

中国第四次人口普查资料中死亡指标方法探讨

黄荣清

1990年中国成功地进行了第四次人口普查。这次普查数据,对于了解中国的国情、制订国民经济和社会发展规划具有重要的作用。人口普查之后,继手工汇总数据的发表,最近人口普查办又公布了10%电子计算机汇总资料,100%电子计算机汇总结果也将发表。所以我们的工作,一是如何充分利用普查结果,二是如何对普查资料进行深开发和再加工。本文拟从人口死亡研究的角度,探讨如何利用普查资料确定人口死亡指标问题。

一、关于四普的人口死亡资料

关于人口死亡,第四次人口普查和第三次人口普查(以下简称“四普”和“三普”)一样,在调查时要求报告者申报从调查时点起过去一年半(即从1989年1月1日至1990年6月30日)内本户的死亡人数。但在汇总时,两次普查采取的方式却不一样。三普时,只汇总了普查前一年日历年度(1981年)的死亡人数,而四普汇总了普查前一年半的死亡人数,并且把它划分为三个时期,即1989年上半年、1989年下半年和1990年上半年。由于有了这样的区分,我们不但可以用三普时使用过的方法确定1989年人口死亡指标,并还有可能对人口死亡分布的特点有进一步的了解。同时,通过对三个不同时期人口死亡数的比较,可以估计普查时人口死亡报告的质量。但另一方面,这也带来与三普人口死亡指标不同的计量方法问题。

二、人口死亡分析中普查时的人口数意义和其所对应的死亡人口数

在对一个人口群体作死亡分析时,笔者认为首先该明确以下三点:(1)规定分析的时间范围。由于人口死亡的速度随时间发生变化,通常所说的死亡指标发生在一定的时间范围内才有意义。所以从分析的一开始,就必须明确是在哪个时间区间内来研究人口死亡的,以此来规定期初时间和期末时间。习惯上,我们常取一年作为时间的区间长度。例如根据四普资料,可规定时间范围是在1989年,则考察的期初时间为1989年初,考察的期末时间为1989年末;如果我们要研究普查前一年的人口死亡情况,则考察的期初时间为1989年7月1日,考察的期末时间为1990年6月30日。(2)在规定的时间内,各年龄的死亡人口数。由于四普资料汇总时对人口死亡资料作了划分,所以如果我们在分析1989年的人口死亡时,只要把1989年上半年和1989年下半年的同年龄人口死亡数加总,即可得到1989年各年龄的死亡人口数。类似地,只要把1989年下半年和1990年上半年同年龄死亡人口数加总,即可得到普查前一年的死亡人口数。(3)在规定的时间内,普查时人口数的意义。例如若研究普查前一年的人口死亡,则普查时间为期末,普查时各年龄人口数可看作期末人口数。

这里要特别讨论一下当把死亡分析的时间范围规定在1989年时,1990年普查时人口数的意义。若把时间范围规定在1990年,通常的作法是把普查人口数当作当年的平均人口。笔者认为这只能作为一种处理方法;但不是最好的处理方法,更不是唯一的处理方法。

把普查时的人口数(即年中人口数)当作平均人口,常用于求平均死亡率。平均死亡率的计算公式为:

$$m_x = D_x \sqrt{P_x}$$

这里 m_x 为年龄 x 的平均死亡率, D_x 为年龄 x 的死亡人口数, \bar{P}_x 为年龄 x 的平均人口。它的意

义是指一个人口群体在 x 年龄内的平均死亡比率。把年中人口作为平均人口,仔细推敲存在着不合理的地方。例如,在求1990年的0岁平均死亡率时,分子为死于1990年的0岁人口,则死于1990年上半年0岁的人是在1989年和1990年上半年出生的;死于1990年下半年0岁的人口是在1989年下半年和1990年出生的,而1990年年中人口为出生于1989年下半年和1990年上半年的人。即是说,分子(死亡人口数)和分母(年中人口)并不完全是同一个人口群体。严格说来,死亡率是一比率概念,分子和分母是部分与全体的关系,若分子和分母不属于同一集合,其比率也就失去其意义了。在确定平均死亡率时,最重要的是使分子和分母同属一个人口群体。实际上,年平均人口有许多种估计方法,用年中人口作为年平均人口,仅仅是估计方法之一。这种估计方法,只有当人口年龄构成变动较小才合适,否则就会带来较大误差。笔者在以下的处理中,常把普查时各年龄人口直接看作这个年龄的年中人口数,它和生命表中尚存人口数 $l(x+0.5)$ 相对应,这样可能更合理些。

