MPG difference: Automatic vs. Manual Transmissions

Ariel Lev, 24. May 2015

Executive Summary

With a confidence level of 95%, vehicles equipped with a manual transmission drive between 0.05 to 5.83 more miles per gallon then vehicle which are equipped with automatic transmission.

Exploratory data analysis

The data, extracted from 1974 Motor Trend magazine, comprise gasoline mileage in miles per gallon (MPG), and 10 aspects of automotive design and performance for 32 vehicles.

Table 1.|First three entries out of Motor Trend sample

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1

This particular sample of 32 vehicles has a bias to non-U.S. OEMs: it includes 7 Mercedes, a Porsche, a Ferrari and a Maserati. Therefore we might not expect a universal prediction model to emerge. A t-test comparing the MPG means of automatic vs. manual transmission suggests a significant difference in sample means, with p-value of **2.85e-04** and **95% confidence interval (3.64, 10.85)**

Regression Anaylsis

Obviously in order to study the response MPG depending on transmission type, any linear model must include AM as a term.

```
fit0 <- lm(mpg ~ factor(am), data = mtcars)
```

Unfortunately AM can be accounted only for little of the variance of MPG. Our basic model reports a R square of only 35.98%. In addition, the partial correlation of MPG and AM, controlled by WT, yields hardly **any** correlation between the two.

```
pcor(c("mpg", "am", "wt"), var(mtcars))
```

```
## [1] -0.002837144
```

Following code shows the 7 strongest most correlated features with MPG, given in absoulte values. Correlation of MPG against each one of the features suggests WT as the best single predictor. Physically it makes sense, while mileage can be considered as work applied to a vehicle's body.

```
sort(abs(cor(mtcars)[1, -c(1)]), decreasing = T)[1:7]
```

```
##        wt       cyl      disp        hp      drat        vs        am
## 0.8676594 0.8521620 0.8475514 0.7761684 0.6811719 0.6640389 0.5998324
```

```
# shifting wt by its mean in order to yield meaningful intercept values
fit1 <- update(fit0, . ~ . + I(wt - mean(wt)), data = mtcars)
```

Including WT as a term results not only in a far larger R square (75.28%), but also in a supportive variance analysis showing a 46.12 F-Ratio with a very small p-value suggesting a significant improvement to the prediction force of our model.

```
anv <- anova(fit0,fit1)
data.frame( F = anv$"F"[2], "p-value" = anv$"Pr(>F)"[2])
```

```
##          F      p.value
## 1 46.11506 1.867415e-07
```

A fit of MPG may be subtle to inaccuracy, when ignoring an (at least intuitively) important aspect such as the engine fuel consumption. For a 3rd regressor i was considering either engine displacement or 1/4 mile time. Using the above methods I found out that 1/4 mile time is a slightly better contributor than displacement. When adding a 4th regressor, the procedure loses its strength, and it is hard to improve the model any further.

A residual analysis reveals a mean of 1.71e-16 and p-value of 0.0804277 in Shapiro-Wilk Normality Test, indicating that the residuals are normally distributed. Figure 2 depicts the residuals.

```
# shifting qsec by its mean in order to yield meaningfull intercept values
fit2 <- update(fit1, . ~ . + I(qsec - mean(qsec)), data = mtcars)
summary(fit2)$coefficients
```

```
##                       Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)          18.897941  0.7193542 26.270704 2.855851e-21
## factor(am)1           2.935837  1.4109045  2.080819 4.671551e-02
## I(wt - mean(wt))     -3.916504  0.7112016 -5.506882 6.952711e-06
## I(qsec - mean(qsec))  1.225886  0.2886696  4.246676 2.161737e-04
```

```
# extracting confidence intervals
sumCoef <- summary(fit2)$coefficients
ci <- data.frame(low = c(), high = c())
for (i in 1:4) {
    ci <- rbind(ci, round(c(sumCoef[i, 1], sumCoef[i, 1] + c(-1, 1) * qt(0.975,
        fit2$df) * sumCoef[i, 2]), 2))
}
names(ci) <- c("coeff", "low", "high")
row.names(ci) <- row.names(sumCoef)
ci
```

```
##                      coeff   low  high
## (Intercept)          18.90 17.42 20.37
## factor(am)1           2.94  0.05  5.83
## I(wt - mean(wt))     -3.92 -5.37 -2.46
## I(qsec - mean(qsec))  1.23  0.63  1.82
```

Summary

The p-values imply that all coefficients are significant. An analysis of the above coefficients shows with a 95% confidence that:

+ A vehicle equipped with an automatic transmission, weights 3220lb and makes a 1/4 of mile in 17.8 secs, drives between 17.42 to 20.37 miles per gallon
+ A vehicle equipped with a manual transmission drives between 0.05 to 5.83 more miles per gallons in comparison to a vehicle with an automatic transmission

In addition:

+ Between 5.37 to 2.46 decrease in MPG per 1000lb increase in weight
+ Between 0.63 to 1.82 decrease in MPG per each additional second in 1/4 mile time

Appendix

Figure 1.|Aqua tiles indicate a stronger correlation.

