

# 基于多元线性回归的新能源汽车销量影响因素研究

陈 龙

上海理工大学, 上海

收稿日期: 2023年1月9日; 录用日期: 2023年1月29日; 发布日期: 2023年2月13日

## 摘 要

随着当今世界科技的快速发展, 汽车行业也随之飞速发展。在带给人们出行便利的同时, 燃油车对于世界上不可再生资源的消耗以及其对环境的恶劣影响, 人们不得不选择更加经济环保的新能源汽车, 其中以纯电动汽车发展的最为迅速。现在人们在买车时都会考虑新能源汽车, 这对新能源汽车企业来说既是机遇又是挑战。本文通过2021~2022年我国主流新能源汽车的销量统计, 利用多元线性模型进行线性回归, 来分析影响新能源汽车销量的几个因素, 主要有价格、充电时间、百公里能耗及电池类型等因素, 并通过异方差检验以及多重共线性检验来验证模型的可行性。最终通过标准化回归模型来找出对销量影响的最主要因素是电池充电时间, 为新能源汽车企业以及消费者提供一定的帮助。

## 关键词

新能源汽车, 异方差, 多重共线性检验, 销量

## Research on Influencing Factors of New Energy Vehicle Sales Based on Multiple Linear Regression

Long Chen

University of Shanghai for Science and Technology, Shanghai

Received: Jan. 9<sup>th</sup>, 2023; accepted: Jan. 29<sup>th</sup>, 2023; published: Feb. 13<sup>th</sup>, 2023

## Abstract

With the rapid development of science and technology in today's world, the automobile industry

also develops rapidly. While bringing convenience to people's travel, due to the consumption of non renewable resources in the world and its adverse impact on the environment, people have to choose more economical and environmentally friendly new energy vehicles, among which pure electric vehicles are developing most rapidly. Now people will consider new energy vehicles when buying cars, which is both an opportunity and a challenge for new energy vehicle enterprises. Based on the sales statistics of China's mainstream new energy vehicles from 2021 to 2022, this paper uses the multiple linear model for linear regression to analyze several factors affecting the sales of new energy vehicles, mainly including price, charging time, 100 km energy consumption and battery type, and verifies the feasibility of the model through heteroscedasticity test and multiple collinearity test. Finally, the standardized regression model was used to find out that the most important factor affecting the sales volume was the battery charging time, so as to provide some help for new energy vehicle enterprises and consumers.

## Keywords

New Energy Vehicles, Heteroscedasticity, Multicollinearity Test, Sales Volume

Copyright © 2023 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

随着社会的发展, 当今世界的汽车产业也随着时代飞速发展。汽车产品的快速发展给人们的生活带来了极大的便利。但随着传统燃油车对于能源的消耗以及其排放物对于环境的污染, 已经很难满足当今世界人与自然和谐相处、绿色发展等理念, 使其慢慢退出历史舞台[1]。因此, 更加环保节能的新能源汽车更加受到人们的青睐, 国家也对于新能源汽车的发展创造了极好的环境, 各种新能源汽车新兴品牌应运而生, 对于一个企业的发展, 其产品的销量必然是最重要的评价指标。

过往的论文对于新能源汽车销量的研究很多, 主要方法主要有王小璇的基于 LSSVM (Least Square SVM)的新能源汽车销量研究[2], 白一凡的基于 SARIMA (Seasonal Autoregressive Integrated Moving Average)和 BP 神经网络的新能源汽车销售预测, 以及张娟的基于朴素贝叶斯和 SARLMA (seasonal ARIMA model)的新能源汽车销量预测研究等方法[3]。以前的研究方法的研究样本众多, 数据对结论的影响很小, 但是对于销量的研究时间线太长可能会导致数据的不准确性, 可变因素的增多, 进而影响结论的预测研究。

本文拟对当今市面上常见的新能源汽车品牌的销量进行研究分析, 数据来源于各大网站及各个汽车官网的消费者数据, 主要集中在 2021~2022 年的数据, 时间线较短, 对数据的影响减小, 创新点在于采用多元线性回归模型对汽车的销量进行分析, 本文在此基础上对最初的模型假设作异方差检验和多重共线性检验, 来确保模型的可行性以及准确性, 研究影响其销量的各个因素以及其中最为重要的因素[4]。

## 2. 多元线性回归

在现实世界中, 变量之间大多为不确定关系, 两个变量之间的非确定性关系也称为相关关系。在数理统计中, 研究变量间各种相关关系的分析方法称为回归分析, 常见的回归分析有一元线性回归和多元线性回归。考虑到影响新能源汽车销量的因素有很多, 所以采用多元线性回归分析。

设有  $k$  个自变量  $X_1, X_2, X_3, \dots, X_k$  所对应的每个观测值分别为  $X_{i1}, X_{i2}, \dots, X_{ik}$ ，以及因变量  $y$  对应的观测值  $y_i$  满足关系式：

$$Y_i = \beta_0 + \sum_j \beta_j x_{ij} + \varepsilon_i, i = 1, 2, \dots, n; j = 1, 2, \dots, k$$

其中  $\beta_0$  和  $\beta_j$  为未知待估计的参数， $\varepsilon_i$  为无法观测且满足一定条件的误差项。

### 3. 新能源汽车销量影响因素分析

#### 3.1. 新能源汽车销量常见影响因素

新能源汽车已经成为现代人买车必然会考虑的方向，但是市面上的新能源汽车品牌众多，对于消费者的购买来说也是一个极大的考验。不同的消费者在自己购买车辆时考虑的方面虽然会有所不同，但是总体考虑的几个方面都是相同的。本文根据各类汽车论坛用户讨论的结果，将影响新能源汽车销量的因素主要归纳为以下几种，分别是汽车的价格、行驶里程、充电时长、百公里能耗以及一些汽车的常规参数电池类型、车辆产地和车型。电池类型主要分为磷酸铁锂电池和三元锂电池两类，产地主要是国产新能源汽车和进口合资汽车，车型按照汽车轴距的大小，分为 A、B、C 三种类型[5]。

本文根据在各类汽车论坛，找出在 2021.03~2022.03 时间段之间我国市场主流的几十款车型的销量数据，并且根据网站上购买该车的用户反映以及品牌方提供的数据找出了每一款车型的价格、行驶里程、充电时长、百公里能耗以及汽车的电池类型、车辆产地和车型等信息。根据这些信息作为自变量，来与因变量销量进行回归性分析，进而得出影响新能源汽车销量的因素以及其中最为重要的因素，来对新能源汽车企业的发展作出参考。

#### 3.2. 多元线性回归模型的建立

基于以上对于影响新能源汽车销量主要因素的分析，我们便可以对其各种影响因素做多元线性回归模型分析。我们可以把影响新能源汽车销量的各个变量分为定量指标和定性指标，对于其值会在一定范围内变化的变量称为定量指标，对于其范围在几个不同类型之间的变量称为定性指标[6]。对于这两种指标的分类见下表 1。

Table 1. Overview of eight indicators

表 1. 八个指标的总体情况介绍

变量类型	变量名称	说明
定量指标	销量	反映顾客对产品的关注度
	价格	数据位于 3.28~45.67 之间
	续航里程	数据位于 170~620 之间
	充电时间	数据位于 75~651 之间
定性指标	百公里能耗	数据位于 7.4~21.38 之间
	电池类型	共两类：磷酸铁锂和三元锂
	产地	共两个：国产和进口
	车型	共三款：A、B、C

首先，对于定量指标，我们记各品牌新能源汽车的销量为因变量  $y$ ，记各车型的价格为  $x_1$ ，各车的电池续航里程为  $x_2$ ，每次从 0 开始充电所需要的时间为  $x_3$ ，各车的百公里能耗为  $x_4$ 。

在研究定性指标时，我们为了研究汽车的电池类型、产地以及汽车的大小类型是否会对汽车的销量

造成比较大的影响，我们记电池类型为  $x_5$ ，汽车产地为  $x_6$ ，及汽车大小类型为  $x_7$ 。由于这些变量没有具体的数据，我们引入虚拟变量。分别设置汽车电池为磷酸铁锂电池、汽车产地为国产及汽车车型为 A 型的变量为对照组，其余的变量为虚拟变量。即当样本是虚拟变量时，我们对其回归系数取 1，当其为对照组时，我们取其回归系数为 0。在加入控制变量后，我们就能够对定性指标作回归性分析。

首先我们假设以上的八个变量与新能源汽车的销量都服从多元线性回归模型：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_7 x_7 + \varepsilon_i$$

其中  $\beta_0$  为常数， $\beta_1 - \beta_7$  为各个因素的回归系数， $\varepsilon_i$  为扰动项。

首先我们对影响销量的定量指标即每一款车型的价格、行驶里程、充电时长、百公里能耗作描述性统计，利用 MATLAB 软件，所得结果如下表 2。

**Table 2.** Descriptive statistics

**表 2.** 描述性统计结果

变量	销量	价格	续航里程	充电时间	百公里能耗
最小值	11344	3.28	170	75	7.4
最大值	417731	45.67	620	651	21.38
均值	55904.263	17.983684	484.44737	123.60526	14.654737
中位数	33002.5	14.735	500	116.5	14.45
峰度	3.7309717	1.0730988	-0.722799	5.4379781	0.0485609
偏度	18.063228	3.7190377	3.6884	32.396654	4.4409158
标准差	72610.3	9.4097998	100.84285	90.240981	2.4637071

从描述性统计的结果我们可以看出，研究样本销量的均值大概为 55904 辆，样本的价格平均为 15 万元左右等相关信息。以及与销量有关的各种指标的最值、均值及标准差等数据，让我们对数据的分析有宏观的认识。

对于定性指标，我们使用 STATA 软件分别新能源汽车的电池类型、车辆产地及车型大小描述性统计。分别设置汽车电池为磷酸铁锂电池、汽车产地为国产及汽车车型为 A 型的变量为对照组，其余的变量为虚拟变量。各个变量的描述性统计结果如下表 3、表 4、表 5。

**Table 3.** Descriptive statistics of battery type

**表 3.** 电池类型的描述性统计结果

电池类型	数量	频率	累计频率
磷酸铁锂电池	14	36.84	36.84
三元锂电池	24	63.16	100
总计	38	100	100

**Table 4.** Descriptive statistical results of automobile origin

**表 4.** 汽车产地的描述性统计结果

产地	数量	频率	累计频率
国产	32	84.21	84.21
进口	6	15.79	100
总计	38	100	100

**Table 5.** Descriptive statistics of vehicle size  
**表 5.** 车型大小的描述性统计结果

车型大小	数量	频率	累计频率
小	17	44.74	44.74
中	15	39.47	84.21
大	6	15.79	100
总计	38	100	100

要想分析出对于销量影响的重要变量，还需要进行多元线性回归分析。因此我们对此模型进行方差分析。假设原假设：

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \dots = \beta_7 = 0$$

其中多元线性回归的回归方法采用最小二乘估计法，利用其去寻求对每一个自变量的回归系数，其中最小二乘法原理记：

$$Q(\beta_0, \beta_1, \dots, \beta_7) = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_7 x_{i7})]^2$$

在我们设置原假设的基础上使用 OLS 及普通最小二乘估计法进行回归方程的显著性检验，方差分析结果见表 6。

**Table 6.** ANOVA results  
**表 6.** 方差分析结果

Source	SS	df	MS	F(8, 29) = 13.68
Model	1.54E+11	8	1.93E+10	Prob > F = 0.00
Residual	4.09E+10	29	1.41E+09	R-squared = 0.7905
Total	1.95E+11	37	5.27E+09	Adj R-squared = 0.7327

利用 STATA 对影响销量的定量变量和定性变量进行回归分析，首先对七个自变量前的回归系数进行联合显著性检验，其  $P$  值等于 0.00，小于 0.05，所以根据  $P < 0.05$  可以得出在 95% 的置信水平下拒绝原假设。

其中对于多元线性回归的拟合优度的求解需要先计算回归平方和，误差平方和以及总体平方和，计算公式如下：

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

据此，我们可以得出回归系数  $\beta_i (i=1, 2, \dots, 7)$  不会全部为 0 的。由方差分析结果，我们得到回归平方和  $SSR = 1.54E+11$ ，残差平方和  $SSE = 4.09E+10$ ，总离差平方和  $SST = 1.95E+11$  [7]。根据：

$$R^2 = 1 - \frac{SSE}{SST}$$

$$R_{\text{adjusted}}^2 = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$

由此我们得出拟合优度  $R^2 = 0.7905$ ， $R^2_{\text{adjusted}} = 0.7327$ ，同时，我们对于汽车销量的研究所做的回归模型的变量是解释型回归，我们对于拟合优度的要求并不是很高。对于拟合优度，其可能会与调整后的拟合优度差距会比较大，所以可能存在多重共线性的问题，综合以上因素考虑，我们使用调整后的拟合优度[8]。

**Table 7.** Regression analysis results based on OLS

**表 7.** 基于 OLS 的回归分析结果

销量	Coef.	Std. Err.	<i>t</i>	<i>P</i> > <i>t</i>
价格	2341.56	1423.522	1.64	0.111
续航里程	-101.5178	84.76195	-1.2	0.241
充电时间	614.3058	84.03849	7.31	0
百公里能耗	-5429.381	4552.103	-1.19	0.243
A2	-32589.96	14436.1	-2.26	0.032
B2	26287.96	21498.46	1.22	0.231
C2	-8568.982	17249.78	-0.5	0.623
C3	-49657.16	35338.78	-1.41	0.171
_cons	94264.64	62419.64	1.51	0.142

对于所得回归系数进行显著性 *t* 检验，通过表 7 的基于 OLS 的回归分析结果可以看出，变量  $x_2$  和  $x_4$  的 *P* 值均小于 0.05，所以可以得出汽车的充电时间和汽车的电池类型是显著性影响因素。但由于变量本身的横截面特性导致我们的要求不能够被满足，所以我们对数据进行异方差怀特检验及多重共线性 VIF 检验。

### 3.3. 异方差检验

对于多元线性回归的扰动项，其方差的不全相等会对回归的结果造成很大的影响。由于影响汽车销量的数据均为横截面数据，其本身特性会导致扰动项的方差不全相等，也就是异方差的出现。如果扰动项出现异方差，则会导致 OLS 估计出来的自变量回归系数是无偏的，OLS 估计量也不再是最优线性无偏估计量，也会导致假设检验无法使用。

异方差的检验大多有三种方法，第一种是画出残差与拟合值的散点图，直观的观察是否存在异方差。其次就是使用异方差的假设检验怀特检验，对于本文中的模型，我们采用最为准确的怀特检验进行异方差的检验。

利用 STATA 软件，我们对 OLS 模型进行怀特检验得到：

$$\chi^2(35) = 37.88$$

$$P > \chi^2 = 0.3399$$

我们可以得出 *P* 值的结果大于 0.05，所以在 95% 的置信区间下，不能够拒绝原假设，就可以认为本文销量建立的模型不存在异方差。

### 3.4. 多重共线性检验

在多元线性回归模型中，多重共线性是指如果存在一个自变量能够被其他的自变量线性表出，则称该模型存在严重的多重共线性。为了检验本文的模型中是否存在多重共线性的问题，我们采用方差膨胀

因子 VIF。对于本文的模型，一共有 7 个自变量，所以第  $m$  个自变量的方差膨胀因子的计算公式为：

$$VIF_k = \frac{1}{1 - R_{1-k/M}^2}$$

VIF 的值越大，说明改变量与其他变量的相关性就越大，也就意味着该模型存在多重共线性的影响。我们将影响销量的所有 7 个自变量进行计算，利用 STATA 软件得出的 VIF 计算结果见下表 8。

**Table 8.** Independent variable VIF test results

**表 8.** 自变量 VIF 检验结果

Variable	VIF	1/VIF
价格	4.71	0.212253
C3	4.48	0.223318
百公里能耗	3.3	0.30279
续航里程	1.92	0.521256
C2	1.92	0.521606
B2	1.66	0.60341
充电时间	1.51	0.662185
A2	1.31	0.764697
Mean VIF	2.6	

由结果我们可以看出，影响新能源汽车销量的所有自变量的 VIF 都小于 5，所以不存在多重共线性的影响。所以我们通过简单最小二乘法做得到的多元线性回归结果是符合模型的假设，不存在相关问题。

#### 4. 模型的结果分析

对于影响新能源汽车销量所建立的多元线性回归模型，最终得出我们的假设模型的拟合优度是 0.7327，由于该模型的类型属于解释型模型，因此可以接受此结果。我们也可以得出对于新能源汽车销量影响较大的因素有汽车的充电时间以及汽车的电池类型。其对应的回归系数分别为  $\beta_2 = 614.3$  和  $\beta_4 = -32590$ ，扰动项  $\varepsilon = 94264.6$ 。由此我们本文的线性回归模型：

$$y = 614.3x_2 - 32590x_4 + 94264.6$$

为了更为精准的研究影响新能源汽车销量的重要因素，并且去除量纲对于最终结果的影响，我们可考虑使用标准化回归系数，标准化回归后相应可得到标准化回归系数[9]。

对显著性变量进行标准化回归得到的标准化回归系数的结果见表 9。

**Table 9.** Standardized regression coefficient results

**表 9.** 标准化回归系数结果

	销量	beta
充电时间	614.306***	0.7634
	0	
电池类型	-3.26e+04**	0.2194
	0.032	

(\*\*\*  $p < 0.01$  \*\*  $p < 0.05$  \*  $p < 0.1$ )

对回归进行标准化之后，我们通过 beta 值得大小来判断影响新能源汽车销量的最主要因素，由于

$\beta_2 = 0.7634$ ,  $\beta_4 = 0.2194$  我们可以得出  $\beta_2 > \beta_4$ , 因此对于汽车销量影响最大的因素是汽车的充电时间。

## 5. 总结

本文研究对象为影响新能源汽车销量的主要因素与最关键因素, 通过对网站论坛上找到的数据进行整合及描述性统计, 选取了多元线性回归模型进行分析, 假设检验来计算出对于销量影响最大的自变量, 并且通过对于模型的异方差检验及多重共线性检验来验证模型的正确性。最终在 7 个自变量之中找出对于新能源汽车销量影响最大的两个要素分别是汽车的电池充电时间以及电池的类型。并且通过对模型进行标准化回归后, 得出这两个因素影响更大的因素为电池的充电时间。

新能源汽车是我国乃至全世界未来发展的重中之重, 是大势所趋, 也是人们将来进行买车是必然选择。所以能够找到影响新能源汽车销量的主要因素对于企业以及购车人群来说都是十分重要的。只有找到了关键因素, 企业的销量才会增加, 企业的发展才会快速。对于消费者来说, 在购车是要更加关注该车的的核心指数, 才能买到自己最适合的车。因此, 本文的研究对于新能源汽车的发展及销售有一定的指导意义。

## 参考文献

- [1] 路进乐. 新能源汽车购买决策的影响因素研究[J]. 时代汽车, 2018(4): 79-80.
- [2] 王小璇. 我国新能源汽车保有量影响因素分析及预测方法研究[D]: [硕士学位论文]. 北京: 华北电力大学(北京), 2021.
- [3] 刘云芬. 基于多元回归模型的大学生期末数学成绩影响因素分析[J]. 湖北师范大学学报(自然科学版), 2018, 38(4): 103-106.
- [4] 田君, 高洪波, 张跃强, 王一拓, 胡道中. 电动汽车动力锂离子电池寿命预测方法研究[J]. 电源技术, 2020, 44(5): 767-770.
- [5] 黄炎璐. 基于多元线性回归分析的广西农产品跨境电商发展影响因素实证研究[J]. 桂林航天工业学院学报, 2021, 26(4): 470-476.
- [6] 陈振宇, 张倍源, 席聪霞, 窦珍, 左艾艳. 基于多元线性回归分析的水资源供需问题研究[J]. 新型工业化, 2021, 11(6): 27-28+35.
- [7] 白一凡. 基于 SARIMA 和 BP 神经网络的新能源汽车销售预测[D]: [硕士学位论文]. 湘潭: 湘潭大学, 2019.
- [8] 贾洪军, 王辉. 人力资源素质提升对经济发展贡献的实证检验——基于 STATA 软件的回归分析[J]. 经济研究导刊, 2021(11): 88-90.
- [9] 张娟. 基于数据挖掘的我国新能源汽车市场分析和销量预测研究[D]: [硕士学位论文]. 北京: 北京工业大学, 2020.