

우리나라 자료에 적합한 생명표 작성방법에 대한 연구

박유성¹ · 김성용²

¹고려대학교 통계학과, ²고려대학교 경제통계연구소

(2011년 6월 접수, 2011년 8월 채택)

요약

생명표는 특정 집단의 사망 경험(mortality experience)을 반영하여 각 연령에서의 기대여명을 추정하는 통계적 모형이다. 사망 경험은 사망확률(death probability)을 통해 반영되는데, 사망확률을 추정하기 위해서는 세 가지 사항이 고려되어야 한다. 첫째는 사망률(death rate)로부터 사망확률을 추정하는 방법의 선택이며, 둘째는 사망확률의 불규칙성을 해결하기 위한 평활 방법, 셋째는 초고령 자료의 신뢰성 문제를 해결하기 위한 사망확률 추정 및 확장 문제이다. 본 논문에서는 사망확률 추정방법의 선택을 위해 Chiang 방법, 상수방법, Greville 방법, Reed and Merrell 방법 및 Keyfitz and Frauenthal 방법을 비교하며, 평활 방법으로는 Beers 방법, Greville 방법 및 이동평균 방법을 비교하도록 한다. 또한 초고령에서 사망확률 추정 및 확장을 위해 총 12가지 수학적 함수를 비교한다. 본 논문에서는 각 방법들을 비교함으로써 우리나라에 적합한 생명표 작성 방법을 제시하고, 이를 이용하여 2005년부터 2009년까지의 생명표를 작성하도록 한다. 또한 기간별 성별 기대여명의 역전현상(cross-over)현상을 해결하기 위한 방법을 제시한다.

주요어: 생명표, 사망확률 추정, 평활 방법, 초고령 사망확률, 역전현상.

1. 서론

기대여명(life expectancy)은 특정 연령까지 생존한 사람들의 기대되는 잔여생존연수로, 한 인구의 사망 수준 및 건강 상태를 포괄적으로 보여주는 지표이다. 따라서 기대여명의 추정치는 복지 정책 수립, 보험료 산정, 인구 추계 및 재정 추계 등에 널리 이용되고 있으며, 많은 국가에서 매년 연령별 기대여명을 발표하고 있으며, 우리나라 통계청에서도 매년 이를 발표하고 있다. 기대여명은 생명표로부터 추정되는데, 생명표는 특정 집단의 사망 경험(mortality experience)을 반영하여 각 연령에서의 기대여명을 추정하는 통계적 모형이다 (Chiang, 1984).

생명표는 크게 코호트생명표(cohort life table)와 기간생명표(period life table)로 구분된다. 코호트생명표는 특정 연도에 태어난 집단인 코호트가 시간이 흘러 전부 사망할 때까지 관측된 코호트의 연령별 사망률(cohort age-specific death rate)에 근거하여 각 연령별 기대여명을 추정한다. 이에 반해 기간생명표(period life table)은 특정 시점에서 관측된 한 인구의 기간 연령별 사망률(period age-specific death rate)이 미래에도 지속된다는 가정 하에서 이들의 사망률을 가상의 코호트(hypothetical cohort)에 반영함으로써 연령별 기대여명을 추정한다. 따라서 코호트생명표는 종단면적인(longitudinal)

이 논문은 2010년 정부재원(교육과학기술부 인문사회연구역량강화사업비)으로 한국연구재단의 지원을 받아 연구되었음(NRF-2010-411-B00028).

²교신저자: (339-700) 충남 연기군 조치원을 세종로 2511, 고려대학교 세종캠퍼스 경제통계연구소, 연구 교수.
E-mail: yaba96@korea.ac.kr

사망률을 반영하는 데 반해, 기간생명표는 횡단면적인(cross-sectional) 사망률을 반영한다. 이러한 특성으로 인해 코호트생명표를 작성하기 위해서는 오랜 시간이 걸리며, 현재 생존해 있는 대부분의 코호트에 대해서는 코호트생명표를 작성할 수 없는 한계점이 있다. 따라서 대부분의 나라에서는 기간생명표를 이용하고 있으며 우리나라 통계청의 경우도 기간생명표를 작성하여 발표하고 있다.

또한 생명표는 연령구간을 1세 단위로 하느냐 혹은 5세 및 10세 등의 단위로 하느냐에 따라 완전생명표(complete life table)과 간이생명표(abridged life table)로 구분된다. 과거 1세 단위의 자료의 정밀도가 낮은 경우 간이생명표를 발표하였으나, 이러한 문제가 개선됨에 따라 많은 나라에서 완전생명표를 간이생명표와 함께 발표하고 있다. 우리나라 통계청에서는 1970년부터 간이생명표를 발표해오고 있으며, 1997년부터 완전생명표를 이와 함께 발표하고 있다. 본 논문에서는 기간생명표이며, 동시에 완전생명표(이후 생명표로 표기함)를 작성하는 방법에 대해 다루고자 한다. 생명표의 작성은 정지인구(stationary population) 가정하에서 현재 인구의 기간 연령별 사망률(period age-specific death rate)을 가상의 코호트(통상 10만명)의 연령별 사망확률(age-specific death probability)로 전환하고, 각 연령에서의 생존자를 추정된 후 이들의 향후 생존 평균 년수를 계산함으로써 이루어진다. 따라서 생명표 작성시 가장 중요시 되는 부분은 연령별 사망확률의 생성인데, 이를 위해서는 크게 세 가지 사항이 고려되어야 한다.

첫 번째 문제는 연령별 사망확률의 추정방법이다. 연령별 사망확률을 추정하는 방법에는 여러 가지 방법이 있으며, 이들 방법은 연령구간(완전생명표의 경우 1세단위)에서의 사망자 분포 및 인구 분포에 대한 가정에 따라 관측된 연령별 사망률을 다르게 반영하여 사망확률을 생성한다. Reed와 Merrell (1939)은 실제 코호트의 동태자료를 이용한 사망확률과 관측된 사망률을 비교 분석함으로써 사망률과 사망확률과의 관계식을 도출하였다. Greville (1943)은 로그 변환한 사망률이 연령에 비례한다는 가정하에 사망확률을 계산하는 방법을 제안하였으며, Chiang (1968)은 연령구간에서의 사망자 수가 선형적으로 증가함을 가정하였다. Keyfitz (1966)는 연령구간에서 사망자 분포가 이차함수를 따른다는 가정을 하였으며, Keyfitz와 Frauenthal (1975)는 특정 연령구간에서의 인구 및 사망률이 인접한 연령구간의 인구 및 사망률과 선형 관계를 가지고 있다고 가정함으로써 사망확률을 계산하는 방법을 제안하였다. 이 외에도 사망률이 해당 연령구간에서 동일함을 가정한 상수 방법 등이 있다. 현재 미국국립보건통계국(National Center for Health Statistics; NCHS) 및 우리나라 통계청에서는 Chiang 방법을 이용하고 있다.

세계보건기구(World Health Organization; WHO)에서는 간이생명표 작성시 이러한 방법들이 매우 다른 가정을 하고 있음에도 불구하고, 실제 자료에 적용되었을 때 각 방법별로 생성된 기대여명은 유의한 차이를 보이지 않음을 밝혔으며 (WHO, 1977), Ng와 Gentleman (1995)은 위 방법들 가운데 Chiang 방법, Greville 방법 및 Keyfitz and Frauenthal 방법을 적용하여 간이생명표를 작성한 결과, 방법별로 기대여명의 차이가 없음을 밝히고 있다. 그러나 이러한 연구는 특정 국가들의 경우 사망확률 추정방법에 따른 차이가 없음을 의미하며, 우리나라의 자료를 이용한 사망확률 추정방법에 대한 연구는 이루어지지 않고 있다. 따라서 본 논문에서는 상수방법, Chiang 방법, Reed and Merrell 방법, Greville 방법 및 Keyfitz and Frauenthal 방법 등 총 5 방법을 우리나라의 자료에 적용하여 기대여명의 차이를 살펴보았다. 그 결과, 5세 단위의 간이생명표에서는 기존 연구와 달리 방법별로 차이가 발생하는 것으로 나타났으며, 1세단위의 완전생명표에서는 방법별 차이가 발생하지 않는 것으로 나타났다.

두 번째 문제는 생성된 사망확률의 불규칙성이다. 사망확률은 관측된 사망률로부터 생성되기 때문에, 인구 수나 사망자 수 자료가 불규칙(anomaly)한 경우 사망확률의 안정적인 패턴을 파악하기 어려워진다. 이러한 문제를 해결하기 위해 다양한 평활 방법이 이용된다. 평활 방법은 크게 연령별 사망자 수 및 인구 수를 평활하는 방법 (Beers, 1944; Sprague, 1881; Shryock 등, 1993)과 사망률이나 사망확률을 평

활하는 방법이 있다 (Greville, 1945; Buettner, 2002). 현재 NCHS에서는 Beers 방법을 이용하고 있으며, 우리나라 통계청에서는 Greville의 9차항 계수 방법을 이용하고 있다.

박유성 등 (2009)은 0세부터 95세까지의 우리나라의 주민등록 인구와 사망자료로 생성한 사망률에 대해 Beers 방법 및 Greville의 9차항 계수 방법을 적용하여 비교 연구하였으며, 그 결과 관측된 사망확률에 대한 적합도는 Greville의 9차항 계수 방법이 우수하며, 평활성의 경우 Beers 방법이 더 우수함을 밝혔다.

본 논문에서는 우리나라 사망자 수 및 센서스 인구를 이용하여 연령별 사망률을 구하고 Beers 방법, Greville 방법 및 이동평균 방법을 적용하여 비교하였다. 그 결과 Beers 방법의 경우 박유성 등 (2009)의 결과와는 달리 95세 이상의 연령에서 심각한 변동이 발생하는 것으로 나타났으며, Greville 방법과 이동평균방법의 경우 그러한 문제가 발생하지 않는 것으로 나타난다.

세 번째 문제는 초고령 자료의 신뢰성에 대한 문제로, 초고령(80세 이상)에서의 사망률은 부정확한 연령 보고 등으로 인해 자료의 신뢰성이 확보되지 않는 것으로 알려지고 있다 (Condran 등, 1991; Coale과 Kisker, 1990; Hill, 1999; Boleslawski와 Tabeau, 2001; Buettner, 2002). 그러나 고령화가 가속화됨에 따라 80세 이상 연령에서 뿐만 아니라 100세 이상의 사망률을 파악할 필요가 있게 되었고, 이러한 초고령에서 자료의 신뢰성 문제를 해결하기 위해 수학적 함수를 이용하여 사망률을 100세 이상까지 확장하는 방법들이 제안되었다. 이 방법은 수학적 함수를 신뢰성이 확보된 사망률 자료에 적합하여 모수를 추정 후, 초고령에서 절단된 사망률을 외삽(extrapolation)한다. 이와 같이 확장된 사망률을 사망확률로 전환하여 초고령에서의 기대여명을 구하게 된다.

