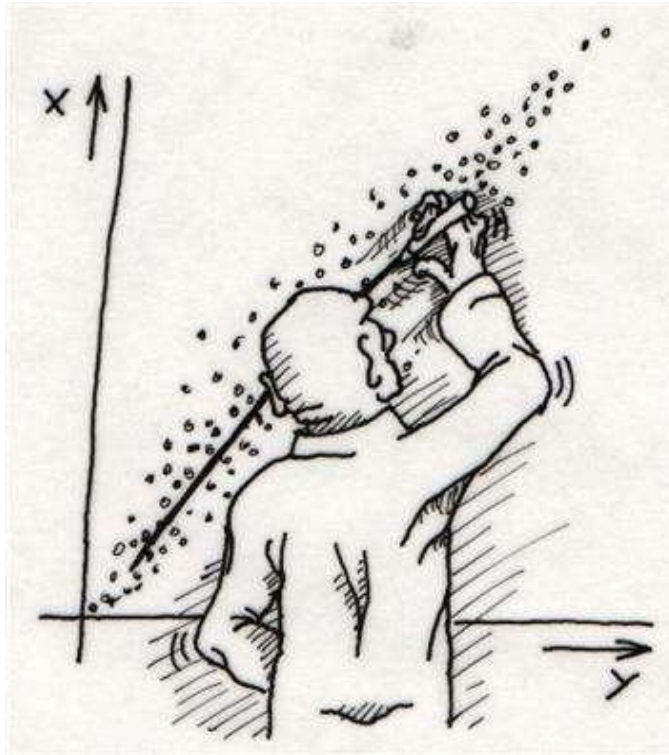


1. Introdução à modelagem estatística aplicada à ecologia populacional



Leonardo L. Wedekin

Curso Ecologia Populacional – USP 2011

Resumo

1. Motivações

Aplicadas

Teóricas

2. Método científico

Revisão sobre inferência

Uso de modelos

3. O modelo linear

Modelos fatoriais

Extensões: GLM e GAM

Implementação no R

Agradecimentos:



USP – Pós-graduação em Ecologia, em especial
aos doutores:

Roberto Munguia-Steyer

Alexandre Adalardo de Oliveira

Glauco Machado

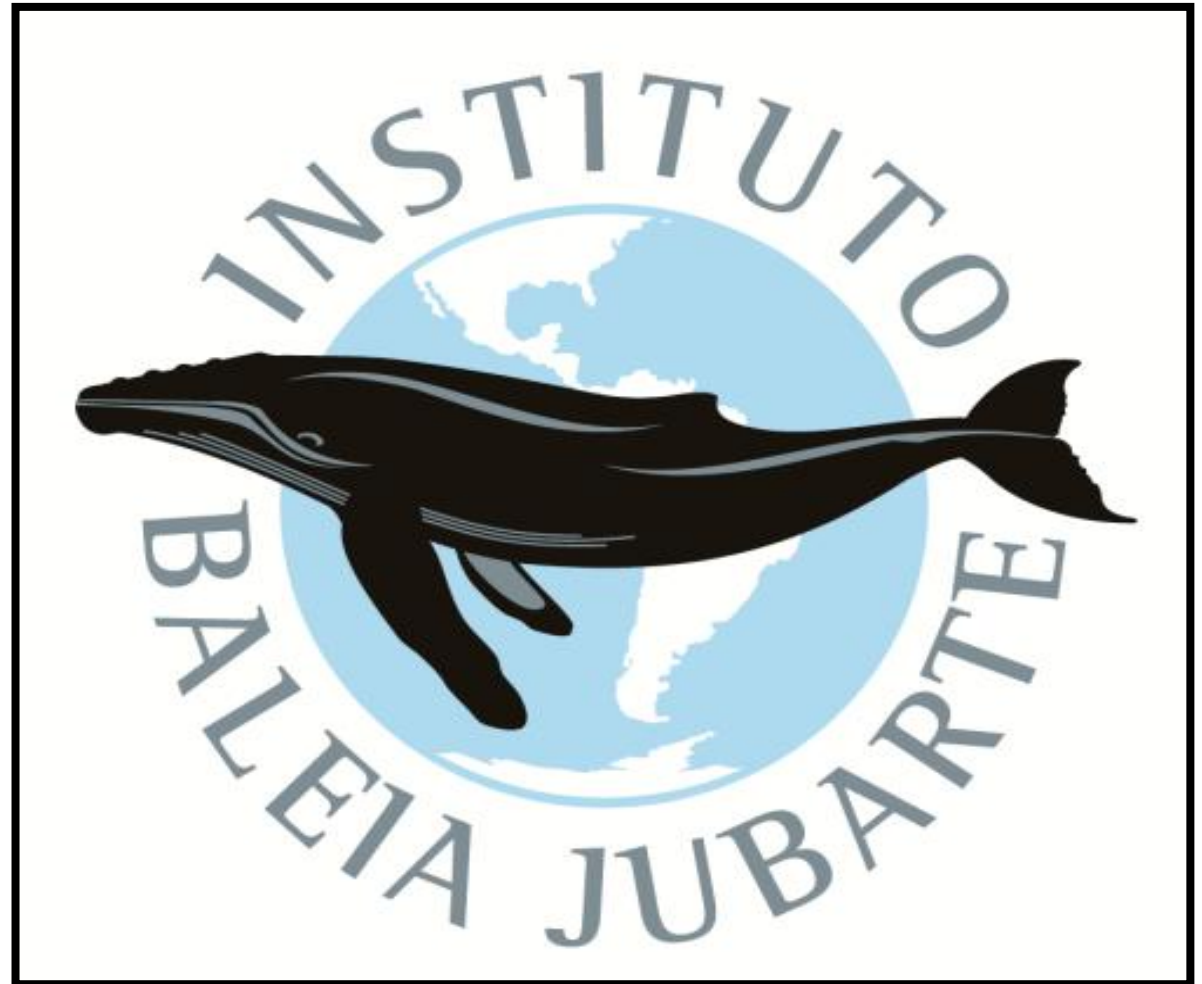
Paulo Inácio de Prado

Agradecimentos:

