

Motor Trend Transmission and MPG Analysis

Peggy Fan

September 20, 2014

Question 1. Is automatic or manual transmission better for mpg (miles per gallon)? To answer this question, I first conducted exploratory analyses to understand the nature of the data and the relationships between MPG and other variables. A quick glance of the mpg data by transmission type below shows us that automatic tends to have lower MPG than automatic. The data sample of automatic cars has smaller standard deviation and more cars than that of manual; in other words, data points of manual cars are fewer with higher variance.

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
mean	2	20.770	5.123	17.150	24.390
sd	2	5.000	1.649	3.834	6.167
N	2	16.000	4.243	13	19

I created plots for all variables by automatic and manual to see if there are any variables that have distinct patterns by transmission type, for which the model should control for to avoid omitted variable bias (see figure 1-3 in the appendix). It seems like there are distinct patterns of mpg by transmission type with the following variables: wt (weight), cyl (cylinder), and drat (rear axle ratio).

Model fitting I first fitted a base model with mpg as the outcome variable and transmission as the independent variable. Then I added each of the three variables mentioned above to the base models, resulting in three additional models. The coefficients on “amManual” measures the change in mpg of manual in comparison with automatic, since automatic equals to zero and is the reference group.

We see that in models 2 and 3, both “wt” and “cyl” are significant at $p < 0.001$ as controls and both models have a significantly higher R-squared from the base model (0.75 v. 0.36). As weight and cylinder are shown to be important variables in explaining the distribution of mpg, I test another model where it includes both weight and cylinder as controls (model 5), and it has an even higher R-squared (0.83).

To find out whether the models of interest, model 2, 3, and 5 are significantly different from the base model, I performed several likelihood-ratio tests. The tests show that all three models are significantly different from the base model at $p < 0.001$.

Residuals and diagnostics I created a residual plot and other diagnostic tests to assess the fit of the model of my choice (see figure 4). The “Residual vs. Fitted” plot is close to a flat line, which means there is no curvature in the data and a linear model is appropriate. The Q-Q plot shows that the residuals are normally distributed. The “Scale-Location” plot shows that the model’s variance is fairly constant because there is no strong up or down tilt. Lastly, the “Residual v. Leverage” plot shows that the data are not likely to suffer from high leverage or high residual data points. Overall, this fitted model seems to be robust.

Using the fitted model (model 5 in table 2), the answer to question 1 is: compared to an automatic car, a manual car are 0.18 higher in mpg after controlling for the weight and the number of cylinders of the car. But this difference is not statistically significant at the 95% confidence level. We could not reject the null hypothesis that automatic is different from manual, thus we cannot definitively say automatic or manual is better for mpg because other aspects of a car affects its mpg as well.

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	17.15*** (1.12)	37.32*** (3.05)	34.52*** (2.60)	-1.95 (7.07)	39.42*** (2.64)
amManual	7.24*** (1.76)	-0.02 (1.55)	2.57 (1.29)	2.81 (2.28)	0.18 (1.30)
wt		-5.35*** (0.79)			-3.13** (0.91)
cyl			-2.50*** (0.36)		-1.51** (0.42)
drat				5.81* (2.13)	
R ²	0.36	0.75	0.76	0.49	0.83
Adj. R ²	0.34	0.74	0.74	0.46	0.81
Num. obs.	32	32	32	32	32

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 2: Regression models

	Model2	Model3	Model5
Model1	0.000	0.000	0.000

Table 3: Likelihood-Ratio Tests

Question 2: What is the MPG difference between automatic and manual transmissions? To quantify the difference between automatic and manual, I conducted an independent t-test, which applies to samples with unequal variances.

	Mean of Automatic	Mean of Manual	P-value	Lower interval	Upper interval
estimate	17.15	24.39	0.00	-11.28	-3.21

T-test shows that the difference in the mean mpg of automatic and manual is significant at 95% confidence level. Given this dataset, the mean mpg for automatic is lower than that of manual. The confidence intervals of the difference in the mean comparing automatic to manual are -11.28 and -3.2.

Appendices:

