

Name: _____

Probability & Statistics: Cryptography

Everyone thought can read this since your hand:

if you can read this, you have a strange mind too Can you read this?

**Only 55 people out of 100 can. I couldn't believe that I could audaciously
understand what I was reading. The phenomenal power of the human
mind, according to research at Cambridge University, it doesn't
matter in what order the letters in a word are, the only important
thing is that the first and last letter be in the right place. The rest can
be a total mess and you can still read it without a problem. This is
because the human mind does not read every letter by itself, but the
word as a whole. Amazing huh? yeah and I always thought spelling
was important!**

- When analyzing an encoded message, not every letter needs to be completely deciphered in order to understand the decoded phrase!!

How do I crack a substitution code?

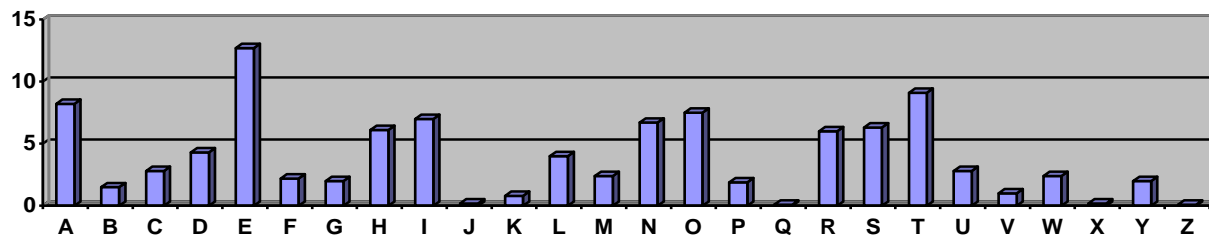
Substitution codes, often called cryptograms, are often used as games. These are solved primarily by two methods. First, the successful decryption of substitution codes will use letter frequency to help solve a difficult code. The twelve most frequently-used letters in the English language are ETAOIN SHRDL, in that order. The least common letters are JXQZ. If you notice a certain letter being used again in again in any given cryptogram, at a frequency much higher than any other letter, it's a good bet that its unencrypted form will be one of the ETAOIN group.

Secondly, pattern recognition is used. The easiest to recognize are single-letter words, which generally can only be A or I (or, very rarely, O). Then there are a limited number of two-letter words such as IN, IS, IT, TO, AN, AT, AS, WE, HE, US, etc. One trick in particular is to look for the common TH- words, i.e. THE, THAT, THEN, THEY, THERE, THEIR.

Finally, when all else fails, simple trial and error can generally fill in the gaps.

Frequency Analysis

Below is a list of average frequencies for letters in the English language. So, for example, the letter E accounts for 12.7% of all letters in English, whereas Z accounts for 0.1%. All the frequencies are tabulated and plotted below.



Please note, these frequencies are averages, and E will not always constitute 12.7% of all the letters in a text, and may not even be the most common letter. The longer the message, the more likely it is that it will obey the average distribution shown above. However, there are exceptions to this rule. In 1969, the French author Georges Perec managed to write a 20-page book called 'La Disparition' without using any words containing the letter E. Amazingly, the book was later translated into English by Gilbert Adair, again avoiding the use of the letter E.

Frequency from greatest to least

ETAOIN SHRDL CUMWFG YPBVK JXQZ

Relative frequencies of letters

By letter

Letter	Frequency
A	0.08167
B	0.01492
C	0.02782
D	0.04253
E	0.12702
F	0.02228
G	0.02015
H	0.06094
I	0.06966
J	0.00153
K	0.00772
L	0.04025
M	0.02406
N	0.06749
O	0.07507
P	0.01929
Q	0.00095
R	0.05987
S	0.06327
T	0.09056
U	0.02758
V	0.00978
W	0.02360
X	0.00150
Y	0.01974
Z	0.00074

By frequency

Letter	Frequency
E	0.12702
T	0.09056
A	0.08167
O	0.07507
I	0.06966
N	0.06749
S	0.06327
H	0.06094
R	0.05987
D	0.04253
L	0.04025
C	0.02782
U	0.02758
M	0.02406
W	0.02360
F	0.02228
G	0.02015
Y	0.01974
P	0.01929
B	0.01492
V	0.00978
K	0.00772
J	0.00153
X	0.00150
Q	0.00095
Z	0.00074

Most common bigrams, or two letter combinations. (in order)

th, he, in, en, nt, re, er, an, ti, es, on, at, se, nd, or, ar, al, te, co, de, to, ra, et, ed, it, sa, em, ro.

Most common trigrams, or three letter combinations. (in order)

the, and, tha, ent, ing, ion, tio, for, nde, has, nce, edt, tis, oft, sth, men

Top 10 beginning of word letters

Letter	Frequency
T	0.1594
A	0.155
I	0.0823
S	0.0775
O	0.0712
C	0.0597
M	0.0426
F	0.0408
P	0.040
W	0.0382

Top 10 end of word letters

Letter	Frequency
E	0.1917
S	0.1435
D	0.0923
T	0.0864
N	0.0786
Y	0.0730
R	0.0693
O	0.0467
L	0.0456
F	0.0408

V

P	D	Z

X	F

A	X	Y	M	T

V	S	T

K	S	V	W	Q	M	N	V	C	W	M

T	M	Q	M	N	L	X	S	V	Q	X	D	S		

Q	D

T	D

V

Q	G	X	S	R

Z	G	M	S

F	K	B	G

V

T	M	Q	M	N	L	X	S	V	Q	X	D	S		

X	F

N	M	W	V	Q	M	T	

Q	D

F	D	L	M	Q	G	X	S	R	

S	D	C	W	M

Z	G	X	B	G

B	V	S

D	S	W	H

K	I	W	X	A	Q

Q	G	M

L	V	S

Z	G	D

L	V	J	M	F

Q	G	M

N	M	F	D	W	P	M

What is the message??

Letter	Tally	Freq.	English	Code
A			E	
B			T	
C			A	
D			O	
E			I	
F			N	
G			S	
H			H	
I			R	
J			D	
K			L	
L			C	
M			U	
N			M	
O			W	
P			F	
Q			G	
R			Y	
S			P	
T			B	
U			V	
V			K	
W			J	
X			X	
Y			Q	
Z			Z	

English	Code
E	
T	
A	
O	
I	
N	
S	
H	
R	
D	
L	
C	
U	
M	
W	
F	
G	
Y	
P	
B	
V	
K	
J	
X	
Q	
Z	

Letter	Tally	Freq.
A		
B		
C		
D		
E		
F		
G		
H		
I		
J		
K		
L		
M		
N		
O		
P		
Q		
R		
S		
T		
U		
V		
W		
X		
Y		
Z		

And finally, the most common words in the English language:

- the
- of
- are
- I
- and
- you
- a
- can
- to
- he
- her
- that
- in
- was
- is
- has
- it
- him
- his