WORD-SQUARE SUPPORT: PART 2

A. ROSS ECKLER
Morristown, New Jersey

This is the second of two articles examining the statistical characteristics of the support for a word square (the sample size needed to produce, on the average, a single word square). In the preceding article, published in the August 1993 **Word Ways**, a method was developed for calculating the support from small sample sizes - ones large enough to typically produce only a few squares, rather than the thousands or millions found by Chris Long. In the course of this investigation, it was discovered that the support level depended on the stockpile of words being sampled. In order to explore this matter further, Leonard Gordon classified the words in the stockpile according to their vowel-consonant patterns: BRACE ccvcv, EIDER vvcvc. (The letter Y was classified as a vowel unless it occurred at the beginning of a word.) The table below, typical of several generated, demonstrates that words beginning with consonants are underrepresented in word squares, and words beginning with vowels are overrepresented. The second column gives the total number of times a word with a specified vc-pattern appeared in a word square (in any position), and the third column, the number of different words with a specified vc-pattern appearing in the sample. The fourth column compares these two

Pattern	Squares	Sample	Ratio	Code	Pattern	Squares	Sample	Ratio	Code
ccccc				A	vcccc	14	1		a
ccccv				B	vcccv	330	13	0.46	b
cccvc	300	12	0.47	C	vccvc	2590	86	0.39	c
cccvv	44	3		D	vccvv	244	12	0.58	d
ccvcc	2515	349	1.63	E	vcvcc	1365	46	0.40	e
ccvcv	1611	134	0.98	F	vcvcv	3743	48	0.15	f
ccvvc	1260	155	1.45	G	vcvvc	574	19	0.39	g
ccvvv				H	vcvvv				h
cvccc	2216	295	1.57	I	vvccc	22	4		i
cvccv	2692	321	1.40	J	vvccv	61	2	0.39	j
cvcvc	6020	422	0.98	K	vvcvc	355	14	0.49	k
cvcvv	592	36	0.72	L	vvcvv	99	2	0.24	l
cvvcc	884	226	3.01	M	vvvcc				m
cvvcv	624	81	1.53	N	vvvcv				n
cvvvc	98	17	2.04	O	vvvvc				o
cvvvv	2	2		P	vvvvv				p

mixes; if, for example, 6 per cent of the words in the sample had pattern vcvcv and 20 per cent of the words appearing in the squares had the pattern vcvcv, then vcvcv is underrepresented in the sample by a factor of 6/20, or 0.30. There were 5647 squares gen-

erated by a sample of 2400 five-letter words. Note the preponderance of the vcvcv and cvcvc patterns in the squares formed; more than one-third of all words in the squares consisted of these two vc-patterns.

To learn more about the usefulness of different vc-patterns in forming word squares, Leonard Gordon tabulated the number of word squares of each pattern formed, out of a total of 4488 squares generated by 1928 words. Each square was identified by a five-letter sequence taken from the Code column in the table above; for example, a word square consisting of words with the vc-patterns depicted at the left would be encoded KfKfK.

cvcvc
vcvcv
cvcvc
vcvcv
cvcvc

Not surprisingly, there are many different five-letter sequences characterizing word squares. The commonest ones encountered in the sample are listed below.

KfKfK 259	JfKEj 71	cNKfK 57	JgKNc 45	EJcFK 37
KfKeJ 128	KfKeI 68	cFKfK 56	IfKFK 43	KfIbL 37
KfKfL 117	KfJcM 60	KfJcN 53	EKfJG 42	JfKFk 33
JeKFc 100	FKcNc 58	IfKEJ 52	JeKFd 40	KfNcN 32

The support associated with the sample is equal to 359, consistent with Chris Long's experimental value of 350. Suppose that one were to use only the vcvcv and cvcvc patterns to generate five-squares. 422 cvcvc pattern words and 48 vcvcv pattern words combine to yield 259 KfKfK squares and 7 fKfKf squares, for a total of 266. This yields a support value of $(422 + 48)/(266)^{1/5}$, or 154, a more than twofold improvement. The mix of patterns can make a large different in the support.

Suppose that one enriches the vc-pattern mix by drawing on vowel-starting words from a larger stockpile than the one for consonant-starting words. If one doubles the vcvcv pattern words to 96, one will, on the average, multiply by four the number of KfKfK squares, and multiply by eight the number of fKfKf squares, yielding a new support of $(422 + 96)/(4\text{x}259 + 8\text{x}7)^{1/5}$, or 128. A second doubling (192 vcvcv pattern words) lowers the support to 114, and a third doubling (384 vcvcv pattern words), to 113. Further doublings raise the support; one concludes that the optimum mix is in the ratio 1:1 for consonant-starting versus vowel-starting words. Using only 22 per cent of the original stockpile, and augmenting this with a like number of words from a larger stockpile, one has lowered the support to less than one-third!

A more realistic approach to lowering the support by enriching the stockpile is to increase the total number of vowel-starting words relative to the consonant-starting ones. For the five-square such a strategy can be easily evaluated. There are 253 vowel-starting and 1675 consonant-starting words; doubling the former lowers the support to 308, doubling it again, to 279, and doubling it a third time, to 275. The improvement is more modest: 0.77 instead of 0.31.

If one turns to six-squares a similar story emerges. The relevant table is found on the next page.

