# Regression Models - Project 2

Kalyani

1/11/2022

Executive 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:

"Is an automatic or manual transmission better for MPG"

"Quantify the MPG difference between automatic and manual transmissions"

Data Processing/briefly analysis

```
library(car)
```

```
## Loading required package: carData
```

```
mtcars$am = as.factor(mtcars$am)
levels(mtcars$am) = c("Automatic", "Manual")
summary(mtcars$mpg)
```

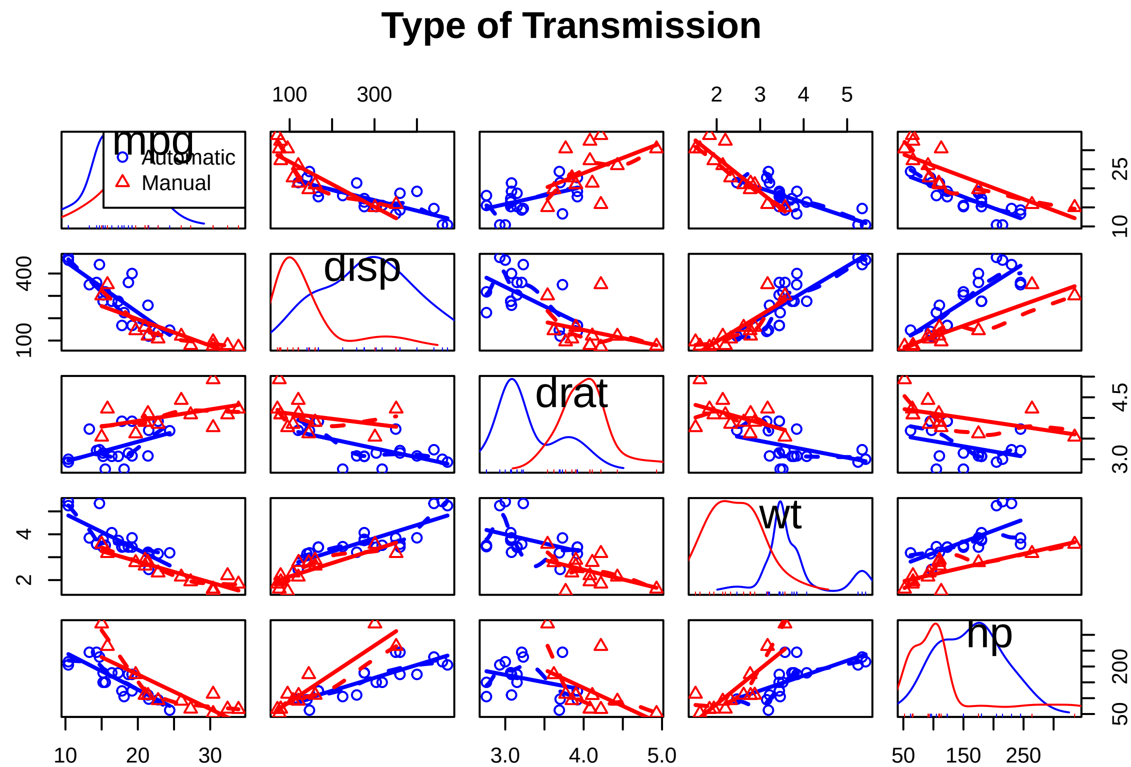
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   15.43   19.20   20.09   22.80   33.90
```

