

## ESTATÍSTICA, DELINEAMENTO & AMOSTRAGEM

- Estatística → estimativa da incerteza
- Amostragem → representatividade da estimativa
- Delineamento → esforço amostral e disposição das amostras no espaço e no tempo.

## AMOSTRAGEM E DELINEAMENTO

- Determinação da unidade amostral (ou amostrador)
- Determinação do esforço amostral
- Alocação das amostras no espaço e no tempo
- Seleção dos tratamentos (=variáveis analisadas)
- Avaliação do número de níveis de cada tratamento
- Análise de custos (projeto piloto)

# AMOSTRAGEM em ESTUDOS BIOLÓGICOS

- Unidade Amostral (Réplica)

- ✓ Tamanho da unidade amostral

- ✓ Forma da unidade amostral

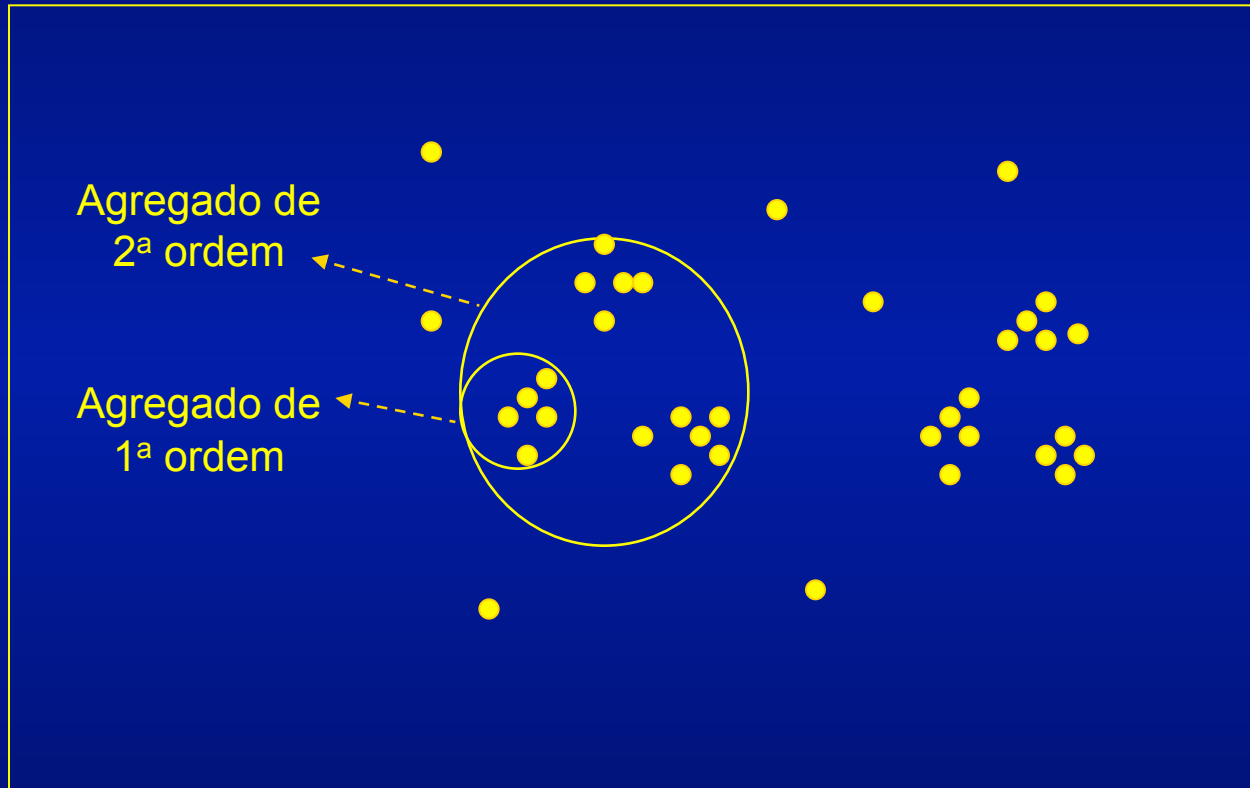
- ✓ Número mínimo de réplicas

- ✓ Alocação espacial e temporal das amostras

# DISTRIBUIÇÃO ESPACIAL AGREGAÇÃO (= MANCHAS)

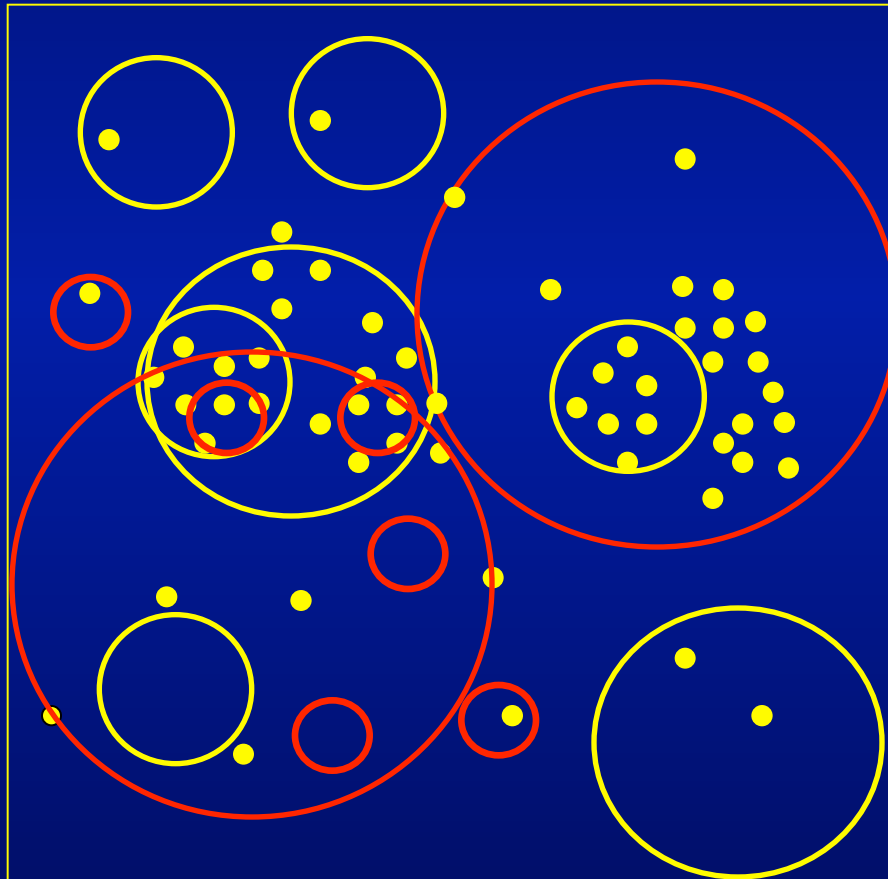
- Objetivo do estudo
  - ✓ Variável explicativa
- Ruído
  - ✓ Variáveis não incluídas (< escala)
  - ✓ Erro amostral

# AGREGAÇÃO (= MANCHAS)

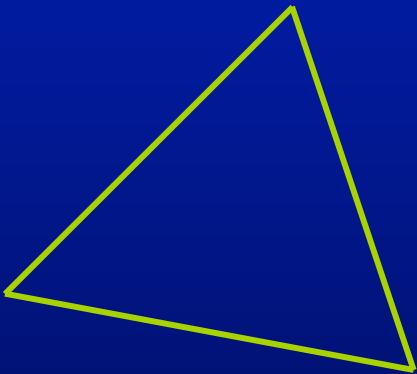
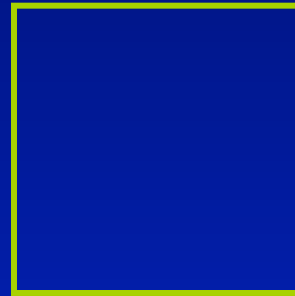
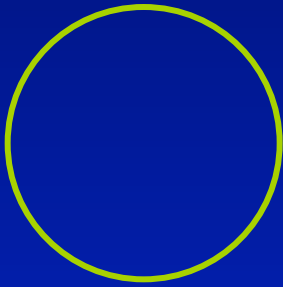


# TAMANHO DA UNIDADE AMOSTRAL

- Unidade Amostral  $>$  Mancha
- Unidade Amostral  $\approx$  Mancha
- Unidade Amostral  $<$  Mancha



# FORMA DA UNIDADE AMOSTRAL



## NÚMERO DE UNIDADES AMOSTRAIS (RÉPLICAS)

- Custo/Benefício
- N° Réplicas × Tratamentos
- N° Réplicas × Precisão (% da média ou diferença absoluta)



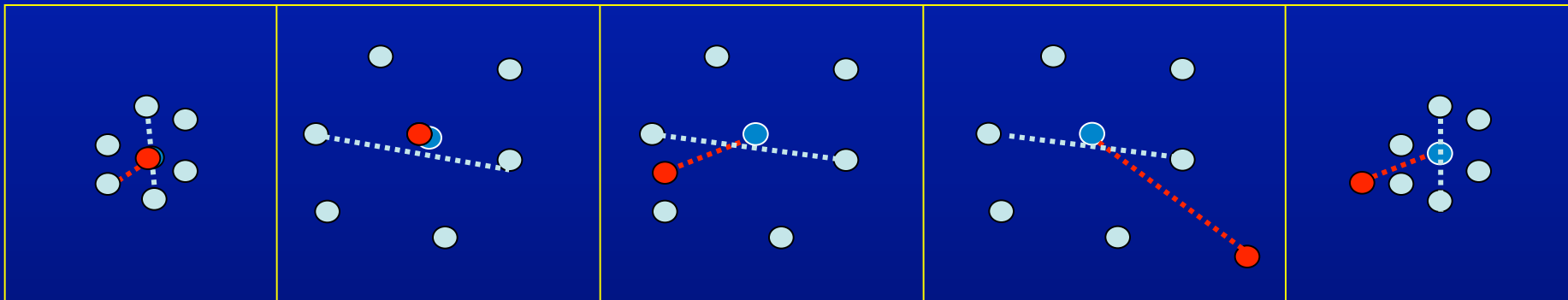
# PRECISÃO & ACURÁCIA

## (1) Precisão

- Repetibilidade dos dados
- Avaliação do erro amostral

## (2) Acurácia

- Avaliação do erro de mensuração (viés, dependência, 'confounding')



● Real

..... Acurácia

● Amostra

..... Precisão

● Média

## PRECISÃO & ACURACIA

Cracas por metro quadrado

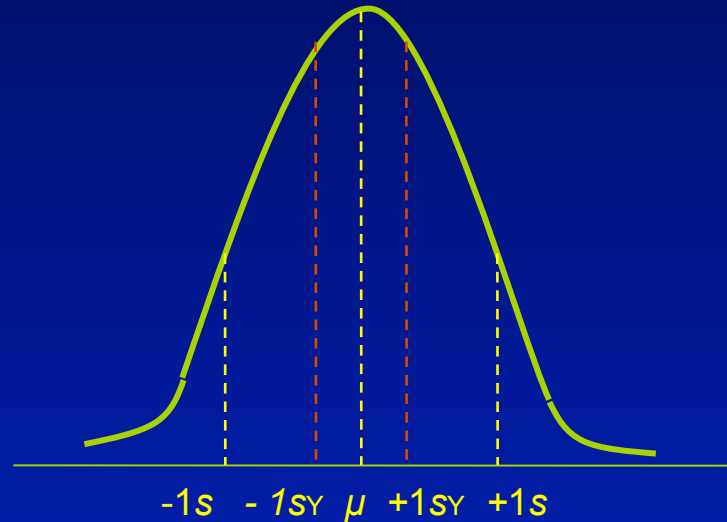
100, 102, 104, 101, 103

MÉDIA = 102

ERRO PADRÃO DA MÉDIA = 2

Erro padrão da média  
ou  
Desvio padrão da média  
ou  
Erro Padrão

$$s_Y = s / \sqrt{n}$$

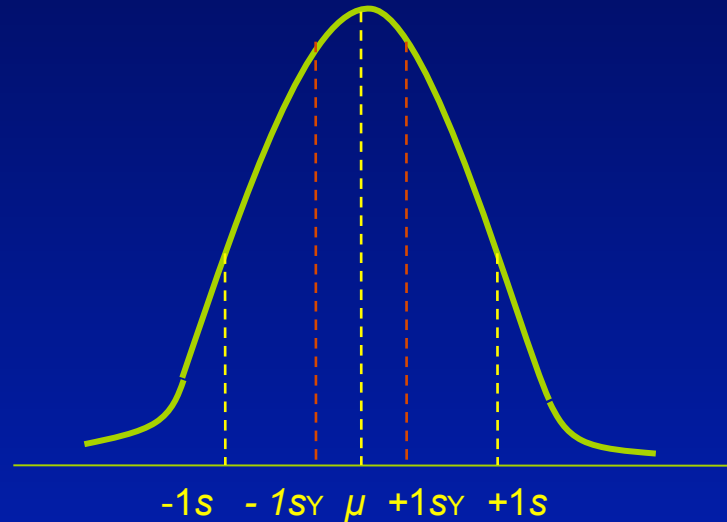


2 3 5 9 4 5 7 10 2 12 5 4 6 8 9 2

$$m = 5,8 \quad s = 3,08 \quad s_Y = 0,77$$

Erro padrão da média  
ou  
Desvio padrão da média  
ou  
Erro Padrão

$$s_Y = s / \sqrt{n}$$



2 3 5 9 4 5 7 10 2 12 5 4 6 8 9 2

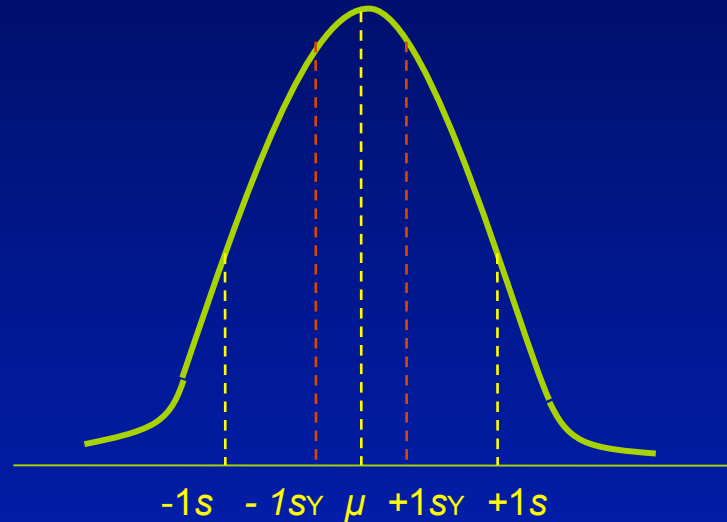
$$m = 5,8$$

$$s = 3,08$$

$$s_Y = 0,77$$

Erro padrão da média  
ou  
Desvio padrão da média  
ou  
Erro Padrão

$$s_Y = s / \sqrt{n}$$



2 3 5 9 4 5 7 10 2 12 5 4 6 8 9 2

Solução Analítica

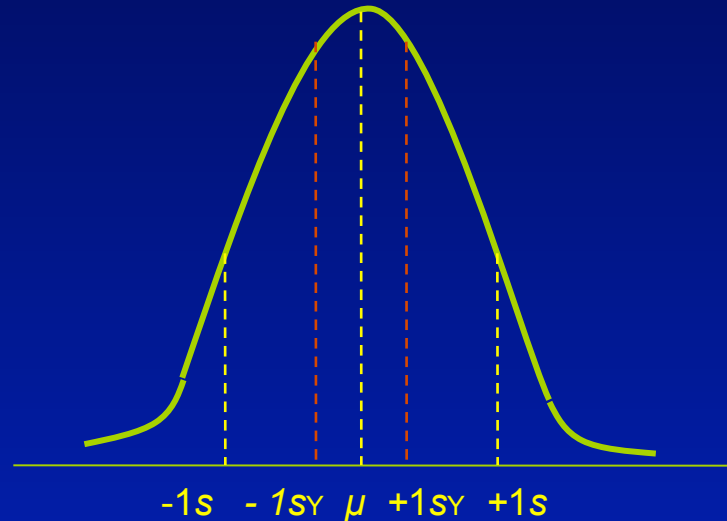
$$m = 5,8$$

$$s = 3,08$$

$$s_Y = 0,77$$

Erro padrão da média  
 ou  
 Desvio padrão da média  
 ou  
 Erro Padrão

$$s_Y = s / \sqrt{n}$$



2 3 5 9 4 5 7 10 2 12 5 4 6 8 9 2

Solução Analítica

$$m = 5,8$$

$$s = 3,08$$

$$s_Y = 0,77$$

$$X_1 = 4,7$$

$$X_2 = 6$$

$$X_3 = 6,2$$

Solução Empírica

$$\bar{X} = 5,6$$

$$s_Y = 0,80$$

## PRECISÃO & ACURÁCIA

Cracas por metro quadrado

100, 102, 104, 101, 103

MÉDIA = 102

ERRO PADRÃO DA MÉDIA = 2

PRECISÃO =  $2/102 = 1,96\%$

VALOR REAL = 110

ACURÁCIA =  $(110 - 102)/110 = 7,8\%$

# PRECISÃO & ACURÁCIA

Cracas por metro quadrado

100, 102, 104, 101, 103

MÉDIA = 102

ERRO PADRÃO DA MÉDIA = 2

PRECISÃO =  $2/102 = 1,96\%$

PRECISÃO = ERRO PADRÃO / MÉDIA

$$P = S_Y / m$$



Precisão (P) = Erro padrão da média / média

$$P = \frac{S_Y}{m}$$

$$S_Y = \frac{s}{\sqrt{n}}$$

Precisão (P) = Erro padrão da média / média

$$P = \frac{S_Y}{m}$$

$$S_Y = \frac{S}{\sqrt{n}}$$

$$P = \left(\frac{S}{\sqrt{n}}\right)/m$$

$\leftrightarrow$

$$m \cdot P = \frac{S}{\sqrt{n}}$$

Precisão (P) = Erro padrão da média / média

$$P = \frac{S_Y}{m}$$

$$S_Y = \frac{S}{\sqrt{n}}$$

$$P = \left(\frac{S}{\sqrt{n}}\right)/m$$

$\leftrightarrow$

$$m \cdot P = \frac{S}{\sqrt{n}}$$

$$\sqrt{n} = \frac{S}{m \cdot P}$$

$\leftrightarrow$

$$n = \left(\frac{S}{m \cdot P}\right)^2$$

Precisão (P) = Erro padrão da média / média

$$P = \frac{S_Y}{m}$$

$$S_Y = \frac{S}{\sqrt{n}}$$

$$P = \left(\frac{S}{\sqrt{n}}\right)/m \quad \leftrightarrow \quad m \cdot P = \frac{S}{\sqrt{n}}$$

$$\sqrt{n} = \frac{S}{m \cdot P} \quad \leftrightarrow \quad n = \left(\frac{S}{m \cdot P}\right)^2$$

Exemplo:

$$m = 120 \quad s = 45 \quad P = 20\%$$

$$n = (45 / 120 \cdot 0,2)^2$$

$$n = (45 / 24)^2$$

Precisão (P) = Erro padrão da média / média

$$P = \frac{S_Y}{m}$$

$$S_Y = \frac{S}{\sqrt{n}}$$

$$P = \left(\frac{S}{\sqrt{n}}\right)/m \quad \leftrightarrow \quad m \cdot P = \frac{S}{\sqrt{n}}$$

$$\sqrt{n} = \frac{S}{m \cdot P} \quad \leftrightarrow \quad n = \left(\frac{S}{m \cdot P}\right)^2$$

