

2. Estimando a sobrevivência: Modelo de Cormack-Jolly-Seber

ϕ - phi

Leonardo L. Wedekin

Curso Ecologia Populacional – USP 2011

Resumo

1. Cormack-Jolly-Seber

Parâmetros

Premissas

2. Testes GOF

Testes das premissas

3. Seleção de modelos

Critério de informação de Akaike

Sobredispersão e QAIC

Testes LRT

4. Outros efeitos no modelo CJS

Grupos

Coortes

Idade

MODELOS DE MARCAÇÃO-CAPTURA-RECAPTURA

Animais são capturados e k **ocasiões de captura** (como anos) e marcados durante um período relativamente curto de captura (como semanas).

Períodos de tempo dependem da história de vida do animal.

MODELOS DE MARCAÇÃO-CAPTURA-RECAPTURA

Baseiam-se na **identificação individual**:

- **Marcas naturais** (pigmentação, cicatrizes, calosidades, DNA microsatélite, etc.)
- **Marcas artificiais** (tags, transponders, mutilação, anilhas, tintas, etc.)

Tipos de dados

- Recapturas de animais vivos
- Recuperação de marcas em animais mortos
- Recatpuras viva e de marcas em animais mortos
- Destino conhecido (*known fate*)

População aberta

Abundância muda com nascimentos, mortes,
imigração e emigração

$$N_{(i + 1)} = N_i + B - D + I - E$$

Onde:

- N_i = tamanho da população na ocasião i
- B = nascimentos
- D = mortes
- I = imigração
- E = emigração

Modelo de Cormack-Jolly-Seber

(Cormack, 1964; Jolly, 1965; Seber, 1965)

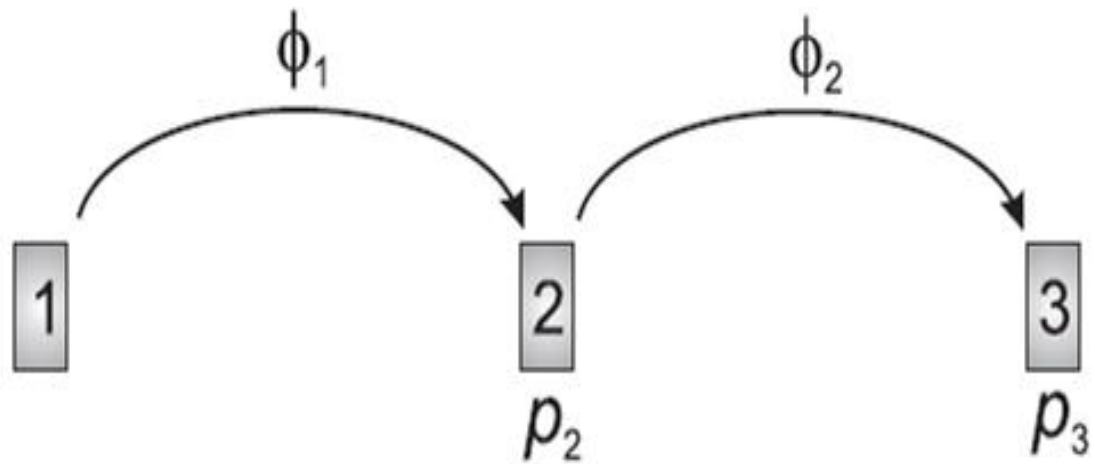


Considera o histórico de capturas a partir da primeira captura do indivíduo

Dois parâmetros são estimados:

- probabilidade de **captura** (p), e...
- probabilidade de **sobrevivência** (ϕ).

Parâmetros CJS



Sobrevivência aparente



Por quê aparente?

