Type 1
The procedure to carry on this portfolio was followed according to the given instructions which are mentioned in the attached sheet.

DATA ANALYSIS
The following table was made by taking the weight of 100 students of classes 10th and 12th from the school's dispensary. (With the use a Microsoft Excel Sheet)

Weights	Number of Students
34	
35	1
36	
37	
38	
39	
40	2
41	1
42	2
43	1
44	2
45	4
46	1
47	3
48	
49	2
50	7
51	
52	
53	2
54	6
55	5
56	4
57	3
58	2
59	3
60	3
61	2
62	6
63	
64	5
65	6
66	1
67	1
68	2
69	
70	3
71	1
72	1
73	3
74	
75	1

76	1
77	1
78	2
79	1
80	1
81	1
82	
83	
84	2
85	
86	
87	
88	
89	
90	3
91	
92	
93	
94	
95	1
96	
97	
98	1
99	
100	

This table was further processed by making a table (shown bellow) containing weights (x_i), number of students having the respective weight (f_i) and the product of both (with the help of Microsoft Excel) of them ($f_i x_i$).

Weights (x_i)	Number of Students (f_i)	$x_i\, f_i$
34		0
35	1	35
36		0
37		0
38		0
39		0
40	2	80
41	1	41
42	2	84
43	1	43
44	2	88
45	4	180
46	1	46
47	3	141
48		0
49	2	98
50	7	350
51		0
52		0

53	2	106
54	6	324
55	5	275
56	4	224
57	3	171
58	2	116
59	3	177
60	3	180
61	2	122
62	6	372
63		0
64	5	320
65	6	390
66	1	66
67	1	67
68	2	136
69		0
70	3	210
71	1	71
72	1	72
73	3	219
74		0
75	1	75
76	1	76
77	1	77
78	2	156
79	1	79
80	1	80
81	1	81
82		0
83		0
84	2	168
85		0
86		0
87		0
88		0
89		0
90	3	270
91		0
92		0
93		0
94		0
95	1	95
96		0
97		0
98	1	98
99		0
100		0

After obtaining this table I calculated the mean, median, mode and standard deviation for this sample of population.

First the mean was calculated (μ) by using the following formula:

μ = (∑fᵢxᵢ)/fᵢ

Where,

μ: Mean

fᵢ: Frequency

xᵢ: Weight of students

∑: Sum total all the fᵢ(frequency) and xᵢ (weight)

Hence, the mean (μ) obtained was: 60.59

After calculating the mean (μ) then the median (M) was found by using the following formula:

[(n+1)/2]

Where,

M: Median

n: Sum total of number of students

Hence, the median (∂) obtained was 50.5

After calculating the median (∂) the mode was found.

This was done by just looking at the table and seeing the highest frequency which is marked with yellow on the table it self. **(i.e. 50)**

Lastly, the standard deviation (σ) by using "STDEV (: ; :)" on the Microsoft Excel Sheet.

Hence, the standard deviation (σ) obtained was 35.17023617

After finding the mean, median, mode and standard deviation for the 100 weights, the same but this time by taking two different class intervals (i.e. 10 & 5) was calculated.

The first one class interval taken was of 10. In the table a class interval of 10 (C.I), the middle value of the class interval range (xi), frequency (fᵢ) and the product of the middle value and the frequency (fᵢxᵢ)

Class Interval (C.I)	Middle Value (xᵢ)	Frequency (fᵢ)	xᵢfᵢ
30-40	35	1	35
40-50	45	18	810
50-60	55	32	1760
60-70	65	26	1690
70-80	75	14	1050
80-90	85	4	340
90-100	95	5	475

From the above table C.I - 10 (of 100 weights) Mean, Median, Mode and Standard Deviation by using same above mentioned formulas.

Hence,

Mean: 61.6

Median: 50.5

Mode (marked in yellow): 55

Standard Deviation: 31.17594522

**From these calculations (above) we can clearly see that the mean obtained is not the same as the mean calculated when all the 100 weights are taken at a time, this is because in the above calculations we have divided it into class intervals (i.e. 10). This shows as that when the data is arranged in a different class interval range it gives us a different mean.

From the same table (of class interval 10), the graph (between Frequency & Middle Value of Weights) obtained was:
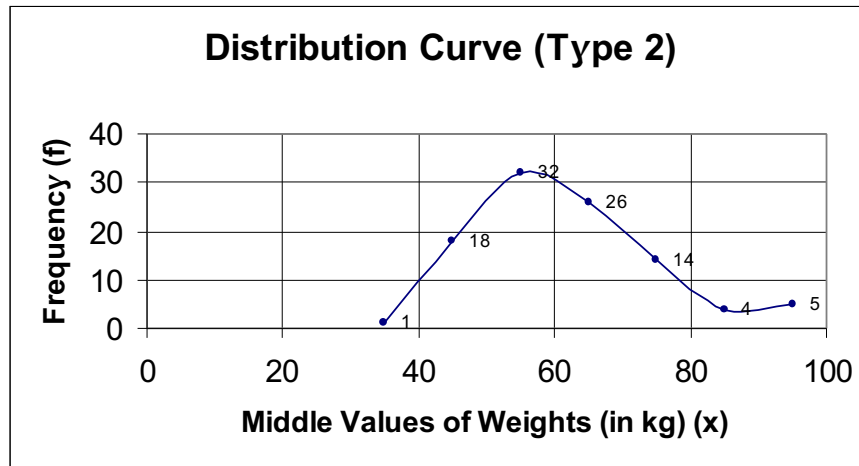
A very important thing that should be taken into consideration is that according to the mean calculated above was 61.6 is not the same as the one which we obtain by looking at the graph (i.e. 55.32) since, in the graph there is a certain minor fall (i.e. at 80 - 90) this is because, at that certain class interval there are very less people with that certain weight.

After the first graph was made for a class interval of 10, I proceeded to do a second graph with a lower class interval (i.e. 5, shown bellow) in order to reduce the difference in mean (by following the same procedure as the class interval of 10).

Class Interval (C.I)	Middle Value (x_i)	Frequency (f_i)	$X_i f_i$
30-35	32.5	0	0
35-40	37.5	1	37.5
40-45	42.5	8	340
45-50	47.5	10	475
50-55	52.5	15	787.5
55-60	57.5	17	977.5
60-65	62.5	16	1000
65-70	67.5	10	675
70-75	72.5	8	580
75-80	77.5	6	465
80-85	82.5	4	330
85-90	87.5	0	0
90-95	92.5	3	277.5
95-100	97.5	2	195

From the above table C.I - 5 (of 100 weights) I calculated the Mean, Median, Mode and Standard Deviation by using same above mentioned formulas.
Hence,
Mean: 61.4

Median: 50.5

**Mode (marked in yellow): 57.5**

Standard Deviation: 33.0932225

**From these calculations (above) one can clearly see that the mean obtained is not the same as the mean calculated when all the 100 weights are taken at a time, this is because in the above calculations we have divided it into class intervals (i.e. 5).

From the same table (of class interval 5), the graph (between Frequency & Middle Value of Weights) obtained was:
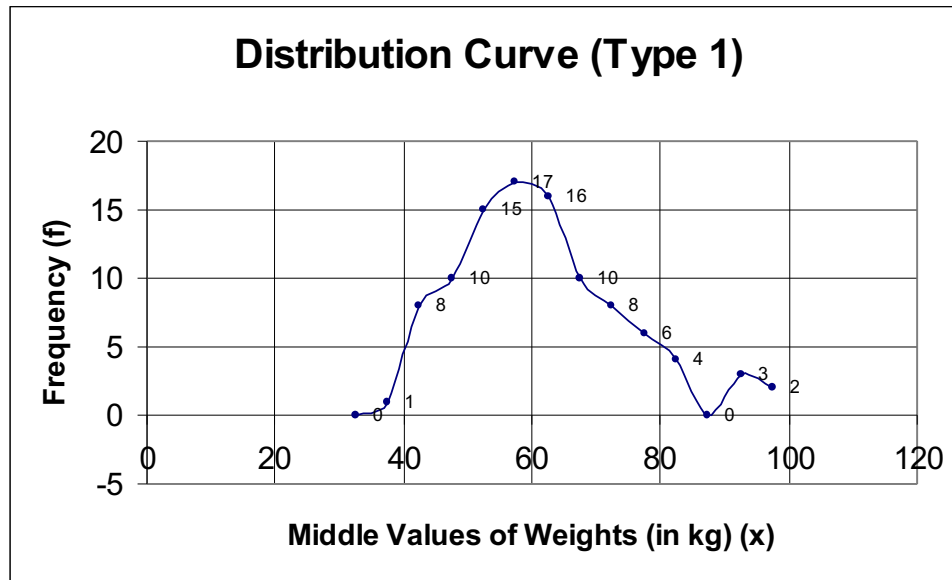
A very important thing that should be taken into consideration is that according to the mean calculated above was 61.4 is not the same as the one which we obtain but looking at the graph (i.e. 57.5) since, in the graph there is a fall (i.e. at 90) this is because, at that class interval there is no one with that weight, but this value is much closer to the mean taken with a class interval of 10 which clearly shows us that less the class interval, will lead to small scattering of data.

