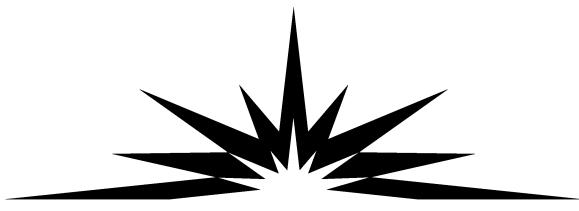


中国可持续能源项目

能 源 基 金 会  
项目资助号: G-1009-13372



**中国建筑能耗基准工具研究**  
**The Development of building energy**  
**performance benchmarking tools in China**

中国建筑科学研究院  
2011 年 10 月

## 摘要

建筑能耗基准是判定和分析能源利用效率水平高低的重要依据。与欧美发达国家相比,我国建筑能耗基准方面的工作基础较差,未建立完善的数据采集制度,建筑能耗及建筑基本情况等基础数据严重匮乏。为获取建筑能耗数据,国内各级政府机构、科研院所近年来开展了大量的工作,目前已经取得一定成果。部分持有或管理多个建筑的大型房地产开发商和物业管理公司,逐渐认识到建筑基础数据的重要性,连续多年开展数据采集整理工作。与此同时,国内多个学者开展了关于建筑能耗数据分析的工作,提出我国建筑能耗的一些特点及存在的问题。但如何建立建筑能耗基准工具,用来判断建筑能源利用效率水平的高低,到目前为止在国内一直没有得到很好的解决。

本文以酒店建筑为例,对建立我国公共建筑能耗基准工具的方法进行研究,计划主要完成以下几项工作:

1、对欧美发达国家几种典型的建筑能耗基准工具的原理、特点和应用进行介绍和比较,对优缺点进行分析。

2、对我国公共建筑能耗数据统计工作进行汇总,以不同气候区域为单位,汇总并分析我国学者已完成关于数据统计及分析工作的成果,得到全国范围内公共建筑能耗特点的共性问题,作为建立公共建筑能耗基准工具的数据依据。

3、对多种数学统计方法进行分类整理,分析其优缺点,结合我国公共建筑能耗数据的特点,选取适合的统计工具。

4、以酒店建筑为例,详细介绍数据样本筛选、自变量选择、模型建立、模型验证、评价方法等关键技术问题的解决过程,并基本建立我国酒店建筑能耗基准工具的模型基础和评价方法。

5、采取多个一元模型的选择,对多元回归模型进行改进,建立性能更优秀的模型。

**关键词:** 公共建筑, 建筑能耗基准, 多变量分析, 多元线性回归分析



## ABSTRACT

Building energy consumption benchmark is important for determining and analyzing the energy efficiency. Compared with the developed countries in Europe and America, the work foundation of building energy consumption benchmark is bad in China, such as, no comprehensive data acquisition system, shortage of building energy consumption and basic fundamental data. For the purpose of acquisition of building energy consumption data, the domestic government, scientific research institutes have made some achievements in recent years. Some of the large real estate developers and property management companies gradually realize the importance of building basic data, and begin to collect data for several years. Meanwhile, many domestic scholars make some data analysis about the characteristics and existing problems of buildings in China. However, it has not been solved properly so far how to establish the benchmark tools to determine the level of building energy efficiency. In this paper, it will take the hotel building for example to complete the following several works for establishing the benchmark tool of public buildings in China.

1. Make a comparison and analysis of the advantages and disadvantages of principles, characteristics and applications between several typical energy consumption benchmark tools in developed countries of Europe and America.
2. Summarize the statistical analysis work of public building energy consumption data in China; Based on the summarization of public building energy consumption characteristics all over the China, support some useful data for establishing public building energy consumption benchmark tools.
3. Classifying mathematical statistical method, analyzing the advantages and disadvantages, combined with the characteristics of public building energy consumption data, select appropriate statistical tools.
4. Take hotel buildings for example, introduce the process of setup the data filters, selecting independent variables, establishing a model, evaluating model and technology of how to solve the key problems. Finally, establish hotel building energy consumption benchmark tool model and evaluation method basically.
5. Improve multiple regression model by choosing some unary model and establish more outstanding model.

**Keywords:** public buildings, building energy consumption benchmarks, Multivariate Statistical Analysis, Multivariate linear regression analysis



## 目 录

<b>ABSTRACT</b> .....	4
<b>第一章 绪论</b> .....	1
<b>1.1 课题研究的背景和意义</b> .....	1
<b>1.2 国内建筑能耗基准研究现状</b> .....	4
1.2.1 通过模拟计算方法建立建筑能耗基准模型 .....	4
1.2.2 采用多元拟合回归技术建立建筑能耗基准模型 .....	6
1.2.3 采用其他统计方法研究建筑能耗定额 .....	8
<b>1.3 国外建筑能耗基准研究现状</b> .....	8
1.3.1 美国能耗基准研究现状 .....	9
1.3.2 与 ENERGY STAR® Benchmarking Tools®类似的其他工具介绍 .....	11
1.3.3 英国能耗基准研究现状 .....	12
1.3.4 德国能耗基准研究现状 .....	13
<b>1.4 课题研究任务</b> .....	17
<b>第二章 公共建筑能耗基准模型的建立方法</b> .....	19
<b>2.1 我国公共建筑能耗分析研究现状调查</b> .....	19
2.1.1 我国公共建筑能耗分析研究数据汇总 .....	20
2.1.2 我国公共建筑能耗分析研究结论汇总 .....	23
<b>2.2 公共建筑能耗基准模型建立思路的提出</b> .....	26
<b>第三章 数学统计方法的介绍及模型选定</b> .....	29
<b>3.1 统计学的基本概念</b> .....	30
3.1.1 统计学的研究对象 .....	30
3.1.2 统计学研究对象的特点 .....	30
3.1.3 统计的功能 .....	32
<b>3.2 统计学的基本分析方法</b> .....	32
3.2.1 单变量分析法 .....	33
3.2.2 多变量分析法 .....	36
3.2.3 统计学工具的选择方法 .....	39
<b>3.3 统计方法的对比和选择</b> .....	41

---

3.3.1	建筑基准模型的基本功能 .....	41
3.3.2	在单变量分析法和多变量分析法分类之间选择 .....	42
3.3.3	在多变量分析法分类中选择合适的分析方法 .....	43
3.4	回归分析法的原理 .....	43
3.4.1	线性回归 .....	43
3.4.2	非线性回归 .....	47
3.4.3	回归方法的计算流程 .....	48
3.5	本章小结 .....	50
第四章	酒店建筑能耗基准模型的建立 .....	51
4.1	数据来源及基本情况 .....	51
4.1.1	数据基本情况 .....	51
4.1.2	HDD 和 CDD 的基础数据来源和计算 .....	52
4.2	回归分析使用软件的介绍 .....	53
4.3	试验性拟合回归分析 .....	54
4.3.1	拟合过程和结果 .....	55
4.3.2	拟合结果分析 .....	57
4.3.3	小结 .....	57
4.4	多元线性拟合回归分析的技术路线 .....	58
4.5	样本的筛选 .....	58
4.5.1	样本筛选的意义 .....	58
4.5.2	样本筛选标准的制定 .....	59
4.6	自变量的筛选 .....	62
4.6.1	散点图分析法 .....	63
4.6.2	相关性分析法 .....	64
4.7	中心化的意义和实现方法 .....	65
4.7.1	中心化的意义 .....	65
4.7.2	中心化的方法 .....	69
4.8	建立电耗基准模型 .....	70
4.9	建立总能耗基准模型 .....	77
4.9.1	一次能源的转化计算 .....	77

4.9.2	拟合的结果及分析	77
<b>4.10</b>	<b>本章小结</b>	80
<b>第五章</b>	<b>多元线性回归模型的改进</b>	81
5.1	改进的非线性多元回归模型	81
5.2	改进模型的分析和建立	82
5.3	酒店建筑能耗基准模型的计算公式	84
5.3.1	五星级酒店建筑能耗基准模型	85
5.3.2	四星级酒店建筑能耗基准模型	85
5.3.3	三星级酒店建筑能耗基准模型	85
5.3.4	汇总	85
<b>第六章</b>	<b>评价方法的建立</b>	87
6.1	建筑能耗评价工具的建立	87
6.2	应用举例	91
6.3	建筑能耗评价工具的应用	92
<b>第七章</b>	<b>结论与展望</b>	95
7.1	结论	95
7.2	创新点	95
7.3	展望	96
参考文献		97

## 第一章 绪论

### 1.1 课题研究的背景和意义

中国的建筑节能事业从上世纪 80 年代开始研究，至今已走过了 30 个年头。由于全球的能源价格飞涨和气候变暖的大趋势，致使我国的建筑节能工作进入了快速发展的“黄金十年”。这期间，从国家到地方，从政府到企业，逐渐加大对建筑节能的重视，积极开展各项技术研究和政策研究，并努力在各城市、各项目中大力推广。“中国式”的经济发展速度在近十年已经震撼了全世界，中国建筑节能的快速发展同样吸引了全世界的目光。有国外学者声称，目前中国关于地源热泵系统应用的研究已经走在了世界的前列。当然这只是一个例证。更多的现实是，虽然目前我国建筑能耗水平实际远低于以美国为代表的发达国家。但由于城镇化进程的加速，我国的建筑业的发展速度惊人。同时，经济的快速发展，人民生活水平的提高，大众对室内舒适度的要求逐渐变高，这已经引起我国建筑能耗的大幅度提升<sup>(1)</sup>。与其等到建筑领域能耗对 GDP 的增长产生负面影响的时刻，不如现在就开始做工作，积极应对建筑能耗的不断攀升。

纵观发达国家建筑节能发展的历程，都经历了“自上而下”、“从自发到引导”的过程。关于建筑节能的一项技术或一种整合的技术体系，首先需要从国家的主管政府部门，按照法令、法规等政策形势进行推广，辅之于经济方面的补助和技术方面的支持。在以后相关法规颁布执行时，都会优先考虑使用那些之前大力推广的技术和技术体系。这一段时期属于“市场培育期”，完全或部分由政府主导并资助进行，让市场中的建筑业主、房地产开发商和科研院所以切身感受到节能技术的优点。当市场成熟后，相关技术应用将变成用户的自觉行为，节能技术所带来的好处已经充分得到用户的认同，包括经济效益和品牌效益等。用户会自觉主动的应用相关的建筑节能技术，积极的要求参与到节能工作中。最终实现个人用户经济利益和全社会综合效益的双赢模式。

2000 年由美国绿色建筑委员会（USGBC）颁布了非政府性的绿色建筑评价标准 LEED，全称“Leadership in Energy and Environmental Design”，中文意思为“领先的能源与环境设计”，进行多次改版和扩充，目前已更新至 LEED2009 版本，

扩充为六个分支，包括新建建筑 (NC)、商业建筑室内设计(CI)、核心筒及外壳 (CS)、绿色社区(ND)、家庭(FOR FAMILY)和学校 (FOR SCHOOL)、医疗、零售。经过多年的发展和市场培育，虽然 LEED 不是全球最早提出的绿色建筑评价体系，但目前 LEED 的市场占有率居全球前列，已在 41 个不同国家和地区应用<sup>(2)</sup>，在美国部分州和一些国家已被列入强制性标准<sup>(3)</sup>。

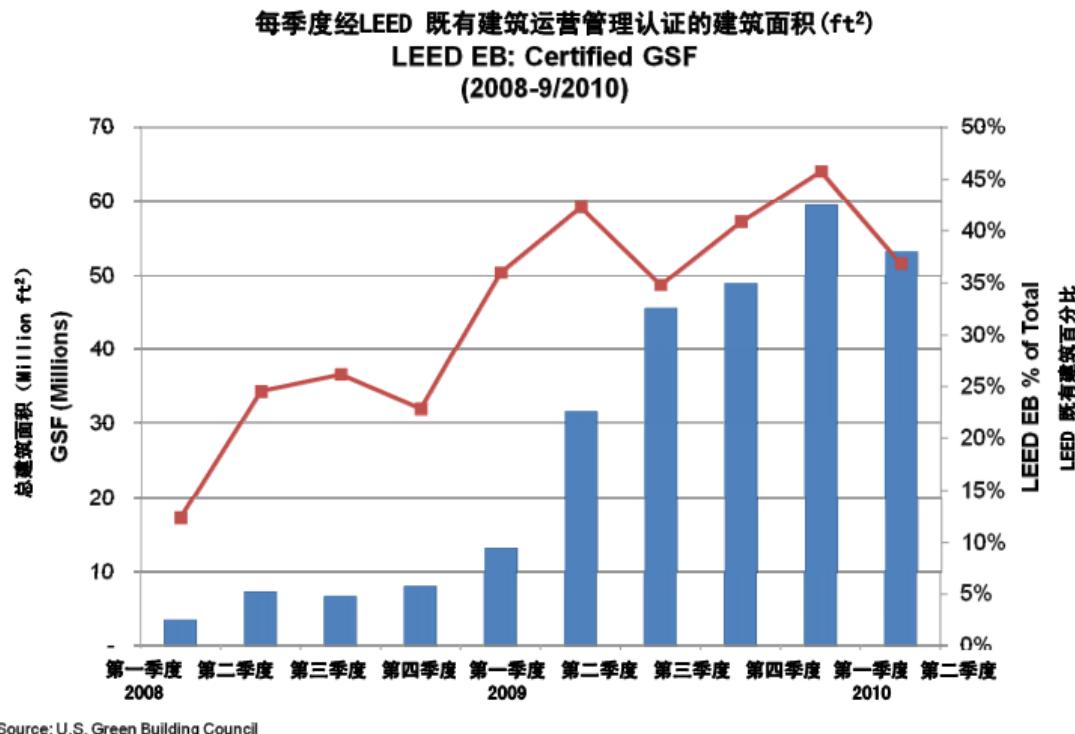


图1 参与 LEED 认证的既有建筑的发展

图 1 给出了在 2008 年~2010 年期间，申请 LEED 认证的既有建筑的发展趋势。有理由相信，从 LEED 制定的 2000 年到 2008 年第一季度之前，LEED 的发展速度是缓慢的。但在这长达八年的时间内，美国绿色建筑委员和美国政府做出了大量关于市场培育和市场挖掘的工作。以州为单位，将 LEED 直接写入各州的法律，作为强制性标准进行推广。在终端用户，包括建筑业主、建筑管理者、设计院，认识到 LEED 的有效作用和价值后，自发性选择 LEED 体系用于指导工作。这是在建筑节能领域一个典型的成功案例。

从我国目前建筑节能领域未来发展来看，建筑的能耗数据及能耗水平的评价，是建筑节能工作的基础。关于数据和评价方法的需求，来自于以下几个方面的社会需求。

(1) 国家各级政府主管部门的需求。

建筑节能相关政策法规的制订需要数据作为有力的支持。哪些地区建筑能耗高，哪种类型建筑能耗高，哪类分项能耗高，解决了这些问题只意味着能够确保节能工作的大方向是正确的，但没有基础建筑能耗数据的指导，那么节能的大方向的正确性就很难保证。同时缺少详细建筑能耗数据，很难对建筑节能工作做出准确的经济性分析，很难有效的判断具体工作的成效。

(2) 建筑业主的需求。

首先，出于运营成本、广告宣传等商业目的，建筑业主有强烈的愿望希望降低建筑能耗，以降低运营成本，愿意在建筑节能方面做工作，并期望收到切实的效果。其次，在确定开展工作前，建筑业主也需要明确所管理建筑的能耗水平在全社会中处于什么位置，属于高能耗还是属于低能耗。如果属于高能耗，那么该建筑的节能潜力空间有多大？最后一点，对于建筑业主也是最重要的。当完成相关的节能工作并取得实际效果后，如何进行有效的宣传，确保对所在的房地产开发商或物业管理公司的品牌起到提升作用。在这一点上，建筑能耗数据的支持和能耗评价体系是至关重要的。

(3) 节能技术及其系统集成商的需求。

前已述及，没有数据支持的宣传是空洞的。这也是目前国内建筑节能市场上存在的问题。对于节能技术及其系统集成商而言，缺乏数据支持的问题也是他们最期望解决的。

我国目前尚未形成一套完善的建筑能耗数据调研与汇总制度。2007年以来，为全面掌握我国建筑能耗实际状况，加强建筑节能的管理，住房和城乡建设部在23个城市组织试行民用建筑能耗统计工作。2010年经国家统计局批准，同意制发民用建筑能耗和节能信息统计报表制度，有效期2年。目前建筑能耗数据主要来自各级政府、各大科研院所、高校和相关企业，根据工作需要进行调研获得的数据，其准确度、可信度参差不齐，加上能源折算方法的不统一、调研方式的不一致，导致虽然关于数据工作，全国做了很多工作，但至今所有数据还无法汇总进行统一计算处理。

值得注意的是，近年来，随着建筑节能工作的深入，建筑能耗数据及评价体

---

---

系作为建筑节能的重要基础工作，已经得到越来越多政府主管部门、科研学者、房地产开发商和建筑节能企业的重视，重新开始做技术数据和评价体系的基础工作。

2006 年 8 月，美国环保署（EPA）与中国住房与城乡建设部（MHURD）科技司签署合作备忘录。在备忘录中，适用于中国公共建筑能耗基准工具的建立是主要的工作之一。

需要什么样的数据，获得数据之后如何进行有公信力的评价，评价体系将会如何指导节能工作，本文拟对以上问题进行探讨，希望能对我国建筑能耗评价工作的开展提供参考，通过提供有力的数据支持，使我国的建筑节能工作更进一步。

## 1.2 国内建筑能耗基准研究现状

建筑能耗基准，为各类型建筑提供建筑能耗指标或评价方法，对某一限定范围内的建筑能源消耗水平具有较高的代表性，通过该基准可以评价不同建筑能源消耗水平的高低，用于指导建筑节能工作。能耗基准在建筑节能领域的很多方面都有应用。包括新建建筑在设计阶段的能耗评价和既有建筑在运行阶段的能耗评价两方面。

### 1.2.1 通过模拟计算方法建立建筑能耗基准模型

在设计阶段应用建筑能耗基准一般使用动态能耗模拟的方法进行计算。按照国家强制建筑节能标准的规定，例如按照《公共建筑节能设计标准》GB50189-2005 中 1.0.3 条款的规定：“按本标准进行的建筑节能设计，在保证相同的室内环境参数条件下，与未采取节能措施前相比，全年采暖、通风、空气调节和照明的总能耗应减少 50%”<sup>(4)</sup>。以 20 世纪 80 年代改革开放初期建造的公共建筑为基准建筑，按照当时的围护结构、暖通空调设备及系统、照明设备的参数，在保持与目前标准约定的室内环境参数的条件下，计算得到的全年暖通空调和照明能耗，作为基准能耗。按照标准规定的围护结构、暖通空调和照明参数，按照同样的室内环境参数，计算得到的建筑能耗应为基准能耗的 50% 以下。新建建筑的设计方案必须满足该标准。

国内学者在利用模拟法计算建筑能耗基准方面做了一些工作。张旭<sup>(5)</sup>等通过 BIN 法对上海地区典型居住建筑全能耗进行计算，通过改变部分参数的方

式计算外墙、外窗传热系数、室内空调设计参数和能效比等因素对建筑能耗的影响。文精卫<sup>(6)</sup>等使用 EnergyPlus 软件对能耗进行模拟,以上世纪八十年代的建筑能效水平为能效基准,对夏热冬冷地区的新建节能建筑的能效水平进行评定。

建筑能耗模拟计算的方法有两大类,一类是建立在稳定传热理论基础上的静态能耗分析法,另一类是建立在不稳定传热理论基础上的动态能耗模拟法<sup>(7)</sup>。

静态能耗分析法用于采暖能耗的模拟,其基本原理是将供暖期或供暖期中的各旬、各月的耗热量按稳态传热理论进行计算,而不考虑各部分围护结构的蓄热效应。静态能耗分析的方法主要有:有效传热系数法、度日法、BIN 方法(变基准温度的度日法)、当量满负荷运行时间法等。

动态能耗模拟法对室内外各种扰量考虑较细,得到的结果也比较准确,但是动态模拟工程非常复杂。利用计算机,可以方便地对不断变化的室外参数作用下建筑物的冷热负荷进行动态计算。

这两种建筑能耗的模拟计算法,无论是静态能耗分析法,还是动态能耗模拟法,都是以建筑能耗相关的物理模型为基础的。在计算时,需要考虑外界各种因素的干扰。静态能耗分析法需要考虑和稳态传热相关的室内外温度和围护结构传热系数等参数的干扰。动态能耗模拟分析法还需要考虑各种外扰和内扰参数的干扰。

由此可以看出建筑能耗模拟计算,是建立在物理模型上的计算方法,要考虑与该物理模型相关的参数。为了进行计算,需要对这些参数进行约束和规定。这些参数也可以理解为“边界条件”,模型在边界条件的限定下进行计算。显然,边界条件选定的合理性,对于最终模拟计算结果的准确性具有重要的影响。

事实上,不只是模型边界条件选定的合理性会影响模拟计算结果的准确性。李骥<sup>(8)</sup>等提出:由于长波辐射难以确定、对流换热系数难以确定、自然通风难以确定和模拟软件中所作的各种算法的简化,导致软件模拟计算的结果可能存在与实际不符的情况。

采用模拟法计算建筑能耗基准,具有一定的适用范围,虽然模拟软件功能强大,使用简便,但绝不是“放之四海而皆准”的万能工具。在设计阶段,建筑没有竣工投入使用,肯定不会有实际能耗数据。因此,模拟法更加适合对新建建筑的

---

---

设计方案进行评定，模拟法不仅可以用于计算基准能耗，而且可以用于预测建筑未来的能耗。

在国外，模拟法计算建筑的基准能耗的方法也有应用。美国政府颁布了《用于减税的公共建筑节能模型及审查规范》<sup>(9)</sup>，通过模拟计算，预测建筑的实际能耗与建筑的基准能耗，通过对比两者的差值是否满足规范要求，来确定减免税收的多少。该规范将建筑分为三种模型：纳税人模型（Taxpayer's Building）、参考模型（Reference Building Model）和推荐模型（Proposed Building Model）。纳税人模型为被审查建筑。参考模型为满足 ASHRAE 90.1-2001 中对照明、围护结构等最低要求的模型，是判定被审查建筑是否合格的下限。建议模型是内部参数得到更加优化的模型，也是能源利用效率较高的模型，政府以此模型来确定被审查建筑是否有资格获得减税的优惠政策。

### 1.2.2 采用多元拟合回归技术建立建筑能耗基准模型

目前在设计阶段，按照相关节能设计标准的要求，设计方案的建筑能耗应进行节能审查，不经过节能审查的设计方案将无法得到批准，以保证设计建筑达到节能标准要求。

众所周知，建筑只有投入运行之后才能产生能耗。在建筑的全部生命周期内，设计、建造阶段，是一个需要持续投资的阶段。在这个阶段中，需要保证建筑按照绿色和可持续性进行设计、施工。在建筑竣工投入使用后，长期的运行阶段才是检验建筑是否节能、是否绿色、是否具有可持续的重要阶段。一个按照绿色标准进行设计的建筑，只有在运行阶段实现“绿色运行”，实现对高效率、高水平的能源利用，这样节能效果才是持续、长期和不断累加的。

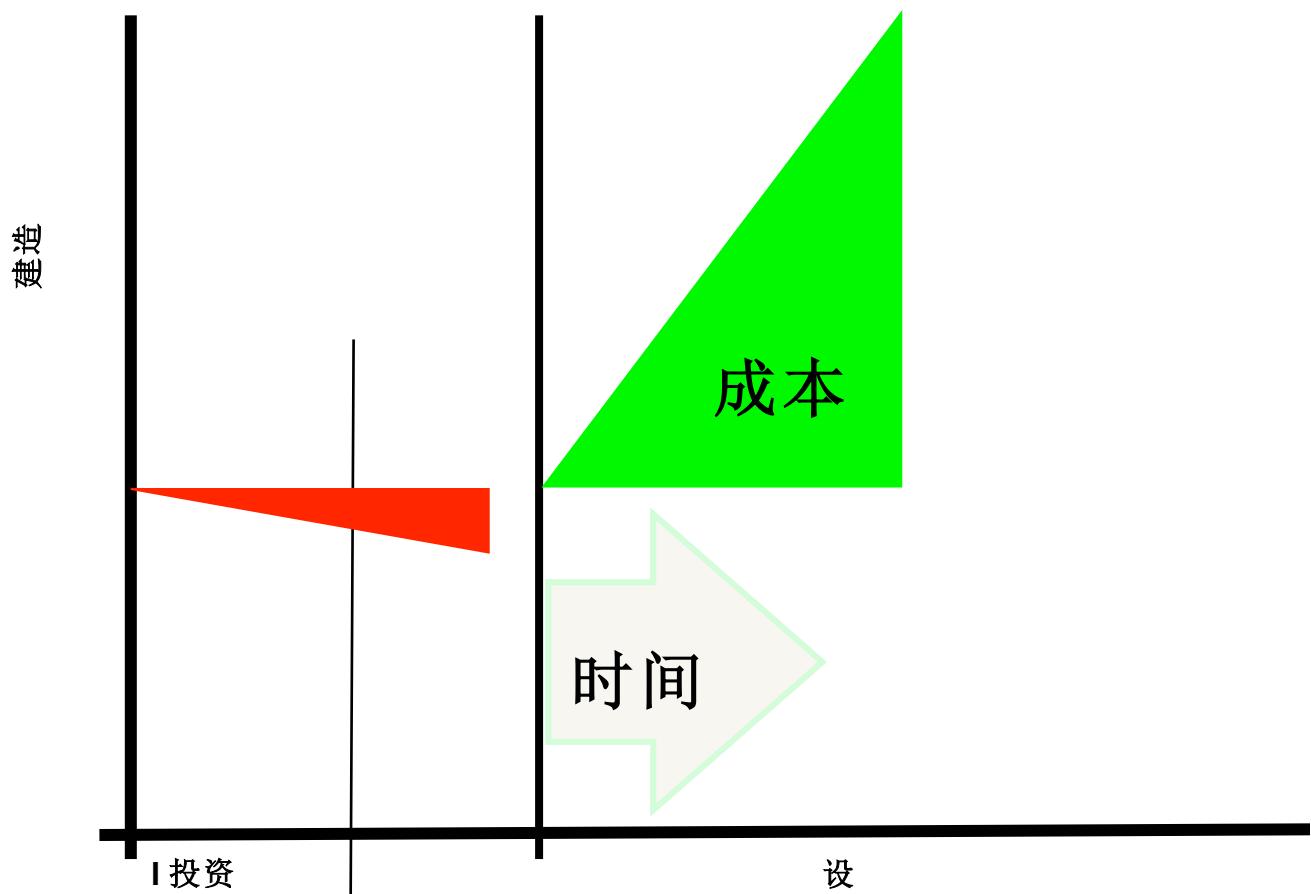


图2 可持续建筑的解释

建筑在运行阶段的实际能耗管理的重要性逐渐得到重视。国内外很多绿色建筑评价体系已经从关注于设计阶段的评价或标识，转入在运行阶段对实测能耗水平的评价和标识。例如，LEED 针对既有建筑的 LEED-EB 评价体系，需要建筑实际能耗水平的评价<sup>(10)</sup>。我国的建筑能效标识也明确提出，应对建筑进行实测能耗的标识<sup>(11)</sup>。

