# Sampling: Need and designs

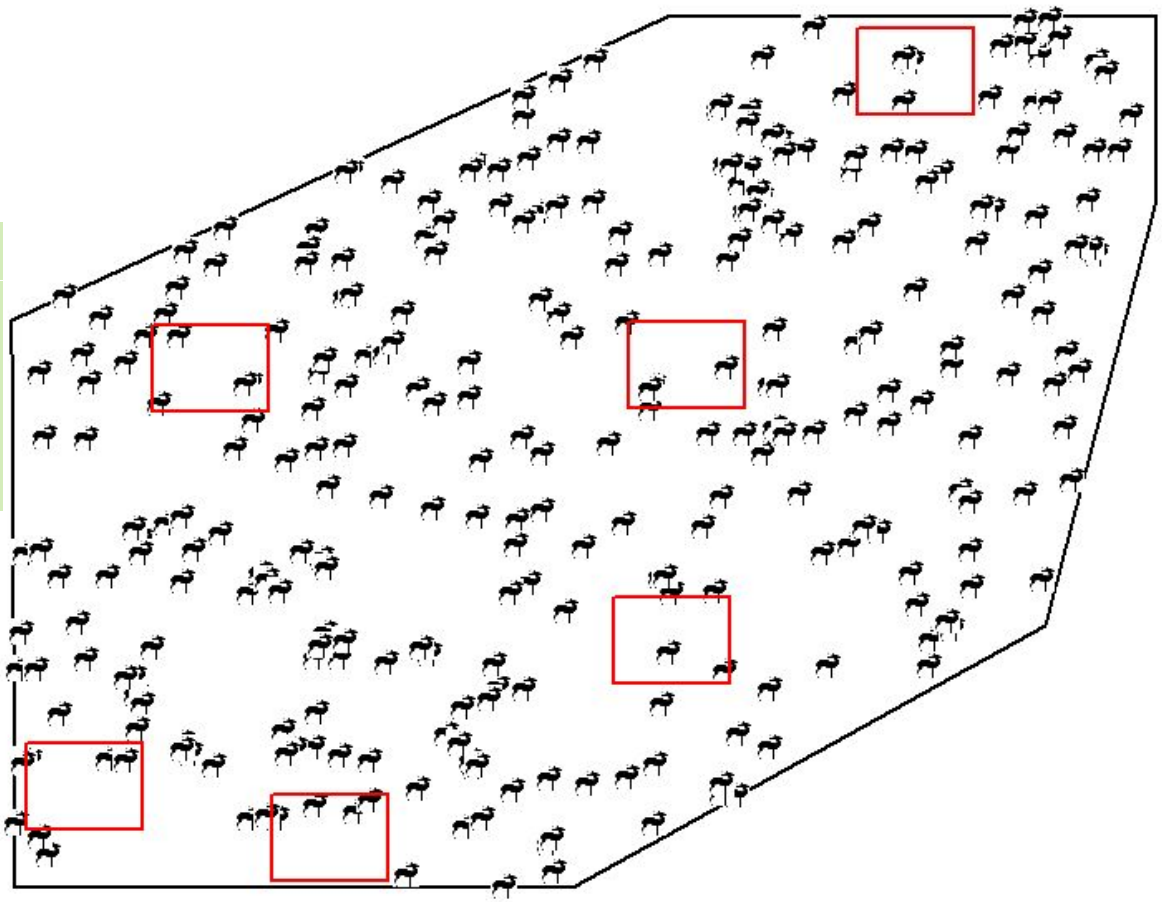
Koustubh Sharma & Justine Shanti Alexander

gef Snow Leopard Trust SNOW LEOPARD NETWORK SNOW LEOPARD The Rufford Small Grants Foundation www.ruffordsmallgrants.org UNDP Empowered lives. Resilient nations.

# Distribution of object of interest

- Random

$a = area\ of\ unit$
$A = Total\ Area$
$c = count$
$N = Abundance$
$D = Density$

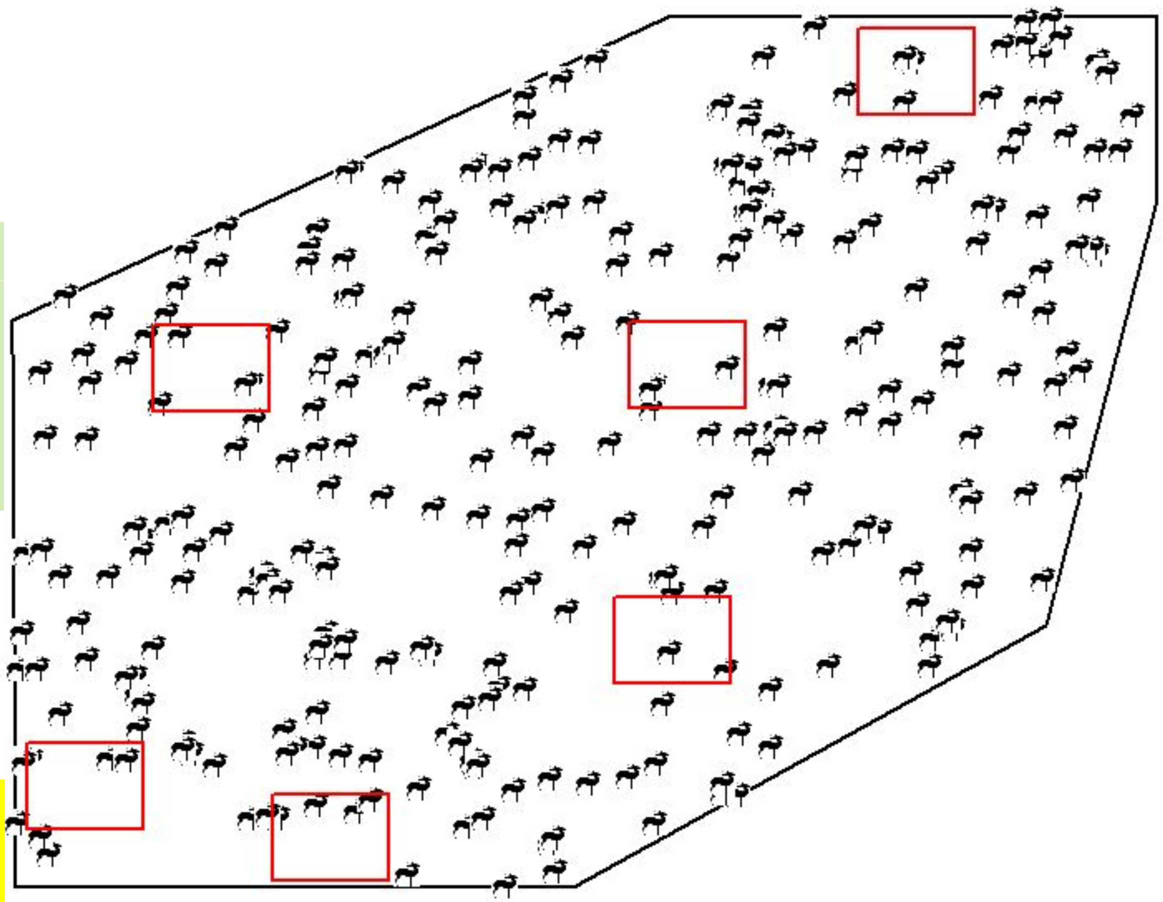


$D = c/a$
$N = D.A$

# Distribution of object of interest

- Random

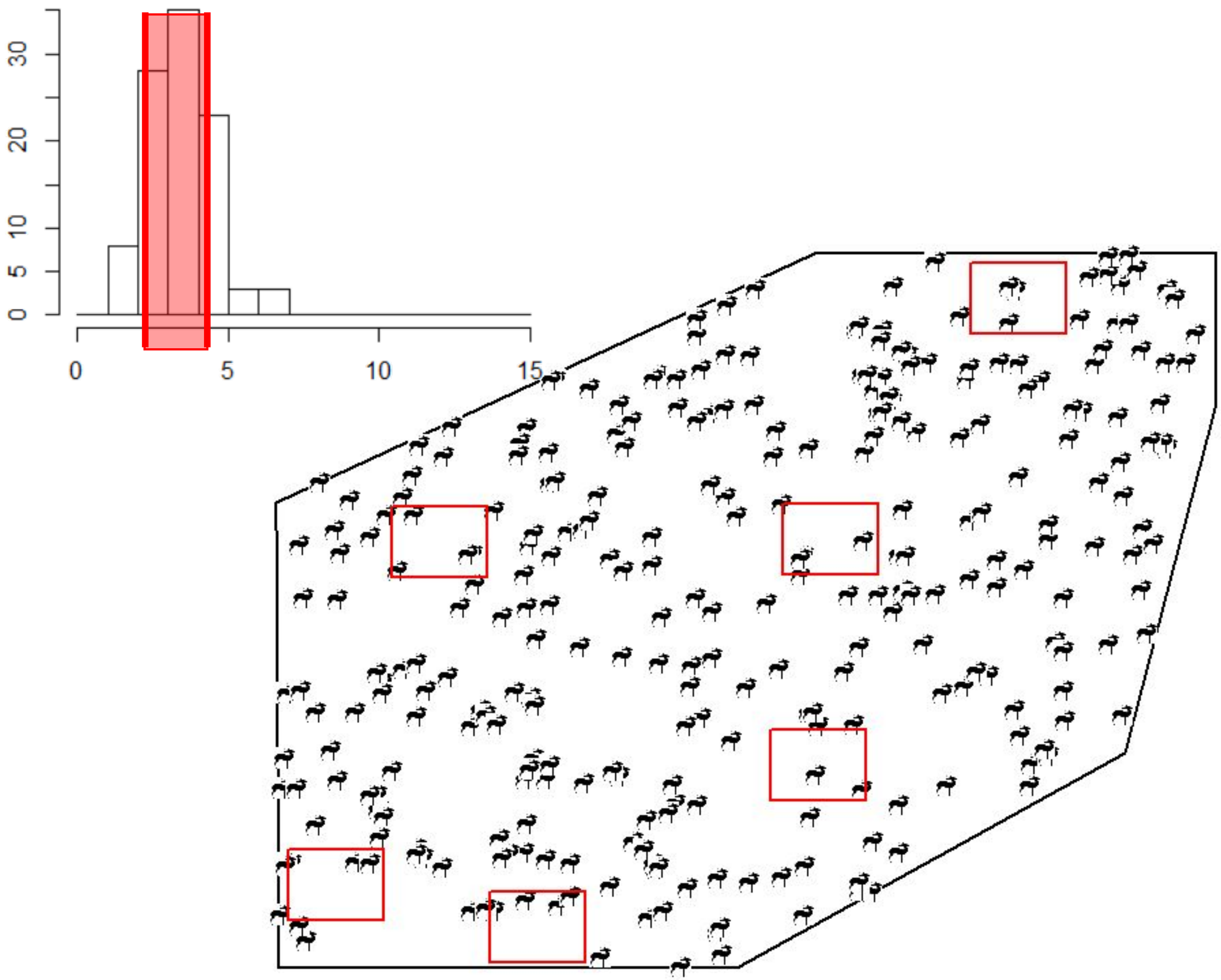
$a = area\ of\ unit$
$A = Total\ Area$
$c = count$
$N = Abundance$
$D = Density$



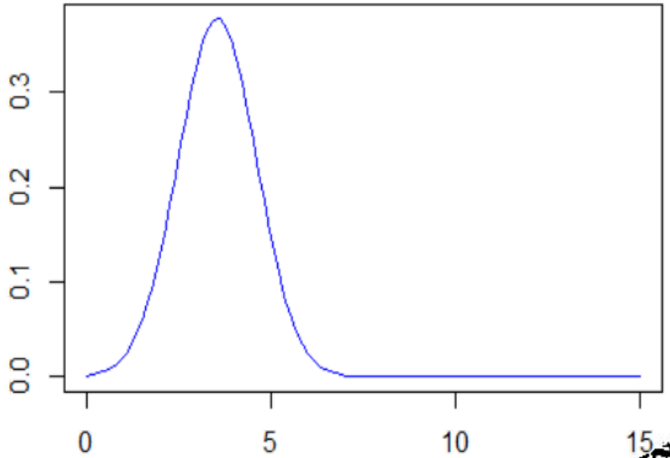
| Unit | Count |
|---|---|
| A | 4 |
| B | 4 |
| C | 3 |
| D | 3 |
| E | 2 |
| F | 5 |
| **Mean** | **3.5** |
| **StDev** | **1.05** |
| **CL** | **2.7-4.3** |

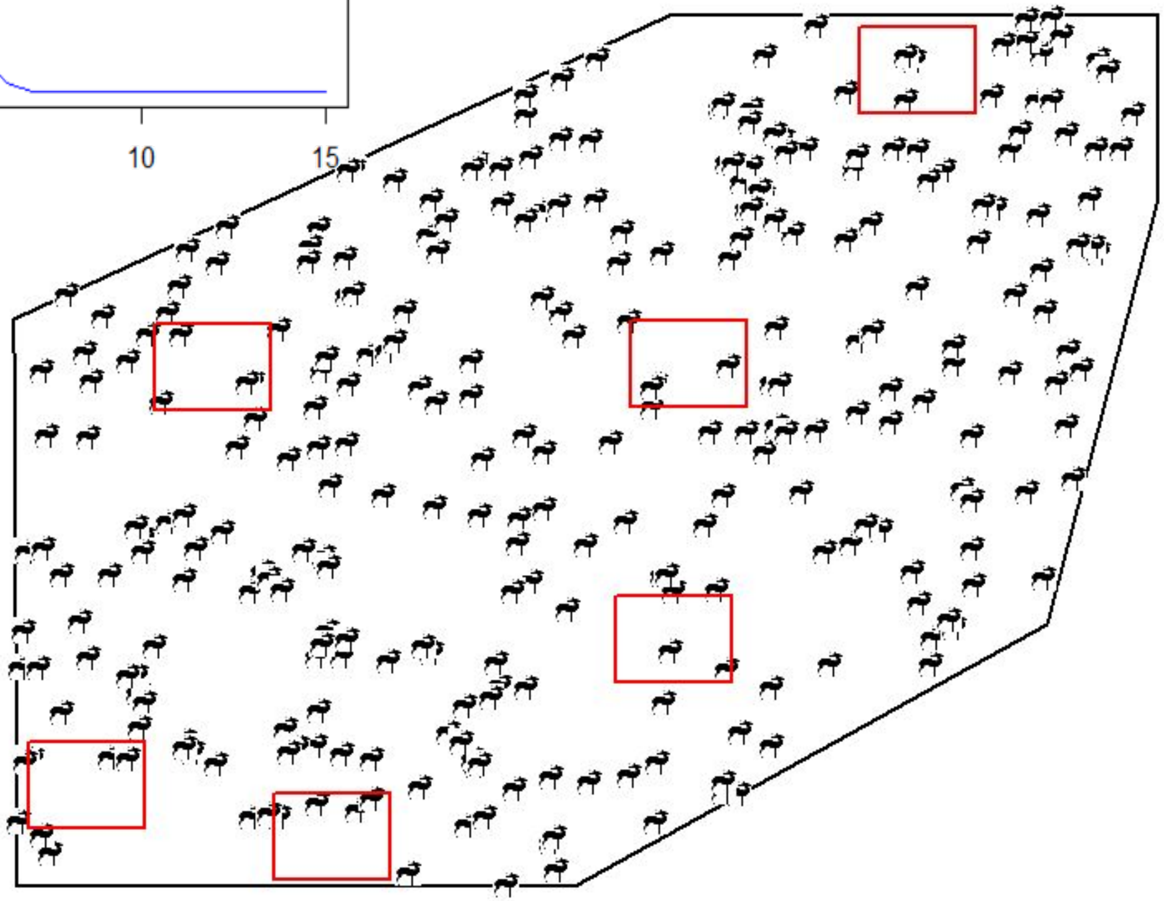
$D = c/a$
$N = D.A$

$\widehat{D} = 35 \pm x$
$\widehat{N} = 350 \pm x$




| Unit | Count |
|---|---|
| A | 4 |
| B | 4 |
| C | 3 |
| D | 3 |
| E | 2 |
| F | 5 |
| **Mean** | **3.5** |
| **StDev** | **1.05** |
| **CL** | **2.7-4.3** |

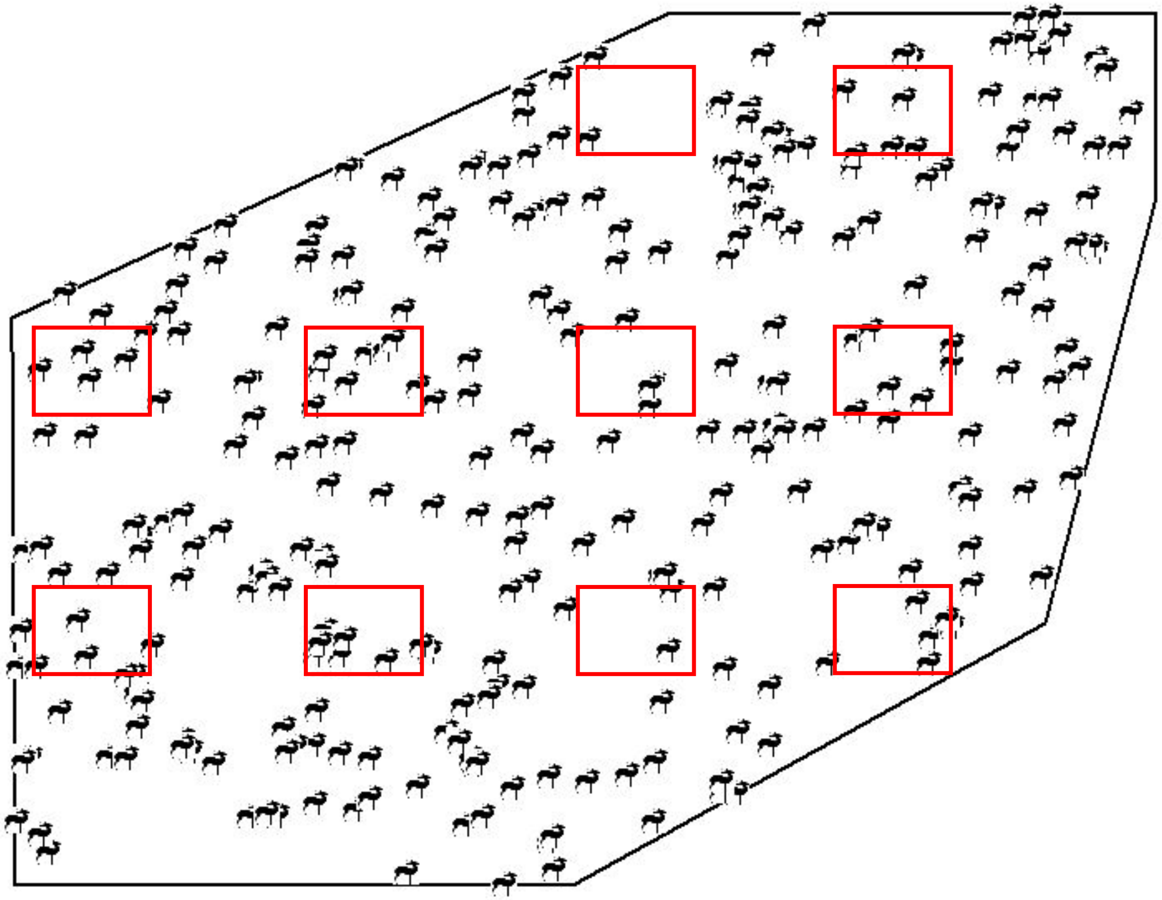


| Unit | Count |
|---|---|
| A | 4 |
| B | 4 |
| C | 3 |
| D | 3 |
| E | 2 |
| F | 5 |
| **Mean** | **3.5** |
| **StDev** | **1.05** |
| **CL** | **2.7-4.3** |

