RegressionModel_Week4_AssignedProject

Ronda Hutchin

10/25/2020

EXECUTIVE SUMMARY: This report is for Motor Trend Magazine, and for the purpose the purpose is to (A) analyzing a data set of a collection of cars, (B) explore the relationship between a set of variables and miles per gallon (MPG) (outcome) and (C) answer the two questions listed below.

(1) "Is an automatic or manual transmission better for MPG"?
(2) "Quantify the MPG difference between automatic and manual transmissions"? ## Examine the data with STR, HEAD and SUMMARY, and save space hiding results, messages and warnings

```
library(ggplot2)
data(mtcars)
library(dplyr)
head(mtcars)
summary(mtcars)
str(mtcars)
```

EXPLORATORY ANALYSIS using boxplot (in appendix A), getting the mean of each type and doing a t-test. The below results has a p-value = 0.001374 which iswithin the 95% confidence interval. It appears the difference between the mean MPG of Automatic and Manual cars is significant, but further analysis is needed.

```
aggregate(mpg ~ am, data = mtcars, mean)
```

```
##   am      mpg
## 1  0 17.14737
## 2  1 24.39231
```

```
Automatic <- mtcars[mtcars$am == "0",]
Manual <- mtcars[mtcars$am == "1",]

t.test(Automatic$mpg, Manual$mpg)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  Automatic$mpg and Manual$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
```

FURTHER ANALYSIS using a Regression model good for "binary" models since we have two transmission types represented by 0 and 1. The results are promising as shown below - note the probability and AIC values. More analysis is needed

```r
fit1<-glm(mtcars$am~mtcars$mpg, family ="binomial")
exp(fit1$coefficients)
```

```
## (Intercept)  mtcars$mpg
## 0.001355579 1.359379288
```

```r
summary(fit1)
```

```
##
## Call:
## glm(formula = mtcars$am ~ mtcars$mpg, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5701  -0.7531  -0.4245   0.5866   2.0617
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -6.6035     2.3514  -2.808  0.00498 **
## mtcars$mpg    0.3070     0.1148   2.673  0.00751 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 43.230  on 31  degrees of freedom
## Residual deviance: 29.675  on 30  degrees of freedom
## AIC: 33.675
##
## Number of Fisher Scoring iterations: 5
```

FURTHER ANALYSIS using Regression model to consider other vairables which is why I chose a multivariable regression model to also consider cyclinders and horsepower. Results are below for fit2 to the last fit7

```r
fit2<-lm(mpg~factor(am)+factor(cyl)+hp, data=mtcars)
anova(fit2)
```

```
## Analysis of Variance Table
##
## Response: mpg
##             Df Sum Sq Mean Sq F value    Pr(>F)
## factor(am)   1 405.15  405.15 55.4722 5.183e-08 ***
## factor(cyl)  2 456.40  228.20 31.2446 9.430e-08 ***
## hp           1  67.30   67.30  9.2141  0.005266 **
## Residuals   27 197.20    7.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ factor(am) + factor(cyl) + hp, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.231 -1.535 -0.141  1.408  5.322
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.29590    1.42394  19.169  < 2e-16 ***
## factor(am)1   4.15786    1.25655   3.309  0.00266 **
## factor(cyl)6 -3.92458    1.53751  -2.553  0.01666 *
## factor(cyl)8 -3.53341    2.50279  -1.412  0.16943
## hp           -0.04424    0.01458  -3.035  0.00527 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.703 on 27 degrees of freedom
## Multiple R-squared:  0.8249, Adjusted R-squared:  0.7989
## F-statistic: 31.79 on 4 and 27 DF,  p-value: 7.401e-10
```

