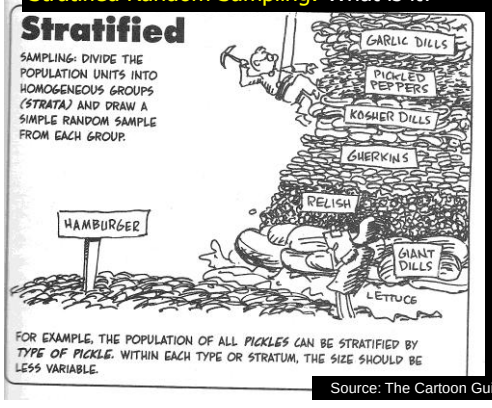
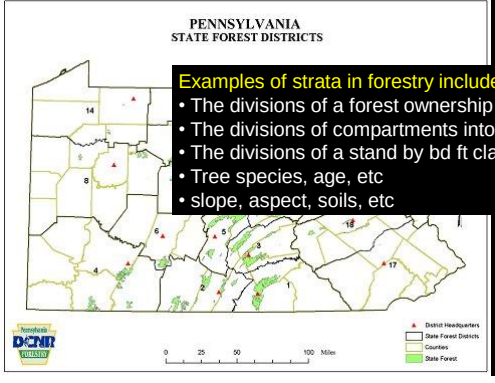
Stratified Random Sampling: What is it?

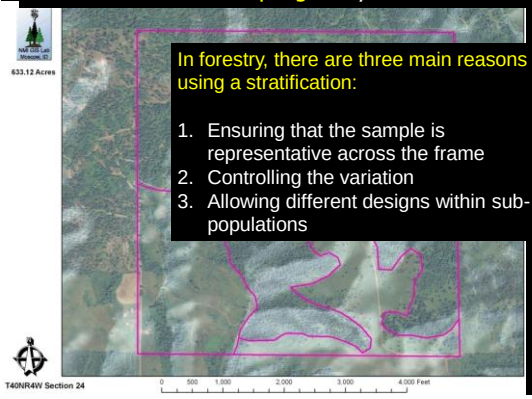


Source: The Cartoon Guide to Statistics

Stratified Random Sampling: What is it?



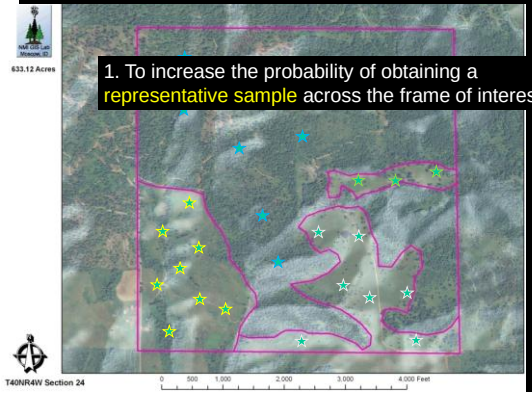
Stratified Random Sampling: Why do we use it?



In forestry, there are three main reasons for using a stratification:

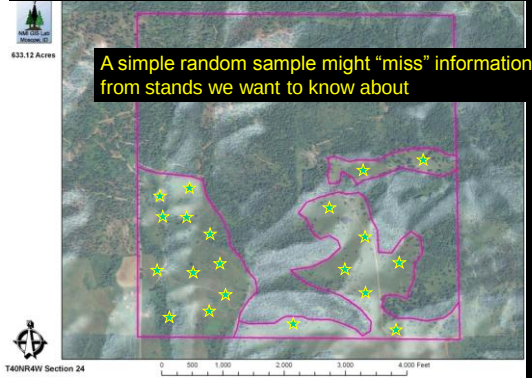
1. Ensuring that the sample is representative across the frame
2. Controlling the variation
3. Allowing different designs within sub-populations

Stratified Random Sampling: Why do we use it?



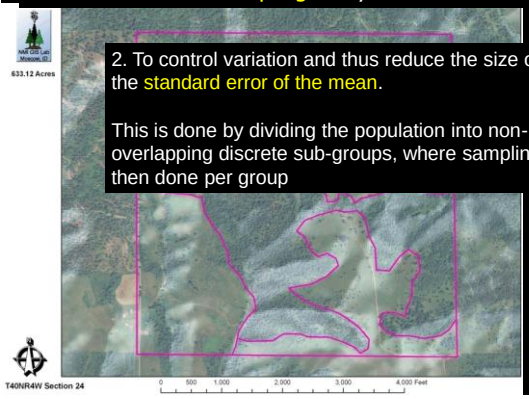
1. To increase the probability of obtaining a representative sample across the frame of interest

Reminder of a Frame: A construct that highlights the boundaries of a population – e.g., the edge of the management boundary



A simple random sample might "miss" information from stands we want to know about

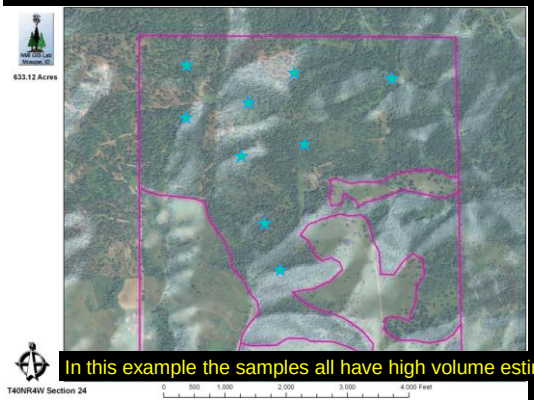
Stratified Random Sampling: Why do we use it?



2. To control variation and thus reduce the size of the standard error of the mean.

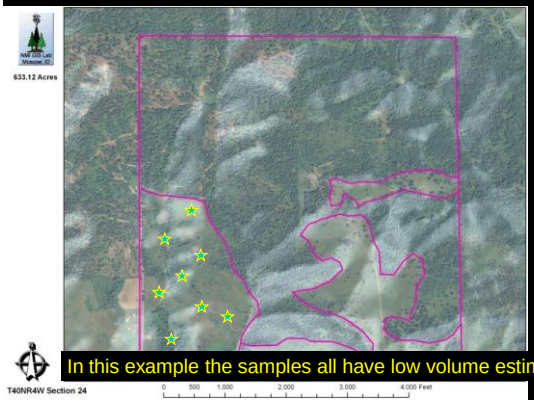
This is done by dividing the population into non-overlapping discrete sub-groups, where sampling is then done per group

In a simple random sample this might be the result:



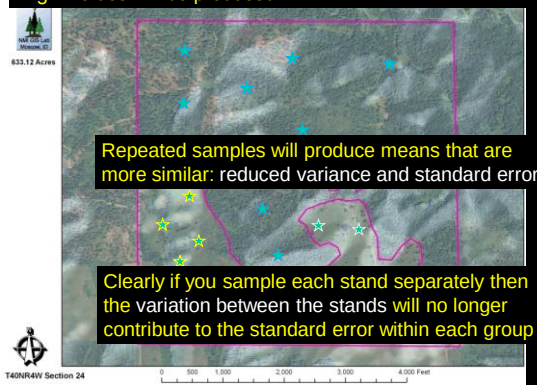
In this example the samples all have high volume estimates

Or this:

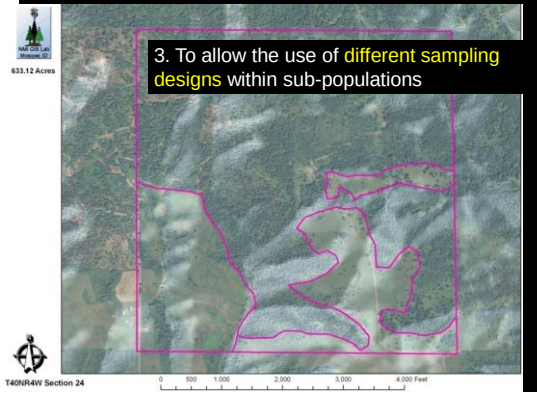


In this example the samples all have low volume estimates

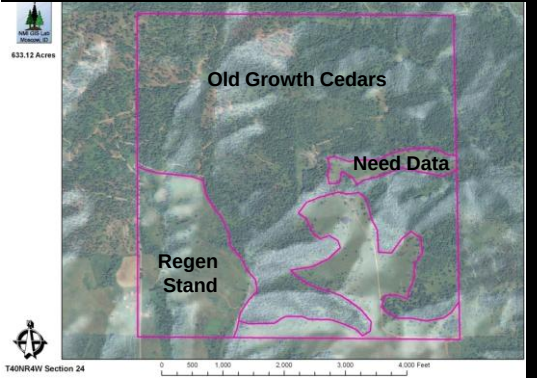
Sampling within each area makes it unlikely that only “low” or “high” values will be produced



Stratified Random Sampling: Why do we use it?



The management objective of each stand may vary and some stands may still need preliminary data



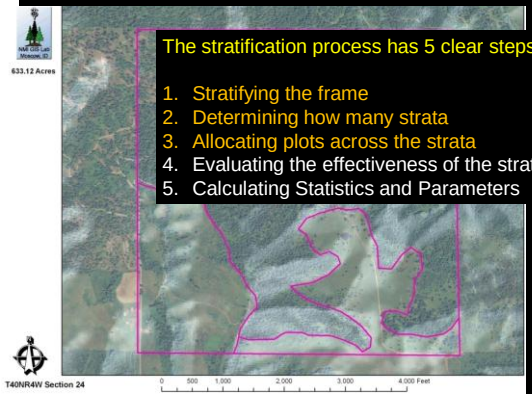
The management objective of each stand may vary and some stands may still need preliminary data

← Insect Damage

Seed Trees →

Image Source: USDA FS

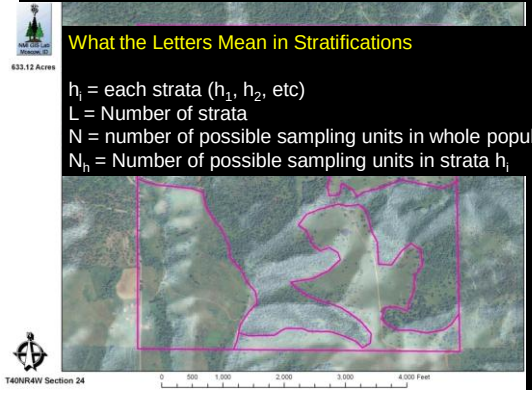
Stratified Random Sampling: How to do it?



The stratification process has 5 clear steps:

1. Stratifying the frame
2. Determining how many strata
3. Allocating plots across the strata
4. Evaluating the effectiveness of the strata
5. Calculating Statistics and Parameters

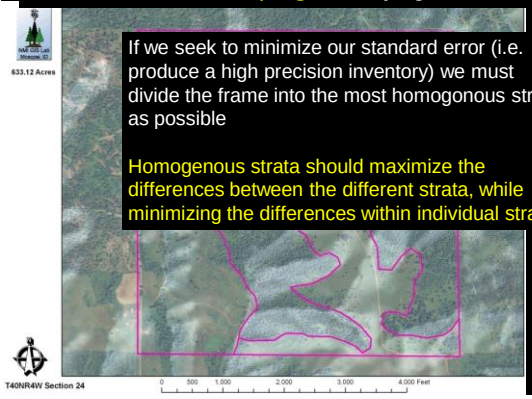
Stratified Random Sampling: How to do it?



What the Letters Mean in Stratifications

h_i = each strata (h_1, h_2, etc)
L = Number of strata
N = number of possible sampling units in whole population
N_h = Number of possible sampling units in strata h_i

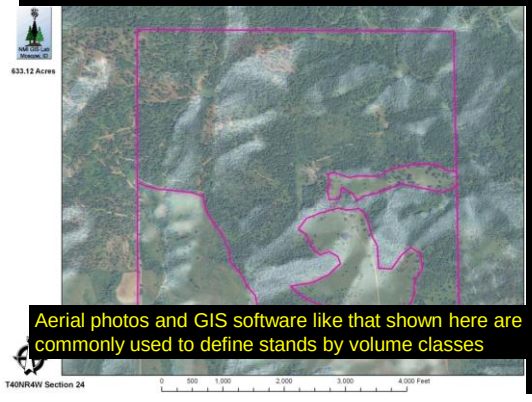
Stratified Random Sampling: Stratifying our frame



If we seek to minimize our standard error (i.e. produce a high precision inventory) we must divide the frame into the most homogonous strata as possible

Homogenous strata should maximize the differences between the different strata, while minimizing the differences within individual strata

