CHAPTER THREE

MEASURES OF CENTRAL TENDENCY

Specific Objectives

At the end of this topic the trainee should be able to:

- Define the measures of central tendency;
- State the properties of the measure of central tendency;
- Determine the measures of central tendency.

Introduction

Measures of central tendency show the tendency of some central value around which data tends to cluster. These are statistical values which tend to occur at the centre of any well ordered set of data. Whenever these measures occur they do not indicate the centre of that data

Objectives of averaging

- To get one single value that describes the characteristics of the entire data.
- To facilitate comparison.

Properties of Good Measure

- It should be easy to understand i.e. since statistical methods are designed to simplify complexity, it is described that an average be such that can be readily understood, its use is bound to be very limited.
- Should be simple to compute.
- Should be based on all observation.
- Should be rigidly defined, i.e. an average should be properly defined so that it has one and only one interpretation.
- Should be capable of further algebraic treatment.
- Should have sampling stability.
- Should not be unduly affected by the presence of extreme values.

N/B. The following statistical terms are commonly used in statistical calculations. They must therefore be clearly understood.

i)Class limits

These are numerical values which limits the extend of a given class i.e. all the observations in a given class are expected to fall within the interval which is bounded by the class limits e.g. 15 & 19 are class limits as in the table of the example above.

ii) Class boundaries

These are statistical boundaries, which separate one class from the other. They are usually determined by adding the lower class limit to the next upper class limit and dividing by 2 e.g. in the above table the class boundary between 19 and 20 is 19.5 which is = $\frac{19+20}{2}$.

iii) Class mid points

These are very important values which mark the center of a given class. They are obtained by adding together the two limits of a given class and dividing the result by 2.

iv) Class interval/width

This is the difference between an upper class boundary and lower class boundary. The value usually measures the length of a given class.

The following are important measures of central tendency

- Arithmetic mean
- Median
- Mode
- Geometric mean
- Harmonic

Arithmetic Mean

The most popular and widely used measure for representing the entire data by one value is an average. Its value is contained by adding together all the observation and dividing this total by the number of observation
These is commonly known as average or mean_it is obtained by first of all summing up the values given and by dividing the total value by the total no. of observations.

I.e. mean $= \frac{\sum X}{n}$

Where x = no. of values

$\sum$ = summation

n = no of observations

Example

The mean of 60, 80, 90, 120

$$\frac{60+80+90+120}{4}$$

$$=\frac{350}{4}$$

$$= 87.5$$

The arithmetic mean is very useful because it represents the values of most observations in the population.
The mean therefore describes the population quite well in terms of the magnitudes attained by most of the members of the population

Computation of the mean from grouped Data i.e. in classes.

The following data was obtained from the manufacturers of electronic cells. A sample of electronic cells was taken and the life spans were recorded as shown in the following table.

Life span hrs	No. of cells (f)	Class MP(x)	X - A = d	fd
1600 – 1799	25	1699.5	-600	-15000
1800 – 1999	32	1899.5	-400	-12800
2000-2199	46	2099.5	-200	-9200
2200 – 2399	58	2299.5(A)	0	0
2400 – 2599	40	2499.5	200	8000
2600 – 2799	30	2699.5	400	12000
2800 – 2999	7	2899.5	600	4200

A = Assumed mean, this is an arbitrary number selected from the data, MP = mid point

$$\text{Arithmetic mean} = \text{assumed mean} + \frac{\sum fd}{\sum f} = 2299.5 + -\frac{12800}{238}$$

= 2299.5 +-53.78

= 2245.72 hours

Example 2 – (use of the coded method)

The following data was obtained from students who were registered in a certain college.
The table shows the age distribution

Age (yrs)	No. of Students (f)	mid points (x)	x-a = d	D/c = u	fu
15 - 19	21	17	-15	-3	-63
20 - 24	35	22	-10	-2	-70
25 - 29	38	27	-5	-1	-38
30 - 34	49	32(A)	0	0	0
35 - 39	31	37	+5	+	31
40 - 44	19	42	+10	+2	38
	193				-102

Required calculate the mean age of the students using the coded method

$$\text{Actual mean} = \text{A(assumed mean)} + \frac{\sum fu}{\sum f} \times c$$

$$= 32 + \frac{-102}{193} \times 5$$

= 29.36 years

<u>Merits of arithmetic mean</u>

- It possess first six our seven characteristics of good average and no other average possess such large number of characteristics
- It is unduly affected by the presence of extreme value.
- It utilizes all the observations given.
- It is a very useful statistic in terms of applications. It has several applications in business management e.g. hypothesis testing, quality control e.t.c.
- It is the best representative of a given set of data if such data was obtained from a normal population.
- The a.m. can be determined accurately using mathematical formulas.

