MTCars Analysis

Dennis Reilly

Wednesday, July 25, 2014

Summary

In this analysis we will seek to answer two questions:

"Is an automatic or manual transmission better for MPG"

"Quantify the MPG difference between automatic and manual transmissions"

The conclusion based on only linear regression is that manual transmissions have slightly better (~3 mpg) fuel economy than automatic transmissions. This analysis was based on a fairly small sample set and additional methods may want to be considered.

Exploration of the data

The first step in this analysis will be to take a look at the data. In figure 1 of the appendix is a boxplot of the mpg variable and the am variable.

This initial plot shows that cars with manual transmission have a higher overall MPG than do cars with automatic transmission. The mean, median, and range for the manual transmisson cars are higher than that of the automatic transmission. There are however other variables that are correlated with MPG (see figure 2) so we will need to take a look at those as well to see if those variables are also affecting MPG.

It may also be true that Automatic transmission is associated with heavier cars with big engines, this would mean that the difference in MPG that we are seeing is not caused by the transmission type but rather by those other factors.

Model Selection

The next step is to run a regression against all available variables to see if any others are influencing mpg.

```
fitall=lm(mpg~., data = mtcars)
```

We will want to eliminate any variables that are not significant predictors of mpg. These additional variables may "overfit" the test data but will most likely create a model that does not perform as well on new data.

In looking at the regression summary (appendix figure 3) it is evident that some of these variables are not significant. Upon initial inspection it looks like only one variable is significant at all (wt). But this could be due to additional "noise" introduced by the insignificant variables (especially with some of the high correlations that we saw in figure 2). To get a better idea of significance, the number of variables will be reduced. Using reverse stepwise model selection we can start eliminating the most insignificant variables until we are left with a better model.

```
step(fitall, direction="backward")
```

This leaves us with the following model:

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

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    9.618     6.9596   1.382 1.779e-01
## wt            -3.917     0.7112  -5.507 6.953e-06
## qsec           1.226     0.2887   4.247 2.162e-04
## am             2.936     1.4109   2.081 4.672e-02
```

In this model the "am" variable has a t value of 2.08 and a P value of .0467, qhich means that it is significant (barely) at the .05 significance level. The wt and qsec variables are highly significant. The estimate for the "am" variable would suggest that cars with manual transmissions have a 2.9358 higher mpg than cars with automatic transmission (if all else is equal).

Residuals

Residuals have been plotted (appendix figure 5) along with other diagnostics. In looking at these plots we do see a slight U-shape in the residual best fit line so we may need to explore adding quadratic terms but there does not apper to be any heteroskedasticity issues. The second plot shows that the results are basically normally distributed (although there is some non-normality displayed by extreme values). The fourth plot shows that there are no observations that exert unacceptable leverage.

There are 2 outliers, they are: Chrysler Imperial, Fiat 128

Certainty

```
##                2.5 % 97.5 %
## (Intercept) -4.63830 23.874
## wt          -5.37333 -2.460
## qsec         0.63457  1.817
## am           0.04573  5.826
```

With 95% confidence, the true mean of the difference in MPG for Manual transmission is between 0.0457 and 5.8259.

Appendix

About the data

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models).

Figure 1

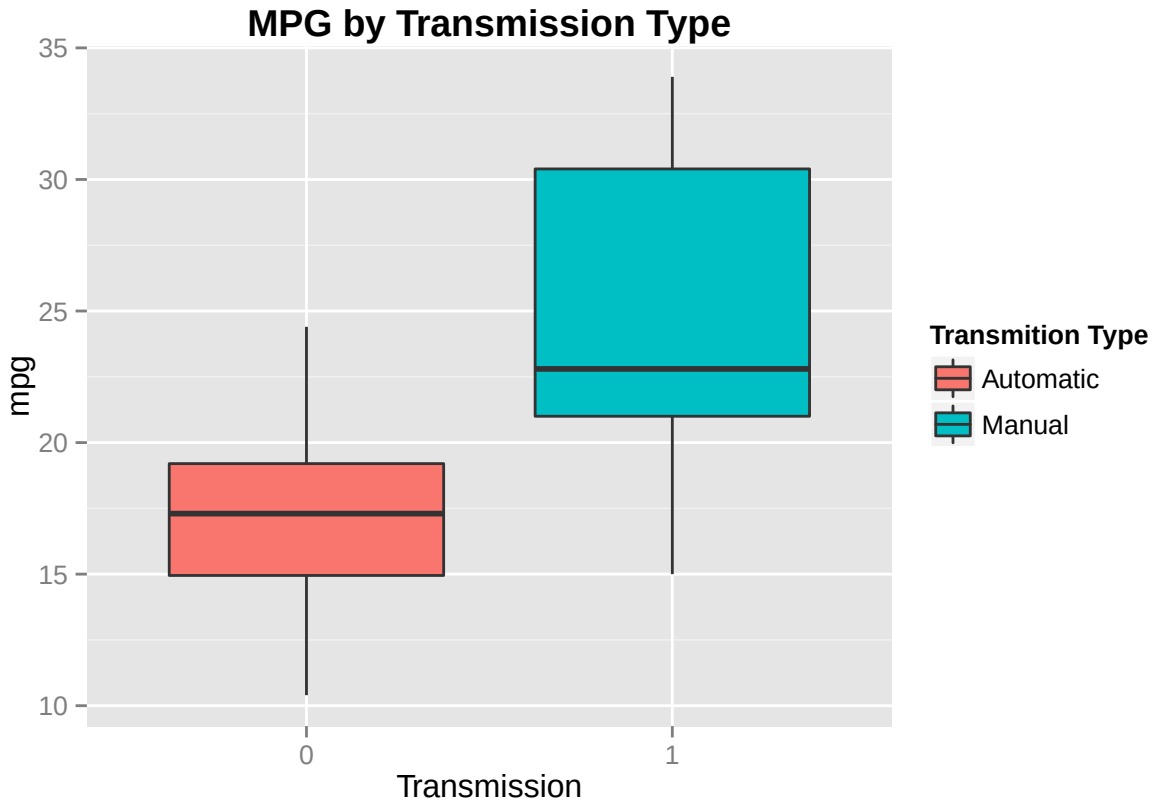

Figure 2

