.
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. 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.

Answer. C)poission

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)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.

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 varients ,A and B.it includs application of statistical hypothesis testing or two sample hypothesis testing as used in the feild of statistics.

14. What is linear regression in statistics?

Answer. Linear regression is one of the most fundamental and widely known as Machine Learning algorithms which people start with building blocks of a linear regression model are :
Discrete /continuous independent variables
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.
the equation of the linear regression is

$$Y= a+bx$$

a= intercept
b= slop of the line
e= error term

5. What are the various branches of statistics?

Answer.5. There are three real branches of statistics
1- Data collection
2-Descriptive statistics
3-inferential statistics

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 feilds while taking the survey. sometimes some attribute 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

13. Is mean imputation of missing data acceptable practice?

Answer. 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 the mean from other samples of the 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.