Analysis of Stratified Surveys

Section 3.7 of Buckland et al. (2001)

Section 2.3 of Buckland et al. (2015)

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

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

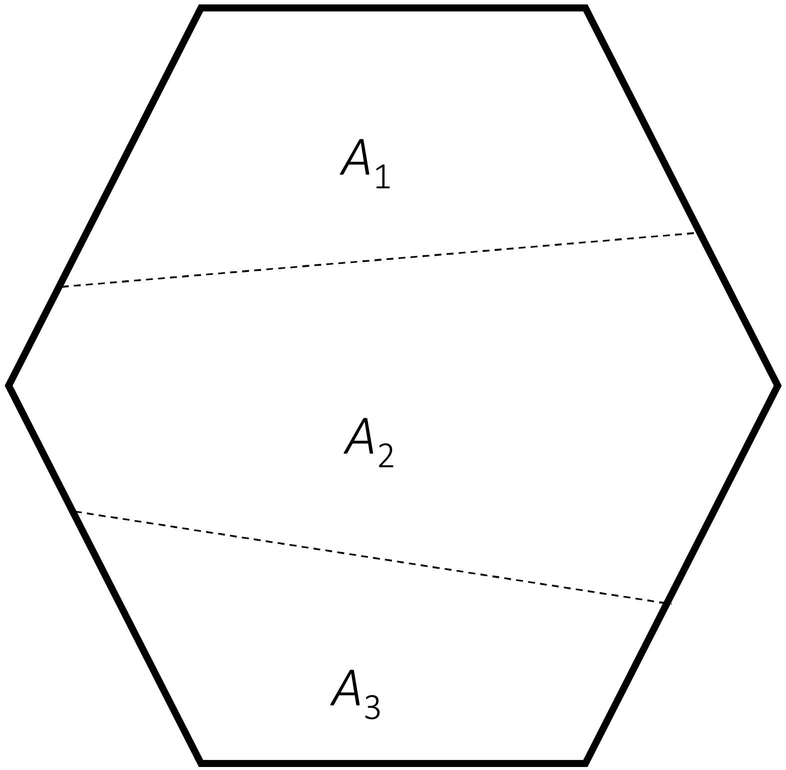
Stratify by:

