

## STATISTICS 2020 solved short

### 01 : What is meant by statistics ? (2020, 2019)

**Ans.** Statistics is a branch of applied mathematics that involves the collection, description, analysis, and inference of conclusions from quantitative data.

An example of statistics is a report of numbers saying how many followers of each religion there are in a particular country.

### 02 : What is difference between simple and composite hypothesis ?

Simple Hypothesis	Composite Hypothesis
If a set contains a single element (i.e., a single value for the parameter), then we have a simple hypothesis.	When a set contains more than one parameter value, then the hypothesis is called a composite hypothesis, because it involves more than one model.

### 03 : Define Cluster sampling. Also give example.

**Ans.** Cluster sampling is a probability sampling method in which you divide a population into clusters, such as districts or schools, and then randomly select some of these clusters as your sample. The clusters should ideally each be mini-representations of the population as a whole.

### 04 : What is secondary data. Also give example.

**Ans.** Secondary data is the data which has already been collected by someone other than the primary user, may be sorted, tabulated and has undergone a statistical treatment. Common sources of secondary data for include censuses, information collected by government departments, organizational records and data that was originally collected for other research purposes.

### 05 : Write any two properties of correlation coefficient.

**Ans.** Correlation Coefficient Properties include,

- 1) Correlation coefficient remains in the same measurement as in which the two variables are.
- 2) The sign which correlations of coefficient have will always be the same as the variance.
- 3) The numerical value of correlation of coefficient will be in between -1 to + 1.

**06 : Find Median: -1, -2, 1, 0, 3, 3, 4, 3, 2, 6.**

**Sol.**

Arranging the numbers in ascending order : -2,-1,0,1,2,3,3,3,4,6

Here,  $n = 10$  which is even

**Median** =  $\frac{1}{2} ( (n/2)^{\text{th}} \text{ term} + ((n/2) + 1)^{\text{th}} \text{ term} )$

$$= \frac{1}{2} ((10/2)^{\text{th}} \text{ term} + ((10/2) + 1)^{\text{th}} \text{ term})$$

$$= \frac{1}{2} (5^{\text{th}} \text{ term} + (5 + 1)^{\text{th}} \text{ term})$$

$$= \frac{1}{2} (5^{\text{th}} \text{ term} + 6^{\text{th}} \text{ term})$$

$$= \frac{1}{2} (2 + 3)$$

$$= \frac{1}{2} \times 5$$

$$= 2.5$$

Hence the median is **2.5**.

**07 : What is standard normal distribution ?**

**Ans.** The standard normal distribution is a special case of the normal distribution . It is the distribution that occurs when a normal random variable has a mean of zero and a standard deviation of one.

The standard normal distribution is centered at zero and the degree to which a given measurement deviates from the mean is given by the standard deviation.

**08 : What is hypothesis testing ? (2020,2018)**

**Ans.** Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. It is used to assess the plausibility



of a hypothesis by using sample data. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis.

**09 : Define non-parametric data.**

**Ans.** Data that does not fit a known or well-understood distribution is referred to as nonparametric data. It means that data is not assumed to come from prescribed models that are determined by a small number of parameters.

Examples of such models include the normal distribution model and the linear regression model.

**10 : What is ratio and interval scale ?**

**Ans.** The interval scale and ratio scale are variable measurement scales. They offer a quantitative definition of the variable attributes.

The difference between interval and ratio scales comes from their ability to dip below zero. Interval scales hold no true zero and can represent values below zero. For example, you can measure temperature below 0 degrees Celsius, such as -10 degrees.

Ratio variables, on the other hand, never fall below zero. Height and weight measure from 0 and above, but never fall below it.

**11 : What is difference between quantitative and qualitative data ?**

Quantitative Data	Qualitative Data
Quantitative data can be counted, measured, and expressed using numbers.	Qualitative data is descriptive and conceptual. It can be categorized based on traits and characteristics.
Quantitative data is statistical and is typically structured in nature.	Qualitative data is non-statistical and is typically unstructured or semi-structured.
Examples of quantitative data include, <ul style="list-style-type: none"><li>• Weight</li><li>• Age</li></ul>	Examples of qualitative data are, <ul style="list-style-type: none"><li>• Color</li><li>• Material</li></ul>

• Height	
----------	--

## 12 : Define attribute ?

**Ans.** An attribute is a characteristic or feature that is measured for each observation (record) and can vary from one observation to another.

