post02-meiying-huang

Meiying Huang

11/30/2017

Introduction:

On December 1, 1969, the United States held its first draft lottery ever since 1942, which gave young men (any year between 1944 and 1950 were assigned lottery) a random number corresponding to their birthdays. As part of the war in Vietnam, the United States government implemented a lottery to draft men of appropriate ages into the US Armed Forces. Under the old system of drafting first the oldest available men age 19 through 25, in 1970, a drawing took place in which capsules corresponding to each day of the year (numbered sequentially as 1-366) were placed into a drum and selected one-by-one. The order of selection determined the order in which men would be drafted by matching each eligible male to the capsule labeled with his birthday; hence, men whose birthdays were selected earlier would be drafted into the military earlier.

This post is focusing on analysing simulating randomization test for correlation as the following question:

#)How the draft worked? ## The method of the draft lottery: the days of the year(including Feb 29) were written on slips of paper. These pieces of paper were then placed in separate plastic capsules that were mixed in a shoebox and then dumped into a deep glass jar. Capsules were drawn from the jar one at a time.

If you had been born in 1950, you would be at least eighteen years old. Would your draft number have been called on December 1, 1969 among the first number and the second number being drew? You can test here to find out.

```
# sicne lottery number 1 is sep 19 : 177
#m1969<- lottery$Month_Number=="9"
#d1969<- lottery$Day_of_year=="263"
m1=9
d1=19
m2=4
d2=24
test_my_bday<- function(m,d){
  if((m == m1)&(d== d1)){
    print("Oops, you was called in 1970 in the first two draws")
  }else if((m==m2) & (d==d2)){
    print("Oops, you was called in 1970 in the first two draws")
   } else{
    print("Lucky you! You was NOT called in 1970 in the first two draws")
    }
}

#let me test my birthday
test_my_bday(4,19)
```

```
## [1] "Lucky you! You was NOT called in 1970 in the first two draws"
```

```
#go ahead to test your birthday here
# text_my_bday(m,d)
```

According to Dan Eberhart’s “A pact with living”, during the first selective lottery, 85,000 American young man between the ages of

19 to 25. Are there any relationship between different variables?

Data: According to the RESULTS FROM LOTTERY DRAWING - Vietnam Era 1970. You can find it on the file 1970 lottery. It is named “data_1970lottery.txt” and is vailable in the “post01.” The columns of the file data_1970lottery have month, month number, day of teh eyar, and draft number. Here will consider two variables recorded on each date, nad sequential date in the year(e.g, Jan 31 transcribed into sequential date as 31, Feb 1 as 32, and December 31 as 366, and so on).

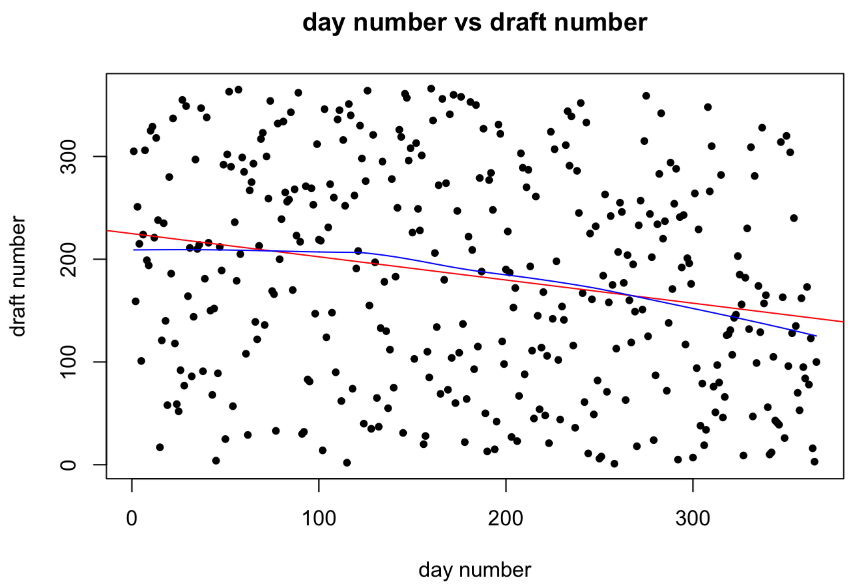
Load the data files (Since this post is being asked to write it in a way which is also easy for the other audience with (almost) no experience in this area. And it is easier for people to run the code without worrying about the wrong paths, directory issue. I stored every file in post02 instead of putting more different files inside).

```
lottery = read.table('data_1970lottery.txt', sep = ',', header = T)
```

First, let’s see if there is any relationship between the draft number and the day number by plotting the draft numver versus day number. In other word, in a perfectly fair, random lottery, what should be the value of the correlation coefficient between draft number and sequential date of birthday?

- Scatterplot

```
attach(lottery)
# Make a scatterplot of draft number against the day number to see if anything appear amiss?
#consider the day number as the independent varialbe x, and draft number is the dependent variable y.
plot(lottery$Day_of_year, lottery$Draft_No, main = "day number vs draft number",
     xlab = "day number", ylab = "draft number", pch= 20)
# include a regression line
abline(lm(lottery$Draft_No ~ lottery$Day_of_year), col = "red")
#LOWESS smoother which uses locally-weighted polynomial regression give the coordinates of the smooth
lines(lowess(lottery$Day_of_year, lottery$Draft_No), col = "blue")
```




```
##From the above scatterplot, we observe that the data points are scattered randomly. The plot shows there seems h
as no relationship between the draft numbers and the day numbers. Therefore we conclude that the day numbers and t
he draft numbers are independent. The scatter plot does not reval much of an association between draft number and
sequential dates.
```

```
#Find the correlations: Compute both the Pearson and Spearman correlations between day number and draft number. Co
mpare their values. Does this make sense? Return: Pearson and Spear- man correlations.

#find Pearson correlations between day number and draft number.
##draft_pearson_cor
draft_pearson_cor<- cor.test(lottery$Day_of_year,
                             lottery$Draft_No,
                              method = "pearson")
draft_pearson_cor
```

```
##
##  Pearson's product-moment correlation
##
## data:  lottery$Day_of_year and lottery$Draft_No
## t = -4.4198, df = 364, p-value = 1.305e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3207751 -0.1260932
## sample estimates:
##        cor
## -0.2256861
```

```
# find Spearman correlations between day number and draft number.
##draft_spearman_cor
draft_spearman_cor<-cor.test(lottery$Day_of_year,
                             lottery$Draft_No,
                              method =  "spearman")
draft_spearman_cor
```

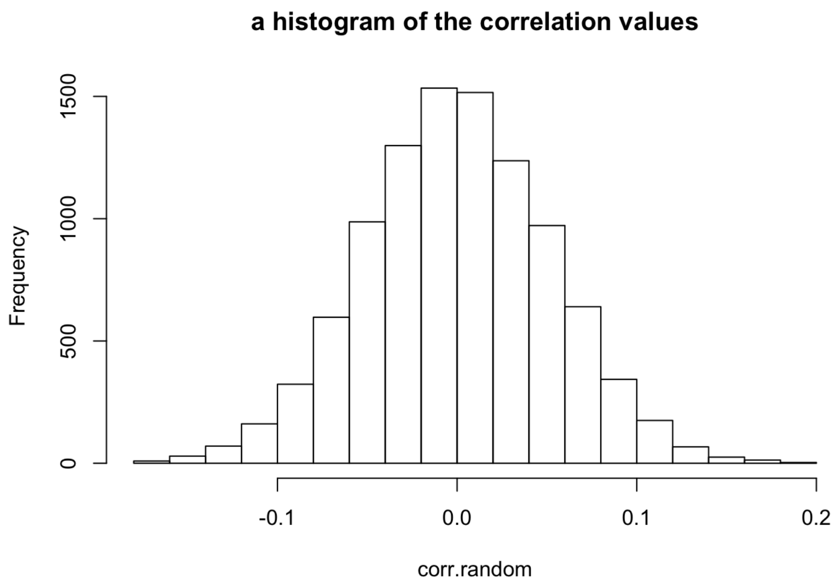
```
##
##  Spearman's rank correlation rho
##
## data:  lottery$Day_of_year and lottery$Draft_No
## S = 10015000, p-value = 1.375e-05
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.2256861
```

```
obs_corr<- cor(lottery$Day_of_year, lottery$Draft_No)
##result: the Pearson and Spearman correlations between day number and draft number are the same -0.2256861. Which
means x(day number) and y(draft number) has a week negative linear relationship, means when one variable tends to
increase as the other decreases. The correlation shows us that the highest value of draft number is tend to associ
ated with the lowest value of day number but the relationship is weak.
```

```
##To determine whether the observed correlation is statistically significant, we need to perform a permutation tes
t with a big enough sample size ( I did it with 10,000 permutations).
#H0(null hypothesis): p = -0.2260414, H1:(alternative hypothesis) p > -0.2260414
draft_num<- lottery$Draft_No
day_num<- lottery$Day_of_year
corr.random =  NULL
for(i in 1:10000){
  draft_random = sample(draft_num, length(draft_num), replace = TRUE )
  corr.random[i] = cor(day_num, draft_random)
}
p_value= sum(corr.random > obs_corr)/10000
p_value
```

