A-level math paper 2: Statistics

This a branch of mathematics dealing with collection, presentation, analysis and interpretation of data

Types of data

(a) Discrete data
Its information collected by counting and usually takes integral values that do not lie within a given range

(b) Continuous data
It is information that takes values within a given range

Discrete or ungrouped data

Measures of central tendency

These are values of the distribution that tend to locate the central values. They include mean, median and mode

(i) Mean or average of a sample

It is denoted by $\bar{x}$ and defined as $\bar{x} = \frac{\sum x}{n}$; where x is the variable given and n is the number of variable

If assumed mean (working mean) a is given then

$\bar{x} = A + \frac{\sum d}{n}$; where d = x – A

If the frequency, f, s given then $\bar{x} = \frac{\sum fx}{\sum f}$ or $\bar{x} = A + \frac{\sum fd}{\sum f}$

(ii) Mode

This is the value of the distribution that appears most

(iii) Median

This is the middle value of the distribution obtained after the values have been arranged either in ascending of descending order.

Median = $\left(\frac{N}{2}\right)^{th}$ value.

Example 1

Given the following sets of values

2, 1, 3, 4, 5, 6, 7, 8, 9, 10, 3, 4, 6, 8, 9, 6, 3, 2

(a) Form a frequency table of ungrouped data

x	f	fx	cf
1	1	1	1
2	2	4	3
3	3	9	6
4	2	8	8
5	1	5	9
6	4	24	13
7	2	14	15
8	2	16	17
9	2	18	19
10	1	10	20
	$\sum f = 20$	$\sum fx = 109$	

(b) Use the table to find the mean and mode

Mean, $\bar{x} = \frac{\sum fx}{\sum f} = \frac{109}{20} = 5.45$

Mode = 6 (the value that appear most

(c) Find the median value

Median = $\left(\frac{N}{2}\right)^{th}$ value = $\left(\frac{20}{2}\right)^{th}$ = 10th value from cumulative frequency, cf ; median = 6

Example 2

Given the information below

x	10	11	12	13	14	15	16	17	18
f	4	2	6	3	7	2	1	2	2

Find

(a) Mean
(b) modal value
(c) median

Solution

x	f	fx	Cf
10	4	40	4
11	2	22	6
12	6	72	12
13	3	39	15
14	7	98	22
15	2	30	24
16	1	16	25
17	2	34	27
18	2	36	29
	$\sum f = 29$	$\sum fx = 387$	

(a) Mean, $\bar{x} = \frac{\sum fx}{\sum f} = \frac{387}{29} = 13.34$

(b) Modal value = 14 (appear most)

(c) Median = $\left(\frac{N}{2}\right)^{th}$ value = $\left(\frac{29}{2}\right)^{th}$ = 14.5th value from cumulative frequency, cf ; ***median = 13***

Example 3

The following are marks obtained by 30 students in a mathematic test marked out of 10

Mark (x)	0	1	2	3	4	5	6	7	8	9	10
Frequency (f)	2	2	3	2	3	4	5	2	1	3	3

(a) Find the mean using a working mean of 5
(b) Find the modal mark and he median

Solution

x	f	D= (x-5)	fd	fc
0	2	-5	-10	2
1	2	-4	-8	4
2	3	-3	-9	7
3	2	-2	-4	9
4	3	-1	-3	12
5	4	0	0	16
6	5	1	5	21
7	2	2	4	23
8	1	3	3	24
9	3	4	12	27
10	3	5	15	30
	$\sum f = 30$		$\sum fx = 5$	

(a) $\bar{x} = A + \frac{\sum fd}{\sum f} = 5 + \frac{5}{30} = 5.167$
(b) Modal value = 6
(c) Median = $\left(\frac{N}{2}\right)^{th}$ value = $\left(\frac{30}{2}\right)^{th}$ = 15th value from cumulative frequency, cf ; median = 5

Measures of dispersion

This is a measure of how the observations are spread out from the mean

(a) Range

It is the difference between the largest value and the smallest value.

Example 4

Find the range of the following set of values

2, 1, 3, 4, 5, 6, 7, 8, 9, 10, 3, 4, 6, 8, 9, 6, 3, 2

Range = 10 – 1 = 9

(b) Variance of a sample

The variance of x is denoted by Var(X) is defined as

Var(X) = $\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$

Var(X) = $\frac{\sum x^2}{n} - \bar{x}^2$

If the frequency is given then

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

= $\frac{\sum fx^2}{\sum f} - \bar{x}^2$

(c) Standard deviation

S.d = $\sqrt{\text{Var(X)}}$

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

= $\sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2}$

Example 5

Find the variance and standard deviation for the following data; 5, 8, 12, 13, 15

Solution

x	x^2
5	25
8	64
12	144
13	169
15	225
$\sum x$ =53	$\sum x^2$ =627

Var(X) = $\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{627}{5} - \left(\frac{53}{5}\right)^2$

= 125.4 - 112.36 = 13.04

S.d = $\sqrt{\text{Var(X)}} = \sqrt{13.04}$ =3.611

Example 5

The frequency distribution table shows the marks of some student from a certain school

x	45	63	65	66	70	72	75	80	88
f	3	5	6	4	6	2	1	2	1

Calculate standard deviation

Solution

x	f	fx	fx^2
45	3	135	6075
63	5	315	19845
65	6	390	25350
66	4	264	17424
70	6	420	29400
72	2	144	10368
75	1	75	5625
80	2	160	12800
88	1	88	7744
	$\sum f$ =30	1991	$\sum fx^2$ =134631

S.d = $\sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$

= $\sqrt{\frac{134631}{30} - \left(\frac{1991}{30}\right)^2}$

= $\sqrt{\frac{134631}{30} - \left(\frac{1991}{30}\right)^2}$

=$\sqrt{4487.7 - 4404.53}$

= 9.12

Using assumed mean to get variance and standard deviation

Var (X) = $\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2$

s.d = $\sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$

Example 6

The frequency distribution table shows the heights of some students at a certain shool

Height	154	155	160	164	171	180
Frequency	4	6	8	5	4	3

Determine the variance and standard deviation of the data using a working mean of 160

Solution

x	f	d = x - A	fd	fd^2
154	4	-6	-24	144
155	6	-5	-30	150
160	8	0	0	0
164	5	4	20	80
171	4	11	44	484
180	3	20	60	1200
	$\sum f$ =30		$\sum fd$ =70	$\sum fd^2$ =2058

Var (X) = $\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2$

= $\frac{2058}{30} - \left(\frac{70}{30}\right)^2$

= 63.156

s.d = $\sqrt{Var(x)}$ = $\sqrt{63.156}$ =7.95

Quartiles

A quartile is a value that divides given values into four equal parts

q_1 is the lower quartile and is defined by

$q_1 = \left(\frac{1}{4}N\right)^{th}$ value where N is the sum of all the variables

q_3 is the upper quartile and is defined by

$q_3 = \left(\frac{3}{4}N\right)^{th}$ value where N is the sum of all the variables

Percentiles

A percentile is a value that divides given values into 100 pats.

