MEASURES OF DISPERSION

WHY STUDY DISPERSION ???

- ❑ An average, such as the mean or the median and others measures of central tendency only locates the center of the data.
- ❑ An average does not tell us anything about the spread of the data

Examples

Consider the series

(i) : 7, 8, 9, 10, 11

(ii) : 3, 6, 9, 12, 15

(iii): 1, 5, 9, 13, 17

In all cases the number of observation, is **$n=5$ and mean is 9.**

Literal meaning of dispersion is " scatteredness ". We study dispersion to have an idea about the homogeneity of heterogeneity of the distribution.

In the above cases:
we say that series (i) is more homogeneous (less dispersed) than the series (ii) or (iii) or
we say that series (iii) is more heterogeneous (more scattered) than the series (i) or (ii).

DISPERSION

- Dispersion refers to the variations of the items among themselves / around an average.
- Greater the variation amongst different items of a series, the more will be the dispersion.
- As per Bowley, *“Dispersion is a measure of the variation of the items”*.

OBJECTIVES OF MEASURING DISPERSION

- To determine the reliability of an average
- To compare the variability of two or more series
- For facilitating the use of other statistical measures
- Basis of Statistical Quality Control

PROPERTIES OF A GOOD MEASURE OF DISPERSION

- Easy to understand
- Simple to calculate
- Uniquely defined
- Based on all observations
- Not affected by extreme observations
- Capable of further algebraic treatment

MEASURES OF DISPERSION

- **Absolute**
  - Expressed in the same units in which data is expressed
  - Ex: Rupees, Kgs, Ltr, Km etc.
- **Relative**
  - In the form of ratio or percentage, so is independent of units
  - It is also called **Coefficient of Dispersion**

METHODS OF MEASURING DISPERSION

- Range
- Interquartile Range & Quartile Deviation
- Mean Deviation
- Standard Deviation
- Coefficient of Variation

RANGE (R)

- It is the simplest measures of dispersion
- It is defined as the difference between the largest and smallest values in the series

$$\boxed{R = L - S}$$

R = Range, L = Largest Value, S = Smallest Value

- Coefficient of Range $= \dfrac{L - S}{L + S}$

Practice Problems – Range

Q1: Find the range & Coefficient of Range for the following data: 20, 35, 25, 30, 15

Solution: R=**L – S =35-15 = 20**

Coefficient of Range = L – S / L +S =20/50 = 0.4

Q2: Find the range & Coefficient of Range:

X	10	20	30	40	50	60	70
F	15	18	25	30	16	10	9

Ans: 60, 0.75

Q3: Find the range & Coefficient of Range:

Size	5-10	10-15	15-20	20-25	25-30
F	4	9	15	30	40

Ans: 25, 5/7

Range

MERITS

- Simple to understand
- Easy to calculate
- Widely used in statistical quality control

DEMERITS

- Can’t be calculated in open ended distributions
- Not based on all the observations
- Affected by sampling fluctuations
- Affected by extreme values

Interquartile Range & Quartile Deviation

- ***Interquartile Range (IQR)*** is the difference between the upper quartile (Q_3) and the lower quartile (Q_1)
- It covers dispersion of middle 50% of the items of the series
- Symbolically, **Interquartile Range** = $Q_3 - Q_1$

- ***Quartile Deviation (QR)*** is half of the interquartile range. It is also called Semi Interquartile Range
- Symbolically, Quartile Deviation = $\frac{Q_3 - Q_1}{2}$

- ***Coefficient of Quartile Deviation***: It is the relative measure of quartile deviation.
- **Coefficient of Q.D.** = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

Practice Problems – IQR & QD

Q1: Find interquartile range, quartile deviation and coefficient of quartile deviation:

28, 18, 20, 24, 27, 30, 15 *Ans: 10, 5, 0.217*

Q2:

X	10	20	30	40	50	60
F	2	8	20	35	42	20

Ans: 10, 5, 0.11

Q3:

Age	0-20	20-40	40-60	60-80	80-100
Persons	4	10	15	20	11

Ans: 14.33, 0.19

$\bar{x}$

MEAN DEVIATION (M.D.)

- It is also called Average Deviation
- It is defined as the arithmetic average of the deviation of the various items of a series computed from measures of central tendency like mean or median or mode.

$$\text{M.D. from Mean} = \text{MDM} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

$$\text{Coefficient of M.D. From mean} = \frac{MDM}{Mean}$$

If frequency is given:

$$\text{M.D.from Mean} = \text{MDM} = \frac{\sum_{i=1}^{n} f_i \,|\, x_i - \bar{x} \,|}{\sum f_i}$$

Example:

Find the mean deviation about the mean for the following data:

