Math AA HL Chapter 19 Statistics

Statistics

Mean ($\bar{x}$) – Average $\left(\bar{x} = \frac{\sum x}{n}\right)$

Median – middle
Mode (modal) – most often
Population – group of people or objects
Sample – part of the population
Discrete data – exact data
Continuous data – not exact, will have some degree of accuracy
Frequency table – Table showing how often certain things occur.
Example of Frequency table

Waist size (inches)	Tally	Frequency	Cumulative frequency	Size × frequency
28	𝍸	3	3	84
30	𝍸	4	7	120
32	𝍸 𝍸	7	14	224
34	𝍸 𝍸	9	23	306
36	𝍸	3	26	108
38	𝍸	3	29	114
40	|	1	30	40
	Total	**30**		**996**

Sometimes it is useful to expand a frequency table with additional columns to organize additional information.

Sometimes it is useful to group data together or when the data is continuous, intervals are used to group the data together. Continuous data will have intervals given as inequalities.
Class interval – the groupings
Class width – the width of the group (usually the same width but not always)
Mid-interval value – the median of each interval.

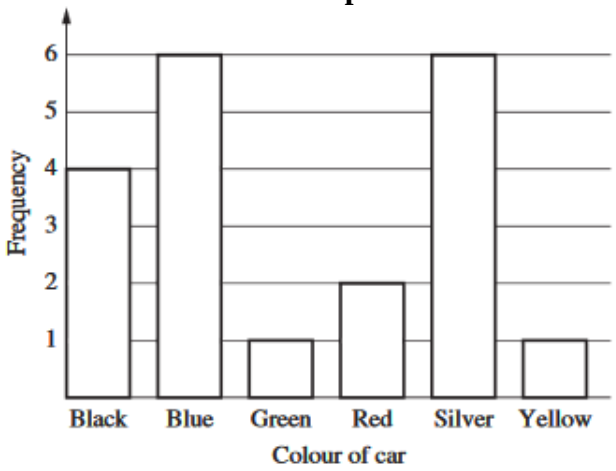
Frequency Distribution – similar to frequency table but horizontal.

Number of languages, *x*	1	2	3	4
Frequency	31	57	42	19

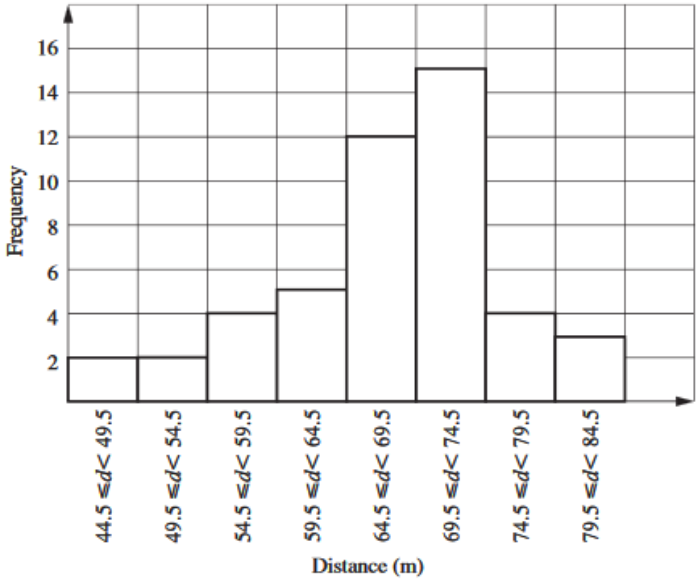
Histograms - similar to a bar chart but …

1. The bars must be adjacent with no spaces between the bars.
2. What is important about the bars is their area, not their height. (We will have equal class widths and so the height can be used to signify the frequency but it should be remembered that it is the area of each bar that is proportional to the frequency.)

