

从数据到模型

全国大学生统计建模大赛获奖论文选

全国大学生统计建模大赛执行委员会 编
国家统计局统计教育培训中心

 中国统计出版社
China Statistics Press

从数据到模型

全国大学生统计建模大赛获奖论文选

全国大学生统计建模大赛执行委员会 编
国家统计局统计教育培训中心

 中国统计出版社
China Statistics Press

(京)新登字 041 号

图书在版编目(CIP)数据

从数据到模型:全国大学生统计建模大赛获奖论文选/全国大学生统计建模大赛执行委员会,国家统计局统计教育培训中心编.--北京:中国统计出版社,2010.7

ISBN 978-7-5037-5969-7

I. ①从… II. ①全… ②国… III. ①统计模型—文集 IV. ①C8-53

中国版本图书馆 CIP 数据核字(2010)第 118576 号

从数据到模型:全国大学生统计建模大赛获奖论文选

作者/全国大学生统计建模大赛执行委员会 国家统计局统计教育培训中心

责任编辑/吕军 孙慧

封面设计/李雪燕

出版发行/中国统计出版社

通信地址/北京市西城区月坛南街 57 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电话/邮购(010)63376907 书店(010)68783172

印刷/河北天普润印刷厂

开本/710×1000mm 1/18

字数/600千字

印张/35.5

版别/2010年7月第1版

版次/2010年7月第1次印刷

书号/ISBN 978-7-5037-5969-7/C·2339

定价/49.00元

版权所有。未经许可,本书的任何部分不准以任何方式在世界任何地区以任何文字翻印、拷贝、仿制或转载。

中国统计版图书,如有印装错误,本社发行部负责调换。

序

国家统计局副局长 李 强

一分耕耘，一分收获。《从数据到模型——全国大学生统计建模大赛论文选》即将付梓之际，编辑同志送来书稿，请我为序，第一句话，我就联想到耕耘与收获的景象。我知道，这样一本书，融汇着从大赛策划到组织培训，从论文写作到大赛答辩整个全程的辛劳努力，可谓“事非经过不知难”。

2008年的全国统计系统建模大赛，我曾作为专家评委组成员，参与论文评审和赛事总结评判。记得我还代表大家提出一个建议：“就是进一步发展建模大赛这种很好的方式，可以在办好全国统计系统建模大赛的基础上扩大到高等院校，进一步扩展，扩大影响，培养更多的优秀统计人才。”可喜可贺的是，这个建议，2009年得以实现，第一届全国大学生统计建模大赛成功举行。

这届大赛秉承创新精神，为大学生们走出课堂，开阔视野，参与实践，搭建了一个很好的平台，为今后继续开展这项活动提供了范例，值得嘉许。统计建模比赛活动，不断扩而广之，范围从国家统计局机关扩展向全国统计系统，再扩展向高等院校，这是一件好事，一件应该鼓励的事。好的活动，应该坚持下去，不断拓展，实现可持续发展。

这本论文选是一面镜子，一个载体，字里行间有激情有理性，有探索有建树。从中可以看出大学生们思维很活跃，知识面很广，基本功也很扎实；可以看出他们在统计、经济、计算机知识的结合与运用上，已经初见功力；还可以感受到当代大学生们，朝气蓬勃，富有个性的创新精神、竞争精神和团队精神。

“风物长宜放眼量”这本论文选也启示我们：当代大学生继续努力下去、奋斗下去，就可以成为我们国家统计战线的后备人才的宝贵资源。以此角度，看全国大学生统计建模大赛，意义就更为宏大和深远；站此高度，开展好这项赛事，统计

系统责无旁贷且惟我们最能搞好也最应该搞好。

大赛不止在赛场。统计是一片充满朝气和生机的天地，一位英国统计学家说：“统计的大部分历史在未来……”书写统计的未来，对于统计、对于有志于统计事业、对于有兴趣钻研统计科学、运用统计科学的大学生们，前面充满希望，前面任重道远。

然而“千里之行，始于足下”，走好每一个台阶，才能一步步达到理想的境界。每一项好的有益的统计活动或竞赛，都是年青人进步的一个台阶，把全国大学生统计建模大赛打造成当代大学生展示风采的舞台，让枯燥的统计数字彰显独特的魅力，让大赛成为具有影响力的统计品牌，就是为大学生们铺设进步的台阶。

