Central Tendency and Dispersion

Past Trends

Attempt	Practical	Theory	Total
May 2018	3	4	7
Nov 2018	10	2	12
Jun 2019	9	3	12
Nov 2019	10	7	17
Nov 2020	4	5	9
Jan 2021	4	5	9
Jul 2021	11	1	12
Dec 2021	6	5	11
Jun 2022	6	3	9
Dec 2022	13	3	16

Central Tendency - Basics

Meaning	• Central Tendency is the **tendency** of a given set of observations to **cluster** around a single **central or middle value.** • The **single value** that **represents** the given set of observations is described as a measure of central tendency.
Different Measures of Central Tendency	• Arithmetic Mean (AM) • Median (Me) • Mode (Mo) • Geometric Mean (GM) • Harmonic Mean (HM)
Types of Formula based Questions	• Discrete Observations • Simple Frequency Distribution • Grouped Frequency Distribution

Arithmetic Mean

Discrete Observations	$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$ $\quad$ $\bar{x} = \frac{\sum x}{n}$
Frequency Distribution	$\bar{x} = \frac{\sum fx}{N}$ In case of simple frequency distribution: x = individual values In case of grouped frequency distribution: x = mid-point of class intervals N = total number of observations: $N = \sum f$
Assumed Mean / Step-Deviation Method	AM using assumed mean / step deviation method $\bar{x} = A + \frac{\sum fd}{N} \times C$

	where $d=\frac{x-A}{C}$, A is assumed mean, C is class length
Property 1	If all the observations are constant, AM is also constant
Property 2	the algebraic sum of deviations of a set of observations from their AM is zero
Property 3	AM is affected both due to change of origin and scale If $y = a + bx$ then $\overline{y} = a + b\overline{x}$
Property 4	Combined AM $\overline{x}_c = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$
General Review	• AM is best measure of central tendency • AM is based on all observations • AM is affected by sampling fluctuations • AM is amenable to mathematical property • AM cannot be used in case of open end classification

PYQ May 18 If the variables x and z are related by z = ax + b where a and b are constant, then $\overline{z} = a + b\overline{x}$

a. True b. False c. Both d. None

Ans: a

PYQ May 18 If each item is reduced by 15 then AM is

a. Reduced by 15 b. Increased by 15

c. Reduced by 10 d. None

Ans: a

PYQ Nov 18 The mean of 20 items of a data is 5 and if each item is multiplied by 3, then the new mean will be

a. 5 b. 10 c. 15 d. 20

Ans: c

PYQ Nov 18 The algebraic sum of the deviations of a set of values from their AM is

a. >0 b. =0 c. <0 d. None

Ans: b

PYQ Nov 18 If the frequencies of three series are 50, 60 and 90 and their means are 12, 15 and 20 respectively, then the mean of their composite series is

a. 16 b. 15.5 c. 16.5 d. 14.5

Ans: c

PYQ Jun 19 The AM of 15 observations is 9 and AM of first 9 observations is 11 then the AM of remaining observations is

a. 11 b. 6 c. 5 d. 9

Ans: b

Extra If assumed mean is 419.5 and sum of product of frequency and deviation from assumed mean is -43. Find the AM (given class length of distribution is 20)

a. 397.51 b. 410.66 c. 416.71 d. 432.55

Ans: c

PYQ Jul 21 The mean of n observations is x. If k is added to each observation, then the new mean is

a. k b. xk c. x-k d. x+k

Ans: d

PYQ Dec 21 If there are 3 observations 15, 20 and 25 then the sum of deviation of the observations from their AM is

a. 0 b. 5 c. -5 d. 10

Ans: a

PYQ Dec 21 If average mark for a group of 30 girls is 80 and for group of boys is 70 and combined average is 76, then how many are in the boy's group?

a. 21 b. 20 c. 22 d. 19

Ans: b

PYQ Jun 22 The mean of 20 observations is 38. If two observations are taken as 84 and 36 instead of 48 and 63 find new means.

