

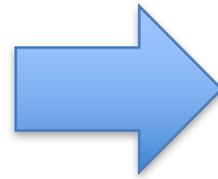
# O papel da estatística na ciência



Camila de Toledo Castanho

2020

# Por que eu preciso saber estatística?



*Ferramenta para tomada de decisão diante de incerteza*



# CIÊNCIA

*Sistema útil para responder perguntas*

# Procedimento Científico

## ESTATÍSTICA

- Planejamento



” To consult the statistician after a experiment is finished is often merely to ask him to conduct a post mortem examination - he can perhaps say what the experiment died of ”.

Ronald Fisher

# Procedimento Científico

## ESTATÍSTICA

- Planejamento
- Síntese dos resultados
- Interpretação diante de incerteza

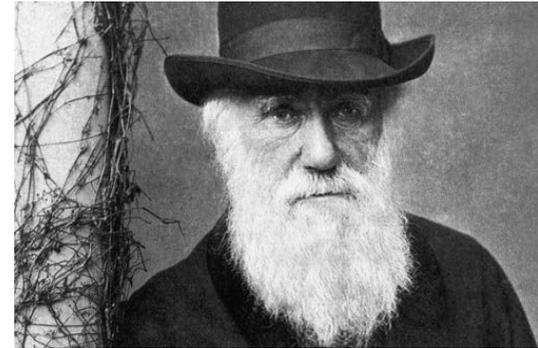


INCERTEZA



VARIAÇÃO NÃO PREVISÍVEL

Varição natural em sistemas biológicos é especialmente importante -> **Teoria da Seleção Natural**





*Em todas as ciências a Estatística é a linguagem comum usada para **planejar** a coleta, **sintetizar** e **interpretar** os dados que coletamos para responder perguntas*

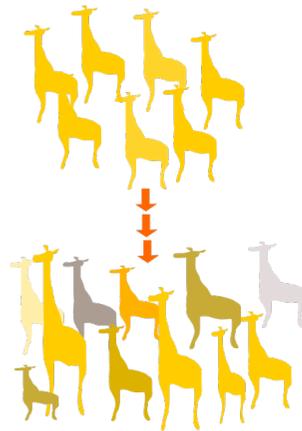
A Ciência em si é mais antiga que a estatística, porém a estatística revolucionou a ciência moderna

DETERMINISMO  
FILOSÓFICO



MODELO ESTATÍSTICO DA  
REALIDADE

(final do sec XX)



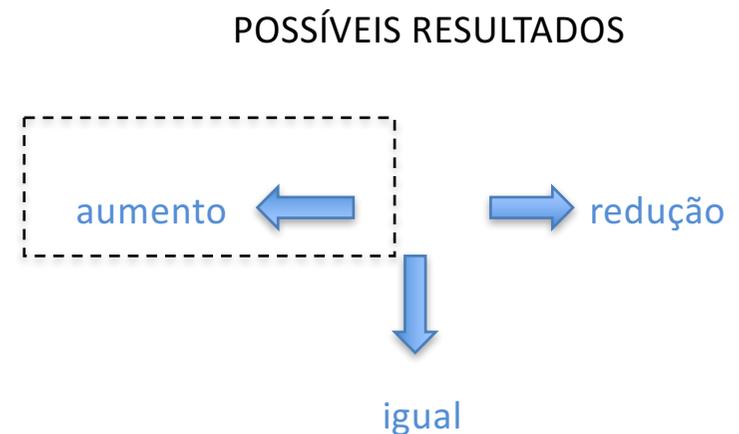
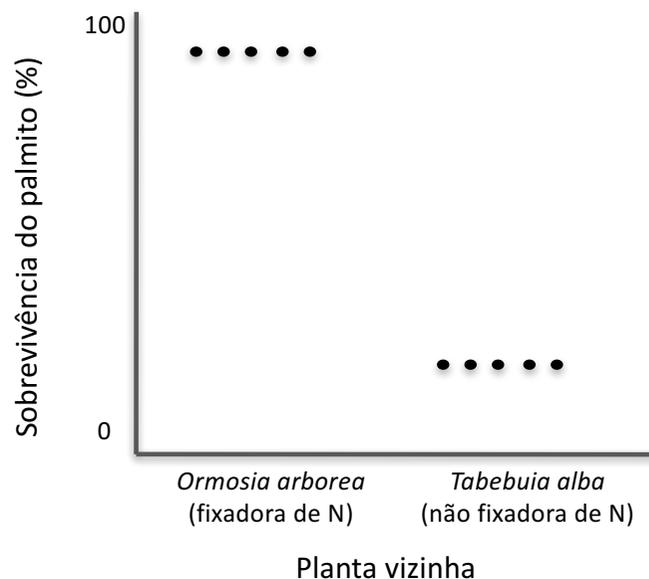
Salsburg, D. 2009. Uma senhora toma chá – como a estatística revolucionou a ciência do século XX. Editora Zahar

# A VARIABILIDADE NA PRÁTICA CIENTÍFICA

**Pergunta:** Plantas fixadoras de nitrogênio afetam o desempenho de plantas vizinhas não fixadoras?

**Hipótese:** Plantas fixadoras de nitrogênio umentam o desempenho de plantas vizinhas não fixadoras?

**Predição:** A sobrevivência do *Euterpe edulis* é maior quando vizinha de *Ormosia arborea* (fixadora) do que quando vizinha de *Tabebuia alba* (não-fixadora).

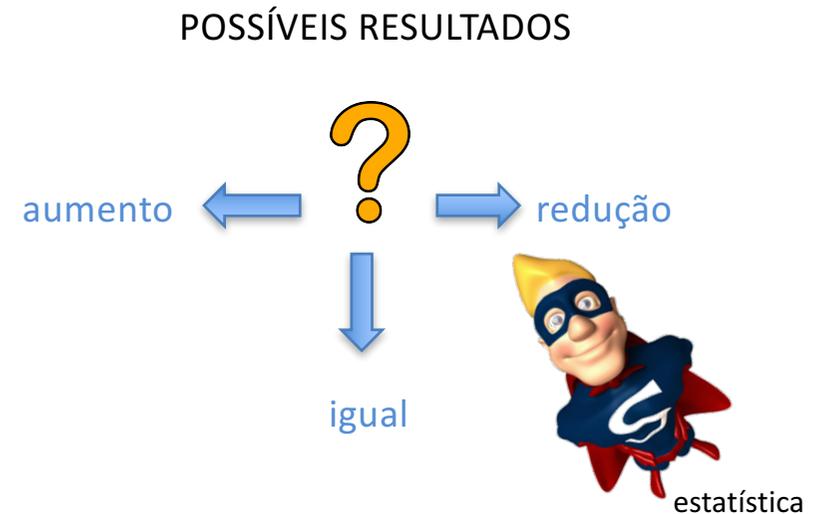
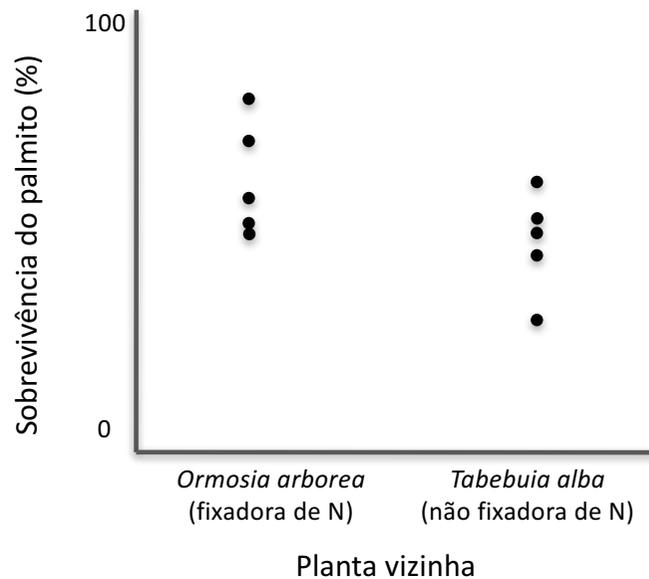


# A VARIABILIDADE NA PRÁTICA CIENTÍFICA

**Pergunta:** Plantas fixadoras de nitrogênio afetam o desempenho de plantas vizinhas não fixadoras?

**Hipótese:** Plantas fixadoras de nitrogênio umentam o desempenho de plantas vizinhas não fixadoras?

**Predição:** A sobrevivência do *Euterpe edulis* é maior quando vizinha de *Ormosia arborea* (fixadora) do que quando vizinha de *Tabebuia alba* (não-fixadora).



