

基于我国民用汽车拥有量的主成分分析

金童, 牟唯嫣

北京建筑大学理学院, 北京

收稿日期: 2022年9月19日; 录用日期: 2022年10月9日; 发布日期: 2022年10月25日

摘要

随着我国社会主义现代化的建设蓬勃发展、国民经济良好增长, 越来越多人们的买车需求加大。本文基于主成分分析法, 对民用汽车拥有量进行定量分析, 并对未来民用汽车拥有量进行总体趋势预测。我国民用车辆的拥有量, 由量变到质变地影响民用交通规划, 使得汽油消耗且大量尾气排除, 既消耗不可再生资源又影响空气质量。则根据目前民用汽车拥有量的分析, 进行汽车数量控制, 制定相适应地环境保护政策, 不可再生资源的合理规划。根据未来民用汽车拥有量进行总体趋势预测, 可提前做好应对措施。故该分析对民用交通规划、环境保护政策、不可再生资源的合理规划、汽车数量管控, 都有着举足轻重的意义。

关键词

主成分分析, 民用汽车拥有量, 线性回归预测

Based on the Principal Component Analysis of Civil Automobile Ownership in China

Tong Jin, Weiyang Mu

School of Science, Beijing University of Civil Engineering and Architecture (BUCEA), Beijing

Received: Sep. 19th, 2022; accepted: Oct. 9th, 2022; published: Oct. 25th, 2022

Abstract

With the vigorous development of China's socialist modernization construction and the good growth of the national economy, more and more people have increased their demand for cars. Based on the principal component analysis method, this paper quantitatively analyzes the own-

ership of civil vehicles and predicts the overall trend of future civil automobile ownership. The ownership of civil vehicles in China affects civil transportation planning from quantity to quality, so that gasoline consumption and a large number of exhaust gas are eliminated, which not only consumes non-renewable resources but also affects air quality. According to the current analysis of civil automobile ownership, the number of vehicles is controlled, appropriate environmental protection policies are formulated, and reasonable planning of non-renewable resources is formulated. According to the overall trend forecast of future civilian car ownership, countermeasures can be taken in advance. Therefore, the analysis is of great significance to civil transportation planning, environmental protection policies, rational planning of non-renewable resources, and control of the number of vehicles.

Keywords

Principal Component Analysis, Civilian Car Ownership, Linear Regression Prediction

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着我国社会主义现代化建设蓬勃发展、国民经济良好增长,越来越多人们的买车需求加大。且汽车产业作为经济发展的不可或缺组成,基于民用汽车拥有量的分析,可进一步对我国总体经济见微知著。此外,该分析对民用交通规划、环境保护政策、不可再生资源的合理规划、汽车数量管控,都有着举足轻重的意义,因此,更多的研究关注到民用汽车拥有量。部分学者对中国民用汽车拥有量及相关经济变量分析。基于 Logistic 时序模型,对未来民用汽车拥有量进行预测,侧重分析了人口总数、工业增加值、钢材产量和全国居民消费水平与民用汽车拥有量之间的关系[1]。部分学者运用指数平滑法建立了汽车拥有量预测模型,侧重分析对交通拥堵、环境、能源等问题的加重的影响[2]。而本文基于主成分分析,是不同的方法进行分析,该方法应用广泛、分析效果良好。且基于线性回归分析进行趋势预测,而线性回归常作为预测手段[3] [4],预测总体趋势简单又效果良好。

2. 方法

主成分分析是一种属于多元统计分析的方法,用于分析若干个因素变量间的相关性,通过分析和探究若干个极具代表性的主成分因子来探究多个因素变量之间的内部结构[5]。主成分因子在不缺失主要信息的情况下,要具有比原始变量具有某些更优越的性能,以达到精简研究原始数据结构且抓住研究问题实质的目的。在多元统计分析中,降维的方法有很多,诸如:聚类分析、主成分分析、因子分析,本文选择主成分分析法进行分析优势显然。该方法应用广泛,可研究城市中主要污染物排放量[6]、邮政和电信业务量情况[7]等。

2.1. 数据描述

本文数据,源自《中国统计年鉴(2021年)》中章节 16-20 民用汽车拥有量的数据[8],本文选取其中九个因子进行主成分分析,选取的因子为:大型载客汽车、中型载客汽车、小型载客汽车、微型载客汽车、重型载货汽车、中型载货汽车、轻型载货汽车、微型载货汽车、其他汽车(单位:万辆)。此外,31个省或直辖市为结果分析的描述对象。

