Stat 101

* Nature of statistic its defination, Importance & Limitation

Nature (origin and development) of statistic, in a sense, is as old as the human society itself. It has been use right on the existence of life on this earth. Its origin can be traced to the old days when it was regarded as the science of state-craft and was the by product of administrative activity of the state. The word "statistic" seems to have been derived from the latin word "status" or the italian word "statista" or the german word "statistik" each of with a means a "political state"

In ancient times, the government use to collect the information regarding the population "property or wealth" or the country. The former enabling the government to have an idea of the man power of the country and the after prowiding its basis for introducing new taxes and levies. Its therefore very important to start a course of this kind with an explanation of the purpose and concept of statistical science.

Statistic is basically a science of

P. T.O.

Collecting, organizing, analysing, presenting and Interpreting data. These five purposes are Interlwined and any one is usually Controled by another to a large extend, but untill the beginning of the 19th century far more attention was paid to the collection and presentation of data than to there Interpretation. Large Volume of data were usually Collected and frequently misInterpreted if Indeed Interpretation was attempted.

However, since that time the Importance of a scientific approach in the Interpretation of data has been realized and greats steps have been made in the development of appropriate methods. Correspondingly, method of collecting and presentation of data have been altered to keep pace with the new methods of analysis and Interpretation especially in this era of Information and Communication Technology where Computers formed the center stage of data analysis until recently. The Interpretation of data hold a central position in statistic and method of collection and presentation are hinged method

of Interpretation.

The growth of statistic especially in the methods of data Interpretation during the William Sealy, Karl pearson, franic yales, Charles edward Spearman ↕ past 60y may be linked with the names of ↕ and in particular Sir Ronald Aylmer fisher. this growth stated in the field of biological research, but the nature of the problems in Counter has Caused the new method to be applied to medical, physycologycal and economic data and to some extend in physics and engineering. example the biologist in his routine work is Comfronted with the difficulty that the measurement after Identical threatment, or two animals or plants apperently Similar in all assets, can give widely different result.

Again the background, statistic has been defined as the science With deals with the Collection, organisation, presentation, analysing and Interpretation of numerical data with aim of drawing a logical Conclution. statistic is the most veritable tool for planning policy formulation and decision macing in

every life, in every sector, decision are taking and plans are made but the success of every plan depends on the accuracy, analysis and Interpretation of the available data this shows that statistic placed a vital role in all planning and decision making process.

Typically, there are two main branches of statistic which are!

Descriptive statistic and Inferential statistic both of these are employed in scientific analysis of data and both are equally important.

* Descriptive statistic - these simply deals with the collection and presentation of data which is usually the first part of statistical analysis.
* Inferential statistics - as the name suggest this Involve drawing the right Conclution from the statistical analysis that has been performed using descriptive statistic.

IMPORTANT OF STATISTIC

in modern times, statistic is the back bone of the decision making and it been introduced in to all sector of life, medical, Industrial, economics, political, among others.

statistic is no longer Consider as a ~~tool~~ mere device for collecting numerical data but as a tool for developing sound techniques for data handling, analysis and making viable Inference for them, some of the importance and uses of statistic are as follows :-

1. statistic and planning - statistic is dispensable to planning all over the world government especially of the developing economic are taking to planning for economic and technological aspect.
2. statistic and economics - statistic and techniques of statistical analysis have proved Immensely Important in solving the varities of economics problems such as wages prices, time series analysis, in fact wide application of mathematics and statistic in the study of economic have lead to the development of new discipline called the economic statistic and econmetrics
3. statistic and business - statistic is an Indispensable tool in product marketing, Inventry analysis and

production Control. Business executives are reline more and more on statistical techniques for studying the needs and desires of the Consum and for many other business purposes. The success of a business man depends so much on the accuracy and precision of his statisti forecasting

4 Statistic and Industry - in the Industryal sector statistic is widely use in quality control, in production engineering as to know wether the product is Comfarm to specification or not. Statistical tools like Inspection plant, control ~~are~~ of client, acceptance Inspection and process Control of extreme Importanie in the Industry. statistical data analysis has also proved useful in research and scietific field like biology, astronomy and medical scien

LIMITATIONS OF STATISTIC

1. If sufficient care is not taken in collecting analysis and Interpreting data, statistical might be misleading.
2. statistical method are only based applicable to

quantity data and not suited for quality phenominon.

3. statistic can not be applied to heteragenous data.

4. statistic doesn't study individual and therefore has no recognised to the individuals items of a series.

5. unlike the laws of physical and natural sciences, statistical laws are only approximation and not exact.

Types of statistical data

There are two general types of statistical data these are :-

1. Qualitative data
2. Quantitative data.

* Qualitative data - In certain investigation, we are only concern with the present or absent of some characteristics in a sect of object or individual. for example if we have record of birth, we are interested only to know whether is male or female and count the number of male babies. Similarly, if a coin is

tossed a number of times we may only n
a number of head in a given set of tosses
This type of data is called qualitative
data. The characteristics used to classi
the individual into different catergories is
called attributed. typically examples of
attributed are eye colour, religion, gend
Nationality.

* Quantitative data - when we are interested
in a variable we either measure the mag.
nitude of each of the individuals or unit
under consideration. Thus types of data is
called quantitative data example weight
of a person, The speed of a car, weight of
a person, age e.t.c

The primary and secondary data

Data may be gathered from 2 main sources
namely.

1. primary data - primary data are those dat
gathered directly from the source, that is, when
data is collected directly from the observe values
for e.g, data collected true questionaire

Interviews, observations or experiment.

2. Secondary data - are those collected ~~data gathered directly~~ from the other source, but essentially not directly from observed values for example data collection through transcription, or documentation from existing record. data collected from magazines, Journals, library, achives among others.

Note

The measure difference between primary and secondary data is that secondary data have to be extracted from data that was probably collected from a different purposes whereas primary data is collected purposely to the users practical need.

METHOD of DATA COLLECTION

Data are generally needed, collected and analysis to provide useful and meaning Information for planning and execution of a survey or experiment depends on the types of data needed which is great influence by the method through with the data are.

collected the decision and choice of method of the data collection should be arrived at after considering the aims and objectives of the survey on experiment, the nature of information needed, the population and study the degree of accuracy desired, practicability of result, time and cost. Thus, the method are discussed as follows

1. mail or postal Questionares method - in this method, prepared questionaires on the subject matter of the survey are sent through an agent such as post office or email to the respondent requesting them back by posting online or directly.
2. personal Interview method - this method requires the interviewer to ask prepaired set of questions from the respondent and record the answer this method allowed the interviewer and the respondent to meet fare to face this method is widely use in Journalism and population census
3. Telephone interview method - with the modern telephone System, an Interviewer Can obtaine

Information from respondent by making use of a telephone some decade ago, telephone interview was most common in advance country where there is good No of telephone substance and good telecommunication system. Now with the advent of mobile phones the telephone interview could be use in the development country.

4. Direct observation method - this is the method of data collection by physical observation or measurement of the unit, item or respondent under study. The researcher record the result using the appropriate measuring instrument

5. Experimental method - In this method the derived data are collected through in the design of an and the success or failure of this method depends on the skills of the experimenter as well as the quality of the instrument used for the experiment.

6. method of registration - In this method, the respondent are required to registered the required

the information at some designated places the vital statistic registeration in many country provides and illustration of this method.

7. Transcription or Documentation method - this method is used when the data is needed for a specified survery - data already in registration files, document among others are usually collected by this method this is a method of collecting secondary data.

Graphical and Diagramatical Representation of data

Graphical representation of data the most convinient and popular way of describing data is using graphical presentation its easier to understand and interprete data when they are present graphically than using words or a frequency table. a graph can present data in a simple and clear way. also it can illustrate the important

aspect of the data. these lead to better analysis and presentation of the data

Statistical graphs or chart served two purposes. The first purpose is the presentation of statistical information on graphs and chart is to make the reader appreciate a simple, evocative display. The second main use is as a private aid to statistical analysis. The statistician will often have recourse to graphs and chart to gain insight in to the structure of the data and to take assumptions which might be made in an analysis. Data can be presented in any of the following forms;

