Regression Models Course Project

Hiro Miyake

October 16, 2016

Executive Summary

In the simplest model where I ignore all other variables, manual transmission has greater miles per gallon (mpg) than automatic transmission by a factor of 1.414307 with a standard error range of (1.294274, 1.545473). On average this corresponds to an increase of 6.927806 mpg. These values were obtained with a log-linear regression, which appears to have reduced heteroskedasticity compared to a simple linear regression. However, for any purchasing decision of a particular vehicle for a desired mpg, it would most likely be prudent to consider all characteristics of a car and not just the transmission type.

Exploratory Analysis

First, I explore the relationship between miles per gallon (mpg) and transmission type (automatic or manual) ignoring all other factors. This is shown in figure 1 of the appendix by the blue circles. This plot suggests a positive relationship between the two parameters.

Models

Linear Regression: I fit a linear model for the relationship between mpg and transmission type, where the transmission type is treated as a factor variable. This can be mathematically written as mpg $= a \times$ transmission_type $+ b$, where a and b are fitting parameters. The fit is shown as the red line in figure 1. This model suggests manual transmission increases the mpg by 7.245 ± 1.764 mpg, where the uncertainty is the standard error. Given that the p-value is very small, i.e., 0.000285, it is highly likely the relationship is true.

```
fit1 = lm(mpg ~ factor(am), data = mtcars)
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ factor(am), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## factor(am)1    7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Log-Linear Regression: I also consider a model where I take the logarithm of the mpg and consider a linear relationship with transmission type. In this case the mathematical expression is $\log(\text{mpg}) = a \times \text{transmission\_type} + b$, where a and b are fitting parameters. This is shown in figure 3. In this case, the mpg of manual transmission is larger than the mpg of automatic transmission by a factor of $e^{0.34664} = 1.414307$, with lower and upper standard error range given by $(e^{0.25795}, e^{0.43533})$=(1.294274, 1.545473). The p-value is 0.0004905, so the relationship is again highly likely to truly exist.

```
fit2 = lm(I(log(mpg)) ~ factor(am), data = mtcars)
summary(fit2)
```

```
##
## Call:
## lm(formula = I(log(mpg)) ~ factor(am), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4749 -0.1213  0.0073  0.1693  0.3779
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.81669    0.05653  49.830  < 2e-16 ***
## factor(am)1  0.34664    0.08869   3.909 0.000491 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2464 on 30 degrees of freedom
## Multiple R-squared:  0.3374, Adjusted R-squared:  0.3153
## F-statistic: 15.28 on 1 and 30 DF,  p-value: 0.0004905
```

Both models suggest a positive relationship between mpg and transmission type. I will determine which is more appropriate by looking at the residuals in the next section.

Residual Analysis

I plot the residuals for the simplest linear model in figure 2. It appears that the residuals may be larger for manual transmission, indicating that there may be heteroskedasticity in the fit. This means that the variance of the data may be different depending on transmission type. This reduces our confidence in deciding that manual transmission indeed has better mpg than automatic transmission from the simple linear regression. However, the residuals for the log-linear model, shown in figure 4, appears to not show heteroskedasticity, which means that we can have more confidence in the results from the log-linear regression.

Conclusions and Uncertainties

Ignoring all other factors, I conclude that manual transmission leads to a higher mpg than automatic transmission by a factor of 1.414307, with standard error range of (1.294274, 1.545473). On average this corresponds to an increase of 6.927806 mpg. However, it is possible other variables may nullify this relationship. Therefore before deciding if a particular car has the desired mpg, one should consider all available factors, and not just the transmission type, before making a purchasing decision.

Appendix

