

인구통계학적 분석을 이용한 우리나라의 센서스 및 동태자료에 대한 질적 평가

전새봄 · 김성용 · 박유성

1. Motivation

(1) 인구의 단조증가현상은 지켜지는가?

〈표 1〉 센서스 여성인구

(단위: 천명)

연 도	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44
2000	1,489	1,613	1,449	1,778	1,820	2,040	2,025	2,069	1,966
2005	1,145	1,515	1,619	1,474	1,746	1,814	2,036	2,047	2,040
2010	1,077	1,151	1,518	1,612	1,430	1,736	1,829	2,039	2,060

-. 2000년 15~19세와 35~39세 코호트만 단조 감소. 이는 센서스에서 coverage error가 발생 하였거나 동태자료에서 문제가 있음을 시사하고 있다.

〈표 2〉 성별 연령별 순이민자수

연 령	남 자			여 자		
	2000	2005	2010	2000	2005	2010
0-4	-4,771	-17,721	-17,581	-3,920	-15,242	-15,461
5-9	-3,698	-17,116	-13,053	-2,967	-15,361	-11,998
10-14	-4,938	-27,247	-16,428	-4,188	-22,648	-14,170
15-19	-4,206	-27,591	-18,081	-5,467	-26,202	-19,437
20-24	-8,159	-29,917	-19,384	-9,154	-48,651	-39,853
25-29	-3,433	-26,287	-15,225	-6,240	-21,636	-1,039
30-34	-1,468	-8,198	-871	-3,637	-17,983	-6,701
35-39	-446	-4,885	-1,514	-3,757	-16,895	-8,215
40-44	-868	-4,843	181	-2,991	-15,142	-4,731

(2) 성비의 규칙성은 지켜지는가?

출생성비는 여성 100에 남성 104~106으로부터 출발하여 20세 후반~30세 전반에 비율이 100:100으로 맞추어진 후 이후 점차적으로 여성 100에 남성 50~60으로 감소하게 된다(Poston, 2006). 1995년 0~19세까지 4개의 코호트의 성비가 증가하고 있으며, 15~19세 코호트의 성비가 20~24세에 증가하였다는 것은 20~24세 센서스 인구의 coverage error가 성별로 큰 차이를 보일 가능성이 높다는 것을 의미한다.

〈표3〉 년도별 연령별 성비

연령	연도					
	1985	1990	1995	2000	2005	2010
0-4	1.0803	1.1120	1.1340	1.1021	1.0607	1.1071
5-9	1.0711	1.0727	1.1074	1.1357	1.0921	1.0958
10-14	1.0670	1.0604	1.0643	1.1142	1.1222	1.0900
15-19	1.0662	1.0391	1.0589	1.0766	1.1033	1.1327
20-24	1.0614	1.0915	1.0830	1.1144	1.0972	1.1366
25-29	0.9922	0.9946	1.0092	1.0087	1.0247	1.0384

(3) 한국의 1970~1975년의 센서스 자료를 이용하여 한국의 사망등록률이 62%~65%라고 추정하여 한국의 사망등록시스템의 후진성을 간접적으로 나타내고 있다(Horiuchi(1981)과 Hill(1987)). Murry et al.(2010)은 이러한 오류를 어느 정도 수정했으나 1970~2005년에 걸친 0~4세에 대한 사망등록률을 50% 내외로 추정하는 심각한 오류를 범하고 있다.

2. 연구목적

- (1) 한국의 센서스의 coverage error는 어느 정도인가? 이 coverage error의 추이는 어떠한가?
- (2) 한국의 사망자 등록시스템을 통한 사망자 등록률은 우수한가? 사망자 등록률의 성별, 연령별 패턴과 추이는 어떠한가?
- (3) 센서스별·성별·연령별 coverage error는 어느 정도인가?
- (4) 이러한 의문점들의 답은 앞에서 열거한 비정상적인 코호트별 인구증가 현상과 성비에서 지적된 문제를 해결할 수 있는가?

3. 분석방법

(1) 센서스(census) 오류(error): 포함오류(coverage error)와 내용오류(content error) 포함오류: 가구나 개인의 센서스에서 누락(omission)되거나 중복(duplication)

내용오류: 가구나 개인의 특성에 대한 부정확한 보고나 기록(incorrect reporting or recording).

(2) 오류의 추정방법: 인구통계학적 분석(Demographic Analysis; DA)과 센서스후 조사(Post Enumeration Survey; PES).

DA: 거시적인 관점에서의 오류측정방법으로 출생, 사망, 이민 등의 동태자료(vital statistics)를 이용하여 소위 “인구 방정식”을 이용한 분석방법. 미국은 1960년부터 PES와 함께 사용하고 WHO(World Health Organization)는 약 100여개 국가의 성인사망률(adult mortality)를 점검하기 위해 DA방법이 사용하고 있다(WHO, 2010)

PES: 미시적 관점에서의 오류측정방법(Robinson et al., 1993)이며 표본의 자료를 센서스 자료와 case-by-case로 비교(matching)하여 오류를 추정하는 방법.

