김일성종합대학학보 (철학, 경제학)
주체107(2018)년 제64권 제1호

과거의 산생특성분석을 위한 산생률거꿀추정방법

허 광 림

위대한 령도자 **김정일**동지께서는 다음과 같이 교시하시였다.

《통계분석사업을 강화하는것은 통계사업을 혁신하는데서 나서는 중요한 과업입니다.》

당정책과 현실발전의 요구에 맞게 통계자료에 대한 분석사업을 결정적으로 개선강화하는것은 사회주의강국건설에서 나서는 중요한 문제의 하나이다. 특히 인구집단의 상태와 그 변화를 반영하는 다양한 인구통계자료는 나라의 경제관리를 개선하고 인민생활을 향상시키는데서 기초적이면서도 매우 중요한 의의를 가지는 자료로서 그에 대한 분석을 과학적으로 진행하는것은 사회경제발전에서 중요한 요구로 나선다.

인구통계자료의 필수적인 구성부분의 하나로 되는 산생자료를 과학적으로 분석하는것은 현재의 산생수준과 앞으로의 변화에 영향을 미치는 요인들을 정확히 파악하여 미래의 산생수준을 예견하고 그에 따라 국가와 사회발전에 필요한 인구규모를 보장하기 위한 적극적이며 현실성있는 대책을 세우는데 이바지하는것으로 하여 통계자료분석에서 중요한 문제로 된다.

일반적으로 미래의 산생수준에 대한 예측은 과거산생수준의 변화상태와 그에 영향을 미친 요인들에 대한 분석을 통하여 얻을수 있다.

산생률거꿀추정방법은 과거의 산생특성을 분석하는데 리용되는 방법의 하나이다.

산생률거꿀추정방법은 본질에 있어서 어느 한 시점에서의 나이별인구수자료가 주어졌을 때 해당 나이의 인구집단의 출생년도로 거슬러 올라가면서 지난 년도들의 산생수준을 추정하는 방법이다.

인구이동이 없는 인구집단에서 x나이의 인구수는 $x-n$년전에 출생한 사람들이 $x+n$살까지 생존한 수이다. 그러므로 출생후 x살까지의 생존확률을 추정할수 있다고 가정한다면 현시점에서의 x나이인구수에 기초하여 $x-n$년전의 출생자수를 추정할수 있다.

산생률거꿀추정방법은 이러한 원리에 기초하여 조사시점 이전년도들의 출생률이나 일반산생률, 총산생률과 같은 산생관련지표값들을 추정하는 방법이다.

산생률거꿀추정방법의 특징은 조사를 통하여 보고되는 나이자료의 정확성과 출생한 어린이에 대한 보고에서의 루락정도에 따라 추정의 정확성이 크게 달라지게 된다는것이다.

조사에서 응답자의 나이보고가 정확하지 않은것으로 하여 생기는 나이분포에서의 오차는 추정되는 산생률결과에 큰 영향을 미칠수 있다. 이로부터 산생률거꿀추정방법은 자료가 주어진 시점에서 10년이나 15년전까지의 기간의 산생률추정에 주로 적용된다. 그보다 더 오랜 기간에 대해서는 자료오차범위가 더 커지는것으로 하여 추정결과의 정확성이 크게 떨어지게 된다.

인구조사에서 응답자들이 여러가지 요인으로 하여 출생한 어린이를 모두 보고하지 않고 그 일부를 응답에서 빠뜨리는 경우에는 그에 기초하여 얻어진 산생률추정값이 실제

값보다 낮아지게 된다. 그러므로 산생률거꿀추정방법을 적용하기 위하여서는 사전에 성별, 나이별자료의 정확성에 대한 평가를 진행하여야 한다.

과거의 산생특성은 우선 년도별출생률을 거꿀추정하는 방법으로 특징지을수 있다.

