# Regression Models Course Project

Khalid

7/17/2020

## 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"
```

```
data("mtcars")
library(ggplot2)
library(dplyr)
```

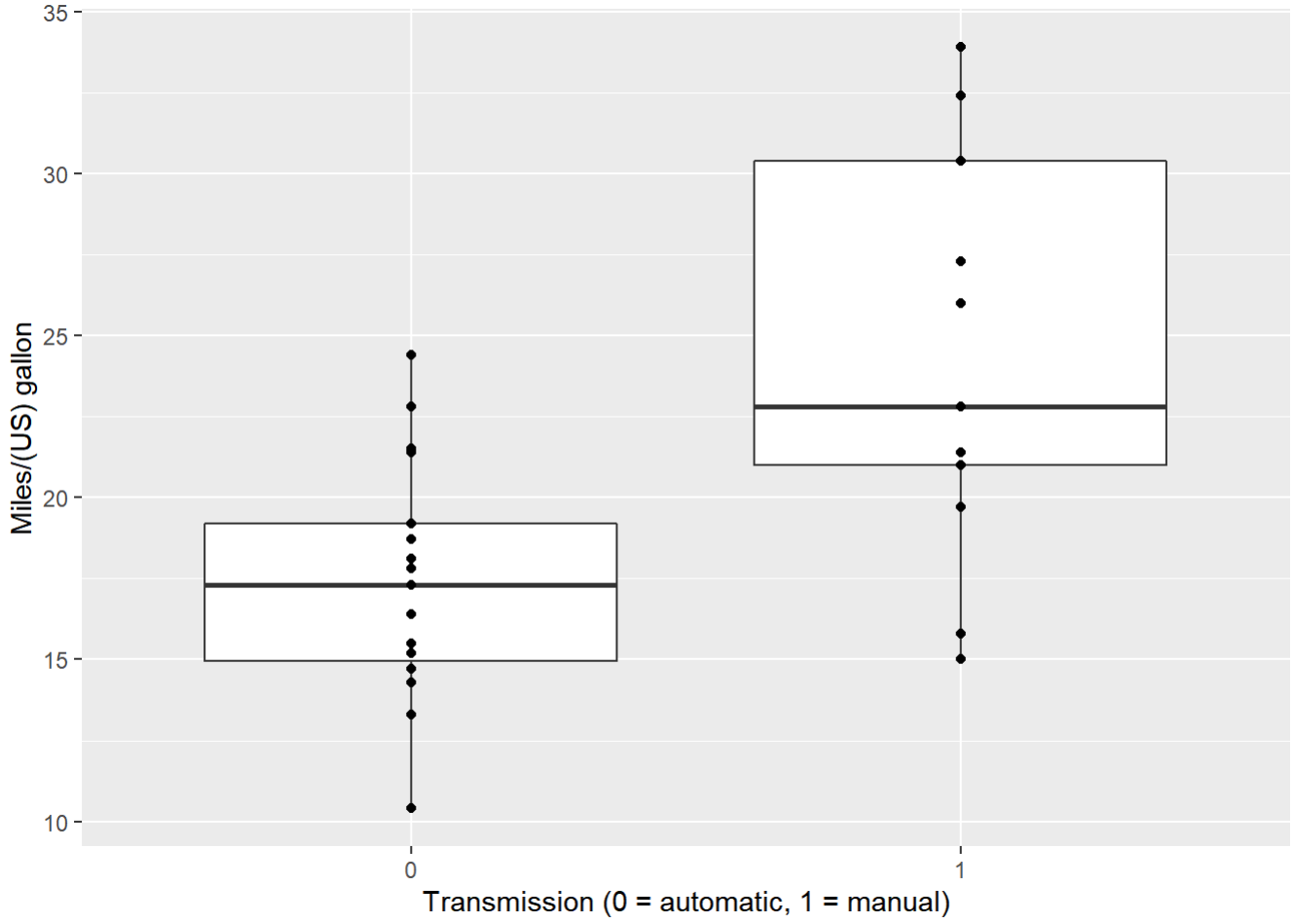
```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

## manual transmission has higher mean of miles per gallon

```
auto_mpg <- mtcars %>% select(mpg, am) %>% mutate(am = as.factor(am))
ggplot(data = auto_mpg, aes(x = am, y = mpg)) +
        geom_boxplot() + geom_point() +
        xlab('Transmission (0 = automatic, 1 = manual)') +
        ylab('Miles/(US) gallon')
```




