Motor Trends : Automatic or Manual transmission?

Yao-Jen Kuo

Sunday, October 25, 2015

1 Executive Summary

You work for *Motor Trend*, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). 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

In short, the purpose of this analysis is to find if there is any relationship between a set of variables and MPG.

2 Data Management

Factor some numeric variables at the beginning.

```
data(mtcars)
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$am <- factor(mtcars$am)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

3 Exploratory Analysis

In *Figure 1* the boxplot suggests a significant difference on fuel consumption between automatic and manual transmission cars. So can we use only transmission type to predict MPG as an univariate linear model?

```
fitUnivariate <- lm(mpg ~ am, data=mtcars)
round(summary(fitUnivariate)$r.squared*100, digits=2)
```

```
## [1] 35.98
```

The model explains only **35.98%** of the MPG variation. In *Figure 2* the pairs graph shows that several variables have high correlation with MPG.

4 Regression Analysis

Therefore, let us build an initial model using all variables and then using stepwise selection for the best predictors.

```r
fitMultivariate <- lm(mpg ~ ., data=mtcars)
fitStepwise <- step(fitMultivariate, direction="both", trace=0)
round(summary(fitStepwise)$r.squared*100, digits=2)#R-squared
```

```
## [1] 86.59
```

```r
summary(fitStepwise)$coef[6,"Pr(>|t|)"]#alpha
```

```
## [1] 0.2064597
```

The final model contains 4 predictors including cyl(number of cylinders), hp(horsepower), weight(weight) and am(transmission type). This model explains **86.59%** of the MPG variation. The transmission has no effect on the fuel consumption since α**=0.2064597>0.05**.

4.1 Residual Analysis and Diagnostics

According to *Figure 3* we can verify the following underlying assumptions:

1. The Residuals vs Fitted plot shows no consistent pattern.
2. The Normal Q-Q plot indicates that the residuals are normally.
3. The Scale-Location plot confirms the constant variance assumption.
4. The Residuals vs Leverage confirms that no outliers are present.

5 Conclusion

```r
c(abs(fitStepwise$coef[2]), abs(fitStepwise$coef[3]),
  abs(fitStepwise$coef[4]), abs(fitStepwise$coef[5]))
```