e. 38.45 f. 41.15 g. 37.55 h. 40.05

Ans: c

Median

<table>
<tr><td>Discrete Observations</td><td>• If n = odd, then middle term
• If n = even, average of two middle terms</td></tr>
<tr><td>Simple Frequency Distribution</td><td>• First make column of less than cumulative frequency
• Apply same formula as discrete</td></tr>
<tr><td>Grouped Frequency Distribution</td><td>Median in case of grouped frequency distribution
<table>
<tr><td>Step 1</td><td>Prepare a less than type cumulative frequency distribution</td></tr>
<tr><td>Step 2</td><td>Calculate $\frac{N}{2}$ and check between which class boundaries it falls and call it as Median Class</td></tr>
<tr><td>Step 3</td><td><table><tr><td>l_1</td><td>N_u</td><td>N_l</td><td>C</td></tr><tr><td>LCB of Median Class</td><td>Cum Freq. of Median Class</td><td>Cum. Freq. of Pre-Median Class</td><td>Class length of Median Class</td></tr></table></td></tr>
<tr><td>Step 4</td><td>Apply Formula
$Me = l_1 + \left(\frac{\frac{N}{2} - N_l}{N_u - N_l} \right) \times C$</td></tr>
</table></td></tr>
<tr><td>Property 1</td><td>For a set of observations, the sum of absolute deviations is minimum, when the deviations are taken from the median.
$\sum(x - \bar{x}) = 0$</td></tr>
<tr><td>Property 2</td><td>Median is also affected by both change of origin and scale.</td></tr>
<tr><td>General Review</td><td>• Median is also called as positional average
• Median is not based on all observations
• Median is not affected by sampling fluctuations
• Median is best measure of central tendency in case of open-end classification</td></tr>
</table>

PYQ Nov 18
PYQ Dec 21

The median of the data 5, 6, 7, 7, 8, 9, 10, 11, 11, 12, 15, 18, 18 and 19 is

a. 10.5 b. 10 c. 11 d. 11.5

Ans: a

PYQ Jun 19

Find median for the below distribution:

X	1	2	3	4	5	6
F	6	9	10	14	12	8

a. 3.5 b. 3 c. 4 d. 5

Ans: c

PYQ Nov 19

The deviations are minimum when taken from

a. Mean b. Median c. Mode d. None

Ans: b

PYQ Nov 19

Find the median of the following:

Class	0-10	10-20	20-30	30-40	40-50
Frequency	2	3	4	5	6

a. 35 b. 32 c. 36 d. 37.5

Ans: b

Partition Values

Meaning

- These may be defined as **values dividing** a given **set of observations** into number of **equal parts**
- When we want to divide the given set of observations into two equal parts, we consider median, similarly there are quartiles, deciles, percentiles

Name of PV	No. of equal parts	No. of PVs	Symbol
Median	2	1	Me
Quartile	4	3	Q_1, Q_2, Q_3
Decile	10	9	$D_1, D_2, \ldots, D_9$
Percentile	100	99	$P_1, P_2, \ldots, P_{99}$

Formula – Discrete Observations

- Rank Calculation $(n+1)p^{th}$ term
- Value of p depends on partition value

#	Median	Quartile	Decile	Percentile
First	1/2	1/4	1/10	1/100
Second		2/4	2/10	2/100
...				
Last		3/4	9/10	99/100

Quartiles Grouped FD

Quartiles in case of Grouped Frequency Distribution: Steps are like median with few modifications.

1st Quartile	3rd Quartile
Find Q_1 class using $\frac{N}{4}$	Find Q_3 class using $\frac{3N}{4}$
$Q_1 = l_1 + \left(\frac{\frac{N}{4} - N_l}{N_u - N_l}\right) \times C$	$Q_3 = l_1 + \left(\frac{\frac{3N}{4} - N_l}{N_u - N_l}\right) \times C$

Deciles Grouped FD

Deciles in case of Grouped Frequency Distribution: Steps are like median with few modifications.

1st Decile	9th Decile
Find D_1 class using $\frac{N}{10}$	Find D_9 class using $\frac{9N}{10}$
$D_1 = l_1 + \left(\frac{\frac{N}{10} - N_l}{N_u - N_l}\right) \times C$	$D_9 = l_1 + \left(\frac{\frac{9N}{10} - N_l}{N_u - N_l}\right) \times C$

Percentiles Grouped FD

Percentiles in case of Grouped Frequency Distribution: Steps are like median with few modifications.

1st Percentile	99th Percentile
Find P_1 class using $\frac{N}{100}$	Find P_{99} class using $\frac{99N}{10}$
$P_1 = l_1 + \left(\frac{\frac{N}{100} - N_l}{N_u - N_l}\right) \times C$	$P_{99} = l_1 + \left(\frac{\frac{99N}{10} - N_l}{N_u - N_l}\right) \times C$