Missão:
**conservação de
cetáceos e seus
habitats**

OSCIP

Fronteira / ciência
aplicada



1. Motivações: ciência aplicada



Conservação da biodiversidade:

Monitoramentos populacionais

Riqueza de espécies


Extinção recente do Golfinho do Rio Yangtze

Lipotes vexillifer †

Guo (2006)



Status de conservação




The IUCN Red List of Threatened Species™ 2008

[Login](#) | [Feedback](#) | [FAQ](#) | [Terms of use](#)

[::About](#) [::Initiatives](#) [::News](#) [::Photos](#) [::Partners](#) [::Sponsors](#) [::Technical Documents](#)

Enter Red List search term(s)



[Home](#) » [Search](#) » Search Results

Explore or refine your search below:

- Keywords
- Taxonomy
- Location
- Systems
- Habitats
- Threats
- Assessment
- History

Results 1 to 1 of 1

Sotalia fluviatilis (Guianian River Dolphin)
Status: Data Deficient [ver 3.1](#)
Pop. trend: unknown

Current search:

[Save / Export Search](#)

Search terms

Show taxa:

Species

Keyword search:

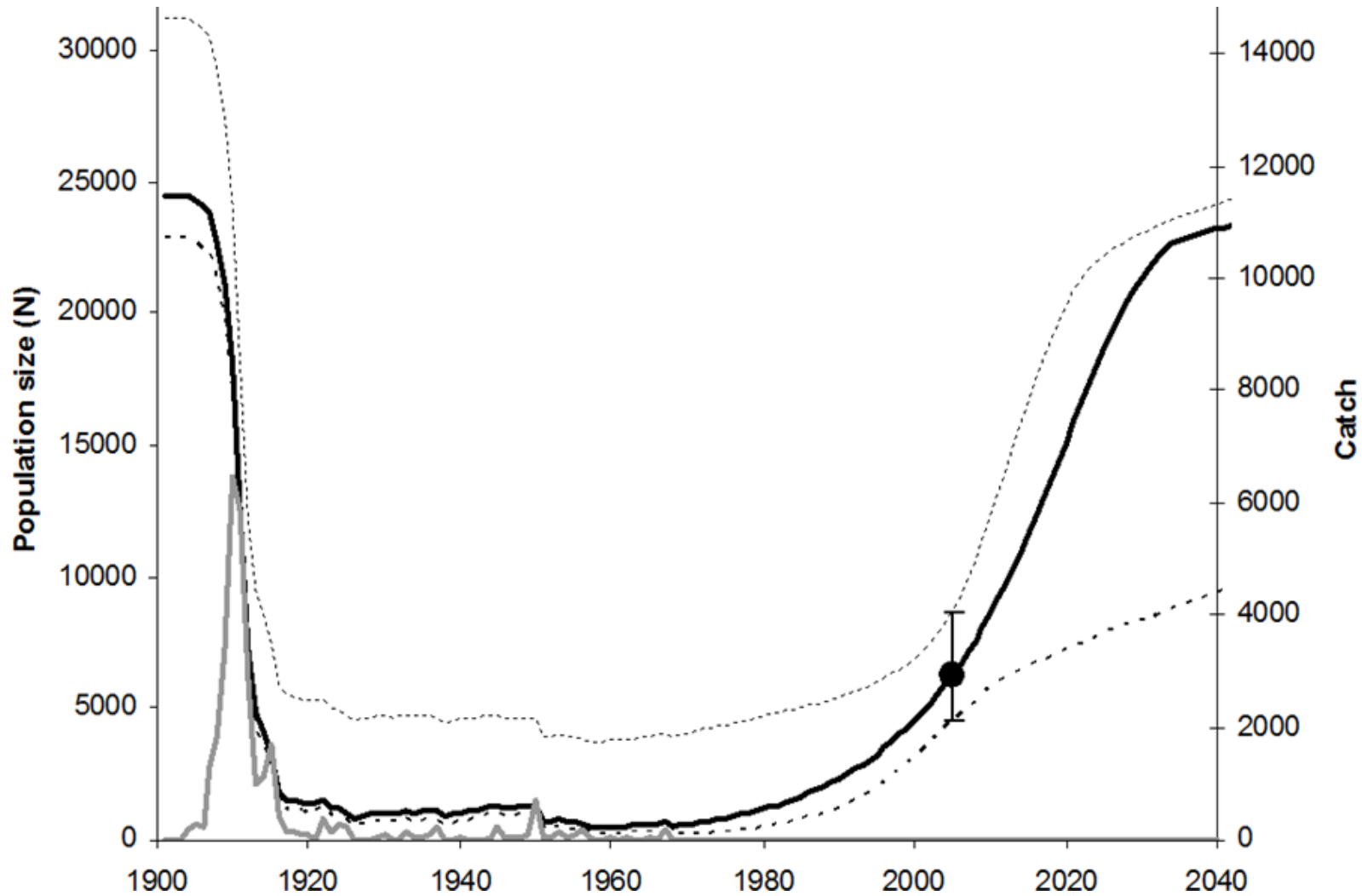
"Sotalia", Exact phrase,
The entire database

Citation: IUCN 2008. 2008 IUCN Red List of Threatened Species. <www.iucnredlist.org>. Downloaded on 13 April 2009.

