Measure Of Dispersion

Computer Oriented Numerical and Statistical Methods

Outline

- Introduction
- Some Definitions
- Measure of Dispersion
- Range
- Quartile Deviation
- Mean Deviation
- Standard Deviation
- Coefficient of Variation
- Standard Deviation and Normal Curve
- Empirical Relationships

Introduction

- An average does not tell the full story. It is hardly fully representation of a mass unless we know the manner in which the individual items scatter around it. A further description of the series is necessary if we are to gauge how representative the average is .

  –G. Simpson & F. Kafka
- An average is a single value which represents a set of values in a distribution.
- It is a central value which typically represents the entire distribution.
- Dispersion, on the other hand, indicates the extent to which the individual values fall away from the average or the central value

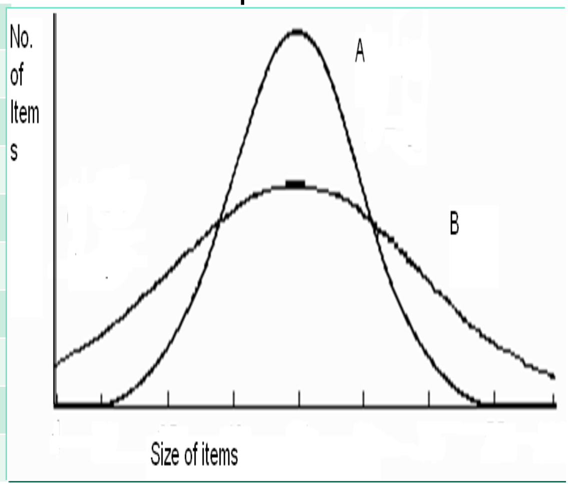
Introduction

- This measure brings out how two distributions with the same average value may have wide differences in the spread of individual values around the central values

Introduction

Size of Item	No. of items (frequency)	
	A	B
0 – 5	1	0
5 – 10	2	3
10 – 15	3	4
15 – 20	5	5
20 – 25	6	8
25 – 30	5	5
30 – 35	3	4
35 – 40	2	3
40 – 50	1	0

- Both have the same mean but the spread-out values



Introduction

- Average is useful to study the state of output, sales, profits, etc, not its variability.
- But to study temperature, the price of shares, rainfall etc. dispersion is found to be more important.
- Dispersion is an important measure sought for describing the character of variability in data.
- While an average discovers the representative value, dispersion finds out how individual values fall apart, on an average, from the representative value.
- The average was derived from the actual values but dispersion is known by averaging the deviations in individual values from some representative value and therefore called am average of the second order.

Some Definitions

- Dispersion is measure of the extent to which the individual items vary. ----- L. R. Connor.
- Dispersion or spared is the degree of the scatter or variation of the variables about a central value. ------ B. C. Brooks & W. F.L. Dick
- The degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data. ------ Spiegel
- Dispersion is the measure of the variations of the items. ------ A. L. Bourey

Uses / Significance of Measure of Variation

- To judge the reliability of central tendency.
- To compare two or more series with regards to their variability.
- To control the variability itself
- To facilitate the use of other statistical measure
- To control quality
- To analysis of time series.

Measure of Dispersion

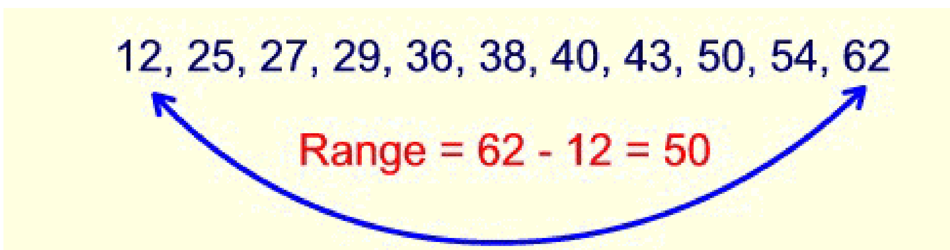
- Measure of Dispersion
  - Algebraic (Absolute and Relative)
    - Range
      - Full Range & Coefficient
      - Quasi Range & Coefficient
      - Quartile Range & Coefficient
      - Percentile Range & Coefficient
    - Quartile Deviation & Coefficient of Q.D.
    - Mean Deviation & Coefficient of M.D.
    - Standard Deviation & Coefficient of S.D.
  - Graphic (Lorenz Curve)

