

# WORKSHEET

## 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) 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?

- 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) Modelling event/time data
- b) Modelling bounded count data
- c) Modelling contingency tables
- d) All of the mentioned

Ans: b) Modelling 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 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) Outliers cannot conform to the regression relationship

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

**10. What do you understand by the term Normal Distribution?**

**Ans:** A normal distribution is the bell-shaped frequency distribution curve of a continuous random variable with mean =0 and standard deviation =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.

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

**Ans:** Following are the ways to handle missing data:

1. Deleting the entire row
2. Deleting the entire column.
3. Replacing Nans with an arbitrary value.

**Imputation techniques:**

1. Replacing Nans with the mean.
2. Replacing Nans with the mode.
3. Replacing Nans with the median.

**12. What is A/B testing?**

**Ans:** A/B testing is a powerful tool in machine learning. It allows us to compare two versions of a model or feature to determine which performs better. By splitting the audience into two groups and presenting each with a different version, we can measure the effect of each version on the desired outcome.

**13. Is mean imputation of missing data acceptable practice?**

**Ans:** No, Mean imputation of missing data is not acceptable practice because mean decreases the variance of our data while increasing bias and Leads to an underestimate of the standard deviation. As a result of the reduced variance, the model becomes less accurate.

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

**Ans:** Linear regression is a data analysis technique that predicts the value of unknown data by using another related and known data value. 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:** The two branches of statistics are: **Descriptive** and **Inferential**.

**Descriptive statistics** deals with the presentation and collection of data.

**Inferential statistics** involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics.