**LY6010 MODULE 3**

**WEEK 3**

**Name : shreyansh Bhalodiya**

**NUID : 002664707**

**Dataset Selected:**

****

**Dimensions of Dataframe : 101 \* 11**

**Renaming data frame columns to short names.**

****

**Part 1**

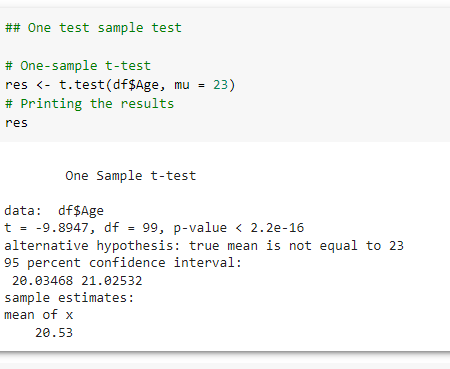
**One Sample T test**

We have a student age column for all 4 years of students. Ideally their age will be in range

18-24 let’s assume their mean is 23 and do one sample test.

Null Hypothesis : mean is equal to 23

Alternative Hypothesis: Mean is not equal to 23



**Results**

We can clearly see that the mean is not equal to 23 hence our alternative hypothesis is true following is our interpretations for other factors used in hypothesis testing.

# mean is equal to 23

# t is the t-test statistic value (t = -57.966),

# df is the degrees of freedom (df= 99),

# p-value is the significance level of the t-test (p-value = 2.2e-16).

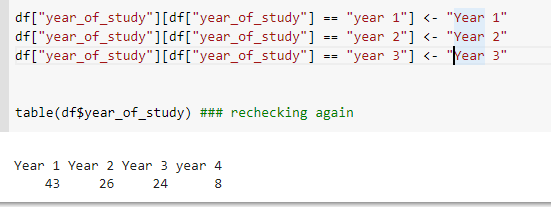
# conf.int is the confidence interval of the mean at 95% (conf.int = [20.03468, 21.02532]);

# sample estimate is the mean value of the sample (mean = 20.53).

**Part 2**

**Hypothesis testing for p-value using two sample test**

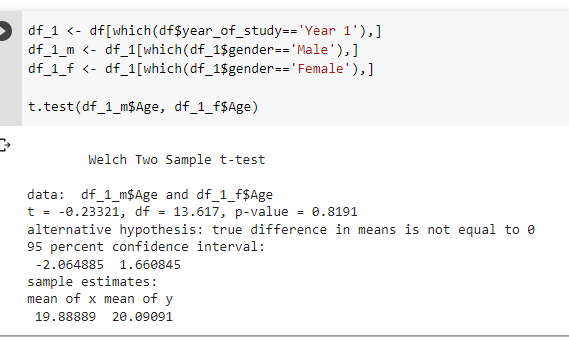
we can see that our values are not aligned in year of study columns. First, let's give a common name for each unique year.



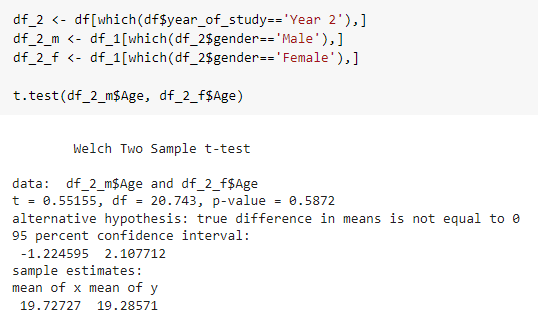
**Problem statement**

**Let's assume that the age of males and females in the same year of study is the same. We will conduct the T-test for year 1 and year 2 assuming that mean age for male and female students is the same for each year.**

**Year 1**

****

**Year 2**

****

For both year students we are assuming that the mean of male and female is the same.

But the same is not true as in year 1 we can see that the mean of male and mean of female is 19.88 and 20.09 which is definitely closer but not equal. We are getting a p-value as 0.8191 which is greater than 0.05.

On the similar lines mean for year 2 male and female students is not the same it's more closer as compared to year 1. For year 2 we got a p-value as 0.5872 which is more than 0.05 but less than 0.8191 the main reason is the mean age for year 2 students is more close as compared to year 1 students.

Overall we must be wondering that means are almost small but still we are high P-value. The main reason is that the overall age of almost all the students is the same hence we have very small width for assumptions.

Hence even a smaller difference in mean ages is giving high value in P-value. We can see the same iin boxplots below.

**YEAR 1 YEAR 2**

