Lecture 13 Sampling Design Issues: Stratified Random Sampling

Comment on the BACI Homework and Critiques

Brief General Comments on Jim Gilliam's Lecture last Thursday.

Important Basic Sampling Designs

Stratified Random Sampling –Major Focus

Special Issues on Sampling of Animals

Detection Probability Models

Methods of Estimating Detection Probability

Later Lectures

- Will consider a few of the BACI Designs you found. This should help us see the breadth and complexity of their use.
- I will consider a few of the papers that you critiqued that bring out important points. I was very happy with the way you handled the critiques and they were of a very high standard.
- I still have a few more to finish grading by Thursday.

Jim Gilliam Lecture Discussion

- He uses a mix of Experiments and Observational Studies and suggests iterating between the two. Nick Haddad made a very similar point.

- My opinion is that a strong overall research strategy will arise if it includes and integrates:

  -Mathematical Modelling of Popn or Community Dynamics

  -Experiments often on a component of the overall puzzle. Strong Inference due to replication, randomisation and control. Often on a small spatial and temporal scale.

  -Observational Studies on larger more realistic spatial and temporal scales. Weaker inference due to possible confounding effects will lead to the need to iterate back to the other approaches to understand and clarify confounded effects.

Jim Gilliam Lecture Discussion

- Observational Studies-

  -He focused a lot on principal components analysis as a way of reducing dimensionality.
- I will explore and illustrate this at some length later in the semester. I will talk about Principal Components Analysis and then using Principle Components in a multiple regression model.

All Sampling Design Topics

Basic Sampling Designs

Simple Random Sampling
Systematic Random Sampling
Stratified Random Sampling

A Brief Review

Simple Random Sampling

We began by considering simple random sampling without replacement which is the simplest probability sampling method.

It is analogous to using a completely random design in experimental design.

It is most useful if our sampling units are homogeneous

Now Consider the Simplest Possible Problem – Estimate the Population Mean (μ)

Simple Random Sampling: Estimation of the Population Mean

Represent the Population $\{y_1, y_2, \ldots, y_N\}$

Population Mean (Finite Population) - Parameter

$$\mu = \frac{\sum_{1}^{N} y_i}{N}$$

Represnt the Sample by $\{y_1, y_2, \ldots, y_n\}$

Sample Mean - Estimate of the Parameter

$$\bar{y} = \frac{\sum_{1}^{n} y_i}{n}$$

$\bar{y}$ is an estimate of μ.

Simple Random Sampling: Standard Error of Sample Mean

Standard Result

$$\text{SE of } \bar{y} = \frac{s}{\sqrt{n}}$$

Finite Population Result

$$\text{SE of } \bar{y} = \frac{s}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N}}\right)$$

Note : The second term is called the finite population correction factor.

Simple Random Sampling: Another Problem-Estimate Population Total

Population Total (Parameter)

This is a very important new parameter in finite populations

$$\tau = \sum_{1}^{N} y_i = N\mu$$

Population Total Estimate

$$\hat{\tau} = \mathrm{N}\overline{\mathrm{y}}$$

$$\mathrm{SE}(\hat{\tau}) = \mathrm{N\,SE}(\overline{\mathrm{y}})$$

$$= \frac{s}{\sqrt{n}}(\sqrt{N(N-n)})$$

Simple Random Sampling

Sample size needed for desired precision (p. 64-65) Example:

$$n = \left(\frac{z_{\alpha/2} CV}{r} \right)^2$$

$$n' = n/(1 + n/N)$$

$$z_{\alpha/2} = 1.96,\ \mathrm{N} = 910,\ \mathrm{CV} = 0.486,\ \mathrm{r} = 0.1$$

$$n = \left(\frac{1.96 x 0.486}{0.10} \right)^2 \approx 91$$

$$n' = \frac{91}{(1 + 91/910)} = 83$$

smaller than 91.

Systematic Random Sampling

Systematic random sampling often gives a better spatial coverage than simple random sampling. Here is an example.

- Think of sampling along a transect of length 100 meters where you start at a random point in first 10 m (7 meters from Excel) and then every 10th meter.The systematic random sample will be

  7,17,27,37,47,57,67,77,87,97

- I also chose a completely random sample of n=10 using Excel

  18,20,33,59,63,85,90,91,92,96

- Notice clumping along transect, random does not mean uniform!!!

NOTE: **Only major danger of systematic random sampling would be if there is some cyclical pattern of response along the transect. This does sometimes happen**

Stratified Random Sampling

Stratified Random Sampling

The population is divided up into L homogeneous strata.