```
plot(mtcars$am, mtcars$mpg, xlab="Automatic/Manual (0/1)", ylab = "Miles per Gallon",
     pch = 21, bg = "lightblue", cex = 2)
lines(mtcars$am, predict(fit1), col = "red", lwd = 2)
```

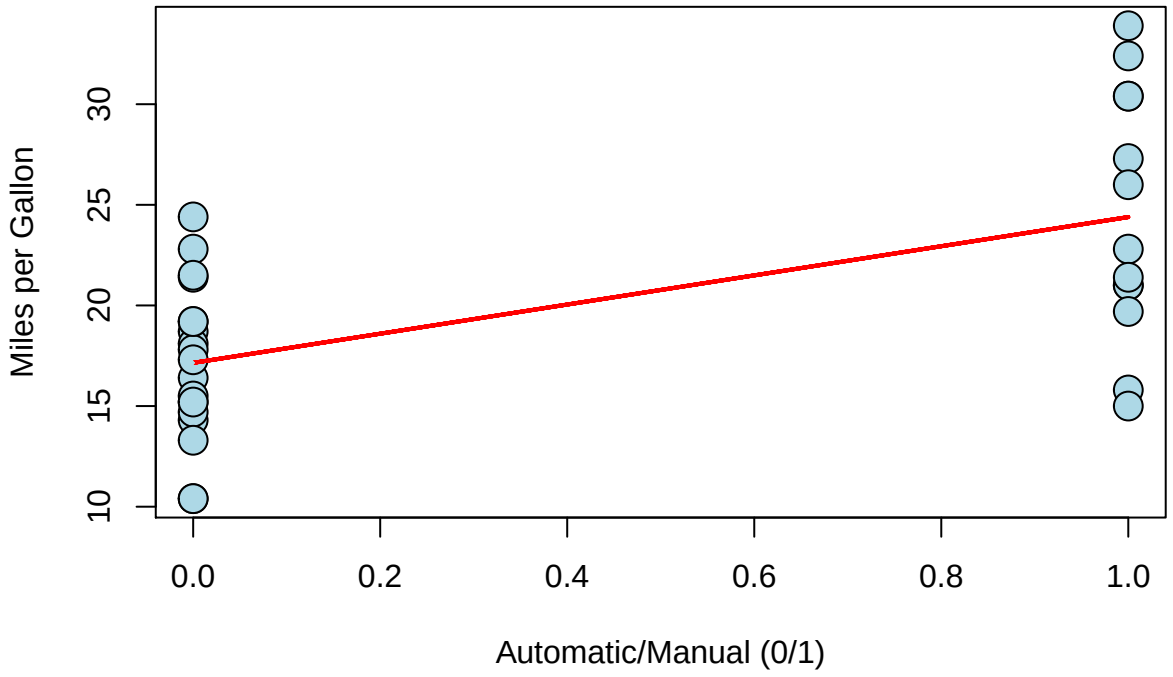

Figure 1: Miles per gallon versus transmission type with a linear model.

```
plot(mtcars$am, resid(fit1), ylab="Residuals", xlab="Automatic/Manual (0/1)",
     pch = 21, bg = "lightblue", cex = 2)
abline(0, 0, col = "red")
```

```
plot(mtcars$am, log(mtcars$mpg), xlab="Automatic/Manual (0/1)", ylab = "log(Miles per Gallon)",
     pch = 21, bg = "lightblue", cex = 2)
lines(mtcars$am, predict(fit2), col = "red", lwd = 2)
```

