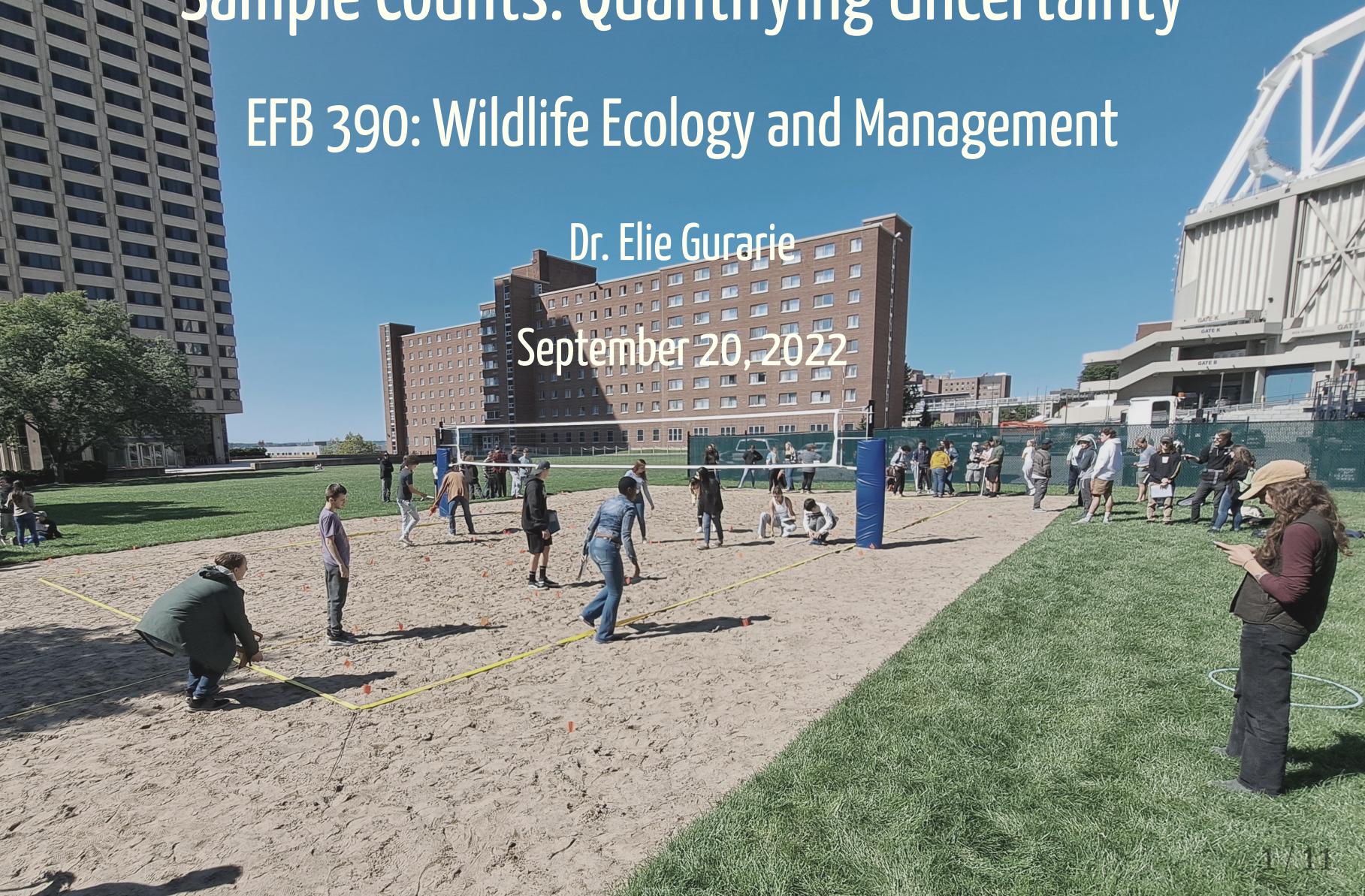


# Sample Counts: Quantifying Uncertainty

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

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

The variance of the estimator can be written as

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

Component-wise,

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

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

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

and

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

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

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

