Question 1: What is the difference between descriptive statistics and inferential statistics? Explain with examples.
Answer : statics parts
1. descriptive - describe the whole data (it consisting and summarising all the data)
   example : virat kholi of run avg (will check all kind of match)
   avg weight of class and height
   delay of each flight
   if we need exact data then descriptive data.

   a. central tendency- mean, median, mode

   b. dispersion - standard dev , variance

   c. symmetricity - skewness , kurtosis

2. inferential - inference you can't take data of trees in a jungle due to lrage amount of numbers,
   it consist of using data that has been measured to form conclusion
   about a population i.e sample data used to fetch data

USE FOR LARGE AMOUNT OF DATA/LOW TIME AND RESCOURSE
then we use sample

   a. probability
   b. CTL

Question 2: What is sampling in statistics? Explain the differences between random and stratified sampling.
Answer: SMAPLING
   1 simple random sample
   2 Stratified sampling
   3 Cluster sampling
   4 Systematic sampling

simple random sample
   random sample- every member have equal probability of selection

   Disadvantages- sample being not a being a part of sample from a certain group
         up and assam might be different

Stratified sampling

strate=layers/group each group get equal chance using different group such as up group/assam grp

Cluster sampling
divides the population or group and some of the group are selected random to from cluster

Systematic sampling
every nth element is selected
example every 3rd person will get 3 marks less

Question 3: Define mean, median, and mode. Explain why these measures of central tendency are important.
Answer:
Mean is the arithmetic average of a set of numbers, calculated by dividing the sum of all values by the number of values.
Median is the middle value in a dataset arranged in ascending or descending order. If the number of values is even, the median is the average of the two middle values.
Mode is the value that appears most frequently in a dataset.

Mean ,Median , Mode are the quantity which helps in to clean a data base and make sure there will be no data left with null values ,if a data contains null values then these measures of central tendency helps that null value to be filled . suppose if a data containing male and female contains the null value then mode will help to fill that data whereas if we have a height column then we can use mean and median to fill that null values .

There is only one issue using mean to find the avg as if the data contains the outliers that will affect the complete data to deal with this median helps to average out the data.

Question 4: Explain skewness and kurtosis. What does a positive skew imply about the data?

Answer: Skewness : is the measure of the symmetric of how the standard deviation is spread over on the graph
Types of Skewness
1 No Skewness
2 Positive Skewness Distribution
3 Negative Skewness Distribution
4 Mild Skewness

1 No Skewness(mean = mode=median)
Data is spread evenly all over the graph

2 Positive Skewness Distribution

Data is spread mostly on the right side of the graph

3 Negative Skewness Distribution
Data is mostly spread over left side of the graph

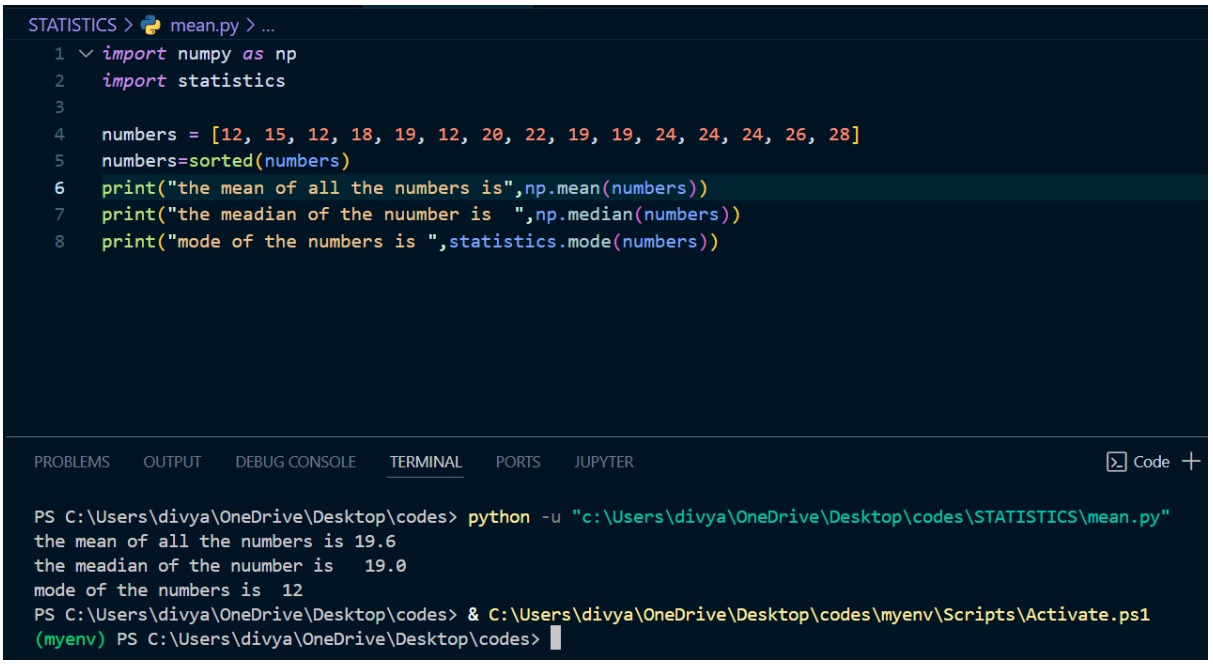
4 Mild Skewness
When skewness lies between -1 to +1

