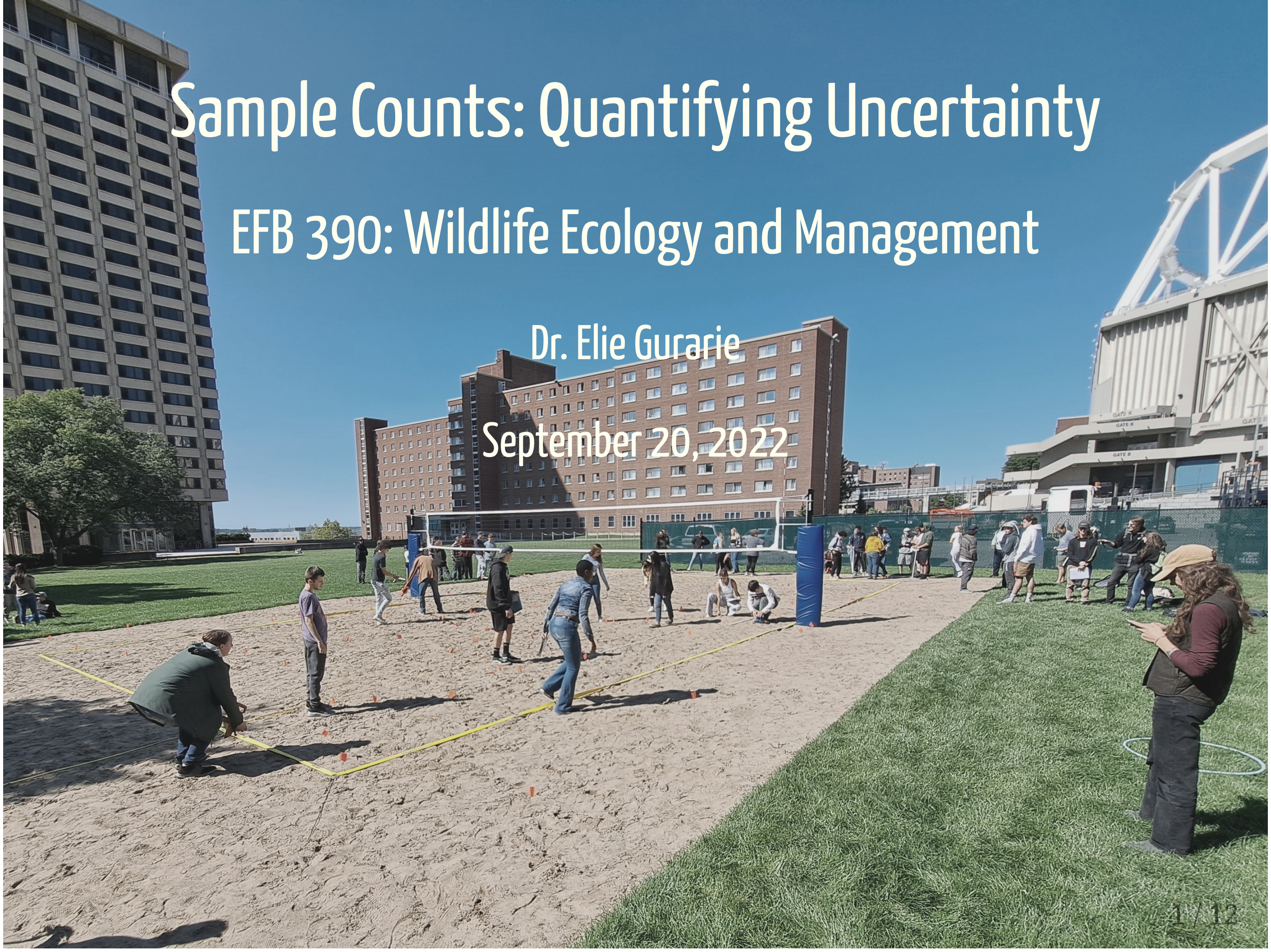
Sample Counts: Quantifying Uncertainty

EFB 390: Wildlife Ecology and Management

Dr. Elie Gurarie

September 20, 2022



The trickiest thing in sampling

Is computing the *precision* (standard errors / confidence intervals)

