Linear Regression Project

Jeff Wolbach

Wednesday, September 17, 2014

Oddly, this document probably will not include any code. If one considers only a naive model of strictly mpg vs. transmission, one can conclude that cars with manual transmissions average 7.2 mpg more than cars with automatic transmissions. However, essentially all of that difference is due to factors other than the type of transmission. When accounting for all relevant variables, the dependence of mpg on transmission type is statistically zero.

The mtcars dataset contains information on fuel consumption and 10 aspects of automobile design and performance for 32 automobiles from the 1973-74 model year. One of these aspects is the type of transmission, which is encoded in the dataset as a numeric variable, am, taking a value of 0 for an automatic transmission and 1 for a manual transmission.

A boxplot can be constructed to show the variation in mpg with the type of transmission (figure A1); it is seen that cars with manual transmissions have generally higher mpg than cars with automatic transmissions. A linear regression model of mpg vs. transmission type yields the equation: mpg = 17.15 + 7.24*am, where the intercept may be interpreted as the average mpg for cars with automatic transmissions (am=0) and the sum of the slope and intercept (24.39) as the average mpg for cars with manual transmissions (am=1). The fit of the equation to the data, and the residuals from the fit, are shown in figure A2.

We should consider other variables that may be related to mpg more strongly than the type of transmission. Figure A3 displays correlation plots of mpg with the other 9 aspects of automobile design given in the dataset, and we may calculate the the correlation coefficient of each design aspect with mpg:

```
##    cyl   disp     hp   drat     wt   qsec     vs     am   gear   carb
## -0.852 -0.848 -0.776  0.681 -0.868  0.419  0.664  0.600  0.480 -0.551
```

I considered five linear regression models for mpg vs. design aspects, including the type of transmission. The models were:

1. mpg vs. am
2. mpg vs. am + wt
3. mpg vs. am + wt + cyl
4. mpg vs. am + wt + cyl + hp
5. mpg vs. am + wt + cyl + hp + carb

The ANOVA analysis for these five models produced:

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + cyl
## Model 4: mpg ~ am + wt + cyl + hp
## Model 5: mpg ~ am + wt + cyl + hp + carb
##   Res.Df RSS Df Sum of Sq     F  Pr(>F)
## 1     30 721
## 2     29 278  1       443 70.72 6.9e-09 ***
## 3     28 191  1        87 13.94 0.00093 ***
```

```
## 4     27 170  1        21  3.36 0.07813 .
## 5     26 163  1         7  1.16 0.29068
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

indicating that the addition of the design aspects hp and carb in models 4 and 5 did not produce a statistically meaningful improvement to model 3, and hence we should only consider the first three models.

In model 1:

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.134e-15
## am             7.245      1.764   4.106 2.850e-04
```

In model 2:

```
##             Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 37.32155     3.0546 12.21799 5.843e-13
## am          -0.02362     1.5456 -0.01528 9.879e-01
## wt          -5.35281     0.7882 -6.79081 1.867e-07
```

and in model 3:

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  39.4179     2.6415 14.9228 7.425e-15
## am            0.1765     1.3045  0.1353 8.933e-01
## wt           -3.1251     0.9109 -3.4309 1.886e-03
## cyl          -1.5102     0.4223 -3.5764 1.292e-03
```

In models 2 and 3, the coefficient on am can not be distinguished from zero. When the weight of the automobile is considered, this dataset does not support a distinction in mpg based on the type of transmission.

