

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

‘Maringá, 30 de Agosto de 2022



# Sumário

<b>1</b>	<b>Apresentação</b>	<b>5</b>
1.1	Resumo . . . . .	5
<b>2</b>	<b>Introdução</b>	<b>7</b>
2.1	Objetivos . . . . .	8
<b>3</b>	<b>Desenvolvimento (Materiais e Métodos)</b>	<b>11</b>
3.1	Modelo de captura-recaptura com heterogeneidade . . . . .	13
3.2	Modelo de captura-recaptura com heterogeneidade e efeito à marcação . . . . .	14



# Capítulo 1

## Apresentação

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

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

### 1.1 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

## Capítulo 2

# 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 and 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 and 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 and 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 et al. [1988], Oliveira and Silva [2007], Salazar [2011], Wang 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 and 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 and 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].

## 2.1 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:



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



## Capítulo 3

# 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 3.1 abaixo.

Tabela 3.1: Exemplo de quantidades observadas em um processo de captura-recaptura.

	j	1	2	3
Animais marcados antes da ocasião	M	0	10	22
Animais capturados (não marcados)	u	10	12	8
Animais recapturados (marcados)	m	0	3	9
Animais selecionados	n	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.

### 3.1 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 ou não captura) e independência das capturas e mesma probabilidade de sucesso (captura) para as quantidades  $u_j$  e  $m_j$ , temos as seguintes distribuições:

$$u_j | N, p_j, M_j \sim \text{Binomial}(N - M_j, p_j), \quad j = 1, 2, \dots, k \quad (3.1)$$

e

$$m_j | p_j, M_j \sim \text{Binomial}(M_j, p_j), \quad j = 2, \dots, k \quad (3.2)$$

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  $j = 1, 2, \dots, k$ ,

### 3.2 Modelo de captura-recaptura com heterogeneidade e efeito à marcação

# Referências Bibliográficas

- John Ashton, Tim Burnett, Ivan Diaz-Rainey, and Peter Ormosi. Known unknowns: How much financial misconduct is detected and deterred? *Journal of International Financial Markets, Institutions and Money*, 74:101389, 2021.
- Sanjib Basu and Nader Ebrahimi. Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence. *Biometrika*, 88(1):269–279, 2001.
- Sheila M Bird and Ruth King. Multiple systems estimation (or capture-recapture estimation) to inform public policy. *Annual Review of Statistics and Its Application*, 5:95, 2018.
- Dankmar Böhning, Irene Rocchetti, Antonello Maruotti, and Heinz Holling. Estimating the undetected infections in the covid-19 outbreak by harnessing capture–recapture methods. *International Journal of Infectious Diseases*, 97:197–201, 2020.
- P. S. Laplace. *Sur les naissances, les mariages et les morts*. Number 693. In: Histoire de L Académie Royale des Sciences, Paris, 1783.
- Frederick Charles Lincoln. Calculating waterfowl abundance on the basis of banding returns. *US Department of Agriculture*, (118), 1930.
- Rachel S McCrea and Byron JT Morgan. *Analysis of capture-recapture data*. CRC Press, 2014.
- Antônio Wilson Soares de Oliveira and Ionizete Garcia da Silva. Distribuição geográfica e indicadores entomológicos de triatomíneos sinantrópicos capturados no estado de goiás. *Revista da Sociedade Brasileira de Medicina Tropical*, 40:204–208, 2007.
- David L Otis, Kenneth P Burnham, Gary C White, and David R Anderson. Statistical inference from capture data on closed animal populations. *Wildlife monographs*, (62):3–135, 1978.
- C. G. J. Petersen. The yearly immigration of young plaice into limfjord from the german sea, etc. *Rept. Danish Biol. Stn*, 6(1-48), 1896.

- Jonathan A Polonsky, Dankmar Böhning, Mory Keita, Steve Ahuka-Mundeke, Justus Nsio-Mbeta, Aaron Aruna Abedi, Mathias Mossoko, Janne Estill, Olivia Keiser, Laurent Kaiser, 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*, 27(12): 3063, 2021.
- Josemar Rodrigues, Heleno Bolfarine, and José Galvão Leite. A bayesian analysis in closed animal populations from capture recapture experiments with trap response. *Communications in Statistics-Simulation and Computation*, 17(2): 407–430, 1988.
- J Andrew Royle, Richard B Chandler, Rahel Sollmann, and Beth Gardner. *Spatial capture-recapture*. Academic Press, 2013.
- Marcelo Ryngelblum and Maria Fernanda Tourinho Peres. Análise da qualidade dos dados das mortes cometidas por policiais no município de são paulo, brasil, 2014-2015. *Cadernos de Saúde Pública*, 37, 2021.
- Luis Ernesto Bueno Salasar. Eliminação de parâmetros perturbadores em um modelo de captura-recaptura. 2011.
- George Arthur Frederick Seber et al. The estimation of animal abundance and related parameters. 1982.
- Xiaoyin Wang, Zhuoqiong He, Dongchu Sun, et al. Bayesian estimation of population size via capture-recapture model with time variation and behavioral response. *Open Journal of Ecology*, 5(01):1, 2015.