

Assignment_03: Hypothesis Testing

Shrawan Kumar Kumawat

Q.1.

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

Solution:

Level of significance = 0.05

Let , the null hypothesis be , $H_0: \mu_1 = \mu_2$ H_0 : (There is no difference in diameters of cutlets between two units)

And alternative hypothesis be , $H_a: \mu_1 \neq \mu_2$ H_a : (There is significant difference in diameters of cutlets between two units)

Here , 2 Samples 2 tail test is applicable.

```
In [13]: import numpy as np
import pandas as pd
data=pd.read_csv('cutlets.csv')
data.head()
```

Out[13]: **Unit A Unit B**

	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

In [10]:

```
In [19]: from scipy import stats
tstatistics, pvalue=stats.ttest_ind(data['Unit A'], data['Unit B'], equal_var=True)

In [21]: if pvalue < 0.05:
    print('Reject the null hypothesis at 5% level of significance and conclude that ,')
else:
    print('do not reject the null hypothesis at 5% level of significance and conc]
```

do not reject the null hypothesis at 5% level of significance and conclude that , there is no difference in diameters of cutlets between two units.

Hence , p-value > 0.05 , at 5% level of significance we do not reject the null hypothesis and conclude that, there is no difference in diameters of cutlets between two units.

-----XX-----

-----XX-----

Q.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 and determine whether there is any difference in average

TAT among the different laboratories at 5% significance level.

.... Minitab File: LabTAT.mtw

Solution

Level of significance = 0.05

null hypothesis is,

H0 : There is no difference in average TAT among the different laboratories.

And alternative hypothesis be,

H1 : There is a difference in average TAT among the different laboratories.

we will use one_way anova test

```
In [24]: import numpy as np
import pandas as pd
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

```
In [29]: statistic,p_val=stats.f_oneway(data['Laboratory 1'],data['Laboratory 2'],data['Laboratory 3'])
p_val
```

Out[29]: 2.1156708949992414e-57

```
In [32]: if p_val < 0.05:
    print('Reject the null hypothesis at 5% level of significance and conclude that ,')
else:
    print('do no reject the null hypothesis at 5% level of significance and conclude that ,')
```

Reject the null hypothesis at 5% level of significance and conclude that , there is a difference in average TAT among the different laboratories

Here , p-value < 0.05 , at 5% level of significance we reject the null hypothesis

and conclude that, there is a difference in average TAT among the different laboratories.

-----+ + + + +-----
-----+ + + + +-----+ + +

Q.3.

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

file : Buyer Ratio.mtw

Solution

Level of significance = 0.05

null hypothesis is,

H0 : There is no significant association between gender and regions.

And alternative hypothesis is,

Ha : There is a significant association between gender and regions.

for that we have to use chi square test

```
In [25]: import numpy as np  
import pandas as pd
```

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

```
Out[27]:
```

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

```
In [28]: data=data.iloc[:,1:]  
data
```

Out[28]:

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

In [29]:

```
chi_square_test_val,pval,dof,expected_table=stats.chi2_contingency(data)
print('p_value is : ',pval)
```

p_value is : 0.6603094907091882

In [33]:

```
if pval < 0.05:
    print('Reject the null hypothesis at 5% level of significance and conclude that ,')
else:
    print('Do not reject the null hypothesis at 5% level of significance and conclude')
```

Do not reject the null hypothesis at 5% level of significance and conclude that, there is no significant association between gender and regions

Here , p-value > 0.05 , at 5% level of significance we do not reject the null hypothesis and conclude that,

there is no significant association between gender and regions.

=====XXXX=====

Q.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

Minitab File: CustomerOrderForm.mtw

Answer :

Level of significance = 0.05

null hypothesis is,

H_0 : There is no significant association between centers and defectives.

And alternative hypothesis is,

H_a : There is a significant association between centers and defectives.

```
In [34]: import numpy as np
import pandas as pd
```

```
In [37]: data=pd.read_csv('CustomerOrderForm.csv')
data
```

Out[37]:

	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 \times 4 columns

```
In [46]: #taking the value_counts in list
error_free=[]
defactive=[]
for i in range(0,4):
    x,y=data.iloc[:,i].value_counts()
    error_free.append(x)
    defactive.append(y)
print(error_free)
print(defactive)
```

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

```
In [50]: # now calculate pvalue
```

```
statistic,p_val,df,exp_tab = stats.chi2_contingency([error_free, defactive])
round(p_val,4)
```

Out[50]: 0.2771

```
In [52]: if pval < 0.05:
    print('Reject the null hypothesis at 5% level of significance and conclude that ,')
else:
    print('Do not reject the null hypothesis at 5% level of significance and conclude')
```

Do not reject the null hypothesis at 5% level of significance and conclude that , there is no significant association between centers and defectives.

Hence , p-value > 0.05 , at 5% level of significance we do not reject the null hypothesis, and conclude that, there is no significant association between centers and defectives.

In []: