Regression Project: Fuel Economy by Transmission Type

Zain Tejani

Sunday, December 21, 2014

Executive Summary

The mtcars dataset was loaded into R to study the relationship between fuel economy (mpg) and transmission type (automatic or manual, defined by the am column). Using multivariate regression, confounders such as weight (wt), acceleration (parametrized as the time to run a quarter mile, column name qsec), and horsepower (hp) were selected to allow for the best prediction of the effect of transmission type on fuel economy. The adjusted R squared value and residual standard error were studied to quantifiably confirm the best choice of parameters.

Analysis

Using the following code, the mtcars data was loaded.

```
library(datasets)
data(mtcars)
```

Using the ?mtcars command in R, one can see that the transmission classification is binary in the “am” column, with a value of 0 representing an automatic gearbox, and 1 representing a manual gearbox.

The column names were explored, to determine which variables could act as confounders to mpg.

```
colnames(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

The development of the regression model was initially based on knowledge of cars and their workings. It stands to reason that heavier cars consume more fuel, so weight was selected as a confounder. Furthermore, cars that accelerate faster tend to burn fuel quicker. Hence, the qsec variable (a drag racing measurement) was selected as aonther confounder.

Beyond this point, confounders were selected based on trial and error, comparing the R-squared values, residual standard error, and Pr(>|t|) values to select the best alternative for the model. After numerous trials, the hp and carb variables were the two best alternatives found.

The coefficient tables for the models can be seen below:

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

```
##                Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 17.44019110  9.3188688  1.871492 0.072149342
## am           2.92550394  1.3971471  2.093913 0.045790788
## wt          -3.23809682  0.8898986 -3.638726 0.001141407
## qsec         0.81060254  0.4388703  1.847021 0.075731202
## hp          -0.01764654  0.0141506 -1.247052 0.223087932
```

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

```
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 12.8971970  7.4724922  1.725957 0.095783676
## am           3.5113918  1.4874727  2.360643 0.025720752
## wt          -3.4343216  0.8200199 -4.188095 0.000268619
## qsec         1.0191298  0.3377635  3.017288 0.005507044
## carb        -0.4886034  0.4212171 -1.159980 0.256211958
```

A comparison of the adjusted R squared values and residual errors can be seen below:

```
a<-rbind(summary(lm(mpg~am+wt+qsec+hp, data=mtcars))[[9]],summary(lm(mpg~am+wt+qsec+carb, data=mtc
ars))[[9]])
b<-rbind(summary(lm(mpg~am+wt+qsec+hp, data=mtcars))[[6]],summary(lm(mpg~am+wt+qsec+carb, data=mtc
ars))[[6]])
c<-cbind(a,b)
colnames(c)<-c("Adjusted R Squared","Residual Standard Error")
rownames(c)<-c("hp","carb")
print(c)
```

```
##      Adjusted R Squared Residual Standard Error
## hp            0.8367919                2.434828
## carb          0.8355852                2.443813
```

As visible in the table above, the hp model has a higher R squared value and lower residual standard error, making it the preferred choice. Furthermore, for the hp model, the standard error in the estimate of the am parameter is lower, allowing for a more accurate prediction. The selection of the hp model also complements the “common sense” approach, as horsepower can act as a confounder to mpg, hence would predict the am parameter better when held constant.