

UNIVERSIDADE ESTADUAL DE MARINGÁ
PROGRAMA DE INICIAÇÃO CIENTÍFICA - PIC
DEPARTAMENTO DE ESTATÍSTICA
Orientador: Prof. Dr. George Lucas Moraes Pezzott
Acadêmico: Márcio Roger Piagio
Acadêmico: Raphael Amaral Luiz

**ESTIMAÇÃO POR MÁXIMA VEROSSIMILHANÇA DO TAMANHO
POPULACIONAL EM MODELOS DE CAPTURA-RECAPTURA**

UNIVERSIDADE ESTADUAL DE MARINGÁ
PROGRAMA DE INICIAÇÃO CIENTÍFICA - PIC
DEPARTAMENTO DE ESTATÍSTICA
Orientador: Prof. Dr. George Lucas Moraes Pezzott
Acadêmico: Márcio Roger Piagio
Acadêmico: Raphael Amaral Luiz

**ESTIMAÇÃO POR MÁXIMA VEROSSIMILHANÇA DO TAMANHO
POPULACIONAL EM MODELOS DE CAPTURA-RECAPTURA**

Relatório contendo os resultados finais do projeto
de iniciação científica vinculado ao Programa PIC-
UEM.

RESUMO

O presente projeto apresenta dois modelos de captura-recaptura: o modelo M_t , que considera probabilidades diferentes no instante de captura dos animais; e o modelo M_{tb} que leva em consideração possíveis diferenças entre as probabilidades de captura e recaptura. O processo inferencial foi definido por obter as estimativas de máxima verossimilhança dos parâmetros dos modelos e os resultados numéricos deste trabalho se basearam em duas aplicações com dados reais de captura-recaptura de ratos e um estudo de simulação em diferentes cenários para avaliar a performance dos estimadores de máxima verossimilhança. No geral, foi possível concluir que o método de captura-recaptura é extremamente válido como uma técnica de amostragem para se estimar o tamanho da população e também é robusto quando se tem pelo menos 60% da população observada na amostra pois seus estimadores apresentam baixo viés e erro quadrático médio.

Palavras-chave: modelo de captura-recaptura, heterogeneidade; efeito comportamental; estimativa de máxima verossimilhança

Conteúdo

1	Introdução	1
1.1	Objetivos:	2
2	Desenvolvimento (Materiais e Métodos)	5
2.1	Modelo de captura-recaptura com heterogeneidade	6
2.2	Modelo de captura-recaptura com heterogeneidade e efeito à marcação	8
2.3	Estimador de máxima verossimilhança	10
3	Resultados e Discussão	13
3.1	Aplicação com dados reais	13
3.1.1	Primeira aplicação	13
3.1.2	Segunda aplicação	15
3.2	Estudo de simulação	17
3.2.1	Modelo M_t	17
3.2.2	Modelo M_{tb}	19
4	Conclusões	25
	Referências	27

1 Introdução

Diversas áreas do conhecimento buscam conhecer o número de elementos em uma dada população. Na ecologia, por exemplo, o monitoramento do número de animais em uma determinada região é de suma importância no estudo e conservação da espécie (SEBER et al., 1982; MCCREA; MORGAN, 2014). Na área da saúde, é necessário avaliar a quantidade de indivíduos com uma determinada característica (doença, usuários de droga, ...) em uma cidade para promover políticas públicas e sociais (BIRD; KING, 2018; BÖHNING et al., 2020). Por outro lado, a confiabilidade de um *software* está relacionada ao número de erros (falhas) que ele apresenta (BASU; EBRAHIMI, 2001).

Entretanto, na maioria das aplicações é impraticável observar todos os elementos da população devido à dificuldade operacional como, dentre outros fatores, populações esquivas, excessivo tempo para execução ou alto custo financeiro. Nestes casos, procedimentos estatísticos inferenciais são necessários para obter estimativas para os parâmetros populacionais e o método de captura-recaptura (CR) mostra-se uma técnica de amostragem útil e robusta.

Resumidamente, o método de captura-recaptura consiste em selecionar elementos desta população em diferentes ocasiões de amostragem. Na primeira ocasião, uma amostra é retirada, os elementos capturados recebem uma marca e, em seguida, todos são devolvidos à população. Após um certo período de tempo, é selecionada uma segunda amostra e realizada a contagem dos elementos marcados (recapturas), e aqueles não marcados recebem uma marca, e todos são devolvidos à população. Este procedimento é repetido em k ($k \geq 2$) ocasiões de amostragem, e em cada ocasião é realizada a contagem do número de elementos selecionados e daqueles previamente marcados, feita a marcação dos elementos não marcados e todos são devolvidos à população. No final do processo faz-se a inferência sobre os parâmetros populacionais baseada no número de elementos capturados e recapturados (OTIS et al., 1978; RODRIGUES; BOLFARINE; GALVÃO LEITE, 1988; OLIVEIRA; SILVA, 2007; SALASAR, 2011; WANG; HE; SUN et al., 2015).

A utilização da técnica de captura-recaptura para estimar o tamanho de uma população pode ser melhor compreendida com um exemplo simples. Considere que N é o tamanho da população a ser estimada e n_1 é o total de animais capturados na primeira época de captura, sendo todos

marcados e devolvidos à população. Logo, nesta população, temos uma proporção n_1/N de animais marcados. Considere que na segunda época de captura tenhamos n_2 animais capturados, dos quais m estejam marcados. A ideia é estimar a proporção de marcados da população (n_1/N) pela proporção de marcados na segunda amostra (m/n_2), isto é,

$$\widehat{\left(\frac{n_1}{N}\right)} = \frac{m}{n_2}$$

onde, resolvendo-se em N , tem-se um estimador \hat{N} para o tamanho populacional, dado por

$$\hat{N} = \frac{n_1 n_2}{m}$$

Na literatura, este estimador é conhecido como estimador de Lincoln-Petersen, em referência aos primeiros pesquisadores a empregarem este método na ecologia, o dinamarquês Petersen (1896), em seu estudo sobre o fluxo migratório de peixes do mar Báltico e Lincoln (1930), ao estimar o tamanho da população de patos selvagens na América do Norte. Contudo, este método foi proposto inicialmente por Laplace (1783) para estimar o tamanho da população francesa.

Atualmente, diversos modelos de captura-recaptura são encontrados na literatura para as mais diversas aplicações. Embora tenha-se maior volume de trabalhos com aplicação na ecologia para a estimação de abundâncias de populações animais (MCCREA; MORGAN, 2014; ROYLE et al., 2013), o método pode ser aplicado em outras áreas do conhecimento. Citamos o trabalho de Polonsky et al. (2021) como um exemplo de aplicação na epidemiologia, onde se buscou estimar a prevalência de Ebola via integridade do rastreamento de contatos durante o surto na República Democrática do Congo, entre os anos de 2018 e 2020. Nas áreas de políticas públicas e sociais, Ryngelblum e Peres (2021) empregaram um estudo sobre a análise da qualidade dos dados das mortes cometidas por policiais no município de São Paulo, Brasil, entre os anos de 2014 e 2015. Outro exemplo é um estudo onde foi possível identificar a má conduta das instituições financeiras e seus funcionários no Reino Unido entre 2004 e 2016 apresentado por Ashton et al. (2021).

1.1 Objetivos:

O método de captura-recaptura vêm sendo aplicado nas mais diversas áreas do conhecimento e, por isso, o objetivo principal deste projeto é estudar essa metodologia de amostragem e a modelagem estatística para os dados provenientes desta técnica. Com relação à modelagem, temos os seguintes objetivos específicos para o projeto:

1. estudar dois tradicionais modelos estatísticos de captura-recaptura da literatura: M_t e M_{tb} (OTIS et al., 1978);
2. discutir métodos de estimação dos parâmetros desses modelos e aplicá-los em dados reais da literatura para exemplificar a metodologia;
3. apresentar um estudo de simulação para avaliar a performance dos estimadores.

2 Desenvolvimento (Materiais e Métodos)

Considere uma população com N indivíduos. Neste projeto, supomos que a população é "fechada" (demográfica e geograficamente), isto é, não há mortes (saídas ou perdas) nem nascimentos (entradas) de indivíduos na população ao longo do processo de captura-recaptura. Em outras palavras, assumimos que a população de estudo é composta pelos mesmos N indivíduos desde o início da primeira ocasião de captura até o final da última.

