

厦门市机动车保有量的预测方法

张东军¹, 邓丽娟¹, 马龙俊¹, 姜伟²

(1. 集美大学航海学院, 福建 厦门 361021; 2. 重庆交通大学交通运输学院, 重庆 400074)

[摘要] 为了预测厦门市未来5年、15年的机动车保有量, 以厦门市历年机动车保有量为研究对象, 选取地区GDP、财政总收入、人均可支配收入、燃料零售价格指数、常住人口、公路通车里程等6个影响较大的数据指标. 采用传统的非线性回归方法, 主成份分析与Logistic模型相结合的方法, 综合考虑两种预测方法及其结果, 预测厦门市2020年、2030年的机动车保有量分别约为220万辆、530万辆.

[关键词] 机动车保有量; 非线性回归; 因子分析; Logistic模型

[中图分类号] U 491.14

Forecasting Method of Motor Vehicle Population in Xiamen

ZHANG Dong-jun¹, DENG Li-juan¹, MA Long-jun¹, JIANG Wei²

(1. Navigation Institute, Jimei University, Xiamen 361021, China; 2. School of Traffic & Transportation, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: In order to forecast motor vehicle population in Xiamen over the next 5 and 15 years, this paper takes the motor vehicle populations in recent years in Xiamen as objectives, then select six major influential factors that contain regional GDP, financial gross income, per capita disposable income, fuel retail price index, resident population as well as traffic mileage of highways. Using the traditional method of nonlinear regression, and the method of combining principal component analysis with Logistic model, the two methods and their outcomes are considered, and the motor vehicle populations for 2020 and 2030 are forecast to be 2.2 m and 5.3 m, respectively.

Keywords: motor vehicle population; nonlinear regression; factor analysis; Logistic model

0 引言

随着社会经济的发展, 人们的生活越来越离不开交通, 交通需求的增长使得机动车的保有量日益增长, 机动车在带给人民便利的同时, 也带来了一系列的问题: 交通拥挤, 大气污染, 交通事故等. 机动车保有量的预测可为政府制定经济发展规划提供参考, 对实施节能环保政策, 提升居民生活品质有重要意义^[1].

机动车保有量的预测方法有回归分析法、趋势外推法、人工神经网络及Logistic分析法^[2]. 前两种预测方法相对比较简单, 人工神经网络作为一类高级机器学习算法, 尚有许多不成熟的地方, 在样本学习过程中可能存在学习不足或过度学习的不稳定性. 目前, 作为基本预测算法加以利用的是非线性

性回归方法, 这种方法较简单, 特别是在样本数量充足的情况下. 近年, 还有学者利用降维的思想, 运用主成份分析 (因子分析) 与 Logistic 模型相结合的方法, 去求解一些预测性问题^[3]. 本文分别运用传统的非线性回归方法和主成份分析与 Logistic 模型结合的方法^[4], 对 2020 年、2030 年厦门市的机动车保有量分别进行预测, 并对两种方法所得出的预测值进行比较, 从而确定厦门市未来 5 年、15 年的机动车保有量情况.

1 指标的选取与相关数据

一个地区或城市的机动车保有量受到城市人口及城市社会经济发展水平及需求等因素的影响^[5]. 本文结合实际情况, 选取了 6 个影响指标, 并统计了 2001 年—2014 年厦门市机动车保有量及 6 个影响指标的变化情况, 见表 1.

表 1 2001—2014 年厦门市机动车保有量及相关数据
Tab.1 The motor vehicle population and related data inXiamen from 2001 to 2014