## VARIAÇÃO COMO NORMA

**Exemplos ecológicos:** número de visitas de polinizadores; diversidade de espécies no Cerrado;

**Exemplos cotidianos:** qualidade de vinhos; filmes de Hollywood; habilidade esportiva



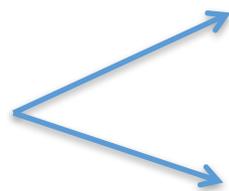
”Não é muito fácil determinar que proporção de um resultado se deve à **habilidade** e que proporção se deve à **sorte**”

”Quando examinamos feitos extraordinários nos esportes – ou em qualquer outra área-, devemos ter em mente que **eventos extraordinários podem ocorrer sem causas extraordinárias.**”

Leonard Mlodinow – O andar do bêbado

# VARIAÇÃO COMO NORMA

**VARIAÇÃO EM QUALQUER FENÔMENO**

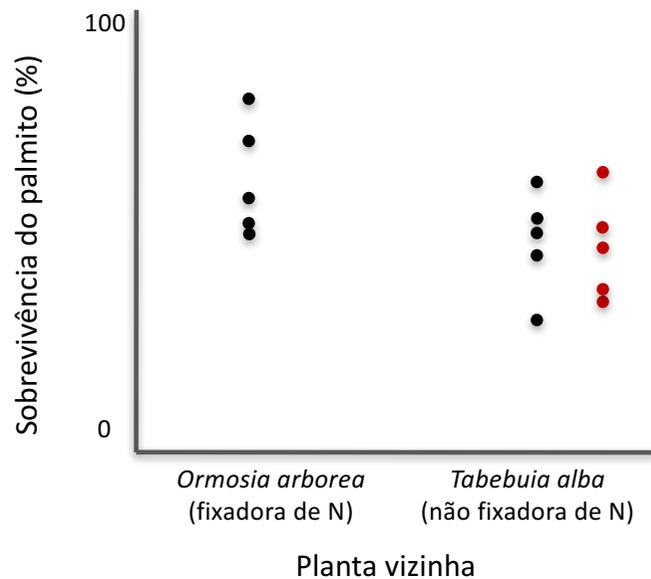


Variação aleatória, ao acaso ( “imprevisível”)

Sorte

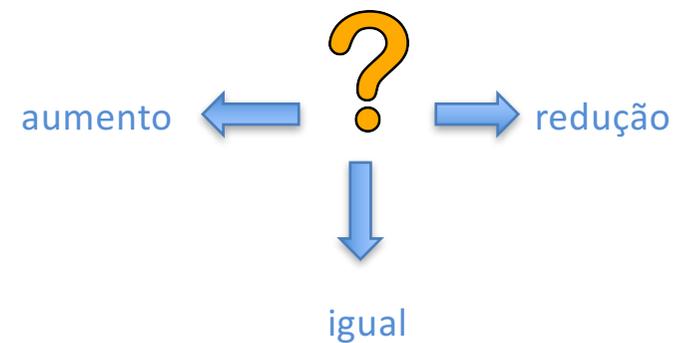
Variação determinística (previsível)

Habilidade



ERRO DE AMOSTRAGEM

POSSÍVEIS RESULTADOS



# ALEATORIEDADE

**Senso-comum:** exprime imprevisibilidade, ausência de causa

**Científico:** reconhece imprevisibilidade porém admite regularidades na ocorrência de eventos cujos resultados não são certos

## IMPÕE LIMITAÇÕES À IMPREVISIBILIDADE



Se 2 moedas forem jogadas qual o número de caras que pode aparecer?

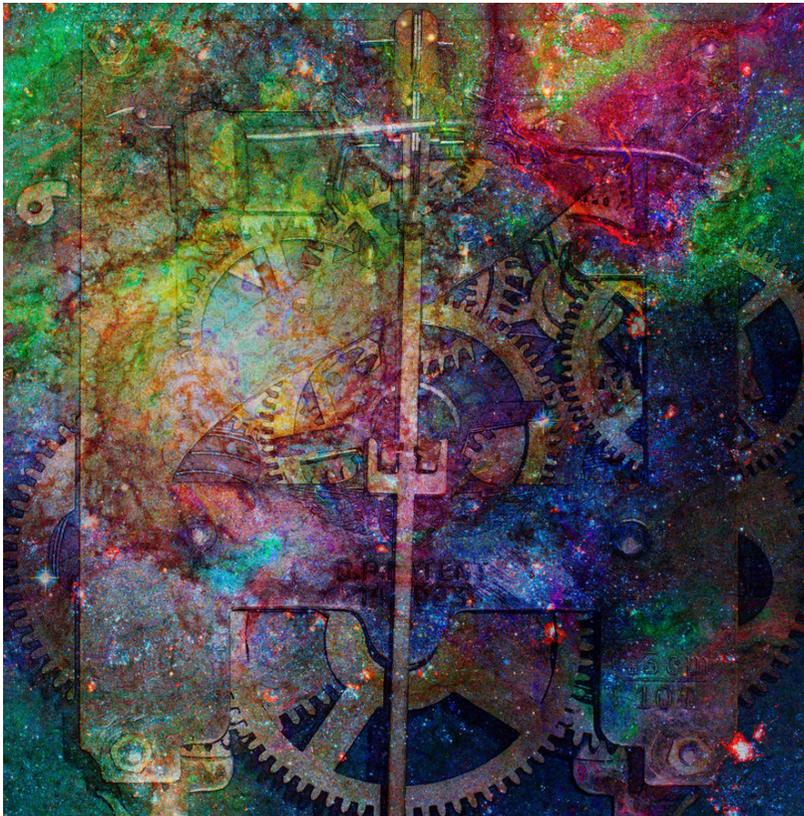
0= 25%

1= 50%

2= 25%

# ALEATORIEDADE

## Histórico das medições – a Lei dos Erros



Tentativa de descrever como as medições variam

Características gerais dos “erros aleatórios”



**Lei do erro:** a idéia de que a distribuição dos erros segue uma lei universal

# ALEATORIEDADE

## Histórico das medições – a Lei dos Erros

- Daniel Bernoulli: em 1777 notou que série de medições divergentes compartilham características comuns

MEDIÇÕES ASTRONÔMICAS X SETAS DE UM ARQUEIRO

Daniel Bernoulli



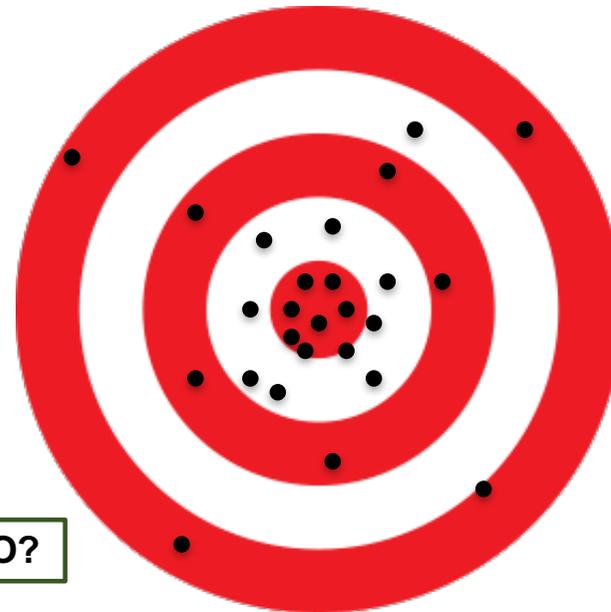
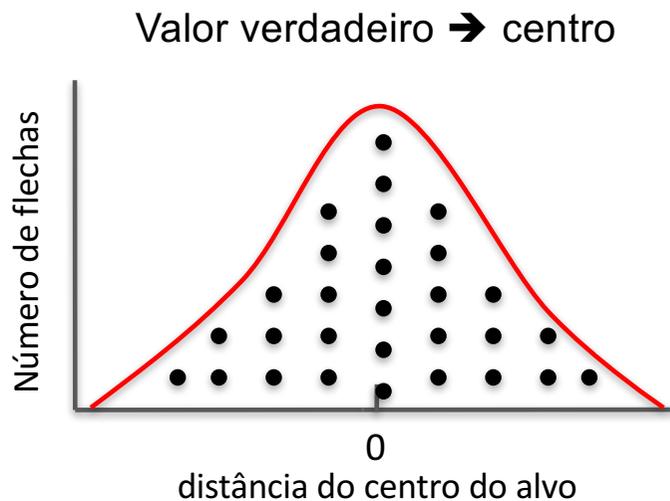
# ALEATORIEDADE

## Histórico das medições – a Lei dos Erros

- Daniel Bernoulli: em 1777 notou que série de medições divergentes compartilham características comuns

MEDIÇÕES ASTRONÔMICAS X SETAS DE UM ARQUEIRO

Daniel Bernoulli



QUAL A FORMA ESPECÍFICA DESTA DISTRIBUIÇÃO?