O processo de captura-marcação-recaptura de animais da população é aplicado da maneira tradicional: em cada ocasião de amostragem, observa-se um número aleatório (não fixado previamente) de animais desta população, registra-se o número de animais marcados e não marcados na amostra, marcando todos os não marcados e devolve-se todos (marcados e não marcados) à população. O processo é repetido em k ocasiões de amostragem. Supomos que não há animais marcados na população antes da primeira ocasião de amostragem, e que os animais não perdem suas marcas durante o processo.

Diante do exposto, definimos

- u_j : número de animais não marcados na j -ésima ocasião de amostragem, $j = 1, 2, \dots, k$; com $\mathbf{u} = (u_1, u_2, \dots, u_k)$
- m_j : número de animais marcados (recapturas) na j -ésima ocasião de amostragem, $j = 2, \dots, k$ com $\mathbf{m} = (m_2, m_3, \dots, m_k)$.
- n_j : número de animais selecionados (marcados e não marcados) na j -ésima ocasião de amostragem, isto é, $n_j = u_j + m_j$, $j = 1, 2, \dots, k$;
- M_j : número de animais marcados na população imediatamente antes da j -ésima ocasião de amostragem, $j = 1, \dots, k$, com $M_1 = 0$ e $M_{j+1} = M_j + u_j$, para $j = 2, 3, \dots, k$.

A seguir, apresentamos um exemplo.

Suponha uma população de tamanho $N = 100$. De fato, temos $M_1 = 0$ pois não temos animais marcados na população antes da primeira época de captura.

Primeira ocasião: *Suponha que capturamos $n_1 = 10$ animais. Consequentemente, temos $u_1 = 10$ animais não marcados na amostra e um total de $m_1 = 0$ animais marcados observados, isto é,*

nenhuma recaptura na primeira ocasião. Marcamos esses 10 animais e devolvemos à população.

Observa-se que agora temos $M_2 = M_1 + u_1 = 0 + 10 = 10$ animais marcados na população.

Segunda ocasião: Após certo tempo para que os animais devolvidos na primeira ocasião possam se misturar aos demais, colhemos uma segunda amostra. Suponha que agora observamos $n_2 = 15$ animais, dos quais $u_2 = 12$ não são marcados e $m_2 = 3$ são marcados, isto é temos três animais recapturados. Então marca-se os 12 não marcados, e todos são devolvidos à população. Neste momento, temos $M_3 = 10 + 12 = 22$ animais marcados na população.

Terceira ocasião: Na terceira ocasião de captura, suponha um total de $n_3 = 17$ animais capturados, sendo $u_3 = 8$ animais não marcados e $m_3 = 9$ marcados (observa-se que a marcação não permite identificar se o animal recapturado foi marcado na primeira ou segunda época de captura, embora poderia realizar uma marcação específica em cada animal para registrar tal histórico). Por fim, marcando os 8 animais e devolvendo todos à população, temos $M_3 = 22 + 8 = 30$ animais observados até o momento.

As quantidades estão apresentadas na Tabela 2.1 abaixo.

Tabela 2.1: Exemplo de quantidades observadas em um processo de captura-recaptura.

	j	1	2	3
Animais marcados antes da ocasião	M_j	0	10	22
Animais capturados (não marcados)	u_j	10	12	8
Animais recapturados (marcados)	m_j	0	3	9
Animais selecionados	n_j	10	15	17

Em três épocas de captura, foi possível observar 30 animais distintos, de um total de $N = 100$ animais. O presente exemplo se encerra comentando que o processo de captura-recaptura poderia ser estendido para mais épocas de captura.

A seguir, apresentamos dois modelos estatísticos para estimação do tamanho populacional utilizando as informações de capturas definidas acima.

2.1 Modelo de captura-recaptura com heterogeneidade

Nesta seção vamos apresentar o modelo de captura-recaptura com heterogeneidade temporal das probabilidades de captura, comumente denotado por M_t . Definimos p_j sendo a probabilidade de um animal (marcado ou não) ser capturado na j -ésima ocasião de amostragem, $j = 1, 2, \dots, k$ com $\mathbf{p} = (p_1, p_2, \dots, p_k)$. Supomos que os animais comportam-se independente uns dos outros, isto é, a captura (ou não) de um animal não altera a probabilidade de captura de qualquer outro. Adicionalmente,

supomos que a captura (ou não) de um animal não altera sua probabilidade de recaptura, ou seja, animais marcados e não marcados tem a mesma probabilidade p_j de serem capturados na ocasião j .

Note que, por se tratar de uma sequência de sucessos ou fracassos (captura o não captura) e independência das capturas e mesma probabilidade de sucesso (captura) para as quantidades u_j e m_j , temos as seguintes distribuições:

$$u_j|N, p_j, M_j \sim \text{Binomial}(N - M_j, p_j), \quad j = 1, 2, \dots, k \quad (2.1)$$

e

$$m_j|p_j, M_j \sim \text{Binomial}(M_j, p_j), \quad j = 2, \dots, k \quad (2.2)$$

Logo, a função de verossimilhança do modelo M_t para os parâmetros N e $\mathbf{p} = (p_1, \dots, p_k)$, dada uma amostra de captura-marcação-recaptura $\mathbf{u} = (u_1, u_2, \dots, u_k)$ e $\mathbf{m} = (m_2, m_3, \dots, m_k)$, é definida por

$$\begin{aligned} L(N, \mathbf{p}|\mathbf{u}, \mathbf{m}) &= p(\mathbf{u}, \mathbf{m}|N, \mathbf{p}) \\ &= p(u_1|N, p_1) \prod_{j=2}^k p(u_j|N, p_j, M_j) p(m_j|M_j, p_j) \\ &= \prod_{j=1}^k \binom{N - M_j}{u_j} p_j^{u_j} (1 - p_j)^{N - M_j - u_j} \times \prod_{j=2}^k \binom{M_j}{m_j} p_j^{m_j} (1 - p_j)^{M_j - m_j} \\ &= \prod_{j=2}^k \binom{M_j}{m_j} \times \prod_{j=1}^k \binom{N - M_j}{n_j - M_j} p_j^{n_j} (1 - p_j)^{N - n_j} \end{aligned}$$

para $N \geq r$, onde $r = M_k + u_k$ é o número de animais distintos observados no estudo, e $0 < p_j < 1$, para $j = 1, 2, \dots, k$, sendo $n_j = u_j + m_j$.

Observe que, como $M_{j+1} = M_j + u_j$, temos

$$\begin{aligned} \prod_{j=1}^k \binom{N - M_j}{u_j} &= \prod_{j=1}^k \frac{(N - M_j)!}{u_j! (N - M_j - u_j)!} \\ &= \prod_{j=1}^k \frac{1}{u_j!} \prod_{j=1}^k \frac{(N - M_j)!}{(N - M_{j+1})!} \\ &= \prod_{j=1}^k \frac{1}{u_j!} \times \frac{N!}{(N - r)!} \end{aligned}$$

Logo, a função de verossimilhança pode ser reescrita em termos proporcionais à

$$L(N, \mathbf{p} | \mathbf{u}, \mathbf{m}) \propto \frac{N!}{(N-r)!} \prod_{j=1}^k p_j^{n_j} (1-p_j)^{N-n_j}$$

para $N \geq r$ e $0 < p_j < 1$, para $j = 1, 2, \dots, k$,

2.2 Modelo de captura-recaptura com heterogeneidade e efeito à marcação

Nesta seção apresentamos o modelo captura-recaptura com heterogeneidade temporal das probabilidades de captura e com efeito da marcação. O modelo foi originalmente proposto por Pollock (1980), e na literatura é referido como o modelo M_{tb} (ver Otis et al. (1978)). O modelo M_{tb} pressupõe, para cada ocasião de amostragem, que todos os animais não marcados têm uma probabilidade de captura e todos os animais marcados têm outra probabilidade de captura, e assume que essas probabilidades variam de uma ocasião de amostragem para outra.

Seja c_j a probabilidade de recapturar na ocasião j um animal marcado. Neste caso, as seguintes distribuições condicionais são definidas para as quantidades m_j com essa nova probabilidade:

$$m_j | c_j, M_j \sim \text{Binomial}(M_j, c_j), \quad j = 2, 3, \dots, k \quad (2.2.1)$$

enquanto que as distribuições para u_j permanecem como previamente definidas na equação? Observe que a modelagem permite probabilidades diferentes entre marcados (p_j) e não marcados (c_j) em uma determinada ocasião j .

