

Regression Models Course Project

suwei

2018-3-13

Executive Summary

This is an report about the assignment of Coursera project: Regression Models. In the report, we explored dataset `mtcars` to answer two interested questions:

1. "Is an automatic or manual transmission better for MPG"
2. "Quantify the MPG difference between automatic and manual transmissions"

The answer is that:

1. In average, the `manual transmission` vehicles have a 7.245 greater MPG than the `automatic transmission` vehicles (confirmed by t-test).
2. In our model, it shows that the `manual transmission` vehicles only have a 1.81 MPG greater than `automatic transmission` vehicles.

Exploratory Analysis

```
library(ggplot2)
data("mtcars")
head(mtcars)
```

| | mpg | cyl | disp | hp | drat | wt | qsec | vs | am |
|-------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| Mazda RX4 | 21.0 | 6 | 160 | 110 | 3.90 | 2.620 | 16.46 | 0 | 1 |
| Mazda RX4 Wag | 21.0 | 6 | 160 | 110 | 3.90 | 2.875 | 17.02 | 0 | 1 |
| Datsun 710 | 22.8 | 4 | 108 | 93 | 3.85 | 2.320 | 18.61 | 1 | 1 |
| Hornet 4 Drive | 21.4 | 6 | 258 | 110 | 3.08 | 3.215 | 19.44 | 1 | 0 |
| Hornet Sportabout | 18.7 | 8 | 360 | 175 | 3.15 | 3.440 | 17.02 | 0 | 0 |
| Valiant | 18.1 | 6 | 225 | 105 | 2.76 | 3.460 | 20.22 | 1 | 0 |

6 rows | 1-10 of 12 columns

```
col_name <- c('cyl', 'vs', 'am', 'gear', 'carb')
mtcars[col_name] <- lapply(mtcars[col_name], factor)

automatic <- mtcars[mtcars$am == 0,]
manual <- mtcars[mtcars$am == 1,]
```

As you can see, in Appendix Plot 1, `automatic trasmission` vehicles are having a lower MPG than `manual transmission`. So we considered it as our hypothesis, and use t-test to confirmed it.

```
test <- t.test(manual$mpg, automatic$mpg)
```

The p-value of the test is 0.0013736, so the test rejects the null hypothesis that there is no difference between manual transmission and automatic transmission. The manual transmission vehicles have a 7.2449393 greater MPG than the automatic transmission vehicles.

Next, let's quantify the MPG difference between automatic transmissions and manual transmissions.

```
model_all <- lm(mpg ~., data = mtcars)
model_best <- step(model_all)
```

```
summary(model_best)
```

```
##
## 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 **
## am            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
```

First, we use all the variables except mpg to build a linear model, and then use the step() function to find the best one. It seems that when using variables: cyl, hp, wt, am to build a model, we can get the best model which its AIC is the lowest one. The R-squared value is 0.8659 confirms that this model explains 86.59% of the variance in MPG. Also the extremely small p-value shows the model is statistically significant. As a result, the manual transmission vehicles only have a 1.81 MPG greater than automatic transmission vehicles.

Last, let's check the residual diagnostics part. As you can see, in Appendix Plot 2:

1. From the **Residuals vs Fitted** plot.
2. From the **Normal Q-Q** plot, residuals are followed the normal distribution.
3. From the **Scale-Location** plot, random distribution of residuals confirms that the variance is almost a constant value.
4. From the **Residuals vs Leverage** plot, no high leverage point exists.

Conclusion

MPG difference exists between manual transmission vehicles and automatic transmission vehicles. A manual transmission vehicle has 1.81 MPG greater than an automatic transmission vehicle.

