Theoratical - Questions/Answers

Q1. What is statistics, and why is it important?

Ans. Statistics is the science of collecting, organizing, analyzing, and interpreting data. It helps in making decisions based on facts and evidence rather than guesses. Importance:

- Helps in understanding and comparing data.
- Useful in business, reseacrch and economics.
- Helps make predictions and informed decisions.

Q2. What are the two main types of statistics?

Ans.

- Descriptive statistic: Summarizes and describes data (like mean, mode, median, graphs).
- Infreential statistics: Make predictions or conclusions about a large group using sample data.

Q3. What are descriptive statistics?

Ans. Descriptive statistics help to describe the main features of a dataset through numbers, tables, or graphs — e.g., mean, median, mode, charts.

Q4. What is inferential statistics?

Ans. Inferential statistics involve drawing conclusions or predictions about a population from a sample using probability theory (e.g., hypothesis testing, regression).

Q5. What is sampling in statistics?

Ans. Sampling is the process of selecting a small group (sample) from a larger population to study it and make conclusions about the whole group.

Q6. What are the different types of sampling methods?

Ans.

- Random Sampling
- Systematic Sampling
- Stratified Sampling
- Cluster Sampling
- Convenience Sampling

Q7. What is the difference between random and non- Random Sampling?

Ans. Basis Random Sampling Non-Random Sampling Selection Every item has an equal chance Items chosen by personal judgment Bias Less chance of bias More chance of bias Example Lottery system Choosing easily available people

Q8. Define and give examples of qualitative and quantitative data.

Ans.

- Qualitative Data: Describe qualities or characteristics (e.g., color,gender,brand).
- Quantitative Data: Expressed in numbers (e.g., age, height, income).

Q9. What are the different types of data in statistics?

1. Qualitative Data
2. Quantitative Data

- Discrete Data
- Continous Data

Q10. Explain nominal, ordinal, interval, and ratio levels of measurement.

Ans. Nominal: Categories without order (e.g., gender, colors).

Ordinal: Ordered categories (e.g., ranks, grades).

Interval: Numeric data with equal intervals, no true zero (e.g., temperature in °C).

Ratio: Numeric data with true zero (e.g., weight, height, income).

Q11. What is the measure of central tendency?

Ans. It shows the center or average value of a dataset — usually measured using mean, median, or mode.

Q12. Define mean, median, and mode. Ans. Mean: Sum of values ÷ number of values.

Median: Middle value when data is arranged in order.

Mode: The most frequently occurring value.

Q13. What is the significance of the measure of central tendency?

Ans. It helps to summarize large data into a single value that represents the whole dataset, making it easier to compare and analyze.

Q14. What is variance, and how is it calculated?

Ans. Variance shows how much data points differ from the mean. Formula:

$$Variance$$

$\sum (x - \bar{x})^2 n$ Variance= n $\sum(x- \bar{x})^2$

Q15. What is standard deviation, and why is it important?

Ans. Standard deviation is the square root of variance. It measures the spread or dispersion of data. If SD is small → data is close to mean; if large → data is widely spread.

Q16. Define and explain the term range in statistics.

Ans. Range = Highest value − Lowest value It shows the difference between extreme values in a dataset.

Q17. What is the difference between variance and standard deviation?

Ans.

Basis Variance Standard Deviation Definition Average of squared deviations Square root of variance Unit Square of data unit Same as data unit Use For theoretical analysis For practical comparison.

Q18. What is skewness in a dataset?

Ans. Skewness shows whether data is symmetrical or tilted toward one side of the mean.

Q19. What does it mean if a dataset is positively or negatively skewed?

Ans.

- Positively Skewed: Tail on the right side; mean > median.
- Negatively Skewed: Tail on the left side; mean < median.

Positively Skewed: Tail on the right side; mean > median.

Negatively Skewed: Tail on the left side; mean < median.

Q20. Define and explain kurtosis.

Ans. Kurtosis measures the peakedness or flatness of a distribution curve.