Sobrevivência aparente



Não é possível distinguir os processos morte e emigração

$$\phi = S (1 - E)$$

E = probabilidade de emigração

1- ϕ representa indivíduos que morreram ou deixaram a população

Geralmente $\phi < S$

Históricos de capturas dos indivíduos

1 para captura
0 para não captura

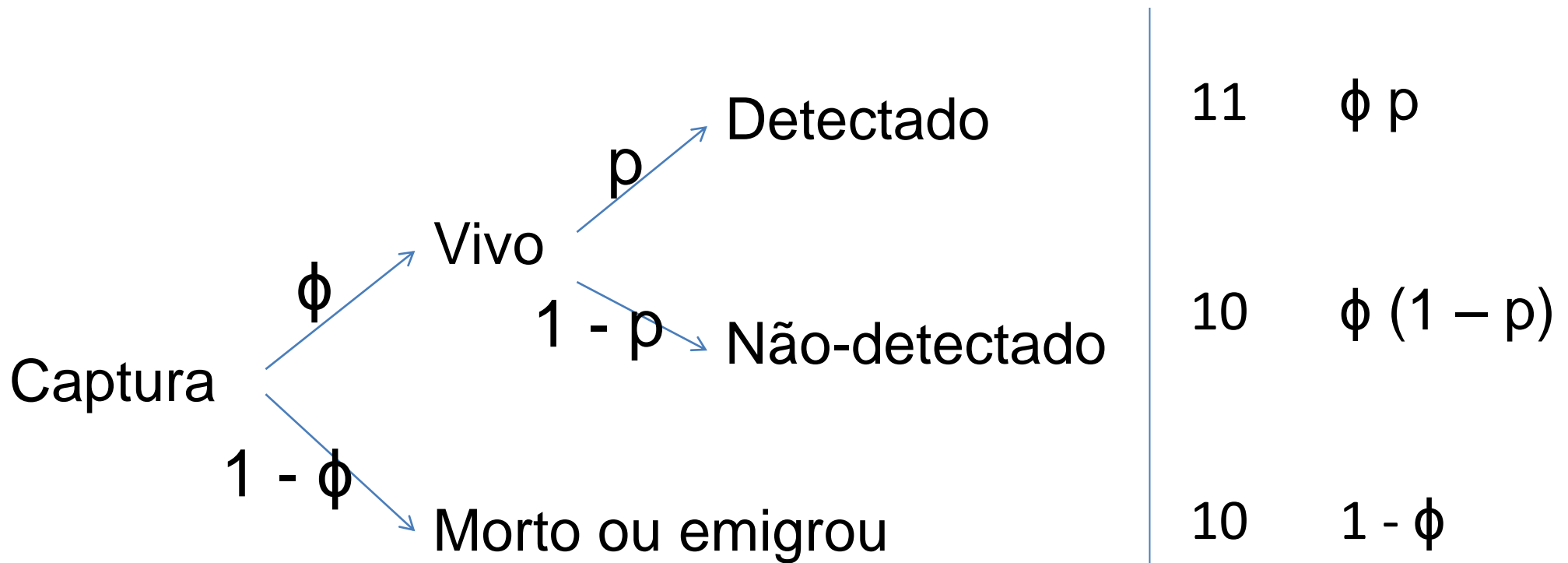
		Ocasião		
		1	2	3
Indivíduos	A	1	0	0
	B	1	1	1
	C	1	0	1
	D	0	1	1
	E	0	0	1
	F	0	0	1

Probabilidade condicional

$$P(A|B)$$

Probabilidade de um evento A
dado o resultado do evento B

Probabilidade de ser detectado condicionada à probabilidade de estar vivo

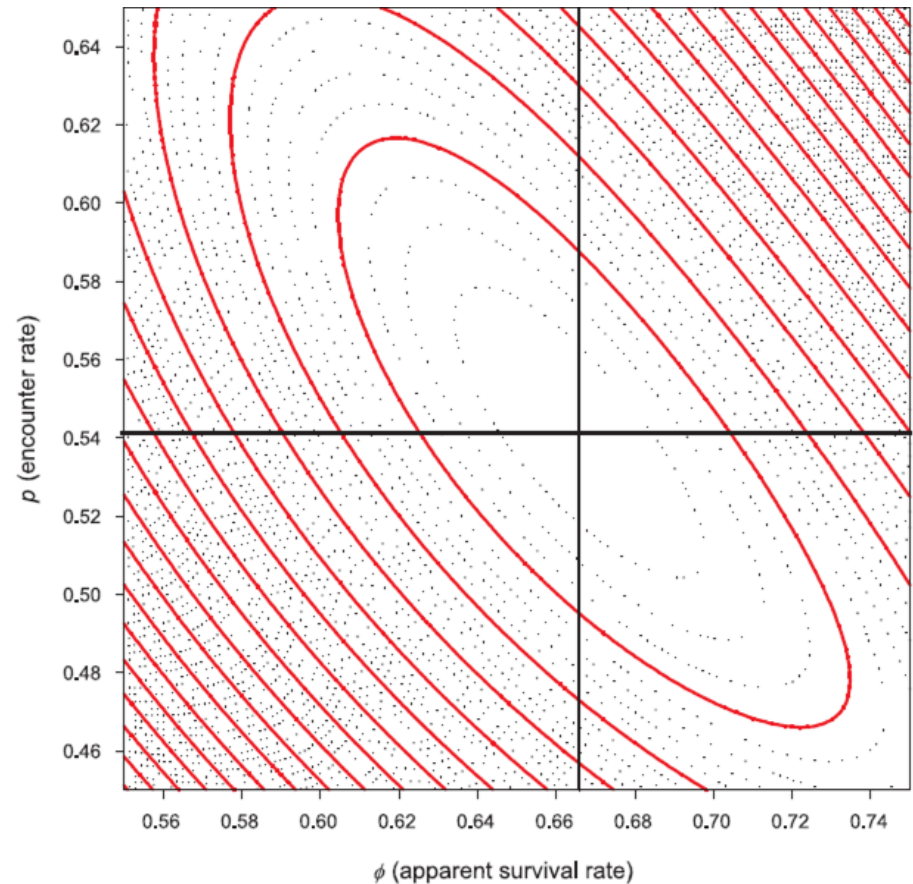


Os dois históricos “10”, que podem ser decorrentes da não sobrevivência ou não detecção do animal são substituídos por outro termo que é a probabilidade do animal não ter sido detectado novamente a partir da primeira captura.

