

Name: Jiahong Hu  
UNI: jh3561  
HW3  
STAT W 4201

**Problem 1: *What are the underlying assumptions for the capture–recapture estimation method mentioned in that paper?***

The underlying assumptions for the capture- recapture estimation methods mentioned in this paper are listed below.

1. The population of rat-inhabited lots being estimated is closed. That is, the total number of rat-inhabited lots does not change throughout the study period.
  - a. In this study, the population if rat-inhabited lots is not closed because property owner and city workers are actively trying to cleanse lots identified by the 311 system. However, the study is designed to wait six months between samples to minimize this problem.
  - b. General speaking, the capture-recapture estimation methods hold the assumption that during the interval between the preliminary marking period and subsequent recapture period, nothing has happened to upset the proportions of marked to unmarked animals. That is, no new individuals were born or immigrated into the population, and none died or emigrated. An exception is when there are equal proportions of deaths or emigrants for marked and unmarked individuals during the sampling period.
2. Rat-inhabited lots reported in the two sample periods are randomly and independently identified from the total population of rat-inhabited lots. This means that:
  - a. Rat-Inhabited lots are equally likely to be reported to 311;
    - i. Although rat-inhabited lots may not be reported to 311 with equal probability due to the differences in land use, geography, culture values or human population density, it is reasonable that reported sightings would be equally likely among inhabited lots since the neighborhood used in this study were constructor to represent communities with similar characteristics.
    - ii. The chances for each individual in the population to be caught are equal and constant for both the initial marking period and the recapture period. That is, marked individuals must not become either easier or more difficult to catch.
  - b. Any lot identified as rat-inhabited during the first sample is as likely to be identified during the second sample period as any other inhabited lot.
    - i. The 6 months buffer serves as a cool down time and every report from the first sample period has been resolved by the beginning of

the second period, which ensure s two samples are independent of each other.

- ii. The chances for each individual in the population to be caught are equal and constant for both the initial marking period and the recapture period. That is, marked individuals must not become either easier or more difficult to catch. Marking an individual does not make them more or less likely to be recaptured relative to unmarked individuals. Furthermore, there cannot be something unique about those animals that were initially marked that make them more conspicuous.
- iii. Sufficient time must be allowed between the initial marking period and the recapture period for all marked individuals to be randomly dispersed throughout the population.

### 3. Animals do not lose their marks

***Problem 2: Give other potential application of the capture–recapture estimation method.***

Capture-recapture estimation is a method commonly used in ecology to estimate an animal population's size. The method is most useful when it is not practical to count all the individuals in the population. For example,

1. Estimating the number of fish in a pond, or more generally, estimating the population size of various species in the wild.
2. Estimating the number of duplicate records on a list or a database.
3. Estimating the number of people needing particular services (i.e. services for children with learning disabilities, services for medically frail elderly living in the community), or with particular conditions (i.e. illegal drug addicts, people infected with HIV etc.) For example, estimating the number of drug addicts in the United States.

Let's discuss the fish example in detail.

Let  $N$  be the total number of fish in a pond, which is the number we are interested to approximate.

We could take a sample of fish, with size  $M$ , in a pond and tag each of the fish selected. A day later, we could take a second sample of fish, with size  $T$ , from the pond, and count the number of fish in this second sample that had been tagged the day before, which is of size  $R$ . Then, using capture and recapture method, we could estimate the total number of fish in the pond  $N = MT/R$ .

***Problem 3 State some limitations of the design of this study.***

1. A limitation of this study is that the fact that the population of rat-inhabited lots is not closed. Property owners and city workers are actively trying to clean lots identified by the 311 systems.
2. Use the count of 311 calls might not be a good indicator of rats' existence in that lot. There might exist large reservoirs of rats but never been reported. The estimation under the study would be underestimated, if that is the case. The accuracy of the estimation under the study is dependent on the level of call rate. If the call rate is very low, there is high possibility that lots with rats are not reported.
3. The study first applies the captures and recapture method to estimate the number of city lots harboring rates. Then, multiple the number of lots by the average number of rats per inhabited lot to recover the population rats.
  - a. The researcher thinks that there is only one colony in a lot and thus the study uses the 40-50 as the average number of rats per inhabited lot. However, the average number of rats per lot used in the study may not be a good estimate because multiple colonies might exit in a large lot. In that case, the real number of rat population would be largely underestimated. Moreover, It is hard to find an accurate estimate of the average numbers of rats per inhabited lot due to large variability between different lots. However, a good estimate of average number of rats per lot is very essential in this study because a wrong number would lead to a poor estimation even if you have a good estimation of inhabited-lots.
  - b. The multiplication method is likely to produce a relative large error because if there are errors in both numbers, the result will produce even large errors.
4. The method is likely to overestimate the population of rats in NYC. This is because we treat every lot with at least one rat sighting as evidence that a full colony of rats inhabited the lot during of the study period.
5. The study assume that a recaptured lot is as likely to be reported as any other rat-inhabited lot, however, due to culture difference in different lots or some other factors, the probability might be different.