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

```
###### Here we are using 2-sample 2-tail test

##### set the hypothesis
#### H0= no difference bet two units(unit A=unit B)
#### Ha= there is a difference

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

data= pd.read_csv('C:/Users/RIG1/Desktop/DS ASSIGNMENTS/QUESTIONS -all
assignments/Ass 3/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

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

```
0    6.8090
1    6.4376
2    6.9157
3    7.3012
4    7.4488
Name: Unit A, dtype: float64
```

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

```
0    6.7703
1    7.5093
2    6.7300
3    6.7878
4    7.1522
Name: Unit B, dtype: float64
```

```
stats.ttest_ind(unit_A,unit_B)
```

```
Ttest_indResult(statistic=0.7228688704678061,
pvalue=0.4722394724599501)
```

```
#### now At 5% significance level comparing p-value
#### o.4722>0.05
#### p-value > alpha


### here we will accept H0
```

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.

```
######### Here we are going to use anova test because here we are
comparing more than 2 variables


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

df=pd.read_csv('C:/Users/RIG1/Desktop/DS ASSIGNMENTS/QUESTIONS -all
assignments/Ass 3/LabTAT.csv')

df.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

```
stats.f_oneway(df.iloc[:,0],df.iloc[:,1],df.iloc[:,2],df.iloc[:,3])
```

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

```
####### NOW COMPARE THIS P-VALUE WITH ALPHA
#### 2.11 > 0.05
## hence, p-value > alpha then, we are accepting H0
```

```
# There is no difference in average TAT among the different
laboratories at 5% significance level.
```

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

```
### here we are using chi-square test

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

df1=pd.read_csv('C:/Users/RIG1/Desktop/DS ASSIGNMENTS/QUESTIONS -all
assignments/Ass 3/BuyerRatio.csv')

df1.head()

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

new_df=df1.iloc[:,1:6]
new_df

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

val=stats.chi2_contingency(new_df)
val

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

no_of_rows=len(new_df.iloc[0:2,0])

no_of_columns=len(new_df.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

######### Now expected value
Expected_value=val[3]
Expected_value
```

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

```
####### now chi-square value/p-value
#### X2=(O-E)2/+...........

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

```
1.5152956451130446
```

```
##### now critical value (using alpha value- 95% and df- 3)
critical_value=chi2.ppf(0.95,3)
critical_value
```

```
7.814727903251179
```

```
## compare p-value and critical value
## 7.814 > 1.515
# critical value > chi-square value
# hence, accept H0


#p-value=1.515 > alpha=0.05
#so, we accept H0
```

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

```
##### using chi-square test

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


df2=pd.read_csv('C:/Users/RIG1/Desktop/DS ASSIGNMENTS/QUESTIONS -all
assignments/Ass 3/Costomer+OrderForm.csv')
df2.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
```

```
print(df2['Phillippines'].value_counts(),df2['Indonesia'].value_counts
(),df2['Malta'].value_counts(),df2['India'].value_counts())
```

```
Error Free    271
Defective      29
Name: Phillippines, dtype: int64 Error Free    267
Defective      33
Name: Indonesia, dtype: int64 Error Free    269
Defective      31
Name: Malta, dtype: int64 Error Free    280
Defective      20
Name: India, dtype: int64
```

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

```
expected = chi2_contingency([[271,267,269,280],[29,33,31,20]])
expected
```

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

```
## p-value 0.277 > alpha value 0.05
# hence accept H0
```