초고령의 사망률 패턴을 나타내기 위해 Gompertz 함수 (Gompertz, 1825), Heligman-Pollard 함수 (Heligman과 Pollard, 1980), Coale and Kisker 함수 (Coale과 Kisker, 1990), 다항함수 및 로지스틱 형태의 함수 (Himes 등, 1994; Beard, 1971; Perks, 1932) 등 다양한 함수들이 제안되어 왔다. NCHS에서는 1997년부터 1999년까지의 생명표의 경우 신뢰도가 높은 85세부터 99세까지의 의료보험자료로 구한 사망확률에 변형된 Coale and Kisker 함수를 적합하여 모수를 추정 후, 이를 이용하여 85세 이상의 사망확률을 추정하였다 (Anderson, 1999). 그러나 2000년 이후의 생명표에서는 Heligman-Pollard 함수를 65세부터 100세까지의 의료보험 자료 및 동태자료로 구한 사망확률에 적합하여 모수를 추정 후, 66세 이상의 사망확률을 추정하였다 (Arias, 2010).

우리나라 통계청에서는 NCHS의 변형된 Coale and Kisker 방법을 85세부터 99세까지의 동태자료로 구한 사망확률에 적합하여 사망확률을 추정한다 (통계청, 2010). 그러나 그림 1.1에서 알 수 있듯이, 우리나라의 경우 85세 이상 사망확률의 경우 매우 불규칙할 뿐만 아니라 미국의 패턴과도 다른 형태를 보이고 있다. 그림 1.1(a)는 미국의 1996년 의료보험데이터로 구한 사망확률의 변화율이며, 그림 1.1(b)는 우리나라의 2005년 센서스 인구 및 동태자료로 구한 사망확률의 변화율이다.

이에 따라 본 논문에서는 미국과 우리나라의 사망확률의 패턴이 다르다는 판단 하에 위에서 언급한 여러 함수를 적용하여 우리나라의 사망률 패턴을 잘 설명해주는 함수를 찾으려 한다. Boleslawski와 Tabeau (2001)은 고연령대에서의 사망률의 연령별 패턴을 파악하기 위해 11개의 수학적 함수를 실제 자료에 적용하여 비교 분석하였다. 이들은 최소한 90세까지 신뢰성이 확보된 자료가 존재하는 경우 Coale and Kisker의 모형을 이용할 것을 추천하였으며 최소한 85세까지의 신뢰할 수 있는 자료가 존재하는 경우 다항함수를 이용하여 사망률을 외삽할 것을 권고하고 있다.

본 논문에서는 Boleslawski와 Tabeau (2001)가 이용한 11개의 함수와 현재 우리나라 통계청에서 이용하고 있는 변형된 Coale and Kisker 방법을 우리나라 2005년 자료에 적용하여 비교 분석하였다. 이때 추정된 사망률이 갖춰야 할 전제조건으로 일치성(consistency)이 있다. 첫 번째 일치성은 사망률의 증

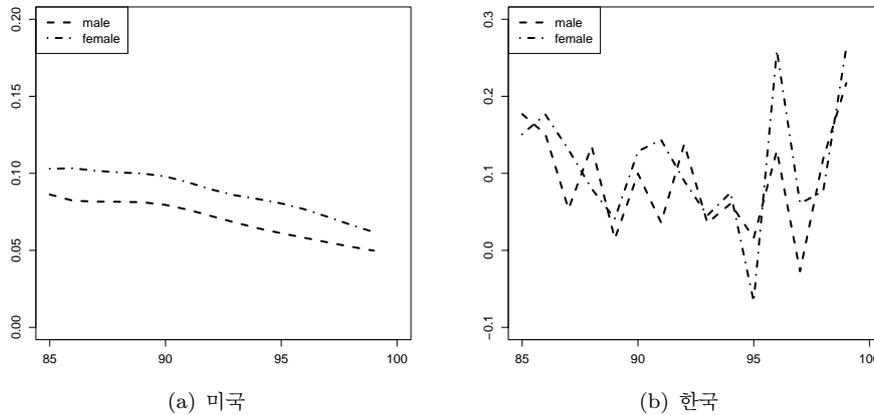


그림 1.1. 미국 및 우리나라의 사망확률의 변화율

가율이 특정 고연령 이후에는 변하지 않거나 감소하는 추세를 유지해야 한다는 것이며 (Condran 등, 1991; Wilmoth, 1995; Boleslawski와 Tabeau, 2001), 두 번째는 동 연령에서 남자의 사망률이 여자의 사망률보다 높아야 한다는 것이다.

본 논문에서는 각 함수의 적합도 및 두 가지의 일치성을 고려하여 우리나라의 2005년부터 2009년까지의 완전생명표를 작성하도록 한다. 2005년의 인구로는 0세부터 100세 이상(100+) 연령으로 정리되어 있는 유일한 인구인 센서스 인구를 이용하도록 하며, 2006년부터 2009년의 인구로는 센서스후 인구(postcensal estimates)를 이용하도록 한다.

이 때 기간별 생명표가 갖춰야 할 전제조건으로, 초고령에서 기간별 사망률의 순서는 0세에서의 기대여명의 순서와 일치해야 한다 (Buettner, 2002). 이는 일반적으로 시간이 지남에 따라 사망률이 낮아진다는 점을 감안할 때, 연도별로 사망률을 구했을 때 이들의 사망률에 역전현상이 없어야 한다는 것을 의미한다. 그러나 연도별로 다항함수를 이용하여 초고령의 사망률을 추정했을 때 이러한 일치성을 만족하지 않았으며, 따라서 본 논문에서는 이러한 일치성을 만족하기 위한 방법을 제안하고, 이를 이용하여 최종적으로 2005년부터 2009년까지의 완전생명표를 작성하도록 한다.

본 논문의 구성은 다음과 같다. 2절에서는 사망률을 사망확률로 전환하는 방법들에 대해 설명하고, 우리나라의 자료에 각 방법들을 적용했을 때의 기대여명의 차이를 살펴보도록 한다. 3절에서는 여러 평활 방법을 우리나라의 자료에 적용하여 각 방법의 장·단점을 파악하도록 한다. 4절에서는 앞서 제시한 사망확률 추정 방법 및 평활 방법을 적용한 후, 초고령의 사망률 패턴을 잘 설명하는 함수를 찾도록 하며, 5절에서는 기간별 사망률의 역전현상을 해결하기 위한 식을 제시하고, 우리나라의 2005년부터 2009년까지의 생명표를 작성하도록 하도록 한다.

2. 연령별 사망확률의 추정 방법 비교

사망확률의 추정 방법을 논의하기에 앞서 연령별 사망률 및 사망확률의 정의에 대해 알아보자. 본 논문에서는 1세단위의 완전생명표 작성에 대해 논의하지만, 일반화를 위해 연령단위를 n 세로 하여 설명하도록 한다. x 세에서 $x+n$ 세까지의 사망확률(${}_nq_x$)은 x 세까지 살아남은 가상의 코호트의 생존자(l_x) 수 가운데 해당 연령구간에서 발생하는 사망자 수(${}_nd_x$)의 비(ratio)로 정의되며, 따라서 ${}_nq_x$ 는 l_x 가 x 세에서 $x+n$ 세에서 사망할 확률을 나타낸다. x 세에서 $x+n$ 세까지의 사망률(${}_nm_x$)은 다음과 같다. l_x 가 x 세

표 2.1. ${}_n a_x$ 추정식

방법	${}_n a_x$
Chiang 방법	$\frac{n}{2}$
상수 방법	$n + \frac{1}{n m_x} - \frac{n}{1 - \exp[-n m_x]}$
Greville 방법	$n + \frac{1}{n m_x} - \frac{n}{1 - \exp[-n m_x - \frac{n^3}{12} m_x^2 \log c]}$
Reed-Merrell 방법	$n + \frac{1}{n m_x} - \frac{n}{1 - \exp[-n m_x - 0.008 n^3 m_x^2]}$ if $x \geq 5, n = 5$
Keyfitz 방법	$n + \frac{1}{n m_x} - \frac{n}{1 - \exp[-n m_x - \frac{n}{48 n P_x} (n P_{x-n} - n P_{x+n}) (n m_{x-n} - n m_{x+n})]}$

에서 $x + n$ 세까지 살아가는 년수를 인년수(the number of person-years, ${}_n L_x$)라고 하며, 이 때 ${}_n m_x$ 는 ${}_n d_x$ 에 대한 ${}_n L_x$ 의 비로 정의된다. 실제 관측되는 기간 연령별 사망률(${}_n M_x$)은 x 세에서 $x + n$ 세까지의 총사망자 수(${}_n D_x$)에 대한 해당연령의 중간인구(${}_n P_x$, mid-year population)에 대한 비로 정의되며 생명표상의 연령별 사망률과는 다르다. 그러나 정지인구(stationary population) 가정 하에서 ${}_n M_x$ 는 ${}_n m_x$ 에 근사하므로, 생명표 작성시 ${}_n m_x$ 대신 ${}_n M_x$ 를 이용하여 ${}_n q_x$ 를 추정한다 (Preston 등, 2000; Siegel과 Swanson, 2004).

생명표에서 연령별 사망률을 이용하여 사망확률을 추정하기 위해, Greville (1943)과 Chiang (1968)은 ${}_n m_x$ 를 ${}_n q_x$ 로 전환하는 다음의 일반식을 제시하였다.

$${}_n q_x = \frac{n \cdot n m_x}{1 + (n - n a_x) n m_x}, \tag{2.1}$$

여기서 ${}_n a_x$ 는 x 세에서 $n + x$ 세까지의 연령구간에서 사망하는 코호트가 평균적으로 사는 기간이다. 따라서 위 식을 이용하여 사망확률을 추정하기 위해서는 ${}_n a_x$ 에 대한 추정이 필요하다. 서론에서 언급한 여러 방법들은 ${}_n a_x$ 에 대한 추정 방법의 차이로 귀결되며, 본 논문에서는 이 가운데 대표적인 방법인 Chiang 방법 (Chiang, 1968), 상수방법, Greville 방법 (Greville, 1943), Reed-Merrell 방법 (Reed와 Merrell, 1939), Keyfitz 방법 (Keyfitz와 Frauenthal, 1975) 등을 비교하도록 한다. 각 방법에서 ${}_n a_x$ 에 대한 추정식은 표 2.1에 제시되어 있다.

