

文章编号: 1004-8227(2010)03-0244-05

# 昆山市人口指标的数学模型

姚鑫<sup>1,2,3</sup>, 杨桂山<sup>1,3</sup>, 孙洪波<sup>1,2,3</sup>, 万荣荣<sup>1,3</sup>

(1. 中国科学院南京地理与湖泊研究所, 江苏 南京 210008; 2. 中国科学院研究生院, 北京 100039;  
3. 中国科学院南京地理与湖泊研究所湖泊与环境国家重点实验室, 江苏 南京 210008)

**摘要:** 数学模型常用于人口指标变化趋势的拟合, 以昆山市 1985~2006 年 7 种人口指标为研究对象, 采用马尔萨斯模型、Logistic 模型和 Keyfitz 模型进行模拟, 结合历史资料分析人口指标变化背后的社会经济因素。研究结果表明: 从整体拟合效果来看, Logistic 模型表现良好, 但这很大程度上在于其三参数优势, 而不是人口增长面临最大人口上限的压力, 即使某一时段反映出人口上限, 该值也可能在下一时段被突破; 由于人口指标增长转折点的存在, 使利用数学模型进行人口预测的可信度降低, 但利用转折点分段分析, 了解其代表的政治、经济因素的影响, 可以为未来的规划决策提供科学参考; 昆山的发展历程明显分为 3 个阶段, 而不是以往认为的 5 个阶段。

**关键词:** 人口指标; 数学模型; 转折点  
**文献标识码:** A

沿海地区人口的增长加剧了区域资源和环境的压力, 制约了经济、社会的可持续协调发展<sup>[1]</sup>。人口问题的研究对于认识区域经济当前发展状况, 以及预测和规划区域经济未来发展趋势, 都有至关重要的意义<sup>[2]</sup>。

昆山市(120°48'21"E~121°09'04"E, 31°06'34"N~31°32'36"N) 位居我国东部沿海经济发达的苏、锡、常地区, 土地总面积为 928 km<sup>2</sup>, 2006 年全市 GDP 达到 932 亿元, 人均 GDP 91 583 元, 已成为以开放型经济为主导, 三大产业协调发展的全国百强县之一<sup>[3]</sup>。同年底, 昆山总户籍人口达到 66 万人, 外来人口达到 78 万人, 仅以户籍人口计算, 昆山人口密度即达到全国平均水平的 5.7 倍<sup>[4]</sup>。经济发展迅速、人口密度高的双重特色使得针对该地区的研究工作具有一定的代表性。

## 1 模型回归

本文选择昆山市 1985 至 2006 年统计年鉴共 22 年的人口数据, 选择依据为: 改革开放以后且已实施计划生育, 经济因素和政策因素的作用无太大波动。具体指标包括总户籍人口、非农业人口、外来

人口、总从业人数、不同产业从业人数等 7 项<sup>①</sup>。

采用回归模型为:

Model 1 自然增长指数模型(马尔萨斯模型):

$$P_t = P_0(1+r)^t \approx ae^{rt};$$

Model 2 Logistic 模型:

$$P_t = \frac{P_m}{1+ae^{-bt}}$$

Model 3 Keyfitz 双曲线模型:

$$\frac{1}{P_t} = \frac{1}{P_0} - \frac{rt}{P_0} = a + bt;$$

以上公式中:  $t$  表示时序, 1984 年  $t=0$ ;  $P_t$  表示  $t$  年的人口;  $P_0$  表示初始人口;  $P_m$  为环境最大承载人口;  $r$  为人口年增长率;  $a, b$  为模型参数。指数模型和双曲线模型参数可以在 SPSS 中通过回归分析计算, Logistic 模型先采用二分逼近法寻找决定系数(拟合优度)最大的  $P_m$ , 再于 SPSS 中进行回归<sup>[5~8]</sup>, Keyfitz 双曲线模型通过变形后于 SPSS 中拟合。得出所有模型参数和决定系数  $R^2$  后再于 EXCEL 中对其还原计算标准差  $SE$ 。

对表 1 进行分析: 综合考虑模型的决定系数及标准误差, Logistic 模型对于  $P_t, P_1, P_2, P_E$  的拟合效果胜过其他两种模型, 对  $P_3$  的拟合决定系数优于