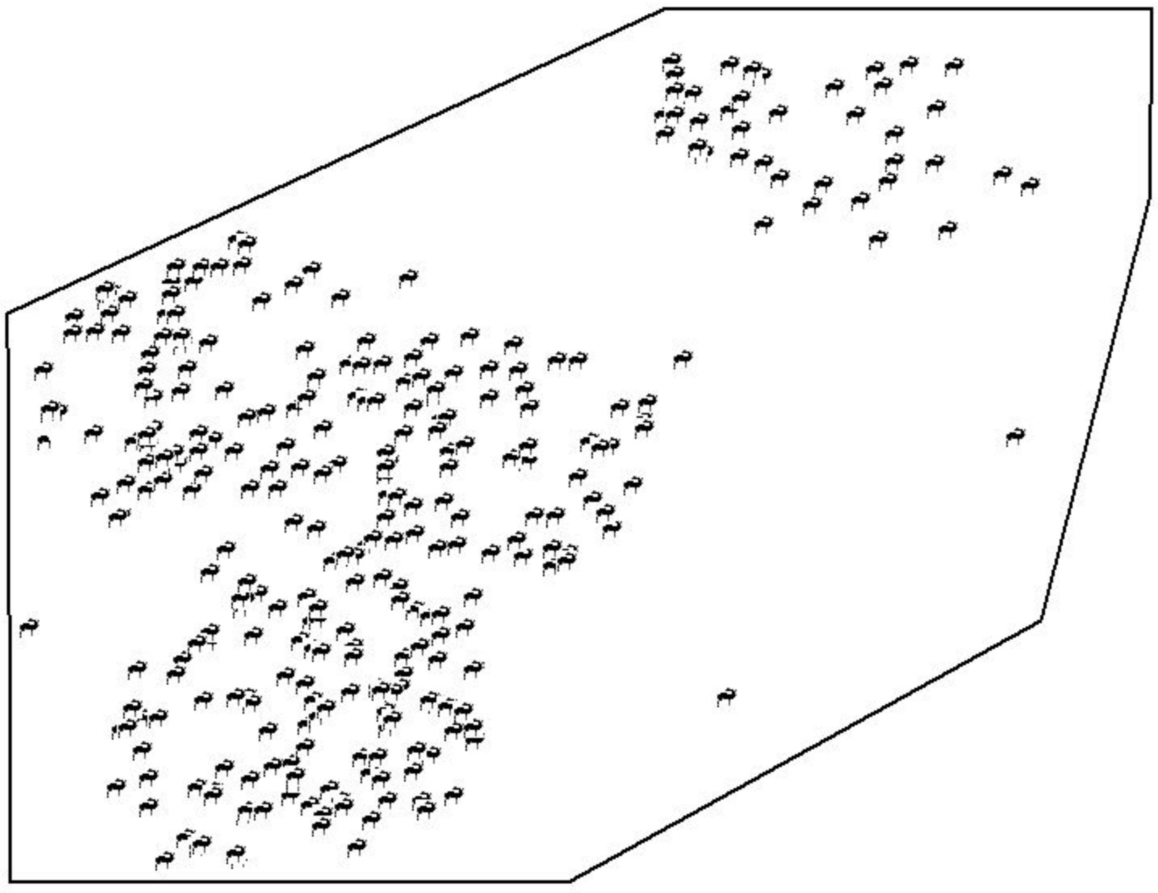
# Uniform sampling

- Similar to random



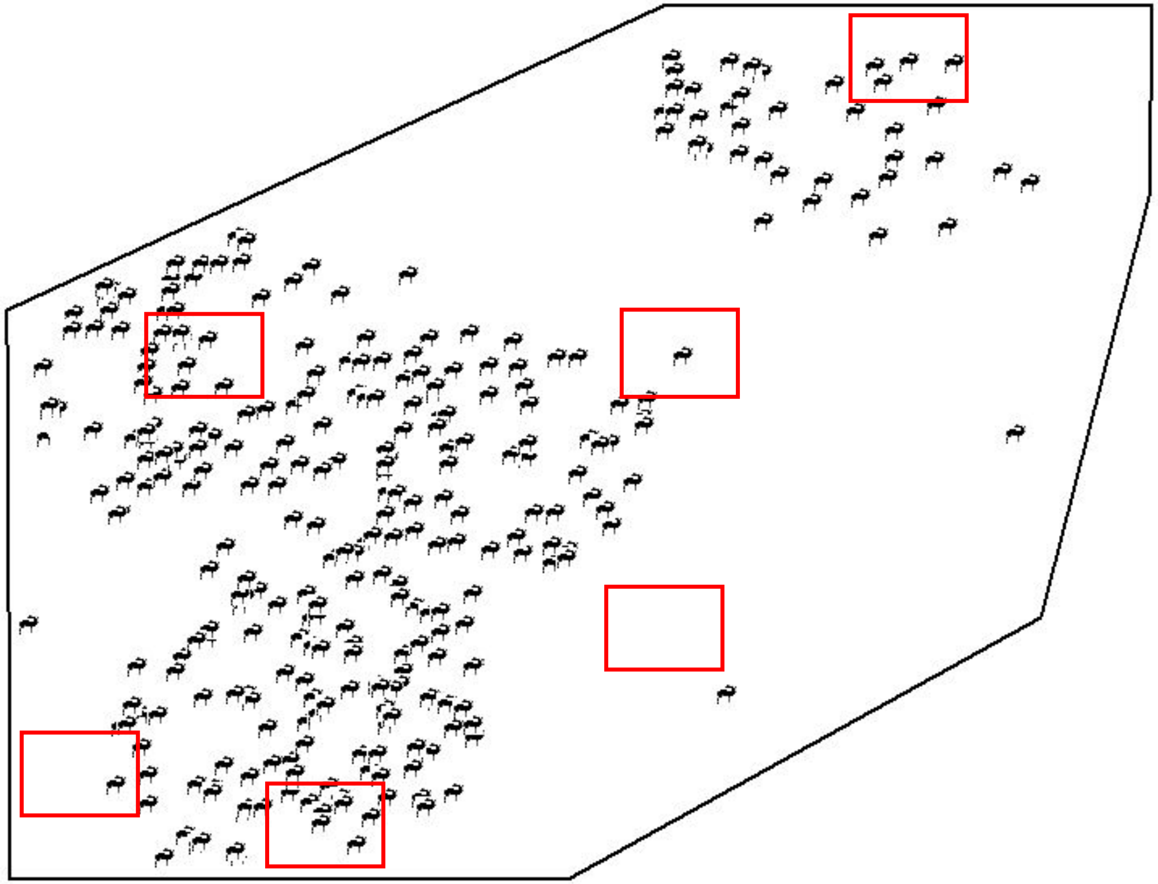
# Distribution of object of interest

- Uneven



# Distribution of object of interest

- Uneven



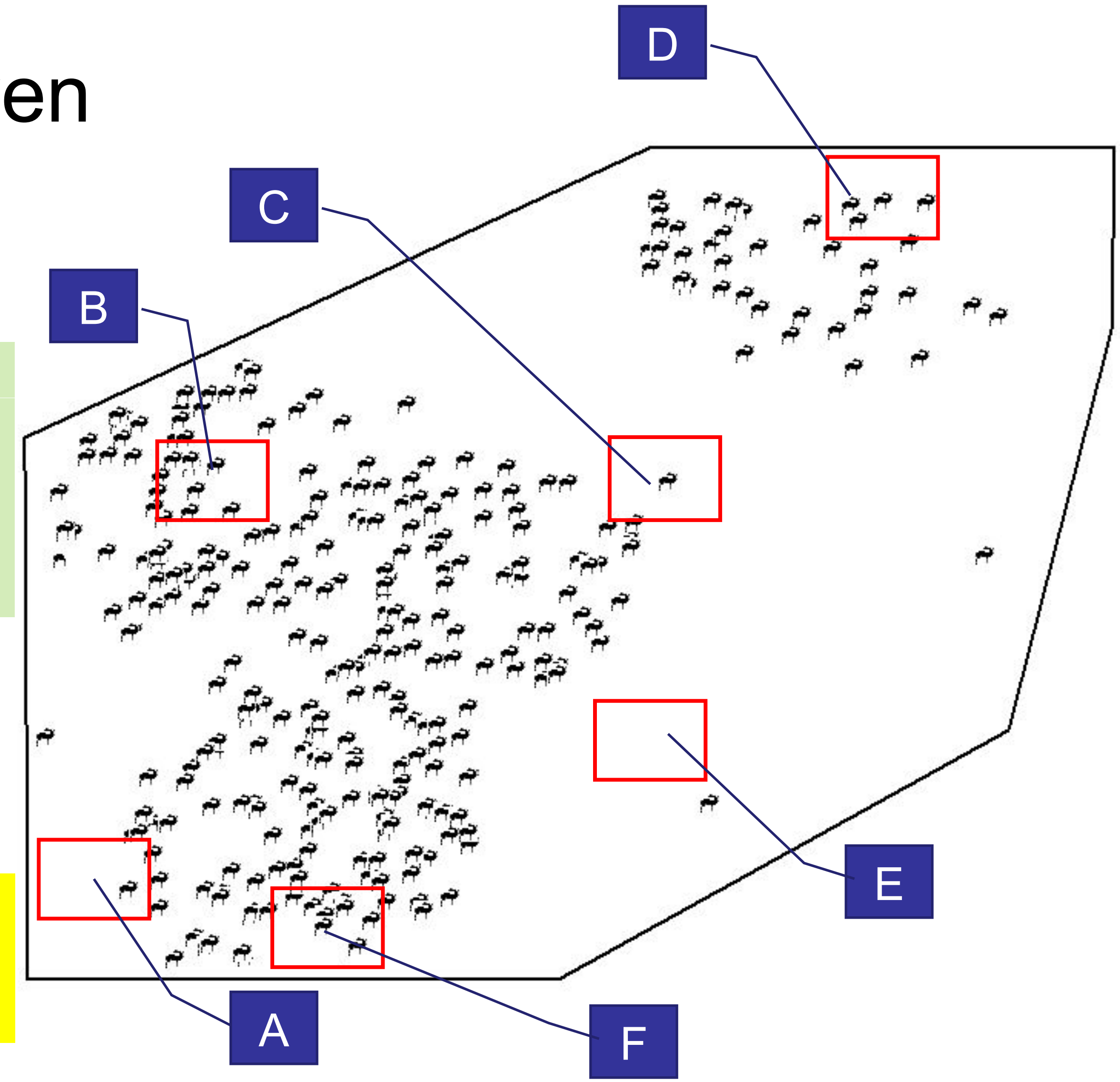
# Distribution of object of interest

- Uneven

$a = area\ of\ unit$
$A = Total\ Area$
$c = count$
$N = Abundance$
$D = Density$

$D = c/a$
$N = D.A$

$D = \_\_?$
$N = \_\_?$




| Unit | Count |
|---|---|
| A | |
| B | |
| C | |
| D | |
| E | |
| F | |
| **Mean** | |
| **StDev** | |
| **CL** | |

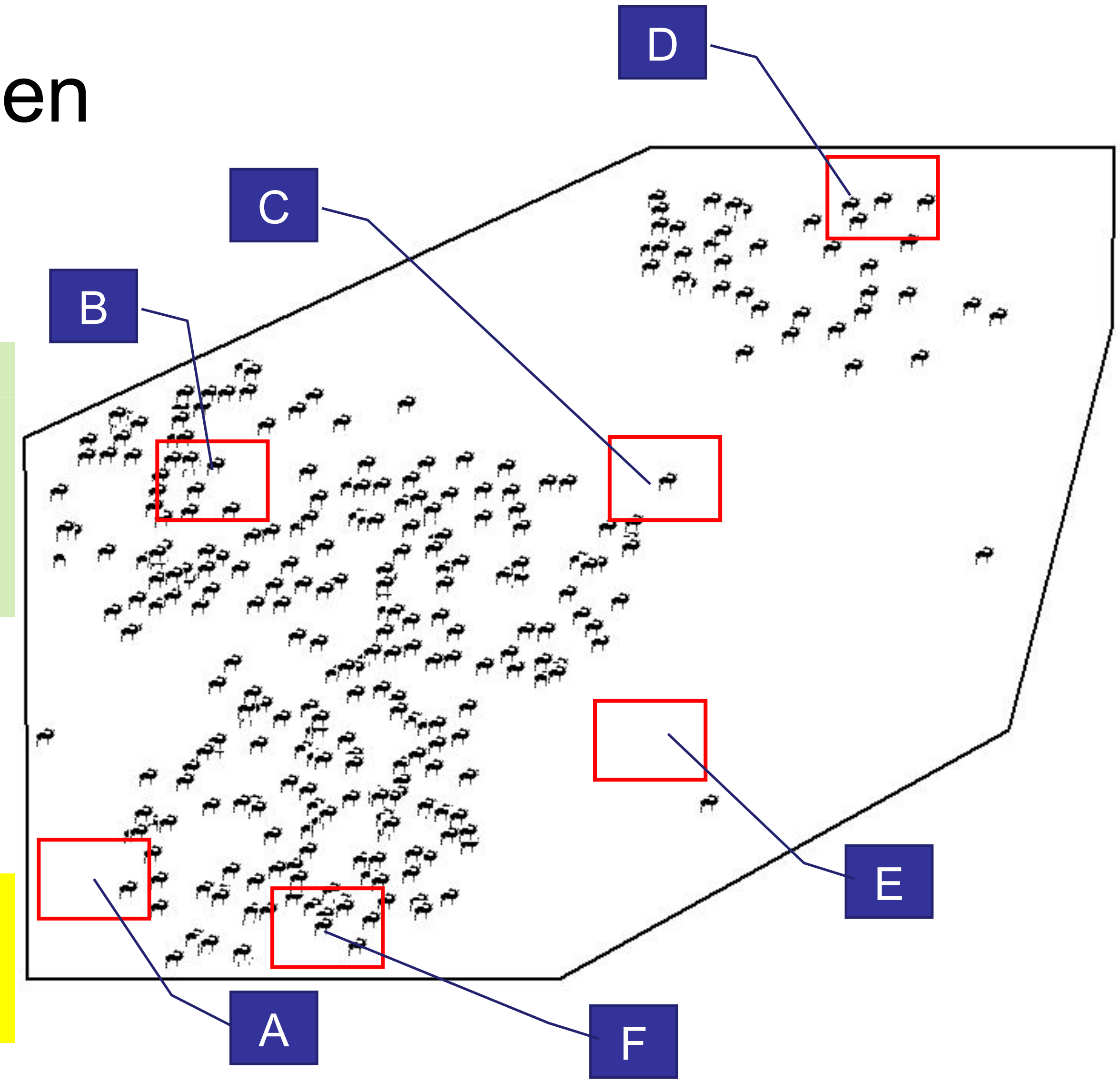
# Distribution of object of interest

- Uneven

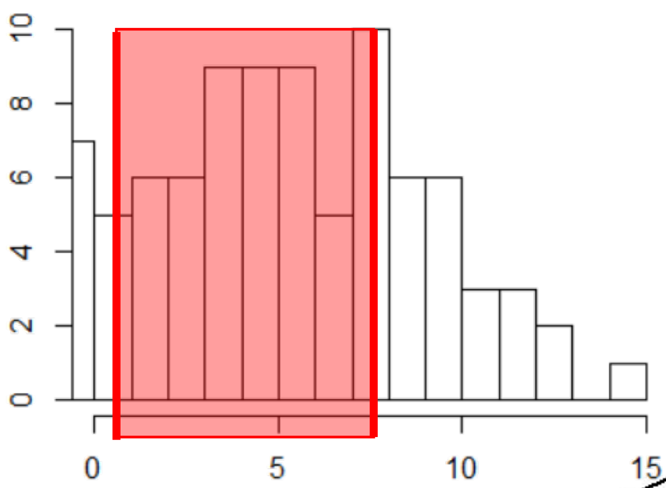
$a = area\ of\ unit$
$A = Total\ Area$
$c = count$
$N = Abundance$
$D = Density$

$D = c/a$
$N = D.A$

$D = \_\_?$
$N = \_\_?$




| Unit | Count |
|---|---|
| A | 1 |
| B | 9 |
| C | 1 |
| D | 4 |
| E | 0 |
| F | 10 |
| **Mean** | |
| **StDev** | |
| **CL** | |




$a = area\ of\ unit$

$A = Total\ Area$

$c = count$

$N = Abundance$

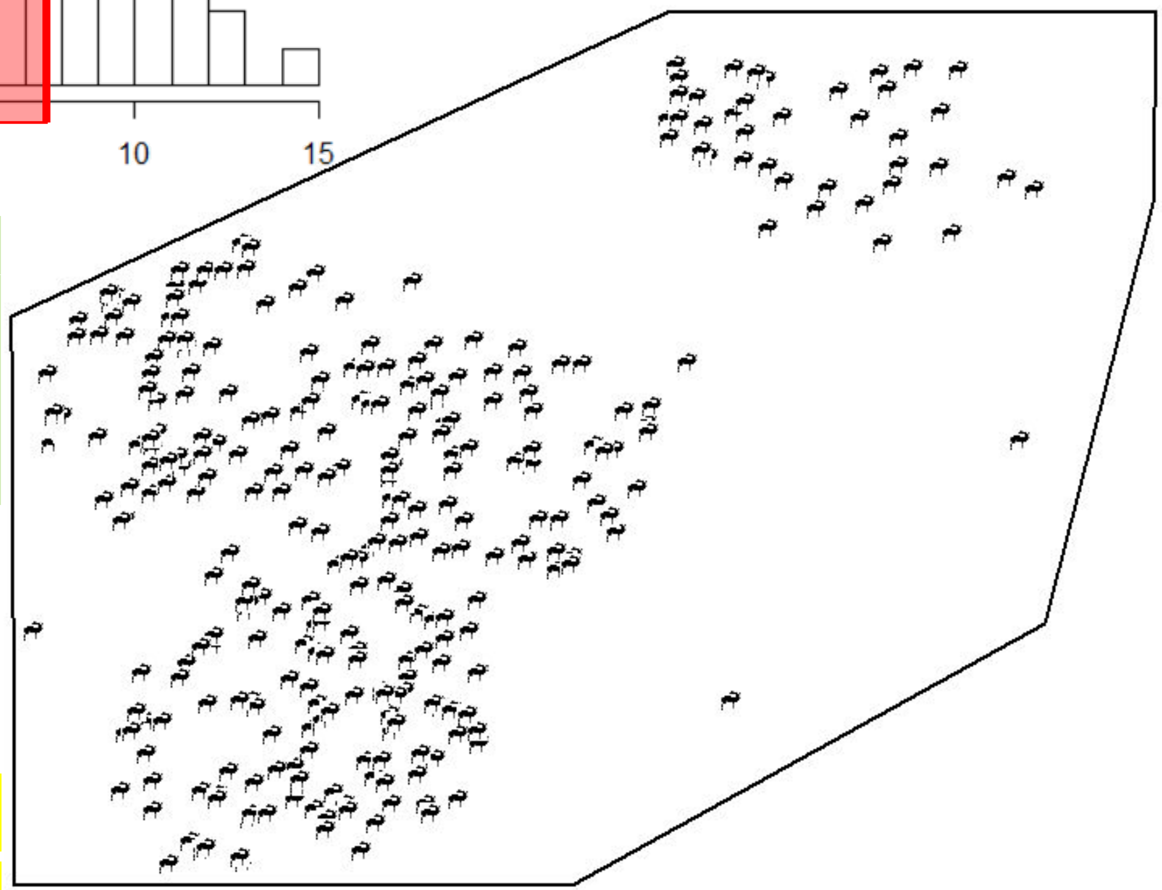
$D = Density$

$D = c/a$

$N = D.A$

$D = 42 \pm x$

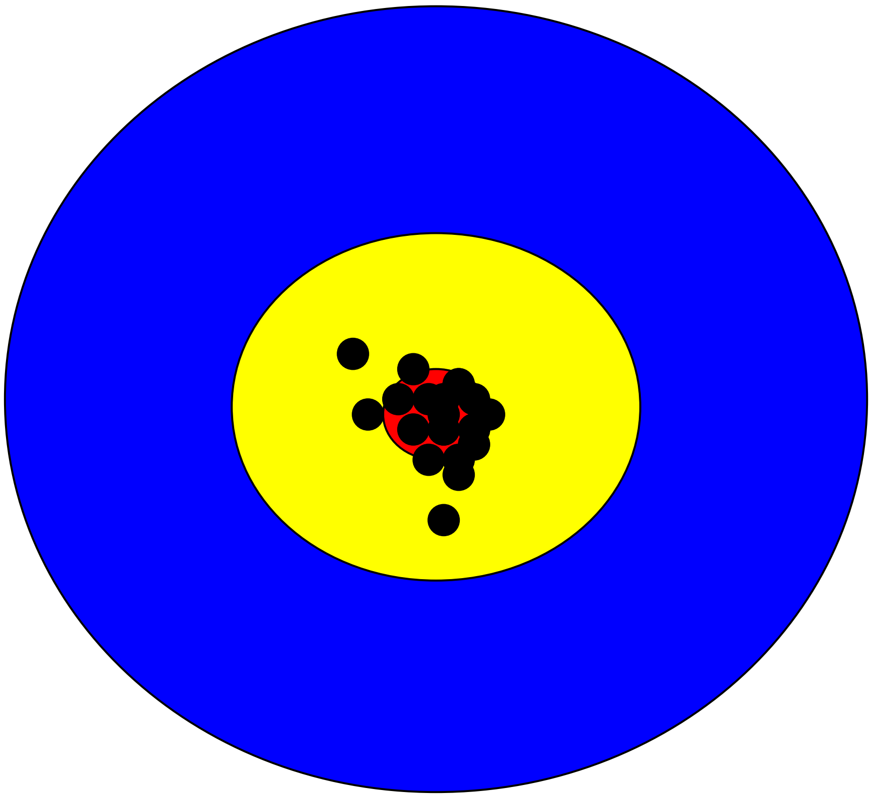
$N = 420 \pm x$

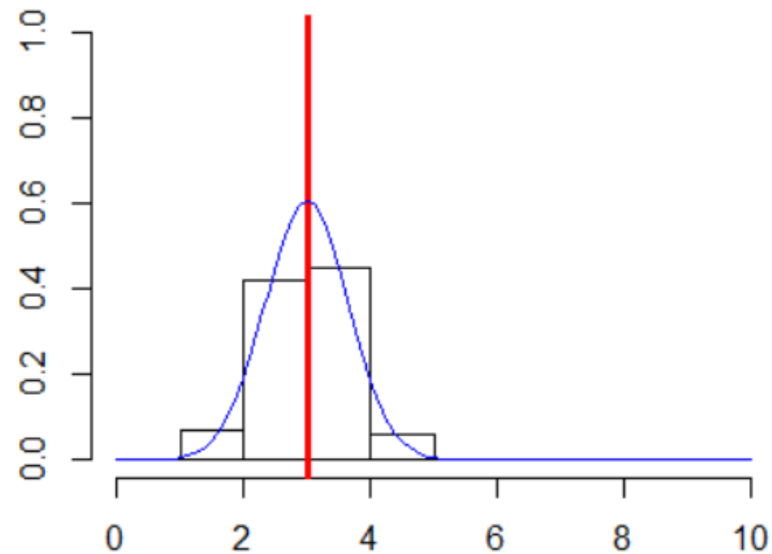


| Unit | Count |
|---|---|
| A | 1 |
| B | 9 |
| C | 1 |
| D | 4 |
| E | 0 |
| F | 10 |
| **Mean** | **4.2** |
| **StDev** | **4.36** |
| **CL** | **0.7-7.6** |