记得全国统计建模大赛有一个响亮的口号：“一样的数据，不一样的创造。”我觉得这个口号很好，它强调了“创造”。这次全国大学生统计建模大赛也很好地体现了“创造”，大学生们的论文凸现了这一特色。这本论文选，是一个很好的展示创造性、创造力的平台。希望这本论文选不仅是这届大赛的一个圆满句号，还是下一届大赛的一个新的起点；希望大学生们奉献的一篇篇论文，像一颗颗充满希望的种子，带来一届届大赛金色的丰收。

目 录

一、全国大学生统计建模大赛一等奖论文

1. 笔记本电脑特征价格指数编制模型

云南财经大学

顾光同、王江、高丽 (1)

2. 金融集聚影响因素空间计量模型及其应用

湖南大学

徐玲、游万海、陈雪梅 (24)

3. 基于结构方程模型的杭州城镇居民食品安全满意度统计评估

浙江财经大学

梁一鸣、张钰烂、董西钊 (41)

二、全国大学生统计建模大赛二等奖论文

1. 中国地区经济增长收敛性分位数回归分析

安徽财经大学

黄蓓、范悍彪、宋峰 (55)

2. 大学生休闲态度、休闲参与和休闲阻碍调查研究

山东工商学院

张凤、肖粤志、许长淑 (78)

3. 用天气发生器对我国主要城市天气指标的模拟和预测

华东师范大学

吴蔚、王磊、李树良 (117)

4. 中国环保投资废气治理效率问题研究

中南财经政法大学

吴淑丽、昌先宇、谭竿荣 (148)

5. 流动性信息与资产收益:基于非参数模型的分析

浙江工商大学

李攀登、刘海燕、高贇玥 (177)

6. 2009年上半年货币投放与未来通货膨胀的分析

安徽大学

柳玲娣、胡月、赵颖 (195)

7. 外部经济变量对我国货币需求模型影响程度的实证分析

中南财经政法大学

李玮、郝婷婷、李双双 (209)

8. 紧凑型城市:中国城市经济可持续发展的新视角
浙江工商大学 夏青 (227)
9. 研究生教育收费改革实证研究
西南财经大学 高勇标、林亮、黄宝辉 (239)
10. 我国各地区教育支出与经济增长的空间计量分析
中央民族大学 宛立杰 (282)
11. 中国内陆甲型 H1N1 流感的预测和控制模型
鲁东大学 刘玉方、律清萍、高培安 (303)

三、全国大学生统计建模大赛三等奖论文

1. 次贷危机下的中国汇率风险
厦门大学 张瑞端、俞滢、栗相如 (321)
2. 5·12 灾后重建“羌绣帮扶计划”的可持续性分析
西南财经大学 徐小寒、郑夏雨、余芳 (338)
3. 休闲服装企业顾客满意度模型的改进及应用研究
天津财经大学 刘超、张津、黄天龙 (362)
4. 成都地铁沿线房价变动研究
西南财经大学 董文亮、王道新、罗媛 (395)
5. 区域软实力的测度及其对区域发展的贡献
浙江工商大学 杨少娜、孙鹏、袁妍 (421)
6. 经济理论对人行为的影响
北京大学 田昊枢、牛启昆、彭沁 (444)
7. 我国社会保障水平与经济发展及人口结构的协调性研究
河北经贸大学 柴亮、李壮壮、党建令 (476)
8. 我国对外贸易与经济增长的 VAR 模型
厦门大学 陈飞、柴家友、陈婷 (494)
9. 中国经济区域能源消费的 panel data 模型
华北电力大学 袁蒂、蒋莉莉、牛胜男 (512)
10. 本科院校考研成功率影响因素的模型构建
中南财经政法大学 朱璐璐、卢苏娟、薛亚楠 (539)

笔记本电脑特征价格指数编制模型 ——基于非线性的 Hedonic 模型

参赛队员:顾光同 王 江 高 丽

参赛学校:云南财经大学

指导老师:费 宇 王力宾

摘 要

商品价格指数的编制是一个古老而又新鲜的问题。商品的异质性问题会使普通价格指数不能真实反映商品的价格变化,特征价格是针对上述问题来构建商品价格指数的一种有效方法,本文应用特征价格法研究异质性明显的商品——笔记本电脑的价格指数编制问题。并基于传统的线性、半对数和双对数三种形式的 Hedonic 模型,以及基于 Box-Cox 变换的非线性的 Hedonic 模型来估计笔记本电脑的价格,从而编制更能反映价格变化的价格指数——笔记本电脑的特征价格指数(HPI),同时验证了本文提出的四个猜想,给出了编制 HPI 指数的建议,为我国笔记本电脑价格指数的编制提供了一种新的思路。

关键词:特征价格法;特征价格指数;特征变量;Hedonic 模型;笔记本电脑 Box-Cox 变换

一、问题的提出

从笔记本电脑的发展趋势及特点来看,笔记本电脑于 1985 年诞生于日本东芝公司,我国笔记本电脑市场萌芽于 1993 年,开始之初,笔记本电脑与台式机电脑的性能价格比差距很大,其产品和同期台式机电脑相比,同样运算能力的产品其价格比台式机高出 2 至 3 倍。随着技术的进步和产业的逐渐成熟,笔记本电脑的性能快速得到提升,价格持续下降,市场需求量也急剧增加。随着市场需求的不断扩大,笔记本电脑的销量在一路增长。据有关资料显示,我国的笔记本电脑

市场销量从1995年5.2万台到2002年超过60万台,而2003年达到150万台,8年增长28倍。截至2005年底,我国笔记本电脑年销售量已突破性的增长到近340万台,占全球笔记本电脑市场的份额约为5.2%。同时,我国笔记本电脑市场的需求增长速度还远远高于台式机市场的需求增长速度^[1]。笔记本电脑在我国短短几年内迅速走进了千家万户,笔记本“挤压”市场并将逐步代替台式机,已成为业界一个不争的断言。根据艾凯数据研究中心提供数据表明^[2],2008年,我国规模以上计算机行业实现销售收入17134亿元,同比增长7.3%,规模继续在世界保持前列,计算机产量占全球比重超过四成。2008年我国笔记本电脑市场规模突破600亿元,我们预计到2012年我国笔记本电脑市场规模将达到1000亿。该中心还提供了2008年1—12月全国计算机类产品生产、出口情况,见表1。

表1 2008年1—12月全国计算机类产品生产、出口情况 (单位:万台)

	生产			出口		
	完成	增速%	同期增速%	完成	增速%	同期增速%
台式计算机	2807.9	-17.5	0.66	768.2	-12.2	7.3
笔记本电脑	10858.7	25.2	46.7	10111.4	38.5	40.5

注:数据来源于艾凯数据研究中心,见文献[2].

由以上分析可看出,笔记本电脑逐渐在普通家庭中得到普及。对于笔记本电脑这类技术比较成熟的商品,商家之间的促销手段和价格战也加快了人们对笔记本电脑的更新换代。我国消费者物价指数的查价商品中也增加了电脑这类商品,可见电子信息商品在国民经济消费中的比重也越来越大,研究构建这些电子信息商品的价格指数也就十分必要。

另一方面,由于笔记本电脑是整机销售,即便是同一种品牌,其特征也随着型号的不同在不断发生变化,而且更新换代的速度很快。新机型不断代替旧机型,许多机型上市一段时间后即停产不再生产,成为退市商品。如此快速的变化,笔记本的观察价格的变化已经不代表实际价格的变化,传统价格指数已经不能反映真实的价格变化趋势,只有编制笔记本电脑的特征价格指数,才能较为准确的反映实际价格的变化。特征价格指数(Hedonic Price Index, HPI)是采用特征法编制的个体价格指数的统称,反映该类商品的“纯价格变化”,其基本思想是以特征理论为基础,建立反映商品特征和商品的价格之间的特征价格函数,根据特征价格函数计算的价格指数。据相关资料显示,对于一般的消费品,我国政府统计部门在编制物价指数的实践中通常采用“纯样本匹配法”编制反映“纯价格变化”的指数。但像一些异质性产品,如笔记本电脑、汽车、住房等高技术、高消费产品来

