**INTERVIEW QUESTIONS**

**16. Can you please tell me formula for skewness?**

Skewness = (3 \* (Mean – Median))/Standard Deviation

**17. Have you applied student T distribution Anywhere?**

Student T distribution is a type of normal distribution wherein it can be used for smaller sample sizes and is approximately normal distributed.

I have used it in a situation where I was supposed to measure the average of salary of just 10 people which was very low. But I found that there is some bell curve kind of a structure approximately. So to measure the average of salary I used T distribution

**18. What do you understand by statistical analysis of data, Give me scenario where you have used statistical analysis in last projects?**

In one of my projects I had a dataset wherein there were lot of columns. At that time as from domain I made an assumption that these two features looks same so I looked up for the correlation between the columns vs the dependent column and it was multicollinear so I removed one of the columns. So this helped me in feature selection step. Similarly I have used Chi square based approach also in one of my projects. I use describe function in most of my projects to get to know about the data better.

**19. Can you please tell me criterion to apply binomial distribution, with example?**

The criteria are :

The observations must be independent of each other,

The number of observations must be fixed,

The probability of success is same for each outcome

A real time example is Lottery ticket where there is only two ways either you reach success or failure

**20. There are 100 people, who are taking this particular 30 days Data science interview preparation course, what is the probability that 10 people will be able to make transition in 1 week? If 50 people were able to make transition in 3 weeks?**

**21. lets suppose I have appeared in 3 interviews, what is the probability that I am able to crack at least 1 interview?**

n(S) = 3

P(A<=1) = 1 – P(A=1) \* P(A=2) \* P(A=3)

= 1 – ½ \* ½ \* ½ = 1 – 1/8 = 7/8

**22. Explain Gaussian Distribution in your own way.**

Gaussian distribution is also called as normal distribution which has a bell shaped curve structured data distribution. This is denoted as

N(mu, sigma\*\*2)

Here always the skewness and kurtosis is 0

**23. What do you understand by 1st, 2nd and 3rd Standard Deviation from Mean?**

According to Empirical Formula,

If the data is normally distributed

The data between the mean and 1st standard deviation is 68% approx.

The data between the mean and 2nd standard deviation is 95% approx.

The data between the mean and 3rd standard deviation is 99.7% approx.

If not normal then we can use Chebyshev’s inequality to find how the data is distributed between mean and 1st, 2nd and 3rd deviation.

**24. What do you understand by variance in data in simple words?**

Variance is spread of dataset. It is a statistical measure which will say how the data is distributed. It shows how much a data is deviated from the mean. It is denoted by sigma

Formula: Variance = (Summation of(X –mu)\*\*2)/N

X = Data value

mu = mean

N = Total Population

**25. If variance of dataset is too high, in that case How you will be able to handle it or decrease it?**

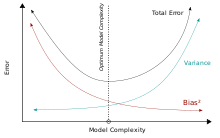
Ensemble models

Increasing the train dataset size

Hyperparameter Tuning

**26. Explain the relationship between Variance and Bias.**

Bias tells us the difference between the average difference between the actual value and predicted value and Variance says how the data is spread. Variance and Bias are inversely proportional which means if bias increase variance will decrease or vice versa.



Overfitting – Low Bias High Variance

Underfitting – High Bias High Variance

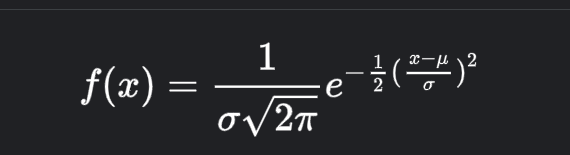
Perfect – Low Bias, Low Variance

**28. What do you understand by Z Value given in Z Table?**

A Z Value in a Z Table will give us how much percentage of data is available to the left of the given Z score

For example for a specific z score 1.25, 89.44% of value lies behind the given z score (left of it)

**29. Do you know a Standard Normal Distribution Formula?**

****

To convert a data into standard normal distributed data the following formula can be used:

**z = (x-mean) / standard deviation.**

**30. Can you please explain critical region in your way?**

It is a region we use in hypothesis testing. We used make a null hypothesis and if the statistic tests like T test Anova test outputs a value in this critical region then we will reject the null hypothesis and accept the alternate hypothesis. So it is also called rejection region.