Ch-14 Measures of Central Tendency & Dispersion

measures of central tendency

- means (mathematical avg)
  - A.M
  - G.M
  - H.M
- Partitional values (positional value)
  - Median
  - Quartile
  - Decile
  - Percentile
- Mode

- data always tend to lie at a central location; different methods known as measures of central tendency

Measure	Individual series	Discrete Series	Continuous series	Properties
Mean: (1) Arithmetic mean ($\bar{x}$) (simple avg of given observations)	$\frac{\sum_{i=1}^{n} x_i}{n}$ or, $\frac{x_1+x_2+x_n}{n}$	$\bar{x} = \frac{\sum fx}{\sum f}$	m: mid point $\bar{x} = \frac{\sum fm}{\sum f}$	• If values are constant, say k, AM also k. eg: 2, 2, 2 AM = 2 • The algebraic sum of deviation of a set of observation from their AM is zero. $\sum(x-\bar{x})=0$ • AM is affected due to change in scale, as well as due to change in origin. (a) scale (b) origin $y=a+bx$, $\bar{y}=a+b(\bar{x})$ • combined mean($\bar{x}$) = $\frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}$
(2) Geometric mean (avg of rates & percentages)	$\sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot x_n}$	$\sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot x_n^{f_n}}$	$\sqrt[n]{m_1^{f_1} \cdot m_2^{f_2} \cdot m_n^{f_n}}$	• all values assumed by a variable ~~is~~ are constant say k, then GM is k. • Logarithm of GM for a set of observation is the AM of the log of observation
3) Harmonic mean (avg of speeds)	$\frac{n}{\sum(1/x)}$	$\frac{N}{\sum(f/x)}$	$\frac{N}{\sum(f/m)}$	• If all values assumed by a variable are constant, k, then HM is k. • Combined HM: $\frac{n_1+n_2}{\frac{n_1}{HM_1}+\frac{n_2}{HM_2}}$
Relationship bet^n AM, GM & HM	$AM \geq GM \geq HM$	$HM = \frac{GM^2}{AM}$		• To calculate avg speed, use H.M.

Measure	Individual series	Discrete series	Continuous series	Property
Partition values (Positional Avg) - divide a given series into equal parts.				
Median	$\left(\frac{N+1}{2}\right)^{th}$ item	$\left(\frac{N+1}{2}\right)^{th}$ item	Median class $= \frac{N}{2}^{th}$ item Median $= l + \frac{\frac{N}{2} - c.f}{f} \times i$	• affected due to change in scale as well as origin. $y = a + bx$, $y_{me} = a + bx_{me}$ • For a set of observation, the sum of absolute deviation is minimum when deviation calculated from median.
Quartile (1-3)	$Q_1 = \left(\frac{n+1}{4}\right)^{th}$ item $Q_2 = 2\left(\frac{N+1}{\cancel{2}4}\right)^{th}$ item $Q_3 = 3\left(\frac{N+1}{4}\right)^{th}$ item	$Q_1 = \left(\frac{N+1}{4}\right)^{th}$ item $Q_2 = 2\left(\frac{N+1}{4}\right)^{th}$ item $Q_3 = 3\left(\frac{N+1}{4}\right)^{th}$ item	$Q_1 = l + \frac{\frac{N}{4} - c.f}{f} \times i$ $Q_2 = l + \frac{\frac{2N}{4} - c.f}{f} \times i$ $Q_3 = l + \frac{\frac{3N}{4} - c.f}{f} \times i$	← Q_1 class $= \frac{N}{4}^{th}$ item Q_2 class $= \frac{2N}{4}^{th}$ item Q_3 class $= \frac{3N}{4}^{th}$ item 2.75^{th} item $= 2^{nd}$ item $+ 0.75$ ($3^{rd} - 2^{nd}$)
Percentile (1-99)	$P_n = \left(\frac{n+1}{100}\right)^{th}$ item	$P_n = \left(\frac{n+1}{100}\right)^{th}$ item	$P_n = l + \frac{\frac{n}{100} - c.f}{f} \times i$	P_n class $= \left(\frac{N}{100}\right)^{th}$ item
Decile (1-9)	$D_n = \left(\frac{n+1}{10}\right)^{th}$ item	$D_n = \left(\frac{n+1}{10}\right)^{th}$ item	$D_n = l + \frac{\frac{n}{100} - c.f}{f} \times i$	D_n class $= \left(\frac{N}{10}\right)^{th}$ item
Mode:	highest occurring value in the data.	$\cancel{l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i}$ highest frequency in data.	$l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$	• Mode is affected due to change in scale as well as due to change in origin.

