

Jolly-Seber방법을 이용한 로버스트 설계

김지혜¹⁾, 홍태경²⁾, 최진식³⁾, 남궁 평⁴⁾

요 약

개방형 모집단에 대한 Mark-Recapture방법은 일반적으로 Jolly-Seber방법이 사용된다. 이 방법은 각 표본에서 모든 동물들에 대한 포획될 확률은 동일하다는 가정을 갖는다. 개방형 모집단에 대한 Mark-Recapture방법을 소개하고 폐쇄형 모집단에서 갖는 문제를 해결하기 위한 개방형 모집단의 로버스트 설계를 이용하였다. Pollock과 Kendall의 Jolly-Seber 모수로부터 유도된 모집단의 성장률 추정에 대하여 SAS와 POPAN의 결과를 비교하였다.

주요용어 : Mark-Recapture, Jolly-Seber, 로버스트, POPAN

1. 서론

Mark-Recapture방법은 이동성이 강한 개체들로 이루어진 모집단을 목표 모집단으로 한다. 이 방법은 폐쇄형 모집단에 대한 경우와 개방형 모집단에서 표식된 개체로부터 일정 기간동안의 생존을 추정하는 Cormack-Jolly-Seber 모형(Cormack 1964: Jolly 1965: Seber 1965)과 유입과 생존에 대해서 추정하는 Jolly-Seber 모형(Jolly 1965: Seber 1965)으로 나눌 수 있다.

본 논문에서는 연구 지역이 지리적으로 폐쇄되었으나 인구 통계학적인 개체 증감을 허용하는 개방형 모집단에서의 새로운 개체의 유입과 생존이 관심의 대상이므로 Jolly-Seber 모형의 모집단 성장률을 알아보고 모수를 추정해 보고자 한다. 특히, 모집단 성장률을 추정함에 있어 Jolly-Seber 방법의 가정이 동등한 포획력과 덧 반응에 대해 이질성을 지니고 있기에 현실적인 문제를 가지고 있으므로 이에 대한 대안으로 로버스트 설계를 적용하여 보기로 한다.

2. Jolly-Seber 방법과 로버스트 설계

2.1. Jolly-Seber 방법

표본 시간 i 전에 표식된 모집단의 추정된 크기를 M_i , 표본 시간 i 바로 전에 추정된 모집단 크기를 N_i , 생존율을 ϕ_i , 모집단의 추가비율을 B_i , 포획력을 p_i 라고 표시하면 근사추정량들은 다음과 같다.

$$\bar{M}_i = m_i + \frac{(R_i + 1)z_i}{r_i + 1} \quad (2.1)$$

-
- 1) 성균관대학교, 일반대학원 통계학과, 석사과정
 - 2) 용인대학교, 컴퓨터 정보처리학과 강사 E-mail : hongstat@skku.edu
 - 3) 동우대학, 컴퓨터정보과 교수 E-mail : jschoi@duc.ac.kr
 - 4) 성균관대학교, 통계학과 교수 E-mail : namkung@skku.ac.kr

Jolly-Seber방법을 이용한 로버스트 설계

$$\bar{N}_i = \frac{(n_i + 1) \bar{M}_i}{m_i + 1} \quad (2.2)$$

$$\bar{\phi}_i = \frac{\bar{M}_{i+1}}{\bar{M}_i + (R_i - m_i)} \quad (2.3)$$

$$\bar{B}_i = \bar{N}_{i+1} - \bar{\phi}_i [\bar{N}_i - (n_i - R_i)] \quad (2.4)$$

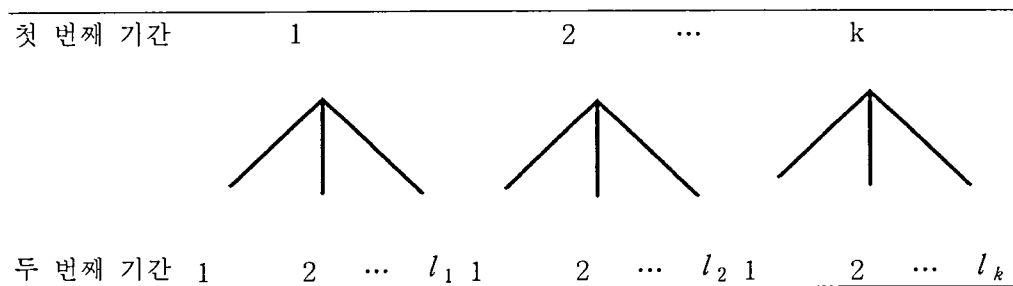
$$\bar{p}_i = \frac{m_i}{\bar{M}_i} \quad (2.5)$$

