[Academic Script]

Central Tendency & Dispersion

Subject:	Business Economics
Course:	B. A. (Hons.), 1st Semester, Undergraduate
Paper No. & Title:	Paper – 102 Statistics for Business Economics
Unit No. & Title:	Unit – 2 Univariate Analysis
Lecture No. & Title:	Lecture – 1 Measures of Central Tendency & Dispersion

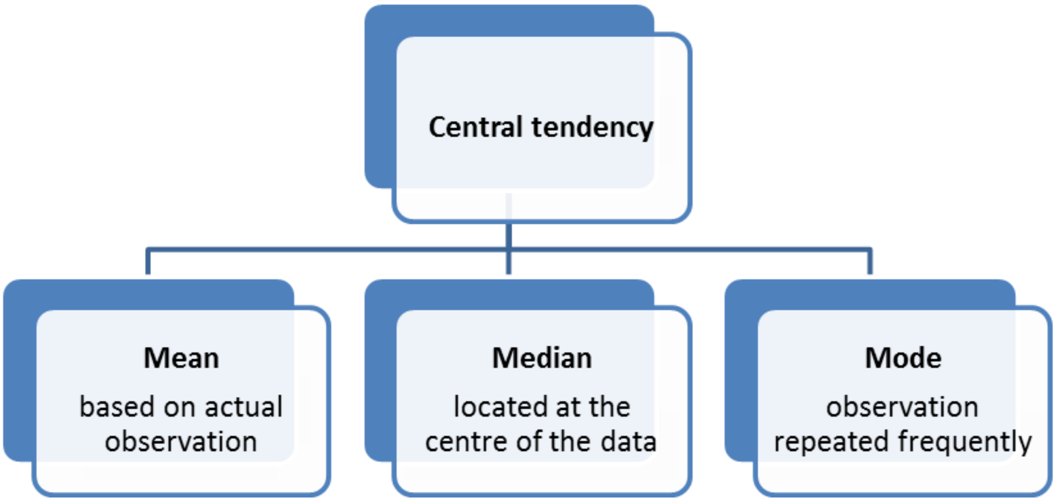
Academic Script

1. Introduction

We have already studied techniques of data collection and data presentation. Huge data is considerably condensed by classification, tabulation, diagrams and graphs.

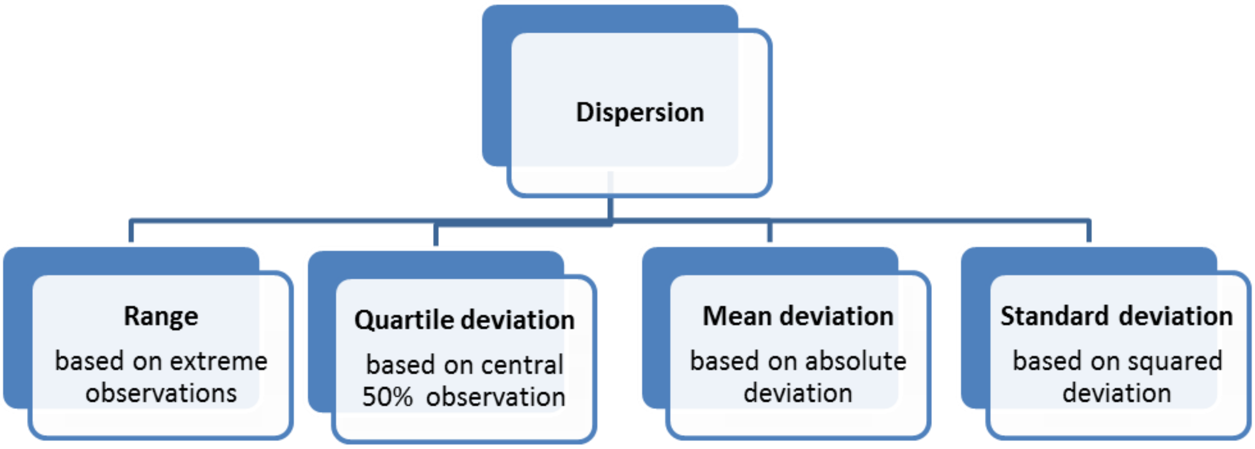
Classification and tabulation of data are very useful but not sufficient to provide summary of data. We require a value which condenses the entire data into single numerical value. A measure capable of representing the important characteristics of the data is known as **measure of average or measures of central tendency**. While **measures of dispersion** indicate the extent to which individual observations fall away from the central value. The measure of dispersion which is expressed in terms of original units of the data is known as **absolute measure** of dispersion. For comparison of two or more data, **relative measure** of dispersion is required. Relative measure is obtained as ratio or percentage, thus it is independent of the units of measurement. Central tendency gives a typical value of the variable, while dispersion gives how much variability there is in the values of the variable. When describing the values of a single variable, it is customary to report on both the central tendency and the dispersion. Not all measures of central tendency and not all measures of dispersion can be used to describe the values of cases on every variable. Depending on the type of data and the purpose of the study, measures of central tendency and dispersion should be used.

2. Central Tendency



Central tendency	**Ungrouped data**	**Discrete data**	**Continuous data**
Mean	$\bar{x} = \frac{\sum x}{n}$	$\bar{x} = \frac{\sum fx}{\sum f}$ Or $\bar{x} = A + \frac{\sum fd}{\sum f}$ where $d = x - A$	$\bar{x} = A + \frac{\sum fd}{\sum f} * c$ where $d = \frac{x-A}{c}$
Median	M $=(\frac{n+1}{2})$th observation	M $=(\frac{n+1}{2})$th observation	$M = L + \frac{\frac{n}{2} - cf}{f} * c$
Mode	Frequently occurring observation	Observation having highest frequency	$M_o = L + \frac{f_1 - f_0}{f_2} * c$

3. Dispersion



Measures of Dispersion	**Ungrouped data**	**Discrete data**	**Continuous data**
Range	R $=X_H - X_L$	R $=X_H - X_L$	R $= U.L. - L.L.$
Quartile deviation	Q.D.$= \frac{Q_3 - Q_1}{2}$	Q.D.$= \frac{Q_3 - Q_1}{2}$	Q.D.$= \frac{Q_3 - Q_1}{2}$
Mean deviation	M.D.$=\frac{\sum\lvert x-\bar{x}\rvert}{n}$	M.D.$=\frac{\sum f\lvert x-\bar{x}\rvert}{n}$	M.D.$=\frac{\sum f\lvert x-\bar{x}\rvert}{n}$
Standard deviation	S $=\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$	S $=\sqrt{\frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2}$	S $=\sqrt{\frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2}$

There are many other derived formulas for mean deviation and standard deviation.

4. Example

Example1: Following are the daily sales of a garment showroom. Find mean, median and mode. Also calculate range, quartile deviation, mean deviation and standard deviation.

Day	1	2	3	4	5	6	7
Sales (Rs.)	51000	53000	49000	50000	51500	49500	51000

Solution:

Mean: $\bar{x} = \frac{\sum x}{n}$

$= \frac{51000+53000+49000+50000+51500+49500+51000}{7} = \frac{355000}{7}$

= 50714.29 Rs.

The mean daily sale is Rs. 50714.29

Median: M =$(\frac{n+1}{2})$th observation

For median observations need to be arranged in ascending order.

49000,49500,50000,51000,51500,51500,53000

Now, M =$\left(\frac{7+1}{2}\right) = 4$ th observation.

= 51000

The median daily sales is Rs. 51000

Mode is the frequently occurring observation in the data.

Here 51500 is occurring maximum number of times i.e. twice.

The modal daily sale is Rs. 51500.

Range is the difference between the highest and lowest observation in the data.

R =$X_H - X_L$

=53000 -49000

=4000 Rs.

Range of daily sales is 4000 Rs.

Quartile deviation: Q.D. = $\frac{Q_3 - Q_1}{2}$

For quartiles observations need to be arranged in ascending order.