원칙적으로 Chiang 방법이라 함은 식 (2.1) 자체를 지칭하나, 대부분의 경우 Chiang 방법은 연령구간 내에서 사망자 수가 선형적으로 증가하는 ${}_n a_x = 1/2$ 인 경우를 지칭하므로 본 논문에서도 이와 같이 적용하도록 한다. 상수 방법은 연령구간 내에서 사망률이 ${}_n m_x$ 로 동일한 경우를 의미하며, Greville 방법은 연령구간 내에서 나이를 먹어감에 따라 사망률이 증가함을 가정하며, $\log c$ 의 값으로 0.095를 이용하도록 한다 (Siegel과 Swanson, 2004).

이러한 사망확률 추정 방법으로 인한 차이는 연령구간이 1세 단위인 경우보다 5세 단위에서 더 크기 때문에 (Keyfitz와 Frauenthal, 1975), 본 논문에서는 우선 위 방법들을 이용하여 5세 단위의 간이생명표를 작성하여 방법별 기대여명을 비교해 보도록 한다. 또한 영·유아의 사망확률인 ${}_1 q_0$ 및 ${}_4 q_1$ 은 Coale and Demeny의 모형 생명표 가운데 동양 국가에 가장 가깝다고 알려진 서쪽 모형(West model)에서 이용되는 식을 이용하도록 한다 (Coale과 Demeny, 1983). 이를 이용하는 이유는 영·유아의 경우 다른 연령대와는 달리 연령구간의 초반에 사망이 집중되어 있다고 알려져 있으나, 우리나라의 경우 이를 파악하기 위한 자료가 존재하지 않기 때문이다. 표 2.2는 이와 같은 설정 하에 각 방법을 적용한 2004-2006년 간이생명표로, 사망자 수는 연령별 임의변동을 제거하기 위해 2004년부터 2006년까지의 평균값을 이용하였으며, 인구 수는 대부분의 나라에서 가장 많이 이용되는 센서스 인구를 이용하였다.

각 방법의 평균수명을 비교해보면 남자의 경우 Keyfitz 방법이 78.63년로 가장 컸으며 Greville 방법 및

표 2.2. 사망확률 추정 방법별 2004-2006년 간이생명표

연령	Chiang		상수		Greville		Reed-Merrell		Keyfitz	
	남자	여자	남자	여자	남자	여자	남자	여자	남자	여자
0	78.42	85.21	78.35	85.14	78.45	85.23	78.45	85.23	78.63	85.50
1	75.27	81.99	75.21	81.92	75.31	82.01	75.31	82.01	75.49	82.29
5	70.40	77.10	70.33	77.03	70.43	77.13	70.43	77.12	70.61	77.41
10	65.47	72.16	65.40	72.09	65.50	72.18	65.50	72.18	65.65	72.42
15	60.52	67.20	60.46	67.13	60.56	67.22	60.55	67.22	60.70	67.46
20	55.64	62.27	55.58	62.20	55.67	62.29	55.67	62.29	55.82	62.54
25	50.79	57.37	50.73	57.30	50.82	57.39	50.82	57.39	50.98	57.65
30	45.98	52.49	45.91	52.41	46.01	52.51	46.01	52.51	46.18	52.78
35	41.20	47.62	41.13	47.55	41.23	47.64	41.23	47.64	41.40	47.92
40	36.52	42.79	36.45	42.72	36.55	42.81	36.55	42.81	36.73	43.09
45	32.00	38.01	31.93	37.93	32.03	38.03	32.03	38.02	32.21	38.30
50	27.67	33.28	27.60	33.21	27.69	33.30	27.69	33.30	27.85	33.55
55	23.51	28.63	23.44	28.55	23.53	28.64	23.53	28.64	23.67	28.86
60	19.50	24.04	19.43	23.97	19.52	24.06	19.52	24.06	19.63	24.25
65	15.84	19.66	15.77	19.58	15.85	19.67	15.85	19.67	15.95	19.84
70	12.50	15.49	12.42	15.41	12.50	15.49	12.50	15.49	12.60	15.65
75	9.44	11.67	9.35	11.59	9.42	11.66	9.42	11.66	9.49	11.78
80	6.89	8.44	6.80	8.35	6.86	8.42	6.85	8.41	6.89	8.49
85	4.89	5.89	4.79	5.78	4.83	5.84	4.83	5.83	4.85	5.87
90	3.52	4.10	3.41	4.00	3.44	4.04	3.44	4.03	3.45	4.04
95	2.55	2.98	2.43	2.86	2.45	2.89	2.45	2.89	2.46	2.90
100	1.68	1.85	1.68	1.85	1.68	1.85	1.68	1.85	1.68	1.85

표 2.3. 각 방법과 Chiang 방법 비교

항목	상수방법		Greville		Reed-Merrell		Reed-Merrell	
	남자	여자	남자	여자	남자	여자	남자	여자
ME	-0.07	-0.08	0.01	0	0.01	0	0.12	0.19
MAE	0.07	0.08	0.03	0.02	0.03	0.02	0.14	0.21

Reed-Merrell 방법의 경우 모두 78.45년으로 그 뒤를 잇고 있으며, Chiang의 방법이 78.42년이며 상수 방법이 78.35년으로 가장 작다. 여자의 경우 역시 이와 동일한 순서를 보이는데 Keyfitz 방법의 경우 85.50년으로 가장 컸으며 Greville 방법 및 Reed-Merrell 방법은 85.23년, Chiang 방법은 85.21년, 마지막으로 상수 방법이 85.14년이다. 평균수명의 차이는 남자의 경우 가장 큰 차이가 0.27년, 여자의 경우 가장 큰 차이가 0.37년인 것으로 나타났다. 연령별 기대여명의 차이를 살펴보기 위해 가장 일반적으로 쓰이는 Chiang 방법을 기준으로 하여 평균오차(Mean Error; ME) 및 평균절대오차(Mean Absolute Error; MAE)를 구해보면 표 2.3과 같다.

ME의 관점에서 남녀 모두 Keyfitz 방법은 Chiang방법에 비교하여 크며, Greville 방법 및 Reed-Merrell 방법의 경우 Chiang 방법과 거의 차이가 없는 것으로 나타났다. 상수 방법은 Chiang 방법보다 작은 것으로 나타난다. MAE의 관점에서 Greville 방법, Reed-Merrell 방법, Keyfitz 방법은 ME보다 MAE가 큰 값을 보이고 있는데, 이는 표 2.2에서 알 수 있듯이 이들 방법의 경우 장년층까지는 Chiang 방법보다 기대여명이 큰데 반해, 초고령층에서 Chiang 방법의 기대여명보다 더 작은 값을 갖고 있기 때문이다.

표 2.4. 사망확률 추정 방법별 2004-2006년 완전생명표

연령	Chiang		상수		Greville		Reed-Merrell		Keyfitz	
	남자	여자	남자	여자	남자	여자	남자	여자	남자	여자
0	74.89	81.63	74.89	81.63	74.89	81.63	74.89	81.63	74.93	81.67
1	74.25	80.98	74.24	80.97	74.25	80.98	74.25	80.98	74.28	81.01
10	65.42	72.12	65.41	72.12	65.42	72.12	65.42	72.12	65.43	72.13
20	55.59	62.23	55.59	62.23	55.59	62.23	55.59	62.23	55.6	62.24
30	45.92	52.45	45.92	52.44	45.93	52.45	45.93	52.45	45.93	52.46
40	36.46	42.75	36.46	42.75	36.47	42.75	36.47	42.75	36.47	42.76
50	27.61	33.24	27.60	33.24	27.61	33.24	27.61	33.24	27.62	33.25
60	19.45	24.01	19.45	24.00	19.45	24.01	19.45	24.01	19.45	24.01
70	12.42	15.44	12.42	15.43	12.42	15.44	12.42	15.44	12.42	15.44
80	6.81	8.38	6.81	8.37	6.81	8.38	6.81	8.38	6.81	8.38
85	4.79	5.80	4.79	5.80	4.79	5.80	4.79	5.80	4.79	5.80
90	3.42	4.00	3.41	4.00	3.41	4.00	3.41	4.00	3.41	4.00
95	2.42	2.84	2.42	2.83	2.42	2.83	2.42	2.83	2.42	2.84
100	1.68	1.85	1.68	1.85	1.68	1.85	1.68	1.85	1.68	1.85

각 방법별에 따른 연령별 기대여명의 차이는 남자의 경우 평균적으로 0.19년이며, 여자의 경우 0.29년로, 평균수명의 차이보다는 작은 것으로 나타난다. 그러나 이러한 차이는 Ng와 Gentleman (1995)이 비교한 Chiang 방법, Greville 방법 및 Keyfitz 방법들의 차이보다 크다. 이들의 결과에 따르면 간이생명표 작성시 평균 수명의 경우 각 방법에 따른 차이가 0.01년에 불과하며, 연령별 차이는 0.01년 미만이다. 따라서 우리나라의 경우 WHO (1977) 및 Ng와 Gentleman (1995)의 결과와 달리 사망확률 추정 방법에 따라 큰 차이를 보이고 있다.

이제 각 방법들을 이용하여 1세단위의 완전생명표를 작성함으로써 사망확률추정방법을 비교해 보도록 하자. 각 방법을 이용하여 완전생명표를 작성하고, 이를 10세 단위로 나타내면 표 2.4와 같다.

표 2.4에 따르면 Chiang 방법, 상수 방법, Greville 방법 및 Reed-Merrell 방법은 평균수명이 성별에 관계없이 동일하다. 그러나 Keyfitz 방법의 경우 이들보다 남녀 모두 0.04년만큼 더 큰 것을 알 수 있다. 그러나 이는 약 14.6일의 차이에 불과하다. 각 연령별 평균 기대여명의 경우, Chiang 방법, 상수 방법, Greville 방법 및 Reed-Merrell 방법은 차이가 없는 것으로 나타났으며, Keyfitz 방법의 경우 이들 방법보다 평균적으로 0.01세가 더 큰 것으로 나타났다. 따라서 간이생명표의 경우와 달리 완전생명표에서는 사망확률 추정 방법별로 차이가 없는 것으로 나타난다.

