

Lecture Notes in Quantitative Biology
Estimating Population Size

Chapter 21

(from 6 December 1990)

revised 1995, + 26 November 1997
used in 1995, skipped in 1996

ReCap Numerical methods
Determining population size
Reasons
Good practices
Methods
Censusing
Complete enumeration
Sampling
Mark-recapture

ReCap. Numerical methods substitute repeated calculations for a single analytic solution.

They are used with increasing frequency to solve problems in quantitative biology, as the price of computing power drops, and the accessibility increases.

Examples:

Monte Carlo estimates of p-values

Finding an optimum (maximum) value

E.g. "mystery statistic" (the mean, minimizes squared deviations)

E.g. Finding "best" cladistic tree

Computing fluid flows, using Newton's laws

Computing genetic combinations

Rate of evolution by assembly of subunits

Numerical methods provide an answer with little in the way of complex mathematics. They provide an answer to the problem at hand, not to a general sets of problems. (Mathematical analysis to arrive at answers to sets of problems)

Today: Estimating population size.

Emphasis has been on model based statistics,
particularly models used in experimental analysis (GLM).

Today brief look at a different approach, used in survey design.

Sometimes called design based. Finite rather than infinite populations.

Wrap-up

Reasons for Determining Population Size

Goal is to determine an important quantity, the numbers of members of a populations. Obviously important in human populations: expected requirements for resources. Use of natural resources. Building schools and hospitals. Also important in exploited populations, such as codfish or trees. What is stock size ? Goal is to set harvest at level that maintains stock size by setting harvest at the renewal rate. Analogy of withdrawing interest from bank account, leaving capital intact. Important in natural populations. Rate of evolution depends on population size. If rate of appearance of new genetic material through mutation is 10^{-6} then takes one generation to produce new gene in population of a million, it takes on average 100 generations in population of 10,000. Knowledge of population size also important in very small populations, at verge of extinction.

Techniques can be classified as enumeration, sampling, mark-recapture

Good Practices.

1. Define the population limits in space (e.g. all sea lions in the N. Pacific).
Define quantity, state type of measurements unit
(N = sea lions in N. Pacific)
(N = sea lions per km of coast)
Define units used in sampling (e.g. quadrats, colonies).
If on ratio scale, state units (e.g. sea lions per colony).
A list of all possible units is called the frame.
This is the spatial range and resolution of the quantity N.
2. State model, assumptions, and methods.
This can often be done by citation to Seber 1984
3. Test assumptions. If this is not possible, evaluate assumptions.
4. Compare different methods, if possible, because different methods differ in their sensitivity to some assumptions.
5. Check indices (widely used, inexpensive) against absolute estimate (drain pond once).
6. Use accessory information to improve estimates.

Good Practices (continued)

7. Provide an estimate of uncertainty, or range of possible estimates.
Avoid giving misleading sense of certainty by reporting a single number.
 1. Theoretical, based on behaviour of organism.
 2. Statistical distribution, based on fit to similar data or situation.
 3. Taylor's Power Law $\text{Var} = \text{Mean}^a$ $1 < a < 3$
 4. Bootstrap

8. Power (Type II error). Is this study doomed to failure ? Can enough samples and information be gathered to obtain an answer of sufficient accuracy and precision, within logistic limits ?
Pilot studies. Use of preliminary results to design better study.

Comprehensive list of methods. (cf Seber)

I Censusing

A Complete enumeration.

B Sampling

1. Relative measures, to evaluate differences. Indices.
2. Absolute
 - a. Quadrat (1) Size and shape
 - (2) Number of quadrats
 - (3) Location - random or stratified random
 - b. Strip transect
 - c. Line transect. Estimate detection function.
 - d. Distance methods
 - (1) from random point
 - (2) from random individuals

II Mark-recapture

- A. Petersen. Closed population, using single mark.
- B. Closed population, using multiple mark.
- C. Open populations. Require estimate of immigration/emigration.
- D. Catch Effort. Individuals "marked" by removal.
- E. Change in ratio methods.

I Censusing

30 minute tour.
Basic idea of each one:
Type of data, assumptions, formal model
Computational route.
Advantages/disadvantages.

A **Complete enumeration.**

Example: People, trees. colonies (sea lions).

Evaluation: How complete is the enumeration ?
All individuals detectable ?
Immigration or emigration ?

Easier to defend. Just prove that no member of the population was missed.

Human populations.

Seems easy, just count up the number of individuals in some registry, such as social insurance numbers.

But existing registry are for other purposes, and some are incomplete.

E.g. SIN number registry incomplete for children.

So a registry is created each census by sending out forms.

Estimates of subpopulations not allowed, until very recently.

Wild populations. No registry. So complete enumeration is expensive.

Usually more practical in small populations, e.g. California Condor

I. Censusing

B. **Sampling.** Produces estimate N .

Harder to defend

More practical because cost is lower.

