Motor Trend - Auto Transmission Analysis

Zhang,Xiaoyu

Saturday, April 25, 2015

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) (outcome). They are particularly interested in the following two questions: - "Is an automatic or manual transmission better for MPG" - "Quantify the MPG difference between automatic and manual transmissions"

Data Descriptive

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). - mpg : Miles/(US) gallon - cyl : Number of cylinders - disp : Displacement (cu.in.) - hp : Gross horsepower - drat : Rear axle ratio - wt : Weight (lb/1000) - qsec : 1/4 mile time - vs : V/S - am : Transmission (0 = automatic, 1 = manual) - gear : Number of forward gears - carb : Number of carburetors

Main Analysis

Exploratory Analysis

```
data(mtcars)
summary(mtcars[,c("mpg","am")])
```

```
##       mpg              am
##  Min.   :10.40   Min.   :0.0000
##  1st Qu.:15.43   1st Qu.:0.0000
##  Median :19.20   Median :0.0000
##  Mean   :20.09   Mean   :0.4062
##  3rd Qu.:22.80   3rd Qu.:1.0000
##  Max.   :33.90   Max.   :1.0000
```

Data Pre-processing

Transform the cylinder and transmission data to factors

```
mtcars <- transform(mtcars,cyl = factor(cyl))
mtcars <- transform(mtcars,am = factor(am,labels = c("Automatic","Manual")))
```

Hypothesis Test

Null: $\beta_{automotic} = \beta_{manual}$

When building linear model, coefficient β_{am} for X_{am} is interpreted as the difference in the mean whem comparing two groups.

```
fit <- lm(mpg ~ am, data = mtcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ 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 ***
## amManual       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
```

Quantify the Difference

Build the full model including all variables and select the variables using backward selection to capture the most variability.

```
fit2 <- lm(mpg ~ .,data = mtcars)
fit2.best <- step(fit2,trace = F,k=log(nrow(mtcars)))
fit2.best
```

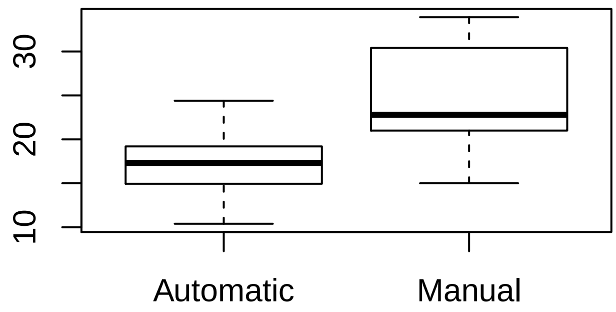
```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Coefficients:
## (Intercept)           wt         qsec     amManual
##       9.618       -3.917        1.226        2.936
```

Result

- As we can see from above, the P-value for β_{am} is 0.000285 which reject the null hypothesis, which means there is a great difference in different type of transmission
- The coefficient β_{am} is positive which means Manual transmission is better for MPGX
- The MPG difference is β_{am}, which is 2.936

Appendix

```r
boxplot(mpg ~ am,data = mtcars)
```




```r
par(mfrow=c(2,2))
plot(fit2.best)
```