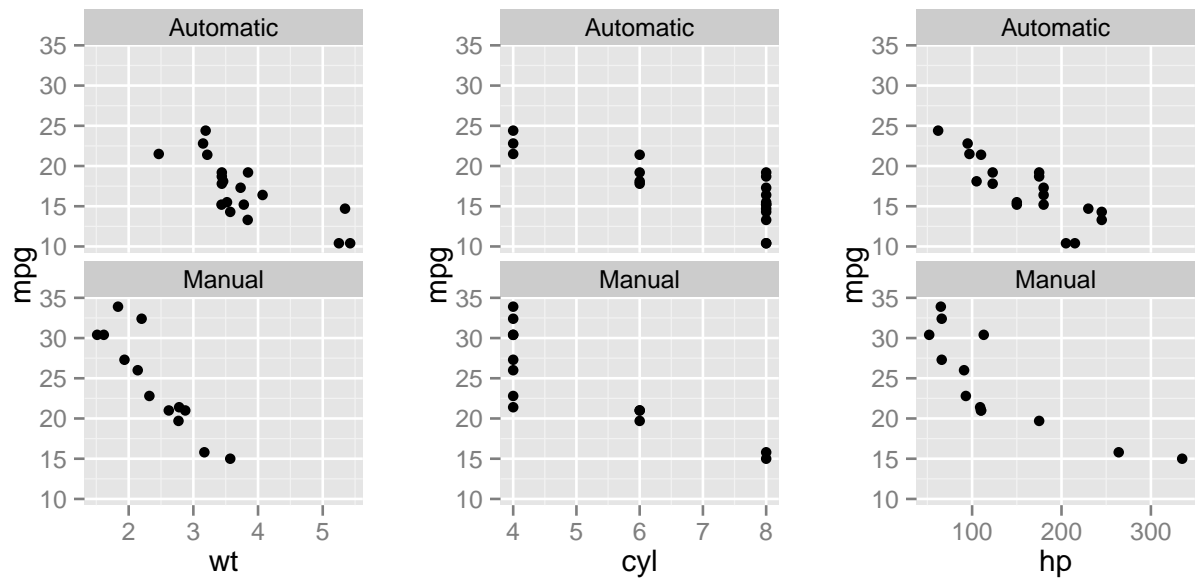


Figure 1.

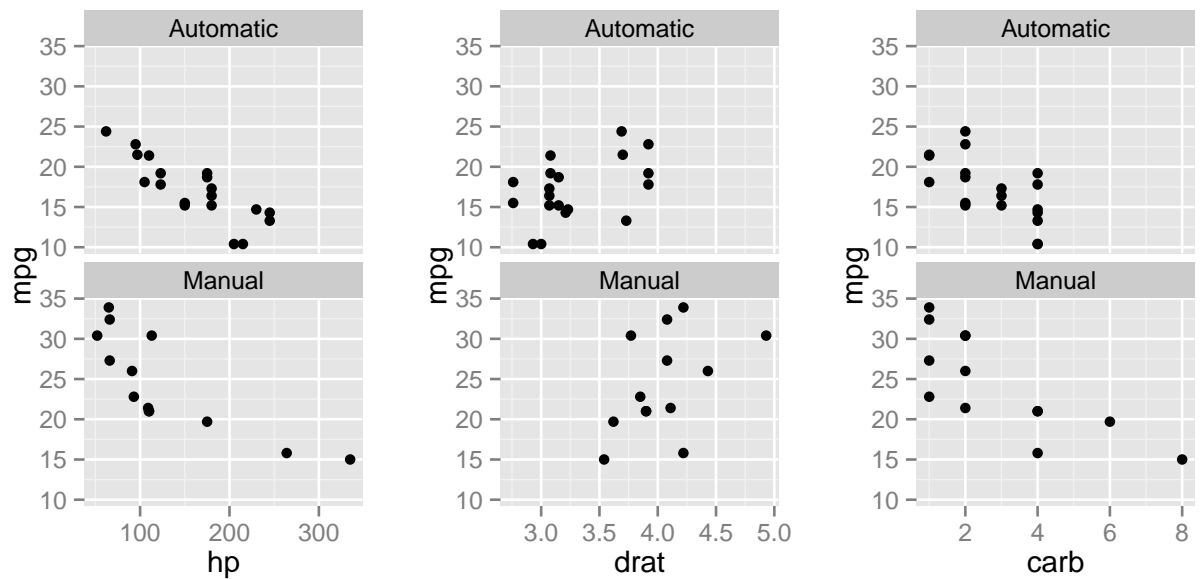


Figure 2.

