Regression Models v1.0

John Rood

Sunday, September 27, 2015

Executive Summary

We have two problems which have been set for us: 1. "Is an automatic or manual transmission better for MPG?" 2. "Quantify the MPG difference between automatic and manual transmissions"

We are to investigate the two problems using the mtcars dataset. Our first hypothesis is that the weight of the vehicle is the determinant of which type of transmission is better for MPG. Investigating this, it would appear that this doesn't answer the question. We then investigate additionally the effect of how "quick" a car is, as measured by its time in the quarter-mile. Again this doesn't clearly answer the question, but a conjecture about this is formulated.

Findings

Exploratory data results and graphs (Appendix)

```
library( scatterplot3d)
attach( mtcars )
#summary( cars )
#head( mtcars )
d <- dim( mtcars )
d
```

```
## [1] 32 11
```

Some models with residuals (plots in Appendix)

```
#The base model
fit0 <- lm(formula = mpg ~ am, data = mtcars)
fit0
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Coefficients:
## (Intercept)           am
##      17.147        7.245
```

```
e0 <- resid( fit0 )

#The obvious models assuming wt is most important (cf forums)
fit1 <- lm(formula = mpg ~ am + wt, data = mtcars)
fit1
```

```
##
## Call:
## lm(formula = mpg ~ am + wt, data = mtcars)
##
## Coefficients:
## (Intercept)           am           wt
##    37.32155     -0.02362     -5.35281
```

```
e1 <- resid( fit1 )

#anova( fit1, fit0 )

fit2 <- lm(formula = mpg ~ am + wt + am * wt, data = mtcars)
fit2
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + am * wt, data = mtcars)
##
## Coefficients:
## (Intercept)           am           wt        am:wt
##      31.416       14.878       -3.786       -5.298
```

```
e2 <- resid( fit2 )

#anova( fit2, fit1 )

#These don't seem to solve our problem.

#What other variable(s) ... ?

#Speed Racer ...
fit3 <- lm(formula = mpg ~ am + wt + qsec, data = mtcars)
fit3
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + qsec, data = mtcars)
##
## Coefficients:
## (Intercept)           am           wt         qsec
##       9.618        2.936       -3.917        1.226
```

```
e3 <- resid( fit3 )

anova( fit3, fit1 )
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + wt + qsec
## Model 2: mpg ~ am + wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
```

```
## 1     28 169.29
## 2     29 278.32 -1   -109.03 18.034 0.0002162 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Speed Racer actually seems somewhat useful!  Could it be the case that this variable actually introduc
```

Conclusions regarding the two main questions.

From the models investigated here the answers to the two main questions are not clear. In the terminology of the grading rubric, the questions are not answerable.

Quantification of uncertainty. Inference.

Acknowledgement.

The discussion in the class forums helped considerably with this project, as far as i am able to understand what i am doing.

