

ESTIMAÇÃO DO TAMANHO DE UMA POPULAÇÃO PELO PROCESSO DE CAPTURA-RECAPTURA COM UM ÚNICO ESTÁGIO DE MARCAÇÃO

Gilson Yuuji SHIMIZU¹
José Galvão LEITE²
Nelson Ithiro TANAKA¹

- RESUMO: Consideramos o problema da estimação do tamanho de uma população animal fechada e das probabilidades de captura, segundo o processo de captura-recaptura com um único estágio de marcação. Tal processo consiste na marcação de um número de animais selecionados aleatoriamente somente na primeira época de amostragem e, posteriormente, na contagem do número de animais marcados e não marcados observados em cada uma entre várias épocas. Em alguns casos este método é mais simples e menos oneroso de ser aplicado do que o processo de captura-recaptura usual, onde os animais capturados são marcados em todas as épocas do processo. Determinamos estimativas de máxima verossimilhança e de máxima verossimilhança condicional do tamanho populacional e das probabilidades de captura, bem como um intervalo de confiança aproximado para o tamanho populacional. Apresentamos exemplos com dados simulados e um exemplo com dados reais.
- PALAVRAS-CHAVE: População fechada; processo de captura-recaptura com um único estágio de marcação; estimativas de máxima verossimilhança e de máxima verossimilhança condicional; intervalo de confiança.

1 Introdução

O método de captura-recaptura, originariamente utilizado para estimar o tamanho de uma população animal, consiste na seleção de animais de uma

¹Departamento de Estatística, Instituto de Matemática e Estatística – IME, Universidade de São Paulo – USP, Rua do Matão, 1010 - Cidade Universitária, CEP: 05508-090, São Paulo, SP, Brasil, E-mail: *gshimizu@ime.usp.br* / *nitanaka@ime.usp.br*

²Departamento de Estatística, Centro de Ciências Exatas e Tecnologia, Universidade Federal de São Carlos-UFSCar, Caixa Postal 676, CEP: 13565-905, São Carlos, SP, Brasil, E-mail: *leite@power.ufscar.br*

determinada área em várias épocas de amostragem. Os animais capturados na primeira amostra são marcados e devolvidos à população. Em cada amostra seguinte, faz-se o registro de uma nova captura (recaptura) para cada animal previamente marcado capturado (recapturado), os animais capturados pela primeira vez (não marcados) são marcados e todos os animais são devolvidos à população. Desse modo, ao final do processo a história ou a trajetória de captura de cada animal é conhecida e a estimativa do tamanho populacional pode ser obtida, em função dos números de animais marcados e não marcados observados nas várias épocas de amostragem. Atualmente este método também é utilizado em outras áreas do conhecimento, como nas áreas de saúde pública, controle de qualidade e controle de erros de programas de computadores. Na área de saúde pública o método pode ser utilizado na estimação do número de indivíduos acometidos por determinada doença, auxiliando no planejamento de medidas preventivas ou corretivas do tipo vacinação, por exemplo. Na área de controle de qualidade o método pode ser aplicado na estimação do número de partículas contidas em um determinado tipo de filme fotográfico e finalmente, com relação à área de controle de erros de programas de computadores, o método pode ser usado na estimação do número de erros remanescentes de um programa, o que pode ter significativo impacto econômico, principalmente em programas críticos. Vários pesquisadores, como Campos; Almeida e Iochida (1998), Fienberg (1972), Micheletti (2003), Nayak (1988), Pollock (1991) e Sanathanan (1972b) publicaram trabalhos neste contexto. Todavia, com relação à aplicação do método na estimação do tamanho de uma população animal, existem situações em que a captura e a marcação dos animais não marcados, em cada uma das épocas de amostragem, podem ser complexas, onerosas e consumirem demasiado tempo. Uma dessas situações ocorre com animais cujo hábitat é de difícil acesso, como os cabritos que vivem em regiões montanhosas. Com o objetivo de contornar esta problemática consideramos o que denominamos processo de captura-recaptura com um único estágio de marcação. Este processo é uma variante do processo de captura-recaptura, em que somente os animais capturados na primeira amostra são marcados, ou seja, todos os animais capturados a partir da segunda amostra não são marcados, mas simplesmente devolvidos à população antes da seleção da próxima amostra. Desse modo amostras “visuais” podem ser tomadas uma vez que, a partir da segunda amostra, os animais não são marcados. Se considerarmos somente duas amostras esse processo coincide com o método de captura-recaptura usual. Ananda (1997) utilizou tal processo para estimar o número de cabritos que habitavam as montanhas do deserto de Mojave, nos Estados Unidos da América. Na determinação da função de verossimilhança ele aproximou a distribuição hipergeométrica pela distribuição binomial, reduzindo assim o problema da estimação do tamanho populacional ao problema clássico da estimação da probabilidade de sucesso da distribuição binomial. Neste artigo generalizamos essa metodologia supondo válidas as premissas: (a) a população considerada é fechada, isto é, não há nascimentos, mortes, emigração ou imigração de elementos da população durante o período de estudo; (b) todo animal marcado não perde sua marca ao longo do processo; (c) todos os animais marcados devem ser contados e

registrados corretamente em cada amostra e (d) todos os animais não marcados previamente devem ser contados e registrados corretamente em cada amostra. Na Seção 2 introduzimos o modelo estatístico e as funções de verossimilhança; na Seção 3 determinamos as estimativas de máxima verossimilhança e de máxima verossimilhança condicional dos parâmetros do modelo; na Seção 4 construímos um intervalo de confiança aproximado para o tamanho populacional; na Seção 5 analisamos a *performance* das estimativas por meio de exemplos com dados simulados e apresentamos um exemplo com dados reais. Finalmente, apresentamos as conclusões.

