

各地区老年人口死亡率的数值模拟^{*}

黄润龙¹, 鲍思顿²

(1. 南京人口管理干部学院 江苏 南京 210042; 2. 美国得克萨斯 A & M 大学社会学系)

摘要: 依据我国第三、第四次人口普查资料, 首先对全国老年人口年龄别死亡率进行了数学模拟, 其次对我国 30 个地区老年人口年龄别死亡率与模型(全国老年人口)值进行了数学分析与模拟。研究表明, 全国及各地区老年人口死亡率均可表达为年龄 X 的简单数学函数。本文还研究了全国和各省市自治区老年死亡人口登记的相性误差。

关键词: 老年人口死亡率; 模型生命表; Logit 模型; 数值模拟

中图分类号: C921 文献标识码: A 文章编号: 1000-4149(2000)03-0017-06

Mathematics Modeling for the Elderly Mortality of by Regions

HUANG Run-long, Dudley L. Poston Jr.

(1. Nanjing College of Population Program Management, Nanjing, Jiangsu Province 210042,

2. Department of Sociology, A & M University, Texas USA)

Abstract: Using the third and fourth national population census data, a mathematics modeling is established to simulate the elderly age-specific mortality separately on national and provincial levels. The result shows that the elderly age-specific mortality both at national level or provincial level can be described as a simple function of age X. The article has also discussed the registered errors of the elderly mortality by national and provincial levels.

Keywords: the elderly mortality; model of life table; the Logit model; mathematics modeling

探索死亡规律、提高对于未来人口发展规律的认知能力是人类永恒的追求。如何用较少的参数来模拟人类死亡过程, 不少人口统计学家、生物统计学家和卫生统计学家进行了不懈的努力, 做了大量的工作。1959 年莱德曼(Ledeman)用因素分析法研究了 157 张实际生命表^[1], 提出至少需要 3 个参数(总死亡率、老年人口死亡率、儿童成人死亡率比)才能解释生命表方差的 93%, 1975 年布拉斯(Brass)提出了两参数相对模型生命表^[2], 1983 年尤班克(Ewbank)对模型进行了改造, 提出了参数意义明了的可压缩模型生命表^[3]。但这些生命表都是对整个人口而言, 对下降速度最快、变化幅度最大的老年人口没有特殊的关注。1994 年海姆斯(Himes)和普雷斯顿(Preston)^[4]提出在中高年龄段男女性年龄别死亡率的 Logit 函数与年龄 x 可能存在着较强的线性相关关系。鉴于中国老龄化发展趋势, 我们曾研究老年人口死亡指标的内在联系、老年人口死亡率随时间、地区的变化规

收稿日期: 1999-01-08

作者简介: 黄润龙, 男, 江苏人, 1991 年毕业于伦敦大学, 获硕士学位, 南京人口管理干部学院副教授。

^{*} 本文属国家统计局科研课题《中国老年人口死亡规律研究》的一部分, 1999 年 6 月通过鉴定, 结项评定为优秀。

律以及我国老年人口死亡率与发达国家情况的差异^[5]。本文将继而研究老年人口死亡率随年龄的变化模型(规律)和老年死亡人口的登记误差。具体是进行中国老年人口死亡率的数学模拟和仿真,探索用合适的函数(Logit 或 Ln 函数)、变量(死亡率 M_x 抑或死亡概率 Q_x)进行拟合,计算相应的拟合误差。

本文中 1981、1989 年全国年龄别死亡人数、死亡率资料分别源于国务院人口普查办公室 1993 年编的《中国 1990 年人口普查资料,第 4 册》。1989 年全国各省市自治区各年龄组死亡人数、死亡率资料是由中国人口情报中心提供的四普人口基本资料经计算加工而得。