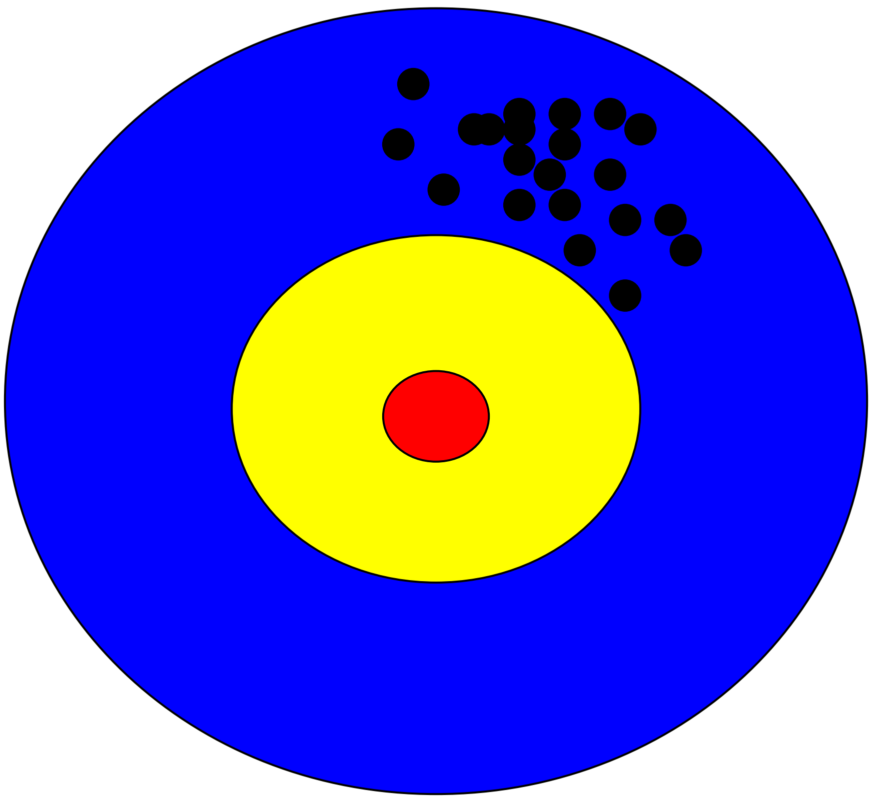
# How good is my estimate?

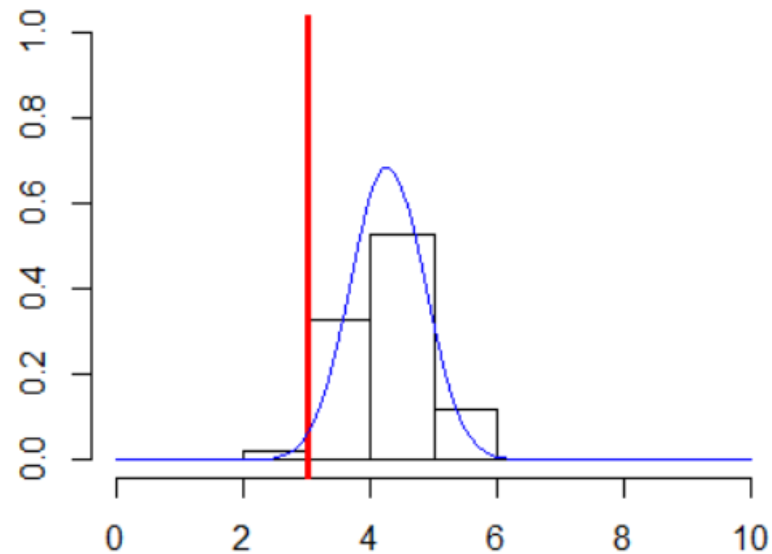
# Accurate and Precise



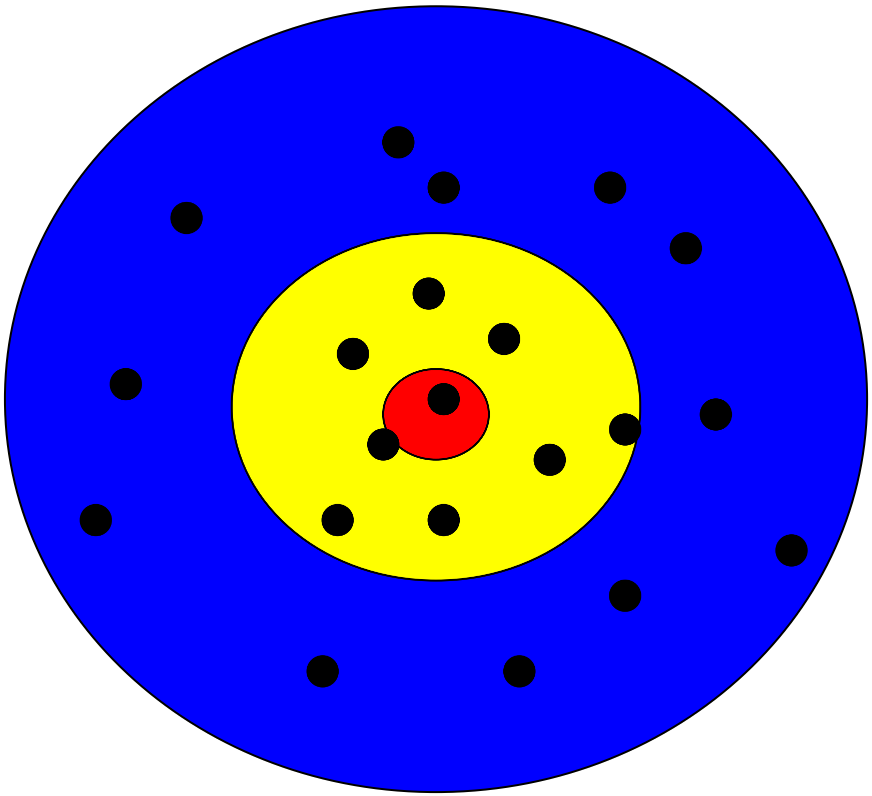


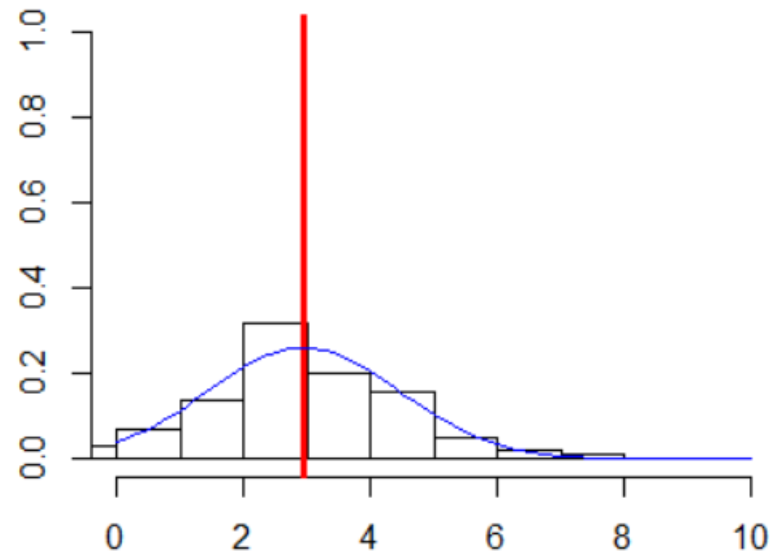
# Inacurate but Precise



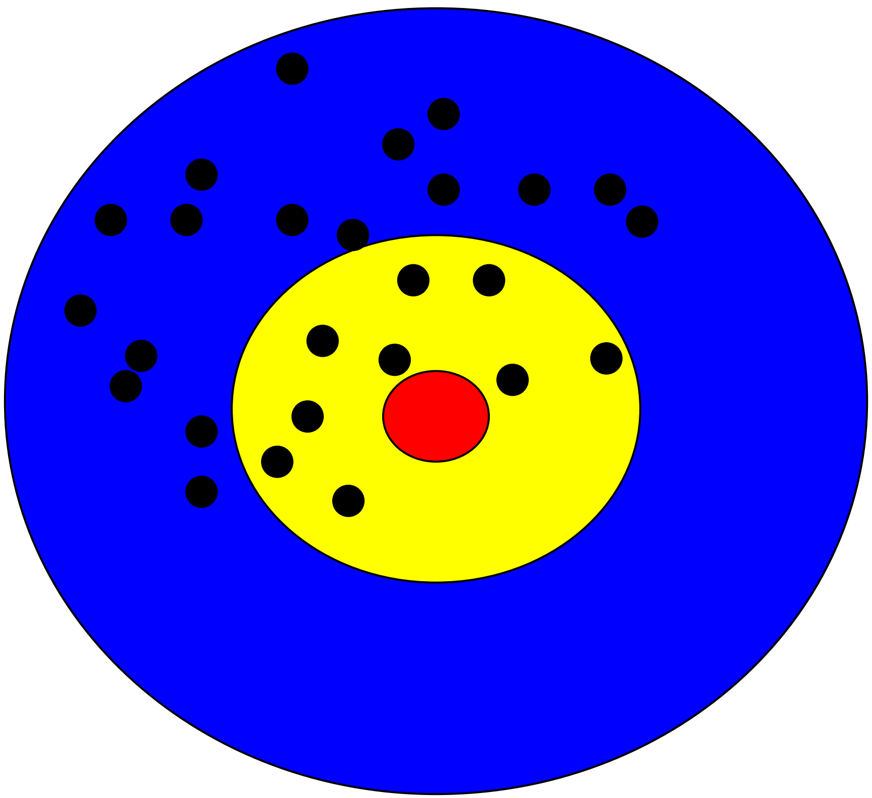


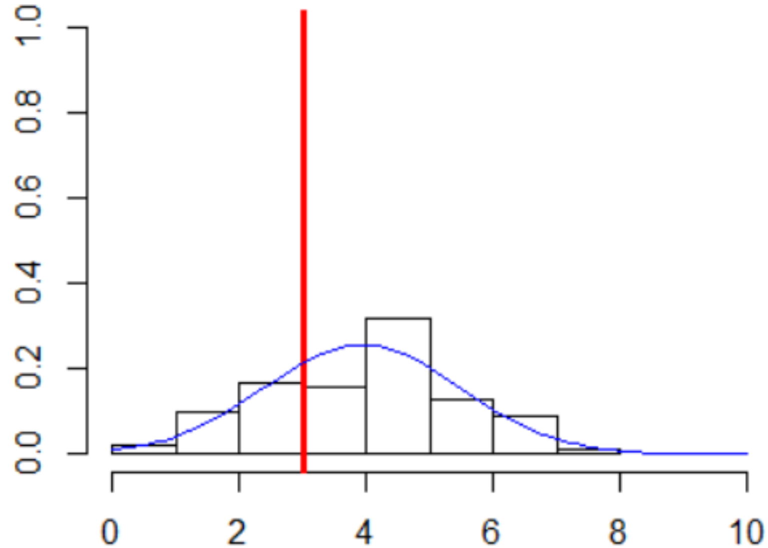
# Accurate but Imprecise





# Inaccurate and Imprecise





# How do we report error?

- Variance $$\frac{\sum(x_i - mean)^2}{(n-1)}$$
- Standard Deviation $$\sqrt{\frac{\sum(x_i - mean)^2}{(n-1)}}$$
- Coefficient of Variation $$\frac{Standard\ deviation}{Mean}$$
- 95% Confidence Interval $$x \pm 1.96\left(\frac{Standard\ deviation}{\sqrt{n}}\right)$$

# How to reduce Error

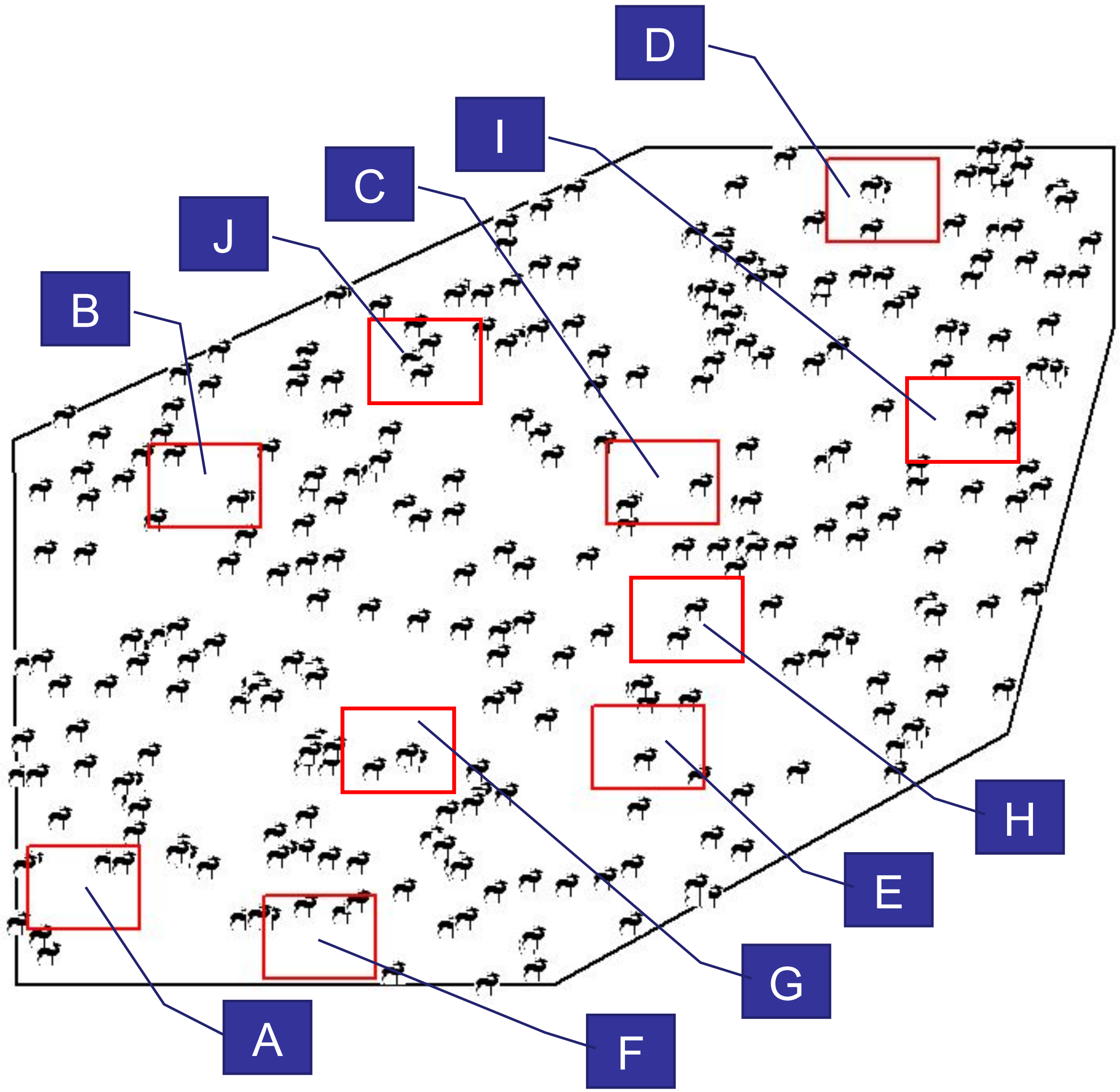
- Increase number of samples (n)
  - Random (or uniform)
    - Needs more resources
  - Stratify
    - Needs ecological information

- Model
    - Needs a good understanding of covariates

# Increase samples

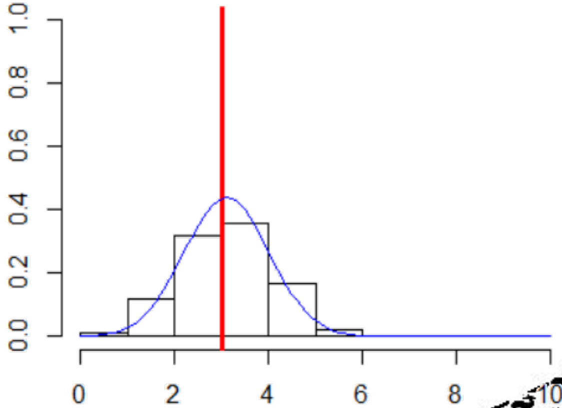
- Variance $$\frac{\sum(x_i - mean)^2}{(n-1)}$$
- Standard Deviation $$\sqrt{\frac{\sum(x_i - mean)^2}{(n-1)}}$$
- Coefficient of Variation $$\frac{Standard\ deviation}{Mean}$$
- 95% Confidence Interval $$x \pm 1.96\left(\frac{Standard\ deviation}{\sqrt{n}}\right)$$