<u>Limitation</u>

- It is difficult to compute mean without making assumption regarding the size of the class- interval of the open end classer.

- If the data is not drawn from a 'normal' population, then the a.m. may give a wrong impression about the population.
- In some situations, the a.m. may give unrealistic values especially when dealing with discrete variables e.g. when working out the average no. of children in a no. of families. It may be found that the average is 4.4 which is unrealistic in human beings.

<u>The Median</u>

This is a statistical value which is normally located at the center of a given set of data which has been organized in the order of magnitude or size e.g. consider the set 14, 17, 9, 8, 20, 32, 18, 14.5, 13. When the data is ordered it will be 8, 9, 13, 14, 14.5, 17, 18, 20, and 32. The middle number/median is 14.5. The importance of the median lies in the fact that it divides the data into 2 equal halves. The no. of observations below and above the median is equal. In order to determine the value of the median from grouped data, the data is grouped the median may be determined by using the following methods

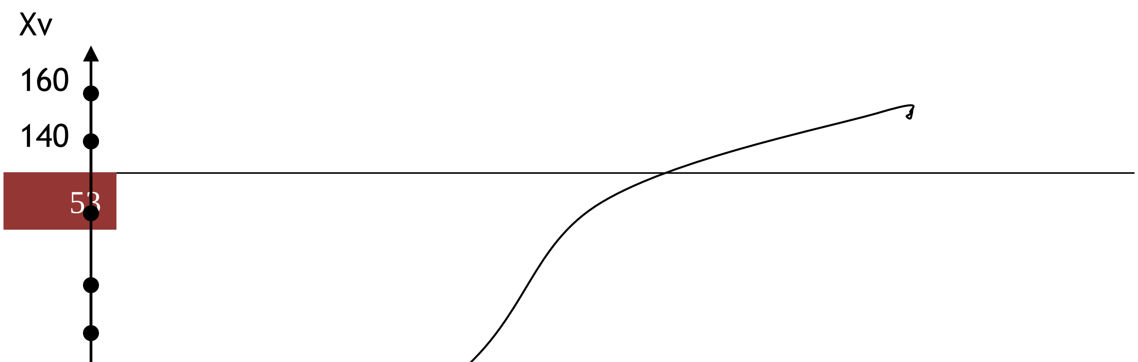
i. Graphical method using the cumulative frequency curve (ogive).
ii. The formula

<u>Example</u>

Referring to the table in 105, determine the median using the methods above

<u>The graphical method</u>

IQ	No of resid	UCB	Cumulative Frequency
0 – 20	6	20	6
20 – 40	18	40	24
40 – 60	32	60	56
60 – 80	48	80	104
80 – 100	27	100	131
100 – 120	13	120	144
120 – 140	2	140	146
	146		



120

100

80

60

40

20

20 40 60 80 100 120 140 160

Value of the median

The position of the median = $\frac{n+1}{2} = \frac{146+1}{2}$

Computation

The formula used is

Media= L +i/f{m-c}

Where L = Lower class boundary of the class containing the median

m = (n/2)th or (n+1/2)th

c= Cumulative frequency of the class before that containing the median.

f = Frequency of the class containing the median

Media= L +i/f{m-c}

$$= 60 + \frac{73.5 - 56}{48} \times 20$$

= 60 + 7.29

= 67.29

Merits of median

- It is not influenced by the presence extreme values
- The sum of the deviation of observation from median is minimum
- It shows the centre of a given set of data
- Knowledge of the determination of the median may be extended to determine the quartiles
- The median can easily be defined
- It can be obtained easily from the cumulative frequency curve

- It can be used in determining the degrees of skewness

Limitation

- Its value is not determined by each and every observation.
- Median is not capable of algebraic treatment.
- Median is less reliable.
- It is affected by sampling fluctuation.
- In some situations where the no. of observations is even, the value of the median obtained is usually imaginary.
- The computation of the median using the formulas is not well understood by most businessmen.
- In business environment the median has got very few applications.

Quartiles, Deciles, Percentiles

The procedure for computing quartiles deciles etc is the same as for median For grouped data the following formula are used for quartiles, deciles and percentiles:

$$Q1 = L + i/f\ (m-c)$$

Where, Q1= First or lower quartile.
L= Lower class limit of the lower quartile group.
i = class interval of the lower quartile group.
f = frequency of the lower quartile group.
m = n/4
c = cumulative frequency of the group preceding the lower quartile group.

