Is automatic or manual transmission better for MPG?

Chris Lill

Executive Summary

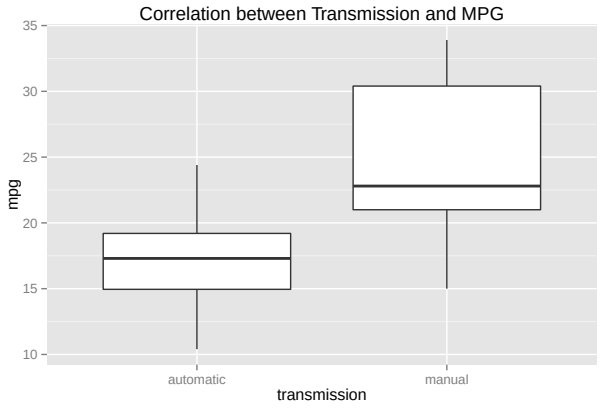
Regression modelling of the mtcars dataset from 1974 shows that weight and horsepower have the greatest impact on miles per gallon (MPG). Holding these constant, there is a 95% probability that the effect of the transmission is between -2.8 and 0.9 MPG. There is no significant difference between MPG for an automatic and manual transmission.

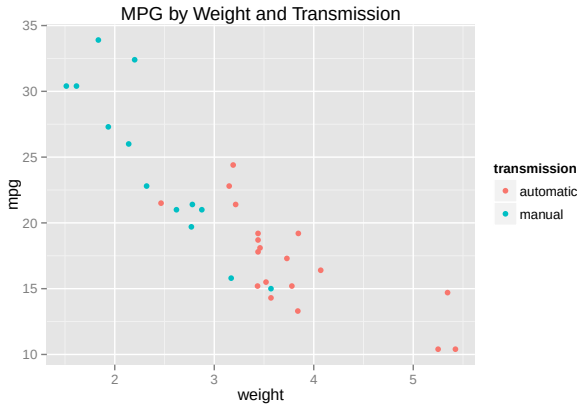
Exploratory Data Analysis

The boxplot below shows that there is a correlation between transmission and miles per gallon. However, this result may be explained by other variables so we need to determine which have the most significant correlation.

The scatterplot below shows a good correlation between weight(wt) and mpg. It also shows that weight and transmission are correlated, so automatic gearboxes tend to be used on heavier cars.

The table in appendix 1 shows that in a linear regression model of mpg containing all other factors, weight is the most significant. There is only a 6% probability that it has no effect on mpg.





Model Selection

To select the most appropriate model, we will select `wt` as the first factor and add other factors incrementally. We'll build a new linear regression model each time and use Anova to see whether the new factor is a significant improvement.

The table in appendix 2 shows that the models with the most significant additional factors are `mpg ~ wt + hp` and `mpg ~ wt * hp`. We will select the `mpg ~ wt + hp` model so that we avoid overfitting on a dataset that only has 32 observations.

The residal diagnostic plots in appendix 3 show that `mpg ~ wt + hp` is a reasonable model given the small dataset. Possible issues with the model are that the Q-Q plot shows that the residuals are not normally distributed, and the plot of residuals vs leverage shows some outliers with high leverage.

Results

Is an automatic or manual transmission better for MPG

To visualise the difference between an automatic and manual transmission we can plot the residuals from our model against transmission(am). The boxplot in appendix 4 shows that there is a sizeable overlap between the automatic and manual data, when holding weight and horsepower constant. This suggests that there is no significant difference between the two groups.

For a more rigorous approach we can test the hypothesis that there is a difference between the means of MPG for automatic and manual cars. The null hypothesis is that there is no difference. This can be calculated with a two sided t-Test assuming equal variance.

The results show that the 95% confidence interval for the difference is from -2.8 to 0.9, with a p-value of 0.32. Therefore we fail to reject the null hypothesis, and conclude that there is no significant difference in MPG between cars with an automatic and manual transmission, holding weight and horsepower constant.

