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

import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
from scipy import stats
```

Fabric Data

In [3]:
```
fabric = pd.read_excel("Fabric data.xlsx")
fabric.head()
```

Out[3]:

	Fabric_length
0	151.2
1	160.3
2	147.5
3	149.2
4	159.2

Step 1. Formulation of Ho,Ha

- Ho: Mean >= 150
- Ha: Mean < 150

Step 2. Select Level of significance

- alpha = 0.05

Step 3. Check for Normality

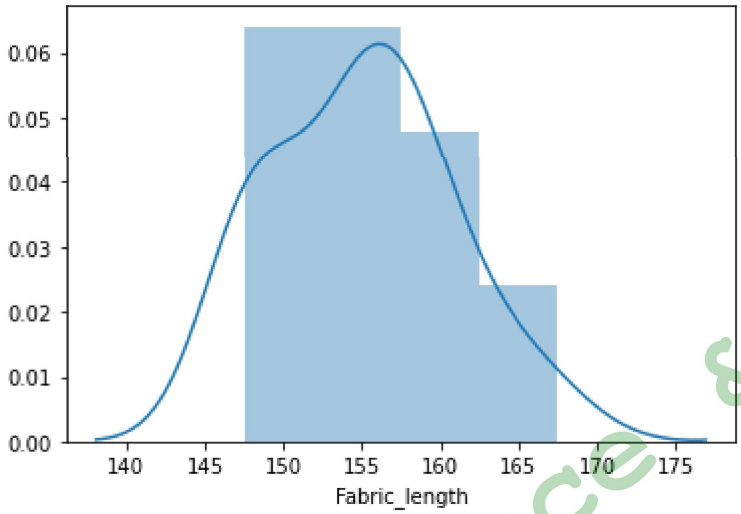
In [4]:
```
# option1 -- Skewness

fabric['Fabric_length'].skew()
```

Out[4]: 0.29650632012590666

In [5]:
```
# option2 -- Density Curve

sns.distplot(fabric['Fabric_length'])
plt.show()
```



In [6]:
```
# option3 -- Shapiro Test

stats.shapiro(fabric["Fabric_length"])
```

Out[6]:
```
ShapiroResult(statistic=0.9397523999214172, pvalue=0.14609353244304657)
```


Step 4. Select the statistical test and calculate p value

- 1 sample Z test

In [7]:
```
fabric["Fabric_length"].mean()
```

Out[7]:
```
155.06399999999996
```

In [8]:
```
Z_cal = (155.06-150)/(4/np.sqrt(25))
print(Z_cal)
```

```
6.325000000000003
```

In [9]:
```
stats.norm.cdf(Z_cal)
```

Out[9]:
```
0.9999999998733837
```

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

- P > alpha
- P High -- Null Fly
- Do not Reject H0.

Bolt Diameter

In [10]:
```
bolt = pd.read_excel("Bolt diameter.xlsx")
bolt.head()
```

Out[10]:

	Diameter
0	11.27
1	12.06
2	12.15
3	9.89
4	10.82

1. Formulation of H0,H1

- Ho: Mean = 10
- Ha: Mean != 10

2. Select Level of significance

- alpha = 0.05

3. Check for Normality

In [11]:
```
bolt['Diameter'].skew()
```

Out[11]: 0.380916607218941

4. Select the Statistical test & Calculate the p value

- 1 sample T test

In [12]:
```
stats.ttest_1samp(bolt.Diameter, 10)
```

Out[12]: Ttest_1sampResult(statistic=4.1146847709314756, pvalue=0.0005896412896356807)

5. Based on p value, Accept or Reject H0

- P < alpha
- P Low -- Null go
- Reject H0 (Average bolt diameter is not equal to 10) --> take action

Creditcard Promotion

In [13]:
```
promotion=pd.read_excel("Promotion.xlsx")
promotion.head()
```

Out[13]:

	InterestRateWaiver	StandardPromotion
0	1989.10	1272.25
1	1808.38	1250.38
2	1153.75	1474.78
3	1745.64	2064.89
4	1008.24	2030.87

1. Formulation of H0,H1

- Ho: Avg of purchases made by FIW < = Avg purchases made by SC =>default/ current/ no action
- Ha: Avg of purchases made by FIW > Avg purchases made by SC =>take action

2. Select Level of significance

- alpha = 0.05

3. Check for Normality

In [14]:
```
print(stats.shapiro(promotion.InterestRateWaiver))
```

```
ShapiroResult(statistic=0.9923660159111023, pvalue=0.22453102469444275)
```

In [15]:
```
print(stats.shapiro(promotion.StandardPromotion))
```

