Representing and Measuring Data

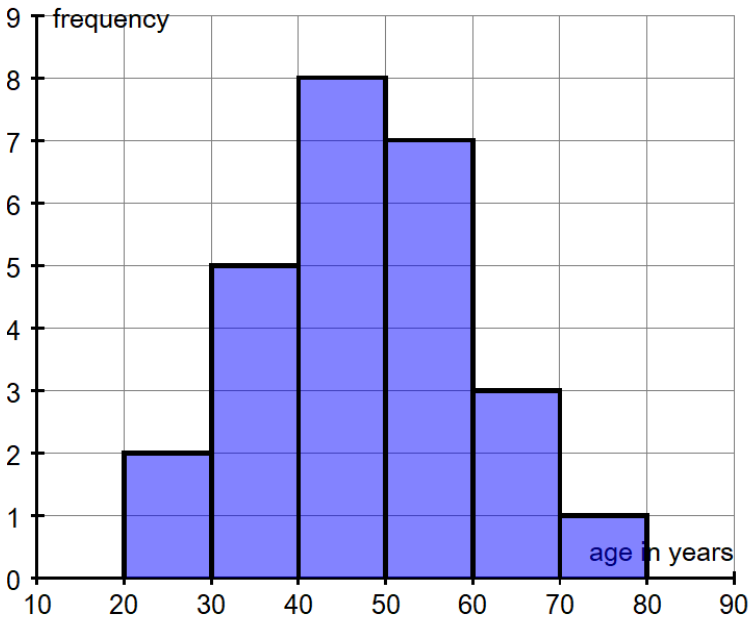
Histogram

These are frequency diagrams. For the IB, we are only concerned with equal class intervals.

For example, here are some data about the ages of teachers in a school

Age	Frequency
20 ≤ x < 30	2
30 ≤ x < 40	5
40 ≤ x < 50	8
50 ≤ x < 60	7
60 ≤ x < 70	3
70 ≤ x < 80	1

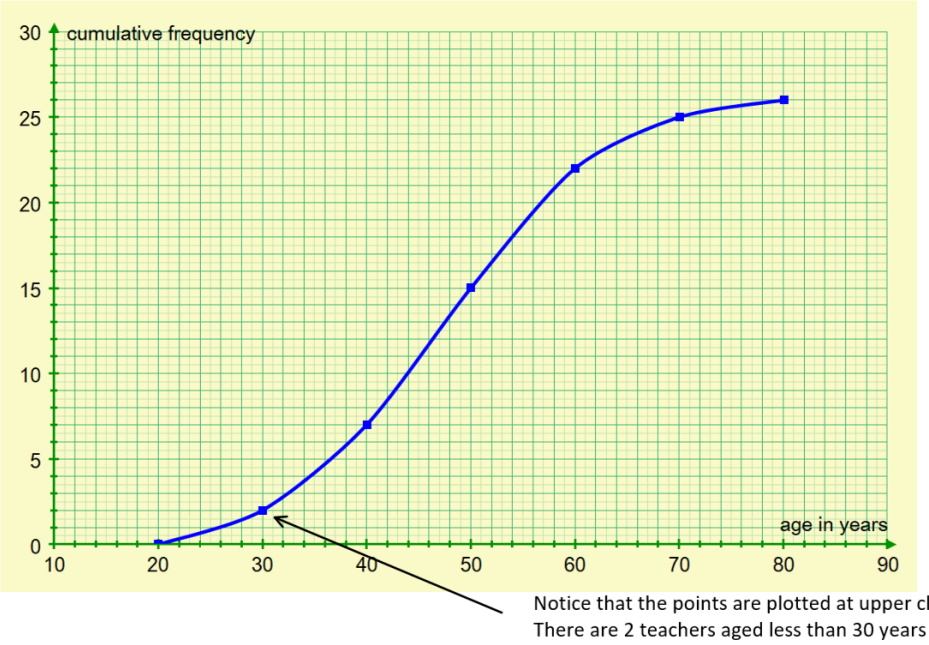
A Histogram to represent the Ages of Teachers



