

基于经济计量回归模型的车辆能耗 分层抽样方法研究

胡莹^{1,2}, 邵春福¹, 王书灵², 孙海瑞²

(1. 北京交通大学 交通运输学院, 北京 100044; 2. 北京交通发展研究中心, 北京 100073)

摘要: 随着车辆保有量的不断上涨,环境的不断恶化,控制车用燃油已成为北京市交通领域节能减排工作的重要措施之一。私人小客车作为车用燃油的主要用户,由于数量庞大,消费自主性较强,目前无法采用全样统计的方式获取其燃油消耗,仅能依靠抽样的方式,利用样本推算其整体燃油消耗。因此,为获取可靠的燃油消耗样本,准确反映全市私人小客车燃油消耗水平,明确能耗显著影响因素和设计科学的分层抽样框架成为了关键。本文利用北京市现有私人小客车能耗调查数据,通过构建计量回归模型分析私人小客车能耗关键影响因素;利用计量学中引入虚拟变量的处理方法,提出了分层抽样框设计原理和阈值划分模型,最终形成分层抽样框架。

关键词: 城市交通;分层抽样方法;经济计量模型;私人小客车;关键影响因素;交通能耗

Vehicle Energy-consumption Stratified Sampling Survey Based on Econometrics Regression Model

HU Ying^{1,2}, SHAO Chun-fu¹, WANG Shu-ling¹, SUN Hai-rui²

(1. School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China;
2. Beijing Transportation Research Center, Beijing 100073, China)

Abstract: With increasing of the vehicles and deteriorating of the environment, controlling the use of vehicle fuel has become one of the important measures in Beijing transport energy saving and emission reduction. The private vehicle energy-consumption cannot count from the whole sample, because there have too large number and the adding oil for the private vehicle is subjective. It only can count from the sampling survey method. Therefore, in order to obtain the reliable data of fuel consumption which accurately reflects the city's private vehicle energy consumption, it is necessary to explore the key factors of energy consumption and design a scientific stratified sampling framework. based on the data of Beijing private vehicle energy consumption survey, this paper constructs a econometric regression model to study the key determinants of energy consumption of private vehicle. This paper proposes the stratified sampling frame and threshold divided model through introducing the dummy variables which uses in econometrics, eventually form a stratified sampling framework.

Keywords: urban traffic; stratified sampling survey; econometric model; private vehicle; key influence factors; traffic energy-consumption

0 引言

近年来,北京市持续遭受不同程度的雾霾天气,居民生活质量和健康状况受到严重影响.统计数据发现,2013年北京市机动车已达534.1万辆,全市机动车能耗为1 921.2万吨标准煤,占全市总能耗的26.6%,其中,社会车辆能耗为594.9万吨^[1].全市机动车污染物排放已经成为本地污染物的主要来源,占31.1%.

为节约能源、改善空气质量,北京市政府印发《北京市2013—2017年清洁空气行动计划》和《北京市2013—2017年机动车污染控制工作方案》,其中重点提出了“控制车用燃油总量,2017年比2012年降低5%以上,促进机动车污染物减排25%”的考核目标.而目前,私人小客车由于其数量庞大,燃油消耗为自主性消费,无法从供给端进行燃油统计,仅能从消费端依靠抽样调查的方式获取特征样本推算整体燃油消耗.因此,如何在调查成本有限的条件下,获取可靠的私人小客车燃油消耗样本,准确反映其消耗水平,成为了亟需解决的问题.

采用分层抽样调查的方式是一种组织实施便利、精度较高,且能够获取准确总体值的一种常用技术.但采用这种技术的前提是必须事前充分掌握整体信息,明确整体的特征,提供一个合理的分层抽样框架.由于受到车辆总体量较大,能耗数据为部分数据和私人信息保密性等条件的限制,实际工作中无法采用事前分层技术获得车辆的分层抽样框.但郭世琪等^[2]的研究表明,只要抽取的样本容量足够大,利用抽样样本进行总体特征分析,获取分层依据的事后分析方法,也可以得到与事前分层抽样同样的效果.