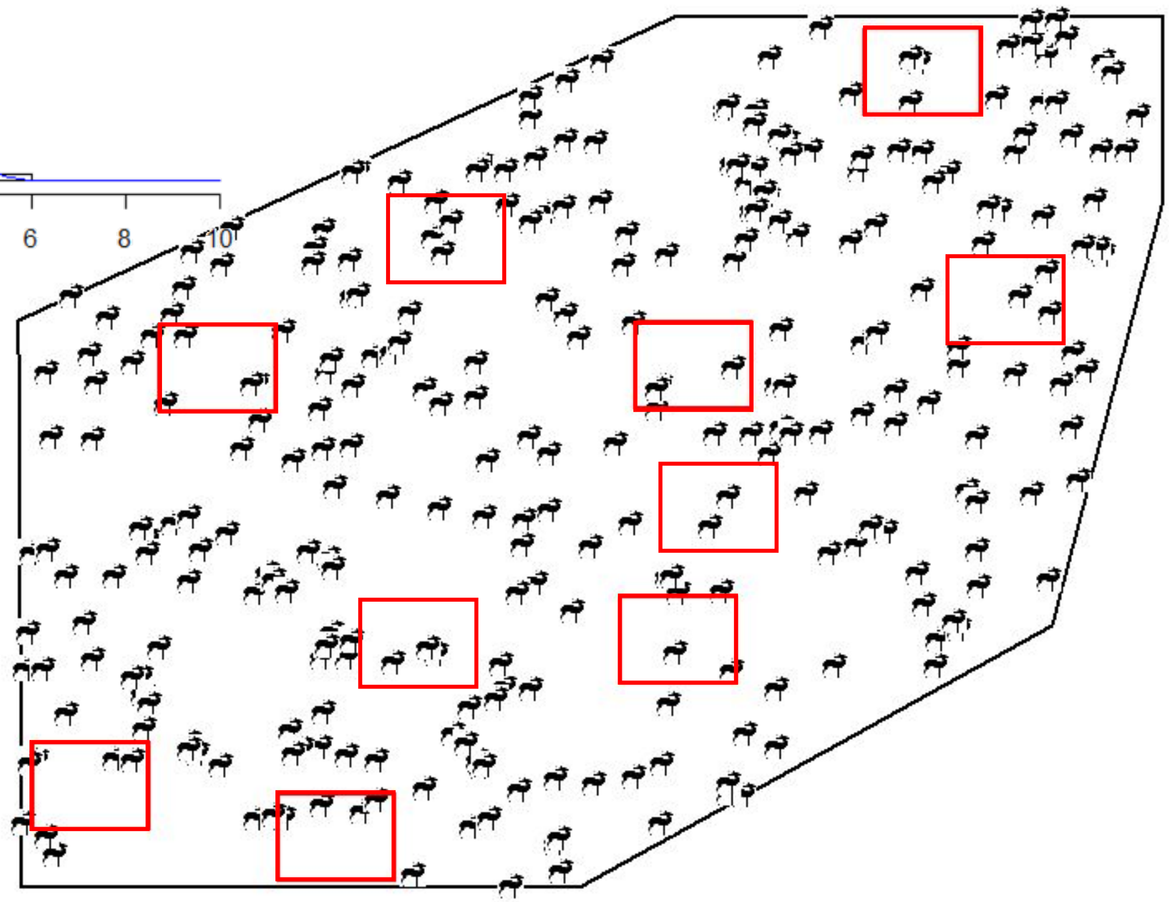
# Increase samples




| Unit | Count |
|---|---|
| A | 4 |
| B | 4 |
| C | 3 |
| D | 3 |
| E | 2 |
| F | 5 |
| G | 3 |
| H | 2 |
| I | 3 |
| J | 4 |
| **Mean** | |
| **StDev** | |
| **CL** | |

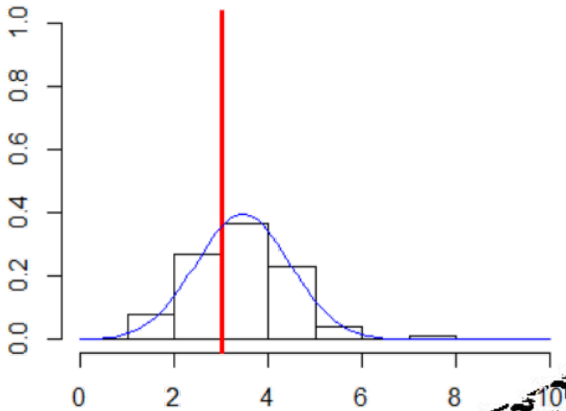
# Increase samples

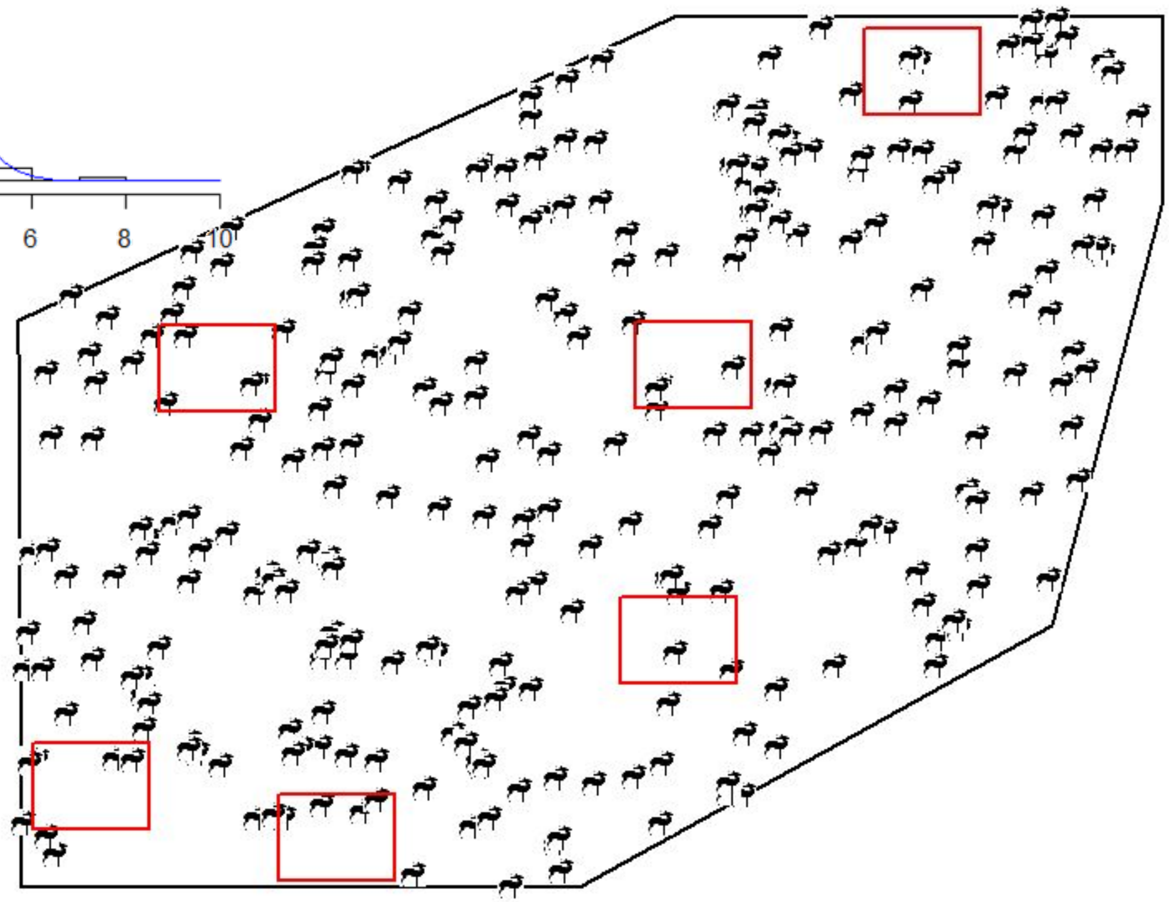


| Unit | Count |
|---|---|
| A | 4 |
| B | 4 |
| C | 3 |
| D | 3 |
| E | 2 |
| F | 5 |
| G | 3 |
| H | 2 |
| I | 3 |
| J | 4 |
| **Mean** | **3.3** |
| **StDev** | **0.95** |
| **CL** | **2.7-3.9** |

# Gets better than before!

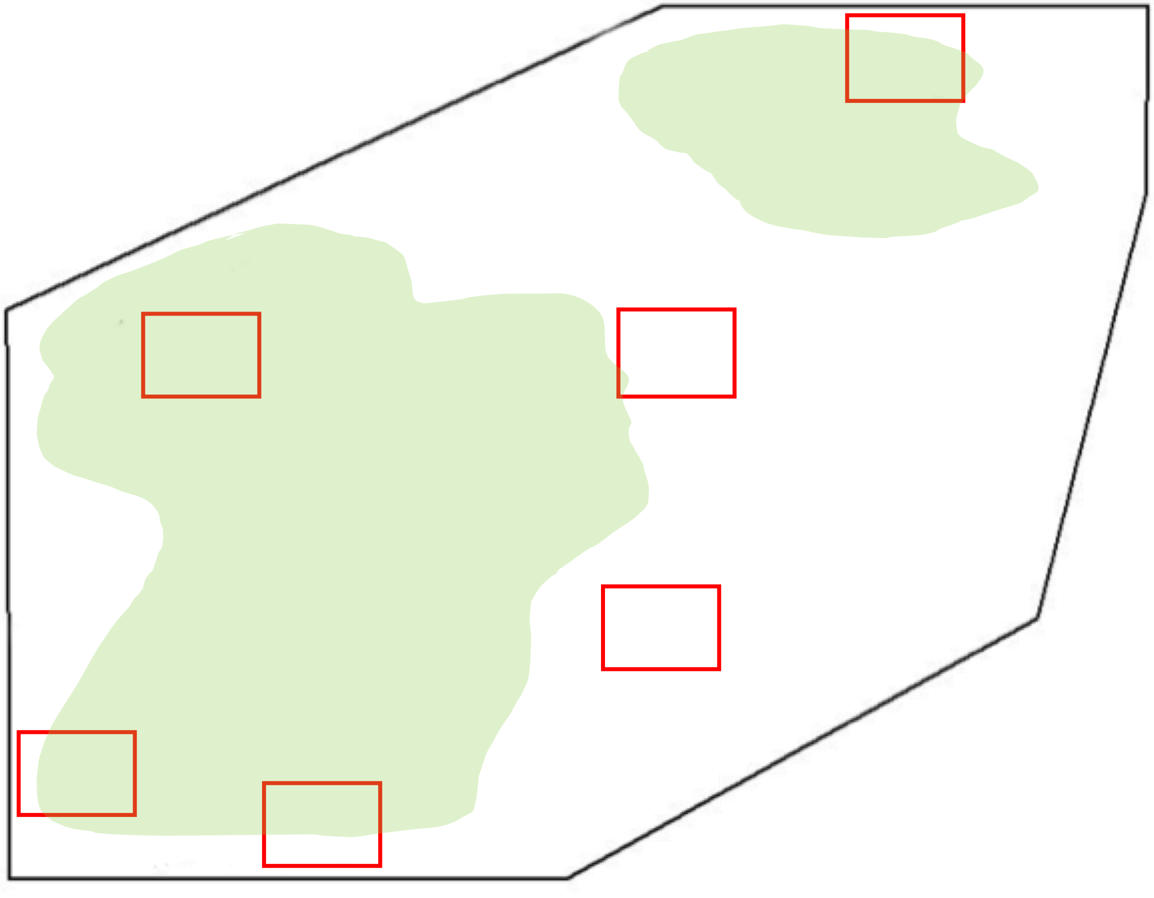


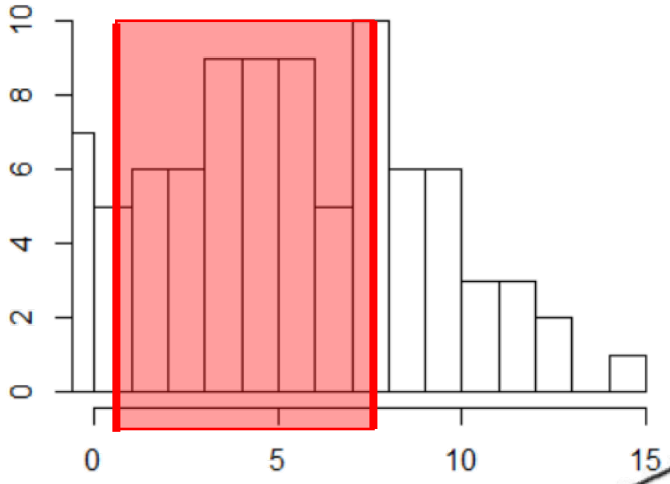
| Unit | Count |
|---|---|
| A | 4 |
| B | 4 |
| C | 3 |
| D | 3 |
| E | 2 |
| F | 5 |
| **Mean** | **3.5** |
| **StDev** | **1.05** |
| **CL** | **2.7-4.3** |

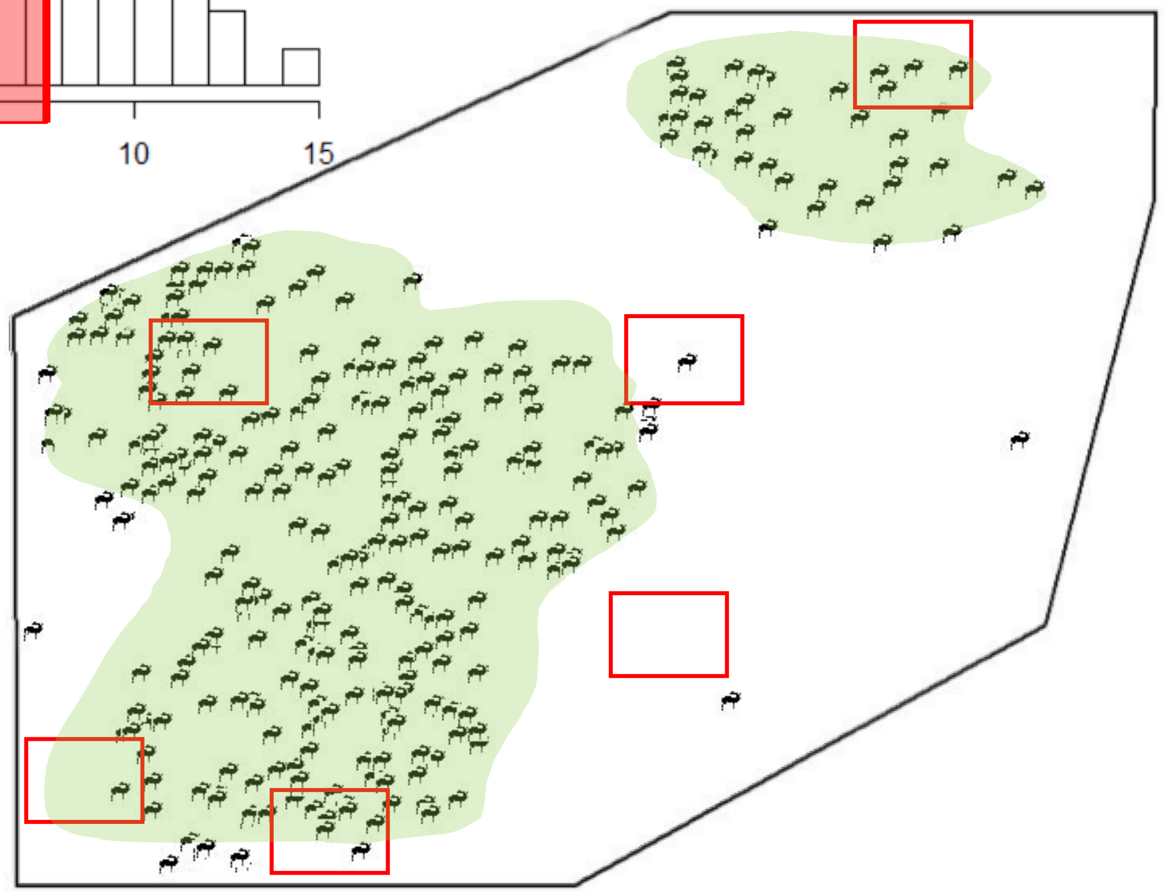
# Habitat

- Uneven distribution



# Habitat

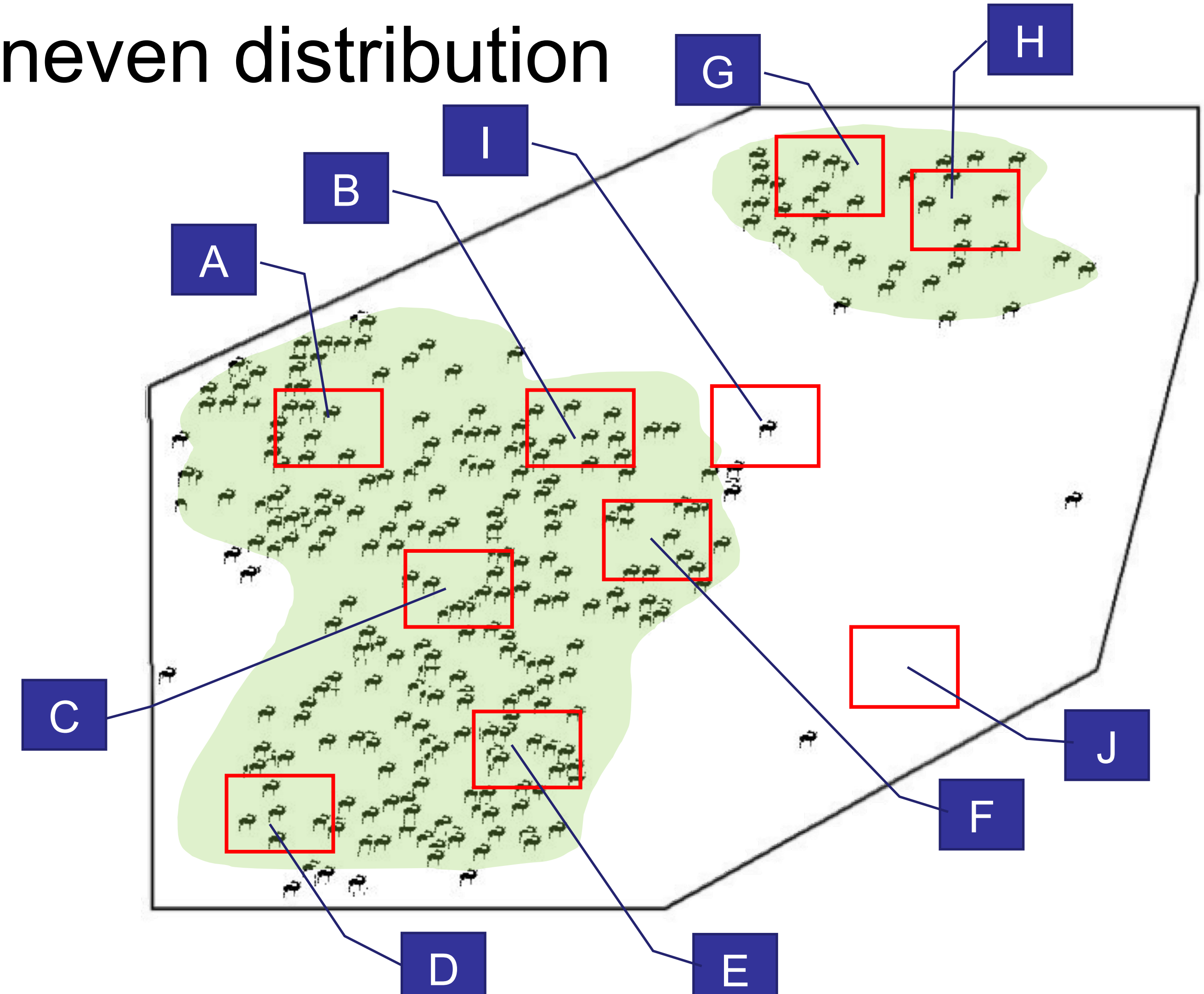


| Unit | Count |
|---|---|
| A | 1 |
| B | 9 |
| C | 1 |
| D | 4 |
| E | 0 |
| F | 10 |
| **Mean** | **4.2** |
| **StDev** | **4.36** |
| **CL** | **0.7-7.6** |

# Habitat

- Uneven distribution




| Unit | Count |
|---|---|
| A | 8 |
| B | 9 |
| C | 10 |
| D | 5 |
| E | 12 |
| F | 11 |
| G | 8 |
| H | 5 |
| **Mean1** | **8.5** |
| **StDev1** | **2.6** |
| I | 1 |
| J | 0 |
| **Mean2** | **0.5** |
| **StDev** | **0.71** |