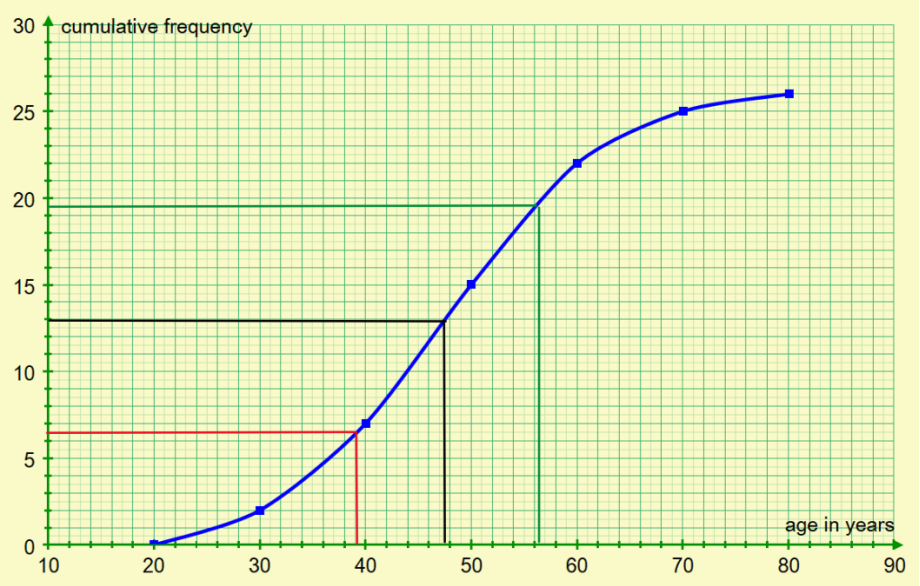
We can put this into a cumulative frequency table

Age	Cumulative Frequency
x < 30	2
x < 40	7
x < 50	15
x < 60	22
x < 70	25
x < 80	26

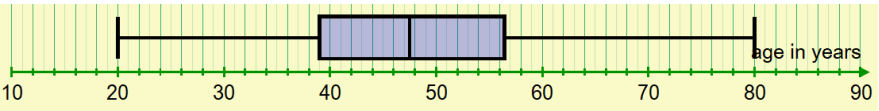
Cumulative Frequency Graph and Box and Whisker Diagram



We can work out the median and quartiles:



Lower quartile = 39
Median = 47.5
Upper quartile = 56.5
And this can be used to plot a box and whisker diagram



Outliers are defined as a data item which is more than 1.5 × interquartile range (IQR) from the nearest quartile. (IQR = Upper quartile - Lower quartile)

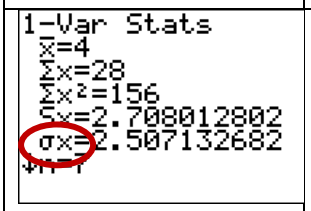
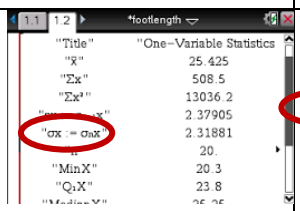
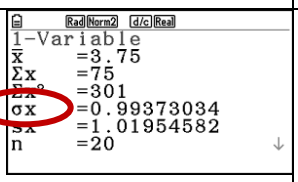
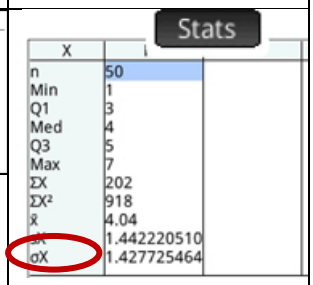
Mean and Standard Deviation & Variance

For Analysis and Approaches course, we are usually given a sample and asked to make calculations about that sample ($\bar{x}$=mean of sample, μ=mean of population). The IB uses the symbol σ to represent standard deviation.

When using grouped data, we use the mid-interval values. So for our teachers, we would use this

Age	Mid-interval	Frequency
20 ≤ x < 30	25	2
30 ≤ x < 40	35	5
40 ≤ x < 50	45	8
50 ≤ x < 60	55	7
60 ≤ x < 70	65	3
70 ≤ x < 80	75	1

Your calculator uses the symbol σx for standard deviation (S_x represents the unbiased estimate of population standard deviation…you won't be asked to calculate this)

Ti 84+	Ti Nspire	Casio	HP Prime
			

Generally, you will be required to use your GDC to make these calculations. However, the IB is keen that you have a good conceptual understanding, so it is useful to understand the formulae

Mean of a data set $\bar{x} = \frac{\sum x}{n}$	Mean from a frequency table $\bar{x} = \frac{\sum f \times x}{\sum f}$
Standard deviation of a data set $\sigma = \frac{\sum (x-\bar{x})^2}{n}$	alternative formula $\sigma = \frac{\sum x^2}{n} - \bar{x}^2$
Standard deviation from a frequency table $\sigma = \frac{\sum f(x-\bar{x})^2}{\sum f}$	alternative formula $\sigma = \frac{\sum f \times x^2}{\sum f} - \bar{x}^2$

Variance = σ^2

Effect of constant changes on the original data

Change to data	Effect on mean	Effect on standard deviation
Add a	Add a	unchanged
Multiply by b	Multiply by b	Multiply by b