中国中部六省预期寿命时序加密估算研究

李 畅,王安丽,龚胜生,孙攸宁
(湖北省地理过程分析与模拟重点实验室 华中师范大学城市与环境科学学院,武汉 430079)

摘要: 年龄组死亡率是利用年龄分组人口数据计算预期寿命的关键参数,而非采样年份的统计年鉴中年龄分组死亡率缺失导致无法计算预期寿命。针对该问题,本文将人口普查数据与统计年鉴数据融合,首次提出一种基于拉格朗日插值的中国省级预期寿命时间序列加强密集度(时序加密)的算法,以解决非采样(即未进行人口普查或1%人口抽样调查)年份省级预期寿命的估算问题。以中国中部六省为例,在所选取年份省级预期寿命估算实验中,绝对精度表明年龄分组人口比例线性插值计算的精度明显高于人口比例抛物线插值和直接插值算法的精度,故为推荐算法。本研究为高时间分辨率下省级预期寿命值的获取提供了一个新的可行思路,为分省较精确地进行预期寿命趋势分析奠定基础。

关键词: 省级预期寿命;时序加密;中国中部六省;拉格朗日插值;线性插值;二次多项式插值
DOI: 10.11821/dlxb202010016

1 引言

健康是人类发展的永恒主题。李克强总理于2016年3月9日参加十二届全国人大四次会议时强调:“健康也是生产力”。健康问题在地理学中一直备受关注[1]。综合反映人口健康状况的预期寿命(Life Expectancy)可以间接体现一个国家或地区现阶段的经济发展水平,也是反映民生状况的核心指标之一。预期寿命是指已经活到一定岁数的人平均还能再活的年数。在不特别指明岁数的情况下,是指0岁人口的平均预期寿命[2]。计算预期寿命需要基于寿命表[3]。寿命表是根据特定人群的年龄组死亡率编制出来的一种统计表(表中其他指标均可由完整年龄分组死亡率推算且年龄组死亡率=年龄组实际死亡数/年龄组平均人口数),反映人群的生命或死亡过程,编制中需要完整的人口与死亡数据。寿命表根据编制的年龄组组距不同,可分为完全寿命表(以0岁为起点直至生命的极限,年龄组组距为1岁)和简略寿命表(除第1年外,各年龄组均大于1年,一般年龄组组距为5岁)[4]。后者年龄分组少,每个年龄组人口数比较多,各年龄组死亡率较为稳定,计算量也较少,是卫生统计中比较常用的方法[5]。预期寿命水平不受人口年龄构成的影响,各地区可以直接比较[6]。

当前,学术界对预期寿命计算的研究多致力于提高采样年份(即人口普查或1%人口

收稿日期: 2019-03-29; **修订日期:** 2020-04-05
基金项目: 国家社会科学基金项目(11AZD117, 12&ZD145); 国家自然科学基金项目(41171408, 41771493, 41101407) [**Foundation:** National Social Science Foundation of China, No.11AZD117, No.12&ZD145; National Natural Science Foundation of China, No.41171408, No.41771493, No.41101407]
作者简介: 李畅(1982-), 男, 湖北武汉人, 博士, 教授, 硕士生导师, 主要从事遥感与地理信息科学理论、方法、技术及其应用研究。E-mail: lcshaka@126.com; lichang@mail.ccnu.edu.cn
通讯作者: 龚胜生(1965-), 男, 湖南涟源人, 博士, 教授, 博士生导师, 主要从事历史地理、健康地理和可持续发展研究。E-mail: shshgong@mail.ccnu.edu.cn

抽样调查年份）预期寿命估算的精度[7-9]。而对于非采样年份预期寿命计算研究尚未见报道。预期寿命受很多因素影响[10-14]，为更好地分析预期寿命变化趋势，则需要对其进行高时间分辨率的研究。然而，寿命表能否编制取决于完整年龄分组死亡率可否获取，其在中国只在采样年份可选取公开的年龄分组人口（含死亡人口）进行计算，而统计年鉴（非采样）数据没有死亡率数据，导致无法计算预期寿命，因此，编制寿命表计算省级预期寿命一直局限于采样年份。然而，人口普查每10年进行一次调查，1%人口抽样调查只在逢“5”的年份进行，中间年份间隔过长，统计年鉴中分年龄段人口虽每年统计一次，但年龄分段过大（如0~14岁、15~64岁、65岁以上）且相应年龄分段死亡人口缺失，无法计算与寿命表一致的年龄分组死亡率，因而无法编制简略寿命表，省级预期寿命也难以在非采样年份计算。

因此，寻求科学合理的方法融合统计年鉴、人口普查和1%人口抽样调查的年龄分组人口数据，进行中间缺失完整年龄分组死亡率年份省级预期寿命时序加密（即时间序列加强密集度，亦估算缺失年份的预期寿命）估算显得十分必要。另外，时间序列上也存在类似于空间上的邻近效应（地理学第一定律），即：时间上（空间上）相近的事物会更相似，并考虑到时空变量的离散函数逼近原理。基于该思路，本文提出基于拉格朗日插值估算非采样年份缺失的完整年龄分组死亡率相关计算指标——年龄分组人口比例（包括年龄组平均人口占所在大年龄段人口比例和年龄组实际死亡人口占总平均人口比例）的方法，进行非采样年份省级预期寿命估算（注：为保证精度，本文仅限于内插估算，不包括外推估算）。考虑到中部地区是贯穿南北和东西部的核心腹地位置，且通过后面实验验证该区域的预期寿命具有一定的空间代表性。所以，最后以中部六省为例，对省级预期寿命估算的绝对精度进行比较，确定最优算法。

2 理论与方法

2.1 时序加密估算

正文本文选用简略寿命表（表1）计算省级预期寿命（e_x）（除0岁组和1~4岁组外，其他年龄组分组间隔一般为5），编制该表时选取由世界卫生组织一直推荐的蒋庆琅法[15]（表1中所有参数只需根据相应年龄组死亡率（${}_nm_x$）依次推算）。表1中年龄组平均人口数（${}_nP_x$）和年龄组实际死亡数（${}_nD_x$）均出自中国2010年人口普查数据安徽部分，x表示

表1 2010年安徽省简略寿命表

Tab. 1 Abridged life table of Anhui province in 2010

年龄组(岁)	平均人口数	实际死亡数	死亡率	死亡概率	尚存人数	死亡人数	生存人年数	生存总人年数	预期寿命(岁)
X~	${}_nP_x$	${}_nD_x$	${}_nm_x$	${}_nq_x$	l_x	${}_nd_x$	${}_nL_x$	T_x	e_x
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	748259	3641	–	0.004866	100000	487	99586	7853094	78.53
1~	2945073	1419	0.000482	0.001925	99513	192	397670	7753507	77.91
5~	3325594	810	0.000244	0.001217	99322	121	496307	7355837	74.06
10~	3557210	973	0.000274	0.001367	99201	136	495666	6859530	69.15
15~	4744232	1846	0.000389	0.001944	99065	193	494845	6363865	64.24
…	…	…	…	…	…	…	…	…	…
95~	16114	2928	0.181705	0.624733	7969	4978	27397	37280	4.68
100~	1996	604	0.302605	1.000000	2990	2990	9882	9882	3.30

年龄组的下限值，n表示年龄组的组距。其余相关参数含义及计算公式[16-17]如下：

① 年龄“X~”表示刚满年龄，比如“1~”即是指刚满1岁的孩童；年龄组死亡率（${}_nm_x$）指某年龄组人口（x岁）在n年内的平均死亡水平，可根据${}_nP_x$和${}_nD_x$计算，公式如下：

$$ {}_nm_x = {}_nD_x / {}_nP_x \tag{1} $$

② 年龄组死亡概率（${}_nq_x$）表示假想的同一时间出生的一代人，刚满x岁的尚存者在今后n年内死亡的可能性，可由${}_nm_x$计算求得。编制寿命表过程中，0岁组死亡概率通常用婴儿死亡率（或校正婴儿死亡率）近似估计，记为${}_1m_0$；最后一个年龄组（用φ表示）的死亡概率为1；其他各年龄组中，在年龄组距$n \leqslant 5$的情况下，${}_nq_x$与${}_nm_x$近似函数关系表达式如下：

$$ {}_nq_x = \begin{cases} {}_1m_0, & x=0 \\ 2n_nm_x/(2+n_nm_x), & 0<x<\varphi \\ 1, & x=\varphi \end{cases} \tag{2} $$

③ 尚存人数（l_x）表示假想的同一时间出生的一代人到刚活满x岁时尚生存的人数。一般假定“0~”岁组的人数为：$l_0=100000$；死亡人数（${}_nd_x$）是假想的同一时间出生的一代人死于各年龄组x~(x+n)的人数。二者的关系表达式如下：

$$ {}_nd_x = l_x \times {}_nq_x \tag{3} $$

$$ l_x + n = l_x - {}_nd_x \tag{4} $$

④ 生存人年数（${}_nL_x$）是指x岁尚存者在今后n年内的生存人年数，即l_x曲线下，x~(x+n)间的面积，其计算公式如下：

$$ {}_nL_x = \begin{cases} l_1 + a_{01}d_0, & x=0 \\ 0.5n(l_x + l_x + n), & 0<x<\varphi \\ l_\varphi / {}_nm_\varphi, & x=\varphi \end{cases} \tag{5} $$

式中：${}_1L_0$（x = 0时）应将“0~”岁组死亡者的平均存活年数统计计算在内，a_0表示每个死亡婴儿的平均存活年数，可以依据婴儿死亡率与a_0关系的经验系数进行推算。a_0的经验系数：男性是0.1450，女性是0.1525，男女合计为0.1500。${}_1d_0$为0岁组的死亡人数；${}_nm_\varphi$为最后一个年龄组的死亡率。

⑤ 生存总人年数（T_x）指的是活满x岁者今后尚能生存的总人年数，也就是x岁及以上各年龄组${}_nL_x$的总和，生存总人年数计算时由下向上累计，计算公式如下：

$$ T_x = \sum {}_nL_x \tag{6} $$

⑥ 预期寿命（e_x）指的是活满x岁者今后尚能存活的年数（即为岁数），计算公式如下：

$$ e_x = \frac{T_x}{l_x} \tag{7} $$

需要说明的是，采样年份中，官方公布的省级预期寿命数值在按上述方法（即蒋庆琅法）计算之前，对年龄组死亡率（${}_nm_x$）结合漏报率进行了处理，因此，按照以上方法直接计算出的省级预期寿命在保持与官方数据总趋势一致下会略有出入。且为保证精度，本文仅限于内插估算，不包括外推估算。

中国省级预期寿命时序加密估算的技术流程如图1所示。考虑到人口普查年份（2000年、2010年）和1%人口抽样调查年份（2005年）具有“真值/参考值”可用作精度验证，故可选为插值节点，按照内插及邻近选取插值节点的原则进行研究。如表2所

示，可获取人口数据的年份包括1995年（部分数据）、2000年、2005年、2010年，可获取官方发布预期寿命的年份包括1990年、2000年、2010年。为保证实验精度，本文进行实验均为内推实验，进行线性插值时需要选取研究年份前后两个时间点为节点进行插值，而进行抛物线插值时需要选取研究年份前后3个时间节点进行插值，又2000年有官方发布的预期寿命作为参考，因此选取2000年作为验证年份。