본 논문에서는 완전생명표의 작성 방법에 대해 다루고자 하므로, 가장 일반적으로 쓰이는 Chiang 방법을 이용하여 이를 작성하도록 한다. 그러나 간이생명표의 경우 각 방법에 따라 기대여명의 차이가 나타나므로, 추후 이에 대한 연구가 필요하다고 여겨진다. 3장에서는 사망확률의 불규칙성을 해결하기 위한 평활 방법을 다루도록 한다.

3. 사망확률의 평활

평활은 사망확률의 불규칙성을 제거함으로써 안정적인 사망 패턴을 찾기 위해 이루어진다. 평활방법으로는 대표적으로 Beers 방법, Greville 방법 및 이동평균방법이 있는데, NCHS에서 이용되는 Beers 방법은 사망자 수와 인구 수 자체를 평활하는 방법으로, 사망자 수 및 인구 수를 5세단위로 합한 후, 이를 1세단위로 재할당함으로써 1세단위의 임의변동을 제거하는 방법이다. 이에 반해 통계청에서 이용되

표 3.1. 경험기대여명과 모형기대여명의 비교

연령	경험		Beers		Greville		이동 평균	
	남자	여자	남자	여자	남자	여자	남자	여자
0	74.89	81.63	74.89	81.63	74.89	81.63	74.64	81.35
1	74.25	80.98	74.05	80.79	74.25	80.98	73.99	80.70
5	70.34	77.07	70.33	77.05	70.35	77.07	70.22	76.92
10	65.42	72.12	65.40	72.11	65.42	72.12	65.29	71.97
20	55.59	62.23	55.57	62.22	55.59	62.23	55.47	62.09
30	45.92	52.45	45.91	52.43	45.93	52.45	45.81	52.30
40	36.46	42.75	36.45	42.74	36.47	42.75	36.35	42.60
50	27.61	33.24	27.59	33.23	27.61	33.24	27.50	33.10
60	19.45	24.01	19.42	23.99	19.45	24.01	19.36	23.87
70	12.42	15.44	12.40	15.42	12.42	15.44	12.33	15.32
80	6.81	8.38	6.78	8.36	6.82	8.38	6.76	8.31
85	4.79	5.80	4.74	5.78	4.80	5.80	4.78	5.77
90	3.42	4.00	3.34	4.01	3.45	4.02	3.42	3.99
95	2.42	2.84	2.15	2.86	2.44	2.82	2.39	2.73
100	1.68	1.85	1.68	1.85	1.68	1.85	1.68	1.85

는 Greville 방법은 사망확률을 직접 평활하는 방법이며, Buettner (2002)는 사망률을 이동평균 방법으로 평활한 후, 이를 사망확률로 전환하였다. 이동평균 방법의 경우 3개월, 5개월, 7개월 이동평균 가운데 최소 2차 미분변동(second-order difference) 기준 (Buettner, 2002)에 따라 가장 좋은 결과를 보이는 7개월 이동평균을 이용하도록 하며, Beers 방법 및 Greville 방법에 대한 자세한 사항은 박유성 등 (2009)를 참고하도록 한다.

본 논문에서는 평활 방법의 차이를 파악하기 위해 각 방법을 이용했을 때의 연령별 기대여명과 평활을 하지 않고 구한 연령별 경험 기대여명 비교해 보도록 한다. 또한 사망확률 추정 방법과 평활 방법에 따른 기대여명의 차이를 살펴보기 위해 2절에서와 동일한 데이터를 이용한다. 표 3.1은 이들을 비교한 것으로 완전생명표를 구한 후 10세 단위로 정리한 것이다.

표 3.1에 따르면 평활방법에 따른 기대여명의 차이는 2절의 사망확률 추정방법에 따른 기대여명의 차이보다 큰 것으로 나타난다. 완전생명표의 경우 사망확률 추정방법에 따라서 기대여명의 차이가 거의 없는 것으로 나타났으나, 평활방법에 따라서는 이보다 큰 차이가 나는 것으로 나타난다. Beers 방법의 경우 남자 95세에서 경험기대여명과 0.27년 차이가 나고 있으며, 이동 평균 방법의 경우 평균수명에서 남자는 경험기대여명과 0.26년, 여자는 0.28년 차이가 발생하는 것으로 나타난다. 세 방법 가운데 Greville 방법이 전 연령대에 걸쳐 경험기대여명에 가장 가까운 것으로 나타났다.

이제, 평활 방법별로 기대여명에 차이가 발생하는 원인에 대해 알아보도록 하자. 그림 3.1은 세 가지의 평활 방법을 적용했을 때와 평활을 적용하지 않았을 때의 사망확률로, 임의변동이 심한 0세부터 10세 및 80세부터 99세까지 연령의 남녀별 그림이다.

위 그림에 따르면 10세 미만의 경우 Greville 방법은 실제 자료와 거의 차이가 없으며, Beers 방법의 경우 0세부터 5세까지의 경우 실제값과 차이를 보이고 있으나 부드러운 곡선의 형태를 띠고 있음을 알 수 있다. 이동평균 방법의 경우 Greville 방법과 Beers 방법의 중간 정도의 형태를 띠고 있다. 그러나 그림 3.1(c) 및 3.1(d)를 보게 되면 Beers 방법은 95세 이상에서 매우 큰 변동을 보이고 있다. 이는 0세부터 95세까지의 사망확률의 평활도를 비교한 박유성 등 (2009)의 결과와는 매우 상이한 결과이다. 이들은 Beers 방법의 경우 평활도는 높으나 적합도가 낮으며, Greville 방법은 평활도는 낮으나 적합도는 높

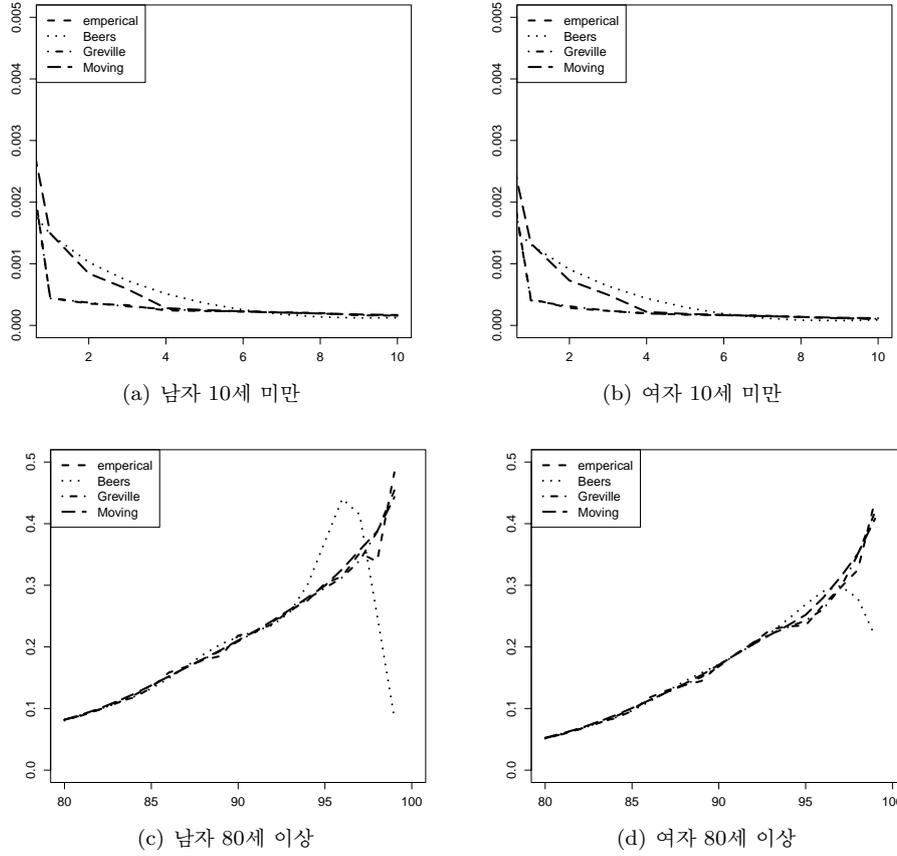


그림 3.1. 평활방법에 따른 영·유아 및 초고령층의 사망확률

표 3.2. Beers 평활 전과 평활 후의 사망자 수 및 인구 수

연령	사망자 수				인구 수			
	평활 전		Beers 적용 후		평활 전		Beers 적용 후	
	남자	여자	남자	여자	남자	여자	남자	여자
95	239	842	234	920	671	3,162	514	2,962
96	166	658	176	690	444	2,130	312	2,034
97	115	502	122	485	268	1,444	234	1,380
98	65	345	72	310	160	890	249	957
99	46	230	27	171	73	409	307	703
100+	66	496	66	496	111	915	111	915

다고 밝히고 있다.

이러한 차이는 99세에서 두드러지게 나타나는데 이는 Beers 방법을 적용했을 때 인구 수는 센서스 인구 보다 크며, 사망자 수는 더 작기 때문에 발생한다. 표 3.2는 95세 이상의 연령에서 성별 인구 수 및 사망자 수와 Beers 방법으로 평활된 인구 수 및 사망자 수의 값이다.

표 3.2에 따르면 Beers 방법을 적용했을 때 99세에서의 사망자 수는 남자의 경우 평활 전 값의 58.6%이

표 3.3. 평활 방법별 적합도(단위 : 10^{-2}) 및 평활도

		Beers	Greville	Moving
적합도	남자	0.1922230	0.0036800	0.0044910
	여자	0.0531070	0.0012570	0.0025720
평활도	남자	1.4295450	0.5111241	0.5367826
	여자	0.9489272	0.5526432	0.5627522

며 여자의 경우 평활전의 74.6%에 불과하나, 인구 수는 남자의 경우 평활 전 값의 423.1%이며, 여자의 경우 평활전 값의 171.8%에 이른다. 즉, Beers 방법을 이용했을 때 99세에서의 사망자 수는 작아지며 인구 수는 커지므로 사망률 및 사망확률이 평활전의 값에 비해 작아지게 되는 것이다.

우리나라의 경우 98세에서 99세가 됨에 따라 사망자 수 및 인구 수가 급격히 감소하고 있다. 우리나라의 경우 남녀 모두 99세 인구는 98세 인구에 비해 46% 미만이나, 미국의 2000년 센서스 결과에 따르면 남녀 모두 99세 인구는 98세 인구에 비해 75% 이상이다. 즉, 초고령에서 우리나라와 미국은 서로 다른 패턴을 보이고 있으며, Beers 방법은 이러한 차이를 반영하지 못하는 것으로 판단된다.

