

Nonrandom Risk: The 1970 Draft Lottery

Norton Starr
Amherst College

Journal of Statistics Education v.5, n.2 (1997)

Copyright (c) 1997 by Norton Starr, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the author and advance notification of the editor.

Key Words: Chi-square; Classroom exercise; Correlation; Exploratory data analysis; Randomness; Regression.

Abstract

The 1970 draft lottery for birthdates is reviewed as an example of a government effort at randomization whose inadequacy can be exhibited by a wide variety of statistical approaches. Several methods of analyzing these data -- which were of life-and-death importance to those concerned -- are given explicitly and numerous others are cited. In addition, the corresponding data for 1971 and for 1972 are included, as are the alphabetic lottery data, which were used to select draftees by the first letters of their names. Questions for class discussion are provided. The article ends with a survey of primary and secondary sources in print.

1. Introduction

1 Good examples make a statistics course come alive, and are memorable afterwards. Among the case histories that have held up well over the years is the 1970 draft lottery, the data for which are still not widely available. Even less accessible are the data for the subsequent lotteries.

2 The Iraq war of 1991 and the recent presence of our troops in the Balkans show that military service continues to expose our soldiers to death and severe injury. Thus the 1970 draft lottery, vivid to many older readers, can be regarded as relevant to our students. Of course, were we living in more peaceful times, it would still be easy to evoke the awful risks to those conscripted for duty in Vietnam. Recall that in an attempt to expose male youth fairly to the risk of being drafted, a lottery was held to allocate birthdates at random: 366 capsules, each containing a unique day of the year, were successively drawn from a container. The first date drawn (September 14) was assigned rank 1, the second date drawn (April 24) was assigned rank 2, and so on. Those eligible for the draft who were born on September 14 were called first for physicals, then those born on April 24 were tapped, and so on.

3 This lottery was a source of considerable discussion before being held on December 1, 1969. Soon afterwards a pattern of unfairness in the results led to further publicity: those with birthdates later in the year seemed to have had more than their share of low lottery numbers and hence were more likely to be drafted. On January 4, 1970, the *New York Times* ran a long article, "Statisticians Charge Draft Lottery Was Not Random," illustrated with a bar chart of the monthly averages ([Rosenbaum 1970a](#)). It described the way the lottery was carried out, and with hindsight one can see how the attempt at randomization

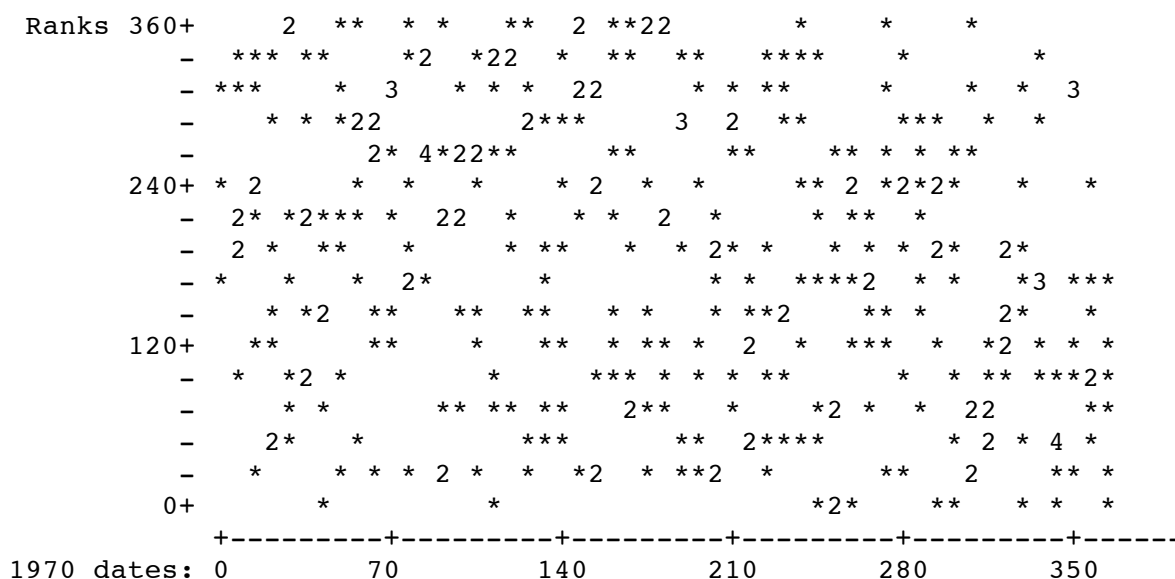
broke down. The capsules were put in a box month by month, January through December, and subsequent mixing efforts were insufficient to overcome this sequencing. The details of the procedure are quoted in [Fienberg \(1971a\)](#) and the first three editions of [Moore \(1979, 1985, 1991\)](#).

