

2015 年(第四届)全国大学生统计建模大赛 参赛论文

基于 Malthus 模型的人口老龄化预测及影响因素分析¹

参 赛 队 员 _____ 王莹 王燕霞 张惠宁

指 导 教 师 _____ 杜永强

学 校 院 系 _____ 天津商业大学 统计系

日 期 _____ 2015 年 6 月 30 日

¹ 注:该论文获得由中国统计教育学会举办的“2015 年(第四届)全国大学生统计建模大赛”大数据统计建模类本科生组三等奖。

摘要

自从我国 20 世纪 80 年代实行计划生育政策以来，人口出生率迅速减少，为控制人口的增长带来了显著的成效。但随之而来的是人口老龄化问题的逐年加剧。人口老龄化给中国的经济、社会、政治、文化等方面的发展带来了深刻影响，庞大的老年群体的养老、医疗、社会服务等方面需求的压力也越来越大。目前，这些压力还只是初现端倪，随着人口老龄化快速发展，这些压力的影响将会更加深刻、更加普遍。

本文通过绘制“人口金字塔”分析了我国 1995-2013 年人口老龄化的现状；通过建立 Malthus 人口预测模型，对我国 2015-2025 年人口情况进行了预测，得出我国人口老龄化的严重性；分别以教育经费的支出、人口密度、国内生产总值、卫生费用总支出为自变量，以老年人口比重为因变量，建立多元线性回归模型，对我国人口老龄化的影响因素进行了分析。实证表明：一是教育费用的总支出和政府社会卫生总支出是影响我国老年人口比重最重要的因素；二是由教育费用的总支出和政府社会卫生总支出所建立的回归模型对预测我国未来老年人口比重具有较高的精度。

关键词：老龄化，人口金字塔，Malthus 人口模型，多元线性回归模型

1 研究背景和研究意义

1.1 研究背景

联合国教科文组织规定，一个国家 60 岁以上的人口占其人口总数的 10%或以上，或一个国家 65 岁以上的人口占其人口总数的 7%或以上，则该国家就进入了老龄化社会^[1]。

我国现阶段人口老龄化严重。依照联合国教科文组织对老龄化社会的界定，我国在 2000 年已经进入了老龄化国家的行列。2010 年第六次人口普查资料显示，全国 0-14 岁的少年儿童占总人口的比例是 16.60%，60 岁以上老年人口占总人口的 13.26%，其中 65 岁及以上占总人口的 8.87%。而 80 年代“婴儿潮”期间出生的人口，将在 2045 年前后陆续进入老年阶段，中国的人口老龄化进程会突然加速而逐渐达到人口老龄化高峰。

劳动力减少、养老负担加重、政府财政负担加重。老年人口比重的增加意味着青壮年劳动力的减少，社会生产力的下降；此外我国人口老龄化具有速度快、人口基数大、未富先老等特点，其发展的后劲很强，再加上我国社会保障体系本身存在诸多不完善之处，导致老龄化问题给中国带来了养老负担。

我国政府对人口老龄化问题高度重视。应对老龄化挑战，我国政府决定要把老龄化社会作为 21 世纪一个重要国情认真对待。把解决老龄化社会各种矛盾和问题纳入全面建设小康社会和社会主义现代化建设的总体发展战略中来，做好应对老龄化社会的各项准备。强调要大力发展老龄产业，制定老龄产业行业发展规划，颁布实施国家对老龄产业的扶持保护政策，

建立老龄产业发展管理体制。另外，在 2011 年起，我国将以黑龙江、吉林、辽宁、江苏和浙江为试点，允许夫妻双方只要有一人是独生子女就可生二胎。足见国家对人口老龄化的高度重视。

1.2 研究意义

通过建立“人口金字塔”，分析现阶段我国老龄化问题的严重性。自从我国实行计划生育以来，我国人口出生率明显下降，而老年人口所占比例逐年增加。在 2000 年我国 65 岁以上人口已经占到了我国总人口的 7%，我国进入了老龄化社会。这势必导致我国劳动力下降，阻碍了社会生产力的发展。同时老年人口的养老问题变得越来越突出，青壮年一代面临着较为繁重的赡养老人的压力，给家庭和社会带来了极大的经济负担。

通过建立 Malthus 人口预测模型，预测我国未来 10 年人口老龄化状况，旨在分析我国未来人口老龄化的趋势。中国作为人口大国，老龄化的形势更加严峻。到本世纪 40 年代后期，全国老年人口将超过 4.3 亿人，老年人口将占总人口的三分之一左右，我国将成为老龄化程度最严重的国家之一。据联合国统计，到 2050 年，中国 80 岁以上的老龄人口将达到一亿人，占世界的五分之一。人口老龄化给我国政府带来了严峻的挑战。甚至有人断言，未来十年，老龄化对发达国家的影响超过金融危机。而在中国，这一问题也不可忽视，为此人口老龄化问题亟需解决。

通过建立多元线性回归模型，找到引起我国老龄化的主要影响因素，提出解决策略。老龄化状况的日趋严重，严重阻碍了我国经济的发展，因此解决老龄化问题迫在眉睫。为此找出影响我国老龄化问题的主要因素是解决这一问题的关键所在。

2 研究问题

问题 1：我国人口老龄化的现状如何？

解决问题 1 的思路：收集 1995 年到 2013 年我国人口各年龄段男女人数，利用 excel 绘制出“人口金字塔”结构图及不同年龄段人口数量的折线图，进而分析我国人口结构的变化趋势。

问题 2：我国未来 10 年(即 2015-2025 年)老年人口数量为多少？

解决问题 2 的思路：根据 1982 年到 2013 年 1-14 岁、15-64 岁、65 岁以上三个年龄段人口数量的统计，建立 Malthus 人口预测模型，预测未来 10 年我国人口数。

问题 3：影响我国老龄化的因素是什么？

解决问题 3 的思路：结合文献[1][4]影响我国老龄化的直接因素分别为人口的出生率和死亡率。本文以出生率和死亡率为立足点，参照国内外相关文献[9-15]筛选出影响出生率和死亡率的四个因素。即教育经费的支出、人口密度、人均国内生产总值、卫生费用总支出。这四个因素与出生率和死亡率的对应关系见表 1。分别以教育经费的支出、人口密度等为自变量，以老年人口比重为因变量，建立多元线性回归模型，对我国人口老龄化的影响因素进行分析。

表 1 影响人口老龄化的因素

直接因素	实际因素	如何影响
------	------	------

出生率	教育经费支出	投资力度的加大说明国民受教育程度提高,传统观念有所改变,进而减少了人口的出生率 ^[4] 。
	人口密度	人口密度大时,由于空间与自然资源的限制,会导致出生率的下降 ^[4] 。
死亡率	人均国内生产总值	人均国内生产总值越高,说明人民的生活水平越好,人的寿命就得到延长,导致死亡率降低。
	卫生费用总支出	投入的资金越多,公众享受到的服务和保障力度越大,就越有益于人的寿命的延长,死亡率越小。