이제 각 방법의 평활도 및 적합도를 구하여 이들 가운데 우리나라에 적합한 평활 방법을 찾아보도록 하자. 적합도 및 평활도로는 London (1981)이 제시한 평균제곱오차(Mean Squared Error; MSE) 및 3차 절대 차분함을 이용하며 값이 작을수록 적합도 및 평활도가 더 좋음을 의미한다. 일반적으로 평활도가 높으면 적합도가 떨어지고, 평활도가 낮으면 적합도가 높아지는 반비례의 관계가 성립한다. 따라서 유사한 적합도를 보이며 평활도가 더 좋거나, 혹은 유사한 평활도를 보이면서 적합도가 더 좋은 방법을 찾도록 한다. 적합도 및 평활도를 구하기 위한 식은 다음과 같다.

$$\text{적합도} = \frac{\sum_{x=0}^{100} (v_x - \hat{v}_x)^2}{101}, \quad \text{평활도} = \sum_{x=0}^{100} |\hat{v}_{x+3} - 3\hat{v}_{x+2} + 3\hat{v}_x + 1 - \hat{v}_x| \quad (3.1)$$

이 값이 작을수록 적합도 및 평활도가 우수하다고 평가한다. 표 3.3은 세 가지 평활법을 이용하여 0세부터 100+의 전 연령대에 대해 적합도 및 평활도를 구한 결과이다.

표 3.3에 따르면 Greville 방법이 평활도와 적합도에서 우수한 것으로 나타났다. 따라서 Greville 방법이 기대여명과, 95세 이상 연령에서의 평활도에서도 우수한 것으로 나타나, Greville 방법이 우리나라 사망 확률에 가장 적합한 것으로 나타났다.

4. 초고령에서 사망률의 추정

본 논문에서는 사망률의 사망확률 전환방법으로 Chiang 방법을 이용하고, 사망확률을 평활하기 위한 방법으로 Greville 방법을 이용할 것을 제안하였다. 이제 이들 방법을 우리나라 2005년 인구 및 사망자 자료에 적용한 후, 우리나라 초고령에서의 사망률을 잘 반영하는 함수를 찾아보도록 하자.

서론에서 언급한 바와 같이 초고령에서의 사망률은 부정확한 연령보고 등으로 신뢰성이 확보되지 않는 것으로 알려져 있다. 이는 $m_x (= {}_1m_x)$ 의 연령별 변화율인 $k_x = \log(m_x/m_{x-1})$ 의 변화에서도 잘 나타나는데, 그림 4.1은 우리나라 2005년 센서스 인구 및 연령별 사망자 자료를 이용한 60세부터 99세까지의 평활전의 k_x 와 Greville 방법으로 평활한 후 생성한 k_x 의 그림이다.

위 그림에 따르면 평활전의 k_x 는 임의변동이 매우 심한 것으로 보이며, 평활 후의 k_x 는 임의변동이 줄어든 것으로 파악된다. 또한 이 경우 남녀 모두 75세 이후에 감소하는 추세를 보이고 있으며, 이는 초

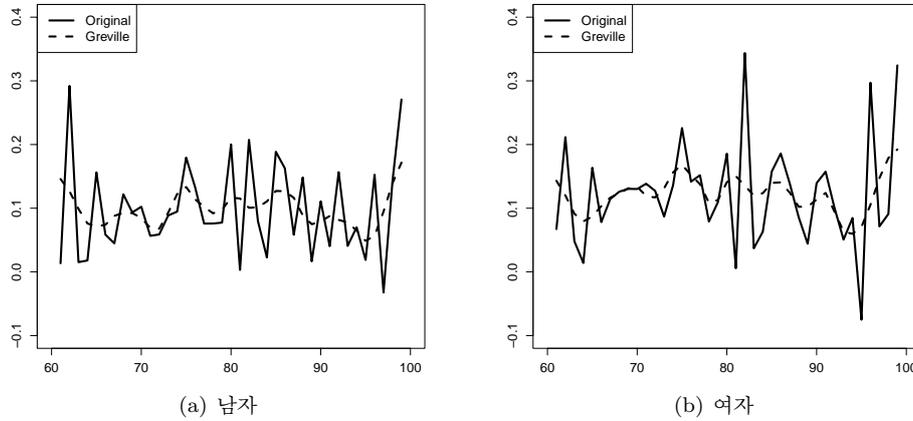


그림 4.1. 평활 전·후의 2005년 k_x

고령층에서 연령이 증가할수록 k_x 가 감소한다는 기존의 연구와 일치함을 알 수 있다 (Coale과 Kisker, 1986; Coale과 Kisker, 1990; Wilmoth, 1995; Boleslawski와 Tabeau, 2001). 그러나 95세 이상 연령에서는 평활 전 뿐만 아니라 평활 후에도 k_x 가 증가하고 있으며, 이는 기존 연구 결과와 일치하지 않는다. 따라서 본 논문에서는 안정적인 초고령층의 사망확률을 구하기 위해 Boleslawski와 Tabeau (2001)가 비교한 11개의 수학적 함수와 현재 우리나라에서 이용하고 있는 변형된 Coale-Kisker 모형(MCK 모형으로 표기)을 2005년 자료에 적용하여 비교 분석하여 우리나라 자료에 가장 적합한 함수를 찾으려 한다. 본 논문에서 비교하고자 하는 총 12개의 수학적 함수 및 추정해야 할 모수의 수, 그리고 특징은 표 4.1과 같다. 우리나라 자료에 적합한 함수를 선택하기 위한 기준으로는 적합도, 일치성 및 경험기대여명과의 차이를 고려하도록 한다.

적합도의 척도로는 적합구간에서 사망률 m_x 와 사망률의 변화율 k_x 에 대한 오차제곱합(Sum of Squared Error; SSE)과 MAPE(Mean Absolute Percentage Error; MAPE)를 이용하도록 한다. SSE와 MAPE가 작을수록 좋은 모형을 의미하며 MAPE는 실제값 대비 추정치가 몇 % 벗어났는지를 측정하는 척도이다.

두 번째 기준인 일치성은 추정된 사망률이 갖춰야 할 전제조건으로 두 가지 사항을 고려하도록 한다. 첫 번째 사항은 사망률의 증가율이 특정 고연령 이후에는 변하지 않거나 감소하는 추세를 유지해야 한다는 것으로, 이러한 패턴은 양질의 자료를 확보하고 있는 여러 선진국의 50년 이상의 자료로부터 발견되었다 (Boleslawski와 Tabeau, 2001). 그림 4.1에서 볼 수 있듯이 우리나라의 경우에도 이러한 패턴을 보이며, 적합도에 문제가 없는 함수들 가운데 사망률의 증가율이 감소 패턴을 보이는 함수를 선택하도록 한다. 두 번째 사항은 동 연령에서 남자의 사망률이 여자의 사망률보다 높아야 한다는 것이다.

세 번째 기준은 위 함수를 이용하여 추정한 모형기대여명(model life expectancy)과 함수를 이용하지 않고 구한 경험기대여명(empirical life expectancy)간의 차이가 작아야 한다는 점이다.

위 세 가지 기준에 의한 함수를 찾기 위해서는 각 함수의 모수를 추정해야 하며, 이를 위한 목적함수로

$$\sum_{x=65}^{95} \left(\frac{m_x - \hat{m}_x}{m_x} \right)^2 \tag{4.1}$$

를 이용하도록 한다. 여기서 m_x 는 Greville 방법으로 평활된 사망률이며, \hat{m}_x 는 표 4.1의 평활 후 각 함

표 4.1. 12개 함수식 및 그 특징

함수	식	식	특징
GOMP	$m_x = be^{ax}$	2	연령이 늘어남에 따라 사망률 증가 가정 (Gompertz, 1925)
MAKE	$m_x = c + be^{ax}$	3	GOMP 곡선의 40대 이하 과소추정 문제 보완 (Makeham, 1860)
CK	$m_x = e^{a+bx+cx^2}$	3	k_x 가 연령에 대한 선형함수임을 가정 (Coale과 Kisker, 1990)
HPC	$m_x = \frac{be^{ax}}{1+be^{ax}}$	2	사망률의 로지스틱 값이 연령에 대한 선형함수임을 가정 (Himes 등, 1994)
BRD	$m_x = \frac{be^{ax}}{1+ce^{ax}}$	3	HPC 모형에 모수 1개 추가 (Beard, 1971)
PRS	$m_x = \frac{d+be^{ax}}{1+ce^{ax}}$	4	HPC 모형에 모수 2개 추가 (Perks, 1932)
HP	$q_x = \frac{be^{ax}}{1+be^{ax}}$	4	사망확률의 로지스틱 값이 연령에 대해 선형함수임을 가정 (Heligman과 Pollard, 1980)
WEIB	$m_x = bx^a$	2	초고령층의 사망률이 지수함수와 로지스틱 함수의 중간형태임을 가정 (Weibull, 1951)
S-WEIB	$m_x = c + bx^a$	3	WEIB 모형에서 모수 1개 추가
QUAD	$m_x = a + bx + cx^2$	3	사망률이 연령에 대한 2차 함수임을 가정
CUB	$m_x = a + bx + cx^2 + dx^3$	4	사망률이 연령에 대한 3차 함수임을 가정
MCK	$\log(q_x/q_{x-1}) = k_{85} + (x - 85)s$	1	$k_{85} = \log(q_{88}/q_{81})/7$ 이며, 85세에서 99세의 $\log(q_x/q_{x-1})$

수에 의해 추정된 사망률이다. 각 함수의 모수를 추정하기 위한 적합구간은 65세부터 95세 연령까지를 이용하는데, 이는 그림 4.1에서 볼 수 있듯이 95세 이상에서는 k_x 의 임의변동이 심하기 때문이다. 그러나 MCK 모형의 경우 NCHS 및 통계청과 동일하게 85세부터 99세까지를 적합구간으로 한다. 표 4.2는 12개의 함수를 이용하여 65세부터 115세까지의 사망률을 추정한 후 각 함수의 적합도를 비교한 결과이다.

