Regression Models Course Project

Executive Summary

Per the analysis below, the relationship between mpg and the available data is best catpured using a model that regresses weight and horse power against mpg. For a given horse power and weight, manual transmission have a higher mpg. The incremental effect of a manual transmission is approximatley 2.08 additional miles/gallon.

Synopsis

This analysis will explore the relationships between 11 variables for 32 automobiles collected by Motor Trend magazine in 1974 and answer the following questions:

- Is an automatic or manual transmission better for MPG?
- What is the MPG difference between automatic and manual transmissions?

Loading and Reviewing Raw Data

The dataset used is the mtcars dataframe in the R datasets package.

A description of the variables in this dataset can be found here.

Exploration

A pairs comparison of the variables in the dataset can be found here. Some variables appear highly correlated with mpg.

Model Selection

Initial Transmision Model

A model must balances bias and variability in model estimates. Below is summary of a regression using only transmission. It captures a low percentage of the observered variation in mpg values.

	Adj.R.Sqred	F.Stat	Resid.Std.Err
mpg ~ factor(am)	0.3385	16.8603	4.9020

Table 1: Transmission Only Model Fit

Selection Of Variables

To selectively identify model variables, I used the variable with the highest correlations with mpg, weight.

	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	-0.85	-0.85	-0.78	0.68	-0.87	0.42	0.66	0.60	0.48	-0.55

Table 2: Mpg Correlations

I then examined the next three hightest correlations and used the one with the smallest correltain with weight, horse power.

	cyl	disp	hp
wt	0.78	0.89	0.66

Table 3: Weight Correlations

Analysis of Variance

To check if any additional variables should be added, an analysis was done to see the impact of adding one additional varible. None of the other variables had statistically significant p-values. Only mpg, weight and horse power will be used in the model to analyize transmission impact.

	Adj.R.Sqred	F.Stat	Resid.Std.Err	anova.P.Value
mpg ~ .	0.8066	13.9325	2.6502	
mpg ~ wt	0.7446	91.3753	3.0459	
mpg ~ wt + hp	0.8148	69.2112	2.5934	0.0015
mpg ~ wt + hp + factor(am)	0.8227	48.9600	2.5375	0.1413
mpg ~ wt + hp + factor(cyl)	0.8361	40.5253	2.4402	0.0736
mpg ~ wt + hp + drat	0.8194	47.8839	2.5613	0.1988
mpg ~ wt + hp + qsec	0.8171	47.1528	2.5778	0.2546
mpg ~ wt + hp + factor(vs)	0.8150	46.5163	2.5924	0.3207
mpg ~ wt + hp + factor(gear)	0.8112	34.3033	2.6186	0.4949
mpg ~ wt + hp + factor(carb)	0.7894	17.5995	2.7659	0.9084
mpg ~ wt + hp + disp	0.8083	44.5655	2.6389	0.9285

Analysis Of Transmission

For the transmission variable, a value of 0 means automatic and 1 means manual. The intercept captures the mpg for automatic transmission. The factor(am)1 coefficient captures manual transmission marginal impact.See the Executive Summary for remaining analysis.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	34.0029	2.6427	12.8669	0.0000
wt	-2.8786	0.9050	-3.1808	0.0036
hp	-0.0375	0.0096	-3.9018	0.0005
factor(am)1	2.0837	1.3764	1.5139	0.1413

A plot of the residuals for this model can be found here.