年份 Year	地区 GDP Regional GDP/万元	财政总收入 Fiscalgross income /万元	人均可支配 Per capita disposable income/元	燃料零售价 格指数 Fuel retail price index (2001 年 = 100)	常住人口 Resident population /万人	公路通车 里程 Traffic mileage of highway /km	机动车保 有量 Motor vehicle population /辆
2001	5 018 706	914 984	10 812.96	100.0	205	1 051.00	187 789
2002	5 583 268	1 105 019	11 364.96	101.7	219	1 456.00	207 433
2003	6 483 570	1 263 092	11 767.68	99.1	232	1 459.00	236 935
2004	7 596 934	1 492 249	12 915.12	109.0	245	1 469.00	270 719
2005	8 877 149	1 603 600	14 442.67	120.7	258	1 493.00	314 823
2006	10 065 830	2 097 252	16 402.75	136.6	273	1 516.99	367 136
2007	11 737 984	2 752 266	18 513.17	158.9	288	1 704.46	433 116
2008	14 025 849	3 484 363	21 502.58	167.0	304	1 741.44	507 009
2009	16 107 098	4 101 378	23 947.57	194.1	326	1 836.58	568 474
2010	17 372 349	4 514 073	26 130.58	176.2	330	1 873.03	632 179
2011	20 600 738	5 260 215	29 253.14	202.3	356	1 908.22	714 784
2012	25 393 132	6 518 284	33 565.31	229.2	361	1 916.36	801 161
2013	28 151 706	7 394 619	37 576.05	236.8	367	1 923.59	907 190
2014	30 181 565	8 250 968	41 360.4	237.0	373	2 003.70	1 030 666

数据来源: 厦门统计年鉴 2001 年—2014 年
Source: The Statistical Yearbook in Xiamen from 2001 to 2014

2 非线性回归预测

2.1 非线性回归模型构建与验证

根据表 1 的统计数据, 绘制机动车保有量与时间的曲线关系, 如图 1 所示. 由图 1 可以看出机动车保有量 (纵坐标) 随着年度 (横坐标) 逐年急剧增加, 且曲线走势较光滑, 具有二次曲线的特征. 根据这一特点, 假定历年机动车保有量 (Y) 与时间 (X) 为二次曲线关系, 其回归方程可设为:

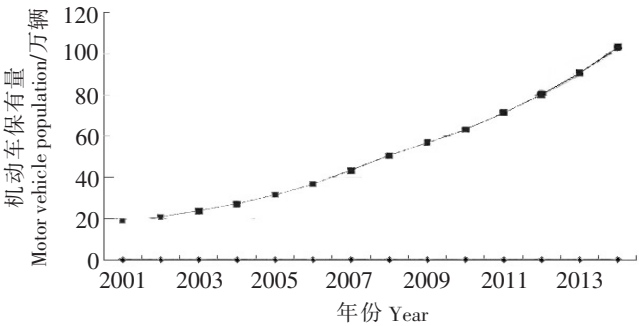


图 1 2001—2014 年厦门市机动车保有量
Fig.1 The motor vehicle population in Xiamen from 2001 to 2014

$$Y = a + b_1X + b_2X^2$$

(1)

令: $X = X_1, X^2 = X_2$, 则, 上式为:

$$Y = a + b_1X_1 + b_2X_2$$

(2)

式中的参数 a, b_1, b_2 待定. 为计算参数, 列表计算见表 2.

表 2 回归模型参数计算表

Tab.2 The calculation of regression model parameters

									万辆
年份 Year	Y	X = X ₁	X ₁ ²	X ² = X ₂	X ₂ ²	X ₁ X ₂	X ₁ Y	X ₂ Y	Y ²
2001	18.778 9	1	1	1	1	1	18.778 9	18.778 9	352.647 085 2
2002	20.743 3	2	4	4	16	8	41.486 6	82.973 2	430.284 494 9
2003	23.693 5	3	9	9	81	27	71.080 5	213.241 5	561.381 942 3
2004	27.071 9	4	16	16	256	64	108.287 6	433.150 4	732.887 769 6
2005	31.482 3	5	25	25	625	125	157.411 5	787.057 5	991.135 213 3
2006	36.713 6	6	36	36	1 296	216	220.281 6	1 321.689 6	1 347.888 425 0
2007	43.311 6	7	49	49	2 401	343	303.181 2	2 122.268 4	1 875.894 695 0
2008	50.700 9	8	64	64	4 096	512	405.607 2	3244.857 6	2 570.581 261 0
2009	56.847 4	9	81	81	6 561	729	511.626 6	4604.639 4	3 231.626 887 0
2010	63.217 9	10	100	100	10 000	1 000	632.179 0	6 321.790 0	3 996.502 880 0
2011	71.478 4	11	121	121	14 641	1 331	786.262 4	8648.886 4	5 109.161 667 0
2012	80.116 1	12	144	144	20 736	1 728	961.393 2	11536.718 4	6 418.589 479 0
2013	90.719 0	13	169	169	28 561	2 197	1 179.347 0	15 331.511 0	8 229.936 961 0
2014	103.066 6	14	196	196	38 416	2 744	1 442.932 4	20 201.053 6	10 622.724 040 0
合计 Total	717.941 4	105	1 015	1 015	127 687	11 025	6 839.855 7	74 868.615 9	46 471.242 800 0