년도별출생률을 거꿀추정하기 위하여서는 5살간격으로 된 15살미만의 인구수, 두 시점 또는 한 시점에서 계산된 총인구수와 인구장성률, 생명표에서의 어린이생존지표(구체적으로는 생명표에서의 ${}_5L_0$, ${}_5L_5$, ${}_5L_{10}$)값자료가 있어야 한다. 이와 함께 계산을 간단히 하기 위하여 연구되는 인구집단에서는 적어도 거꿀추정이 진행되는 기간에는 인구이동이 없다고 가정한다. 추정을 위하여 요구되는 자료와 가정에 기초하여 5살나이간격의 출생률은 다음과 같은 방법으로 추정할수 있다.

첫째로, 생명표를 리용하여 어린이생존지표들을 계산한다.

세 나이구간(0−4, 5−9, 10−14살)에서 인구의 출생률을 거꿀추정하기 위하여 먼저 해당 나이구간들에서의 인년수(${}_5L_0$, ${}_5L_5$, ${}_5L_{10}$)들을 계산한다. 이 값을 계산하기 위하여 연구되는 인구집단의 경험적인 사망자료나 적합한 모형생명표의 사망률자료를 리용할수 있다. 이때 매 나이구간안에서의 사망수준은 일정하다고 가정한다.

만약 사망수준이 과대추정되면 산생수준도 과대추정되며 반대로 과소추정되면 산생수준도 과소추정된다. 출생후 첫해의 사망변화가 다른 나이구간에 비하여 큰 경우에는 ${}_5L_0$을 ${}_1L_0$와 ${}_4L_0$의 합으로 계산할수도 있다.

둘째로, 중간년인구수를 추정한다.

5살나이간격으로 된 자료에서 첫 세구간(0−4, 5−9, 10−14살)의 년도별출생률을 추정하기 위해서는 그 세나이구간의 중간시점에서의 총인구수(중간년인구수)를 추정하여야 한다. 조사시점으로부터 d년전(우의 세나이구간에서 보면 2.5, 7.5, 12.5년전)의 총인구수인 $N_{(t-d)}$를 추정하는 가장 쉬운 방법은 인구장성률 r가 고정되여있다는 가정하에서 이 장성률을 t시점에서의 총인구수에 적용하는것이다. 즉

$$N_{(t-d)}=N_{(t)}\times\exp(-d\times r)$$

이다.

장성률 r는 두 시점(t_0, t_1)에서의 총인구수(P_0, P_1)자료가 주어졌을 때 아래와 같이 계산할수 있다.

$$r=\ln(P_1/P_0)/(t_1-t_0)$$

이에 기초하여 조사전 2.5년전의 인구수는

$$N_{(t-2.5)}=N_{(t)}\times\exp(-2.5\times r)$$

와 같으며 다른 시점에서의 인구수도 이와 같은 방법으로 계산할수 있다.

셋째로, 조사전 매 5년기간에서의 출생자수를 추정한다.

$B_{(t-5,\ t)}$를 조사전 t−5년부터 t년 기간에서의 매해 출생자수라고 하면 우에서 설정된 세나이구간에 해당한 t값은 각각 0, 5, 10로 된다. 또한 ${}_5N_{x(t)}$를 조사시점에서 x살부터 x+5살나이의 인구수라고 하면 $t-5$년부터 t년사이의 매 5년구간에서의 평균출생자수는 다음과 같다.

$$B_{(t-5,\ t)}={}_5N_{0(t)}\times l_0/{}_5L_0$$

$$B_{(t-10,\ t-5)}={}_5N_{5(t)}\times l_0/{}_5L_5$$

$$B_{(t-15,\ t-10)}={}_5N_{10(t)}\times l_0/{}_5L_{10}$$

다음 매 나이구간에서의 출생률(CBR)은 매 나이구간에서의 출생자수를 대응되는 중간년인구수로 나누어 계산한다. 즉

$$CBR_{(t-5,\ t)}=B_{(t-5,\ t)}/N_{(t-2.5)}$$

이다.

나이별인구수자료가 1년간격으로 주어져있다면 이와 같은 방법으로 년도별출생률을 추정할수 있다.