```
summary(mtcars$am)
```

```
## Automatic    Manual
##        19        13
```

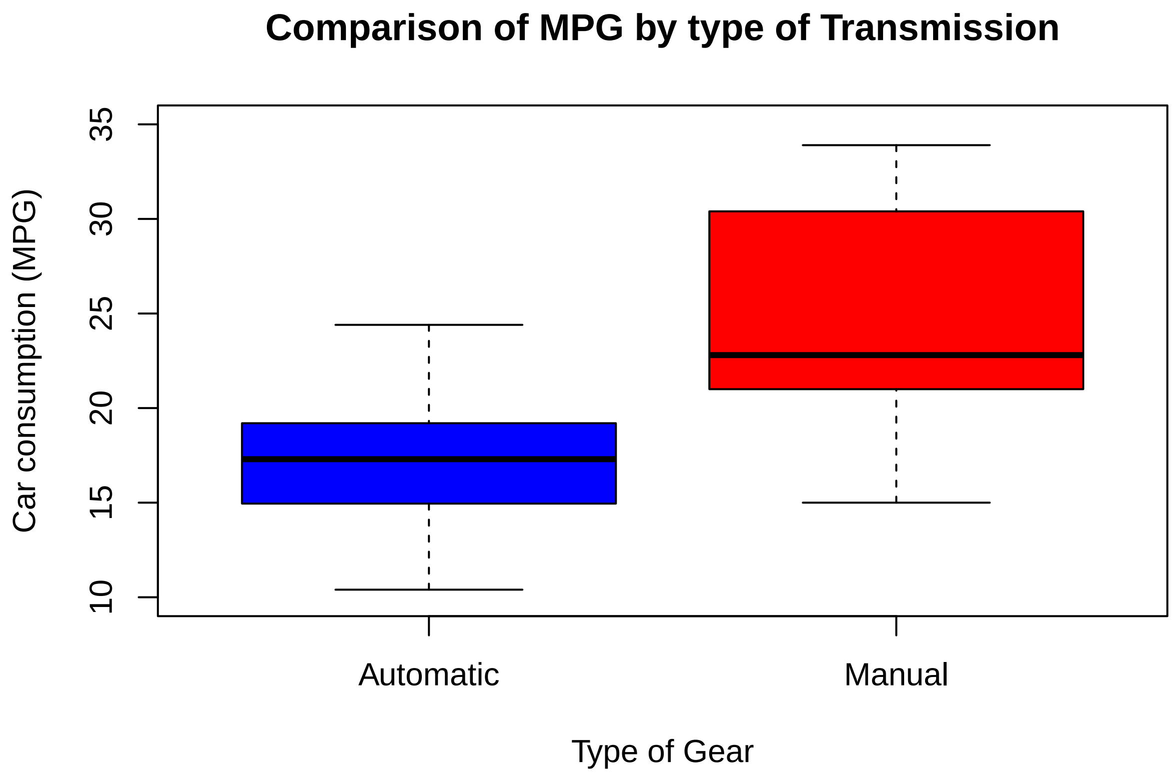
#Exploratory analysis A Scatterplot to check the relations of MPG variables

```
scatterplotMatrix(~mpg+disp+drat+wt+hp|am, data=mtcars,
                  col = c("blue", "red"),
                  main="Type of Transmission")
```




Boxplot to check what the type of transmission could be more efficiently to gas consumption.

```
amboxplot = boxplot(mpg ~ am, data=mtcars, main = "Comparison of MPG by type of Transmission",
              xlab = "Type of Gear",
              ylab = "Car consumption (MPG)",
              ylim = c(10, 35),
              col = c("blue", "red"))
```




A correlation analysis of MPG variable and others.

```
cor(mtcars[, -c(9)])[1, ]
```

```
##        mpg        cyl       disp         hp       drat         wt       qsec
##  1.0000000 -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594  0.4186840
##         vs       gear       carb
##  0.6640389  0.4802848 -0.5509251
```

Linear models Fit 1 - am

