Motor Trend MPG Analysis

Mikhail Romadanovskiy

25 Jul 2015

Executive summary

In this report, we will analyze `mtcars` data set and explore the relationship between a set of available variables and miles per gallon (MPG). The data was extracted from the 1974 **Motor Trend** US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). There are 2 main questions to consider: what type of transmission is better for mpg, and how it can be discribed in numeric terms. The t-test shows that mpg performance difference between cars with automatic and manual transmission. And it is about 7 MPG more for cars with manual transmission than those with automatic transmission. We fit several linear regression models and select the one with highest Adjusted R-squared value. Given model shows that cars that are lighter in weight with a manual transmission and cars that are heavier in weight with an automatic transmission will have higher MPG values.

Contex

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

- Is an automatic or manual transmission better for MPG
- Quantify the MPG difference between automatic and manual transmissions

Since p-Value is pVal the automatic and manual transmissions are from different populations. And the mean for MPG of manual transmitted cars is about result$estimate[[2]] - result$estimate[[1]] more than that of automatic transmitted cars.

Regression Model

Testing simple model.

Simple model can explain **34%** of variance of MPG. Adjusted R-squared value indicates that we need to add other variables to the model to make better predictions.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.1474	1.1246	15.25	0.0000
am1	7.2449	1.7644	4.11	0.0003

Finding proper set of explanation variables

Let's use all the availible data to make a regression model.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	23.8791	20.0658	1.19	0.2525
cyl6	-2.6487	3.0409	-0.87	0.3975
cyl8	-0.3362	7.1595	-0.05	0.9632
disp	0.0355	0.0319	1.11	0.2827
hp	-0.0705	0.0394	-1.79	0.0939
drat	1.1828	2.4835	0.48	0.6407
wt	-4.5298	2.5387	-1.78	0.0946
qsec	0.3678	0.9354	0.39	0.6997
vs1	1.9309	2.8713	0.67	0.5115
am1	1.2121	3.2135	0.38	0.7113
gear4	1.1144	3.7995	0.29	0.7733
gear5	2.5284	3.7364	0.68	0.5089
carb2	-0.9794	2.3180	-0.42	0.6787
carb3	2.9996	4.2935	0.70	0.4955
carb4	1.0914	4.4496	0.25	0.8096
carb6	4.4776	6.3841	0.70	0.4938
carb8	7.2504	8.3606	0.87	0.3995

Full regression model can explain **78%** of MPG variance, but none of the coeficients are significant. We could find more effective set of variables to explain mpg.

```
stepModel <- step(fullModel, k = log(nrow(mtcars)))
xtable(summary(stepModel))  # results='hide'
mpgModel <- lm(mpg ~ wt + qsec + am, data = mtcars)
```

Linear model mpg ~ wt + qsec + am - 1 explains **83%** of MPG variance, but consists only significant at 5% s.l. coefficients.

```
xtable(anova(mpgModel, simpleMpgModel, fullModel))
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	28	169.29				
2	30	720.90	-2	-551.61	34.36	0.0000
3	15	120.40	15	600.49	4.99	0.0018

```r
xtable(confint(mpgModel))
```

	2.5 %	97.5 %
(Intercept)	-4.64	23.87
wt	-5.37	-2.46
qsec	0.63	1.82
am1	0.05	5.83

