Comparing Mileage by Transmission Type

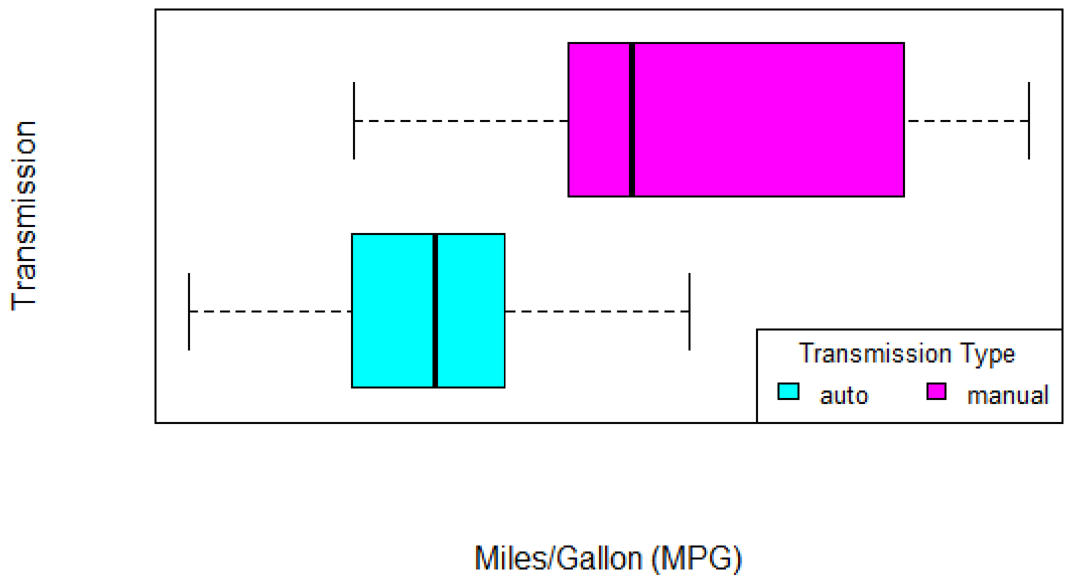
DocOfi

October 16, 2015

Executive Summary

There is an 11.6% increase in mileage (mpg) for cars with manual transmission compared to those with automatic transmission (p = 0.009; t value = 2.83, 95%CI 3.17 to 19.98), while holding weight (wt) and number of cylinders(cyl) constant. There is 2.4% decrease in mileage for cars with automatic transmission for every 1 unit increase in weight (wt), while holding the number of cylinders and type of transmission constant. There is 2.7% decrease in mileage for cars with 6 cylinders and 4.8 % decrease for cars with 8 cylinders compared to 4 cylinder cars while holding weight and type of transmission constant. There was a significant p value (0.007) for the interaction between weight and transmission suggesting that relationship between miles per gallon and weight varies by type of transmission.

A brief description of the data



For this analysis we will be using the dataset mtcars, which is included with every standard installation of R. The data comprises fuel consumption and 10 aspects (number of cylinders (cyl), engine displacement (mpg), gross horsepower (hp), rear axle ratio (drat), weight (wt), quarter mile time (qsec), type of transmission (am), number of forward gears (drat), and number of carburetors (carb)) of automobile design and performance for 32 automobiles (1973-74 models). We will be zeroing particularly on the problem: "Which type of transmission (automatic or manual) produces better mileage (more miles per gallon, MPG)".

Mileage (Miles/Gallon or MPG) by Transmission Type

Shapiro.Wilk.normality.test	Automatic	Manual
statistic.w	0.98	0.95
p.value	0.89	0.54

Assuming normality of our data (shapiro.wilk's normality test of 0.9 and 0.54) and that random sampling was performed, the difference in the average mileage between cars with automatic and manual transmission is significant with a p-value of 0.0014 (t-stat= 3.77, 95% CI = -11.280194 -3.209684).

	t	deg.f	p.val	low.CI	upp.CI	auto	manual
t.test_mpg~am	-3.77	18.33	0	-11.28	-3.21	17.15	24.39

Fitting our Model

Among the continuous variables, the most correlated to mileage is weight (wt) and we use that as our initial predictor together with the type of transmission (factor variable am).

	mpg	disp	hp	drat	wt	qsec
mpg	1	-0.8475514	-0.7761684	0.6811719	-0.8676594	0.418684

Fitting the other variables in our models resulted in the model *mpg ~ wt* factor(am) + factor(cyl)* with the best fit (*R squared value = 0.877*) while maintaining a significant p value in all the coefficients and a confidence interval that does not include 0.