当采用事后分层方法进行私人小客车燃油特征估算时,对车辆百公里油耗的关键影响因素的判断会影响分层框的构建.因此,本文在搜集整理北京市现有交通能耗随机抽样获取的样本数据的基础上,引入统计分析和计量回归分析方法,对车辆百公里油耗的关键影响因素进行判断,并对分层阈值进行划分.

1 基础数据及方法分析

1.1 基础数据情况分析

目前,北京市关于私人小客车能耗调查的数据主要包括3类,每年开展的小样本调查数据、5年开展1次的大样本调查数据和不定期更新的数据.数据来源的调查类型和样本容量如表1所示.

表1 数据基本情况
Tabel 1 The data situation

| 调查类型 | 数据来源 | 数据规模 |
|-------|--------------|---------------------|
| 小样本调查 | 2012年街访调查 | 1 042 |
| | 2013年街访调查 | 1 030 |
| | 2011年入户调查 | 1 630 |
| | 2012年入户调查 | 800 |
| 大样本调查 | 2013年入户调查 | 714 |
| | 北京市第3次交通综合调查 | 7.4万户 (14 042有车) |
| | 北京市第4次交通综合调查 | 4.7万户 (13 991有车) |
| 其他数据 | 中国汽车燃料消耗量网站 | 29 153 |
| | 易车网 | 55 |

根据张卫华等^[3]研究的结论,汽车自身的特性,如车型、车龄、排量等,以及道路状况、驾驶行为都会对车辆的百公里油耗产生重要影响.同时,考虑事后分析方法需基于一定样本量的基础上.因此,本文重点利用大样本调查和部分小样本数据开展研究工作.

1.2 研究方法分析

关键影响因素分析一般包括^[3-4]统计和计量两种递进的分析方法.统计侧重于将一堆看似杂乱无章的数据整理成表格、图形和分类的形式,并比较不同数据之间的差异.而计量主要以一定的经济理论和统计资料为基础,运用数学、统计学方法与电脑技术,以建立经济计量模型为主要手段,定量分析具有随机性特性的经济变量关系.

(1) 统计方法.

在统计分析中,研究变量间相互关系常用的分析方法包括画散点图、计算简单Pearson相关系数、进行简单的双变量单因素方差分析或进行非参数分析.本文利用统计学的方法初步判定车辆特征与百公里油耗的表象关系.

(2) 计量方法.

计量分析是研究变量在数据上的因果关系,通过实际经验与回归结果,判定变量间的因果关系,以及不同变量对目标变量的边际影响效应的大小.计量分析方法的本质是通过回归方程的设置来构建一个模拟的控制实验环境,在此环境中研究兴趣变量之间的关系.本文将采用计量分析方法来分析车辆特征数据与百公里油耗数据之间的因果关系,并进行阈值的划分.