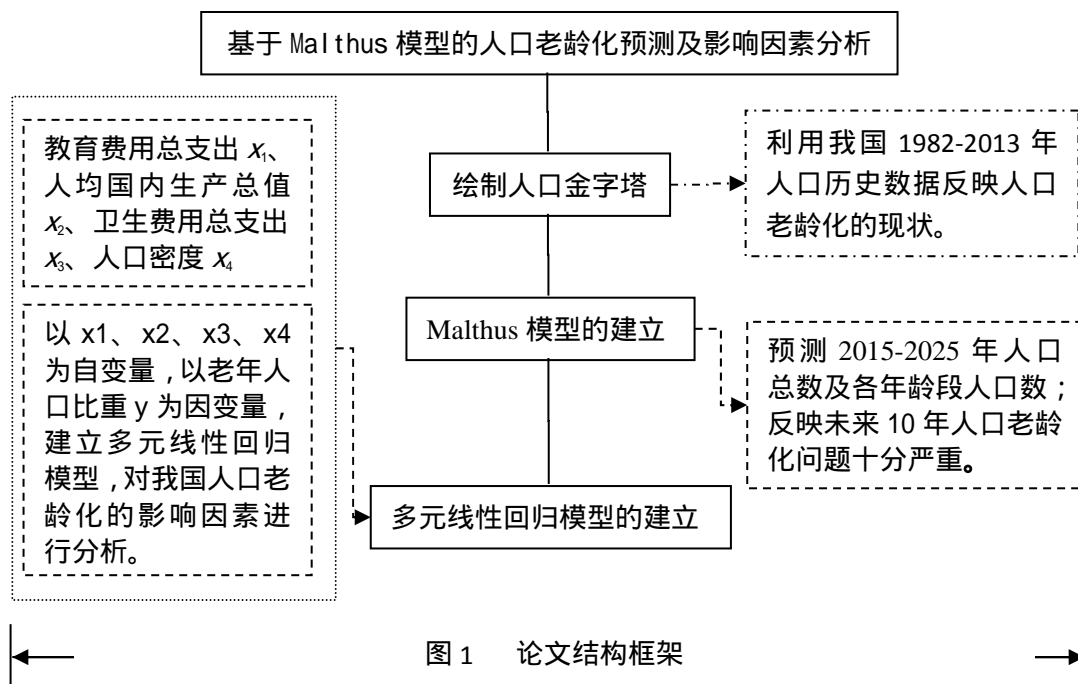


图 1 论文结构框架

3 实现路径

3.1 人口老龄化的现状分析

“人口年龄金字塔”可以清楚地表示出不同年份中各个年龄段和性别人口在总人口中所占的比例。通过收集我国 1982-2013 年在不同年龄段的人口数量[2][5],画出了 1997 年、2000 年、2003 年、2006 年、2009 年、2012 年我国的人口年龄金字塔(见图 1),从图 1 中可以看出人口老龄化的严重性。

图 1 左边为特定年龄组男性人口占全部人口的百分比,右边为女性人口的百分比。从图 1 可知,2000 年之前人口年龄结构形成的金字塔具有“宽大的底座和比较狭长的顶尖”,呈现上窄下宽的形状,意味着年轻人口所占的比例较大。“宽大的底座”意指 0-14 岁的青少年人口数量较其他各年龄段的人口数量要多,“比较狭长的顶尖”意指 65 岁以上老年人口数量较少。而到了 2000 年以后,人口金字塔的底座已经缩小,顶尖变宽,表明少年儿童人口比重减小,老年人口比重增大,人口结构趋于老化。