2. Data Analysis

4 The beauty of the 1970 draft lottery data is that students can confirm the nonrandomness in the lottery process by a wide variety of approaches. The dataset is ideal for computer laboratory experiments and for graphical exploration. A few Minitab examples will illustrate the ease with which one can display the lottery's bias. Also analyzed are the corresponding data for the 1971 lottery, which featured a much improved, if more complicated, randomization. (The 1970 draft lottery applied to eligible men aged 19 to 26 prior to January 1, 1970, and so included births taking place in some leap years. The 1971 draft lottery applied to men born in 1951, so only 365 days were involved.)

5 The dataset `draft70yr.dat` has the days of the year coded as consecutive numbers 1 through 366 in the first column (call it C1), the corresponding 1970 ranks in the second (call it C2), and the coding for the months (1 through 12) in the third column. The Minitab command `PLOT C2 C1` generates Figure 1, which is scarcely revealing, even in higher resolution representations.

Figure 1. Scatterplot of the 1970 Draft Lottery Data.



6 A comparable scatterplot of the 1971 data (contained in `draft71yr.dat`) similarly lacks obvious bias. One could at this stage test for a trend by regressing the ranks C2 on the dates C1. The slope for the 1971 regression line is not significantly different from zero, but for the 1970 data, one obtains a sample slope of $-.226$, which is significantly different from zero with a p-value that is zero to four decimal places. Because C1 and C2 contain permutations of the same numbers, $-.226$ is also the correlation coefficient, and its significance level is equivalent to that reported by [Moore and McCabe \(1993\)](#) for the correlation coefficient.

7 Various groupings of the data yield results that are, for many readers, more convincing evidence of unfairness. Among the ways of aggregating the data, the most natural seems to be grouping the ranks by month; this was done in the original *New York Times* article of January 4, 1970 ([Rosenbaum 1970a](#)), and

is done in almost every text on and criticism of the lottery. For each month, consider the ranks assigned to its dates. For example, January 1, 2, and 3 were the 305th, 159th, and 251st capsules drawn, so these three ranks are among the numbers one would use to determine the mean or median lottery rank for January. The 1970, 1971, and 1972 monthly data are given in draft70mn.dat, draft71mn.dat, and draft72mn.dat, respectively, where each dataset has 12 columns of ranks, with the 31 January ranks in the first column, and so on.