收稿日期: 2009-03-09; 修回日期: 2009-08-16

基金项目: 国家自然科学基金(40601098)

作者简介: 姚鑫(1982~), 男, 江苏省兴化人, 博士研究生, 主要研究方向为资源利用。E-mail: anklestone@sohu.com

① 昆山市统计局, 昆山市统计年鉴, 昆山: 昆山市统计局, 1984~2006。

表 1 人口指标数学模型的决定系数及标准误差表

Tab. 1 Determination Coefficients and Standard Errors of Mathematical Models of Population Variables

		$P_T$	$P_N$	$P_I$	$P_1$	$P_2$	$P_3$	$P_E$
$M_1$	$R^2$	0.859	0.968	0.951	0.939	0.788	0.820	0.712
	SE	10 860	20 583	55 724	19 092	30 921	14 204	38 800
$M_2$	$R^2$	0.859 <sup>b</sup>	0.967 <sup>b</sup>	0.951 <sup>b</sup>	0.939 <sup>b</sup>	0.893 <sup>a</sup>	0.910 <sup>a</sup>	0.712 <sup>b</sup>
	SE	10 579	20 649	54 716	9 413	27 961	14 297	36 680
$M_3$	$R^2$	0.874	0.965	0.621	0.850	0.768	0.662	0.721
	SE	10 114	57 643	647 426	4 871 470	31 635	17 601	39 932

(a)  $P_T$  (Total Local Population) 表示总户籍人口;  $P_N$  (Nonagricultural Population) 表示非农业人口;  $P_I$  (Immigratory Population) 表示外来人口;  $P_1$  (First Industry Population) 表示第一产业从业人数;  $P_2$  (Second Industry Population) 表示第二产业从业人数;  $P_3$  (Third Industry Population) 表示第三产业从业人数;  $P_E$  (Employed Population) 表示总从业人数;  $M_1$  表示 Model 1;  $M_2$  表示 Model 2;  $M_3$  表示 Model 3;  $M_4$  表示 Model 4;  $a/b$  表示 Logistic 模型存在/不存在合理  $P_m$  值; 不存在合理  $P_m$  的含义即  $P_m$  不收敛: 对于任意  $P^{m'} > P_m$ , 有  $R^{2'} \geq R^2$   
 (b) 以上模型皆满足  $P < 0.001$

指数模型, 同时标准误差接近指数模型, 另外其对于  $P_T$  和  $P_N$  的拟合效果在 3 种模型中也居于第二。但进一步的分析表明, 只有对  $P_2$  和  $P_3$  进行拟合时存在合理的  $P_m$  值, 表明 Logistic 模型的拟合效果好并不一定反映人口增长受到环境、资源压力而具有增长上限, 可能是使用二分逼近法增加了模型参数  $P_m$  ( $P_m$  不收敛的情况以虚拟  $P_m$  为 100 000 000 进行计算), 使拟合精度超过其他双参数模型, 这种情况下的 Logistic 模型只具有数学意义而不具有社会经济含义。实际上当饱和值不收敛时, SPSS 默认 Logistic 模型为指数模型, 所以二者  $R^2$  相同, SE 的差异证明了 Logistic 模型的三参数优势。

指数模型反映了不受任何约束时自然状况下种

群中个体数量的增加, 一般适用于种群发展的初期。符合指数模型的增长方式可能说明如下信息: 基数小、高增长率、增长压力小<sup>[9]</sup>。所有指标中,  $P_N$  的增长以指数模式拟合效果最好。

Keyfitz 双曲线模型通过假设  $P_m$  足够大, 而对 Logistic 模型作形式变换, 以减小模型拟合的  $P$  值, 从而解决一些用 Logistic 模型拟合时因  $P$  值较大, 导致模型不能适用的问题。双曲线模型对  $P_T$  拟合效果较好, 符合双曲线模型反映了总户籍人口增长迅速, 呈资源掠夺式<sup>[6]</sup>。

综上, 兼顾 SPSS 对 Logistic 模型的处理方式, 表 2 采用指数模型拟合  $P_N$ 、 $P_I$ 、 $P_1$ 、 $P_E$ , 用 Logistic 模型拟合  $P_2$ 、 $P_3$ , 用 Keyfitz 双曲模型拟合  $P_T$ 。