- 두 추정치의 불일치는 센서스 결과가 질적 문제를 내포하고 있다는 것을 의미하고 DA나 PES가 방법론적인 문제가 있음을 시사하게 된다(U.S. Census Bureau, 2004).
- 본 연구에서는 5년마다 시행된 한국 센서스 자료(1985~2010년)를 이용하여, 센서스의 포함 오류(coverage error)와 동태자료의 질적 평가를 위한 인구통계학적 분석을 하고자 한다.

4. 인구통계학적 분석방법(Demographic Analysis)

- 센서스의 완성도와 사망자료의 완성도의 측정

(1) General Growth Balance(GGB) 방법

기본가정: 폐쇄인구, 연령분포의 정상성, 인구증가율이 연령에 의존하지 않음, 센서스와 사망자등록의 완성도(completeness)가 연령에 무관하게 동일.

- Hill(1987)은 인접한 두 센서스자료에 대하여 N_1° 와 N_2° 를 각각 첫 번째와 두 번째의 관찰된 센서스 인구라고 정의하고 D° 를 두 센서스 사이에서 발생한 관찰된 사망자수라고 정의한 후,

$$N_1 = N_1^\circ / K_1 \quad N_2 = N_2^\circ / K_2 \quad D = D^\circ / K_3$$

으로 참값(N_1, N_2, D)과 관찰값의 관계를 설정하였다. K_1 : 첫 번째 센서스의 완성도, K_2 : 두 번째 센서스의 완성도, K_3 : 사망자등록 완성도

$$\frac{N^\circ(a)}{N^\circ(a+)} - r^\circ(a+) = \frac{1}{t} \log \frac{K_1}{K_2} + \frac{(K_1 K_2)^{1/2}}{K_3} \frac{D^\circ(a+)}{N^\circ(a+)}$$

의 관계로부터 K_1 대비 K_2 과 K_3 를 구한다.

(2) Synthetic Extinct Generations(SEG) 방법

사망자 자료의 완성도를 측정하기 위한 방법으로

기본가정: 폐쇄인구, 연령분포의 정상성, 인구증가율이 연령에 의존하지 않음, 센서스와 사망자등록의 완성도(completeness)가 연령에 무관하게 동일, $K_1=K_2$.

- Vincent (1951)

$$N_t(a) = \int_a^\infty D_{t+x-a}(x) dx$$

- Bennett and Horiuchi(1981, 1984)는 인구의 연령분포가 정상성이라는 가정에서

$$N_t(a) = \int_a^\infty D_t(x) e^{\int_a^x r(u) du} dx$$

를 도출하였다. $N_t^\circ(a)$ 와 $D_t^\circ(x)$ 를 각각 시점 t에서의 센서스 및 사망자수 관찰치라고 정의할 때

$N_t(a)$ 의 추정치를 $\hat{N}_t(a) = \int_a^\infty D_t^\circ(x) \exp\left[\int_a^x r(u) du\right] dx$ 로 놓으면

$$\begin{aligned} \frac{\hat{N}_t(a)}{N_t^\circ(a)} &= \frac{\int_a^\infty D_t^\circ(x) e^{\int_a^x r(u) du} dx}{N_t^\circ(a)} = \frac{K_3 \int_a^\infty D_t(x) e^{\int_a^x r(u) du} dx}{N_t^\circ(a)} \\ &= \frac{N_t^\circ(a) K_3 / K_1}{N_t^\circ(a)} = \frac{K_3}{K_1} \end{aligned}$$

- $K_1 \neq K_2$ 이면 $r(a)$ 의 추정치가 편향(bias)되어 사망등록의 완성도 K_3 가 편향되게 추정되게 된다(Bennett and Horiuch, 1981; Hill, You and Choi, 2009).

(3) Extension

- Bennett and Horiuch(1981)은 $\hat{N}_t(a)/N_t^*(a)$ 가 연령 a 에 의존하지 않도록 연령별 인구증가율 $r(a)$ 에 적절한 상수를 더해주는 확장된 SEG(extended SEG; ESEG)를 제안하였다.
- Hill, You and Choi(2009)는 GGB-SEG 방법을 제안하였다.

(4) 연령구간의 선택

- GGB, SEG 그리고 GGB-SEG는 모두 연령선호(age preference) 또는 연령쌓음(age heaping)이 없다는 가정과 폐쇄인구 가정으로 인해 인구의 국제이동이 없다는 가정을 하고 있다.
- 이러한 가정을 근사적으로나마 충족시키기 위해서는 필연적으로 모형을 적합할 연령구간의 선택이 매우 중요하다.

Hill, You and Choi(2009): GGB나 SEG의 가정을 위배한 유형에 따라 총 98개의 시나리오 시뮬레이션을 실시한 결과, 5~65세보다는 15~55세가 GGB와 SEG의 적합구간으로 좀 더 좋은 결과를 보였다. 국제이동이 없을 경우, GGB-SEG가 가장 좋은 결과를 보여 주었으며 국제이동이 있는 경우에는 GGB와 SEG가 매우 불안정하여 30~65의 연령 구간을 제안하고 있다.