2.1.1 针对采样年份（2000、2005、2010年） 从中国人口普查数据和1%人口抽样调查数据中提取${}_nP_x$和${}_nD_x$，利用基于蒋庆琅法的简略寿命表计算${}_nm_x$（公式（1）），进而算出相应年份的e_x。

2.1.2 针对非采样年份 对于本文研究时段1997—2013年的非采样年份中，有的年份的年龄分段如表3中2011年平均人口抽样数据的年龄分段，更多年份的人口年龄分段则是如表3中2001年平均人口抽样数据的年龄分段。为解决表4中年龄分段和简略寿命表（表1）中年龄分组不一致的问题，可按照表1中年龄分组对非采样年份相应年龄分组的平均人口占所在大年龄段（如0~14岁、15~64岁、65岁以上）人口的比例p_1和相应年龄分组死亡人口占总平均人口的比例p_2进行拉格朗日插值（注：非采样年份均采用与表1一致的年龄分组进行插值与编制简略寿命表）。非采样年份省级预期寿命时序加密估算的具体步骤可细分如下：

① 确定非采样年份插值节点。对于各年份的${}_nP_x$（包括采样年份），首先可根据表4中2001年年龄分段合并为三大年龄段：0~14岁、15~64岁、65岁以上；对于各年份的${}_nD_x$，由于其只在采样年份对外公布（年龄分组如表4所示），非采样年份仅可获取分年龄段平均人口抽样数据的合计人口（表3）。因此，为最大限度利用公开的统计数据，可将采样年份分年龄组数据中各${}_nP_x$占其所在的年龄段（0~14岁、15~64岁、65岁以上）人口数的比例和各${}_nD_x$占总平均人口的比例（即为图2中的a，b，c）作为非采样年份的插值节点。例如安徽省0岁组平均人口的插值节点（2000年：4.4898%）

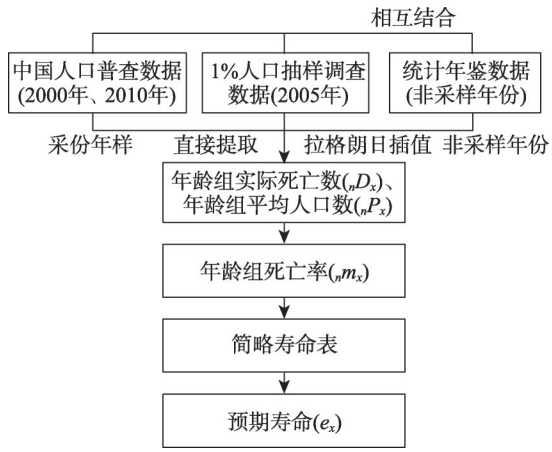


图1 中国省级预期寿命时序加密估算流程图

Fig. 1 The flow chart of time-series estimation of provincial life expectancy in China

表2 可获取人口数据及相应年份

Tab. 2 Available population data and corresponding years

年份	数据类型		
	官方发布预期寿命	分年龄组人口数据	分年龄组1%人口抽样数据
1990	√	×	×
1995	×	×	√
2000	√	√	×
2005	×	×	√
2010	√	√	×
2015	×	×	√

表3 非采样年份安徽省分年龄段平均人口抽样数据

Tab. 3 Sampling data of average population by age group in non-sampling years of Anhui province

2001年		2011年	
年龄段(岁)	${}_nP_x$(人)	年龄段(岁)	${}_nP_x$(人)
0~14	14720	0~4	15701
		5~9	15651
		10~14	15616
15~64	41016	…	…
		60~64	13560
≥ 65	4926	≥ 65	28820
合计	60662	合计	252638

表4　采样年份安徽省分年龄组人口数据及插值节点

Tab. 4　Population data and interpolation nodes by age group in sampling years of Anhui province

年龄组(岁)	2000年(人口普查)				2005年(1%抽样)				2010年(人口普查)			
X~	${}_nP_x$(人)	p_1(%)	${}_nD_x$(人)	p_2(%)	${}_nP_x$(人)	p_1(%)	${}_nD_x$(人)	$p_{2(\%)}$	${}_nP_x$(人)	$p_{1(\%)}$	${}_nD_x$(人)	p_2(%)
0	675278	4.4898	18780	0.0318	9893	5.2885	156	0.0193	748259	7.075	3641	0.0061
1~	2602905	17.3063	3398	0.0058	35996	19.2423	24	0.003	2945073	27.8464	1419	0.0024
5~	5216742	34.6853	2687	0.0046	55090	29.4493	16	0.002	3325594	31.4443	810	0.0014
10~	6545272	43.5185	2251	0.0038	86088	46.0199	21	0.0026	3557210	33.6343	973	0.0016
…	…	…	…	…	…	…	…	…	…	…	…	…
100~	904	0.0202	193	0.0003	37	0.0453	3	0.0004	1996	0.0328	604	0.001
合计	58999948	–	340562	–	810095	–	5027	–	59500468	–	351844	–

为0岁组的平均人口（675278）与其所在的合并后的0~14岁大年龄段人口（675278+2602905+5216742+6545272）之商；0岁组死亡人口的插值节点（2000年：0.0318%）为0岁组的死亡人口（18780）与总平均人口（58999948）之商。同理可由采样年份计算其他年龄组的插值节点。为减少误差，插值节点应参照非采样的年份按照尽量内插及选取已知最邻近节点的原则来确定。