Disclaimer: To make use of this information, please check the Copyright and Data Disclaimer.

Feedback: If you see any errors or have any questions or suggestions on what is shown on this page, please fill in the feedback form so that we can correct or extend the information provided

Trajetória da população



(Zerbini *et al.*, no prelo)

Risco de extinção

Version 9.50



VORTEX

**A Stochastic Simulation
of the Extinction Process**

Mudanças climáticas



Espécies dependentes do gelo

- **Ecosistema do Mar de Gelo no Oceano Austral**
- **Ártico**

Ecologia populacional



“... é dedicada ao estudo de indivíduos de uma mesma espécie, como eles compõem as populações em que vivem, e como estas populações mudam ao longo do tempo. Na ecologia populacional nós buscamos informações sobre parâmetros demográficos chave, tipicamente relacionados com a mortalidade, movimentos e produtividade... Nós estamos interessados também em tentar relacionar parâmetros demográficos com influências internas, como o aquecimento global, mudanças no habitat.” (King *et al.*, 2010)

Parâmetros demográficos

Sobrevivência (S ou ϕ)

Taxa de crescimento (λ)

Recrutamento (f)

Abundância (N)

Densidade (D)

Etc...

1. Motivações: teóricas



Níveis de organização biológica:

meta-comunidades

comunidades

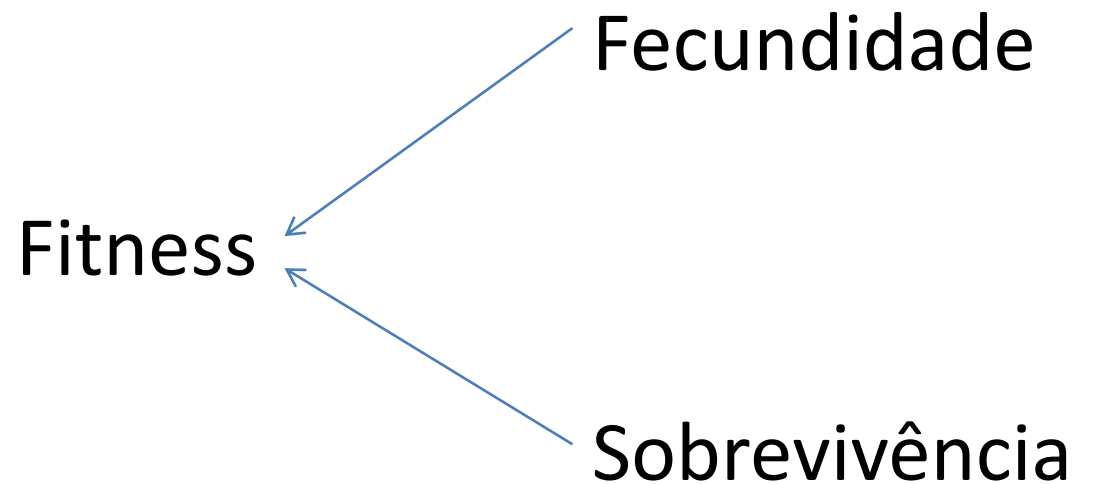
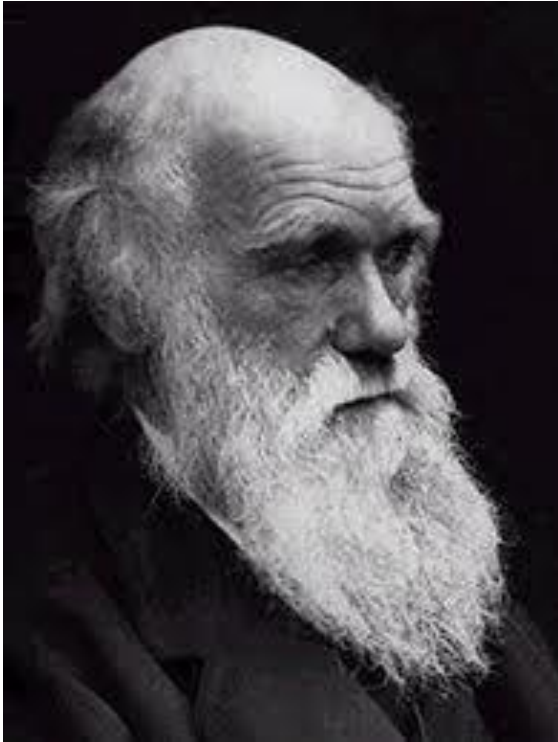
meta-populações

populações

Seleção natural



Charles Darwin (1802-1882)



2. Método científico



Como observar a população, coletar dados empíricos?

- **Censo:** contagem completa

Mas raramente se aplica...



