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

1 sample Z test

Case Study 1 : The length of 25 samples of a fabric are taken at random. Historic Mean and standard deviation are 150 cm and 5.5 cm respectively. Company manager wants to test if the current mean is equal to the the historic mean. Assume α to be 0.05

Step 1. Formulation of Ho,Ha

- Ho: Mean = 150
- Ha: Mean != 150

Step 2. Select Level of significance

- alpha = 0.05

In [2]:
```
df1 = pd.read_excel("Fabric data.xlsx")
df1.head()
```

Out[2]:

	Fabric_length
0	151.2
1	160.3
2	147.5
3	149.2
4	159.2

Step 3. Check for Normality

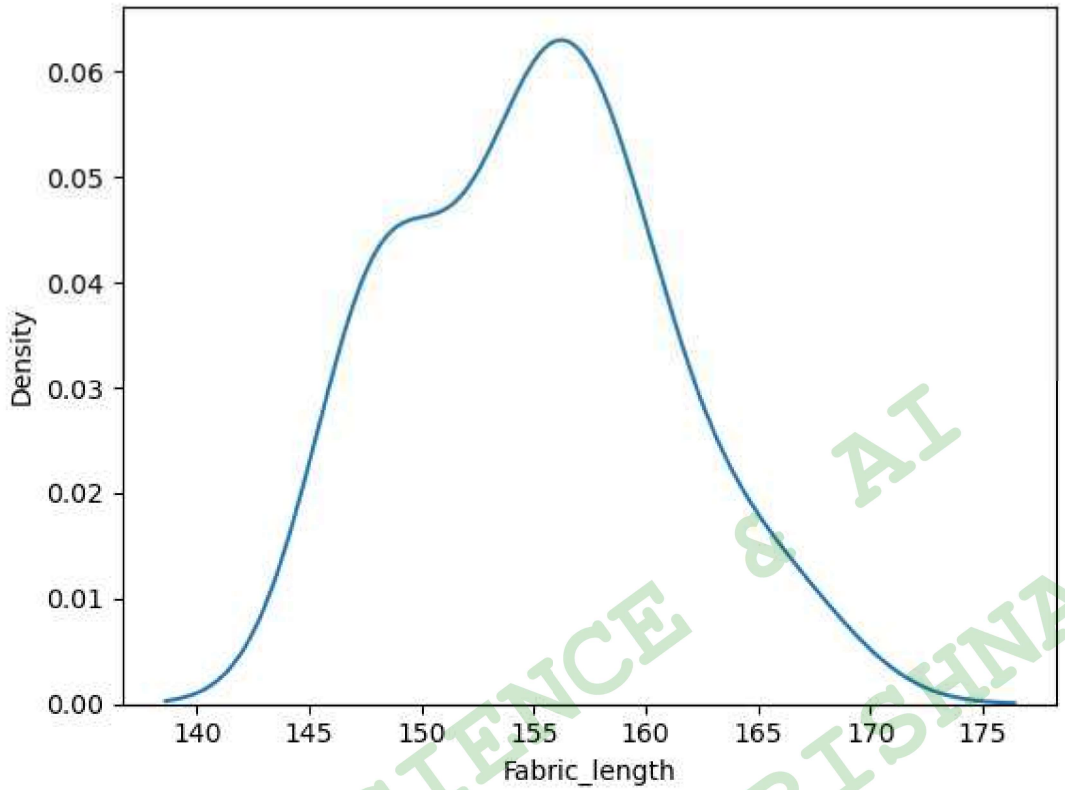
In [3]:
```
df1['Fabric_length'].skew()
```

Out[3]:
```
0.29650632012590666
```

or

In [4]:
```
import seaborn as sns
sns.kdeplot(df1['Fabric_length'])
```

Out[4]:
```
<Axes: xlabel='Fabric_length', ylabel='Density'>
```



or

Shapiro test (to check the data is normal or not)

- H0: Data is normal (p>0.05)
- H1: Data is not normal (p<0.05)

In [5]:
```
from scipy import stats

stats.shapiro(df1['Fabric_length'])
```

Out[5]:
```
ShapiroResult(statistic=0.9397523999214172, pvalue=0.1460934281349182)
```

Step 4. Select the statistical test and calculate p value

- 1 sample Z test (two tail rejection)

In [6]:
```
from statsmodels.stats.weightstats import ztest
ztest = ztest(df1["Fabric_length"], value=150,  alternative = 'two-sided')
ztest
```

Out[6]:
```
(4.488987748287781, 7.156241253356764e-06)
```

Step 5. Based on p value, Accept or Reject H0

- P < alpha
- P Low -- Null go
- Reject H0 --> Mean!=150 (action required)

1 Sample t test

Case Study 2 : The mean diameter of the bolt manufactured should be 10mm to be able to fit into the nut. 20 samples are taken at random from production line by a quality inspector. Conduct a test to check with 90% confidence that the mean is not different from the specification value.