Pattern	Squares	Sample	Ratio	Code
ccccvc		1		C
ccccvv	4	2		D
cccvcc	114	43	1.49	E
cccvcv	145	25	0.68	F
cccvvc	113	18	0.63	G
cccvvv		1		H
ccvccc	473	143	1.20	I
ccvccv	674	191	1.12	J
ccvcvc	1207	322	1.06	K
ccvcvv	109	24	0.87	L
ccvvcc	314	115	1.45	M
ccvvcv	329	83	1.00	N
ccvvvc	15	13	3.43	O
cvcccc	32	18	2.23	Q
cvcccv	327	178	2.16	R
cvccvc	3631	1375	1.50	S
cvccvv	282	156	2.23	T
cvcvcc	1540	559	1.44	U
cvcvcv	2177	460	0.84	V
cvcvvc	528	303	2.27	W
cvcvvv	57	17	1.18	X
cvvccc	49	61	4.93	a
cvvccv	304	97	1.26	b
cvvcvc	1271	332	1.03	c
cvvcvv	101	43	1.69	d
cvvvcc	25	28	4.44	e
cvvvcv	28	10	1.41	f
cvvvvc	6	9		g
cvvvvv		1		h

Pattern	Squares	Sample	Ratio	Code
vccccvc	270	71	1.04	k
vcccvv	18	5	1.10	l
vccvcc	855	196	0.91	m
vccvcv	2210	140	0.25	n
vccvvc	286	80	1.11	o
vccvvv	40	4	0.40	p
vcvccc	153	25	0.65	q
vcvccv	745	40	0.21	r
vcvcvc	2051	185	0.36	s
vcvcvv	267	17	0.25	t
vcvvcc	195	24	0.49	u
vcvvcv	565	19	0.13	v
vcvvvc	10	6	2.38	w
vvcccc	15	5	1.32	
vvcccv	8	2		
vvccvc	96	31	1.28	
vvccvv	6	6		
vvcvcc	96	21	0.87	
vvcvcv	125	20	0.63	
vvcvvc	36	15	1.65	
vvvcvc	64	5	0.31	

As before, most of the ratios for consonant-starting words are greater than one, and ost of the ratios for vowel-starting words are less than one, indicating a deficiency of the latter in the sample. There were 3661 squares generated by a sample of 5546 six-letter words (out of a total stockpile of 25915). From these figures, one can calculate a support of $5546/(3661)^{1/6} = 1408$. This is significantly less than the support of 1726 estimated for the 6-square in the article in the August **Word Ways.** The earlier support was based on common six-letter words, which evidently have less pattern-variety than do rarer six-letter words.

The commonest of the 6-squares are listed below:

UsVsVK	89	UsVsUI	35	SnKVsU	30	SnMcnS	27	VsSnbo	25
SnJSnc	84	VsVsVs	34	SrUKnS	30	VsUrSm	26	UsVsUJ	24
SnISnS	82	SnKSvS	31	SvUcnS	28	KVncnc	26	KVqSnS	24
SmJSnK	66	SvSVsU	31	IVsSNS	28	SmJWnK	26	SvVcnc	23

If one uses only the vcvcvc and cvcvcv patterns to generate 6-squares, then 185 + 460 = 645 words generate 34 VsVsVs and one sVsVsV squares, leading to a support estimate of $645/(35)^{1/6} = 357$, four times better than before. A doubling of the vcvcvc pattern words from 185 to 370 lowers the support to $830/(280)^{1/6} = 324$,

and a further doubling to $1200/(2240)^{1/6} = 331$, lowering the support to about 23 per cent of its original value. This is an even more dramatic improvement than the one achieved for the 5-square.

It is more realistic to enrich all of the vowel-starting words and settle for a more modest reduction in the support. Leonard Gordon conducted two such experiments, one in which he tripled the sample size of vowel-starting words, and another in which he took all words in this 25915-word stockpile having vowel beginnings and endings, but used only one-sixth of his remaining stockpile. In the first experiment, he found that 4831 words generated 3505 squares for a support estiate of 1236; in the second, 4884 words generated 8397 squares for a support estimate of 1083. These results are similar to the ones achieved if one is restricted to the 3903 words in the sample (out of an original 5546) with patterns ccvccc, ccvccv, ccvcvc, cvccvc, cvcvcc, cvcvcv, cvvcvc, vccvcc, vccvcv and vcvcvc. 706 squares using only these patterns were found, and the support was estimated to be 1308, a decidedly modest improvement of 1408. However, doubling the number of vowel-starting words lowers the support to 1137, redoubling the number, to 1055, and re-redoubling the number, to 1072.

What does all this say about the search for a 10-square? Although any ten-letter word added to the stockpile increases the probability that a 10-square will be found, vowel-starting words increase this probability somewhat more rapidly than do consonant-starting ones. This becomes important when computer processing time (which increases, probably, as the tenth power of saple size) begins to play a limiting role. If one can find a 10-square with, say, 200,000 words from a vowel-starting-enriched stockpile, it will take far less computer time to process these for a potential 10-square than it will using an unenriched stockpile of 250,000 or more. Where these additional vowel-starting words will come from is another matter; there may not be a sufficient number available to make a 10-square likely. Two possibilities are place names in the Times Index-Gazetteer (augmented by the Omni Gazetteer, mentioned in the February 1992 **Word Ways**), and ten-letter US surnames taken from telephone directories (use PhoneDisc, mentioned in the May 1991 **Word Ways**).