**Questions Set Statistics**

1. Define data. How is data different from information?
2. What is a line chart, and when would you use it?
3. Explain the difference between a scatter chart and a line chart.
4. What is the purpose of a bar graph? How does it differ from a histogram?
5. Describe when you would use a pie chart and its potential limitations.
6. Define and differentiate between the mean, median, and mode.
7. What are measures of dispersion, and why are they important?
8. Explain the concept of skewness in a dataset.
9. What is a uniform distribution?
10. Describe the binomial distribution and provide an example of its application.
11. What is a standard normal distribution?
12. What is random sampling and how does it differ from stratified sampling?
13. Explain the Central Limit Theorem and its significance.
14. What is hypothesis testing, and how do you interpret p-values?
15. Define Type 1 and Type 2 errors.
16. Describe different data collection techniques and their pros and cons.
17. Explain the process of data cleaning and why it is crucial.
18. Explain the central limit theorem
19. How would you describe a 'p-value
20. What is sampling?
21. What is the normal distribution?
22. Where is inferential statistics used?
23. Difference between population and sample
24. What does standard deviation mean?
25. What is an outlier?
26. What is the confidence interval?
27. Describe hypothesis testing
28. What are descriptive statistics?
29. What are left-skewed and right-skewed distributions
30. What is root cause analysis?
31. How do data scientists use statistics
32. Explain the central limit theorem and give examples of when you can use it in a real-world problem?
33. Describe briefly the hypothesis testing and p-value in layman’s term? And give a practical application for them?
34. What is the meaning of selection bias and how to avoid it

#### **Section 1: Understanding Data and Visualization**

**1.1. Definitions and Concepts**

1. **Question:** Define data. How is data different from information?

**Expected Answer:** Data refers to raw facts and figures that are unprocessed, while information is data that has been processed or interpreted to have meaning.

1. **Question:** What is a line chart, and when would you use it?

**Expected Answer:** A line chart displays data points connected by straight lines. It is used to show trends over time or continuous data.

1. **Question:** Explain the difference between a scatter chart and a line chart.

**Expected Answer:** A scatter chart displays individual data points and is used to observe relationships or correlations between two variables, while a line chart connects data points with lines to show trends.

1. **Question:** What is the purpose of a bar graph? How does it differ from a histogram?

**Expected Answer:** A bar graph is used to compare quantities across different categories, while a histogram shows the distribution of a continuous variable by grouping data into bins.

1. **Question:** Describe when you would use a pie chart and its potential limitations.

**Expected Answer:** A pie chart shows proportions or percentages of a whole. It is best used for categorical data with limited categories, but can be less effective with many categories or small differences.

**1.2. Practical Assignment**

* **Assignment:** Given a dataset containing monthly sales figures for different product categories, create a line chart to visualize sales trends over the year. Additionally, create a bar graph comparing sales across product categories.

#### **Section 2: Characterizing Data with Descriptive Statistics**

**2.1. Central Tendency and Dispersion**

1. **Question:** Define and differentiate between the mean, median, and mode.

**Expected Answer:** Mean is the average value, median is the middle value when data is sorted, and mode is the most frequent value.

1. **Question:** What are measures of dispersion, and why are they important?

**Expected Answer:** Measures of dispersion include range, variance, and standard deviation. They describe the spread or variability of the data, which helps understand the consistency of the data.

**2.2. Probability Basics and Distributions**

1. **Question:** Explain the concept of skewness in a dataset.

**Expected Answer:** Skewness measures the asymmetry of a data distribution. Positive skew indicates a longer tail on the right, while negative skew indicates a longer tail on the left.

1. **Question:** What is a uniform distribution?

**Expected Answer:** A uniform distribution is one where all outcomes are equally likely, and each value within a range has the same probability of occurring.

1. **Question:** Describe the binomial distribution and provide an example of its application.

**Expected Answer:** The binomial distribution describes the number of successes in a fixed number of independent Bernoulli trials. An example is flipping a coin multiple times and counting the number of heads.

1. **Question:** What is a standard normal distribution?

**Expected Answer:** A standard normal distribution is a normal distribution with a mean of 0 and a standard deviation of 1. It is used as a reference for comparing different data distributions.

**2.3. Practical Assignment**

* **Assignment:** Given a dataset with exam scores, calculate the mean, median, mode, variance, and standard deviation. Create a histogram to visualize the distribution and determine the skewness of the data.

#### **Section 3: Predicting from Data with Inferential Statistics**

**3.1. Sampling and Hypothesis Testing**

1. **Question:** What is random sampling and how does it differ from stratified sampling?

**Expected Answer:** Random sampling involves selecting a sample where each member of the population has an equal chance of being chosen. Stratified sampling divides the population into subgroups and samples from each subgroup to ensure representation.

1. **Question:** Explain the Central Limit Theorem and its significance.

**Expected Answer:** The Central Limit Theorem states that the distribution of the sample mean will be approximately normal if the sample size is large enough, regardless of the population distribution. It is significant for making inferences about population parameters.

1. **Question:** What is hypothesis testing, and how do you interpret p-values?

**Expected Answer:** Hypothesis testing is a method to determine if there is enough evidence to reject a null hypothesis. A p-value indicates the probability of obtaining the observed results if the null hypothesis is true; a low p-value suggests strong evidence against the null hypothesis.

1. **Question:** Define Type 1 and Type 2 errors.

**Expected Answer:** Type 1 error occurs when a true null hypothesis is incorrectly rejected (false positive), while Type 2 error occurs when a false null hypothesis is not rejected (false negative).

What is pdf,pmf, cdf ?

How would you interpret a z-score in the context of a normal distribution?

<https://www.simplilearn.com/tutorials/data-analytics-tutorial/data-analyst-interview-questions>

<https://www.datacamp.com/blog/statistics-interview-questions>

<https://intellipaat.com/blog/interview-question/statistics-interview-questions/>