MEASURES
OF
CENTRAL TENDENCY

Jose M C

Measures of Central Tendency

When we have a set of observations on some variable, a natural phenomenon can be observed within the data set that most of the observations are clustered around some central value. This tendency of observations to show a clustering around some central value is called central tendency. The so called central value around which the observations are clustered is called measure of central tendency or average or measures of location.

Measures of Central Tendency

According to Simpson and Kafka a measure of central tendency is ‘a typical value around which other figures aggregate’.

According to Croxton and Cowden an average is ‘a single value within the range of the data that is used to represent all the values in the series’. Since an average is somewhere within the range of data, it is sometimes called ‘a measure of central value’.

Measures of Central Tendency

According to Prof Bowley "Measures of central tendency (averages) are statistical constants which enable us to comprehend in a single effort the significance of the whole."

In general terms, central tendency is a statistical measure that determines a single value that accurately describes the centre of the distribution and represents the entire distribution of scores.

Measures of Central Tendency

The goal of central tendency is to identify the single value that is the best representative for the entire set of data. The following are the commonly used average or central tendency:

- Mean or Arithmetic mean or simple mean
- Geometric mean
- Harmonic mean
- Median
- Mode

Arithmetic mean, Geometric mean and Harmonic means are usually called mathematical averages

Mode and Median are called positional averages.

Measures of Central Tendency

Desirable properties of a good measure of Central Tendency

1. Simplicity: The fundamental feature of the average is that it should be easy to calculate and simple to follow.

2. Representation: Average should represent the entire mass of data.

3. Rigidly Defined: Averages should be rigidly defined. If it is so, instability in its value will be no more and would always be a definite figure.

4. Algebraic Treatment: Averages should always be amenable or capable of further algebraic treatment.

5. Clear and Stable Definition: A good average should have a clear and stable definition.

Measures of Central Tendency

Desirable properties of a good measure of Central Tendency - Contd

6. Absolute Number: A good average should be an absolute number.

7. Effect of fluctuations of Sampling: A good average should not be affected by fluctuations of sampling. In other words, average calculated from different samples from a population should be equal.

8. Based on all values of a variable: An average is said to be a true representative only when it is based on all the values of a variable.

9. No Effect of Extreme values: For a good average, it should not be unduly affected by extreme values. If it is so, it will not be a true representative.

10. Value can be found by Graphic Method: A good average is one which can found by arithmetic as well as graphic method.

12. Possible to find Central Tendency for open ended classes: A good average is one which can be calculated even when the class intervals are open ended.

Measures of Central Tendency

OBJECTIVES : The main aim of this unit is to study the frequency distribution. After going through this unit one should be able to

- describe measures of central tendency ;
- calculate

  Mean or Arithmetic Mean

  Median

  Mode

  Geometric Mean

  Harmonic Mean

Measures of Central Tendency

Arithmetic Mean or Mean

Arithmetic mean of a set of observations is their sum divided by the number of observations.

Arithmetic mean = $\dfrac{Sum\ of\ observations}{Number\ of\ observations}$

Case (1): Arithmetic mean from raw data

Let $x_1, x_2, x_3, \ldots \ldots \ldots . x_n$ be n observations. Then the Arithmetic mean denoted by $\bar{x}$ is defined as $\bar{x} = \dfrac{x_1 + x_2 + \ \ldots\ldots\ldots + x_n}{n} = \dfrac{\sum_{i=1}^{i=n} x_i}{n} = \dfrac{\sum x}{n}$

Calculate the A. M of the following observations:
12, 28, 36,14,25,20,15,40, 33,17

Answer: $\bar{x} = \dfrac{x_1 + x_2 + \ \ldots\ldots\ldots + x_n}{n} = \dfrac{\sum x}{n}$

$\bar{x} = \dfrac{12+28+36+14+25+20+15+40+33+17}{10} = \dfrac{240}{10} = \mathbf{24}$

$\bar{x} = \dfrac{x_1 + x_2 + \ \ldots\ldots\ldots + x_n}{n} = \dfrac{\sum x}{n}$

Sum of observations $\sum x$ = n $\bar{x}$

Measures of Central Tendency

Arithmetic Mean or Mean

Case (2): Arithmetic mean from ungrouped frequency distribution

Let $x_1, x_2, x_3, \dots\dots\dots. x_n$ be n observations and let $f_1, f_2, f_3, \dots\dots\dots. f_n$ be the corresponding frequencies. Then the Arithmetic mean denoted by $\bar{x}$ is defined as,

$$\bar{x} = \frac{f_1x_1+f_2x_2+\ \dots\dots\dots+f_nx_n}{f_1+f_2+\ \dots\dots\dots+f_n}$$

$$= \frac{\sum_{i=1}^{i=n} f_ix_i}{\sum_{i=1}^{i=n} f_i}$$

$$= \frac{\sum_{i=1}^{i=n} f_ix_i}{N} = \frac{\sum f_ix_i}{N}$$

$$\bar{x} = \frac{f_1x_1+f_2x_2+\ \dots\dots\dots+f_nx_n}{f_1+f_2+\ \dots\dots\dots+f_n} = \frac{\sum f.x}{N}$$

Where, N is the total frequency or the total number of observations in the data set.

n stands for the number of observations in raw data

and

N stands for the total number of observations in frequency distribution (Ungrouped/ Grouped as the case may be)

$\sum fx = N\,\bar{x}$

$\sum fx$ indicate the total sum of all the observations

Measures of Central Tendency

Arithmetic Mean or Mean

Calculate the mean wage from the following data:

Daily wage of worker (Rs):	**550**	**625**	**750**	**825**	**925**	**1000**	**1200**
Number of workers:	**12**	**23**	**37**	**45**	**33**	**24**	**16**

Method 1: $\overline{x} = \dfrac{f_1x_1+f_2x_2+\ \ldots\ldots\ldots+f_nx_n}{f_1+f_2+\ \ldots\ldots\ldots+f_n}$

$$= \frac{12x550+13x625+37x750+45x825+33x925+24x1000+16x1200}{12+23+37+45+33+24+16}$$

$$= \frac{159575}{190} = \mathbf{839.87}$$

x	f	f x
550	12	6600
625	23	14375
750	37	27750
825	45	37125
925	33	30525
1000	24	24000
1200	16	19200
	190	**159575**

Method 2: Arithmetic Mean $\overline{x} = \dfrac{\sum f.x}{N}$

$$= \frac{159575}{190}$$

$$= \mathbf{839.87}$$

Mean daily wage of worker = Rs 839.87

Measures of Central Tendency

Arithmetic Mean or Mean

Case (2): Arithmetic mean from grouped frequency distribution

Note: For calculating Arithmetic mean from grouped frequency distribution, we assume that all the observations in a class are having a value equal to mid value of the class.

Let $x_1, x_2, x_3, \ldots \ldots \ldots . x_n$ be the mid values of the classes and let $f_1, f_2, f_3, \ldots \ldots \ldots . f_n$ be the corresponding frequencies. Then the Arithmetic mean denoted by $\bar{x}$ is defined as, $\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \ldots\ldots\ldots + f_n x_n}{f_1 + f_2 + \ldots\ldots\ldots + f_n}$

$$= \frac{\sum_{i=1}^{i=n} f_i x_i}{\sum_{i=1}^{i=n} f_i}$$

$$= \frac{\sum_{i=1}^{i=n} f_i x_i}{N}$$

$$= \frac{\sum f_i x_i}{N}$$

$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \ldots\ldots\ldots + f_n x_n}{f_1 + f_2 + \ldots\ldots\ldots + f_n} = \frac{\sum f . x}{N}$,

where, N is the total frequency

n stands for the number of observations in raw data

and

N stands for the total frequency (in Ungrouped/ Grouped as the case may be)

Measures of Central Tendency

Arithmetic Mean or Mean

Calculate the A. M from the following data

Class:	**70–80**	**80–90**	**90–100**	**100–110**	**110–120**	**120–130**	**130–140**	**140–150**
Frequency:	**12**	**18**	**35**	**42**	**50**	**45**	**20**	**8**

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

$= \frac{25400}{230} = \mathbf{110.44}$

Class	Frequency	Mid X	f x
70 - 80	12	75	900
80 - 90	18	85	1530
90 -100	35	95	3325
100-110	42	105	4410
110-120	50	115	5750
120-130	45	125	5625
130-140	20	135	2700
140-150	8	145	1160
	230		**25400**

Note: If open ended class are given, then it is usually assumed that open end class has the same width as the adjacent class.
In the above problem, for the first class, if it is given as 'less than 80' instead of 70 – 80, we assume the width to be 10 and the class will be taken as 70 - 80

Measures of Central Tendency

Arithmetic Mean or Mean

Short Cut method or Step deviation method for calculating A.M or Discuss the effect of Change of origin and scale on Arithmetic mean **(Case of Raw Data)**

Answer: Let $x_1, x_2, x_3, \dots \dots \dots . x_n$ be the observations.

Then the A.M is given by $\bar{x} = \frac{\sum \boldsymbol{x}}{\boldsymbol{n}}$

Consider a transformation of the form $d_i = \frac{x_i - A}{c}$, where A and c are any two constants. A is called the assumed mean which can be some observation or suitable value which comes in the middle of minimum and maximum of the observations.

From $d_i = \frac{x_i - A}{c}$, we get $x_i = A + c\, d_i \quad \rightarrow \quad \Sigma x_i = \Sigma A + \Sigma\, c\, d_i$

$\therefore \Sigma x_i = nA + c\, \Sigma d_i \quad \rightarrow \quad \frac{\Sigma x_i}{n} = \frac{nA}{n} + \frac{c\, \Sigma d_i}{n} \quad \rightarrow \boldsymbol{\bar{x} = A + c\, \bar{d}}$

If $d_i = \frac{x_i - A}{c}$, then $\boldsymbol{\bar{d}} = \frac{\boldsymbol{\bar{x} - A}}{\boldsymbol{c}}$.

