# Sampling Designs

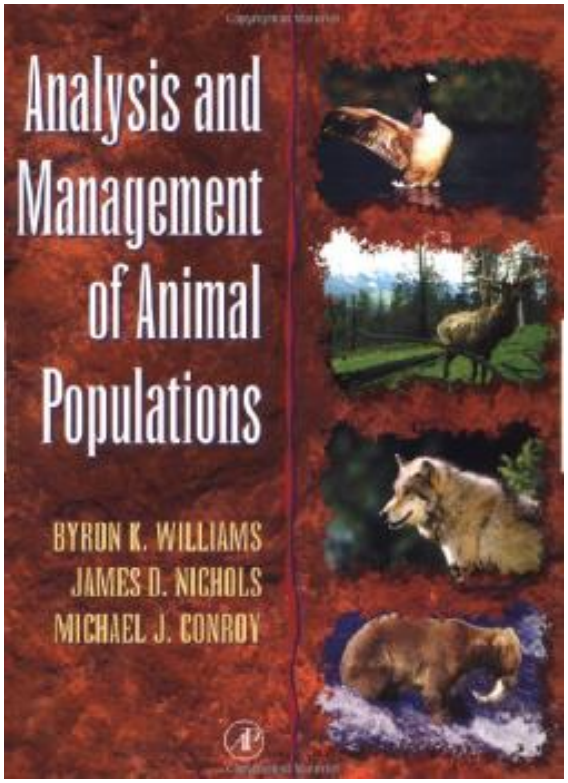
## Outline

Sampling design features

- Replication
- Randomization
- Control of variation

Some designs

- Simple random sampling
- Stratified random sampling
- Other (cluster, systematic, double)



# Define study objectives

- What are you trying to achieve
- Quantities that are being estimated (abundance, occ.)

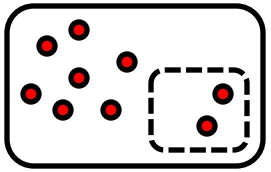
- Spatial (size) and temporal scope (duration)

- Establish criteria of reliability: bias and precision

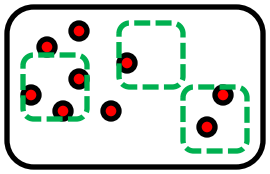
- Constraints: costs, logistical limitations

# Design issue

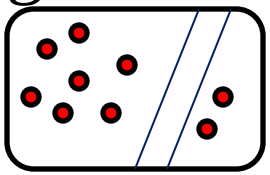
- Define target and sample population



- Replication: multiple samples from sampled population



- Randomization: selection of random samples. Guards against the systematic influence of unrecognized sources of variation



- Estimate reliability: bias and precision

# Detection probability & SV

Fraction of total area (A)

$$C = \alpha \times \beta \times N$$

Detection probability

$$N = \frac{C}{\alpha\beta}$$

# Sampling designs

- Simple Random
- Stratified
- Cluster
- double

# Simple random sampling

- *n* sampling units selected from list of *N* units
- Each unit has probability *n/N* of being selected
- With or without replacement.
- Population mean $\bar{Y} = ( y_1 + ... + y_N )/N$
- Unbiased estimate is sample mean $\bar{y} = \sum_{i=1}^{n} y_i/n$
- With replacement, the variance
- Estimated by $\hat{v}ar(\bar{y}) = s^2/n$

$$s^2 = \sum_{i=1}^{n} ( y_i - \bar{y} )^2/(n-1).$$

# Simple Random Sampling

- Sampling without replacement, need to accommodate for finite population:

$$\hat{v}ar(\bar{y}) = \frac{s^2}{n}(1 - n/N),$$

| | | |
|---|---|---|
| **8** | **4** | **5** |
| **6** | **8** | **3** |
| **3** | **6** | **5** |