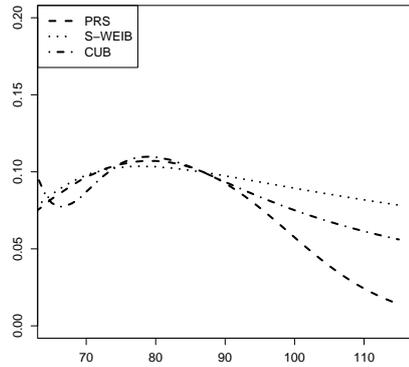
표 4.2에 따르면 남자는 BRD, QUAD 및 MCK의 경우 m_x 의 MAPE가 10% 이상으로 체계적 편이가 나타나고 있으며, SSE 역시 다른 함수에 비해 높은 값을 가지고 있다. 또한 BRD의 경우 k_x 의 MAPE 및 SSE가 높은 것으로 나타난다. 여자의 경우 QUAD 및 MCK는 m_x 의 MAPE가 10% 이상으로 체계적 편이가 나타나며, MCK는 m_x 의 SSE 또한 높다. 따라서 본 논문에서는 남자의 경우 총 12개의 함수 가운데 BRD, QUAD 및 MCK를 제외한 9가지 모형과, 여자의 경우 QUAD 및 MCK를 제외한 총 10개의 함수를 고려하도록 한다.

다음으로 두 번째 기준인 일치성에 대해 알아보자. 두 가지 일치성 가운데 우선 초고령에서 k_x 의 감소 여부를 먼저 파악하도록 한다. 적합도에 문제가 없는 함수들 가운데 남녀 모두 PRS, S-WEIB, CUB가 이러한 패턴을 보이고 있다. 그림 4.2는 이들 함수의 성별 k_x 이다.

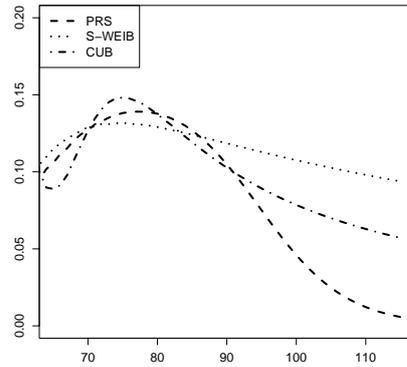
남자의 사망률이 여자의 사망률보다 더 높아야 한다는 두 번째 일치성을 살펴보기 위해, 위에서 선택된 세 가지의 함수를 이용하여 평활된 사망률을 115세까지 외삽(extrapolation) 하도록 한다. 이 때 외삽의 최초시점은 85세 이상의 연령에서 평활된 사망률과 함수로 추정한 사망률의 차이가 최소인 연령으로

표 4.2. 각 함수별 적합도(단위: 10^{-2})

함수	남자				여자			
	m_x		k_x		m_x		k_x	
	MAPE	SSE	MAPE	SSE	MAPE	SSE	MAPE	SSE
COMP	3.94	0.24	21.54	1.39	6.13	0.47	20.40	2.09
MAKE	3.43	0.35	20.87	1.37	6.19	0.51	20.45	2.11
CK	4.59	0.16	21.60	1.39	5.36	0.12	18.98	1.74
HPC	5.56	0.26	18.50	1.24	4.06	0.09	16.26	1.31
BRD	62.23	61.40	100.00	28.69	4.32	0.14	17.06	1.42
PRS	2.31	0.06	14.79	0.75	1.36	0.01	11.02	0.68
HP	4.33	0.16	19.72	1.22	4.76	0.24	18.20	1.61
WEIB	7.06	0.39	22.19	1.71	6.08	0.15	19.85	1.93
S-WEIB	2.90	0.19	18.49	1.07	3.91	0.27	17.16	1.44
QUAD	11.26	1.78	25.43	2.47	15.96	3.17	28.49	4.91
CUB	1.88	0.09	15.60	0.77	2.43	0.04	13.51	0.92
MCK	19.65	17.90	46.43	2.30	17.79	11.23	36.58	2.32



(a) 남자



(b) 여자

그림 4.2. 남녀 3개 모형의 k_x

표 4.3. 함수별 최고 연령

함수명	남자		여자	
	최고 연령	사망률	최고 연령	사망률
PRS	115	0.78469	115	0.50095
S-WEIB	107	1	106	1
CUB	110	1	113	1

결정하였다. 즉, 남자의 경우 PRS는 85세부터, S-WEIB는 92세부터, CUB의 경우 역시 92세부터 외삽한다. 여자의 경우 PRS는 88세부터, S-WEIB는 92세부터, CUB의 경우 85세부터 외삽한다. 각 함수를 이용하여 115세까지의 사망률을 확장하고, 이 때 사망률이 100%에 이르는 연령 및 이에 못 미치는 경우 115세에서의 사망률은 표 4.3과 같다.

표 4.3에 따르면 PRS는 남자의 경우 115세에서의 사망률이 0.785이며, 여자의 경우 역시 0.500에 불과하다. 그러나 우리나라의 경우 97년부터 2007년까지 110세 이상의 사망자 수는 남자의 경우 40명

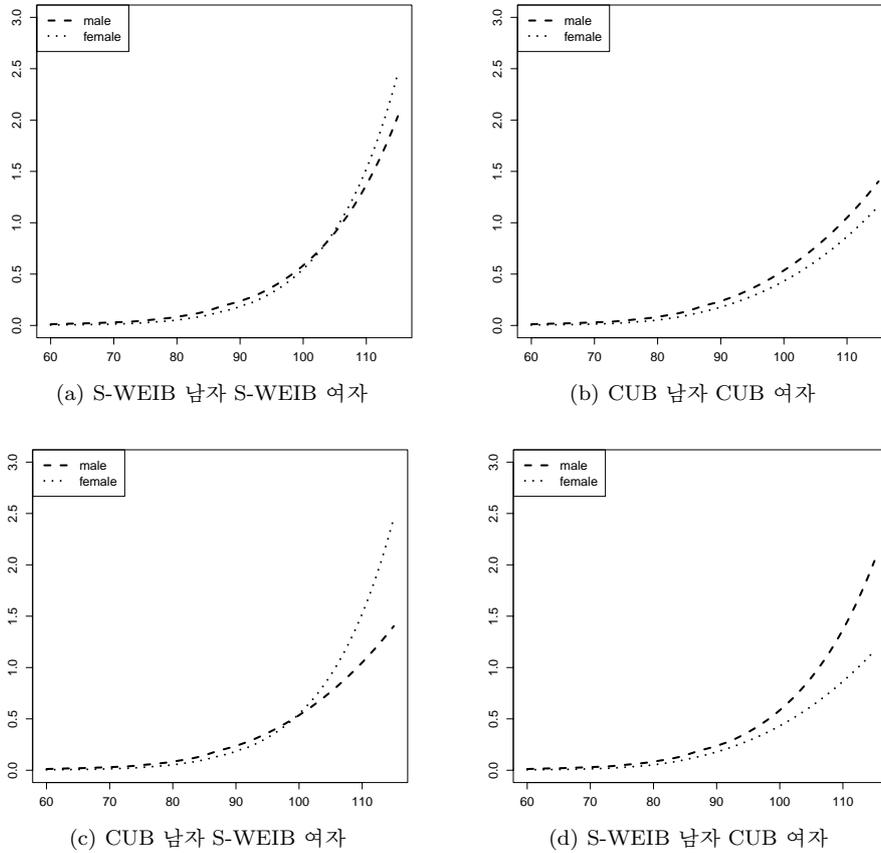


그림 4.3. S-WEIB 와 CUB 함수의 남녀 조합

이며, 여자의 경우 215명이다. 10년 동안 총 사망자 수가 남자의 경우 1,487,355이며, 여자의 경우 1,195,892임을 감안할 때, PRS는 사망률을 지나치게 과소 추정하는 것으로 판단된다. 따라서 PRS 모형을 제외하고, S-WEIB와 CUB를 이용하여 두 번째 일치성을 파악해보도록 한다. 그림 4.3은 남녀별로 60세 이상의 S-WEIB와 CUB 모형의 사망률 조합이다.

위 그림에 따르면 여자의 경우 S-WEIB를 이용하면, 남자의 함수로 S-WEIB나 CUB 가운데 어떤 것을 이용하던지 상관없이 초고령에서 여자의 사망률이 남자의 사망률보다 높아지는 역전현상이 발생한다. 따라서 남녀별 일치성을 만족하는 함수로 여자는 CUB를 선택하도록 하며, 남자는 S-WEIB와 CUB를 선택한다.

이제 세 번째 기준인 경험 기대여명과 모형 기대여명(model life expectancy)의 차이를 파악함으로써 S-WEIB와 CUB 함수를 비교해 보도록 하자. 표 4.4은 115세까지의 사망률 확장 없이 100세까지의 사망률을 평활한 후, 이를 이용하여 구한 경험기대여명과 각 함수를 이용하여 115세까지 사망률을 확장한 후 구한 10세 단위의 기대여명이다.

그림 4.1에서 볼 수 있듯이 우리나라의 사망률은 평활하더라도 95세 이상에서 임의변동이 심하므로 95세 미만의 연령에 대하여 S-WEIB와 CUB 함수를 이용한 모형기대여명과 경험기대여명을 비교하

표 4.4. 함수별 최고 연령

연령	남자			여자	
	경험	S-WEIB	CUB	경험	CUB
0	74.92	74.92	74.92	81.59	81.63
1	74.25	74.24	74.25	80.92	80.96
10	65.42	65.41	65.42	72.06	72.10
20	55.59	55.58	55.58	62.18	62.22
30	45.94	45.93	45.93	52.42	52.46
40	36.49	36.48	36.49	42.74	42.78
50	27.63	27.62	27.62	33.23	33.28
60	19.45	19.44	19.44	24.00	24.05
70	12.42	12.41	12.42	15.44	15.49
80	6.80	6.79	6.79	8.41	8.47
85	4.76	4.74	4.75	5.84	5.93
90	3.42	3.37	3.39	4.04	4.15
95	2.49	2.32	2.40	2.80	2.91
100	1.56	1.57	1.72	1.85	2.08

도록 한다. 표 4.4에 따르면 남자의 경우 CUB 함수가 95세 미만 전 연령에서 S-WEIB 보다 경험기대여명과의 차이가 작음을 알 수 있다. 따라서 남녀 모두 최종적으로 CUB 함수를 선택하여 초고령의 사망률을 추정하도록 한다. 5장에서는 이 방법을 이용하여 2005년부터 2009년까지의 우리나라 생명표를 구해보도록 하자.