2 Modelo estatístico e funções de verossimilhança

Suponhamos uma população fechada constituída de N animais, N desconhecido, e que cada animal, independentemente dos demais, tenha probabilidade p_i de ser capturado na $(i + 1)$ -ésima amostra, $i = 0, 1, 2, \dots, s$, onde $s + 1$ é o número de amostras selecionadas da população durante todo o processo, $s \geq 1$. Na primeira amostra registramos o número de animais capturados, marcamos e devolvemos todos eles à população. Nas demais amostras registramos os números dos animais capturados e dos marcados (recapturados), mas os animais capturados pela primeira vez não são marcados antes de serem devolvidos à população, isto é, somente os animais capturados na primeira amostra são marcados. Suponhamos também que todo animal marcado não perca sua marca ao longo do tempo. Denotemos por

n_i o número de animais capturados na $(i + 1)$ -ésima amostra;

m_i o número de animais recapturados na $(i + 1)$ -ésima amostra entre os animais marcados na primeira amostra, $i = 0, 1, 2, \dots, s$, ($m_0 = 0$);

$\mathbb{P} = (p_0, p_1, \dots, p_s)$ o vetor de probabilidades de capturas;

$D = (n_0, n_1, m_1; \dots; n_s, m_s)$ o vetor de dados amostrais e

$r = n_0 + \max\{n_1 - m_1, n_2 - m_2, \dots, n_s - m_s\}$ o número mínimo de animais distintos na população.

Logo, supondo que as amostras sejam selecionadas independentemente, segue que as distribuições de probabilidades conjunta de $(n_0, n_1, m_1; \dots; n_s, m_s)$ e de (n_0, n_1, \dots, n_s) , dados \mathbb{P} e N , são dadas por

$$\begin{aligned} P(n_0, n_1, m_1; \dots; n_s, m_s | \mathbb{P}, N) &= P(n_0 | \mathbb{P}, N) \prod_{i=1}^s P(n_i - m_i, m_i | n_0, \mathbb{P}, N) \\ &= \binom{N}{n_0} \left[\prod_{i=1}^s \binom{n_0}{m_i} \binom{N - n_0}{n_i - m_i} \right] \prod_{i=0}^s p_i^{n_i} (1 - p_i)^{N - n_i} \end{aligned} \quad (1)$$

e

$$P(n_0, n_1, \dots, n_s | \mathbb{P}, N) = \prod_{i=0}^s \binom{N}{n_i} p_i^{n_i} (1 - p_i)^{N - n_i}, \quad (2)$$

respectivamente.

De (1) e (2) temos

$$\begin{aligned}
 & P(m_1, m_2, \dots, m_s | n_0, n_1, \dots, n_s, \mathbf{p}, N) \\
 &= \prod_{i=1}^s P(m_i | n_0, n_1, \dots, n_s, \mathbf{p}, N) = \prod_{i=1}^s \frac{\binom{N-n_0}{n_i-m_i} \binom{n_0}{m_i}}{\binom{N}{n_i}} \quad (3)
 \end{aligned}$$

e de (1) segue que a função de verossimilhança é tal que

$$\begin{aligned}
 L(\mathbf{p}, N | D) &= L(\mathbf{p}, N | n_0; n_1, m_1; \dots; n_s, m_s) \\
 &\propto \binom{N}{n_0} \left[\prod_{i=1}^s \binom{N-n_0}{n_i-m_i} \right] \prod_{i=0}^s p_i^{n_i} (1-p_i)^{N-n_i}, \quad (4)
 \end{aligned}$$

$0 < p_i < 1$, $i = 0, 1, 2, \dots, s$, $N \geq r$.

Denotemos por $L_1(N | m_1, m_2, \dots, m_s)$ a função de verossimilhança correspondente à distribuição de probabilidades condicional de (m_1, m_2, \dots, m_s) , dados (n_0, n_1, \dots, n_s) , N e \mathbf{p} . Segue de (3) que

$$L_1(N | m_1, m_2, \dots, m_s) \propto \prod_{i=1}^s \frac{\binom{N-n_0}{n_i-m_i}}{\binom{N}{n_i}}, \quad (5)$$

$N \geq r$, e de (2) segue que a função de verossimilhança correspondente à distribuição de probabilidades de n_0, n_1, \dots, n_s , dados \mathbf{p} e N , é dada por

$$L_2(\mathbf{p}, N | n_0, n_1, \dots, n_s) = \prod_{i=0}^s \binom{N}{n_i} p_i^{n_i} (1-p_i)^{N-n_i}, \quad (6)$$