Note: From x values we calculate $\bar{x} = \frac{\sum \boldsymbol{x}}{\boldsymbol{n}}$

From d values we calculate $\boldsymbol{\bar{d}} = \frac{\sum \boldsymbol{d}}{\boldsymbol{n}}$

If $x_i = A + c\, d_i$, then $\bar{x} = A + c\, \bar{d}$

If $d_i = \frac{x_i - A}{c}$, then $\bar{d} = \frac{\bar{x} - A}{c}$

If $d_i = \frac{1}{k} x_i$, then $\bar{d} = \frac{1}{k} \bar{x}$

Measures of Central Tendency

Arithmetic Mean or Mean

Short Cut method or Step deviation method for calculating A.M or Discuss the effect of Change of origin and scale on A. M **(Case of Frequency distribution)**

Answer: Let $x_1, x_2, x_3, \ldots \ldots \ldots . x_n$ be the observations (in the case of ungrouped frequency distribution) or mid values of classes (in the case of grouped frequency distribution) and let $f_1, f_2, f_3, \ldots \ldots \ldots . f_n$ be the corresponding frequencies.

Then the A.M is given by $\bar{x} = \frac{\sum \boldsymbol{f_i x_i}}{\boldsymbol{N}}$

Consider a transformation of the form $d_i = \frac{x_i - A}{c}$, where A and c are any two constants. A is called the assumed mean which can be some observation or midvalue in the middle of the data set

From $d_i = \frac{x_i - A}{c}$, we get $x_i = A + \text{c } d_i \quad \rightarrow \quad \Sigma f_i x_i = \Sigma A f_i + \Sigma \text{ c } f_i d_i$

$\therefore \Sigma f_i x_i = \text{A } \Sigma f_i + \text{c } \Sigma f_i d_i \rightarrow \quad \frac{\Sigma f_i x_i}{N} = \frac{\text{A } \Sigma f_i}{N} + \frac{\text{c } \Sigma f_i d_i}{N}$

$\therefore \ \boldsymbol{\bar{x}} = \textbf{A} + \textbf{c} \frac{\Sigma \boldsymbol{f_i d_i}}{\boldsymbol{N}}$ **or** $\boldsymbol{\bar{x}} = \textbf{A} + \textbf{c } \boldsymbol{\bar{d}}$**,** where $\bar{d}$ is calculated as $\bar{d} = \frac{\Sigma f_i d_i}{N}$

Measures of Central Tendency

Arithmetic Mean or Mean

Calculate A.M for the following data using direct method and short cut method.

11, 22, 33, 44, 55, 66, 77, 88, 99

Answer:

Direct Method $\bar{x} = \frac{\sum \boldsymbol{x}}{\boldsymbol{n}} = \frac{11+22+33+44+55+66+77+88+99}{9} = \frac{495}{9} = \mathbf{55}$

Short cut Method 1

Let $d_i = \frac{x_i - 55}{11}$

Then $x_i = 55 + 11 d_i$

$\bar{x} = 55 + 11\, \bar{d}$

$\bar{d} = \frac{\sum d}{n} = \frac{0}{9} = 0$

$\therefore \bar{x} = 55 + 11 \times 0 = 55$

x	$d_i = \frac{x_i - 55}{11}$
11	-4
22	-3
33	-2
44	-1
55	0
66	1
77	2
88	3
99	4
	0

Short cut Method 2

Let $d_i = \frac{1}{11} x_i$

Then $\bar{d} = \frac{1}{11} \bar{x}$

$\therefore \frac{1}{11} \bar{x} = \bar{d} \rightarrow \bar{x} = 11\, \bar{d}$

$\bar{d} = \frac{\sum d}{n} = \frac{45}{9} = 5$

$\bar{x} = 11 \times 5 = 55$

x	$d_i = \frac{x_i}{11}$
11	1
22	2
33	3
44	4
55	5
66	6
77	7
88	8
99	9
	45

Measures of Central Tendency

Arithmetic Mean or Mean

Calculate A.M for the following data using direct method and short cut method.

Mark of students:	15	21	23	27	35	41	43	48
Number of students:	11	19	20	25	29	23	15	13

Direct Method

$$\overline{x} = \frac{\sum f.x}{N}$$

$$= \frac{4913}{155}$$

$$= 31.697$$

Short cut Method

Let $d_i = \frac{x_i - 31}{2}$

$$\overline{x} = A + c\,\frac{\sum f_i d_i}{N}$$

$$= 31 + 2 \text{ x } \frac{54}{155}$$

$$= 31 + 0.6967$$

$$= 31.697$$

Mark (x)	No. of Students (f)	f.x
15	11	165
21	19	399
23	20	460
27	25	675
35	29	1015
41	23	943
43	15	645
47	13	611
	N = 155	Σ fx=4913

Table for direct method

Mark (x)	No. of Students (f)	$d_i = \frac{x_i - 31}{2}$	$f_i d_i$
15	11	-8	-88
21	19	-5	-95
23	20	-4	-80
27	25	-2	-50
35	29	2	58
41	23	5	115
43	15	6	90
47	13	8	104
	155		$\Sigma f_i d_i = 54$

Table for short cut method

Measures of Central Tendency

Arithmetic Mean or Mean

Calculate A.M for the following data using direct method and short cut method.

Mark of students:	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79
Number of students:	11	19	20	25	29	23	15	13

Direct Method

$$\overline{x} = \frac{\sum f.x}{N}$$

$$= \frac{6157.5}{155}$$

$$= 39.726$$

Short cut Method

Let $d_i = \frac{x_i - 34.5}{10}$

$$\text{A.M } (\overline{x}) = \text{A} + \text{c} \frac{\sum f_i d_i}{N}$$

$$= 34.5 + 10 \text{ x } \frac{81}{155}$$

$$= 34.5 + 5.2258$$

$$= 39.726$$

Mark	No. of Students (f)	Mid Value (X)	f.x	$d_i = \frac{x_i - 34.5}{10}$	$\Sigma f_i d_i$
0 - 9	11	4.5	49.5	-3	-33
10 - 19	19	14.5	275.5	-2	-38
20 - 29	20	24.5	490	-1	-20
30 -39	25	34.5	862.5	0	0
40 - 49	29	44.5	1290.5	1	29
50 - 59	23	54.5	1253.5	2	46
60 - 69	15	64.5	967.5	3	45
70 - 79	13	74.5	968.5	4	52
	N = 155		Σfx = 6157.5		Σfd = 81

Measures of Central Tendency

Arithmetic Mean or Mean

The A.M for the following data is known to be 67.45 inches. Find the missing frequency f.

Height in inches:	60 – 62	63 – 65	66 – 68	69 -71	72 – 74
Number of students:	15	54	f	81	24

$$\overline{x} = \frac{\sum f.x}{N}$$

$$67.45 = \frac{11793+67f}{174+f}$$

$$67.45\,(174 + f) = 11793 + 67\,f$$

$$67.45\,f - 67\,f = 11793 - 11736.3$$

$$0.45\,f = 56.7$$

$$f = \frac{56.7}{0.45} = 126$$

Missing frequency (f) = 126

Mark	No. of Students (f)	Mid Value (X)	f.x
60 - 62	15	61	915
63 - 65	54	64	3456
66 - 68	f	67	67 f
69 - 71	81	70	5670
72 - 74	24	73	1752
	N = 174 + f		Σfx = 11793 +67f

Measures of Central Tendency

Arithmetic Mean or Mean

State and Prove the properties of Arithmetic Mean

Property 1: The algebraic sum of deviations of observations from A.M is Zero

Proof: Let $x_1, x_2, x_3, \ldots\ldots\ldots x_n$ be n observations.

Then the Arithmetic mean denoted by $\bar{x}$ is given by $\bar{x} = \frac{\sum \boldsymbol{x}}{\boldsymbol{n}}$.

*D*eviation of observations from the A.M are , $(x_1 - \bar{x}), (x_2 - \bar{x}), \ldots\ldots\ldots (x_n - \bar{x})$

Algebraic sum of deviations $= (x_1 - \bar{x}) + (x_2 - \bar{x}) + \ldots + (x_n - \bar{x})$

$= (x_1 + x_2, + x_3, + \cdots \ldots\ldots + x_n) - n\bar{x}$

$= \sum \boldsymbol{x} - n\bar{x}$

$= n\bar{x} - n\bar{x}$

$= \mathbf{0}$

Note: For frequency distribution with respective frequencies $f_1, f_2, f_3, \ldots\ldots\ldots f_n$,

Algebraic sum of deviations $= f_1(x_1 - \bar{x}) + f_2(x_2 - \bar{x}) + \ldots + f_n(x_n - \bar{x})$

$= (f_1 x_1 + f_2 x_2 + \cdots \ldots\ldots + f_n x_n) - \bar{x}\,(f_1 + f_2 + \ldots . + f_n)$

$= \sum \boldsymbol{f}\boldsymbol{x} - \bar{x}\,N = N\,\bar{x} - N\,\bar{x} = 0$

Measures of Central Tendency

Arithmetic Mean or Mean

State and Prove the properties of Arithmetic Mean

Property 2: Sum of squares of deviations of observations from A.M is minimum

Proof: Let $x_1, x_2, x_3, \ldots\ldots\ldots. x_n$ be n observations.

Let A be a constant from which deviations of observations are taken.

Deviations of observations from A are are , $(x_1 - A), (x_2 - A), \ldots\ldots\ldots. (x_n - A)$

$Sum\ of\ squares$ of deviations (S) of observations from y A is given by,

$$S = (x_1 - A)^2 + (x_2 - A)^2 + \ldots\ldots.+ (x_n - A)^2$$

$$= \sum (x_i - A)^2$$

$S = \sum (x_i - A)^2$ is a minimum when $\frac{dS}{dA} = 0$ and $\frac{d^2S}{dA^2} > 0$

$$\frac{dS}{dA} = 0 \rightarrow 2\sum (x_i - A)^{2-1}(0 - 1) = 0$$

$$ie,\ \sum (x_i - A) = 0 \rightarrow \Sigma x_i = \Sigma A \rightarrow n\bar{x} = nA \rightarrow A = \bar{x}$$

$$\frac{d^2S}{dA^2} = \frac{d}{dA}\left(\frac{dS}{dA}\right) = \frac{d}{dA}\left(-2\sum (x_i - A)^1\right) = -2\,\Sigma\,(0\text{-}1) = 2n > 0$$

$\therefore Sum\ of\ squares\ of\ deviations\ of\ observations\ is\ a\ minimum\ when\ A = \bar{x}$

Measures of Central Tendency

Arithmetic Mean or Mean

Weighted Arithmetic Mean

Sometimes we wish to find the mean or average of numbers or observations having varying importance in the data set. For this, we assign more importance, or weight, to some of the numbers. Engineering entrance exam with different weightage for different subjects is an example of such a situation.

Suppose the grade of students in an examination is based on a midterm and a final exam, each of which is based on 100 possible points. However, the final exam will worth 60% of the grade and the midterm only 40%. How could you determine an average score that would reflect *these different weights? The average you need is the weighted average, given by,* ***Weighted average*** $= \frac{\sum w_1 x_i}{\sum w_i}$, where x_i is a data value and w_i is the weight assigned to that data value.

Measures of Central Tendency

Arithmetic Mean or Mean

Calculate the simple and Weighted Average price per ton of an item during January to June. Account for difference if any between the two.

Month:	Jan	Feb	Mar	Apr	May	June
Price per ton (Rs):	42.50	51.25	50.00	52.00	44.25	54.00
Qty Purchased (Tons):	25	30	40	50	10	45

Note: In order to find the weighted average price per ton, the Quantity purchased should be taken as weight

Simple Arithmetic average price $= \frac{\Sigma x}{n}$

$$= \frac{294}{6} = \mathbf{49.00}$$

Weighted Arithmetic Average Price $= \frac{\Sigma wx}{\Sigma w}$

$$= \frac{10072.50}{200} = \boldsymbol{50.36}$$

Simple average and weighted average are equal only when all weights are equal. The difference is because the weights are not equal.

Month	Price (x)	Qty (w)	w.x
Jan	42.50	25	1062.50
Feb	51.25	30	1537.50
Mar	50.00	40	2000.00
Apr	52.00	50	2600.00
May	44.25	10	442.50
Jun	54.00	45	2430.00
	Σ x = 294.00	**Σ w = 200**	**Σ wx = 10072.50**

Measures of Central Tendency

Arithmetic Mean or Mean

Calculate the missing frequencies in the following distribution if $\bar{x}$ = 11.09,N =60

Class:	9.3-9.7	9.8-10.2	10.3-10.7	10.8-11.2	11.3-11.7	11.8-12.2	12.3-12.7	12.8-13.2	Total
Frequency:	2	5	?	?	14	6	3	1	60

We have, $\bar{x} = A + c\,\bar{d}$

Given that $\bar{x} = 11$, N = 60

$$\bar{d} = \frac{\Sigma f.d}{N} = \frac{23 - f_1}{60}$$

ie, $11.09 = 11 + 0.5\left(\frac{23 - f_1}{60}\right) = 11 + \left(\frac{23 - f_1}{120}\right)$

$11.09 - 11 = \left(\frac{23 - f_1}{120}\right)$

$0.09 \times 120 = 23 - f_1$

$10.8 = 23 - f_1 \rightarrow \quad f_1 = 23 - 10.8 = 12.2$

Since the frequency to be an integer **$f_1 = 12$**

Also $2+5+f_1+f_2+14+6+3+1 = 60$

$31+f_1+f_2 = 60 \rightarrow f_2 = 60-31-f_1 = 60-31-12 = 17$

$f_1 = 12$ and $f_2 = 17$

Class Limits	Frequency f	Mid X	$d=\frac{x-11}{0.5}$	f.d
9.3-9.7	2	9.5	- 3	- 6
9.8-10.2	5	10.0	- 2	- 10
10.3-10.7	f_1	10.5	- 1	$-f_1$
10.8-11.2	f_2	11.0	0	0
11.3-11.7	14	11.5	1	14
11.8-12.2	6	12.0	2	12
12.3-12.7	3	12.5	3	9
12.8-13.2	1	13.0	4	4
Total	60			$23-f_1$

Measures of Central Tendency

Arithmetic Mean or Mean

Combined Arithmetic Mean

If k groups contain $n_1, n_2, \ldots\ldots\, n_k$ observations with means $\bar{x}_1, \bar{x}_2, \ldots\ldots\ldots\ldots\bar{x}_k$, then the mean of the combined group of $n_1 + n_2 + \ldots\ldots + n_k$ observations is given by, $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots\ldots + n_k\bar{x}_k}{n_1 + n_2 + \ldots\ldots + n_k}$.

The mean of n_1 observations is given by $\bar{x}_1 = \dfrac{Sum\ of\ n_1\ observations}{n_1}$.

There fore Sum of n_1 observations = $n_1\ \bar{x}_1$

Similarly, Sum of n_2 observations = $n_2\ \bar{x}_2$

……………………………………………

Sum of n_k observations = $n_k\ \bar{x}_k$

Sum of $n_1 + n_2 + \ldots\ldots + n_k$ observations = $n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots\ldots + n_k\ \bar{x}_k$

AM of $n_1 + n_2 + \ldots\ldots + n_k$ observations ($\bar{x}$) = $\dfrac{\text{Sum of } n_1 + n_2 + \ldots\ldots + n_k \text{ observations}}{n_1 + n_2 + \ldots\ldots + n_k}$

$$= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots\ldots + n_k\bar{x}_k}{n_1 + n_2 + \ldots\ldots + n_k}.$$

For two groups we have

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

Measures of Central Tendency

Arithmetic Mean or Mean

Question: There are two branches of an establishment employing 100 and 80 persons respectively. If the arithmetic means of the monthly salaries paid by the two branches are 27500 and 22500 respectively find the arithmetic mean of monthly salary of employees of the establishment as a whole.

Answer: For two groups, the combined mean is given by $\boldsymbol{\bar{x}} = \dfrac{\boldsymbol{n_1\bar{x}_1 + n_2\bar{x}_2}}{\boldsymbol{n_1 + n_2}}$

We have $n_1 = 100$, $n_2 = 80$, $\bar{x}_1 = 27500$, $\bar{x}_2 = 22500$

The mean monthly salary of employees in the establishment $(\overline{\boldsymbol{x}}) = \dfrac{\boldsymbol{n_1\bar{x}_1 + n_2\bar{x}_2}}{\boldsymbol{n_1 + n_2}}$

$$\frac{100 \text{ x } 27500 + 80 \text{ x } 22500}{100 + 80} = \frac{\mathbf{45500}}{\mathbf{180}} = \mathbf{25278}$$

Question: The arithmetic means of the monthly salaries paid all employees in factory is Rupees 50000. The mean monthly salary paid to male and female employees are Rupees 52000 and 42000. Obtain the percentage of male to female employees in the company.

We have $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \rightarrow (n_1 + n_2)\ \bar{x} = n_1\bar{x}_1 + n_2\bar{x}_2$

$\therefore (n_1 + n_2)\ 50000 = n_1\ 52000 + n_2 42000$

$50000\ n_1 + 50000\ n_2 = 52000\ n_1 + 42000\ n_2$

i.e, $8000\ n_2 = 2000\ n_1$ $\rightarrow \dfrac{n_1}{n_2} = \dfrac{8000}{2000}$ $\rightarrow \boldsymbol{n_1 : n_2} = \mathbf{4 : 1}$

Measures of Central Tendency

Arithmetic Mean or Mean

Merits and demerits of Arithmetic mean :

Merits:

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.
3. If the number of items is sufficiently large, it is more accurate and more reliable.
4. It is a calculated value and is not based on its position in the series.
5. It is amenable for further mathematical treatment
6. Of all averages, it is affected least by fluctuations of sampling.
7. It provides a good basis for comparison.

Demerits:

1. It cannot be obtained by inspection nor located through a frequency graph.
2. It cannot be used in the study of qualitative phenomena not capable of numerical measurement i.e. Intelligence, beauty, honesty etc.,
3. It can ignore any single item only at the risk of losing its accuracy.
4. It is affected very much by extreme values.
5. It cannot be calculated for open-end classes.
6. It may lead to fallacious conclusions, if the details of the data from which it is computed are not given.

Measures of Central Tendency

Median

Positional Averages: Positional averages are based on (1) the position of the given observation in a series, arranged in an ascending or descending order (2) the number of positions occupied by various observations. Averages of position can be that observation which is positioned at the center in an arranged data or that observation which occupies the maximum number of positions.

Median: Median is defined as the middle most observation when they are arranged in the ascending or descending order of magnitude.

The median is that value of the variable which divides the group into two equal parts, one part comprising all values greater than median, and the other, all values less than median.

The number of observations smaller than median is same as the number of observations greater than it. Median gives the value of the most central observation and hence it is treated as real measure of central tendency, especially in psychological and achievement tests. For example, to find the student of average intelligence we use the measure Median.