建立运行阶段的能耗基准模型，国外一般采用多元回归的方法。国内学者在这方面做了一些研究工作。

谢艳群<sup>(12)</sup>通过调查实测获得长沙市 57 个住宅能耗样本，通过多元线性回归分析建立建筑总能耗模型和单位面积建筑能耗模型，模型的拟合优度不高，从“样

---

---

本数较少、能耗影响因素的多样性、调查数据的不完整及获得数据的误差等方面”分析造成模型拟合优度不高的原因。

严智勇<sup>(13)</sup>用 13 栋办公建筑的能耗建立多元线性回归模型, 变量包括建筑面积、人数、建筑朝向、窗墙比等参数, 该模型拟合优度较好。

郑晓卫<sup>(14)</sup>等人简要介绍了对上海市商用建筑信息数据库和商用建筑能耗预测模型的研究工作, 通过上海市商用建筑信息数据库, 收集了约 90 栋建筑的能耗信息, 并计划建立能耗预测模型。

韩连华<sup>(15)</sup>从计算理论的角度, 提出了用多种数学模型建立建筑能耗模型的可能。包括采用关联挖掘技术的 APriori- rule 模型和数据挖掘的 ID3 模型。

### 1.2.3 采用其他统计方法研究建筑能耗定额

在国内, 政府主管部门为了便于建筑能耗的监督和管理, 提出应开展建筑能耗定额的研究工作。国内学者主要从统计方法, 提出了一些建立建筑能耗定额的思路和方法。

明雷<sup>(16)</sup>等提出针对标杆建筑能耗的拆分和模拟计算, 建立重庆市公共建筑能耗技术定额的方法。

周智勇<sup>(17)</sup>等提出编制公共建筑能耗定额的三种方法: 平均值法、二次平均值法和回归法, 分析各类方法的异同。在回归法分析中, 采用 22 栋建筑的能耗数据建立了模型。

徐强<sup>(18)</sup>提出用建筑能耗的统计概率分布来确定上海市公共建筑能耗定额的思路, 采用上四分之一分位数、下四分之一分位数或均值加标准偏差的方法确定能耗定额的限制。

徐强<sup>(19)</sup>等提出基于多元线性拟合回归模型的“归一化”方法, 考虑在各自变量完全独立的情况下, 所有自变量分别对建筑能耗所产生影响最大值的总和作为能耗定额的限制。使用 81 栋建筑, 对宾馆、办公和商场类建筑建立了由 2 个自变量组成的回归模型。

刘刚<sup>(20)</sup>等提出用定额水平法建立深圳市公共建筑能耗定额的思路, 即使用算术平均值加标准偏差作为定额限值。

## 1.3 国外建筑能耗基准研究现状

欧美发达国家经过多年的发展,已经建立了相对成熟的建筑能耗基准评价方法和管理制度。一般采用基于网络的在线评估工具,便于能耗评估工作的推广。用户只需要将包括建筑能耗在内的一些基本信息输入到在线评估工具中,可以迅速的计算出该建筑的能耗等级。美国能源部对全球部分国家和地区使用的建筑能耗评价工具进行了统计<sup>(21)</sup>,在北美地区的能耗基准评价工具有17种、亚太地区4种和欧洲地区3种。这些评价工具,包括由各国家政府制定的评价工具,例如:澳大利亚政府制定的NABERS(National Australian Built Environment Rating System)评价体系<sup>(22)</sup>和奥地利能源署制定公司级别能源基准工具(Energy Benchmarking at Company Level);也包括仅适用于各地区、由地方政府和科研机构制定的评价工具,例如:应用于美国加州的Cal-Arch工具,由美国橡树岭国家实验室制定的、应用于爱荷华州、科罗拉多州和佛罗里达州所有学校建筑的“Oak Ridge比较工具”(Oak Ridge benchmarking tools);还包括为大型公司、联锁商业公司制定的评价工具,例如:针对太平洋电气公司(PG&E)的客户制定的CustomNet评价工具,为ABB公司提供的Energy Profiler Online工具等。

以下将选取各地区、具有代表性的几种建筑能耗基准评价工具进行介绍,最后给出各种工具特点的比较和分析。

### 1.3.1 美国能耗基准研究现状

美国的ENERGY STAR® Benchmarking Tools®是ENERGY STAR®项目的一部分<sup>(23)</sup>。ENERGY STAR®是由美国环保署(EPA)和能源部(DOE)联合主持的项目,旨在通过一系列高效节能产品和服务实现能源费用的节约和保护环境。该项目始于1992年,由美国环保署发起,作为一种自愿性标识推向市场,用于判定产品是否高效节能有利于减少温室气体排放。电脑和显示器成为第一批获得该标识的产品。到1995年,该项目已扩展到对办公用设备和住宅用制冷和制热设备的认证。到1996,该项目与美国能源部合作开展工作,将认证范围扩大为办公设备、照明、家用电器,并扩展到包括新建住宅、商业和工业建筑在内的建筑行业中。已有超过30000座商业建筑采用“能源之星”进行测评,其中3200多座建筑得到了ENERGY STAR®认证。

美国环保署建立Benchmarking Tools,在全国范围内,与其他类似建筑相比,

---

---

以帮助建筑业主获取自己所在建筑能源效率的水平。该方法建立起 1~100 的评分体系，可以便于用户很快理解自己所在建筑的能耗水平。50 分表示平均水平，75 分表示不错的水平。拥有多栋建筑的业主或管理者也可以与全国范围内其他类似建筑比较的同时，在所管理的这些楼之间进行能耗水平的比较。并且，通过这些比较，建筑业主和管理者可以方便获得最佳节能潜力，即该建筑能耗的目标水平值，以确定哪些具体的措施可以实施。

最初针对于报告建筑的 Benchmarking Tools 于 1999 年发布，目前已经扩展到包括酒店、医院、学校在内的 13 种建筑类型。目前，该评价方法适用于全美国 60% 的建筑面积，并在持续更新和完善。

每一种建筑都有独特的属性，并对该类建筑能耗水平造成影响。基于这种思路，EPA 意图建立起准确、平等和真实数据的评价方法。该方法包括以下必要的组成部分。

a、使用具有代表性的全美建筑能耗数据库 CBECS<sup>(24)</sup> (Commercial Buildings Energy Consumption Survey)。美国能源部下属的建筑能源信息局在全美范围内，每隔四年对约 6000 栋建筑的能耗进行数据更新。该数据库是评价方法的基础，但并非全部。

b、为保证比较的公平性，统一将所有能源形式转换为一次能源，可以考虑制造及输送效率。

c、采用基于数学统计的方法建立建筑能耗基准模型，将各种影响到建筑能耗的关键性因素考虑在内，包括建筑面积、气象参数、运行时间等便利。

d、利用全国性建筑能耗数据库确定建筑能源利用效率的概率分布情况，以此建立查询表格，根据建筑实际能耗与模型预测能耗的比值查询其所对应的水平。

Energy Star 给出美国各类型建筑能耗基准确定方法，根据建筑的功能和数据统计分析的需要，给出相应的数据筛选范围。以办公建筑 (Energy Star 的办公建筑包括普通办公建筑、银行、金融中心以及法院) 为例，Energy Star 要求办公建筑面积在 5,000 到 1,000,000 平方英尺之间，需连续 12 个月以上的能耗数据，每周设备运行需要 30 小时以上等。Energy Star 应用最小二乘法，以耗能强度 (Energy Use Intensity) (以下简称 EUI) 作为因变量，影响因子作为自变量，拟

合自变量与因变量之间的回归方程。

### 1.3.2 与 ENERGY STAR® Benchmarking Tools®类似的其他工具介绍

美国 ENERGY STAR® Benchmarking Tools®基于数学统计进行评分的技术思路,由于该体系建立时间早、影响范围广和市场推广好等因素,在世界范围内很多地区得到较好的认可,很多国家和地区的建筑能耗评价方法都采取了类似或完全相同的技术思路,即基于庞大的建筑能耗数据库、采用数学统计的方法建立模型、基于网络的在线评分系统。以下是对这些类似工具的介绍。

美国、加拿大地区还有应用在某个地区或者某类型建筑的评价工具,例如,由劳伦斯伯克利实验室开发的建筑评价工具 Cal-Arch,与 Energy Star 的方法类似,应用 CEUS(California Commercial End-Use Survey)数据库(包括 PGE、SCE、SDGE 和加州电业公司提供的调查数据),通过输入建筑的参数,它统计出相似建筑能耗指标的分布情况,应用直方分布图的形式比较被测建筑的 EUI 与相似建筑的 EUI。

在欧盟的 EU-ASEAN Energy Facility 项目的支持下,由新加坡国立大学开发了 Energy Smart Tool 工具<sup>(25)</sup>。该工具与 ENERGY STAR® Benchmarking Tools®的功能和技术路线基本一致,但由于新加坡地域较小,采用的建筑能耗数据库包括了 95 栋办公建筑、30 栋酒店建筑和 59 个工业建筑。与 ENERGY STAR® Benchmarking Tools®不同的是,该工具不需要提供连续一年十二个月每月的能耗和费用数据。需要用户提供的自变量也略有不同。

由于认识到建筑能源效率的重要性,香港机电工程署(EMSD)下属的能源效率办公室进行了在香港对选定建筑组群开发能耗指标和基准工具的咨询性研究。该项目提供了一定深度的建筑能耗数据和建筑内各设备的运转效率参数,不仅可以便于政府部门制订未来建筑节能减排的目的,也为用户提供了为建筑能耗水平进行评分的在线工具<sup>(26)</sup>。与 ENERGY STAR® Benchmarking Tools®不同的是,该工具需要用户提供更加详细的建筑内设备及系统运行参数等信息,包括日常室内维持温湿度的数值,主要冷冻设备运行情况等参数。

在其他国家和地区,还有这样类似的工具。例如,加拿大的在线能耗基准评价工具 e.Review 和商业建筑能耗基准评价工具 EnerPro。

---

---

在一些大型企业和高校，也建立了以统计数据模型为基础的基准评价工具。例如，美国加州的太平洋煤电公司（PG&E）为其用户所建立的 CustomNet 工具<sup>(27)</sup>，美国劳伦斯伯克利国家实验室为评价实验室能耗水平所建立的 Labs21 工具<sup>(28)</sup>等。在美国加州 CalArch 工具网站上，归纳了 17 种全球范围内类似的建筑能耗评价基准工具<sup>(29)</sup>。

### 1.3.3 英国能耗基准研究现状

英国的“政府能效最佳实践项目”<sup>(30)</sup>，英文全称为“Government Energy Efficiency Best Practice programme”，简称为“EEBPP”，始于 1989 年 4 月。该评价方法与美国 ENERGY STAR® Benchmarking Tools®都属于采用数据统计的方式进行建模，都需要大量的、实际的建筑能耗数据，都认为前 25% 水平为较理想水平。但在技术体系和评价方法上存在较多的不同点，包括以下内容。

a、将同一类建筑进行二次分类

建筑类型划分更加细致。例如，将同一类医院建筑根据特点的不同分为四类(9)：大型医疗机构（Teaching and specialist）、专科医院（Acute and maternity）、医疗站（Cottage）和疗养院（Long Stay）；将同一类办公建筑根据特点的不同分为四类：自然通风的多孔建筑（cellular）、自然通风的敞开式建筑（open-plan）、标准空调建筑（standard）和豪华空调建筑（prestige）。

b、分别对能耗量和能耗费用水平的高低进行评价

定义了基准指标，分别采用 £/100m<sup>3</sup>/a 和 GJ/100m<sup>3</sup>/a 两个不同的单位，对建筑能源费用和能源消耗量的水平进行评价。

c、分别对化石能源和电力消耗进行评价

未将不同能源统一折算为一次能源。分别给出对应的、化石能源和电力的消耗量和费用的基准指标数值。

d、分别计算各分项能耗的指标，通过合成，得到总的建筑能耗指标

为了计算总的能耗指标，需要对该项目的能耗进行分解，对其中部分分项能耗指标的选取过程进行选取和计算。

对于受其他因素影响较小的能耗，如供热能耗和基础能耗，该方法直接给出了表格，根据医院类型和能源类型直接选取对应指标

对于容易受到外界其他因素的干扰,如 HVAC 系统能耗,给出在不同系统配置情况下的能耗指标,需要根据实际情况进行选择和计算。

由于引入了用户自定义的这些分项能耗指标,才使计算出来的总能耗指标更加科学、合理,符合该建筑的运营特点。

通过总能耗值和能源费用值,分别与基准指标进行比较,比较并分析化石能源和电力的消耗和费用高的潜在因。通过各分项能耗实际值与指标值的比较,还能更加明确能耗的重点和潜力,有利于指导该项目下一步建筑节能工作的开展。

#### e、解决复合型建筑的能耗指标确定

例如,在办公建筑能耗评价方法中引入了复合型办公建筑的判定<sup>(31)</sup>。由于办公建筑经常由多种类型建筑复合而成,该方法认为不同类型建筑的运营时间不同,因此需要根据各类型建筑面积所占比例,对该复合建筑的运营时间进行修正。修正公式如下:

$$\left( \left( \frac{A}{100} \times \frac{48}{I} \right) + \left( \frac{B}{100} \times \frac{58}{I} \right) + \left( \frac{C}{100} \times \frac{62}{I} \right) + \left( \frac{D}{100} \times \frac{67}{I} \right) \right) = Z$$

其中, A、B、C、D: 分别是 4 大类型办公建筑所占比例, %;

I: 该复合型建筑的实际运营时间;

48、58、62、67: 分别是 4 大类型办公建筑运营时间的典型值;

Z: 该复合型建筑运营时间的修正系数。

#### 1.3.4 德国能耗基准研究现状

德国建筑节能的法律法规非常全面,这与德国整体的法律体系非常健全是分不开的。相关的法律基础包括: BGB-《德国民法典》、EnEV-《建筑节能法规》和 VOB-《建筑工程承包、发包条例》等。技术基础包括《供暖设备法规》、VOB 标准、DIN 标准和 VDI 标准等<sup>(32)</sup>。

为了对建筑能耗进行管理和评定,德国的建设部与德国工程师协会共同制定并发布了 VDI 3807 系列标准。VDI 3807 标准全名为《建筑能源消耗的特征数值》,由四部分标准文件组成,分别为:

VDI 3807-1: 基础。包括定义与计算方法;

VDI 3807-2: 供热和电力整体能耗的特征数值。针对这两类能耗给出基准值

的计算方法；

VDI 3807-3：建筑内及附近地区水资源消耗的特征数值；

VDI 3807-4：电力消耗的特征数值。对电力系统的能耗进行详细分析和节能潜力评估。

在 VDI 3807-2 中，对建筑整体能耗提出了采用与基准值进行对比的方式评定建筑的能耗水平。首先提出“平均值”和“指导值”的概念。“平均值”，不是通常意义理解的算术平均值，而是通过数学统计的方式，建立建筑能耗频率分布模型，该模型中分布最密集的区域对应的为平均值。该指标比算术平均值更具有代表性，因为算术平均值容易受到异常点的干扰，具体数值会朝异常值方向偏移。详见图 3。

“指导值”，可理解为建筑节能的目标值，是建筑能耗分布模型中，能耗位于较低的 25% 区域内建筑的算术平均值。详见图 4。

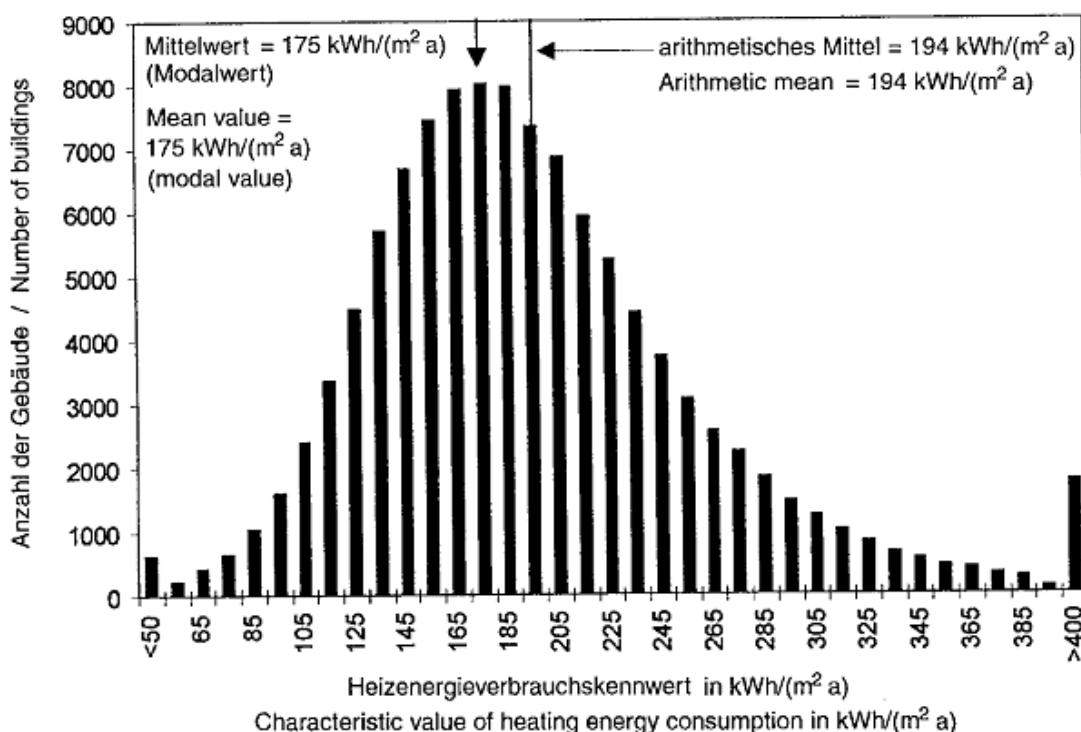


图3 建筑能耗频率分布模型和平均值的定义

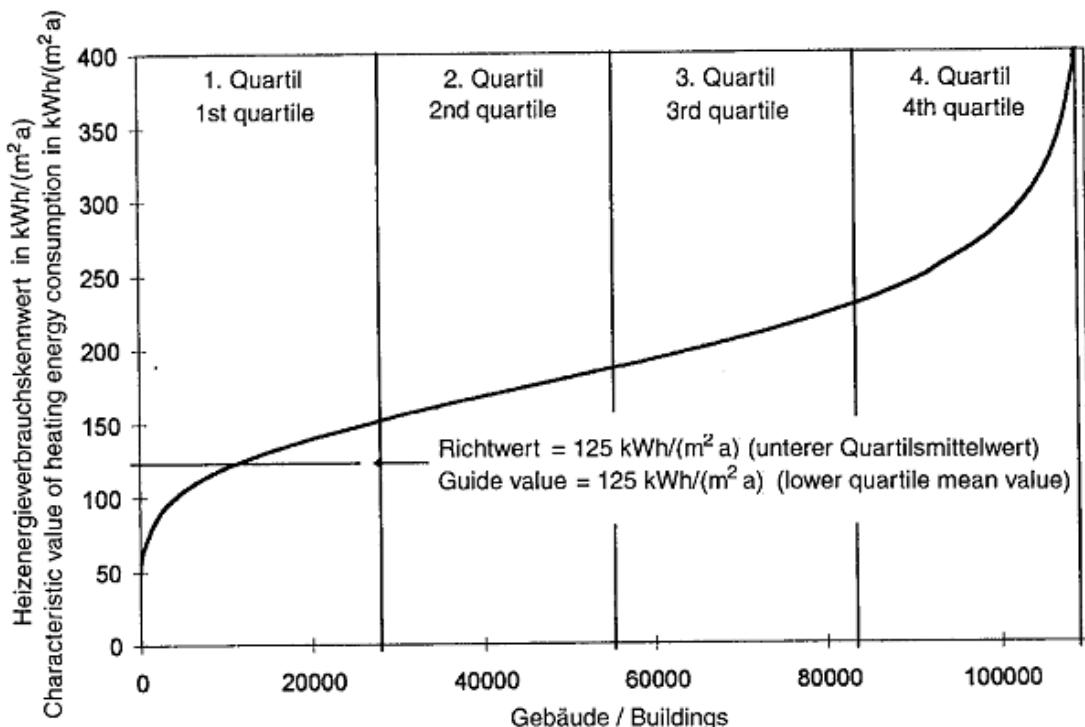


图4 建筑能耗数量分布模型和指导值的定义

建筑能耗频率分布模型和数量分布模型的建立和建筑能耗平均值、指导值的计算,都是基于大量实际的建筑能耗数据。德国工程师协会使用了不同来源、1993年~1995年期间的建筑能耗数据,基本涵盖德国全境。

建筑类型的差异对建筑能源消耗的影响较大,因此德国对建筑进行了非常细致的分类,共分为15大类,按照上述的数据和计算方法,分别计算出供热和电力消耗的平均值和指导值。这15大类将建筑分为法院、政府建筑、医院、学校等建筑。对其中部分建筑大类进行了更加细致的二次分类。例如,将政府建筑又细化为高级州政府和地方政府建筑、普通行政建筑、大型会议中心、人才中心、税务大楼和警察局,也分别计算出对应的平均值和指导值。

本标准还对普通住宅建筑计算出基准值。显然,由于住宅建筑的差异更加突出,在计算基准值时,将住宅进行了类似细致的分类。

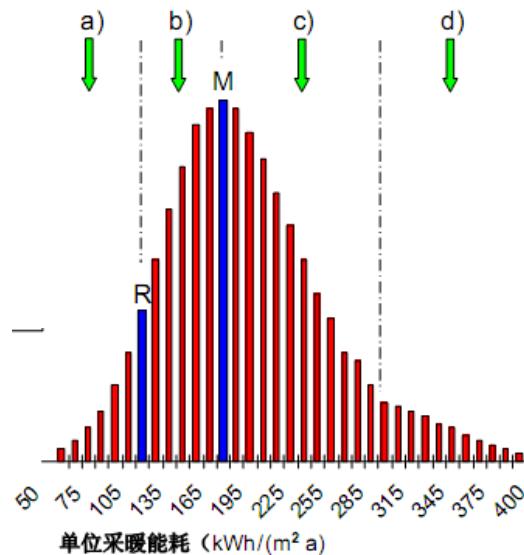


图5 建筑能耗频率模型和建筑能耗水平的判定区域

建筑能耗基准的建立可用于对建筑的能耗水平进行评估,指出哪些建筑应该进行改造,哪些建筑处于良好状态。在本标准中,采用建筑能耗频率分布模型对所有建筑进行了分类。如图5所示,在建筑能耗频率模型中,划分四块区域,包括:

- a 区:  $E_{VH} < R$ , 优秀, 表示该区域内建筑能耗水平不错, 不需要改进;
- b 区:  $R < E_{VH} < M$ , 良好, 有提升空间, 不需要改进;
- c 区:  $M < E_{VH} < R+M$ , 普通, 需要采取改进措施;
- d 区:  $E_{VH} > R+M$ , 较差, 必须采取改进措施。

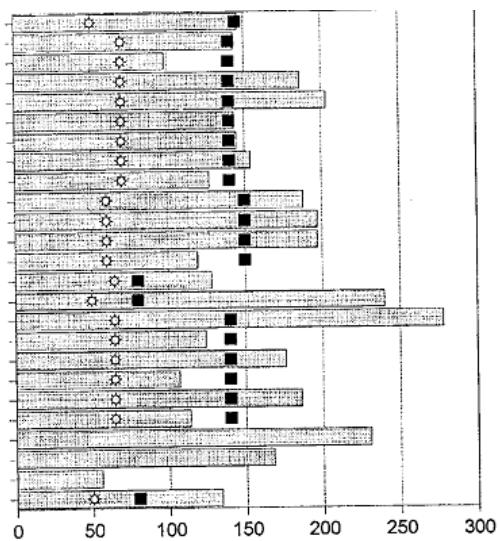


图6 建筑能耗评估( $\text{kWh}/\text{m}^2 \text{a}$ )

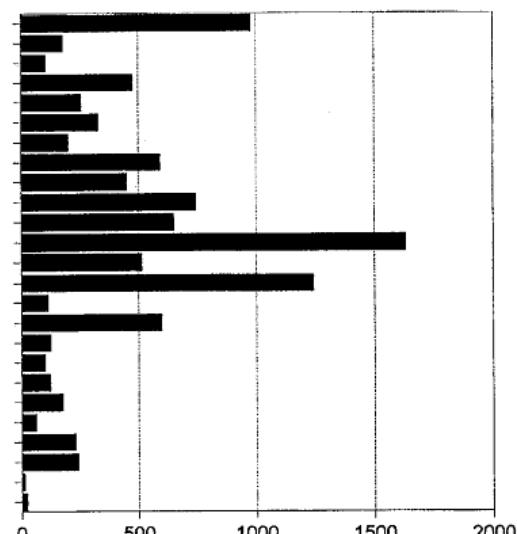


图7 建筑节能潜力( $\text{MWh}/\text{a}$ )

