Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

Question – 1 :-

Problem Statement:- A wholesale distributor operating in different regions of India has information on annual spending of several items in their stores across different regions and channels. The data consists of 440 large retailers' annual spending on 6 different varieties of products in 3 different regions (Gujarat, West Bengal, Other) and across different sales channel (Hotel, Retail).

Pre-Requisites :-

Code :-

```
import pandas as pd
Dataset = pd.read_excel("./T1LA Dataset - 1.xlsx")
Dataset["Total"] = Dataset["Fresh"] + Dataset["Milk"] + Dataset["Grocery"] + Dataset["Frozen"] +
Dataset["Detergents_Paper"] + Dataset["Delicassen"]
```

1) Use methods of descriptive statistics to summarize data. Which Region and which Channel seems to spend more? Which Region and which Channel seems to spend less?

Code :-

```
Q_1_Dataset_Channel = Dataset.groupby("Channel")["Total"].mean()
Q_1_Dataset_Region = Dataset.groupby("Region")["Total"].mean()
print(f"Channel Wise Spendings :- \n{Q_1_Dataset_Channel}")
print(f"Region Wise Spendings :- \n{Q_1_Dataset_Region}")
```

Output:-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> & "C:/Program Files/Python312/p
ython.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_1/Q_1.py"
Channel Wise Spendings :-
Channel
1    26844.191275
2    46619.232394
Name: Total, dtype: float64
Region Wise Spendings :-
Region
1    30997.571429
2    33086.978723
3    33789.870253
Name: Total, dtype: float64
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1>
```

2) There are 6 different varieties of items are considered. Do all varieties show similar behavior across Region and Channels?

Code :-

```
Q_2_Dataset = Dataset.groupby(["Channel" , 'Region'])['Total'].agg(['mean' , 'median'])
print(f"Channel and Region wise Analysis :- {Q_2_Dataset}")
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> & "C:/Program Files/Python312/python.
exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_1/Q_1.py"
Channel and Region wise Analysis :-                          mean    median
Channel Region
1       1       26073.593220  22287.0
        2       25683.928571  22134.5
        3       27213.635071  19789.0
2       1       47137.277778  41989.0
        2       43996.736842  36446.0
        3       47004.971429  37085.0
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1>
```

3) Based on a descriptive measure of variability, which item shows the most inconsistent behavior? Which items show the least inconsistent behavior?

Code :-

```
Items = ["Fresh", "Milk", "Grocery", "Frozen", "Detergents_Paper", "Delicassen"]
Standatd_Deviation = []

for item in Items :
    Standatd_Deviation.append(Dataset[item].std())
Q_3_Answer = pd.DataFrame()
Q_3_Answer["Items"] = Items
Q_3_Answer["Standatd Deviation"] = Standatd_Deviation
print("Product Wise Standard Deviation :- ")
print(Q_3_Answer.to_string(index=False))
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> & "C:/Program Files/Python312/python.
exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_1/Q_1.py"
Product Wise Standard Deviation :-
           Items  Standatd Deviation
           Fresh        12647.328865
            Milk         7380.377175
         Grocery         9503.162829
          Frozen         4854.673333
Detergents_Paper         4767.854448
      Delicassen         2820.105937
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1>
```

4) Are there any outliers in the data?

Code :-

```
def detect_outliers_iqr(data):
    Q1 = data.quantile(0.25)
    Q3 = data.quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = data[(data < lower_bound) | (data > upper_bound)]
    return list(outliers)


Items = ["Fresh", "Milk", "Grocery", "Frozen", "Detergents_Paper", "Delicassen"]
outliers_list = []

for item in Items:
    outliers_list.append(detect_outliers_iqr(Dataset[item]))

Q_4_Answer = pd.DataFrame({"Items": Items, "Outliers": outliers_list})

print(Q_4_Answer.to_string(index = False))
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> & "C:/Program Files/P
ython312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_1/Q_1
.py"
           Items

                                                                                                                  Outliers
           Fresh
                                                                                                [43088, 56159, 44466, 40721, 43265, 56082, 76237, 42312, 45640, 112151
, 47493, 56083, 53205, 49063, 68951, 40254, 42786, 39679, 38793, 39228]
            Milk                                                                                                                                                                  [36423, 2048
4, 15729, 22044, 54259, 21412, 29892, 38369, 20959, 46197, 73498, 27472, 16729, 15726, 25862, 29627, 43950, 28326
, 16599, 23133, 17972, 23527, 20655, 25071, 16784, 18664, 21858, 16687]
         Grocery
                                                                                                [25957, 23998, 55571, 28921, 26866, 59598, 45828, 28540, 92780, 32114, 32034, 28986, 34792, 26870
, 24708, 23596, 39694, 36486, 33586, 24773, 26839, 67298, 26316, 30243]
          Frozen [9408, 10002, 9510, 7782, 10643, 8872, 8132, 9735, 8693, 35009, 18028, 8853, 16538, 8195, 8425,
16745, 36534, 7888, 18711, 8321, 11422, 10155, 16919, 10303, 8692, 8366, 12569, 60869, 7849, 11559, 8170, 15601,
9584, 8164, 9927, 8620, 13223, 9806, 17866, 15348, 15082, 13486, 13135]
Detergents_Paper                                                                                                                                                                 [9529, 24171, 13583, 1
7740, 26701, 24231, 12034, 40827, 20070, 18906, 12591, 11577, 13726, 9836, 9959, 14235, 12420, 19410, 13308, 1859
4, 10069, 11783, 17120, 12408, 9606, 38102, 15469, 12218, 12638, 14841]
      Delicassen
                                                                                                [7844, 5185, 4478, 4334, 16523, 5778, 5206, 4626, 5864, 4985, 6465, 14472, 14351, 5130, 4430, 6
250, 8550, 47943, 6854, 6372, 5121, 5609, 4100, 4829, 5120, 5137, 4365]
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1>
```

5) Based on this report, what are the recommendations?

- Here there are too many outliers are there so for analysis purpose we should consider Z-Score

Question – 2 :-

1) Write a code in C/C++/Java/Python to take numbers as an input and gives output five points summary of box plot. It should also declare outlier if any.

Code :-

```
def Calculate_Median(Data) :
    if(len(Data) % 2 == 0) :
        return (Data[len(Data) // 2] + Data[len(Data) // 2 - 1]) / 2
    else :
        return Data[(len(Data) // 2)]

def Five_Plot_Summary(Data) :
    Q2 = Calculate_Median(Data)
    Q1 = 0
    Q3  = 0
    Lower_Quartile = []
    Upper_Quartile = []
    if(len(Data) % 2 == 0) :
        Lower_Quartile = Data[:((len(Data) // 2))]
        Upper_Quartile = Data[len(Data) // 2 :]

    else :
        Lower_Quartile = Data[:((len(Data) // 2))]
        Upper_Quartile = Data[(len(Data) // 2) + 1 :]

    Q1 = Calculate_Median(Lower_Quartile)
    Q3 = Calculate_Median(Upper_Quartile)

    IQR = Q3 - Q1
    Min = min(Data)
    Max= max(Data)
    Lower_Limit = Q1 - (1.5 * IQR)
    Upper_Limit = Q3 + (1.5 * IQR)
    Outliers = []

    for i in Data :
        if(i < Lower_Limit or i > Upper_Limit):
            Outliers.append(i)
```

```
    print(f"Five Point Summary of Box Plot For {Data} :\nMinimum = {Min}\nMaximum = {Max}\nQ1 = {Q1}\nQ2 = {Q2}\nQ3 = {Q3}\nIQR = {IQR}\nLower Limit = {Lower_Limit}\nUpper Limit = {Upper_Limit}\nOutliers = {Outliers}\n")


Input = input("Enter A Space Seprated Value :-")
Data = []

for i in Input.split(" "):
    Data.append(float(i))

Data_Sorted = Data.sort()
Five_Plot_Summary(Data_Sorted)
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_1/Q_2_1.py"
Enter A Space Seprated Value :-5 15 16 20 21 25 26 27 30 30 31 32 32 34 35 38 38 41 43 66
Five Point Summary of Box Plot For [5.0, 15.0, 16.0, 20.0, 21.0, 25.0, 26.0, 27.0, 30.0, 30.0, 31.0, 32.0, 32.0, 34.0, 35.0, 38.0, 38.0, 41.0, 43.0, 66.0] :
Minimum = 5.0
Maximum = 66.0
Q1 = 23.0
Q2 = 30.5
Q3 = 36.5
IQR = 13.5
Lower Limit = 2.75
Upper Limit = 56.75
Outliers = [66.0]
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

2) The nine measurements that follow are furnace temperature recorded on successive batches in a semiconductor manufacturing process (Units are in °C): 511, 513, 509, 515, 510.5, 509.4, 514, 512, 513.3

1) Calculate the sample mean, sample variance and sample standard deviation.

B) i) Sample mean $(\bar{x}) = \frac{\Sigma x_i}{N}$

$= \frac{4607.2}{9}$

$= 511.91$

Sample Variance :-

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
511	-0.91	0.8281
513	1.09	1.1881
509	-2.91	8.4681
515	3.09	9.5481
510.5	-1.41	1.9881
509.4	-2.51	6.3001
514	2.09	4.3681
512	0.09	0.0081
513.3	1.39	1.9321
		34.6289

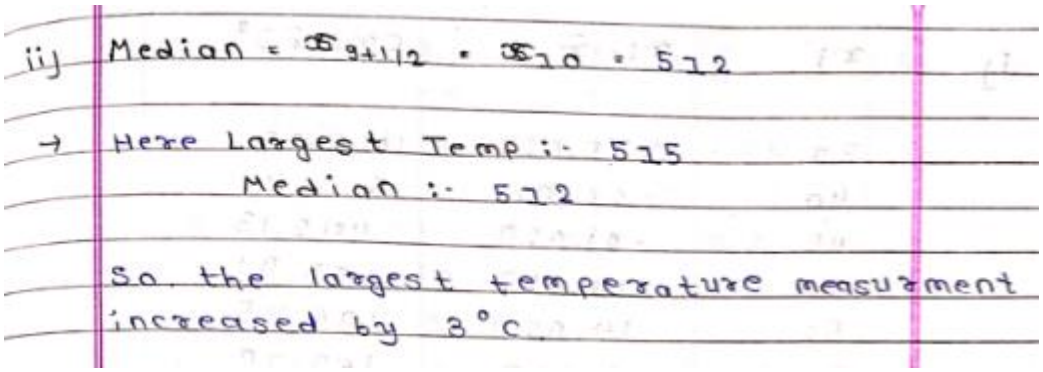
$s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1} = \frac{34.6289}{8} = 4.329$

Sample variance $(s^2) = 4.329$

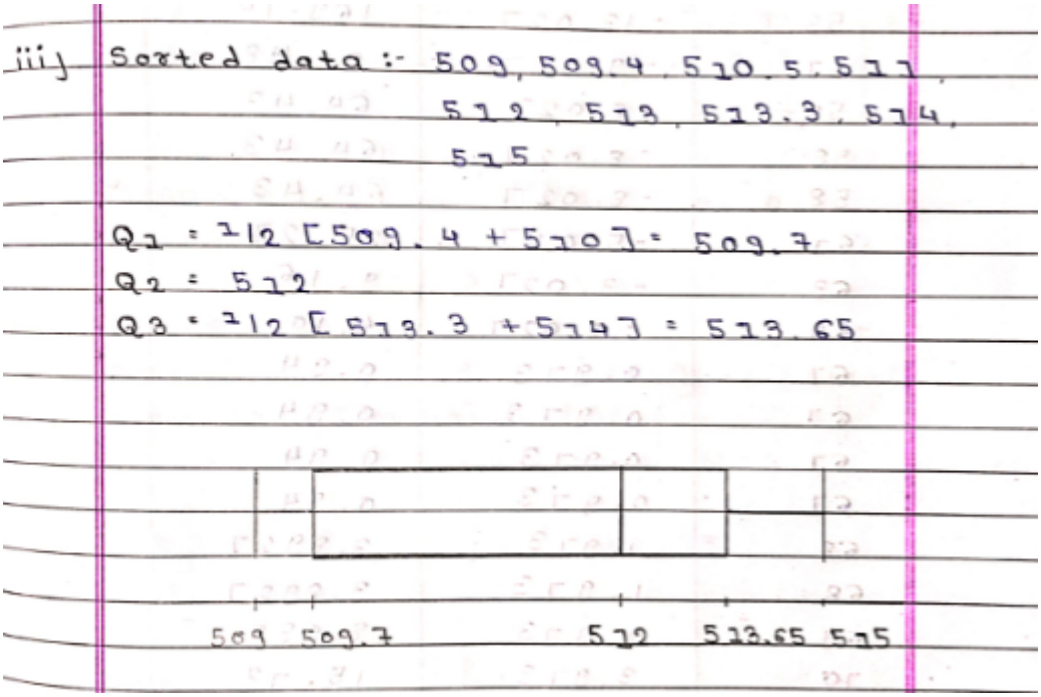
Sample standard deviation $(s) = (4.329)^{1/2}$

$= 2.08$

2) Find the median. How much could the largest temperature measurement increase without changing the median value?



