

AULA 2:

Método de máxima verossimilhança

Um Exemplo



Uma espécie tem probabilidade de ocorrência de 50% por sítio.

Para uma região com 12 sítios, qual a probabilidade da espécie ocorrer em todos os sítios?

E em nenhum deles?

Dados:

$p = 50\%$ por sítio

$N = 12$ sítios

Qual a probabilidade da espécie ocorrer em

$x = 12$ sítios?

Dados:

$p = 50\%$ por sítio

$N = 12$ sítios

Qual a probabilidade da espécie ocorrer em $x = 12$ sítios?



0,0002441406

Ó, mortal!

(Com precisão de 10 casas decimais)

Função de probabilidade

$$f(x | N, p)$$

N, p são **parâmetros**

x é a **variável**

Função de probabilidade

$$f(x | N, p)$$

N, p são **parâmetros**

x é a **variável**

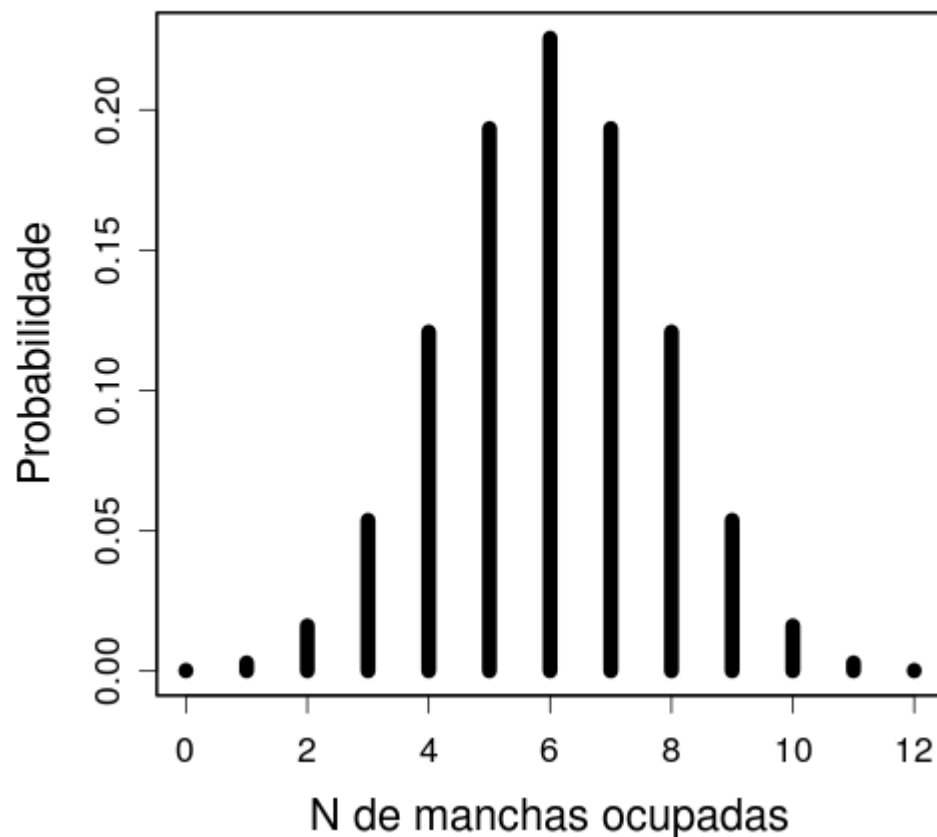
exemplo:

$$f(x=12 | N=12, p=0,5) = 0,0002441406$$

Resposta*: distribuição Binomial

$$f(x) = \frac{12!}{12!(12-x)!} 0,5^x (1-0,5)^{12-x}$$

Binomial, N=12, p=0,5



N = 12 locais

p = ocupação = 0,5

* assumindo p constante e independência das ocorrências

ou **Modelo Binomial**

$$P(x) = \binom{N}{x} p^x (1-p)^{N-x}$$

PARÂMETROS

N = número de tentativas

p = probabilidade de sucesso por tentativa

Obs: $\binom{N}{x} = \frac{N!}{x!(N-x)!}$

Dados:

$x = 6$ de $N=12$ sítios ocupados

Qual é a probabilidade de ocupação por sítio?



Dados:

$x = 6$ de $N=12$ sítios ocupados

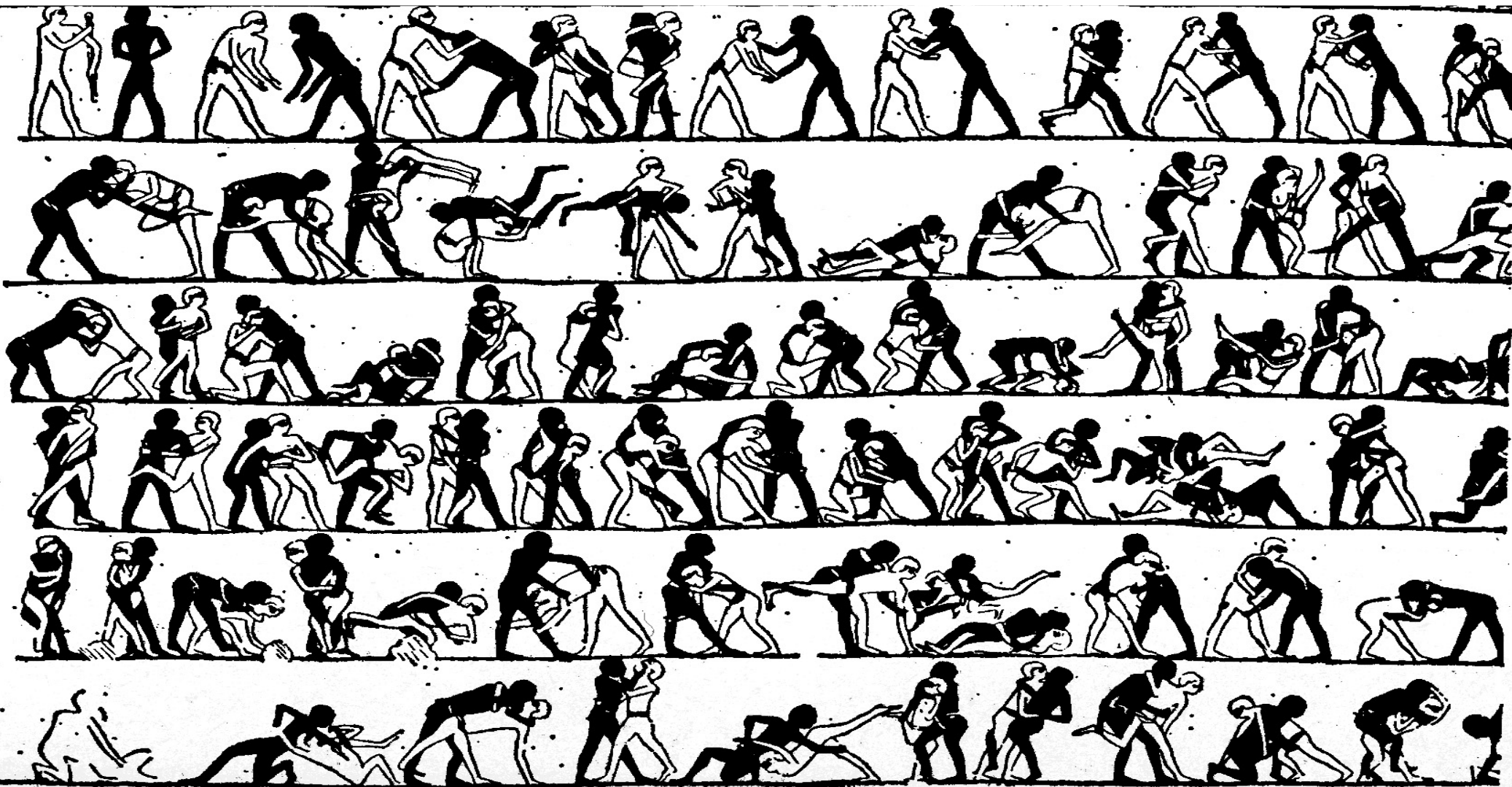
Qual o melhor palpite para a probabilidade de ocupação por sítio?