respectively. The expectation of Eq. 4.17 is

$$\begin{aligned} E\left({}_1\widehat{\boldsymbol{\Psi}}_k\right) =& E_{s_w}\left\{\frac{M_k}{m_{2k}}\sum_{i=1}^{m_{2k}} E_{s_{L_i}}\left[\frac{N_{ki}}{n_{ki}} E_{s_O}\left(\frac{n_k}{n_{Ok}} E_{s_A}\left(\frac{N_{Om}}{n_{Am}}\psi_{Aki}\right)\right)\right]\right\} \\ =& E_{s_w}\left\{\frac{M_k}{m_{2k}}\sum_{i=1}^{m_{2k}} E_{s_{L_i}}\left[\frac{N_{ki}}{n_{ki}} E_{s_O}\left(\frac{n_k}{n_{Ok}}\psi_{Oki}\right)\right]\right\} \\ =& E_{s_w}\left[\frac{M_k}{m_{2k}}\sum_{i=1}^{m_{2k}} E_{s_{L_i}}\left(\frac{N_{ki}}{n_{ki}}\psi_{Lki}\right)\right] = E_{s_w}\left(\frac{M_k}{m_{2k}}\sum_{i=1}^{m_{2k}}\boldsymbol{\Psi}_{ki}\right) = \boldsymbol{\Psi}_k. \end{aligned}$$

The variance of the estimator can be written as

$$\begin{aligned} V\left({}_1\widehat{\boldsymbol{\Psi}}_k\right) =& \underbrace{V_{s_w}\left\{E_{s_{L_i}}\left[E_{s_O}\left(E_{s_A}\left({}_1\widehat{\boldsymbol{\Psi}}_k\right)\right)\right]\right\}}_{V_1} + \underbrace{E_{s_w}\left\{V_{s_{L_i}}\left[E_{s_O}\left(E_{s_A}\left({}_1\widehat{\boldsymbol{\Psi}}_k\right)\right)\right]\right\}}_{V_2} \\ &+ \underbrace{E_{s_w}\left\{E_{s_{L_i}}\left[V_{s_O}\left(E_{s_A}\left({}_1\widehat{\boldsymbol{\Psi}}_k\right)\right)\right]\right\}}_{V_3} + \underbrace{E_{s_w}\left\{E_{s_{L_i}}\left[E_{s_O}\left(V_{s_A}\left({}_1\widehat{\boldsymbol{\Psi}}_k\right)\right)\right]\right\}}_{V_4}. \end{aligned}$$

Component-wise,

$$V_1 = V_{s_w}\left(\frac{M_k}{m_{2k}}\sum_{i=1}^{m_{2k}}\boldsymbol{\Psi}_{ki}\right) = M_k\left(\frac{M_k}{m_{2k}}-1\right)\frac{\sum_{i=1}^{M_k}\left(\boldsymbol{\Psi}_{ki}-\overline{\boldsymbol{\Psi}}_k\right)\left(\boldsymbol{\Psi}_{ki}-\overline{\boldsymbol{\Psi}}_k\right)^T}{M_k-1},$$

$$\begin{aligned} V_2 =& E_{s_w}\left[\left(\frac{M_k}{m_{2k}}\right)^2\sum_{i=1}^{m_{2k}} V_{s_{L_i}}\left(\frac{N_{ki}}{n_{ki}}\psi_{Lki}\right)\right] \\ =& \frac{M_k}{m_{2k}}\sum_{i=1}^{M_k} N_{ki}\left(\frac{N_{ki}}{n_{ki}}-1\right)\frac{N_{ki}\left[\operatorname{diag}\left(\mathbf{P}_{ki}\right)-\mathbf{P}_{ki}\mathbf{P}_{ki}^T\right]}{N_{ki}-1}, \end{aligned}$$