# Habitat

- Uneven distribution

$a = 0.1km^2$

$A = 10km^2$

$A1 = 3.5km^2$

$A2 = 6.5km^2$

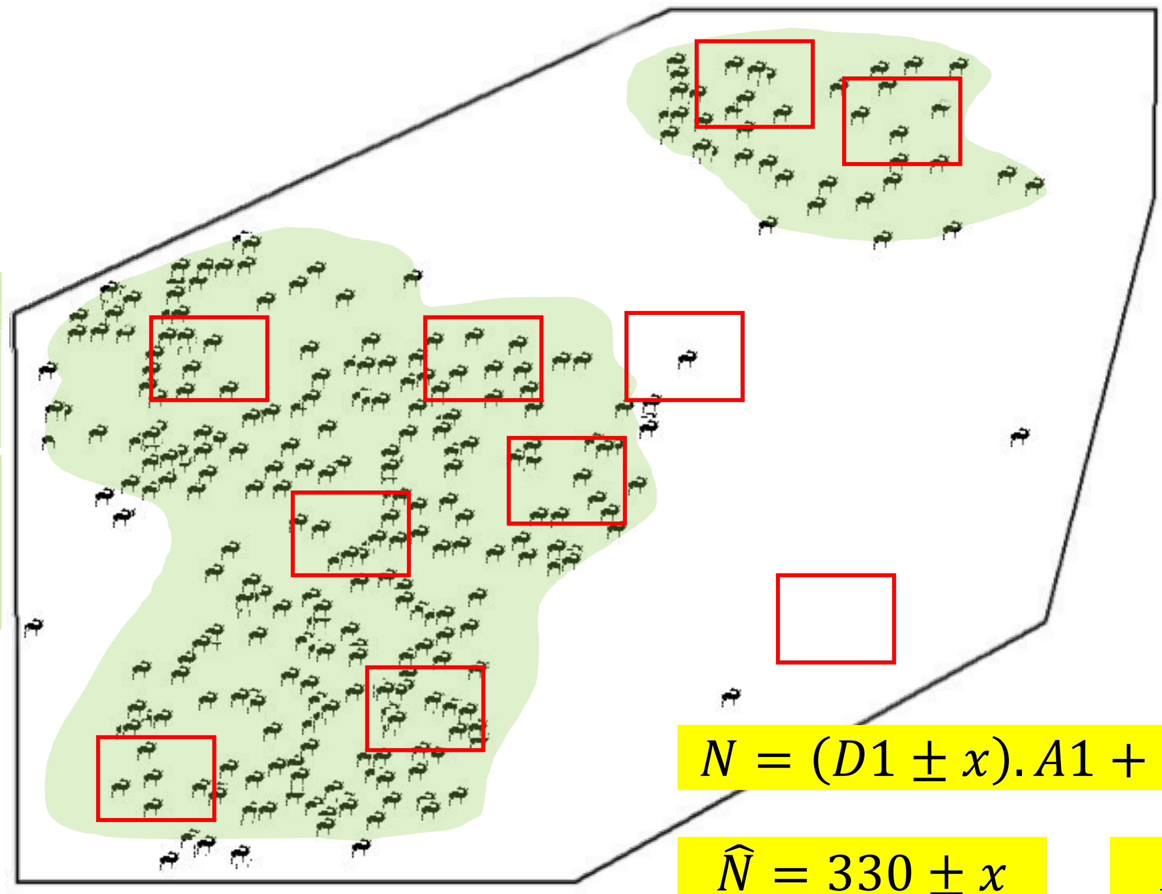
$c = count$

$N = Abundance$

$D = Density$

$D = c/a$

$N = D.A$



| | |
|---|---|
| Mean1 | 8.5 |
| StDev1 | 2.6 |
| Mean2 | 0.5 |
| StDev | 0.71 |
| Mean | 3.3 |
| StDev | 1.6 |

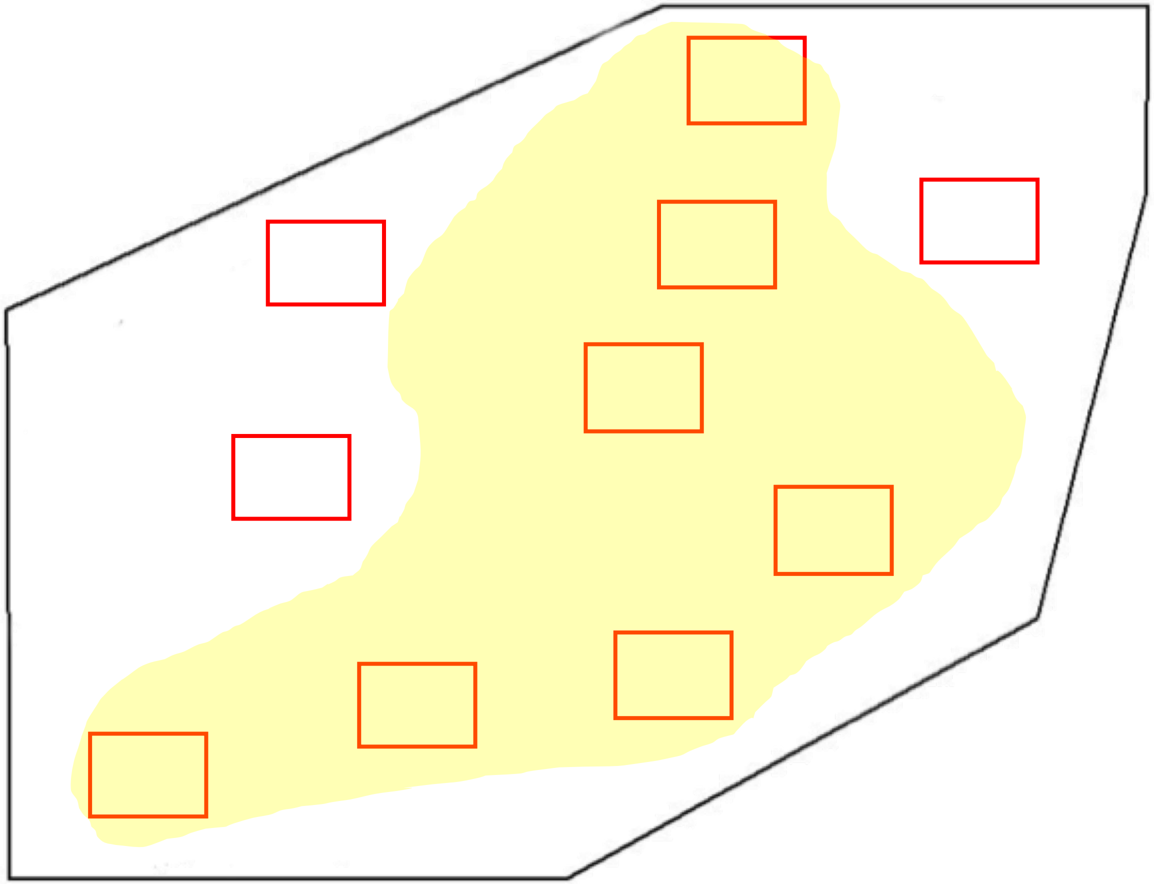
$$N = (D1 \pm x).A1 + (D2 \pm x2).A2$$

$$\widehat{N} = 330 \pm x$$

$$\widehat{D} = 33 \pm x$$

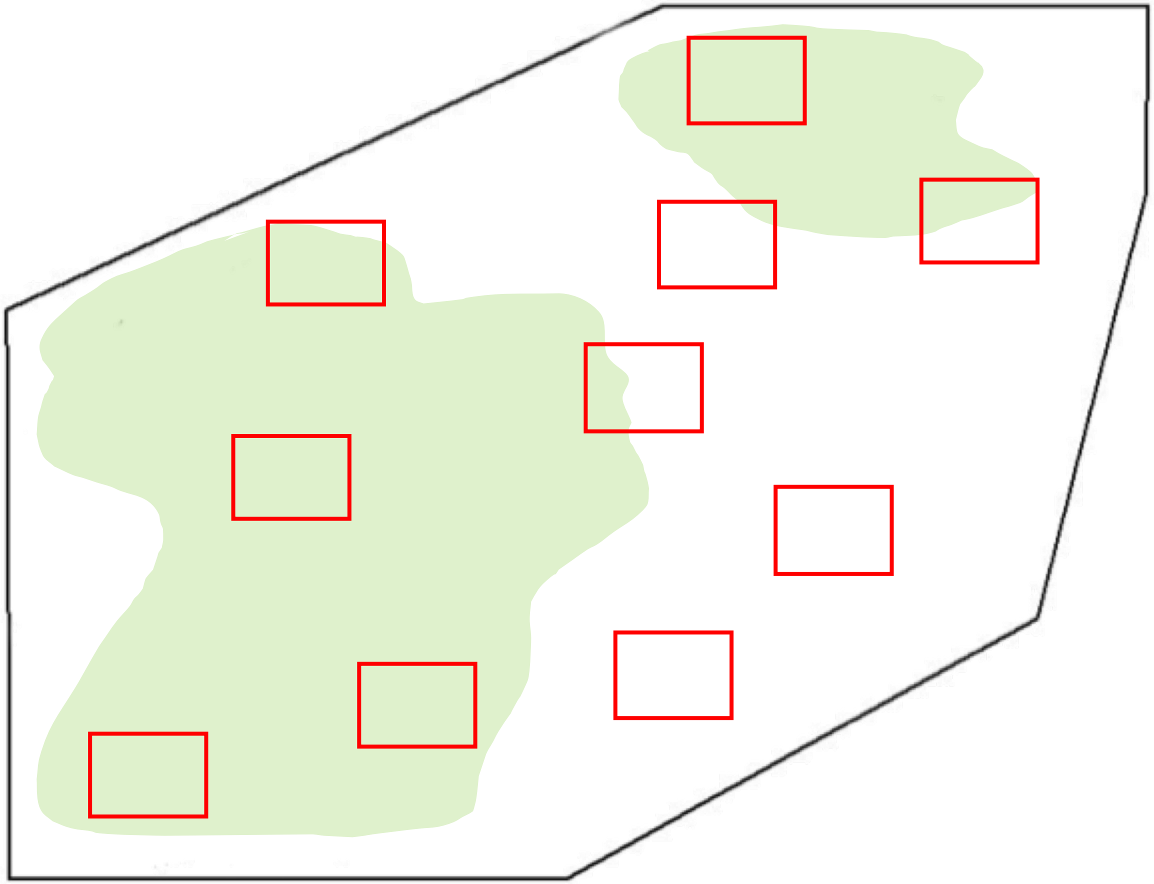
# Question!

- What if our stratification is wrong?



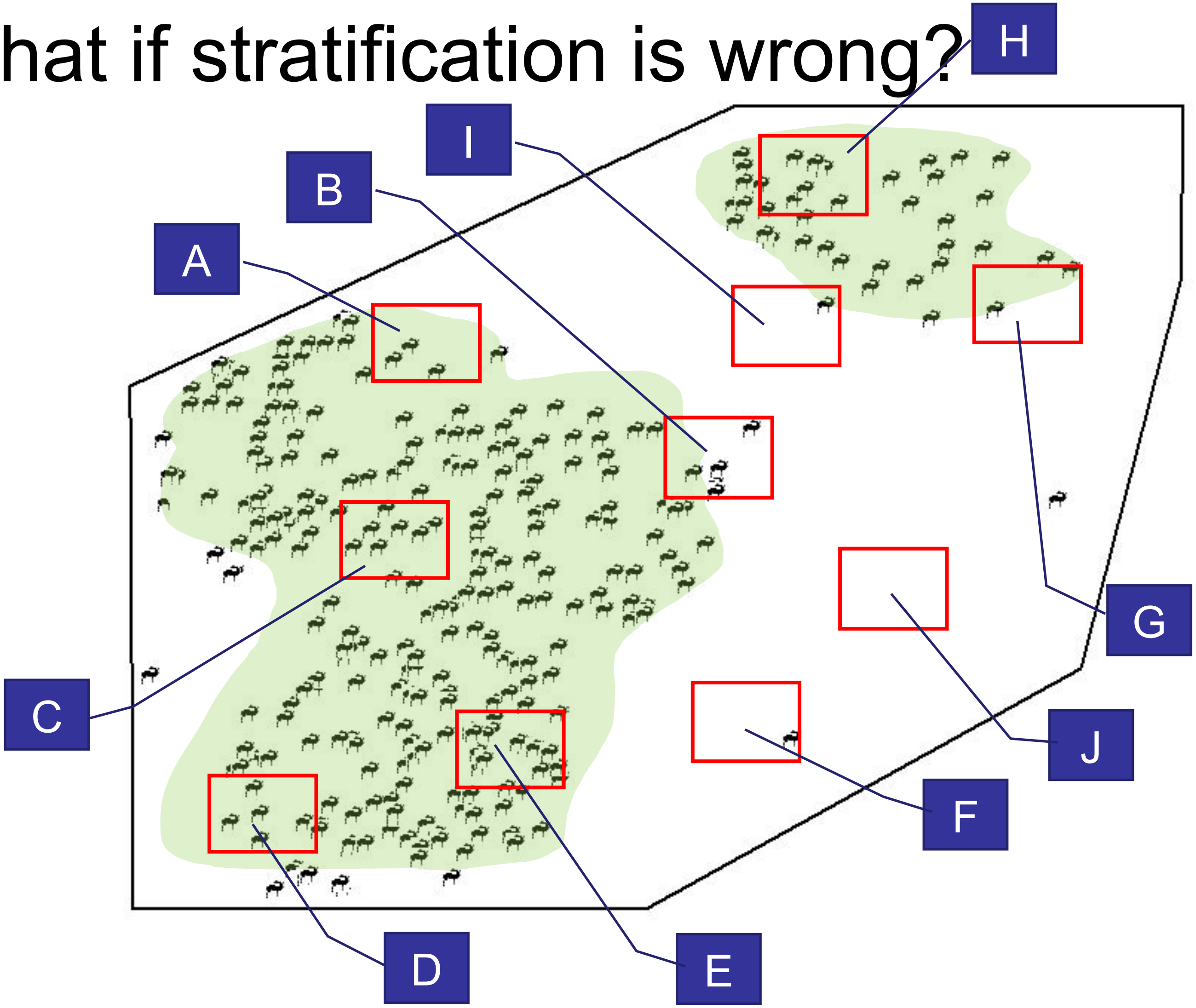
# Question!

- What if stratification is wrong?

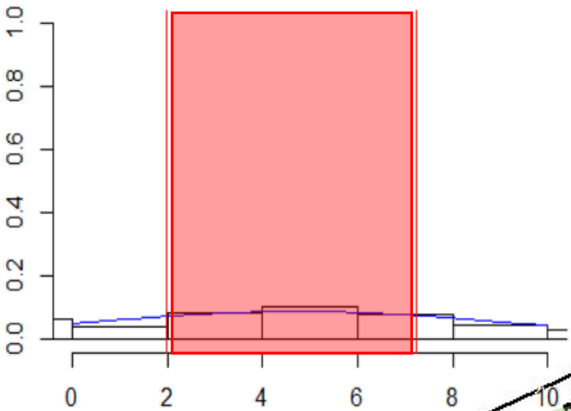


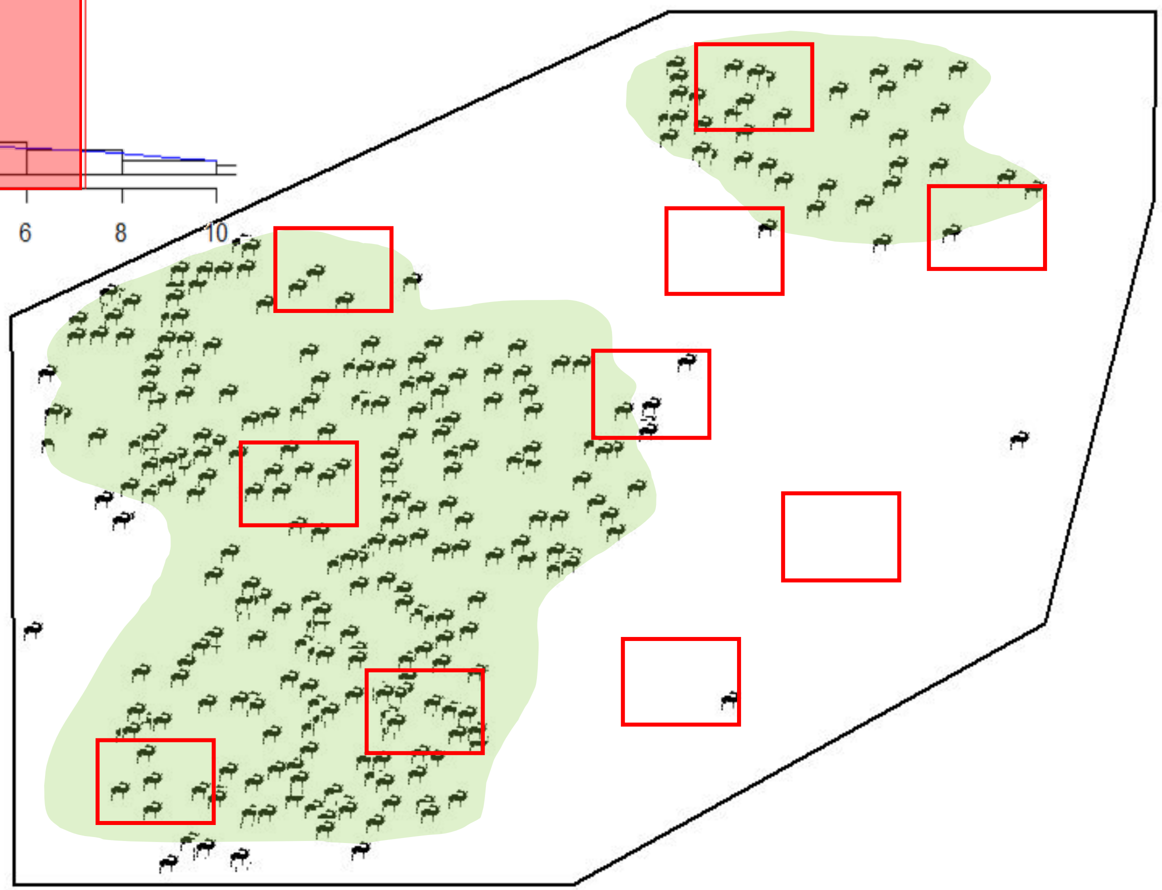
# Question!

- What if stratification is wrong?