一、全国老年人口年龄别死亡率的数学模拟

全国 50 岁以上男女性年龄别死亡(概)率的 Logit 函数、Ln 函数与年龄 x 之间的线性相关关系计算表明(表 1),死亡率比死亡概率更适合于相对模型的计算与拟合;Logit 函数对死亡概率的模拟优于 Ln 函数的模拟结果;1989 年男女性年龄别死亡率的模型值与年龄 x 之间的线性相关系数高于相应 1981 年相关系数,即 1989 年模型优于 1981 年;对女性模型的拟合优于男性模型,对 5 岁组模拟的线性相关系数略高于 1 岁组的情况,但考虑到 1 岁组的样本是 5 岁组的 5 倍,有理由认为是 1 岁组的拟合效果为好。由此可建立如下以年龄 X 为函数的死亡率模型:

表 1 普查(登记)死亡率与理论死亡率的相关系数*

| 年份 | 年龄组 | 死亡率 指标 | Ln(x) | | | Logit(x) | | |
|------|------|-----------|----------|----------|----------|----------|----------|----------|
| | | | 合计 | 男性 | 女性 | 合计 | 男性 | 女性 |
| 1981 | 1 岁组 | $Q(x)$ | 0.992725 | 0.983065 | 0.997046 | -0.99536 | -0.98597 | -0.99809 |
| | 1 岁组 | $M(x)$ | 0.994374 | 0.984853 | 0.996827 | -0.99612 | -0.98664 | -0.99795 |
| 1989 | 1 岁组 | $M(x)$ | 0.998507 | 0.997926 | 0.999189 | -0.99930 | -0.99911 | -0.99946 |
| | 5 岁组 | $M(x)$ | 0.998682 | 0.998163 | 0.999415 | -0.99946 | -0.99930 | -0.99972 |

* 1981 年龄组为 50~99 岁,1 岁组为 50 组样本。1989 年龄组为 50~89 岁,1 岁组为 40 组样本;5 岁组为 7 组样本。

$$\text{男性: } M_x^m = 6.95733 \times 10^{-5} * e^{0.092545X} \quad (r^2=0.9958) \quad (1)$$

$$\text{或 } M_x^m = \frac{1}{1 + e^{9.89627 - 0.09828X}} \quad (r^2 = -0.9982) \quad (2)$$

$$\text{女性: } M_x^f = 3.66939 \times 10^{-5} * e^{0.096263X} \quad (r^2 = 0.9984) \quad (3)$$

$$\text{或 } M_x^f = \frac{1}{1 + e^{10.452978 - 0.10048X}} \quad (r^2 = -0.9990) \quad (4)$$

下图就给出了实际年龄别死亡率及 Ln、Logit 模型死亡率。由图 1 可见,Logit 变换后的死亡率比 Ln 变换后的死亡率更接近于实际死亡率,尤其是 85 岁以上的高龄老人的死亡率。而男女性年龄别死亡率的性别比,由(1)、(3)式之比可得:

$$\frac{x \text{ 岁男性死亡率}}{x \text{ 岁女性死亡率}} = 1.8960454 * \text{EXP}(-0.003718 * X) \quad (X=50, 51, \dots, 89) \quad (5)$$

