

一种新的预测人群死亡率方法的应用

发布日期: 2002-01-04

侯长荣/余松林/陈心广

传统的人群死亡率预测方法需假定人群的期望寿命有一上限,或采用控制人群期望寿命增长速度的方式,才能使其预测结果趋于合理。例如:使用最佳寿命表法或模型寿命表等方法[1,2]。为解决传统方法的不足之处,美国人口学Ronald D. Lee和Lawrence R. Carter在1992年提出用一种新方法来预测人群的死亡率[3]。

新方法不需考虑医疗卫生或社会因素对死亡率变化的影响。它的优点是相对方便简单,不需要控制很多参数,它融合了统计学的时间序列和人口学模型的方法,根据死亡率的历史变动情况,预测其未来的趋势。它允许人群年龄别死亡率无人为限制地按指数下降,也不需对期望寿命的减速增长施任何假设条件。

一、方法介绍

新方法引进了三个参数,参数 $k(t)$ 死亡率水平指数--反映历年死亡率相对强度,参数 $a[x]$ --反映各年龄别死亡率对数变化的基数,参数 $b[x]$ --反映各年龄别死亡率对数变化趋势。通过对未来年份死亡率水平指数 $k(t)$ 的预测,预测出未来年份的死亡率。

(一) 模型结构

预测死亡率的模型结构为:

$$M[x](t) = \exp[a[x] + b[x] \times k(t) + \varepsilon[x](t)] \quad (1)$$

将(1)式两边取对数得:

$$\ln[M[x](t)] = a[x] + b[x] \times k(t) + \varepsilon[x](t) \quad (2)$$

式中 $M[x](t)$ 表示第 t 年份, x 年龄组的死亡率。 $k(t)$ 表示第 t 年份的死亡率水平指数,它表示不同时间人群死亡率水平的相对高低。 $\varepsilon[x](t)$ 是误差项,反映的实际数据为不能被模型解释的那部分信息。 $a[x]$ 描述的是各年龄别死亡率对数变化的基数;根据 $k(t)$ 的变化, $b[x]$ 侧面告诉我们各年龄别死亡率对数下降或上升的速率。理论上讲, $b[x]$ 在某些年龄别可能出现负值,即表明该年龄别死亡率对数变化和其他年龄别死亡率对数变化是反方向的。

(二) 参数的估计方法

(1) $a[x]$ 的估计将(2)式写成方程为:

将方程(3)两边对 t 求和得:

式中 i 是死亡率实际观察数据的开始年, j 是死亡率实际观察的结束年。在方程(3)中假定和分别符合的条件分布。对于不同年份 t ,同一年龄组的和是相同的。即可将(4)式转化为:

又因,那么

即 x 年龄组各个时间的死亡率对数的均数就是 x 年龄组的 $a[x]$ 的估计值。

(2) $k(t)$ 的估计

将方程(3)移项后两边求和得:

当t不变时，也就是说对于同一年份，不同年龄组 值是不变的。
又因 所以有：

即第t年份的死亡率水平指数k(t)的估计值等于该年份各年龄组死亡率对数与其相应年龄组的差值之和。

(3)b[, x]的估计

根据以上 和 的估计值，用最小二乘法求解方程

以 为自变量，以每一年龄别各个年份的 为一组反映变量，建立一个不含常数项的线性回归方程[5]，估计出每个年龄组的b[, x]。但因该方法的 $\varepsilon[, x](t)$ 残差项较大，不够理想。为了减小残差，将(9)式改写为(10)式后，改用非线性方程的(Marquardt)解方程(10)。对于某一年龄组 $\exp(a[, x])$ 是一常数项，若干个M和若干个k进行指数曲线拟合，求出该年龄组的 。

(三) 计算未来年份死亡率水平指数k(t)的预测值

k(t)的预测值是采用自回归累计移动平均过程[3, 4](Autoregressive Intergrated Moving Average Process)ARIMA(p, d, q)时间序列模型预测。p是模型的自回归阶数，d是模型的差分阶数，q是模型的移动平均阶数。本研究预测k(t)所选择的参数均为p=0, d=1, q=1，即ARIMA(0, 1, 1)一阶差分移动平均模型。

由于k(t)是线性非平稳过程，因此需转变为线性平稳过程 $\Delta k(t)$ 后来处理。

一阶差分移动平均模型的一般表达式为：

$$\Delta k(t) = \mu - \theta[, 1] \times \varepsilon(t-1) + \varepsilon(t) \quad (12)$$

式中 μ 是过程 $\Delta k(t)$ 的算术平均数， $\mu = \sum [\Delta k(t)] / (n-1)$ ，n是k(t)的观察个数； $\theta[, 1]$ 是一阶移动平均常数 $(-1 < \theta[, 1] < 1 =)$ ； $\varepsilon(t)$ 是第t年份的随机误差，

