Statistics is a broad aspect of mathematics that collates, presents, analyses and interpretes data. Data can be described as an organised set of measurements or figures. All operations in statistics are carried out using tables, charts, graphs and statistical formulae. While there are myriad of tables and graphs at the workshop, at this point, we will examine the measures of central tendency and the measures of dispersion.

Measures of Central Tendency

Mean: The mean, also called the arithmetic mean, is the average mark (or score) of a set of data. The mean, commonly represented as $\bar{x}$, is also called average in mathematics.

The mean of an ungrouped data is expressed as $\bar{x} = \frac{\sum x}{n}$, where $\sum x$ is the addition of all the individual elements of the data set and n is the total number of data elements in the data. For a grouped data, the mean is expressed as

$\bar{x} = \frac{\sum fx}{\sum f}$: $\sum fx$ is the sum of the product of frequency (f) and class mark (x) for individual classes, and $\sum f$ is the sum of all the frequencies of all the classes.

Median: The median of a data set is the number that falls in the middle of the data set after the data set must have been carefully arranged in ascending or descending order of magnitude.

Mode: The mode of an ungrouped data set is the number (or element) that appears most in the data set. In order words, mode is the mark that has the highest frequency (most occurring element) in a data set.

Measures of Dispersion

The root word of dispersion is 'disperse,' which has to do with something being spread around (to distribute widely). For this reason, measures of dispersion provide a more detailed account of the distribution of data elements (scores) in a data. Measures of dispersion do **not** concentrate on giving information about the centre marks only, like the measures of central tendency; however, measures of dispersion show data variation over a great part (if not all parts) of the data set. Examples of these measures are range, interquartile range, semi-interquartile range, mean deviation, standard deviation, variance etc.

Range

Range measures the difference between the highest mark (score) and the least mark of the data set.

Interquartile-range and Semi-interquartile range

Quartile relates to the word 'quarter' which means a portion out of four equal divisions, and the concept of quartiles in statistics is built around this, but on a scale of 100%. Thus the first quartile in statistics is the first quarter of 100%, which can also be called the 25th percentile

since $\frac{100\%}{4} = 25\%$. The cumulative frequency curve below will serve as a good example to explain the quartiles.

In the cumulative frequency curve (figure 12.1), the highest point on the vertical axis is 80. This is because the sum total of all the frequencies of the data set is 80. Then, to locate the first quartile (which is the 25th percentile), first find 25% *of* 80 = 20; trace from the 20 mark on the vertical axis to the curve, then read the corresponding value on the horizontal axis, andi this is 40 *hours*. The second quartile can be seem as two quarter $\left(\frac{2}{4}\right)$ and $\frac{2}{4}$ *of* 100% = 50%; thus,, the second quartile is the 50th percentile, 50%« *of* 80 = 40. So, trace the 40 mark on the verticall

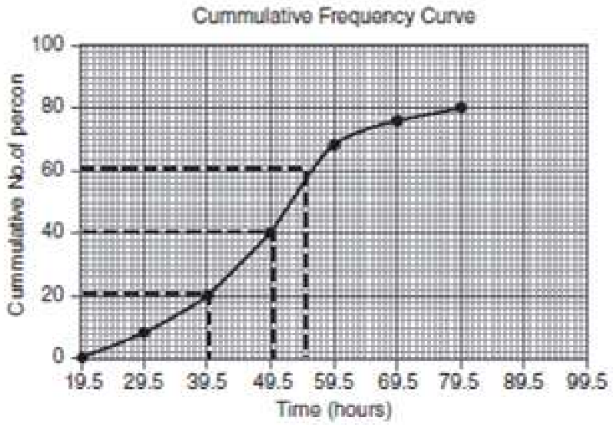

Fig. 12.1

axis to meet the ogive (curve) and read the corresponding value on the horizontal axis; the 50th percentile on the ogive is 49 *hours*. The second quartile (which is the 50th percentile) is the median: it divides the distribution of data into two equal halves such that there are equal number of data elements on both sides. By the same explanation, the third quartile will be the 75th percentile,75% o*f* 80 = 60. So, trace the 60 mark on the vertical axis to meet the ogive (curve) and read the corresponding value on the horizontal axis; the 75th percentile on the ogive is 56.5 *hours*.

Having understood the quartiles, the interquartile ange is the **difference** between the third quartile and the first quartile. Hence, from the ogive, interquartile range $Q3 - Q1 = 56.5 - 40 = 16.5$ *hours* . The

semi-interquartile range is expressed as $\frac{Q_3 - Q_1}{2}$.

Mean Deviation

Just like its name implies, the mean deviation is calculated by summing the product of the absolute values of the deviation of each mark (or class mark) and corresponding frequencies (that is $\Sigma f d$), and then dividing this

by the sum of all the frequencies as $\left(\frac{\Sigma f|d|}{\Sigma f}\right)$. In summary mean deviation is the average of the absolute deviation of each of the data set from the mean. *Mean deviation* $= \frac{\Sigma f|x - \bar{x}|}{\Sigma f} = \frac{\Sigma f|d|}{\Sigma f}$.
Where x represents the class marks.

Note that the absolute value of any number is its positive equivalence. For example, the absolute value of $-8 = |-8| = 8$, *the absolute value of* $+16 = |+16| = 16$, *the absolute value of* $-0.5 = |-0.5| = 0.5$.

Variance

Variance is the average deviation of all the data elements from the mean mark. Again you can view variance as adding together the deviation of each mark form the mean mark (this is achieved by subtracting each mark from the mean mark and then summing up all the resulting deviation). This sum is then divided by the total number of elements of the data set (marks). The difference between the mean deviation and the variance is that mean deviation calculates the **absolute** average deviation and **not** the actual average deviation calculated as variance. The standard deviation is the square root of variance.

The standard deviation is the square root of variance.
For ungrouped data, *Variance* $= \frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2$,
and *Standard deviation* $= \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$.
Should you be working with a grouped data,
Variance $= \frac{\Sigma f d^2}{\Sigma f} - \left(\frac{\Sigma f d}{\Sigma f}\right)^2$, and *Standard deviation* $= \sqrt{\frac{\Sigma f d^2}{\Sigma f} - \left(\frac{\Sigma f d}{\Sigma f}\right)^2}$. Come along to learn more at the workshop.

Statistics

1. The table shows the distribution of the ages (in years) of a group of people.

Ages (in years)	25–27	28–30	31–33	34–36	37–39	40–42
Frequency	2	9	15	19	10	5

(a) Draw a histogram of the distribution.

(b) Using your histogram, estimatethe modal age.

(c) Estimate the median, expressing youranswer correct to 1 decimal place.

(WAEC)

Workshop

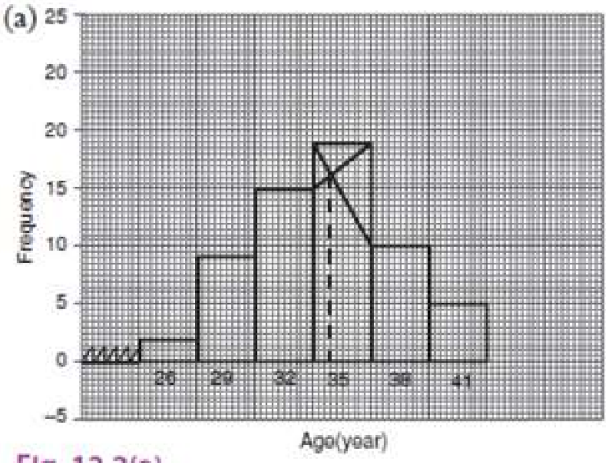

Fig. 12.2(a)

(b) From Figure 12.2(a), the modal class is the class with the tallest bar, which is class, 34 – 36. To know the modal age; draw a straight line from the top-right edge of the bar of the modal class to the top-right edge of the bar to its left. Also, draw a straight line from the top left edge of the bar of the modal class to the top left edge of the bar to its right (as shown above, in Figure 12.2(a). Note the point where the two lines intersect. Trace this point vertically to the horizontal axis and read the corresponding value on the horizontal axis. This value is the modal age. The modal class is 34 –36, hence, the thick line marking the beginning of the bar to the left, is the mid-point between 32 and 35, which is point 33.5 (lower boundary of the modal class) on the horizontal axis, so the modal age is a little more than the mid point between 33.5 and 35. The modal age is at about point 34.4. Therefore, the modal age of the distribution using the histogram, is 34.4 years.

