ELEMENTARY STATISTICS

Summary:

1. *The following terms are computed as follows:*

For a set of ***n*** *values* ***x***	*For frequency distribution table,*
Mean $\bar{x} = \frac{\sum x}{n}$	*Mean* $\bar{x} = \frac{\sum fx}{\sum f}$
Variance $= \frac{\sum (x - \bar{x})^2}{n}$ $= \frac{\sum x^2}{n} - \bar{x}^2$	*Variance* $= \frac{\sum f(x - \bar{x})^2}{\sum f}$ $= \frac{\sum fx^2}{\sum f} - \bar{x}^2$
Standard deviation $= \sqrt{\text{variance}}$	*Standard deviation* $= \sqrt{\text{variance}}$

NOTE: *Graphically,* $SD = \frac{1}{2} \times$ *range of the middle* ***68%*** *or* $\frac{2}{3}$ *of the data*

2 (i) *Mode is the mostly occuring value.There can be more than one mode in a given data.*

(ii) *In case of grouped data with equal class width, the modal class is the one with the highest frequency.*

(iii) *In case of grouped data with unequal class widths, the modal class is the one with the highest frequency density,*

$$\left(\textbf{Frequency density} = \frac{\text{class frequency}}{\text{class width}} \right)$$

(iv) *For grouped data,* $\textbf{mode} = L + \left(\frac{D_1}{D_1 + D_2} \right) c$

In case of equal class width
L = *lower boundary of the modal class* D_1 ***=*** *modal frequency – premodal frequency* D_2 ***=*** *modal frequency – post modal frequency* ***C =*** *modal class width*
In case of unequal class widths
L = *lower boundary of the modal class* D_1 ***=*** *modal frequency density – premodal frequency density* D_2 ***=*** *modal frequency density – post modal frequency density* ***C =*** *modal class width*

NOTE: *The modal class must be determined before calculating the mode*

3 (i) *Median is the middle value when the given data is listed in order of magnitude. If the number of items is even, the average of the middle two is used.*

(ii) *For grouped data,* $\textbf{median} = L + \left(\frac{\frac{1}{2}\sum f - Cf_b}{f_w} \right) c$ *, where*

L = *lower boundary of the median class*

$\sum f$ = *total frequency*

Cf_b = *cumulative frequency before the median class*

f_w = *frequency within the median class*

C = *median class width*

NOTE:

(a) *The median class must be determined before calculating the median*

(b) *The median class corresponds to the cumulative frequency of* $\frac{1}{2}\sum f$

4. (i) *The lower quartile* Q_1 *is the value for which* ***25%*** *of the data falls below it.*

(ii) *The upper quartile* Q_3 *is the value for which* ***75%*** *of the data falls below it.*

(iii) *Interquartile range* $= Q_3 - Q_1$

(iv) *Semi–interquartile range* $= \frac{1}{2}\left(Q_3 - Q_1\right)$

(v) *The* J^{th} *percentile* P_J *is the value for which* ***J%*** *of the data falls below it.*

(vi) *The* J^{th} *decile* D_J *is the value for which* $\frac{J}{10}$ *of the data falls below it.*

(vii) *For grouped data, the following terms are computed as follows:*

Lower quartile $Q_1 = L + \left(\frac{\frac{1}{4}\sum f - Cf_b}{f_w}\right)c$
Upper quartile $Q_3 = L + \left(\frac{\frac{3}{4}\sum f - Cf_b}{f_w}\right)c$

The J^{th} percentile $P_J = L + \left(\frac{\frac{J}{100}\sum f - Cf_b}{f_w} \right) c$
The J^{th} decile $D_J = L + \left(\frac{\frac{J}{10}\sum f - Cf_b}{f_w} \right) c$

NOTE:

(a) *The quartile class must be determined before calculating the required quartile*

(b) *The lower and upper quartile classes correspond to the cumulative frequencies*

of $\frac{1}{4}\sum f$ *and* $\frac{3}{4}\sum f$ *respectively.*

(c) *The* J^{th} *percentile and decile classes correspond to the cumulative frequencies*

of $\frac{J}{100}\sum f$ *and* $\frac{J}{10}\sum f$ *respectively.*

5 (i) *The cumulative frequency curve* ***or*** *an ogive is a curve where cumulative frequencies are plotted against the upper class boundaries.*

(ii) *A histogram with equal class widths consists of bars with frequency as the height of each and class boundaries as the width of each.*

(iii) *A histogram with unequal class widths consists of bars with frequency density as the height of each and class boundaries as the width of each.*

(iv) *A frequency polygon is a line graph drawn by plotting frequency* ***or*** *frequency density against class mid values.*

NOTE:

(a) *The points are joined by straight lines.*
(bi) *The polygon extends to the next lower and higher classes with zero frequencies*

EXAMPLES:

1. *The lengths in cm of* ***6*** *metal rods were as follows:* ***31, 28, 30, 33, 25, 30***
Find the:

(i) mean length
(ii) standard deviation
(iii) mode
(iv) median
(v) semi–interquartile range

Solution:

							Sum
x	*31*	*28*	*30*	*33*	*25*	*30*	***177***
x^2	*961*	*784*	*900*	*1089*	*625*	*900*	***5259***

(i) *mean* $\bar{x} = \frac{177}{6}$

$= 29{\cdot}5$

(ii) *Standard deviation* $\sigma = \sqrt{\frac{5259}{6} - \left(\frac{177}{6}\right)^2}$

