Regression Models

Alex MacCalman

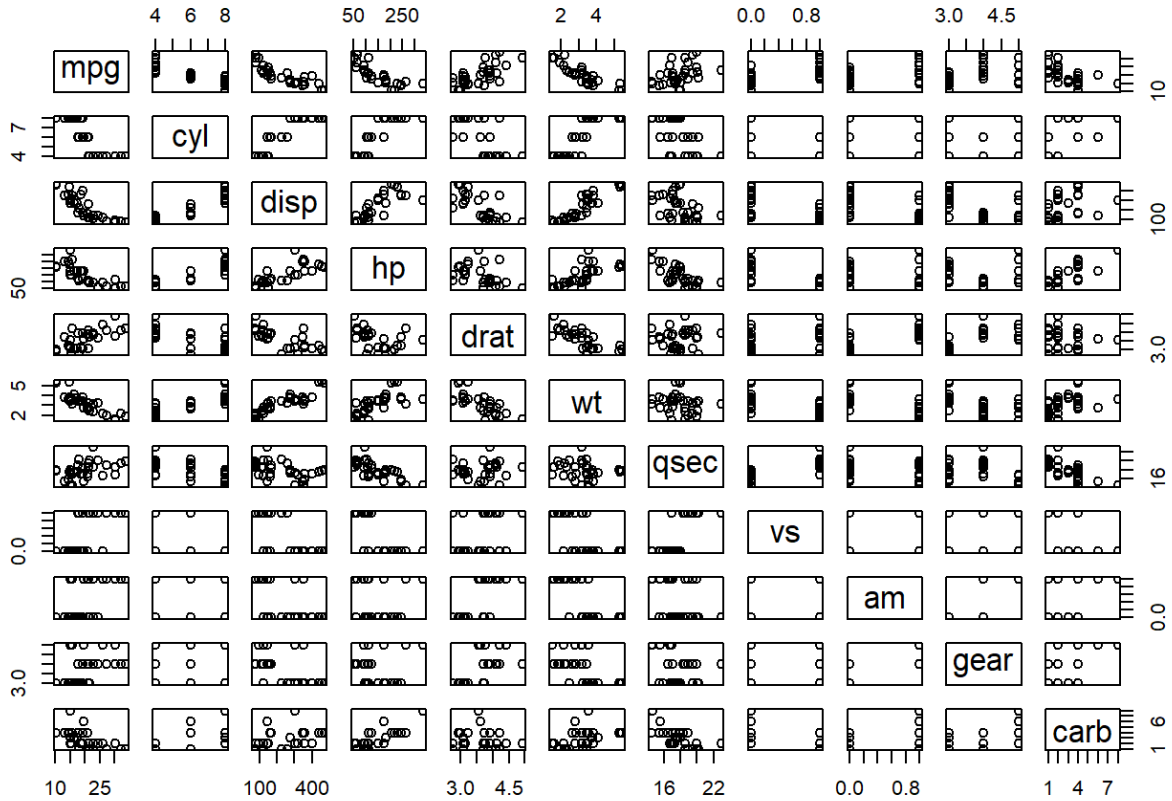
5/28/2020

The report analyzes the Motor Trend Car Road Tests (mtcars) data set. The 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). we will answer the following two questions:
“Is an automatic or manual transmission better for MPG”
“Quantify the MPG difference between automatic and manual transmissions”

