Q.1. What do you understand by P Value?

Solution: It is probability of observing a value of a statistic test that is more extreme than what was observed in the sample, assuming null hypothesis is true.

Q2. Where we can use chi square and have used this test anywhere in your application?

Solution: chi square test is used on categorical data to find relation between categorical features.

Q3. Can we use Chi square with Numerical dataset? If yes, give example. If no, give Reason?

Solution: No , reason: chi square test is used to find relation between independent groups (categories) also chi square test can only be used on actual numbers and not on percentages, proportions, means, etc. Numerical data have continuous variable while chi square requires discrete independent variables .

Q.4 What do you understand by ANOVA Testing?

Solution: Statistical Technique developed to study significance of difference of ‘Means’ of more than 2 or more samples.

Q.5. Give me a scenario where you can use Z test and T test.

Solution: When Population std. Deviation is given and sample size (n>=30) we use Z test and when Population Standard Deviation is not given and sample size (n<=30) we use T test.

Q.6. What do you understand by inferential Statistics?

Solution: Inferential statistics allows you to make predictions from data. With inferential statistics, you take data from samples and make generalizations about a population using various techniques like Hypothesis testing and various tests like Z test ,T test, chi square test, Anova test ,etc.

Q.7. When you are trying to calculate Std Deviation or Variance, why you used N-1 in Denominator? (Hint: Basel Connection)

Solution: to avoid bias estimation or to avoid bias due to finite sample size because our sample std Deviation or variance might get too much away from actual population std deviation or variance.

Q.8. What do you understand by right skewness, Give example?

Solution: The right-skewed distribution has a long right tail. Right-skewed distribution are also called positive-skew distributions. That’s because there is a long tail in the positive direction on the number line. The mean is also to the right of the peak. And most values are clustered around the left tail of the distribution.

Q.9. What is difference between Normal distribution and Std Normal Distribution and Uniform Distribution?

Solution: Normal distribution: normal distribution can have any real values for the mean and standard deviation. It is probability distribution which peaks out in the middle and gradually decreases towards both ends of axis. Also known as Gaussian distribution and bell curve.

Standard Normal Distribution: It is also kind of Normal Distribution with mean of zero and a standard deviation of 1. It is also a bell curve.

Uniform Distribution: It is a probability distribution where probability of x is constant. That is all points in range are equally likely to occur consequently it looks like a rectangle.

Q.10. What is different kind of Probabilistic distributions you heard of?

Solution: Normal distribution, binomial distribution, chi square distribution and Poisson distribution.

Q.11. What do you understand by symmetric dataset?

Solution: A symmetric distribution is a type of distribution where the left side of the distribution mirrors the right side . The normal distribution is symmetric. In a symmetric distribution, the mean, mode and median all fall at the same point.

Q12. Can you please tell me criterion to apply binomial distribution, with example?

Solution: criteria to apply binomial distribution:

1)Experiment should repeated for fix number of times.

2)The trials are independent.

3)Trials have 2 mutually exclusive outcomes either success or failure.

4)The probability of success is same for all trials.

Example: The number of heads obtained in tossing a coin n times.

Q13. lets suppose I have appeared in 3 interviews, what is the probability that I am able to crack at least 1 interview?

Solution: by using binomial distribution p(x>=1)=p(x=1)+p(x=2)+p(x=3)

Formula= n Cx Px(1-P)n-x

P(x=1)+P(x=2)+P(x=3) = 0.875 or 87.5%.

Q14. Explain Gaussian Distribution in your own way.

Solution: Gaussian Distribution is also called normal distribution. The Probability is high and forms peak at center and gradually decreases along the ends. The curve is symmetric at the center. The total area under the curve is 1.The mean, median, mode are all equal.

Q15. What do you understand by 1st,2nd and 3rd Standard Deviation from Mean?

Solution: According to empirical rule, 68% of observations falls within the first standard deviation (µ±Ϭ), 95% within the first two standard deviations(µ±2Ϭ), and 99.7% within the first three standard deviations(µ±3Ϭ). It indicates spread of data about the mean.

Q16. What do you understand by variance in data in simple words?

Solution: Variance measures how far each number in the set is from the mean and thus from every other number in the set.

Q17. If variance of dataset is too high, in that case How you will be able to handle it or decrease it?

Solution: We must add bias, to reduce variance of an Estimate 1)Repeat the estimate on many different small samples of data from the domain and calculate the mean of the estimates.

2)Another approach would be to dramatically increase the size of the data sample on which we estimate the population mean.

Q18. Explain the relationship between Variance and Bias.

Solution: Variance and Bias are inversely related , if Variance increases then bias decreases and vice-versa.