

Quantitative Analysis of Economic Impact on Population Migration in Xinjiang

Zhen Li, Naiyu Feng

Henan University, Kaifeng Henan
Email: gengtao@henu.edu.cn

Received: Sept. 10th, 2015; accepted: Sept. 24th, 2015; published: Sept. 29th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

What kind of rules that economy in the ethnic minority regions impacts on the flow of population of ethnic minorities is ambiguous. For a long time, the qualitative analysis, limited to the theory apply and thick lines, is the main method. And this method is described with lack of objective data analysis and historical trend. This paper tries to use quantitative method to study the problem of population migration. Firstly, we analyze the spatial construction and historical changes of migration among provinces. Secondly, influence factors of population migration among provinces in China are focused on. In this paper, the suitable theory model of population migration for ethnic areas is built based on the classical population migration model analysis. By analysis on the influencing factors, the model is proved reasonable. Proposing the hypothesis of the coefficient of the model through the principal component regression analysis by MATLAB, the model is proved reasonable, and the empirical model of population migration in ethnic minority areas is got at last by SPSS 19.0.

Keywords

Population Migration, Ethnic Minority Areas, Correlation Coefficient, System Clustering, Multiple Linear Regression

经济对新疆地区人口迁移影响定量分析

李 珍, 冯迺郁

河南大学, 河南 开封
Email: gengtao@henu.edu.cn

收稿日期: 2015年9月10日; 录用日期: 2015年9月24日; 发布日期: 2015年9月29日

摘要

民族地区经济对少数民族人口的流动影响具有什么样的规律, 长期以来, 仅限于理论套用和粗线条的定性分析, 缺少客观性数据分析和历史性态势描述。本文试图使用定量方法以新疆地区为例研究经济因素对人口迁移问题的影响。首先, 分析了省际间人口迁移的空间格局和历史演变。其次, 建立统计学模型研究我国人口迁移圈的影响因素。本文在经典人口迁移模型分析基础上, 构建了适合民族地区人口迁移的理论模型。利用Matlab对模型假设的影响因子进行相关性分析, 确定模型的合理性, 通过SPSS 19.0主成分回归确定模型系数, 得到民族地区人口迁移的实证模型。

关键词

人口迁移, 民族地区, 相关系数, 系统聚类, 多元线性回归

1. 研究的背景及意义

人口迁移历史同人类历史一样久远, 人口迁移的范围同人类的生存活动空间一样宽广。国际间、国境内无时不在进行着频繁的人口迁移。带着不同使命和目的的人们进行了不同方式的迁移, 却达到了一个共同的宏观效果——推动社会进步、传播先进的生产技术、促进文化的交流和发展。任何一个国家, 从民族的形成到社会形态的变化, 无不伴随着人口迁移。发达资本主义国家工业化的完成, 是在大规模人口迁移的前提下实现的。随着二次世界大战炮声的平息, 广大的发展中国家从帝国主义列强的统治和奴役下解放了出来, 发展自己的民族工业, 实现工业化也同样毫无例外地需要通过人口大迁移来完成。人口迁移以它对人类进步的巨大贡献证明了它的意义和价值。我国少数民族人口流动是各种因素综合作用的结果, 其发展态势也是对我国不同时期民族地区社会经济发展状况的客观反映。而在这些因素当中民族地区经济对少数民族人口的流动影响具有什么样的规律, 长期以来, 仅限于理论套用和粗线条的定性分析, 缺少客观性数据分析和历史性态势描述。因此基于权威性数据, 通过统计学、数学模型来定量分析少数民族人口流动与民族地区经济发展情况, 以数字权重说明关联度, 以图表模型印证发展态势, 使问题研究更加科学直观, 显示发展规律, 将是本领域一次开拓性的尝试[1]-[4]。

当代中国人口的迁移和流动问题是伴随中国经济发展和社会转型出现的一个重大现实问题, 是未来很长一段时间内人口和发展研究方面最有活力的领域之一。研究涉及经济学、人口学、地理学、社会学、统计学、数学等若干学科的交叉探讨与研究, 第六次人口普查数据为现实问题的研究提供了新的定量解构基础, 因此该问题具有特别的研究价值, 本文所用到的符号及符号意义见表1。