2.2. 可行性分析

本文使用 SPSS26.0 多元统计分析软件分析所用数据。先对 9 个变量进行球形检验, 由表 1, 可知本论文中的 KMO 取样适切性量数观测值为 0.700, 这个值处于 0~1 区间内, 较靠近于 1, 可以说明用这些指标主成分分析出来的结果是可信的。此外, Bartlett 球形检验的 p 值为 0, 说明可以用主成分分析法进行数据分析[9]。

Table 1. Test of KMO and Bartlett

表 1. KMO 和巴特利特检验

KMO 取样适切性量数	巴特利特球形度检验		
	近似卡方	自由度	显著性
0.700	233.814	36	0.000

2.3. 模型处理与分析

从表 2, 即公因子方差分析表中, 可以看出, 本文采取的提取方法为主成分分析法, 多数信息提取都大于等于 50.4%, 总体信息损失较小, 其中: 大型载客汽车、中型载货汽车、轻型载货汽车、其他汽车的提取程度大于等于 70.3%。

Table 2. Common factor variances

表 2. 公因子方差

因子	初始	提取
大型载客汽车	1.000	0.955
中型载客汽车	1.000	0.472
小型载客汽车	1.000	0.573
微型载客汽车	1.000	0.540
重型载货汽车	1.000	0.658
中型载货汽车	1.000	0.736
轻型载货汽车	1.000	0.873
微型载货汽车	1.000	0.483
其他汽车	1.000	0.877

从表 3, 可以认识到 1~9 个主成解释原始变量总方差的情况, 保留特征根大于 1 的主成分的基础上。据此, 本文保留了两个主成分, 这两个主成分集中了原始 9 个变量信息的 68.6%, 基本保留了原来 9 个指标的信息, 这样由 9 个指标转化为 2 个新指标, 达到了降维的目的。此外, 亦可由表 1: 即公因子方差分析表的提取度作平均得到, 即 $(0.955 + 0.472 + 0.573 + 0.540 + 0.658 + 0.736 + 0.873 + 0.483 + 0.877)/9 = 68.6\%$ 。保留两个主成分, 亦可从图 1, 即碎石图可以看出, 到第三个主成分, 直线斜率大大降低, 即保留两个主成分是合理的。总体, 可见效果较为良好[8]。

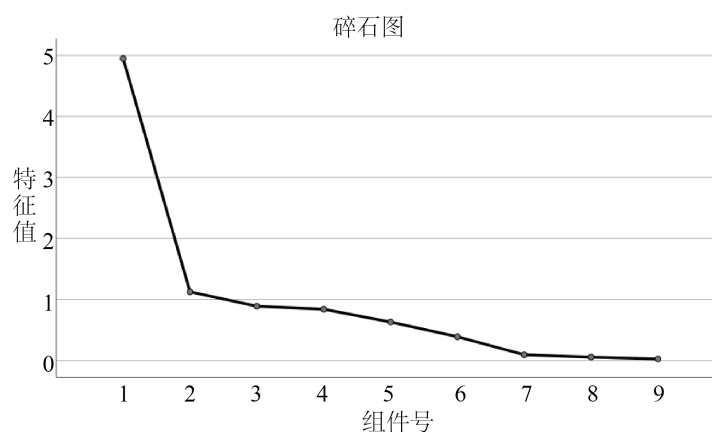


Figure 1. Gravel map

图 1. 碎石图

Table 3. Cumulative contribution rate of variance

表 3. 方差累积贡献率

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积 %	总计	方差百分比	累积 %
1	5.036	55.955	55.955	5.036	56	56
2	1.132	12.579	68.534	1.132	12.6	68.6
3	0.904	10.049	78.583			
4	0.868	9.644	88.227			
5	0.657	7.301	95.528			
6	0.251	2.792	98.320			
7	0.086	0.956	99.276			
8	0.041	0.456	99.733			
9	0.024	0.267	100.000			

