**Hypothesis Testing Exercise**

1. A F&B manager wants to determine whether there is any significant difference in the diameter of the cutlet between two units. A randomly selected sample of cutlets was collected from both units and measured? Analyze the data and draw inferences at 5% significance level. Please state the assumptions and tests that you carried out to check validity of the assumptions.

Minitab File : **Cutlets.mtw**

Ans :- To check whether there is "significant difference" in diameter, between unit's of cutlets, 2 test need to be performed:

1**. "Normality test".**

Tests to see if the data/diameter of the cutlet is normally distributed.

"Assumptions"

Observations in each sample diameter is not dependent and identically distributed.

"Interpretation"

o If p >0.05 or 5%(given): the data set is normally distributed.

o p <0.05: the data set is not Normally distributed.

**2. Variance test.**

Tests whether the means of 2 or more are not dependent samples or data are significantly different.

Assumptions

o Observations in each/every sample diameter are not dependent and identically distributed.

o Observations in each/every sample diameter are normally distributed.

o Observations in each /every sample have the same variance.

Interpretation

· If p >0.05: the Mean of the sample is equal.

· p < 0.05: the Mean of sample is not Equal.

Sample Output:

DiameterOfCutlet1\_st=0.232, DiameterOfCutlet1\_p=0.890

Data set of DiameterOfCutlet1 is normally distributed.

DiameterOfCutlet2\_st2=1.007, DiameterOfCutlet2\_p2=0.605

Data set of DiameterOfCutlet2 is normally distributed.

stat=0.019, p=0.891

Variance is same between the sample.

Program code:

#import module

from scipy.stats import \*

#Sample of daimeter of 2 cutlets 1 & 2.

DiameterOfCutlet1=[6.8090,6.4376,6.9157,7.3012,7.4488,7.3871,6.8755,7.0621,6.6840,6.8236]

DiameterOfCutlet2=[6.7703,7.5093,6.7300,6.7878,7.1522,6.8110,7.2212,6.6606,7.2402,7.0503]

alphaa=0.05 #given 5%

#Step 1-------> Normality Test of dataset.

DiameterOfCutlet1\_st,DiameterOfCutlet1\_p=stats.normaltest(DiameterOfCutlet1)

print('\nDiameterOfCutlet1\_st=%.3f, DiameterOfCutlet1\_p=%.3f' %(DiameterOfCutlet1\_st,DiameterOfCutlet1\_p))

if DiameterOfCutlet1\_p> alphaa:

print("\nData set of DiameterOfCutlet1 is normally distributed.")

else:

print("Data set of DiameterOfCutlet1 is not normally distributed.")

DiameterOfCutlet2\_st2, DiameterOfCutlet2\_p2=stats.normaltest(DiameterOfCutlet2)

print('\nDiameterOfCutlet2\_st2=%.3f, DiameterOfCutlet2\_p2=%.3f' %(DiameterOfCutlet2\_st2,DiameterOfCutlet2\_p2))

if DiameterOfCutlet2\_p2> alphaa:

print("\nData set of DiameterOfCutlet2 is normally distributed.")

else:

print("Data set of DiameterOfCutlet2 is not normally distributed.")

#2) Variance Test

#Create Hypothesis for variances of Cutlet 1 & 2 diameter's"

# Example of the Analysis of Variance Test

stat, p =stats. f\_oneway(DiameterOfCutlet1,DiameterOfCutlet2)

print('\nstat=%.3f, p=%.3f' % (stat, p))

if p > alphaa:

print('\nVariance is same between the sample.\n')

else:

print('Variance is not same between the sample.')

Code description:

1. Import scipy module.

2. Collect the diameter of 2 cutlets in 2 separate lists.

3. Check whether the data is normally distributed or not.

4. If yes Carry forward in our test.

5. Check for variance test as well.

6. If both test pass means there is no significant size difference between cutlets.

2. A hospital wants to determine whether there is any difference in the average Turn Around Time (TAT) of reports of the laboratories on their preferred list. They collected a random sample and recorded TAT for reports of 4 laboratories. TAT is defined as sample collected to report dispatch.