After obtaining the two graphs of different class intervals (i.e. 10 & 5) while
We can compare the shape of the curves, the standard deviation and the mean between them.

Distribution Curve Comparison
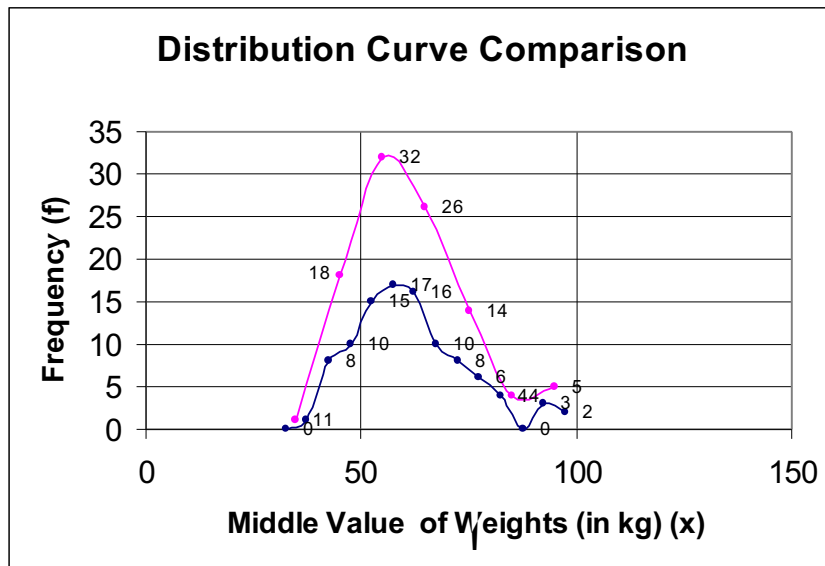
The two curves are not symmetrical due to various reasons:
The curve with class interval of 10 has a small fall at the interval of 80 – 90 this is because of the very less amount of people with this weight. The same thing occurs in the second graph for the class interval of 5 at the interval of 85 – 90. This is the reason why the curves have a non – asymmetrical shape.
The two graphs can also be compared in terms of mean the red curve which is of the class with of 5 shows us a closer (which is the actual one) than the blue line that is of a class interval of 10.
The two graphs can also be compared by the standard deviation as we can see that the data is more compressed in the curve made by the standard deviation of 10 the data is more compressed than the curve made with the class interval of 5 which hence, shows us that greater the class interval, smaller the dispersion of data and smaller the class interval greater is the scattering of data.

To prove all these above mentioned points a random sample of 40 weights was taken (with the class width of 10 and 15).

The table containing the 40 random weights with a (class width of 15) taken is the shown bellow:

Class Interval (C.I)	Middle Value	Frequency	fx
40-55	47.5	9	427.5
55-70	62.5	17	1062.5
70-85	77.5	12	930
85-100	92.5	2	185

For the above random sampling of 40 weights the mean, median, mode and standard deviation was calculated by using the same formulas used for the 100 weights.
Hence,
Mean = 65.25

Median = 20.5

Mode (marked in yellow) = 62.5

Standard Deviation = 35.46915006

**From these calculations (above) one can clearly see that the mean obtained is not the same as the mean calculated when all the 100 weights are taken at a time, this is because in the above calculations we have divided it into a class interval (i.e.15).

