Statistics Revision Sheet – Question 6 of Paper 2

The Statistics question is concerned mainly with the following terms.

The **Mean** and the **Median** and are two ways of measuring the average.

$$\textbf{Mean} = \frac{sum\,of\,values}{no.of\,values}$$

**Median** = Middle value   **Mode** = Most common value

Frequency Distribution Tables and **Cumulative Distribution Tables** are ways of displaying large amounts of data in a table form.

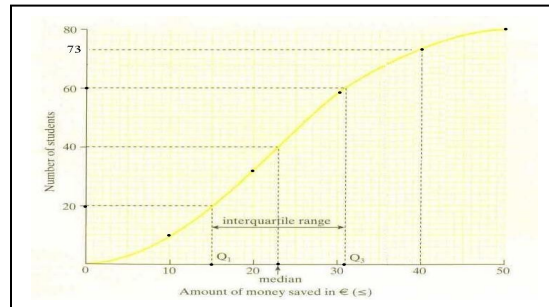| Age of Entrants	15	16	17	18
No. of People	3	5	8	4

Inter-quartile Range measure the spread of the data.

Inter-quartile Range = $Q_3 - Q_1$

The **Histogram** and the **Cumulative frequency curve** are graphs used to display data.
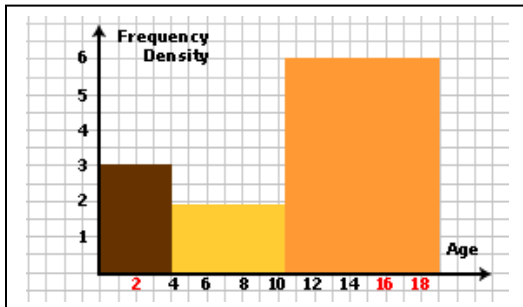
Histogram

Cumulative frequency curve

The Mean, Mode and Median –

Mean – This is the average of a set of values and is found by getting the sum of the values and dividing by the number of values

$$\text{Mean} = \frac{\sum x}{n} \quad \left(\frac{sum\,of\,values}{no.of\,values}\right) \qquad \text{(sum of values means all the values added together)}$$

Mode – This is the value that occurs most frequently

Median – this is the middle value when the values have been arranged in order from the lowest to the highest. If there are two middle values then the median is the average of these.

Example – Find the **(i)** mean, **(ii)** mode and **(iii)** median of the following set of numbers:
3, 3, 5, 6, 6, 4, 3, 7, 8, 9

(i) $\text{Mean} = \dfrac{sum\,of\,values}{no.of\,values} = \dfrac{3+3+5+6+6+4+3+7+8+9}{10} = \dfrac{54}{10} = 5.4$

We add up all the numbers (54) and divide it by the number of numbers, (10).

(ii) Mode = 3
3 occurs three times, this is more than any other number.

(iii) Median – *First line up the numbers from the lowest to the highest.*
3, 3, 3, 4, 5, 6, 6, 7, 8, 9

$$\text{Median} = \frac{5+6}{2} = \frac{11}{2} = 5.5 \qquad \textit{There are two middle terms 5 and 6}$$
We average these by adding together and dividing by 2

Sometimes we will be given the mean and asked instead to find a missing value

Example – The mean of the numbers 4, 8, 6, x, 5 and 3 is 5. Find the value of x.

$\text{Mean} = \dfrac{sum\,of\,values}{no.of\,values} = 5$ *Write down the formula and let it equal the mean, 5*

$\dfrac{4+8+6+x+5+3}{6} = 5$ *Add the no. of values (including the x) and divide the no. of numbers, 6. Let it equal 5, the mean.*

$4+6+6+x+5+3 = 5(6)$ *Multiply across by 6*

$24 + x = 30$ *Simplify*

$x = 30 - 24$ *x's to one side, numbers to the other*

x = 4   *Answer*

Frequency Distribution –

This is a table, which summarises data showing how often each value occurs.

To find the **Mean** of a frequency distribution we find the sum of the values multiplied by their frequency and divide this by the sum of the frequencies.

$$\frac{\sum fx}{\sum f}$$ <u>Sum of the values multiplied by their frequency</u>
Sum of the frequencies

Example – Below is a Frequency Distribution showing the marks obtained (out of 5) by 30 students in a class. Calculate the mean.
(3 students got 0 marks, 7 students got 1 mark, 5 students got 2 marks etc)

Mark	0	1	2	3	4	5
No. of Students	3	7	5	6	8	1

$$\text{Mean} = \frac{\sum fx}{\sum f} = \frac{(0\times3)+(1\times7)+(2\times5)+(3\times6)+(4\times8)+(5\times1)}{3+7+5+6+8+1}$$

This gives the total marks
This gives the total people

$$\frac{3+7+10+18+12+10}{3+7+5+6+8+1} = \frac{60}{30} = 2$$

Mean = 2   Therefore the average score in the class was 2 marks.