Analyze the data and determine whether there is any difference in average TAT among the different laboratories at 5% significance level.

Minitab File: LabTAT.mtw

Ans:

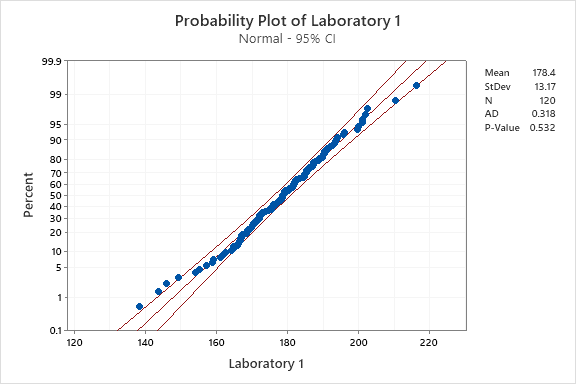
Solution: input is discrete in more than 2 categories so we will go with ANOVA test

**Normality test:**

**Hypothesis for lab 1**

Ho= data are normally distributed

Ha= data are not normally distributed

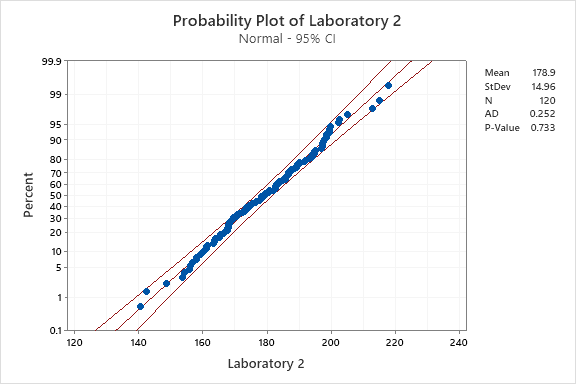


p-value is 0.532>0.05

p high Ho fly accept null hypothesis hence data are normally distributed

**Hypothesis for lab 2**

Ho= data are normally distributed

