"BORN TO RUN" … CHI-SQUARE TESTS: MATHEMATICS IN MUSIC

by

Paul M. Sommers

March 2010

MIDDLEBURY COLLEGE ECONOMICS DISCUSSION PAPER NO. 10-07

DEPARTMENT OF ECONOMICS
MIDDLEBURY COLLEGE
MIDDLEBURY, VERMONT 05753

http://www.middlebury.edu/~econ

"BORN TO RUN" … CHI-SQUARE TESTS: MATHEMATICS IN MUSIC

by

Paul M. Sommers

Department of Economics
Middlebury College
Middlebury, Vermont 05753
psommers@middlebury.edu

"BORN TO RUN" … CHI-SQUARE TESTS:
MATHEMATICS IN MUSIC

If everyone writes differently, then songwriters should be no exception. For example, Bruce Springsteen's writing style is probably different from Michael Jackson's. And, one could use, say, the relative frequency of letters in their song lyrics to show their possible similarities or more probable differences.

In this brief note, we compare letter frequencies (as well as the frequencies of vowels and consonants) in Bruce Springsteen's *Greatest Hits* album (released February 27, 1995 on Columbia Records) and the best-selling album of all time, Michael Jackson's *Thriller* (released November 30, 1982 on Epic Records). The relative frequencies of letters in the English language are based on the work of mathematician and cryptologist Robert E. Lewand [1, p. 36].[1] The program used to count letters in a text is www.amstat.org/publications/jse/secure/v7n2/count-char.cfm .[2] The source of Springsteen's lyrics on his *Greatest Hits* album (and audio snippets of the eighteen songs) can be found at www.brucespringteen.net/albums/gh.html . And, the source of Michael Jackson's lyrics on his *Thriller* album can be found at www.songlyrics.com/jackson-michael/thriller . *Thriller* won eight Grammy Awards at the 1984 Grammys, including "Album of the Year."

Table 1 shows the distribution of letters, vowels, and consonants in Springsteen's "Born to Run", the first track on his *Greatest Hits* album. (No distinction is made between lower case letters and upper case; for example, "*y*" and "*Y*" are treated alike as two "y"s.) How well does the observed frequency distribution conform to Lewand's twenty-six letter expected frequencies? The null hypothesis is that the observed distribution can in fact be represented by Lewand's relative frequencies (multiplied by 1178, the total number of letters in the song's lyrics), and that any discrepancies between them are due to chance. The value of the test statistic is:

$$\chi^2_{Calculated} = \sum_{i=1}^{26} \frac{(O_i - E_i)^2}{E_i} = 70.703 .$$

Since $\chi^2_{.001,25} = 52.62$, we can reject the null hypothesis. We note, for example, that there are disproportionately many (few) "*w*"s ("*f*"s). But, when all vowels (*a*, *e*, *i*, *o*, *u*, and *y*) and consonants are separately summed, the observed frequency in each of the two groups (472 vowels, 706 consonants) are practically *identical* to the expected frequency. Springsteen's "Streets of Philadelphia" (see Table 2) actually does follow Lewand's individual letter relative frequencies (that is, $\chi^2_{Calculated} = 34.164$, $p = .104$) and it is the only one of the eighteen tracks on Springtseen's *Greatest Hits* album that does. Table 3 summarizes the vowel and consonant comparisons for each of the album's songs. Only one track (namely, "Murder Incorporated") has disproportionately many vowels (and few consonants), in large part due to the twenty-six times Springsteen uses the triplet "*y*", "*o*", "*u*" or "*you*".

The lyrics to Michael Jackson's title track "Thriller" are analyzed letter by letter in Table 4. The calculated χ^2 is 166.949 (and obviously exceeds $\chi^2_{.001,25} = 52.62$). Yet, the number of vowels (consonants) account for about 40 (60) percent of the 1683 letters in the title song's lyrics. Table 5 shows that the single "Thriller" is the exception. For seven of the nine tracks, Jackson's lyrics are top-heavy with vowels. Curiously, the two exceptions, "Thriller" and "Human Nature", were *not* written by Michael Jackson.[3]

Concluding Remarks

The most notable difference between Bruce Springsteen's *Greatest Hits* and Michael Jackson's songs on his *Thriller* album is Jackson's predilection for vowels.

The chi-square tests employed in this note to compare theoretical letter frequencies as well as the theoretical 40-60 split between vowels and consonants could be used, say, to compare compositions by Paul McCartney and those by fellow Beatle John Lennon, not to mention those co-written by Lennon and McCartney but where there is (still) some dispute as to who should have been listed first.

Table 1. Distribution of letters, vowels and consonants in Bruce Springsteen's "Born to Run"

Letter	Observed frequency (O_i)	R.E. Lewand's relative frequency (p_i)	Expected frequency ($E_i = 1178 \cdot p_i$)
a	85	.08167	96.207
b	28	.01492	17.576
c	19	.02782	32.772
d	49	.04253	50.100
e	144	.12702	149.630
f	11	.02228	26.246
g	26	.02015	23.737
h	57	.06094	71.787
i	82	.06966	82.059
j	0	.00153	1.802
k	15	.00772	9.094
l	56	.04025	47.415
m	27	.02406	28.343
n	83	.06749	79.503
o	86	.07507	88.432
p	17	.01929	22.724
q	0	.00095	1.119
r	81	.05987	70.527
s	68	.06327	74.532
t	102	.09056	106.680
u	43	.02758	32.489
v	14	.00978	11.521
w	53	.02360	27.801
x	0	.00150	1.767
y	32	.01974	23.254
z	0	.00074	0.872
vowels*	472	.40074	472.072
consonants	706	.59925	705.917

*Vowels include *a*, *e*, *i*, *o*, *u*, and *y*.

