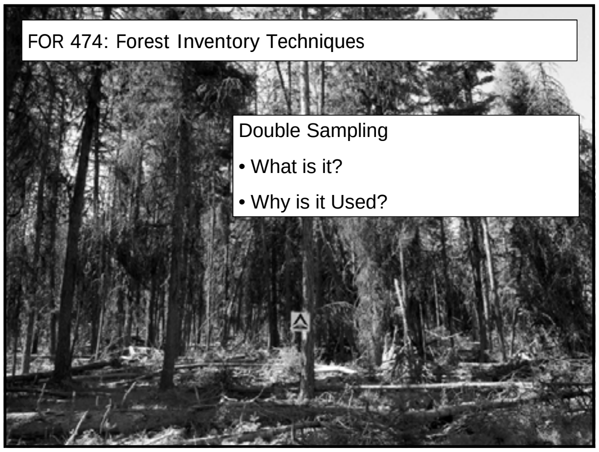
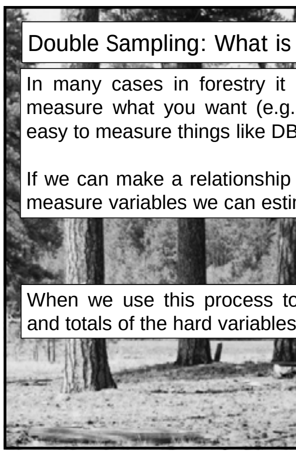


FOR 474: Forest Inventory Techniques

Double Sampling

- What is it?
- Why is it Used?

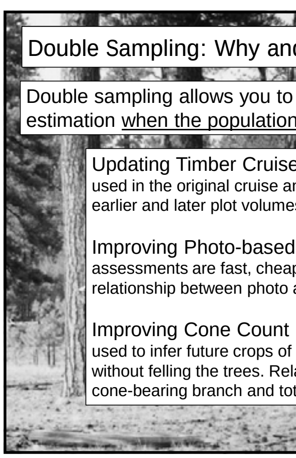


Double Sampling: What is it?

In many cases in forestry it is too expensive or difficult to measure what you want (e.g. total stand volume) but really easy to measure things like DBH

If we can make a relationship between the easy and hard-to-measure variables we can estimate the hard variables.

When we use this process to estimate the population mean and totals of the hard variables this is called Double Sampling.



Double Sampling: Why and Where is it used?

Double sampling allows you to use regression and ratio estimation <u>when the population mean or total is unknown</u>

Updating Timber Cruises: Re-measure a sample of the plots used in the original cruise and calculate relationship between earlier and later plot volumes.

Improving Photo-based Timber Cruises: As photo assessments are fast, cheap, and cover large areas, establish relationship between photo and ground cruise.

Improving Cone Count Estimates: Cone counts are often used to infer future crops of seed trees but are difficult to do without felling the trees. Relationships between cone count on a cone-bearing branch and total cone count are often used.

Source: Johnson p811

Double Sampling: Regression

We will assume that relationships are always linear.

We have a large sample of easy-to measure *primary* variable - say DBH - of size n_p

We have a smaller sub-sample of the hard-to-measure *secondary* variable – say maximum tree heights - of size n_s

To produce an unbiased estimate of tree height at any DBH, the regression between DBH and height using the smaller sample n_s can be calculated.

Double Sampling: Regression

Estimate of Population Mean:

$$\bar{y}_{np} = \bar{y}_{ns} + b\left(\bar{x}_{np} - \bar{x}_{ns}\right)$$

Height mean = height sample mean + slope * (sample mean for large sample – sample mean for smaller sub-sample)

Estimate of Population Variance:

$$s^2_{yx} = \frac{1}{n_s - 2}\left[\sum_i^{n_s}(y_i - \bar{y}_s)^2 - b^2\sum_i^{n_s}(x_i - \bar{x}_s)^2\right]$$

Double Sampling: Sample Size via Budget Constraint

C_s = cost of locating and evaluating 1 sample unit of the smaller secondary sample (n_s)
C_p = cost of locating and evaluating 1 sample unit of the large primary sample (n_p)