$p = 0,5$

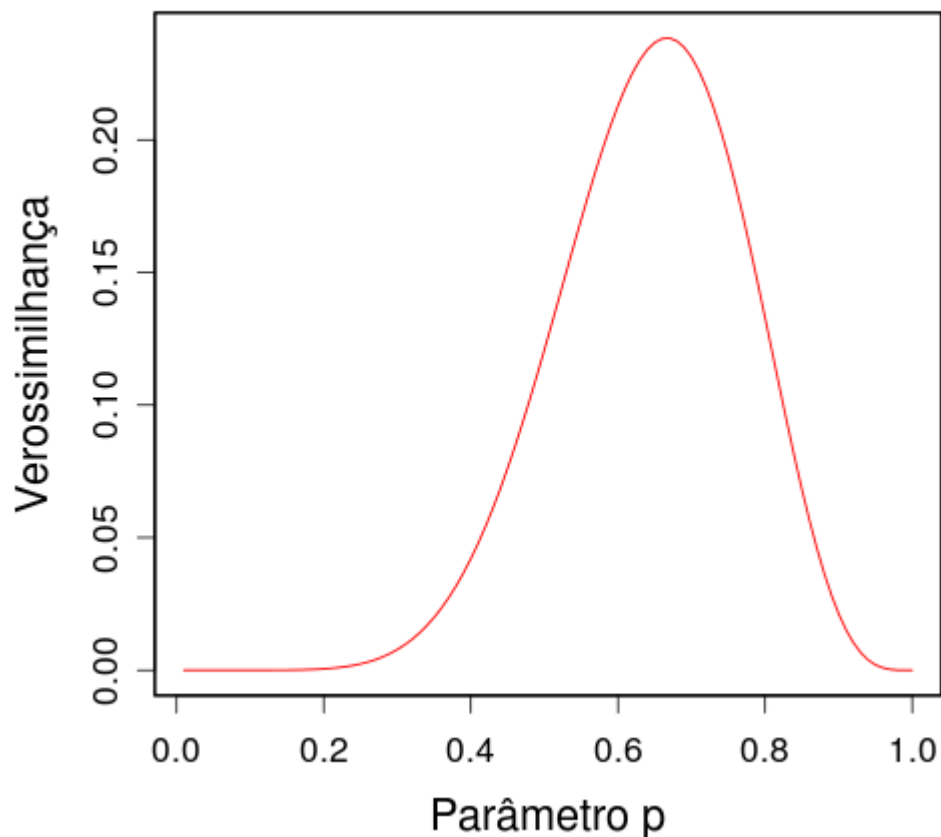
Ó, mortal!

Hora da animação!



Função de Verossimilhança

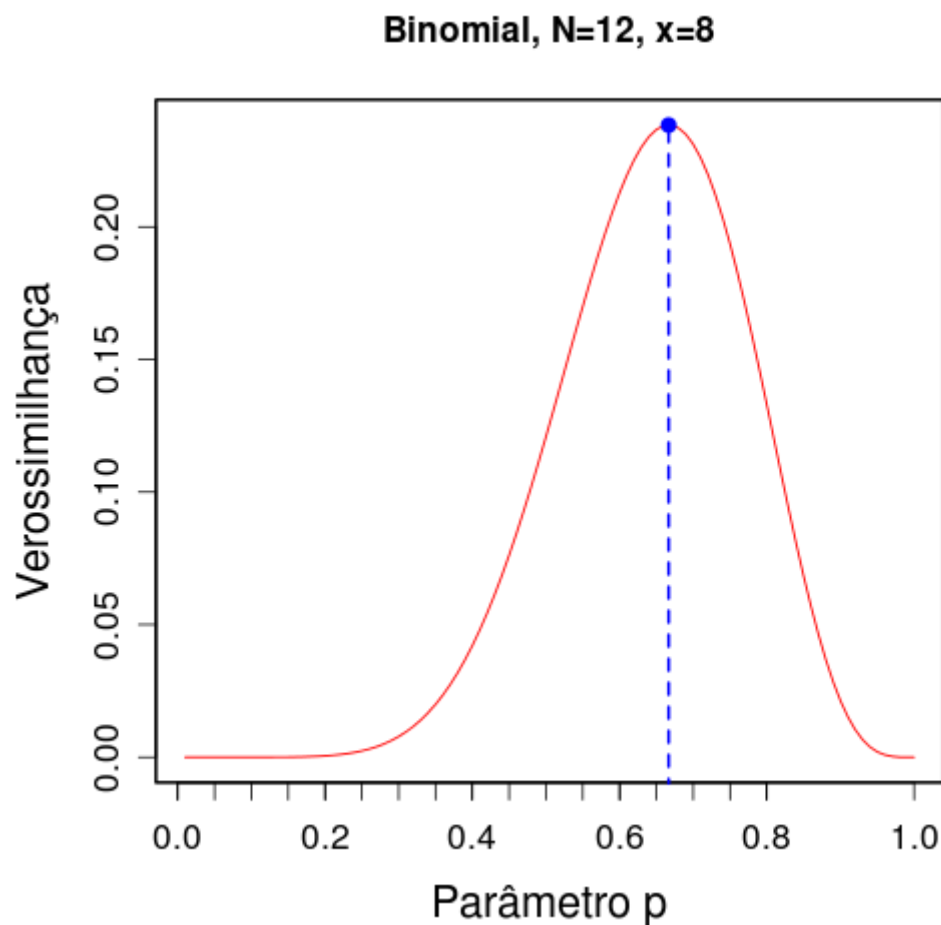
Binomial, $N=12$, $x=8$



$N = 12$ locais
 $x = 8$ ocupados

$$L(p) = \frac{12!}{12!(12-8)!} p^8 (1-p)^{12-8}$$

Estimadores de Máxima Verossimilhança (MLEs)



O valor mais plausível de um parâmetro, dadas as observações.

OU

O valor de parâmetro do modelo que atribui a maior probabilidade a um conjunto de dados.

Lei da Verossimilhança (Um Enunciado Informal*)

Dado que:

- Há mais de uma explicação para um conjunto de dados.
- Cada explicação atribui uma probabilidade diferente aos dados.