3) Construct a box plot of the data.



Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

3) The following data are the joint temperature of the O-rings (°F) for each test firing or actual launch of the space shuttle rocket motor (from the presidential commission on the space shuttle challenger accident, Vol 1, pp. 129-131):

84, 49, 64, 63, 40, 70, 83, 78, 67, 52, 45, 67, 68, 53, 70, 67, 69, 75, 79, 61, 58, 70, 68, 81, 58, 76, 67, 79, 72, 75, 73, 76, 70, 58, 57, 31

1) Compute the mean, mode and standard deviation. Construct the dot plot of the graph

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
	31	-35.027	1226.89
	40	-26.027	677.4
	45	-21.027	442.13
	49	-17.027	289.91
	52	-14.027	196.75
	53	-13.027	169.70
	57	-9.027	81.48
	58	-8.027	64.43
	58	-8.027	64.43
	58	-8.027	64.43
	61	-5.027	25.27
	63	-3.027	9.16
	64	-2.027	4.10
	67	0.973	0.94
	67	0.973	0.94
	67	0.973	0.94
	67	0.973	0.94
	68	1.973	3.8927
	68	1.973	3.8927
	69	2.973	8.838
	70	3.973	15.78
	70	3.973	15.78
	70	3.973	15.78
	70	3.973	15.78
	70	3.973	15.78
	72	5.973	35.67
	73	6.973	48.62
	75	8.973	80.51

Marwadi University **Faculty of Technology** **Department of Information and Communication Technology**		
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

75	8.973	80.52
76	9.973	99.46
76	9.973	99.46
78	11.973	143.35
79	12.973	168.298
79	12.973	168.298
81	14.973	224.190
83	16.973	228.082
84	17.973	323.028
		5174.973

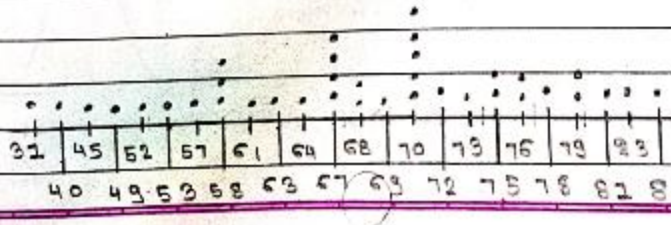
Mean $(\bar{x}) = \frac{\sum x_i}{N} = \frac{2443}{37} = 66.027$

Mode = 70

Standard Deviation $(SD) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N-1}}$

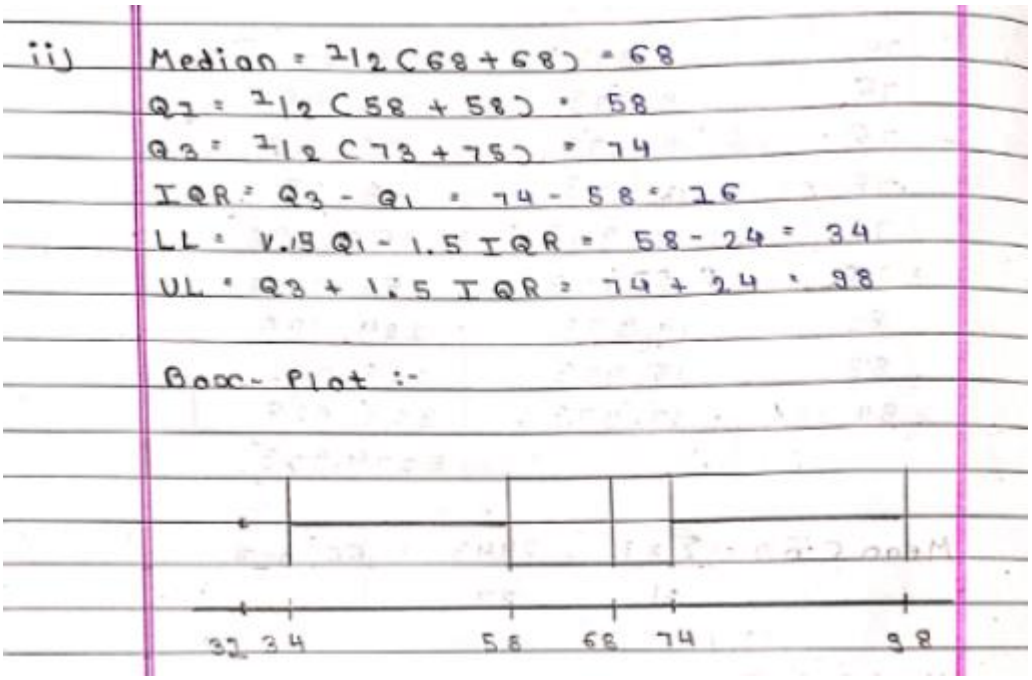
$= \sqrt{\frac{5174.973}{36}} = 143.75$

Dot plot :-

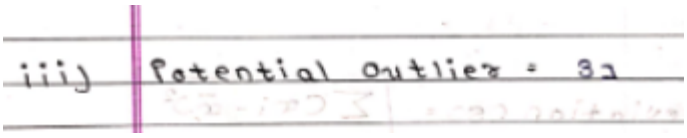


Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

2) Computer the median, upper and lower quartile and prepare box plot.



3) Is there any potential outlier? if yes, find the outlier data



4) For this data suggest the what are the two extremes out of which data will be considered as extreme outliers?

- There is Only 1 Outlier is this in this dataset

5) Set aside the lowest observation (31°F) and recompute the quantities in part (ii). Comment on your findings. How "different are the other temperatures from this lowest value?

- It 31° F is Already a Lowest Observation and the calculation will remain same

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

4) Write a script in python to find SD and Variance of any given data, test it by applying three different data set. Submit script and output of the code

Code :-

```
import math

def Calculate_SD(data_unsorted) :
    data = sorted(data_unsorted)
    mean = 0
    for i in data :
        mean = mean + i
    mean = mean / len(data)
    Sigma_Xi_x = 0
    for i in data :
        Sigma_Xi_x = Sigma_Xi_x + (i - mean) ** 2
    Sigma_Xi_x = Sigma_Xi_x / (len(data) - 1)
    Standard_Deviation = math.sqrt(Sigma_Xi_x)
    return Standard_Deviation
Input = input("Enter A Space Seprated Value :-")
Data = []
for i in Input.split(" "):
    Data.append(float(i))
Standard_Deviation = Calculate_SD(Data)
Variance = Standard_Deviation ** 2
print(f"For Data :- {Data} :- Standard Deviation = {Standard_Deviation}\nVariance = {Variance}")
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> & "C:/Program Files/Python312/python.
exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_1/Q_2_4.py"
Enter A Space Seprated Value :-1.1 2.3 3.5 4.7 5.9 7.1 8.3 9.5 10.7 11.9 13.1 14.3 15.5 16.7 17.9 19.1 20.3 21.5 22.7 23.9 25.1 2
6.3 27.5 28.7 29.9 31.1 32.3 33.5 34.7 35.9
For Data :- [1.1, 2.3, 3.5, 4.7, 5.9, 7.1, 8.3, 9.5, 10.7, 11.9, 13.1, 14.3, 15.5, 16.7, 17.9, 19.1, 20.3, 21.5, 22.7, 23.9, 25.1
, 26.3, 27.5, 28.7, 29.9, 31.1, 32.3, 33.5, 34.7, 35.9] :- Standard Deviation = 10.564090116995406
Variance = 111.60000000000001
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> & "C:/Program Files/Python312/python.
exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_1/Q_2_4.py"
Enter A Space Seprated Value :-0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10.0 10.5 11.0 11.5 12
.0 12.5 13.0 13.5 14.0 14.5 15.0
For Data :- [0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 10.5, 11.0, 11.
5, 12.0, 12.5, 13.0, 13.5, 14.0, 14.5, 15.0] :- Standard Deviation = 4.401704215414752
Variance = 19.375
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> & "C:/Program Files/Python312/python.
exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_1/Q_2_4.py"
Enter A Space Seprated Value :-0.9 1.8 2.7 3.6 4.5 5.4 6.3 7.2 8.1 9.0 9.9 10.8 11.7 12.6 13.5 14.4 15.3 16.2 17.1 18.0 18.9 19.8
 20.7 21.6 22.5 23.4 24.3 25.2 26.1 27.0
For Data :- [0.9, 1.8, 2.7, 3.6, 4.5, 5.4, 6.3, 7.2, 8.1, 9.0, 9.9, 10.8, 11.7, 12.6, 13.5, 14.4, 15.3, 16.2, 17.1, 18.0, 18.9, 1
9.8, 20.7, 21.6, 22.5, 23.4, 24.3, 25.2, 26.1, 27.0] :- Standard Deviation = 7.923067587746553
Variance = 62.774999999999984
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1>
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

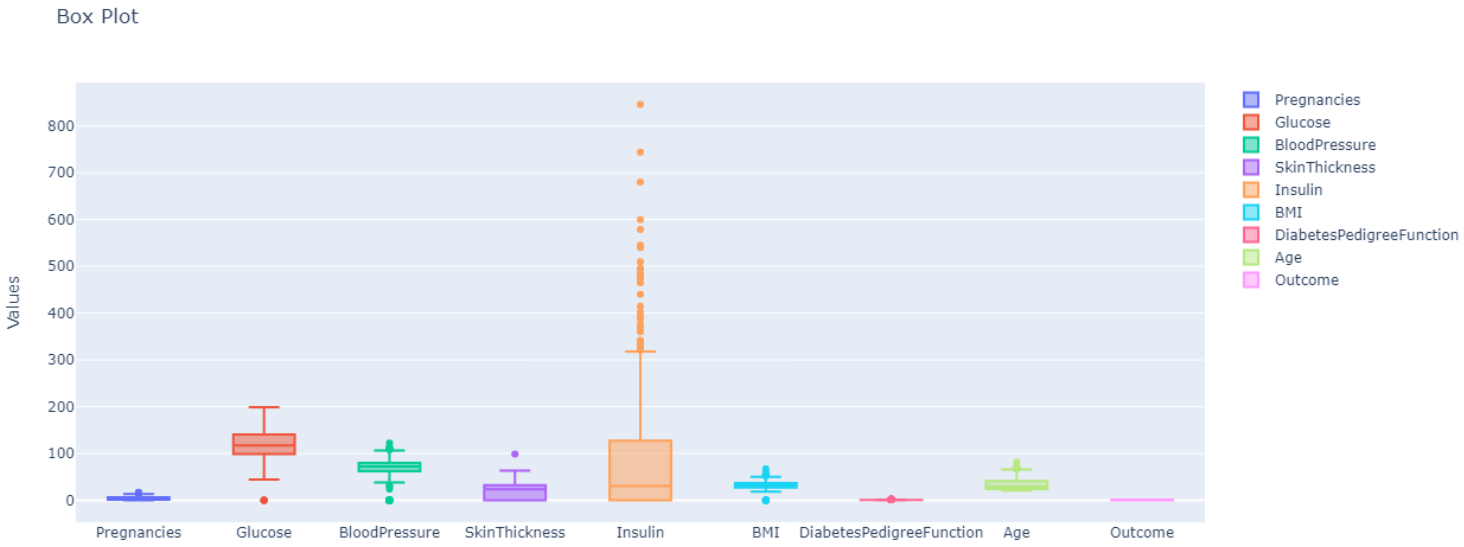
5) Using Python programming script and data (T1LA_dataset_2_diabetes.csv/excel) plot the box and whisker plot for all features and class. Write your inference from the box plots. Submit scripts and outputs.

Code :-

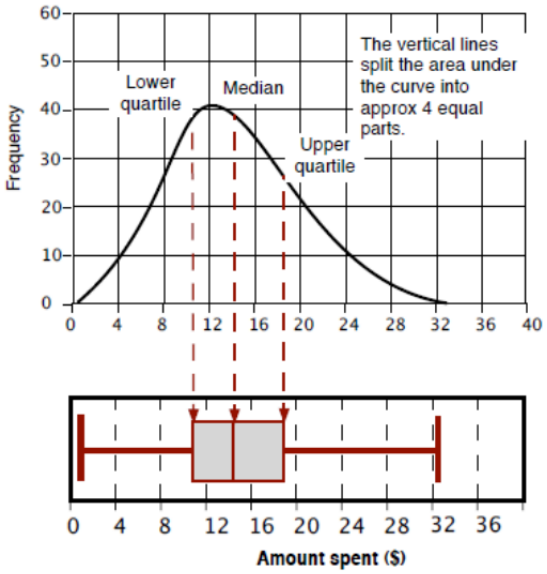
```
import pandas as pd
import plotly.graph_objs as go
Dataset = pd.read_excel("./T1LA Dataset - 1.xlsx")
def Calculate_Median(Data_Unsorted):
    Data = sorted(Data_Unsorted)
    if len(Data) % 2 == 0:
        return (Data[len(Data) // 2] + Data[len(Data) // 2 - 1]) / 2
    else:
        return Data[len(Data) // 2]
def Five_Plot_Summary(Data_Unsorted):
    Data = sorted(Data_Unsorted)
    Q2 = Calculate_Median(Data)
    Q1 = Calculate_Median(Data[: len(Data) // 2])
    Q3 = Calculate_Median(Data[len(Data) // 2 :])
    IQR = Q3 - Q1
    Min = min(Data)
    Max = max(Data)
    Lower_Limit = Q1 - (1.5 * IQR)
    Upper_Limit = Q3 + (1.5 * IQR)
    Outliers = [i for i in Data if i < Lower_Limit or i > Upper_Limit
for column in Dataset.columns:
    Data = list(Dataset[column])
    print(f"Five Point Summary of Box Plot For {column} :\n")
    Five_Plot_Summary(Data)
Traces = []
for column in Dataset.columns:
    Trace = go.Box(y=Dataset[column], name=column)
    Traces.append(Trace)
layout = go.Layout(title="Box Plot", yaxis=dict(title="Values"))
Figure = go.Figure(data=Traces, layout=layout)
Figure.write_html("Q_2_5.html")
Figure.show()
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