Stratified Random Sampling: Stratifying our frame



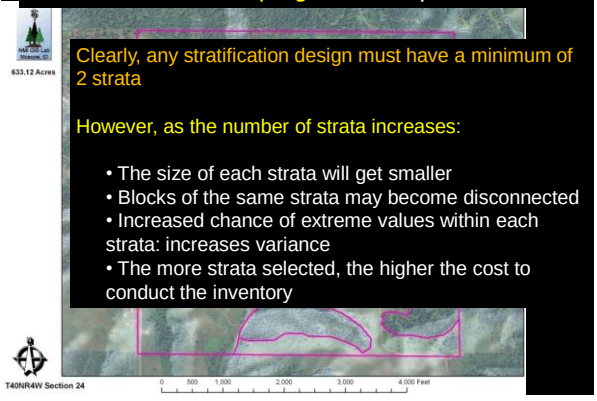
Aerial photos and GIS software like that shown here are commonly used to define stands by volume classes

Stratified Random Sampling: Stratifying our frame

Common stratification variables include: LAI, topography (slope, aspect, elevation), density, etc

Image Source: Purdue University

Stratified Random Sampling: How Many Strata?

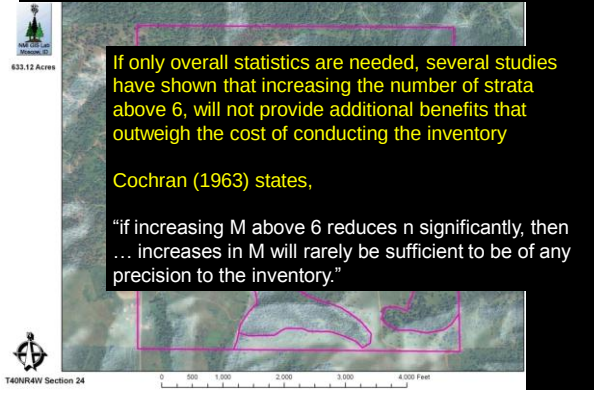


Clearly, any stratification design must have a minimum of 2 strata

However, as the number of strata increases:

- The size of each strata will get smaller
- Blocks of the same strata may become disconnected
- Increased chance of extreme values within each strata: increases variance
- The more strata selected, the higher the cost to conduct the inventory

Stratified Random Sampling: How Many Strata?

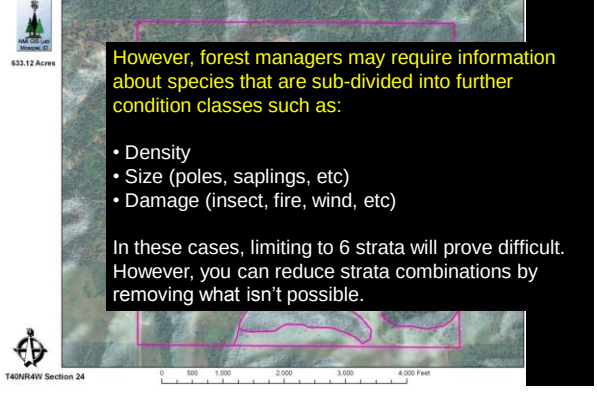


If only overall statistics are needed, several studies have shown that increasing the number of strata above 6, will not provide additional benefits that outweigh the cost of conducting the inventory

Cochran (1963) states,

"if increasing M above 6 reduces n significantly, then … increases in M will rarely be sufficient to be of any precision to the inventory."

Stratified Random Sampling: How Many Strata?

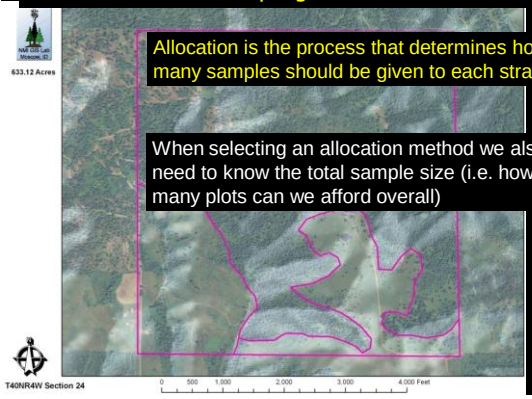


However, forest managers may require information about species that are sub-divided into further condition classes such as:

- Density
- Size (poles, saplings, etc)
- Damage (insect, fire, wind, etc)

In these cases, limiting to 6 strata will prove difficult. However, you can reduce strata combinations by removing what isn't possible.