Range

- The range method is based on any two boundary values of a distribution.
- It is not concerned with the rest of the values nor with the concentration on individual values.
- The following are the various types of range.
  - Full Range
  - Quasi Range
  - Inter-Quartile or Quartile Deviation
  - Percentile Range

Full Range

- This is defined as the difference between the largest /highest and the smallest/lowest values in the distribution.
- Range (R) = $X_{max} – X_{min}$ or Range (R) = $X_n – X_1$
- where X_n is the last item and X_1 is the first item of a series arrange in an ascending order of magnitude.



Full Range Example

- The two sets below have the same mean and median (7). Find the range of each set.

Set A	1	2	7	12	13
Set B	5	6	7	8	9

- Solution
- Range of Set A: 13 – 1 = 12.
- Range of Set B: 9 – 5 = 4.

Quasi – Range

- Quasi – Range refers to the difference between the values leaving the extreme values.
- If we leave two extreme values it will be $R = X_{n-1} - X_2$ where n is the total number of items. X_{n-1} is the last but one item and X_2 is next to the first item.
- If more than two extreme values are left out, the expression will be $R = X_{n-2} - X_3$

Coefficient Of Range

- The coefficient of range is derived by dividing a given range by sum of the values of the two boundary values taken into account for calculating range.

$$\text{Coefficient of range} = \frac{\text{Highest value} - \text{Lowest value}}{\text{Highest value} + \text{Lowest value}}$$

$$= \frac{X_n - X_1}{X_n + X_1}$$

- Coefficient of range is more suitable than range for comparison purpose.

Inter-Quartile Range Or Quarter Deviation (Q.D.)

- Q.D. is a measure of dispersion based on the upper quartile (Q_3) and the lower quartile (Q_1).
- It is also called semi-inter-quartile range because it represents the average difference between two quartiles. Quarter Deviation (Q.D.) = $(Q_3 – Q_1) / 2$
- Q.D. as defined above is only an absolute measure of dispersion.
- For comparative studies of variability of two distributions we need an absolute measure which is known as coefficient of quartile deviation and is given by

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Percentile Range

- Percentile range is based on the values of any two percentile.
- For example percentile range ($P_{90} – P_{10}$) is based on values of the 90^{th} and 10^{th} percentile.

Mean Deviation (M.D.)

- The range and quartile deviation are positional measure of dispersion and are based on the position of certain items in a distribution.
- The M.D. or average deviation is a measure of dispersion that is based on all items.
- The M.D. is the arithmetic mean of the deviations of the individual values from the average of the given data.
- The average which is frequently used in computing the mean deviation is mean or median. Also only the absolute values of the dispersion are used.

Mean Deviation (M.D.)

- "Average deviation is the average amount of scatter, of the items in a distribution from either the mean or the median, ignoring the signs of the deviation. The average that is taken of the scatter is an A.M., which accounts for the fact that this measure is often called the M.D."
- M.D. is denoted by δ (delta). The sign of average taken is deviation is used as subscript i.e.
  - $\delta_{\bar{x}}$ (mean deviation from mean)
  - δ_{Md} (mean deviation from median)
  - δ_{Mo} (mean deviation from mode)

Coefficient Of Mean Deviation

- M.D. calculated by any measure of central tendency is an absolute measure.
- When it is divided by the average used for calculating it, we get coefficient of mean deviation which will give a relative measure of dispersion suitable for comparing two or more series which are expressed in different units or expressed in the same units but of different order of magnitude.