```
fit1 = lm(mpg ~ am, data=mtcars)
rmse1 = sqrt(sum(fit1$residuals ^ 2) / nrow(mtcars))
rsq1 = summary(fit1)$r.squared
```

Fit 2 - mpg

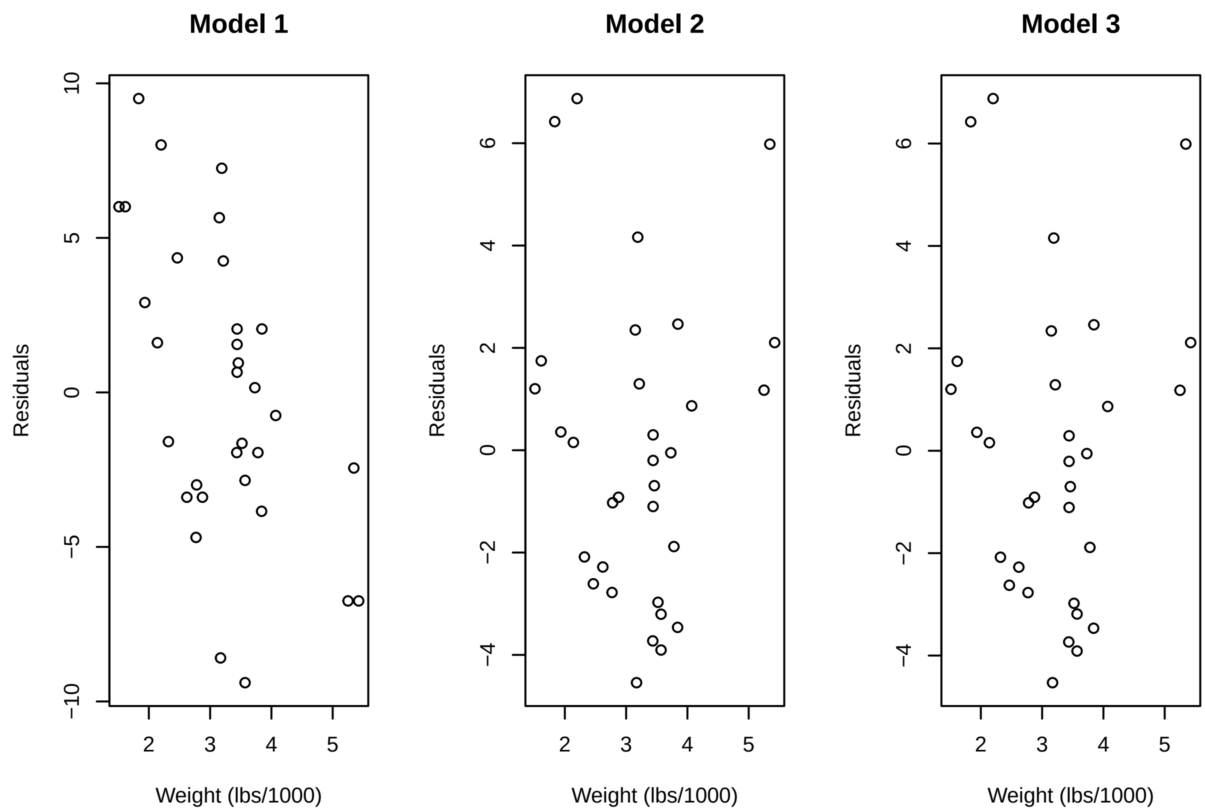
```
fit2 = lm(mpg ~ wt, data=mtcars)
rmse2 = sqrt(sum(fit2$residuals ^ 2) / nrow(mtcars))
rsq2 = summary(fit2)$r.squared
```

Fit 3 - am+wt

```
fit3 = lm(mpg ~ am + wt, data=mtcars)
rmse3 = sqrt(sum(fit3$residuals ^ 2) / nrow(mtcars))
rsq3 = summary(fit3)$r.squared
```

#Residual Analysis Residuals Plot

```
par(mfcol = c(1, 3))
plot(mtcars$wt, resid(fit1), main = "Model 1", xlab = "Weight (lbs/1000)", ylab = "Residuals")
plot(mtcars$wt, resid(fit2), main = "Model 2", xlab = "Weight (lbs/1000)", ylab = "Residuals")
plot(mtcars$wt, resid(fit3), main = "Model 3", xlab = "Weight (lbs/1000)", ylab = "Residuals")
```




The residuals for model 1 exhibit a linear pattern. Model 1 has larger residuals than the other two models.The residuals of models 2 and 3 are almost identical.

««««< table of comparison - Root Mean Squared Error

```
print(paste('Model 1 = ', rmse1))
```

```
## [1] "Model 1 =  4.74636900427014"
```

```
print(paste('Model 2 = ', rmse2))
```

```
## [1] "Model 2 =  2.94916268595503"
```

```
print(paste('Model 3 = ', rmse3))
```

```
## [1] "Model 3 =  2.94915081645569"
```

Table of comparisation - R2

```
print(paste('Model 1 = ', rsq1))
```

```
## [1] "Model 1 =  0.359798943425465"
```

```
print(paste('Model 2 = ', rsq2))
```

```
## [1] "Model 2 =  0.752832793658264"
```

```
print(paste('Model 3 = ', rsq3))
```

```
## [1] "Model 3 =  0.752834783202689"
```

Model 1 does not predict MPG very well, models 2 and 3 have very similar performance characteristics. The R2 values reveals the fact that adding the transmission type to model 2 does not add any predictive power.

Coefficients

```
coef2 = summary(fit2)$coef
coef2
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 37.285126   1.877627 19.857575 8.241799e-19
## wt          -5.344472   0.559101 -9.559044 1.293959e-10
```

Coefficient Model 3:

```
coef3 = summary(fit3)$coef
coef3
```

```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## amManual    -0.02361522  1.5456453 -0.01527855 9.879146e-01
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```

```
transmission_ci <- coef3[2, 1] + c(-1, 1) * qt(.975, df = fit3$df) * coef3[2, 2]
transmission_ci
```

```
## [1] -3.184815  3.137584
```

Conclusion: The 95% confidence interval for this coefficient is rather large compared to its estimated value, namely (-3.1848, 3.1376). To provide a basis for comparison, an increase in weight of 1000 lbs would lower the MPG by an average of 5.3528.