图 6 给出某区域内所有建筑总能耗的评估情况，■代表平均值，□代表指导值。通过建筑实际能耗与平均值和指导值的比较，能够得到每个建筑大概的节能空间的大小，见图 7 所示。

VDI 3807-2 不适于对于用能系统形式复杂的建筑进行整体能耗的评价。需要针对各分项能耗进行详细分析，此时应采用标准 VDI 3807-4。该标准与英国 EEBPP 项目的技术思路类似，将能耗细分，并分别计算分项能耗的基准值。不同的是，英国 EEBPP 项目采用实测数据的统计计算分析能耗的基准值，而 VDI 3807-4 采用简化模拟的方法进行计算：假设设备按照某一合理的运行方式运行，根据设备额定功率和运行参数，计算理想状态下该设备的能耗值，作为基准值，与实际能耗值比较，用于计算该设备分项能耗的节能潜力，同时也用于协助确定节能措施。该标准认为这种计算方法具有一定的精度，但同时也承认，这种简化模拟计算法必须与现场调研和实测数据相结合，期望“最迟在三年之内，用分项能耗的实测值代替简化计算值”。

#### 1.4 课题研究任务

归纳总结国外关于建筑能耗基准的研究工作，发现由于研究时间长、配套的制度体系已经建立健全，在国外该领域已经得到了成熟的发展。基本形成了完善的建筑能耗数据调查制度和包含大量数据的建筑能耗数据库。美国、英国和德国在建筑能耗基准模型建立时，虽然采用了不同的方法和数学模型，但都是基于对大量建筑实际能耗数据的统计和分析，并且数据调查工作和模型建立工作会定期更新，以保持与全社会建筑能耗技术水平发展速度一致。该建筑能耗基准建立的体系证明是完善、实用的，多年来在全球各国的广泛应用，也证明了该技术体系的特点，可以作为我国开展该项工作的主要指导。

从国内目前建筑能耗基准研究领域所做的工作可以看出，人们已经认识到在建筑的运行阶段对能耗情况进行监管的重要意义。虽然很多学者和相关的政府部门在该领域进行了一定工作，但并未形成类似国外的完善的数据调查制度及完备的建筑能耗数据库，最终建立建筑能耗基准模型和评价方法。更多的研究处于试验性探索阶段，研究国外技术路线在中国的适用性问题。

---

---

本课题拟参照国外技术路线，通过对数学统计模型方法的学习和分类，突破本专业工程师所掌握数学统计知识的局限，按照合理的解释，结合本专业知识和数学统计知识，选择适当的数学模型；通过一定数量规模的酒店类建筑能耗数据和基本信息，建立适合我国建筑能耗技术特点的基准模型和评价方法。

## 第二章 公共建筑能耗基准模型的建立方法

在确定建立能耗基准模型的技术路线前,有必要对我国目前在公共建筑能耗分析方面所做的研究工作进行汇总,虽然不能完全代表所有的研究工作,但本文仍希望通过概况总结,为提出能耗基准模型技术路线,提供一些应用的背景知识。

### 2.1 我国公共建筑能耗分析研究现状调查

《公共建筑节能设计标准》条文说明 1.0.2 节,将建筑划分为民用建筑和工业建筑,其中民用建筑又分为居住建筑和公共建筑。公共建筑则包含办公建筑(包括写字楼、政府部门办公楼等)、商业建筑(如商场、金融建筑等)、旅游建筑(如旅馆饭店、娱乐场所等)、科教文卫建筑(包括文化、教育、科研、医疗、卫生、体育建筑等)、通信建筑(如邮电、通讯、广播用房)以及交通运输用房(如机场、车站建筑等)。

根据现行《民用建筑热工设计规范》将我国建筑气候划分为严寒地区、寒冷地区、夏热冬冷地区、夏热冬暖地区及温和地区,详见图 8。由于不同气候区温湿度、降水量、日照及各季节时间长短的不同,导致其建筑能耗存在一定的差异。

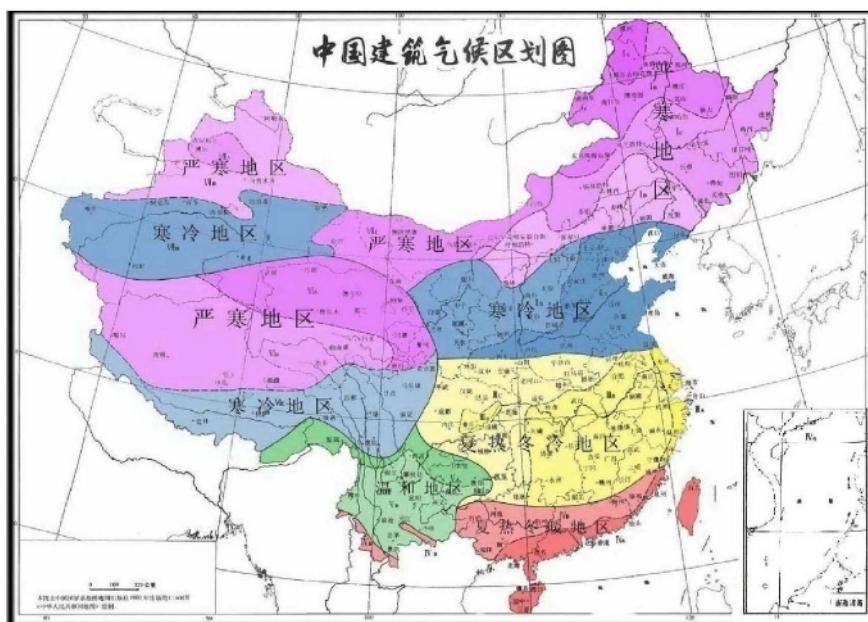


图8 全国建筑气候分区

本节拟在国内部分学者的调查研究基础上,汇总各学者的调查数据,用于比较我国寒冷地区、夏热冬冷地区和夏热冬暖地区办公建筑(办公楼和写字楼)、

商业建筑（商场和超市）、旅游建筑（宾馆酒店）和科教文卫建筑（学校和医院）的建筑能耗。

### 2.1.1 我国公共建筑能耗分析研究数据汇总

选取寒冷地区、夏热冬冷地区和夏热冬暖地区典型城市天津、大连、西安、长沙、杭州、成都、浙江、广州、福州和深圳等 10 座城市<sup>(33)(34)(35)(36)(37)(38)(39)(40)(41)(42)(43)</sup>，共 1043 栋建筑为研究对象，使用 2007 年建筑能耗数据。不同气候区、各类公共建筑样本分布如图 9 和图 10 所示。寒冷地区使用的能源主要为电、煤和燃气，夏热冬冷地区消耗的能源有电、燃气、柴油、煤、热力，夏热冬暖地区使用的能源有电、燃气和柴油，且这三类气候区中电能均为主要能源。

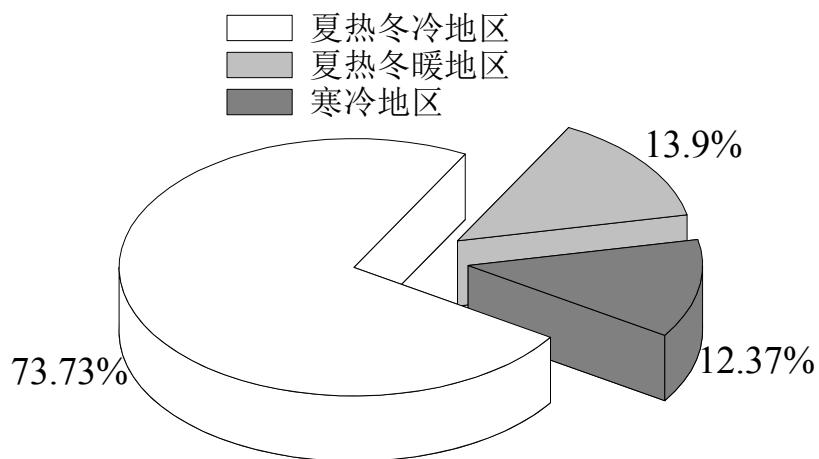


图9 不同气候区调查样本分布

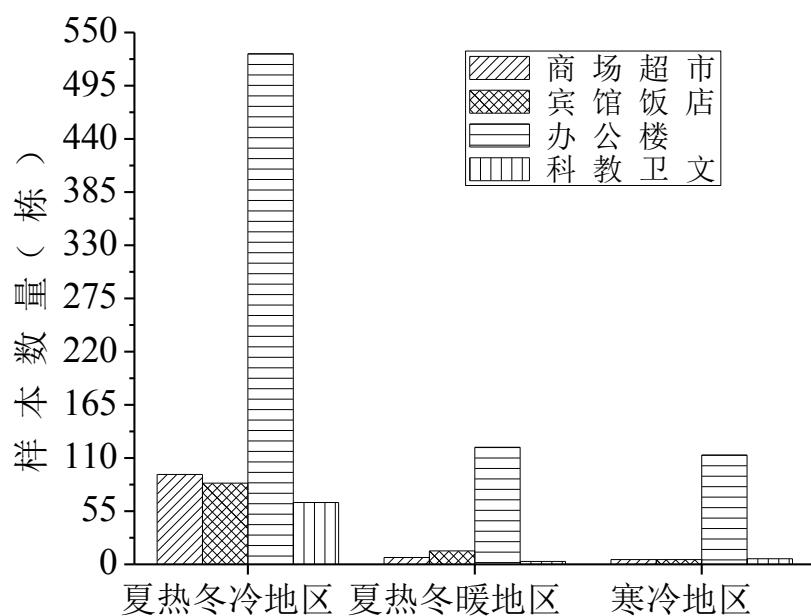


图10 不同气候区各类公共建筑样本分布

图 11 为 3 个气候分区公共建筑单位建筑面积年平均能耗。寒冷地区公共建筑单位建筑面积年平均能耗范围为  $18.11\sim127.12\text{kgce}/(\text{m}^2\cdot\text{a})$ ，平均值为  $46.05\text{kgce}/(\text{m}^2\cdot\text{a})$ ；夏热冬冷地区单位建筑面积年平均能耗范围为  $2.89\sim404.59\text{kgce}/(\text{m}^2\cdot\text{a})$ ，平均值为  $64.74\text{kgce}/(\text{m}^2\cdot\text{a})$ ；夏热冬暖地区为  $4.88\sim141.85\text{kgce}/(\text{m}^2\cdot\text{a})$ ，平均值为  $44.86\text{kgce}/(\text{m}^2\cdot\text{a})$ 。

表 1 给出寒冷地区、夏热冬冷地区和夏热冬暖地区不同功能公共建筑单位建筑面积年平均能耗的范围。图 12 为 3 个气候区不同功能建筑单位建筑面积年平均能耗值。由图 5 可知，(1) 3 个气候区不同功能建筑单位建筑面积年平均能耗有相同的变化趋势，从高到低依次为：商业建筑、旅游建筑、办公建筑和科教文卫建筑。寒冷地区、夏热冬冷地区和夏热冬暖地区中商业建筑平均单位建筑面积年能耗是科教文卫建筑的倍数依次为：4.27、3.45、2.37。(2) 对同一功能建筑而言，夏热冬冷地区四类建筑的能耗均最高，寒冷地区商业建筑和科教文卫建筑的能耗最高最低，而办公建筑和旅游建筑能耗最低的为夏热冬暖地区。不同气候区、不同类型公共建筑的能耗存在较明显的差异，应按照不同的建筑用能定额进行建筑节能评价。

表1 不同气候区各类公共建筑单位建筑面积年平均能耗的范围

气候区	单位建筑面积年平均能耗的范围 ( $\text{kgce}/(\text{m}^2\cdot\text{a})$ )			
	办公建筑	商业建筑	旅游建筑	科教文卫建筑
寒冷地区	21.02~114.28	78.25~127.12	18.11~86.87	30.16~72.78
夏热冬冷地区	34.86~151.59	95.89~404.59	64.35~182.5	2.89~150.14
夏热冬暖地区	4.88~80.27	31.24~141.85	33.64~139.59	16.14~51.55

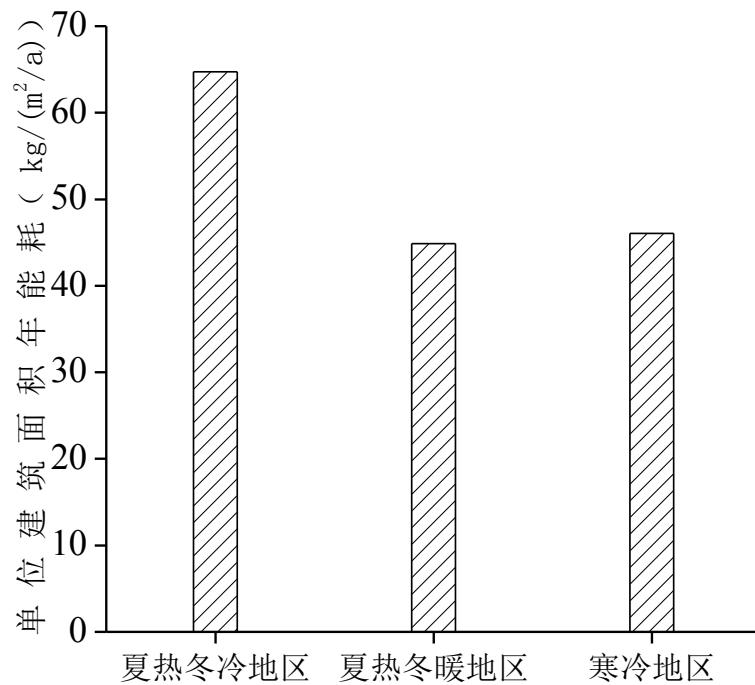


图11 不同气候区公共建筑单位建筑面积能耗值

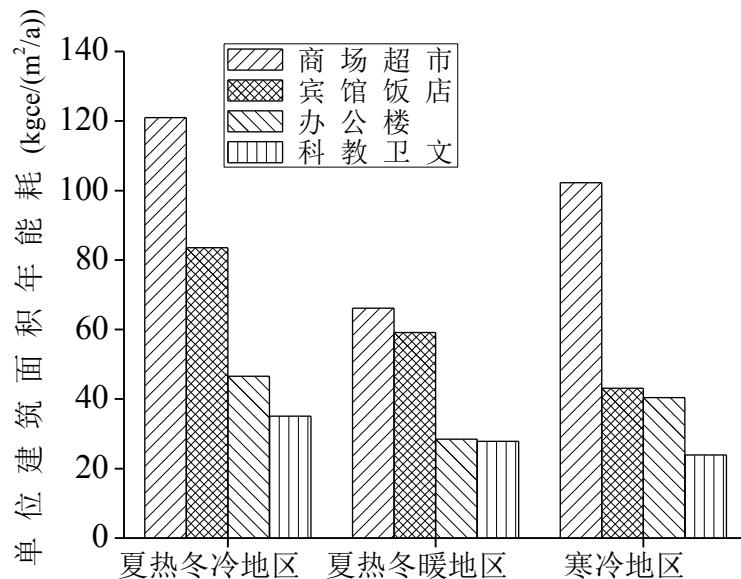


图12 不同气候区各类公共建筑单位建筑面积能耗值

此外，通过对不同地区公共建筑电能消耗量的统计发现，夏热冬冷地区单位建筑面积年平均电耗为  $33.92\text{kgce}/(\text{m}^2\cdot\text{a})$ ，占其总能耗的 52.39%；寒冷地区为  $38.64\text{kgce}/(\text{m}^2\cdot\text{a})$ ，占其总能耗的 59.45%。由于夏热冬暖地区的几个城市中有部分没有给出电能消耗的详细情况，所以未对该气候区电能进行统计。由此可发现，我国建筑能耗中存在的普遍现象是电能消耗在公共建筑能耗中所占的比例非常大，为了有效控制和降低公共建筑能耗，耗电设备的选取及室内温湿度环境的控

制应首先得到关注。

### 2.1.2 我国公共建筑能耗分析研究结论汇总

国内在公共建筑能耗分析研究方面的结论可分为以下几类：

(1) 提出某地区某类型建筑全年单位面积能耗的平均值

王树健<sup>(44)</sup>等提出重庆市商场建筑平均能耗面积密度为  $295.61 \text{ kW h}/(\text{m}^2 \cdot \text{a})$ ;

(2) 进行地区间某类型建筑单位面积全年建筑能耗比较

“能耗密度分别为北京的 1.13 倍，天津的 1.28 倍，日本神户的 1.05 倍”<sup>(44)</sup>；

“所调查的公共建筑总体平均单位面积能耗低于国外相似气候条件下的同类型建筑”<sup>(45)</sup>。

(3) 相同建筑类型的不同建筑能耗差异巨大

“不同商场之间能耗差异很大，最高能耗商场建筑能耗指标是最小能耗建筑的 1.94 倍，相互之间相差较大，节能潜力巨大”<sup>(44)</sup>。

谭颖等<sup>(46)</sup>提出“这 15 家商场类建筑能源，……，可见相互之间的差异是非常大的”。

(4) 建筑之间进行能耗比较的不合理性

龙惟定<sup>(47)</sup>等提出“在建筑节能评价时，一般采用单位面积年平均一次能耗量作为评价的指标，本文就是采用了这一办法。但是平均能耗量并不能反映建筑物的使用水平。试想，一家五星级宾馆和一家无星级的普通旅馆，能耗怎么可能一样呢？把它们放在同一标准下进行比较，只能得出带有片面性的结论。”

我们可以看到，建筑能耗分析工作给出数据，和给出的最终结论，都是从宏观角度提出。或者是给出多个建筑的平均能耗，或者是该平均能耗与其他地区相比的结果如何。但这些结论和数据如何用于指导一栋具体的建筑进行节能潜力评估，或者进行节能改造，或者进行优化运行？均值只能代表被调查样本的特征，很难对整体进行描述。特别是当样本数较少时，再拿均值当作一个重要的指标，多少有些不严谨。本文第三章对于均值的意义有详细的解释，不再赘述。

我国公共建筑能耗分析工作如果仅从宏观分析着手，给出一些不痛不痒的数据，不进入一栋建筑中做具体分析，那其发展和作用将十分有限。

从很多具体的建筑能耗分析中很容易看出，即使是同样的建筑类型，不同单

体建筑的单位面积能耗差异常常是巨大的，可能会高达两倍。那么此时，使用单一基准数值，不论是用算术平均值，或是用中位数，来衡量建筑的能耗水平，都存在一定偏差。

选取我国寒冷地区 16 栋办公建筑在 2009 年 12 月~2010 年 11 月连续 12 个月的建筑能耗数据为例，对公共建筑能耗特点进行分析说明。图 13 给出了这 16 栋办公建筑全年单位面积的实际能耗，按照从小到大顺序排列的情况。

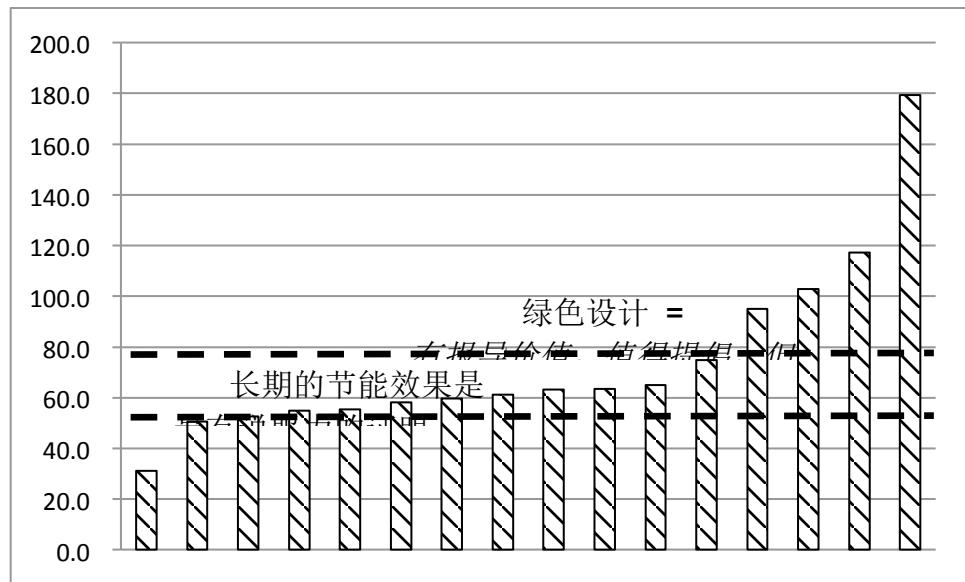


图13 办公建筑单位面积能耗 (kWh/m<sup>2</sup>/年)

16 栋办公建筑的总能耗平均为 240.5 万 kWh，最大值为 537.8 万 kWh，最小值为 62.2 万 kWh；单位面积能耗平均为 79.6kWh/m<sup>2</sup>/年，最大值为 179.3kWh/m<sup>2</sup>/年，最小值为 50.7kWh/m<sup>2</sup>/年。

从图 13 中可以看到：

- (1) 建筑之间的能耗差异巨大，可达 3.5 倍；
- (2) 绝大多数建筑能耗水平接近；
- (3) 由于能耗异常点的存在，导致平均值 (79.6kWh/m<sup>2</sup>/年) 偏移。如果以平均值作为能耗指标，以此作为准绳判定其他建筑是否节能，那么从图 13 中可以看出，11 栋建筑（占 68.8%）是节能的，5 栋建筑（占 31.2%）有较大提升空间。但由于建筑能耗差异巨大，将能耗最高建筑的能耗值，从 179.3 降至 79.6，从技术上和从经济上分析，都存在巨大困难。那么用平均值作为建筑能耗判定的依据，其公平性还值得商榷。

(4) 25%分位点的定义为: 25%的建筑能耗值小于或等于该数值。从图中可以看出, 绝大多数建筑能耗在 25%分位点附近, 其代表性比平均值要高。但由于建筑能耗的巨大差异, 存在上述(3)中所述公平性的问题。

(5) 由于建筑运营特点的不同, 建筑之间能耗差异巨大是普遍现象。

详细比较能耗最高建筑 TF 和能耗最低建筑 HB 的分项能耗, 特殊用电能耗的逐月对比详见图 14。TF 建筑的特殊用电能耗远高于 HB 建筑。经过核查发现, TF 建筑位于信息产业园区, 该建筑包括 IT 机房等信息设备机房, 特殊用电高直接造成总能耗偏高。而 IT 机房的节能空间有限, 并且 IT 机房的设置完全属于建筑运营的需要, 不能因此判定 TF 建筑能源利用效率低。

对应的 HB 建筑属于普通办公建筑, 无 IT 机房, 特殊能耗较低。但 HB 建筑内人数为 730 人, TF 建筑内人数为 3000 人。因此, 从这个角度上来分析, HB 建筑能耗低并不能证明其能源利用效率高。

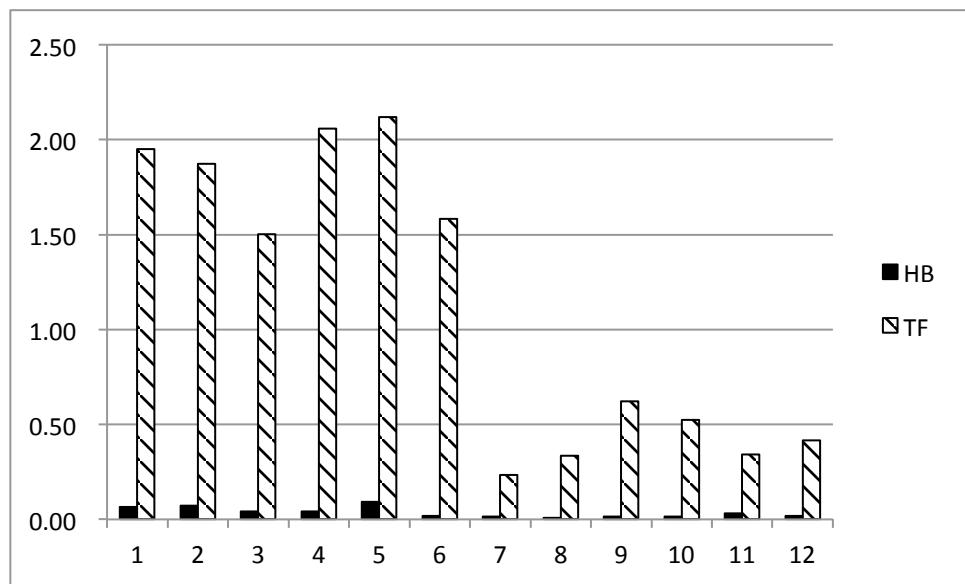


图14 TF 建筑和 HB 建筑特殊能耗逐月分布图

选取另外两栋能耗差异较大的建筑 FE 和 KF 进行比较。建筑 FE 和 KF 的单位面积年能耗为  $59.8\text{kWh}/\text{m}^2/\text{年}$  和  $117.3\text{kWh}/\text{m}^2/\text{年}$ , 后者的能耗几乎等于前者的两倍。详细核查两栋建筑的情况, 发现建筑 KF 的照明插座能耗和特殊能耗几乎为 FE 大厦的两倍。通过图 15 逐月比较两座建筑能耗分布发现, 几乎每月 KF 建筑的能耗都比 FE 建筑高。现场调研发现, FE 建筑的运营时间为每天 9 小时, 每周 5 天, 每周运营时间 45 小时。而 KF 建筑中有部分业主经常晚上和周末加班,

因此运营时间为每天 12 小时，每周 7 天，每周运营时间 84 小时，远高于 FE 建筑。但 KF 建筑这种长时间的运营方式是根据其运营特定形成的，无法甚至很难改变，不能因此认为 KF 建筑的建筑能耗高。

