Chapter 3
Statistics

Average and Mean:

DEF: The mean of a set of values is defined as the sum of all the values divided by the number of values.

Formula:

Mean = $\dfrac{Sum.of.all.the.values}{Number.of.values}$ **or** $\bar{x} = \dfrac{\sum x}{n}$

where: $\bar{x}$, **(x bar) is the symbol for mean**
$\sum$ **means the "the sum of"**
n is the number of values.

Median:

DEF: When the values are arranged in ascending or descending order, the median is the middle number.

Example:
Find the (i) mean and (ii) median of the following array of numbers.
3, 8, 4, 10, 8, 7, 2. 5, 5, 9, 5

Answer:
(i) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{3+8+4+10+8+7+2+5+5+9+5}{11} = \dfrac{66}{11} = 6$

(ii) Median: First arrange the numbers from smallest to biggest.
2, 3, 4, 5, 5, 5, 7, 8, 8, 9, 10

The middle number is 5 so the <u>median = 5</u>

<u>Mean and Median of a Frequency Distribution</u>

Mean:

Use the formula - $\bar{x} = \frac{\sum fx}{\sum f}$ **where f = frequency and x = values.**

Median:

When the values are arranged in order of size, the median an then be read directly from the table.

<u>Example:</u>

Find the (i) mean and (ii) median of the following frequency distribution.

Value	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Frequency	6	8	10	7	3	5	4	2	1	1

<u>Answer:</u>

(i) Mean

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{6(0)+8(1)+10(2)+7(3)+3(4)+5(5)+4(6)+2(7)+1(8)+1(9)}{6+8+10+7+3+5+4+2+1+1}$$

$$= \frac{0+8+20+21+12+25+24+14+8+9}{47} = \frac{141}{47} = \underline{3}$$

<u>Mean = 3</u>

(ii) Median

There are 47 values altogether. The median then lays on the 24th value.

The 24th value lays where x = 2

So the <u>median = 2</u>

Grouped Frequency Distribution.

Sometimes it is impossible to write down all the values so they get arranged into suitable groups called *class intervals*.

To answer these questions we use the same formula as with frequency distribution, but we must find the *mid-interval values* before we can calculate the mean.

Example:

Find the (i) mean and (ii) median of the following frequency distribution.

Value	**0 - 2**	**2 - 6**	**6 - 12**	**12 - 20**
Frequency	4	4	6	11

Answer:

Firstly, re-draw the table using the mid-interval values.

Value	**1**	**4**	**9**	**16**
Frequency	4	4	6	11

(i)

$$\overline{x} = \frac{\sum fx}{\sum f} = \frac{4(1)+4(4)+6(9)+11(16)}{4+4+6+11} = \frac{4+16+54+176}{25} = \frac{250}{25} = 10$$

Mean = 10

(ii) There are 25 values altogether. The middle value is 13. The 13^{th} value lays in the 6 – 12 class interval.

Weighted Mean

Weighted mean = $\overline{x}_w = \dfrac{\sum wx}{\sum w}$ **where w = weight and x = value**

Example:

A composite index number is constructed by taking the weighted mean.

The following table gives the index and weighting for each of the four commodities:

Commodity	Food	Fuel	Mortgage	Clothing
Index	120	110	90	80
Weight	8	5	4	3

Calculate the composite index number.

Answer:

Use the weighted mean formula

$$\overline{x}_w = \frac{\sum wx}{\sum w} = \frac{8(120)+5(110)+4(90)+3(80)}{8+5+4+3}$$

$$= \frac{960+550+360+240}{20} = \frac{2110}{20} = 105.5$$

So the composite number = <u>105.5</u>

Standard Deviation

The standard deviation gives an indication of how far the values as are spread out from the mean. The larger the standard deviation the more spread out the values are.

We use the formula: $\sigma = \sqrt{\frac{\sum d^2}{n}}$

Where: σ **= symbol for standard deviation**

d = deviation, or $x - \bar{x}$

n = the number of values of x.

Example:

Calculate the standard deviation, correct to two decimal places, of the following array of numbers: 1, 2, 3, 6, 8

Answer:

First calculate the mean, $\bar{x}$.

$$\bar{x} = \frac{1+2+3+6+8}{5} = \frac{20}{5} = 4$$

Secondly, make a table.

x	d = $x - \bar{x}$	d^2
1	1 – 4 = -3	9
2	2 – 4 = -2	4
3	3 – 4 = -1	1
6	6 – 4 = 2	4
8	8 – 4 = 4	16

$$\sum d^2 = 34$$

Now use the formula:

$$\sigma = \sqrt{\frac{\sum d^2}{n}} = \sqrt{\frac{34}{5}} = \sqrt{6.8} = \underline{\underline{2.61}}$$

Standard Deviation of a Frequency Distribution.

We use the formula: $\sigma = \sqrt{\frac{\sum fd^2}{\sum f}}$ **where d =** $x - \bar{x}$ **and** f **= frequency**

Example.

Calculate the standard deviation, correct to two decimal places, of the following frequency distribution.

Value	**0 - 10**	**10 - 20**	**20 - 30**	**30 - 50**	**50 – 80**
Frequency	**10**	**19**	**25**	**30**	**16**

Answer:

Firstly re-draw the table with the mid-interval values.

Value	**5**	**15**	**25**	**40**	**65**
Frequency	**10**	**19**	**25**	**30**	**16**

Now calculate the mean.

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{10(5) + 19(15) + 25(25) + 30(40) + 16(65)}{10 + 19 + 25 + 30 + 16} = \frac{3200}{100} = 32$$

Now calculate the standard deviation

f	**x**	**d**	**d^2**	**fd^2**
10	**5**	**5 – 32 = -27**	**$(-27)^2 = 729$**	**10 × 729 = 7290**
19	**15**	**15 – 32 = -17**	**$(-17)^2 = 289$**	**19 × 289 = 5491**
25	**25**	**25 – 32 = -7**	**$(-7)^2 = 49$**	**25 × 49 = 1225**
30	**40**	**40 – 32 = 8**	**$(8)^2 = 64$**	**30 × 64 = 1920**
16	**65**	**65 – 32 = 33**	**$(33)^2 = 1089$**	**16 × 1089 = 17424**

$$\sum fd^2 = 33350$$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f}} = \sqrt{\frac{33350}{100}} = \sqrt{333.5} = \underline{\underline{18.26}}$$

Histograms.

A histogram is similar to a bar chart, but there is no spaces between the bars and in a histogram the area of each rectangle represents the frequency. Also the sum of the areas equals the sum of the frequency.

Example:

Construct a histogram to represent the data of the following frequency distribution table.

Interval	*0 – 20*	*20 – 30*	*30 – 50*	*50 – 80*	*80 – 120*
Frequency	*20*	*8*	*22*	*21*	*20*

Answer:

The smallest interval is 20 – 30 (with 10 between them) so we give it a base of 1.

0 – 20 and 30 – 50 (20 between them) so we give them a base of 2.

50 – 80 is next (30 between them) so we give it a base of 3.

80 – 120 is last (40 between them) so we give it a base of 4.

New Table:

Interval	0 – 20	20 – 30	30 – 50	50 – 80	80 – 120
Frequency	20	8	22	21	20
Base	2	1	2	3	4
Height = $\frac{Frequency}{Base}$	$\frac{20}{2} = 10$	$\frac{8}{1} = 8$	$\frac{22}{2} = 11$	$\frac{21}{3} = 7$	$\frac{20}{4} = 5$

Given the Histogram.

Sometimes we are given a histogram and we have to work out the different frequencies represented by the rectangles. We will be given the area of one rectangle and its height. From this we can work out the base.
We can then work out the other frequencies from the formula:
Area = base × height.

Example:

The distribution of contributions, in euros, given to a charity by a number of people is shown in a histogram below.

Amount in €	**0 – 10**	**10 – 30**	**30 – 40**	**40 – 70**	**70 – 80**
Number of People		**30**			

Answer:

1. **The area between 10 – 30 is 30 and the height is 5.**
2. **Base =** $\frac{Area}{Height} = \frac{30}{5} = 6$

   This rectangle has two marked units so each marked unit has a measurement of 3.
3. **0 – 10, 30 – 40 and 70 – 80 have bases of 3 and 40 – 70 has a base of 9.**
4. **0 – 10 : area = height × base = 4 × 3 = 12**

   30 – 40 : area = height × base = 4 × 3 = 12

   40 – 70 : area = height × base = 3 × 9 = 27

   70 – 80 : area = height × base = 3 × 3 = 9

Cumulative Frequency.

Cumulative frequency is where the frequencies are added up. Each accumulated frequency is the combined total of all the previous frequencies up to that particular value.

In these questions we can be asked to draw a cumulative frequency curve which is known as an Ogive and asked to find the median and interquartile range.

- **The LOWER quartile (Q1) is ¼ of the distribution.**
- **The MEDIAN (Q2) is ½ of the distribution.**
- **The UPPER quartile (Q3) is ¾ of the distribution.**

The interquartile range = Upper quartile - Lower quartile.

Example:

A group of people form a club and over a period of time they contribute to a fund to purchase equipment. The records show the contributions as follows:

Contributions, €	0 – 10	10 – 20	20 – 30	30 – 40	40 – 60
No of People	5	10	25	40	20

(i) Represent the data on a cumulative frequency curve.

(ii) Find the Median

(iii) Find the interquartile range.

(iv) Find the number of contributions below €18

(v) Find the number of contributions between €45 and €55

Answer:

Firstly draw a cumulative frequency table.

Contributions, €	**<0**	**<10**	**<20**	**<30**	**<40**	**<60**
No of People	**0**	**5**	**15**	**40**	**80**	**100**

(i)