Então:

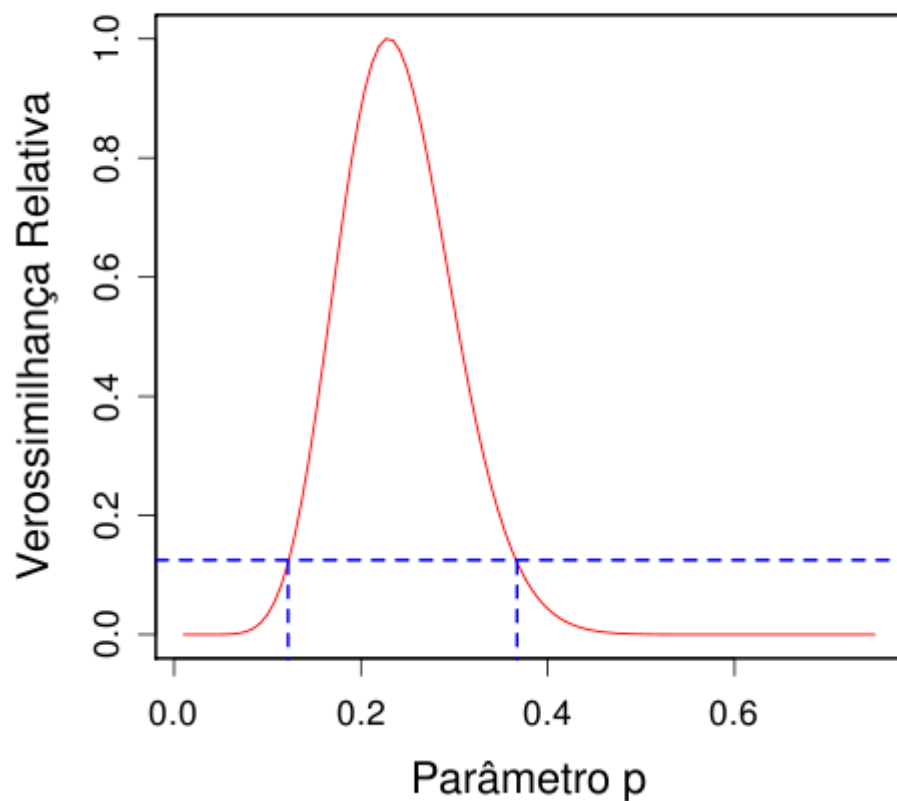
A EXPLICAÇÃO MAIS PLAUSÍVEL SERÁ AQUELA QUE ATRIBUIR A MAIOR PROBABILIDADE AOS DADOS.

