BRAC UNIVERSITY
Inspiring Excellence

Measure of Dispersion or Variation

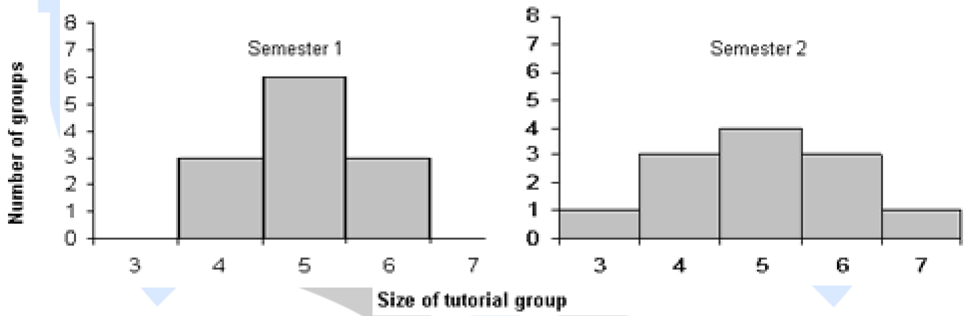
Measures of average (such as the median and mean) represent the typical value for a dataset. Within the dataset the actual values usually differ from one another and from the average value itself. The extent to which the central value are good representatives of the values in the original dataset depends upon the variability or dispersion in the original data. Dispersion is the spread or scatter of item values from a measure of central tendency. Dispersion is usually measured as an average of deviations about some central value. Dispersion thus is a type of average and is sometimes called a second order average. Datasets are said to have high dispersion or variation when they contain values considerably higher and lower than the central value.

Group 1	49	50	50	51
Group 2	0	0	100	100

Example 1: let us consider two groups of students with score in a particular examination as shown in the table. The AM for each group is 50. It is clear from the data that the first group consists of near average intelligent student and the 2nd group is made up of very bright and very dull students. It is evident that the distributions of both groups have the same AM. But they differ in variation from $\overline{X}$; such variation is usually measured by the measure of dispersion.

Example 2: In figure 1 the number of different sized tutorial groups in semester 1 and semester 2 are presented. In both semesters the mean and median tutorial group size is 5 students, however the groups in semester 2 show more dispersion (or variability in size) than those in semester 1.

Figure 1: Comparison of tutorial group sizes



Dispersion within a dataset can be measured or described in several ways including the range, inter-quartile range and standard deviation.

Characteristics of a good measure of variation or dispersion:
The following are the characteristics of an ideal measure of variation or dispersion

1. It should be easy to understand.
2. It should be easy to calculate.
3. It should be based upon all observations.
4. It should be rigidly defined.
5. It should be unduly affected by extreme values.
6. It should be suitable for further algebraic treatment.
7. It should be less affected by sampling fluctuation.

Purpose of measure of dispersion or variation:
Measure of dispersion is important for the following purpose.

1. To determine the reliability of an average.
2. To compare the variability.
3. To compare two or more series with regard to their variability.
4. To facilitate the use of other statistical measures.
5. It is one of the most important quantities used to characterize a frequency distribution.

Types of measure of dispersion:
Measure of dispersion or variation may be either absolute or relative.
Absolute measure of variation is expressed in the same statistical unit in which the original data are given such as takas, kilograms, tones, etc. and may be used to compare the variation in two distributions, provided the variables are expressed in the same units and of same average size.

On the other hand, often it is necessary to compare the distribution in two or more different frequency distributions having variables expressed in different units. In such a case dispersion is calculated by dividing the absolute measure of dispersion by a measure of central tendency – which generates pure number that are independent of the unit of measurement. The resultant numerical value is a ***relative measure*** of dispersion.

Which measures of Dispersion to choose

- When dealing with data, ones' **objective** is ***"only to determine"*** the variation of ***single set of variable / information*** **-** s/he can / will choose to use **Absolute measure of dispersion**.

- When dealing with data, ones' **objective** is ***"to determine and compare"*** the variations of ***multiple set of variables / information*** having ***expressed*** in ***same / different unit(s)*** **-** s/he can / will choose to use **Relative measure of dispersion**.

Different types of Absolute and Relative measure of dispersion are listed below:

Absolute measure of dispersion	Relative measure of dispersion
1. Range 2. Quartile deviation 3. Variance and Standard deviation	1. Coefficient of range 2. Coefficient of quartile deviation 3. Coefficient of variation and standard deviation

And so, on

These measures are discussed below:
Range and Coefficient of Range:
The range of a set of data values is the difference between the highest and the lowest values in the set. If X_l & X_s the smallest and the largest values respectively in a set then the range "R" is defined as $R = X_l - X_s$.

For group data the range is taken either as the difference between the lower boundary of the first class and the upper boundary of the last class or as the difference between the highest and the lowest mid-values.

The coefficient of dispersion corresponding to range called coefficient of range and it is obtained by

Coefficient of range = $\dfrac{X_l - X_s}{X_l + X_s}$; Where X_l =Largest value and X_s = Smallest value

Quartile Deviation and Coefficient of Quartile Deviation:
Quartiles divide the observations in to four equal parts, when observations are arranged in order of magnitudes median, denoted by Q_2, is the middle most observation and Q_1 & Q_3 are the middle most observations of the lower and upper half respectively.

