

## 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

**Answer: (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

**Answer: (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

**Answer: (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

**Answer: (C) The square of a standard normal random variable follows what is called chi-squared distribution**

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

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

**Answer: (C) Poisson**

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

- a) True
- b) False

**Answer: (A) True**

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

**Answer: (D) None of the mentioned**

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

**Answer: (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

**Answer: (B) Outliers can be the result of spurious or real processes**

**Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.**

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

**Answer: Normal distribution can also be called as the bell curve or Gaussian distribution, illustrates like how numbers are grouped around an average value**

- The majority of numbers are near the average with fewer numbers spread out toward the ends.
- It's characterized by its average value and how much the numbers vary, known as the standard deviation.
- Many natural measurements, such as heights or test scores, often match this pattern, which makes it a valuable tool in statistics.

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

Answer: Missing data can be handled in below mentioned ways

1. Mean/Median/Mode Imputation
2. Forward/Backward Fill
3. Remove Missing Data
4. K-Nearest Neighbors (KNN) Imputation
5. Multiple Imputation
6. Interpolation
7. Regression Imputation
8. Algorithms that Support Missing Data

12. What is A/B testing?

Answer: A/B testing is used to see which of two versions is better. A group is split in half: one half uses version A, the other half uses version B. By looking at both results, you see which one is best.

13. Is mean imputation of missing data acceptable practice?

Answer: Using the average value to fill in missing data is common, but whether it's okay depends on the data:

- **Advantages:** It's easy and keeps the sample size the same. It works well when missing data are randomly spread out.
- **Considerations:** However, using the average can make things seem more uniform than they really are if the missing data aren't random. This can affect how the data appears to be connected.
- **Alternatives:** Depending on what's needed, other methods like using the middle value, most common value, or more complex methods like using multiple guesses or patterns might be better. These can handle missing data more accurately by considering how things are spread out and connected.

14. What is linear regression in statistics?

Answer: Linear regression in statistics showcases how two things relate. It assumes a straight-line link between something we can change (X) and something we measure (Y). The goal is to find the best line that fits the points on a graph. This helps predict Y when we know X.

15. What are the various branches of statistics?

Answer: Here are the list of main branches:

1. Descriptive Statistics
2. Probability Theory
3. Social Statistics
4. Inferential Statistics
5. Biostatistics
6. Econometrics
7. Statistical Computing