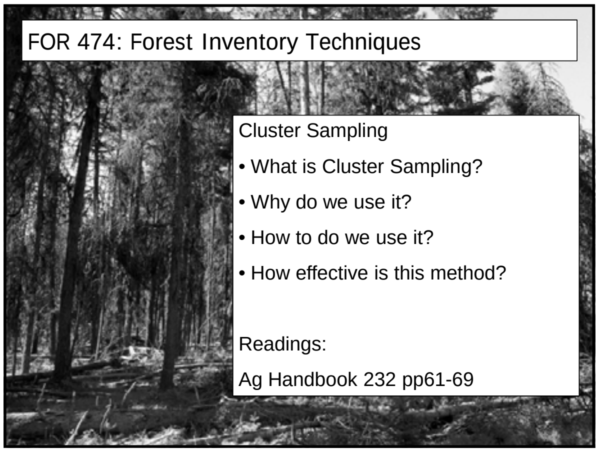
C_F = Cost Function = $(n_s{*}C_s)+(n_p{*}C_p)$

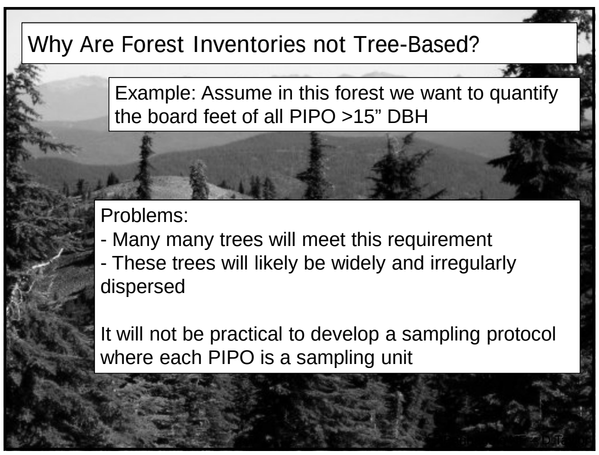
$s^2_{Y.X}$ = Estimated variance about the regression

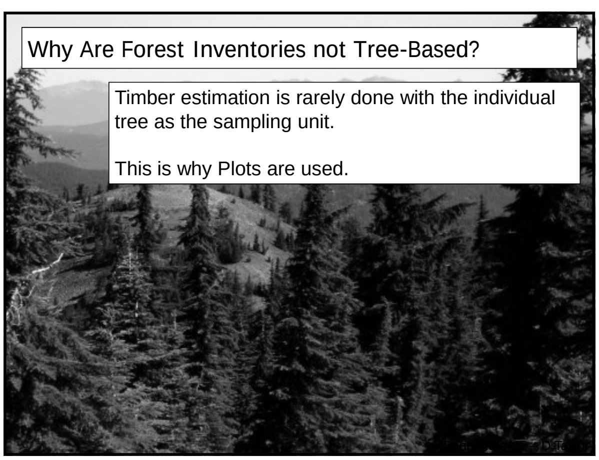
s^2_{XY} = Covariance

$$n_{p(needed)} = \frac{C_F\sqrt{\frac{c_s s^2_{XY}}{s^2_X}}}{c_s\sqrt{c_p\left(\frac{s^2_Y - s^2_{XY}}{s^2_X}\right) + c_p\left(\frac{c_s - s^2_{XY}}{s^2_X}\right)}}$$

$$n_{s(needed)} = \frac{C_F\sqrt{c_p s^2_{Y.X}}}{c_s\sqrt{c_p s^2_{Y.X}} + c_p\left(\frac{c_s s^2_{XY}}{s^2_X}\right)}$$


FOR 474: Forest Inventory Techniques
Cluster Sampling
• What is Cluster Sampling?
• Why do we use it?
• How to do we use it?
• How effective is this method?
Readings:
Ag Handbook 232 pp61-69


Why Are Forest Inventories not Tree-Based?
Example: Assume in this forest we want to quantify the board feet of all PIPO >15" DBH
Problems:
- Many many trees will meet this requirement
- These trees will likely be widely and irregularly dispersed
It will not be practical to develop a sampling protocol where each PIPO is a sampling unit


Why Are Forest Inventories not Tree-Based?
Timber estimation is rarely done with the individual tree as the sampling unit.
This is why Plots are used.