P_{10} is the 10^{th} percentile and is defined as

$P_{10} = \left(\frac{10}{100}N\right)^{th}$ value where N is the sum of all the variables

P_{90} is the 10^{th} percentile and is defined as

$P_{90} = \left(\frac{90}{100}N\right)^{th}$ value where N is the sum of all the variables

Deciles

A decile is a value that divides given values into 10 pats.

D_5 is the 5th decile and is defined as

$D_5 = \left(\frac{5}{10}N\right)^{th}$ value where N is the sum of all the variables

Example 7

The table below shows the marks obtained by 20 students in a mathematics test marked out of 20

Marks	10	11	12	13	14	15	16	17	18	19	20
Number of students	1	2	2	2	2	4	2	1	2	1	1

Find:

(a) Mean mark
(b) Standard deviation
(c) 60th percentile
(d) Interquartile range

Solution

x	f	cf	fx	fx^2
10	1	1	10	100
11	2	3	22	242
12	2	5	24	288
13	2	7	26	338
14	2	9	28	392
15	4	13	60	900
16	2	15	32	512
17	1	16	17	289
18	2	18	36	648
19	1	19	19	361
20	1	20	20	400
	$\sum f$ =20		$\sum fx$ =294	$\sum fx^2$ =4470

(a) Mean, $\bar{x} = \frac{\sum fx}{\sum f} = \frac{294}{20} = 14.7$

(b) s.d = $\sqrt{\frac{\sum fd^2}{\sum f} - \bar{x}^2}$

$= \sqrt{\frac{4470}{20} - 14.7^2}$

= 2.722

(c) 60th percentile

$= \left(\frac{60}{100}\ x\ 20\right)^{th}$ value

= 12th value from cf =15

(d) $q_1 = \left(\frac{1}{4}\ x\ 20\right)^{th}$ value = 5th value from cf = 12

$Q_3 = \left(\frac{3}{4}\ x\ 20\right)^{th}$ value = 15th value from cf = 16

Interquartile range = 16 -12 = 4

Example 8

Given the following scores

8, 6, 8, 9, 10, 6, 4, 5, 6, 4, 4, 6, 8, 7, 10, 8, 6, 11, 12,8

(a) Form a frequency distribution table of ungrouped data.

(b) Find the standard deviation
(c) Calculate semi-quartile range
(d) Determine the range of 45th and 90th percentile.

Solution

x	f	cf	fx	fx^2
4	3	3	12	48
5	1	4	5	25
6	5	9	30	180
7	1	11	7	49
8	5	16	40	320
9	1	17	9	81
10	2	19	20	200
11	1	20	11	121
12	1	21	12	144
	$\sum f$ =20		$\sum fx$ =146	$\sum fx^2$ =1168

(b) s.d = $\sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$

$= \sqrt{\frac{1168}{20} - \left(\frac{146}{20}\right)^2}$ =2.26

(c) $q_1 = \left(\frac{1}{4}\ x\ 20\right)^{th}$ value

= 5th value from cf = 6

$q_3 = \left(\frac{3}{4}\ x\ 20\right)^{th}$ value

= 15th value, from cf = 8

Semi-interquartile range = $\frac{8-6}{2} = 1$

(d) 45th percentile = $\left(\frac{45}{10}\ x\ 20\right)^{th}$ value = 9th value from cf = 6

90th percentile = $\left(\frac{90}{100}\ x\ 20\right)^{th}$ value = 18th value from cf = 10

The range between the 90th percentile and 45th percentile = 10 – 6 = 4

Revision exercise 1

1. The data below represents the length of leaves in cm: 4.5 4.4, 6.2, 9.4, 8.2, 12.6, 10.0, 8.8, 3.8 and 13.6. find the;
   (a) Mean
   (b) Standard deviation
2. The concentration in m per litre of a trace element in 7 randomly chosen samples of water from spring wells were: 240.8, 237.3, 236.6, 2333.9 and 232.5. Determine the mean and the variance of the concentration of the trace elements per litre.
3. The table below shows the length of flowers from a certain plant to the nearest 0.5cm.

Length (cm)	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0
Number of flowers	4	9	11	8	10	7	2	3

   Find the:
   (a) Mean
   (b) Mode
   (c) The median
   (d) Standard deviation
4. The marks scored by 11 students in a test are:52, 61, 78, 49, 47, 79, 54, 58, 62, 73, 72
   Find;
   (a) Median,
   (b) Mean,
   (c) Interquartile range
   (d) Semi-quartile range

Solutions to revision exercise 1

1. The data below represents the length of leaves in cm: 4.5 4.4, 6.2, 9.4, 8.2, 12.6, 10.0, 8.8, 3.8 and 13.6. find the;
   (a) Mean
   (b) Standard deviation

   Solution

x	x^2
4.5	20.25
4.4	19.36
6.2	38.44
9.2	84.64
8.2	67.24
12	144
10	100
8.8	77.44
3.8	14.44
13.6	184.96
$\sum x$ =80.7	$\sum x^2$ =750.77

Mean, $\bar{x} = \frac{\sum x}{n} = \frac{80.7}{10} = 8.07$

s.d = $\sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{750.77}{10} - (8.07)^2}$ =3.155

2. The concentration in m per litre of a trace element in 7 randomly chosen samples of water from spring wells were: 240.8, 237.3, 236.6, 2333.9 and 232.5. Determine the mean and the variance of the concentration of the trace elements per litre.

   Solution

x	x^2
240.8	57984.64
237.3	56311.29
236.7	56026.89
234.2	54849.64
236.6	55979.56
233.9	54709.21
232.5	54056.25
$\sum x$ =1652	$\sum x^2$ =389917.5

Mean, $\bar{x} = \frac{\sum x}{n} = \frac{1652}{7} = 236$

Var(x) = $\frac{\sum x^2}{n} - \bar{x}^2$

$= \frac{389917.5}{7} - 236^2$

= 55,702.5 – 55,696

= 6.5

3. The table below shows the length of flowers from a certain plant to the nearest 0.5cm.

Length (cm)	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0
Number of flowers	4	9	11	8	10	7	2	3

Find the:
(a) Mean
(b) Mode
(c) The median
(d) Standard deviation

Solution

x	f	fx	fx^2	cf
7.5	4	30	225	4
8.0	9	72	576	13
8.5	11	93.5	794.75	24
9.0	8	72	648	32
9.5	10	95	902.5	42
10.0	7	70	700	49
10.5	2	21	220.5	51
11.0	3	33	363	54
	$\sum f$ =54	$\sum fx$ =486.5	$\sum fx^2$ =4429.75	