```
##       cyl6       cyl8         hp         wt
## 3.03134449 2.16367532 0.03210943 2.49682942
```

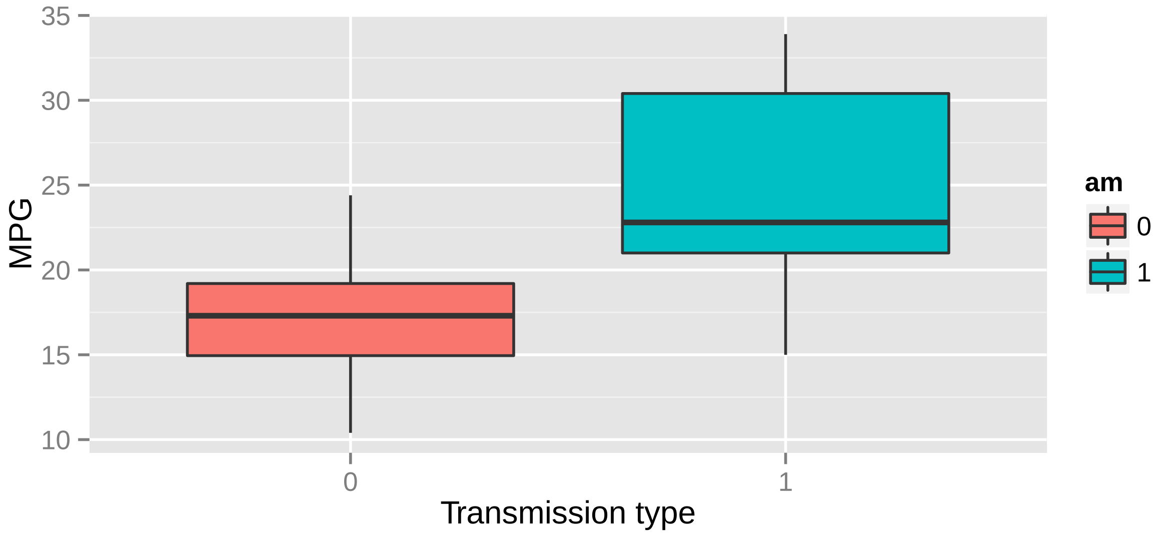
1. If a car has 6 or 8 cylinders, everything else same, the MPG decreases by **3.031344** and **2.163675**, respectively.
2. One unit of increase on gross horsepower, everything else same, results a decrease of **0.03210943** on MPG.
3. 1,000 lb increase on the weight of a car, everything else same, results a decrease of **2.496829** on MPG.
4. This analysis was not able to find significant link between the transmission type and fuel consumption. It is less likely that driving a manual car will save you any money on gasoline.

6 Appendix

6.1 Figure 1

Boxplot of MPG vs. Transmission

