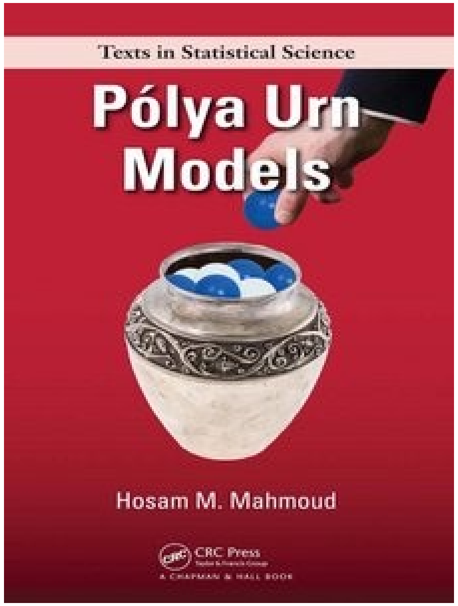
AULA 1:

Introdução à Verossimilhança

Conceitos

1. Verossimilhança
2. Função de Verossimilhança
3. Log Verossimilhança Negativa
4. Método da Máxima Verossimilhança
5. Estimadores de Máxima Verossimilhança (MLEs) e suas propriedades
6. Curva e Superfície de Verossimilhança
7. Critério de informação de Akaike (AIC)
8. Seleção de modelos

O que é Verossimilhança?



Observação:

um sorteio, uma bola branca.

Hipóteses

H_1: Há apenas bolas brancas na urna.

H_2: Metade das bolas da urna são brancas e metade são azuis.

Problema:

Identificar a hipótese mais plausível, **dada a observação**.

Lei da Verossimilhança
(Um Enunciado Informal*)

Dado que:

- Há mais de uma explicação para um conjunto de dados.
- Cada hipótese 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

Qual a Força da Evidência?

- Um sorteio de uma bola.
- A bola sorteada é branca.

$$\boxed{\begin{aligned} P(x=1|H_1)&=1{,}0 \\ P(x=1|H_2)&=0{,}5 \end{aligned}}$$

H_1 é $\frac{1{,}0}{0{,}5}=2$ vezes mais plausível que H_2

E para observações múltiplas?

- Dois sorteios de uma bola cada.
- Em ambos tivemos uma bola branca.

$$\boxed{\begin{array}{c} P(x_1=1, x_2=1|H_1)=1{,}0\times 1{,}0=1{,}0 \\ \\ P(x_1=1, x_2=1|H_2)=0{,}5\times 0{,}5=0{,}25 \end{array}}$$

H_1 é $\frac{1{,}0}{0{,}25}=4$ vezes mais plausível que H_2

Função de Verossimilhança

Qualquer função proporcional ao produto das probabilidades que um modelo atribui a cada valor dos dados*

$$\boxed{L \propto P(x_1|H) \times P(x_2|H) \times \ldots P(x_n|H)}$$

* Sob a premissa de que os dados são realizações independentes de um mesmo processo.

Função de Log-Verossimilhança

Logaritmo de uma função de verossimilhança, ou seja: qualquer função proporcional à soma dos logaritmos das probabilidades que um modelo atribui a cada valor dos dados*

$$\boxed{\mathrm{LL} \propto \ln P\left(x_1 \middle| H\right) + \ln P\left(x_2 \middle| H\right) + \ldots \ln P\left(x_n \middle| H\right)}$$

* Sob a premissa de que os dados são realizações independentes de um mesmo processo.

RESUMINDO ...

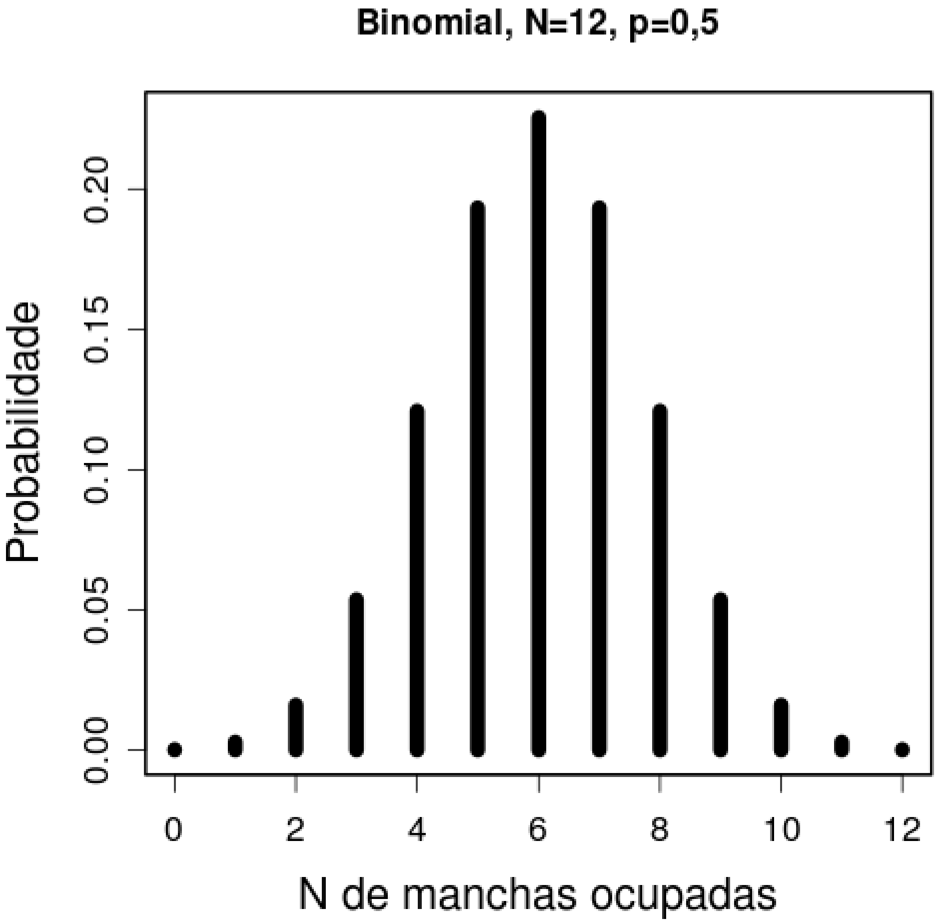
SE:

1. Temos dados que podem ser explicados por mais de uma hipótese, e
2. Cada hipótese é um modelo que atribui alguma probabilidade aos dados

ENTÃO:

Podemos expressar o quão plausível uma hipótese é <u>em relação às outras</u> por meio de uma função, chamada verossimilhança (ou pelo seu logaritmo, chamada função de log-verossimilhança)..

Um Exemplo: Distribuição Binomial

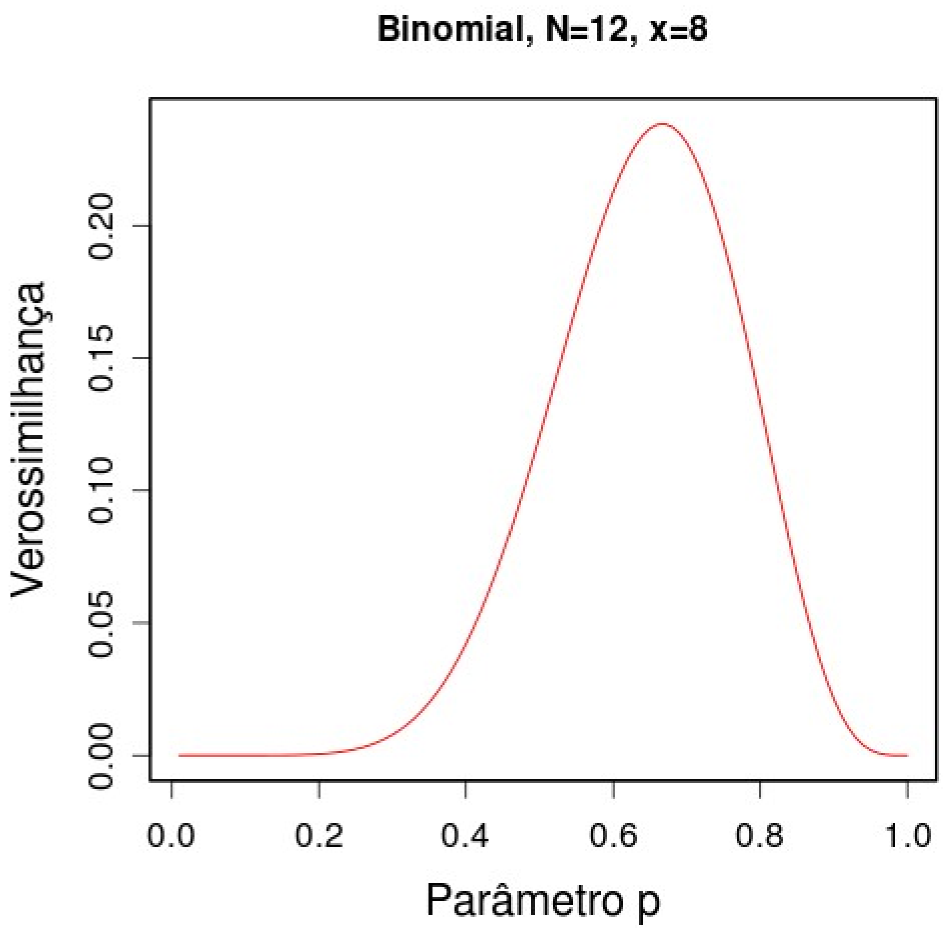


N = 12 locais

p = ocupação = 0,5

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

Função de Verossimilhança



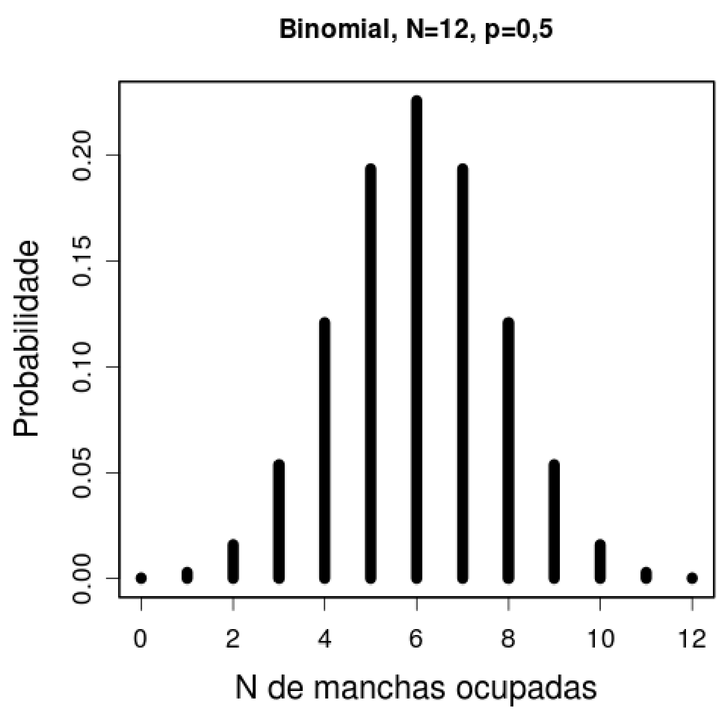
N = 12 locais

x = 8 ocupados

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

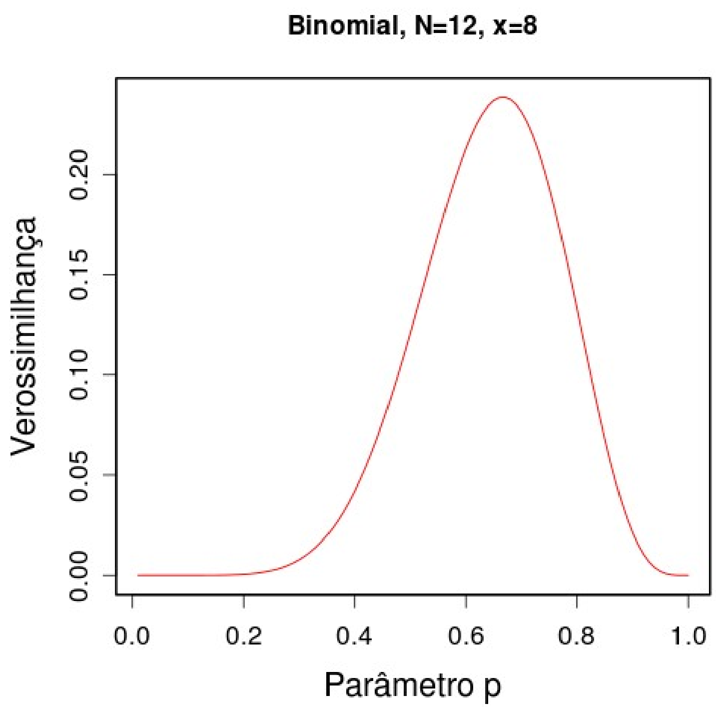
Probabilidade x Verossimilhança

PARÂMETROS FIXOS



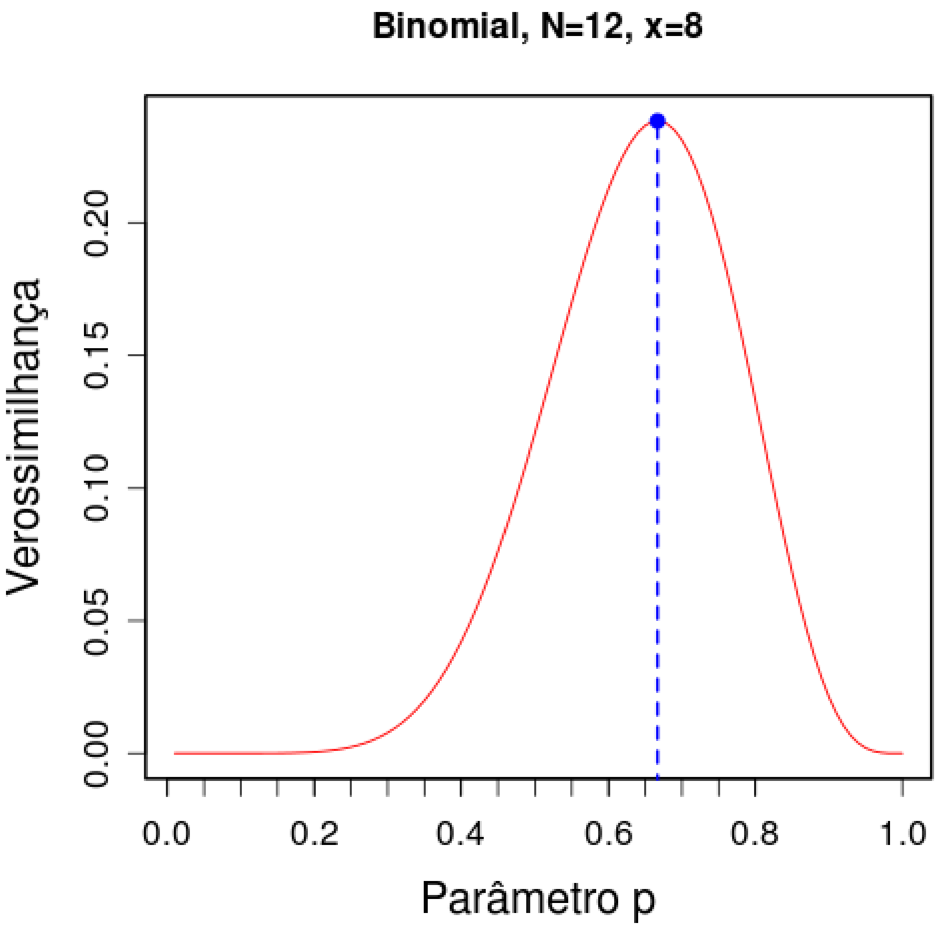
- Funções de uma variável aleatória
- Parâmetros conhecidos
- Somam (ou integram) um
- Podem ser discretas ou contínuas