Appendix

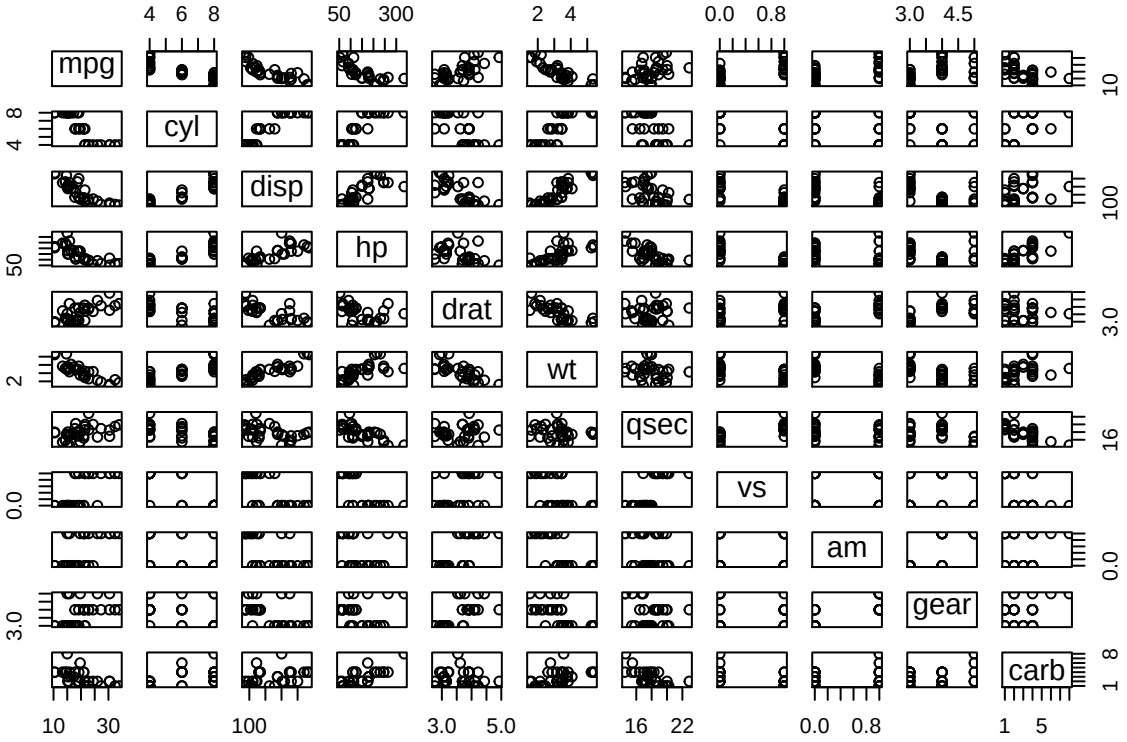
Meaning of the Dataset

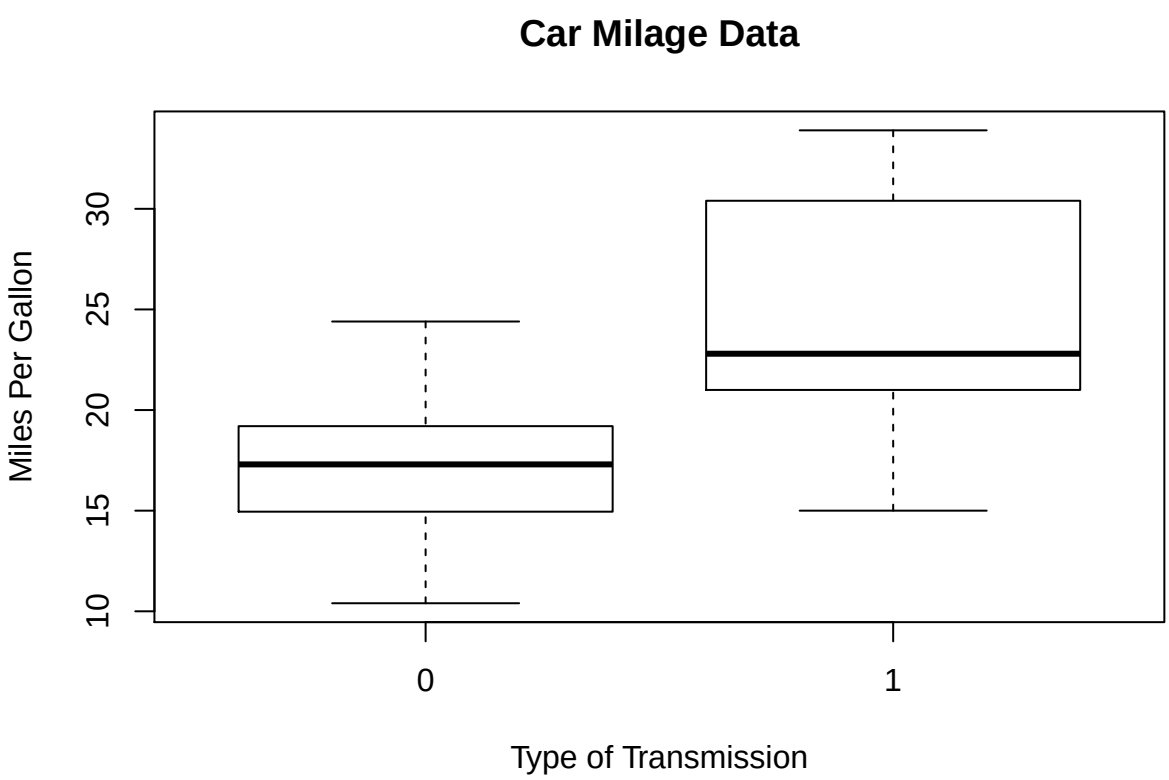
By using help(mtcars), we find (in part):

A data frame with 32 observations on 11 variables.

[, 1] mpg Miles/(US) gallon
[, 2] cyl Number of cylinders
[, 3] disp Displacement (cu.in.)
[, 4] hp Gross horsepower
[, 5] drat Rear axle ratio
[, 6] wt Weight (lb/1000)
[, 7] qsec 1/4 mile time
[, 8] vs V/S –cylinder configuration, V-shaped or straight
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears
[,11] carb Number of carburetors

Exploratory plots (graphs)




Car Milage Data
Miles Per Gallon
10
15
20
25
30
0
1
Type of Transmission

Plots of residuals