Q1: Bernoulli random variables take (only) the values 1 and 0.
A1: a) True

Q2: 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?
A2: a) Central Limit Theorem

Q3: Which of the following is incorrect with respect to use of Poisson distribution?
A3: b) Modeling bounded count data

Q4: Point out the correct statement.
A4: d) All of the mentioned

Q5: _____ random variables are used to model rates.
A5: c) Poisson

Q6: Usually replacing the standard error by its estimated value does change the CLT.
A6: b) False

Q7: Which of the following testing is concerned with making decisions using data?
A7: b) Hypothesis

Q8: Normalized data are centered at_____and have units equal to standard deviations of the original data
A8: a) 0

Q9: Which of the following statements is incorrect with respect to outliers?
A9: c) Outliers cannot conform to the regression relationship

Q10: What do you understand by the term Normal Distribution?
A10: A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the mean of the distribution.
The normal distribution is also known as a Gaussian distribution or probability bell curve. It is symmetric about the mean and indicates that values near the mean occur more frequently than the values that are farther away from the mean.

Q11: How do you handle missing data? What imputation techniques do you recommend?
A11: Deleting the Missing values
Imputing the Missing Values
- Replacing with an arbitrary value
- Replacing with the mean
- Replacing with the mode
- Replacing with the median
- Replacing with the previous value – forward fill
- Replacing with the next value – backward fill
- Interpolation

Q12: What is A/B testing?
A12: A/B testing, also known as split testing, is a marketing experiment wherein you split your audience to test a number of variations of a campaign and determine which performs better. In other words, you can show version A of a piece of marketing content to one half of your audience, and version B to another.

Q13: Is mean imputation of missing data acceptable practice?
A13: Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.
Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

Q14: What is linear regression in statistics?
A14: Regression analysis is an important statistical method for the analysis of medical data. It enables the identification and characterization of relationships among multiple factors. It also enables the identification of prognostically relevant risk factors and the calculation of risk scores for individual prognostication.

Q15: What are the various branches of statistics?
A15: There are three real branches of statistics: data collection, descriptive statistics and inferential statistics