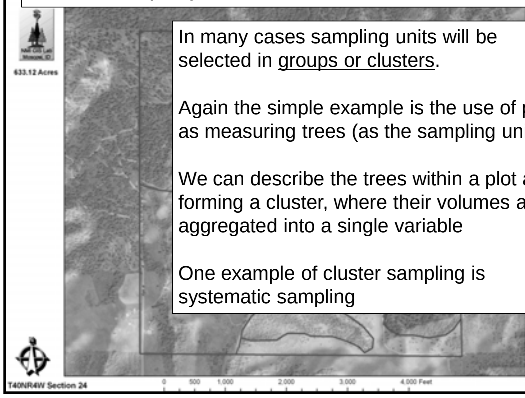
Cluster Sampling: What is it?

In many cases sampling units will be selected in groups or clusters.

Again the simple example is the use of plots, as measuring trees (as the sampling unit)

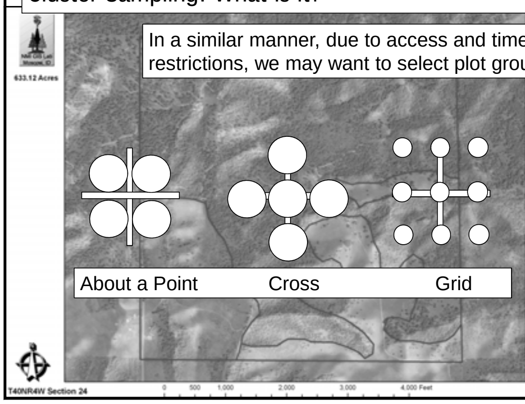
We can describe the trees within a plot as forming a cluster, where their volumes are aggregated into a single variable

One example of cluster sampling is systematic sampling



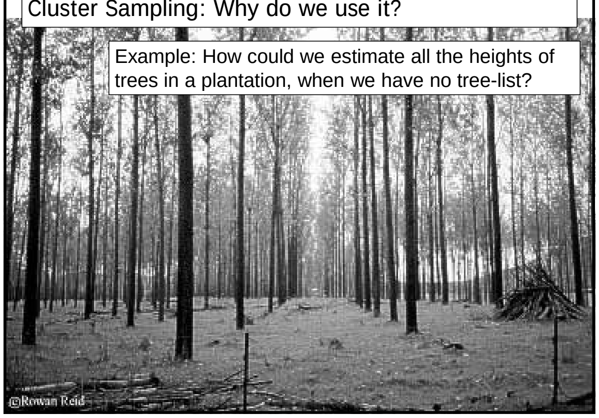
Cluster Sampling: What is it?

In a similar manner, due to access and time restrictions, we may want to select plot groups:

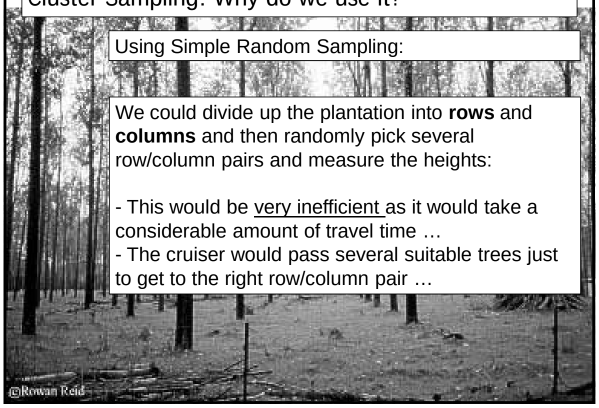


Cluster Sampling: Why do we use it?

Example: How could we estimate all the heights of trees in a plantation, when we have no tree-list?



Cluster Sampling: Why do we use it?