Ha= data are not normally distributed 

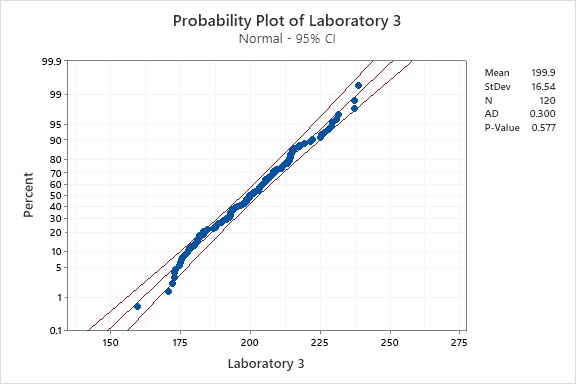
p-value is 0.733>0.05

phigh Ho fly accept null hypothesis hence data are normally distributed

**Hypothesis for lab 3**

Ho= data are normally distributed

Ha= data are not normally distributed



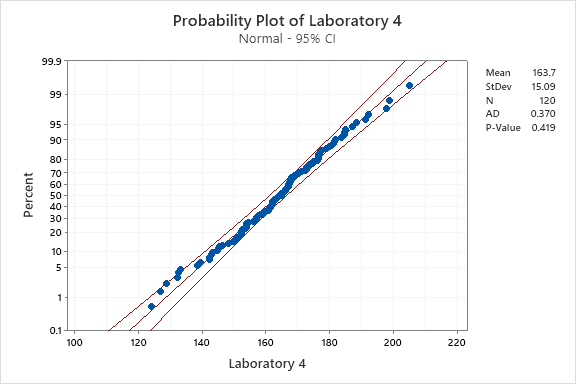
p-value is 0.577>0.05

p high Ho fly accept null hypothesis hence data are normally distributed

**hypothesis for lab 4**

Ho= data are normally distributed

Ha= data are not normally distributed



p-value is 0.419 >0.05

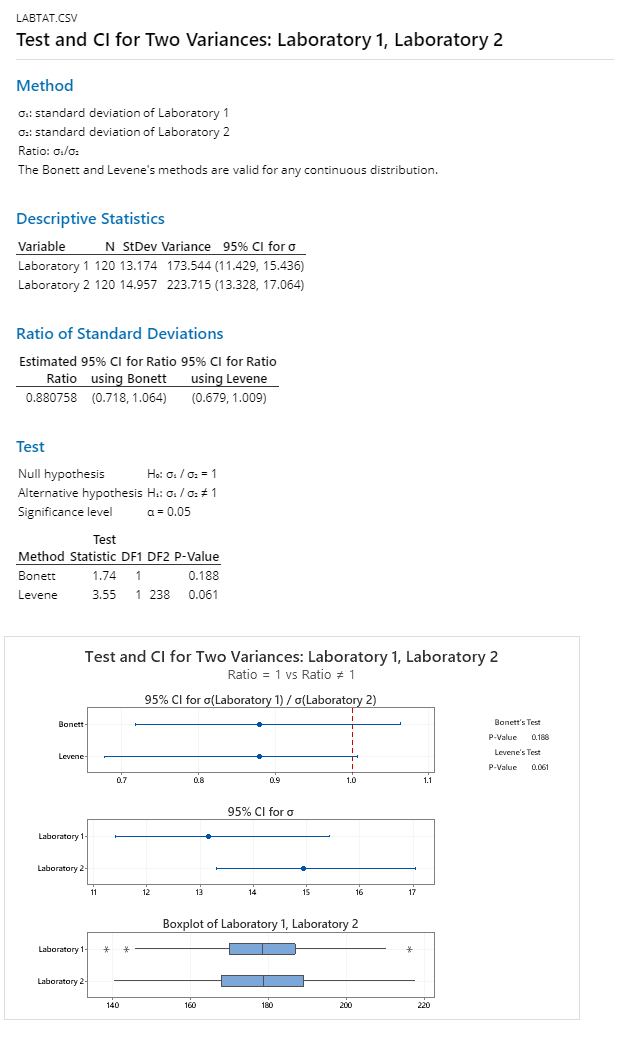
p high Ho fly accept null hypothesis hence data are normally distributed