$$\begin{aligned} V_3 =& E_{s_w}\left\{\left(\frac{M_k}{m_{2k}}\right)^2 E_{s_{L_i}}\left[V_{s_O}\left(\sum_{i=1}^{m_{2k}}\frac{N_{ki}}{n_{ki}}\frac{n_k}{n_{Ok}}\psi_{Oki}\right)\right]\right\} \\ =& E_{s_w}\left\{\left(\frac{M_k}{m_{2k}}\right)^2\sum_{i=1}^{m_{2k}} E_{s_{L_i}}\left[\left(\frac{N_{ki}}{n_{ki}}\right)^2 V_{s_O}\left(\frac{n_k}{n_{Ok}}\psi_{Oki}\right)\right]\right\} \\ &+ E_{s_w}\left\{\left(\frac{M_k}{m_{2k}}\right)^2\sum\sum_{i\neq j}^{m_{2k}} E_{s_{L_i}}\left[\frac{N_{ki}}{n_{ki}}\frac{N_{kj}}{n_{kj}} Cov_{s_O}\left(\frac{n_k}{n_{Ok}}\psi_{Oki}, \frac{n_k}{n_{Ok}}\psi_{Okj}\right)\right]\right\} \\ =& E_{s_w}\left\{\left(\frac{M_k}{m_{2k}}\right)^2\sum_{i=1}^{m_{2k}} E_{s_{L_i}}\left[\left(\frac{N_{ki}}{n_{ki}}\right)^2 n_k\left(\frac{n_k}{n_{Ok}}-1\right)\frac{n_k\left[\operatorname{diag}\left(\mathbf{p}_{Lki}\right)-\mathbf{p}_{Lki}\mathbf{p}_{Lki}^T\right]}{n_k-1}\right]\right\} \\ &- E_{s_w}\left\{\left(\frac{M_k}{m_{2k}}\right)^2\sum\sum_{i\neq j}^{m_{2k}} E_{s_{L_i}}\left[\frac{N_{ki}}{n_{ki}}\frac{N_{kj}}{n_{kj}} n_k\left(\frac{n_k}{n_{Ok}}-1\right)\frac{n_k\mathbf{p}_{Lki}\mathbf{p}_{Lkj}^T}{n_k-1}\right]\right\} \end{aligned}$$

and

$$\begin{aligned} V_4 =& E_{s_w}\left\{\left(\frac{M_k}{m_{2k}}\right)^2 E_{s_{L_i}}\left[E_{s_O}\left(V_{s_A}\left(\sum_{i=1}^{m_{2k}}\frac{N_{ki}}{n_{ki}}\frac{n_k}{n_{Ok}}\frac{N_{Om}}{n_{Am}}\psi_{Aki}\right)\right)\right]\right\} \\ =& E_{s_w}\left\{\left(\frac{M_k}{m_{2k}}\right)^2 E_{s_{L_i}}\left[\sum_{i=1}^{m_{2k}} E_{s_O}\left[\left(\frac{N_{ki}}{n_{ki}}\frac{n_k}{n_{Ok}}\right)^2 V_{s_A}\left(\frac{N_{Om}}{n_{Am}}\psi_{Aki}\right)\right]\right.\right. \\ &\left.\left. + \sum\sum_{i\neq j}^{m_{2k}} E_{s_O}\left[\frac{N_{ki}}{n_{ki}}\frac{N_{kj}}{n_{kj}}\left(\frac{n_k}{n_{Ok}}\right)^2 Cov_{s_A}\left(\frac{N_{Om}}{n_{Am}}\psi_{Aki}, \frac{N_{Om}}{n_{Am}}\psi_{Akj}\right)\right]\right]\right\} \\ =& E_{s_w}\left\{\left(\frac{M_k}{m_{2k}}\right)^2\sum_{i=1}^{m_{2k}} E_{s_{L_i}}\left[\left(\frac{N_{ki}}{n_{ki}}\frac{n_k}{n_{Ok}}\right)^2 N_{Om}^2\left(\frac{N_{Om}}{n_{Am}}-1\right)\frac{\left[\operatorname{diag}\left(\mathbf{p}_{Oki}\right)-\mathbf{p}_{Oki}\mathbf{p}_{Oki}^T\right]}{N_{Om}-1}\right]\right\} \\ &- E_{s_w}\left\{\left(\frac{M_k}{m_{2k}}\right)^2\sum\sum_{i\neq j}^{m_{2k}} E_{s_{L_i}}\left[\frac{N_{ki}}{n_{ki}}\frac{N_{kj}}{n_{kj}}\left(\frac{n_k}{n_{Ok}}\right)^2 N_{Om}^2\left(\frac{N_{Om}}{n_{Am}}-1\right)\frac{\mathbf{p}_{Oki}\mathbf{p}_{Okj}^T}{N_{Om}-1}\right]\right\} \end{aligned}$$

where $\mathbf{P}_{ki} = \boldsymbol{\Psi}_{ki}/N_{ki}$, $\mathbf{p}_{Lki} = \psi_{Lki}/n_k$ and $\mathbf{p}_{Oki} = \psi_{Oki}/N_{Om}$.