为了以下讨论的方便,先来看一下在普查时 x 年龄的人口在过去一年内相对应的死亡人口。我们借助于图来说明。

把水平线表示为时间轴(见图1),设与时间轴成 45° 的一条条平行线为出生于同一时刻的人口群体,如图1中 A_0A_0' 和 A_1A_1' 分别表示出生于1990年年中以前 $x+1$ 年和 x 年的人口群体在不同时间(年龄)的人口数(注意在直线上每一点代表的人口数都不同),它们与人口普查时的时点直线 T_0T_0' 相交于 $Q'Q_2$,则线段 $Q'Q_2$ 表示人口普查时 x 岁年龄的人口。类似地,直线 A_1A_1' 、 A_0A_0' 与人口普查前一年的时间线 T_1T_1' 相交于 S_1S' ,线段 S_1S' 表示人口普查前一年的 $x-1$ 岁人口,用平行四边形 $S_1S_1'Q_2Q'$ 表示与人口普查时 x 岁同时出生的人口群体在人口普查前一年内死去的人口;人口普查前一年内所有死于 $x-1$ 岁和 x 岁的人口用和 T_0T_0' 、 T_1T_1' 相交并与过年龄点 $x-1$ 和 $x+1$ 和年龄轴相垂直的矩形 $S_1S_2Q_2Q_1$ 来表示。其中 $S_1S_1'O'O$ 表示在1989年下半年死于 $x-1$ 岁的人数, $S'S_2O_2O'$ 表示在1989年下半年死于 x 岁半的人数, $OO'Q'Q_1$ 表示在1990年上半年死于 $x-1$ 岁的人数, $O'O_2Q_2Q'$ 表示1990年上半年死于 x 岁的人数。若死亡人口数分布是均匀的,我们可以用比较平行四边形 $S_1S_1'Q_2Q'$ 和矩形 $S_1S_2Q_2Q_1$ 的面积来估计与人口普查时 x 岁的人口相对应的死亡人口数,用 S 表示面积,于是有

$$S \square S_1S_1'Q_2Q' = \frac{3}{4} S \square S_1S_1'O'O + \frac{1}{4} S \square S'S_2O_2O' + \frac{1}{4} S \square OO'Q'Q_1 + \frac{3}{4} S \square O'O_2Q_2Q'.$$

换句话说,与人口普查时 x 岁的人口同一出生时刻的人口群体在人口普查前一年内死亡人口数为

$$D_x = \frac{3}{4} D(x-1, \text{II}) + \frac{1}{4} D(x, \text{II}) + \frac{1}{4} D(x-1, \text{III}) + \frac{3}{4} D(x, \text{III}).$$

其中II表示1989年下半年,III表示1990年上半年, $D(x, \text{II})$ 表示在1989年下半年 x 岁的死亡人

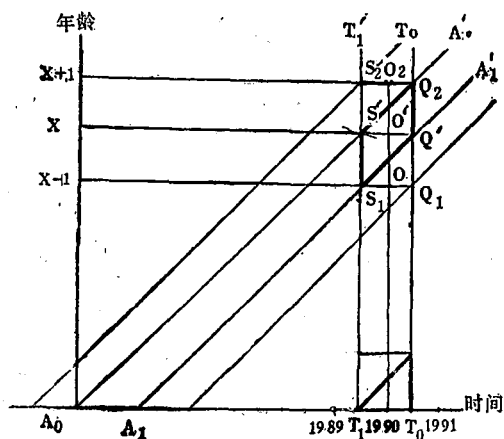


图1 普查时 x 岁年龄人口和其相对应的普查前一年内的死亡人口关系

数, 其余以此类推。

比较特殊的人口普查时0岁的人口在过去一年内相对应的死亡人口数为

$$D_0 = \frac{3}{4} D(0, \text{II}) + \frac{1}{4} D(0, \text{III}).$$

三、利用四普资料确定人口死亡指标的几种方法

在作人口死亡研究时, 关键是确定年龄别死亡指标。我们只要确定了一种指标, 那么无论是年龄别平均死亡率, 还是年龄别死亡概率等, 就可以用一种指标推算另一种指标, 并进而计算其它综合指标。