# Simple Random Sampling

- Population Total: $\hat{Y} = N\bar{y}$

- Estimate:

$$\hat{v}ar(\hat{Y}) = N^2 \hat{v}ar(\bar{y}),$$

- Williams et al. (2002), p. 63

# Sample Size

$$n = \left( \frac{z_{\alpha/2}}{r} CV \right)^2 \qquad n' = \frac{n}{1 + n / N}$$

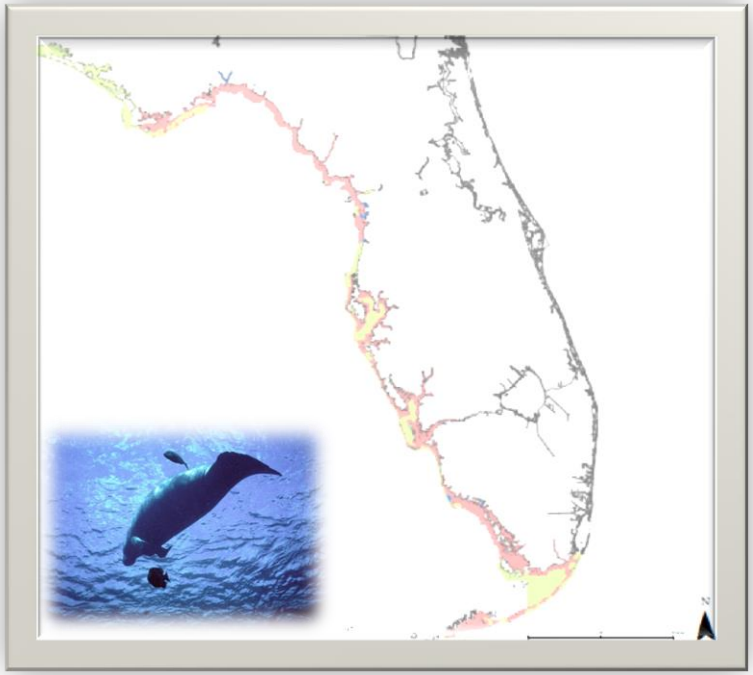
*r is the desired precision as % of true mean*
*z_alpha/2: upper alpha point from the standard normal*

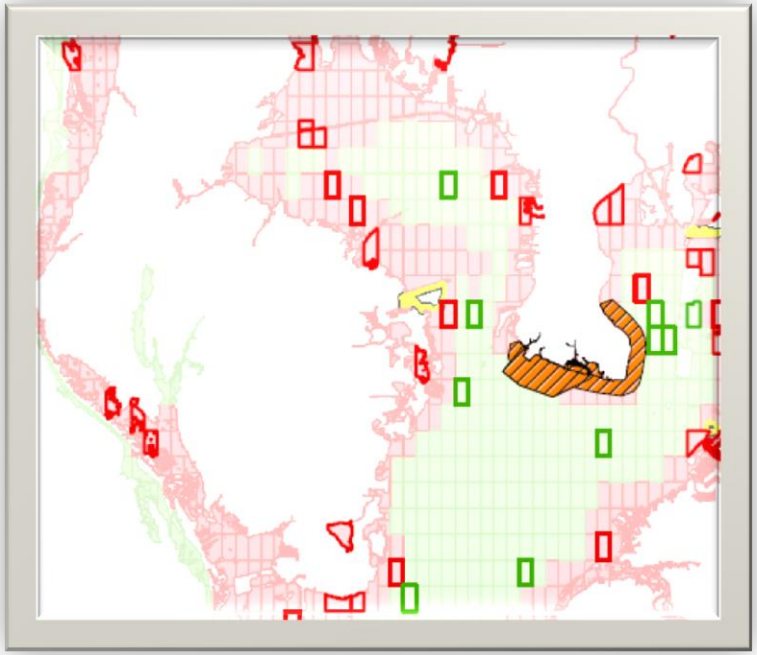
- Total area: N plots (N=910), the goal to estimate the mean number of Pine Beetles per plot within 10% of true value (95% of the time [i.e., alpha=0.05])
- CV based on pilot~0.486

$$n = (\frac{1.96}{.1} 0.486)^2 \sim 91 \qquad n' = \frac{91}{1 + 91/910} \sim 83$$

