An Inversion Formula for Putnam Data

Keith A. Brandt
Department of Mathematics, Computer Science, and Physics
Rockhurst University
Kansas City, MO 64110

Donald L. Vestal, Jr
Department of Computer Science, Mathematics, and Physics
Missouri Western State College
Saint Joseph, MO 64507

Below are the score and rank data from the 62nd William Lowell Putnam Mathematical Competition held in December, 2001. The chart is reproduced exactly as it is presented in the mailings from the Competition organizers.

Score	Rank	Score	Rank	Score	Rank	Score	Rank	Score	Rank
101	1	61	37	44	123	32	225.5	13	606
100	2	60	47.5	43	125	31	252	12	655
86	3	59	56	42	130.5	30	287.5	11	759
80	4.5	58	59.5	41	143.5	29	314.5	10	970
79	6	57	62.5	40	164	28	324.5	9	1130.5
77	7.5	55	64	39	181.5	26	330	8	1154
73	9	54	66	38	190.5	24	332.5	5	1162.5
72	11	53	69	37	195	23	340	4	1166
71	14	52	72.5	36.9	197	22	365	3	1181.5
70	16.5	51	78	36.8	198	21	414.5	2	1252
69	19	50	93.5	36.7	199	20	494	1	1469.5
68	23.5	49	109	36.6	200	19	555	0	2292
67	27.5	48	115	36	202	18	570.5		
66	29	47	118.5	35	204	17	575		
63	30	46	120.5	34	205	15	577		
62	32	45	122	33	209	14	586.5		

-Putnam Scores and Ranks-

Note that the chart does not include the frequencies of the scores. This is an interesting omission because the frequencies are used to determine the rankings. Suppose a student, Putnam Pat, scored a 20. According to the table above, Pat's is the 59th highest score, giving a rank of 494.

Can anything more be said? Our goal in this note is to recover the frequencies from the rankings. We first choose some notation and explain how the rankings are determined. For $i = 1, 2, \ldots, 76$ (76 being the number of different scores), let s_i denote the i^{th} score (where s_1 is the highest), let n_i be the number of students with score s_i, and let c_i be the number of students with score s_i or higher. Thus n_i is the i^{th} frequency and c_i is the i^{th} cumulative frequency. Finally, let r_i be the rank of the score s_i.

Since the rankings for the top three scores are 1, 2, and 3, it is clear that these scores were obtained by one student each. The rank 4.5 assigned to the fourth score of 80 indicates that more than one student had this score. In particular, 4.5 is the average of the places 4 and 5, so it follows that two students had this score. Here is the formula for the rankings:

$$r_i = \sum_{k=1}^{i-1} n_k + \frac{n_i + 1}{2} = c_{i-1} + \frac{n_i + 1}{2} \qquad 1 \le i \le 76. \tag{1}$$

For example, $r_6 = 7.5$ means that two students scored 77, since $7.5 = 6 + \frac{2+1}{2}$. We will derive a simple formula for n_i in terms of the r_i.

First note that $r_i - r_{i-1} = \frac{1}{2}(n_i + n_{i-1})$, so $n_i + n_{i-1} = 2(r_i - r_{i-1})$. Thus,

$$n_i - n_{i-2} = (n_i + n_{i-1}) - (n_{i-1} + n_{i-2}) = 2(r_i - 2r_{i-1} + r_{i-2}),$$

and similarly,

$$n_i + n_{i-3} = (n_i - n_{i-2}) + (n_{i-2} + n_{i-3}) = 2(r_i - 2r_{i-1} + 2r_{i-2} - r_{i-3}).$$

Continuing in this fashion, we obtain

$$n_i + (-1)^i n_1 = 2\left(r_i - 2r_{i-1} + 2r_{i-2} \mp \cdots + (-1)^i 2r_2 + (-1)^{i-1} r_1\right).$$

Finally, since $n_1 = 2r_1 - 1$, we have

$$\begin{aligned} n_i &= 2\left(r_i - 2r_{i-1} + 2r_{i-2} \mp \cdots (-1)^i 2r_2 + (-1)^{i-1} 2r_1\right) + (-1)^i \\ &= 2r_i + 4\sum_{k=1}^{i-1} (-1)^k r_{i-k} + (-1)^i \qquad 1 \le i \le 76. \end{aligned} \tag{2}$$

With this formula (and a little help from a computer), we can present all the relevant data for the Putnam Competition.