根据交通运输系统工程回归模型的相关计算方法^[6], 结合表 2 中数据, 可以确定回归预测模型中的系数:

$$L_{11} = \sum_{i=1}^{14} (X_{1i} - \overline{X_1})^2 = \sum_{i=1}^{14} X_{1i}^2 - \frac{(\sum_{i=1}^{14} X_{1i})^2}{n} = 1015 - (105 \times 105)/14 = 227.5;$$

$$L_{22} = \sum_{i=1}^{14} (X_{2i} - \overline{X_2})^2 = \sum_{i=1}^{14} X_{2i}^2 - \frac{(\sum_{i=1}^{14} X_{2i})^2}{n} = 127687 - (1015 \times 1015)/14 = 54\,099.5;$$

$$L_{12} = L_{21} = \sum_{i=1}^{14} (X_{1i} - \overline{X_1})(X_{2i} - \overline{X_2}) = 11025 - (105 \times 1015)/14 = 3\,412.5;$$

$$L_{Y1} = \sum_{i=1}^{14} (X_{1i} - \overline{X_1})(Y_i - \bar{Y}) = 6839.8557 - (105 \times 717.9414)/14 = 1\,455.2952;$$

$$L_{Y2} = \sum_{i=1}^{14} (X_{2i} - \overline{X_2})(Y_i - \bar{Y}) = 74868.6159 - (1015 \times 717.9414)/14 = 23\,035.364\,4;$$

$$L_{YY} = \sum_{i=1}^{14} (Y_i - \bar{Y})^2 = 46471.2428 - 717.9414^2/14 = 9\,654.1104.$$

再将上述计算结果代入式:

$$\begin{cases} L_{11}b_1 + L_{21}b_2 = L_{Y1}; \\ L_{21}b_1 + L_{22}b_2 = L_{Y2}. \end{cases} \tag{3}$$

得,
$$\begin{cases} 227.5b_1 + 3\,412.5b_2 = 1\,455.2952, \\ 3\,412.5b_1 + 54\,099.5b_2 = 23\,035.364\,4. \end{cases}$$

解上述方程组得: $b_1 = 0.185\,004\,313\,186\,813\,2$; $b_2 = 0.414\,126\,510\,989\,011$.

根据公式 (2), 则可将其变形为: $a = \bar{Y}_1 - b_1 \bar{X}_1 - b_2 \bar{X}_2$.

根据表 2, 可计算 \bar{Y}_1 为 51.281 528 57, \bar{X}_1 为 7.5, \bar{X}_2 为 72.5, 则:

$$a = 51.281\,528\,57 - 0.185\,004\,313\,186\,813\,2 \times 7.5 - 0.414\,126\,510\,989\,011 \times 72.5 = 19.869\,824\,174\,395\,603\,5.$$

将参数 a, b_1, b_2 的值带入公式 (1), 即可得到所求回归方程:

$$Y = 19.869\,824\,174\,395\,603\,5 + 0.185\,004\,313\,186\,813\,2X + 0.414\,126\,510\,989\,011X^2 \tag{4}$$

2.2 非线性回归模型验证

利用所求的回归方程, 计算理论预测值, 并与实际 Y 值进行比较, 见表 3.

表 3 机动车保有量理论值与实际值比较
Tab.3 Theoretical value and actual value of motor vehicle population

万辆

年份 Year	实际保有量 Actual value	理论保有量 Theoretical value	年份 Year	实际保有量 Actual value	理论保有量 Theoretical value
2001	18.778 9	20.469 0	2008	50.700 9	47.854 0
2002	20.743 3	21.896 3	2009	56.847 4	55.079 1
2003	23.693 5	24.152 0	2010	63.217 9	63.132 5
2004	27.071 9	27.235 9	2011	71.478 4	72.014 2
2005	31.482 3	31.148 0	2012	80.116 1	81.724 1
2006	36.713 6	35.888 4	2013	90.719 0	92.262 2
2007	43.311 6	41.457 1	2014	103.066 6	103.628 7

由表 3 可见, 用所求的回归方程式计算的预测值与实际值非常接近, 说明在其他条件不变的情况下, 此方程是可信的.

