WORD-SQUARE SUPPORT: PART 1

A. ROSS ECKLER
Morristown, New Jersey

In "How Many Words Support a Square?" in the May 1992 **Word Ways**, I defined the **support** of a word square to be the number of words required, on the average, to form a single word square, and showed that it could be experimentally calculated by means of the scaling formula

$$\text{Support} = (\text{number of words used})/(\text{number of squares found})^{1/n}$$

where n is the word-length (the square size). In "Mathematics of Square Construction" in the February 1993 **Word Ways**, Chris Long derived experimental values of approximately 90 and 350 for 4-squares and 5-squares, respectively, based on computer runs involving typically thousands of words and millions of squares. (He also derived theoretical values for the support of 63 and 250, about 70 per cent as large, based on a highly-idealized model of word formation which assumes letter-frequencies to be probabilistically and positionally independent.)

It is the purpose of this article to examine the statistical variations in the support when the number of words is so small that only a few squares can be expected to result. A successor article will show how the experimental value of the support is strongly influenced by the nature of the word-stock, in particular the mix of words having various vowel-consonant patterns. I am indebted to Leonard Gordon for the computer runs generating the statistics cited below, as well as some of the insights provided.

The data below were generated by drawing samples of various sizes from a stockpile of 1512 common four-letter words. The first column lists the number of squares found, and the second and succeeding columns list the number of times a sample of n words was drawn yielding that number of squares.

Plotting the average number of squares against the sample size, one finds that this is equal to 1 for a sample size of approximately 131, which can then be termed the experimental support for a 4-square relative to this stockpile of words. Note that it is not the same as the experimental support of 90 discovered by Chris Long using a different stockpile; this discrepancy, which cannot be explained away by statistical fluctuations (such as 5 heads in one throw of 10 coins, and 7 heads in the next throw) was the motivation leading to the successor article.

If one substitutes n=131 into the scaling formula, one can calculate the expected number of squares found (call this **a**) for different sample sizes. The formula 1 - exp(-.7a) turns out to

n	100	110	120	130	140	150
0	162	141	140	207	89	65
1	20	37	32	91	53	44
2	12	15	14	49	18	30
3	5	6	6	31	16	23
4	1	1	4	9	11	14
5			3	4	5	8
6			1	3	1	4
7					2	5
8						7
9				1	3	
12				1	1	
17					1	
Total	200	200	200	400	200	200
Average	0.32	0.45	0.58	1.02	1.41	1.84

n	100	110	120	130	140	150
a	0.34	0.50	0.70	0.97	1.30	1.72
Observed Prob	.19	.29	.30	.48	.55	.68
1-exp(-.7a)	.21	.30	.39	.49	.60	.70

be a good approximation for the observed probability (from the table above) that one or more squares will be found. The one anomalous value is the observed probability of forming one or more squares from a sample of size 120; it is unlikely to be so close to the corresponding value for a sample of size 110 (.30 vs. .29).

To gain further insight into the probability of square-formation from small sample sizes, Leonard Gordon ran a more detailed study. He drew 60 words at random and noted if any squares could be constructed from them. He then drew additional words one at a time, each time attempting to form squares with the augmented set, until he had sampled 170 words. Repeating this process 100 times, he found the following probabilities that he had formed at least one square by the time the nth sample value was included.

n	100	110	120	130	140	150
Observed Prob	.21	.32	.44	.53	.64	.75

These are in reasonable agreement with the earlier observed probabilities (except for the previously-mentioned anomalous value). One can with some degree of confidence use the empirical rule to assess the likelihood of a square being formed with a small number of words. For example, if one draws a sample of 69, there is a probability of 1 - exp(-.7(.077)) = 0.05 that a square will be found. In the light of this prediction, it seems that I was extraordinarily lucky to find a square (actually, a set of four) after sampling only 68 words from Kucera and Francis (see the May 1992 **Word Ways**). However, it may well be that the Kucera-Francis list differed in vowel-consonant patterns from the Gordon sampling; the implications of such differences are explored in the next article.

More to the point, it would be of interest to know to what extent the empirical formula may be relied upon to predict success with much larger squares - say, of size 10. Using Chris Long's theoretical scaling rule, one might guess that there is a 5 per cent chance of success at finding a 10-square if the sample size is (69/63)(247718) = 271000. However, as will be demonstrated in the next article, there may be somewhat greater hope. Comprehensive word lists seem to generate lower supports than do common ones.

The story for 6-squares is similar to the one for 4-squares. From a sample of 9058 6-letter words, Leonard Gordon drew 195 samples ranging from 1370 to 1909 in size, determining for each the number of squares that could be formed. Using techniques similar to the ones for the 4-square analysis, I estimate a support value of 1726.