- High kurtosis → sharp peak.
- Low kurtosis → flat curve.

Q21. What is the purpose of covariance?

Ans. Covariance shows the direction of relationship between two variables (how they change together).

22. What does correlation measure in statistics?

Ans. Correlation measures the strength and direction of the relationship between two variables.

Q23. What is the difference between covariance and correlation?

Ans. Basis Covariance Correlation Meaning Direction of relationship Direction + strength Range Any number -1 to +1 Unit Depends on variables No unit (standardized)

Q24. What are some real-world applications of statistics?

Ans.

- Business: Market analysis, sales forecasting.
- Health: Medical research, disease studies.
- Government: Census, population studies.
- Sports: Player performance analysis.

Practical Questions/Answers

```
#Ans1. Calculate the mean, median, and mode of a dataset
import numpy as np
from statistics import mode

data = [12, 15, 20, 20, 25, 30, 35]

mean = np.mean(data)
median = np.median(data)
mode_value = mode(data)

print("Mean:", mean)
print("Median:", median)
print("Mode:", mode_value)
```

```
Mean: 22.428571428571427
Median: 20.0
Mode: 20
```

```
##Ans2. Compute the variance and standard deviation of a dataset
import numpy as np

data = [10, 12, 23, 23, 16, 23, 21, 16]

variance = np.var(data)
std_deviation = np.std(data)

print("Variance:", variance)
print("Standard Deviation:", std_deviation)
```

```
Variance: 24.0
Standard Deviation: 4.898979485566356
```

Ans3. Create a dataset and classify into nomial, ordinal, interval, and ratio types

Data type : Nomianl, Ordinal, Interval, Ratio.

Example: Gender-Male/Feamle, Grades- A,B,C, Temperature (degree celcious), weight(kg).

Explantaion: Names or labels, ordered but unequal gaps, Equal gaps, no true zero, Equal gaps, has a true zero

```
##Ans4. Implement random and stratified sampling
import pandas as pd
from sklearn.model_selection import train_test_split

data = pd.DataFrame({
    'Age': [18, 20, 22, 24, 26, 28, 30, 32],
    'Gender': ['M', 'F', 'M', 'F', 'M', 'F', 'M', 'F']
})

# Random Sampling
random_sample = data.sample(n=4)

# Stratified Sampling (based on Gender)
train, test = train_test_split(data, test_size=0.5, stratify=data['Gender'])

print("Random Sample:\n", random_sample)
print("\nStratified Sample:\n", test)
```

