RegressionProject

Hnin Su PHYU

8/27/2020

##Overview You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are 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" ##Dataset

```r
library(datasets)
data(mtcars)
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
```

##Processing Data

```r
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- as.factor(mtcars$am)
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 : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...
##  $ 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  : Factor w/ 2 levels "0","1": 1 1 2 2 1 2 1 2 2 2 ...
##  $ am  : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

##Exploratory analysis with Appendix: Figures The mean of the variant cars (by transmission types) for their MPG:

```r
mtcars$am<-as.factor(mtcars$am)
levels(mtcars$am)<-c("Auto", "Man.")
aggregate(mpg ~ am, data = mtcars, mean)
```

```
##     am      mpg
## 1 Auto 17.14737
## 2 Man. 24.39231
```

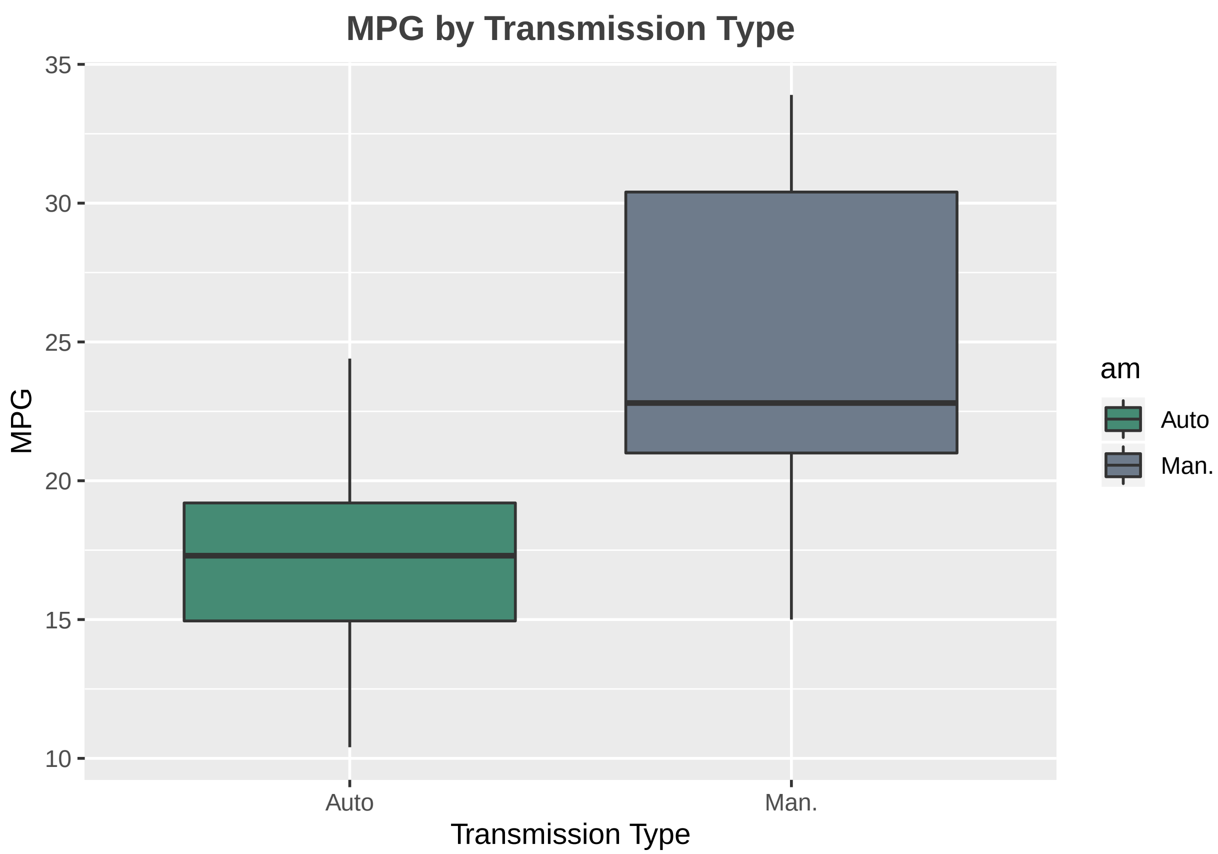
We can see from this simple view that the mean average of MPG's for Automatic cars outstrips that of Manual.

We should see if this variance is significant - via running a t-test:

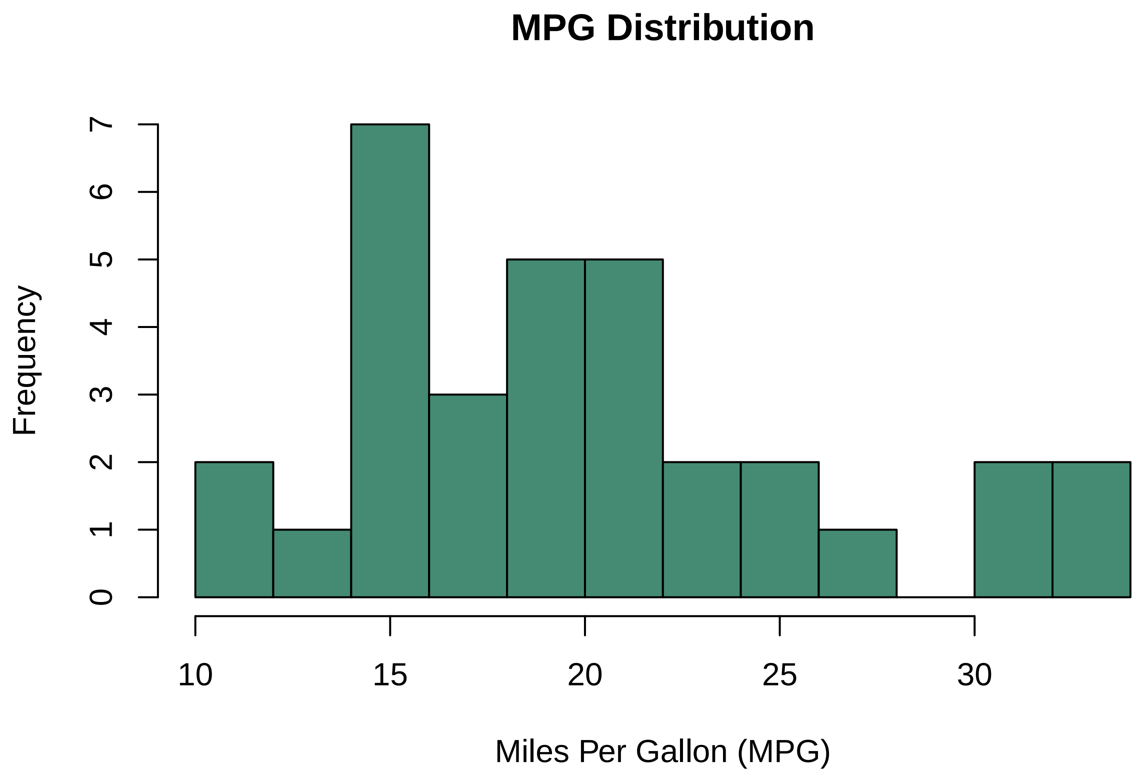
```r
Auto.Data <- mtcars[mtcars$am == "Auto",]
Manu.Data <- mtcars[mtcars$am == "Man.",]