由于四普人口死亡资料的汇总特点, 以及对普查时人口的不同处理方法, 我们可以有以下几种不同的方法确定年龄别死亡资料。

(一) 计算1989年的年龄别死亡指标

1. 由普查资料估计1989年年中人口, 把它看作平均人口, 计算平均死亡率。

由上面的讨论可知, 普查时 x 岁年龄人口在过去一年内所对应的死亡人口为 $\frac{3}{4} D(x-1, \text{II}) + \frac{1}{4} D(x, \text{II}) + \frac{1}{4} D(x-1, \text{III}) + \frac{3}{4} D(x, \text{III})$, 所以普查前一年(即1989年年中)的 $(x-1)$ 岁人口为

$$P(x-1, 1989) = P(x, 1990) + \frac{3}{4} D(x-1, \text{II}) + \frac{1}{4} D(x, \text{II}) + \frac{1}{4} D(x-1, \text{III}) + \frac{3}{4} D(x, \text{III}), \quad (x=1, 2, \dots, w-2)$$

这里 w 表示普查时开区间年龄。由于

$$P(w-1, 1989) = P(w, 1990) + \frac{3}{4} D(w-1, \text{II}) + \frac{1}{4} D(w, \text{II}) + \frac{1}{4} D(w-1, \text{III}) + \frac{3}{4} D(w, \text{III})$$

$$P(w, 1989) = P(w+1, 1990) + \frac{3}{4} D(w, \text{II}) + \frac{1}{4} D(w+1, \text{II}) + \frac{1}{4} D(w, \text{III}) + \frac{3}{4} D(w+1, \text{III}).$$

.....

把以上式子等号左右分别相加

$$\begin{aligned} \sum_{x=w-1}^{\infty} P(x, 1989) &= \sum_{x=w}^{\infty} P(x, 1990) + \frac{3}{4} \sum_{x=w-1}^{\infty} D(x, \text{II}) + \frac{1}{4} \sum_{x=w}^{\infty} D(x, \text{II}) \\ &+ \frac{1}{4} \sum_{x=w-1}^{\infty} D(x, \text{III}) + \frac{3}{4} \sum_{x=w}^{\infty} D(x, \text{III}) = \sum_{x=w}^{\infty} P(x, 1990) + \frac{3}{4} D(x-1, \text{II}) \\ &+ \sum_{x=w}^{\infty} \left[\frac{3}{4} D(x, \text{II}) + \frac{1}{4} D(x, \text{II}) \right] + \frac{1}{4} D(x-1, \text{III}) + \sum_{x=w}^{\infty} \left[\frac{1}{4} D(x, \text{III}) + \frac{3}{4} D(x, \text{III}) \right] \\ &= \sum_{x=w}^{\infty} P(x, 1990) + \frac{3}{4} D(x-1, \text{II}) + \frac{1}{4} D(x-1, \text{III}) + \sum_{x=w}^{\infty} D(x, \text{II}) + \sum_{x=w}^{\infty} D(x, \text{III}). \end{aligned}$$

由于当 $x \rightarrow \infty$ 时 $P(x, 1990)$, $D(x, \text{II})$, $D(x, \text{III})$ 都趋于零, 所以

$$P^+(w-1, 1989) = P^+(w, 1990) + \frac{3}{4}D(w-1, II) + \frac{1}{4}D(w-1, III) \\ + D^+(w, II) + D^+(w, III).$$