$= 2{\cdot}5$

(iii) *mode =* ***30***

(iv) ~~25~~, ~~28~~, 30, 30, ~~31~~, ~~33~~

$\text{median} = \frac{30 + 30}{2}$

$= 30$

(v) 25, (28), 30, 30, (31), 33

Q_1 media Q_3

$$\text{Semi–interquartile range} = \frac{31 - 28}{2}$$

$$= 1{\cdot}5$$

2. *The lengths* ***a, b, 8, 7,5*** *in cm of five metal rods are distributed with a mean of* ***6*** *and variance of* ***2****. Given that* ***a > b****, find the values of* ***a*** *and* ***b****.*

Solution:

						Sum
x	*A*	*B*	*8*	*7*	*5*	***a + b + 20***
x^2	a^2	b^2	*64*	*49*	*25*	$a^2 + b^2 +$ ***138***

If $\frac{a + b + 20}{5} = 6$

$\Rightarrow$

$a + b = 10$ ----------(i)

If $\frac{a^2 + b^2 + 138}{5} - 6^2 = 2$

$\Rightarrow a^2 + b^2 = 52$ ----------(ii)

If $b = 10 - a$

$\Rightarrow a^2 + (10 - a)^2 = 52$

$a^2 - 10a + 24 = 0$

$a = \frac{10 \pm \sqrt{100 - 96}}{2}$

$a = 4$ *or* 6

$b = 10 - a$

a	*4*	*6*
b	*6*	*4*

∴ a = 6, b= 4

3. *The ages in years of* ***100*** *students were as follows:*

Age	***12***	***1***	***16***	***18***	***20***	***22***

		4				
Frequency	15	25	18	22	12	8

Find the:

(i) mean age
(ii) standard deviation
(iii) mode
(iv) median
(v) semi–interquartile range

Solution:

Age (x)	f	fx	fx^2	Cf
12	15	180	2160	15
14	25	350	4900	40
16	18	288	4608	58
18	22	396	7128	80
20	12	240	4800	92
22	8	176	3872	100
Sum	**100**	**1630**	**27468**	

(i) mean $\bar{x} = \frac{1630}{100}$

= 16·3

(ii) Standard deviation $\sigma = \sqrt{\frac{27468}{100} - \left(\frac{1630}{100}\right)^2}$

= 2·9983

(iii) mode = **14** **(ie 14** *occurs the most* **)**

(iv) median position $= \frac{100}{2} = 50$

.: median **= 16**

(v) *lower quartile position* $= \frac{100}{4} = 25$

$\therefore$ *lower quartile* ***= 14***

upper quartile position $= \frac{3}{4} \times 100 = 75$

$\therefore$ *upper quartile* ***= 18***

Semi–interquartile range $= \frac{18 - 14}{2}$

$= 2$

4. *A certain frequency distribution with standard deviation* ***2·5*** *has the following results:* $\sum fx = 177$ *and* $\sum fx^2 = 5259$. *Find its:*

(i) $\sum f$

(ii) *mean*

Soln:

(i) *If* $\sqrt{\frac{5259}{n} - (\frac{177}{n})^2} = 2 \cdot 5$

$\Rightarrow \frac{5259}{n} - \frac{31329}{n^2} = 6 \cdot 25$

$6 \cdot 25n^2 - 5259n + 31329 = 0$

$n = \frac{5259 \pm \sqrt{(-5259)^2 - 783225}}{12 \cdot 5}$

$n = 6$ *or* $835 \cdot 44$

$\therefore \sum f = 6$

(ii) *mean* $= \frac{177}{6} = 29 \cdot 5$

5. *The prices in thousands of shillings of* ***50*** *bags were as follows:*

4·2	3·1	2·8	4·0	2·3	3·7	3·3	4·4	2·5	3·0
3·6	4·3	3·2	2·4	4·1	3·4	2·7	4·2	4·8	2·6
2·2	3·0	4·1	4·6	3·7	2·9	4·3	2·0	3·2	4·0
4·7	2·6	3·8	2·3	4·0	3·3	2·7	4·5	2·4	3·6
2·0	3·5	2·7	3·2	2·1	4·2	3·0	4·1	2·8	4·7

(a) *Starting with the lowest class limit of* ***2·0*** *thousand shillings, form a frequency*
distribution table with class intervals of ***0·5*** *thousand shillings.*

(b) *Calculate the:*

(i) mean price (ii) variance (iii) mode
(iv) median (v) semi–interquartile range
(vi) 60^{th} *percentile (vii)* 4^{th} *decile*
(viii) number of bags cheaper than ***3·3*** *thousand shillings.*

(c) *Display the data on a histogram and use it to estimate the:*
(i) mode
(ii) median

(c) *Draw an ogive for the data and use it to estimate the:*
(i) median
(ii) semi–interquartile range
(iii) 40^{th} *to* 90^{th} *percentile price range*
(iv) range of the prices of the middle ***60%*** *of the bags*
(v) 3^{rd} *to* 6^{th} *decile price range.*
(vi) standard deviation

Solution:

Price	***Tallys***	*f*	*x*	*fx*	fx^2	***Cf***
2·0 – 2·4	𝍸 ///	*8*	*2·2*	*17·6*	*38·72*	*8*
2·5 – 2·9	𝍸 ////	*9*	*2·7*	*24·3*	*65·61*	*17*
3·0 – 3·4	𝍸 𝍸	*10*	*3·2*	*32*	*102·4*	*27*
3·5 – 3·9	𝍸 /	*6*	*3·7*	*22·2*	*82·14*	*33*
4·0 – 4·4	𝍸 𝍸 //	*12*	*4·2*	*50·4*	*211·68*	*45*
4·5 – 4·9	𝍸	*5*	*4·7*	*23·5*	*110·45*	*50*
Sum		***50***		***170***	***611***	

(b) (i) $\text{mean} = \frac{170}{50}$

$= 3{\cdot}4$

(ii) $\text{Variance} = \frac{611}{50} - \left(\frac{170}{50}\right)^2$

$= 0{\cdot}66$

(iii) $\text{mode} = 3 \cdot 95 + \left(\frac{6}{6 + 7}\right) \times 0 \cdot 5$

$= 4{\cdot}1808$

(iv) $\text{median position} = \frac{50}{2} = 25$

$$\text{median} = 2 \cdot 95 + \left(\frac{\frac{50}{2} - 17}{10}\right) \times 0 \cdot 5$$

$\therefore$ median $= 3{\cdot}35$

(v) $\text{lower quartile position} = \frac{50}{4} = 12 \cdot 5$

$$\text{lower quartile, } Q_1 = 2 \cdot 45 + \left(\frac{\frac{50}{4} - 8}{9}\right) \times 0 \cdot 5$$

$= 2{\cdot}7$

$\text{upper quartile position} = \frac{3}{4} \times 50 = 37 \cdot 5$

$$\text{upper quartile, } Q_3 = 3 \cdot 95 + \left(\frac{\frac{3}{4} \times 50 - 8}{9}\right) \times 0 \cdot 5$$

$= 4{\cdot}1375$

$$\therefore \text{Semi–interquartile range} = \frac{4 \cdot 1375 - 2 \cdot 7}{2}$$

$$= 0 \cdot 71875$$

(vi) 60^{th} percentile position $= \frac{60}{100} \times 50 = 30$

$$60^{th} \text{ percentile, } P_{60} = 3 \cdot 45 + \left(\frac{\frac{60}{100} \times 50 - 27}{6} \right) \times 0 \cdot 5$$

$$= 3 \cdot 7$$

(vii) 4^{th} decile position $= \frac{4}{10} \times 50 = 20$

$$4^{th} \text{ decile, } D_4 = 2 \cdot 95 + \left(\frac{\frac{4}{10} \times 50 - 17}{10} \right) \times 0 \cdot 5$$

$$= 3 \cdot 1$$

(viii)

Price	2·95	3·3	3·45
Cf	17	x	27

$$\frac{x - 17}{3 \cdot 3 - 2 \cdot 95} = \frac{27 - 17}{3 \cdot 45 - 2 \cdot 95}$$

$$x = 24$$

6. The weights in kg of **70** students were as follows:

Weights	40 – 44	45 – 49	50 – 54	55 – 59	60 – 64	65 – 69	70 – 74
Frequency	5	8	16	12	14	9	6

(a) *Calculate the:*

(i) mean weight

(ii) standard deviation

(iii) number of students whose weights exceed the mean weight

(b) *(i) Display the data on a histogram and superimpose a frequency polygon.*

(ii) Use your histogram to estimate the mode

(c) *Draw an ogive for the data and use it to estimate the:*

(i) number of students weighing below ***51kg***

(ii) percentage of students weighing above ***61kg***

(iii) probability that the weight of a student picked at random lies between

62·5kg *and* ***68·5kg.***

(iv) number of students whose weights lie within one standard deviation of

the mean

(v) weight exceeded by ***20%*** *of the students.*

(vi) standard deviation

Solution:

Weight	*f*	*x*	*fx*	fx^2	***Cf***
40 – 44	*5*	*42*	*210*	*8820*	*5*
45 – 49	*8*	*47*	*376*	*17672*	*13*
50 – 54	*16*	*52*	*832*	*43264*	*29*
55 – 59	*12*	*57*	*684*	*38988*	*41*
60 – 64	*14*	*62*	*868*	*53816*	*55*
65 – 69	*9*	*67*	*603*	*40401*	*64*
70 – 74	*6*	*72*	*432*	*31104*	*70*
Sum	***70***		***4005***	***234065***	

(a) (i) $mean = \frac{4005}{70}$

$= 57{\cdot}2143$

(ii) $SD = \sqrt{\frac{234065}{70} - \left(\frac{4005}{70}\right)^2}$

$= 8{\cdot}3852$

(iii)

weight	*54·5*	*57·2143*	*59·5*
Cf	*41*	***x***	*29*

$$\frac{x - 41}{57 \cdot 2143 - 54 \cdot 5} = \frac{29 - 41}{59 \cdot 5 - 54 \cdot 5}$$

$x = 34{\cdot}48568$

$\Rightarrow$ *Required No = 35*

7. *The heights in cm of* ***100*** *plants were as follows:*

Heights	*30 – 39*	*40 – 49*	*50 – 54*	*55 – 69*	*70 – 79*	*80 – 84*
Frequency	*11*	*22*	*15*	*27*	*12*	*13*

(a) *Calculate the:*

(i) mode

(ii) median

(iii) number of plants shorter than ***71·5cm***

(b) *Display the data on a histogram and use it to estimate the mode*

(c) *Plot a cumulative frequency curve for the data and use it to estimate the:*

(i) median

(ii) range of the heights of the middle ***50%*** *of the plants.*

(iii) shortest height for sale if the top ***10%*** *of the plants qualify for sale.*