DADOS FIXOS



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

Estimadores de Máxima Verossimilhança (MLEs)



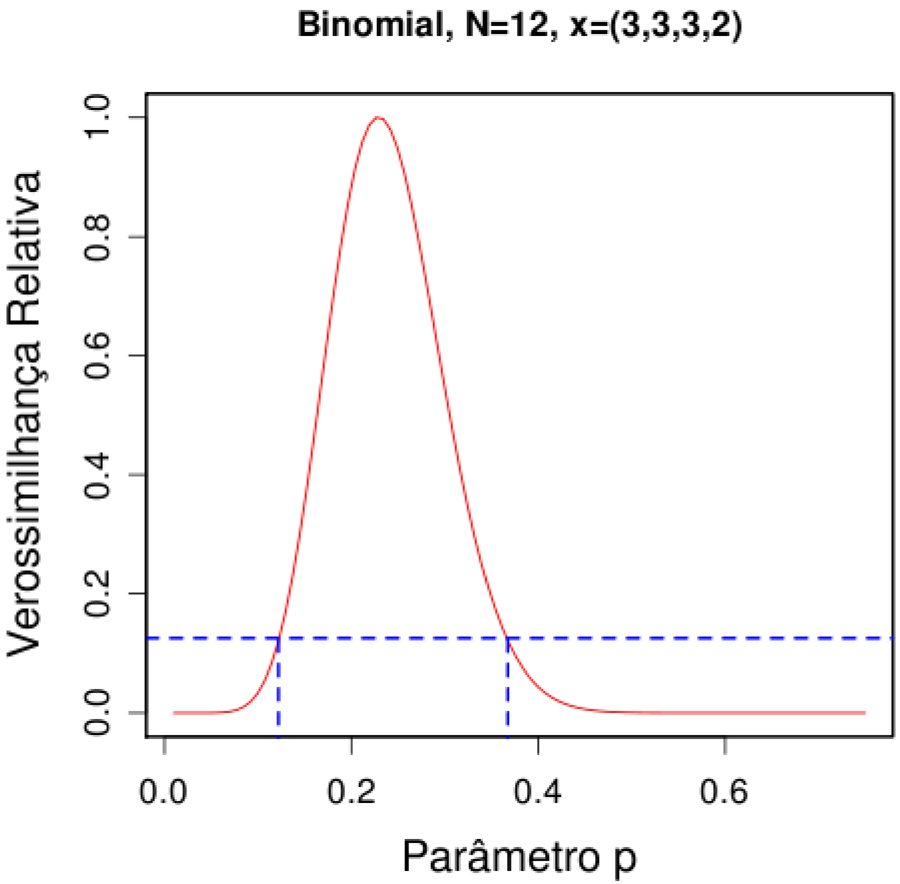
O valor mais plausível de um parâmetro, dadas as observações.

OU

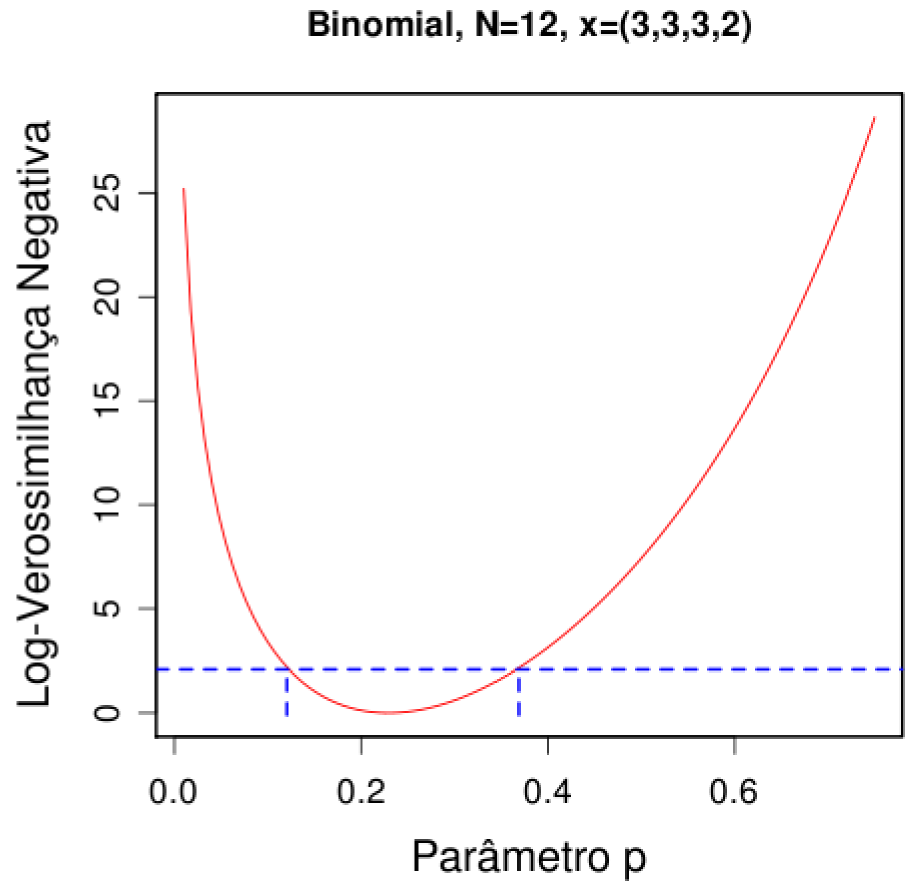
O valor de parâmetro do modelo que atribui a maior probabilidade a um conjunto de dados.