+号表示年龄为开区间, 由上式可知, 若1990年年中人口的开区间年龄为W, 则对1989年年中人口, 我们可取的开区间年龄为W-1。

把1989年的年中人口看作1989年的年平均人口, 则1989年的年龄别平均死亡率为

$$m(x, 1989) = D(x, 1989) / P(x, 1989).$$

由平均死亡率计算年龄别生存概率可用如下公式

$$P_x = \begin{cases} \exp(-m_x) & (1) \text{在年龄区间两端} \\ 1 - \frac{m_x}{1+0.5m_x} = (1-0.5m_x)/(1+0.5m_x) & (2) \text{其它} \end{cases}$$

在年龄区间两端, 即婴幼儿期和高龄期, 由于死亡率变化较大, 死亡人口在前期较后期多, 所以生存概率用平均死亡率的指数函数来估计; 在其它年龄区间, 死亡率和死亡人口变化较小, 在假设人口呈线性变化时, 则生存概率可用(2)式来计算。

2. 把普查时的人口看作年中人口, 根据制作生命表的方法, 计算生存概率。

设1989年x岁的年中人口为 $P(x-0.5, 1989)$, 一年以后, 即普查时人口为 $P(x+0.5, 1990)$, $l(x)$ 表示生命表中x岁的尚存人口数, 则

$$\frac{P(x-0.5, 1989)}{P(x+0.5, 1990)} = \frac{l(x-0.5)}{l(x+0.5)} \quad (3)$$

$$P(x-1, 1989) \left(1 - \frac{l(x)}{l(x-1)}\right) = D(x-1, 1989) \quad (4)$$

对于熟悉生命表的人来说, 上面两个公式的意义是很明显的, 把(3)×(4)式, 利用公式

$$\frac{P(x-1, 1989)}{P(x-0.5, 1989)} = \frac{l(x-1)}{l(x-0.5)}$$

对乘积进行整理后可得

$$\frac{l(x)}{l(x-1)} + \frac{D(x-1, 1989)}{P(x+0.5, 1990)} \cdot \frac{l(x+0.5)}{l(x-1)} = 1 \quad (5)$$

由于 $l(x+0.5)/l(x-1) = \frac{l(x+0.5)}{l(x)} \cdot \frac{l(x)}{l(x-1)}$, 把它代入(5)式并进行整理后可得

$$\frac{l(x-1)}{l(x)} = \left(1 + \frac{D(x-1, 1989)}{P(x+0.5, 1990)} \cdot \frac{l(x+0.5)}{l(x)}\right)^{-1} \quad (6)$$

由于 $l(x-1)/l(x) = P_{x-1}$ 为x-1的生存概率, 在老幼年龄两端, $l(x)$ 按几何级数变化, 则 $l(x+0.5) = \sqrt{l(x+1)l(x)}/l(x) = \sqrt{l(x+1)/l(x)} = \sqrt{P_x}$; 在其它区

间, $l(x)$ 按算术级数变化, 则 $l(x+0.5)/l(x) = \frac{\frac{1}{2}[l(x+1)+l(x)]}{l(x)} = \frac{1}{2}(1+P_x)$,

把 $l(x+0.5)/l(x)$ 写成 P_x' , 则(6)式可写成

$$P_{x-1} = \left(1 + \frac{D(x-1, 1989)}{P(x+0.5, 1990)} P_x'\right)^{-1} \quad (6')$$

若我们知道了 x 岁的生存概率 p_x ，由 x 所在的年龄区间，可估计 p_x' ，由(6')式计算 p_{x-1} ，用样方法估计 p'_{x-1} ，计算 p_{x-2} ，这样递推下去可求出直到0岁的存活概率。

按照上面的方法，为了计算出各年龄的生存概率，对最末端的年龄需要给出一个估计数。人们要问这样求出的各年龄生存概率是否会有误差？可以证明，即使估计数有误差，在递推过程中，由末端年龄生存概率的误差带来的其它年龄别存活概率误差收敛于零，并且其收敛速度非常快。