Appendix

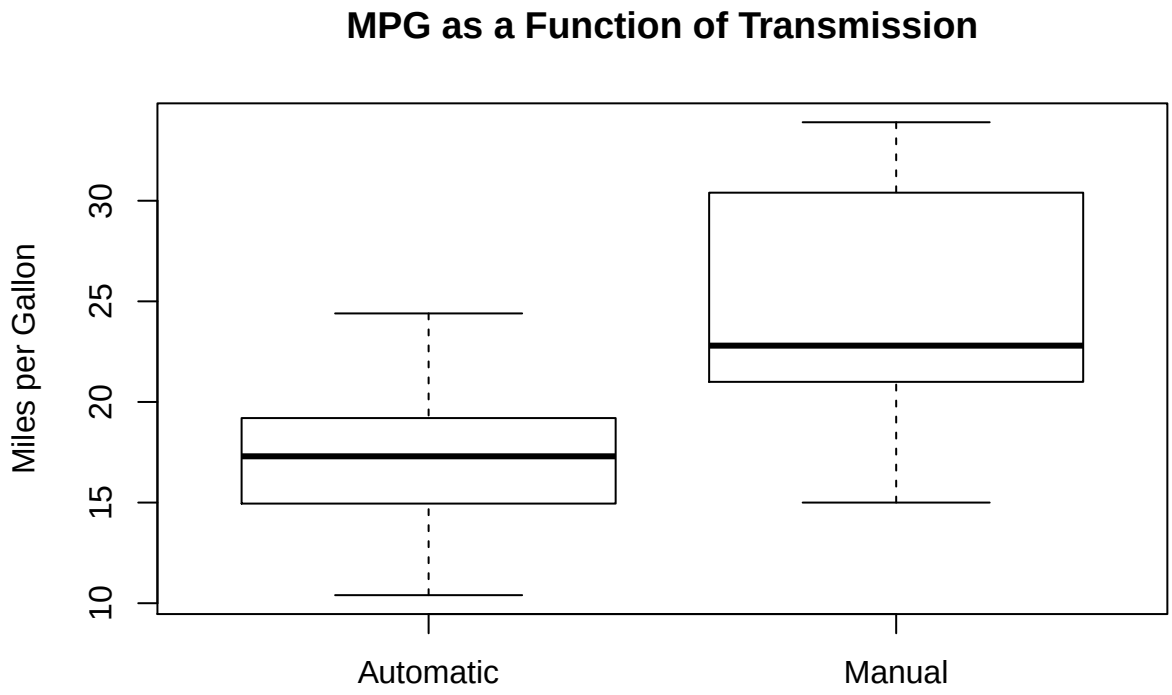

Figure A1

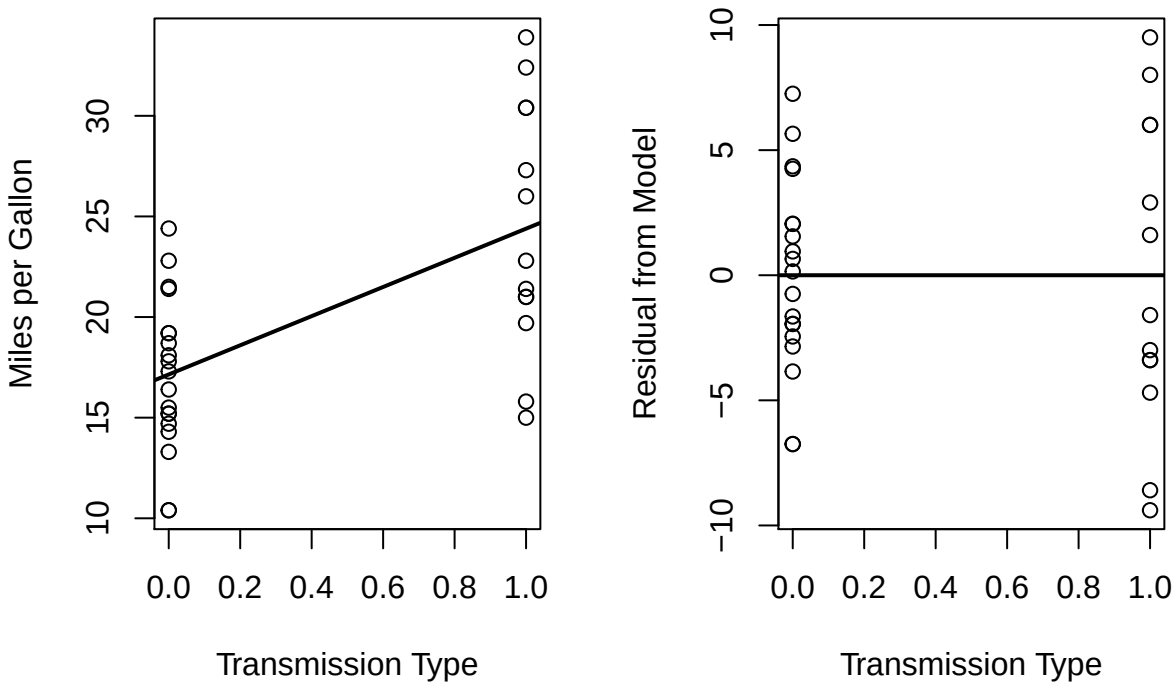
Miles per Gallon
10
15
20
25
30
0.0
0.2
0.4
0.6
0.8
1.0
Transmission Type
Residual from Model
−10
−5
0
5
10
0.0
0.2
0.4
0.6
0.8
1.0
Transmission Type