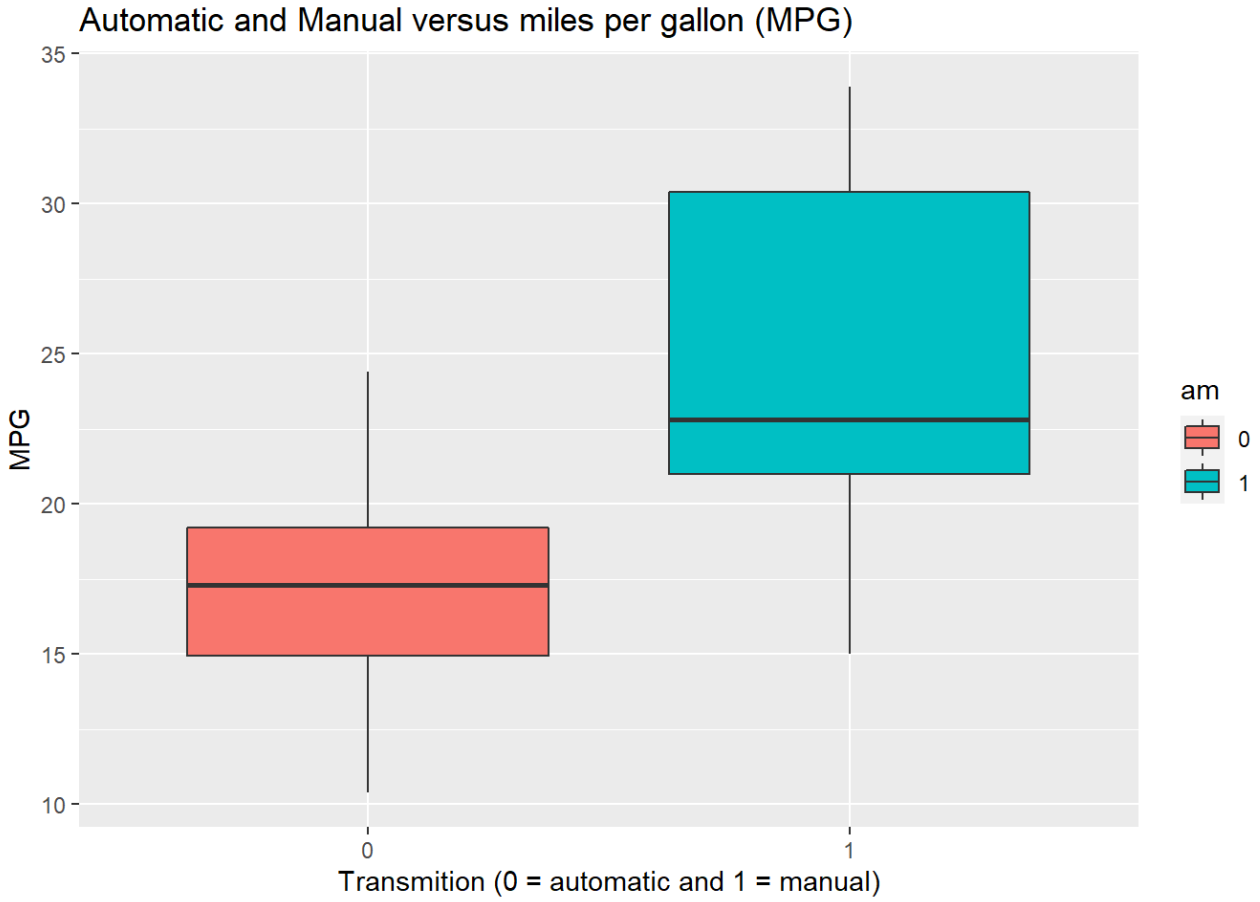
First we will load libraries.

Now we load the data and look at a scatter plot matrix and blox plot of am where o = automatic and 1 = manual.

```
df <- mtcars
pairs(~., df)
```



```
df$am <- as.factor(df$am)
ggplot(df, aes(x=am, y=mpg)) +geom_boxplot(aes(fill=am)) +
  labs(title = "Automatic and Manual versus miles per gallon (MPG)") + xlab("Tr
ansmition (0 = automatic and 1 = manual)") +
  ylab("MPG")
```



Here we perform a t test to see if there is a significant difference between tranmission types.

```
auto <- df %>% select(mpg, am) %>% filter(am == "0")
man <- df %>% select(mpg, am) %>% filter(am == "1")
t.test(auto$mpg, man$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  auto$mpg and man$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
```