(a)Mean = $\frac{\sum fx}{\sum f}$

$= \frac{486.2}{54} = 9.0$

(b) mode = 8.5

(c) Median = $\left(\frac{54}{2}\right)^{th}$ value

= 27th value

From cf the 27th value = 9

(d) s.d = $\sqrt{\frac{\sum x^2}{\sum f} - \bar{x}^2}$ = $\sqrt{\frac{4429.75}{54} - \bar{x}^2}$ = 1

4. The marks scored by 11 students in a test are:52, 61, 78, 49, 47, 79, 54, 58, 62, 73, 72
   Find;
   (a) Median
   Arrange values in ascending order
   47, 49, 52, 54, 58, 61, 62, 72,73,78,79
   (a) Median =$middle\ number$ = 6th value = 61
   (b) Mean,
   $\sum x$ = 52+ 61+78 + 49 + 47 + 79 + 54 + 58 + 62 + 73 + 72 =685
   Mean = $\frac{\sum x}{n} = \frac{685}{11}$ = 62.273
   (b) Interquartile range
   $q_1 = \left(\frac{1}{4}\ x\ 11\right)^{th}$ value = 2.75th value = 52

   $q_3 = \left(\frac{3}{4}\ x\ 11\right)^{th}$ value = 8.25th value = 73

   Interquartile range = 73 -52 = 21

   (c) Semi-quartile range = $\frac{Interquartile\ range}{2} = \frac{21}{2} = 10.5$

Continuous or grouped data.

This is data whose scores or values are said to be continuous and take interval values

Example 9

Class	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70 – 79	80 - 89
Number of students	4	5	7	3	6	4	1

Draw a frequency table

Terms used

(a) Class: these are limits of distribution. In the table above, the classes are: (20 – 29), (30 – 39), (40 – 49), (50 – 59), (60 – 69), (70 – 79), (80 – 89).

(b) Class mark (mark)
This is the mid-point value of the class. It is normally denoted by x. in the table above, the class marks are 24.5, 34.5, 44.5 …..

(c) Class boundary
These are continuous class limits. In the above table the first class boundary is (20-0.5) – (29 +0.5). In this case, the lower class boundary is 19.5 and upper class boundary is 29.5

For class interval 2.0 – 2.9, the class boundary is (2.0 -0.05) – (2.9 + 0.05) = 1.95-2.95.

(d) Class width or class interval
This is the width of each class boundary.
It is given by;
Class width = upper class boundary – lower class boundary
In the table above, class width = 29.5 – 19.5 = 10

Solution

Class	Frequency	Class width	Class mark, x	Class boundary
20 – 29	4	10	24.5	19.5 - 29.5
30 – 39	5	10	34.5	29.5 - 39.5
40 – 49	7	10	44.5	39.5 - 49.5
50 – 59	3	10	54.5	49.5 - 59.5
60 – 69	6	10	64.5	59.5 - 69.5
70 – 79	4	10	74.5	69.5 - 79.5
80 – 89	1	10	84.5	79.5 - 89.5

Example 10

The data below shows the heights in centimetres of 70 students.

Height (cm)	130 – 135	135 – 140	140 – 145	145 – 150	150 – 160	160 – 170	170 – 180
Number of students	10	12	8	9	11	15	5

Construct a frequency distribution table for the above data

Class	Class boundary	Class width	Class mark, x	Frequency, f
130 – 135	130 – 135	5	132.5	10
135 – 140	135 – 140	5	137.5	12
140 – 145	140 – 145	5	142.5	8
145 – 150	145 – 150	5	147.5	9
150 – 160	150 – 160	10	155	11
160 – 170	160 – 170	10	165	15
170 - 180	170 - 180	10	175	5

Example 11

Use the data below to construct a frequency distribution table

Marks	20-<30	30-<40	40-<50	50-<60	60-<70	70-<80	80-<90
Number of students	10	14	8	18	11	15	5

Solution

Class	Class boundary	Class width	Class mark, x	Frequency, f
20-<30	20 – 30	10	25	10
30-<40	30 – 40	10	35	14
40-<50	40 – 50	10	45	8
50-<60	50 – 60	10	55	18
60-<70	60 – 70	10	65	11
70-<80	70 – 80	10	75	15
80-<90	80 – 90	10	85	5

Example 12

The table below shows the ages of 35 people

Age	0-	5-	10-	15-	20-	30-	40-
Frequency	4	6	3	5	7	2	8

Draw a frequency table for the data

Solution

Class	Class boundary	Class width	Class mark, x	Frequency, f
0-	0 – 5	5	2.5	4
5-	5 – 10	5	7.5	6
10-	10 – 15	5	12.5	3
15-	15 – 20	5	17.5	5
20-	20 – 30	10	25	7
30-	30 – 40	10	35	2
40-	40 – 45	5	42.5	7

Note : the last class width is 5 since it is the most common

Example 13

The table below shows the marks of 40 students

Marks	-20	-30	-40	-50	-60	-65	-70
Frequency	8	4	7	10	2	2	7

Draw a frequency table

Solution

Class	Class boundary	Class width	Class mark, x	Frequency, f
-20	10 – 20	10	15	8
-30	20 – 30	10	25	4
-40	30 – 40	10	35	7
-50	40 – 50	10	45	10
-60	50 – 60	10	55	2
-65	60 – 65	5	62.5	2
-70	65- 70	5	67.5	7

Note: 1- the first class width is 10 because it is the most common

2- it is also acceptable for the first class to start from zero. i.e. (0 – 20)

Example 14

The data below shows the length in minutes made of different phone calls made by Airtel clients

Length (minutes)	<20	<30	<35	<40	<50	<60
Frequency	4	20	32	42	48	60

Construct a frequency table for the data.

Solution

Class	Class boundary	Class width	Class mark, x	Frequency, f
<20	10 – 20	10	15	4
<30	20 – 30	10	25	16
<35	30 – 35	5	32.5	12
<40	35 – 40	5	37.5	10
<50	40 – 50	10	45	6
<60	50 – 60	10	55	2

Measure of central tendency

(a) Mean or average of sample

The mean of grouped data is given by

$\bar{x}=\frac{\sum fx}{\sum f}$ where f = frequency, x is the mid-mark

(b) Median of grouped data

Median of grouped data is defined by

$$\text{Median} = L_b + \left(\frac{\frac{\sum f}{2} - c.f_b}{f}\right) C$$

Where L_b = lower class boundary of the median class

C= class width of the median class

f = frequency of the median class

$c.f_b$= cumulative frequency before that of the median class

Example 15

The ages of people in the town are as follows

Age	0-<5	<15	<30	<50	<70	<90
Number of people	44	81	105	147	158	160

Calculate the

(i) Mean
(ii) Median

Solution

Age	x	f	fx	cf	Class boundary
0-<5	2.5	44	110	44	0-<5
5-<15	10	37	370	81	5-<15
15-<30	22.5	24	540	105	15-<30
30-<50	40	42	1680	146	30-<50
50-<70	60	11	660	158	50-<70
70-<90	80	2	160	160	70-<90
		$\sum f$ =160	$\sum fx$ =3520		

