

# Lecture 11 Sampling Design Issues

---

## Today's Lecture

### When is Sampling Design Important?

Experiments

Observational Studies

### Important Basic Sampling Designs

Simple Random Sampling

Systematic Random Sampling

Stratified Random Sampling

# Reminder

---

- ◆ Read Jim Gilliam's paper as he is visiting on Thursday

# All Sampling Design Topics

---

## Basic Sampling Designs

Simple Random Sampling

Systematic Random Sampling

Stratified Random Sampling

## More Advanced Sampling Designs

Cluster and Multi-Stage Sampling

Double Sampling

Adaptive Sampling

## Special Sampling Designs for Animals

The centrality of detection probability

Capture-Recapture, Distance Sampling and other techniques discussed very briefly as this is focus of ST 506.

# When Does Sampling Design Become Important?

---

## “True” or “Quasi” Experiments

Because the experimental unit is too large or too difficult to be censused (or measured) completely.

This is almost always the case in practice.

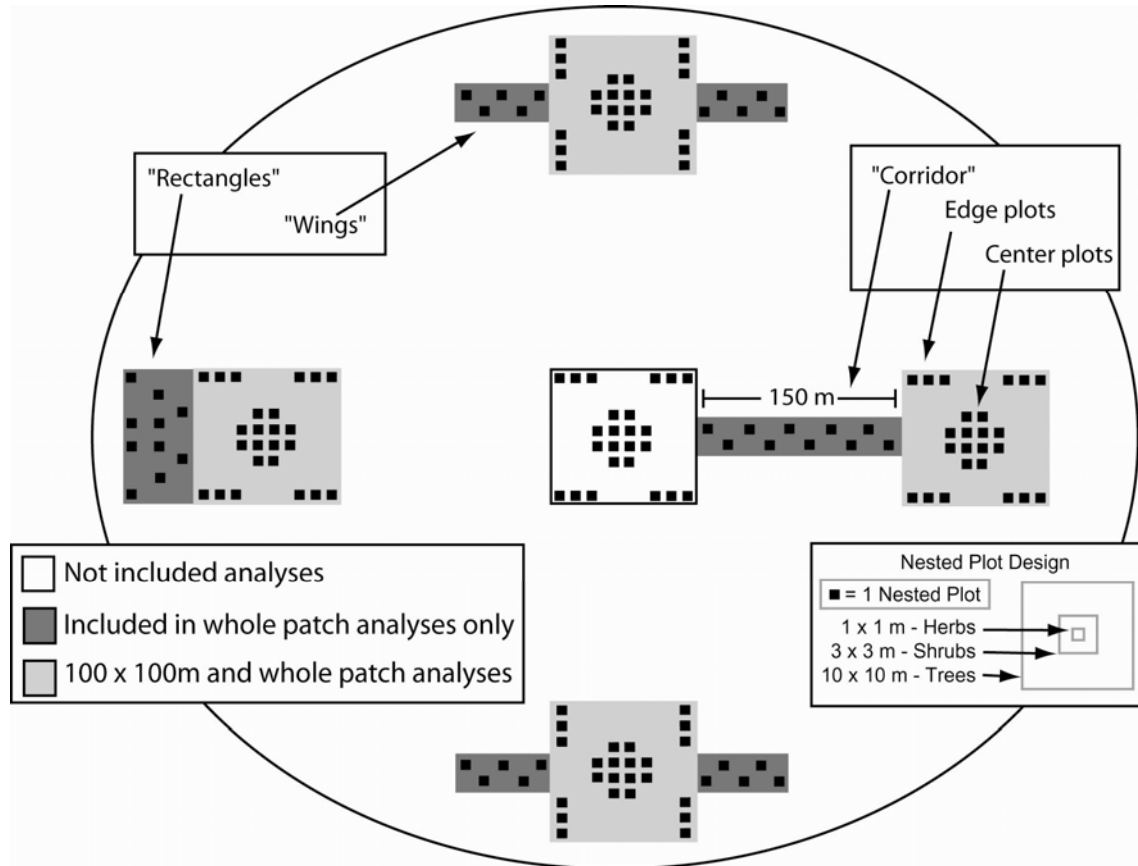
Sometimes the sampling design will even need to have a nested structure depending on what is measured.