- 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



Total size of study region

$$A = A_1 + A_2 + A_3$$

Estimate density in each sub-region

$$\hat{D}_1, \hat{D}_2, \hat{D}_3$$

Abundance in each sub-region is given by

$$\hat{N}_1 = A_1 \hat{D}_1$$

$$\hat{N}_2 = A_2 \hat{D}_2$$

$$\hat{N}_3 = A_3 \hat{D}_3$$

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

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) density is

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

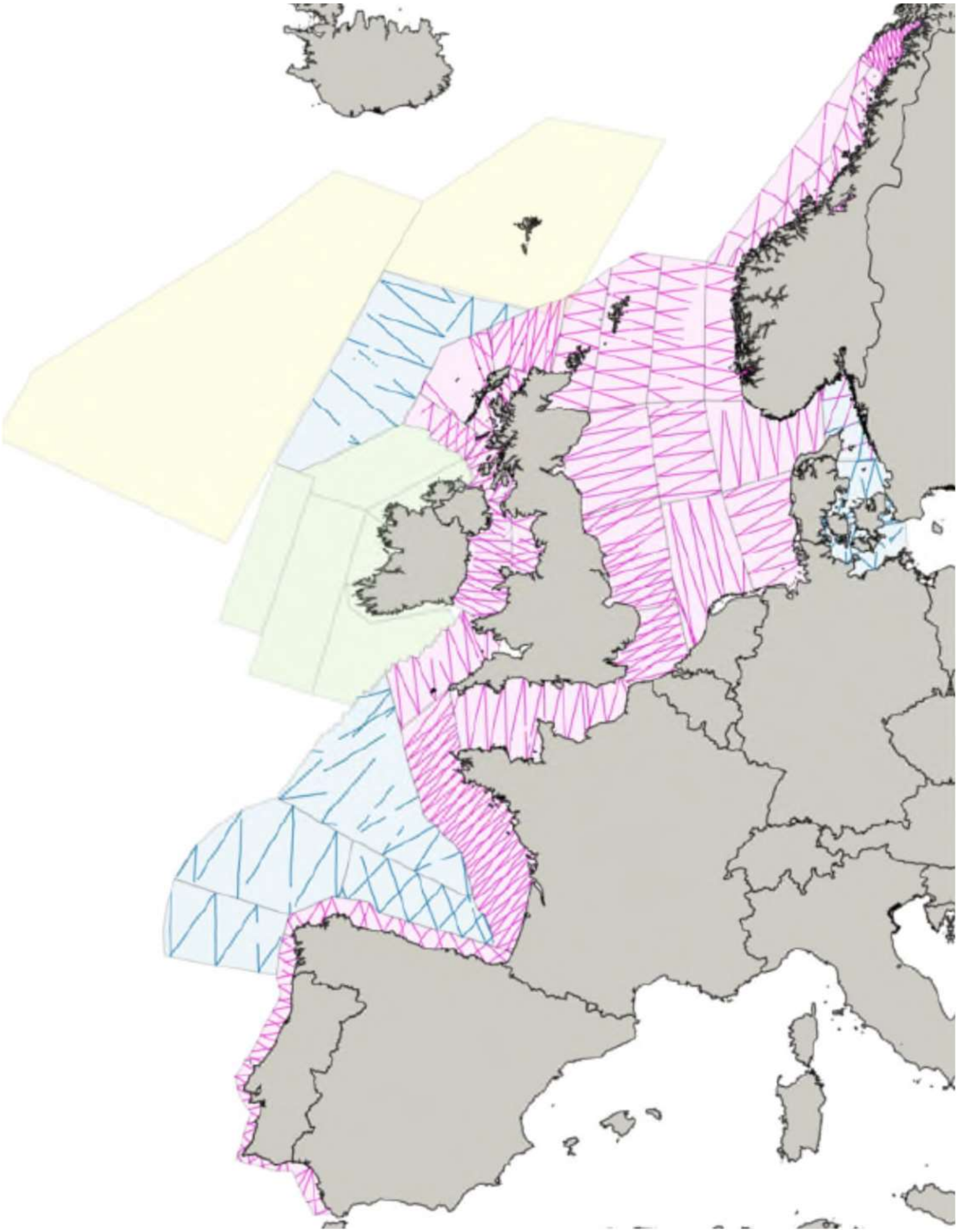
$$= \left(\frac{A_1}{A}\right)\hat{D}_1 + \left(\frac{A_2}{A}\right)\hat{D}_2 + \left(\frac{A_3}{A}\right)\hat{D}_3$$

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

Note form of equation

Example: SCANS III (2016)