Exemplo:

$$m = 120 \quad s = 45 \quad P = 20\%$$

$$n = (45 / 120 \cdot 0,2)^2$$

$$n = (45 / 24)^2$$

$$n = 3,5 \quad (68\%)$$

ou

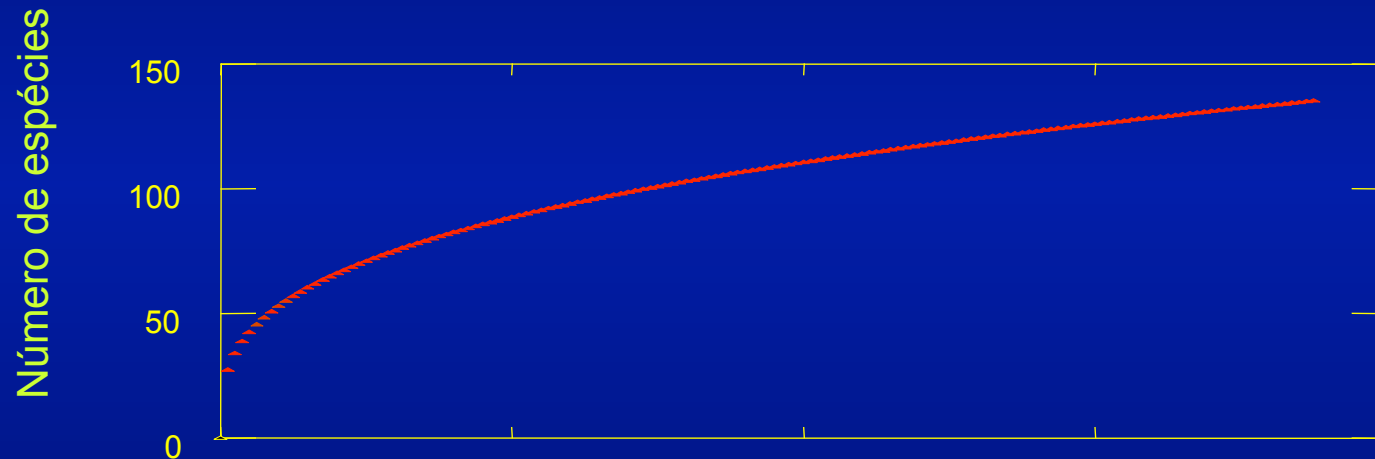
$$n = 14 \quad (2s) \quad (t \rightarrow 95\%)$$

PRECISÃO DESEJADA	10%	20%	30%	40%	50%
<i>Spiochaetopterus nonatoi</i>	85	21	9	5	3
<i>Lumbrineris</i> sp.	90	22	10	6	3
<i>Magelona pappilicornis</i>	136	34	15	8	5
<i>Amphiodia atra</i>	188	47	20	12	8
<i>Nucula puelcha</i>	218	55	24	14	9
<i>Amphiodia rüsei</i>	257	64	29	16	10
<i>Eunoe papillosa</i>	351	88	39	22	14
<i>Diplodonta danieli</i>	357	89	40	22	14
<i>Pectinaria (Pectinaria) laelia</i>	454	114	50	28	18
<i>Tellina</i> sp.	522	131	58	33	21
<i>Melaniella</i> sp.	585	146	65	37	23
<i>Ctena pectinella</i>	596	149	66	37	24
<i>Entodesma</i> sp.	629	157	70	39	25
<i>Amphitalamus vallei</i>	642	160	71	40	26
<i>Mooreonuphis lineata</i>	808	202	90	50	32
<i>Batea catharinensis</i>	1328	332	148	83	53
D10	16	4	2	1	1
D TOTAL	18	4	2	1	1
POLYCHAETA	20	5	2	1	1
PELECYPODA	63	16	7	4	3
GASTROPODA	149	37	17	9	6
OPHIUROIDEA	140	35	16	9	6
AMPHIPODA	223	56	25	14	9

# REPLICAGEM PARA ESTIMATIVA DA RIQUEZA

## Curvas espécie-área

## Curvas de acumulação de espécies

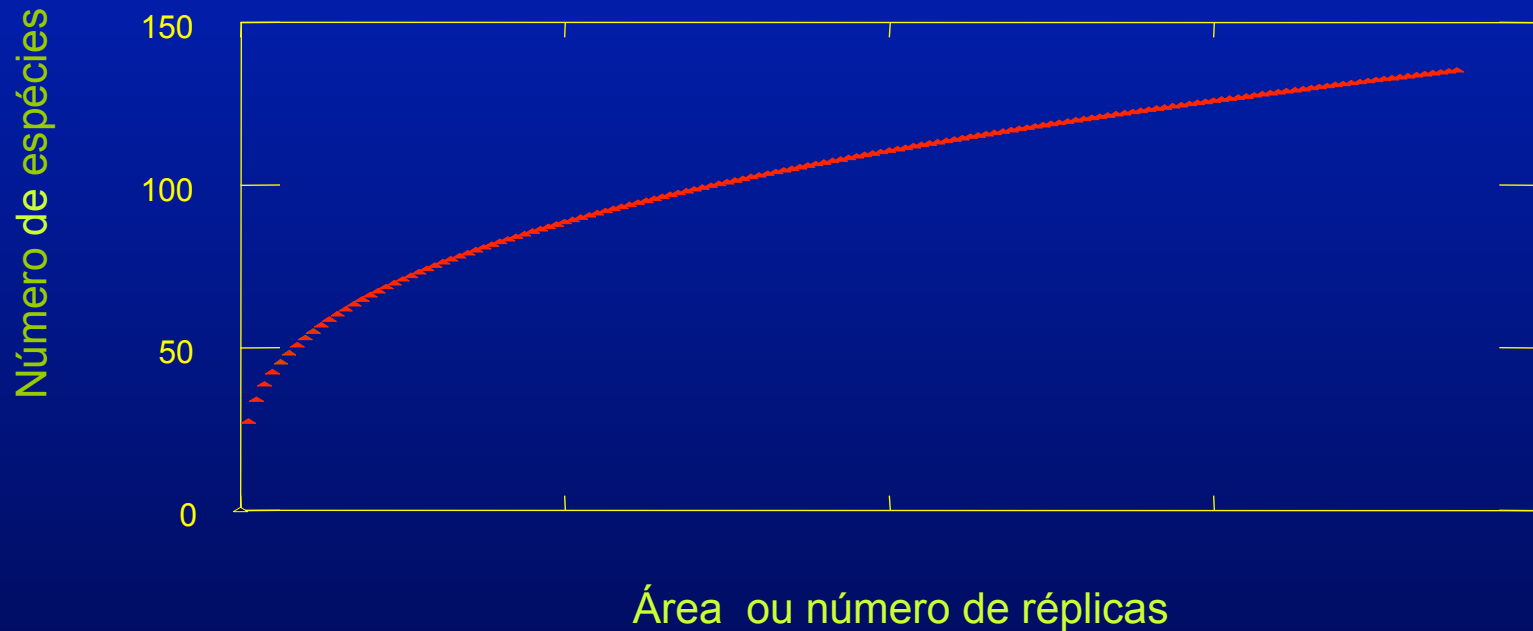


Área amostral = número de réplicas ?

# Curvas espécie-área/acumulação de espécies

Por que cresce ?

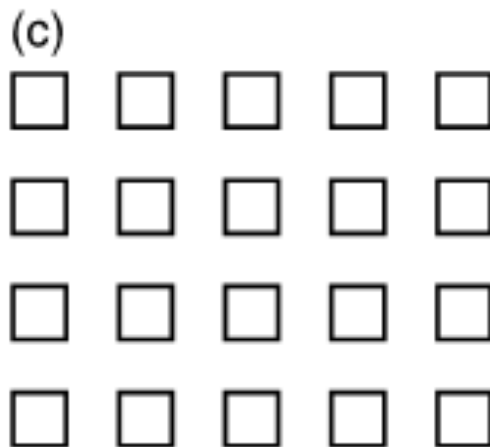
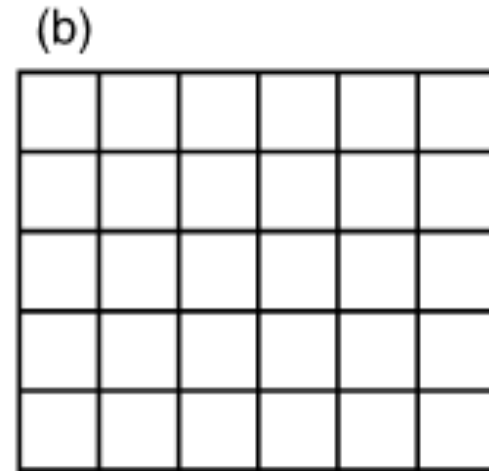
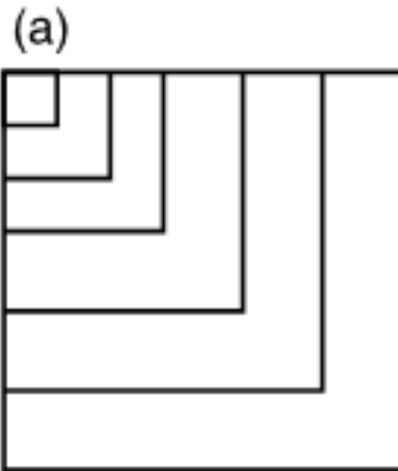
- > espécies menos freqüentes
- > espécies e novos ambientes (heterogeneidade)



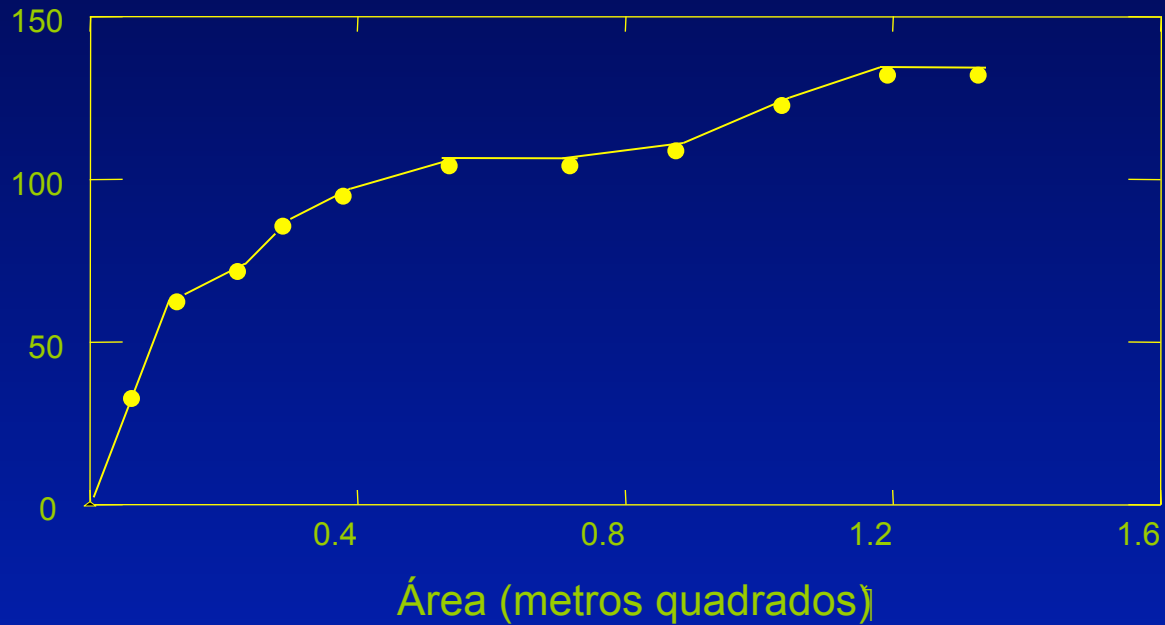


(a), (b), (c) = curva de acumulação de espécies

(d) = curva espécie-área

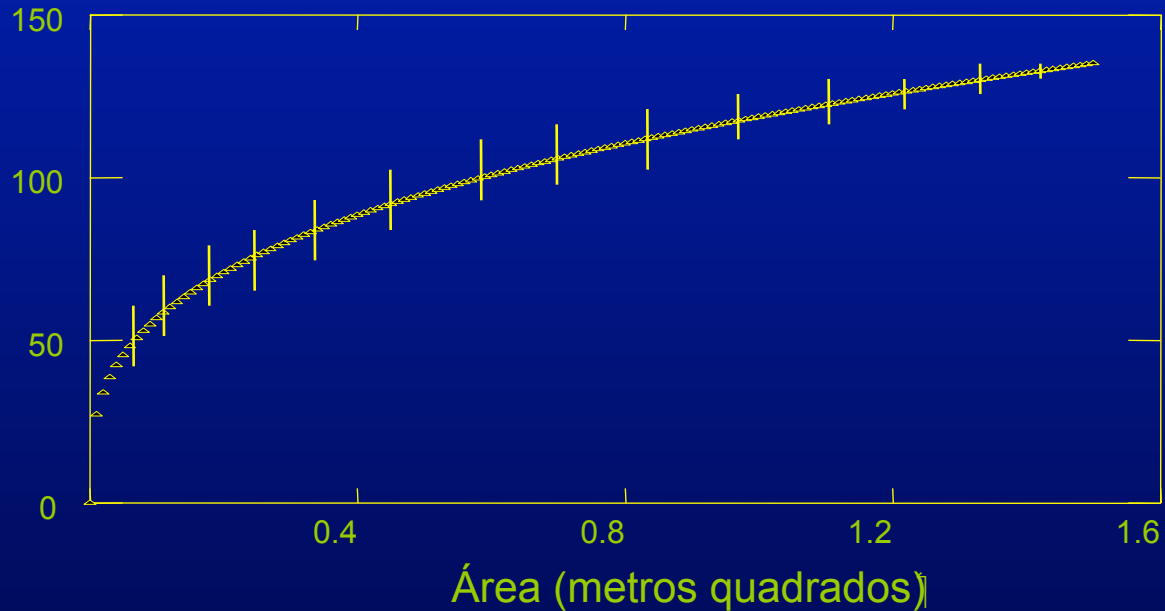


Número de espécies



CURVA DE ACUMULAÇÃO DE SPP.  
(baseada nas amostras)

Número de espécies

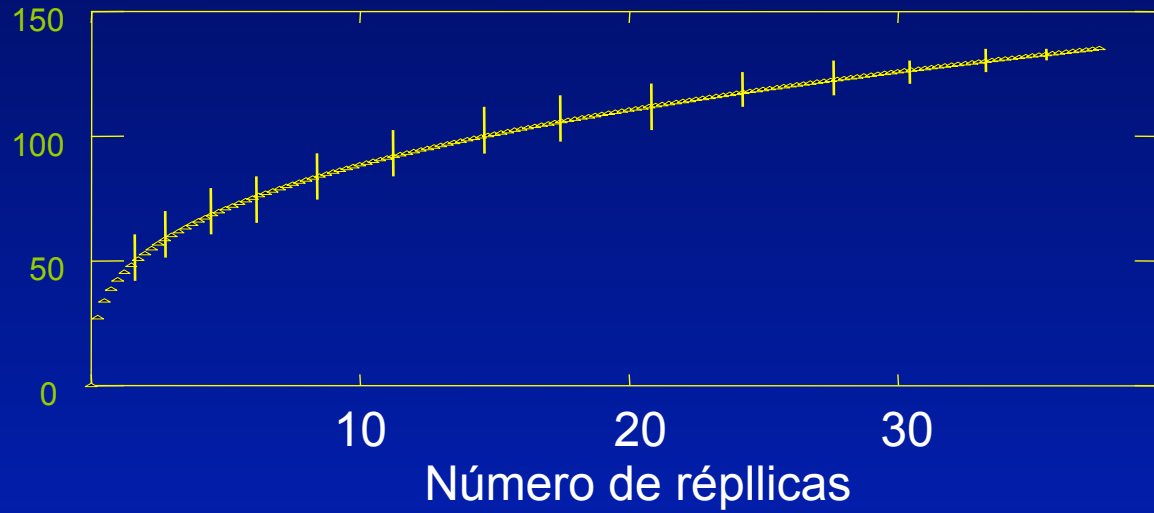


CURVA DE RAREFAÇÃO DE SPP.  
(baseada nas amostras)  
\*pode ser baseada nos indivíduos tb.

# CURVA DE RAREFAÇÃO DE SPP.

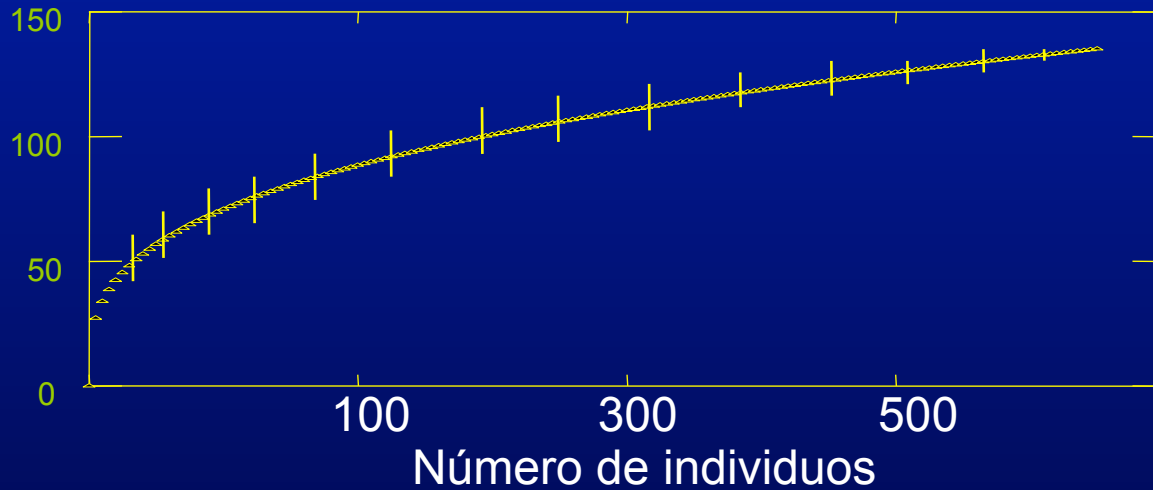
- Baseada nas amostra

Número de espécies



- Baseada em indivíduos

Número de espécies



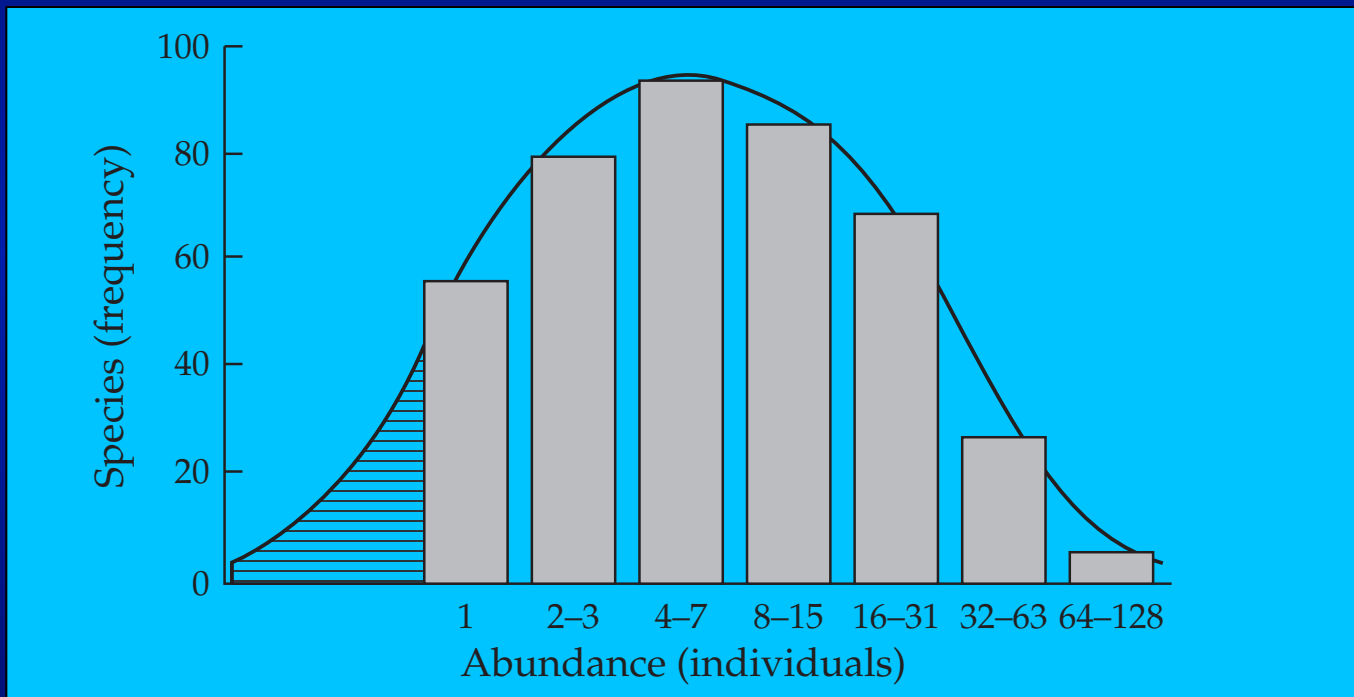
## MÉTODOS PARA ESTIMAR O ESFORÇO AMOSTRAL

- (1) Ajustando uma distribuição estatística para os dados de abundância (distribuição log-normal, broken stick, etc.)
- (2) Extrapolando uma curva de acumulação de espécies pela sua assintótica (curva do coletor ajustando uma função p. ex.)
- (3) Estimando o número asintótico de espécies com estimadores não paramétricos (Chao1, Chao2, Jackknife, ACE, ICE, Jackknife, etc.)

