

ENGLISH LANGUAGE APPROXIMATE LETTER FREQUENCIES

Source: <http://www.cryptograms.org/letter-frequencies.php>

by frequency		alphabetical order	
E	0.1257	A	0.0800
T	0.0909	B	0.0154
A	0.0800	C	0.0258
O	0.0759	D	0.0432
I	0.0692	E	0.1257
N	0.0690	F	0.0235
S	0.0634	G	0.0198
H	0.0624	H	0.0624
R	0.0596	I	0.0692
D	0.0432	J	0.0014
L	0.0406	K	0.0074
U	0.0284	L	0.0406
C	0.0258	M	0.0256
M	0.0256	N	0.0690
F	0.0235	O	0.0759
W	0.0222	P	0.0180
G	0.0198	Q	0.0012
Y	0.0190	R	0.0596
P	0.0180	S	0.0634
B	0.0154	T	0.0909
V	0.0098	U	0.0284
K	0.0074	V	0.0098
X	0.0018	W	0.0222
J	0.0014	X	0.0018
Q	0.0012	Y	0.0190
Z	0.0008	Z	0.0008

Most common bigrams (these account for approx. 1/3 of all bigrams)

TH	0.0388	OU	0.0129
HE	0.0368	ED	0.0128
IN	0.0228	HA	0.0127
ER	0.0218	TO	0.0117
AN	0.0214	OR	0.0115
RE	0.0175	IT	0.0113
ND	0.0157	IS	0.0111
ON	0.0142	HI	0.0109
EN	0.0138	ES	0.0109
AT	0.0134	NG	0.01015

The above statistics were produced by analyzing 4 GB of English-language literature in Project Gutenberg (approximately 9,481 different works of literature).

Top 10 beginning-of-word letters: TAISOCMFPW

Top 10 end-of-word letters: ESDTNYROLF