2.3 非线性回归法预测结果

如果要求厦门市 2020 年、2030 年的机动车保有量, 则分别将 $X = 20, X = 30$ 代入公式 (4), 得:

$$\begin{aligned} Y(X = 20) &= 19.869\,824\,174\,395\,603\,5 + 0.185\,004\,313\,186\,813\,2 \times 20 + 0.414\,126\,510\,989\,011 \times 20^2 \\ &= 19.869\,824\,174\,395\,603\,5 + 3.700\,086\,263\,736\,264 + 165.650\,604\,395\,604\,4 \\ &\approx 189.220\,5 \text{ (万辆)} \end{aligned}$$

$$\begin{aligned} Y(X = 30) &= 19.869\,824\,174\,395\,603\,5 + 0.185\,004\,313\,186\,813\,2 \times 30 + 0.414\,126\,510\,989\,011 \times 30^2 \\ &= 19.869\,824\,174\,395\,603\,5 + 5.550\,129\,395\,604\,396 + 372.713\,859\,890\,109\,9 \\ &\approx 398.133\,8 \text{ (万辆)} \end{aligned}$$

由此法计算可知, 厦门市 2020 年、2030 年的机动车保有量分别是 189 万辆、398 万辆.

3 主成份分析与 Logistic 模型预测

3.1 主成份分析与 Logistic 模型的构建

主成份分析是利用降维的思想, 通过原始指标的线性组合, 产生一系列互不相关的综合性指标,

从中选出少数几个综合指标，并使它尽可能多地反映原始指标的信息，进而用这较少的几项综合性指标来刻画整体^[7]。此方法可以克服多个影响因素模拟预测时收敛速度较慢的缺点。

首先，利用 SPSS. 22 软件可得到因子分析共同度^[8]，见表 4。表 4 中第二行是因子分析初始解下的变量共同度，表示对原 6 个变量如果采用主成份分析法提取所有 6 个特征根，那么原有的变量方差都可以被解释，变量的共同度均为 1。第三行列出了按指定提取条件提取特征根时的共同度。可以看到，除公路通车里程外的变量的绝大部分信息可被因子解释，即变量信息丢失较少。

对上文中确定的 6 个影响因素进行建模分析，然后利用标准化后的数据计算因素相关系数矩阵；求解相关系数矩阵 R 的特征值和特征矢量，确定主成份；选取 $m(m < 6)$ 个主成份，使得累积方差贡献率超过某一定值^[9]。

计算得到相关系数矩阵的特征值、累加方差贡献率，结果见表 5。第一个因子的特征根值为 5.762，解释了原有 6 个变量总方差的 96.027%。因此，可以只选取第一个因子作为主因子即可^[10]。其中，下文所指的“元件”是 SPSS 工具语言，而其概念与上文的“影响指标”相同。

表 4 因子分析共同度
Tab.4 Factor analysis

变量 Variable	地区 GDP Regional GDP /万元	财政总收入 Fiscal gross income /万元	人均可支配 Per capita disposable income/元	燃料零售价格 指数 Fuel retail price index (2001 年 = 100)	常住人口 Resident population /万人	公路通车里程 Traffic mileage of highway/km
起始 Origin	1.000	1.000	1.000	1.000	1.000	1.000
提取 Capture	0.978	0.977	0.979	0.979	0.974	0.874

提取方法：主体元件分析。
Method: principal component analysis.

表 5 主成份分析法所得影响因子累积贡献率
Tab.5 Cumulative contribution rate of impact factors

元件 Component	起始特征值 Starting eigenvalues			元件 Component	起始特征值 Starting eigenvalues		
	总计 Total	变异的 Variant /%	累加 Summation /%		总计 Total	变异的 Variant /%	累加 Summation /%
1	5.762	96.027	96.027	4	0.010	0.174	99.961
2	0.189	3.143	99.170	5	0.002	0.032	99.993
3	0.037	0.617	99.788	6	0.000	0.007	100.000

提取方法：主体元件分析。
Method: principal component analysis.