Todos eles permitem o cálculo do número mínimo de réplicas, área amostral ou n. de indivíduos para uma amostragem representativa da comunidade

## Estimativas de riqueza “real” (= não amostrada)

Ajustando uma distribuição log-normal de abundância



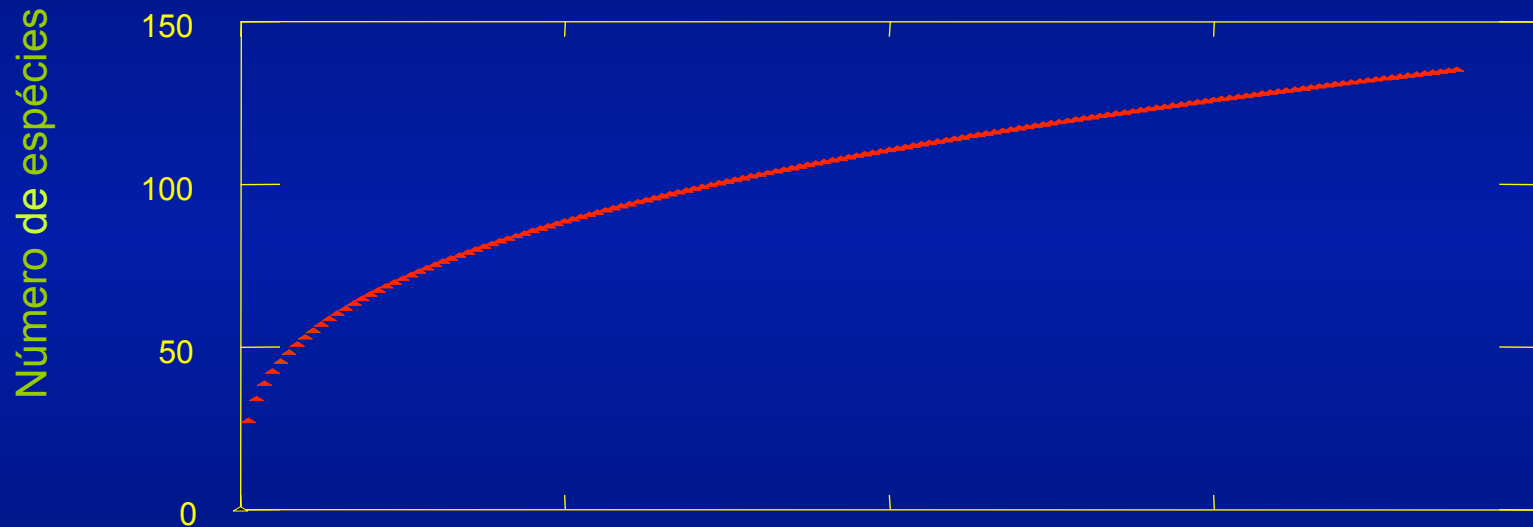
## MÉTODOS PARA ESTIMAR O ESFORÇO AMOSTRAL

- (1) Ajustando uma distribuição estatística para os dados de abundância (distribuição log-normal, broken stick, etc.)
- (2) Extrapolando uma curva de acumulação de espécies pela sua assintótica (curva do coletor ajustando uma função p. ex.)
- (3) Estimando o número asintótico de espécies com estimadores não paramétricos (Chao1, Chao2, Jackknife, ACE, ICE, Jackknife, etc.)

Todos eles permitem o cálculo do número mínimo de réplicas, área amostral ou n. de indivíduos para uma amostragem representativa da comunidade

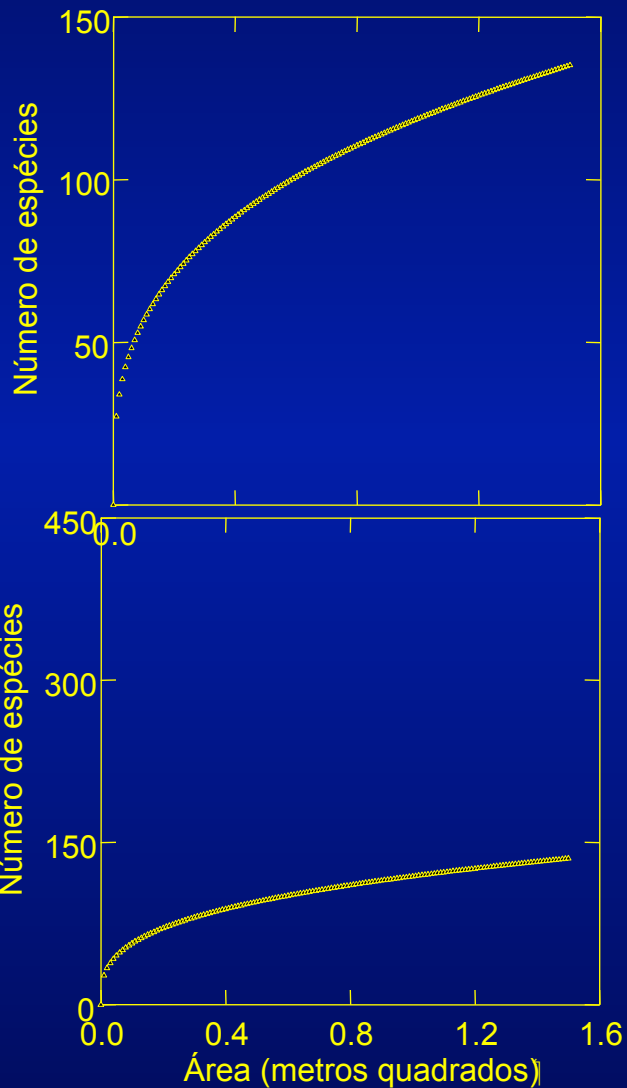
# Estimativas de riqueza “real” (= não amostrada)

## Curva de acumulação ou rarefação



Área amostral  
ou número de réplicas  
ou número de indivíduos

**CRITÉRIO: ESTABILIZAÇÃO DA CURVA**



**CRITÉRIO: % DE ESPÉCIES**

Área (m <sup>2</sup> )	nº réplicas	% do total de espécies
0,12	15	50%
0,20	25	60%
0,32	40	70%
0,52	65	80%
0,88	110	90%
<b>1,16</b>	<b>145</b>	<b>95%</b>

**CRITÉRIO: PONTO DE MOLINIER (20% de incremento na área amostral)**

Área (m <sup>2</sup> )	nº réplicas	Incremento na % total de espécies
0,13	17	20%
0,30	37	10%
0,59	74	5%
<b>1,03</b>	<b>132</b>	<b>4%</b>
1,44	180	2%



## MÉTODOS PARA ESTIMAR O ESFORÇO AMOSTRAL

- (1) Ajustando uma distribuição estatística para os dados de abundância (distribuição log-normal, broken stick, etc.)
- (2) Extrapolando uma curva de acumulação de espécies pela sua assintótica (curva do coletor ajustando uma função p. ex.)
- (3) Estimando o número asintótico de espécies com estimadores não paramétricos (Chao1, Chao2, Jackknife, ACE, ICE, Jackknife, etc.)

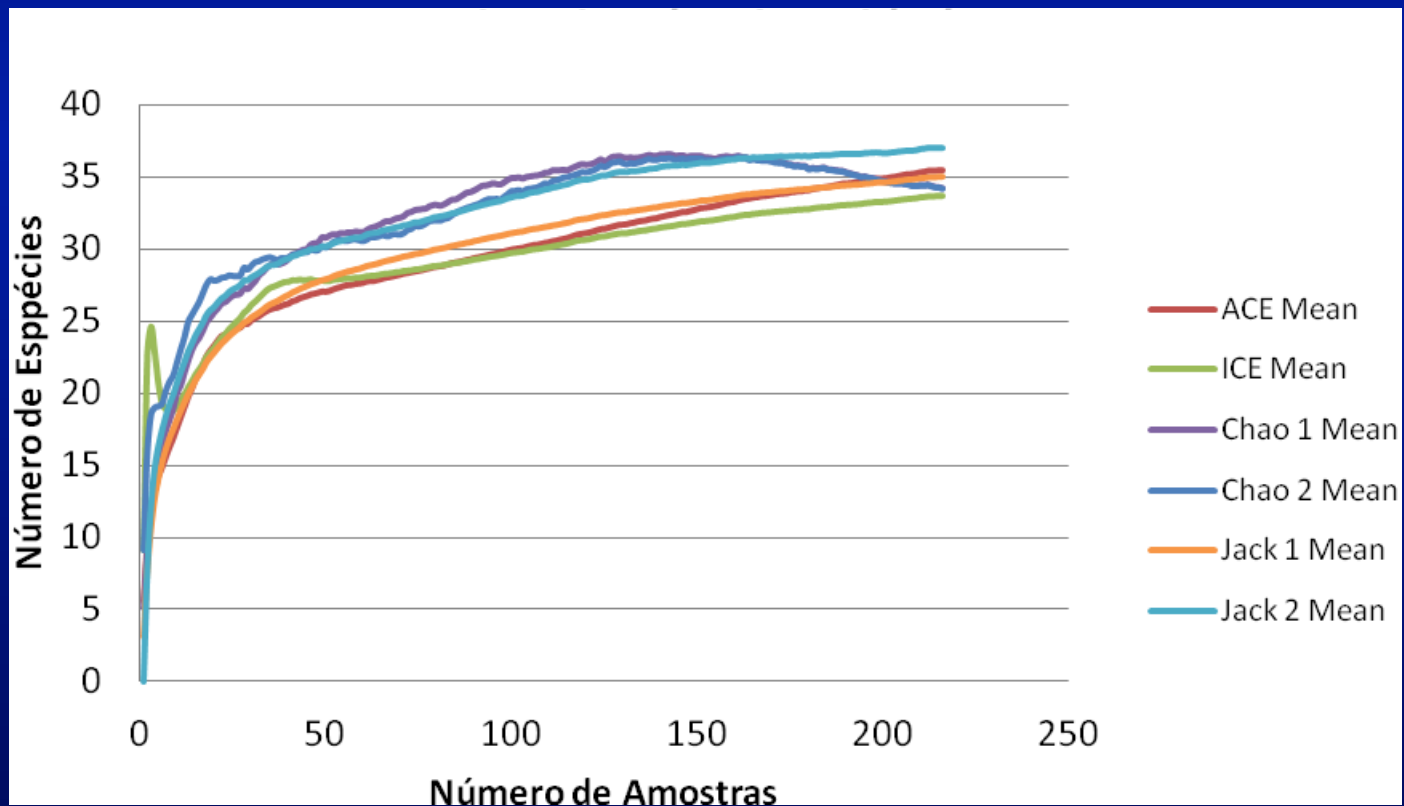
Todos eles permitem o cálculo do número mínimo de réplicas, área amostral ou n. de indivíduos para uma amostragem representativa da comunidade

# Estimativas de riqueza “real” (= não amostrada)

## Estimadores assintóticos de riqueza

$S_{est} = S_{obs} +$  função de singletons ( $f_1$ ) e doubletons ( $f_2$ )

(Ex.  $S_{jackknife2} = S_{obs} + 2f_1 - f_2$ )



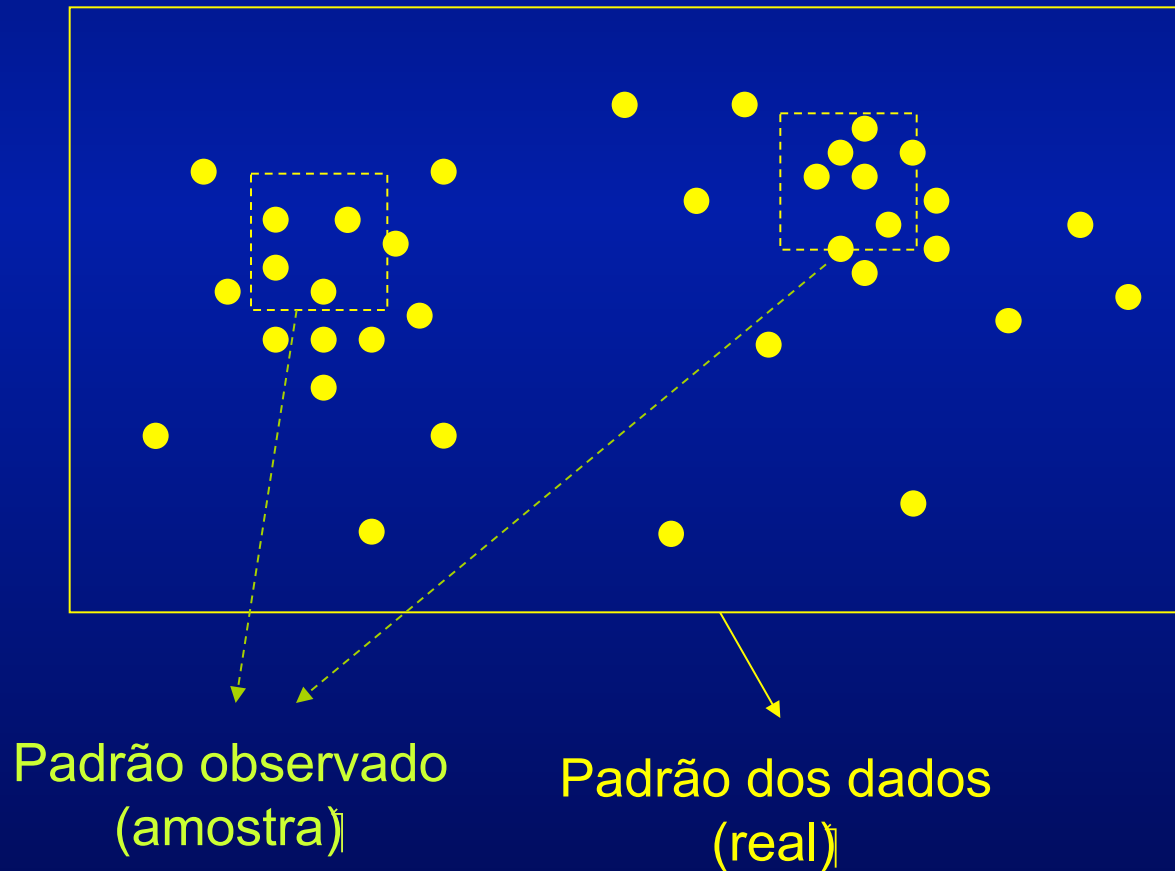
## ALOCAÇÃO DAS AMOSTRAS & ESCALAS

- Tipos gerais de amostragem
- Padrões de distribuição espacial e temporal (= manchas)
- Escala = Amplitude & Resolução
- Escala de amostragem & padrão de manchas (resposta)
- Escala de amostragem & escala dos processos (explicativos)
- Distribuição das réplicas por 'tratamento'

# ESCALAS

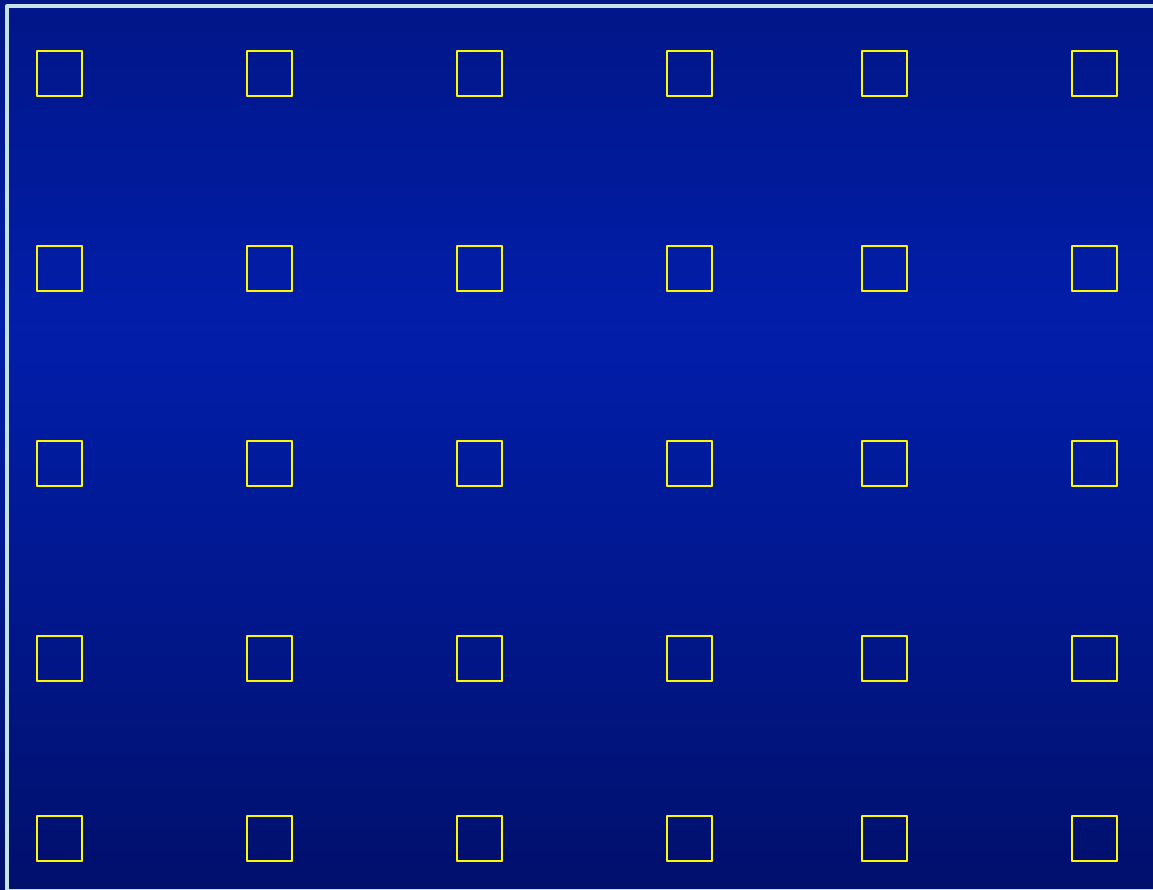
Padrões de distribuição espacial e temporal