Measures of Central Tendency

Median

Calculation of Median from Raw data : Arrange the given observations in the increasing or decreasing order of magnitude.

If the number of observations (n) is odd, median is the middle most one.

Note:

Median is not $\frac{n+1}{2}$

Median = $\left(\frac{n+1}{2}\right)^{th}$ observation in the arranged set of observations

Example: Find median for the following data 25, 18, 27, 10, 8, 30, 42, 20, 53

Solution: Arranging the data in the increasing order, we get, 8, 10, 18, 20, 25, 27, 30, 42, 53

Number of observations (n) = 9, which is an odd number.

Hence Median is the middle most observation

Median = $\left(\frac{n+1}{2}\right)^{th}$ **observation in the arranged set**

= $\left(\frac{9+1}{2}\right)^{th}$ **observation in the arranged set**

= $(5)^{th}$ **observation in the arranged set = $\underline{25}$**

Measures of Central Tendency

Median

Calculation of Median from Raw data : Arrange the given observations in the increasing or decreasing order of magnitude.

If the number of observations (n) is odd, median is the middle most one.

Median = $\left(\frac{n+1}{2}\right)^{th}$ observation in the arranged set of observations

If the number of observations is even, median is the mean of middle two observations.

$$\text{Median} = \frac{\left(\frac{n}{2}\right)^{th}\text{observationin the arranged set} + \left(\frac{n}{2}+1\right)^{th}\text{observation in the arranged set}}{2}$$

Example: Find median for the following data 3, 9, 12, 40, 28, 10, 2, 32

Solution: Arranging the data in the increasing order 2, 3, 9, 10, 12, 28, 32, 40

Here the number of observations (n) =8 which is an even number.

Hence median is the mean of the middle two items namely $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observations.

$$\text{Median} = \frac{\left(\frac{n}{2}\right)^{th}\text{observationin the arranged set} + \left(\frac{n}{2}+1\right)^{th}\text{observation in the arranged set}}{2}$$

$$= \frac{4^{th}\text{observationin the arranged set} + 5^{th}\text{observation in the arranged set}}{2} = \frac{10+12}{2} = \underline{\mathbf{11}}$$

Measures of Central Tendency

Median

The following table represents the marks obtained by a batch of 10 students in certain class test two subjects in Statistics and Mathematics. Indicate in which subject is the level of knowledge higher ?

Serial Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Marks (Statistics)	53	55	52	32	30	60	47	46	35	**28**
Marks (Mathematics)	**57**	**45**	**24**	**31**	**25**	**84**	**43**	**80**	**32**	**72**

Solution: For these types of questions, median is the most suitable measure of central tendency.

The mark in the two subjects are first arranged in increasing order as follows:

Serial Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Marks in Statistics	28	30	32	35	46	47	52	53	55	**60**
Marks in Mathematics	**24**	**25**	**31**	**32**	**43**	**45**	**57**	**72**	**80**	**84**

$$\text{Median} = \frac{\mathbf{5}^{th}\text{observationin the arranged set} + \mathbf{6}^{th}\text{observation in the arranged set}}{2}$$

$$\text{Median Mark for Statistics} = \frac{\mathbf{5}^{th}\text{ item} + \mathbf{6}^{th}\text{ item}}{2} = \frac{\mathbf{46+47}}{2} = \mathbf{46.5}$$

$$\text{Median Mark for Mathematics} = \frac{\mathbf{5}^{th}\text{ item} + \mathbf{6}^{th}\text{ item}}{2} = \frac{\mathbf{43+45}}{2} = \mathbf{44}$$

The level of knowledge in Statistics is more than that of Mathematics.

Measures of Central Tendency

Median

Grouped Data:

In a grouped distribution, we present the frequencies corresponding to observati0ns or frequency of the various classes. Grouping can be in the form of a discrete frequency distribution or a continuous frequency distribution. Whatever may be the type of distribution, for calculating the median, cumulative frequencies have to be calculated.

Cumulative frequency : (c f)

Cumulative frequency of each class is the sum of the frequency of the class and the frequencies of the pervious classes, i.e., adding the frequencies successively, so that the last cumulative frequency gives the total number of items.

Discrete Series:

Step1: Find cumulative frequencies.

Step2: Find $\left(\frac{N+1}{2}\right)$

Step3: See in the cumulative frequencies the value just greater than $\left(\frac{N+1}{2}\right)$

Step4: Then the corresponding value of x is median.

Measures of Central Tendency

Median

Calculation of Median from ungrouped frequency distribution: Arrange the given observations in the increasing order of magnitude along with the corresponding frequencies. The cumulative frequency (Less than type) is calculated for each of the observations. If N is the total frequency, the observation corresponding to the cumulative frequency $\left(\frac{N+1}{2}\right)$ gives the median.

Median = $\left(\frac{N+1}{2}\right)^{th}$ observation among the arranged set

= Observation corresponding to the Cum. frequency $\left(\frac{N+1}{2}\right)$

Example: Calculate median from the following data

X: 12 17 23 34 40 56 61 65 70

f: 8 12 17 25 33 28 17 11 9

Calculate Median		
X	f	c.f
12	8	8
17	12	20
23	17	37
34	25	62
40	**33**	**95**
56	28	123
61	17	140
65	11	151
70	9	160
	160	

From table, we get N = 160 and $\left(\frac{N+1}{2}\right)$ = 80.5, which implies that Median will be just midway between 80^{th} and 81^{th} observations. From table we can see that the value of observation from 63^{rd} to 95^{th} is 40. Hence Median = 40

Measures of Central Tendency

Median

Calculation of Median from grouped frequency distribution: When the observations are grouped into different classes, the median can be calculated using the formula, Median $(M_e) = l + \frac{\left(\frac{N}{2} - m\right) c}{f}$, *where,*

l is the true lower limit of median class

N is the total frequency

m is the cumulative of the class preceding median class

c is width of the median class

f is the frequency of the median class

Example: Calculate Median for the following data

Class:	**0-10**	**10-20**	**20-30**	**30-40**	**40-50**	**5-60**	**60-70**	**70-80**	**80-90**
f:	**8**	**12**	**17**	**25**	**33**	**28**	**17**	**11**	**9**

N = 160, N/2 = 80, l = 40, m = 62, c = 10, f = 33

$$\text{Median } (M_e) = l + \frac{\left(\frac{N}{2} - m\right) c}{f} = 40 + \frac{(80 - 62)\,10}{33} = \mathbf{45.45}$$

Calculate Median		
Class	f	c.f
0 - 10	8	8
10 - 20	12	20
20 – 30	17	37
30 – 40	25	62
40 – 50	**33**	**95**
50 – 60	28	123
60 – 70	17	140
70 – 80	11	151
80- 90	9	160
	160	

Measures of Central Tendency

Median

Example: Calculate the most suitable average for the following data giving reasons for your choice .

Wage (Rs)/hour: $\leq$ 100 100-200 200-300 300-400 $\geq$ 400

No of workers 28 67 75 54 36

Answer: Median is the appropriate average to be calculated from a frequency distribution with open ended class.

N = 260, N/2 = 130, l = 200, m = 94, c = 100, f = 75

$$\text{Median } (M_e) = l + \frac{\left(\frac{N}{2} - m\right) c}{f}$$

$$= 200 + \frac{(130 - 94)\ 100}{75}$$

$$= 200 + \frac{36\ x\ \ 100}{75}$$

$$= 200 + 48 = \underline{\underline{\mathbf{248}}}$$

Calculate Median		
Class	f	c.f
$\leq$ 100	28	28
100 - 200	66	94
200 – 300	**75**	**169**
300 – 400	55	224
$\geq$ 400	36	260
	260	

Measures of Central Tendency

Median

Determination of Median from ogives : When the observations are grouped into different classes, the median can be determined graphically from ogive/ogives. The median of a grouped frequency distribution can be determined from the ogive/ogives by the following step by step procedure:

Median from Less than ogive: First draw a less ogive by taking class boundaries on the horizontal axis (X-axis) and less than cumulative frequencies on the vertical axis (Y-axis). Less than cumulative frequencies are plotted against the class boundaries.

From the point N/2 on the vertical axis, draw a horizontal line to meet the ogive. Then draw a perpendicular to horizontal axis from the point of intersection. The point at which this perpendicular meets the horizontal axis will give the median.

Median from Greater than ogive: Follow the above procedure except greater than cumulative frequencies are plotted against the class boundaries.

Median from Greater than ogive: Draw the two ogives and draw a perpendicular from the point of intersection of the two ogives to meet the horizontal axis. The foot of the perpendicular meets the horizontal axis at Median.

