STATISTICS SET 1

1. Bernoulli random variables take (only) the values 1 and 0.

   True.

2. Which of the following theorem states that the distribution of averages of independent variables, properly normalized, becomes that of a standard normal as the sample size increases?
   Central limit theorem.
3. Which of the following is incorrect with respect to use of Poisson distribution?
   Modelling bounded count data.
   Modelling contingency tables.
4. Point out the correct statement.
   All of the mentioned.
5. ____random variables are used to model rates.
   Poisson.
6. Usually replacing the standard error by its estimated value does change the CLT.
   False.
7. Which of the following testing is concerned with making decisions using data?
   Hypothesis.
8. Normalized data are centred at___ and have units equal to standard deviations of the original data.
   0
9. Which of the following statement is incorrect with respect to outliers?
   Outliers cannot conform to the regression relationship.
10. What do you understand by the term Normal Distribution?
    The normal distribution, also known as the Gaussian distribution, is the most important probability distribution in statistics for independent random variables. It is a continuous probability distribution that is symmetrical around its mean.
11. How do you handle missing data? What imputation techniques do you recommend?
    There are three main approaches to handle missing data:(1) Imputation- where values are filled in the place of missing data, (2) Omission – where samples with invalid data are discarded from further analysis and (3) analysis-by directly applying methods unaffected by the missing values.
    The simplest imputation method is replacing missing values with the mean or median values of the dataset at large.
12. What is A/B testing?
    An experiment to compare two competing options (A, B). To determine what is the better option.
13. Is mean imputation of missing data acceptable practice?
    True, imputing the mean preserves the mean of the observed data. So, if the data are missing completely at random, the estimate of the mean remain unbiased. Mean imputation does not preserve relationships between variables such as correlations, mean imputation is not a good solution.
14. What is linear regression in statistics?
    Linear regression is an attempt to model the relationship between two variables by fitting a linear equation to observed data, where one variable is considered to be an explanatory variable and the other as a dependent variable.
15. What are the various branches of statistics?

The two major areas of statistics are known as descriptive statistics, which describes the properties of sample and population data, and inferential statistics, which uses those properties to test hypothesis and draw conclusions.