Small Cetaceans in European Atlantic waters and the North Sea



CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

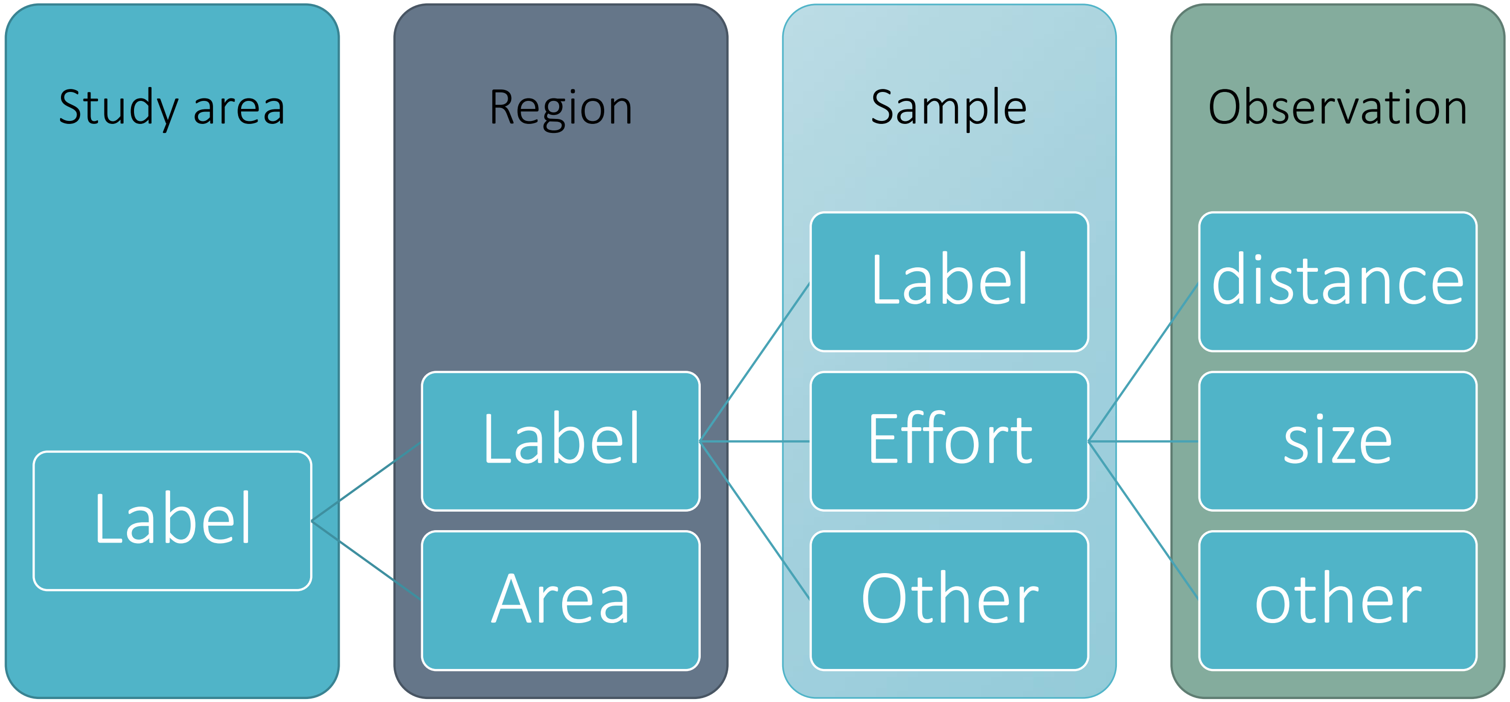
Example of stratified data

	Region.Label	Area	Sample.Label	Effort	distance	size
1	Ideal	84734	1	86.75	0.10	1
2	Ideal	84734	1	86.75	0.22	1
3	Ideal	84734	1	86.75	0.16	3
:	:	:	:	:	:	:
45	Ideal	84734	13	75.63	0.03	1
46	Marginal	630582	14	83.33	0.49	1
47	Marginal	630582	14	83.33	1.94	1
48	Marginal	630582	14	83.33	1.10	1
:	:	:	:	:	:	:
99	Marginal	630582	25	107.72	NA	1

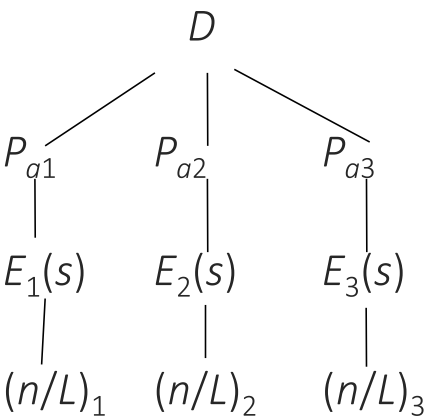
CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Data organisation hierarchy



Example: Full geographic stratification



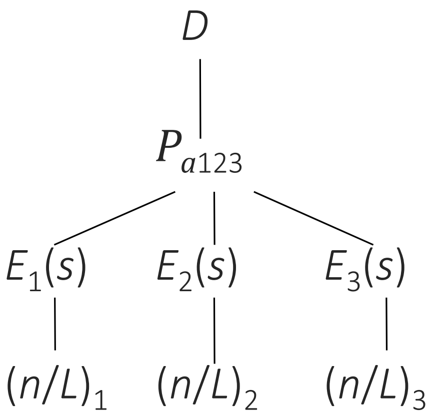
Select strata and fit `ds` to each strata
E.g. (many ways to perform selection)

```
ideal.dat <- whales[whales$Region.Label=="Ideal", ]
whales.ideal <- ds(data=ideal.dat, key="hr")
```

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Example: P_a pooled



Data contains different strata in `Region.Label`
The `ds` function performs this stratification by default

```
whale.pool <- ds(whales, key="hr”)
```

E.g. Part of output from `summary(whale.pool)`

```
Summary …
Abundance:
  Label      Estimate        se        cv      lcl       ucl       df
1 Marginal 12181.313 4638.5533 0.3807926 5499.920 26979.371 12.96742
2 Ideal     3653.313  910.0737 0.2491091 2181.589  6117.879 17.96142
3 Total    15834.626 4834.1865 0.3052921 8388.389 29890.769 15.25427
```