General principle: The bigger the sample, the smaller the error.

k sample flames with counts n_i, each of area a out of large area A

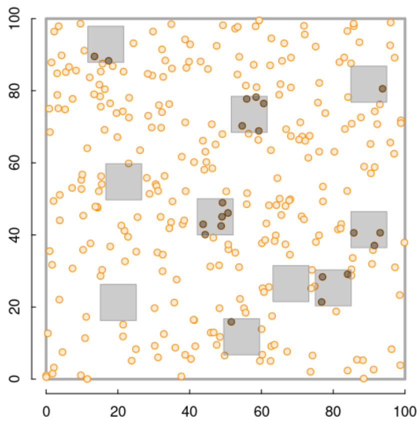
total area sampled is much less than total Area:

$$a_s = \sum_{i=1}^{k} a_i = k \times a \ll A$$

then:

$$\widehat{N} = \frac{A}{a_s} \sum c_i = \frac{A}{k \times a} \sum c_i$$

$$SE(\widehat{N}) = \frac{A}{a_s} \sqrt{\sum n_i} = \frac{A}{a} \frac{\sqrt{\sum n_i}}{k}$$

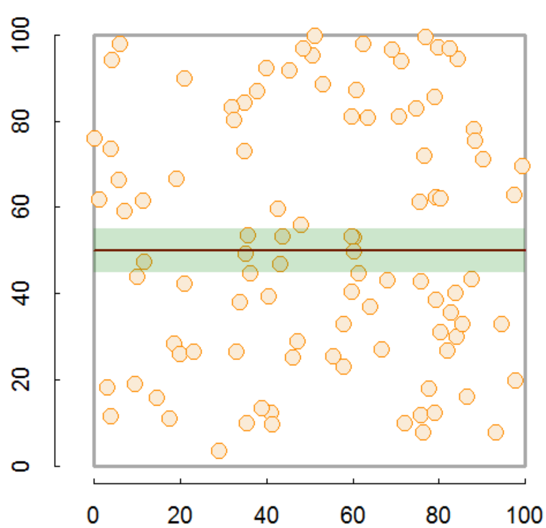


- $\widehat{N} = \frac{100 \times 100}{10 \times 10 \times 10} \times 21 = 210$
- $SE(\widehat{N}) = \frac{100 \times 100}{10 \times 10 \times 10} \sqrt{21} = 45.8$
- $95\%\ C.I. = \widehat{N} \pm 1.96 \times SE(\widehat{N}) = \ldots$
- Coefficient of Variation = $\frac{SE(\widehat{N})}{\widehat{N}} = \ldots$

Example - single transect, simple formula

$$SE(\widehat{D}) = \frac{1}{a}\sqrt{\sum n_i} \text{ and } SE(\widehat{N}) = A \times SE(\widehat{D})$$

$n = 8; a = 1000; A = 10,000$



point estimates

$$\hat{d} = 8/1,000 = .008$$

$$\widehat{N} = \hat{d} \times A = 80$$

standard errors:

$$SE(\widehat{D}) = \frac{\sqrt{8}}{1000} = 0.0028$$

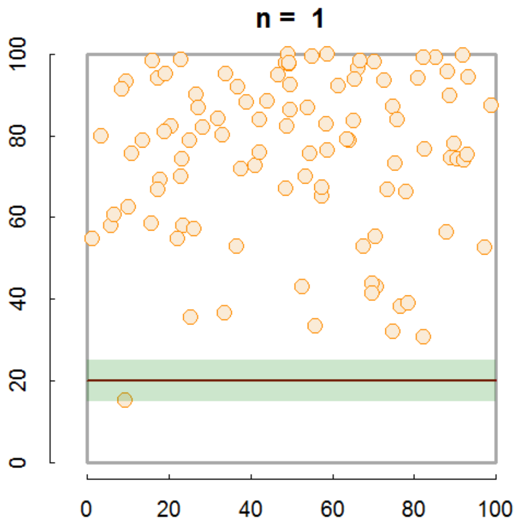
$$SE(\widehat{N}) = 0.0028 \times 10,000 = 28.28$$

