Automatic vs Manual Transmission for MPG, using mtcars Data, for Regressions Models Coursera Course

CRR

Sunday, July 26, 2015

EXECUTIVE SUMMARY

This project explored the relationship between automobile MPG (miles per gallon) and other variables given in the mtcars data set, especially transmission type, using linear models. It was determined (based on the corresponding coefficient for the linear model selected of +1.56, where the variable is 1 for manual and 0 for automatic) that (1) manual transmission was better than automatic for MPG and (2) that manual transmissions added about 1.56 MPG over automatic, when other variables considered in the best model found were held constant. Model selection was made using anova results.The 95% confidence interval for the transmission variable coefficient was about -1.40 to +4.52; that this range includes negative values reflects that other variables must be taken into account to properly model MPG, but the range is skewed well positive, making the conclusions reached here reasonable overall and in most specific cases, accurate as well.

DATA PROCESSING - Exploring Data, Fitting Linear Models - see code chunk comments for details

```
##NOTE: Output here hidden to reduce report size
##Automatic vs. manual transmission for mpg, mtcars data, for Regression Models Coursera Course
##Getting data and basic exploration of the data
data(mtcars);head(mtcars, 1);str(mtcars)
##?mtcars - this gives useful info, but commented out here to avoid confusion when apply knitr
##Looking at relationships between mpg and other variables; Variable selections
pairs(mtcars)
##Various linear models for mpg from transmission only to all variables in mtcars data
fit1<-lm(mpg~am, data=mtcars);summary(fit1)
fit2<-lm(mpg~cyl+disp+hp+wt+am, data=mtcars);summary(fit2)
fit3<-lm(mpg~cyl+disp+hp+wt+vs+am, data=mtcars);summary(fit3)
fit4<-lm(mpg~., data=mtcars);summary(fit4)
```

RESULTS - Model Selection (Anova) & Statistics (confidence intervals etc) - see code chunk comments for details

```
##Model testing and selecting best of these - used anova on nested models and p value progression
anova(fit1, fit2, fit3, fit4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + disp + hp + wt + am
## Model 3: mpg ~ cyl + disp + hp + wt + vs + am
## Model 4: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
```

```
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     26 163.12  4    557.78 19.8538 6.809e-07 ***
## 3     25 159.52  1      3.60  0.5123    0.4820
## 4     21 147.49  4     12.03  0.4281    0.7867
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

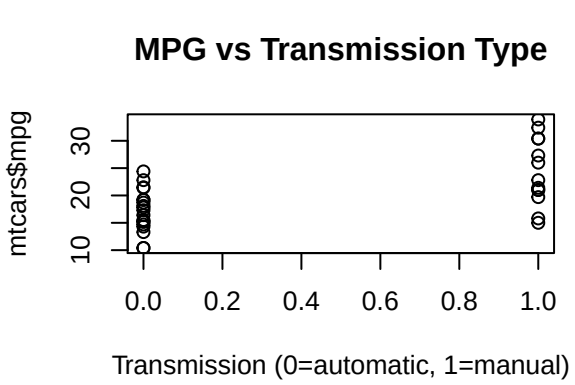
```
##Summary (w/statistics), coefficients & 95% confidence intervals for coeffs of model chosen
summary(fit2);coef(fit2);confint(fit2)
```

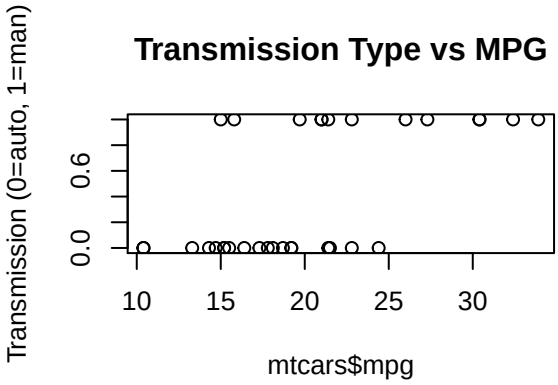
```
##
## Call:
## lm(formula = mpg ~ cyl + disp + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5952 -1.5864 -0.7157  1.2821  5.5725
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.20280    3.66910  10.412 9.08e-11 ***
## cyl         -1.10638    0.67636  -1.636  0.11393
## disp         0.01226    0.01171   1.047  0.30472
## hp          -0.02796    0.01392  -2.008  0.05510 .
## wt          -3.30262    1.13364  -2.913  0.00726 **
## am           1.55649    1.44054   1.080  0.28984
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.505 on 26 degrees of freedom
## Multiple R-squared:  0.8551, Adjusted R-squared:  0.8273
## F-statistic:  30.7 on 5 and 26 DF,  p-value: 4.029e-10

## (Intercept)         cyl        disp          hp          wt          am
## 38.20279869 -1.10637984  0.01225708 -0.02796002 -3.30262301  1.55649163

##                   2.5 %       97.5 %
## (Intercept) 30.66086287 45.7447345009
## cyl         -2.49664757  0.2838878816
## disp        -0.01180588  0.0363200383
## hp          -0.05657652  0.0006564778
## wt          -5.63285880 -0.9723872244
## am          -1.40457258  4.5175558419
```

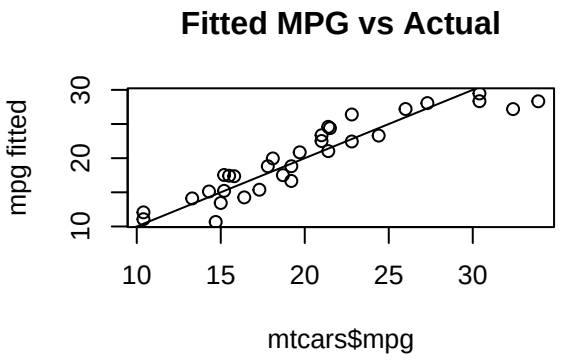
APPENDIX - Plots to evaluate/provide diagnostics for model chosen

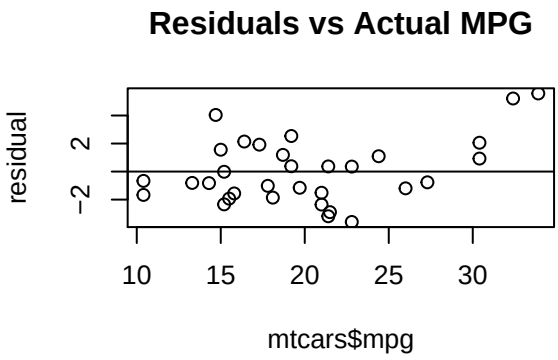
```
##Plots to evaluate/provide diagnosics for model chosen
par(mfrow=c(2,2))
plot(mtcars$am,mtcars$mpg, xlab="Transmission (0=automatic, 1=manual)", main="MPG vs Transmission Type")
plot(mtcars$mpg,mtcars$am, ylab="Transmission (0=auto, 1=man)", main="Transmission Type vs MPG")
```

```r
plot(mtcars$mpg,fit2$fitted,ylab="mpg fitted", main="Fitted MPG vs Actual");abline(0,1)
plot(mtcars$mpg,fit2$resid, ylab="residual", main="Residuals vs Actual MPG");abline(0,0)
```

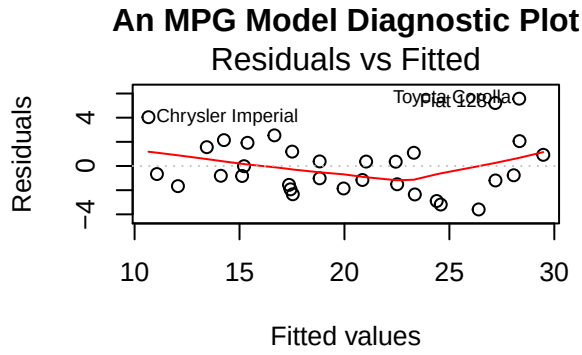


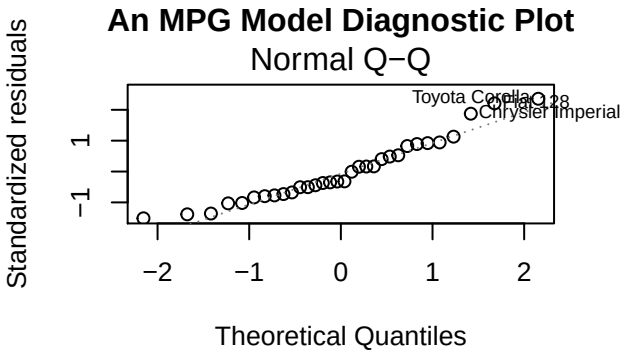


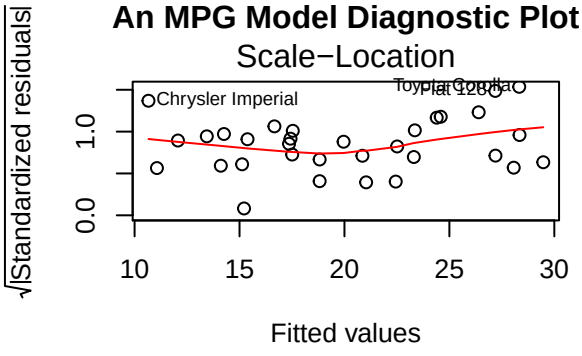


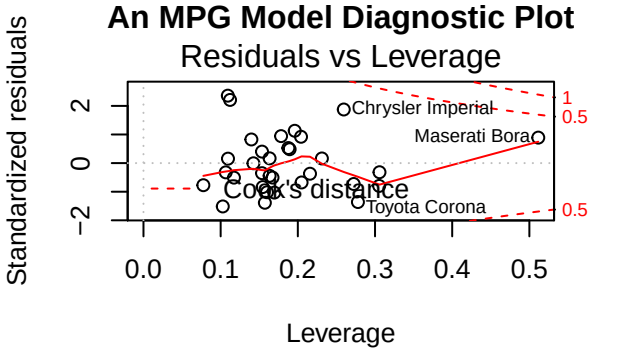