Decile:

$$D2 = L + i/f\ (m-c)$$

Where, D2 = second decile
m = 2n/10

Percentile:

$$P40 = L + i/f\ (m-c)$$

Where, P40 = fortieth percentile.
m = 40n/100

Example

The profit earned by 100 companies during 2003-04 are given bellow

Profits (Kshs.)	Number of companies
20-30	4
30-40	8

40-50	18
50-60	30
60-70	15
70-80	10
80-90	8
90-100	7

Calculation Q median, D4 and P80 and interpret the values.

<u>Solution</u>

Calculation of Q1, Q2, Q4, and P80

Profits (Kshs.)	f	cf
20-30	4	4
30-40	8	12
40-50	18	30
50-60	30	60
60-70	15	75
70-80	10	85
80-90	8	93
90-100	7	100

Q1 =size of N/4th observation = 100/4 =25th observation
Hence q1 lies in the class 40-50
Q1= L+ i/f (m-c)

= 40 +10/18(25-12)
= 40 + 7.22 = 47.22
= 47.22

D4 = L+ i/f (m-c)
= 50 +10/30 (40-30)
= 50 + 3.33
= 53.33

Thus 40 percent of the companies earn an annual profit of kshs.53.33 or less

P80 = size of 80/100 the observation = 80 the observation
P80 lies in the class 70-80
P80 = L+ i/f (m-c)
= 70 + 10/10(80-75)
= 70 +5

= 75

This means that 80 percent of the companies earn an annual profit of kshs.75 or less

20 percent of the companies earn an annual profit of more than kshs. 75

Mode

This is one of the measures of central tendency. The mode is defined as a value within a frequency distribution which has the highest frequency. Sometimes a single value may not exist as such in which case we may refer to the class with the highest frequency. Such a class is known as a **modal class**

The mode is a very important statistical value in business activities quite often business firms tend to stock specific items which are heavily on demand e.g. footwear, clothes, construction materials (beams, wires, iron sheets e.t.c.

The mode can easily be determined form ungrouped data by arranging the figures given and determining the one with the highest frequency.

When determining the values of the mode from the grouped data we may use the following methods;-

i. The graphical method which involves use of the histogram
ii. The computation method which involves use of formula

Example

In a social survey in which the main purpose was to establish the intelligence quotient (IQ) of resident in a given area, the following results were obtained as tabulated below:

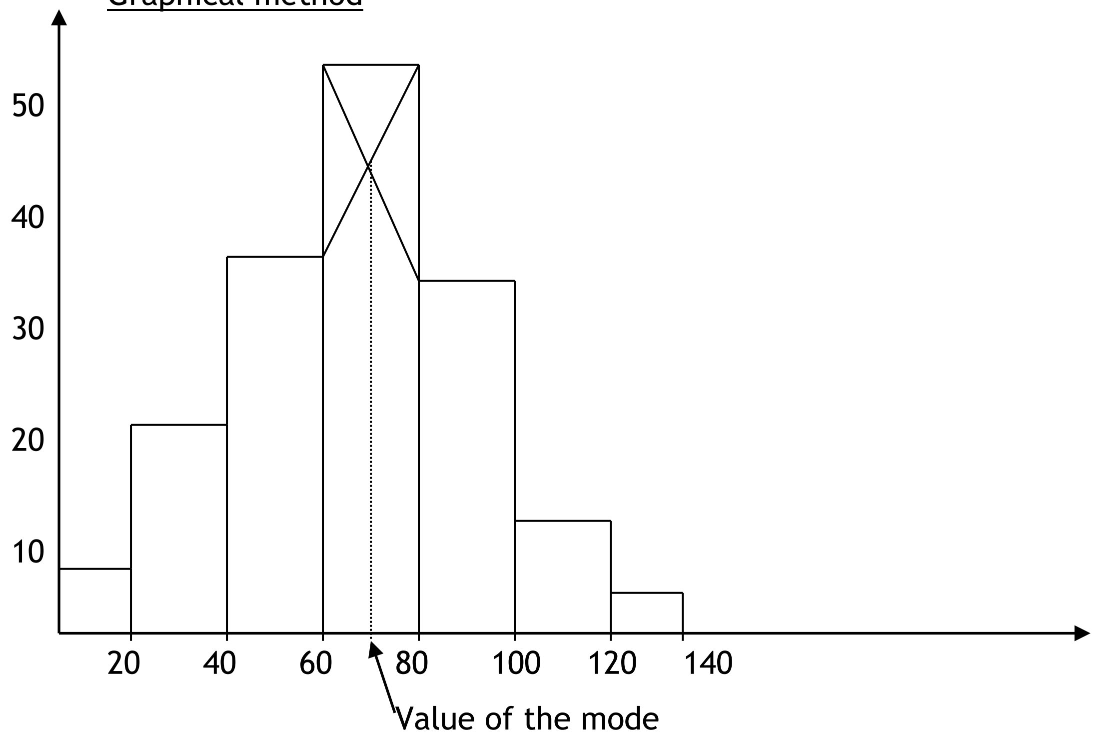
IQ	No. of residents	Upper class bound	CF
1 – 20	6	20	6
21 – 40	18	40	24
41 – 60	32 fo	60	56
61 – 80	48 f1	80	104
81 – 100	27 f2	100	131
101 – 120	13	120	144
121 – 140	2	140	146