1. pie chart
2. Bar chart
3. Histogram
4. frequency polygon
5. Cummulative frequency Curve (Ogive)

pie chart

With pie chart data are represented with a cycle diuided into sectors with angl in each sector propositional to the freque in that particular group. Angle subtended at the center of the cycle is 360° the angle in each group is computed by the formular below;

$$\text{Angle} = \frac{\text{frequency}}{\text{total}} \times 360^\circ$$

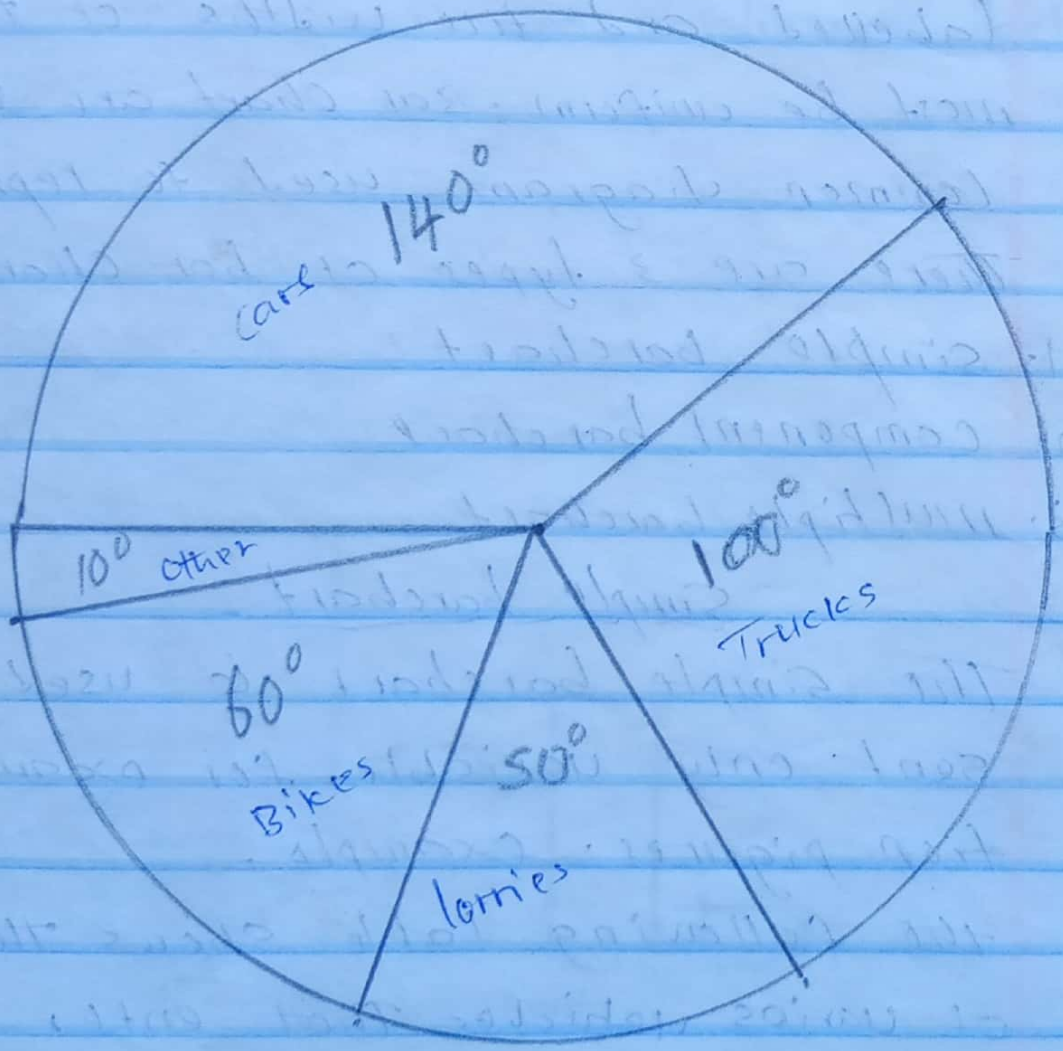
example

the following table shows the distribution of various vehicles that enter a teaching hospital in a day. Represente the Information in a pie chart

Vehicle	number of vehicle
Cars	70
Trucks	50
Lorries	25
Bikes	30
others	5
Total	180

Soln

Vehicles	Frequency	Angle
Cars	70	$\frac{70}{180} \times 360^\circ = 140^\circ$
Trucks	50	$\frac{50}{180} \times 360^\circ = 100^\circ$
Lorries	25	$\frac{25}{180} \times 360^\circ = 50^\circ$
Bikes	30	$\frac{30}{180} \times 360^\circ = 60^\circ$
Others	5	$\frac{5}{180} \times 360^\circ = 10^\circ$
Total	180	360°



pie chart

Bar chart

with bar chart data are represented with a series of equally spaced rectangles called bar. The bars are drawn either vertically or horizontally such that there height correspond or are proportional to the frequency in each group. for clearity, both axis of the chart most be properly labelled and the widths of the bars most be uniform. Bar chart are the most common diagram used to represent data there are 3 types of bar chart namely;

1. simple barchart
2. component barchart
3. multiple barchart

Simple barchart

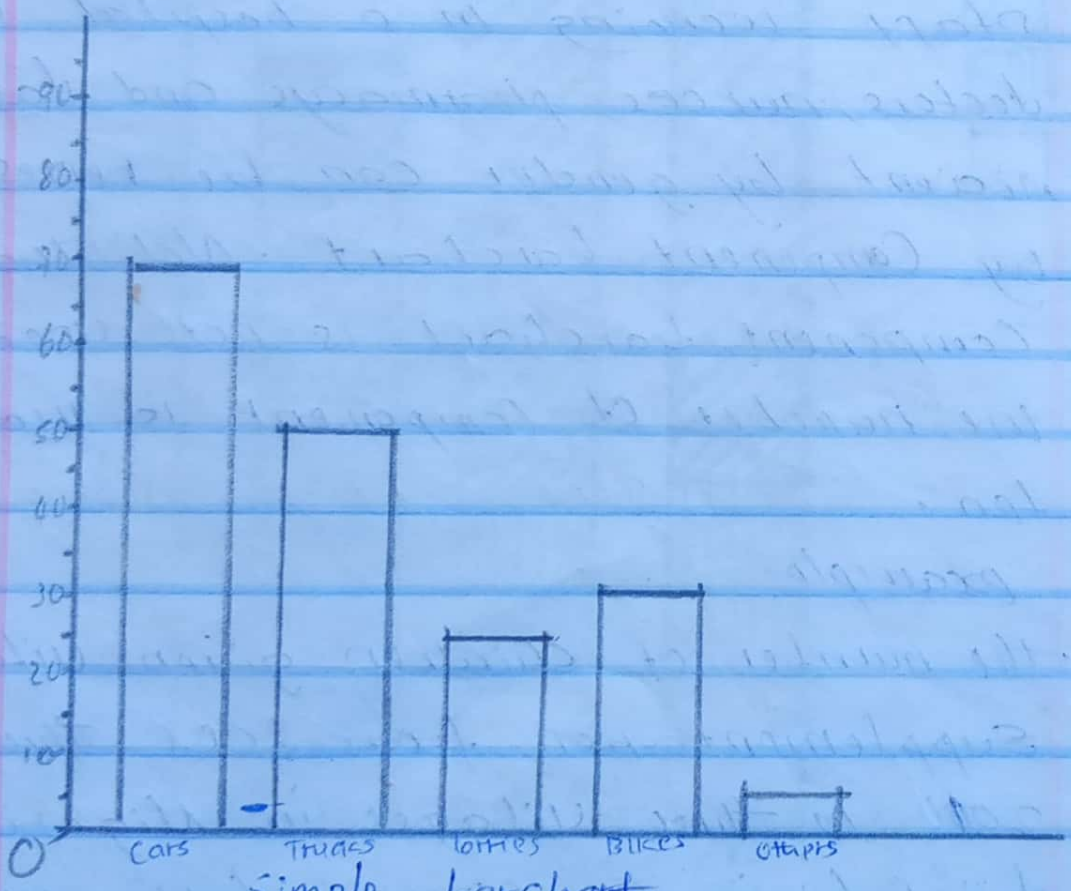
The simple barchart is used to represent only variable, for example population figures. example -

the following table shows the distribution of varios vehicles that enter a teaching hospital in a day. Represent the Information in a bar chart

/

Vehicles	Number of Vehicles
Cars	70
Trucks	50
Lorries	25
Bikes	30
others	5
total	180

