

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:-** Didn't understand the question

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) 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

**Answer:-** a)

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:-** a) Probability

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?

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:-** d) None of the mentioned

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

**Answer:-** 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?

**Answer:-** Missing data (or missing values) is defined as the data value that is not stored for a variable in the observation of interest. The problem of missing data is relatively common in almost all research and can have a significant effect on the conclusions that can be drawn from the data.

The imputation technique:-

a-Missing completely at random:- Missing completely at random (MCAR) is defined as when the probability that the data are missing is not related to either the specific value which is supposed to be obtained or the set of observed responses. MCAR is an ideal but unreasonable assumption for many studies performed in the field of anaesthesiology.

b-Missing at random:- Missing at random (MAR) is a more realistic assumption for the studies performed in the anaesthetic field. Data are regarded to be MAR when the probability that the

responses are missing depends on the set of observed responses, but is not related to the specific missing values which is expected to be obtained.

c- Missing not at random:- If the characters of the data do not meet those of MCAR or MAR, then they fall into the category of missing not at random (MNAR).

12. What is A/B testing?

**Answer:-** A/B testing is 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 drive business metrics.

13. Is mean imputation of missing data acceptable practice?

**Answer:-** Bad practice in general. If just estimating means: mean imputation preserves the mean of the observed data. Leads to an underestimate of the standard deviation. Distorts relationships between variables by “pulling” estimates of the correlation toward zero

14. What is linear regression in statistics?

**Answer:-** 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  $y = c + b \cdot x$ , where  $y$  = estimated dependent variable score,  $c$  = constant,  $b$  = regression coefficient, and  $x$  = score on the independent variable.

15. What are the various branches of statistics?

**Answer:-** Statistics may be divided into two main branches:

(1) Descriptive Statistics (2) Inferential Statistics

### **(1) Descriptive Statistics**

Descriptive statistics deals with the collection of data, its presentation in various forms, such as tables, graphs and diagrams and finding averages and other measures which would describe the data.

**For example:** Industrial statistics, population statistics, trade statistics, etc. Businessmen make use of descriptive statistics in presenting their annual reports, final accounts, and bank statements.

## **(2) Inferential Statistics**

Inferential statistics deals with techniques used for the analysis of data, making estimates and drawing conclusions from limited information obtained through sampling and testing the reliability of the estimates.

**For example:** Suppose we want to have an idea about the percentage of the illiterate population of our country. We take a sample from the population and find the proportion of illiterate individuals in the sample. With the help of probability, this sample proportion enables us to make some inferences about the population proportion. This study belongs to inferential statistics.