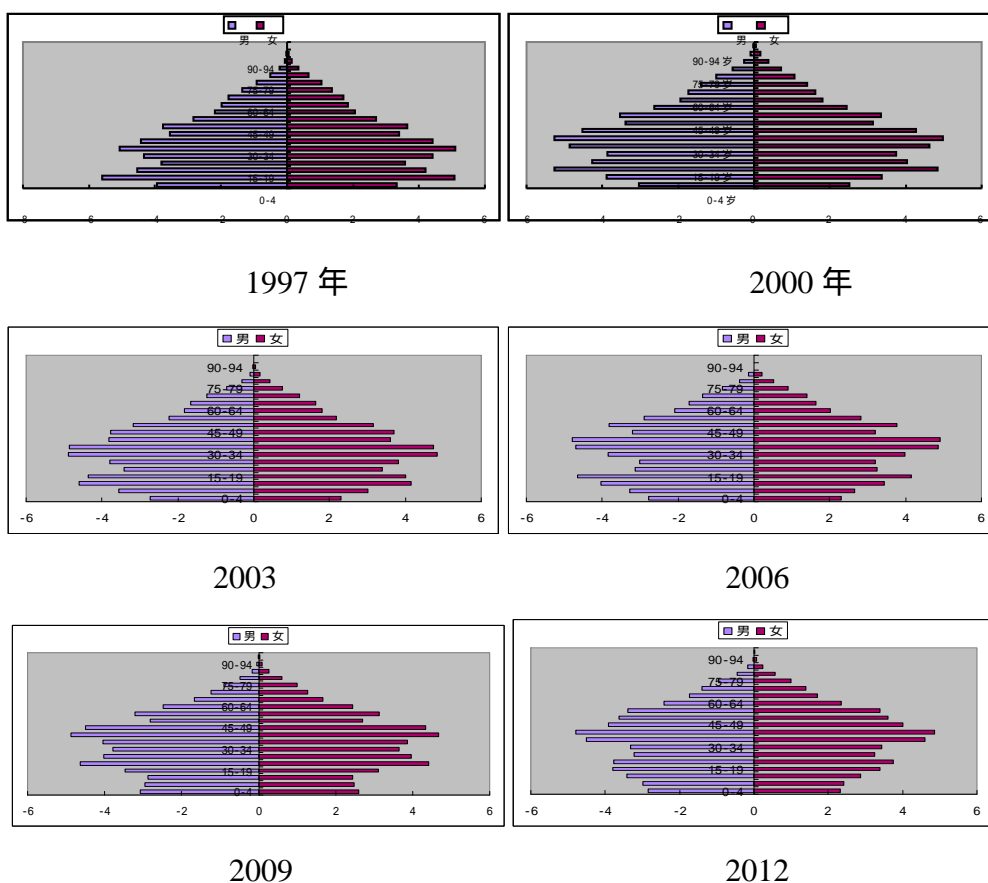


图2 人口年龄金字塔

为了更清晰的分析我国人口老龄化的现状，利用 excel 画出如下折线图，见图 3。

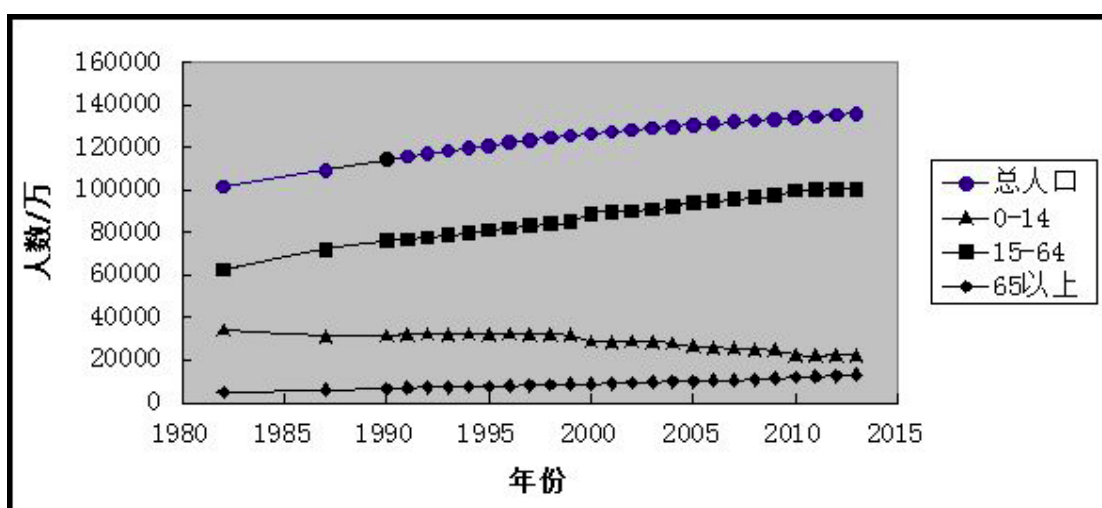


图3 我国不同年龄段人口折线图

从图 3 可知：总人口数量和年龄段在 15-64 岁的人口数量都在逐年递增，并且增长趋势相同(图 3 中的“圆圈”所连的折线与“方形格”所连折线接近平行)。从图 3 中的“三角”所连折线可知，我国 0-14 岁人口数量明显减少。从图 3 中的“菱形”所连折线可知，

我国 65 岁以上的人口数量呈上升趋势。

通过“人口金字塔”可以直观的看出具体某一年男女各年龄段人口所占比例的大小，可以清晰地反映各个年份的人口结构情况。在绘制了多个人口金字塔之后，又可以清楚地看出我国人口结构模型发生的显著变化。由起初的“下宽上窄”变为如今的“中间宽两头窄”即所占人口比例较大的年龄段在逐渐上移，由 1997 年的 0-14 岁上升为如今的 30-50 岁年龄段，照此趋势发展未来几年我国人口所占比例最大的人口将很有可能变为 65 岁以上年龄段人口。由此可见我国人口老龄化的严重性。

通过绘制我国不同年龄段人口折线图，可以清楚地观察到从 80 年代到现如今我国各年龄段的人口变化趋势，老年人口数量的逐年增加，青少年人口数量的逐年递减，最终将导致老年人口比重高于青少年人口数量的比重，社会老龄化现象加重。

综上所述，通过横向、纵向两个角度分析了现如今我国人口老龄化的现状，那么在未来的几年内我国老龄化现象会有所改善，还是会持续恶化呢？针对这一问题，很有必要对我国未来的人口状况进行预测，进而分析我国未来老龄化的状况。

3.2 我国人口老龄化的预测

根据 1982 年到 2013 年 1-14 岁、15-64 岁、65 岁以上三个年龄段人口数量的统计，建立 Malthus 人口预测模型^[3]，预测未来 10 年我国老年人口数。

3.2.1 Malthus 模型的构建

(1) 模型建立的假设条件

假设收集的数据均真实有效。假设未来几年内无战争、重大自然灾害等因素的影响。人口的迁入与迁出率保持稳定。

(2) 模型建立

假设在一个小的单位时间间隔 Δt 内，人口出生率为 b ，人口死亡率为 c 。 $t + \Delta t$ 时刻的人数 $P(t + \Delta t)$ 等于 t 时刻人数 $P(t)$ 加上在 Δt 间隔内新出生的人数减去死亡的人数。即，

$$P(t + \Delta t) = P(t) + bP(t)\Delta t - cP(t)\Delta t \quad (1)$$

或

$$\frac{\Delta P}{\Delta t} = bP - cP = kP \quad (2)$$

其中 k 是常数，称为内禀增长率。当 k 为正数时说明出生率大于死亡率，当 k 为负数时说明出生率小于死亡率。因此， k 反映了人口的净增长比率。

由式(2)，在一个时间段内人口的平均年龄变化率与人口数量成比例。用瞬时变化率来逼近平均变化率，可以得到下面的微分方程，其中 t_0 为初始时刻， P_0 为初始时刻对应的人口数。

$$\frac{dP}{dt} = kP, P(t_0) = P_0, t \geq t_0 \quad (3)$$

通过对式(3)进行求解，解得

$$P(t) = P_0 e^{k(t-t_0)} \quad (4)$$

式(4)就是 Malthus 人口增长模型^[3]。

由式(4)可以求出，

$$k = \left(\frac{1}{t - t_0} \right) \ln \frac{P}{P_0} \quad (5)$$

表 2 1982 年到 2013 年总人口、各分段人口和比重数据(单位：万人)

年份	总人口	0-14 岁		15-64 岁		65 岁以上	
		人口数	比重	人口数	比重	人口数	比重
1982	101654	34145.5786	33.59	62517.21	61.5	4991.2114	4.91
1983	102133	34011	33.3	63596	62.27	4526	4.43
1984	103211	33568	32.52	64786	62.77	4857	4.72
1985	105843	32005	30.24	68567	64.78	5271	4.98
1986	107533	31984	29.74	70023	65.18	5526	5.14
1987	109300	31347.24	28.68	71984.98	65.86	5967.78	5.4
1988	110058	31366	28.5	72569	65.94	6123	5.56
1989	112598	31542	28.01	74745	66.38	6311	5.6
1990	114333	31658.8077	27.69	76305.8442	66.74	6368.3481	5.57
1991	115823	32094.5533	27.71	76790.649	66.3	6937.7977	5.99
1993	118517	32177.3655	27.15	79050.839	66.7	7288.7955	6.15
1994	119850	32359.5	27	79868.04	66.64	7622.46	6.36
1995	121121	32218	26.6	81393.312	67.2	7509.502	6.2
1996	122389	32311	26.39	82245.408	67.2	7833	6.41
1997	123626	32093.3096	25.96	83447.55	67.5	8085.1404	6.54
1998	124761	32064	25.7	84338	67.6	8359	6.7
1999	125786	31949.644	25.4	85157.122	67.7	8679	6.9
2000	126743	29012	22.89	88910	70.1	8821	6.96
2001	127627	28716.075	22.5	89849.408	70.4	9062	7.1
2002	128453	28774	22.4	90302	70.3	9377	7.3
2003	129227	28559	22.1	90976	70.4	9692	7.5
2004	129988	27947	21.5	92184	70.9	9857	7.6
2005	130756	26504	20.3	94197	72	10055	7.7
2006	131448	25961	19.8	95068	72.3	10419	7.9
2007	132129	25660	19.4	95833	72.5	10636	8.1

2008	132802	25166	19	96680	72.7	10956	8.3
2009	133450	24659	18.5	97484	73	11307	8.5
2010	134091	22259	16.6	99938	74.5	11894	8.9
2011	134735	22164	16.5	100283	74.4	12288	9.1
2012	135404	22287	16.5	100403	74.1	12714	9.4
2013	136072	22329	16.4	100582	73.9	13161	9.7

