11) The growth (in centimetres) of two varieties of plant after 20 days is shown in this table. Construct a back-to-back stem and leaf plot for the data, and compare the distributions.

Variety 1	Variety 2
20 12 39 38 41 43 51 52 59 55	18 45 62 59 53 25 13 57 42
53 59 50 58 35 38 23 32 43 53	55 56 38 41 36 50 62 45 55

12) The math and reading achievement scores from the National Assessment of Educational Progress for selected states are listed below. Construct a back-toback stem and leaf plot with the data and compare the distributions.

Math	Reading
52 66 69 62 61 63 57 59 59	65 76 76 66 67 71 70 70 66
55 55 59 74 72 73 68 76 73	61 61 69 78 76 77 77 77 80

13) Draw a back to back stem and leaf to compare the reaction time of boys and girls to a certain stimuli

Maths	Reading
0.14, 0.19, 0.18, 0.09, 0.19, 0.23, 0.16	0.18, 0.24, 0.16, 0.22, 0.19, 0.19, 0.25,
0. 22, 0.15, 0.16, 0.20, 0.16, 0.16, 0.11	0.22, 0.21, 0.16 , 0.22, 0.18, 0.21, 0.22
0.15, 0.21, 0.23, 0.22, 0.23, 0.18	0.22, , 0.25, 0.17, 0.22, , 0.19, 0.19

3 NUMERICAL SUMMARIES

A numerical summary for a set of data is referred to as a statistic if the data set is a sample and a parameter if the data set is the entire population.

Numerical summaries are categorized as measures of location and measures of spread. Measures of location can further be classified into measures of central tendancy and measures of relative positioning (quantiles).

3.1 Measures of Location

Before discussing the measures of location, its important to consider summation notation and indexing

Index (subscript) Notation: Let the symbol x_i (read 'x sub t'*i*) denote any of the n values $x_1, x_2, ..., x_n$ assumed by a variable X. The letter i in x_i (i=1,2,. . . ,n) is called an index or subscript. The letters j, k, p, q or s can also be used.

Summation Notation: $\sum_{i=1}^{n} x_i = x_1 + x_2 + ... + x_n$

Example: $\sum_{i=1}^{n} X_i Y_i = X_1 Y_1 + X_2 Y_2 + ... + X_N Y_N$ **and**

$$\sum_{i=1}^{n} aX_i = aX_1 + aX_2 + ... + aX_N = a(X_1 + X_2 + ... + X_N) = a\sum_{i=1}^{n} X_i$$

3.1.1 Measures of Central Tendency (Averages)

A Measures of Central Tendency of a set of numbers is a value which best represents it. There are three different types of Central Tendencies namely the mean, median and mode. Each has advantages and disadvantages depending on the data and intended purpose.

Arithmetic Mean

The arithmetic mean of a set of values $x_1, x_2, ..., x_n$, denoted $\overline{x}$ if the data set is a sample, is found by dividing the sum of the set of numbers with the actual number of values. Ie

$$\overline{\mathrm{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + ... + x_n}{n}$$

Example 1 Find the mean of 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10.

Solution

Sum of values: 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = 55

Number of values = 10 Mean of values $\overline{x} = \frac{55}{10} = 5.5$

Note: If the numbers $x_1, x_2, ..., x_n$ occur $f_1, f_2, ..., f_n$ times respectively, (occur with frequencies $f_1, f_2, ..., f_n$), the arithmetic mean is, given by

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

where n is the total frequency. This is the formular for the mean of a grouped data.

Example 2 The grades of a student on six examinations were 84, 91, 72, 68, 91 and 72. Find the arithmetic mean.

