Regression Models Course Project

Anit Bhandari

August 23, 2016

Executive Summary

This analysis seeks to determine if the automatic or manual transmission is better for MPG and to quantify the difference. Using an ordinary least squares regression model on the mtcars dataset we can conclude:

- After adjusting for vehicle weight and quarter mile time, the manual transmissions provides a statistically significant 2.9 MPG advantage over the automatic transmission.

About the Data[1]

This analysis was preformed using data extracted from the 1974 Motor Trend US magazine. It comprises fuel consumption for 32 1973–74 model automobiles and ten aspects of automobile design and performance:

- Number of cylinders
- Displacement (cu.in.)
- Gross horsepower
- Rear axle ratio
- Weight (lb/1000)
- 1/4 mile time
- V/S
- Transmission
- Number of forward gears
- Number of carburetors

Data Processing

The only adjustment made to the data was relabeling the transmission type variable (am) to make it easier to interpret throughout the analysis.

```
data(mtcars)
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
```

Exploratory Data Analysis

Mean MPG for this dataset is **20.09** with a standard deviation of **6.03**. There are 19 cars with an automatic transmission and 13 cars with a manual transmission.
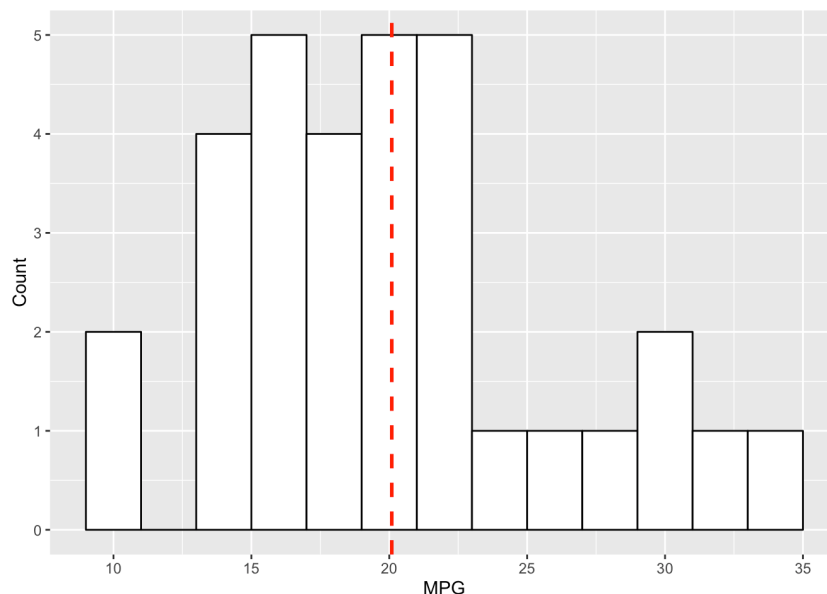

MPG histogram

Regression Analysis

Single Variable Analysis

In order to tease out the effect of the type of transmission we first modeled MPG as a function of the type of transmission. Our null hypothesis is that there is no difference between the two types and the alternate is that there is a difference. In this model, the intercept represents cars with an automatic transmission:

```
fit <- lm(mpg ~ am, data = mtcars)
```
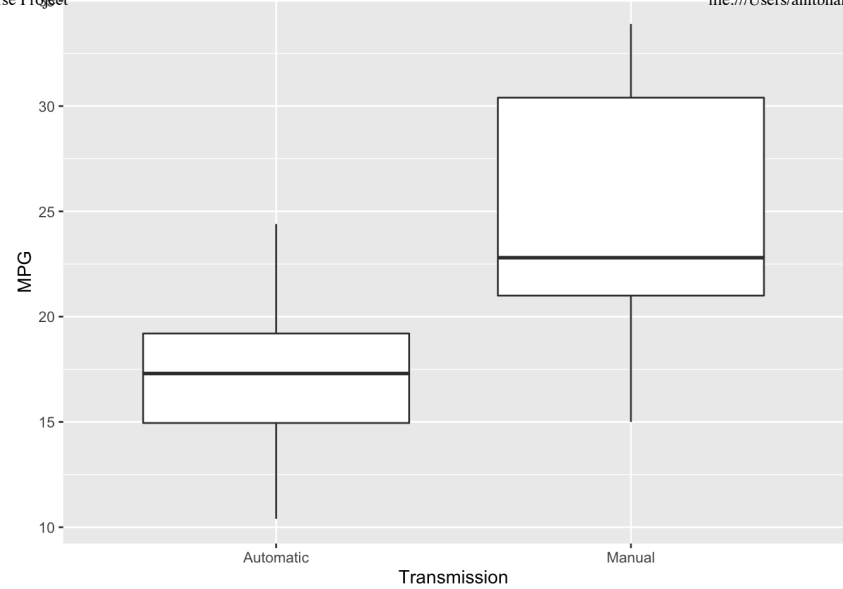
MPG by type of transmission