1. Coefficient of M.D. = M.D. / Mean or Median or Mode
   if the result is desired in percentage, then
2. Coefficient of Mean variation = (M.D. / Mean or Median or Mode) * 100

Methods of Computation

- If $X_1, X_2, \ldots, X_n$ are n given observations, then the M.D. about an average A, say is given by
  M.D. (about an average A) $= \frac{1}{n}\sum |X - A|$

$$= \frac{1}{n}\sum |d|$$

where |d| = |X – A|
A is any one of the average mean (M), median (Md) or Mode (Mo)

Methods of Computation (Steps)

1. Calculate the average A of the distribution by the usual method.
2. Take the deviation d = X – A of each observation from the average A.
3. Ignore the negative signs of the deviations, taking all the deviation to be positive to obtain the absolute deviations. $|d| = |X - A|$
4. Obtain the sum of the absolute deviations obtain in step(3).
5. Divide the total obtained in step(4) by n, the number of observations. The result gives the value of the mean deviation about the average A.

Mean Deviation for Group Data

- In case of frequency distribution or grouped or continuous frequency distribution, mean deviation about an average A is given by M.D. (about the average A) = $1 / N * \Sigma f*|X - A| = 1 / N * \Sigma f*|d|$
  - Where X is the value of the variable or it is the mid value of the class interval (in case of grouped or continuous frequency distribution) f is the corresponding frequencies, $N = \Sigma f$ is the total frequency and $|X - A|$ is the absolute values of the deviation $d = (X - A)$ of the given values of X from the average A (mean, median, mode)

Mean Deviation for Group Data (Steps)

1. Calculate the average A of the distribution by the usual method.
2. Take the deviation d = X – A of each observation from the average A.
3. Ignore the negative signs of the deviations, taking all the deviation to be positive to obtain the absolute deviations. |d| = |X – A|
4. Multiply the absolute deviation |d| = |X – A| by the corresponding frequency f to get f*|d|
5. Divide the total in step (4) by N, the total frequency.

- The resulting value is the mean deviation about the average A.

Standard Deviation (S.D.)

- The concept of S.D. was first suggest by Karl Pearson in 1893.
- It may be defined as the positive square root of the arithmetic mean of the square deviations of given observations from their arithmetic mean.
- In short S.D. may be defined as "Root-Mean-Square-Deviation from mean."
- It is denoted by σ (sigma).
- It is by far the most important and a widely used measure of studying dispersion.
- For a set of N observations $X_1, X_2, \ldots, X_N$ with mean $\overline{x}$
- Deviation from mean : $(X_1 - \overline{X}), (X_2 - \overline{X}), \ldots\ldots, (X_N - \overline{X})$

Standard Deviation (S.D.)

- Square – Deviations from mean :

$$(X_1-\overline{X})^2,(X_2-\overline{X})^2,......,(X_N-\overline{X})^2$$

- Mean Square Deviation from mean :

$$\frac{1}{N}[(X_1-\overline{X})^2+(X_2-\overline{X})^2+......+(X_N-\overline{X})^2]$$

$$=\frac{1}{N}\sum(x-\bar{x})^2$$

- Root-mean square deviation from mean i.e.

S. D. (σ) = $\sqrt{\frac{1}{N}\sum(x_i-\bar{x})^2}$

- The square of S.D. i.e. σ^2 is known as variance

$\therefore$ varaince = $(S.D.)^2$

Standard Deviation (S.D.)

- In case the data is grouped in the form of frequency distribution, the individual measurements of the variable or the mid-value of the class intervals (as the case may be) are weighted by the corresponding frequency.
- For e.g. if X_1, X_2,…. Are the individual measurements and if f_1, f_2, …. are the corresponding frequencies, then X_1 is weighted by f_1, X_2 by f_2 and so on.
- $\therefore$ S.D. for the frequency distribution is

$$\sigma=\sqrt{\frac{1}{N}\sum f*(x-\bar{x})^2}\ ;\ N=\sum f$$

Comparison Between M.D. And S.D.

M.D.	S.D.
Deviation are calculated from mean, median or mode.	These are calculated from the arithmetic mean only.
The algebraic sign have to be ignored- only values of deviation are taken.	Since the deviations are squared the plus and minus signs need not be omitted.
It is based on simple average of the sum of absolute deviations.	It is based on the square root of the average of the squared deviation.

Comparison Between M.D. And S.D.