(c) On Figure 12.2(b), the median of the distribution can be read on the horizontal axis from a vertical line dividing the whole histogram into two equal areas. The class widths are equal for all the classes, so we can take the width of each bar to be b. Area = Length × Breadth, Area of the first bar = $2 \times b = 2b$

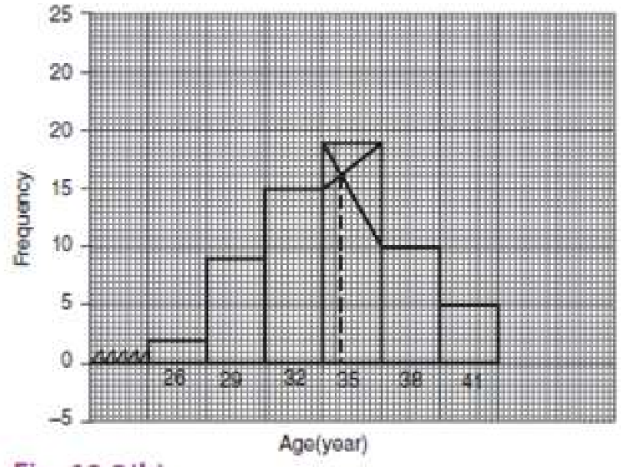

Fig. 12.2(b)

Area of the second bar = $9b$

Area of the third bar = $15b$;

Area of the fourth bar = $19b$

Area of the fifth bar = $10b$;

Area of the sixth bar = $5b$.

Then, the area of the histogram

= $2b + 9b + 15b + 19b + 10b + 5b = 60b$.

Half of the area = $\frac{60b}{2} = 30b$.

Back to the histogram, the area of the **first three** bars is $2b + 9b + 15b = 26b$ (not up to half) and the area of the first four bars = $26b + 19b = 45b$ (greater than $30b$). This means the vertical line dividing the histogram into two equal parts should be within the bar, with area $19b$ in the histogram (that is, class 34 – 36). Therefore, the median class is 34 – 36. As we learnt earlier, the area of the first three bars is $26b$, so we have to add an area of $4b$ to this $26b$ from the bar having the area $19b$, so as to divide the histogram into two equal areas. Now $19b - 4b = 15b$, so we will need to vertically divide the bar with area $19b$ in the ratio 4:15, in such a way that the $4b$ fraction will beto the side of the three bars with area $26b$, adding it up to $30b$, which will be half the area of the histogram.

On Figure 12.2(b), the thick line vertically divides the whole histogram into two equal halves, so the median mark will be the point where this vertical line crosses the horizontal axis. Now, the beginning of the bar for the median class is the lower class boundary of the class, which is 33.5. The thick line indicating the median mark is a little **less** than half way between the 33.5 and 35 marks. The mid-point between 33.5 and 35 is 34.25, but the median age is a little less than this value. Therefore, the median age is about 34.2. Correct to 1 decimal place, the median age is 34.2 years.

2. If the mean and the variance of the numbers 1, 4, x, y, 10 are 6 and 10 respectively, find the values of x and y. **(WAEC)**

Workshop

The mean (average) of the numbers is 6,

therefore, $\frac{1+4+x+y+10}{5} = 6$;

$\frac{15+x+y}{5} = 6$; $15 + x + y = 30$; $x + y = 15$ ..... (i)

$\left(\begin{matrix}\textit{The variance of an}\\ \textit{ungrouped data}\end{matrix}\right) = \frac{\Sigma d^2}{n} - \frac{\Sigma(x-x)^2}{n}$, where,

for this question, mean $\bar{x} = 6$ and $n = 5$. (*Note:* Σ *means summation (i.e addition)*).

$\frac{\Sigma(x-x)^2}{n}$

$= \frac{(1-6)^2 + (4-6)^2 + (x-6)^2 - (y-6)^2 + (10-6)^2}{5}$

$= 10$;

$$\frac{(-5)^2+(-2)^2+(x-6)^2+(y-6)^2+(4)^2}{5}=10;$$

$(-5)^2+(-2)^2+(x-6)^2+(y-6)^2+(4)^2=10\times 5;$

$25+4+(x-6)^2+(y-6)^2+16=50;$

$29+x^2-12x+36+y^2-12y+36+16=50;$

$x^2-12x+y^2-12y=50-117;$

$x^2+y^2-12x-12y=-67$ .....*(ii)*

Recall that $x+y=15$ .....*(i)*

$x^2+y^2-12x-12y=-67$ .....*(ii)*

From equation *(i)* above, $y=15-x$; put $y=15-x$ into *(ii)* to get

$x^2+(15-x)^2-12x-12(15-x)=-67;$

$x^2+225-30x+x^2-12x-180+12x=-67;$

$2x^2-30x+45=-67;$ $2x^2-30x+112=0$. Divide the equation through, by 2, to get $x^2-15x+56=0;$ $x^2-8x-7x+56=0;$ $x(x-8)-7(x-8)=0;$ $(x-8)(x-7)=0;$ $(x-8)=0$ *or* $(x-7)=0;$ $x=8$ *or* $x=7$. Recall that $y=15-x$, when $x=7$, $y=15-7=8$ and when $x=8$, $y=15-8=7$.

Therefore, the possible values of x, y are 7, 8 and 8, 7 respectively.

3. The table shows the distribution of teachers in 100 schools in a state.

No. of Teachers	No. of Schools
10–19	3
20–29	12
30–39	17
40–49	29
50–59	21
60–69	10
70–79	6
80–89	2

(a) Using an assumed mean of 44.5 calculate:

(i) correct to the nearest whole number, the mean number of teachers per school;

(ii) correct to 3 significant figures, the standard deviation of the distribution.

(b) Find the range of the distribution. **(WAEC)**

Workshop

(a) Assumed mean, $m = 44.5$.

Mean, $\bar{x} = \frac{\Sigma f d}{\Sigma f} + m$, and

$$\begin{pmatrix}\text{standard}\\ \text{deviation}\end{pmatrix} = \sqrt{\frac{\Sigma f d^2}{\Sigma f} - \left(\frac{\Sigma f d}{\Sigma f}\right)^2}.$$

Hence, to know the mean and standard deviation of these data, we need to first of all, find Σf, $\Sigma f d$ and $\Sigma f d$ 2. These three can be calculated from the table below.

Class Limit	Frequency (f)	Class Marks (x)	$d = x - m$	d 2	fd	fd 2
10 – 19	3	14.5	–30	900	–90	2 700
20 – 29	12	24.5	–20	400	–240	4 800
30 – 39	17	34.5	–10	100	–170	1 700
40 – 49	29	44.5	0	0	0	0
50 – 59	21	54.5	10	100	210	2 100
60 – 69	10	64.5	20	400	200	4 000
70 – 79	6	74.5	30	900	180	5 400
80 – 89	2	84.5	40	1 600	80	3 200
	$\Sigma f = 100$				$\Sigma fd = 170$	= 23 900

(i) $\text{Mean } \bar{x} = \frac{\Sigma fd}{\Sigma f} + m = \frac{170}{100} + 44.5 = 46.2.$

Therefore, the mean number of teachers correct to the nearest whole number is 46 *teachers*.