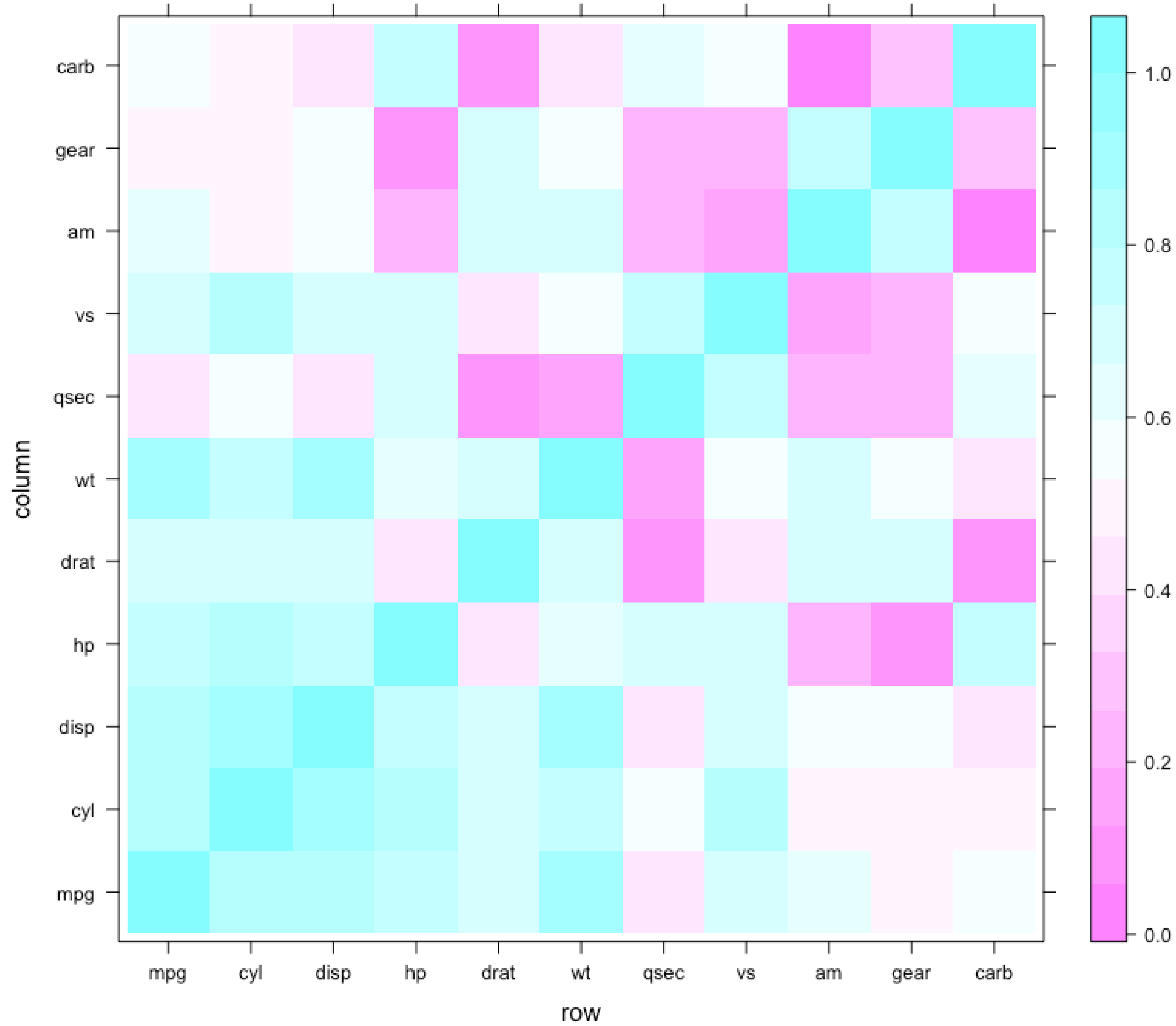

Figure 2. | Residuals plot showing homoscedasticity and normality.