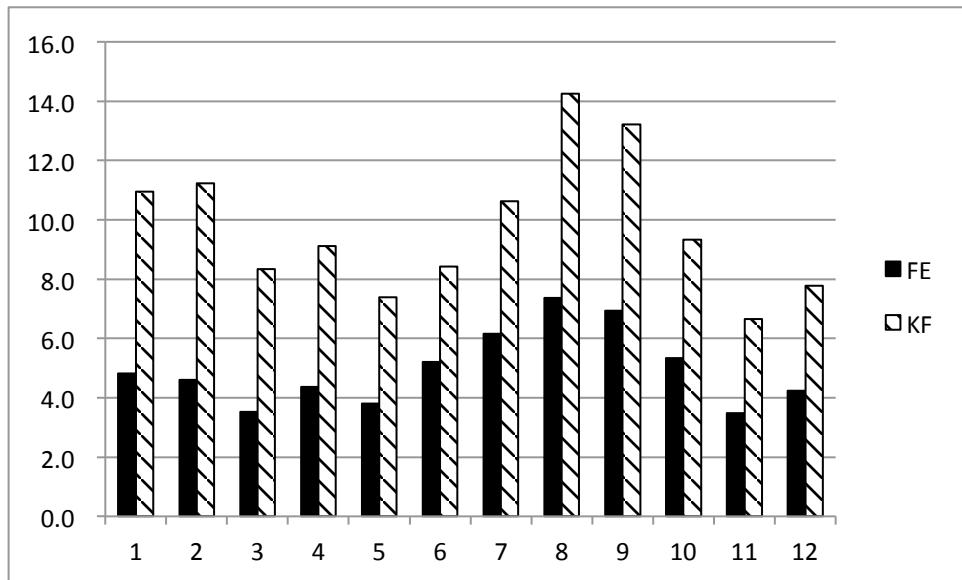


图15 FE 建筑和 KF 建筑能耗逐月分布图 ( $\text{kWh}/\text{m}^2$ )

## 2.2 公共建筑能耗基准模型建立思路的提出

上节已经提及，建筑能耗差异巨大是普遍情况。造成该现象的原因千差万别。显然，用建筑能耗模拟计算软件很难对现实中各种影响因素进行模拟和表述，因此单纯以模拟计算以确定能耗基准是不合理的。但可以结合实际建筑能耗数据，用模拟计算软件对一栋建筑进行详细的模拟分析，进而发现哪些不合理因素造成能耗异常，从而发现节能潜力和措施。

通过文献综述的总结可以看到，美国、英国和德国的技术思路，都是以大量实测数据为依据，采用统计学方法建立建筑能耗基准模型，从大量建筑能耗数据中挖掘特征信息建立模型，针对每一栋建筑的运营特点，用该模型计算能耗基准值。这种建立能耗基准的技术思路，是与我国建筑实际能耗差异巨大的实际情况相符合的。因此，具有充分的公平性和说服力。

目前需要解决两个问题：第一，对建筑能耗产生影响的因素众多，应选择什么类型的因素，用于建立能耗基准模型；第二，国外流行的统计分析方法是多元线性拟合回归分析，但为什么要选择该分析方法，统计学分析工具还有哪些，是否还有更适用的方法。

首先解决第一个问题。对建筑能耗能产生影响的因素众多，可分类为以下几类：

### 第一类：建筑固有特性

建筑固有特性，指受建筑用途所限定，该建筑在实际运行中无法改变的参数，包括建筑面积、酒店运营时间、上班期间工作人员数、是否准备商务用餐等变量。而这些变量对建筑能耗有很大影响，但无法改变。

### 第二类：耗能系统固有特性

耗能系统固有特性，指从技术角度分析，与建筑能耗相关的各类参数。包括围护结构类别、主要耗能设备的类型、是否使用一些节能设备等变量。这些变量可以通过技术措施进行改变，例如外墙加保温、更换高效的冷源设备等。

### 第三类：运行管理控制参数

运行管理控制参数，指该建筑在实际运行时，与耗能系统相关的设定参数和控制策略等。包括室内设定温度和湿度、室内灯具的开启情况、冷源系统的运行策略、耗能系统的维护和保养等变量。这些变量受人为因素影响，在不同建筑之间差异巨大。

公共建筑能耗基准模型，要求具有强大的数据支持，具有公正性和准确性，便于理解，便于推广。由于影响建筑能耗的变量众多，应该选取合理的影响变量，消除不同建筑之间变量的差异，平等地比较建筑能耗。从影响变量的分类分析可以看出，耗能系统固有特性和运行管理控制参数属于可更改变量，可以修正或改善，只有第一类的建筑固有特性是受到建筑用途和运营特点所限制，是无法改变的，因此，应选择这类变量作为建立建筑能耗建筑模型的基础。

从另一个角度分析，建立模型不考虑第二类耗能系统固有特性，对于那些由于耗能系统运行效率不高，造成整个建筑能耗异常的楼宇，保留了从技术上改进和改造的空间。例如，一栋建筑的冷源采用风冷冷水机组，该机组固有 COP 较低，无提升空间。可以通过更换 COP 更好的水冷冷水机组达到降低能耗的效果。例如，一栋建筑建成年代较早，由于围护结构性能差所导致的建筑能耗异常，可以通过增加外墙或内墙保温的措施进行改进，降低建筑整体能耗。如果，将这类因素考虑在基准模型内，那么在分析时，会将这些不合理技术因素进行平均化处

---

---

理，从而掩盖了建筑能耗异常的原因。

建立模型不考虑第三类运行管理控制参数，对于那些由于管理水平低下，耗能系统处于非正常运行的系统，或未按照最优化策略运行的系统，造成整个建筑能耗异常的楼宇，保留了从管理和运行上改进和改造的空间。例如，冬季室内设定温度过高或夏季室内设定温度过低，必然会造成建筑能耗偏高，通过简单的方式即可降低能耗。例如，新风机组未按照优化策略运行，冬季和夏季新风量过大，同样会造成能耗巨大。

因此，不考虑第二类和第三类参数变量对建筑能耗的影响，可以促进建筑的业主和管理者从技术改进和管理水平提高的角度考虑提高建筑的能源利用效率，降低建筑能耗。

计划在第三章详细介绍统计方法的主要分类，对典型的统计方法进行介绍，提出根据应用环境选择最适合统计工具的方法，并最终确定本课题拟采用的统计学分析工具。

### 第三章 数学统计方法的介绍及模型选定

统计学，是基础数学的一个分支<sup>(48)</sup>。通过对数据进行分析，从中提取有用的信息，用于解决实际工作中出现的各种问题。统计学解决问题的思想可以概括为，当某一学科的问题在该专业范围内很难或无法用专业知识进行解释时，可以从实际数据出发，通过对一定数量的样本进行分析，用分析所得到的样本特征，“无限”接近于该问题的本质，对该问题进行有一定偏差的解释。因此，在目前人类科学技术还存在很多空白的情况下，统计学能够解决相当大范围的问题，这也是统计学近年来得到广泛应用的原因之一。

目前，数据的统计分析已成为广大自然科学与社会科学研究人员、经济管理人员和各类工程技术人员最有用的分析工具之一。它研究的主要对象是客观事物变量间的统计关系。通过建立在对客观事物进行大量试验和观察的基础上，用来寻找隐藏在那些看上去是不确定的现象中的统计性规律。它是分析统计模型研究变量间相互关系的密切程度、结构状态、模型预测的一种有效的工具。在各个行业中，数据的统计分析都发挥着重要的作用。例如，用于人口增长的预测，天气变化的预测，股票变化的分析等。在建筑行业，统计学也发挥着重要作用。例如，通过数学模型对房地产市场价格的预测，对建筑发展速度的预测等。

在建筑能耗数据分析方面，国内学者已经做出一些工作。从这些文献中可以看出，目前关于建筑能耗数据的分析方面，相关学者使用较多的数学统计方法都相对简单，所包含的信息较少。包括下文将介绍的平均指标法和指数分析法等，或者对样本数据进行计算，求算术平均值，求中位值，以此作为一定范围内建筑能耗的重要特征。或采用时间序列分析法，分析一段时间内建筑能耗变化的趋势，再结合专业知识分析该趋势的合理性。

建筑能耗信息和建筑基础信息，包括的数据量大、信息多。如何从这些数量巨大、结构复杂的数据中抽丝剥茧，获取关键性特征信息，用于指导建筑节能相关工作，是需要解决的问题。

在建筑能耗预测中，数据的产生并不是完全任意的。那些看上去没有规律的建筑能耗数据的离散分布，在其背后有一个真正的数据生成过程。本章的任务是

---

---

通过介绍统计学基本知识和分类，结合建筑能耗数据形成的特点，选取最合适的方法，用于指导数据分析工作。

### 3.1 统计学的基本概念

#### 3.1.1 统计学的研究对象

统计学一般研究的对象包括社会政治、经济、文化、人民生活等领域中所体现出的各种现象。例如国民财富与资产、人口与劳动力资源、生产与消费、财政与金融等。通过在数据方面对这些基本的现象进行认识，可以获取对整个社会的基本信息的认识。在现代社会发展的过程中，如果不能准确、及时、全面、系统地掌握各种经济现象所体现出数量的特征和变化，便无法有效地进行正确的规划与决策，更谈不上有效的管理和充分的分析研究。从这个角度来说，社会发展的程度越高，对统计的要求就越重要。

通过各种统计数据来说明某项工作的发展状况，具体而又方便，证据充分，结论清晰。因此有必要通过挖掘各种数据信息（统计指标）来反映研究对象的发展情况及变动规律。这些数据信息既可以是静态的（横向的数据比较），也可以是动态的（随时间变化的纵向数据的比较）。

##### （1）横向的数据比较

横向的数据比较可以反映在同一时间内，某一现象样本整体的规模和结构分布情况。例如，据统计，大型公共建筑总面积不足民用建筑的 5%，却占民用建筑总用电量的 30%，其用电量达  $100\sim300\text{ kWh}/(\text{m}^2\cdot\text{a})$ ，为住宅建筑用电量的 10 倍以上。

例如，截止 2005 年底，某市共有星级饭店 229 家（其中五星级饭店 9 家，四星级饭店 25 家），商务楼宇 160 座，规划总建筑面积达 3247 万  $\text{m}^2$ 。

这些都显示了在同一时间段上相关数据结构的分布情况。

##### （2）纵向数据的比较

历史发展的纵向数据反映的是现象发展的变动规律。如 2008 年某办公建筑单位建筑面积全年电耗为  $80.5\text{ kWh}/\text{m}^2$ ，比 2007 年增长了 14.1%。

#### 3.1.2 统计学研究对象的特点

##### （1）数据性

统计学的方法是通过研究各种现象的数据，以获取信息。因此，统计活动的首要问题是数据问题。这也说明了统计的特点是用客观、具体、准确的数据来描述和认识客观现象的特征、性质和规律。没有数据也就没有统计这样的认识工具。统计就是用数据的各种组合，来反映客观事物的现状、依存关系和变化发展趋势。如数量上的多少，可以反映客观事物发展的规模和水平；数量的对比关系和相关关系，可以反映客观事物之间发展和变化的速度。

#### （2）整体性

统计学是对客观事物的样本整体数量情况进行研究，并不是关注于个别现象的数量特征。当部分样本点数据出现偏差和波动时，统计学观点并不认为该现象能代表整个样本的特点。但当相当数量的样本点出现偏差和波动时，统计学观点就认为这些由特别现象所组成的特点，能够对整体样本特征产生影响。因此，统计就要对整体各单位普遍存在的事实进行大量的观察和综合分析，得出反映现象整体的数量特征。如，要研究城镇建筑能耗的消耗水平，目的不在于了解个别建筑能耗的消耗情况，而是要反映各市、各区县建筑能耗消耗水平的数量特征。

#### （3）离散性

离散性是指整体各样本特征的变化存在着差异，而且这些差异并不是由于某种特定的原因造成的。例如，一个建筑中运行时间长短不同，建筑面积大小不同，所处城市气候条件不同，运行管理水平不同，导致即使是相同类型建筑，依然存在建筑能耗差异巨大的现象。只有通过统计分析才能对其表现出来的差异总结出规律性。

由于各种现象具有广泛而复杂的联系，各种先决条件各不相同，因而会有特殊性和偶然性，所以对现象整体的数量研究，必须运用统计的方法，从各样本数据中归纳概括出共同、普遍的特征。

#### （4）具体性

统计分析的数据是对实际问题的反映，是该领域这一问题的体现和表象。因此统计分析不能独立于实际问题对应的科学知识，必须与相关的专业知识一起，从各个角度对问题进行分析。这也说明，统计学虽然功能强大，但绝不是万能的，也不是孤立存在的。与专业技术工具之间是相辅相成的关系。

---

### 3.1.3 统计的功能

统计的功能按照从基础到深入的层次顺序，依次包括信息传递、咨询、职能监督三大职能。

#### （1）信息传递

这是统计最基本的功能，是指根据科学的统计指标体系和统计调查方法，系统地收集、整理、传递、储存和提供信息，展示该对象特征的功能。

#### （2）咨询

咨询是统计在传递信息时延续和深化，是指利用所掌握丰富的数据信息，运用科学的方法和技术手段，深入开展综合分析和各种专题研究，为相关活动的决策提供科学的咨询建议和决策方案。在当前社会经济建设中，通过各种统计服务方式，为社会各界和生产经营以及消费者提供有关的信息咨询。各级各部门提供统计资料，是各级部门和决策者了解情况、指导工作、制定政策方面的重要依据。

#### （3）职能监管

统计的职能监管功能是对信息传递和咨询功能的深化，指在相关部门的执行阶段，通过反馈的数据信息来评判、检验和调整方案，即根据统计调查和统计分析资料，准确地从整体上反映在一定时间、地点条件下的运行状况，并对其进行全面、系统的检查、监督和预警，以发现运行过程中的反映的状况及其原因，在发出预警信号的同时，提出相应的对策和措施。

## 3.2 统计学的基本分析方法

统计学工具在各个行业中应该如何选择，如何应用。是否存在万能的统计学工具，能解决所有行业中的问题？结论当然是：不存在。在不同的应用领域中，数据量、数据结构、需要解决的问题都不相同，不可能存在一种可以解决所有问题的统计学工具。由于统计学的这种与应用领域联系紧密的特点，派生出各种各样的统计学理论和工具，用于在不同的场合解决不同的问题。因此，目前最重要的是要结合建筑能耗基准模型的特点，选择合适的统计学工具。

统计学分支很多，数据分析工具很多。包括评价指标法、时间序列法、指数分析法等简单的数据分析工具，还包括单因素 ANOVA（方差分析）、卡方检验分析、Friedman 检验分析、多元 logistic 回归分析等高级数据分析工具。本文限

于篇幅，不可能对所有统计工具进行逐一介绍，仅选取几种典型常见的工具进行介绍，最后给出统计学工具的选择方法。

本文通过对大量的统计学工具的总结，将统计学工具分为两大类：单变量分析法和多变量分析法。以下将根据这两大分类，分别介绍几种典型的工具。

### 3.2.1 单变量分析法

顾名思义，是只针对单一变量进行分析的方法。例如，对某一城市办公建筑的能耗特点进行分析。由于只为了获取“办公建筑能耗”变量的特点，可采用单变量分析法进行分析。单变量分析法也是统计学里原理简单、使用广泛的工具之一。常见的单变量分析法包括平均指标法、时间序列法、指数分析法和抽样推断法。

#### 3.2.1.1 平均指标法

平均指标又称平均数，是统计中十分重要也是经常使用的综合指标。指在相同样本整体内将各单位某一数量标志表现的离散性抽象化，用以反映在特定时间、地点、条件下的样本整体一般水平的综合指标。其特征通常表现为：一般距离平均值远的样本比较少，而距离平均值近的或者接近平均值的样本比较多，所以平均指标反映了样本整体分布的集中趋势或一般水平。

平均指标是把各个单位标志表现之间的差异抽象化，从而说明样本整体的一般水平。如某个地区的办公建筑的平均能耗就是把各办公建筑之间的不同能耗的差异抽象化，用以说明该地区办公建筑能耗的一般水平。

计算平均指标的各个单位必须具有同类性质，这是计算平均指标的前提。因为只有本质相同的现象计算平均数才能正确反映客观实际情况，如果把不同性质的个体混杂在一起，则由此计算平均数只会掩盖事物的本质区别，得出错误结论。

从样本整体的分布情况看，不管用什么技术方法计算的平均数，结果都靠近中间部分，而不是在两头。这就说明多数样本值集中在平均数附近，所以，平均指标是标志集中趋势的测度数，是反映样本整体中倾向的代表值。

平均指标法包括算术平均数、调和平均数、几何平均数、加权平均数等，通过计算各种不同的平均数来反映不同的条件下和不同考虑因素的平均指标。与此同时为了更好地执行统计功能还需计算方差和标准差以更好地反映数据的真实意义。

---

---

由于平均指标法原理简单、使用方便、能够表示样本整体的性能，在建筑能耗分析中，一般用于计算某一范围内所有建筑能耗的平均值，以此来衡量该区域建筑能耗的总体水平。

对于统计问题而言，总体的均值不可能获得，因为无法获得所有样本的数据，因此只能采用有限样本对总体进行估计。于是就有“样本均值”这一概念，而且所有“均值”都应是“样本均值”。通过计算样本的算术平均值可以得到样本均值，而且随着样本数量的增大，样本均值可无限趋近于总体均值。

在建筑能耗的统计数据中，最常见的就是样本均值。例如某市办公建筑的平均能耗、某区酒店建筑的平均能耗等。很多人会把自己建筑的能耗与样本均值进行比较，然后得到结论：“达到了全市平均水平”。但仔细考虑一下，这样的均值能有多大概率上的意义？

与均值类似的另一个概念就是中位数，但在建筑能耗分析中使用较少。因为样本均值只是代表样本的算术平均值，而中位数能在一定程度上表示样本的分布特点，因此中位数有较强的统计意义。如果总体的分布是对称的，则样本均值和中位数数值上相等。但如果不是对称分布，那么仅提供建筑能耗的样本均值显然是很片面的，必须有其他指标来进行补充说明<sup>(49)</sup>。不幸的是，国内目前还在大量使用单一样本均值用于说明建筑能耗水平。在 1.3.4 节对德国建筑能耗基准的研究中，已经说明，德国目前采用中位数而非均值作为指标。

### 3.2.1.2 时间序列分析

时间序列分析，是指从时间的发展变化角度，研究各不同事物在不同时间的发展状况，探索其随时间推移的演变趋势和规律，解释其数量变化和时间关系，预测客观事物在未来时间上可能达到的数量和规模。

时间序列对于统计分析有着极其重要的作用，它是对现象进行动态研究与分析的依据，其作用主要表现在以下几方面：通过时间序列的编制和分析，可以从事物在不同时间上的量变过程中，认识事物的发展变化方向、程度、趋势和规律，为指定政策、编制计划提供可行的依据；通过对时间序列资料的研究，可以发现研究现象发展变化的规律和未来趋势，以便对其进行预测分析；通过时间序列可以进行不同地区的同类现象进行分析比较。

由此可见,时间序列分析的主要内容包括:根据时间序列计算各种分析指标,描述现象所处的状态。总结时间序列变动的规律和趋势,并以此对现象的未来进行预测分析。

由于加入是时间元素,时间序列分析法与平均指标法不同,主要用于分析建筑能耗随时间的变化特点和趋势,也能用于预测未来建筑能耗变化的情况。

### 3.2.1.3 指数分析法

指数是用于经济分析的一种特殊的统计分析方法,主要用于反映事物的相对变化程度。

随着社会经济的发展,指数的应用范围在不断扩大,相应地,指数的概念也在不断扩展,指数的概念有广义和狭义之分。

广义上的指数是指反映社会经济现象数量变动的全部相对数。如比较相对数、动态相对数(发展速度)、计划完成程度相对数等。

狭义上的指数是指综合反映数量上不能直接加总的复杂社会经济现象样本整体数量综合变动程度的相对数。由于复杂样本整体中各种事物的性质不同,使用价值及计量单位不同,即不能同度量,所以无法将它们的数量加以汇总,因而也就无法通过计算一般的相对数来反映其数量的综合变动程度,而应运用专门的特殊的方法才能解决此类问题。例如,对于工业产品的产量,如电视、汽车、钢材等,用速度指标即可分别测定其变动程度。但要把品种繁多的工业产品作为一个样本整体来考察,以反映全部产品的总变动,则是速度指标是不能直接测定的,这时就要通过编制相应的产量指数、销售指数、单位成本指数、价格指数等,以此来反映样本整体数量的综合变动。

统计指数反映的是事物发展变化的相对程度,它可以度量样本整体在不同空间或时间上的相对变化。对于狭义的指数而言,它反映的是复杂社会经济现象样本整体内部各个组成成分的综合变动,而不是其中某一个组成成分的变动。统计指数是样本整体水平的代表数值之一。

在建筑能耗分析中,指数分析法一般不使用单一指数指标。例如,在国家标准《空调调节系统经济运行》GB/T 17981-2007<sup>(50)</sup>中,单位面积空调能耗、单位面积耗冷量、空调系统能效比、冷冻水输送系数等一系列指数指标,用于衡量各

---

---

分项系统的用能情况。指数分析法在绿色建筑评价体系中也应用广泛。

### 3.2.2 多变量分析法

变量统计分析 (Multivariate Statistical Analysis) 简称多变量分析, 为统计学的一支, 常用于管理科学、社会科学和生命科学等领域中。多变量分析主要用于分析拥有多个变量的样本, 探讨样本彼此之间的关联性或结构, 而有别于传统统计方法所着重的参数估计以及假设检定。由于多变量分析方法需要复杂且大量的计算, 因此多借助电脑来进行运算, 常用的统计套装软件有 SAS、SPSS、Statistica 等<sup>(51)</sup>。由于计算技术和硬件的迅速发展, 以及多变量分析的验证性日趋受到重视, 因此得到了迅速、广泛的应用。学者 Gatty 指出: “在任何应用科学领域的研究中, 研究问题如果没有用多变量分析处理, 则结果会是相当肤浅的。”

#### (1) 多变量方差分析<sup>(52)</sup>

MANOVA 是 Multivariate Analysis of Variance 的简称, 适用于同时探讨一个或多个自变量与多个 ( $\geq 2$  个) 以上因变量间因果关系的统计方法, 依照研究者所操作自变量的个数, 可以分为单因素 (一个自变量) 或多因素 (两个以上自变量) MANOVA。

#### (2) 主成分分析

主成分分析的主要功能在分析多个变量间的相关, 以建构变量间的总体性指标 (overall indicators)。这些变量没有自变量和因变量的区别。当研究者测量一群彼此间具有高度相关的变量时, 在进行显著性检验前, 为避免变量数过多, 造成解释上的复杂与困扰, 常会先进行主成分分析, 在尽量不丧失原有信息的前提下, 抽取少数几个主成分, 作为代表原来变量的总体性指标, 达到资料缩减 (data reduction) 的功能, 也可解释为具有“降维”的功能。

#### (3) 因子分析

因子分析与主成分分析常被研究者混用, 因为二者功能都是通过对变量间的相关分析, 以达到简化数据功能。但不同的是, 主成分分析是在找出变量间最佳线性组合 (linear combination) 的主成分, 以说明变量间最多的变异量; 而因子分析, 则在于找出变量间共同的潜在结构 (latent structure) 或因子, 以估计每一个变量在各因子上的负荷量 (loading)。在因子分析中, 也不存在自变量和因

变量之分。

#### (4) 典型相关分析

典型相关分析的主要功能在分析两个变量间的相关性。进行多元回归分析的目的，是在分析一个或多个自变量与一个因变量间的关系，而典型相关中因变量也可以是多个；也就是说，典型相关是多元回归分析的功能扩展。

#### (5) 聚类分析

聚类分析的主要功能为对观测值的相似性或差异性进行分类(classification)，以形成几个性质不同的类别，简化解释的工作，达到组内同质、组间异值的目的。进行聚类分析时，并无自变量和因变量的区分。

#### (6) 判别分析

判别分析是多变量分析中应用相当广泛的统计方法，它可以用来对样本进行分类的工作；也可以用来了解不同类别样本在某些变量上的差异情形；同时也可根据不同类别的样本在某些变量的实际表现，用来预测新的样本属于某一类别的概率。因此，在行为科学中，常见的研究者单独使用判别分析，建立判别函数(discriminant function)，以对新样本进行预测；或是多变量方差分析的检验值达到显著性水平后，比较不同组别样本在因变量平均数的差异情形；或是聚类分析后，检验聚类分析的正确性。进行判别分析时，自变量是定距以上层次变量，至于因变量通常是离散变量。

#### (7) 多维量表分析

多维量表分析基本上也是一种分类的统计方法，他在市场上普遍被应用。当研究者想要解释一群受试者（例如消费者）对一组客体（例如商品）在某些变量上相似性的测量中所包含的信息，此时多维量表分析就是一个相当适用的方法。研究者只要将这一组客体在变量上的测量值转化成多维度的几何表征，就能够将这些客体有效地显示在这个几何空间中，达到分类的目的，同时也可以进一步解释这些几何表征所代表的潜在结构或意义。进行多维量表分析时，并无自变量和因变量的区分，同时变量可以是等距以上变量，也可以是定类或定序变量。

#### (8) 线性结构方程

线性结构方程是一个相当具有变通与弹性的统计方法，随着研究者对变量间

---

---

关系界定的差异, LISREL 的常见名称包括协方差结构分析, 潜变量分析、线性结构模型或验证性因子分析。LISREL 可视为多元回归分析与因子分析两个方法论的整合模型, 让研究者可以探讨变量间的线性关系(回归分析), 并对可测量显变量与不可测量的潜变量(因子分析)的因果模型作假设检验。

#### (9) 逻辑斯蒂回归分析

逻辑斯蒂回归可视为传统多元回归分析的一个特列。它和多元回归分析一样, 都具有解释自变量与因变量之间的关系, 并可进行预测。所不同的是在进行多元回归分析时, 包括自变量与因变量都必须是定距以上层次变量; 但在进行逻辑斯蒂回归分析时, 自变量仍是定距以上层次变量, 因变量则是二分的定类变量或三分定类变量或定序变量。

