ESTIMAÇÃO PONTUAL PARA MODELOS DE CAPTURA-RECAPTURA A TEMPO DISCRETO.

Ruan da Silva Vianna[1], Hugo Henrique Kegler dos Santos[2], Douglas Rodrigues Pinto[3]

Introdução

O processo de amostragem caracterizado como captura-recaptura é frequentemente utilizado para estimar o tamanho populacional de uma dada região. Ele consiste em selecionar uma amostra de elementos de uma população, marcá-los e registrá-los, e devolvê-los à população de origem, dando-lhes tempo para que se misturem novamente aos demais. Uma segunda amostra é retirada, conta-se o número de elementos marcados e não marcados, marcando os sem marcas e depois, os indivíduos são devolvidos a população. Essas seleções, ou épocas de amostragem, repetem-se até que seja feita a estimação do tamanho populacional.

O método foi inicialmente proposto por Laplace (1783), cujo o objetivo era estimar o número de habitantes da França. Petersen (1991) utilizou o método, dessa vez em zoologia, estudando o fluxo migratório de peixes no mar Báltico. Lincoln (1930) utilizou o processo para estimar o tamanho populacional de patos selvagens na América do Norte. Esses pesquisadores utilizaram o processo com apenas duas épocas de amostragem. O estimador pontual para o método em duas etapas ficou conhecido como Lincoln-Petersen e se baseia em igualar a proporção de elementos marcados na primeira etapa com a proporção de elementos marcados na segunda etapa.

Os diversos modelos de captura-recaptura se diferenciam a partir das suposições sobre a população, sobre os tempos de captura, e sobre o método de captura.

Objetivos

Descrever, ilustrar e sequenciar as fases de um processo de captura-recaptura para o tempo discreto. Ressalta-se também como objetivos específicos, estudar os estimadores de

1 Universidade Federal Fluminense (UFF), ruanvianna@id.uff.br
2 Universidade Federal Fluminense (UFF), hugosantos@id.uff.br
3 Universidade Federal Fluminense (UFF), douglasrp@id.uff.br

máxima verossimilhança para o tamanho populacional de todos os modelos que são apresentados ao longo do texto.

Material e Métodos

São apresentados os modelos de captura-recaptura para populações fechadas que surgem em consequência do método adotado e também do comportamento da população frente a captura. Ao todo três tipos de modelos são avaliados: modelo hipergeométrico, modelos binomial sem mudança de probabilidade e modelo binomial com mudança de probabilidade. As notações abaixo são usadas nos três modelos.

- N é o número de elementos da população.
- K é o número de etapas do processo.
- n_i é o número de elementos capturados na i-ésima etapa do processo.
- u_i é o número de marcados na etapa i-ésima etapa do processo.
- m_i é o número de remarcados na etapa i-ésima etapa do processo.
- p é a probabilidade de um elemento ser selecionado na primeira vez.
- c é probabilidade de um elemento ser selecionado uma I-ésima vez.
- D_{k+1} é o total de elementos distintos marcados em todas as etapas.

Agora é apresentado o logaritmo do núcleo da função de verossimilhança dos trêsmodelos citados em relação a N, que é o parâmetro que deseja-se estimar.

1.Modelo Hipergeométrico: $\log(K(N))=-(k-1)\log(N!)-\log((N-D_{k+1})!)+\Sigma^{k}_{i=1}\log((N-n_i)!)$.

2.Modelo Binomial sem Mudança de Probabilidade: $\log(K(N,p))=\log(N!)-\log((N-D_{k+1})!)+\Sigma^{k}_{i=1}n_i\log(p)+\Sigma^{k}_{i=1}(N-n_i)\log(1-p)$.

3.Modelo Binomial com Mudança de Probabilidade: $\log(K(N,p,c))=\log(N!)-\log((N-D_{k+1}!)+\Sigma^{k}_{i=1}u_i\log(p)+\Sigma^{k}_{i=1}m_i\log(c)+\Sigma^{k}_{i=1}(N-D_i-u_i)\log(1-p)+\Sigma^{k}_{i=1}(D_i-m_i)\log(1-c)$.