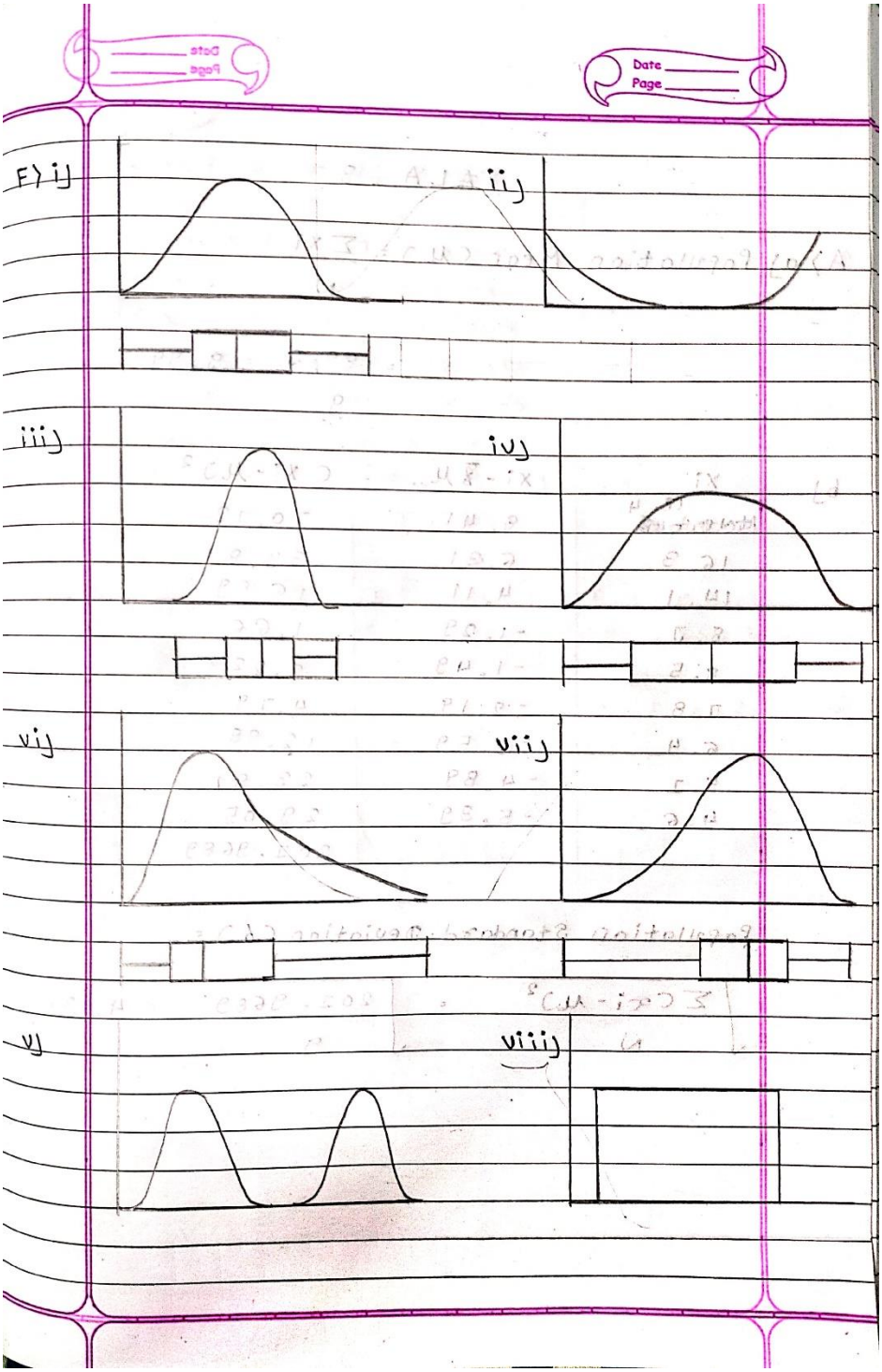
Output :-



6) We can correlate box plot and frequency plot as follows



Marwadi University **Faculty of Technology** **Department of Information and Communication Technology**		
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030



Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

7) The data of 30 customers on credit card usage in INR1000, sex (1:male, 2: female) and whether they have done shopping or banking (1: yes , 2: no) with credit card are given in file (T1LA_dataset_3_CC_Expenses_Exercise.csv/excel)

1) Import the file to Python
2) Compute descriptive summary of variable Credit Card Usage
3) Check whether the average usage varies with sex?
4) Check whether the average credit card usage vary with those who do shopping with credit card and those who don't do shopping?
5) Check whether the average credit card usage vary with those who do banking with credit card and those who don't do banking?
6) Compute the aggregate average of usage with sex & shopping?
7) Compute the aggregate average of usage with all three factors?

Code :-

```
import math

import pandas as pd

Dataset = pd.read_excel("./T1LA Dataset - 3.xlsx")
print(Dataset.columns)


def Calculate_Median(Data):
    if len(Data) % 2 == 0:
        return (Data[len(Data) // 2] + Data[len(Data) // 2 - 1]) / 2
    else:
        return Data[len(Data) // 2]


def Calculate_SD(data):
    mean = sum(data) / len(data)
    Sigma_Xi_x = sum((i - mean) ** 2 for i in data)
    Sigma_Xi_x = Sigma_Xi_x / (len(data) - 1)
    Standard_Deviation = math.sqrt(Sigma_Xi_x)
    return Standard_Deviation


def Descriptive_Statistics(Data):
    Frequency_Distribution = {}

    for i in Data:
        if i in Frequency_Distribution:
            Frequency_Distribution[i] += 1
        else:
```

```
        Frequency_Distribution[i] = 1

Frequency_Distribution = pd.DataFrame(
    list(Frequency_Distribution.items()), columns=["Observation", "Frequency"]
)

Frequency_Distribution = Frequency_Distribution.sort_values(
    by="Frequency", ascending=False
).reset_index()

Mean = sum(Data) / len(Data)
Median = Calculate_Median(Data)
Q1 = Calculate_Median(Data[: len(Data) // 2])
Q2 = Calculate_Median(Data)
Q3 = Calculate_Median(Data[len(Data) // 2 :])
Minimum = min(Data)
Maximum = max(Data)
IQR = Q3 - Q1
Lower_Limit = Q1 - (1.5 * IQR)
Upper_Limit = Q3 + (1.5 * IQR)
Range = Maximum - Minimum
No_of_Observation = len(Data)
Standard_Deviation = Calculate_SD(Data)
Variance = Standard_Deviation**2
Mode = [
    Frequency_Distribution["Observation"][i]
    for i in range(len(Frequency_Distribution["Observation"]))
    if Frequency_Distribution["Frequency"][i]
    == Frequency_Distribution["Frequency"].iloc[0]
]

print("Description of Crediti Card Usage")
print("Mean: ", Mean)
print("Median:", Median)
print("Mode:", Mode)
print("Q1:", Q1)
print("Q2:", Q2)
print("Q3:", Q3)
print("Minimum:", Minimum)
print("Maximum:", Maximum)
print("IQR:", IQR)
print("Lower Limit:", Lower_Limit)
print("Upper Limit:", Upper_Limit)
print("Range:", Range)
print("No. of Observations:", No_of_Observation)
print("Standard Deviation:", Standard_Deviation)
```

```
    print("Variance:", Variance)
Data = sorted(list(Dataset["Amount"]))
Descriptive_Statistics(Data)

Male = Dataset[Dataset["Gender"] == "M"]
Female = Dataset[Dataset["Gender"] == "F"]

if (
    (Male["Amount"].mean() <= 1.2 * Female["Amount"].mean())
    and (Male["Amount"].mean() >= 0.8 * Female["Amount"].mean())
) or (
    (Female["Amount"].mean() <= 1.2 * Male["Amount"].mean())
    and (Female["Amount"].mean() >= 0.8 * Male["Amount"].mean())
):
    print("The Average Uses are not Varies with Sex")
else:
    print("The Average Uses are Varies with Sex")

print(set(Dataset["Exp Type"]))


Do_Shopping = Dataset[
    (Dataset["Exp Type"] == "Grocery") | (Dataset["Exp Type"] == "Food")
]

Dont_Do_Shopping = Dataset[
    (Dataset["Exp Type"] != "Grocery") | (Dataset["Exp Type"] != "Food")
]

if (
    (Do_Shopping["Amount"].mean() <= 1.2 * Dont_Do_Shopping["Amount"].mean())
    and (Do_Shopping["Amount"].mean() >= 0.8 * Dont_Do_Shopping["Amount"].mean())
) or (
    (Dont_Do_Shopping["Amount"].mean() <= 1.2 * Do_Shopping["Amount"].mean())
    and (Dont_Do_Shopping["Amount"].mean() >= 0.8 * Do_Shopping["Amount"].mean())
):
    print("The Average Uses are not Varying")
else:
    print("The Average Uses are Varying")

Do_Banking = Dataset[(Dataset["Exp Type"] == "Bills")]
Dont_Do_Banking = Dataset[(Dataset["Exp Type"] != "Bills")]

if (
    (Do_Banking["Amount"].mean() <= 1.2 * Dont_Do_Banking["Amount"].mean())
    and (Do_Banking["Amount"].mean() >= 0.8 * Dont_Do_Banking["Amount"].mean())
```

```
) or (
    (Dont_Do_Banking["Amount"].mean() <= 1.2 * Do_Banking["Amount"].mean())
    and (Dont_Do_Banking["Amount"].mean() >= 0.8 * Do_Banking["Amount"].mean())
):
    print("The Average Uses are not Varying")
else:
    print("The Average Uses are Varying")

Shopping_And_Gender = Dataset[(Dataset["Exp Type"] == "Grocery") | (Dataset["Exp Type"] == "Food")]
Aggregate_Average_for_Gender_and_Shopping = Shopping_And_Gender.groupby(["Gender", "Exp
Type"])['Amount'].mean()
print(Aggregate_Average_for_Gender_and_Shopping)

Shopping_And_Gender_And_Banking = Dataset[
    (Dataset["Exp Type"] == "Grocery")
    | (Dataset["Exp Type"] == "Food")
    | (Dataset["Exp Type"] == "Bills")
]
Aggregate_Average_for_Gender_and_Shopping_and_Banking = (
    Shopping_And_Gender_And_Banking.groupby(["Gender", "Exp Type"])["Amount"].mean()
)
print(Aggregate_Average_for_Gender_and_Shopping_and_Banking)
```