Levantamentos



contagem incompleta (amostragem)



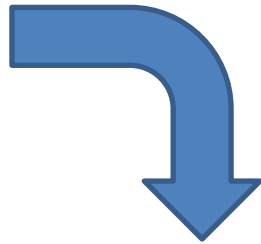
estimativa por inferência



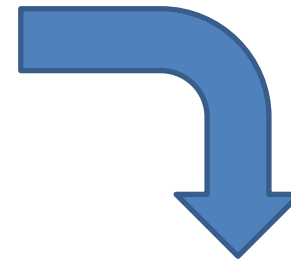
Estimativa de um parâmetro



n amostras



fórmula (estimador)

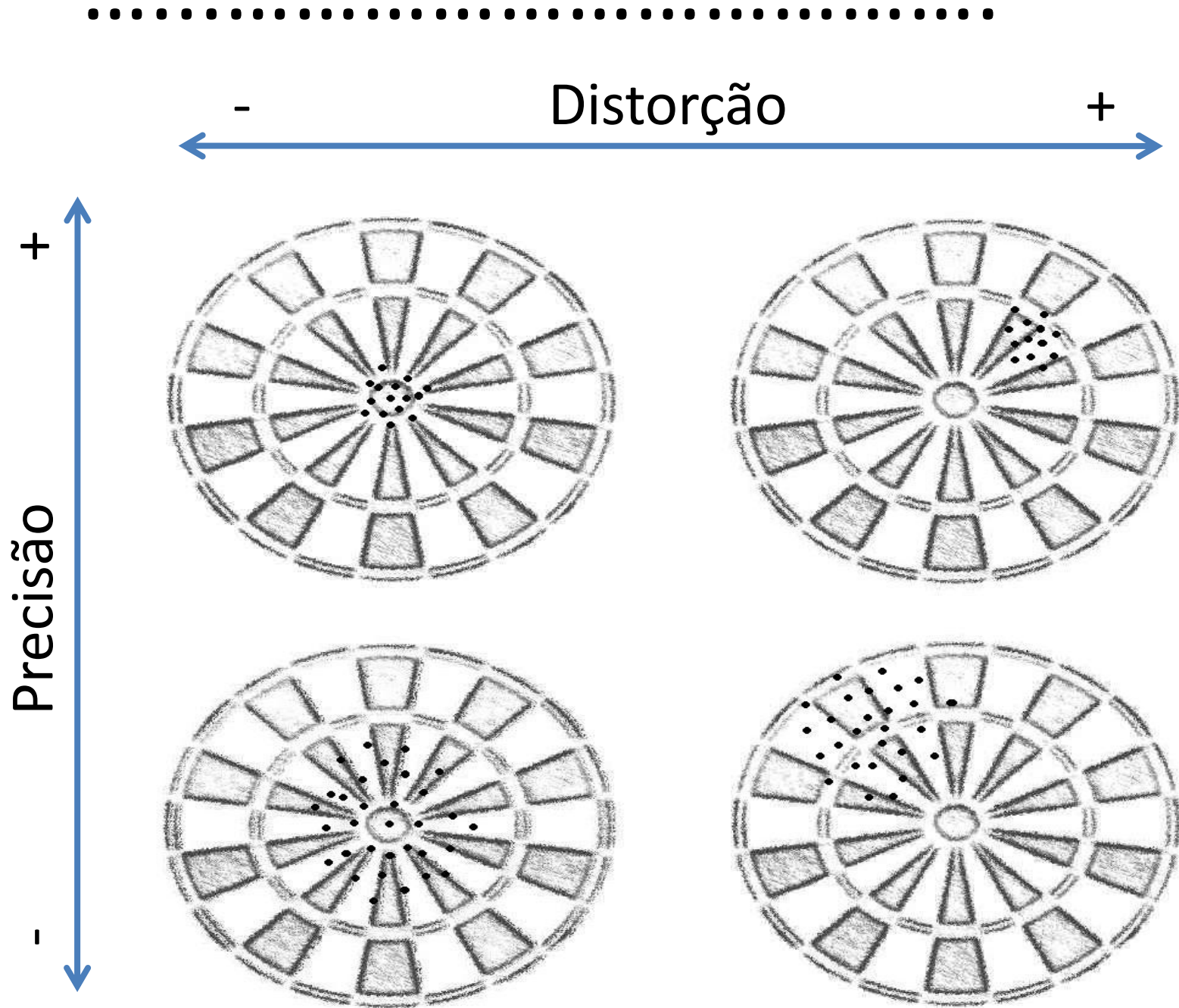


Estimativa pontual...

\hat{N}

...de uma verdade desconhecida.

Precisão e distorção



Definições

.....

- ***Precisão***: segundo o método científico, é o grau na qual repetidas mensurações de um fenômeno sobre condições similares mostra os mesmos resultados.
 - Pode ser definida como uma medida de erro ao redor da média de uma estimativa.
- ***Distorção / acurácia***: grau de proximidade da mensuração de uma quantidade em relação ao verdadeiro valor desta quantidade.

