Metodologia Ecológica
Abordagens e Métodos Estatísticos em Ecologia

Aula 3

- Probabilidades e amostras independentes

O cálice sagrado: independência entre observações

- Como obter?

- Relação com 'graus de liberdade'

Prof. Marcus Vinícius Vieira - Instituto de Biologia UFRJ

Eventos e resultados

- A probabilidade de um **evento** único: captura de presas por plantas carnívoras

- Evento – processo simples em que há um início e fim bem definidos

- Exemplo:
  - Plantas carnívoras: apenas alguns dos insetos que visitam a planta caem em sua armadilha. Como estimar a probabilidade de um inseto que visita a planta cair em sua armadilha?
  - Evento – visita a planta
  - Resultados possíveis – captura ou fuga (**resultado discreto**)

Sarracenia purpurea

Jargões básicos

- O resultado é discreto: 1 – captura; 2- fuga.
- O conjunto de todos os possíveis resultados é o **espaço amostral**.
- No caso, o espaço amostral é um **conjunto discreto**.
- Cada inseto que visita a planta consiste em uma **tentativa** ou **replica**.
- O conjunto de replicas forma um **experimento** (no sentido estatístico, mais amplo).

Exemplo

- Visitas (= tentativas) = 3000
- Capturas = 30
- Probabilidade que um inseto visitando a planta seja capturado: 30/3000 = 0.01
- Esta probabilidade no entanto, é limitada ao espaço amostral analisado.
- Como podemos tornar o resultado mais geral?

  Como inferir se esta é a probabilidade “real”, na população como um todo?

  → Delinear um experimento para inferir a probabilidade **na população** que um inseto visitando uma planta seja capturado

Limitações da definição de probabilidade

- Restrita à forma como determinamos, medimos o evento
- Ex.: sensores nas faces da moeda
- Plantas carnívoras: variação entre indivíduos, tipos de presas, época, etc.

Matemática das Probabilidades

- Definindo o espaço amostral:

Visita = {(*Captura*), (*Fuga*)}

- Primeiro axioma de probabilidade:

Axioma 1: A soma de todas as probabilidades de resultados dentro de um espaço amostral = 1,0

- Os eventos têm que ser **mutuamente exclusivos** e o espaço amostral **completo** (*exhuastive*) para que seja verdadeiro

Besouro de manchas laranjas (hipotético):

- ao longo da vida tem apenas duas proles, digamos *a* e *b*
- *a* = primeira prole   *b* = segunda prole
- Como representar o espaço amostral da aptidão deste besouro?
- $\Omega = \{ a, b \}$
- Digamos que *a* e *b* podem ser 2, 3 ou 4
- Como seria o espaço amostral então?

Espaço amostral das proles do Besouro de manchas laranjas

- {(2, 2), (2, 3), (2, 4), (3, 2), (3, 3), (3, 4), (4, 2), (4, 3), (4, 4)}

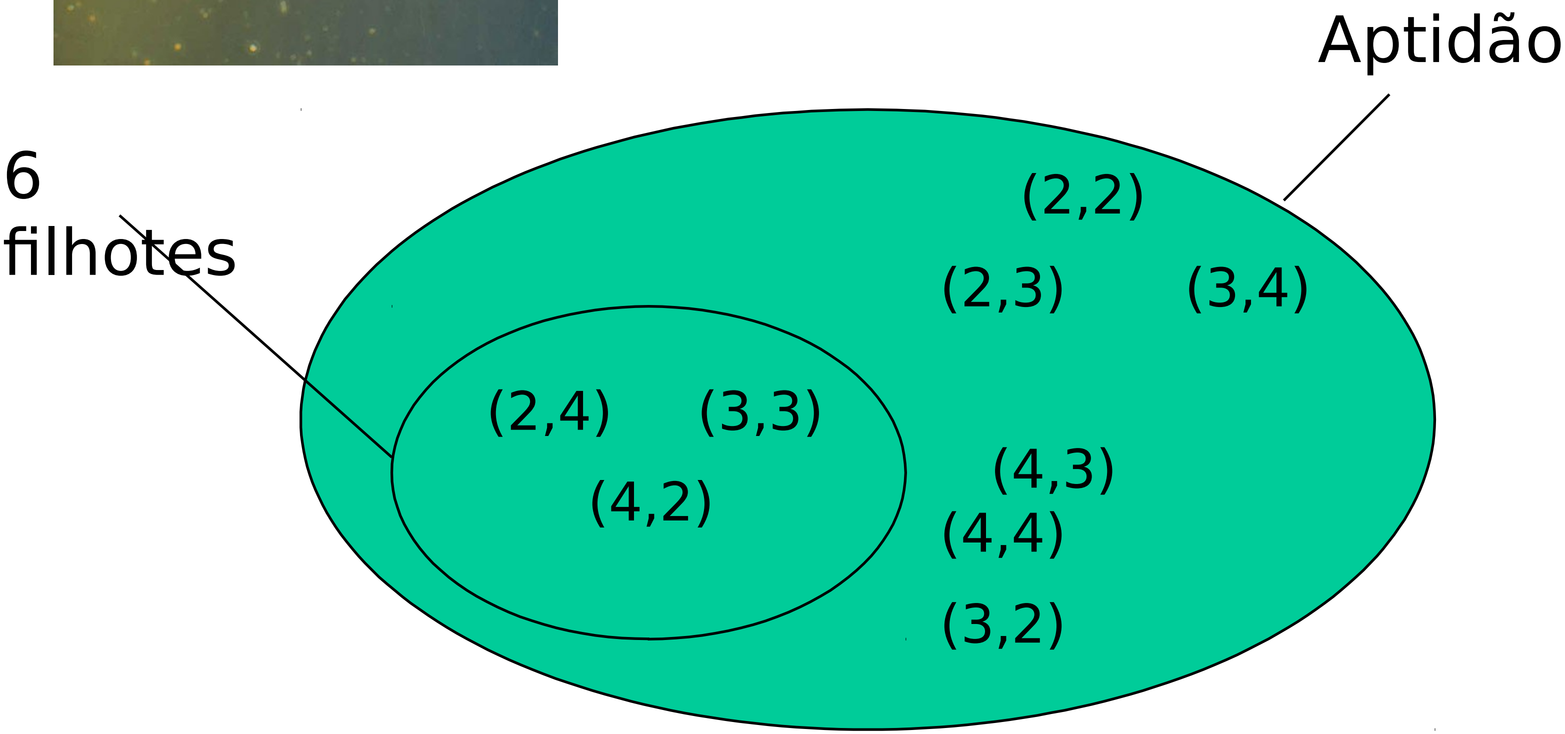
- Assumindo que a probabilidade de cada um é igual
- e seguindo o Axioma 1, qual a probabilidade de cada resultado?