**Variance test**

**Hypothesis for lab 1 and lab 2**

Ho=variance of TAT of lab 1 is equal to variance of TAT of lab 2

Ha=variance of TAT of lab 1 is not equal to variance of TAT of lab 2



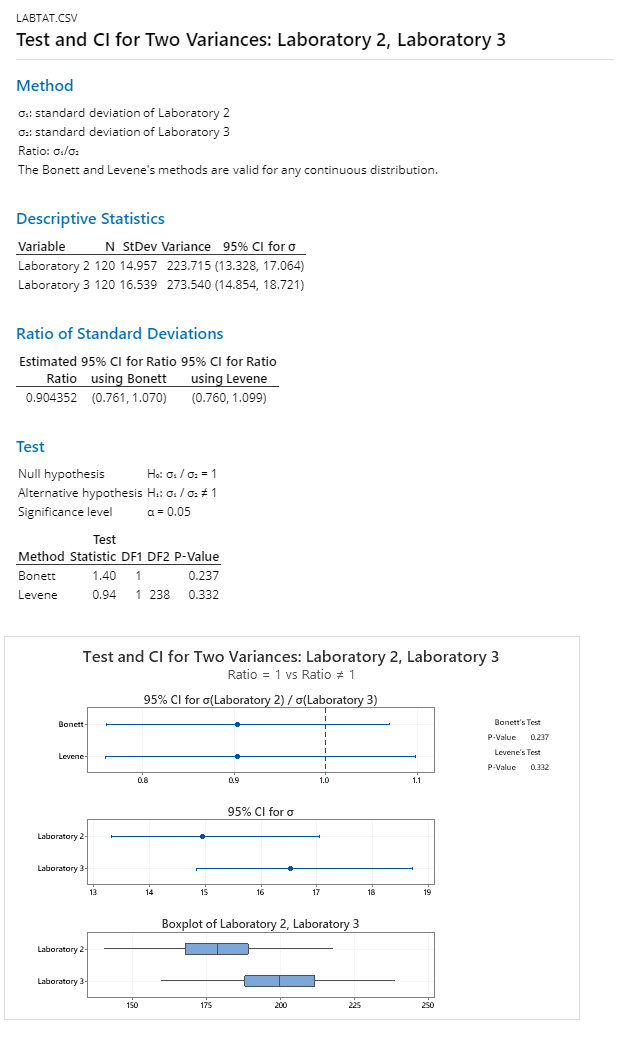
p-value is 0.061 >0.05

p high Ho fly accept null hypothesis hence variance of TAT of lab 1 is equal to variance of TAT of lab 2

**Hypothesis for lab 2 and lab 3**

Ho=variance of TAT of lab 2 is equal to variance of TAT of lab 3

Ha=variance of TAT of lab 2 is not equal to variance of TAT of lab 3



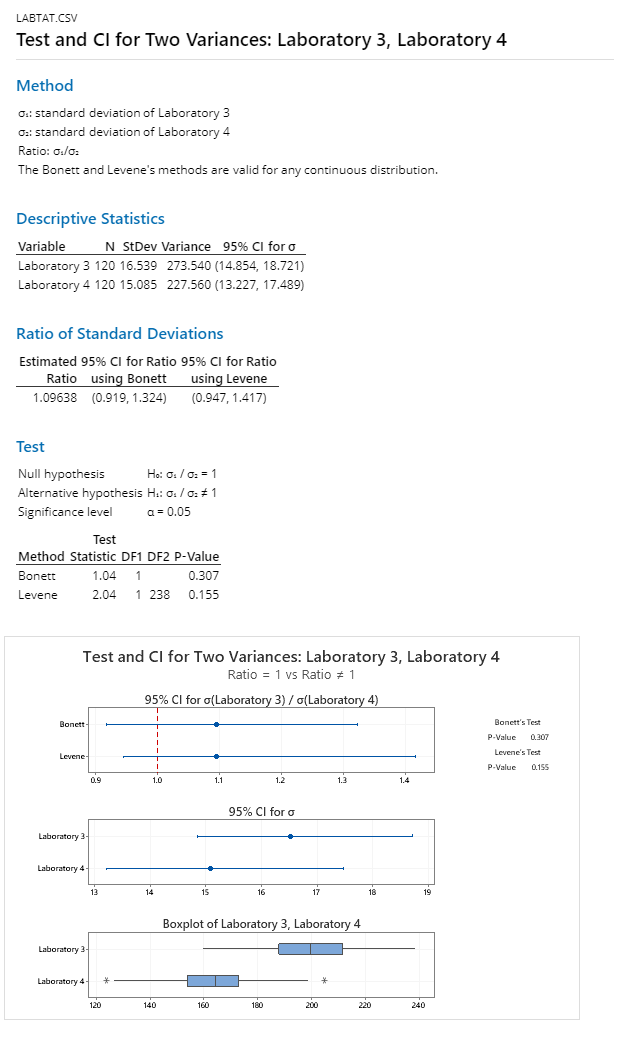
p-value is 0.332>0.05

p high Ho fly accept null hypothesis hence variance of TAT of lab 2 is equal to variance of TAT of lab 3

