

## # LAB\_ASSIGNMENT\_03

Using Numpy,pandas and Matplotlib demonstrate the following operations for the sample dataset given,

- i) Central tendency
- ii) Dispersion and Distribution
- iii) ANOVA
- iv) Hypothesis testing

## # PRELIMINARY INFORMATION

"""

```
import pandas as pd
import numpy as np
df=pd.read_excel('IBM-313 Marks.xlsx')
print(df.head())
```

```
print(df.columns)
```

## """# CENTRAL TENDENCY"""

```
import scipy
from scipy import stats
data = df['Total']
print('MEAN = ',scipy.mean(data))
print('MEDIAN = ',scipy.median(data))
print('MODE = ',stats.mode(data))
```

```
from scipy import stats
x=df['Total']
y=np.array(x)
print('Percentile = ',np.percentile(y,30))
```

## """# DISPERSION"""

```
range=max(y)-min(y)
print("RANGE = ",range)
```

```
Q1 = np.percentile(y,25)
Q3 = np.percentile(y,75)
print("IQR = ",Q3-Q1)
print("VARIANCE = ",np.var(y))
```

```
import statistics
print("POPULATION STANDARD DEVIATION = ",statistics.pstdev(y))
print("SAMPLE STANDARD DEVIATION = ",statistics.stdev(y))
```

```
from scipy.stats import skew
print(skew(y))
```

## """# HYPOTHESIS TESTING"""

```
pp_df=pd.read_excel('passport.xlsx')
```

```
pp_df.head()

print(list(pp_df.processing_time))

import math
def z_test(p_mean,p_std,sample):
    z_score = (sample.mean() - p_mean)/(p_std/math.sqrt(len(sample)))
    return z_score,stats.norm.cdf(z_score)

z_test(30,12.5,pp_df.processing_time)

"""# ANOVA"""

anova_df = pd.read_excel('discounts.xlsx')

anova_df.head()

import seaborn as sns
sns.distplot(anova_df['discount_0'],label = 'No Discount')
sns.distplot(anova_df['discount_10'],label = '10% Discount')
sns.distplot(anova_df['discount_20'],label = '20% Discount')
plt.legend()

from scipy.stats import f_oneway
f_oneway(anova_df['discount_0'],anova_df['discount_10'],anova_df['discount_20'])
```