Miles per Gallon (MPG) – Automatic Transmission vs Manual Transmission

Executive Summary

The purpose of this project is to explore the relationship between a set of variables and miles per gallon (MPG). Specifically, we want to answer the following two questions:

1. Is an automatic or manual transmission better for MPG“?
2. How different is the MPG between automatic and manual transmissions?

This study shows that the answer depends on the car’s weight (**wt**) and the 1/4 mile time (**qsec**), a performance benchmark for car acceleration. For light weighted cars, manual tranmission is better for MPG. However, for heavier cars, automatic transmission is better for MPG. For a given **qsec**, manual transmission is always better for MPG. For the range of data considered in this study, the difference in MPG between autmoatic and manual transmission is between -7 and 7.

Data source

The **mtcars** data set in R’s **datasets** package is used. **mtcars** has 11 variables and 32 observations, in which 19 cars have automatic transmission and 13 cars have manual transmission. It is a small data set. It is important that we don’t overfit the model by using too many predictors. A new data frame **mtcars2** is copied from the original data **mtcars**. **mtcars2** is the same as **mtcars** except that the variables **am** and **vs** are factored.

The first model

Let **Y** denote the car’s MPG, **T** the car’s transmission (1 for automatic; 2 for manual), and **X** the other predictor of the car’s MPG. We can start with the following regression model:

Y = C_0 + C_1 **T***1 + C*2,1 **X**, for cars with automatic transmission

Y = C_0 + C_1 **T***2 + C*2,2 **X**, for cars with manual transmission

Which variable should be used? It is desirable to pick the one that has the largest absolute correlation with MPG.

```
which.max(abs(cor(mtcars)[2:11, 1])) + 1
```

```
## wt
##  6
```

Consequently, the first model can be generated using the following codes:

```
mpg_formula <- mpg ~ am + am:wt
mpg_lm <- lm(mpg_formula, data = mtcars2)
summary(mpg_lm)
```

```
##
## Call:
## lm(formula = mpg_formula, data = mtcars2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.600 -1.545 -0.533  0.901  6.091
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   31.416      3.020   10.40  4.0e-11 ***
## am1           14.878      4.264    3.49   0.0016 **
## am0:wt        -3.786      0.786   -4.82  4.6e-05 ***
## am1:wt        -9.084      1.212   -7.49  3.7e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.59 on 28 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.815
## F-statistic: 46.6 on 3 and 28 DF,  p-value: 5.21e-11
```

The plot is given in Fig 1 in Appendix.

Final Model and Variable Selection

Can the first model be improved? Let **Y** denote the car's MPG, **T** the car's transmission (1 for automatic; 2 for manual), **X__2** the car weight (**wt**), and **X__3** the added variable. The new model may be written as:

Y = C_0 + C_1 **T**$1 + C$2,1 **X__2** + C_3,1 **X__3**, for cars with automatic transmission

Y = C_0 + C_1 **T**$2 + C$2,2 **X__2** + C_3,2 **X__3**, for cars with manual transmission

The following method is used to search for **X__3**. The model that produces the smallest p-value of the anova table may be desirable.

```r
fit_cyl <- anova(mpg_lm, update(mpg_lm, mpg ~ am + am:(wt + cyl)))
fit_disp <- anova(mpg_lm, update(mpg_lm, mpg ~ am + am:(wt + disp)))
fit_hp <- anova(mpg_lm, update(mpg_lm, mpg ~ am + am:(wt + hp)))
fit_drat <- anova(mpg_lm, update(mpg_lm, mpg ~ am + am:(wt + drat)))
fit_qsec <- anova(mpg_lm, update(mpg_lm, mpg ~ am + am:(wt + qsec)))
fit_vs <- anova(mpg_lm, update(mpg_lm, mpg ~ am + am:(wt + vs)))
fit_gear <- anova(mpg_lm, update(mpg_lm, mpg ~ am + am:(wt + gear)))
fit_carb <- anova(mpg_lm, update(mpg_lm, mpg ~ am + am:(wt + carb)))
```

The model that including **qsec** produces the smallest p-value (0.002) among all. Furthermore, that p-value is significant small. To ensure that the 1/4 mile time (**qsec**) is a good choice, a plot of **MPG** vs **qsec** is given in Fig 2 in Appendix. Note that for a given **qsec**, manual transmission is always better for MPG.

