Motor Trend Tests_Transmission Changes of Automatic and Manual on Miles/(US)gallon

by Mingzhu Ye

Executive Summary

Motor Trend, a magazine about the automobile industry is interested in exploring the relationship between a set of variables and miles per gallon (MPG) (mpg is the outcome). We're asked to focus on solving two questions of interest. From the regression model and statistical inference we can discover manual transmission is better for MPG. Cars with Manual transmission get 1.8 more miles per gallon on average compared to cars with Automatic transmission.

Data Processing

First, let's take a look at the basic structure of mtcars data frame! As we can see, there are 11 variables totally.

```
##                mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
```

Second, let's transfer some variables(cyl, vs, am, gear, carb) from numerics into factors.

Exploratory

1.Generally, We'll discover various relationships between variables in the processed dataset 'mtcars' (shown in Appendix **figure.1**). We leaned from figure.1 that the variables cyl, disp, hp, drat, wt, vs and am have a strong correlation with mpg.

2.Since we are interested in the effects of transmission type on mpg, a boxplot was produced to show the difference between automatic and manual in terms of mpg (Appendix **figure.2**). In figure.2, it's clear manual transmission tends to have higher mpg.

Let's take a further look at the data in regression analysis section by fitting a linear model.

Regression Analysis

Our first linear model 'allmodel' includes all variables as predictors and mpg as outcome:

```
allmodel <- lm(mpg~., data = mtcars)
```

Secondly, we'll build a best match model by AIC in a Stepwise Algorithm. It produces list of the best predictors. As shown in 'bestmodel' summary, the most significant predictors in determining the mpg are cyl, hp, wt and am. (shown in Appendix **figure.3**)

```r
bestmodel <- step(allmodel, direction = "both", trace = 0)
```

Then let's build a base linear model using am as predictor and mpg as outcome.

```r
basemodel <- lm(mpg ~ am, data = mtcars)
```

Next we compare the basemodel and the bestmodel which we obtained above. As shown in compare results, the P-value is highly significant and we reject the null hypothesis that cyl, hp and wt don't contribute to the accuracy of the model. This means am, cyl, hp and wt significantly contribute to the outcome mpg.(shown in Appendix **figure.4**)

```r
compare <- anova(basemodel, bestmodel)
```

Model Residuals and Diagnostics

The residuals from the bestmodel are plotted in Appendix **figure.5**

Observasion
-The Residuals vs Fitted plot showing no pattern between the residuals and fitted values indicating the good independence condition.
-The Normal Q-Q plot concluded points almost falling on the line indicates the residuals are normally distributed.
-The scale-Location plot concluded points scattered in a constant pattern indicates constant variance.
-There are some distinct points in the Residuals vs Leverage plot indicating top right values have increased leverage.

Statistical Inference

In this section, we perform a T-test on the two subsets of mpg data: manual and automatic transmission assuming that the transmission data has a normal distribution. And also, test the null hypothesis that they come from the same distribution. Codes are shown in Appendix **figure.6**.

Observation
Based on the t-test results, p-value (0.001374) is much lower than 0.05, so we reject the null hypothesis that the mpg distributions for manual and automatic transmissions are the same.

Conclusion

-Manual transmission is better for MPG. Cars with Manual transmission get 1.8 more miles per gallon on average compared to cars with Automatic transmission.
-mpg decreases 0.032 with every increase of hp.
-mpg decreases 2.497 with every 1000lb increase of wt.
-mpg decreases 3 with every increase of cyl6 and 2 with every increase of cyl8.
-Adjusted R^2 is 0.8401 indicating more than 84% of the variability is explained by the above model.