$0 < p_i < 1$, $N \geq \max\{n_0, n_1, \dots, n_s\}$. Então, de (4), (5) e (6) temos

$$L(\mathbf{p}, N | D) = L_1(N | m_1, m_2, \dots, m_s) L_2(\mathbf{p}, N | n_0, n_1, \dots, n_s). \quad (7)$$

3 Estimativas de máxima verossimilhança

Nesta seção determinamos as estimativas de máxima verossimilhança, $(\hat{\mathbf{p}}, \hat{N})$, e de máxima verossimilhança condicional, $(\hat{\mathbf{p}}_c, \hat{N}_c)$, de (\mathbf{p}, N) . Nos teoremas 3.1 e 3.2 determinamos $(\hat{\mathbf{p}}, \hat{N})$ para o caso particular em que $p_i = p$, $i = 0, 1, 2, \dots, s$, e para o caso geral, respectivamente. De maneira análoga, determinamos $(\hat{\mathbf{p}}_c, \hat{N}_c)$ nos teoremas 3.3 e 3.4. Em particular, mostramos que \hat{N} e \hat{N}_c são infinitos quando nenhum animal marcado for observado durante o processo. As demonstrações dos teoremas e corolários desta seção podem ser encontradas em Shimizu (2000).

Teorema 3.1. Suponhamos que $p_i = p$, $0 < p < 1$, $n_i \geq 1$, $i = 0, 1, \dots, s$. Então,

$$\hat{p} = \frac{\sum_{i=0}^s n_i}{(s+1)\hat{N}} \text{ e } \hat{N} \text{ é aproximadamente igual a solução da equação}$$

$$\left(1 - \frac{\sum_{i=0}^s n_i}{(s+1)\hat{N}}\right)^{s+1} = \left(1 - \frac{n_0}{\hat{N}}\right) \prod_{i=1}^s \left(1 - \frac{n_i - m_i}{\hat{N} - n_0}\right), N > r.$$

Do teorema 3.1 segue o seguinte corolário.

Corolário 3.1. Suponhamos que $p_i = p$, $0 < p < 1$, $n_i \geq 1$, $i = 0, 1, \dots, s$. Se $m_i = 0$, $i = 1, 2, \dots, s$, então $\hat{N} = \infty$.

Teorema 3.2. Suponhamos que $n_i \geq 1$, $i = 0, 1, \dots, s$. Então, $\hat{\mathbf{p}} = (\hat{p}_0, \hat{p}_1, \dots, \hat{p}_s)$, onde $\hat{p}_i = \frac{n_i}{\hat{N}}$, $i = 0, 1, \dots, s$, e \hat{N} é aproximadamente igual a solução da equação

$$\prod_{i=0}^s \left(1 - \frac{n_i}{\hat{N}}\right) = \left(1 - \frac{n_0}{\hat{N}}\right) \prod_{i=1}^s \left(1 - \frac{n_i - m_i}{\hat{N} - n_0}\right), N > r.$$

Do teorema 3.2 segue o seguinte corolário.

Corolário 3.2. Suponhamos que $n_i \geq 1$, $i = 0, 1, \dots, s$. Se $m_i = 0$, $i = 1, 2, \dots, s$, então $\hat{N} = \infty$.

Uma outra forma de estimar (\mathbf{p}, N) é pela metodologia da máxima verossimilhança condicional introduzida por Sanathanan (1972a), uma vez que a função de verossimilhança $L(\mathbf{p}, N|D)$ pode ser escrita como

$$L(\mathbf{p}, N|D) = L_1(N|m_1, m_2, \dots, m_s) L_2(\mathbf{p}, N|n_0, n_1, \dots, n_s)$$

(veja a relação (7)). Nesse contexto, o método da máxima verossimilhança condicional consiste primeiramente na determinação da estimativa de máxima verossimilhança de N , \hat{N}_c , com relação a $L_1(N|m_1, m_2, \dots, m_s)$, na substituição de N por \hat{N}_c em $L_2(\mathbf{p}, N|n_0, n_1, \dots, n_s)$ e, em seguida, na determinação da estimativa de máxima verossimilhança de \mathbf{p} , $\hat{\mathbf{p}}_c$, com relação a $L_2(\mathbf{p}, \hat{N}_c|n_0, n_1, \dots, n_s)$.

De (5), para se determinar um ponto de máximo de $L_1(N|m_1, m_2, \dots, m_s)$ basta determinar um ponto de máximo de $K(N) = \prod_{i=1}^s \frac{\binom{N-n_0}{n_i-m_i}}{\binom{N}{n_i}}$, $N \geq r$, onde a função $K(N)$ é o kernel de $L_1(N|m_1, m_2, \dots, m_s)$. Então, igualando a “derivada em relação a N do $\ln K(N)$ ” a zero temos $K(N) = K(N-1)$, e como

$$K(N) = \prod_{i=1}^s \frac{\binom{N-n_0}{n_i-m_i}}{\binom{N}{n_i}} = K(N-1) = \prod_{i=1}^s \frac{\binom{N-1-n_0}{n_i-m_i}}{\binom{N-1}{n_i}}$$

$$\Leftrightarrow \left(1 - \frac{n_0}{N}\right)^s = \prod_{i=1}^s \left(\frac{1 - \frac{n_0+n_i-m_i}{N}}{1 - \frac{n_i}{N}}\right), \quad (8)$$

$N > r$, segue que \widehat{N}_c é aproximadamente igual a solução de (8).