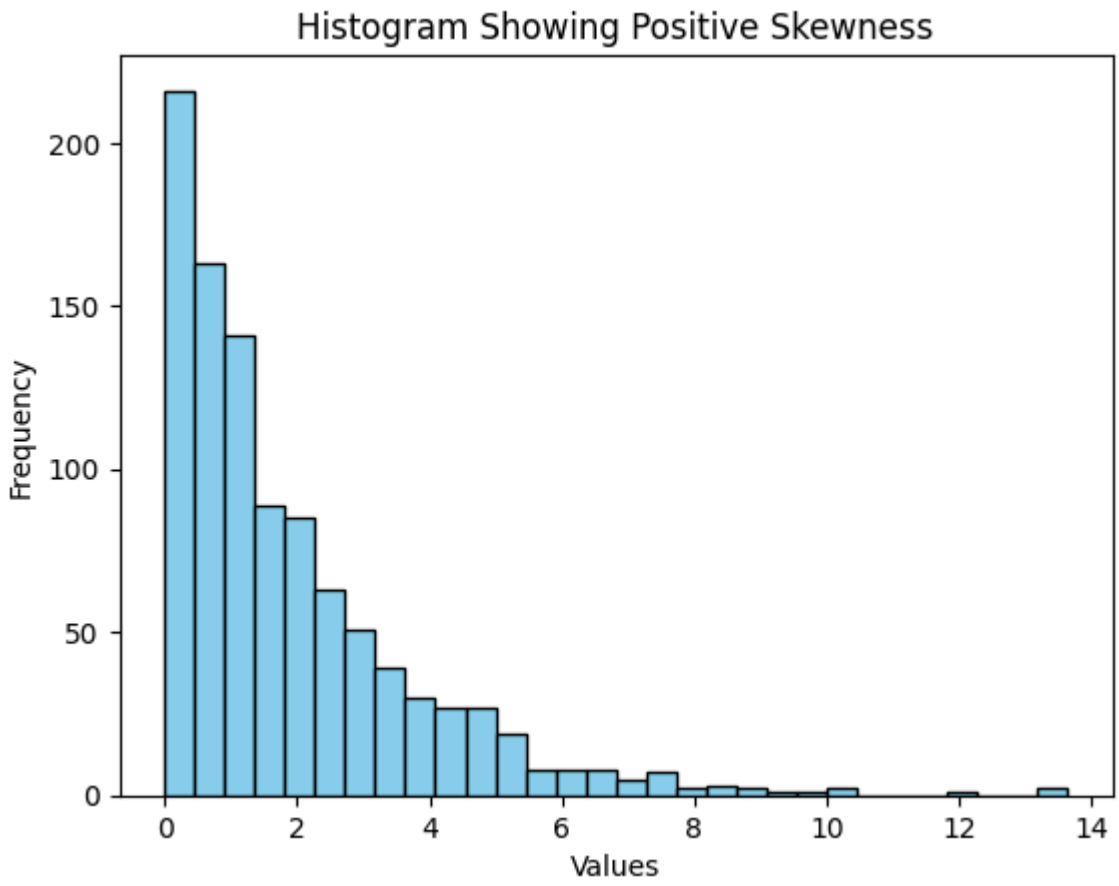
```
Random Sample:
    Age Gender
5   28      F
3   24      F
2   22      M
4   26      M

Stratified Sample:
    Age Gender
4   26      M
2   22      M
3   24      F
7   32      F
```

```
##Ans5. Python function to calculate the range of a dataset
def find_range(data):
    return max(data) - min(data)

data = [5, 10, 15, 20, 25]
print("Range:", find_range(data))
```

```
Range: 20
```

```
#Ans6. Plot a histogram to visualize skewness
import matplotlib.pyplot as plt
import numpy as np

data = np.random.exponential(scale=2, size=1000)
plt.hist(data, bins=30, color='skyblue', edgecolor='black')
plt.title("Histogram Showing Positive Skewness")
plt.xlabel("Values")
```

```
plt.ylabel("Frequency")
plt.show()
```