The final model is as follows:

```r
mpg_formula_fin <- mpg ~ am + am:(wt + qsec)
mpg_lm_fin <- lm(mpg_formula_fin, data = mtcars2)
summary(mpg_lm_fin)
```

```
##
## Call:
## lm(formula = mpg_formula_fin, data = mtcars2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.683 -1.322 -0.375  1.069  4.091
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.249      6.992    1.61  0.11974
## am1            8.926     12.666    0.70  0.48723
## am0:wt        -2.996      0.691   -4.34  0.00019 ***
## am1:wt        -6.754      1.349   -5.01  3.3e-05 ***
## am0:qsec       0.945      0.307    3.08  0.00481 **
## am1:qsec       1.181      0.464    2.54  0.01729 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.12 on 26 degrees of freedom
## Multiple R-squared:  0.897,  Adjusted R-squared:  0.877
## F-statistic: 45.1 on 5 and 26 DF,  p-value: 5.39e-12
```

As it is shown in the summary statistics of the final model, the final model produces smaller residual standard error, higher R-square and adjusted R-square than the first model.

A comparison of residuals of the two models are plotted in Fig 3 and 4 (Appendix)

Is an automatic or manual transmission better for MPG?

Based on the plot of the first model (Fig 1, Appendix), the answer of question 1 depends on the car's weight. For cars with a weight less than 2.8081 tons, manual tranmission is better for MPG. However, for cars with a weight greater than 2.8081 tons, automatic transmission is better for MPG.

To quantify the differences, the following routine is used:

```
wtarr <- seq(from = 1.5, to = 5.5, by = 0.5)
qsecarr <- seq(from = 15, to = 22, by = 0.5)
nwtarr <- length(wtarr)
nqsecarr <- length(qsecarr)

wtarr2 <- rep(rep(wtarr, each = nqsecarr), times = 2)
qsecarr2 <- rep(qsecarr, times = nwtarr * 2)
amarr2 <- factor(rep(c(0, 1), each = nwtarr * nqsecarr))

p_dataframe <- data.frame(mpg = rep(0, times = nwtarr * nqsecarr * 2), wt = wtarr2,
    qsec = qsecarr2, am = amarr2)

p_dataframe$predict <- predict(mpg_lm_fin, newdata = p_dataframe)

diffmat <- matrix(0, nrow = nwtarr, ncol = nqsecarr)
for (i in 1:nwtarr) {
    for (j in 1:nqsecarr) {
        diffmat[i, j] = p_dataframe$predict[(i - 1) * nqsecarr + j] - p_dataframe$predict[(i -
            1) * nqsecarr + j + nwtarr * nqsecarr]
    }
}
```

The contours of the difference (MPG of automatic transmission - MPG of manual transmission) matrix are plotted in Fig 5 (Appendix). The magnitude and the sign of the difference depend on the car's weight (**wt**) and 1/4 mile time (**qsec**). Within the scope of this data set, the difference ranges between (-7, 7).

Appendix

```
layout(matrix(1:2, ncol = 2))
plot(mpg ~ wt, data = mtcars2, col = as.integer(mtcars2$am), main = "Fig 1- MPG vs wt")
abline(lm(mpg ~ wt, data = mtcars2, subset = am == 0), col = 1)
abline(lm(mpg ~ wt, data = mtcars2, subset = am == 1), col = 2)
legend("topright", legend = c("auto", "manual"), col = c(1, 2), pch = 1, lty = 1)
plot(mpg ~ qsec, data = mtcars2, col = as.integer(mtcars2$am), main = "Fig 2- MPG vs qsec")
legend("topright", legend = c("auto", "manual"), col = c(1, 2), pch = 1)
```

```
layout(matrix(1:2, ncol = 2))
mpg_resid <- residuals(mpg_lm)
mpg_fitted <- fitted(mpg_lm)
plot(mpg_fitted, mpg_resid, xlab = "Fitted", ylab = "Residuals", main = "Fig 3- First model",
    col = as.integer(mtcars2$am), ylim = c(-3, 7))
mpg_resid <- residuals(mpg_lm_fin)
mpg_fitted <- fitted(mpg_lm_fin)
plot(mpg_fitted, mpg_resid, xlab = "Fitted", ylab = "Residuals", main = "Fig 4- Final model",
    col = as.integer(mtcars2$am), ylim = c(-3, 7))
```

```
layout(matrix(1))
contour(x = wtarr, y = qsecarr, z = diffmat, xlab = "wt", ylab = "qsec", labcex = 1.5,
    main = "Fig 5- difference in MPG (auto-manual)")
points(mtcars2$wt, mtcars2$qsec, col = as.integer(mtcars2$am))
legend("topright", legend = c("auto", "manual"), col = c(1, 2), pch = 1)
```

