In [28]:
```
import pandas as pd
import numpy as np
from scipy.stats import t
```

In [26]:
```
data= pd.read_csv("Crop_recommendation.csv")
data.head()
```

Out[26]:

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

2sample t-test for rice and jute

For humidity

H0: There is no difference between rice and jute, mean humidity of rice - mean humidity of jute =0

H1: There is no difference between rice and jute, mean humidity of rice - mean humidity of jute !=0

In [65]:
```
alpha = 0.05
df_rice = data[data["label"]=="rice"]
df_jute = data[data["label"]=="jute"]
```

In [66]:
```
mean_humidity_rice = df_rice["humidity"].mean()
sd_humidity_rice = df_rice["humidity"].std()
mean_humidity_jute = df_jute["humidity"].mean()
sd_humidity_jute = df_jute["humidity"].std()
```

```
print(mean_humidity_rice,  sd_humidity_rice)
print(mean_humidity_jute, sd_humidity_jute)
```

```
82.27282153889999 1.4183811592213085
79.6398642063 5.50779761705286
```

In [67]:
```
n1 = df_rice.shape[0]
n2 = df_jute.shape[0]
print(n1 , n2)
```

```
100 100
```

In [68]:
```
std_error =  np.array((sd_humidity_rice**2 / n1) +  (sd_humidity_jute**2/n2))
std_error = np.sqrt(std_error)
std_error
```

Out[68]: 0.5687498545340223

In [69]:
```
dof = ((sd_humidity_rice**2/n1) + (sd_humidity_jute**2/n2))**2/ (((sd_humidity_rice**2/n1)**2/(n1-1)) + (sd_humidity_jute**2/n
dof
```

Out[69]: 112.07342230894034

In [70]:
```
test_statistic = (mean_humidity_rice - mean_humidity_jute)/std_error
test_statistic
```

Out[70]: 4.629376713875685

In [71]:
```
#get the critical value for the specified degrees of freedom and conf level
critical_value_right = t.ppf(1-alpha/2,dof)
critical_value_left = t.ppf(alpha/2,dof) #
p_value = 1-t.cdf(test_statistic,dof)
print("Critical value left : "+ str(critical_value_left) , "Critical value right:" + str(critical_value_right))
print("P value: "+ str(p_value))
if(test_statistic > critical_value_right or test_statistic < critical_value_left):
    print("We reject the null hypotheses because there is significant evidence against it. Hence there is a significant differ
else:
    print("There is not enough to reject the null hypotheses. Hence there is no significant difference in the humidity require
```

```
Critical value left : -1.9813576402765938 Critical value right:1.9813576402765931
P value: 4.965259592371041e-06
We reject the null hypotheses because there is significant evidence against it. Hence there is a significant difference in the
humidity requirements for rice and jute
```

For temperature

H0: The temperature requirements are similar. mean temp of rice - mean temp of jute = 0

H1: The temperature requirements are different. mean temp of rice - mean temp of jute !=0

In [43]:
```
mean_temp_rice = df_rice["temperature"].mean()
sd_temp_rice = df_rice["temperature"].std()
mean_temp_jute = df_jute["temperature"].mean()
sd_temp_jute = df_jute["temperature"].std()

print(mean_temp_rice,  sd_temp_rice)
print(mean_temp_jute, sd_temp_jute)
```

```
23.6893322105 2.0312719543716664
24.958375826499996 1.185138311343944
```

In [44]:
```
n1 = df_rice.shape[0]
n2 = df_jute.shape[0]
print(n1 , n2)
```

```
100 100
```

In [50]:
```
alpha = 0.05
std_error = np.sqrt((sd_temp_rice**2/n1) + (sd_temp_jute**2/n2))
std_error
```

Out[50]:
```
0.23517267208653442
```

In [51]:
```
dof = ((sd_temp_rice**2/n1) + (sd_temp_jute**2/n2))**2/ (((sd_temp_rice**2/n1)**2/(n1-1)) + ((sd_temp_jute**2/n2)**2/(n2-1)))
dof
```