利用表 2 中的历史数据计算适合我国不同年龄段人口数量的内禀增长率 k 。

利用 1982-2013 年的数据，计算出不同年份各年龄段的 k 值，并求其平均值作为总人口数、0-14 岁人口数、15-64 岁人口数和 65 岁以上人口数的内禀增长率 k 。

例如，计算 0-14 岁年龄段的内禀增长率 k 。将表 2 中 1982-1983 年 0-14 岁人口数量数据 34145.5786、34011 分别代入上文式(5) 得 1982 年 0-14 岁年龄段的内禀增长率 k 为 -0.004，列入表 3。同理，可得 1983-2011 年 0-14 岁年龄段的内禀增长率 k ，列入表 3 的相应行和相应列。将 1982-2011 年 0-14 岁年龄段的内禀增长率 k 的算术平均值作为 0-14 岁年龄段的内禀增长率 k ，列入表 3 最后一行。

同理，可以得到其他年龄段的 k 值列入表 3 最后一行的相应列。

表 3 各年龄段的 k 值

年份	k_{0-14} 岁	k_{15-64} 岁	k_{65} 岁以上	k 总人口
1982	-0.004	0.017	-0.098	0.005
1983	-0.013	0.019	0.071	0.010
1984	-0.048	0.057	0.082	0.025
1985	-0.001	0.021	0.047	0.016
1986	-0.020	0.028	0.077	0.016
1987	0.001	0.008	0.026	0.007
1988	0.006	0.030	0.030	0.023
1989	0.004	0.021	0.009	0.015
1990	0.014	0.006	0.086	0.013
1991	0.008	0.011	0.040	0.012
1992	-0.005	0.018	0.010	0.011
1993	0.006	0.010	0.045	0.011
1994	-0.004	0.019	-0.015	0.011
1995	0.003	0.010	0.042	0.010
1996	-0.007	0.015	0.032	0.010
1997	-0.001	0.011	0.033	0.009
1998	-0.004	0.010	0.038	0.008

1999	-0.096	0.043	0.016	0.008
2000	-0.010	0.011	0.027	0.007
2001	0.002	0.005	0.034	0.006
2002	-0.008	0.007	0.033	0.006
2003	-0.022	0.013	0.017	0.006
2004	-0.053	0.022	0.020	0.006
2005	-0.021	0.009	0.036	0.005
2006	-0.012	0.008	0.021	0.005
2007	-0.019	0.009	0.030	0.005
2008	-0.020	0.008	0.032	0.005
2009	-0.102	0.025	0.051	0.005
2010	-0.004	0.003	0.033	0.005
2011	0.006	0.001	0.034	0.005
平均值	-0.015	0.016	0.031	0.010

将各年龄段对应的内禀增长率 k 分别代入式(4)，得到各年龄段对应的人口预测模型，列入下文的表 4 第 3 列。

利用表 4 中 65 岁以上的模型 $p(t) = p_{\alpha_{65\text{以上}}} e^{(t-t_0)0.031}$ 预测 2012 年与 2013 年 65 岁以上人口数量，列入表 5 第 3 列。并与 2012 年与 2013 年 65 岁以上人口数量的实际值作比较，得到相对误差。

由表 5 可得，平均相对误差仅为 0.58%，所以本文建立的预测模型有较好的效果。

表 4 不同年龄段人口预测模型

年龄段	k	模型预测
总人口	0.010	$p(t) = p_{0\text{总}} e^{(t-t_0)0.010}$
0-14 岁	-0.015	$p(t) = p_{\alpha_{0-14}} e^{(t-t_0)(-0.015)}$
15-64 岁	0.016	$p(t) = p_{\alpha_{15-64}} e^{(t-t_0)0.016}$
65 岁以上	0.031	$p(t) = p_{\alpha_{65\text{以上}}} e^{(t-t_0)0.031}$

表 5 2012-2013 年 65 岁以上人口的预测结果与实际值比较

年份	实际统计值	预测值	相对误差%	平均误差%
2012	12714	12810.96	0.76	0.58
2013	13161	13214.32	0.41	

3.2.2 人口预测

根据表 4 的预测模型,用 2013 年的数据作为初始数据,计算 2015-2025 年总人口、0-14 岁人口、15-64 岁人口、65 岁以上人口的预测值。

例如:计算 2020 年 0-14 岁人口的预测值,将 $t_0=2013$ 、 $p_0=22329$ 代入表 4 中 0-14 岁的人口预测模型得 $p(2020)=20103.35$,列入表 6 中;其他每年各年龄段的预测值同理。

运用 Malthus 模型,本文主要以 65 岁及以上人口数和占总人口比重作为研究老龄化的指标,预测 2015-2025 年的老龄化情况,预测在 2015 年总人口为 138820.84 万人,到 2025 年达到 153420.75 万人,说明人口增长呈现较稳定的状态。同时 65 岁以上人口数量不断增多,而且所占总人口的比重也不断上升并且超过了 7%,已达到老龄化标准。预测出在 2015 年中国 65 岁以上老龄人口超过 10%,到 2025 年 65 岁以上老年人口将高达 12.4%,65 岁以上人口比重不断上升,说明老龄化情况将越来越严重。

表 6 人口老龄化预测模型对 2015-2030 年人口情况的预测结果(万人)

年份	0-14 岁	15-64 岁	65 岁以上	总人口	65 岁以上人口所占比例
2015	21669.08	103852.68	14002.81	138820.84	0.101
2016	21346.47	105527.68	14443.69	140216.01	0.103
2017	21028.66	107229.71	14898.46	141625.20	0.105
2018	20715.58	108959.18	15367.55	143048.56	0.107
2019	20407.17	110716.55	15851.40	144486.22	0.110
2020	20103.35	112502.26	16350.49	145938.33	0.112
2021	19804.05	114316.77	16865.29	147405.04	0.114
2022	19509.20	116160.55	17396.31	148886.48	0.117
2023	19218.75	118034.07	17944.04	150382.82	0.119
2024	18932.62	119937.80	18509.02	151894.19	0.122
2025	18650.75	121872.24	19091.78	153420.75	0.124

为了更好的说明问题,将 65 岁以上的老年人所占的比率绘制成散点图如下:

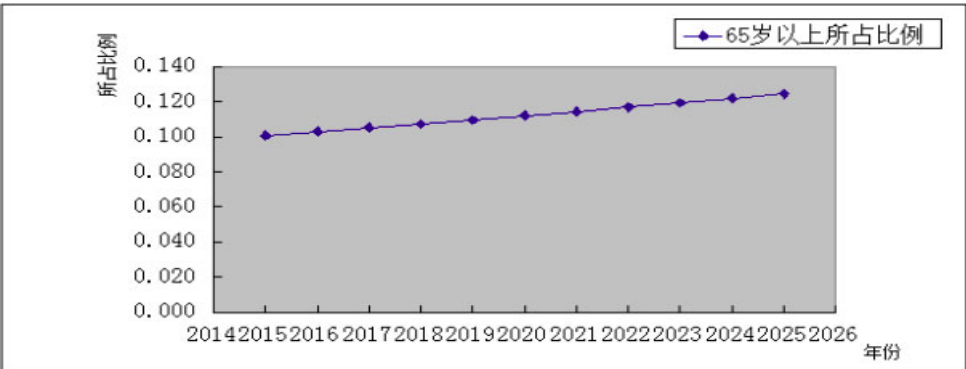


图 4 2015-2025 年 65 岁人口占总人口比例散点图

由图 4 更直观看出未来数年我国 65 岁以上老年人所占总人口比例呈直线上升趋势,人口老龄化问题十分严重。这势必会给我国政府带来承重的财政负担,影响我国经济、政治、文化等各个领域的发展,因此,解决我国人口老龄化问题迫在眉睫。但是在解决问题之前必须要清楚地认识到引起我国老龄化的重要因素是哪些?