2. 经典人口迁移数学模型

1) 重力模型

由于人口迁移本身的复杂性以及详细准确的迁移数据较为缺乏, 使得人口迁移模型与生育、死亡模型相比显得不够成熟。从研究社会经济现象的客观需要出发, 可选择适当的方法来对人口迁移现象进行估计。1949年, 美国社会学家吉佛把“万有引力定律”引入推拉模型, 并应用于人口迁移研究。重力模型是最早的迁移模型, 一个多世纪以来, 重力模型一直被应用于人口迁移分析。其出发点是认为大多数迁移所越过的距离很短; 当吸收迁移的中心地点距离迁出地愈远时, 由于迁移中遇到种种困难, 迁移人数迅速下降; 而每个迁移流都将产生一个反向迁移流。简言之, 两地之间迁移人口与两地人口规模成正比, 与两地之间距离成反比。基于此提出了重力模型(gravity model):

Table 1. Symbols and their meanings in this paper
表 1. 本文所用到的符号及符号意义

符号	符号的意义
Y	迁入人口数
X_1	迁入来源省份总人口数
X_2	迁入来源省份人均 GDP
X_3	迁入来源省份少数民族人口数
X_4	迁入来源省份接受高等教育人口数
X_5	迁入迁出省会间铁路距离
F_{ij}	第 j 个省份的第 i 个指标值
\bar{F}_i	i 指标的均值
s_i	i 指标的标准差
\tilde{F}_{ij}	标准化后的第 j 个省份的第 i 个指标值

$$M_{ij} = K \frac{P_i P_j}{D_{ij}} \quad (1)$$

式中, M_{ij} 为 i 地与 j 地之间的人口迁移量, P_i 和 P_j 分别为两地人口规模, D_{ij} 为两地之间的距离, K 为常数, α 为距离衰减系数。

α 表示迁移量受距离的影响程度, α 值越大, 预测的迁移量受两地间距离的影响也越大; $\alpha = 0$ 时, 两地间的距离对迁移量没有影响。当距离衰减作用很大时, 表明大多数人距离出发地不远; 当距离对人口迁移量影响不大时, 那么两地的人口规模将成为决定人口迁移量的主导因素。距离对迁移的影响主要用来描述迁移产生的交通成本、心理成本以及不确定性。心理成本源于与家庭和朋友的分离, 迁移的距离越远, 团聚的机会越少, 心理成本越高[5]-[9]。

2) 马尔科夫链模型

马尔可夫过程理论是随机过程理论的一种, 它描述了这样一种过程, 该过程在某一时刻(或空间位置)将达到的状态, 仅依赖于目前所处的状态, 而与以往状态无关; 因此它描述了问题依据内在的规律而发生变化, 反映了事物动态的演变过程。马尔可夫过程不仅能够描述时间序列的变动过程, 也能够描述结构变动的过程。因此, 马尔可夫过程为我国省际间人口流动的动态分析提供了有效的依据[10]。

若人口迁移的变化在一段时间内很平稳, 即人口迁移变化不大, 则人口迁移变化过程可看成是马尔科夫链规律变化, 可利用人口迁移概率矩阵求出一定时期人口分布的变化。因此我们也称马尔科夫链模型为迁移矩阵模型。

设初始人口分布为向量 M_0 , 人口迁移概率矩阵为 $P = (P_{ij})$, 下一期人口分布为 M_1 则有:

$$M_1 = M_0 P \quad (2)$$

当迁移概率矩阵稳定不变时, 由马尔科夫链的性质可求出其后 n 个时间周期以后的人口分布:

$$M_n = M_{n-1} P = M_{n-2} P^2 = \dots = M_0 P^n \quad (3)$$

式中, P^n 为人口迁移概率的 n 次幂, P 是迁移概率矩阵, 其每一个元素 P_{ij} 表示从 i 到 j 地转移概率。

$$P = (P_{ij}) = \begin{pmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{m1} & \cdots & P_{mn} \end{pmatrix} \quad (4)$$

$$P_{ij} = \frac{M_{ij}}{\sum_{i=1}^n M_{ij}} \quad (5)$$

3. 人口迁移问题的分析

根据第五次、第六次全国人口普查数据, 进入 21 世纪以来五个民族自治区(西藏自治区、新疆维吾尔自治区、宁夏回族自治区、内蒙古自治区、广西壮族自治区)人口迁移量逐年增多, 各民族自治区迁入的人口均有显著增加。五个民族自治区迁入迁出人口规模如表 2。