# Stratification

Strata 1: Aggregation
Strata 2: medium density
Strata 3: low density

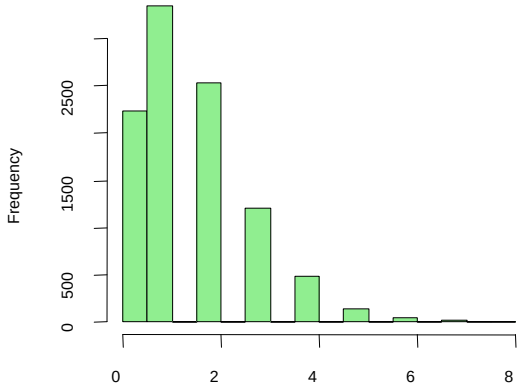


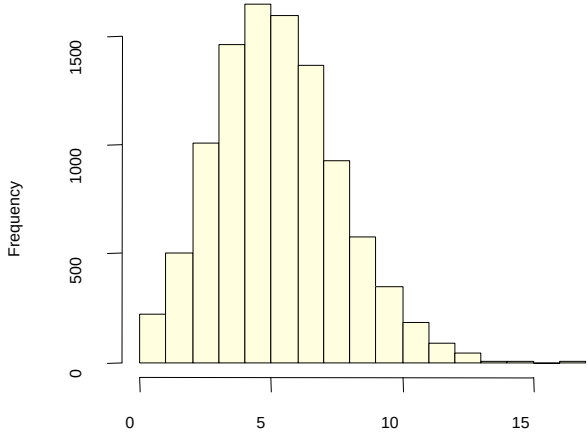


# Design: spatial variation

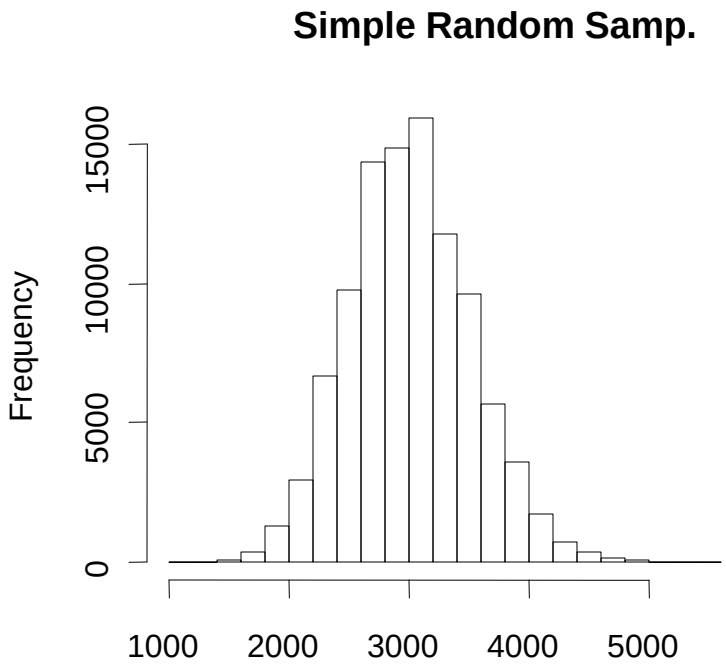
## 300 sites are sampled to estimate N (3003)

- 5040 sites have no animals
- 1000 sites have on average 1.5 animals
- 260 sites have on average 5.78 animals