M.D.	S.D.
It is simple to calculate when mean is a round number. The short-cut method is somewhat cumbersome.	This is somewhat complex because of squaring of the deviations but it is suitable in all cases- whether the mean is a round number or a fraction, since a short-cut method is also available.
It lacks mathematical prosperities since only absolute values are considered.	It is mathematically sound on account of the fact that algebraic signs are not ignored.

Computation Of S.D.

- Ungrouped Data :
  - Direct Method
  - Short-cut method
- Direct Method :

a) The procedure of computing S.D. from ungrouped data is given below:

1) Obtain the arithmetic mean of the given data.
2) Obtain the deviation of each value from the arithmetic mean, or $x = X - \overline{X}$
3) Square each deviation to make it positive, or x^2

Computation Of S.D.

- Direct Method :

4) Obtain sum of the deviations squared, or Σx^2.
5) Find the variance (σ^2) by dividing the sum by the number of observations (n) in the data i.e. $\sigma^2 = \Sigma x^2 / n$
6) Extract the square root of the variance to find S.D.

$$\sigma = \sqrt{\frac{\sum x^2}{n}}$$

Computation Of S.D.

- Direct Method :

b) The second direct method of computing the S.D. is the one where the values of the variable are used directly and no deviations need to be calculated. The formula is

$$\sigma = \sqrt{\frac{\sum X^2}{n} - \overline{X}^2}$$

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

- i.e S.D. is the square root of the average of the sum of the squared values minus the square of the average of all the values.

Computation Of S.D (Short Cut Method)

1. One of the value, generally the value in the middle, is taken as assumed or working mean (A).
2. Obtain the deviation of each item from the assumed average, dx = X – A, and take the total of the deviations, Σdx.
3. The deviations are squared up and totaled to obtain Σdx^2.
4. The standard deviation is obtained by using any of the following formula

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

$$(b)\ \sigma = \sqrt{\frac{\sum d^2}{n} - \left(\overline{X} - A\right)^2}$$

Computation Of S.D (Grouped Data [Direct Method])

1. Obtain the A.M. of the given data.
2. Find the deviation of the item (in case of discrete series) or the mid-points of each class interval (in-case of continuous series) from the A.M. or

$$x = X - \overline{X} \;\; or \;\; x = m - \overline{X}$$

3. Obtain the total of deviations squared for each class fx^2.
4. Obtain the sum of the deviations squared or Σfx^2.
5. Σfx^2 is divided by the number of items. Then extract the square root of the quotient to obtain the S.D.

$$\sigma = \sqrt{\frac{\sum fx^2}{N}} \; ; N = \sum f$$

Computation Of S.D (Grouped Data [Short-Cut Method])

- The short-cut method of computing the S.D. for grouped data is basically the same as the short-cut method for ungrouped data.
- The only difference is d^2 and d values in each class are multiplied by corresponding frequency of that class

$$\therefore \sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \; ; N = \sum f$$

Computation Of S.D (Grouped Data [Step-Deviation Method])

- In order to simply further the calculation work, some common factor may be taken out from all the deviations.
- The formula for computing S.D. stands modified as follows :

$$\sigma = h * \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \quad ; N = \sum f$$

- Where d' = (X-A)/h (for discrete series)
  d' = (m-A)/h (for continuous series) and h is a common factor, A is the assumed average and m is the mid-value of the class-interval

S.D Of Combined Series

- If two groups contains n_1 and n_2 observations with mean $\overline{X_1}$ *and* $\overline{X_2}$ and S.D. σ_1 and σ_2 respectively, the S.D. (σ_{12}) of the combined group is given by

$$\sigma_{12} = \sqrt{\frac{n_1 * (\sigma_1^2 + d_1^2) + n_2 * (\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

where σ_{12} = combined S.D. of the two groups

$$d_1 = \overline{X_{12}} - \overline{X_1} \text{ and } d_2 = \overline{X_{12}} - \overline{X_2}$$

$$and\ \overline{X_{12}} = \frac{n_1 * \overline{X}_1 + n_2 * \overline{X}_2}{n_1 + n_2}$$

S.D (Combined Group Example)

- The mean and the S.D. of a sample of size 10 were found to be 9.5 and 2.5 respectively. Later on an additional observation became available. This was 15.0 and was included in the original sample. Find the mean and S.D. of the 11 observations.