#### (10) 对数线性方程

在基本统计学中, 当研究者面对探讨两个定类或定序变量间关系的研究问题时, 都是以卡方检验来进行假设检验。当问题的性质是探讨两个定类变量间是否独立或是关联强度时, 是以卡方独立性检验来进行假设检验。进行卡方独立性检验时, 研究者必须将样本在两个定类变量上的反应, 建立二维列联表(contingency table), 以进一步根据列联表中各单元格(cell)的次数反应, 进行显著性检验。但当研究者面对三个或三个以上的定类变量时, 所建立的多元列联表间变量关联的分析, 卡方独立性检验将无法解决这样的问题, 此时适合的方法就是对数线性模型。利用对数线性模型来解决多元列联表的问题的目的, 主要就在于探讨构成列联表的多个定类变量间的关系, 进而在精简原则下构建拟合的解释模型, 并根据所建立的模型估计单元格参数值, 以了解各变量效果对单元格次数的影响。

#### (11) Logit 对数线性模型

在对数线性模型中, 多个定类变量间是互为因果的关系(即相关关系), 并无自变量与因变量的区分, 研究目的在于探讨变量间的关联强度和性质。但有时研究者会面临变量间有自变量和因变量的区分的情境。在基本统计学中, 当研究者面对的问题性质是两个定类变量间有自变量和因变量的区别, 目的在于探讨两个变量间的因果关系时, 多是以卡方齐性检验来进行假设检验。但自变量个数在两个以上时, 卡方齐性检验就不再适用, 而必须改用 logit 对数线性模型方法来

对数据进行分析。Logit 对数线性模型的功能与多元回归分析相当类似，都可以用来探讨与解释因变量与自变量间的关系，但不同的是，多元回归分析的变量都是定距以上层次变量，通常以最小二乘法进行模型估计与检验；logit 对数线性模型的变量都是定类变量，通常以最大似然估计法进行模型估计与检验。

### 3.2.3 统计学工具的选择方法

表 2<sup>(53)</sup>给出了根据自变量、因变量的数量和特点，选择统计学工具通用的方法。

表2 统计学工具的选择

因变量 个数	自变量个数	因变量 性质	应采用的方法
1 个	0 自变量， 1 个总体	连续且 正态	单样本 t 检验分析
		有序或 连续	单样本中位数检验分析
		二分类	二项式检验分析
		分类	卡方优度拟合 t 分析
1 个	1 自变量， 2 水平， 组间独立	连续且 正态	2 独立样本 t 检验 分析
		有序非	Wilcoxon-Mann Whitney 检验分析
		分类	卡方检验分析
			Fisher 精确检验分析
1 个	1 自变量, 2 个或 2 个以上水平， 组间独立	连续且 正态	单因素 ANOVA (方差分 析)
		有序或 连续	Kruskal Wallis 分析
		分类	卡方检验分析

因变量个数	自变量个数	因变量性质	应采用的方法
1 自变量， 2 水平， 组间相关/配对或配伍	1 自变量， 2 水平， 组间相关/配对或配伍	连续且正态	配对 t 检验分析
		有序或连续	Wilcoxon 符号秩检验分析
		分类	McNemar 分析
	1 自变量, 2 个或 2 个以上水平, 组间相关/配对或配伍	连续且正态	单因素重复测量 ANOVA (方差分析)
		有序或连续	Friedman 检验分析
		分类	重复测量 logistic 回归分析
	2 个或以上自变量, 组间独立	连续且正态	析因 ANOVA (方差分析)
		有序或连续	普通 logistic 回归分析
		分类	因子 logistic 回归分析
1 个连续性自变量	1 个连续性自变量	连续且正态	相关性分析
		有序或连续	简单线性回归分析
	1 或多个连续性自变量和/或 1 个或多个自变量组	连续且正态	非参数相关分析
		分类	简单 logistic 回归分析
	1 或多个连续性自变量和/或 1 个或多个自变量组	连续且正态	多元拟合回归分析
		有序或连续	协方差分析
		分类	多元 logistic 回归分析
			判别分析

因变量个数	自变量个数	因变量性质	应采用的方法
2个或以上	1个自变量, 2个或2个以上水平, 组间独立	连续且正态	one-way MANOVA (多变量方差分析)
2个或以上	2个或以上自变量	连续且正态	多变量多重线性回归分析
2组或2组以上	0	连续且正态	典型相关分析
2组以上	0	连续且正态	因子分析

表2给出了全面的方法,但不太便于初学者选择。因此用图16再给出一个简单的选择方式。

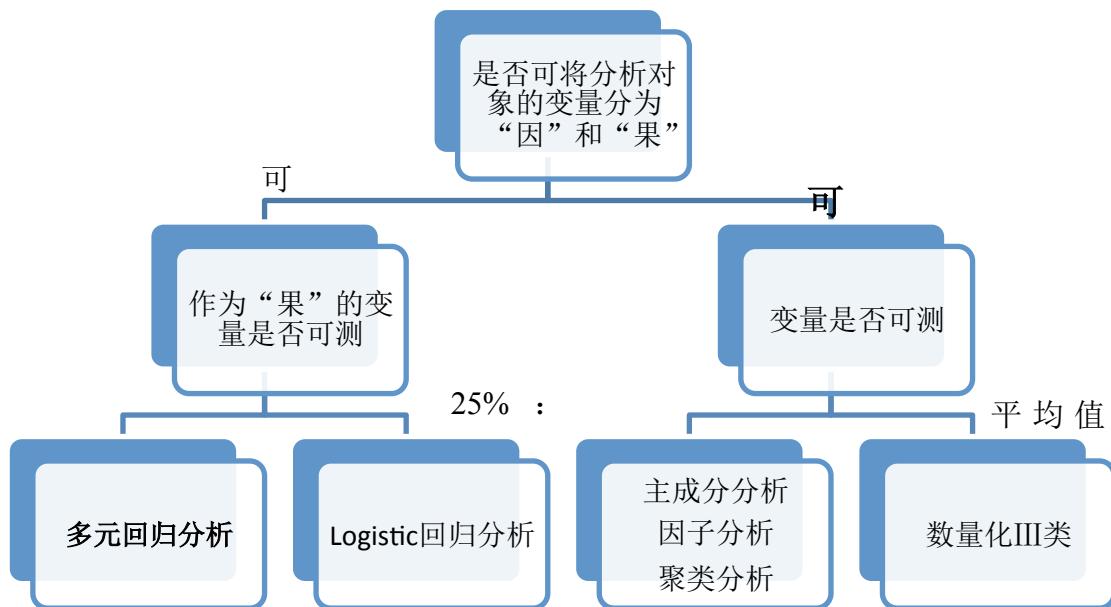


图16 多变量分析工具的选择方法<sup>(54)</sup>

### 3.3 统计方法的对比和选择

#### 3.3.1 建筑基准模型的基本功能

根据文献综述所归纳国外的研究成果,初步确定建筑基准模型应能够以大量建筑能耗数据和建筑运行参数为基础,通过数学方法,将对建筑能耗产生影响的主要变量做归一化处理,尽量消除外界因素对建筑能耗的影响,保证各建筑能在

---

---

公平的基准条件下进行能源利用效率的比较。

确定建筑基准模型的基本功能后，可以明确研究对象的特征。建筑能耗为单一的研究对象，存在 2 个或 2 个以上可能对建筑能耗产生影响的因素，定义为自变量。

### 3.3.2 在单变量分析法和多变量分析法分类之间选择

统计方法主要分为单变量分析和多变量分析两大类。各分析方法都有自己的特点和使用范围，通过对比这些方法最终找到能够充分满足需要的最适用的方法是这节的重点。

通过上述介绍可以发现：单变量分析法由于原理简单、使用方便，因此应用广泛。但单变量分析法也有缺点。平均指标反映了样本整体分布的集中趋势或一般水平。只能进行同类现象计算，反映样本整体分布的趋势；时间序列分析法从时间的发展变化角度，研究各不同事物在不同时间的发展状况，探索其随时间推移的演变趋势和规律，解释其数量变化和时间关系，预测客观事物在未来时间上可能达到的数量和规模。是对样本整体的未来发展的推测；指数分析法主要用于反映事物某一方面的相对变化程度，基本上是采用一些统计公式或方法，比如工业生产指数、居民消费价格指数、股票价格指数等。

采用单一变量衡量建筑能耗生成的复杂过程，存在信息量缺乏的问题。平均指标法、时间序列法和指标分析法，只能对建筑能耗的其中一种特点进行描述，显然是不充分的。多变量分析法更加适合建筑能耗基准模型的建立和分析工作。

表 3 中给出了各单变量分析法优缺点比较的总结。

表3 各单变量分析法优缺点的比较

名称	优点	缺点	适用范围
平均指标法	简单，明了，反应样本整体特征	指标单一，无法反应样本之间的离散情况	简单分析和比较
时间序列法	体现样本随时间变化的动态特性	无法体现整体特性	动态分析
指数分析法	能够有针对性体现	指标单一，仅能体现	详细表现某一现象

名称	优点	缺点	适用范围
	某一特征	单一特征	

显然,单变量分析法的分析结果信息量缺乏,未考虑外界因素对建筑能耗的影响。综合考虑,多变量分析法更加适合。

### 3.3.3 在多变量分析法分类中选择合适的分析方法

按照表 2 的方法进行选择。

- a、具有 1 个因变量(建筑能耗);
- b、1 个或更多自变量(众多影响因子);
- c、变量之间不独立,自变量是连续的,呈状态分布的;
- d、选择“多元回归分析”。

按照图 16 的方法,也能得到相同的选择。

- a、可以将描述建筑能耗特征的所有变量分为“因”和“果”,对建筑能耗产生影响的因子是“因”,建筑能耗是“果”;
- b、作为“果”的建筑能耗是可以测量的;
- c、选择“多元回归分析”。

## 3.4 回归分析法的原理

回归分析原理:当某几个变量之间存在相互影响、相互制约关系时,即其中一个量的变化会影响另一个变量也发生某种程度的变化,但是这种变化不是精确的,只能粗略地表示出另一个变量的估计值,变量间的这种关系称为相关关系。当变量间存在相关关系时称为两变量相关,对于存在相关关系的两变量  $x$  和  $y$ ,用函数  $y=f(x)$  最优拟合,对于每一个确定变量  $x$  的值都可以通过函数  $f(x)$  获得一个与之相对应的  $y$  的估计值。这种方法称为回归分析,  $f(x)$  称为回归方程。

回归分析法主要分为两类:线性回归、非线性回归。线性回归又分为一元线性回归和多元线性回归。非线性回归有多项式回归以及其他种类非线性回归。

### 3.4.1 线性回归

#### 3.4.1.1 一元线性回归的模型

一元线性回归模型又称为简单直线回归方程,是根据两个变量的成对数据,

---



---

配合直线方程式，再根据自变量的变动值，来推算因变量的估计值的一种统计分析方法。

对于所要考察的变量  $y$  来说，其主要的影响因素只有自变量  $x$  一个，且  $y$  与  $x$  呈线性相关关系，则在变量  $y$  与  $x$  之间建立数学模型为：

$$Y = \beta_0 + \beta_1 x$$

式中， $Y$  为理论值； $\beta_0$  表示在  $Y$  轴上的截距，代表直线的起点值； $\beta_1$  表示直线的斜率，又称  $Y$  对  $x$  的回归系数，它的实际意义是说明  $x$  变化每一个单位时， $y$  就平均增加  $\beta_1$  个单位；若  $x$  前面是负号，说明  $x$  每增加一个单位时， $Y$  就平均减少  $\beta_1$  个单位。因此，回归系数  $\beta_1$  前面的正负号还分别表示变量  $Y$  与  $x$  的正相关还是负相关。 $\beta_0$  和  $\beta_1$  都叫做待定系数，它要根据  $Y$  和  $x$  的实际资料求解，一旦解出  $\beta_0$  和  $\beta_1$  的值，则表明变量之间的一般关系的回归直线就随之确定。

对于一元线性回归方程，实际观察值  $y_i$  与估计值  $\hat{y}_i$  之差  $y_i - \hat{y}_i = y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i$  映了观察值  $y_i$  与回归直线  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$  的偏离程度。令  $Q(\beta, \beta) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$ ， $Q$  表示所有观察值  $y_i$  与回归直线  $\hat{y}_i$  的偏离平方和，刻画了所有观察值与回归值的偏离程度。

取  $Q$  达到极小值，求出它关于  $\hat{\beta}_0$ ， $\hat{\beta}_1$  的偏导数并令其为零，即可求出  $\hat{\beta}_0$ ，

$$\hat{\beta}_1 \text{ 的最小二乘估计为 } \begin{cases} \hat{\beta}_0 = \bar{y} - \bar{x} \hat{\beta}_1 \\ \hat{\beta}_1 = L_{xy} / L_{xx} \end{cases} \text{，其中：} \begin{cases} L_{xx} = \sum_{i=1}^{dif} (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2 \\ L_{xy} = \sum_{i=1}^{dif} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \end{cases}$$

$$\begin{cases} \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \\ \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \end{cases} \text{ 由此得到关于样本的一元线性回归方程 } \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

### 3.4.1.2 多元线性回归模型

如果变量  $x_1, x_2, \dots, x_p$  与随机变量  $y$  之间存在着相关关系, 通常就意味着每当  $x_1, x_2, \dots, x_p$  取定值后,  $y$  便有相应的概率分布与之对应。随机变量  $y$  与相关变量  $x_1, x_2, \dots, x_p$  之间的概率模型为

$$y = C_0 + C_1 x_1 + C_2 x_2 + \dots + C_n x_n$$

其中, 随机变量  $y$  称为因变量,  $x_1, x_2, \dots, x_p$  称为自变量。 $f(x_1, x_2, \dots, x_p)$  为一般变量  $x_1, x_2, \dots, x_p$  的确定性关系,  $\varepsilon$  为随机误差。正是因为随机误差项  $\varepsilon$  的引入, 才将变量之间的关系描述为一个随机方程, 使得我们可以借助随机数学方法研究  $y$  与  $x_1, x_2, \dots, x_p$  的关系。由于客观现象是错综复杂的, 一种现象很难用有限个因素来准确说明, 随机误差项可以概括表示由于人们的认识以及其他客观原因的局限而没有考虑的种种偶然因素。随机误差项主要包括下列因素的影响:

- a、由于人们认识的局限或时间、费用、数据质量等制约未引入回归模型但又对回归因变量  $y$  有影响的因素;
- b、样本数据的采集过程中变量观测值的观测误差的影响;
- c、理论模型设定误差的影响;
- d、其他随机因素的影响。

上式清楚地表达了变量  $x_1, x_2, \dots, x_p$  与随机变量  $y$  的相关关系, 它由两部分组成: 一部分是确定性函数关系, 由回归函数  $f(x_1, x_2, \dots, x_p)$  给出; 另一部分是随机误差项  $\varepsilon$ 。由此可见上式准确地表达了相关关系那种既有联系又不确定的特点。

对于多元线性回归方程, 设影响因变量  $Y$  的自变量个数为  $p$ , 并且分别记为  $x_1, x_2, \dots, x_p$ , 所谓多元线性模型是指这些自变量对  $Y$  的影响是线性的, 即有

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$
 其中  $\varepsilon \sim N(0, \sigma^2)$ ,  $\beta_0, \beta_1, \beta_2, \dots, \beta_p, \sigma^2$  是与  $x_1, x_2, \dots, x_p$  无关的未知参数, 称上式为因变量  $Y$  对自变量  $x_1, x_2, \dots, x_p$  的多元线性回归方程。

记  $n$  组样本分别为  $(x_{i1}, x_{i2}, \dots, x_{ip}), (i = 1, 2, \dots, n)$ , 令

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$

此公式即为关于样本的多元线性回归方程的求解公式。

### 3.4.1.3 多元线性回归模型的检验

假设  $H_0: \beta_0 = \beta_1 = \cdots = \beta_k = 0$

#### a、F 检验法

当  $H_0$  成立时,  $F = \frac{U/k}{Q_e/(n-k-1)} \sim F(k, n-k-1)$

如果  $F > F_{1-\alpha}(k, n-k-1)$ , 则拒绝  $H_0$ , 认为  $y$  与  $x_1, \dots, x_k$  之间显著地有线性关系; 否则就接受  $H_0$ , 认为  $y$  与  $x_1, \dots, x_k$  之间线性关系不显著.

F 检验是否通过是根据查表得来的, 统计学里有一张 F 检验表, 通过拟合出来的值同表上的值相比较得出是否通过 F 检验。具体结果为: 如果拟合值大于表中所查的值则通过 F 检验, 否则不通过。

#### b、r 检验法

定义  $R = \sqrt{\frac{U}{L_{yy}}} = \sqrt{\frac{U}{U+Q_e}}$  为  $y$  与  $x_1, x_2, \dots, x_k$  的多元相关系数或复相关系数。

由于  $F = \frac{n-k-1}{k} \frac{R^2}{1-R^2}$ , 故用 F 和用 R 检验是等效的。

#### c、t 检验法

t 检验法是在给定的显著水平  $\alpha$  下, 确定临界值  $t_{\alpha/2}(n-k-1)$ , 或者计算 t 值所对应的相伴概率值 P 的大小。

如果  $|t| > t_{\alpha/2}(n-k-1)$ , 或  $P < \alpha$ , 则原假设  $H_0$  不成立, 认为该回归系数与因变量之间存在显著的线性关系, 它的变化能够反映因变量的变化, 应保留在回归预测模型中。

如果  $|t| < t_{\alpha/2}(n-k-1)$ , 或  $P > \alpha$  则原假设  $H_0$  成立, 认为该回归系数与零差异不显著, 该自变量与因变量间无显著的线性关系, 它的变化不能够反省因变量的线性变化, 应从回归模型中剔除。

因此, 当  $P$  越小时则该自变量越接近零差异, 说明该自变量越能反映因变量的变化。

### 3.4.2 非线性回归

在一元回归分析中, 如果因变量与自变量的关系为非线性的, 但是又找不到适当的函数曲线来拟合, 则可以采用一元多项式回归。多项式回归的最大优点就是可以通过增加  $x$  的高次项对实测点进行逼近, 直至满意为止。多项式回归可以处理相当一类非线性问题, 它在回归分析中占有重要的地位, 因为任意函数都可以分段用多项式来逼近。

多项式回归有以下 4 种情况:

$$\text{linear (线性): } y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$$

$$\text{pure quadratic (纯二次): } y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{j=1}^n \beta_{jj} x_j^2$$

$$\text{interaction (交叉): } y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{1 \leq j \neq k \leq m} \beta_{jk} x_j x_k$$

$$\text{quadratic (完全二次): } y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{1 \leq j, k \leq m} \beta_{jk} x_j x_k$$

由以上式子可以看出, 多项式拟合能够进行分段拟合, 包含的范围比较广, 但当自变量的个数大于 3 个之后, 其稳定性较差, 因此多项式回归的方法经常用于一元的数据拟合。

由于多元非线性回归的应用少, 这里只介绍一元非线性回归。一元非线性拟合首先对两个变量  $x$  和  $y$  作  $n$  次试验观察得  $(x_i, y_i), i = 1, 2, \dots, n$  画出散点图, 根据散点图确定须配曲线的类型.然后由  $n$  对试验数据确定每一类曲线的未知参数  $a$  和  $b$ .采用的方法是通过变量代换把非线性回归化成线性回归, 即采用非线性回归线性化的方法。

通常选择的六类曲线如下:

---

---

a、双曲线  $\frac{1}{y} = a + \frac{b}{x}$

b、幂函数曲线  $y = ae^b$ , 其中  $x > 0$ ,  $a > 0$

c、指数曲线  $y = ae^{bx}$  其中参数  $a > 0$ .

d、倒指数曲线  $y = ae^{b/x}$  其中  $a > 0$ ,

e、对数曲线  $y = a + b \ln x$ ,  $x > 0$

f、S型曲线  $y = \frac{1}{a + be^{-x}}$

非线性拟合同多项式拟合一样, 其主要应用的领域为一元数据拟合, 当自变量的个数增多时其方程的稳定性较差。

### 3.4.3 回归方法的计算流程

通过对以上的回归方法的介绍以及目前所掌握的数据可以发现, 多元线性回归方法的回归方程稳定性好, 数据意义清晰, 便于理解, 与建筑能耗基准模型的特点类似, 因此是建立建筑能耗基准模型适合的方法。

#### 3.4.3.1 逐步回归分析的主要思路

在实际问题中, 人们总是希望从对因变量  $y$  有影响的诸多变量中选择一些变量作为自变量, 应用多元回归分析的方法建立“最优”回归方程以便对因变量进行预报或控制。所谓“最优”回归方程, 主要是指希望在回归方程中包含所有对因变量  $y$  影响显著的自变量, 而不包含对  $y$  影响不显著的自变量的回归方程。逐步回归分析正是根据这种原则提出来的一种回归分析方法。它的主要思路是在考虑的全部自变量中按其对  $y$  的作用大小, 显著程度大小或者说贡献大小, 由大到小地逐个引入回归方程, 而对那些对  $y$  作用不显著的变量可能始终不被引入回归方程。另外, 已被引入回归方程的变量在引入新变量后也可能失去重要性, 而需要从回归方程中剔除出去。引入一个变量或者从回归方程中剔除一个变量都称为逐步回归的一步, 每一步都要进行  $F$  检验, 以保证在引入新变量前回归方程中只含有对  $y$  影响显著的变量, 而不显著的变量已被剔除。

逐步回归分析的实施过程是, 每一步都要对已引入回归方程的变量计算其偏回归平方和, 然后选一个偏回归平方和最小的变量, 在预先给定的  $F$  水平下进

行显著性检验, 如果显著则该变量不必从回归方程中剔除, 这时方程中其它的几个变量也都不需要剔除(因为其它的几个变量的偏回归平方和都大于最小的一个更不需要剔除)。相反, 如果不显著, 则该变量要剔除, 然后按偏回归平方和由小到大地依次对方程中其它变量进行 F 检验。将对  $y$  影响不显著的变量全部剔除, 保留的都是显著的。接着再对未引入回归方程中的变量分别计算其偏回归平方和, 并选其中偏回归平方和最大的一个变量, 同样在给定 F 水平下作显著性检验, 如果显著则将该变量引入回归方程, 这一过程一直继续下去, 直到在回归方程中的变量都不能剔除而又无新变量可以引入时为止, 这时逐步回归过程结束。

### 3.4.3.2 逐步回归分析的主要计算步骤

#### a、确定 F 检验值

在进行逐步回归计算前要确定检验每个变量是否显著的 F 检验水平, 以作为引入或剔除变量的标准。F 检验水平要根据具体问题的实际情况来定。一般地, 为使最终的回归方程中包含较多的变量, F 水平不宜取得过高, 即显著水平  $\alpha$  不宜太小。F 水平还与自由度有关, 因为在逐步回归过程中, 回归方程中所含的变量的个数不断在变化, 因此方差分析中的剩余自由度也总在变化, 为方便起见常按  $n-k-1$  计算自由度。n 为原始数据观测组数, k 为估计可能选入回归方程的变量个数。例如  $n=15$ , 估计可能有  $2 \sim 3$  个变量选入回归方程, 因此取自由度为  $n=11$ , 查 F 分布表, 当  $\alpha=0.1$ , 自由度  $f_1=1$ ,  $f_2=11$  时, 临界值  $F_{\alpha}=3.23$ , 并且在引入变量时, 自由度取  $f_1=1$ ,  $f_2=n-k-2$ , F 检验的临界值记  $F_1$ , 在剔除变量时自由度取  $f_1=1$ ,  $f_2=n-k-2$ , F 检验的临界值记  $F_2$ , 并要求  $F_1 \geq F_2$ , 实际应用中常取  $F_1=F_2$ 。

#### b、逐步计算

如果已计算  $t$  步(包含  $t=0$ ), 且回归方程中已引入 1 个变量, 则第  $t+1$  步的计算为:

(a)计算全部自变量的偏回归平方和  $V'$ 。

(b)在已引入的自变量中, 检查是否有需要剔除的不显著变量。这就要在已引入的变量中选取具有最小  $V'$  值的一个并计算其 F 值, 如果  $F < F_2$ , 表示该变量不显著, 应将其从回归方程中剔除, 计算转至(c)。如  $F > F_2$  则不需要剔除变量, 这时则考虑从未引入的变量中选出具有最大  $V'$  值的一个并计算 F 值, 如果  $F > F_1$ ,

---

---

则表示该变量显著，应将其引入回归方程，计算转至(c)。如果  $F \leq F_1$ ，表示已无变量可选入方程，则逐步计算阶段结束，计算转入(c)。

(c)剔除或引入一个变量后，相关系数矩阵进行消去变换，第  $t+1$  步计算结束。其后重复(a)~(c)再进行下步计算。由上所述，逐步计算的每一步总是先考虑剔除变量，仅当无剔除时才考虑引入变量。实际计算时，开头几步可能都是引入变量，其后的某几步也可能相继地剔除几个变量。当方程中已无变量可剔除，且又无变量可引入方程时，第二阶段逐步计算即告结束，这时转入第三阶段。

### c、其他计算

主要是计算回归方程入选变量的系数、复相关系数及残差等统计量。

逐步回归选取变量是逐渐增加的。选取第  $L$  个变量时仅要求与前面已选的  $L-1$  个变量配合起来有最小的残差平方和，因此最终选出的  $L$  个重要变量有时可能不是使残差平方和最小的  $L$  个，但大量实际问题计算结果表明，这  $L$  个变量常常就是所有  $L$  个变量的组合中具有最小残差平方和的那一个组合，特别当  $L$  不太大时更是如此，这表明逐步回归是比较有效的方法。引入回归方程的变量的个数  $L$  与各变量贡献的显著性检验中所规定的  $F$  检验的临界值  $F_1$  与  $F_2$  的取值大小有关。如果希望多选一些变量进入回归方程，则应适当增大检验水平  $\alpha$  值，即减小  $F_1 = F_2$  的值，特别地，当  $F_1 = F_2 = 0$  时，则全部变量都将被选入，这时逐步回归就变为一般的多元线性回归。相反，如果  $\alpha$  取得比较小，即  $F_2$  与  $F_1$  取得比较大时，则入选的变量个数就要减少。此外，还要注意，在实际问题中，当观测数据样本容量  $n$  较小时，入选变量个数  $L$  不宜选得过大，否则被确定的系数  $b_i$  的精度将较差。