Bar Graph

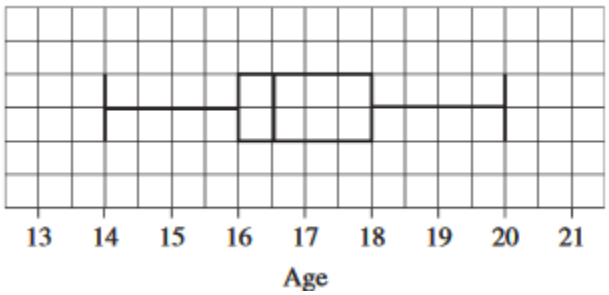


Histogram



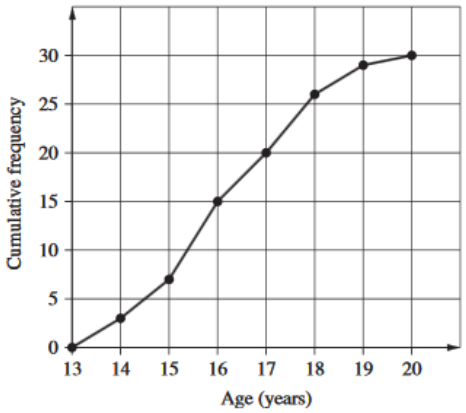
Box and whisker plots

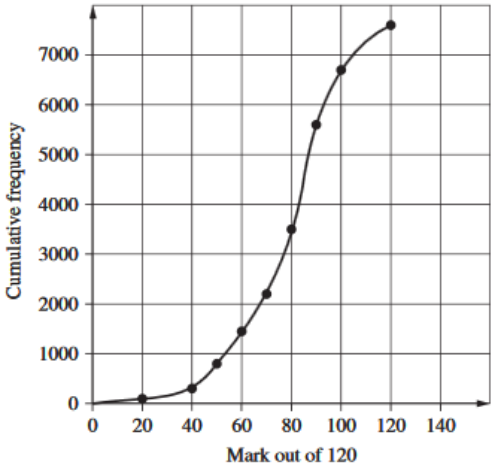
Quartile (Q) – where the data set is divided into 4 equal parts
First quartile (Q_1) – cut off for the lowest 25%
Mid quartile (Q_2) – the median of the data
Third quartile (Q_3) – cut off for the highest 25%



Cumulative frequency diagrams

A cumulative frequency diagram, or ogive, is another diagram used to display frequency data. Cumulative frequency goes on the y-axis and the data values go on the x-axis. The points can be joined by straight lines or a smooth curve. The graph is always rising (as cumulative frequency is always rising) and often has an S-shape.



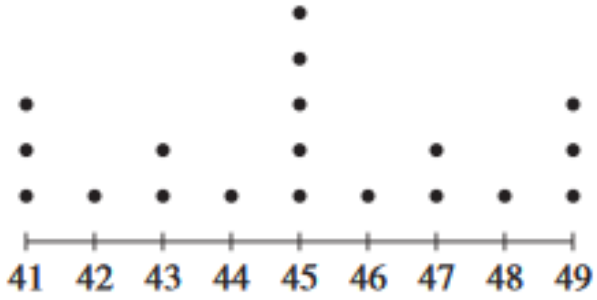


Stem and leaf diagram

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

$n = 24$ key: 6 | 1 means 61 grams

Dot plot



Range – difference between the high and the low
Interquartile range – the difference between the upper quartile (Q_3) and lower quartile (Q_1)
Semi-interquartile range – half the interquartile range

Standard deviation (s) – the measure of dispersion connected to the mean. $s = \sqrt{\frac{\sum (X - \bar{X})^2}{n}}$

Or $s = \sqrt{\frac{\sum x^2}{n} - \left(\bar{x}\right)^2}$

Standard deviation of a population (σ) is $\sigma = \sqrt{\frac{n}{n-1}} \times s$

Variance (σ^2) – the square of the standard deviation $\sigma^2 = \frac{n}{n-1} s^2$

For large samples use $s = \sqrt{\frac{\sum_{i=1}^{k} f_i (X_i - \bar{X})^2}{n}}$

Transformations of statistical data

Consider the effect of these transformations:
• Adding on a constant c to each data item
• Multiplying each data item by a constant k.

Adding on a constant c to each data item
The mean is the original mean + c
The standard deviation is unaltered.

Multiplying each data item by a constant k
The mean is multiplied by k.
The standard deviation is multiplied by k.