Stratified Random Sampling: Allocation



Allocation is the process that determines how many samples should be given to each strata

When selecting an allocation method we also need to know the total sample size (i.e. how many plots can we afford overall)

Stratified Random Sampling: Allocation

Volume class
I
II
III
IV
V
Total

In equal allocation the sample size is the same in each of the strata

For a sample size of 100, equal allocation would give each class a sample size of 20 regardless of its area.

Avery and Burkhart Chapter 3

Stratified Random Sampling: Allocation

Volume class	Stratum area (acres)
I	15
II	45
III	110
IV	60
V	70
Total	300

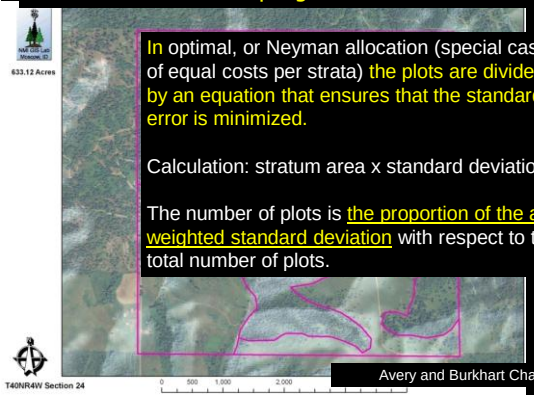
In proportional allocation the number of plots per strata is proportional to the area of the strata, making the sampling intensity constant over all strata.

In proportional allocation each class receives as many samples as its area is in % to the total area.

If we had 100 samples, Class 1 would get 15/300 * 100 = 5

Avery and Burkhart Chapter 3

Stratified Random Sampling: Allocation



In optimal, or Neyman allocation (special case of equal costs per strata) the plots are divided by an equation that ensures that the standard error is minimized.

Calculation: stratum area x standard deviation

The number of plots is the proportion of the area weighted standard deviation with respect to the total number of plots.

Avery and Burkhart Chapter 3

Stratified Random Sampling: Allocation

Volume class	Stratum area (acres)	Std. dev. (cords/acre)	Area × std. dev.
I	15	20	300
II	45	70	3,150
III	110	35	3,850
IV	60	45	2,700
V	70	25	1,750
Total	300	···	11,750

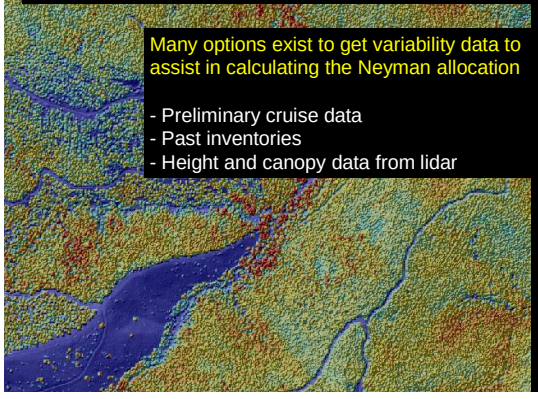
Using Neyman allocation class 1 only gets (out of 100):

300/11750 * 100 = 3 plots.

$$\frac{A_{h_1} * s_{h_1}}{\sum_{h_i}^{n}(A_{h_1} * s_{h_1})} * n = n_h$$

Avery and Burkhart Chapter 3

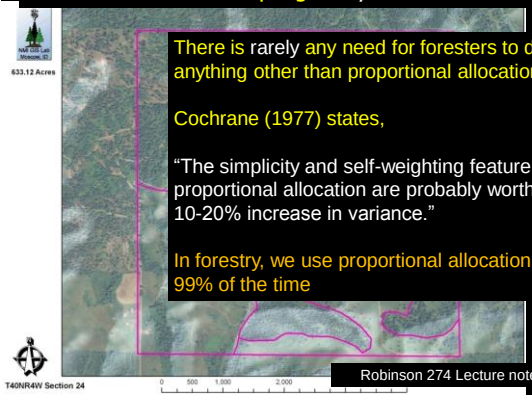
Stratified Random Sampling: Allocation



Many options exist to get variability data to assist in calculating the Neyman allocation

- Preliminary cruise data
- Past inventories
- Height and canopy data from lidar

Stratified Random Sampling: Why % Allocation?