Doing the same thing with `dht2`

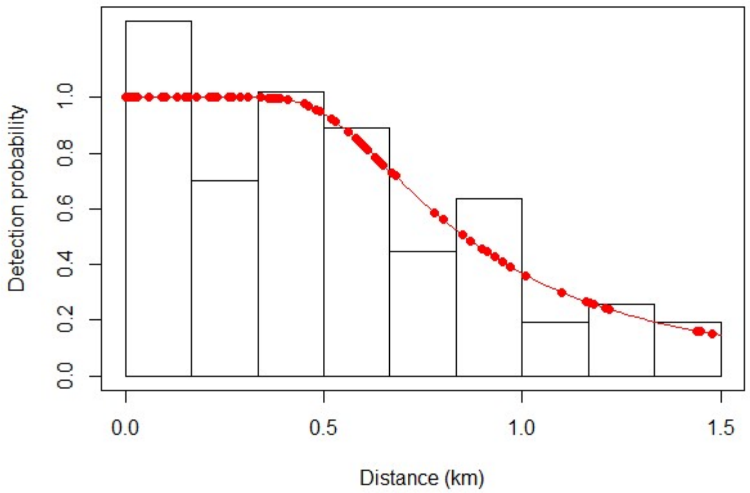
The same stratification (i.e. pooled P_a) can be achieved with:

```
dht2(model=whale.pool, flatfile=whales,
        strat_formula=~Region.Label,
        stratification="geographical")
```

The `dht2` function can be used for more complicated stratification (as we will see later)

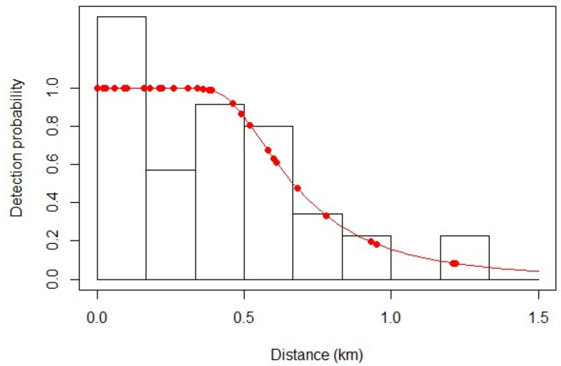
Pooled *vs* Stratified P_a

Pooled n=88

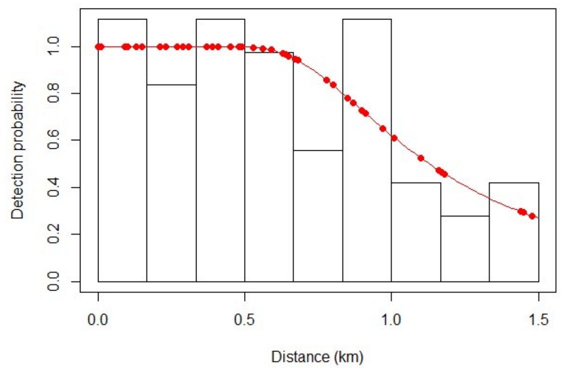


Stratified

Ideal habitat n=39



Marginal habitat n=49



CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

It is a Model Selection Problem

	Pooled	Stratum 1	Stratum 2	Stratum Sum
Log likelihood $\log_e(L)$	-180.490	-72.699	-104.676	-177.375
No. parameters (q)	2	2	2	4
AIC	364.980	149.398	213.352	362.75

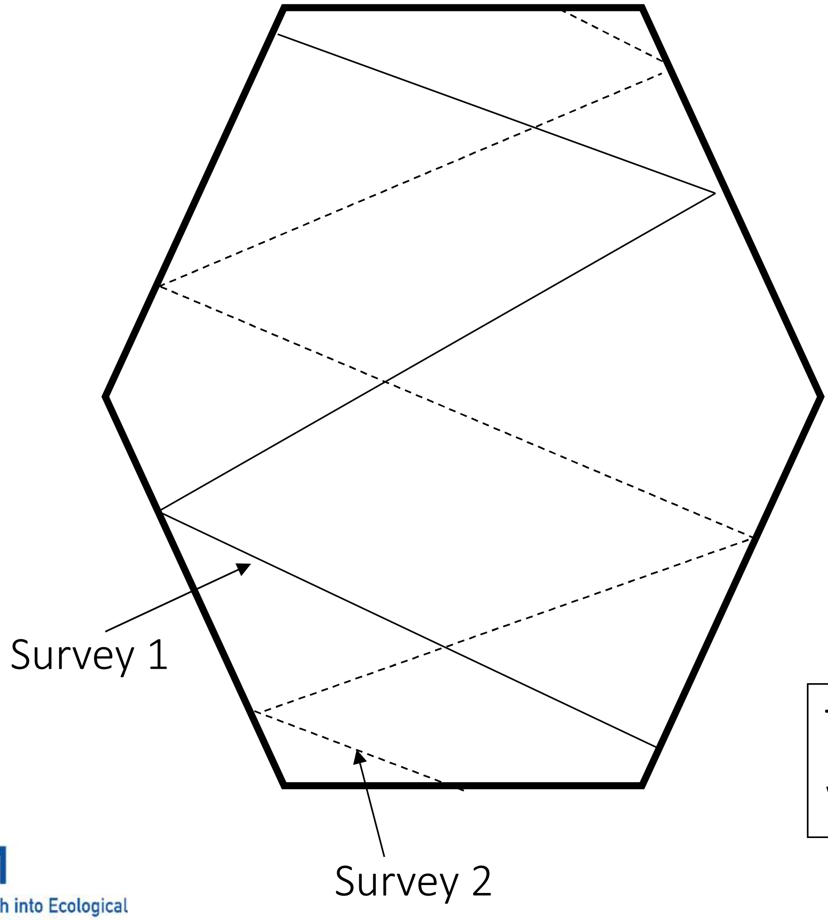
Criterion for stratification of P_a:
Fit separate P_a for each strata if

