S5 - MATHEMATICS

Topic 1: Descriptive Statistics

Lesson 1: Representation of Data

Competences

In this lesson, you should be able to learn how to:

(i) form a frequency table from a set of raw data.
(ii) draw a histogram for equal and unequal classes and use the histogram to determine the mode.
(iii) draw a frequency polygon on its axes and on a histogram.
(iv) draw an Ogive and use it to determine the median.

Introduction

Statistics is a branch of science that deals with the collection, interpretation, presentation and analysis of data. When numerical information is collected, it's called raw data before it is arranged.

After collection, data is arranged in a table which shows the different scores and their number of occurrences (frequencies). This is called a **frequency table**. Frequency tables are of two types; **Ungrouped** and **Grouped** frequency table.

In O-level you saw how frequency tables are generated. The ungrouped frequency table is formed when the spread of data is small. For example, if a class is given a test marked out of 10 marks, each of the scores 0 to 10 is taken individually in the frequency table.

Ungrouped frequency table

The table shows the marks of students in a class in a certain test.

Mark	Freq.
0	3
1	5
2	4
3	4
4	6
5	5
6	3
7	4
8	3
9	2
10	1

3 students got 0 mark, 5 students got 1 mark, 4 students got 2 marks etc.

While a grouped frequency table is formed when the spread of data is large i.e. scores ranging from 1 to 100. If a test is marked out of 100, the marks will be grouped in classes of our convenience and each mark counted in its respective class.

A Grouped frequency table

The marks of students in a class with many students can be grouped for convenience as shown in the table.

Class	Freq.
10–19	4
20–29	3
30–39	7
40–49	10
50–59	6
70–79	5
80–89	3
90–99	2

When a grouped frequency table is made, some other concepts are generated; these include:

(i) **Class limits:** are the end marks in a given class. They are of two types: The lower-class limit and the upper-class limit. For example, in the class 20–29, 20 is the lower-class limit and 29 is the upper-class limit.

(ii) **Mid mark (x):** is the mark taken to represent a given class. It is got by adding the class limits and divide by 2 (i.e.) in the class 30–34, $x = \frac{30+34}{2} = 32$

(iii) **Class interval:** is the size of a given class. It is denoted by letter i and got by subtracting the lower-class limit from the upper-class limit and adding 1 i.e. in the class 80–89, i = (89 – 80) + 1 = 9 + 1 = 10.

(iv) **Class boundaries:** are the real class limits of a given class. They are got from the class limits by subtracting 0.5 from the lower-class limit and adding 0.5 to the upper-class limit to get the lower-class boundary and upper-class boundary respectively. That is, in the class 65–69, the class boundaries are (65 – 0.5, and 69 + 0.5) = 64.5 – 69.5.

A grouped frequency table

MARKS	FREQUENCY	MID MARK (X)	CLASS BOUNDARIES
30–34	11	32	29.5–34.5
35–39	17	37	34.5–39.5
40–44	13	41	39.5–44.5
45–49	24	47	44.5–49.5
50–54	22	52	49.5–54.5
55–59	15	57	54.5–59.5
60–64	8	62	59.5–64.5
65–69	4	67	64.5–69.5

A frequency table can be represented by class boundaries instead of class limits.

Marks	20 - < 30	30 - < 40	40 - < 50	50 - < 60	60 - < 70
Frequency	5	8	15	12	6

Marks	20 - 30	30 - 40	40 - 50	50 - 60	60 – 70
Frequency	5	8	15	12	6

Weight	< 20	< 30	< 40	< 45	< 55
Frequency	11	15	8	9	10

After representing data in a tabular form, it can further be represented graphically in different forms. Among them, we have bar graph, histogram, ogive, frequency polygon, pie chart etc.

Histogram

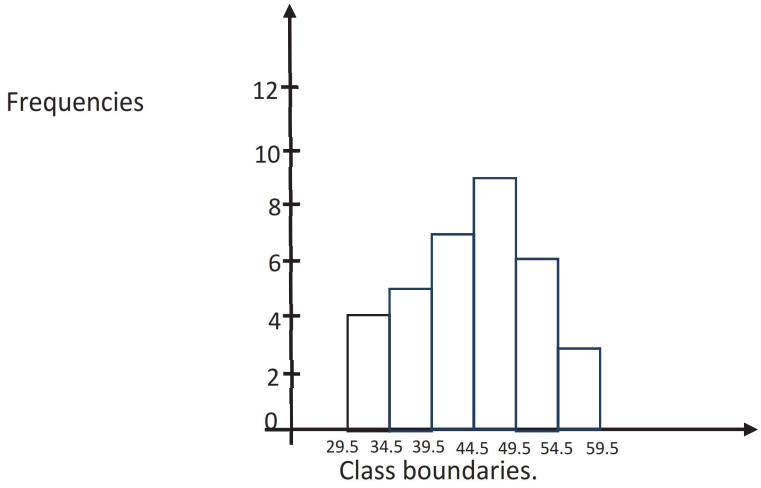
Histogram is a graph where class frequencies are plotted against class boundaries. Histograms are of two types. One of equal class interval and another of unequal class interval.