Escala de amostragem & padrão dos dados



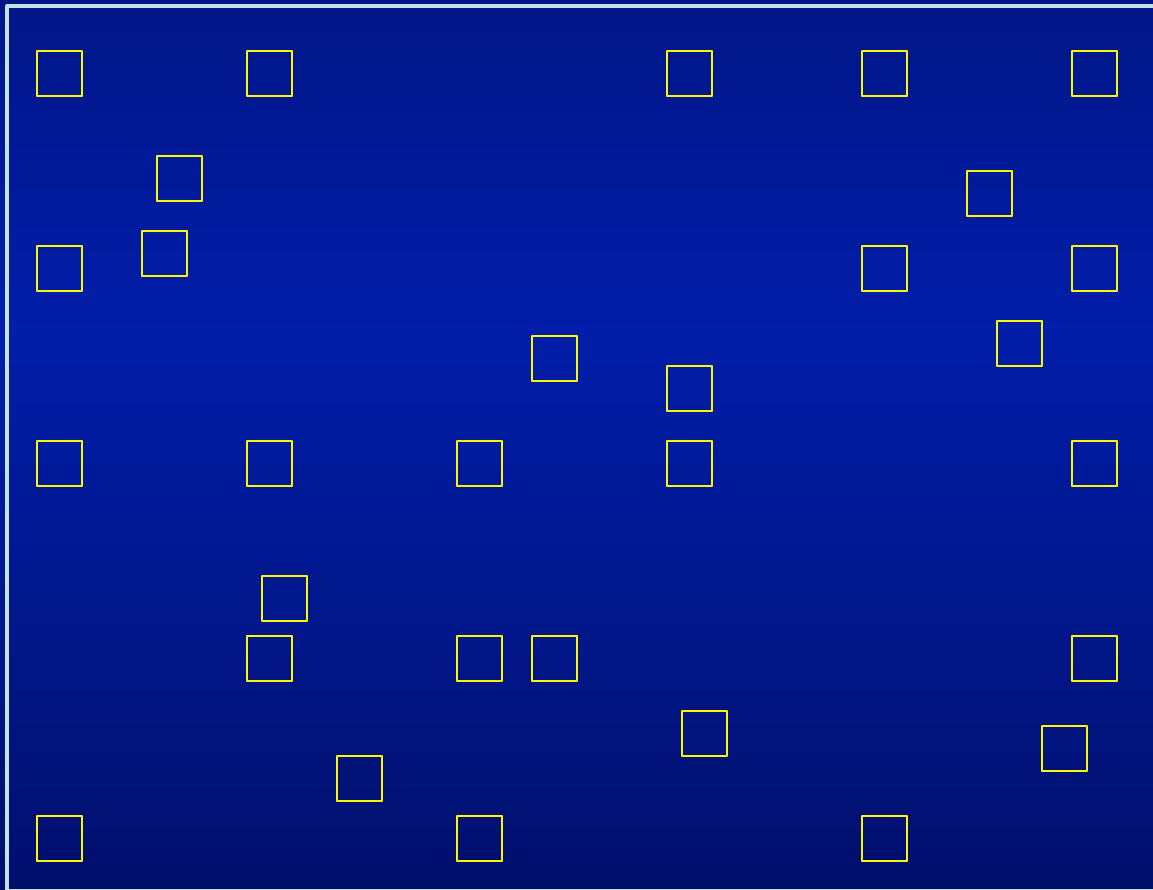
## ALOCAÇÃO DAS AMOSTRAS

- Amostragem sistemática (30 réplicas)



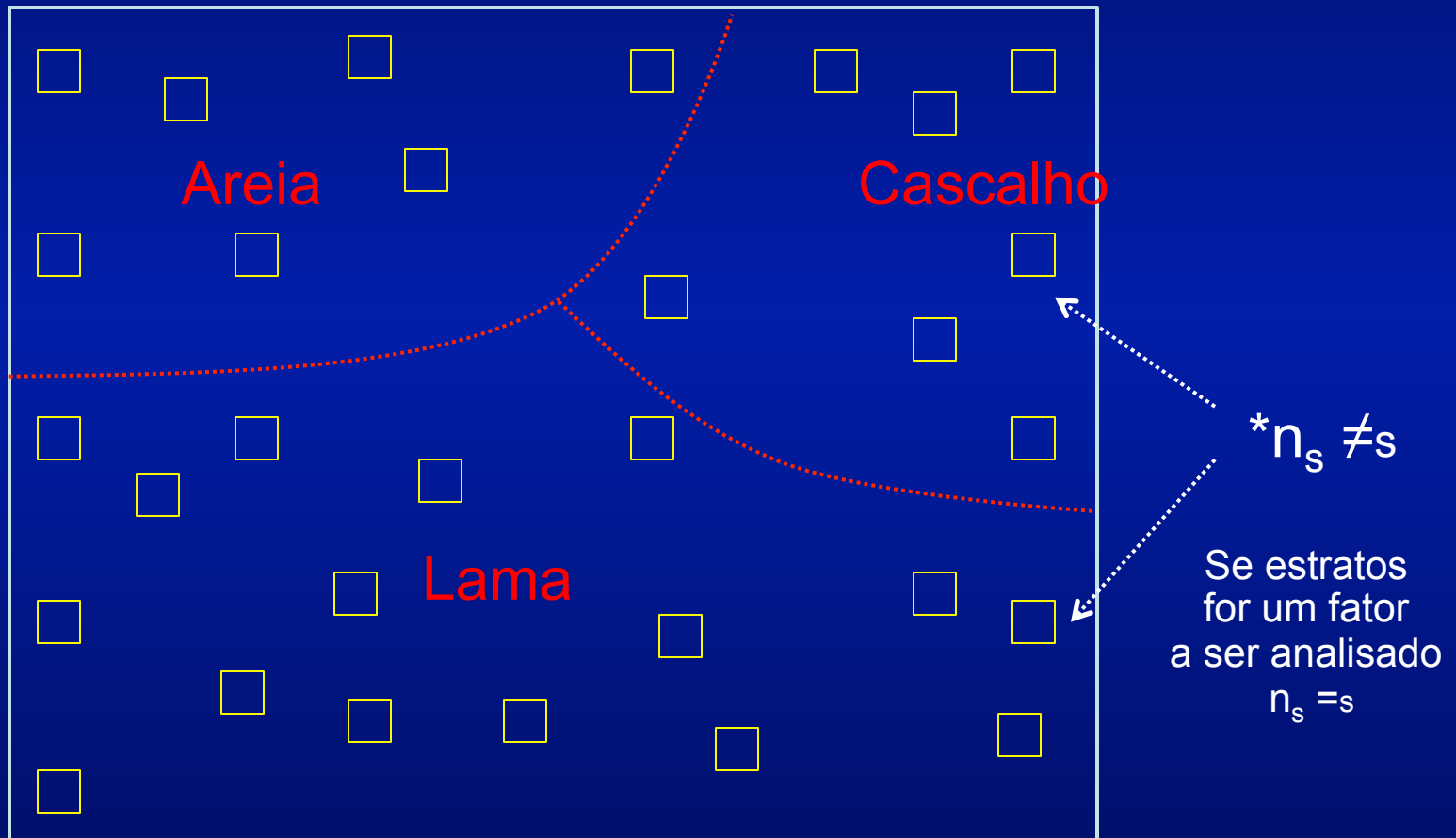
## ALOCAÇÃO DAS AMOSTRAS

- Amostragem aleatória (30 réplicas)



## ALOCAÇÃO DAS AMOSTRAS

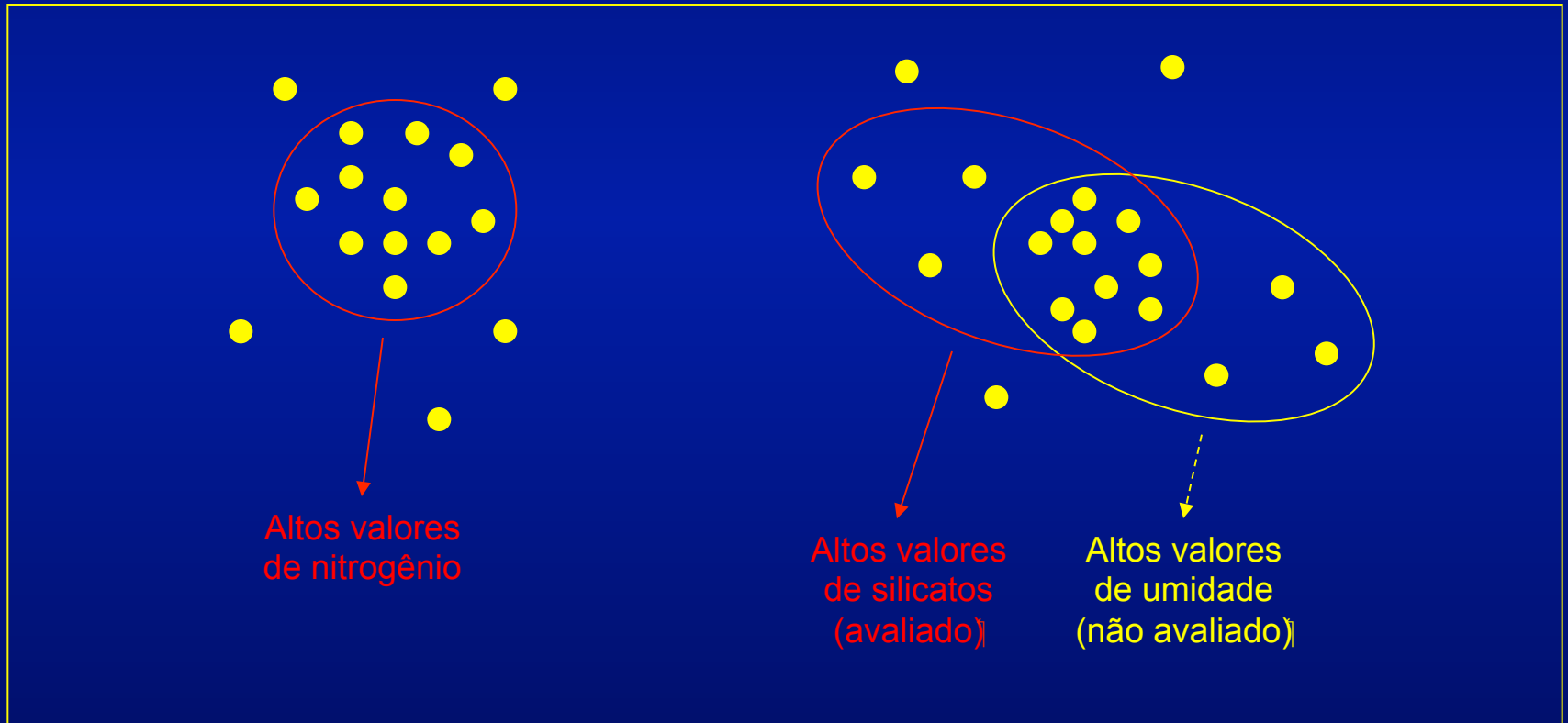
- Amostragem aleatoriamente estratificada (30 réplicas)



# ESCALAS

Padrões de distribuição espacial e temporal

Escala dos processos → Variáveis explicativas

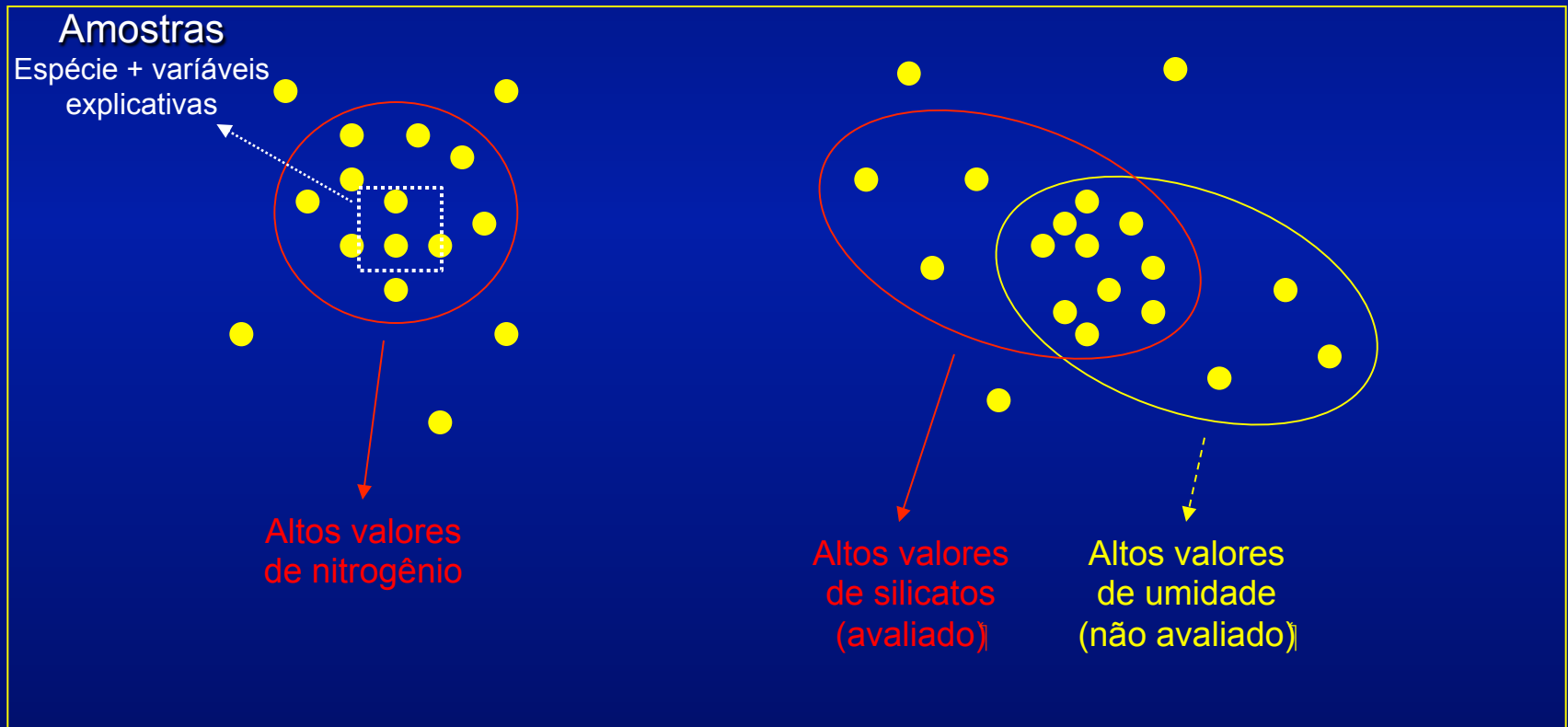




# ESCALAS

Padrões de distribuição espacial e temporal

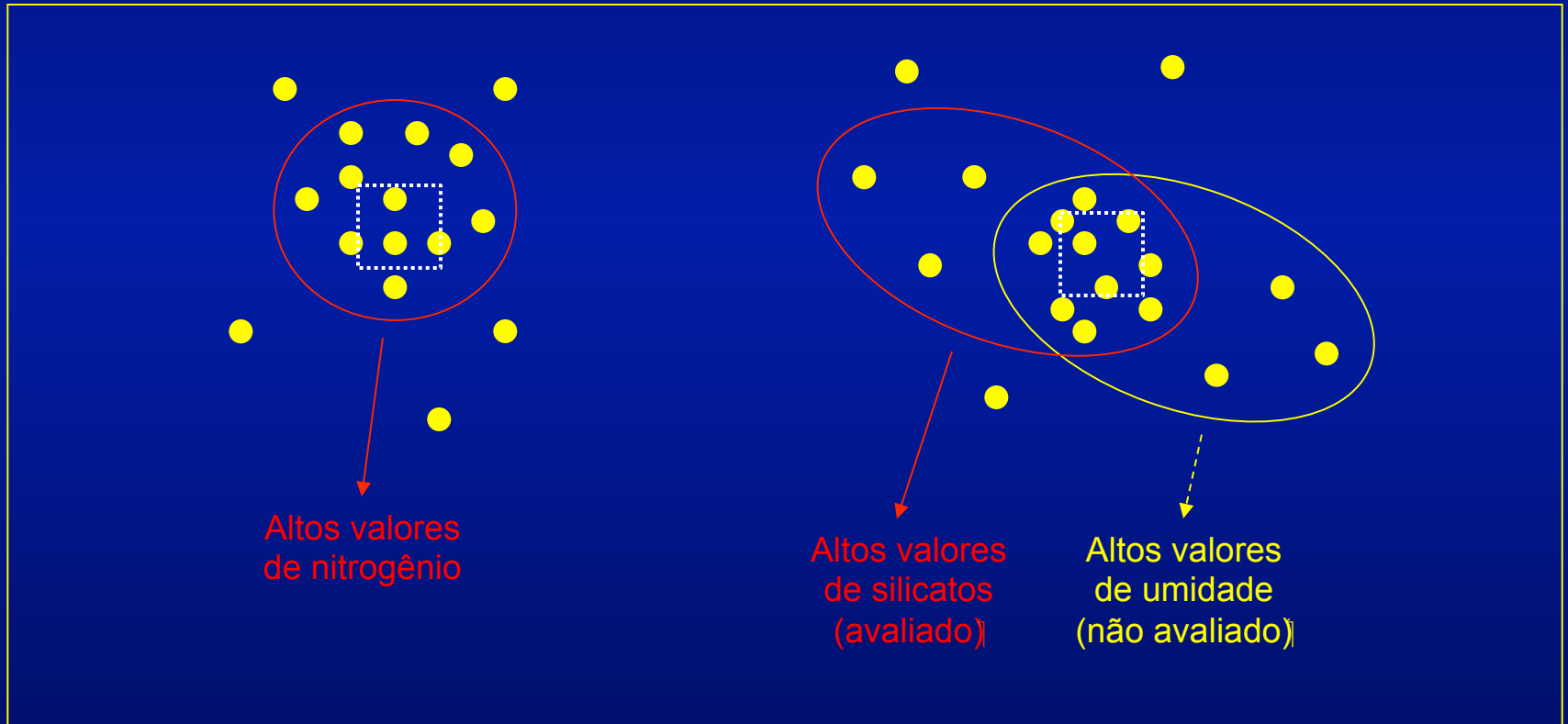
Escala dos processos → Variáveis explicativas



# ESCALAS

Padrões de distribuição espacial e temporal

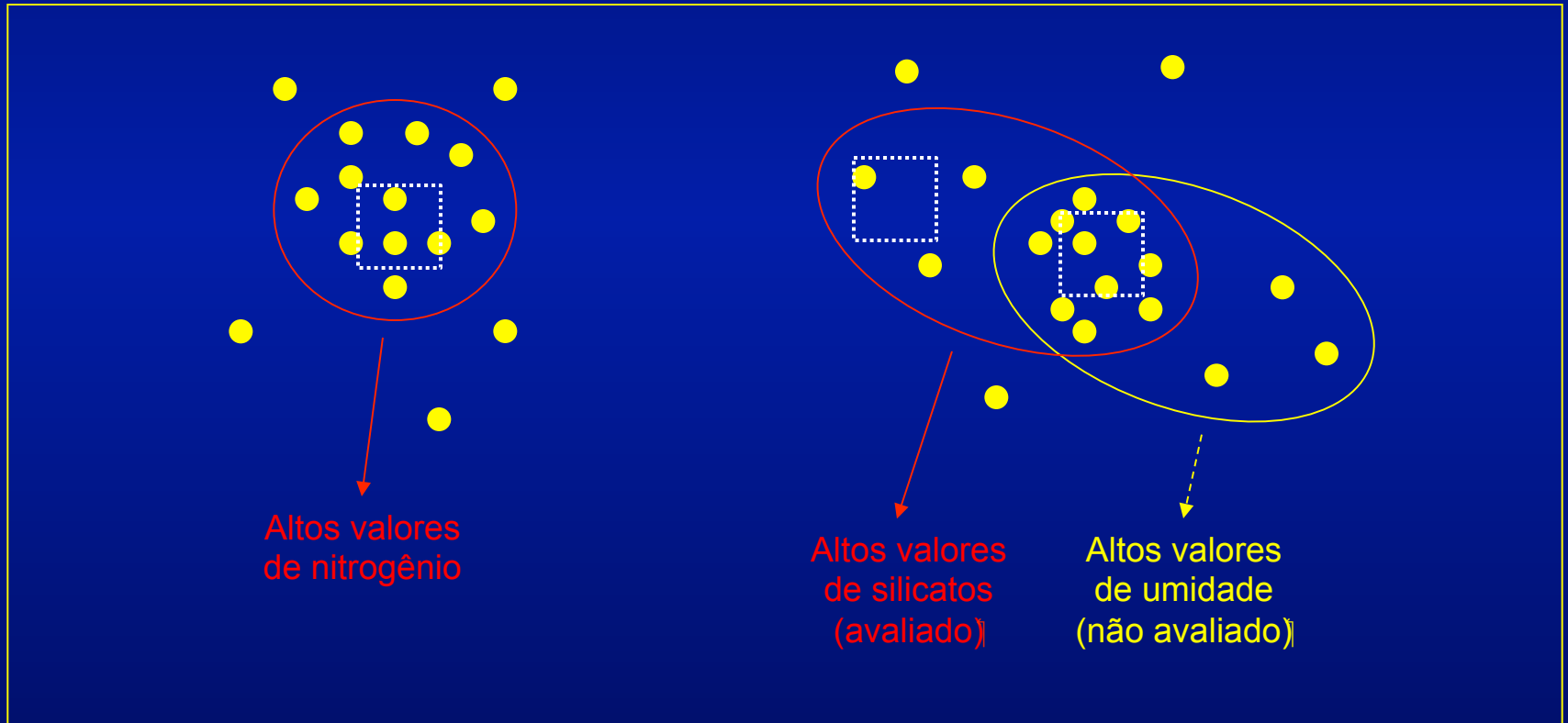
Escala dos processos → Variáveis explicativas



