Designing Efficient Sampling Plans for Ecology and Conservation

Edward F. Connor

Department of Biology

San Francisco State University

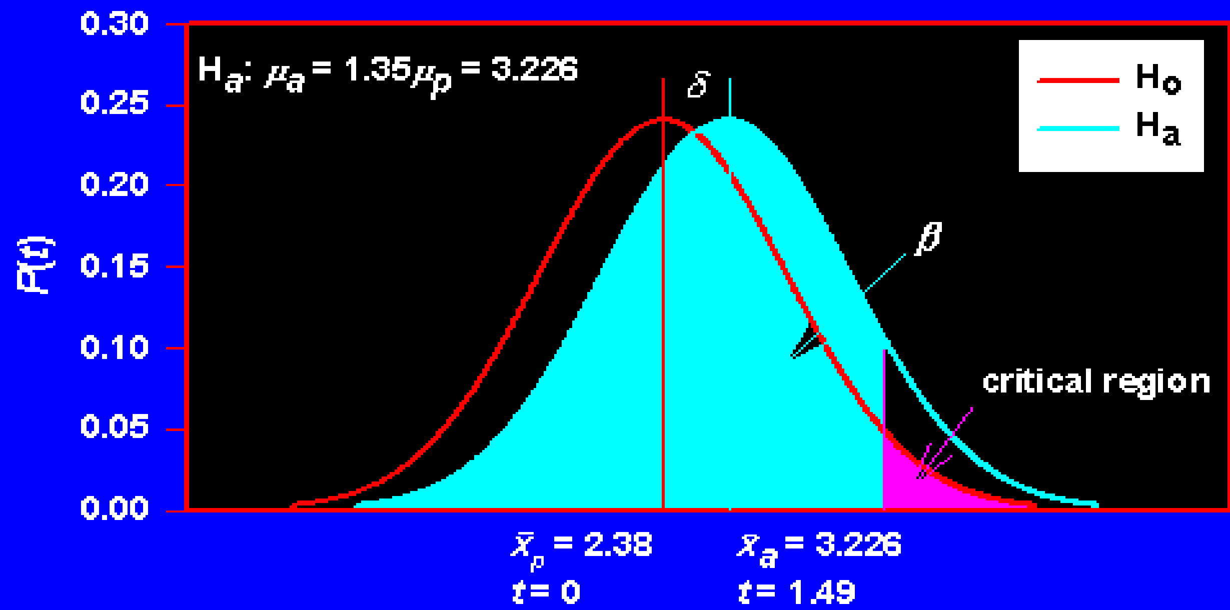
Issues in Study Design

- Sampling ***intensity*** (how much to sample)
- Choosing among different ***sampling designs*** (between and within-subjects designs)
- Determining sampling ***frequency*** (how often to sample)
- Determining which study design is most ***cost effective***

Sampling Intensity
Choosing a Sample Size

- Obtain a preliminary estimate of σ
- Conjecture the minimally interesting difference in treatment means, or effect size, δ
- Set the Type I error rate, α
- Set the desired power for the test (1- β)
- Calculate necessary sample size by approximation, web-based calculators, software, or simulation

Figure 2 - Distribution of t under H_o and H_a



Recommended Calculators for Sample Size Determination

Web Based

Author	Location	URL
R. Lenth	University of Iowa	http://www.stat.uiowa.edu/~rlenth/Power/index.html
Rollin Brant	U Calgary	http://stat.ubc.ca/~rollin/stats/ssize/index.html
M. Friendly	York University	http://www.math.yorku.ca/SCS/Online/power/

Software

Author	Package	Availability
W. Dupont & W. Plummer	PS	http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize
R. O'Brien	UnifyPow	http://www.bio.ri.ccf.org/power.html
A. Buchner, F. Faul, & E. Erdfelder	G*Power	http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/index.html
J. Hintze	PASS	http://www.ncss.com/pass.html (for purchase)

Choosing a Sampling Design

ANOVA Designs

- Between Subjects*
- Within Subjects*

Regression Designs

- Between Subject

Mixed Model Designs

- Between Subject
- Within Subject
- Covariates

Difference between Separate Groups and Paired t-test

Separate Groups *t*-test

$$t = \frac{(\bar{x}_a - \bar{x}_b) - (\mu_a - \mu_b)}{\sqrt{S^2_{\bar{x}_a - \bar{x}_b}}}$$

$$s^2_{\bar{x}_a - \bar{x}_b} = \frac{s_a^2}{n} + \frac{s_b^2}{n}$$

$$df = 2(n-1)$$

Paired *t*-test

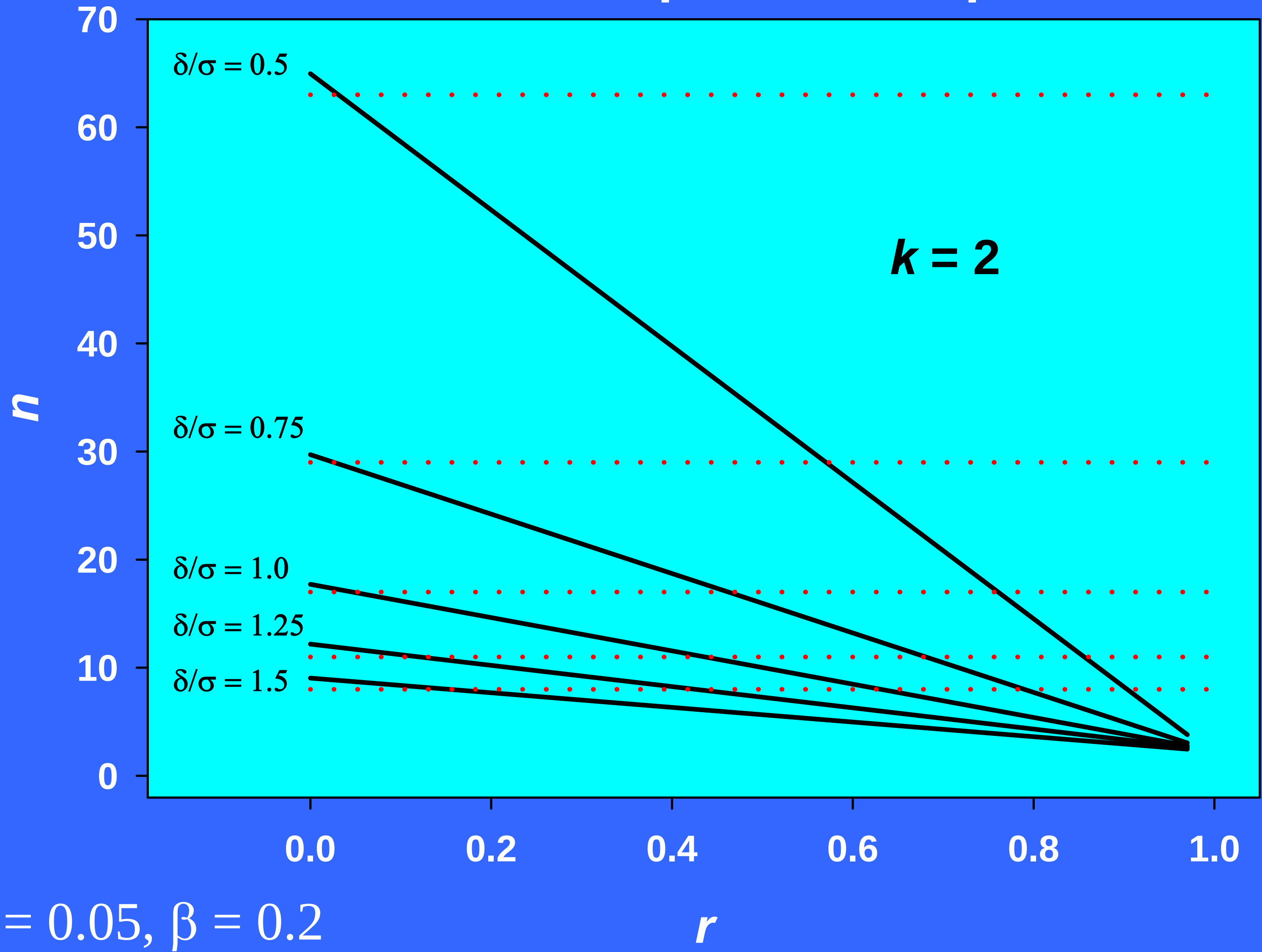
$$t = \frac{\bar{d} - u}{\sqrt{S^2_{\bar{d}}}}$$

$$s^2_{\bar{d}} = s^2_{\bar{x}_a - \bar{x}_b} = \frac{s_d^2}{n} = \frac{s_a^2}{n} + \frac{s_b^2}{n} - \frac{2 r_{ab} s_a s_b}{n}$$

