Analysis of Stratified Surveys
Stratification

• Why stratify?

• Stratification by:
  • Geographic area
  • Survey
  • Species / cluster size

• Limitations of Distance
Stratification is used to:
• reduce variance and improve precision
• and for producing estimates in regions of interest

Stratify by:
• AREA or GEOGRAPHIC REGION
  - the study region is partitioned into smaller regions
• SURVEY
  - used when different surveys cover the same geographic area
• POPULATION/SPECIES/CLUSTER SIZE
  - same geographic region with different ‘sub-stocks’ in it
Area/Geographic stratification

Estimate density in each sub-region

\[ \hat{D}_1, \hat{D}_2, \hat{D}_3 \]

Abundance in each sub-region is given by

\[ \hat{N}_1 = A_1 \hat{D}_1 \]
\[ \hat{N}_2 = A_2 \hat{D}_2 \]
\[ \hat{N}_3 = A_3 \hat{D}_3 \]

Total size of study region

\[ A = A_1 + A_2 + A_3 \]
Total abundance is

\[ \hat{N} = \hat{N}_1 + \hat{N}_2 + \hat{N}_3 \]

\[ = A_1 \hat{D}_1 + A_2 \hat{D}_2 + A_3 \hat{D}_3 \]

Overall (Global in Distance) density is

\[ \hat{D} = \frac{\hat{N}}{A} = \frac{A_1 \hat{D}_1 + A_2 \hat{D}_2 + A_3 \hat{D}_3}{A_1 + A_2 + A_3} \]

\[ = \left( \frac{A_1}{A} \right) \hat{D}_1 + \left( \frac{A_2}{A} \right) \hat{D}_2 + \left( \frac{A_3}{A} \right) \hat{D}_3 \]

\[ = \sum_{i=1}^{3} \left( \frac{A_i}{A} \right) \hat{D}_i \]

Note form of equation
Example: SCANS II (2005)
SCANS II survey effort
Example of stratified data
Example: Full geographic stratification

\[ D = f_1(0) \times f_2(0) \times f_3(0) \]
\[ = E_1(s) \times E_2(s) \times E_3(s) \]
\[ = (n/L)_1 \times (n/L)_2 \times (n/L)_3 \]

Model Definition Properties: [Full Stratification]

- Analysis Engine: CDS - Conventional distance sampling
- Estimate:
  - Detection function
  - Cluster size
  - Multipliers
  - Variance
  - Misc.
- Stratum definition:
  - No stratification
  - Use layer type: Stratum
  - Post-stratify, using: Stratum, Area
- Sample definition (for encounter rate):
  - Use layer type: Sample
- Quantities to estimate and level of resolution:
  - Level of resolution of estimates:
    - Global
    - Stratum
    - Sample
  - Global density estimate is:
    - Mean
    - of stratum estimates
    - weighted by Stratum area

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews
Example: $f(0)$ pooled

\[ D \]
\[ f_{123}(0) \]
\[ E_1(s) \]
\[ E_2(s) \]
\[ E_3(s) \]
\[ (n/L)_1 \]
\[ (n/L)_2 \]
\[ (n/L)_3 \]
Pooled vs Stratified $f(0)$

Pooled $n=88$

Stratified

Ideal Habitat $n=39$

Marginal Habitat $n=49$
It is a Model Selection Problem

<table>
<thead>
<tr>
<th></th>
<th>Pooled</th>
<th>Stratum 1</th>
<th>Stratum 2</th>
<th>Stratum Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood $\log_e(L)$</td>
<td>-180.490</td>
<td>-72.699</td>
<td>-104.676</td>
<td>-177.375</td>
</tr>
<tr>
<td>No. parameters (q)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>AIC</td>
<td>364.980</td>
<td>149.398</td>
<td>213.352</td>
<td>362.75</td>
</tr>
</tbody>
</table>

Criterion for stratification of $f(0)$:
Fit separate $f(0)$ for each strata if

$$AIC_{pooled} > \sum AIC_{stratum}$$
Non-geographic stratification -- Stratification by survey

Let $L_i$ be effort for survey $i$

Global density is given by

$$
\hat{D} = \left( \frac{L_1}{L_1 + L_2} \right) \hat{D}_1 + \left( \frac{L_2}{L_1 + L_2} \right) \hat{D}_2
$$

$$
= \sum_{i=1}^{2} \left( \frac{L_i}{L} \right) \hat{D}_i
$$

This is the same form as before, but weighting factor now depends on effort.
Stratification by survey

Global density estimate is the mean of stratum estimates weighted by total effort in stratum.
Stratification by species

\[ \hat{D} = \hat{D}_{sp1} + \hat{D}_{sp2} \]

Species 1

Species 2
Limitations in Distance

- Distance cannot currently do multilevel stratification in one run

- Two runs are necessary
  - Estimate $f(0)$, $E[s]$ and $n/L$ by stratum
  - Combine strata 1 and 2 to estimate $f_{12}(0)$

- Care must be taken when calculating CVs because the density estimates for stratum 1 and 2 have an estimated $f(0)$ in common
Alternatives to stratification in Distance

- Small sample sizes can lead to low precision in stratum-specific estimates
- An alternative approach to reducing bias due to heterogeneity is Multiple Covariates Distance Sampling (MCDS)
  - Covariates, other than distance, are incorporated into the scale parameter of the detection function
  - MCDS can be used to fit the detection function at multiple levels e.g. stratum-specific density estimates can be obtained even if you don't have enough data to fit separate detection functions for each stratum
  - MCDS methods are covered in an upcoming lecture.