② 计算非采样年份的p_1和p_2。基于非采样年份的插值节点（a, b, c），以拉格朗日二次多项式（抛物线）插值为例（图2）得出非采样年份相应年龄组的p_1、p_2。以安徽省为例，由采样年份计算出0岁组平均人口的插值节点分别为4.4898%（2000年）、5.2885%（2005年）、7.0750%（2010年）；0岁组死亡人口的插值节点分别为0.0318%（2000年）、0.0193%（2005年）、0.0061%（2010年），3个插值节点可确定一条抛物线（图2），可将非采样年份带入，求出对应年龄组平均人口占所在大年龄段的比例和相应年龄组死亡人口占总平均人口的比例，例如所求得2007年0岁年龄组人口占对应年龄段的比例为5.8558%。图2中对非采样年份插值p_1、p_2的结果中若出现负值，说明各年份对应的该年龄组数据不适用于该插值思想，该年龄组插值方式可改为线性插值或负值部分使用邻近插值，不对其他年龄组产生影响；插值过程中合并后的最后一个年龄组实际死亡数以不超过前一年龄组的为宜[15, 18]。

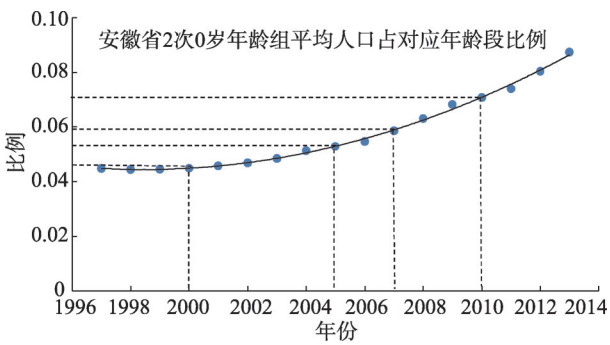


图2　拉格朗日插值中年份及对应年龄分组人口比例示意图

Fig. 2　Proportions estimation of population by age group based on Lagrange interpolation for unknown years

拉格朗日插值计算公式的一般形式[19]如下：

$$L_j(x)=\sum_{k=0}^{j}y_k\frac{\omega_{j+1}(x)}{(x-x_k)\omega'_{j+1}(x_k)},\quad (k=0,1,\cdots,j) \tag{8}$$

$$\omega_{j+1}(x)=(x-x_0)(x-x_1)\cdots(x-x_j) \tag{9}$$

$$\omega'_{j+1}(x_k)=(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_j) \tag{10}$$

式中：$L_j(x)$表示拉格朗日j次插值函数（本文中对应待插的p_1或p_2）；x为自变量（本文中对应相应的待插年份）；y_k为选取的已知插值节点对应的年龄组比例（即已知的p_1或p_2）。

当j = 1时为线性函数，当j = 2时为抛物线（图2），实际应用中，可根据相关数据特点，选取合适的j值。

③ 确定非采样年份的${}_nP_x$和${}_nD_x$。将非采样年份某一年龄组的p_1乘以其在统计年鉴中位于的大年龄段抽样人口数即为该年龄组的${}_nP_x$，如安徽省按照②中方法插值出2001年0岁组的p_1为0.045705，而0岁组在统计年鉴中位于的大年龄段（0~14岁）抽样人口为14720，因此${}_nP_x$ = 0.045705×14720 = 673（人）；将该年龄组的p_2乘以分年龄段抽样人口的合计人口，即为该年龄组的${}_nD_x$，如安徽省插值出2001年0岁组的p_2为0.000293，而统计年鉴中2001年安徽省分年龄段抽样人口的合计人口为60662，因此，${}_nD_x$ = 0.000293×60662 = 18（人）。

④ 计算非采样年份预期寿命。根据公式（1）计算非采样年份完整年龄分组死亡率（${}_nm_x$），代入简略寿命表（表5），计算e_x（安徽省2001年的预期寿命为74.16岁）。

表5 2001年安徽省简略寿命表

Tab. 5 Abridged life table of Anhui province in 2001

年龄组(岁)	平均人口数	实际死亡数	死亡率	死亡概率	尚存人数	死亡人数	生存人年数	生存总人年数	预期寿命(岁)
X~	${}_nP_x$	${}_nD_x$	${}_nm_x$	${}_nq_x$	l_x	${}_nd_x$	${}_nL_x$	T_x	e_x
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	673	18	–	0.026419	100000	2642	97754	7416350	74.16
1~	2526	3	0.001201	0.004792	97358	466	388499	7318595	75.17
5~	4866	2	0.000486	0.002428	96892	235	483870	6930096	71.52
10~	6655	2	0.000319	0.001594	96656	154	482897	6446226	66.69
15~	4800	3	0.000581	0.002902	96502	280	481811	5963329	61.79
…	…	…	…	…	…	…	…	…	…
90~	32	6	0.180710	0.622377	15491	9641	53352	83606	5.40
95~	9	2	0.193354	1.000000	5850	5850	30254	30254	5.17

2.2 精度验证

为计算省级预期寿命，拉格朗日插值总共可分为4种方式：年龄分组人口比例线性插值（2个比例值作为插值节点）、预期寿命直接线性插值（2个预期寿命值作为插值节点）、年龄分组人口比例抛物线插值（3个比例值作为插值节点）、预期寿命直接抛物线插值（3个预期寿命值作为插值节点）。为确定最优的插值方式，且2015年由于数据暂时未能获取而无法计算预期寿命，从而未将其作为插值节点。本文以安徽、河南、湖北、湖南、江西、山西六省为例，以2000年简略寿命表法计算所得预期寿命为标准（记为s），分别进行绝对精度对比试验。