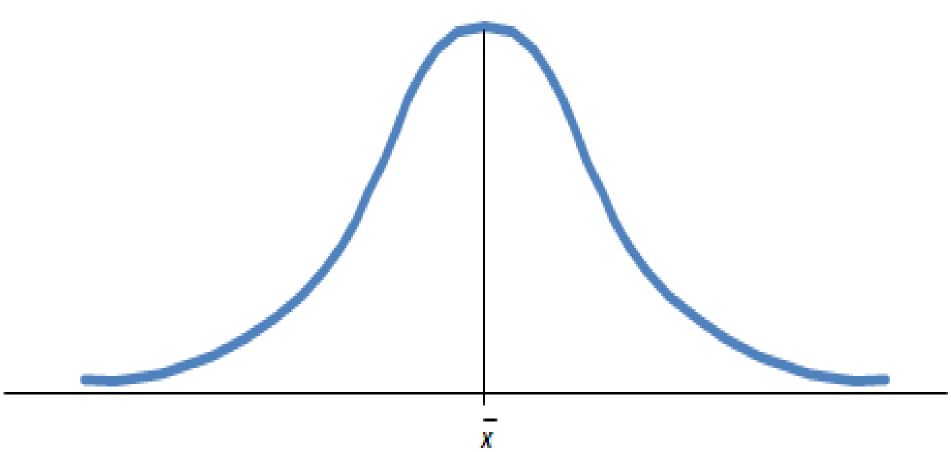
S.D And Normal Curve

- Most distributions are symmetrical bell-shape.
- The frequency curved formed from such distributions may be regarded as approximations to an important well known curve known as the 'normal curve'.
- Mean $\pm$ S.D. will indicate the range within which a given percentage of values of the distributions are likely to fall i.e. nearly 68.27% will lie within mean $\pm$ 1 S.D., 95.45% within mean $\pm$ 2 S.D. and 99.73% within mean $\pm$ 3 S.D. under the normal curve.

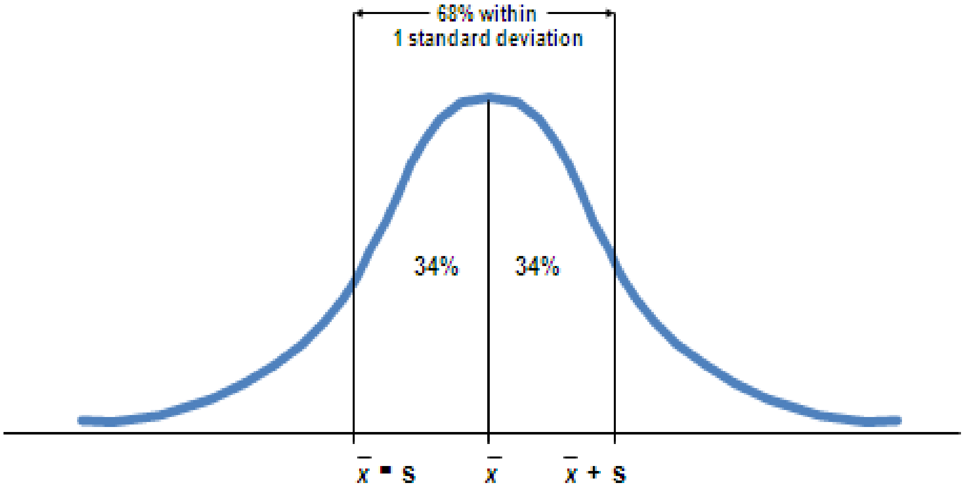
Relationship Among 3 measure of Variability (For a Normal Distribution)

Measure of Variability	% of item included within given range of mean $\overline{X}$			Size of measure of variability to S.D. at
	± One deviation	± two deviation	± three deviation	
Quartile Deviation	50.0	82.3	95.7	0.6748
Mean Deviation	57.5	88.9	98.3	0.7979
Standard Deviation	68.27	95.45	99.73	1.0000

Relationship Among 3 measure of Variability (For a Normal Distribution)