在实际作法中，为了减少盲目性，在年龄末端可用方法1求出其存活概率，再用方法2算出直至0岁的存活概率。

由表1可以知道，这二种算法的计算结果基本上没有多大差别。

表1 中国1989~1990年的几种人口死亡指标

	1989年		1989~1990年		1990年	
	1	2	3	4	5	6
粗死亡率(‰)						
男女合计	5.796	5.798	6.245	6.182	6.274	6.620
男	6.212	6.213	6.621	6.552	6.661	6.950
女	5.356	5.358	5.846	5.790	5.867	6.268
婴儿死亡率(‰)						
男女合计	21.50	21.32	27.09	26.99	24.23	30.51
男	20.43	20.07	25.35	25.27	22.63	28.39
女	22.70	22.50	29.03	28.91	26.00	32.87
0岁期望寿命(岁)						
男女合计	71.28	71.28	70.26	70.57	70.25	69.54
男	69.32	69.31	68.47	68.77	68.45	67.90
女	73.43	73.40	72.16	72.49	72.20	71.24

*1~6表示按照本文中1~6方法对《中国1990年人口普查10%抽样资料》计算的结果。

(二) 计算普查前一年(1989.7.1~1990.6.30)的死亡指标

把人口普查时的人口数看作期末人口数，把1989年年中人口数看作期初人口数，可以有以下两种算法(为表述方便，方法序号接上页顺序编排，下同)：

3. 估计平均人口数，计算年龄别平均死亡率

由上面1989年年中人口数的计算方法，可得到期初人口数，于是平均人口数

$$\bar{P}_x = \frac{1}{2} [P(x, 1989) + P(x, 1990)] \quad (x=0, 1, \dots, w-1)$$

由此可算出平均死亡率

$$m_x = \frac{D(x, \text{II}) + D(x, \text{III})}{\bar{P}_x} \quad (x=0, 1, \dots, w-1)$$

4. 由生命表制作方法计算年龄别存活概率

把人口普查时年龄 x 岁人数看作期末人口数，则它和生命表中尚存人数 $l(x+1)$ 相对应，于是有

$$\frac{D(x, \text{II}) + D(x, \text{IV})}{P(x, 1990)} = \frac{d(x)}{l(x+1)}$$

其中 $d(x)$ 表示生命表中 x 岁的死亡人数, $d(x) = l(x) - l(x+1)$, 代入上式可得

$$\frac{l(x) - l(x+1)}{l(x+1)} = \frac{D(x, \text{II}) + D(x, \text{III})}{P(x, 1990)},$$

所以

$$p_x = \left(1 + \frac{D(x, \text{II}) + D(x, \text{III})}{P(x, 1990)} \right)^{-1}.$$

(三) 计算1990年的人口死亡指标

由于人口普查是在1990年年中作的调查, 所以普查结果本身不能提供1990年全年死亡人口资料。在此情况下可以这样处理: 一种是把人口普查时的人数看作是过去一年来同一个出生群死亡后剩下的人口。此方法类似于求1989年死亡指标, 所不同的是在求1989年人口死亡指标时, 没有直接的人口数可利用, 而在求1990年人口死亡指标时没有直接的死亡人口数可利用。另外一种处理方法是, 根据1990年上半年的死亡人口资料, 直接来估计1990年全年的死亡人口数。在求1990年人口死亡指标时, 两种方法都把普查人口数当作年中人口数。

5. 把普查人口看作1989年年中人口经过一年后的留存人口, 与相应的生命表尚存人数函数相对应:

$$\frac{P(x-0.5, 1989) - P(x+0.5, 1990)}{P(x+0.5, 1990)} = \frac{l(x-0.5) - l(x+0.5)}{l(x+0.5)},$$

其中 $P(x-0.5, 1989) - P(x+0.5, 1990)$ 可以用前述的 D_x 来估计:

$$D_x = \frac{3}{4} D(x-1, \text{II}) + \frac{1}{4} D(x, \text{II}) + \frac{1}{4} D(x-1, \text{III}) + \frac{3}{4} D(x, \text{III}),$$

$$(x=1, 2, \dots, w-1)$$

$$D_0 = \frac{1}{4} D(0, \text{II}) + \frac{3}{4} D(0, \text{III}),$$

$$\text{于是 } \frac{l(0) - l(0.5)}{l(0.5)} = \frac{D_0}{P(0.5, 1990)},$$

$$\frac{l(x-0.5) - l(x+0.5)}{l(x+0.5)} = \frac{D_x}{P(x+0.5, 1990)} \quad (x=1, 2, \dots, w-1)$$

