Project

We are here to answer 2 questions:

- "Is an automatic or manual transmission better for MPG"
- "Quantify the MPG difference between automatic and manual transmissions"

To answer this questions we'll be using data from *mtcars data set in R project*

So let's rename the variables first:

```
y <- mtcars$mpg
x <- mtcars$am
z <- (1-mtcars$am)
```

Now let's split the data in 2 (manual and automatic) and make a explanatory analyses

Calling manual data as

```
manual_mpg<-y*x
manual_mpg<-manual_mpg[manual_mpg!=0] #getting values only when cars are manual
n_manual<-length(manual_mpg) #calculating the sample size
```

Doing the same for automatic cars we have:

```
automatic_mpg<-y*z
automatic_mpg<-automatic_mpg[automatic_mpg!=0]
n_automatic = length(automatic_mpg)
```

Manual Explanatory Statistics

```
summary(manual_mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   21.00   22.80   24.39   30.40   33.90
```

Standard Deviation

```
sd(manual_mpg)
```

```
## [1] 6.166504
```

Automatic Explanatory Statistics

```
summary(automatic_mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   14.95   17.30   17.15   19.20   24.40
```

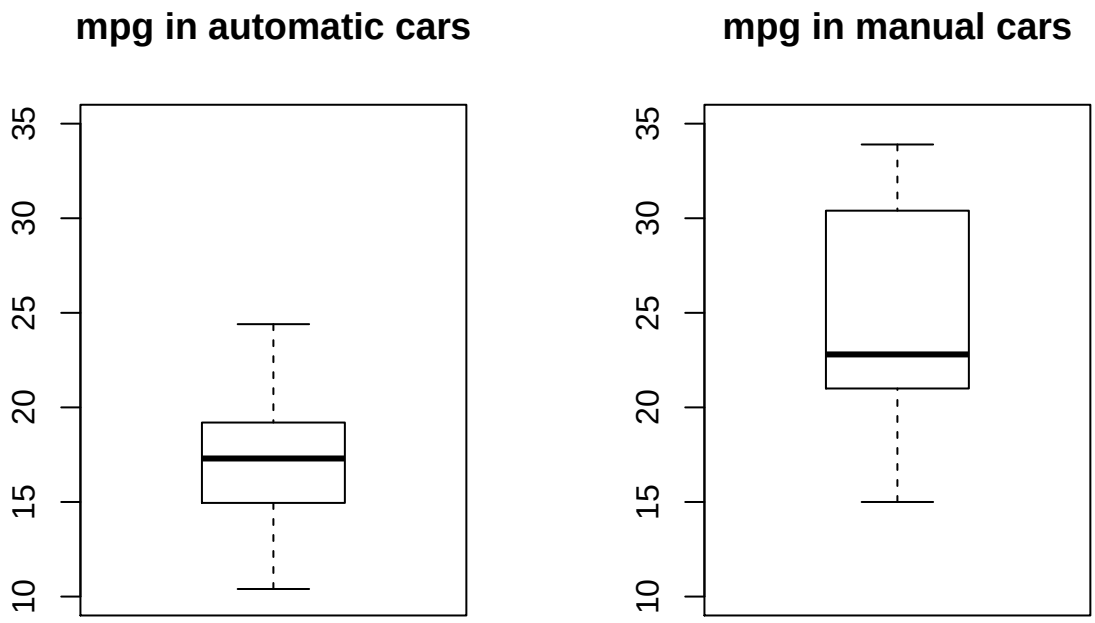
Standard Deviation

```r
sd(automatic_mpg)
```

```
## [1] 3.833966
```

Comparing both sets of data:

```r
par(mfcol=c(1,2))
boxplot(automatic_mpg, main = "mpg in automatic cars",ylim = c(10, 35))
boxplot(manual_mpg, main = "mpg in manual cars",  ylim = c(10, 35))
```



We can clearly see from the data above that:

- Automatic cars have a more consistent performance, since the variation is lower
- Manual cars have a better performance in general. We can see that the manual car having the poorest performance for mpg is almost as good as the average for automatic cars. Also in average automatic is better (we have 17.1 mpg for automatic and 24.4 for manual).
- Checking the data as quartiles we have 75% of manual cars with performance lower than 30.4 mpg, while for automatic the same 75% is lower than 19.2

So based on the affirmative above, we can say that in general manual cars are a little better than automatic, but we still don't know how better, or if there's any chance of not being always true.