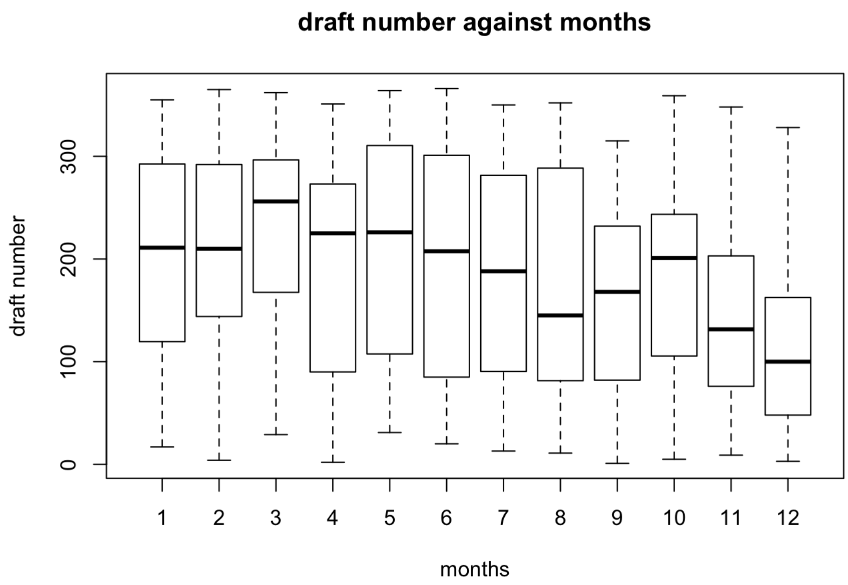
```
## [1] 1
```

```
#Provide a histogram of the correlation values arising from your permutations and annotate the true correlation. C
ompute the empirical p-value. Is the correlation significant? Return: empirical p-value.
hist(corr.random, main = "a histogram of the correlation values")
abline(v= obs_corr, col = "red", lwd = 2)
```




```
#result: assuming significance level=5%, since here p_value=1 which is greater than 5%, so fail to reject the H0(n
ull hypothesis). Thus, the observed correlation is statistically significant.
```

```
#plot a boxplot of draft number against months to see (if any) what trend here has?
month_num<- lottery$Month_Number
boxplot(draft_num ~ month_num,
        main = "draft number against months",
        xlab =  "months", ylab = "draft number")
```



```
##December has the losest median draft number and has less the lowest minimum and lowest maximum draft number. It'
s more likely to have an average around 200 draft number.
##The average draft number per month increases from January to March, and from August to October. Decreases from M
arch to August, and from October to December. There's no clear increasing or decreasing pattern through out the mo
nths in a year.
```

```
#Why is the Kruskal-Wallis test more appropriate than ANOVA for this data? Perform the Kruskal-Wallis test to see
if draft numbers are different across months and describe what you would conclude based on the result. Return: KW
test p-value.
kruskal.test(lottery$Month ~ draft_num)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  lottery$Month by draft_num
## Kruskal-Wallis chi-squared = 365, df = 365, p-value = 0.4902
```

```
kw_p_val = 0.4902
```

```
#How many pairwise comparisons are possible? For each comparison, perform a Mann- Whitney U test. Return: the set
of (uncorrected) p-values.
#since we have n=366, comparison=n(n-1)/2=66,795 with the pairwise comparison method.
# a helpful function is combn
mw_p_vec1<- wilcox.test(draft_num, month_num, correct = FALSE)
mw_p_vec1
```

```
##
##  Wilcoxon rank sum test
##
## data:  draft_num and month_num
## W = 131760, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
## p-value1 < 2.2e-16, which is less than 0.05. So reject the null hypothesis, means that these two groups(draft n
umber and month number do not have statistical equality mean)
sum.rank.draft = sum(rank(c(draft_num, month_num)) [1:366])
w1=sum.rank.draft- length(draft_num)*(length(draft_num)+1)/2
w1
```

```
## [1] 131755
```

```
sum.rank.month = sum(rank(c(draft_num, month_num)) [367: 732])
w2 = sum.rank.month -  length(month_num)*(length(month_num)+1)/2
w2
```

```
## [1] 2201
```

```
w3= sum(rank(c(draft_num,month_num))[1:366])
w3
```

```
## [1] 198916
```

```
w4 = sum(rank(c(draft_num,month_num))[367:732])
w4
```

```
## [1] 69362
```

```
##in the group(draft and month number) since the compute interval(w3=198916, w4=69362) is contained within the tab
ulated interval(w1=131755, w2= 2201), we conclude by accepting the hypothesis H0 of equality of means.

mw_p_vec2<- wilcox.test(draft_num, day_num, correct = FALSE)
mw_p_vec2
```

```
##
##  Wilcoxon rank sum test
##
## data:  draft_num and day_num
## W = 66978, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
```

```r
##p-value = 1, greater than 0.05. then we can accepct the null hypothesis H0 of statistical equality of the means of two group(draft number and day number.)
sum.rank.draft2 = sum(rank(c(draft_num, day_num)) [1:366])
w11=sum.rank.draft2- length(draft_num)*(length(draft_num)+1)/2
w11
```

```
## [1] 66978
```

```r
sum.rank.day = sum(rank(c(draft_num, day_num)) [367: 732])
w22 = sum.rank.day -  length(day_num)*(length(day_num)+1)/2
w2
```

```
## [1] 2201
```

```r
w33= sum(rank(c(draft_num,day_num))[1:366])
w33
```

```
## [1] 134139
```

```r
w44 = sum(rank(c(draft_num,day_num))[367:732])
w44
```

```
## [1] 134139
```

```r
# Which comparisons are significant using the Bonferroni (FWER) correction with α = 0.05? What about the Benjamini-Hochberg (FDR) correction at α = 0.05? Return: the set of corrected p-values for both methods.
kw_p_val = 0.4902
p1 = kw_p_val
fwer_p_vec1<- p.adjust(p1, method = "bonferroni", n= length(p1))
fwer_p_vec1
```

```
## [1] 0.4902
```

```r
fdr_p_vec1<- p.adjust(p1, method = "fdr", n= length(p1))
fdr_p_vec1
```

```
## [1] 0.4902
```

```r
p2 = 1
fwer_p_vec2<- p.adjust(p2, method = "bonferroni", n= length(p2))
fwer_p_vec2
```

```
## [1] 1
```

```r
fdr_p_vec2<- p.adjust(p2, method = "fdr", n= length(p2))
fdr_p_vec2
```

```
## [1] 1
```

```r
#Perform Fisher’s exact test with the following categories in the table: January-June vs July-December and draft numbers of 1-182 vs 183-366. Provide the table and compare the p-value to that from the Kruskal-Wallis test. Return: p-value.
m1to6= c(lottery$Month_Number<7, lottery$Month_Number)
m1to6
```

```
##   [1]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
##  [24]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
##  [47]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
##  [70]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
##  [93]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [116]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [139]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [162]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  0  0
## [185]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [208]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [231]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [254]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [277]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [300]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [323]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [346]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  1
## [369]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [392]  1  1  1  1  1  1  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2
## [415]  2  2  2  2  2  2  2  2  2  2  2  2  3  3  3  3  3  3  3  3  3  3  3
## [438]  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  4  4  4
## [461]  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
## [484]  4  4  4  4  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5
## [507]  5  5  5  5  5  5  5  5  5  5  5  5  6  6  6  6  6  6  6  6  6  6  6
## [530]  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  7  7  7  7
## [553]  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
## [576]  7  7  7  7  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8
## [599]  8  8  8  8  8  8  8  8  8  8  8  8  9  9  9  9  9  9  9  9  9  9  9
## [622]  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9 10 10 10 10
## [645] 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
## [668] 10 10 10 10 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
## [691] 11 11 11 11 11 11 11 11 11 11 11 12 12 12 12 12 12 12 12 12 12 12 12
## [714] 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12
```

```
sum(m1to6)
```

```
## [1] 2566
```

```
m7to12= c(lottery$Month_Number>6, lottery$Month_Number)
m7to12
```

```
##   [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [24]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [47]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [70]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [93]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [116]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [139]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [162]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  1
## [185]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [208]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [231]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [254]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [277]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [300]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [323]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [346]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [369]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## [392]  1  1  1  1  1  1  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2
## [415]  2  2  2  2  2  2  2  2  2  2  2  2  3  3  3  3  3  3  3  3  3  3  3
## [438]  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  4  4  4
## [461]  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
## [484]  4  4  4  4  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5
## [507]  5  5  5  5  5  5  5  5  5  5  5  5  6  6  6  6  6  6  6  6  6  6  6
## [530]  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  7  7  7  7
## [553]  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
## [576]  7  7  7  7  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8
## [599]  8  8  8  8  8  8  8  8  8  8  8  8  9  9  9  9  9  9  9  9  9  9  9
## [622]  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9 10 10 10 10
## [645] 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
## [668] 10 10 10 10 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
## [691] 11 11 11 11 11 11 11 11 11 11 11 12 12 12 12 12 12 12 12 12 12 12 12
## [714] 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12
```

```
sum(m7to12)
```

```
## [1] 2568
```

```
low_draft= c(lottery$Draft_No< 183, lottery$Draft_No)
low_draft
```

```
##   [1]   0   1   0   0   1   0   0   0   0   0   0   0   0   0   1   1   0
##  [18]   1   1   0   0   0   1   1   1   1   0   1   0   1   0   1   1   0
##  [35]   0   0   0   1   1   0   0   1   1   1   1   1   0   0   0   1   0
##  [52]   0   0   1   0   1   0   0   0   0   1   1   0   0   0   1   1   0
##  [69]   0   0   1   0   0   0   1   1   1   0   0   0   0   0   0   0   0
##  [86]   1   0   0   0   0   1   1   0   1   1   0   0   1   0   0   0   1
## [103]   0   1   0   0   1   0   1   0   0   1   0   0   1   0   0   1   0
## [120]   0   0   0   0   1   0   0   1   1   0   0   1   1   1   0   1   1
## [137]   1   1   0   1   0   0   0   0   1   0   0   0   0   0   1   0   0
## [154]   0   0   1   1   1   1   0   0   0   1   0   1   0   1   0   1   0
## [171]   1   0   1   0   1   0   1   1   1   0   0   0   1   0   1   0   0
## [188]   0   1   1   0   0   0   1   1   0   0   1   1   0   0   0   1   1
## [205]   1   1   1   0   0   1   0   0   0   1   1   0   1   1   1   1   1
## [222]   1   1   0   1   0   0   1   1   1   1   0   0   0   0   1   1   0
## [239]   0   0   1   1   0   1   0   1   1   0   1   1   1   0   0   1   1
## [256]   0   1   1   1   0   0   0   1   1   0   1   1   0   1   1   0   0
## [273]   1   0   0   1   0   0   1   1   0   0   0   0   0   1   1   0   1
## [290]   0   0   1   0   0   0   1   0   0   1   1   0   1   0   1   1   1
## [307]   1   0   0   0   1   1   1   1   0   1   1   1   1   1   1   1   1
## [324]   0   0   1   1   1   0   1   0   1   0   1   1   1   0   1   1   1
## [341]   1   1   1   1   1   1   0   1   1   0   1   0   1   0   1   1   1
## [358]   1   1   1   1   1   1   1   1   1 305 159 251 215 101 224 306 199
## [375] 194 325 329 221 318 238  17 121 235 140  58 280 186 337 118  59  52
## [392]  92 355  77 349 164 211  86 144 297 210 214 347  91 181 338 216 150
## [409]  68 152   4  89 212 189 292  25 302 363 290  57 236 179 365 205 299
## [426] 285 108  29 267 275 293 139 122 213 317 323 136 300 259 354 169 166
## [443]  33 332 200 239 334 265 256 258 343 170 268 223 362 217  30  32 271
## [460]  83  81 269 253 147 312 219 218  14 346 124 231 273 148 260  90 336
## [477] 345  62 316 252   2 351 340  74 262 191 208 330 298  40 276 364 155
## [494]  35 321 197  65  37 133 295 178 130  55 112 278  75 183 250 326 319
## [511]  31 361 357 296 308 226 103 313 249 228 301  20  28 110  85 366 335
## [528] 206 134 272  69 356 180 274  73 341 104 360  60 247 109 358 137  22
## [545]  64 222 353 209  93 350 115 279 188 327  50  13 277 284 248  15  42
## [562] 331 322 120  98 190 227 187  27 153 172  23  67 303 289  88 270 287
## [579] 193 111  45 261 145  54 114 168  48 106  21 324 142 307 198 102  44
## [596] 154 141 311 344 291 339 116  36 286 245 352 167  61 333  11 225 161
## [613]  49 232  82   6   8 184 263  71 158 242 175   1 113 207 255 246 177
## [630]  63 204 160 119 195 149  18 233 257 151 315 359 125 244 202  24  87
## [647] 234 283 342 220 237  72 138 294 171 254 288   5 241 192 243 117 201
## [664] 196 176   7 264  94 229  38  79  19  34 348 266 310  76  51  97  80
## [681] 282  46  66 126 127 131 107 143 146 203 185 156   9 182 230 132 309
## [698]  47 281  99 174 129 328 157 165  56  10  12 105  43  41  39 314 163
## [715]  26 320  96 304 128 240 135  70  53 162  95  84 173  78 123  16   3
## [732] 100
```

```
length(low_draft)
```

```
## [1] 732
```

```
high_draft= c(lottery$Draft_No> 182, lottery$Draft_No)
high_draft
```

```
##   [1]   1   0   1   1   0   1   1   1   1   1   1   1   1   1   0   0   1
##  [18]   0   0   1   1   1   0   0   0   0   1   0   1   0   1   0   0   1
##  [35]   1   1   1   0   0   1   1   0   0   0   0   0   1   1   1   0   1
##  [52]   1   1   0   1   0   1   1   1   1   0   0   1   1   1   0   0   1
##  [69]   1   1   0   1   1   1   0   0   0   1   1   1   1   1   1   1   1
##  [86]   0   1   1   1   1   0   0   1   0   0   1   1   0   1   1   1   0
## [103]   1   0   1   1   0   1   0   1   1   0   1   1   0   1   1   0   1
## [120]   1   1   1   1   0   1   1   0   0   1   1   0   0   0   1   0   0
## [137]   0   0   1   0   1   1   1   1   0   1   1   1   1   1   0   1   1
## [154]   1   1   0   0   0   0   1   1   1   0   1   0   1   0   1   0   1
## [171]   0   1   0   1   0   1   0   0   0   1   1   1   0   1   0   1   1
## [188]   1   0   0   1   1   1   0   0   1   1   0   0   1   1   1   0   0
## [205]   0   0   0   1   1   0   1   1   1   0   0   1   0   0   0   0   0
## [222]   0   0   1   0   1   1   0   0   0   0   1   1   1   1   0   0   1
## [239]   1   1   0   0   1   0   1   0   0   1   0   0   0   1   1   0   0
## [256]   1   0   0   0   1   1   1   0   0   1   0   0   1   0   0   1   1
## [273]   0   1   1   0   1   1   0   0   1   1   1   1   1   0   0   1   0
## [290]   1   1   0   1   1   1   0   1   1   0   0   1   0   1   0   0   0
## [307]   0   1   1   1   0   0   0   0   1   0   0   0   0   0   0   0   0
## [324]   1   1   0   0   0   1   0   1   0   1   0   0   0   1   0   0   0
## [341]   0   0   0   0   0   0   1   0   0   1   0   1   0   1   0   0   0
## [358]   0   0   0   0   0   0   0   0   0 305 159 251 215 101 224 306 199
## [375] 194 325 329 221 318 238  17 121 235 140  58 280 186 337 118  59  52
## [392]  92 355  77 349 164 211  86 144 297 210 214 347  91 181 338 216 150
## [409]  68 152   4  89 212 189 292  25 302 363 290  57 236 179 365 205 299
## [426] 285 108  29 267 275 293 139 122 213 