The simple bar chart in diagram below



simple barchart

Component barchart

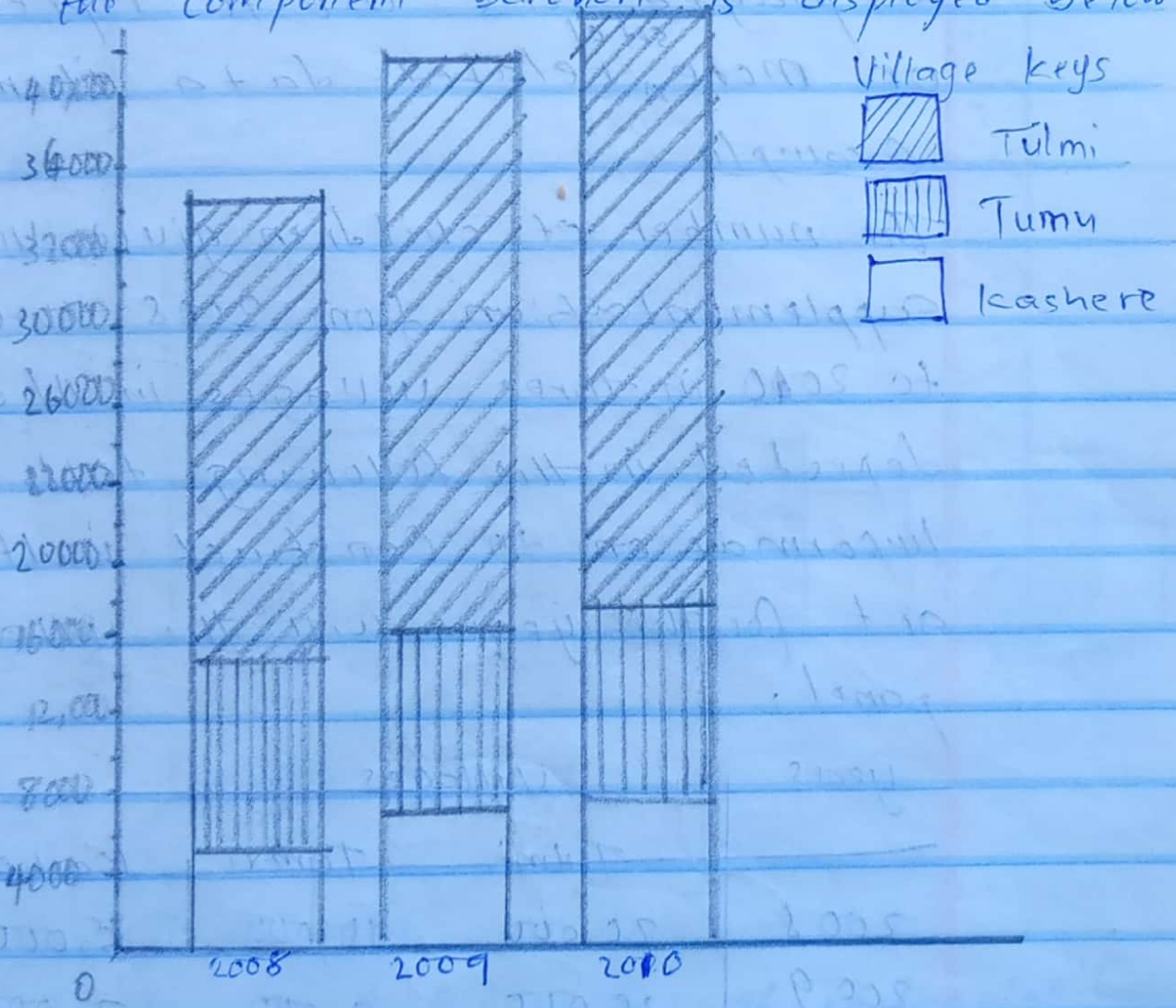
- The component barchart is used to represent two or more variable, for examples population figures by gender where male and female population are displayed. its used to represent the magnitude of a giving phenomena which is further subdivided into its various component. for example the Number of professionay staff working in a hospital like doctors, nurses, pharmacys and labotechnicient by gender can be represented by Component barchart. Not that the Component barchart is not used when the number of component is more than ten.

example

- the number of children given vitamin A supplementation from 2008 throughout to 2010 in three vilages in Nigeria is depicted in the following table. used the information to construct Component barchat for the years with the vilages as panel.

years	Villages			
	Tulmi	Tumu	kashere	total
2008	20,000	10,000	5,000	35,000
2009	28,000	9,000	7,000	44,000
2010	31,000	9,500	7,500	48,000
total	79,000	28,500	19,500	127,000

the Component barchart is displayed below



Component barchart

Multiple barchart

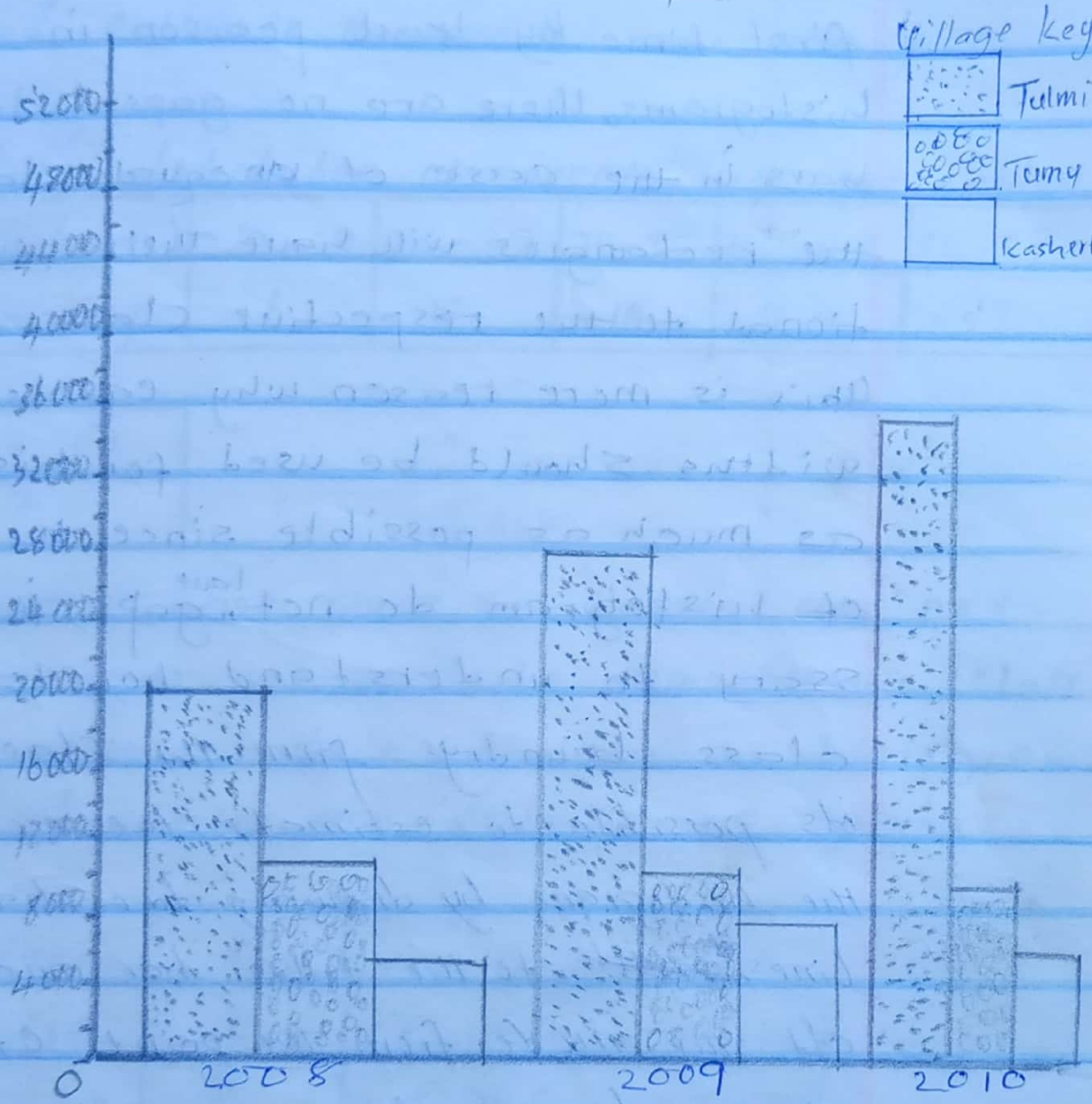
The multiple barchart are also used to represent two or more variable with different bars side by side, with different shedins or colours. for example population figures by gender where male and female population are displayed its use to represent two or more set of related data items.

Example

The number of children given Vitamin A supplementation from 2008 through to 2010 in three villages in Nigeria is depicted in the following table. used the information to construct multiple barchart for the years with the villages as panel.

years	villages			
	Tulmi	Tumu	Kashere	total
2008	20,000	10,000	5,000	35,000
2009	28,000	9,000	7,000	44,000
2010	31,000	9,500	7,500	48,000
total	78,000	28,500	19,500	127,000

the multiple Barchart displayed below



multiple barchart

Histogram

A histogram is the series of rectangles having areas proposional to the frequency of each class. the term histogram was used for the

first time by Karl pearson in 1895. in histograms, there are no gaps between the bars in the case of unequal classwidth, the rectangles will have their width proportional to the respective class sizes. this is more reason why equals class widths should be used for histogram as much as possible since the bars of histogram do not have gaps, it is necessary to understand how to obtain class boundrys from the class limits. its possible to estimate the mode from the histogram by drawing two diagonal lines to the tallest bar and reading off the mode from the point of intercept of the diagonal

example

an opthalmologies from the national eye cen observed the following data on the distribution of 50 people by age who suffer from eye problem in a particular village use the data to draw a histogram

and estimate the modal age from your histogram.