# Question!

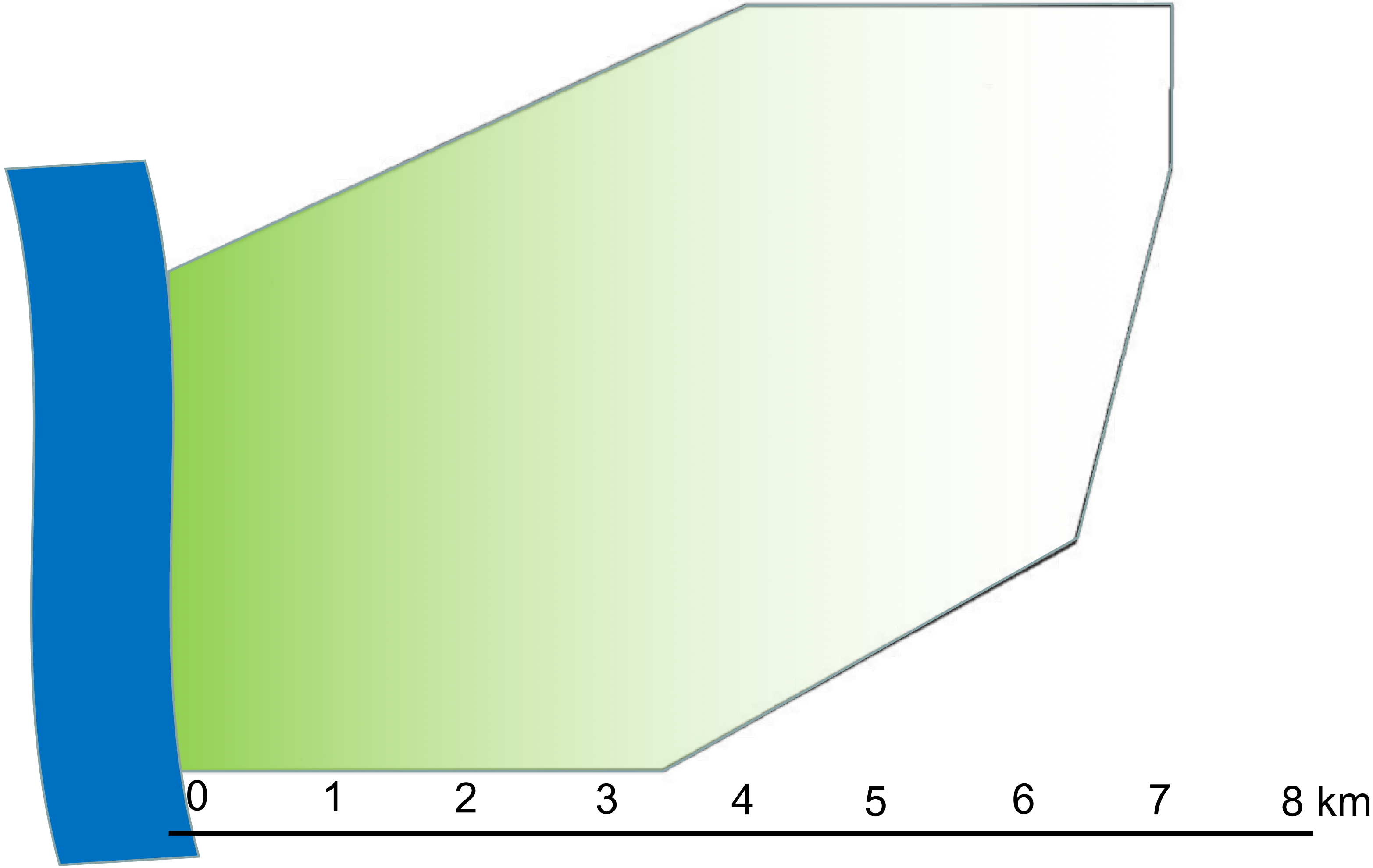


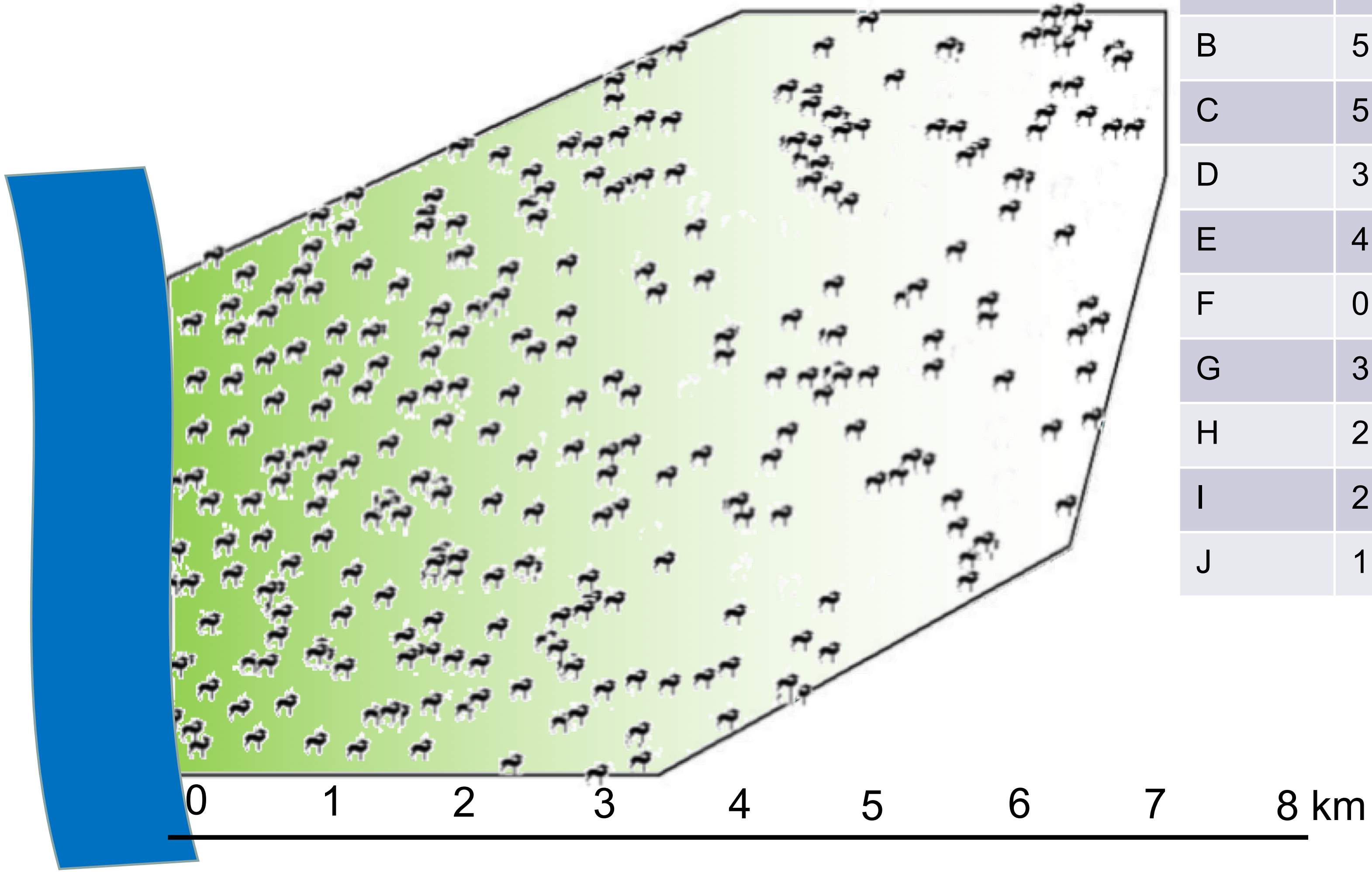
| Unit | Count |
|---|---|
| A | 3 |
| B | 5 |
| C | 7 |
| D | 5 |
| E | 14 |
| F | 1 |
| G | 2 |
| H | 8 |
| I | 1 |
| J | 0 |
| **Mean** | **4.6** |
| **StDev** | **4.25** |
| **CL** | **2.0-7.2** |

# What if habitat not explicit?

- Gradient of distance from a river

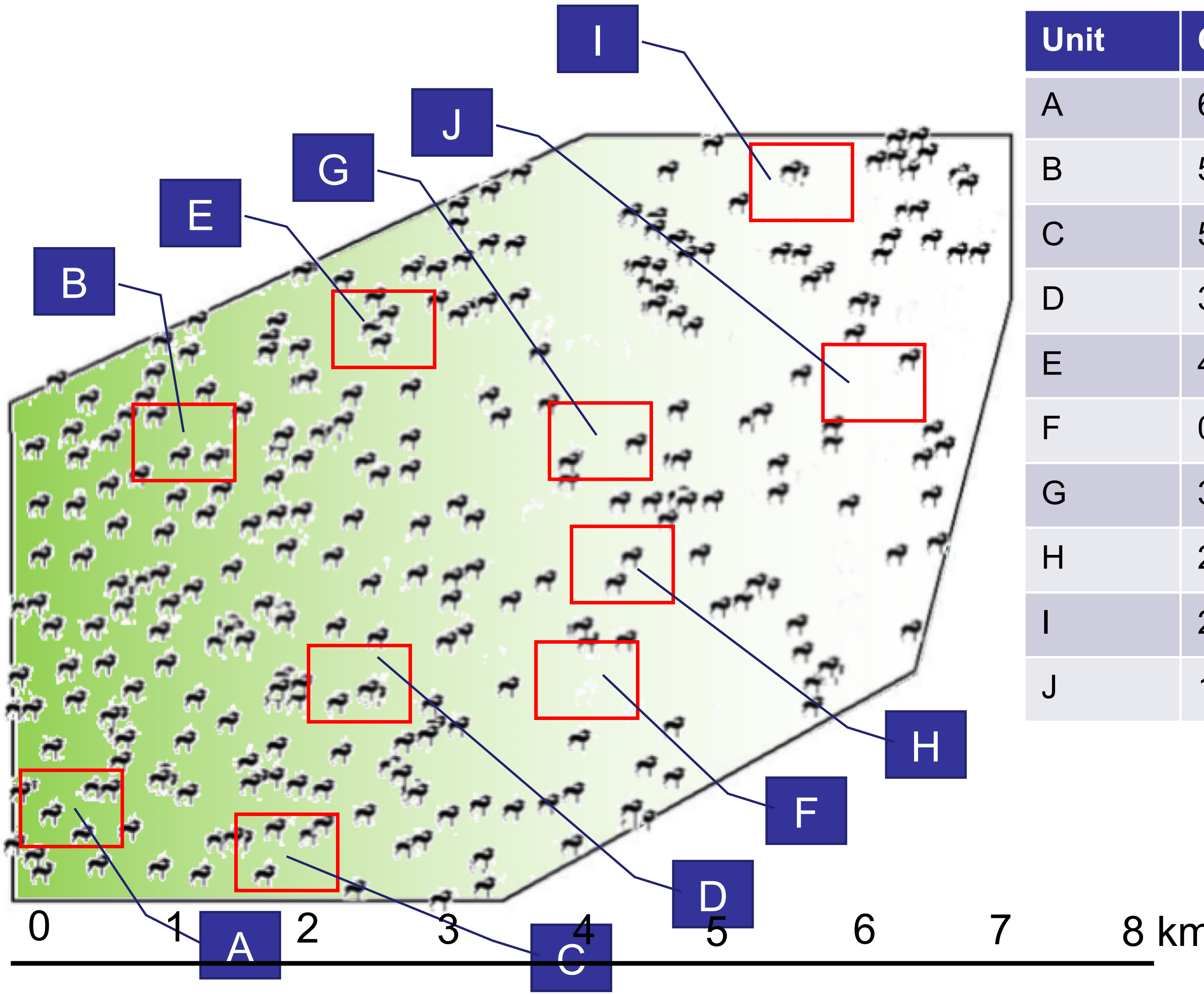


# What if habitat not explicit?



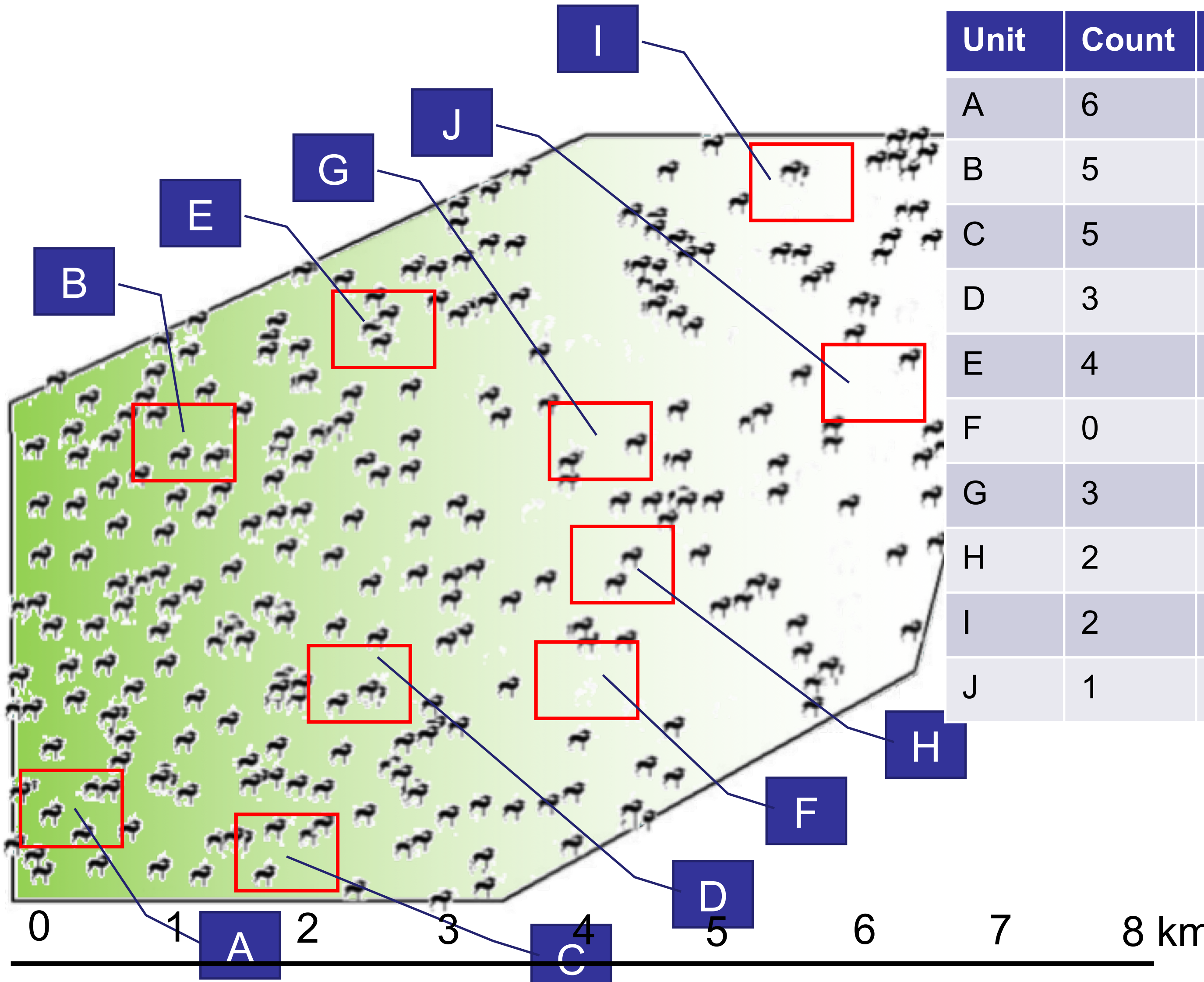

| Unit | Count |
|---|---|
| A | 6 |
| B | 5 |
| C | 5 |
| D | 3 |
| E | 4 |
| F | 0 |
| G | 3 |
| H | 2 |
| I | 2 |
| J | 1 |

# What if habitat not explicit?



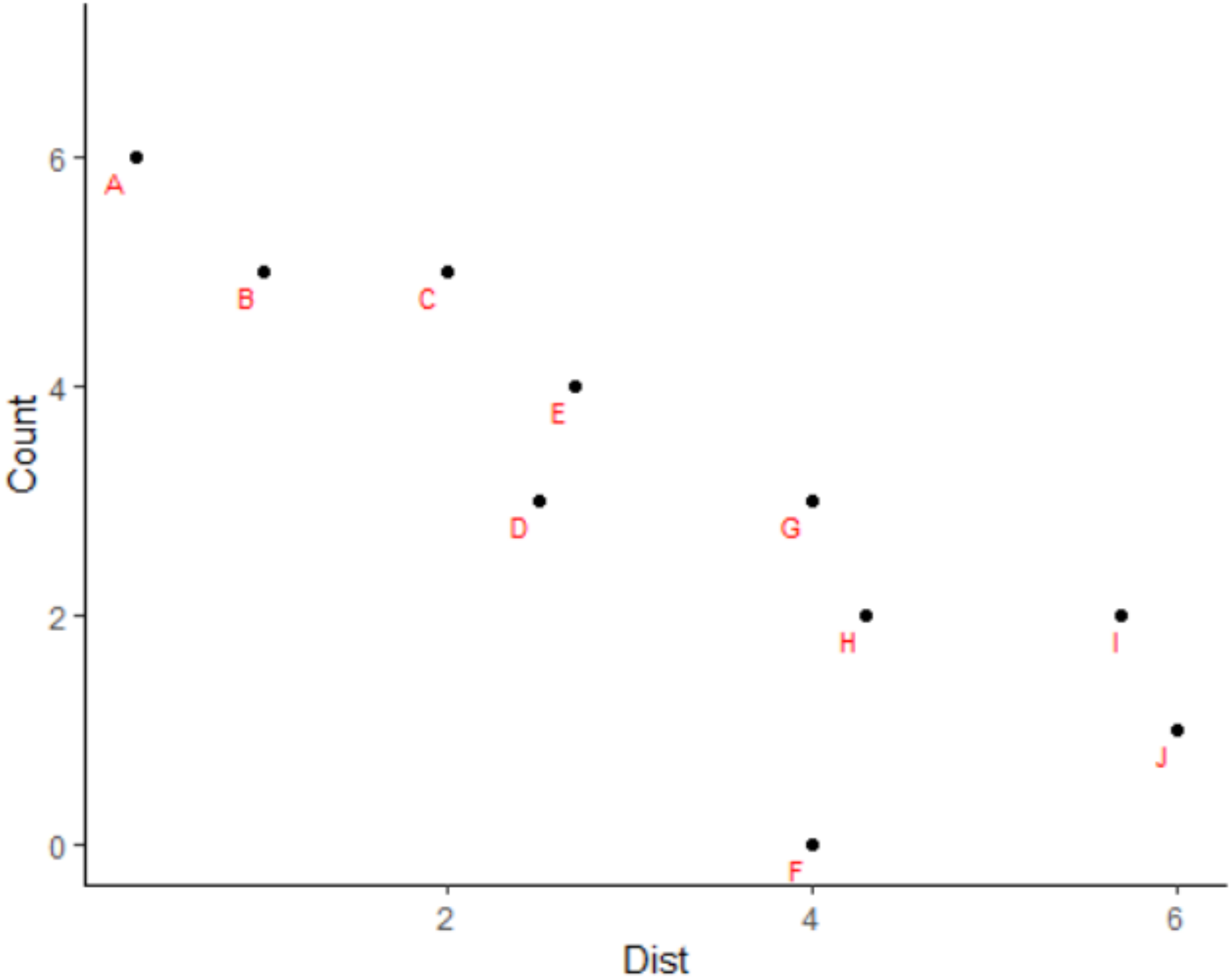

| Unit | Count |
|---|---|
| A | 6 |
| B | 5 |
| C | 5 |
| D | 3 |
| E | 4 |
| F | 0 |
| G | 3 |
| H | 2 |
| I | 2 |
| J | 1 |

# What if habitat not explicit?



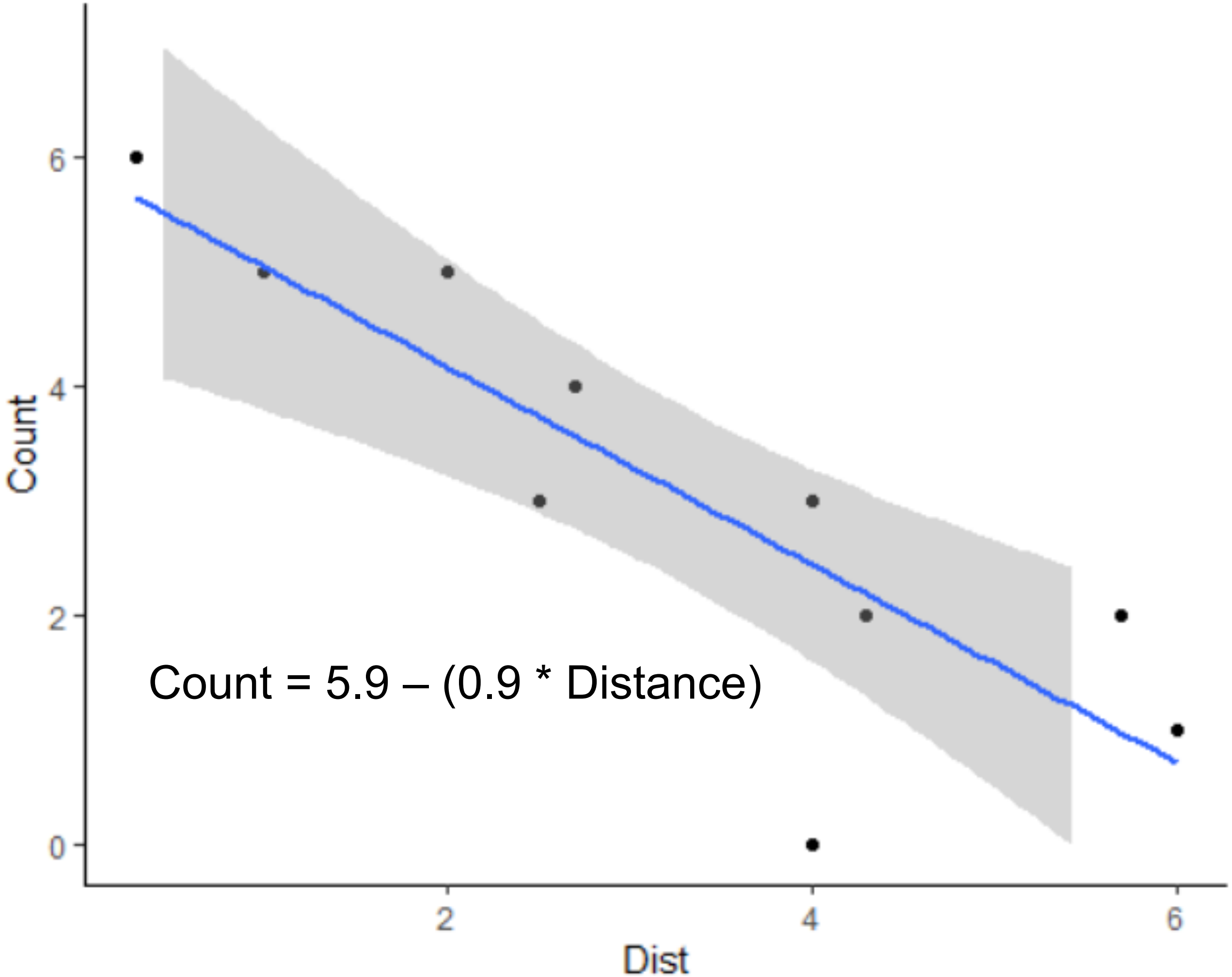

| Unit | Count | Dist |
|---|---|---|
| A | 6 | 0.3 |
| B | 5 | 1 |
| C | 5 | 2 |
| D | 3 | 2.5 |
| E | 4 | 2.7 |
| F | 0 | 4 |
| G | 3 | 4 |
| H | 2 | 4.3 |
| I | 2 | 5.7 |
| J | 1 | 6 |