Murry et. al(2010): 78개의 연령구간을 세 개의 서로 다른 시뮬레이션 환경에 적용하여 GGB는 40~70세, SEG는 55~80세, 그리고 GGB-SEG는 50~70세 연령구간에서 가장 좋은 결과를 보인다고 보고하고 있다.

- 본 연구에서는 Hill, You and Choi(2009)가 제안한 3개의 연령구간과 Murry et. al(2010)이 제안한 3개의 연령구간을 한국자료에 적용하여 이들 연령구간을 비교하고 그 특성을 살펴보고자 한다.

(5) Age/Period/Cohort (APC)모형

Preston, Elo, Foster and Fu(1998)은 연령분포의 정상성이나 폐쇄인구의 가정이 없이 연령별 센서스 완성도를 측정하고 코호트별 인구를 재구축(reconstruction)하기 위한 Age/Period/Cohort (APC)모형을 고려하였다.

전제조건: 다수의 센서스 자료가 존재해야만 하고 사망자와 국제인구이동 자료의 신뢰도가 높다.

〈표 4〉 APC에서 추정해야 할 모수

연 령	1985(τ_1)	1990(τ_2)	1995(τ_3)	2000(τ_4)	2005(τ_5)	2010(τ_6)
0-4(α_1)	γ_1	γ_{18}	γ_{19}	γ_{20}	γ_{21}	γ_{22}
5-9(α_2)	γ_2					
⋮	⋮					
80-84(α_{17})	γ_{17}					

X_{it}° : 센서스 시점 t 에서 관찰된 코호트 i 의 인구수

D_{it} : γ_i 가 측정된 시점부터 t 시점까지 사망한 코호트 i 의 누적사망자수와 누적순이민자수의 합

$$X_{it}^\circ = \alpha_j \tau_t (\gamma_i - D_{it}) + \epsilon_{it}$$

$SSE = \sum_{i,t} (\alpha_j \tau_t (\gamma_i - D_{it}))$ 를 최소화 하는 α_j, τ_t 그리고 γ_i 를 추정.

5. 새로운 인구통계학적 분석방법

(1) Modified SEG(MSEG)

Bennett and Horiuchi(1981, 1984)가 제안한 식을 전개하면

$k \geq 1$ 에 대해

$$\begin{aligned} N(a-k) &= \int_{a-k}^\infty D(x) \exp\left[\int_{a-k}^x r(u) du\right] dx \\ &= \int_a^\infty D(x) \exp\left[\int_a^x r(u) du + \int_{a-k}^a r(u) du\right] dx + \int_{a-k}^a D(x) \exp\left[\int_{a-k}^x r(u) du\right] dx \\ &= N(a) \exp\left[\int_{a-k}^a r(u) du\right] + \int_{a-k}^a D(x) \exp\left[\int_{a-k}^x r(u) du\right] dx \end{aligned}$$

이 된다.

- Bennet and Horiuchi(1981)가 $k = 5$ 를 선택하여 5세단위 인구를 사용한 근본적인 이유는 연령선호(age preference), 연령쌓음(age heaping), 또는 연령올림(age overstatement) 등으로 인한 연령오류(age misreporting)을 최소화하기 위해서이다(United Nations, 1983; Hill, You and Choi, 2009).

⇒ 근사식을 이용한 오류 발생

－ 연령오류가 없다면 $k = 1$ 을 택하여

$$N(a-1) = N(a)\exp[r(a-1)] + D(a-1)\exp[r(a-1)]$$

으로 근사식이 필요 없게 된다.

⇒SEG를 수정된 SEG(Modified SEG; MSEG)라고 정의.

(2) GGB-APC방법

－ APC (Age-Period-Cohort) 모형의 문제점:

(i) X_{it}° 의 값이 작은 연령대(즉, 60대 이후 연령대)의 경우, 기존의 목적함수는 오차가 크게 나타날 가능성이 크다

(ii) 미국의 medicare 자료 등과 같은 보조정보가 충분히 있지 않으면 추정해야 할 모수가 너무 많아 자유도가 너무 작다.

목적 함수를
$$\sum_{i,t} \frac{(X_{it}^\circ - \alpha_j \tau_t (\gamma_j - D_{it}))^2}{X_{it}^\circ}$$
로 하고, 두 번째 문제인 자유도를 늘이기 위해

GGB에 추정된 τ_t 를 이용한다.

⇒ GGB와 결합된 APC 방법을 GGB-APC 명명.

6. 자료 분석

－ 1985년부터 2010년까지 5년마다 실시된 센서스 자료와 통계청이 발표한 월별 사망자 자료, 그리고 2000년부터 제공되는 국제인구이동 자료.