The arithmetic mean $\overline{\mathrm{x}} = \frac{1}{n}\sum_{i=1}^{n} f_i x_i == \frac{1(84) + 2(91) + 2(72) + 1(68)}{1 + 2 + 2 + 1} = 79.67$

Example 3 If 5, 8, 6 and 2 occur with frequencies 3, 2, 4 and 1 respectively, the arithmetic mean is $\overline{\mathrm{x}} = \frac{1}{n}\sum_{i=1}^{n} f_i x_i = \frac{3(5) + 2(8) + 4(6) + 1(2)}{3 + 2 + 4 + 1} = 5.7$

Exercise

1 Find the mean of 9, 3, 4, 2, 1, 5, 8, 4, 7, 3

2 A sample of 5 executives received the following amount of bonus last year: sh 14,000, sh 15,000, sh 17,000, sh 16,000 and sh y. Find the value of y if the average bonus for the 5executives is sh 15,400

Properties of the Arithmetic Mean

(1) The algebraic sum of the deviations of a set of numbers from their arithmetic mean is zero, that is $\sum_{i=1}^{n}\left(x_i - \overline{\mathrm{x}}\right) = 0$.

(2) $\sum_{i=1}^{n}\left(x_i - a\right)^2$ is minimum if and only if $a = \overline{x}$.

(3) If n_1 numbers have mean $\overline{x}_1$, n_2 numbers have mean $\overline{x}_2$,…, n_k numbers have mean $\overline{x}_k$, then the mean of all the numbers called the combined mean is given by

$$\overline{\mathrm{x}}_c = \frac{n_1\overline{\mathrm{x}}_1 + n_2\overline{\mathrm{x}}_2 + ... + n_k\overline{\mathrm{x}}_k}{n_{11} + n_2 + ... + n_k} = \frac{\sum n_i \overline{\mathrm{x}}_i}{\sum n_i}$$

Solution

Boundaries	4.5-9.5	9.5-14.5	14.5-19.5	19.5-24.5	24.5-29.5	29.5-34.5	34.5-39.5
Mid pts (x)	7	12	17	22	27	32	37
Frequency	5	12	32	40	16	9	6
Xf	35	144	544	880	432	288	222
CF	5	17	49	89	105	114	120

Mean $\bar{x} = \frac{\sum fx}{n} = \frac{35+144+...+222}{120} = \frac{2545}{120} \approx 21.2083$

$n = 120$ thus Median $= 60.5^{th}$ Value $\Rightarrow$ Median class is 19.5-24.5 thus

$$\text{Median} = \text{LCB} + \left(\frac{\left(\frac{n+1}{2}\right) - Cf_a}{f}\right) \times i = 19.5 + \left(\frac{60.5-49}{40}\right) \times 5 = 20.9375$$

The modal class (class with the highest frequency) is 19.5-24.5 therefore

$$\text{Mode} = \text{LCB} + \left(\frac{f - f_a}{2f - f_a - f_b}\right) \times i = 19.5 + 5\left(\frac{40-32}{80-32-16}\right) = 20.75$$

Exercise

1. Find the mean median and mode for the following data: 9, 3, 4, 2, 9, 5, 8, 4, 7, 4
2. Find the mean median and mode of 1, 2, 2, 3, 4, 4, 5, 5, 5, 5, 7, 8, 8 and 9
3. The number of goals scored in 15 hockey matches is shown in the table.

No of goals	1	2	3	4	5
No of matches	2	1	5	3	4

Calculate the mean number of goals cored

4. The table shows the heights of 30 students in a class calculate an estimate of the mean mode and median height.

Height (cm)	140<x<144	144<x<148	148<x<152	152<x<156	156<x<160	160<x<164
No of students	4	5	8	7	5	1

5. Estimate the mean, median and mode for the following frequency distribution:

Class	1-4	5-8	9-12	13-16	17-20	21-24
frequency	10	14	20	16	12	8

Class	40-59	60-79	80-99	100-119	120-139	140-159	160-179	180-199
freq	5	12	32	40	16	9	6	

The Empirical Relation between the Mean, Median and Mode

$$MEAN - MODE = 3(MEAN - MEDIAN)$$

The above relation is true for unimodal frequency curves which are asymmetrical.