Decompondo as probabilidades

- $P_r \{\text{histórico de captura} \mid \text{primeira captura}\}$
- O produto das probabilidades de cada histórico de capturas segue então uma distribuição multinomial
- Os parâmetros do modelo podem então ser obtidos com base nas frequências de cada histórico de capturas e suas probabilidades

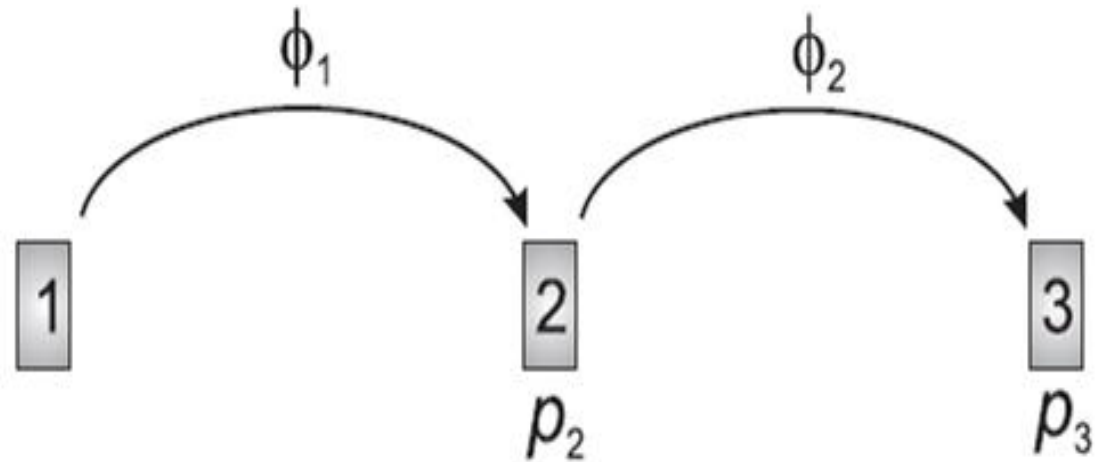
Estimação dos parâmetros por máxima verossimilhança



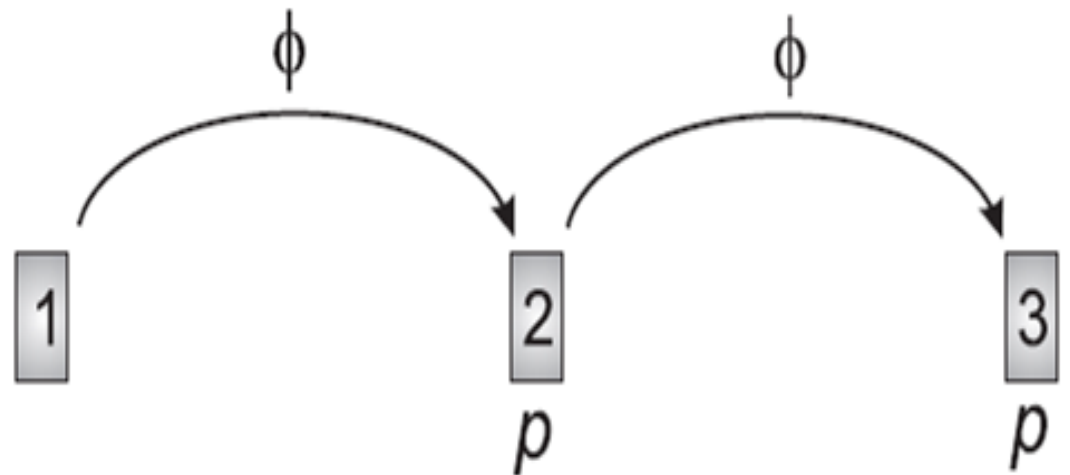
$\ln L(\text{parâmetros do modelo}) = \text{somatório das frequências de cada histórico de captura multiplicado pela } \ln \text{ da sua probabilidade}$

Modelos básicos

Variação temporal
 $\phi(t)$ $p(t)$



Constante no tempo
 $\phi(\cdot)$ $p(\cdot)$





Estimando a sobrevivência e probabilidade de captura da baleia-jubarte no Banco dos Abrolhos entre 1989 e 2008

