FISH 621 Laboratory #2: Simple Mark Recapture

Curry Cunningham 2022

Instructions

The purpose of this lab is to:

- Increase our familiarity with uncertainty based on probability functions.
- Explore our standard estimators for simple mark-recapture estimators:
  - Petersen
  - Chapman
  - Bailey
- Implement likelihood profiling and our simple estimators.
- Examine the implications of changing sample size.

If you have a question during the lab, please un-mute yourself and ask, or type it into the chat box. There is a high likelihood that someone else has the same question. It is more fun if we all learn together in our distance-learning world.

I have posted the lecture slides to the ***Canvas site***, so you can reference this material as you work through the lab.

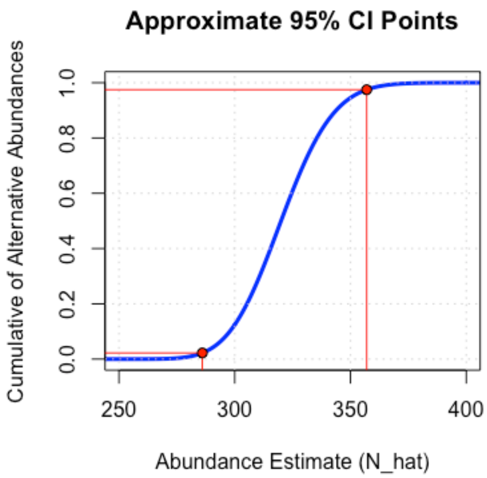
This and all other labs will be graded based on your attendance and participation.

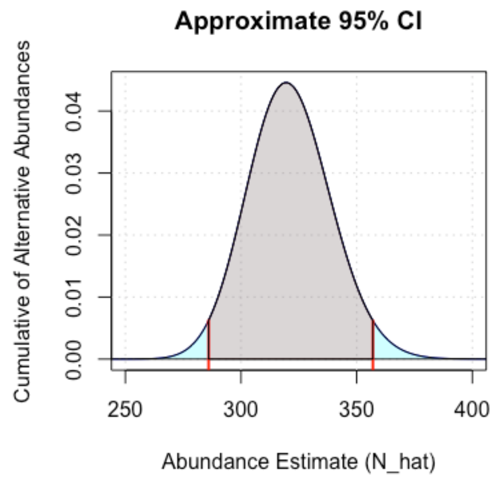
Lab Contents

- **621_Lab 1_Simple Mark Recapture.pdf** (this file)
- **621_Lab 1_Simple Mark Recapture.R** R script with exercises.

Exercise 2: Uncertainty in Abundance Estimates with Known P(Observation)

In this exercise we will explore what we can learn about uncertainty in our estimate of population size from our simple estimator for abundance given a known probability of observing an individual within the population: $\widehat{N} = \frac{n}{p}$ where n is the number of individuals detected and p is the detection probability known from a prior experiment.

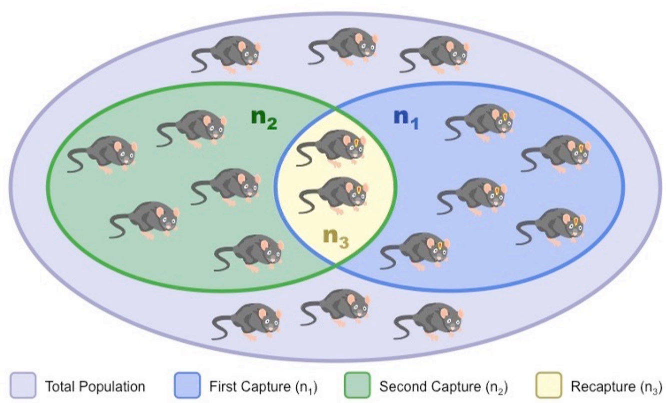




Exercise 3: Simple Mark-recapture Estimators

In lecture this week we have discussed several of our simple mark-recapture abundance estimators, applicable in the context of an experiment, where some number of individuals are captured and marked at time t_1, then released and some number of individuals are captured at time t_2 and inspected for marks.

In this case our number of captured and marked individuals is n_1, our number of sampled individuals at t_2 is n_2, and our observed number of marked individuals is m_2.