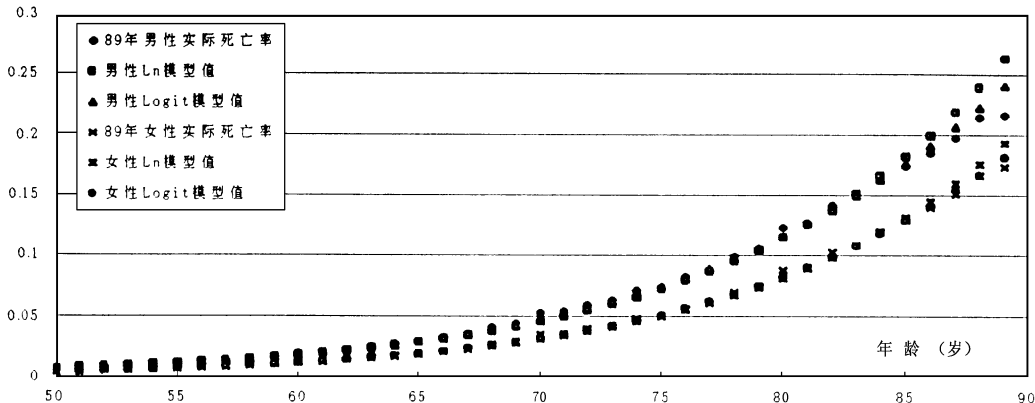


图 1 实际死亡率与模拟死亡率随年龄变动

这说明了 50 岁以上 X 岁年龄别死亡率的性别比, 随着年龄 X 增大呈指数递减。(5) 式中的指数也可用幂级数的形式进行展开。

二、各地区老年人口死亡率的模型模拟

对各地区老年人口而言, 海姆斯和普雷斯顿发现实际(地区)死亡率的罗杰特(Logit)变换值与模型值(全国)同样呈现严格的直线相关关系, 并可由此建立两参数相对模型生命表。作者并计算了西方 16 个工业国家不同时期的模型参数。根据这一思路, 我们取 1989 年 50~89 岁老年人口死亡率, 即(2)、(4)式作为模型值, 将同期中国各地区男、女性登记死亡率进行 Logit 变换。 $\alpha < 0$ 表示老年人口总体死亡率偏高; $\alpha > 0$ 则相反, 当然这是相对高低不是绝对的。 $\beta < 1$ 表示老年人口与成年人口死亡率之比偏大; $\beta > 1$ 则相反。1989 年我国 30 个省市老年人口死亡率的模型参数计算见表 2, 由表可见, 我国 30 个地区 α 值很小, 男、女性都分别在 -0.0005 至 0.0005 之间, 而不像工业化国家男、女性 α 值分别在 -0.6406 到 0.6028 之间; β 值则变动幅度较大。计算同时表明各省市男女性模型参数也有很大的一致性, 全国 30 个地区男、女性 β 值的相关系数为 0.923; 而中国老年人口死亡率主要受 β 值的影响, 死亡人口中 60 岁以上老年人口比例与 β 值有较高的相关系数, 男性的相关系数为 -0.744; 女性为 -0.784, 即 β 值偏高说明老年人口死亡比例偏低; 反之则偏高。于是得各地区 X 岁老人年龄别死亡率的估计值为:

$$M_x = \frac{1}{1 + e^{2\alpha} \times \left(\frac{1}{M_s} - 1\right)^\beta}$$

由于 α 近似等于 0, 所以 $\exp(2\alpha)$ 十分接近于 1。由(2)、(4)式求出 X 岁人口死亡率的模型值 M_s , 略加整理获得我国各地区男女性老年人口死亡率的估计式如(6)、(7)式所示。因此, 由表 2 可选取某地区的 β 值, 并代入公式(6)、(7)得到该地区各年龄别 X 死亡率的估计值。并由各地区每一组年龄别死亡率的估计值与实际登记值, 进而可求得该地区的相关关系 r。它表示该地区两参数模型对于实际死亡率的拟合程度, r 越大说明拟合程度越逼真。1981 年中国男、女性老年人口死亡率的估计值与实际值的相关系数为 0.9991、0.9990(n=50)。1989 年相关系数分别为 0.9991、0.9995(n=40)。这说明两参数相对模型对中国老年人口死亡率拟合的程度相当高。由表 2 的最后行可见, 不管男性还是女性, 1981 年到 1989 年 β 值呈明显的增加趋势, 如果变化是线性的, 是否可推算 2000 年乃至以后的全国及各地区老年人口死亡率呢? 这有待于第五次人口普查的死亡率调查资料的证实。再次, 就 1989 年各地区而言, 相关系数最低的地区是在新疆(男 0.9674、女 0.9632)、西藏(0.9862、0.9878)、青海(0.9923、0.9867)。

$$\text{男性: } M_x^m = \frac{1}{1 + (e^{9.89627 - 0.09828X})^\beta} \quad (r^2 = -0.9982) \quad (6)$$

$$\text{女性: } M_x^f = \frac{1}{1 + (e^{10.452978 - 0.10048X})^\beta} \quad (r^2 = -0.9990) \quad (7)$$