* Para enunciado formal e mais detalhes, disciplina BIE5781

Função de Log-Verossimilhança Negativa

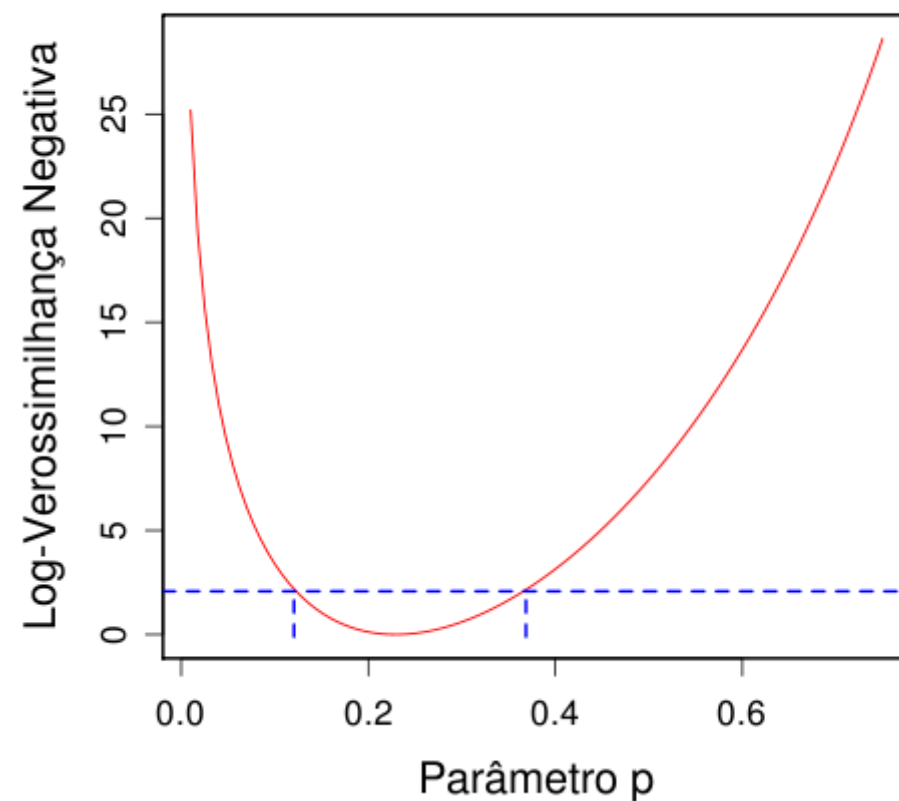
Verossimilhança

Binomial, $N=12$, $x=(3,3,3,2)$



Log-Verossimilhança

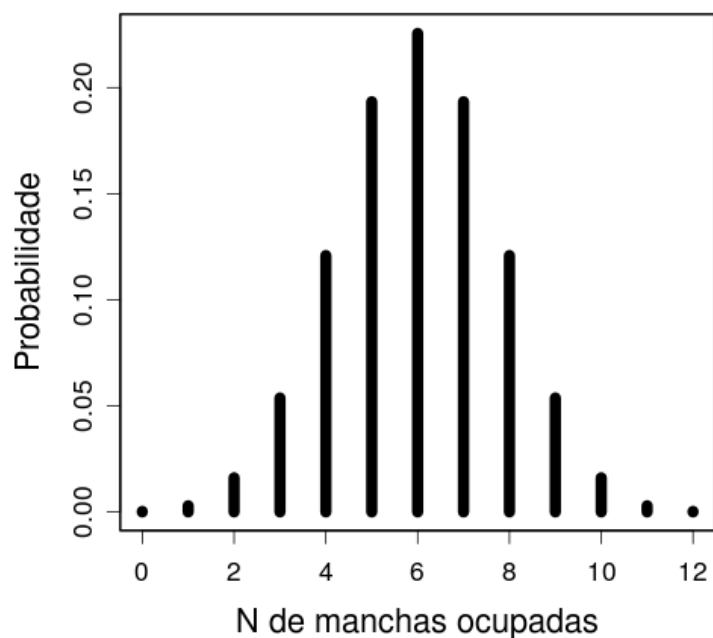
Binomial, $N=12$, $x=(3,3,3,2)$



Probabilidade x Verossimilhança

PARÂMETROS FIXOS

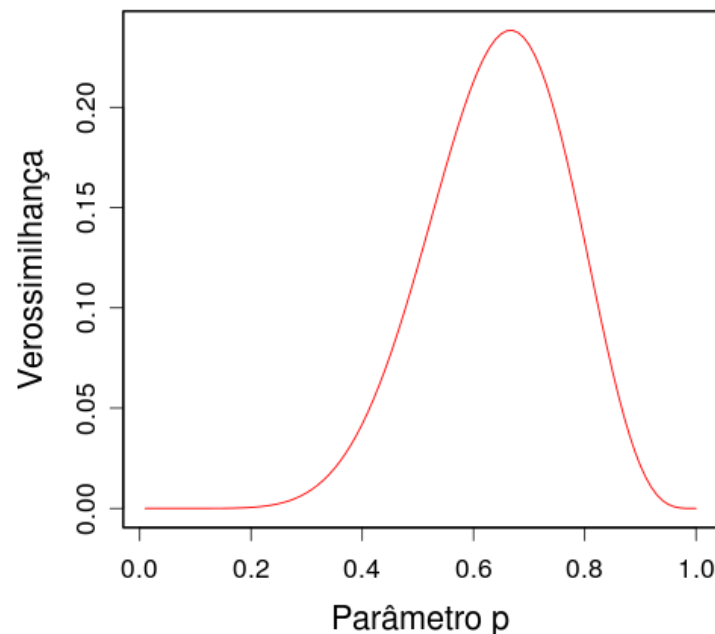
Binomial, $N=12$, $p=0,5$



- Funções de uma variável aleatória
- Parâmetros conhecidos
- Somam (ou integram) um
- Podem ser discretas ou contínuas

DADOS FIXOS

Binomial, $N=12$, $x=8$

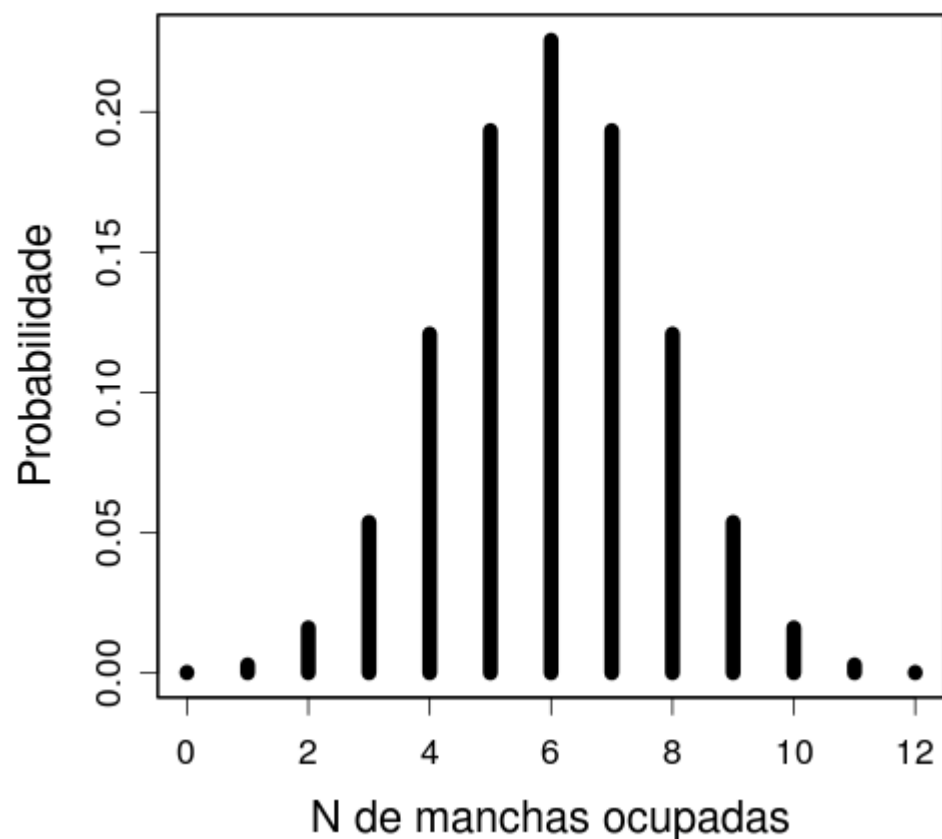


- Funções dos parâmetros
- Dados conhecidos
- Não precisam ter integral um
- São contínuas

Probabilidade

PARÂMETROS FIXOS

Binomial, $N=12$, $p=0,5$

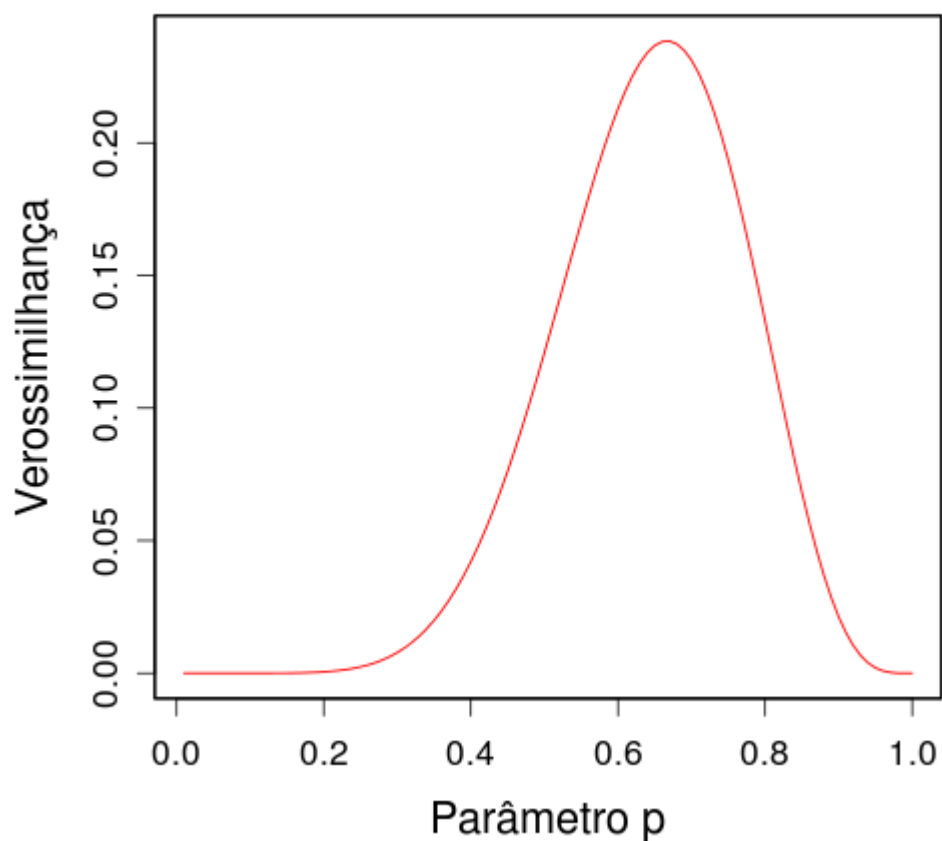


- Funções de uma variável aleatória
- Parâmetros conhecidos
- Somam (ou integram) um
- Podem ser discretas ou contínuas

Verossimilhança

DADOS FIXOS

Binomial, $N=12$, $x=8$



- Funções dos parâmetros
- Dados são conhecidos
- Não integram um
- São funções contínuas