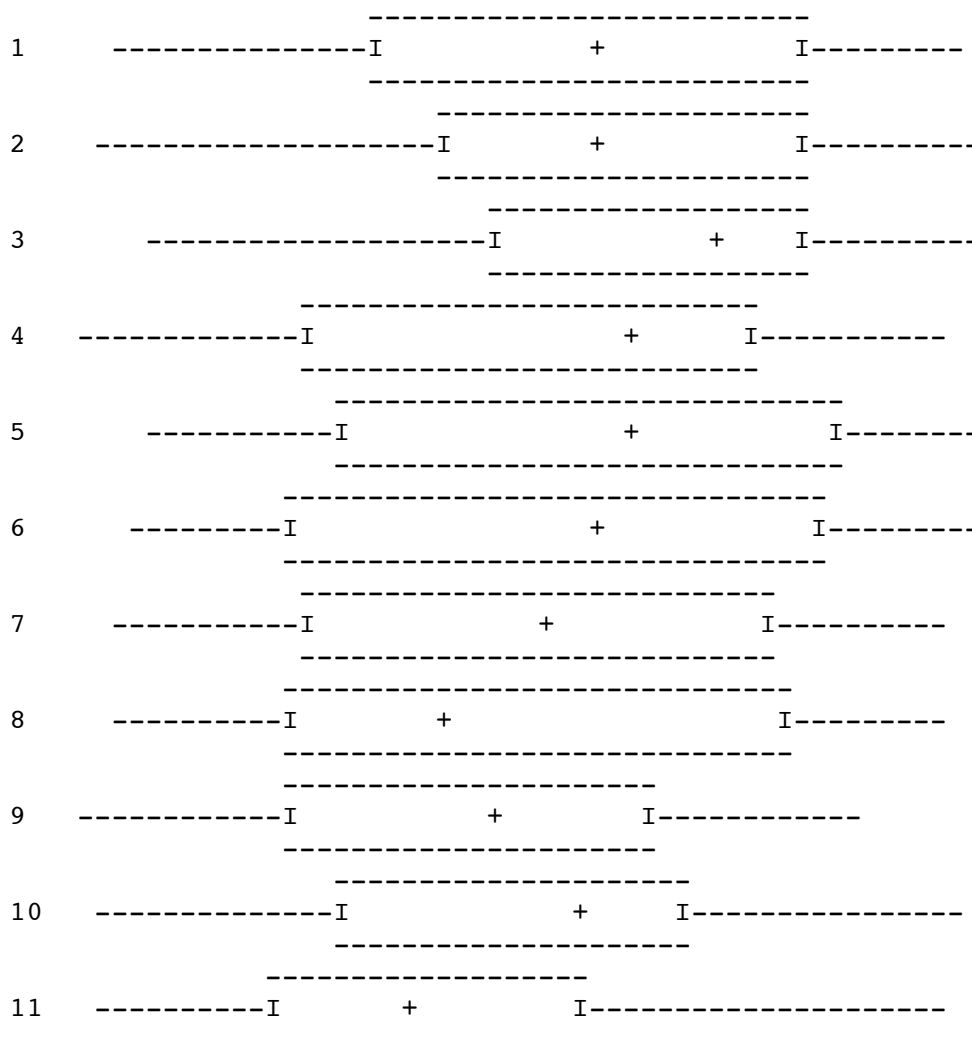
8 The monthly ranks in draft70mn.dat can be obtained from draft70yr.dat by means of Minitab's UNSTACK command, using column C3, which codes the months, as the subscripts.

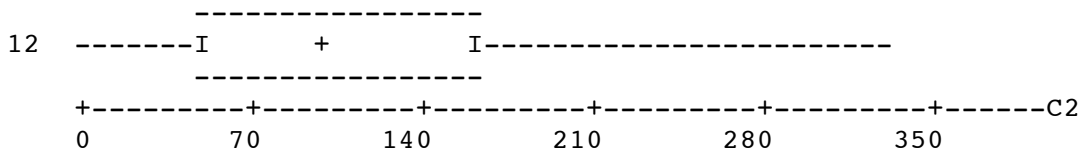
```
MTB > unstack c2 c11-c22;
SUBC> subscripts c3.
```

9 One can compute the means for each month and carry out a regression of the means against the consecutive month numbers 1 through 12. One can also just look at the means and see their striking decline toward the end of the year. More revealing than these twelve means are the monthly boxplots in Figure 2, which can be generated by the following Minitab commands, using the stacked data in draft70yr.dat and the month codings in C3.

```
MTB > boxplot c2;
SUBC> by c3.
```