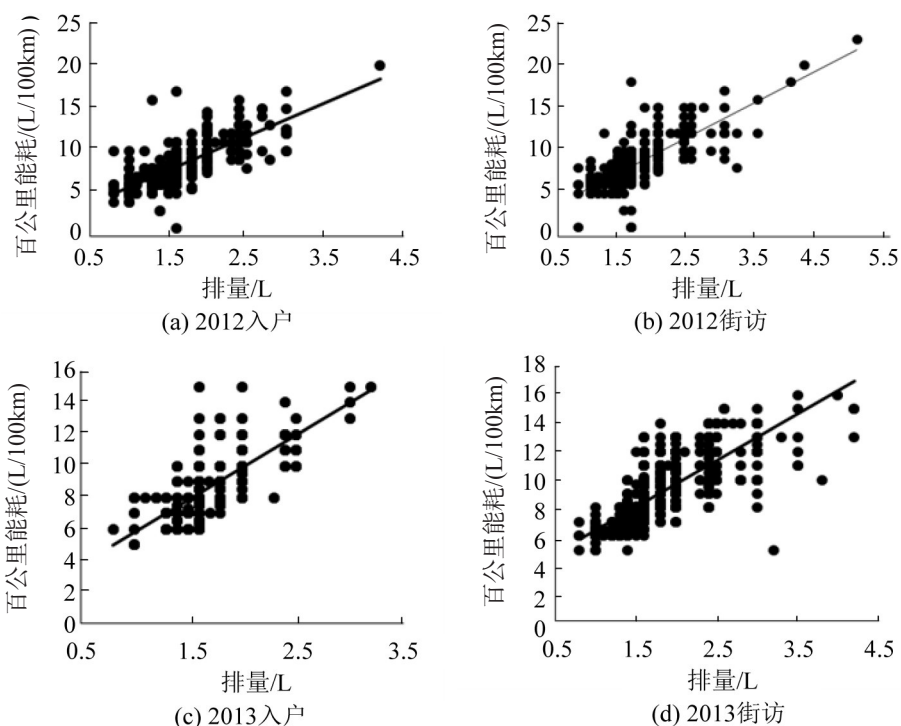


图1 排量与百公里油耗散点图

Fig. 1 A scatter diagram of displacement and one hundred kilometers energy consumption

表2 排量和百公里能耗的相关系数

Tabel 2 The correlation coefficient of displacement and one hundred kilometers energy consumption

| 数据来源 | 2012 入户 | 2013 入户 | 2012 街坊 | 2013 街坊 |
|------|---------|---------|---------|---------|
| 相关系数 | 0.694 0 | 0.756 0 | 0.521 5 | 0.743 2 |

(2) 车龄分析.

利用入户和街坊调查数据,对车龄进行散点描述和相关系数分析.结果表明:车龄对百公里能耗的影响并不是很明显,但是车龄是否为百公里油耗的显著影响因素还需要进一步的计量分析.具体结果如图2和表3所示.

2 关键影响因素分析

2.1 统计分析

为初步明确百公里能耗的可能影响因素,判断因素间的表象关系,本文对排量、车龄和车辆类型3个因素,分别采用散点描述、相关系数和非参数检验等方法进行了分析.

(1) 排量.

利用入户和街坊调查数据,进行散点描述和相关系数分析.结果显示:排量与百公里能耗呈现显著正相关,为可能的影响因素,具体结果如图1和表2所示.

表3 车龄的百公里能耗的相关系数

Tabel 3 The correlation coefficient of vehicle age and one hundred kilometers energy consumption

| 数据来源 | 2012 入户 | 2013 入户 | 2012 街坊 | 2013 街坊 |
|------|----------|---------|----------|----------|
| 相关系数 | -0.057 0 | 0.029 7 | -0.016 3 | -0.068 3 |

(3) 车辆类型.

利用大样本调查数据,对车辆类型进行了简单统计分析和非参数检验分析,结果表明车辆类型是百公里能耗的可能影响因素.具体结果如表4和表5所示.

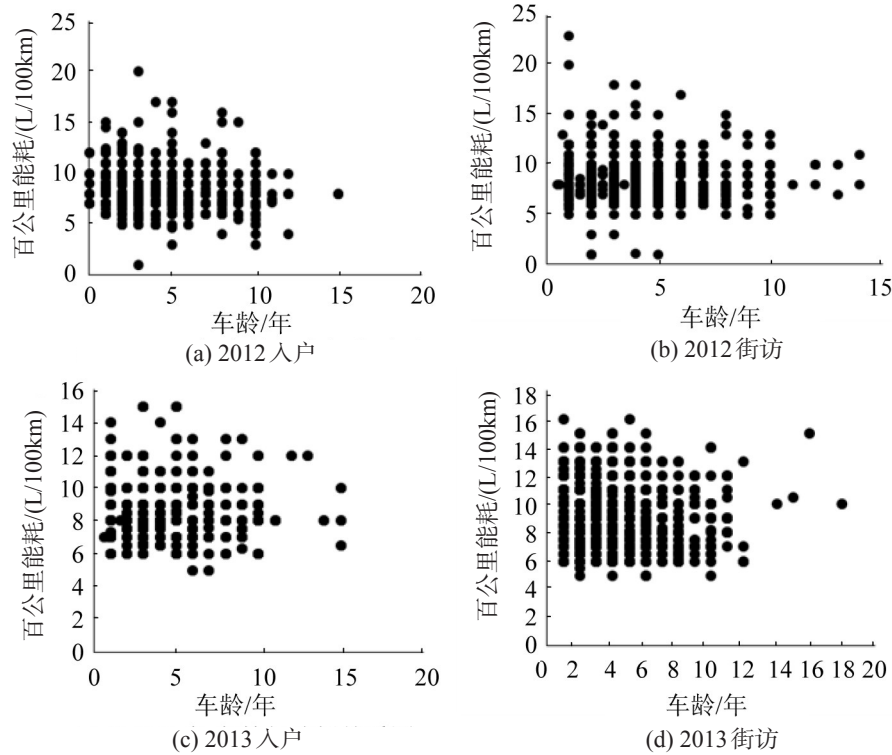


图2 车龄与百公里油耗散点图

Fig. 2 A scatter diagram of vehicle age and one hundred kilometers energy consumption

表4 车辆类型和百公里能耗的统计分析情况

Tabel 4 The statistical analysis result of vehicle types and one hundred kilometers energy consumption

| 车辆类型 | 百公里能耗均值(L/100 km) | 误差值/% |
|------|-------------------|-------|
| 小轿车 | 10.91 | 32.11 |
| 越野车 | 16.07 | |

表5 Kruskal—Wall H非参数检验结果

Tabel 5 The Kruskal—Wall H nonparametric test results

| 车辆类型 | 样本数/辆 | 秩和 |
|------|--------|----------|
| 小轿车 | 10 248 | 5.30e+07 |
| 越野车 | 132 | 878 863 |

chi-squared = 32.070 with 1 d.f. probability = 0.000 1

chi-squared with ties = 32.572 with 1 d.f. probability = 0.000 1

2.2 计量分析

(1) 计量模型构建.

由于实际中每辆车的差异,以及驾驶者驾驶行为的差异会导致调查数据中存在异方差的可能性,从而影响计量模型中回归系数的显著性结果.因此,本文利用怀特检验和BP检验进行了异方差分析,提出了本文基于“OLS+稳健标准差”验证法

的计量模型.具体模型为

$$Y_i = \alpha + \beta_0 A_i + \beta_1 B_i + \beta_2 C_i + \gamma \quad (1)$$

式中: Y_i —— t 样本的百公里油耗;

A_i —— t 样本排量;

B_i —— t 样本车龄;

C_i —— t 样本车辆类型;

α ——分类虚拟变量;

β_{0-2} ——变量的系数;

γ ——常数.

(2) 应用分析.

利用调查数据,在数据通过异方差检验后,利用“OLS+稳健标准差”估计模型,对车龄、车辆类型和排量与百公里能耗间的关系进行了因果分析.检验结果和模型回归结果如表6和表7所示.

表6 异方差检验结果

Tabel 6 The test results by heteroscedastic

| 检验方法 | 原假设 | 检验结果 |
|------|-----|------------------------------------|
| 怀特检验 | 同方差 | chi2(8)=139.19 Prob > chi2=0.000 0 |
| BP检验 | 同方差 | chi2(1)=24.15 Prob > chi2=0.000 0 |

表7 “OLS+稳健标准差”回归结果

Tabel 7 The regression results by “OLS + difference robust standard”

| 变量 | 系数 | 标准误 | t | P> t | [95% 置信区间] |
|----|--------------|-----------|-------|-------|------------|
| 排量 | 3.689 327*** | 0.140 546 | 26.25 | 0 | 3.413 409 |
| 车龄 | -0.005 72 | 0.018 28 | -0.31 | 0.755 | -0.041 6 |
| 类型 | 0.611 518*** | 0.322 724 | 1.89 | 0.059 | -0.022 05 |
| 常数 | 1.740 505*** | 0.422 807 | 4.12 | 0 | 0.910 458 |

注:***表示1%的显著水平;**表示5%的显著水平;*表示10%的显著水平。

从表7可以看出,排量和车辆类型呈现三颗星,是百公里油耗的显著性影响因素,而车龄不是显著影响因素。

2.3 小结

通过散点描述、相关系数和非参数检验等统计学方法对排量、车龄和车辆类型等3个因素的分析可以初步明确:排量和车辆类型对百公里能耗的影响很显著,车龄不是很显著。在此基础上,构建基于“OLS+稳健标准差”验证法的计量模型,进一步分析百公里能耗与车龄、车辆类型和排量之间的因果关系,通过分析表明,排量和车辆类型为百公里油耗的显著性影响因素,在分层抽样框设计中需要考虑。