(a) Mean, $\bar{x} = \frac{\sum fx}{\sum f} = \frac{3520}{160}$ = 22years

(b) Median = $L_b + \left(\frac{\frac{\sum f}{2} - c.f_b}{f}\right) C$

$\frac{\sum f}{2} = \frac{160}{2} = 80$

Median class boundary is 5 -<15, f = 37 and C = 10

∴ Median = $5 + \left(\frac{80-44}{37}\right) x10$ = 14.73 years

(c) Mode of grouped data

Mode of grouped data with equal class width is defined as

Mode = $L_b + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) C$

Where

L_b = lower class boundary of modal class

C = Class width of the modal class

Δ_1 = Modal frequency (pre-modal frequency)

Δ_2 = Modal frequency (post modal frequency)

Example 16

The table below shows the weight of 250 students at The Science Foundation College

Weight (kg)	44.0 – 47.9	48.0 – 51.9	52.0 – 55.9	56.0 – 59.9	60.0 – 63.9	64.0 – 67.9	68.0 – 71.9	72.0 – 75.9
Frequency	3	17	50	45	46	57	23	9

Find

(i) Average weight

(ii) Median weight

(iii) Modal weight

Solution

Class	Class boundary	x	f	fx	cf
44.0 - 47.9	43.95 - 47.95	45.95	3	137.85	3
48.0 - 51.9	47.95 - 51.95	49.95	17	849.15	20
52.0 - 55.9	51.95 - 55.95	53.95	50	2697.5	70
56.0 -59.9	55.95 - 59.95	57.95	45	2607.75	115
60.0 - 63.9	59.95 - 63.95	61.95	46	2849.7	161
64.0 - 67.9	63.95 - 67.95	65.95	57	3759.15	218
68.0 - 71.9	67.95 - 71.95	69.95	23	1608.85	241
72.0 - 75.9	71.95 - 75.95	73.95	9	665.55	250
			250	15175.5	

(i) Mean, $\bar{x} = \frac{\sum fx}{\sum f} = \frac{15175.5}{250} = 60.702kg$

(ii) Median = $L_b + \left(\frac{\frac{\sum f}{2} - c.f_b}{f}\right) C$

$\frac{\sum f}{2} = \frac{250}{2} = 125$

Median class boundary is 59.95 - 63.95, f = 46 and C = 4

∴ Median = $59.95 + \left(\frac{125-115}{46}\right) x\ 4$ = 60.82kg

(iii) Modal class boundary is 63.95 – 67.95, since 57 is the highest frequency and C = 4

$\Delta_1 = 57 - 46 = 11$ and $\Delta_2 = 57 - 23 = 34$

$Mode = 63.95 + \left(\frac{11}{11+34}\right) x\ 4 = 64.93kg$

(d) Mode of grouped data with unequal class width

Mode of grouped data with unequal class width defined as

Mode = $L_b + \left(\frac{\Delta_f.d_1}{\Delta_f.d_1 + \Delta_f.d_2}\right) C$

Modal class is determined from the highest frequency density

Frequency density = $\frac{frequency}{class\ width}$

Where

L_b = lower class boundary of modal class

C = Class width of the modal class

$\Delta_f.d_1$ = Modal frequency density (pre-modal frequency)

$\Delta_f.d_2$ = Modal frequency density (post modal frequency)

Example 17

Given the data below

Marks (x)	10-19	20-24	25-34	35-39	40-54	55-64	65-79
Frequency (f)	4	6	7	3	8	6	6

Find the mode

Solution

Class boundary	Class width	f	Frequency density
9.5 - 19.5	10	4	0.4
19.5 - 24.5	5	6	1.2
24.5 - 34.5	10	7	0.7
34.5 - 39.5	5	3	0.6
39.5 - 54.5	15	8	0.53
54.5 - 64.5	10	6	0.6
64.5 - 79.5	15	6	0.4
		$\sum f$ =40	

Mode = $L_b + \left(\frac{\Delta_f . d_1}{\Delta_f . d_1 + \Delta_f . d_2}\right) C$

$= 19.5 + \left(\frac{1.2-0.4}{(1.2-0.4)+(1.2-0.7)}\right) x\ 5$

= 22.58

Example 18

The table shows the weights (kg) of 150 patients who visited a certain health centre.

Weight (kg)	0 – 9	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59	60 - 69
Frequency (f	30	16	24	32	28	12	8

Calculate

(a) Mean
(b) Mode
(c) Median

Class	Class boundary	class mark (x)	f	fx	cf
0 - 9	0 - 9.5	4.5	30	135	30
10 - 19	9.5 - 19.5	14.5	16	232	46
20 - 29	19.5 - 29.5	24.5	24	588	70
30 - 39	29.5 -39.5	34.5	32	1104	102
40 - 49	39.5 - 49.5	44.5	28	1246	130
50 - 59	49.5 - 59.5	54.5	12	654	142
60 - 69	59.5 - 69.5	64.5	8	516	150
			$\sum f$ =150	$\sum fx$ =4475	

(a) Mean $\bar{x} = \frac{\sum fx}{\sum f} = \frac{4475}{150}$ = 29.83kg

(b) Mode = $L_b + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) C$

Modal class boundary is 29.5 – 39.5, since 32 is the highest frequency and C = 10

$\Delta_1 = 32 - 24 = 8$ and $\Delta_2 = 32 - 28 = 4$

$Mode = 29.5 + \left(\frac{8}{8+4}\right) x\ 10 = 36.17kg$

(c) Median = $L_b + \left(\frac{\frac{\sum f}{2} - c.f_b}{f}\right) C$

$\frac{\sum f}{2} = \frac{150}{2} = 75$

Median class boundary is 29.5 -39.5, f = 32 and C = 10

∴ Median = $29.5 + \left(\frac{75-70}{32}\right) x\ 10$ = 30.06kg

Measure of dispersion of grouped data

(i) Variance of the sample

The variance of grouped data denoted by Var(x) is defined as

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

Or

Var(x) = $\frac{\sum fx^2}{\sum f} - \bar{x}^2$

Or

Var(x) = $\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2$ where d = x – A and A = suggested mean

(ii) s.d = $\sqrt{Var(x)}$

Example 19

The table below shows the number of crimes committed by students

Number of crimes	5-<10	10-<20	20-<30	30-<50	50-<100
Number of students	10	15	25	40	26

Calculate the variance and standard deviation for the number of crimes committed

Solution

Number of crime	x	f	fx	fx^2
5-<10	7.5	10	75	562.5
10-<20	15	15	225	3375
20-<30	25	25	625	15625
30-<50	40	40	1600	64000
50-<100	75	25	1875	140625
		$\sum f$ =115	$\sum fx$ =4400	$\sum fx^2$ =224187.5

Var(x) = $\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2 = \frac{224187.5}{115} - \left(\frac{4400}{115}\right)^2 = 485.56$

s.d = $\sqrt{\text{Var(x)}} = \sqrt{485.56} = 22.04$

Example 20

The table below shows the weight of 250 students at a certain day school

Weight (kg)	44.0 – 47.9	48.0 – 51.9	52.0 – 55.9	56.0 – 59.9	60.0 – 63.9	64.0 – 67.9	68.0 – 71.9	72.0 – 75.9
Frequency	3	17	50	45	46	57	23	9

Using assumed mean of 57.95, find

(a) average weight
(b) variance
(c) standard deviation

Solution

weight	x	f	d = x - A	fd	fd^2
43.95-47.95	45.95	3	-12	-36	432
47.95-51.95	49.95	17	-8	-136	1088
51.95-55.95	53.95	50	-4	-200	800
55.95-59.95	57.95	45	0	0	0
59.95-63.95	61.95	46	4	184	736
63.95-67.95	65.95	57	8	456	3648
67.95-71.95	69.95	23	12	276	3312
71.95-75.95	73.95	9	16	144	2304
		$\sum f$ =250		$\sum fd$ =688	$\sum fd^2$ =12320

(a) $\bar{x} = A + \frac{\sum fd}{\sum f} = 57.95 + \frac{688}{250} = 60.702kg$

(b) Var(x) = $\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2$

$= \frac{12320}{250} - \left(\frac{688}{250}\right)^2 = 41.71kg$

(c) S.d = $\sqrt{Var(x)} = \sqrt{41.71} = 6.46kg$

Percentile and quartile of grouped data.

Percentile

This a value that divides a given distribution into 100 equal parts

The 60th percentile for instance is defined as

$P_{60} = L_b + \left(\frac{\frac{60}{100}\sum f - c.f_b}{f}\right)C$

Where

L_b = lower class boundary of the 60th class

C= class width

F = frequency of the 60th class

$c.f_b$ = cumulative frequency before that one of the 60th class

Quartiles

This a value that divides a given distribution into 4 equal parts

The lower quartile denoted q_1 for instance is defined as

$$q_1 = L_b + \left(\frac{\frac{1}{4}\sum f - c.f_b}{f}\right)C$$

Where

L_b = lower class boundary of the q_1 class

C= class width

f = frequency of the q_1 class

$c.f_b$ = cumulative frequency before that one of the q_1 class

The upper quartile denoted q_1 for instance is defined as

$$q_3 = L_b + \left(\frac{\frac{1}{4}\sum f - c.f_b}{f}\right)C$$

Where

L_b = lower class boundary of the q_3 class

C= class width

f = frequency of the q_3 class

$c.f_b$ = cumulative frequency before that one of the q_3 class

Interquartile range = $q_3 - q_1$

Semi-interquartile range = $\frac{q_3 - q_1}{2}$

Example 21

The following table shows the marks obtained by to students in a physics test marked out of 100

Marks (%)	20 -29	30-39	40-49	50-59	60-69	70-79	80-89	90-100
Number of students	4	6	2	5	7	8	5	2

Find

(a) Mean
(b) Standard deviation
(c) Median and mode
(d) Semi-interquartile range
(e) 40^{th} and 85^{th} percentile range

Solution

Class boundary	x	f	fx	fx2	cf
19.5-29.5	24.5	4	98	2401	4
29.5-39.5	34.5	6	207	7141.5	10
39.5-49.5	44.5	2	89	3960.5	12
49.5-59.5	54.5	5	272.5	14851.25	17
59.5-69.5	64.5	7	451.5	29121.75	24
69.5-79.5	74.5	8	596	44402	32
79.5-89.5	84.5	5	422.5	35701.25	37
89.5-99.5	94.5	3	283.5	26790.75	40
		$\sum f$ =40	$\sum fx$ =2420	$\sum fx^2$ =164370	

(i) Mean $\bar{x} = \frac{\sum fx}{\sum f} = \frac{2420}{40} = 60.5\%$

(ii) S.d = $\sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2} = \sqrt{\frac{164370}{40} - \left(\frac{2420}{40}\right)^2}$ = 21.19%

(iii) **Median** = $L_b + \left(\frac{\frac{\sum f}{2} - c.f_b}{f}\right)C$

$\frac{\sum f}{2} = \frac{40}{2} = 20$

Median class boundary is 59.5-69.5, f = 7 and C = 10

∴ Median = $59.5 + \left(\frac{20-17}{7}\right) x\ 10$ = 63.786%

Mode = $L_b + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)C$

Modal class boundary is 69.5-79.5, since 8 is the highest frequency and C = 10

$\Delta_1 = 8 - 7 = 1$ and $\Delta_2 = 8 - 5 = 3$

$Mode = 69.5 + \left(\frac{1}{1+3}\right) x\ 10 = 72\%$

(iv) $q_1 = L_b + \left(\frac{\frac{1}{4}\sum f - c.f_b}{f}\right)C$

$\frac{\sum f}{4} = \frac{40}{4} = 10$, L_b =29.5, f= 6, C = 10

$q_1 = 29.5 + \left(\frac{10-4}{6}\right)x10 = 39.5$

$q_3 = L_b + \left(\frac{\frac{3}{4}\sum f - c.f_b}{f}\right)C$

$\frac{3\sum f}{4} = \frac{3\ x\ 40}{4} = 30$, L_b =69.5, f= 8, C = 10

$q_1 = 69.5 + \left(\frac{30-24}{8}\right)x10 = 77\%$

Semi-quartile range = $\frac{q_3 - q_1}{2} = \frac{77-39.5}{2} = 18.75\%$

(v) $P_{40} = L_b + \left(\frac{\frac{40}{100}\sum f - c.f_b}{f}\right)C$

$\frac{40\sum f}{100} = \frac{40\ x\ 40}{100} = 16$, L_b =49.5, f= 5, C = 10

$P_{40} = 49.5 + \left(\frac{16-12}{5}\right)x10 = 57.5\%$

(vi) $P_{85} = L_b + \left(\frac{\frac{40}{100}\sum f - c.f_b}{f}\right)C$

$\frac{85\sum f}{100} = \frac{85\ x\ 40}{100} = 34$, L_b = 79.5, f= 5, C = 10

$P_{85} = 79.5 + \left(\frac{34-32}{5}\right)x10 = 83.5\%$

40^{th} and 85^{th} range = 83.5 – 57.5 = 26%

Example 22

Given the information in the table

Class	20-29	30-34	35-44	45-64	65-74	75-84
Frequency	5	5	12	20	10	8

Find

(a) Mean value
(b) Standard deviation
(c) Mode
(d) Median
(e) Interquartile range
(f) 90^{th} percentile

