**Statistics**

**1. Bernoulli random variables take (only) the values 1 and 0.**

a) True b) False

**Answer:** 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 states that as the sample size increases the distribution become more normal.

**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:** D)All of the mentioned

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

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

**Answer:** C:poisson random variable are used to model rates.

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

a) True

b) False

**Answer:** True

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

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

**Answer:** Hypothesis testing is concerned with making decisions using data

**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

**Answer:** Normalized data are centered at 0 and have units equal to standard deviations of the original data

**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:** C (Outliers can conform to the regression relationship)

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

**Answer:** Normal Distribution is also called as Gaussian distribution.

In normal distribution mean it only work with continious data.

In normal distribution every event is independent of each other.

it is also called is bell curve or called as symetric distribution.

In normal distribution mean median and mode value are almost equal to each other.

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

**Answer:** There is two data imputation techniques to handle missing data:

Average imputation and common-point imputation.

Average imputation:

Average imputation is Simply calculate the mean of the observed values for that variable for all individuals who are non-missing.

Common-point imputation:

Common-point imputation uses the middle point or the most commonly chosen value. For example, on a five-point scale, the substitute value will be 3. Something to keep in mind when utilising this method is the three types of middle values: mean, median and mode, which is valid for numerical data (it should be noted that for non-numerical data only the median and mean are relevant).

**12. What is A/B testing?**

**Answer:** A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

For instance, let’s say you own a company and want to increase the sales of your product. Here, either you can use random experiments, or you can apply scientific and statistical methods. A/B testing is one of the most prominent and widely used statistical tools.

In the above scenario, you may divide the products into two parts – A and B. Here A will remain unchanged while you make significant changes in B’s packaging. Now, on the basis of the response from customer groups who used A and B respectively, you try to decide which is performing better.

**13. Is mean imputation of missing data acceptable practice?**

**Answer:** Mean imputation does not preserve the relationships among variables.

True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased.

**14. What is linear regression in statistics?**

**Answer:** Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task.Linear regression algorithm shows a linear relationship between a dependent (lables) and one or more independent (y) variables, hence called as linear regression.It includes the only one features and create the model to see the relationship between features and labels.This is called as SLR. It having feature(indipendent variable) and label(dependent variable)

**15. What are the various branches of statistics?**

**Answer:** The two main branches of statistics are descriptive statistics and inferential statistics.

1.Discriptive:if we am able to describe any data then it is called as discriptive.it provides exact and accurate information.if you are not able to make decision from millions of dataset you take sample data from the millions of data and analysis the data.the sample data(small amount of data) from the millians of data is called as discriptive(basically discribe the small amount of data)

2.Inferential:if the data is too big to discribe,so we can pick the sample(randomly) data from the total of data size So here i picked the sample data from the whole data,discribe it and inferrence(refer) it to my whole dataset(Population).Inferential means if you not able to discribe the data due to data in millions so you take the sample from millions of data discribe it and inference to refer it to the millions of data(sample data).