本文以新疆维吾尔自治区为样本地区对其人口迁移的基本情况加以分析。

新疆维吾尔自治区省际人口迁入数量逐年增多, 可以占到全国省际总迁入人口的 2.086%。迁移人口来源于全国各地, 但主要是陇海铁路沿线和长江、黄河流域人口稠密地区。表 3 统计了新疆省际人口迁移来源地分布。

Table 2. The scale of the moving population for five national regions
表 2. 五个民族自治区迁入迁出人口规模

	迁入人口(人)	迁出人口(人)	净迁移人口(数)	迁入人口占比
内蒙古自治区	82,859	64,759	18,100	56.13%
广西壮族自治区	59,999	282,053	-222,054	17.54%
西藏自治区	9197	6249	2948	59.54%
宁夏回族自治区	23,922	15,066	8856	61.36%
新疆维吾尔自治区	84,050	28,669	55,381	74.54%

Table 3. Distribution of the source of the population migration in Xinjiang province
表 3. 新疆省际人口迁移来源地分布

省份	占比	省份	占比	省份	占比
甘肃	21.03%	湖南	1.81%	北京	0.57%
河南	20.73%	浙江	1.5%	吉林	0.52%
四川	17.9%	山西	1.12%	贵州	0.44%
陕西	6.01%	广东	1.09%	内蒙古	0.44%
山东	3.78%	青海	1.08%	上海	0.36%
安徽	3.74%	福建	0.93%	广西	0.24%
重庆	3.69%	黑龙江	0.74%	天津	0.19%
江苏	2.9%	辽宁	0.65%	港澳台	0.11%
宁夏	2.17%	江西	0.62%	海南	0.1%
河北	1.9%	云南	0.6%	西藏	0.03%

我国西部地区与东部地区在经济上还存在较大差距, 从某一地区迁移至另一地区的人口数量必然和两地的经济状况相关。而地区人均 GDP 能够很好地反映一个地区的经济状况, 因此模型构建时需要考虑迁入地和迁出地的地区人均 GDP。

我国民族地区的少数民族比例很高, 而少数民族由于语言文字及宗教信仰等方面和汉族有较大差异, 所以在流动性上和汉族不同。因此模型构建时需要考虑迁入地和迁出地的少数民族人口数。

我国民族地区的人口文化素质和东部的非民族地区有一定差距, 通常来说, 一个人受教育的程度越高, 就越有可能迁移。因此在模型构建时需要考虑迁入地和迁出地的接受高等教育人口数。

我国人口迁移是各种因素综合作用的结果, 同时也是对我国不同时期社会经济发展状况的反映, 更是对人口发展状况的反映。我国民族地区的人口迁移问题除有特殊性外, 也与一般情况的迁移有相似之处。如区域间的迁移受迁入地和迁出地的人口分布的限制, 及迁入地和迁出地之间地理距离等等, 这些因素也需要考虑到民族地区的人口迁移中。其中, 考虑到省会城市经济发展水平、基础设施建设都比较好, 省际迁移大部分是围绕省会城市迁移, 而且火车仍是人们国内进行长途旅行的主要交通工具。基于此本文对迁入地和迁出地之间的地理距离用两地区省会城市间铁路距离来表示[5]。

整理的数据如表 4 所述。

4. 实证分析

本文首先要对迁入人口数这一因变量的所有自变量做了相关分析。考虑到数据的复杂性, 人工处理相关矩阵显然不合理, 这里借助 Matlab 软件 Corrcoef 函数求解标准化矩阵的相关系数矩阵。

处理后的相关系数矩阵如表 5。

Table 4. Data table
表 4. 整理的数据表

地区	迁入人口数(人)	总人口数(人)	人均 GDP(元)	少数民族人口数(人)	接受高等教育人口数(人)	省会间铁路距离(公里)
甘肃	17,675	25,575,254	12,882	2,410,498	1,923,250	1948
河南	17,423	94,023,567	21,073	1,128,283	6,015,570	3017
四川	15,044	80,418,200	17,289	4,907,951	5,368,247	3026
陕西	5051	37,327,378	20,497	186,636	3,940,303	2544
山东	3177	95,793,100	35,893	651,393	8,328,700	3803
安徽	3143	59,501,000	16,656	374,856	3,985,000	3737
重庆	3101	28,846,200	20,219	1,851,926	2,493,000	3358
江苏	2437	78,659,903	43,907	259,577	8,506,817	3774