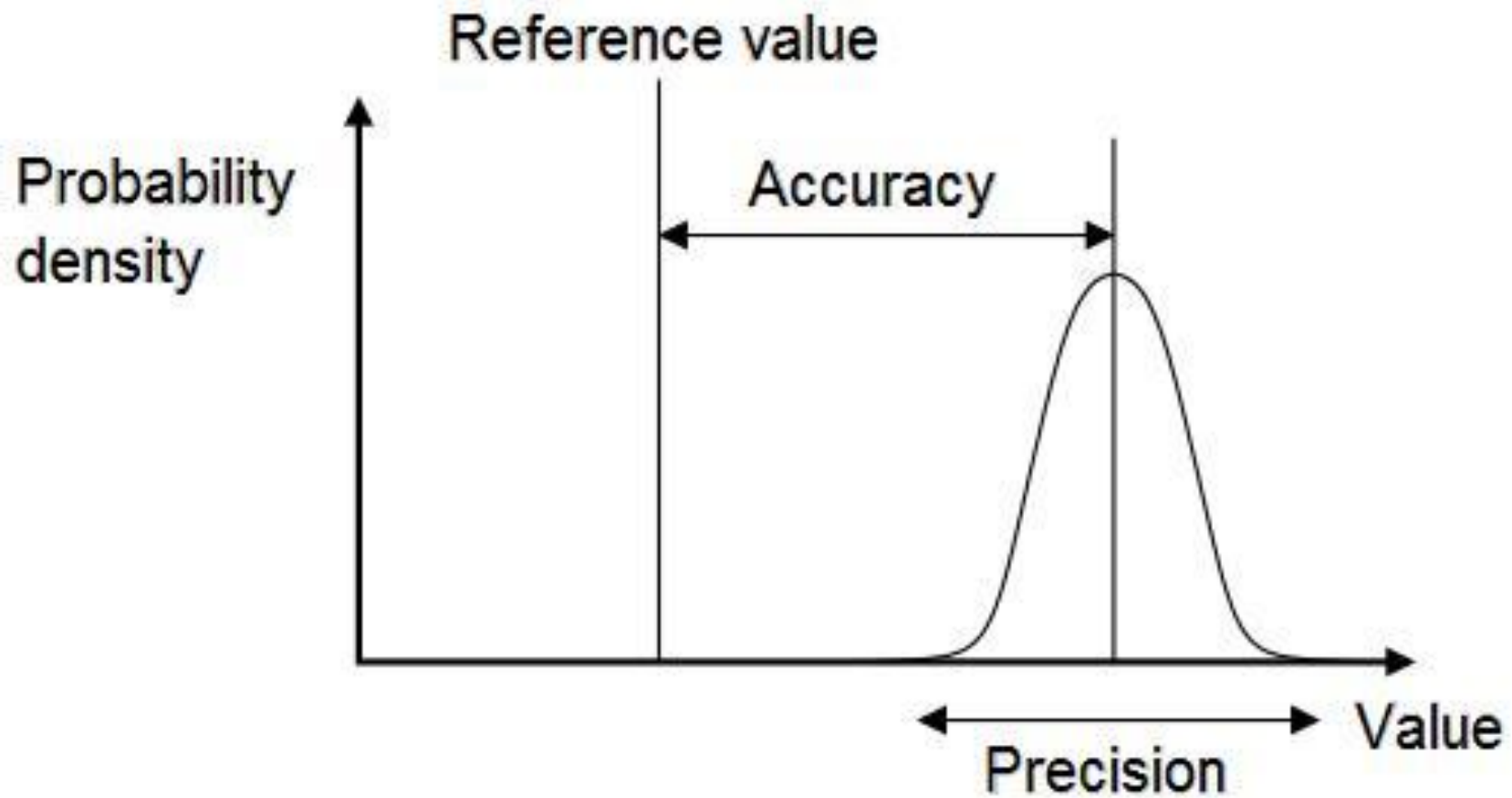
Medidas de precisão



- **Variância:** medida de dispersão em torno de um valor esperado (média ou estimativa)
- **Erro padrão:** desvio padrão estimado
- **Intervalo de confiança:** limite inferior e superior com 95% de probabilidade de incluir o valor real da população;
- **Coefficiente de variação:** erro padrão / estimativa pontual. Medida relativa, comparável entre estudos.