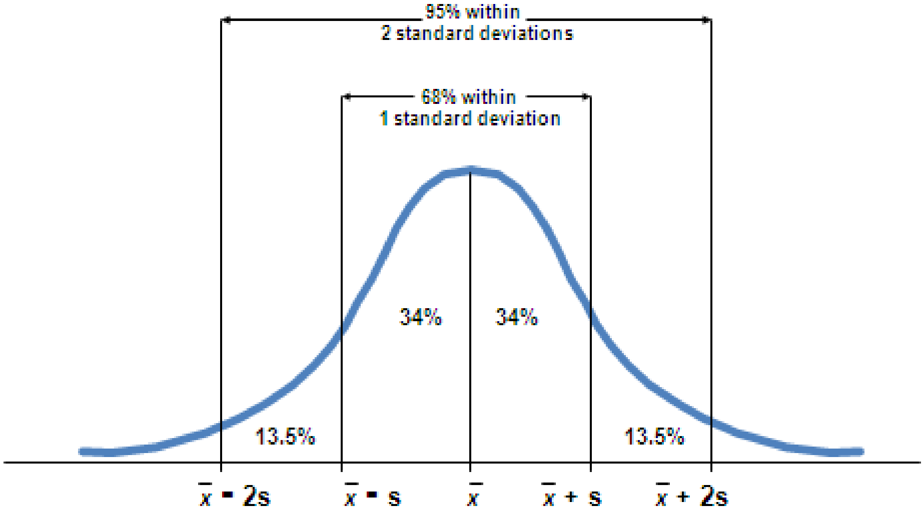


Relationship Among 3 measure of Variability (For a Normal Distribution)



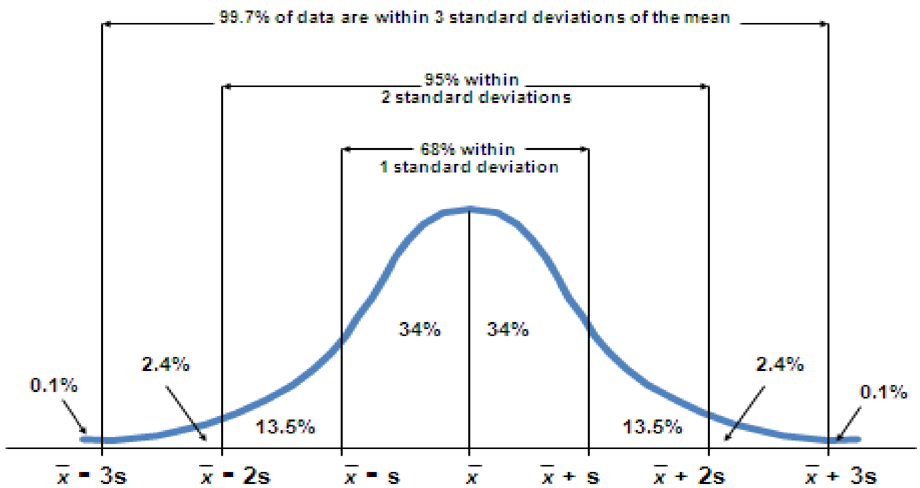
Minal Shah

Relationship Among 3 measure of Variability (For a Normal Distribution)



Minal Shah

Relationship Among 3 measure of Variability (For a Normal Distribution)



Empirical Relationship

1. Quartile Deviation ≈ 2/3 S.D. ≈ 0.6745σ
2. M.D. ≈ 4/5 S.D. ≈ 0.7979σ
3. Q.D. ≈ 5/6 M.D. = 0.8333 M.D.
4. Range = 6 * S.D. or 6σ

- These can further also be expressed as

5. 3 Q.D. ≈ 2 S.D.
6. 5 M.D. ≈ 4 S.D.
7. 6 Q.D. ≈ 5 M.D.

- S.D. ensures highest degree of reliability and Q.D. the lowest 4 S.D. ≈ 5 M.D. ≈ 6 Q.D.


THANKS!
Thanks!
Thanks
THANKS
Thanks
THANKS
THANKS
THANKS