Novamente se $m_i = 0$, $i = 1, 2, \dots, s$, então a única solução de (8) é $\widehat{N}_c = \infty$.

Existem situações em que n_i é muito menor do que N , $i = 1, 2, \dots, s$, e n_0 é tal que a ordem de magnitude de $\frac{n_0}{N}$ não é tão próxima de zero. Notando que $\frac{n_i}{N} \cong 0$, $i = 1, 2, \dots, s$, podemos aproximar $L_1(N|m_1, m_2, \dots, m_s)$ por

$$\begin{aligned} & L_1(N|m_1, m_2, \dots, m_s) \\ &= \prod_{i=1}^s \frac{\binom{N-n_0}{n_i-m_i} \binom{n_0}{m_i}}{\binom{N}{n_i}} \\ &\cong \prod_{i=1}^s \binom{n_i}{m_i} \left(\frac{n_0}{N}\right)^{m_i} \left(1 - \frac{n_0}{N}\right)^{n_i-m_i} \\ &\propto \left(\frac{n_0}{N}\right)^{\sum_{i=1}^s m_i} \left(1 - \frac{n_0}{N}\right)^{\sum_{i=1}^s n_i - \sum_{i=1}^s m_i}, \end{aligned} \quad (9)$$

$N \geq r$. Segue que a estimativa de máxima verossimilhança condicional de $\frac{n_0}{N}$ é tal que

$$\left(\frac{n_0}{N}\right)_c \cong \frac{\sum_{i=1}^s m_i}{\sum_{i=1}^s n_i}$$

e, pela propriedade da invariância, a estimativa de máxima verossimilhança condicional de N é dada por

$$\widehat{N}_c \cong \widehat{N}_A = \frac{n_0 \sum_{i=1}^s n_i}{\sum_{i=1}^s m_i},$$

onde \widehat{N}_A é a estimativa obtida por Ananda (1997).

Uma vez obtida a estimativa de máxima verossimilhança condicional, \widehat{N}_c , de N pela solução da equação (8), podemos então obter a estimativa de máxima verossimilhança condicional para o vetor de probabilidades \mathbf{p} , maximizando agora a função de verossimilhança $L_2(\mathbf{p}, \widehat{N}_c | n_0, n_1, \dots, n_s)$. Os teoremas 3.3 e 3.4 a seguir fornecem estas estimativas para o caso de probabilidades de captura iguais e para o caso geral, respectivamente.

Teorema 3.3. Suponhamos que $p_i = p$, $0 < p < 1$, $n_i \geq 1$, $i = 0, 1, \dots, s$. Então

$\hat{p}_c = \frac{\sum_{i=0}^s n_i}{(s+1)\hat{N}_c}$ é a estimativa de máxima verossimilhança condicional de p obtida de $L_2(p, \hat{N}_c | n_0, n_1, \dots, n_s)$, onde \hat{N}_c é aproximadamente igual a solução da equação (8).

Teorema 3.4. Suponhamos que $n_i \geq 1$, $i = 0, 1, \dots, s$. Então

$$\hat{\mathbb{P}}_c = (\hat{p}_{c0}, \hat{p}_{c1}, \dots, \hat{p}_{cs}) = \left(\frac{n_0}{\hat{N}_c}, \frac{n_1}{\hat{N}_c}, \dots, \frac{n_s}{\hat{N}_c} \right)$$

é a estimativa de máxima verossimilhança condicional de \mathbb{p} obtida de

$$L_2(\mathbb{p}, \hat{N}_c | n_0, n_1, \dots, n_s),$$

onde \hat{N}_c é aproximadamente igual a solução da equação (8).

Os teoremas 3.3 e 3.4 são válidos quando \hat{N}_c for substituído por \hat{N}_A . Logo, a estimativa de máxima verossimilhança condicional do vetor de probabilidades de captura \mathbb{p} obtida de $L_2(\mathbb{p}, \hat{N}_A | n_0, n_1, \dots, n_s)$ é dada, neste caso, por $\hat{p}_A = \frac{\sum_{i=0}^s n_i}{(s+1)\hat{N}_A}$ ou $\hat{\mathbb{P}}_A = \left(\frac{n_0}{\hat{N}_A}, \frac{n_1}{\hat{N}_A}, \dots, \frac{n_s}{\hat{N}_A} \right)$.