$$\frac{-2\,\mathrm{cov}_{ab}}{n}$$

$$df = (n-1)$$

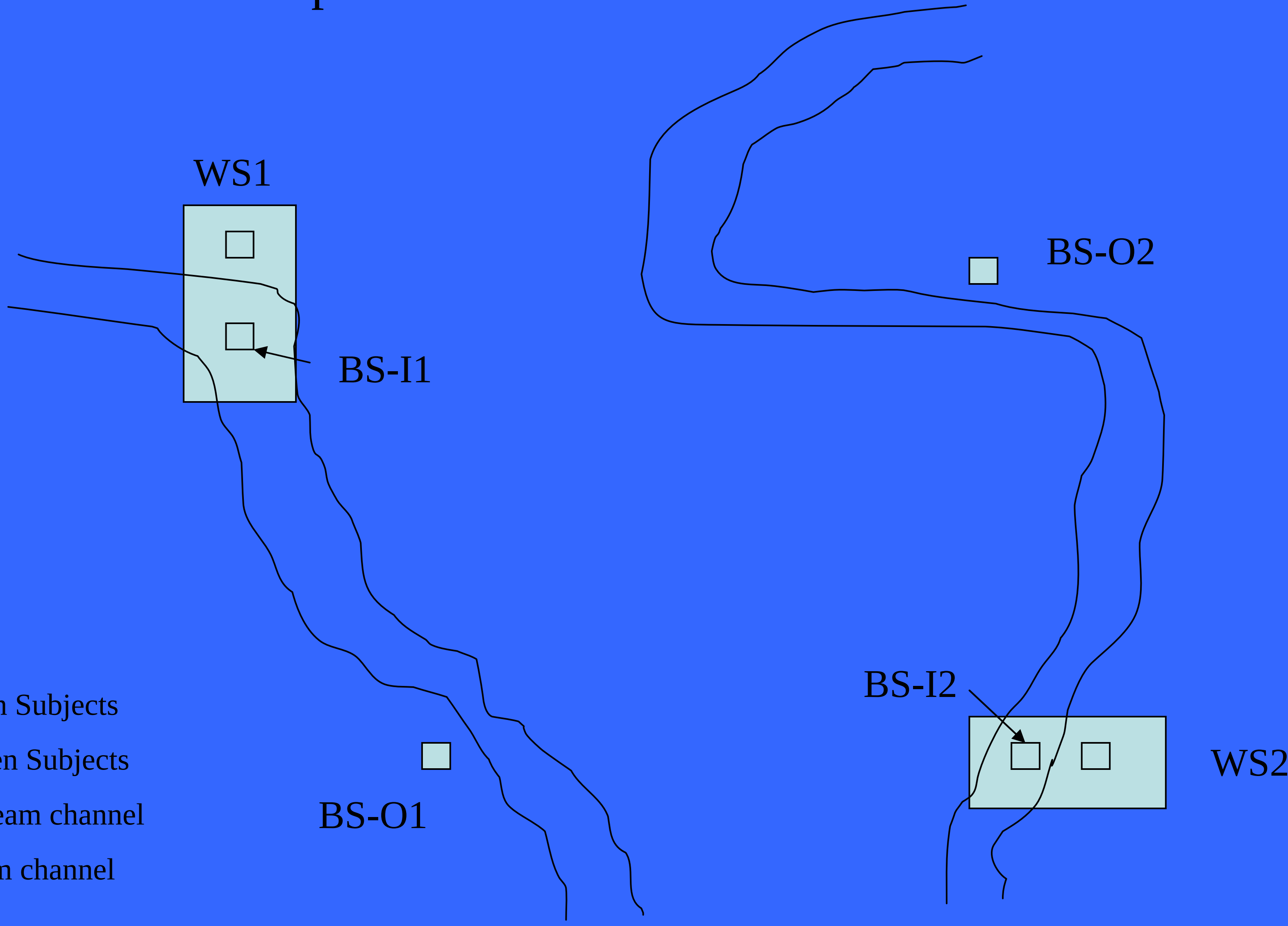
Paired versus Separate Groups *t*-test



$\alpha = 0.05$, $\beta = 0.2$

Estimating flowering plant species richness in upland regions and seasonal stream channels in Freeman Meadow (Lakes Basin, California)

Preliminary Data Collection
Meadow Example



WS= Within Subjects

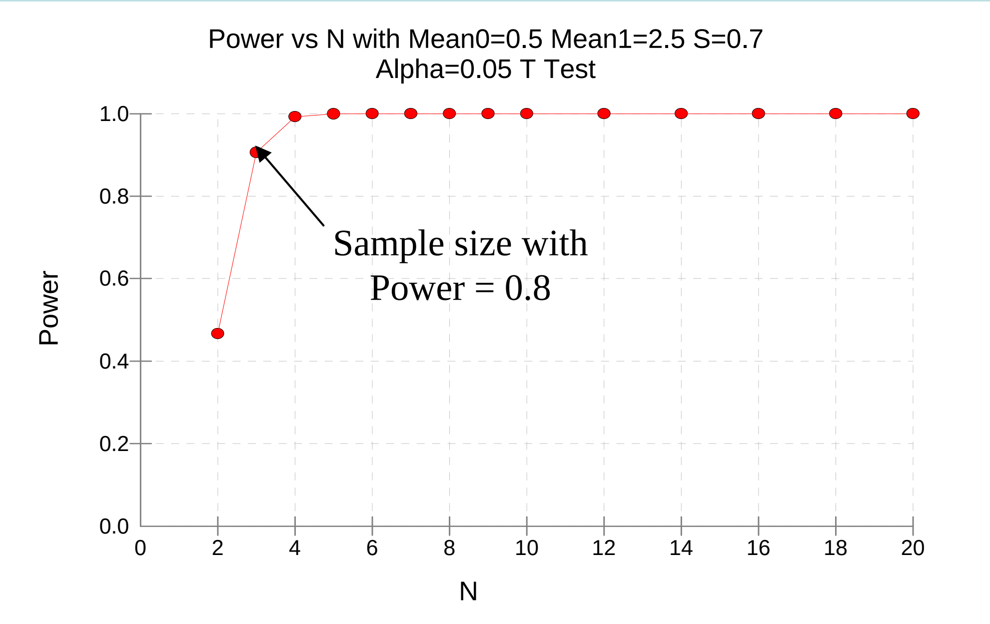
BS = Between Subjects

O = out of stream channel

I = In stream channel

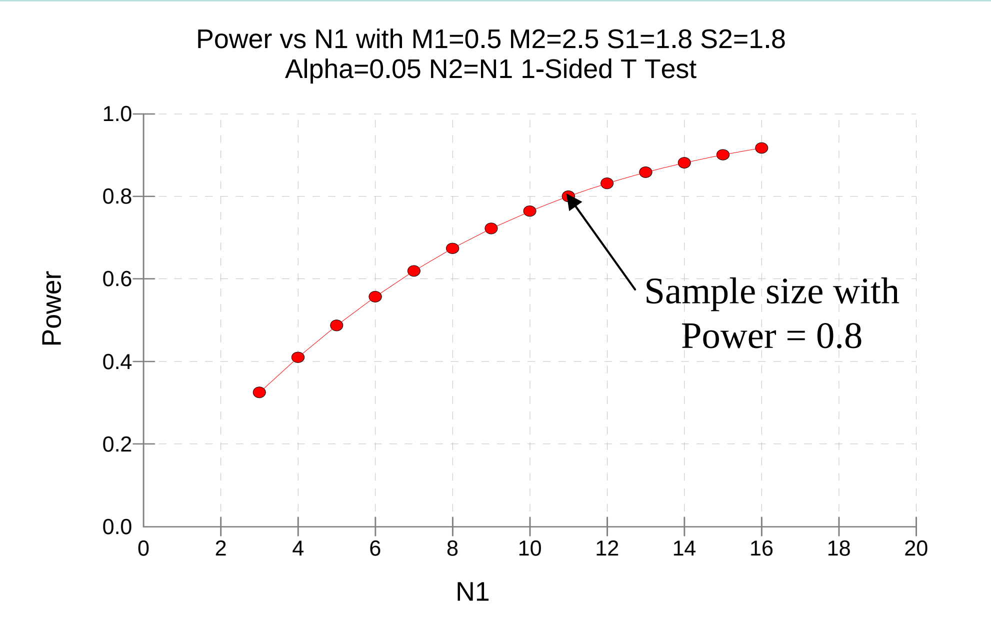
Sample Size Calculations for Meadow Example Using PASS