表 2 不同人口指标的数学模型

Tab. 2 Mathematical Models of Different Population Variables

	模型方程	$R^2$	模型参数		
			$a$	$b$	$P_m$
$P_T$	$\frac{1}{P_t} = a + bt$	0.874	$1.88 \times 10^{-6}$	$-1.48 \times 10^{-8}$	/
$P_N$	$P_t = ae^{bt}$	0.968	$6.30 \times 10^4$	0.083 0	/
$P_I$	$P_t = ae^{bt}$	0.951	$2.73 \times 10^3$	0.260 0	/
$P_1$	$P_t = ae^{bt}$	0.939	$2.15 \times 10^5$	-0.080 0	/
$P_2$	$\ln(\frac{1}{P_t} - \frac{1}{P_m}) = a + bt$	0.893	-11.2	-0.090 5	800 000
$P_3$	$\ln(\frac{1}{P_t} - \frac{1}{P_m}) = a + bt$	0.910	-9.93	-0.143 0	250 000
$P_E$	$P_t = ae^{bt}$	0.712	$2.43 \times 10^5$	0.032 0	/

## 2 典型指标分析

选择总户籍人口、非农业人口、外来人口和第二产业从业人数进行典型指标分析。

### 2.1 总户籍人口( $P_T$ )

昆山市总户籍人口从 1985 年的 537 940 人增加至 2006 年的 666 359 人, 22 年共增加 128 419

人, 总增长率为 23.9%, 年均增长率达到 0.102%, 对比同期江苏省的户籍人口年均增长率 0.94%<sup>[8]</sup>, 可以看出, 昆山市的户籍人口增长速度略高于江苏省平均水平。符合 Keyfitz 曲线的拟合结果表明: 整体而言, 户籍人口仍然有很大上升空间。

图 1 为昆山市户籍人口散点图及双曲线拟合结果, 分析图 1 可以看出, 大约以 2000 年为界, 户籍人口在此前后的增长方式并不相同, 据此推测可能该时

段有某种社会经济因素的作用,使得户籍人口增长偏离原有曲线。为验证此推断,下面以 2000 年为界,进行分段模拟,结果显示 1985~1999 年段,户籍人口增长符合 Logistic 曲线,  $R^2 = 0.999$ ,  $P_m$  为 600 000; 2000~2006 年,增长方式符合指数模型,  $R^2 = 0.972$ 。这个结果可以基本确认户籍人口在 2000 年前后属于不同的增长方式,2000 年前,人口增长逐渐减缓,面临 Logistic 曲线 600 000 人口上限的压力,2000 年,可能由于某种社会经济因素的影响,人口增长方式突破了原有的人口上限 600 000,进入了新的指数增长阶段。

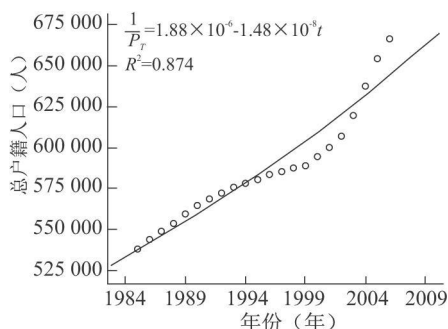


图 1 昆山市户籍人口增长图

Fig. 1 Increasing of Total Population in Kunshan

## 2.2 非农业人口( $P_N$ )

非农业人口和农业人口共同构成户籍人口,昆山市的非农业人口从 1985 年的 76 583 人增加至 2006 年的 382 045 人,总增长率达 398.8%,最高年增长率发生在 2000 年,为 43.3%。

从图 2 可以看出,2000 年也是非农业人口增长转折点,对其分段模拟结果显示:1985~2000 年,增长方式为指数形式,  $R^2$  为 0.989; 2000 年以后,同样为指数增长方式,  $R^2$  为 0.995。并且参数  $b$  值分别为 0.065 和 0.050,说明虽然同样为指数增长,但 2000 年以后增长速度不如 2000 年以前。