# ESCALAS

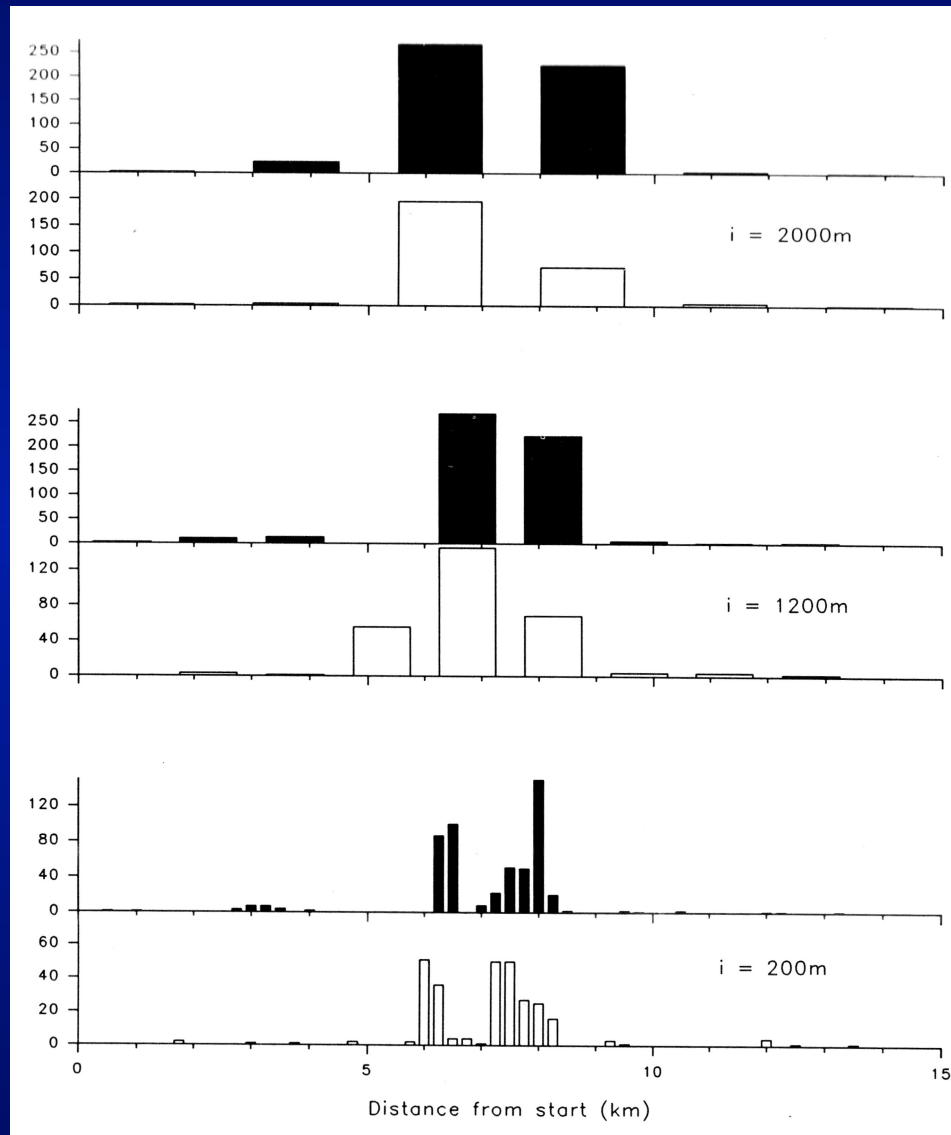
Padrões de distribuição espacial e temporal

Escala dos processos → Variáveis explicativas



# ESCALAS & PROCESSOS ESPACIAIS

## Relação predador/presa em aves marinhas



# ESCALAS & PROCESSOS ESPACIAIS



*Echinometra lucunter*



*Schizoporella errata*

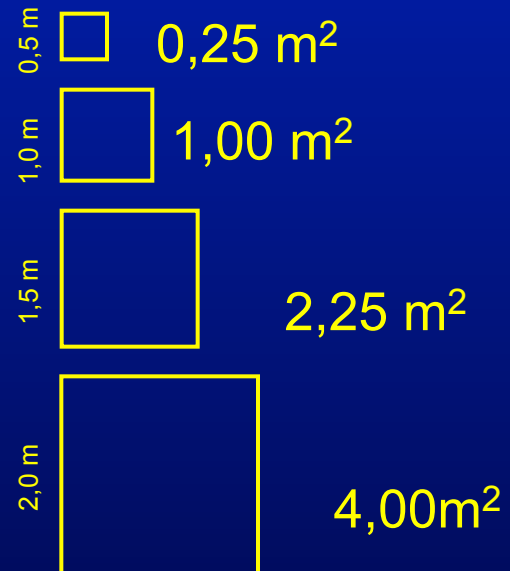


pastagem

competição



Algas filamentosas

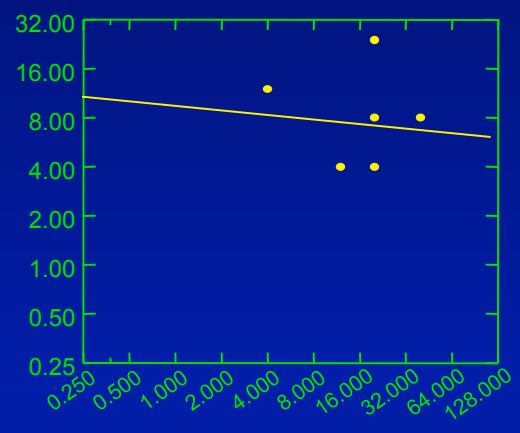


# ESCALAS & PROCESSOS ESPACIAIS

*Schizoporella errata*

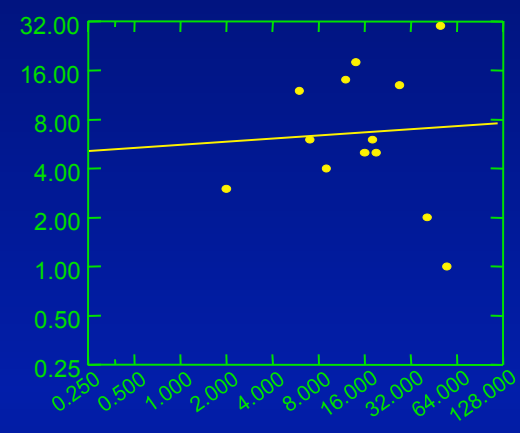
0,25 m<sup>2</sup>

R<sup>2</sup> = 0,03



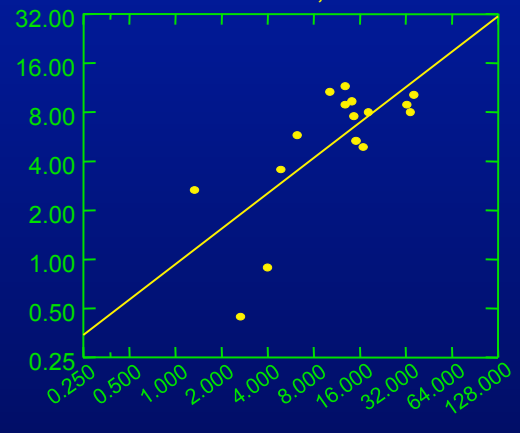
1 m<sup>2</sup>

R<sup>2</sup> = 0,39



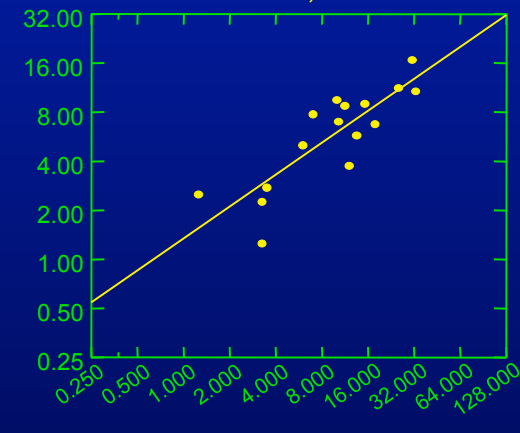
2,25 m<sup>2</sup>

R<sup>2</sup> = 0,85



4 m<sup>2</sup>

R<sup>2</sup> = 0,91

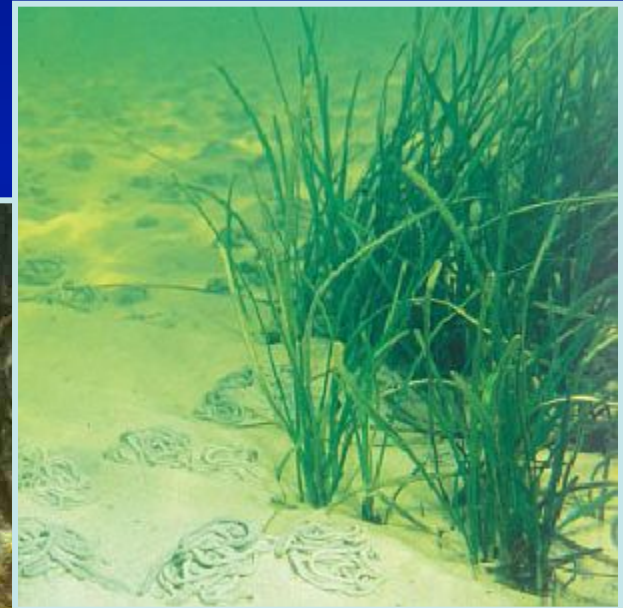


*Echinometra lucunter*

# ESCALAS & PROCESSOS ESPACIAIS



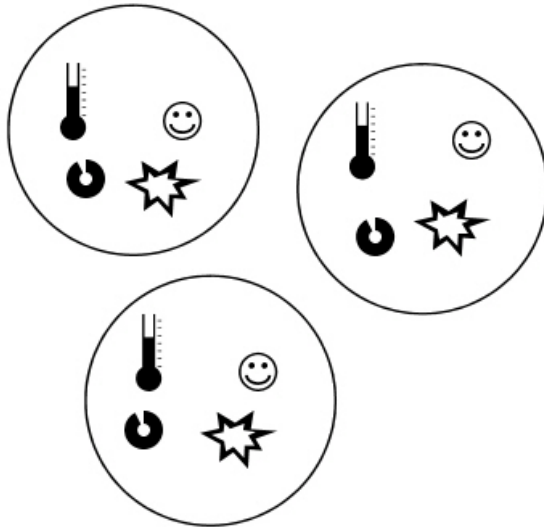
Gramas marinhas  
&  
fauna de sedimentos





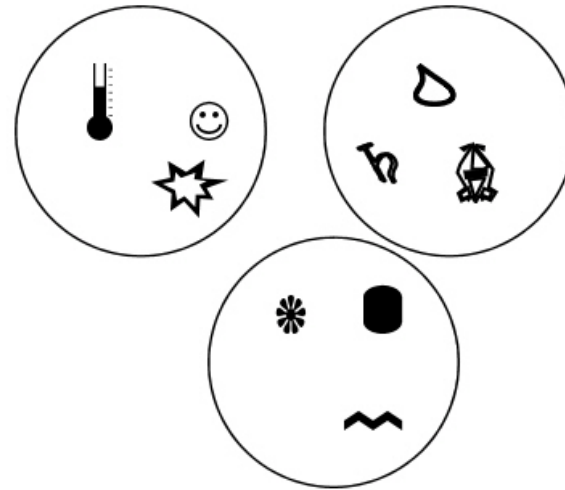
# ESCALAS & PROCESSOS ESPACIAIS

## Riqueza específica



Núm espécies p/ amostra = 4

Núm espécies p/ área = 4



Núm espécies p/ amostra = 3

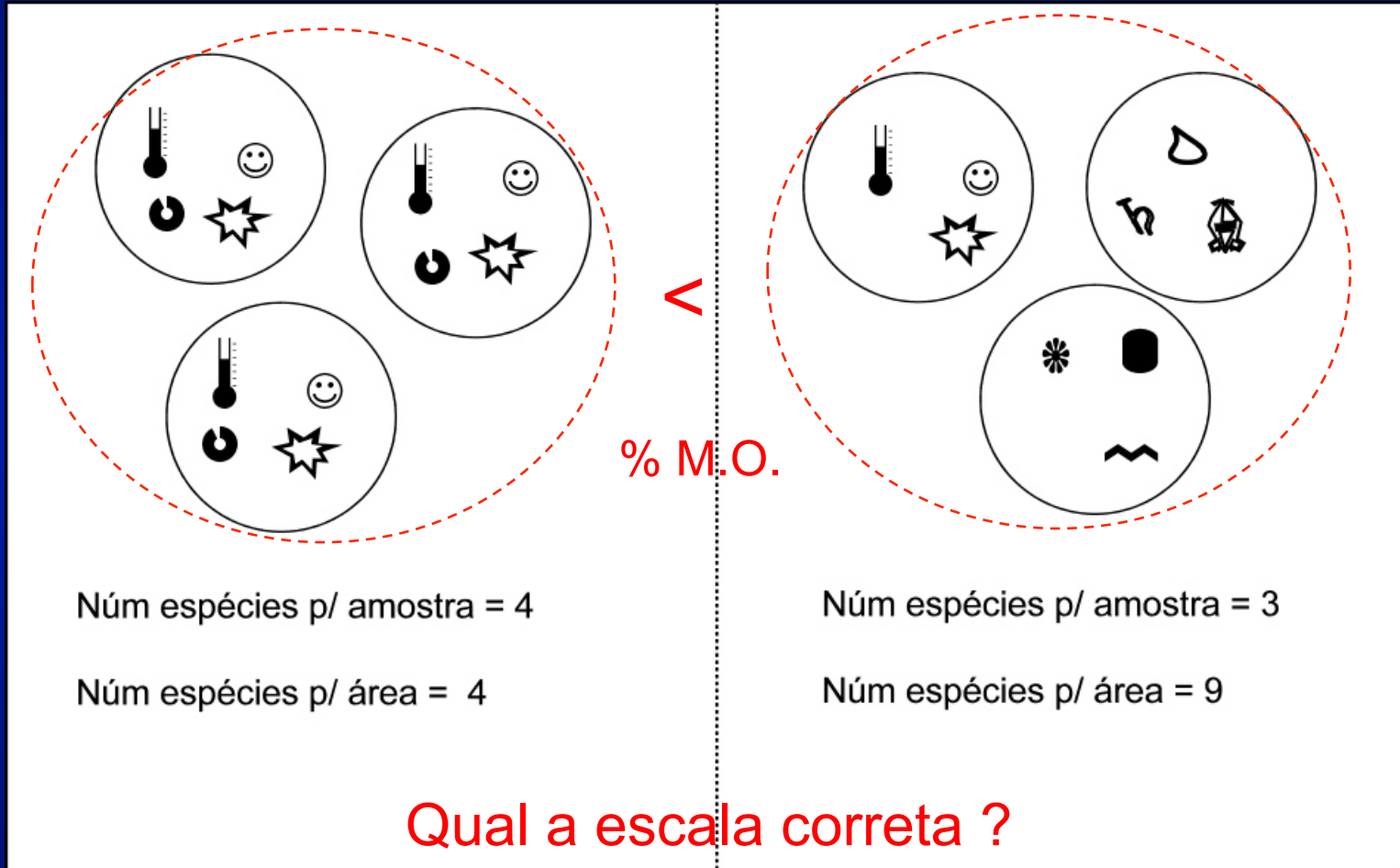
Núm espécies p/ área = 9

Qual a escala correta ?