Table 5. Correlation coefficient factor
表 5. 因子相关系数

	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Y	1.000	0.049	-0.547	0.613	-0.275	-0.696
X ₁	0.049	1.000	0.536	-0.025	0.873	0.573
X ₂	-0.547	0.536	1.000	-0.463	0.867	0.646
X ₃	0.613	-0.025	-0.463	1.000	-0.269	-0.350
X ₄	-0.275	0.873	0.867	-0.269	1.000	0.651
X ₅	-0.696	0.573	0.646	-0.350	0.651	1.000

从表格中可以看出因变量迁入人口数(Y)除了与迁入来源省份总人口数(X_1)相关性不强外,跟其他自变量的相关性都比较强。可以建立因变量与众多自变量之间的线性回归模型。

通过对不同地区因变量的分析,发现上述8个省区的情况是不同的,在进行多元回归拟和时拟合方案也应该是不同的。据此,本文依据各省区因变量指标将8个省区进行系统聚类分析。系统聚类方法是研究多要素事物分类的数量方法。其过程是找出能度量这些数据或指标之间相似程度的统计量;然后以统计量作为划分类型的依据,把一些相似程度大的变量(或样品)首先聚合为一类,而把另一些相似程度较小的变量(或样品)聚合为另一类,直到所有的变量(或样品)都聚合完毕,最后根据各类之间的亲疏关系,逐步画成一张完整的分类系统图,又称谱系图。进行类别合并的准则是使得类间差异最大,而类内差异最小。

运用 SPSS 19.0 统计分析软件完成聚类过程,最后得到8个省区聚类的谱系图(图1),本文以距离15为分界点,将8个省区分为两类:

经过系统聚类分析(如表6),本文将第1类省区为例用自变量指标拟合一条直线说明迁入人口数(Y)和自变量(X)的关系。要拟合出的多元线性回归模型:

$$y = a_0 + \sum_{i=1}^n a_i z_i^* + \varepsilon \quad (6)$$

其中, ε 服从期望为0,方差为 σ^2 的正态分布,并满足基本假定。

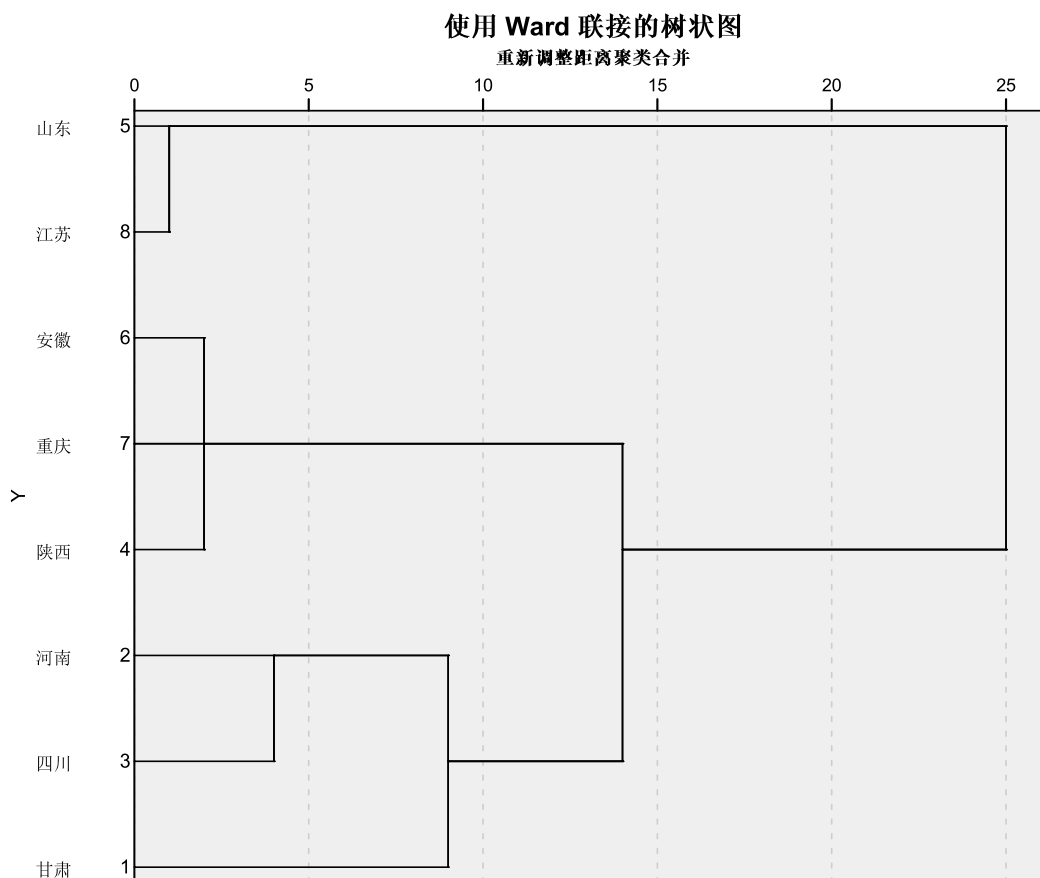


Figure 1. The pedigree chart of system clustering of 8 provinces

图1. 8个省区系统聚类的谱系图

Table 6. The analysis results of system clustering of 8 provinces
表 6. 8 个省区系统聚类分析结果

类别	省区	个数
1	甘肃、四川、河南、陕西、重庆、安徽	6
2	山东、江苏	2

多元线性回归方程为

$$y = \hat{a}_0 + \sum_{i=1}^n \hat{a}_i z_i^* \quad (7)$$

用 SPSS 19.0 统计分析软件中的分析 - 回归 - 线性工具, 求解出的结果:

$$Y = 31436.36 + 0.001X_1 + 0.398X_2 + 0.001X_3 - 0.007X_4 - 10.536X_5 \quad (8)$$

为说明各指标自变量与迁入人口数目的密切关系, 本文继续对求解得到的五元线性回归模型进行了显著性检验:

A) 判定系数

拟合优度 R 检验

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (9)$$

当 R 越大, 说明 Y 与 X_1, \dots, X_m 之间的关系越密切。

从表 7 中看到, 复相关系数 $R = 1$, 多重判定系数 $R^2 = 1$ 。模型的拟合度相当好。

B) 回归方程显著性检验

F 检验

因变量(Y)与自变量(X)是否存在如方程(8)所示的线性是需要检验的。显然, 如果求出的所有回归系数 $|\hat{a}_i| (i=1, \dots, m)$ 都很小, 那么 Y 与 X_1, \dots, X_m 之间的线性关系就不明显。所以可令原假设为

$$H_0 : a_i = 0 (i=1, \dots, m)$$

当 H_0 成立时,

$$F = \frac{SSR/m}{SST/n-m-1} \quad (10)$$

其中, $SST = SSR + SSE$, SST 是总离差平方和, SSE 是残差平方和, SSR 是回归平方和。在显著性水平 α 下有上 α 分点 $F_\alpha(n, n-m-1)$, 当 $F < F_\alpha(n, n-m-1)$ 时, 则拒绝 H_0 ; 否则, 接受 H_0 。

从表 8 中看可以通过 F 检验与 t 检验, 说明模型整体是显著的。

C) 回归系数显著性检验

运用最小二乘原理, 求回归系数 a_0, a_1, \dots, a_n ;

运用最小二乘原理, 使得残差平方和;

$$E = \sum_{i=1}^n \varepsilon_i^2 \quad (11)$$

最小, 从而求出回归系数。

Table 7. Models
表 7. 模型汇总

模型	R	R 方	调整 R 方	标准 估计的误差
1	1.000 ^a	1.000	.	.

a. 预测变量: (常量), X_5, X_3, X_4, X_2, X_1 ; b. 因变量: Y 。

Table 8. Variance analysis table
表 8. 方差分析表

模型	平方和	df	均方	F	Sig.
1	回归	2.582E8	5	51642315.90	.a
	残差	0.000	0	.	.
	总计	2.582E8	5	.	.

