Capture & Recapture

Dr. Mujtaba A. Tahir
Basrah University
Agriculture College

How many fish are in the lake?

- **Student: “Throw many sticks of dynamite into lake and then count the dead fish!”**

- **Instead use capture-recapture method.**

- **Dictionary of Epidemiology defines capture-recapture as a method of estimating the size of a target population (or a subset of the population) using overlapping and presumably incomplete but intersecting sets of data about the population.**

Level of Statistical Sophistication

- Assumptions are fairly easy to understand.
- The derivation of the Lincoln-Peterson equation is not difficult.
- For children, proportions and basic algebra.
- For undergraduates, 95% CI's.
- For graduate students, log-linear models.

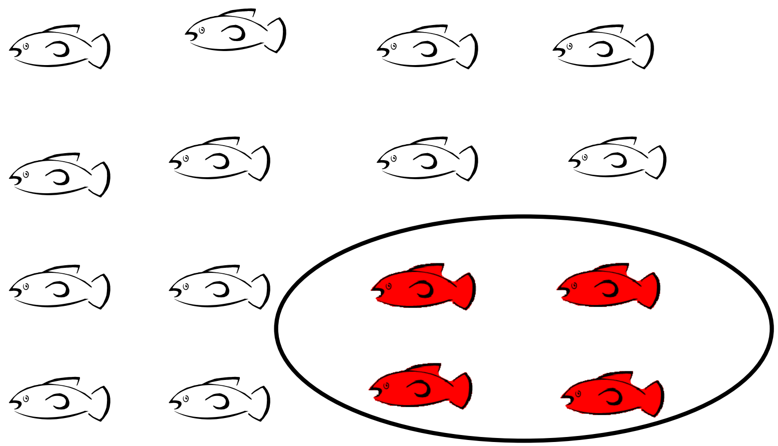
Brief History of Capture-Recapture Method

- 1662 – Used to estimate population of London.
- 1802 – Pierre Laplace explains formulas he used to estimate the population of France.
- 1896 – Peterson used this method to estimate populations of Danish fish.
- 1930 – Lincoln used it to estimate waterfowl in the U.S.
- 1949 – Sekar & Deming use these principles to estimate birth and death rates in India.
- 1954 – Chapman modifies Lincoln-Peterson formula.
- 1963 – Wittes & Sidel publish the first use of capture-recapture in epidemiology by estimating the number of hospital patients using methicillin.
- 1972 – Feinberg publishes paper on the use of log-linear models to analyze data from a multiple-list capture-recapture.

Capture-recapture Method

Phase I

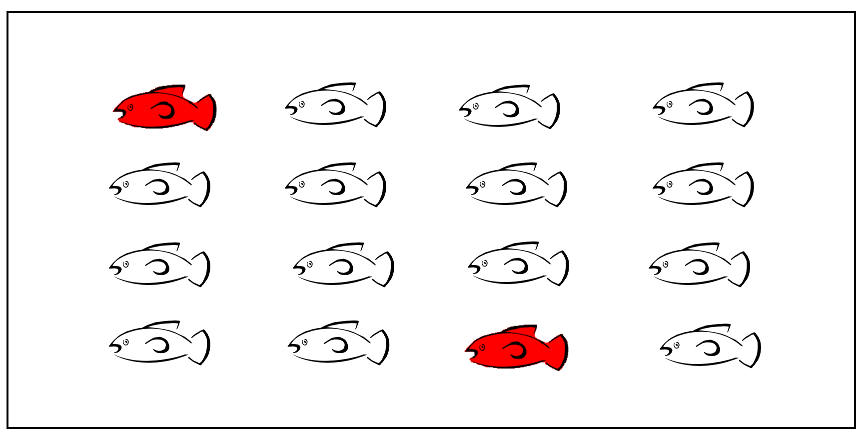
Capture
Mark
Release



M = 4

Phase II

Recapture
& Count



C = 16

R = 2

M, C, & R are used to estimate N (number of fish in lake)

Assumptions for Capture-Recapture Method

1. All the individuals (animals) have an equal chance of being caught in Phase I.

2. The population is closed. No emigration or immigration. (No animals can leave the area and no new animals can enter the area.)

3. There are no births or deaths during the time-period between Phase I and Phase II.

4. After the animals are marked they must be randomly redistributed into the population. (They must completely mix into the population.)

5. The probability that a marked animal can be caught in Phase II must remain the same as in Phase I. This means the marking cannot cause an injury and slow the animal. The marked animals cannot "learn" from Phase I and therefore are more likely be able to avoid capture in Phase II.

Violation of Assumption #2 & Possibly #3

50 king penguins with bands were compared to 50 penguins with implanted microchips. The banded birds had 40% fewer chicks and a 16% lower survival rate.