**Hypothesis for lab 3 and lab 4**

Ho=variance of TAT of lab 3 is equal to variance of TAT of lab 4

Ha=variance of TAT of lab 3 is not equal to variance of TAT of lab 4



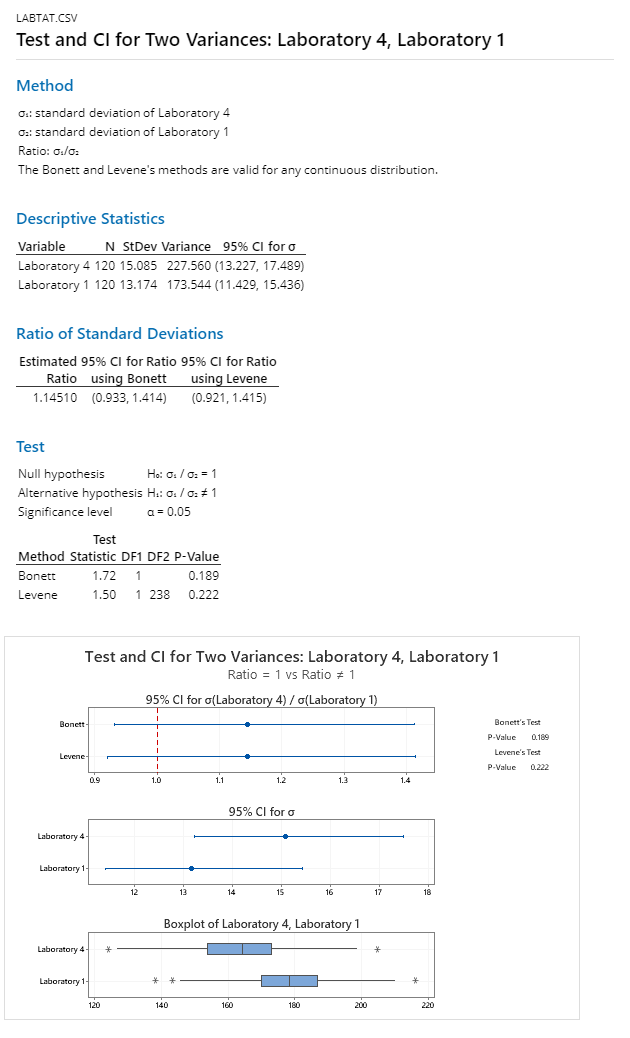
p-value is 0.155>0.05

p high Ho fly accept null hypothesis hence variance of TAT of lab 3 is equal to variance of TAT of lab 4

**Hypothesis for lab 4 and lab 1**

Ho=variance of TAT of lab 4 is equal to variance of TAT of lab 1

Ha=variance of TAT of lab 4 is not equal to variance of TAT of lab 1



p-value is 0.222>0.05

p high Ho fly accept null hypothesis hence variance of TAT of lab 4 is equal to variance of TAT of lab 1

