In [18]:
```
import numpy as np
import pandas as pd
import scipy.stats as stats
```

Hypothesis Testing Exercise

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.

In [19]:
```
df=pd.read_csv('Cutlets.csv')
```

In [20]:
```
df
```

Out[20]:

	Unit A	Unit B
0	6.8090	6.7703
1	6.4376	7.5093
2	6.9157	6.7300
3	7.3012	6.7878
4	7.4488	7.1522
5	7.3871	6.8110
6	6.8755	7.2212
7	7.0621	6.6606
8	6.6840	7.2402
9	6.8236	7.0503
10	7.3930	6.8810
11	7.5169	7.4059
12	6.9246	6.7652
13	6.9256	6.0380
14	6.5797	7.1581
15	6.8394	7.0240
16	6.5970	6.6672
17	7.2705	7.4314
18	7.2828	7.3070
19	7.3495	6.7478
20	6.9438	6.8889
21	7.1560	7.4220
22	6.5341	6.5217
23	7.2854	7.1688
24	6.9952	6.7594
25	6.8568	6.9399
26	7.2163	7.0133
27	6.6801	6.9182
28	6.9431	6.3346
29	7.0852	7.5459
30	6.7794	7.0992
31	7.2783	7.1180
32	7.1561	6.6965

	Unit A	Unit B
33	7.3943	6.5780
34	6.9405	7.3875

H0 == Mean for Y1 and Y2 are equal (There is no significance difference between diameter of the Culets)

H1 == Mean for Y1 and Y2 are not equal (There is a significance difference between diameter of the Culets)

In [21]:
```
unit_A=df['Unit A']
```

In [22]:
```
unit_B=df['Unit B']
```

In [23]:
```
stats.ttest_rel(unit_A,unit_B)
```

Out[23]:
```
TtestResult(statistic=0.7536787225614316, pvalue=0.4562300768038412, df=34)
```

TtestResult(statistic=0.7536787225614316, pvalue=0.4562300768038412, df=34)

pvalue > α hence we accept null hypothesis

therefore there is no significance difference between diameter of the cultes

---

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.

In [24]:
```
df=pd.read_csv('LabTAT.csv')
```

In [25]:
```
df
```

Out[25]:

	Laboratory 1	Laboratory 2	Laboratory 3	Laboratory 4
0	185.35	165.53	176.70	166.13
1	170.49	185.91	198.45	160.79
2	192.77	194.92	201.23	185.18
3	177.33	183.00	199.61	176.42
4	193.41	169.57	204.63	152.60
...	...	...	...	...
115	178.49	170.66	193.80	172.68
116	176.08	183.98	215.25	177.64
117	202.48	174.54	203.99	170.27
118	182.40	197.18	194.52	150.87
119	182.09	215.17	221.49	162.21

120 rows × 4 columns

H0 == Mean for Y1,Y2,Y3 & Y4 are equal (there is no difference in average TAT among the different laboratories)

H1 == Mean for Y1,Y2,Y3 & Y4 are not equal (there is a difference in average TAT among the different laboratories)

In [26]:
```
stats.f_oneway(df['Laboratory 1'],df['Laboratory 2'],df['Laboratory 3'],df['Labo
```

Out[26]:
```
F_onewayResult(statistic=118.70421654401437, pvalue=2.1156708949992414e-57)
```

F_onewayResult(statistic=118.70421654401437, pvalue=2.1156708949992414e-57)

pvlaue < α hence we reject null hypothesis

therefore there is a difference in average TAT among the different laboratories

---

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

In [27]:
```
df=pd.read_csv('BuyerRatio.csv')
```

In [28]:
```
df
```

Out[28]:

	Observed Values	East	West	North	South
0	Males	50	142	131	70
1	Females	435	1523	1356	750

H0 = male-female buyer rations are similar across regions

H1 = male-female buyer rations are not similar across regions

Assume α=5%

In [29]:
```
stats.chi2_contingency([df['East'],df['West'],df['North'],df['South']])
```

Out[29]:
```
Chi2ContingencyResult(statistic=1.5959455386610577, pvalue=0.6603094907091882,
dof=3, expected_freq=array([[  42.76531299,  442.23468701],
       [ 146.81287862, 1518.18712138],
       [ 131.11756787, 1355.88243213],
       [  72.30424052,  747.69575948]]))
```

Chi2ContingencyResult(statistic=1.5959455386610577, pvalue=0.6603094907091882, dof=3)

pvalue > α hence null hypothesis is accepted

Therfore male-female buyer rations are similar across regions

---

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

In [30]:
```
df=pd.read_csv('Costomer+OrderForm.csv')
```

In [31]:
```
df
```

Out[31]:

	Phillippines	Indonesia	Malta	India
0	Error Free	Error Free	Defective	Error Free
1	Error Free	Error Free	Error Free	Defective
2	Error Free	Defective	Defective	Error Free
3	Error Free	Error Free	Error Free	Error Free
4	Error Free	Error Free	Defective	Error Free
...	...	...	...	...
295	Error Free	Error Free	Error Free	Error Free
296	Error Free	Error Free	Error Free	Error Free
297	Error Free	Error Free	Defective	Error Free
298	Error Free	Error Free	Error Free	Error Free
299	Error Free	Defective	Defective	Error Free

300 rows × 4 columns

H0 = Defective % does not varies by centre

H1 = Defective % varies by centre

In [32]:
```
stats.chi2_contingency([df.Phillippines.value_counts(),df.Indonesia.value_counts
```

Out[32]:
```
Chi2ContingencyResult(statistic=3.8589606858203545, pvalue=0.2771020991233144,
dof=3, expected_freq=array([[271.75,  28.25],
       [271.75,  28.25],
       [271.75,  28.25],
       [271.75,  28.25]]))
```

statistic=3.8589606858203545, pvalue=0.2771020991233144

pvalue > α hence null hypothesis is accepted

therfore Defective % does not varies by centre

In [ ]:
```

```