

# Car Transmission Analysis

*Phat Doan*

*Sunday, May 24, 2015*

## Executive Summary

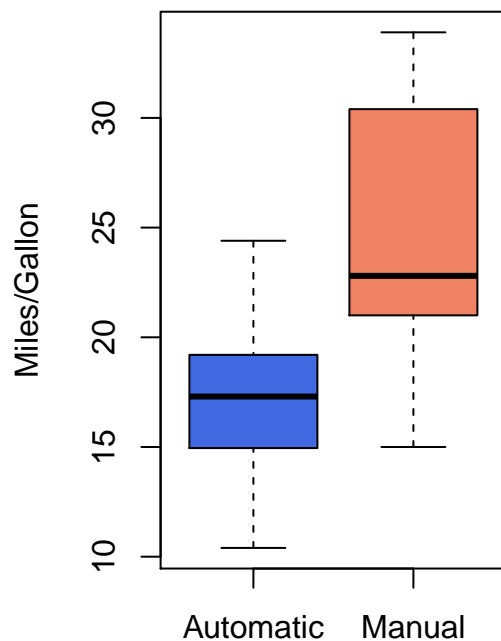
This analysis was performed on Motor Trend Cars data to determine whether there is a significant difference of miles/gallon (mpg) between automatic and manual transmission cars.

We found that transmission may play a role in car milage. Manual transmission cars are 2.936 mpg higher on average than the automatic cars, all else being equal.

## Exploratory Data Analysis

Box-plotting the mpg for each transmission type reveals that there seems to be a difference in the distribution of *mpg* for each type of *transmission*. The correlation matrix (see Appendix - Fig 1) shows that there is a relationship of **+0.6** between *transmission* and *mpg*, meaning manual cars have higher mpg than automatic ones.

```
par(mar = c(2, 4, 1, 1))
boxplot(mpg ~ am, data = mtcars, xlab = "Transmission Type",
        ylab = "Miles/Gallon", col = c("royalblue", "salmon2"),
        names = c("Automatic", "Manual"))
```



## Regression Models

A model with all the available variables is built and we then examine the p-values of the regressors:

```
fit_all <- lm(mpg ~ ., data = mtcars)
summary(fit_all)$coef
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	12.30337416	18.71788443	0.6573058	0.51812440
## cyl	-0.11144048	1.04502336	-0.1066392	0.91608738
## disp	0.01333524	0.01785750	0.7467585	0.46348865
## hp	-0.02148212	0.02176858	-0.9868407	0.33495531
## drat	0.78711097	1.63537307	0.4813036	0.63527790
## wt	-3.71530393	1.89441430	-1.9611887	0.06325215
## qsec	0.82104075	0.73084480	1.1234133	0.27394127
## vs	0.31776281	2.10450861	0.1509915	0.88142347
## am	2.52022689	2.05665055	1.2254035	0.23398971
## gear	0.65541302	1.49325996	0.4389142	0.66520643
## carb	-0.19941925	0.82875250	-0.2406258	0.81217871

Because no variable is significant at 95% confidence, we deploy **backward elimination** - removing insignificant variables until only significant variables are left in the model.

```
step(fit_all, direction = "backward", test = "F")
```

The resulted model is  $\text{mpg} \sim \text{wt} + \text{qsec} + \text{am}$

```
fit <- lm(formula = mpg ~ wt + qsec + am, data = mtcars)
summary(fit)$coef
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	9.617781	6.9595930	1.381946	1.779152e-01
## wt	-3.916504	0.7112016	-5.506882	6.952711e-06
## qsec	1.225886	0.2886696	4.246676	2.161737e-04
## am	2.935837	1.4109045	2.080819	4.671551e-02

It can be inferred from the above table that a) Transmission type is significant in explaining mpg and b) Manual transmission cars are 2.936 mpg higher on average than the automatic cars, all else being equal.

## Model Residual Plot & Diagnosis

A residual plot and other diagnosis plots are done. In the residual plot (Appendix - Fig 2), the residuals are not randomly scattered but follow a curve, suggesting that there might be important confounding variables that were not taken into account. The Q-Q plot show that data is somewhat normal.