# Visualizing on a scatterplot



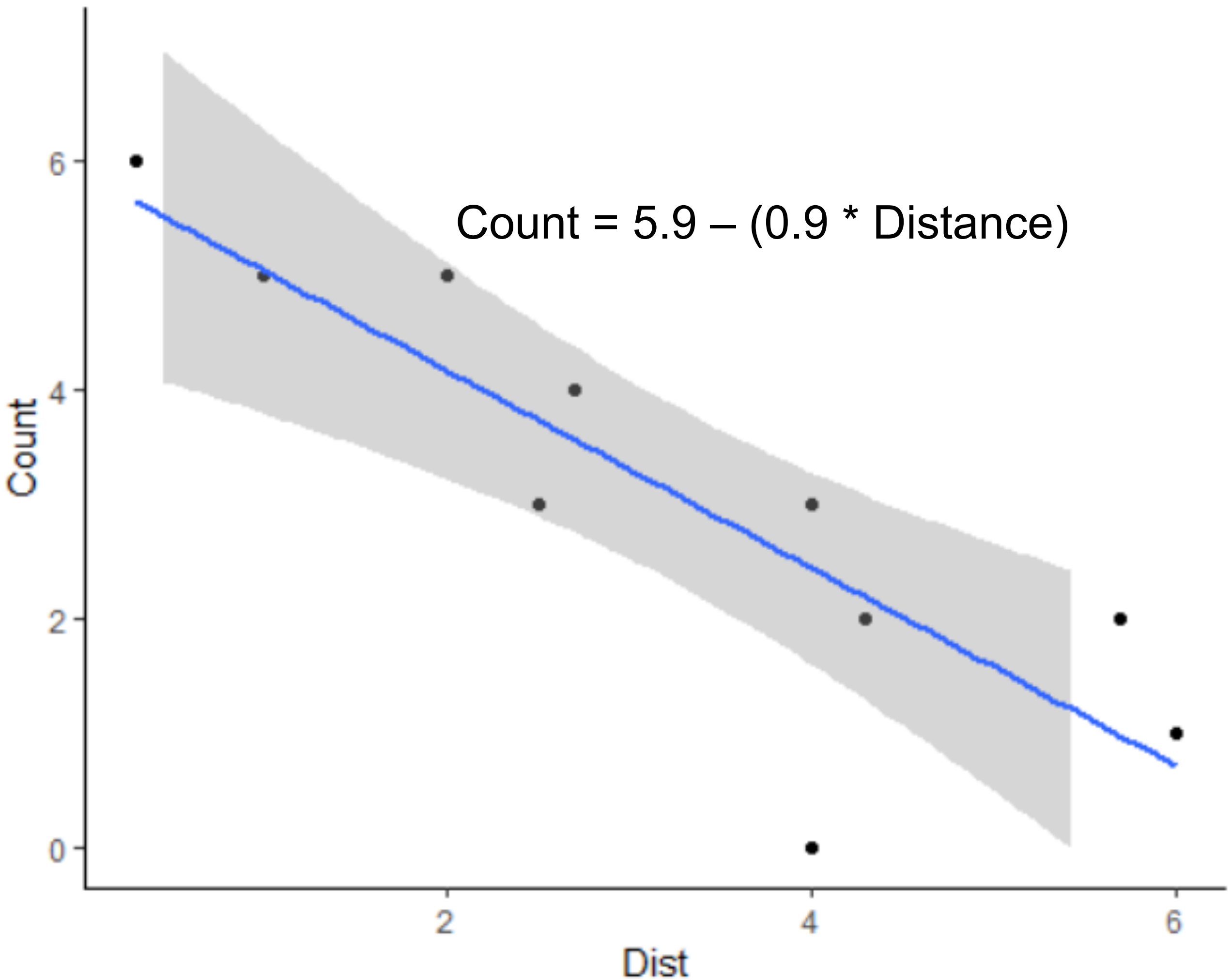

| Unit | Dist | Count |
|---|---|---|
| A | 0.3 | 6 |
| B | 1 | 5 |
| C | 2 | 5 |
| D | 2.5 | 3 |
| E | 2.7 | 4 |
| F | 4 | 0 |
| G | 4 | 3 |
| H | 4.3 | 2 |
| I | 5.7 | 2 |
| J | 6 | 1 |

# We model!




| Unit | Dist | Count |
|---|---|---|
| A | 0.3 | 6 |
| B | 1 | 5 |
| C | 2 | 5 |
| D | 2.5 | 3 |
| E | 2.7 | 4 |
| F | 4 | 0 |
| G | 4 | 3 |
| H | 4.3 | 2 |
| I | 5.7 | 2 |
| J | 6 | 1 |

# Modeled count!



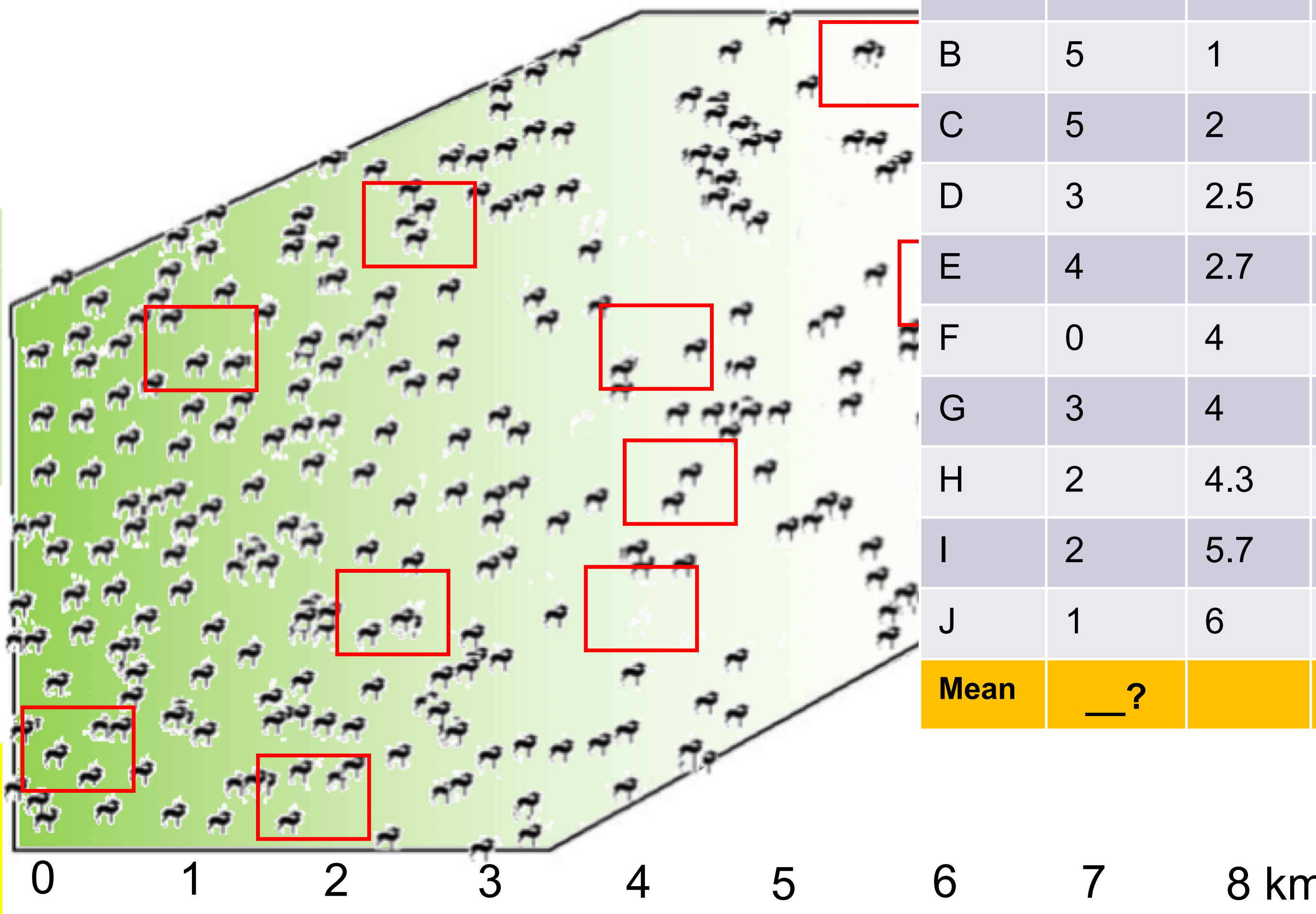

| Unit | Dist | Count | $\hat{c}$ |
|---|---|---|---|
| A | 0.3 | 6 | __? |
| B | 1 | 5 | **5.0** |
| C | 2 | 5 | **4.1** |
| D | 2.5 | 3 | **3.7** |
| E | 2.7 | 4 | __? |
| F | 4 | 0 | **2.3** |
| G | 4 | 3 | **2.3** |
| H | 4.3 | 2 | **2.0** |
| I | 5.7 | 2 | **0.8** |
| J | 6 | 1 | **0.5** |

# Average D vs Modeled D

$a = area\ of\ unit$
$A = Total\ Area$
$c = count$
$N = Abundance$
$D = Density$

$D = c/a$
$N = D.A$

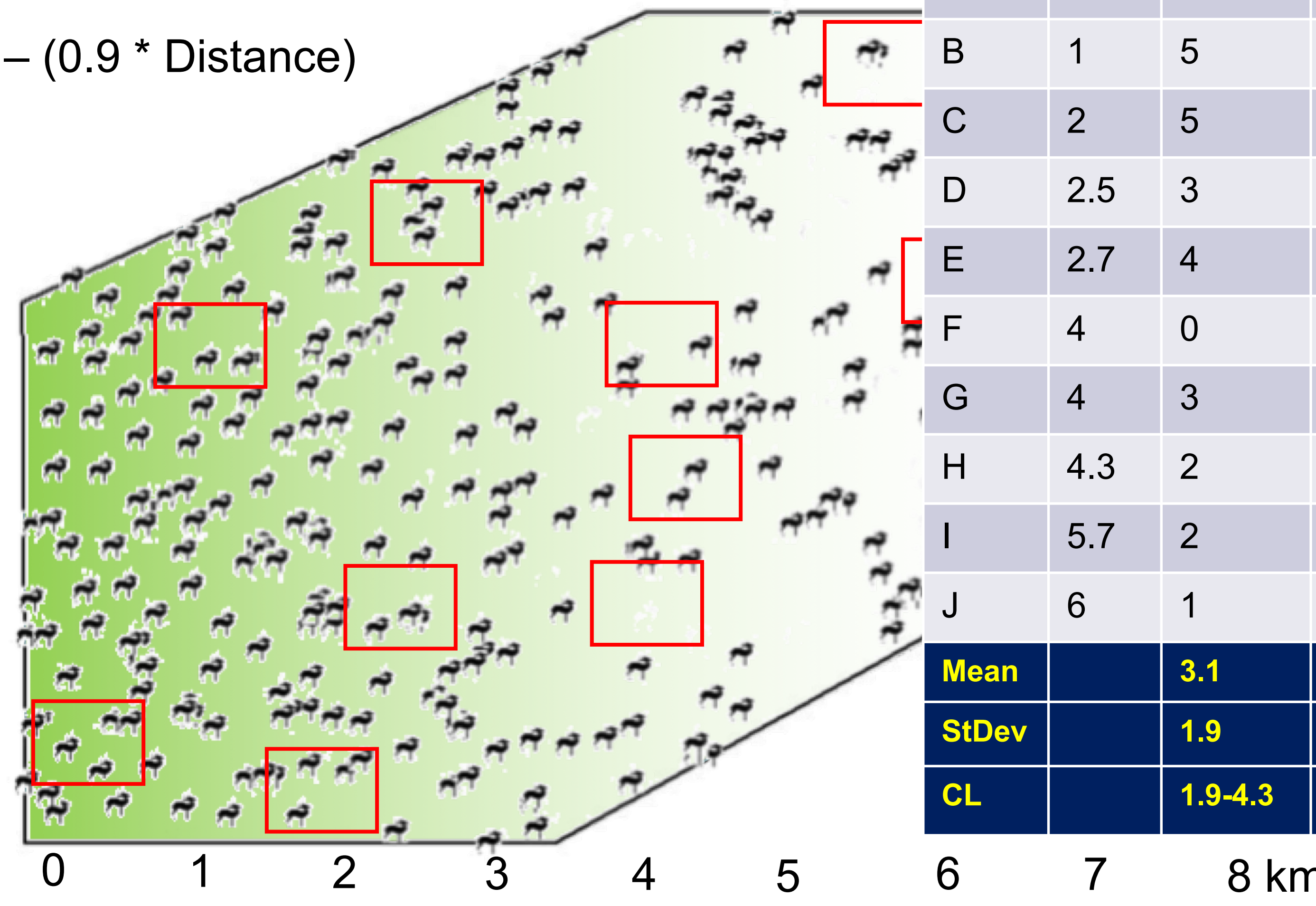
$D =$ __?
$N =$ __?




| Unit | Count | Dist | $\hat{c}$ |
|---|---|---|---|
| A | 6 | 0.3 | **5.6** |
| B | 5 | 1 | **5.0** |
| C | 5 | 2 | **4.1** |
| D | 3 | 2.5 | **3.7** |
| E | 4 | 2.7 | **3.5** |
| F | 0 | 4 | **2.3** |
| G | 3 | 4 | **2.3** |
| H | 2 | 4.3 | **2.0** |
| I | 2 | 5.7 | **0.8** |
| J | 1 | 6 | **0.5** |
| **Mean** | **__?** | | **__?** |

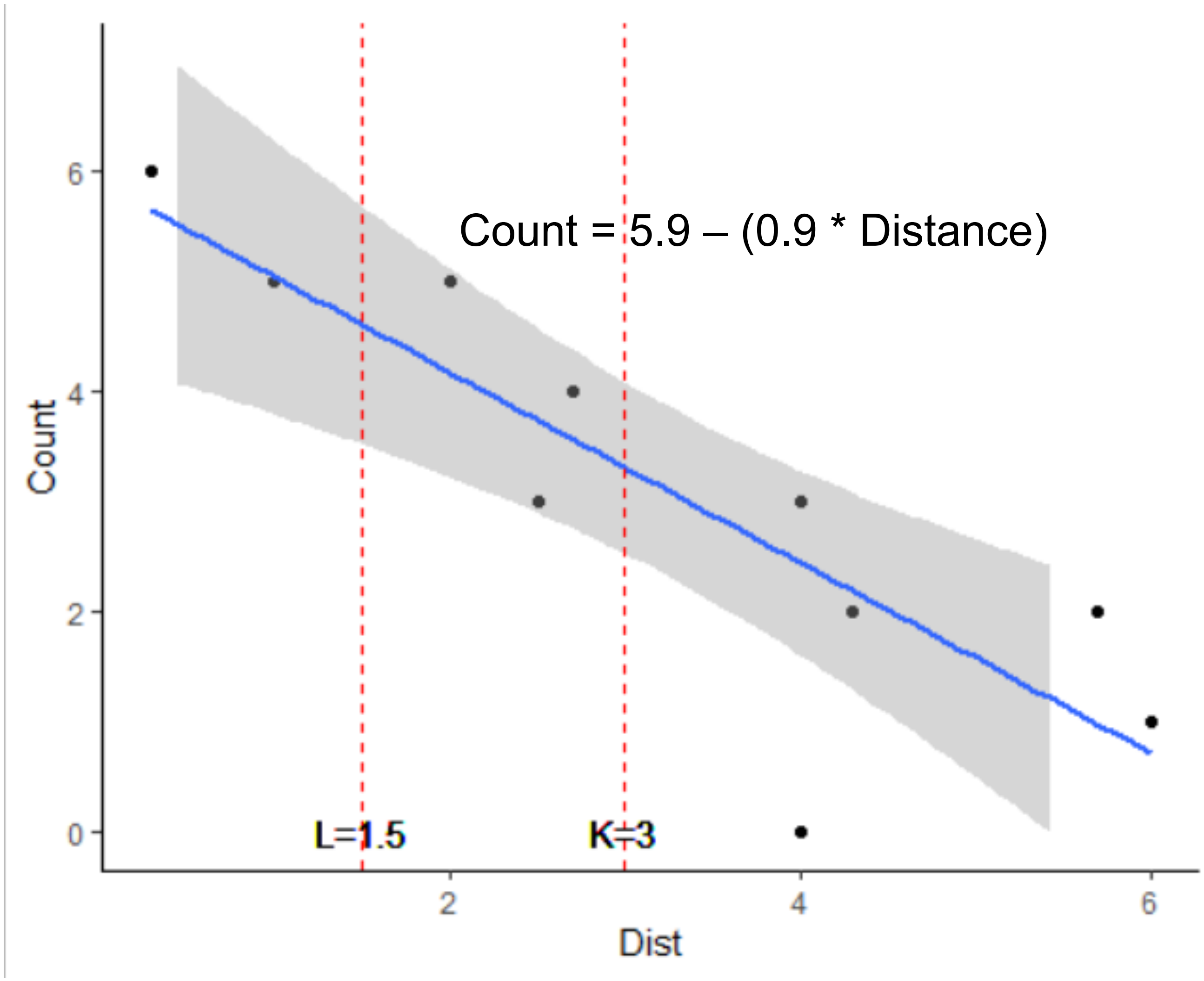
# Modeled count (or density)

Count = 5.9 – (0.9 * Distance)




| Unit | Dist | Count | $\hat{C}$ |
|---|---|---|---|
| A | 0.3 | 6 | **5.6** |
| B | 1 | 5 | **5.0** |
| C | 2 | 5 | **4.1** |
| D | 2.5 | 3 | **3.7** |
| E | 2.7 | 4 | **3.5** |
| F | 4 | 0 | **2.3** |
| G | 4 | 3 | **2.3** |
| H | 4.3 | 2 | **2.0** |
| I | 5.7 | 2 | **0.8** |
| J | 6 | 1 | **0.5** |
| **Mean** | | **3.1** | **3.0** |
| **StDev** | | **1.9** | **1.7** |
| **CL** | | **1.9-4.3** | **1.9-4.0** |

