

Transmission and Fuel Efficiency

John James

April 15, 2016

Introduction

Data

Table 1 presents the summary statistics for the continuous and categorical variables in the Motor Trend Cars dataframe. There are 32 observations of 11 variables. Miles per gallon range from 10.4 to 33.9.

Methods

A univariate analysis was undertaken to establish the strength of the correlation between transmission type (am) and fuel efficiency, measured in miles per gallon (mpg). Next, a multivariate regression analysis model selection process was undertaken whereby each variable was added to the univariate linear model, one-by-one, in order of correlation (Table 2) with the dependent variable, mpg. This series of nested models was evaluated using Anova tests and all models in which their additional variables had no significant effect on the mpg (p.value < 0.05) were discarded. Lastly, to obtain constant variance, variables highly correlated with transmission type were removed and replaced by predictors with high correlation with mpg and a relatively low correlation with transmission type.

Analysis

Univariate Analysis

A linear regression was calculated to predict mpg based on transmission type (variable named, am). The linear equation produced an adjusted R^2 of 0.338 ($F(1, 30)=16.86$, $p < 0$). The low R^2 revealed a lack of fit with the model. Further, correlation amongst the omitted variables and the dependent variable (Table 2) was high, suggesting multicollinearity was a problem.

Multivariate Analysis

Regression Model

To isolate the effect of transmission type on fuel efficiency, a multiple linear regression was calculated to predict mpg, based upon transmission type (am), weight (wt), and horsepower (hp). A significant regression equation was found ($F(3, 28)=48.96$, $p < 0$), with an adjusted R^2 of 0.823. Predicted mpg was equal to $34.003 + 2.084 * (\text{Transmission}) - 2.879 * (\text{wt}) - 0.037 * (\text{hp})$, where Transmission was coded as 0 = Auto, 1 = Manual, wt is the weight of the vehicle measured in 1000lb, and hp was gross horsepower. Miles per gallon increased by 2.084 mpg for cars with manual transmissions, increased by -2.879 mpg for each 1000 lb of weight, and increased by -0.037 mpg for each unit of horsepower. Though weight and horsepower were significant predictors of mpg with p-values 0.004 and 0.001 respectively, transmission type was not. The p-value for transmission type is 0.141, significantly above our significance factor of $\alpha = 0.05$.

Regression Model Assessment

The following assumptions hold for this model

Existence

The relationship between dependent and independent variables exists. The model predictors produce an adjusted R^2 of 0.823, thereby explaining 82.3 percent of the variability in mpg.

Linearity

The partial residual plots shown in Figure 1 illustrates the linear relationship between the predictors and the outcome (mpg) over the spectrum of values studied.

Independence

- For given values of explanatory variable x, the y-values are independent of each other
- The explanatory variables are independent of each other - hence the name "independent variables"

Normality

For given values of the explanatory variables, the y-values are normally distributed #####*Constant Variance The distribution of y-values has equal variance at each value of x