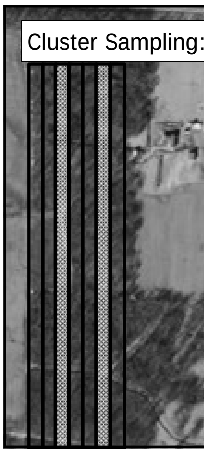


Using Simple Random Sampling:

We could divide up the plantation into **rows** and **columns** and then randomly pick several row/column pairs and measure the heights:

- This would be <u>very inefficient</u> as it would take a considerable amount of travel time …
- The cruiser would pass several suitable trees just to get to the right row/column pair …

Cluster Sampling: An Efficient Alternative?



In cluster sampling we instead measure all the trees in a row (or plot)

- We can select random sample of rows
- We measure all heights within those sample rows

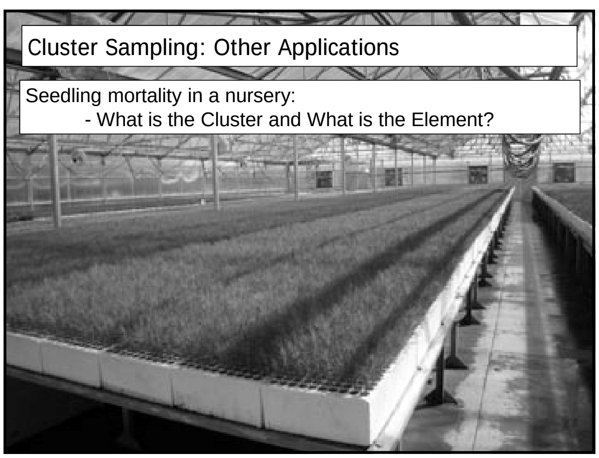
In this case:

trees = element
rows = cluster

Cluster sampling is more efficient than simple random sampling for plantations!

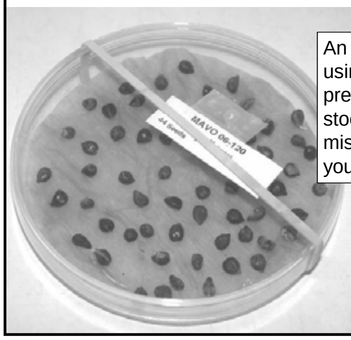
Cluster Sampling: Other Applications

Seedling mortality in a nursery:
- What is the Cluster and What is the Element?



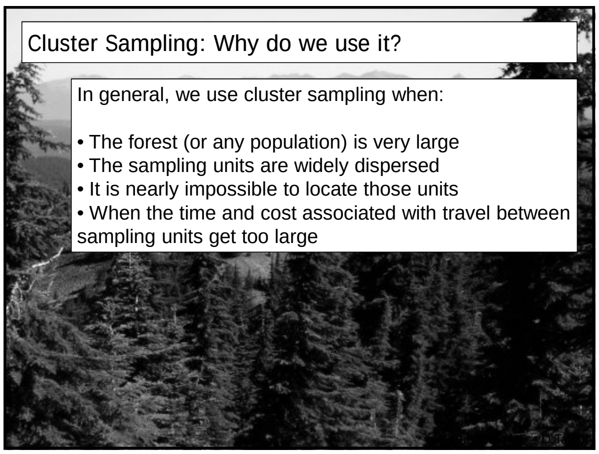
Cluster Sampling: Other Applications

Germination percentage of a batch of seed:



An additional advantage of using cluster sampling is to prevent you losing an entire stock of seed to disease or mistakes (i.e. not putting all your eggs in one basket)

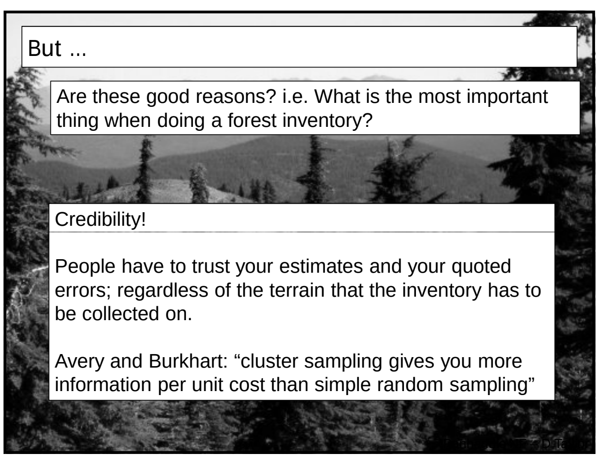
Cluster Sampling: Why do we use it?