说,人们对其的需求是缺乏弹性的,它们不仅已成为人们必须的生活消费品,而且其更新换代速度加快,产品品种、规格和价格等的变化十分明显。相应的,不同时期观察到的此类产品市场价格水平的波动可能源于市场供求情况的改变,也可能源于市场中产品品质、技术的整体提升或物业聚集地的整体变化,还可能源于各期交易量中不同特征产品组成结构的变化。这就违背了价格指数编制中的“同质可比性”原则。此时,特征价格理论无疑为解决这一难题带来了福音。与传统价格指数编制方法相比,HPI方法是一种严密的、含有较少主观色彩的方法,用统计手段科学地将特征变化过程加以量化,避免了主观原因引起的测量误差。而且随着笔记本电脑市场的市场化及信息化程度的不断地提高,由于本身的优越性,特征价格法编制笔记本电脑的特征价格指数在我国具有广阔的推广前景。

二、研究现状及存在的问题

特征价格指数法可以追溯到上世纪30年代的Court^[3]发表的“Hedonic Price Indexes with Automobile Examples”,之后该方法及其经济上的解释等相关理论被建立起来了。上世纪60年代的Griliches^[4]利用Hedonic方法编制了汽车产业的价格指数。随着技术的发展特征理论逐步完善,美国学者Lancaster(1966)提出的特征消费理论,又被称为Lancaster偏好理论。与萨缪尔森等经济学家在考察偏好和效用从个体行为出发的做法不同,Lancaster从产品的差异出发,分析了构成产品的基本“元素”空间,认为对产品的需求并不是基于产品本身,而是因为产品所内含的特征,并且认为商品(特别是像住宅、汽车和劳动力这样的异质商品)具有一系列的特征,这些特征结合在一起形成影响效用的特征包,商品是作为内在特征的集合来出售的。家庭购买和使用这些商品,把它们转化为效用,效用水平的高低取决于商品所包含的各种特征的数量和质量。美国经济学家Rosen^[5]等学者扩展了Lancaster理论,其将竞价理论融入到Lancaster特征消费理论中,就商品特征提出了特征供需均衡模型。在完全竞争市场条件下,Rosen以消费者效用最大化和生产者利润最大化作为目标,从理论上分析了异质性产品市场的短期均衡和长期均衡,为特征价格理论的建模、特征价格函数的估计奠定了基础。

在国内,根据所搜集的文献,王力宾^[6]是中国较早将特征价格法应用到实践中来的学者,他将特征价格法运用到房地产价格分析研究中,温海珍、贾生华(2004)连续发表三篇论文定量分析房地产属性对房地产价格的影响,并建立杭州市总体市场和细分市场的特征价格模型。夏祥谦、王力宾^[7]在研究汽车特征价格指数时,采用汽车功率、扭矩、最高车速和整备质量作为特征变量,得到特征价格函数,并指出特征价格模型剔除了“异质性”因素,能很真实的反映汽车价格的变

动趋势。鉴于国内外学者的研究成果,我们可以肯定本文所研究对象的合理性及科学性;也相信其研究成果对现实的指导有很高的实践意义。王力宾^[8]是国内较早研究计算机特征价格指数编制方法的,不过用的不是我国的样本数据;次年,王晓玲、王力宾^[9]运用我国笔记本电脑的数据资料编制笔记本电脑特征价格指数。徐国祥、牟嫫^{[10][11]}以2005年1月至2006年12月我国计算机硬盘价格为例,编制并实证了我国计算机硬盘特征价格指数,分析了硬盘价格的变化趋势。

上述学者在计算机这类商品方面做了大量研究工作,为后续研究做了很好的铺垫,但是,我们发现一个共同存在的问题:大多是根据经验,直接采用适合于其他商品(汽车、住房、轨道交通等)的传统模型的三种形式(简单线性模型、双对数模型、半对数模型)中的一种拟合较好的模型来进行指数编制的研究,比如,雷怀英(2007,2008)在《统计研究》中分别发表的关于质量调整的 Hedonic 价格指数研究以及关于汽车价格指数研究中的价格模型也是直接选用了半对数模型进行指数编制。这样的经验方式存在几个不足之处(这里指的是值得商榷的地方,更合理的是我们觉得可以进行改进的地方):

(1)不同商品有明显的异质性会导致根据经验的模型不是万能的;

