Statistics Worksheet

1) Bernoulli random variables take (only) the values 1 and 0.
Ans - 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?
Ans - a) Central Limit Theorem

3) Which of the following is incorrect with respect to use of Poisson distribution?
Ans - b) Modeling bounded count data

4) Point out the correct statement.
Ans - d) All of the mentioned

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

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

7) Which of the following testing is concerned with making decisions using data?
Ans - b) Hypothesis

8) Normalized data are centered at_____and have units equal to standard deviations of the original data.
Ans - a) 0

9) Which of the following statement is incorrect with respect to outliers?
Ans - c) Outliers cannot conform to the regression relationship

10) What do you understand by the term Normal Distribution?
Ans - Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, the normal distribution appears as a "bell curve".

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

Ans - When dealing with missing data, we use two primary methods to solve the error: imputation or the removal of data. The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low.
Imputation Techniques commonly used,

1. Listwise or case deletion. ...
2. Pairwise deletion. ...
3. Mean substitution. ...
4. Regression imputation. ...
5. Last observation carried forward. ...
6. Maximum likelihood. ...
7. Expectation-Maximization. ...
8. Multiple imputation.

As many more.

12) What is A/B testing?

Ans - A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

13) Is mean imputation of missing data acceptable practice?

Ans - Mean imputation is typically considered terrible practice since it ignores feature correlation.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.

14) What is linear regression in statistics?

Ans - Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15) What are the various branches of statistics?

Ans - There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.