## Appendix

Fig 1 - Correlation Matrix

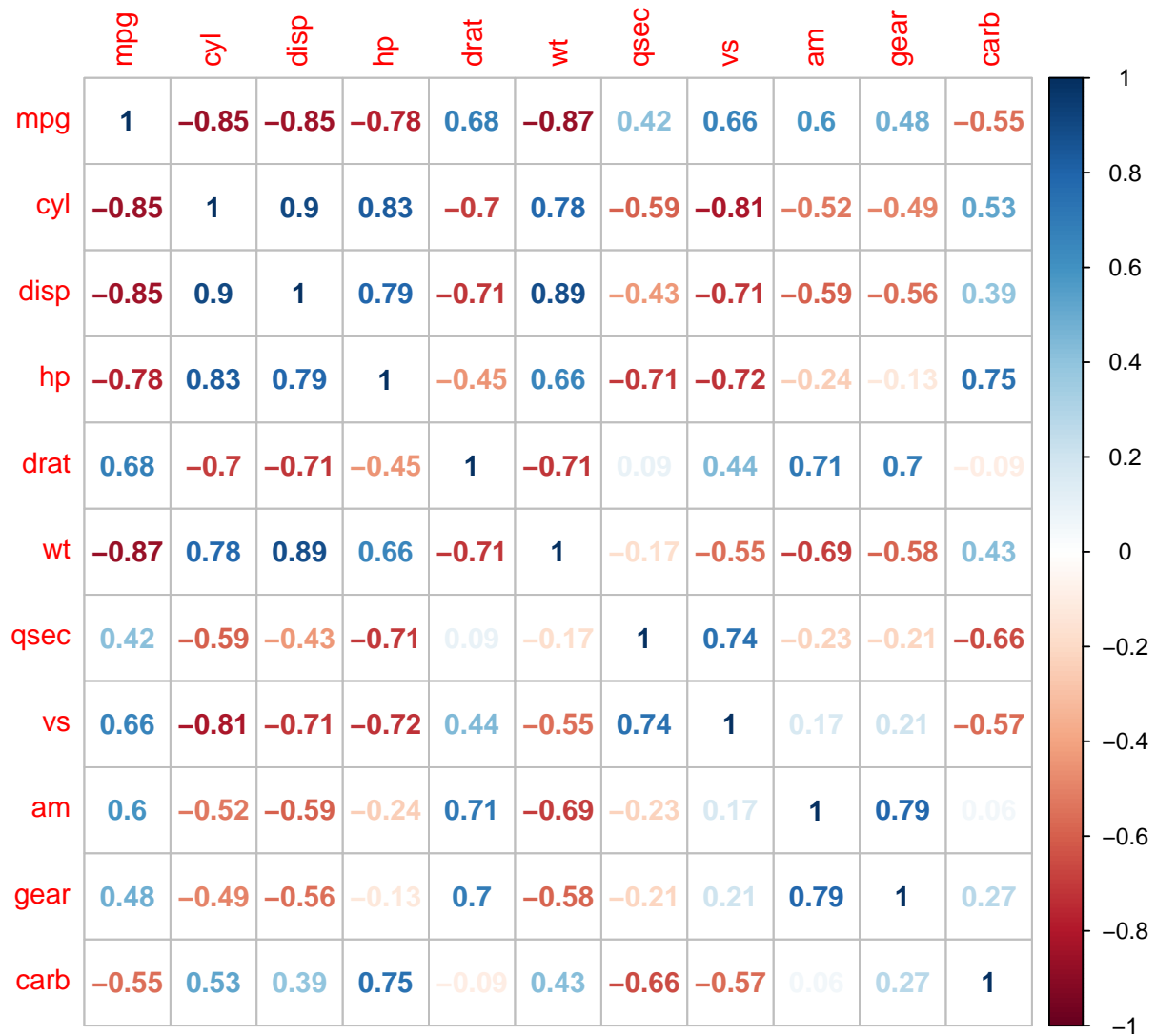


Fig 2 - Diagnosis Plots

