BUSINESS STATISTICS & ANALYTICS

Unit 1

MBA/BBA/B.com /B.Tech /UGC Net

By

Dr. Anand Vyas

Meaning of statistics

a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.

Scope of statistics,

If we take the various statistical methods in consideration, we can define statistics as a science in which we study: **Numerous methods of collecting, editing, classifying, tabulating and presenting facts using graphs and diagrams. Several ways of condensing data regarding various social, political, and economic**

Types of Statistics

Two types of statistical methods are used in analyzing data: **descriptive statistics and inferential statistics**.

Descriptive statistics summarize the characteristics of a data set.

Inferential statistics allow you to test a hypothesis or assess whether your data is generalizable to the broader population.

Functions of statistics

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.

limitations of statistics

1. **Statistics cannot be applied to heterogeneous data.**
2. **It does not deal with individual items.**
3. **Laws are not exact**
4. **Results are true only on average**
5. **To Many methods to study problems**

Meaning of Central tendency

- Central tendency is **a descriptive summary of a dataset through a single value that reflects the center of the data distribution**.

Measures of Central tendency – Mean, Median and Mode

Mean :

Mean is the most common measure of central tendency. Mean can be defined for all ratio-scale and interval-scale data. To calculate the mean, simply add all of your numbers together. Next, divide the sum by however many numbers you added. The result is your mean or average score.

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

Median

Median is the middle number of any data series which has been sorted in ascending (i.e. lowest to highest) or descending (i.e. highest to lowest) order. If the series is odd, then the median is exactly the middle number. However, if the series is even, then it is the average of the middle two numbers. Median is defined for ordinal data too, along with interval-scale or ratio-scale data.
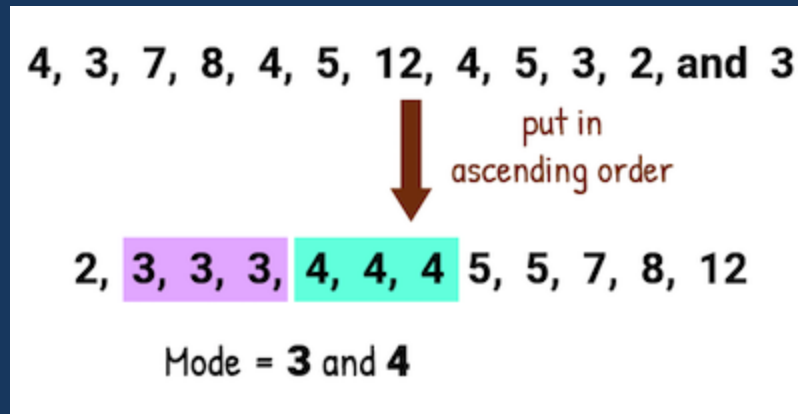
1, 3, 3, **6**, 7, 8, 9

Median = 6

1, 2, 3, **4, 5**, 6, 8, 9

Median = $(4 + 5) \div 2$

= 4.5

Mode

In statistics, the mode in a list of numbers refers to the integers that occur most frequently. Unlike the median and mean, the mode is about the frequency of occurrence. There can be more than one mode or no mode at all; it all depends on the data set itself.

4, 3, 7, 8, 4, 5, 12, 4, 5, 3, 2, and 3

put in ascending order

2, 3, 3, 3, 4, 4, 4, 5, 5, 7, 8, 12

Mode = **3** and **4**

Measures of Dispersion

- In statistics, dispersion is the extent to which a distribution is stretched or squeezed. Common examples of measures of statistical dispersion are the variance, standard deviation, and interquartile range. For instance, when the variance of data in a set is large, the data is widely scattered

- Measures of Dispersion – Range, Inter quartile range, Mean deviation, Standard deviation,

Range

Range refers to the difference between each series' minimum and maximum values. The range offers us a good indication of how dispersed the data is, but we need other measures of variability to discover the dispersion of data from central tendency measurements. A range is the most common and easily understandable measure of dispersion. It is the difference between two extreme observations of the data set. If X_{max} and X_{min} are the two extreme observations then

Range = $X_{max} - X_{min}$

Quartile Deviation

The quartiles divide a data set into quarters. The first quartile, (Q_1) is the middle number between the smallest number and the median of the data. The second quartile, (Q_2) is the median of the data set. The third quartile, (Q_3) is the middle number between the median and the largest number.

Quartile deviation or semi-inter-quartile deviation is

$Q = \frac{1}{2} \times (Q_3 - Q1)$

Mean Deviation

Mean deviation is the arithmetic mean of the absolute deviations of the observations from a measure of central tendency. If x_1, x_2, ... , x_n are the set of observation, then the mean deviation of x about the average A (mean, median, or mode) is

Mean deviation from average A = $1⁄n [\sum_i |x_i - A|]$

For a grouped frequency, it is calculated as:

Mean deviation from average A = $1⁄N [\sum_i f_i |x_i - A|]$, N = $\sum f_i$

Here, x_i and f_i are respectively the mid value and the frequency of the ith class interval.

No.			
5	6	1	
3	6	3	
7	6	1	
8	6	2	
4	6	2	
9	6	3	
36		12	Mean Deviations 2
6		6	

Standard Deviation

A standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean. It is denoted by a Greek letter sigma, σ. It is also referred to as root mean square deviation. The standard deviation is given as

Standard Deviation Formula

Population	Sample
$$\sigma = \sqrt{\dfrac{\sum(X - \mu)^2}{N}}$$	$$s = \sqrt{\dfrac{\sum(X - \overline{x})^2}{n - 1}}$$
X - The Value in the data distribution μ - The population Mean N - Total Number of Observations	X - The Value in the data distribution $\overline{x}$ - The Sample Mean n - Total Number of Observations

4	-1	1		
2	-3	9		
5	0	0		
8	3	9		
6	1	1		
25		20	S.D. =√5	2.236
5		4 (N-1)		

Coefficient of Variation

The coefficient of variation (CV) is **a statistical measure of the relative dispersion of data points in a data series around the mean**. In finance, the coefficient of variation allows investors to determine how much volatility, or risk, is assumed in comparison to the amount of return expected from investments.

Formula

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

σ = population standard deviation

μ = population mean

	Regular Test	Randomized Answer
Mean	59.9	44.8
SD	10.2	12.7
CV	17.03	28.35

$$\text{coefficient of variation} = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100\%$$

- Skewness and Kurtosis.
- **Skewness essentially measures the symmetry of the distribution, while kurtosis determines the heaviness of the distribution tails.**" The understanding shape of data is a crucial action. It helps to understand where the most information is lying and analyze the outliers in a given data.

Skewness measures **the deviation of a random variable's given distribution from the normal distribution, which is symmetrical on both sides**. A given distribution can be either be skewed to the left or the right. Skewness risk occurs when a symmetric distribution is applied to the skewed data.