Sample size

Estimating the required sample size when designing a distance sampling survey.

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Sample size

- Aim for at least 60-80 sightings for fitting the detection function
- and at least 20 lines or points for estimating encounter rate *n/L* or *n/k*
- Whether reliable estimates can be obtained from smaller samples depends on the data

Sample size – continued

More observations are required:

- if detection function is spiked
- if population is highly aggregated
- for point transect sampling

Increasing sample size using repeat counts

If a line is sampled three times,

- pool the distance data from the three visits
- enter survey effort as three times the line length.

If a point is sampled three times,

- enter survey effort as 3.

Determining total line length

Pilot study: n_0 animals (or clusters) counted from lines totalling L_0 in length.

Total line length required in main survey is

$$L = \left(\frac{q}{\left[cv_t\left(\hat{D}\right) \right]^2} \right) \times \frac{L_0}{n_0}$$

Where $cv_t\left(\hat{D}\right)$ is the target cv (e.g. 10% is 0.1) and…

Determining line length (cont)

q is approximately $\dfrac{V(n)}{n} + \dfrac{nV[\hat{f}(0)]}{[\hat{f}(0)]^2}$

Pilot studies are typically too small to estimate q. If past similar data sets are not available, assume $q = 3$.

Line length example

A pilot study yields n_0 = 20 observations from lines of total length 5km. We require a CV of 10%, and assume q = 3.

$$L = \frac{3}{0.1^2} \times \frac{5}{20} = 75\text{km}$$

Estimated sample size is

$$n = L \times \frac{n_0}{L_0} = 75 \times \frac{20}{5} = 300$$

Determining line length (cont)

If pilot survey is sufficiently large, calculate line length for main survey as

$$L = \frac{L_0[cv(\hat{D}_0)]^2}{[cv_t(\hat{D})]^2}$$

where

$cv(\hat{D}_0)$ is the cv of estimated density obtained from the pilot survey, and L is total line length in the main survey

Point transects: number of points

or

$$k = \left(\frac{q}{[cv_t(\hat{D})]^2} \right) \times \frac{k_0}{n_0}$$

$$k = \frac{k_0[cv(\hat{D}_0)]^2}{[cv_t(\hat{D})]^2}$$

where k_0 points in the pilot survey yielded n_0 detections, or estimated density of $\hat{D}_0$