여기서 m_i 는 i 번째 표본에서 포획된 동물 중 표식이 있는 동물 수를, n_i 는 i 번째 표본에서 포획된 총 동물 수, R_i 는 i 번째 표본 이후에 풀어 준 n_i 의 수, r_i 는 i 에서 풀어준 동물 R_i 중 다시 포획된 동물 수, z_i 는 i 이전에 포획되어 i 에서는 포획되지 않고 그 뒤에 다시 포획되는 동물 수를 의미한다.

2.2 로버스트 설계

로버스트 설계(Pollock 1982)는 개방형 모집단과 폐쇄 모집단이 결합된 혼합 설계이다.

<그림 2-1>은 초기 k 개의 추출시점(예: 년, 월)을 가진 Mark-Recapture 추출 실험을 묘사하고 있다. k 개 각각의 기간 안의 두 번째 추출 시점 l_i 는 각각의 기간 동안 폐쇄되어 있다. 또한, 모든 초기 기간 안의 두 번째 기간의 수는 동등할 필요는 없다.



<그림 2-1> 로버스트 설계 방법

설계는 모집단이 각각의 초기 추출기간 안의 두 번째 추출기간 동안 일정하다고 가정한 각각의 초기 추출기간에 대해 모집단 크기 N_1, N_2, \dots, N_k 와 생존율 $\phi_1, \phi_2, \dots, \phi_{k-2}$ 그리고 초기 추출기간 사이에 모집단에 진입하는 새로운 개체 수 B_1, B_2, \dots, B_{k-2} 를 추정할 수 있다. 이러한 방법의 각각의 추정량은 다음과 같다.

$$p_i^* = 1 - \prod_{j=1}^{l_i} q_{ij} \quad (2.6)$$

$$\mathcal{U}_i = \frac{u_i}{p_i^*}, \quad i=1, 2, \dots, k \quad (2.7)$$

$$\mathcal{M}_i = \frac{m_i}{p_i^*}, \quad i=2, \dots, k \quad (2.8)$$

$$\mathcal{N}_i = \mathcal{U}_i + \mathcal{M}_i, \quad i=1, 2, \dots, k \quad (\mathcal{N}_1 = \mathcal{U}_1) \quad (2.9)$$

$$\mathcal{B}_i = \mathcal{N}_{i+1} + \mathcal{O}_i \mathcal{N}_i, \quad i=1, 2, \dots, k-1 \quad (2.10)$$

3. 모집단 증가율 추정

모집단은 공간과 시간에 따라 크기가 변화하므로 모집단 증가율은 다음과 같이 나타낸다.

$$\begin{aligned} \Delta N &= \text{증가} - \text{감소} \\ \text{모집단의 순증가} &= \frac{N_{i+1}}{N_i} = \lambda \quad i=2, \dots, k-2 \end{aligned} \quad (3.1)$$

또한, B_i 는 앞선 정의에 의해 i 에는 존재하지 않으나, $i+1$ 에는 존재하는 수이므로 다음과 같은 관계가 성립한다.

$$N_{i+1} = N_i \Phi_i + B_i \quad (3.2)$$

$$\lambda_i = \frac{B_i}{N_i} + \Phi_i \quad (3.3)$$

만약, N_{i+1} 이 시점 $i+1$ 에 모집단의 개체 수라면, $\frac{B_i}{N_{i+1}}$ 는 모집단의 개체가 i 와 $i+1$ 사이
에 모집단에 진입하는 비율이다. 따라서 $\frac{B_i}{N_{i+1}}$ 가 개체가 모집단에 진입하는 확률이라면,
 $1 - \frac{B_i}{N_{i+1}}$ 는 개체가 이미 모집단에 있을 확률이다. 개체가 $i-1$ 에 존재하였으며, 시점 i 모집
단에 있을 상위 확률 γ_i 는 $N_{i+1} = N_i \Phi_i + B_i$ 이므로 다음이 성립된다.

