**Statistics worksheet assignment solution**

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)**

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)**

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

a) True b) False

**answer:- 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

**answer:- 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

**answer:-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 ment

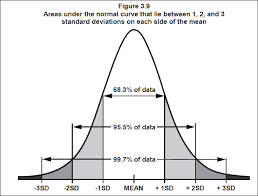
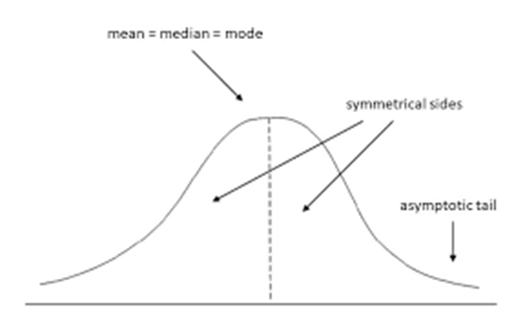
**answer:- c (Outliers cannot conform to the regression relationship)**

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

Ans. A  [normal distribution](https://www.statisticshowto.com/probability-and-statistics/normal-distributions/), sometimes called the bell curve, is a distribution that occurs naturally in many situations. For example, the bell curve is seen in tests like the SAT and GRE. The bulk of students will score the [average](https://calculushowto.com/average-value-of-a-function/#def)(C), while smaller numbers of students will score a B or D. An even smaller percentage of students score an F or an A. This creates a distribution that resembles a bell (hence the nickname). The bell curve is [symmetrical](https://www.calculushowto.com/symmetry-of-a-function/). Half of the data will fall to the left of the [mean](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-median-mode/#mean); half will fall to the right.  
Many groups follow this type of pattern. That’s why it’s widely used in business, statistics and in government bodies like the [FDA](https://www.fda.gov/default.htm):

* Heights of people.
* Measurement errors.
* Blood pressure.
* Points on a test.
* IQ scores.
* Salaries.

The [empirical rule](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/empirical-rule-2/) tells you what percentage of your data falls within a certain number of [standard deviations](https://www.statisticshowto.com/probability-and-statistics/standard-deviation/) from the [mean](https://www.statisticshowto.com/mean):  
• 68% of the data falls within one [standard deviation](https://www.statisticshowto.com/probability-and-statistics/standard-deviation/) of the [mean](https://www.statisticshowto.com/mean).  
• 95% of the data falls within two [standard deviations](https://www.statisticshowto.com/probability-and-statistics/standard-deviation/) of the [mean](https://www.statisticshowto.com/mean).  
• 99.7% of the data falls within three [standard deviations](https://www.statisticshowto.com/probability-and-statistics/standard-deviation/) of the [mean](https://www.statisticshowto.com/mean).

Properties of a normal distribution

* The [mean, mode and median](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-median-mode/) are all equal.
* The curve is symmetric at the center (i.e. around the mean, μ).
* Exactly half of the values are to the left of center and exactly half the values are to the right.
* The total area under the curve is 1.

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

* Mean imputation. Simply calculate the mean of the observed values for that variable for all individuals who are non-missing.
* Mode, Median
* Substitution.
* Hot deck imputation.
* Cold deck imputation..
* Regression imputation.
* Stochastic regression imputation..
* Interpolation and extrapolation.

For numerical type data using Mean imputation technique and for categorical type data using Mode or Median.

Q.12 what is A\B testing?

Ans.

* . Two-sample hypothesis testing
* Randomized experiments with two variants: A and B
* A: control; B: variation
* User-experience design: identify changes to web pages that increase clicks on a banner
* Current website: control; NULL hypothesis
* New version: variation; alternative hypothesis

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

* Bad practice in general
* If just estimating means: mean imputation preserves the mean of the observed data
* Leads to an underestimate of the standard deviation
* Distorts relationships between variables by “pulling” estimates of the correlation toward zero

Q.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. A linear regression line has an equation of the form Y = a + bX, where X is the explanatory variable and Y is the dependent variable.

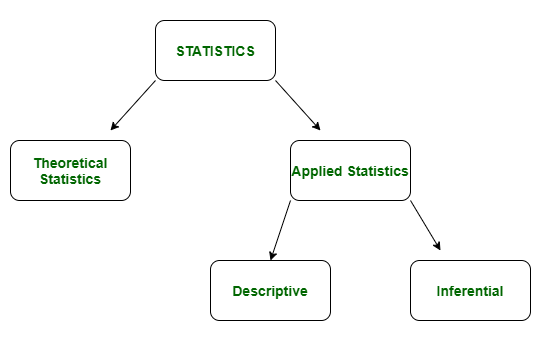
Linear regression is the next step up after correlation. It is used when we want to predict the value of a variable based on the value of another variable. The variable we want to predict is called the dependent variable (or sometimes, the outcome variable).

Q.15 What are the various branches of statistics?

Ans. [Statistics](https://www.geeksforgeeks.org/introduction-of-statistics-and-its-types/) is a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. It is basically a collection of quantitative data.

**Types of Statistics –**

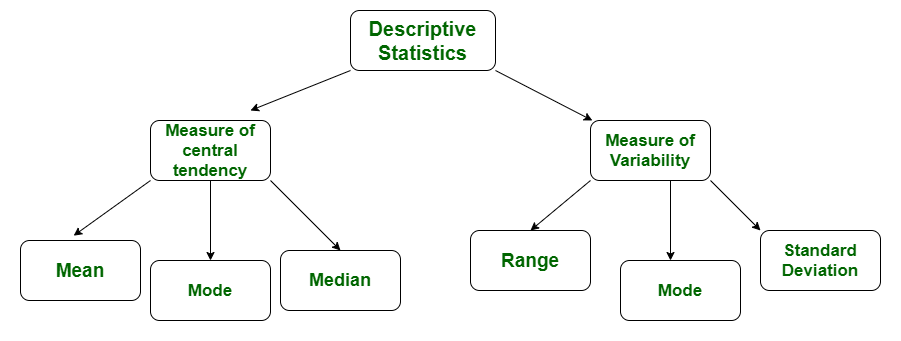
* Theoretical Statistics
* Applied Statistics



**1.**[Descriptive Statistics](https://www.geeksforgeeks.org/descriptive-statistic/)**:**  
Descriptive statistics is a term given to the analysis of data that helps to describe, show and summarize data in a meaningful way. It is a simple way to describe our data. Descriptive statistics is very important to present our raw data ineffective/meaningful way using numerical calculations or graphs or tables. This type of statistics is applied on already known data.

**Types of Descriptive Statistics –**

* Measure of Central Tendency
* Measure of Variability



[2. Inferential Statistics](https://www.geeksforgeeks.org/introduction-of-statistics-and-its-types/)**:**  
In inferential statistics predictions are made by taking any group of data in which you are interested. It can be defined as a random sample of data taken from a population to describe and make inference about the population. Any group of data which includes all the data you are interested is known as population. It basically allows you to make predictions by taking a small sample instead of working on whole population.

