Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Bernoulli random variables take (only) the values 1 and 0.

Ans- (True)

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

Ans- A (Central Limit Theorem)

3. Which of the following is incorrect with respect to use of Poisson distribution?

Ans- B (Modelling bounded count data)

4. Point out the correct statement.

Ans- A (The exponent of a normally distributed random variables follows what is called the log-normal distribution)

5. _____ random variables are used to model rates.

Ans- C (Poisson)

6. Usually replacing the standard error by its estimated value does change the CLT.

Ans- B(False)

7. 1. Which of the following testing is concerned with making decisions using data?

Ans- B(Hypothesis)

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.

Ans- A (0)

9. Which of the following statement is incorrect with respect to outliers?

Ans- C (Outliers cannot confirm to the regression relationship)

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?
Ans- It is a probability function that specifies how values of a variable are distributed is called the normal distribution. Also, it is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme. Height is one simple example of something that follows a normal distribution.

11. How do you handle missing data? What imputation techniques do you recommend?

Ans- As we know that a data frame can consists of many rows where each row can have values for various columns. if a value corresponding to a column is not present, it is considered to be a missing value. a missing value is denoted as NaN.

In the real-world dataset, it is common for an object to have some missing attributes. There may be several reasons for that. In some cases, data was not collected properly resulting in missing data i.e. Some people did not fill all the fields while taking the survey. sometimes some attributes are not relevant to all. For example, if a person is unemployed then salary attribute will be irrelevant and hence may not have been filled up.

The two most common strategies for handling missing values explained in this section are:

1. drop the object having missing values.

2.fill or estimate the missing values

12. What is A/B testing?
Ans- A/B testing is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B.it includes application of statistical hypothesis testing or two sample hypothesis testing as used in the field of statistics.

13. Is mean imputation of missing data acceptable practice?

Ans- Replace missing values with the mean value from the rest of column (columns, not rows). A column represents a single feature, it only makes sense to take mean from other samples with same feature.

Fast and easy won't affect mean or sample size of overall data set.

. Median may be a better choice than mean when outliers are present

. But it's generally pretty terrible.

. only works on column level, misses correlations between features.

. can’t use on categorical features (imputing with most frequent value can work in this case though).

. not very accurate.

14. What is linear regression in statistics?
Ans- It is one of the most fundamental and widely known Machine Learning algorithms
The building blocks of a linear regression model are:

Discrete /continuous independent variables- It is a best fit regression line.

Continuous dependent variable i.e. A linear regression model predicts the dependent variable using a regression line based on the independent variables.

Equation is Y= a+bx (a= intercept, b= slop of the line, e= error term)

15. What are the various branches of statistics?

Ans- There are three real branches of statistics

-Data collection, Descriptive statistics and inferential statistics