a. 预测变量: (常量), X_5, X_3, X_4, X_2, X_1 ; b. 因变量: Y 。

Table 9. Coefficient a
表 9. 系数 a

模型	非标准化系数		标准系数	t	Sig.	
	B	标准误差	试用版			
(常量)	31436.360	0.000	.	.	0.000	
1	X_1	0.001	0.000	2.216	.	0.000
	X_2	0.398	0.000	0.173	.	0.000
	X_3	0.001	0.000	0.135	.	0.000
	X_4	-0.007	0.000	-1.605	.	0.000
	X_5	-10.536	0.000	-0.919	.	0.000

a. 因变量: Y 。

表9中给出了模型对5个变量的偏回归系数是否等于0的t检验结果, P值均小于显著性水平0.05, 因此认定偏相关系数 β 显著不等于0, 每一个指标都和因变量线性相关显著。

5. 结论

本文以历年中国统计年鉴及全国人口普查资料为原始数据, 建立了具有民族地区特色的人口迁移模型。模型从经济、地理、受教育程度及地区民族占比等方面考虑影响民族地区人口迁移的影响因素。运用整理的统计数据, 利用 Matlab 软件和 SPSS 19.0 统计分析软件对模型假设的影响因子进行相关性分析, 论证了模型的合理性。

将通过系统聚类分析得到的新疆维吾尔自治区第 1 类迁移来源地区和第 2 类迁移来源地区进行多元线性回归拟合, 并进行了拟合优度检验和显著性检验。得到了如下的迁入人口数(Y)和自变量(X)的多元回归方程。

第 1 类地区的回归方程:

$$Y = 31436.36 + 0.001X_1 + 0.398X_2 + 0.001X_3 - 0.007X_4 - 10.536X_5$$

第 2 类地区的回归方程:

$$Y = -93865.069 + 25.517X_5$$

通过对数据分析得到如下结果:自本文统计到的 2000 年第五次全国人口普查数据起, 民族自治区人口迁移量逐年增多, 从相关系数矩阵中可以看出, 就本文所研究的样板民族地区——新疆维吾尔自治区而言, 迁入人口依赖于迁入来源省份人均 GDP, 受地理位置、交通等限制, 往往向距离迁出地区较近的地方迁移, 再者迁入人口受迁入来源省份民族人口规模和受教育程度的影响。在这些因素之中, 经济因素占据主要地位, 迁徙的基本规律受经济因素制约, 因此, 经济基础决定了人类社会化活动规律。论文通过建立适当的数学模型, 可以解决许多无法量化的社会人口流动问题, 分析少数民族流动的影响因素和态势结构, 判断影响该地区经济发展的因子与权重。通过本课题的研究, 探讨少数民族人口流动的影响因素, 并结合民族地区人口流动的实际特点, 进一步探求民族地区经济发展规律。

参考文献 (References)

- [1] Weidlich, W. and Hagg, G. (1983) Concepts and models of a quantitative sociology: The dynamics of interacting population. Springer, Berlin.
- [2] Ommeren, J.V. (1999) Job moving, residential moving, and commuting: A search perspective. *Journal of Urban Economics*, **46**, 230-253.
- [3] Raphael, S. and Riker, D.A. (1999) Geographic mobility, race and wage differentials. *Journal of Urban Economics*, **45**, 17-46.
- [4] 王桂新 (1996) 中国人口迁移与区域经济发展关系之分析. *人口研究*, **6**, 9-16.
- [5] 王桂新 (1993) 我国省际人口迁移与距离关系之探讨. *人口与经济*, **2**, 3-8.
- [6] 刘振亚 (1990) 发展中国家的口迁移问题. *农村经济与社会*, **4**, 48-53.
- [7] 黄荣清, 赵显人, 等 (2004) 20 世纪 90 年代中国各民族人口的变动. 民族出版社, 北京, 175-212.
- [8] Burro, R. and Sala-i-Martin, X. (1995) *Economic Growth*. McGraw Hill, New York.
- [9] 姜启源, 谢金星 (2003) *数学模型*. 第三版, 高等教育出版社, 北京, 294-326.
- [10] 张羚广, 蒋正华, 林宝 (2006) *人口信息分析技术*. 中国社会科学出版社, 北京.