3.1.2 Other Types of Means

These will include weighted, harmonic and geometric means.

The Weighted Arithmetic Mean

The weighted arithmetic mean of a set of n numbers $x_1, x_2, ..., x_n$ having corresponding weights $w_1, w_2, ..., w_n$ is defined as

$$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_{11} + w_2 + \ldots + w_n} = \frac{\sum w_ix_i}{\sum w_i}$$

Example 1 Consider the following table with marks obtained by two students James (mark x) and John (mark y). The weights are to be used in determining who joins the engineering course whose requirement is a weighted mean of 58% on the four subjects below;

Subject	Maths	English	History	Physics	Total
Mark x	25	87	83	30	225
Mark y	70	45	35	75	225
Weight	3.6	2.3	1.5	2.6	10

Working the products of the marks and the weights we get

Subject	Maths	English	History	Physics	Total
Wx	90	200.1	124.5	78	492.6
Wy	252	103.5	52.5	195	603

Now $\bar{x}_w = \frac{\sum w_ix_i}{\sum w_i} = \frac{492.6}{10} = 49.26$ and $\bar{y}_w = \frac{\sum w_iy_i}{\sum w_i} = \frac{603}{10} = 60.3$

Clearly John qualifies but James does not.

Example 2 If a final examination is weighted 4 times as much as a quiz, a midterm examination 3 times as much as a quiz**,** and a student has a final examination grade of 80, a midterm examination grade
of 95 and quiz grades of 90, 65 and 70, the mean grade is

$$\bar{X} = \frac{1(90) + 1(65) + 1(70) + 3(95) + 4(80)}{1 + 1 + 1 + 3 + 4} = \frac{830}{10} = 83.$$

Question A tycoom has 3 house girls who he pays Ksh 4,000 each per month, 2 watch men who he pays Ksh 5,000 each and some garden men who receives Ksh 7,000 each. If he pays out an average of Ksh 5,700 per month to these people, find the number of garden men.

Question A student's grades in laboratory, lecture, and recitation parts of a computer course were 71, 78, and 89, respectively.

(a) If the weights accorded these grades are 2,4, and 5, respectively, what is an average grade?
(b) What is the average grade if equal weights are used?

The Geometric and Harmonic Means

Let $x_1, x_2, \ldots., x_n$ be the sample values, the geometric mean GM is given by

$$GM = \sqrt[n]{x_1 \times x_2 \times \ldots. \times x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

and the harmonic mean is given by

$$HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots. + \frac{1}{x_n}} = \frac{n}{\sum\left(\frac{1}{x_i}\right)}.$$

The Relation between the Arithmetic, Geometric and Harmonic Means:

$$HM \le GM \le \bar{x}.$$

The formulas for geometric and harmonic means of a frequency distribution are respectively given by;

$$GM = \sqrt[n]{x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_n^{f_n}} = \sqrt[n]{\prod_{i=1}^{n} x_i^{f_i}} \Rightarrow \log(GM) = \frac{1}{n}\sum_{i-1}^{n} f_i \log x_i$$

and

$$HM = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots + \frac{f_n}{x_n}} = \frac{n}{\sum\left(\frac{f_i}{x_i}\right)} = \left[\frac{1}{n}\sum\left(\frac{f_i}{x_i}\right)\right]^{-1}$$

where $n = \sum f_i$ and x_i are the midpoints

Example 1 Find the harmonic and the geometric mean of the numbers 2,4 and 8

Solution The geometric mean $GM = \sqrt[3]{2\times4\times8} = \sqrt[3]{64} = 4$ and

the harmonic mean $HM = \dfrac{3}{\frac{1}{2}+\frac{1}{4}+\frac{1}{8}} = \dfrac{3}{\frac{7}{8}} = \dfrac{34}{7} \approx 3.43$

Example 2 Find the harmonic and geometric mean of the frequency table below

x	13	14	15	16	17
f	2	5	13	7	3

Solution

The harmonic mean $HM = \dfrac{30}{\frac{2}{13}+\frac{5}{14}+\frac{13}{15}+\frac{7}{16}+\frac{3}{17}} \approx 15.$ and

The geometric mean $GM = \sqrt[30]{13^2 \times 14^5 \times 15^{13} \times 16^7 \times 17^3} \approx 15.09837$

Exercise:

1. Find the harmonic and the geometric mean of the numbers 10, 12, 15, 5 and 8
2. The number of goals scored in 15 hockey matches is shown in the table below . Calculate the harmonic and geometric mean number of goals scored.