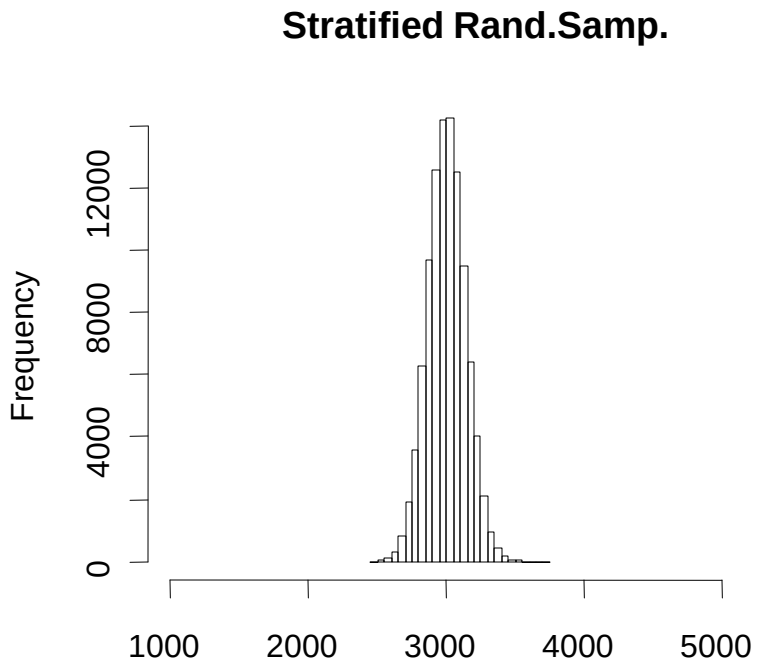


# Stratified random sampling




Abundance: 3006 (2079 to 4053)




Abundance: 3003 (2735 to 3274)

# Stratified Random Sampling

$$\bar{y}_{st} = \sum_{i=1}^{I} W_i \bar{y}_i \qquad W_i = \frac{N_i}{N}$$

*y_bar_st: mean number of organisms*

*Stratum (i)*

*Ni: number of plots in stratum i*

*N: total # of plots*

*yi-bar:  mean number of organisms per plot*

# Stratified Random Sampling

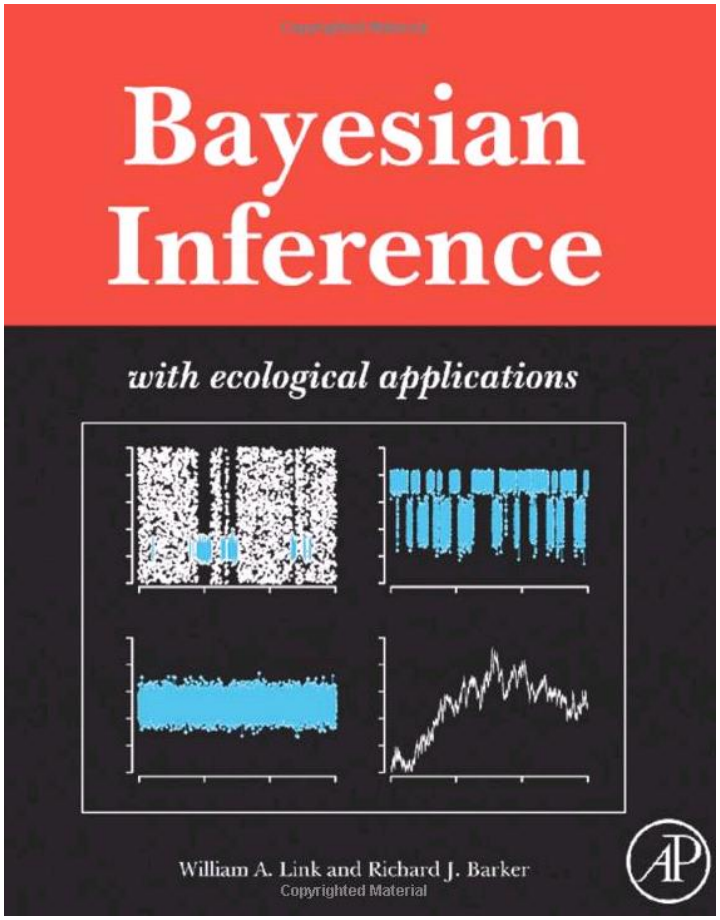
- Sampling variance is

$$\hat{v}ar(\bar{y}_{st}) = \frac{1}{N^2}\sum_i N_i\left(N_i - n_i\right)\frac{s_i^2}{n_i}$$

$$s_i^2 = \sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2/\left(n_i - 1\right)$$