# Damschen Corridor Study on Plant Species Richness

---

- ◆ To study the plant community and how dispersal may be aided by corridors.
- ◆ Recall exptl design was a Randomised block design (8 Blocks) with 3 treatments.
- ◆ Recall treatment factor applied to the patches has these types
  - Connected
  - Isolated Winged
  - Isolated Rectangular
- ◆ A Key point was the structure of the expt units
- ◆ Another key point is now the structure of the sampling units.
- ◆ This is a very good example of the complexities of sampling design.

**Figure 1.3.** Patch and plot design. Here the point is that the black squares represent one nested plot with a 1 x 1 m, 3 x 3 m, and 10 x 10 m plot sampling herbs, shrubs, and trees, respectively.



# Damschen Corridor Study: Experimental Design Summary

---

- ◆ Experimental unit is the patch.
- ◆ Allocation of Units-  
Randomised Block Design.
- ◆ Treatment Structure-  
Connected  
Isolated Winged  
Isolated Rectangular

# Damschen Corridor Study: Sampling Design Summary

---

- ◆ The sampling unit varies depending on the variable measured due to the costs of the measurements
  - Whole Patch (census)
  - 10x10m (Trees)
  - 3x3m (Shrubs)
  - 1x1m (Herbs)
- ◆ Here the sampling units have a nested structure with the smaller ones being inside the larger ones and this is quite common in many examples.
- ◆ There are different designs for choosing the actual sampling units to be measured
- ◆ Key point is the structure of the sampling units and that it may not be trivial to get an estimate of the value of the variable for the whole experimental unit.

# When Does Sampling Design Become Important?

---

## **.In Observational Studies**

One or Multiple Populations will usually be sampled for a wide variety of purposes and especially Inventory and Monitoring Studies in Fisheries and Wildlife.

Our speakers will provide many examples: For example,

- Simons- Old Growth & Second Growth Forests Bird Sampling.
- Hightower- Sampling Fisheries Populations
- Wentworth- Plant Communities Will Be Sampled so that Classification Techniques can be applied.

# All Sampling Design Topics

---

## Basic Sampling Designs

Simple Random Sampling

Systematic Random Sampling

Stratified Random Sampling

## More Advanced Sampling Designs

Cluster and Multi-Stage Sampling

Double Sampling

Adaptive Sampling

## Special Sampling Designs for Animals

The centrality of detection probability

Capture-Recapture, Distance Sampling and other techniques discussed very briefly as this is focus of ST 506.

# Sampling Design Definitions

---

- ◆ **Population**- consists of  $N$  sampling units with  $N$  Known and Finite. Sampling units are labeled  $1, 2, 3, \dots, N$ . To begin assume that the sampling units are all of the same size. Think of them of plots in study area
- ◆ **Sample** –consists of  $n$  sampling units. Drawn without replacement by some probabilistic method like simple random or stratified random sampling.

# Sampling Design Definitions

---

- ◆ Variables of interest- e.g numbers of individual animals or plants are measured on the individual sampling units (plots) ( $y_i$ )
- ◆ Population parameters like Population Mean ( $\mu$ ) or Population Total ( $\tau$ ) can be estimated from the sample values.

# Probability Sampling

---

We need to use probability based sampling methods because they allow us to be able to study the properties of the samples we collect.

A probability based sampling procedure is one where we know the probability of drawing each sample.

We discourage convenience samples or purposive samples because there we do not know their properties.

# Simple Random Sampling

---

We begin by considering simple random sampling without replacement which is the simplest probability sampling method. In this case each of the distinct possible samples has the same probability.

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 ( $\mu$ )

# Simple Random Sampling: Estimation

---

Now we consider the estimation of the finite population parameters.

1. Estimate the Population Mean ( $\mu$ )
2. Estimate the Population Total ( $\tau$ )

# Simple Random Sampling: Estimation of the Population Mean

---

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

Population Mean (Finite Population) - Parameter

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

Represent the Sample by  $\{y_1, y_2, \dots, y_n\}$

Sample Mean - Estimate of the Parameter

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

$\bar{y}$  is an estimate of  $\mu$ .

# Simple Random Sampling: Properties of Sample Mean as an Estimate of the Popn Mean

---

First

$\bar{y}$  is an Unbiased Estimate of  $\mu$ .

Second

What is the SE of  $\bar{y}$ ?

# Simple Random Sampling: Standard Error of Sample Mean

---

Standard Result from say ST511

Applies to Infinite Populations

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

where

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

# 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

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

Population Total Estimate

$$\hat{\tau} = N\bar{y}$$

$$\text{SE}(\hat{\tau}) = N \text{SE}(\bar{y})$$

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

# Simple Random Sampling

## Cottontail Rabbit Example P64 Williams et al.(2002)

---

### **Finite Population:**

There is a set of  $N = 1000$  sampling units in the population.