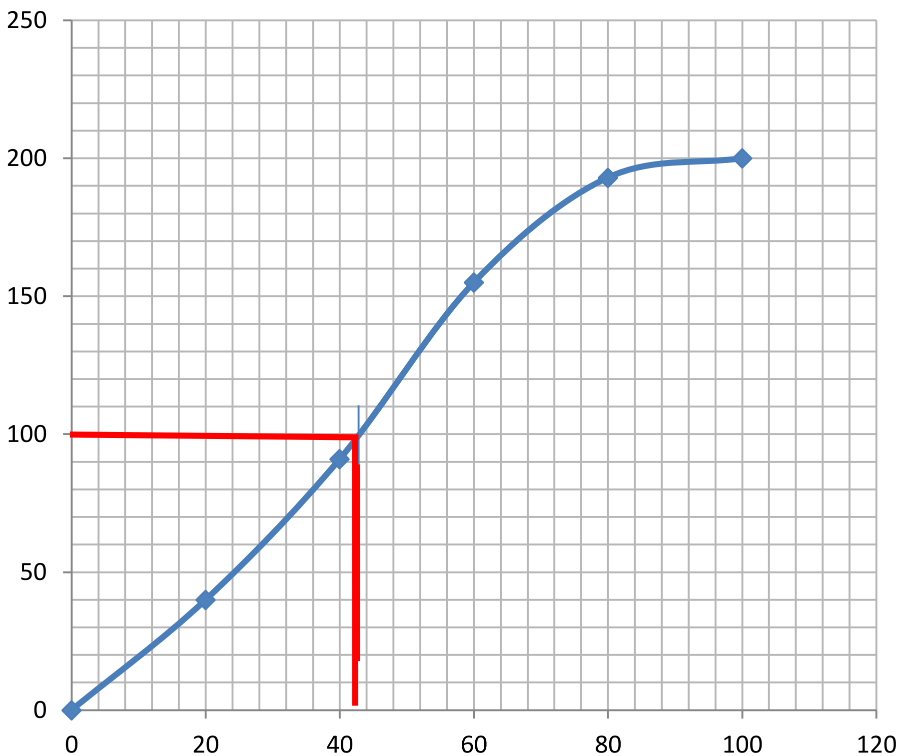
Measures of Central Tendency

Median

Find the median graphically using less than ogive.

Weekly Wages in Rs	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
Number of workers	40	51	64	38	7

Answer: The table on the right gives the less than cumulative frequencies



Median = 43

Wages	Frequency	Upper Class boundaries	Less than cumulative frequency
0 -20	40	**0**	**0**
20-40	51	**20**	**40**
40-60	64	**40**	**91**
60-80	38	**60**	**155**
80-100	7	**80**	**193**
		100	**200**

Less than cumulative frequency is the number of observations below each upper bound

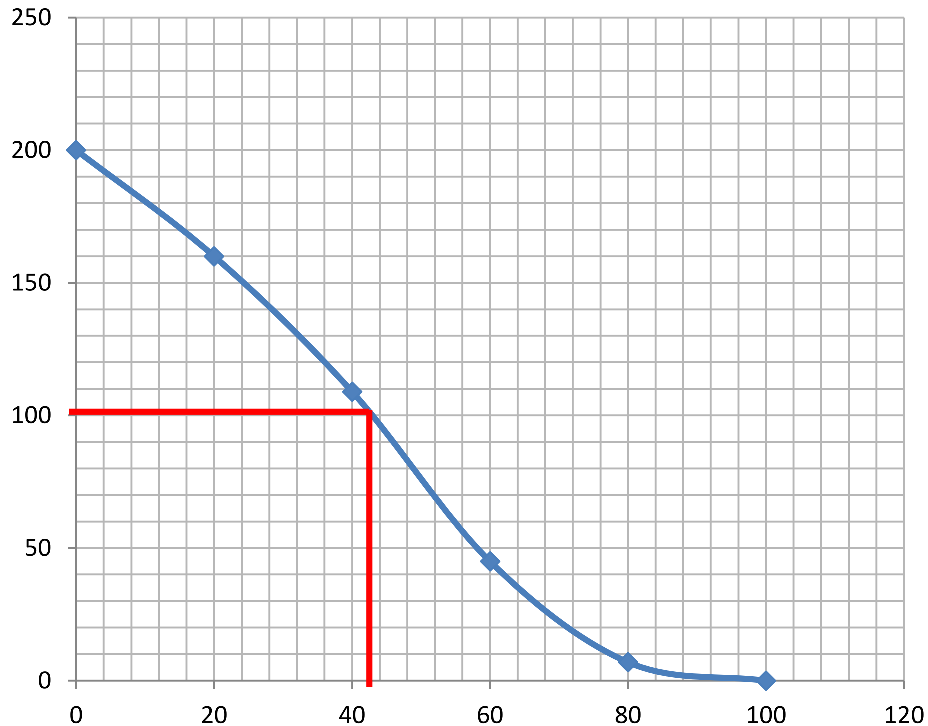
Measures of Central Tendency

Median

Find the median graphically using greater than ogive.

Weekly Wages in Rs	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
Number of workers	40	51	64	38	7

Answer: The table on the right gives the less than cumulative frequencies



Wages	Frequency	Lower Class boundaries	Greater than cumulative frequency
0 -20	40	**0**	**200**
20-40	51	**20**	**160**
40-60	64	**40**	**109**
60-80	38	**60**	**45**
80-100	7	**80**	**7**
		100	**0**

Greater than cumulative frequency is the no. of observations above each lower bound

Median = 43

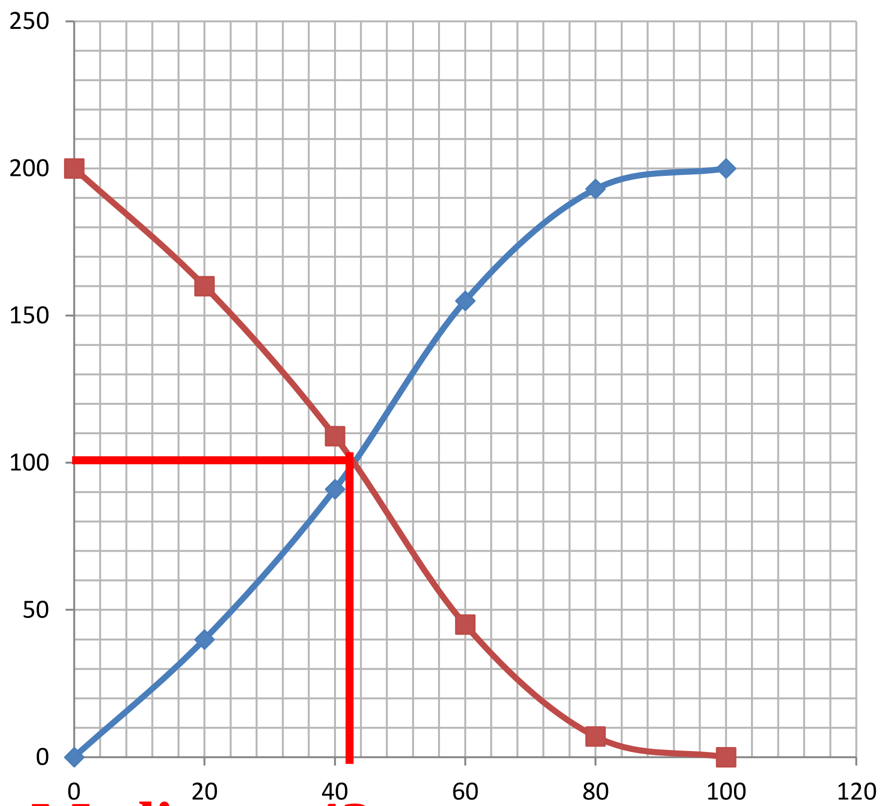
Measures of Central Tendency

Median

Find the median graphically using greater than ogive.

Weekly Wages in Rs	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
Number of workers	40	51	64	38	7

Answer: The table on the right gives the less than cumulative frequencies



Median = 43

Wages	Frequency	Lower Class boundaries	Less than cumulative frequency	Greater than cumulative frequency
0 -20	40	**0**	**0**	**200**
20-40	51	**20**	**40**	**160**
40-60	64	**40**	**91**	**109**
60-80	38	**60**	**155**	**45**
80-100	7	**80**	**193**	**7**
		100	**200**	**0**

Measures of Central Tendency

Median

Find the mean and median. Also locate median graphically using ogives.

Central value of class	60	65	70	75	80	85	90	95
No. of observations	4	5	31	39	114	30	25	2

Answer: The data gives mid values of classes and frequencies. Since there is a common difference of 5 between the mid values, we can assume that the classes have a width of 5.

Also the class boundaries will lie exactly mid way between the two successive mid values as Shown in the table.

(Median class is given in red color)

Mid values	Frequency	Class boundaries	d = $\frac{x-75}{5}$	f.d	Less than c.f	Greater than c.f
60	4	57.5 – 62.5	-3	- 12	4	250
65	5	62.5 – 67.5	-2	- 10	9	246
70	31	67.5 – 72.5	-1	- 31	40	241
75	39	72.5 – 77.5	0	0	79	210
80	114	77.5 – 82.5	1	114	193	171
85	30	82.5 – 87.5	2	60	223	57
90	25	87.5 – 87.5	3	75	248	27
95	2	87.5 – 92.5	4	8	250	2
	250			204		