Eventos complexos x eventos compartilhados ou

Adição x multiplicação de probabilidades

- Eventos complexos: Evento X pode ser resultado do Evento A **ou** Evento B **ou** Evento C
- Portanto, o Evento X poder ocorrer de mais de uma forma ou caminho
- Ex. Besouros de manchas laranjas (Gotteli & Ellison). Número total de filhotes ao longo da vida, sendo que tem apenas duas proles (litters), com um mínimo de 2 e um máximo de 4 filhotes por prole.
- Qual a probabilidade de ter um sucesso reprodutivo ao longo da vida de 6 filhotes? (5 min. para resolver)

- {(2,4), (3,3), (4,2)}



- *6 filhotes* é um subconjunto próprio do conjunto *Aptidão*
- *6 filhotes* $\subset$ *Aptidão*

Eventos complexos

- **Segundo axioma de probabilidade:**
- **Axioma 2**: a probabilidade de um evento complexo é igual à soma das probabilidades dos resultados que compõem este evento.
- O evento “6 proles ao longo da vida” é complexo.

Eventos compartilhados ou simultâneos

- Qual a probabilidade de ocorrência do Evento A **e** Evento B **e** Evento C?
- Probabilidade simultânea de dois ou mais eventos.
- SE FOREM INDEPENDENTES, PREMISSA CRUCIAL, basta multiplicar suas probabilidades.

Eventos independentes

- Dois eventos são independentes se o resultado de um evento não é afetado pelo resultado do outro.
- Ex. Qual a probabilidade de um indivíduo do besouro ter proles de 2 e 4 filhotes ao longo da vida? Basta multiplicar a probabilidade de ter uma primeira prole de 2 pela de ter uma prole de 4 na segunda.
- Sendo 1/3 cada uma (Diagrama de Venn), então 1/9.
- Se forem independentes, a multiplicação é igual à interseção entre dois conjuntos de eventos:

$$P(A \cap B) = P(A).P(B)$$

Eventos independentes

$P(A \cap B) = P(A).P(B)$

$P(2,4) = P(2).P(4) = 1/3 \times 1/3$

$P(2,4) = 1/9$

- Se forem F e S forem independentes!

Asclepias e largartas de borboletas Monaca

- P(Resistente) = 0,20
- P(não R) = 1 – P(R) = 0,80

- P(Lagarta) = 0,70
- P(não L) = 1 – P(L) = 0,30

PREMISSAS:

- Todas as plantas e lagartas podem dispersar e alcançar todas as manchas de habitat, independentemente uma da outra.
- Populações locais de *Asclepias* sempre persistem na ausência de lagartas,
- mas quando lagartas estão presentes só populações resistentes persistem.

Probabilidades de eventos compartilhados (mas independentes!)

Evento compartilhado	**Cálculo**	***Asclepias* resistente ?**	**Lagartas presentes ?**
População suscetível e nenhuma lagarta	[1 - P(R)] x [1 - P(C)] = (1,0 - 0,2) x (1,0 - 0,7) = 0,24	Não	Não
População suscetível e lagartas presentes	[1 - P(R)] x [P(C)] = (1,0 - 0,2) x (0,7) = 0,56	Não	Sim
População resistente e nenhuma lagarta	[P(R)] x [1 - P(C)] = (0,2) x (1,0 - 0,7) = 0,06	Sim	Não
População resistente lagartas presentes	[P(R)] x [P(C)] = (0,2) x (0,7) = 0,14	Sim	Sim

- Quatro eventos compartilhados. Formam um conjunto fechado?

  0,24 + 0,56 + 0,06 + 0,14 = 1,00

- Qual a probabilidade de persistência de *Asclepias* com estas probabilidades?

  0,24 + 0,06 + 0,14 = 0,44

Aptidão = {(2,2) (2,3) (2,4) (3,2) (3,3) (3,4) (4,2) (4,3) (4,4)}

F = {1a. Prole com 2 filhotes}

S = (2a. Prole com 4 filhotes}

$F \cup S$ = ? $F \cap S$ =?