49000,49500,50000,51000,51500,51500,53000

Q_1 =$(\frac{n+1}{4})$th observation

$=(\frac{7+1}{4})$th observation

= 2nd observation

= 49500

Q_3 $=3(\frac{n+1}{4})$th observation

$=3(\frac{7+1}{4})$th observation

= 6th observation

= 51500

So, Q.D. $= \frac{Q_3 - Q_1}{2} = \frac{51500-49500}{2} = \frac{2000}{2} = 1000$

Thus quartile deviation of daily sales is Rs. 1000

Mean deviation: M.D. $= \frac{\sum|x-\bar{x}|}{n}$

Daily sales (x)	$x - \bar{x}$	$\lvert x - \bar{x}\rvert$
51000	285.71	285.71
53000	2285.71	2285.71
49000	-1714.29	1714.29
50000	-714.29	714.29
51500	785.71	785.71
49500	-1214.29	1214.29
51000	285.71	285.71
		7285.71

$\bar{x}$ = 50714.29 (We have already calculated)

M.D. $= \frac{\sum|x-\bar{x}|}{n}$

$= \frac{7285.71}{7}$

= 1040.82 Rs.

Thus mean deviation from mean of daily sales is Rs. 1040.82

Standard deviation: S = $\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$ or $\sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$

Daily sales (x)	$d = x - A$ ($A = 50000$)	d^2
51000	1000	1000000
53000	3000	9000000
49000	-1000	1000000
50000	0	0
51500	1500	2250000
49500	-500	250000
51000	1000	1000000
	5000	**14500000**

A is the assumed mean and we can assume any value. Let us take A = 50000

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

= $\sqrt{\frac{14500000}{7} - \left(\frac{5000}{7}\right)^2}$

= $\sqrt{2071428.57 - (714.29)^2}$

= $\sqrt{2071428.57 - 510210.20}$

= $\sqrt{156128.37}$

= 1249.49 Rs.

Thus standard deviation of the daily sales is Rs. 1249.49.

5. Which Measure to use?

Which measure to select and use depend on the variable's level of measurement.

Table below gives us the "big picture" for this chapter. The statistics we gain as we move from nominal to ordinal to interval/ratio are in the table.

Table: Measures of Central Tendency and Dispersion by Level of Measurement

Level of Measurement	***Measure of Central tendency***	***Measure of Dispersion***
Nominal	Mode	*Quartile deviation*
Ordinal	Median Mode	Range Quartile deviation
Interval / Ratio	Mean Median Mode	Range Quartile deviation Mean deviation Standard deviation

Modes are relatively simple in nature and only require that the values that make up a variable be different. That is a property that the values of nominal, ordinal, and interval/ratio variables all have. Therefore, we can calculate mode for all three levels of measurement.

Median and range provide more information about the scores on variables, but these statistics only make sense if the values of a variable have rank order. Since rank order is a property possessed by the values of ordinal and interval/ratio variables but not nominal variables, we may only calculate median, range and inter-quartile range for ordinal and interval/ratio variables.

Mean, variance, and standard deviation provide still more information about the scores on variables, but these statistics require the values of the variable to form a numeric scale with a fixed unit of measurement. Since only interval/ratio variables have this property, mean, variance, and standard deviation may only be calculated for interval/ratio variables.

Usually, we should use a statistic that makes the fullest use of the information packed into the variable's values. Here are the standard approaches: For a nominal variable, report the mode and just a few valid percents. For an ordinal variable, report the median, the minimum, and the maximum. For an interval/ratio variable, report the mean and the standard deviation.

The exception has to do with interval/ratio variables whose distributions are badly skewed. A skewed distribution occurs when there are a few extreme scores on a variable but only in one direction. For example, when all but a few students score in the 80s and 90s in an exam but those few others score below 50, the mark's distribution is negatively skewed. Or when most employees in a company receive annual salaries between 5 million and 6 million but a few executives receive salaries in excess of 6 million, the salary distribution is positively skewed. The extreme cases pull the mean in their direction. The result is that the mean, although mathematically correct, is not a good measure of central tendency. The median, however, is not so sensitive to extreme values. The addition of one or two extreme cases has only a minor effect on the median, but they have a major effect on the mean. Therefore, report the median instead of the mean when the distribution is badly skewed.

6. Summary

A measure capable of representing the important characteristics of the data is known as measure of averages. As such an average will have a tendency to be somewhere at the centre of the data, it is known as measure of central tendency.

Mean, median and mode are measures of central tendency.

Measure of dispersion indicates the extent to which individual observations fall away from the central value.

Range, quartile deviation, mean deviation and standard deviation are measures of dispersion.