Acrescentando um parâmetro

Inspeções

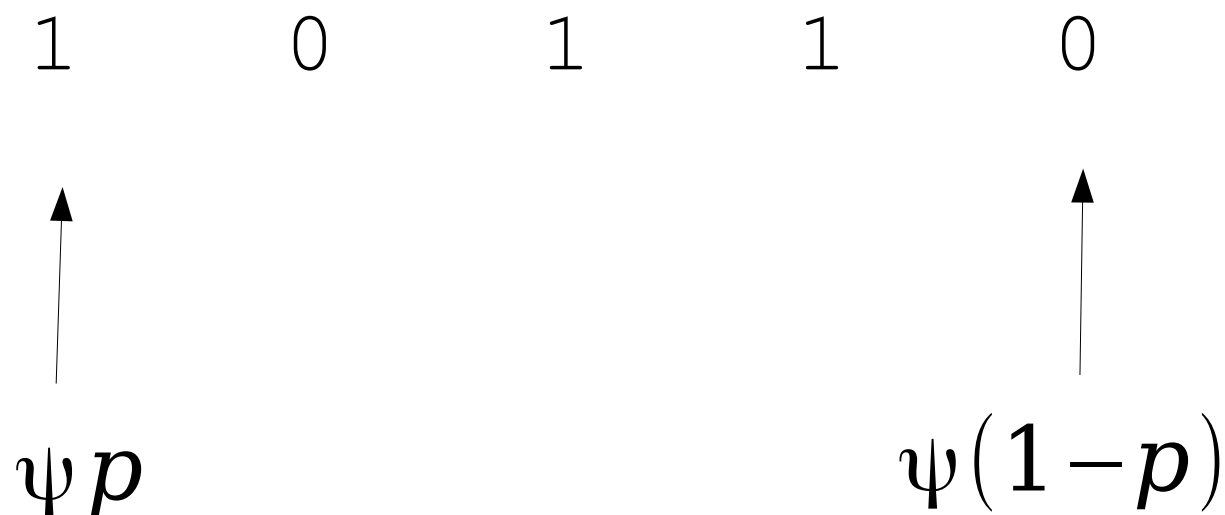
		Inspeções					TOTAL
		[, 1]	[, 2]	[, 3]	[, 4]	[, 5]	
Sítios	[1,]	1	0	1	1	0	3
	[2,]	1	0	0	0	1	2
	[3,]	0	0	0	0	0	0
	[4,]	0	1	0	0	0	1
	[5,]	0	1	1	1	0	3
	[6,]	0	0	0	0	1	1
	[7,]	0	0	0	0	0	0
	[8,]	0	0	0	0	0	0
	[9,]	1	1	1	0	0	3
	[10,]	0	0	0	0	0	0
	[11,]	0	0	1	1	0	2
	[12,]	1	1	1	1	0	4

Para seguir adiante

A probabilidade de dois eventos independentes é o produto das probabilidades de cada evento:

$$P(A \& B) = P(A) \times P(B)$$

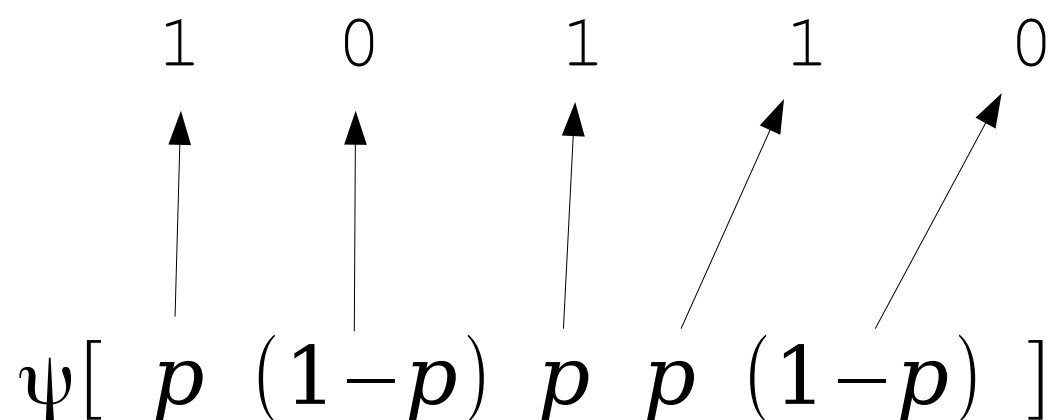
Acrescentando um parâmetro



Ψ : probabilidade de ocupação

p : probabilidade de detecção, **dado que está presente**

Histórico de registros : $n > 0$



Ψ : probabilidade de ocupação

p : probabilidade de detecção, **dado que está presente**

Probabilidade de n registros

$$P(n) = \psi p^n (1-p)^{(N-n)}$$

Ψ : probabilidade de ocupação

p : probabilidade de detecção, **dado que está presente**

n : número de inspeções com presenças observadas

N : total de inspeções

Para seguir adiante

A probabilidade de dois eventos alternativos é a soma das probabilidades de cada evento:

$$P(A \text{ ou } B) = P(A) + P(B)$$

Histórico de registros: $n=0$

0 0 0 0 0

$$\psi [(1-p) (1-p) (1-p) (1-p) (1-p)] + (1-\psi)$$

Ψ : probabilidade de ocupação

p : probabilidade de detecção, **dado que está presente**

Probabilidade de zero registros

$$P(0) = \psi (1-p)^N + (1-\psi)$$

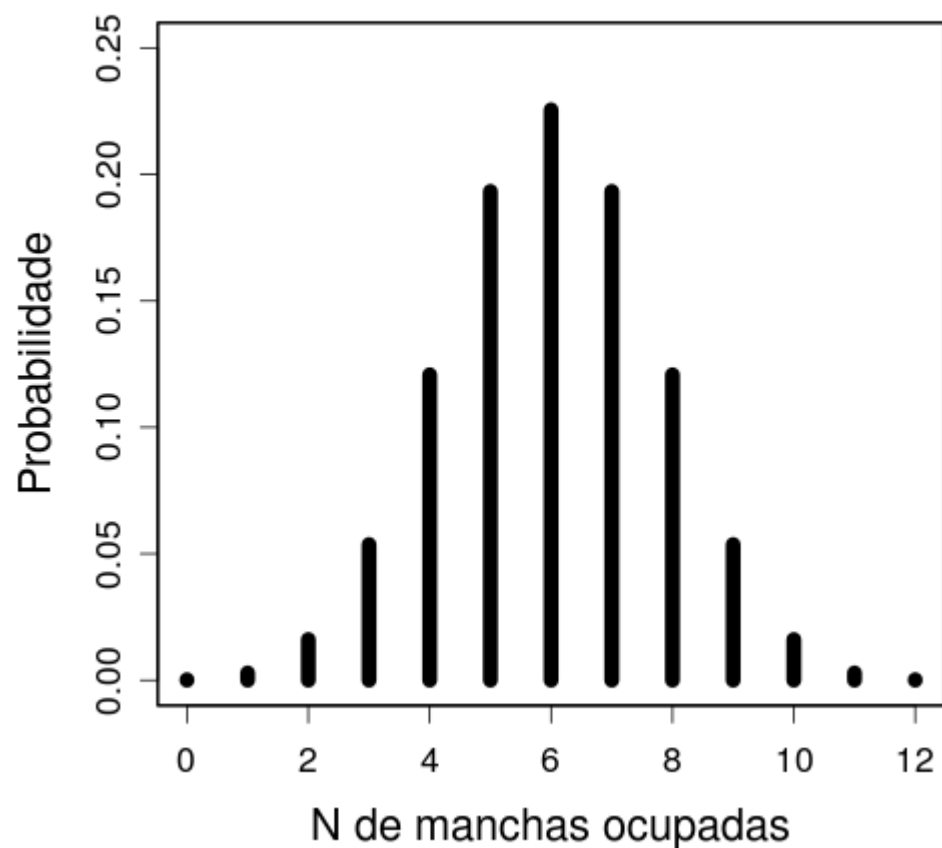
Ψ : probabilidade de ocupação

p : probabilidade de detecção, **dado que está presente**

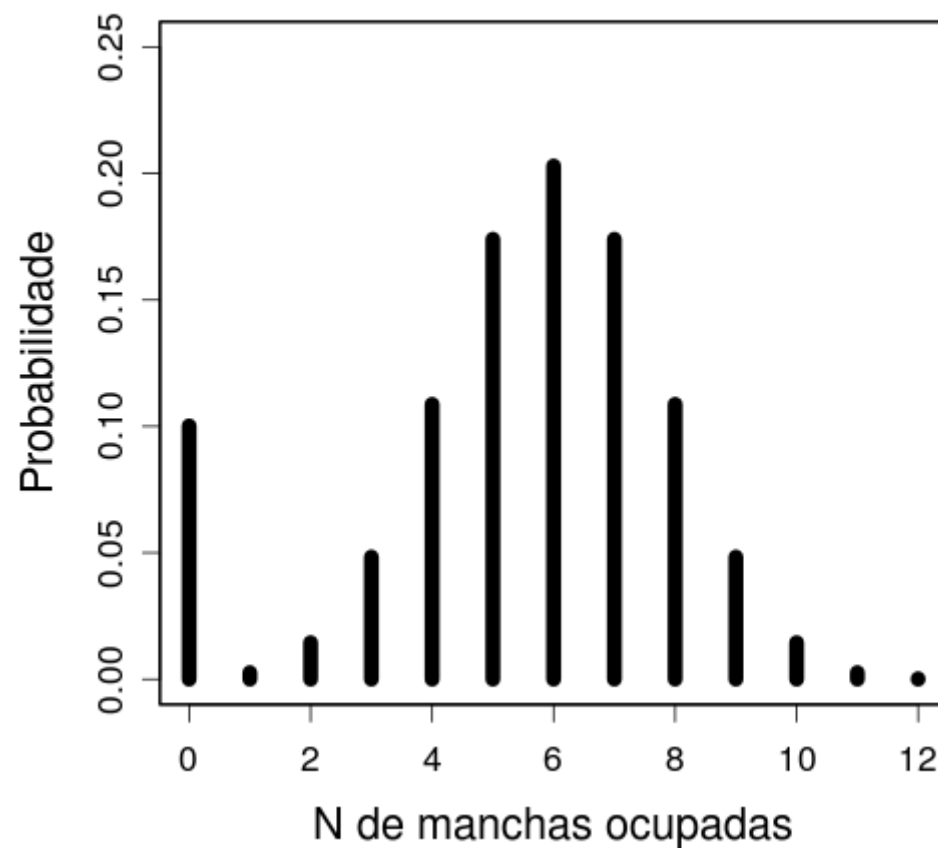
N : total de inspeções

O modelo neste caso: binomial com excesso de zeros

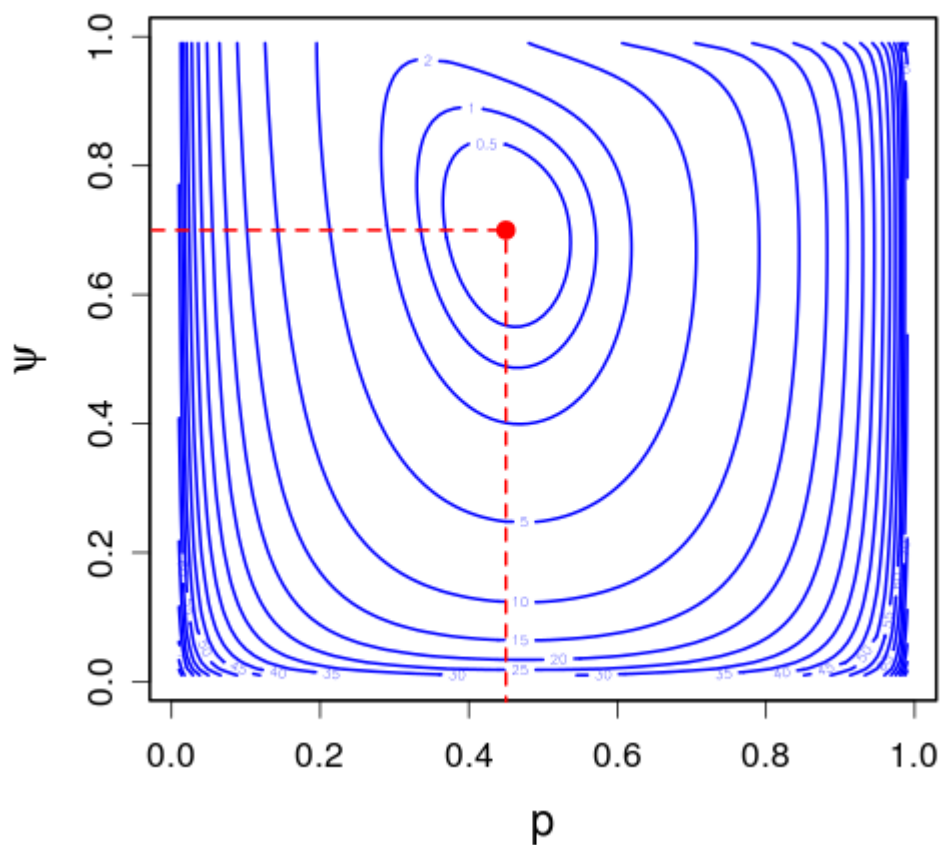
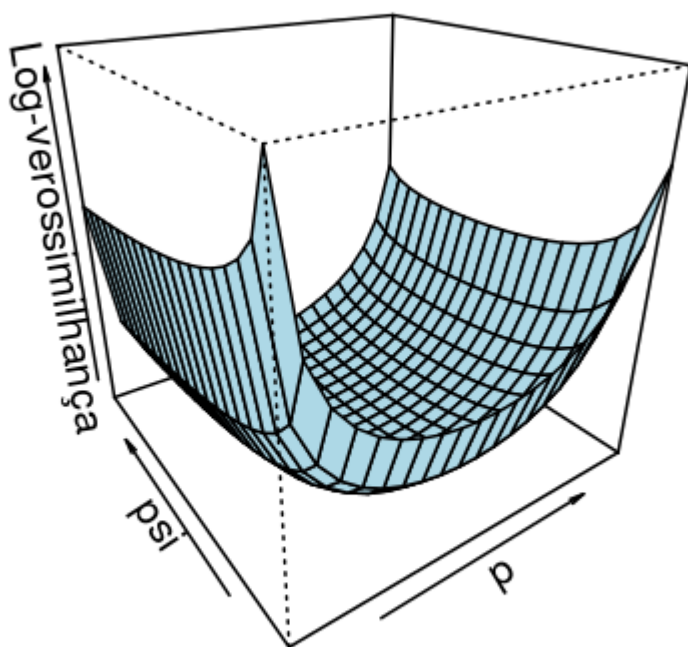
Binomial, $N=12$, $p=0,5$