$$AIC_{pooled} > \sum_{strata} AIC_{stratum}$$

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

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

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Stratification by survey (with common detection function)

Need to use the `dht2` function:

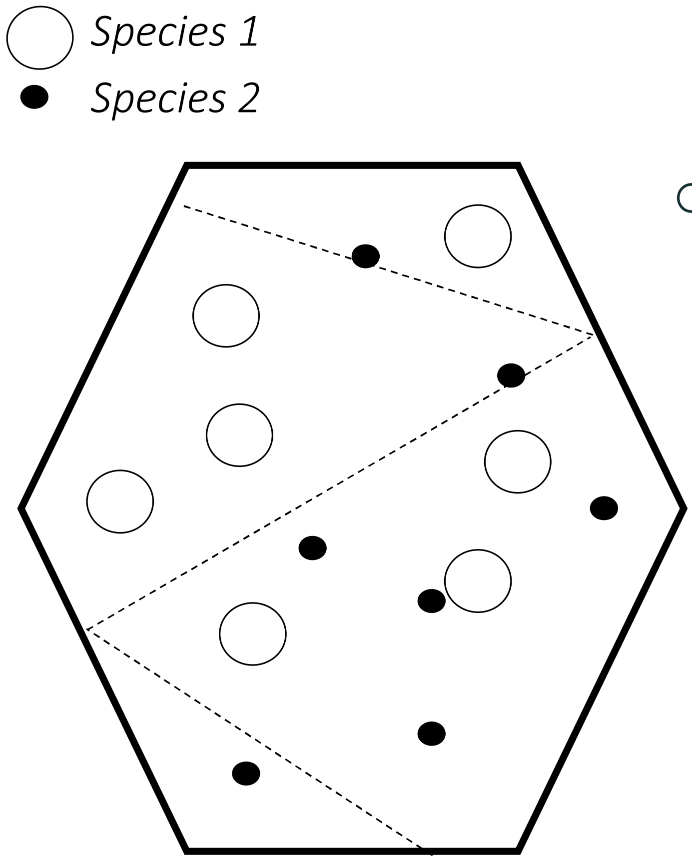
```
dht2(model=whale.pool, flatfile=whales,
     strat_formula=~Region.Label,
     stratification="replicate")
```

This represents different surveys

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Stratification by species (with common detection function)



$$\hat{D} = \hat{D}_{sp1} + \hat{D}_{sp2}$$

```
dht2(model=whales.pool,
     flatfile=whales,
     strat_formula=~species,
     stratification="object")
```

Need this column in data

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

When to use `dht2` for density or abundance estimates

- Use density or abundance estimates from **`ds`** when:
  - no stratification
  - simple geographical stratification (i.e. specified in `Region.Label`)
  - no multipliers/cue counts
  - Typical encounter rate variance estimators

    Rather than the estimators used to produce better variance estimates under systematic surveys (from precision lecture)
- Call **`dht2`** after detection function fitting with **`ds`** in other situations

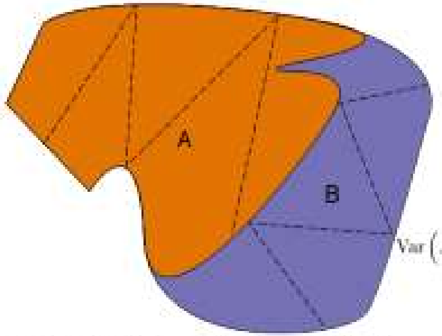
Stratification in `dht2`

(Density edition)