## p-value less than 5% in the full model

```
D <- mtcars %>% mutate(cyl = as.factor(cyl), vs = as.factor(vs), am = as.factor(am), gear = as.f
actor(gear), carb = as.factor(carb))
fit_All <- lm(mpg ~ ., data = D)
summary(fit_All)$coef[, 4]
```

```
## (Intercept)        cyl6        cyl8        disp          hp        drat
##  0.25252548  0.39746642  0.96317000  0.28267339  0.09393155  0.64073922
##          wt        qsec         vs1         am1       gear4       gear5
##  0.09461859  0.69966720  0.51150791  0.71131573  0.77332027  0.50889747
##       carb2       carb3       carb4       carb6       carb8
##  0.67865093  0.49546781  0.80956031  0.49381268  0.39948495
```

# Still no coefficients with a significant p-value after removing the cyl variable

```
which.max(summary(fit_All)$coef[, 4])
```

```
## cyl8
##    3
```

```
fitD <- D %>% select(-cyl); fitRm <- lm(mpg ~ ., data = fitD); summary(fitRm)$coef[, 4]
```

```
## (Intercept)        disp          hp        drat          wt        qsec
##   0.4158127   0.2145504   0.1357694   0.2914041   0.1020825   0.5372086
##         vs1         am1       gear4       gear5       carb2       carb3
##   0.5622658   0.4964455   0.8004203   0.5903340   0.7423912   0.5839796
##       carb4       carb6       carb8
##   0.7337118   0.8632349   0.6856502
```

```
which.max(summary(fitRm)$coef[, 4]) #the carb variable
```

```
## carb6
##    14
```

```
summary(fitRm)$coef[, 4]; which.max(summary(fitRm)$coef[, 4]) #the gear variable
```

```
## (Intercept)        disp          hp        drat          wt        qsec
##   0.4158127   0.2145504   0.1357694   0.2914041   0.1020825   0.5372086
##         vs1         am1       gear4       gear5       carb2       carb3
##   0.5622658   0.4964455   0.8004203   0.5903340   0.7423912   0.5839796
##       carb4       carb6       carb8
##   0.7337118   0.8632349   0.6856502
```

```
## carb6
##    14
```

# the vs variable

```
summary(fitRm)$coef[, 4]; which.max(summary(fitRm)$coef[, 4])
```

```
## (Intercept)        disp          hp        drat          wt        qsec
##   0.4158127   0.2145504   0.1357694   0.2914041   0.1020825   0.5372086
##         vs1         am1       gear4       gear5       carb2       carb3
##   0.5622658   0.4964455   0.8004203   0.5903340   0.7423912   0.5839796
##       carb4       carb6       carb8
##   0.7337118   0.8632349   0.6856502
```

```
## carb6
##    14
```

```
fitD <- fitD %>% select(-vs); fitRm <- lm(mpg ~ ., data = fitD)
```

# the drat variable

```
summary(fitRm)$coef[, 4]; which.max(summary(fitRm)$coef[, 4])
```

```
## (Intercept)        disp          hp        drat          wt        qsec
##   0.4482557   0.2506065   0.1532772   0.2807226   0.1001228   0.4077272
##         am1       gear4       gear5       carb2       carb3       carb4
##   0.5977575   0.6606758   0.4634669   0.5304308   0.7733536   0.5530478
##       carb6       carb8
##   0.9359528   0.8108093
```

```
## carb6
##    13
```

```
fitD <- fitD %>% select(-drat); fitRm <- lm(mpg ~ ., data = fitD)
```

# the disp variable

```
summary(fitRm)$coef[, 4]; which.max(summary(fitRm)$coef[, 4])
```

```
## (Intercept)        disp          hp          wt        qsec         am1
##  0.24378942  0.32186848  0.23268872  0.06274242  0.29117666  0.39348137
##       gear4       gear5       carb2       carb3       carb4       carb6
##  0.50023151  0.37483685  0.73841023  0.74098041  0.75340764  0.85737776
##       carb8
##  0.91471539
```

```
## carb8
##    13
```

```
fitD <- fitD %>% select(-disp); fitRm <- lm(mpg ~ ., data = fitD)
```

# the hp variable

```
summary(fitRm)$coef[, 4]; which.max(summary(fitRm)$coef[, 4])
```

```
## (Intercept)          hp          wt        qsec         am1       gear4
##  0.17333490  0.44184490  0.06589603  0.43637202  0.29963713  0.74306335
##       gear5       carb2       carb3       carb4       carb6       carb8
##  0.56757385  0.83792586  0.79011804  0.55391767  0.44018695  0.67555247
```

```
## carb2
##     8
```

```
fitD <- fitD %>% select(-hp); fitRm <- lm(mpg ~ ., data = fitD)
```

```
summary(fitRm)$coef[, 4]
```

```
## (Intercept)          wt        qsec         am1       gear4       gear5
##  0.24802943  0.01969676  0.23152066  0.27153772  0.58228588  0.59222913
##       carb2       carb3       carb4       carb6       carb8
##  0.85545551  0.69979185  0.44292140  0.41518282  0.32564877
```