ZI Binomial, $N=12$, $p=0,5$, $\psi=0,9$



Dois parâmetros: Superfície de Verossimilhança



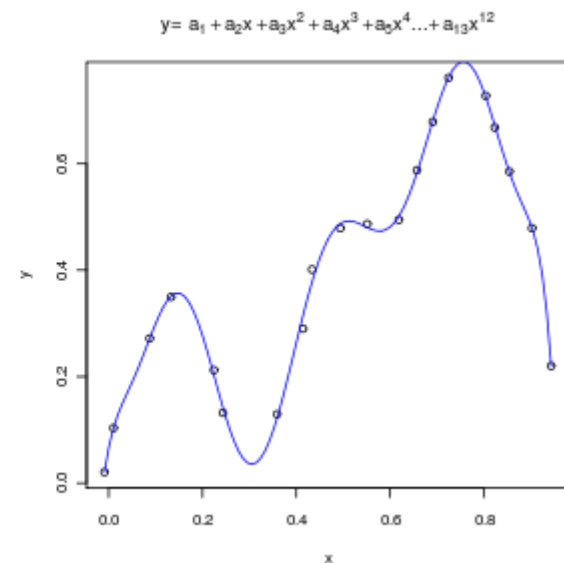
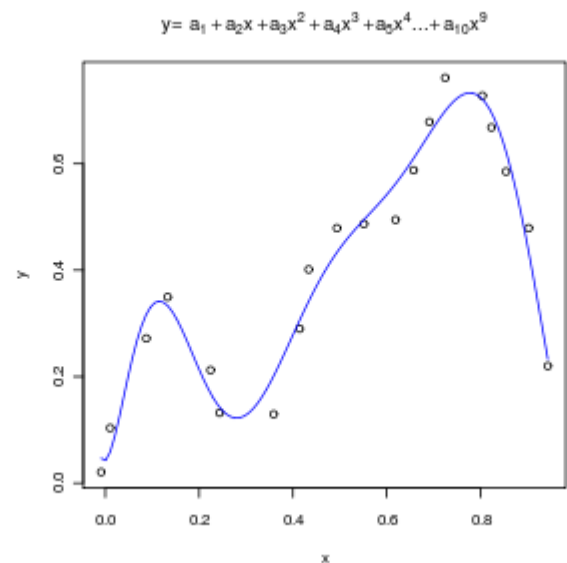
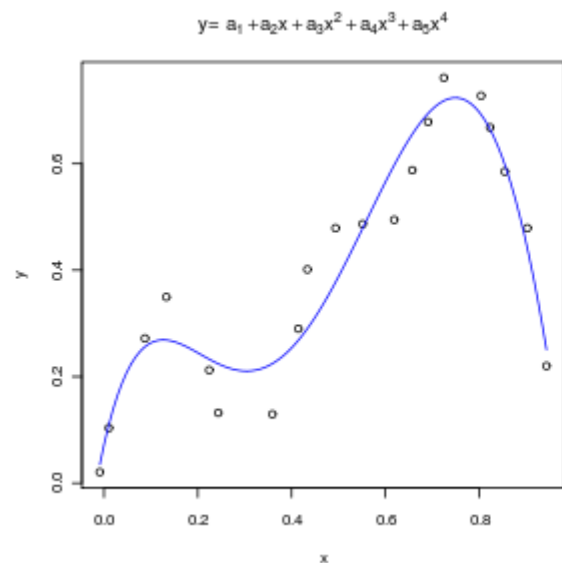
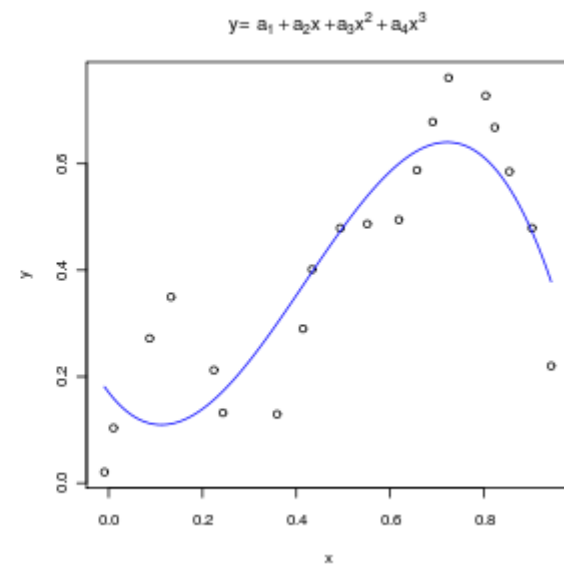
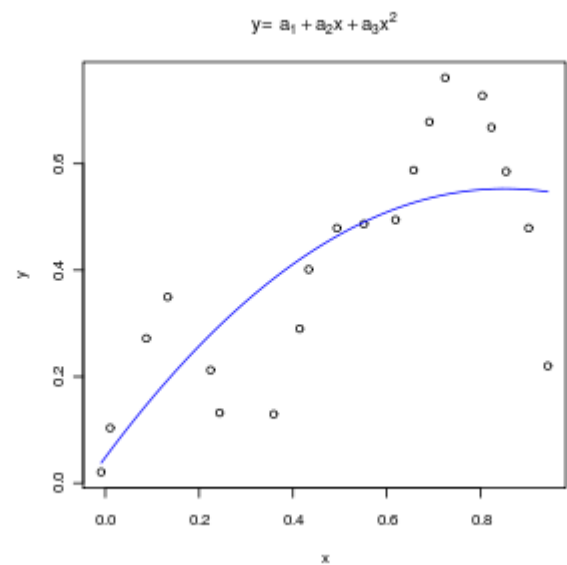
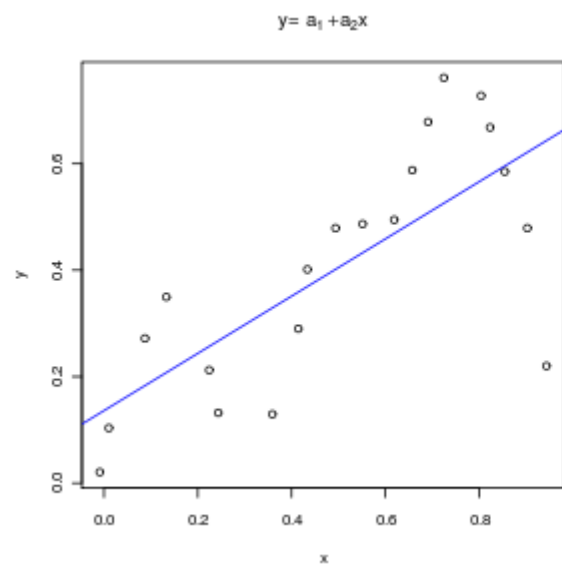
Parcimônia!