Measures of Central Tendency

Median

To find the mean

We have $\sum fd = 204, N = 250, where\ d = \frac{x - A}{c} = \frac{x - 75}{5}$

Arithmetic mean, $\bar{x}$ = A + c $\frac{\sum f\ d}{N}$

$= 75 + 5 \text{ x } \frac{204}{250} = \mathbf{\underline{79.08}}$

To find the median

N = 250, N/2 = 125, l = 77.5, m = 79, c = 5, f = 114

Median (M_e) $= l + \frac{\left(\frac{N}{2} - m\right) c}{f}$

$= 77.5 + \frac{(125 - 79)5}{114}$

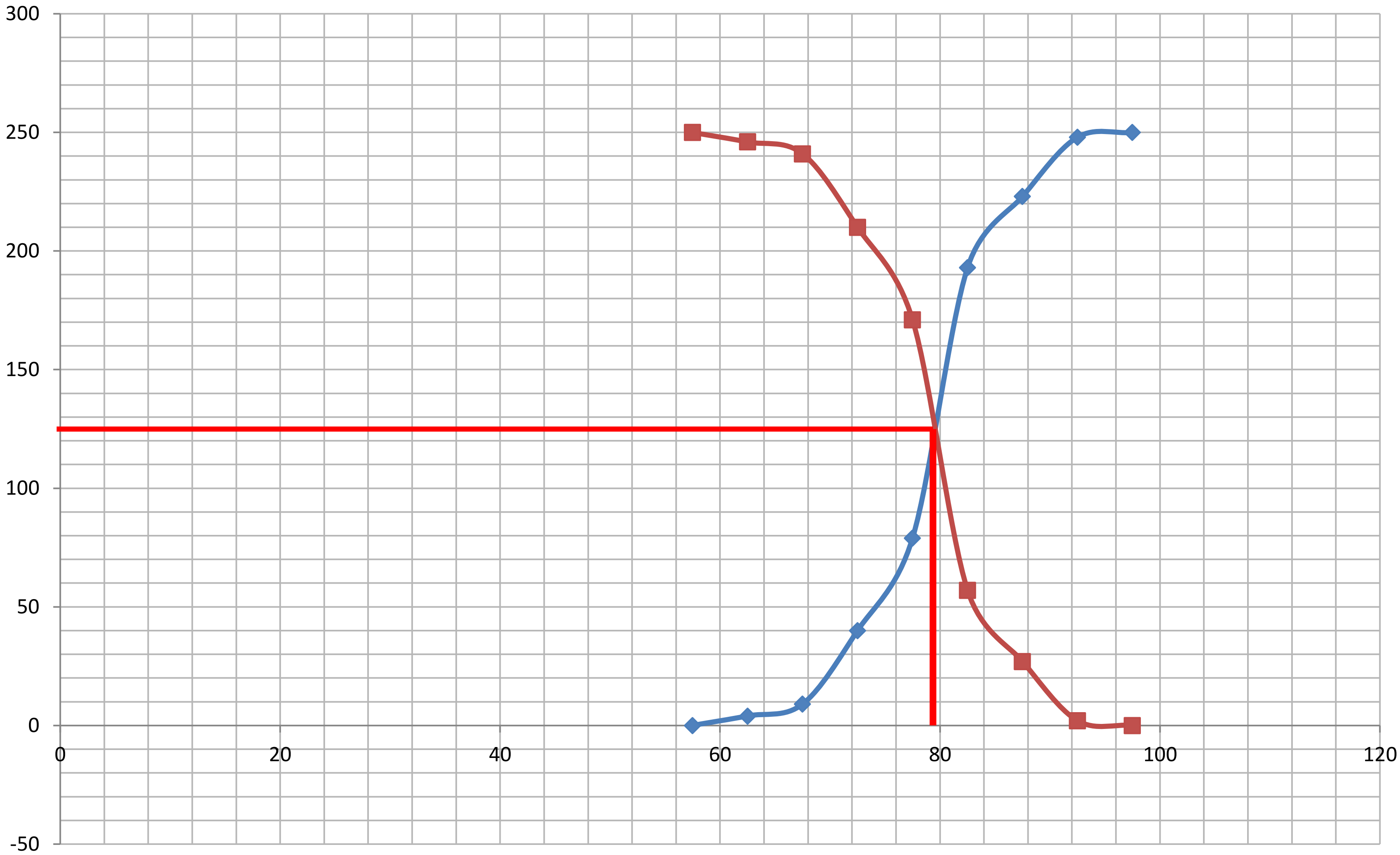
$= 77.5 + \frac{46\ x\ \ 5}{114}$

= **79.52**

Measures of Central Tendency

Median

To find the median using ogives



Measures of Central Tendency

Median

Example: It is known that median of the following data set having a total frequency of 1000 is 413.11. Calculate the missing frequencies.

Values	300-325	325-350	350-375	375-400	400-425	425-450	450-475	475-500
Frequency	5	17	80	?	326	?	88	9

Let **a** and **b** denote the frequencies of the classes 375-400 and 425-450.

We are given that total frequency = 1000. i.e., 5+17+80+**a**+326+**b**+88+9 = 1000

i.e., 525 + **a** + **b** = **1000** $\rightarrow$ **a** + **b**= 1000 – 525 = 475 $\rightarrow$ **b**= 475 - **a** ………(1)

Since the median is given to be 413.11, the median class is 400 – 425

Median is calculated using the formula, Median (M_e) $= l + \dfrac{\left(\frac{N}{2} - m\right)c}{f}$

l = 400, $\frac{N}{2}$ = 500, m = 5+17+80+**a** = 102+**a**, c = 25, f = 326

$\therefore\ 400 + \dfrac{(500 - \overline{102+a})\,25}{326} = 413.11 \rightarrow \dfrac{(398 - a)25}{326} = 13.11 \rightarrow 398 - a = \dfrac{13.11 \times 326}{25}$

i.e., 398 – a = 171 $\rightarrow$ 398 – 171 = a $\rightarrow$ a = 227 ………………**(2)**

Using (2) in (1) we get **b**= 475 - 227 = **248** $\therefore$ a = 227 and b = 248

Measures of Central Tendency

Mode

Mode of a given set of observations is the most frequently occurring observation. Or, mode is that value which occurs with the maximum frequency. Mode is the most typical or prevalent value, and at times represent the true characteristic of the distribution as a measure of central tendency. Mode is the value of the variable which is predominant in the series.

The modal size of shoe for students of a particular age is that size of the shoe which the largest number of students of that age use, and as such this size may be considered as the representative size of that age. Mode is commonly used in business because it is most likely to occur. Meteorological studies and forecasting also depend upon mode as a suitable average.

From a raw data, mode can be determined as the value which occurs the maximum number of times. In an ungrouped frequency distribution, mode is that value having the maximum frequency if it is not repeated and if it is not at the beginning or at the end of the distribution.

Measures of Central Tendency

Mode

Calculation of Mode for continuous frequency distribution.

In case of continuous frequency distribution, Mode is given by the formula:

$$\text{Mode } (M_o) = l + \frac{c(f - f_1)}{(2f - f_1 - f_2)}$$

Where, l is true lower limit of modal class

c is width of modal class

f is the frequency of modal class

f_1 is frequency of the class just preceding modal class

f_2 is the frequency of the class just after modal class

Note: When the frequency distribution has classes of unequal width, the above formula is not suitable and in such cases we use the empirical relation to find mode.

Mean – Mode = 3(Mean – Median) or **Mode = 3 Median – 2 Mean**

The above empirical relationship is suitable to find mode from unimodal distributions of moderate skewness

Measures of Central Tendency

Mode

Calculate Modal wage from the following data.

Wage (Rs)	500-599.99	600- 699.99	700-799.99	800-899.99	900-999.99	1000-1099.99	1100-1199.99
No of workers	18	30	36	54	40	25	12

$$\text{Mode } (M_o) = l + \frac{c(f - f_1)}{(2f - f_1 - f_2)}$$

Where, l is true lower limit of modal class, $\boldsymbol{l}$ **= 799.995**

c is width of modal class, **c = 100**

f is the frequency of modal class, **f = 54**

f_1 is frequency of the class just preceding modal class, $\boldsymbol{f_1}$ **= 36**

f_2 is the frequency of the class just after modal class, $\boldsymbol{f_2}$ **= 40**

$$\text{Mode } (M_o) = l + \frac{c(f - f_1)}{(2f - f_1 - f_2)} = 799.995 + \frac{100(54 - 36)}{(2\ x\ 54 - 36 - 40)}$$

$$= 799.995 + \frac{100\ x\ 18}{(108 - 76)} = 799.995 + \frac{1800}{32}$$

$$= 799.995 + 56.25 = \mathbf{856.245}$$

Measures of Central Tendency

Mode

Calculate mean, Median and Mode from the following data.

Value	Less than 10	Less than 20	Less than 30	Less than 40	Less than 50	Less than 60	Less than 70	Less than 80
Frequency	14	35	66	108	140	165	183	190

Ware given the less than cumulative frequency distribution. We have to find out the frequency of each class as given in the table for calculating Mean and Median

Note: The table gives the details for calculation of A.M by Direct and Short cut methods

Value	Frequency	Class	Frequency	Mid X	f.x	$d = \frac{x-35}{10}$	f.d	c.f
< 10	14	00 –10	14	5	70	- 3	-42	14
< 20	35	10 –20	21	15	315	- 2	-42	35
< 30	66	20 – 30	31	25	775	- 1	-31	66
< 40	108	30 – 40	42	35	1470	0	0	108
< 50	140	40 – 50	32	45	1440	1	32	140
< 60	165	50 – 60	25	55	1375	2	50	165
< 70	183	60 – 70	18	65	1170	3	54	183
< 80	190	70 - 80	7	75	525	4	28	190
			190		7140		49	

Measures of Central Tendency

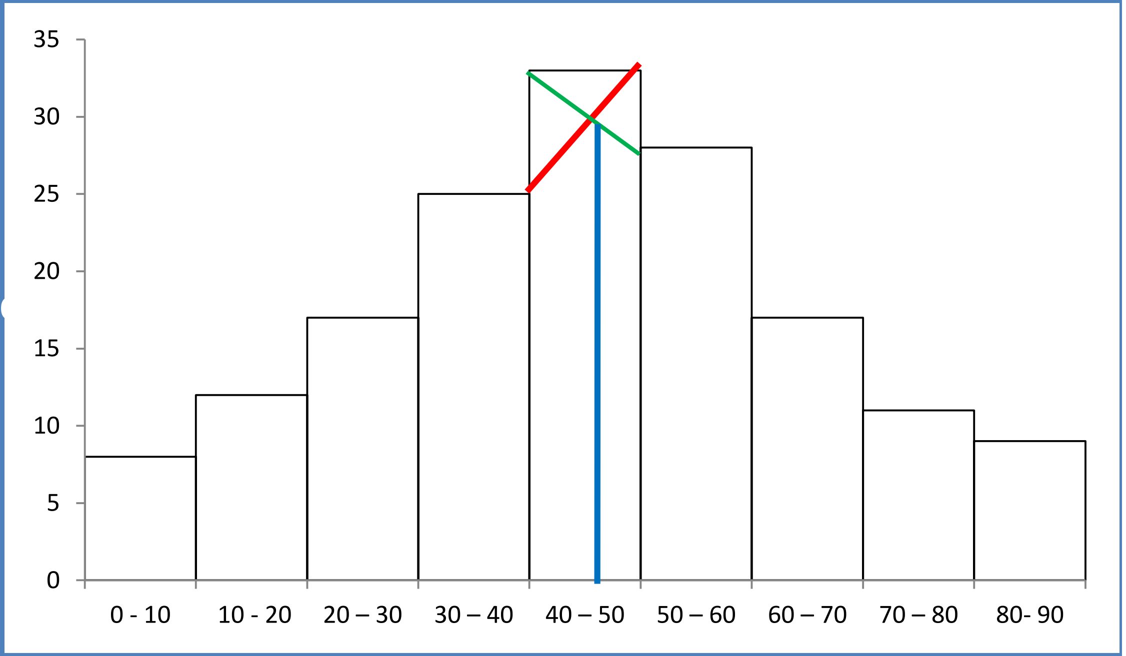
Mode

Step 1: Draw a histogram for the given data

Step 2: The top right corner of the highest rectangle is joined by a straight line to the top right corner of the preceding rectangle.