Appendix

figure.1

```r
pairs(mpg ~ ., data = mtcars)
```

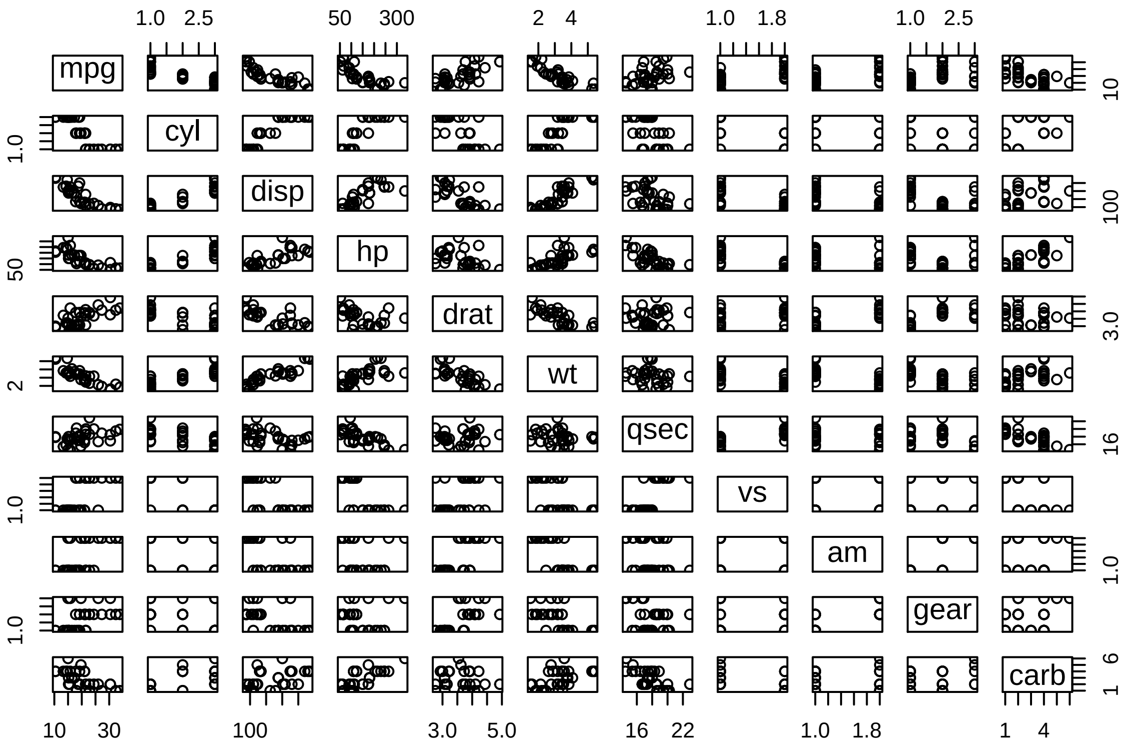

figure.2

```r
boxplot(mpg ~ am, data = mtcars, col = (c("yellow","steelblue")),
        ylab = "Miles Per Gallon", xlab = "Transmission Type")
```