在 SPSS 软件中输入指令，则可得到因子碎石图，如图 2 所示。由图 2 可见，第一个特征值明显大于后面的特征值，说明提取第一个因子是合适的。

通过载荷系数大小可以分析不同公共因子的主要指标的区别，表 6 显示第一个元件的相关指标载荷系数。由于第一个元件在 6 个指标上的载荷值都很大，说明它综合反映了该地区综合经济发展值，可以作为综合经济发展值看待。

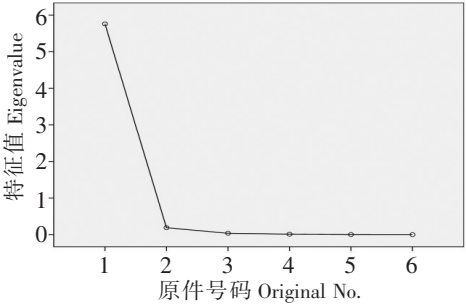


图 2 因子碎石图
Fig.2 Factor scree plot

表 6 因子载荷矩阵
Tab.6 Factor loading matrix

因素 Factor	燃料零售价格 指数 Fuel retail price index (2001 年 = 100)	人均可支配 Per capita disposable income/元	地区 GDP Regional GDP /万元	财政总收入 Fiscal gross income /万元	常住人口 Resident population /万人	公路通车里程 Traffic mileage of highway/km
载荷系数 Coefficient	0.990	0.989	0.989	0.988	0.987	0.935

擷取方法：主体元件分析。
 Method: principal component analysis.

在原数据浏览窗口中新增了变量“FAC1_1”，表示不同年份的综合经济发展值。采用主成份分析法可以得出因子得分系数。其中，模型的整体拟合优度值 R^2 为 0.981，统计量 F 为 612.040，概率 P 值小于显著性水平 0.05，说明该模型有统计学意义，利用统计学知识，得到的统计方程：

$$y = 1/(1/U + b_0 \times b_1^x)^{[10]}.$$
 (5)

其中： x 为“FAC1_1”（综合经济发展变量）值； y 为机动车保有量； U 为上限默认值 1 000； b_0, b_1 参数分别为 0.021, 0.557.

3.2 Logistic 模型的模拟与预测

输入拟合步骤，可直接得出图 3。由图 3 可见，实际数据的散点分布、线性模拟以及 Logisitic 方法的拟合情况。相比之下，Logistic 回归方程的预测曲线更能诠释各年的机动车保有量数值，另一方面也进一步说明 Logistic 方程较线性回归拟合效果更好。

选择曲线估计中的线性回归选项，得到综合经济发展变量“FAC1_1”时间序列的同时，可以直接导出综合经济发展变量与时间关系的预测方程：

$$\text{FAC1_1} = -476.52 + 0.2374 \times t,$$
 (6)
 其中 t 代表年份。

于是，可以利用公式 (6)，可分别求得 t 为 2020 和 2030 的“FAC1_1”值；

$$\text{FAC1_1}(t = 2020) = -476.52 + 0.2374 \times t = 3.028;$$

$$\text{FAC1_1}(t = 2030) = -476.52 + 0.2374 \times t = 5.402.$$

分别将对应 2020 年、2030 年的“FAC1_1”值代入公式 (5)，则：

$$y(2020) = 1/(1/1000 + 0.021 \times 0.557^x) = 1/(1/1000 + 0.021 \times 0.557^{3.028}) = 218.8191 \text{ (万辆)};$$

$$y(2030) = 1/(1/1000 + 0.021 \times 0.557^x) = 1/(1/1000 + 0.021 \times 0.557^{5.402}) = 529.1365 \text{ (万辆)}.$$
 由此方法计算可知，厦门市 2020 年、2030 年的机动车保有量分别是 219 万辆、529 万辆。

4 结论

由于机动车保有量增加会带来诸多问题，为预测机动车保有量，本文采用了非线性回归模型、主成份分析与 Logistic 相结合的方法，利用 2001 年至 2014 年的相关数据分别进行了预测。

非线性回归模型只考虑时间与厦门市机动车保有量的关系，得出了到 2020 年、2030 年的机动车保有量将分别达到 189.2205 万辆、398.1338 万辆。这种方法使用简单，需要原始数据较少，但考虑