317 323 136 300 259 354 169 166
## [443]  33 332 200 239 334 265 256 258 343 170 268 223 362 217  30  32 271
## [460]  83  81 269 253 147 312 219 218  14 346 124 231 273 148 260  90 336
## [477] 345  62 316 252   2 351 340  74 262 191 208 330 298  40 276 364 155
## [494]  35 321 197  65  37 133 295 178 130  55 112 278  75 183 250 326 319
## [511]  31 361 357 296 308 226 103 313 249 228 301  20  28 110  85 366 335
## [528] 206 134 272  69 356 180 274  73 341 104 360  60 247 109 358 137  22
## [545]  64 222 353 209  93 350 115 279 188 327  50  13 277 284 248  15  42
## [562] 331 322 120  98 190 227 187  27 153 172  23  67 303 289  88 270 287
## [579] 193 111  45 261 145  54 114 168  48 106  21 324 142 307 198 102  44
## [596] 154 141 311 344 291 339 116  36 286 245 352 167  61 333  11 225 161
## [613]  49 232  82   6   8 184 263  71 158 242 175   1 113 207 255 246 177
## [630]  63 204 160 119 195 149  18 233 257 151 315 359 125 244 202  24  87
## [647] 234 283 342 220 237  72 138 294 171 254 288   5 241 192 243 117 201
## [664] 196 176   7 264  94 229  38  79  19  34 348 266 310  76  51  97  80
## [681] 282  46  66 126 127 131 107 143 146 203 185 156   9 182 230 132 309
## [698]  47 281  99 174 129 328 157 165  56  10  12 105  43  41  39 314 163
## [715]  26 320  96 304 128 240 135  70  53 162  95  84 173  78 123  16   3
## [732] 100
```

```
length(high_draft)
```

```
## [1] 732
```

```
m1to6_lowdraft= c(m1to6[draft_num< 183])
a11= sum(m1to6_lowdraft)
a11
```

```
## [1] 1407
```

```
m1to6_highdraft= c(m1to6[draft_num> 182])
a12= sum(m1to6_highdraft)
a12
```

```
## [1] 1159
```

```
m7to12_lowdraft= c(m7to12[draft_num< 183])
a21= sum(m7to12_lowdraft)
a21
```

```
## [1] 1445
```

```
m7to12_highdraft= c(m7to12[draft_num> 182])
a22= sum(m7to12_highdraft)
a22
```

```
## [1] 1123
```

```
p1=a11/(a11+a12)
p1
```

```
## [1] 0.5483242
```

```
p2= a21/(a21+a22)
p2
```

```
## [1] 0.5626947
```

```
test1<- matrix(c(a11,a12,a21,a22), nrow = 2,
               dimnames= list(month= c("1to6", "7to12"),
                              draft= c("low", "high")))

fet_p_val<- fisher.test(test1, alternative = "two.sided")
fet_p_val
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  test1
## p-value = 0.312
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.8437613 1.0549228
## sample estimates:
## odds ratio
##   0.943487
```

```
#result: Since p-value = 0.312, the p-value is not small enough to reject the null hypothesis,  conclusion will continue to hold even if the dependence is not properly accounted for. So what would you tell the congress when they call you to testify about the fairness of the lottery? What are the advantages and disadvantages of the Kruskal-Wallis test against Fisher’s exact test in this case?
#
```

```
#for further research:
#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.)
```


```
#Reference
#1,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.

#2, Landscaper.net (2009). The Military Draft and 1969 Draft Lottery for the Vietnam War. Last modified 2009-03-24. Confirmed 2011-05-26.

#3, "The vietnam lotteries" Last Updated June 18, 2009 https://web.archive.org/web/20120915225120/https://www.sss.gov/lotter1.htm

#4, RESULTS FROM LOTTERY DRAWING - Vietnam Era 1970 #https://web.archive.org/web/20120915232321/http://www.sss.gov/LOTTER8.HTM

#5, Eberhart, Dan. A pact with the living. 10-26-2016. Print

#6, Draft lottery (1969) https://en.wikipedia.org/wiki/Draft_lottery_(1969)#cite_note-SSS70-1

#7, “Would your draft number have been called?” USA Today, Gannett Satellite Information Network, 2017, www.usatoday.com/vietnam-war/draft-picker.

#8,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").
```