The mode (most common) is 4 marks. 8 people got 4 marks.
More people got 4 marks than any other mark.

Example – The mean of the following frequency distribution is 3. Calculate the value of x.

Mark	0	1	2	3	4	5	6
No. of Students	1	8	12	11	x	5	4

$$\text{Mean} = \frac{\sum fx}{\sum f} = \frac{(0\times1)+(1\times8)+(2\times12)+(3\times11)+(4\times x)+(5\times5)+(6\times4)}{1+8+12+11+x+5+4} = 3$$

$$\frac{0+8+24+33+4x+25+24}{41+x} = 3$$ *We sub values into the formula and let it = 3*

$$\frac{114+4x}{41+x} = 3$$ *Simplify*

$$114+4x = 3(41+x)$$ *Multiply across by (41 + x)*

$$114+4x = 123+3x$$ *Simplify*

$$4x-3x = 123-114$$ *x's to one side, numbers to the other*

$$x = 9$$ ***Answer***

3

`Grouped Frequency Distributions –`

A grouped frequency distribution shows the frequency of a range of values. Instead of having a heading for every score in a test from 1 to 30 we could group them into those who got $0 – 10$ those who got $10 – 20$ and those who got $20 – 30$.

To find the **mean** we must first rewrite the table inserting a **mid-interval value** instead of a range of values.
We then find the sum of the mid-interval values multiplied by their frequency all divided by the total number of values.

Mid-interval Value – If we have a group interval of $5 – 15$, the mid interval would be $\frac{5+15}{2} = \frac{20}{2} = 10$.
We add the two values and divide by 2.

Example – The time taken for 20 students to run a race is given below. Calculate the mean.

Time	12-14	15-17	18-20	21-23
No. of people	3	5	8	4

Rewrite the table using mid-interval values.

Mid interval	$\frac{12+14}{2} = 13$	$\frac{15+17}{2} = 16$	$\frac{18+20}{2} = 19$	$\frac{21+23}{2} = 22$
Time	13	16	19	22
No. of people	3	5	8	4

$$\text{Mean} = \frac{\sum fx}{\sum f} = \frac{(13 \times 3) + (16 \times 5) + (19 \times 8) + (22 \times 4)}{3+5+8+4}$$

$$\frac{13+90+152+88}{20} = \frac{343}{20} = 17.15$$

Mean = 17.15 minutes

IF YOU GET ASKED A QUESTION ON WEIGHTED MEANS TREAT IT LIKE A FREQUENCY DISTRIBUTION QUESTION.
WEIGHTS ARE LIKE FREQUENCIES.

Histograms –

We use histograms to illustrate frequency distributions. It is similar to a bar chart though the widths of the bars MAY be different sizes.

Make sure that the Histogram is well labelled.
The variables (things being measured) go on the horizontal (x) axis.
The frequencies (number of people usually) go on the vertical (y) axis.
Just remember that the top of your table goes on the bottom of your graph.

When the class intervals are equal the histogram is just like a bar chart.
If the class intervals are unequal in measure the widths of the columns will change.

Example of a Histogram where intervals are **equal** –

Age	0-10	10-20	20-30	30-40
No. of people	4	8	10	7
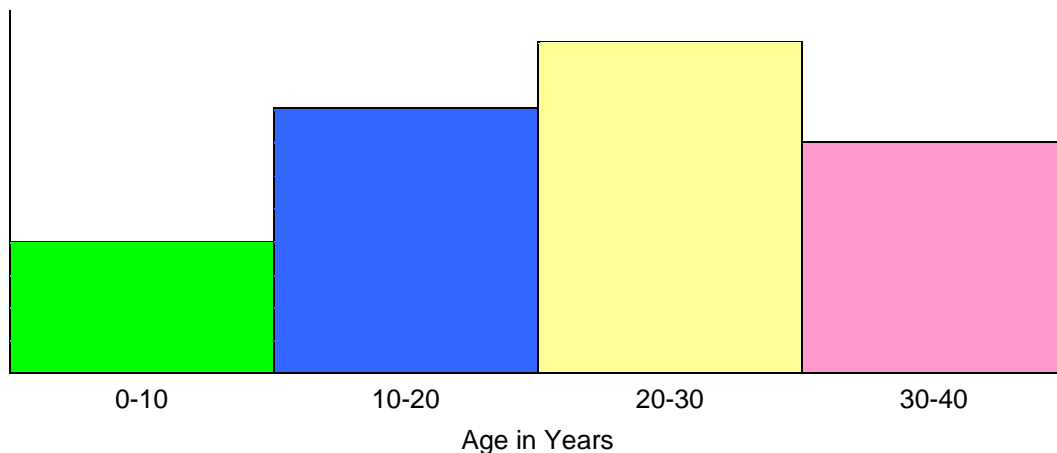
The class intervals are equal (all 10 units wide) so the width of the columns remains equal across the histogram.