(2)没有很好解决被解释变量的正态性检验,事实上,很多商品价格即使用自然对数变换,仍然不能理所当然地把这些变换后的变量看作为近似正态,所以,上述研究在模型选择上存在一些可改进的地方;

(3)即使有学者在研究特征价格指数时提到了可用 Box-Cox 变换来选择模型,但是为何要进行变换,也就是对变量的变换必要性即假设检验问题至今没有引入 Hedonic 模型选择中;

(4)还有一些不足,可能是本文中也可能存在的,需要进一步研究。

三、模型构建前的准备

(一)若干假设

本文从研究的背景以及前辈们对所研究商品的特征指数中的不足出发,我们提出四个猜想:

(1)被解释变量即使通过取自然对数变换,仍然可能不满足正态性假定;

(2)在特征价格指数研究中,存在非线性 Hedonic 模型;

(3)对某些商品的特征价格指数研究的模型选择中,数据变换是必要的,而且变换的假设检验的引入也是必要的;

(4)针对某些商品的 Hedonic 模型,基于 Box-Cox 变换所得的非线性 Hedonic 模型优于传统三种 Hedonic 模型,并适合用于特征价格指数的编制和经济意义解

释。

本文将在编制笔记本电脑特征价格指数的实证性研究中逐一验证上述猜想。国外学者 Laurice 和 Bhattacharya(2005)认为在研究美国南加州地区三郡的房价特征价格时,增加二次的及立方的函数形式这样的非线性模型对改善模型的解释程度有相当良好的帮助作用。Griliches(1967)甚至建议可利用 Box-Cox 转换函数方法,在有关的统计结构下,从这些不同的特征函数形式选出其中一种形式。这建议说明,特征价格函数除了线性关系外,还存在非线性关系。Box 和 Draper(1969)也证明性的说明了,若在正态分布的线性模型下,采用 Box-Cox 变换函数后,将使误差得到正确的修正,更有修匀线性模型的效果。Raimond Maurer, Martin Pitzer, Steffen Sebastian(2004)指出根据所观察的数据得出,Box-Cox 变换是确定特征价格函数最恰当的方法,并且在指数形成的基础上,选择特征价格的方程式,应优先考虑 Box-Cox 模型。因此,以上这些学者的观点说明,我们提出的猜想是合理的,我们的研究是具有实际意义的。

(二)数据的来源

根据 IT 权威中关村在线发布的 2009 年第一季度中国笔记本市场研究报告^[17]资料显示,笔记本电脑用户的关注度状况,第一季度中国市场关注度从高到低的品牌为:联想(31.6%)、惠普(18.6%)、华硕(9.9%)、戴尔(9.8%)、神州(8.2%)、宏基(6.1%)、索尼(3.4%)、东芝(2.9%)、三星(2.4%)。前面 9 个品牌就占了关注度的 92.9%,其他被关注的品牌中我们还注意到苹果的关注度为 1.7%,方正的关注度为 0.8%,清华同方的关注度为 0.68%,富士通的关注度为 0.4%。本文就从这十三个关注度累积达到 96.5% 品牌出发,进行编制基于非线性 Hedonic 模型的笔记本电脑的特征价格指数的实证研究。

用于构建特征价格指数的笔记本电脑数据来源于中关村在线、太平洋电脑网以及 IT 主流资讯平台 IT168 公布的 2009 年 1 月至 9 月的上述 13 个品牌的报价和相应特征参数。9 期数据,共 580 个样本,其中第一季度 176 个样本(1 月—3 月样本数分别为 62、56 和 58),第二季度 177 个样本(4 月—6 月样本数分别为 54、60 和 63),第三季度 227 个样本(7 月—9 月样本数分别为 72、66 和 89)。

(三)特征变量的选择

构建特征价格指数,首先须选择正确的特征指标,所选特征指标既是笔记本电脑特征的代表,又是消费者较为敏感、能够引起价格变化的指标。因此,我们按照下面三个方面选择特征变量,并对数据进行量化处理。

1. 借鉴国内外的研究经验

从国内有关笔记本电脑的特征变量的选择来看,随着时间的推移以及研究侧

重点的不同,笔记本电脑特征变量的选择也在不断地发生变化,王晓玲、王力宾(2007)对特征变量的考虑最为全面,他们从基本规格、存储设备、显示屏、外观特征、电池规格、显卡类型、品牌等方面进行了特征变量的选择与组合,选取了处理器主频、内存容量、硬盘容量、重量、电池工作时间、品牌作为特征变量。但是,他们所研究的笔记本电脑样本数据是按年份收集的,时间跨度大,Hedonic模型也是经验性地采用传统模型。