$n = 100$  sampling units in the sample.

The sampling units are chosen by simple random sampling without replacement.

There are 1600 total animals counted in the 100 plots in the sample.

$$\hat{\mu} = \bar{y} = \sum_{i=1}^n y_i / n = 1600 / 100 = 16,$$

$$\text{Note : } s^2 = 40$$

**The population total estimate is:**

$$\hat{\tau} = N \bar{y} = 1000 \times 16 = 16000$$

**We are assuming all animals are detected for the moment.**

## Simple Random Sampling

### Cottontail Rabbit Example P64 Williams et al.(2002)

---

$$\hat{\tau} = N \bar{y} = 1000 \times 16 = 16,000$$

$$\begin{aligned} \text{Var}(\hat{\tau}) &= N^2 \text{Var}(\bar{y}) = N(N-n) \frac{s^2}{n} \\ &= 1000(1000-100) \frac{40}{100} = 360,000 \end{aligned}$$

$$SE(\hat{\tau}) = \sqrt{\text{Var}(\hat{\tau})} = 600$$

$$95\% \text{ CI } \hat{\tau} \pm 1.96 \times SE(\hat{\tau})$$

$$16,000 \pm 1.96 \times 600$$

$$16,000 \pm 1,176$$

# Simple Random Sampling

---

Is this a Reasonable Design to Use is a Key Question? Populations often have a lot of heterogeneity in which case stratified random sampling is better.

Sample Size Calculation – How large a sample do we need to take is another key design issue??

# Simple Random Sampling

---

**Planning Your Study. Do a pilot study if possible!!!**

**Sample size needed for desired precision (p. 64)**

**$r$  is desired half width as a proportion of the popn mean  $\mu$**

$\bar{y} \pm r\mu$  with  $(1 - \alpha)$  confidence level

$$n = \left( \frac{z_{\alpha/2} CV}{r} \right)^2 \text{ standard result from say ST 511}$$

Estimated  $CV = s / \bar{y}$  from pilot study or previous survey

$n' = n / (1 + n / N)$  takes account of sampling without replacement

## Simple Random Sampling

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

---

$$n = \left( \frac{z_{\alpha/2} CV}{r} \right)^2 \qquad n' = n / (1 + n / N)$$

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

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

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

smaller than 91.

# Systematic Random Sampling

---

**Is it Better than Simple Random Sampling?**

**The answer, I think, is that it depends on the situation.**

# 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

# Simple and Systematic Random Sampling Summary

## Simple and Systematic Random Sampling Summary

---

**N sampling units in population, n in the sample**

**Good if the population is homogeneous.**

**The key result is the finite population correction factor on the variance and standard error of the sample mean due to the sampling being without replacement.**

**Applications to surveys of human popns, area surveys to estimate animal population abundance**

**Modify to Systematic Random Sampling sometimes**

**Now we move onto considering how to sample heterogeneous populations using stratified random sampling to improve the precision over simple random sampling.**

# Stratified Random Sampling

---

- ◆ Very Widely used option.
- ◆ Useful when the population is **heterogeneous** and it is possible to establish strata which are reasonably **homogeneous** within each one.

**Note:** Analogous to Randomised Complete Block Design in Experimental Design.

# Stratified Random Sampling

## Stratified Random Sampling

---

**The population is divided up into  $L$  homogeneous strata.**

**The stratum sizes are  $N_1, N_2, \dots, N_L$**

**Within each stratum a simple random sample of size  $n_1, n_2, \dots, n_L$  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

# Stratified Random Sampling

## Stratified Random Sampling

---

Think of breaking up a region into 3 habitat strata



$N_1$  units  
sample  
 $n_1$

$N_2$  units  
sample  
 $n_2$

$N_3$  units  
sample  $n_3$

# Stratified Random Sampling

---

## Why Stratify ?

**-Strata of Interest**

**-Increase Efficiency of Overall Population**

**Estimators by choice of strata that are homogeneous within each.**

# Stratified Random Sampling

---

## How to Stratify and How Many Strata?

**Pick Homogeneous Strata**

**Usually Use 5-10 Strata as it becomes unwieldy with  
too many**

# 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.
-



# 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.
-

# More Details on Stratified Random Sampling in Later Lecture

---