From the same table (of class width 15), the graph (between Frequency & Middle Value of Weights) obtained was:

Distribution Curve (Type 1)
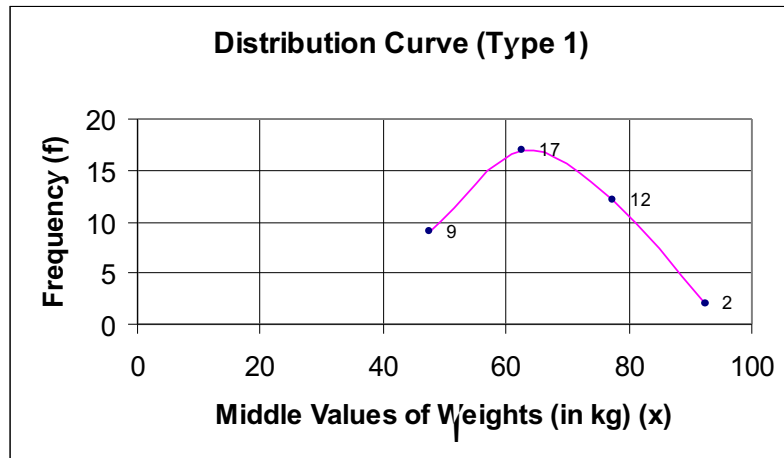
Here, we can clearly see that this is not a perfect normal distribution curve since; the above curve starts only at a height of nine and then has a gradual fall at the class width of 85 – 100.
In this graph we can also see that the mean (on the graph) is not the same to the actual 100 weight mean. The mean obtained from this graph is 62.5 where as the actual mean calculated for this class width is 65.125.

After taking the above class width (i.e. 15) the class with of 10 was taken in order to prove some of the above made points.

The table containing the 40 random weights with a (class width of 10) is shown bellow:

Class Interval (C.I)	Middle Value	Frequency	fx
40-50	45	6	270
50-60	55	9	495
60-70	65	11	715
70-80	75	8	600
80-90	85	4	340
90-100	95	2	190

For the above random sampling of 40 weights the mean, median, mode and standard deviation was calculated by using the same formulas used for the 100 weights.
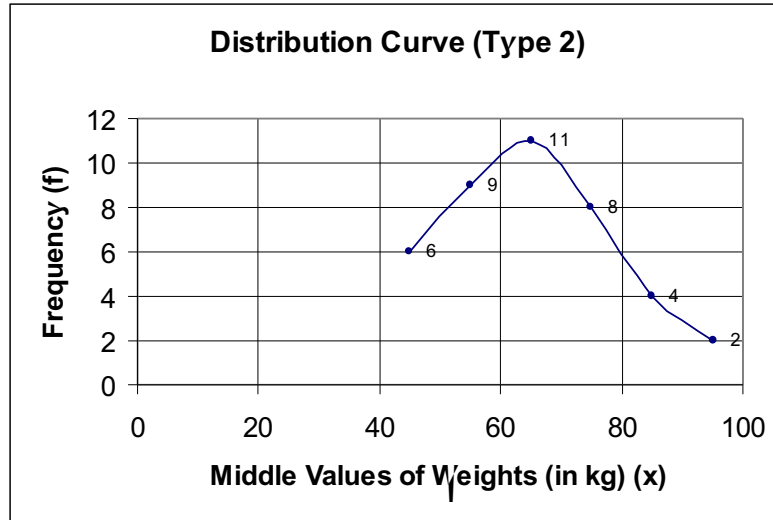Hence,
Mean = 65.25

Median = 20.5

Mode (marked in yellow) = 65

Standard Deviation = 35.46915006

** Again here, from the above calculations we can clearly see that the mean obtained is not the same (as the mean taken for 100 weights).

From the same table (of class width 10), the graph (between Frequency & Middle Value of Weights) obtained was:

Distribution Curve (Type 2)

** Here again we can see that the mean that can be seen in the graph is not the same as the on obtained (i.e. 65.11) for 100 weights, but it can also be said that it is closer to the actual mean.

After obtaining the two graphs of different class intervals (i.e. 15 & 10), we can compare the shape of the curves, the standard deviation and the mean between them.

Distribution Curve (Comparison)
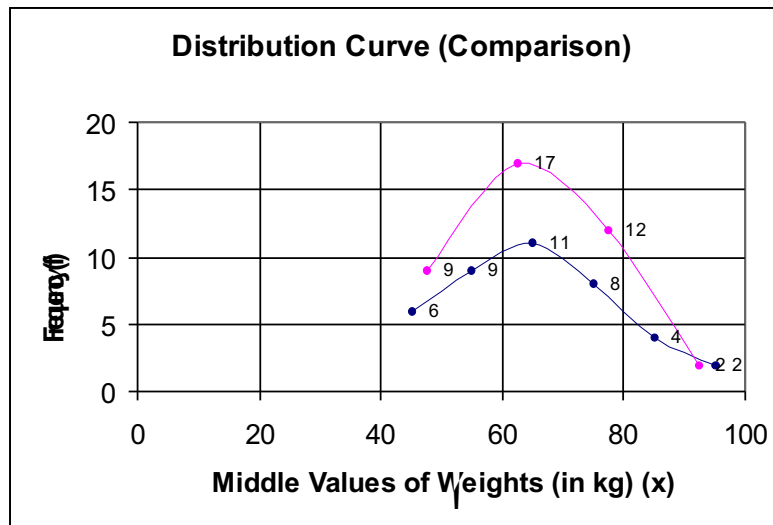
** From this curve we can yet see that both the curve are not perfectly symmetrical but more than the others (of all 100 weights at a time). This is because; these two curves, the red (i.e. C.I – 10) and the blue (C.I – 15) start from a height of 6 and 7 and have a minor fall at 2 and 4 respectively. Hence, their shape is not symmetrical enough.
These two graphs can again be compared by also looking at their standard deviations the red curve (i.e. C.I – 10), is more compressed but the distortion of data is more hence the standard deviation is more than the blue curve (i.e. C.I – 15).

This standard deviation greater than the actual standard deviation for 100 weights. Hence, it has been again shown that greater the mean greater the standard deviation.

<u>QUESTION & ANSWERS</u>

1. What are possible errors in your random gathering of your weights

2. What is the shape of the POPULATION distribution, and why?

The shape of the population distribution can best be described as a non symmetric distribution curve since it has various falls in the graph due to absence of a person or less person of that certain weight.

3. How do the means for each of the four sampling distributions compare to the population mean?

The means of the different samples of class intervals differ to the total sample (i.e. of 100 weights), since the mean in each sample was taken for different class widths.

4. How do the standard deviations compare?

The standard deviations for every sample were different from the actual one. These all differ since, different class intervals. A very important thing was noticed among the different standard deviations which is that grater the class width taken smaller the scattering of date (i.e. standard deviation) and the

smaller the class width the greater the scattering of data (i.e. standard deviation).

5. How do the shapes of the 2 distributions relate to each other?

The shapes of the two graphs all of them have skewness (Skewness is an asymmetrical frequency distribution in which the values are concentrated on one side of the central tendency and trail out on the other side) to the right, which can also be called as positive skewness. This positive skewness is because; the mean is greater than the mode or the median.
This is also positive skewness since, the large tail of the distribution lies toward the higher values of the variable (i.e. right).

6. What will be the relationship between any population mean and the mean of a sampling distribution taken from it?

The main difference between a population mean and a mean of a sample of the population is the standard deviation (i.e. the scattering of data), the mean itself and the mode. This is again because of, the class width taken for each sample of the population.

7. What will be the relationship between their shapes?

The relation between their shapes is that all the samples have a certain fall at a particular value this is due to the reason that there was no one with that certain weight or very less people with that certain weight.
Also, another relationship is that all the curves are positively skewed.
It can also be said that this curves are not symmetrical because they do not follow the fall under the formula given for a normal distribution curve that is:
_____

8. What will be the relationship between their standard deviations?

The only relation between the standard deviations is that the samples of the population have a class interval and also that greater the class width less is the standard deviation (i.e. scattering of data) and vice versa.

9. What causes the placement of the measures of center for the population? Why is one further to one side or the other? What is the cause?