

Assignment 1(Statistics Worksheet 1)

1)Bernoulli random variable take(only) the values 1 and 0

Ans : True

2)Which of the following theorem states that the distribution of average of iid variables properly normalised, becomes that of a standard normal as the sample size increases

Ans : Central limit theorem

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

Ans : Modeling bounded count data

4)Point out the correct statement.

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

5)_____ random variables are used to model rates

Ans : Poisson

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

Ans : False

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

Ans : Hypothesis

8)Normalized data are centered at _____and have units equal to standard deviations of the original data.

Ans : 0

9) Which of the following statement is incorrect with respect to outliers?

Ans : Outliers cannot conform to the regression relationship

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

Ans : Normal distributions are symmetrical, but not all symmetrical distributions are normal. Many naturally-occurring phenomena tend to approximate the normal distribution. In finance, most pricing distributions are not, however, perfectly normal. The normal distribution is the most common type of distribution assumed in technical stock market analysis and in other types of statistical analyses. The standard normal distribution has two parameters: the mean

and the standard deviation. The normal distribution model is important in statistics and is key to the Central Limit Theorem (CLT).

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

12)What is A/B testing?

ANS :A/B testing (also known as split testing or bucket testing) is a method of comparing two versions of a webpage or app against each other to determine which one performs better. A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal. Running an A/B test that directly compares a variation against a current experience lets you ask focused questions about changes to your website or app and then collect data about the impact of that change.

13)Is mean imputation of missing data acceptable practice?

14)What is linear regression in statistics?

Ans : Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model. Before attempting to fit a linear model to observed data, a modeler should first determine whether or not there is a relationship between the variables of interest.

15) What are the various branches of statistics

Ans : Statistics have majorly categorised into two types:

1)Descriptive statistics

2)Inferential statistics

Descriptive Statistics

In this type of statistics, the data is summarised through the given observations. The summarisation is one from a sample of population using parameters such as the mean or standard deviation.

Descriptive statistics is a way to organise, represent and describe a collection of data using tables, graphs, and summary measures. For example, the collection of people in a city using the internet or using Television.

Descriptive statistics are also categorised into four different categories:

Measure of frequency

Measure of dispersion

Measure of central tendency

Measure of position

The frequency measurement displays the number of times a particular data occurs. Range, Variance, Standard Deviation are measures of dispersion. It identifies the spread of data.

Central tendencies are the mean, median and mode of the data. And the measure of position describes the percentile and quartile ranks.

Inferential Statistics

This type of statistics is used to interpret the meaning of Descriptive statistics. That means once the data has been collected, analysed and summarised then we use these stats to describe the meaning of the collected data. Or we can say, it is used to draw conclusions from the data that depends on random variations such as observational errors, sampling variation, etc.

Inferential Statistics is a method that allows us to use information collected from a sample to make decisions, predictions or inferences from a population. It grants us permission to give statements that goes beyond the available data or information. For example, deriving estimates from hypothetical research.