F = {(2,2) (2,3) (2,4)}

S = {(2,4) (3,4) (4,4)}

$F \cup S$ = {(2,2) (2,3) (2,4) (3,4) 4,4)}

$F \cap S$ = {(2,4)}

F = {(2,2) (2,3) (2,4)}

S = {(2,4) (3,4) (4,4)}

F ∪ S = {(2,2) (2,3)
(2,4) (3,4)
4,4)}

F ∩ S = {(2,4)}

F ∪ S = F + S - F ∩ S

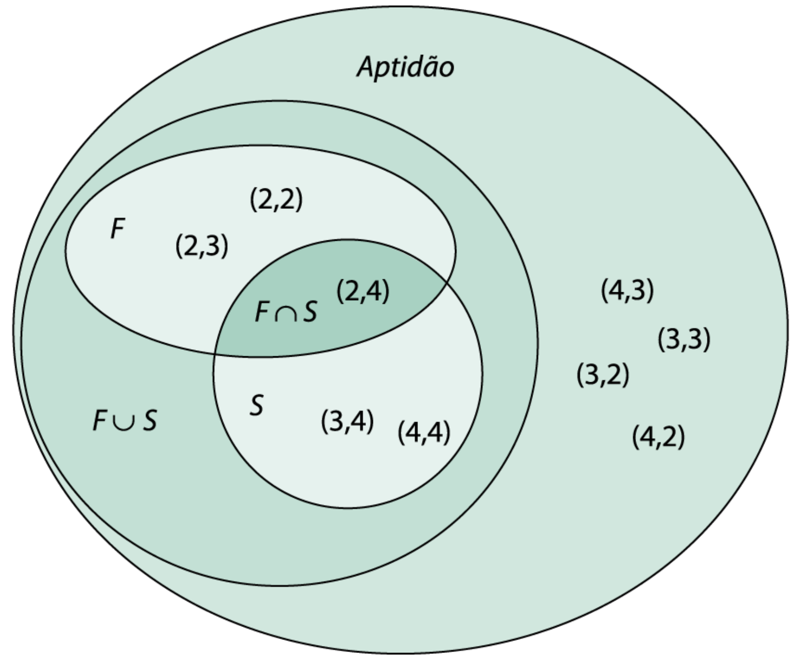


Figura 1.3 Diagrama de Venn ilustrando uniões e interseções de conjuntos. Cada anel representa um conjunto diferente do número de descendentes produzidos por um par de proles de uma espécie hipotética do besouro girinídeo, como na Figura 1.2. O anel maior representa o conjunto *Aptidão,* que compreende todos os resultados reprodutivos possíveis do besouro. O anel menor *F* é o conjunto de todos os pares de proles com exatamente dois descendentes na primeira prole. O anel menor *S* é o conjunto de todos os pares de proles com exatamente 4 descendentes na segunda prole. A área na qual os anéis se sobrepõem representa a interseção entre *F* e *S* ($F \cap S$) e contém os elementos comuns a ambos. O anel que compreende *F* e *S* representa a união entre *F* e *S* ($F \cup S$) e contém todos os elementos de ambos os conjuntos. Note que a união de *F* e *S* não duplica o elemento em comum (2,4). Em outras palavras, a união de dois conjuntos é a soma dos elementos de ambos, menos seus elementos em comum. Assim, $(F \cup S) = (F + S) - (F \cap S)$.

Probabilidades de eventos complexos: Probabilidade da união

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$F = \{(2,2)\ (2,3)\ (2,4)\}$$

$$S = \{(2,4)\ (3,4)\ (4,4)\}$$

$$F \cup S = \{(2,2)\ \ (2,3)\ (2,4)\ (3,4)\ 4,4)\}$$

$$\mathrm{F} \cap \mathrm{S} = \{(2,4)\}$$

- $F \cup S = F + S - F \cap S$

Eventos independentes

$P(A \cap B) = P(A).P(B)$

$P(2,4) = P(2).P(4) = 1/3 \times 1/3$

$P(2,4) = 1/9$

- Se forem F e S forem independentes!

E se os eventos forem dependentes?

- Retornando ao besouro de manchas laranjas: qual a probabilidade de tendo tido 2 filhotes na primeira prole, ter um total de 6 proles ao longo da vida?

- Espaço amostral antes:

  {(2, 2), (2, 3), (2, 4), (3, 2), (3, 3), (3, 4), (4, 2), (4, 3), (4, 4)}

- Agora:

  {(2, 2), (2, 3), (2, 4)}

Probabilidades de eventos dependentes ou condicionais

- P (1ª. Prole = 2) = ?

(1ª. Prole = 2) = {(2,4), (2,3), (2,2)}

P (1ª. Prole = 2) = 3/9 = 1/3

- P (6 proles | 1a. Prole = 2) = ?

(6 proles | 1a. Prole = 2) = { (2,4) }

P(6 proles | 1a. Prole = 2) = 1/3

MAS ORIGINALMENTE…. P (2,4) = 1/9

SE FOSSEM INDEPENDENTES!

http://blogdoxandro.blogspot.com/2013/08/tiras-n4928-recruta-zero-greg-e-mort_25.html

Como obter pontos independentes?

Cuidados no delineamento amostral

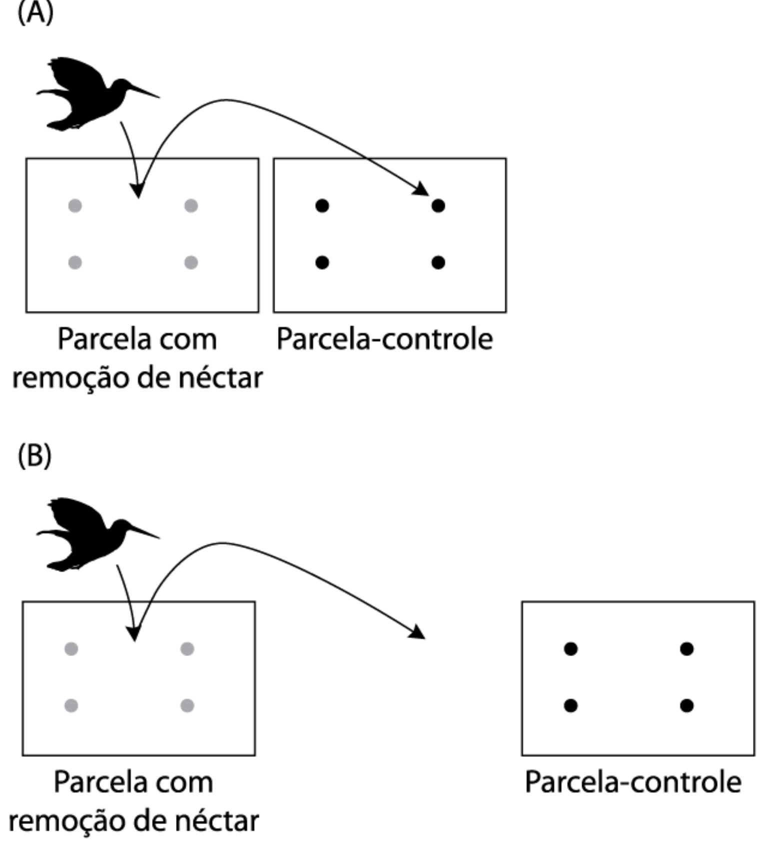


Figura 6.5 O problema de falta de independência em estudos ecológicos é ilustrado por um delineamento experimental em que beija-flores forrageiam em busca de néctar em parcelas controle e em parcelas das quais o néctar foi removido de todas as flores. (A) Em um esboço sem independência, as parcelas com remoção de néctar e controle são adjacentes uma a outra e beija-flores que entram na parcela de remoção de néctar imediatamente a deixam e começam a forragear na parcela-controle adjacente. Como consequência, os dados coletados na parcela controle não são independentes dos dados coletados na parcela com remoção de néctar: as respostas em um tratamento influenciam as respostas do outro. (B) Se o esboço for modificado de forma que as parcelas sejam amplamente separadas, os beija-flores que deixarem a parcela com remoção de néctar não necessariamente entrarão na parcela-controle. As duas parcelas são independentes e os dados coletados em uma parcela não serão influenciados pela presença da outra parcela. Apesar de ser fácil ilustrar o problema potencial da falta de independência, na prática pode ser muito difícil saber por antecipação a escala espacial e temporal que garantirá independência estatística.