PYQ Nov 19

For 899, 999, 391, 384, 390, 480, 485, 760, 111, 240
Rank of Median is

a. 2.75 b. 5.5 c. 8.25 d. None

Ans: b

PYQ Nov 20

50th Percentile is equal to

a. Median b. Mode c. Mean d. None

Ans: a

PYQ Jun 22

The 3rd Decile for the numbers
15, 10, 20, 25, 18, 11, 9, 12 is

a. 13 b. 10.7 c. 11 d. 11.5

Ans: b

PYQ Jun 19

The QD of 6 numbers 15, 8, 36, 40, 38, 41 is equal to

a. 12.5 b. 25 c. 13.5 d. 37

Ans: c

Mode

Meaning	Mode is the **value** that **occurs the maximum** number of times
Special Thing about Mode	• If two or more observations are having maximum frequency then there are **multiple modes** [multimodal distribution] • If there are **exactly two** modes then distribution is called as **Bimodal** Distribution • If all observations are having same frequency then distribution has **no mode** • We can say that Mode is **not rigidly defined**
Grouped Frequency Distribution	• Find Modal Class: Class with highest frequency and obtain below values f_{-1}: frequency of pre modal class; f_0: frequency of the modal class; f_1: frequency of the post modal class • Apply Formula $$Mo = l_1 + \left(\frac{f_0 - f_{-1}}{2f_0 - f_{-1} - f_1}\right) \times C$$
Property 1	If all the observations are constant, mode is also constant
Property 2	Mode is also affected both due to change of origin and scale
General Review	• Mode is not based on all observations • Mode is not rigidly defined • Mode is not amenable to Mathematical Property

PYQ Nov 19

Find the mode of the following data:

Class	3-6	6-9	9-12	12-15	15-18	18-21
Frequency	2	5	10	23	21	12

a. 25 b. 4.6 c. 14.6 d. 13.5

Ans: c

PYQ Jan 21
PYQ Jul 21

If y = 3 + 4.5x and mode for x is 20, then the mode for y is

a. 3.225 b. 12 c. 24.5 d. 93

Ans: d

Relationship between Mean, Median and Mode

In case of Symmetric Distribution	**Mean = Median = Mode**
In case of Moderately Skewed Distribution (Empirical relationship)	**Mean – Mode = 3 (Mean – Median)**

PYQ May 18

Relation between Mean, Median and Mode is

a. Mean – Mode = 2 (Mean – Median)
b. Mean – Median = 3 (Mean – Mode)
c. Mean – Median = 2 (Mean – Mode)
d. Mean – Mode = 3 (Mean – Median)

Ans: d

PYQ Nov 18
PYQ Jun 19
PYQ Dec 21

If in a moderately skewed distribution, the values of mode and mean are 32.1 and 35.4 respectively then the value of median is

a. 34.3 b. 33.3 c. 34 d. 33

Ans: a

PYQ Jun 19

For a symmetric distribution

a. Mean = Median = Mode
b. Mode = 3 Median – 2 Mean
c. Mode = 1/3 Median = 1/2 Mean
d. None

Ans: a

PYQ Dec 21

For a moderately skewed distribution, the median is twice the mean, then the mode is _____ times the median.

a. 3 b. 2 c. 2/3 d. 3/2

Ans: b

Geometric Mean

Definition	For a given set of n **positive observations**, the geometric mean is defined as the n^{th} root of the product of the observations
Formula – Discrete	$G = \left(x_1 \times x_2 \times \ldots \times x_n\right)^{1/n}$
Formula – Frequency Distribution	$G = \left(x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_n^{f_n}\right)^{1/N}$
Property 1	Logarithm of G for a set of observations is the AM of the logarithm of the observations $\log G = \frac{1}{n}\sum \log x$
Property 2	If all the observations are constant, GM is also constant
Property 3	$\text{GM of } z = \text{GM of } x \times \text{GM of } y$
Property 4	$\text{GM of } z = \frac{\text{GM of } x}{\text{GM of } y}$

PYQ Nov 18

The GM of 3, 6, 24, 48 is

a. 8 b. 12 c. 24 d. 6

Ans: b

PYQ Dec 21

If two variables are related by c = ab then GM of c is equal to

a. GM of a + GM of b
b. GM of a × GM of b
c. GM of a – GM of b
d. GM of a/ GM of b

Ans: b

Harmonic Mean

Definition	For a given set of **non-zero** observations, harmonic mean is defined as the **reciprocal of the AM of the reciprocals of the observation**
Formula – Discrete	$H = \frac{n}{\Sigma(\frac{1}{x})}$
Formula – Frequency Distribution	$H = \frac{N}{\Sigma(\frac{f}{x})}$
Property 1	If all observations are constant HM is also constant
Property 2	$\text{Combined HM} = \frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$