Example of a Histogram where intervals are **unequal** –

Minutes	5-15	15-25	25-45	45-75
No. of People	10	50	80	60

The class intervals are unequal and therefore we must adjust the widths of the columns. They will not be the same.

5 – 15   interval of 10
15 –25   interval of 10
25 – 45   interval of 20
45 – 75   interval of 30

> *The smallest interval is 10 so we use this as the unit base of our histogram.*
> *Interval of 10 = 1 unit wide*
> *Interval of 20 = $\dfrac{20}{10} = 2\,units\ wide$*
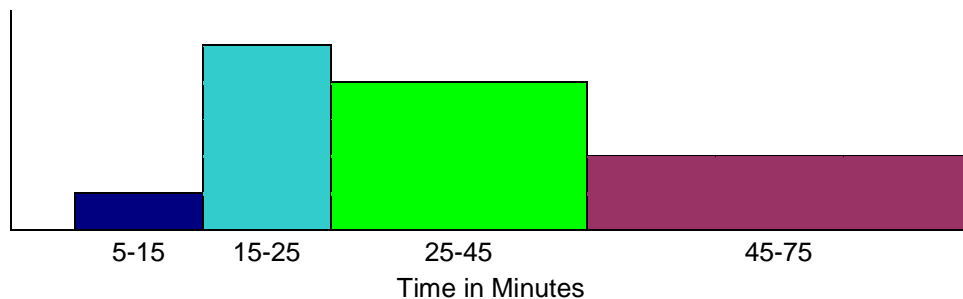> *Interval of 30 = $\dfrac{30}{10} = 3\,units\ wide$*

The height of each interval is found by dividing the number of people in each interval by the amount of units wide each interval is as calculated above.

5 – 15 is 1 unit wide so $\dfrac{10}{1} = 10$ *Height = 10*

15 –25 is 1 unit wide so $\dfrac{50}{1} = 50$ *Height = 50*

25 – 45 is 2 units wide so $\dfrac{80}{2} = 40$ *Height = 40*

45 – 75 is 3 units wide so $\dfrac{60}{3} = 20$ *Height = 20*

5-15   15-25   25-45   45-75

Time in Minutes

Using the Histogram to calculate the frequencies

We will sometimes be given the histogram and asked to calculate the frequencies. A histogram is made up of columns (rectangles).
We will be given the area of one of the rectangles and use this to calculate the others.

Example – Use the Histogram below to complete the frequency table.

Age in Years	0-10	10-20	20-30	30-40	40-60	60-100
No. of Pupils				20		
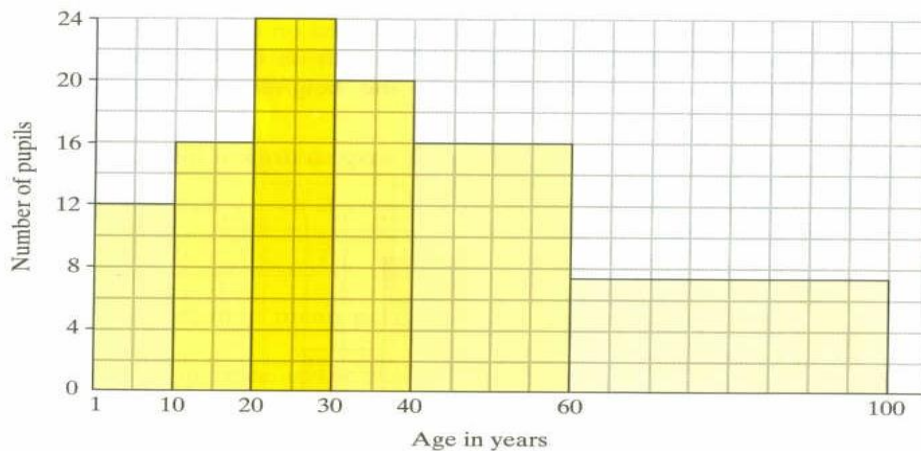
We are given the 4th rectangle (30 – 40) and told that the frequency is 20.
Since we can see that its height is 20 in the graph and we are told that it is 20 in the table, the base is 1 unit wide.
(the narrowest column is always 1 unit wide).

