1. Bernoulli random variables take (only) the values 1 and 0.
a) True
b) False
ans:a

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?
a) Central Limit Theorem
b) Central Mean Theorem
c) Centroid Limit Theorem
d) All of the mentioned
ans:a

3. Which of the following is incorrect with respect to use of Poisson distribution?
a) Modeling event/time data
b) Modeling bounded count data
c) Modeling contingency tables
d) All of the mentioned
ans:a

4. Point out the correct statement.
a) The exponent of a normally distributed random variables follows what is called the log- normal distribution
b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
c) The square of a standard normal random variable follows what is called chi-squared distribution
d) All of the mentioned
ans:d

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

ans:c

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False
ans:b

7. 1. Which of the following testing is concerned with making decisions using data?
a) Probability
b) Hypothesis
c) Causal
d) None of the mentioned
ans:b

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0
b) 5
c) 1
d) 10
ans:a

9. Which of the following statement is incorrect with respect to outliers?
a) Outliers can have varying degrees of influence
b) Outliers can be the result of spurious or real processes
c) Outliers cannot conform to the regression relationship
d) None of the mentioned
ans:c

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

A normal distribution is also called as probability bell curve.
In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3. Normal distributions are symmetrical in nature.

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

Handling missing data is important as many machine learning algorithms do not support data with missing values.
If missing data is less in number NaN values are replaced with mean in continuous data and mode in Categorical data
Simple imputer
Knn imputer
Iterative imputer

12. What is A/B testing?

A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

13. Is mean imputation of missing data acceptable practice?

No, the imputation technique depends on the size of data and percentage of missing values on total data, if the missing values are high mean is not the recommended as it will give same repeated result in the column, hence analysis will be meaningless.

14. What is linear regression in statistics?

It is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) 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.

15. What are the various branches of statistics?

1. Descriptive statistics.
2. Inferential statistics.