设 $l(0) = 1.0$, 可计算出

$$l(0.5) = 1.0 / \left(1 + \frac{D_0}{P(0.5, 1990)} \right),$$

$$l(x+0.5) = l(x-0.5) / (1 + D_x / P(x+0.5, 1990))$$

$$(x=1, 2, 3, \dots, w)$$

由计算出的 $l(0.5), l(1.5), \dots, l(w-0.5)$ 进行插值就可求出 $l(1), l(2), \dots, l(w-1)$ 。本文中所用的插值方法为拉格朗日插值多项式

$$l(x) = \sum_{k=1}^3 \prod_{\substack{i=1 \\ i \neq k}}^3 \left(\frac{x - x_i}{x_k - x_i} \right) l(x_k).$$

其中 $x_1 = x-1, x_2 = x-0.5, x_3 = x+0.5$ ($x=1, 2, \dots, w-1$)。则年龄别存活概率:

$$p_x = l(x+1) / l(x).$$

6. 利用1990年上半年的死亡人口资料

$$\text{因为 } \frac{D(x, \text{III})}{P(x+0.5, 1990)} = \frac{l(x) - l(x+0.5)}{l(x+0.5)},$$

则

$$\frac{1(x+0.5)}{1(x)} = \left(1 + \frac{D(x, W)}{P(x+0.5, 1990)}\right)^{-1}$$

当 x 在年龄区间两端, $1(x+0.5)/1(x) = \sqrt{P_x}$; 在其它年龄区间 $1(x+0.5)/1(x) = \frac{1}{2}(1-P_x)$ 。在确定了存活概率 P_x 后, 可求出平均死亡率

$$m = \begin{cases} -1_n p_x & (1) \text{ 在年龄区间两端} \\ \frac{2(1-p_x)}{1+p_x} & (2) \text{ 在其它年龄} \end{cases}$$

表1是使用上述6种方法, 利用1990年全国人口普查10%抽样资料求出的几个全国人口死亡指标。由表1可以知道, 关于各种死亡指标, 第1、2种方法非常接近, 3~5比较接近。其中由第1、2种方法求出的结果所表示的死亡力最小, 方法6求出的死亡力最大。在上述方法中, 使用资料都是第四次普查时的人口数据, 但死亡人口用的资料却有所不同。方法1、2用的是普查中1989年的死亡人口数, 而方法3、4、5用的是1989年下半年、1990年上半年的死亡人口数, 方法6却仅仅用1990年上半年死亡人口数。可见, 方法的不同(只要基本合理)并不会使人口死亡指标与实际死亡力水平出现较大误差, 造成误差的主要来源在于资料本身。如果死亡报告中存在漏报、误报现象, 就不可避免地产生误差。从抽样得到的1989年上半年~1990年上半年的死亡报告人数来看, 1989年上半年为321 336人、1989年下半年为325 113人、1990年为376 568人。由此1989年下半年较1989年上半年死亡人口多1.18%, 而1990年上半年死亡数较1989年下半年死亡数多15.83%、较1989年上半年多17.19%。死亡报告数有这么大的差别, 除了说明有漏报现象存在外, 是难以用其它原因来解释的。关于死亡漏报程度, 漏报死亡人数等问题, 由于本文的目的只是讨论死亡指标的计算方法, 限于篇幅这里没有作详细分析。

(本文责任编辑: 徐 莉)

(作者工作单位: 北京经济学院人口经济研究所)

勘 误

本刊1992年第2期61页表3第三栏“瑞典”之下应为“英国”。同页表3无数字, 应为:

表3 1982年1%人口生育率抽样调查已婚有偶育龄妇女未避孕原因构成

未采取避孕措施的已婚 有偶育龄妇女人数	其 中							
	有 指 标		不 孕		绝 育		其 它	
	人 数	%	人 数	%	人 数	%	人 数	%
49 392	17 838	36.12	3 165	6.41	7 520	15.23	20 869	42.25