3 分层抽样框设计

3.1 分层抽样设计原理

根据计量分析结果,车辆类型和排量为影响私人小客车百公里油耗的关键影响因素,因此,以车辆类型为第1层分层标志,排量为第2层的分层标志。

全市私人小客车车辆类型按照车体结构主要包括越野车和小轿车两类,且两类间存在显著差异,因此第1层分为越野车和小轿车。而由于第2层排量范围较广,需要对两类车的排量阈值进行分层方法研究。

3.2 分层抽样阈值划分模型

本文基于现有调查数据,采用事后分层抽样的研究思路,采用计量学中虚拟变量引入的方式,通过统计和计量的方式,建立阈值划分模型。其中,主要步骤如下:

Step 1 利用现有数据,进行统计分析,判断直观的划分区间,提出的划分方案 n ,及划分区的数目 j 。

Step 2 将方案的区间划分数目 j ,设置虚拟变量 i ,其中 $i=j-1$ 。

Step 3 构建虚拟变量计量分析模型,具体模型为

$$Y_i = \alpha + \sum_{i=j-1} \beta_i a_{ii} + u \quad (2)$$

式中: Y_i —— i 样本的百公里油耗;

α ——分类截距;

β_i ——第 i 个虚拟变量的系数;

a_{ii} ——第 i 样本是虚拟变量,表示 i 样本是否属于第 i 个区间,若 $a_{ii} = 1$,则属于,若 $a_{ii} = 0$ 则不属于;

u ——常数。

Step 4 利用步骤3构建模型,对方案 n 进行计量分析,最终通过计量结果,得出最优分层方案。

3.3 应用分析

利用阈值划分模型,分别对越野车和小轿车排量阈值进行划分。

(1) 越野车排量阈值划分。

Step 1 利用历史数据,对越野车的排量和油耗进行直观观察,关系图如图3所示。

通过关系图3可看出,越野车排量和油耗之间并没有显著变化趋势,但是依然可以看出排量为2.0 L, 2.4 L, 3.6 L和3.7 L为直观变化点。

Step 2 根据直观变化点,提出了3种划分越野车排量阈值的方案。具体的划分方案如表8所示。

表8 排量的阈值划分情景

Tabel 8 The scenarios for the SUV displacement

| 方案 | 阈值划分 |
|-----|--------------------------------|
| 方案1 | [1,2.0], (2.0,3.7], (3.7,+) |
| 方案2 | [1, 2.4], (2.4, 3.6], (3.6, +) |
| 方案3 | [1,2.0], (2.0,+) |

Step 3 利用引入虚拟变量的方式,构建计量模型,对方案进行比选。

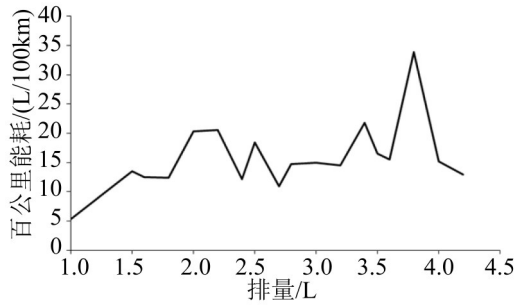


图3 越野车排量与油耗的关系图

Fig. 3 The relationship of SUV displacement and energy consumption

Step 4 通过模型,对3组方案进行计量回归,其计算结果如表9所示.

表9中回归结果显示,3种方案的阈值划分都

没有通过显著性检验,并且R²都非常小,表明越野车中,排量不需要分层.

(2) 小轿车排量阈值划分.

Step 1 利用历史数据,直观观测小轿车排量与油耗的关系,如图4所示.