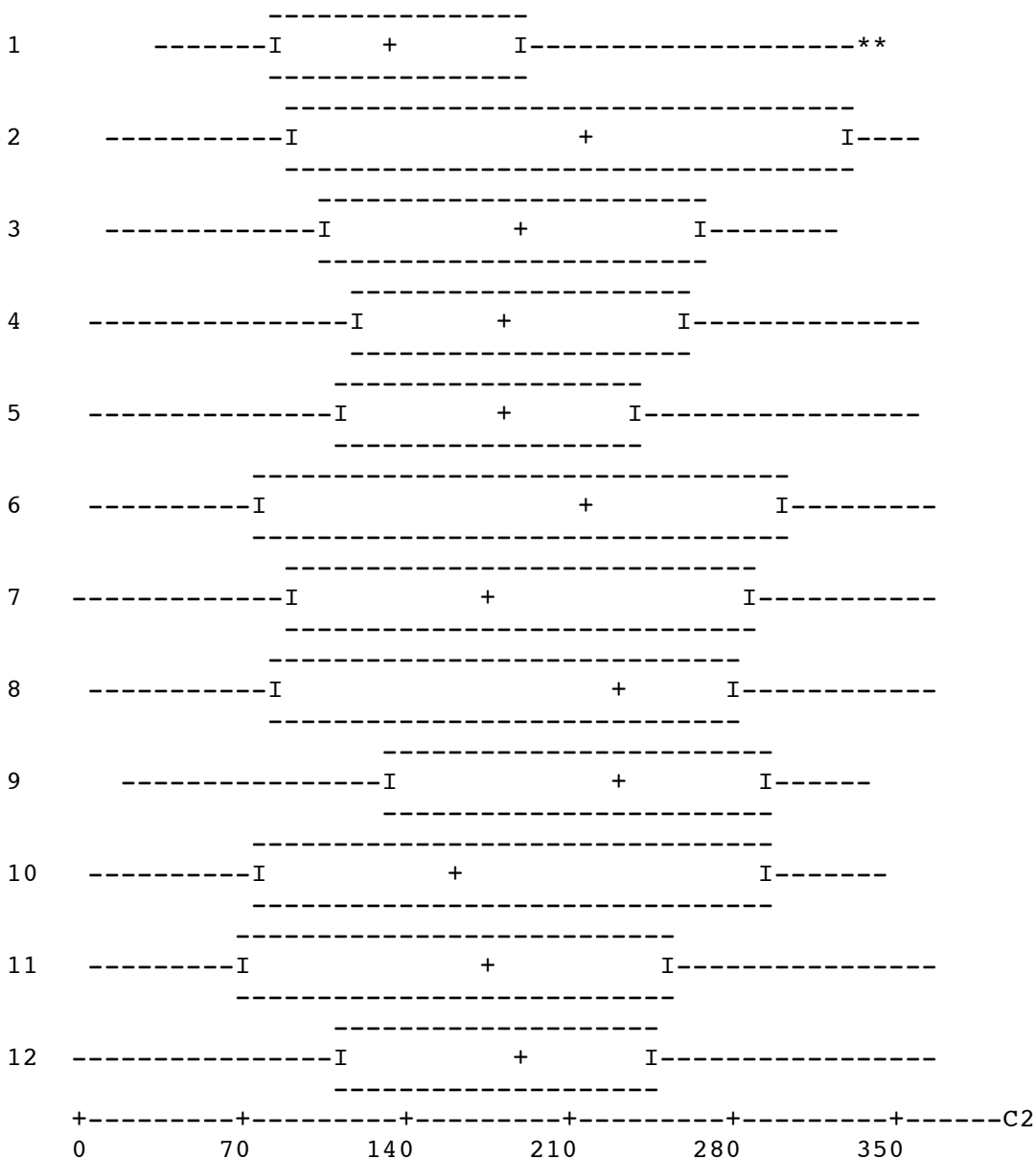
Figure 2. Side-by-Side Boxplots of the 1970 Draft Lottery Data.





