# Objective Questions on Statistics

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 above

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 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

# Write brief answer of below questions

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

Ans: Normal distribution or Gaussian distribution or Bell curve is a statistical term used to evaluate a probability distribution of a random variable.

In this, the mean, median, and mode are equal, and the data is symmetrically distributed around the mean.

The normal distribution is a particularly useful model for data analysis and statistical inference because of its well-understood properties and mathematical tractability.

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

Ans: As we know that every data is important in data processing. so, we must have to handle the missing data because it can be possible that the missing data can affect the accuracy of the results.

Therefore, some of the technique are as follows to handle the missing data -

(a) Remove the Row : In this, usually we remove the entire row in which data is missing and do the analysis with rest of the data in table. This technique will be useful when the missing data is very small and that doesn't going to affect the overall analysis. Although, this is not used frequently because it can be possible that it lead to loss of important information.

(b) Imputation : In this technique, we use to substitute the missing data by predicting from available data.

It can be done in two ways i.e Univariant and Multivariant.

Now, in Univariant there can be different techniques used based on column where data is missing. i.e Numerical or Categorical.

In Numerical, we can fill the missing data by mean or median of the available data and in Categorical, we can fill the missing data by calculating mode of available data or can write missing.

In my opinion, Imputation of data will be better technique than deletion of row because it might possible that we can lose the important informations.

12. What is A/B testing?

Ans: It is a statistical technique used in product development to compare the performance of two versions of same thing to find out which is more effective.

Ex- let's take an example of of a Manufacturing company's webpage in which the contact button is on the bottom that is responsible for generating sales. Now he analyst think that if the contact button will be on the top, it will lead the sales.

So, to evaluate the which option is more productive, we run both versions parallelly and users to both randomly. And then, measure the how visiters response on both versions and which version lead to

better sales.

13. Is mean imputation of missing data acceptable practice?

Ans: well, it is totally depend on size of missing data and accuracy of the Imputation technique.

If the missing data is not related to other available data then the mean Imputation can results in incorrect statistical inference .

It can be only acceptable practice when the amount of missing dais very small accuracy of Imputation technique are met. Before choosing any Imputation technique, must have to understand the nature and amount of missing data.

14. What is linear regression in statistics?

Ans: Linear regression is a statistical method used to build the relationship between a dependent variable or outcomes and independent attributes or features. There can be one or more independent attributes.

Two type of linear regression -

(a) Simple linear regression: In this mode of linear regression, there is only one independent attribute or feature and the graph between the independent and dependent variables are linear or straight line.

i.e $y = mx + c$

Here, 'y' is dependent variable, 'x' is independent variable, m is slope of graph and 'c' is intercept.

(b) Multiple Linear Regression: In this mode of regression, there is more than one independent attributes or features and the graph between both independent and dependent variables are plane or hyperplane.

$y = b0 + b1x1 + b2x2 +.....+bmxn$

Here, bi = Regression Coefficient

xi = Independent variable

y = output variable

bi depend on the how much important is related independent attribute. If the particular independent attribute is playing most important role in calculating of output or dependent variable then the regression coefficient associated with this attribute must be highest and vice versa.

Linear regression is a powerful tool for understanding and quantifying the relationship between variables and making predictions based on data.

15. What are the various branches of statistics?

Ans: Basically, field of statistics is divided into two broad categories -

(a) Descriptive Statistics : In this mode of statistics, we perform the test on entire group or population and analyse the result.

Ex - Sales of any organisation in any fiscal year, Average score of cricketers in a series

(b) Inferential Statistics : In this mode of statistics, we perform the test on a sample from an entire population and after analysing the results implement to entire population.

Ex - Exit poll of any election