The stratum sizes are N_1, N_2……… N_L

Within each stratum a simple random sample of size n_1, n_2……… n_L is taken.

It is important to realize that the sampling is independent in the different strata.

Note: Analogous to randomised complete block design in experimental design

Focus of Todays Lecture

Stratified Random Sampling

Why Stratify ?

-Strata of Interest

-Increase Efficiency of Overall Population Estimators

Stratified Random Sampling

How to Stratify and How Many Strata?

Pick Homogeneous Strata

Usually Use 5-10 Strata

Stratified Random Sampling

Purposes of Stratification.

-Strata of Interest

-Increase Efficiency of Overall Population Estimators

How to Stratify and How Many Strata?

Pick Homogeneous Strata to increase efficiency

Usually Use 5-10 Strata. If too many then the sample size within strata is too low

Stratified Random Sampling

Estimation Methods

-Individual stratum means, totals and variances.

-Over all population means and totals

Estimation of Overall Population Mean (μ)

- $$\bar{y}_{st} = \sum_{i=1}^{I} W_i \bar{y}_i$$

  This is a weighted average of the individual stratum means. Here

  $$W_i = N_i / N$$

Estimation of Overall Population Mean (μ)

$$\bar{y}_{st} = \sum_{i=1}^{I} W_i \bar{y}_i$$

•

$$Var(\bar{y}_{st}) = \sum_{i=1}^{I} W_i^2 Var(\bar{y}_i)$$

$$Var(\bar{y}_i) = \left[\frac{N_i - n_i}{N_i}\right] \frac{s_i^2}{n_i}$$

Estimation of Overall Population Mean (μ) :Example Williams et al (2002) P66

$$N_1 = 90, N_2 = 100, N_3 = 400, N_4 = 320, N = 910$$

$$n_1 = 21, n_2 = 21, n_3 = 21, n_4 = 20, n = 83$$

•

$$W_1 = 0.10, W_2 = 0.11, W_3 = 0.44, W_4 = 0.35 \sum_{1}^{4} W_i = 1$$

$$\bar{y}_1 = 20.50, \bar{y}_2 = 15.00, \bar{y}_3 = 30.00, \bar{y}_4 = 21.00$$

$$\bar{y}_{st} = 0.10(20.5) + 0.11(15) + 0.44(30) + 0.35(21)$$

$$= 24.25$$

Estimation of Overall Population Mean (μ) : Example Williams et al (2002)

$$Var(\bar{y}_{st}) = \sum_{i=1}^{I} W_i^{2} Var(\bar{y}_i)$$

$$Var(\bar{y}_i) = \left[\frac{N_i - n_i}{N_i}\right]\frac{s_i^2}{n_i}$$

- $$N_1 = 90, N_2 = 100, N_3 = 400, N_4 = 320, N = 910$$

$$n_1 = 21, n_2 = 21, n_3 = 21, n_4 = 20, n = 83$$

$$W_1 = 0.10, W_2 = 0.11, W_3 = 0.44, W_4 = 0.35$$

$$s_1^{2} = 16.0, s_2^{2} = 9.0, s_3^{2} = 9.0, s_4^{2} = 4.0.$$

$$Var(\bar{y}_{st}) = 0.11 \text{(Check yourself if you like)}$$

Example based on Helicopter Survey

Stratified Random Sample

Very Widely Used

Mule Deer Helicopter Example

Kufeld et al (1980) Journal of Wildlife Management, 44, 632-639.

- 8 strata of different sizes based on different regions in different habitats.
- Sampling Unit is a plot where a complete count of mule deer is made.
- Very good example of use of the standard sampling methodology. Individual stratum estimates and then overall population estimates (sum of the stratum estimates).
- Reasonable precision of estimates
- No adjustment for detectability. Assumes all animals seen in each plot so there is likely a negative bias on the estimates.

Table 1. Estimates of the wintering mule deer population based on a stratified random sample of eight strata on the Uncompangr Plateau, Colorado, 1977–79 (from Kufeld et al. 1980).