表9 越野车计量模型回归结果

Tabel 9 The regression results through the SUV Econometrics model

| 方案 | 区间① | 区间② | 区间③ | R ² |
|-----|-------|------|-------|----------------|
| 方案1 | -1.94 | | -2.51 | 0.008 4 |
| 方案2 | | 2.19 | -0.15 | 0.006 1 |
| 方案3 | 0.73 | | | 0.004 3 |

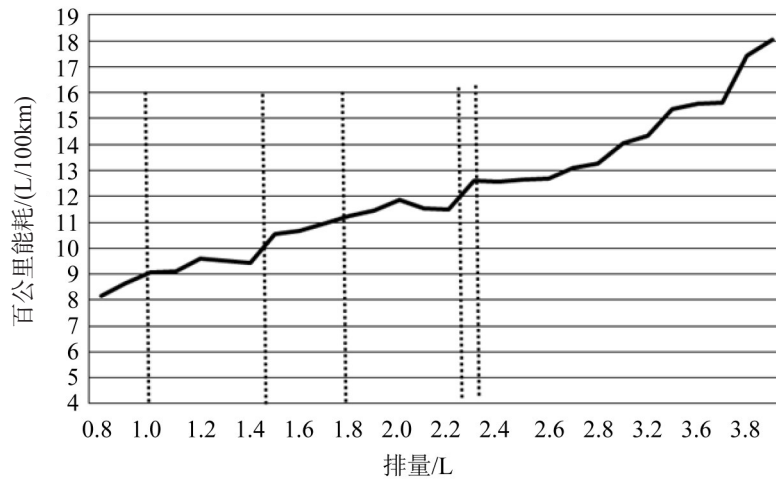


图4 小轿车排量与油耗的关系图

Fig. 4 The relationship of sedan displacement and energy consumption

图4显示,小轿车的排量与百公里油耗间呈现正相关的关系,且1.0 L以下的车辆百公里油耗均在9 L/100km以下,车辆排量达到1.5 L后在百公里油耗提高至10 L/100km,当达到1.8 L排量时,百公里油耗在11 L/100km,当排量变为2.3 L和2.4 L以上后,车辆的百公里油耗均在12 L/100km以上.由此可以初步判断,小轿车排量的变化点主要为1.0 L、1.5 L、1.8 L、2.3 L.结合实际车辆常见排量值,选择1.0 L、1.5 L、1.6 L、1.8 L、2.3 L和2.4 L作为小轿车排放量变化点.

Step 2 在Step1的基础上,提出了8种划分小轿车排量阈值的方案.具体的划分方案如表10所示.

Step 3 利用设置虚拟变量的方法,构建小轿

车排量阈值划分计量模型.

Step 4 根据计量模型获取小轿车排量的阈值的最优方案,8项方案的回归结果如表11所示.

表10 小轿车排量阈值划分方案

Tabel 10 The scenarios for the sedan displacement

| 方案 | 阈值划分 |
|-----|-----------------------------|
| 方案1 | (<1,[1,1.6],[1.6,2.4],≥2.4) |
| 方案2 | (≤1,(1,1.6),(1.6,2.4),>2.4) |
| 方案3 | (<1,[1,1.8],[1.8,2.4],≥2.4) |
| 方案4 | (≤1,(1,1.8),(1.8,2.4),>2.4) |
| 方案5 | (<1,[1,1.5],[1.5,2.4],≥2.4) |
| 方案6 | (<1,[1,1.5],[1.5,2.3],≥2.3) |
| 方案7 | (<1,[1,2.4],≥2.4) |
| 方案8 | (<1,[1,2],≥2) |

表 11 小轿车计量模型回归结果
Tabel 11 The regression results through the sedan econometrics model

| 方案 | 区间① | 区间② | 区间③ | R ² |
|------|----------------|----------------|----------------|----------------|
| 方案 1 | -5.552 553 *** | -4.733 276 *** | -3.083 521 *** | 0.4 |
| 方案 2 | -5.215 5 | -4.656 089 | -3.084 068 | 0.406 |
| 方案 3 | -5.548 767 *** | -3.947 585 *** | -0.986 963 *** | 0.457 |
| 方案 4 | -5.225 696 *** | -3.861 684 *** | -0.980 009 *** | 0.479 |
| 方案 5 | -5.550 177 *** | -4.835 548 *** | -3.196 863 *** | 0.379 |
| 方案 6 | -5.550 177 *** | -4.835 548 *** | -3.196 863 *** | 0.379 |
| 方案 7 | -5.563 081 *** | -3.610 642 *** | | 0.237 |
| 方案 8 | -4.953 537 *** | -3.328 762 *** | | 0.446 |