Note that this calculation is slightly conservative because it assumes 30 degrees of freedom. The correct calculation would have 28 degrees of freedom, calculated as 32 observations less 2 for the model and 2 for the t-test. Inspection of old-fashioned t-statistic tables shows that the t-statistic is 2.048 for 28df, compared with 2.042 for 30df. Therefore this error is less than 1% and can be ignored.

```
fit.wt.hp <- lm(mpg ~ wt + hp, data = mtcars)
t <- t.test(resid(fit.wt.hp) ~ mtcars$am, var.equal = TRUE)
t
```

```
##
##  Two Sample t-test
##
## data:  resid(fit.wt.hp) by mtcars$am
## t = -1.0168, df = 30, p-value = 0.3174
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.760359  0.925366
## sample estimates:
## mean in group automatic    mean in group manual
##              -0.3727330               0.5447636
```

Quantify the MPG difference between automatic and manual transmissions

Assuming our model of `mpg ~ wt + hp`, the difference between the mean of the residuals for the groups is -0.9 mpg, with automatic transmissions showing as marginally more efficient. Given the small sample size and the confidence interval above, this is not a significant difference.

```
t$estimate["mean in group automatic"] - t$estimate["mean in group manual"]
```

```
## mean in group automatic
##              -0.9174966
```

Appendix

Appendix 1: Regression model for mpg with all factors

Table 1: Coefficients and their significance for each factor

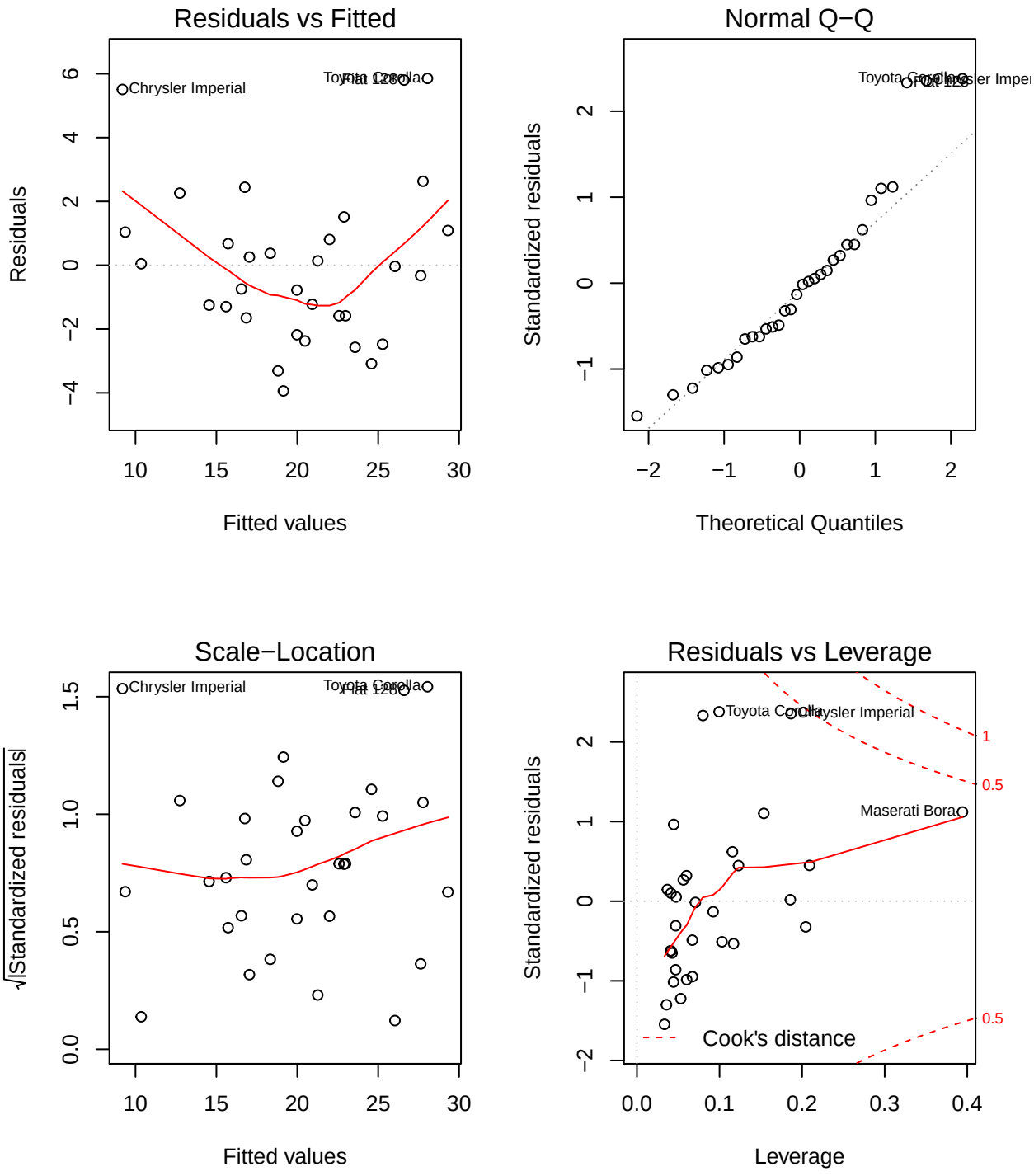
	Coefficient	P-value
wt	-3.72	0.06
am	2.52	0.23
qsec	0.82	0.27
hp	-0.02	0.33
disp	0.01	0.46
(Intercept)	12.30	0.52
drat	0.79	0.64
gear	0.66	0.67
carb	-0.20	0.81
vs	0.32	0.88
cyl	-0.11	0.92