final abundance estimate:

$$\widehat{N} = 80$$

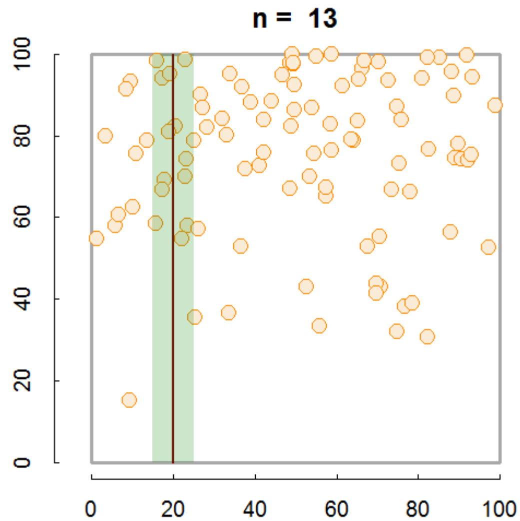
$$95\%\ CI(\widehat{N}) = \widehat{N} \pm 1.96 \times SE(\widehat{N}) = \{24.5, 135\}$$

This is why you *want* lots of transects:



To capture variation!

This is also why you go along the gradient of variation:



gradient - means slope of (steepest) change

More complex formulae

from Fryxell book Chapter 12:

Table 13.3 Estimates and their standard errors for animals counted on transects, quadrats, or sections. The models are described in the text.

Model	Density	Numbers
Simple		
Estimate	$D = \Sigma y / \Sigma a$	$Y = A \times D$
Standard error of estimate (SWR)	$SE(D)_1 = 1/a \times \sqrt{[(\Sigma y^2 - (\Sigma y)^2/n)/(n(n-1))]}$	$SE(Y) = A \times SE(D)_1$
Standard error of estimate (SWOR)	$SE(D)_2 = SE(D)_1 \times \sqrt{[1 - (\Sigma a)/A]}$	$SE(Y) = A \times SE(D)_2$
Ratio		
Estimate	$D = \Sigma y / \Sigma a$	$Y = A \times D$
Standard error of estimate (SWR)	$SE(D)_3 = n/\Sigma a \times \sqrt{[(1/n(n-1))(\Sigma y^2 + D^2 \Sigma a^2 - 2D\Sigma ay)]}$	$SE(Y) = A \times SE(D)_3$
Standard error of estimate (SWOR)	$SE(D)_4 = SE(D)_3 \times \sqrt{[1 - (\Sigma a)/A]}$	$SE(Y) = A \times SE(D)_4$
PPS		
Estimate	$d = 1/n \times \Sigma(y/a)$	$Y = A \times d$
Standard error of estimate (SWR)	$SE(D) = \sqrt{[(\Sigma(y/a)^2 - (\Sigma(y/a))^2/n)/(n(n-1))]}$	$SE(Y) = A \times SE(d)$

SWR, sampling with replacement; SWOR, sampling without replacement. Notation is given in Section 13.5.1.

These are used when **sampling areas** are unequal, and account for differences when sampling **with replacement** or **without replacement**.

Simple-SWR

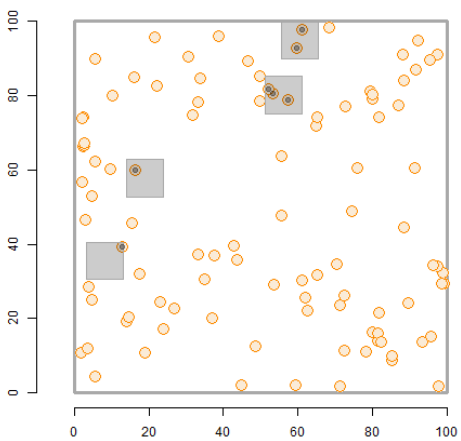
- **Simple:** Equal sized sampling frames a_i all equal
- **SWR:** Sampling 'with replacement', i.e. frames *OVERLAP*; some individuals counted more than once..