De maneira similar ao modelo M_t , temos a função de verossimilhança para $N, \mathbf{p} = (p_1, p_2, \dots, p_k)$ e $\mathbf{c} = (c_2, \dots, c_k)$ dada por

$$\begin{aligned} L(N, \mathbf{p}, \mathbf{c} | \mathbf{u}, \mathbf{m}) &= p(\mathbf{u}, \mathbf{m} | N, \mathbf{p}, \mathbf{c}) \\ &= \prod_{j=1}^k \binom{N-M_j}{u_j} p_j^{u_j} (1-p_j)^{N-M_j-u_j} \times \prod_{j=2}^k \binom{M_j}{m_j} c_j^{m_j} (1-c_j)^{M_j-m_j} \\ &\propto \frac{N!}{(N-r)!} \prod_{j=1}^k p_j^{u_j} (1-p_j)^{N-M_j+1} \prod_{j=2}^k c_j^{m_j} (1-c_j)^{M_j-m_j} \end{aligned}$$

para:

- $N \geq r$, $0 < p_j < 1$ para $j = 1, \dots, k$
- $0 < c_j < 1$ para $j = 2, \dots, k$

Contudo, o modelo acima não é identificável pois temos $2k - 1$ estatísticas suficientes em $\mathbf{u} = (u_1, u_2, \dots, u_k)$ e $\mathbf{m} = (m_2, m_3, \dots, m_k)$ para $2K$ parâmetros $(N, \mathbf{p}, \mathbf{c})$, isto é, tem-se mais parâmetros que dados. Portanto, restrições no espaço paramétrico são necessárias.

A restrição proposta por Wang (2002) considera a razão da chance de capturas e recapturas da seguinte forma:

$$\frac{c_j}{1 - c_j} = w \times \frac{p_j}{1 - p_j}, \quad j = 2, 3, \dots, k \quad (2.2.3)$$

onde existe três situações para a razão de chances w :

- $w > 1$: probabilidade de capturar um animal marcado é maior que a probabilidade de capturar um animal não marcado (efeito “feliz” da marcação);
- $w = 1$: probabilidade de capturar um animal marcado é igual à probabilidade de capturar um animal não marcado, (sem efeito à marcação);
- $w < 1$: probabilidade de capturar um animal marcado é menor que a probabilidade de capturar um animal não marcado, (efeito “triste” à marcação).

O efeito “feliz” à marcação pode ser gerado por agrados durante a captura, como petiscos. O efeito “triste” pode ser derivado por possíveis situações de desconforto durante a marcação como, por exemplo, *tags* e pulseiras apertadas.

Consequente, consideramos as seguintes transformações:

$$c = \log(w) \quad (2.2.4)$$

$$\eta_{j0} = \log\left(\frac{p_j}{1 - p_j}\right), \quad j = 1, 2, \dots, k \quad (2.2.5)$$

$$\eta_{j1} = \log\left(\frac{c_j}{1 - c_j}\right), \quad j = 2, 3, \dots, k \quad (2.2.6)$$

com $\eta_{j0} \in (-\infty, \infty)$, $\eta_{j1} \in (-\infty, \infty)$ e $c \in (-\infty, \infty)$.

Logo, temos a seguinte relação entre os parâmetros:

$$\eta_{j1} = c + \eta_{j0}, \quad j = 2, 3, \dots, k$$

diminuindo o espaço paramétrico para $k + 2$ parâmetros: $(N, \boldsymbol{\eta} = (\eta_{10}, \eta_{20}, \dots, \eta_{k0}), c)$. Portanto,

se $k \geq 3$, temos que a quantidade de estatísticas suficientes será pelo menos igual ao número de parâmetros e, conseqüentemente, garantimos a identificabilidade do modelo.

Como $p_j = e^{\eta_{j0}} / (1 + e^{\eta_{j0}})$ e $c_j = e^{c+\eta_{j0}} / (1 + e^{c+\eta_{j0}})$, podemos reescrever a função de verossimilhança para os novos parâmetros da seguinte forma:

$$L(N, \boldsymbol{\eta}, c | \mathbf{u}, \mathbf{m}) \propto \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{e^{\eta_{j0}n_j + cm_j}}{(1 + e^{\eta_{j0}})^{N-M_j} (1 + e^{c+\eta_{j0}})^{M_j}}$$

para $N \geq r$, $\eta_{j0} \in (-\infty, \infty)$, $j = 1, 2, \dots, k$, e $c \in (-\infty, \infty)$.

2.3 Estimador de máxima verossimilhança

Seja $\boldsymbol{\theta}$ um vetor de $d \geq 1$ parâmetros desconhecidos presentes em uma função de verossimilhança $L(\boldsymbol{\theta} | \mathbf{x})$ relacionado a uma amostra observada \mathbf{x} . A estimativa de máxima verossimilhança de $\boldsymbol{\theta}$, denotada por $\hat{\boldsymbol{\theta}}$, é o valor de $\boldsymbol{\theta}$ que maximiza a função de verossimilhança L , isto é,

$$\hat{\boldsymbol{\theta}} = \arg \max L(\boldsymbol{\theta} | \mathbf{x})$$

onde a maximização é feita sobre todo o espaço paramétrico de $\boldsymbol{\theta}$ (CASELLA; BERGER, 2021).

Em problemas envolvendo mais de um parâmetro e/ou funções de verossimilhanças complexas, em geral, não se obtém uma forma analítica para os estimadores de máxima verossimilhança, recorrendo-se a métodos iterativos de otimização/maximização como, por exemplo, o algoritmo de Newton–Raphson. Neste trabalho, utilizamos a função `optim`¹ que emprega o algoritmo L-BFGS-B² do software R Core Team (2022) como auxílio computacional para encontrar as estimativas de máxima verossimilhança dos modelos apresentados. O L-BFGS-B é um algoritmo de memória limitada para resolver grandes problemas de otimização não linear sujeitos a limites simples nas variáveis. Como o parâmetro N é um valor inteiro, aproximamos a estimativa dada pela função `optim` para o inteiro mais próximo.

A função de verossimilhança utilizada para maximização segundo o modelo M_t é exibida abaixo:

```
1 # --- Funcao de verossimilhanca: Modelo Mt
2 Likelihood_function_Mt <- function(parameters){
3   N = parameters[1]
4   eta = parameters[2:(K+1)]
5   log.like = lgamma(N+1) - lgamma(N - r+1) + sum(eta*n) -
6     sum((N-M)*log(1 + exp(eta))) - sum(M*log(1 + exp(eta)))
```

¹Disponível em: <https://www.rdocumentation.org/>

²Disponível em: <https://dl.acm.org/doi/>


```

7   return(-log.likelihood)
8 }

```

A função de verossimilhança utilizada para maximização segundo o modelo M_{tb} é exibida a seguir. Observa-se que a função de verossimilhança de M_t foi reparametrizada nos parâmetros η_j do modelo M_{tb} pois tais parâmetros não possuem restrições no espaço paramétrico sendo que, com a propriedade da invariância dos estimadores de máxima verossimilhança, podemos sem problemas obter posteriormente as estimativas das probabilidades de captura p_j , assim como feito para o modelo M_{tb} .

```

1  # --- Funcao de verossimilhanca: Modelo Mtb
2  Likelihood_function_Mtb <- function(parameters){
3    N = parameters[1]
4    eta = parameters[2:(K+1)]
5    c = parameters[K+2]
6    log.likelihood = lgamma(N+1) - lgamma(N - r+1) + sum(eta*n) + c*sum(m) -
7      sum((N-M)*log(1 + exp(eta))) -
8      sum(M*log(1 + exp(c + eta)))
9    return(-log.likelihood)
10 }

```


3 Resultados e Discussão

Neste capítulo apresentaremos resultados numéricos para a metodologia estudada. Dividimos o capítulo em duas partes: na primeira parte, ilustraremos a metodologia em duas aplicações com dados reais de captura-recaptura; e na segunda parte realizaremos um estudo de simulação Monte Carlo para verificar o comportamento dos estimadores de máxima verossimilhança – com foco no estimador do parâmetro N – dos dois modelos apresentados, M_t e M_{tb} .

