Quantifying the Effect of Transmission Type on Gas Mileage Using Linear Regression

Connor Claypool

29 June 2018

Executive Summary

This analysis used data from the 1974 *Motor Trend* magazine and a linear regression model to estimate the effect of transmission type on gas mileage (MPG), while taking into account the effect of other relevant variables. The estimate of the model is that, given a specific 1/4 mile time and weight,

$$\text{manual MPG} = \text{automatic MPG} + 14.1 - 4.1 \times \text{weight}$$

where weight is given in thousands of pounds. According to this rule, manual cars have better MPG at weights lower than about 3400 lbs, while automatic cars perform better at greater weights. Additionally, this difference is greater the farther the weight is from 3400 lbs. However, there is a fair amount of uncertainty in the model's estimate - the margin of error for the number 14.1 is around 7, and the margin of error for the number 4.1 is around 2.5. This means that the relationship as estimated by the model is far from definite.

Analysis

The difference between the distributions of MPG values based on transmission type is clearly demonstrated in a box plot of this data (see Appendix). However, this plot conveys no information as to the effect of confounding variables on this difference. To determine the influence transmission type has after adjusting for the impact of other relevant variables, a linear regression model was fit using transmission type, weight and quarter-mile time as predictors of MPG. These features were selected using a backwards elimination model selection strategy in which insignificant predictors are iteratively removed. The coefficients of this model and their corresponding p-values are shown in the table below.

```
library(xtable)
model <- lm(mpg ~ am + wt + qsec, mtcars)
print(xtable(model, digits = 4, caption = "Model Coefficients"), comment = F)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.6178	6.9596	1.3819	0.1779
am	2.9358	1.4109	2.0808	0.0467
wt	-3.9165	0.7112	-5.5069	0.0000
qsec	1.2259	0.2887	4.2467	0.0002

Table 1: Model Coefficients

The fitted model has an adjusted R-squared of 0.8336, indicating a fairly high goodness of fit. However, this model can be improved upon by including the interaction of transmission type and weight. The coefficients of this model are shown in the table below.

```
model <- lm(mpg ~ am * wt + qsec, mtcars)
print(xtable(model, digits = 4, caption = "Model Coefficients"), comment = F)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.7231	5.8990	1.6482	0.1109
am	14.0794	3.4353	4.0985	0.0003
wt	-2.9365	0.6660	-4.4090	0.0001
qsec	1.0170	0.2520	4.0354	0.0004
am:wt	-4.1414	1.1968	-3.4603	0.0018

Table 2: Model Coefficients

This model has a better adjusted R-squared of 0.8804. Furthermore, the table below, which summarizes the distribution of hat values for this model, shows that this model is not distorted by any extreme outliers.

```
hatvalue_summary <- t(as.matrix(summary(hatvalues(model))))
print(xtable(hatvalue_summary, caption = "Distribution of Hat Values"),
      comment = F, include.rownames = F)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.05	0.09	0.13	0.16	0.22	0.37

Table 3: Distribution of Hat Values

Finally, plotting the residuals against the fitted values and the theoretical normal quantiles (see Appendix) shows that they do not follow any clear pattern and are approximately normally distributed, in line with the assumptions of linear regression.

Results

According to the coefficients estimated by the fitted model, the difference in mean MPG based on transmission type, for a specific weight and quarter mile time, can be represented by the equation

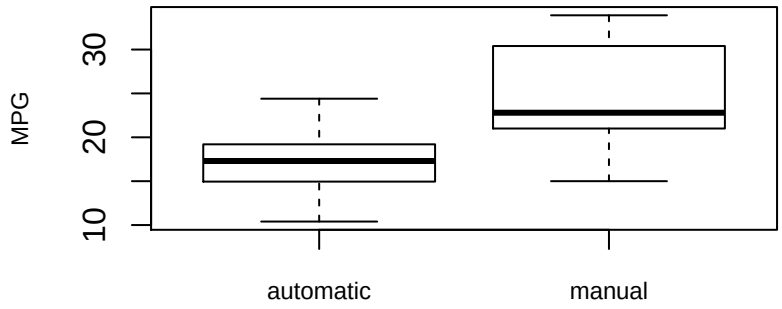
$$\text{MPG}_{manual} = \text{MPG}_{auto} + 14.0794 - 4.1414 \times \text{weight}$$

This equation indicates that manual transmissions are associated with higher MPG at weights lower than approximately 3400 lbs, with the difference becoming less pronounced the closer the weight is to this value. At greater weights, automatic transmissions are associated with higher MPG, with this difference becoming larger as weight increases. However, the 95% confidence intervals for the two model coefficients used in the above equation are wide; the first is between 7.0309 and 21.128, and the second is between -6.597 and -1.6857. This means there is a significant amount of uncertainty in this estimated relationship between transmission type and gas mileage.

Appendix

```
data(mtcars)
boxplot(mpg ~ am, mtcars, xaxt = "n", ylab = "MPG", cex = 0.75, cex.lab = 0.75)
axis(side = 1, at = c(1, 2), labels = c("automatic", "manual"), cex.axis = 0.75)
title(main = "Figure 1: MPG by Transmission Type",
      line = 0.75, cex.main = 0.75, font.main = 1)
```

Figure 1: MPG by Transmission Type




```
par(mfrow=c(1, 2))
plot(model, which=c(1,2), caption = c("Figure 2: Residuals vs Fitted",
                                      "Figure 3: Normal Q-Q"),
     font.main = 1, cex.lab = 0.75, cex.caption = 0.75)
```

Figure 2: Residuals vs Fitted

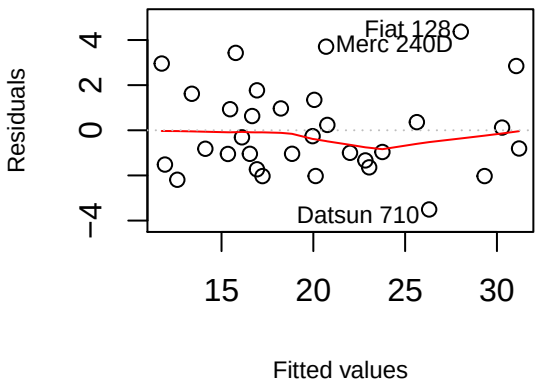


Figure 3: Normal Q−Q