1.Where you have used Hypothesis Testing in your Machine learning Solution

Evaluate the mutual excusive statement by using sample data on population. Based upon the tests we can decide either Null hypothesis or Alternate Null hypothesis is true.

2. What kind of statistical tests you have performed in your ML Application

To decide Null hypothesis or Alternate hypothsis is true,we do tests:

One sample proportion test,CHI Square test,T test,ANNOVA test,correlation

3. What do you understand by P Value? And what is use of it in ML?

Based upon P value we can decide to reject the null hypothesis or accept the null hypothesis

|  |  |  |
| --- | --- | --- |
|  | Null Hypothesis  H0  True | Null Hypothesis  H0  False |
| Accept the Null hypothesis  H0 | Correct decision | Type 2 Error |
| Reject the Null hypothesis  H0 | Type I Error | Correct decision |

4. Which type of error is severe Error, Type 1 or Type 2? And why with example

Type 1 Error: When we reject the null hypothesis(H0) ,if H0 is true then we get type 1 error

Type 2 Error: When we accept the null hypothesis(H0),if H0 is False then we get type 2 error

We need to reduce the both error and It depends upon the problem statement

5. Where we can use chi square and have used this test anywhere in your application

To find whether these 2 categorical variables are independent or dependent, we use CHI square test.Based upon P value we can can accept Null hypothesis(H0) or reject

6. Can we use Chi square with Numerical dataset? If yes, give example. If no, give Reason?

We can’t perform for data type numerical (we can find the relationship and strength from pearson correlation coefficient or Spearman rank correlation. For object data type we use CHI square test

7. What do you understand by ANOVA Testing?

If we want to see any relationship between categorical variables(more than 2 categories) and numerical variables we can perform ANOVA test.Finding the mean within the column and also finding the mean between the columns.Based upon F statistics we can prove whether mean across is same or not.

8. Give me a scenario where you can use Z test and T test.

9. What do you understand by inferential Statistics?

Based upon sample data, we do test and come up with conclusion. This conclusion that works for population data.

10. When you are trying to calculate Std Deviation or Variance, why you used N-1 in Denominator? (Hint: Basel Connection)

By doing denominator n-1 in calculating the sample variance,we get bigger value variance at some point it converges with true mean

11. What do you understand by right skewness, Give example?

The tail skewed at the right side and its mean will be greater than median and mode.for example wealth distribution,few people revenue income will more than 100cr

12. What is difference between Normal distribution and Std Normal Distribution and Uniform Distribution?

Normal distribution : Mean can any value and it is symmetrical.mean = median =mode

Standardad Normal distribution: mean = 0 and sd = 1.

Uniform distribution: It’s a continuous probability distribution that events likely to occur

13. What is different kind of Probabilistic distributions you heard of?

1. Normal distribution
2. Bernouil distribution
3. Binomial distribution
4. Poisson distribution
5. Exponential distribution

14. What do you understand by symmetric dataset?

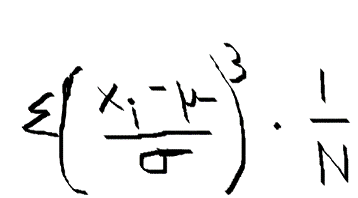
Mean will be centre and divides that one halve is mrror of other halve. The values appear at regular frequencies and often the mean,median and mode occur at same point

15. In your last project, were you using symmetric data or Asymmetric Data, if its asymmetric, what kind of EDA you have performed?

Logarithmic Transformation

16. Can you please tell me formula for skewness?

Population Skewness:



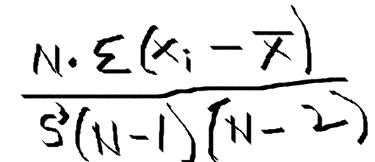
Xi individual score

μ population mean

σ population standard deviation

N population size

Sample Skewness:



Xi individual score

X sample mean;

S sample standard deviation

N sample size

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

Based upon statistical analysis of sample data we will have decision for population data.

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

P(x<=1) = 1/3 = 0.33

22. Explain Gaussian Distribution in your own way.

Mean =median = mode and the distribution is symmetric

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

1st standard deviation = about 68% of data lies in this region

2nd Standard deviation = about 95% of data lies in this region

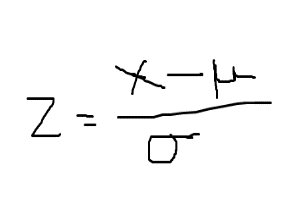
3rd Satandard deviation = about 99% of data lies in this region

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

Variance describes the spread of data from the mean

26. Explain the relationship between Variance and Bias.

29. Do you know a Standard Normal Distribution Formula?



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

Calculated Sample space value which lies on rejecting the null hypothesis region is called critical region

31. How you can define Machine Learning?

Machine Learning focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy

32.What do you understand Labelled training dataset?

The target dataset datas will be in labelled for example SPAM detection mail

33. What kind of Machine learning algorithm would you used to walk robot in various unknown area?

Reinforcement learning

34. What do you understand by Gradient decent? How will you explain Gradient decent to a kid?

Gradient descent is an optimization algorithm used to find the values of parameters (coefficients) of a function(f) that minimizes a cost function (cost).

If we are on top of hill and we have to get down to bottom of hill.Then we have to move step into near by rock and keep continuing until we reach close to bottom of hill.At some point we know that we are close to bottom of hill then step slowly to reach the bottom of the hill.

35. Can you please explain diff between regression & classification?

Regression: Our independent variables depend on dependent target variables then we say linear regression.

Classification: Our target variables classified based upon labels.

36.If I give you 2 columns of any dataset, what will be the steps will be involved to check the relationship between those 2 columns?

If the two columns datatype are numerical we can find whether correlation exist between the columns

If the two column datatype are categorical we can perform CHI Square test and there is any difference exist or not based upon hypothesis test we can find

37. Can you please explain 5 diff kind of strategies at least to handle missing values in dataset?

1.Mean /Median/Mode imputation

2.Label imputation

3.End of distribution imputation

4.Frequency categories imputation

38. What is your strategy to handle categorical dataset? Explain with example.

* If the two column datatype are categorical we can perform CHI Square test and there is any difference exist or not based upon hypothesis test we can find
* Handling the null values
* We convert categorical data by using one of the technique one-Hot encoding

39. Explain 5 different encoding techniques

One Hot encoding

Label encoding

40. How do you define some features are not important for ML model? What strategy will you follow

Lasso regression also helpful in finding feature selection.we can find the high correlation of features and either of the feature is selected for feature selction

41. If your model is overfitted, what you will do next?

We can do regularization either L1 norm or L2 norm

42.How can you avoid overfitting in decision tree?

By using regularization we can avoid overfitting

43. Give me list of 10 activation functions with explanation

1. Sigmoid Function
2. Tangent(Tanh) Function
3. Softmax Function
4. Softsign Function
5. ReLU(Rectified Linear Unit) Function
6. ELUs(Exponential Linear Units) Function

44. What do you understand by data warehousing?

Storing and maintain the historic and current data for the future forecasting.

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

Based upon sample mean ,sample standard deviation and sample size we can build confidence interval so the population mean lies within that for example 95% confidence interval. Z value can get from Z table and multiplied with standard error(SD/(n)^1/2) to get an marginal error.Sample mean is subtracted and added from marginal error to get an confidence interval.

46. What do you understand by multicollinearity and homoscedasticity in Dataset?

Multicollinearity:

Multicollinearity is the occurrence of high intercorrelations among two or more independent variables

Homoscedasticity:

Homoscedasticity describes a situation in which the error term is the same across all values of the independent variables