3.3 影响我国老龄化的因素分析

引起人口老龄化最直接的因素是出生率和死亡率,出生率下降会导致人口年龄金字塔的底部变窄即相对老龄化,而死亡率下降则会导致金字塔的顶部增宽即绝对老龄化。但仅仅通过出生率和死亡率来研究人口老龄化太过宏观,因为这两个因素本身是由众多其它间接因素所决定的,所以本文从出生率和死亡率这两个方面出发,从微观的角度对影响人口老龄化的间接影响因素进行分析研究。借鉴国内外相关文献^{[4][8-18]}选取了教育经费、人均国内生产总值、政府和社会卫生支出、人口密度等 4 个影响出生率和死亡率的因素进行定量分析,所选取的因素对人口老龄化的影响详见上文表 1。

3.3.1 样本的选取与数据来源

我国在 1983 年实行计划生育政策以来,人口出生率明显下降,这可能是导致我国进入老龄化社会的原因所在。为此,本文选取了 1982 年至 2013 年的相关数据进行影响老龄化因素的分析。

我国 1998-2013 年教育经费、人均国内生产总值、政府和社会卫生支出、人口密度等数据均来源于 1983-2014 年《中国统计年鉴》^[5],数据列入表 7。

表 7 影响老龄化因素的相关数据

年份	老年人口 比重(%)	教育经费 (亿元)	人均国内 生产总值 (元)	政府和社会 卫生支出 (亿元)	人口密度(人 /平方公里)
1982	4.91	725.82	528	177.53	106
1983	4.43	485.75	583	207.42	107
1984	4.72	619.14	695	242.07	109
1985	4.98	769.6	858	279.00	110
1986	5.14	879.84	963	315.90	112
1987	5.4	1093.66	1112	379.58	114
1988	5.56	1250.32	1366	488.04	116
1989	5.6	1292.87	1519	615.50	117
1990	5.57	1260.82	1644	747.39	119
1991	5.99	1806.77	1893	893.49	121
1992	6.16	867.05	2311	1096.86	122
1993	6.15	1059.94	2998	1377.78	123
1994	6.36	1488.78	4044	1761.24	125
1995	6.2	1877.95	5046	2155.13	126

1996	6.41	2262.34	5846	2709.42	127
1997	6.54	2531.73	6420	3196.71	129
1998	6.7	2949.06	6796	3678.72	130
1999	6.9	3349.04	7159	4047.50	131
2000	7.0	3849.08	7858	4586.63	132
2001	7.1	4637.66	8622	5025.93	133
2002	7.3	5480.03	9398	5790.03	134
2003	7.5	6208.27	10542	6584.10	135
2004	7.6	7242.60	12336	7590.29	135
2005	7.7	8418.84	14185	8659.91	136
2006	7.9	9815.31	16500	9843.34	137
2007	8.1	12148.07	20169	11573.97	138
2008	8.3	14500.74	23708	14535.40	138
2009	8.5	16502.71	25608	17541.92	139
2010	8.9	19561.85	30015	19980.39	140
2011	9.1	23869.29	35198	24345.91	140
2012	9.4	27695.97	38459	28119.00	141
2013	9.7	31455.62	41908	31668.95	141

3.3.2 回归模型的建立

符号说明： x_1 表示教育费用的支出， x_2 表示人均国内生产总值， x_3 表示卫生费用支出， x_4 表示人口密度， y 表示老年人口比重（65岁以上人口所占总人口的比重）。

(1) x_1 - x_4 与 y 是否呈现线性关系？

在做多元线性回归时需检验自变量与因变量之间是否存在线性关系。为此需要判断自变量教育经费、人均国内生产总值、政府和社会卫生支出、人口密度分别与老年人口比重是否呈线性关系。根据表 7 中的数据，利用 R 软件做出教育经费、人均国内生产总值、政府和社会卫生支出、人口密度分别与老年人口比重的散点图，即图 5。观察 4 个影响因素分别与老年人口比重是否呈线性相关关系。

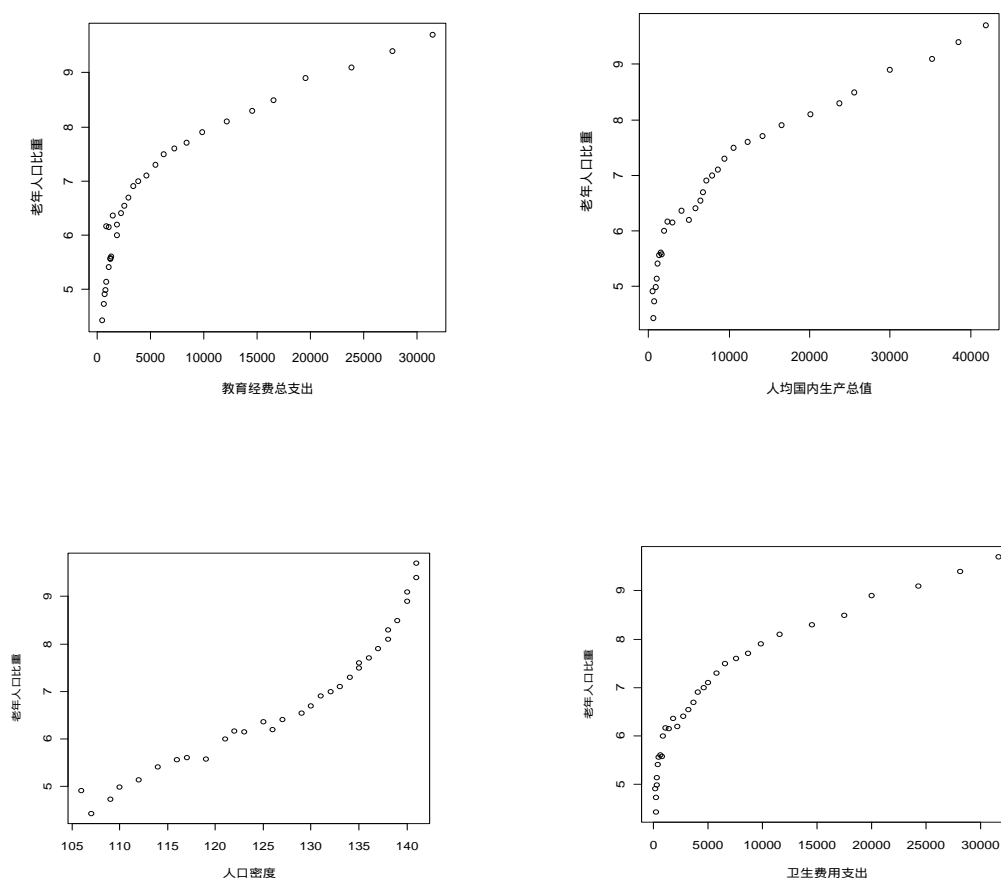


图 5 影响因素与老年人口比重散点图

从上面的散点图可以看出：教育经费支出、人均国内生产总值、卫生费用总支出和人口密度对人口老龄化均具有正向影响。人口密度与老年人口比重呈线性相关关系，教育经费支出、人均国内生产总值和卫生费用总支出与老年人口比重的线性相关性较差。