6.1 GGB 와 SEG를 이용한 완성도

<표 5> 30~65세와 40~70세 연령구간에 대한 적합도

		1985-1990	1990-1995	1995-2000	2000-2005	2005-2010
남	30-65	0.975	0.984	0.979	0.971	0.986
	40-70	0.990	0.981	0.977	0.983	0.982
여	30-65	0.980	0.982	0.979	0.972	0.976
	40-70	0.995	0.993	0.986	0.975	0.986

〈표 6〉 첫 번째 센서스 대비 두 번째 센서스 및 사망자 등록률 완성도

성별	연령구간	완성도	1985 ~1990	1990 ~1995	1995 ~2000	2000 ~2005	2005 ~2010
남	30-65	K_2	1.018	0.993	0.974	0.989	1.000
		K_3	1.048	0.971	1.150	1.130	0.969
	40-70	K_2	1.025	0.982	0.979	0.987	0.997
		K_3	1.027	1.016	1.122	1.139	0.988
여	30-65	K_2	0.997	0.988	0.984	0.999	1.005
		K_3	1.080	1.129	1.146	1.103	1.035
	40-70	K_2	0.992	0.981	0.993	0.999	0.999
		K_3	1.088	1.183	1.051	1.106	1.133

- K_2 는 연령구간과 성별에 관계없이 $\pm 3\%$ 내의 값을 가지고 있어 한국 센서스의 완성도는 매우 우수한 것으로 나타났다.
- 사망자 등록률의 완성도는 남자의 경우(30~65세 기준), 1995~2000년과 2000~2005년의 사망자 등록률의 과대산정된 것으로 보이며, 여자의 경우(40~70세 기준) 1990~1995년, 2000~2005년, 그리고 2005~2010년에서 다소 과대산정된 것으로 판단할 수 있다.

〈표 7〉 Bennett and Horiuchi 의 SEG 와 ESEG 에 의한 K_3

		1985-1990	1990-1995	1995-2000	2000-2005	2005-2010
남	SEG	1.053	0.745	0.689	0.702	0.646
	ESEG	0.867	0.745	0.689	0.702	0.544
여	SEG	0.718	0.716	0.705	0.759	0.850
	ESEG	0.615	0.716	0.705	0.759	0.687

- 1985~1990년 남자의 SEG 추정치를 제외하고 모두 K_3 가 1보다 현격하게 낮은 값을 보이고 있다. 이러한 결과는 사망자 등록을 해야만 매장이 가능한 우리나라 사망자 등록 시스템을 고려할 때 거의 불가능한 사망자 등록률의 완성도일 뿐만 아니라 〈표 6〉에서 제시된 GGB의 K_3 와 비교하더라도 문제점이 많은 것으로 보인다.

6.2 연령오류(age misreporting)와 수정된 SEG(MSEG)

〈표 8〉 우리나라의 연령비

성별	연령비	1985	1990	1995	2000	2005	2010
남	5-9	95.7	105.7	87.1	112.5	108.3	88.9
	10-14	108.7	96.3	105.9	86.2	110.7	107.8
	15-19	99.1	104.3	95.7	105.1	87.2	111.3
	20-24	102.7	103.6	110.1	102.1	110.0	89.6
	25-29	107.4	97.4	94.8	100.4	93.5	103.3
	30-34	94.9	112.5	102.7	99.1	105.0	96.6
	35-39	98.2	95.3	112.9	103.4	99.7	104.6
	40-44	93.7	95.7	93.9	112.3	103.4	100.9
	45-49	108.7	95.3	96.7	93.1	111.8	103.3
	50-54	101.0	106.8	94.2	96.5	92.4	110.9
	55-59	89.7	102.2	108.5	94.9	97.0	92.4
	60-64	101.6	87.1	100.2	107.7	95.3	96.4
	65-69	97.2	103.2	87.0	100.3	107.1	96.3
	70-74	92.9	92.6	101.0	86.5	100.2	108.2
75-79	91.3	88.8	88.0	95.6	83.2	95.6	
여	5-9	95.9	106.8	86.3	109.8	109.6	88.7
	10-14	108.8	95.8	107.5	85.5	108.3	109.9
	15-19	98.9	108.0	97.1	108.7	87.6	109.4
	20-24	99.7	96.5	105.0	95.4	106.2	85.4
	25-29	114.0	104.3	99.2	106.1	95.9	106.5
	30-34	92.5	110.8	101.9	98.6	105.5	96.9
	35-39	96.5	94.4	113.6	103.7	100.4	104.9
	40-44	93.7	93.1	92.2	111.6	102.4	101.3
	45-49	106.5	96.1	95.2	93.0	111.8	102.2
	50-54	101.0	104.8	94.4	94.5	92.5	111.2
	55-59	97.4	102.7	106.6	95.3	95.2	92.6
	60-64	100.9	95.5	101.8	106.4	95.5	94.4
	65-69	94.9	102.5	96.6	102.8	106.8	97.0
	70-74	99.5	93.5	102.1	97.3	104.0	108.1
75-79	101.2	99.3	91.7	99.9	96.0	103.4	

〈표 9〉 연령비 점수(ARS)와 연령정밀지수(AAI)

		1985	1990	1995	2000	2005	2010
ARS	남	5.17	5.98	6.93	6.13	7.22	6.30
	녀	4.24	5.05	5.72	5.97	6.16	6.78
AAI		4.70	5.52	6.32	6.05	6.69	6.54

