

STATISTICS WORKSHEET-1

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

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

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

a) Modelling bounded count data

4. Point out the correct statement

d) All of the mentioned

5. _____ random variables are used to model rates

b) Poisson

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

a) True

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

b) Hypothesis

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

a) 0

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

c) Outliers cannot conform to the regression relationship

Q10 and Q15 are subjective answer type questions

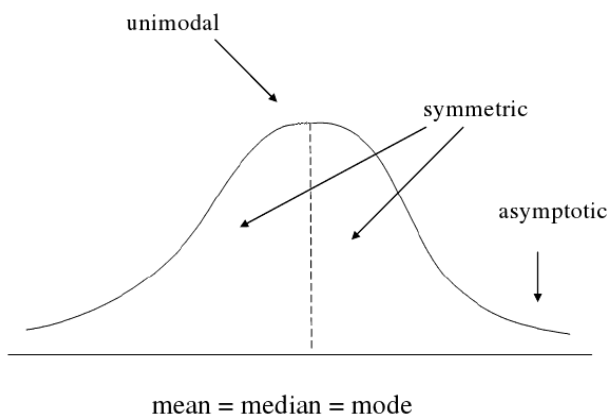
10. What do you understand by the term Normal Distribution?

The normal distribution, also known as the Gaussian distribution, is the most important probability distribution in statistics for independent, random variables.

A **normal distribution** is a bell-shaped frequency distribution curve. Most of the data values in a normal distribution tend to cluster around the mean. The further a data point is from the mean, the less likely it is to occur.

As with any probability distribution, the normal distribution describes how the values of a variable are distributed. It is the most important probability distribution in statistics because it accurately describes the distribution of values for many natural phenomena.

Characteristics of Normal Distribution



Here, we see the four characteristics of a normal distribution. Normal distributions are **symmetric**, **unimodal**, and **asymptotic**, and the **mean**, **median**, and **mode** are all equal.

A normal distribution is perfectly symmetrical around its center. That is, the right side of the center is a mirror image of the left side. There is also only one mode, or peak, in a normal

The **center of a normal distribution** is located at its peak, and 50% of the data lies above the mean, while 50% lies below. It follows that the mean, median, and mode are all equal in a normal distributio

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

Missing data can be dealt with in a **variety of ways**. We use **imputation**. Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.

imputation techniques

- Mean imputation.
- Substitution.
- Hot deck imputation.
- Cold deck imputation.
- Regression imputation.

12. What is A/B testing?

A/B testing is basically **statistical hypothesis testing, or, in other words, statistical inference**. It is an analytical method for making decisions that estimates population parameters based on sample statistics.

A/B testing is one of the most popular controlled experiments used to optimize web marketing strategies. It allows decision makers to choose the best design for a website by looking at the analytics results obtained with two possible alternatives A and B.

A/B testing is **not limited by web pages only**, you can A/B test your emails, popups, sign up forms, apps and more. Nowadays, most MarTech software comes with an A/B testing function built-in.

13. Is mean imputation of missing data acceptable practice?

The process of replacing null values in a data collection with the data's mean is known as **mean imputation**. It is a popular solution to missing data, despite its drawbacks. Mainly because it's easy.

But **that doesn't make it a good solution**, and it may not help you find relationships with strong parameter estimates. Even if they exist in the population. On the other hand, there **are many alternatives to mean imputation** that provide much more accurate estimates and standard errors,

1. Mean imputation does **not preserve the relationships among variables**.
2. A second reason is applies to any type of single imputation. Any statistic that uses the imputed data will have a **standard error that's too low**.

14. What is linear regression in statistics?

Linear regression strives to show the **relationship between two variables** by applying a linear equation to observed data. One variable is supposed to be an **independent variable**, and the other is to be a **dependent variable**. Linear regression analysis is the most widely used of all statistical techniques.

The use of linear regression model is important for the following reasons:

- a. **Descriptive** – It helps in analyzing the strength of the association between the outcome (dependent variable) and predictor variables
- b. **Adjustment** - It adjusts for the effect of covariates or the confounders
- c. **Predictors** – It helps in estimating the important risk factors that affect the dependent variable
- d. **Extent of prediction** – It helps in analyzing the extent of change in the independent variable by one “unit” would affect the dependent variable
- e. **Prediction** – It helps in quantifying the new cases.

15. What are the various branches of statistics?

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

Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.

Descriptive Statistics

Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis , Descriptive statistics are used to do various kinds of analysis on different studies.

Example of Descriptive Statistics The average score of the college students in the math test .The average age of the people who voted for the winning candidate in the last election .The average length of the statistics book.

Inferential Statistics

Inference statistics are techniques that enable statisticians to use the **information collected** from the sample to conclude, bring decisions, or predict a defined population.

Inference statistics often speak in terms of **probability** by using descriptive statistics. Besides, a statistician uses these techniques for data analysis, drafting, and making conclusions from limited information.