In general, we use cluster sampling when:

- The forest (or any population) is very large
- The sampling units are widely dispersed
- It is nearly impossible to locate those units
- When the time and cost associated with travel between sampling units get too large

But ...



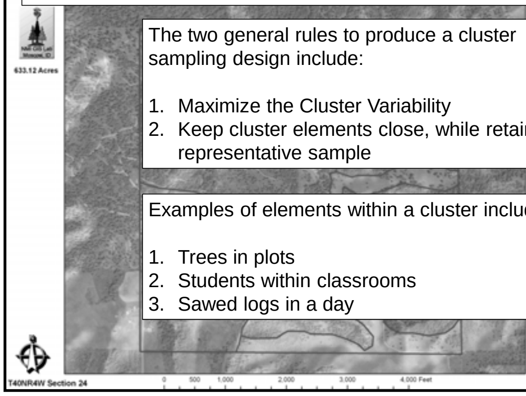
Are these good reasons? i.e. What is the most important thing when doing a forest inventory?

Credibility!

People have to trust your estimates and your quoted errors; regardless of the terrain that the inventory has to be collected on.

Avery and Burkhart: “cluster sampling gives you more information per unit cost than simple random sampling”

Cluster Sampling: How do we do it?



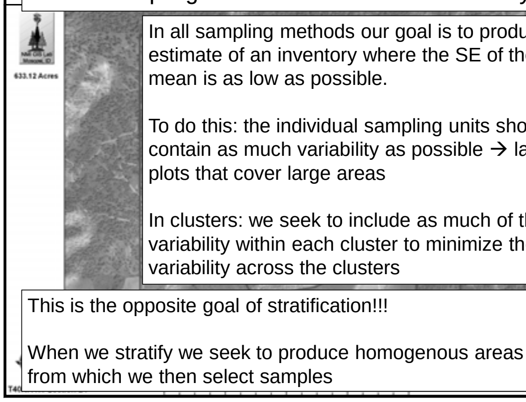
The two general rules to produce a cluster sampling design include:

1. Maximize the Cluster Variability
2. Keep cluster elements close, while retaining a representative sample

Examples of elements within a cluster include:

1. Trees in plots
2. Students within classrooms
3. Sawed logs in a day

Cluster Sampling: Maximize the Cluster Variability



In all sampling methods our goal is to produce an estimate of an inventory where the SE of the mean is as low as possible.

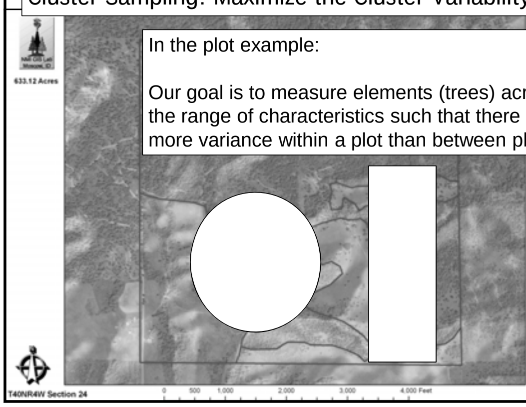
To do this: the individual sampling units should contain as much variability as possible → large plots that cover large areas

In clusters: we seek to include as much of the variability within each cluster to minimize the variability across the clusters

This is the opposite goal of stratification!!!

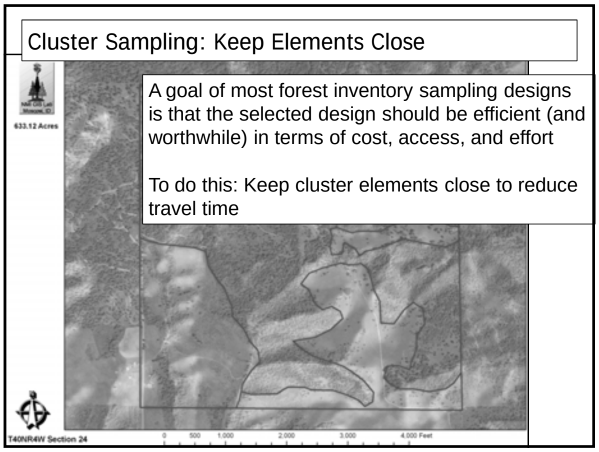
When we stratify we seek to produce homogenous areas from which we then select samples