Relationship betn mean, median & mode.

For symmetric data;

$$\text{Mean} = \text{Median} = \text{Mode}$$

For skew symmetric data; $\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median})$

$\text{Mode} = 3\,\text{Median} - 2\,\text{mean}$

For positively skewed data: $\text{Mean} > \text{Median} > \text{Mode}$

For negatively skewed data: $\text{Mean} < \text{Median} < \text{Mode}$

Mean	Median	Mode
• **Arithmetic mean** ÷ - best measure of Central Tendency - rigidly defined, based on all observations - easy to comprehend, simple to calculate - amenable to mathematical property - affected by sampling fluctuation - mean cannot be advocated for open end classification • **GM & HM** ÷ - rigidly defined, based on all observations - difficult to comprehend & compute - limited application for the computation of avg rates & ratio etc	- rigidly defined and easy to comprehend & compute. - not based on all observations - ~~not much~~ - not much affected by sampling fluctuation - most appropriate measure of central tendency for an open ended classification. Mode = 3median − 2mean	- most popular measure of central tendency. - no mathematical property - affected by sampling fluctuation.

Measures of dispersion : scatteredness of the data

Measure	Formula	Coefficient	Properties
Range :			
Individual series	[Largest value (L) − smallest value (S)]	$\frac{L-S}{L+S} \times 100$	- remains unaffected due to change in origin but affected in the same ratio due to change in a scale.
continuous series (exclusive)	[Upper most class boundary − Lower most class boundary]	$\frac{UMCB-LMCB}{UMCB+LMCB} \times 100$	$R_y = \lvert b \rvert \times R_x$ - All observation assumed by a variable are constant, Range is 0.
Mean deviation : (mean of deviation (difference)) Avg of absolute deviation from an appropriate measure of central tendency.	Mean deviation (M.D) $\frac{\sum \lvert x-\bar{x} \rvert}{n}$ (about median) $\frac{\sum \lvert x - Md \rvert}{N}$ (mode) $\frac{\sum \lvert x-\bar{\bar{Z}} \rvert}{N}$	$\frac{M.D \text{ about } \bar{x}}{\bar{x}} \times 100$ $\frac{M.D \text{ about } Md}{Md} \times 100$ $\frac{M.D \text{ about } \bar{\bar{Z}}}{Md} \times 100$	- unaffected due to change in origin but affected in the same ratio due to a change in scale. $MD_y = \lvert b \rvert \cdot MD_x$ - M.D is minimum when deviation is calculated from median. - All observation are constant, MD = 0

Measures	Formula	Coefficient	Properties
Standard deviation (square root of variance) [variance = deviation ka square ka mean] absolute deviation $= \lvert x-\bar{x} \rvert$	Individual: S.D (σ) $= \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n}}$ S.D $= \sqrt{\dfrac{\Sigma x^2}{N} - (\bar{x})^2}$ discrete / Continuous: S.D $= \sqrt{\dfrac{\Sigma f(x-\bar{x})^2}{N}}$ S.D $= \sqrt{\dfrac{\Sigma f(x)^2}{N} - (\bar{x})^2}$	Coefficent of variance: $\dfrac{S.D}{A.M} \times 100$ • S.D of two numbers (a & b) $\dfrac{\lvert a-b \rvert}{2}$ • S D of n natural no. $\sigma = \sqrt{\dfrac{n^2-1}{12}}$	• If all observation assumed by a variable are constant, S.D is zero. • S.D remains unaffected due to change in origin but affected in the same ratio due to the change in scale. (s) $S.D_y = \lvert b \rvert\, S.D_x$ • Combined S.D S.D $= \sqrt{\dfrac{n_1 S_1^2 + n_2 S_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1+n_2}}$
Quartile deviation	Q.D $= \dfrac{Q_3-Q_1}{2}$	$\dfrac{Q_3-Q_1}{Q_3+Q_1} \times 100$	• Most appropriate for open ended series • unaffected due to change in origin but affected due to change in scale. $QD_y = \lvert b \rvert\, QD_x$