Results

References

Appendix

Figures

Figure 2: Global Test of Model Assumptions

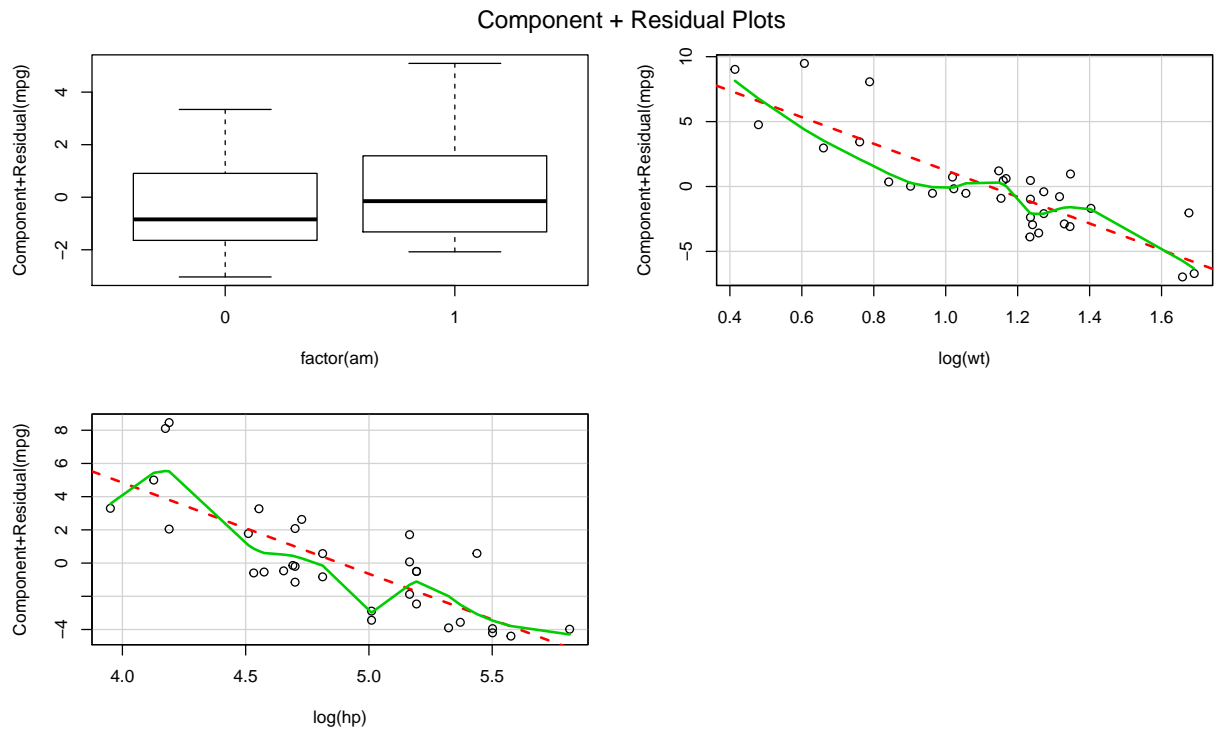


Figure 1: Partial Residual Plots

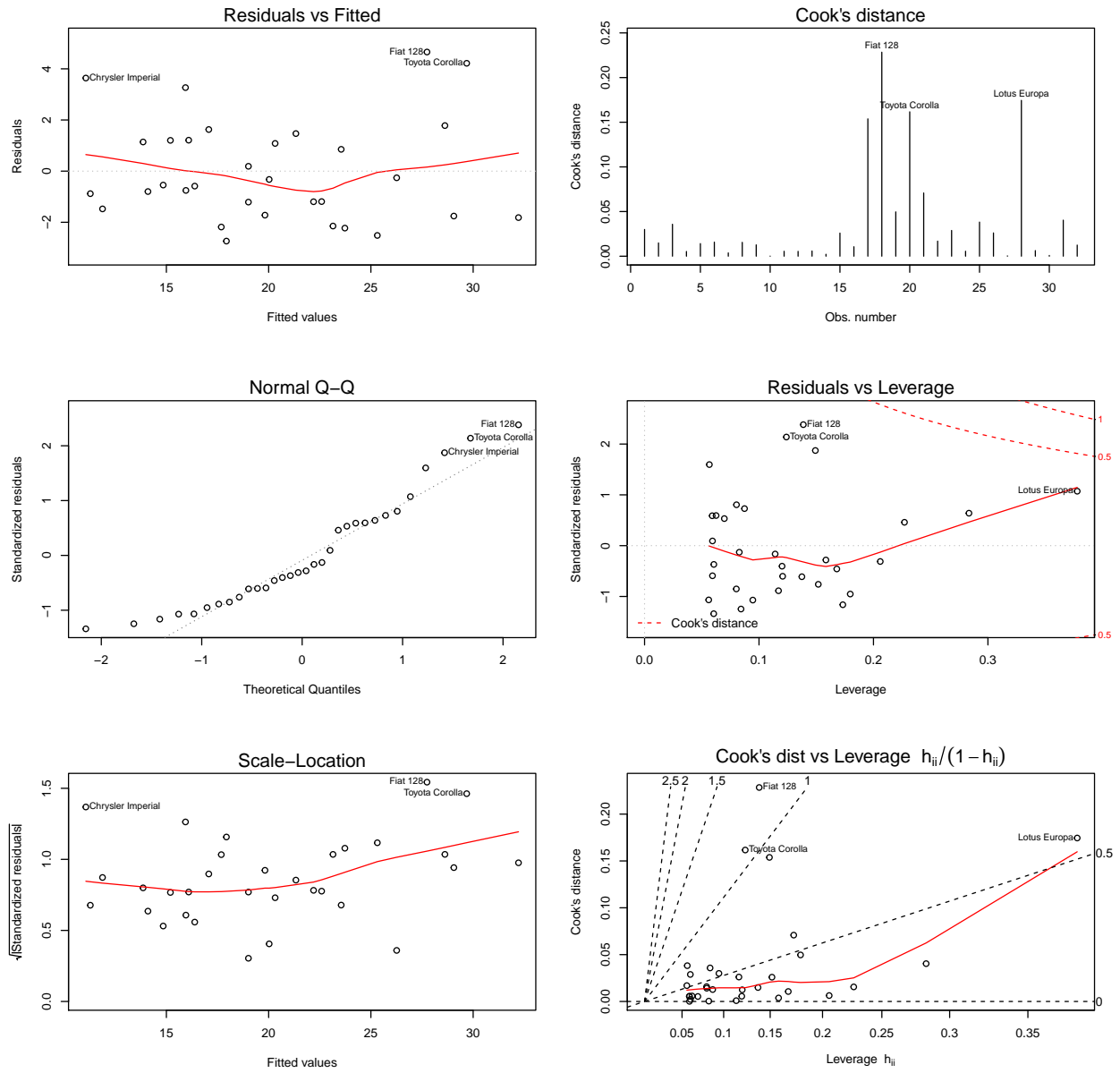


Figure 3: Fit plots for Multiple Regression Model

Tables

Table 1: Summary Statistics for Motor Trend Cars Data Frame

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear
mean	20.09062	6.1875	230.7219	146.6875	3.596563	3.21725	17.84875	0.4375	0.40625	3.6875
median	19.20000	6.0000	196.3000	123.0000	3.695000	3.32500	17.71000	0.0000	0.00000	4.0000
minimum	10.40000	4.0000	71.1000	52.0000	2.760000	1.51300	14.50000	0.0000	0.00000	3.0000
maximum	33.90000	8.0000	472.0000	335.0000	4.930000	5.42400	22.90000	1.0000	1.00000	5.0000
s.size	32.00000	32.0000	32.0000	32.0000	32.000000	32.00000	32.00000	32.0000	32.00000	32.0000

Table 3: Nested Anova Evaluation Model Set 3

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
30	720.8966	NA	NA	NA	NA
29	245.4393	1	475.4573110	72.8808330	0.0000000
28	220.5530	1	24.8862675	3.8147103	0.0616460
27	169.9978	1	50.5552515	7.7493999	0.0098812
26	169.6178	1	0.3799207	0.0582364	0.8111984

Table 4: Nested Anova Evaluation Model Set 4

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
30	720.8966	NA	NA	NA	NA
29	333.6786	1	387.217959	55.1314184	0.0000003
28	245.6514	1	88.027197	12.5331589	0.0019392
27	240.4557	1	5.195707	0.7397557	0.3994521
26	204.6202	1	35.835544	5.1022024	0.0346572
25	197.0111	1	7.609130	1.0833746	0.3097796
24	184.9960	1	12.015105	1.7106897	0.2050332
23	150.7149	1	34.281052	4.8808764	0.0384000
22	148.8475	1	1.867420	0.2658800	0.6114952
21	147.4944	1	1.353055	0.1926457	0.6652064

Table 5: Nested Anova Evaluation Model Set 5

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
30	720.8966	NA	NA	NA	NA
29	333.6786	1	387.217959	49.2017275	0.0000002
28	245.6514	1	88.027197	11.1851479	0.0025127
27	204.9699	1	40.681570	5.1691908	0.0314861
26	204.6202	1	0.349682	0.0444322	0.8346948