(ii) $S.D = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}$

$= \sqrt{\frac{23\,900}{100} - \left(\frac{170}{100}\right)^2} = \sqrt{239 - 2.89}.$

Standard deviation $= \sqrt{236.11} = 15.4$ *teachers* (correct to 3 significant figures).

(b) The range of this distribution, is the difference between the class mark of the uppermost class and the class mark of the lowest class.

Therefore *Range* $= 84.5 - 14.5 = 70$.

4. Below is the cumulative frequency table of the life-span of 100 rabbits in a controlled environment.

Life Span in Days	Cumulative No. of Rabbits
25	5
50	21
75	40
100	60
125	80

150	91
175	98
200	10

(a) Draw a cumulative frequency curve of the distribution.

(b) Use your curve to estimate:

(i) the semi-interquartile range,

(ii) the number of rabbits still alive after 130 days.

(c) Using the cumulative frequency table, copy and complete the table below.

No. of Days	No. of Rabbits
0–25	5
26–50	
51–75	
76–100	
101–125	
126–150	
151–175	7
176–200	

(d) Find the probability that a rabbit chosen at random in the environment will live beyond 125 days. (WAEC)

Workshop

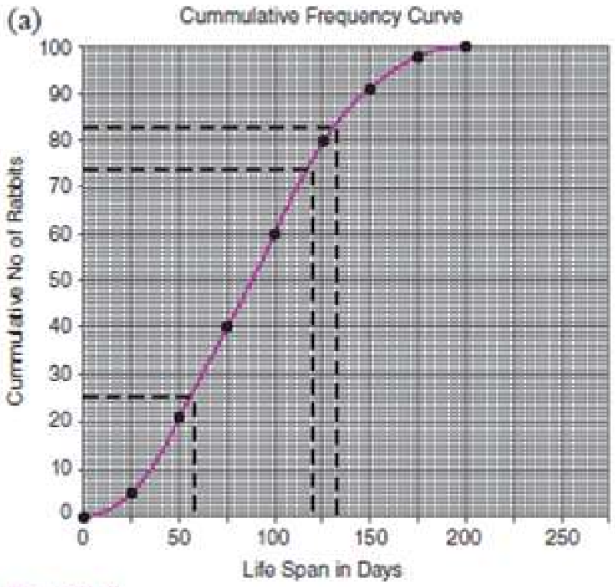

Fig. 12.3

(b) (i) Interquartile range $= (Q_3 - Q_1)$.

Semi-interquartile range $= \frac{(Q_3 - Q_1)}{2}$,
where Q_1 is the first quartile, and Q_3 is the third quartile.

For this question, the cumulative frequency of the distribution is 100. Thus, Q_1 can be read on the ogive, by first calculating 25% of the cumulative frequency, which is $\frac{25}{100} \times 100 = 25$. Then, trace the 25 mark on the cumulative frequency curve to the ogive, and then read its corresponding value on the horizontal axis; this is Q_1. Q_3 can be known from the graph (figure 12.3) by, calculating 75% of the cumulative frequency, which is $\frac{75}{100} \times 100 = 75$. Then, trace the 75 mark on the cumulative frequency curve to the ogive, and then read its corresponding value on the horizontal axis; this is Q_3. By this explanation from the graph, Q_1 is 56 and Q_3 is 118. Semi-interquartile range $= \frac{118 - 56}{2} = 31$.

Therefore, the semi-interquartile range of the distribution is 31 *days*.

Note that semi-interquartile range will bear the unit of Q_1 and Q_3.

(ii) Also, from the graph, the number of rabbits that are dead after 130 days can be known by tracing 130 mark on the horizontal axis to the ogive, then reading the corresponding cumulative frequency on the vertical axis. This gives 83; that is, the number of rabbits that cannot live beyond 130 days is 83 rabbits. Since we have 100 rabbits in all, the number of rabbits that are still alive, after 130 days is 100 – 83 = 17. So, 17 rabbits are still alive, after 130 *days*.

*Note that in this problem, 75 and 25 % of the total frequency (100) happen to be equal to 75 and 25 respectively. This **does not** apply in all cases. So, note that Q_1 and Q_3 are, respectively, 25% and 75% of the total frequency, which may **not** be 100. Also, note that since you were told to solve b)i and b)ii using your curve, you must not use any other method to solve it.*

(c) From the cumulative frequency table, the cumulative frequency of class 0-25 is 5, while the cumulative frequency of classes 0-50 is 21; therefore, the frequency of class 26-50 will be 21 – 5 = 16. The cumulative frequency of class 0-75 is 40, while the cumulative frequency of classes 0-50 is 21, then the frequency of class 51-75 will be 40 – 21 = 19. Also, frequencies 20, 20, 11 and 2, in the table

below, were calculated using the same method used to calculate frequencies 16 and 19 for class 26-50 and class 51-75 respectively.

No. of Days	No. of Rabbits
0–25	5
26–50	**16**
51–75	**19**

No. of Days	No. of Rabbits
76–100	**20**
101–125	**20**
126–150	**11**
151–175	7
176–200	**2**

(d) From the cumulative frequency table in the question, the number of rabbits in the environment, that will **not** live beyond 125 days is 80 *rabbits*. Because there are 100 rabbits in the environment, then the number of rabbits that will live beyond 125 days will be $100 - 80 = 20$ *rabbits*.

$$\Pr(\textit{that an event } E \textit{ will occur}) = \frac{\textit{number of elements in event space}}{\textit{number of elements in sample space}}.$$

$$\Pr\begin{pmatrix}\textit{that a rabbit selected will}\\ \textit{live beyond 125 days}\end{pmatrix} = \frac{\textit{number of rabbits that will live beyond 125 days}}{\textit{total number of rabbits in the environment}}$$

$$= \frac{20}{100} = \frac{1}{5}.$$

5. A scientist measured the length, *x mm* of the tails of 100 small reptiles of the same species. The table below is the summary of the results.

Tail length (x *mm*)	Frequency
$75.0 < x \leq 76.0$	8
$76.0 < x \leq 76.5$	12
$76.5 < x \leq 77.0$	28
$77.0 < x \leq 77.5$	22
$77.5 < x \leq 78.0$	12
$78.0 < x \leq 79.0$	10
$79.0 < x \leq 80.0$	8

(a) Using a scale of 2 *cm* to 5 units on the frequency axis, draw a histogram of the distribution.

(b) From your histogram, estimate, correct to one decimal place, the mode of the distribution. **(WAEC)**

Workshop

(a) From the table in question, the lower class boundary of the first class is 75.0 while the upper class boundary of this class is 76.0; 75.0 and 76.0 are the class boundaries and **NOT** the class limits. This is so because, the inequality signs shows that the class mark x – for the class, $75.0 < x \leq 76.0$ – is more than 75.0, and less than or equal to 76.0 (i.e. x cannot be more than 76.0 and x is greater than 75.0).

Therefore, from the format of this question, 75.0 and 76.0 are the respective lower and upper class boundaries of the class, $75.0 < x \leq 76.0$. Recall that;

Class width = (*Upper class Boundary*) – (*Lower class Boundary*); Then, the class width for the first class will be; 76.0 – 75.0 = 1.0. The class width for class $76.0 < x \leq 76.5$ (second class) is 76.5 – 76.0 = 0.5.

You will observe that the class width of the classes are not the same, therefore, we have to plot a frequency density on the vertical axis of the histogram and **NOT** the conventional frequency. We can calculate the frequency density for each class as

$$\left(\begin{matrix}\text{Frequency} \\ \text{density}\end{matrix}\right) = \frac{\text{Frequency}}{\text{Class Width}}$$

The frequency density for the classes is calculated in the table below.