t.test(Auto.Data$mpg, Manu.Data$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  Auto.Data$mpg and Manu.Data$mpg
## 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 of x mean of y
##  17.14737  24.39231
```

```r
library(ggplot2)
## Box Plot  of MPG by Transmission Type
g <- ggplot(aes(x = am, y = mpg), data = mtcars)
g <- g + geom_boxplot(aes(fill = am))
g + labs(x = "Transmission Type", y = "MPG", title = "MPG by Transmission Type") +
        theme(plot.title = element_text(color="grey25", face="bold",hjust=0.5)) +
        scale_fill_manual(values=c("aquamarine4","lightsteelblue4"))
```




```
## Histgram of MPG
hist(mtcars$mpg, breaks=12, xlab="Miles Per Gallon (MPG)", main="MPG Distribution", col="aquamarine4")
```



##Regression modelling

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

```
##
## 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 ***
## amMan.         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
```

Since Adjusted R^2 = 0.3385, thus only 33.85% of the regression variance of this model is explained.

```r
multiVariableRegressionModel <- lm(mpg ~ am + wt + hp, data = mtcars)
summary(multiVariableRegressionModel)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## amMan.       2.083710   1.376420   1.514 0.141268
## wt          -2.878575   0.904971  -3.181 0.003574 **
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

Compared to single variable regression model, the Adjusted R^2 = 0.823, thus approximately 82.3% of the regression variance of this model is explained.

```r
anova(singleVariableRegressionModel, multiVariableRegressionModel)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + hp
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     28 180.29  2    540.61 41.979 3.745e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value is 3.745e-09, we can now reject the H0 and state that our Multivariable Model is considerably different than the Single Linear Regression Model.

```r
t.test(mtcars$mpg~mtcars$am)
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg by mtcars$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 Auto mean in group Man.
##           17.14737           24.39231
```

With assumption that all other conditions remain unchanged. Since p-value = 0.001374 which is less than 0.05, we conclude that manual transmission is better than automatic transmission for MPG and reject the null hypothesis that there is no difference in MPG.

##Executive Summary The key finding of this analysis is that manual transmission on average has a better miles per gallon (mpg) than its counterpart i.e. automatic transmission.