# WORKSHEET



**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: 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?

1. Central Limit theorem

## Central Mean theorem

1. Centroid Limit theorem
2. AII of the above mentioned

Ans : a) Central Limit theorem

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

1. Modeling event/time data

## Modeling bounded count data

1. Modeling contingency tables
2. AII of the mentioned

Ans: b)Modelling bounded count data

4 Point out the correct statement.

## The exponent of a normally distributed random variables follows what is called the log- normal

distribution

## Sums of normally distributed random variables are again normally distributed even if the variables are dependent

1. The square of a Standard normal random variable follows what is Called chi-squared distribution

# All of the mentioned

Ans: d) All of the mentioned

## random variables as need

* 1. Empirical

## Binomial

* 1. Poisson
  2. All of the mentioned

Ans: c)Poisson

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

1. True
2. False

Ans: b) False

1. Which of the following testing is concerned with making decisions using data?
   1. Probability
   2. Hypothesis
   3. Casual
   4. None of the mentioned

Ans: b) null hypothesis

 4. Normalized data are centered at and have units equal to standard deviations of the original data.

a)0 b)5

## 1 d) 10

Ans: a)0

1. Which of the following statement is incorrect with respect to outliers.?
   1. Outliers can have varying degrees of influence
   2. Outliers can be the result of spurious or real processes
   3. Outliers cannot conform to the regression relationship
   4. None of the mentioned

Ans: c) Outliers cannot conform to the regression relationship

# WORKSHEET



**Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.**

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

Ans : Normal Distribution is a term that comes time and again in the field of Statistics. It examines the data with respect to it’s point near the mean. Normal Distribution is represented by a bell shaped curve diagram. It says that the frequency of occurrence of an event is more at the mean and it lessens, when the data is away from the mean. Normal distribution plays a significant part in Data Science, as far as Hypothesis Testing is concerned.

A Normal Distribution is unique when compared to other distributions is that a normal distribution retains it’s bell shape. Some of the important characteristics of a normal distribution are:-

* The end result of the sum of two normal distributions is also a normal distribution.
* The end result of a product of two normal distributions is also a normal distribution.
* The Fourier Transformation of a normal distribution normal.

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

* **Ignore the records with missing values.**

Many tools ignore records with missing values. When the percentage of records with missing values is small, we could ignore those records.

* **Substitute a value such as mean**.

When the percentage is large and also when it makes sense to do something to avoid bias modeling results, substituting a value (e.g. mean, median) is a commonly used way. But this method could cause bias distribution and variance. That’s where the following imputation methods come in.

* **Predict missing values.**

Depending on the type of the imputed variable (i.e. continuous, ordinal, nominal) and missing data pattern (i.e. monotone, non-monotone), below are a few commonly used models. If you plan to do it in SAS, there are SAS codes that you can write to identify the missing data pattern.

1. What is A/B testing?

Ans: A/B testing is to compare two versions of webpage with each other and check which one performs better. It allows you to make more out of your existing traffic so you need to constantly test and optimize your web page to bring best result in conversions.

1. Is mean imputation of missing data acceptable practice?

Ans : True, imputing the **mean** preserves the mean of the observed **d**ata. So if the data are missing completely at random, the estimate of the mean remains unbiased. That's a good thing. ... Since most research studies are interested in the relationship among variables**,**mean imputation is not a good solution.

1. What is linear regression in statistics?

Ans: Linear regression is used when the response variable(output) is a continuous variable. It will be a quantitative variable and take real or integer values

15.What are the various branches of statistics?

Ans: 1) Statistics

2)Inferential Statistics