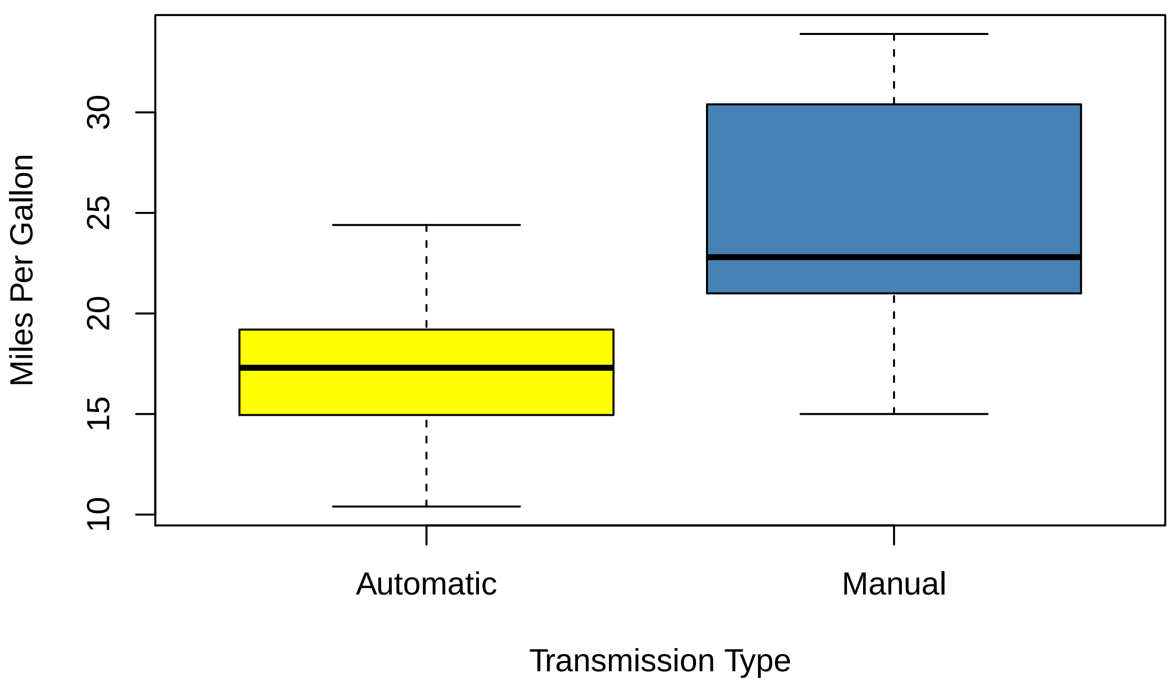

figure.3

```r
bestmodel <- step(allmodel, direction = "both", trace = 0)
summary(bestmodel)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

figure.4

```r
compare <- anova(basemodel, bestmodel)
compare
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 151.03  4    569.87 24.527 1.688e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

figure.5

```r
par(mfrow = c(2,2))
plot(bestmodel)
```

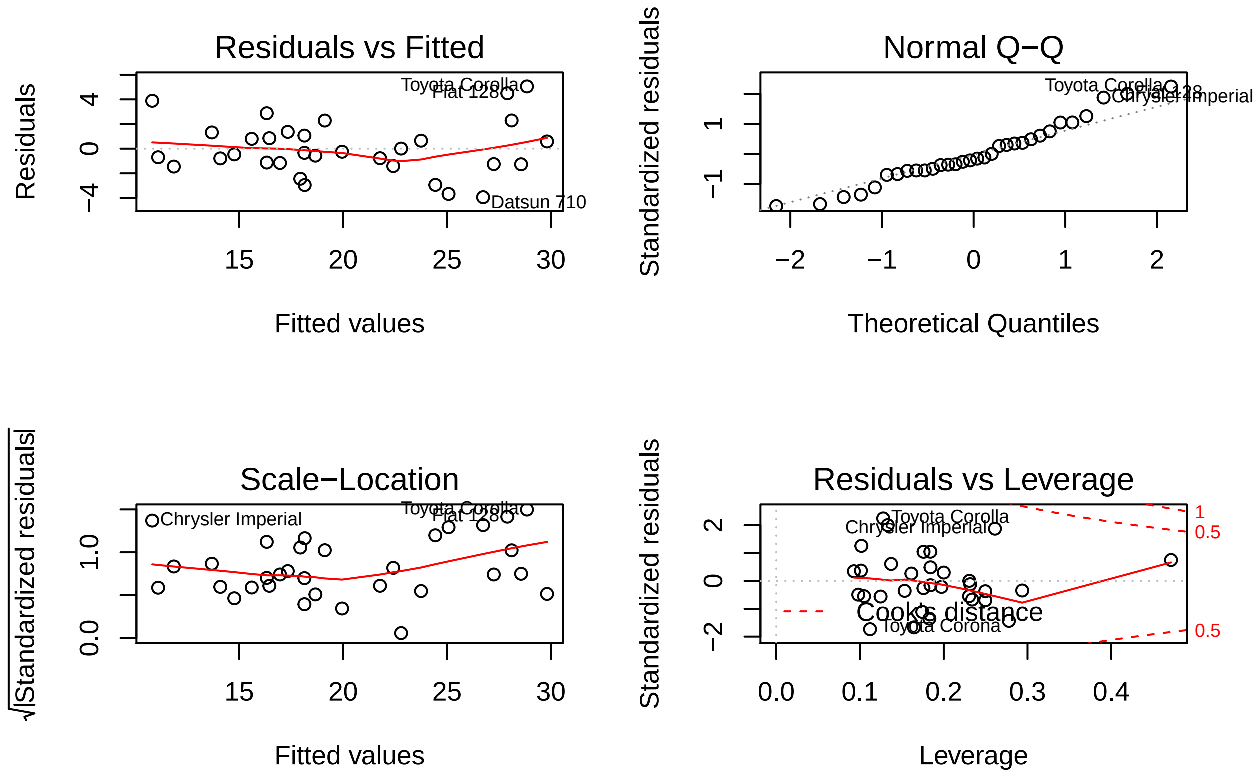

figure.6

```r
t.test(mpg ~ am, data = mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group Automatic    mean in group Manual
##                17.14737                24.39231
```