

Statistics Interview Questions

(in simple language by Sahil Josan)

Q1. What are the most important topics in statistics ?

Ans: Statistics can be divided into 2 types

- 1) Inferential Statistics
- 2) Descriptive Statistics

Mean, Median, Mode, Histogram, 5 Number summary, Measure of central tendency, percentiles and Quartiles, normal distribution, Z-score, central limit theorem etc...

Q2. What is EDA ?

Ans: It is an approach to analyze the data using visual techniques. It is used to discover trends, patterns or to check assumptions with the help of statistical summary and graphical representation

Q3. What are Quantitative Data and Qualitative data?

Ans: Quantitative is numerical data

Ex: Height, Weight etc

- 1) Discrete ----> No of students in class = 50
- 2) Continuous ----> Height = 32

Qualitative is categorical data

Ex: Gender

- 1) Nominal ----> Data that doesn't follow any order
- 2) Ordinal ----> Data that follows order

Q4. What is meaning of KPI in statistics ?

Ans: KPI :Key Point Indicator or Key Performance Indicator

- It is a quantifiable measure of performance over time for a specific objective.
- KPI's provide targets for teams to shoot for milestones to gauge progress, and insights that help people across the organization to take better decisions.
- Tools of KPI – **Power BI, Tableau** etc.

Q5. What is the univariate , bivariate, and multivariate analysis ?

Ans:

- **Univariate Analysis:** In this analysis we consider only one variable at a time.
Univariate can be graphically illustrated by:
 1. Bar Charts
 2. Histogram
 3. Pie charts
- **Bivariate Analysis:** In this , we consider two variables at a time.
Bivariate can be graphically illustrated by:
 1. Scatter plot
 2. Box plot
- **Multivariate Analysis:** In this, we consider more than two variables
Multivariate can be graphically illustrated by:
 1. Heat map
 2. Pair plot

Q6. How would you approach the data that's missing more than 30% of it's values ?

Ans: In Statistics if missing values are less than 50 to 90, we have to drop the feature, because the statistical analysis won't be accurate even though we performed.

- If the feature is not so important we can drop that feature
- If the feature is highly important, we can insert the value using different methods like:
 1. Replacing null value's with random numbers
 2. Replacing null value's with mean values
 3. If we have lot of outliers we can replace them with median values

Q7. Give an example whether the median is better or the mean?

Ans:

- **Mean** is the average
 - **Median** is positional average
- If the distribution is skewed one median is better than mean

Example: We can consider the salaries of employees. Some of them are freshers and some will be experienced, so their salary will be different. Now because of the high salaries of experienced people we will not get correct mean, but with the median we can find some nearby values.

Q8. Difference between Descriptive and Inferential Statistics?

Ans:

Descriptive Stats is a term given to the analysis of data that helps to describe, show and summarize data in a meaningful way

Types of Descriptive Statistics:

- Measure of Central Tendency
- Measure of Variability

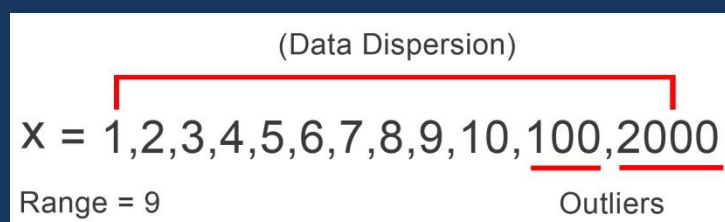
Inferential Stats: In inferential statistics, predictions are made by taking any group of data in which you are interested. It can be defined as a random sample of data taken from a population to describe and make inferences about the population

Q9. What are the methods for dispersion of data in statistics ?

Ans: Range, Interquartile Range, Variance and Standard Deviation

Q10. Is range sensitive to outliers ?

Ans: Yes, range is sensitive to outliers as range is nothing but the difference between the largest and the smallest observation in the data. If there is any outlier either on negative or positive side that will definitely increase the range.



Q11. What are the scenarios where outliers are kept in data ?

Ans:

1. When we want to see the trend changes, because outliers are new start.
2. When we want to release new trend, we need to understand the failure cases like where it is failed and how it is failed to make the product better

Q12. Bessel's Correction ?

Ans: In Statistics, Bessel's correction is the use of $n-1$ instead of n in several formulas, including the sample variance and standard deviation, where n is the number of observations in a sample. This method corrects the bias in the estimation of the population variance. It also partially corrects the bias in the estimation of population standard deviation, thereby providing more accurate results.

Q13. What is the benefit of Boxplot ?

Ans: Box plot gives us the understanding of lower fence, higher fence, Interquartile Range and **Outliers**