Score	Rank	Frequency	Cumulative Frequency
101	1	1	1
100	2	1	2
86	3	1	3
80	4.5	2	5
79	6	1	6
77	7.5	2	8
73	9	1	9
72	11	3	12
71	14	3	15
70	16.5	2	17
69	19	3	20
68	23.5	6	26
67	27.5	2	28
66	29	1	29
63	30	1	30
62	32	3	33
61	37	7	40
60	47.5	14	54
59	56	3	57
58	59.5	4	61
57	62.5	2	63
55	64	1	64
54	66	3	67
53	69	3	70
52	72.5	4	74
51	78	7	81
50	93.5	24	105
49	109	7	112
48	115	5	117
47	118.5	2	119
46	120.5	2	121
45	122	1	122
44	123	1	123
43	125	3	126
42	130.5	8	134
41	143.5	18	152
40	164	23	175
39	181.5	12	187

Score	Rank	Frequency	Cumulative Frequency
38	190.5	6	193
37	195	3	196
36.9	197	1	197
36.8	198	1	198
36.7	199	1	199
36.6	200	1	200
36	202	3	203
35	204	1	204
34	205	1	205
33	209	7	212
32	225.5	26	238
31	252	27	265
30	287.5	44	309
29	314.5	10	319
28	324.5	10	329
26	330	1	330
24	332.5	4	334
23	340	11	345
22	365	39	384
21	414.5	60	444
20	494	99	543
19	555	23	566
18	570.5	8	574
17	575	1	575
15	577	3	578
14	586.5	16	594
13	606	23	617
12	655	75	692
11	759	133	825
10	970	289	1114
9	1130.5	32	1146
8	1154	15	1161
5	1162.5	2	1163
4	1166	5	1168
3	1181.5	26	1194
2	1252	115	1309
1	1469.5	320	1629
0	2292	1325	2954

-Frequencies and Cumulative Frequencies of the Putnam Scores-

Now we can easily say more about the results of the competition. For example, here is the information relevant to Putnam Pat: there were 99 (since $n_{59} = 99$) students who had a score of 20; there were 444 (since $c_{58} = 444$) students who had a score above 20; and since 2954 students participated, there were 2954 – 543 = 2411 (since $c_{59} = 543$) students who scored below 20. Furthermore, the scores resulting in prize money are more apparent. The Putnam fellows (the top five) had scores 101,100, 86, and 80. The $1000 winners (the next ten) had scores 79, 77, 73, 72, and 71. The $250 winners (the next eleven) had scores 70, 69, and 68.

Notice that this inversion formula can be derived using linear algebra: the recurrence in equation (1)

$$\begin{aligned} r_1 &= \frac{n_1}{2} + \frac{1}{2} \\ r_2 &= n_1 + \frac{n_2}{2} + \frac{1}{2} \\ r_3 &= n_1 + n_2 + \frac{n_3}{2} + \frac{1}{2} \\ &\vdots \\ r_{76} &= n_1 + n_2 + \ldots + n_{75} + \frac{n_{76}}{2} + \frac{1}{2} \end{aligned}$$

can be written as a matrix equation: $\vec{r} = \mathrm{A}\vec{n} + \overrightarrow{1/2}$, where

$$\vec{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_{76} \end{bmatrix}, \qquad \vec{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{76} \end{bmatrix}, \qquad \overrightarrow{1/2} = \begin{bmatrix} 1/2 \\ 1/2 \\ \vdots \\ 1/2 \end{bmatrix},$$

and

$$\mathrm{A} = \begin{bmatrix} 1/2 & 0 & 0 & \cdots & 0 \\ 1 & 1/2 & 0 & \cdots & 0 \\ 1 & 1 & 1/2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1/2 \end{bmatrix}.$$

Solving the linear system for $\vec{n}$ yields $\vec{n} = \mathrm{A}^{-1}\left(\vec{r} - \overrightarrow{1/2}\right)$. So we can find the formula for the n_i if we find the inverse of the matrix A. Not surprisingly, because of the simple form of the matrix A, it is not hard to find the inverse, which is also a lower triangular matrix:

$$\mathrm{A}^{-1} = \begin{bmatrix} 2 & 0 & 0 & 0 & \cdots & 0 \\ -4 & 2 & 0 & 0 & \cdots & 0 \\ 4 & -4 & 2 & 0 & \cdots & 0 \\ -4 & 4 & -4 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -4 & 4 & -4 & 4 & \cdots & 2 \end{bmatrix}.$$

The equation for $\vec{n}$ becomes

$$\vec{n} = \begin{bmatrix} 2 & 0 & 0 & 0 & \cdots & 0 \\ -4 & 2 & 0 & 0 & \cdots & 0 \\ 4 & -4 & 2 & 0 & \cdots & 0 \\ -4 & 4 & -4 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -4 & 4 & -4 & 4 & \cdots & 2 \end{bmatrix} \begin{bmatrix} r_1 - 1/2 \\ r_2 - 1/2 \\ \vdots \\ r_{76} - 1/2 \end{bmatrix},$$

giving us

$$\begin{aligned} n_1 &= 2\left(r_1 - 1/2\right) = 2r_1 - 1 \\ n_2 &= -4\left(r_1 - 1/2\right) + 2\left(r_2 - 1/2\right) = -4r_1 + 2r_2 + 1 \\ n_3 &= 4\left(r_1 - 1/2\right) - 4\left(r_2 - 1/2\right) + 2\left(r_3 - 1/2\right) = 4r_1 - 4r_2 + 2r_3 - 1 \\ &\vdots \\ n_{76} &= -4\left(r_1 - 1/2\right) + 4\left(r_2 - 1/2\right) - \ldots + 2\left(r_{76} - 1/2\right) = -4r_1 + 4r_2 - \ldots + 2r_{76} - 1, \end{aligned}$$

which is our solution (see (2)).