Within Subjects



Number of measurements required = 6

Between Subjects



Number of measurements required = 22

Difference between Within-subjects and Between-subjects ANOVA

Between Subjects

$$F = \frac{MS_{treat}}{MS_{error}}$$

$$dfs = \frac{k-1}{k(n-1)}$$

$$MS_{error} = \overline{\text{var}}$$

Within Subjects

$$F = \frac{MS_{treat}}{MS_{error}}$$

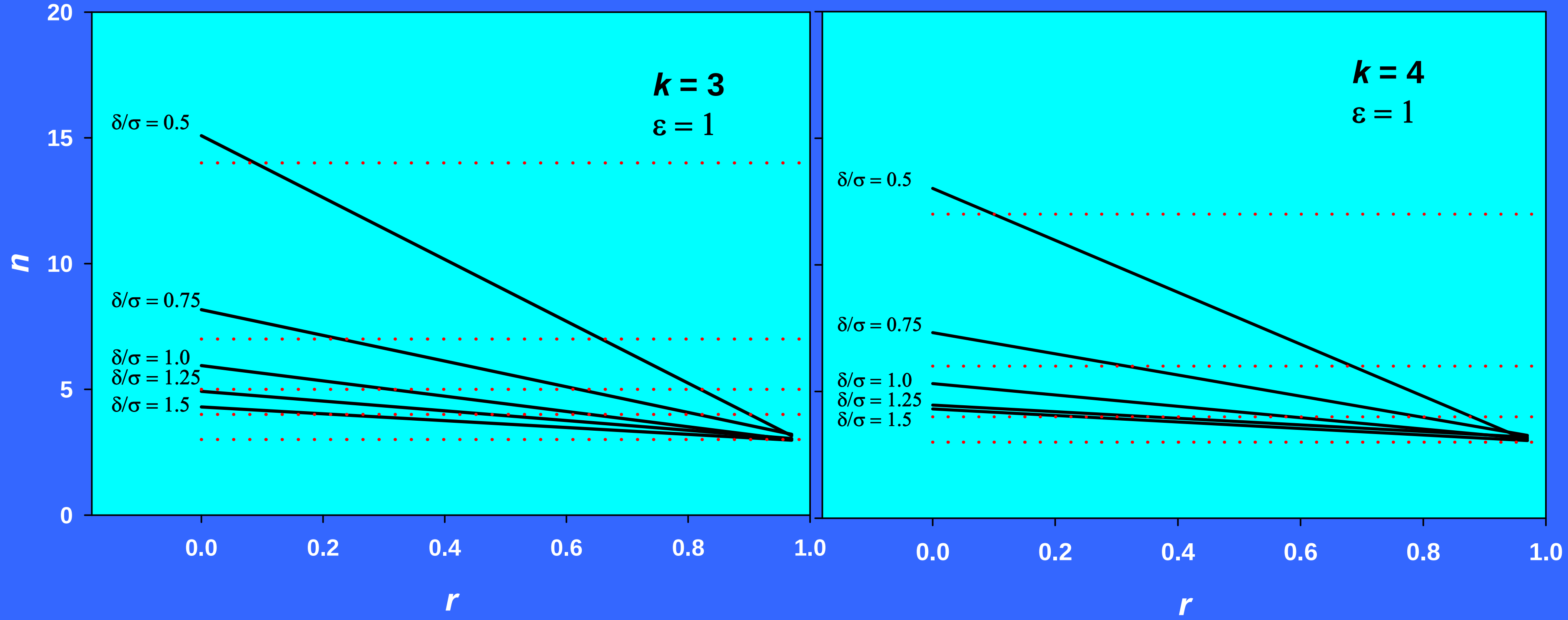
$$dfs = \frac{k-1}{(n-1)(k-1)}$$

$$MS_{error} = \overline{\text{var}} - \overline{\text{cov}}$$

$$r' = \frac{\overline{\text{cov}}}{\overline{\text{var}}}$$

$$MS_{error} = \overline{\text{var}}(1-r')$$

Within versus Between Subjects ANOVA



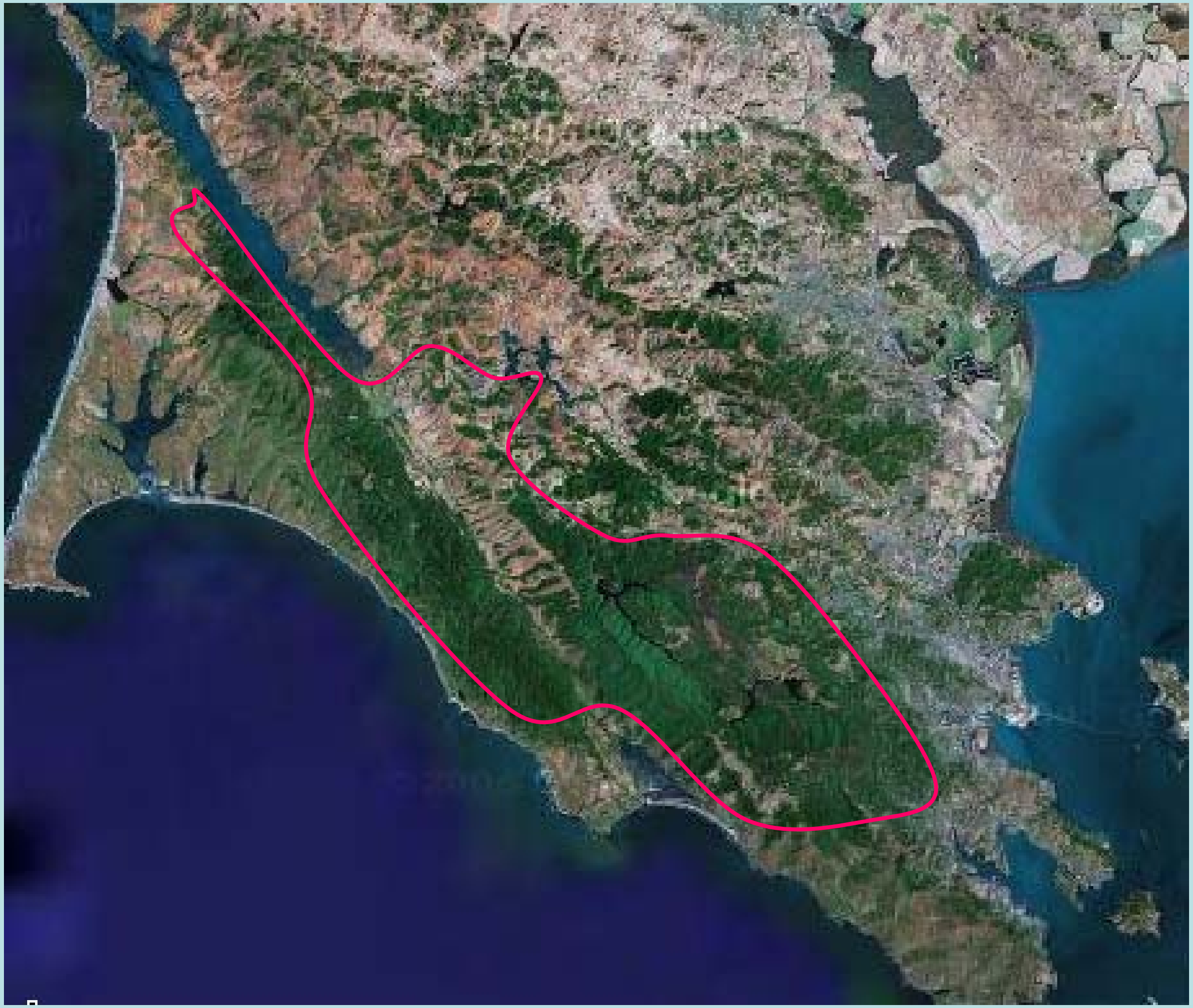
$\alpha = 0.05, \beta = 0.2$

Advantage of Within-subjects Design

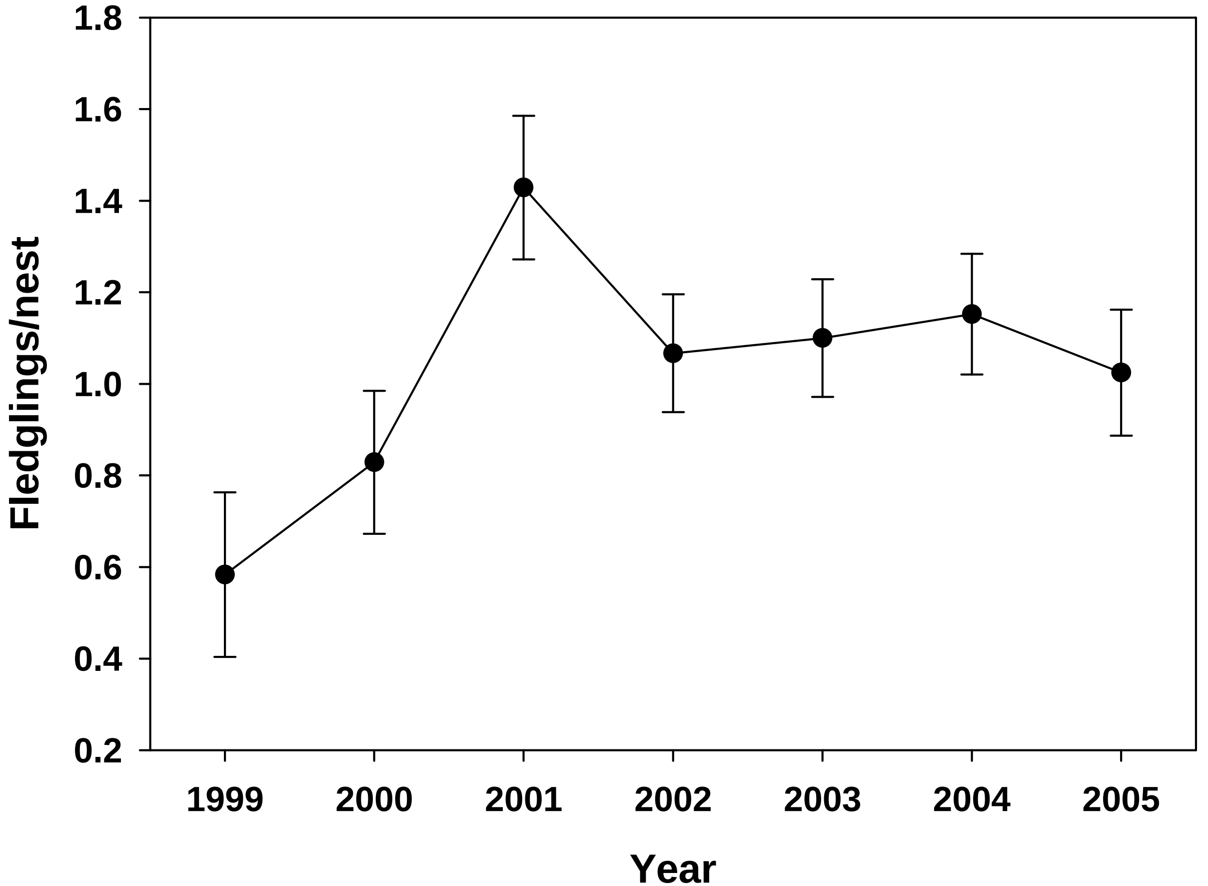
For modest levels of correlation within a subject, ***within subjects designs*** will be more powerful than the equivalent between subjects design.

When testing the same H_o against the same H_a (δ fixed), with the same α, β, ***within subjects designs*** are more efficient designs since σ will most likely be sufficiently smaller.

