Independent t-test

The independent t test is used to test if there is any statistically significant difference between two means. Use of an independent t test requires several assumptions to be satisfied. The assumptions are listed below

* The variables are continuous and independent
* The variables are normally distributed
* The variances in each group are equal

When these assumptions are satisfied the results of the t test are valid. Otherwise they are invalid and you need to use a non-parametric test. When data is not normally distributed you can apply transformations to make it normally distributed.

For this set of exercises we will use a motor trend car road tests data set. This data is already available in R as [mtcars](https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html). The data consists of fuel consumption and vehicle characteristics related to design and the level of performance. Our interest in this exercise is to test if there are any significant differences in miles per gallon achieved between manual and automatic transmission vehicles.

## 1. Inspect the structure of the data

#Display the first few rows of the data to understand its structure  
head(mtcars)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

## 2. Label the am (0,1) variable into automatic and manual categories

# The am variable is coded as 0/1. Give it meaningful labels  
mtcars$am = factor(mtcars$am, levels = c(0,1), labels = c("Automatic", "Manual"))

## 3. Check data labeling was successful

# Check the labelling was successful  
head(mtcars)

## mpg cyl disp hp drat wt qsec vs am gear  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 Manual 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 Manual 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 Manual 4  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 Automatic 3  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 Automatic 3  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 Automatic 3  
## carb  
## Mazda RX4 4  
## Mazda RX4 Wag 4  
## Datsun 710 1  
## Hornet 4 Drive 1  
## Hornet Sportabout 2  
## Valiant 1

## 4. Attach mtcars data so that its variables are easily accessible

# Make the mtcars columns easily available  
attach(mtcars)

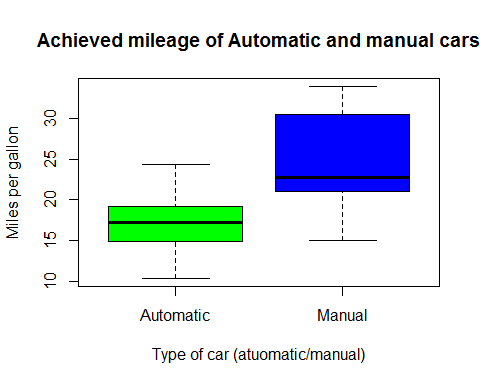
## 5. Generate descriptive statistics for each group

# Calculate mean for each category of cars  
aggregate(mpg~am,data=mtcars,FUN = mean)

## am mpg  
## 1 Automatic 17.14737  
## 2 Manual 24.39231

## 6. Generate box plot for each group

# Plot box plots for each category to understand data distribution and identify outliers   
boxplot(mpg~am,main = "Achieved mileage of Automatic and manual cars", xlab = "Type of car (atuomatic/manual)",col = (c("green","blue")), ylab = "Miles per gallon")



## 7. Apply a log transformation to stabilize data variance

#Take the log of mpg data to stabilize variance  
log.transformed.mpg = log(mpg)

## 8. Perform a t test on the transformed variable

#Perform a t test assuming equal variance  
t.test(log.transformed.mpg~am,var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: log.transformed.mpg by am  
## t = -3.9087, df = 30, p-value = 0.0004905  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.5277597 -0.1655209  
## sample estimates:  
## mean in group Automatic mean in group Manual   
## 2.816692 3.163332

## 9. Interpret the results

* Manual cars had a higher mean miles per gallon at 24.4 as compared to automatic cars which had a mean miles per gallon of 17.1
* The box plots did not reveal presence of any outliers
* The t test showed there was a significant difference in miles per gallon between manual and automatic cars. Manual cars give better mileage Interpret the results