Tail length (x mm)	Class Mark	Class Width	Frequency	$\left(\begin{matrix}\text{Frequency} \\ \text{density}\end{matrix}\right) = \frac{\text{Frequency}}{\text{Class Width}}$
$75.0 < x \leq 76.0$	75.50	1.0	8	8
$76.0 < x \leq 76.5$	76.25	0.5	12	24
$76.5 < x \leq 77.0$	76.75	0.5	28	56
$77.0 < x \leq 77.5$	77.25	0.5	22	44

Tail length (x mm)	Class Mark	Class Width	Frequency	$\left(\begin{matrix}\text{Frequency} \\ \text{density}\end{matrix}\right) = \frac{\text{Frequency}}{\text{Class Width}}$
$77.5 < x \leq 78.0$	77.75	0.5	12	24
$78.0 < x \leq 79.0$	78.50	1.0	10	10
$79.0 < x \leq 80.0$	79.50	1.0	8	8

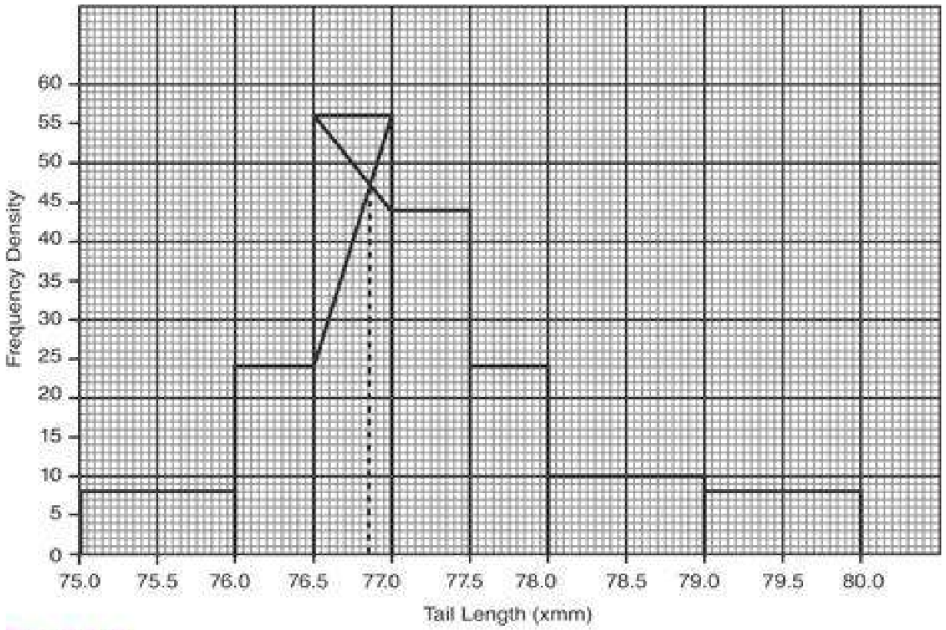

Fig. 12.4

You will observe from the histogram in Figure 12.4 that the classes with class width of 1, have bars that are twice the width of the classes with class width of 0.5.

Please, note that the frequency density is plotted against the upper class boundaries, i.e. 9.5, 19.5, etc, and not the upper class limits, 9, 19, etc.

In addition, from the histogram, the modal class (*class with the highest frequency*) is $76.5 < x \leq 77.0$. To know the modal age; a straight line was drawn from the top right edge of the bar of the modal class to the top right edge of the bar to the left of the modal class bar. A straight line was also drawn from the top left edge of the bar of the modal class to the top left edge of the bar to the right of the modal class bar.

The point of intersection of the two lines was then traced **vertically** to the horizontal axis, and the corresponding value read; this is the modal age of the distribution.

As indicated on the graph, the line marking the middle of the bar representing the modal class is 76.75 *mm*, while the thick line marking the end of this bar is 77.0. Looking closely at the diagram, the modal length is half way between 76.75 and 77.0. The length halfway between these two values is 76.875. Therefore, the mode of the distribution, correct to one decimal place is 76.9 *mm*.

6. The deviation of a set of numbers from an assumed mean 7 are:
−4, −2, −1, 0, 1, 3, 4, 5. Calculate, correct to two decimal places, the:
(a) mean,
(b) standard deviation of the numbers. **(WAEC)**

Workshop

Because **none** of the values of deviation in question is equal, then the data, from which the assumed mean was subtracted to get this deviation, are also **not** equal. Hence, the frequency of each mark in the data is 1. Therefore,
$\Sigma fd = \Sigma(1)d = \Sigma d$. This is because frequency f, of each mark, x, is 1, i.e $f = 1$, for each mark.
$\Sigma d = -4 + (-2) + (-1) + 0 + 1 + 3 + 4 + 5$;
$\Sigma d = -7 + 13 = +6$.
$\Sigma d2 = (-4)^2 + (-2)^2 + (-1)^2 + 0^2 + 1^2 + 3^2 + 4^2 + 5^2$
$= 16 + 4 + 1 + 0 + 1 + 9 + 16 + 25 = 72$.
The frequency of each mark in the data is 1 and there are 8 marks in all, so, $\Sigma f = 8$.

(a) For **ungrouped** data,

$$\bar{x} = \text{Assumed Mean} + \frac{\Sigma d}{\Sigma f} = 7 + \frac{6}{8} = 7.75.$$

Note that, for ungrouped data, $\bar{x} = \text{Assumed Mean} + \frac{\Sigma d}{\Sigma f}$, *while for grouped data,* $\bar{x} = \text{Assumed Mean} + \frac{\Sigma fd}{\Sigma f}$.

(b) The Standard deviation of an ungrouped data is expressed as

$$\sqrt{\frac{\Sigma d^2}{\Sigma f} - \left(\frac{\Sigma d}{\Sigma f}\right)^2} = \sqrt{\frac{72}{8} - \left(\frac{6}{8}\right)^2}$$
$$= \sqrt{9 - 0.5625} = \sqrt{8.4375}$$
$$= 2.9$$

Note that for grouped data, standard deviation $= \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}$.

7. The following table shows the distribution of nurses in some hospitals in a state.

No. of Nurses	No. of Hospitals
0–9	2
10–19	7
20–29	9
30–39	11
40–49	13
50–59	14
60–69	18
70–79	15
80–89	8
90–99	3

(a) Construct a cumulative frequency table for the distribution.

(b) Use your table to draw a cumulative frequency curve.

(c) From your curve, estimate the number of hospitals having at least 55 nurses.

(d) If an hospital is selected at random, what is the probability that:

(i) it has between 40 and 49 nurses,

(ii) it has less than 35 nurses? **(WAEC)**

Workshop

(a)

No. of Nurses	Cumulative frequency
Not more than 0	0
Not more than 9.5	2
Not more than 19.5	9
Not more than 29.5	18
Not more than 39.5	29
Not more than 49.5	42
Not more than 59.5	56
Not more than 69.5	74
Not more than 79.5	89
Not more than 89.5	97
Not more than 99.5	100

(b)

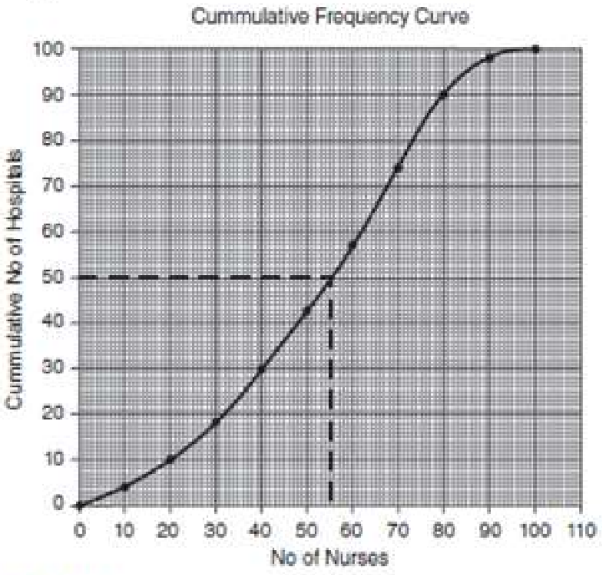

Fig. 12.5

(c) Convince yourself, from the ogive in figure 12.5, that the number of hospitals having, **at most**, 55 nurses is 50 hospitals. This was obtained by tracing the 55 nurses mark on the horizontal axis to the ogive, then the corresponding value of the cumulative number of hospitals was read on the vertical axis. Since the total number of hospitals is 100, and the number of hospitals having **at most** (*a maximum of*) 55 nurses is 50, the number of hospitals having **at least** (*a minimum of*) 55 nurses will be = 100 – 50 = 50.