There are four stratification options in dht2, this cheatsheet shows how density and its variance are calculated and gives examples of when to use them

Geographical `(stratification="geographical")`

Each stratum represents a different geographical area, we want the total density over all the areas



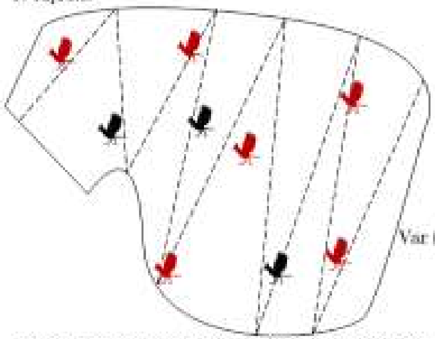
$$\hat{D} = \sum_{m=1}^{M} \frac{A_m}{A_{\text{total}}} \hat{D}_m$$

$$\text{Var}\left(\hat{D}\right) = \sum_{m=1}^{M} \left(\frac{A_m}{A_{\text{total}}}\right)^2 \text{Var}\left(\hat{D}_m\right)$$

Example: estimates are required for areas "A" and "B", above as well as an estimate of total density in and its variance.

Object `(stratification="object")`

Objects are of different "classes", for example sex, species or life stage. Post-stratification is then required to obtain the density of individuals across all the classes of objects.



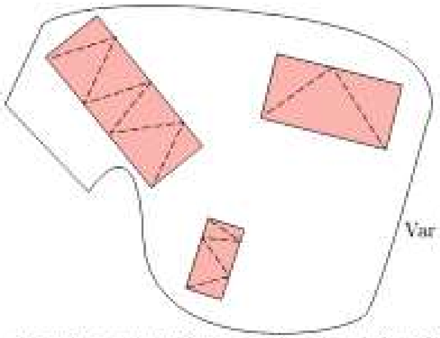
$$\hat{D} = \sum_{m=1}^{M} \hat{D}_m$$

$$\text{Var}\left(\hat{D}\right) = \sum_{m=1}^{M} \text{Var}\left(\hat{D}_m\right)$$

Example: if you have stratified by colour (red/black), but also want a total number of animals.

Effort-weighted sum `(stratification="effort_sum")`

Strata are from surveys (perhaps using different designs) but you don't have many replicates and/or want an estimate of "average variance" and average density



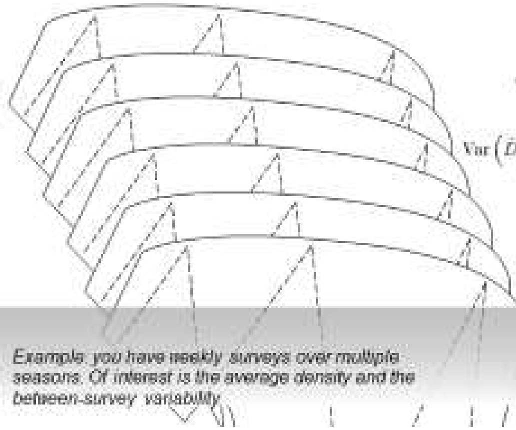
$$\hat{D} = \sum_{m=1}^{M} \frac{L_m}{\sum_{k=1}^{M} L_k} \hat{D}_m$$

$$\text{Var}\left(\hat{D}\right) = \sum_{m=1}^{M} \left(\frac{L_m}{\sum_{k=1}^{M} L_k}\right)^2 \text{Var}\left(\hat{D}_m\right)$$

Example: surveys (red) were made and are believed to be representative of the larger study area so they can be summed (weighted by the amount of effort) to obtain an average density.

Replicate `(stratification="replicate")`

Many replicate surveys have been conducted and the average density weighted by amount of effort is required, along with variance between the surveys



$$\hat{D} = \sum_{m=1}^{M} \frac{L_m}{\sum_{k=1}^{M} L_k} \hat{D}_m$$

$$\text{Var}\left(\hat{D}\right) = \sum_{m=1}^{M} \frac{\left(\hat{D}_m - \hat{D}\right)^2}{M-1}$$