Parameter	Year	Stratum 1	2	3	4	5	6	7	8	Total
Area (km^2)	All	133	285	262	317	150	207	103	231	1,688
No. quadrats										
Total	All	206	440	405	489	232	319	159	357	2,607
Sample	All	12	35	23	38	13	25	9	38	193
Deer population estimates	1977	189	289	581	1,361	428	2,386	724	5,440	11,401
	1978	275	1,307	352	3,861	892	4,772	901	5,524	17,884
	1979	155	1,446	669	2,458	339	5,589	247	6,182	17,085
90% CL	1977									±2,205
	1978									±4,042
	1979									±2,951

Example based on Helicopter Survey : Detection Probability Issues

- There is no adjustment for detectability. Assumes all animals seen in each plot. This is likely untrue!
- There is likely a negative bias on the population estimates
- We now take a short detour into the whole issue of the estimation of detection probability because it clearly is so central to design in wildlife and fisheries studies.
- ST 506 will involve study of this topic in far more detail.

Components of Sampling Design

Standard Finite Sampling Methods

Need to understand some of the key concepts of finite population sampling. Absolutely crucial for any graduate student in ecology. However, often cannot apply without modification to animal sampling.

Special Issues on Sampling of Animals

Plants don't move! Animals do and are often hard to detect. We have to adjust animal counts to obtain good inferences. This requires detection probability be estimated. A very difficult and complex task.

Covered in detail in ST 506 but it has to be mentioned here as it can be so central to your design. Again absolutely crucial that you be aware of this issue

OVERVIEW OF METHODS OF ASSESSING ANIMAL POPULATIONS

Direct methods of monitoring

Census Method

Count all animals in the population
(a usually unrealistic practice)

Sampling Methods

Count animals in sampling units (or areas)

Absolute Abundance

Estimate popn size by adjusting for 'unseen' animals and only part of area being sampled

OUR MAIN FOCUS

Relative Abundance

Use incomplete count as an 'index' of population size

CHEAPER BUT PROBLEMS

CENSUS

Usually it is impossible to count all the members of a wildlife population for practical reasons.

- Exceptions are small, localized populations of highly visible and valuable endangered species.

- Examples are the California Condor the Puerto Rican Parrot (popn size in wild is only around 40 birds)

ABSOLUTE ABUNDANCE ESTIMATION

Not all area sampled and not all animals 'seen'

$$\hat{N} = C/\alpha\hat{\beta}$$

C = count of animals seen

α = fraction of area sampled (relates to (n/N) in s r sampling)

$\hat{\beta}$ = estimate of the fraction of animals seen or caught

Note: there are many ways to estimate β . For example, capture-recapture, line transects etc.

ABSOLUTE ABUNDANCE ESTIMATION

Example: Aerial survey of caribou in Alaska.

C = count of caribou seen = 551

α = fraction of area sampled= 0.1 (say n=150, N=1500 plots)

$\hat{\beta}$ = estimate of the fraction of animals seen or caught=0.5

Note: there are many ways to estimate β. For example, two independent observers in the plane.We will discuss details of this later

$$\hat{N} = C / \alpha \hat{\beta}$$

$$= 551 / (0.1 \text{ x } 0.5)$$

$$= 551 \, x \, 10 \, x \, 2$$

$$= 11{,}020$$

ABSOLUTE ABUNDANCE ESTIMATION

CAPTURE METHODS

- Capture-Recapture
- Removal and Catch-Effort
- Change-in-Ratio or Selective Removal

COUNT METHODS

- Line Transects
- Variable Circular Plots
- Double Sampling (e.g., ground, aerial)
- Counts from multiple observers with mapping

All can be viewed as methods to estimate β in

$$\hat{N} = C/\alpha\hat{\beta}$$

Complete Counts on a Subsample of Plots

Example : Aerial Survey with Complete Ground Counts on a Subsample of Plots

Description:

Key Assumption:

The "complete" count is a truly complete, accurate count on the same plot, at the same time as the aerial count.

Key Questions:

1. Is it possible to do an accurate, complete count? In aerial survey applications it is often impossible to do a complete ground count even in a few plots.

2. If it is possible, is it possible at reasonable expense for a reasonable size sample? Otherwise $\hat{\beta}$ based on a very small subsample, will be very imprecise.

Complete Counts on a Subsample of Plots

Example : Aerial Survey with Complete Ground Counts on a Subsample of Plots

Numerical Example : Thompson (1992) and p. 251-252 Williams et al.