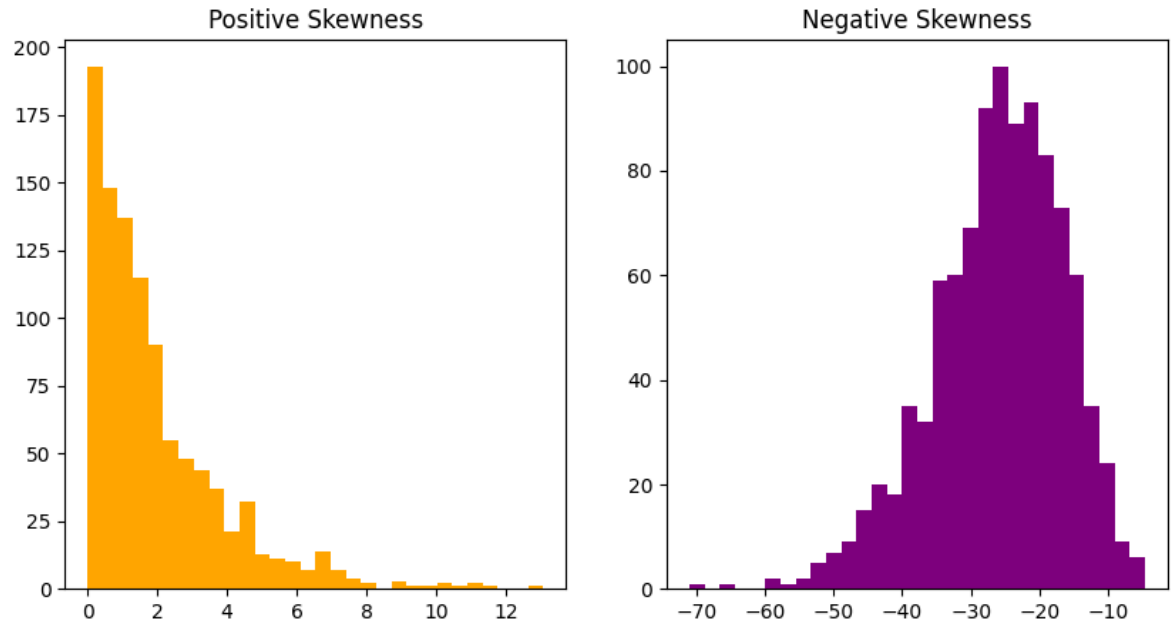



```
#Ans.7 Calculate skewness and kurtosis using Python
from scipy.stats import skew, kurtosis
import numpy as np

data = np.random.normal(0, 1, 1000)

print("Skewness:", skew(data))
print("Kurtosis:", kurtosis(data))
```

```
Skewness: -0.023834517424674517
Kurtosis: 0.09000006302372787
```

```
# Ans8. Demonstrate positive and negative skewness
import numpy as np
import matplotlib.pyplot as plt

# Positive Skew
pos_skew = np.random.exponential(scale=2, size=1000)

# Negative Skew
neg_skew = np.random.normal(5, 1, 1000) ** 2

plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
plt.hist(pos_skew, bins=30, color='orange')
plt.title("Positive Skewness")
```

```
plt.subplot(1,2,2)
plt.hist(-neg_skew, bins=30, color='purple')
plt.title("Negative Skewness")
plt.show()
```