（1）年龄分组人口比例线性插值与预期寿命直接线性插值分别选取1995年、2005年年龄分组人口比例作为线性插值节点，按照2.1小节中非采样年份省级预期寿命计算思路来计算2000年的预期寿命，记为y_1；选取1995年、2005年简略寿命表法计算所得预期寿命值作为线性插值节点，直接插值出2000年的预期寿命值，记为t_1。通过比较$s-y_1$与$s-t_1$的大小，从而确定更优算法。

（2）预期寿命直接抛物线插值与年龄分组人口比例抛物线插值。分别选取1995年、2005年、2010年年龄分组人口比例作为抛物线插值节点，同样按照上述非采样年份省级预期寿命计算思路来计算2000年的预期寿命，记为y_2；选取1995年、2005年简略寿命表计算所得预期寿命与2010年简略寿命表法计算所得的预期寿命值作为抛物线插值节点，

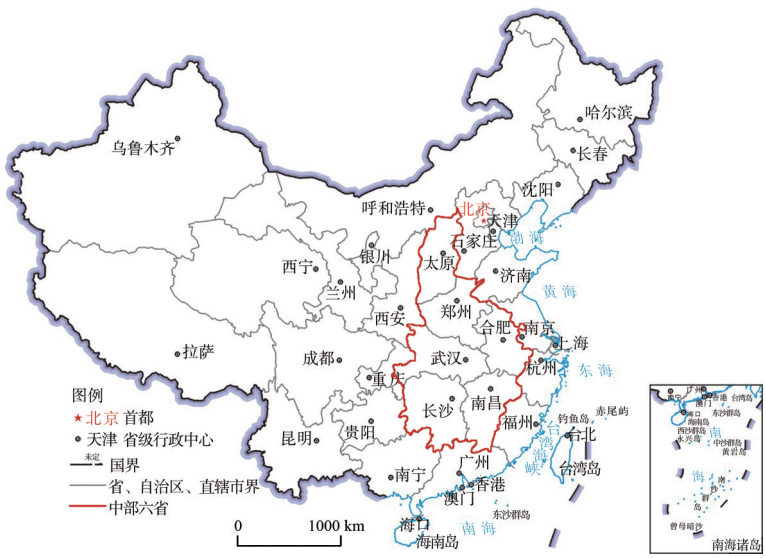
直接插值出2000年的预期寿命值，记为t_2。通过比较$s-y_2$与$s-t_2$的大小，从而确定更优算法。

（3）各种算法误差累计比较。将各种算法所计算得到的值与2000年简略寿命表计算所得预期寿命值之差进行累加，比较误差的累加值，从而确定更优算法。

3　算法验证

3.1 研究区及数据概况

本文选择中国中部六省（山西、河南、安徽、湖北、江西、湖南）为实验区域进行算法精度验证，其位于中国大陆中部腹地，承东启西，联接南北（图3），人口密集，总面积约为102.8万km^2，占全国的10.7%。



注：基于自然资源部标准地图服务网站审图号为GS(2019)1823号的标准地图制作，底图无修改。

图3　中国中部六省位置示意图

Fig. 3　Location of six Central China provinces

由2010年全国人口普查数据（表6）可知，中部六省的总人口数为35672万，约占全国人口数的26.76%，所占比例较大，其人口普查的平均预期寿命与中国全国的平均预期寿命在同一可比水平上。

采样年份的人口普查数据（2000年、2010年，但不包括港澳台）和1%人口抽样调查数据（2005年）下载自中华人民共和国国家统计局官网；非采样年份的抽样人口数据（与简略寿命表年龄分组不一致）源自1998—2014年中国

表6　2010年六省及全国普查人口数与各省人口数占比

Tab. 6　Census population of China and six Central China provinces with the proportion of each province in 2010

省份	2010年普查人口数(万人)	各省人口占全国人口比例
河南	9402	0.07
湖南	6570	0.05
安徽	5950	0.04
湖北	5723	0.04
江西	4456	0.03
山西	3571	0.03
中部六省	35672	0.27
全国	133281	1.00

知网数据库的各省统计年鉴或者《中国人口统计年鉴》；六省行政边界源自国家基础地理信息系统1∶400万数据库。

为测试中部六省预期寿命与全国之间的空间关系，并验证是否能在全国具有代表性，本实验对其进行了空间分析。本文选取人口普查年份2000年和2010年进行了Local Moran's *I*聚类。检验结果如图4和图5所示。两年结果均表明部分西部地区和东部沿海地区有特殊值聚类现象，西部为低低型聚类，东部为高高型聚类，故没有表现出空间分异性。而中部六省的预期寿命没有通过显著性检验，即没有显著的空间集聚性，故在一定程度可以代表中国大部分省份的预期寿命水平。