Migração annual da baleia-jubarte



Identificação individual por marcas naturais



Saídas embarcadas foto-identificação



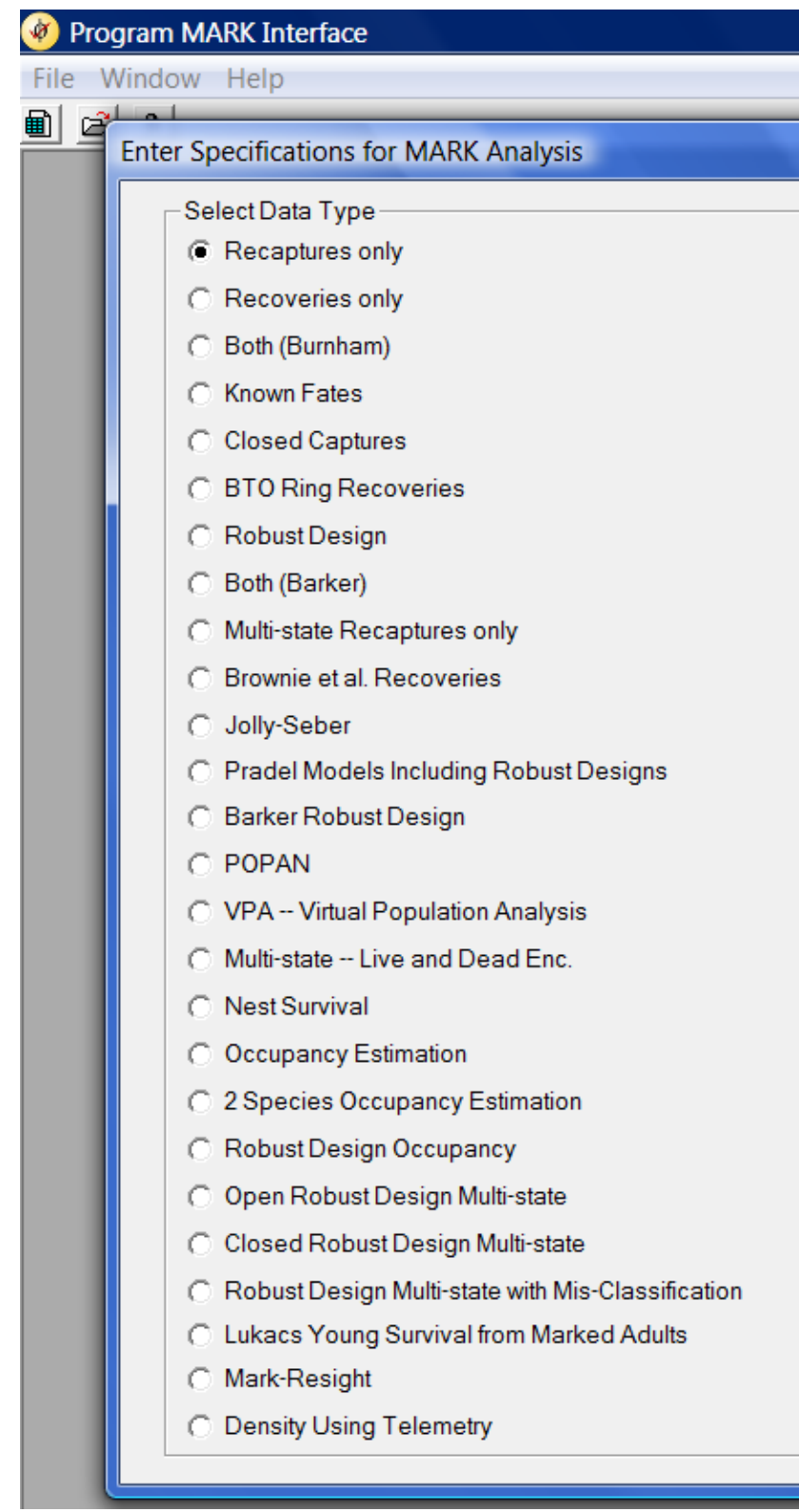
Softwares



MARK

www.phidot.org

Opções:



Preparando o arquivo



- Excel
 - concatenar células (função “**concatenate**”)
 - salvar como formato de texto (“txt”)
- Mudar extensão para “**.INP**”

```
Baleias_1989-2008 - Notepad
File Edit Format View Help
/* Baleia-jubarte Abrolhos, 1989-2008, 20 ocasioes */
1000000000001000000000 1;
10110000000000000011000 1;
100000000000000000000000 1;
1000001000010000000000 1;
100000000000000000000000 1;
1010000001100000000000 1;
100000000000000000000000 1;
1000010110000100000 1;
100000000000000000000000 1;
1000000000010000000000 1;
11000000001100011001 1;
```



Número de histórico de capturas

Matriz pode ser condensada de maneira a quantificar o número de histórico de capturas iguais, sem repetí-los na matriz

PIM / PIM Chart



Design matrix



Função de ligação



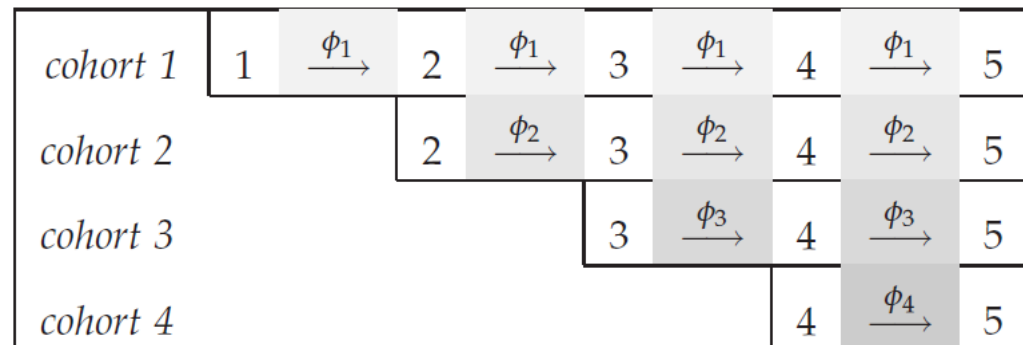
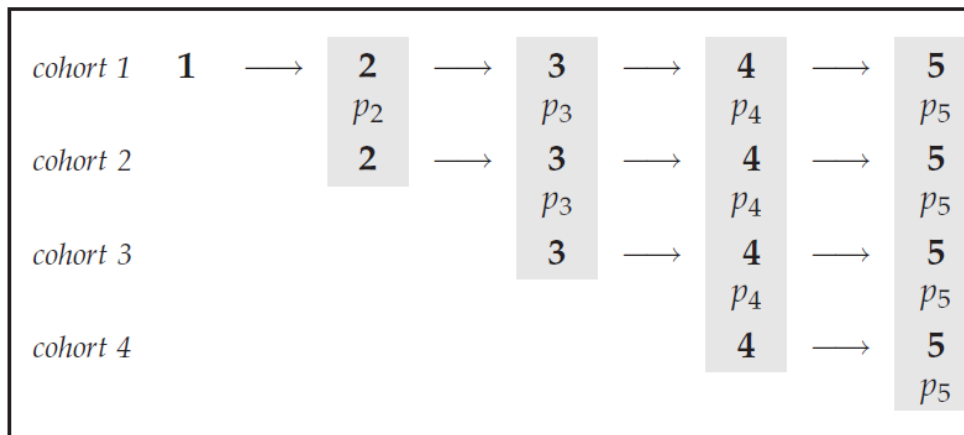
Parâmetro