PYQ Nov 20

Given the weights for the numbers 1, 2, 3, …, n is respectively $1^2, 2^2, 3^2, \ldots, n^2$ then weighted HM is

a. $\frac{2n+1}{4}$ b. $\frac{2n+1}{6}$ c. $\frac{2n+1}{3}$ d. $\frac{2n+1}{2}$

Ans: c

PYQ Nov 20

The HM of A and B is 1/3 and HM of C and D is 1/5. The HM of A, B, C and D is

a. 8/15 b. 1/4 c. 1/15 d. 5/3

Ans: b

PYQ Jan 21

If there are two groups with n_1 and n_2 observations and H_1 and H_2 are respective HMs, then HM of combined observations is

a. $\frac{n_1H_1 + n_2H_2}{n_1 + n_2}$ b. $\frac{n_1H_1 + n_2H_2}{H_1 + H_2}$

c. $\frac{n_1 + n_2}{n_1H_1 + n_2H_2}$ d. $\frac{(n_1 + n_2)H_1H_2}{n_1H_2 + n_2H_1}$

Ans: d

Use of GM and HM

Both	Both are used for calculating average rates
GM	Appropriate for rates having percentages
HM	Appropriate for rates other than percentages

PYQ Nov 20

A fire engine rushes to a place of fire accident with a speed of 110kmph and after the completion of operation returned to the base at a speed of 35kmph. The average speed per hour in per direction is obtained as _____ of speeds

a. AM b. GM c. HM d. None

Ans: c

Relationship between AM, GM, and HM

Relation	**Scenario** / **Relation** When all the observations are same / AM = GM = HM When observations are distinct / AM > GM > HM In question is silent / $AM \geq GM \geq HM$
Special Relation	If there are only two observations: $AM \times HM = (GM)^2$

PYQ Nov 20 If the AM and HM of two numbers are 6 and 9 respectively, then GM is

a. 7.35 b. 8.5 c. 6.75 d. None

Ans: a

Weighted Average

When to use	If the observations are not of equal importance and we need to treat observations according to their hierarchical importance, then we use Weighted Average
Formulas	Weighted AM: $\frac{\sum wx}{\sum w}$ Weighted GM: $\left(x_1^{w_1} \times x_2^{w_2} \times x_3^{w_3} \times \ldots \times x_n^{w_n}\right)^{\frac{1}{\sum w}}$ Weighted HM: $\frac{\sum w}{\sum\left(\frac{w}{x}\right)}$

Measures of Dispersion

Meaning of Measure of Dispersion	• Dispersion for a given set of observations may be defined as • the **amount of deviation** of the observations, • usually, from an **appropriate** measure of **central tendency**	
Types of Measure of Dispersion	Absolute Measures of Dispersion	• These are with units • These are not useful for comparison of two variables with different units. • Example: Range, Mean Deviation, Standard Deviation, Quartile Deviation
	Relative Measures of Dispersion	• These are unit free measures • These are useful for comparison of two variables with different units. • Example: Coefficient of Range, Coefficient of Mean Deviation, Coefficient of variation, Coefficient of Quartile Deviation

Range

Discrete – Formula	$L - S$ where L: Largest Observation, S: Smallest Observation
Grouped Frequency Distribution – Formula	$L - S$ where Largest Observation = **UCB** of last class interval, Smallest Observation = **LCB** of first-class interval
Coefficient of Range	$\frac{L-S}{L+S} \times 100$
Property 1	• **Not affected** by change of **origin** • Affected by **change of scale (only value)** • **No impact of sign** of change of scale • Note: **Measure of Dispersion can never be negative**
General Review	• Not Based on All Observations • Easy to Compute

PYQ Nov 18 If the range of a set of values is 65 and maximum value in the set is 83, then the minimum value in the set is

a. 74 b. 9 c. 18 d. None

Ans: c

PYQ Jul 21 If the relationship between x and y is given by 2x + 3y = 10 and the range of y is 10, then what is the range of x?

a. 10 b. 18 c. 8 d. 15

Ans: d

PYQ Dec 21 The marks secured by 5 students in a subject are 82, 73, 69, 84, 66. What is the coefficient of Range

a. 0.12 b. 12 c. 120 d. 0.012

Ans: b

Mean Deviation

Meaning	• Mean deviation is defined as the • **arithmetic mean** of the • **absolute deviations** of the observations • from an **appropriate measure** of central tendency
Formula – Discrete	$MD_A = \frac{1}{n}\Sigma\lvert x - A\rvert$ where, A = Appropriate Central Tendency Measure
Formula – Frequency Distribution	$MD_A = \frac{1}{N}\Sigma f\lvert x - A\rvert$
Coefficient of Mean Deviation	Coefficient of Mean Deviation: $\frac{\text{Mean Deviation about A}}{A} \times 100$
Property 1	Mean Deviation takes its **minimum value** when deviations are taken from **Median**
Property 2	Change of Origin – **No Affect**, Change of Scale – **Affect of value not sign**
General Review	• Based on **all observations** • Improvement over Range • **Difficult to compute** • **Not amenable to Mathematical Property** because of usage of **Modulus**