非农业人口的增加来源主要有 3 个途径,即自

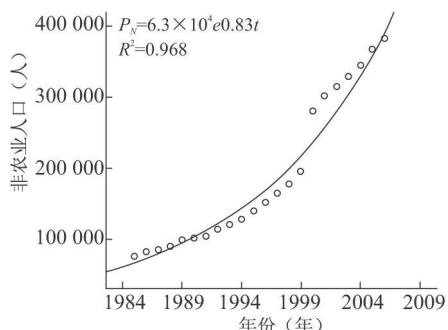


图 2 昆山市非农业人口增长图

Fig. 2 Increasing of Nonagricultural Population in Kunshan

然增长、外地非农业人口迁入和本地农业人口转为非农业人口。先作一个合理的假设:非农业人口自然增长率等于户籍人口自然增长率。那么结合户籍人口的分析结论,2000 年之前,非农业人口的增长主要原因不可能是自然增长或外界迁入,否则户籍人口上涨规律不会体现出人口上限。具体数值也证明了这一点,2000 年以前的数年内,非农业人口增加数大于总户籍人口增加,甚至达到几倍的规模,这反映了昆山在上个世纪末农业人口转为非农业人口的大规模城镇化进程。进入新世纪后,按照户籍人口增长规律,非农业人口在前文所提未知因素作用下,有了新的发展动力,但增长速度反而不如从前,表明城市化进程到 2000 年已经完成,2000 年的 43.3% 增长可能是城市化和新增长因素双重作用下的一个突飞猛进。

## 2.3 外来人口( $P_I$ )

昆山市对外来人口的统计始于 1990 年,为 10 063 人,至 2006 年已经增加至 780 956 人,总增长率达到 7 660.7%,2006 年的总户籍人口 666 359 人。外来人口整体增加的过程伴随着个别年份的减少,如 1993~1994 年以及 1996~1997 年也经历了一定规模的下降。图 3 为昆山市外来人口散点分布及指数模型拟合状况。

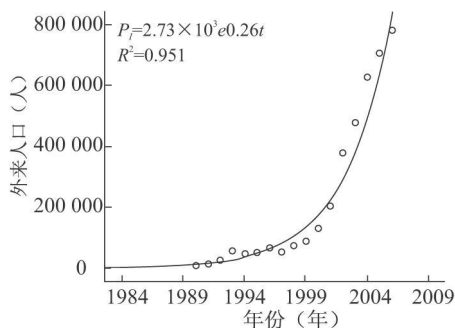


图 3 昆山市外来人口增长图

Fig. 3 Increasing of Immigratory Population in Kunshan

图中可以看出,昆山外来人口的增加基本符合指数增长模式,2000 年前外来人口缓慢增加,2000 年后曲线开始加速上升。但这也可以解释为外来人口基数较小,导致前期增长速度慢,仅从数据分布上看不出 2000 年外界因素的影响,故不对此分段分析。

## 2.4 第二产业从业人数( $P_2$ )

1985 年昆山市第二产业从业人数为 57 755 人,至 2006 年底增加为 359 181 人,总增长率为 521.9%。二分逼近法可以找到合适的  $P_m$ , 约为 800 000 人。仅从模型分析,可以认为  $P_2$  增长属于 Logistic 模式,反映了环境对人数增长的压力。 $P_2$

的散点分布及 Logistic 曲线拟合结果如图 4。

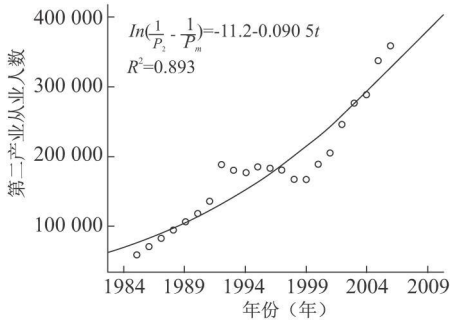


图 4 昆山市第二产业从业人数变化图  
Fig. 4 Increasing of the Second Industry  
Population in Kunshan

图 4 表明, 散点分布模式与典型 Logistic 的 S 形有很大出入, 而呈现出近似 N 形。散点分布明显分为 3 段: 1985~1991 年, 1992~1999 年, 2000~2006 年。按照对  $P_T$  的处理方式, 分段模拟结果如下: 1985~1991 年段, 增长符合指数模式,  $R^2$  为 0.993,  $b$  为 0.137; 1992~1999 年段, 为指数模式,  $R^2$  为 0.559,  $b$  为 -0.014, 略微下降; 2000~2006 年段, 增长再次呈现指数模式,  $R^2$  为 0.982,  $b$  为 0.111。