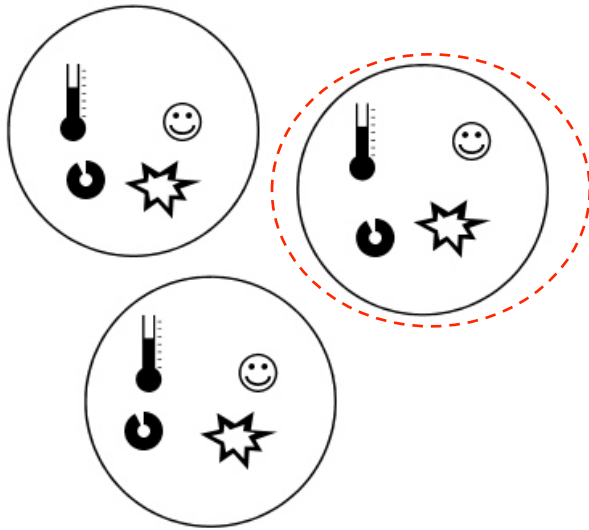
# ESCALAS & PROCESSOS ESPACIAIS

## Riqueza específica



# ESCALAS & PROCESSOS ESPACIAIS

## Riqueza específica



≠



Núm espécies p/ amostra = 4

Núm espécies p/ área = 4

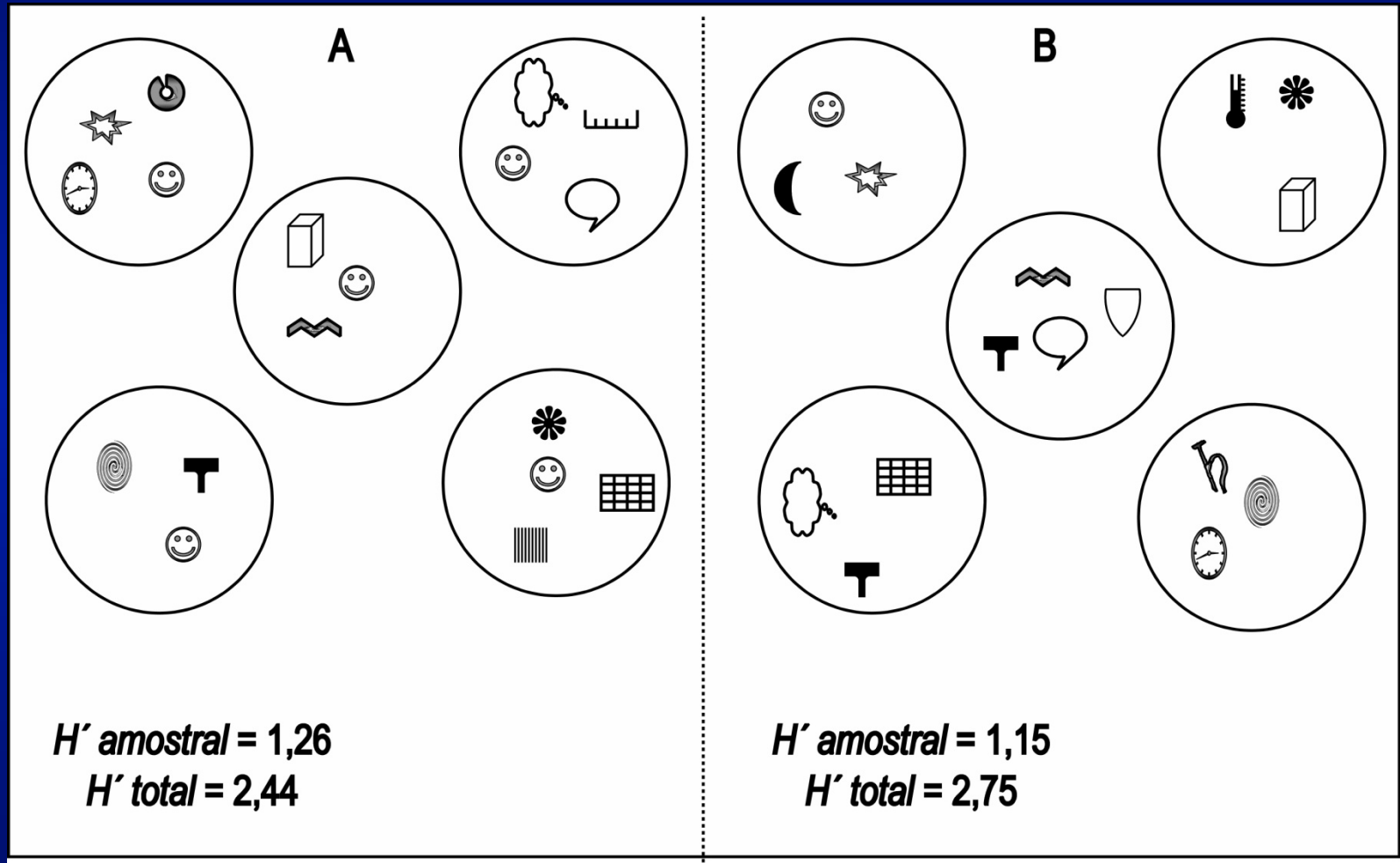
Núm espécies p/ amostra = 3

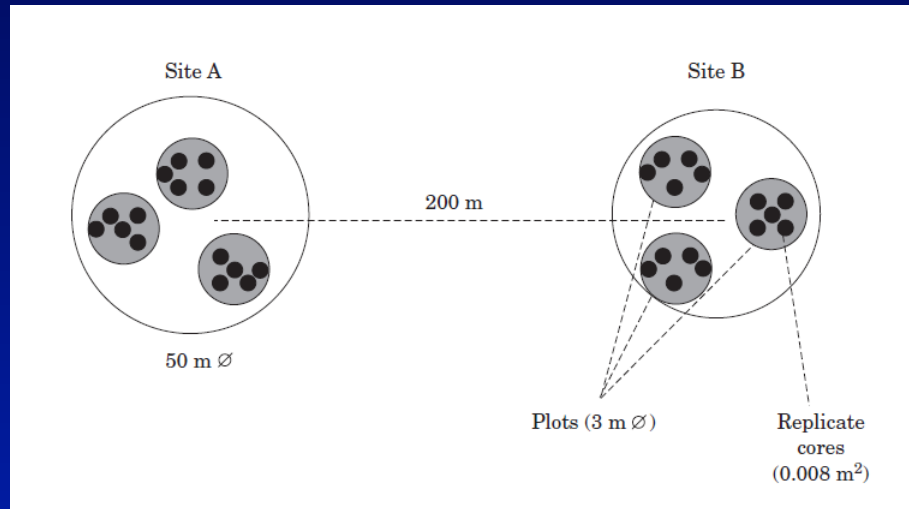
Núm espécies p/ área = 9

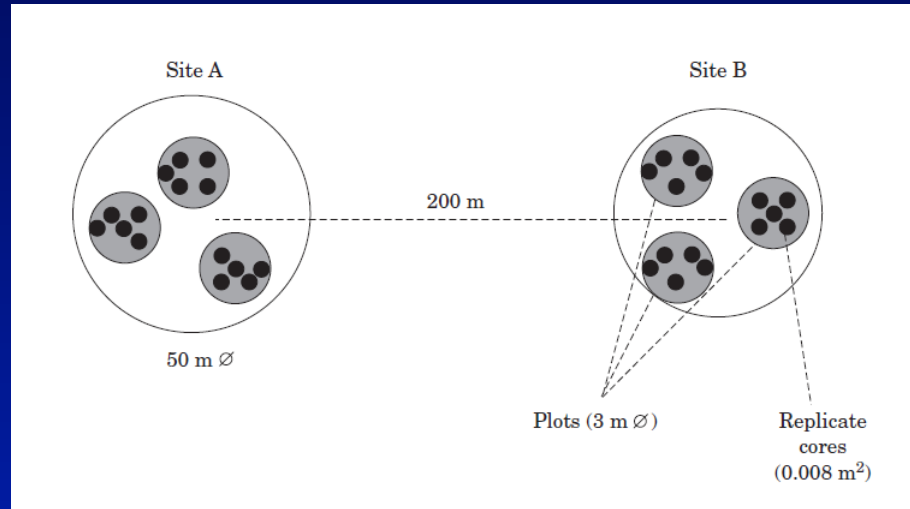
Qual a escala correta ?

# ESCALAS & PROCESSOS ESPACIAIS

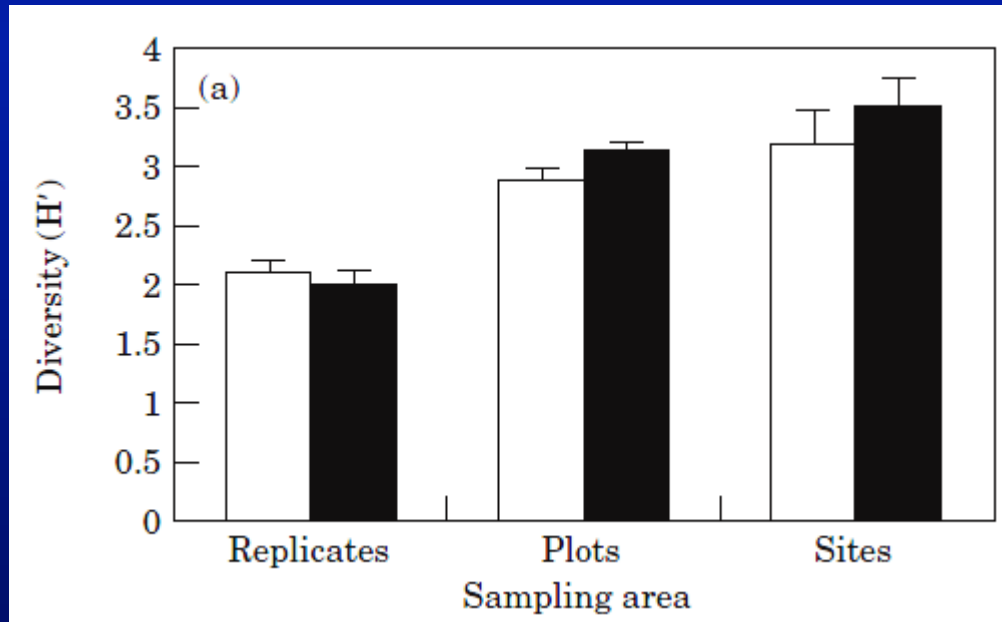
## Diversidade específica







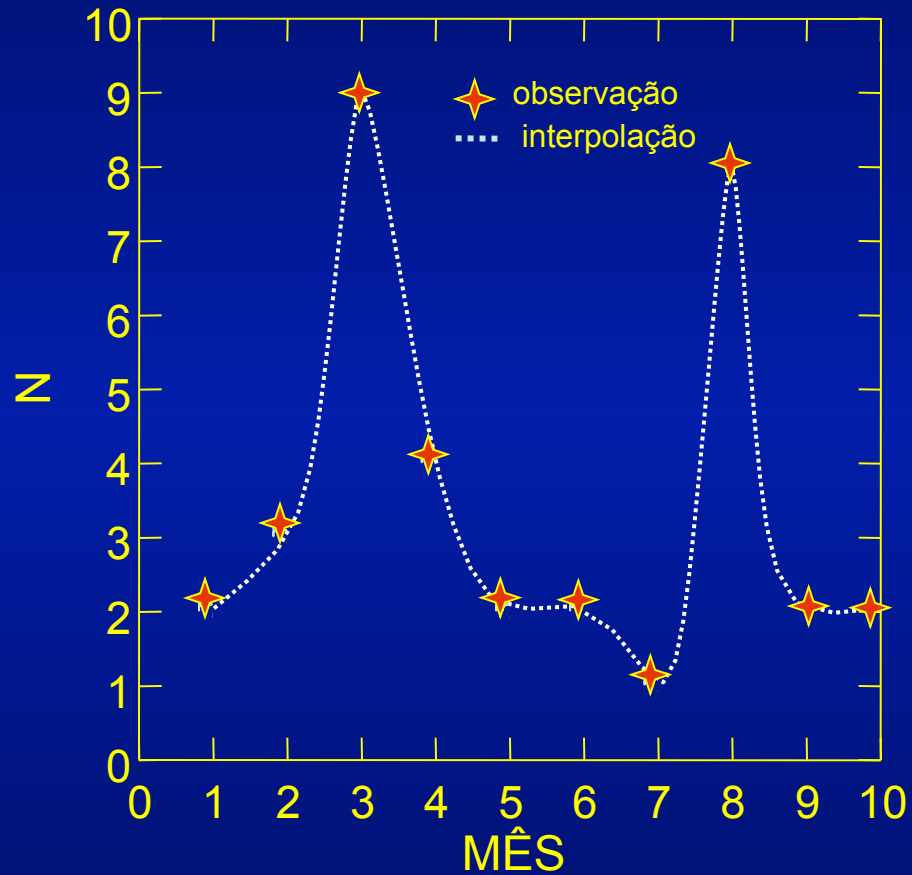
Diversidade  $\alpha$   
 &  
 Diversidade  $\beta$



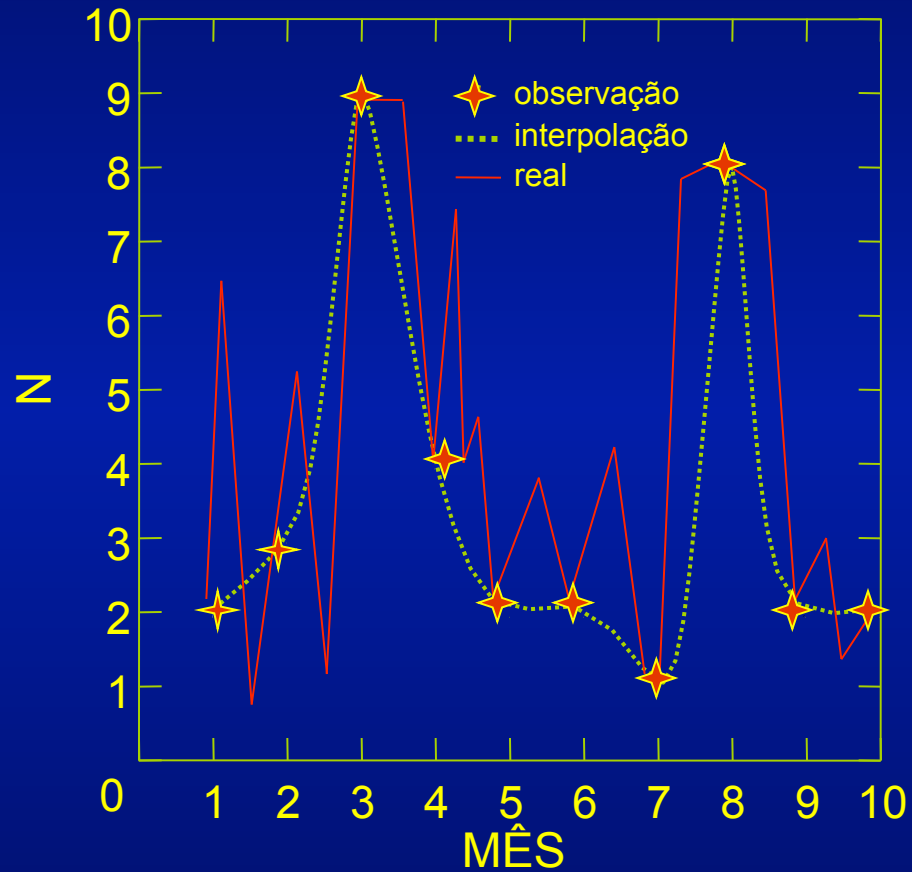
# AMOSTRAGEM TEMPORAL

- Quantas épocas amostrar (resolução) ?
- Por quanto tempo amostrar (amplitude) ?
- Por que 12 meses ? Por que 7 dias ? Por que 2 anos ?
- Sazonalidade
- Contrastes
- Tempo: Fator *fixo* ou *aleatório* ?
- Falta de independência em estudo temporal

# RESOLUÇÃO TEMPORAL

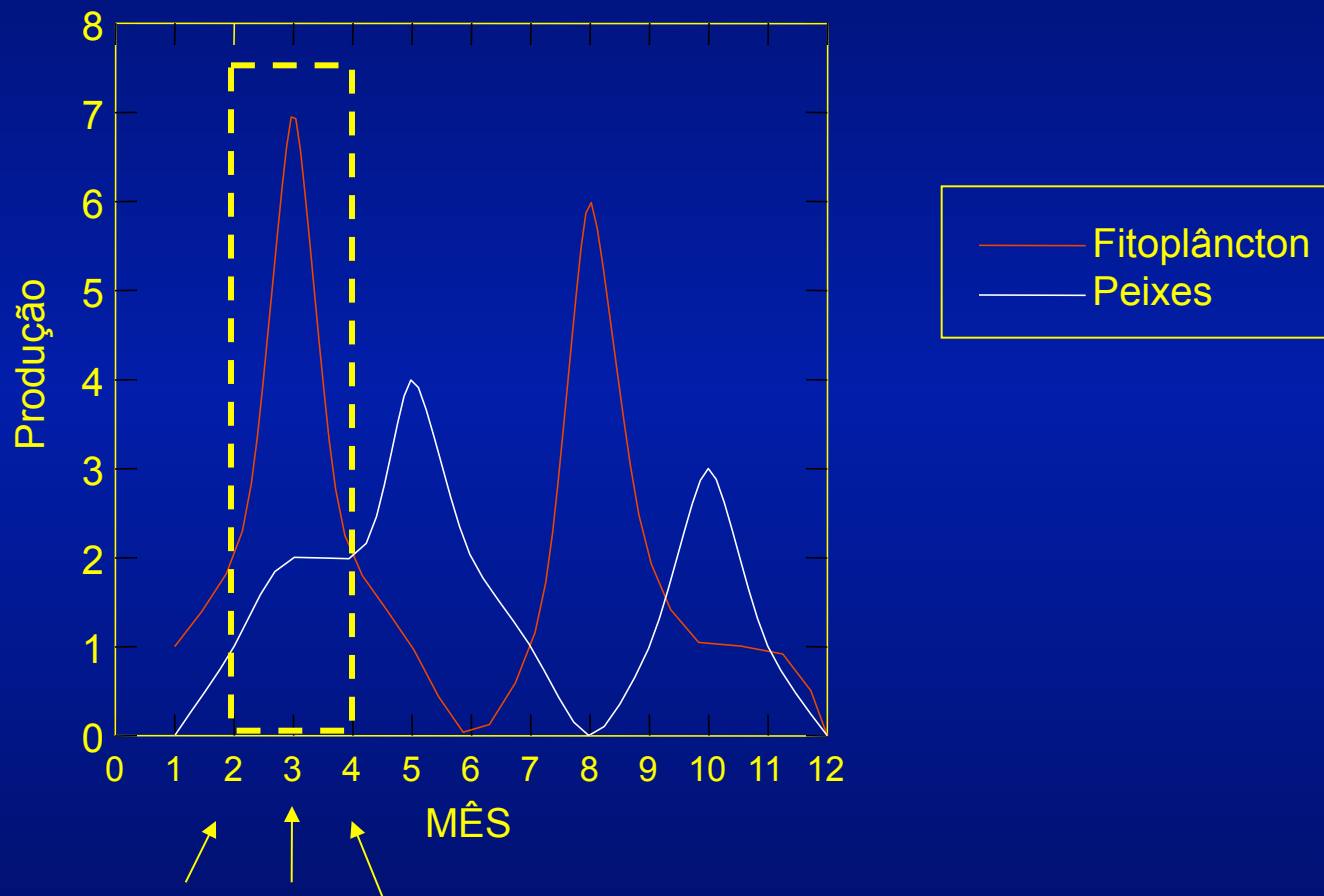


# RESOLUÇÃO TEMPORAL

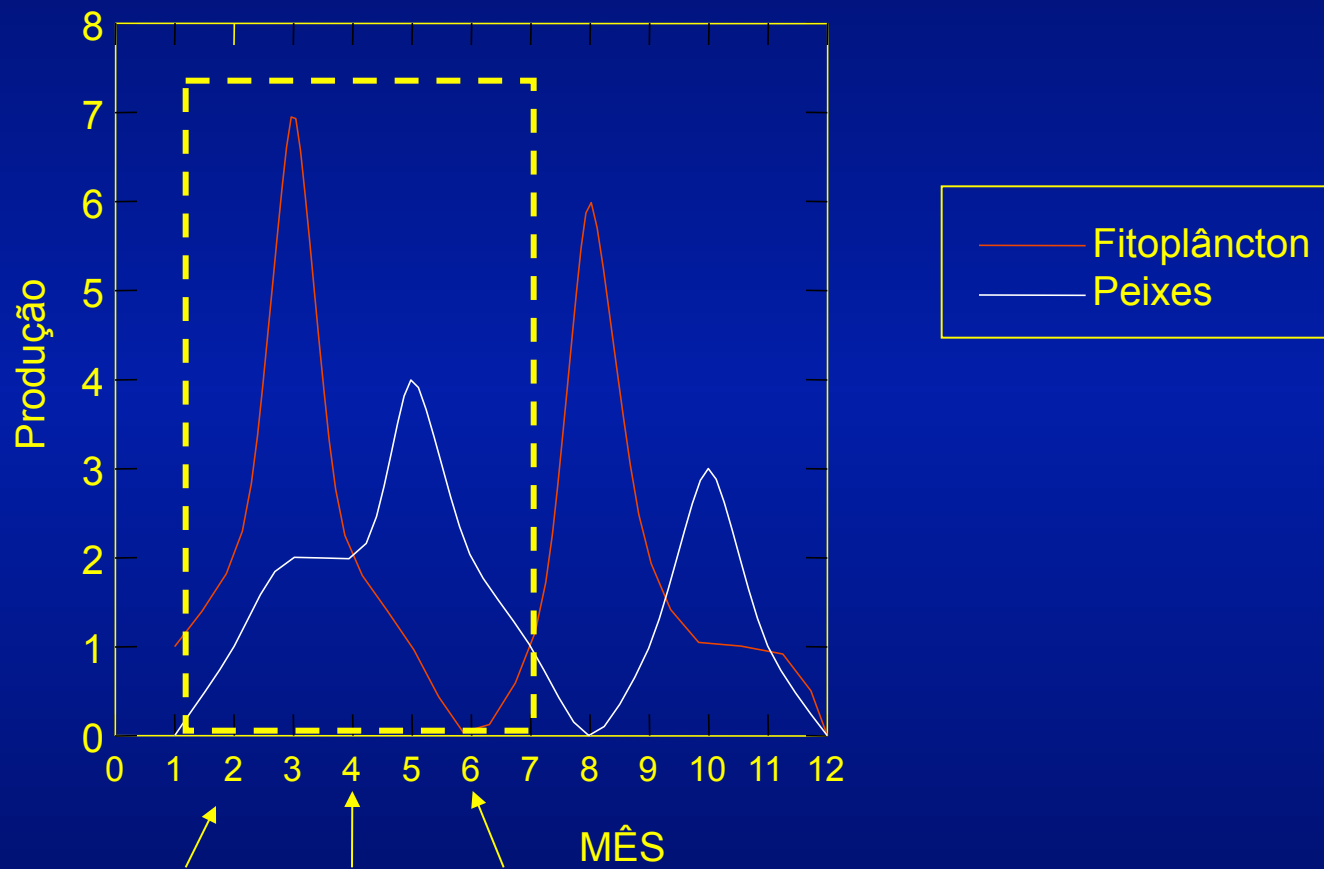




# ESCALAS & PROCESSOS TEMPORAIS



# ESCALAS & PROCESSOS TEMPORAIS



## AMOSTRAGEM CORRETA



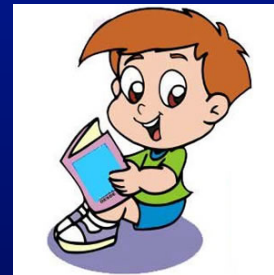
NÃO EQUALIZAÇÃO = PSEUDOREPLICAÇÃO

# PSEUDOREPLICAÇÃO ( 'Confounding' )

- Alocação incorreta das amostras
- Ausência de independência amostral
  - ✓ violação de premissa básica da amostragem
- Exemplos:

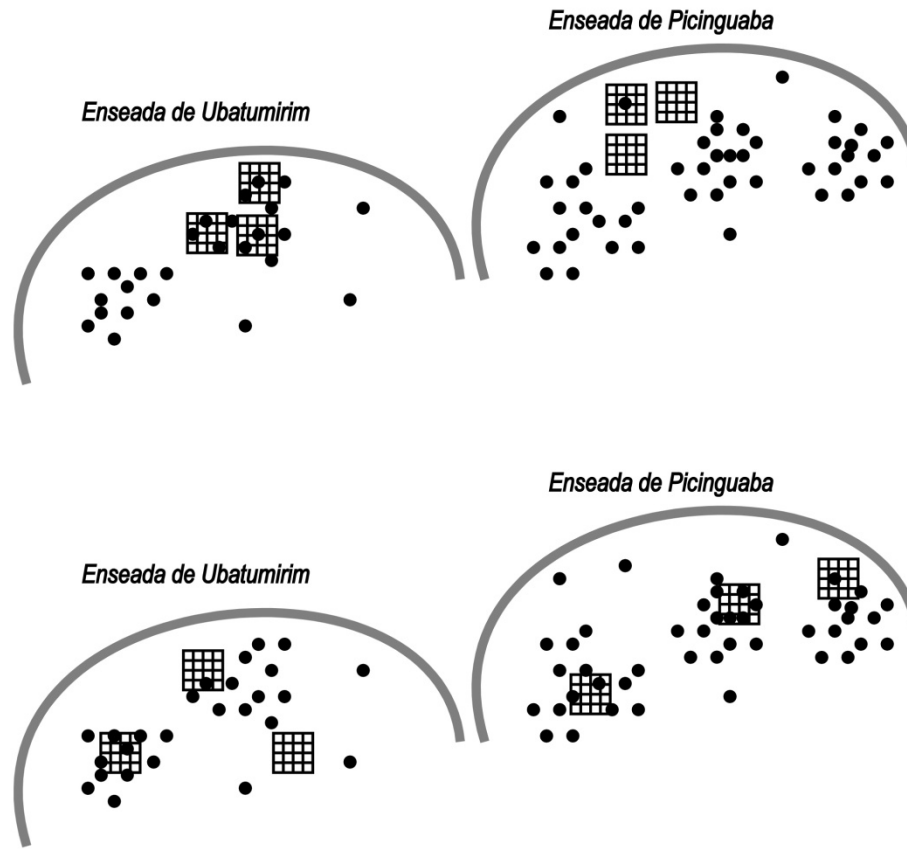
- ✓ Produção Primária em lagos

- ✓ Leitura dos alunos



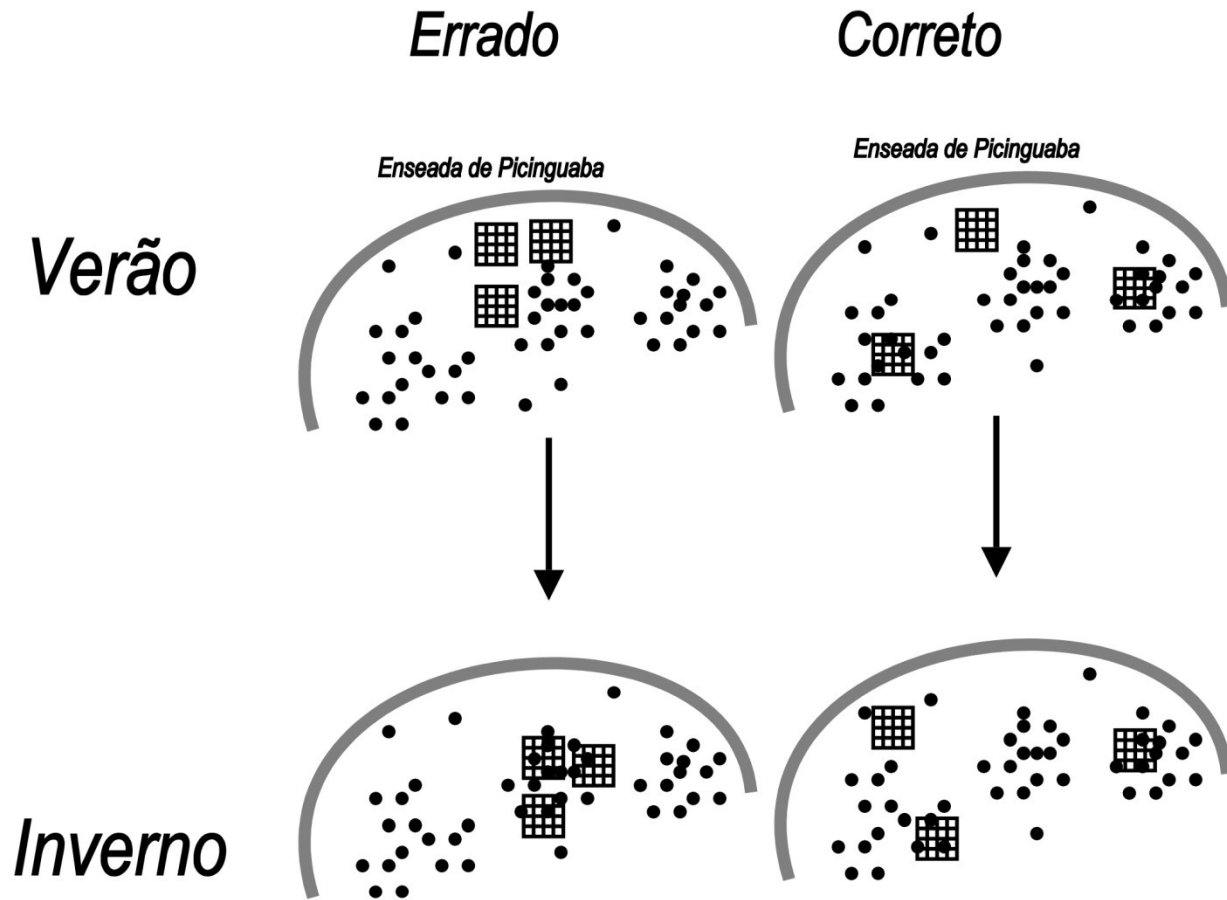
# PSEUDOREPLICAÇÃO ESPACIAL

## Estudo espacial



# PSEUDOREPLICAÇÃO ESPACIAL

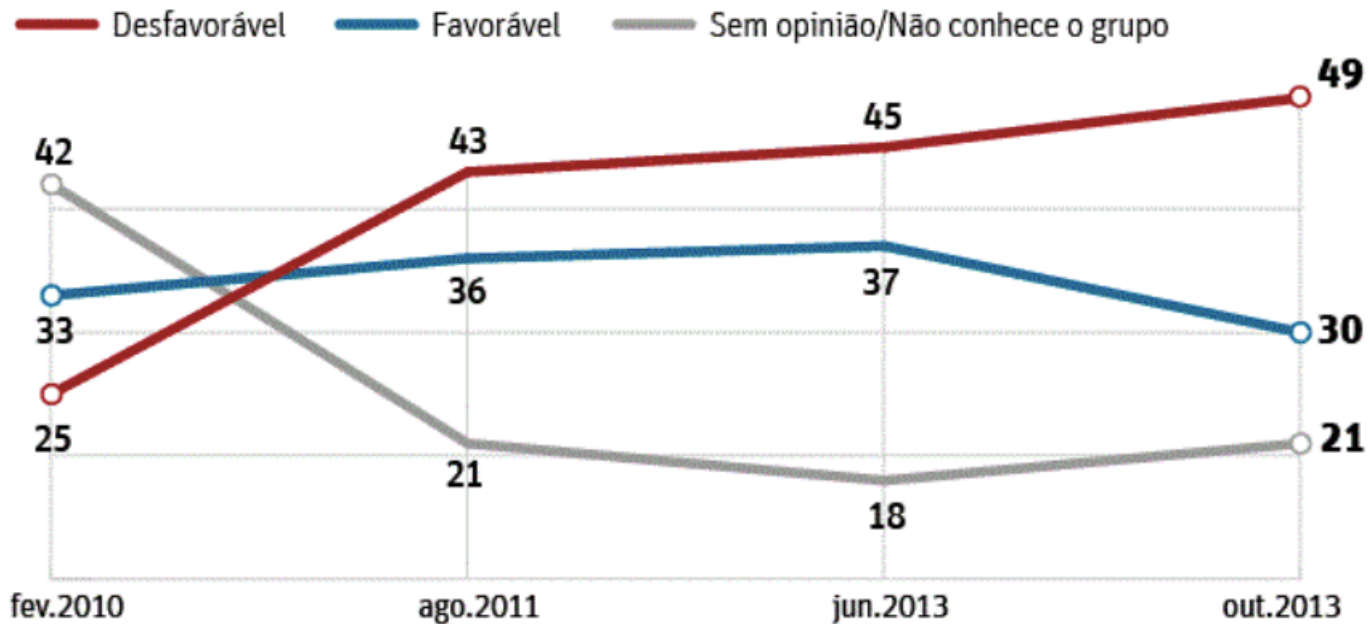
## Estudo Temporal



# PSEUDOREPLICAÇÃO TEMPORAL

## CAI APROVAÇÃO DO MOVIMENTO

Opinião dos americanos sobre o Tea Party, em %

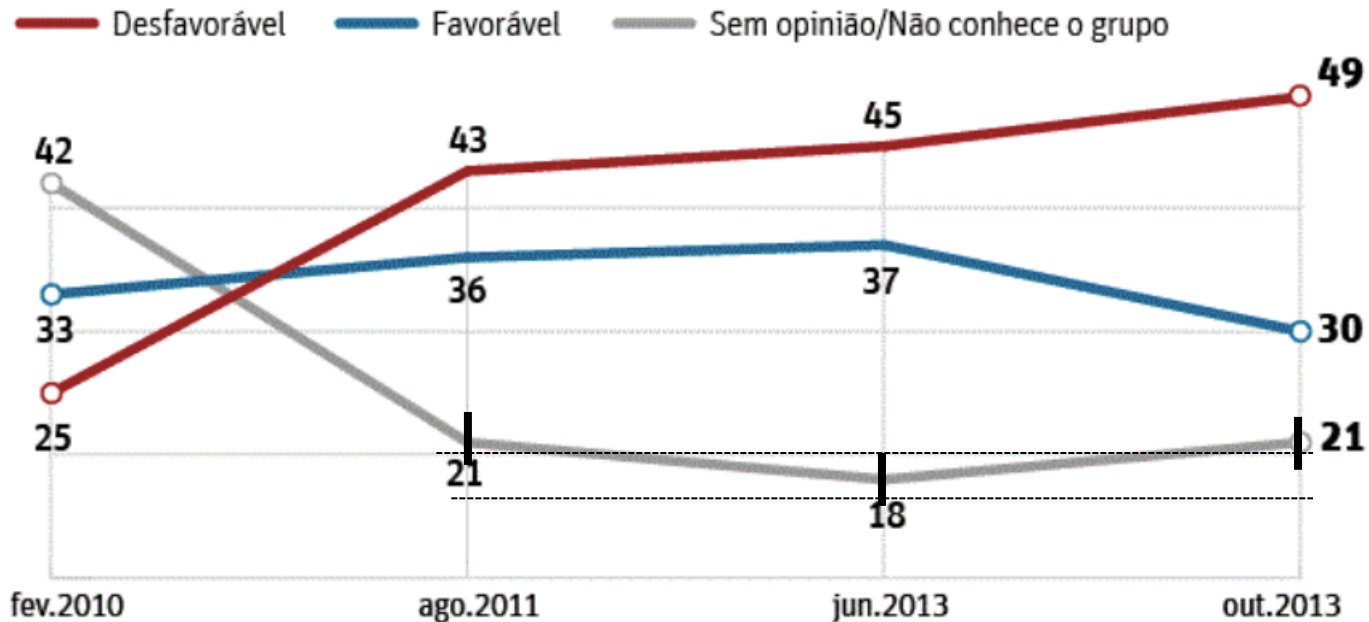


Fonte: Pew Research Center

# PSEUDOREPLICAÇÃO TEMPORAL = Não

## CAI APROVAÇÃO DO MOVIMENTO

Opinião dos americanos sobre o Tea Party, em %



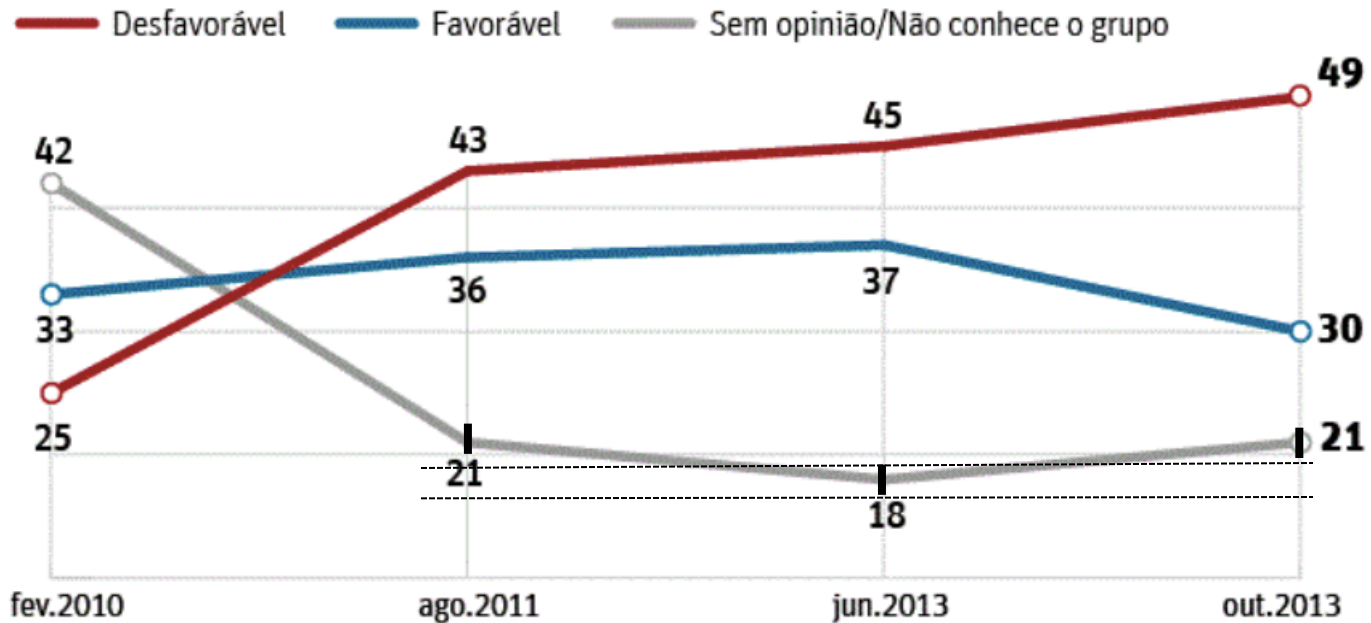
Fonte: Pew Research Center



# PSEUDOREPLICAÇÃO TEMPORAL = Sim

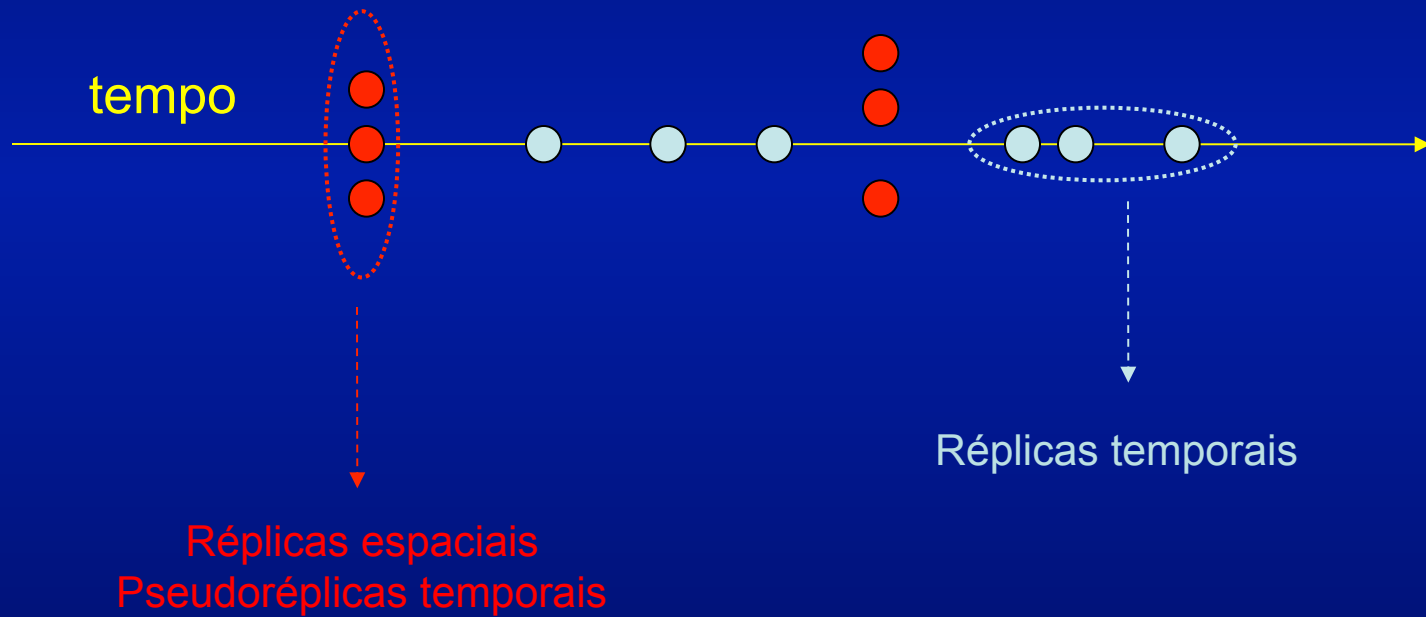
## CAI APROVAÇÃO DO MOVIMENTO

Opinião dos americanos sobre o Tea Party, em %

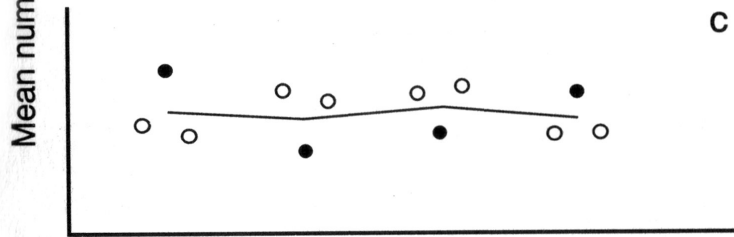
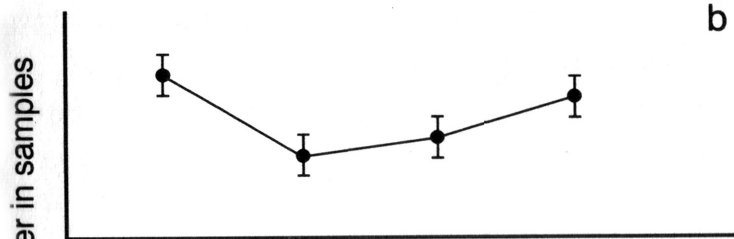
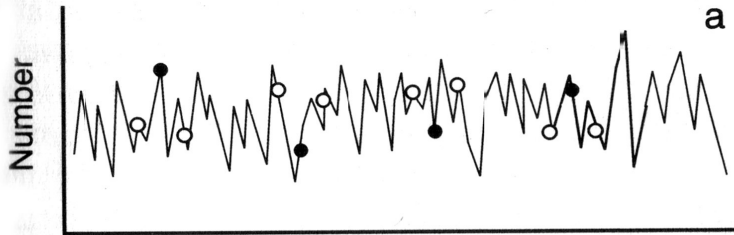