10 The boxplots clearly show a decline in ranks in the latter third of the year, reflecting the inadequate mixing of the capsules that were added last to the mixing bowl. Other aspects of the boxplots are discussed by [Witmer \(1992\)](#). The 1971 boxplots in Figure 3 lack the evident bias of the 1970 display.

Figure 3. Side-by-Side Boxplots of the 1971 Draft Lottery Data.



11 The most primitive quantitative breakdown of the data, suggested by [Fienberg \(1973\)](#), is a two-way table exploring whether ranks above the median were as likely to fall in the first half as in the last half of the year. To construct this table in Minitab, first code both the ranks and the days of the year as zeros and ones, depending on whether they fall in the first or last halves; then invoke the TABLE command with a CHISQUARE subcommand. (In draft70yr.dat, the 366 ranks for 1970 are in C2, while the days of

the year, numbered 1 through 366, are in increasing order in C1.)

```
MTB > read 'draft70yr.dat' c1-c3
Entering data from file: draft70yr.dat
  366 rows read.
MTB > #Convert days and ranks to fractions whose
MTB > #rounded values are zero or one as desired.
MTB > let c11 = (c1-1)/366
MTB > let c12 = (c2-1)/366
MTB > round c11 c21
MTB > round c12 c22
MTB > table c22 c21;
SUBC> chis.
```

	0	1	ALL
0	74	109	183
1	109	74	183
ALL	183	183	366

CHI-SQUARE = 13.388 WITH D.F. = 1 # p < .0005

For 1971, the corresponding test shows no evidence of bias:

```
ROWS: C22      COLUMNS: C21
```

	0	1	ALL
0	94	88	182
1	88	95	183
ALL	182	183	365

CHI-SQUARE = 0.463 WITH D.F. = 1 # p = .50

12 Bob Hayden ran a two-sample t-test to compare the ranks in the first half and the last half of the year, providing yet another way of confirming the distinction between the 1970 and the 1971 data. Readers and their students may be similarly creative in working with this classic example.

13 The sources in the reference list include a variety of analyses, with varying levels of sophistication. These range from [Fienberg's](#) original article in *Science*, through [Moore and McCabe's](#) discovery and inference in *Introduction to the Practice of Statistics*, to [Fienberg's](#) elementary exposition in *Statistics by Example*, and [Witmer's](#) treatment in his supplementary text, *Data Analysis, An Introduction*. Also included are references to the original data as published by the [U. S. Government](#).

14 It is not widely known that there was a second drawing on December 1, 1969, held to rank the twenty-six letters of the alphabet. "The order of selection from among men born on the same date would be determined by the order in which the first letters of their last, first and middle names were drawn" ([U.S. Selective Service System 1970](#), p. 7). These alphabetic data (draftalpha.dat) are a good, simple source for elementary analyses.

15 The 1972 monthly lottery data (draft72mn.dat) were taken from the corresponding Selective Service report. (The stacked data for 1972 are contained in draft72yr.dat.) This constitutes a new dataset to which students can apply the methods that confirm the already known properties of the 1970 and 1971 lotteries.

3. Questions

16 Here are some questions for possible classroom use:

1. Assuming males are born with equal likelihood throughout the year, was the lottery really necessary? Or was it carried out largely to convey a sense of fairness in an essentially stochastic context where the stakes happened to be very serious?
2. Should attention have been given to twins, triplets, and other siblings to avoid multiple impacts on a given family?
3. Were those born on February 29 treated unfairly in the 1970 lottery?
4. The alphabetic data for 1970 seem, to the naked eye, to have their own significant bias: only three letters from the first half of the alphabet were among the first thirteen chosen. Is this apparent lack of randomness statistically significant? If so, is it of practical significance? If the answer to the latter is "no," then why was a permutation of the alphabet used in the first place?
5. How should men lacking a middle or even a first name be handled? (This is a real-life missing data issue! Students might be encouraged to find out what was actually done.)