Histogram with Equal Class Interval

Example

Draw a histogram for the data below.

Class	Freq.
30–34	4
35–39	5
40–44	7
45–49	9
50–54	6
55–59	3



Histogram with Unequal Class Width

Here histogram is drawn with:

a) frequency density on the vertical axis.
b) class boundaries on the horizontal axis.

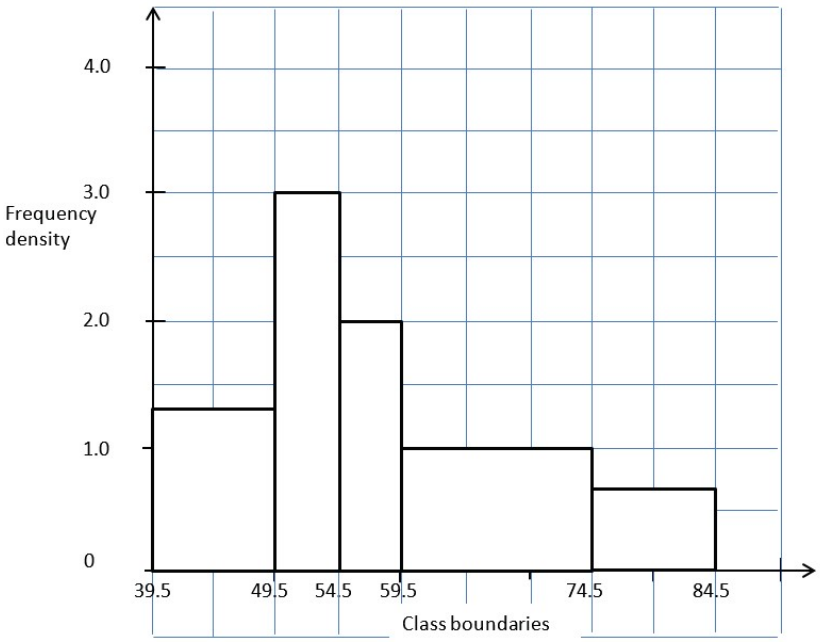
Frequency density is obtained by dividing frequency of that class with its class width.

Example

Draw a histogram for the frequency table below.

Class	Frequency	Class width, i	Frequency density $(\frac{f}{i})$	Class boundaries
40–49	13	10	1.3	39.5–49.5
50–54	15	5	3.0	49.5–54.5

Class	Frequency	Class width, i	Frequency density ($\frac{f}{i}$)	Class boundaries
55–59	10	5	2.0	54.5–59.5
60–74	15	15	1.0	59.5–74.5
75–84	7	10	0.7	74.5–84.5