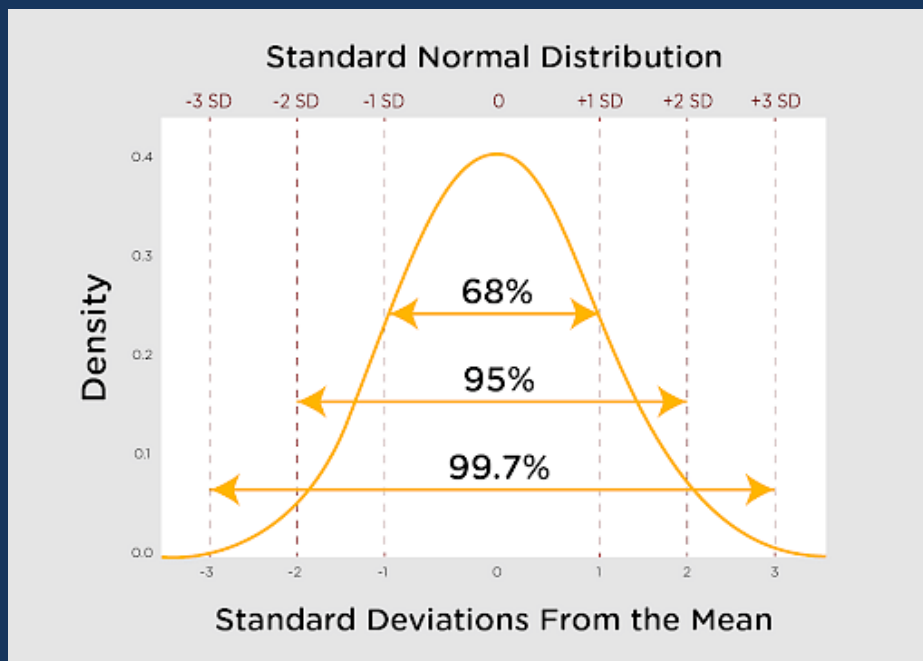
Q14. What is the difference between 1st, 2nd and 3rd Quartile ?

Ans:

- **First quartile:** 25% from smallest to largest of numbers.
- **Second quartile:** Between 25.1% and 50% (till median)
- **Third quartile:** 51% to 75% (above the median)
- **Fourth quartile:** 25% of largest numbers.

Q15. What is the empirical rule ?

Ans: The empirical rule, also known as the three-sigma rule or the 68-95-99.7 rule, is a statistical rule that states that almost all observed data for a normal distribution will fall within three standard deviations (denoted by σ) of the mean or average



Q16. What is Skewness ?

Ans: If the bell curve is shifted to the left or the right, it is said to be skewed. Skewness can be quantified as a representation of the extent to which a given distribution varies from a normal distribution. A normal distribution has a zero skew. A distribution can have right (or positive), left (or negative), or zero skewness

Q17. What is Central Limit Theorem ?

Ans: The central limit theorem states that if you take sufficiently large samples from a population, the samples' means will be normally distributed, even if the population is not normally distributed. Sample size equal to or greater than 30 are often considered sufficient for the central limit theorem to hold.

Q18. Different types of probability distribution ?

Ans:

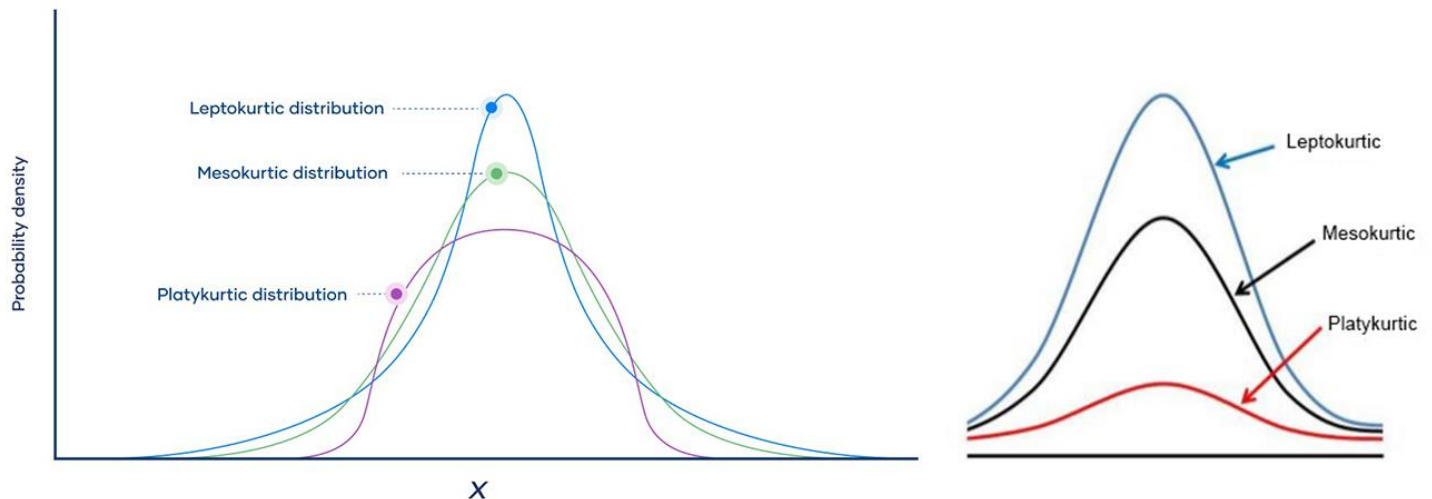
1. Normal Distribution
2. Log-Normal Distribution
3. Power Law Distribution
4. Bernollis Distribution
5. Binomial Distribution

Q19. What is Kurtosis ?

Ans: The sharpness of the peak of a frequency – distribution curve is known as “kurtosis”.

There are 3 types of kurtosis:

- Leptokurtic distribution (Positive Kurtosis)
- Mesokurtic distribution (Zero Kurtosis)
- Platykurtic distribution (Negative Kurtosis)



Q20. What kind of data doesn't have normal distribution ?

Ans: The data which is left (or) right skewed