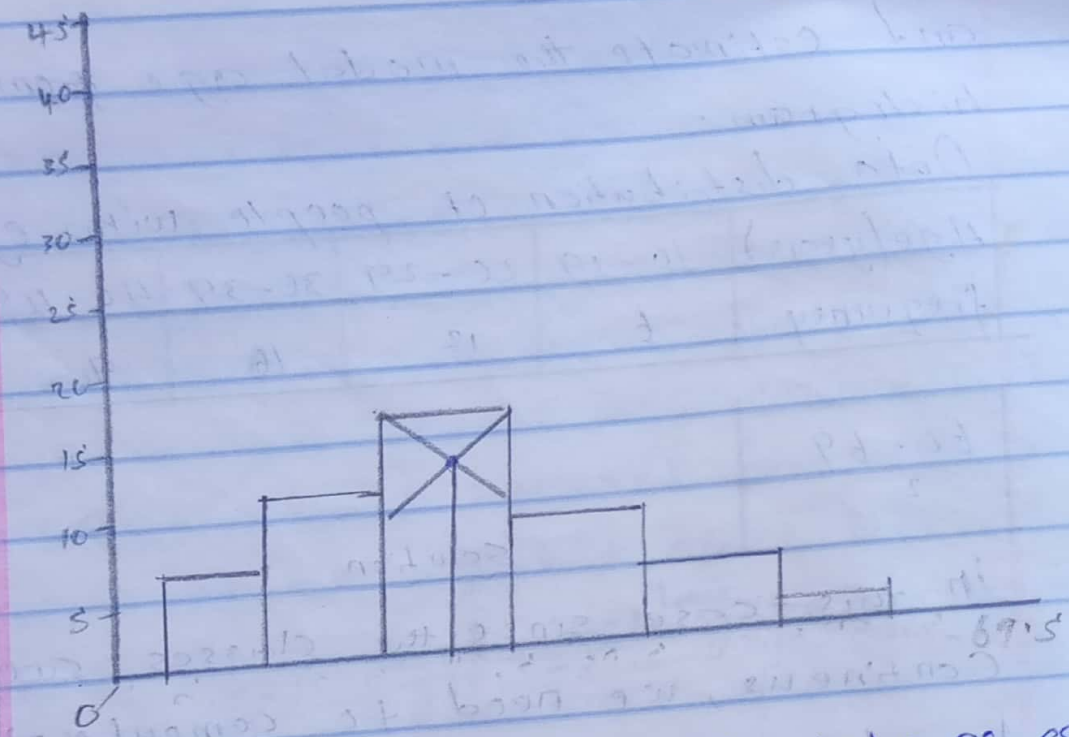
Data distribution of people with eye problem

Age(years)	10-19	20-29	30-39	40-49	50-59	60-69
frequency	6	12	10	9	5	2

Solution

in this case since the classes are not continuous, we need to computer the class boundries before drawing the histogram. the boundries are obtained as follows.

Age	f	class boundry
10-19	6	9.5-19.5
20-29	12	19.5-29.5
30-39	10	29.5-39.5
40-49	9	39.5-49.5
50-59	5	49.5-59.5
60-69	2	59.5-69.5



The modal age is 33.1 approximately as estimate from the histogram.

frequency polygon

A frequency polygon is a series of connected lines segment in which the point connection represent the values in the classes against their respective frequency for group data, the point of connection are the mid point of each class against their respective frequency. Hence, the frequency polygon is a diagram drawn

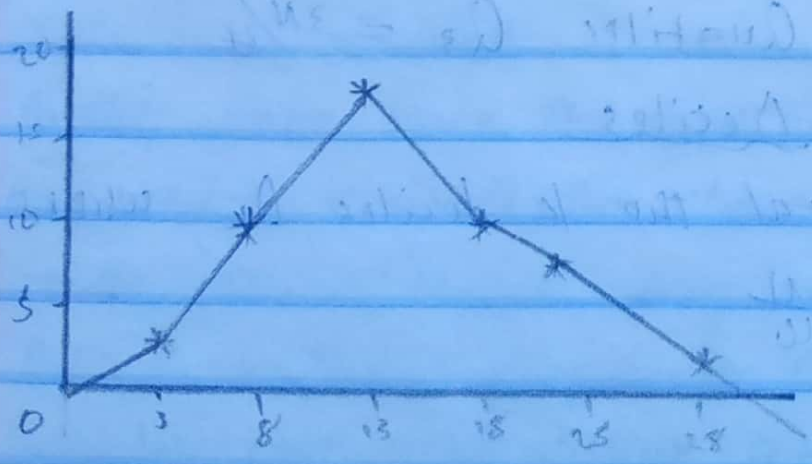
by joining of the mid point of the tops of histogram bars with straight lines.

Example

An Opthalmologist from the national eye centre observed the following data on the distribution of 50 people by age who suffered from eye problem In a particular village use the data to draw a frequency polygon.

Age	No of patient frequency	mid point
1—5	3	3
6–10	10	8
11–15	18	13
16–20	10	18
21–25	7	23
26–30	2	28

the frequency polygon is displayed below.



$P = \frac{N}{100}$

$P_K = \frac{KN}{100}$

Deciles $= \frac{N}{10}$

$D_8 = \frac{8N}{10}$

$Q_1 = \frac{N}{4}$ $\quad Q_3 = \frac{3N}{4}$

$Q_2 = \frac{2N}{4} = \frac{N}{2}$

Cummulative frequency Curve (ogive)

When frequency are added successively, they are called cummulative frequency the curve obtained by plotting the Cummulative frequency againts the upper class boundry is called the cummulative frequency curve (ogive) the curve can be used to estimate partition values such as Qutiles, Deciles, parcentiles. the position of the Quatiles, deciles and parcentiles are first located and subsequently estimated from the ogive those positions ban be located by the following formulie.

Quatiles

Lower Qutiles $Q_1 = N/4$

middle Quatiles $Q_2 = 2N/4 = N/2$

upper Quatiles $Q_3 = 3N/4$

Deciles

In general the k deciles D_k where

$D_k = \frac{kN}{10}$

percentiles

In general the k^{th} percentile is P_k where

$P_k = \frac{kN}{100}$

where N represents the total frequency

Example

the table below shows the distribution of marks obtained by 200 candidates in a practical biology examination. make a cummulative frequency table for the data and construct the cummulative frequency curve. Hence, use the curve to estimate the median marks.

marks(100%)	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
frequency	2	5	11	20	26	40	40	36	9	3

Solution

in this case since the classes are not continuous we need to compute the class boundry, as well as the cummulative frequency before drawing the cummulative frequency curve the class boundry and cumulative frequency are obtained

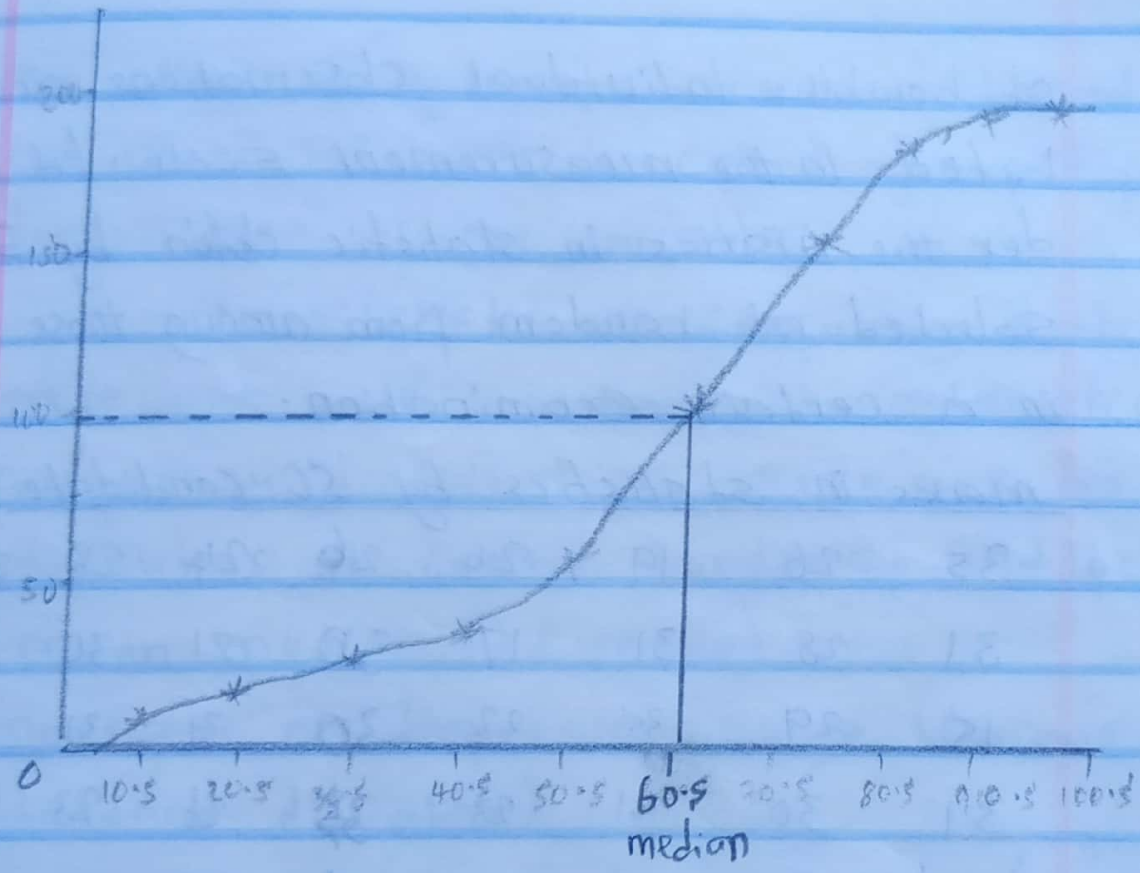
as follows

Mgodes	frequency	Class boundry	C.f
1-10	2	0.5 - 10.5	2
11-20	5	10.5 - 20.5	7
21-30	11	20.5 - 30.5	18
31-40	20	30.5 - 40.5	38
41-50	26	40.5 - 50.5	64
51-60	42	50.5 - 60.5	106
61-70	46	60.5 - 70.5	152
71-80	36	70.5 - 80.5	188
81-90	9	80.5 - 90.5	197
91-100	3	90.5 - 100.5	200

The median is located as follows

median $Q_2 = \frac{2N}{4} = \frac{2 \times 200}{4}$ $Q_2 = \underline{100}$

the median can now be estimated
the curve at the point where C.f is
the cummulative frequency curve is d
below :



frequency distribution

frequency distribution - is an organized tablation/graphical representation of the number of individuals in each categories on the scale of measurement. It allows the reseacher to have a glance at the entire data conveniently. It shows whether the observations are high or low and also whether they are concentrated in one area or spread out across the entire scale. thus, frequency distribution present a picture

of how the individual observations are distri-buted in the measurement scale. Let us consider the marks in statistic obtain by 50 candid selected at random from among those appear in a certain examination.