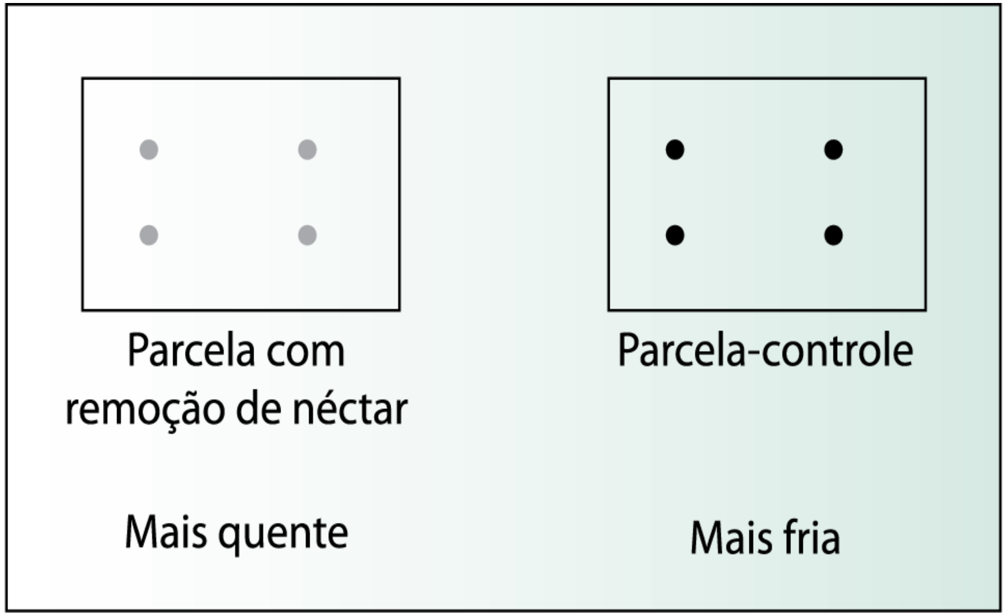


Figura 6.6 Um delineamento experimental confundido. Como na Figura 6.5, o estudo estabelece parcelas-controle e com remoção de néctar para avaliar as respostas no forrageamento de beija-flores. Nesse delineamento, embora as parcelas sejam suficientemente afastadas para garantir independência, foram colocadas em pontos diferentes ao longo de um gradiente termal. Por consequência, os efeitos dos tratamentos são confundidos com diferenças de temperatura. O resultado líquido é que o experimento compara dados de uma parcela quente com remoção de néctar com dados de uma parcela-controle fria.

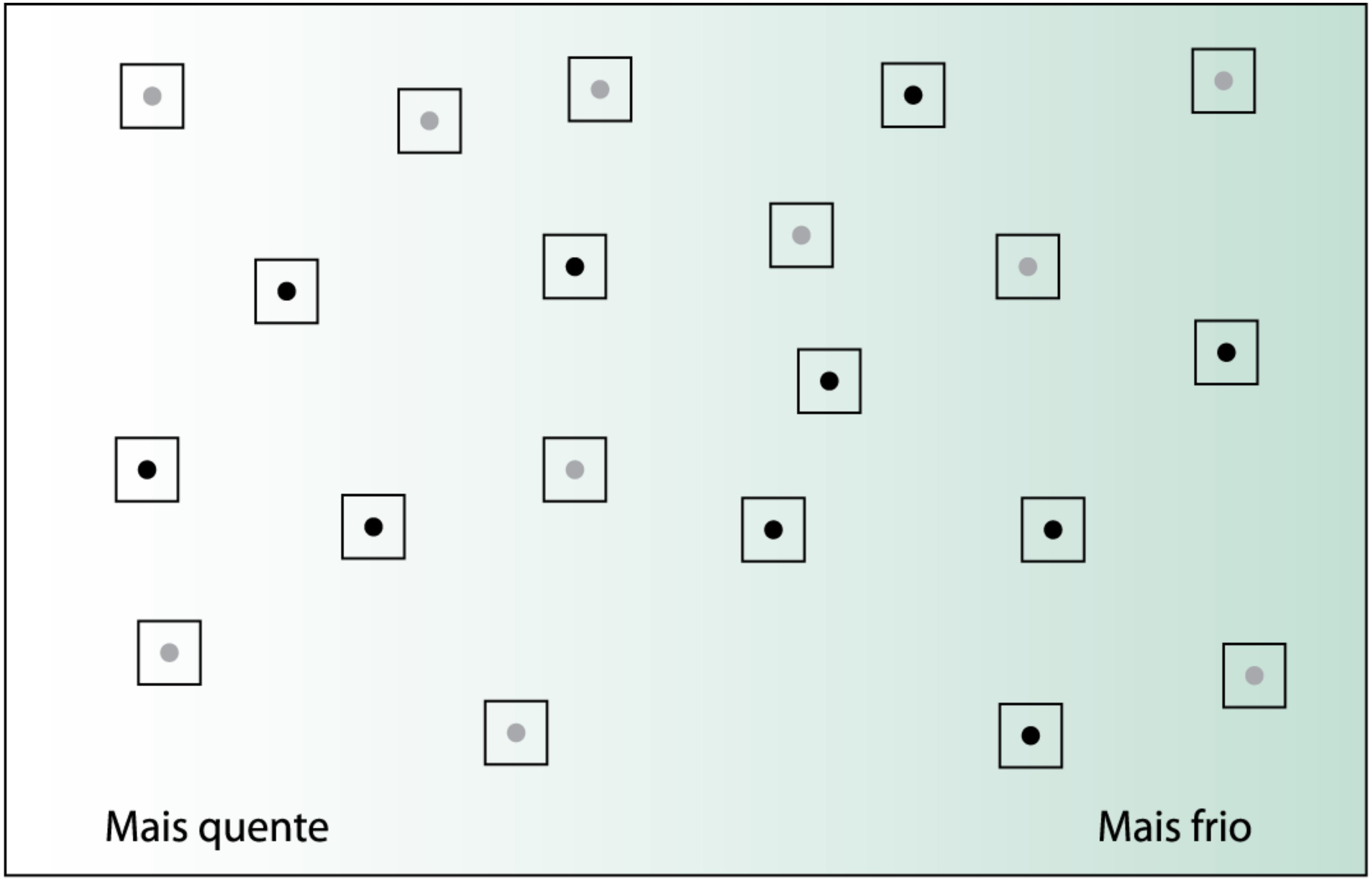


Figura 6.7 Um delineamento experimental devidamente replicado e aleatorizado. O estudo estabelece parcelas como na Figura 6.6. Cada quadrado representa uma réplica da parcela-controle (pontos pretos) ou da parcela com remoção de néctar (pontos cinzas). As parcelas são separadas por distância suficiente para garantir independência, mas sua localização dentro do gradiente de temperatura foi aleatorizada. Há 10 réplicas para cada um dos dois tratamentos. A escala espacial desse desenho é maior que na Figura 6.6.