PYQ Jan 21 Find the coefficient of mean deviation about mean for the data: 5, 7, 8, 10, 11, 13, 19

a. 17.28 b. 28.57 c. 32.11 d. 18.56

Ans: c

PYQ Jun 22 Mean Deviation about Mode from the data: 3, 10, 10, 4, 7, 18, 5

a. 4.39 b. 4.70 c. 4.14 d. 5.24

Ans: c

Standard Deviation

Meaning	• Improvement over Mean Deviation • It is defined as the **root mean square deviation** when the deviations are <u>taken from the AM</u> of the observations
Formula – Discrete	$\sigma_x = SD_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$ $\sigma_x = SD_x = \sqrt{\frac{\Sigma x^2}{n} - (\bar{x})^2}$

Formula – Frequency Distribution	$\sigma_x = SD_x = \sqrt{\frac{\sum f(x-\bar{x})^2}{N}}$ $\sigma_x = SD_x = \sqrt{\frac{\sum fx^2}{N} - (\bar{x})^2}$
Coefficient of Variation	$\frac{SD_x}{\bar{x}} \times 100$
SD for any two numbers	$SD = \frac{Range}{2}$
SD for first n natural numbers	$s = \sqrt{\frac{n^2-1}{12}}$
Property 1	If all the observations are constant, SD is **ZERO**
Property 2	No effect of change of origin but affected by change of scale in the magnitude (ignore sign)
Property 3	$SD_c = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$ $d_1 = \bar{x}_c - \bar{x}_1 \; d_2 = \bar{x}_c - \bar{x}_2$

PYQ May 18
PYQ Jun 19
If the SD of the 1st n natural number is $\sqrt{30}$
a. 19 b. 20 c. 21 d. None
Ans: a

PYQ Nov 18
If the variance of 5, 7, 9 and 11 is 4, then the coefficient of variation
a. 15 b. 25 c. 17 d. 19
Ans: b

PYQ Nov 18
Marks obtained by a student in monthly tests are 30, 35, 25, 20, 15. Find SD of marks
a. 25 b. $\sqrt{50}$ c. $\sqrt{30}$ d. 50
Ans: b

PYQ Jun 19
If variance is 100 and coefficient of variation is 20% then AM is
a. 60 b. 70 c. 80 d. 50
Ans: d

PYQ Jun 19
SD is _____ times of $\sqrt{MD \times QD}$
a. 2/3 b. 4/5 c. $\sqrt{\frac{15}{8}}$ d. $\sqrt{\frac{8}{15}}$
Ans: c

PYQ Jun 19 The sum of mean and SD of a series is a + b, if we add 2 to each observation of the series then the sum of mean and SD is

a. a + b + 2 b. 6 – a + b
c. 4 + a – b d. a + b + 4

Ans: a

PYQ Nov 19
PYQ Jan 21 Find SD of 1, 2, 3, 4, 5, 6, 7, 8, 9

a. 2.58 b. 60/9 c. 60/3 d. 3.20

Ans: a

PYQ Jul 21
PYQ Jun 22 If the numbers are 5, 1, 8, 7, 2 then the coefficient of variation is

a. 56.13% b. 59.13% c. 48.13% d. 44.13%

Ans: b

PYQ Jun 22 AM and Coefficient of variation of x is 10 and 40. What is the variance of 30-2x

a. 64 b. 56 c. 49 d. 81

Ans: a

Quartile Deviation

Formula	$QD_x = \frac{Q_3 - Q_1}{2}$
Calculation	Quartiles are calculated same as we studied in Central Tendency
Coefficient of Quartile Deviation	$\frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$
General Review	• It is the **best measure** of dispersion for **open-end** classification • It is also **less affected** due to sampling fluctuations • Like other measures of Dispersion, **QD** is also not affected by change of origin but affected by scale ignoring sign
Relationship between SD, MD and QD	$4SD = 5MD = 6QD$ or $SD : MD : QD = 15 : 12 : 10$

PYQ Jun 19 Coefficient of QD is 1/4 then Q_3 / Q_1 is

a. 5/3 b. 4/3 c. 3/4 d. 3/5

Ans: a