2. 根据消费者的喜好

从消费者角度来看,配置、产品类型、品牌、外观以及功能是消费者购买笔记本电脑时必定考虑的四大因素,因此笔记本电脑特征的选择主要从这四方面着手,所选变量要尽可能地反映消费者的消费理念。

中关村的第一季度资料显示^[17],硬件配置已经成为制约消费的最大因素,有近36%会看重笔记本电脑的配置。由于笔记本电脑的功能扩展能力较强,因此功能因素所占的比例仅占13%。而着重笔记本电脑的款式和售后服务的消费者各占4%,只有极个别消费者重视笔记本电脑的产地。14.1英寸屏幕大小是笔记本市场主流需求趋势,获得65.6%的关注比例;迅驰2笔记本关注度最高,获得19.5%关注比例,其次是奔腾双核机型;独立显卡笔记本更胜一筹,以66.8%的关注比例领先于集成显卡机型。

从配置来看,消费者主要关注处理器标称主频、标配内存容量、处理器型号、显存、显卡类型、硬盘容量等指标;从产品类型来看,主要关注家用和商用等类型;从品牌来看,消费者主要关注13个品牌;从外观来看,主要关注笔记本电脑的屏幕尺寸以及厚薄、轻重等;从功能来看,主要关注蓝牙、无线上网、指纹识别、内置摄像头等功能。

3. 选取的特征变量与笔记本电脑价格具有较高的相关关系

利用统计软件SPSS15.0,将第二步初步选定的定量的一些特征变量与价格的相关性进行分析,结果见表2,结果显示笔记本重量和屏幕尺寸2个变量与价格(P)的相关性较小,因此在构建Hedonic模型直接将这些变量从模型中剔除。

表2 笔记本特征变量与价格的相关系数表

特征变量	标称主频 (GHZ)	标配内存容量 (GB)	硬盘容量 (GB)	笔记本重量 (kg)	屏幕尺寸 (英寸)
Pearson 相关系数	0.275(**)	0.474(**)	0.256(**)	0.053	-0.019
伴随概率	0.000	0.000	0.000	0.197	0.643

注:**表示在显著水平 $\alpha=0.01$ 处极显著;*表示在显著水平 $\alpha=0.05$ 处显著。

根据以上分析,从笔记本定量的特性来看,可初步将标称主频、标配内存容量和硬盘容量等变量纳入模型,分别用 X_1 、 X_2 、 X_3 表示。由于变量之间存在较高的相关关系,会引起多重共线性问题,而笔记本电脑配置存在性能搭配问题,比如标称主频高其标配内存容量大些,硬盘容量也大一些。从相关系数看,标配内存容量和标称主频之间的相关系数为 0.392,伴随概率为 0.000,通过了显著性检验,另外内存容量和硬盘容量二者的相关系数为 0.558,伴随概率为 0.000,通过了显著性检验。因此,在建模过程中,还需根据实际情况对变量进行再一次的筛选,以避免多重共线性问题的出现。

从定性的特征变量来看,首先考虑不同类型的笔记本其产品定位的不同以及面向的消费群的不同,其降价的幅度与快慢也不同,因此引入虚拟变量以反映笔记本的不同产品类型,如将笔记本分为家用和商用两种,引入 1 个虚拟变量,用 C_1 来表示, $C_1=1$,表示家用, $C_1=0$,表示商用。

其次,从笔记本配置方面考虑,消费者比较关心处理器型号,因此,引入 1 个虚拟变量,用 C_2 来表示, $C_2=1$,表示迅驰 2 及以上, $C_2=0$,表示奔腾双核、赛扬双核及单核等。

另外,消费者比较关心显卡是不是独立的,尤其对图像处理功能要求高的消费者更关心这个问题,通过收集数据发现,现在显卡情况有集成显卡、中低端独立显卡和中高端独立显卡,因此,引入 2 个虚拟变量,用 C_3 和 C_4 来表示, $C_3=1$,表示中低端独立显卡或中高端独立显卡, $C_3=0$,表示集成显卡, $C_4=1$,表示中低端独立显卡, $C_4=0$,表示集成显卡或中高端独立显卡,即 $C_3=1, C_4=1$ 表示中低端独立显卡; $C_3=0, C_4=0$ 表示集成显卡。