Marks in statistics by 50 candidate

35	26	19	24	26	24	28	24	22
31	28	31	17	30	21	30	31	19
15	29	35	22	30	31	31	30	28
31	30	21	23	35	26	23	31	35
17	35	18	31	37	27	17	24	15

marks (x)	tally	frequency
15 – 19	𝍸 | | | |	9
20 – 24	𝍸 𝍸 |	11
25 – 29	𝍸 𝍸	10
30 – 34	𝍸 𝍸 𝍸	15
35 – 39	𝍸	5
		50

Such a table showing the distribution of frequencys in the different class is called frequency table and the manner in which

the class frequencys are distributed over the class Intervals is called grouped frequency distribution of the variable.

Characteristic of frequency distribution

there are four Important characteristic of frequency distribution there are as follows

1. measures of central tendency and location (mean, median, and mode).
2. measures of dispersion (Range, Varians and standard deviation)
3. The extent of symmetry/asymmetry (skewness)
4. The flatness or peakedness (kurtosis)

measures of central tendancy

The measures of central tendancy are otherwise called measures of location, they are scores or numerical values in central part of the distribution they are values somewhere midway between the highest and the Lowest observation which are used to represent the entire Information. The following are the five measures of central tendency that are in common use -

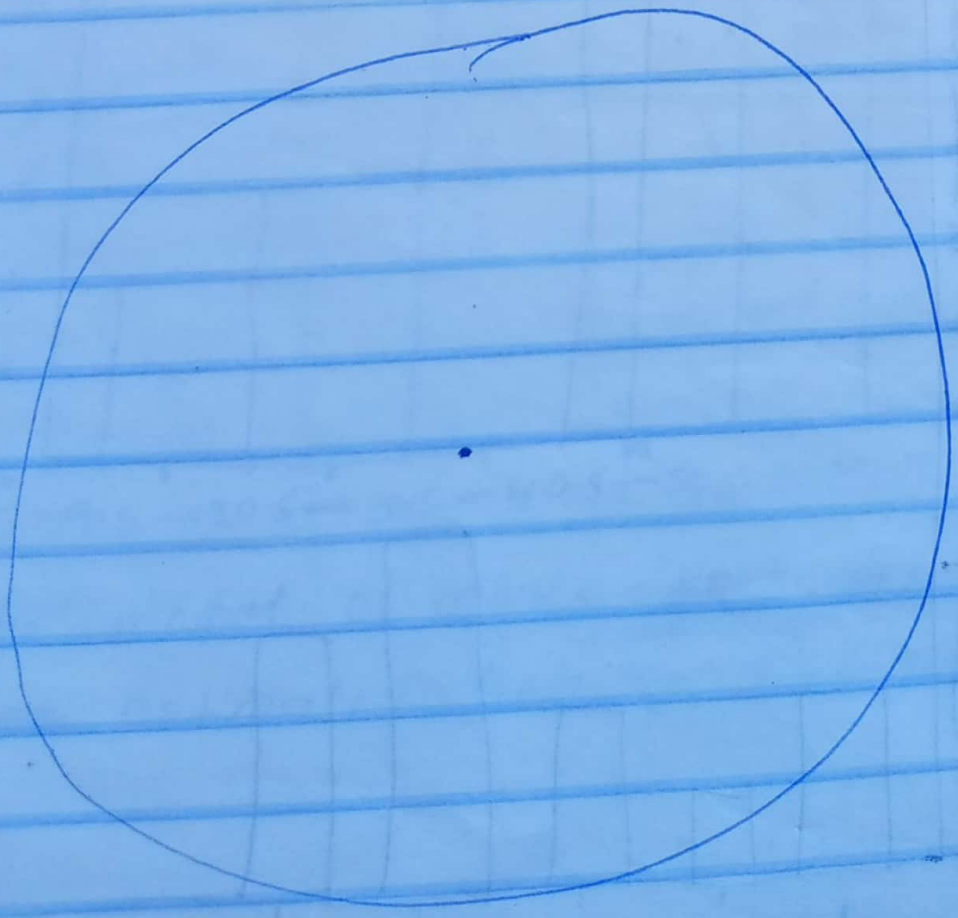
1. Arithmetic Mean or Simply mean.
2. median
3. mode
4. Geometric mean
5. Harmonic mean

1.	food Items	Amounts (₦)
	Rice	3,000
	meat	6,000
	oil	500
	Vegetables	500
	Beans	2,000
	yam	2,500
	fish	2000
	miscellanous	1,800

Soln

1.

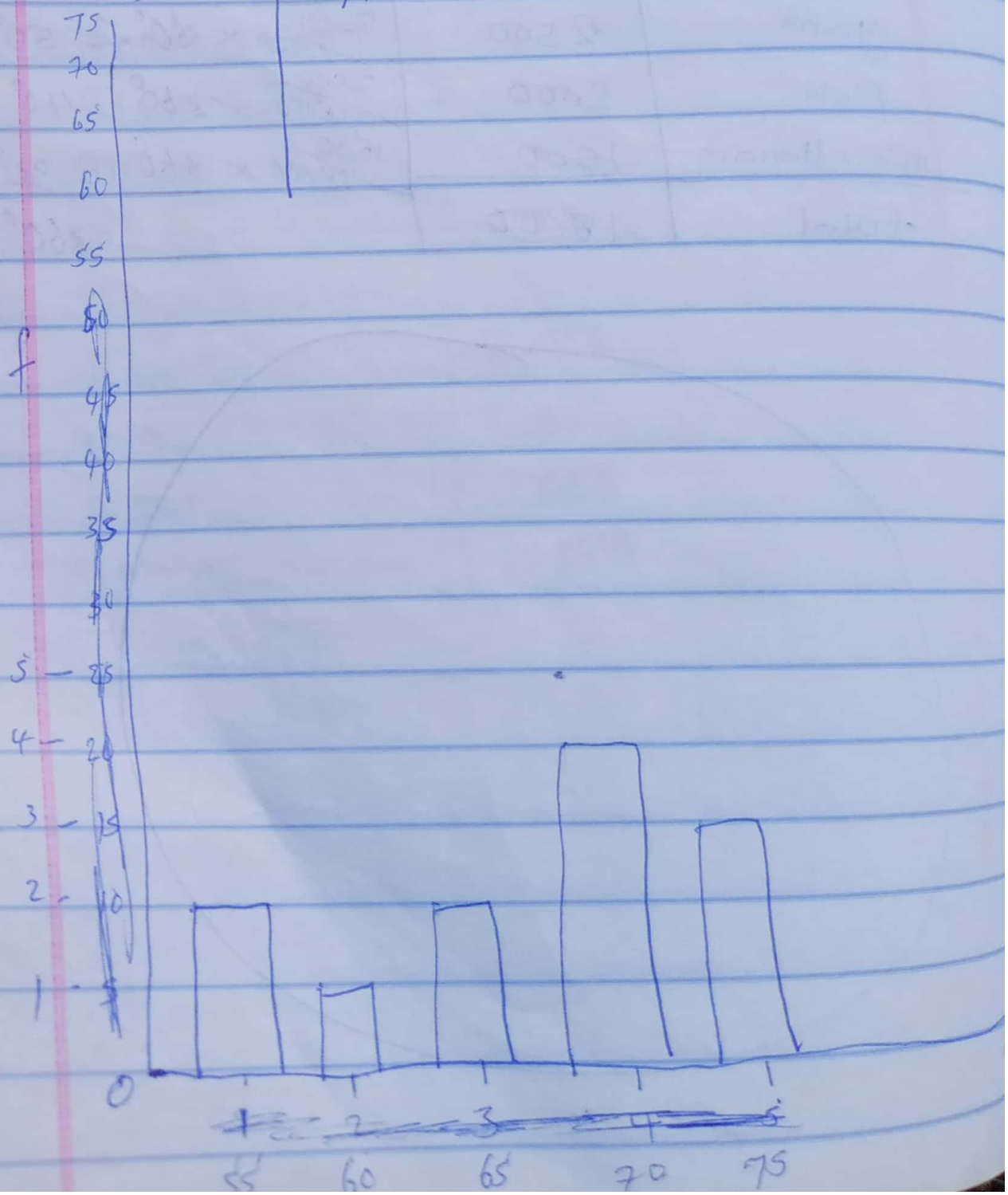
food items	Amounts	Angles°
Rice	3,000	$\frac{3000}{18,000} \times 360^\circ = 60^\circ$
meat	6,000	$\frac{6000}{18,000} \times 360^\circ = 120^\circ$
oil	500	$\frac{500}{18,000} \times 360^\circ = 10^\circ$
Vegetables	500	$\frac{500}{18,000} \times 360^\circ = 10^\circ$
Beans	2,000	$\frac{2,000}{18,000} \times 360^\circ = 40^\circ$
yam	2,500	$\frac{2,500}{18000} \times 360^\circ = 50^\circ$
fish	2,000	$\frac{2000}{18000} \times 360^\circ = 40^\circ$
miscellanous	1500	$\frac{1500}{18,000} \times 360^\circ = 30^\circ$
total	18,000	360°