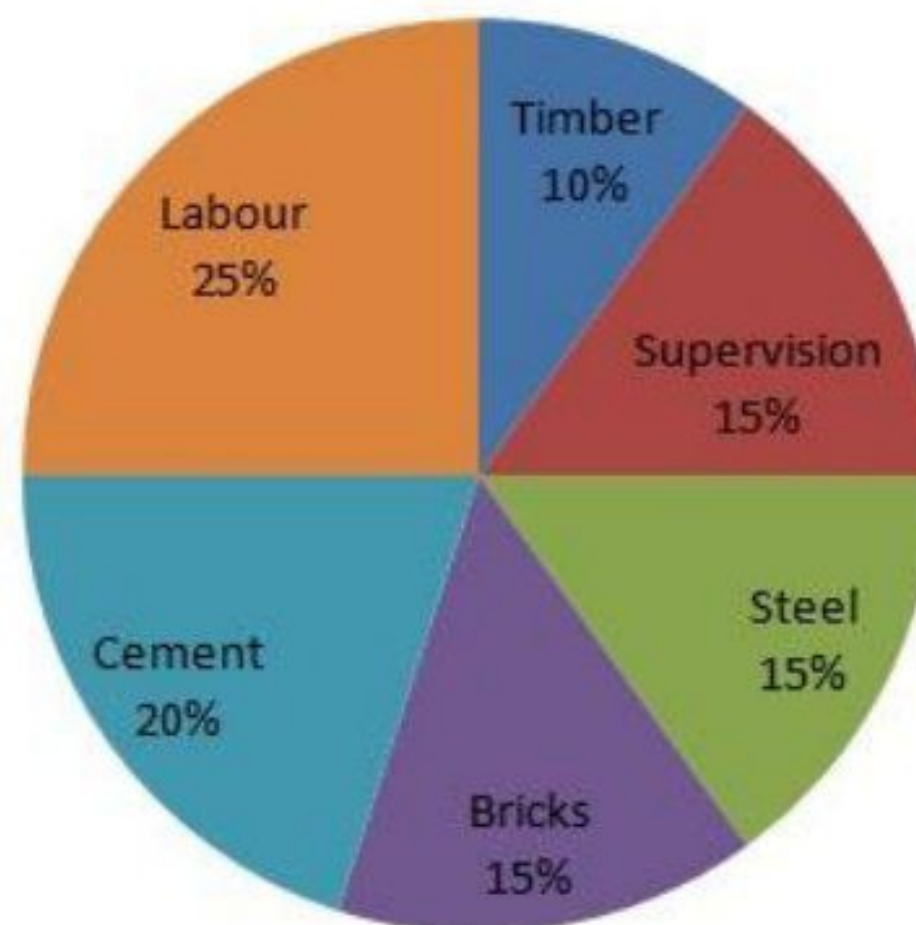
In statistics classifying data based on attributes or characteristic is known qualitative classification of data. Examples of attributes are region, caste etc.

## 13 : What is pie chart ?

**Ans.** A pie chart (or a circle chart) is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. In a pie chart, the arc length of each slice is proportional to the quantity it represents.

Here is an example of a pie chart.

### Cost of Construction of House



**14 : In a symmetrical distribution  $Q_1=140$  and Median=150. Find quartile deviation.**



**Sol.**

**Given data,**  $Q1 = 140$

Median = 150

Q.D = ?

**Formula :**  $Q.D. = (Q3 - Q1) / 2$

So, first we have to find Q3,

$$\text{Median} = \{1/2 (n+1)\}^{\text{th}} = 150$$

$$\begin{aligned} Q3 &= \{3/4 (n+1)\}^{\text{th}} \\ &= [3/2 \{1/2 (n+1)\}]^{\text{th}} \\ &= 3/2 (\text{Median}) \\ &= 3/2 (150) = \mathbf{225} \end{aligned}$$

Now, we have Q1 and Q3,

So,  $Q.D. = (Q3 - Q1) / 2$

$$Q.D = (225 - 140) / 2 = 85/2 = \mathbf{42.5}$$

**15 : Given  $\sum X = 180$ ,  $\sum X^2 = 6660$  and  $n = 5$ . Find coefficient of variation.**

**Sol.**

**Given data,**  $\sum x = 180$

$$\sum x^2 = 6660$$

$$n = 5$$

In order to find coefficient of variation, we must know standard deviation ( $\sigma$ ) and  $\bar{x}$ .

