Motor Trend Analysis

2019-05-13

1. Summary

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"

2. Exploratory Data Analyses

```
data(mtcars)
dim(mtcars)
```

```
## [1] 32 11
```

As we can see, the *mtcars* data set has 11 variables and 32 observations.

Based on the data description, *am*, *vs*, *cyl*, *gear*, *carb* are discrete. (Code is in the appendix part 1)

In order to explore the relationship between a set of variables and MPG, we can draw scatterplots for each two variables. (In the appendix part 2)

3. Is an automatic or manual transmission better for MPG

In order to find whether there is difference between automatic and manual transmission for MPG, we can draw a boxplot to show it. (In the appendix part 3)

We can also fit a linear regression model to check it.

```
fit1 <- lm(mpg ~ am, mtcars)
(fit1_summ <- summary(fit1))
```

```
##
## 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 p-value is smaller than 0.05, so we reject the null hypothesis, which means there exists difference between automatic and manual transmissions. Actually, the coefficient of *Manual* is 7.2449, which means manual transmission has 7.2449 MPG larger than automatic, thus the automatic is better than the manual.

However, we notice that the R-squared is only 0.3598, which means only 35.98% variance in MPG can be explained by transmission types. Therefore, we will consider to add more variables into the linear model.

4. Quantify the MPG difference between automatic and manual transmissions

First of all, fit a full model. Then we will do a stepwise regression for it, where the choice of predictor is carried out automatically by comparing AIC (the smaller AIC, the better model). Details are in the appendix part 4.

```
fit2 <- lm(mpg ~ ., mtcars)
fit3 <- step(fit2, trace=0)
(fit3_summ <- summary(fit3))
```

```
##
## 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
```

Finally, we get a best model with four predictors: number of cylinders, gross horsepower, weight, and transmission. The R-squared shows there is 86.59% variance can be explained, which is good enough.

The diagnostic plots also show this model is good. (In the appendix part 5)

If other variables are fixed, then the MPG will decrease 0.03211 as gross horsepower increases one unit, the MPG will decrease 2.49683 as weight increases 1000 lbs, the manual transmission will lead the MPG has 1.89921 higher than automatic transmission. As for number of cylinders, compring to 4 cylinders, 6 cylinders will decrease 3.03134 MPG, and 8 cylinders will decrease 2.16368 MPG.

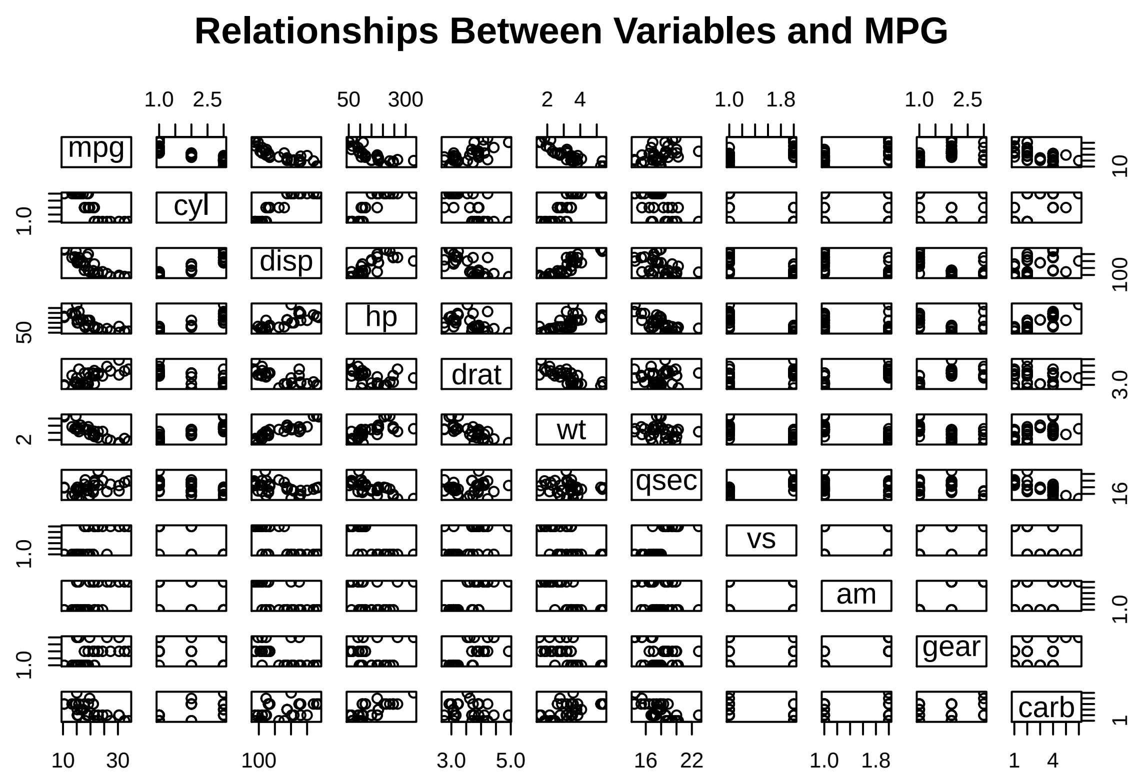
Appendix

1. Code for transform numeric variables to factor variables

