Statistics assignment

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.
d) All of the mentioned

5 - _____ random variables are used to model rates.
c) Poisson

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

7 - 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?
A variable is normally distributed when its data points' (observations) spread follow a belle curve shape

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

When there are missing/null values (NaN) in a dataset, we have three ways of handling them - first, we can leave them and treat then as "missing" or "NaN" if the concerned variable is categorical, but there should be evidence that leaving them will not bias dataset

- secondly, we can remove the data either by row or column, but again there should be evidence that removing them will not bias dataset

- thirdly, we can replace them with a relevant data point value (imputation), accordingly whether it is a numerical or categorical variable. Here, we should make sure that the replacement is true/coherent to the concerned variable.

As imputation techniques, we can k-nearest neighbour (KNN), or replace the NaN with the column mean, median for example (if numerical variable) and with mode for example if it is categorical variable

12 - What is A/B testing?

A/B testing is an experiment where two or more versions of a variable are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

13 - Is mean imputation of missing data acceptable practice?
Mean imputation is not acceptable because it is very likely not to preserve relationship between variables

14 - What is linear regression in statistics?
Linear regression is the prediction of a known target/label/dependent variable based on known feature(s)/independent variable(s), using the best fit line [ordinarily least square: minimizing the sum of square of errors (difference between expected and predicted values)]

15 - What are the various branches of statistics?

The branches of statistics are
- descriptive statistics which helps to understand data and the associated patterns
- inferential statistics which helps to generate new data using sampled data
- predictive statistics which helps to predict future based on past occurrences
- prescriptive statistics which helps with corrective measures