3.1 Aplicação com dados reais

Nesta seção vamos aplicar os dois modelos de captura-recaptura estudados, M_t e M_{tb} , em dois conjuntos de dados reais da literatura. Para fins comparativos, iremos escolher como melhor ajuste aos dados o modelo que apresentar os menores valores dos critérios de informação AIC (Critério de Informação de Akaike) e BIC (Critério de Informação Bayesiano), definidos respectivamente por

$$\text{AIC} = -2 \log L(\hat{\boldsymbol{\theta}}|\mathbf{x}) - 2d$$

$$\text{BIC} = -2 \log L(\hat{\boldsymbol{\theta}}|\mathbf{x}) - d \log(n)$$

onde $\hat{\boldsymbol{\theta}}$ é a estimativa de máxima verossimilhança do vetor paramétrico $\boldsymbol{\theta}$ de dimensão d , e \mathbf{x} é o conjunto de dados contendo n observações (BURNHAM; WHITE; ANDERSON, 1995). Em nossos modelos, n é o número total de distintos capturados, denotado neste projeto por r .

3.1.1 Primeira aplicação

Como parte de um estudo de ciências populacionais, Coulombe, Ridgway e Evans (1965) conduziram um estudo com armadilhas para ratos domésticos selvagens *Mus musculus*. Um quadrado 10×10 foi usado com armadilhas espaçadas de 3 metros entre elas e as armadilhas foram verificadas pelos pesquisadores duas vezes ao dia, de manhã e à noite, durante 5 dias, totalizando $k = 10$ épocas de captura. Um total de $r = 173$ ratos diferentes foram capturados. Os dados estão disponíveis em Otis et al. (1978) e são exibidos na Tabela 3.1.

Tabela 3.1: Dados reais de captura-recaptura de ratos (OTIS et al., 1978).

	j	1	2	3	4	5	6	7	8	9	10
Animais marcados antes da ocasião	M_j	0	68	102	128	140	156	159	171	171	173
Animais capturados (não marcados)	u_j	68	34	26	12	16	3	12	0	2	0
Animais recapturados (marcados)	m_j	0	27	36	40	58	38	64	36	74	39
Animais selecionados	n_j	68	61	62	52	74	41	76	36	76	39

Observe a baixa frequência de animais não marcados (u_j) nas últimas ocasiões de captura, indicando que poucos animais da população ainda não foram vistos no estudo. Os tamanhos amostrais (n_j) diferentes também indicam heterogeneidade ao longo do tempo.

A Tabela 3.2 a seguir exhibe as estimativas dos dois modelos, M_t e M_{tb} .

Tabela 3.2: Estimativas de máxima verossimilhança dos parâmetros dos modelos M_t e M_{tb} na Aplicação 1.

Estimativas	M_t	M_{tb}
\hat{N}	175	176
\hat{c}	-	0.0858
\hat{p}_1	0.3886	0.3864
\hat{p}_2	0.3486	0.3391
\hat{p}_3	0.3543	0.3410
\hat{p}_4	0.2971	0.2826
\hat{p}_5	0.4229	0.4039
\hat{p}_6	0.2343	0.2196
\hat{p}_7	0.4343	0.4129
\hat{p}_8	0.2057	0.1913
\hat{p}_9	0.4343	0.4115
\hat{p}_{10}	0.2229	0.2074

Primeiramente, destacamos que as estimativas para o tamanho populacional dos dois modelos foram bem próximas, 175 animais para M_t e 176 para M_{tb} , e não muito superiores ao número total de animais distintos capturados, $r = 173$ animais. Isto é, estima-se que apenas dois ou três animais não foram vistos durante o estudo de acordo com os modelos M_t e M_{tb} , respectivamente. A heterogeneidade entre as probabilidades de captura é observada nas estimativas dos parâmetros p_j , que variaram entre 0,2057 e 0,4343 para o modelo M_t e de 0,1913 a 0,4129 no modelo M_{tb} . O parâmetro de efeito comportamental à marcação foi estimado em $\hat{c} = 0,0858$, bem próximo de zero que indica não efeito da marcação na chance de recapturar o animal. Essa evidência foi confirmada pelos critérios AIC e BIC exibidos na Tabela 3.3 onde escolhemos o modelo M_t como sendo de melhor ajuste por apresentar menores valores dos critérios.

Para fins ilustrativos, reajustamos os dados considerando somente as $k = 5$ primeiras épocas de captura, o que totaliza $r = 156$ animais capturados. A Tabela 3.4 exhibe as novas estimativas dos

Tabela 3.3: Critérios AIC e BIC para os modelos M_t e M_{tb} na Aplicação 1.

Critério	M_t	M_{tb}
AIC	112.74	114.55
BIC	147.42	152.39

dois modelos, M_t e M_{tb} , de acordo com esses dados restritos. Observe que a estimativa de N para o modelo M_t se manteve consistente com $\hat{N} = 173$, enquanto que o modelo M_{tb} elevou a estimativa para $\hat{N} = 184$. Os critérios AIC e BIC neste caso foram de 57.20 e 75.50 para o modelo M_t e 58.55 e 79.90 para o modelo M_{tb} , ainda favorecendo o modelo M_t para ajuste.

Tabela 3.4: Estimativas de máxima verossimilhança dos parâmetros dos modelos M_t e M_{tb} na Aplicação 1 ($k = 5$).

Estimativas	M_t	M_{tb}
\hat{N}	173	184
\hat{c}	-	0.3298
\hat{p}_1	0.3886	0.3864
\hat{p}_2	0.3526	0.3042
\hat{p}_3	0.3584	0.2965
\hat{p}_4	0.3006	0.2376
\hat{p}_5	0.4277	0.3431

3.1.2 Segunda aplicação

Nesta seção vamos utilizar dados reais de captura-recaptura de ratos veados que foram coletados por S. Hoffman em meados de julho de 1974, inicialmente utilizado por Otis et al. (1978) e discutidos no contexto bayesiano por Wang, He, Sun et al. (2015). Os dados foram coletados da seguinte forma: a captura foi em 5 manhãs consecutivas e 110 ratos distintos foram capturados, onde presumiu-se que exista resposta comportamental à marcação. Os resultados são apresentados na Tabela 3.5.

Tabela 3.5: Dados reais de captura-recaptura de ratos veados (WANG; HE; SUN et al., 2015)

	j	1	2	3	4	5
Animais marcados antes da ocasião	M_j	0	37	68	77	98
Animais capturados (não marcados)	u_j	37	31	9	21	12
Animais recapturados (marcados)	m_j	0	23	49	44	57
Animais selecionados	n_j	37	54	58	65	69

A Tabela 3.6 exibe as estimativas de máxima verossimilhança dos parâmetros, onde nota-se a diferença significativa entre as estimativas para $\hat{N} = 113$ para o modelo M_t e de $\hat{N} = 215$ para o modelo M_{tb} . De fato, o efeito comportamental à marcação nesta aplicação foi de $\hat{c} = 2,5195$,

considerada alta, o que pode estar influenciando as estimativas dos demais parâmetros entre os dois ajustes.

Tabela 3.6: Estimativas de máxima verossimilhança dos parâmetros dos modelos M_t e M_{tb} na Aplicação 2.

Estimativas	M_t	M_{tb}
\hat{N}	113	215
\hat{c}	-	2.5196
\hat{p}_1	0.3274	0.1721
\hat{p}_2	0.4779	0.1579
\hat{p}_3	0.5133	0.1121
\hat{p}_4	0.5752	0.1201
\hat{p}_5	0.6106	0.1012

De fato, temos que a probabilidade de capturar um animal não marcado (sendo a mesma para o animal marcado no modelo M_t) na j -ésima ocasião é:

$$p_j = \frac{e^{\eta_j}}{1 + e^{\eta_j}}, \quad j = 1, 2, \dots, k$$

e que a probabilidade de capturar um animal marcado pelo modelo M_{tb} na j -ésima ocasião é:

$$c_j = \frac{e^{c+\eta_j}}{1 + e^{c+\eta_j}}, \quad j = 2, 3, \dots, k.$$

Logo, pela propriedade da invariância dos estimadores de máxima verossimilhança, temos pelo modelo M_{tb} que as probabilidades estimadas de recapturas são $\hat{c}_2 = 0.6996$, $\hat{c}_3 = 0.6106$, $\hat{c}_4 = 0.6290$ e $\hat{c}_5 = 0.5831$, muito superiores que os \hat{p}_j exibidos na tabela anterior. De fato, temos que a razão de chances está de $\hat{w} = \exp(\hat{c}) = \exp(2.5196) = 12,4236$ de recaptura para a probabilidade de captura, evidenciando um forte efeito (“feliz”) no comportamento do animal devido à marcação. Isso é consistente com os dados, uma vez que há mais recapturas do que as capturas iniciais na maioria das ocasiões de amostragem.