Therefore $Q_2 - Q_1$ and $Q_3 - Q_2$ gives us some measure of dispersion. The AM of these two measures give us the quartile deviation and is denoted by QD and is defined as

$$QD = \frac{(Q_2 - Q_1) + (Q_3 - Q_2)}{2} = \frac{Q_3 - Q_1}{2}$$

The coefficient of variation corresponding to quartile deviation is called the coefficient of quartile deviation and is defined as

$$\text{Coefficient of } QD = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Iftekhar M S Kalam
Assistant Professor, MNS

Test yourself
Range and coefficient of range
Quartile deviation and co efficient of quartile deviation

Problem 1:
The following data represents Annual wages of two Factories X and Y. for the given information

i. Construct back to back stem leaf plot.
ii. Determine range and coefficient of range. (in ‘000 Tk)
iii. Determine the quartile deviation and coefficient of Co-efficient of quartile deviation.

Table 1: Annual wages of Factory X workers
(in ‘000 Tk)

91	70	74	79	86	93
60	71	76	79	87	96
112	72	127	79	87	62
68	72	77	79	90	76
69	73	77	85	48	157

Table 2: Annual wages of Factory Y workers
(in ‘000 Tk)

97	78	85	92	97	105
72	79	85	92	97	107
112	79	87	92	97	72
113	80	90	96	68	75
78	82	90	97	100	

Hints:

- Arrange the data from lowest to Highest
- Construct Stem and Leaf for the two groups
- For each group determine
  - Median (or 2nd quartile)
  - Median of the values lower than median (i.e. 1st quartile)
  - Median of the values greater than median (i.e. 3rd quartile)

Test yourself
Range and coefficient of range
Quartile deviation and co efficient of quartile deviation
(Solution)

Answer 1:

i. The stem and leaf plot-

Factory Y Leaf	STEM	Factory X Leaf
	4	8
	5	
8	6	0,2,8,9
2,2,5,8,8,9,9	7	0,1,2,2,3,4,6,6,7,7,9.9,9,9
0,2,5,5,7	8	5,6,7,7
0,0,2,2,2,6,7,7,7,7,7	9	0,1,3,6
0,5,7	10	
2,3	11	2
	12	7
	13	
	14	
	15	7

For the given information	For Factory X (in ‘000 Tk)	For Factory Y (in ‘000 Tk)
Largest value (L)	157 tk	113
Smallest value (S)	48 tk	68
Number of observations (n)	30	29
Median or 2nd Quartile (Q_2)	78 tk	90 tk
1st Quartile (Q_1)	71.5	79 tk
3rd Quartile (Q_3)	87 tk	97 tk

ii. Calculation of Range and Coefficient of Range:

For the given information	For Factory X (in ‘000 Tk)	For Factory Y (in ‘000 Tk)
Largest value (L)	157	113
Smallest value (S)	48	68
Range (R)	$R_X = L_X\text{-}S_X = (157 - 48)Tk = 109Tk$	$R_Y = L_Y\text{-}S_Y = (113 - 68)Tk = 45Tk$
Coefficient of Range ($Co\ R$)	$Co.R_X = \frac{L_X - S_X}{L_X + S_X} = \frac{(157 - 48)\cancel{Tk}}{(157 + 48)\cancel{Tk}} = 0.53$ ***(Unit free quantity***	$Co.R_y = \frac{L_Y - S_Y}{L_Y + S_Y} = \frac{(113 - 68)\cancel{Tk}}{(113 + 68)\cancel{Tk}} = 0.24$ ***(Unit free quantity***

iii. Calculation of Quartile deviation and Coefficient of quartile deviation:

For the given information	For Factory X (in '000 Tk)	For Factory Y (in '000 Tk)
Number of observations (n)	30	29
Median or 2nd Quartile (Q_2)	78 tk	90 tk
1st Quartile (Q_1)	71.5	79 tk
3rd Quartile (Q_3)	87 tk	97 tk
Quartile deviation (QD)	$QD_X = \frac{(Q_3 - Q_1)}{2}$ $= \frac{(87 - 71.5)}{2} Tk$ = 7.75 tk	$QD_Y = \frac{(Q_3 - Q_1)}{2}$ $= \frac{(97 - 79)}{2} Tk$ = 9 tk
Coefficient of Quartile deviation $(Co.QD)$	$Co.QD_X = \frac{(Q_3 - Q_1)}{(Q_3 + Q_1)}$ $= \frac{(87 - 71.5)\cancel{tk}}{(87 + 71.5)\cancel{tk}}$ = 0.097 ***(Unit free quantity)***	$Co.QD_Y = \frac{(Q_3 - Q_1)}{(Q_3 + Q_1)}$ $= \frac{(97 - 79)\cancel{tk}}{(97 + 79)\cancel{tk}}$ = 0.102 ***(Unit free quantity)***

For any queries related to this presentation please contact
IFTEKHAR Mohammad Shafiqul Kalam
Assistant Professor
Department of Mathematics and Natural Sciences
Email: imskalam@bracu.ac.bd