Letter Frequencies in *Letter Head*

www.atlas-games.com/letterhead

One of the outstanding features of *Letter Head* is that the number of cards displaying each letter exactly reflects how frequently that letter actually occurs in English. Thus, for example, since 5% of the letters in any random sampling of words are H's, then you can be sure that 5% of the cards in the *Letter Head* deck are also H's — or as close to 5% as can be approximated with rounded-off whole numbers. This unique feature of *Letter Head* allows for especially fluid and enjoyable game action because words often seem to crop up in players' cards almost supernaturally. Playtesters have commented that there always seems to be just the right amount of each letter; this is because, in fact, there *is* just the right amount of each letter.

If you've always assumed that *all* word games have these same frequencies, think again. Every other word game we've ever come across has serious, sometimes gross distortions in their letter frequencies — often negatively affecting how the games are played.

How can we be so sure that our calculations are correct? That's the question this document addresses. Fortunately, most of the work has already been done for us. Over the years, many studies have been made to determine how frequently each letter of the alphabet occurs in English (mostly for the purposes of cryptanalysis, or code-breaking). When we were figuring out how many cards of each letter to include in the *Letter Head* deck, we did not rely on any one particular study of letter frequencies. Why? Because there are many studies to choose from, and — despite the fact that they all claim to be authoritative — they all more or less disagree with each other. To get at the truth, we decided to make use of the information found in *all* of the best studies: we *averaged together* 16 different reliable analyses of letter frequencies. By so doing, we greatly broadened our statistical base and the resulting combined totals pinpoint with a very high degree of accuracy the true frequency-of-occurrence of all letters in the English language. The chart on the next page shows the totals of all the original studies and our averaged results.

Frequency Distribution Analysis

Total number of each letter's occurrences in the English language per 1000 letters

