7 기타 표집관련 기법

7.1 포획-재포획 추정(capture recapture estimation)

N : 모집단 크기, n_1 : 첫 번째 표본의 크기, n_2 : 두 번째 표본의 크기, m : 두 번째 표본에서 꼬리 달린 개체의 수, $y_i = 1\,(i=1,\ \cdots, N)$, $x_i = \begin{cases} 1, & i\text{번째 물고기에 고리표가 달렸으면} \\ 0, & i\text{번째 물고기에 고리표가 없으면} \end{cases}$

$$\tau_y = \sum_{i=1}^{N} y_i = N,\ \ \tau_x = \sum_{i=1}^{N} x_i = n_1, \quad \text{관측값 : } \left\{(y_1, x_1),\ \cdots, (y_{n_2},\ x_{n_2})\right\}$$

$$\sum_{i=1}^{n_2} y_i = n_2,\ \ \sum_{i=1}^{n_2} x_i = m,\ \ b = \frac{\sum_{i=1}^{n_2} y_i}{\sum_{i=1}^{n_2} x_i} = \frac{n_2}{m}$$

가정 1) 경과 시간에 상관없이 모집단 크기는 항상 일전(단힌 모집단)

2) 각 표본은 모집단으로부터 단순임의추출, 두 표본은 서로 독립

- $\hat{N} = \hat{\tau}_y = b\tau_x = \dfrac{n_1 n_2}{m}$ (편향추정량),

$$\widehat{Var}(\hat{N}) = (\tau_x)^2\, \widehat{Var}(b) = (\tau_x)^2 \frac{N - n_2}{N} \frac{s_e^2}{n_2 \bar{x}^2} \qquad \text{(유한모집단수정 무시)}$$

$$= n_1^2 \frac{1}{n_2 (m/n_2)^2} \frac{n_2(n_2 - m)}{m(n_2 - 1)}$$

$$\bar{x} = \frac{m}{n_2},$$

$$s_e^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - bx_i)^2 = \frac{1}{n_2 - 1}\left\{ m\left(1 - \frac{n_2}{m}(1)\right)^2 + (n_2 - m)\left(1 - \frac{n_2}{m}(0)\right)^2 \right\} = \frac{n_2(n_2 - m)}{m(n_2 - 1)}$$

$$\widehat{Var}(\hat{N}) \approx \frac{n_1^2 n_2 (n_2 - m)}{m^3} \quad (n_2 \approx n_2 - 1)$$

- $N^* = \dfrac{(n_1 + 1)(n_2 + 1)}{m + 1} - 1$(덜 편향추정량), $\widehat{Var}(N^*) = \dfrac{(n_1+1)(n_2+1)(n_1 - m)(n_2 - m)}{(m+1)^2 (m+2)}$

예 7.1 $n_1 = 300,\ n_2 = 100,\ m = 10$

$\hat{N} = 3000,\ \ \widehat{Var}(\hat{N}) = 810000, \widehat{se}(\hat{N}) = 900$

$N^* = 2763,\ \ \widehat{Var}(N^*) = 546464,\ \ \ \widehat{se}(N^*) = 739$

- $\hat{N} = \hat{\tau}_y = b\tau_x = \dfrac{n_1 n_2}{m},\ \ \widehat{Var}(\hat{N}) = \dfrac{n_1^2 n_2 (n_2 - m)}{m^2 (m+1)}$ (역표집)

7.2 사각표집

관심 지역에 존재하는 개체들이 총수 : N

총면적 : $A = Ka$, 관심지역의 면적 a의 K개의 사각으로 분할한다.

$y_i (i = 1, \cdots, k)$: 사각 안에 존재하는 관심 개체들의 수

표본밀도 : $\hat{\lambda} = \dfrac{\sum_{i=1}^{k} y_i}{ka} = \dfrac{\bar{y}}{a}$

- $\hat{N} = \hat{\lambda} A$, $\widehat{Var}(\hat{\lambda}) = \left(\dfrac{1}{a}\right)^2 \widehat{Var}(\bar{y}) = \left(\dfrac{1}{a}\right)^2 \left(\dfrac{K-k}{K}\right)\dfrac{s^2}{k} \approx \left(\dfrac{1}{a}\right)^2 \dfrac{s^2}{k}$, $s^2 = \dfrac{1}{k-1}\sum_{i=1}^{k}(y_i - \bar{y})^2$

$$\widehat{Var}(\hat{N}) = K^2 \frac{s^2}{k}$$

- 포아송 과정

$$\widehat{Var}(\hat{N}) = A^2 \frac{\hat{\lambda}}{ak}$$

- 적재사각표집(stocked-quadrat sampling)

$$\hat{N} = \hat{\lambda} A, \ \hat{\lambda} = -\left(\frac{1}{a}\right)\log\left(\frac{q}{k}\right) \ (q \text{ : 비적재사각의 개수}), \ \widehat{Var}(\hat{N}) = \frac{K^2}{k}(\exp(a\hat{\lambda}) - 1)$$

예 7.2 남산 숲 A=20만평 a=1000평 단위의 K=200개 구획으로 분할한 뒤 k=20개 구획들을 임의 표집 하여 토끼의 숫자를 세었다.