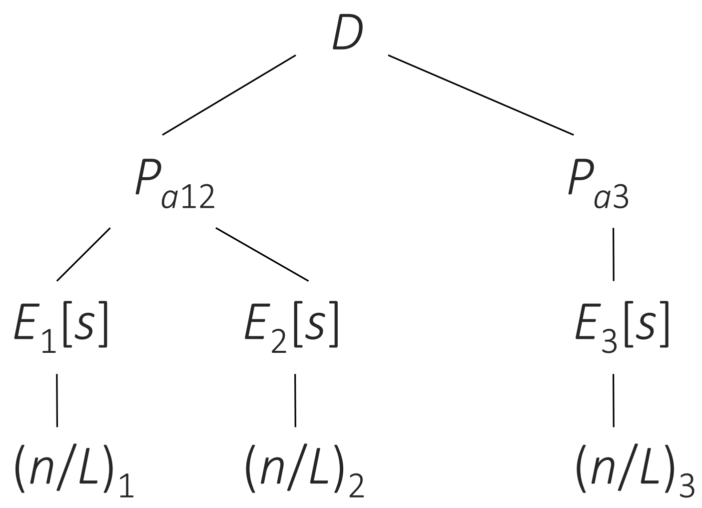
Example: you have weekly surveys over multiple seasons. Of interest is the average density and the between-survey variability.

Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Limitations in Distance

- Distance cannot currently do multilevel stratification in one command
- Two runs are necessary
  - Estimate P_a, $E[s]$ and n/L by stratum
  - Combine strata 1 and 2 to estimate P_{a12}
- Care must be taken when calculating CVs because the density estimates for stratum 1 and 2 have an estimated P_a in common



CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Alternatives to stratification

- 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

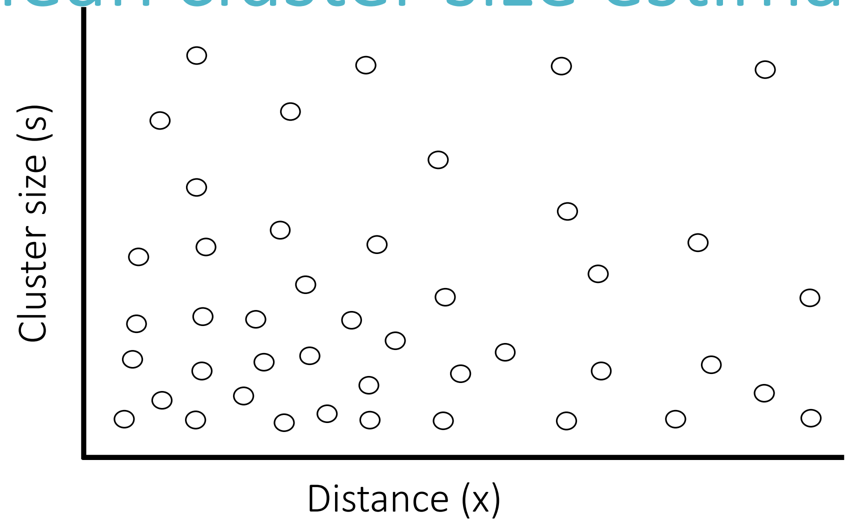
    - MCDS methods are covered in the next lecture.