Fonte: Pew Research Center

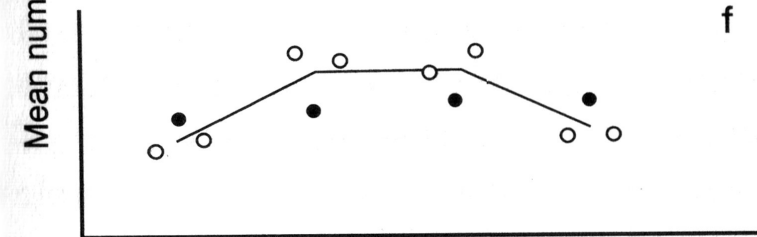
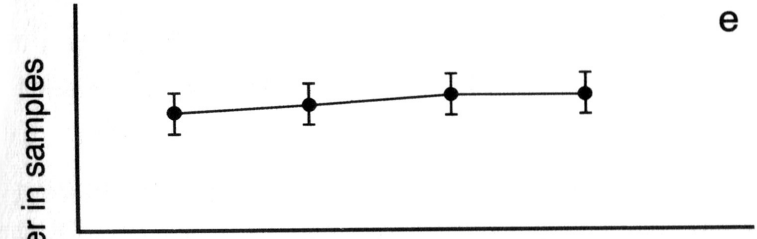
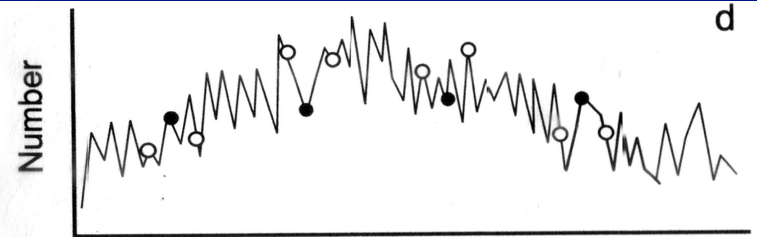
# PSEUDOREPLICAÇÃO TEMPORAL



# PSEUDOREPLICAÇÃO TEMPORAL

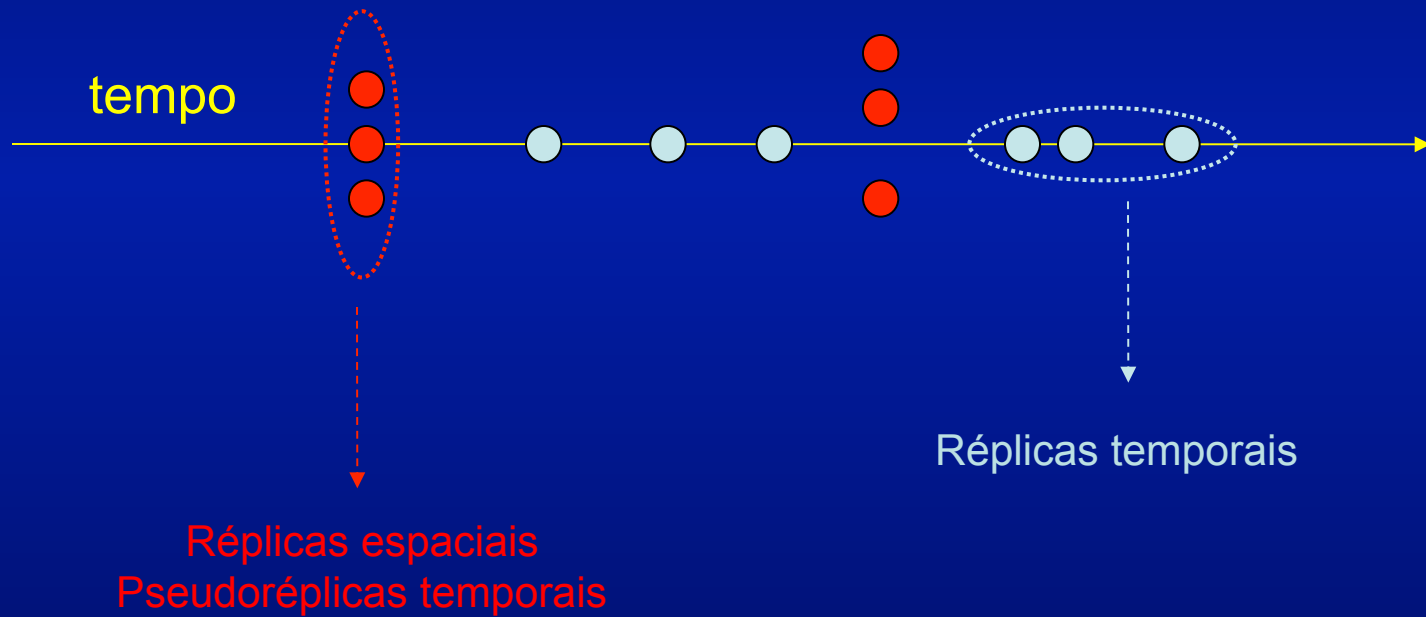


Época do ano



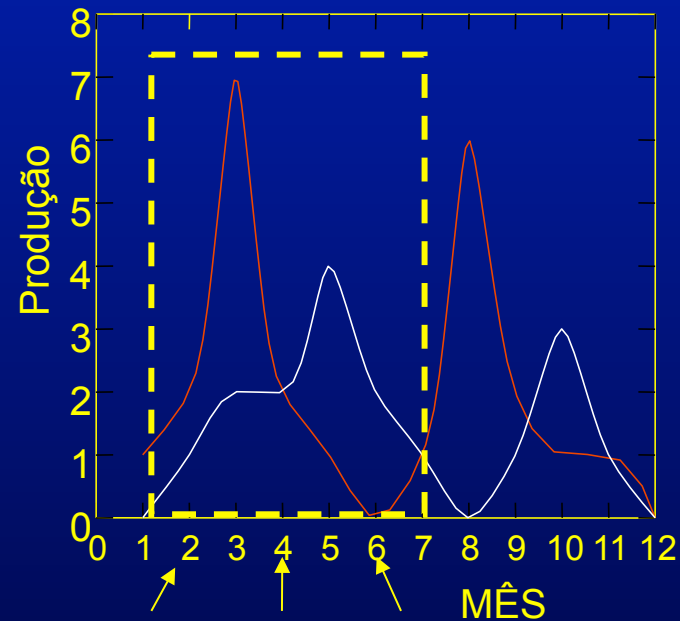
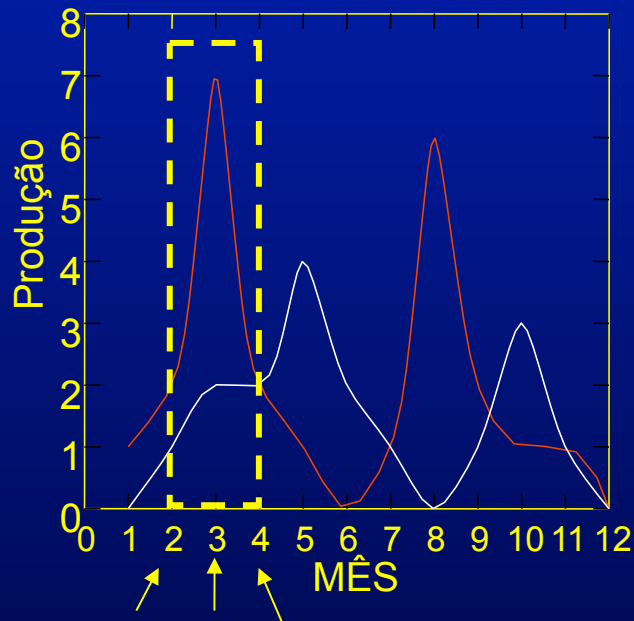
Época do ano

# PSEUDOREPLICAÇÃO TEMPORAL



# PSEUDOREPLICAÇÃO TEMPORAL

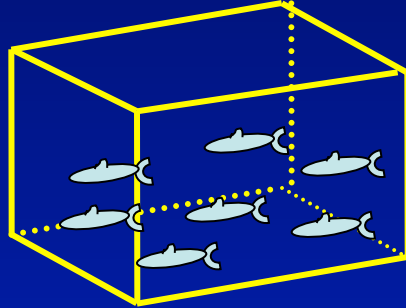
- Gravidade do erro depende da escala temporal:
  - Ciclo de vida do organismo (Nematódeo x Baleia)
  - Duração do processo (janela de leitura)



# PSEUDOREPLICAÇÃO ESPACIAL

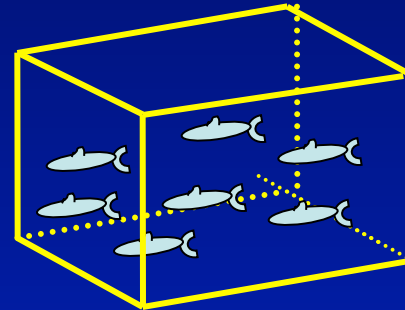
Experimento manipulativo

Aquário 1

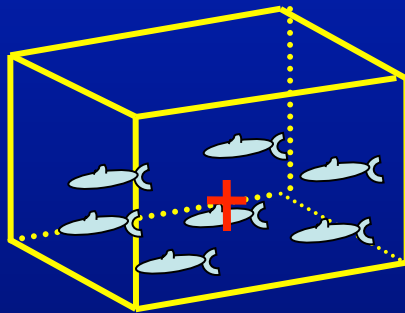


20° C

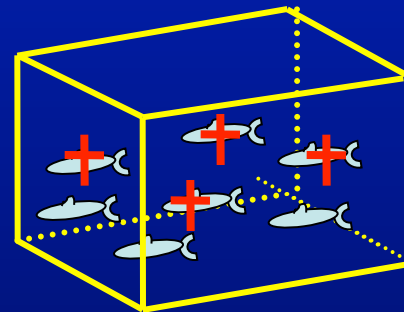
Aquário 2



30° C



20° C

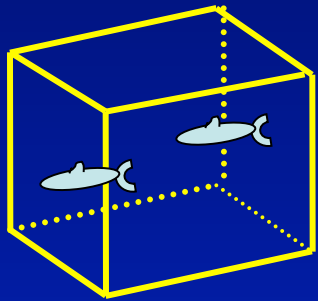


30° C

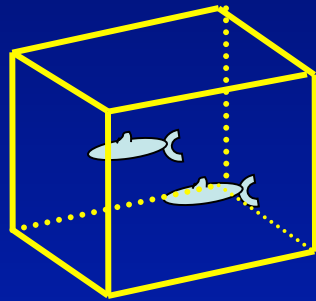
TEMPERATURA ou AQUÁRIO ?

# PSEUDOREPLICAÇÃO ESPACIAL

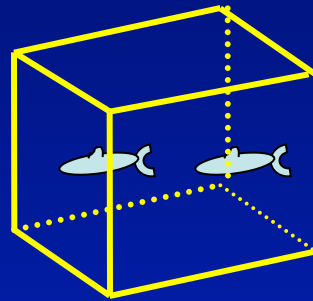
## Experimento manipulativo



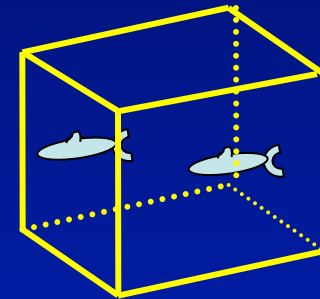
20° C



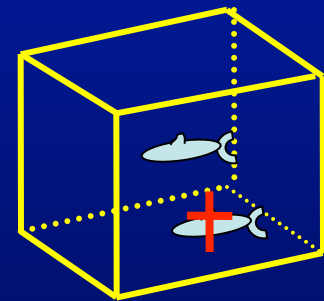
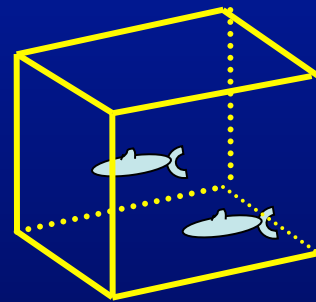
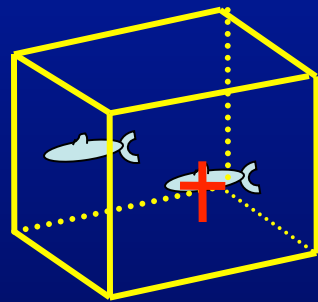
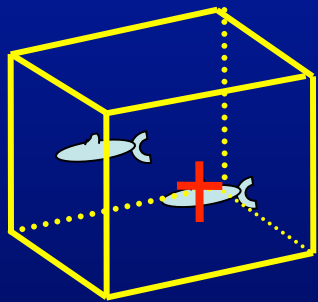
30° C



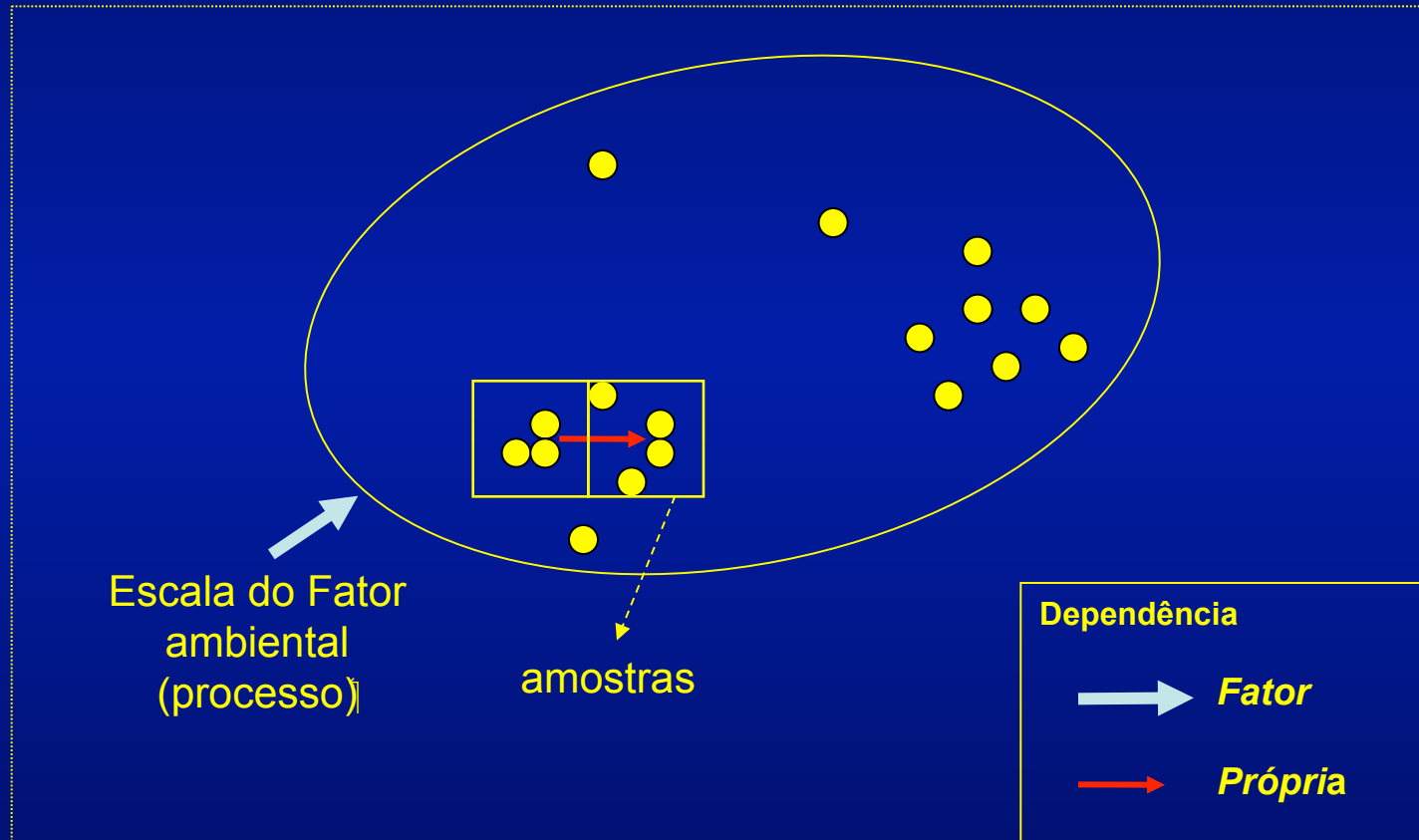
20° C



30° C

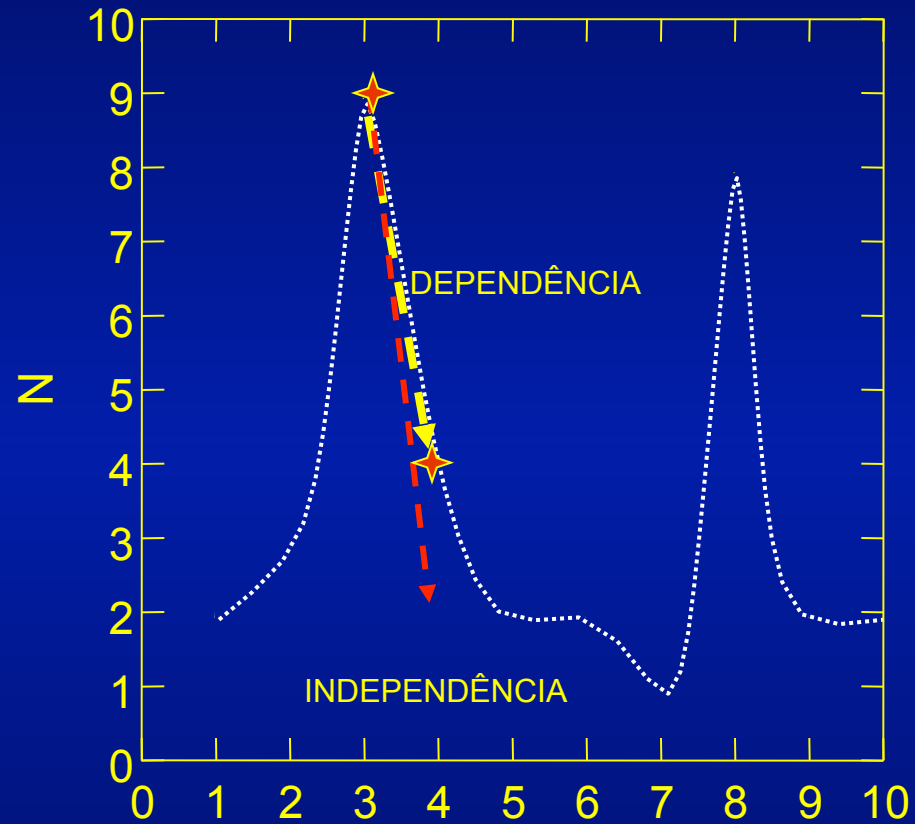


# DEPENDÊNCIA AMOSTRAL (ESPACIAL)





# DEPENDÊNCIA AMOSTRAL (TEMPORAL)



# PSEUDOREPLICAÇÃO (o pós-Hulbert, 1984)

- Extrapolação × Generalização
- Não independência ('confounding')
- Exageros
- Soluções:
  - ✓ Controle maior das variáveis (estudos exploratórios)
  - ✓ Estudos em múltiplas escalas (hierárquicos)
  - ✓ Análises com controle da dependência
    - Análise de séries temporais
    - Estatística espacial
    - Modelos mixtos – fatores aleatórios para controlar