itle: “Regression Model Assignment”
uthor: “Yanyu”
ate: “October 21, 2017”
utput: pdf_document

Motor Trend Analysis

Summary

The dataset is a extract from 1974 Motor Trend US magazine compromises the fuel consupmtion and 10 aspects of automobile design and performance.

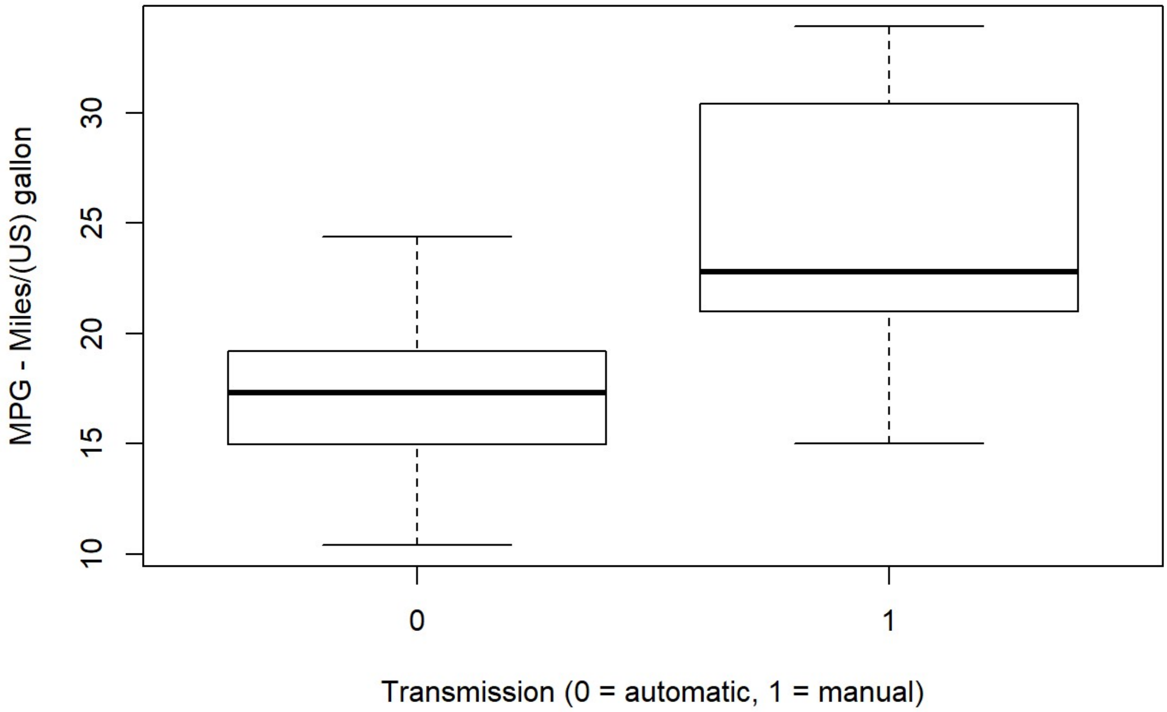
```
head(mtcars)
```

```
##                    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
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```
mtcars$am <- factor(mtcars$am)
boxplot(mpg ~ am, data = mtcars,
        xlab="Transmission (0 = automatic, 1 = manual)",
        ylab="MPG - Miles/(US) gallon")
```



According on the boxplot chart, it’s quite obvious that the mean and distribution of the Miles per Gallon (MPG) in manual transmission is higher than automatic transmission.

Simple Linear Regression Model

MPG (mpg) vs Transmission (am)

```
fit <- lm(mpg ~ am  , mtcars)
summary(fit)$coeff
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am1          7.244939   1.764422  4.106127 2.850207e-04
```

```
summary(fit)$adj.r.squared
```

```
## [1] 0.3384589
```

The p-value 0.000285 is lower than 5% confidence level. We can reject the null hypothesis that ??1 = 0. In other words, there is significant relationship between MPG and transmission type in the linear regression model with confidence level > 95%. However, the Adjusted R-squared value is quite low that only 33.85% of regression variance can be explained by this model. It may lead to higher chance of under fitting with this model. In other words, there might need tobe more variables that we take into consideration.

Multivariate Linear Regression Model

Covariates Selection

```
all_fit <- lm(mpg ~ ., data = mtcars)
all_fit$coefficients
```

```
## (Intercept)         cyl        disp          hp        drat          wt
## 12.30337416 -0.11144048  0.01333524 -0.02148212  0.78711097 -3.71530393
##        qsec          vs         am1        gear        carb
##  0.82104075  0.31776281  2.52022689  0.65541302 -0.19941925
```

```
summary(all_fit)$adj.r.squared
```

```
## [1] 0.8066423
```

Weight (wt) and Gross Horsepower (hp) are selected because they are the other 2 variables with lowest p-values and standard errors.

```
wt_fit <- lm(mpg ~ wt, data = mtcars)
wt_fit$coefficients
```

```
## (Intercept)          wt
##   37.285126   -5.344472
```

```
summary(wt_fit)$adj.r.squared
```

```
## [1] 0.7445939
```

Obviously, weight (wt) variable has near to zero empirical p-values and very high Adjusted R-squared value at 74.46%.

```
hp_fit <- lm(mpg ~ hp, data = mtcars)
hp_fit$coefficients
```

```
## (Intercept)          hp
## 30.09886054 -0.06822828
```

```
summary(hp_fit)$adj.r.squared
```

```
## [1] 0.5891853
```

Obviously, horsepower (hp) variable has near to zero empirical p-values and high Adjusted R-squared value at 58.92%. So, weight (wt) and horsepower (hp) are best covariates to be associated into the regression model.

MPG (mpg) vs Transmission (am) + Weight (wt) + Horsepower (hp)

```
best_fit <- lm(mpg ~ am + wt + hp, data = mtcars)
best_fit$coefficients
```

```
## (Intercept)         am1          wt          hp 
## 34.00287512  2.08371013 -2.87857541 -0.03747873
```

```
summary(best_fit)$adj.r.squared
```

```
## [1] 0.8227357
```

In this Multivariate Regression Model, the p-value is near to zero, hence null hypothesis is rejected. This time, the R-squared value (82.27%) is much higher than the previous Simple Regression Model.

```
par(mfrow = c(2,2))
plot(best_fit)
```