LECTURES FOR WEEK 11

Stratified Random Sampling:

So far, we have discussed random and systematic location of cruise plots in the context of simple random sampling. These methods work well when the area sampled is homogeneous. Foresters, however, often cruise areas with different forest cover types (i.e., stands). These different stands make the area heterogeneous in terms of cover types. In this case, stratified random sampling can be used to calculate a more precise cruise.

In stratified random sampling, the units of the population (e.g., stands) are grouped together based on similar criterion (e.g., overstory tree type). Each unit (stratum) or stand is then cruised and the stratum estimates are combined to give an estimate for the entire area. We will illustrate this methodology with an example.

EXAMPLE: You cruise a 500 acre tract using variable radius plots and determine the following cubic foot volumes per acre for cruise points in three timber types:

I. Black Cherry – Maple (SAF Cover Type 28)

<table>
<thead>
<tr>
<th>PLOT VOLUMES (cubic foot volume per acre)</th>
</tr>
</thead>
<tbody>
<tr>
<td>570       640       480       560       510</td>
</tr>
<tr>
<td>590       670       600       780       700</td>
</tr>
</tbody>
</table>
II. Yellow-Poplar – White Oak – Northern Red Oak (SAF Cover Type 59)

<table>
<thead>
<tr>
<th>PLOT VOLUMES (cubic foot volume per acre)</th>
</tr>
</thead>
<tbody>
<tr>
<td>520</td>
</tr>
<tr>
<td>760</td>
</tr>
</tbody>
</table>

III. White Oak – Black Oak – Northern Red Oak (SAF Cover Type 52)

<table>
<thead>
<tr>
<th>PLOT VOLUMES (cubic foot volume per acre)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
</tr>
<tr>
<td>540</td>
</tr>
</tbody>
</table>

**Mean:** First, calculate the mean cubic foot volume per acre (CFV/A) for each stratum using the simple random procedure we used earlier:

\[
\overline{X}_I = \frac{6100}{10} = 610 \text{ cfv/a}
\]

\[
\overline{X}_{II} = \frac{7370}{10} = 737 \text{ cfv/a}
\]

\[
\overline{X}_{III} = \frac{3040}{10} = 304 \text{ cfv/a}
\]

The mean of the stratified sample can now be computed by:

\[
\overline{X}_{ST} = \frac{\sum_{h=1}^{L} N_h \times \overline{X}_h}{N},
\]

where: 

\( L \) = The number of strata

\( N_h \) = The size of stratum \( h \) (\( h = 1, 2, \ldots, L \)) in acres.

\( N = \) The total size of the tract in acres (\( N = \sum_{h=1}^{L} N_h \)).
If the strata sizes are:

\[ \begin{align*}
N_I &= 250 \text{ acres} \\
N_{II} &= 100 \text{ acres} \\
N_{III} &= 150 \text{ acres}
\end{align*} \]

Total Acres: \( N = 500 \) acres.

Then the stratified sample mean cubic volume per acre is:

\[
\overline{X}_{st} = \frac{250 \times 610 + 100 \times 737 + 150 \times 304}{500} = 543.6 \text{ cfv/a}
\]

If you used the simple random sampling formula to calculate the mean, you would get:

\[
\overline{X}_S = \frac{16,510}{30} = 550.3 \text{ cfv/a}
\]

which is close to the stratified mean.

