C7W4 Assignment

Nic Neo

10 July 2020

Problem Statement

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:
1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions.

Describing the data

Source: mtcars description
The mtcars data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). This dataset consists of a data frame with 32 observations on 11 (numeric) variables.
[, 1] mpg Miles/(US) gallon
[, 2] cyl Number of cylinders
[, 3] disp Displacement (cu.in.)
[, 4] hp Gross horsepower
[, 5] drat Rear axle ratio
[, 6] wt Weight (1000 lbs)
[, 7] qsec 1/4 mile time
[, 8] vs Engine (0 = V-shaped, 1 = straight)
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears
[,11] carb Number of carburetors

Process the Data

```r
# Load the relevant data and libraries
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

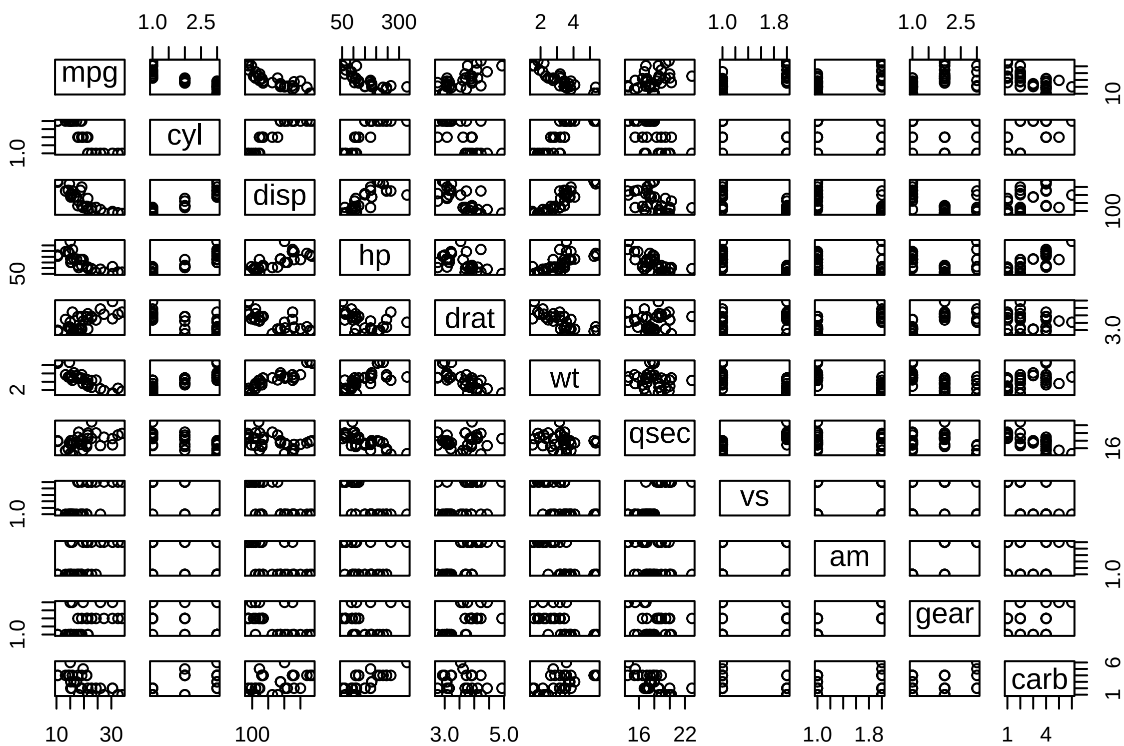
```r
library(MASS)
data(mtcars)
# Convert categorical variables to factors
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$am <- factor(mtcars$am, labels=c('Automatic','Manual'))
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
# inspect the variables in the dataset
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 "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
##  $ carb: Factor w/ 6 levels "1","2","3","4",..: 4 4 1 1 2 1 4 2 2 4 ...
```

Basic Exploratory Data Analysis

Check the relationships between the variables

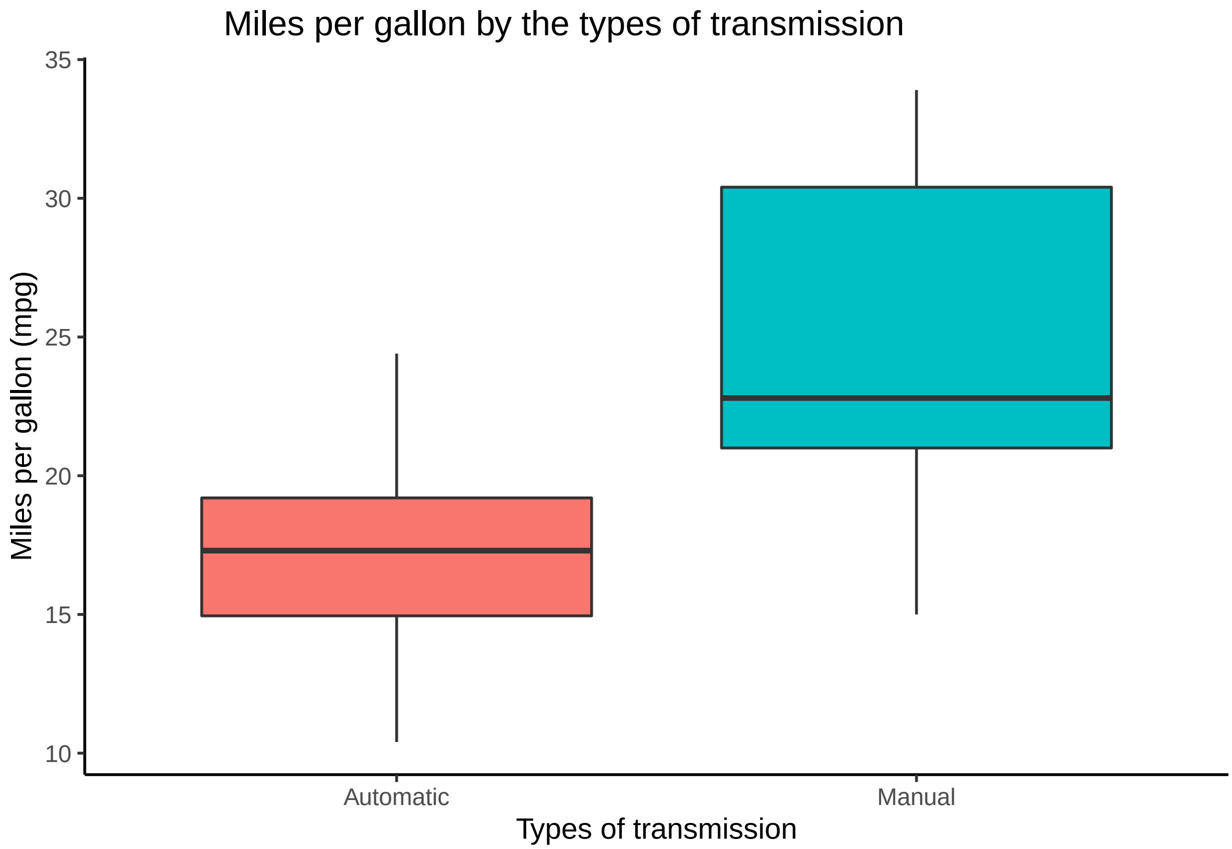
```r
pairs(mpg ~ ., data = mtcars)
```



Get rough gauge of the mean and spread of MPG by transmission type