- 남녀 모두 1995년 5~9세의 코호트, 1985년 25~29세 코호트, 그리고 1985년 남성 55~59세 코호트를 제외하고는 100%를 크게 벗어나지 않는 것으로 보인다.
- 1985년 남성 55~59세 코호트는 1950년 20~24 세로 6.25 전쟁에 의한 남성인구의 감소가 주 요인인 것으로 보이며(Poston, 2006),
- 1985년 25~29세 코호트는 1956~1960년생으로 소위 베이비 부머(baby boomer)의 핵심 연령이고,
- 1995년 5~9세 코호트는 1986~1990년 생으로 우리나라의 출산률이 급격하게 줄어든 연령이다.
- 〈표 9〉는 Shryock and Siegel(1976) 가 제안한 연령비 함수(Age Ratio Score; ARS)(=|연령비-100|의 평균)와 Smith(1992)가 제안한 연령정밀지수(Age Accuracy Index; AAI)(= AAI는 남자와 여자의 ARS의 단순평균)
- 여성보다는 남성의 ARS가 약간 크게 나타나고 있으며 2005년 7.22%를 평균적으로 벗어나 있으며 AAI는 최대 6.69% 벗어나 있다.
 - ⇒ 연령비가 100%에서 크게 벗어나는 경우는 코호트적인 특성으로 인한 것으로 인한 것이며, 우리나라는 연령오류가 없거나 심각하지 않은 것으로 판단된다.
- 남성의 경우 1985년 5~9세 코호트가 2000년에 대폭 증가하는 현상, 1985년 15~19세 코호트가 1995년에 대폭 감소하는 현상은 좀더 면밀한 검토가 필요하다.

〈표 10〉 MSEG, GGB-SEG에 의한 사망자 등록률의 완성도

성별	방법	연령	1985 ~1990	1990 ~1995	1995 ~2000	2000 ~2005	2005 ~2010
남	MSEG		1.300	1.016	0.967	0.967	0.916
	GGB-MSEG	30~65	1.079	0.934	1.093	1.091	0.913
	GGB-MSEG	40~70	1.141	1.018	1.049	1.067	0.937
	GGB	30~65	1.048	0.971	1.150	1.130	0.969
	GGB	49~70	1.027	1.016	1.120	1.139	0.988
	Mean		1.119	0.991	1.076	1.079	0.945
여	MSEG		1.110	1.039	1.018	1.056	1.039
	GGB-MSEG	30~65	1.025	1.03	1.12	1.117	0.989
	GGB-MSEG	40~70	1.003	1.135	1.06	1.061	1.047
	GGB	30~65	1.080	1.129	1.146	1.103	1.035
	GGB	49~70	1.088	1.183	1.051	1.106	1.133
	Mean		1.061	1.103	1.079	1.088	1.049

- 1985~1990년의 K_3 를 제외하고는(특히, 남자의 경우) 사망자 등록률의 완성도에 큰 문제가 없는 것으로 보인다.
- SEG의 K_3 값과 비교할 때 큰 차이를 보이고 있어, Bennett and Horiuchi (1981)의 SEG를 적용하기에는 우리나라의 경우 문제가 있는 것으로 보인다.

6.3 GGB-APC에 의한 연도별 연령별 누락률의 추정

〈표 11〉 GGB-APC 모형에서의 연도효과(period effect) τ_i

성 별	연령구간	1985	1990	1995	2000	2005	2010
남	30-65	1.027	1.045	1.038	1.012	1	1.000
	40-70	1.028	1.054	1.035	1.013	1	0.997
여	30-65	1.033	1.029	1.017	1.000	1	1.005
	40-70	1.035	1.027	1.008	1.001	1	0.999

〈표 12〉 연령효과(α_j)의 성별 추정치

연 령	남	녀	연 령	남	녀
0-4	1.017	0.982	45-49	0.989	0.995
5-9	1	1	50-54	0.984	0.981
10-14	1.022	1.025	55-59	0.980	0.982
15-19	1.040	1.028	60-64	0.941	0.946
20-24	1.060	1.010	65-69	0.921	0.963
25-29	0.997	1.017	70-74	0.929	0.988
30-34	0.980	0.978	75-79	0.920	0.998
35-39	0.966	0.992	80-84	0.886	1.004
40-44	0.979	0.999			

- 연령효과는 5~9세를 1로 했을 때의 상대적인 연령효과이다. Preston et al.(1998)이 5~9세를 기준으로 하였듯이, 한국의 경우도 5~9세가 확령인구로 진입하는 연령으로 비교적 정확한 인구통계로 인식되고 있다.
- 남자의 경우 0~24세까지는 과대 산정(overcount) 또는 중복이 있는 것으로 나타났으며, 25세 이후는 과소 산정(undercount) 또는 누락이 있는 것으로 나타났다. 특히 60세 이후에 누락률(omission rate)이 매우 높게 나타난 것은 향후 센서스와 센서스 후 조사의 설계 및 정확도 검증시 유의해야 할 부분이다.

- －. 여성의 경우, 0~4세는 누락이, 5~29세까지는 중복이, 그리고 30세 이후는 누락이 있는 것으로 나타나고 있다. 그러나 여성의 경우, 남성과 달리 60세 이후에 누락률이 높아지는 경향은 나타나지 않는 것으로 보인다.