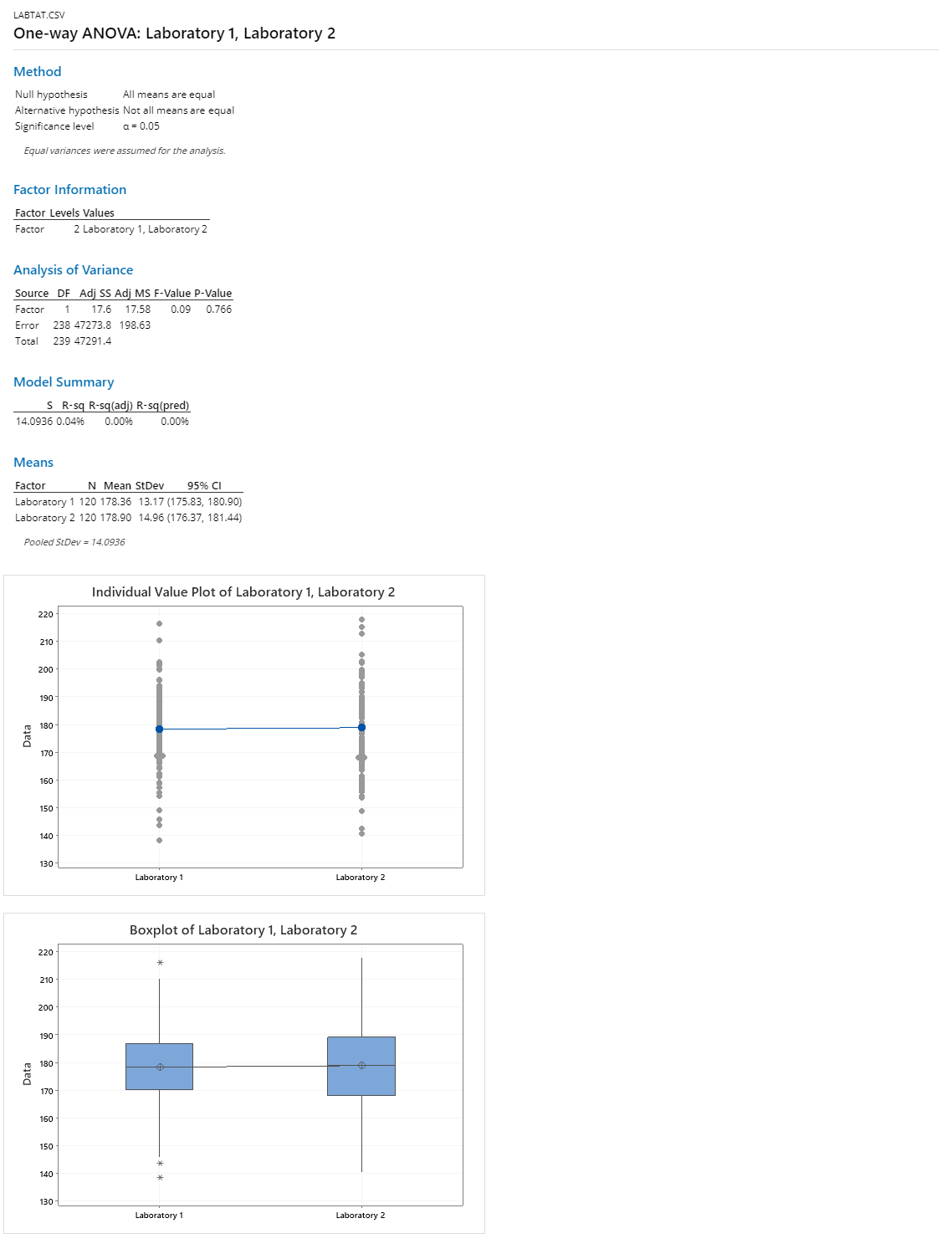
**ANOVA one way test**

Ho=average TAT of all sample is same

Ha=average of TAT of all sample is not same

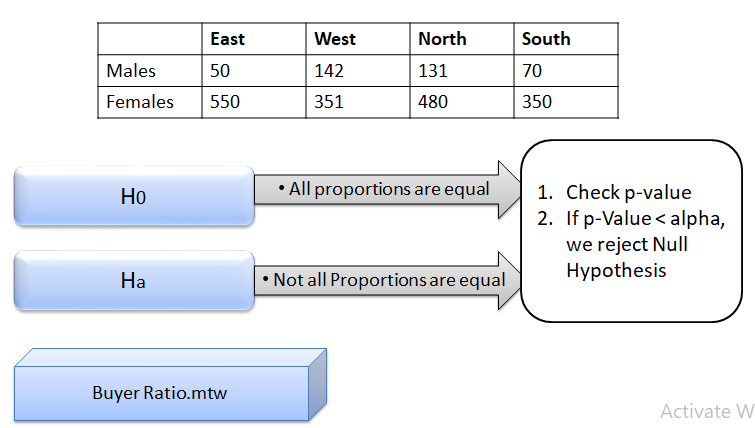
p-0.00<0.05

p low Ho go accept alternate hypothesis and reject null hypothesis

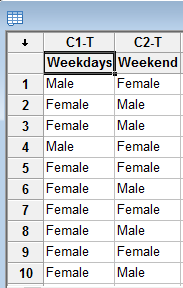


**inferences :** average of TAT of all samples is not same so there is difference in average turn around time of labs.

3. Sales of products in four different regions is tabulated for males and females. Find if male-female buyer rations are similar across regions.