```
mtcars$am <- factor(mtcars$am, levels=0:1, labels = c("Automatic", "Manual"))
mtcars$vs <- factor(mtcars$vs, levels=0:1, labels = c("Vshaped", "Straight"))
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb)
```

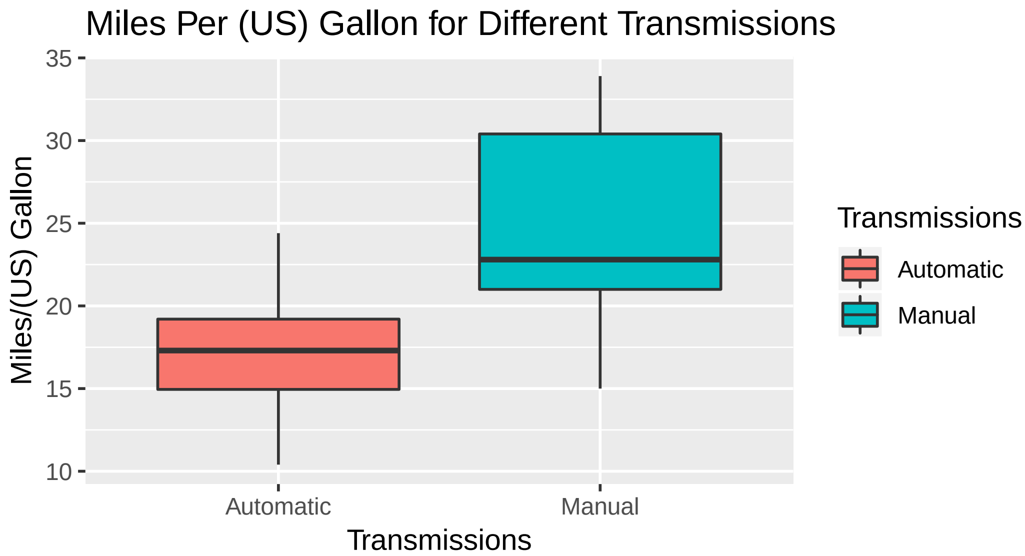
2. Scatterplots for each two variables.

```
pairs(mpg ~ ., data = mtcars, main="Relationships Between Variables and MPG")
```



3. Boxplot for the difference between automatic and manual transmission for MPG.