Sample Range	Average	Number of Squares Found							Average	Ln
		0	1	2	3	4	5	6		
1370-1499	1446	25	5	3	0	0	0	0	0.33	-1.10
1500-1599	1558	25	8	4	0	0	1	0	0.55	- .59
1600-1679	1637	24	10	2	3	0	1	0	0.70	- .36
1680-1749	1714	18	14	4	3	2	0	0	1.07	.07
1750-1909	1800	16	13	8	2	1	0	2	1.21	.19

Note that the ratio of the theoretical Long value of the support to this experimental value is 992/1726 = 0.57, a little larger than the corresponding value of 63/131 = 0.48 for the 4-square. This suggests that the mix of various vowel-consonant patterns has changed; the support increase due to word-commonness is less marked. There is a probability of 0.05 that a sample of size 1120 will yield at least one 6-square; the corresponding prediction for a 10-square has increased to (1120/992)(247718) = 280000.

Additional evidence for this support value for the 6-square based on common words is provided by a study by Leonard Gordon in which 6403 words were sampled from 9058 common ones (out of a full stockpile of 25915 6-letter words). This sample yielded 2888 squares and a support of $6403/2888^{1/6} = 1692$, in fine agreement with the 1726 above.

APPENDIX

The following table shows how additional 4-squares were formed as words were added one at a time to the sample from 60 to 170. The first line reads as follows: when the 164th word was added to the sample, two squares were found; when the 167th word was added, three more squares were found. If the number is 60, it means that the parenthesized number of squares was found before the 60th word was added to the sample.

164(2)167(3)
147(8)156(8)159(2)167(4)
168(4)
102(1)137(1)142(1)169(1)
60(2)96(1)125(1)129(1)142(2)144(1)
  146(1)149(1)160(1)
159(1)165(2)
60(1)169(1)
105(1)144(1)160(2)
170(0)
170(0)
138(1)161(1)162(1)
142(1)165(2)169(1)
123(1)137(1)

87(1)137(1)163(2)
95(1)109(2)126(1)153(1)
133(1)155(1)166(1)
66(1)118(1)170(3)
60(1)
133(4)138(1)169(1)
105(1)117(2)136(2)
153(2)
136(1)158(1)
141(2)157(1)
170(0)
113(1)125(1)146(1)
138(3)143(1)145(3)148(1)154(1)
 156(3)158(3)160(2)168(3)
128(1)
110((1)134(1)155(1)162(1)
163(3)165(4)
170(0)
129(4)142(1)158(1)
122(1)128(1)135(1)160(1)161(1)
 165(1)
124(3)126(1)159(1)167(1)
170(0)
118(1)125(2)126(2)
94(1)164(1)165(1)
170(0)
141(1)
165(1)
170(0)
110(3)120(1)122(1)
146(1)167(2)170(1)
170(0)
124(3)164(1)
113(1)128(1)131(1)137(1)140(1)
 142(1)144(1)162(1)165(1)
74(1)120(1)130(1)
96(1)133(1)167(6)168(1)
60(1) 76(1)85(1)95(1)115(1)127(1)
 139(1)162(1)165(1)
85(1)110(1)125(1)127(1)146(1)
120(1)124(1)134(1)
155(1)
109(1) 150(1)159(1)
113(4)123(1)126(4)166(1)
64(1)72(1)119(1)159(1)163(1)
134(1)136(1)157(1)
156(1)
134(2)135(3)140(1)157(1)160(1)170(1)
159(1)
110(2)150(1)154(1)
101(2)168(1)
116(1)146(1)154(1)156(2)161(1)
 164(2)165(1)
161(1)
142(1)154(1)
113(1)169(2)
82(1)117(2)137(2)166(2)
134(1)157(1)
73(1)136(1)
118(2)143(1)163(1)165(2)
108(1)
143(1)
170(0)
109(1)120(1)163(1)165(10167(1)
127(1)150(1)
122(1)127(1)152(1)165(3)
141(2)170(2)
134(2)156(1)
170(0)
122(1)161(2)
145(4)
111(1)120(1)149(1)
68(1)120(1)151(1)159(1)163(1)
95(1)115(1)116(2)137(1)160(1)
141(1)
170(0)
113(2)121(1)134(1)154(2)
117(1)133(1)135(1)146(1)152(1)168(1)
60(1)153(1)154(4)155(1)158(2)
165(1)
156(1)
99(1)114(1)140(1)153(1)164(1)
65(1)98(1)160(3)
134(4)161(1)169(3)
103(1)168(1)
160(1)167(1)
116(2)169(2)
82(3)100(1)126(3)129(1)145(2)157(1)
 159(1)167(1)
80(1)99(1)100(1)112(1)125(4)142(1)
170(0)
132(2)136(1)140(1)151(1)163(1)
147(6)151(3)155(1)161(1)164(1)168(1)