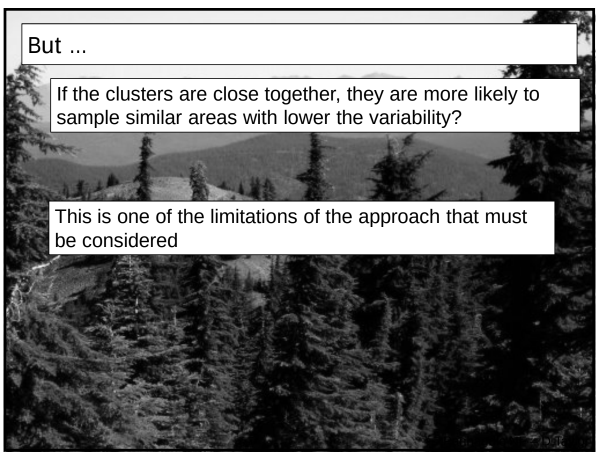
Cluster Sampling: Maximize the Cluster Variability

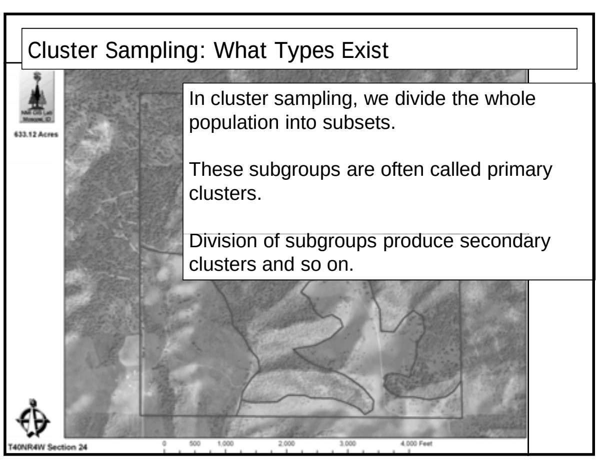


In the plot example:

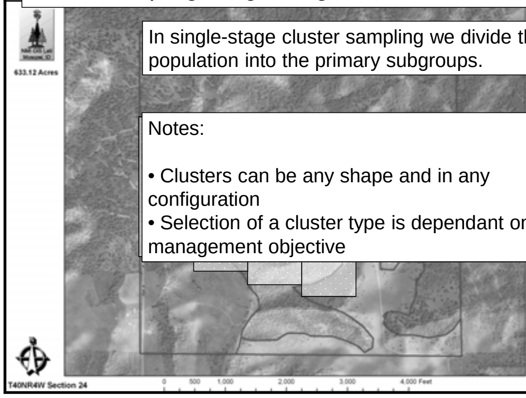
Our goal is to measure elements (trees) across the range of characteristics such that there is more variance within a plot than between plots.


Cluster Sampling: Keep Elements Close
A goal of most forest inventory sampling designs is that the selected design should be efficient (and worthwhile) in terms of cost, access, and effort
To do this: Keep cluster elements close to reduce travel time
633.12 Acres
T40NR4W Section 24
0 500 1,000 2,000 3,000 4,000 Feet


But ...
If the clusters are close together, they are more likely to sample similar areas with lower the variability?
This is one of the limitations of the approach that must be considered


Cluster Sampling: What Types Exist
In cluster sampling, we divide the whole population into subsets.
These subgroups are often called primary clusters.
Division of subgroups produce secondary clusters and so on.
633.12 Acres
T40NR4W Section 24
0 500 1,000 2,000 3,000 4,000 Feet

Cluster Sampling: Single-Stage

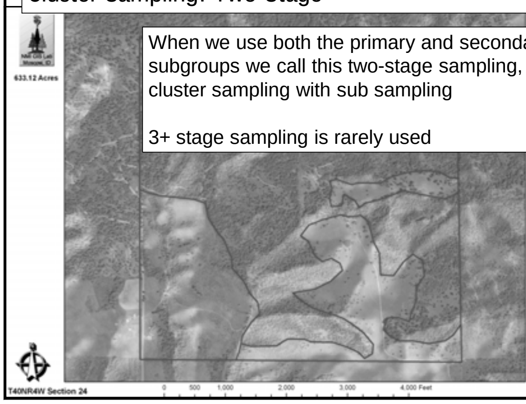


In single-stage cluster sampling we divide the population into the primary subgroups.

