

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

- a) True
- b) False

Ans. b) False

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) Hypothesis

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) 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

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?

Ans. In a normal distribution, data is symmetrically distributed with no skew. When plotted on a graph, the data follows a bell shape, with most values clustering around a central region and tapering off as they go further away from the center. Normal distributions are also called Gaussian distributions or bell curves because of their shape.

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

Ans. Multiple imputation is considered a good approach for data sets with a large amount of missing data. Instead of substituting a single value for each missing data point, the missing values are exchanged for values that encompass the natural variability and uncertainty of the right values.

12. What is A/B testing?

Ans. [A/B testing](#), also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

13. Is mean imputation of missing data acceptable practice?

Mean Imputation: the missing value is replaced for the mean of all data formed within a specific cell or class. This technique isn't a good idea because the mean is sensitive to data noise like outliers. Tavares and Soares [2018] compare some other techniques with mean and conclude that mean is not a good idea.

14. What is linear regression in statistics?

Ans. Linear regression is used to predict the relationship between two variables by applying a linear equation to observed data. There are two types of variable, one variable is called an independent variable, and the other is a dependent variable.

15. What are the various branches of statistics?

Ans. The two main branches of statistics are [descriptive statistics](#) and [inferential statistics](#). Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.

[Descriptive statistics](#) deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid [biases](#) that are so easy to creep into the [experiment](#).

[Inferential statistics](#), as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.