RegressionModelsAssignment

JB

2023-12-01

#Executive Summary This report explores the relationship between transmission type (automatic or manual) and miles per gallon (MPG) in the Motor Trend dataset. We address two main questions: 1) Is an automatic or manual transmission better for MPG? 2) What is the quantified MPG difference between automatic and manual transmissions? From our analysis we can say the automatic transmission yields better MPG than manual. When holding the weight and accelration of the car constant, cars with manual transmission add 14.079 + (-4.141)*wt more MPG (miles per gallon) on average than cars with automatic transmission.

Loading data and packages

```
library(MASS)
library(ggplot2)
data(mtcars)
```

Data manipulation

```
str(mtcars)
```

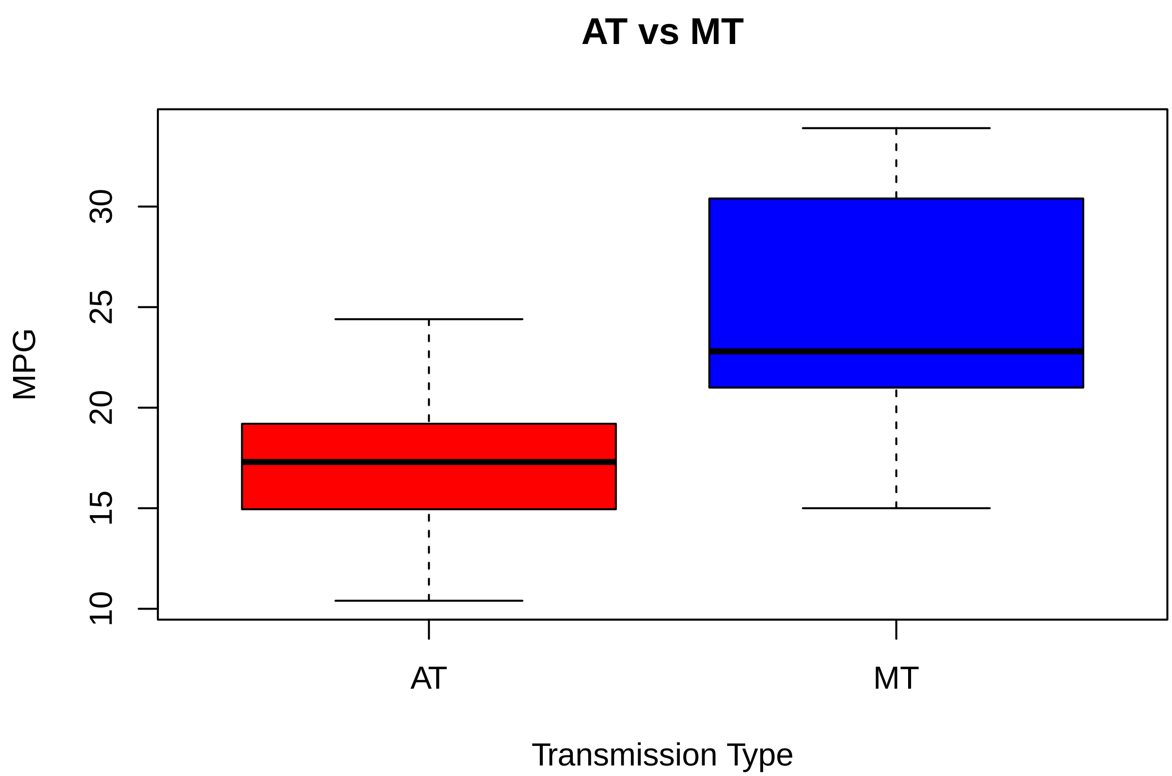
```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 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$am = as.factor(mtcars$am)
levels(mtcars$am)=c("AT", "MT")
```

We change the am variable from a numeric class to a factor class to improve the model interpretability.

Comparison of the effect of automatic or manual transmission on MPG

```
boxplot(mpg ~ am, data = mtcars, col = c("red", "blue"), xlab = "Transmission Type", ylab = "MPG", main
```




```
t_test_result<- t.test(mtcars$mpg ~ mtcars$am)
t_test_result$p.value
```

```
## [1] 0.001373638
```

```
t_test_result$estimate
```

```
## mean in group AT mean in group MT
##         17.14737         24.39231
```

```
levels(mtcars$am)=as.numeric(c(1,0))
mtcars$am=as.numeric(mtcars$am)

model1 <- lm(mpg ~ factor(am), data = mtcars)
summary(model1)
```

```
##
## Call:
```

```
## lm(formula = mpg ~ factor(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 ***
## factor(am)2    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
```

The plot shows that manual transmission yields better mileage than automatic but we will perform a statistical t-test to confirm that result. Our null hypothesis is that AT and MT have no different effect on MPG. Since the p-value of the test is lower than 0.05 the Null hypothesis is rejected, therefore we can say that MT and AT are from different populations. We see that for automatic transmission cars (am=0) the milege is 17.147 mpg holding other variables constant. For manual transmission cars, there is 7.245 units increase in mpg (in comparison to AT) while holding other variables constant.

Correlation between MPG vs. other attributes of the Cars

We use the cor() function on the variables from MTCARS dataset to understand the correlation between the variables, we extracted especially the MPG vs all data as it is the most relevant. This information helps us to decide the variables that may influence the MPG during regression analysis.

```
Correltion<- cor(mtcars[, c(1:11)])
MPG_cor<-Correltion[,1]
MPG_cor
```

```
##        mpg        cyl       disp         hp       drat         wt       qsec
##  1.0000000 -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594  0.4186840
##         vs         am       gear       carb
##  0.6640389  0.5998324  0.4802848 -0.5509251
```

We see that some variables are negatively correlated to MPG it e.g. (cyl) number of cylinders or (wt) car weight. Other have positive effect on the mileage like (qsec) accelaration speed or (am) transmission. This doesn't however reveal any information on causation.

Regression model MPG vs ALL

For building a linear regression model, appropriate variable is absolute necessary. Including too few variables may cause bias and including too many would cause high variance and over fitting issues. in such situations model selection becomes ideal. In our case we use step() function to choose the best model based on AIC in a stepwise algorithm.

```r
model_step<- step(lm(mpg~., mtcars))
```

```r
summary(model_step)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.6819     8.0101   0.834 0.411235
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Through the step function over lm() model, we see that variables like "car weight (wt)" and "Acceleration Speed (qsec)" turns out to be good predictors od Milege mpg. These variables where able to explain 84% of variation in Milege (mpg).

Regression model MPG vs WT:AM

From the correlation matrix we can see that there appears to be a strong interaction between variables WT and AM. This might be due to the fact that automatic cars weight more than manuals. The interaction will be checked by the following regression model