```r
ggplot(mtcars, aes(x = factor(am), y = mpg)) +
  geom_boxplot(aes(fill = factor(am)), show.legend = FALSE) +
  labs(x = "Types of transmission",
       y = "Miles per gallon (mpg)",
```

```
      title = "Miles per gallon by the types of transmission"
      ) +
  scale_x_discrete(labels = c("Automatic","Manual")) +
  theme_classic() +
  theme(plot.title = element_text(hjust = 0.3))
```



Brief Conclusion from boxplot(w/o hypothesis testing yet):
Claim: Manual transmission seems to have a higher fuel efficiency (mpg).
Evidence to support claim: Test the hypthesis with a regression analysis

Regression Analysis

Fit linear regression model to see the relationship between type of transmission and the fuel efficiency (mpg)

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

```
##
## 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 ***
## amManual       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 R-squared estimate shows that only about 36% of the variation in mpg is explained by the model, indicating that we need to control other variables in the model. Hence needing the multivariable regression model.

Fit multivariable regression model with all variables using the `stepAIC()` function to determine which variables to include in the final model, with smaller AIC values indicating better model fit.

```r
# Fit model
multiModel <- lm(mpg ~ ., data = mtcars)
# Select model
bestFit <- stepAIC(multiModel, direction = "both")
```

```r
summary(bestFit)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

We see that the final model consists of `cyl`, `hp` and `wt` as covariates, with `mpg` as outcome and `am` as predictor. In addition, about 84% of variance in mpg is now explained by the model's adjusted R-squared value, indicating better model fit.

Inference

1. The expected mpg with manual transmission is about 1.81 greater than auto transmission.

2. However, the p-value for manual versus auto transmission (reference) is about 0.206, much greater than the standard threshold of 0.05. Hence, we cannot make that conclusion that the manual transmission results in better mpg, compared to auto transmission.

Confirm claims with confidence interval

```
confint(bestFit, "amManual", level = 0.95)
```

```
##              2.5 %   97.5 %
## amManual -1.060934 4.679356
```

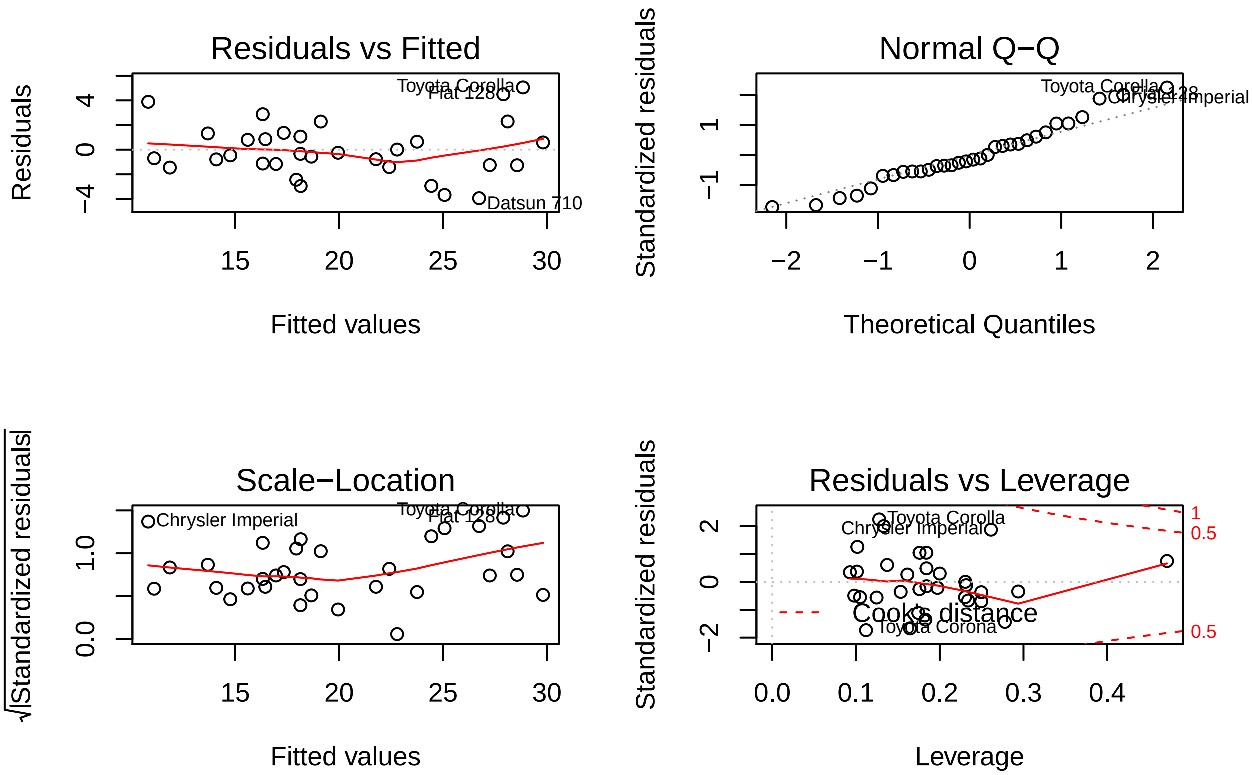
The confidence interval overlaps with 0, indicating non-signficance.

Conclusions

The difference in mpg between auto and manual transmission is not statistically significant. We cannot conclude that either type of transmission has better mpg than the other.

Appendix - Diagnostics

```
# Plot residuals
par(mfrow = c(2, 2))
plot(bestFit)
```




```r
# Top 3 most influential points on slope coefficients
influence <- dfbetas(bestFit)
head(sort(influence[,6], decreasing = TRUE), 3)
```

```
##      Toyota Corona          Fiat 128 Chrysler Imperial
##          0.7305402         0.4292043         0.3507458
```