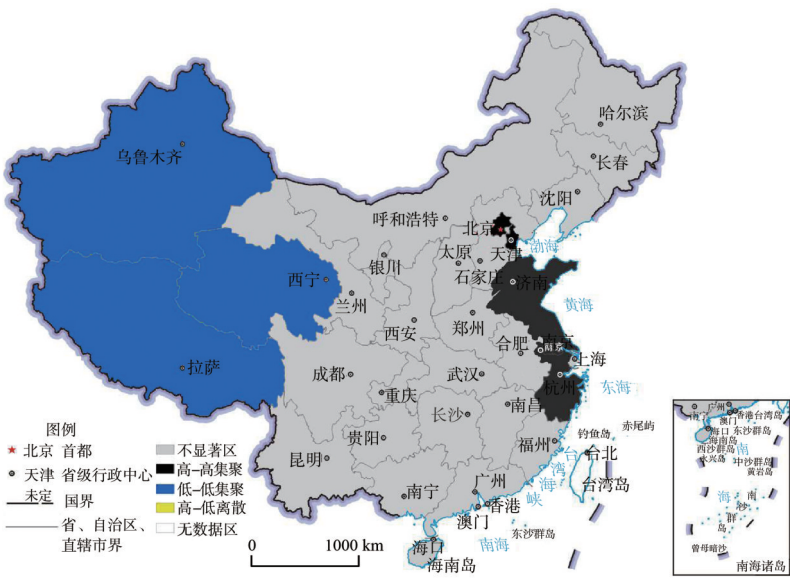


图4 2000年Local Moran's *I*聚类和异常值分析

Fig. 4 Analysis of clustering and outlier based on Local Moran's *I* in 2000

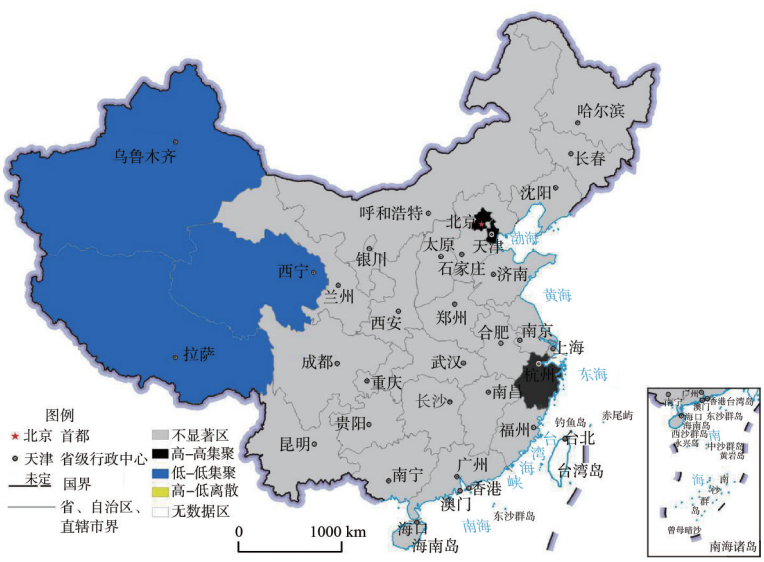


图5 2010年Local Moran's *I* 聚类和异常值分析

Fig. 5 Analysis of clustering and outlier based on Local Moran's *I* in 2010

为进一步测试预期寿命的空间分异性问题，本文采用地理探测器[20]q值进行检验。2000年和2010年中部六省的预期寿命具有非常大的空间分异性，分别为：q_{2000} = 0.984，q_{2010} = 0.973，与前文Local Moran's I聚类在中部六省的空间聚集性不显著的结果是一致的，正是因为空间分异，所以空间不聚集。

3.2 算法精度比较

验证实验包括4部分，分别比较年龄分组人口比例线性插值与预期寿命直接线性插值；预期寿命直接抛物线插值与年龄分组人口比例抛物线插值；预期寿命直接线性插值与预期寿命直接抛物线插值；年龄分组人口比例线性插值与年龄分组人口比例抛物线插值。并在最后对实验结果进行汇总，以此来进行绝对精度验证。

实验一：以2000年简略寿命表计算所得预期寿命值作为参考值，基于年龄分组人口比例线性插值与预期寿命直接线性插值的结果如表7所示。

表7表明，对于2000年省级预期寿命，以简略寿命表计算预期寿命作为检验标准，选取1995年、2005年简略寿命表计算所得预期寿命值作为线性插值节点，湖北省和湖南省直接插值出2000年的预期寿命值误差（$s-t_1$）较小；河南、江西、山西3省年龄分组人口比例线性插值所得误差（$s-y_1$）较小。简略寿命表计算所得预期寿命直接线性插值的误差绝对值累加为4.74，年龄分组人口比例线性插值的误差绝对值累加为3.65，可以看出线性比例插值结果相较于预期寿命直接插值更为精确。

实验二：以2000年简略寿命表计算所得预期寿命值作为参考值，基于年龄分组人口比例抛物线插值与预期寿命直接抛物线插值的结果如表8所示，因安徽、湖北、江西人口数据缺失，六省共有3组（河南、湖南、山西）不同方法下获取的预期寿命值与参考值的对比结果。

表8表明，对于2000年省级预期寿命，年龄分组人口比例抛物线插值（$s-y_2$）的3省误差绝对值累计为2.91明显小于预期寿命直接抛物线插值（$s-t_2$）的3省误差绝对值累计

表7　中部六省线性插值方式下获取的2000年预期寿命值比较(岁)

Tab. 7　Comparison of life expectancy for six Central China provinces obtained by linear interpolation in 2000 (years)

省份	年龄分组人口比例线性插值(y_1)	简略寿命表计算所得预期寿命直接线性插值(t_1)	简略寿命表计算所得2000年预期寿命(s)	$s-y_1$	$s-t_1$
安徽	–	–	73.55	–	–
河南	74.45	74.61	73.84	0.61	0.77
湖北	72.44	73.62	73.45	1	0.18
湖南	71.52	71.69	73.06	1.54	1.37
江西	71.34	72.85	70.90	0.43	1.95
山西	72.59	73.13	72.67	0.07	0.47

表8　中部六省抛物线插值方式下获取的2000年预期寿命值比较(岁)

Tab. 8　Comparison of life expectancy for six Central China provinces gained by parabolic interpolation in 2000 (years)

省份	年龄分组人口比例抛物线插值(y_2)	预期寿命直接抛物线插值(t_2)	简略寿命表计算所得2000年预期寿命(s)	$s-y_2$	$s-t_2$
安徽	–	–	73.55	–	–
河南	74.98	76.78	73.84	−1.14	−2.94
湖北	–	76.40	73.45	–	−2.95
湖南	71.79	73.93	73.06	1.27	−0.87
江西	–	73.90	70.90	–	−3
山西	72.17	73.69	72.67	0.5	−1.02