Step 3: The top left corner of the highest rectangle is joined by a straight line to the top left corner of the succeeding rectangle.

Step 4: A perpendicular to the horizontal axis is drawn from the point of intersection of the two lines in step 2 and step 3.



Step 5: The value on the horizontal axis at the foot of the perpendicular indicate the mode of the given distribution.

Comparison of Mean, median and mode

1. Use: The arithmetic Mean is comparatively stable and is widely used than the Median and Mode. It is suitable for general purposes, unless there is any particular reason to select any other type of average. As far as the simplicity is concerned mode is the simplest of three.

Mode is the most usual or typical item and it can be located by inspection also. Median divides the set of observations/curve into two equal parts and is simpler than the mean. In certain cases Median is as stable as the mean.

2. Algebraic manipulation: Mean lends itself to algebraic manipulation. For example, we can calculate aggregate when the number of items and the average of the series is given. We can calculate the combined mean of two groups when the individual means and number observations are known.

Median and Mode cannot be algebraically manipulated.

3. Extreme and abnormal items: Presence of extreme and abnormal items can lead to certain misleading conclusion in case of mean. As for Mode and Median are concerned, they are not much influenced by the presence of abnormal items in the series. Median/mode should be used in such cases because they are least influenced.

Comparison of Mean, median and mode

4. Qualitative expression: Mean cannot be used when the data is qualitative or is not capable of numerical expressions. With the help of Median we can measure quantities which are capable of numerical expression. We can measure the intelligence or health of boys etc. Similarly, mode is the average that proves useful for non-numerical data.

5. Presence of Skewness: In case of a symmetrical curve, the value of mean, median and mode would coincide. The value of median and mean changes with the presence of positive or negative skewness. The value of mean changes to a greater extent than the value of median because it is affected by the position and value of every item.

6. Fluctuations of sampling: Mean is least affected by fluctuations of sampling. If the number of items is large, the abnormalities on the one side cancel the abnormalities on the other. Median distributes the curve into two equal parts and is affected by the fluctuations of sampling. Mode is affected to a great extent than even the median.

7. Classes with open end: Indeterminate mid-values will lead to inaccurate value of mean. Median and mode are not much affected by the presence of open end classes, except in case of extremely skewed curves.

8. Scales of measurement: When data are on interval scale the suitable measure of central tendency is mean. Median is suitable when data are on ordinal scale. Mode is calculated when data are on nominal scale.

Measures of Central Tendency

Geometric Mean

Geometric mean of a set of n observations is the n^{th} root of their product.

The Geometric mean (G.M) denoted by G of n observations $x_1, x_2, x_3, \ldots\ldots x_n$ is given by, **G =** $\boldsymbol{(x_1.x_2.x_3 \ldots\ldots x_n)^{\frac{1}{n}}}$

For the computation of G.M, we take logarithm on both sides so that,

Log G = log $(x_1.x_2.\ldots\ldots x_n)^{\frac{1}{n}} = \frac{1}{n}(log x_1 + log x_2 + \ \ldots + log x_n) = \frac{1}{n}(\sum_1^n log x_i)$

G = Antilog $\boldsymbol{\left[\frac{1}{n}(\sum_1^n log x_i)\right]}$

In case of frequency distributions, the Geometric mean G is given by,

G = $\boldsymbol{\left(x_1{}^{f_1}.x_2{}^{f_2}.x_3{}^{f_3} \ldots\ldots\ldots x_n{}^{f_n}\right)^{\frac{1}{N}}}$**,** where x_i, i=1,2,..n are the observations or mid values of classes and f_i, i=1,2,..n are the frequencies

Using Logarithm on both sides we get **G = Antilog** $\boldsymbol{\left(\frac{1}{N}\sum_1^n f_i log x_i\right)}$

Geometric Mean from Raw Data

G = $(x_1.x_2.x_3 \ldots\ldots x_n)^{\frac{1}{n}}$

G = Antilog $\left[\frac{1}{n}(\sum_1^n log x_i)\right]$

Geometric Mean from Frequency Distributions

G = $\left(x_1{}^{f_1}.x_2{}^{f_2}.x_3{}^{f_3} \ldots\ldots\ldots x_n{}^{f_n}\right)^{\frac{1}{N}}$

G = Antilog $\left(\frac{1}{N}\sum_1^n f_i log x_i\right)$

Measures of Central Tendency

Geometric Mean

Note 1: Geometric mean is used to

- find the growth rate (Population, interest etc),
- Construct index numbers

Note 2: if any one of the observations is zero, the Geometric mean becomes zero and if any one of the observations is negative, the geometric mean becomes imaginary

Note 3: Geometric mean of the combined group.

If n_1 and n_2 are the sizes and G_1 and G_2 are the Geometric means of two groups of observations, then the Geometric mean (G) of the combined group is given by,

$$\text{Log G} = \frac{n_1.logG_1+n_2.logG_2}{n_1+n_2} \text{ or G} = \text{Antilog}\left(\frac{n_1.logG_1+n_2.logG_2}{n_1+n_2}\right)$$

G.M is rigidly defined, based on all observations, suitable for further mathematical treatment, not much affected by fluctuations in sampling, gives weightage to smaller items

G.M is not easy to understand and also not easy to calculate

Measures of Central Tendency

Geometric Mean

Qn. 1 Calculate Geometric mean of the following data: 93, 97, 107,111, 118, 120.

X	Log X
93	1.9685
97	1.9868
107	2.0294
111	2.0453
118	2.0719
120	2.0792
Sum	**12.1810**

Geometric Mean from raw data is is given by

G = Antilog $\left[\frac{1}{n}\left(\sum_{1}^{n} log\, x_i\right)\right]$

$\therefore$ G = Antilog $\left[\frac{1}{6}(12.1810)\right]$

= Antilog(2.0301)

= 107.1766

Qn. 2 Calculate Geometric mean of the following data: 50, 100,1920, 143740, 204980, 154910, 1206740

Qn. 3 Calculate Geometric mean of the following data: 123, 32.1, 4.3, 0.51, .087, 0.00987

Measures of Central Tendency

Geometric Mean

Qn. 4 Calculate Geometric mean of the following data:

Value:	118	120	97	107	111	93
Frequency:	4	1	2	6	5	2

Geometric mean of an ungrouped frequency distribution is given by,

G = Antilog $\left(\frac{1}{N}\sum_1^n f_i log x_i\right)$

$= \text{Antilog}\left(\frac{1}{20} 40.6801\right)$

$= \text{Antilog} (2.0340)$

$=$ **108.1**

Value (X)	Frequency (f)	Log X	f. log X
118	4	2.0719	8.2875
120	1	2.0792	2.0792
97	2	1.9868	3.9735
107	6	2.0294	12.1763
111	5	2.0453	10.2266
93	2	1.9685	3.9370
	20		**40.6801**

Measures of Central Tendency

Geometric Mean

Qn. 5 Calculate Geometric mean of the following data:

Class:	0-9	10-19	20-29	30-39	40-49	50-59	60-69
Frequency:	14	23	37	45	33	22	16

Geometric mean from grouped frequency distribution is given by,

G = Antilog $\left(\frac{1}{N}\sum_{1}^{n} f_i \log x_i\right)$

$= \text{Antilog}\left(\frac{1}{190} 278.00\right)$

$= \text{Antilog}\ (1.4632)$

$=$ **29.05**

Class	f	MidX	Log X	f. log X
0 - 9	14	4.5	0.6532	9.1450
10 – 19	23	14.5	1.1614	26.7115
20 – 29	37	24.5	1.3892	51.3991
30 – 39	45	34.5	1.5378	69.2019
40 – 49	33	44.5	1.6484	54.3959
50 – 59	22	54.5	1.7364	38.2007
60 – 69	16	64.5	1.8096	28.9530
	190			278.00

Measures of Central Tendency

Harmonic Mean

Harmonic mean of a set of observations is the reciprocal of the arithmetic mean of their reciprocals.