```
cor(mtcars)
```

```
##          mpg     cyl    disp      hp     drat      wt    qsec      vs
## mpg   1.0000 -0.8522 -0.8476 -0.7762  0.68117 -0.8677  0.4187  0.6640
## cyl  -0.8522  1.0000  0.9020  0.8324 -0.69994  0.7825 -0.5912 -0.8108
## disp -0.8476  0.9020  1.0000  0.7909 -0.71021  0.8880 -0.4337 -0.7104
## hp   -0.7762  0.8324  0.7909  1.0000 -0.44876  0.6587 -0.7082 -0.7231
## drat  0.6812 -0.6999 -0.7102 -0.4488  1.00000 -0.7124  0.0912  0.4403
## wt   -0.8677  0.7825  0.8880  0.6587 -0.71244  1.0000 -0.1747 -0.5549
## qsec  0.4187 -0.5912 -0.4337 -0.7082  0.09120 -0.1747  1.0000  0.7445
## vs    0.6640 -0.8108 -0.7104 -0.7231  0.44028 -0.5549  0.7445  1.0000
## am    0.5998 -0.5226 -0.5912 -0.2432  0.71271 -0.6925 -0.2299  0.1683
## gear  0.4803 -0.4927 -0.5556 -0.1257  0.69961 -0.5833 -0.2127  0.2060
## carb -0.5509  0.5270  0.3950  0.7498 -0.09079  0.4276 -0.6562 -0.5696
##            am    gear     carb
## mpg   0.59983  0.4803 -0.55093
## cyl  -0.52261 -0.4927  0.52699
## disp -0.59123 -0.5556  0.39498
## hp   -0.24320 -0.1257  0.74981
```

```
## drat  0.71271  0.6996 -0.09079
## wt   -0.69250 -0.5833  0.42761
## qsec -0.22986 -0.2127 -0.65625
## vs    0.16835  0.2060 -0.56961
## am    1.00000  0.7941  0.05753
## gear  0.79406  1.0000  0.27407
## carb  0.05753  0.2741  1.00000
```

Figure 3

```
summary(fitall)$coef
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788  0.6573  0.51812
## cyl         -0.11144    1.04502 -0.1066  0.91609
## disp         0.01334    0.01786  0.7468  0.46349
## hp          -0.02148    0.02177 -0.9868  0.33496
## drat         0.78711    1.63537  0.4813  0.63528
## wt          -3.71530    1.89441 -1.9612  0.06325
## qsec         0.82104    0.73084  1.1234  0.27394
## vs           0.31776    2.10451  0.1510  0.88142
## am           2.52023    2.05665  1.2254  0.23399
## gear         0.65541    1.49326  0.4389  0.66521
## carb        -0.19942    0.82875 -0.2406  0.81218
```

Figure 4

```
summary(bestfit)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.481 -1.556 -0.726  1.411  4.661
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.618      6.960    1.38  0.17792
## wt            -3.917      0.711   -5.51    7e-06 ***
## qsec           1.226      0.289    4.25  0.00022 ***
## am             2.936      1.411    2.08  0.04672 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.46 on 28 degrees of freedom
## Multiple R-squared:  0.85,   Adjusted R-squared:  0.834
## F-statistic: 52.7 on 3 and 28 DF,  p-value: 1.21e-11
```

Figure 5

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