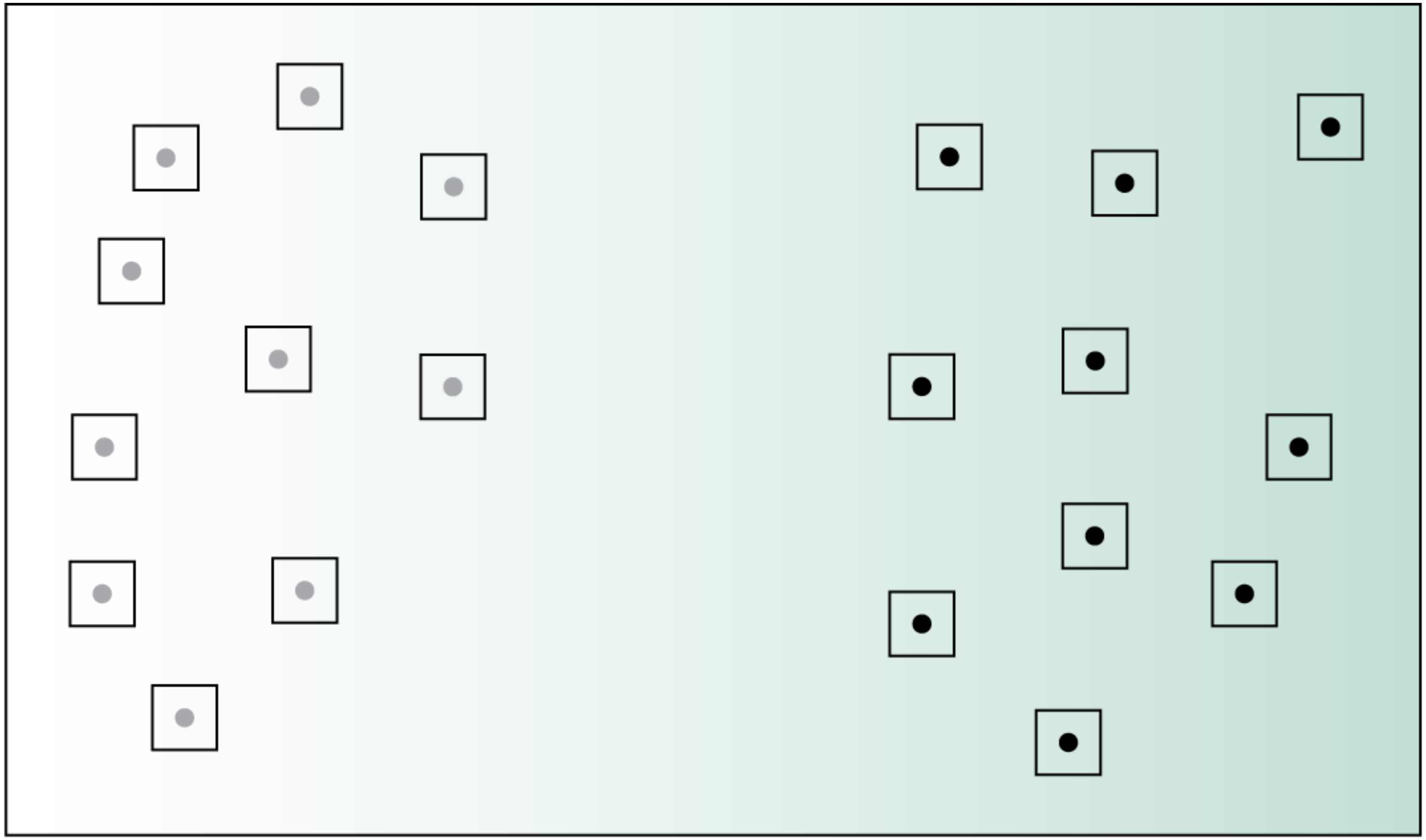
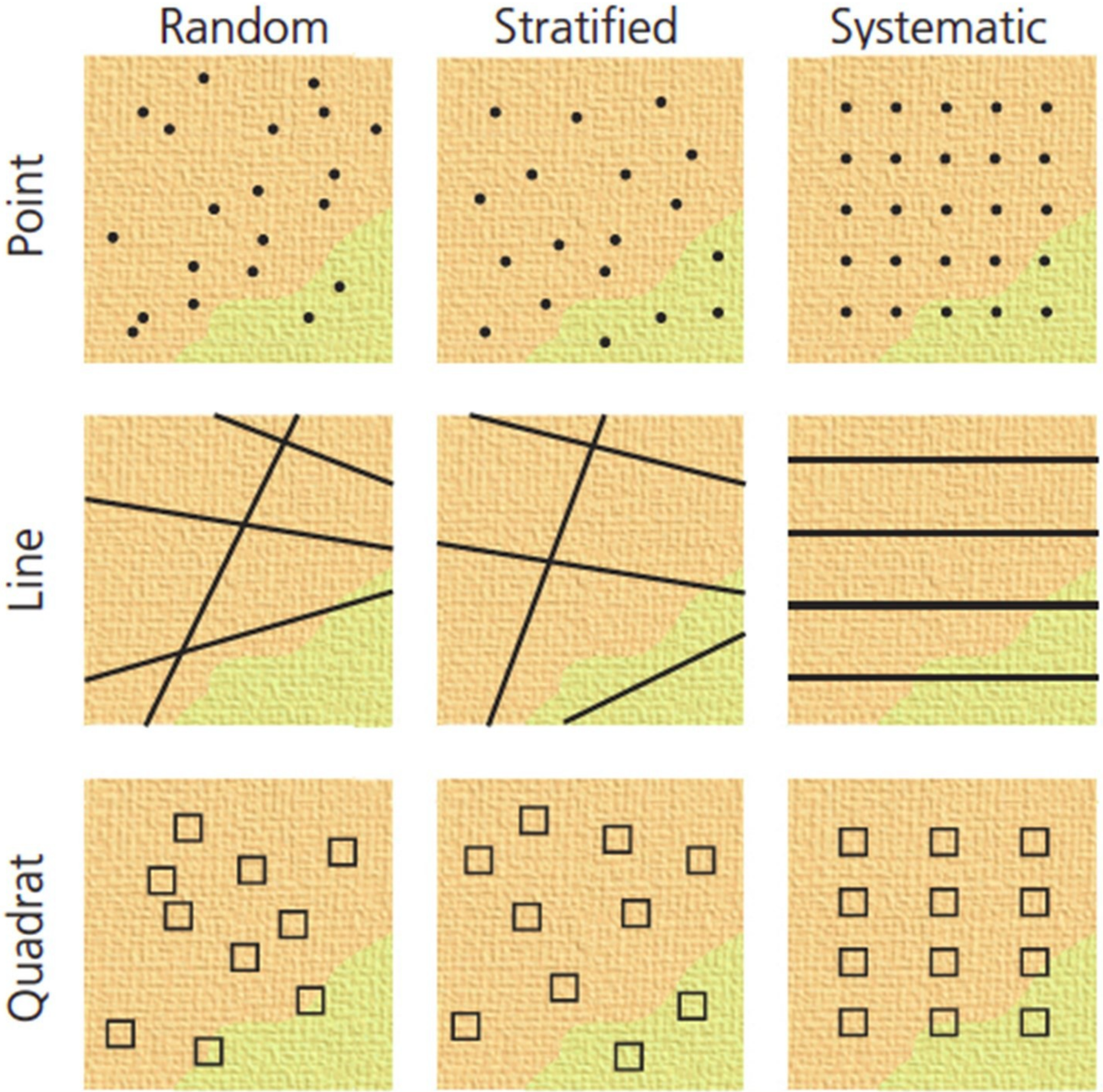
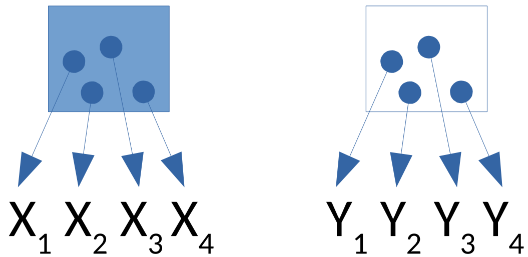