Solution

Class boundary	class width	x	f	f.d	fx	fx^2	cf
19.5-29.5	10	24.5	5	0.5	122.5	3001.25	5
29.5-34.5	5	32	5	1	160	5120	10
34.5-44.5	10	39.5	12	1.2	474	18723	22
44.5-64.5	20	54.5	20	1	1090	59405	42
64.5-74.5	10	69.5	10	1	695	48302.5	52
74.5-84.5	10	79.5	8	0.8	636	50562	60
			$\sum f$ =60		$\sum fx$ =3177.5	$\sum fx^2$ =185113.8	

(a) Mean $\bar{x} = \frac{\sum fx}{\sum f} = \frac{3177.5}{60} = 52.96$

(b) S.d = $\sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2} = \sqrt{\frac{185113.8}{60} - \left(\frac{3177.5}{60}\right)^2}$ = 16.75

(c) Mode = $L_b + \left(\frac{\Delta_f . d_1}{\Delta_f . d_1 + \Delta_f . d_2}\right) C$

=34.5 + $\left(\frac{1.2-1}{(1.2-1)+(1.2-1)}\right) 10$ = 39.5

(d) **Median** = $L_b + \left(\frac{\frac{\sum f}{2} - c.f_b}{f}\right) C$

$\frac{\sum f}{2} = \frac{60}{2} = 30$

Median = 44.5 + $\left(\frac{30-22}{20}\right) x\ 20$ = 552.5

(e) $q_1 = L_b + \left(\frac{\frac{1}{4}\sum f - c.f_b}{f}\right) C$

34.5 + $\left(\frac{\frac{1}{4} x\ 60-10}{20}\right) x\ 10$ = 38.67

$Q_3 = L_b + \left(\frac{\frac{3}{4}\sum f - c.f_b}{f}\right) C$

64.5 + $\left(\frac{\frac{3}{4} x\ 60-42}{10}\right) x\ 10$ = 67.5

Interquartile range = 67.5 – 38.67 = 28.83

(vii) $P_{90} = L_b + \left(\frac{\frac{90}{100}\sum f - c.f_b}{f}\right)C$

$= 74.4 + \left(\frac{\frac{90}{100}(60) - 52}{8}\right)x\ 10 = 77$

Graphs

(a) Grouped data with equal class width

(i) Histogram

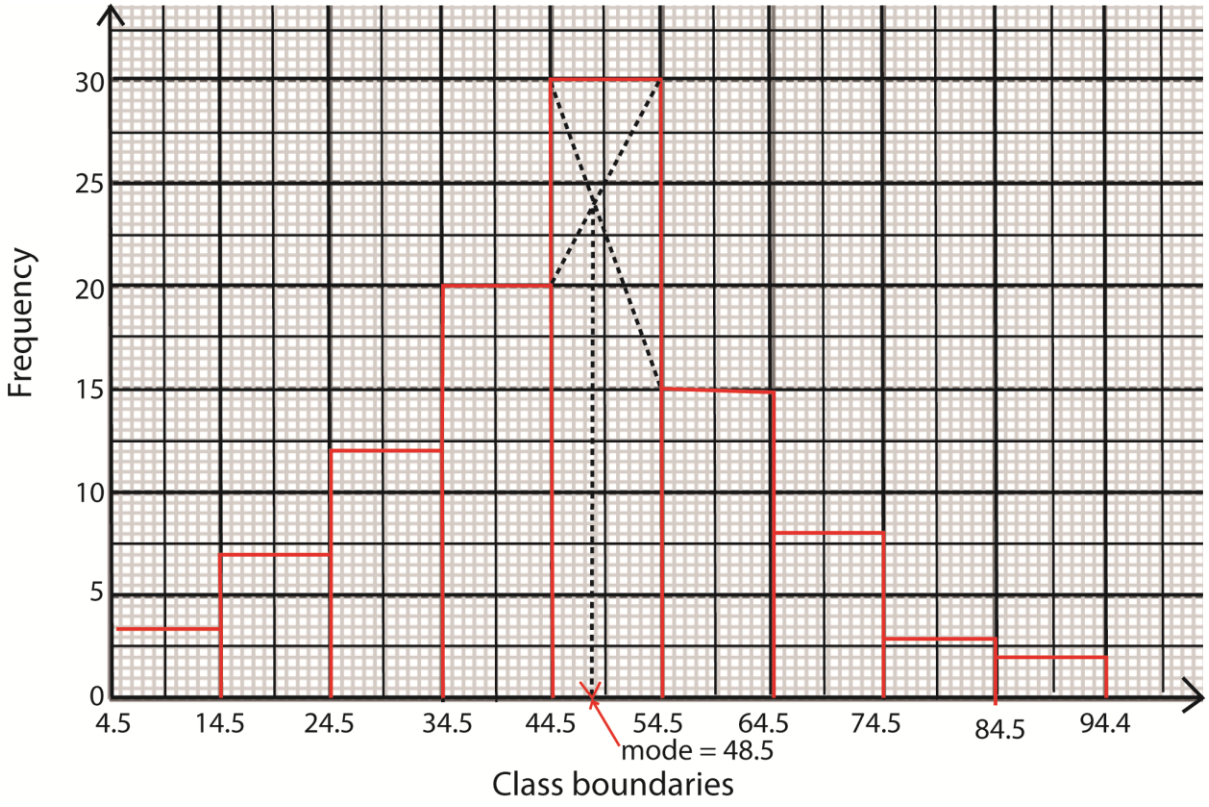
This is a graph consisting of vertical bars. It is a graph of frequency against class boundary. The area of the bar is equal to the frequency. Histogram is used to obtain the mode

Example 23

Given the data below

Marks	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-85	85-94
Frequency	3	7	12	20	30	15	8	3	2

Draw a histogram and use it to determine the mode



(ii) Cumulative frequency curve (Ogive)

This is a curve of cumulative frequency against class boundaries. An Ogive can be used to determine the median, quartiles, percentiles and deciles

Note: The values of cumulative frequency are plotted against the upper class boundaries and first value of the lower class boundary is plotted against cumulative frequency = 0