(2) 对自变量做怎样的转换可使其与因变量呈线性关系？

正如上文所述，教育经费支出、人均国内生产总值和卫生费用总支出与老年人口比重的线性相关性较差。本文将教育经费支出、人均国内生产总值和卫生费用总支出做对数变换，并利用 SPSS 分别与老年人口比重做拟合。从拟合结果可以看出拟合程度 R^2 均大于 0.95，拟合程度较高。

其中，原假设 H_0 为 b_1 为 0，即因变量与自变量无关，显著性检验的 p 值为 $0.000 < 0.05$ ，故拒绝原假设，说明老年人口比重与四个自变量具有线性相关关系，所以本文可以认为教育费用的对数、人均国内生产总值的对数、卫生总支出的费用的对数和人口密度分别与人口老龄化比重呈线性关系，表 8 是我们利用 SPSS 拟合得出的数据结果，为了更加清晰直观的看出拟合效果，还绘制了拟合后的曲线图像见图 6。结果如下：

表 8 曲线拟合程度汇总表

方程	模型汇总					参数估计	
	R 方	F	df1	df2	sig	常数	b_1

log(x ₁)与 y	.969	943.174	1	30	.000	-2.381	2.612
log(x ₂)与 y	.959	943.174	1	30	.000	-2.167	2.411
log(x ₃)与 y	.960	717.915	1	30	.000	-.154	2.034
x ₄ 与 y	.914	320.281	1	30	.000	-8.937	.124

(3)各因子间是否相互独立？

表9 各因子间的相关系数矩阵

因子	X1	X2	X3	X4
X1	1.000	-0.309	-0.524	0.627
X2	-0.309	1.000	-0.281	-0.419
X3	-0.524	-0.281	1.000	-0.232
X4	0.627	-0.419	-0.232	1.000

由表9可知，各因子间的相关系数均小于0.6，可以认为影响人口老龄化的因素，即教育费用、人均国内生产总值、卫生费用总支出和人口密度之间是相互独立的。满足多元线性回归的前提条件。

(4)回归模型的建立

由上述的曲线拟合知道与log x₁、log x₂、log x₃ 与x₄分别和y呈线性关系，由此建立多元线性回归模型：

$$y=c+a_1*\log(x_1)+a_2*\log(x_2)+a_3*\log(x_3)+a_4*x_4+e \quad (7)$$

其中c为常数项；a₁~a₄表示因变量y对自变量log(x₁)~log(x₃) 与x₄的回归系数；e为残差。

利用SPSS逐步回归法,即对全部因子按其对y影响程度的大小，从大到小地依次逐个地引入回归方程，在引入的因子中选出对y的影响显著性大的，对y显著性小的剔除掉，进而得出多元线性回归模型^[7]：

$$y=-1.559+1.511*\log(x_1)+0.892*\log(x_3) \quad (8)$$

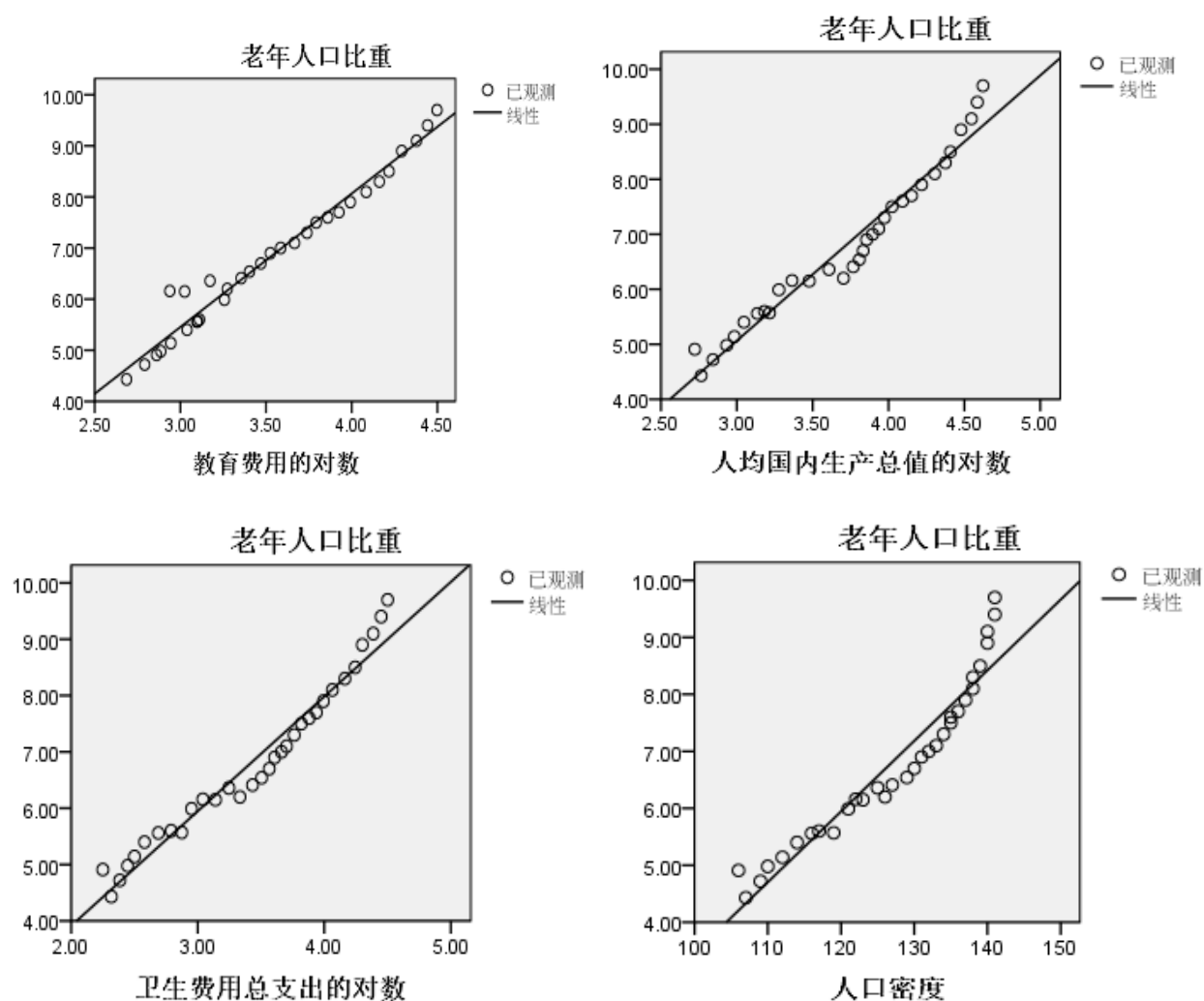


图 6 影响因素对数与老年人口比重拟合效果图

表 10 回归系数的显著性检验

模型	非标准化系数		标准系数	t	sig.
	b	标准误差	试用版		
(常量)	-1.559	.304		-5.137	.000
教育费用的对数	1.511	.260	.569	5.806	.000
卫生费用支出对数	.892	.204	.430	4.381	.000

a. 因变量: 老年人口比重.