There is rarely any need for foresters to do anything other than proportional allocation

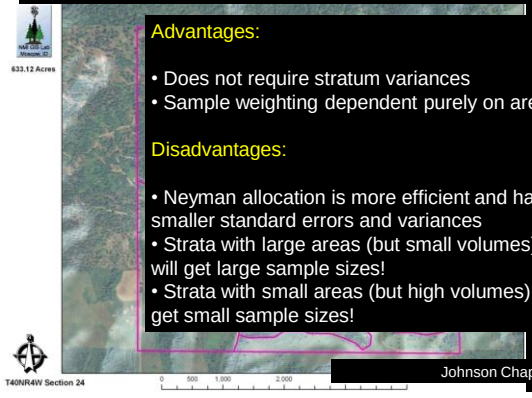
Cochrane (1977) states,

"The simplicity and self-weighting feature of proportional allocation are probably worth a 10-20% increase in variance."

In forestry, we use proportional allocation 99% of the time

Robinson 274 Lecture notes p57

Stratified Random Sampling: Why % Allocation?



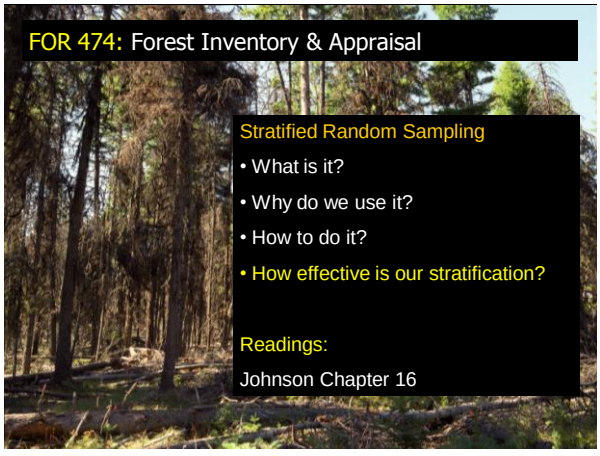
Advantages:

- Does not require stratum variances
- Sample weighting dependent purely on area

Disadvantages:

- Neyman allocation is more efficient and has smaller standard errors and variances
- Strata with large areas (but small volumes) will get large sample sizes!
- Strata with small areas (but high volumes) will get small sample sizes!

Johnson Chapter 16



FOR 474: Forest Inventory & Appraisal

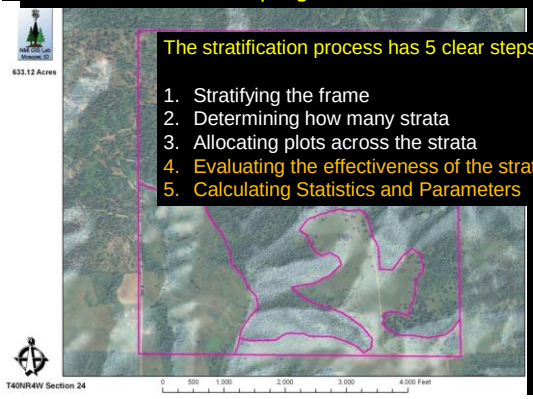
Stratified Random Sampling

- What is it?
- Why do we use it?
- How to do it?
- How effective is our stratification?

Readings:

Johnson Chapter 16

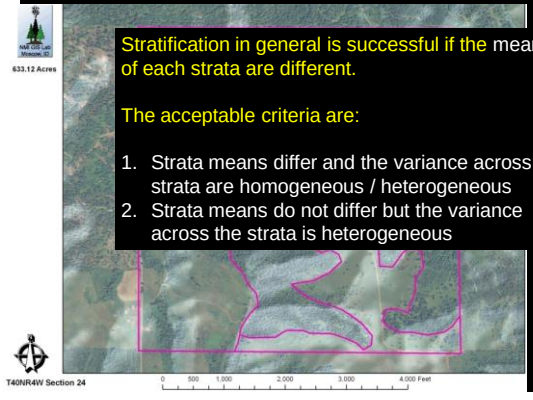
Stratified Random Sampling: How to do it



The stratification process has 5 clear steps:

1. Stratifying the frame
2. Determining how many strata
3. Allocating plots across the strata
4. Evaluating the effectiveness of the strata
5. Calculating Statistics and Parameters

Stratified Random Sampling: Effectiveness



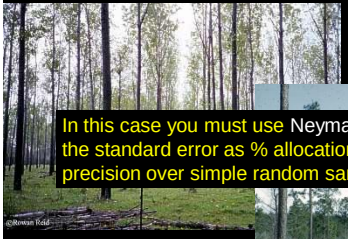
Stratification in general is successful if the means of each strata are different.