Table 2. Distribution of letters, vowels and consonants in Bruce Springsteen's "Streets of Philadelphia"

Letter	Observed frequency (O_i)	R.E. Lewand's relative frequency (p_i)	Expected frequency ($E_i = 608 \cdot p_i$)
a	54	.08167	49.655
b	7	.01492	9.071
c	11	.02782	16.915
d	26	.04253	25.858
e	76	.12702	77.228
f	17	.02228	13.546
g	12	.02015	12.251
h	33	.06094	37.052
i	54	.06966	42.353
j	0	.00153	0.930
k	9	.00772	4.694
l	39	.04025	24.472
m	15	.02406	14.629
n	47	.06749	41.034
o	39	.07507	45.643
p	8	.01929	11.728
q	0	.00095	0.578
r	24	.05987	36.401
s	39	.06327	38.468
t	47	.09056	55.061
u	12	.02758	16.769
v	7	.00978	5.946
w	18	.02360	14.349
x	0	.00150	0.912
y	14	.01974	12.002
z	0	.00074	0.450
vowels*	249	.40074	243.650
consonants	359	.59925	364.344

*Vowels include *a*, *e*, *i*, *o*, *u*, and *y*.

Table 3. Bruce Springsteen's *Greatest Hits* Album

Song	Vowels		Consonants		Calculated	
	Observed	Expected	Observed	Expected	χ^2	*p*-value
"Born to Run"	472	472.072	706	705.917	<.001	>.999
"Thunder Road"	678	655.210	957	979.774	1.322	.250
"Badlands"	478	452.836	652	677.153	2.333	.127
"The River"	440	448.829	680	671.160	0.290	.590
"Hungry Heart"	197	181.936	257	272.060	2.081	.149
"Atlantic City"	474	466.862	691	698.126	0.182	.670
"Dancing in the Dark"	490	483.693	717	723.295	0.137	.711
"Born in the USA"	340	322.195	464	481.797	1.641	.200
"My Hometown"	388	379.100	558	566.891	0.348	.555
"Glory Days"	337	343.033	519	512.958	0.177	.674
"Brilliant Disguise"	424	402.744	581	602.246	1.871	.171
"Human Touch"	475	461.252	676	689.737	0.683	.408
"Better Days"	482	466.461	682	697.527	0.863	.353
"Streets of Philadelphia"	249	243.650	359	364.344	0.196	.658
"Secret Garden"	277	254.470	358	380.524	3.328	.068
"Murder Incorporated"	476	413.564	556	618.426	15.728	**<.001**
"Blood Brothers"	353	349.045	518	521.947	0.075	.785
"This Hard Land"	471	486.899	744	728.089	0.867	.352

Table 4. Distribution of letters, vowels and consonants in Michael Jackson's "Thriller"

Letter	Observed frequency (O_i)	R.E. Lewand's relative frequency (p_i)	Expected frequency ($E_i = 1673 \cdot p_i$)
a	95	.08167	136.634
b	12	.01492	24.961
c	32	.02782	46.543
d	53	.04253	71.153
e	201	.12702	212.504
f	32	.02228	37.274
g	39	.02015	33.711
h	121	.06094	101.953
i	123	.06966	116.541
j	0	.00153	2.560
k	15	.00772	12.916
l	108	.04025	67.338
m	25	.02406	40.252
n	95	.06749	112.911
o	144	.07507	125.592
p	9	.01929	32.272
q	0	.00095	1.589
r	138	.05987	100.163
s	107	.06327	105.851
t	170	.09056	151.507
u	68	.02758	46.141
v	10	.00978	16.362
w	17	.02360	39.483
x	0	.00150	2.510
y	53	.01974	33.025
z	6	.00074	1.238
vowels*	684	.40074	670.438
consonants	989	.59925	1002.545

*Vowels include *a*, *e*, *i*, *o*, *u*, and *y*.

Table 5. Michael Jackson's *Thriller* Album

Song	Vowels Observed	Vowels Expected	Consonants Observed	Consonants Expected	Calculated χ^2	*p*-value
"Wanna Be Startin' Somethin'"	1376	1119.668	1418	1674.305	97.917	**<.001**
"Baby Be Mine"	574	513.348	707	767.639	11.956	**<.001**
"The Girl Is Mine"	498	464.858	662	695.130	3.942	**.047**
"Thriller"	684	670.438	989	1002.545	0.457	.499
"Beat It"	800	758.200	1092	1133.781	3.844	**.050**
"Billie Jean"	815	771.825	1111	1154.156	4.029	**.045**
"Human Nature"	325	314.180	459	469.812	0.621	.431
"PYT (Pretty Young Thing)"	445	376.295	494	562.696	20.931	**<.001**
"The Lady in My Life"	455	388.718	515	581.273	18.858	**<.001**

Reference

1. R.E. Lewand, *Cryptological Mathematics*, Washington, DC: Mathematical Society of America, 2000.

Footnotes

1. The frequency table used here is of letters that appear in English documents of *all* types. The author is not aware of frequency tables of the letters that appear in specific types of documents, like song lyrics. And, one might expect to find possibly different frequency tables for Broadway show tunes and rock 'n' roll music.

2. The program counts characters (e.g., letters) and groups of characters (e.g., pairs or triplets of letters) in a text. Simply, use the **Copy** and **Paste** commands from the **Edit** menu of your browser.

3. "Thriller" was written by Rod Templeton (who also wrote "Baby Be Mine" and "The Lady In My Life") and "Human Nature" was written by John Bettis and Steve Porcaro.