表10给出了回归方程中的参数和常数项的估计值。其中常数项系数的估计值为-1.559，回归系数为1.511和0.892，由此建立了两个自变量，教育费用的对数 $\log(x_1)$ 与卫生费用总支出的对数 $\log x_2$ 与因变量老年人口比重 y 的具体关系式，见式(8)。

3.3.3 回归模型的检验

为了检验模型的显著性，我们分别进行了t检验、拟合优度检验、F检验。

t检验即回归系数的显著性检验的目的是为了检验某个相关系数是否为0，对于原假设 H_0 为 $a_j=0$ ，回归系数t检验的t统计量观察值为5.806和4.381，均小于 $t_{0.99}(30)=2.4573$ ，落在了拒绝域 $|t| > t_{0.99}(30)$ ，而且假设性检验的p值均为0.000小于0.01，所以拒绝原假设，可以认为回归系数有显著意义。

在拟合优度检验中(表11)，因变量y与多个自变量之间的复相关系数R为0.991，决定系数 R^2 为0.981，又因为 R^2 越接近1，模型拟合越好，说明y与多个因子的线性相关程度较高，同时调整决定系数为0.980，消除了自变量个数不同的影响。综上，说明 x_1 与 x_3 对老年人口比重影响较为明显，相关程度较高，模型拟合程度较好。

表 11 R^2 检验模型汇总

模型	R	R方	调整R方	标准估计的误差
2	.991 ^a	.981	.980	.20160

a.预测变量：（常量），教育费用的对数，卫生费用总支出的对数。

进一步由以下的回归方程的多因子方差分析表(表12)知道回归平方和为62.365，残差平方和为1.179，剩余均方为0.041，说明组间偏差平方和大于组内偏差平方和，即偏差主要由各因子的不同引起的。F统计量检验的原假设 H_0 为 $a_1=a_2=0$ ，即因变量的相关系数相等且均为0，假设性检验的P值为 $0.000 < 0.05$ ，所以拒绝原假设，可以认为老年人口比重与教育费用总支出的对数和卫生费用总支出的对数这两个因子所组成的回归方程是显著的。

表 12 多因子方差分析表

模型	平方和	df	均方	F	sig
回归	62.365	2	31.183	767.222	.000 ^b
1 残差	1.179	29	.041		
总计	63.544	31			

a.因变量\：老年人口比重。

b.预测变量：（常量），预测变量：（常量），教育费用的对数，卫生费用总支出的对数。

在回归标准化残差的直方图中（图7），标准化残差呈正态分布，不存在极端值。除此之外，还可以从回归标准化的正态p-p图中（图8）可以看到观测值的残差分布与假设的正态分布相比较，标准化残差点大体分布在直线附近，更进一步说明标准化残差呈正态分布。

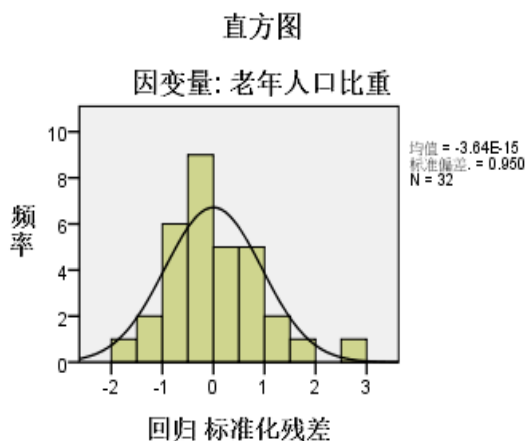


图7 残差分布直方图

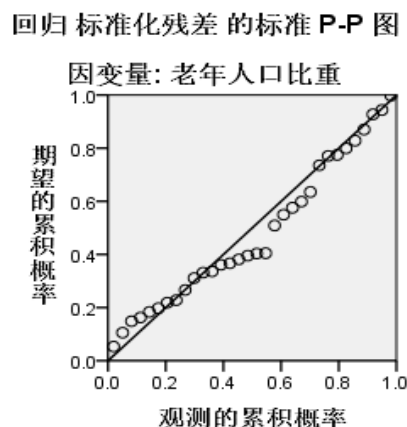


图8 p-p概率图

3.3.4 模型的应用价值

前面已经对模型的合理性进行了理论上的检验,接下来要用模型来预测2012-2013年的老龄人口比重,并与2012-2013年老龄人口比重的实际数值比较,从而检验模型的真实性和预测能力。

表 13 2012 年老年人口预测值与实际值比较

年份	教育费用总支出(亿元)	政府和社会卫生总支出(亿元)	教育费用总支出对数	政府和社会卫生总支出对数	老年人口比重预测值	老年人口比重实际值	绝对误差	相对误差
2012	27695.97	28119.00	4.44	4.45	9.12	9.4	0.28	0.023
2013	31455.62	31668.95	4.50	4.50	9.32	9.7	0.38	0.039

从预测结果可以看出,模型的预测值与实际值相差很小,可以认为该模型具有很好的预测价值,可以应用于今后我国人口老龄化程度的预测。

有学者用灰色预测模型对我国的人口老龄化趋势进行预测,但其不能够反映自变量与因变量的具体关系,当某个自变量出现极端值时,就会导致预测的结果与实际值偏差较大。而本文建立的多元线性回归模型,预测的结果更加接近真实值,不会由于极端值的出现而使预测结果与真实结果出现较大偏差,可以很好的反映自变量与因变量的具体关系。

4 结束语

本文利用Malthus模型对我国的人口老龄化进行了预测,发现我国的人口老龄化现象日趋严重,由此带来的一系列后果势必会影响我国社会经济的发展,同时也会给社会的发展带来一系列的问题。分别以教育经费的支出、人口密度、国内生产总值、卫生费用总支出为自变量,以老年人口比重为因变量,建立多元线性回归模型,对我国人口老龄化的影响因素进行了分析。研究表明:一是教育费用的总支出和政府社会卫生总支出是影响我国老年人口比重最重要的因素;二是由教育费用的总支出和政府社会卫生总支出所建立的回归模型对预测我国未来老年人口比重具有较高的精度。