6, 7, 10, 12, 13, 4, 8, 12

Mean of the given data = $\frac{\textit{Sum of all terms}}{\textit{Total number of terms}}$

$$\bar{x} = \frac{6+7+10+12+13+4+8+12}{8}$$

$$= \frac{72}{8}$$

$$= 9$$

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
6	$6 - 9 = -3$	$\lvert -3 \rvert = 3$
7	$7 - 9 = -2$	$\lvert -2 \rvert = 2$
10	$10 - 9 = 1$	$\lvert 1 \rvert = 1$
12	$12 - 9 = 3$	$\lvert 3 \rvert = 3$
13	$13 - 9 = 4$	$\lvert 4 \rvert = 4$
4	$4 - 9 = -5$	$\lvert -5 \rvert = 5$
8	$8 - 9 = -1$	$\lvert -1 \rvert = 1$
12	$12 - 9 = 3$	$\lvert 3 \rvert = 3$
		$\sum_{1}^{8} \lvert x_i - \bar{x} \rvert = 22$

Mean deviation about mean

$$= \frac{\sum \lvert x_i - \bar{x} \rvert}{8}$$

$$= \frac{22}{8}$$

$$= 2.75$$

Example:

Find mean deviation about the mean for the following data :

x_i	2	5	6	8	10	12
f_i	2	8	10	7	8	5

First we will calculate mean

x_i	f_i	$f_i x_i$	$\|x_i - \bar{x}\|$	$f_i\|x_i - \bar{x}\|$
2	2	$2 \times 2 = 4$	$\|2 - 7.5\| = \|-5.5\| = 5.5$	$2 \times 5.5 = 11$
5	8	$5 \times 8 = 40$	$\|5 - 7.5\| = \|-2.5\| = 2.5$	$8 \times 2.5 = 20$
6	10	$6 \times 10 = 60$	$\|6 - 7.5\| = \|-1.5\| = 1.5$	$10 \times 1.5 = 15$
8	7	$8 \times 7 = 56$	$\|8 - 7.5\| = \|0.5\| = 0.5$	$7 \times 0.5 = 3.5$
10	8	$10 \times 8 = 80$	$\|10 - 7.5\| = \|2.5\| = 2.5$	$8 \times 2.5 = 20$
12	5	$12 \times 5 = 60$	$\|12 - 7.5\| = \|4.5\| = 4.5$	$5 \times 4.5 = 22.5$
	$\sum f_i = 40$	$\sum f_i x_i = 300$		$\sum f_i\|x_i - \bar{x}\| = 92$

$$\text{Mean}(\bar{x}) = \frac{\sum f_i x_i}{\sum f_i}$$

$$\bar{x} = \frac{300}{40}$$

$$\bar{x} = 7.5$$

$$\text{Mean deviation about mean} = \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i}$$

Putting $\sum f_i |x_i - \bar{x}| = 92$, $\sum f_i = 40$

$$\text{M.D. } (\bar{x}) = \frac{1}{40} \times 92$$

$$= 2.3$$

Mean Deviation

Merits

- Simple to understand
- Easy to compute
- Less effected by extreme items
- Useful in fields like Economics, Commerce etc.
- Comparisons about formation of different series can be easily made as deviations are taken from a central value

Demerits

- Ignoring '±' signs are not appropriate
- Not accurate for Mode
- Difficult to calculate if value of Mean or Median comes in fractions
- Not capable of further algebraic treatment
- Not used in statistical conclusions.

Standard Deviation

- Most important & widely used measure of dispersion
- First used by Karl Pearson in 1893
- Also called root mean square deviations
- It is defined as the square root of the arithmetic mean of the squares of the deviation of the values taken from the mean
- Denoted by σ (sigma)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

- Coefficient of S.D. = $\frac{\sigma}{\bar{x}}$

- **Variance** = $(\text{S.D.})^2 = \sigma^2$

Calculation of Standard Deviation

Individual Series

- Actual Mean Method

Discrete Series

- Actual Mean Method

Continuous Series

- Actual Mean Method

Standard Deviation – Individual Series Actual Mean Method

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

Find the Standard deviation and Variance of the following data:6, 8, 10, 12, 14, 16, 18, 20, 22, 24.

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
6	-9	81
8	-7	49
10	-5	25
12	-3	9
14	-1	1
16	1	1
18	3	9
20	5	25
22	7	49
24	9	81
150		330

$$\text{Mean} = \frac{\text{Sum of observation}}{\text{Total Number of observation}}$$

Mean= 150/10 =15

$$\text{S.D.} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{330}{10}} = 5.74$$

Variance=(S.D.)^2=33

Standard Deviation – Discrete Series Actual Mean Method

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{N}}$$

STANDARD DEVIATION – CONTINUOUS SERIES

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{N}}$$

Example : Discrete Series

Find the variance and standard deviation for the following data:

First we will calculate mean

x_i	f_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
4	3	$4 - 14 = -10$	$(-10)^2 = 100$	$3 \times 100 = 300$
8	5	$8 - 14 = -6$	$(-6)^2 = 36$	$5 \times 36 = 180$
11	9	$11 - 14 = -3$	$(-3)^2 = 9$	$9 \times 9 = 81$
17	5	$17 - 14 = 3$	$(3)^2 = 9$	$5 \times 9 = 45$
20	4	$20 - 14 = 6$	$(6)^2 = 36$	$4 \times 36 = 144$
24	4	$24 - 14 = 10$	$(10)^2 = 100$	$4 \times 100 = 400$
32	1	$32 - 14 = 18$	$(18)^2 = 324$	$1 \times 324 = 324$
	$\sum f_i = 30$			$\sum f_i(x_i - \bar{x})^2 = 1374$

$$\text{Mean}(\bar{x}) = \frac{\sum f_i x_i}{\sum f_i}$$

$$\bar{x} = \frac{420}{30}$$

$$\bar{x} = 14$$

Now, finding variance

$$\sum f_i(x_i - \bar{x})^2 = 1374$$

$$\sum f_i = 30$$

$$\text{Variance } (\sigma^2) = \frac{\sum f_i(x_i - \bar{x})^2}{\sum f_i}$$

$$= \frac{1374}{30}$$

$$= 45.8$$

$$\text{Standard deviation } (\sigma) = \sqrt{45.8}$$

$$= 6.76$$

Example :continuous series

Calculate the mean, variance and standard deviation for the following distribution :

Class	Frequency (f_i)	Mid – point (x_i)	$f_i x_i$
30 – 40	3	35	35 × 3 = 105
40 – 50	7	45	45 × 7 = 315
50 – 60	12	55	55 × 12 = 660
60 – 70	15	65	65 × 15 = 975
70 – 80	8	75	75 × 8 = 600
80 – 90	3	85	85 × 3 = 255
90 – 100	2	95	95 × 2 = 190
	$\sum f_i = 50$		$\sum f_i x_i = 3100$

$\sum f_i x_i = 3100$

$\sum f_i = 50$

$$\text{Mean } (\overline{x}) = \frac{\sum f_i x_i}{\sum f_i}$$

$$= \frac{3100}{50}$$

$$= 62$$

Continue Example:

Finding Variance and Standard Deviation

Class	Frequency (f_i)	Mid – point (x_i)	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
30 – 40	3	35	$(35 - 62)^2 = 729$	$3 \times 729 = 2187$
40 – 50	7	45	$(45 - 62)^2 = 289$	$7 \times 289 = 2023$
50 – 60	12	55	$(55 - 62)^2 = 49$	$12 \times 49 = 588$
60 – 70	15	65	$(65 - 62)^2 = 9$	$15 \times 9 = 135$
70 – 80	8	75	$(75 - 62)^2 = 169$	$8 \times 169 = 1352$
80 – 90	3	85	$(85 - 62)^2 = 529$	$3 \times 529 = 1589$
90 – 100	2	95	$(95 - 62)^2 = 1089$	$2 \times 1089 = 2187$
	$\sum f_i = 50$			Sum = 10050

$$\sum f_i(x_i - \bar{x})^2 = 10050$$

$$\sum f_i = 50$$

$$\text{Variance } (\sigma^2) = \frac{1}{N}\sum f_i(x_i - \bar{x})^2$$

$$= \frac{1}{50} \times 10050$$

$$= \mathbf{201}$$

$$\text{Standard deviation } (\sigma) = \sqrt{201}$$

$$\mathbf{(\sigma) = 14.17}$$

Combined Standard Deviation

- It is the combined standard deviation of two or more groups as in case of combined arithmetic mean
- It is denoted by σ_{12}

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

where σ_{12} = Combined SD

σ_1 = SD of first group

σ_2 = SD of second group

$d_1 = \overline{X_1} - \overline{X_{12}}$

$d_2 = \overline{X_2} - \overline{X_{12}}$

$\overline{X_{12}}$ = Combined Mean

Example: Two samples of sizes 100 & 150 respectively have means 50 & 60 and SD 5 & 6. Find the Combined Mean & Combined Standard Deviation.

Solution: Given that:

$n_1=100$ and $n_2=150$

$\overline{x_1} = 50$

$\overline{x_2} = 60$

$\sigma_1 = 5$

$\sigma_2 = 6$

$$\overline{x_{12}} = \frac{n_1\overline{x_1} + n_2\overline{x_2}}{n_1 + n_2} = 56$$

$$d_1 = \overline{x_1} - \overline{x_{12}} = -6$$

$$d_2 = \overline{x_2} - \overline{x_{12}} = 4$$

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

$$\sigma_{12} = 7.46$$

Coefficient of Variation (C.V.)

- It was developed by Karl Pearson.
- It is an important relative measure of dispersion.
- It is used in comparing the variability, homogeneity, stability, uniformity & consistency of two or more series.
- Higher the CV, lesser the consistency.

$$C.V. = \frac{\sigma}{\overline{X}} * 100$$

Practice Problems

Q1: The scores of two batsmen A & B in ten innings during a certain match are:

A	32	28	47	63	71	39	10	?	96	14
B	19	31	48	53	67	90	10	62	40	80

Which batsmen have minimum coefficient of variation?

Ans: B

Q2: Sum of squares of items is 2430 with mean 7 & N = 12. Find coefficient of variation. *Ans: 176.85%*