1. Formulation of H0,H1

- Ho: Mean = 10 (no action)
- Ha: Mean != 10 (action)

2. Select Level of significance

- alpha = 0.1

In [7]:
```
df2 = pd.read_excel("Bolt diameter.xlsx")
df2.head()
```

Out[7]:

	Diameter
0	10.02
1	10.00
2	9.98
3	9.99
4	10.01

3. Check for Normality

In [8]:
```
df2['Diameter'].skew()
```

Out[8]:
```
0.0
```

4. Select the Statistical test & Calculate the p value

- 1 sample T test

In [9]:
```
ttest = stats.ttest_1samp(df2['Diameter'], 10 , alternative='two-sided')
ttest
```

Out[9]:
```
TtestResult(statistic=0.0, pvalue=1.0, df=19)
```

5. Based on p value, Accept or Reject H0

- P > alpha
- P High -- Null fly
- Do not Reject H0, continue with production

ANOVA Test

Case Study 3 : A marketing organization outsources their back-office operations to three different suppliers. The contracts are up for renewal and the CMO wants to determine whether they should renew contracts with all suppliers or any specific supplier. CMO want to

renew the contract of supplier with the least transaction time. CMO will renew all contracts if the performance of all suppliers is similar.

1. Formulation of H0,H1

- Ho: No difference in Mean transaction time of all suppliers
- Ha: difference in Mean transaction time of all suppliers

2. Select Level of significance

- alpha = 0.05

In [10]:
```
df3 = pd.read_excel("Contract Renewal.xlsx")
df3.head()
```

Out[10]:

	SupplierA	SupplierB	SupplierC
0	6.15	7.87	7.41
1	6.22	5.21	3.61
2	6.76	7.94	7.23
3	4.29	7.36	5.53
4	7.08	6.17	3.97

3. Check for Normality

In [11]:
```
df3["SupplierA"].skew(), df3["SupplierB"].skew(), df3["SupplierC"].skew()
```

Out[11]:
```
(0.12032960598073747, -0.06982967935882636, 0.25116833883092665)
```

4. Select the Statistical test & Calculate the p value

- ANOVA Test (F-Test)

In [12]:
```
anova_test = stats.f_oneway(df3["SupplierA"],df3["SupplierB"],df3["SupplierC"])
anova_test
```

Out[12]:
```
F_onewayResult(statistic=2.280378701368123, pvalue=0.10373295731933224)
```

5. Based on p value, Accept or Reject H0

- P > alpha
- P High -- Null fly
- Do not Reject H0, Renew contracts all the 3 Suppliers

1 Proportion Test

Case Study 4: The people carry out a poll to find the acceptability of new football coach. It was decided that if the support rate for the coach for the entire population was truly less than 90%, the coach would be fired. 2000 people participated and 496 people supported the new coach. Conduct a test to check if the new coach should be fired with 95% level of confidence.

1. Formulation of H0,H1

- H0:Coach not to be fired (No action) -No diiference
- H1:Coach to be fired (action)

2. Select Level of significance

- alpha = 0.05

4. Select the Statistical test & Calculate the p value

- 1 proportion test

In [13]:
```
binom_test = stats.binomtest(496, 2000, 0.9)
binom_test
```

Out[13]:
```
BinomTestResult(k=496, n=2000, alternative='two-sided', statistic=0.248, pvalue=0.
0)
```

5. Based on p value, Accept or Reject H0

- P < alpha
- P Low Null Go
- Reject H0

Chi-Square Test

Case Study 5 : Johnnie Talkers soft drinks division sales manager has been planning to launch a new sales incentive program for their sales executives. The sales executives felt that adults (>40 yrs) won't buy, children will & hence requested sales manager not to launch the program. Analyze the data & determine whether there is evidence at 5% significance level to support the hypothesis

1. Formulation of H0,H1

- H0 -> Proportions of Adults = Proportions of Children
- H1 -> Proportions of Adults != Proportions of Children

2. Select Level of significance

- alpha = 0.05

4. Select the Statistical test & Calculate the p value

In [14]:
```
df5=pd.read_excel("JohnyTalkers.xlsx")
df5.head()
```

Out[14]:

	Person	Drinks
0	Adults	Did Not Purchase
1	Adults	Did Not Purchase
2	Adults	Did Not Purchase
3	Adults	Did Not Purchase
4	Adults	Did Not Purchase

In [15]:
```
df5.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1220 entries, 0 to 1219
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   Person  1220 non-null   object
 1   Drinks  1220 non-null   object
dtypes: object(2)
memory usage: 19.2+ KB
```

In [16]:
```
count = pd.crosstab(df5["Person"],df5["Drinks"],margins=True)
count
```

Out[16]:

Drinks	Did Not Purchase	Purchased	All
Person			
Adults	422	58	480
Children	588	152	740
All	1010	210	1220

In [17]:
```
chisquare_test = stats.chi2_contingency(count)
chisquare_test
```

Out[17]:
```
Chi2ContingencyResult(statistic=14.613224681541515, pvalue=0.005574456386286158, d
of=4, expected_freq=array([[ 397.37704918,   82.62295082,  480.        ],
       [ 612.62295082,  127.37704918,  740.        ],
       [1010.        ,  210.        , 1220.        ]]))
```

5. Based on p value, Accept or Reject H0

- P < alpha
- P low Null go
- Reject H0
- Accept alternate hypothesis i.e. Unequal proportions