Table 6: Nested Anova Evaluation Model Set 5

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
30	720.8966	NA	NA	NA	NA
29	207.9715	1	512.9250783	88.3196128	0.0000000
27	156.2067	2	51.7647995	4.4566422	0.0250799
26	156.2017	1	0.0049811	0.0008577	0.9769265
25	133.7481	1	22.4536214	3.8662472	0.0633007
24	133.5881	1	0.1600021	0.0275505	0.8698365
23	128.9270	1	4.6611241	0.8025903	0.3809765
22	128.0461	1	0.8808933	0.1516794	0.7010517
21	125.9364	1	2.1097349	0.3632713	0.5534685
20	116.1520	1	9.7843428	1.6847477	0.2090651

Table 2: MPG Correlation Table (Absolute Value)

Variable	Correlation
Weight	0.8676594
Cylinders	0.8521620
Displacement	0.8475514
HP	0.7761684
Rear Axle Ratio	0.6811719
VS	0.6640389
Carburetors	0.5509251
Gears	0.4802848
Qtr Mile	0.4186840

Table 7: Transmission Correlation Table (Absolute Value)

Variable	Correlation
Carburetors	0.0575344
VS	0.1683451
Qtr Mile	0.2298609
HP	0.2432043
Cylinders	0.5226070
Displacement	0.5912270
Weight	0.6924953
Rear Axle Ratio	0.7127111
Gears	0.7940588

Table 8: Multiple Regression Summary

```
##
## Call:
## lm(formula = mpg ~ factor(am) + carb + vs, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.2803 -1.2308  0.4078  2.0519  4.8197
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  19.5174     1.6091  12.130 1.16e-12 ***
## factor(am)1    6.7980     1.1015   6.172 1.15e-06 ***
## carb         -1.4308     0.4081  -3.506 0.00155 **
## vs            4.1957     1.3246   3.168 0.00370 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.962 on 28 degrees of freedom
## Multiple R-squared:  0.7818, Adjusted R-squared:  0.7585
## F-statistic: 33.45 on 3 and 28 DF,  p-value: 2.138e-09
```

Code

```
anchors("index")
```

label	type	number
mtStats	Table	1
cordf1	Table	2
lm12Linearity	Figure	1
lm12global	Figure	2
plotlm12	Figure	3
anovaSet3	Table	3
anovaSet4	Table	4
anovaSet5	Table	5
anovaSet6	Table	6
cordf2	Table	7
linearModel	Table	8