4 Intervalo de confiança aproximado para o tamanho populacional

Construímos nesta seção um intervalo de confiança aproximado para N . De (9) temos

$$P(m_1, m_2, \dots, m_s | n_0, n_1, \dots, n_s, N) \cong \prod_{i=1}^s \binom{n_i}{m_i} \left(\frac{n_0}{N} \right)^{m_i} \left(1 - \frac{n_0}{N} \right)^{n_i - m_i}$$

e

$$P(m_i | n_0, n_1, \dots, n_s, N) \cong \binom{n_i}{m_i} \left(\frac{n_0}{N} \right)^{m_i} \left(1 - \frac{n_0}{N} \right)^{n_i - m_i},$$

$i = 1, 2, \dots, s$, isto é, dados n_0, n_1, \dots, n_s e N , temos que m_1, m_2, \dots, m_s são independentes e identicamente distribuídos com distribuição aproximadamente *Binomial*($n_i, \frac{n_0}{N}$), $i = 1, 2, \dots, s$. Então, dados n_0, n_1, \dots, n_s e N , $\sum_{i=1}^s m_i$ tem distribuição aproximadamente Binomial com parâmetros $\sum_{i=1}^s n_i$ e $\frac{n_0}{N}$, ou seja, a distribuição de $\sum_{i=1}^s m_i$ é aproximadamente igual à distribuição de uma soma de $\sum_{i=1}^s n_i$

váriáveis aleatórias independentes e identicamente distribuídas com distribuição de Bernoulli de parâmetro $\frac{n_0}{N}$. Logo, pelo Teorema Central do Limite,

$$\frac{\sum_{i=1}^s m_i - \left(\sum_{i=1}^s n_i\right) \frac{n_0}{N}}{\sqrt{\left(\sum_{i=1}^s n_i\right) \frac{n_0}{N} \left(1 - \frac{n_0}{N}\right)}} \text{ tem distribuição aproximadamente Normal Padrão.}$$

Então, um intervalo de confiança $100(1 - \alpha)\%$ aproximado para $\frac{n_0}{N}$ é dado por

$$\left(\left(\frac{\sum_{i=1}^s m_i}{\sum_{i=1}^s n_i} \right) + z_{\alpha/2} \sqrt{\frac{\sum_{i=1}^s m_i \left(\sum_{i=1}^s n_i - \sum_{i=1}^s m_i \right)}{\left(\sum_{i=1}^s n_i \right)^3}}, \right. \\ \left. \left(\frac{\sum_{i=1}^s m_i}{\sum_{i=1}^s n_i} \right) - z_{\alpha/2} \sqrt{\frac{\sum_{i=1}^s m_i \left(\sum_{i=1}^s n_i - \sum_{i=1}^s m_i \right)}{\left(\sum_{i=1}^s n_i \right)^3}} \right),$$

o que implica que um intervalo de confiança $100(1 - \alpha)\%$ aproximado para N é dado por

$$\left(\frac{\frac{n_0}{\left(\frac{\sum_{i=1}^s m_i}{\sum_{i=1}^s n_i} \right) - z_{\alpha/2} \sqrt{\frac{\sum_{i=1}^s m_i \left(\sum_{i=1}^s n_i - \sum_{i=1}^s m_i \right)}{\left(\sum_{i=1}^s n_i \right)^3}},}{\frac{n_0}{\left(\frac{\sum_{i=1}^s m_i}{\sum_{i=1}^s n_i} \right) + z_{\alpha/2} \sqrt{\frac{\sum_{i=1}^s m_i \left(\sum_{i=1}^s n_i - \sum_{i=1}^s m_i \right)}{\left(\sum_{i=1}^s n_i \right)^3}}}, \right) \right), \quad (10)$$

onde $z_{\alpha/2}$ é o quantil $\alpha/2$ da distribuição normal padrão.

5 Estudo do desempenho dos estimadores

Nesta seção, a fim de analisar o desempenho das estimativas de N e p estudadas anteriormente, apresentamos exemplos com dados simulados. Apresentamos também um exemplo com dados reais.

Nos exemplos com dados simulados, consideramos $s = 4$, isto é, cinco épocas de amostragem e atribuímos vários valores para o tamanho populacional e para o vetor de probabilidades de captura.

No primeiro exemplo supomos $N = 500$, $p_i = p$, $i = 0, 1, 2, 3, 4$, e atribuímos diversos valores a p , no segundo, supomos $N = 500$ e assumimos p_i 's diferentes, no terceiro, consideramos diversos tamanhos populacionais e p_i 's diferentes e, no quarto exemplo, construímos um intervalo de confiança aproximado para N com dados do Exemplo 3.

5.1 Exemplo 1

Para $p_i = p$, $i = 0, 1, 2, 3, 4$, e $N = 500$ os dados amostrais foram gerados de acordo com as seguintes distribuições de probabilidades:

$$\begin{aligned} n_0 &\sim \text{Binomial}(500, p), \\ m_i &\sim \text{Binomial}(n_0, p), \\ n_i &= m_i + d_i, \text{ onde } d_i \sim \text{Binomial}(500 - n_0, p), i = 1, 2, 3, 4. \end{aligned}$$

A Tabela 1 mostra as estimativas \hat{N} , \hat{p} , \hat{N}_c , \hat{p}_c , \hat{N}_A e \hat{p}_A para diversos valores atribuídos a p .