Output :-

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

```
D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> python Q_2_7.py
Index(['index', 'City', 'Date', 'Card Type', 'Exp Type', 'Gender', 'Amount',
       'Customer_Age', 'Education_Level', 'Marital_Status'],
      dtype='object')
Description of Crediti Card Usage
Mean:  166179.31717191666
Median: 157300
Mode: [248858, 23427]
Q1: 79638
Q2: 157300
Q3: 233088.5
Minimum: 1024
Maximum: 998077
IQR: 153450.5
Lower Limit: -150537.75
Upper Limit: 463264.25
Range: 997053
No. of Observations: 10127
Standard Deviation: 124109.42253663488
Variance: 15403148762.376974
The Average Uses are not Varies with Sex
{'Entertainment', 'Bills', 'Grocery', 'Fuel', 'Food'}
The Average Uses are not Varying
The Average Uses are Varying
Gender  Exp Type
F       Food        152933.868185
        Grocery     152821.389019
M       Food        151890.358226
        Grocery     150911.378616
Name: Amount, dtype: float64
Gender  Exp Type
F       Bills       259785.090574
        Food        152933.868185
        Grocery     152821.389019
M       Bills       148789.771725
        Food        151890.358226
        Grocery     150911.378616
Name: Amount, dtype: float64

D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1>
```

8) Refer the research paper "Data analysis using Box and Whisker Plot for Lung Cancer" presented in International Conference on Innovations in Power and Advanced Computing Technologies [i-PACT2017] Write down five important/useful learning (which was not covered in any of the problem in this assignment) Suggested extra reading : Multidimensional box plot, different box plot width types

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

- Here Multidimensional Scatter Plot is used to visualize the relationship between Years of smoked to Year and age.
- Here Control Charts are used to analyze of Patient's cancer status.
- GG Plot is Used to analyze the relation between Smoking Status and Years Smoked
- A histogram is used to predict the Lung Cancer Chances.
- Here Box Plots are used to find the pattern between various attribute.

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

Question – 3:-

1) Population. The Census of India provides a variety of statistical information on different aspects of the Indian population. According to the Population Enumeration Data for the 2011 census, provided on the official website for the Indian census censusindia.gov.in, the top 9 metro cities on the basis of their population are as follows.
City Population (in million) Mumbai 18.4 Delhi 16.3 Kolkatta 14.1 Chennai 8.7 Bangalore 8.5 Hyderabad 7.8 Ahemdabad 6.4 Pune 5.1 Surat 4.6

a. Compute the population mean enrolment, μ, of the cities. (Round your answer to two decimal places.)
b. Compute σ. (Round your answer to two decimal places.)
c. Letting x denote a city, specify the standardized variable, z, corresponding to x.
d. Without performing any calculations, give the mean and standard deviation of z. Explain your answers.

Answer :-

A) a) Population Mean (μ) = $\frac{\Sigma x_i}{N}$

$= \frac{89.9}{9} = 9.99$

b)

x_i	$x_i - \mu$	$(x_i - \mu)^2$
18.4	8.41	70.72
16.3	6.31	39.81
14.1	4.11	16.89
8.7	-1.29	1.66
8.5	-1.49	2.22
7.8	-2.19	4.79
6.4	-3.59	12.88
5.1	-4.89	23.91
4.6	-5.39	29.05
		201.9689

Population Standard Deviation (σ) =

$\sqrt{\frac{\Sigma (x_i - \mu)^2}{N}} = \sqrt{\frac{201.9689}{9}} = 4.737$

c)

City Popu. (x_i)	Z - Score
18.4	1.7753
16.3	1.3320
14.1	0.8676
8.7	-0.2723
8.5	-0.3145
7.8	-0.4623
6.4	-0.7578
5.1	-1.032
4.6	-1.1378

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

2) Write a code in C/C++/Java/Python which can take input numbers and calculate Z score (Zi)against each data(Xi).

Code :-

```
import math
def Calculate_Population_Mean(Data):
    Population_Mean = sum(Data) / len(Data)
    return Population_Mean
def Calculate_Population_SD(Data):
    Mean = Calculate_Population_Mean(Data)
    Sum = 0
    for observation in Data:
        Sum += observation * observation
    Population_SD = math.sqrt((Sum / len(Data)) - (Mean**2))
    return Population_SD
def Calculate_Z_Score(Data):
    Population_Mean = Calculate_Population_Mean(Data)
    Population_SD = Calculate_Population_SD(Data)
    Z_Scores = []
    for Observation in Data:
        Z_Scores.append((Observation - Population_Mean) / Population_SD)
    return Z_Scores
Input = input("Enter A Space Seprated Value :-")
Data = []
for i in Input.split(" "):
    Data.append(float(i))
Z_Scores = Calculate_Z_Score(Data)
print(f"The Z-Scores for {Data} is :-\n{Z_Scores}")
```

Output:-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> & "C:/Program Files/Python312/python.
exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_1/Q_3_2.py"
Enter A Space Seprated Value :-5 15 16 20 21 25 26 27 30 30 31 32 32 34 35 38 38 41 43 66
The Z-Scores for [5.0, 15.0, 16.0, 20.0, 21.0, 25.0, 26.0, 27.0, 30.0, 30.0, 31.0, 32.0, 32.0, 34.0, 35.0, 38.0, 38.0, 41.0, 43.0
, 66.0] is :-
[-2.0481296593650064, -1.2369892002105833, -1.1558751542951353, -0.8314189706333429, -0.7503049247178948, -0.4258487410561025, -0.
34473469514065436, -0.2636206492252063, -0.020278511478862022, -0.020278511478862022, 0.06083553443658607, 0.14194958035203414, 0
.14194958035203414, 0.30417767218293035, 0.3852917180983784, 0.6286338558447226, 0.6286338558447226, 0.8719759935910669, 1.034204
0854219632, 2.899827141477269]
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1>
```

3) Refer the link
https://www.linkedin.com/advice/1/what-benefits-drawbacks-using-z-scores-standardize Topic “What are the benefits and drawbacks of using z-scores to standardize your data for predictive modeling?”
Write down learning from it.

- By using Z-score we can reduce the effect of outliers and skewness and the Z-score makes data more interpretable because it has a mean of 0 and a Standard Deviation of 1. Using Z score it is difficult to explain the result and z score makes distribution normal but it is not there every time. There are many alternative methods is also there which can be used instead of z-score like min-max scaling.
- We can compute z score easily in python import pandas calculate mean standard deviation and iterate loop through a list and append z score in a new list

Question – 4:-

1) Write a code in C/C++/Java/Python which can take input numbers and declared the following requirement
- **(i)** Mean, median, mode
- **(ii)** Distribution is unimodal, bimodal or multimodal
- **(iii)** Declare the shape of distribution normal(central)/left skewed/right skewed by comparing mean and median

Code :-

```
import pandas as pd
def Calculate_Frequency_Distribution(Data):
    Frequency_Distribution = {}
    for i in Data:
        if i in Frequency_Distribution:
            Frequency_Distribution[i] += 1
        else:
            Frequency_Distribution[i] = 1
    Frequency_Distribution = pd.DataFrame(
        list(Frequency_Distribution.items()), columns=["Observation", "Frequency"]
    )
    Frequency_Distribution = Frequency_Distribution.sort_values(by="Frequency",
ascending=False).reset_index(drop=True)
    return Frequency_Distribution
Input = input("Enter space-separated values: ")
Data = [float(i) for i in Input.split()]

Frequency_Distribution = Calculate_Frequency_Distribution(Data)
Mean = sum(Data) / len(Data)
Median = 0
Mode = [
    Frequency_Distribution["Observation"][i]
    for i in range(len(Frequency_Distribution["Observation"]))
    if Frequency_Distribution["Frequency"][i]
    == Frequency_Distribution["Frequency"].iloc[0]
```

```
]

if len(Data) % 2 == 0:
    Median = (Data[len(Data) // 2] + Data[len(Data) // 2 - 1]) / 2
else:
    Median = Data[len(Data) // 2]
if Median < Mean:
    Shape = "Right Skewed"
elif Median > Mean:
    Shape = "Left Skewed"
else:
    Shape = "Normal Distribution"
if len(Mode) == 1:
    Distribution = "Unimodal"
elif len(Mode) == 2:
    Distribution = "Bimodal"
else:
    Distribution = "Multimodal"
print("Frequency Distribution:")
print(Frequency_Distribution.to_string(index=False))
print()
print("Mean:", Mean)
print("Median:", Median)
print("Mode:", Mode)
print("Distribution:", Distribution)
print("Shape:", Shape)
```

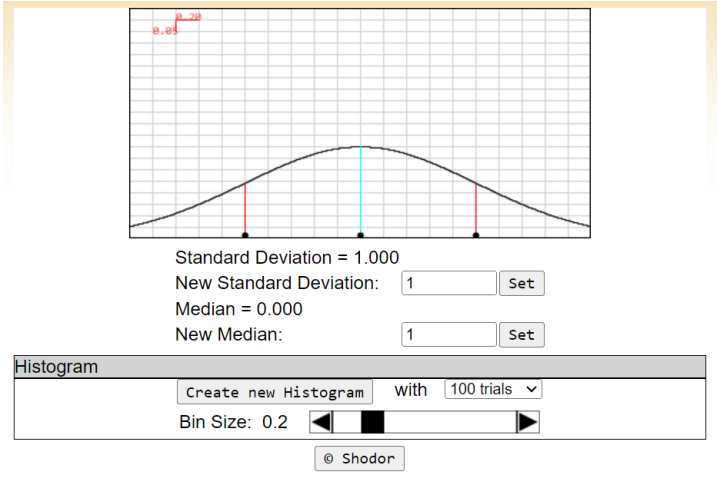
Output :-

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

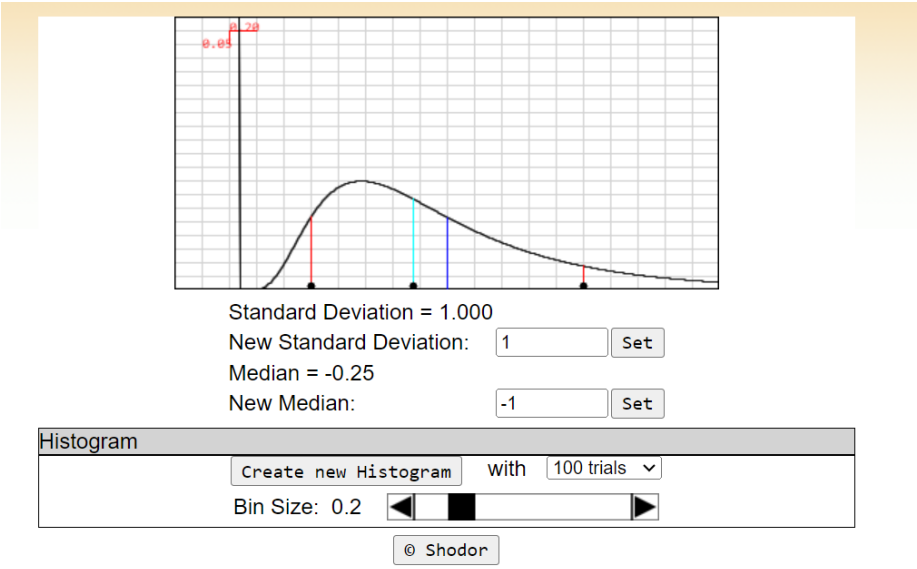
```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_1/Q_4_1.py"
Enter space-separated values: 5 15 16 20 21 25 26 27 30 30 31 32 32 34 35 38 38 41 43 66
Frequency Distribution:
 Observation  Frequency
        30.0          2
        38.0          2
        32.0          2
        31.0          1
        43.0          1
        41.0          1
        35.0          1
        34.0          1
         5.0          1
        15.0          1
        27.0          1
        26.0          1
        25.0          1
        21.0          1
        20.0          1
        16.0          1
        66.0          1

Mean: 30.25
Median: 30.5
Mode: [30.0, 38.0, 32.0]
Distribution: Multimodal
Shape: Left Skewed
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1>
```

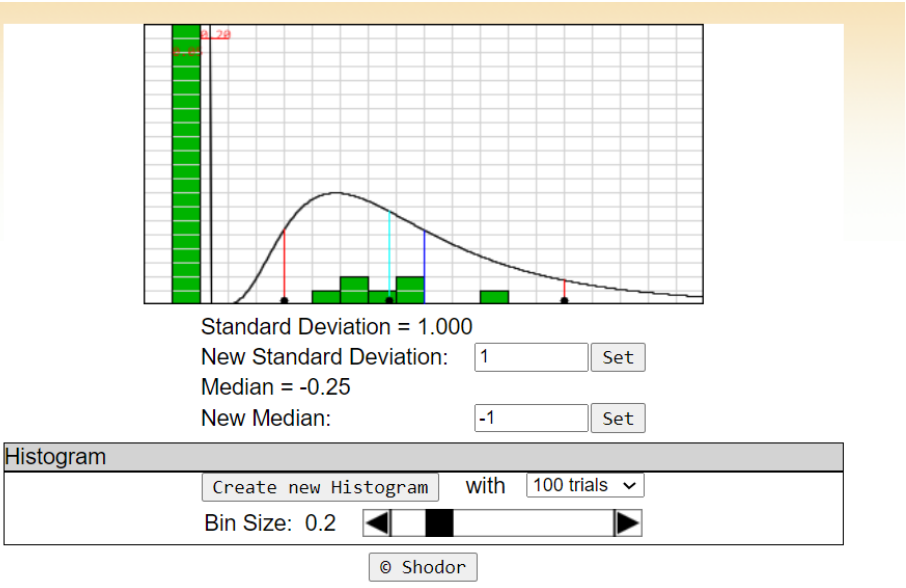
2) Refer the activity applet (http://www.shodor.org/interactivate/activities/SkewDistribution/) Explore the relationship between central tendency, skewness, relation of mean and median for various shapes. Paste at least 8 different scenario with two -three line conclusion on each. Guideline for the activity is mention in instructor and help menu. Also answer the following questions. This applet lets you see how the results of an experiment with a lot of trials might look if the mean, median, and mode aren't the same. Experiment with moving the median and standard deviation lines to change the shape of the graph. Try to answer these questions:



(i) Move the median line a small amount away from the center, about half a box. Where is the mode? Did it move farther than the median? The same amount? Not as far?



(ii) Create a 100-trial histogram with the median off-center. Is the mode of the histogram where you expect it to be? The median? Repeat the trial a few times. Also try with a 1000-trial histogram, and experiment with the bin size. Are the median and mode close to where the curve has them?

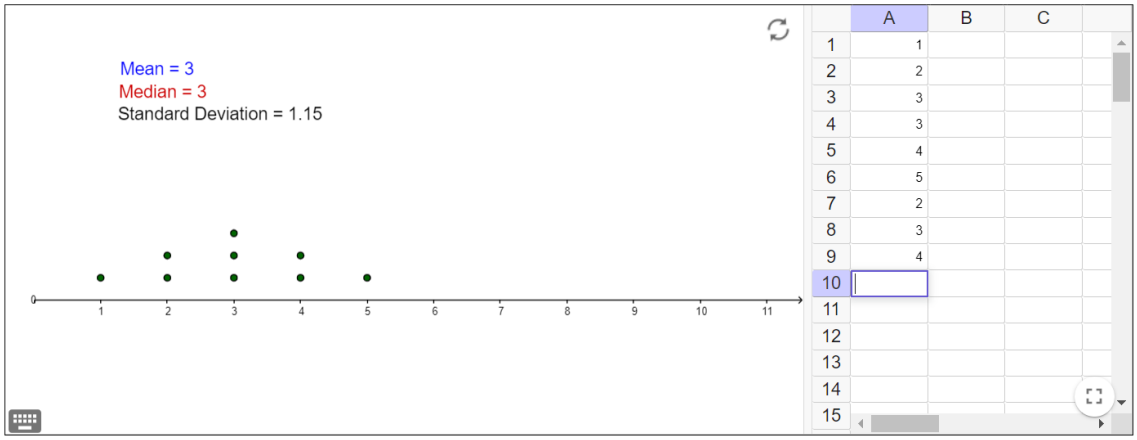


(iii) Do you think it would be easier or harder to recognize this kind of distribution than the normal distribution in an experiment?

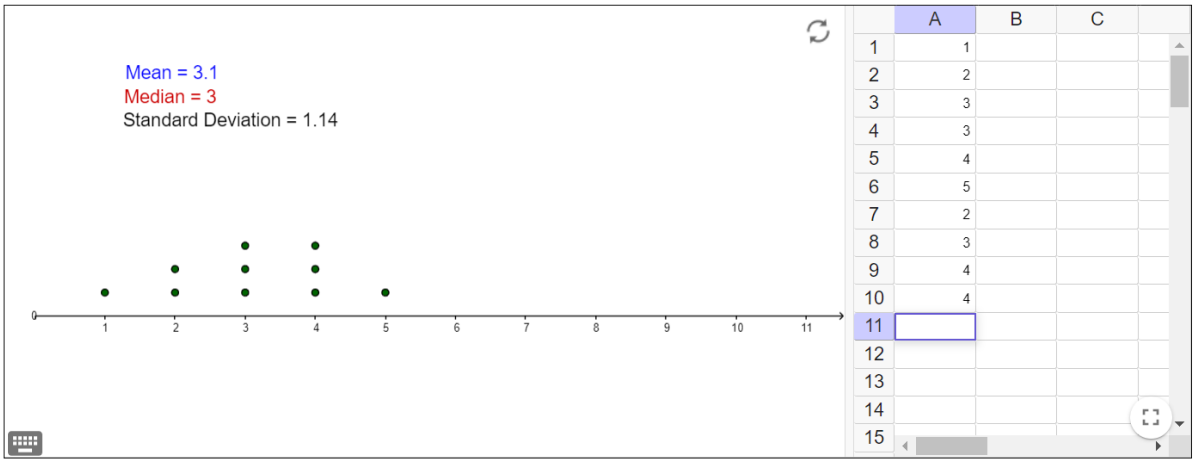
- It would be difficult to recognize the distribution with respect to normal distribution

3) Using the link (https://www.geogebra.org/m/BxqJ4Vag) give three case study by entering suitable data(min 30 data).

(i) Mean = Median

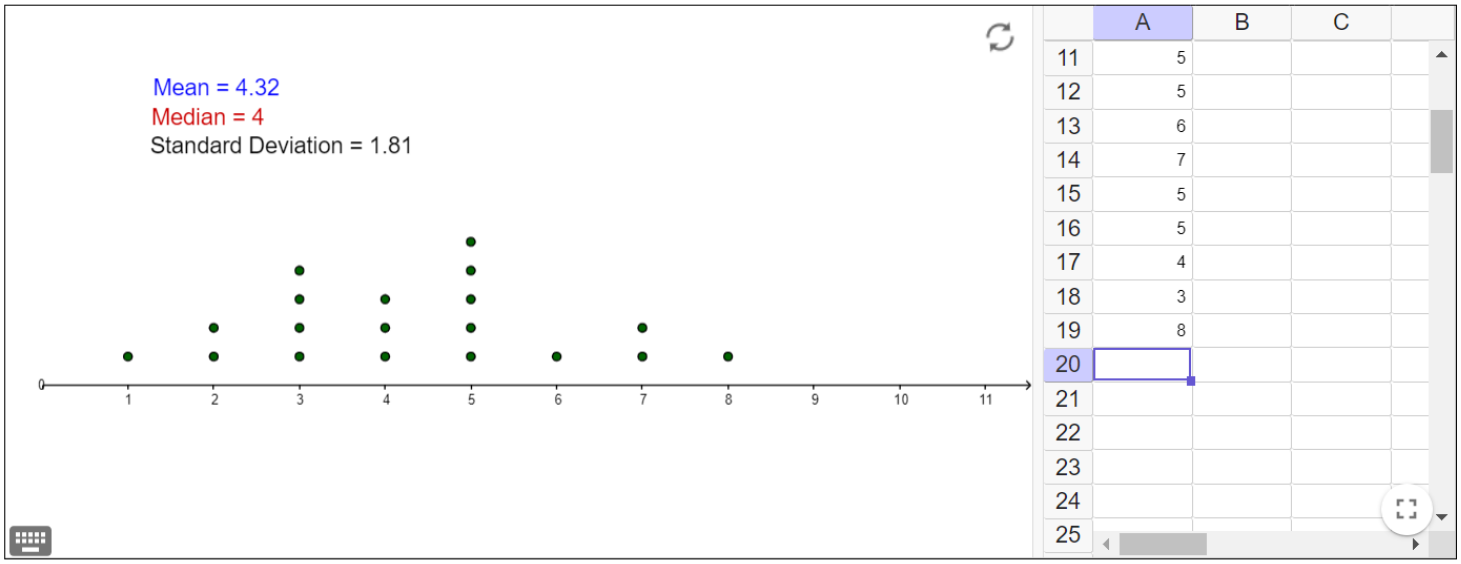


(ii) Mean > Median

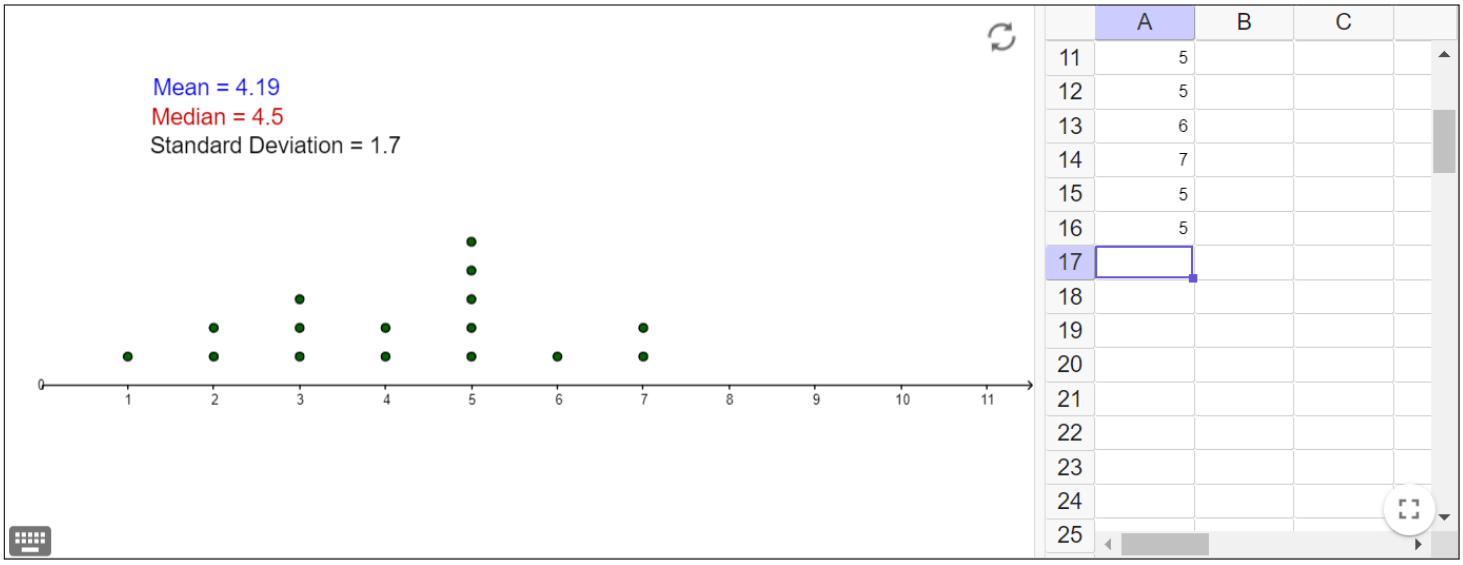


Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

(iii) Mean >> Median

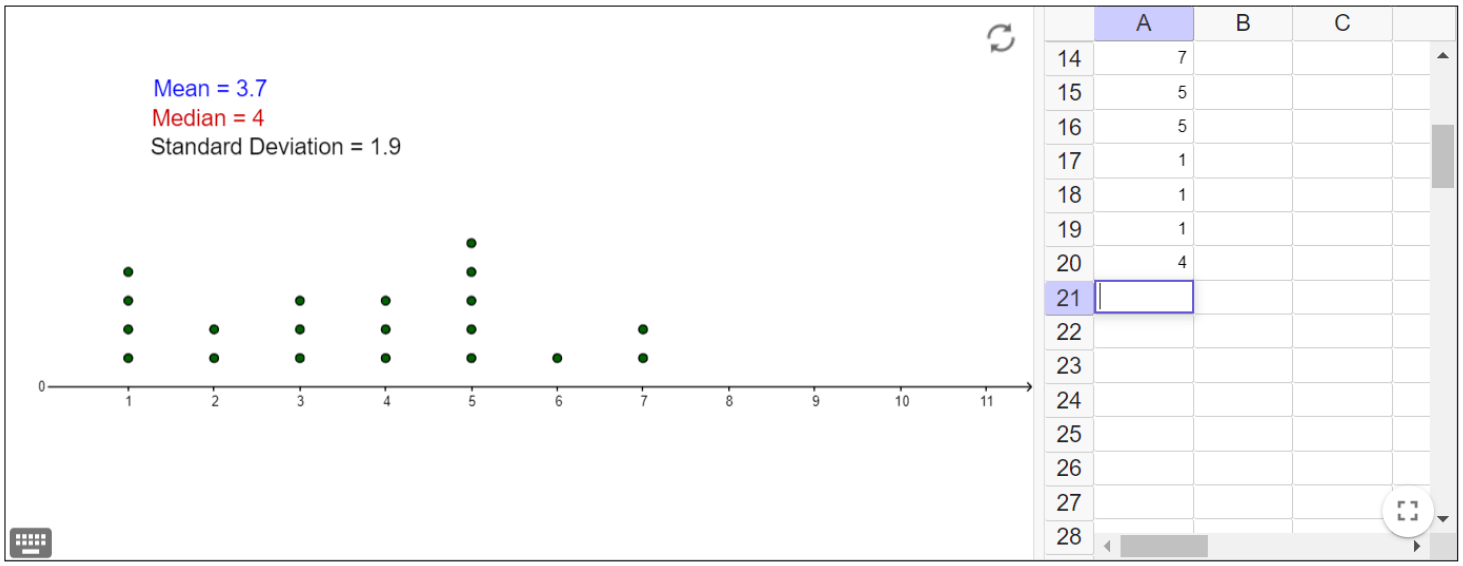


(iv) Mean < Median

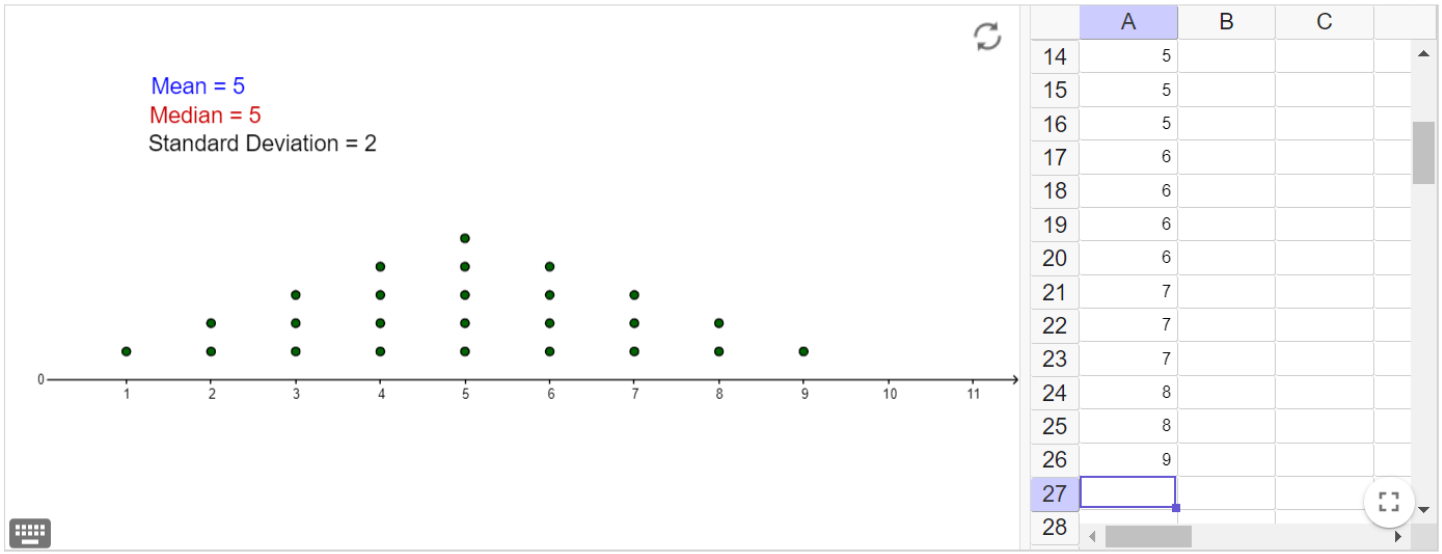


Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

(v) Mean << Median

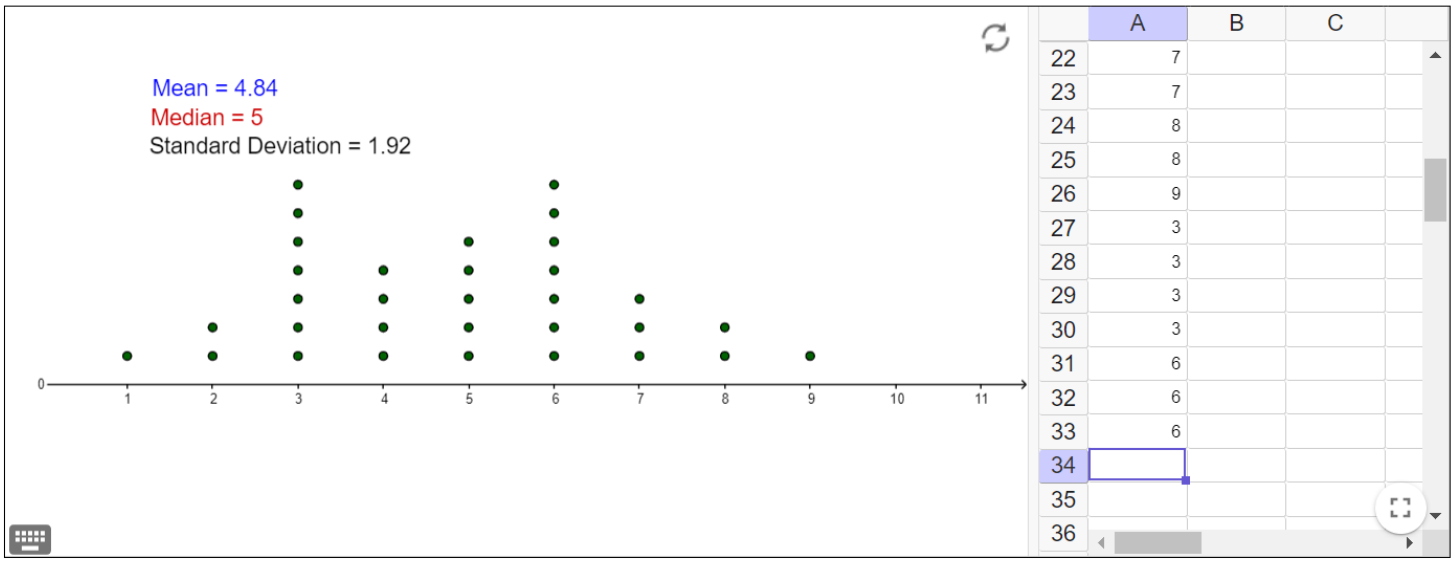


(vi) Unimodal Distribution



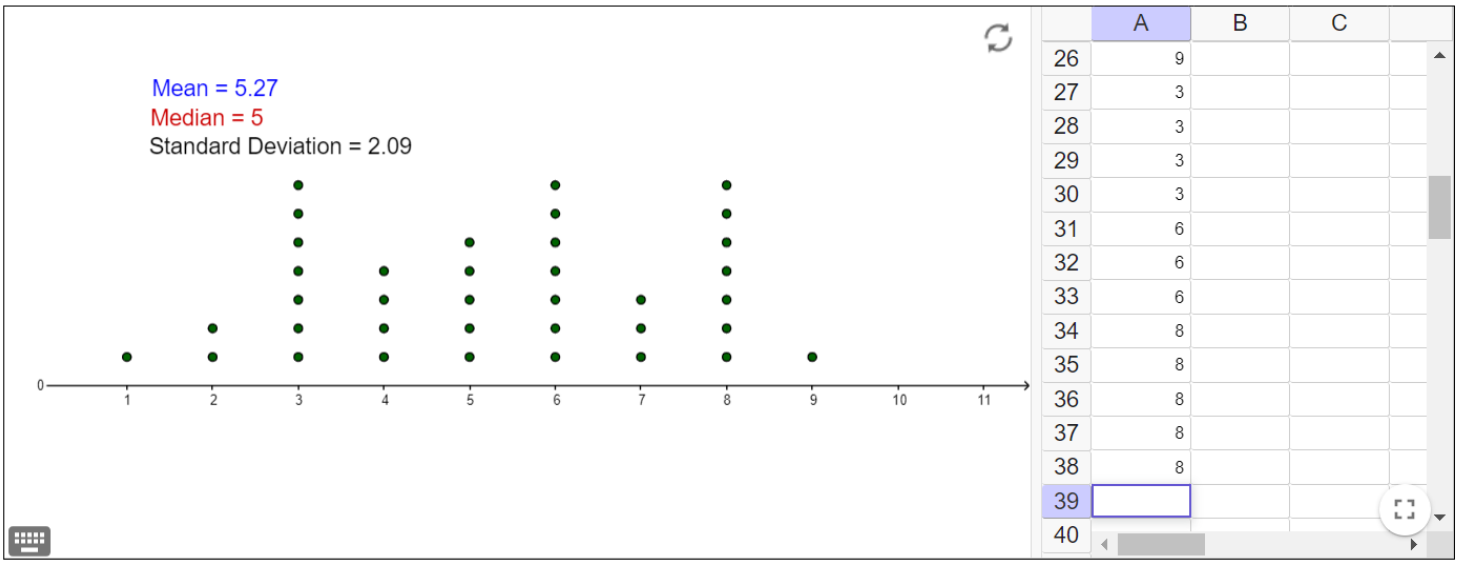
Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

(vii) Bimodal distribution



	A
22	7
23	7
24	8
25	8
26	9
27	3
28	3
29	3
30	3
31	6
32	6
33	6
34	
35	
36	

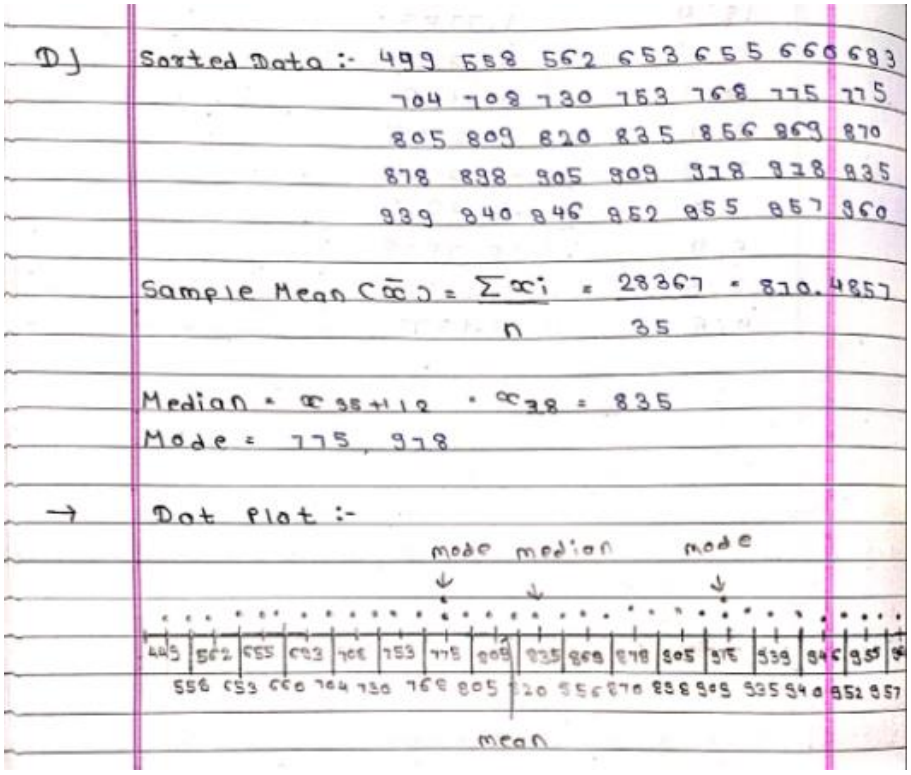
(viii) Multimodal distribution



	A
26	9
27	3
28	3
29	3
30	3
31	6
32	6
33	6
34	8
35	8
36	8
37	8
38	8
39	
40	

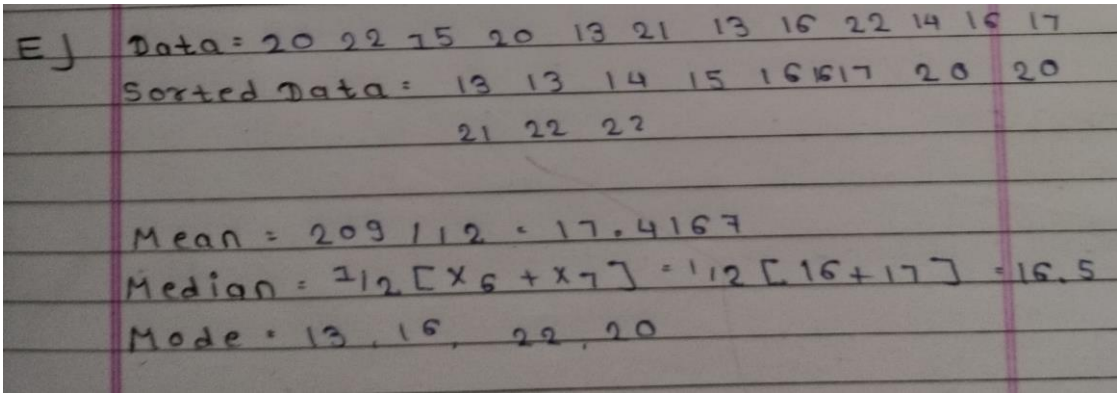
Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

4) The following data are direct solar intensity measurements (watts per meter square). On different days at a location in southern Spain: 562, 869, 708, 775, 775, 704, 809, 856, 655, 805, 878, 909, 918, 558, 768, 870, 918, 940, 946, 660, 820, 898, 935, 952, 957, 693, 835, 905, 939, 955, 960, 499, 653, 730, 753. Calculate sample mean, mode and median. Prepare dot diagram manually and indicate mean, median, mode in it.



5) Write a code to generate 12 different numbers randomly from the range entered by the user. (You can also ask to enter a choice of integer, fractional value with one digit in fraction, or fractional value with two digits in fraction).

(i) Calculate the mean, median, and mode Integer outcome of this code manually on paper



(ii) Calculate the mean median and mode of choice 2/3 of the code.

Code :-