The t test indicates that the true difference in the means between automatic and manual transmission effect on mpg is not zero. There are better mpg for manual transimision. Nopw we'll fit some linear models.

```
fit <- lm(mpg~am, df)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = df)
##
## 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 ***
## am1            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
```

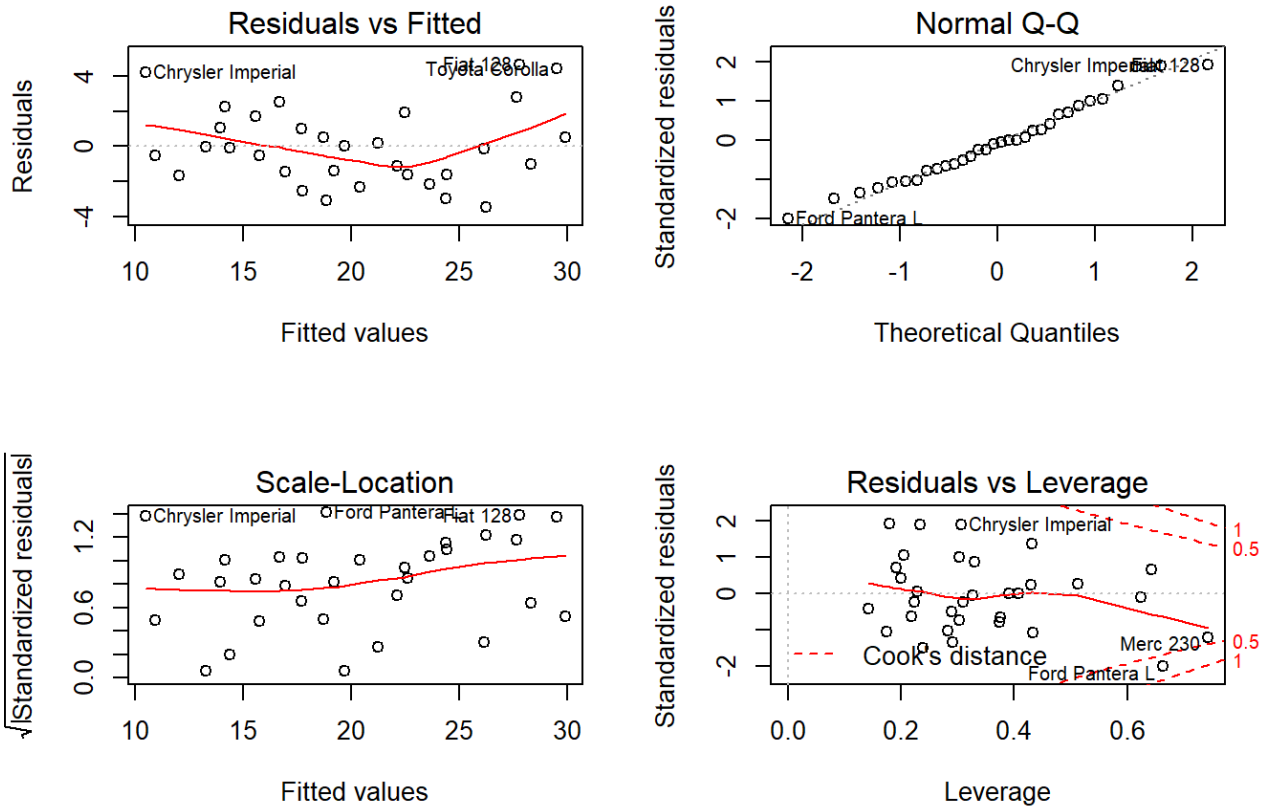
We can see that the manual transmision has a 7.245 mpg better performace than the automatic transition. Included only the am variable explains ~ 36% of the variable. Next we will fit all variables to then see the impact of am in the presence of the other variables.

```
fit2 <- lm(mpg~., df)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am1          2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

In the presence of the other variables and while they are held constant, automatic only has a 2.52 mpg improvement over manual. Now we will take a look at the residuals of fit2.

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



The plots show that the data is normal and that there is no heteroscedasticity.