Northern Spotted Owl

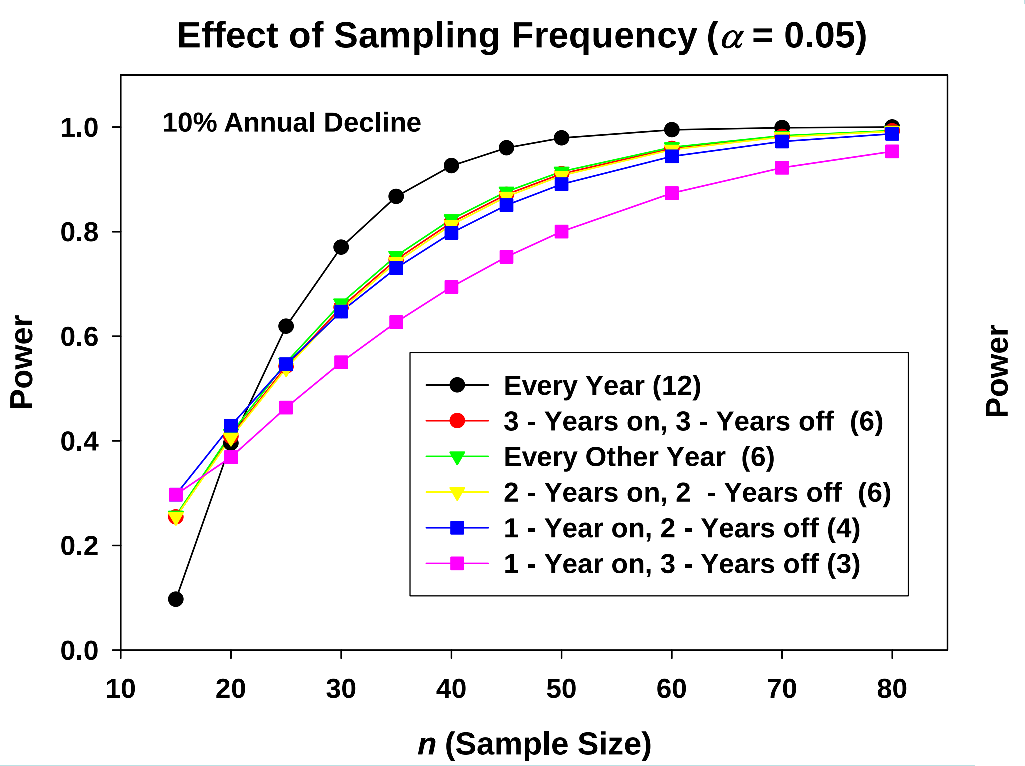


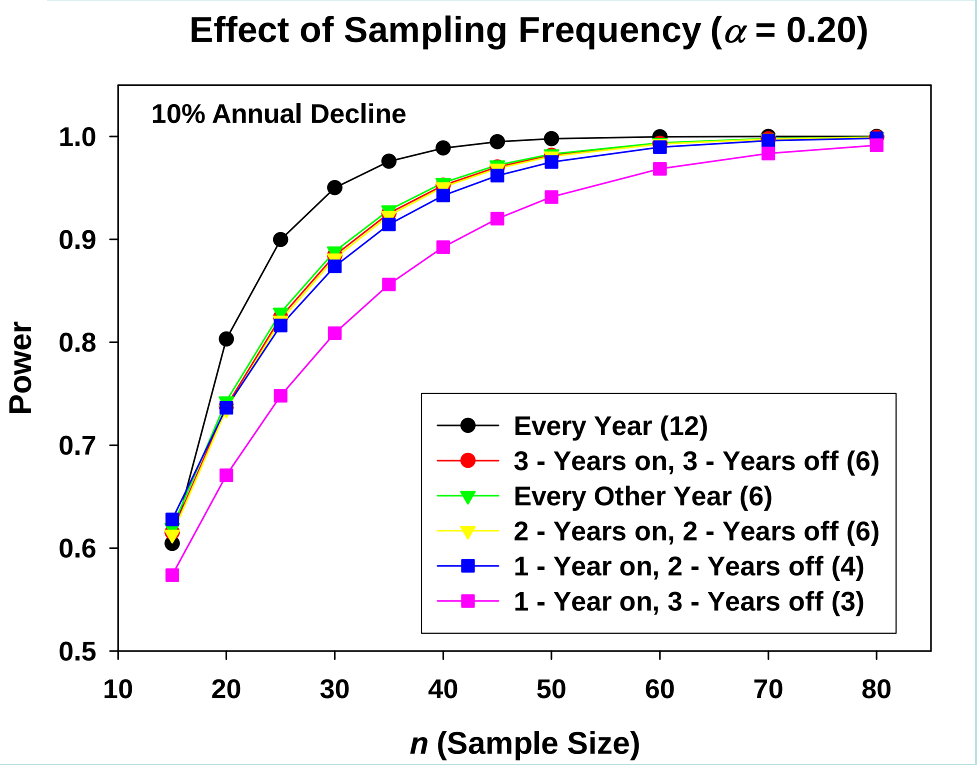
Northern Spotted Owl
Point Reyes National Seashore, Muir Woods and Golden Gate National Recreation Area



Spotted Owl Fecundity Sample Size Calculation

Effect of varying sampling frequency





Trade off between Sampling Intensity and Frequency

- Achieve approximately equal power using fewer sites sampled more years
- Or more sites sampled fewer years
- For Spotted Owl more cost effective to sample more sites fewer years (20/yr x 12 yrs = 240 surveys, 25/yr x 4 yrs = 100 surveys

In Search of the Optimal Quadrat Size and Shape – What is Optimal?

- Sampling variance
- Sampling time/replicate
- Transit time/replicate
- Minimizes bias

Field effort = (p x n) + (d x t)

p = processing time per quadrat
n = sample size (number of quadrats)
d = total distance traveled between quadrats
t = rate of movement between quadrats

Field	**Transit rate through vegetation (1.1125 m/sec)**	**t**
	Processing time for each size quadrat	**p**
Computer	**Estimate *σ* from simulated population** **Use *σ* to calculate estimate of required sample size**	**n**
	For estimated n determine shortest distance between quadrats	**d**

Field

- Using a within-subjects design (n = 4) we estimated the time required to set-up and sample a variety of quadrat sizes and shapes (p)
- We estimated the time required to walk a variety of distances through vegetation on a fixed bearing (t)

Time Costs for Different Size and Shape Quadrats – Carmen Valley, CA

Computer

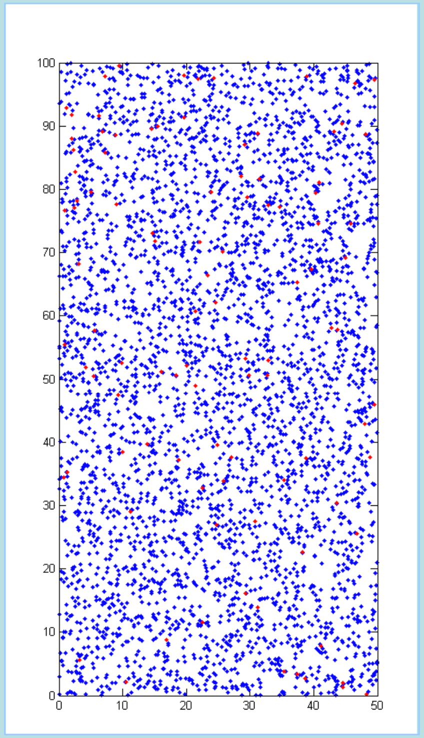
- We estimated the between quadrat variance in plant density for each size and shaped quadrat for each simulated plant population

- We used the estimated variance to calculate the sample size (n) required to place a (1 - α) confidence interval on the mean population density that was no greater than $\pm 0.3 \times \mu$

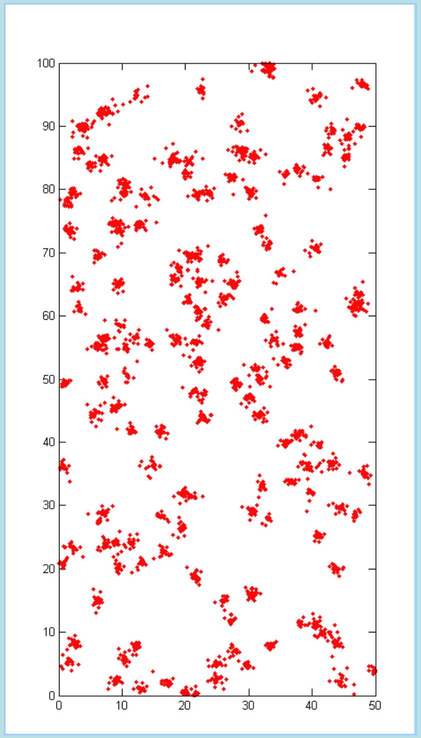
Computer

Simulated plant populations with a mean density of 0.8 plants/m^2 in a 50 x 100 m region with different spatial structures