# VARIÁVEL ALEATÓRIA

## Variável

A propriedade que será medida por observações individuais. Ex: o comprimento ou número de indivíduos, o PH, a riqueza de espécies em parcelas

## Variáveis das aulas de algebra

$$x - 2 = 6$$

$$y = 3 + 2x$$

Variáveis das quais é possível **atribuir valores**

## Variáveis aleatórias

- Uma variável cujo valor não é conhecido antes que uma amostra seja feita;
- São formas de mapear resultados de processos aleatórios para número;
- Quantificação dos resultados de um processo aleatório.

X= soma da face superior de 2 dados

Y= média da altura de 5 alunos aleatórios de Planeco

Z= riqueza de espécies de borboletas em fragmentos de cerrado

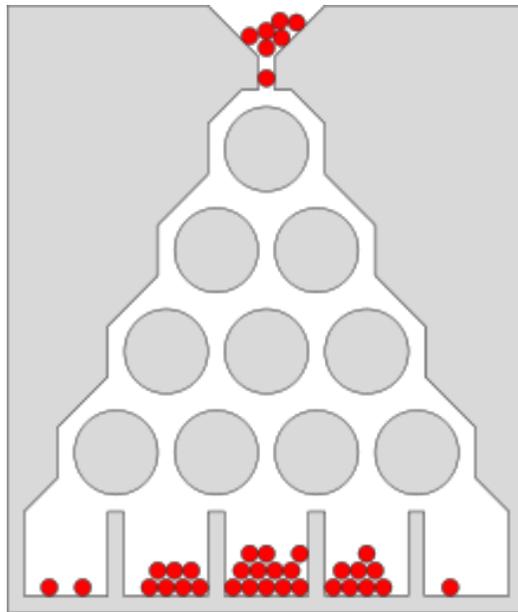
Podem assumir **muitos valores diferentes com diferentes probabilidades.**

A maior parte das variáveis em biologia são aleatórias, embora variáveis preditoras em modelos podem ser definidas como fixas antecipadamente.

# PROBABILIDADE

*Medida da freqüência esperada que um evento ocorra a longo prazo*

- Aplicada em situações que envolvem **incerteza**
- Interpretação estatística clássica (abordagem frequentista)
- Fundamental para interpretar os testes de hipótese estatísticos (P e intervalo de confiança)



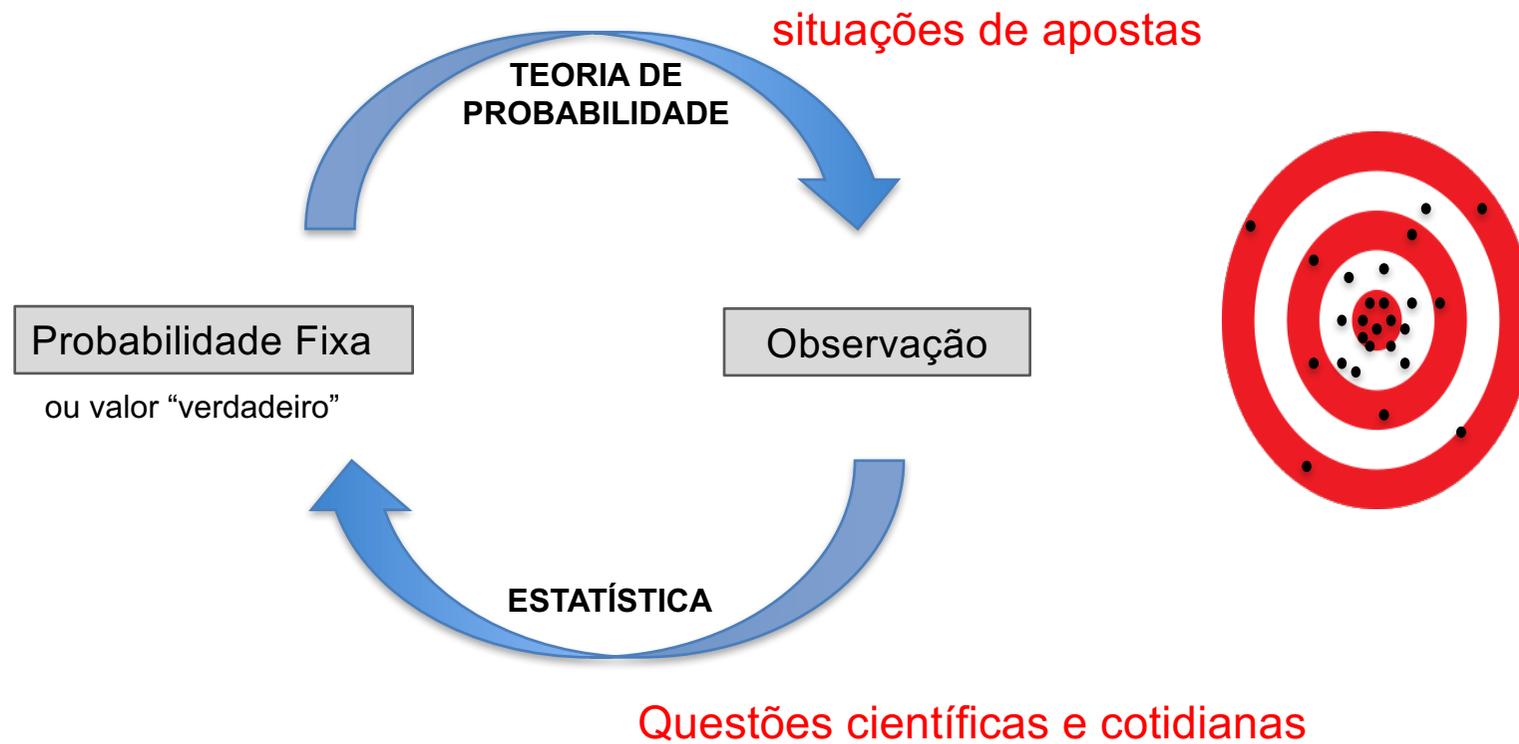
$$P = \frac{\text{número de resultados de interesse}}{\text{número de observações}}$$

