1. If the variance of a variable/column is 0 then what does it mean? Can we use that variable for our analysis?

Ans) If the variance is zero it means that there is no variability in the data set and all the numbers in the data set are same. Suppose for example if we have a data set with 10 numbers and all the numbers are same (say all the numbers are 5 i.e. 5,5,5,5,5,5,5,5,5,5) then according to the formula of variance: -

$$\sigma^2 = \frac{\sum(X-\mu)^2}{N}$$

The variance will be zero.

We cannot use that variable for our analysis as it is not carrying any meaningful information.

2. Calculate mean, median, mode, variance and standard deviation for column A

A	B	C	D
7	5	2003	2003
6	8	1976	1976
7	5	2001	2002
7	5	1915	1970
8	5	2000	2000
5	5	1993	1995
8	5	2004	2005
7	6	1973	1973
7	5	1931	1950
5	6	1939	1950
5	5	1965	1965

Mention all step by step formula calculations in the answer sheet.

Ans) Mean = $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

$$= \frac{7+6+7+7+8+5+8+7+7+5+5}{11}$$

$$= \frac{72}{11}$$

= 6.5454

Arranging the data of column A in ascending order, we have: -

5,5,5,6,7,**7**,7,7,7,8,8

Clearly, the middle number is 7 (marked in bold and underlined above)

Therefore, the Median is 7.

7 appears the most number of times (5 times)

So Mode is 7.

Variance of the population = $\sigma^2 = \sum \frac{(X-\mu)^2}{N}$

Where, N = 11
μ = 6.5454

$$\text{Variance} = \frac{(7-6.5454)^2 + (6-6.5454)^2 + (7-6.5454)^2 + (7-6.5454)^2 + (8-6.5454)^2 + (5-6.5454)2 + (8-6.5454)^2 + (7-6.5454)^2 + (7-6.5454)^2 + (5-6.5454)^2 + (5-6.5454)^2}{11}$$

$$= \frac{0.2066+0.2974+0.2066+0.2066+2.1158+2.3882+2.1158+0.2066+0.2066+2.3882+2.3882}{11}$$

$$= \frac{12.7266}{11}$$

= 1.1569

Variance of the sample = $\sigma^2 = \sum \frac{(X-\mu)^2}{N-1}$

$$= \frac{12.7266}{11-1}$$

$$= \frac{12.7266}{10}$$

= 1.2726

Standard Deviation of the population = $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Where n = 11, $\bar{x}$ = 6.5454

s = √1.1569

= 1.0755

Standard Deviation of the sample = $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Where n = 11, $\bar{x}$ = 6.5454

s = √1.2726

= 1.1281

Since, Standard Deviation is the square root of the Variance.

3. In a group of 12 scores, the largest score is increased by 36 points. What effect will this have on the mean of the scores?

Ans)

Mean = $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i$ …………….. (i)

New Mean = $\bar{\mathbf{x}}' = \frac{1}{n}\left(\sum_{i=1}^{n}(x_i) + a\right)$ …………… (ii)

Where n = 12 and a = 36

Now subtracting (i) from (ii), we get: -

$\bar{X}' - \bar{X} = \frac{1}{n}\left(\sum_{i=1}^{n}(x_i) + a\right) - \frac{1}{n}\sum_{i=1}^{n} x_i$

$= \frac{1}{n}\sum_{i=1}^{n} x_i$ + a/n - $\frac{1}{n}\sum_{i=1}^{n} x_i$

$= \quad a/n$

$= \quad 36/12$

$= \quad 3$

Therefore, the new mean will be increased by 3.

4. Explain the difference between Data (Singular) and Data (Plural) with examples?

Ans) • **Data (singular)**: The value of the variable associated with one element of a population or sample. This value may be a number, a word, or a symbol.
• **Data (plural)**: The set of values collected for the variable from each of the elements belonging to the sample.

For example, let us consider the following table: -

A	B	C	D
7	5	2003	2003
6	8	1976	1976
7	5	2001	2002
7	5	1915	1970
8	5	2000	2000

In column D, if we take the first value i.e. 2003, then it is called a data (singular) because this value (or in this case, the number) is associated with one element of the given sample.

Similarly, in column D, if we take all the values i.e. 2003, 1976, 2002, 1970 and 2000, then they are called a data (plural) because these are the set of values collected for the variable from each of the elements belonging to the given sample.

5. How the inferential statistics helps to make decisions out of it?

Ans) After we are done with descriptive statistics, we use inferential statistics to draw conclusion on a large set of data. It helps us to estimate decisions and draw an initial insight of the data so that it can be presented to the client.

For example, suppose we want to know the average female heights of the people of India. It is almost impossible to collect all the female heights of India. So select a sample which covers the larger population of females of India, collect their heights and infer about the whole female population. Inferential statistics are techniques that allow us to use these samples to make generalizations about the populations from which the samples were drawn.