为4.83，说明省级预期寿命直接抛物线插值的结果误差会更大，因此，年龄分组人口比例抛物线插值的结果整体要优于预期寿命直接抛物线插值的结果。

综上实验对比，可以发现对于2000年省级预期寿命，① 除安徽省数据缺失外其他5省预期寿命直接抛物线插值误差（$s-t_2$）的绝对值明显整体大于预期寿命直接线性插值误差（$s-t_1$）的绝对值：（$s-t_1$）的六省误差绝对值累计为4.74，（$s-t_2$）的3省误差绝对值累计为4.83。因此，省级预期寿命直接抛物线插值的结果误差会更大，主要原因可能为线性直接插值选用1995年和2005年简略寿命表计算所得的预期寿命，而预期寿命直接抛物线插值则是参考了1995年、2005年与2010年的预期寿命，预期寿命直接抛物线插值所得到的结果偏高。② 因安徽、湖北、江西3省数据缺失无法计算，对河南、湖南、山西3省年龄分组人口比例线性插值误差（$s-y_1$）与年龄分组人口比例抛物线插值误差（$s-y_2$）的比较来看，（$s-y_2$）的绝对值总和为2.91大于（$s-y_1$）的总和为2.22，说明年龄分组人口比例抛物线插值的误差相较于年龄分组人口比例线性插值更大。

以2000年简略寿命表计算所得的预期寿命值作为真值，选取六省中河南、湖南、山西3组各种方法下获取预期寿命值与真值进行比较（表9）。由表9可以看出，（$s-y_1$）的绝对值求和最小，即年龄分组人口比例线性插值所得到的结果最接近真值。

表9 2000年中部六省各种方法下获取的预期寿命值与简略寿命表法计算所得预期寿命差值比较(岁)

Tab. 9 Comparison of life expectancy values obtained by various methods and calculated by abridged life table for six Central China provinces in 2000 (years)

省份	真值与年龄分组人口比例线性插值结果之差($s-y_1$)	真值与预期寿命直接线性插值结果之差($s-t_1$)	真值与年龄分组人口比例抛物线插值结果之差($s-y_2$)	真值与预期寿命直接抛物线插值结果之差($s-t_2$)
河南	0.61	0.77	−1.14	−2.94
湖南	1.54	1.37	1.27	−0.87
山西	0.07	0.47	0.50	−1.02
绝对值求和	2.22	2.61	2.91	4.83

4 结论与展望

预期寿命作为反映人口健康状况的综合性指标是医学地理学的重要研究内容[21]。本文针对非采样年份省级预期寿命计算的完整年龄分组死亡率缺失问题（统计年鉴中人口数据年龄分段与简略寿命表年龄分组不一致且缺乏相应年龄分段的死亡人口致使年龄组死亡率无法计算），提出了基于拉格朗日插值的省级预期寿命时序加密算法。

此算法将可获取的人口普查和1%人口抽样调查数据与统计年鉴人口数据相结合，在最大限度利用现有数据的基础上，依据时间序列的邻近效应以及时空变量的离散函数逼近原理，选取拉格朗日插值对缺失的年龄分组人口（含死亡人口）进行估算，并选择具有空间代表性的中部六省，通过两组绝对精度对比实验说明年龄分组人口比例线性插值计算的精度明显高于人口比例抛物线插值和直接插值算法的精度，因此可作为非采样年份省级预期寿命估算的推荐算法，该算法使得省级预期寿命的预测以及趋势分析成为可能。

本文提出的省级预期寿命时序加密算法可有效地估算非采样年份省级预期寿命并较好地反映其发展趋势。然而，由于社会统计数据本身存在误差，而且拉格朗日插值在处理数据时也会存在误差和不确定性，致使估算出的预期寿命值相比真实值不可避免地存

在小的偏差。另一方面，由于本文采用的是最经典的预期寿命算法，其本身也存在一定的局限性。因此，今后一方面应着眼于提高社会统计数据的质量，以期借鉴更先进的算法改进估算方法；另一方面，为更好地验证并提高模型的精确性，可搜寻更多相关数据并将研究区域推广至全国。

参考文献(References)

[1] Tan Jian'an, Li Ribang, Zhu Wenyu. The progress in medical geography of China and its prospect. Acta Geographica Sinica, 1990, 45(2): 187-201. [谭见安, 李日邦, 朱文郁. 我国医学地理研究的主要进展和展望. 地理学报, 1990, 45(2): 187-201.]

[2] Lin Muxi, Huang Taiyan. Dictionary of National Economics. Beijing: Economic Science Press, 2014: 484. [林木西, 黄泰岩. 国民经济学辞典. 北京: 经济科学出版社, 2014: 484.]

[3] Namboodiri N K, Suchindran C M. Life Table Techniques and Their Applications. Orlando: Academic Press, 1987: 275.

[4] Zhou Shikai. Health Statistics. 3rd ed. Beijing: People's Medical Publishing House, 1987: 202. [周士楷. 卫生统计学(第三版). 北京: 人民卫生出版社, 1987: 202.]

[5] Ding Yuanlin, Gao Ge. Health Statistics: Case Edition. Beijing: Science Press, 2008: 364. [丁元林, 高歌. 卫生统计学: 案例版. 北京: 科学出版社, 2008: 364.]

[6] Samji H, Cescon A, Hogg R S, et al. Closing the gap: Increases in life expectancy among treated HIV-positive individuals in the United States and Canada. Plos One, 2013, 8(12): e81355.