Função de Log-Verossimilhança Negativa

Verossimilhança



Log-Verossimilhança



Acrescentando um parâmetro

Inspeções

Sítios

	[,1]	[,2]	[,3]	[,4]	[,5]	TOTAL
[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

Acrescentando um parâmetro

1 0 1 1 0

p $(1-p)$

Histórico de registros : n>0

1 0 1 1 0

$$\psi\left[\ p\ \ (1-p)\ \ p\ \ p\ \ (1-p)\ \right]$$

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 visita com presenças observadas

N : total de visitas

Histórico de registros: n=0

0 0 0 0 0

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

Ψ : probabilidade de ocupação

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

Probabilidade de zero registros

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

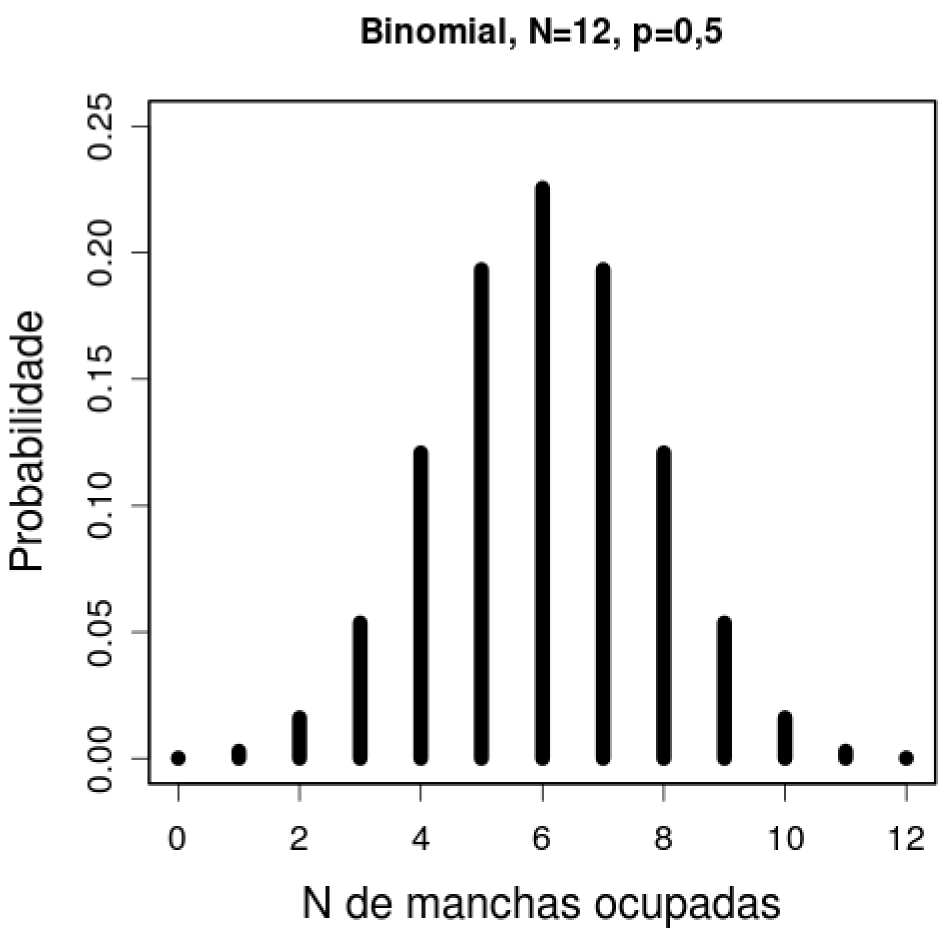
Ψ : probabilidade de ocupação

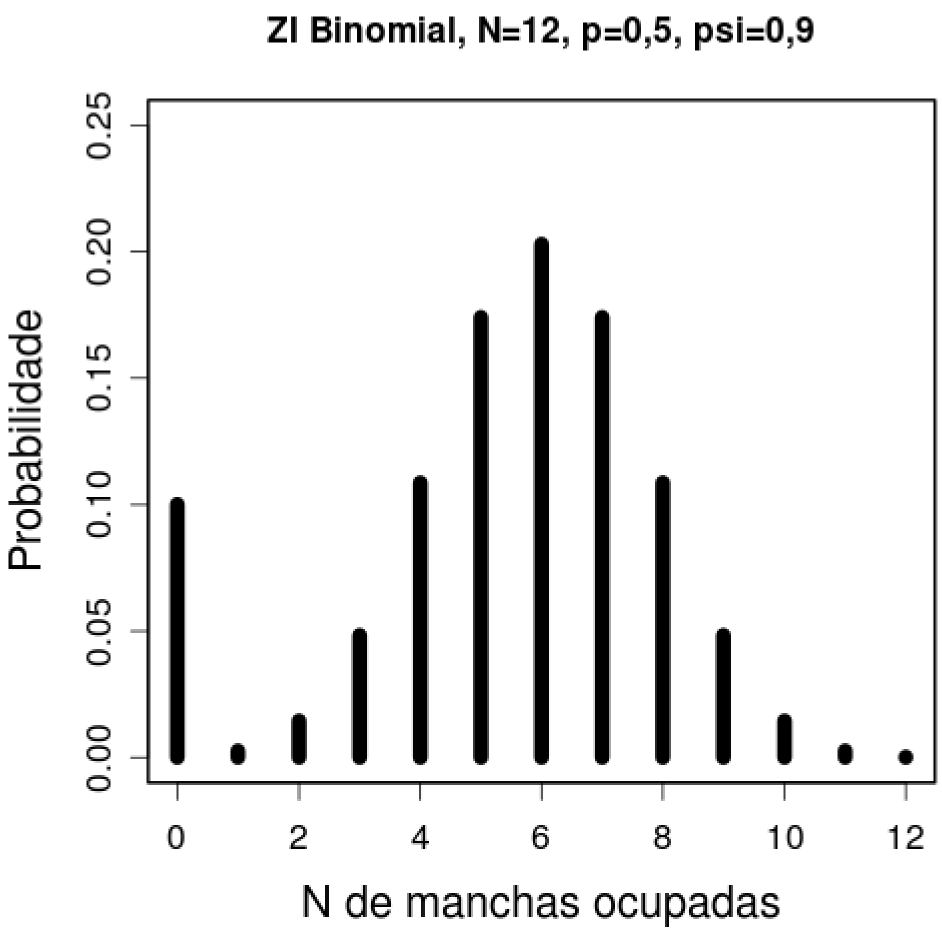
p : probabilidade de detecção, **dado que está presente**