Aerial Survey alone

$n = 20$ plots out of $\mathrm{N} = 100$ plots in whole area

240 moose detected

$\frac{100x240}{20} = 1200$ moose estimated if ignore detection probability

$$\alpha = \frac{20}{100} = 0.2$$

Ground Survey on Subsample of 5 plots

$$\hat{\beta} = \frac{56}{70} = \frac{\text{air count}}{\text{ground count}} = 0.80$$

Estimation of Population Size

$$\hat{N}^* = \frac{1200}{\hat{\beta}} = \frac{1200}{0.80} = 1500$$

or

$$\hat{N}^* = \frac{C}{\alpha\hat{\beta}} = \frac{240}{0.2x0.8} = 1500$$

Complete Counts on a Subsample of Plots

Example : Arctic Breeding Bird Survey with some plots searched twice.

All plots – standard search

Some plots – more thorough search, problem is if they are really complete counts

Another Method-We will return to Aerial Surveys!

CAPTURE-RECAPTURE MODELS

LINCOLN-PETERSEN MODEL

N - Population size

n_1 - No. of marked animals in the population

n_2 - Sample size

m_2 - No. of marked animals in the sample

Sample   Population

$$(m_2/n_2) \approx (n_1/N)$$

$$\hat{N} = n_1 n_2 / m_2$$

CHAPMAN'S MODIFICATION

TO REDUCE BIAS

$$\frac{m_2 + 1}{n_2 + 1} = \frac{n_1 + 1}{N + 1}$$

$$\hat{N}_c = \frac{(n_1 + 1)(n_2 + 1)}{(m_2 + 1)} - 1$$

This estimator is approximately unbiased.

Unbiased: If you repeated the study many times and took the average N, it would be equal to N.

PETERSEN MODEL ASSUMPTIONS

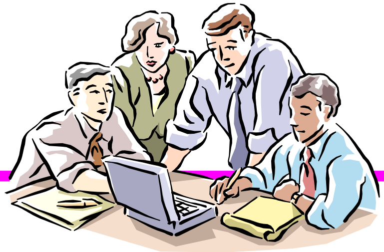
1. Closure

2. Equal Catchability

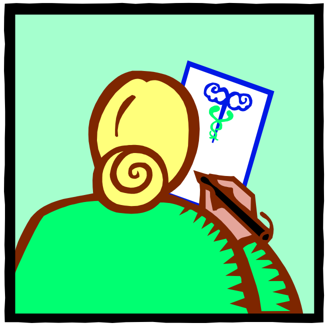
3. Zero Mark Loss (Marking is definitive)

Now returning to Aerial Surveys

Multiple Observers



Two independent observers



Apply the Lincoln-Petersen model!!

n_1 - seen by first observer

n_2 - seen by second observer

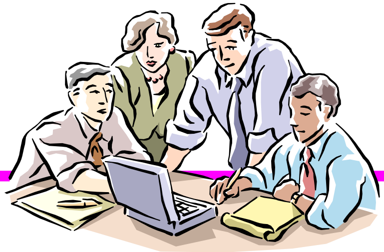
m_2 - seen by both observers

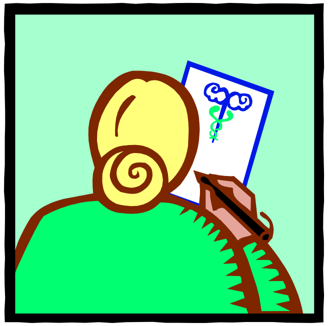
$$\hat{N} = \frac{n_1 n_2}{m_2}$$

$$\hat{N} = \frac{n_1}{\hat{p}_1}$$

$$\hat{p}_1 = \frac{m_2}{n_2} = \hat{\beta}$$

Multiple Observers





Two independent observers

Key Assumptions

1. No matching errors
2. Closed population
3. The observers are really independent

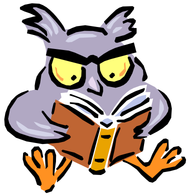
Multiple Observers

Two independent observers

Examples: aerial surveys

-Two observers in same plane

-One observer in plane, one on ground



Examples: Bird point counts

-Two observers at the same points

Multiple Observers

Two independent observers

Example: Henny and Burnham (1971) JWM.

Osprey nests- $\alpha = 1$ i.e., total area is sampled , there is one air observer and one ground observer.