〈표 13-1〉 인구의 재구성(남성)

연령	1985	1990	1995	2000	2005	2010
0-4	1,861,488	1,584,389	1,757,958	1,587,997	1,215,600	1,125,146
5-9	1,929,789	1,850,806	1,578,300	1,750,496	1,567,859	1,246,572
10-14	2,132,872	1,923,054	1,846,246	1,571,829	1,721,896	1,549,641
15-19	2,127,885	2,122,490	1,915,331	1,837,426	1,541,533	1,700,146
20-24	2,148,266	2,110,039	2,107,729	1,898,323	1,803,142	1,518,729
25-29	2,096,658	2,127,137	2,092,889	2,091,324	1,863,405	1,778,200
30-34	1,597,711	2,071,602	2,105,636	2,075,494	2,072,261	1,854,245
35-39	1,314,126	1,571,485	2,043,056	2,082,036	2,054,397	2,057,829
40-44	1,137,369	1,282,808	1,539,685	2,008,002	2,049,929	2,033,891
45-49	1,064,057	1,091,228	1,245,446	1,501,310	1,964,880	2,019,966
50-54	819,818	1,004,586	1,039,291	1,201,946	1,459,557	1,923,524
55-59	569,216	760,129	938,939	980,945	1,156,051	1,416,198
60-64	457,992	507,551	694,368	864,969	917,018	1,105,602
65-69	316,543	384,363	440,042	620,421	782,563	844,949
70-74	202,267	241,731	306,673	366,207	537,115	688,763
75-79	109,421	134,889	170,924	227,302	287,759	446,604
80-84	39,708	58,786	80,135	107,499	153,680	209,789

〈표 13-2〉 인구의 재구성(여자)

연령	1985	1990	1995	2000	2005	2010
0-4	1,775,318	1,466,717	1,591,994	1,487,320	1,166,239	1,097,769
5-9	1,845,251	1,766,984	1,462,328	1,586,468	1,469,850	1,152,395
10-14	2,070,911	1,840,333	1,764,079	1,457,378	1,563,058	1,454,466
15-19	2,058,452	2,065,064	1,836,408	1,756,791	1,430,094	1,542,040
20-24	2,084,485	2,050,206	2,059,231	1,824,477	1,706,076	1,387,508
25-29	2,012,759	2,074,957	2,043,170	2,048,530	1,798,551	1,698,937
30-34	1,505,803	2,002,417	2,066,604	2,033,429	2,025,018	1,786,068
35-39	1,228,582	1,495,671	1,992,198	2,054,396	2,009,348	2,009,423
40-44	1,090,939	1,216,941	1,485,338	1,978,146	2,028,918	1,994,959
45-49	1,055,479	1,073,894	1,204,216	1,472,520	1,957,659	2,018,040
50-54	900,223	1,033,062	1,054,995	1,189,400	1,457,857	1,946,180
55-59	710,192	873,117	1,007,409	1,033,222	1,172,415	1,444,357
60-64	570,360	677,136	840,711	976,880	1,007,005	1,154,358
65-69	417,604	527,790	634,463	798,872	937,587	976,448
70-74	315,725	364,385	469,812	576,654	742,832	887,149
75-79	201,995	249,947	294,743	392,773	498,828	669,943
80-84	97,619	136,130	171,838	210,042	297,793	405,029

〈표 14-1〉 연도별 연령별 센서스 누락률(남성)

연령	1985	1990	1995	2000	2005	2010
0-4	-3.187%	-8.250%	-3.480%	-3.240%	-1.754%	-1.495%
5-9	-4.718%	-7.413%	-2.989%	-4.420%	-5.221%	0.264%
10-14	-7.691%	-6.398%	-3.530%	-2.674%	-5.199%	-6.364%
15-19	-4.464%	-6.380%	-3.609%	-3.995%	-5.217%	-6.901%
20-24	-1.714%	-8.031%	-5.818%	-6.404%	-5.885%	-6.561%
25-29	3.427%	-1.563%	0.696%	1.653%	0.273%	-1.365%
30-34	0.510%	-3.324%	-1.897%	0.353%	0.599%	-0.651%
35-39	-0.773%	-4.655%	-2.851%	-1.674%	-0.546%	-0.117%
40-44	2.587%	-2.462%	-2.542%	-1.055%	-1.561%	-1.812%
45-49	2.020%	-0.885%	-1.273%	0.348%	0.154%	-1.207%
50-54	1.260%	1.013%	1.011%	1.410%	2.310%	1.883%
55-59	1.540%	-0.114%	1.658%	2.216%	2.578%	4.075%
60-64	3.998%	2.568%	3.065%	3.408%	2.188%	4.595%
65-69	3.206%	2.292%	4.555%	4.453%	3.521%	1.405%
70-74	6.148%	3.610%	4.418%	5.164%	4.448%	2.358%
75-79	5.707%	5.461%	6.496%	7.549%	6.328%	8.735%
80-84	9.802%	7.155%	12.444%	14.196%	12.417%	12.785%