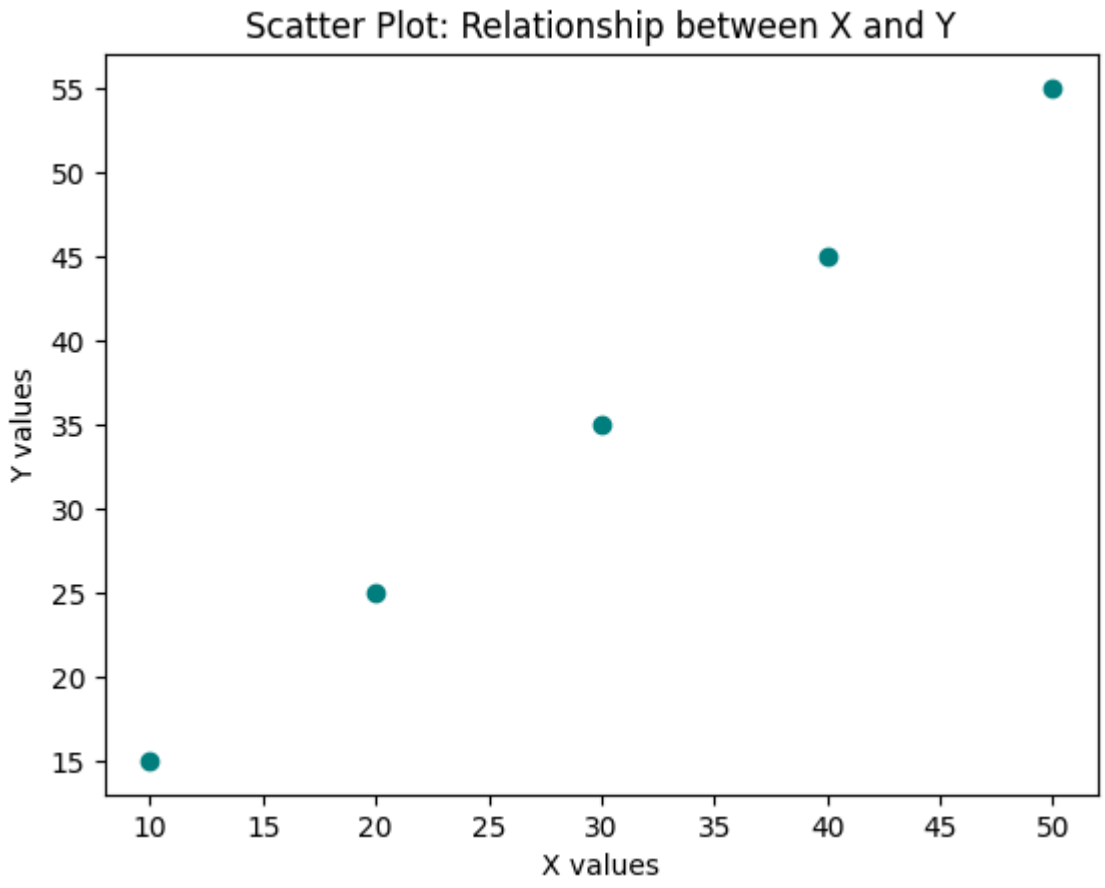

```
#Ans9. Calculate covariance between two datasets
import numpy as np

x = [1, 2, 3, 4, 5]
y = [2, 4, 6, 8, 10]

cov_matrix = np.cov(x, y)
print("Covariance Matrix:\n", cov_matrix)
```

```
Covariance Matrix:
 [[ 2.5  5. ]
 [ 5.  10. ]]
```

```
#Ans10.Calculate correlation coefficient between two datasets
import numpy as np

x = [1, 2, 3, 4, 5]
y = [2, 4, 6, 8, 10]

corr = np.corrcoef(x, y)
print("Correlation Coefficient:\n", corr)
```

```
Correlation Coefficient:
 [[1. 1.]
 [1. 1.]]
```

```
#Ans11.Create a scatter plot to visualize relationship
import matplotlib.pyplot as plt

x = [10, 20, 30, 40, 50]
y = [15, 25, 35, 45, 55]

plt.scatter(x, y, color='teal')
plt.title("Scatter Plot: Relationship between X and Y")
plt.xlabel("X values")
plt.ylabel("Y values")
plt.show()
```




```
#Ans12. Compare simple random and systematic sampling
import pandas as pd

data = pd.DataFrame({'Values': range(1, 21)})

# Simple Random Sampling
random_sample = data.sample(n=5)

# Systematic Sampling (every 4th element)
systematic_sample = data.iloc[::4]
```

```
print("Random Sample:\n", random_sample)
print("\nSystematic Sample:\n", systematic_sample)
```

```
Random Sample:
     Values
17      18
0        1
11      12
10      11
9       10

Systematic Sample:
     Values
0        1
4        5
8        9
12      13
16      17
```

```
#Ans13.Calculate mean, median, and mode of grouped data
import pandas as pd
from statistics import mean, median, mode

data = [5,5,6,6,6,7,7,8,8,9,9,9]
print("Mean:", mean(data))
print("Median:", median(data))
print("Mode:", mode(data))
```

```
Mean: 7.083333333333333
Median: 7.0
Mode: 6
```

```
#Ans14. Simulate data and calculate central tendency & dispersion
import numpy as np

data = np.random.randint(10, 100, 20)

mean = np.mean(data)
median = np.median(data)
std = np.std(data)
var = np.var(data)

print("Data:", data)
print("Mean:", mean)
print("Median:", median)
print("Variance:", var)
print("Standard Deviation:", std)
```

```
Data: [96 68 32 54 95 40 60 24 51 64 71 69 91 43 81 85 13 72 86 29]
Mean: 61.2
Median: 66.0
Variance: 583.8599999999999
Standard Deviation: 24.163195152959386
```

```
Start coding or generate with AI.
```