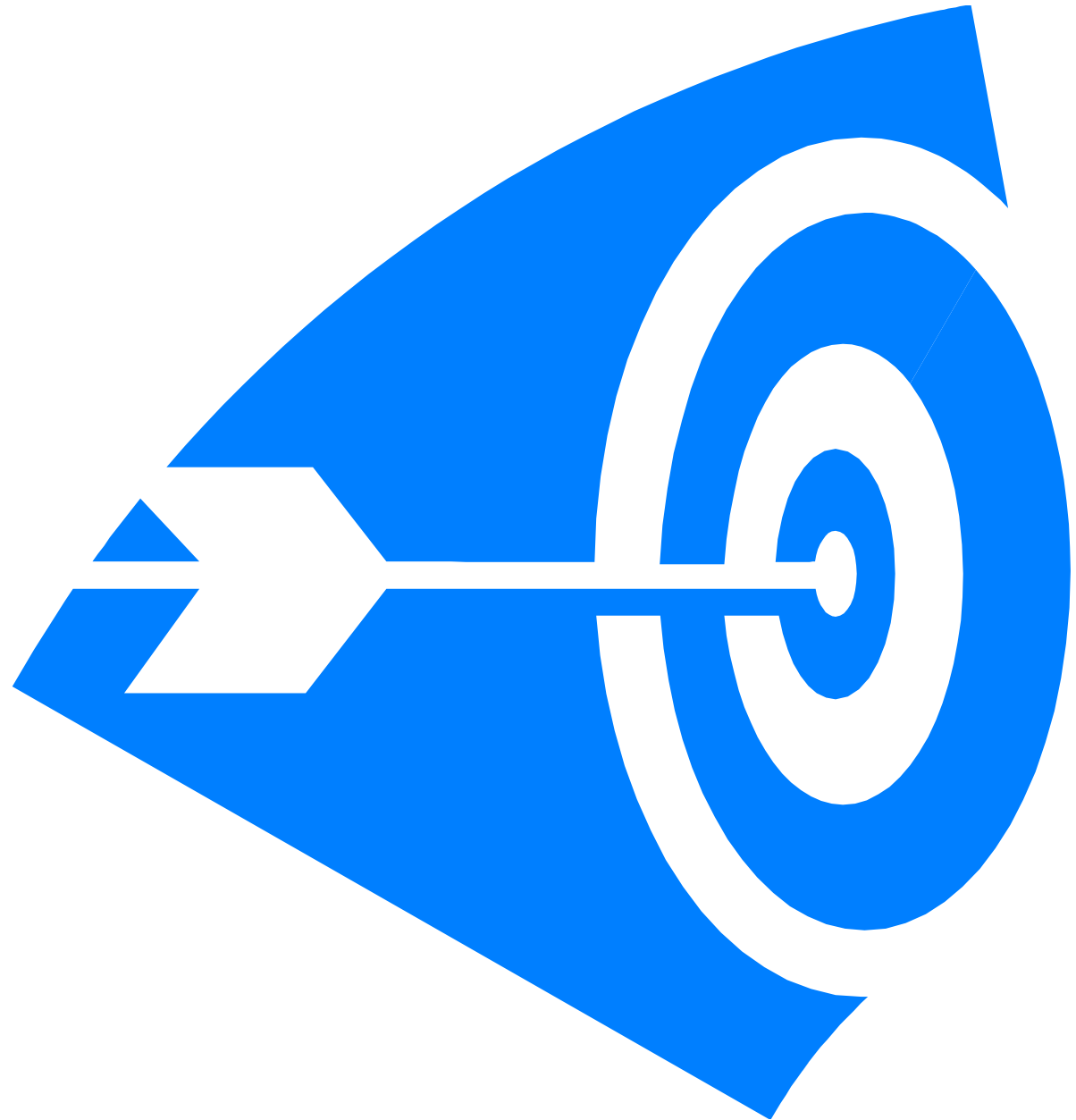
Outra forma de enxergar

.....



Precisão aceitável

- Coeficiente de variação abaixo de 20%;
- CV abaixo de 10% considerado muito bom.

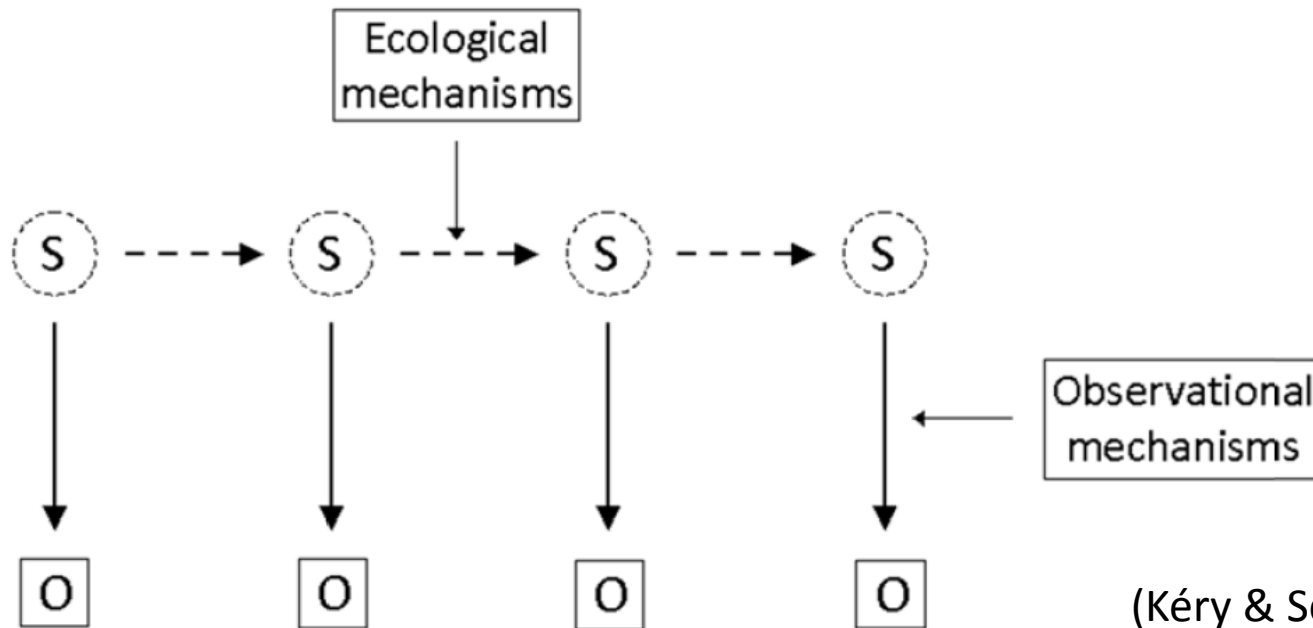


Fontes de variação



- **Variações naturais temporais e espaciais:** *process* ou *population variation*
- **Variações amostrais:** observador, coleta de dados

Separando a observação do processo



(Kéry & Schaub, no prelo)

Exemplos:

- Em marcação-recaptura separamos a probabilidade de captura dos outros parâmetros;
- Em *Distance sampling* estimamos primeiro a probabilidade de detecção para depois estimar os outros parâmetros;
- Em modelos hierárquicos (Royle & Dorazio, 2008) ou *state-space* os dois componentes são explicitamente formulados.

Detecção imperfeita ou distorção por detecção



Observadores não detectam,
animais se escondem,
mudam de comportamento,
condições ambientais se alteram...

