Homework 3

SDGB 7844, Prof. Nagaraja

Yuwen(Suyi) Wu 3:30 PM - 5:30 PM

11/01/2018

Question 1

Simulate the capture-recapture method for a population of size N = 5, 000 when n1 = 100 and n2 = 100 using the sample() function (we assume that each individual is equally likely to be "captured"). Determine m2 and calculate N^LP using Eq.1. (Hint: think of everyone in your population as having an assigned number from 1 to 5,000, then when you sample from this population, you say you selected person 5, person 8, etc., for example.)

Solution:

```
N <- 5000
n1 <- 100
n2 <- 100
s1 <- sample(x = 1:5000, size = 100, replace = TRUE)
s1 ## check for number captured

##   [1] 3247  798 4129 2384 3414 1316  684   79 3958 3617 2333 1473 2694
1776
##  [15] 3036 1045 3333 1021 2055 3648 3847 4391 1540 2047 4059 1217 1182
455
##  [29] 1074 4790   16 4837 2779 2712  771 2571 1723 4626 3576 4898  927
1522
##  [43] 4699 1764 1706 3765 2066 2817 4177 2850  463 4209 4911 1468 2781
352
##  [57] 4975 3225  311 2872 4466  893 2120 1514  523 2236 3390  593 4942
4395
##  [71]  533 1637  842  293 2420 2997 1875 2763 1579 1341 4653 2276 2530
3098
##  [85] 1158 4448  608  292 4991 3460 1307 2547  671 1623 1668  894 1615
144
##  [99] 1223 3745

s2 <- sample(x = 1:5000, size = 100, replace = TRUE)
s2 ## check for number captured in the second time

##   [1]  414 2402 2122  792 1260 2334 4048 1023 2123 4426 4037 1245 3114
883
##  [15] 4319 2527 3051 4836 2686 1692  432 3400   89  709 1672 3211 4755
2959
##  [29] 3763 4381 4963 3932  611 4407 4259  409 2206 1936 3364 1539 4381
```

```
4545
##  [43]  722 3831  363 1314 2885  473 3255 1196 2238 4860 1732 3196   29
3324
##  [57]  705 3144 3715 3989  899 1636 1319  233 4715 4545 2538 2738  256
3733
##  [71] 3182 4200 2685 2260 2548 4193  572  983 3542 1801 2915 3566 4697
4034
##  [85] 2902 3569 2293 4913  374 2237 2542  800 4477 4322 2027  593 1898
4568
##  [99] 4671 2933
```

```
s3 <- intersect (s1,s2)
s3 ## find the fish been captured twice
```

```
## [1] 593
```