Below is the sample data for Male and Female on weekday and weekends



***Inputs***are **2 discrete variables**.

***Output***is **Discrete**as we are trying to find out if proportions of male and female walking in to the store is same or not

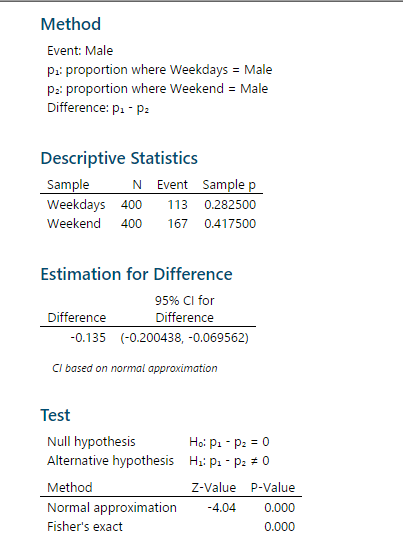
We proceed with **2-proportion test**

Create hypothesis

Ho= Proportions of Male and Female are same

Ha= Proportions of Male and Female are not same

Stat — >Basic Statistics— >2-proportion test



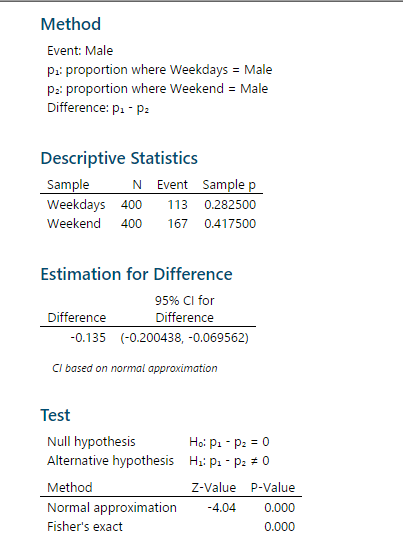
P-value is less than 0.05 and hence we fail to reject Null.Hence proportions of Male and Female are not same

Now we will try to find out whose proportion is higher. We create another hypothesis

Ho= Proportions of Male is less than or equal to Female

Ha= Proportions of Male is greater than Female

Stat->Basic Statistics->2 proportion test.



P-value <0.05 and hence we reject null.**Hence proportion of Male is greater than Female.**

**4. TeleCall uses 4 centers around the globe to process customer order forms. They audit a certain % of the customer order forms. Any error in order form renders it defective and has to be reworked before processing. The manager wants to check whether the defective % varies by centre. Please analyze the data at 5% significance level and help the manager draw appropriate inferences**

Minitab File: CustomerOrderForm.mtw

 Solutions:

Input are 4 discrete variables(phillippines, Indonesia, Malta, India)

And output is also discrete we are trying to find out whether the defective percentage varies by centre.

Ho= % of defective of all countries are equal i.e % Phillipness=%Indonesia=%Malta=%India

Ha =at lest one defective% is not equal

Here we perform chi-square test

p-value is 0.277>0.05

p high Ho fly hence accept null hypothesis i.e percentage defective in all countries are equal

**inferences**: as defective percentage is same in all countries there is no need of any action .

Fantaloons Sales managers commented that % of males versus females walking in to the store differ based on day of the week. Analyze the data and determine whether there is evidence at 5 % significance level to support this hypothesis.

Minitab File: Fantaloons.mtw