〈표 14-2〉 연도별 연령별 센서스 누락률(여성)

연령	1985	1990	1995	2000	2005	2010
0-4	-0.252%	-5.551%	-0.876%	-0.119%	1.851%	1.941%
5-9	-2.419%	-5.180%	-0.467%	-1.621%	-2.958%	0.089%
10-14	-4.364%	-5.011%	-1.896%	0.548%	-3.430%	-4.202%
15-19	-1.460%	-5.353%	-2.134%	-1.176%	-2.988%	-4.354%
20-24	1.220%	-2.465%	-0.349%	0.247%	-2.299%	-2.975%
25-29	-1.491%	-4.494%	-0.793%	0.435%	-0.825%	-2.143%
30-34	-1.299%	-3.025%	-0.829%	0.415%	-0.557%	-2.345%
35-39	-2.246%	-3.692%	-1.903%	-0.728%	-1.845%	-1.446%
40-44	1.123%	-0.584%	-0.397%	0.571%	-0.573%	-3.157%
45-49	0.885%	-0.189%	0.119%	1.140%	0.960%	-0.526%
50-54	1.647%	1.729%	1.944%	2.093%	2.041%	1.886%
55-59	0.426%	1.306%	1.775%	2.422%	1.822%	2.732%
60-64	0.683%	2.253%	2.356%	2.572%	1.567%	2.591%
65-69	0.360%	0.615%	1.823%	2.138%	1.457%	-0.253%
70-74	1.617%	0.712%	0.206%	1.186%	0.588%	-0.669%
75-79	-3.155%	0.273%	-0.146%	0.905%	0.522%	-0.549%
80-84	-3.821%	-3.076%	-1.764%	0.200%	0.758%	-1.092%

－ 남성: 50세를 기준으로 50세보다 젊은 경우에는 대부분 중복(과대산정)이 발생하고, 50세 이상의 나이에서는 누락(과소산정)이 발생했다. 80~84세의 누락률은 심각할 정도로 크게 나타났다. 2010년: 55~64세, 75~84세의 누락률이 다른 해보다 높게 나타났다.

－ 여성: 45세를 기준으로 45세 이하의 중복이, 그 이상은 누락이 많은 것으로 나타났다. 특히

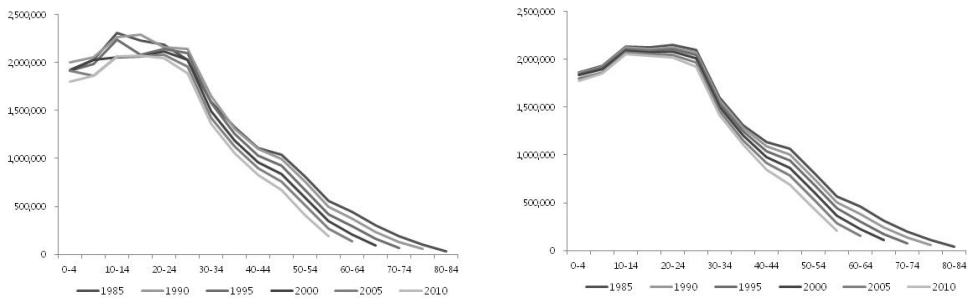
2010년 센서스의 경우 65세 이상이 모두 중복이 일어나고 있으며, 이는 다른 센서스에서는 나타나지 않는 현상이다. 이와 같이 남성과 여성 모두에서 2010년 누락률의 이상현상이 2010년에 처음으로 실시된 인터넷 센서스의 영향이 아닌지 면밀한 검토가 필요하다.

- 연령비의 이상현상은? 1985년 5~9세 남성 코호트가 2000년에 대폭상승. 1985년 남성 15~19세 남성코호트가 1995년에 대폭감소.

6.4 인구 특성의 이상 현상은 해결되었는가?

(1) 센서스

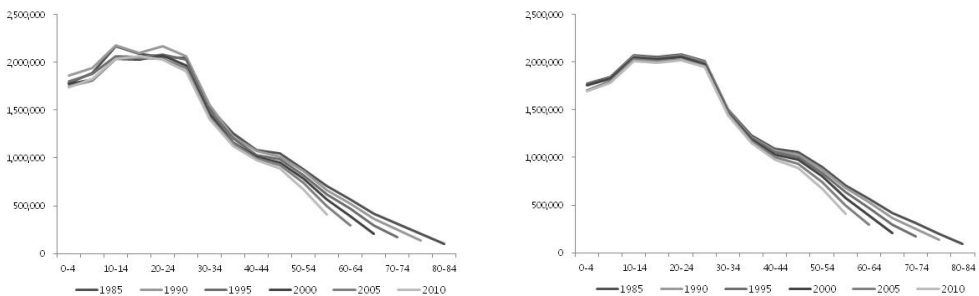
(2) 재구성



〈그림 2-1〉 연도별 연령별 인구(남성)

