



Akash Kamerkar [!\[\]\(c8d96c8885d3000a912c2582004aed63\_img.jpg\)](#)

# Must-Known 20 Statistics Questions for Data science Interview Preparation : A Freshers' Guide



Akash Kamerkar



# 1.What is the difference between mean and median?

The mean is the average of a set of values, while the median is the middle value when the data is arranged in ascending or descending order.



Akash Kamerkar



## 2.What is standard deviation?

Standard deviation measures the spread or dispersion of a dataset around the mean.



Akash Kamerkar



### 3. What is a correlation coefficient?

A correlation coefficient measures the strength and direction of the linear relationship between two variables.



Akash Kamerkar



## 4. What is the difference between a population and a sample?

A population is the complete set of individuals or objects of interest, while a sample is a subset of the population used to make inferences about the entire population.



Akash Kamerkar



## 5. What is a p-value?

A p-value is the probability of obtaining a test statistic as extreme as, or more extreme than, the observed result, assuming the null hypothesis is true.



Akash Kamerkar



## 6.What is the Central Limit Theorem?

The Central Limit Theorem states that, for a large sample size, the sampling distribution of the sample mean will be approximately normally distributed regardless of the shape of the population distribution.



Akash Kamerkar



## 7. What is the difference between Type I and Type II errors?

Type I error occurs when we reject the null hypothesis when it is true, while Type II error occurs when we fail to reject the null hypothesis when it is false.



Akash Kamerkar



## 8. What is the difference between a parametric and non-parametric test?

Parametric tests assume certain properties about the population distribution, while non-parametric tests do not make any assumptions about the distribution.



Akash Kamerkar



## **9.What is the difference between a one-tailed and a two-tailed test?**

A one-tailed test is directional and tests for a difference in a specific direction, while a two-tailed test is non-directional and tests for a difference in either direction.



Akash Kamerkar



## 10. What is the difference between covariance and correlation?

Covariance measures the linear relationship between two variables, whereas correlation measures both the strength and direction of the linear relationship.



Akash Kamerkar



## 11. What is the law of large numbers?

The law of large numbers states that as the sample size increases, the sample mean approaches the population mean.



Akash Kamerkar



## 12. What is a sampling distribution?

A sampling distribution is the probability distribution of a sample statistic, such as the mean or standard deviation, based on multiple samples taken from the same population.



Akash Kamerkar



## 13. What is the purpose of hypothesis testing?

Hypothesis testing is used to make inferences about a population based on a sample and to determine whether there is enough evidence to support or reject a claim about the population.



Akash Kamerkar



## 14. What is the difference between a null hypothesis and an alternative hypothesis?

The null hypothesis represents the default assumption or claim, while the alternative hypothesis represents the claim that we are trying to find evidence for.



Akash Kamerkar



## 15. What is a confidence interval?

A confidence interval is an estimate of a population parameter that provides a range of values within which the true population parameter is likely to fall, along with a level of confidence.



Akash Kamerkar



## 16. What is the difference between a t-test and a z-test?

A t-test is used when the population standard deviation is unknown, or the sample size is small, while a z-test is used when the population standard deviation is known, or the sample size is large.



Akash Kamerkar



## 17. What is the difference between probability and odds?

Probability represents the likelihood of an event occurring, while odds represent the ratio of the probability of success to the probability of failure.



Akash Kamerkar



## 18.What is the difference between a random variable and an observation?

A random variable is a variable that can take on different values with certain probabilities, while an observation is a specific value of the random variable.



Akash Kamerkar



## 19. What is the difference between a continuous and a discrete random variable?

A continuous random variable can take on any value within a specified range, while a discrete random variable can only take on specific, isolated values.



Akash Kamerkar



## 20.What is the difference between parametric and non-parametric regression?

Parametric regression assumes a specific functional form of the relationship between the independent and dependent variables, while non-parametric regression makes no assumptions about the functional form.