- **Aplicada ao conceito de variável aleatória**

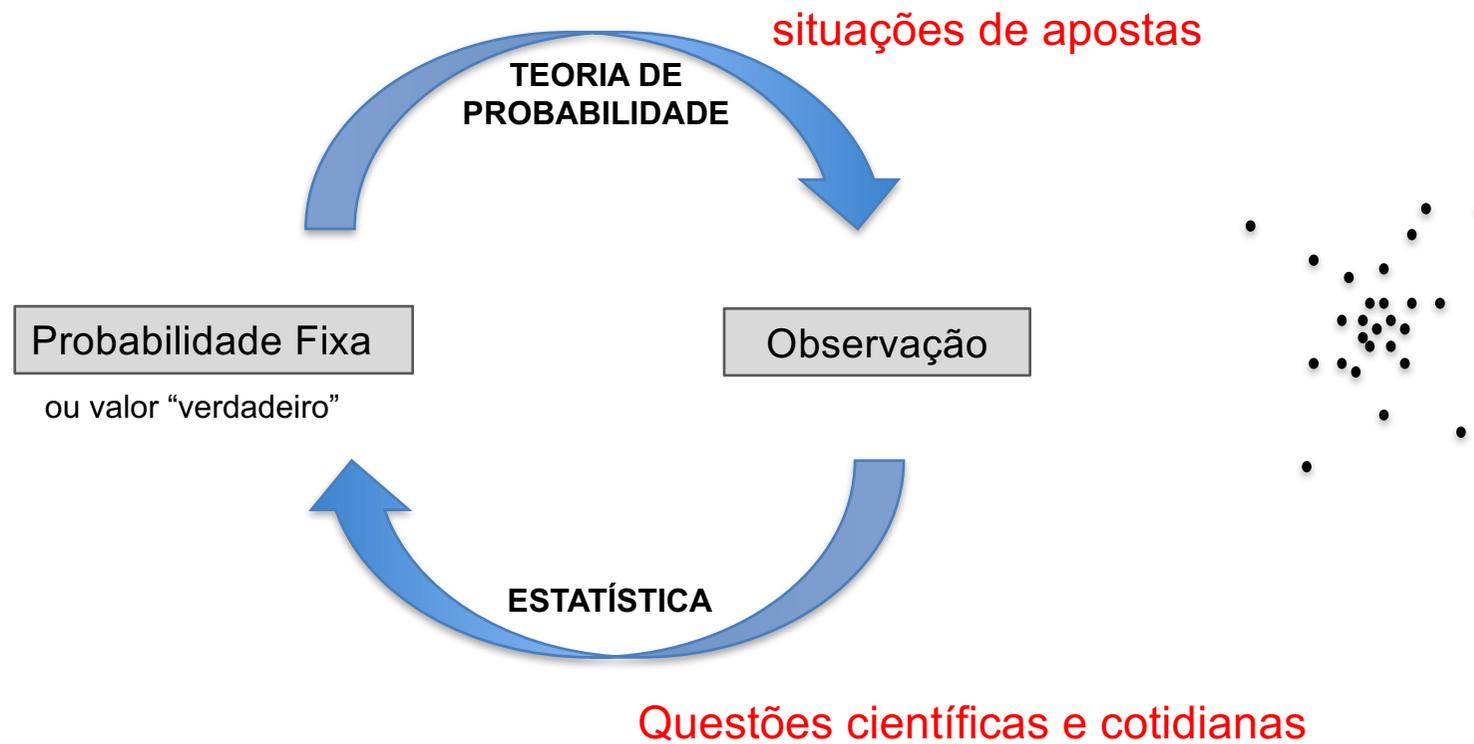
X= número de filhas entre 3 crianças

P (X=2)

# PROBABILIDADE



# PROBABILIDADE



# VARIÁVEL ALEATÓRIA

SUCESSOS	COMBINAÇÕES POSSÍVEIS
0	
1	 ou 
2	

Processo aleatório: lançamento de duas moedas honestas

Variável aleatória: número de coroas

# VARIÁVEL ALEATÓRIA

SUCESSOS	COMBINAÇÕES POSSÍVEIS	P
0		
1	 ou 	
2		

Processo aleatório: lançamento de duas moedas honestas

Variável aleatória: número de coroas

# VARIÁVEL ALEATÓRIA

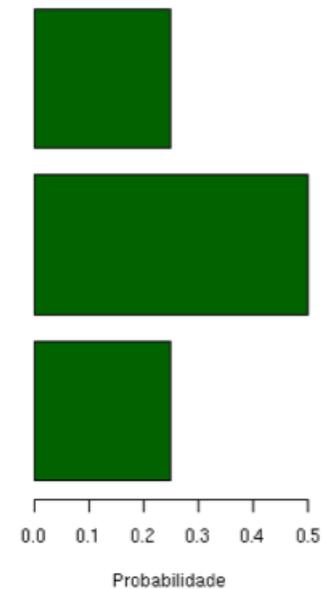
SUCESSOS	COMBINAÇÕES POSSÍVEIS	P
0		1/4
1	 ou 	2/4
2		1/4

Processo aleatório: lançamento de duas moedas honestas

Variável aleatória: número de coroas

# VARIÁVEL ALEATÓRIA

SUCESSOS	COMBINAÇÕES POSSÍVEIS	P
0		1/4
1	 ou 	2/4
2		1/4

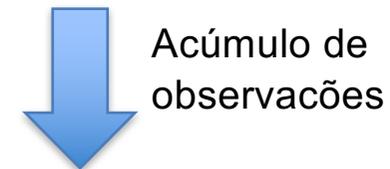
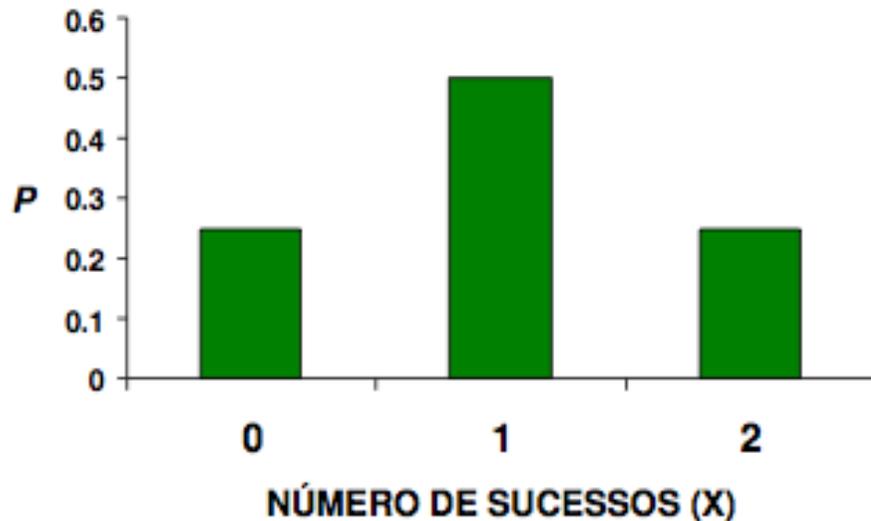


Processo aleatório: lançamento de duas moedas honestas

Variável aleatória: número de coroas

# VARIÁVEL ALEATÓRIA

O resultado de uma única observação é um evento incerto

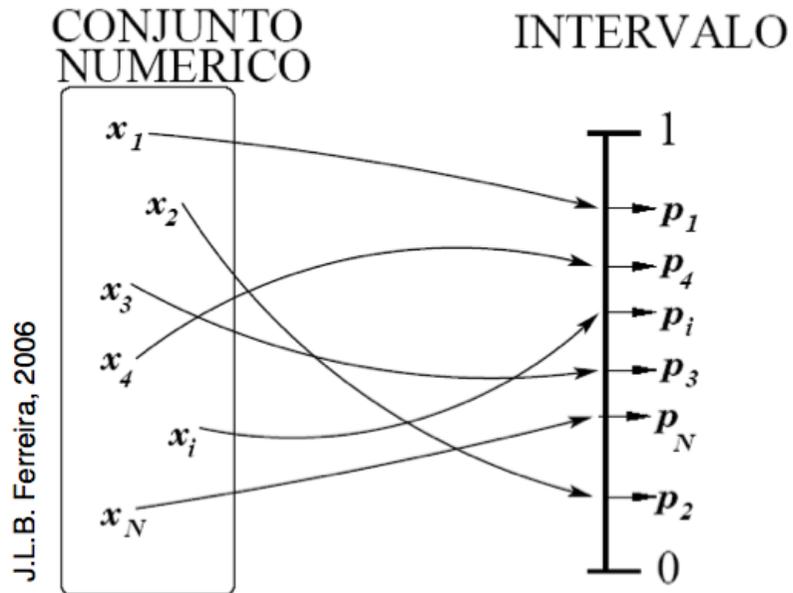


Acúmulo de observações

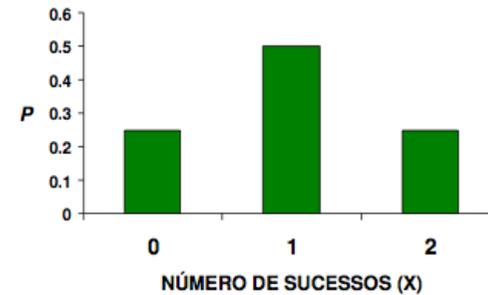
Padrões regulares na frequência na distribuição dos eventos

- **Uma variável aleatória terá uma distribuição de probabilidade associada**
- Diferentes valores da variável (espaço amostral) estão no eixo x e as probabilidades relativas desses possíveis valores no eixo y

# VARIÁVEL ALEATÓRIA



$$f(x) = P(X = x_i)$$



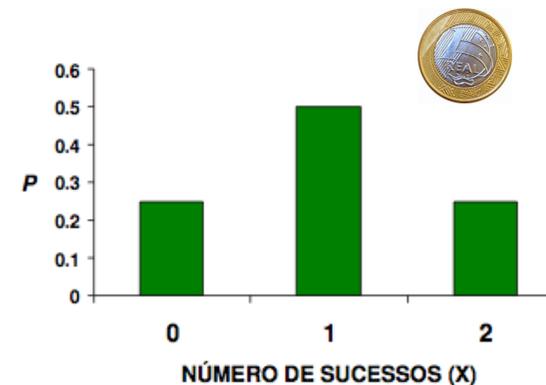
Função matemática que nos informa sobre a probabilidade de um número observado assumir certo valor