	Estimate	Std.Error	t.value	P.Value	2.5 %	97.5 %
(Intercept)	29.775	10.483	10.482836	0.000	23.936	35.613
wt	-2.399	-2.842	-2.842116	0.009	-4.134	-0.664
factor(am)1	11.569	2.830	2.830083	0.009	3.166	19.971
factor(cyl)6	-2.710	-1.996	-1.996371	0.056	-5.500	0.080
factor(cyl)8	-4.776	-3.070	-3.069814	0.005	-7.974	-1.578
wt:factor(am)1	-4.068	-2.911	-2.911075	0.007	-6.940	-1.196

Regression Diagnostics

The plots referred to in this section may be viewed in the Appendix section The sum of our residuals is -3.330669110^{-16}. The points on the plot of the Residuals vs. Fitted values are randomly scattered and the non-constant Variance Score Test is not significant (0.113) which suggest that the error variance does not changes with the level of the fitted values (test for heteroscedasticity).

The Q-Q plot and the shapiro.test for normality show a normal distribution of the residuals (p = 0.103). The Scale-Location plot and the Reidual vs Leverage plot identified 3 points of interest which depart from the cluster of data points. We further examine these datapoints for there influence in our model using the function influence. measure.

Using the outliertest function, we have identified Fiat 128 as an outlier.

In the interest of reproducible research, codes for this analysis are available upon request

Appendix

fig.1 Plotting mileage (mpg) vs weight by transmission (am)

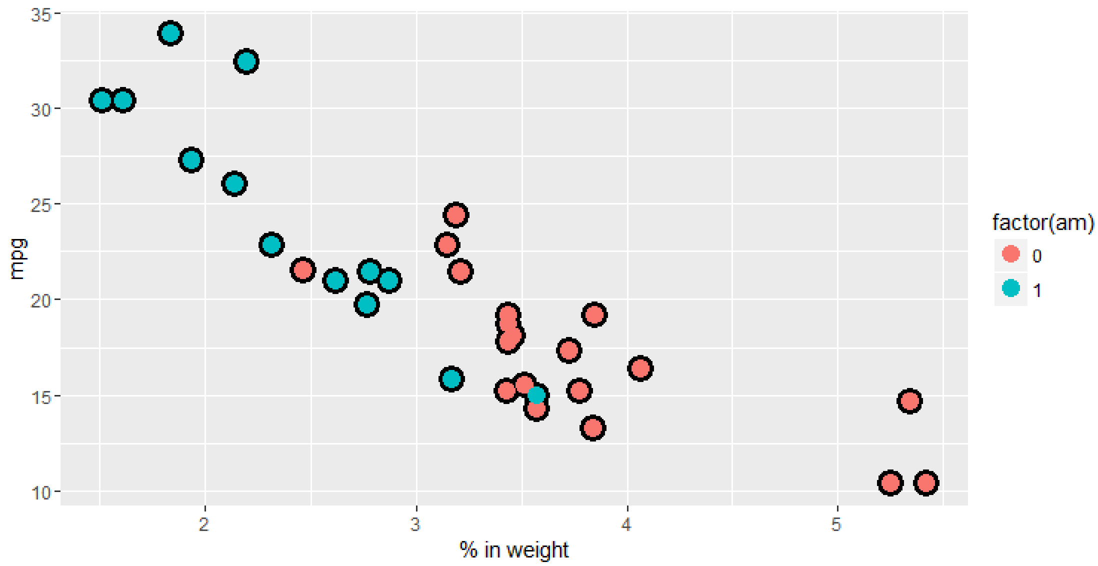


fig.2 plotting mpg and weight by transmission type with our regression line mpg ~ wt