[7] Shu Xingyu, Wen Yong, Zong Zhanhong. Indirect estimation and evaluation of China's average life expectancy. Population Journal, 2014, 36(5): 18-24. [舒星宇, 温勇, 宗占红. 对我国人口平均预期寿命的间接估算及评价: 基于第六次全国人口普查数据. 人口学刊, 2014, 36(5): 18-24.]

[8] Zhang Wenjuan, Wei Meng. The evaluation of the mortality and life expectancy of Chinese population. Population Journal, 2016, 38(3): 18-28. [张文娟, 魏蒙. 中国人口的死亡水平及预期寿命评估: 基于第六次人口普查数据的分析. 人口学刊, 2016, 38(3): 18-28.]

[9] Yang Dongliang, Wang Xiaolu. Inter-provincial differences in life expectancy and spatial dependence characteristics of China's population. Social Science Front, 2016(4): 172-179. [杨东亮, 王晓璐. 中国人口预期寿命的省际差异与空间相依特征. 社会科学战线, 2016(4): 172-179.]

[10] Gong Shengsheng, Chu Huan, Zhang Tao. The spatial-temporal changes of population longevity indicators of Anhui Province during 1990-2010. Areal Research and Development, 2015, 34(4): 160-168. [龚胜生, 储环, 张涛. 1990—2010年安徽省人口长寿水平的时空变化. 地域研究与开发, 2015, 34(4): 160-168.]

[11] Gong Shengsheng. A preliminary study on the geographical distribution and environmental background of the longevity area in ancient China. Journal of Chinese Historical Geography, 1997(3): 227-251. [龚胜生. 中国古代长寿点区的地理分布及其环境背景的初步研究. 中国历史地理论丛, 1997(3): 227-251.]

[12] Gilligan A M, Skrepnek G H. Determinants of life expectancy in the Eastern Mediterranean Region. Health Policy Planning, 2015, 30(5): 624.

[13] Olshansky S J, Antonucci T, Berkman L, et al. Differences in life expectancy due to race and educational differences are widening, and many may not catch up. Health Affairs, 2012, 31(8): 1803.

[14] Austin K F, Mckinney L A. Disease, war, hunger, and deprivation: A cross-national investigation of the determinants of life expectancy in less-developed and Sub-Saharan African nations. Sociological Perspectives, 2012, 55(3): 421-447.

[15] Chiang C L. Life Table and Mortality Analysis. Life Table & Mortality Analysis, 1978: 91-113.

[16] Kang Xiaoping. Practical Health Statistics. Beijing: Peking University Medical Press, 2010: 276. [康晓平. 实用卫生统计学. 北京: 北京大学医学出版社, 2010: 276.]

[17] Zhou Renyu. Traditional Chinese Medicine (TCM) Statistics. New Century 2nd ed. Beijing: China Press of Traditional Chinese Medicine, 2008: 239. [周仁郁. 中医药统计学(新世纪第2版). 北京: 中国中医药出版社, 2008: 239.]

[18] The Panel. A Book for Qualification Examination of National Practicing Physicians: Examination-oriented Guidance for Public Health Practitioner in 2014. Beijing: China Concorde Medical University Press, 2014: 1291. [本书专家组. 2014国家执业医师资格考试用书: 公卫医师应试指导. 北京: 中国协和医科大学出版社, 2014: 1291.]

[19] Milne W E. Numerical Calculus. Princeton: Princeton University Press, 2015: 126-135.

[20] Wang Jinfeng, Xu Chengdong. Geodetector: Principle and prospective. Acta Geographica Sinica, 2017, 72(1): 116-134. [王劲峰, 徐成东. 地理探测器: 原理与展望. 地理学报, 2017, 72(1): 116-134.]

[21] Tan Jianan. Health, environment, development: The theme of contemporary medical geography. Acta Geographica Sinica, 1994, 61(Suppl.1): 710-718. [谭见安. 健康、环境、发展: 当代医学地理的主题. 地理学报, 1994, 61(Suppl.1): 710-718.]

Time-series estimation of provincial life expectancy in China: A case study of six provinces in central China

LI Chang, WANG Anli, GONG Shengsheng, SUN Youning
(Key Laboratory for Geographical Process Analysis & Simulation, Hubei Province, and College of Urban and Environmental Science, Central China Normal University, Wuhan 430079, China)

Abstract: Age-specific mortality rate is a key parameter to estimate life expectancy based on age-group population. However, it is impossible to estimate life expectancy in non-sampling years (i.e., without census or 1% population sampling survey) due to the loss of age-specific mortality rate in statistical yearbooks. To estimate time-series life expectancy at China's provincial level in the non-sampling years, this paper firstly proposes a time-series estimation algorithm based on Lagrange interpolation by combining census data with population data from statistical yearbooks. We selected six provinces in central China as study areas and estimated provincial time-series life expectancy in non-sampling years by four algorithms, i.e., linear interpolation and quadratic polynomial interpolation in direct and indirect ways. And the absolute accuracy of estimating time-series life expectancy indicates that the accuracy of linear interpolation for proportions of population by age group (i.e. indirect method) is significantly higher than that of quadratic polynomial interpolation (i.e. indirect method) and time-series interpolation of life expectancy (i.e. direct method) based on two methods, which is proposed as a recommendation algorithm. This study provides a new and feasible way to acquire the provincial time-series life expectancy in non-sampling years, which lays a foundation for the more accurate trend analysis of life expectancy in China.

Keywords: provincial life expectancy; time-series estimation; non-sampling year; six provinces of central China; Lagrange interpolation; linear interpolation; quadratic polynomial interpolation