The 1st rectangle has base 1 and height 8   Frequency = 8
The 2nd rectangle has base 1 and height 16   Frequency = 16
The 3rd rectangle has base 1 and height 24   Frequency = 24
The 4th rectangle has base 1 and height 20   Frequency = 20
The 5th rectangle has base 2 and height 16   Frequency = 2 x 16 = 32
The 6th rectangle has base 4 and height 7   Frequency = 4 x 7 = 28

Age in Years	0-10	10-20	20-30	30-40	40-60	60-100
No. of Pupils	8	16	24	20	32	28

Cumulative Frequency Curve/ Ogive –

A cumulative frequency table differs from a normal table in that we have a running total of the values as the intervals increase.
We often use a normal frequency table to construct our cumulative frequency table.
We do this by drawing the table again and adding the previous term each time to the frequency values.

The cumulative frequency curve (ogive) is used to illustrate this data. The curve will always rise in this type of graph.

We use the curve to estimate the **median** and **inter-quartile** ranges.

To find the **median**, we go to the midway point of y axis draw a line across to hit our curve and bring this vertically down to read the value of the x axis.

The **inter-quartile** range = $Q_3 - Q_1$

Q_1 is found by going a quarter of the way up the y axis, drawing a line across to hit our curve and bring this vertically down to read the value of the x axis.

Q_3 is found by going a three quarters of the way up the y axis, drawing a line across to hit our curve and bring this vertically down to read the value of the x axis.

Example – Find the (i) Median, (ii) the Inter-quartile range and (iii) number of students who saved over €40 by drawing a cumulative frequency curve (ogive) of the following table.
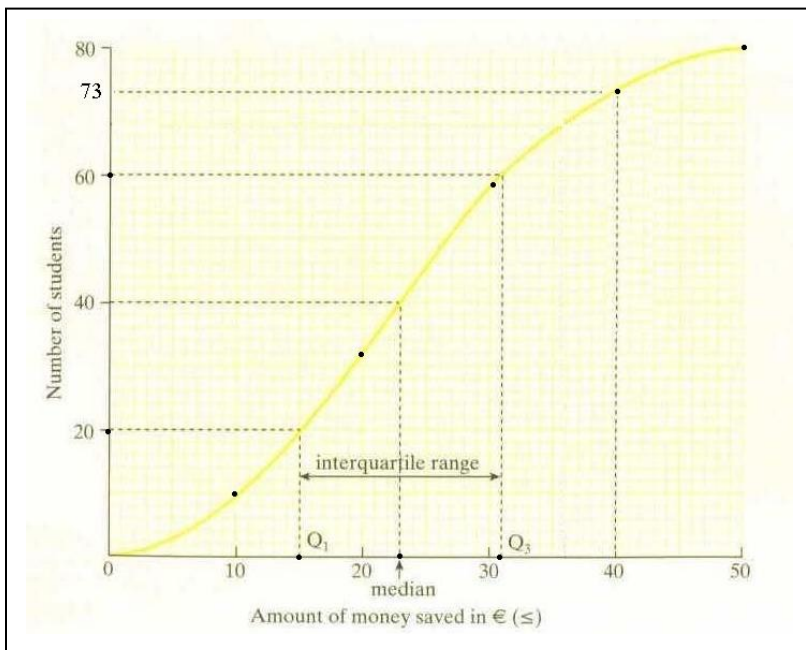
Amount saved	0-10	10-20	20-30	30-40	40-50
No. of Students	10	22	26	16	6

First construct the cumulative frequency table.

Amount saved	≤ 10	≤ 20	≤ 30	≤ 40	≤ 50
No. of Students	10	32	58	74	80
How?	*10*	*(10+22)*	*(32+26)*	*(58+16)*	*(74+6)*

The points then for the curve are (10, 10), (20, 32), (30, 52), (40, 74) and (50, 80)

(i) Median – we go up half way on the y-axis draw a line across then down to the x-axis.
= €23

(ii) $Q_1 = 15$
         $Q_3 = 31$

Inter-quartile = $Q_3 - Q_1 =$
      31 – 15 = 16

(iii) Saved more than €40

Go to 40 on the x-axis, draw vertical line up to meet the curve and read across the value on the y-axis to get 73.

80 – 73 = 7 people saved more than €40