MPG of Auto vs Manual Transmissions

Kent Mok

October 22, 2015

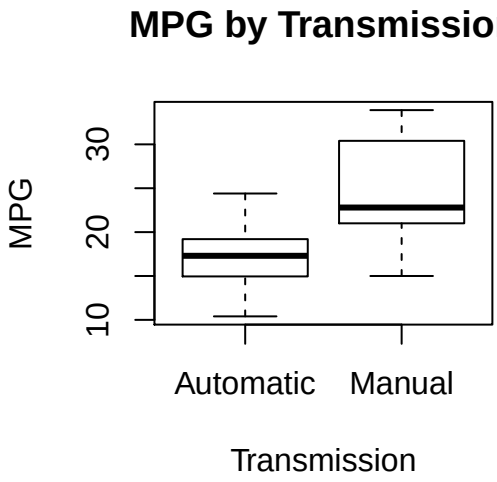
Executive Summary

Motor Trends provided fuel consumption data in 1974 for 32 automobiles along with 10 aspects of automobile design and performance. The analysis below uses the data to answer the questions "Is an automatic or manual transmission better for MPG?" It also attempts to "quantify the MPG difference between automatic and manual transmissions." Regression models are used to model the expected gas mileage for automatic versus manual transmissions. Relevant and significant aspects of automotive design are also included for more accurate regression. The final model achieves an adjusted R-squared of **82.27%** (model explains 82.27% of the total variance) and concludes that manual transmissions are better for gas mileage than automatic transmissions. The average improvement in gas mileage gained by choosing manual over automatic transmissions is **2.084 MPG**.

Exploratory Data Analyses

A quick preview of the Motor Trends MPG data provides the gas mileage for various cars as well as other characteristics of the various cars.

A boxplot of the MPG data versus the transmission type shows that the manual transmission is better for MPG. An initial look shows the average MPG of automatic transmissions at 17.15 and the average MPG of manual transmissions at 24.39.



Regression Models

Simple Linear Regression

The initial model tackles the main question with a linear model and am as the only regressor: $mpg = \beta_0 + \beta_1 am$. For this regression, $\beta_0 =$ **17.147** and $\beta_1 =$ **7.245**.

Multivariate Regression

A review of the correlation of the MPG variable to the other variables show that the cyl, disp, hp, and wt variables have the highest correlation with MPG. These variables should be considered for inclusion in the model. It is also noted that cyl has very high correlation with disp and hp as well.

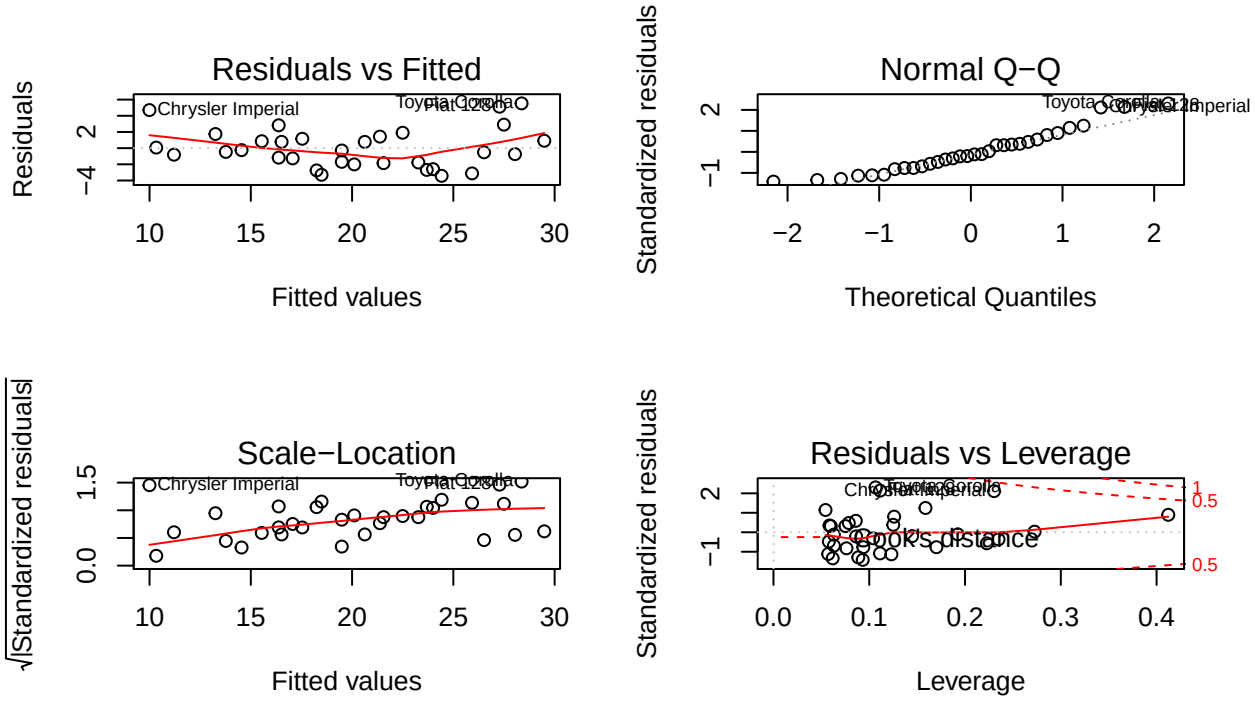
By adding doing trial regressions with each variable added to the base model, the hp variable shows to yield in the lowest RSS and p-value. Including hp reduces the Residual Sum of Squares from **720.90** down to **245.44** with a p-value of **2.92 x 10^{-8}**.