Example 24

Given the data below

Marks	20-29	30-39	40-49	50-59	60-69	70-79
Frequency	4	6	12	8	7	3

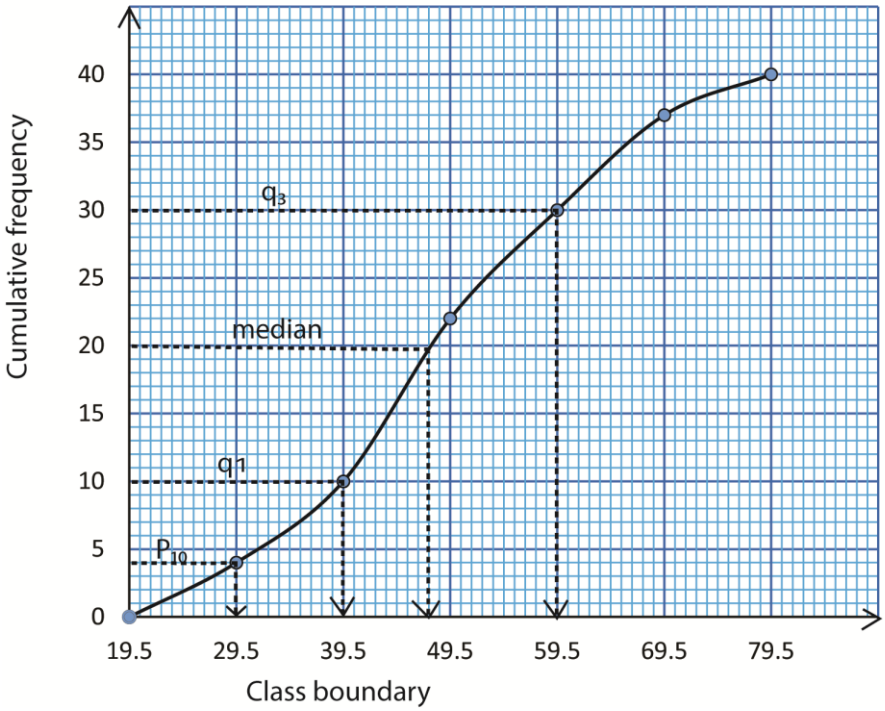
Draw an Orgive and use it to determine

(a) Median
(b) Interquartile range
(c) 10^{th} percentile

Solution

Class boundary	19.5 – 29.5	29.5–39.5	39.5–49.5	49.5 – 59.5	59.5 – 69.5	69.5 – 79.5
cf	4	10	22	30	37	40

An Orgive



(a) The median = $\left(\frac{N}{2}\right)^{th} = \left(\frac{40}{2}\right)^{th}$ = 20^{th} value from the graph = 48.5

(b) $q_1 = \left(\frac{N}{4}\right)^{th} = \left(\frac{40}{4}\right)^{th}$ = 10^{th} value ; from the graph q_1 = 39.5

$q_3 = \left(\frac{3N}{4}\right)^{th} = \left(\frac{3\ x\ 40}{4}\right)^{th}$ = 30^{th} value; from the graph q_3 = 59.5

Interquartile range = 59.5 – 39.5 = 20

(c) $P_{10} = \left(\frac{10N}{100}\right)^{th} = \left(\frac{10\ x\ 40}{100}\right)^{th}$ = 4^{th} value ; from the graph P_{10} = 29.5

Ungrouped data with unequal class width

(i) Histogram

This is a graph of frequency density against class boundary

Note that frequency density = $\frac{Frequency}{class\ width}$

Orgive

This is a graph of cumulative frequency against the class boundary

Example 25

The data shows the length in centimetres for different calendars produced by a printing press. A cumulative frequency distribution was formed

Length (cm)	<20	<30	<35	<40	<50	<60
Cumulative frequency	4	20	32	42	48	50

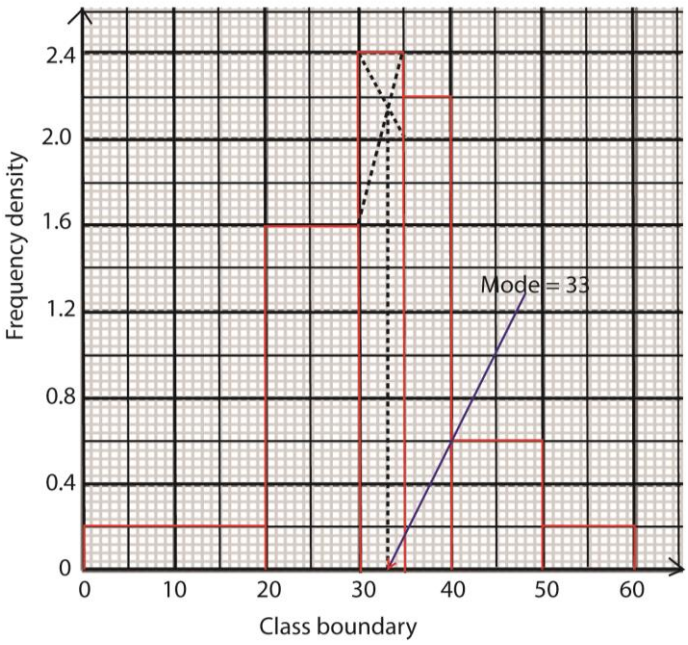
(a) Construct a frequency table.
(b) Find the mean length of the calendars
(c) Draw a histogram and use it to estimate the modal length
(d) Draw an Orgive and use it to estimate the median length.

Solution

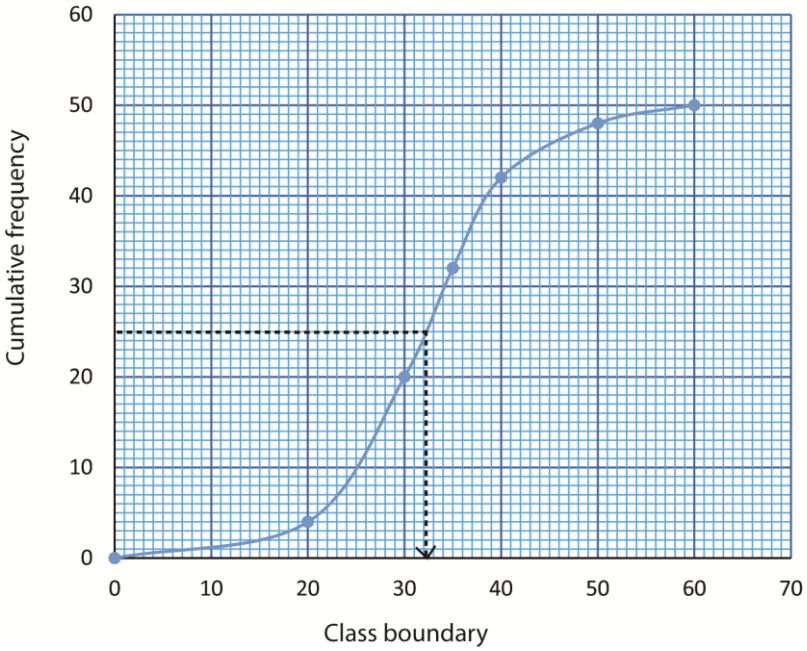
(a) Frequency table

Class boundary	x	f	fx	class width	frequency density	cf
0 - 20	10	4	40	20	0.2	4
20 - 30	25	16	400	10	1.6	20
30 - 35	32.5	12	390	5	2.4	32
35 - 40	37.5	10	375	5	2	42
40 - 50	45	6	270	10	0.6	48
50-60	55	2	110	10	0.2	50
		$\sum f$50	$\sum fx$ =1585			

(b) Mean $\bar{x} = \frac{\sum fx}{\sum f} = \frac{1585}{50} = 31.7$



An Orgive



Median length is the $\left(\frac{N}{2}\right)^{th} = \left(\frac{50}{2}\right)^{th}$ =25th value, from the graph = 32