Tabela 1 - Valores de \hat{N} , \hat{p} , \hat{N}_c , \hat{p}_c , \hat{N}_A e \hat{p}_A segundo amostras geradas para diferentes valores de p

p	\hat{N}	\hat{p}	\hat{N}_c	\hat{p}_c	\hat{N}_A	\hat{p}_A
0,95	500,41	0,95	500,40	0,95	499,94	0,95
0,90	498,82	0,91	498,79	0,91	498,40	0,91
0,80	498,67	0,80	498,56	0,80	498,37	0,80
0,40	494,81	0,40	493,74	0,40	493,54	0,40
0,30	467,84	0,31	467,46	0,31	467,55	0,31
0,20	470,24	0,21	467,82	0,21	469,33	0,21
0,10	518,89	0,10	513,45	0,10	513,71	0,10
0,09	604,09	0,07	595,67	0,07	592,36	0,07
0,08	633,36	0,06	626,89	0,06	628,00	0,06
0,07	743,20	0,04	730,38	0,04	729,00	0,04
0,06	675,18	0,04	662,20	0,04	661,25	0,04
0,05	483,01	0,05	470,99	0,05	472,50	0,05
0,04	401,25	0,05	390,27	0,05	392,00	0,05
0,03	707,30	0,02	677,57	0,02	680,00	0,02

Podemos observar nesse exemplo que as estimativas de N segundo o modelo de Ananda (1997) e as estimativas de máxima verossimilhança condicionais são próximas das estimativas de máxima verossimilhança. Contudo, as estimativas do tamanho populacional e das diferentes probabilidades de captura são inadequadas quando a probabilidade de captura é pequena, isto é, quando $p \leq 0,09$, com exceção do caso onde $p = 0,05$.

5.2 Exemplo 2

Para $\mathbb{p} = (p_0, p_1, p_2, p_3, p_4)$ e $N = 500$ os dados amostrais foram gerados de acordo com as seguintes distribuições de probabilidades:

$n_0 \sim \text{Binomial}(500, p_0); m_i \sim \text{Binomial}(n_0, p_i); n_i = m_i + d_i, \text{ onde}$
 $d_i \sim \text{Binomial}(500 - n_0, p_i), i = 1, 2, 3, 4.$

Atribuimos a \mathbf{p} os valores

$$(0, 6; 0, 4; 0, 3; 0, 5; 0, 6);$$

$$(0, 2; 0, 1; 0, 05; 0, 15; 0, 2) \text{ e}$$

$$(0, 04; 0, 02; 0, 05; 0, 02; 0, 04)$$

As respectivas estimativas de máxima verossimilhança e máxima verossimilhança condicional obtidas para \mathbf{p} e N foram

$$\hat{\mathbf{p}} = \hat{\mathbf{p}}_c = (0, 62; 0, 41; 0, 33; 0, 49; 0, 60) \text{ e } \hat{N} = \hat{N}_c = 510, 37;$$

$$\hat{\mathbf{p}} = \hat{\mathbf{p}}_c = (0, 24; 0, 12; 0, 08; 0, 19; 0, 31) \text{ e } \hat{N} = \hat{N}_c = 368, 50 \text{ e}$$

$$\hat{\mathbf{p}} = \hat{\mathbf{p}}_c = (0, 04; 0, 03; 0, 1; 0, 02; 0, 06) \text{ e } \hat{N} = \hat{N}_c = 316, 61.$$

e as respectivas estimativas segundo o modelo de Ananda (1997) obtidas para \mathbf{p} e N foram

$$\hat{\mathbf{p}}_A = (0, 62; 0, 41; 0, 33; 0, 49; 0, 60) \text{ e } \hat{N}_A = 511, 53;$$

$$\hat{\mathbf{p}}_A = (0, 23; 0, 12; 0, 08; 0, 19; 0, 30) \text{ e } \hat{N}_A = 373, 64 \text{ e}$$

$$\hat{\mathbf{p}}_A = (0, 04; 0, 03; 0, 1; 0, 02; 0, 06) \text{ e } \hat{N}_A = 322.$$

Nesse exemplo, onde as probabilidades de capturas são diferentes, notamos que as estimativas de máxima verossimilhança, \hat{N} e \hat{p} , e as estimativas de máxima verossimilhança condicionais, \hat{N}_c e \hat{p}_c , de N e p são iguais.

Novamente, as estimativas de N , segundo o modelo de Ananda (1997), são bem próximas das estimativas de máxima verossimilhança (e conseqüentemente das estimativas de máxima verossimilhança condicionais) de N . Entretanto, quando as probabilidades de captura são pequenas todas as estimativas são inadequadas devido a um número reduzido de capturas e recapturas.

5.3 Exemplo 3

Nesse exemplo supomos probabilidades de captura pequenas (exceto p_0) e atribuimos a N valores moderados e grandes. Supondo então $\mathbb{p} = (p_0, p_1, p_2, p_3, p_4) = (0, 25; 0, 01; 0, 02; 0, 01; 0, 03); N = 500, 1.000, 10.000 \text{ e } 100.000$ constatamos que $\hat{N}_c = \hat{N}$ e $\hat{p}_c = \hat{p}$. As estimativas \hat{N} e \hat{p} estão descritas na Tabela 2.