表 2 1989 年我国各省市老年人口相对死亡模型参数*

| 地区 | 男性 | | 女性 | | 地区 | 男性 | | 女性 | |
|------------------|----------|---------|----------|---------|------------------|----------|---------|----------|---------|
| | α | β | α | β | | α | β | α | β |
| 北京 | 0.0001 | 0.8860 | 0.0001 | 0.9093 | 河南 | 0.0000 | 1.0158 | 0.0000 | 1.0069 |
| 天津 | 0.0005 | 0.8886 | 0.0002 | 0.9520 | 湖北 | 0.0001 | 1.0065 | 0.0000 | 1.0059 |
| 河北 | 0.0000 | 0.9598 | 0.0000 | 0.9711 | 湖南 | 0.0001 | 0.9799 | 0.0001 | 0.9958 |
| 山西 | 0.0001 | 0.9328 | 0.0002 | 0.9525 | 广东 | -0.0001 | 1.0189 | -0.0001 | 0.9823 |
| 内蒙古 | 0.0000 | 0.9924 | 0.0001 | 1.0030 | 广西 | -0.0002 | 1.0601 | -0.0002 | 1.0610 |
| 辽宁 | 0.0000 | 0.9865 | 0.0000 | 0.9671 | 海南 | -0.0003 | 1.1395 | -0.0003 | 1.0447 |
| 吉林 | 0.0000 | 1.0737 | 0.0000 | 1.0605 | 四川 | 0.0000 | 1.0270 | 0.0000 | 1.0333 |
| 黑龙江 | -0.0002 | 1.1260 | -0.0002 | 1.1441 | 贵州 | -0.0001 | 1.0191 | 0.0000 | 1.0119 |
| 上海 | 0.0001 | 0.8638 | 0.0002 | 0.8411 | 云南 | 0.0001 | 0.9658 | 0.0001 | 0.9831 |
| 江苏 | 0.0000 | 0.9986 | 0.0000 | 0.9633 | 西藏 | -0.0005 | 1.2904 | -0.0005 | 1.3030 |
| 浙江 | 0.0000 | 0.9510 | 0.0001 | 0.9053 | 陕西 | 0.0001 | 0.9403 | 0.0001 | 0.9884 |
| 安徽 | -0.0001 | 1.0556 | -0.0001 | 0.9990 | 甘肃 | 0.0001 | 0.9896 | 0.0001 | 1.0238 |
| 福建 | -0.0001 | 1.0517 | 0.0000 | 1.0026 | 青海 | 0.0000 | 1.0352 | 0.0000 | 1.0824 |
| 江西 | 0.0001 | 0.9624 | 0.0001 | 0.9561 | 宁夏 | 0.0000 | 0.9941 | 0.0001 | 0.9596 |
| 山东 | 0.0001 | 0.9640 | 0.0001 | 0.9403 | 新疆 | -0.0005 | 1.1990 | -0.0005 | 1.2627 |
| 全国 ⁸⁹ | 0.0000 | 0.9981 | 0.0000 | 0.9988 | 全国 ⁸¹ | 0.0001 | 0.9697 | 0.0001 | 0.9813 |

* 各省市死亡人口原始资料来自中国人口情报中心。

为了解模型确切的拟合状态,我们将全国分为经济发达地区、经济中等发达地区和经济欠发达地区三类以检验死亡率数据具体模拟的质量,随机分别抽取上海、福建和宁夏为代表,计算老年人口实际年龄别死亡率和模型死亡率。计算表明,老年人口死亡率在中等发达地区模拟得最好,误差为最小;在我国西部经济欠发达地区拟合得次之;在上海、北京、天津等地区经济发达地区拟合质量最差。分析其间原因是多方面的,其既可能源于模型本身的缺陷,模型拟合能力不强,又可能源于老年人口的死亡登记性误差。由于我国地区性发展不平衡,社会经济的地区差异较大,我国老年人口死亡率差异也很大。我国西部经济文化相对落后,老年人口死亡登记差错较大,而在我国东部发达地区的老年人口死亡率已十分接近发达国家相应水平,该模型缺乏预测模拟能力,从而暴露模型本身的弱点。但总的说来低年龄段和女性模拟质量较好。