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$\sum x^2/n = 6660/5 = 1332$$

$$(\sum x/n)^2 = (180/5)^2 = (36)^2 = 1296$$

$$\text{So, } \sigma = \sqrt{(1332-1296)} = \sqrt{36} = \mathbf{6}$$

Similarly,  $\bar{x} = \Sigma x/n = 180/5 = 36$

Now putting values in formula,

**Coefficient of variation (C.V) =  $(\sigma/\bar{x}) \times 100\%$**

$$C.V = (6/36) \times 100\%$$

$$C.V = 16.66\%$$

### **16 : What is SPSS ?**

**Ans.** SPSS is short for *Statistical Package for the Social Sciences*. It is a software package that is used for the management and statistical analysis of social science data by various kinds of researchers.

## **STATISTICS 2019 solved short questions**

### **01 : Define statistics. (2020, 2019)**

**Ans.** Statistics is a branch of applied mathematics that involves the collection, description, analysis, and inference of conclusions from quantitative data.

An example of statistics is a report of numbers saying how many followers of each religion there are in a particular country.

### **02 : What is proper definition of Data ?**

**Ans.** Data are individual pieces of factual information, often numeric, that are collected through observation and used for the purpose of analysis. It is the raw information from which statistics are created.

### **03 : Write three statistical uses.**

**Ans.** Three uses of statistics are given below.

1. Statistics helps in providing a better understanding and accurate description of nature's phenomena.



2. Statistics helps in the proper and efficient planning of a statistical inquiry in any field of study.
3. Statistics helps in collecting appropriate quantitative data.

**04 : Define Variable.**

**Ans.** A variable is any characteristics, number, or quantity that can be measured or counted. A variable may also be called a data item.

Age, sex, business income and expenses, country of birth, capital expenditure, class grades, eye color and vehicle type examples of variables.

**05 : Write any four merits of mode.**

**09 : Write three advantages of mode.**

**Ans.** The merits of mode are given below.

1. It is easy to compute. In some cases, it can be computed merely by inspection.
2. Mode can be located on the graph also.
3. It is not affected by extreme values. It can be obtained even if the extreme values are not known.
4. It is the most observed data point and is easy to understand.

**06 : Differentiate between discrete and continuous variable ?**

<b>Discrete Variable</b>	<b>Continuous Variable</b>
A variable that takes only whole number as its value is called discrete variable.	A variable that can take any value, within a reasonable limit is called a continuous variable.
These variables increase in jumps or in complete numbers.	These variables assume a range of values or increase in fractions and not in jumps.
For example- Number of people in a family, number of students in a class, etc.	For example- age, height, weight, etc.

**07 : Define class boundaries.**

**Ans.** Class boundaries are the numbers used to separate classes. They are not part of the classes or the dataset. The size of the gap between classes is the difference between the upper class limit of one class and the lower class limit of the next class.

For example,

Class	Frequency	Class boundaries
180 – 199	2	179.5, 199.5
200 – 219	5	199.5, 219.5
220 – 239	12	219.5, 239.5

**08 : Define Arithmetic mean.**

**Ans.** In statistics, the arithmetic mean, or simply the mean or the average, is the sum of a collection of numbers divided by the count of numbers in the collection.

$$\bar{x} = \sum x/n$$

**10 : Define quartile.**

**Ans.** A quartile is a statistical term that describes a division of observations into four defined intervals based on the values of the data and how they compare to the entire set of observations.