```
import random
import pandas as pd
def Generate_Numbers() :
    Numbers = []
    while True :
        Choice = int(input("Enter Your Choice Your Generating Data :-\n1)Integer\n2)Fraction with
One Deciaml\n3)Fraction with Two Decimal\n"))
        if(Choice in [1,2,3]) :
            break
        else :
            print("Enter A Valid Choice")
    while True:
        Lower_Limit = int(input("Enter The Lower Limit :-"))
        Upper_Limit = int(input("Enter The Upper Limit :- "))
        if(Upper_Limit > Lower_Limit) :
            break;
        else :
            print("Upper Limit Should Be Greater Than Lower Limit :-")
    if(Choice == 1) :
        for i in range(0,12) :
            Numbers.append(random.randint(Lower_Limit,Upper_Limit))
    elif (Choice == 2) :
        for i in range(0,12) :
            Numbers.append(round(random.uniform(Lower_Limit,Upper_Limit),1))
    else :
        for i in range(0, 12):
            Numbers.append(round(random.uniform(Lower_Limit, Upper_Limit), 2))
    return Numbers
def Calculate_Mean(Data) :
    return sum(Data) / len(Data)
def Calculate_Median(Input_Data):
    Data = sorted(Input_Data)
    if len(Data) % 2 == 0:
        return (Data[len(Data) // 2] + Data[len(Data) // 2 - 1]) / 2
    else:
        return Data[(len(Data) // 2)]
def Calculate_Mode(Data) :
```