n : número de sítios com registros

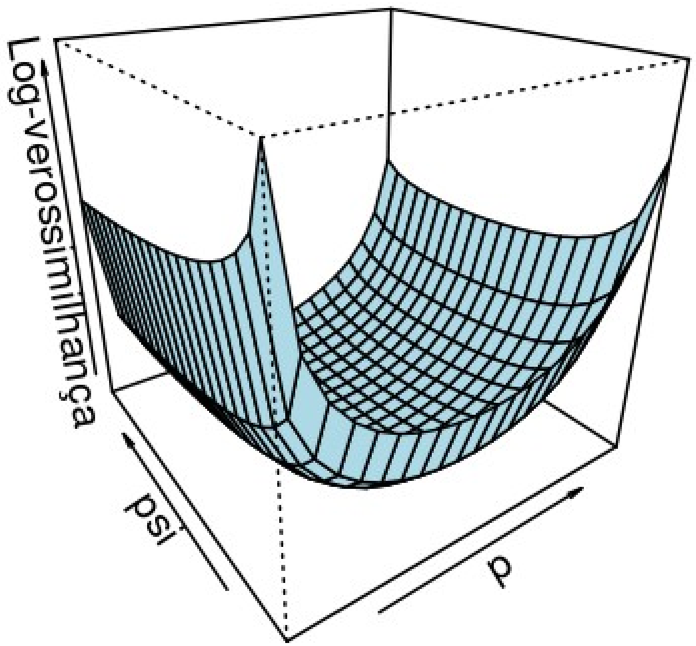
N : total de sítios

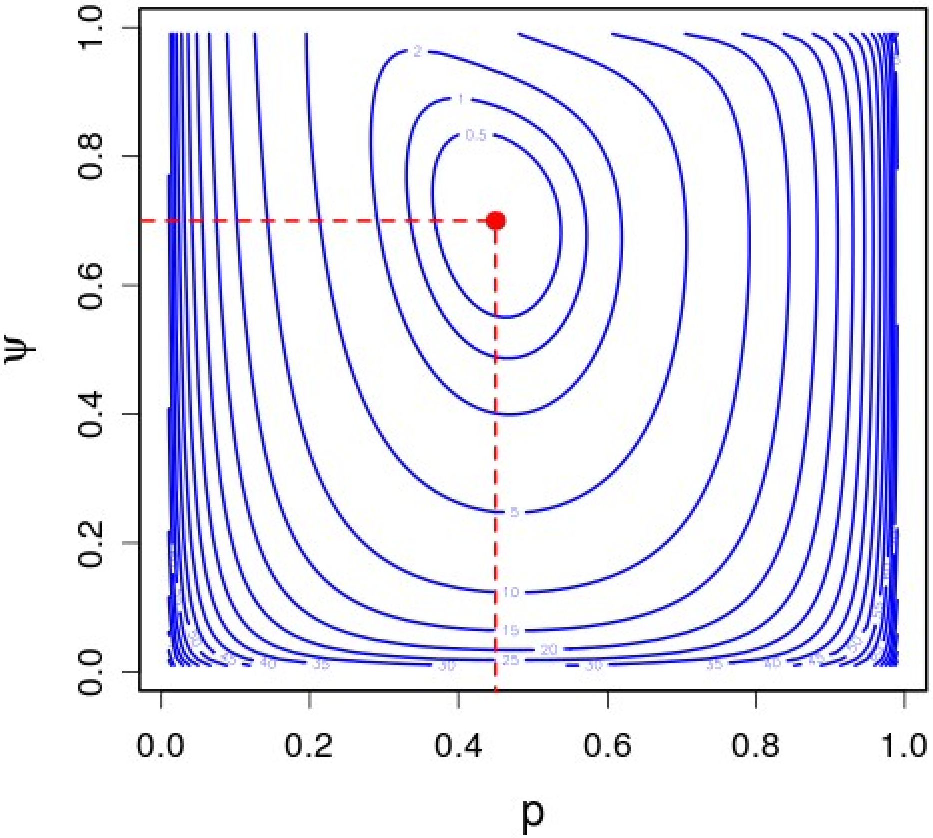