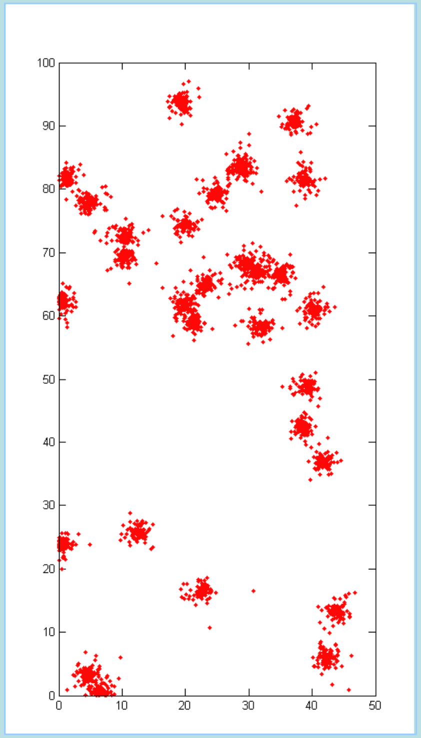
Aggregation



Random = 1



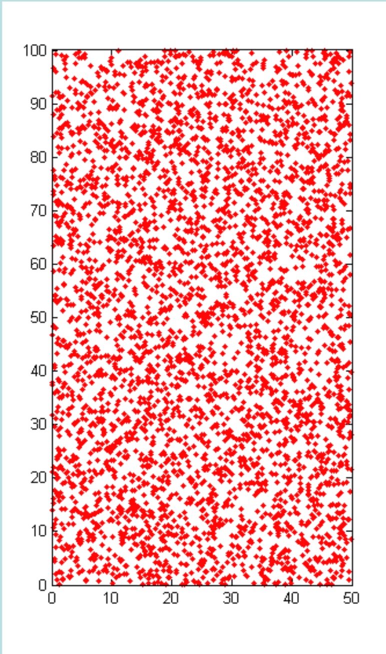
Low aggregation = 5



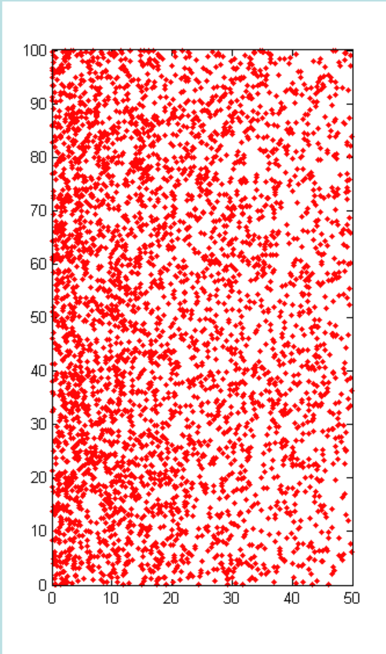
High aggregation = 10

Computer

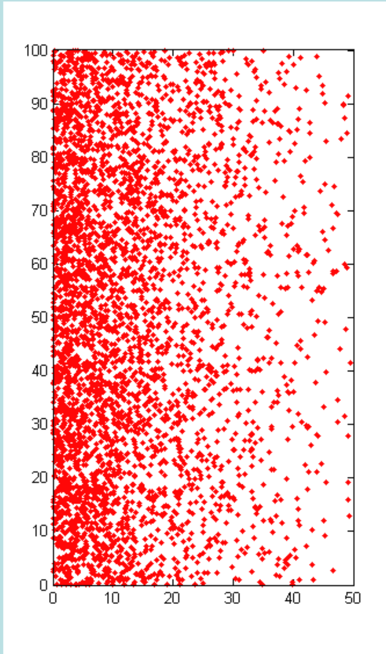
Gradient



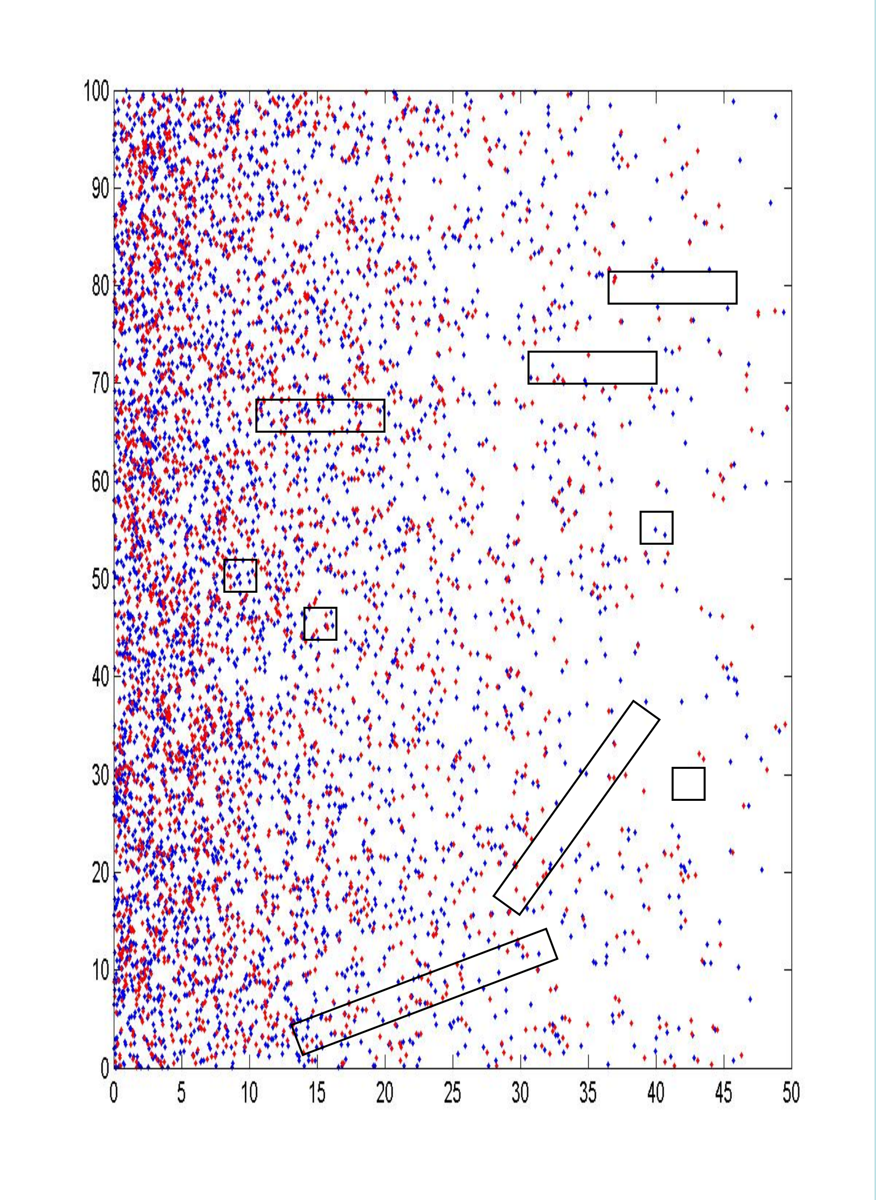
No Gradient



Weak Gradient



Strong Gradient



Shape - Rectangularity

Length/Width

Square = 1

0.25 x 50m = 200

Rotation

0^o along gradient

90^o perpendicular to gradient

From Zar p. 105

$$P(\bar{x} - t_{(1-\alpha/2,n-1)} \, s/\sqrt{n} \le \mu \le \bar{x} + t_{(1-\alpha/2,n-1)} \, s/\sqrt{n}) = 1-\alpha$$

$$n = \frac{s^2 t^2_{(1-\alpha/2,n-1)}}{d^2}$$

$$n = \frac{s^2 t^2_{(1-\alpha/2,n-1)}}{0.3\bar{x}^2}$$

Conditional Probability of having the correct width interval (*W*), given that it is a 1 - *α* interval (*V*) (Jiroutek et al 2003)

$$\Pr(W \mid V) = \Pr\left\{\left[\left(U - L\right) \le \delta\right] \mid \left(L \le \theta \le U\right)\right\} = \frac{\Pr\left(W \cap V\right)}{\Pr\left(V\right)}$$

Where $\Pr\left(V\right) \ge \left\{1 - \alpha\right\}$

$$\Pr\left(W\right) = \left\{\left(U - L\right) \le \delta\right\}$$

- **Let U and L indicate the upper and lower confidence interval bounds, respectively**

- **Let *δ* indicate the desired Confidence Interval width**

$$\Pr(W \mid V) \geq \int_0^{x_1} \left[\left\{ \Phi\left(c_1\sqrt{x}\right) - \Phi\left(-c_1\sqrt{x}\right) \right\} \frac{f_{x^2}\left(x; v_e\right)}{(1-\alpha)} \right] dx$$