Applying the same method with am and hp held as regressors shows the wt to have the next most significant impact on the mpg model. Including wt bring the Residual Sum of Squares from **245.44** down to **220.55** with a p-value of **0.08636**.

Further analysis modeling with the remaining variables of cyl and disp do not appear to add value to the model, so they are left out. The final model for mpg can be shown as $mpg = \beta_0 + \beta_1 am + \beta_2 hp + \beta_3 wt$, where $\beta_0 = 34.003$, $\beta_1 = 2.084$, $\beta_1 = -0.037$, and $\beta_1 = -2.879$. Going from automatic to manual transmission improves the gas mileage by **2.084 MPG**, and not the optimistic 7.245 MPG as estimated by the initial model.

Regression Diagnostics

Below are the diagnostic plots for the final regression model.



The residuals plot shows that the residuals are small in value and stay close to the horizontal line at y = 0 for the length of the fitted values. There are no significant outliers, which shows that the model has done a decent job.

The Normal Q-Q plot shows that all the points fall close to the dotted line, which shows that the errors are normally distributed so they satisfy one of the assumptions required for least-squares regression.

Appendix

Preview of MPG data:

```
head(mtcars)
```

```
##                   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
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Correlation matrix for MPG data:

```
cor(mtcars)
```

```
##             mpg        cyl       disp         hp        drat         wt
## mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684  0.68117191 -0.8676594
## cyl  -0.8521620  1.0000000  0.9020329  0.8324475 -0.69993811  0.7824958
## disp -0.8475514  0.9020329  1.0000000  0.7909486 -0.71021393  0.8879799
## hp   -0.7761684  0.8324475  0.7909486  1.0000000 -0.44875912  0.6587479
## drat  0.6811719 -0.6999381 -0.7102139 -0.4487591  1.00000000 -0.7124406
## wt   -0.8676594  0.7824958  0.8879799  0.6587479 -0.71244065  1.0000000
## qsec  0.4186840 -0.5912421 -0.4336979 -0.7082234  0.09120476 -0.1747159
## vs    0.6640389 -0.8108118 -0.7104159 -0.7230967  0.44027846 -0.5549157
## am    0.5998324 -0.5226070 -0.5912270 -0.2432043  0.71271113 -0.6924953
## gear  0.4802848 -0.4926866 -0.5555692 -0.1257043  0.69961013 -0.5832870
## carb -0.5509251  0.5269883  0.3949769  0.7498125 -0.09078980  0.4276059
##             qsec         vs          am       gear        carb
## mpg   0.41868403  0.6640389  0.59983243  0.4802848 -0.55092507
## cyl  -0.59124207 -0.8108118 -0.52260705 -0.4926866  0.52698829
## disp -0.43369788 -0.7104159 -0.59122704 -0.5555692  0.39497686
## hp   -0.70822339 -0.7230967 -0.24320426 -0.1257043  0.74981247
## drat  0.09120476  0.4402785  0.71271113  0.6996101 -0.09078980
## wt   -0.17471588 -0.5549157 -0.69249526 -0.5832870  0.42760594
## qsec  1.00000000  0.7445354 -0.22986086 -0.2126822 -0.65624923
## vs    0.74453544  1.0000000  0.16834512  0.2060233 -0.56960714
## am   -0.22986086  0.1683451  1.00000000  0.7940588  0.05753435
## gear -0.21268223  0.2060233  0.79405876  1.0000000  0.27407284
## carb -0.65624923 -0.5696071  0.05753435  0.2740728  1.00000000
```

Basic linear model with am as regressor:

```
mdl <- lm(mpg ~ am, data = mtcars)
summary(mdl)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Multivariate model with am plus one additional regressor:

```
mdlwt <- lm(mpg ~ am + wt, data = mtcars)
anova(mdl, mdlwt)
mdlcyl <- lm(mpg ~ am + cyl, data = mtcars)
anova(mdl, mdlcyl)
mdldisp <- lm(mpg ~ am + disp, data = mtcars)
anova(mdl, mdldisp)
mdlhp <- lm(mpg ~ am + hp, data = mtcars)
anova(mdl, mdlhp)
```

Multivariate model with am, hp, and one additional regressor:

```
## mdlhpwt <- lm(mpg ~ am + hp + wt, data = mtcars)
summary(mdlhpwt)
```

```
##
## Call:
## lm(formula = mpg ~ am + hp + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## am           2.083710   1.376420   1.514 0.141268
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## wt          -2.878575   0.904971  -3.181 0.003574 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

```
## anova(mdlhp, mdlhpwt)
## mdlhpcyl <- lm(mpg ~ am + hp + cyl, data = mtcars)
## anova(mdlhp, mdlhpcyl)
## mdlhpdisp <- lm(mpg ~ am + hp + disp, data = mtcars)
## anova(mdlhp, mdlhpdisp)
```

Multivariate model with am, hp, wt, and one additional regressor:

```
mdlhpwtcyl <- lm(mpg ~ am + hp + wt + cyl, data = mtcars)
anova(mdlhpwt, mdlhpwtcyl)
mdlhpwtdisp <- lm(mpg ~ am + hp + wt + disp, data = mtcars)
anova(mdlhpwt, mdlhpwtdisp)
```