**Formulas :**

1. *First Quartile* (Q1) =  $((n+1)/4)$ th
2. *Second Quartile* (Q2) =  $((n+1)/2)$ th
3. *Third Quartile* (Q3) =  $(3(n+1)/4)$ th

**11 : What is difference between parameter and statistics ?**

Parameter	Statistics
-----------	------------



It is a measure of a characteristic of an entire population, based on all the elements within that population.	It's a measure of characteristic saying something about a fraction (a sample) of the population under study.
A parameter is a fixed, unknown numerical value, describing the whole population.	The statistic is a known number and a variable which depends on the portion of the population.
20% of U.S. senators voted for a specific measure. Since there are only 100 senators, you can count what each of them voted.	50% of people living in the U.S. agree with the latest health care proposal. Researchers can't ask hundreds of millions of people if they agree, so they take samples or part of the population and calculate the rest.

## **12 : What is meant by sampling ?**

**Ans.** Sampling is a process used in statistical analysis in which a predetermined number of observations are taken from a larger population.

The methodology used to sample from a larger population depends on the type of analysis being performed, but it may include simple random sampling or systematic sampling.

## **13 : Define Sampling without replacement.**

**Ans.** In sampling without replacement, each sample unit of the population has only one chance to be selected in the sample.

For example, if one draws a simple random sample such that no unit occurs more than one time in the sample, the sample is drawn without replacement.

## **14 : What is one sample T test ?**

**Ans.** A one sample test of means compares the mean of a sample to a pre-specified value and tests for a deviation from that value.

For example, we might know that the average birth weight for white babies in the US is 3,410 grams and wish to compare the average birth weight of a sample of black babies to this value.

**15 : What is linear regression model ?**

**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.

**16 : Differentiate between linear regression and multiple regression ?**

<b>Linear Regression</b>	<b>Multiple Regression</b>
Linear regression models the relationship between a dependent variable and one or more explanatory variables using a linear function.	Multiple regression, on the other hand, is a broader class of regressions that encompasses linear and nonlinear regressions with multiple explanatory variables.

## **STATISTICS 2019 (B) solved short questions**

**1. Define population with examples.**

**Ans.** In statistics, a population is a set of similar items or events which is of interest for some question or experiment. A statistical population can be a group of existing objects (e.g. the set of all stars within the Milky Way galaxy) or a hypothetical and potentially infinite group of objects conceived as a generalization from experience (e.g. the set of all possible hands in a game of poker).

**2. What is inferential statistics ?**

**Ans.** Inferential statistics makes inferences and predictions about a population based on a sample of data taken from the population in question. The goal of inferential statistics is to reach conclusions that extend beyond the immediate data



alone. For instance, we use inferential statistics to try to infer from the sample data what the population might think.

### **3. Define frequency distribution.**

**Ans.** Frequency distribution in statistics is a representation that displays the number of observations within a given interval. The representation of a frequency distribution can be graphical or tabular so that it is easier to understand.

### **4. Write any four demerits of mode.**

**Ans.** Some demerits of Mode are given below,

- It is not well defined.
- It is not based on all the values.
- It is stable for large values so it will not be well defined if the data consists of a small number of values.
- It is not capable of further mathematical treatment.

### **5. Write three empirical relations between Mean, Median and Mode.**

**Ans.** Empirical relationship between mean, median and mode is:

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$

$$\Rightarrow \text{Mode} - \text{Mean} = 3 \text{ Median} - 2 \text{ Mean} - \text{Mean}$$

$$\Rightarrow \text{Mode} - \text{Mean} = 3 \text{ Median} - 3 \text{ Mean}$$

$$\Rightarrow \text{Mode} - \text{Mean} = 3 [\text{Median} - \text{Mean}]$$

### **6. Define Harmonic Mean and give any example.**

**Ans.** The harmonic mean is a type of numerical average. It is calculated by dividing the number of observations by the reciprocal of each number in the series. Thus, the harmonic mean is the reciprocal of the arithmetic mean of the reciprocals. The harmonic mean of 1, 4, and 4 is:  $3 \left( \frac{1}{1} + \frac{1}{4} + \frac{1}{4} \right) = 3 \frac{1}{1}$ .

### 7. Distinguish between primary and secondary data.

Primary Data	Secondary Data
Primary data is the type of data that is collected by researchers directly from main sources.	Secondary data is the data that has already been collected through primary sources and made readily available for researchers to use for their own research.
An example of primary data is the national census data collected by the government.	An example of secondary data is the data collected from online sources.