Required

Calculate the modal value of the IQ's tabulated above using

i. The graphical method and
ii. Formula

Graphical method



Computation method

$$\text{Mode} = L + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times c$$

Where L = Lower class boundary of the class containing the mode

f_0 = Frequency of the class below the modal class
f_1 = Frequency of the class containing the mode
f_2 = frequency of the class above the modal class
c = Class interval

$$\text{Therefore Mode} = 60.50 + \left(\frac{(48 - 32)}{2(48) - 32 - 27}\right) \times 20$$

$= 69.14$

Merits of mode

- Mode is not affected by extreme values.
- At can be easily used to decide qualitative phenomenon.
- Mode happens to be meaningful as an average.
- It can be determined from incomplete data provided the observations with the highest frequency are already known.
- The mode has several applications in business.

- The mode can be easily defined.
- It can be determined easily from a graph.

Limitation of mode

- Mode cannot always be computed
- If the data is quite large and ungrouped, determination of the mode can be quite cumbersome
- Use of the formula to calculate the mode is unfamiliar to most business people
- The mode may sometimes be non existent or there may be two modes for a given set of data. In such a case therefore a single mode may not exist

Geometric mean

This is a measure of central tendency normally used to measure industrial growth rates. It is defined as the nth root of the product of 'n' observations or values

- i.e. $GM = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n}$

Example

In 1995 five firms registered the following economic growth rates; 26%. 32% 41% 18% and 36%.

Required

Calculate the GM for the above values

$$GM = \sqrt[5]{26 \times 32 \times 41 \times 18 \times 36}$$

$$= \tfrac{1}{5}\left[Log\,26 + Log\,32 + Log\,41 + Log\,18 + Log\,26\right]$$

No.	Log
26	1.4150
32	1.5052
41	1.6128
18	1.2533
36	1.5563
	7.3446

Therefore Log of GM = 1/5 x 7.3446 = 1.46892

So GM = Antilog of 1.46892

= 29.43

Merits

i. It makes use of all the values given (except when x = 0 or negative)
ii. It is the best measure for industrial growth rates

Demerits

i. The determination of the GM by using logarithms is not familiar process to all those expected to use it e.g. managers
ii. If the data contains **zeros** or **-ve** values, the GM ceases to exist

Harmonic mean

This is a measure of central tendency which is used to determine the average growth rates for natural economies. It is defined as the reciprocal of the average of the reciprocals of all the values given by HM.

$$HM = \frac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \ldots \frac{1}{x_3}\right)}$$

Example

The economic growth rates of five countries were given as 20%, 15%, 25%, 18% and 5%
Calculate the harmonic mean

$$\text{The HM} = \frac{1}{\frac{1}{5}\left(\frac{1}{20} + \frac{1}{15} + \frac{1}{25} + \frac{1}{10} + \frac{1}{5}\right.}$$

$$= \frac{1}{0.2(0.05 + 0.07 + 0.04 + 0.10 + 0.2)}$$

$$= \frac{1}{0.092}$$

10.86%

Merits – same as the arithmetic mean
Demerits – same as the arithmetic mean

Weighted mean

- This is the mean which uses arbitrarily given weights
- It is a useful measure especially where assessment is being done yet the conditions prevailing are not the same. This is particularly true when assessment of students is being done given that the subjects being taken have different levels of difficulties.

Examples

The following table shows that marks scored by a student doing section 3 and 4 of CPA

Subject	Scores (x)	Weight (w)	wx
STAD	65	50	3250
BF	63	40	2520
FA2	62	45	2340
LAW	80	35	2800
QT	69	55	3795
FA3	55	60	3300
		w = 285	wx = 18005

Weighted mean

$$\frac{Ewx}{Ew}$$

$$= \frac{18005}{285}$$

$$= 63.17\%$$