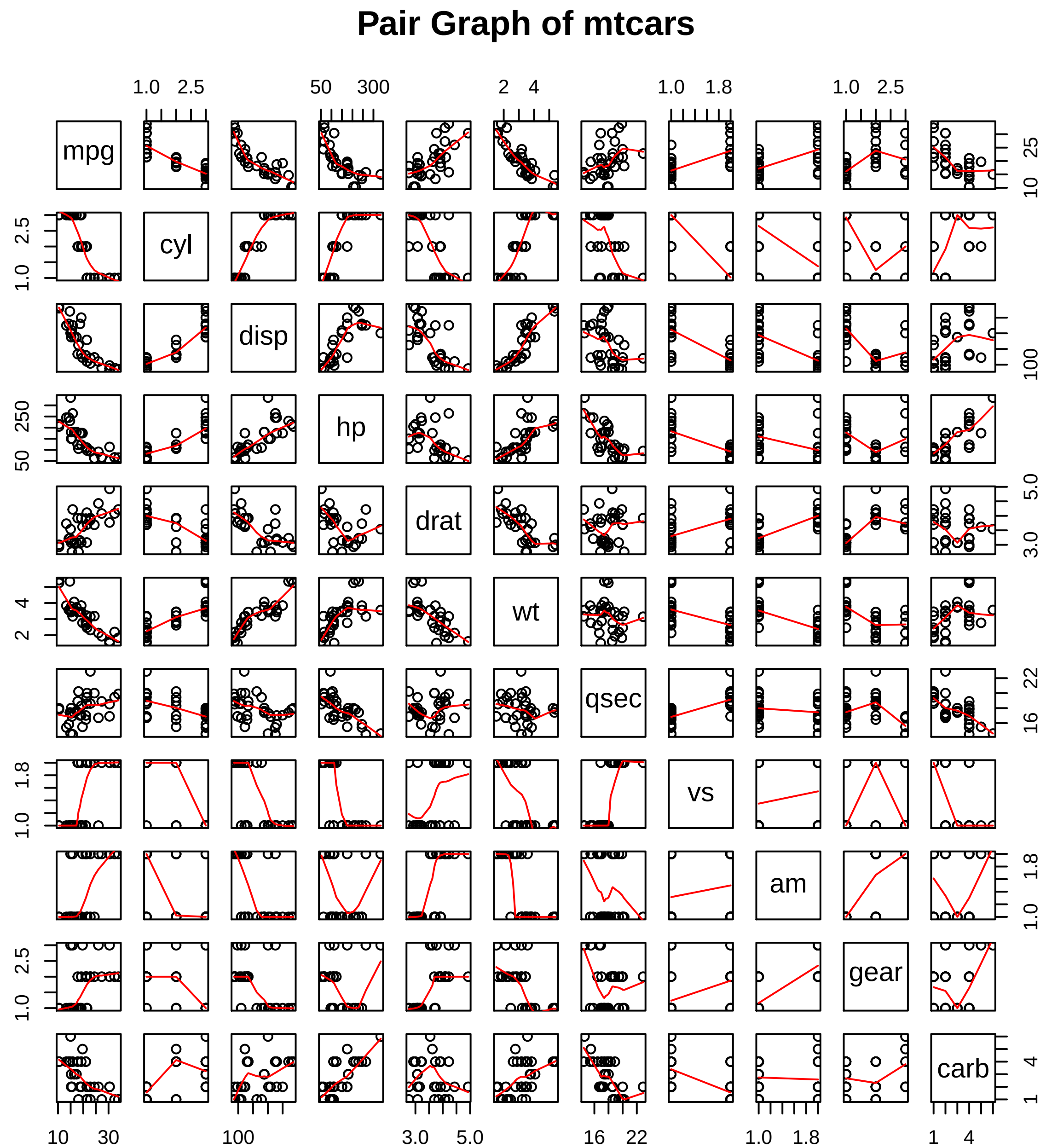
```r
ggplot(mtcars, aes(x=am, y=mpg, fill=am)) +
  geom_boxplot() +
  xlab("Transmission type") +
  ylab("MPG")
```



6.2 Figure 2

Pairs Graph of `mtcars`

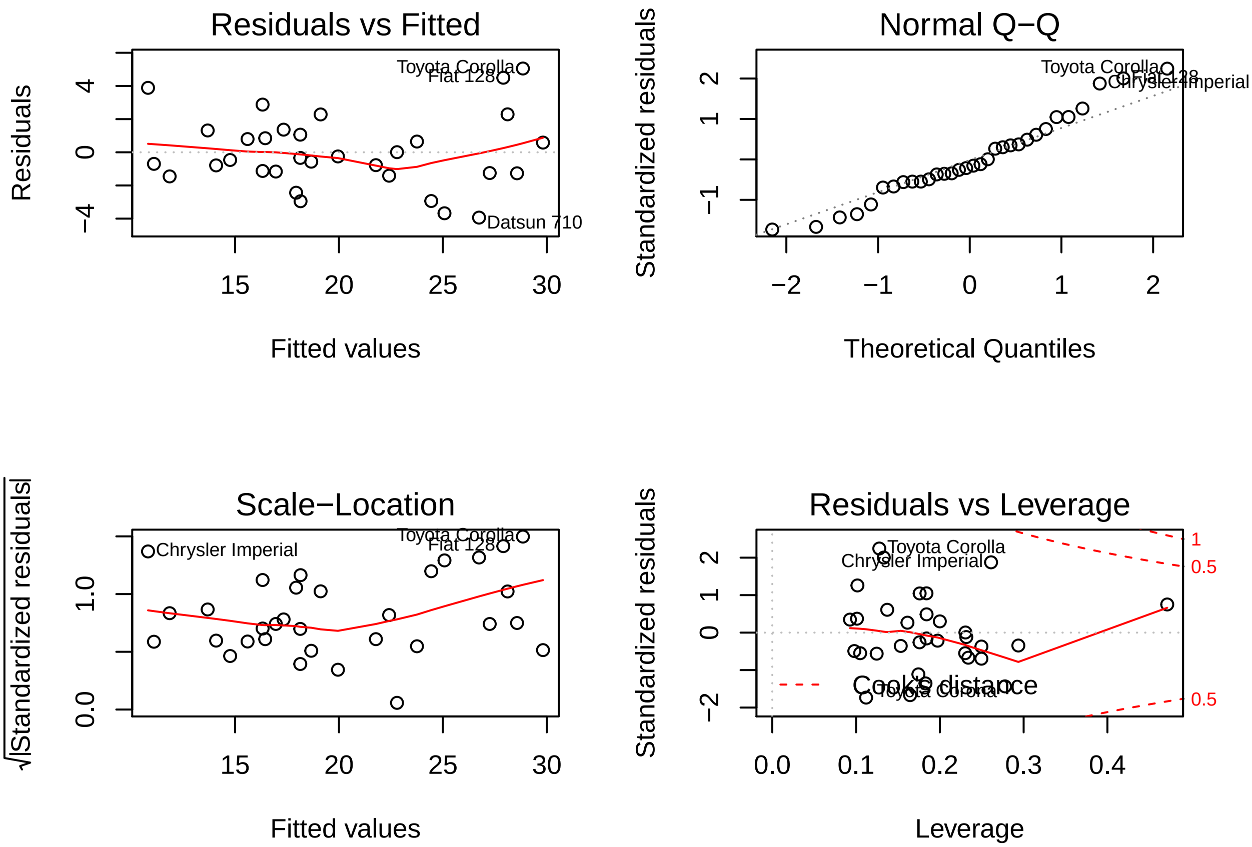
```r
pairs(mtcars, panel=panel.smooth, main="Pair Graph of mtcars")
```



6.3 Figure 3

Residual Analysis and Diagnostics

```
par(mfrow = c(2,2))
plot(fitStepwise)
```


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Datsun 710
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Fiat 128
Chrysler Imperial
Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals vs Leverage
Standardized residuals
Leverage
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's distance
1
0.5
0.5