### 8. What is difference between statistics and statistic.

Statistics	Statistic
Statistics is a mathematical science concerned with data collection, presentation, analysis, and interpretation.	Statistic is a single item in a statistical study. It is A quantity calculated from the data in a sample, which characterizes an important aspect in the sample.
Classified facts of a numerical nature regarding any topic.	For instance, Mean or Standard Deviation.

### 9. Write down the properties of variance.

**Ans.** The variance,  $\text{var}(X)$  of a random variable  $X$  has the following properties.

- 1)  $\text{Var}(X + C) = \text{Var}(X)$ , where  $C$  is a constant.
- 2)  $\text{Var}(CX) = C^2 \cdot \text{Var}(X)$ , where  $C$  is a constant.
- 3)  $\text{Var}(aX + b) = a^2 \cdot \text{Var}(X)$ , where  $a$  and  $b$  are constants.

If  $X_1, X_2, \dots, X_n$  are  $n$  independent random variables, then

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n).$$

### 10. Define variable.



**Ans.** A variable is any characteristics, number, or quantity that can be measured or counted. A variable may also be called a data item. Age, sex, business income and expenses, country of birth, capital expenditure, class grades, eye colour and vehicle type are examples of variables.

**11. What is meant by sampling design ?**

**Ans.** A sample design is the framework, or road map, that serves as the basis for the selection of a survey sample and affects many other important aspects of a survey as well. The sample design provides the basic plan and methodology for selecting the sample.

**12. Define sampling with replacement.**

**Ans.** When a sampling unit is drawn from a finite population and is returned to that population, after its characteristic(s) have been recorded, before the next unit is drawn, the sampling is said to be “with replacement”

**13. What is null hypothesis ?**

**Ans.** A null hypothesis is a type of hypothesis used in statistics that proposes that there is no difference between certain characteristics of a population (or data-generating process).

**14. Define regression with examples.**

**Ans.** Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

**15. What is range of correlation ? Also write down the formula of correlation coefficient.**

**Ans.** Possible values of the correlation coefficient range from -1 to +1, with -1 indicating a perfectly linear negative, i.e., inverse, correlation (sloping

downward) and +1 indicating a perfectly linear positive correlation (sloping upward). A correlation coefficient close to 0 suggests little, if any, correlation.

Formula of correlation coefficient is given below,

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

## 16. What are the measurement scales ?

**Ans.** Measurement scale, in statistical analysis, the type of information provided by numbers. Each of the four scales (i.e., **nominal**, **ordinal**, **interval**, and **ratio**) provides a different type of information.

## STATISTICS 2018 solved short questions

### 01 : Differentiate between descriptive and inferential statistics.

Descriptive statistics	Inferential statistics
Descriptive statistics uses the data to provide descriptions of the population, either through numerical calculations or graphs or tables.	Inferential statistics makes inferences and predictions about a population based on a sample of data taken from the population in question.

### 02 : Differentiate between nominal and ordinal scale.

Nominal Scale	Ordinal Scale
Nominal scale is a naming scale, where variables are simply “named” or labeled, with no specific order.	Ordinal scale has all its variables in a specific order, beyond just naming them.
Examples of nominal data include country, gender, race, hair color etc.	Example of ordinal data include having a position in class as “First” or “Second”.



### 03 : Define correlation.

**Ans.** Correlation is a statistical measure that expresses the extent to which two variables are linearly related (meaning they change together at a constant rate). It's a common tool for describing simple relationships without making a statement about cause and effect.

An example of positive correlation would be height and weight. Taller people tend to be heavier.

### 04 : Differentiate between parameter and statistic.

Parameter	Statistic
Parameters are numbers that summarize data for an entire population.	Statistics are numbers that summarize data from a sample, i.e. some subset of the entire population.

### 05 : Write two disadvantages of Mode.

**Ans.** Some disadvantages of Mode are given below,

- It is not well defined.
- It is not based on all the values.
- It is stable for large values so it will not be well defined if the data consists of a small number of values.
- It is not capable of further mathematical treatment.

### 06 : What is Classification. Also give example.

**Ans.** The process of arranging data into homogenous groups or classes according to some common characteristics present in the data is called classification.

For example: During the process of sorting letters in a post office, the letters are classified according to the cities and further arranged according to streets.

