1. Bernoulli random variables take (only) the values 1 and 0. **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?
   a) Central Limit Theorem
3. Which of the following is incorrect with respect to use of Poisson distribution?
   b) Modeling bounded count data
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**
5. _____ random variables are used to model rates. **c) Poisson**
6. 10. Usually replacing the standard error by its estimated value does change the CLT. **b) False**
7. 1. Which of the following testing is concerned with making decisions using data?
   b) Hypothesis
8. Normalized data are centered at_____and have units equal to standard deviations of the original data. **a) 0**
9. Which of the following statement is incorrect with respect to outliers?
   c) Outliers cannot conform to the regression relationship

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

In normally distributed data, there is a constant proportion of data points lying under the curve between the mean and a specific number of standard deviations from the mean

In Normal distribution all data will be covered under bell shaped curve.The shape of the normal distribution is perfectly symmetrical.

This means that the curve of the normal distribution can be divided from the middle and we can produce two equal halves. Moreover, the symmetric shape exists when an equal number of observations lie on each side of the curve.

The mean, median, and mode are equal

In normally distributed data, there is a constant proportion of data points lying under the curve between the mean and a specific number of standard deviations from the mean.

Thus, for a normal distribution, almost all values lie within 3 standard deviations of the mean

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

A real-world dataset can have lot of reasons to have missing data.

There are three categories of missing data,

Missing completely at random (MCAR), Missing at random (MAR), Not missing at random (NMAR)

To handling missing data we are using various imputation techniques,

Imputation Using (Mean/Median) Values

the mean/median of the non-missing values in a column and then replacing the missing values within each column separately and independently from the others

Imputation Using (Most Frequent) or (Zero/Constant) Values

by replacing missing data with the most frequent values within each column

Imputation Using k-NN

very useful in making predictions about the missing values by finding the *k's* closest neighbours to the observation with missing data and then imputing them based on the non-missing values in the neighbourhood

Imputation Using Multivariate Imputation by Chained Equation (MICE)

filling the missing data multiple times. Multiple Imputations (MIs) are much better than a single imputation as it measures the uncertainty of the missing values in a better way

Imputation Using Deep Learning

categorical and non-numerical features. It is a library that learns Machine Learning models using Deep Neural Networks to impute missing values in a data frame.

12. What is A/B testing?

A/B testing is one of the most popular controlled experiments used to optimize web marketing strategies. It allows decision makers to choose the best design for a website by looking at the analytics results obtained with two possible alternatives A and B

A and B. Visitors of a website are randomly served with one of the two. Then, data about their activity is collected by web analytics. Given this data, one can apply statistical tests to determine whether one of the two designs has better efficacy

13. Is mean imputation of missing data acceptable practice?

Mean imputation is not acceptable in practice. It will work fine only on small datasets. Mean imputation does not preserve the relationships among variables

14. What is linear regression in statistics?

Regression is a supervised learning technique that supports finding the correlation among variables. A regression problem is when the output variable is a real or continuous value

Regression shows a line or curve that passes through all the data points on a target-predictor graph

Linear regression is a quiet and simple statistical regression method used for predictive analysis and shows the relationship between the continuous variables. Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis)

The linear regression example is y=mx+c

15. What are the various branches of statistics

Statistics is the study and the manipulation of the data, including methods for data collecting, organisation, analysis and conclusion.
Descriptive and inferential statistics are the two main areas of statistics
descriptive statistics, which deals with the presentation and collecting of data

Generally, descriptive statistics can be categorized into

Measures of central tendency
Measures of variability
measures of tendency and measures of variability easily use graphs, tables, and general discussions

Mean, Median, Mode

Measures of Variability
The measure of variability helps the statisticians in analysing the distribution that comes from a particular data set. Quartiles, ranges, variances, and standard deviation are the variability variables.