```r
par(mfrow=c(2,2));plot(fit2, main="An MPG Model Diagnostic Plot");par(mfrow=c(1,2))
```

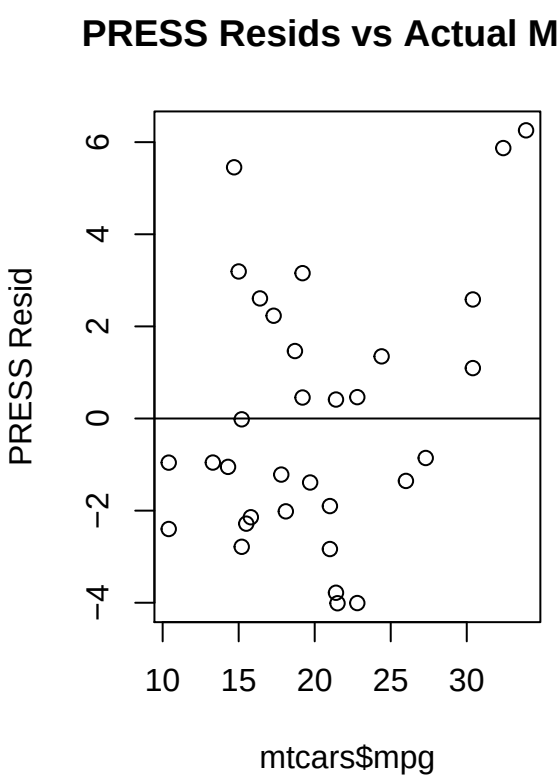


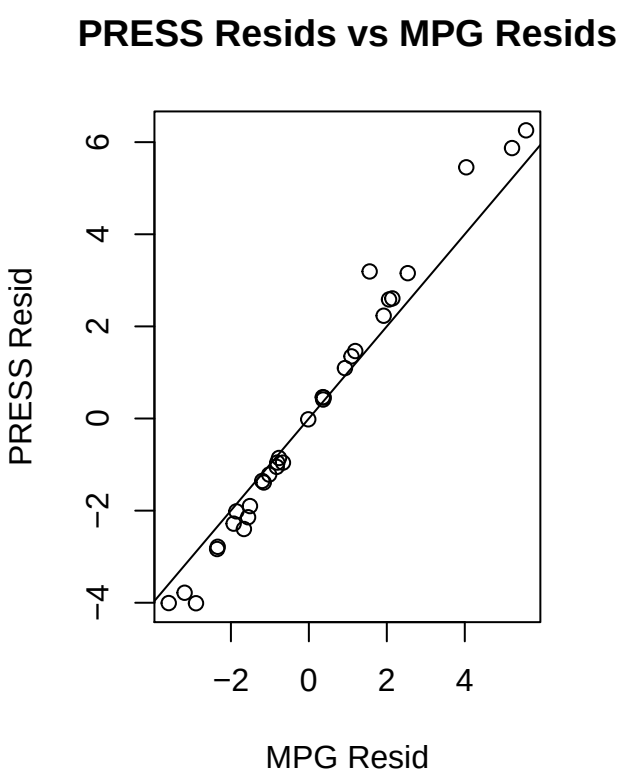







```
plot(mtcars$mpg,resid(fit2)/(1-hatvalues(fit2)),ylab="PRESS Resid", main="PRESS Resids vs Actual MPG");a
plot(resid(fit2),resid(fit2)/(1-hatvalues(fit2)), xlab="MPG Resid", ylab="PRESS Resid", main="PRESS Resi
```


PRESS Resids vs Actual MPG
PRESS Resid
6
4
2
0
−2
−4
10
15
20
25
30
mtcars$mpg


PRESS Resids vs MPG Resids
PRESS Resid
6
4
2
0
−2
−4
−2
0
2
4
MPG Resid