ockham wielding razor

MODELO	Parâmetros	LL
H1	ρ	-36,5
H2	ρ, ψ	-24,5

Super-parametrização



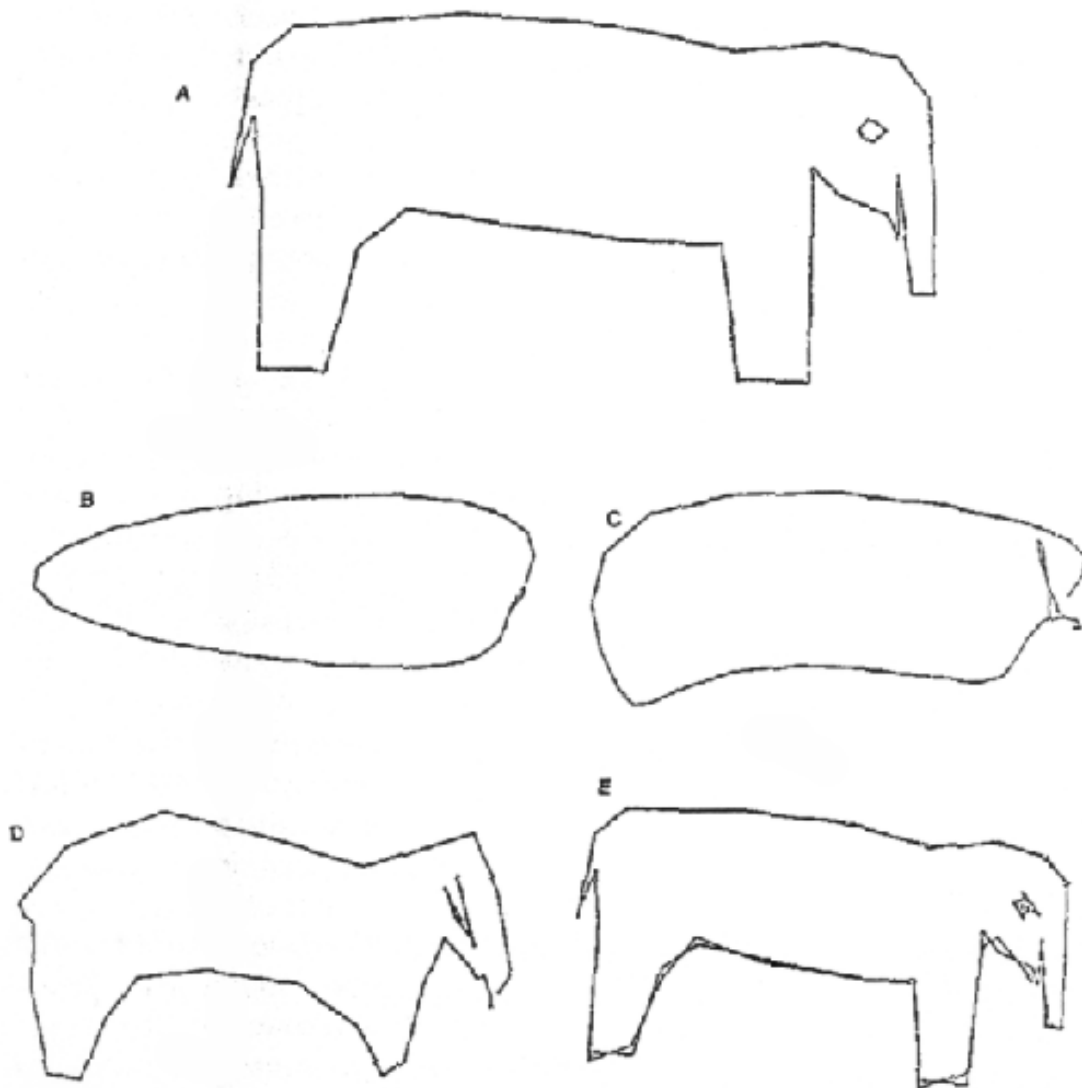


FIGURE 1.2. "How many parameters does it take to fit an elephant?" was answered by Wei (1975). He started with an idealized drawing (A) defined by 36 points and used least squares Fourier sine series fits of the form $x(t) = \alpha_0 + \sum \alpha_i \sin(it\pi/36)$ and $y(t) = \beta_0 + \sum \beta_i \sin(it\pi/36)$ for $i = 1, \dots, N$. He examined fits for $K = 5, 10, 20,$ and 30 (shown in B–E) and stopped with the fit of a 30 term model. He concluded that the 30-term model "may not satisfy the third-grade art teacher, but would carry most chemical engineers into preliminary design."

"Me dê 5
parâmetros e
ajusto um
elefante"

J. Von Neumann

Veja também:

J. Wei, "Least Square Fitting of an Elephant,"
CHEMTECH, 5(2), 1975 pp. 128–129.

<http://demonstrations.wolfram.com/FittingAnElephant/>

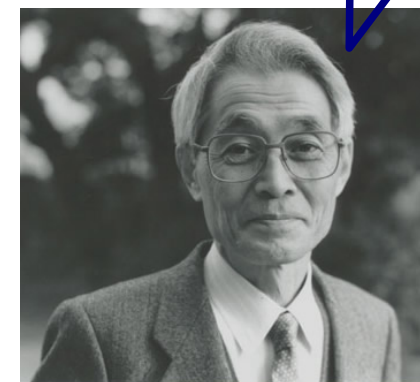
AIC

AIC = - 2 x Log-Verossimilhança + 2 x n de parâmetros

MODELO	Parâmetros	LL	AIC
H1	ρ	-36,5	75,0
H2	ρ, ψ	-24,5	53,0

H2 vence!

Hirotsugu Akaike (1927-2010)



Os 3 passos da modelagem estatística



1. Defina os modelos concorrentes.
2. Busque o melhor ajuste de cada modelo: combinação de parâmetros que maximiza a verossimilhança.
3. Fique com o melhor modelo (menor AIC).

RESUMO

- Modelos estatísticos: descrevem a probabilidade de que sua variável assumira um certo valor.
- Os modelos diferem quanto aos seus parâmetros.

RESUMO

- Se temos dados mas não conhecemos os parâmetros a função de probabilidade torna-se uma função de verossimilhança.
- É mais conveniente usar a função de log-verossimilhança negativa.

RESUMO

- Uma vez obtidos os dados, cada modelo terá uma verossimilhança máxima.
- Os valores dos parâmetros que maximizam a verossimilhança são as estimativas de máxima verossimilhança dos parâmetros (MLEs).

RESUMO

- Para modelos com mais de um parâmetro temos superfícies de verossimilhança.
- O AIC permite comparação de modelos que diferem quanto ao número de parâmetros

Para saber mais

Royle & Dorazio 2008. Hierarchical modeling and inference in ecology. Capítulo 2 - Essentials of statistical inference.

Johnson, J. B. & Omland, K. S. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19:101-108.

Página da Disciplina de Modelagem Estatística Eco-USP: <http://cmq.esalq.usp.br/BIE5781/>