Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

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

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

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

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.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

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

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

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

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
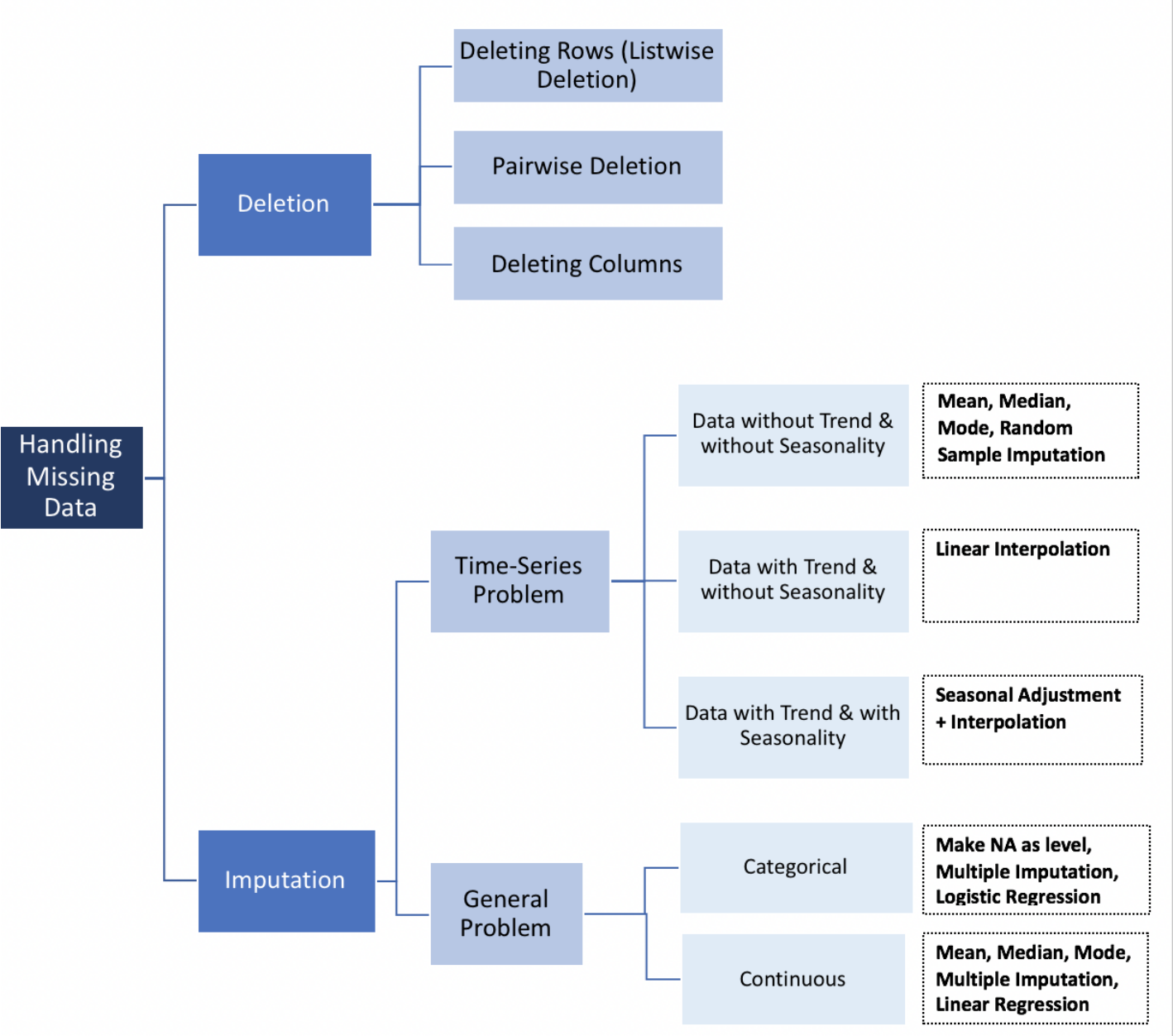c) Outliers cannot conform to the regression relationship
d) None of the mentioned

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

The distribution of a variable which looks like a bell-shaped curve is called as Normal Distribution. It is also called Gaussian distribution because it was discovered by Carl Friedrich Gauss. It is a continuous probability distribution and is one of the most common distributions you will come across in statistical studies. It is also important to note here that it is a continuous distribution.

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

Handling Missing Data
- Deletion
  - Deleting Rows (Listwise Deletion)
  - Pairwise Deletion
  - Deleting Columns
- Imputation
  - Time-Series Problem
    - Data without Trend & without Seasonality → **Mean, Median, Mode, Random Sample Imputation**
    - Data with Trend & without Seasonality → **Linear Interpolation**
    - Data with Trend & with Seasonality → **Seasonal Adjustment + Interpolation**
  - General Problem
    - Categorical → **Make NA as level, Multiple Imputation, Logistic Regression**
    - Continuous → **Mean, Median, Mode, Multiple Imputation, Linear Regression**

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?

- Bad practice in general
- If just estimating means: mean imputation preserves the mean of the observed data
- Leads to an underestimate of the standard deviation
- Distorts relationships between variables by "pulling" estimates of the correlation toward zero

14. What is linear regression in statistics?
Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable.

15. What are the various branches of statistics?
The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.
Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.
Inferential statistics, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.