과거의 산생특성은 또한 일반산생률과 총산생률을 거꿀추정하는 방법으로 특징지을수 있다.

일반산생률은 총출생자수를 산생적령기녀성수로 나눈 값이다. 그러므로 일반산생률을 계산하기 위하여서는 성인사망자료에 기초하여 현 시점에서의 산생적령기녀성수로부터 과거시점에서의 산생적령기녀성수를 추정하여야 한다.

그리고 총산생률을 추정하기 위하여서는 주어진 해의 총출생자수에 대한 추정값과 함께 산생적령기녀성들의 산생나이에 대한 자료가 있어야 한다. 이 자료를 얻을수 있는 보다 쉬운 방법은 산생나이분포에 대한 독립적인 추정값을 리용하여 총출생자수를 어머니나이집단들에 배분하는것이다.

년도별일반산생률을 추정하기 위하여서는 한살간격의 0－14살나이구간 인구수, 5살간격의 15－64살나이구간 녀성인구수, 0－14살나이 어린이들의 동시집단생존확률(L_x), 조사전 15년기간의 매 5년구간별 성인녀성들의 생존비(${}_5L_{x\text{-}5}/{}_5L_x$)자료가 요구된다.

이와 함께 총산생률을 추정하기 위하여서는 하나 또는 두개의 나이별산생분포가 요구되는데 한개의 산생분포를 리용하는 경우는 추정되는 전체 기간에 대하여 산생분포가 일정하다고 보는 경우이고 두개의 산생분포를 리용하는 경우는 조사시점과 추정하려는 시점에서 산생분포가 서로 다르다고 보는 경우이다.

일반산생률이나 총산생률의 추정값은 산생분포에 따라 추정된 출생자수와 그에 대응되는 산생적령기녀성수에 의하여 결정된다. 여기서도 거꿀추정이 진행되는 기간에 한해서는 인구이동이 없다고 가정한다.

이와 같은 자료와 가정에 기초하여 일반산생률과 총산생률은 다음과 같은 방법으로 추정할수 있다.

첫째로, 조사전 기간에 대한 년도별출생자수를 추정한다.

어느 조사에서나 x나이의 인구수는 조사전 x+0.5년을 중간시점으로 하는 12달기간에 태여난 출생자들중에서 조사시점까지 생존해있는 사람들의 수를 반영한다. 이것을 공식으로 표시하면 아래와 같다.

$$B_{x+0.5}=N_x/{}^cL_x,\ (0\leqq x\leqq 14)$$

이 공식에서 리용된 생존지표 ${}^cL_x$는 동시발생집단의 인년수이다. 이 값은 조사시점이 전년도들로부터 조사시점년도로 넘어오면서 주어진 해의 사망수준에 따라 달라진다. 주어진 조사자료가 일제조사자료인 경우에 이 값들은 태여난적이 있는 어린이수와 생존어린이수에 기초하여 간접추정할수 있으며 산생조사자료가 주어진 경우에는 동시발생집단의 산생경력에 대한 분석으로부터 직접 얻을수도 있다.

조사이전년도들에 대한 산생분포를 얻기 위하여 곰퍼츠모형을 리용할수 있는데 이때 t기간의 산생분포는 다음과 같이 표시할수 있다.(여기서 T=0은 조사전 0－4년기간을, T=5

는 조사전 5－9년기간을, T=10은 조사전 10－14년기간을 의미한다.)

$$Y_{x,\ T}=\alpha_T+\beta_T\times Y^s_x$$

여기서 Y는 로지스틱함수로서

$$Y_x=(1/2)\times\ln((1-l(x))/l(x)) \quad (1)$$

로 표시된다.

l_x는 l_0의 값이 1(생명표에서 이 값은 보통 100 000으로 계산)인 생명표에서 출생후 만 x살까지 생존하는 비률을 의미한다. 따라서 방정식 (1)은

$$l_x=(1+\exp(2Y_x))^{-1}\text{로 되며 } l_{x,\ T}=(1+\exp(2(\alpha_T+\beta_T\times Y^s_x)))^{-1} \quad (2)$$

로 된다.