## 3.5 本章小结

本章首先介绍了统计学的作用、原理。根据变量的多少将统计学分为两大类：单变量分析和多变量分析。通过对常用和典型方法的介绍和比较，提出如何选择统计学分析方法。根据本文拟建立的能耗基准模型的基本功能，确定选择多元回归分析法。

对回归分析法的原理和基本计算流程进行介绍，为下一步建立基准模型的工作提供基础知识。

## 第四章 酒店建筑能耗基准模型的建立

多元线性拟合回归方法是一种数学统计工具, 可以用于从大量看上去没有规律的数据中挖掘样本数据潜在的特征信息。作为一种功能强大的数学工具, 多元线性拟合回归不能脱离专业领域的知识独立解决问题, 两者必须有机结合, 互相补充, 才能有效获得真实的建筑能耗数据特征信息。

本章将详细说明对酒店类建筑能耗数据进行多元线性拟合回归的分析过程、发现的问题及解决的办法, 最终建立相应的建筑能耗基准模型, 也给出拟合回归具体的方法及注意事项。

### 4.1 数据来源及基本情况

#### 4.1.1 数据基本情况

数据包括全国范围内各星级酒店共 470 栋, 其中五星级酒店 230 栋, 四星级酒店 180 栋, 三星级酒店 60 栋。数据不仅包括建筑全年各类能源形式的总消耗量, 也包括了一些与运营有关的变量。数据结构见表 4。

表4 建筑能耗和基本信息数据结构

变量名称	变量类型	变量单位
联系人信息	/	/
建筑名称	/	/
地址	/	/
城市	/	/
邮编	/	/
建成年代	数字	年
总建筑面积	数字	$m^2$
星级	数字	/
平均年入住率	数字	%
房间总数	数字	间

变量名称	变量类型	变量单位
上班期间工作人员数	数字	人
冰箱及制冷设备数	数字	台
是否提供餐饮服务	布尔值	是/否
空调区域比例	数字	%
供热区域比例	数字	%
是否提供洗衣服务	布尔值	是/否
游泳池类型	数字	/
游泳池面积	数字	$m^2$
是否有 SPA、零售店、酒吧等	布尔值	是/否
各类能源全年总消耗量（2009 年）	数字	$kWh/m^3/t$
HDD	数字	/
CDD	数字	/

#### 4.1.2 HDD 和 CDD 的基础数据来源和计算

从原始数据中发现，由于对 HDD（供热度日数）和 CDD（制冷度日数）理解的不同，所提交上来的关于 HDD 和 CDD 的数据存在较大偏差。包括有的建筑提交的 HDD 和 CDD 相等；同一城市内不同建筑所提交的 HDD 和 CDD 值偏差巨大。因此，为了规范 HDD 和 CDD，方便进行统一的计算，应对 HDD 和 CDD 进行统一的定义和计算。定义以下两个变量：

HDD18：室外基准温度为 18℃ 时的采暖度日数；

CDD26：室外基准温度为 26℃ 时的制冷度日数。

由于建筑能耗数据全部为 2009 年全年能耗值，因此应根据 2009 年全年气象参数情况计算实际 HDD 和 CDD 值，不应使用历年气象参数的平均值。由于国家相关机构对气象参数的保密制度，因此使用网站 [www.wunderground.com](http://www.wunderground.com) 提供的 2009 年气象参数，对全国各相关城市的 HDD18 和 CDD26 进行统一计算。该网站为民间组织，由世界各地拥有气象设备的单位和个人自愿参与并分享气象数

据。

详细的结算结果见表 5。

表5 2009 年全国相关城市 CDD26 和 HDD18 数值

城市	CDD26	HDD18	城市	CDD26	HDD18
北京	189	3292	兰州	39	4393
长春	59	5447	南京	295	2107
长沙	360	1570	青岛	132	2854
成都	154	1438	泉州	460	561
重庆	294	1434	三亚	572	38
大连	57	3460	上海	346	1780
东莞	541	546	汕头	374	484
福州	295	994	沈阳	62	4488
广州	541	546	深圳	492	384
桂林	315	1192	石家庄	196	3078
海口	568	198	苏州	346	1780
杭州	344	1742	太原	142	3489
哈尔滨	71	6042	天津	172	3089
合肥	306	1982	武汉	339	1708
呼和浩特	95	4836	厦门	460	561
昆明	41	1139	西安	197	2620
昆山	346	1780	郑州	168	2386
珠海	492	384	/	/	/

## 4.2 回归分析使用软件的介绍

---

---

SPSS，全称为 Statistical Product and Service Solutions，统计产品与服务解决方案软件。SPSS 是世界上最早的统计分析软件，由美国斯坦福大学的三位研究生于 20 世纪 60 年代末研制，迄今 SPSS 软件已有 40 余年的成长历史。全球约有 25 万家产品用户，它们分布于通讯、医疗、银行、证券、保险、制造、商业、市场研究、科研教育等多个领域和行业，是世界上应用最广泛的专业统计软件<sup>(55)</sup>。

SPSS 是世界上最早采用图形菜单驱动界面的统计软件，它最突出的特点就是操作界面极为友好，输出结果美观漂亮。它将几乎所有的功能都以统一、规范的界面展现出来，使用 Windows 的窗口方式展示各种管理和分析数据方法的功能，对话框展示出各种功能选择项。用户只要掌握一定的 Windows 操作技能，粗通统计分析原理，就可以使用该软件为特定的科研工作服务。SPSS 采用类似 EXCEL 表格的方式输入与管理数据，数据接口较为通用，能方便的从其他数据库中读入数据。其统计过程包括了常用的、较为成熟的统计过程，完全可以满足非统计专业人士的工作需要。输出结果十分美观，存储时则是专用的 SPO 格式，可以转存为 HTML 格式和文本格式。对于熟悉老版本编程运行方式的用户，SPSS 还特别设计了语法生成窗口，用户只需在菜单中选好各个选项，然后按“粘贴”按钮就可以自动生成标准的 SPSS 程序。极大的方便了中、高级用户。

SPSS for Windows 的分析结果清晰、直观、易学易用，而且可以直接读取 EXCEL 及 DBF 数据文件，现已推广到多种各种操作系统的计算机上，它和 SAS、BMDP 并称为国际上最有影响的三大统计软件。在国际学术界有条不成文的规定，即在国际学术交流中，凡是用 SPSS 软件完成的计算和统计分析，可以不必说明算法，由此可见其影响之大和信誉之高。

SPSS for Windows 由于操作简单，已经在我国的社会科学、自然科学的各个领域发挥了巨大作用。该软件还可以应用于经济学、生物学、心理学、地理学、医疗卫生、体育、农业、林业、商业、金融等各个领域。

### 4.3 试验性拟合回归分析

采用回归方法，对数据进行多元线性回归，通过拟合能耗与自变量之间的关

系, 得出在标准条件下的预测建筑能耗。通过预测的能耗与实际能耗的比较来判断出建筑是否节能。

进行多元线性回归的首要条件是要有足够的数据, 这样才能反映样本整体的一般特性, 拟合出来的曲线才能更加接近真实值, 更加具有代表性。

原始的酒店建筑能耗数据包括了全国四百多栋酒店详细的基本情况和能耗数据, 这样的数据量已经能够满足进行多元线性回归的条件, 因此在确定使用的数学统计方法之后, 采用输入法对数据进行拟合回归分析。

#### 4.3.1 拟合过程和结果

##### (1) 自变量和因变量的选择

试验性拟合采用的方法是输入法 (Enter Method), 将样本数据中所有相关自变量全部放入回归方程。由于建筑能耗数据是分星级统计的, 一般也会认为酒店的星级水平对建筑能耗水平有一定影响, 所以将星级当作自变量放入回归方程中。

方程中包含的自变量如下:

- 全年总能耗, kWh;
- 建筑面积,  $m^2$ ;
- 酒店星级;
- 平均入住率;
- 客房数;
- 人员数;
- 冷库数量; walk-in
- 开放式冰箱数量, open;
- 封闭式冰箱数量, close;
- 是否提供洗衣服务, yes/no;
- 是否提供健康俱乐部服务, yes/no;
- 是否提供车库服务, yes/no;
- 是否准备商务用餐, yes/no;
- 是否有零售, yes/no;
- 泳池面积,  $m^2$ ;

- 
- 供热面积比例;
  - 供冷面积比例;
  - 冰箱总数

在因变量的选择上, 由于本文的目的是为了拟合出酒店的能耗, 因此使用 2009 年总能耗作为拟合方程的因变量。

## (2) 拟合的结果

在这次拟合的中, 由于不清楚到底哪些自变量对建筑总能耗的影响程度大, 而且也不能无理由地剔除自变量, 因此将这些自变量全部代入方程中参加计算。

表6 各自变量与建筑总能耗的相关性分析

变量	相关性指标	变量	相关性指标
总面积	0.838	洗衣服务	0.147
星级	0.552	健康服务	0.415
入住率	<b>-0.192</b>	车库服务	0.183
客房数	0.592	商务用餐	0.239
LN 雇员数	0.854	零售店	<b>-0.097</b>
冷库数	0.781	泳池面积	0.276
开放式冰箱数	<b>-0.016</b>	供热面积比例	<b>-0.055</b>
封闭式冰箱数	0.318	制冷面积比例	<b>-0.077</b>
冰箱总数	0.343	/	/

由表 6 可以发现某些自变量的相关性为负值, 也就是说这些自变量的数值越低, 平均能耗越高, 说明这些自变量和建筑总能耗是成反向相关的, 从实际情况上来看, 这显然是不合理的。

对模型进行检验的其中一种方法是拟合优度检验, 即用样本决定系数的大小来衡量模型的拟合优度。拟合优度  $Adjusted R^2$  越大, 说明回归方程拟合原始数据  $y$  的观测值的效果越好<sup>(56)</sup>。用输入法进行拟合回归分析, 分析结果模型的拟

合结果见表 7。

表7 模型概述

统计方法	模型拟合优度	F 值
输入法	0.806	39.790

#### 4.3.2 拟合结果分析

最终曲线的拟合优度 Adjusted  $R^2$  为 0.806, 其效果还是不错的。但是很多自变量未通过 p 值检验, 且出现了很多不合理的现象。由于曲线拟合的优劣并不完全由拟合优度 Adjusted  $R^2$  来决定, 因此, 从综合的角度来看, 此次分析的效果并不理想。

分析造成本次分析效果不理想的原因, 发现了以下问题:

a、虽然样本所包含自变量数据庞大, 但其中某些数据存在明显不合理, 致使拟合的结果不好。为保证最终数据拟合的效果, 需要对原始数据进行筛选。

b、曲线拟合需要找到对能耗影响大的因素作为自变量。同时, 自变量太多也会引起曲线的不稳定性增加。当样本容量  $n$  较小, 变量个数  $p$  较大时,  $F$  或  $t$  检验自由度太小, 尽管  $R^2$  会较大, 但是估计的效果很不稳定。在这种情况下, 需要解决如何从很多自变量中选取合适自变量的问题。

c、尝试将星级当作自变量代入模型中进行回归分析, 但总体效果较差。一般来说, 不同星级酒店由于管理水平、服务水平、豪华水平的差异, 必然导致建筑能耗水平的差异。这样很容易认为“星级”一定是对建筑能耗影响巨大的变量。考虑到不同星级酒店建筑能耗的差异, 可以引起模型中变量结构的不同, 即模型中最合适的自变量组合会不同。因此, 将不同类型的自变量强制组合成同一个数据模型, 很难得到优秀的回归拟合结果。有必要分开对不同星级酒店建筑能耗进行分析。

#### 4.3.3 小结

通过本次试验性的数据拟合, 发现了很多问题, 比如对于自变量多、样本数量较大的数据, 应建立合理的筛选标准, 剔除其中一些明显不合理的样本, 避免数据错误, 影响拟合回归的效果。还发现应选择合适的自变量参与数据拟合回归, 并不是自变量越多越能说明问题。自变量太多反而会对方程的稳定性产生影响。

---

---

通过以上发现的问题，明确了数据的拟合回归工作并不是简单学会如何使用 SPSS 软件，绘制几张图表就能解决问题。应通过统计学工具与专业领域知识的协同工作，不断发现问题并解决问题，周而复始、循环上升，最终解决问题的过程。

#### 4.4 多元线性拟合回归分析的技术路线

通过试验性拟合回归分析中发现的问题以及分析的小结，对下一步的拟合回归分析做以下说明：

a、由于数据单位的问题，回归拟合分析首先使用单位最统一、数据完整性最好的耗电量作为因变量。

b、考虑到建筑面积与总能耗值的强相关性，高达 0.838，这也与建筑能耗的特点一致。为避免建筑面积强相关性对模型的影响，本次回归模型统一使用单位面积能耗作为因变量，该模型将仅体现除建筑面积外，其他自变量对建筑能耗的综合影响程度。

经过多次反复的拟合回归分析，统计学参考书籍<sup>(57)</sup>，对分析结果进行判断，确定了具体工作的技术路线。

- 样本的筛选
- 自变量的初次筛选
- 对数据进行中心化处理
- 模型的分析和修正

#### 4.5 样本的筛选

##### 4.5.1 样本筛选的意义

并不是数据库中的所有数据都适合参加数据分析。一种情况，由于数据获取过程中可能产生的种种问题，导致数据出现的未知偏差，甚至产生错误。这样的数据是不能参与分析的。另一种情况，样本中异常点的存在，虽然不属于错误数据，但该异常点与其他大部分样本相比有较大偏离，参与数据拟合分析后，对结果产生不利影响。这样的数据也应该删除。具体对上述两种情况进行分别说明。

a、数据错误对数据分析的影响

关键数据的缺失，造成该样本无效，例如总能耗、建筑面积和人员数等数据的缺失。

数据呈现明显错误，造成该样本无效，例如客房面积大于建筑总面积，折合单位电费 50 元/kWh。

### b、异常点对数据分析的影响

异常点与大多数样本点偏离巨大，例如如图 17 所示，红框内所指示的四个样本在图上的分布位置远远偏离其他样本点。这样的数据代入模型中计算，会导致模型结果偏差较大。

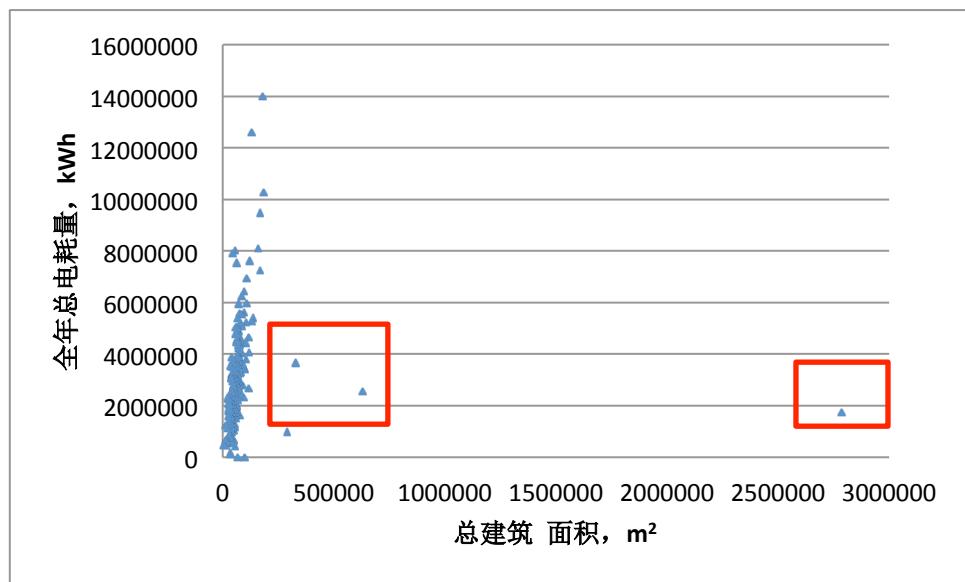


图17 五星级酒店建筑面积与总能耗散点图

#### 4.5.2 样本筛选标准的制定

结合上一节的结果，分析相关文献<sup>(58)</sup>，应确定样本数据的筛选标准，对原始的样本数据进行筛选，保证拟合回归分析的效果。

将数据筛选的标准分为三类：

➤ 建筑定义的限制标准

通过酒店建筑的定义，对样本进行筛选，例如星级变量为五星级，必须至少有一间客房等

➤ 建筑正常运营的限制标准

建筑内运营参数的合理范围，例如人员密度、房间密度和建筑能耗密度的合

---

---

理范围。

➤ 数据分析的限制标准

对关键性变量的要求，例如必须有建筑能耗数据、建筑总面积数据和人员总数等。

建筑定义和数据分析的限制标准比较容易理解和确定，但合理确定建筑正常运营的限制标准是个技术难题。可使用观察和分析散点图分布的方法

针对自变量范围的选择，首先做了新的分析，然后通过参考文献及相关分析确定了新的自变量的筛选范围。

a、单位面积建筑能耗数据的范围选取，在第二次拟合时，该变量的取值范围是大于 10 kWh 小于 400 kWh，具体选取方法和新的改进如下：

绘制以总建筑面积为横坐标、总用电能耗为纵坐标的散点图，如图 18 所示，由于不合理点的存在，导致散点图未出现有规律性分布。剔除异常点前，做散点图的一元线性拟合曲线，如图中虚线所示，并未反映出散点图的分布趋势。

在去掉几个明显偏离点后，做散点图的一元线性拟合曲线，如图中实线所示，充分体现了散点图的分布特点。

剔除散点图的分布变为如图 19 所示。做散点的一元线性趋势线，该趋势线经过原点，趋势线的斜率为 51.86，代表单位面积建筑能耗的平均值水平。做两条如图所示经过原点的直线，将绝大多数样本点包括在两条直线的范围内，对应的单位面积能耗为  $23 \text{ kWh/m}^2$  和  $180 \text{ kWh/m}^2$ ，并以此确定合理样本点的筛选依据。

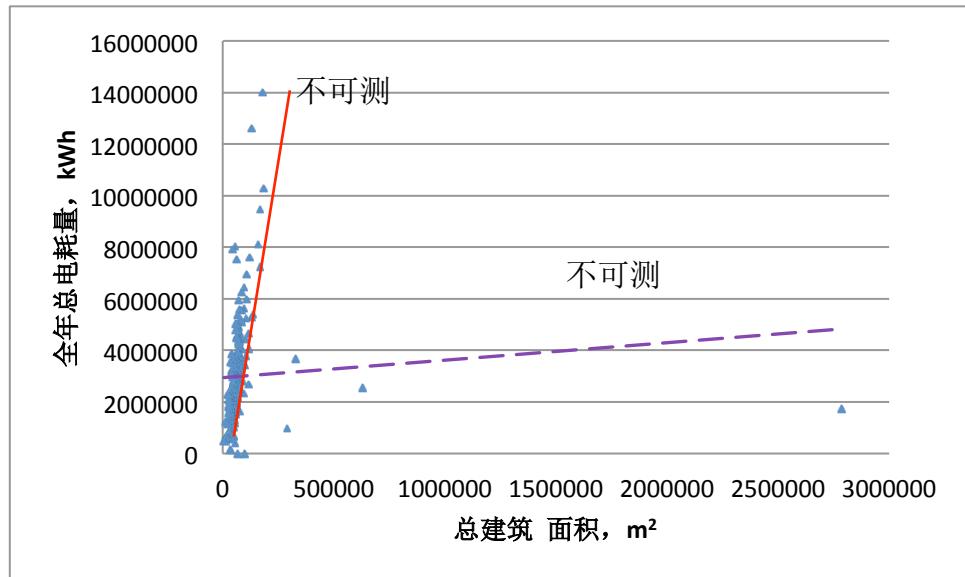


图18 五星级酒店建筑面积与总能耗散点图 1

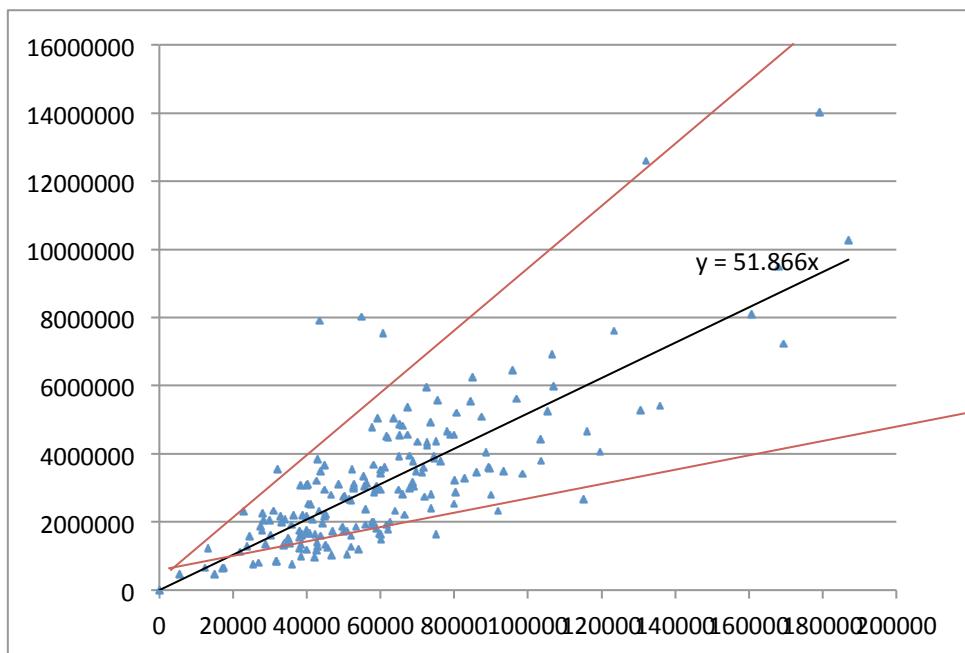


图19 五星级酒店建筑面积与总能耗散点图 2

b、千平米的房间数的选取。

五星级酒店标准对标间面积有严格的规定，要求“标间必须大于 20 平米”，根据这个原则和上节所述的方法，确定每一千平米房间数的取值范围为 3~16。

c、千平米人员数。

同理，确定每一千平米人员数的取值范围为 3~16。

确定如下的数据筛选标准。见表 8。

表8 数据筛选标准及筛选结果

筛选条件	解释	筛选标准类型	剩下建筑数量
五星级	选取全国范围内五星级酒店	酒店定义	316
至少有一名工作人员	确定是否是酒店	酒店定义	310
必须至少有一间客房	确定是否是酒店	酒店定义	300
开放时间必须大于200天	确定酒店开放天数是否合理	运营情况限制	292
员工数大于3小于26(人/ $1000m^2$ )	酒店人员密度是否合理	运营情况限制	261
房间数大于3小于25(间/ $1000m^2$ )	酒店房间的密度是否合理	运营情况限制	246
平均耗电量大于10kWh 小于 400kWh	酒店平均面积耗电量是否合理	数据分析限制	198
采暖、制冷面积的比例	必须具有供热、供冷面积比例	数据分析限制	179
必须具有 HDD18、CDD26	应有 HDD18、CDD26 数据	数据分析限制	158
冷库数量	要具有冷库	数据分析限制	142
是否准备商务用餐	应有该变量	数据分析限制	102

#### 4.6 自变量的筛选

在多元线性拟合回归分析中，并不是自变量越多越好。由于自变量之间不可避免的相关性，过多的自变量反而容易造成模型预测值的增大，破坏模型的稳定

性。

对自变量的筛选按照由粗到细的顺序进行筛选：

- a、观测各自变量与因变量分布的散点图；
- b、各自变量对因变量的贡献及相关性分析；
- c、拟合回归分析中对各自变量的 p 检验是否能通过。

通过以上标准对自变量进行筛选后，保留对因变量贡献较大的自变量，建立模型。

#### 4.6.1 散点图分析法

具体做法如下，首先选择单位面积能耗作为因变量。由于自变量种类较多，需要将每种自变量与因变量进行变化趋势图的比较，通过它们之间变化趋势的相关性来判断是否选择此自变量。

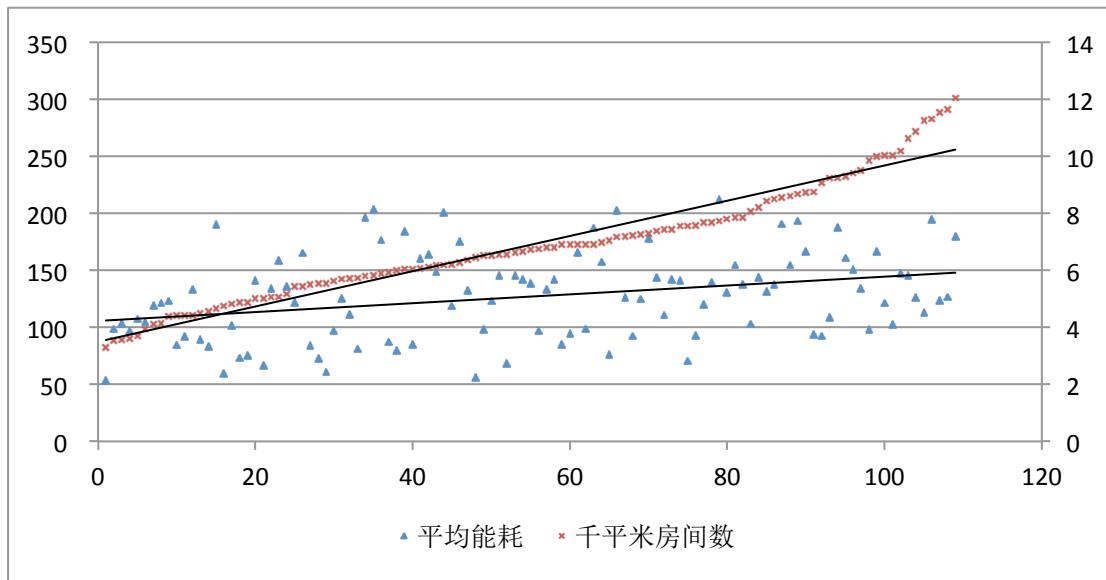


图20 单位面积能耗与客房密度的比较

从图 20 中可以看出，各样本建筑客房密度（用棕色十字叉表示）按照升序排列，观察单位面积平均能耗（用蓝色三角表示）的变化趋势。由于本模型采用单位面积平均能耗作为因变量，不考虑建筑面积对建筑能耗的影响。其他变量对建筑能耗的影响程度一般小于建筑面积，因此，从图 20 上看到单位面积的建筑能耗基本与客房密度保持一致的变化趋势。可以选定客房密度作为候选自变量。