参考文献

- [1] 蒋正华. 中华人口老龄化现象及对策[J]. 求是. 2005, (06).
- [2] 中国人口网
[DB/OL] http://www.china.com.cn/aboutchina/zhuanti/zgrk/node_7042806.htm 2015-06.
- [3] Frank R. Giordano, William P. Fox, Steven B. Horton 著. 叶其孝, 姜启源译. 数学建模 (原书第 4 版) [M]. 出版社: 机械工业出版社, 2009: 289-291.
- [4] 陈仁爱, 刘婷. 基于多元线性回归模型的人口老龄化影响因素研究[J]. 科技视界, 2012, (11).
- [5] 中国统计年鉴[DB/OL] <http://www.stats.gov.cn/tjsj/nds/> 2015-06.
- [6] 薛薇著. SPSS 统计分析方法及应用 (第 3 版) [M]. 电子工业出版社, 2013.
- [7] 高祥宝, 董寒青. 数据分析与 SPSS 应用[M], 北京: 清华大学出版社, 2007: 345-350.
- [8] 任强. 沃夫冈人口密度和生育率: 一项探索性分析[J]. 中国人口科学, 2003, (05).
- [9] 刘向红; 李红英. 影响我国人口老龄化的社会经济因素分析[J]. 期刊论文-科技情报开发与经济, 2007, (06).
- [10] 袁俊; 吴殿廷. 吴铮争中国农村人口老龄化的空间差异及其影响因素分析[J]. 期刊论文-中国人口科学, 2007, (03).
- [11] 李培; 邓慧慧. 京津冀地区人口迁移特征及其影响因素分析[J]. 期刊论文-人口与经济, 2007, (06).
- [12] Wu, Cangping, He Ling and Sun Huifeng. "The Theoretical and Practical Significance of the Proposition 'Getting Old before Getting Rich'" ("未富先老" 命题提出的理论价值和现实意义). *Population Research*, 2007, no. 4.
- [13] Yin, Weimin. "Prospects for the Development of the Social Security System in China" (展望中国社会保障事业发展前景), August 23, 2010. Retrieved from http://news.xinhuanet.com/politics/2010-08/23/c_12476236_2.htm, January 25, 2011.
- [14] Zhang, Kaidi, Guo Ping and Wang Haitao. "Reflections on Studies of Population Aging" (对人口老龄化研究的反思). In *Selected Papers Delivered at the Symposium on Dealing with the Challenges of an Aging Society* (应对老龄化社会的挑战研讨会论文), ed. Shanghai Forum. Shanghai: May 2010.
- [15] Zhang, Xiulan and Xu Yuebin. "Constructing China's Development-oriented Family Policy" (建构中国的发展型家庭政策). *Zhongguo Shehui Kexue*, 2003, no. 6.

附录

R 软件画散点图的代码

附录 1 教育费用总支出与老年人口比重的散点图：

```
教育费用总支出=c(725.82,485.75,619.14,769.6,879.84,1093.66,1250.32,1292.87,1260.82,
1806.77,867.05, 1059.94, 1488.78, 1877.95, 2262.34, 2531.73, 2949.06, 3349.04,
3849.08, 4637.66, 5480.03, 6208.27, 7242.60, 8418.84, 9815.31, 12148.07, 14500.74,
16502.71, 19561.85, 23869.29, 27695.97, 31455.62)

老年人口比重=c(4.91,4.43,4.72,4.98,5.14,5.4,5.56,5.6,5.57,5.99,6.16,6.15,6.36,
6.2,6.41,6.54,6.7,6.9,7.0,7.1,7.3,7.5,7.6,7.7,7.9,8.1,8.3,8.5,8.9,9.1,9.4,9.7)

plot(教育费用总支出,老年人口比重)
```

附录 2 人均生产总值与老年人口比重的散点图：

```
人均国内生产总值=c(528,583,695,858,963,1112,1366,1519,1644,1893,2311,2998,4044,
5046,5846,6420,6796,7159,7858,8622,9398,10542,12336,14185,16500,20169,23708,
25608,30015,35198,38459,41908)

老年人口比重=c(4.91,4.43,4.72,4.98,5.14,5.4,5.56,5.6,5.57,5.99,6.16,6.15,6.36,
6.2,6.41,6.54,6.7,6.9,7.0,7.1,7.3,7.5,7.6,7.7,7.9,8.1,8.3,8.5,8.9,9.1,9.4,9.7)

plot(人均国内生产总值,老年人口比重)
```

附录 3 政府与社会卫生总支出与老年人口比重的散点图：

```
政府与社会卫生支出=c(177.53,207.42,242.07,279.00,315.90,379.58,488.04,615.50,747.39,
893.49,1096.86,1377.78,1761.24,2155.13,2709.42,3196.71,3678.72,4047.50,4586.63,
5025.93,5790.03,6584.10,7590.29,8659.91,9843.34,11573.97,14535.40,17541.92,
19980.39,24345.91,28119.00,31668.95)

老年人口比重=c(4.91,4.43,4.72,4.98,5.14,5.4,5.56,5.6,5.57,5.99,6.16,6.15,6.36,
6.2,6.41,6.54,6.7,6.9,7.0,7.1,7.3,7.5,7.6,7.7,7.9,8.1,8.3,8.5,8.9,9.1,9.4,9.7)

plot(人均国内生产总值,老年人口比重)
```

附录 4 人口密度与老年人口比重的散点图：

```
人口密度=c(106,107,109,110,112,114,116,117,119,121,122,123,125,126,127,129,130,
131,132,133,134,135,135,136,137,138,138,139,140,140,141,141)

老年人口比重=c(4.91,4.43,4.72,4.98,5.14,5.4,5.56,5.6,5.57,5.99,6.16,6.15,6.36,
6.2,6.41,6.54,6.7,6.9,7.0,7.1,7.3,7.5,7.6,7.7,7.9,8.1,8.3,8.5,8.9,9.1,9.4,9.7)

plot(人口密度,老年人口比重)
```

附录 5 1993-2013 年人口结构数据

年份	总人口	男	所占比重%	女	所占比重%
1993	118517	60472	51.02	58045	48.98
1994	119850	61246	51.1	58604	48.9
1995	121121	61808	51.03	59313	48.97
1996	122389	62200	50.82	60189	49.18
1997	123626	63131	51.07	60495	48.93
1998	124761	63940	51.25	60821	48.75
1999	125786	64692	51.43	61094	48.57
2000	126743	65437	51.63	61306	48.37
2001	127627	65672	51.46	61955	48.54
2002	128453	66115	51.47	62338	48.53
2003	129227	66556	51.50	62671	48.50
2004	129988	66976	51.52	63012	48.48
2005	130756	67375	51.53	63381	48.47
2006	131448	67728	51.52	63720	48.48
2007	132129	68048	51.5	64081	48.5
2008	132802	68357	51.47	64445	48.53
2009	133450	68646	51.44	64803	48.56
2010	134091	68748	51.27	65342	48.73
2011	134735	69068	51.26	65667	48.74
2012	135404	69395	51.25	66009	48.75
2013	136072	69728	51.24	66344	48.76