$\bar{y} = \dfrac{0(10) + 1(4) + 2(3) + 3(2) + 4(1)}{20} = 1$, $s^2 = 1.58$, $\hat{\lambda} = 1/1000 = 0.001$,

1) $\hat{N} = \hat{\lambda} A = 0.001 \times 200000 = 200$, $\widehat{Var}(\hat{N}) = K^2 \dfrac{s^2}{k} = (2000)^2 \dfrac{1.58}{20} = 3160$

2) 포아송 모형 $\widehat{Var}(\hat{N}) = A^2 \dfrac{\hat{\lambda}}{ak} = (200000)^2 \dfrac{0.001}{1000(20)} = 2000$

3) 적재사각표집 $\hat{\lambda} = -\left(\dfrac{1}{a}\right)\log\left(\dfrac{q}{k}\right) = -\dfrac{1}{1000} log \dfrac{10}{20} = 0.000693$, $\hat{N} = \hat{\lambda} A = 0.000693 \times 200000 = 139$

$$\widehat{Var}(\hat{N}) = \frac{K^2}{k}(\exp(a\hat{\lambda}) - 1) = \frac{200^2}{20}(\exp(1000 \times 0.000693) - 1) = 2000$$

7.3 임의화 반응

$\pi = p(\text{카드1을 선택})$, $\phi = p(\text{예})$

$$\phi = p(\text{예}) = p(\text{예}|\text{카드}\,1)p(\text{카드}\,1) + p(\text{예}|\text{카드}\,2)p(\text{카드}\,2)$$

$$= p_s \pi + (1 - p_s)(1 - \pi)$$

- $\hat{p}_s = \dfrac{\hat{\phi} - (1-\pi)}{2\pi - 1} (\pi \neq 1/2)$, $\hat{\phi} = m/n$, $\widehat{Var}(\hat{p}_s) = \dfrac{1}{(2\pi-1)^2 n^2}\widehat{Var}(m) = \dfrac{1}{(2\pi-1)^2}\dfrac{1}{n}\left(\dfrac{m}{n}\right)\left(1 - \dfrac{m}{n}\right)$

예. 카드 1 : 나는 외도 그룹입니다.(70%)

카드 2 : 나는 일편단심 그룹이다.

300명중에서 “예”응답이 186명이다

$\hat{p}_s = \dfrac{\hat{\phi} - (1-\pi)}{2\pi - 1} \dfrac{186/300 - (1 - 0.7)}{2(0.7) - 1} = 0.8$, $\widehat{Var}(\hat{p}_s) = \dfrac{1}{(2(0.7)-1)^2}\dfrac{1}{300}(0.62)(0.38) = 0.005$

예7.3

카드1 : ‘예’라고 대답하시오. (100장)

카드2 : ‘아니오’라고 대답하시오. (100장)

카드 3 : 외도 하십니까 ? (300장)

$\phi = p(\text{예}) = p(\text{예}|\text{카드}1 \text{ or } 2)p(\text{카드}1 \text{ or } 2) + p(\text{예}|\text{카드}3)p(\text{카드}3)$

$= p_I(1-\pi) + p_s\pi, \quad p_I = p(\text{예}|\text{카드}1 \text{ or } 2)$

- $\hat{p_s} = \dfrac{\hat{\phi} - (1-\pi)p_I}{\pi}$, $\hat{\phi} = m/n$, $\widehat{Var}(\hat{p_s})\dfrac{\widehat{Var}(\hat{\phi})}{\pi^2} = \dfrac{1}{\pi^2}\dfrac{\hat{\phi}(1-\hat{\phi})}{n} = \dfrac{1}{\pi}\dfrac{1}{n}\left(\dfrac{m}{n}\right)\left(1-\dfrac{m}{n}\right)$

$\hat{p_s} = \dfrac{0.68-(1-0.6)(0.5)}{0.6} = 0.8(\hat{\phi} = 204/300 = 0.68,\ \pi = 30/50 = 0.6,\ p_I = 10/20 = 0.5)$

$\widehat{se}(\hat{p_s}) = \sqrt{\dfrac{1}{0.6^2}\dfrac{0.68(0.32)}{300}} = 0.045$

예 7.4 (2) 동전을 던지십시오. 앞면이 나왔습니까 ?

임의화 기구로서 8개 빨간 구슬과 2개의 흰 구슬이 들어있는 주머니와 100원 짜리 동전을 준비.

100명의 응답자 중 30명이 “예”라고 대답하였다면, 혼외정사를 하는 사람의 비율은 얼마인가?

$\hat{\phi} = 30/100 = 0.3,\ \pi = 8/10 = 0.8,\ p_I = 1/2 = 0.5$

$\hat{p_s} = \dfrac{0.3-(1-0.8)(0.5)}{0.8} = 0.25,\ \widehat{se}(\hat{p_s}) = \sqrt{\dfrac{1}{0.8^2}\dfrac{0.3(0.7)}{100}} = 0.057$

7.4 상호관입표집(interpenetrating sampling)과 임의그룹법(random group method)

$$\hat{\theta} = \sum_{i=1}^{R}\hat{\theta}_i/R,\ \widehat{Var}(\hat{\theta}) = s_r^2/R = \frac{1}{R}\frac{\sum_{i=1}^{R}(\hat{\theta}_i - \hat{\theta})^2}{R-1}$$

예 7.5 재학생 수 10000명인 으뜸대학교에서 1400명을 단순임의표집하여 키를 측정한다.

$\hat{\mu} = \dfrac{171+\cdots+172}{7} = 171.43,\ \widehat{se}(\hat{\mu}) = \sqrt{\dfrac{1}{7}\dfrac{(171-171.43)^2+\cdots+(172.-171.43)^2}{7-1}} = 0.48$

7.5 드문 사건에 대한 표집

7.5.1 불비례배정 층화표집(disproportional allocation)

7.5.2 이상표집(two-phase sampling)

7.5.3 다중틀 조사(multiple frame survey)

7.5.4 망조사(network sampling)

7.5.5 눈덩이 조사(snowball sampling)

7.6 복잡한 조사설계