Dados as três funções de verossimilhança, é observado que as duas últimas apresentam parâmetros de pertubação, ou seja, a função de verossímil é para um vetor de parâmetros e como o objetivo ´estimar N, os outros denominam-se parâmetros de pertubação.

Resultados e Discussão

Ao tentar estimar os três estimadores de máxima verossimilhança, observou-se que nenhum possuía uma forma fechada, então foi necessário auxílio de métodos numéricos para aproximação dos mesmos, utilizando-se aproximações via Monte Carlo (MC) com 1000 replicações em cada modelo. Para simulação foi utilizado o software R como instrumento de análise. Para encontrar o máximo feito o uso da função optmise que retorna o máximo.

Após esta etapa, calculou-se os limites inferiores e superiores do intervalo de confiança de Wald, considerando o nível de significância de 5%.

São apresentadas três tabelas em forma de figura com as estimativas simuladas usando MC.

$N = 10000$

Cenário		Média de $\hat{N}$	DP de $\hat{N}$	L. Inferior	L. Superior
I	n = 15 e k = 15	8763.34	1898.21	8643.88	8882.80
II	n = 20 e k = 20	9229.68	1461.96	9138.61	9320.75
III	n = 20 e k = 30	9529.94	1242.39	9452.94	9606.95
IV	n = 30 e k = 20	9580.12	1205.02	9504.90	9655.33
V	n = 30 e k = 30	9847.61	1045.13	9782.83	9912.38

$N = 10000$

Cenário		Média de $\hat{N}$	DP de $\hat{N}$	L. Inferior	L. Superior
I	p = 0.01 e k = 30	3600.21	42.77	3597.55	3602.87
II	p = 0.05 e k = 30	8853.96	39.90	8851.48	8856.43
III	p = 0.05 e k = 50	10072.61	33.20	10070.55	10074.68
IV	p = 0.1 e k = 30	10091.71	23.91	10090.22	10093.19
V	p = 0.1 e k = 50	10016.34	6.94	10015.91	10016.78

$N = 10000$

Cenário		Média de $\hat{N}$	DP de $\hat{N}$	L. Inferior	L. Superior
I	p = 0.01, c = 0.05 e k = 20	4670.85	40.05	4625.52	4716.17
II	p = 0.05, c = 0.01 e k = 20	8687.10	9.84	8675.96	8698.23
III	p = 0.05, c = 0.1 e k = 30	9478.48	5.86	9471.84	9485.11
IV	p = 0.1, c = 0.05 e k = 30	9879.98	1.56	9878.20	9881.75
V	p = 0.1, c = 0.01 e k = 40	9959.33	0.85	9958.36	9960.29

Figura 1: Estimativas Médias, DP e Limites do Intervalo de Confiança dos três modelos apresentados respectivamente.

Conclusão

Considerando o que foi discutido, pode-se afirmar que não existe um fator que contribua para uma melhor estimação, e sim um conjunto de fatores. Não necessariamente o número de etapas de um processo vai produzir melhores resultados.

Em todos os cenários abordados, o intervalo de confiança não contempla o real valor do tamanho populacional em 95% das vezes, contudo, nota-se que as estimativas são próximas de N. Os dois últimos modelos apresentam baixa dispersão entre as estimativas, corroborando em o estimador não ser viciado

De fato, os modelos estão estimando bem o tamanho populacional, porém é necessário a continuação do estudo para poder evidenciar se os estimadores são viesados ou não, e se são eficientes.

Referências

[1] LAPLACE, P. S. Sur les naissances, les marriages, et les morts. Historie de L'Academie Royale des Sciences, 1783.

[2] LINCOLN, F. C. Calculating waterfowl abundance on the basis of banding returns. 1930.

[3] POLLOCK, K. H. Modeling capture, recapture and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present and future. Journal of the American Statistical Association, 1991.

[4] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2017. Dispon´ıvel em: <https://www.R-project.org/>.