# Clearly the p-values less than 0.05. three coefficients, wt, qseq, and Transmission

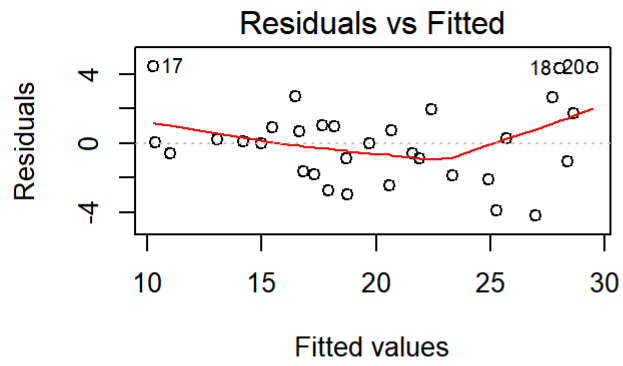
```
summary(fitRm)
```

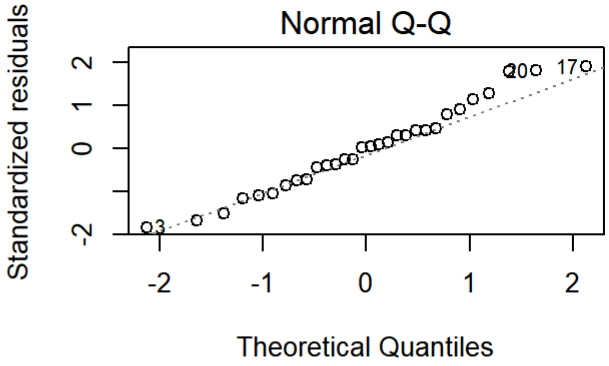
```
##
## Call:
## lm(formula = mpg ~ ., data = fitD)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1808 -1.6559  0.0333  1.0185  4.4595
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.2958    12.0317   1.188   0.2480
## wt           -2.9595     1.1722  -2.525   0.0197 *
## qsec          0.8164     0.6626   1.232   0.2315
## am1           2.7640     2.4477   1.129   0.2715
## gear4         1.5930     2.8513   0.559   0.5823
## gear5         1.7036     3.1322   0.544   0.5922
## carb2        -0.3547     1.9235  -0.184   0.8555
## carb3        -0.9960     2.5477  -0.391   0.6998
## carb4        -2.4592     3.1446  -0.782   0.4429
## carb6        -3.5205     4.2352  -0.831   0.4152
## carb8        -5.1181     5.0852  -1.006   0.3256
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.723 on 21 degrees of freedom
## Multiple R-squared:  0.8617, Adjusted R-squared:  0.7959
## F-statistic: 13.09 on 10 and 21 DF,  p-value: 6.494e-07
```

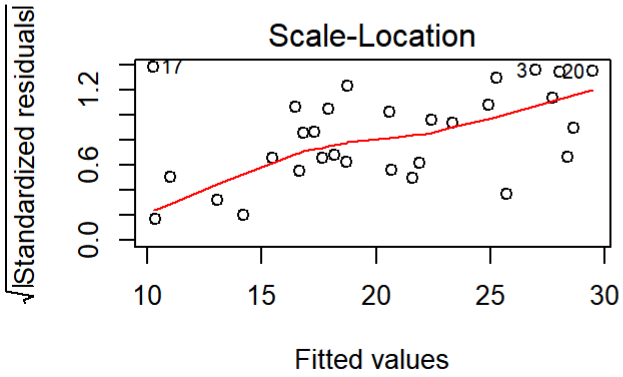
```
par(mfrow = c(2, 2))
plot(fitRm)
```

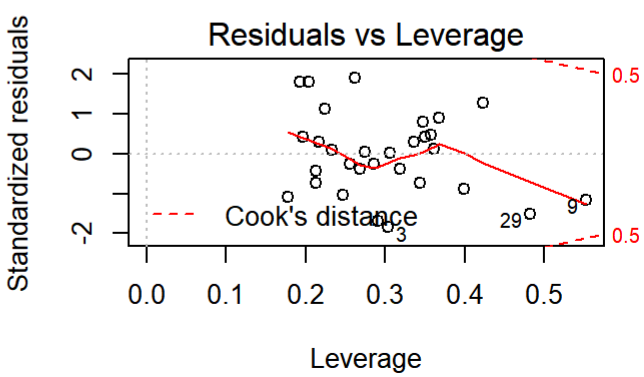
```
## Warning: not plotting observations with leverage one:
##   30, 31

## Warning: not plotting observations with leverage one:
##   30, 31
```




# The Conclusion:

```
the manual one will have an average of 2.9358 higher miles/gallon than the automatic car
```