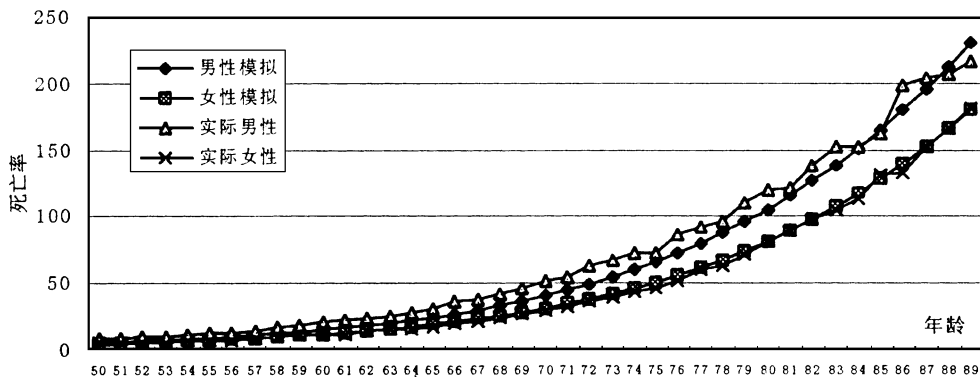


图2 福建省老年人口死亡率的数值模拟

三、老年死亡人口登记性误差

死亡的登记性误差主要来源于死亡人口的漏报和死亡年龄的误报两个方面。考虑到我国户口管理虽然较严格,但第一类型误差来源比较复杂,这里仅用最简单的模型讨论第二类误差。如果取3个年龄组死亡率的滑动算术平均为年龄别死亡率的真实值,观察(登记)值与真实值商与1的差异称为登记性误差相对百分比,表示登记死亡率误差的相对程度。1989年全国老年人口死亡率登记性误差相对百分比随年龄的变化分析表明(图4),58岁以下登记误差很小,58岁以上偶岁年龄登记性误差为正,奇数年龄登记性误差明显为负;登记误差较大的是80、88、82、70等吉利年龄组;各年龄组男女死亡人口登记误差正负号基本一致,但男性死亡登记性误差明显大于女性死亡登记性误差。在实际生活中,我国人民普遍使用的农历(虚岁)年龄,对死者的年龄有考大、虚报1~2岁的习惯,并且喜欢取吉祥或偶数的年龄或容易记忆的年龄。而我国死亡人口年龄登记也并不十分严格。这些都是造成死亡年龄登记差错的主要根源。1981年资料分析显示出十分相似的结果