是第t-1年份的随机误差。

对模型配合的适度检验采用波特曼图(Portmeanteau test)卡方检验，计算公式如下：

式中 $r[, s]$ 是滞后时期数为s的残差的自相关函数估计值； ι 为包括在本研究中的自相关函数的个数($\iota=3$)； ρ 为模型中的参数个数($\rho=2$)。

$X[2, (\iota-\rho)]$ 服从自由度为 $\iota-\rho$ 的 $\chi[2]$ 分布。

时间序列k(t)的变化趋势的预测是采用EXECUSTAT软件包[4]。

(四) 计算未来年份的各年龄别预期死亡率M[, x](t)

将以上预测的k(t)值以及参数a[, x]、b[, x]的估计值 、 代入下式：

即得到未来几年的各年龄别死亡率。

例如1995年0岁组人群的预期死亡率的估计为：

(*以1988年作为t=0年，1989年t=1，依此类推，故1995年t=7。)

二、 应用实例

(一) 资料来源

表1是1988~1994年我国农村死亡登记点男性人口死亡率的资料，由卫生部统计信息中心提供，其中1991年的资料因故缺。

表1 全国死亡登记点1988~1994年农村男性人口各年龄组死亡率

年龄	1988	1989	1990	1992	1993	1994
0~	0.02453	0.02427	0.02452	0.02289	0.02217	0.02138
1~	0.00271	0.00249	0.00220	0.00210	0.00180	0.00158
5~	0.00091	0.00084	0.00085	0.00076	0.00078	0.00086
10~	0.00051	0.00045	0.00044	0.00043	0.00047	0.00052
15~	0.00091	0.00078	0.00070	0.00065	0.00060	0.00056
20~	0.00157	0.00148	0.00121	0.00129	0.00124	0.00125
25~	0.00133	0.00133	0.00129	0.00143	0.00151	0.00149
30~	0.00181	0.00166	0.00152	0.00142	0.00163	0.00181
35~	0.00251	0.00261	0.00230	0.00226	0.00227	0.00223
40~	0.00357	0.00344	0.00327	0.00333	0.00360	0.00367
45~	0.00470	0.00469	0.00468	0.00479	0.00509	0.00526
50~	0.00776	0.00756	0.00753	0.00735	0.00782	0.00773
55~	0.01302	0.01337	0.01210	0.01179	0.01170	0.01165
60~	0.02433	0.02366	0.02267	0.02233	0.02228	0.02226
65~	0.03502	0.03633	0.03388	0.03312	0.03291	0.03307
70~	0.06090	0.06420	0.05917	0.06300	0.06229	0.06247
75~	0.08411	0.09094	0.08637	0.08745	0.08214	0.08009
80~	0.14217	0.15339	0.14788	0.14552	0.13904	0.13913
85~	0.23995	0.27467	0.23165	0.23475	0.17839	0.18684
合计	0.00691	0.00680	0.00688	0.00686	0.00687	0.00691

(二) 模型配合结果

1. $a[, x]$ 估计值的计算结果

根据(6)式0~岁组的 $\ln q_0$ 等于1988年至1994年该人群0~岁组死亡率自然对数的算术平均数。

依此类推, 计算出 $a[, x]$ 、 $b[, x]$ 估计值见表2。

2. $k(t)$ 估计值的计算结果

根据表2中 $a[, x]$ 的估计值, 套用(8)式计算出 $k(t)$, 例如:

1990年农村男性人群的 $k(1990)$ 就等于1990年农村男性各个年龄组的死亡率的自然对数与其年龄组的 $a[, x]$ 之差值的之和。

依此类推, 计算出的 $k(t)$ 的估计值见表3。

表2 $a[, x]$ 、 $b[, x]$ 的估计结果

年龄	$a[, x]$	$b[, x]$	年龄	$a[, x]$	$b[, x]$
0~	-3.7454	0.0014	45~	-5.3291	-0.0373
1~	-6.1476	0.1955	50~	-4.8815	0.0049
5~	-7.0984	0.0602	55~	-4.3967	0.0649
10~	-7.6563	0.0175	60~	-3.7807	0.0477
15~	-7.2446	0.1788	65~	-3.3874	0.0464
20~	-6.6157	0.1260	70~	-2.7811	0.0022
25~	-7.0984	-0.0566	75~	-2.4673	0.0246
30~	-6.4312	0.0625	80~	-1.9378	0.0213
35~	-6.0529	0.0757	85~	-1.5064	0.1369
40~	-5.6721	0.0068			