Analysis of Populations in Clusters

CREEM
Centre for Research into Ecological and Environmental Modelling

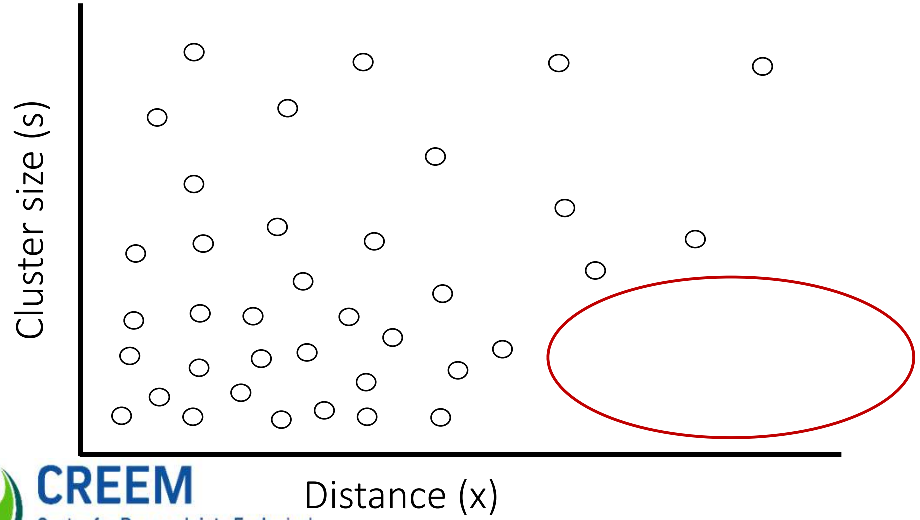
University of St Andrews

Mean cluster size estimation



No Size Bias

- Mean of observed sizes does not change with distance



Size Bias

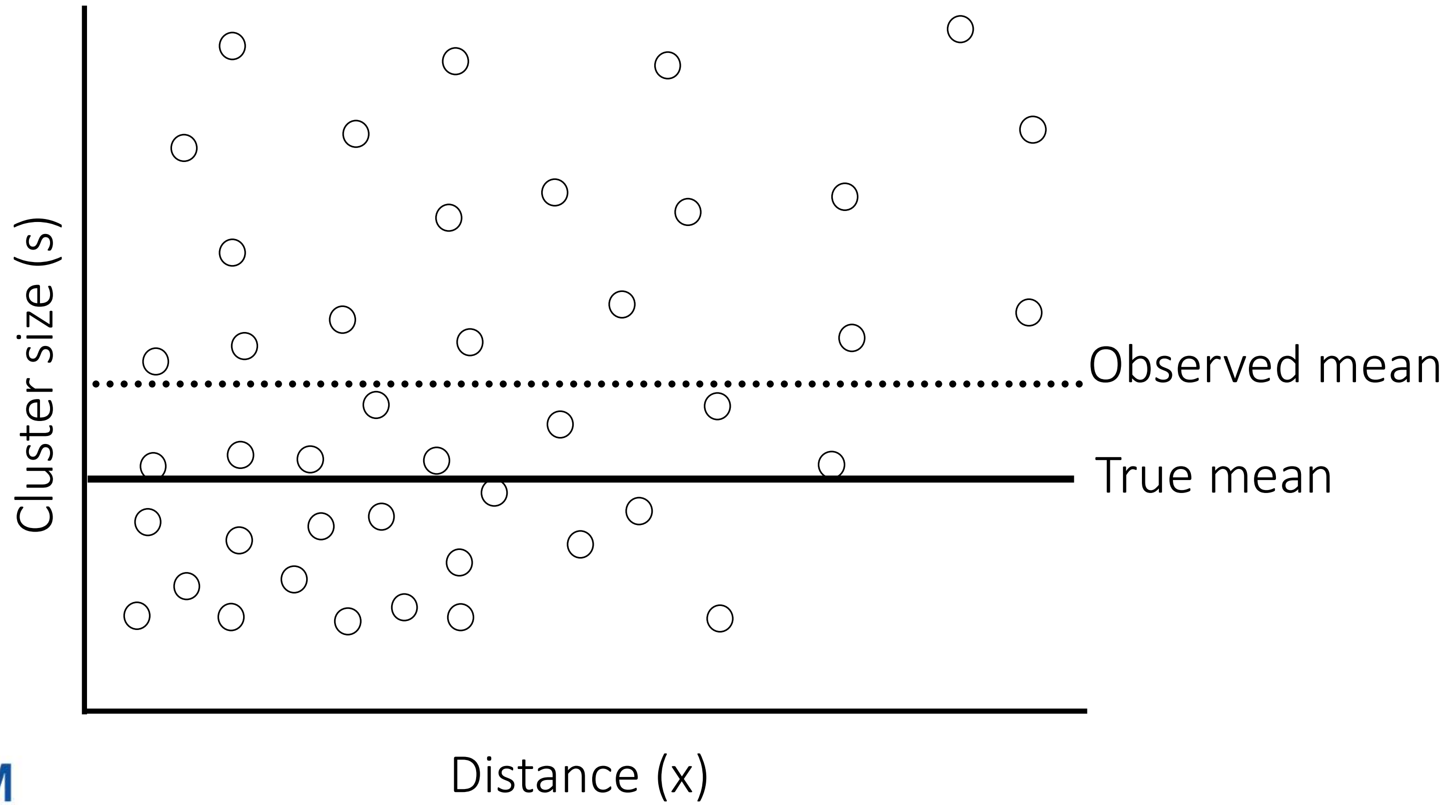
- Smaller clusters less detectable at larger distances
- Mean observed cluster size **increases** with distance

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Effect of size bias on sample mean

If size bias is present, $\hat{E}(s) = \bar{s}$ will be positively biased:



Remedy to size bias

- Recognize that detection of cluster
- *Depends upon cluster size*
- Model the dependence in the detection function
- *using covariates in the detection function*
- Details tomorrow