The acceptable criteria are:

1. Strata means differ and the variance across the strata are homogeneous / heterogeneous
2. Strata means do not differ but the variance across the strata is heterogeneous

Stratified Random Sampling: Effectiveness

Strata means do not differ and the variance across the strata is heterogeneous

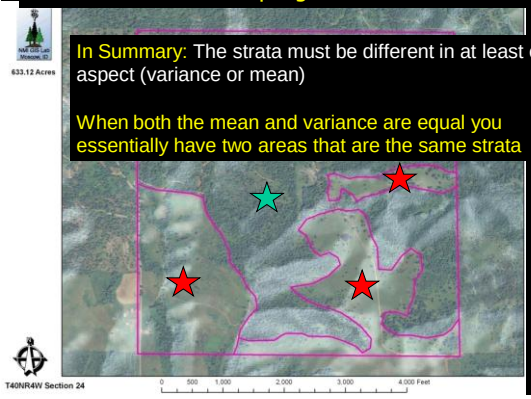


← Low Variance

In this case you must use Neyman Allocation to reduce the standard error as % allocation will not increase the precision over simple random sampling

High Variance →

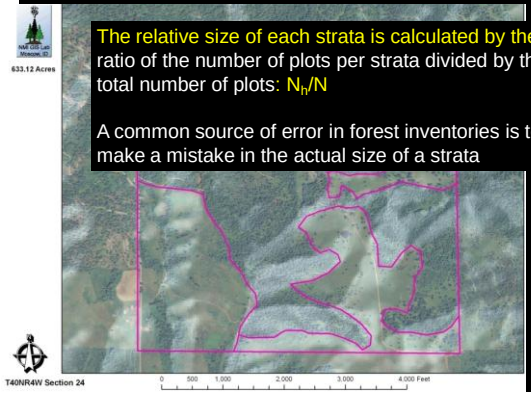
Stratified Random Sampling: Effectiveness



In Summary: The strata must be different in at least one aspect (variance or mean)

When both the mean and variance are equal you essentially have two areas that are the same strata

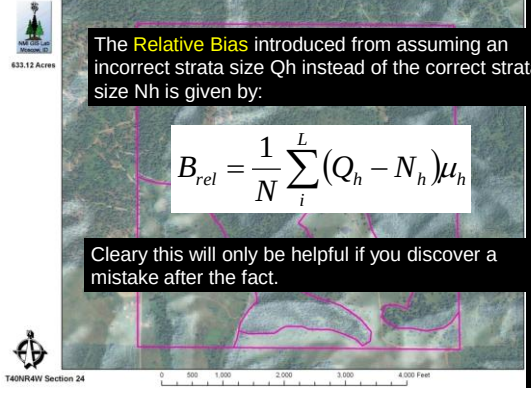
Stratified Random Sampling: Strata Size



The relative size of each strata is calculated by the ratio of the number of plots per strata divided by the total number of plots: N_h/N

A common source of error in forest inventories is to make a mistake in the actual size of a strata

Stratified Random Sampling: Strata Size Error

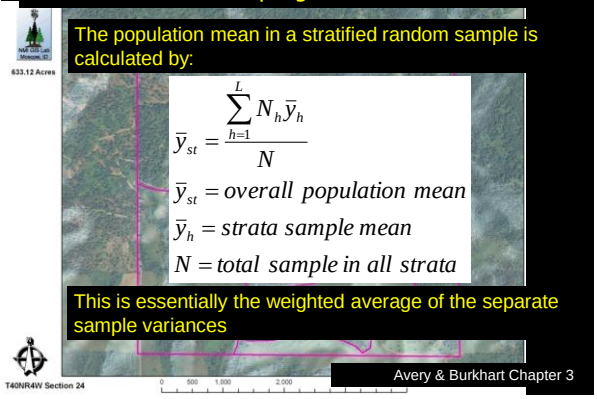


The Relative Bias introduced from assuming an incorrect strata size Qh instead of the correct strata size Nh is given by:

$$B_{rel} = \frac{1}{N}\sum_{i}^{L}(Q_h - N_h)\mu_h$$

Cleary this will only be helpful if you discover a mistake after the fact.