从功能来看,现代电子产品越来越追求无线传输,是否支持蓝牙是消费者购机考虑的因素,因此,引入 1 个虚拟变量,用 C_5 来表示, $C_5=1$,表示支持蓝牙,否则表示不支持蓝牙;现在网上联络即聊天已经是“家常便饭了”,消费者也很关心,笔记本是否有内置摄像头,因此,引入 1 个虚拟变量,用 C_6 来表示, $C_6=1$,表示有内置摄像头,否则表示没有内置摄像头。

从品牌来看,不同品牌,消费者关注度不同,三大关注品牌为联想、惠普、华硕,其关注度分别为 31.6%、18.6%、和 9.9%,共占市场上关注度总额的 60%以上,因此,引入 3 个虚拟变量,用 C_7 、 C_8 和 C_9 来表示, $C_7=1$,表示联想, $C_7=0$ 表示其他品牌, $C_8=0$ 表示惠普, $C_8=1$,表示其他品牌, $C_9=0$ 表示华硕, $C_9=1$,表示其他品牌,即 $C_7=1, C_8=1, C_9=1$ 表示联想; $C_7=0, C_8=0, C_9=1$ 表示惠普; $C_7=0, C_8=1, C_9=0$ 表示华硕; $C_7=0, C_8=1, C_9=1$ 表示其他品牌。

当然,还有些特征如指纹识别、人脸识别、双键盘以及迅驰 4 等高端的功能和

配置虽说是顺应潮流,逐渐成为购买笔记本的消费者喜爱的配置,但从收集到的数据资料来看,现阶段只有少数的高档笔记本才配有这些装置,因此本文暂不把这些变量引入模型。

综合以上分析,首批选定包含基本规格、类型、功能特征、品牌四个方面的 12 个特征变量,见表 3。

表 3 笔记本特征变量及其定义

特征分类	特征变量定义(单位)	代码
基本规格	1. 标称主频(GHZ)	X_1
	2. 标配内存容量(GB)	X_2
	3. 硬盘容量(GB)	X_3
	4. 处理器型号	C_2
	5. 集成显卡	C_3
	6. 中低端独立显卡	C_4
类型	7. 产品类型	C_1
功能特征	8. 蓝牙	C_5
	9. 摄像头	C_6
品牌	10. 联想	C_7
	11. 惠普	C_8
	12. 华硕	C_9

四、笔记本电脑的传统 Hedonic 模型的构建

(一)传统模型的三种形式

Hedonic 回归文献中经常使用的函数形式有三种:线性模型、半对数模型和双对数模型,也就是本文所指的三种传统 Hedonic 模型。

(1)线性模型

$$P_k^t = \beta_t + \sum_{i=1}^N a_i X_{ik}^t + \varepsilon_k^t \quad t=1, \dots, T; k=1, \dots, K^t \quad (1)$$

(2)半对数模型

$$\ln P_k^t = \beta_t + \sum_{i=1}^N a_i X_{ik}^t + \varepsilon_k^t \quad t=1, \dots, T; k=1, \dots, K^t \quad (2)$$

(3)双对数模型

$$\ln P_k^t = \beta_t + \sum_{i=1}^N a_i \ln X_{ik}^t + \varepsilon_k^t \quad t=1, \dots, T; k=1, \dots, K^t \quad (3)$$

这三种模型中, P_k^t 指在 t 时期第 k 种商品的价格, X_{ik}^t 表示 t 时期第 k 种商品的第 i 个特征, β_t 是常数项, ε_k^t 是误差项, a_i 是特征变量系数, 表示第 i 个单位特征对价格的影响程度。在 Hedonic 模型中, 暗含着一个假设条件: 消费者购买的是一组特征的组合, 由此可将特征变量的系数 a_i 解释为特征的“隐含价格”或“影子价格”。因为特征的“价格”也就不能直接观察到, 从而只能通过观察由不同特征组合成的商品的价格来估计各单位特征的“隐含价格”或“影子价格”。^[12]