유아시기가 아닌 다른 나이구간에서 T기간의 x살부터 x+1살사이의 생존인년수를 표시하는 $L_{x,\ T}$는 로지스틱척도에서 선형적으로 감소하며 $L_{x,\ T}$가 $Y_{x,\ T}$와 $Y_{x+1,\ T}$의 평균이라는 가정에 기초하여 근사하게 계산될수 있다. 이것을 방정식 (2)에 기초하여 표시하면 다음과 같다.

$$L_{x,\ T}\approx l_{x+0.5,\ T}=(1+\exp(2((Y_{x,\ T}+Y_{x+1,\ T})/2)))^{-1}=(1+\exp(Y_{x,\ T}+Y_{x+1,\ T}))^{-1},\ (0<x\leqq 14) \quad (3)$$

모형생명표에서 l_x의 값들은 5살까지는 한살간격으로 표시되고 그 이상나이부터는 5살간격으로 표시된다. 실례로 생명표가 x=5과 x=10에서의 값을 가지고있을 때 $L_{9,\ t}$의 추정값은 다음과 같이 표시할수 있다.

$$L_{9,\ T}\approx l_{9.5,\ T}=(1+\exp(2((1/10)Y_{5,\ T}+(9/10)Y_{10,\ T})))^{-1}=(1+\exp(0.2\,Y_{5,\ T}+1.8\,Y_{10,\ T}))^{-1}$$

유아들의 사망이 주로 출생후 첫 몇일이나 몇주기간에 집중된다는것을 고려할 때 사망수준이 중간정도이거나 높은 수준에 있는 인구집단에서 출생후 첫해의 생존인년수는 대략적으로 다음과 같이 표시할수 있다.

$$L_{0,\ T}=0.3+0.7(\exp(2Y_{1,\ T}))^{-1} \quad (4)$$

T기간에 한나이에서 그 다음나이까지 생존하는 비인 $P_{x,\ T}$는 $L_{x,\ T}$를 리용하여 얻을수 있다.

$$P_{x,\ T}=L_{x,\ T}/L_{x-1,\ T},\ (0<x\leqq 14)$$

$$P_{0,\ T}=L_{0,\ T} \quad (5)$$

한살나이간격과 5년간격으로 된 어린이생존에 대한 추정값이 우에서 설명한 방법이나 기타 다른 방법으로 주어지면 한살간격의 동시발생집단의 년도별생존추정값은 다음과 같이 계산할수 있다.

$P_{a,\ T}$를 T기간에서의 a살과 a+1살사이의 생존비라고 하고 조사전 0－4년기간을 T=0으로, 조사전 5－9년기간을 T=5로, 조사전 10－14년기간을 T=10으로 표시하자. 그리고 $S_{a,\ T}$를 T기간에서 T+1기간까지에서의 a살과 a+1살사이의 생존비라고 하자.($0<t\leqq 14$) 선형보간법을 리용하여 중간년들의 생존비를 추정하면 다음과 같다.

$$S_{a,\ T}=P_{a,\ 0},\ (0\leqq t\leqq 2)$$

$$S_{a,\ T}=P_{a,\ T}(1-(t-T-2)/5)+P_{a,\ T+5}((t-T-2)/5),\ (2<t-T<8)$$

$$S_{a,\ T}=P_{a,\ 10},\ (12<t<15)$$

이에 기초하여 조사전 x년부터 x+1년기간에 출생한 출생자들가운데서 조사시점까지 생존하는 비률인 ${}^cL_x$는 다음과 같이 계산한다.

$${}^cL_x=S_{0,\ x}S_{1,\ x-1}\cdots S_{x-1,\ 1}S_{x,\ 0} \quad (6)$$

조사전 매 년도에서의 출생자수는 그해의 중간시점에 집중되므로 다음과 같이 계산할수 있다.

$$B_{x+0.5}=N_x/{}^cL_x,\ (0\leqq x\leqq 14)$$

여기서 N_x는 조사에서 보고된 x나이의 어린이수이다.

둘째로, 5살간격의 나이집단별녀성들의 중간년인구수를 추정한다.

조사시점에서의 15－64살나이 녀성들의 생존비는 사망수준이 성인나이의 중심구간에서 보통 낮기때문에 쉽게 계산할수 있다. 그러므로 사망수준에 대한 근사한 추정값이라고 하여도 그것을 리용하여 현재의 인구수로부터 과거시점의 인구수에 대한 만족한 추정값을 얻을수 있다.

일반적으로 성인나이구간에서의 임의의 5살간격의 나이집단에서 사망의 절대적인 변동은 작으며 따라서 Y_x와 Y_{x+5}사이를 선형보간하는 방법으로 ${}_5L_x$의 근사값을 얻을수 있다. 이 값에 기초하여 T시점(T=5, 10, 15)에서 5살간격의 나이집단들사이의 생존비를 추정할수 있다.

$${}_5P_{x,\ t}={}_5L_{x+5}/{}_5L_x=(1+\exp(2\alpha T+\beta_T(Y^s_x+Y^s_{x+5})))/(1+\exp(2\alpha_T+\beta_T(Y^s_{x+5}+Y^s_{x+10}))) \quad (7)$$

T=0에서 계산된 인구수를 가지고 시작할 때 조사진행 T+5년전에 5살간격의 매 나이집단에서의 녀성들의 수는 다음과 같이 계산할수 있다.

$${}_5N^f_{x,\ T=5}={}_5N^f_{x+5,\ T}/{}_5P_{x.\ t},\ (10\leqq \mathrm{x}\leqq 60,\ T=0,\ 5,\ 10) \quad (8)$$

인구의 나이구조는 일반적으로 천천히 변하므로 5살나이간격으로 된 10－64살나이구간에서 조사전 매해의 중간년도 녀성인구수는 방정식 (8)에서 계산된 조사진행 0, 5, 10, 15년전의 인구추정값들사이의 선형보간을 통하여 얻을수 있다. 실례로 조사전 8.5년시점에서 20－24살나이구간의 녀성수는 다음과 같이 계산할수 있다.

$${}_5N_{20,\ 8.5}=0.3\times{}_5N_{20,\ 5}+0.7\times{}_5N_{20,\ 10}$$

셋째로, 계산된 자료에 기초하여 일반산생률을 추정한다.

조사전 x-0.5년시점에 중심을 둔 주어진 해의 일반산생률(GER)은 다음과 같이 계산한다.

$$GER_{x+0.5}=B_{x+0.5}/\sum_{a=15,\ 5}^{45}{}_5N_{a,\ x+0.5},\ (0\leqq x\leqq 14) \quad (9)$$

여기서 분모는 출생이 일어난 기간의 중간년시점에서의 15－49살사이 녀성들의 총수이다.

넷째로, 나이별산생비와 총산생률을 추정한다.

총산생률을 추정하기 위해서는 간접표준화와 류사한 방법으로 연구되는 인구집단의 산생나이분포형태를 리용하여야 한다.

이러한 산생나이분포형태를 설명하는 자료는 거꿀추정방법을 리용하여 산생분석이 진행되는 조사자료에서 수집된 최근 출생자수에 대한 자료에서 얻을수 있다. 같은 인구에 대한 이전의 일제조사나 선택조사를 통해서 이전시기의 산생형태를 얻을수 있다면 두 시점에서의 산생나이형태사이를 보간하거나 필요한 경우에 보외법을 리용하여 년도별총산생률을 추정해볼수도 있다.