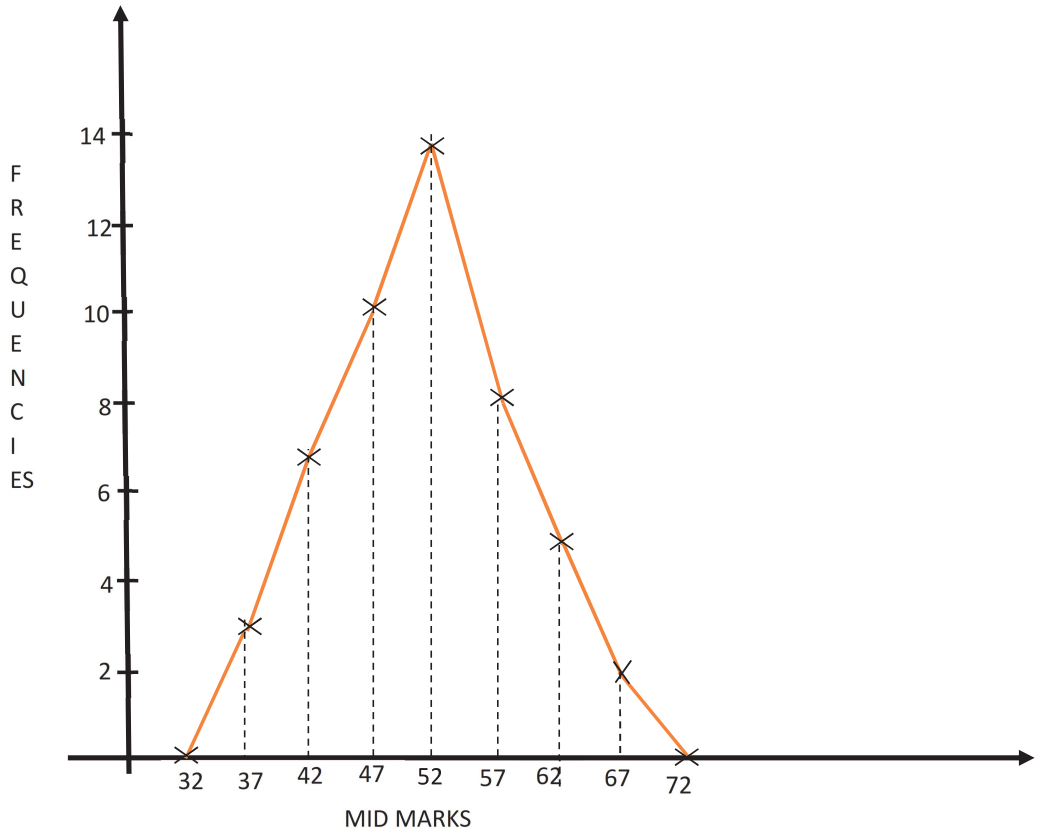
Frequency Polygon

Frequency polygon is drawn with frequencies on the vertical axis and mid marks on the horizontal. It can also be put on the same graph with the histogram.

Example

Draw a frequency polygon for the data below.

Class	**Frequency**	**Mid mark**
35–39	3	37
40–44	7	42
45–49	10	47
50–54	14	52
55–59	8	57
60–64	5	62
65–69	2	67



Cumulative Frequency Curve (Ogive)

The cumulative frequency curve is a curve drawn with cumulative frequency on the vertical axis and class boundaries on the horizontal axis. We plot upper class boundaries and the cumulative frequencies and join those points with a smooth curve.

Example 3

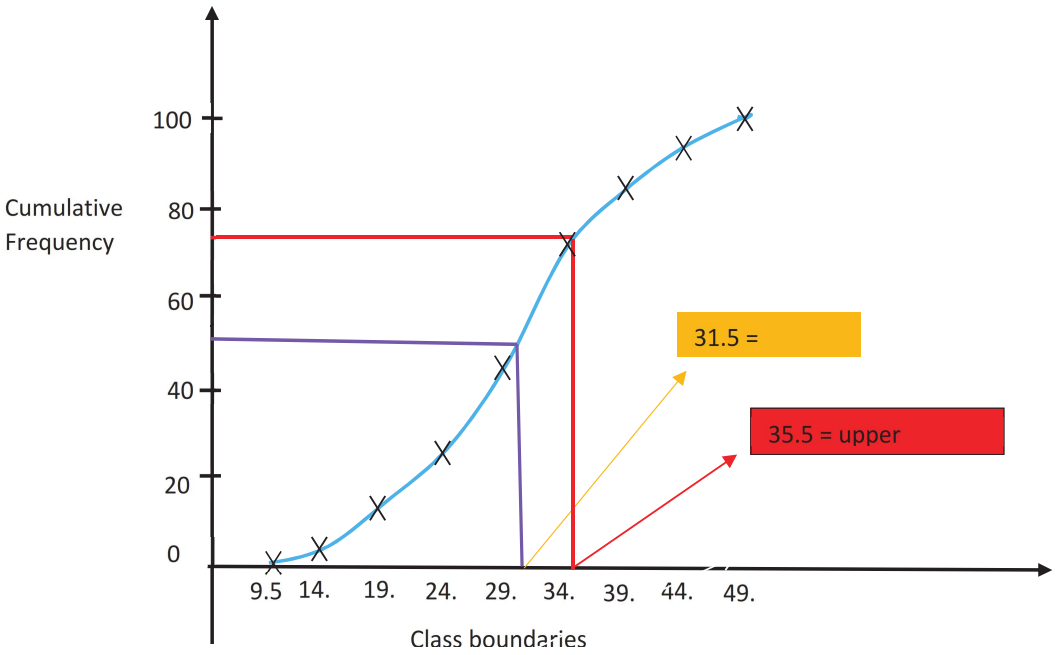
The frequency distribution table shows the weights of 100 children measured to the nearest kilogram.

Weight	Frequency
10–14	5
15–19	9
20–14	12
25–29	18
30–34	25
35–39	15
40–44	10
45–49	6

Draw a cumulative frequency curve for the data.

Solution

Class Boundaries	Frequency	Cumulative Frequency
9.5	0	0
14.5	5	5
19.5	9	14
24.5	12	26
29.5	18	44
34.5	25	69
39.5	15	84
44.5	10	94
49.5	6	100



Exercise

1. The lengths in cm of 40 metal rods were as follows:

Lengths	Frequency
30 – < 35	8
35 – < 40	5
40 – < 55	12
55 – < 60	9
60 – < 65	6

(a) Calculate the:
  (i) mean length.
  (ii) upper quartile.
(b) Display the data on a histogram and use it to estimate the mode.

2. The times (minutes) taken by the taxi to move from Kampala to Gomba were recorded over a certain period of time and grouped as follows:

Time (minutes)	Frequency (f)
80–84	10
85–89	15
90–94	35
95–99	40
100–104	28
105–109	15
110–114	4
115–119	2
120–124	1

(a) Calculate the mean time of travel from Kampala to Gomba by taxi.
(b) Draw a cumulative frequency curve for the data. Use it to estimate the:
  (i) median time for the journey.
  (ii) semi interquartile range of time of travel from Kampala to Gomba.

Lesson 2: Measures of Central Tendency and Measures of Dispersion

Competences

In this lesson, you will learn how to:

(i) calculate the mean, mode and median.
(ii) calculate the mean deviation, variance and standard deviation.

In O-Level, you saw the measures of central tendency by calculating the mean, mode and median. You used the formulae:

Mean $= \frac{\sum f(x)}{\sum f}$, or using assumed mean. Mean $= X_A + \frac{\sum f(d)}{\sum f}$

Mode $= L_o + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)C$ and Median $= L_o + \left(\frac{\frac{N}{2} - F_b}{f_m}\right)C$.

Measures of Dispersion

These are measures used to find out how the observations are spread out from the average. These include: mean deviation, quartile range, variance and standard deviation.

(a) Mean deviation

The mean deviation of a set of numbers is given by:

$\sum_1^n \frac{|x_i - M|}{n}$ where M is the mean

Find the mean deviation of the set of numbers 24, 43, 38, 28, 36, 40, 26, 37.

$M = \frac{\sum x}{n} = \frac{24+43+38+28+36+40+26+37}{8} = 34$

$\text{Mean deviation} = \frac{|24-34|+|43-34|+|38-34|+|28-34|+|36-34|+|40-34|+|26-34|+|37-34|}{8}$

$= \frac{10+9+4+6+2+6+8+3}{8} = \frac{48}{8}$

$= 6.$

(b) Variance

Variance is the sum of the mean deviations squared divided by the number of observations.

$\text{Variance} = \frac{\sum_1^n (x_i - M)^2}{n}$.

If the observations appear with the frequencies (repeated values), then variance is given by:

$\text{Variance} = \frac{\sum_1^n f(x_i - M)^2}{n}$.

A simplified form of the formula for variance used for computations is:

$\text{Variance} = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$ or $\text{Variance} = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$

Standard deviation is the positive square root of the variance.

i.e. standard deviation $= \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$.

Example

The frequency distribution table shows the marks scored by students in a certain school.

x	55	63	65	66	70	72	75	80	90
f	2	2	3	1	2	2	4	3	1

Calculate the standard deviation of the data.

x	**f**	**x^2**	**fx**	**fx^2**
55	2	3025	110	6050
63	2	3969	126	7938
65	3	4225	195	12675
66	1	4356	66	4356
70	2	4900	140	9800

72	2	5184	144	10368
75	4	5625	300	22500
80	3	6400	240	19200
90	1	8100	90	8100
	$\sum f = 20$		$\sum fx = 1411$	$\sum fx^2 = 100987$

Variance $= \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$

$= \frac{100987}{20} - \frac{1411}{20}$

$= 72.05$

Standard deviation $= \sqrt{72.05}$

$=$ **8.488**

Exercise

1. The table below shows the distribution of marks of students in a test.

Score	Frequency
$20 \leq x < 30$	4
$x < 45$	3
$x < 50$	9
$x < 65$	21
$x < 75$	3
$x < 80$	5
$x < 100$	14

(a) Draw a histogram and use it to estimate the modal mark.

(b) Calculate the: (i) mean score.
(ii) standard deviation score.

2. The heights of a group of workers in a factory were recorded as shown in the frequency table below.

Height (cm)	Frequency
170 – 175	19
175 – 180	36
180 – 185	70
185 – 190	64
190 – 195	39
195 - 200	22

(a) Estimate the mean and standard deviation of the worker's height.
(b) Plot an ogive.
(c) Use the ogive to estimate the median and interquartile range for the data.