# Stratified Random Sampling with Bayesian methods

- More flexible method
- Not based on asymptotic approximation
- Hierarchical modeling
- Finite population



# Sample size stratified random sampling

- *Proportional allocation:* sample units are allocated according to relative sizes of strata

$$n_i = n\frac{N_i}{N}$$

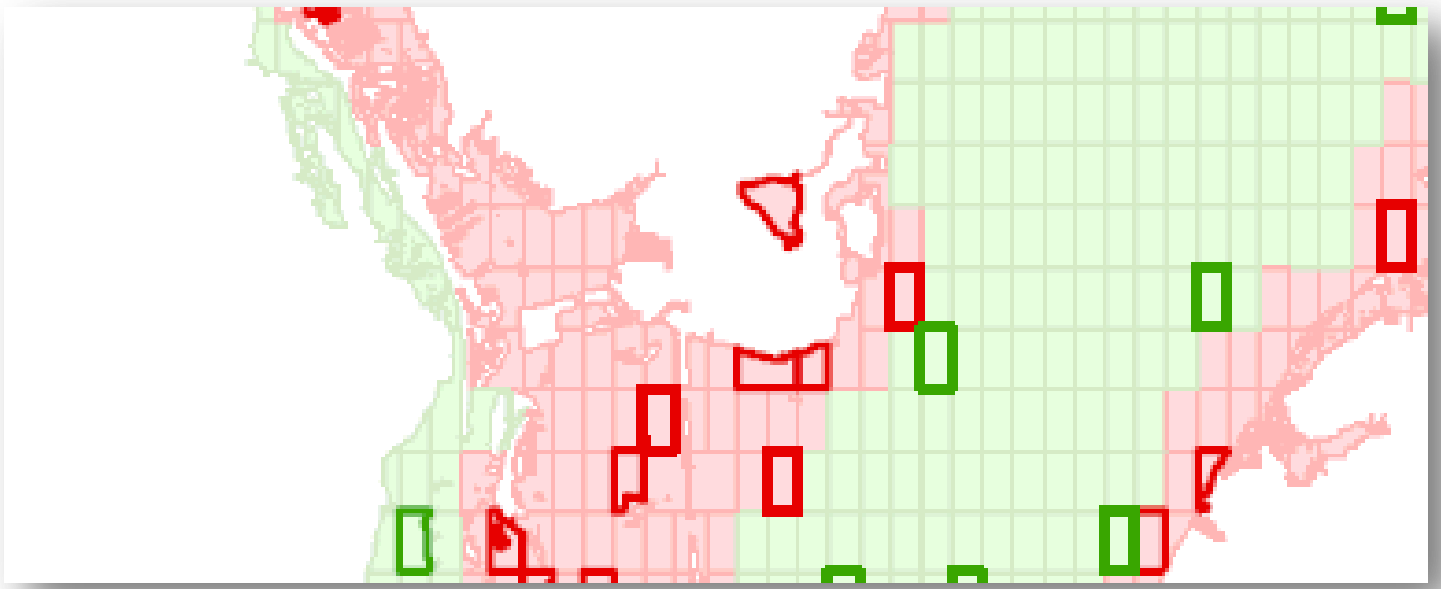
- *Optimal allocation* allocate samples to strata in a manner that minimizes variance