如果认为从业人数变化反映第二产业的发展状况, 那么昆山市第二产业的发展阶段性就很明确: 前后两个指数增长期, 中间为平缓调整期。一、三段之间进行比较表明: 最早的指数增长模式比现在的指数增长模式更快; 中间段上下波动频繁, 模型决定系数偏低, 表明可能存在多种外界因素的交互作用。

综上所述, 用 22 年数据建立的 Logistic 模型的模拟效果好, 可能只是 3 种增长模式的综合体现, 因而散点分布可以完全背离 Logistic 曲线的 S 形而呈 N 形态, 那么前面根据 Logistic 模型分析出的最大人口上限只具有数学意义, 并不反应真实的环境压力。

## 2.5 讨论

从以上指标的分析中, 可以看出人口指标的增长模式中经常出现转折点, 转折点的典型特征是: 前后的增长方式都呈现很好的统计规律, 但利用全部数据拟合模型的效果不佳, 因而根据长时间序列的人口数据建模以预测未来人口变化的可行性就大大降低。

进一步结合昆山的实例进行分析, 一般认为昆山市的发展分为 5 个阶段: 20 世纪 80 年代, 乡镇工业崛起, 农业转工业; 90 年代前期, 以 1992 年国务院批准成立国家级经济技术开发区为带动, 乡镇工业转为外向型经济; 90 年代后期, 特别是 1997 年金融风暴的影响, 昆山主攻台资, 开始向加工企业转

型; 2000~2006 年, 国家第一个出口加工区成立, 加工企业向高科技、集中化的方向发展; 2007 年后, 昆山的发展进入新的阶段。

结合这些资料再分析人口指标的变化, 就有如下的分析: 第一阶段, 1985~1991 年, 昆山的乡镇企业大规模发展 ( $P_2$  呈指数增长), 同时城镇化进程逐步加快 ( $P_N$  呈指数增长), 但此时的昆山对外来人员的吸引力有限 ( $P_I$  基数很低,  $P_T$  仍处于 Logistic 增长段), 总就业人数增长缓慢, 不能满足第二产业发展的需求, 主要靠其他产业人员流动来满足第二产业发展的需要; 第二阶段, 1992~1999 年, 国家级经济开发区的成立带动了企业转型, 前期过度发展的乡镇企业经济转型为外向型经济, 第二产业经历一定时间的调整期 ( $P_2$  波动明显), 受 1997 年金融风暴的影响明显 ( $P_2$  指数拟合  $b$  为负值), 同时昆山的城镇化到了最后阶段 ( $P_N$  进入指数增长末期), 但昆山对外来务工人员的吸引力仍然有限 ( $P_T$  处于 Logistic 增长末期, 且  $P_I$  尚未进入指数增加段); 2000 年以后, 城镇化进程基本完成 (2000 年  $P_N$  出现 43.3% 的历年最大值, 此后进入缓慢增长期), 国家级出口加工区成立, 对外来人口吸引力有了质的飞跃 ( $P_T$  进入指数增长期,  $P_I$  大幅度增加, 甚至大于  $P_T$ ), 昆山开始进入全面发展阶段 ( $P_2$  进入新的指数增长期)。

同时, 基于以上分析, 可以认为人口增长转折点主要由一些具有重大影响的政策、经济因素造成, 如国家级经济技术开发区、国家级出口加工区的成立, 金融风暴也存在一定影响, 但仍不足以形成人口增长的转折点。转折点出现的概率可能跟以下几个因素相关: 采用数据的时间跨度, 越长出现转折点几率越大; 研究区域的开放程度, 封闭区域人口流动性低, 受影响小; 研究区域的尺度, 小尺度区域更容易受政策因素影响, 也更易出现转折点; 偶然因素, 包括政治和经济上的偶然事件。

## 3 结论

人口指标研究常用的几种数学模型中, Logistic 模型的拟合效果整体较好, 但这并不一定表明人口指标的变化面临 Logistic 模型理论中的人口上限的压力, 而可能只是 Logistic 模型三参数优势的体现。即使如 2000 年以前总户籍人口的变化所示, 真实反映了人口增长上限, 该上限也可以在一定政策、经济因素的影响下被突破, 并不是一成不变的。相比而言, 指数模型参数计算更方便, 实用性较强, 多数情