Appendix 2: Comparison of different regression models for mpg

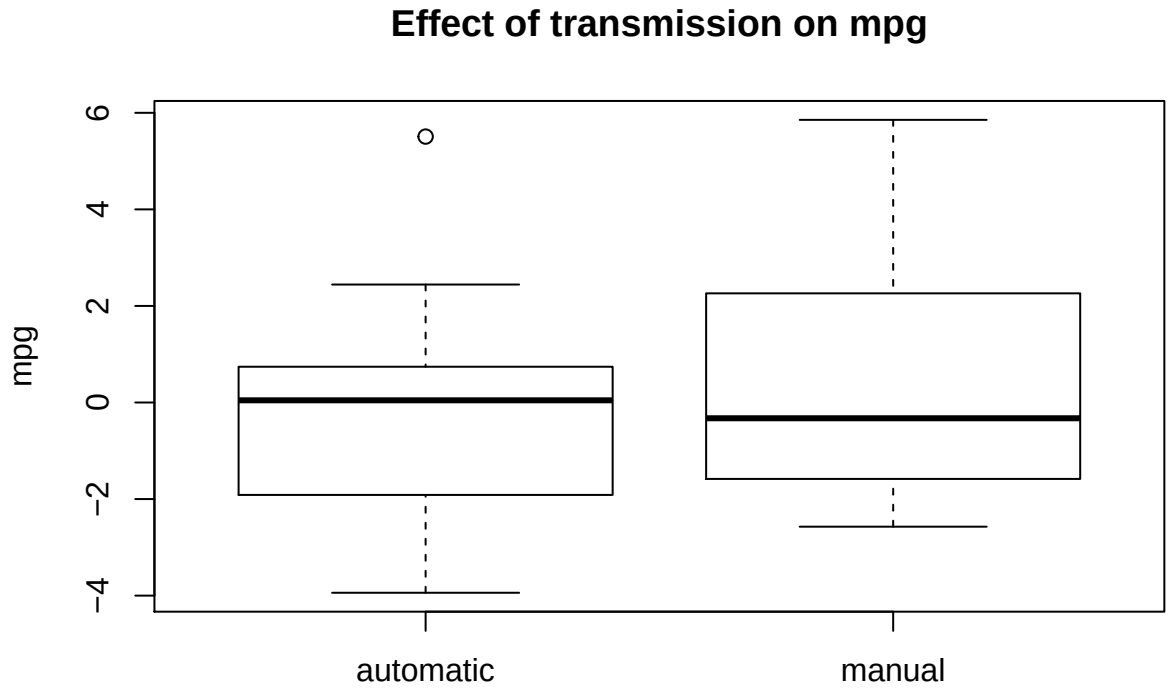
Table 2: Significance of additional factor in each model

Possible model	Compared with	P-value
mpg ~ wt + am	mpg ~ wt	0.988
mpg ~ wt + factor(cyl)	mpg ~ wt	0.003
mpg ~ wt + disp	mpg ~ wt	0.064
mpg ~ wt + hp	mpg ~ wt	0.001
mpg ~ wt * hp	mpg ~ wt + hp	0.001
mpg ~ wt + hp + am	mpg ~ wt + hp	0.141
mpg ~ wt + hp + factor(cyl)	mpg ~ wt + hp	0.074
mpg ~ wt + hp + disp	mpg ~ wt + hp	0.929
mpg ~ wt + hp + drat	mpg ~ wt + hp	0.199
mpg ~ wt + hp + qsec	mpg ~ wt + hp	0.255
mpg ~ wt + hp + factor(vs)	mpg ~ wt + hp	0.321
mpg ~ wt + hp + factor(gear)	mpg ~ wt + hp	0.495
mpg ~ wt + hp + factor(carb)	mpg ~ wt + hp	0.908

Appendix 3: Residual diagnostics for mpg ~ wt + hp



Appendix 4: Effect of transmission on mpg, holding weight and horsepower constant



Holding weight and horsepower constant