Notes:

• Clusters can be any shape and in any configuration
• Selection of a cluster type is dependant on the management objective

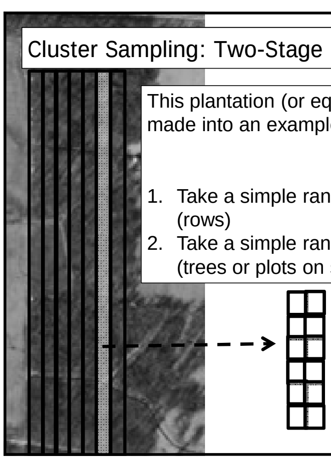
Cluster Sampling: Two-Stage



When we use both the primary and secondary subgroups we call this two-stage sampling, or cluster sampling with sub sampling

3+ stage sampling is rarely used

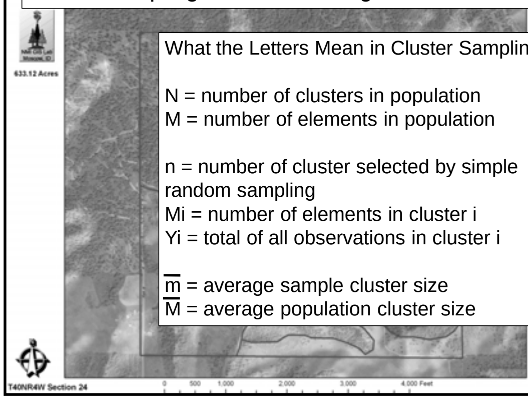
Cluster Sampling: Two-Stage



This plantation (or equally nursery) case can be made into an example of Two-Stage sampling:

1. Take a simple random sample of clusters (rows)
2. Take a simple random sample of elements (trees or plots on say a grid)

Cluster Sampling: More Confusing Letters



What the Letters Mean in Cluster Sampling

N = number of clusters in population
M = number of elements in population

n = number of cluster selected by simple random sampling
Mi = number of elements in cluster i
Yi = total of all observations in cluster i

$\overline{m}$ = average sample cluster size
$\overline{M}$ = average population cluster size

Cluster Sampling: Calculating the Mean

Example: Assess how many seedlings are alive (each cluster can have 100 possible alive)

Cluster	1	2	3	4	5	Total
Seedlings	68	79	82	91	84	404

$$\bar{p} = \frac{\sum_{i=1}^{n} p_i}{n}$$

p (mean mortality %) = Σ(seedlings) / #clusters
= 404 / 5
= 80.8 %

Cluster Sampling: Calculating the Variance

Example: Assess how many seedlings are alive (each cluster can have 100 possible alive)

Cluster	1	2	3	4	5	Total
Seedlings	68	79	82	91	84	404

$$s_p^2 = \frac{\sum_{i=1}^{n} p_i^2 - \frac{\left(\sum_{i=1}^{n} p_i\right)^2}{n}}{(n-1)}$$

$s^2 = [(68^2 + 79^2 + 82^2 + 91^2 + 84^2) - (404)^2/5] / 4$
$= 70.7$

Cluster Sampling: Calculating the Standard Error

Symbols: n = number of clusters sampled, N = total number of clusters possible

	Cluster	1	2	3	4	5	Total	
	Seedlings	68	79	82	91	84	**404**	

$$s_{\bar{y}} = \sqrt{\frac{s_y^2}{n}\left(1 - \frac{n}{N}\right)}$$

SE = √[s²/n ~~* (1-n/N)]~~ ignore finite population correction
= √[70.7/5]
= 3.76

FOR 474: Forest Inventory Techniques

Task: Fill in the Cluster Sampling spreadsheet