O efeito significativo no parâmetro comportamental no ajuste final é descrito também pelos critérios AIC e BIC apresentados na Tabela 3.7, onde observamos menores valores de ambos critérios para o modelo M_{tb} e, assim, sendo escolhido como melhor ajuste aos dados da Tabela 3.5.

Tabela 3.7: Estimativas de máxima verossimilhança dos parâmetros dos modelos M_t e M_{tb} na Aplicação 2.

Critério	M_t	M_{tb}
AIC	89.10	67.37
BIC	105.31	86.28

3.2 Estudo de simulação

Para estudar o comportamento dos estimadores de máxima verossimilhança, utilizaremos as medidas de desempenho: Viés, Viés Relativo e Erro Quadrático Médio das estimativas obtidas dos parâmetros para um número M de amostras de captura-recaptura simuladas.

Genericamente, suponha que temos um parâmetro θ e para cada uma das M amostras geradas, temos uma estimativa de máxima verossimilhança $\hat{\theta}_m$ para θ , $m = 1, 2, \dots, M$. Suponha que o verdadeiro valor de θ usado na geração das amostras seja $\theta_{\text{verd.}}$. Assim, definimos o Viés, Viés Relativo e Erro Quadrático Médio, respectivamente, por

$$\text{Viés} = \frac{1}{M} \sum_{m=1}^M \hat{\theta}_m - \theta_{\text{verd.}} \quad (3.1.1)$$

$$\text{ViésR} = \frac{1}{M} \sum_{m=1}^M \frac{\hat{\theta}_m - \theta_{\text{verd.}}}{\theta_{\text{verd.}}} \quad (3.1.2)$$

$$\text{EQM} = \frac{1}{M} \sum_{m=1}^M \left(\hat{\theta}_m - \theta_{\text{verd.}} \right)^2 \quad (3.1.3)$$

3.2.1 Modelo M_t

Para o estudo do comportamento do estimador de máxima verossimilhança do parâmetro N no modelo M_t , fixamos o verdadeiro tamanho populacional nas simulações em $N = 100, 300$ e 500 e consideramos números diferentes de épocas de capturas, sendo $k = 5, 6, \dots, 15$. Para os parâmetros p_j , consideramos quatro cenários diferentes na geração destes valores, ilustrados na Tabela 3.8. A ideia é avaliar as estimativas variando cenários onde a probabilidade de captura é baixa (P1) até uma probabilidade de captura relativamente alta (P4).

Tabela 3.8: Cenários para geração das probabilidades de captura p_j .

Cenário	Distribuição
P_1	$p_1, \dots, p_k \sim \text{Uniforme}(0, 05; 0, 15)$
P_2	$p_2, \dots, p_k \sim \text{Uniforme}(0, 10; 0, 20)$
P_3	$p_3, \dots, p_k \sim \text{Uniforme}(0, 15; 0, 25)$
P_4	$p_3, \dots, p_k \sim \text{Uniforme}(0, 20; 0, 30)$

Observe que, com a combinação dos diferentes tamanhos populacionais N , épocas de capturas k e probabilidades de captura p_j 's, temos um total de $3 \times 11 \times 4 = 132$ diferentes cenários para avaliarmos no estudo de simulação. Para cada um dos 132 cenários, foram geradas 10.000 amostras de captura-recaptura.

A função implementada em código R para geração dos dados é exibida a seguir. Observe que

3 Resultados e Discussão

a função é genérica para ambos os modelos (M_t e M_{tb}) sendo que, quando estivermos gerando dados do modelo M_t , basta fixar $c = 0$ durante a geração dos dados.

```
1 # --- Funcao para gerar as amostras :
2
3 gera_dados <- function(N,K,pj,c,seed){
4   set.seed(seed)
5   u = m = n = M = c()
6   etaj = log(pj/(1-pj))
7   cj = exp(etaj+c)/(1+exp(etaj+c)) # probabilidade de recaptura
8   u[1] = rbinom(1,N,pj[1])          # numero de animais capturados
9   m[1] = 0                          # numero de animais marcados
10  n[1] = u[1]+m[1]                  # numero de animais selecionados
11  M[1] = 0                          # numero de animais marcados na populacao
12  # segunda ocasio em diante:
13  for(i in 2:K){
14    M[i] = M[i-1] + u[i-1]
15    u[i] = rbinom(1,N-M[i],pj[i])
16    m[i] = rbinom(1,M[i],cj[i])
17    n[i] = u[i] + m[i]
18  }
19  r = M[K] + u[K]
20  return(list(u,m,M,n,r))
21 }
```

Na Tabela 3.9 temos a relação em porcentagem do número médio de animais distintos observados dividido pelo tamanho da população em cada um dos cenários. Por exemplo, no cenário com $k = 5$, $N = 100$ e Cenário P_1 , temos que, em média, o número de animais distintos observados nas 10.000 amostras está em torno de 41,4% do total da população, isto é, 41,4 animais em $N = 100$. Já para o caso de $k = 15$, $N = 500$ e Cenário P_4 , captura-se em torno de 98,7% dos animais (média de 493,5 animais), isto é, quase todos eles são vistos durante o estudo. Em geral, observamos que com o crescimento de k ou cenários onde a probabilidade de captura é maior, tem-se em média mais animais capturados, como esperado.

Primeiramente, vale ressaltar que nem todas as amostras resultaram em estimativas válidas para os parâmetros. Especificamente, quando $N = 100$ e Cenário P_1 , obteve-se um total de 42 amostras sem estimativas quando $k = 5$ e cinco amostras sem estimativas quando $K = 6$, número relativamente baixo se comparado ao total de amostras simuladas (10.000). É conhecido em modelos de captura-recaptura que quando não se observa animais marcados em nenhuma época, a estimativa de máxima verossimilhança de N é infinito, ou extremamente elevada quando o número de recapturados

Tabela 3.9: Relação em porcentagem do número médio de animais distintos observados dividido pelo tamanho da população no primeiro estudo de simulação.