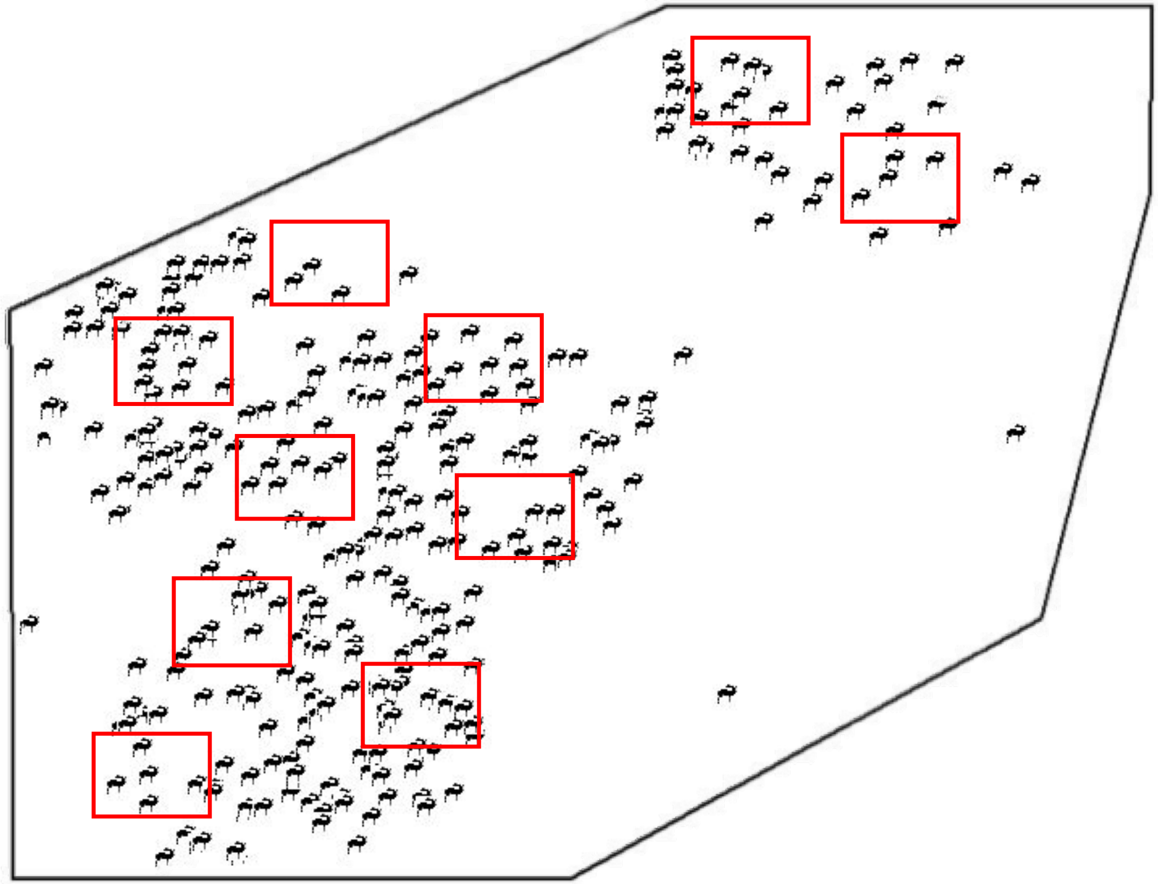
# Predict the count!




| Unit | Dist | Count | $\widehat{C}$ |
|---|---|---|---|
| A | 0.3 | 6 | **5.6** |
| B | 1 | 5 | **5.0** |
| C | 2 | 5 | **4.1** |
| D | 2.5 | 3 | **3.7** |
| E | 2.7 | 4 | **3.5** |
| F | 4 | 0 | **2.3** |
| G | 4 | 3 | **2.3** |
| H | 4.3 | 2 | **2.0** |
| I | 5.7 | 2 | **0.8** |
| J | 6 | 1 | **0.5** |
| K | 3 | NA | **__?** |
| L | 1.5 | NA | **__?** |

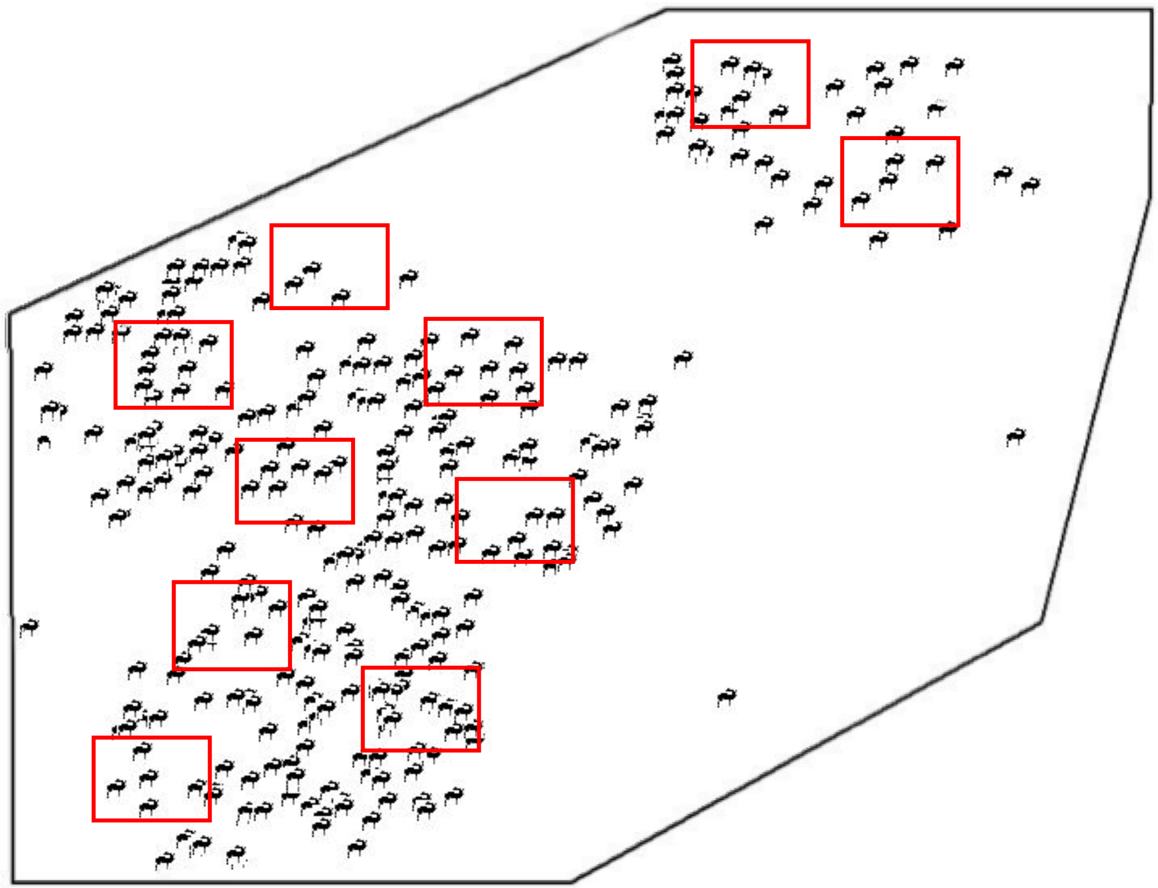
# Question!

- What if we only sample where animals are?



# Question!

- What if we only sample where animals are?



| Unit | Count |
|---|---|
| A | 11 |
| B | 3 |
| C | 9 |
| D | 7 |
| E | 8 |
| F | 9 |
| G | 11 |
| H | 5 |
| I | 8 |
| J | 4 |

$a = area\ of\ unit$
$A = Total\ Area$
$c = count$
$N = Abundance$
$D = Density$

$D = c/a$
$N = D.A$

$D = \_\_?$
$N = \_\_?$

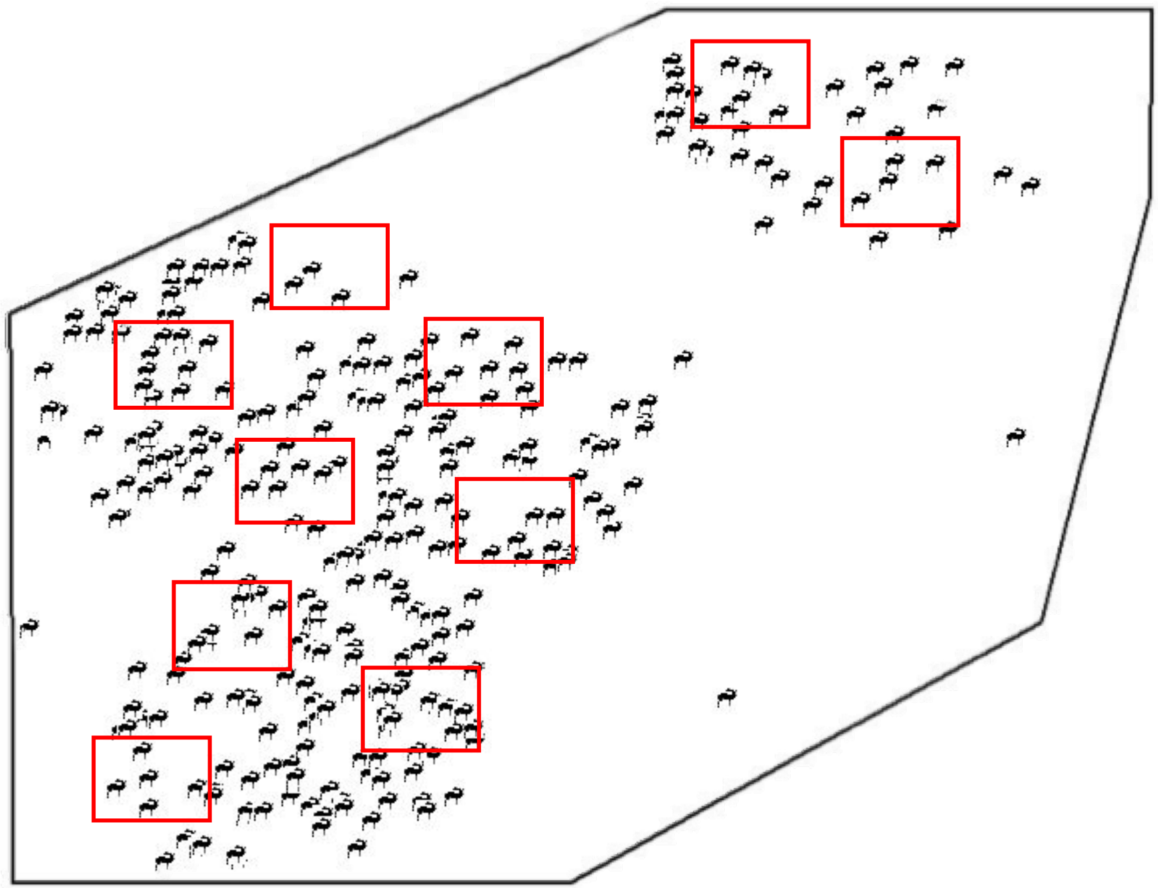
# Answer!

- We bias…

$a = area\ of\ unit$
$A = Total\ Area$
$c = count$
$N = Abundance$
$D = Density$

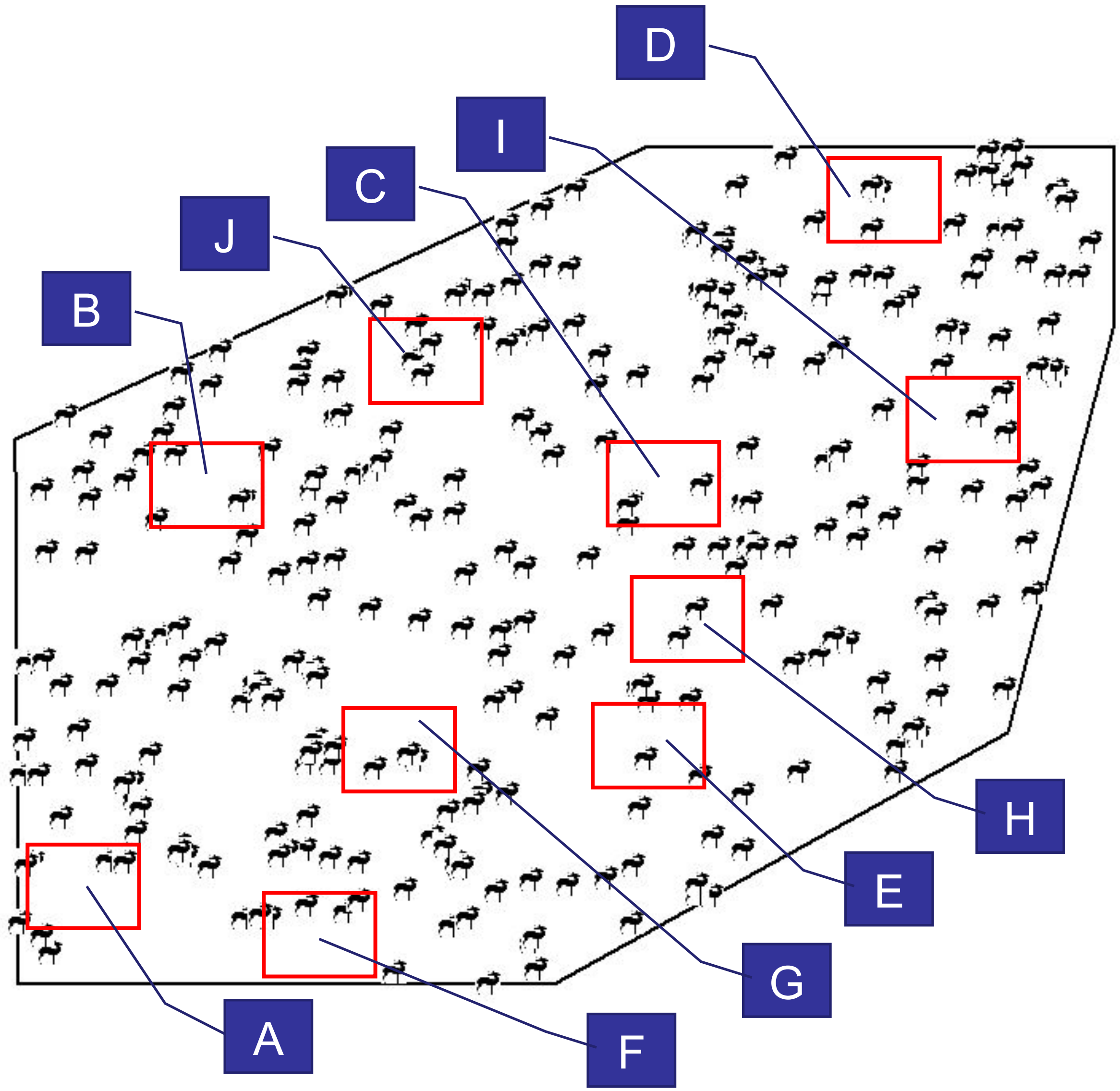
$D = c/a$
$N = D.A$

$D = \_\_?$
$N = \_\_?$



| Unit | Count |
|---|---|
| A | 11 |
| B | 3 |
| C | 9 |
| D | 7 |
| E | 8 |
| F | 9 |
| G | 11 |
| H | 5 |
| I | 8 |
| J | 4 |
| **Mean** | **7.5** |
| **StDev** | **2.8** |

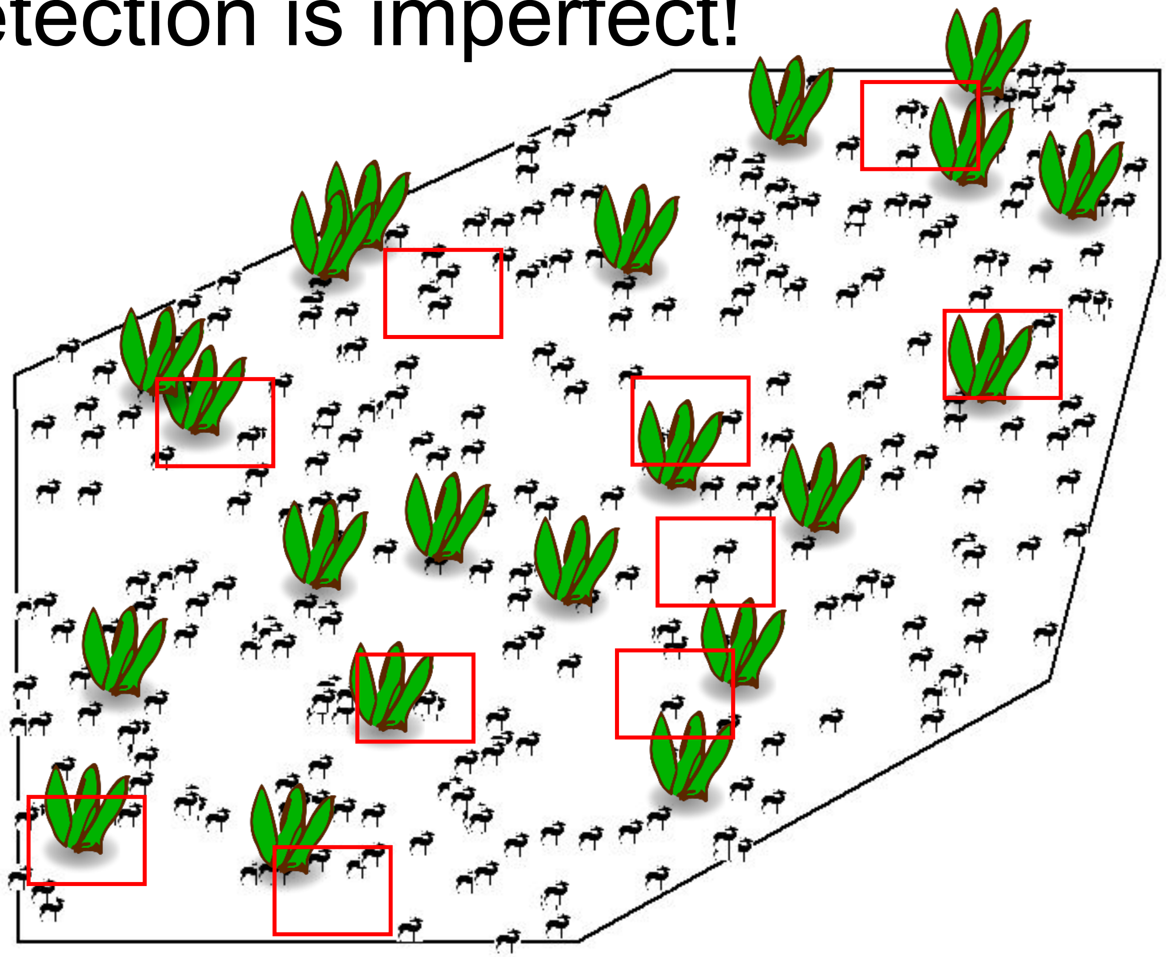
# What if…




| Unit | Count |
|---|---|
| A | 4 |
| B | 4 |
| C | 3 |
| D | 3 |
| E | 2 |
| F | 5 |
| G | 3 |
| H | 2 |
| I | 3 |
| J | 4 |
| **Mean** | **3.3** |
| **StDev** | **0.95** |
| **CL** | **2.7-3.9** |

# What if…

- Detection is imperfect!



| Unit | Count | C1 |
|---|---|---|
| A | 4 | 2 |
| B | 4 | 3 |
| C | 3 | 1 |
| D | 3 | 3 |
| E | 2 | 2 |
| F | 5 | 4 |
| G | 3 | 2 |
| H | 2 | 2 |
| I | 3 | 2 |
| J | 4 | 4 |
| **Mean** | **3.3** | **2.5** |
| **StDev** | **0.95** | **0.97** |
| **CL** | **2.7-3.9** | **1.9-3.1** |

# Negative Bias

- Unknown number of animals NOT seen
- An unknown proportion of total animals SEEN

$$D = \frac{N}{A} = \frac{\hat{C}}{A}$$

Count

Area

# Detection Probability (p)

- Unknown number of animals NOT seen
- An unknown proportion of total animals SEEN

$$D = \frac{N}{A} = \frac{\hat{C}}{pA}$$

Count

Area

Detection Probability

# Detection Probability (p)

- Unknown number of animals NOT seen
- An unknown proportion of total animals SEEN

$$D = \frac{N}{A} = \frac{\hat{C}}{pA}$$

Count

Detection Probability

# Design-based inference

- Fewer assumptions
- Easy to understand outputs
- Estimators obtained computationally through sample design using weights & other auxiliary information
- Have ‘good’ statistical properties

# Model based

- Useful in case of non-sampling errors or discontinuities in survey design
- Valuable in understanding patterns
- Need ecological information about covariates
- Wrong selection of covariates may lead to spurious inference