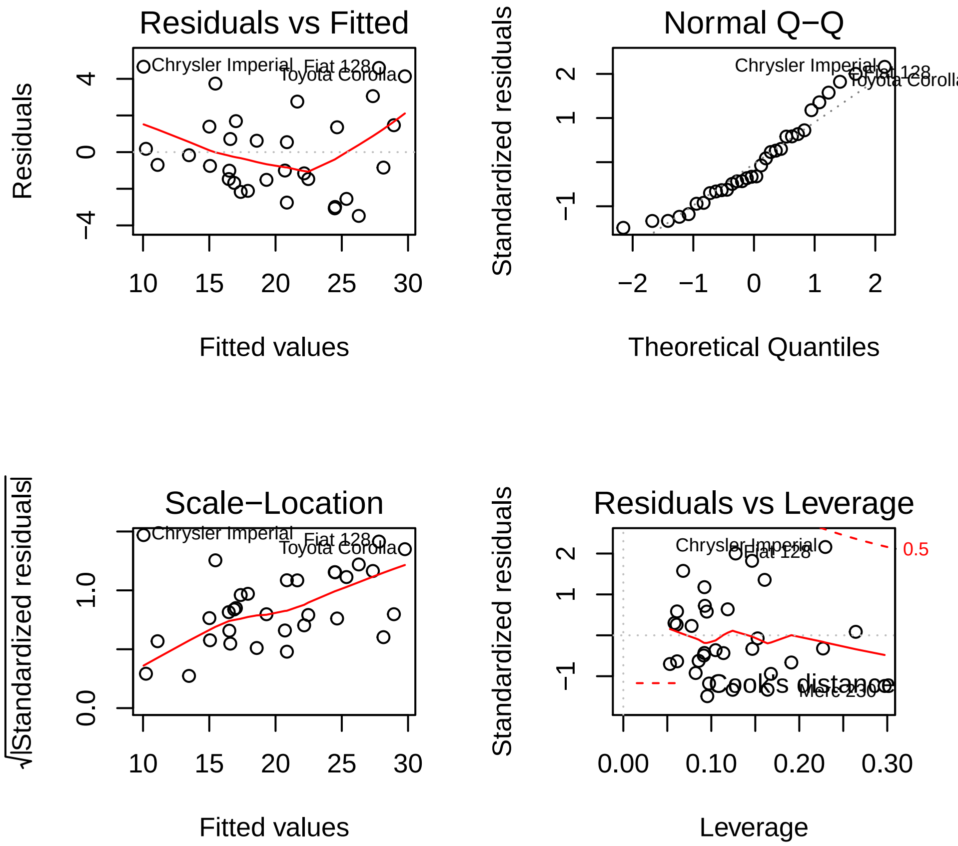

Figure 2: Residual Diagnostic Plots