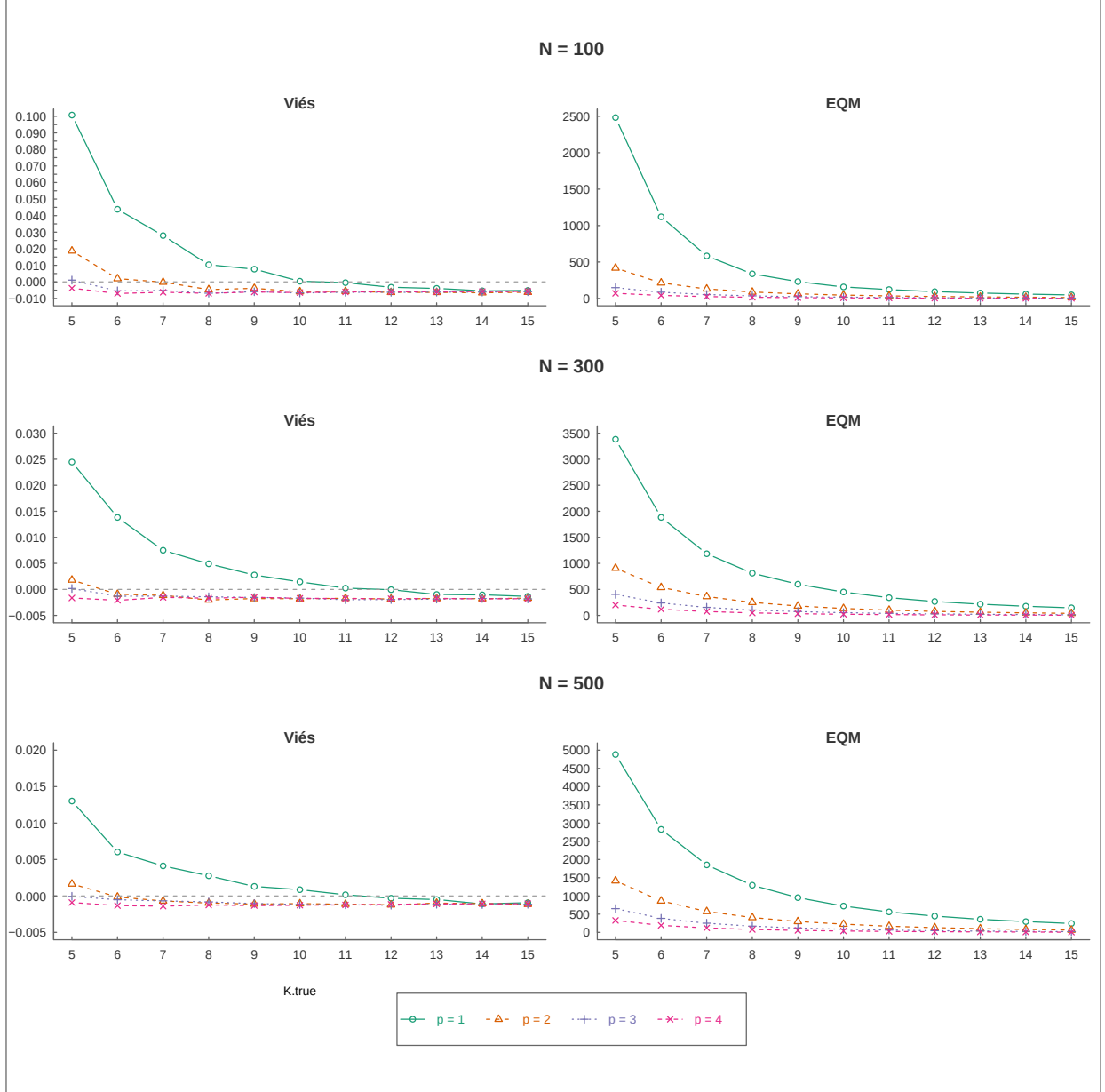
k	Cenário P_1			Cenário P_2			Cenário P_3			Cenário P_4		
	$N = 100$	300	500	100	300	500	100	300	500	100	300	500
5	41.4	40.5	40.9	55.8	55.6	55.6	67.3	67.3	67.3	76.3	76.3	76.3
6	47.2	46.5	46.8	62.4	62.3	62.3	73.8	73.8	73.8	82.1	82.2	82.2
7	52.5	51.8	52.1	68.0	67.9	67.9	79.0	79.0	79.1	86.6	86.7	86.7
8	57.1	56.7	56.9	72.8	72.7	72.7	83.2	83.2	83.2	89.9	90.0	90.0
9	61.4	61.0	61.2	76.9	76.8	76.8	86.5	86.6	86.6	92.4	92.5	92.5
10	65.2	64.9	65.0	80.3	80.3	80.3	89.2	89.3	89.3	94.3	94.4	94.4
11	68.7	68.4	68.5	83.2	83.3	83.3	91.4	91.4	91.4	95.7	95.8	95.8
12	71.8	71.5	71.7	85.7	85.8	85.8	93.1	93.1	93.1	96.8	96.8	96.8
13	74.6	74.3	74.5	87.9	87.9	87.9	94.4	94.5	94.5	97.6	97.6	97.6
14	77.1	76.9	77.0	89.7	89.7	89.7	95.5	95.6	95.6	98.2	98.2	98.2
15	79.4	79.2	79.4	91.2	91.3	91.3	96.4	96.5	96.5	98.6	98.7	98.7

é muito baixo que, em geral, leva a não convergência dos processos de maximização. Geralmente, baixos números de animais recapturados são observados em cenários com N e k pequenos e baixa probabilidade de captura, em concordância com os cenários observados sem estimativas no nosso estudo de simulação. De fato, as medidas de desempenho foram calculadas utilizando somente as estimativas válidas para os parâmetros.

Discutemos em sequência o comportamento do estimador de máxima verossimilhança de N . A Figura 3.1 abaixo apresenta os Viéses Relativo e o EQM para as estimativas de N do modelo M_t . Visualmente, observamos que conforme aumentam o número de épocas de captura, a estimação do tamanho da população torna-se mais precisa no sentido que diminui o viés relativo e o erro quadrático médio. Como esperado, o Cenário P_1 acarreta em viéses e EQM's maiores se comparado aos demais cenários pois tem probabilidades de captura p_j mais baixas, variando de 0,05 a 0,15. Em geral, podemos assumir uma boa precisão das estimativas para $k \geq 10$ para todos os cenários, até mesmo para o cenário P_1 , o que corresponde aproximadamente pelo menos 60% dos animais capturados da população (veja Tabela 3.9).

3.2.2 Modelo M_{tb}

Com o propósito de analisar o comportamento das estimativas dos parâmetros N e c para o modelo M_{tb} , foi realizado um estudo de simulação com 10000 amostras simuladas para cada valor verdadeiro fixado do parâmetro de efeito comportamental, $c = (-1.5, -1.0, -0.5, 0, 0.5, 1.0, 1.5)$, número de épocas de captura, $k = (5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15)$, e cada tamanho populacional, $N = (100, 300, 500)$. Fixamos a geração dos $p_j \sim \text{Uniforme}(0, 10; 0, 20)$ para todos os cenários acima. Considerando a combinação das quantidades acima, temos $3 \times 7 \times 11 = 231$ cenários diferentes.

Figura 3.1: Medidas de desempenho das estimativas do tamanho da população, N para o modelo M_t .

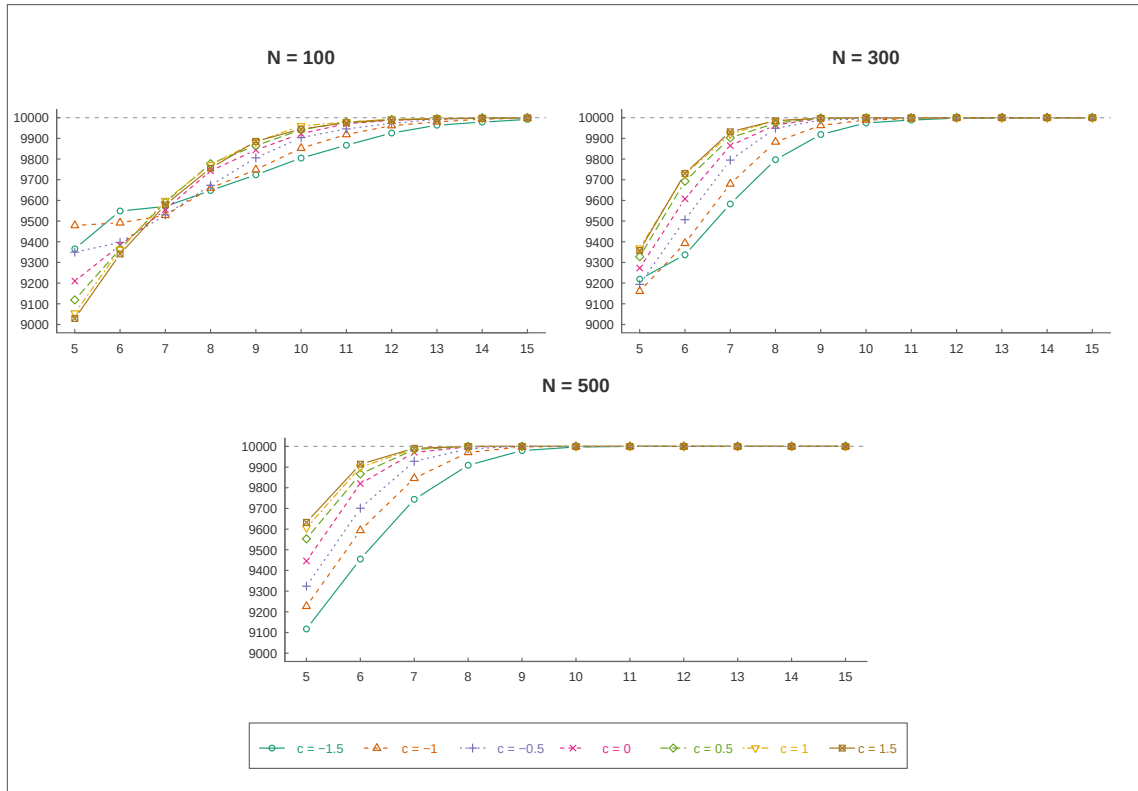
A Tabela 3.10 exibe a porcentagem média de animais distintos observados em cada cenário. Note que são bem parecidos ao Cenário P_2 do estudo de simulação anterior, pois utilizamos a mesma variação para os p_j . Outro fato que observamos é que não há diferença no número de animais distintos observados de acordo com o parâmetro c e por isso exibimos na tabela apenas quatro cenários diferentes para c . De fato, esse parâmetro modifica a probabilidade de capturar os animais marcados, que não influencia a probabilidade de capturar novos animais.