PIM



- Indexando os parâmetros

Tempo 



Exercício 2.1



- European dipper (*Cinclus cinclus*)



Exercício 2.1



- Arquivo “Dipper_male.INP”
- Construir modelo tempo dependente usando a PIM
- Rodar estimativas / interpretar parâmetros
- Gerar gráficos de ϕ e p

A grande lei



Quanto maior a probabilidade de
captura, melhor...

Duas perguntas fundamentais:

(1) O modelo de ajustou bem aos dados?

(2) Qual o melhor modelo?

2. Testando as premissas do CJS

- 1) **Todo animal marcado em uma determinada ocasião de captura tem a mesma probabilidade de recaptura;**
- 2) **Todo animal marcado tem a mesma probabilidade de sobreviver entre uma ocasião e a próxima;**
- 3) Marcas não se modificam ou se perdem;
- 4) Períodos amostrais são curtos e instantâneos, animais marcados são soltos imediatamente;
- 5) Toda emigração da área de estudo é permanente;
- 6) O destino de uma animal com relação à sua captura e sobrevivência independe dos destinos dos outros animais.

Verificando as premissas

- Burnham *et al.* (1987: 65) descreve dois testes (2 e 3) de “bondade de ajuste” (*Goodness of fit* – GOF) baseados na distribuição Qui-quadrado rodados no programa Release dentro do MARK

O modelo se ajustou bem aos dados?



TESTE 2: computado a partir da matriz triangular, parâmetros são específicos de cada ocasião de captura. Testa homogeneidade de recapturas.

Verifica se a probabilidade de um indivíduo ser capturado na ocasião $(i + 1)$ é uma função do indivíduo ter sido capturado ou não na ocasião (i) , considerando que este indivíduo sobreviveu de (i) para $(i + 1)$.

O modelo se ajustou bem aos dados?

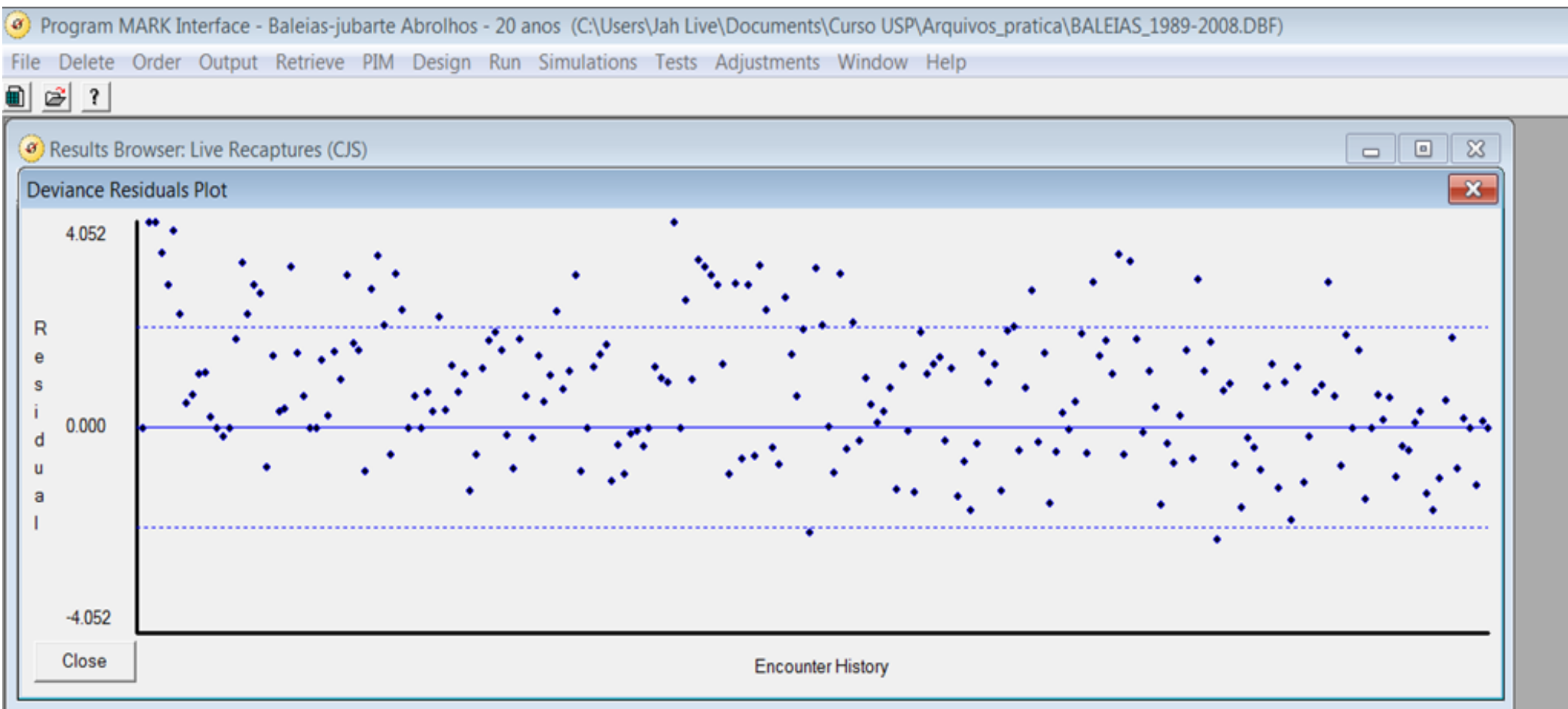
TESTE 3: parâmetros são dependentes dos históricos de indivíduos marcados em uma ocasião de captura específica. Testa homogeneidade de sobrevivência.