So let's solve this:

Let's try to predict mpg based on information about being automatic or not. For this we'll use a multivariate linear regression model where our outcome is mpg performance and our output is a binary variable (1 if manual, 0 if automatic)

```r
fit<-lm(y~ x + z)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x + z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## x              7.245      1.764   4.106 0.000285 ***
## z                 NA         NA      NA       NA
## ---
## 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
```

```r
sumCoef <- summary(fit)$coefficients
#sumCoef[1,1] + sumCoef[2,1] + c(-1, 1) * qt(.975, df = fit$df) * sumCoef[2, 2]
#sumCoef[1,1] + c(-1, 1) * qt(.975, df = fit$df) * sumCoef[2, 2]
```

From the output above, we have a model that instead of having 2 betas, has only one but considering both singularities as it says in the output. Therefore our model can be written as:

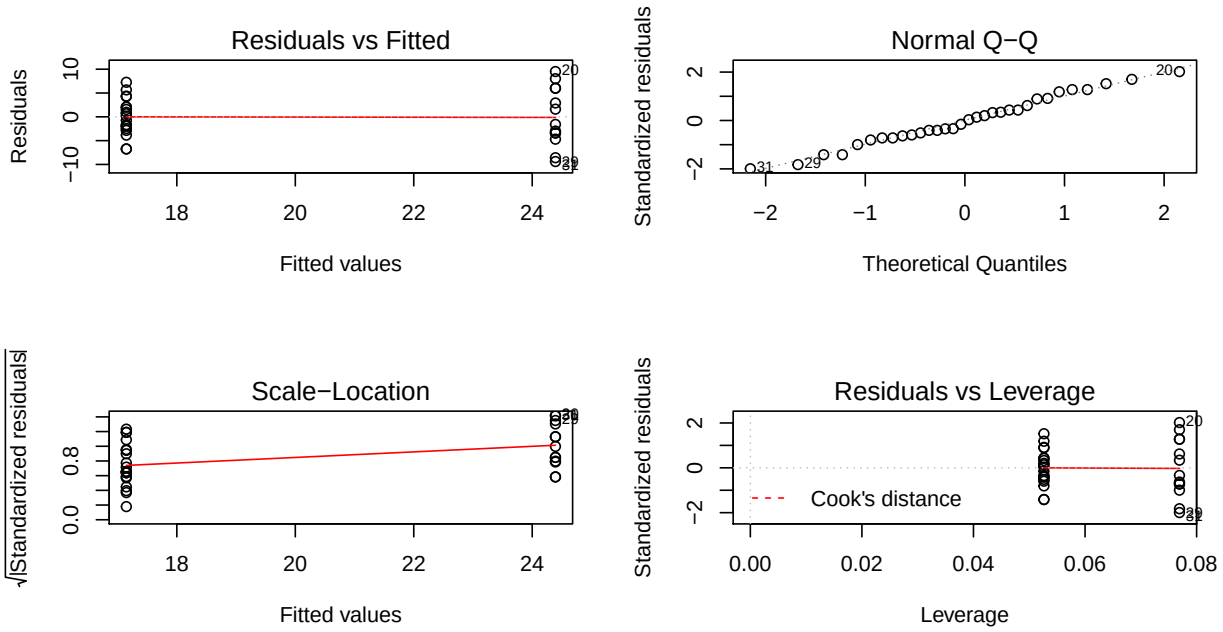
mpg = 17.15 + 7.24x

For manual cars we have mpg = 24.39 , while for automatic cars, we have mpg = 17.15.

Both Betas have passed the T- test , ie, they both have a small p-value, lower than 0.5%.

When we check the residuals, we have:

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



So based on data above we can see that the model fits well since the residuals follow a normal distribuition and they are consistent as well. The standard variario is almost zero (if you considere that there are two poles.)

Let's try exclude the intercept to see if it changes anything about the coefficients:

```
fit2<-lm(y~ x + z - 1)
summary(fit2)
```

```
##
## Call:
## lm(formula = y ~ x + z - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##   Estimate Std. Error t value Pr(>|t|)
## x   24.392      1.360   17.94  < 2e-16 ***
## z   17.147      1.125   15.25 1.13e-15 ***
## ---
## 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:  0.9487, Adjusted R-squared:  0.9452
## F-statistic: 277.2 on 2 and 30 DF,  p-value: < 2.2e-16
```

```
sumCoef <- summary(fit2)$coefficients
```

And we have the same model for obvious reasons. We didn't really exclude the intercept, we just called it beta2.

So based on descriptive analyses and on the regression model we can conclude that the manual cars have a better transmission for mpg than automatic cars.

To quantify this difference, I'd use the odd ratio that in this case is

beta(for manual cars)/beta(automatic cars) =

```
sumCoef[1,1]/sumCoef[2,1]
```

```
## [1] 1.42251
```

So manual cars have a performance for mpg 42% better in average compared to automatic cars