Data

$n_1 = 51$ seen by air observer

$n_2 = 63$ seen by ground observer

$m_2 = 41$ seen by both observers

Two independent observers

$$\hat{N} = \frac{(n_1+1)(n_2+1)}{(m_2+1)} - 1$$

$$= \frac{52x64}{42} - 1 = 78.24$$

$$\text{SE}(\hat{N}) = 3.11$$

$$\hat{\beta}_{\text{air}} = \frac{n_1}{\hat{N}} = \frac{51}{78.24} = 0.65$$

$$\hat{\beta}_{ground} = \frac{n_2}{\hat{N}} = \frac{63}{78.24} = 0.81$$

Modeling Availability & Perception Detection Processes

Two Processes

Availability- animals have to be available to be detected.In many animal populations not all animals are available.(eg dugong-underwater, birds - don’t sing)

Perception- even if animals are available then they still have to be detected. This is also uncertain.(a dugong on the surface may still be missed)

MODELING OVERALL DETECTION PROBABILITY

The probability of detection is made up of a probability that the area is sampled plus an availability process and a detection process

P [animal detected] = P[area sampled]x

P[animal available] x

P [animal detected l animal available]

$$\hat{p} = p_{area}\hat{p}_a\hat{p}_d$$

Estimating Availability & Perception Detection Processes

Availability- **Very hard to estimate-special studies. For example one can use the robust capture-recapture design. Often assume all animals are available and do nothing which may not be good depending on the application!**

Perception- **The methods I mentioned earlier can be used to estimate this quantity and often it assumed to be the total detection probability.**

Some References on Detection Probability

Buckland, S.T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2001). Introduction to Distance Sampling. Oxford University Press, Oxford.

Pollock, K. H., Marsh, H., Bailey, L. L., Farnsworth, G. L., Simons, T. L., and Alldredge, M. W. (2004) Separating components of detection probability in abundance estimation: An overview with diverse examples. In W. L Thompson Ed. Sampling Rare and Elusive Species: Concepts, Designs and Techniques for Estimating Population Parameters. Island Press, Washington D. C.

Bailey, L. L., T. R. Simons, and K. H. Pollock.(2004). Estimating detectability parameters for plethodon salamanders using the robust capture-recapture design. Journal of Wildlife Management, 68: 1-13.

Pollock, K. H., Marsh, H., Lawler, I. and Alldredge, M. W. (2006). Modeling availability and perception processes for strip and line transects: an application to dugong aerial surveys. Journal of Wildlife Management 70, 255-262.

Simons, T.R., M.W. Alldredge, K. H. Pollock, and J. M. Wettroth. (2007). Experimental analysis of the auditory detection process on avian point counts. The Auk 124(3): 986-999.

Stratified Random Sampling

Total Sample Size and Sample Allocation Issues for Strata?

Stratified Random Sampling: Sample Allocation Rules to Obtain Better Precision

-Equal Allocation – **allocate the sample size equally in all the strata.**

-Proportional Allocation – **allocate proportional to the size of the strata-very widely used**

-Optimal – **allocate proportional to size and stratum variances and inversely proportional to costs in the different strata.**

Stratified Random Sampling: Sample Allocation Rules to Obtain Better Precision

Equal Allocation – allocate the sample size equally in all the strata.

- Not usually sensible unless all strata are equal size in terms of overall estimates precision.
- However, maybe good if you want to compare stratum means as your primary focus

Stratified Random Sampling: Sample Allocation Rules to Obtain Better Precision

Proportional Allocation – allocate proportional to the size of the strata-very widely used

- Commonly used-takes account of stratum sizes being different.

$$n_i = n(N_i / N) = nW_i$$

Stratified Random Sampling: Sample Allocation Rules to Obtain Better Precision

Proportional Allocation –

$$n_i = n(N_i / N) = nW_i$$

Example with 4 strata

$n_1 = 83x0.10,\ n_2 = 83x0.11,\ n_3 = 83x0.44,\ n_4 = 83x0.35$

$n_1 = 8.3, n_2 = 9.1, n_3 = 36.5,\ n_4 = 29.05$

Note the weights given earlier in lecture

Use rounded values

$n_1 = 8, n_2 = 9, n_3 = 37, n_4 = 29$

for total $\mathrm{n} = 83$

Stratified Random Sampling: Sample Allocation Rules to Obtain Better Precision

Optimal Allocation –

- Takes account of stratum sizes, different variances, and different costs of sampling in different strata (see p. 67 of Williams et al. (2002) reference)

- Optimal Allocation is not used as much as proportional allocation but it can result in a gain in precision if the costs and variances are known or well estimated from a prior study or a pilot survey.

Thursday Ted Simons Lecture

Simons, T.R., M.W. Alldredge, K. H. Pollock, and J. M. Wettroth.(2007).Experimental analysis of the auditory detection process on avian point counts. The Auk 124(3): 986-999.