Stratified Random Sampling: Parameters & Statistics



The population mean in a stratified random sample is calculated by:

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N}$$

$\bar{y}_{st} = overall\ population\ mean$

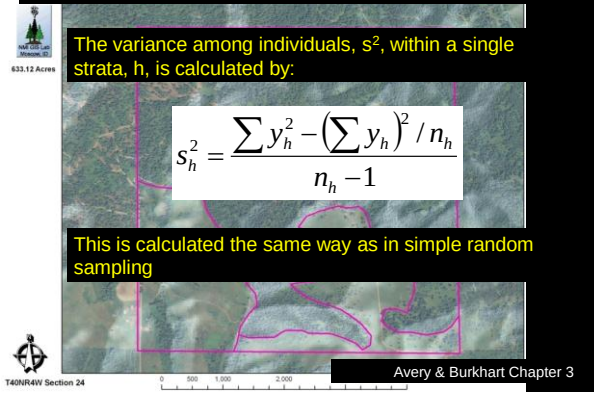
$\bar{y}_h = strata\ sample\ mean$

$N = total\ sample\ in\ all\ strata$

This is essentially the weighted average of the separate sample variances

Avery & Burkhart Chapter 3

Stratified Random Sampling: Parameters & Statistics



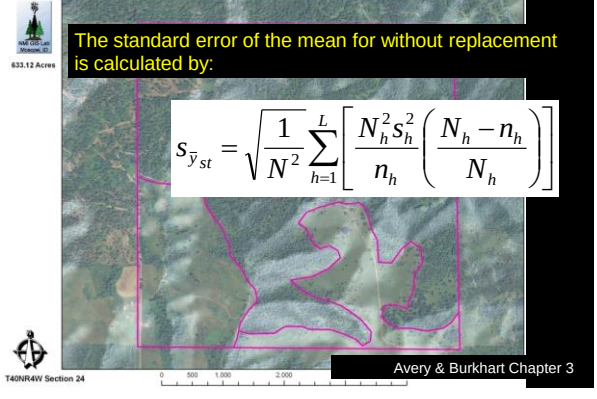
The variance among individuals, s^2, within a single strata, h, is calculated by:

$$s_h^2 = \frac{\sum y_h^2 - \left(\sum y_h\right)^2 / n_h}{n_h - 1}$$

This is calculated the same way as in simple random sampling

Avery & Burkhart Chapter 3

Stratified Random Sampling: Parameters & Statistics

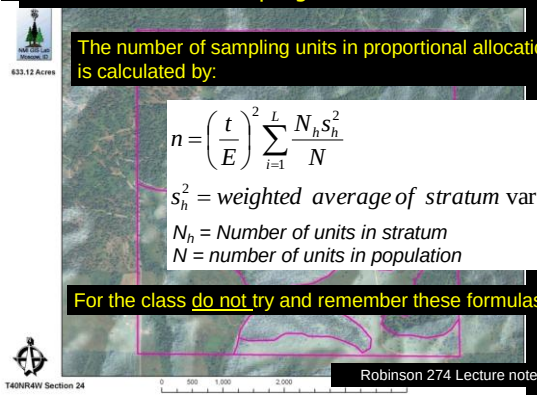


The standard error of the mean for without replacement is calculated by:

$$s_{\bar{y}\,st} = \sqrt{\frac{1}{N^2} \sum_{h=1}^{L} \left[\frac{N_h^2 s_h^2}{n_h} \left(\frac{N_h - n_h}{N_h} \right) \right]}$$

Avery & Burkhart Chapter 3

Stratified Random Sampling: Parameters & Statistics



The number of sampling units in proportional allocation is calculated by:

$$n = \left(\frac{t}{E}\right)^2 \sum_{i=1}^{L} \frac{N_h s_h^2}{N}$$

$$s_h^2 = weighted\ \ average\ of\ \ stratum\ \text{var}$$

N_h = *Number of units in stratum*
N = *number of units in population*

For the class do not try and remember these formulas!

Robinson 274 Lecture notes p57