Revision Exercise 2

UNEB 2002/2/7

The table below shows cumulative distribution of ages (in years of 400 student

Age(years)	<12	<13	<14	<15	<16	<17	<18	<19
Cumulative frequency	0	27	85	215	320	370	395	400

(a) Construct a cumulative frequency curve

(b) Use the curve to estimate

(i) Median age (Answer 14.9)

(ii) 20^{th} and 80^{th} percentile range (Ans. 2.1)

UNEB 2002/2/14

The table below shows the time taken by students to solve a mathematics problem

Time (mins)	5-9	10-14	15-19	20-24	25-29	30-34
Frequency	5	14	30	17	11	3

(a) Draw a histogram and use it to estimate the modal time. (ans. 17.3)

(b) Find the mean and standard deviation of solving the problem (Mean = 18.5mins, s.d = 5.9896 (4D))

UNEB 2004/2/14

The frequency distribution table shows the heights of s.6 students measured to the nearest cm;

Height	149-152	153-156	157-160	161-164	165-168	169-172	173-176
Frequency	5	17	20	25	15	6	2

(a) Calculate

(i) Mean height (Ans. 160.9cm)

(ii) Standard deviation (Ans. 5.5873)

(b) Draw a cumulative frequency curve and use it to estimate the median (Ans. 161cm) and range of height of the middle 60% of the candidates. (Ans. 10cm)

UNEB 2005/2/15

The table below the weights of some S.5 students from a certain school

Weight	50-53	54-57	58-61	62-65	66-69	70-73	74-77	78-81
Number of student	3	8	12	18	11	5	2	1

(a) Calculate

(i) Mean (63.1kg)

(ii) Standard deviation of students' weight (6kg)

(b) Draw a cumulative frequency curve and use it to estimate

(i) Median weight (63.1kg)

(ii) Number of students with weight between 58.9kg and 66.7kg (29students

UNEB 2006/2/12

The table below is the distribution of weights of a group of animals

Mass (kg)	Frequency
21-25	10
26-30	20
31-35	15
36-40	10
41-50	30
51-60	45
66-74	5

(a) Draw a cumulative frequency curve to estimate semi-quartile range (24kg)

(b) Find

(i) Mode (28.8333kg)

(ii) Standard deviation (11.772)

UNEB 2008/2/9

The table below shows the amount of money (in thousands of shillings) that was paid out as allowances to participants during a certain workshop

Amount (shs'000s)	110-114	115-119	120-129	130-134	135-144	145-159
Number of participants	13	20	32	17	16	12

(a) Draw a histogram and use it to estimate the modal allowance (shs. 11800)

(b) Calculate the:

(i) Median allowance (shs. 126,375/=)

(ii) Mean allowance (shs. 128,000/=)

UNEB 2009/2/11

The table below shows the income of 40 factory workers in millions of shillings per annum

1.0	1.1	1.0	1.2	5.4	1.6	2.0	2.5
2.1	2.2	1.3	1.7	1.8	2.4	3.0	2.2
2.7	3.5	4.0	4.4	3.9	5.0	5.4	5.3
4.4	3.7	3.6	3.9	5.2	5.1	5.7	1.5
1.6	1.9	3.4	4.3	2.6	3.8	5.3	4.0

(a) Form a frequency distribution table with class interval of 0.5millin shillings starting with the lowest limit of 1million shillings

(b) Calculate the

(i) Mean income (shs. 3,175,000)

(ii) Standard deviation (shs. 1,413,992.574)

(c) Draw a histogram to represent the above data. Use it to estimate the modal income
Modal income: (shs. 5, 200,000)

UNEB 2010/2/10

The table below shows the mars obtained by students in a physic test

Marks (%)	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	60-69	70-74
Frequency	9	12	10	17	13	25	18	14	8	8

(a) Draw a histogram and use it to estimate the modal mark. (52.5)

(b) Find the

(i) Mean mark (49.4627)

(ii) Standard deviation (12.424)

UNEB 2012/2/9

The table below shows the marks obtained in an examination by 200 candidates

Marks(%)	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89
Number of candidates	18	34	58	42	24	10	6	8

(a) Calculate the

(i) Mean mark (40.2%)

(ii) Modal mark (35.5%)

(b) Draw a cumulative frequency curve for the data. Hence estimate the lowest mark for a distinction one if the top 5% of the candidates qualify for the distinction. (75%)

UNEB 2013/2/1

A class performed on an experiment to estimate the diameter of a circular object. A sample of five students had the following results in centimetre. 3.13, 3.16, 2.94, 3.33 and 3.0.

Determine the sample;

(i) Mean (3.11)

(ii) Standard deviation (0.1356)(05marks)

UNEB 2017/2/12

The times taken for 55 students to have their lunch to the nearest minute are given in the table below

Time (minutes)	3 -4	5-9	10-19	20 – 29	30 – 44
Number of students	2	7	16	21	9

(a) Calculate the mean time for the student to have lunch. (mean=20.65) (04marks)

(b) (i) Draw a histogram for the given data

(ii) Use your histogram to estimate the modal time for the students to have lunch. (08marks) (modal time = 22 minutes)

UNEB 2018/2/9

1. The frequency distribution below shows the age of 240 students admitted to a certain University.

Age (years)	Number of student
18 - < 19	24
19 -< 20	70
20 -< 24	76
24 -< 26	48
26 -< 30	16
30 -< 32	6

(a) Calculate the mean age of the students. (mean =22.1458)(04mark)

(b) (i) Draw a histogram for the given data

(ii) Use the histogram to estimate the modal age (modal age = 19.58) (08mark)

UNEB 2019/2/1

The table shows the masses of bolts bought by a carpenter.

Mass (grams)	98	99	100	101	102	103	104
Number of bolts	8	11	14	20	17	6	4

Calculate the:

(a)median mass (101g)

(b) mean mass of the bolt(100.7625g) (05mark

UNEB 2019/2/10

The table below shows the marks obtained in a mathematic test by a group of student

marks	5 -<15	15-<25	25-<35	35-<45	45-<55	55-<65	65-<75	75-<100
Number of students	5	7	19	17	7	4	2	3

(a) Construct a cumulative frequency (O give) for the data (05 marks)

(b) Use your Ogive to find the

(i) Range between the 10^{th} and 70^{th} percentiles (26)

(ii) Probability that a student selected at random scored below 50 marks. 0.8125) (07 marks)

UNEB 2020/2/9

The table below shows the marks obtained by 100 students in a mathematics test

Marks	20-<40	40-<50	50-<55	55-<60	60-<70	70-<90	90-<100
Number of students	5	15	10	15	25	25	5

(a) Calculate the mean mark (63.125)

(b) Construct a cumulative frequency curve (Ogive) and use it to find the

(i) Median mark (61.5)

(ii) Range of the middle 40% of the mark (15)

Thank You

Dr. Bbosa Science