Tabela 2 - Valores de \hat{N} , \hat{p} , \hat{N}_A e \hat{p}_A segundo amostras geradas para $\mathbf{p}=(0,25; 0,01; 0,02; 0,01; 0,03)$ e diferentes valores de N

	$N = 500$	$N = 1.000$	$N = 10.000$	$N = 100.000$
\hat{N}	652, 82	837, 21	9920, 60	98149, 72
\hat{p}_0	0, 172	0, 276	0, 246	0, 253
\hat{p}_1	0, 005	0, 013	0, 010	0, 010
\hat{p}_2	0, 017	0, 031	0, 019	0, 020
\hat{p}_3	0, 003	0, 011	0, 010	0, 010
\hat{p}_4	0, 029	0, 041	0, 030	0, 031
\hat{N}_A	653, 33	840, 00	9918, 29	98145, 55
\hat{p}_{A0}	0, 171	0, 275	0, 246	0, 253
\hat{p}_{A1}	0, 005	0, 013	0, 099	0, 010
\hat{p}_{A2}	0, 017	0, 031	0, 019	0, 020
\hat{p}_{A3}	0, 003	0, 011	0, 010	0, 010
\hat{p}_{A4}	0, 029	0, 040	0, 030	0, 031

Ainda para os mesmos valores de N e supondo o vetor de probabilidades de captura $\mathbf{p} = (0, 05; 0, 01; 0, 02; 0, 01; 0, 03)$, obtivemos as estimativas de N e \mathbf{p} apresentadas na Tabela 3.

Tabela 3 - Valores de \hat{N} , \hat{p} , \hat{N}_A e \hat{p}_A segundo amostras geradas para $\mathbf{p}=(0,05; 0,01; 0,02; 0,01; 0,03)$ e diferentes valores de N

	$N = 500$	$N = 1.000$	$N = 10.000$	$N = 100.000$
\hat{N}	574, 39	656, 05	11793, 73	99537, 27
\hat{p}_0	0, 031	0, 061	0, 040	0, 049
\hat{p}_1	0, 005	0, 015	0, 010	0, 010
\hat{p}_2	0, 023	0, 038	0, 019	0, 020
\hat{p}_3	0, 005	0, 012	0, 010	0, 010
\hat{p}_4	0, 023	0, 035	0, 025	0, 030
\hat{N}_A	576, 00	660, 00	11791, 24	99553, 49
\hat{p}_{A0}	0, 031	0, 061	0, 040	0, 049
\hat{p}_{A1}	0, 005	0, 015	0, 010	0, 010
\hat{p}_{A2}	0, 023	0, 038	0, 019	0, 020
\hat{p}_{A3}	0, 005	0, 012	0, 009	0, 010
\hat{p}_{A4}	0, 023	0, 035	0, 025	0, 030

Com base nos resultados obtidos nesse exemplo constatamos que, no caso de uma população grande as estimativas de N e \mathbf{p} são boas, mesmo que as probabilidades de observar um animal durante o processo sejam pequenas. Contudo, quanto maior for a probabilidade de captura na etapa inicial do processo, (p_0),

possivelmente mais animais marcados teremos e, conseqüentemente, melhores serão as estimativas.

5.4 Exemplo 4

Nesse exemplo construímos intervalos de confiança aproximados para N segundo os dados do exemplo 3. Supondo:

$$\mathbb{P} = (p_0, p_1, p_2, p_3, p_4) = (0, 25; 0, 01; 0, 02; 0, 01; 0, 03) \text{ e}$$

$$N = 500, 1.000, 10.000 \text{ e } 100.000,$$

obtivemos os intervalos (expressão (10)) descritos na tabela Tabela 4.

Tabela 4 - Intervalos de confiança aproximados de 95% para N

N	IC de 95% para N
500	(378,01; 2405,02)
1.000	(619,56; 1303,94)
10.000	(8764,30; 11422,25)
100.000	(94342,35; 102268,26)

As amplitudes dos intervalos de confiança obtidos são razoavelmente grandes, possivelmente pelo fato das probabilidades de captura dos animais serem pequenas. Em geral, esta é uma característica dos modelos do método de captura-recaptura.

5.5 Exemplo com dados reais

Nesse exemplo determinamos estimativas do tamanho de uma população de cabritos que habitavam, em 1986, as montanhas do deserto de Mojave, USA. Os dados se encontram em He e Ananda (2002). Para este conjunto de dados consideramos o caso em que as probabilidades de captura são diferentes. Este conjunto de dados é descrito na Tabela 5.

Tabela 5 - Dados reais sobre um estudo de cabritos de montanha (1986)

i	0	1	2	3	4	5	6	7	8	9	10	11	12
n_i	13	26	19	32	4	29	34	29	32	13	5	34	35
m_i	0	3	4	4	2	5	9	9	7	7	0	8	4

Fonte: He e Ananda (2002).

Na Tabela 6 são dadas as estimativas \hat{N} , ($= \hat{N}_c$) e \hat{N}_A .

O intervalo de confiança aproximado de 95% para N obtido a partir da expressão (10) é dado por (50, 1; 78, 6). Observamos que, para o tamanho populacional a estimativa de máxima verossimilhança é igual a estimativa de máxima verossimilhança condicional. Como observado anteriormente, a estimativa de máxima verossimilhança está muito próxima da estimativa obtida por Ananda (1997).