$$\lambda_i = \frac{\Phi_i}{\gamma_{i+1}} \quad (3.4)$$

Φ_i 와 γ_{i+1} 을 추정하기 위하여 개체 수 N 없이도 λ_i 의 추정량들은 추정할 수 있다. 즉,
시점 i 모집단에 이미 존재하는 개체 수에 대한 i 와 $i+1$ 사이의 모집단에 새로 진입하는 개
체수의 비로써 조사를 시작할 때 초기에 생존한 구성원에 대한 다음과 같은 구간 진입비

Jolly-Seber방법을 이용한 로버스트 설계

$f_i = \frac{B_i}{N_i}$ 를 사용하면 $\lambda_i = \phi_i + f_i$ 이므로 γ_{i+1} 는 다음과 같다.

$$\gamma_{i+1} = \frac{\phi_i}{\phi_i + f_i} \quad (3.5)$$

4. MARK 자료를 이용한 모집단 증가율 추정

4.1 POPAN을 이용한 모집단 증가율 추정

POPAN을 이용하여 자료 분석을 하였다. POPAN은 "POPulation ANalysis" 의 약자로서 모집단 분석을 위한 소프트웨어이다. 여기서는 POPAN에 의한 모수 추정과 SAS에 의한 모수 추정에 대한 모집단 증가율을 비교해 보았다.

<표 4-1> POPAN을 이용한 모집단 증가율 λ_i

시 점	모집단 순증가		B_i 에 의한 증가		ϕ_i 와 γ_{i+1} 에 의한 증가	
	λ_i	$s. e (\lambda_i)$	λ_i	$s. e (\lambda_i)$	λ_i	$s. e (\lambda_i)$
1						
2	1.7179	0.1153	1.6854	0.1138	1.6854	0.1138
3	1.1160	0.0693	1.1442	0.0709	1.1442	0.0709
4	1.4101	0.0663	1.3956	0.0662	1.3956	0.0662
5						
6						

4.2 SAS를 이용한 모집단 증가율 추정

표식된 모집단의 크기 M 의 추정량 $\hat{M}_i = \frac{R_{i+1}z_i}{r_{i+1}} + m_i$ 을 이용하여 모집단 크기 \hat{N}_i 과 모집단 증가율 그리고 생존률과 개체수를 구한 후 모집단 증가율을 구해 보면 <표 4-2>와 같다.

<표 4-2> \hat{M}_i 를 이용한 모집단 증가율 λ_i

시 점	모집단 순증가	B_i 에 의한 증가	ϕ_i 와 γ_{i+1} 에 의한 증가
1			
2	1.6991	1.6991	1.6991
3	1.1145	1.1144	1.1144
4	1.4084	1.4084	1.4084
5	0.6995	0.6996	0.6996
6			

$\hat{M}_i = m_i + \frac{R_i z_i}{r_i}$ 을 이용한 모집단 증가율은 <표 4-3>과 같다.

<표 4-3> \widehat{M}_i'' 를 이용한 모집단 증가율 λ_i

시 점	모집단 순증가	B_i 에 의한 증가	ϕ_i 와 γ_{i+1} 에 의한 증가
1			
2	1.7000	1.6996	1.6996
3	1.1141	1.1142	1.1142
4	1.4083	1.4083	1.4083
5	0.6994	0.6995	0.6995
6			

다음은 로버스트 설계 하에서 추정량들에 대한 모집단 증가율을 구해 보고자 한다. 로버스트 설계하의 모집단 크기는 $\widehat{M}_i = \frac{m_i}{\widehat{p}_i}$ ($i=2, \dots, k$)이며, \widehat{p}_i^* 는 $1 - \prod_{j=1}^i \left(1 - \frac{m_j}{\widehat{M}_j}\right)$ 이다.

위 통계량은 포획력이 시점마다 다르다는 것을 이용하여 그 시점에 맞는 포획력을 모수에 적용한 것이며, 본 논문에서 가장 주목해서 보고자하는 추정량들이다. 또한, 포획력을 구하기 위한 추정량 \widehat{M}_i 은 <표 4-2>에서 구한 추정량을 사용하였다.