1. If  $a_s \ll A$  (i.e. low sampling intensity)

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

**remember:**  $n_s = \sum n_i$  is the total sample count

in our example:  $SE = 100^2 / (10 \times 10^2) \sqrt{30} = 54.8$

# 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:**

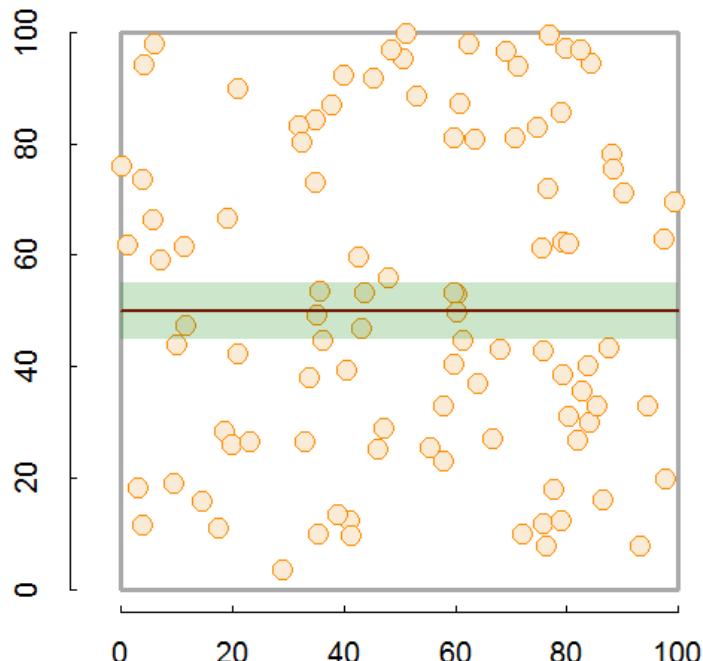
$$SE(\widehat{D}) = \frac{\sqrt{8}}{1000} = 0.0028$$

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

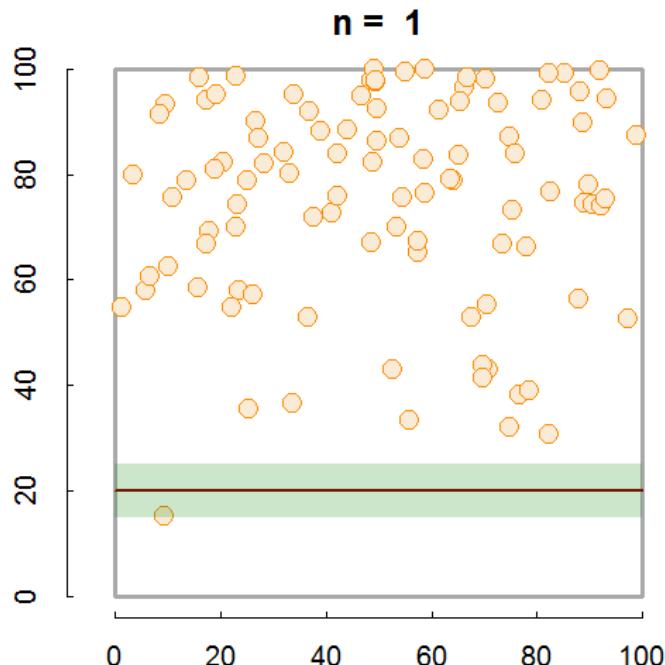
**final abundance estimate:**

$$\widehat{N} = 80$$

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

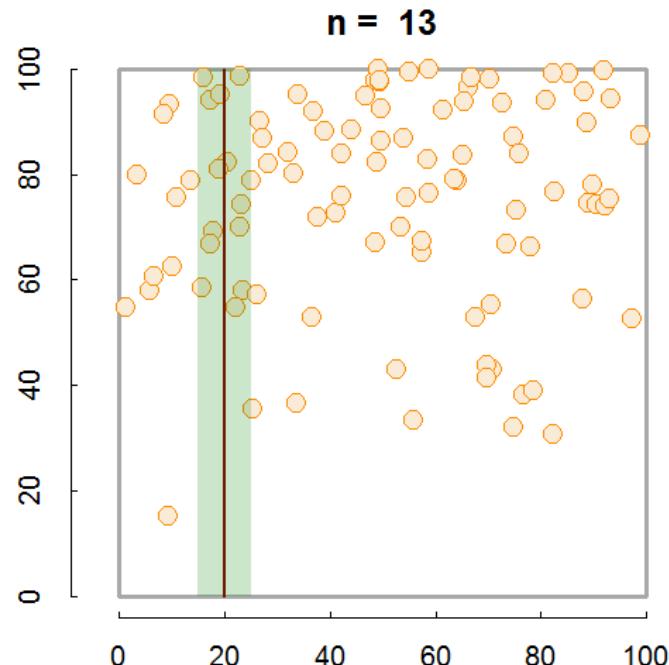


This is why you *want* lots of transects:



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

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

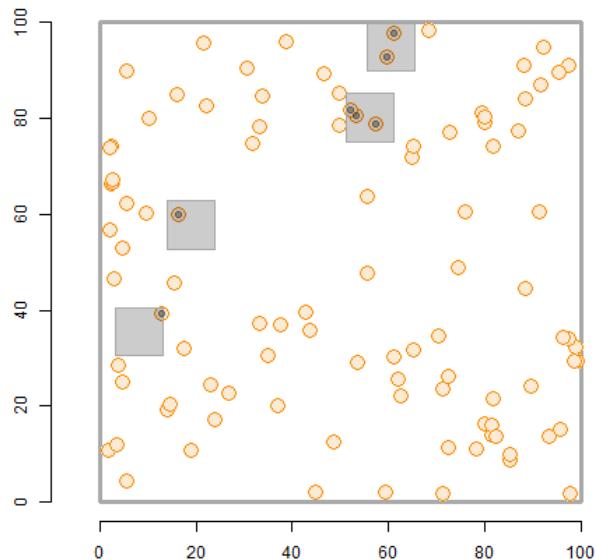
- **Simple:** Equal sized sampling frames  $a_i$  all equal
- **SWR:** Sampling 'with replacement', i.e. frames *OVERLAP*; some individuals counted more than once..

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

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

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

# Example:



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

a = 100; A = 10,000

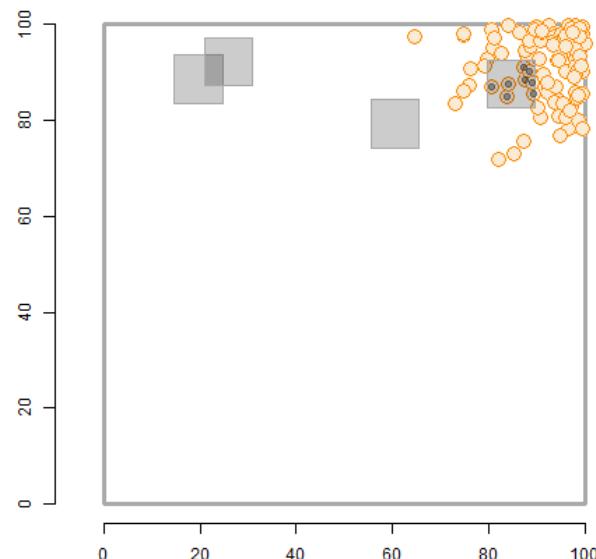
- $\widehat{N} = \frac{2+3+1+1}{100} \times 10,000 = 70$
- $SE(\widehat{N}) = \frac{10,000}{100 \times \sqrt{4 \times 3}} \times \sqrt{(1 + 1 + 9 + 4) - \frac{(1+1+3+2)^2}{4}} = 48$
- $SE(\widehat{N}) = \frac{50}{\sqrt{3}} \times \sqrt{15 - \frac{49}{4}} = 48$
- $\widehat{N} = 70; 95\%CI = (-23, 164)$

*Anything wrong with this confidence interval?*

# Example: More Heterogeneity

**data:**

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



$$\widehat{N} = 80$$

$$SE(\widehat{N}) =$$

$$\frac{50}{\sqrt{3}} \times \sqrt{(0 + 0 + 0 + 8)^2 - \frac{(0+0+0+8^2)}{4}} = \\ = \frac{50}{\sqrt{3}} \times \sqrt{48} = \frac{50}{\sqrt{3}} 4\sqrt{3} = 200$$

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

**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 → smaller errors → narrower confidence intervals → more precision.

The error estimates take into account **sample randomness**, but also **heterogeneity**. The more **heterogeneous** the distribution the larger the errors

- unless (*spoiler alert*) you take that into account in your **distribution modelling** ...