

## STATISTICS WORKSHEET-1

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

- a) True
- b) False

**Ans : a)True**

2. Which of the following theorems states that the distribution of averages of iid variables, properly normalised, 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)Central Limit Theorem**

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

**Ans : b) Modeling bounded count data**

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)All of the mentioned**

5. \_\_\_\_\_ random variables are used to model rates.

- a) Empirical
- b) Binomial
- c) Poisson
- d) All of the mentioned

**Ans : c)Poisson**

6. Usually replacing the standard error by its estimated value does change the CLT.

- a) True
- b) False

**Ans : b) False**

7. 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) Hypothesis**

8. Normalized data are centred at \_\_\_\_\_ and have units equal to standard deviations of the original data.

- a) 0
- b) 5
- c) 1
- d) 10

**Ans : a) 0**

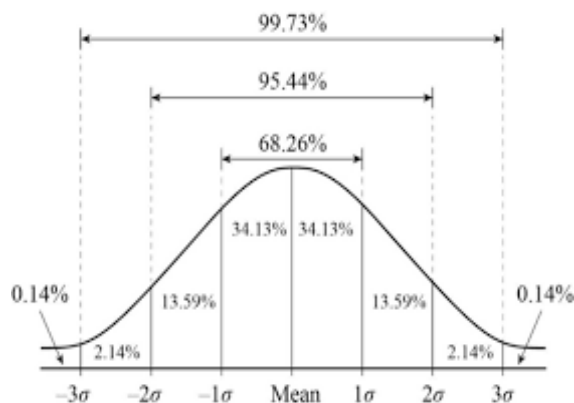
9. Which of the following statements is incorrect with respect to outliers?

- 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) Outliers cannot conform to the regression relationship**

### 10)What do you understand by the term Normal Distribution?

- Ans :-1)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.
- 2)In graph form, normal distribution will appear as a bell curve
- 3)In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3.
- 4)Normal distributions are symmetrical , but not all symmetrical distributions are normal.
- 5)In reality , most pricing distributions are not perfectly normal.
- 6)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.



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

Ans :-Handling missing data is very important in the model building process. The method to detect missing data is understanding data thoroughly. The basic method used to handle missing data is by using fillna() method in Python. Advanced Imputation Techniques:

- 1)Simple imputation
- 2)Knn imputation
- 3)Iterative imputation

### 12)What is A/B testing?

Ans :-

A/B testing (also known as bucket testing or split-run testing) is a user experience research methodology. ... A/B testing is a way to compare two versions of a single variable, typically by testing a subject's response to variant A against variant B, and determining which of the two variants are more effective

### 13) Is mean imputation of missing data acceptable practice?

Ans :-

Yes, it can be accepted if NaN values are lesser and it is a continuous data or else there will be a data leakage

**14. What is linear regression in statistics?**

Ans :-

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they—indicated by the magnitude and sign of the beta estimates—impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula  $y = mx + c$ , where  $y$  = estimated dependent variable score,  $c$  = constant,  $m$  = regression coefficient, and  $x$  = score on the independent variable

**15. What are the various branches of statistics?**

Branches of statistics

1. Descriptive statistics.
2. Inferential statistics.