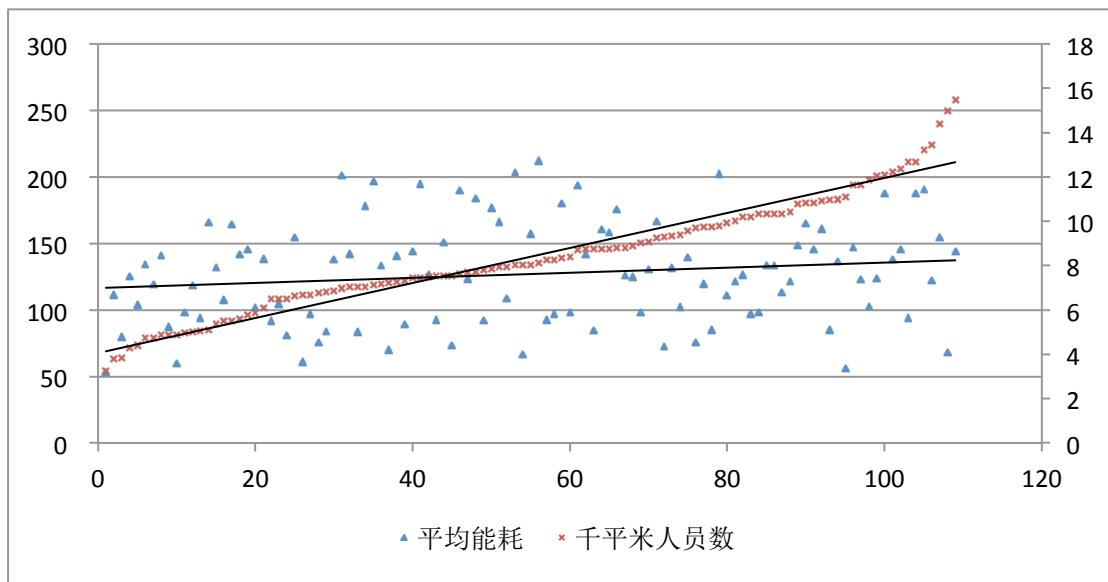


图21 单位面积能耗与人员密度的比较

结合自变量数据的完整性，依次剔除的自变量包括：建成年代、星级、是否提供洗衣服务、是否有 SPA、是否有零售店、是否有酒吧、游泳池类型、游泳池面积等变量，如表 9 所示。

表9 自变量的保留和剔除

保留的自变量	剔除的自变量
人员密度、房间密度、冷库密度、室内游泳池、商务餐、供热乘积、供冷乘积、HDD 和 CDD	建成年代、星级、是否提供洗衣服务、是否有 SPA、是否有零售店、是否有酒吧、游泳池类型、游泳池面积等变量

#### 4.6.2 相关性分析法

做各自变量与因变量的相关性分析，结果见表 10。

表10 各自变量与因变量的相关性分析

自变量	相关性	自变量	相关性
人员密度	0.425	供热乘积	0.005
房间密度	0.303	供冷乘积	0.040
冷库密度	0.561	HDD	0.013
室内游泳池	-0.100	CDD	0.035
商务餐	0.125	/	/

分析上表：

- a、室内游泳池与单位面积能耗的相关性为负值，该变量应予以剔除；
- b、供热乘积、供冷乘积、HDD 和 CDD 相关性指标都偏低。从专业角度解释，这几个自变量都应作为衡量天气因素对建筑能耗所产生影响大小的自变量，暂不剔除；在建立模型时通过 p 值检验的方式决定取舍；
- c、供热乘积和供冷乘积作为一组自变量，HDD 和 CDD 作为一组自变量，应在两组之间进行选择，而不是同时使用。

通过各个自变量与因变量之间的图形比较，并且参考了 Energy Star 的相关筛选的步骤及结果，确定参加模型分析的待选自变量如下：

- ◆ 人员密度，(人/1000m<sup>2</sup>)
- ◆ 房间密度，(间/1000m<sup>2</sup>)
- ◆ 冷库密度，(个/1000m<sup>2</sup>)
- ◆ 供热乘积 = 供热面积比例 × HDD18
- ◆ 供冷乘积 = 供冷面积比例 × CDD26
- ◆ HDD18
- ◆ CDD26
- ◆ 是否提供商务餐

## 4.7 中心化的意义和实现方法

### 4.7.1 中心化的意义

为了检验各个自变量的显著性以及对方程的贡献如何，首先进行了全部进入法的拟合。拟合结果见表 11 和表 12。

表11 模型概述

统计方法	模型拟合优度	F 值
输入法	0.460	11.325

表12 自变量显著性检验

模型	非标准化系数		Beta	t	显著性
	B	标准误			
1 (常量)	-19.619	29.972		-.655	<b>.514</b>
	人员密度	6.615	1.310	.440	.000
	房间密度	-1.475	1.757	-.077	.203
	冷库密度	161.307	24.050	.510	.000
	商务餐	17.227	6.965	.184	.015
	供热乘积	-.026	.012	-.822	.033
	供冷乘积	.013	.063	.040	.840
	HDD	.033	.013	1.106	.011
	CDD	.087	.067	.311	.200

该模型 p 值检验结果中可以发现, 常数项的 P 为 0.514, 未通过显著性检验。通过其他多次的回归分析, 所建立的预测模型中的常数项的 p 值都较大, 无法通过显著性检验。在去掉常数项后, 建立预测模型, 模型概述见表, 自变量检验结果见表 13 和表 14。

表13 模型概述

统计方法	模型拟合优度	F 值
输入法	0.955	261.9

表14 自变量显著性检验

模型	非标准化系数		Beta	t	显著性
	B	标准误			
1 人员密度	6.425	1.274	.425	5.044	.000
	房间密度	-1.782	1.688	-.096	.294
	冷库密度	160.842	23.971	.291	.000
	商务餐	15.753	6.572	.152	.018
	供热乘积	-.024	.011	-.369	.041
	供冷乘积	.012	.063	.026	.854
	HDD	.029	.011	.507	.010
	CDD	.063	.056	.164	.267

从第二次拟合的效果来看, 模型的拟合优度由原来的 0.460 增长到现在的

0.955, 说明现在的拟合曲线的拟合程度已经非常理想。

从表 14 可以看出, 部分自变量的显著性都有明显的下降, 说明在去除常数项之后, 部分自变量的效果都有很大的改进, 除了供冷乘积、房间密度和 CDD 这几个自变量的显著性还有些偏大之外, 其他的自变量的显著性都已经非常优秀。

因此, 会产生一种错觉, 认为模型中常数项的出现, 对模型拟合优度和各自变量的显著性检验造成不利影响, 应消除常数项。

绘制人员密度和单位面积能耗的散点图, 并用 S 函数拟合回归, 做回归曲线, 如图 22 所示。用 S 函数做拟合回归, 模型的拟合优度和 F 值检验均优于一元线性拟合回归模型 (详见第五章)。该模型的拟合优度和 F 值见表 15。

表15 模型概况

模型编号	拟合优度 Adjusted R <sup>2</sup>	平均相对误差	F 值
S 型曲线模型	0.210	0.285	29.783

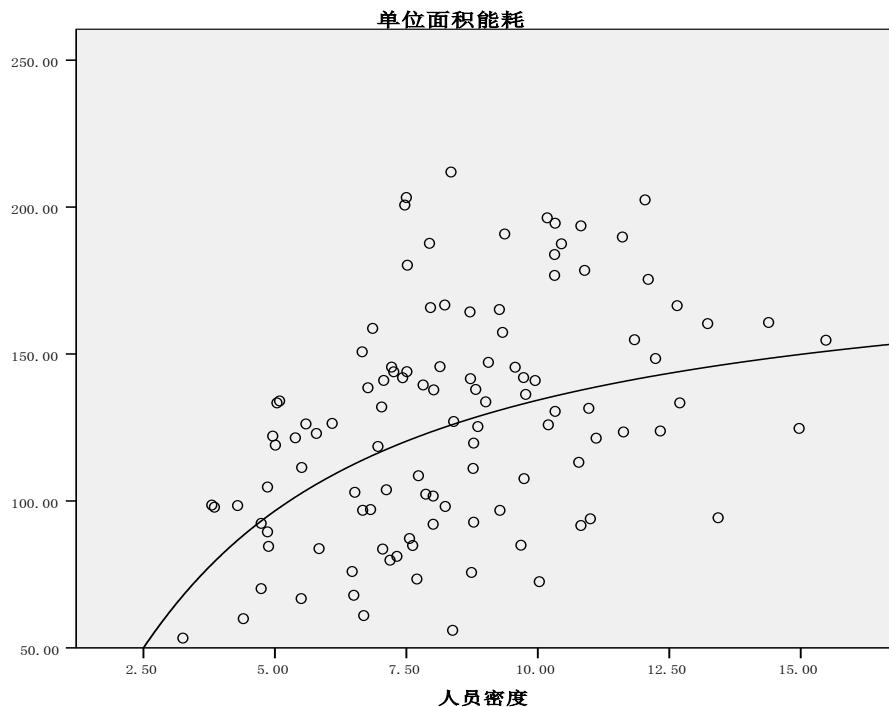


图22 S 模型的散点图和拟合曲线

由于模型的拟合优度较小, 考虑到常数项的干扰, 去掉常数项后重新建立预测模型, 改进模型的拟合优度和 F 值见表 16, 散点图和拟合曲线见图 22。

表16 模型概况

模型编号	拟合优度 Adjusted R <sup>2</sup>	平均相对误差	F 值
改进的 S 型曲线模型	0.871	1.724	745.965

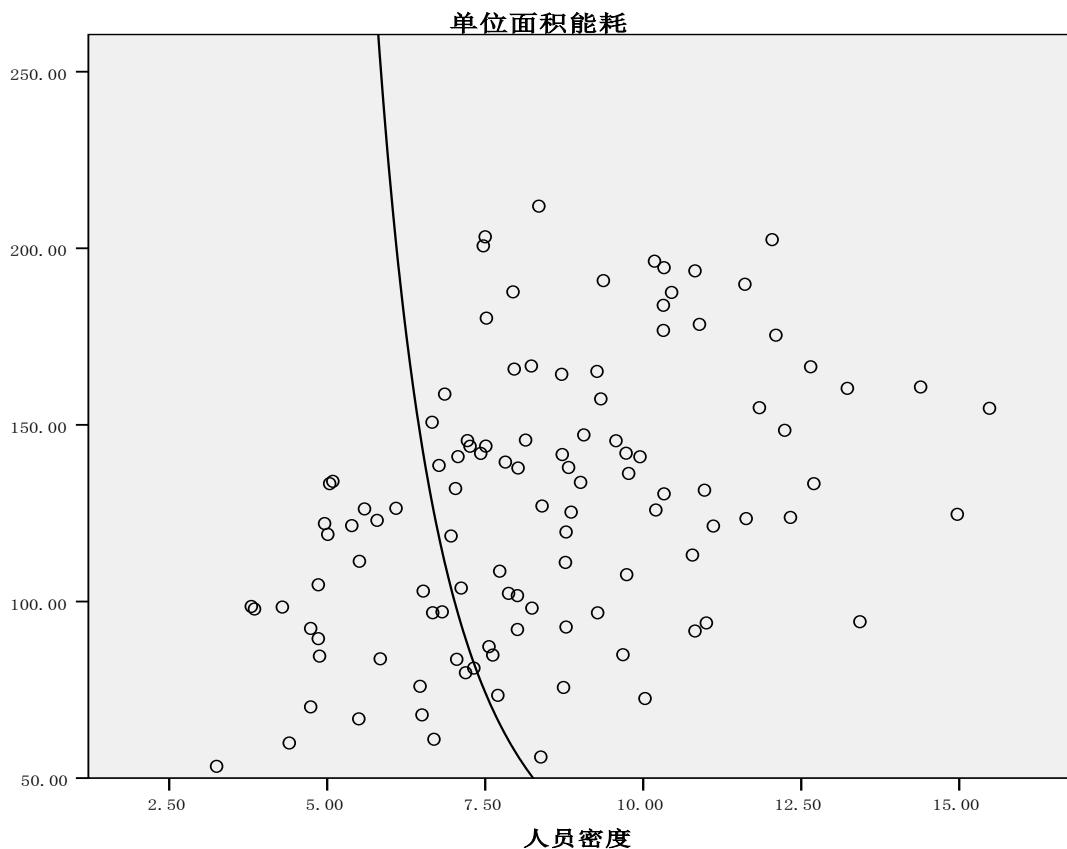


图23 改进 S 模型的散点图和拟合曲线

表17 普通和改进的 S 型曲线模型模型概述

模型编号	拟合优度 Adjusted R <sup>2</sup>	平均相对误差	F 值
S 型曲线模型	0.210	0.285	29.783
改进的 S 型曲线模型	0.871	1.724	745.965

从表 17 中可以看出，改进后的 S 型曲线模型拟合优度得到了迅速提升，从 0.210 提升为 0.871，F 值检验也从 29.783 提升为 745.965。这两个关键性参数的改善似乎说明了去掉常数项后模型性能的优越性。但从图 23 中，可以看出，拟合曲线的趋势完全没有体现出散点图分布的特点。

在 SPSS 的帮助文档中这样解释<sup>(59)</sup>：“在默认情况下，回归模型包含常数项。去掉常数项后，意味着强迫模型通过原点，而这是很少出现的情况。某些通过原

点的拟合模型的预测结果无法与其他不通过原点的模型相比。例如，在这种情况下， $R^2$ 就不具有可比性。”这说明了在表 17 中两个模型  $R^2$ 、 $F$  值的比较无意义。

从影响建筑能耗的机理上分析，通过多元线性拟合回归公式作为模型预测建筑能耗，公式中每一项都能分别产生对建筑能耗的影响。公式中常数项的意义可理解为对其他自变量解释力的补充。当所有自变量为零时，常数项表示了建筑的基础能耗。模型在去掉常数项后，认为模型所包括的自变量能够解释所有建筑能耗的变化情况，这是不合理的。因此，作为模型一个必要组成部分的常数项，不能去掉。

在多元线性回归中，因为涉及多个自变量，自变量的单位往往不同，这给利用回归方程进行结构分析带来一定困难。并且多元回归涉及的数据量很大，可能由于舍入误差而使计算结果不理想。尽管计算机能使我们保留更多位的小数，但是舍入误差肯定会出现。因此对原始数据进行一些处理，尽量避免大的误差是有实际意义的。

产生舍入误差有两个主要原因：一是回归分析计算中，各变量数据的量级有很大差异，例如 10000000 和 0.1，大小差异悬殊的数据出现在同一个计算中；二是病态矩阵的影响，会产生较大的误差。

将数据进行中心化能减少运算的自变量个数，同时能减少误差，因此很多时候会将数据进行中心化处理。

中心化的意义是，通过中心化消去常数项，以此来保证回归分析能够得到较高的  $R^2$ 。然后通过计算得到正常公式内的常数项。在变换过程中， $R^2$  不变。

#### 4.7.2 中心化的方法

多元线性拟合回归模型的一般形式为：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

其经验回归方程为：

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p$$

此经验回归方程经过样本中心  $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p; \bar{y})$ ，将坐标原点移至样本中心，即坐标变换：

$$\begin{aligned}x_{ij}^{'} &= x_{ij} - \bar{x}_j, i = 1, \dots, n; j = 1, \dots, p \\y^{'} &= y_i - \bar{y}, i = 1, \dots, n\end{aligned}$$

上述经验方程即变换为

$$\hat{y}^{'} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{'} + \hat{\beta}_2 x_2^{'} + \dots + \hat{\beta}_p x_p^{'}$$

此即为中心化经验回顾方程。

中心化回归方程的常数项为 0，而回归系数的最小二乘估计值  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$  保持不变，这是因为坐标系的平移变换只改变直线的截距，不改变直线的斜率。

中心化后的回归方程只包含了  $p$  个参数估计值  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$ ，比原回归方程减少了一个未知的参数。在变量较多时，减少一个未知参数，计算工作量都会减少许多，对计算来说能减少部分运算。

最终再由公式  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 - \dots - \hat{\beta}_p \bar{x}_p$  求出常数项  $\hat{\beta}_0$ 。

#### 4.8 建立电耗基准模型

用逐步回归法，剔除  $p$  值显著性检验不合格的自变量供热乘积和供冷乘积。由于在多次拟合回归中，房间密度的  $p$  值一直较大，但考虑房间的情况对建筑能耗有较大影响，因此在模型中，保留房间密度作为自变量。

建立的单位面积能耗预测模型自变量统计信息见表 18，模型概述见表 19，自变量显著性验证见表 20。

表18 自变量的统计信息

变量名	平均值	最大值	最小值
单位面积能耗	127.0	211.9	53.3
人员密度	8.4	15.5	3.3
房间密度	6.9	12.1	3.3
冷库密度	0.2	0.7	0.0
是否提供商务餐	0.2	1.0	0.0
HDD	1917.0	6042.0	38.0
CDD	319.9	572.0	62.0

表19 单位面积能耗模型概述

拟合优度 Adjusted R <sup>2</sup>	平均相对误差	F 值
0.431	28.869	14.787

表20 自变量显著性验证

模型	非标准化系数		Beta	t	显著性
	B	标准误			
1	z人员密度	2.404	.666	.341	3.607
	z房间密度	.546	.883	.059	.618
	z冷库密度	54.556	11.606	.389	4.701
	z商务餐	3.932	3.501	.091	1.123
	zHDD	.006	.002	.461	2.693
	zCDD	.067	.022	.522	3.043

表21 ANOVA 分析

模型	平方和	df	均方	F	显著性
1	回归	73945.241	6	12324.207	.000(a)
	残差	85843.609	103	833.433	
	合计	159788.85	109		