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Total
i	E	T	A	O	N	I	R	S	H	D	L	U	C	M	P	F	Y	W	G	B	V	J	K	X	Q	Z	
	130	90	80	80	70	65	65	60	60	40	35	30	30	30	20	20	20	15	15	15	10	5	5	5	2.5	2.5	1000
ii	E	T	O	A	N	I	R	S	H	D	L	U	C	M	P	F	Y	W	G	B	V	K	J	X	Z	Q	
	128	85	81	78	69	67	65	62	59	40	37	31	30	29	22	20	20	18	17	14	11	7.4	5.7	2.7	1.7	.8	1001.3
iii	E	T	O	A	N	I	R	S	H	D	L	C	F	U	M	P	Y	W	G	B	V	K	X	J	Q	Z	
	130.5	90.2	82.1	78.1	72.8	67.7	66.4	64.6	58.5	41.1	36.0	29.3	28.8	27.7	26.2	21.5	15.1	14.9	13.9	12.8	10.0	4.2	3.0	2.3	1.4	0.9	1000
iv	E	T	A	O	I	N	S	R	H	D	L	C	U	M	W	F	Y	G	P	B	V	K	X	J	Q	Z	
	127	91	82	75	70	67	63	62	61	43	40	28	28	24	23	22	20	20	19	15	10	8	2	1	1	1	1003
v	E	T	R	I	N	O	A	S	D	L	C	H	F	U	P	M	Y	G	W	V	B	X	K	Q	J	Z	
	120.7	90.6	83.5	76.5	76.5	74.4	72.4	58.3	40.2	36.2	33.2	33.2	30.2	30.2	27.2	25.2	21.1	18.1	14.1	13.1	11.1	5	3	3	2	1	1000
vi	E	T	A	O	N	R	I	S	H	D	L	F	C	M	U	G	Y	P	W	B	V	K	X	J	Q	Z	
	131	105	82	80	71	68	63	61	53	38	34	29	28	25	25	20	20	20	15	14	9	4	2	1	1	.7	999.7
vii	E	T	N	R	O	A	I	S	D	L	H	C	F	P	U	M	Y	G	W	V	B	X	Q	J	K	Z	
	130	92	79	76	75	74	74	61	42	36	34	31	28	27	26	25	19	16	16	15	10	5	3	3	2	1	1000
viii	E	T	A	I	N	O	S	H	R	D	L	U	C	M	F	W	Y	P	G	B	V	K	J	Q	X	Z	
	107	82	78	75	74	72	72	57	56	42	38	32	30	29	25	20	20	18	18	17	13	9	6	5	5	2	1002
ix	E	A	I	O	R	T	N	S	L	C	U	M	P	D	H	G	Y	B	F	K	V	W	X	Z	J	Q	
	102.5	91.3	85.4	73.6	71.8	70.9	64.7	59.0	58.1	46.9	34.5	34.0	31.9	29.9	28.8	23.0	22.3	21.6	12.0	10.0	9.7	8.2	3.3	2.7	2.0	1.8	999.9
x	E	T	A	I	R	O	N	S	L	H	C	D	U	P	M	G	Y	F	W	B	V	K	X	J	Q	Z	
	103.6	91.1	86.1	85.0	74.1	72.9	63.7	57.8	52.8	44.0	37.1	35.2	28.3	25.8	25.1	24.5	21.4	18.2	17.6	11.9	11.3	5.7	3.1	1.9	1.3	1.0	1000.5
xi	E	T	A	O	N	I	R	S	H	D	L	C	U	M	P	F	Y	W	G	B	V	K	J	X	Q	Z	
	130	90	80	80	70	65	65	60	60	40	35	30	30	30	20	20	20	15	15	15	10	5	5	5	2.5	2.5	1000
xii	E	T	A	O	N	I	S	R	H	L	D	C	U	P	F	M	W	Y	B	G	V	K	Q	X	J	Z	
	123	96	80	79	72	72	66	60	51	40	37	32	31	23	23	22	20	19	16	16	9	5	2	2	1	.9	998
xiii	E	T	N	R	O	I	A	S	D	H	L	C	F	U	P	M	Y	W	G	V	B	X	K	Q	J	Z	
	130	93	78	77	74	74	73	63	44	35	35	30	28	27	27	25	19	16	16	13	9	5	3	3	2	1	1002
xiv	E	T	A	O	N	I	H	S	R	D	L	U	M	W	C	F	Y	G	P	B	V	K	X	J	Q	Z	
	126.2	90.6	81.4	77.4	71.2	68.2	66.2	63.1	62.1	46.8	36.6	27.5	25.4	23.4	22.4	22.4	20.4	20.4	16.3	13.2	8.1	7.1	1	1	1	0.5	999.9
xv	E	T	A	O	N	I	R	S	H	D	L	U	C	M	P	F	Y	W	G	B	V	K	J	X	Z	Q	
	127.5	98.0	68.6	68.6	68.6	68.6	68.6	68.6	49.0	29.4	29.4	29.4	29.4	29.4	29.4	29.4	19.6	19.6	19.6	19.6	4.9	4.9	4.9	4.9	4.9	4.9	1000
xvi	E	T	A	O	N	R	I	S	H	D	L	F	C	M	U	G	Y	P	W	B	V	K	X	J	Q	Z	
	125.6	96.6	77.3	77.3	67.6	67.6	58.0	58.0	48.3	38.6	29.0	29.0	29.0	29.0	19.3	19.3	19.3	19.3	14.5	14.5	14.5	9.7	9.7	9.7	9.7	9.7	1000
Averaged Totals																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Total
Rank	**E**	**T**	**A**	**O**	**N**	**I**	**R**	**S**	**H**	**D**	**L**	**C**	**U**	**M**	**F**	**P**	**Y**	**G**	**W**	**B**	**V**	**K**	**X**	**J**	**Q**	**Z**	
Total	1972.6	1452.0	1262.2	1222.3	1135.1	1134.4	1088.1	997.4	798.0	627.2	608.1	496.3	456.9	433.3	385.0	367.4	316.2	291.8	270.3	229.7	171.6	93.0	63.7	53.5	43.9	34.0	
Avg.	**123.3**	**90.8**	**78.9**	**76.4**	**70.9**	**70.9**	**68.0**	**62.3**	**49.9**	**39.2**	**38.0**	**31.0**	**28.6**	**27.1**	**24.1**	**23.0**	**19.8**	**18.2**	**16.9**	**14.4**	**10.7**	**5.8**	**4.0**	**3.3**	**2.7**	**2.1**	**1000.3**

Refer to page 5 for bibliographic information for the 16 studies, labeled i-xvi, summarized here.