Usually not possible to cover population range,
so samples used to estimate.

1. Relative measures, to evaluate differences.

Used to examine trends through time, from place to place.

Catchability assumed constant, but unit of effort cannot
by used calculate absolute density, numbers/area.

Example: Catch per unit effort, indices of abundance.

Evaluation: are samples comparable ?
(though not representative).

I. Censusing

A. Sampling

2. Absolute.

Can calculate absolute density.

Example: scallops per dredge haul, trees per hectare.

Evaluation: Are samples representative and random ?

a Quadrat (1) Size and shape. Choice decided by logistics.

(2) Number of quadrats. " " "

(3) Location - random or stratified random
to be representative, to make inferences.

Notation: N = Population size

\hat{N} = estimate of N

A = area of population L^2

$D = N/A$ L^{-2}

a = quadrat area Typically L^2

but other units of effort possible

for example are swept per unit time

s = number of quadrats

$S' = A/s$ = number of quadrats possible

$p = sa/A$ = proportion of area sampled by s quadrats

pA = area sampled

$n = \sum x_i$ = number collected

\bar{x} = average per quadrat.

This is an estimate of N/A if random encounter:

$$\hat{N} = n/p = \bar{x} A/a = \bar{x} S'$$

Example:

$$n = 30 \quad p = 10\% \quad \text{then } \hat{N} = 30/0.1 = 300$$

I. Censusing

A. Sampling

2. Absolute

a. quadrat

Report uncertainty of estimate. Several possible methods.

To set confidence limits, we need the distribution of possible outcomes.

Could use bootstrap methods to compute limits from distribution generated by randomization, if we have several samples.

Could use a probability model, the binomial distribution.

If binomial sampling then $\text{Var}[N] = N \cdot q/p = N \cdot (1-p)/p$

Confidence limits:

Think of N as number of trials,

p as probability of success on each trial,

giving us a binomial distribution with $k = 300$ and $p = 0.1$

Cumulative distribution function, go from prob to outcome.

Use spreadsheet or package (e.g. Minitab) to make calculations.

```
MTB > invcdf .5;
SUBC> binomial 300 .1.
      K P(X LESS OR = K)   K P(X LESS OR = K)
      29      0.4719      30      0.5484

MTB > invcdf .025;
SUBC> binomial 300 .1.
      K P(X LESS OR = K)   K P(X LESS OR = K)
      19      0.0171      20      0.0287

MTB > invcdf .975;
SUBC> binomial 300 .1.
      K P(X LESS OR = K)   K P(X LESS OR = K)
      40      0.9746      41      0.9834
```

The 95% confidence limits around the observed count of 30 are 20 and 40. The 95% confidence limits around the estimate $N = 30/0.1 = 300$ are

$$L = 20/0.1 = 200 \quad U = 40/0.1 = 400$$

2. Absolute.

b Strip transect. Same as quadrat.

c. Line transect.

Estimate detection function. To correct for differential detection.

d. Distance methods (1) from random point

(2) from random individuals

II Mark-recapture

A. Petersen. Closed population, single mark.

Based on idea that probability of capture initially = probability of recapture

Probability
initial capture = Probability
recapture

$$\frac{n_{\text{initial capture}}}{N} = \frac{n_{\text{marked}}}{n_{\text{second capture}}}$$

Hence: $N = n_{\text{initial capture}} \cdot \frac{n_{\text{second capture}}}{n_{\text{marked}}}$

The estimate N is obtained by boosting the original number captured $n_{\text{initial capture}}$ upward by the ratio of $n_{\text{second capture}}/n_{\text{marked}}$

Advantages: Quick, requires only one type of mark, provides a good estimate if assumptions met.

Disadvantages. Assumptions often not met.

B. Closed population multiple mark

Use two or more marks to obtain more information

Two tags at same time to evaluate tag loss

different marks should give same estimate

tags can be in different proportions, depending on cost

Two more tags in sequence.

One tag at time 1

Next tag at time 2

Still another tag at time 3

More information allows mortality and recruitment also to be estimated.

These effects can then be removed to obtain less biased estimate of N

Usually better to carry out over short period, to reduce bias due to mortality and recruitment.

Pick times with low mortality and recruitment.

Advantages: mortality and tag loss can be taken into account

Disadvantage: more work, more chance of tag effects.

C. Open populations.

Immigration/emigration has to be measured, to remove biasing effect on estimate of N

This can be done with accessory information on rates of movement.

This can also be accomplished by using multiple marks (sequential) in a well designed study.

D. Catch/Effort methods.

Individuals are captured and removed. Ie. "marked"

Regress catch per unit effort against trial.

Intercept ("trial 0") is estimate of population size.

E. Change in ratio methods. Probability changes.

Application: How would you determine the number of newly settled codfish ?

You can ask me anything you like. I'll tell you, if it is known.

From this decide on methods that can be used.