$$n_i = n\left(\frac{N_i\sigma_i}{\sum_i N_i\sigma_i}\right).$$

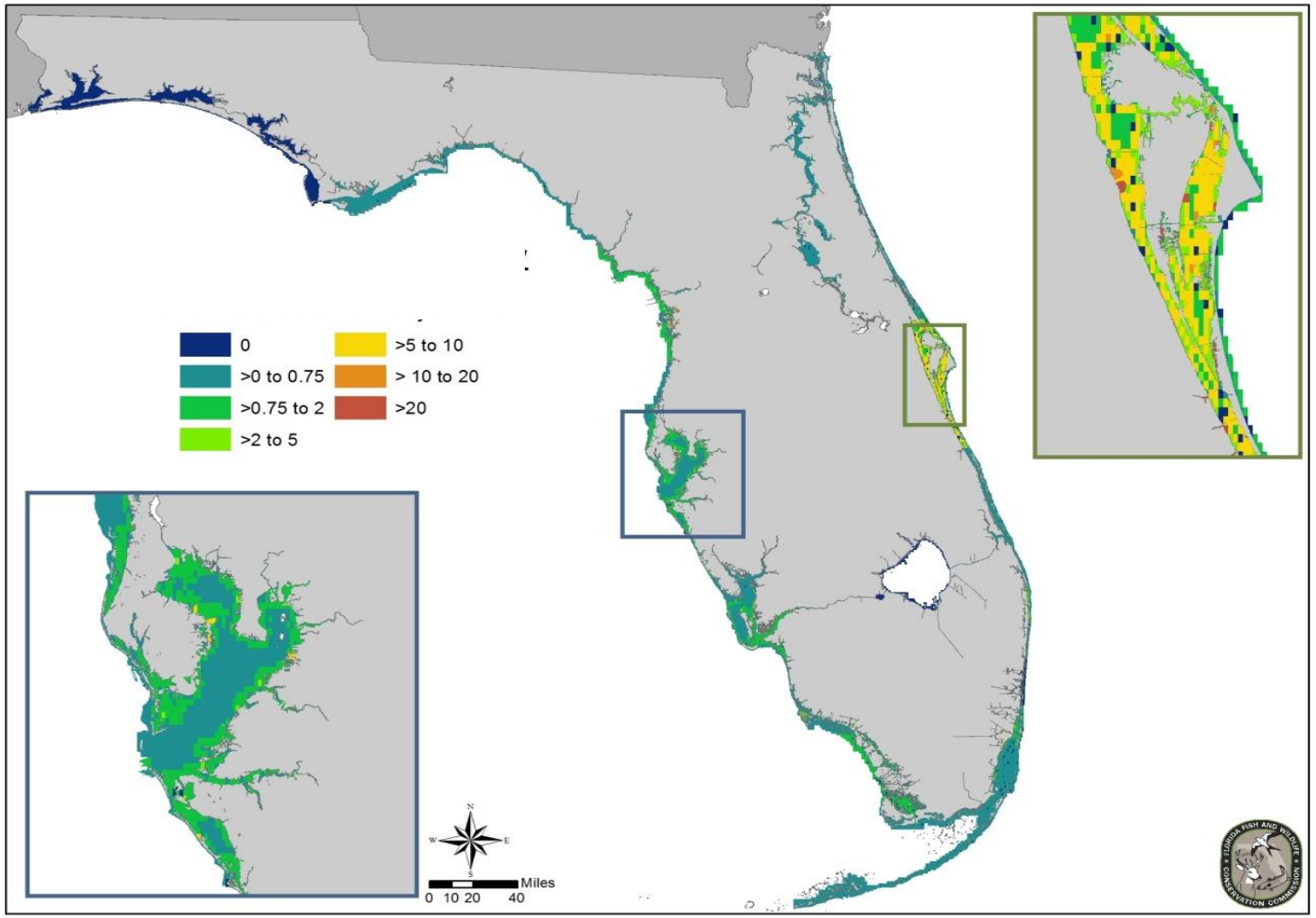
# Unequal areas

- If plots are unequal in size

$$\hat{N}_s = \sum_{i=1}^{I} \frac{\bar{y}_i}{\bar{a}_i} A_i$$



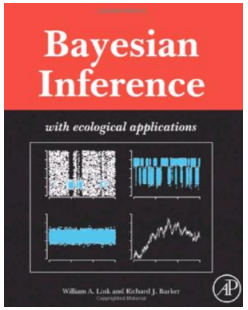
# Example



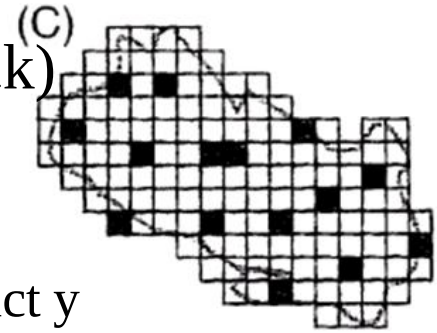
# Systematic Sampling

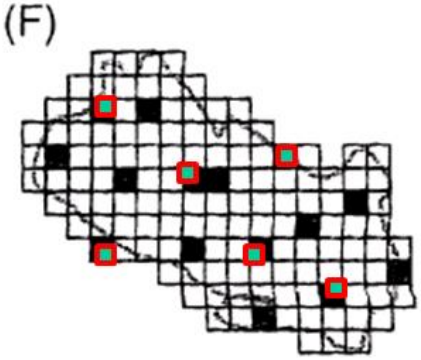
- Very popular among ecologists because simple
  - Sample at regular time intervals
- Treat as simple random samples
- Potential flaw:  variances underestimated
  - Measurements along environmental gradients
  - Correlated measures among adjacent individuals.
- Variation on the theme:  random sampling from grids
  - Random starts (still systematic)