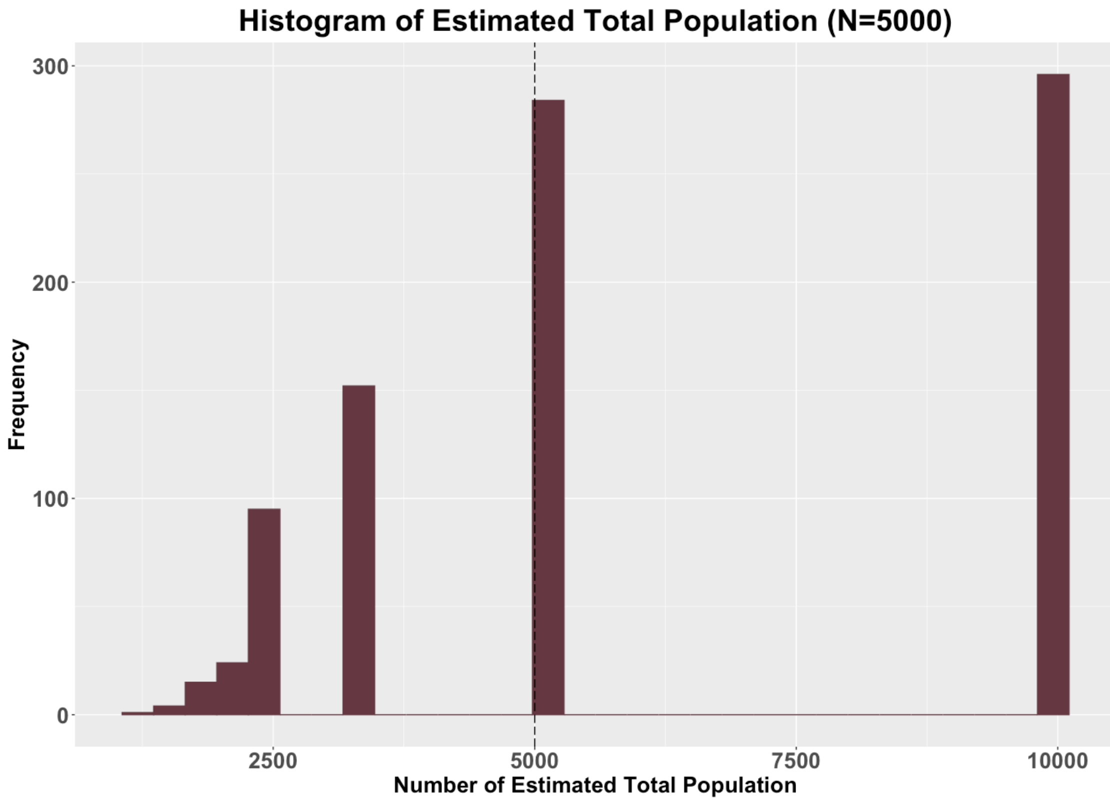
```
m2 <- length(s3)
Nlp <- n1*n2/m2
Nlp
```

```
## [1] 10000
```

Question 2

Write a function to simulate the capture-recapture procedure using the inputs: N, n1, n2, and the number of simulation runs. The function should output in list form (a) a data frame with two columns: the values of m2 and N^LP for each iteration and (b) N. Run your simulation for 1,000 iterations for a population of size N =5,000 where n1 = n2 = 100 and make a histogram of the resulting N^LP vector2. Indicate N on your plot.

Solution:



Question 3

What percent of the estimated population values in question 2 were infinite? Why can this occur?

Solution:

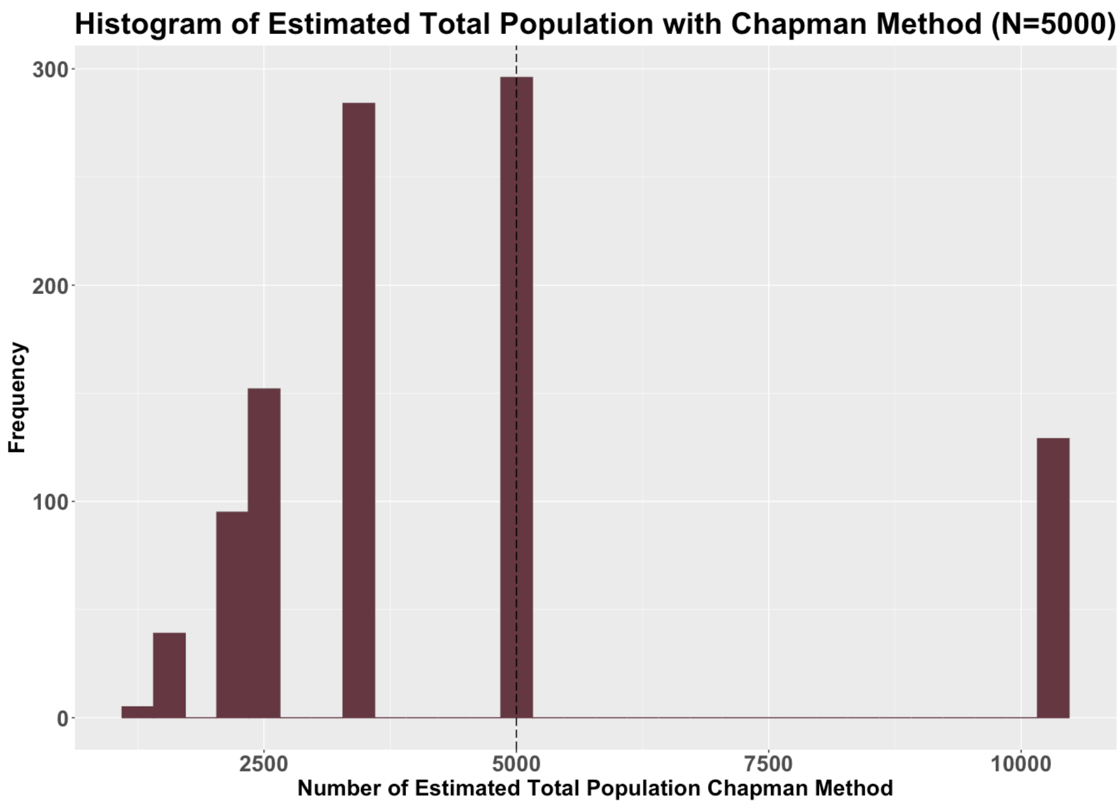
```
number <- length(which(is.infinite(list_Result$`Sim Result`$`NLP`)))
number/1000

## [1] 0.129
```

There are about 12.5% of estimated population were infinite. That is because when the intersection of n1 and n2 is zero, the estimated population value is infinite.

Question 4

An alternative to the Lincoln-Peterson estimator is the Chapman estimator: N^C = (((n1+1)(n2+1)/(m2 + 1)) - 1 Use the saved m2 values from question 2 to compute the corresponding Chapman estimates for each iteration of your simulation. Construct a histogram of the resulting N^C estimates, indicating N on your plot.

Solution:



Question 5

Solution:

```
n.sim <- 1000
sum(list_Result$`Sim Result`$NLP)/n.sim ## calculate the estimator with
infinite

## [1] Inf

NLP <- filter(list_Result$`Sim Result`, NLP != Inf) ## filter the infinite
value
NLP ## checking data

## # A tibble: 871 x 2
##       m2    NLP
##    <int>  <dbl>
##  1     1 10000
##  2     2  5000
##  3     1 10000
##  4     1 10000
##  5     1 10000
##  6     3  3333.
##  7     2  5000
##  8     3  3333.
##  9     2  5000
```

```
## 10     1 10000
## # … with 861 more rows

mean(NLP$NLP) ## calculate estimator without infinite value

## [1] 5974.892

bia.nlp <- abs(mean(NLP$NLP) - N) ## calculate bias of estimator
bia.nlp

## [1] 974.892

sum(N_C.vector/n.sim) ## calculate estimator in Chapman Method

## [1] 4440.726

bia.nc <- abs(sum(N_C.vector/n.sim) - N) ## calculate bias of estimator
bia.nc

## [1] 559.2741
```

As the result, Both estimators are biased when n1,n2 = 100.

Question 6

Based on your findings, is the Lincoln-Peterson or Chapman estimator better? Explain your answer.

Solution:

The chapmen estimators is better because the smaple size and the m2 we captured are not large enough. In lincoln-peterson method, we omited lots of infinite numbers, in this case, so we have bigger biased. For small size data, the estimator would be unbiased when n1 and n2 larger. Therefore, in this question, Chapman estimator would be better.

Question 7

Explain why the assumptions (a), (b), and (c) listed on the first page are unrealistic.

Solution:

(a)each individual is independently captured It is unrealistic because most fish are living together. In this case, the fish might not captured independently and fish would be captured together with other fish nearby. (b)each individual is equally likely to be captured We cannot make sure that every one is equally likely to be captured,and in realistic, captured fish is the fish that we can find. If someone hide for a long time, the exact population will be biased. (c)there are no births, deaths, immigration, or emigration of individuals (i.e., a closed population) This is not realistic because we could not make sure no death and birth in fish group. The fish also would migrate to other place or hide at the begining. So, this is unrealistic.