```r
amIntWtModel <- lm(mpg ~ wt + qsec + am + wt:am,
    data = mtcars)
xtable(summary(amIntWtModel))  # results hidden
```

	Estimate	Std. Error	t value	Pr($>\lvert t\rvert$)
(Intercept)	9.7231	5.8990	1.65	0.1109
wt	-2.9365	0.6660	-4.41	0.0001
qsec	1.0170	0.2520	4.04	0.0004
am1	14.0794	3.4353	4.10	0.0003
wt:am1	-4.1414	1.1968	-3.46	0.0018

The measure of how much an observation has effected the estimate of a regression coefficient, we get sum((abs(dfbetas(mpgModel)))>1) Therefore, the above analyses meet all basic assumptions of linear regression and well answer the questions.

Appendix: Figures

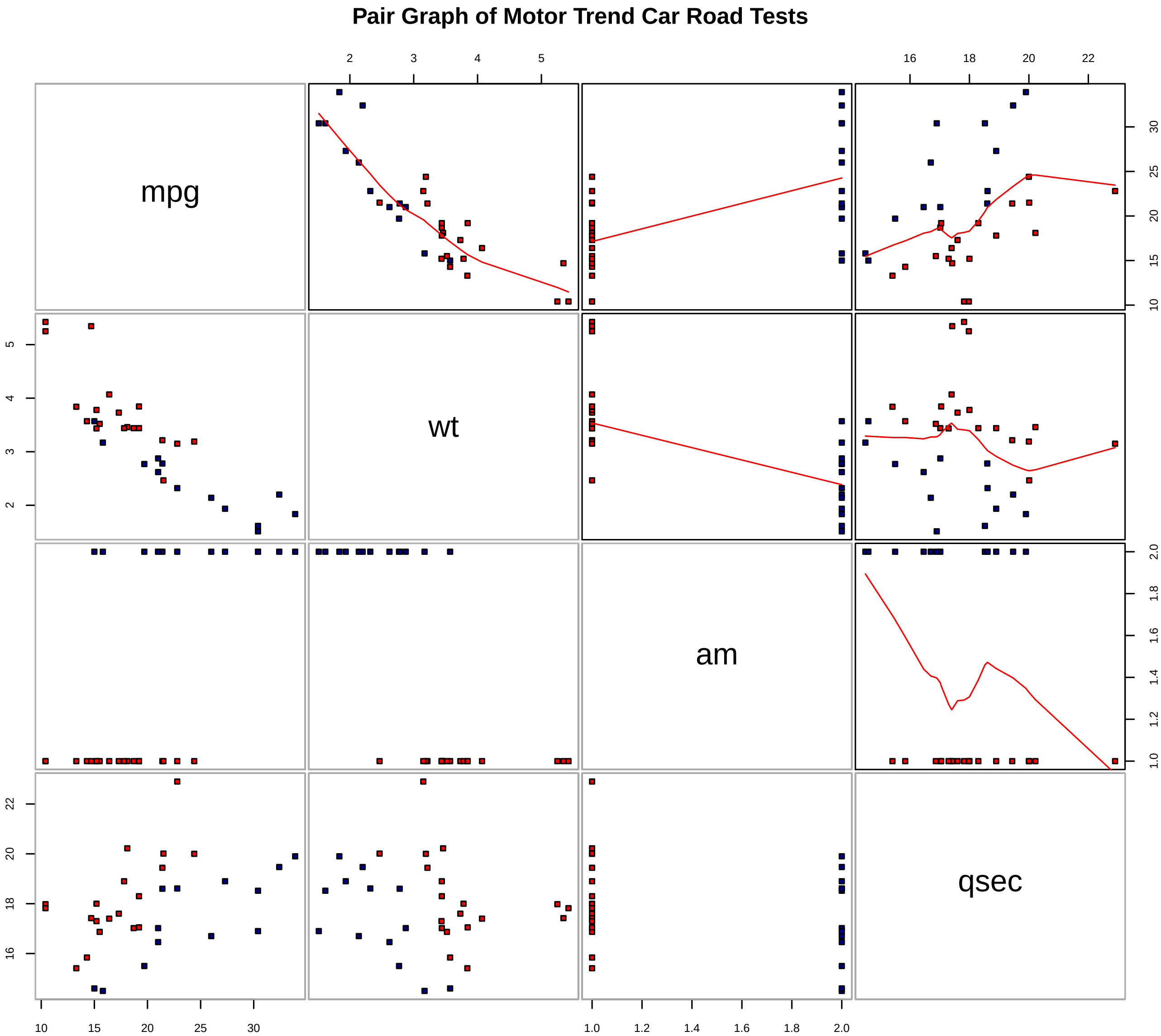

Figure 1: Selected Dimensions Pair Graph

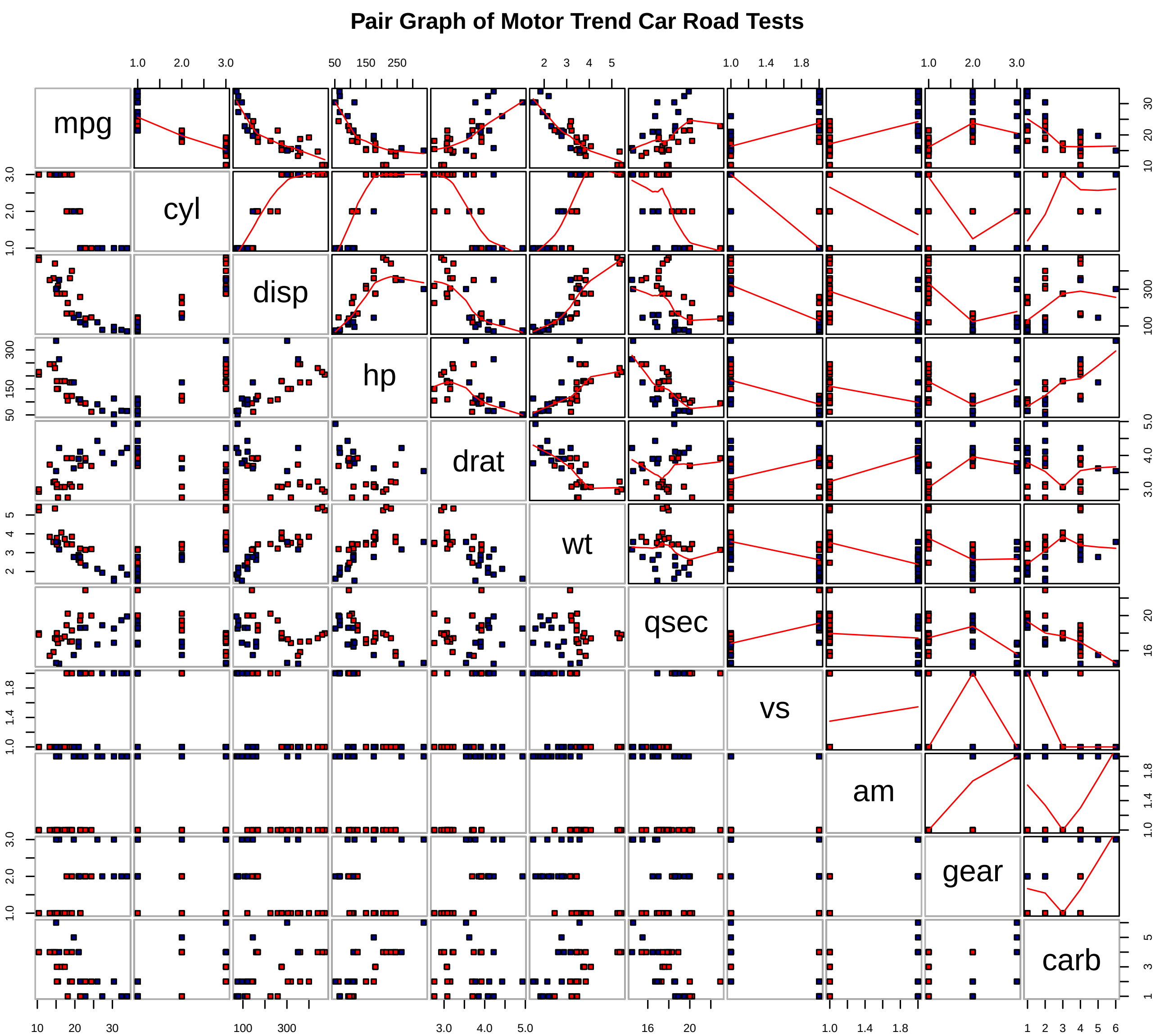

Figure 2: Full Pair Graph

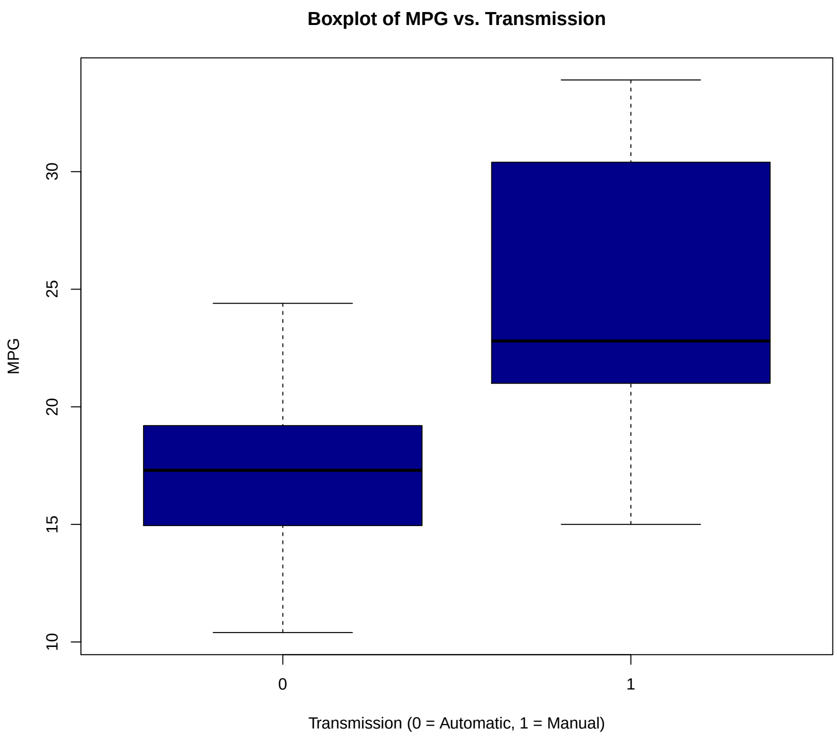

Figure 3: Box Plot

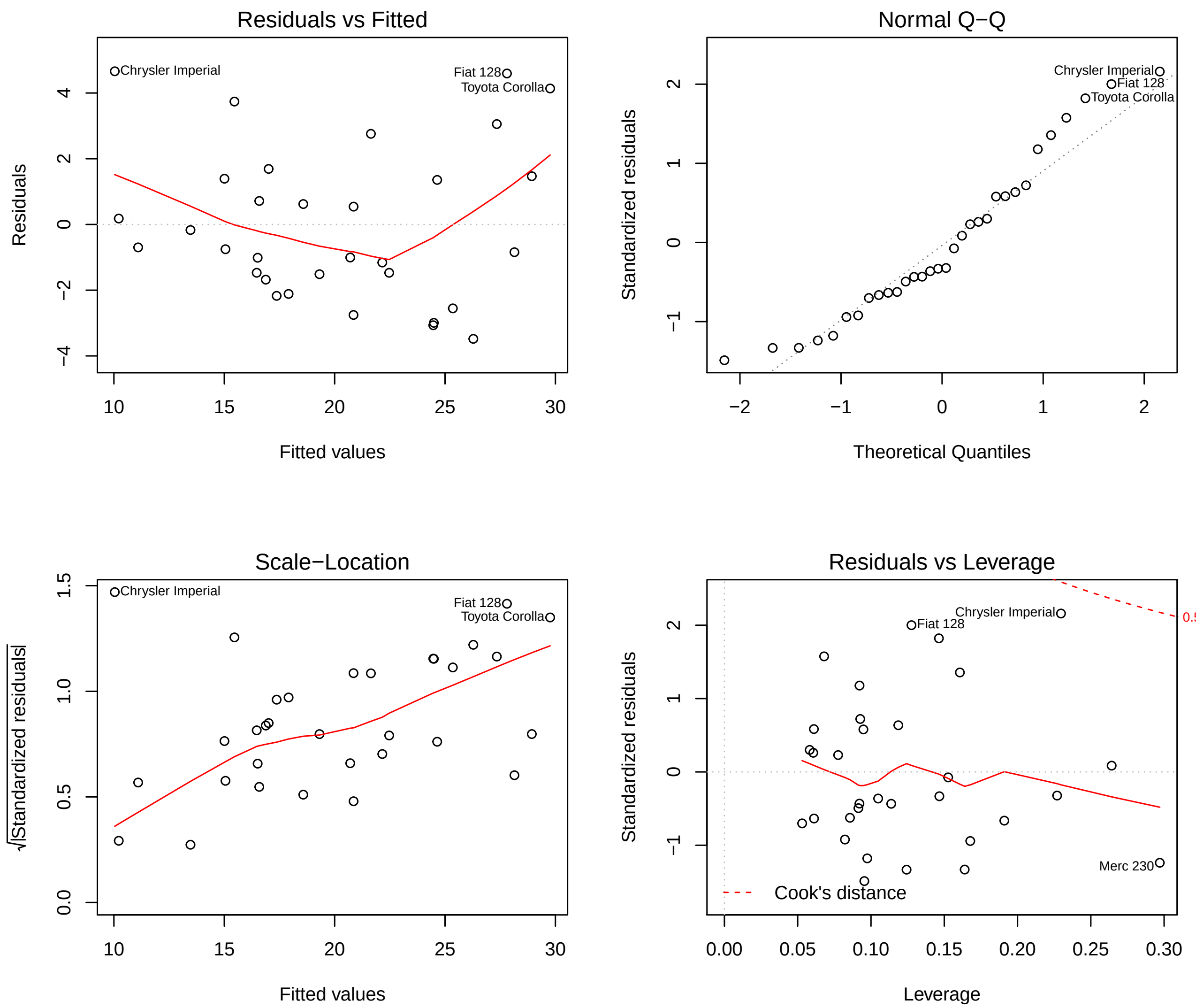

Figure 4: Residuals Analysis