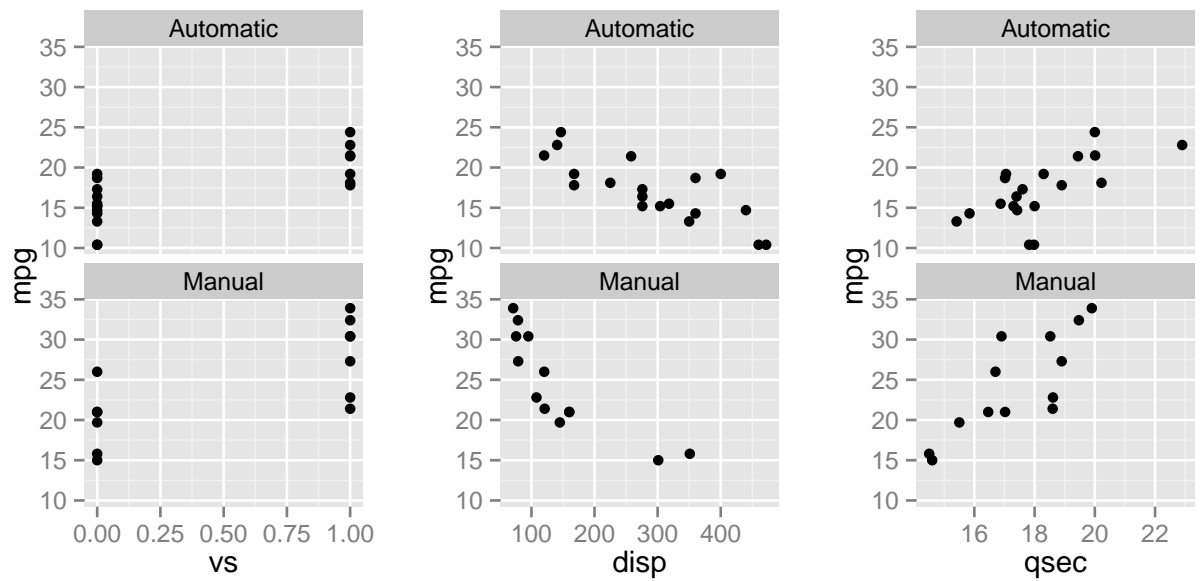


Figure 3.

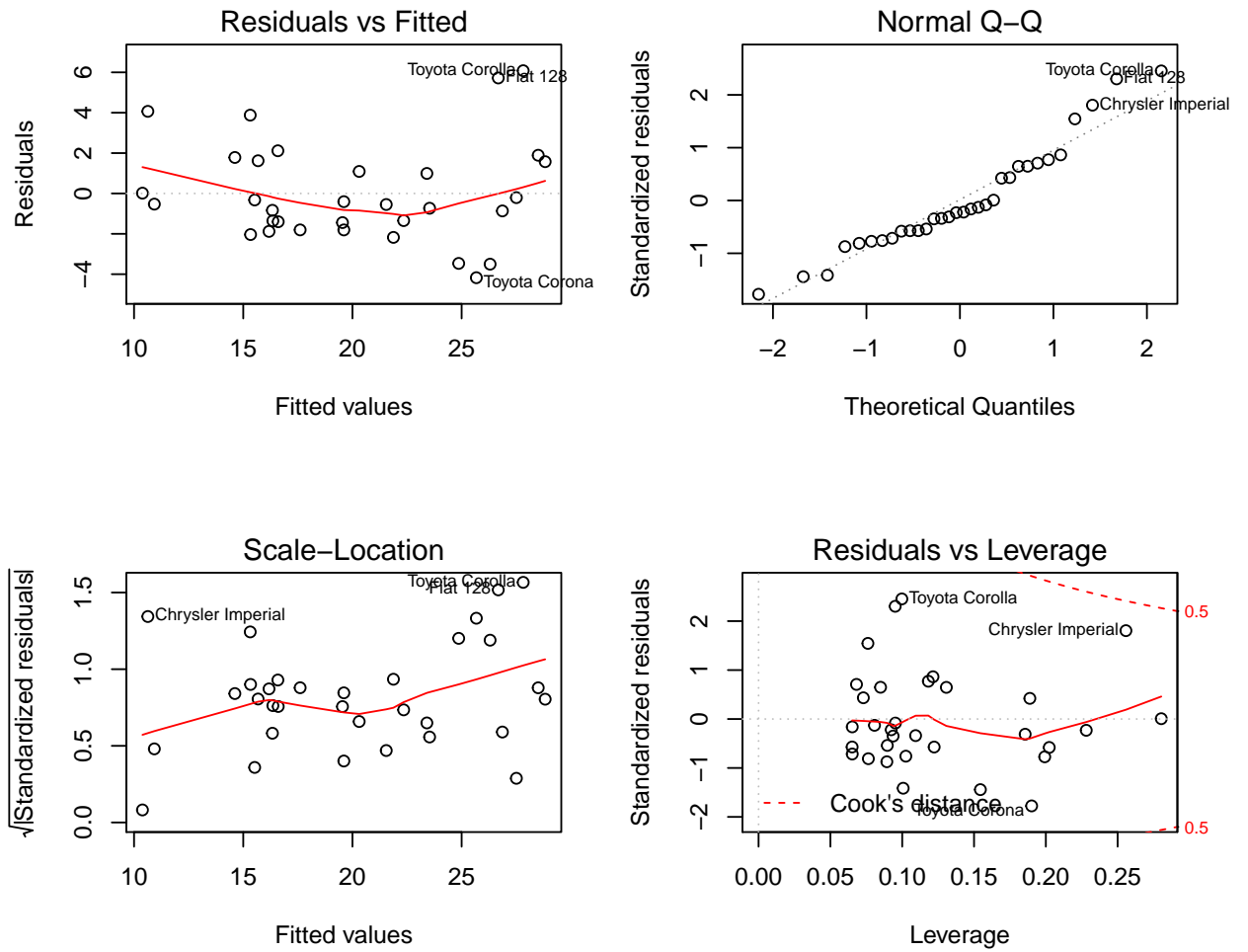


Figure.4