a. What do you mean by statistics?

Statistics is the discipline that concerns the collection, organization, analysis, interpretation and presentation of data. ... **Statistics** deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments. See glossary of probability and **statistics**.

b. What are the two types of statics?

Two types of statistical methods are used in analyzing data:
descriptive **statistics** and inferential **statistics**. Descriptive **statistics** are used to synopsize data from a sample exercising the mean or standard deviation. Inferential **statistics** are used when data is viewed as a subclass of a specific population

c. What is statistics and its uses?

Statistics are the sets of mathematical equations that we used to analyze the things. It keeps us informed about, what is happening in the world around us. ... They **use** their **statistical** skills to collect the relevant data. Otherwise, it results in a loss of money, time and data.

d. What are the advantages of statistics?

they are familiar to library staff and managers. they can be analysed relatively quickly. information is collected in a standardised way. they are usually straightforward to analyse.

e. What are the real life application of statistics?

Statistics are used behind all the medical study. **Statistic** help doctors keep track of where the baby should be in his/her mental development. Physician's also use **statistics** to examine the effectiveness of treatments. **Statistics** are very important for observation, analysis and mathematical prediction models.

f. What is the scope of statics?

Statistical data and techniques of **statistical** analysis are immensely useful in solving economical problems such as wages, price, time series analysis, demand analysis. It is an irreplacable tool of production control. ... Industry **statistics** are widely used in equality control.

g. What are the scope and limitation of statistics?

The **scope** of the science of **statistic** is restricted by certain **limitations** : 1. The use of **statistics** is limited to numerical studies: **Statistical** methods cannot be applied to study the nature of all type of phenomena.

h. What is the nature of statistics?

What is the nature of statistics – science or art? Answer: **Statistics** is both science and art. **Statistical** methods are systematic and have a general application which makes it a science. Further, the successful application of these methods requires skills and experience of using the **statistical** tools.

i. What are the statistics?

Statistics is a form of mathematical analysis that uses quantified models, representations and synopses for a given set of experimental data or real-life studies. Statistics studies methodologies to gather, review, analyze and draw conclusions from data. Some statistical measures include the following:

- Mean
- Regression analysis
- Skewness
- Kurtosis
- Variance
- Analysis of variance

j. Discussion

Mean

A mean is the mathematical average of a group of two or more numerals. The mean for a specified set of numbers can be computed in multiple ways, including the arithmetic mean, which shows how well a specific commodity performs over time, and the geometric mean, which shows the performance results of an investor's portfolio invested in that same commodity over the same period.

Regression Analysis

Regression analysis determines the extent to which specific factors such as interest rates, the price of a product or service, or particular industries or sectors influence the price fluctuations of an asset. This is depicted in the form of a straight line called linear regression.

Skewness

Skewness describes the degree a set of data varies from the standard distribution in a set of statistical data. Most data sets, including commodity returns and stock prices, have either positive skew, a curve skewed toward the left of the data average, or negative skew, a curve skewed toward the right of the data average.

Kurtosis

Kurtosis measures whether the data are light-tailed (less outlier-prone) or heavy-tailed (more outlier-prone) than the normal distribution. Data sets with high kurtosis have heavy tails, or outliers, which implies greater investment risk in the form of occasional wild returns. Data sets with low kurtosis have light tails, or lack of outliers, which implies lesser investment risk.

Variance

Variance is a measurement of the span of numbers in a data set. The variance measures the distance each number in the set is from the mean.

Variance can help determine the risk an investor might accept when buying an investment.

Ronald Fisher developed the analysis of variance method. It is used to decide the effect solitary variables have on a variable that is dependent. It may be used to compare the performance of different stocks over time.

k.What is the difference between sample and population in statictics?

The main difference between a population and sample has to do with how observations are assigned to the data set. A population includes all of the elements from a set of data. A sample consists one or more observations drawn from the population.

l.What is a population and sample in statistics?

"population" data sets and "sample" data sets. A population data set contains all members of a specified group (the entire list of possible data values). ... A sample data set contains a part, or a subset, of a population. The size of a sample is always less than the size of the population from which it is taken.

m. What is population in data analysis?

A population, in statistics and other areas of mathematics, is a discrete group of people, animals or things that can be identified by at least one common characteristic for the purposes of data collection and analysis. To gather information about a large population, data is usually gathered from a sample.

n. What are some examples of population?

Population is the number of people or animals in a particular place. An example of population is over eight million people living in New York City.
.

Problem on Statictics

Mean, Mode, Median, and Standard Deviation

The Mean and Mode

The *sample mean* is the average and is computed as the sum of all the observed outcomes from the sample divided by the total number of events. We use x as the symbol for the sample mean. In math terms,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x$$

where n is the sample size and the x correspond to the observed valued.

Example

Suppose you randomly sampled six acres in the Desolation Wilderness for a non-indigenous weed and came up with the following counts of this weed in this region:

34, 43, 81, 106, 106 and 115

We compute the sample mean by adding and dividing by the number of samples, 6.

$$\frac{34 + 43 + 81 + 106 + 106 + 115}{6} = 80.83$$

We can say that the sample mean of non-indigenous weed is 80.83.

The *mode* of a set of data is the number with the highest frequency. In the above example 106 is the mode, since it occurs twice and the rest of the outcomes occur only once.