Out[51]:
```
159.4018314883493
```

In [52]:
```
test_statistic = (mean_temp_rice - mean_temp_jute) / std_error
test_statistic
```

Out[52]: -5.3962205928971

In [75]:
```
critical_value_right = t.ppf(1-alpha/2,dof)
critical_value_left = t.ppf(alpha/2,dof)
print("Critical value left:", critical_value_right, "||  Critical value right:",critical_value_left)
p_value = 1-t.cdf(test_statistic, dof)
print("P value:" +str(p_value))

if(test_statistic > critical_value_right or test_statistic < critical_value_left):
    print("Reject null Hypotheses. There is significant difference between the temperature requirements for rice and jute")
else:
    print("fail to reject null hypotheses as there is isnt enough evidence to support the fact that the temperature requiremen
```

```
Critical value left: 1.9813576402765931 ||  Critical value right: -1.9813576402765938
P value:4.965259592371041e-06
Reject null Hypotheses. There is significant difference between the temperature requirements for rice and jute
```

2 sample t-test for banana and grapes

1. Humidity

H0: There is no difference between the humidity requirements for banana and grapes. mean humidity of banana = mean humidity of grapes

H1: There is a difference between them. mean humidity of banana - mean humidity of grapes !=0

In [78]:
```
alpha = 0.05
df_banana = data[data["label"]=="banana"]
df_grape = data[data["label"]=="grapes"]
```

In [79]:
```
mean_humidity_banana = df_banana["humidity"].mean()
sd_humidity_banana = df_banana["humidity"].std()
```

```
mean_humidity_grape = df_grape["humidity"].mean()
sd_humidity_grape = df_grape["humidity"].std()

print(mean_humidity_banana,  sd_humidity_banana)
print(mean_humidity_grape, sd_humidity_grape)
```

```
80.3581225811 2.8054808335375014
81.87522752119999 1.177111249739479
```

In [80]:
```
n1 = df_banana.shape[0]
n2 = df_grape.shape[0]
print(n1 , n2)
```

```
100 100
```

In [81]:
```
std_error =  np.sqrt((sd_humidity_banana**2 / n1) +  (sd_humidity_grape**2/n2))
std_error
```

Out[81]:
```
0.3042419037806842
```

In [82]:
```
dof = ((sd_humidity_banana**2/n1) + (sd_humidity_grape**2/n2))**2/ (((sd_humidity_banana**2/n1)**2/(n1-1)) + (sd_humidity_grap
dof
```

Out[82]:
```
132.8088631514806
```

In [83]:
```
test_statistic = (mean_humidity_banana - mean_humidity_grape)/std_error
test_statistic
```

Out[83]:
```
-4.986508831451448
```

In [88]:
```
#get the critical value for the specified degrees of freedom and conf level
critical_value_right = t.ppf(1-alpha/2,dof)
critical_value_left = t.ppf(alpha/2,dof)
if( test_statistic >0 ):
    p_value = 1-t.cdf(test_statistic,dof)
else:
    p_value = t.cdf(test_statistic,dof)

print("Critical value left : "+ str(critical_value_left) , "Critical value right:" + str(critical_value_right))
print("P value: "+ str(p_value))
```

```
if(test_statistic > critical_value_right or test_statistic < critical_value_left):
    print("We reject the null hypotheses because there is significant evidence against it. Hence there is a significant differ
else:
    print("There is not enough to reject the null hypotheses. Hence there is no significant difference in the humidity require
```

```
Critical value left : -1.9779873987352643 Critical value right:1.9779873987352639
P value: 9.428234358219145e-07
We reject the null hypotheses because there is significant evidence against it. Hence there is a significant difference in the
humidity requirements for banana and grape
```

