National Mathematics Day Activity

Title of chapter/article	Frequency analysis
Author(s)	The Australian Association of Mathematics Teachers (AAMT) Inc.
Copyright owner	The Australian Association of Mathematics Teachers (AAMT) Inc.
Year of publication	2012

Each cipher and code activity has a suggested level: lower primary, upper primary or junior secondary. However, many of the activities can be enjoyed by students (and teachers!) of all ages.

For more information about this resource, please contact:

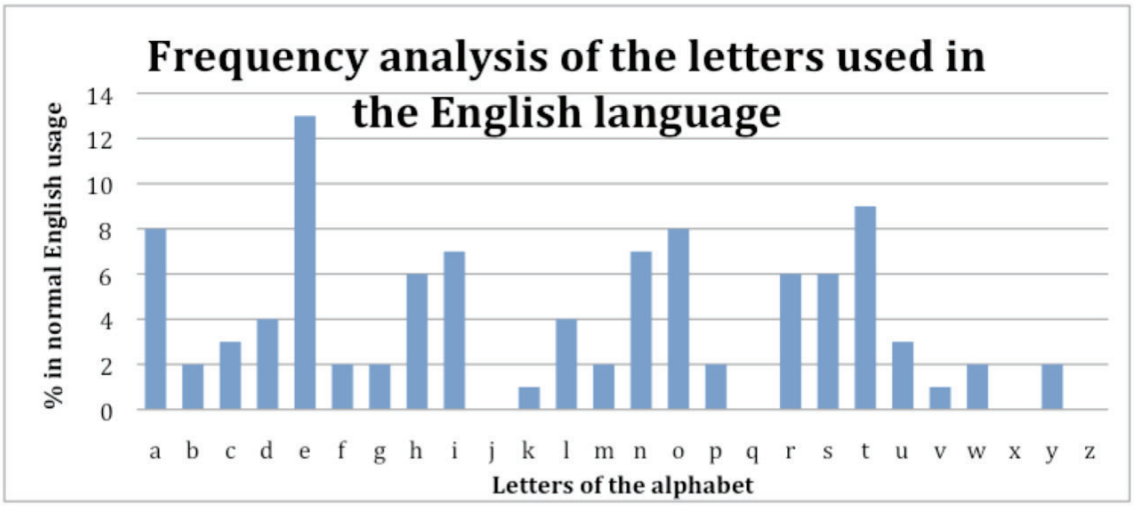
The Australian Association of Mathematics Teachers Inc.
ADDRESS GPO Box 1729, Adelaide SA 5001
PHONE +61 8 8363 0288
FAX +61 8 8362 9288
EMAIL office@aamt.edu.au
INTERNET www.aamt.edu.au

Frequency analysis

[secondary]

Substitution ciphers, where one letter is replaced by a different letter, symbol or number, can be solved using *frequency analysis*. This was the birth of cryptanalysis (code breaking) and was used to incriminate, among others, Mary, Queen of Scots.

It uses the fact that certain letters in the English language are used more often than others to spell words.



Al-Kindi, a 9th century Arab polymath, studied the Koran and discovered that Arabic has a characteristic letter frequency. This is the first known recorded explanation of frequency analysis.

- Look at the graph above. Which is the most frequently occurring letter?
- Which letters do not appear to be used? Can you explain this?
- What comment can you make about the frequencies of vowels and consonants?

- Choose a book. Start at any page and count out 100 letters. Record in order all the letters that you read on the tally chart. For example, ‘substitution’ would be one tally for each of ‘b’, ‘o’ and ‘n’, two tallies for each of ‘s’, ‘u’ and ‘i’, and three tallies for ‘t’.

Letter	Tally	Frequency	Letter	Tally	Frequency
a			n		
b			o		
c			p		
d			q		
e			r		
f			s		
g			t		
h			u		
i			v		
j			w		
k			x		
l			y		
m			z		
			Total		100

- Produce a bar graph, either by hand or electronically, of your results.

  How does your graph compare to the ‘Frequency analysis of the letters used in the English language’ graph above?

  As a class, combine your results and graph them. Is this graph more or less like the graph above? Why do you think this might be?

- Create a short text using some sort of substitution code.

  Ask a partner to ‘crack the code’ using frequency analysis.

  Some other tips that might help are

  – repeated letters—the most common are ‘ss’, ‘ee’, ‘tt’, ‘ff’, ‘ll’, ‘mm’ and ‘oo’ (in that order)

  – one-letter words—there are only two, ‘a’ and ‘I’

  – two-letter words (there are a few more of these!)

  – the most common three letter words—‘and’ and ‘the’

- Investigate/play Scrabble and look at the values given to different letters. Does this match with the frequency analysis table above? Can you explain why/why not?