4. Getting the Data

17 The file [draft.txt](#) is a documentation file containing a brief description of the datasets. The following files contain the raw data:

[draft70yr.dat](#)

[draft71yr.dat](#)

[draft72yr.dat](#)

[draft70mn.dat](#)

[draft71mn.dat](#)

[draft72mn.dat](#)

[draftalpha.dat](#)

Acknowledgments

I thank Bob Hayden for advice and encouragement during the preparation of this article.

Appendix - Key to Variables in Draft Lottery Datasets

The datasets `draft70yr.dat`, `draft71yr.dat`, and `draft72yr.dat` contain the lottery data for 1970, 1971, and 1972, respectively, in stacked format. Values are aligned and delimited by blanks. Note that the 1971 dataset contains only 365 days. (In this appendix, "column" refers to a vertical line, one character wide, in the data listing, and not to a Minitab column.)

Columns

```

1 - 3 Day of the year from 1 to 366
6 - 8 Rank assigned to day
11 - 12 Month of the year between 1 and 12

```

The datasets `draft70mn.dat`, `draft71mn.dat`, and `draft72mn.dat` contain the lottery data for 1970, 1971, and 1972, respectively, in unstacked format. Values are aligned and delimited by blanks. The number of ranks for each month equals the number of days in that month and varies from 28 to 31.

Columns

3 - 5 Ranks assigned to days in January
 7 - 9 Ranks assigned to days in February
 11 - 13 Ranks assigned to days in March
 15 - 17 Ranks assigned to days in April
 19 - 21 Ranks assigned to days in May
 23 - 25 Ranks assigned to days in June
 27 - 29 Ranks assigned to days in July
 31 - 33 Ranks assigned to days in August
 35 - 37 Ranks assigned to days in September
 39 - 41 Ranks assigned to days in October
 43 - 45 Ranks assigned to days in November
 47 - 49 Ranks assigned to days in December

The dataset `draftalpha.dat` contains the alphabetic lottery data. Values are aligned and delimited by blanks. There are no missing values.

Columns

1 - 2 Integers from 1 to 26
 5 Permutation of the 26 letters of the alphabet
 8 - 9 Integers between 1 and 26 corresponding to the
 letters in column 5

References

Eckholm, E. (1986), "Status in Draft Linked to Suicide," *New York Times*, March 7, p. A16. This summarizes the Hearst, Newman, and Hulley article cited below.

Fienberg, S. E. (1971a), "Randomization and Social Affairs: The 1970 Draft Lottery," *Science*, 171, 255-261. The best and most comprehensive single resource on this issue. Includes an interesting historical sketch of randomness in social affairs.

----- (1971b), Comment on "Draft Lottery: Validity of Randomness," by C. J. Scheirer, *Science*, 172, 630-631.

----- (1973), "Randomization for the Selective Service Draft Lotteries," in *Statistics by Example: Finding Models*, eds. F. Mosteller, W. H. Kruskal, R. F. Link, R. S. Pieters, and G. R. Rising, Reading, MA: Addison-Wesley, pp. 1-13. (Note: There is a typographical error in Table 4, p. 13. The 1971 draft rank for May 11 is given as 243, the same rank as for November 1. It should be 293, as a study of the Rosenblatt and Filliben article cited below or a glance at the corresponding U. S. Selective Service report indicates.) A very accessible article, with a variety of analyses and a good set of questions for class use.

Hearst, N., Newman, T. B. and Hulley, S. B. (1986), "Delayed Effects of the Military Draft on Mortality," *New England Journal of Medicine*, 314(10), March 6, 620-624. Demonstrates a higher suicide rate among those with low lottery numbers than among those with high ranks. The question of the possible effect of the hidden variable, actual military service, is addressed as well.