그러나 산생수준이 시간에 따라 변한다고 하여 언제나 두개의 산생나이분포형태가 요구되지는 않는다. 일반적으로 산생나이분포형태는 짧은 기간동안에 급격히 변하지 않고 점차적으로 변화하게 되므로 총산생률추정값은 산생분포형태에 대한 가정의 정확성에 크게 영향을 받지 않는다. 따라서 보통 한개의 산생분포형태를 가지고도 거꿀추정을 진행할

수 있다. 이 경우에는 산생수준이 추정되는 기간의 중간시점에서의 산생분포형태가 주어지면 보다 합리적이라고 볼수 있다.

근사한 수학적모형을 리용하여 산생률이 추정되면 표준으로 선택된 산생나이분포형태에 기초하여 얻어진 모형파라메터 α와 β는 산생나이분포형태를 규정한다. 이 α와 β값들은 총산생률추정에 리용되는 산생나이분포의 형태를 결정하는데 리용될수 있다. 조사시점이전의 매 년도들에 대하여 매 나이집단에서 일어나는 나이집단별산생률이 추정되면 이 비률들은 매 년도별로 매 나이집단의 녀성인구에 적용되여 그 나이집단녀성들이 낳은 어린이수를 추정하는데 리용된다. 이때 총산생률은 녀성 1인당 어린이 1명과 같다.

총산생률이 1인 나이별산생분포형태가 규정되면 조사시점으로부터 15년전까지기간의 매 년도별로 매 나이집단들(a=15, 20, …, 45)에 속하는 녀성들이 낳는 예측된 어린이수는 다음과 같이 계산할수 있다.

$$ {}_5B^*_{a,\ x+0.5} = {}_5N_{a,\ x+0.5} \times {}_5f^*_{a,\ x+0.5},\ (0 \leqq x \leqq 14) $$

따라서 총산생률이 1일 때 x년도의 총출생자수는 다음과 같다.

$$ B^*_{x+0.5} = \sum_{a=15}^{45} {}_5N_{a,\ x+0.5} \times {}_5f^*_{a,\ x+0.5},\ (0 \leqq x \leqq 14) $$

첫째 단계에서 얻은 년도별 실제적인 출생자수에 대한 추정값인 $B_{x+0.5}$를 $B^*_{x+0.5}$로 나누면 총산생률의 추정값이 얻어진다.

$$ TF_x+0.5 = B_{x+0.5}/B^*_{x+0.5},\ (0 \leqq x \leqq 14) $$

주어진 해의 나이별산생률에 대한 추정값은 x년도의 a나이집단녀성들의 산생률인 ${}_5f^*_{a,\ x}$에 그해의 총산생률추정값을 곱해주는 방법으로 얻을수 있다.

이렇게 추정된 일반산생률과 총산생률을 리용하여 과거의 산생수준이 어떻게 변동되였는가를 파악할수 있으며 이것은 산생수준변화에 미친 요인들의 영향정도를 파악하고 앞으로의 그에 대한 과학적인 예측을 진행하는데 도움을 준다.

산생률의 거꿀추정방법에서 중요하게 제기되는 문제의 하나는 이 방법을 리용하여 얻은 추정결과가 오차를 가지는 중요한 리유가 조사에서 응답자가 나이보고를 틀리게 하거나 보고되여야 할 어린이들이 조사보고에서 루락되는것과 관련되여있다는것이다. 이와 같은 리유들로 하여 조사시점 이전기간의 어린이사망이 과대추정되면 산생도 과대추정되고 반대로 과소추정되면 산생도 과소추정되게 된다. 그러므로 산생분석에서 거꿀추정방법을 리용할 때에는 주어진 자료의 특성과 이 방법의 우단점에 대하여 옳게 파악하고 리용하는것이 중요하다.

이와 같은 지난 시기의 산생특성을 과학적으로 분석하기 위한 방법들을 더 깊이 연구하고 인구통계분석에 적극 활용해나가는것은 사회경제발전과 국가의 인민적시책을 과학적으로 조직집행하여 사회주의강국건설을 다그쳐나가는데 적극 이바지할수 있게 한다.

실마리어 산생특성분석, 산생률거꿀추정