(二) 基于传统模型的 Hedonic 模型构建

许多学者根据经验就直接选取半对数模型, 这主要是该模型与双对数模型相比有一个优点^[12]: 半对数模型能够处理一个或更多的特征 X_{ik}^t 为零的情况, 而双对数模型则不能。下面为了和我们提出的基于 Box-Cox 非线性模型比较优劣, 利用合并的数据(三个季度的数据放在一起, 这里我们假定特征价格不受时间的影响)以及统计软件 SPSS15.0 依次对笔记本电脑的传统 Hedonic 模型建模, 其间用逐步回归方法消除特征间的共线性以及剔除不显著的特征变量, 得到显著的线性模型、半对数模型和双对数模型(这里只对定量特征取对数, 虚拟变量不加处理)的判决系数 R^2 分别为: 0.515、0.600、0.587。由此可知, 半对数模型优于其他两个传统模型。我们又假设, 季度内的特征价格不受时间的影响, 利用半对数模型建立合并数据和季度数据的 Hedonic 模型, 利用逐步回归方法, 得到回归输出结果见表 4 所示。由该表可得基于合并数据的笔记本电脑的半对数 Hedonic 模型为

$$\ln \hat{P} = 8.581 + 0.144X_2 - 0.305C_1 - 0.238C_2 - 0.067C_4 + 0.325C_5 + 0.110C_6 - 0.139C_7$$

第一季度的半对数 Hedonic 模型为

$$\ln \hat{P} = 8.467 + 0.207X_2 - 0.354C_1 - 0.155C_2 + 0.393C_5 + 0.128C_6$$

第二季度的半对数 Hedonic 模型为

$$\ln \hat{P} = 8.363 + 0.142X_2 + 0.001X_3 - 0.216C_1 - 0.252C_2 - 0.174C_3 + 0.368C_5 + 0.153C_6 - 0.184C_7$$

第三季度的半对数 Hedonic 模型为

$$\ln \hat{P} = 8.433 + 0.110X_1 + 0.124X_2 - 0.321C_1 - 0.265C_2 + 0.249C_5$$

表 4 基于半对数回归模型(2)的输出结果

数据类型	纳入模型(2)的特征变量	特征变量系数	t 值	P 值 (t 值)	VIF (膨胀因子)
合并数据	(Constant)	8.581	182.454	0.000	
	C5	0.325	13.336	0.000	1.146
	C2	-0.238	-7.801	0.000	1.290
	C1	-0.305	-10.856	0.000	1.063
	X2	0.144	9.630	0.000	1.191
	C7	-0.139	-3.625	0.000	1.026
	C6	0.110	-3.356	0.001	1.096
	C4	-0.0670	-2.900	0.004	1.033
F 值=8.409 P 值(F 值)=0.004 R(相关系数)=0.775 R ² (判决系数)=0.600 DW 值(自相关检验)=1.554 残差平方和=42.378 AIC=-1501.51 BIC=-1466.61					
第一季度数据	(Constant)	8.467	88.209	0.000	
	C5	0.393	8.430	0.000	1.119
	X2	0.207	5.950	0.000	1.315
	C1	-0.354	-6.504	0.000	1.041
	C2	-0.155	-2.517	0.013	1.374
	C6	0.128	2.186	0.030	1.053
F 值=4.779 P 值(F 值)=0.030 R(相关系数)=0.783 R ² (判决系数)=0.612 DW 值(自相关检验)=1.564 残差平方和=14.421 AIC=-428.32 BIC=-409.29					
第二季度数据	(Constant)	8.363	100.057	0.000	
	C5	0.368	9.123	0.000	1.161
	C2	-0.252	-5.027	0.000	1.328
	X2	0.142	4.508	0.000	1.529
	C1	-0.216	-4.603	0.000	1.072
	C3	-0.174	-3.852	0.000	1.464
	X3	0.001	2.948	0.004	1.752
	C6	0.153	2.778	0.006	1.135
	C7	-0.184	-2.230	0.027	1.042
F 值=4.971 P 值(F 值)=0.027 R(相关系数)=0.817 R ² (判决系数)=0.668 DW 值(自相关检验)=1.536 残差平方和=10.309 AIC=-485.23 BIC=-482.24					