Fitting linear model: mpg ~ am			
	Estimate	Std. Error	t value	Pr(>|t|)
amManual	7.245	1.764	4.106	0.000285
(Intercept)	17.15	1.125	15.25	1.134e-15

With a P-value of about 0.0003 we reject the null hypothsis. So at first blush it would appear that there is a statistically significant difference and that cars with a manual transmission have 7.2 MPG advantage than those with automatic transmissions.

This model has an adjusted R squared of about 0.34 which suggests there are other factors we are not taking into consideration. For example the figure 3 illustrates that the vehicle weight has a negative relationship with fuel efficiency. What happens to the transmission effect when we control for other variables?
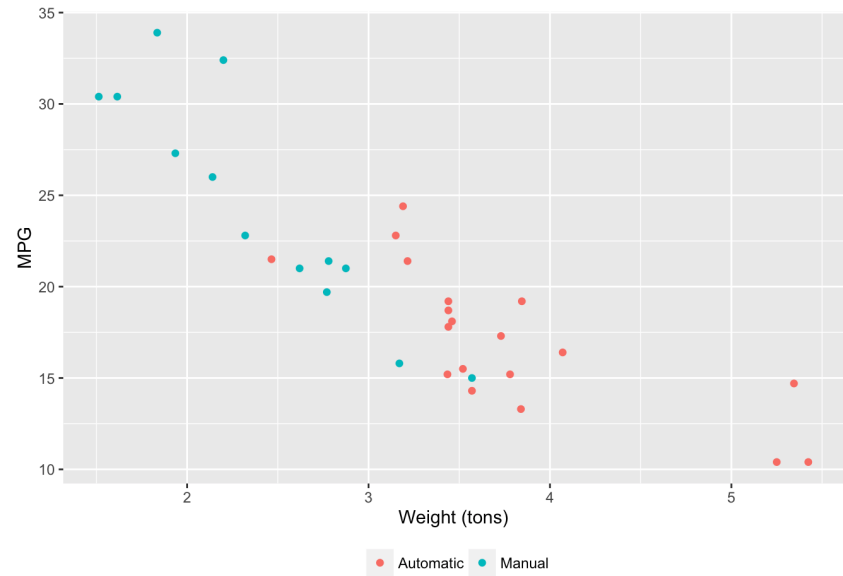
MPG by Vehicle Weight

Multi-Variable Analysis

R's step function was employed to determine which variables should be included in the model. This function uses a stepwise algorithm to select variables based on the Akaike information criterion (AIC).

```
best.fit <- step(lm(mpg ~ ., data=mtcars), direction=c("both"), trace=0)
```

| Fitting linear model: mpg ~ wt + qsec + am			
	Estimate	Std. Error	t value	Pr(>|t|)
wt	-3.917	0.7112	-5.507	6.953e-06
qsec	1.226	0.2887	4.247	0.0002162
amManual	2.936	1.411	2.081	0.04672
(Intercept)	9.618	6.96	1.382	0.1779

Once again our null hypothesis is that there is no difference between the two types of transmission and the alternate is that there is a difference. With a P-value of about 0.047 we would reject the null hypothsis. This model has an adjusted R squared of about 0.83 which is a significant improvement relative to the single variable model[2]. The manual transmission gives about a 2.9 miles per gallon gain in fuel economy relative to the automatic transmission.
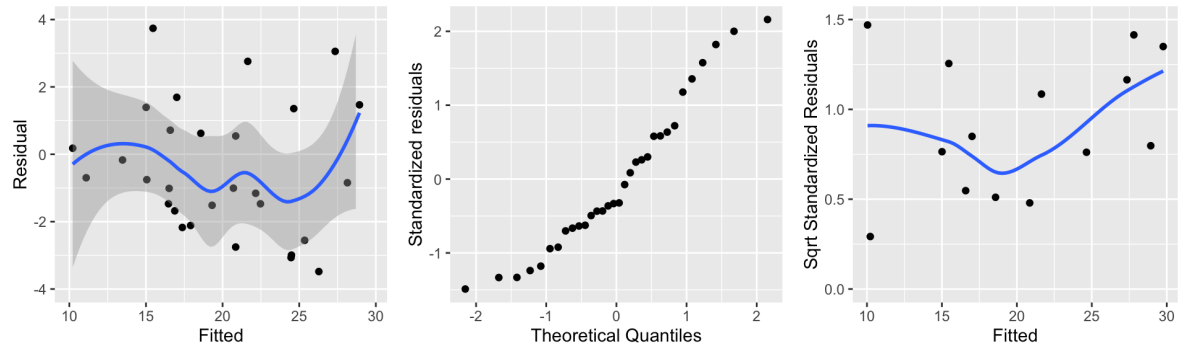
Appendix 1: Sample Data