```
ggplot(mtcars, aes(x=factor(am), y=mpg, fill=factor(am))) + geom_boxplot() +
    labs(x="Transmissions", y="Miles/(US) Gallon",
         title="Miles Per (US) Gallon for Different Transmissions") +
    guides(fill=guide_legend(title="Transmissions"))
```



As the boxplot shows, the automatic transmission is significant better than manual transmission for MPG.

4. Details for stepwise regression.

```
step(fit2)
```

```
## Start:  AIC=76.4
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##        Df Sum of Sq    RSS    AIC
## - carb  5   13.5989 134.00 69.828
## - gear  2    3.9729 124.38 73.442
## - am    1    1.1420 121.55 74.705
## - qsec  1    1.2413 121.64 74.732
## - drat  1    1.8208 122.22 74.884
## - cyl   2   10.9314 131.33 75.184
## - vs    1    3.6299 124.03 75.354
## <none>              120.40 76.403
## - disp  1    9.9672 130.37 76.948
## - wt    1   25.5541 145.96 80.562
## - hp    1   25.6715 146.07 80.588
##
## Step:  AIC=69.83
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear
##
##        Df Sum of Sq    RSS    AIC
## - gear  2    5.0215 139.02 67.005
## - disp  1    0.9934 135.00 68.064
## - drat  1    1.1854 135.19 68.110
## - vs    1    3.6763 137.68 68.694
## - cyl   2   12.5642 146.57 68.696
## - qsec  1    5.2634 139.26 69.061
## <none>              134.00 69.828
```

```
## - am    1   11.9255 145.93 70.556
## - wt    1   19.7963 153.80 72.237
## - hp    1   22.7935 156.79 72.855
##
## Step:  AIC=67
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - drat  1    0.9672 139.99 65.227
## - cyl   2   10.4247 149.45 65.319
## - disp  1    1.5483 140.57 65.359
## - vs    1    2.1829 141.21 65.503
## - qsec  1    3.6324 142.66 65.830
## <none>              139.02 67.005
## - am    1   16.5665 155.59 68.608
## - hp    1   18.1768 157.20 68.937
## - wt    1   31.1896 170.21 71.482
##
## Step:  AIC=65.23
## mpg ~ cyl + disp + hp + wt + qsec + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - disp  1    1.2474 141.24 63.511
## - vs    1    2.3403 142.33 63.757
## - cyl   2   12.3267 152.32 63.927
## - qsec  1    3.1000 143.09 63.928
## <none>              139.99 65.227
## - hp    1   17.7382 157.73 67.044
## - am    1   19.4660 159.46 67.393
## - wt    1   30.7151 170.71 69.574
##
## Step:  AIC=63.51
## mpg ~ cyl + hp + wt + qsec + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - qsec  1     2.442 143.68 62.059
## - vs    1     2.744 143.98 62.126
## - cyl   2    18.580 159.82 63.466
## <none>              141.24 63.511
## - hp    1    18.184 159.42 65.386
## - am    1    18.885 160.12 65.527
## - wt    1    39.645 180.88 69.428
##
## Step:  AIC=62.06
## mpg ~ cyl + hp + wt + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - vs    1     7.346 151.03 61.655
## <none>              143.68 62.059
## - cyl   2    25.284 168.96 63.246
## - am    1    16.443 160.12 63.527
## - hp    1    36.344 180.02 67.275
## - wt    1    41.088 184.77 68.108
##
```

```
## Step:  AIC=61.65
## mpg ~ cyl + hp + wt + am
##
##        Df Sum of Sq    RSS    AIC
## <none>              151.03 61.655
## - am    1     9.752 160.78 61.657
## - cyl   2    29.265 180.29 63.323
## - hp    1    31.943 182.97 65.794
## - wt    1    46.173 197.20 68.191

##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Coefficients:
## (Intercept)         cyl6         cyl8           hp           wt
##    33.70832     -3.03134     -2.16368     -0.03211     -2.49683
##    amManual
##     1.80921
```

5. Diagnostic plots for the final model.

```r
par(mfrow = c(2, 2))
plot(fit3)
```