**STAT 3021 Practical Assignment Part 3**

Group Members: Linxin Li, Zhenyu Fan

**Introduction**

On March, we went to an auto show where exhibited lots of cool cars. And one of us want to buy a car. However, keeping a car is expensive, and therefore we found some information on the car’s power system and Fuel consumption. Specifically, we want to identify how is the number of cylinders in a car could affect the fuel consumption.

In our discussions about the project, we believe that the more cylinders a car has, the less fuel it consumes. Because when a car has more cylinders, it uses the fuel more efficiently. Statistically, this would mean that if μ4 = the mean fuel use of a car with 4 cylinders and μ6 = the mean fuel use of a car with 6 cylinders. I hypothesize that

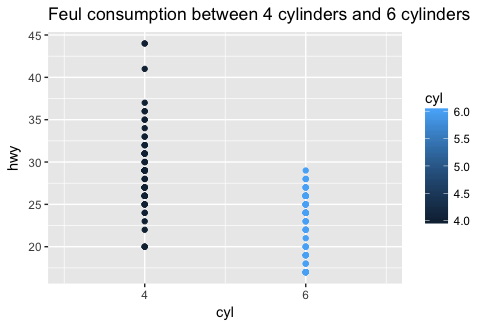
H0: μ4 = μ6

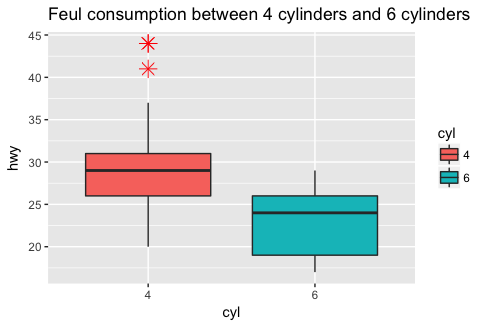
Hα: μ4 > μ6

To answer the above hypothesis, we gathered data by searching from website.

**Graphs and Descriptive Statistics**

First, we obtained a car sample with several independent observations, and we divided the sample into 2 parts based on their number of cylinders. Then we need to conduct a 2-sample T test to see if these 2 parts can actually be from the same fuel consumption population.

The dot plot below gives a comparison of the cost of oil on highway mile per gallon for 4 cylinders and 6 cylinders.

In addition, the box plot below gives a comparison of the consumption of fuel for cars with difference number of cylinders. Clearly, the median of 4 cylinder cars seems larger than the 6 cylinder cars. In addition, there is smaller variability in 4 cylinder cars since the range and the IQR both seem smaller even with 2 outliers above.

Numerical summary statistics allow us to compare data set in details.

$`4` 4 cylinder cars

Min. 1st Qu. Median Mean 3rd Qu. Max.

20.0 26.0 29.0 28.8 31.0 44.0

$`6` 6 cylinder cars

Min. 1st Qu. Median Mean 3rd Qu. Max.

17.00 19.00 24.00 22.82 26.00 29.00

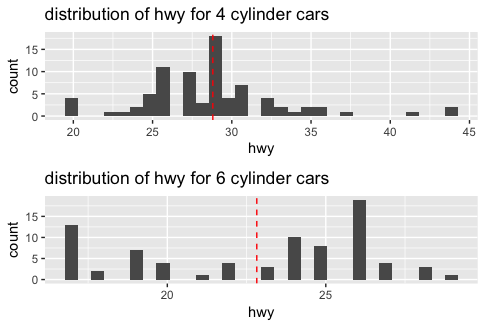
**Verifying necessary data conditions**

To conduct a two samples t-test for means, the two conditions that are required are:

1. both samples should be independent and random

2. both populations should be approximately normally distributed

The way I collected data, I believe the two samples are independent, since there should be no relation between the information i collect from a certain car to another one. Also, I have random samples, since I collected a very board brands of cars manufactured all around the world.

To check the second condition, I created 2 histogram to see if they are normally distributed. The graph for 4 cylinder cars are in a pretty good shape. However 6 cylinder cars are in normal distribution shape with only a huge pike on the left. Therefore, with a little uncertainty, I would conclude it is normally distributed and proceed the analysis with caution in mind.

(red dashed line is the mean of the population)

**Conducting a Hypothesis Test**

We can to check whether the consumption of fuel for 6 cylinder cars are truly less than that of 4 cylinder cars, we need to conduct a t test. With

H0: μ4 = μ6

Hα: μ4 > μ6

with alpha value = 0.05

Then, list my current statistics:

mean for 4 cylinder cars : 29

mean for 6 cylinder cars : 24

standard deviation for 4 cylinder cars : 4.51

standard deviation for 4 cylinder cars : 3.68

number of observations for 4 cylinder cars : 81267213.png

number of observations for 4 cylinder cars : 79

By using the formula of 2 sample t test:

We can calculate our t value which is about 7.69. Then, with this t value, we can find the possibility that our null hypothesis is correct. This p value is significantly small (about 8.2\*e-13)

**Conclusion and Summary**

Although not what I expected at first, but such small p value is definitely less than our alpha value. We therefore, based on our statistics, can conclude that we did not fail the null hypothesis and these 2 samples are likely from the same population.