Q1 to Q9----

1) a
2) a
3) c
4) d
5) c
6) a
7) b
8) a
9) c

10) What do you understand by the term Normal Distribution?

In statistics, Normal distribution is Gaussian, Gauss, or Laplace-Gauss distribution, it was first discovered in 1733 by English Mathematician De-moivres who obtain the continuous as a limiting form of Binomial Distribution.

Definition: Random variable x is said to follow normal distribution with parameters μ and σ if its pdf is given by probability law

F (x: μ, σ) = 1/σ√π{exp[-1/2((x-μ)/σ) ^2)]}
Where μ is mean and σ^2 is the variance of normal distribution

Standard Normal distribution: It is a special case of Normal Distribution as it has μ=0 and σ=1, it is also known as unit normal distribution.

Properties of Normal Distribution:

1. The normal distribution is the only distribution whose cumulants beyond mean and variance are zero.
2. The normal distribution is symmetric about its mean and non-zero over thr entire real line.

3. The value of normal distribution is practically zero when the value of x lies more than a few standard deviations away from mean.

   The main advantage of the normal distribution is that it is symmetric and bell-shaped. This shape is useful because it can be used to describe many populations
   Another main advantage of normal distribution is central limit theorem, In this we can use inferential statistics method that assumes normality, even if the data in our sample doesn't follow normal distribution.

11) How do you find missing data and what imputation techniques do u recommend?

Missing Data or missing values occurs when we don't remember to store data values for certain variables. Data can go missing due to incomplete entries, wrong entries, lost files, improper instruments, and many other reasons. By observing such things, we can find the missing data values in our whole dataset.

Imputation Techniques: It is a technique of replacing a missing value with another value based on reasonable estimate. We can replace these missing values by the mean or mode of the whole column, or we can drop the whole column if it has only zero values or a greater number of missing values.

12) What is A/B testing?

A/B testing is also known as bucket or split-run testing, it is a user experience research methodology. A/B consists of a random experiment with two variants A and B. It includes two-sample hypothesis testing or statistical hypothesis testing used in the field of statistics. A/B is used to compare two versions of single variable. It tests the subject's response towards variant A and towards variant B and to conclude which of the two variants is more effective.

13) Is mean imputation of missing data acceptable technique?

Mean imputation is the replacement of missing observations with the mean of the non-missing observations for that variable, in my opinion it's a good technique for replacing the missing values as it is easy and a simple way, but it has some drawbacks also that mean imputation does not preserve the relationships among variables, it also leads to an underestimate of standard errors, so sometimes it is a good option but not every time.

14) What is linear regression in statistics?

If the variables in a bivariate distribution are related, we will find that the points in the scatter diagram will cluster around some curve called the "curve of regression". If the curve is a straight line, then it is called line of regression and there is said to be linear regression between the variables

The line of regression is the line which gives the best estimate to the value of one variable for any specific value of the other variable. Thus, the line of regression is the line of best fit and obtain nu the principal of least square.

15) What are the various branches of statistics?

The two main branches of statistics are Descriptive statistics and Inferential statistics.

Descriptive statistics deals with the presentation and collection of data. Different areas of study need different kind of analysis using descriptive statistics.

Inferential statistics involves drawing right conclusion from the statistical analysis that has been performed using descriptive statistics. Predictions of the future and generalizations about population by studying a smaller sample comes under the preview of inferential statistics.