**DISTRIBUIÇÃO DE PROBABILIDADES**

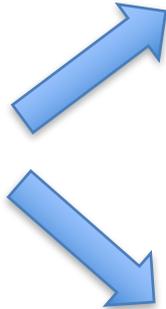
# VARIÁVEL ALEATÓRIA

- O conceito de **distribuição probabilística** **estabelece limitações à aleatoriedade** e nos dá certa capacidade de prever eventos futuros aleatórios
- Podemos falar sobre probabilidades de valores e não sobre certezas
- Os resultados de experimentos individuais são aleatórios, no sentido de que são imprevisíveis
- As distribuições de probabilidade, no entanto, nos permitem descrever a natureza matemática dessa aleatoriedade



# VARIÁVEL ALEATÓRIA

VARIÁVEIS  
ALEATÓRIAS



## Variáveis aleatórias discretas

Assumem apenas certos valores, usualmente inteiros

- presença ou ausência de uma dada espécie (0 ou 1)
- número de sobreviventes entre  $n$  indivíduos
- número de descendentes, folhas, pernas (números inteiros)

## Variáveis aleatórias contínuas

Assumem qualquer valor dentro de um intervalo

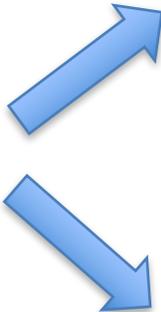
- biomassa de uma ave
- área da folha consumida por um herbívoro
- $O_2$  dissolvido numa quantidade de água

Não há limite teórico  
à precisão que  
podemos obter de  
uma variável  
contínua

# DISTRIBUIÇÕES PROBABILÍSTICAS

- A maior parte dos procedimentos estatísticos dependem de conhecermos a distribuição probabilística da variável que estamos analisando;
- Há diversas distribuições as quais são definidas matematicamente e algumas dessas descrevem adequadamente a distribuição de variáveis em biologia.

VARIÁVEIS  
ALEATÓRIAS



## Variáveis aleatórias discretas

Assumem apenas certos valores, usualmente inteiros.

**Distribuições:** Bernoulli, **Binomial**, Poisson etc

## Variáveis aleatórias contínuas

Assumem qualquer valor dentro de um intervalo.

**Distribuições:** Uniforme, **Normal**, Exponencial, Lognormal, etc

# DISTRIBUIÇÕES PROBABILÍSTICAS

## Variável aleatória de Bernoulli

Variável discreta que possui apenas dois resultados possíveis

- cara/coroa em um lançamento de moeda
- reprodução ou não de um besouro
- presença/ausência de uma espécie
- menino/menina em um nascimento
- sobreviveu/morreu em um período de tempo

$$X \sim \text{Bernoulli}(p)$$

# DISTRIBUIÇÕES PROBABILÍSTICAS - BINOMIAL

## Variável aleatória Binomial

Número de sucessos em N experimentos Bernoulli

Muitas tentativas Bernoulli → variável aleatória binomial

$$X \sim \text{Bin}(n, p)$$

n= número de tentativas

p= probabilidade de sucesso

Ex: número de caras em 10 lançamentos de moeda

*Um dos tipos mais comuns de variáveis aleatórias encontradas em estudos biológicos*

- número de besouros reprodutivos em uma população de x indivíduos
- número de sobreviventes em um aquário com x peixes
- número de meninas em uma família de 3 crianças - EXEMPLIFICAR

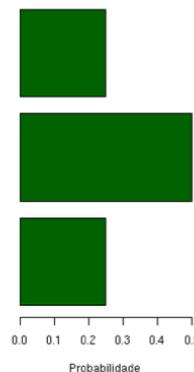
# DISTRIBUIÇÕES PROBABILÍSTICAS - BINOMIAL

A probabilidade de obter X sucessos para uma variável aleatória binomial é:

$$P(X) = \frac{n!}{X! (n-X)!} p^X (1-p)^{n-X}$$

n= número de tentativas    X= número de sucessos  
(X ≤ n)

SUCESSOS	COMBINAÇÕES POSSÍVEIS	P
0		1/4
1	 ou 	2/4
2		1/4



$$\frac{n!}{X! (n-X)!} = \binom{n}{X}$$

## Coeficiente binomial

Indica o número de vias possíveis de se obter dada combinação de sucessos e insucessos

# DISTRIBUIÇÕES PROBABILÍSTICAS - BINOMIAL

## A distribuição binomial

$$P(X) = \frac{n!}{X! (n-X)!} p^X (1-p)^{n-X}$$



Calcular a probabilidade de cada um dos possíveis resultados de uma variável Bernoulli



Distribuição binomial

Exemplo:

Variável aleatória Bernoulli com  $p=0,5$

Tentativas:  $n=25$

Sucesso: cara/filha



Moedas honestas (cara/coroa)



Nascimento (filho/filha)

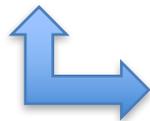


Espaço amostral

Número de caras/filhas =  $\{0, 1, 2, 3, \dots, 24, 25\}$

# DISTRIBUIÇÕES PROBABILÍSTICAS - BINOMIAL

$$P(X) = \frac{n!}{X!(n-X)!} p^X(1-p)^{n-X}$$



- cada resultado de  $X_i$  é um valor da variável aleatória binomial
- a probabilidade de cada resultado é obtido pela **função de distribuição de probabilidade**

*Função (ou regra) que fornece a probabilidade de cada resultado dentro do espaço amostral*

TABELA 2.1 Probabilidades binomiais para  $p = 0,5$  com 25 tentativas

Número de sucessos ( $X$ )	Probabilidade de ( $X$ ) em 25 tentativas ( $P(X)$ )
0	0,00000003
1	0,00000075
2	0,00000894
3	0,00006855
4	0,00037700
5	0,00158340
6	0,00527799
7	0,01432598
8	0,03223345
9	0,06088540
10	0,09741664
11	0,13284087
12	0,15498102
13	0,15498102
14	0,13284087
15	0,09741664
16	0,06088540
17	0,03223345
18	0,01432598
19	0,00527799
20	0,00158340
21	0,00037700
22	0,00006855
23	0,00000894
24	0,00000075
25	0,00000003
S = 1,00000000	

# DISTRIBUIÇÕES PROBABILÍSTICAS - BINOMIAL

Distribuição de probabilidades de uma variável aleatória binomial

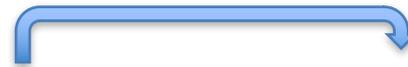
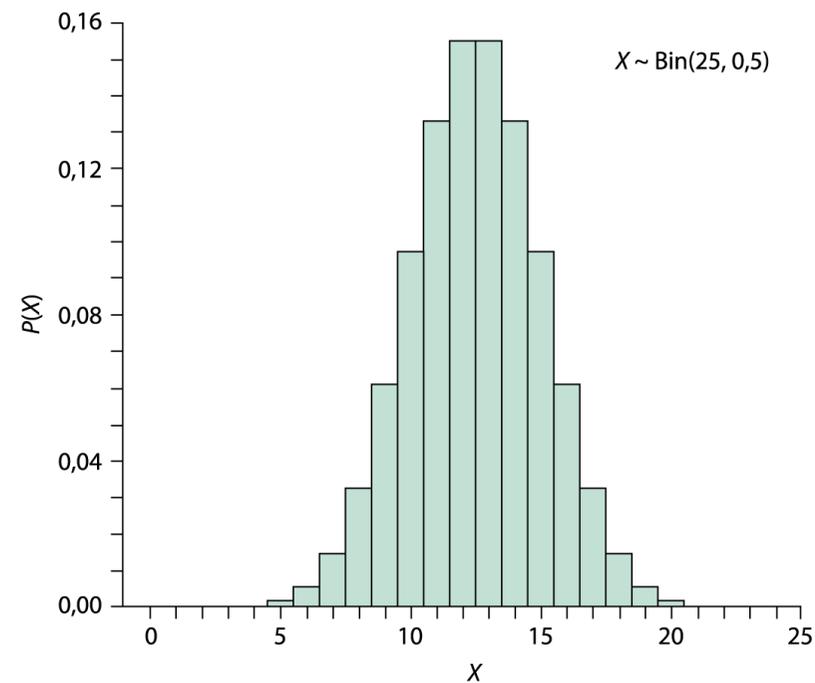


