Statistics Worksheet 1

Name: Aabha Pravin Tathed

Batch: Internship 25

Answers

1. A)True
   The Bernoulli distribution arises as the result of a binary outcome.
2. A) Central Limit Theorem
   The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution
3. B) Modeling bounded count data
   Poisson distribution is used to show how many times an event is likely to occur over a specified event
4. D) All of the Above
   The exponent of normally distributed random variables follows what is called the log- normal distribution. Sums of normally distributed random variables are again normally distributed even if the variables are dependent and The Square of a standard normal random variable follows what is called chi-squared distribution.
5. C) Poisson
   Poisson random variables are used to model rates
6. B) False
   Usually replacing the standard error by its estimated value doesn't change the CLT.
7. B) Hypothesis
   Hypothesis testing is concerned with making decisions using data
8. A) 0
   Normalized data are centred at 0 and have units equal to standard deviations of the original data.
9. c) Outliers cannot conform to the regression relationship
   Outliers can confirm to the regression relationship
10. In normal distribution the values are evenly distributed both above and below the average. Its graph looks like bell curve and it is symmetrical to the centre, which is the right side of the centre is exactly same as left side of the centre.
    For example, classroom test results. Many students will get average marks, some students will get below average and some will get above average.
11. There are some types of missing data like missing completely at random, missing at random, not missing at random.
    Common methods to handle missing data
    1. Complete case analysis
       It completely removes the rows which has missing values

2. Arbitrary Value Imputation
   It handles numerical values.It groups the missing values in a column and assign them to a new value that is far away from the range of that column.
3. Mean/median imputation
   It replaces the missing value with the variable with the Mode of that column

12. It is a process where decision is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not. Prediction is made that one will perform better than two then, data sets from both pages are observed and compared to determine if two is a statistically significant improvement over one.
13. It is not recommended to use mean imputation in practice. Large number of data will have impact on Mean, it will improve power but it will not give proper results. Results will be biased.
14. Linear regression is used to predict the value of a variable based on the value of another variable. The variable which is to be predicted is dependent variable and the variable which will be used to predict the other variable's value is independent variable.
15. The two main braches of statistics are descriptive statistics and inferential statistics. Descriptive statistics describes the properties of sample and population data and inferential statistics uses those properties to test, predict and to find conclusions.