Detectabilidade



Espécies/indivíduos podem estar presentes em
uma área mas não serem detectados
(problema da falsa-ausência)

Detectabilidade



Falsos negativos

Dupla contagem

Falsos positivos

Erros de classificação

Modelagem Hierárquica

Royle & Dorazio (2008)



Dois componentes são descritos:

(a) modelo do parâmetro **ecológico** de interesse que descreve sua variação (temporal, espacial,...)

(b) modelo do processo de **observação** que contém uma descrição probabilística do mecanismo que produz o dado observável

Modelos na ecologia populacional

Objetivos de um modelo incluem:

- *Previsão*

- *Explicação*

- *Generalização*

Importância dos modelos



“Toda inferência na ecologia é feita através de algum modelo. Para alguma observação fazer sentido, todo mundo precisa de um modelo, independente se o observador sabe disto ou não”

(Kéry, 2010)

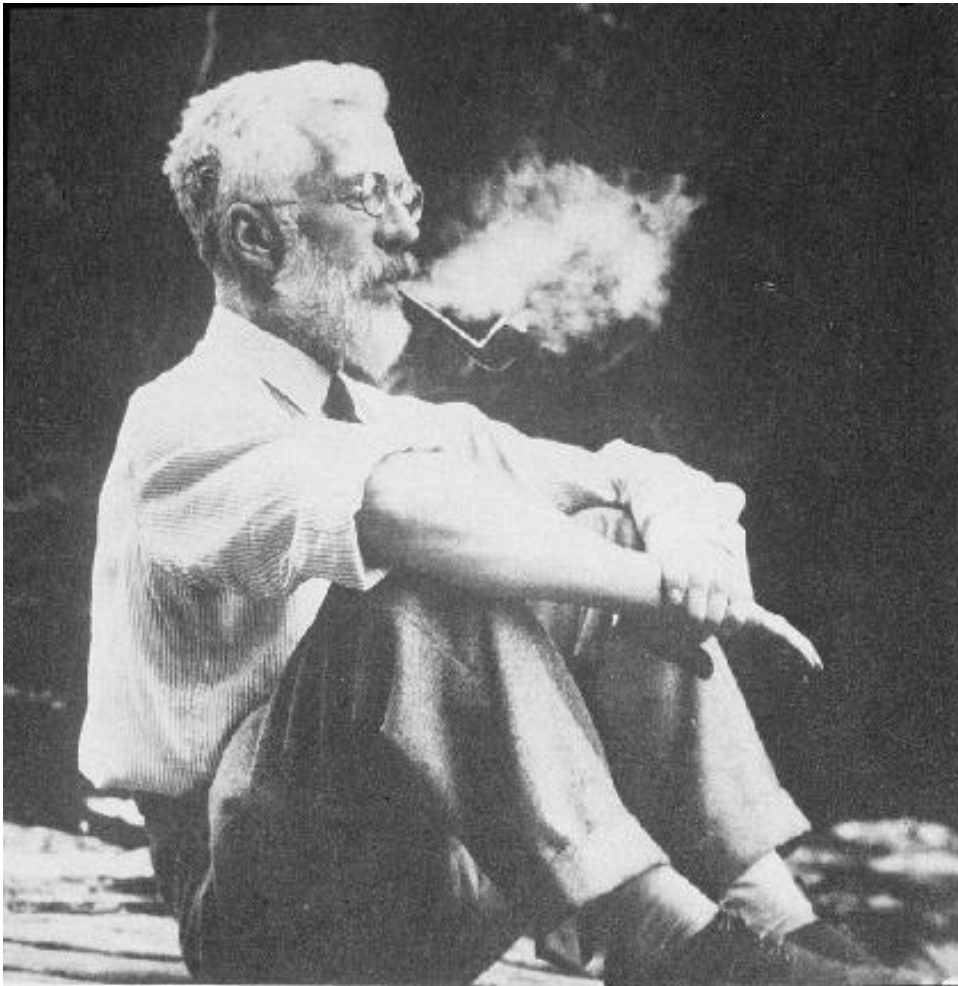
O quê são modelos?



“... o objetivo dos métodos estatísticos é a **redução dos dados**. Uma quantidade de dados, que geralmente pelo seu grande tamanho é impossível de entrar na mente, é para ser substituído por relativamente poucas quantidades que devem representar adequadamente o todo, ou que, em outras palavras, deve conter o máximo possível, ou idealmente toda a informação relevante contida nos dados originais”

(Fisher, 1922)

Ronald A. Fisher (1890-1962)



A natureza é complexa...



...e entendê-la exige simplificá-la.

PORTANTO...



Modelos são aproximações, abstrações, generalizações ou reduções da realidade

Refletem aspectos chave de sistemas ecológicos de interesse

“Todos os modelos são errados e alguns são úteis”

George Box

Tipos de análise

(Gotelli & Ellison, 2004)



Baseada no **desenho** (*design-based*)

Baseada no **modelo** (*model-based*)

CRÍTICA AO TESTE DE HIPÓTESES

(Anderson *et al.*, 2000)



- É válido usá-los, eles funcionam! Ufaa...
- Mas têm pouca utilidade e carregam pouca informação.
- Portanto, diminuem a velocidade do progresso da ciência.

É fato!



**EXISTEM DIFERENTES FORMAS DE VER E
ANALISAR OS DADOS**

PERGUNTA NECESSÁRIA



Qual modelo deve ser usado para melhor aproximar a realidade, baseado em dados de boa qualidade e relevantes para a questão?