pie

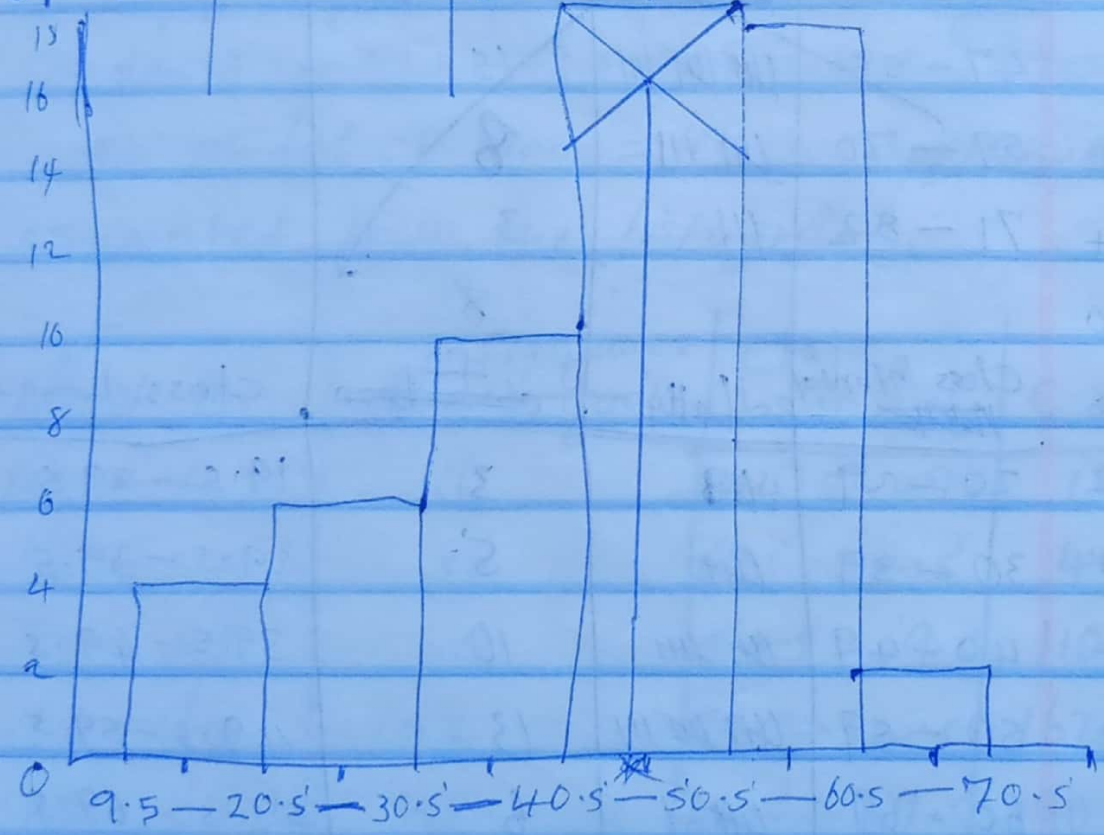
2.

frequency	kg
2	55
1	60
2	65
4	70
3	75



3.

marks	f	class boundry
10 - 20	4	9.5 – 20.5
20 – 30	6	20.5 – 30.5
30 - 40	10	30.5 – 40.5
40 - 50	20	40.5 – 50.5
50 - 60	18	50.5 – 60.5
60 - 70	2	60.5 – 70.5



the modal of marks = 45.5 approximately

~~from~~ estimated from the histogram

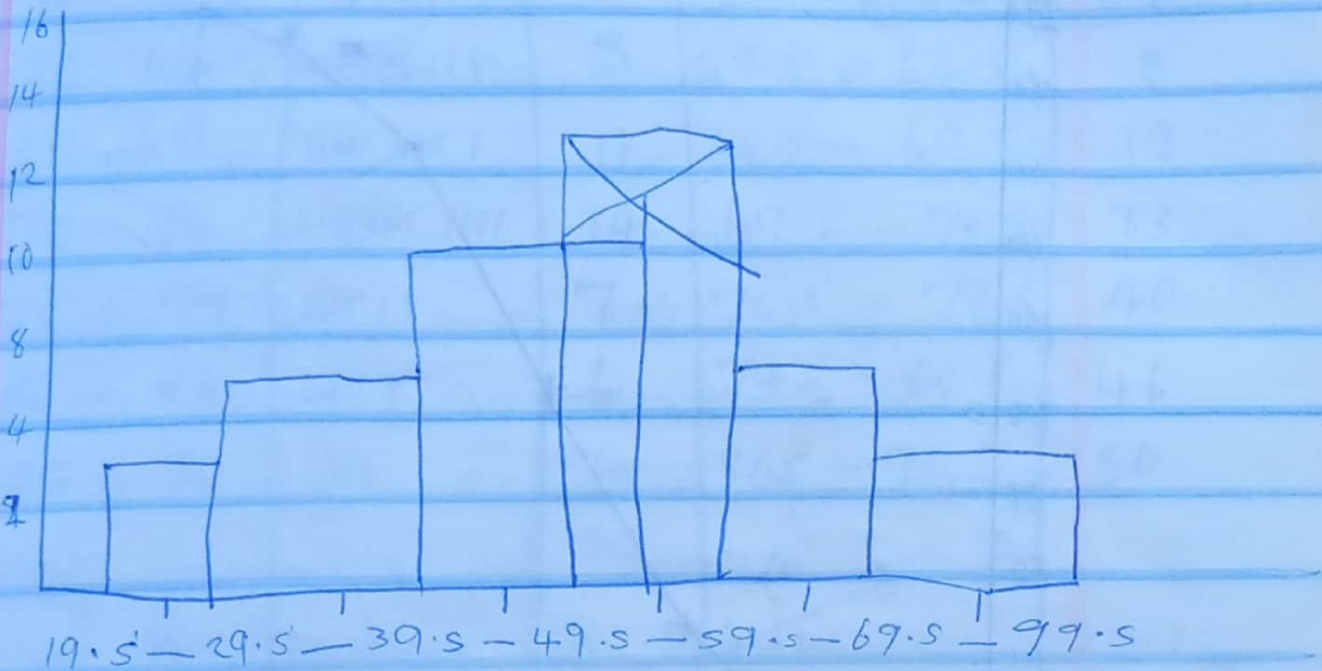
4. marks in statistic of 40 student

41	52	37	56	63	48	65	46	79	71
54	32	51	66	74	23	35	61	63	55
58	44	49	53	45	57	56	38	43	47
59	28	50	49	67	56	36	45	56	26

23
79
23 – 28
29 – 34
35 – 40
41 – 46
47 – 52
53 – 58
59 – 64
65 – 70
71 – 76
77 – 82

marks	tally	frequency
23 – 34	IIII	4
35 – 46	IIII IIII	10
47 – 58	IIII IIII IIII	15
59 – 70	IIII III	8
71 – 82	III	3
		8

Class Interval	tally	f	class boundry
20 – 29	III	3	19.5 – 29.5
30 – 39	IIII	5	29.5 – 39.5
40 – 49	IIII IIII	10	39.5 – 49.5
50 – 59	IIII IIII III	13	49.5 – 59.5
60 – 69	IIII I	6	59.5 – 69.5
70 – 80	IIII	3	69.5 – 79.5
		40	