Many studies do not give numerical analyses of the individual letters, but simply present them ranked in order of frequency. To make our survey even more thorough, we also mathematically combined 13 of these studies, making use of both the mean and the average of each letter's ranking. The final averaged ranking of these less authoritative studies agreed exactly with the ranking of the chart on the previous page, confirming our findings.

Here are the final totals, giving the frequencies of the letters of the alphabet expressed as percentages — the number of times each letter will occur out of a random sampling of 100 letters.

Final Totals

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Total
	E	**T**	**A**	**O**	**N**	**I**	**R**	**S**	**H**	**D**	**L**	**C**	**U**	**M**	**F**	**P**	**Y**	**G**	**W**	**B**	**V**	**K**	**X**	**J**	**Q**	**Z**	
%	12.33	9.08	7.89	7.64	7.09	7.09	6.80	6.23	4.99	3.92	3.80	3.10	2.86	2.71	2.41	2.30	1.98	1.82	1.69	1.44	1.07	.58	.40	.33	.27	.21	100%

The number of *Letter Head* cards depicting each letter and the point values assigned to each letter are derived as closely as possible from these final totals, which are an extremely accurate approximation of the frequency of each letter in the English language.

The *Letter Head* deck has 120 letter cards (plus 4 wild cards). The chart below reveals how the distribution of letters within the deck was calculated:

Calculation of *Letter Head* letter distributions and point values

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Total
	E	**T**	**A**	**O**	**N**	**I**	**R**	**S**	**H**	**D**	**L**	**C**	**U**	**M**	**F**	**P**	**Y**	**G**	**W**	**B**	**V**	**K**	**X**	**J**	**Q**	**Z**	
	12.33	9.08	7.89	7.64	7.09	7.09	6.80	6.23	4.99	3.92	3.80	3.10	2.86	2.71	2.41	2.30	1.98	1.82	1.69	1.44	1.07	.58	.40	.33	.27	.21	100%
x 1.2																											x 1.2
	14.80	10.90	9.47	9.17	8.51	8.51	8.16	7.48	5.99	4.70	4.56	3.72	3.43	3.25	2.89	2.76	2.38	2.18	2.03	1.73	1.28	.70	.48	.40	.32	.25	
cards	15	11	9	9	8	8	8	7	6	5	4	4	3	3	3	3	2	2	2	2	1	1	1	1	1	1	120

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Total
	E	**T**	**A**	**O**	**N**	**I**	**R**	**S**	**H**	**D**	**L**	**C**	**U**	**M**	**F**	**P**	**Y**	**G**	**W**	**B**	**V**	**K**	**X**	**J**	**Q**	**Z**	
12.33 ÷																											12.33 ÷
	12.33	9.08	7.89	7.64	7.09	7.09	6.80	6.23	4.99	3.92	3.80	3.10	2.86	2.71	2.41	2.30	1.98	1.82	1.69	1.44	1.07	.58	.40	.33	.27	.21	100%
	1.00	1.36	1.56	1.61	1.74	1.74	1.81	1.98	2.47	3.15	3.24	3.98	4.31	4.55	5.12	5.36	6.23	6.77	7.30	8.56	11.52	21.26	30.83	37.36	45.67	58.71	
pts.	1	1	2	2	2	2	2	2	2	3	3	4	4	5	5	5	6	7	7	9	12	13	15	18	20	25	

The first row shows how many letters would be in a deck of 100 cards — based exactly on the true frequency of each letter as determined by the analyses given earlier. All numbers are then multiplied by a factor of 1.2, and the resulting totals extended out to two decimal places are the precise number of cards of each letter that should be in a deck of 120 cards. Since it's not possible to have 2.76 P's, these exact frequencies are then rounded off to the nearest whole number. Fortunately, most letters are quite near to a whole number, and thus deviate only slightly from their natural frequencies. X, J, Q and Z were rounded up and assigned one card each, even though in theory they should all be rounded down to 0. To compensate for these extra letters, N, I and L were rounded down instead of up to keep the number of cards at 120. (A comparison of different possible deck sizes has revealed that 120 letter cards produces the least amount of variation from the actual frequencies; decks with 116, 124 or 128 cards all require greater statistical distortion.)