A Figura 3.2 exibe a quantidade de estimativas que convergiram para cada cenário. Diferente do que se observou no primeiro estudo de simulação, o modelo M_{tb} apresentou maior proporção de amostras que não tiveram estimativas válidas, principalmente para valores de k pequenos. Não se observou diferença aparente para os diferentes valores do parâmetro N . Contudo, se observou no

Tabela 3.10: Relação em porcentagem do número médio de animais distintos observados dividido pelo tamanho da população no segundo estudo de simulação.

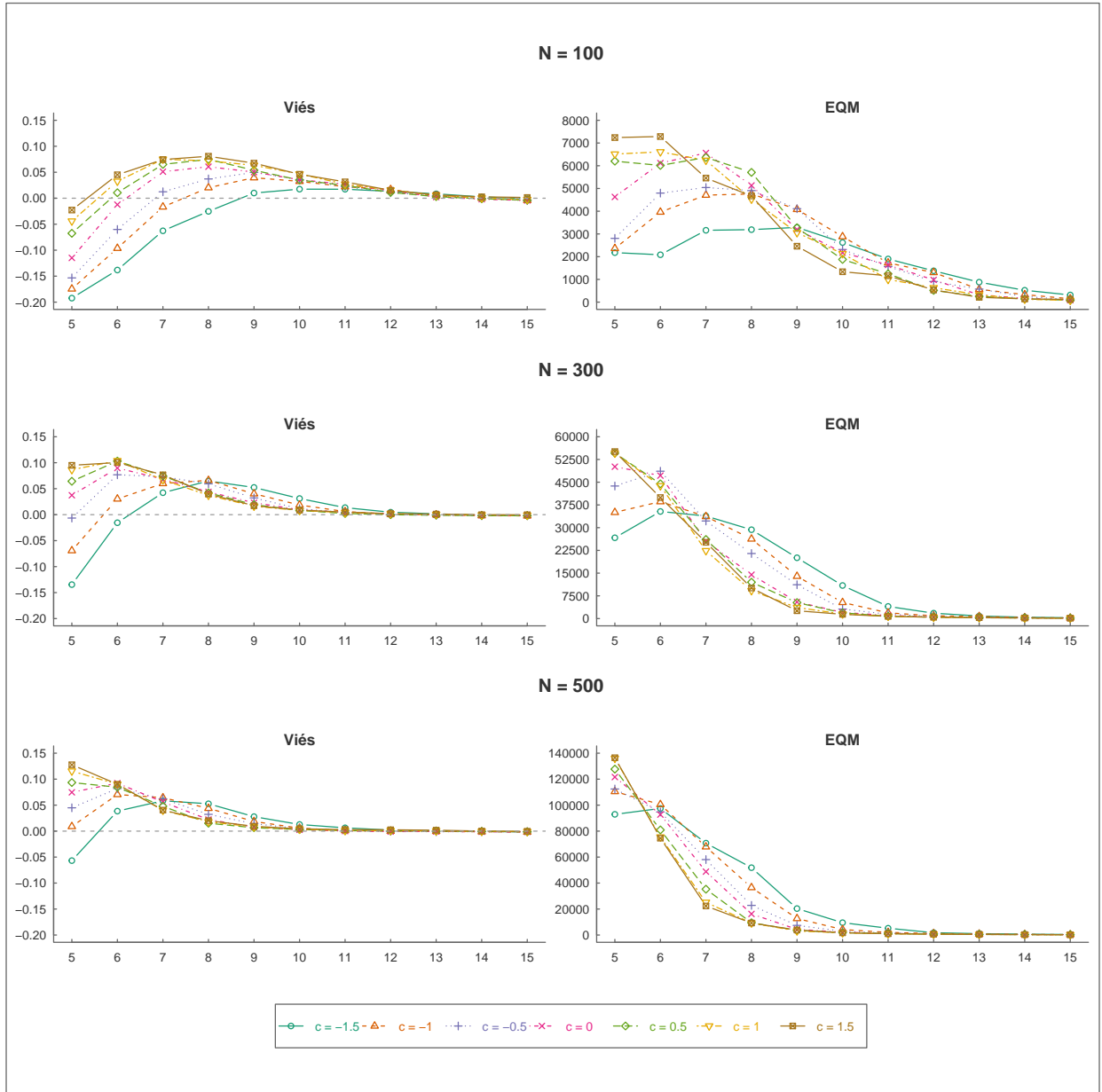
k	$c = -1,5$			$c = -0,5$			$c = 0,5$			$c = 1,5$		
	$N = 100$	300	500	100	300	500	100	300	500	100	300	500
5	55.8	55.6	55.7	55.8	55.6	55.7	55.8	55.6	55.7	55.8	55.6	55.6
6	62.4	62.3	62.3	62.4	62.3	62.3	62.4	62.3	62.3	62.4	62.3	62.3
7	68.0	67.9	67.9	68.0	67.9	67.9	68.0	67.9	68.0	68.0	67.9	67.9
8	72.8	72.7	72.8	72.8	72.7	72.8	72.8	72.7	72.8	72.8	72.7	72.7
9	76.9	76.8	76.8	76.9	76.8	76.8	76.9	76.8	76.8	76.9	76.8	76.8
10	80.3	80.3	80.3	80.3	80.3	80.3	80.3	80.3	80.3	80.3	80.3	80.3
11	83.2	83.2	83.2	83.2	83.2	83.2	83.2	83.2	83.3	83.2	83.3	83.3
12	85.7	85.7	85.8	85.7	85.7	85.8	85.7	85.7	85.8	85.7	85.8	85.8
13	87.9	87.9	87.9	87.9	87.9	87.9	87.9	87.9	87.9	87.9	87.9	87.9
14	89.7	89.7	89.7	89.7	89.7	89.7	89.7	89.7	89.7	89.7	89.7	89.7
15	91.2	91.2	91.3	91.2	91.2	91.3	91.2	91.2	91.3	91.2	91.3	91.3

geral que pelo menos 90% das amostras convergiram e que este número aumenta significativamente com o aumento de k .

Figura 3.2: Quantidade das estimativas válidas de cada cenário para o modelo M_{tb} .

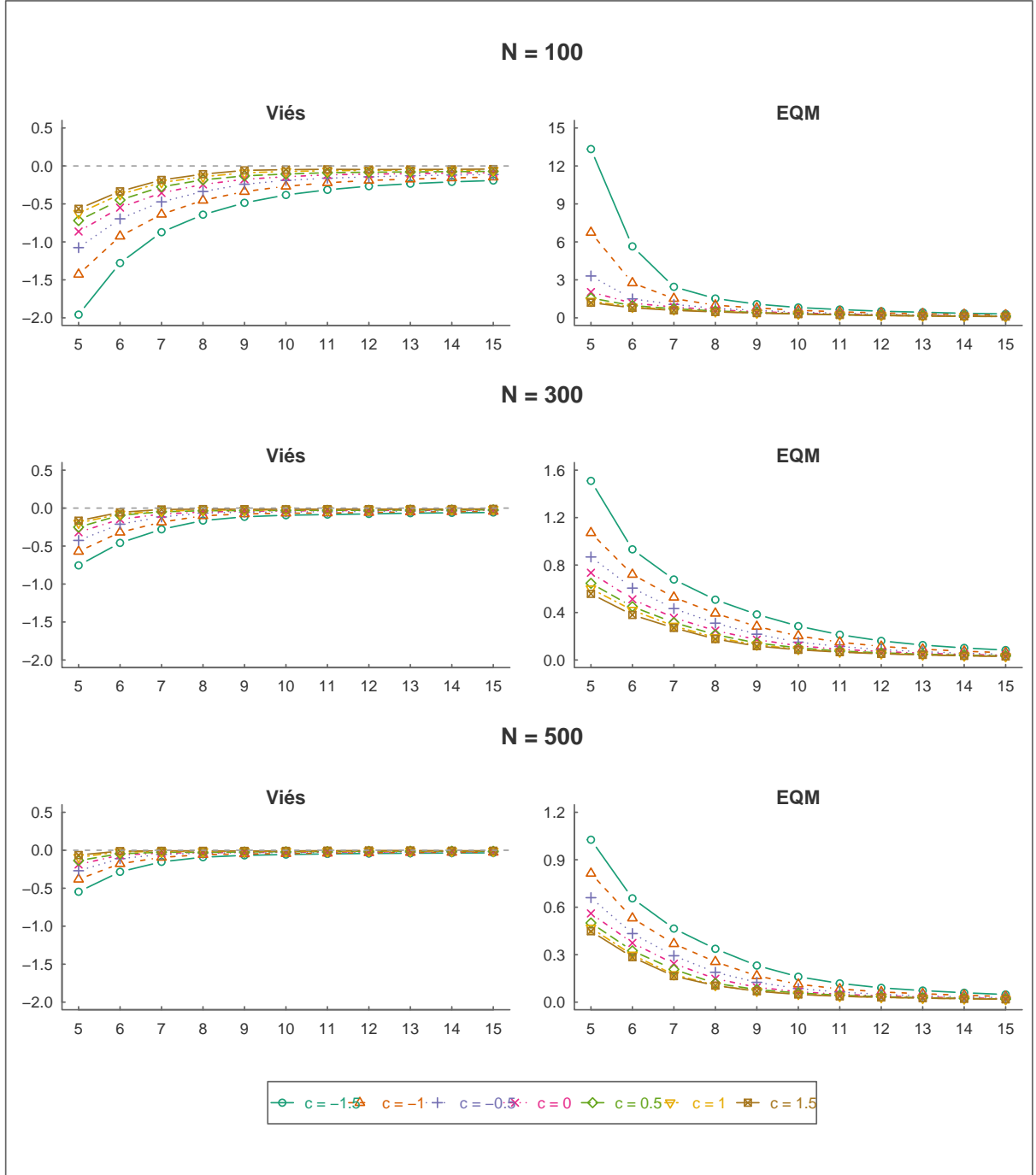
Analisando a Figura 3.3, podemos verificar um comportamento interessante do estimador do tamanho populacional: quando $N = 100$, tem-se subestimativas para N quando k é pequeno, passando para estimativas viesadas (viés positivo) quando k está em torno de 8 e quando k está acima de 10 tem-se que praticamente os vieses relativos convergem para zero. Quando $N = 300$ ou $N = 500$, as subestimativas são geradas praticamente só quando c verdadeiro é negativo. No geral, observa-se a forte dependência do estimador de N com o valor verdadeiro do parâmetro c , e que o viés relativo tende a zero e EQM tende a diminuir quando k aumenta, principalmente para $k \geq 10$.