Solution:

Height	*f*	*c*	*fd*	***cf***

30 – 39	11	10	1·1	11
40 – 49	22	10	2·2	33
50 – 54	15	5	3	48
55 – 69	27	15	1·8	75
70 – 79	12	10	1·2	87
80 – 84	13	5	2·6	100
Sum	**100**			

(a) (i) $mode = 49 \cdot 5 + \left(\frac{0 \cdot 8}{0 \cdot 8 + 1 \cdot 2} \right) \times 5$

$= \mathbf{51 \cdot 5}$

(ii) $median\ position = \frac{100}{2} = 50$

$$median = 54 \cdot 5 + \left(\frac{\frac{100}{2} - 48}{27} \right) \times 15$$

$\therefore median = \mathbf{55 \cdot 6111}$

(iii)

Height	69·5	71·5	79·5
Cf	75	**x**	87

$$\frac{x - 75}{71 \cdot 5 - 69 \cdot 5} = \frac{87 - 75}{79 \cdot 5 - 69 \cdot 5}$$

$x = 77 \cdot 4$

$\therefore \mathbf{x = 78}$

8. *The marks of* **160** *students in a test were as follows:*

Marks	30 –	40 –	50 –	60 –	70 –	80 –
Frequency	5	42	61	37	15	0

(a) *Calculate the:*

(i) standard deviation of the distribution

(ii) mode

*(iii) number of students who scored above **68** marks*

***(b)** Display the data on a histogram and use it to estimate the mode*

***(c)** Plot a cumulative frequency curve for the data and use it to estimate the:*

*(i) range of the marks of the middle **70%** of the students.*

(ii) standard deviation

*(iii) lowest mark for a distinction one if the top **5%** of the students qualify for*

distinction.

Solution:

Marks	f	x	fx	fx^2	**Cf**
30 – 40	5	35	175	6125	5
40 – 50	42	45	1890	85050	47
50 – 60	61	55	3355	184525	108
60 – 70	37	65	2405	156325	145
70 – 80	15	75	1125	84375	160
Sum	**160**		**8950**	**516400**	

(a) (i) $SD = \sqrt{\frac{516400}{160} - \left(\frac{8950}{160}\right)^2}$

$= 9{\cdot}9245$

(ii) $mode = 50 + \left(\frac{19}{19 + 24}\right) \times 10$

$= 54{\cdot}4186$

(iii)

Marks	60	68	70
Cf	52	**x**	15

$$\frac{x - 52}{68 - 60} = \frac{15 - 52}{70 - 60}$$

$x = 22{\cdot}4$

$\Rightarrow$ *Required No = 23*

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

Lengths	***Frequency***
30 and under 35	*8*
35 and under 40	*5*
40 and under 55	*12*
55 and under 60	*9*
60 and under 65	*6*

(a) *Calculate the:*

(i) mean length

(ii) modal length

(iii) number of metal rods shorter than ***45cm.***

(b) *(i) Display the data on a histogram and superimpose a frequency polygon.*

(ii) Use your histogram to estimate the mode

(c) *Draw an ogive for the data and use it to estimate the:*

(i) median.

(ii) semi–interquartile range

(iii) 40^{th} *to* 60^{th} *percentile length range*

Solution:

Lengths	*f*	*x*	*fx*	*c*	*fd*	***Cf***
30 – < 35	*8*	*32·5*	*260*	*5*	*1·6*	*8*
35 – < 40	*5*	*37·5*	*187·5*	*5*	*1*	*13*

40 – < 55	12	47·5	570	15	0·8	25
55 – < 60	9	57·5	517·5	5	1·8	34
60 – < 65	6	62·5	375	5	1·2	40
Sum	**40**		**1910**			

(a) (i) $mean = \frac{1910}{40}$

$= 47{\cdot}75$

(ii) $mode = 55 + \left(\frac{1}{1 + 0{\cdot}6}\right) \times 5$

$= 58{\cdot}125$

(iii)

Lengths	40	45	55
Cf	13	**x**	25

$$\frac{x - 13}{45 - 40} = \frac{25 - 13}{55 - 40}$$

$$x = 17$$

10. *The lengths in inches of* **40** *nails were as follows:*

Lengths (h)	frequency
$3{\cdot}0 \leq h < 3{\cdot}5$	8
$3{\cdot}5 \leq h < 4{\cdot}0$	5
$4{\cdot}0 \leq h < 5{\cdot}5$	12
$5{\cdot}5 \leq h < 6{\cdot}0$	9
$6{\cdot}0 \leq h < 6{\cdot}5$	6

(a) *Display the data on a histogram and use it to estimate the mode*

(b) *Calculate the:*

(i) standard deviation of the distribution

(ii) number of nails shorter than **4·5** *inches.*

(c) *Draw an ogive for the data and use it to estimate the:*

*(i) length exceeded by **75%** of the nails.*

(ii) number of nails whose lengths lie within one standard deviation of the mean.