Point values for each letter are derived similarly. We start by calculating how much rarer than E each letter is. This is done simply by dividing E's frequency (12.33%) by each letter's frequency. This formula shows that S, for example, is almost precisely twice as rare as E — and if we check their frequencies, we see that E occurs a little over 12 times out of 100, while S occurs a little over 6 times out of 100, which confirms

our calculation. By then assigning the most common letter, E, a point value of 1, we can derive each letter's approximate point value by rounding off its "rarity quotient" to the nearest whole number. This process produces, as closely as possible, a value for each letter that accurately reflects its rarity and difficulty of play. As with our calculation of card totals, there is distortion with the rarest letters, but this time in the opposite direction. K, X, J, Q and Z all occur so infrequently that their point values would be unreasonably high. Since we rounded these letters up to give them card totals that actually exceed their true frequency of occurrence, we have diminished their point values accordingly. Assigning each of these ultra-rare letters point values somewhat lower than pure mathematics would dictate enables the game-action of *Letter Head* to run more smoothly, since the true outrageously high point values for J (37),Q (46) and Z (59) would give an unfair advantage to the player who randomly drew them during the game. This way, the curve generated by the point values of all the letters has been flattened out to increase the game's playability.

The preceding three pages of charts and calculations illustrate in detail how the card totals and point values of the *Letter Head* deck were determined through careful and thoughtful analysis, not whim or ignorance. But this whole process is invisible to the player, who only knows that the deck in his hand seems to always have the right amount of each letter, and that the scoring seems to make sense.

The chart below sums up all of our research, and is the basis for the contents of the *Letter Head* deck. In addition to the 120 letter cards, we have also included four wild cards, bringing the total size of the playing deck to 124 cards.

Official card totals and point values for *Letter Head*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
	E	**T**	**A**	**O**	**N**	**I**	**R**	**S**	**H**	**D**	**L**	**C**	**U**	**M**	**F**	**P**	**Y**	**G**	**W**	**B**	**V**	**K**	**X**	**J**	**Q**	**Z**	*
cards	**15**	**11**	**9**	**9**	**8**	**8**	**8**	**7**	**6**	**5**	**4**	**4**	**3**	**3**	**3**	**3**	**2**	**2**	**2**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**4**
pts.	***1***	***1***	***2***	***2***	***2***	***2***	***2***	***2***	***2***	***3***	***3***	***4***	***4***	***5***	***5***	***5***	***6***	***7***	***7***	***9***	***12***	***13***	***15***	***18***	***20***	***25***	*

* = wild card

References

i *The Codebreakers*, David Kahn, 1967

ii *Secret Messages*, P.B. Thomas, 1928

iii *Cryptanalysis*, Helen Fouché Gaines (frequencies proportionally adjusted to reflect number of occurrences per 1000 letters).

iv *Cyber Systems, The Protection of Communications*, H. Beeker and F. Piper, 1982 (frequencies proportionally adjusted to reflect number of occurrences per 1000 letters)

v *Encyclopedia Britannica* cryptology article (frequencies proportionally adjusted to reflect number of occurrences per 1000 letters)

vi *Cloak and Cypher*, D. Moore, 1962

vii *The Code Book*, Michael Marotta, 1979

viii E.C. Brewer, 1888

ix Computerized frequency analysis of the entire *Ispell Dictionary* by Rich Lawson, 1998

x Statistical frequency analysis of the *Oxford English Dictionary* by H.R. Lawson, 1998

xi anonymous untitled study, 1967 (frequencies proportionally adjusted to reflect number of occurrences per 1000 letters)

xii Meaker, 1939

xiii unattributed source found on the Internet, 1997

xiv Study giving letter frequencies in the entirety of *A Tale of Two Cities*, Charles Dickens (frequencies proportionally adjusted to reflect number of occurrences per 1000 letters).

xv *Laughlin's Fact Finder*, 1969 (frequencies proportionally adjusted to reflect number of occurrences per 1000 letters)

xvi *Secret Codes and Ciphers*, Bernice Kohn, 1968 (frequencies proportionally adjusted to reflect number of occurrences per 1000 letters)