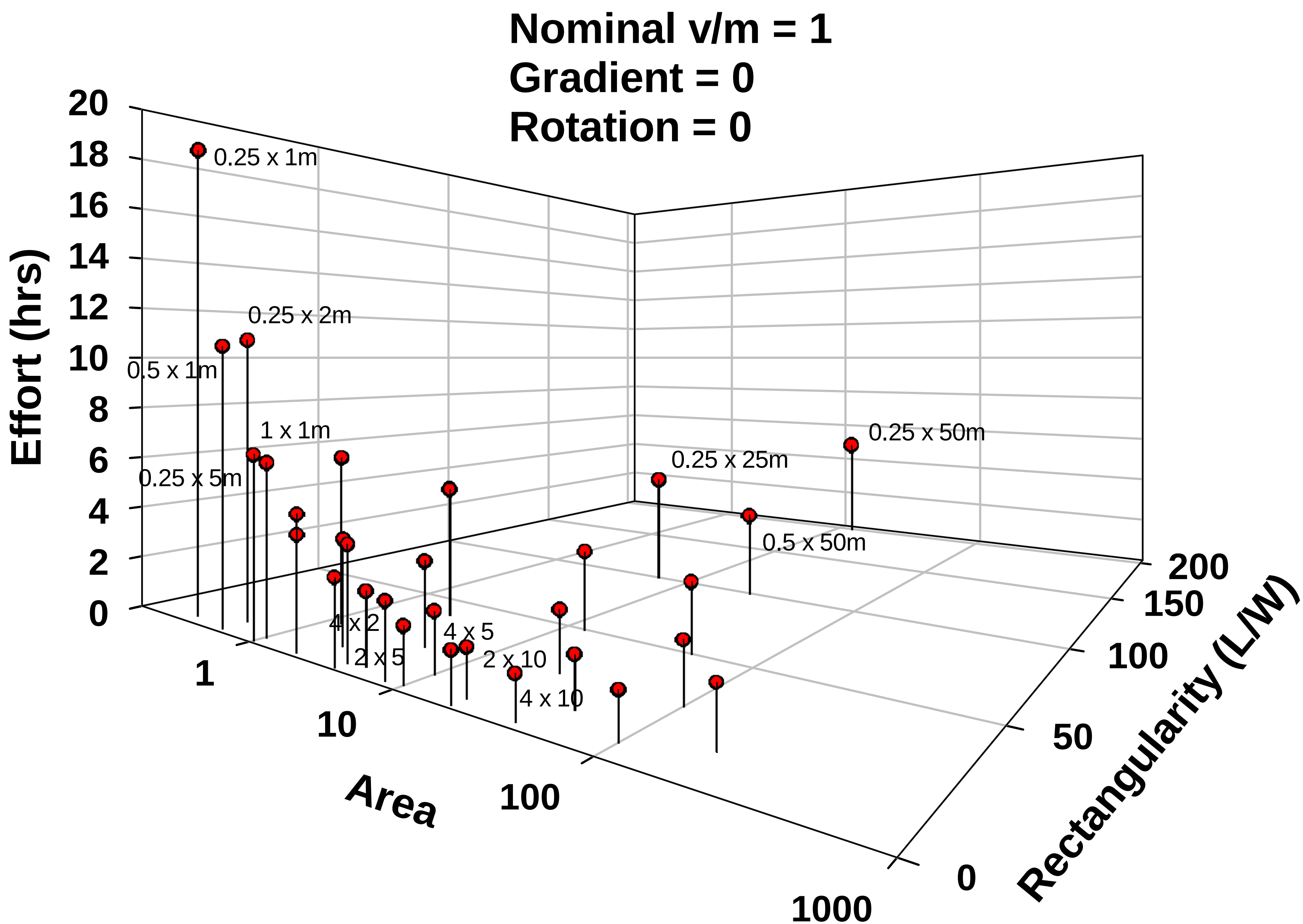
Where $\Phi(\bullet)$ **indicates the CDF of a standard normal variate**

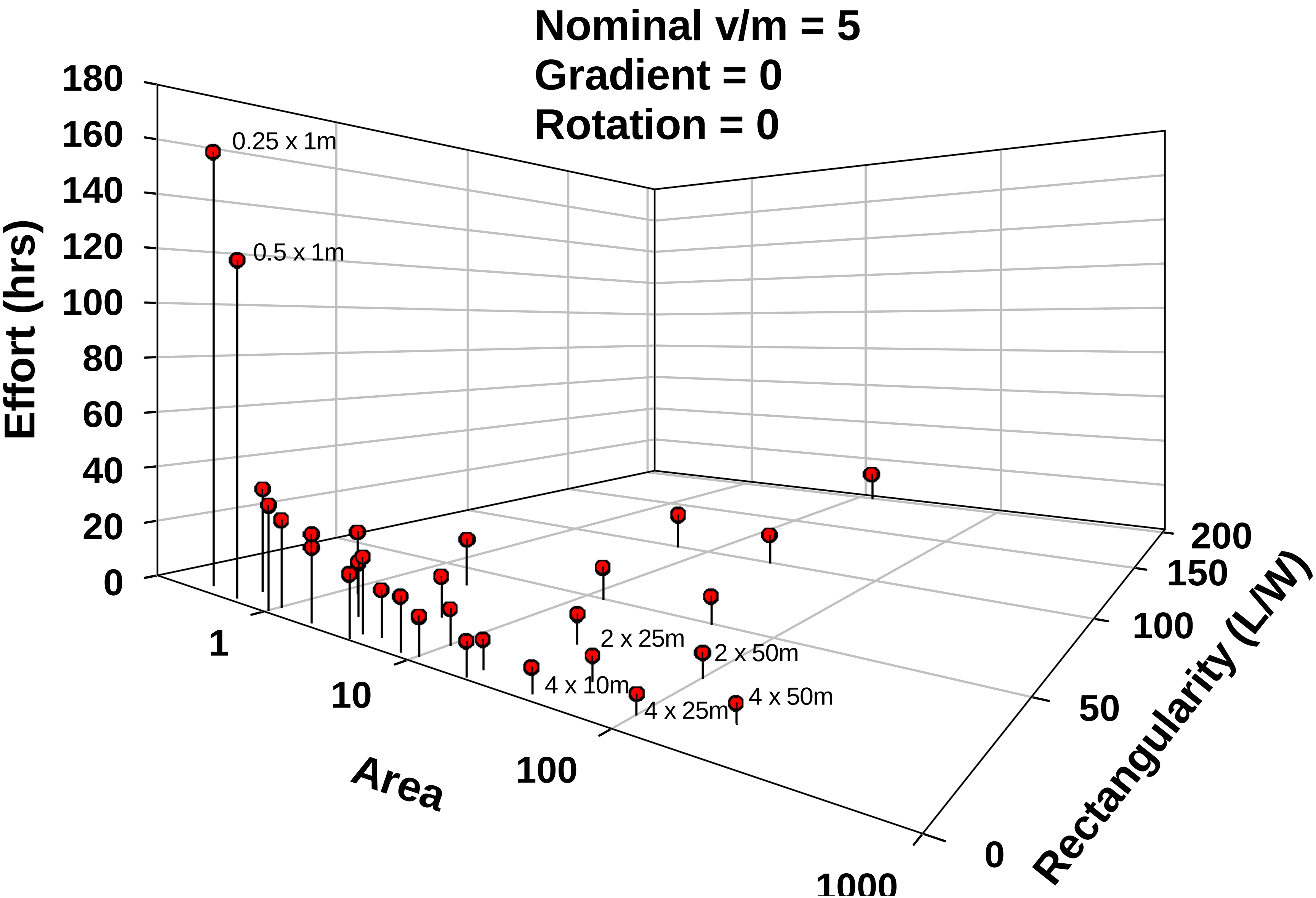
$f_{x^2}\left(x; v_e\right)$ **indicates a central chi-square density function with** v_e **degrees of freedom**

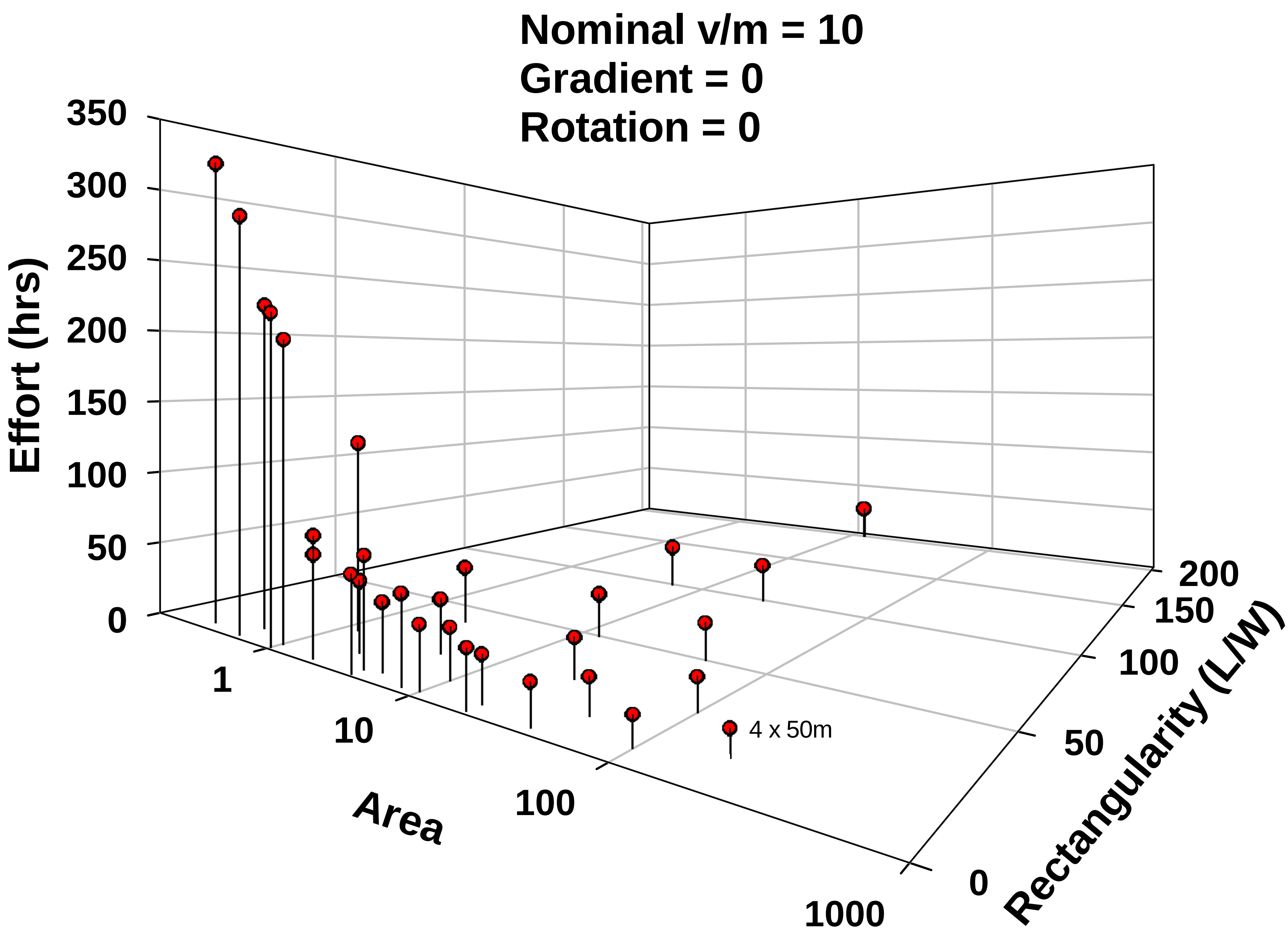
$$c_1 = \sqrt{F_{crit} / v_e} \qquad x_1 = \frac{v_e \delta^2}{\left(4\sigma^2 F_{crit} / N\right)}$$