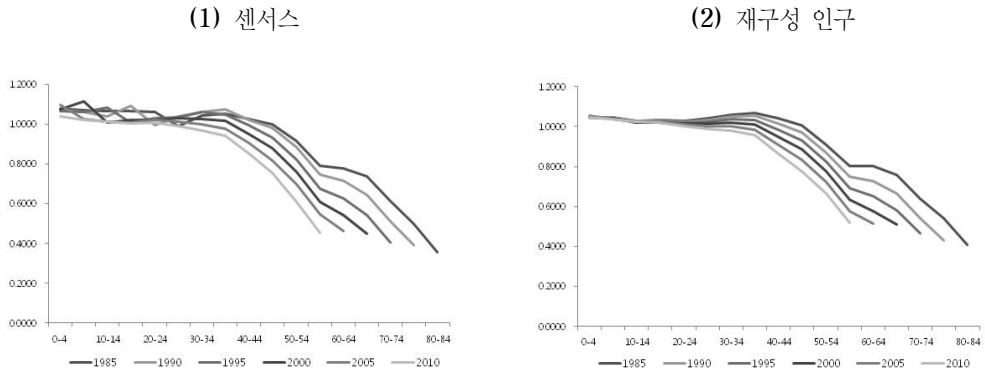
(1) 센서스

(2) 재구성



〈그림 2-2〉 연도별 연령별 인구(여성)

- 센서스 자료의 경우 남녀 모두 45세 미만에서 역전현상(cross-over)가 발생하지만 재구성한 인구의 경우 이와 같은 역전현상이 발생하지 않음을 알 수 있다.



〈그림 3〉 연도별 연령별 성비

〈표 16〉 0~4세의 사망자 등록률의 완성도

성 별	방 법	연 령	1985 -1990	1990 -1995	1995 - 2000	2000 -2005	2005 -2010
남	MSEG		1.297	0.773	0.802	0.791	0.838
	GGB-MSEG	30-65	0.912	0.606	1.004	0.949	0.832
	GGB-MSEG	40-70	1.015	0.753	0.924	0.925	0.873
여	MSEG		1.081	0.772	0.821	0.917	0.971
	GGB-MSEG	30-65	0.958	0.741	1.042	1.009	0.890
	GGB-MSEG	40-70	0.928	0.940	0.889	0.922	0.986

- Murray et al. (2010)이 추정 한 우리나라 사망자 등록률의 완성도인 K_3 가 0.5 내외라는 연구결과는 〈표 16〉을 보면 타당한 결과라고 보기에는 어려워 보인다.
- 그러나 1990~1995년의 K_3 의 값이 특히 작게 나오는 현상은 문제가 있는 것으로 보이며, 이로 인해 〈표 14〉에서 1995년 남녀 모두 0~4세의 중복률이 높게 나오는 것으로 판단된다.

7. 결론

- (1) GGB 방법에 적합한 연령구간은 30~65세 또는 40~70세 연령구간으로 나타났으며,
- (2) Bennett and Horiuchi (1981, 1984)가 제안한 SEG 방법은 근사식의 문제점이 있는 것으로 나타났으며, MSEG가 좀 더 타당한 것으로 나타났다.
- (3) GGB에 의한 센서스 완성도는 비교적 만족할 만한 수준으로 나타났으며, 사망자 등록률의 완성도 K_3 는 MSEG, 30~65세를 이용한 GGB-MSEG, 그리고 40~70세를 이용한 GGB-MSEG의 평균을 이용한 결과, 이 역시 우수한 완성도를 보였다.
- (3) Preston et al. (1998)은 APC 모형을 응용하여 제시한 GGB-APC 모형을 통해 구한 연령별 완성도와 이를 통해 구한 연도별 연령별 누락률은 비정상적인 코호트별 인구증가 현상과 성비, 그리고 연령비의 이상 현상을 설명하는데 매우 유용한 것으로 나타났다.
- (4) GGB-APC 모형에 의해 추정된 누락률은 센서스의 완성도에 의해 발생할 수 있으나 한편으로는 사망자 등록률의 완성도 또는 인구이동의 정확성 때문에 발생할 가능성도 있다.
- (5) 중복률이 높게 나타나는 0~9세, 누락률(특히, 남성)이 매우 높은 75세이상에 대해 센서스의 누락률 조사를 위한 센서스후 조사(PES)시 면밀한 조사설계가 진행되어야 하며, 동시에 해당 연령대의 동태자료의 정확성을 재조사할 필요가 있다.
- (6) 이를 토대로 과거의 센서스 자료의 수정 및 재구축을 해야 할 것이다.
- (7) 2010년의 센서스 자료는 GGB-APC 모형에 의해 이상 현상이 드러나고 있다. 이러한 이상 현상은 연령비에 의해서도 발견되고 있다. 이러한 이상 현상이 2010년 처음으로 실시된 인터넷 센서스 때문에 발생한 것이 아닌지 면밀한 점검과 검증이 필요하다.
- (8) 현행의 센서스가 중단되고 등록 센서스가 실시될 경우 등록 센서스의 완성도와 질적인 우수성을 평가할 수 있는 유일한 방법은 인구통계학적 방법이며 이에 대한 활용이 적극적으로 시행되어야 할 것이다.