STATISTICS WORKSHEET-1

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 Ans:A
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 Ans:A
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 Ans:D
5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned Ans:C
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False Ans:A
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 Ans:B
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 Ans:A
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 Ans:C

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

10. What do you understand by the term Normal Distribution?
11. How do you handle missing data? What imputation techniques do you recommend?
12. What is A/B testing?
13. Is mean imputation of missing data acceptable practice?
14. What is linear regression in statistics?
15. What are the various branches of statistics?

10Ans) Normal distribution is a continuous probability distribution wherein values lie in a symmetrical fashion mostly situated around the mean. It is also known as Gaussian Distribution and it is proper term for probability bell curve. In Normal Distribution mean is Zero and Standard Deviation is 1. It has Zero Skew.

The normal distribution is the most common type of distribution assumed in technical stock market analysis and in other types of statistical analyses. The standard normal distribution has two parameters: the mean and the standard deviation. For a normal distribution, 68% of the observations are within +/- one standard deviation of the mean, 95% are within +/- two standard deviations, and 99.7% are within +- three standard deviations. The normal distribution model is motivated by Central limit theorem. This theory states that averages calculated from independent, identically distributed random variables have approximately normal distributions, regardless of the type of distribution from which the variables are sampled (provided it has finite variance). Normal distribution is sometimes confused with symmetrical distribution. Symmetrical distribution is one where a dividing line produces two mirror images, but the actual data could be two humps or a series of hills in addition to the bell curve that indicates a normal distribution

11Ans)Imputation Techniques are used to handle missing data.To handle them mean, median and mode are techniques.

12Ans)It is a kind of experiment which includes two samples or two variants and it uses statistical hypothesis or sample hypothesis.It is also used to find right price of product.

13Ans)Mean imputation is bad practice in general for missing data.

14Ans)It is used to predict the value of variable and and linear regression is the most commonly used predictive analysis.

15Ans)There are two main branches in Statistics:
1.Descriptive Statistics
2.Inferential Statistics
Statistics are commonly used in applications and professions.

Descriptive Statistics:
In Descriptive statistics two categories are there, they are measures of central tendency and variability.
Measures of tendency are mean, median and mode.
Measures of variability are standard deviation, variance ,Quartiles and ranges.

Inferential Statistics:
This is mainly used for data analysis, writing and drawing conclusions.This is accomplished by taking samples and determining their reliability.

Some different types of inferential statistics are:

1.Regression Analysis
2.Analysis of Variance
3.Analysis of covariance
4.Statistical significance
5.Correlation analysis.