Appendix

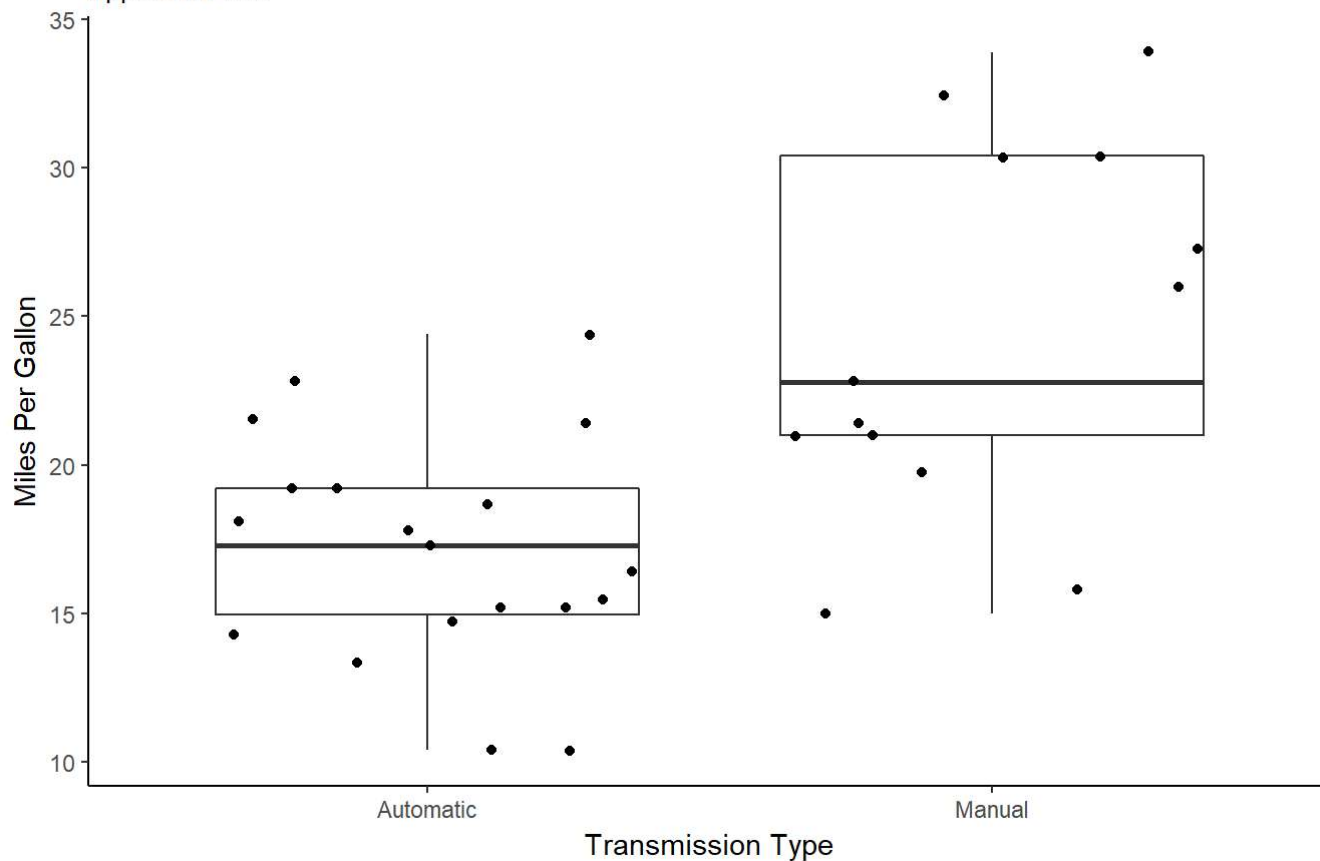
```

box_plot <- ggplot(data = mtcars, aes(x = am, y = mpg)) + geom_boxplot()
box_plot <- box_plot + geom_jitter() + theme_classic()
box_plot <- box_plot + ggtitle("Boxplot of MPG by transmission type", subtitle = "Appendix Plot 1") +
  xlab("Transmission Type") + ylab("Miles Per Gallon") + scale_x_discrete(labels=c("1" = "Manual", "0"
    = "Automatic"))
box_plot

```

Boxplot of MPG by transmission type

Appendix Plot 1



Appendix Plot 2

```

par(mfrow = c(2, 2))
plot(model_best)

```