(Burnham & Anderson, 2001)

“A modelagem na ciência
permanece, no mínimo
parcialmente, uma arte.”

(McCullagh & Nelder, 1989)

TRÊS PRINCÍPIOS GERAIS GUIAM

(Burnham & Anderson, 2001)



- PARCIMÔNIA
- MÚLTIPLAS HIPÓTESES DE TRABALHO
- FORÇA DA EVIDÊNCIA

Princípio da parcimônia



A navalha de Occam (*Occam's razor*), elimina tudo que é desnecessário...

Quanto mais simples, melhor...

Mas não tão simples...



Um modelo deve explicar um fenômeno ecológico de interesse de forma correta e precisa.

MODELAGEM



Ao adicionarmos parâmetros/variáveis aos modelos, o modelo pode se ajustar perfeitamente aos dados, e se aproximar da realidade observada.

EM COMPENSAÇÃO,

Mais parâmetros significam **mais ajuste e menos distorção**, mas **menor precisão**.

Múltiplas hipóteses de trabalho



Ao invés de uma única hipótese de trabalho (como no teste de hipóteses), trabalha-se com múltiplas hipóteses ou modelos biologicamente plausíveis

Múltiplas hipóteses de trabalho



Faz algum sentido trabalhar com dezenas ou centenas de hipóteses e modelos alternativos?

Múltiplas hipóteses de trabalho



Esta abordagem permite que hipóteses novas ou mais elaboradas sejam adicionadas continuamente nos modelos, enquanto hipóteses sem fundamento sejam gradualmente abandonadas.

Força de evidência



Não existe dicotomia, mas diferentes forças de evidência para cada hipótese / modelo

Componentes da variação



Variância da estimação/modelos

Variância da observação

Variância espacial ou do processo

3. O modelo linear e suas extensões

$$y = \alpha + \beta x$$

Modelo linear clássico

$$Y = \alpha + \beta X + \varepsilon$$

Onde:

α = intercepto / onde a linha cruza o eixo X

β = coeficiente regressão / inclinação (+ ou -)

Y = variável resposta/dependente

X = variável explicatória/independente

ε = erro residual

Premissas



- Relação linear entre X e Y
- Normalidade
- Variância constante
- Independência

Regressão múltipla

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + \varepsilon$$

Onde:

β_0 = intercepto / onde a linha cruza o eixo X

β_i = coeficiente regressão / inclinação (+ ou -)

Y = variável resposta/dependente

X_i = variável explicatória/independente

ε = erro residual

Modelo linear generalizado (GLM)

$$E(Y) = f(\beta_0 + \beta_1 X)$$

Onde:

E = distribuição assumida da variável resposta (estrutura do erro)

$f(z)$ = função de ligação (*link function*)

$(\beta_0 + \beta_1 X)$ = componente linear

> Permite outras distribuições de erro relaxando premissas do modelo linear clássico e conferindo flexibilidade ao modelo

Funções de ligação (*link function*)

- Variável resposta assume outras distribuições da família exponencial, como por exemplo:
 - Números positivos e inteiros (Poisson)
 - Resposta binária 0 ou 1 (Binomial)

Error	Canonical link
normal	<i>identity</i>
poisson	<i>log</i>
binomial	<i>logit</i>
Gamma	<i>reciprocal</i>

Modelo linear / ANOVA

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

Onde: Y = variável resposta contínua

β_0 = intercepto

β_1 = inclinação da reta

x = variável código (*dummy*) = 0 ou 1

ε = erro residual

Modelos fatoriais

Notação em matriz

$$y = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \cdot \\ \cdot \\ \cdot \\ Y_{1k} \\ Y_{21} \\ Y_{22} \\ \cdot \\ \cdot \\ \cdot \\ Y_{2k} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{1k} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{2k} \end{bmatrix} = \mathbf{X}\beta + \varepsilon$$

Modelo aditivo generalizado (GAM)

- Extensão não paramétrica dos GLM;
- O que determina o formato da função são os dados e não funções ou distribuições específicas definidas *a priori*;
- Vários métodos de “suavização” (*smooth*) e grau de suavização (graus de liberdade).

??? QUIZ ???



Quanto mais parâmetros em um modelo:

Aumenta ou diminui a precisão?

Aumenta ou diminui a distorção?

Um super-ajuste do modelo (mais parâmetros)

causa qual tipo de erro (de forma análoga ao

teste de hipóteses)? Tipo I ou II?

Implementação no R



1. Ler e incorporar tabela ao espaço de trabalho
2. Ajustar modelo de regressão linear aos dados
3. Interpretar parâmetros do modelo (betas)
4. Checar erros residuais e outras premissas
5. Ajustar outros modelos (GLM Poisson, GAM)

Exercício pra casa



Repita a análise demonstrada usando o script (“**Roteiro_pratica_1.txt**”) e os dados do tamanho da população brasileira entre 1950 a 2010 (“**Pop.brasil.txt**”)

Leituras recomendadas!



Gotelli & Ellison (2004): um *primer* sobre **estatística ecológica** baseada no desenho

Burnham & Anderson (2002) sobre filosofia científica e método de **seleção de modelos e inferência**

McCullagh & Nelder (1989) sobre **modelos generalizados lineares**

Leituras recomendadas!



Faraway (2006) sobre **modelos lineares e extensões**

Sutherland (2006) para **métodos de estudo** aplicados a diferentes grupos taxonômicos