```
ShapiroResult(statistic=0.9919784665107727, pvalue=0.19155508279800415)
```

Variance test - Levene Test

- Ho: Variances are equal
- Ha: Variances are not equal

In [16]:
```
stats.levene(promotion.InterestRateWaiver, promotion.StandardPromotion)
```

Out[16]:
```
LeveneResult(statistic=1.1334674473666406, pvalue=0.2875528565130808)
```

4. Select the Statistical test & Calculate the p value

- 2 Sample t test for equal variance

In [17]:
```
stats.ttest_ind(promotion.InterestRateWaiver,promotion.StandardPromoti
_var=True)
```

Out[17]:
```
Ttest_indResult(statistic=2.260425163136941, pvalue=0.02422584468584312)
```

5. Based on p value, Accept or Reject H0

- P < alpha
- P Low -- Null go
- Reject H0 (FIW<=SC) --> take action

In [18]:
```
### 2 Sample t test for unequal variance
#stats.ttest_ind(promotion.InterestRateWaiver,promotion.StandardPromotion,equa
l_var=False)
```

Contract Renewal

In [19]:
```
cof=pd.read_excel("Contract Renewal.xlsx")
cof.head()
```

Out[19]:

	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

1. Formulation of H0,H1

- Ho: Average time by all suppliers are equal
- Ha: Average time by all suppliers are not equal

2. Select Level of significance

- alpha = 0.05

3. Check for Normality

In [20]:
```
print(stats.shapiro(cof.SupplierA))
print(stats.shapiro(cof.SupplierB))
print(stats.shapiro(cof.SupplierC))
```

```
ShapiroResult(statistic=0.9940784573554993, pvalue=0.8961844444274902)
ShapiroResult(statistic=0.9912325143814087, pvalue=0.6483432650566101)
ShapiroResult(statistic=0.9904154539108276, pvalue=0.5719023942947388)
```

In [21]:
```
############ Variance test ############
stats.levene(cof.SupplierA,cof.SupplierB,cof.SupplierC)  # All 3 suppl
being checked for variances
```

Out[21]:
```
LeveneResult(statistic=0.25183988720942463, pvalue=0.7775071819400866)
```

In [22]:
```
############ Variance test #########
print(stats.levene(cof.SupplierA, cof.SupplierB))
print(stats.levene(cof.SupplierB, cof.SupplierC))
print(stats.levene(cof.SupplierC, cof.SupplierA))
```

```
LeveneResult(statistic=0.03382395609148779, pvalue=0.8542383784793752)
LeveneResult(statistic=0.24333902699144816, pvalue=0.6222596854892893)
LeveneResult(statistic=0.4712013339466693, pvalue=0.49310323292578306)
```

4. Select the Statistical test & Calculate the p value

- Anova Test (F test)

In [23]:
```
stats.f_oneway(cof.SupplierA ,cof.SupplierB ,cof.SupplierC)
```

Out[23]:
```
F_onewayResult(statistic=2.280378701368123, pvalue=0.10373295731933224)
```

5. Based on p value, Accept or Reject H0

- P > alpha
- P High Null Fly
- All the 3 suppliers have equal mean transaction time

Football

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 [24]:
```
stats.binom_test(482, 2000, 0.25)
```

Out[24]:
```
0.36615115821892597
```

5. Based on p value, Accept or Reject H0

- P > alpha
- P High Null Fly
- Do not Reject H0

Johny Talkers

In [25]:
```
df=pd.read_excel("JohnyTalkers.xlsx")
df.head()
```

Out[25]:

	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 [26]:
```
df["Person"].value_counts()
```

Out[26]:
```
Children    740
Adults      480
Name: Person, dtype: int64
```

In [27]:
```
df["Drinks"].value_counts()
```

Out[27]:
```
Did Not Purchase    1010
Purchased            210
Name: Drinks, dtype: int64
```

In [28]:
```
# crosstable
pd.crosstab(df["Person"],df["Drinks"],margins=True)
```

Out[28]:

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

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

- 2 proportion test

In [29]:
```
from statsmodels.stats.proportion import proportions_ztest

count = np.array([58, 152]) #How many adults and childeren are purchasing
nobs = np.array([480, 740]) #Total number of adults and childern are there

proportions_ztest(count,nobs,alternative='two-sided')
```

Out[29]:
```
(-3.8227247718795447, 0.00013198507287726183)
```

5. Based on p value, Accept or Reject H0

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

Bahaman

In [30]:
```
Bahaman=pd.read_excel("Bahaman.xlsx")
Bahaman.head()
```

Out[30]:

	Defective	Country
0	0	India
1	0	India
2	0	India
3	0	India
4	1	India

```
In [31]: count=pd.crosstab(Bahaman["Defective"],Bahaman["Country"])
         count
```

Out[31]:

Country / Defective	Bangladesh	China	India	Srilanka
0	183	179	175	178
1	17	21	25	22

1. Formulation of H0,H1

- H0 : No difference in proportion between countries
- H1 : difference in proportion between countries

2. Select Level of significance

- alpha = 0.05

4. Select the Statistical test & Calculate the p value

- Chisquare test

```
In [32]: stats.chi2_contingency(count)
```

```
Out[32]: (1.7243932538050184,
          0.6315243037546223,
          3,
          array([[178.75, 178.75, 178.75, 178.75],
                 [ 21.25,  21.25,  21.25,  21.25]]))
```

```
In [33]: Chisquares_results=stats.chi2_contingency(count)
         print('p-value:',Chisquares_results[1])
```

```
p-value: 0.6315243037546223
```

5. Based on p value, Accept or Reject H0

- P > alpha
- P high Null fly
- Accept H0