Therefore, the number of hospitals having at least 55 nurses is 50 hospitals.

(d) Recall that, $\Pr\begin{pmatrix}\textit{that an event } E\\ \textit{will happen}\end{pmatrix}$

$$= \frac{\textit{number of elements in event space}}{\textit{number of elements in sample space}}.$$

(i) Then, $\Pr\begin{pmatrix}\textit{that an hospital selected}\\ \textit{at random has between}\\ 40 \textit{ and } 49 \textit{ nurses}\end{pmatrix}$

$$= \frac{\textit{number of hospitals that have between 40 and 49 nurses}}{\textit{Total possible number of nurses}} = \frac{13}{100}.$$

(ii) From the table in question,

$\begin{pmatrix}\textit{the number of hospitals that have}\\ \textit{less than, or equal to, 29 nurses}\end{pmatrix}$

$= 2 + 7 + 9 = 18.$

Also, from the table

$\begin{pmatrix}\textit{the number of hospitals that}\\ \textit{have from 30 to 29 nurses}\end{pmatrix} = 11.$

Now, class 30 – 39 can be divided into 2 equal **subclass** of 30 – 34 and 35 – 39. The number of hospitals for the 30 – 39 class is 11, and since this class has been divided into two equal subclasses, each of the subclasses of the number of nurses in the class 30 – 39 will have $\frac{11}{2}$ *hospital* allotted to them. Thus, the number of hospitals in the class 30 – 39, that have less than 35 nurses, will be the

number of hospitals having 30 – 34 nurses, which is $\frac{11}{2}$ *hospital*. Therefore the **total** number of hospitals, that have **less** than 35 nurses, will be

$$\begin{pmatrix}\text{the number of hospitals}\\ \text{that have less than, or}\\ \text{equal to, 29 nurses}\end{pmatrix} + \begin{pmatrix}\text{the number of hos-}\\ \text{pitals that have from}\\ \text{30 to 34 nurses}\end{pmatrix}$$

$= 18 + \frac{11}{2} = 23\frac{1}{2}$ *hospitals*. But the number of hospitals cannot be a fraction, as there is nothing like half hospital. Then, we need to round $23\frac{1}{2}$ up to a whole number. $23\frac{1}{2} = 23.5$, which is approximately equal to 24. The number of hospitals, that have less than 35 nurses are 24 *hospitals*; therefore,

$$\Pr\begin{pmatrix}\text{that an hospital}\\ \text{selectedat random}\\ \text{has less than 35 nurses}\end{pmatrix} =$$

$$\frac{\text{number of hospitals that haveless than 35 nurses}}{\text{Total possible number of nurses}}$$

$$= \frac{24}{100} = \frac{6}{25}.$$

8. The table shows the distribution of the ages of women in a village who had malaria during pregnancy and those who did not.

Age (years)	15–24	25–34	35–44	45–54	55–64
No. with Malaria	10	40	35	13	2
No. free of Malaria	10	20	35	30	5

(a) Calculate the mean age of **all** the women.

(b) On the same graph sheet, draw separately, the cumulative frequency curves of women:

(i) who had malaria;

(ii) without malaria.

(c) Use your curves to determine the number of women above the mean age, who:

(i) had malaria;

(ii) did not have malaria. **(WAEC)**

Workshop

(a) We are to calculate the mean age of **all** the women, so the frequency of each class will be calculate by adding the frequency of women with malaria during pregnancy to the frequency of women free of malaria during pregnancy, for each of the classes.

Age (years)	Class Mark (x)	Frequency (f)	fx
15 – 24	19.5	20	390.0
25 – 34	29.5	60	1 770.0
35 – 44	39.5	70	2 765.0
45 – 54	49.5	43	2 128.5
55 – 64	59.5	7	416.5
		$\Sigma f = 200$	$\Sigma fx = 7\,470$

The mean age $\bar{x}$ of **all** the women will be

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{7470}{200} = 37.35 \text{ years.}$$

Therefore, the mean age of all the women is 37.35 years.

(b) The cumulative frequency table of the Women with Malaria during pregnancy is drawn below.

Women with Malaria	Cumulative frequency
Not exceeding 14.5	0
Not exceeding 24.5	10
Not exceeding 34.5	50
Not exceeding 44.5	85
Not exceeding 54.5	98
Not exceeding 64.5	100

The cumulative frequency table of the Women free of Malaria during pregnancy is drawn below.

Women free of Malaria	Cumulative frequency
Not exceeding 14.5	0
Not exceeding 24.5	10
Not exceeding 34.5	30
Not exceeding 44.5	65
Not exceeding 54.5	95
Not exceeding 64.5	100

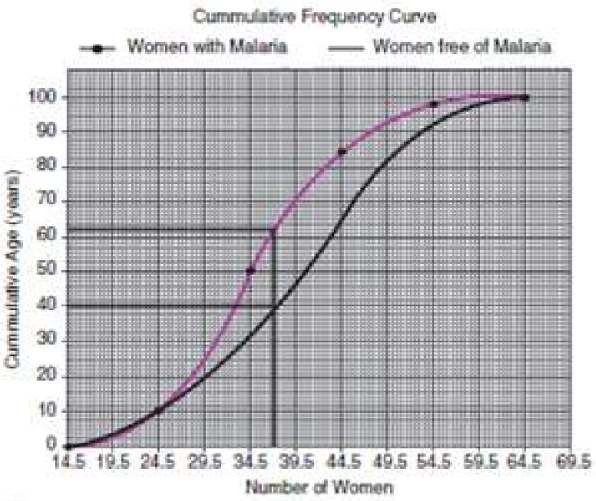

Fig. 12.6

(c) The mean age of this distribution was evaluated as 37.35.

(i) From Figure 12.6, the number of women **below** the mean age that had malaria during pregnancy, can be known by tracing the 37.35 mark (the mean) on the horizontal axis, to the ogive of the

women with malaria; the corresponding value read on the cumulative frequency axis, is the number of women **below** the mean age who had malaria during pregnancy. From the graph, the number of women **below** the mean age who had malaria during pregnancy is 62. The total number of women with malaria is 100, hence the number of women **above** the mean age who had malaria during pregnancy = 100 – 62 = 38. Therefore, the number of women **above** the mean age, who had malaria during pregnancy is 38.

(ii) The number of women **below** the mean age, that are free of malaria can be known by tracing the 37.35 mark on the horizontal axis, to the ogive of the women free of malaria; the corresponding value read on the cumulative frequency axis is the number of women **below** the mean age who are free of malaria. As shown on the graph, the number of women **below** the mean age who are free of malaria is 40. But the total number of women free of malaria is 100; thus, the number of women **above** the mean age, who are free of malaria = 100 – 40 = 60. Therefore, the number of women **above** the mean age, who are free of malaria is 60.

9.

Mass (kg)	20–29	30–39	40–49	50–59	60–69	70–79
No.of Sheep	187	252	196	130	88	47

The table shows the distribution of the masses of sheep in a pen.

(a) Using an assumed mean of 34.5, find:

(i) the mean;

(ii) the standard deviation, of the distribution.

(b) If a sheep is taken at random from the pen, what is the probability that it's mass is at least 50 kg? **(WAEC)**

Workshop

(a) Assumed mean, $m = 34.5$. To calculate mean and standard deviation, we need to know Σf, Σfd and Σfd^2, and these can be calculated from the table below.

Class Limit	Frequency (f)	Class Centre (x)	Deviation $d = x - m$	d^2	fd	fd^2
20 – 29	187	24.5	–10	100	–1 870	18 700
30 – 39	252	34.5	0	0	0	0
40 – 49	196	44.5	10	100	1 960	19 600
50 – 59	130	54.5	20	400	2 600	52 000
60 – 69	88	64.5	30	900	2 640	79 200
70 – 79	47	74.5	40	1 600	1 880	75 200
	$\Sigma f = 900$				$\Sigma fd = 7\,210$	$\Sigma fd^2 = 244\,700$

(i) Mean $\bar{x} = \frac{\Sigma fd}{\Sigma f} + m = \frac{7\,210}{900} + 34.5$

$= 42.5$ kg.

Therefore, the mean mass of the sheep is 42.5 kg.

(ii) Standard Deviation,

$$S.D = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2};$$

$$S.D = \sqrt{\frac{244\,700}{900} - \left(\frac{7\,210}{900}\right)^2}$$

$$= \sqrt{271.89 - 64.18}$$

$$= \sqrt{207.71} = 14.41\text{kg}.$$

Therefore, the standard deviation of the distribution is 14.41 kg.

(b) Recall that $\Pr\begin{pmatrix}\textit{that an event,}\\ \textit{E, will occur}\end{pmatrix}$

$$= \frac{\textit{number of elements in event space}}{\textit{number of elements in sample space}}$$

So, $\Pr\begin{pmatrix}\textit{that a sheep's mass}\\ \textit{is at } \textbf{least}\ 50\,kg\end{pmatrix}$

$$= \frac{\textit{Number of sheeps that are at least } 50\,kg}{\textit{Total number of sheeps available}}$$

$$= \frac{130 + 88 + 47}{900} = \frac{265}{900} = \frac{53}{180}.$$

Therefore, the probability that a sheep taken at random from the pen is at **least** 50 kg is $\frac{53}{180}$.

10. The table shows the distribution of the ages of a group of people in a village.

Age (in years)	Frequency
15 – 18	40
19 – 22	33
23 – 26	25
27 – 30	10
31 – 34	8
35 – 38	4

Using an assumed mean of 24.5, calculate the mean of the distribution. **(WAEC)**

Workshop

Assumed mean, $m = 24.5$.

Class Limit	Frequency (f)	Class Centre (x)	$d = x - m$	fd
15 – 18	40	16.5	–8	–320
19 – 22	33	20.5	–4	–132
23 – 26	25	24.5	0	0
27 – 30	10	28.5	4	40
31 – 34	8	32.5	8	64
35 – 38	4	36.5	12	48
	$\Sigma f = 120$			$\Sigma fd = -300$

Recall that the assumed mean, $m = 24.5$.

$$\text{Mean, } \bar{x} = \frac{\Sigma fd}{\Sigma f} + m = \frac{-300}{120} + 24.5$$

$$= -2.5 + 24.5$$

$$= 22 \text{ years}.$$

11. The table below shows the distribution of hours spent at work by the employee of a factory in a week.

Time (hours)	No. of Persons
20 – 29	8
30 – 39	11
40 – 49	23
50 – 59	25
60 – 69	8
70 – 79	5

(a) Draw an ogive for the distribution.

(b) Using your graph, estimate the:

(i) median;

(ii) lower quartile;

(iii) 40th percentile;

(iv) number of employees that spentat least 50 hours 30 minutes. **(WAEC)**

Workshop

(a)

Time (Hours)	Number of Persons (Frequency)
20 – 29	8
30 – 39	11
40 – 49	23
50 – 59	25
60 – 69	8
70 – 79	5

Recall that an ogive in statistics is a graph of cumulative frequency against upper class boundaries.

Time (Hours) Upper Class Boundary	Cumulative Frequency
Not exceeding 19.5	0
Not exceeding 29.5	8
Not exceeding 39.5	19
Not exceeding 49.5	42
Not exceeding 59.5	67
Not exceeding 69.5	75
Not exceeding 79.5	80

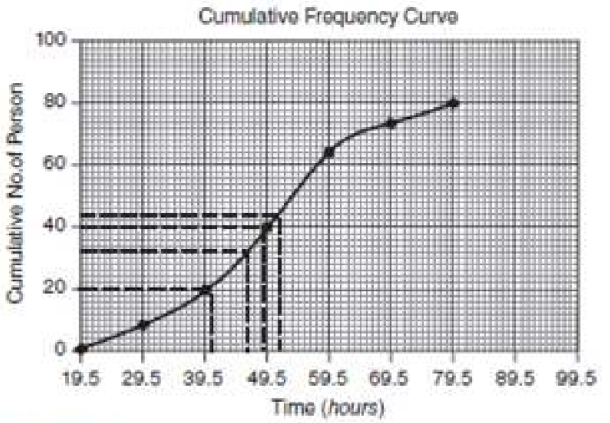

Fig. 12.7

(a) (i) The median is the 50th Percentile, to know the median one should first find 50% of the total frequency. The total frequency of the distribution is 80, 50% of 80 is 40, so, to get the median on the ogive, trace the 40 mark on the cumulative frequency axis to the ogive and read its corresponding value on the horizontal axis. This corresponding value on the horizontal axis is the median of the distribution. By this explanation, from the graph, the median of the distribution is 49.0 hours.

Note that, on the horizontal axis, the smaller divisions between the 19.5 and 29.5 mark are divided into 5 small boxes; 29.5 – 19.5 = 10 units, so that each of the 5 divisions between 19.5 and 29.5 will be $\frac{10}{5} = 2$ *units each. Then, the first small unit after 19.5 mark will be 19.5 + 2 = 21.5, followed by 21.5 + 2 = 23.5 up till the 29.5 mark. This explanation applies to other divisions on the horizontal axis.*

(ii) The lower quartile ($Q1$) is the 25^{th} percentile. 25% of 80 is 20, so, to get the 25th percentile, trace the 20 mark on the cumulative frequency axis to the ogive and read its corresponding value on the horizontal axis. This corresponding value on the horizontal axis is the 25th percentile of the distribution. From the graph, the 25th percentile of the distribution is 40 hours.

(iii) 40% of 80 is 32, thus, to get the 40^{th} percentile, trace the 32nd mark on the cumulative frequency axis to the ogive and then read its corresponding value on the horizontal axis. From the graph, the 40th percentile is 45.5 hours.

(iv) 50 hours 30 minutes is equal to 50.5 hours. The number of employees that spent **at most** 50.5 hours can be deduced from the graph by tracing the 50.5 mark on the horizontal axis to the ogive, then reading the corresponding value on the cumulative frequency axis. From the graph, the number of employees that spent **at most** 50.5 hours is 44. Now, we have 80 people in all. The number of employees that spent **at least** 50.5 hours will be 80 – 44 = 36. Therefore, the number of employee that spent at least 50 hours 30 minutes is 36.

12. The table shows the distribution of ages of workers in a company.

Age (in years)	No. of workers
17 – 21	12
22 – 26	24
27 – 31	30
32 – 36	37
37 – 41	45
42 – 46	25
47 – 51	10
52 – 56	7

(a) Using an assumed mean of 39, calculate the:

(i) mean;

(ii) standard deviation of the distribution.

(b) If a worker is selected at random from the company for an award, what is the probability that he is at most 36 years old? **(WAEC)**

Workshop

(a) Given that assumed mean $A = 39$, we can calculate mean as, $\bar{x} = A + \frac{\Sigma fd}{\Sigma f}$. Also, we can calculate the standard deviation as $S.D = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}$. Thus, to know the mean, $\bar{x}$, and standard deviation S.D, we need to know Σfd, Σf and Σfd^2, and we can obtain these from the table below.

Age (Years)	Class Mark X (years)	Frequency (f) (No. of Workers)	$d = X - A = X - 39$	fd	d^2	fd^2
17 – 21	19	12	– 20	– 240	400	4 800
22 – 26	24	24	– 15	– 360	225	5 400
27 – 31	29	30	– 10	– 300	100	3 000
32 – 36	34	37	– 5	– 185	25	925
37 – 41	39	45	0	0	0	0
42 – 46	44	25	5	125	25	625
47 – 51	49	10	10	100	100	1 000
52 – 56	54	7	15	105	225	1 575
		$\Sigma f = 190$		$\Sigma fd = -755$		$\Sigma fd^2 = 17\,325$

Note that Σfd^2 is the addition of all values of fd^2 in the fd^2 column of the table and we also got Σf and Σfd in this same way.

(i) Mean $\bar{x} = A + \frac{\Sigma fd}{\Sigma f} = 39 + \frac{-755}{190}$

$= 39 + (-3.97) = 39 - 3.97$

$= 35.03.$

Therefore, the mean of the distribution is 35.03 years.

(ii) Standard deviation,

$$S.D. = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2} = \sqrt{\frac{17325}{190} - \left(\frac{-755}{190}\right)^2}$$

$= \sqrt{91.184 - (-3.974)^2}$

$= \sqrt{91.184 - (-3.974 \times -3.974)};$

$= \sqrt{91.184 - (15.793)}$

$= \sqrt{91.184 - 15.793} = \sqrt{75.391} = 8.683.$

Therefore, the standard deviation of the distribution is 8.68 years.

(b) Probability of an event happening

$$= \frac{number\ of\ elements\ in\ event\ space}{number\ of\ elements\ in\ sample\ space}$$

The workers that are **at most** 36 years are workers that are either 36 years or less

than 36 years. So, number of elements in event space will be the number of workers that are at most 36 years old

$$= \begin{pmatrix} number\ of\ workers\ that \\ are\ age\ 36\ and\ below \end{pmatrix}$$

$= 12 + 24 + 30 + 37 = 103$. Number of elements in sample space is the total number of available workers = 190.

Then, Pr

$$= \begin{pmatrix} that\ the\ worker\ selected\ at\ random \\ from\ the\ company\ of\ 190\ workers\ for \\ the\ award\ is\ \textbf{at most}\ 36\ years\ old \end{pmatrix}$$

$= \frac{103}{190} = 0.542.$

Therefore, the probability that a worker selected at random from the company for an award is at most 36 years old is 0.542.

Correlation and Regression

1. The table shows the marks obtained by ten students in theory and practical tests in Biology. **(WAEC)**

Student	A	B	C	D	E	F	G	H	I	J
Theory	79	63	84	46	77	73	56	58	49	69
Practical	56	42	59	35	54	62	47	51	24	49

(a) Plot the scatter diagram for the data.

(b) Draw an eye-fitted line of best fit.

(c) Using your graph, estimate the mark in theory when the practical mark is 50.

Workshop

Theory	Practical
79	56
63	42
84	59
46	35
77	54
73	62
56	47
58	51
49	24
69	49

(a)

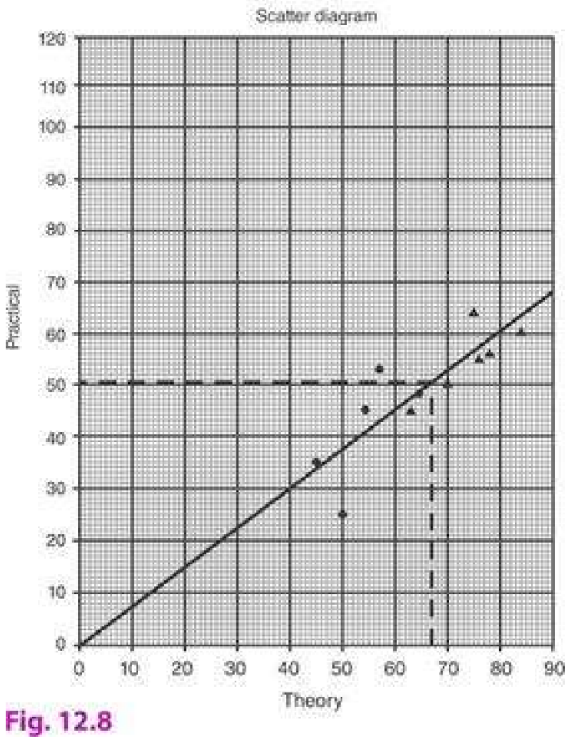

Fig. 12.8

(b) You have to make the eye fitted line of best fit as accurate as possible such that the scattered diagrams are reasonably, well distributed about the line. To achieve this, the line of best fit must trend (be directed) in the general direction of the group of scattered points as shown in figure 12.8.

(c) From the graph the mark in theory when the practical mark is 50 is 68.

2. The table shows the marks obtained by some candidates in Physics (y) and Mathematics (x) tests.

Mathematics	43	46	48	39	30	60	8	45	40
Physics	54	53	63	30	44	75	20	33	49

(a) (i) Represent this information on a scattered diagram;

(ii) Find $\bar{x}$ and $\bar{y}$, the mean of x and y respectively;

(iii) Draw the line of best fit to pass through $(\bar{x}, \bar{y})$.

(b) Find the equation of the line in: (a)(iii).

(c) Use your equation in (b) to find, correct to one decimal place, the mark in Physics for a candidate who scored 28 in Mathematics. **(WAEC)**

Workshop

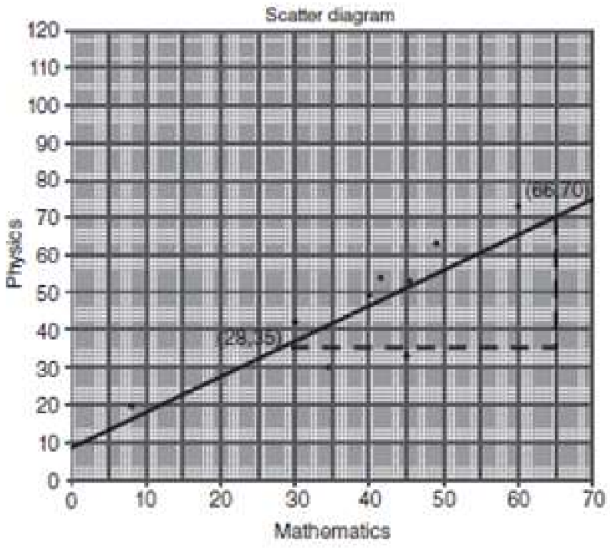

Fig. 12.9

(i) On the scattered diagram (Figure 12.9), the scores in Mathematics are on the Horizontal axis (commonly called x-axis), the mean of x will be the average of the scores in Mathematics. So,

$$\bar{x} = \frac{(43 + 46 + 48 + 39 + 30 + 60 + 8 + 45 + 40)}{9}$$

$$= 39.89.$$

The mean of y will be the average of the scores in Physics, thus,

$$\bar{y} = \frac{(54 + 53 + 63 + 30 + 44 + 75 + 20 + 33 + 49)}{9}$$

$$= 46.78.$$

(iii) The line of best fit as drawn in the diagram **must** pass through point $(\bar{x}, \bar{y})$ = (39.89, 46.78). Also, you must make the eye fitted line of best fit as accurate as possible. To achieve this; the line of best fit must trend (be directed) in the general direction of the group of scattered points as shown in Figure 12.9.

(b) The general equation of a line in the gradient intercept form is given by; $y = mx + c$. Where m is the gradient of the line and c is the intercept of the line on the vertical axis. To know m, we need to know the coordinates of two points on the line. From Figure 12.9, the two points chosen are points (28, 35) and (66, 70). Given two points (x_1, y_1) and (x_2, y_2) on a line, the gradient m, of the line will

be $m = \frac{y_2 - y_1}{x_2 - x_1}$. Therefore, gradient,

m of the graph $= \frac{70 - 35}{66 - 28} = \frac{35}{38} = 0.92$.

Also, from the diagram, the intercept of the line on the vertical axis is $c = 9.0$. Having known m and c, the equation of the line is $y = 0.92x + 9.0$.

(c) The scores in mathematics are the x-values. We want to find y when x is 28. To get this we simply put $x = 28$ in the equation of the line in (b) above.

$y = 0.92(28) + 9.0$; $y = 25.76 + 9.0$;

$y = 34.76 = 34.8$ to 1 decimal place.

Therefore, the mark in physics for a candidate who scored 28 in mathematics is 34.8, correct to 1 decimal place.

3.

Participants	A	B	C	D	E	F	G	H	I	J
Event No. 1	2	6	1	10	3	4	8	7	9	5
Event No. 2	3	8	2	9	4	1	7	5	10	6

(a) Calculate the Spearman's rank correlation coefficient;

(b) Interprete your result. **(WAEC)**

Workshop

Event No. 1 (x)	Ranking (R_x)	Event No. 2 (y)	Ranking (R_y)
10	1	10	1
9	2	9	2
8	3	8	3
7	4	7	4
6	5	6	5
5	6	5	6
4	7	4	7
3	8	3	8
2	9	2	9
1	10	1	10

In the table above, x is arranged in descending order (i.e arranging from the largest, down to the smallest). The biggest event was ranked 1, the following event ranked 2, up till the smallest event that was ranked 10. The same was done for events y. In another table, we will now arrange the events as they were written in the question, and also write the rank of each event in front of it. The rank of each event can be obtained from the previous table, then we calculate for ΣD^2 as below.

X	Y	R_x	R_y	$D = R_x - R_y$	D^2
2	3	9	8	1	1
6	8	5	3	2	4
1	2	10	9	1	1
10	9	1	2	–1	1
3	4	8	7	1	1
4	1	7	10	–3	9
8	7	3	4	–1	1
7	5	4	6	–2	4
9	10	2	1	1	1
5	6	6	5	1	1

$\sum D^2 = 24.$

Then, $\Sigma D^2 = 24$, $N = 10$, where N is the number of data in each event.

(a) The spearman's rank correlation co-efficient, r_k is given by

$$r_k = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6(24)}{10(10^2-1)}$$

$$r_k = 1 - \frac{144}{990} = \frac{846}{990} = 0.85.$$

(b) From the result above, $r_k = 0.85$, and this is close to +1; thus, makes the two sets of data highly positively correlated, which shows a good prediction of one of the sets of data, knowing the other set of data.

4. In a competition, two judges X and Y awarded the following marks to the competitors.

Competitors	A	B	C	D	E	F	G	H	I	J
Marks by X	73	47	50	85	60	78	62	70	80	55
Marks by Y	72	59	52	80	62	75	68	66	82	55

Calculate, correct to 2 decimal places, the spearman's rank correlation coefficient, of the distribution. **(WAEC)**

Workshop

Marks by X	Ranking (R_x)
85	1
80	2
78	3
73	4
70	5
62	6
60	7
55	8
50	9
47	10

Marks by Y	Ranking (R_y)
82	1
80	2
75	3
72	4
68	5
66	6
62	7
59	8
55	9
52	10

The table was formed by arranging the marks awarded by judge X in descending order (starting with the biggest mark down to the smallest). The highest mark was ranked 1, the following mark ranked 2, up till

the smallest mark that was ranked 10. The same was done for marks awarded by judge *Y*. In another table, we will now arrange the marks as they were in the question, we will also write the rank of each mark in front of it. The rank of each mark will be gotten from the previous table, then we calculate for ΣD^2.

Mark by X	Mark by Y	Ranking (R_x)	Ranking (R_y)	$D = R_x - R_y$	D^2
73	72	4	4	0	0
47	59	10	8	2	4
50	52	9	10	−1	1
85	80	1	2	−1	1
60	62	7	7	0	0
78	75	3	3	0	0
62	68	6	5	1	1
70	66	5	6	−1	1
80	82	2	1	1	1
55	55	8	9	−1	1
					$\sum D^2 = 10$

$\Sigma D^2 = 10$, N is the number of competitors,which is 10; $N = 10$.

Spearman's rank correlation coefficient r_k, for this distribution is given by

$$r_s = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6(10)}{10(10^2-1)}$$

$$= 1 - \frac{60}{60 \times 99} = 1 - \frac{60}{990} = \frac{930}{990} = 0.94.$$

r_k = 0.94, which is close to + 1, thus, the two sets of marks, *X* and *Y*, are highly positively correlated. Positive correlation simply means that in general, when mark *X* increases, mark *Y* will also increase, almost linearly (*i.e almost like an increase in x causes an increase in y on a straight line* (*linear*) *graph*). So, the Spearman's rank correlation coefficient of this distribution is 0.94 correct to 2 decimal places.

5. The table shows the marks obtained by ten students in Biology and Physics tests.

Biology	63	53	83	61	57	72	47	65	68	58
Physics	67	45	71	73	47	69	50	70	42	52

Calculate, correct to 2 decimal places, the spearman's rank correlation coefficient. **(WAEC)**

Workshop

Biology Score (X)	Ranking (R_x)	Physics Score (Y)	Ranking (R_y)
83	1	73	1
72	2	71	2
68	3	70	3
65	4	69	4
63	5	67	5
61	6	52	6
58	7	50	7
57	8	47	8
53	9	45	9
47	10	42	10

The marks by x have been arranged in descending order, (starting with the highest mark down to the lowest). The biggest mark was ranked 1, the following mark ranked 2, up till the smallest mark that was ranked 10. The same was done for marks by Y. In the table below, the marks are written as they appear in the question; the rank of each mark is written next to it. The rank of each mark is obtained from the previous table, then ΣD^2 is calculated as shown:

x	y	R_x	R_y	$D = R_x - R_y$	D^2
63	67	5	5	0	0
53	45	9	9	0	0
83	71	1	2	−1	1
61	73	6	1	5	25
57	47	8	8	0	0
72	69	2	4	−2	4
47	50	10	7	3	9
65	70	4	3	1	1
68	42	3	10	−7	49
58	52	7	6	1	1

$\Sigma D^2 = 90$

$N = 10$ (*where N is the number of marks in each of events x and y*) and $\Sigma D^2 = 90$.
The Spearman's rank correlation coefficient, r_k, is expressed as

$$r_k = 1 - \frac{6\Sigma D^2}{N(N^2-1)} = 1 - \frac{6(90)}{10(10^2-1)}$$

$$= 1 - \frac{540}{10(99)} = 1 - \frac{540}{990}$$

$$= 1 - \frac{54}{99} = \frac{45}{99} = 0.45$$

correct to two decimal places.