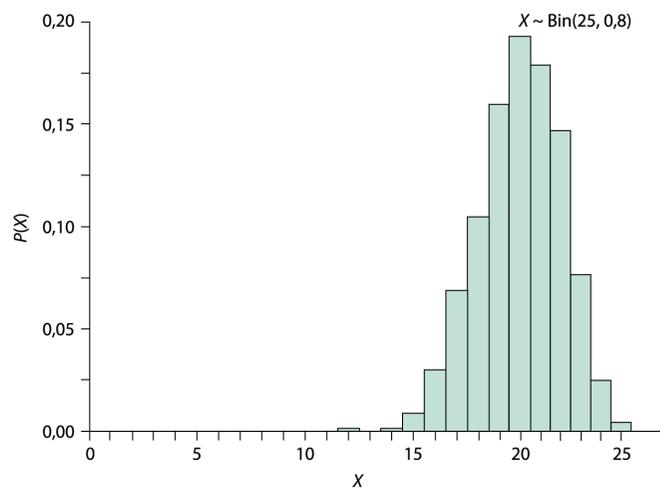
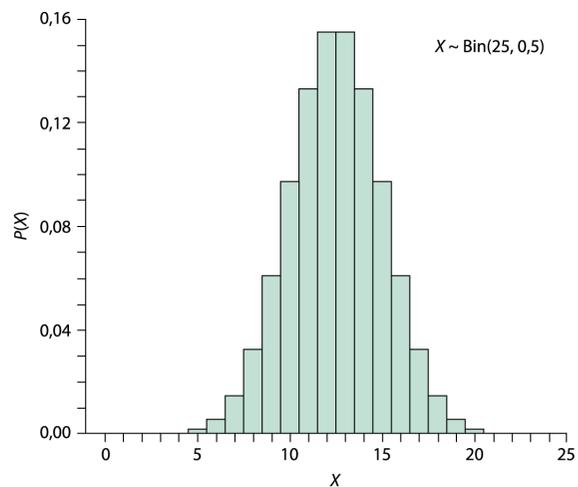
TABELA 2.1 Probabilidades binomiais para  $p = 0,5$  com 25 tentativas

Número de sucessos ( $X$ )	Probabilidade de ( $X$ ) em 25 tentativas ( $P(X)$ )
0	0,00000003
1	0,00000075
2	0,00000894
3	0,00006855
4	0,00037700
5	0,00158340
6	0,00527799
7	0,01432598
8	0,03223345
9	0,06088540
10	0,09741664
11	0,13284087
12	0,15498102
13	0,15498102
14	0,13284087
15	0,09741664
16	0,06088540
17	0,03223345
18	0,01432598
19	0,00527799
20	0,00158340
21	0,00037700
22	0,00006855
23	0,00000894
24	0,00000075
25	0,00000003

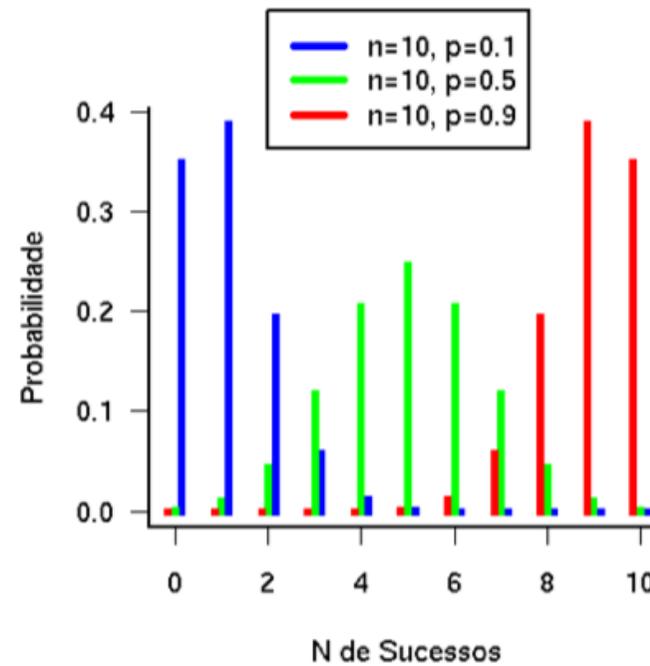
S = 1,00000000



# DISTRIBUIÇÕES PROBABILÍSTICAS - BINOMIAL



- distribuição não é necessariamente simétrica
- a forma depende do número de tentativas ( $n$ ) e da probabilidade de sucesso ( $p$ )

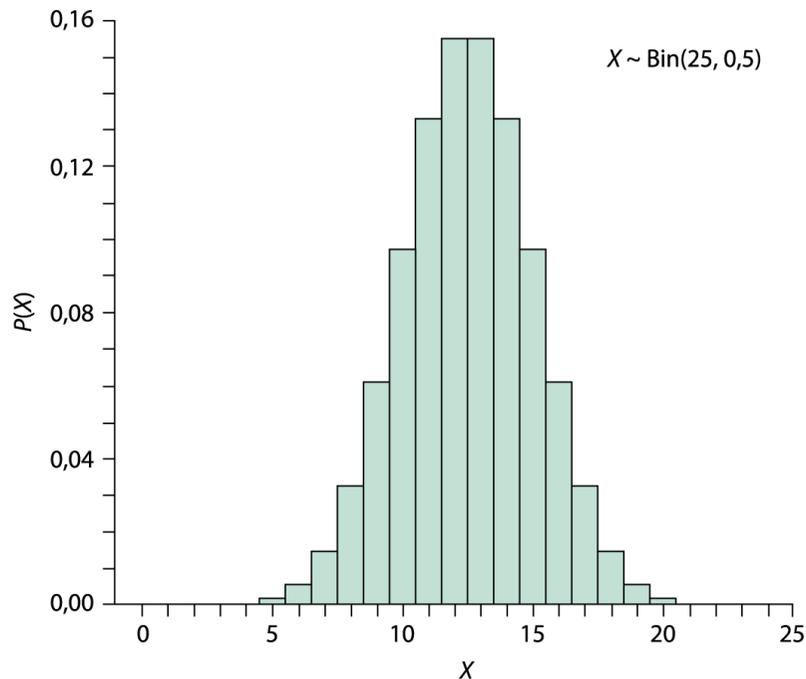


# DISTRIBUIÇÕES PROBABILÍSTICAS - NORMAL

VARIÁVEIS ALEATÓRIAS  
DISCRETAS



É possível definir o espaço amostral total como o **conjunto de resultados discretos possíveis**



VARIÁVEIS  
CONTÍNUAS



Não podemos identificar TODOS os resultados possíveis, pois existem milhares e é difícil obter a probabilidade de um valor em particular

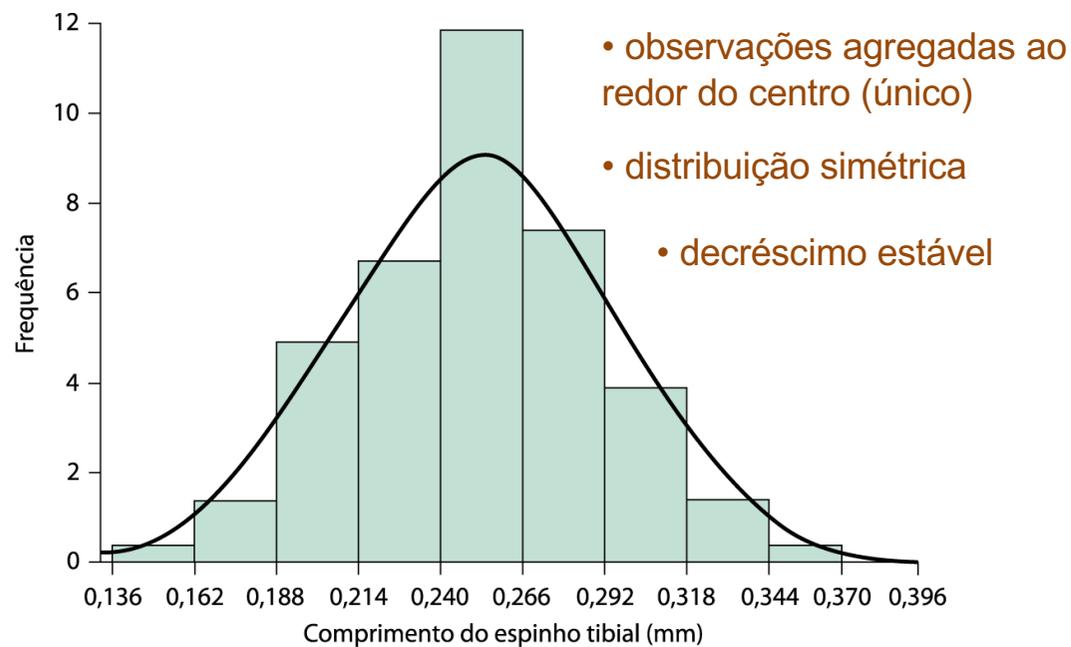
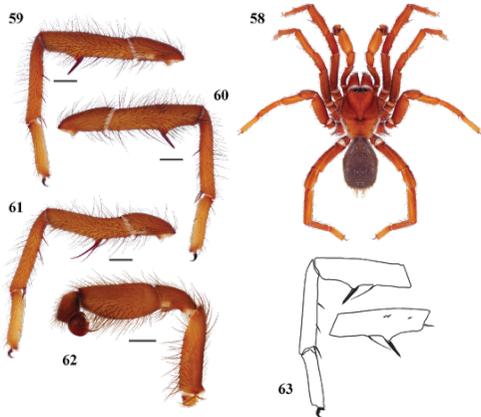