**Standard Errors**: Next, calculate the variance \( (s^2) \) of cubic foot volume per acre (CFV/A) for each stratum using the simple random methodology we used earlier:

\[ s_I^2 = 8111.1 \text{ (cfv/a)}^2 \]
\[ s_{II}^2 = 15,556.7 \text{ (cfv/a)}^2 \]
\[ s_{III}^2 = 12,204.4 \text{ (cfv/a)}^2 \]

With these variances, calculate the stratified standard error of cubic foot volume per acre for the sample by:

\[
SE_{st} = \sqrt{\sum_{h=1}^{L} \left( w_h^2 \times \frac{s_h^2}{n_h} \right)},
\]

where \( n_h \) = number of cruise plots

\( w_h = \) weight factor = \( N_h / N \).
So, the stratified standard error for this example is:

\[ SE_{SR} = \sqrt{\left(\frac{250}{500}\right)^2 \frac{1000.1}{10} + \left(\frac{100}{500}\right)^2 \frac{15,567.7}{10} + \left(\frac{150}{500}\right)^2 \frac{12,204.4}{10}} = 19.4 \text{ cfv/a}, \]

If you used the simple random sampling formula to calculate the standard error, you would get:

\[ SE_S = 38.9 \text{ cfv/a}, \]

which is greater than the stratified standard error. Thus, stratifying the sample will improve our overall cruise precision in this example (why?).

**95% Confidence Interval:** Calculate the 95% confidence interval for cubic foot volume per acre:

\[ \overline{X}_{ST} \pm t_{0.05, n-1 = 29} * SE_{ST} \Rightarrow 543.6 \pm 2.045*19.4 \text{ or } 543.6 \pm 39.7 \text{ cfv/a} \]

**Cruise Precision:** Calculate the cruise precision for this stratified cruise:

\[ \text{Precision} = \frac{SE_{ST} * t_{0.05, 29}}{\overline{X}_{ST}} = \frac{19.4 * 2.045}{543.6} * 100 \geq 7\% \]

\[ \text{Precision (SRS)} \geq 15\% \]

**Sample Size:** You can calculate the number of plots needed in each stratum that are necessary to achieve a specified statistical objective. First, determine the total number of plots necessary by:

\[ n = \frac{t^2 \left( \sum_{h=1}^{1} w_h * s_{xh} \right)^2}{E^2}, \]
where: \( E = \) allowable error in absolute units
all other variables defined as before.

So, for our example, we want to find the number of plots needed to be 95% confident that we are within plus or minus 10% of the true cubic foot volume per acre. In absolute units, \( E = 543.6 \text{ cfv/a} \times (0.10) = 54.4 \text{ cfv/a} \).

\[
\begin{align*}
n \approx & \frac{2^2 \left( \frac{250}{500} \sqrt{8111.1} + \frac{100}{500} \sqrt{15,556.7} + \frac{150}{500} \sqrt{12,204.4} \right)^2}{54.4^2} \\
& \approx 15 \text{ plots}
\end{align*}
\]

For our example, we actually collected more sample points than necessary to achieve this statistical objective.

Continuing with our example, we can now allocate (i.e., optimum allocation) these 15 plots to the three strata with the formula:

\[
n_h = \frac{w_h \cdot s_{sh} \cdot n}{\sum_{h=1}^{s} w_h \cdot s_{sh}}
\]

So,

\[
\begin{align*}
n_1 = & \frac{\left( \frac{250}{500} \right) \sqrt{8111.1}}{\sum \left( \frac{250}{500} \right) \sqrt{8111.1} + \left( \frac{100}{500} \right) \sqrt{15,556.7} + \left( \frac{150}{500} \right) \sqrt{12,204.4}} \times 15 = 45.03 \times 15 \approx 7 \text{ plots}
\end{align*}
\]

\[
\begin{align*}
n_2 = & \frac{\left( \frac{100}{500} \right) \sqrt{15,556.7}}{103.12} \times 15 \approx 4 \text{ plots}
\end{align*}
\]
You will notice that the total number of plots equals 16 if you add up the strata plot numbers. This number is greater than 15 because you should always round up the result when calculating sample size.

You should also know that if you had an estimate of variability and you defined your strata before the cruise, you can calculate your sample size before the cruise and allocate the plots to each stratum with the same methodology.

\[
 n_{\text{III}} = \left( \frac{150}{500} \right) \sqrt{\frac{12204.4}{103.12}} \times 15 \geq 5 \text{ plots}
\]