Tabela 6 - Estimativas clássicas de N

Método	\hat{N} ($= \hat{N}_c$)	\hat{N}_A
Estimativa pontual de N	63,9	61,2

Conclusões

Os resultados obtidos neste artigo, pelos exemplos com dados simulados, nos permitem concluir que as metodologias da máxima verossimilhança, da máxima verossimilhança condicional e de Ananda (1997) produziram estimativas aproximadamente iguais para o tamanho populacional e para as probabilidades de captura. Por outro lado, quando as probabilidades de captura são muito baixas e, portanto, poucos animais são capturados durante todo o processo, verificamos que essas estimativas são inadequadas e no caso de nenhum animal marcado ser observado ($\sum_{i=1}^s m_i = 0$) essas estimativas não existem. Gostaríamos de ressaltar dois fatos. Primeiro, pequenas probabilidades de captura implicam estimativas não muito boas para o tamanho populacional e para as probabilidades de captura também é uma característica dos diferentes modelos da metodologia de captura-recaptura tradicional. Para uma discussão sobre esse assunto ver, por exemplo, White et al. (1982). Segundo, três fontes de variação nas probabilidades de captura podem comprometer a acurácia das estimativas, quando da aplicação do método de captura-recaptura: (a) as probabilidades de captura variam de amostra para amostra; (b) as probabilidades de captura em amostras subsequentes variam para determinados animais por causa de respostas comportamentais ao processo de captura inicial e (c) heterogeneidade individual nas probabilidades de captura, como pela idade ou sexo do animal. Neste artigo tratamos do primeiro caso.

Agradecimentos

O primeiro autor agradece o financiamento parcial da Capes. O segundo autor agradece o financiamento do CNPq – Bolsa de Produtividade em Pesquisa, Processo 302813/2004-7. Os autores agradecem aos revisores pelas valiosas sugestões apresentadas.

SHIMIZU, G. Y.; LEITE, J. G.; TANAKA, N. I. Estimation of population size by a single tagging capture-recapture process. *Rev. Mat. Estat.*, São Paulo, v.24, n.4, p.35-49, 2006.

- **ABSTRACT:** We consider the estimation of the size of a closed animal population and of the related capture probabilities, based on a capture-recapture process with a single tagging stage. Such process consists in tagging a number of randomly selected animals, replacing them and in each one of several samplings done subsequently in the population, counting both marked and non marked animals. No animals are marked after the first sampling. In some cases this method is cheaper and easier to implement than the usual capture-recapture process. We obtain maximum and conditional maximum likelihood estimates for the population size, for the capture probabilities and approximate confidence interval for the population size. We present examples with both simulated and real data.
- **KEYWORDS:** Closed population; single tagging capture-recapture process; maximum and conditional maximum likelihood estimates; confidence interval.

Referências

- ANANDA, M. M. A. Bayesian methods for mark-resighting surveys. *Commun. Stat. – Theory Methods*, New York, v.26, n.3, p.685-697, 1997.
- CAMPOS, J. J. B.; ALMEIDA, H. G. G.; IOCHIDA, L. C. Incidência de diabetes mellitus dependente (tipo I) na cidade de Londrina, PR-Brasil. *Arq. Bras. Endocrinal. Metabol.*, São Paulo, v.42, p.36-44, 1998.
- FIENBERG, S. E. The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika*, London, v.59, p.591-603, 1972.
- HE, W.; ANANDA, M. M. A. Estimation of population size in closed animal populations from mark-resighting surveys. *Appl. Math. Comput.*, Reading, v.125, p.387-398, 2002.
- MICHELETTI, L. R. *Aplicação da metodologia de verossimilhança na prevalência do diabetes*. 2003. 48f. Dissertação (Mestrado em Estatística) – Centro de Ciências Exatas e Tecnologia, Departamento de Estatística, Universidade Federal de São Carlos, São Carlos. 2003.
- NAYAK, T. K. Estimating population size by recapture sampling. *Biometrika*, London, v.75, p.113-120, 1988.
- POLLOCK, K. H. Modeling capture, recapture and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present and future. *J. Am. Stat. Assoc.*, New York, v.86, n.413, p.225-238, 1991.
- SANATHANAN, L. P. Estimating the size of a multinomial population. *Ann. Math. Stat.*, An Arbor, v.43, n.1, p.142-152, 1972a.
- SANATHANAN, L. P. Models and estimation methods in visual scanning experiments. *Technometrics*, Washington, v.14, n.4, p.813-829, 1972b.
- SEBER, G. A. F. *The Estimation of Animal Abundance*. 2nd. ed., New York: Macmillan, 1982. 654p.

SHIMIZU, G. Y. *Estimação do tamanho de uma população fechada: método de captura-recaptura com um único estágio de marcação*. 2002. 50f. Dissertação (Mestrado em Estatística) – Centro de Ciências Exatas e Tecnologia, Departamento de Estatística, Universidade Federal de São Carlos, São Carlos. 2002.

WHITE, G. C. et al. *Capture-recapture and removal methods for sampling closed populations*. New Mexico: Los Alamos National Laboratory, 1982. 235p.

Recebido em 09.05.2006.

Aprovado após revisão em 08.01.2007.