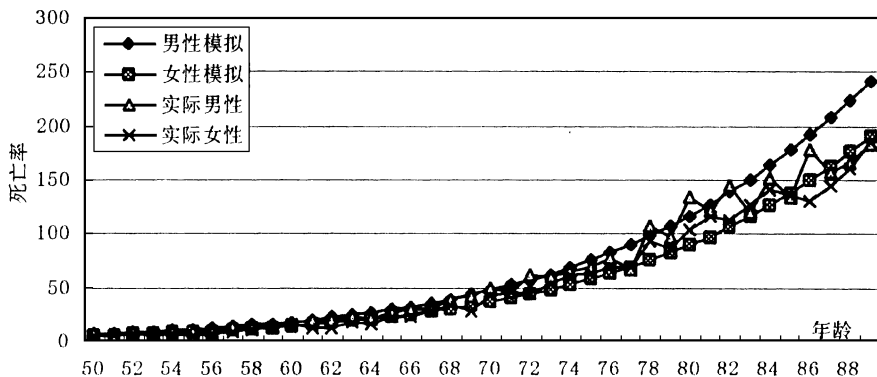


图3 宁夏老年人口死亡率的数值模拟

误差百分比

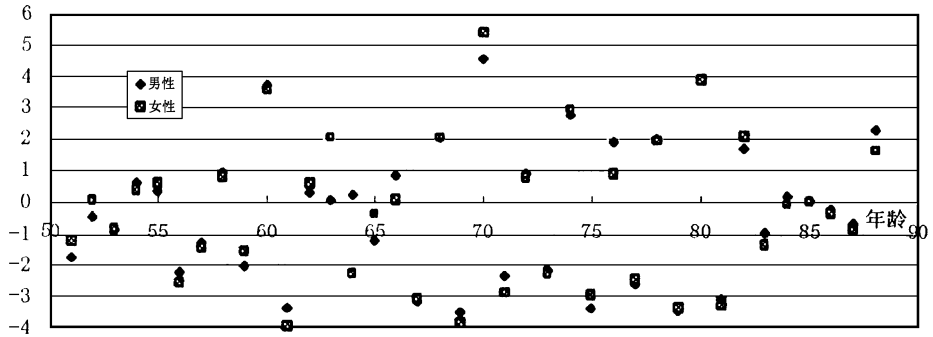


图4 死亡登记误差百分比随年龄的变动

果,并且这种规律一直可延伸到92岁。综合分析表明,无论是1981年资料还是1989年死亡率资料,无论是女性死亡率还是男性死亡率资料,从65~88岁年龄别死亡率稳定地存在着两岁周期性偏差现象,偶岁组高报死亡人口,奇岁组有低报死亡人口数据,仅有84~86岁组略有例外。这是因为汉语中四、死同音,且在部分地区85岁是作为一个吉祥年龄。5岁年龄组的死亡率曲线经过了平滑处理以后,登记性差错大为降低。如果

$$STD = \sqrt{\frac{1}{30} \sum_{x=59}^{88} \left(\frac{M_x}{M_x} - 1 \right)^2}$$

将59~88岁死亡率偏差的平均平方和定义为老年死亡人口登记性误差指数Std,1989年我国男、女性老年人口死亡登记误差指数为2.38%、2.49%;而1981年则为3.67%和3.83%。即80年代我国59~88岁老年人口死亡登记的平均误差为2%~4%。各地区老年死亡人口登记性误差指数统计表明(表3),登记误差指数较大的是在新疆(男18.7%、女20.2%)、西藏(11.0%、10.8%)、宁夏(9.9%、8.3%)、青海(7.6%、7.9%)和海南地区(6.7%、8.7%),尤其在新疆地区。而其他地区死亡人口登记性误差指数较小。注意到新疆、西藏和宁夏是我国少数民族比较集中的地区,而青海和海南地区又是我国经济欠发达地区,因此可以认为死亡登记性误差的大小是与这些地区的社会经济、文化发展等因素有关。

表3 1989年我国各省市59~88岁老年死亡人口登记性误差指数 %

| | 男性 | 女性 | 男性 | 女性 | 男性 | 女性 | | |
|-----|------|------|----|------|------|----|-------|-------|
| 北京 | 3.12 | 3.27 | 浙江 | 2.65 | 2.33 | 海南 | 6.65 | 8.67 |
| 天津 | 4.40 | 3.97 | 安徽 | 3.84 | 3.04 | 四川 | 2.58 | 2.92 |
| 河北 | 4.08 | 3.50 | 福建 | 2.65 | 2.24 | 贵州 | 3.13 | 4.26 |
| 山西 | 3.92 | 3.55 | 江西 | 3.90 | 2.60 | 云南 | 2.94 | 3.54 |
| 内蒙古 | 4.83 | 3.96 | 山东 | 2.50 | 3.35 | 西藏 | 11.03 | 10.75 |
| 辽宁 | 2.60 | 3.45 | 河南 | 2.86 | 3.39 | 陕西 | 3.48 | 2.92 |
| 吉林 | 2.66 | 4.86 | 湖北 | 2.22 | 3.09 | 甘肃 | 3.98 | 3.64 |
| 黑龙江 | 3.47 | 4.50 | 湖南 | 3.36 | 3.04 | 青海 | 7.57 | 7.91 |
| 上海 | 4.06 | 3.14 | 广东 | 3.22 | 2.98 | 宁夏 | 9.86 | 8.26 |
| 江苏 | 2.40 | 3.28 | 广西 | 2.66 | 3.20 | 新疆 | 18.71 | 20.24 |

原始资料来自中国人口情报中心。

由于新疆、西藏、宁夏、青海和海南等地在整个中国的人口比例及死亡人口比例极小,故对全国死亡人口资料影响很小,可以认为1989年中国人口死亡率基本可靠,但在局部研究新疆、西藏、宁夏、青海和海南老年死亡率时应引起适当的注意。