Figure A2

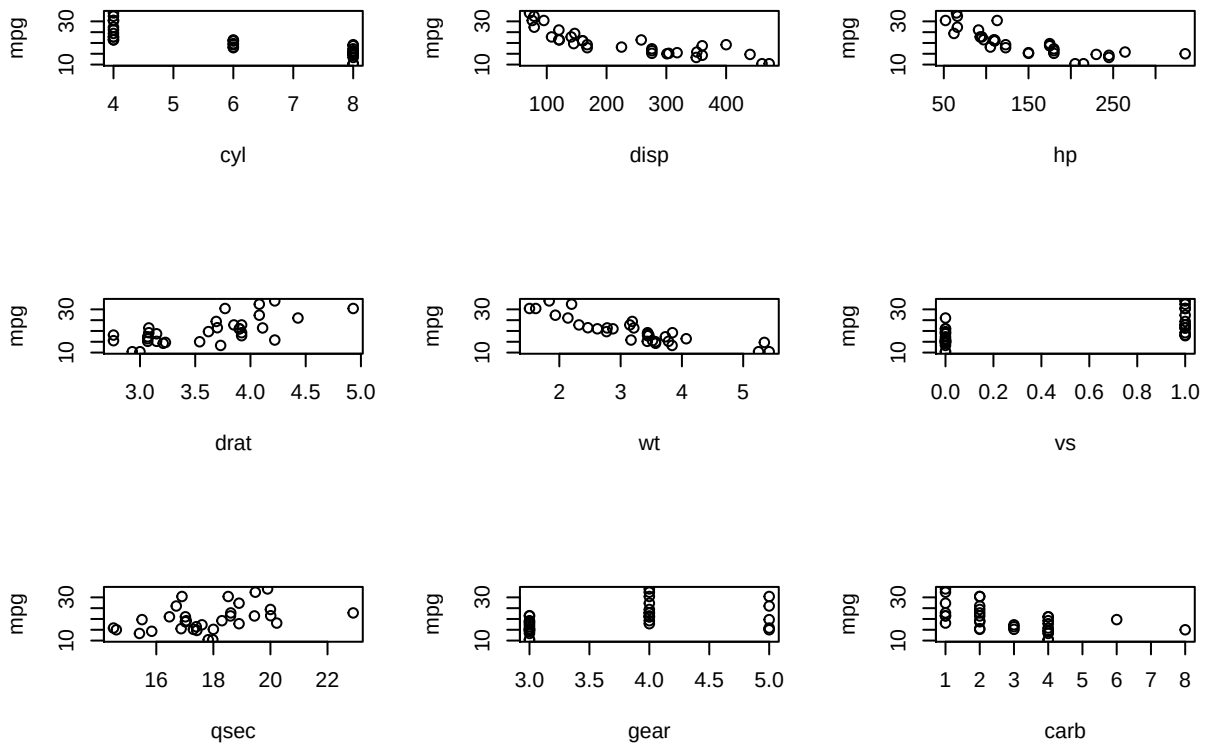

Figure A3

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.