$$SE(\widehat{D}) = \frac{1}{a_i\sqrt{k(k-1)}} \times \sqrt{\sum n_i^2 - \left(\sum n_i\right)^2 / k}$$

$$SE(\widehat{N}) = A \times SE(\widehat{D})$$

variable	meaning	in book
k	number of units sampled	n
a_i	the area of a *single* unit	a
n_i	an individual sample count	y
A	total study area	

Example:



- $\widehat{N} = \frac{2+3+1+1}{100} \times 10,000 = 70$
- $SE(\widehat{N}) = \frac{10,000}{100\times\sqrt{4\times 3}} \times \sqrt{(1+1+9+4) - \frac{(1+1+3+2)^2}{4}}$
- $SE(\widehat{N}) = \frac{50}{\sqrt{3}} \times \sqrt{15 - \frac{49}{4}} = 48$
- $\widehat{N} = 70;\ \ 95\%\text{CI} = (-23, 164)$

Anything wrong with this confidence interval?

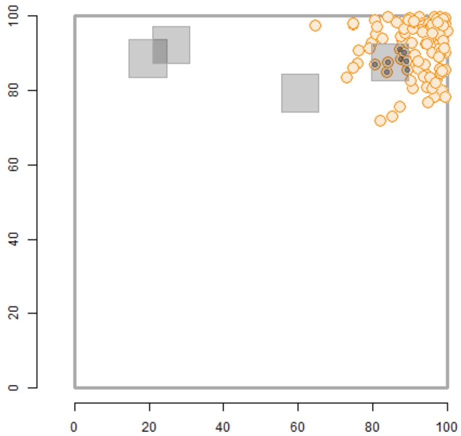
data: counts = {2,3,1,1}

a = 100; A = 10,000

Example: More Heterogeneity

data:

- counts = {0,0,0,8}
- a = 100; A = 10,000



$\widehat{N} = 80$

`

$SE(\widehat{N}) =$

$\frac{50}{\sqrt{3}} \times \sqrt{(0+0+0+8)^2 - \frac{(0+0+0+8^2)}{4}} =$

$= \frac{50}{\sqrt{3}} \times \sqrt{48} = \frac{50}{\sqrt{3}} 4\sqrt{3} = 200$

`

$$95\%\text{CI} = (-130, 470)$$

Enormous confidence intervals, because of enormous variability in samples!

Simple - SWOR

- **SWOR:** Sampling *without* replacement, i.e. design guarantees no individual is counted more than once.

$$SE(\widehat{D_{swor}}) = SE(\widehat{D_{swr}}) \times \sqrt{1 - a/A}$$

The larger the proportion sampled (*coverage*) - the smaller the **sampling error.**

Ratio (SWR/SWOR)

Ratio: unequal sample frames (e.g. both hula hoops and meter squares).

- $\widehat{D} = \sum n_i / \sum y_i$ (same as before)
- Standard errors: more complicated … see formulae.

Take-aways

- Is it **very important** to quantify uncertainty! But also, can be **hard** (and **disheartening**).
- Larger samples & higher coverage $\rightarrow$ smaller errors $\rightarrow$ narrower confidence intervals $\rightarrow$ more precision.
- The error estimates take into account **sample randomness**, but also **heterogeneity**. The more **heterogeneous** the distribution the larger the errors.
- Which is why … you take that heterogeneity into in your **estimates!**

One more formula … for combining estimates

If you have multiple sub-count estimates (e.g. one for each of r sub-region):

- $\widehat{N}_1, \widehat{N}_2, \ldots, \widehat{N}_r,$

and each estimate has a standard error:

- $SE(\widehat{N}_1), SE(\widehat{N}_2), \ldots, SE(\widehat{N}_r)$

Then …

… the **total** estimate will be:

$$\widehat{N} = \sum_{i=1}^{r} \widehat{N}_i$$

and the standard error will be:

$$SE(\widehat{N}) = \sqrt{\sum_{i=1}^{r} SE(\widehat{N}_i)^2}$$

Will the estimate be more precise? **You get to test this out in the field!**