1 line, 1 intercept, 1 slope

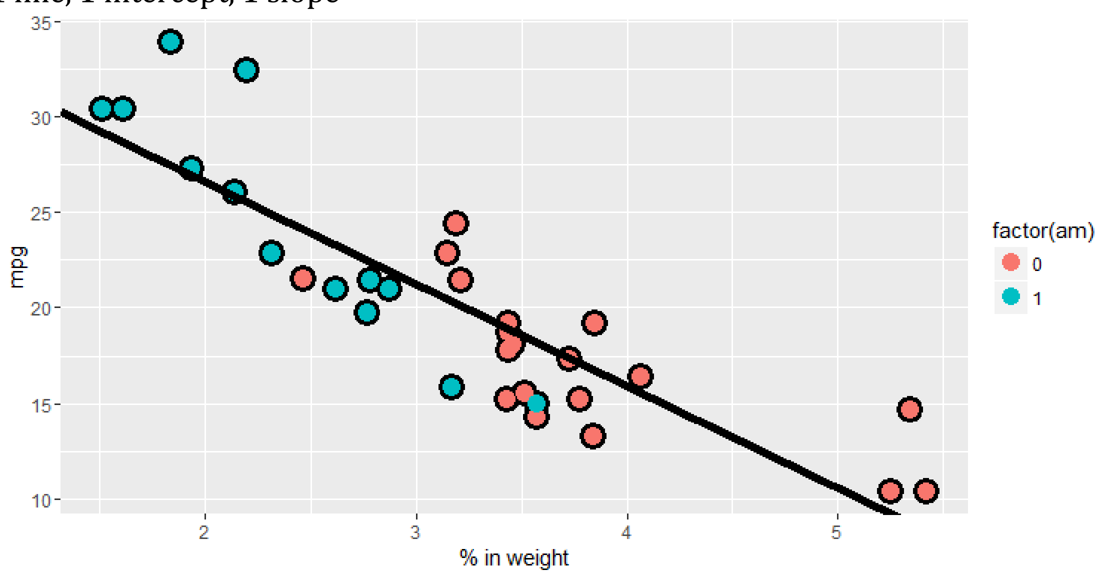


fig.3 plotting mpg and weight by transmission type with our regression line mpg ~ wt + factor(am)

2 lines, 2 intercepts, 1 slope *the lines are very close to each other*

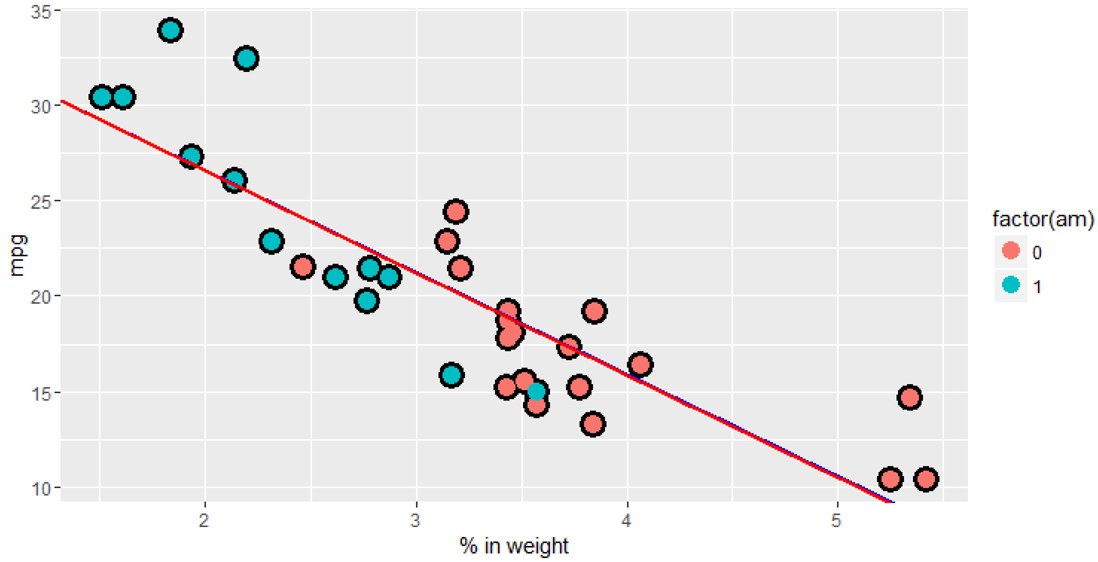


fig.4 plotting mpg and weight by transmission type with our regression line mpg ~ wt * factor(am)