况能满足研究需要,且当 Logistic 模型的  $P_m$  不收敛时,基本可以代替 Logistic 模型。Keyfitz 曲线模型对  $P_r$  拟合效果优于其他两个模型,但由于其整体决定系数并非很好,远不如分段拟合结果,表明这可能仅仅是数据上的巧合。

目前,人口指标研究采用数学模型主要是为了进行预测,本研究表明,由于人口增长转折点的存在,使得长时间序列数据的变化失去统计规律,从而使利用模型预测的结果可信度降低。正确判别转折点,区别随机扰动,用最近的转折点后的数据进行建模可能更有价值,但具体数学方法仍需要进一步研究。

分析转折点的社会经济学意义在于:通过研究区域历史资料,找出其所代表的政策、经济因素,揭示它们对社会和产业经济造成的实际影响,为今后的规划决策提供科学的参考。

以往认为昆山发展历史分为 5 个阶段,对人口指标的数学模型研究结果表明昆山市的发展历程应分为 3 个阶段:1985~1991 年,以乡镇企业崛起为特征

的城镇化期;1992~1999 年,金融风暴压力下经济结构转型的调整期;2000 年以后,社会经济新的发展期。

## 参考文献:

- [1] 段学军,陈雯.江苏省可持续发展分析、评价与战略对策[J].长江流域资源与环境,2003,12(3):199~204.
- [2] 李作臣.论城市人口预测和城市资源与环境容量的关系[J].广州大学学报(综合版),2001,12(5):18~21.
- [3] 虞沈冠.江苏省昆山市人口、资源、环境与可持续发展的战略思考[J].中国人口·资源与环境,1996,6(1):33~38.
- [4] 龙花楼,刘彦随,唐国平,等.昆山土地利用变化与新农村建设[J].北京大学学报(自然科学版),2007,43(6):790~796.
- [5] 汤江龙,赵小敏.土地利用规划中人口预测模型比较研究[J].中国土地科学,2005,19(2):14~20.
- [6] 陈彦光,余斌.人口增长的常用数学模型及其预测方法[J].华中师范大学学报(自然科学版),2006,40(3):452~456.
- [7] 陈彦光.人口与资源预测中 Logistic 模型承载力参数的自回归估计[J].自然资源学报,2009,24(6):1107~1114.
- [8] 杨丽霞,杨桂山,苑韶峰.数学模型在人口预测中的应用[J].长江流域资源与环境,2006,15(3):288~291.
- [9] 周一星,陈彦光.城市与城市地理[M].北京:人民教育出版社,2003.

## MATHEMATICAL MODELING OF POPULATION GROWTH IN KUNSHAN

YAO Xin<sup>1,2,3</sup>, YANG Gu+shan<sup>1,3</sup>, SUN Hong-bo<sup>1,2,3</sup>, WAN Rong-rong<sup>1,3</sup>

(1. Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China; 2. Graduate University of Chinese Academy of Sciences, Beijing 100039, China; 3. State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China)

**Abstract:** Mathematical models are widely used to simulate the development trend of population. In this paper, four mathematical models, named as Malthusian model, Logistic model, linear regression model and Keyfitz model were studied on their capability of modeling the historical development trend of seven population variables of Kunshan, Jiangsu Province from 1985 to 2006. Results showed that, generally, Logistic model was better than other models according to determination coefficients. Then each index was analyzed separately with an aim to explain the differences between simulated and recorded numbers. From the results, it was concluded that: 1) logistic model fitted well primarily because it had three parameters while other models had two, and the additional parameter  $P_m$ , which should stand for the upper limit of population increasing, did not have real sense in most cases; 2) the prediction of future population by the use of mathematical modeling showed unfeasible for the existence of turning points in the development trend of population variables; 3) combined with historical information of the research area, turning points could give a good description on the effects of political and economic factors occurred simultaneously; 4) the development course of Kunshan should be divided into three stages, the developing stage, the adjustment stage and the overall developing stage, rather than five stages generally accepted previously.

**Key words:** population variable; mathematical model; turning point