由表 4, 即成分矩阵(因子载荷阵)表, 可得到基于通过求得的主成分的原始变量标准化系数, 则得到各变量的近似标准化表达式: 标准化的大型载客汽车 $\approx 0.977 * (\text{主成分 } 1) + 0.033 * (\text{主成分 } 2)$; 标准化的中型载客汽车 $\approx 0.684 * (\text{主成分 } 1) - 0.070 * (\text{主成分 } 2)$; 标准化的小型载客汽车 $\approx -0.093 * (\text{主成分 } 1) + 0.751 * (\text{主成分 } 2)$; 标准化的微型载客汽车 $\approx -0.083 * (\text{主成分 } 1) - 0.730 * (\text{主成分 } 2)$; 标准化的重型载货汽车 $\approx 0.810 * (\text{主成分 } 1) - 0.055 * (\text{主成分 } 2)$; 标准化的中型载货汽车 $\approx 0.856 * (\text{主成分 } 1) + 0.057 * (\text{主成分 } 2)$; 标准化的轻型载货汽车 $\approx 0.934 * (\text{主成分 } 1) - 0.035 * (\text{主成分 } 2)$; 标准化的微型载货汽车 $\approx 0.681 * (\text{主成分 } 1) + 0.138 * (\text{主成分 } 2)$; 标准化的其他汽车 $\approx 0.935 * (\text{主成分 } 1) - 0.044 * (\text{主成分 } 2)$ 。

Table 4. Composition matrix

表 4. 成分矩阵

	主成分 1	主成分 2
大型载客汽车	0.977	0.033
中型载客汽车	0.684	-0.070

Continued

小型载客汽车	-0.093	0.751
微型载客汽车	-0.083	-0.730
重型载货汽车	0.810	-0.055
中型载货汽车	0.856	0.057
轻型载货汽车	0.934	-0.035
微型载货汽车	0.681	0.138
其他汽车	0.935	-0.044

由图 2, 可看出小型载客汽车、微型载客汽车因子分布在组件 1 附近, 其余因子分布在组件 2 附近, 这与成分矩阵的系数是一致的。

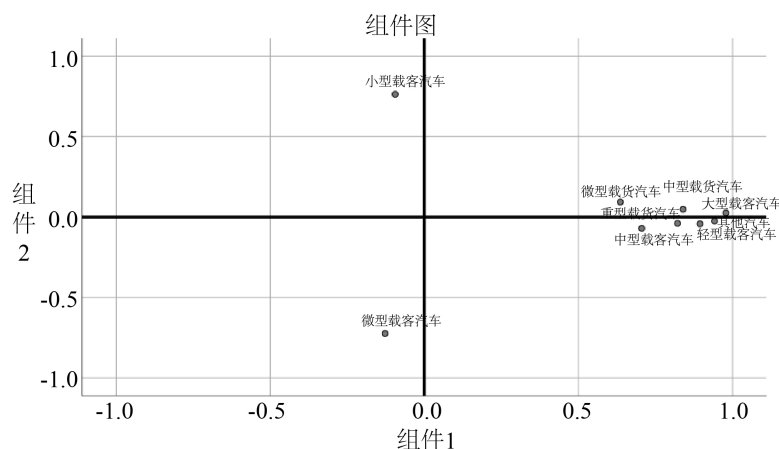


Figure 2. Factor component plot
图 2. 因子组件图

由表 4, 并不能直接得到各个主成分用标准化后的原始变量表达的表达式, 需要在此基础上, 各列除以对应的各列特征根的平方根, 由总方差解释表可知, 对应的特征根为: 5.036、1.132, 则可得到表 5, 即主成分分析的主成分的系数。

Table 5. Coefficients of principal components 1 and 2
表 5. 主成分 1 与 2 的系数

	主成分 1 系数	主成分 2 系数
大型载客汽车	0.194	0.029
中型载客汽车	0.136	-0.062
小型载客汽车	-0.018	0.663
微型载客汽车	-0.016	-0.645
重型载货汽车	0.161	-0.049
中型载货汽车	0.170	0.050
轻型载货汽车	0.185	-0.031
微型载货汽车	0.135	0.122
其他汽车	0.186	-0.039