- $\dfrac{n_1}{N} = \dfrac{m_2}{n_2}$
  - The ratio of the number marked individuals n_1 to the total (unknown) population size N
    - Is equal to the ratio of the number of marked individuals m_2 (recaptures) to total individuals caught in t_2
- Petersen Estimator
  - $\hat{N} = \dfrac{n_1 n_2}{m_2} = \dfrac{n_2}{m_2/n_1} = \dfrac{n_2}{\hat{p}_2} = \dfrac{\#\, sampled}{est.prob.of\ being\ sampled}$

Chapman Estimator

- Chapman (1951) recommended the alternative estimator
  - Called the ”Chapman estimator”
    - $N^* = \frac{(n_1+1)(n_2+1)}{(m_2+1)} - 1$
- The Chapman estimator is unbiased if $n_1 + n_2 > N$
  - And has small bias otherwise, given by
    - $E(N^*) - N = -Ne^{-(n_1+1)(n_2+1)/N} = -Nb$

- Using a Poisson approximation Chapman showed that
  - $var(N^*) \approx N^2 \left(\frac{1}{\mu} + \frac{2}{\mu^2} + \frac{6}{\mu^3}\right)$ ***"Almost equal to"***
  - But, in practice we use the variance estimator
    - $\widehat{var}(N^*) = v^* = \frac{(n_1+1)(n_2+1)(n_1-m_2)(n_2-m_2)}{(m_2+1)^2(m_2+2)}$

Bailey’s Binomial Model cont.

- It can be shown that $\widehat{N}$ is once again the ***maximum likelihood estimator*** (MLE)
  - But is biased!
- Bailey (1951, 1952) recommended the alternative estimator, the “***Bailey estimator***”
  - $\widehat{N}_1 = \frac{n_1(n_2+1)}{(m_2+1)}$
    - Which has small bias of order $e^{-\mu}$
- The ***variance*** of the Bailey estimator is
  - $\widehat{var}(\widehat{N}_1) = v_1 = \frac{{n_1}^2(n_2+1)(n_2-m_2)}{(m_2+1)^2(m_2+2)}$
    - With small bias

Exercise 4: Hypergeometric Distribution

In this exercise we will explore the properties of the hypergeometric distribution, the syntax for its related functions in R, and how to fit the known (n_1, n_2, m_2) and unknown (N) parameters of our simple mark-recapture experiment within this framework. We will also explore its association the Bailey's Binomial Model which assumes sampling at time t_2 is with replacement.

Theory

- Values n_1 and n_2 are fixed by design
  - Therefore m_2 is a random variable
    - When assumptions or simple mark-recapture are satisfied, then the distribution of m_2, given n_1 and n_2, is the geometric distribution
- Geometric distribution
  - $f(m_2|n_1, n_2, N) = \binom{n_1}{m_2}\binom{N-n_1}{n_2-m_2}/\binom{N}{n_2}$
    - Where, "x choose y" is $\binom{x}{y} = \frac{x!}{y!(x-y)!}$
  - Expected value (***unbiased***)
    - $E(m_2|n_1, n_2, N) = \frac{n_1 n_2}{N} = \mu$
  - Variance
    - $var(m_2|n_1, n_2, N) = \frac{n_1 n_2}{N}\left(1-\frac{n_1}{N}\right)\left(\frac{N-n_2}{N-1}\right)$

- An approximately ***unbiased*** variance estimator
  - Based only on our estimated or known quantities is
    - $\widehat{var}(m_2|n_1, n_2, N) = m_2 \frac{\hat{N}}{\hat{N}-1}\left(1-\frac{n_1}{\hat{N}}\right)\left(1-\frac{n_2}{\hat{N}}\right)$

Bailey's Binomial Model

- If the experimental design is modified so that sampling is ***with replacement***
  - Then the ***binomial model*** is the appropriate statistical distribution
- $f(m_2|n_1, n_2, N) = \binom{n_2}{m_2} {p_1}^{m_2}(1-p_1)^{n_2-m_2}$
  - Where the probability of being marked is: $p_1 = \frac{n_1}{N}$
    - And $\binom{n_2}{m_2}$ is "n_2 choose m_2"

Bailey's Binomial Model cont.

- $f(m_2|n_1, n_2, N) = \binom{n_2}{m_2} {p_1}^{m_2}(1-p_1)^{n_2-m_2}$
- In this case the number of captures and recaptures
  - During the second time period
    - May include ***repeat captures*** of the same individual
  - If attention were ***restricted*** to ***unique*** individuals
    - Then the hypergeometric would be the correct distribution
- Under Bailey's model
  - The ***expect value*** of m_2 is
    - $E(m_2|n_1, n_2, N) = n_2 p_1 = n_2 \frac{n_1}{N} = \mu$, which is ***unbiased***
  - The ***variance*** of m_2 is
    - $var(m_2|n_1, n_2, N) = n_2 p_1(1-p_1) = E(m_2|n_1, n_2, N)(1-p_1)$

Exercise 5: Sample Design for Simple Mark-Recapture

In this portion of the lab our challenge is identify trade-offs in the variance of our Chapman and Bailey mark-recapture estimators, as we change our sample design (n1 and n2). Let's, assume our population has N=1,000 individuals.

Please explore a range of possible levels of effort for tagging n1 and recapture n2, under the assumption it costs $2.00 to capture and tag a fish at time t1 and $1.00 to capture a fish at time t2 and determine if it is marked.

Explore how the CV in your estimators change across ranges of n1 and n2, as a function of the associated total cost. ***We will discuss your results at the end of the lab.***