Dentro dos indivíduos capturados na ocasião (i), verifica quantos foram capturados novamente e quando.

Componentes do teste 3

- **Teste 3.SR:** verifica para cada ocasião de captura (ou coorte), se existe diferença entre os indivíduos novos marcados (na ocasião i) ou marcados em ocasiões anteriores serem capturados nas ocasiões subsequentes ($> i$).
- **Teste 3.Sm:** verifica se daqueles animais capturados novamente, quando eles foram capturados depende se eles foram marcados na ocasião (i) ou antes dela.

Análise dos desvios residuais



Exercício 2.2

Dipper:

- Rodar programa Release (dentro do Mark)
- Observar matriz triangular (*m-array*)
- Rodar testes 2 e 3
- Checar erros residuais

3. Seleção de modelos

- Nos métodos atuais grande ênfase é dada à seleção de modelos, sendo um aspecto crítico da análise de dados;
- Estimativa dos parâmetros assume uma menor importância relativa, enquanto identificar processos biológicos significantes na área de estudo comparando diferentes modelos assume uma maior importância.

Informação Kullback-Leibler

- S. Kullback & R. A. Leibler (1951) descreveram uma medida de distância entre a realidade conceitual (f) e um modelo que aproxima esta realidade (g) – distância Kullback-Leibler (K-L)
 - $I(f, g)$ = informação perdida quando um modelo é usado para aproximar a realidade ou distância entre o modelo e a realidade

H. Akaike (1973; 1974)

- Estabeleceu a relação entre a distância K-L e a máxima verossimilhança dentro de uma abordagem de otimização