The Harmonic mean (H.M) denoted by H of n observations $x_1, x_2, x_3, \ldots\ldots x_n$ is given by, $\mathbf{H} = \dfrac{1}{\frac{1}{n}\left(\frac{1}{x_1}+\frac{1}{x_2}+\ \ldots\ldots+\frac{1}{x_n}\right)} = \dfrac{n}{\sum\left(\frac{1}{x_i}\right)}$

In case of frequency distributions, the Harmonic mean H is given by, $\mathbf{H} = \dfrac{1}{\frac{1}{N}\left(\frac{f_1}{x_1}+\frac{f_2}{x_2}+\ \ldots\ldots+\frac{f_n}{x_n}\right)} = \dfrac{N}{\sum\left(\frac{f_i}{x_i}\right)}$ where x_i, i=1,2,..n are the observations or mid values of classes and f_i, i=1,2,..n are the frequencies

Note: H.M gives larger weights to smaller observations.

H.M is preferable to A.M as an average when there are few extremely large or smaller observations. H.M is useful in averages involving time, rate and price.

Measures of Central Tendency

Harmonic Mean

Qn.1If a car travels at 40 km/hr from A to B and at 50 km/hr from B to A, can you accept the statement, "Average speed of the car is 45 km/hr, the A.M of 40 and 50".

Answer: **Average speed of the car =** $\frac{\textit{Total distance travelled}}{\textit{Total Time of Travelling}}$

Let distance from A to B be D kms.

Time taken to travel D kms @ 40 km/hr = $\frac{Distance}{Time} = \frac{D\ km}{40\ km/hr} = \frac{D}{40}$ hr

Time taken to travel D kms @ 50 km/hr = $\frac{Distance}{Time} = \frac{D\ km}{50\ km/hr} = \frac{D}{50}$ hr

Total distance travelled = D +D = 2D Kms

Average speed = $\frac{\textit{Total distance travelled}}{\textit{Total Time of Travelling}} = \frac{2\ D}{\frac{D}{40}+\frac{D}{50}} = \frac{2}{\frac{1}{40}+\frac{1}{50}} = \frac{1}{\frac{1}{2}\left(\frac{1}{40}+\frac{1}{50}\right)}$ = H.M

Average speed = $\frac{1}{\frac{1}{2}\left(\frac{1}{40}+\frac{1}{50}\right)} = \frac{2\ x\ 40\ x\ 50}{40+50} = \frac{4000}{90} =$ **44.44**

Hence it is not possible to accept the statement that average speed of is 45 Km/Hr.

Note: A.M can be used to find the average speed if the duration of journey at both speeds are the same. (In that case, distance travelled will be different)

Measures of Central Tendency

Harmonic Mean

Qn.2 Calculate Harmonic mean of the following data: 12,15,21,25,32,37,40,45,48,50

Harmonic Mean (H) $= \dfrac{n}{\sum\left(\frac{1}{x_i}\right)}$

x	12	15	21	25	32	Sum
$1/x$	.083	.066	.048	0.04	.031	.268

Harmonic Mean, H $= \dfrac{5}{0.268} =$ **18.66**

Qn.3 Calculate Harmonic mean of the following data:

Weight (Kgs):	40 – 50	50 – 60	60 -70	70 – 80	80 – 90
No. of persons:	23	45	57	43	32

Harmonic Mean (H) $= \dfrac{N}{\sum\left(\frac{f_i}{x_i}\right)} = \dfrac{200}{3.156} =$ **63.37**

Class	f	Mid X	f/x
40 – 50	23	45	0.511
50 – 60	45	55	0.818
60 – 70	57	65	0.876
70 – 80	43	75	0.573
80 – 90	32	85	0.376
	200		3.156

Qn.4 (Assignment) Calculate Harmonic mean:

Size of item:	80	121	76	84	93
No if items:	12	34	45	34	18

Assignment 1

Calculate Mean, Median, Mode, G.M and H.M

(a) 9,21, 32,15, 11, 67,43, 8, 55, 60

(b) 0.24, 0.34, 0.15, 0.74, 0.45, 0.61

(c) 121.34, 145.23, 176.02, 154.74, 111.34, 185.54

(d) 125, 105, 95, 90, 75,120, 115, 80, 100, 120

(e) 0.012, 0.234, 0.00987, 0.0024, 0.0176, 0.000342

(f) 1100, 1250, 900, 1200, 1300, 1450, 1500, 950

Assignment 2

Calculate Mean, Median, Mode, G.M and H.M

(a)

Age:	21	42	38	64	53	61	47	55	
Number:		04	14	10	38	19	34	17	12

(b)

No of wrong calls per day:		0	1	2	3	4	5
Number of days:	12	17	23	37	42	31	

(c)

Value:	115	121	127	134	136	140	146	167
Frequency:	11	26	32	45	39	23	19	11

(d)

Class:	0 – 20	20 – 40	40 – 60	60 – 80	80 - 100
Frequency:	15	34	41	29	17

(e)

Class:	130 – 134	135 – 139	140 – 144	145 – 149	150 – 154	155 - 159
Frequency:	11	28	44	39	21	16

Measures of Central Tendency - Additional Problems

Qn. 1 Find the mean of the first n natural numbers.

$$\text{Arithmetic mean } (\bar{X}) = \frac{\sum x_i}{n} = \frac{1+2+3+\ \ldots\ldots+n}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{(n+1)}{2}$$

Qn. 2 Find the weighted arithmetic mean of the first n natural numbers where the weights being the numbers itself.

$$\text{Weighted arithmetic mean } (\bar{X}) = \frac{\sum w_i x_i}{\sum w_i} = \frac{1.1+2.2+\cdots+n.n}{1+2+3+\ \ldots..+n} =$$

$$\frac{1^2+2^2+\ \ldots\ldots+n^2}{1+2+3+\ \ldots..+n} = \frac{\frac{n(n+1)(2n+1)}{6}}{\frac{n(n+1)}{2}} = \frac{(2n+1)}{3}$$

Qn. 3 Find the mean of a, a+d, a+2d, …….(a+2nd)

$$\text{Arithmetic mean } (\bar{X}) = \frac{a+(a+d)+(a+2d)+\ \ldots\ldots+(a+2nd)}{2n+1}$$

$$= \frac{\frac{(2n+1)}{2}[2a+\{(2n+1)-1\}d]}{2n+1} = = \frac{\frac{(2n+1)}{2}[2a+2nd]}{2n+1} = \mathbf{a+nd}$$

Measures of Central Tendency - Additional Problems

Qn. 4 The frequencies of variable values 0,1,2,3….n are given by nC_x $q^{n-x}p^x$, x = 0,1,2,…..n and (p+q)=1. Find the arithmetic mean.

$$\text{Arithmetic mean } (\bar{X}) = \frac{\sum f_i x_i}{N} = \frac{\sum_{x=0}^{n} \mathrm{n}C_x \; q^{n-x}p^x \,.\, x}{\sum_{x=0}^{n} \mathrm{n}C_x \; q^{n-x}p^x}$$

$$= \frac{\sum_{x=0}^{n} \; \frac{n!}{x!(n-x)!} q^{n-x}p^x \,.\, x}{\mathrm{n}C_0 \; q^{n-0}p^0 + \mathrm{n}C_1 \; q^{n-1}p^1 + \cdots . + \mathrm{n}C_n \; q^{n-n}p^n}$$

$$= \frac{\sum_{x=0}^{n} \; \frac{n(n-1)!}{x\,(x-1)![(n-1)-(x-1)]!} q^{(n-1)-(x-1)}p \; p^{x-1} \,.\, x}{(q+p)^n}$$

$$= \mathrm{np} \sum_{x=1}^{n} \; \frac{(n-1)!}{(x-1)![(n-1)-(x-1)]!} q^{(n-1)-(x-1)} \; p^{x-1}$$

$$= \mathrm{np} \sum_{x=1}^{n} \; (n-1)C_{(x-1)} q^{(n-1)-(x-1)} \; p^{x-1}$$

$$= \mathrm{np} \; (q+p)^{n-1} = \boldsymbol{np}$$

Measures of Central Tendency - Additional Problems

Qn. 5 The variable takes values a, ar, ar^2, ar^3,……ar^{n-1}each with frequency unity. Show that the A.M is A = $\frac{a(1-r^n)}{n(1-r)}$, G.M is G = a$r^{\frac{n-1}{2}}$, H.M is H = $\frac{an\,(1-r)r^{n-1}}{1-r^n}$ and AH = G^2

Ans: A. M (A) = $\frac{\sum x_i}{n} = \frac{a+ ar+ ar^2+ \ \ldots\ldots +ar^{n-1}}{n} = \frac{\frac{a(1-r^n)}{(1-r)}}{n} = \boldsymbol{\frac{a(1-r^n)}{n(1-r)}}$

G.M (G) = (a. ar. ar^2. ar^3……a$r^{n-1}$$)^{\frac{1}{n}}$ =$\left\{a^n\, r^{1+2+3+\ \ldots\ldots + (n-1)}\right\}^{\frac{1}{n}}$

$= a\left\{r^{\frac{(n-1)n}{2}}\right\}^{\frac{1}{n}} \boldsymbol{= ar^{\frac{n-1}{2}}}$

H.M (H) = $\left\{\frac{\frac{1}{a}+\frac{1}{ar}+\frac{1}{ar^2}+\ \ldots\ldots + \frac{1}{ar^{n-1}}}{n}\right\}^{-1} = \left\{\frac{1}{n}\left[\frac{\frac{1}{a}\left(\frac{1}{r^n}-1\right)}{\left(\frac{1}{r}-1\right)}\right]\right\}^{-1} = \boldsymbol{\frac{an\,(1-r)r^{n-1}}{(1-r^n)}}$

A.H = $\boldsymbol{\frac{a(1-r^n)}{n(1-r)}} \cdot \boldsymbol{\frac{an\,(1-r)r^{n-1}}{(1-r^n)}} = a^2.\boldsymbol{r^{n-1}} = \left\{\boldsymbol{ar^{\frac{n-1}{2}}}\right\}^2 = G^2$