Figura 6.8 Um delineamento replicado, porém confuso. Como nas Figuras 6.5, 6.6 e 6.7, o estudo estabelece parcelas experimentais de controle e com remoção de néctar para avaliar as respostas no forrageamento de beija-flores. Cada quadrado representa uma parcela-controle (pontos pretos) ou parcelas com remoção de néctar (pontos cinza). Se os tratamentos forem replicados, mas não atribuídos aleatoriamente, o delineamento ainda confunde tratamento com gradientes ambientais adjacentes. A replicação combinada com randomização e espaçamento suficiente entre as réplicas (Figura 6.7) é a única garantia contra a falta de independência (Figura 6.5) e fatores confundidos (Figuras 6.6 e 6.8).

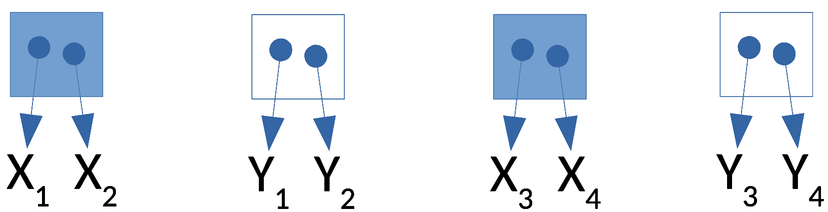


https://www.geography-fieldwork.org/a-level/before-starting/methods/sampling/

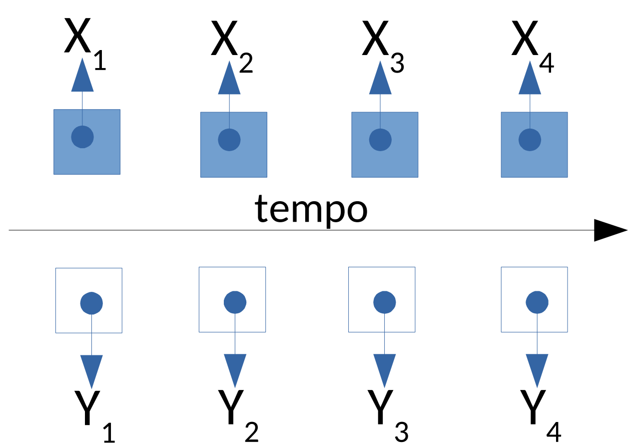
Pseudoreplicação simples

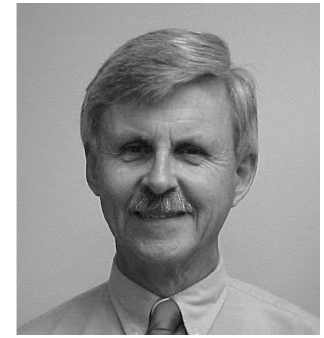


Pseudoreplicação ‘sacrificial’



Pseudoreplicação temporal





Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187–211.

Graus de liberdade

- Uma amostra com *n* = 5, média = 4
- Qual será a soma dos cinco números?
- Quantos valores o primeiro número poderia ter?
  - Digamos que fosse 2
- Quantos valores o segundo número poderia ter?
  - Digamos que fosse 7
- E os próximos poderiam ser 4 e 0
- Quantos valores poderia ter o último número?