Binomial com Excesso de Zeros





Superfície de Verossimilhança

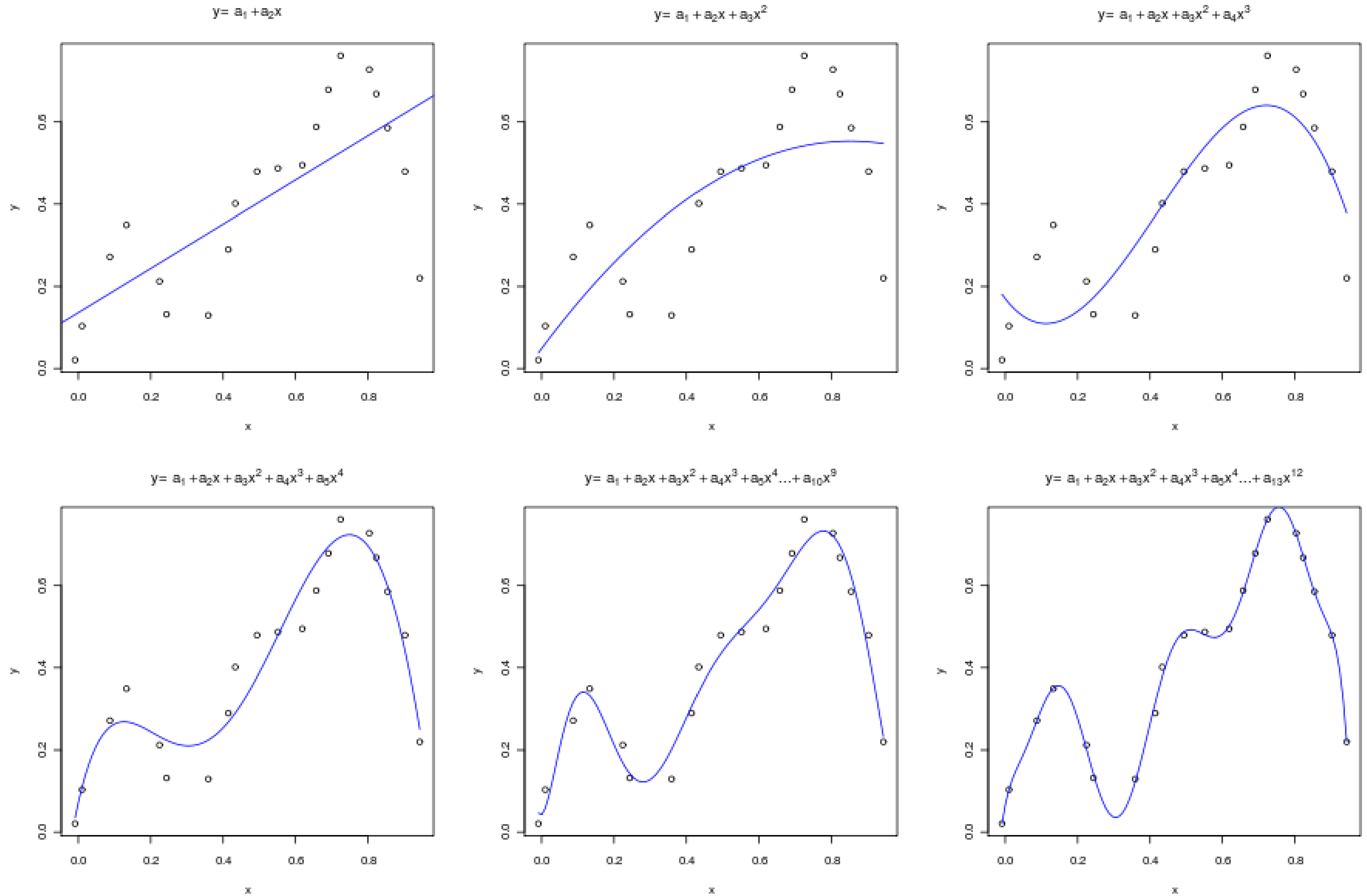




Parcimônia!