```
    Frequency_Distribution = {}
    for i in Data:
        if i in Frequency_Distribution:
            Frequency_Distribution[i] += 1
        else:
            Frequency_Distribution[i] = 1
    Frequency_Distribution = pd.DataFrame(
        list(Frequency_Distribution.items()), columns=["Observation", "Frequency"]
    )
    Frequency_Distribution = Frequency_Distribution.sort_values(
        by="Frequency", ascending=False
    ).reset_index(drop=True)
    Mode = [
        Frequency_Distribution["Observation"][i]
        for i in range(len(Frequency_Distribution["Observation"]))
        if Frequency_Distribution["Frequency"][i]
        == Frequency_Distribution["Frequency"].iloc[0]
    ]
    return Mode
def Print_Result(Data , Mean ,Median, Mode) :
    print(f"Statistical Value for {Data}")
    print("Mean: ", Mean)
    print("Median:", Median)
    print("Mode: ", Mode)
Data = []
Data = Generate_Numbers()
Mean = Calculate_Mean(Data)
Median = Calculate_Median(Data)
Mode = Calculate_Mode(Data)
Print_Result(Data , Mean , Median , Mode)
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1> & "C:/Program Files/Python312/python.
exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_1/Q_4_5.py"
Enter Your Choice Your Generating Data :-
1)Integer
2)Fraction with One Deciaml
3)Fraction with Two Decimal
4
Enter A Valid Choice
Enter Your Choice Your Generating Data :-
1)Integer
2)Fraction with One Deciaml
3)Fraction with Two Decimal
5
Enter A Valid Choice
Enter Your Choice Your Generating Data :-
1)Integer
2)Fraction with One Deciaml
3)Fraction with Two Decimal
1
Enter The Lower Limit :-12
Enter The Upper Limit :- 23
Statistical Value for [20, 22, 15, 20, 13, 21, 13, 16, 22, 14, 16, 17]
Mean:  17.416666666666668
Median: 16.5
Mode:  [20, 22, 13, 16]
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_1>
```

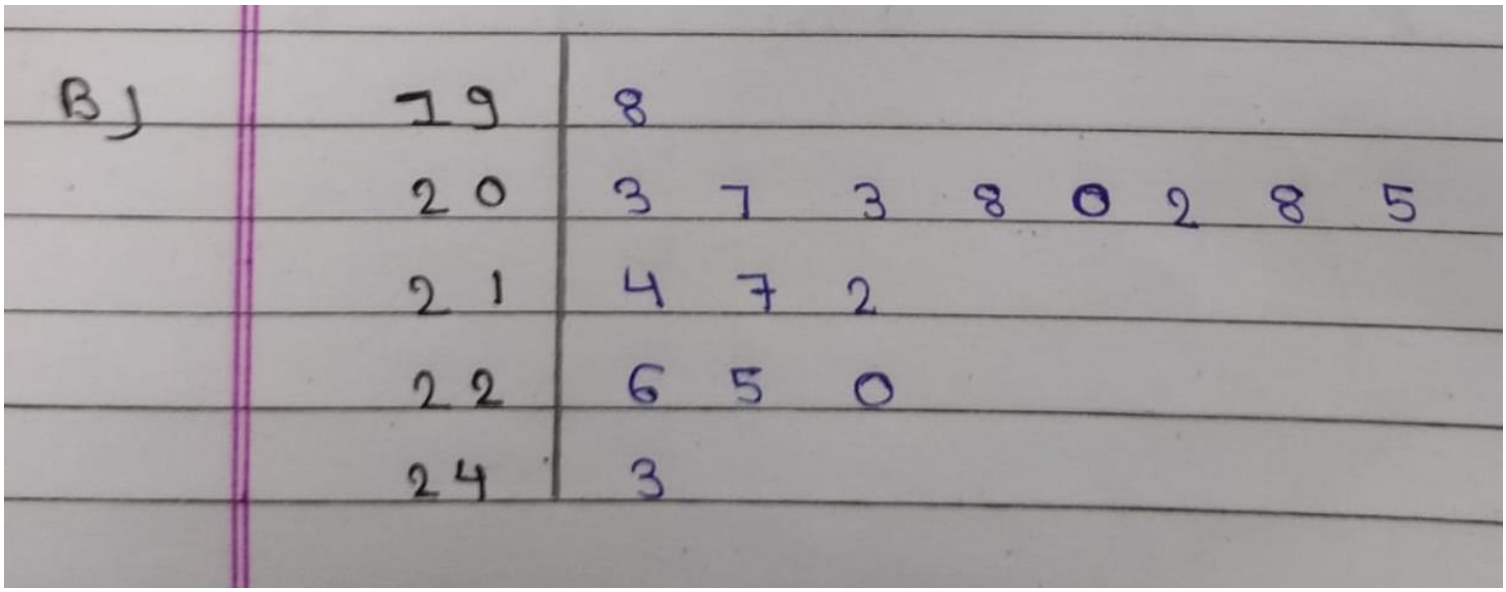
Question – 5:-

1) Refer the following link and enlist all possible types of graphs used in statistical data analysis. Write a brief conclusion about each graph in your own words to represent the advantages/characteristics of a particular graph. Take one case study data for each graph separately and represent it using any online/offline tool. Write inference from the graph below the chart. (You can refer other learning resources as well for the topic) (i) https://www.intellspot.com/types-graphs-charts/ (ii) https://piktochart.com/blog/types-of-graphs/ (iii) https://www.datapine.com/blog/different-types-of-graphs-chartsexamples/ (iv) https://www.studysmarter.co.uk/explanations/math/statistics/statisticalgraphs

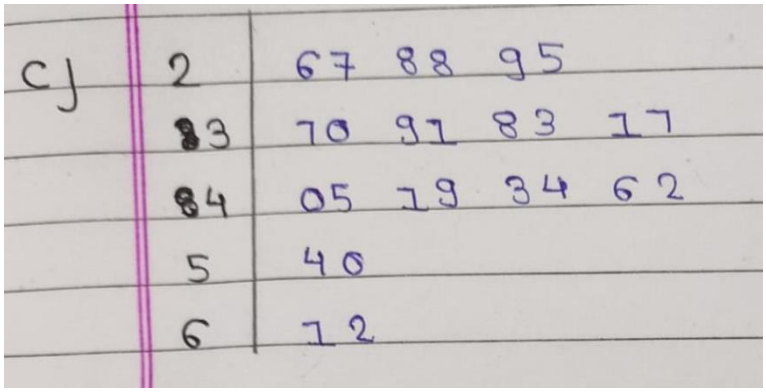
- There are different charts available for visualizing data .like line charts are used to analysis trends and predict future trends ,Bar graph is used when we want to compare the data and it is grouped in specific category. Pie chart is used to analyze the distribution with respect to total data , Histogram used to visualize the distribution, scatter plot used to find the relation between two variables .

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

2) The following are figure on an oil well's daily production in barrels : 214, 203, 226, 198, 243, 225, 207, 203, 208, 200, 217, 202, 208, 212, 205 and 220. Construct a stem-and-leaf display with the stem labels 19, 20,…., and 24.

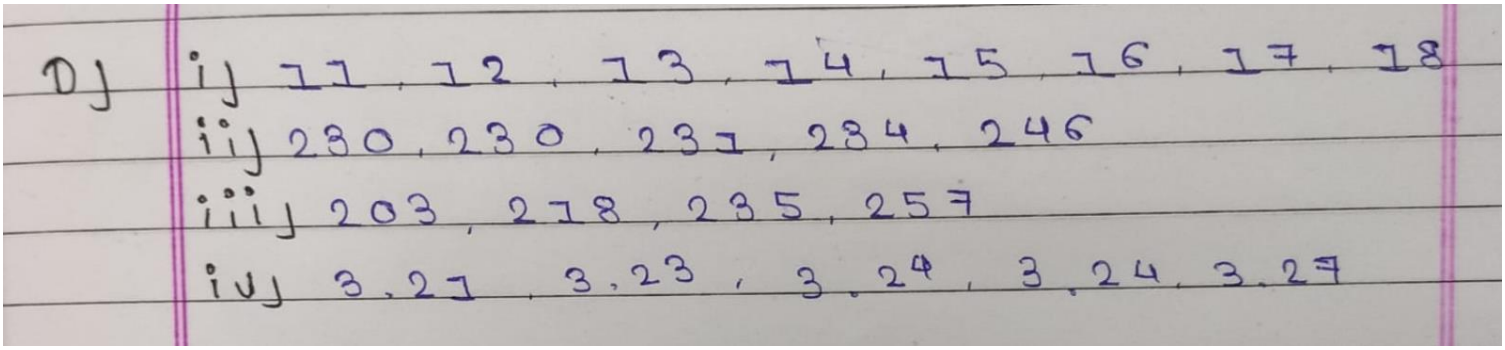


3) The following are determinations of a river's annual maximum flow in cubic meters per second: 405, 355, 419, 267, 370, 391, 612, 383, 434, 462, 288, 317, 540, 295, and 508. Construct a stem-and-leaf display with two-digit leaves.

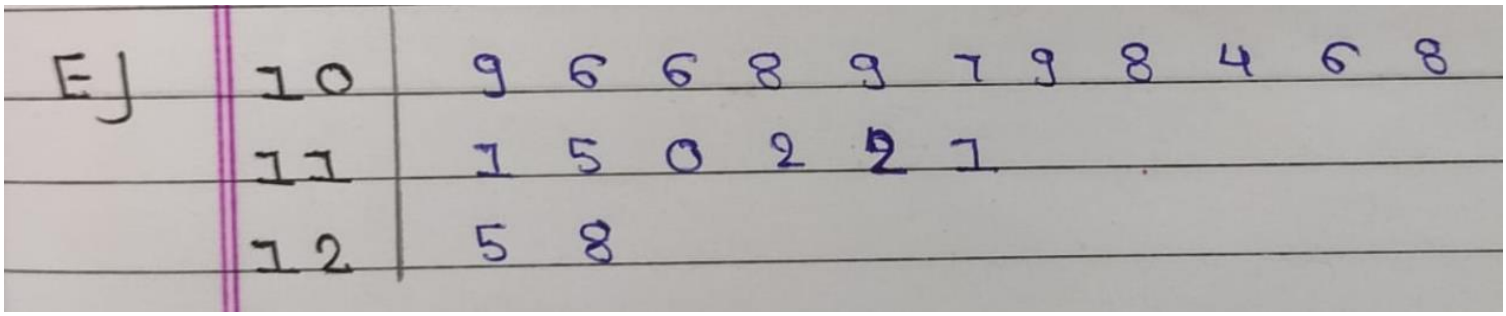


Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

4) List the data that correspond to the following stems of stem-and-leaf displays: (i) 1 | 1 2 3 4 5 7 8 . Leaf unit = 1.0 (ii) 23 | 0 0 1 4 6. Leaf unit = 1.0 (iii) 2 |03 18 35 57. First leaf digit unit =10.0 (iv) 3.2 | 1 3 4 4 7. Leaf unit = 0.01



5) The following are the IQs of 20 applicants to an undergraduate engineering program: 109, 111, 106, 106, 125, 108, 115, 109, 107, 109, 108, 110, 112, 104, 110, 112, 128, 106, 111, 108. Construct a five-stem display with one digit leaves.



6) Use the link (https://www.calculatorsoup.com/calculators/statistics/stemleaf.php) Enter five different use case (which represent various types of stem-leaf) data and take screenshot of plot and write inference from each of it.

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

Stem	Leaf
0	3 6
1	
2	2 5 6
3	3 5 6 8
4	2 5 5 6 7 7 7 8 8 9
5	2 3 4 4 6 6 7 8 8
6	4 5 5 7 8 8 9
7	4 5
8	7
9	9

Stem	Leaf
0	3 6
1	2 3 4 5
2	2 5 6
3	3 5 6 8
4	2 5 5 6 7 7 7 8 8 9
5	2 3 4 4 6 6 7 8 8
6	4 5 5 7

Stem	Leaf
0	3 6
1	2 2 3 3 4 4 5 5 6 7
2	2 5 6
3	3 5 6 8
4	2 5 5 8 8 9
5	6 6 7 8 8
6	4 5 5 7

Stem	Leaf
0	3 6
1	2 2 6 7
2	2 5 6
3	3 5 6 8
4	1 2 2 3 3 4 5 5 5 6 8 8 9
5	6 6 7 8 8
6	4 5 5 7

Stem	Leaf
0	3 3 3 5 6 6
1	2 2 6 7
2	2 5 6
3	2 3 4 5 6 7 8 8
4	1 2 2 3 3 4 5 5 5 6 8 8 9
5	6 6 7 8 8
6	4 5 5 7

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030

Question – 6:-

1) Exam Scores. Consider the following sample of exam scores, arranged in increasing order. 28 57 58 64 69 74 79 80 83 85 85 87 87 89 89 90 92 93 94 94 95 96 96 97 97 97 97 98 100 100 The sample mean and sample standard deviation of these exam scores are 85 and 16.1, respectively. solve the following problems. a. Compare the percentage of the observations that actually lie within two standard deviations to either side of the mean with that given by Chebyshev's rule with k = 2. b. Repeat part (a) with k = 3.

2) Alcohol Consumption. In the Global Status Report on Alcohol and Health 2014, the World Health Organization reports that in 2010, the worldwide consumption was 6.2 liters of pure alcohol per person (15 years or older), which amounts to 13.5 grams of pure alcohol a day. The countrywise total, recorded and unrecorded alcohol per capita consumption (APC), in 2010 (in liters of pure alcohol) is given in the report. The following data is the APC for the African region (45 Countries). 0.1 0.2 0.6 0.7 1 1.1 2.1 2.3 3.4 3.6 3.6 3.9 4 4.3 4.4 4.8 5.6 5.7 6 6.8 7.5 8.4 8.7 10.8 11 10.9 10.1 9.8 9.8 9.3 8.4 7.7 7.1 6.9 6.6 6.5 4.7 4.2 4 3.8 2.5 2.3 1.8 1.1 0.3 The sample mean and sample standard deviation of the consumption are 5.08 liters and 3.26 liters, respectively. A histogram of the consumption is bell-shaped. Modeling your solutions, solve the following problems.

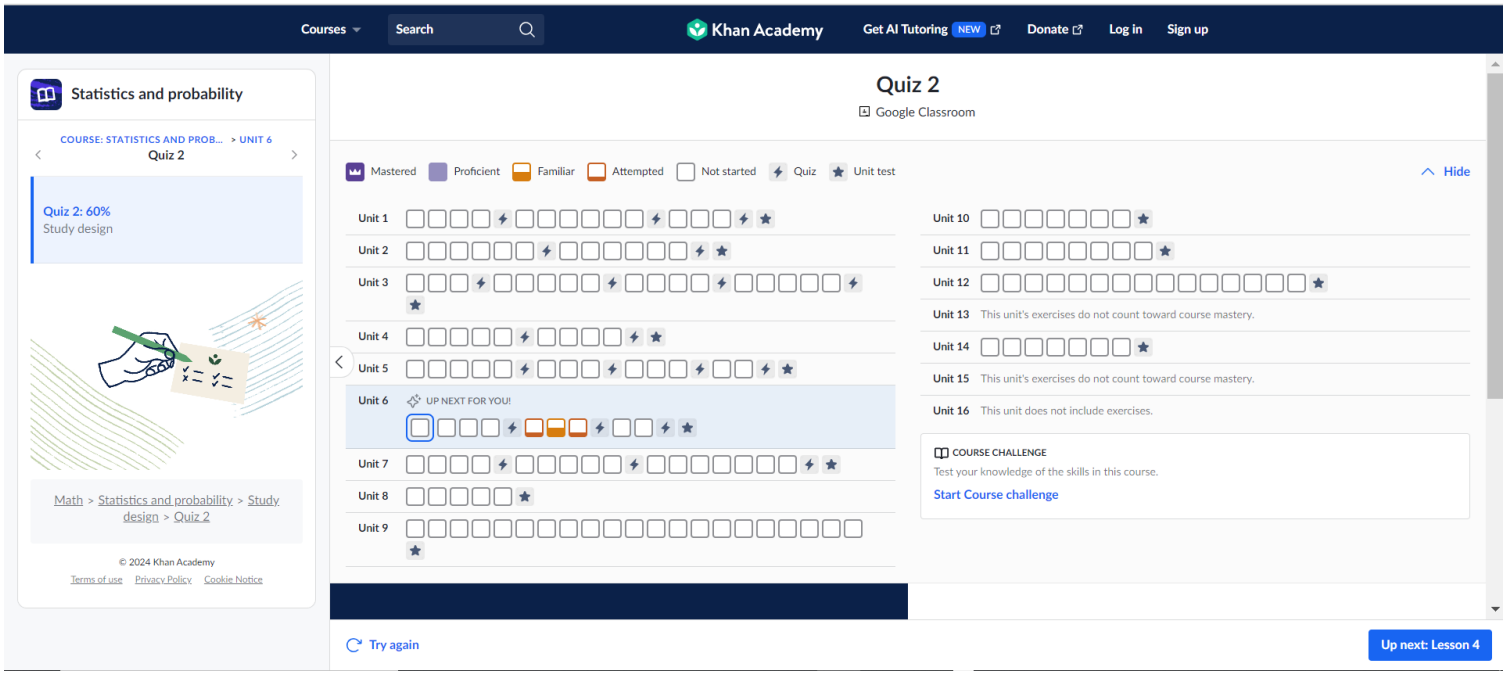
a. Is it reasonable to apply the empirical rule to estimate the percentages of observations that lie within one, two, and three standard deviations to either side of the mean?
b. Use the empirical rule to estimate the percentages of observations that lie within one, two, and three standard deviations to either side of the mean.
c. Use the data to obtain the exact percentages of observations that lie within one, two, and three standard deviations to either side of the mean.
d. Compare your answers in parts (b) and (c).

3) Refer the following link and conclude your learning from it in a maximum five lines. (https://www.linkedin.com/pulse/application-chebyshevs-theorem-swanand-marathe)

- The First thing to use Chebyshev's Theorem can be applied over any type of distribution where the Empirical rule can be applied to normal distribution only. The Percentage that we got in Empirical rule will approximate on the other hand Chebyshev's rule gives assurity that $1 - 1 / k^2$ percen data will be in the range of mean +- k(SD).

Question – 7:-

1) Refer the Lesson 3 of module 6 of the course Statistics and Probability on khan academy. https://www.khanacademy.org/math/statistics-probability/designingstudies/sampling-methods-stats/v/probability-decisions

Output :-

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 1 (Statistics)**	
Tutorial - 1	**Date:-** 02-03-2024	**Enrollment No:-** 92200133030