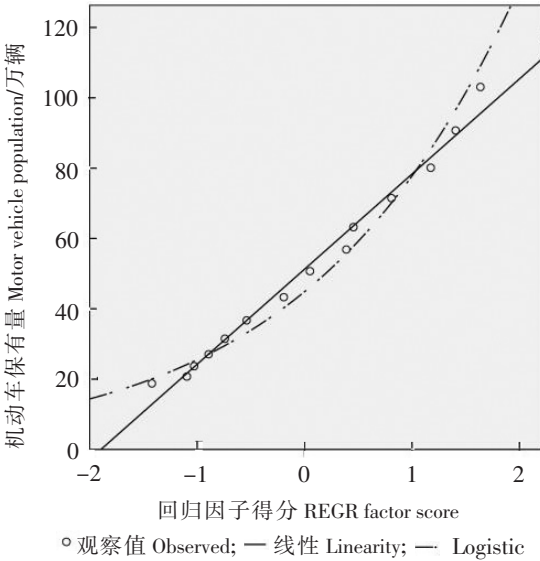


图 3 拟合图
 Fig.3 Fitting figure

的影响因素有限,对多因素影响下的问题处理效果不佳。

而采用主成份分析与 Logistic 模型相结合的方法是为了更充分的运用数据所表达的信息。先利用因子分析法,得出了因子累计贡献率、因子碎石图、因子载荷图,因子得分系数。由上述信息可以产生一个新的综合经济发展变量 FAC1_1,它相当于将 6 个影响因素进行归一。然后,对这一变量进行曲线拟合,得出模型描述表与估计;再结合 Logistic 模型得出预测方程,对机动车保有量进行预测。该法所得到的厦门市 2020 年、2030 年的机动车保有量分别是 218.8191 万辆、529.1365 万辆。Logistic 模型利用降维的方法,将众多的经济指标综合为少量的公共因子,能代表尽可能多的数据的意义,即用数据指标对复杂问题加以细化,使步骤化繁为简,模型预测精度比较高,较非线性模型而言,具有更高的可信度。

随着厦门自由贸易试验区、东南航运中心、厦门高铁等项目的落成,笔者认为机动车的保有量将与时间脱离线性关系,所以仅从时间序列考虑是不恰当的。综合比较两种方法,主成份分析与 Logistic 模型相结合的方法更能贴合现实,因此预测厦门市 2020 年、2030 年的机动车保有量分别为 220 万辆、530 万辆左右。但由于样本的选取有限,不可控(变量)因素存在,也不能对预测值给予充分地验证。所以,预测的数据准确性只能有待未来的检验。希望在以后的预测中,可以考虑其他外部条件的特殊变化,如交通政策的改变、环境的变化等,希望文中预测的结果能为政府和城市道路交通规划,以及相关部门制定相应的规划政策时提供一定的参考。

[参 考 文 献]

- [1] 徐亚丹. 基于状态趋势预测方法的城市机动车保有量预测. 科技通报, 2012, 28(9): 11-14.
- [2] 王立颖. 机动车保有量预测方法综述. 辽宁警专学报, 2015(1): 74-78.
- [3] 赖国毅. SPSS 17.0 中文版常用功能与应用实例精讲. 北京: 电子工业出版社, 2010.
- [4] 蒋艳梅, 赵文平. Logistic 模型在我国私人汽车保有量预测中的应用研究. 工业技术经济, 2010(11): 99-104.
- [5] 许伶俐. 我国汽车保有量的预测研究—基于主成份分析与协整回归模型分析. 大连: 东北财经大学, 2011.
- [6] 刘舒燕. 交通运输系统工程. 北京: 人民交通出版社, 2012.
- [7] 杨维忠. SPSS 统计分析与行业应用案例详解. 北京: 清华大学出版社, 2011.
- [8] 杨光霞. SPSS 数据统计与分析. 北京: 清华大学出版社, 2014.
- [9] 牟振华, 李美玲, 赵庆双. 基于神经网络的山东省机动车保有量预测. 山东建筑大学学报, 2009, 24(3): 229-232.
- [10] 王璐. SPSS 统计分析基础、应用与实战精粹. 北京: 化学工业出版社, 2012.

(责任编辑 陈 敏 英文审校 周云龙)