Solution:

Length	f	c	fd	x	fx	fx^2	cf
3·0 – 3·5	8	0·5	16	3·25	26	84·5	8
3·5 – 4·0	5	0·5	10	3·75	18·75	70·3125	13
4·0 – 5·5	12	1·5	8	4·75	57	270·75	25
5·5 – 6·0	9	0·5	18	5·75	51·75	297·5625	34
6·0 – 6·5	6	0·5	12	6·25	37·5	234·375	40
Sum	**40**				**191**	**957·5**	

(b) (i) $SD = \sqrt{\frac{957 \cdot 5}{40} - \left(\frac{191}{40}\right)^2}$

= 1·0662

(ii)

Length	4·0	4·5	5·5
Cf	13	**x**	25

$$\frac{x - 13}{4 \cdot 5 - 4 \cdot 0} = \frac{25 - 13}{5 \cdot 5 - 4 \cdot 0}$$

x = 17

11. *The cumulative distribution of the ages in years of **98** employees were as follows:*

Age (years)	**< 15**	**< 20**	**< 30**	**< 40**	**< 50**	**< 60**	**< 65**
Cumulative Frequency	**0**	**17**	**39**	**69**	**87**	**92**	**98**

***(a)** Draw an ogive for the data and use it to estimate the semi-interquartile range*

(b) *Calculate the mean age of the employees.*

Solution:

(b)

Ages	***Cf***	***f***	***x***	***fx***
15 – < 20	*17*	*17*	*17·5*	*297·5*
20 – < 30	*39*	*22*	*25*	*550*
30 – < 40	*69*	*30*	*35*	*1050*
40 – < 50	*87*	*18*	*45*	*810*
50 – < 60	*92*	*5*	*55*	*275*
60 – < 65	*98*	*6*	*62·5*	*375*
Sum		***98***		***3357·5***

$$mean = \frac{3357 \cdot 5}{98}$$

$$= 34{\cdot}2602$$

12. *The height distribution of* ***50*** *players in a certain Football club were:*

Height (cm)	***> 160***	***> 164***	***> 168***	***> 172***	***> 176***	***> 180***
No of players	***50***	***45***	***37***	***23***	***7***	***0***

Form a frequency table for the above data starting with a class of ***160 – < 164.***

Hence calculate the mean height and standard deviation.

Solution:

Ages	***No of players***	***f***	***x***	***fx***	fx^2

160 – < 164	50	5	162	810	131220
164 – < 168	45	8	166	1328	220448
168 – < 172	37	14	170	2380	404600
172 – < 176	23	16	174	2784	484416
176 – < 180	7	7	178	1246	221788
Sum		**50**		**8548**	**1462472**

$$mean = \frac{8548}{50}$$

= 170·96

$$SD = \sqrt{\frac{1462472}{50} - \left(\frac{8548}{50}\right)^2}$$

= 4·70302

EER:

1. *The lengths in cm of* ***5*** *metal rods were as follows:* ***8, 5, 2, 6, 4***
Find the:

(i) *mean length*
(ii) *standard deviation*

[Ans: (i) 5 (ii) 2]

2. *A certain frequency distribution with standard deviation* ***7*** *has the following results:* $\sum fx = 288$ *and* $\sum fx^2 = 7500$. *Find the total number of items that were considered*

[Ans: 12]

3. *Two samples of size* ***100*** *and* ***150*** *respectively have mean* ***50*** *and* ***60*** *and standard deviation* ***5*** *and* ***6.*** *Calculate the combined mean and standard deviation*

[Ans: 56, 7·4565]

4. *The weights in kg of* ***8*** *boxes were as follows:* ***11, 5, 2, 3, 11, 5, 8, 11***
Find the:

(i) *mean weight*
(ii) *standard deviation*
(iii) *semi–interquartile range*

[Ans: (i) 7 (ii) 3·5 (iii) 3·5]

5. *A certain frequency distribution with standard deviation* ***7*** *has the following results:* $\sum fx = 177$ *and* $\sum fx^2 = 5259$. *Find its* $\sum f$

[Ans: 6]

6. *A random sample of* ***6*** *books had the following masses in grams:*

31, 28, 30, 33, 25, 30

Find the sample:

(i) *mean mass*
(ii) *standard deviation*

[Ans: (i) 29·5 (ii) 2·5]

7. *The table below shows the mass of boys in a certain school*

Mass (kg)	***15***	***20***	***25***	***30***	***35***
Number of boys	***5***	***6***	***10***	***20***	***9***

Calculate the mean mass

[Ans: 27·2]

8. *The number of children in* ***30*** *families were as follows:*

No of children per family	***1***	***2***	***3***	***4***	***5***
No of families	***10***	***x***	***4***	***y***	***1***

Given that the average number of children per family is ***2·1,*** *find the values of* ***x*** *and* ***y***

[Ans: x = 12, y = 3]

9. *The weights in kg of* ***80*** *boys were as follows:*

Weights	***50 – 59***	***60 – 64***	***65 – 69***	***70 – 79***
Frequency	***18***	***20***	***x***	***y***

Given that ***60%*** *of the boys have their weights below* ***66·5kg,*** *find the values of* ***x*** *and* ***y***

[Ans: x = 15, y = 17]

10. *The heights in cm of* ***80*** *plants were as follows:*

Heights	***0 – 9***	***10 – 19***	***20 – 29***	***30 – 39***	***40 – 49***	***50 – 59***
Frequency	***7***	***13***	***11***	***16***	***18***	***15***

Draw an ogive for the data and use it to estimate the:

(i) *median height*

(ii) *interquartile range*

(iii) 10^{th} *to* 90^{th} *percentile range*

(iv) *range of the heights of the middle* ***60%*** *of the plants.*

(v) 4^{th} *decile*

[Ans: (i) 35·5 (ii) 27 (iii) 44 (iv) 37·5 (v) 30]