**07 : Define hypothesis testing.**

**Ans.** Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. It is used to assess the plausibility of a hypothesis by using sample data. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis.

**08 : Differentiate between point and interval estimation.**

Point Estimation	Interval Estimation
A point estimate is a single value estimate of a parameter.	An interval estimate gives you a range of values where the parameter is expected to lie.
For instance, a sample mean is a point estimate of a population mean.	A confidence interval is the most common type of interval estimate.

**09 : Define probability with examples.**

**Ans.** Probability is the likelihood that an event will occur and is calculated by dividing the number of favorable outcomes by the total number of possible outcomes.

The simplest example is a coin flip. There is a 50% chance the outcome will be heads, and there is a 50% chance the outcome will be tails.

Another plain example is that of rolling a die. A fair die has six-sides, with each side numbered from 1 to 6. Additionally, each side is equally likely to turn up when rolled. And so, the probability of getting each side is 1 out of 6.

**10 : Write down the types of probability sampling and non probability sampling.**

Probability Sampling	Non-Probability Sampling
----------------------	--------------------------



Probability sampling involves random selection, allowing you to make strong statistical inferences about the whole group.	Non-probability sampling involves non-random selection based on convenience or other criteria, allowing you to easily collect data.
There are four main types of probability sample. <ol style="list-style-type: none"> <li>1. Simple random sampling</li> <li>2. Systematic sampling</li> <li>3. Stratified sampling</li> <li>4. Cluster sampling</li> </ol>	The main types of Non-Probability sampling are : <ol style="list-style-type: none"> <li>1. Convenience sampling</li> <li>2. Voluntary response sampling</li> <li>3. Purposive sampling</li> <li>4. Snowball sampling</li> </ol>

### 11 : Write down the formula of pooled sample t-test.

**Ans.** The equal variance t-test is used when the number of samples in each group is the same, or the variance of the two data sets is similar.

The following formula is used for calculating t-value and degrees of freedom for equal variance t-test:

$$T\text{-value} = \frac{mean1 - mean2}{\frac{(n1-1) \times var1^2 + (n2-1) \times var2^2}{n1+n2-2}} \times \sqrt{\frac{1}{n1} + \frac{1}{n2}}$$

**Where:**

Mean1 and mean2 = Average values of each of the sample sets

var1 and var2 = Variance of each of the sample sets

n1 and n2 = Number of records in each sample set

### 12 : Define Skewness.

**Ans.** Skewness refers to a distortion or asymmetry that deviates from the symmetrical bell curve, or normal distribution, in a set of data. If the curve is shifted to the left or to the right, it is said to be skewed.

In other words, Skewness is a measure of the symmetry of a distribution. A distribution is skewed if it is asymmetrical.

**13 : Define level of significance.**

**Ans.** The significance level, also denoted as alpha or  $\alpha$ , is the probability of rejecting the null hypothesis when it is true.

For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference.

**14 : In a moderately skewed distribution, the value of median is 50.55 and the mode is 16.24. Find the value of mean.**

**Sol.** For a moderately skewed distribution,

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$

So,  $16.24 = 3 (50.55) - 2 \text{ Mean}$

$$16.24 = 151.65 - 2 \text{ Mean}$$

$$16.24 - 151.65 = - 2 \text{ Mean}$$

$$- 135.41 = - 2 \text{ Mean}$$

Multiplying by “-” on both sides

$$135.41 = 2 \text{ Mean}$$

$$135.41/2 = \text{Mean}$$

$$\text{Mean} = 67.705$$

**15 : Define interquartile range.**

**Ans.** The interquartile range (IQR), also called the midspread, middle 50%, or H-spread, is a measure of variability, based on dividing a data set into quartiles. It is a measure of statistical dispersion, being equal to the difference between 75<sup>th</sup> and 25<sup>th</sup> percentiles, or between upper and lower quartiles,  $\text{IQR} = Q3 - Q1$ .

**16 : Define sample.**



**Ans.** In statistics, a sample is a set of individuals or objects collected or selected from a statistical population by a defined procedure. It refers to a smaller, manageable version of a larger group. It is a subset containing the characteristics of a larger population. The elements of a sample are known as sample points, sampling units or observations.