Did The Men's Dates of Birth Determine/Affect The Order of Call to Military Service in Vietnam War?

Liqian Situ

Table of Contents

1. Introduction

On December 1, 1969, two lotteries were conducted to determine the order of call to military services in the Vietnam War. Men born between 1994 and 1950 are being "draw" from the lottery.

The main purpose of this paper is to find out whether the lottery is fair. In other words, did the birth dates determine/affect the order of call to military service in the Vietnam War? Or the lottery numbers were drawn randomly regardless of the dates of birth? Or can we predict the probability of a man's lottery number by knowing his birthday? From our analysis, we believe that comparing men born in the first three quarters of the year, men born in the fourth quarter form 1944 to 1950 seem more likely to be assigned lottery numbers belonging to the first or the second lottery quarter, and they were less likely to fall into the category of the fourth lottery quarter.

2. The Data and The Model

Figure 1 is a scatterplot of the lottery number against the day of a year. Men born in the first 100 days of the year were more likely assigned lottery numbers from the range of 200 to 300. Similarly, for those born in last 100 days of the year, they were more likely assigned lottery numbers from the range of 1 to 100.

Figure 1: Scatter Plot of Draft Order vs. The Day of A Year

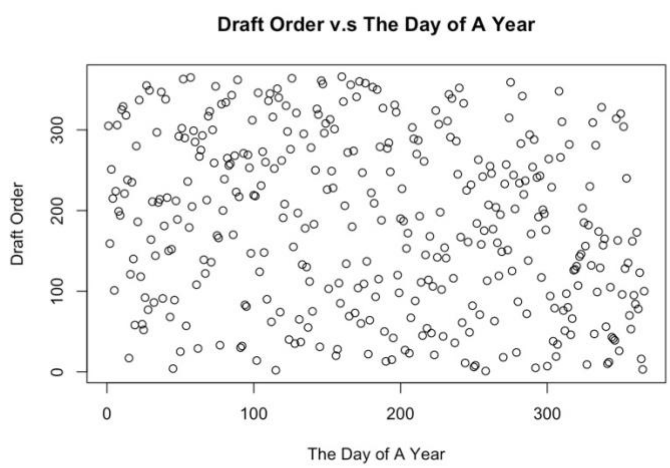

Table 1: Birth Quarter vs. Lottery Quarter

	Lottery Quarter				
Birth Quarter	First	Second	Third	Fourth	Total
First	15	19	28	29	91
Second	23	15	24	29	91
Third	24	26	22	20	92
Fourth	29	31	18	14	92
Total	91	91	92	92	366

By eyeballing Table 1, the days in the first quarter of the year were more likely assigned into the third or the fourth quarter of the lottery. In contrast, the days in fourth quarter seem more likely be assigned to the first or the second lottery quarter. In addition, the birth dates of third quarter of year seem split evenly into each category of the lottery quarters.

However, we cannot conclude the fairness of the lottery by solely checking the scatter plot or eyeballing the distribution of data. Here we employ the idea of testing the independent to sort out the association between birth quarter and lottery quarter. If the Pearson's chi-square test of independence fails to reject the independent relationship between birth quarter and lottery quarter, the lottery is fair because a man's lottery number was not depending on his date of birth. In other words, we cannot predict probability of which lottery quarter a man will be classified by knowing his birthday. In contrast, if we reject the independent relationship between birth quarter and lottery quarter, further study is needed for determining the association between these quarter of the year and the lottery quarter.

3. Test of Independence and Results

3.1 The Chi-Square Test of Independence

First we consider the following hypothesis: H_0: birth quarter and lottery quarter are independent vs. H_a: birth quarter and lottery quarter are not independent. In term of statistics, we should consider a test of $H_0: \pi_{ij} = \pi_{i+}\pi_{+j}$ vs. $H_a: \pi_{ij} \neq \pi_{i+}\pi_{+j}$, where $i, j = 1, \dots, 4$, are the number of rows and columns in Table 1, and $\pi_{ij} = \Pr(birth\ quarter = i \cap lottery\ quarter = j)$. Since the underline probability structure is multinomial, we can easily estimate the value of π_{ij}, and then compute the Pearson chi-square statistic X^2 and the likelihood ratio test statistic G^2. We use R to calculate these two quantities, as shown in Appendix 4b. Giving $X^2 = 20.4362$, with $df = 9$ and $G^2 = 21.1064$, with $df = 9$, we reject the claim that birth quarter and lottery quarter are independent based on 5% significant level.

Like other statistical tests, when we have a very small $p - value$, chi-square test of independence indicates strong evidences of association but provides very little information on the nature/strength of the association (Agresti). In the next section, we analyze the standardized Pearson residuals to further investigate the nature/strength of the association between birth quarter and lottery quarter.

3.2 The Analysis of Residuals

A standardized Pearson residual is defined as $r_{ij} = \frac{n_{ij} - \hat{\mu}_{ij}^0}{\sqrt{\hat{\mu}_{ij}^0 (1 - \hat{\pi}_{i+})(1 - \hat{\pi}_{+j})}}$. Under the null hypothesis, $\{r_{ij}\}$ are asymptotically distributed as standard normal. We also learned that $|\, r_{ij}| > 2$ suggests the violation of independence in category (i, j). We compute this quantity in R (see Appendix 5), and we find lack of fit of the independence in following categories: (1^{st} birth quarter, 1^{st} lottery quarter), (2^{nd} birth quarter, 2^{nd} lottery quarter), (4^{th} birth quarter, 2^{nd} lottery quarter) and (4^{th} birth quarter, 4^{th} lottery quarter). Especially, we see large positive residuals for the 4^{th} birth quarter and the 2^{nd} lottery quarter, and large negative residuals for the 4^{th} lottery quarter and the 4^{th} lottery quarter. In order to confirm our finding and gain more insights about the associations, we run partitioning chi-square tests in the next section.

3.3 Partitioning Chi-Square

Table 3: Subtables Used in Partitioning Chi-Squared for Table 1

Subtable 1			Subtable 2			Subtable 3		
	First	Second		First+Second	Third		First+Second+Thrid	Fourth
Frist	15	19	First	34	28	First	62	29
Second	23	15	Second	38	24	Second	62	29

Subtable 4			Subtable 5			Subtable 6		
	First	Second		First+Second	Third		First+Second+Thrid	Fourth
Frist+ Second	38	34	Frist+ Second	72	52	Frist+ Second	124	58
Third	24	26	Third	50	22	Third	72	20

Subtable 7			Subtable 8			Subtable 9		
	First	Second		First+Second	Third		First+Second+Thrid	Fourth
Frist+ Second+ Third	62	60	Frist+ Second+ Third	122	74	Frist+ Second+ Third	196	78
Fourth	29	31	Fourth	60	18	Fourth	78	14

To understand the association in Table 1 better, and obtain more insights from the analysis of residuals, we partition Table 1 into 9 subtables based on the partitioning method of Lancaster (1949). In other words, we partition G^2 into 9 components.

The Subtable 1 in Table 3 compares the first birth quarter to the second birthday quarter on whether the lottery quarter is first or second. For this subtable, $G^2 = 1.95$, with $df = 1$. Subtable 2 compares these two birth quarters on the proportion of time the draft order was assigned to the third lottery quarter, rather than the first or second. Here $G^2 = 0.53$, with $df = 1$. Subtable 3 compares these two birth quarters on the proportion of time was assigned to the fourth lottery quarter, rather than the first, second or third. This subtable has $G^2 = 0$, with $df = 1$. Based on the significant level at 0.05, there is not enough evidence to suggest that men born in the first quarter from 1944 to 1950 were more likely (or less likely) to be assigned to a particular lottery quarter.

Next we combine the first and the second birth quarters compare to the third birth quarter. With $df = 1$, the G^2 for Subtable 4, 5 and 6 are $0.27, 2.55$ and 3.17, respectively. Hence, based on significant level at 0.05, these results indicate that there is no evidence to support the men born in the third quarter of the year would be more likely (or less likely) to be assigned lottery numbers belonging to a particular lottery quarter. From the analysis of residuals, we also obtain the result of independent relationships between the third birth quarter and the lottery quarters.

Last but not least, we combine the first three-birth quarters compare to the fourth birth quarter. Subtable 8 in Table 3, giving the $G^2 = 5.62$ with $df = 1$, I compare them with the (First+Second, Third) classification. Similarly, in Subtable 9, giving the $G^2 = 6.93$, with $df = 1$, we compare them with the (First+Second+Thrid, Fourth) classification. Based on the significant level at 0.05, comparing men born in the first three quarters of the year, men born in the fourth quarter form 1944 to 1950 seem more likely to be assigned lottery numbers belonging to the first or the second lottery quarter, and they were less likely to fall into the category of the fourth lottery quarter.

4. Conclusions

From our analysis, we do not find any dependence between the first three quarters of the year and the lottery quarter. However, men's who born in the fourth quarter of the years were more likely to be assigned a lottery number belonging to the first or the second lottery quarter, and were less likely to be drawn into the fourth lottery quarter. In conclusion, the lottery is not fair to men who born in the fourth quarter of the year.

Reference

Agresti, A. (1996). Inference for contingency tables. In An introduction to categorical data analysis (2nd ed.). New York: Wiley.