The *population mean* is the average of the entire population and is usually impossible to compute. We use the Greek letter μ for the population mean.

Median, and Trimmed Mean

One problem with using the mean, is that it often does not depict the typical outcome. If there is one outcome that is very far from the rest of the data, then the mean will be strongly affected by this outcome. Such an outcome is called and *outlier*. An alternative measure is the median. The *median* is the middle score. If we have an even number of events we take the average of the two middles. The median is better for describing the typical value. It is often used for income and home prices.

Example

Suppose you randomly selected 10 house prices in the South Lake Tahoe area. Your are interested in the typical house price. In $100,000 the prices were

2.7, 2.9, 3.1, 3.4, 3.7, 4.1, 4.3, 4.7, 4.7, 40.8

If we computed the mean, we would say that the average house price is 744,000. Although this number is true, it does not reflect the price for available housing in South Lake Tahoe. A closer look at the data shows that the house valued at 40.8 x $100,000 = $4.08 million skews the data. Instead, we use the median. Since there is an even number of outcomes, we take the average of the middle two

$$\frac{3.7 + 4.1}{2} = 3.9$$

The median house price is $390,000. This better reflects what house shoppers should expect to spend.

There is an alternative value that also is resistant to outliers. This is called the *trimmed mean* which is the mean after getting rid of the outliers or 5% on the top and 5% on the bottom. We can also use the trimmed mean if we are concerned with outliers skewing the data, however the median is used more often since more people understand it.

Example:

At a ski rental shop data was collected on the number of rentals on each of ten consecutive Saturdays:

44, 50, 38, 96, 42, 47, 40, 39, 46, 50.

To find the sample mean, add them and divide by 10:

$$\frac{44 + 50 + 38 + 96 + 42 + 47 + 40 + 39 + 46 + 50}{10} = 49.2$$

Notice that the mean value is not a value of the sample.

To find the median, first sort the data:

38, 39, 40, 42, 44, 46, 47, 50, 50, 96

Notice that there are two middle numbers 44 and 46. To find the median we take the average of the two.

$$\text{Median} = \frac{44 + 46}{2} = 45$$

Notice also that the mean is larger than all but three of the data points. The mean is influenced by outliers while the median is robust.

Variance, Standard Deviation and Coefficient of Variation

The mean, mode, median, and trimmed mean do a nice job in telling where the center of the data set is, but often we are interested in more. For example, a pharmaceutical engineer develops a new drug that regulates iron in the blood. Suppose she finds out that the average sugar content after taking the medication is the optimal level. This does not mean that the drug is effective. There is a possibility that half of the patients have dangerously low sugar content while the other half have dangerously high content. Instead of the drug being an effective regulator, it is a deadly poison. What the pharmacist needs is a measure of how far the data is spread apart. This is what the variance and standard deviation do. First we show the formulas for these measurements. Then we will go through the steps on how to use the formulas.

We define the *variance* to be

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})^2$$

and the *standard deviation* to be

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})^2}$$

Variance and Standard Deviation: Step by Step

1. Calculate the mean, x.
2. Write a table that subtracts the mean from each observed value.
3. Square each of the differences.
4. Add this column.
5. Divide by n -1 where n is the number of items in the sample This is the *variance*.
6. To get the *standard deviation* we take the square root of the variance.

Example

The owner of the Ches Tahoe restaurant is interested in how much people spend at the restaurant. He examines 10 randomly selected receipts for parties of four and writes down the following data.

44, 50, 38, 96, 42, 47, 40, 39, 46, 50

He calculated the mean by adding and dividing by 10 to get

x = 49.2

Below is the table for getting the standard deviation:

x	x - 49.2	$(x - 49.2)^2$
44	-5.2	27.04
50	0.8	0.64
38	11.2	125.44
96	46.8	2190.24

42	-7.2	51.84
47	-2.2	4.84
40	-9.2	84.64
39	-10.2	104.04
46	-3.2	10.24
50	0.8	0.64
Total		**2600.4**

Now

$$\frac{2600.4}{10 - 1} = 288.7$$

Hence the variance is 289 and the standard deviation is the square root of 289 = 17.

Since the standard deviation can be thought of measuring how far the data values lie from the mean, we take the mean and move one standard deviation in either direction. The mean for this example was about 49.2 and the standard deviation was 17. We have:

$$49.2 - 17 = 32.2$$

and

$$49.2 + 17 = 66.2$$

What this means is that most of the patrons probably spend between $32.20 and $66.20.

The sample standard deviation will be denoted by s and the population standard deviation will be denoted by the Greek letter σ.

The sample variance will be denoted by s^2 and the population variance will be denoted by σ^2.

The variance and standard deviation describe how spread out the data is. If the data all lies close to the mean, then the standard deviation will be small, while if the data is spread out over a large range of values, s will be large. Having outliers will increase the standard deviation.

One of the flaws involved with the standard deviation, is that it depends on the units that are used. One way of handling this difficulty, is called the *coefficient of variation* which is the standard deviation divided by the mean times 100%

$$CV = \frac{\sigma}{\mu} 100\%$$

In the above example, it is

$$\frac{17}{49.2} 100\% = 34.6\%$$

This tells us that the standard deviation of the restaurant bills is 34.6% of the mean.