Graus de liberdade

- De forma geral:
- g.l. = *n* – número de parâmetros estimados dos dados
- Retornando à variância,
- A média já foi estimada dos dados
- Então sobram *n* – 1 graus de liberdade para estimar a variância

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(Y_i - \bar{Y} \right)^2$$

Percentis e *box plots*

- Percentis mais usados: 90, 75, 50 (mediana)
- Coeficiente de variação
- Coeficiente de dispersão

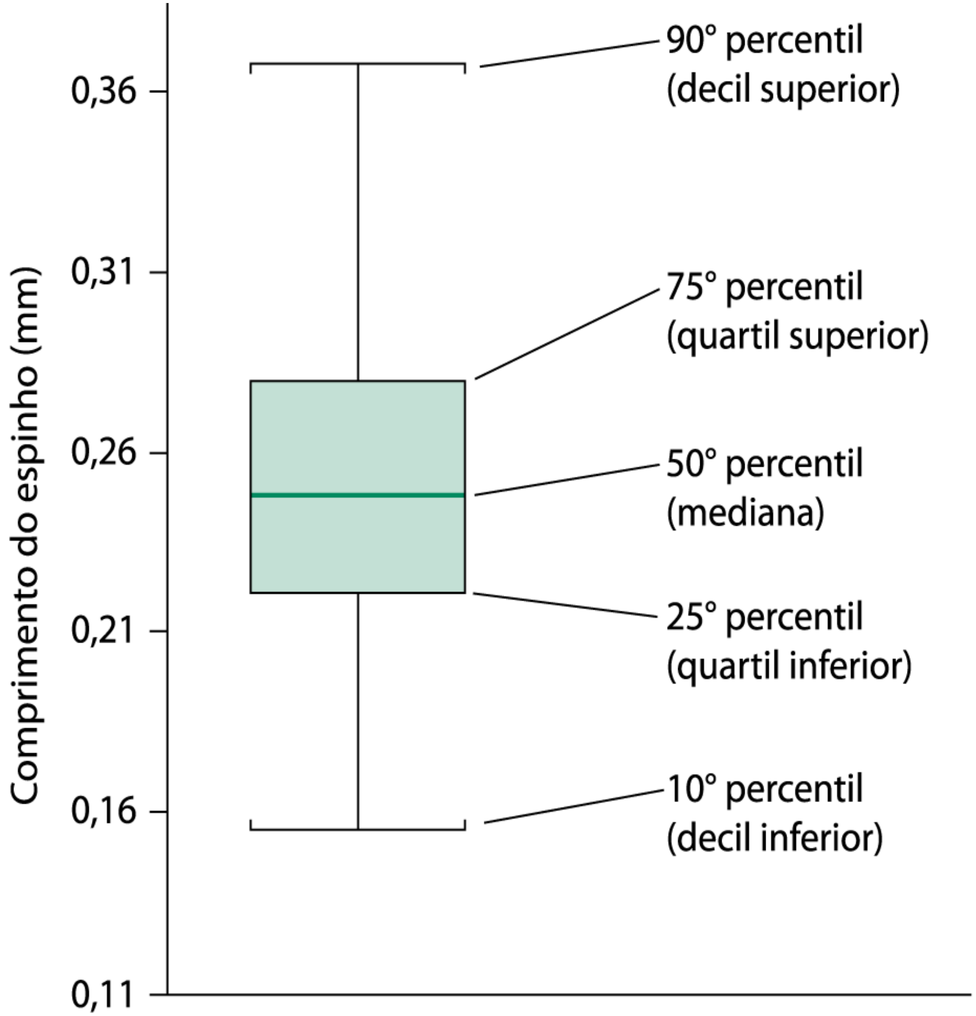


Figura 3.6 *Box plot* ilustrando os quantis dos dados da Tabela 3.1 ($n = 50$). A linha indica o 50° percentil (mediana) e a "caixa" engloba 50% dos dados, a partir do 25° até o 75° percentil. As linhas verticais se estendem do 10° ao 90° percentil.

Intervalos de confiança para *média*

P(média – 1,96ep ≤ μ ≤ média + 1,96ep) = 0,95

Intervalos de confiança generalizados

$$\bar{x} \pm t \bullet \frac{\sigma}{\sqrt{n}}$$

Distribuição *t* (de Student)

Intervalos de confiança para *média*

$P\,(-\,1{,}96 \leq Z \leq +\,1{,}96) = 0{,}95$

média – 1,96 ep ≤ μ ≤ média + 1,96 ep

$$\bar{x} \pm t \bullet \frac{\sigma}{\sqrt{n}}$$

Distribuição *t* (de Student)

