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

Answer. A) 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?

Answer. A) central limit theorem

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

Answer. b) Modeling bounded count data

4. Point out the correct statement?

Answer. D) All of the mentioned.

5. _____ random variables are used to model rates.

Answer. C) Poisson

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

Answer. b) False

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

Answer. b) Hypothesis

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

Answer. A)0

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

Answer. C) Outliers cannot conform to the regression relationship.

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

Answer. A normal distribution 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.
The normal distribution is in bell shaped curve. Normal distribution

have:-

. Mean → 0

. standard deviation → 1

. skewness → 0

. kurtosis → 3

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

Answer. As we know that a dataframe 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 field's 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?

Answer. 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 felid of statistics.

13. Is mean imputation of missing data acceptable practice?

Answer. It is acceptable when the missing value proportion is not large enough.
But, when the missing values are large enough and you impute them with the
mean, the standard errors will be lesser than what they actually would have been.
Small standard errors can lead to small p-values and this can create problems for us,
because some variables will start appearing significant, which are ideally not significant

14. What is linear regression in statistics?

Answer: Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and

in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

Naming the Variables. There are many names for a regression's dependent variable. It may be called an outcome variable, criterion variable, endogenous variable , etc The independent variables can be called exogenous variables, predictor variables, or regressors.

Three major uses for regression analysis are (1) determining the strength of predictors, (2) forecasting an effect, and (3) trend forecasting.

15. What are the various branches of statistics?

Answer 15. There are three real branches of statistics :-

1- Data collection : Data collection is all about how the actual data is collected. For the most part, this needn't concern us too much in terms of the mathematics (we just work with what we are given), but there are significant issues to consider when actually collecting data.

2- Descriptive statistics : Descriptive statistics is the part of statistics that deals with presenting the data we have. This can take two basic forms – presenting aspects of the data either visually (via graphs, charts, etc.) or numerically (via averages and so on).

3- Inferential statistics: Inferential statistics is the aspect that deals with making conclusions about the data. This is quite a wide area; essentially you are asking 'What is this data telling us, and what should we do?