Figura 3.3: Medidas de desempenho das estimativas do tamanho da população, N , para o modelo M_{tb} .



A Figura 3.4 mostra um comportamento de subestimativas para o parâmetro c , principalmente para valores de c negativos. Contudo, observa-se para crescimento de N e k , o estimador torna-se mais preciso no sentido que viés e EQM tendem a zero.

Figura 3.4: Medidas de desempenho das estimativas do parâmetro de efeito comportamental à marcação, c , para o modelo M_{tb} .



4 Conclusões

O presente projeto buscou estudar dois modelos tradicionais de captura-recaptura, o modelo M_t de heterogeneidade das probabilidades de captura entre as épocas de amostragem e o modelo M_{tb} que considera um parâmetro adicional no modelo M_t que leva em conta a possível diferença entre as probabilidades de captura de um animal marcado para um não marcado. O presente relatório buscou apontar as principais características destes dois modelos e obter as estimativas de máxima verossimilhança dos parâmetros.

Os exemplos numéricos apresentados foram baseados em dados reais e dados simulados. Nos dados reais, dois conjuntos de dados foram utilizados para estimar a população de ratos. Pôde-se observar nos resultados que as estimativa do tamanho da população sempre foram acima do número de animais observados, indicando ganho inferencial à população pelos modelos de captura-recaptura. Por fim, concluímos que os resultados do estudo de simulação indicaram boas propriedades dos estimadores de máxima verossimilhança no sentido que quanto mais épocas de captura mais precisas são as estimativas (diminuição de viés e erro quadrático médio). Contudo, estudos mais aprofundados devem ser dirigidos para apontar a dependência entre os estimadores dos parâmetros do modelo M_{tb} .

Por fim, ressaltamos que pesquisas complementares podem ser direcionadas para os modelos como, por exemplo, obtenção de intervalos de confiança, principalmente para o parâmetro N . Adicionalmente, um estudo de simulação para verificar o comportamento da seleção correta de modelos via critérios AIC e BIC poderia ser proposto futuramente junto com outros critérios como o teste de razão de verossimilhanças, por exemplo.

Referências

- ASHTON, John et al. Known unknowns: How much financial misconduct is detected and deterred? **Journal of International Financial Markets, Institutions and Money**, Elsevier, v. 74, p. 101389, 2021.
- BASU, Sanjib; EBRAHIMI, Nader. Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence. **Biometrika**, Oxford University Press, v. 88, n. 1, p. 269–279, 2001.
- BIRD, Sheila M; KING, Ruth. Multiple systems estimation (or capture-recapture estimation) to inform public policy. **Annual Review of Statistics and Its Application**, Europe PMC Funders, v. 5, p. 95, 2018.
- BÖHNING, Dankmar et al. Estimating the undetected infections in the Covid-19 outbreak by harnessing capture–recapture methods. **International Journal of Infectious Diseases**, Elsevier, v. 97, p. 197–201, 2020.
- BURNHAM, Kenneth P; WHITE, Gary C; ANDERSON, David R. Model selection strategy in the analysis of capture-recapture data. **Biometrics**, JSTOR, p. 888–898, 1995.
- CASELLA, George; BERGER, Roger L. **Statistical inference**. [S.l.]: Cengage Learning, 2021.
- COULOMBE, Harry N; RIDGWAY, Sam H; EVANS, William E. Respiratory water exchange in two species of porpoise. **Science**, American Association for the Advancement of Science, v. 149, n. 3679, p. 86–88, 1965.
- LAPLACE, P. S. **Sur les naissances, les mariages et les morts**. [S.l.]: In: Histoire de L Académie Royale des Sciences, Paris, 1783.
- LINCOLN, Frederick Charles. Calculating waterfowl abundance on the basis of banding returns. **US Department of Agriculture**, n. 118, 1930.
- MCCREA, Rachel S; MORGAN, Byron JT. **Analysis of capture-recapture data**. [S.l.]: CRC Press, 2014.

- OLIVEIRA, Antônio Wilson Soares de; SILVA, Ionizete Garcia da. Distribuição geográfica e indicadores entomológicos de triatomíneos sinantrópicos capturados no Estado de Goiás. **Revista da Sociedade Brasileira de Medicina Tropical**, SciELO Brasil, v. 40, p. 204–208, 2007.
- OTIS, David L et al. Statistical inference from capture data on closed animal populations. **Wildlife monographs**, JSTOR, n. 62, p. 3–135, 1978.
- PETERSEN, C. G. J. The yearly immigration of young plaice into Limfjord from the German sea, etc. **Rept. Danish Biol. Stn**, v. 6, n. 1-48, 1896.
- POLLOCK, Kenneth H. Capture-recapture models: a review of current methods, assumptions and experimental design. North Carolina State University. Dept. of Statistics, 1980.
- POLONSKY, Jonathan A et al. Novel use of capture-recapture methods to estimate completeness of contact tracing during an Ebola outbreak, Democratic Republic of the Congo, 2018–2020. **Emerging infectious diseases**, Centers for Disease Control e Prevention, v. 27, n. 12, p. 3063, 2021.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2022. Disponível em: <https://www.R-project.org/>.
- RODRIGUES, Josemar; BOLFARINE, Heleno; GALVÃO LEITE, José. A Bayesian analysis in closed animal populations from capture recapture experiments with trap response. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 17, n. 2, p. 407–430, 1988.
- ROYLE, J Andrew et al. **Spatial capture-recapture**. [S.l.]: Academic Press, 2013.
- RYNGELBLUM, Marcelo; PERES, Maria Fernanda Tourinho. Análise da qualidade dos dados das mortes cometidas por policiais no Município de São Paulo, Brasil, 2014-2015. **Cadernos de Saúde Pública**, SciELO Brasil, v. 37, 2021.
- SALASAR, Luis Ernesto Bueno. Eliminação de parâmetros perturbadores em um modelo de captura-recaptação. Universidade Federal de São Carlos, 2011.
- SEBER, George Arthur Frederick et al. The estimation of animal abundance and related parameters. Blackburn press Caldwell, New Jersey, 1982.
- WANG, Xiaoyin. **Bayesian analysis of capture-recapture models**. [S.l.]: University of Missouri-Columbia, 2002.
- WANG, Xiaoyin; HE, Zhuoqiong; SUN, Dongchu et al. Bayesian Estimation of Population Size via Capture-Recapture Model with Time Variation and Behavioral Response. **Open Journal of Ecology**, Scientific Research Publishing, v. 5, n. 01, p. 1, 2015.