Usar intervalos de valores

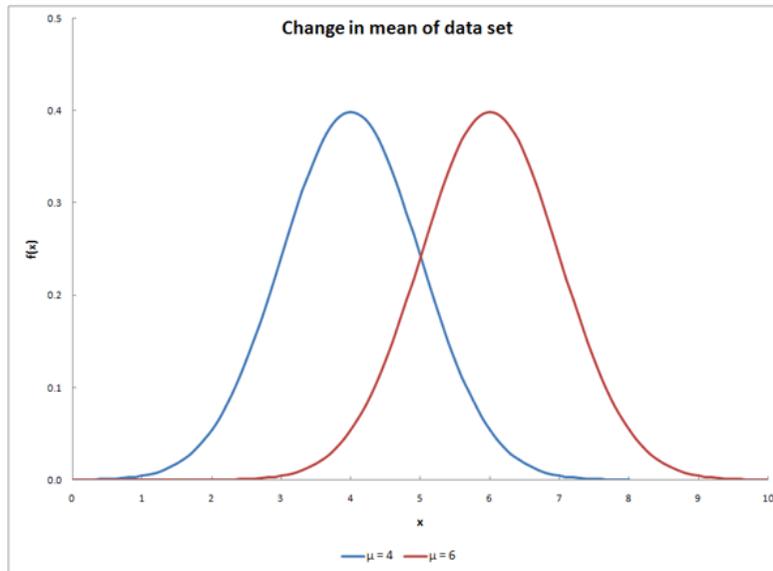
# DISTRIBUIÇÕES PROBABILÍSTICAS - NORMAL

- distribuição de probabilidades mais familiar → curva de sino ou curva de “Gauss”
- provê as bases teóricas para muitos dos testes comumente utilizados biologia: ANOVA e regressão linear
- muitos dados biológicos se ajustam à distribuição normal

Aracnólogo mediu 50 espinhos



# DISTRIBUIÇÕES PROBABILÍSTICAS - NORMAL

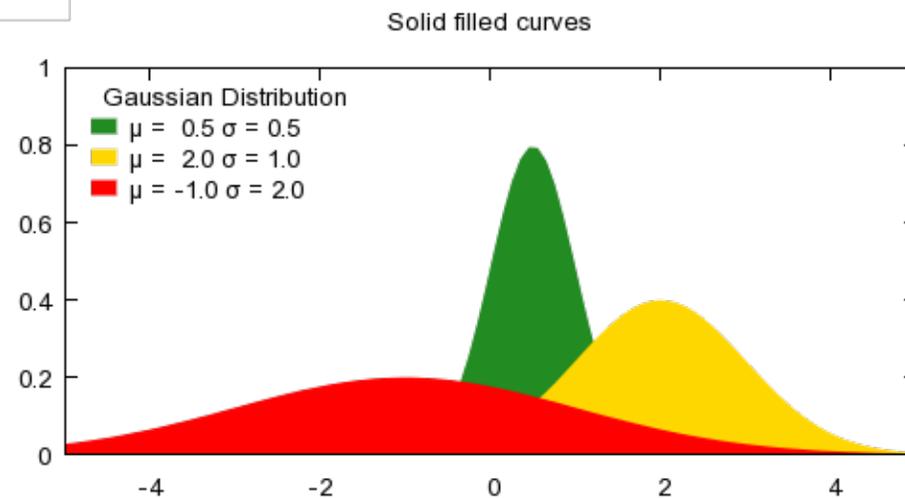


$$X \sim N(\mu, \sigma)$$

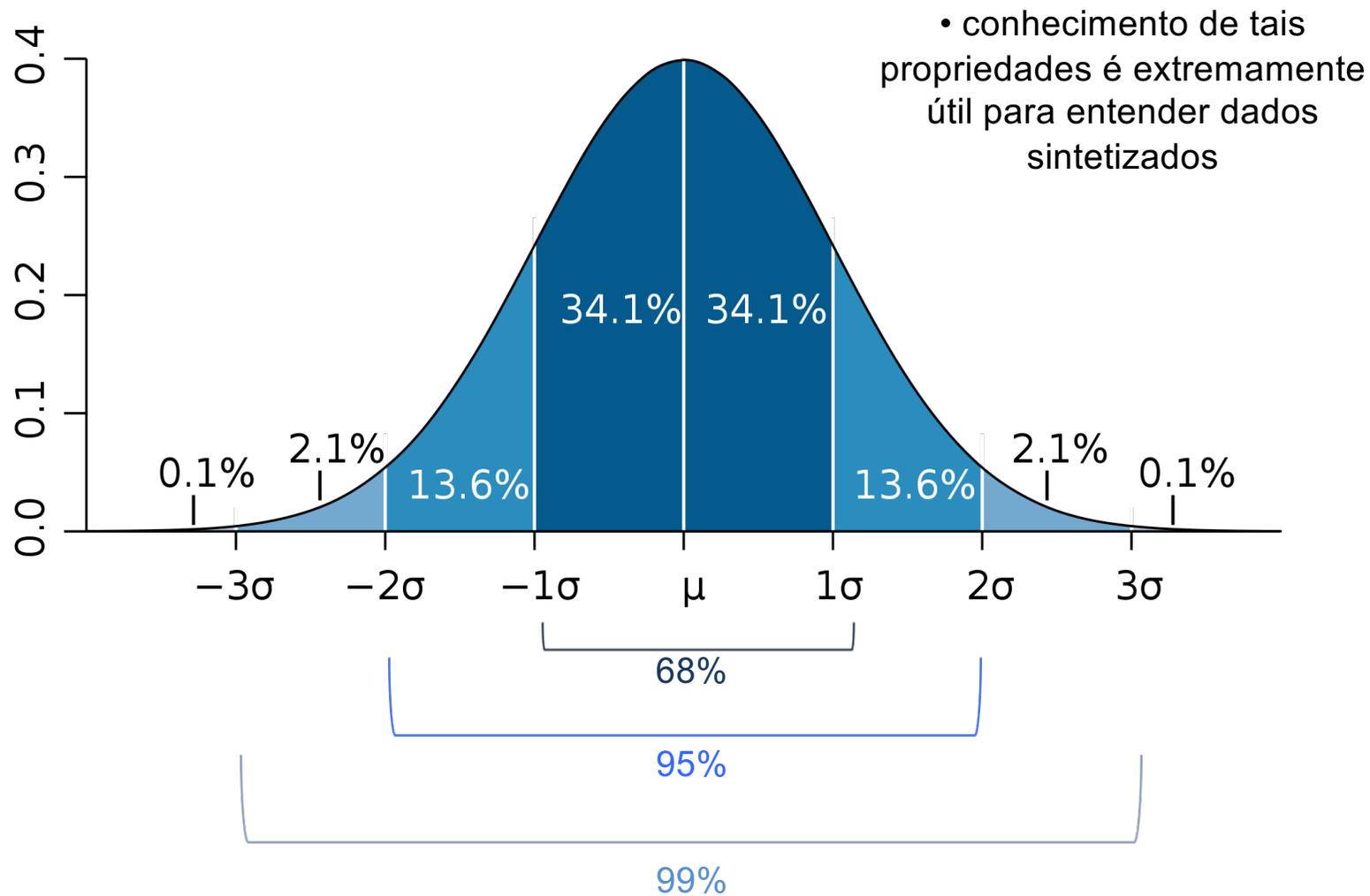
$\mu$  (mi) = média

$\sigma$  (sigma) = desvio padrão

- um único pico central ( $\mu$ )
- simétrica em relação à média
- área sob a curva soma 1 ou 100%