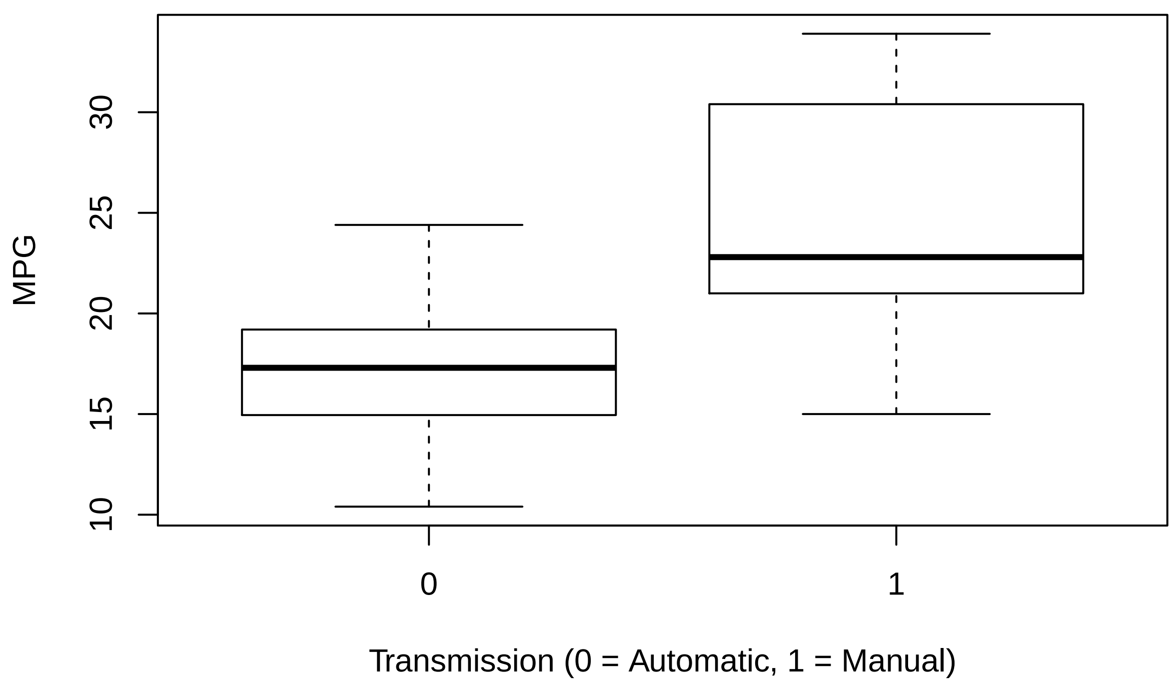
```r
fit3<-lm(mpg ~ factor(am) + factor(cyl), data=mtcars)
fit4<-lm(mpg ~ factor(am) + factor(cyl)+ hp, data=mtcars)
fit5<-lm(mpg ~ factor(am) + factor(cyl)+ wt +hp, data=mtcars)
fit6<-lm(mpg ~ factor(am) + factor(cyl)+ wt +hp +disp, data=mtcars)
fit7<-lm(mpg ~ factor(am) + factor(cyl)+ wt +hp +disp+qsec+factor(gear)+factor(carb), data=mtcars)
anova(fit3,fit4,fit5,fit6,fit7)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am) + factor(cyl)
## Model 2: mpg ~ factor(am) + factor(cyl) + hp
## Model 3: mpg ~ factor(am) + factor(cyl) + wt + hp
## Model 4: mpg ~ factor(am) + factor(cyl) + wt + hp + disp
```

```
## Model 5: mpg ~ factor(am) + factor(cyl) + wt + hp + disp + qsec + factor(gear) +
##     factor(carb)
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     28 264.50
## 2     27 197.20  1    67.297 9.0855 0.007814 **
## 3     26 151.03  1    46.173 6.2337 0.023102 *
## 4     25 150.41  1     0.617 0.0833 0.776403
## 5     17 125.92  8    24.489 0.4133 0.897325
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

CONCLUSION for Question (1) & (2): Transmission type is significant in it's relationship with MPG - it has lower value significance when considered with weight AND qsec (acceleration) - cars with manual transmission add 14.079 + (-4.141)*wt more MPG (miles per gallon) on average than cars with automatic transmission. The multivairable regression test results suggest that fuel efficiency is approx. 2.94 MPG higher for Manual cars over Automatic. ## APPENDIX Results of boxplot and multivariable analysis

```r
with(mtcars, boxplot(mpg~am, xlab="Transmission (0 = Automatic, 1 = Manual)", ylab="MPG"))
```



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