No of goals	1	3	5	6	9
No of matches	2	1	5	3	4

3. Find the harmonic and geometric mean of the frequency table below

Class	0-29	30-49	50-79	80-99
Frequency	20	30	40	10

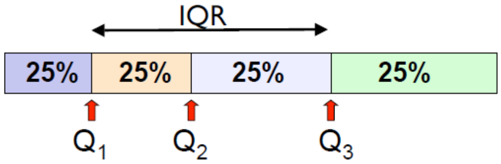
3.1.3 Measures of Relative Positioning (Quantiles)

These are values which divide a sorted data set into N equal parts. They are also known as quantiles or N-tiles. The commonly used quantiles are; **Quartiles, Deciles and Percentiles** These 3 divides a sorted data set into four, ten and hundred divisions, respectively. These measures of position are useful for comparing scores within one set of data. You probably all took some type of college placement exam at some point. If your composite math score was say 28, it might have been reported that this score was in the 94th percentile. What does this mean? This does not mean you received a 94% on the test. It does mean that of all the students who took that exam, 94% of them scored lower than you did (and 6% higher).

Remark For a set of data you can divide the data into three quartiles (Q_1, Q_2, Q_3), nine deciles ($D_1, D_2, \ldots D_9$) and 99 percentiles ($P_1, P_2, \ldots, P_{99}$). To work with percentiles, deciles and quartiles - you need to learn to do two different tasks. First you should learn how to find the percentile that corresponds to a particular score and then how to find the score in a set of data that corresponds to a given percentile.

Quartiles

They divide a sorted data set into 4 equal parts and we have lower, middle and upper quartiles denoted Q_1, Q_2 and Q_3 respectively. The lower quartile Q_1 separates the bottom 25% from the top 75%, Q_2 is the median and Q_3 separates the top 25% from the bottom 75% as illustrated below .



The Kth quartile is given by: $Q_k = \frac{k}{4}(n+1)^{th}$ value where k=1,2,3

Deciles and Percentiles

Similarly the Kth Deciles D_k and the Kth Percentiles P_k are respectively given by;

$$D_k = \tfrac{k}{10}(n+1)^{th} \text{ Value and } P_k = \tfrac{k}{100}(n+1)^{th} \text{ value}$$

NB For ungrouped data we may be forced to use linear interpolation for us to get the required Kth quantile. However for grouped data the Kth Value is given by

$$K^{th}Value = \text{LCB} + \left(\frac{K - Cf_a}{f}\right) \times i$$

where LCB, i and f are the lower class boundary. class interval and frequency of the class containing the K^{th} value. Cf_a is the cumuilative frequency of the class above this particular class

Example 1 Find the lower and upper quartiles, the 7th decile and the 85th percentile of the following data. 3, 6, 9, 10, 7, 12, 13, 15, 6, 5, 13

Solution

Sorted data: 3, 5, 6, 6, 7, 9, 10, 12, 13, 13, 15 Here n=11

$Q_1 = \frac{1}{4}(11+1)^{th} = 3^{rd}\, value = 6$ Similarly $Q_3 = \frac{3}{4}(11+1)^{th} = 9^{th}\, value = 13$

$$D_7 = \tfrac{7}{10}(11+1)^{th} = 7.7^{th}\, value = \underbrace{7^{th}\, value + 0.7(8^{th}\, value - 7^{th}\, value)}_{\text{linear interpolation}} = 10 + 0.7(12-10) = 11.4$$

$$P_{85} = \tfrac{85}{100}(11+1)^{th} = 10.2^{th}\, value = \underbrace{10^{th}\, value + 0.2(11^{th}\, value - 10^{th}\, value)}_{\text{linear interpolation}} = 13 + 0.2(15-13) = 13.4$$