# DISTRIBUIÇÕES PROBABILÍSTICAS - NORMAL



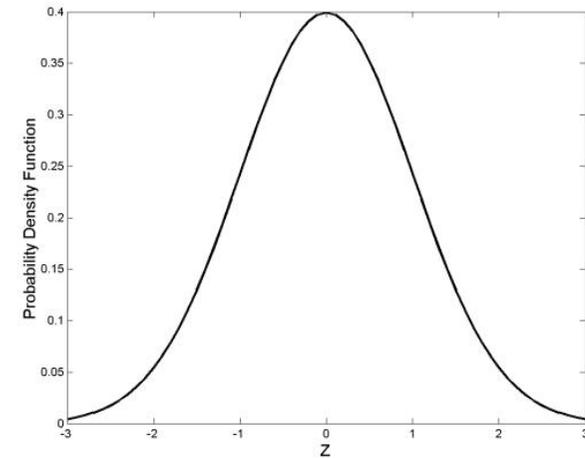
# DISTRIBUIÇÕES PROBABILÍSTICAS - NORMAL

- Muitas distribuições normais diferentes podem ser criadas com diferentes valores de  $\mu$  e  $\sigma$

Distribuição normal padronizada

$$\mu = 0 \text{ e } \sigma = 1$$

Para evitar confusão a variável padronizada é denominada  $z$  (e  $x$  para representar variáveis do mundo real)



## RESUMO DA AULA

PAPEL DA  
ESTATÍSTICA

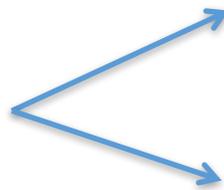


Ferramenta usada para  
tomada de decisões  
diante de **incerteza**

Variação não previsível



FONTES DE VARIAÇÃO

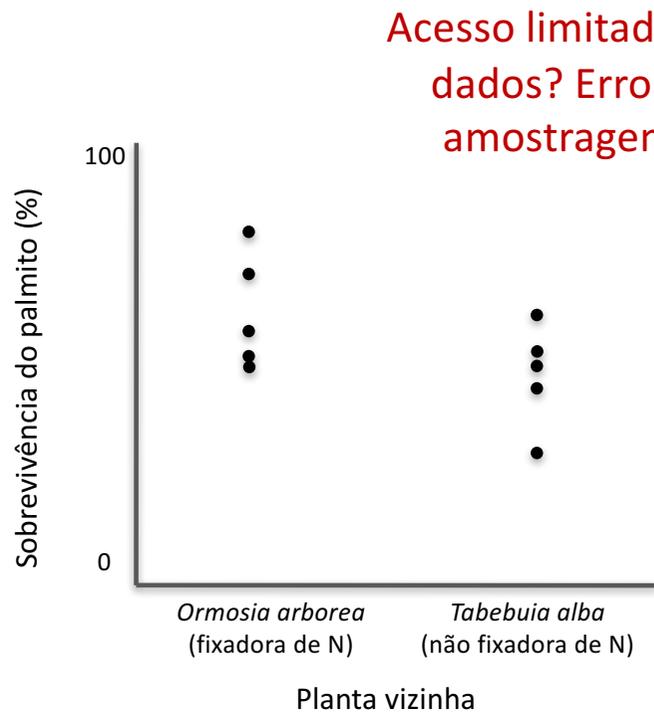
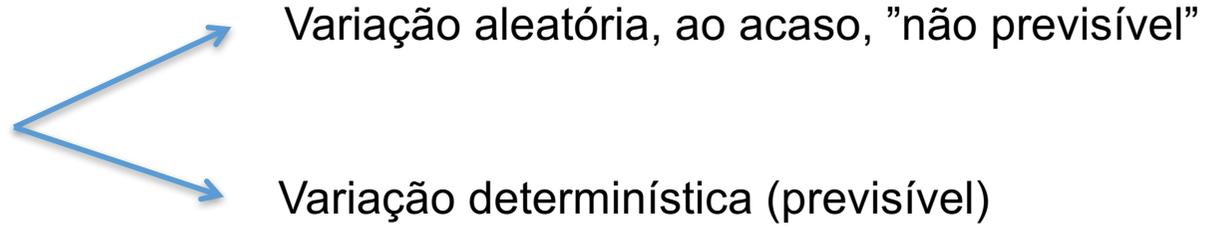


Variação aleatória, ao acaso, "não previsível"

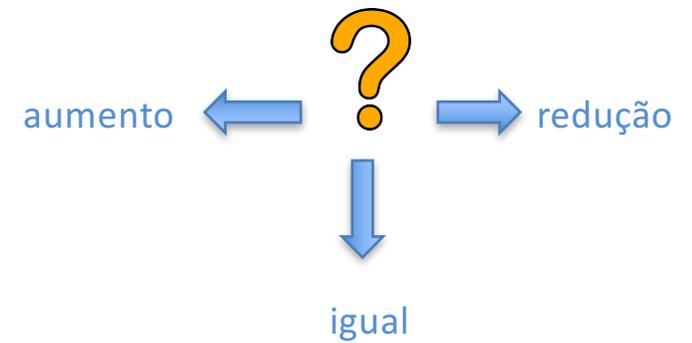
Variação determinística (previsível)

**ERRO DE  
AMOSTRAGEM –  
acesso limitado aos  
dados**

FONTES DE VARIAÇÃO



POSSÍVEIS RESULTADOS



# RESUMO DA AULA

PAPEL DA  
ESTATÍSTICA

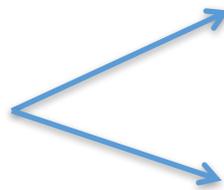


Ferramenta usada para  
tomada de decisões  
diante de **incerteza**

Variação não previsível



FONTES DE VARIAÇÃO



Variação aleatória, ao acaso, "não previsível"

**ERRO DE  
AMOSTRAGEM**

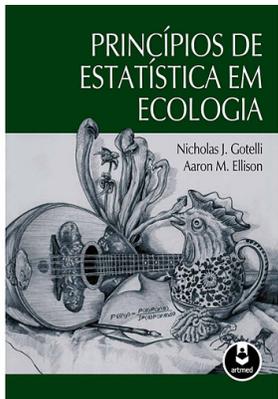
Variação determinística (previsível)

Reconhecer a variação aleatória é fundamental  
para identificarmos se existe variação  
determinística

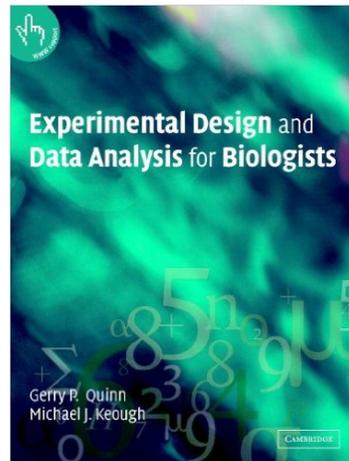


Importância de entendermos o  
conceito de variável aleatória e  
distribuição probabilística

## REFERÊNCIAS E LEITURAS



Gotelli NJ, Ellison AM. 2011. **Princípios de Estatística em Ecologia**. Artmed



Quinn GP, Keough MJ. 2002. **Experimental Design and Data Analysis for Biologists**. Cambridge University Press.

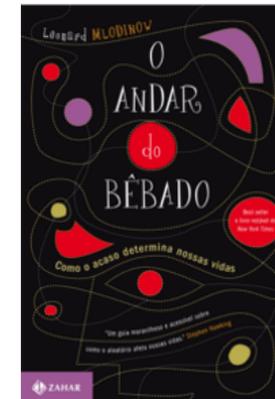


**Vídeo-aulas sobre variáveis aleatórias**

<https://www.khanacademy.org/math/statistics-probability/random-variables-stats-library>



Salsburg, D. 2009. **Uma senhora toma chá – como a estatística revolucionou a ciência do século XX**. Editora Zahar



Leonard Mlodinow. 2009. **O andar do bêbado – como o acaso determina nossas vidas**. Editora Zahar