$$\mathbf{AIC} = -2 \log(\mathcal{L}(\hat{\theta} \mid \mathbf{data})) + 2K.$$



Critério de Informação de Akaike

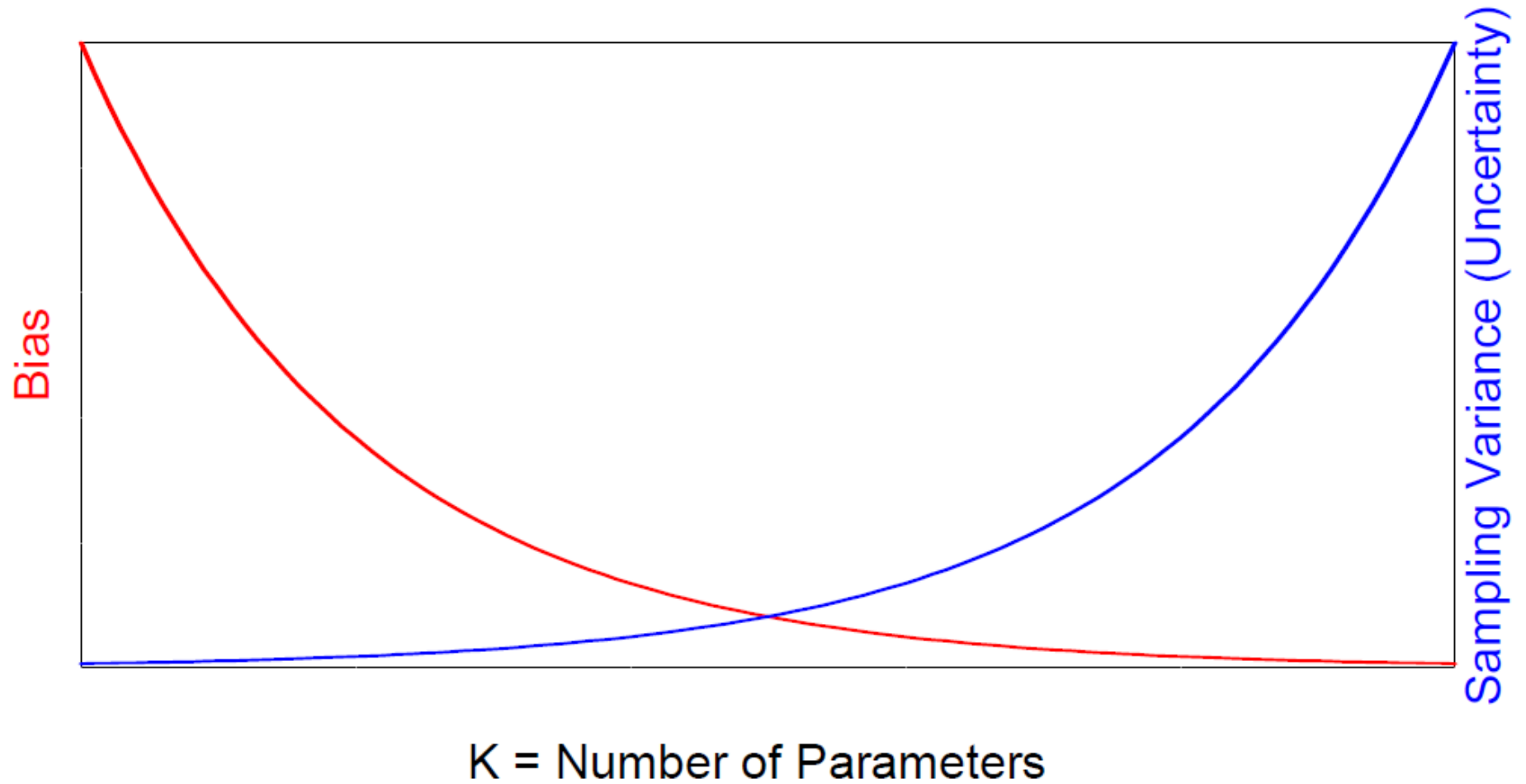
AIC – Critério de Informação de Akaike

$$\text{AIC} = -2 \log(\mathcal{L}(\hat{\theta} \mid \text{data})) + 2K.$$



Número de
parâmetros

Otimização



Seleção do melhor modelo



- Quanto menor o AIC, menor a distância entre o modelo e a realidade
- AIC pode ser computado para cada modelo
- Assim, devemos comparar e escolher o modelo com menor valor de AIC

Delta AIC

$$\Delta_i = \text{AIC}_i - \min \text{AIC}$$

- Valores de AIC re-escalados de forma que o modelo com menor AIC passe a ter $\text{AIC} = 0$
 - $\Delta_i \leq 2$: suporte substancial ao modelo
 - $\Delta_i = 4 - 7$: menor suporte considerável ao modelo
 - $\Delta_i > 10$: sem suporte

AICc



- Correção para modelos com amostra pequena relativa ao número de parâmetros
- Como AIC e AICc convergem com amostras grandes **AICc sempre deve ser usado**

Peso de Akaike (*Akaike weight*)

- Indica a força de evidência de um modelo
- Pode ser interpretado com a probabilidade daquele modelo ser o melhor dentre os modelos candidatos

$$w_i = \frac{\exp(-\frac{1}{2}\Delta_i)}{\sum_{r=1}^R \exp(-\frac{1}{2}\Delta_r)}$$

Sobredispersão



Pode acontecer e ser um problema quando trabalhamos com erros binomiais ou Poisson

Tende a ocorrer quando:

- Importantes variáveis explanatórias não foram incorporadas ao modelo;**
- Distribuição é diferente da binomial ou Poisson.**

Sobredispersão



Resulta em variância/precisão subestimada

Sobredispersão



Quantificando a sobredispersão



Fator de inflação da variância (VIF ou \hat{c})

Calculando c-hat

Forma mais simples através do GOF do modelo global:

$$\hat{c} = \chi^2 / df$$

Quasi-verossimilhança (*Quasi-likelihood*)

- QAIC (Quasi-AIC): tende a favorecer modelos com menos parâmetros

$$\text{QAIC}_c = - \left[2 \log(\mathcal{L}(\hat{\theta})) / \hat{c} \right] + 2K + \frac{2K(K+1)}{n-K-1}$$

Inflando a variância



- Infla-se a variância multiplicando-se pelo valor de $c\text{-hat}$;
- Alarga-se o intervalo de confiança multiplicando-se pela raiz quadrada de $c\text{-hat}$

Seleção do melhor modelo



Teste de razão de verossimilhança **(*Likelihood Ratio Test* - LRT)**

Teste de hipótese para modelos aninhados, ou seja, um modelo contém o outro, ou é um caso simplificado do outro

Teste de razão de verossimilhança (LRT)



$\ln L(\text{parâmetros de um modelo geral})$

vs.

$\ln L(\text{parâmetros de um modelo reduzido})$

Tamanho da diferença entre log-verossimilhanças
indica se o modelo reduzido deve ser preferido

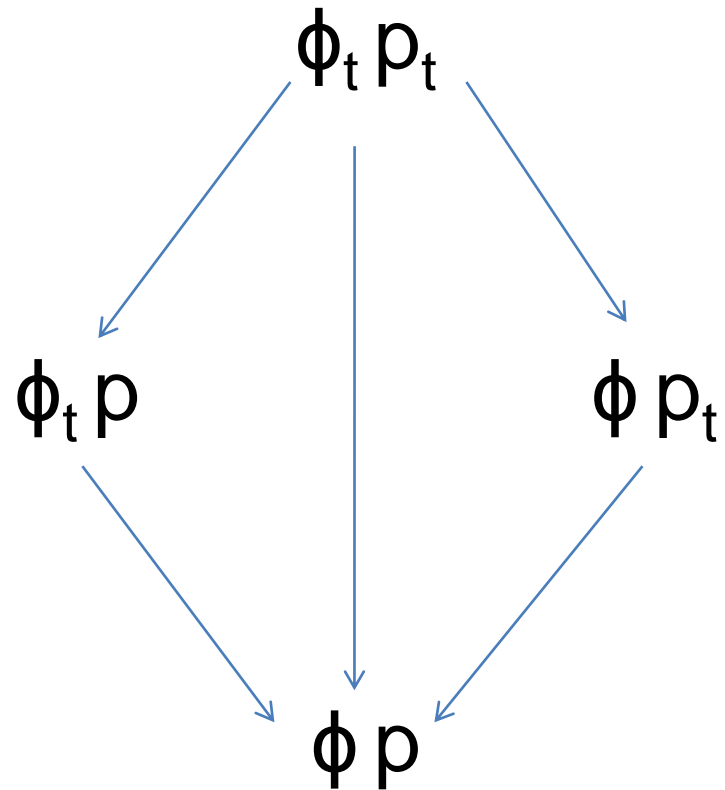
Teste de razão de verossimilhança (LRT)



