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

Hypothesis Formulation

1. Null Hypothesis (H0): Diameter of Cultet between 2 units is same that is there is no any sigificant difference between Diameter of Cultet between 2 units.
2. Alternative Hypothesis (H1): Diameter of Cultet between 2 units is different that is there is sigificant difference between Diameter of Cultet between 2 units.

1. Import Necessary Libraries

```
import pandas as pd
import numpy as np
from scipy.stats import f_oneway
from scipy import stats
```

2. Import Data

```
cutlets_data = pd.read_csv('Cutlets.csv')
cutlets_data.head()
```

	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

3. Data Understanding

```
cutlets_data.shape
```

```
(35, 2)
```

```
cutlets_data.describe()
```

	Unit A	Unit B
count	35.000000	35.000000
mean	7.019091	6.964297
std	0.288408	0.343401
min	6.437600	6.038000
25%	6.831500	6.753600
50%	6.943800	6.939900
75%	7.280550	7.195000
max	7.516900	7.545900

4. Perform Anova Test

```
stats, p_value = f_oneway(cutlets_data['Unit A'],cutlets_data['Unit B'])
p_value
```

```
0.47223947245995734
```

5. Conclusion

```
# At the 5 % significance level
if p_value < 0.05:
  print('Reject the Null Hypothesis because there is significant difference between diameter of cutlets of 2 units. It means Diameter
else:
  print('Do not Reject the Null Hypothesis because there is no any significant difference between diameter of cutlets of 2 units. It
```

```
Do not Reject the Null Hypothesis because there is no any significant difference between diameter of cutlets of 2 units. It mea
```

==========

=

Example 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 & determine whether there is any difference in average TAT among the different laboratories at 5% significance level.

Hypothesis Formulation

1. Null Hypothesis (H0): There is no any significant difference between average TAT among the different Laboratories.
2. Alternative Hypothesis (H1): There is significant difference between average TAT among the different Laboratories.

1. Import Data

```
tat_data = pd.read_csv('LabTAT.csv')
tat_data
```

	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

```
tat_data.describe()
```

	Laboratory 1	Laboratory 2	Laboratory 3	Laboratory 4
count	120.000000	120.000000	120.000000	120.00000
mean	178.361583	178.902917	199.913250	163.68275
std	13.173594	14.957114	16.539033	15.08508
min	138.300000	140.550000	159.690000	124.06000
25%	170.335000	168.025000	188.232500	154.05000
50%	178.530000	178.870000	199.805000	164.42500
75%	186.535000	189.112500	211.332500	172.88250
max	216.390000	217.860000	238.700000	205.18000

Perform ANOVA Test

```
stats, p_value_tat = f_oneway(tat_data['Laboratory 1'],tat_data['Laboratory 2'],tat_data['Laboratory 3'],tat_data['Laboratory 4'])
p_value_tat
```

```
2.1156708949992414e-57
```

Conclusion

```
if p_value_tat < 0.05:
  print('Reject the Null Hypothesis because There is significant difference between average TAT among the different Laboratories.')
else:
  print('Do not Reject the Null Hypothesis because There is no any significant difference between average TAT among the different Lab
```

```
Reject the Null Hypothesis because There is significant difference between average TAT among the different Laboratories.
```

==========

=

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

H0: All proportions are equal.

H1: Not all Proportions are equal

Import Data

```
sales_data = pd.read_csv('BuyerRatio.csv')
sales_data
```

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

```
data = [[50,142,131,70],[435,1523,1356,750]]
```

Perform Chi-Squared Test

```
chi_score, p_value_sales, dof, expected_table = stats.chi2_contingency(observed = data)
print('Chi_Score         :',chi_score.round(5))
print('p_value_sales     :',p_value_sales.round(5))
print('Degree of Freedom :',dof)
print('Expected Table    :\n',expected_table)
```

```
Chi_Score         : 1.59595
p_value_sales     : 0.66031
Degree of Freedom : 3
Expected Table    :
 [[  42.76531299  146.81287862  131.11756787   72.30424052]
 [ 442.23468701 1518.18712138 1355.88243213  747.69575948]]
```

Conclusion

```
# Level of Significance is 5% so alpha =0.05
if p_value_sales < 0.05:
  print("Reject the Null Hypothesis because there is significant difference between buyer's ratio of male & female")
else:
  print("We do not Reject the Null Hypothesis because there is no significant difference between buyer's ratio of male & female")
```

```
We do not Reject the Null Hypothesis because there is no significant difference between buyer's ratio of male & female
```

Example 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.

Hypothesis Formulation

H0 (Null Hypothesis): Customer order forms defective % does not varies by centre

H1 (Alternate Hypothesis): Customer order forms defective % varies by centre

Import Data

```
customer_data = pd.read_csv('/content/Costomer+OrderForm.csv')
customer_data
```

	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

```
customer_data.Phillippines.value_counts()
```

```
Error Free    271
Defective      29
Name: Phillippines, dtype: int64
```

```
customer_data.Indonesia.value_counts()
```

```
Error Free    267
Defective      33
Name: Indonesia, dtype: int64
```

```
customer_data.Malta.value_counts()
```

```
Error Free    269
Defective      31
Name: Malta, dtype: int64
```

```
customer_data.India.value_counts()
```

```
Error Free    280
Defective      20
```

```
Name: India, dtype: int64
```

Perform Independent Chi-Squared Test

```
chi_score, p_value_sales, dof, expected_table = stats.chi2_contingency(observed = np.array([[271,267,269,280],[29,33,31,20]]))
print('Chi_Score         :',chi_score.round(5))
print('p_value_sales     :',p_value_sales.round(5))
print('Degree of Freedom :',dof)
print('Expected Table    :\n',expected_table)
```

```
Chi_Score         : 3.85896
p_value_sales     : 0.2771
Degree of Freedom : 3
Expected Table    :
 [[271.75 271.75 271.75 271.75]
 [ 28.25  28.25  28.25  28.25]]
```

Conclusion

```
# At Level of Significance is 5% so alpha =0.05
if p_value_sales < 0.05:
  print("Reject the Null Hypothesis because there is significant difference. Thus, customer order forms defective % varies by centre.
else:
  print("Do not reject the Null Hypthesis because there is no any significant difference. Thus, customer order forms defective % does
```

```
Do not reject the Null Hypthesis because there is no any significant difference. Thus, customer order forms defective % does no
```

Inference:

As (p_value = 0.2771) > (α = 0.05):

Do not reject the Null Hypthesis because there is no any significant difference. Thus, customer order forms defective % does not varies by centre.

THE END!!!!