```
plot(mtcars$am, resid(fit2), ylab="Residuals", xlab="Automatic/Manual (0/1)",
     pch = 21, bg = "lightblue", cex = 2)
abline(0, 0, col = "red")
```

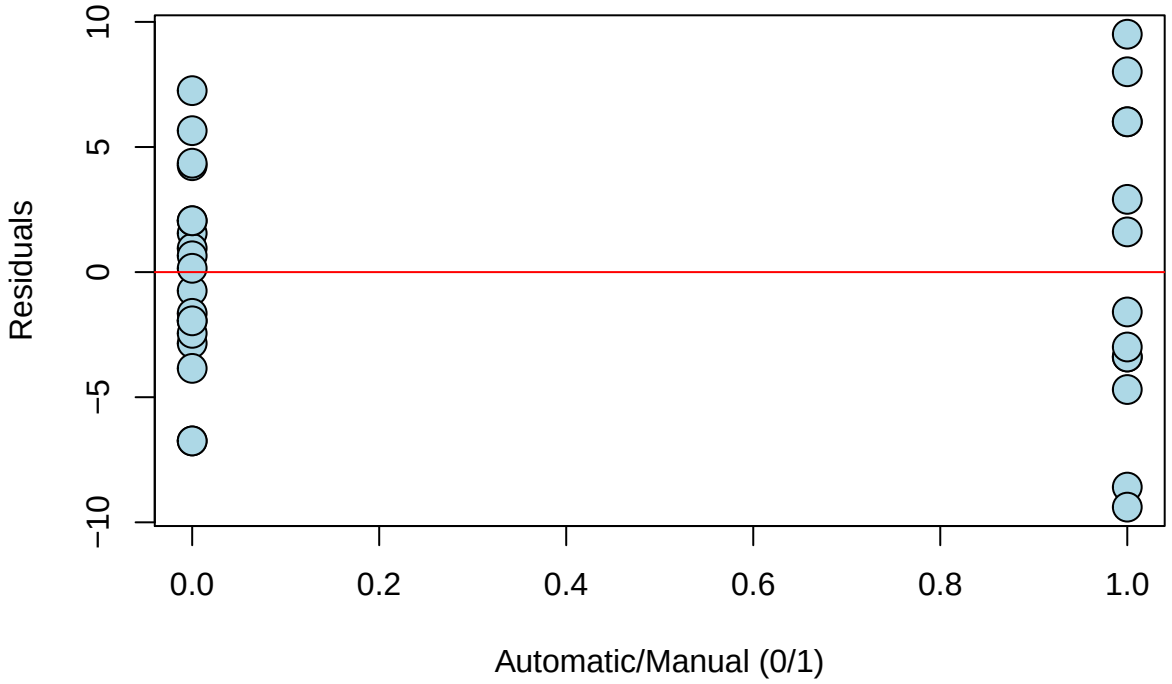

Figure 2: Residual of linear model of mpg versus transmission type.

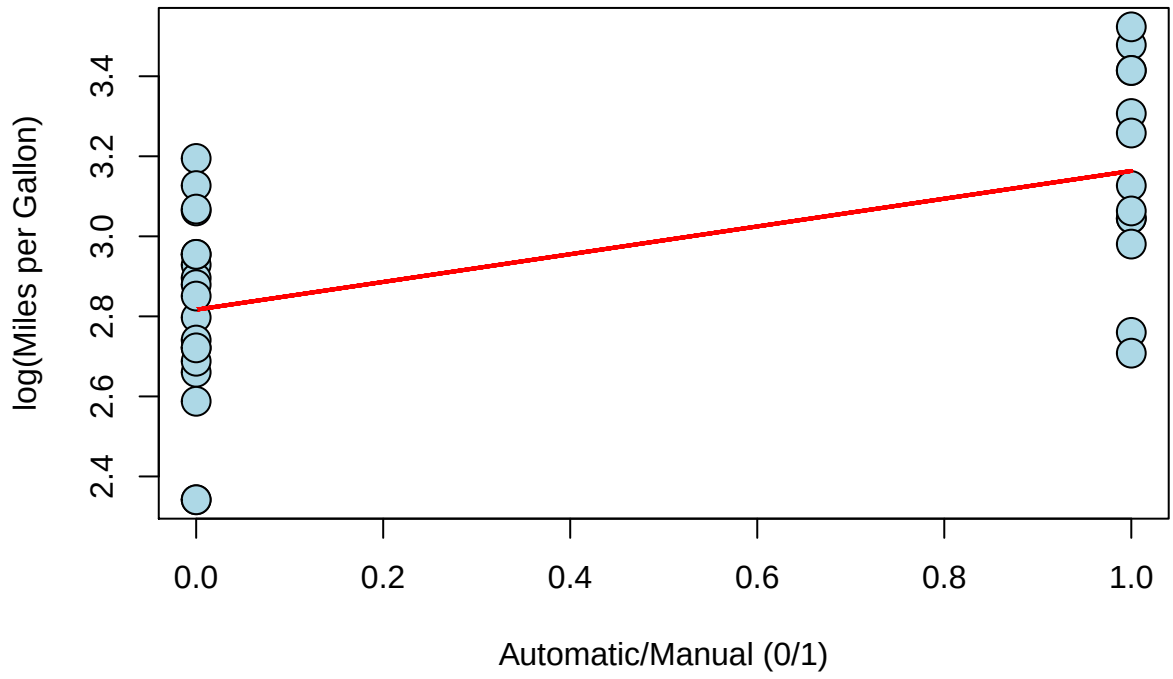

Figure 3: Logarithm of miles per gallon versus transmission type with a linear model.

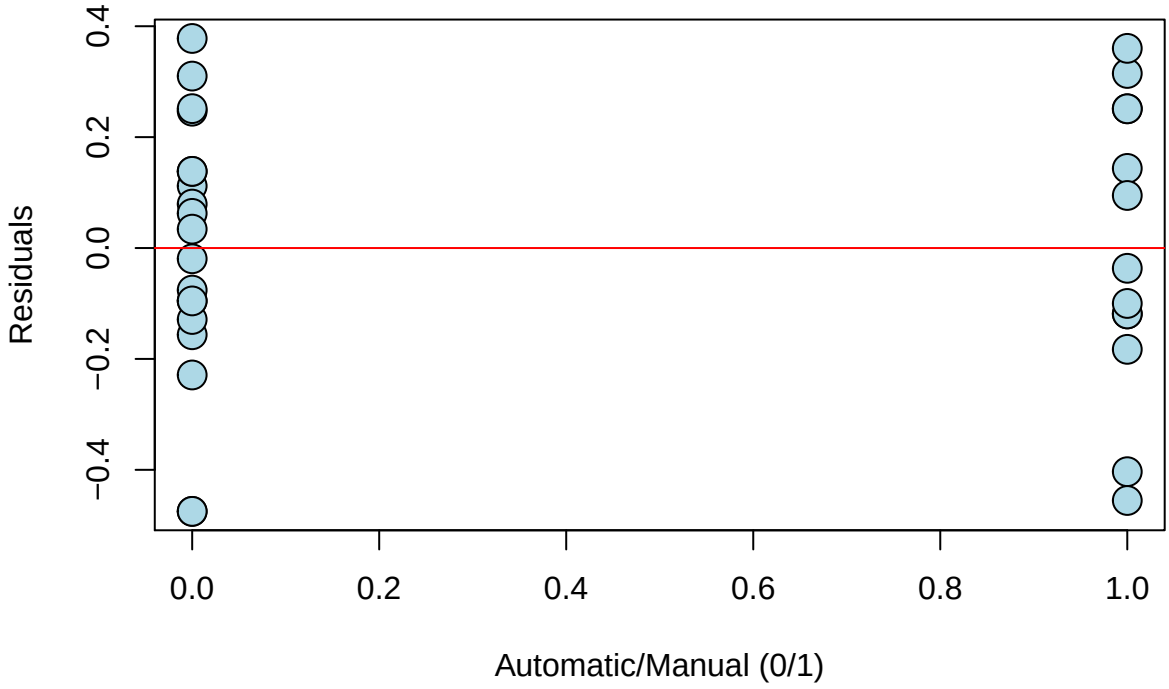

Figure 4: Residual of linear model of log(mpg) versus transmission type.