5. 2005년부터 2009년의 생명표 작성

우리나라의 경우 매 5년마다 센서스가 이뤄지고 있기 때문에 그 사이 기간의 생명표는 센서스후 인구(postcensal estimates)를 이용하여 생명표를 구하고 있다. 미국의 경우 매 10년마다 센서스가 이뤄지고 있으며, 미국의 경우에도 그 사이 기간의 경우 동일하게 생명표를 작성한다. 이 때 작성되는 생명표는 임시적인(preliminary) 성격을 가지고 있으며, 센서스가 이뤄진 후에는 센서스간 인구(intercensal estimates)를 이용하여 생명표를 수정하고 있다 (Anderson과 Arias, 2003). 본 논문에서는 2005년은 센서스 인구, 2006년부터 2009년까지는 센서스후 인구를 이용하여 생명표를 구해보도록 한다. 단, 각 연도별 자료는 Greville 방법으로 평활하며, 평활 후 각각 CUB 함수를 적합하여 초고령에서의 사망률을 확장한다. 이 때 각 연도별로 확장된 사망률의 시작 연령은 4절과 동일하게 85세 이상 연령에서 확장된 사망률과 관측된 사망률의 차이가 가장 작은 연령으로 설정하였다.

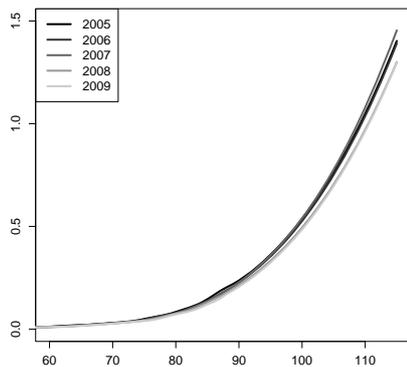
기간별로 생명표를 구했을 때, 동일한 연령에서 나중의 기대여명이 이전의 기대여명보다 더 낮아지는 역전현상(crossover)이 발생할 수 있다. 실제로 CUB 함수를 이용하여 2005년부터 2009년까지의 생명표를 작성한 결과, 표 5.1과 같이 역전현상이 발생하였다.

위 표에 따르면 남녀 평균수명은 시간이 지남에 따라 더 커짐에 반해, 남자의 경우 2007년의 85세 이상 연령에서의 기대여명이 2006년 85세 이상 연령에서의 기대여명보다 작으며, 100세 이상의 경우 2005년의 기대여명보다 작다. 여자의 경우는 2006년 95세 이상 연령에서의 기대여명이 2005년의 기대여명보다 작다. 이와 같이 초고령에서 기대여명의 역전현상이 일어나는 이유는 초고령에서 사망률에 역전현상이 일어나기 때문이다. 그림 5.1은 남녀별 2005년에서 2009년까지 85세에서 115세까지의 사망률이다.

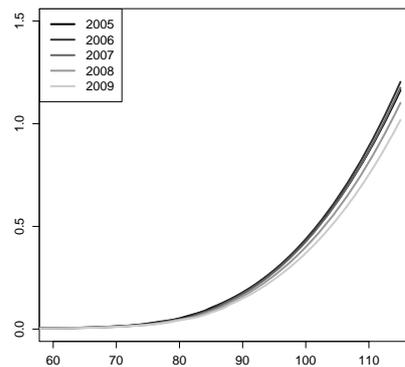
위 그림에서 남녀 모두 사망률의 역전현상이 발생한 것을 볼 수 있다. 그러나 Buettner (2002)은 초고

표 5.1. 2005-2009년 생명표

연령	2005		2006		2007		2008		2009	
	남자	여자								
0	74.92	81.63	75.53	82.20	75.92	82.50	76.37	83.06	76.79	83.51
1	74.25	80.96	74.85	81.49	75.22	81.79	75.62	82.32	76.04	82.77
10	65.42	72.10	66.00	72.64	66.37	72.92	66.76	73.44	67.17	73.88
20	55.58	62.22	56.17	62.75	56.55	63.02	56.93	63.55	57.34	64.00
30	45.93	52.46	46.48	52.95	46.88	53.28	47.27	53.80	47.68	54.27
40	36.49	42.78	36.98	43.23	37.37	43.61	37.75	44.12	38.20	44.62
50	27.62	33.28	28.06	33.69	28.42	34.08	28.78	34.59	29.23	35.09
60	19.44	24.05	19.84	24.44	20.14	24.79	20.49	25.30	20.92	25.78
70	12.42	15.49	12.69	15.79	12.89	16.10	13.14	16.55	13.52	17.01
80	6.79	8.47	7.01	8.63	7.09	8.89	7.31	9.21	7.60	9.61
85	4.75	5.93	4.96	6.00	4.95	6.18	5.17	6.45	5.38	6.79
90	3.39	4.15	3.48	4.16	3.45	4.26	3.65	4.47	3.74	4.72
95	2.40	2.91	2.45	2.90	2.41	2.97	2.58	3.13	2.63	3.33
100	1.72	2.08	1.75	2.05	1.70	2.10	1.85	2.22	1.88	2.37



(a) 남자



(b) 여자

그림 5.1. CUB 함수를 이용한 연도별 사망률

령에서의 기간별 사망률의 순서가 0세에서의 기대여명의 순서와 일치해야 한다고 밝히고 있으며, 이는 고정된 연령과 성(gender)에서 시간이 지날수록 사망률이 감소 또는 최소한 증가하지 않아야 하는 종단면적인(longitudinal) 관점에서의 일치성으로 해석될 수 있다. 따라서 본 논문에서는 기간별 사망률의 종단면적 일치성을 해결하기 위한 방법을 제안하고자 한다.

그림 5.2는 2005년부터 2009년까지의 사망률을 65세부터 10세단위로 그린 사망률의 시계열이다. 여기서 65세부터 85세까지의 사망률은 관측된 사망률을 평활한 값이며, 95세 및 105세의 사망률은 CUB 함수를 이용하여 확장한 값이다.

위 그림에 따르면 남녀 모두 실제 사망률인 65세부터 85세까지의 사망률은 시간이 지남에 따라 선형적으로 감소하는 패턴을 보임을 알 수 있다. 지면상 생략하지만 45세부터 85세까지의 사망률 역시 이와 동일한 패턴을 보이고 있다. 그러나 95세 및 105세에서의 확장된 사망률의 경우 특정 연도에서 사망률이 과거보다 더 높아지는 역전현상이 발생하고 있다. 따라서 본 논문에서는 확장된 사망률의 연령대에

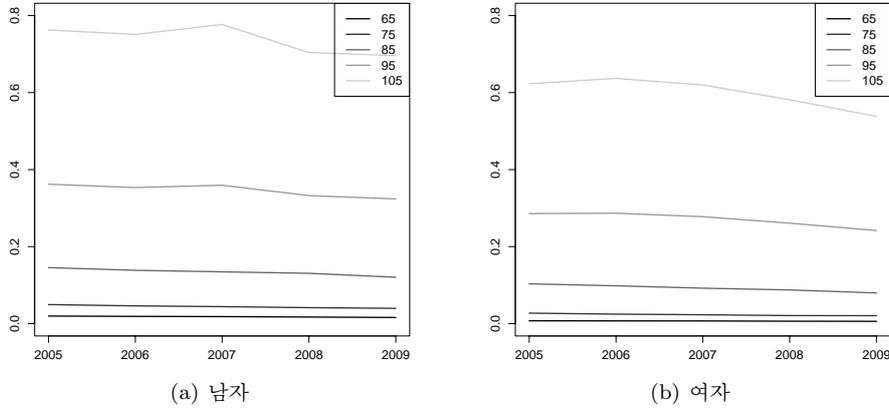


그림 5.2. 연령별 사망률의 시계열

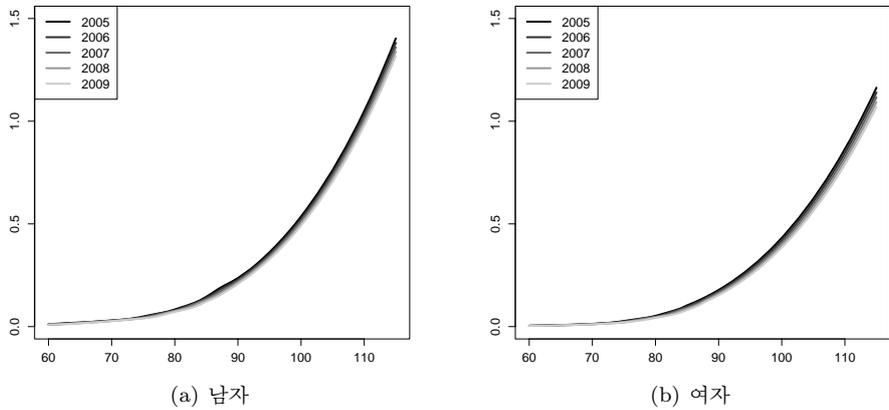


그림 5.3. 연도별 보정 사망률

서 발생하는 중단면적인 일치성의 문제를 해결하고, 연령별 사망률의 선형적 감소추세를 반영하기 위해 다음의 조건부 시계열 단순회귀모형을 고려하고자 한다.

$$\hat{m}_{x,t} = \alpha_0 + \alpha_1 t + \epsilon_t, \tag{5.1}$$

여기에서 $\hat{m}_{x,t}$ 는 시점 t 에서 확장된 x 세의 사망률이다. 위 식은 $\alpha_1 \leq 0$ 을 만족해야 하며 이회귀식은 가장 최근의 센서스 시점과 이에 대응하는 사망률을 지나야 한다. 예를 들어, 본 논문에서는 가장 최근의 센서스자료가 2005년이므로, $\hat{m}_{x,2005} = \hat{\alpha}_0 + \hat{\alpha}_1 \times 2005$ 가 성립되어야 한다. 이러한 조건을 부여하는 이유는 2006년부터 2009년까지의 사망률은 센서스후인구에 기반하므로 임시적 성격을 가지고 있는 반면에 2005년의 경우 확정된 인구이기 때문이다. $\alpha_1 \leq 0$ 인 이유는 연령별 사망률의 시계열이 감소해야 하는 중단면적 시계열의 특성을 반영하기 위함이다. 그림 5.3은 위 식으로 확장된 사망률을 보정한 후의 2005년부터 2009년까지의 사망률이다.

