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.

Minitab File : **Cutlets.mtw**

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

```
# Load the dataset
data = pd.read_csv("Cutlets.csv")
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

```
UnitA = pd.Series(data.iloc[:,0])
UnitA
```

```
0     6.8090
1     6.4376
2     6.9157
3     7.3012
4     7.4488
5     7.3871
6     6.8755
7     7.0621
8     6.6840
9     6.8236
10    7.3930
11    7.5169
12    6.9246
13    6.9256
14    6.5797
15    6.8394
16    6.5970
17    7.2705
18    7.2828
```

```
UnitB = pd.Series(data.iloc[:,1])
UnitB
```

```
0     6.7703
1     7.5093
2     6.7300
3     6.7878
4     7.1522
5     6.8110
6     7.2212
7     6.6606
8     7.2402
9     7.0503
10    6.8810
11    7.4059
12    6.7652
13    6.0380
14    7.1581
15    7.0240
16    6.6672
17    7.4314
18    7.3070
```

```
# 2 Sample - 2 Tail ttest :   stats.ttest_ind(array1,array2) >- ind = independent samples
p_value = stats.ttest_ind(UnitA,UnitB)
p_value
```

```
Ttest_indResult(statistic=0.7228688704678063, pvalue=0.4722394724599501)
```

```
p_value[1]  # 2 tail probability
```

```
0.4722394724599501
```

```
# compare p_value with α = 0.05 (At 5% significance level)
```

Hypothesis Testing Exercise

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**

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

```
data = pd.read_csv("LabTAT.csv")
data.head()
```

	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

```
# Annova ftest statistics : stats.f_oneway(Column-1,Column-2,Column-3,Column-4)
p_value=stats.f_oneway(data.iloc[:,0],data.iloc[:,1],data.iloc[:,2],data.iloc[:,3])
p_value
```

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

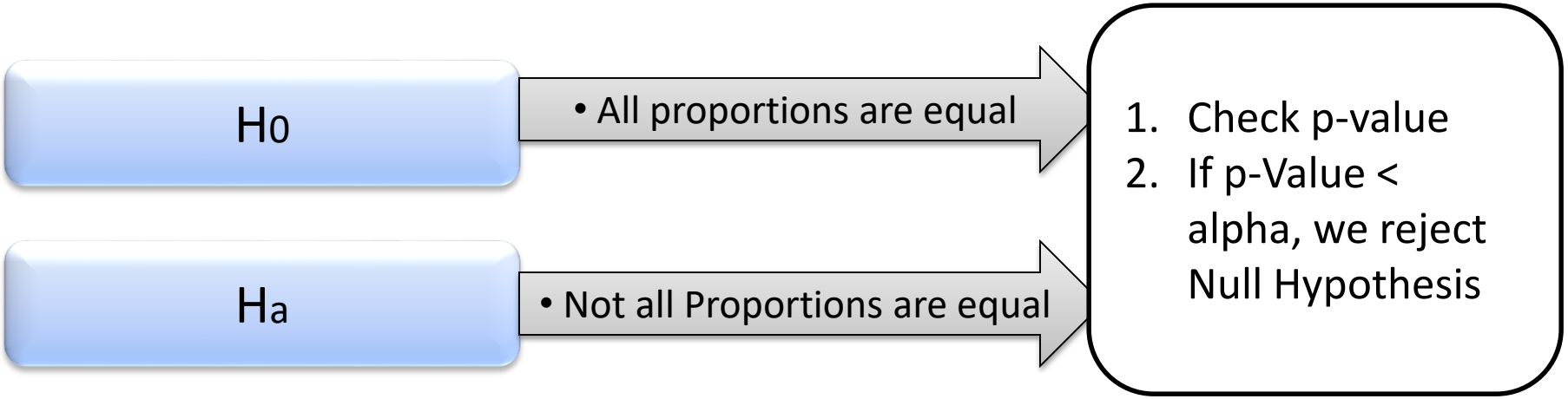
```
p_value[1]  # compare it with  a = 0.05
```

```
2.1156708949992414e-57
```

Hypothesis Testing Exercise

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

	East	West	North	South
Males	50	142	131	70
Females	550	351	480	350



Buyer Ratio.mtw

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

```
df = pd.read_csv("BuyerRatio.csv")
df
```

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

```
df_table = df.iloc[:,1:6]
df_table
```

	East	West	North	South
0	50	142	131	70
1	435	1523	1356	750

```
df_table.values
```

```
val = stats.chi2_contingency(df_table)
```

```
val
```

```
(1.595945538661058,
 0.6603094907091882,
 3,
 array([[  42.76531299,  146.81287862,  131.11756787,   72.30424052],
        [ 442.23468701, 1518.18712138, 1355.88243213,  747.69575948]]))
```

```
type(val)
```

```
tuple
```

```
no_of_rows = len(df_table.iloc[0:2,0])
no_of_columns = len(df_table.iloc[0,0:4])
degree_of_f  = (no_of_rows-1)*(no_of_columns-1)
print('Degree of Freedom',degree_of_f)
```

```
Degree of Freedom 3
```

```
Expected_value = val[3]
```

```
Expected_value
```

```
array([[  42.76531299,  146.81287862,  131.11756787,   72.30424052],
```

```
from scipy.stats import chi2
chi_square= sum([(o-e)**2/e for o,e in zip(df_table.values,Expected_value)])
chi_square_statestic= chi_square[0]+ chi_square[1]
chi_square_statestic
```

```
1.5152956451130446
```

```
critical_value= chi2.ppf(0.95,3)
critical_value
```

```
7.814727903251179
```

```
if chi_square_statestic >= critical_value:

    print('Dependent (reject H0)')
else:

    print('Independent (fail to reject H0)')
```

```
Independent (fail to reject H0)
```

```
pvalue = 1-chi2.cdf(chi_square_statestic,3)
pvalue
```

```
0.6787446296467897
```

```
if pvalue <= 0.05:
      print('Dependent (reject H0)')
else:
       print('Independent (fail to reject H0)')
```

```
Independent (fail to reject H0)
```

Hypothesis Testing Exercise

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**

```
import pandas as pd
import numpy as np
from scipy import stats
from scipy.stats import norm
from scipy.stats import chi2_contingency
```

```
# Load the dataset
data=pd.read_csv("Costomer+OrderForm.csv")
data.head()
```

	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

```
data.Phillippines.value_counts()
```

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

```
data.Indonesia.value_counts()
```

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

```
data.Malta.value_counts()
```

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

```
data.India.value_counts()
```

```
Error Free    280
Defective      20
Name: India, dtype: int64
```

```
# make contingency table
obs = np.array([[271,267,269,280],[29,33,31,20]])
obs
```

```
array([[271, 267, 269, 280],
       [ 29,  33,  31,  20]])
```

```
# Chi2 contengency independence test
chi2_contingency(obs) # o/p is (Chi2 stats value, p_value, df, expected obsvations)
```

```
(3.858960685820355,
 0.2771020991233135,
 3,
 array([[271.75, 271.75, 271.75, 271.75],
        [ 28.25,  28.25,  28.25,  28.25]]))
```

```
# Compare p_value with α = 0.05
```