Kitchens, L. J. (1998), *Exploring Statistics* (2nd ed.), Pacific Grove, CA: Brooks/Cole pp. 20, 216

(Exercise 3.97), and 682 (Exercise 10.62). Includes a data disk with both the 1970 and 1971 data.

Larsen, R. J., and Stroup, D. F. (1976), *Statistics in the Real World*, New York: Macmillan, pp. 241-245. Nonparametric approaches: Kruskal-Wallis test on the monthly ranks and Wilcoxon/Mann-Whitney rank sum test on the ranks in the first and last halves of the year.

Moore, D. S. (1979, 1985, 1991, and 1997), *Statistics: Concepts and Controversies* (1st, 2nd, 3rd, and 4th eds.), New York: Freeman. Includes a description of the 1971 procedure, of which Moore says, "It's awful, but it's random."

Moore, D. S., and McCabe, G. P. (1993), *Introduction to the Practice of Statistics* (2nd ed.), New York: Freeman, pp. 105-107, 447-448. Uses a median trace to raise the possibility of unfairness and helps confirm this with a test of the correlation coefficient.

Mosteller, F., Fienberg, S. E., and Rourke, R. E. K. (1983), *Beginning Statistics with Data Analysis*, Reading, MA: Addison-Wesley, pp. 183-184. Interesting remarks on 1940, 1970, and 1971 lotteries.

Rosenbaum, D. E. (1970a), "Statisticians Charge Draft Lottery Was Not Random," *New York Times*, January 4, p. 66. "If the results occur less frequently" than 5% of the time, "then the statisticians conclude that some causative factor was involved." This makes an interesting contrast to the subsequent delicacy with which the *New York Times* characterizes the statistics involved in polls ("How the Poll Was Conducted").

----- (1970b), "Draft Officials Redesign Lottery Procedures to Make the System More Random," *New York Times*, June 25, p. 17. Describes the 1971 lottery procedure. "We would like to have a drawing this year that appears impartial, both to those professionally curious and to those whose lives are involved." (The spokesperson's use of "appears" shows a keen sensitivity to the issues involved.)

----- (1970c), "Draft Lottery for Youths Born in 1951 to Be Conducted Today; New System Is Hailed," *New York Times*, July 1, p. 19. Some remarks comparing the 1970 and 1971 lotteries.

----- (1970d), "Second Draft Lottery Selects Call-Up Order for 1971," *New York Times*, July 2, pp. 1, 12. Describes the procedure in detail, gives the rank for each day of the year, and declares that the alphabetic priority will be the one determined in the previous lottery.

Rosenblatt, J. R., and Filliben, J. J. (1971), "Randomization and the Draft Lottery," *Science*, 171, pp. 306-308. A description and analysis of the 1971 lottery, concluding that the process was effectively random.

United States Selective Service System (1970), "Semi-Annual Report of the Director of Selective Service for the Period July 1 to December 31, 1969 to the Congress of the United States," Washington: U. S. Government Printing Office, pp. 5-10. Describes the procedures and gives results by date for each month as well as by rank from first selected to last. Also gives the alphabetic lottery results. The birthdate lotteries for future years are given in successive reports, while the alphabet permutation used after 1970 is that determined in the 1970 lottery.

Witmer, J. A. (1992), *Data Analysis: An Introduction*, Englewood Cliffs, NJ: Prentice-Hall, pp. 21-24. Displays a beautifully informative set of monthly boxplots, along with the revealing median trace. These reflect the sequential mixing of the numbered capsules, which was a major source of bias.

Norton Starr

Department of Mathematics and Computer Science
Box 2239 Amherst College
Amherst, MA 01002-5000

nstarr@amherst.edu

[Return to Table of Contents](#) | [Return to the JSE Home Page](#)