F_{crit} **indicates the critical value from a central *F* distribution with 1 numerator and** v_e **denominator degrees of freedom**

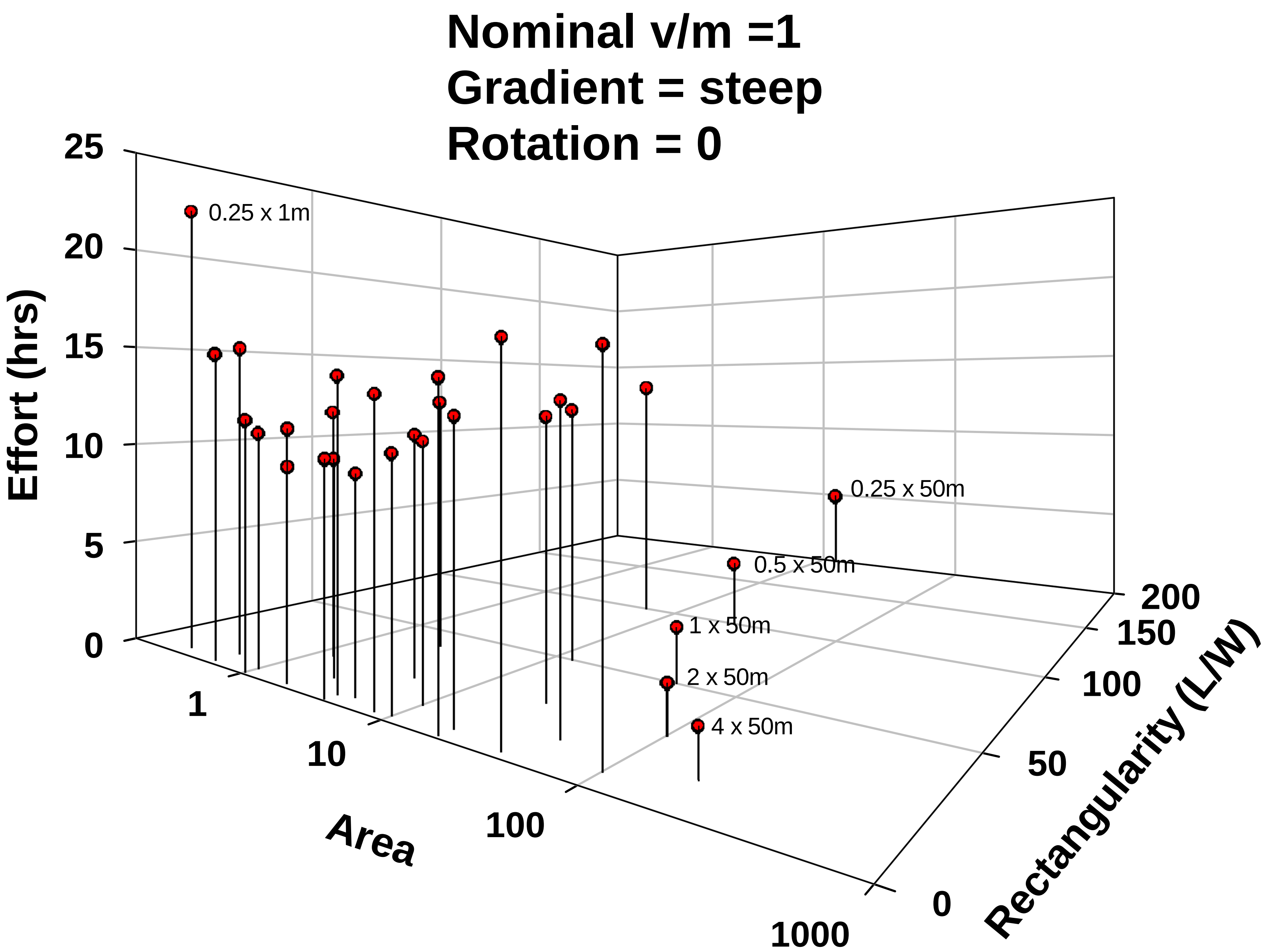
$v_e = N - r$ **where N indicates the estimated sample size and *r* = 2 for a two-sided test**

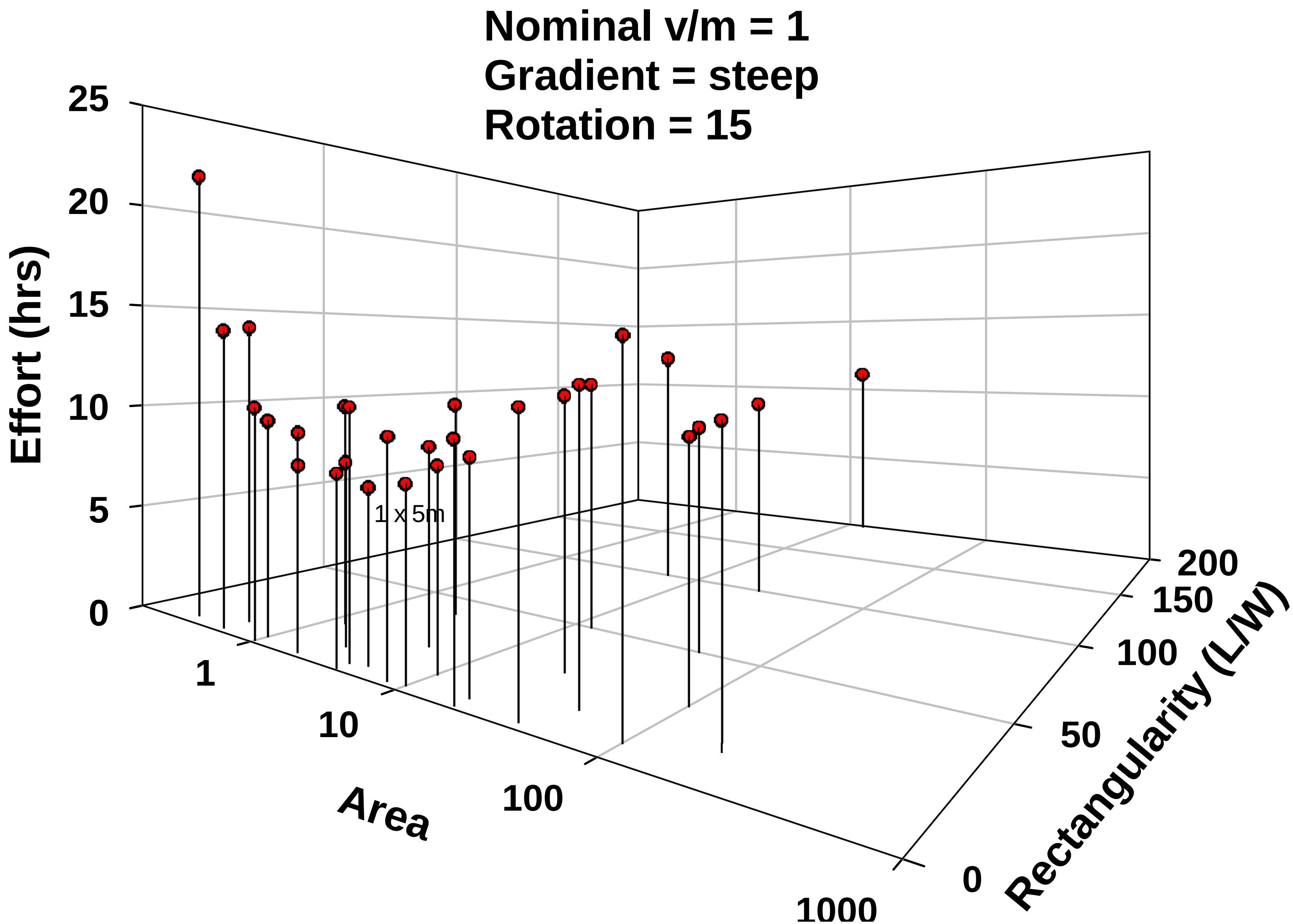

Nominal v/m = 1
Gradient = 0
Rotation = 0
Effort (hrs)
20
18
16
14
12
10
8
6
4
2
0
0.25 x 1m
0.25 x 2m
0.5 x 1m
1 x 1m
0.25 x 5m
4 x 2
2 x 5
4 x 5
2 x 10
4 x 10
0.25 x 25m
0.25 x 50m
0.5 x 50m
Area
1
10
100
1000
Rectangularity (L/W)
0
50
100
150
200