四、结论与讨论

21世纪我国老年人口仍将以较快速度增长。兹以江苏为例,1990~1995年全省人口以每年1.0%速度增长,而60、80、100岁以上的老年人口却以每年4.1%、5.2%和28%的速度增长。老年人口的研究是21世

纪人口研究的重要课题。依据我国第三、第四次人口普查资料, 本文对全国及各地区老年死亡率进行了数学分析和模拟, 并分析了死亡人口登记性误差。研究表明:

1. 死亡率比死亡概率更适合于对相对模型的计算与拟合; Logit 函数对死亡率的模拟优于 Ln 函数的模拟结果; 1989 年男女性年龄别死亡率的模型与年龄 x 之间的线性相关系数高于 1981 年的系数。

2. 全国及各地区老年人口死亡率均可表达为年龄 X 的简单数学函数。我国各地区 α 值本身数值很小, 男、女性都分别在 -0.0005 至 0.0005 之间; β 值则变动幅度较大。计算表明各省市男、女性模型参数也有很大的一致性; 而死亡人口中老年比例与 β 值有较高的相关系数。

3. 无论是 1981 年资料还是 1989 年死亡率资料, 无论是女性全国死亡率还是男性死亡人口资料, 从 65 ~ 88 岁年龄别死亡率稳定地存在着两岁周期性偏差现象, 偶岁组高报死亡人口, 奇岁组有低报死亡人口现象, 仅有 84 ~ 86 岁组略有例外。在海南、新疆、宁夏和西藏等地死亡人口登记误差都较大。

参考文献:

- [1] S. Ledeman and J. Breas. Les dimensions de la mortalite. Population, 1959, 5.
- [2] W. Brass. Use of the Logit System, Univ. Of North Carolina, Poplab, 1975.
- [3] D. C. Ewbank, J. C. Gomez M. A. Stoto. A Reducible four Parameter System of Model Life Table. Population Studies 1983. Vol 37, 105-127.
- [4] Christine L. Himes, Samuel H. Preston, Gretchen A. Condran. A relational Model at Older Ages in the Low Mortality Countries. Population Studies, 1994. Vol. 48; 269-291.
- [5] 黄润龙, 虞沈冠, 鲍思顿. 中国老年人口死亡变动规律研究. 南京人口管理干部学院学报, 1999(4).

(上接第 16 页)

参考文献:

- [1] 张善余. 我国省际人口迁移模式的重大变化. 人口研究, 1990(1)2~8.
- [2] 杨彦彦. 八十年代中国人口迁移的转变. 人口与经济, 1992(5). 12~16.
- [3] 王桂新. 中国的人口移动と经济开发. 统计, 1994. 第 12 号. 8~14.
- [4] 王桂新. 中国省际人口迁移地域结构探析. 中国人口科学, 1996(1)22~29.
- [5] Shryock, Henry S., Jacob S. Siegel and Associates(1976): The Methods and Materials of Demography. New York. Academic Press; 394~395.
- [6] 大友笃. 地域分析入门(改正版). 东洋经济新报社, 1997. 146~147.
- [7] WANG Guixin. The Rural-Urban Migration of Labor Force and Regional Development in China. Proceedings of the International Seminar on Human Resources and Regional Development in The APEC Member Economies, December 2-5, 1997, Tokyo and Hiroshima, Japan. March 1998, Institute of Developing Economies, Tokyo; 1998. 80-103.
- [8] 同[7]
- [9] 王桂新. 我国市镇乡人口省际迁移选择过程的空间模式及其差异. 华东师大学报(哲社版), 1993(5)8~15.
- [10] 王桂新. 我国省际人口迁移迁入目的地选择过程的空间模式及其差异. 人口与经济, 1995(6)40~49.
- [11] 同[9]
- [12] 同[10]
- [13] 同[7]
- [14] [15] 王桂新. 中国の人口の地域分布に関する研究. 日本女子大学纪要, 1994. 第 5 号. 159~188.