5 Highly Skewness
When skewness lies over (-1 to +1) this range

kurtosis
It helps to measure the peak of the data and helps to understand the shape of frequency distribution.
Types of kurtosis
1 Mesokurtic (kurtosis=3)
2 Platykurtic(kurtosis<3)
3 Leptokurtic (kurtosis>3)

Question 5: Implement a Python program to compute the mean, median, and mode of a given list of numbers. numbers = [12, 15, 12, 18, 19, 12, 20, 22, 19, 19, 24, 24, 24, 26, 28] (Include your Python code and output in the code box below.)

Answer :

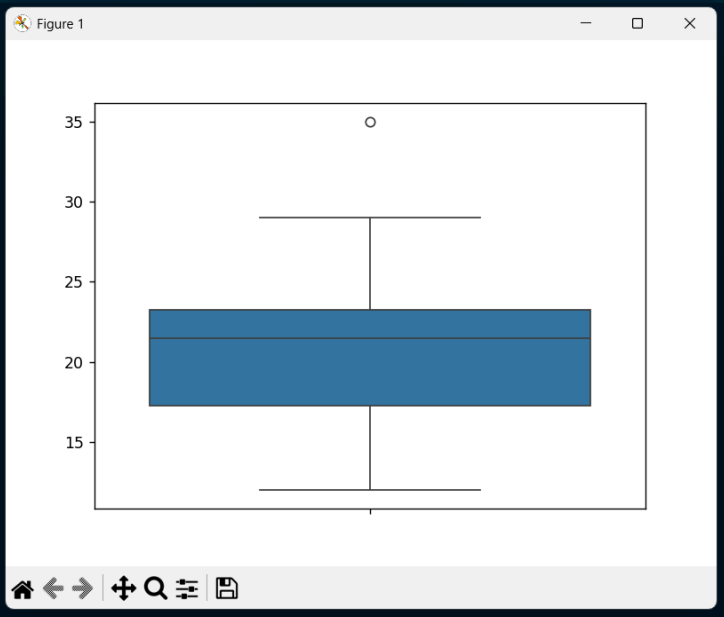


Question 6: Compute the covariance and correlation coefficient between the following two datasets provided as lists in Python: list_x = [10, 20, 30, 40, 50] list_y = [15, 25, 35, 45, 60]

Answer:

Question 7: Write a Python script to draw a boxplot for the following numeric list and identify its outliers. Explain the result: data = [12, 14, 14, 15, 18, 19, 19, 21, 22, 22, 23, 23, 24, 26, 29, 35]

```
ATISTICS > mean.py > ...
1  import seaborn as sns
2  import matplotlib.pyplot as plt
3  data = [12, 14, 14, 15, 18, 19, 19, 21, 22, 22, 23, 23, 24, 26, 29, 35]
4  sns.boxplot(data)
5  plt.show()
```



Question 8: You are working as a data analyst in an e-commerce company. The marketing team wants to know if there is a relationship between advertising spend and daily sales. ● Explain how you would use covariance and correlation to explore this relationship. ● Write Python code to compute the correlation between the two lists: advertising_spend = [200, 250, 300, 400, 500] daily_sales = [2200, 2450, 2750, 3200, 4000]

Answer:

```
STATISTICS > mean.py > ...
1  import numpy as np
2  advertising_spend = [200, 250, 300, 400, 500]
3  daily_sales = [2200, 2450, 2750, 3200, 4000]
4  correlation = np.corrcoef(advertising_spend, daily_sales)[0,1]
5  print(correlation)
6
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS  JUPYTER

PS C:\Users\divya\OneDrive\Desktop\codes> python -u "c:\Users\divya\OneDrive\Desktop\codes\STATISTICS\mean.py"
PS C:\Users\divya\OneDrive\Desktop\codes> & C:\Users\divya\OneDrive\Desktop\codes\myenv\Scripts\Activate.ps1
(myenv) PS C:\Users\divya\OneDrive\Desktop\codes> python -u "c:\Users\divya\OneDrive\Desktop\codes\STATISTICS\mean.py
"
0.9935824101653329
(myenv) PS C:\Users\divya\OneDrive\Desktop\codes>
```

Question 9: Your team has collected customer satisfaction survey data on a scale of 1-10 and wants to understand its distribution before launching a new product. ● Explain which summary statistics and visualizations (e.g. mean, standard deviation, histogram) you'd use. ● Write Python code to create a histogram using Matplotlib for the survey data: survey_scores = [7, 8, 5, 9, 6, 7, 8, 9, 10, 4, 7, 6, 9, 8, 7]

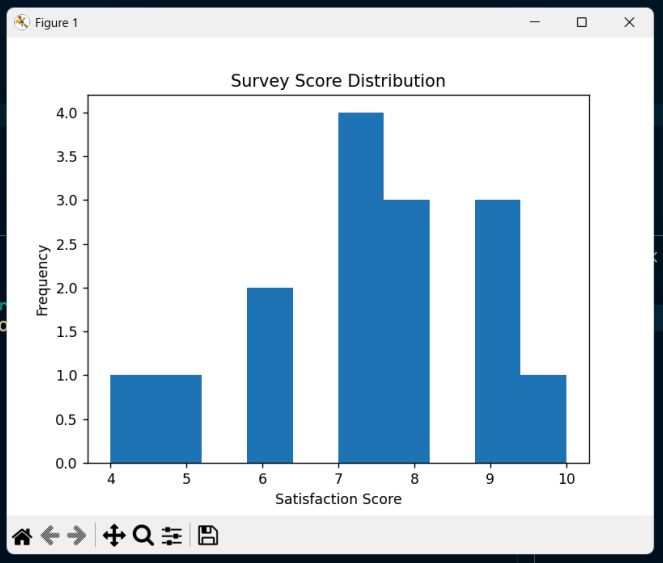
```python
import matplotlib.pyplot as plt

survey_scores = [7, 8, 5, 9, 6, 7, 8, 9, 10, 4, 7, 6, 9, 8, 7]
plt.hist(survey_scores)
plt.title('Survey Score Distribution')
plt.xlabel('Satisfaction Score')
plt.ylabel('Frequency')

plt.show()
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS  JUPYTER

PS C:\Users\divya\OneDrive\Desktop\codes> python -u "c:\User
PS C:\Users\divya\OneDrive\Desktop\codes> & C:\Users\divya\O
(myenv) PS C:\Users\divya\OneDrive\Desktop\codes> python -u
"
0.9935824101653329
(myenv) PS C:\Users\divya\OneDrive\Desktop\codes> python -u
"
(myenv) PS C:\Users\divya\OneDrive\Desktop\codes> python -u
"
```



From my POV mean will the best static to visualizations data .