注:回归系数中的三颗星表示模型的构建通过了1%的显著性验证。

模型结果显示,随着排量的升高,小轿车的百公里油耗逐渐升高,且方案4回归系数最高,且显著性最强,因此方案4为小轿车排放阈值划分的最优方案。

(3) 小 结

通过分层阈值划分模型,确定了小汽车分层抽样调查框如表12所示。

表 12 私人小汽车的分层抽样框
Table 12 The stratified sampling framework for vehicle

| 分层指标 | (0,1] | (1,1.8] | (1.8,2.4] | (2.4,+) |
|------|-------|---------|-----------|---------|
| 小轿车 | | | | |
| 越野车 | | | | |

4 结 论

本文在梳理北京市现有私人小客车能耗统计调查方法和数据的基础上,采用事后分层抽样法,利用计量和统计相结合的方式,构建的基于计量回归模型的分层抽样调查方法。

该方法主要通过构建基于“OLS+稳健标准差”验证法的计量模型,明确了北京市私人小客车的百公里能耗关键影响因素为车辆类型和排量,并利用引入虚拟变量的处理方法,建立了分层抽样框设计原则和阈值划分模型,并利用该模型最终提出了北京市私人小客车分层抽样框。

通过本研究,构建的基于计量回归模型的分层抽样调查方法,既能定量确定出能耗的关键影响因素,又能提供准确的事后分层依据,为调查实

施和精确估计北京市私人小客车的总体能耗提供了技术与理论支撑。

参考文献:

- [1] 2014年北京市交通领域节能减排年度报告[R]. 北京:北京市交通行业节能减排中心. [2014 Beijing transport energy and environment annual report[R]. Beijing: Beijing Transport Energy and Environment Center.]
- [2] 郭世琪,冯士雍,等. 事后分层方法及其在交通运输抽样调查中的应用[J]. 数理统计与管理, 1994(6): 1-5. [GUO S J, FENG S Y, et al., Post-stratification method and its application in transportation sampling survey[J]. Application of Statistics and Management, 1994(6): 1-5.]
- [3] 张卫华,王炜,等. 京石高速公路及其辅道汽车油耗比较分析[J]. 公路交通科技, 2003, 20(3): 182-185. [ZHANG W H, WANG W, et al. Comparison analysis of vehicle fuel consumption on jing-shi expression and its corresponding parallel road[J]. Journal of Highway and Transportation Research and Development, 2003, 20(3): 182-185.]
- [4] 杨青云,赵培英,杨冬青,等. 数据质量评估方法研究[J]. 计算机工程与应用, 2004, 40(9): 3-4. [YANG Q Y, ZHAO P Y, YANG D Q, et al. Research on data quality assessment methodology[J]. Computer Engineering and Application, 2004, 40(9): 3-4.]
- [5] 刘洪,黄燕,等. 统计数据质量及其评估方法[J]. 统计与决策, 2006(3): 30-31. [LIU H, HUANG Y, et al. Statistical data quality and its evaluation methods[J]. Statistics and Decision, 2006(3): 30-31.]
- [6] 蔡风田,谢元芒,等. 汽车运行油耗的影响因素与汽车节能技术[J]. 交通节能与环保, 2006(1): 10-15. [CAI F T, XIE Y M, et al. Automotive operation the influence factors of fuel consumption and energy saving technology[J]. Transport Energy Saving and Environment, 2006(1): 10-15.]
- [7] 颜卫中,张维群,等. 均匀设计在多目标抽样调查方案设计中的应用[J]. 统计与决策, 2006(9): 147-148. [YAN W Z, ZHANG W Q, et al. Uniform design in the application of multi-objective sampling scheme design[J]. Statistics and Decision, 2006(9): 147-148.]
- [8] 艾小青,金勇进,等. 抽样调查下样本随机性的检验[J]. 统计研究, 2009, 26(9): 28-31. [AI X Q, JIN Y J, et al. Hypothesis testing of the sample's randomness in sampling surveys[J]. Statistical Research, 2009, 26(9): 28-31.]