그림 5.3에서 알 수 있듯이 보정 후의 기간별 사망률은 역전현상이 발생하지 않음을 알 수 있으며, 보정된 사망률을 이용하여 2006년부터 2009년의 생명표를 구해보면 기대여명에서도 역시 역전현상이 일어

표 5.2. 보정된 사망률을 이용한 2005-2009년 생명표

연령	2005		2006		2007		2008		2009	
	남자	여자								
0	74.92	81.63	75.51	82.25	75.94	82.55	76.37	83.07	76.78	83.45
1	74.25	80.96	74.83	81.55	75.24	81.84	75.63	82.33	76.03	82.70
10	65.42	72.10	65.98	72.69	66.39	72.97	66.77	73.45	67.15	73.82
20	55.58	62.22	56.15	62.80	56.58	63.08	56.93	63.56	57.32	63.93
30	45.93	52.46	46.46	53.00	46.91	53.34	47.27	53.81	47.67	54.20
40	36.49	42.78	36.96	43.29	37.39	43.67	37.76	44.13	38.19	44.55
50	27.62	33.28	28.04	33.75	28.44	34.14	28.79	34.61	29.22	35.02
60	19.44	24.05	19.82	24.49	20.17	24.85	20.49	25.32	20.91	25.71
70	12.42	15.49	12.66	15.85	12.92	16.16	13.14	16.57	13.51	16.94
80	6.79	8.47	6.96	8.70	7.14	8.97	7.32	9.23	7.58	9.52
85	4.75	5.93	4.89	6.10	5.03	6.28	5.18	6.47	5.34	6.67
90	3.39	4.15	3.47	4.26	3.54	4.37	3.62	4.49	3.71	4.61
95	2.40	2.91	2.45	2.99	2.50	3.06	2.55	3.14	2.60	3.23
100	1.72	2.08	1.75	2.12	1.78	2.18	1.82	2.23	1.85	2.29

나지 않아 종단면적 일치성 역시 만족하게 된다. 표 5.2는 보정된 사망률을 이용하여 구한 2005년부터 2009년까지의 생명표를 10세 단위로 나타낸 표이다.

6. 결론 및 향후 과제

본 논문에서는 완전생명표를 작성하기 위해 고려되어야 하는 세 가지 제반 사항인 사망확률의 추정방법, 평활방법, 초고령에서의 사망률의 확장 방법에 대해 다루었다. 첫 번째 사항인 사망률의 추정방법의 경우 5가지의 방법을 비교하였으며, 간이생명표에서는 방법별로 차이가 발생하는 것으로 나타났다. 이 때 각 방법에 따라 남자의 경우 평균수명이 0.27년 차이가 났으며, 여자의 경우 0.37년 차이가 발생하였다. 본 논문에서는 이러한 차이가 발생하는 원인에 대해서는 밝히지 않았으나, 추후 코호트 자료 및 동태자료를 이용하여 그 원인을 밝힐 필요가 있다고 판단된다. 완전생명표에서는 방법별 차이가 없는 것으로 나타났다. 두 번째 사항인 평활방법의 경우 Beers 방법은 우리나라에 적합하지 않음을 보였으며, 세 가지의 방법을 비교했을 때 Greville의 방법이 가장 우수한 것으로 나타났다. 세 번째 사항인 초고령에서의 사망률을 추정하기 위한 함수로는 총 12개의 함수 가운데 CUB 함수가 가장 우리나라 자료에 적합한 것으로 나타났다. 이러한 결론에 따라 본 논문에서는 Chiang 방법, Greville 방법, CUB 함수를 이용하여 2005년부터 2009년까지의 생명표를 제시하였으며, 이 때 발생하는 기간별 역전현상을 해결하기 위한 보정식을 제시하였다. 그러나 본 논문에서 구한 2006년부터 2009년의 생명표는 센서스 후 인구에 기반한 것이므로 임시적이며, 2010년 센서스 결과 발표 후 센서스간 인구를 구하여 이를 재추정해야 할 것이다.

본 논문에서 이용한 자료는 2005년 센서스 자료에 기반한 것이므로, 본 논문에서 제시한 결과는 한시적일 수 밖에 없다. 따라서 추후 발표되는 센서스 자료 및 동태자료에 기반하여 이를 재평가할 필요가 있다고 판단된다.

참고문헌

- 박유성, 박성근, 최보승, 김기환 (2009). Beers와 Greville 보정방법에 의한 사망확률 보정방법 비교 연구, *Journal of the Korean Data Analysis Society*, 11, 97-110.

- 통계청 (2010). <2009년 생명표>, 통계청.
- Anderson, R. N. (1999). A method for constructing complete annual U.S. life tables, *Vital Health Statistics, Series 2*, No.129, NCHS.
- Anderson, R. N. and Arias, E. (2003). The effect of revised populations on mortality statistics for the United States, 2000, *National Vital Statistics Reports*, **51**, No 9.
- Arias, E. (2010). United states life table 2006, *National Vital Statistics Reports*, **58**, No 21.
- Beard, R. E. (1971). Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes, In W.Brass (Eds.), *Biological Aspects of Demography*, Taylor and Francis, London.
- Beers, H. S. (1944). Six-term formulas for routine actuarial interpolation, *American Institute of Actuaries*, **33**, 245–260.
- Boleslawski, L. and Tabeau, E. (2001). Comparing theoretical age patterns of mortality beyond the age of 80, In E. Tabeau, A. Van Den Berg Jeths, and C. Heathcote (Eds.), *Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective*, Kluwer Academic Publishers, Dordrecht.
- Buettner, T. (2002). Approaches and experiences in projecting mortality patterns for the oldest-old, *North American Actuarial Journal*, **6**, 14–29.
- Chiang, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics*, John Wiley and Sons, New York.
- Chiang, C. L. (1984). *The Life Table and its Applications*, Robert E. Krieger Publishing Co., Malabar, Florida.
- Coale, A. J. and Demeny, P. (1983). *Regional Model Life Tables and Stable Populations*, Academic Press, New York.
- Coale, A. J. and Kisker, E. E. (1986). Mortality crossovers: Reality or bad data?, *Population Studies*, **40**, 389–401.
- Coale, A. J. and Kisker, E. E. (1990). Defects in data on old-age mortality in the United States: New procedures for calculating schedules and life tables at the highest ages, *Asian and Pacific Population Forum*, **4**, 1–31.
- Condran, G., Himes, C. L. and Preston, S. H. (1991). Old age mortality patterns in low-mortality countries: An evaluation of population and death data at advanced ages, 1950 to present, *Population Bulletin of the United Nations*, **30**, 23–61.
- Gompertz, B. (1825). On the nature of the function expressive of the law of mortality, *Philosophical Transactions*, **27**, 513–585.
- Greville, T. N. E. (1943). Short methods of constructing abridged life tables, *The Record of the American Institute of Actuaries*, **32**, 29–42.
- Greville, T. N. E. (1945). Actuarial Note : Some Extensions of Mr. Beers's Method of Interpolation, *American Institute of Actuaries*, **34**, 21–34.
- Heligman, L. and Pollard, J. H. (1980). The age pattern of mortality, *Journal of the Institute of Actuaries*, **107**, 49–80.
- Hill, K. (1999). The Measurement of Adult Mortality: An Assessment of Data Availability, Data Quality and Estimation Methods. In Health and Mortality Issues of Global Concerns, *Proceedings of the Symposium on Health and Mortality*, Brussels: Population and Family Study Centre.
- Himes, C. L., Preston, S. H. and Condran, G. (1994). A relational model of mortality at older ages in low mortality countries, *Population Studies*, **2**, 269–291.
- Keyfitz, N. (1966). A life table that agrees with the data, *Journal of the American Statistical Association*, **61**, 305–12.
- Keyfitz, N. and Frauenthal, J. (1975). An improved life table method, *Biometrics*, **31**, 889–899.
- London, R. L. (1981). In defense of minimum-Ro linear compound graduation, and a simple modification for its improvement, *ARCH, Chicago: Society of Actuaries*, 75–88.
- Makeham, W. (1860). On the law of mortality and the construction of annuity tables, *Journal of the Institute of Actuaries*, **13**, 325–358.
- Ng, E. and Gentleman, J. F. (1995). The impact of estimation method and population adjustment on Canadian life table estimates, *Health Reports*, **7**, 15–21.
- Perks, W. (1932). On some experiments in the graduation of mortality statistics, *Journal of the Institute of Actuaries*, **63**, 12–27.

- Preston, S. H., Heuveline, P. and Guillot, M. (2000). *Demography: Measuring and Modeling Population Processes*, Blackwell Publishing Ltd, Oxford.
- Reed, L. J. and Merrell, M. (1939). A short method for constructing an abridged life table, *American Journal of Hygiene*, **30**, 33–62.
- Shryock, H. S., Siegel, J. S. and Associates (1993). Interpolation: Selected General Methods, In: D.J. Bogue, E.E. Arriaga, D.L. Anderton and G.W. Rumsey (Eds.), *Readings in Population Research Methodology. Vol. 1: Basic Tools. Social Development, Center / UN Population Fund*, Chicago.
- Siegel, J. S. and Swanson, D. A. (2004). *The Methods and Materials of Demography*, Elsevier Academic Press, San Diego.
- Sprague, T. B. (1881). Explanation of a new formula for interpolation, *Journal of the Institute of Actuaries*, **22**, 1880–1881.
- Weibull, W. (1951). A statistical distribution function of wide applicability, *Journal of Applied Mechanics*, **18**, 293–297.
- Wilmoth, J. R. (1995). Are mortality rates falling at extremely high ages? An investigation based on a model proposed by Coale and Kisker, *Population Studies*, **49**, 281–295.
- World Health Organization (1977). *Manual of Mortality Analysis*, WHO, Geneva.

A Method for Construction of Life Table in Korea

Yousung Park¹ · Seongyong Kim²

¹Department of Statistics, Korea University

²Economics and Statistics Institute, Korea University Sejong Campus

(Received June 2011; accepted August 2011)

Abstract

The life table is a statistical model for life expectancy and reflects mortality experiences exposed to a particular group of people. The following three issues are prerequisite for constructing the life table : a selection of how to estimate the death probability from observed death rates, a graduation method to smooth irregularity of the death probabilities, and an extension method of the death probabilities for oldest-old ages. To construct the life table that is fittest to Korean mortality experiences, we examine five estimation methods such as Chiang's and Greville's for the death probability, three graduation techniques including Beer's and Greville's formulae, and twelve mathematical functions for the extension of death probabilities for oldest-old ages. We also propose a method to resolve the cross-over problem arising from construction the life table.

Keywords: Life table, death probability, smoothing, death rates at oldest-old age.

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-411-B00028).

²Corresponding author: Research Professor, Economics and Statistics Institute, Korea University Sejong Campus, 2511 Sejong-Ro, Jochiwon-Eup, Yeongi-Gun, Chungnam 339-700, Korea.

E-mail: yaba96@korea.ac.kr