<표 4-4> 로버스트 설계의 \widehat{M}_i' 를 이용한 모집단 증가율 λ_i

시 점	모집단 순증가	B_i 에 의한 증가	ϕ_i 와 γ_{i+1} 에 의한 증가
1			
2	1.6991	3.6359	3.6359
3	1.1145	2.6381	2.6381
4	1.4084	3.2706	3.2706
5	0.6995	2.2443	2.2443
6			

<표 4-4>와 마찬가지로 <표 4-3>에서 사용된 표식 된 개체 수 추정량 \widehat{M}_i 은 \widehat{p}_i^* 를 추정하는데 사용하면 다음과 같은 결과가 나온다.

<표 4-5> 로버스트 설계의 \widehat{M}_i'' 를 이용한 모집단 증가율 λ_i

시 점	모집단 순증가	B_i 에 의한 증가	ϕ_i 와 γ_{i+1} 에 의한 증가
1			
2	1.6997	3.6362	3.6362
3	1.1141	2.6376	2.6376
4	1.4083	3.2705	3.2705
5	0.6994	2.2440	2.2440
6			

5. 결론

로버스트 설계 하에서 구해진 모집단 증가율은 로버스트 설계를 하지 않았을 때보다 1시점

(분석된 자료의 경우 6시점 중 5시점)이 더 추정될 수 있었다. POPAN으로는 Jolly-Seber 방법의 기본 모수만 추정할 수 있으나 표준오차를 쉽게 구할 수 있다. 그러나 관심을 가졌던 모집단 증가율에 대해서는 SAS를 이용할 수밖에 없었다. 전체적으로 모집단 순증가는 로버스트 설계와 로버스트 설계가 아닌 경우 모두 비슷한 결과가 나왔으나, \hat{B}_i , $\hat{\phi}_i$ 그리고 $\hat{\gamma}_{i+1}$ 의 모집단 증가율은 로버스트 설계가 약 2정도 크게 추정되었는데, 그 원인은 추정에 사용된 추정량들을 통해 추론해 볼 수 있다. 즉, 추정식에 사용된 추정량 \hat{B}_i , $\hat{\phi}_i$ 그리고 $\hat{\gamma}_{i+1}$ 의 영향으로 이러한 결과가 나타났다고 볼 수 있다.

참고 문헌

- [1] 남궁 평 (1999) , 최신 표본조사설계와 분석, 탐진
- [2] Arnason, A.N and Schwarz, C.J. (1999). POPAN-5 (Euring 1997) paper: Using POPAN-5 to analyse banding data. 46 (suppl.), S157-168
- [3] Carl J. Schwarz and J. Brian Dempson (1994), Mark-Recapture Estimation of a Salmon Smolt Population, , 50, 98-108, March.
- [4] Carl J. Schwarz (2001), The Jolly-Seber Model : More Than Just Abundance. Volume 6, Number 2, Pages 195-205
- [5] David L. Borchers, Walter Zucchini, and Rachel M. Fewster (1998), Mark-Recapture Models for Line Transect Surveys, 54, 1207-1220 December.
- [6] James D. Nichols & James E. Hines (2002), Approaches for the direct estimation of , and demographic contributions to , using capture-recapture data, Vol. 29, Nos. 1-4, 2002, 539-568
- [7] Kenneth H. Pollock (1982), A capture-recapture design robust to unequal probability of capture, 46(3).
- [8] R. Pradel (1996), Utilization of Capture-Mark-Recapture for the Study of Recruitment and Population Growth Rate, 52, 703-709,
- [9] Schwarz, C.J. and Stobo, W.T. (1997), Estimating Temporary Migration using the Robust Design, , 53, 178-194
- [10] Trent L. McDonald and Steven C. Amstrup (2001), Estimation of Population Size Using Open Capture-Recapture Models, Volume 6, Number 2, Pages 206-220.
- [11] William L. Kendall, Kenneth H. Pollock and Cavell Brownie (1995), A Likelihood-Based Approach to Capture-Recapture Estimation of Demographic Parameters Under the Robust Design, 51, 293-308 March 1995.