2. Temperature

H0: There is no difference between the temperature requirements for banana and grapes. mean temperature of banana = mean temperature of grapes

H1: There is a difference between them. mean temperature of banana - mean temperature of grapes !=0

In [89]:
```
mean_temp_banana = df_banana["temperature"].mean()
sd_temp_banana = df_banana["temperature"].std()
mean_temp_grape = df_grape["temperature"].mean()
sd_temp_grape = df_grape["temperature"].std()

print(mean_temp_banana,  sd_temp_banana)
print(mean_temp_grape, sd_temp_grape)
```

```
27.3767983057 1.428359021184807
23.849575120049998 9.73864872199479
```

In [ ]:
```
n1 = df_banana.shape[0]
n2 = df_grape.shape[0]
print(n1 , n2)
```

In [90]:
```
std_error =  np.sqrt((sd_temp_banana**2 / n1) +  (sd_temp_grape**2/n2))
std_error
```

Out[90]: 0.9842839449255016

In [91]:
```
dof = ((sd_temp_banana**2/n1) + (sd_temp_grape**2/n2))**2/ (((sd_temp_banana**2/n1)**2/(n1-1)) + (sd_temp_grape**2/n2)**2/(n2-
dof
```

Out[91]: 103.25737225359428

In [92]:
```
test_statistic = (mean_temp_banana - mean_temp_grape)/std_error
test_statistic
```

Out[92]: 3.583542334338258

In [93]:
```
#get the critical value for the specified degrees of freedom and conf level
critical_value_right = t.ppf(1-alpha/2,dof)
critical_value_left = t.ppf(alpha/2,dof)
if( test_statistic >0 ):
    p_value = 1-t.cdf(test_statistic,dof)
else:
    p_value = t.cdf(test_statistic,dof)

print("Critical value left : "+ str(critical_value_left) , "Critical value right:" + str(critical_value_right))
print("P value: "+ str(p_value))

if(test_statistic > critical_value_right or test_statistic < critical_value_left):
    print("We reject the null hypotheses because there is significant evidence against it. Hence there is a significant differ
else:
    print("There is not enough to reject the null hypotheses. Hence there is no significant difference in the temperature requ
```

```
Critical value left : -1.9832053951259196 Critical value right:1.9832053951259192
P value: 0.00025952851450905534
We reject the null hypotheses because there is significant evidence against it. Hence there is a significant difference in the
temperature requirements for banana and grape
```

1 Sample t-test

__H0: ph of mango = 7.5__

__H1:ph of manog != 7.5__

In [95]:
```
alpha =0.5
pop_mean = 7.5
df_mango = data[data["label"] == "mango"]
mean_mango = df_mango["ph"].mean()
```

```
sd_mango = df_mango["ph"].std()
std_error = sd_mango/np.sqrt((df_mango.shape)[0])
dof = (df_mango.shape)[0]-1
test_statistic = (mean_mango - pop_mean)/ std_error
test_statistic
```

Out[95]: -24.637475315453777

In [98]:
```
crit_val_r = t.ppf(1-alpha/2, dof)
crit_val_l = t.ppf(alpha/2, dof)
print("Critical value left:" + str(crit_val_l), "Critical value right:"+  str(crit_val_r))

if(test_statistic <0):
    p_value = t.cdf(test_statistic,dof)
else:
    p_value = 1-t.cdf(test_statistic, dof)

print("P value is :"+ str(p_value))


if(test_statistic < crit_val_l or test_statistic > crit_val_r):
    print("Reject Null hypotheses. There is significant difference between the ph of Mango and the population")
else:
    print("Fail to reject the null hypotheses as there is not enough evidence for it. The ph of Mango is same as the populatio
```

```
Critical value left:-0.6768731561845536 Critical value right:0.6768731561845536
P value is :2.440047669327577e-45
Reject Null hypotheses. There is significant difference between the ph of Mango and the population
```

In [ ]: