Statistics Basics| Assignment

Question 1: What is the difference between descriptive statistics and inferential statistics? Explain with examples.

Answer: Descriptive statistics summarize data you already have. For example, calculating the average age of survey participants.

Inferential statistics help you make predictions about a larger group using a sample. Like estimating national preferences from a small survey.

Question 2: What is sampling in statistics? Explain the differences between random and stratified sampling.

Answer: Sampling means picking a subset from a population to study.

Random sampling gives everyone an equal chance. **Stratified sampling** splits the population into groups and samples from each—ensuring fair representation.

Question 3: Define mean, median, and mode. Explain why these measures of central tendency are important.

Answer:

- **Mean** is the average.
- **Median** is the middle value.
- **Mode** is the most frequent value.
- They help describe the center of your data and spot patterns.

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

Answer: Skewness shows if data leans left or right. A **positive skew** means most values are low, with a few high outliers.

Kurtosis tells how sharp the peak is. High kurtosis means more extreme values.

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:

```
import statistics as stats

numbers = [12, 15, 12, 18, 19, 12, 20, 22, 19, 19, 24, 24, 24, 26, 28]

print("Mean:", stats.mean(numbers))
print("Median:", stats.median(numbers))
print("Mode:", stats.mode(numbers))

```

```
Mean: 19.6
Median: 19
Mode: 12
```

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] (Include your Python code and output in the code box below.)

Answer:

```python
import numpy as np

list_x = [10, 20, 30, 40, 50]
list_y = [15, 25, 35, 45, 60]

cov = np.cov(list_x, list_y)[0][1]
corr = np.corrcoef(list_x, list_y)[0][1]

print("Covariance:", cov)
print("Correlation:", corr)

```

```
Covariance: 275.0
Correlation: 0.995893206467704
```

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] (Include your Python code and output in the code box below.)

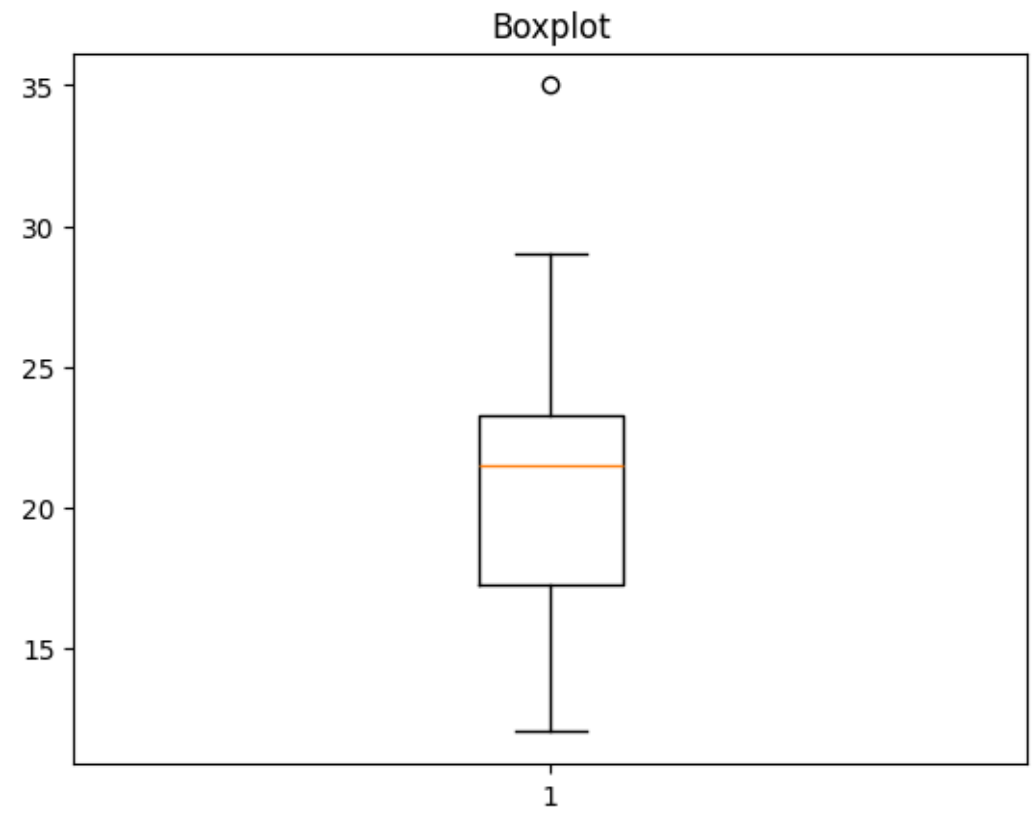
Answer:

```python
import matplotlib.pyplot as plt

data = [12, 14, 14, 15, 18, 19, 19, 21, 22, 22, 23, 23, 24, 26, 29, 35]

plt.boxplot(data)
plt.title("Boxplot")
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] (Include your Python code and output in the code box below.)

Answer:

```
[4]  1 advertising_spend = [200, 250, 300, 400, 500]
     2 daily_sales = [2200, 2450, 2750, 3200, 4000]
     3
     4 import numpy as np
     5 corr = np.corrcoef(advertising_spend, daily_sales)[0][1]
     6 print("Correlation:", corr)
     7
```

```
Correlation: 0.9935824101653329
```

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] (Include your Python code and output in the code box below.)

Answer:

```python
import matplotlib.pyplot as plt
import numpy as np

survey_scores = [7, 8, 5, 9, 6, 7, 8, 9, 10, 4, 7, 6, 9, 8, 7]

mean = np.mean(survey_scores)
std_dev = np.std(survey_scores)

print("Mean:", mean)
print("Standard Deviation:", std_dev)

plt.hist(survey_scores, bins=6, edgecolor='black')
plt.title("Survey Score Distribution")
plt.xlabel("Score")
plt.ylabel("Frequency")
plt.show()

```

```
Mean: 7.333333333333333
Standard Deviation: 1.577621275493231
```