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21	6	160	110	3.9	2.62	16.46	0	Manual	4	4
Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02	0	Manual	4	4
Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	Manual	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	Automatic	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.44	17.02	0	Automatic	3	2
Valiant	18.1	6	225	105	2.76	3.46	20.22	1	Automatic	3	1
Duster 360	14.3	8	360	245	3.21	3.57	15.84	0	Automatic	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.19	20	1	Automatic	4	2
Merc 230	22.8	4	140.8	95	3.92	3.15	22.9	1	Automatic	4	2
Merc 280	19.2	6	167.6	123	3.92	3.44	18.3	1	Automatic	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.44	18.9	1	Automatic	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.07	17.4	0	Automatic	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.73	17.6	0	Automatic	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.78	18	0	Automatic	3	3
Cadillac Fleetwood	10.4	8	472	205	2.93	5.25	17.98	0	Automatic	3	4
Lincoln Continental	10.4	8	460	215	3	5.424	17.82	0	Automatic	3	4
Chrysler Imperial	14.7	8	440	230	3.23	5.345	17.42	0	Automatic	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.2	19.47	1	Manual	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	Manual	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.9	1	Manual	4	1
Toyota Corona	21.5	4	120.1	97	3.7	2.465	20.01	1	Automatic	3	1
Dodge Challenger	15.5	8	318	150	2.76	3.52	16.87	0	Automatic	3	2
AMC Javelin	15.2	8	304	150	3.15	3.435	17.3	0	Automatic	3	2
Camaro Z28	13.3	8	350	245	3.73	3.84	15.41	0	Automatic	3	4
Pontiac Firebird	19.2	8	400	175	3.08	3.845	17.05	0	Automatic	3	2
Fiat X1-9	27.3	4	79	66	4.08	1.935	18.9	1	Manual	4	1
Porsche 914-2	26	4	120.3	91	4.43	2.14	16.7	0	Manual	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.9	1	Manual	5	2
Ford Pantera L	15.8	8	351	264	4.22	3.17	14.5	0	Manual	5	4

Appendix 2: Residual Plots

Single Variable Model
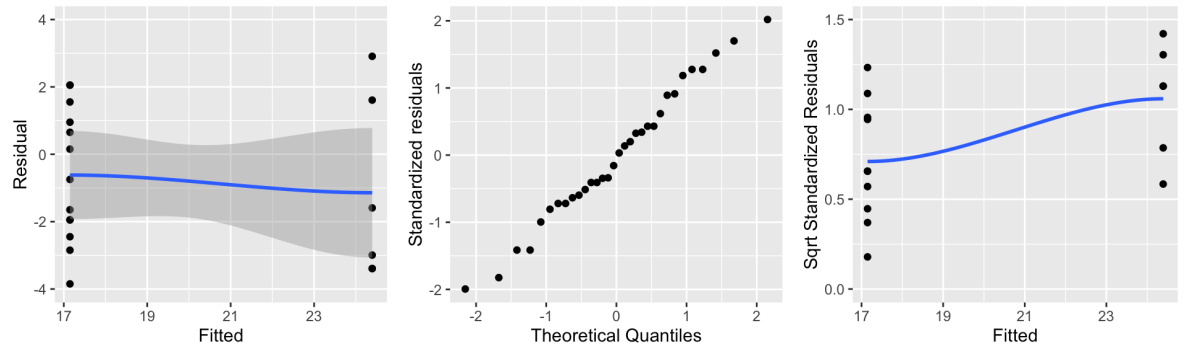
Multi-Variable Model

1. A sample of the data is found in Appendix 1↵

2. See Appendix 2 for a residual plots of the models↵