

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.**

1. a) True
2. b) False

Answer:

1. a

**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?**

1. a) Central Limit Theorem
2. b) Central Mean Theorem

3. c) Centroid Limit Theorem
4. d) All of the mentioned

Answer:

1. a

3. Which of the following is incorrect with respect to use of Poisson distribution?

1. a) Modelling event/time data
2. b) Modelling bounded count data
3. c) Modelling contingency tables
4. d) All of the mentioned

Answer:

2. b

4. Point out the correct statement.

1. a) The exponent of a normally distributed random variables follows what is called the log-normal distribution
2. b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent

3. c) The square of a standard normal random variable follows what is called chi-squared distribution
4. d) All of the mentioned

Answer:

3. c

5. _____ random variables are used to model rates.

1. a) Empirical
2. b) Binomial
3. c) Poisson
4. d) All of the mentioned

Answer:

3. c

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

1. a) True
2. b) False

Answer:

2. b

7. Which of the following testing is concerned with making decisions using data?

1. a) Probability
2. b) Hypothesis
3. c) Causal
4. d) None of the mentioned

Answer:

2. b

8. Normalised data are centred at _____ and have units equal to standard deviations of the original data.

1. a) 0
2. b) 5
3. c) 1
4. d) 10

Answer:

1. a

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

1. a) Outliers can have varying degrees of influence
2. b) Outliers can be the result of spurious or real processes
3. c) Outliers cannot conform to the regression relationship
4. d) None of the mentioned

Answer:

3. c

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

A normal distribution is a way to describe data that clusters around a central point, like the average. It looks like a bell curve, where most of the data points are close to the middle, and fewer points are further

away. It's symmetric, meaning the left side is a mirror image of the right side.

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

When data is missing, we can handle it by either ignoring the missing parts or filling them in. Some techniques for filling in missing data, called imputation, include:

- **Mean/Median Imputation:** Replace missing values with the average (mean) or middle (median) value.
- **Mode Imputation:** Use the most frequent value (mode) for filling in missing data.
- **K-Nearest Neighbors (KNN) Imputation:** Use values from data points that are similar to the one with missing data.

12. What is A/B testing?

A/B testing is a method to compare two versions of something to see which one works better. You split your audience into two groups: Group A sees the original version, and Group B sees a new version. By comparing the results, you can decide which version is more effective.

13. Is mean imputation of missing data acceptable practice?

Mean imputation, where missing values are replaced with the average value, is simple and easy. However, it's not always the best method because it can distort the data. More advanced techniques might give better results, but mean imputation is acceptable if done carefully.

14. What is linear regression in statistics?

Linear regression is a technique used to understand the relationship between two variables. It helps to find a straight line that best fits the data points. This line can then be used to predict the value of one variable based on the value of another.

15. What are the various branches of statistics?

Statistics has several branches, including:

- **Descriptive Statistics:** Summarising and describing the main features of a dataset using measures like mean, median, and mode.
- **Inferential Statistics:** Making predictions or inferences about a larger population based on a sample of data.

- **Exploratory Data Analysis (EDA):** Analysing data sets to find patterns and insights.
- **Applied Statistics:** Applying statistical methods to solve real-world problems in fields like medicine, business, and engineering.