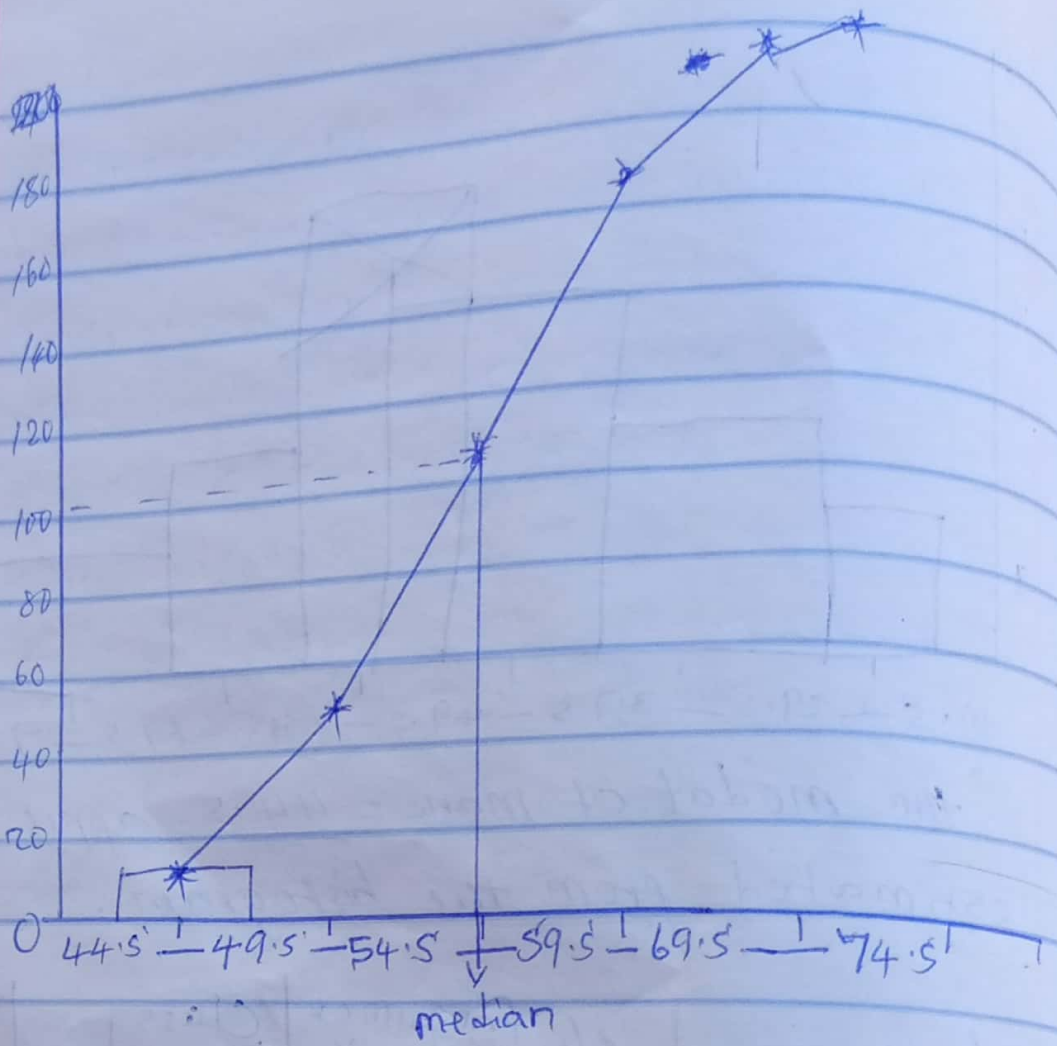


the modal of mark = 44.5 approximately estimated from the histogram.

S. No. of eggs	~~No. of girls~~ frequency	Class boundry	C.f
45 — 49	10	44.5 - 49.5	10
50 - 54	36	49.5' - 54.5'	46
55 - 59	64	54.5' - 59.5'	110
60 - 64	52	59.5' - 64.5	162
65 - 69	28	64.5' - 69.5	190
70 - 74	10	69.5' - 74.5	200

the median is located as follows.

$Q_{\frac{N}{2}} = \frac{2N}{4}$ $Q_2 = \frac{200 \times 2}{4}$ $Q_2 = \frac{400}{4}$ $Q_2 = 100$



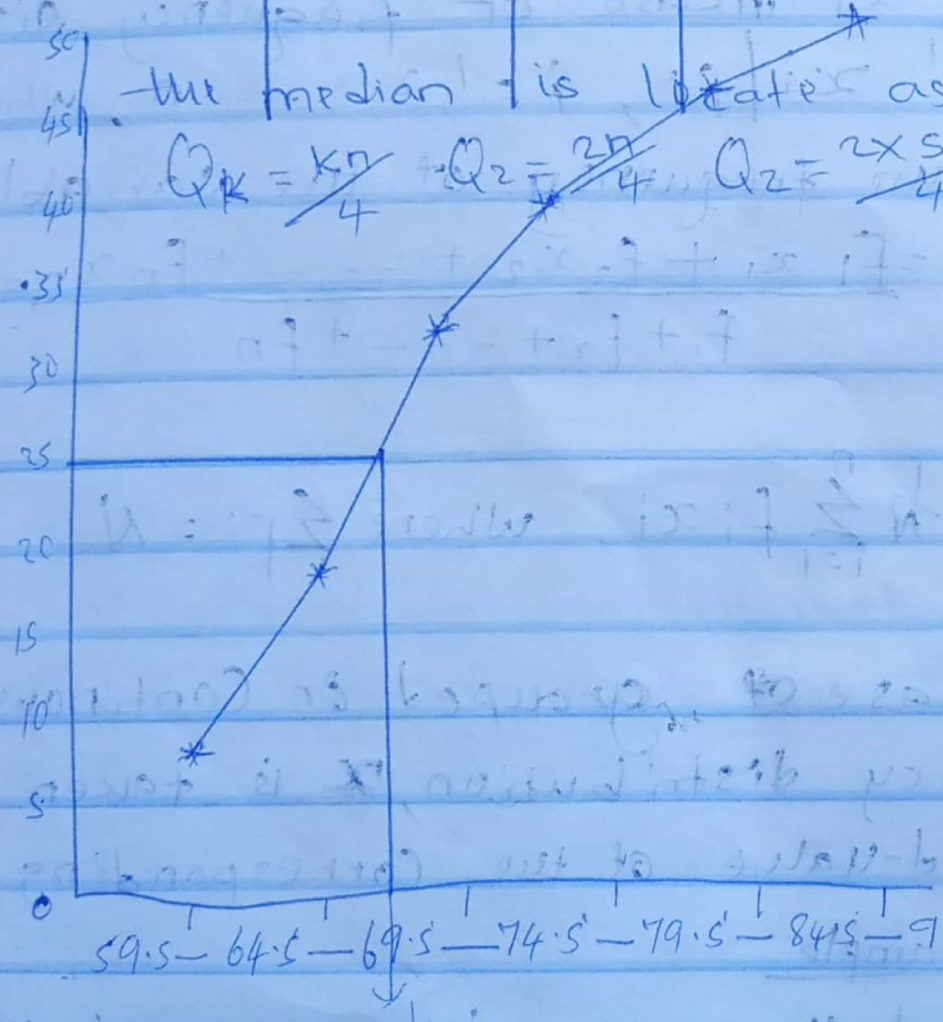
for 7th deciles $D_k = \frac{kN}{10}$

$D_7 = \frac{7 \times 200}{10}$ $D_7 = \frac{1400}{10}$ $D_7 = 140$

6. Weight of 50 pregnant women in Statistic

74	84	65	86	71	75	66	82	64
78	79	81	69	88	67	82	73	70
60	65	68	61	74	69	73	77	62
64	75	71	62	87	73	67	72	64
82	80	70	73	71	74	67	71	85

Class Interval	tally	f	class boundry	C.f
60-64	𝍸 III	8	59.5 - 64.5	8
65-69	𝍸 𝍸 I	11	64.5 - 69.5	19
70-74	𝍸 𝍸 IIII	14	69.5 - 74.5	33
75-79	𝍸 II	7	74.5 - 79.5	40
80-84	𝍸 I	6	79.5 - 84.5	46
85-90	IIII	4	845 - 90.5	50
		50		



the median is located as follows:

$Q_k = \frac{kn}{4}$ $Q_2 = \frac{2n}{4}$ $Q_2 = \frac{2 \times 50}{4}$ $Q_2 = 25$

for 6th deciles $D_k = \frac{kn}{10}$ $D_6 = \frac{6 \times 50}{10}$ $D_6 = \frac{300}{10}$

$D_6 = 30$

for 30th percentile $P_k = \frac{kn}{100}$ $P_{30} = \frac{30 \times 50}{100}$ $P_{30} = \frac{1500}{100}$

$P_{30} = 15$

Arithmetic mean

Arithmetic mean of set of observation is there sum divided by the number of observation.

example - the arithmetic mean $\bar{x}$ of n of observation $x_1, x_2 \dots x_n$ is given by $\bar{x} = \frac{1}{n}(x_1 + x_2 - - - - x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$ incase of frequency distribution $x_i | f_i$, $i = 1, 2, 3 \dots n$, where f_i is the frequency of the variable x_i

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + - - - - + f_n x_n}{f_1 + f_2 + - - - + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i x_i, \text{ where } \Sigma f_i = N$$

Incase of grouped or Continous frequency distribution, x is taken as the mid-value of the corresponding class

Example

find the arithmetic mean of the following frequency distribution.