Diferença possui distribuição de Qui-Quadrado com graus de liberdade igual à diferença de parâmetros entre os dois modelos

Diferença significativa indica que modelo geral deve ser preferido, enquanto diferença não significativa indica que o modelo mais simples deve ser escolhido

Teste de razão de verossimilhança (LRT)



Exercício 2.3

Dipper:

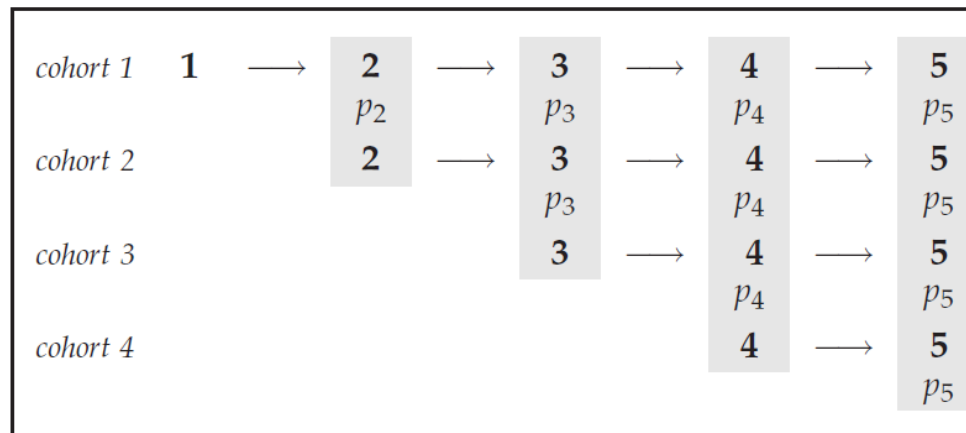
- Comparar AIC
- Rodar LRT entre modelos aninhados
- Comparar resultados AIC e LRT
- Calcular fator de inflação da variância (\hat{c})
- Corrigir modelos
- Adicione um efeito de enchente ocorrido no ano de 1983

4. Outros efeitos

- Efeitos de grupo
 - Sexo, colônias, etc...
- Efeito de coorte
- Efeito da idade

PIM

- Efeitos de grupo, coorte e idade podem ser facilmente determinados pela PIM



<i>cohort 1</i>	1	$\xrightarrow{\phi_1}$	2	$\xrightarrow{\phi_1}$	3	$\xrightarrow{\phi_1}$	4	$\xrightarrow{\phi_1}$	5
<i>cohort 2</i>			2	$\xrightarrow{\phi_2}$	3	$\xrightarrow{\phi_2}$	4	$\xrightarrow{\phi_2}$	5
<i>cohort 3</i>					3	$\xrightarrow{\phi_3}$	4	$\xrightarrow{\phi_3}$	5
<i>cohort 4</i>							4	$\xrightarrow{\phi_4}$	5

Efeito de grupo



- Duas PIMs: parâmetros duplicados para cada grupo

Efeito de coorte



- Grupo de indivíduos marcados pela primeira vez na mesma ocasião

PIM

1	1	1	1
	2	2	2
		3	3
			4

Efeito de corte * tempo

PIM

1	2	3	4
	5	6	7
		8	9
			10

Efeito da idade



- Primeira(s) diagonal(is) da PIM

1	2	2	2
	1	2	2
		1	2
			1

Exercício 2.4

- European dipper (*Cinclus cinclus*)
- 7 ocasiões nos anos de 1981 a 1987



Exercício 2.4



Usando os dados “**Dipper_sex.INP**”:

- Construa modelos com e sem efeitos do sexo nos parâmetros
- Construa modelos com e sem efeitos do tempo nos parâmetros
- Compare os modelos, escolha o melhor

Exercício 2.5

- Swift (*Apus apus*)



Exercício 2.5



Usando os dados “aa.INP”:

- Verifique as premissas e o ajuste do modelo global
- Construa modelos com e sem efeitos do sexo nos parâmetros
- Construa modelos com e sem efeitos do tempo nos parâmetros
- Compare os modelos, escolha o melhor

Leituras recomendadas!

Amstrup *et al.* (2005) para uma **revisão de métodos**

Cooch & White (2004): um guia completo sobre o **Mark** e seus modelos

Lebreton *et al.* (1992): texto clássico e obrigatório sobre os métodos de estimar a **sobrevivência**, com análises dos exercícios 2.4 e 2.5

Leituras recomendadas!

Burnham & Anderson (2002) sobre **AIC**, **sobredispersão** e **seleção de modelos**