11. *The weights in kg of* ***50*** *students were as follows:*

Weights	*20 – 24*	*25 – 29*	*30 – 34*	*35 – 39*	*40 – 44*	*45 – 49*
Frequency	*8*	*9*	*10*	*6*	*12*	*5*

(a) *Calculate the:*

(i) *variance*

(ii) *40 th percentile*

(b) *Display the data on a histogram and use it to estimate the:*

(i) *modal weight*

(ii) *median weight*

(c) *Display the data on an ogive and use it to estimate the standard deviation*

[Ans: a (i) 66 (ii) 31 b (i) 42 (ii) 33·5 (c) 8·75]

12. *The weights in kg for a group of animals were as follows:*

Weight (kg)	*Frequency*
10 – 19	*20*
20 – 24	*20*
25 – 29	*15*
30	*14*
31 – 34	*16*
35 – 39	*10*
40 – 59	*10*

(a) *Calculate the:*

(i) *mean*

(ii) *standard deviation of the distribution*

(b) *Draw a histogram for the data and use it to estimate the modal weight*

[Ans: a(i) 28 (ii) 9·8525 (b) 30]

13. *The lengths in inches of* ***40*** *nails were as follows:*

Lengths	***3·0–3·5***	***3·5–4·0***	***4·0–5·5***	***5·5–6·0***	***6·0–6·5***
Frequency Density	***16***	***10***	***8***	***18***	***12***

(a)
Display the data on a histogram and use it to estimate the mode

(b) *Calculate the:*

(i) *standard deviation of the distribution*

(ii) *number of nails shorter than* ***4·5*** *inches.*

[Ans: (a) 5·8 (b) (i)1·0662 (ii) 17]

14. *The table below shows the weights in kg of* ***36*** *boys*

Weights	***60 –***	***65 –***	***70 –***	***75 –***	***80 –***	***85 – 90***
Frequency	***6***	***10***	***8***	***6***	***2***	***4***

(a) *Calculate the:*

(i) *mean*

(ii) *mode*

(b) *Draw an ogive for the data and use it to estimate the:*

(i) *median*

(ii) *range of the weights of the middle* ***50%*** *of the boys*

(iii) *standard deviation*

[Ans: a(i) 72·5 (ii) 68·3333 b(i) (i) 71·5 (ii) 9·5 (iii) 7·5]

15. *The table below shows the marks scored by students in a certain test*

Marks	***< 20***	***< 30***	***< 35***	***< 40***	***< 50***	***< 60***
Cumulative Frequency	***4***	***20***	***32***	***42***	***48***	***50***

(a) *Construct a frequency distribution table*

(b) *Draw a histogram for the data and use it to estimate the modal mark*
(c) *Calculate the mean mark*

[Ans: (b) 33 (c) 31·7]

16. *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 mean and mode of the distribution*
(b) *Display the data on a histogram and use it to estimate the mode*

[Ans: (a) 47·75, 58·125 (b) 58]

17. *The table below shows the marks of* ***50*** *boys in a certain test*

Marks	*< 20*	*< 30*	*< 40*	*< 50*	*< 60*	*< 70*
No of boys	*0*	*8*	*20*	*34*	*44*	*50*

(a) *Calculate the:*

(i) *variance for the distribution.*
(ii) *number of boys weighing below* ***48kg***

(b) (i) *Display the data on a histogram and superimpose a frequency polygon.*

(ii) *Use your histogram to estimate the mode*

[Ans:(a) (i) 154·56 (ii) 32 (b)(ii) 46·7]

18. *The marks of* ***100*** *students were as follows:*

Marks	*Frequency*
Below ***5***	*6*
5 *and under* ***10***	*11*
10 *and under* ***15***	*17*
15 *and under* ***25***	*28*
25 *and under* ***30***	*20*
30 *and under* ***35***	*15*
35 *and under* ***40***	*3*

(a) *Calculate the:*

(i) *upper quartile mark*

(ii) *variance of the distribution*

(b) *Display the data on a histogram and use it to estimate the mode*

[Ans: (a) (i) 28·25 (ii) 88·96 (b) 27·5]

19. *The heights in* ***cm*** *of* ***100*** *recruits were as follows:*

Height (cm)	*No of recruits*
148 – 153	*6*
153 – 158	*11*
158 – 163	*17*
163 – 173	*28*
173 – 178	*20*
178 – 183	*15*
183 – 188	*3*

(a) *Calculate the:*

(i) *mode*

(ii) *variance of the distribution*

(b) *Display the data on an ogive and use it to estimate the:*

(i) *semi–interquartile range*

(ii) *height exceeded by* ***15%*** *of the recruits.*

(iii) *range of the heights of the middle* ***20%*** *of the recruits.*

[Ans: a(i) 175·7273 (ii) 88·96 b(i) 8 (ii) 179 (iii) 7·25]

20. *The lifetimes in hours of* ***80*** *solar bulbs were as follows:*

Lifetimes	*No of bulbs*
$0 \leq T < 10$	*14*
$10 \leq T < 20$	*19*
$20 \leq T < 30$	*15*
$30 \leq T < 40$	*20*
$40 \leq T < 50$	*12*

(a) *Calculate the variance of the distribution*

(b) *Draw an ogive for the data and use it to estimate the lifetime*

exceeded
*by **75%** of the bulbs.*

[Ans: (a) 178·6094 (b) 13]

***21.** The marks of **40** boys were as follows:*

Marks	***> 30***	***> 35***	***> 40***	***> 55***	***> 60***	***> 65***
No of boys	***40***	***32***	***27***	***15***	***6***	***0***

***(a)** Calculate the:*
(i) mean
(ii) mode

***(b)** Draw an ogive for the data and use it to estimate the:*
(i) median
*(ii) range of the marks of the middle **60%** of the students.*

[Ans: a(i) 47·75 (ii) 58·125 b(i) 50 (ii) 24]

***22.** The marks of **160** students in a test were as follows:*

Marks	***30 –***	***40 –***	***50 –***	***60 –***	***70 –***	***80 –***
Frequency	***5***	***42***	***61***	***37***	***15***	***0***

***(a)** Calculate the:*
(i) mean
(ii) standard deviation of the distribution
(iii) mode
(iv) median
(v) lower quartile
(vi) upper quartile
(vii) 80^{th} percentile
(viii) 4^{th} decile
*(ix) number of students who scored above **68** marks*

***(b)** Display the data on a histogram and use it to estimate the mode*

(c) *Plot a cumulative frequency curve for the data and use it to estimate the:*

(i) *range of the marks of the middle* ***70%*** *of the students.*

(ii) *lowest mark for a distinction one if the top* ***5%*** *of the students qualify for distinction.*

(iii) *number of students whose marks lie within one standard deviation of the mean.*

[Ans: (a) (i) 55·9375 (ii) 9·9245 (iii) 54·4186 (iv) 55·4098 (v) 48·3333 (vi) 63·2432 (vii) 65·4054 (viii) 52·7869 (ix) 23 (b)(i) 21 (ii) 74 (iii) 106]

23. *The weights in kg for a group of animals were as follows:*

Weight (kg)	*Frequency*
21 – 25	*10*
26 – 30	*20*
31 – 35	*15*
36 – 40	*10*
41 – 50	*30*
51 – 65	*45*
66 – 75	*5*

(a) *Calculate the:*

(i) *mode*

(ii) *standard deviation of the distribution*

(b) *Draw a cumulative frequency curve for the data and use it to estimate the semi–interquartile range*

[Ans: a(i) 28·8333 (ii) 13·3565 (b) 12]

24. *The allowances in thousands of shillings to participants during a workshop were as follows:*

Amount (Sh'000s)	*No of participants*
110 – 114	*13*

115 – 119	20
120 – 129	32
130 – 134	17
135 – 144	16
145 – 159	12

(a) *Calculate the mean and median allowance.*

(b) *Display the data on a histogram and use it to estimate the modal allowance.*

[Ans: (a) 128, 126·375 (b) 117·5]

25. *The marks of* **40** *students were as follows:*

26	11	10	12	14	16	20	25
21	22	13	17	18	27	30	32
27	35	40	44	39	28	37	26
44	37	36	39	28	46	32	15
16	19	34	43	26	38	48	40

(a) *Form a frequency distribution table with a lower class of* **10 – 14.**

(b) (i) *Display the data on a histogram and superimpose a frequency polygon*

(ii) *Use your histogram to estimate the modal mark.*

(c) *Calculate the:*

(i) *mean*

(ii) *standard deviation*

(iii) *semi-interquartile range*

[Ans: b(i) 27·5 c (i) 28·375 (ii) 10·5468 (iii) 9·3452]

26. *The ages in years of* **40** *students were as follows:*

12	13	14	12	15	14	13	16	14	15
13	14	16	15	14	12	13	14	15	13
15	16	15	14	15	12	15	13	12	15
13	15	12	15	16	14	15	14	16	14

(a) *Form an ungrouped frequency distribution table for the data*

(b) *Find the:*

(i) mode

(ii) median
(iii) standard deviation
(iv) semi–interquartile range

[Ans: b(i) 15 (ii) 14 (iii) 14·075 (iv) 1·2527]

27. *The expenditure on* ***40*** *students in thousands of shillings were as follows:*

10	***11***	***10***	***12***	***14***	***16***	***20***	***25***
21	***22***	***13***	***17***	***18***	***24***	***30***	***32***
27	***35***	***27***	***30***	***16***	***26***	***25***	***29***
32	***37***	***36***	***17***	***26***	***28***	***22***	***15***
16	***19***	***34***	***33***	***38***	***26***	***27***	***34***

(a) *Form a frequency distribution table with class intervals of* ***5000*** *shillings and*
the lowest class limit being ***10,000***
(b) (i) *Display the data on a histogram and superimpose a frequency polygon*
(ii) *Use your histogram to estimate the modal expenditure.*
(c) *Calculate the:*
(i) *mean expenditure*
(ii) *standard deviation*

[Ans: b(ii) 27·5 (c) (i) 24 (ii) 3·9370]

28. *The prices in thousands of shillings for* ***40*** *bags were as follow:*

3·6	***5·2***	***3·8***	***4·3***	***5·3***	***3·1***	***3·8***	***4·0***	***5·4***	***4·6***
3·7	***3·0***	***4·2***	***3·7***	***3·2***	***4·8***	***4·6***	***3·5***	***4·1***	***3·9***
5·3	***4·8***	***4·4***	***5·0***	***3·3***	***4·2***	***3·6***	***5·3***	***3·8***	***3·0***
5·2	***3·9***	***4·7***	***3·8***	***5·1***	***4·0***	***3·2***	***4·2***	***5·3***	***3·8***

(a) *Starting with the lowest class limit of* ***3·0*** *thousand shillings, form a frequency*
distribution table with class intervals of ***0·5*** *thousand shillings.*
(b) *Calculate the median and standard deviation.*
(c) *Display the data on a histogram and use it to estimate the mode*

[Ans: (b) 4·075, 0·6936 (c) 3·75]

29. *At* ***7:30am*** *daily a bus leaves town* ***P*** *for town* ***Q.*** *The times in minutes*

taken to cover the journey were recorded over a period of time and grouped as follows.

Time (min)	*Frequency*
40 – 44	*12*
45 – 49	*15*
50 – 54	*35*
55 – 59	*40*
60 – 64	*28*
65 – 69	*20*

(a) *calculate the mode, mean and standard deviation for the distribution*

(b) *Draw an ogive for the data and use it to estimate the:*

(i) *median time for the journey*

(ii) *semi-interquartile range for the distribution*

(iii) *number of times the bus arrived in town* ***Q*** *between* ***8:20 – 8:36am***

[Ans: (a) 55·9706, 55·9, 7·1150 b (i) 56 (ii) 5·25 (iii) 108]

30. *The cumulative distribution of the heights in cm of* ***400*** *students were as follows:*

Height (cm)	*< 100*	*< 110*	*< 120*	*< 130*	*< 140*	*< 150*	*< 160*	*< 170*
Cumulative Frequency	*0*	*27*	*85*	*215*	*320*	*370*	*395*	*400*

(a) *Draw an ogive for the data and use it to estimate the:*

(i) *median height*

(ii) *interquartile range*

(iii) 10^{th} *to* 90^{th} *percentile range*

(b) (i) *Calculate the standard deviation for the data.*

(ii) *Draw a histogram for the above data.*

[Ans: a(i) 129 (ii) 16 (iii) 34 b(i) 13·1114]

31. *The marks scored by* ***200*** *students in a test were as follows:*

Marks	*Frequency*

Below 10	23
10 and under 20	67
20 and under 30	42
30 and under 40	25
40 and under 50	21
50 and under 60	15
60 and under 70	7

(a) *Calculate the:*
(i) upper quartile mark
(ii) standard deviation
(iii) number of students who passed, if the pass mark was ***38.***

(b) Draw a cumulative frequency curve for the data and use it to estimate the:

(i) lowest mark for a distinction one if the top ***5%*** *of the students qualify for distinction.*

(ii) number of students who scored between ***15*** *and* ***45*** *marks.*

[Ans: a(i) 37·2 (ii) 15·9899 (iii) 48 b(i) 58 (ii) 106]

32. *The heights in* ***cm*** *of* ***40*** *plants were as follows:*

Heights H	No of Plants
$30 \leq H < 35$	8
$35 \leq H < 40$	5
$40 \leq H < 55$	12
$55 \leq H < 60$	9
$60 \leq H < 65$	6

(a) Calculate the:
(i) mode
(ii) number of plants shorter than ***45cm***

(b) Draw an ogive for the data and use it to estimate the:
(i) median
(ii) range of the heights of the middle ***50%*** *of the plants.*

[Ans: a(i) 58 (ii) 17 b(i) 49 (ii) 21]

33. *The marks scored by* ***20*** *students in a test are as follows:*

60 80 70 90 50 60 80 90 80 60
90 70 60 70 80 80 60 70 60 70

(a) Form an ungrouped frequency distribution table for the data
(b) Find the:
(i) mode
(ii) standard deviation
(iii) median
(iv) semi–interquartile range

[Ans: (i) 60 (ii) 11·5217 (iii) 70 (iv) 10]

34. *A set of* ***9*** *observations had the following results:* $\sum x = 290$ *and* $\sum x^2 = 8469$. *Find the mean and variance of the distribution.*

[Ans: 29, 5·9]

35. *The lengths* ***a, b, 8, 7,5*** *in cm of five metal rods are distributed with a mean of* ***6*** *and variance of* ***2.*** *Given that* ***a > b,*** *find the values of* ***a*** *and* ***b.***

[Ans: a = 6, b = 4]

36. *A certain frequency distribution gave the following results:* $\sum f = 20$, $\sum fx = 563$ *and* $\sum fx^2 = 16143$. *Find the mean and standard deviation of the distribution.*

[Ans: 28·15, 3·8376]

37. *A certain frequency distribution gave the following results:* $\sum f = 30$, $\sum f(x - \bar{x})^2 = 182 \cdot 3$ *and* $\sum fx^2 = 1025$. *Find the mean of the distribution.*

[Ans: 5·3]

38. *A set of ten observations had the following results:* $\sum (x - \bar{x})^2 = 60$ *and* $\sum x^2 = 285$. *Find the mean of the distribution.*

[Ans: 5]

39. *Show that the standard deviation of the first* ***n*** *integers is* $\sqrt{\frac{n^2 - 1}{12}}$. *Hence obtain the mean and standard deviation of the integers* ***1, 2, – – – – 26.***

[Ans: 13·5, 7·5]

40. *Two samples of equal size respectively have mean* μ_1 *and* μ_2 *and standard deviation* σ_1 *and* σ_2 *. Show that the combined variance*

$$\sigma^2 = \frac{2(\sigma_1^2 + \sigma_2^2) + (\mu_1 - \mu_2)^2}{4}$$