x:	1	2	3	4	5	6	7
f:	5	9	12	17	14	10	6

solution

x	f	fx
1	5	5
2	9	18
3	12	36
4	17	68
5	14	70
6	10	60
7	6	42
	$\Sigma f = 73$	$\Sigma fx = 299$

$$\therefore \bar{x} = \frac{1}{N} \Sigma fx = \frac{299}{73} = 4.09$$

example 2

calculate the arithmetic mean of the marks from the following table:

marks:	0–10	10–20	20–30	30–40
No of stu:	12	18	27	20

40–50	50–60
17	6

marks	No. of student	mid point x	fx
0 – 10	12	5	60
10 – 20	18	15	270
20 – 30	27	25	675
30 – 40	20	35	700
40 – 50	17	45	765
50 – 60	6	55	330
	$\Sigma f = 100$		$\Sigma fx = 2,800$

Arithmetic mean

$$\bar{x} = \frac{1}{N} \Sigma fx = \frac{2,800}{100} = 28$$

$$\bar{x} = 2.8$$

Assume mean method

Assume mean – If the values of x or (and) f are large, the calculation of mean by the formula above is quite time-consuming and tedious. The arithmetic is reduced to a great extend by taking the diviations of the given values from any arbitraly point "A" as explained be

Let $d_i = x_i - A$, then $f_i d_i = f_i(x_i - A)$

$$f_i d_i = f x_i - A f_i$$

Summing both sides over i from 1 to n we get

$$\sum_{i=1}^{n} f_i d_i = \sum_{i=1}^{n} f_i x_i - A \sum_{i=1}^{n} f_i$$

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

$$\frac{\sum_{i=n}^{n} f_i d_i}{N} = \bar{x} - A$$

$$\therefore \bar{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N}$$

or $\boxed{\bar{x} = A + \frac{1}{N} \Sigma f_i d_i}$

Any number can solve the purpose of arbitraly point 'A' but usually the value of x corresponding to the middle part of the distribution will be much more convinient Incase of the grouped or continous

frequency distribution the arithmetic is reduced to a still greater extend by taken $d_i = \frac{x_i - A}{h}$

where A is an arbitrary point and h is the common magnitude or class interval in this case we have

$h d_i = x_i - A$ and proceeding as by as above, we get

$$\bar{x} = A + \frac{h}{N} \sum_{i=1}^{n} f_i d_i$$

Example

calculate the mean for the following frequency distribution

class Interval	0-8	8-16	16-24	24-32	32-40	40-48
frequency	8	7	16	24	15	7

Solution

Take $A = 28$

$h = 8$

$d = \frac{x-A}{h}$

Class Interval	mid point (x)	frequency (f)	$d=\frac{(x-A)}{h}$	fd
0 – 8	4	8	$\frac{4-28}{8}=-3$	−24
8 – 16	12	7	$\frac{12-28}{8}=-2$	−14
16 – 24	20	16	$\frac{20-28}{8}=-1$	−16
24 – 32	28	24	$\frac{28-28}{8}=0$	0
32 – 40	36	15	$\frac{36-28}{8}=1$	15
40 – 48	44	7	$\frac{44-28}{8}=2$	14
		$\Sigma f = 77$		$\Sigma fd = -25$

$$\bar{x} = A + \frac{h}{N}\sum_{i=0}^{n} f_i d_i$$

$A = 28 \quad N = \Sigma f = 77 \quad h = 8 \quad \Sigma fd = -25$

$$\bar{x} = 28 + \frac{8}{77} \times (-25)$$

$$\bar{x} = 28 + (-2.59)$$

$$\bar{x} = 28 - 2.59$$

$$\bar{x} = 25.41$$

Advantages of Arithmetic Mean

1. It is rigidly defined
2. It is easy to calculate and widely to understood.
3. It can be determined when only the total

value and the number of observation are available

4. It is liable to mathematical precision

Disadvantages

1. It is affected by extreme values
2. It may not correspond to any value in the data set
3. It requires assumption when open intervals values are involved
4. Arithmetic means it can not be used if we dealing with qualitative characteristics which can not be majored quantitatively Interligent, honesty e.t.c in such cases median is the only average to be used

Median =

L = lower class boundry of d median class
h = magnitude of the class Interval
f = frequency of the median class
N = total frequency = Σf
Cf = cummulative frequency of the class before the median

$N/2$ $\quad$ median $= L + \frac{h}{f}\left(\frac{N}{2} - cf\right)$ class

L = lower class boundry ~~other~~ of the median

Median – the median is the middle number in a given set ~~numb~~ data of a frequency distribution when the number are arranged in order of magnitude ascending or descending order. Thus for example, for an orderly arranged set $(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$, the median is x_4. On the other hand, the median in the orderly arranged set, $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ is $\frac{(x_4 + x_5)}{2}$ in case of distrid

frequency distribution median is obtained by Considering the cummulative frequency. The steps for calculating median are given below.

1. find $N/2$, where $N = \sum_{i=1}^{n} f_i$
2. See the cummulative frequency just greater $N/2$.
3. The corresponding value of x is the median

Example

Obtained the median for the fflwing frequency distribution:

x	1	2	3	4	5	6	7	8	9
f	8	10	11	16	20	25	15	9	6

x	f	CF
1	8	8
2	10	18
3	11	29
4	16	45
5	20	65
6	25	90
7	15	105
8	9	114
9	6	120
	$\Sigma f = 120$	

Hence $N = 120 \Rightarrow N/2 = \frac{120}{2} = 60$

Cummulative frequency just greater than $N/2$ is 65 and the value of x corresponding to 65 is 5. therefore median is = 5

in the case of continous f.d, the class corresponding to the C.f just greater than $N/2$ is called the median class and the value of median is obtaind by the fol formula:

$$\text{median} = L + \frac{h}{f}\left[\frac{N}{2} - c.f_b\right]$$

Where L = Is the lower class boundry of the median class.

f = is the frequency of the median class.

h = is the magnitude of the median class.

C.Fb = is the cummulative frequency before the median class.

$N = \Sigma f$

Example

find the median wage of following distribution.

Wages (#)	20-30	30-40	40-50	50-60
No of Labourers	3	5	20	10

60-70

5

solution

wages (#)	No of Lab	C.f	c-b
20-30	3	3	19.5 - 30.5
30-40	5	8	29.5 - 40.5
40-50	20	28	39.5 - 50.5
50-60	10	38	49.5 - 60.5
60-70	5	43	59.5 - 70.5

$\Sigma f = 43$

Hence $N/2 = 43/2 = 21.5$

$$m = L + \frac{h}{f}\left(\frac{N}{2} - C.f_b\right)$$

$L = 39.5$, $h = 10$, $f = 20$, $C.f_b = 8$, $N = 43$

$$m = 39.5 + \frac{10}{20}(21.5 - 8) =$$

$$m = 39.5 + 6.75$$

$$m = 46.25$$

Example 2 Exercise

Compute the median of the data below which depits the hourly wage (₦) of the 60 dunior workers in a pharmaceutical firm

Hourly wage (₦)	10 – 19	20 – 29	30 – 39	40 – 49
No of workers	4	6	10	20

50 – 59	60 – 69
18	2

Solution

H·W (#)	No of W	C·f	c·b
10-19	4	4	9·5 - 19·5
20-29	6	10	19·5 - 29·5
30-39	10	20	29·5 - 39·5
40-49	20	40	39·5 - 49·5
50-59	18	58	49·5 - 59·5
60-69	2	60	59·5 - 69·5

$\Sigma f = 60$

Hence $N/2 = 60/2 = 30$

$$m = L + \frac{h}{f}\left(\frac{N}{2} - C \cdot f_b\right)$$

Advantages of median.

1. It is not affected by extreme value.
2. Computation is very easy
3. Since it is an actual value, it affairs representable and realistic.

Dis Advantages

1. for group data, the median is only an estimate.
2. It does not take all value in to account in its computation
3. It is not useful or practicable in qualitative data

Mean Deviation $= \frac{\sum_{i=1}^{n} (x - \bar{x})}{N}$

Variance $\sigma w = \frac{\sum_{i=1}^{n} (x - \bar{x})^2}{N}$

Standard devi- $\sigma w = \sqrt{\text{Variance}} = \sqrt{\sum_{i=1}^{n} (x - \bar{x})^2}$

let $w = \{x_1, x_2, x_3, x_4, x_5 \; 2, 6, 5, 7, 5\}$

$\bar{x} = \sum_{i=1}^{n} \frac{\text{Soln}}{(w)} = \frac{2+6+5+7+5}{5} = \frac{25}{5} = 5$

Mean deviation $= \frac{(2-5)+(6-5)+(5-5)+(7-5)+(5+5)}{5}$

$= \frac{(-3)+(1)+(0)+(2)+(0)}{5} = \frac{0}{5} = 0$

Variance $\sigma_w^2 = \frac{(2-5)^2+(6-5)^2+(5-5)^2+(7-5)^2+(5+5)^2}{5}$

$= \frac{(-3)^2+(1)^2+(0)^2+(2)^2+(0)^2}{5} = \frac{9+1+4}{5}$

$= \frac{14}{5}$

Standard deviation $\sigma_W = \sqrt{14/5} = 0.7$