Violation of Assumption #5

- **Capture and tag(mark) M fish and release.**
- **At later date C fish are caught.**
- **R of these fish were previously tagged (recaptured).**
- **Let N = estimated # of fish in lake.**
- **Lincoln-Peterson method:**

$$N = (M \times C) / R$$

Example Using Lincoln-Peterson Equation

80(M) fish are captured, marked (tagged) and released.
60(C) fish are later captured,
of which 12(R) were recaptured (previously tagged).

$$N = (60 \times 80) / 12 = 400$$

Derivation of Lincoln-Peterson Formula

Let M = # marked in Phase I

C = # captured during Phase II

R = # recaptured in Phase II

N = # animals in target population

What is the proportion of animals in population that are marked?

M/N

What is the proportion in Phase II that were marked?

R/C (which is an estimate of M/N)

Set M/N = R/C , Cross multiply , M x C = R x N

Solve for N

N = (M x C) / R

More Advanced Topics for Undergraduates: Unbiased estimators & 95% Confidence Interval

The Lincoln-Peterson equation:
N = (M x C) / R is a biased estimator for N.
It slightly overestimates N and what should you do if R = 0?

N = [(M+1)(C+1) ÷ (R+1)] – 1 is an unbiased estimator with standard error

SE = SQRT [(M+1)(C+1)(M-R)(C-R) ÷ (R+1)(R+1)(R+2)]

As usual, the 95% CI for N is N±1.96(SE)

Redoing the last example where M = 80, C = 60, R = 12, & N = 400 yields

an unbiased estimate of N = 379 with SE = 82.6

and a 95% CI for N of (217, 541)

SQRT = square root

Demonstrations of Capture-Recapture :

- Grasshoppers or Crickets

(mark with white out – not PETA approved)

- White beans (mark with a black pen)

- Marbles

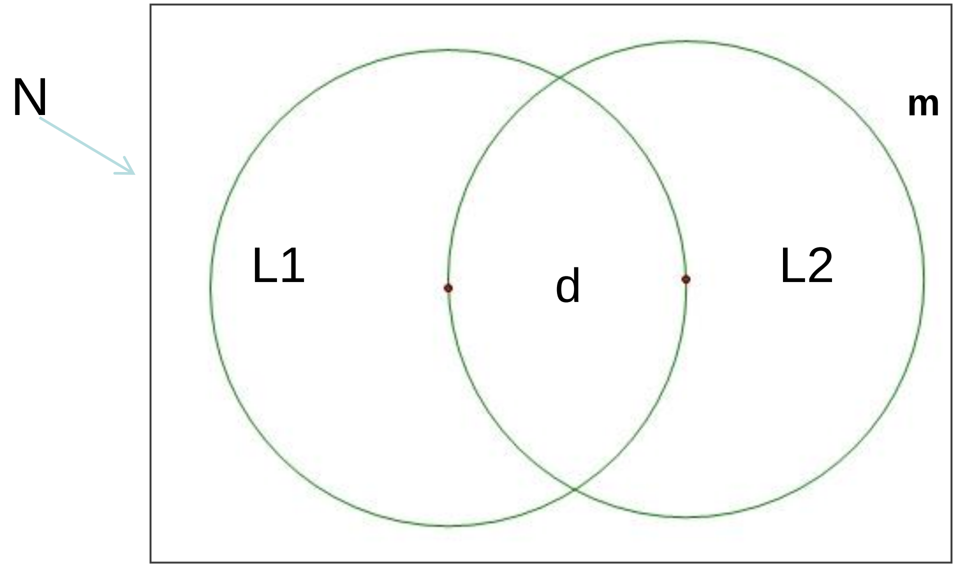
I use marbles and when I capture a fish (marble) I replace it with a red marble.

Results of 100 trials* using marbles to demonstrate the capture-recapture method and formulas with M =20, C=30 (Sorted by R).
There were actually 90 marbles in the container (lake).

R	# OCCUR	$\hat{N}$	$\hat{N}$_unbiased	L_95%CI	U_95%CI
1		600	325	0	663
2		300	216	29	403
3	4	200	162	42	282
4	6	150	129	45.9	212
5	9	120	108	46.5	169
6	23	100	92	45.7	138
7	19	86	80	44.3	116
8	24	75	71	42.8	100
9	6	67	64	41.2	87
10	3	60	58	39.6	77
11	5	55	53	36.7	68
12	1	50	49	35.4	61
Mean of 100 trials		94.2	86.1		

Performed by Brigid Norton

Two-list Capture-Recapture – substitute "caught" with "being on list"



L1 = # people on List1
L2 = # people on List2

d = # duplicates
(people on both lists)

m = # missing
(people on neither list)

N = total population

Chapman's formulas

$$N = [(L1+1)(L2+1) \div (d+1)] - 1$$

$$SE = SQRT\ [(L1+1)(L2+1)(L1-d)(L2-d) \div (d+1)(d+1)(d+2)]$$

95% CI for N is $N \pm 1.96(SE)$

SQRT = square root

from Gill (2002)