02052021 - Hypothesis Testing

February 27, 2021

0.1 Data Exploration

Check if there exists relation between one categorical and one categorical data

[1]:
```
import numpy as np
import pandas as pd
import scipy.stats as stats
```

[4]:
```
data = pd.read_csv('trainl.csv')
data.head()
```

[4]:

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area_Rural	Property_Area_Semiurban	Property_Area_Urban	Loan_Status
0	LP001002	Male	No	0	Graduate	No	5849	0.0	128	360	1	0	0	1	Y
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128	360	1	1	0	0	N
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66	360	1	0	0	1	Y
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120	360	1	0	0	1	Y
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141	360	1	0	0	1	Y

H0: Gender and Loan Status are Independet i.e. There is no relation between gender of the customer and the Loan Status

H1: Gender and Loan Status are not Independet i.e. There is relation between gender of the sutomer and the Loan Status

[6]:
```
tbl = pd.crosstab(data.Gender, data.Loan_Status)
display (tbl)
```

```
Loan_Status    N    Y
Gender
Female        37   75
Male         155  347
```

[8]:
```
chi_square, p_value, degrees_of_freedom, expected_frequencies = stats.
 ↪chi2_contingency(tbl)

print(expected_frequencies) # if the H0 is true what shout the rable look like
print ()
print(degrees_of_freedom)
print ()
print(p_value)
```

```
[[ 35.0228013  76.9771987]
 [156.9771987 345.0228013]]

1

0.7391461310869638
```

[10]:
```
alpha = 0.05

print (p_value < alpha)

if p_value < alpha:
    print("reject null hypothesis")
else:
    print("we do not rej null hypothesis")
```

```
False
we do not rej null hypothesis
```

Result :: There is no relation between gender of the customer and the Loan Status

1 if p_value is less than alpha, we reject the null hypothesis

Ho: Education and Loan Status are independent

Ha: Education and Loan Status are not independent

[11]:
```
tbl = pd.crosstab(data.Education, data.Loan_Status)
display (tbl)
print ()

chi_square , p_value, degrees_of_freedom, expected_frequencies=stats.
 ↪chi2_contingency(tbl)

print(p_value)
print ()
print(degrees_of_freedom)
print ()
print(expected_frequencies)
```

```
Loan_Status     N    Y
Education
Graduate      140  340
Not Graduate   52   82



0.04309962129357355

1

[[150.09771987 329.90228013]
 [ 41.90228013  92.09771987]]
```

[12]:
```
alpha = 0.05

print (p_value < alpha)

if p_value < alpha:
    print("reject null hypothesis")
else:
    print("we do not rej null hypothesis")
```

```
True
reject null hypothesis
```

Result :: Rej Ho. Education and Loan Status are not independent. There is relationship between education level and Loan Status.

[ ]:
```

```