MODELO	Parâmetros	LL
H1	p	36,5
H2	p, ψ	24,5

Super-parametrização



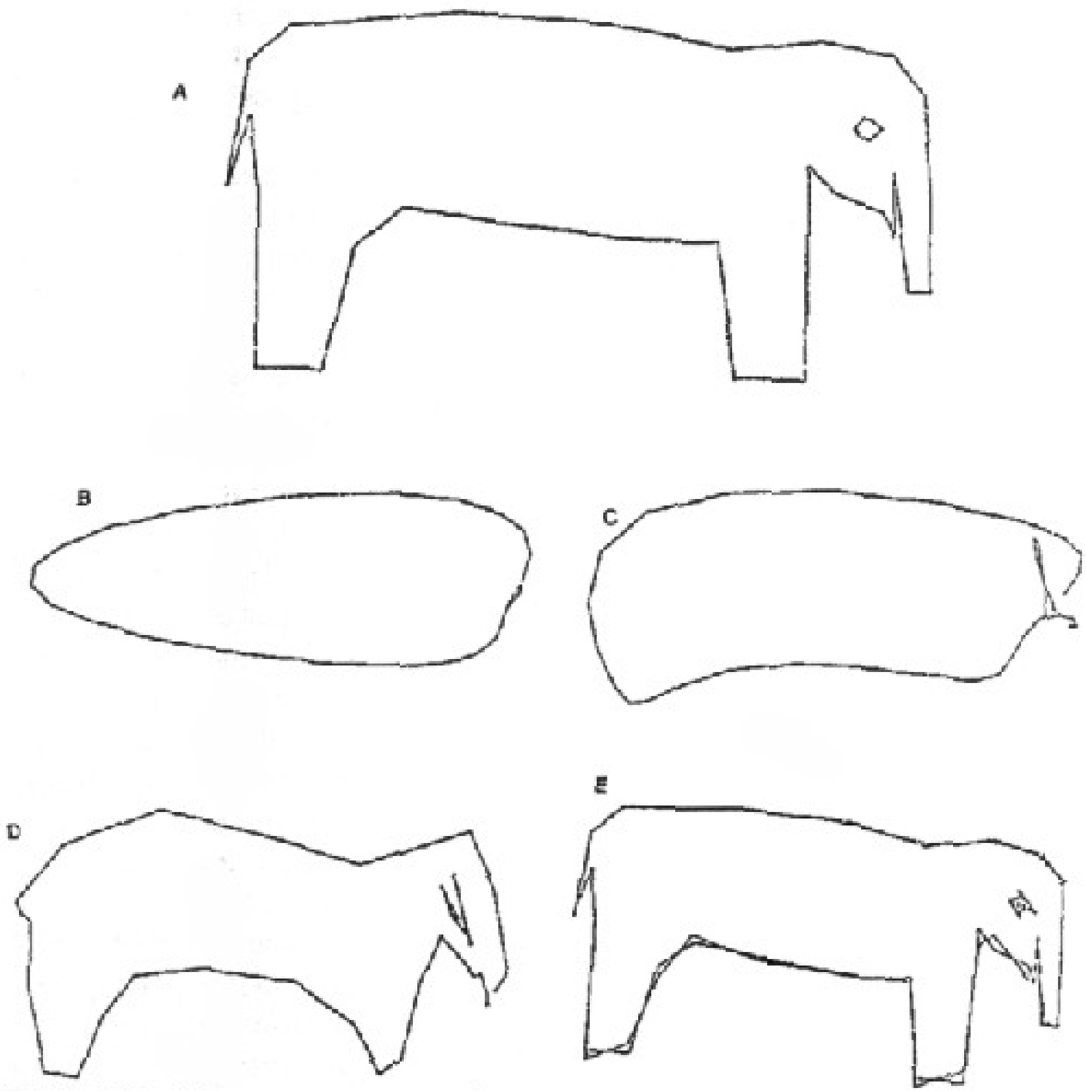


FIGURE 1.2. "How many parameters does does it take to fit an elephant?" was answered by Wel (1975). He started with an idealized drawing (A) defined by 36 points and used least squares Fourier sine series fits of the form $x(t) = \alpha_0 + \sum \alpha_i \sin(it\pi/36)$ and $y(t) = \beta_0 + \sum \beta_i \sin(it\pi/36)$ for $i = 1, \ldots, N$. 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

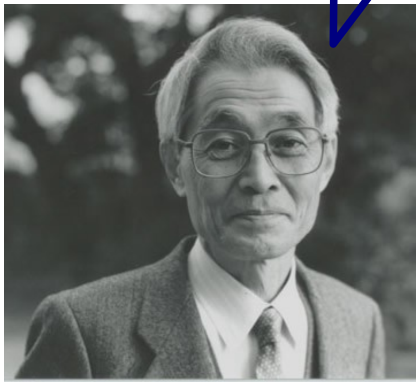
Veja também:
J. Wei, "Least Square Fitting of an Elephant," CHEMTECH, 5(2), 1975 pp. 128–129.
http://demonstrations.wolfram.com/FittingAnElephant

AIC

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

MODELO	Parâmetros	LL	AIC
H1	p	36,5	75,0
H2	p, 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 assuma 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

Burnham, K. P., & Anderson, D. R. (2002). Model Selection and Multimodel Inference: A Practical-Theoretic Approach, 2nd ed. New York, Springer-Verlag.

Bolker, B. (2008). Ecological Models and Data in R. Princeton, Princeton University Press.

Página da Disciplina de Modelagem Estatística USP: http://cmq.esalq.usp.br/wiki/doku.php?id=biometria:verossim:start