2 lines, 2 intercepts, 2 slopes. Interaction between weight and transmission

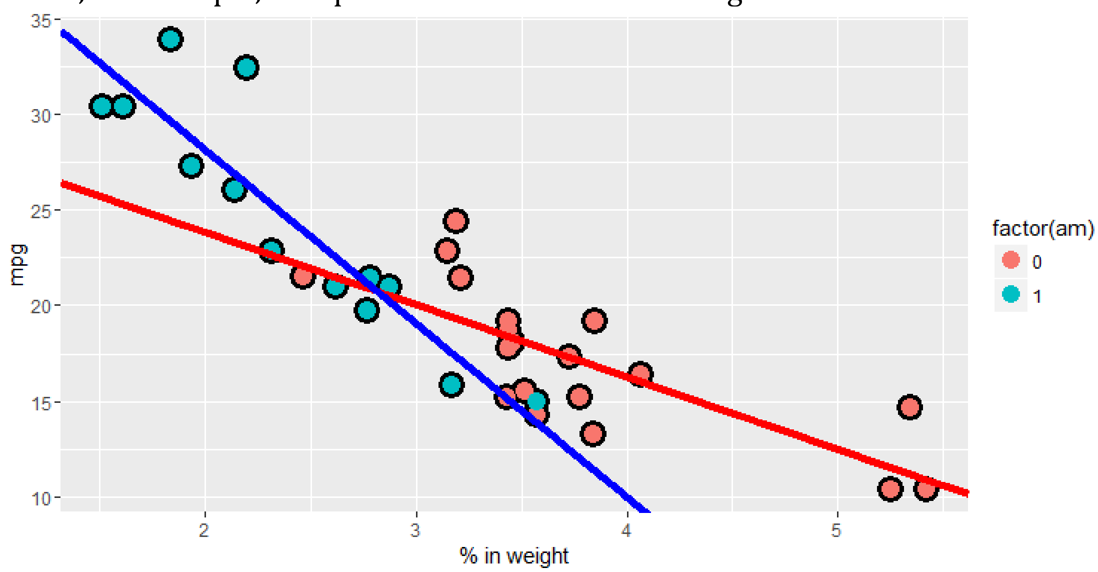


fig.5 plotting mpg and weight by transmission type with our regression line mpg ~ wt * factor(am)

2 lines, 2 intercepts, 2 slopes. Interaction between weight and transmission and adjustments for number of cylinders.

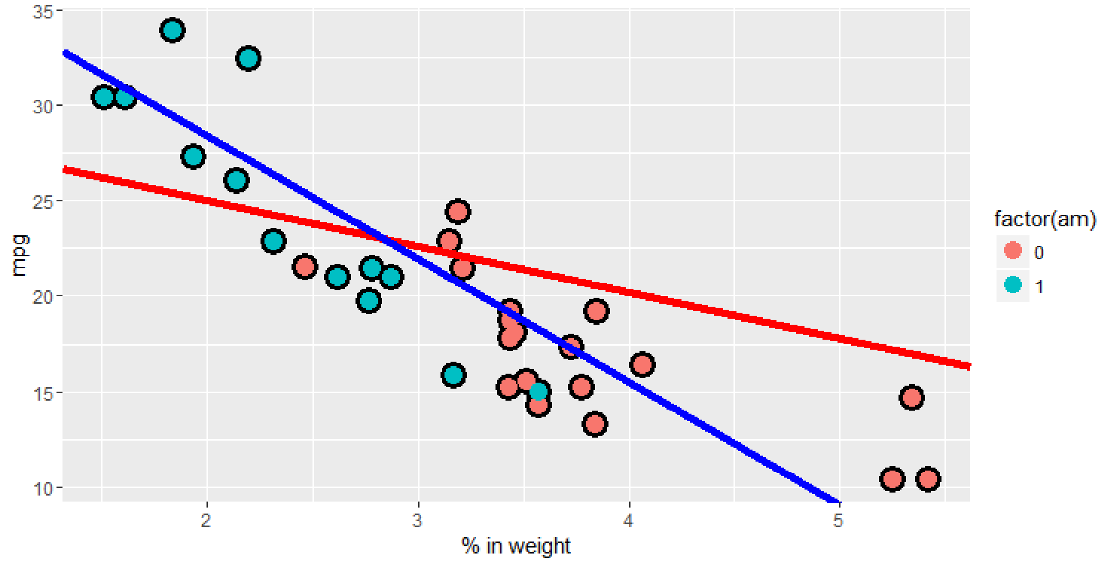


fig.6-10 plot of residuals