```r
modelAMinWT<- lm(mpg~ wt + qsec + am + wt:am, mtcars)
summary(modelAMinWT)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am + wt:am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5076 -1.3801 -0.5588  1.0630  4.3684
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -4.356      7.501  -0.581 0.566233
## wt             1.205      1.598   0.754 0.457412
## qsec           1.017      0.252   4.035 0.000403 ***
```

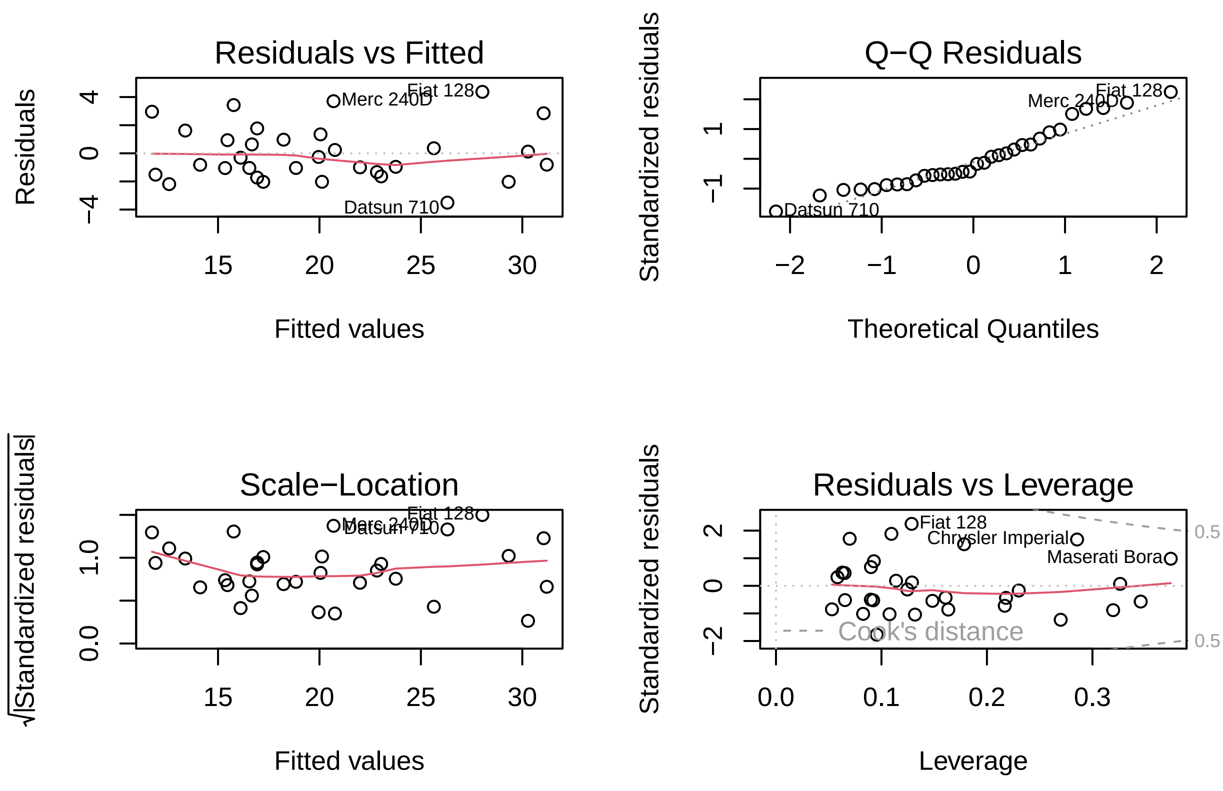
```
## am            14.079      3.435   4.099 0.000341 ***
## wt:am         -4.141      1.197  -3.460 0.001809 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.084 on 27 degrees of freedom
## Multiple R-squared:  0.8959, Adjusted R-squared:  0.8804
## F-statistic: 58.06 on 4 and 27 DF,  p-value: 7.168e-13
```

This model has the Residual standard error as 2.084 on 27 degrees of freedom. And the Adjusted R-squared value is 0.8804, which means that the model can explain about 88% of the variance of the MPG variable. All of the coefficients are significant at 0.05 significant level, which is satisfactory.

Based on the Adjusted R-squared Value the chosen model is the modelAMinWT.

Residual plots

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



Looking at the residual plots, we observe that The Residuals vs. Fitted plot shows no consistent pattern, supporting the accuracy of the independence assumption. The Normal Q-Q plot indicates that the residuals are normally distributed because the points lie closely to the line.The Scale-Location plot confirms the constant variance assumption, as the points are randomly distributed. The Residuals vs. Leverage argues that no outliers are present, as all values fall well within the 0.5 bands.

Summary

```r
summary(modelAMinWT)$coef
```

```
##              Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) -4.356375  7.5014311 -0.5807392 0.5662325009
## wt           1.204846  1.5980694  0.7539382 0.4574115203
## qsec         1.016974  0.2520152  4.0353659 0.0004030165
## am          14.079428  3.4352512  4.0985148 0.0003408693
## wt:am       -4.141376  1.1968119 -3.4603403 0.0018085763
```

Based on this information we can say that when wt and qsec remain constant cars with manual transmission add 14.079 + (-4.141)*wt more MPG (miles per gallon) on average than cars with automatic transmission.