由表 5, 得到前两个主成分的线性组合为: 主成分 1 = 0.194 * 标准化的大型载客汽车 + 0.136 * 标准化的中型载客汽车 - 0.018 * 标准化的小型载客汽车 - 0.016 * 标准化的微型载客汽车 + 0.161 * 标准化的重型载货汽车 + 0.170 * 标准化的中型载货汽车 + 0.185 * 标准化的轻型载货汽车 + 0.135 * 标准化的微型载货汽车 + 0.186 * 标准化的其他汽车。主成分 2 = 0.029 * 标准化的大型载客汽车 - 0.062 * 标准化的中型载客汽车 - 0.663 * 标准化的小型载客汽车 - 0.645 * 标准化的微型载客汽车 - 0.049 * 标准化的重型载货汽车 + 0.050 * 标准化的中型载货汽车 - 0.031 * 标准化的轻型载货汽车 + 0.122 * 标准化的微型载货汽车 - 0.039 * 标准化的其他汽车。

可以看出: 第一主成分的线性组合中大型载客汽车、中型载客汽车、重型载货汽车、中型载货汽车、轻型载货汽车、微型载货汽车、其他汽车, 这些变量相当, 所以, 第一主成分可以看成上述变量的综合变量。第一主成分可以看作中、大型载客汽车与载货车的拥有量, 该主成分占比 56%, 第二主成分可以看作小型载客汽车、微型载客汽车拥有量, 该主成分占比 12.6%, 说明民用汽车中, 以家庭或个体为单位的车辆拥有者占比是很大的。基于标准化后的数据, 代入到主成分计算公式再除以对应的方差贡献率, 可得表 6, 即主成分得分表, 该表按照综合得分进行排序。

Table 6. Principal component score
表 6. 主成分得分

地区	1	2	综合
重庆	-0.7725	3.9462432	43.64422
广东	3.059176	0.8960141	18.3295
山东	2.312781	-0.037199	5.589999
江苏	1.591547	0.0806388	5.075859
西藏	-1.19946	0.6125935	3.964946
四川	0.434653	0.2226756	3.707627
浙江	0.706377	0.1279381	3.318904
湖北	0.311198	0.2034428	3.163755
河南	0.919767	0.0148731	2.566299
辽宁	0.373233	0.131249	2.490009
河北	1.002942	-0.0225	2.350432
云南	-0.10779	0.1882075	1.896844
江西	-0.30496	0.2064423	1.594541
上海	-0.45222	0.2341157	1.53137
北京	0.393878	0.0232709	1.294524
贵州	-0.51906	0.2133346	1.116969
新疆	-0.33202	0.166726	1.064631
陕西	-0.34089	0.1656377	1.028931
甘肃	-0.65441	0.2280752	0.935167

Continued

吉林	-0.65315	0.2183522	0.825962
山西	-0.43216	0.1363494	0.452506
海南	-1.20355	0.2826046	0.136575
宁夏	-1.09162	0.2379771	-0.08835
青海	-1.17417	0.2436468	-0.23765
天津	-0.85207	-0.710861	-10.4421
湖南	0.170876	-1.022192	-11.3812
黑龙江	-0.23211	-1.016983	-12.3699
安徽	0.226323	-1.150243	-12.7183
福建	-0.41283	-1.166079	-14.5653
广西	-0.18005	-1.478905	-17.5785
内蒙古	-0.58772	-2.175423	-26.6979

可以看出, 工业发达地区, 民用车拥有量最高, 以重庆、山东、江苏为代表; 其次是经济发达地区, 以上海、北京为例; 最后是偏远地区民用车拥有量最少, 以广西、内蒙古为典型。虽然经济最为发达的北京、上海、广州并为在前, 很大一部分原因在于, 北京作为首都, 核心作用为经济、政治、文化中心, 而非首都功能, 近年都在往外地调整。例如, 成立雄安新区, 且京津冀一体化, 河北的民用车拥有量较高, 是十分符合政策与实际的。

3. 线性回归预测

通过对已有的 2005 年~2020 年的民用车拥有量进行回归预测, 可得下图中, 各类民用车的拥有量的未来总趋势。从图 3~11, 可以看出中、微型载客、载货民用车拥有量均呈现减少的趋势。

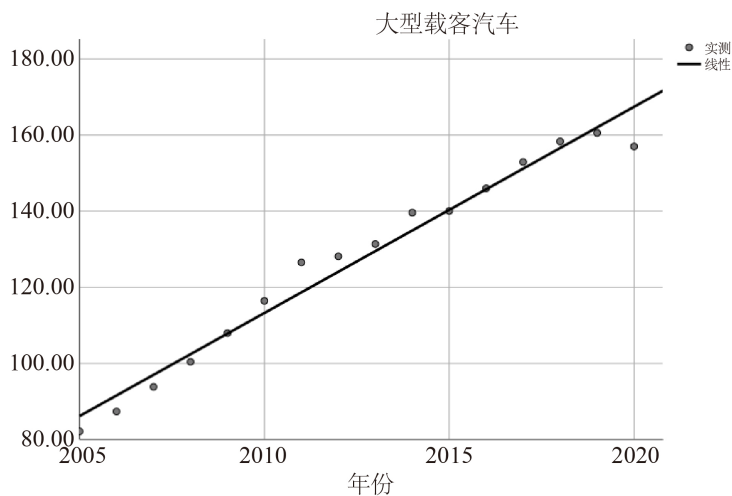


Figure 3. Large passenger cars
图 3. 大型载客汽车

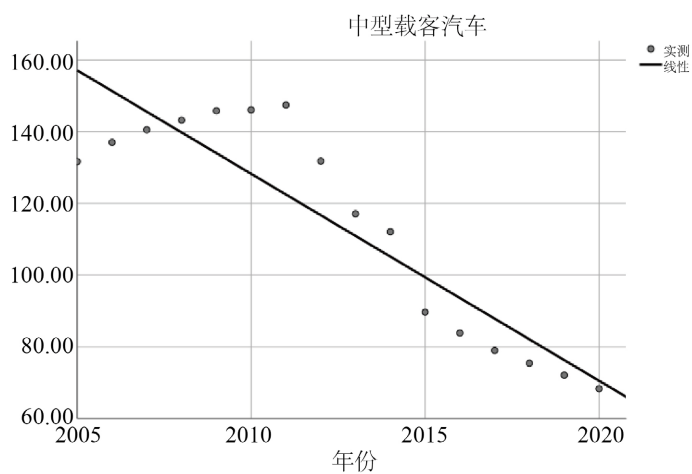


Figure 4. Medium-sized passenger cars

图 4. 中型载客汽车

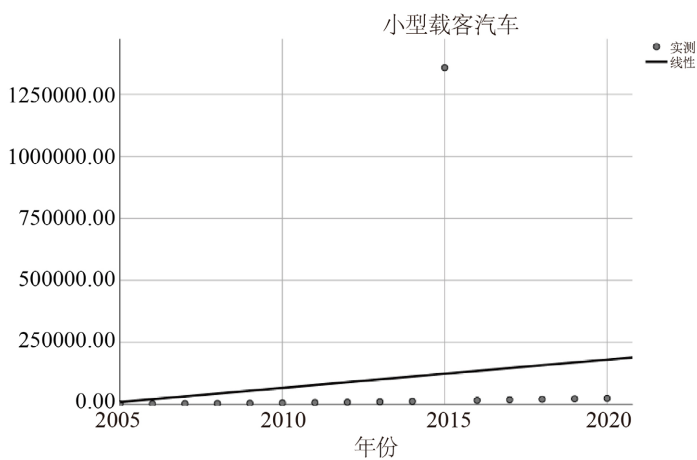


Figure 5. Small passenger cars

图 5. 小型载客汽车

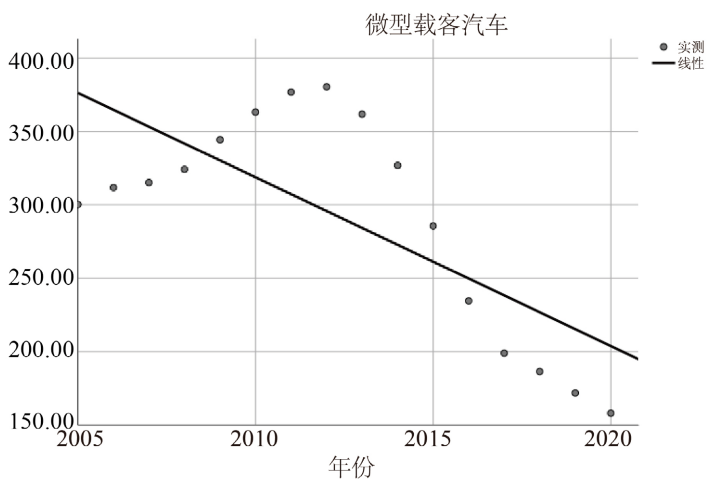


Figure 6. Miniature passenger car

图 6. 微型载客汽车

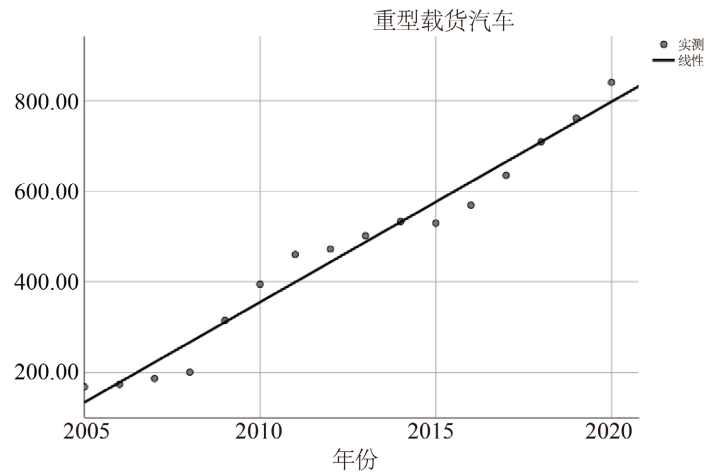


Figure 7. Heavy-duty trucks
图 7. 重型载货汽车

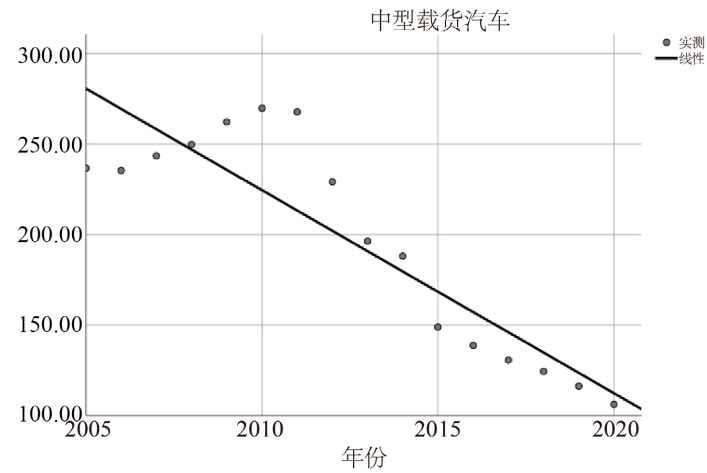


Figure 8. Medium-sized trucks
图 8. 中型载货汽车

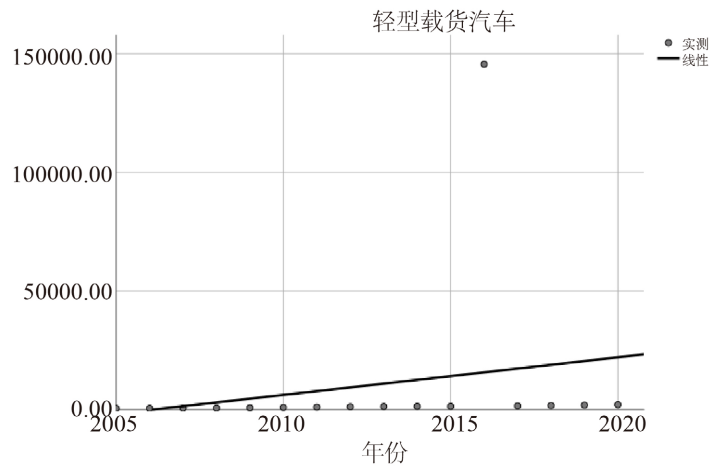


Figure 9. Light trucks
图 9. 轻型载货汽车

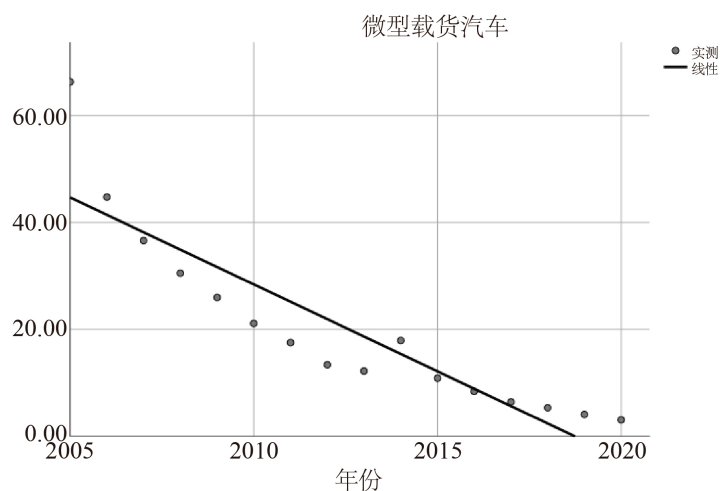


Figure 10. Miniature trucks

图 10. 微型载货汽车

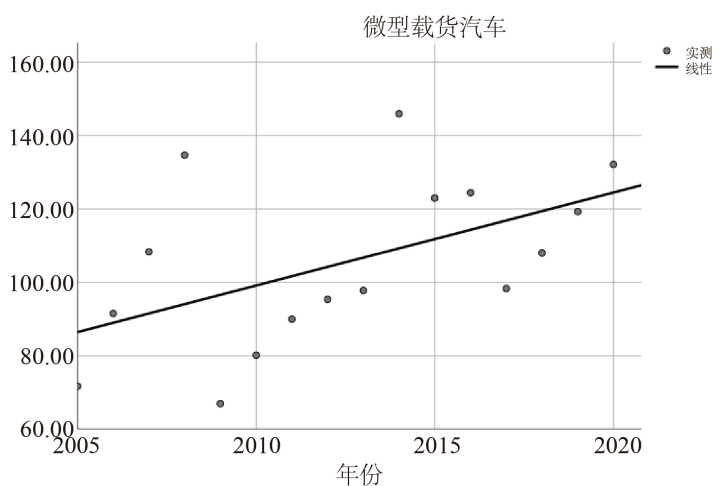


Figure 11. Other cars

图 11. 其他汽车

4. 总结

可以看出, 工业发达地区, 民用车拥有量最高, 以重庆、山东、江苏为代表; 其次是经济发达地区, 以上海、北京为例; 最后是偏远地区民用车拥有量最少, 以广西、内蒙古为典型。各类民用车的拥有量的未来总趋势, 中、微型载客、载货民用车拥有量均呈现减少的趋势, 其他类型民用车仍是增长趋势。我国交通规划仍需进一步优化, 适应未来更多车辆涌入规划道路。此外, 仍需进一步发展新能源, 以此缓解车辆增加造成的汽车尾气对环境的污染。大城市的轨道交通的发展, 缓解了大城市汽车道路拥堵的情况, 未来其他城市可借鉴该做法。

参考文献

- [1] 王英, 胡晓华. 中国民用汽车拥有量多层次分析[J]. 海南师范大学学报(自然科学版), 2018, 31(2): 206-210.
- [2] 陈永胜, 周林芳. 基于指数平滑法与马尔科夫链的汽车拥有量预测模型[J]. 长春工程学院学报(自然科学版), 2020, 21(1): 124-128.

- [3] 刘一鹤, 牟唯嫣, 金童. 线性回归方法在空气质量影响因素分析中的应用[J]. 应用数学进展, 2022, 11(8): 5936-5950.
- [4] 曲培元, 赵志斌, 陈浩, 徐东昕, 刘国军. 基于多元线性回归分析方法的汽车油耗(MPG)预测模型[J]. 统计学与应用, 2022, 11(2): 206-215.
- [5] 党耀国, 米传民, 钱吴永. 应用多元统计分析[M]. 北京: 清华大学出版社, 2012: 112-146.
- [6] 于淼, 金童. 基于主要城市中主要污染物排放量的主成分分析(英文) [J]. 农业科学与技术: 英文版, 2017, 18(7): 1260-1262.
- [7] 金童, 牟唯嫣, 贾晓芳, 李泽好. 基于因子分析法的我国邮政和电信业务量情况分析[J]. 统计学与应用, 2021, 10(6): 989-996.
- [8] 国家统计局, 编. 中国统计年鉴 2021 [EB/OL]. <http://www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm>, 2021-09.
- [9] 何晓群. 多元统计分析(第四版) [M]. 北京: 中国人民大学出版社, 2015: 113-127.