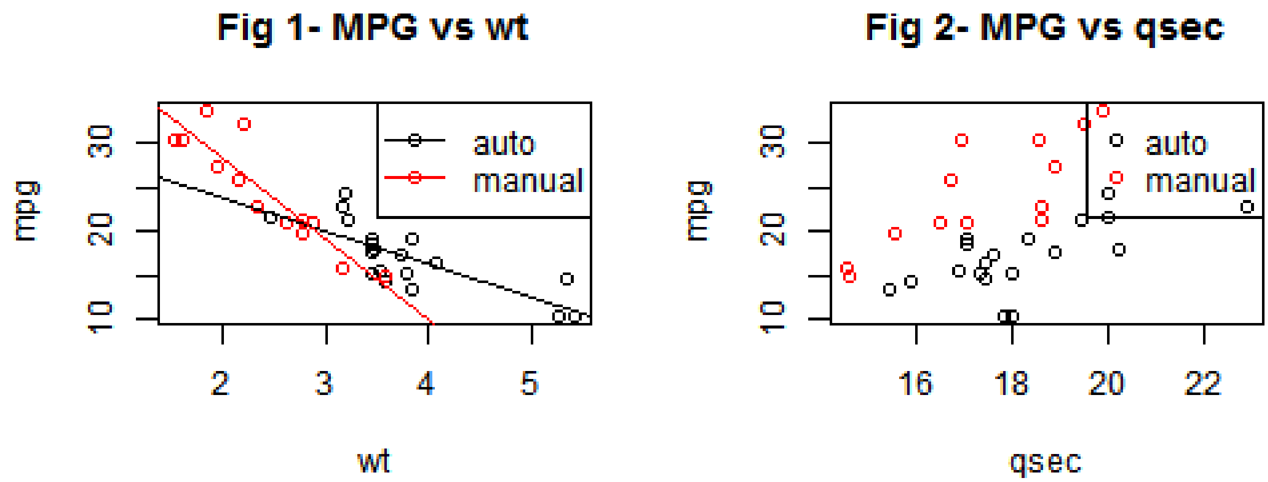

Figure 1: plot of chunk unnamed-chunk-7

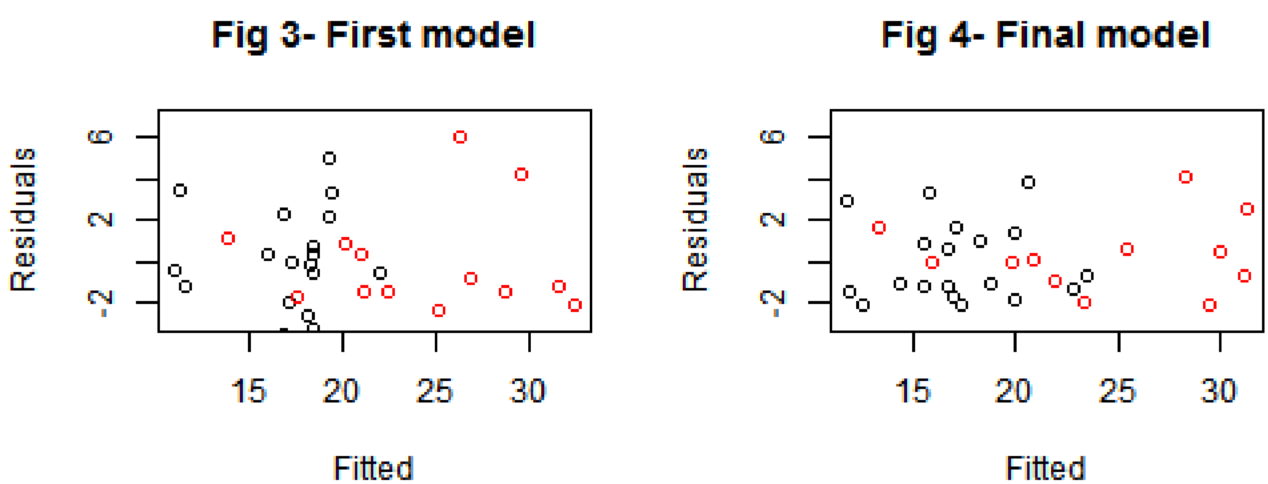

Figure 2: plot of chunk unnamed-chunk-8

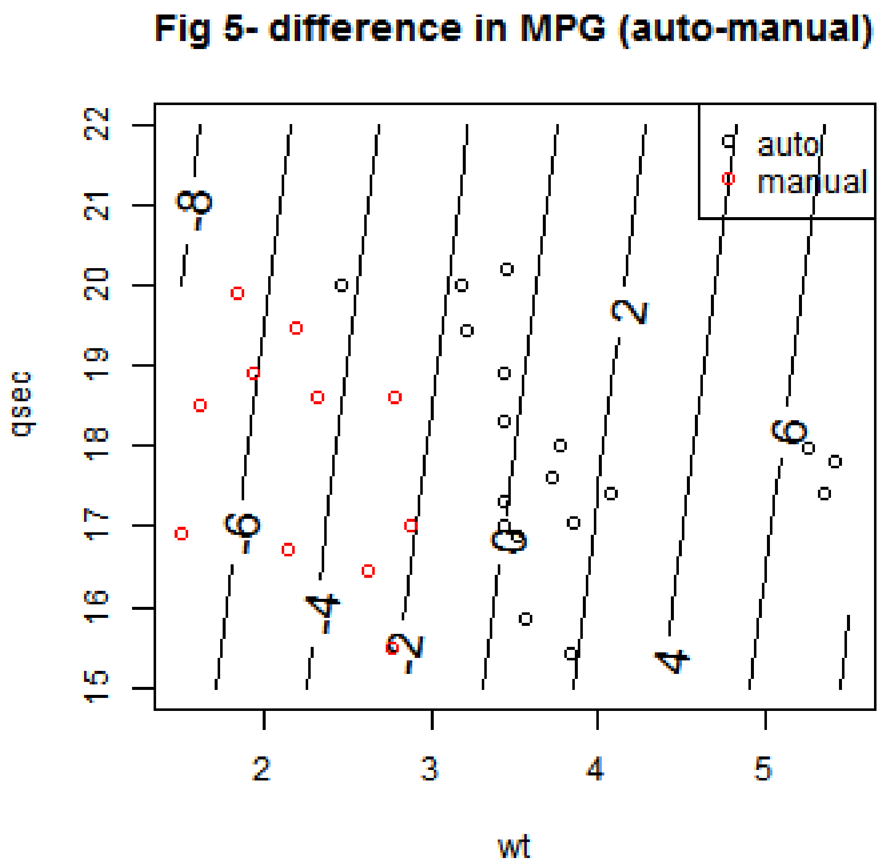

Figure 3: plot of chunk unnamed-chunk-9