表3 $k(t)$ 的估计值结果

$k(t)$ 的估计值满足 $k(t) > 0$ 条件。 $k(t)$ 是第 t 年份各年龄组死亡率的综合反映, $k(t)$ 值的大小可

反应该年份死亡水平的高低，k(t)值越小死亡水平就越低。

3. b[, x]的估计值计算结果

根据a[, x]、k(t)的估计值，拟合指数曲线方程(10)，例如：农村男性0~岁组的 就是以1988年至1994年的农村男性0~岁组的死亡率与农村男性0~岁组的exp 之比做为反应变量，以1988年至1994年的农村男性的 做自变量，做不含常数项的指数曲线拟合得到的回归系数。求出的 列于表2。表中 的值与直线回归方法求得的b[, x]估计值的相对差值小于10%，方程显著性检验的P值均小于0.05。b[, x]的大小是反映年龄组死亡率变化的快慢，b[, x]越大说明该年龄死亡率变化速度越快，反之就越慢。当b[, x]是负数时，表示该年龄组死亡率越来越小。

4. 未来年份k(t)值的预测结果及检验

利用ARIMA(0, 1, 1)模型，即(12)式的一阶差分移动平均模型，对k(t)值的预测结果见表4。用预测方程表示为：

$$\Delta k(t) = -0.2853 - (-1.8759) \times \varepsilon(t-1) \quad (15)$$

表4 未年来份k(t)值的预测结果

年份	k(t)	年份	k(t)
1995	-1.1382	2000	-2.5648
1996	-1.4236	2001	-2.8501
1997	-1.7089	2002	-3.1354
1998	-1.9942	2003	-3.4208
1999	-2.2795	2004	-3.7061

对k(t)的预测结果进行波特曼图X[2]检验（波特曼图X[2]值是对残差序列中的前3个自相关函数的波特曼图X[2]值）。

P值越大预测效果越好。

5. 各年龄别死亡率预测结果见表5。

表5 1995~2000年各年龄别死亡率的预测结果

年龄	1995	1996	1997	1998	1999	2000
0~	0.023589	0.023579	0.023570	0.023560	0.023551	0.023541
1~	0.001712	0.001619	0.001531	0.001448	0.001370	0.001295
5~	0.000772	0.000759	0.000746	0.000733	0.000720	0.000708
10~	0.000464	0.000461	0.000459	0.000457	0.000455	0.000452
15~	0.000583	0.000554	0.000526	0.000500	0.000475	0.000451
20~	0.001160	0.001119	0.001080	0.001042	0.001005	0.000969
25~	0.001490	0.001515	0.001539	0.001564	0.001590	0.001616
30~	0.001500	0.001473	0.001447	0.001422	0.001397	0.001372
35~	0.002157	0.002111	0.002066	0.002022	0.001978	0.001936
40~	0.003414	0.003407	0.003401	0.003394	0.003388	0.003381
45~	0.005059	0.005113	0.005168	0.005223	0.005279	0.005335
50~	0.007543	0.007533	0.007522	0.007512	0.007501	0.007491
55~	0.011441	0.011231	0.011025	0.010823	0.010624	0.010429
60~	0.021602	0.021309	0.021021	0.020737	0.020457	0.020180
65~	0.032058	0.031636	0.031220	0.030810	0.030404	0.030005
70~	0.061815	0.061777	0.061738	0.061699	0.061660	0.061622
75~	0.082472	0.081895	0.081322	0.080753	0.080188	0.079628
80~	0.140571	0.139719	0.138872	0.138031	0.137195	0.136364
85~	0.189717	0.182448	0.175459	0.168738	0.162275	0.156059

6. 预测结果评价

根据表6资料进行预测结果评价。

表6 1997年全国死亡登记点农村男性人口死亡率

年龄组	死亡率	年龄组	死亡率	年龄组	死亡率
0~	0.020210	30~	0.002041	65~	0.031330
1~	0.001246	35~	0.002019	70~	0.057675

5~	0.000752	49~	0.003702	75~	0.077101
10~	0.000510	45~	0.005272	80~	0.121432
15~	0.000520	50~	0.007442	85~	0.150195
20~	0.001000	55~	0.010733		
25~	0.001464	60~	0.020176		

剩余平方和 $[5]RSS=(0.020210-0.023570)[2]+(0.001246-0.001531)[2]+.....+(0.150195-0.175459)[2]=0.000989$

残差分析，显示预测结果比较满意。

(摘自《中国卫生统计》2000年第10期)

[回到页首](#)

[关闭该页](#)

[单位简介](#) | [合作机会](#) | [联系我们](#) | [E-Mail密码修改](#)

为了本系统能够更好的为您服务，建议使用IE4.0或以上版本浏览器。Best View: 800*600

版权所有: 上海市人口和计划生育委员会信息中心

网站制作: 上海棕榈电脑系统有限公司

Copyright?2000 Shanghai Municipal Population and Family Planning Commission. All Rights Reserved.