Major Statistics

Objective :-

1. According to a study, the daily average time spent by a user on a social media website is 50 minutes. To test the claim of this study, Ramesh, a researcher, takes a sample of 25 website users and finds out that the mean time spent by the sample users is 60 minutes and the sample standard deviation is 30 minutes.

    Based on this information, the null and the alternative hypotheses will be:

    Ho = The average time spent by the users is 50 minutes
    H1 = The average time spent by the users is not 50 minutes

    Use a 5% significance level to test this hypothesis

    Ans:- P = 50 , p = 60 and n = 25

2. Height of 7 students (in cm) is given below. What is the median?
    168 170 169 160 162 164 162.

    Ans:- Arranging the values in ascending order
           160, 162, 162, 164, 168, 169, 170
           Median = 164 (Middle Value)

3. Below are the observations of the marks of a student. Find the value of mode.
    84 85 89 92 93 89 87 89 92

    Ans:- Arranging the values in ascending order
           84 85 87 89 89 89 92 92 93
           Mode = 89 (The values that appears most frequently)

4. From the table given below, what is the mean of marks obtained by 20 students?

Marks Xi	No. of studentsfi
3	1
4	2
5	2
6	4
7	5
8	3
9	2
10	1
Total	20

Ans:-

Marks Xi	3	4	5	6	7	8	9	10	Total	
No of Std fi	1	2	2	4	5	3	2	1	20	
fixi	3	8	10	24	35	24	18	10	132	

Mean = $\Sigma fixi / N$ = 132/20

= 6.6

5. For a certain type of computer, the length of time between charges of the battery is normally distributed with a mean of 50 hours and a standard deviation of 15 hours. John owns one of these computers and wants to know the probability that the length of time will be between 50 and 70 hours.

Ans:-

Mean (μ) = 50

Standard deviation (σ) = 15

If x=50 then Z= $(x - \mu)/\sigma$ = 50-50/15

Z = 0

If x=70 then Z= $(x - \mu)/\sigma$ = 70-50/15

Z = 1.33

In standard form the values of z lies b/w 0< Z< 1.333

By z table the area is 0.4082

10. Download the Cell Phone Survey Dataset and perform the below mentioned operations on the dataset :-

●Find Mean of Signal strength column using Pandas and Statistics library.

```
In [1]: %pip install scipy

Requirement already satisfied: scipy in d:\users\lenovo\anaconda3\lib\site-pa
ckages (1.7.3)
Requirement already satisfied: numpy<1.23.0,>=1.16.5 in d:\users\lenovo\anaco
nda3\lib\site-packages (from scipy) (1.21.5)
Note: you may need to restart the kernel to use updated packages.
```

```
In [2]: import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

%matplotlib inline
import statistics as stats

from scipy.stats import norm

import random
```

```
In [5]: df = pd.read_csv('Cell Phone Survey.csv')

print(type(df))

df.head()

<class 'pandas.core.frame.DataFrame'>
```

Out[5]:

	Gender	Carrier	Type	Usage	Signal strength	Value for the Dollar	Customer Service
0	M	AT&T	Smart	High	5	4	4
1	M	AT&T	Smart	High	5	4	2
2	M	AT&T	Smart	Average	4	4	4
3	M	AT&T	Smart	Very high	2	3	3
4	M	AT&T	Smart	Very high	5	5	2

- Find the Median of Customer Service column using Pandas and Statistics library.

- Find Mode of Signal strength column using Pandas and Statistics library.

- Find Standard deviation of Customer Service column using Pandas and Statistics library.

```
In [6]: df_mean = df['Signal strength'].mean()
        print("Mean of Signal strength column is" ,df_mean)

        Mean of Signal strength column is 3.3076923076923075
```

```
In [7]: df_Median = df['Customer Service'].median()
        print("Median of Customer Service column is" ,df_Median)

        Median of Customer Service column is 3.0
```

```
In [8]: df_Mode = df['Signal strength'].mode()
        print("Mode of Signal strength column is" ,df_Mode)

        Mode of Signal strength column is 0    3
        Name: Signal strength, dtype: int64
```

```
In [9]: df_stdev = df['Customer Service'].std()
        print("Standard deviation of Customer Service column is" ,df_stdev)

        Standard deviation of Customer Service column is 0.9623375261979594
```

- Find Variance of Customer Service column using Pandas and Statistics library
- Calculate Percentiles of Value for the Dollar column using Numpy.
- Calculate Range of Value for the Dollar column using Pandas.
- Calculate IQR of Value for the Dollar column using Pandas.

```
[36]: df_Variance = df['Customer Service'].var()
      print("Variance of Customer Service column is" ,df_Variance)

      Variance of Customer Service column is 0.9260935143288083
```

```
[39]: df_col = df['Value for the Dollar']
      df_percentile_25 = np.percentile(df_col, 25)
      df_percentile_75 = np.percentile(df_col, 75)
      print("df_percentile_25:",df_percentile_25)
      print("df_percentile_75:",df_percentile_75)

      df_percentile_25: 3.0
      df_percentile_75: 4.0
```

```
[40]: df_range = df['Value for the Dollar'].max() - df['Value for the Dollar'].min()

      print("Range of Value for the Dollar column is" ,df_range)

      Range of Value for the Dollar column is 4
```

```
[41]: df_IQR = df_percentile_75 - df_percentile_25
      print("InterQuartile Range of Value for the Dollar column is" ,df_IQR)

      InterQuartile Range of Value for the Dollar column is 1.0
```

```
[ ]:
```