Example 2

Estimate the lower quartile, 4th decile and the 72nd percentile for the frequency table below

Class	1-4	5-8	9-12	13-16	17-20	21-24
frequency	10	14	20	16	12	8

Solution

Boundaries	0.5-4.5	4.5-8.5	8.5-12.5	12.5-16.5	16.5-20.5	20.5-24.5
C.F	10	24	44	60	72	80

For this data n=80

$$Q_1 = \tfrac{1}{4}(80+1)^{th} = 20.25^{th}\ value = 4.5 + \left(\frac{20.25-10}{14}\right)\times 4 \approx 7.428571$$

$$D_4 = \tfrac{4}{10}(80+1)^{th} = 32.4^{th}\ value = 8.5 + \left(\frac{32.4-24}{20}\right)\times 4 \approx 10.18$$

$$P_{72} = \tfrac{72}{100}(80+1)^{th} = 58.32^{th}\ value = 12.5 + \left(\frac{58.32-44}{16}\right)\times 4 \approx 16.08$$

Exercise

a) Find the lower and upper quartiles, the 7th decile and the 85th percentile of the data.

a) 9, 3, 4, 2, 9, 5, 8, 4, 7, 4 b) 1, 2, 2, 3, 4, 4, 5, 5, 5, 5, 7, 8, 8 and 9

2) The number of goals scored in 15 hockey matches is shown in the table.

No of goals	1	2	3	4	5
No of matches	2	1	5	3	4

Estimate the lower quartile, 4th decile and the 72nd percentile of the number of goals cored

4) The table shows the heights of 30 students in a class calculate an estimate of the upper and lower quartile of the height.

Height (cm)	140<x<144	144<x<148	148<x<152	152<x<156	156<x<160	160<x<164
No of students	4	5	8	7	5	1

5) The grouped frequency table gives information about the distance each of 150 people travel to work.

Height (cm)	0<d<5	5<d<10	10<d<15	15<d<20	20<d<25	25<d<30
No of students	4	5	8	7	5	1

a) Work out what percentage of the 150 people travel more than 20 km to work

b) Calculate an estimate for the median distance travelled to work by the people?

Properties of measures of Location

(i) They are affected by change of origin. Adding or subtracting a constant from each and every observation in a data set causes all the measures of location to shift by the same magnitude. That is $\text{New measure} = \text{old measure} \pm \text{k}$

(ii) They are affected by change of scale. Multiplying each and every observation in a data set by a constant value scales up all the measures of location by the same magnitude.. That is $\text{New measure} = K(\text{old measure})$

Example: Consider the three sets of data A, B and C below

Set A: 65, 53, 42, 52, 53 $\bar{x}_A = 53$ and $\text{Median}_A = 53$

Set B: 15, 3, -8, 2, 3 $\bar{x}_B = 3$ and $\text{Median}_B = 3$

Set C: 45, 9, -24, 6, 9 $\bar{x}_C = 9$ and $\text{Median}_C = 9$

- Notice that set B is obtained by subtracting 50 from each and every observation in set A and clearly $\bar{x}_B = \bar{x}_A - 50$ and $\text{Median}_B = \text{Median}_A - 50$ Therefore $\text{New measure} = \text{old measure} \pm \text{k}$. This is referred to as change of origin.
- Effectively set C is obtained by multiplying each and every observation in set B by 3 and clearly $\bar{x}_C = 3\bar{x}_B$ and $\text{Median}_C = 3\text{Median}_B$ Thus $\text{New measure} = K(\text{old measure})$ This is referred to as change of scale.