# Double Sampling

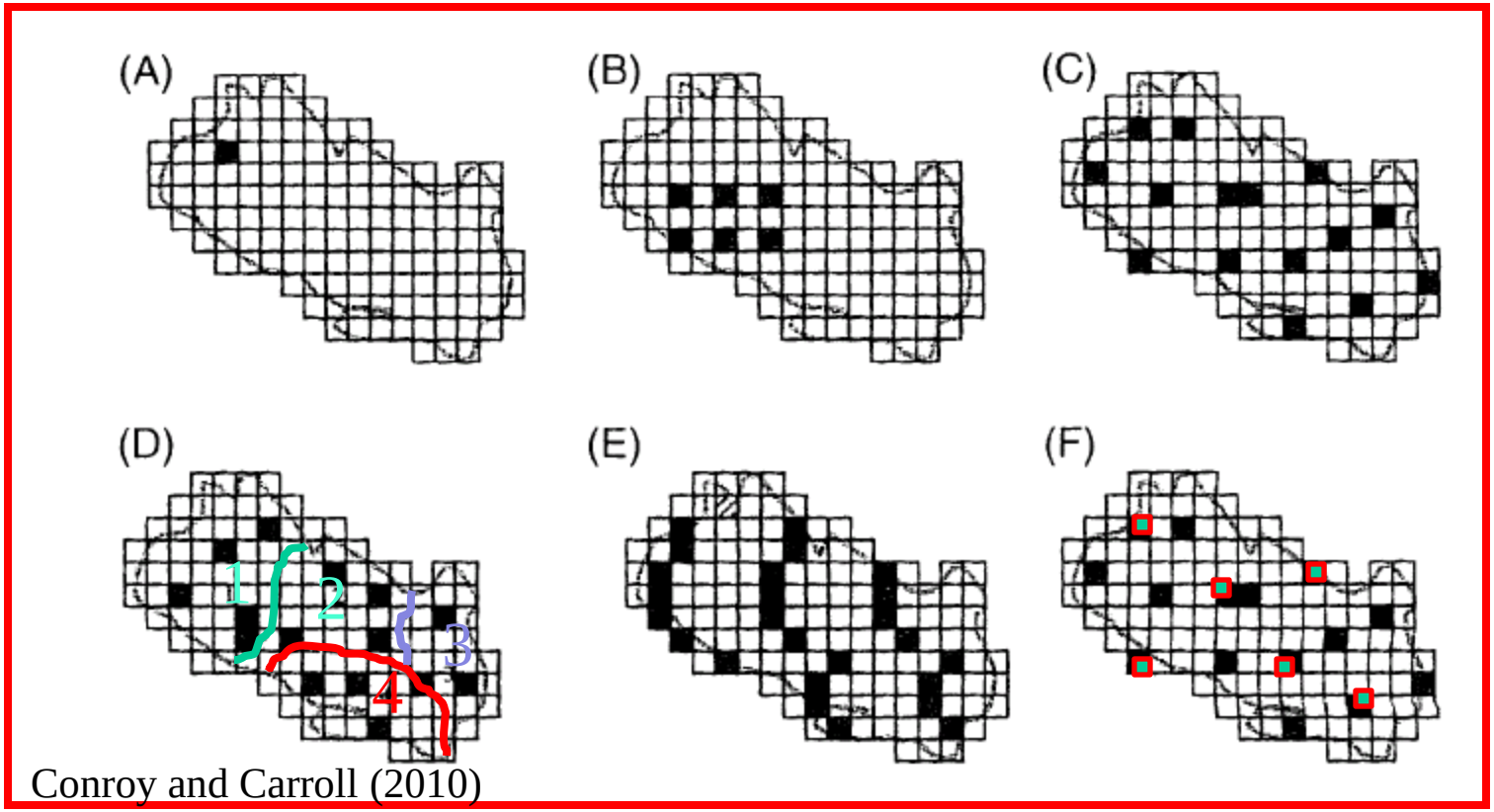



- Often, have extensive (good) sample, and extensive (poorer) sample method
  - Auxiliary variable $(x_i)$ on a sample of $n'$ units (black)
  - Primary variable $(y_i)$ on a subsample of size $n$ (red)
    - Ground counts vs aerial surveys
    - The n units (overlap) are used to build stat model to predict y from x with linear model (Williams et al. 2002, p. 70)
- Underutilized method
  - Cost efficient



# Sampling schemes



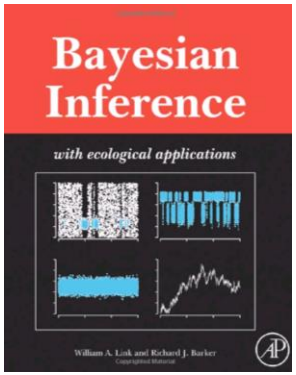

Conroy and Carroll (2010)

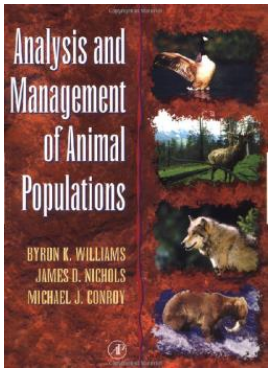
# Cluster sampling

- *M* :number of clusters in a population (M=50 ponds)
  - *m* : size of a sample of clusters (e.g., sample of 5 ponds)

$$\hat{Y} = \frac{M}{m}\sum_{i=1}^{m} y_i \qquad \hat{Y} = \frac{50}{5} \times 334 = 3340$$

- $y_i$ is # nests in pond i



# Take home points

- Design issues (objectives, randomization, replication)
- Spatial variation, detection
- Stratification to improve precision
- Optimization
- Unequal areas
- Cluster, double and other designs

# References


- Cochran, W. G. 1977. Sampling techniques, 3rd edition. Wiley, New York. 428pp.
- Conroy, M. J. and Carroll, J. P. 2010. Quantitative conservation of vertebrates, Wiley-Blackwell
- Skalski, J. R. 1994. Estimating wildlife populations based on incomplete areas surveys. Wildl. Soc. Bull 22:192-203.
- Thompson, W. L., G. C. White, and C. Gowan. 1998. Monitoring Vertebrate Populations. Academic Press, New York.
- Williams et al. 2002, Analysis and management of animal populations. Academic Press.