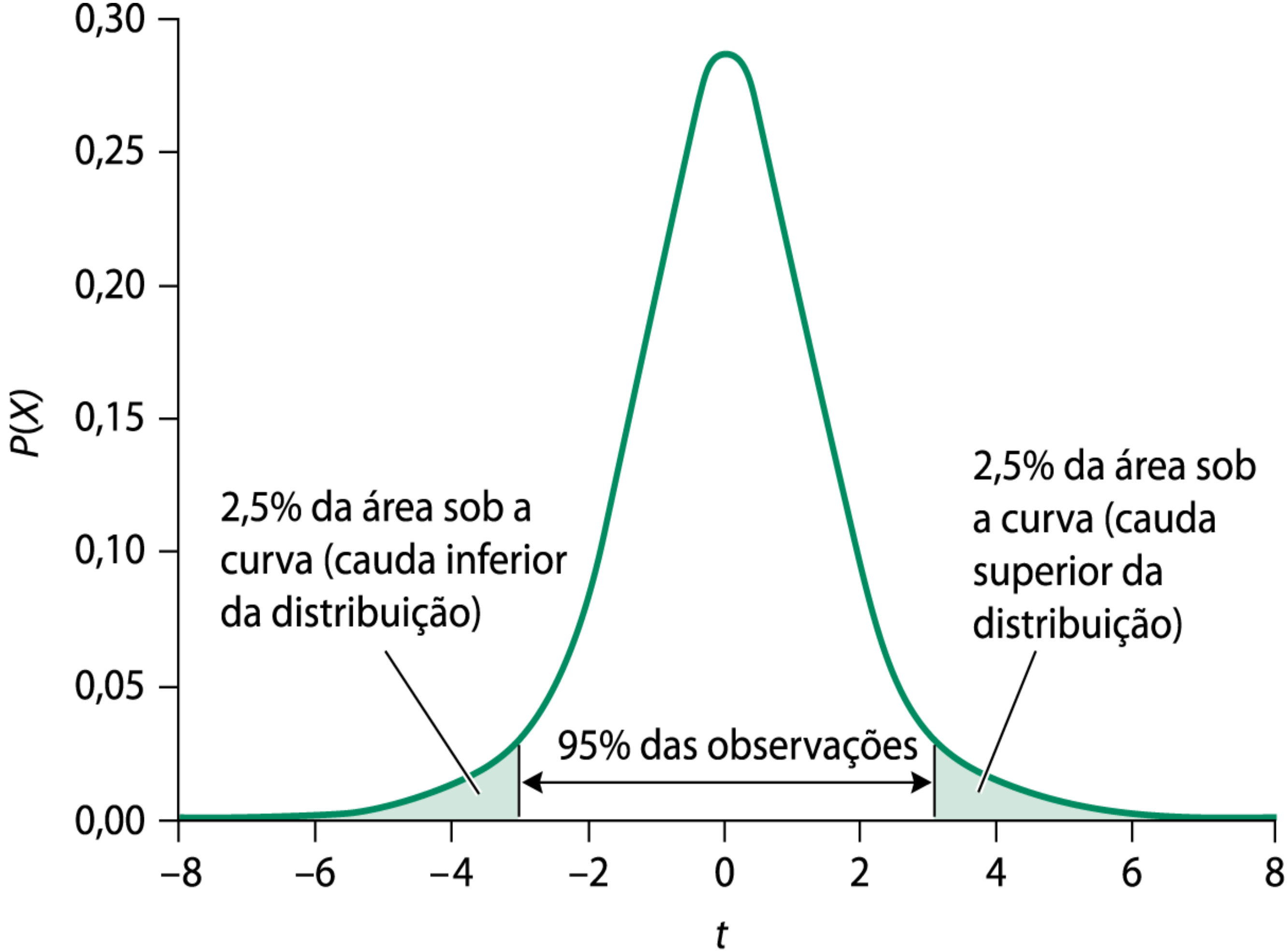


Figura 3.7 Distribuição-*t* ilustrando que 95% das observações, ou massa de probabilidade, caem dentro do percentil de ±1,96 desvio-padrão da média (média = 0). As duas caudas da distribuição contêm cada 2,5% das observações ou massa de probabilidade da distribuição. Elas somam 5% das observações, e a probabilidade $P = 0,05$ de que uma observação caia nessas caudas. Esta distribuição é idêntica à distribuição-*t* ilustrada na Figura 3.5.

Erro Tipo I : α

- Probabilidade de rejeitar Ho quando verdadeira
- Probabilidade de falso negativo
- Para calcular depende:

basta especificar Ho

- Teste conservador

Testes de hipótese nula comuns

- Diferença entre médias de duas amostras:
  - Teste *t*
  - Razão *F*
  - Premissas

Testes *t*

para testar Ho de diferença entre duas médias

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$t = \frac{Diferença\ entre\ médias}{Erro\ padrão\ da\ diferença\ entre\ médias}$$

Erro Tipo II : β

- Probalidade de não rejeitar Ho quando falsa, ou probabilidade de falso positivo, β
- Para calcular depende:
  - Da hipótese alternativa
  - Do tamanho do efeito que se pretende detectar
  - Do tamanho da amostra
  - Do delineamento experimental
- Poder: probabilidade de rejeitar corretamente Ho quando falsa: 1- β
- Princípio de precaução em monitoramento ambiental

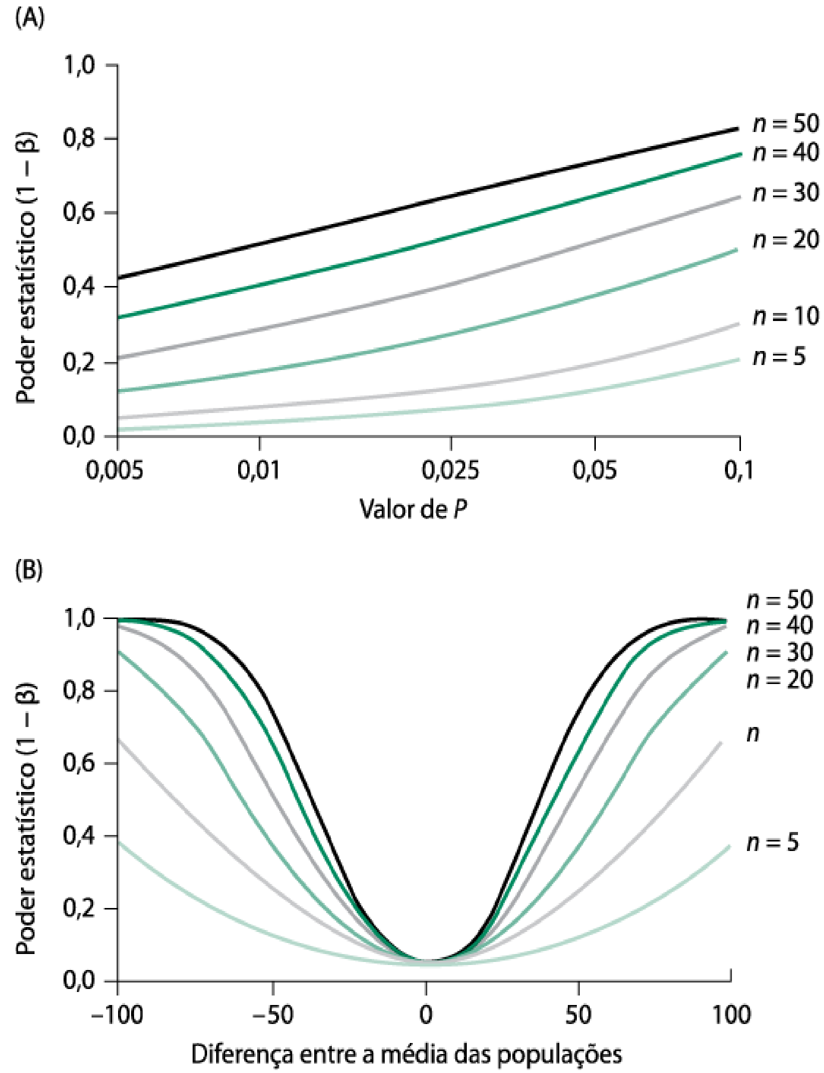


Figura 4.5 Relação entre poder estatístico, valor de *P* e tamanho do efeito observável em função do tamanho amostral. (A) O valor de *P* é a probabilidade de incorretamente rejeitar uma hipótese nula verdadeira, enquanto o poder estatístico é a probabilidade de corretamente rejeitar uma hipótese nula falsa. O resultado geral propõe que, quanto menor o valor de *P* usado para rejeitar a hipótese nula, menor o poder estatístico de corretamente detectar um efeito do tratamento. A um dado valor de *P*, o poder estatístico é maior quando o tamanho amostral é maior. (B) Quanto menor o tamanho do efeito observável do tratamento (i. e., quanto menor a diferença entre o grupo-tratamento e o grupo-controle), maior é o tamanho amostral necessário a um bom poder estatístico para detectar o efeito do tratamento.[21]

e Biologia UFRJ