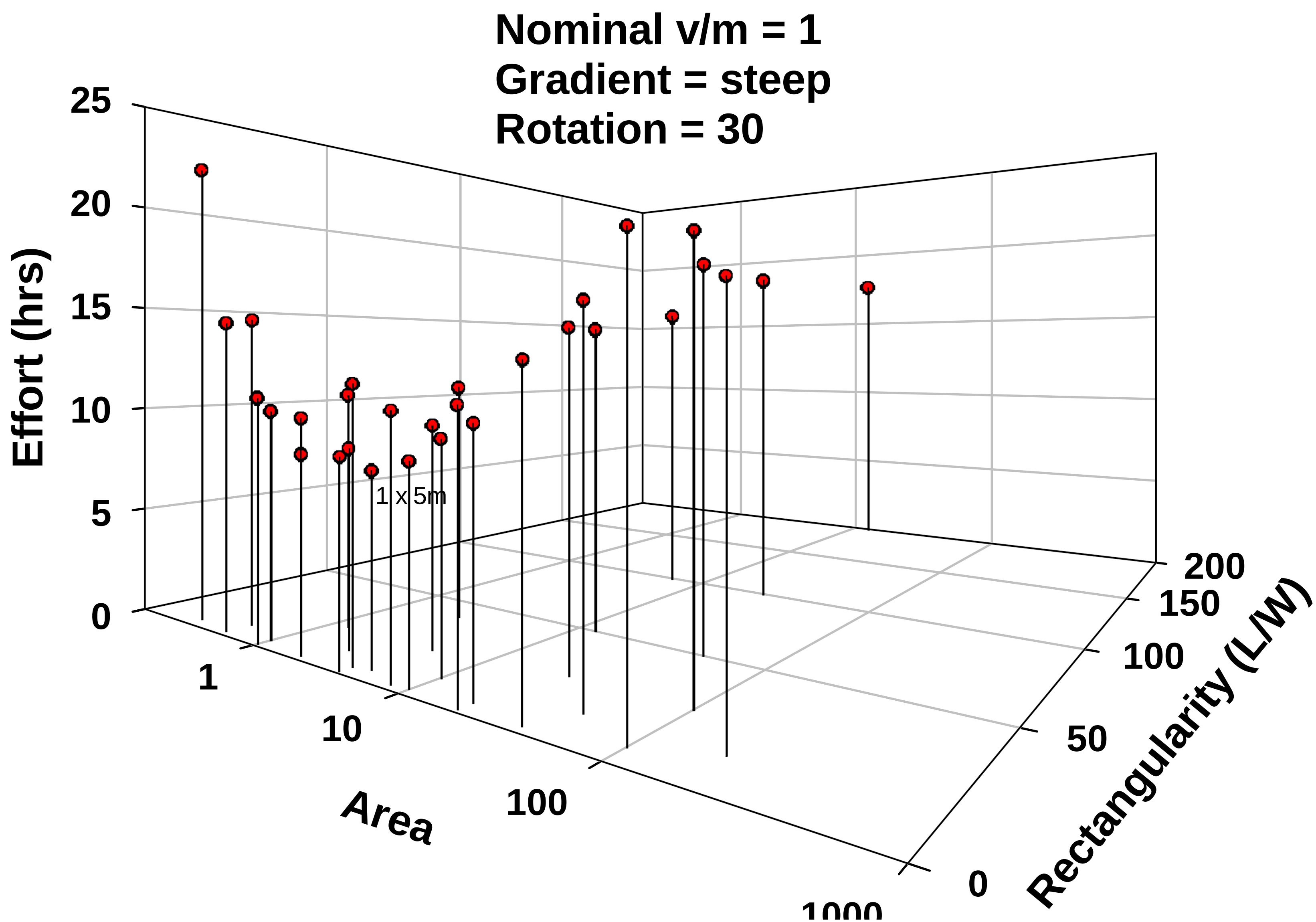

Nominal v/m = 5
Gradient = 0
Rotation = 0
Effort (hrs)
180
160
140
120
100
80
60
40
20
0
0.25 x 1m
0.5 x 1m
2 x 25m
2 x 50m
4 x 10m
4 x 25m
4 x 50m
Area
1
10
100
1000
Rectangularity (L/W)
0
50
100
150
200



Nominal v/m = 10
Gradient = 0
Rotation = 0
Effort (hrs)
350
300
250
200
150
100
50
0
1
10
100
1000
Area
0
50
100
150
200
Rectangularity (L/W)
4 x 50m


Nominal v/m =1
Gradient = steep
Rotation = 0
Effort (hrs)
25
20
15
10
5
0
0.25 x 1m
0.25 x 50m
0.5 x 50m
1 x 50m
2 x 50m
4 x 50m
Area
1
10
100
1000
Rectangularity (L/W)
0
50
100
150
200


Nominal v/m = 1
Gradient = steep
Rotation = 15
Effort (hrs)
0
5
10
15
20
25
1 x 5m
Area
1
10
100
1000
Rectangularity (L/W)
0
50
100
150
200


Nominal v/m = 1
Gradient = steep
Rotation = 30
Effort (hrs)
25
20
15
10
5
0
1 x 5m
1
10
100
1000
Area
0
50
100
150
200
Rectangularity (L/W)

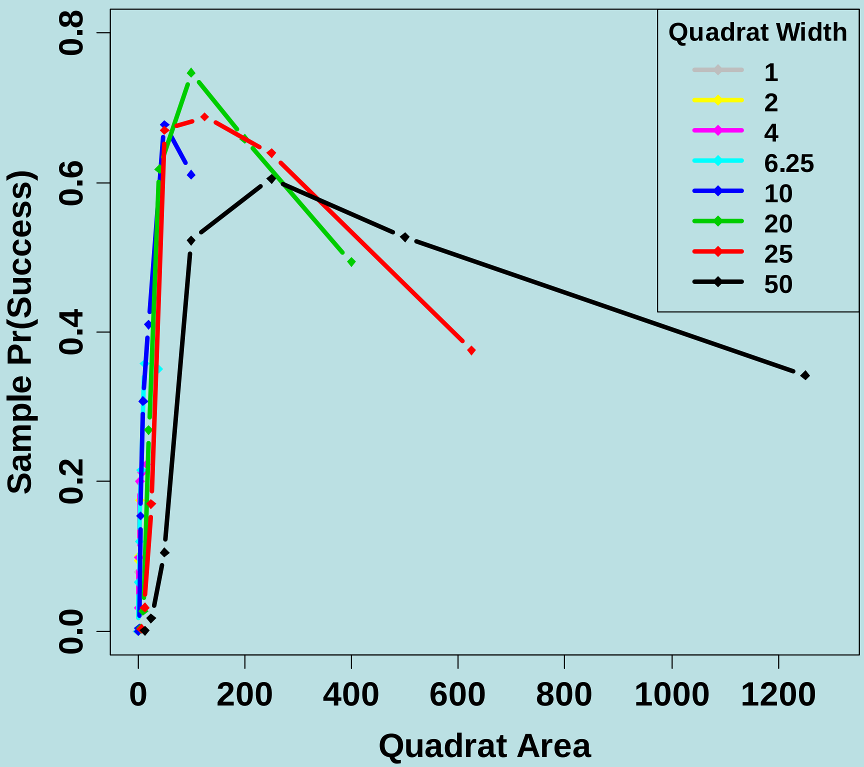
What are the attributes of a quadrat that is optimal under all conditions?

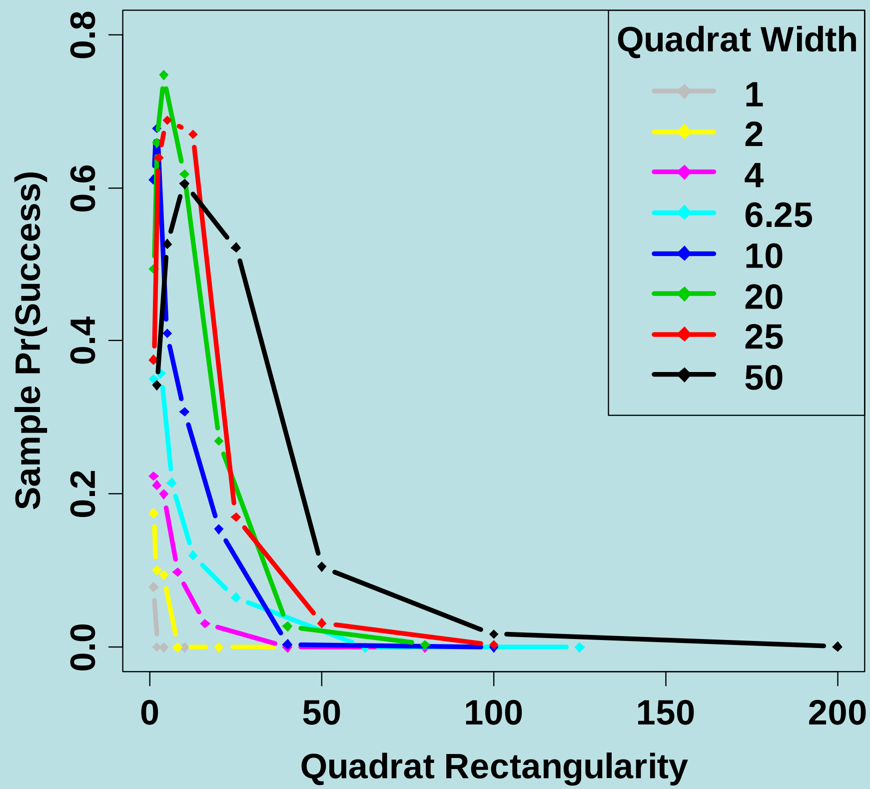
- Used logistic regression; defining a “success” as a quadrat size and shape with effort in the lowest 20% for all quadrats (N=3600 for 36 combinations of size shape, gradient and rotation)

- Optimal Quadrat Size (56-86m^2)

- Optimal Quadrat Shape (L/W = 4/5)

Results of Logistic Regression Analysis





Summary

- Preliminary data is essential to designing cost effective sampling programs
- Within-subject and reduced-frequency sampling designs are generally more cost effective
- Field data combined with computer simulations can provide concrete guidance on sampling methods and study design

Think and calculate before you go to the field

Acknowledgements



Lucas Bohnett **Anastasia Chavez** **Jen Carah**

Numerous students from Biology 315 - Field Methods in Ecology

National Science Foundation
WHERE DISCOVERIES BEGIN

SFSU
San Francisco State University