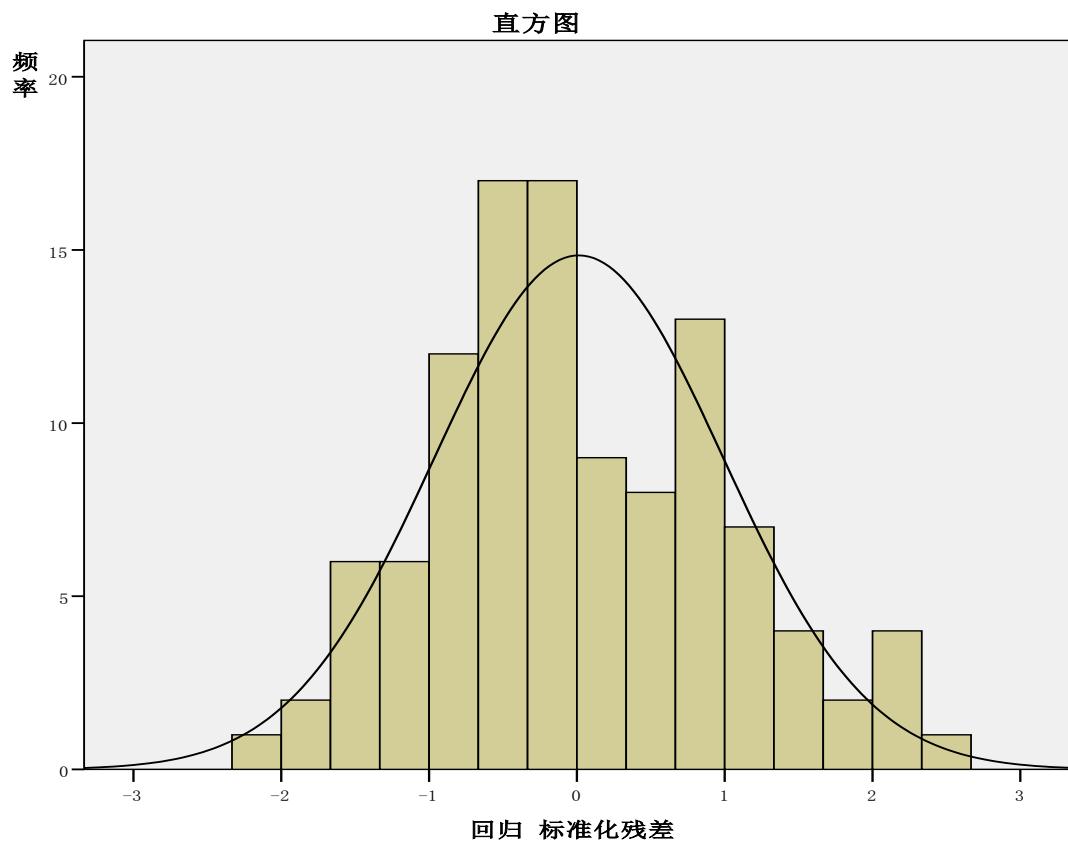


图24 回归标准化残差直方图

回归 标准化残差 的标准 P-P 图

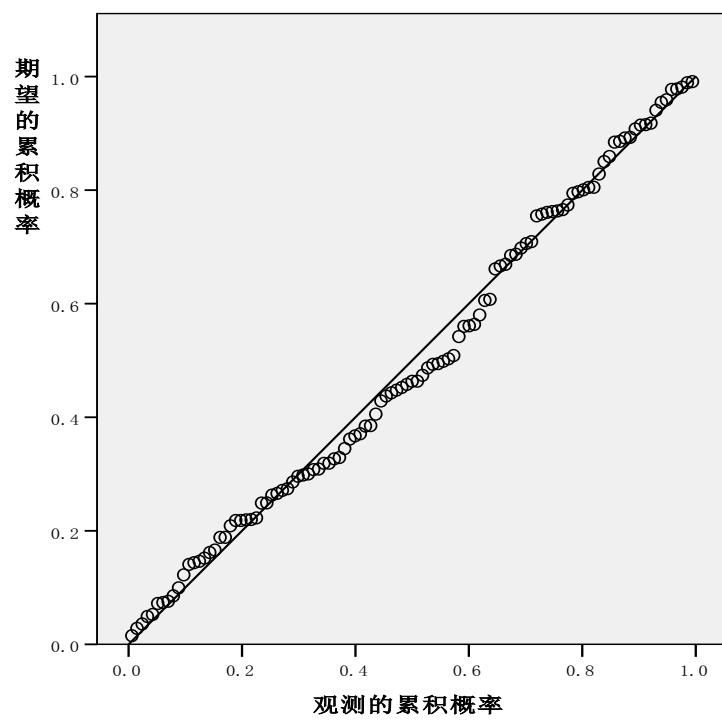


图25 回归标准化残差 P-P 图

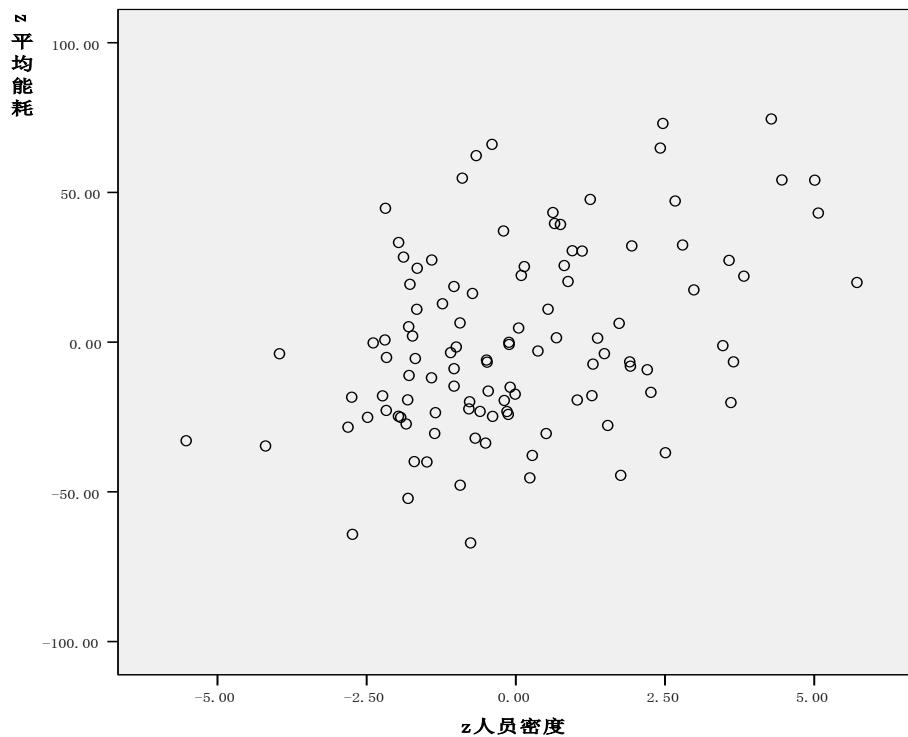


图26 人员密度残差散点图

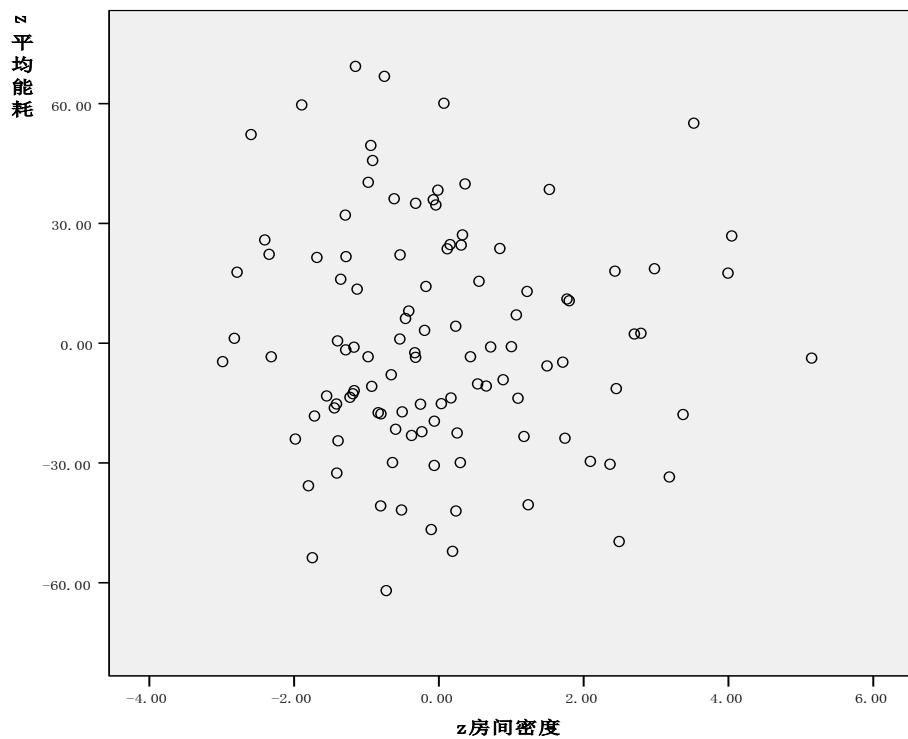


图27 房间密度残差散点图

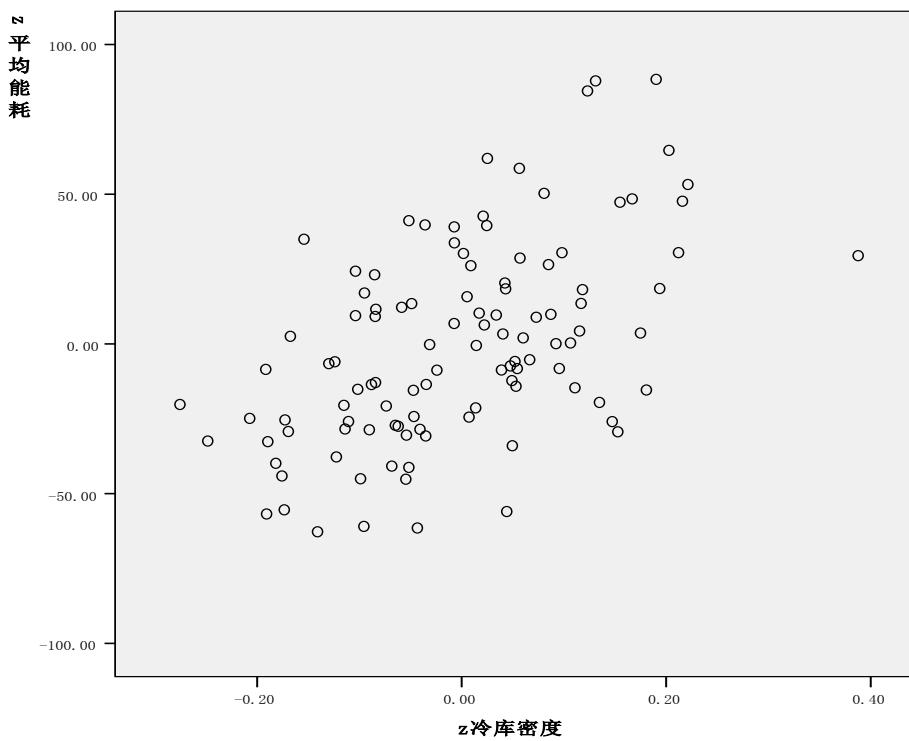


图28 冷库密度残差散点图

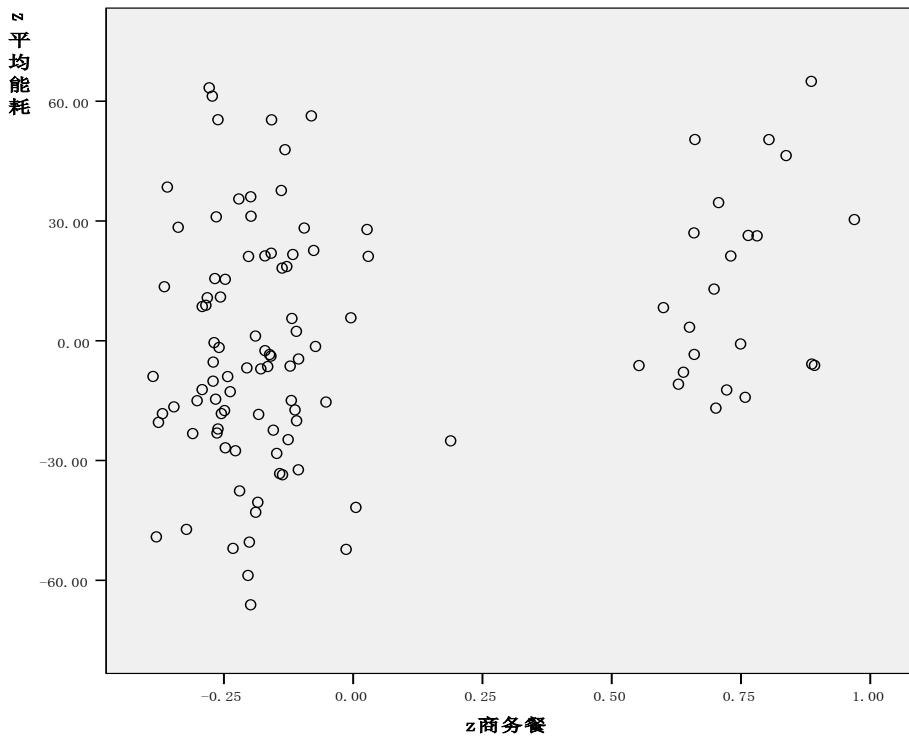


图29 商务餐残差散点图

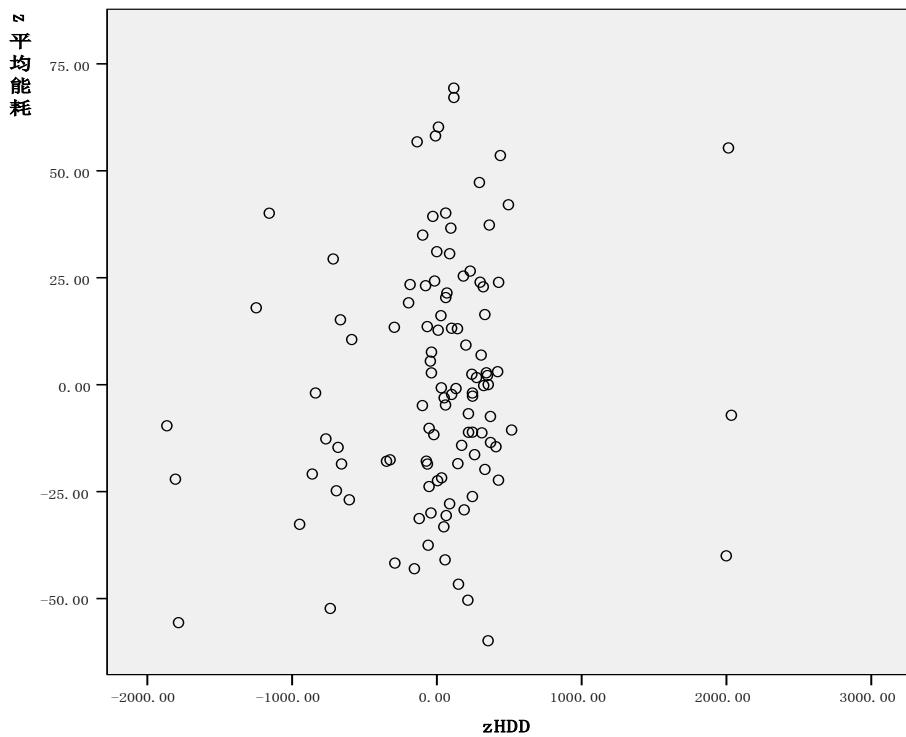


图30 HDD 残差散点图

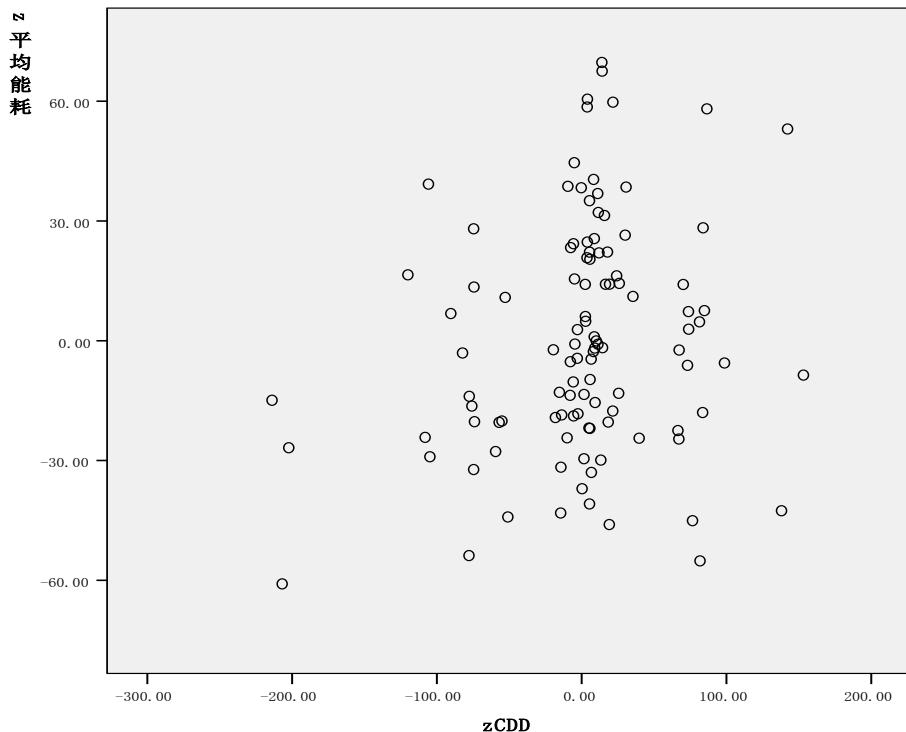


图31 CDD 残差散点图

从图24中可以看到拟合的结果比较理想,最终的R值为0.431,F值为14.787,

通过了 F 检验, 从显著性来看, 除了房间密度之外, 其他自变量的显著性都很好, 证明拟合结果很好。

从回归标准化残差直方图来看, 基本符合正态分布; 从图 24 回归标准化残差 P-P 图来看, 所有的点都几乎在 45° 线上, 这是最理想的结果, 同样证明了拟合结果很好。

图 26-28 分别为人员密度、房间密度、冷库密度残差图, 从图中可以看出, 各个点在图中 0 点周围平均随机分布, 说明这些自变量在进入回归方程时使回归方程与实测值的拟合情况较好。证明这些自变量对

图 29-31 分别为商务餐、HDD、CDD 的残差图, 从图中可以看出, 商务餐、HDD、CDD 的残差图基本在 0 的周围, 从纵坐标上的分布看, 其分布有些散, 但是其依然是均匀散布, 因此影响不大, 而且 HDD、CDD 作为重要的自变量必须将其加入方程中。

由于未考虑建筑总面积作为自变量, 所以本模型拟合优度达到 0.431 已经比较理想, 证明本模型能预测单位面积能耗 43.1% 的情况。

建立的建筑总电耗预测模型的概述见表 22, 自变量显著性验证见表 23。

表22 建筑总电耗模型概述

拟合优度 Adjusted R <sup>2</sup>	平均相对误差	F 值
0.879	1740669	113.711

表23 自变量显著性验证

模型	非标准化系数		Beta	t	显著性
	B	标准误			
1	z建筑总面积	50.259	10.039	.326	5.007
	z人员总数	9726.833	1196.466	.442	8.130
	z客房总数	629.530	1655.642	.021	.380
	z冷库总数	172707.67	25168.536	.285	6.862
	z商务餐	746771.62	377797.02	.069	1.977
	zHDD	643.309	290.496	.164	2.215
	zCDD	7086.001	2720.134	.194	2.605

表24 ANOVA 分析

模型	平方和	df	均方	F	显著性
1 回归	2.41E+015	6	3.45E+014	113.711	.000(a)
残差	3.09E+014	102	3.03E+012		
合计	2.72E+015	109			

## 4.9 建立总能耗基准模型

### 4.9.1 一次能源的转化计算

酒店统计的消耗能源有电、天然气、蒸汽和冷冻水，其中冷冻水不在考虑的范围之内，剩下的三种是考虑的内容，但是蒸汽用量数据的单位不统一，且不清楚其压力情况，很多酒店没有对蒸汽的使用情况进行统计，造成了很多数据空白，因此鉴于以上因素，本文不考虑蒸汽的能耗，最终选定将电和天然气折算为标煤。

电能的当量转化标煤系数为 0.1229kgce/kWh，取煤电的发电效率为 38%，则电能的等价转化标煤系数为 0.3234kgce/kWh；气田天然气的当量转化标煤系数为 1.2143kgce/m<sup>3</sup>。

### 4.9.2 拟合的结果及分析

在解决了以上的问题之后，分别建立了单位面积能耗预测模型和总能耗预测模型。

单位面积能耗预测模型的结果见表 25、表 26 和表 27。

总能耗预测模型的结果见表 28、表 29 和表 30。

表25 单位面积能耗模型概述

拟合优度 Adjusted R <sup>2</sup>	平均相对误差	F 值
0.358	14.290	10.845

表26 单位面积能耗模型自变量显著性验证

模型	非标准化系数		Beta	t	显著性
	B	标准误			
1 z人员密度	2.404	.666	.341	3.607	.000

z房间密度	.546	.883	.059	.618	.538
z冷库密度	54.556	11.606	.389	4.701	.000
z商务餐	3.932	3.501	.091	1.123	.164
zHDD	.006	.002	.461	2.693	.008
zCDD	.067	.022	.522	3.043	.003

表27 单位面积能耗模型 ANOVA 分析

模型	平方和	df	均方	F	显著性
1 回归	13288.529	6	2214.755	10.845	.000(a)
残差	20422.412	100	204.224		
合计	33710.940	106			

表28 总能耗模型概述

拟合优度 Adjusted R <sup>2</sup>	平均相对误差	F 值
0.718	862219	54.091

表29 总能耗模型自变量显著性验证

模型	非标准化系数		Beta	t	显著性
	B	标准误			
1 z建筑总面积	15.983	5.210	.279	3.068	.003
z人员总数	3162.045	613.774	.393	5.152	.000
z客房总数	487.347	829.154	.045	.588	.558
z冷库	65580.435	12560.501	.297	5.221	.000
z商务餐	149939.86	216261.23	.034	.693	.160
zHDD	412.508	144.128	.289	2.862	.005
zCDD	4388.426	1348.749	.331	3.254	.002

表30 总能耗模型 ANOVA 分析

模型	平方和	df	均方	F	显著性
1 回归	2.81E+014	7	4.02E+013	54.091	.000(a)
残差	7.36E+013	99	7.43E+011		
合计	3.55E+014	106			

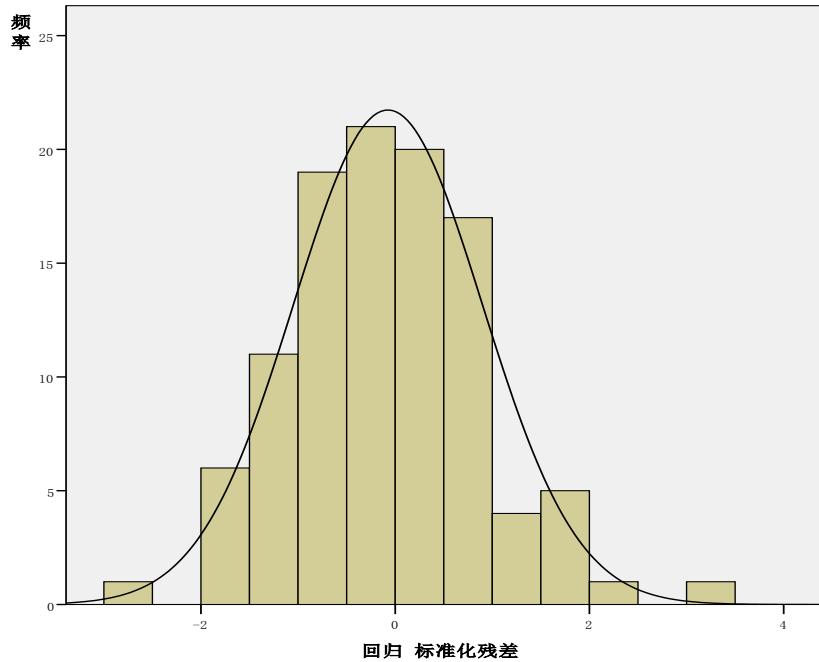


图32 回归标准化残差

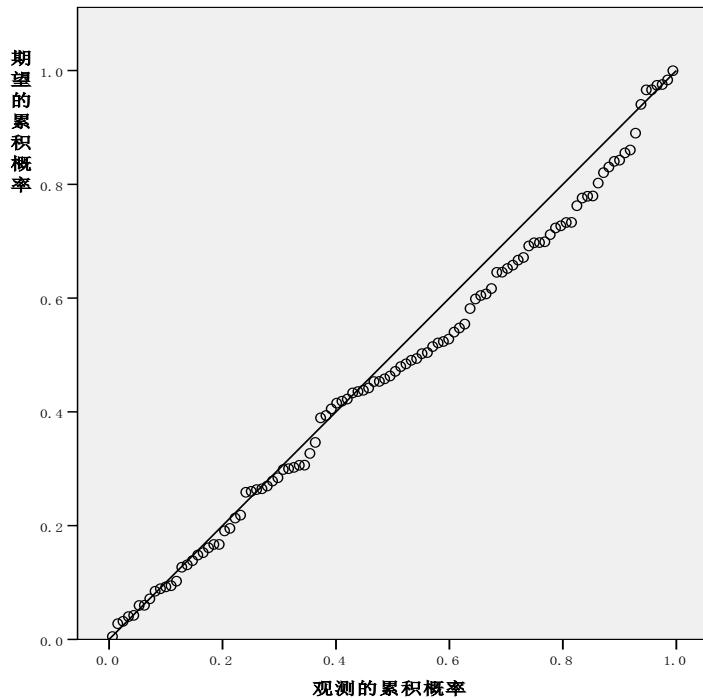


图33 P-P 图

从拟合结果来看,  $R$  为 0.358,  $F$  检验通过, 各个自变量的显著性都很好, 从残差图中可以看出, 残差基本平均分布, 没有出现不合理的残差情况, P-P 图的显示同样证明了结果是优良的。

在拟合之后, 根据 Energy Star 的经验, 本文同样做了总能耗的检验, 最终

---

---

的 R 为 0.718，是很理想的结果。

#### 4.10 本章小结

本章主要介绍了采用多元线性回归工具，建立酒店建筑能耗基准模型的过程。在实际工作中，通过四次较为系统的拟合分析，结合相关文献，确定了技术思路，并基本建立了酒店建筑能耗预测模型的公式，从验证的结果来看是较为理想的。

通过多次拟合发现了原始数据准确度的一些问题，对模型的建立带来的一定困难，例如变量的单位不统一，变量的概念混淆导致数据出现错误等情况。

## 第五章 多元线性回归模型的改进

### 5.1 改进的非线性多元回归模型

由于原理简单、解释力强，多元回归模型在经济、生物、医学等领域得到了广泛的应用。上一章所使用的是目前在国外广泛使用在建筑能耗基准建立的多元线性回归模型，也是由美国 EnergyStar 所提出的经典模型。多元线性回归模型的标准形式为：

$$y = C_0 + C_1x_1 + C_2x_2 + \cdots + C_nx_n$$

用于反映预测量  $y$  与各自变量  $x_1, x_2, \dots, x_n$  之间线性依赖的关系。每一项  $C_nx_n$  中，自变量  $x_n$  的因子  $C_n$  在一定程度上能够反应出  $x_n$  对预测量  $y$  的影响方向和影响程度。多元线性回归模型的拟合优度指标：Adjusted  $R^2$ ，可用于显示预测模型对实际模型的重现程度。该模型便于理解、易于应用和解释的特点，使其应用广泛。

统计学工具是对大量数据进行分析，挖掘数据中潜在特征信息的工具。模型的选择，对最终结果的判定意义重大。我们在完成用多元线性拟合回归模型建立建筑能耗基准的工作后，有必要反思该模型的适用性和改进的空间。

在实际情况下，自变量与因变量简单线性关系的情况较少，更多的时候，呈现出非线性关系。当然，具体表现出何种非线性关系还是未知。这时，需要给定几种简单的函数转化形式，通过不同形式对自变量的组合和变化，挖掘最适合的非线性函数，再代入多元线性拟合回归模型中，采用逐步回归法进行计算。这样可以用于扩展多元回归模型的适用范围。但该方法存在由于自变量较多、函数形式较多，所带来的数据处理量大的难题。如果有  $P$  种函数转化形式，有  $N$  种备选自变量，那么采用完全的排列组合，对每种改进的模型进行验算的话，就一共需要  $P^N$  次模型建立和验算工作。给实际数据处理带来很大的麻烦。

中科院应用数学所的学者项静恬<sup>(60)</sup>等，从数学统计专业的角度提出一种简化的模型改进方式。具体方法如下：

a、根据建筑能耗与各自变量的关系，提供备选的函数变化公式，包括：

表31 常用参考模型

模型编号	模型名称	英文名称	模型公式
1	线性模型	Linear	$y = A + Bx$
2	对数曲线模型	Logarithmic	$y = A + B \ln x$
3	逆曲线模型	Inverse	$y = A + \frac{B}{x}$
4	二次曲线模型	Quadratic	$y = A + Bx + Cx^2$
5	三次曲线模型	Cubic	$y = A + Bx + Cx^2 + Dx^3$
6	混合曲线模型	Compound	$y = AB^x$
7	幂函数曲线模型	Power	$y = Ax^B$
8	S型曲线模型	S	$y = \exp\left(A + \frac{B}{x}\right)$
9	指数曲线模型	Exponential	$y = Ae^{Bx}$
10	逻辑曲线模型	Logistic	$y = \left(\frac{1}{c} + Ae^x\right)^{-1}$

b、将因变量  $y$  依次与各自变量  $x_1, x_2, \dots, x_n$  建立上述一元拟合回归模型，比较各模型  $F$  值大小，最大  $F$  值对应的模型作为该自变量选定的模型，记  $z_n$  作为自变量  $x_n$  的转换形式。

c、将转换形式  $z_n$  代入原多元线性拟合规格模型中，用逐步回归法，对新模型  $y = B_0 + B_1 z_1 + B_2 z_2 + \dots + B_n z_n$  进行计算和验证。

## 5.2 改进模型的分析和建立

选择待选自变量人员密度和房间密度，采用表所述的 10 种模型，分别建立自变量和因变量的一元回归模型，所有模型参数见表 32 和表 33。

表32 人员密度模型

模型编号	拟合优度 Adjusted R <sup>2</sup>	F 值
线性模型	0.183	23.283
对数曲线模型	<b>0.198</b>	<b>25.741</b>

逆曲线模型	0.189	24.205
二次曲线模型	0.216	14.151
三次曲线模型	0.239	10.691
混合曲线模型	0.177	22.313
幂函数曲线模型	0.188	24.118
S型曲线模型	0.177	22.334
指数曲线模型	0.177	22.313
Logistic	0.177	22.313

表33 客房密度模型

模型编号	拟合优度 Adjusted R <sup>2</sup>	F 值
线性模型	0.121	14.356
对数曲线模型	0.136	16.358
逆曲线模型	0.140	16.944
二次曲线模型	0.146	8.835
三次曲线模型	0.147	5.863
混合曲线模型	0.128	15.310
幂函数曲线模型	0.128	15.310
S型曲线模型	0.142	16.247
指数曲线模型	<b>0.141</b>	<b>17.056</b>
Logistic	0.128	15.310

根据表和表的结果,选择对数曲线模型作为人员密度的转化形式,选择幂函数曲线模型作为客房密度的转化形式,即用以下形式代替原自变量。

$$x' = \ln x_{\text{人员密度}}$$

$$x' = \text{EXP}(x_{\text{房间密度}}/100)$$

经过分析计算,发现代入 $x' = x_{\text{房间密度}}^{\beta}$ 的转化形式后,该自变量仍无法通过  $p$  值检验,但代入 $x' = \text{EXP}(x_{\text{房间密度}}/100) + x_{\text{房间密度}}$ ,就能得到相对较好的结果。证明复合函数能充分代表自变量房间密度与因变量的关系,单独拟合该复合函数与因变量的结果也能说明这一点。

根据以上两个自变量变换后的模型结果见表 34 和表 35。

表34 改进的单位面积能耗模型概述

模型	拟合优度 Adjusted R <sup>2</sup>	平均相对误差	F 值
多元线性拟合回归模型	0.358	14.290	10.845
改进的模型	0.409	13.775	11.378
EnergyStar 模型	0.367	/	13.04

表35 改进的单位面积能耗模型自变量显著性验证

模型	非标准化系数		标准化系数 Beta	t	显著性
	B	标准误			
1 (常量)	894.807	481.161		1.860	.066
LN人员密度	20.167	5.266	.353	3.829	.000
房间密度	10.152	5.160	1.106	1.967	.052
exp客房密度	-935.271	481.701	-1.090	-1.942	.055
冷库密度	63.567	11.859	.434	5.360	.000
商务餐	2.555	3.474	.059	.735	.064
HDD	.007	.002	.478	2.884	.005
CDD	.069	.021	.535	3.232	.002

通过改进的模型与原模型的比较,可以看出改进的模型不仅拟合优度从 0.358 上升为 0.409,模型的 F 值从 10.845 上升为 11.378,而且房间密度的显著性指标  $p$  值也从 0.538 下降为 0.052。改进模型性能得到了大幅度提升。

### 5.3 酒店建筑能耗基准模型的计算公式

在完成五星级酒店建筑能耗基准模型的建立后,用同样的方法进行三星级和四星级酒店能耗的拟合。

### 5.3.1 五星级酒店建筑能耗基准模型

$$\begin{aligned}
 y = & -9.950E06 + 19.934x_1 + 1.307E06LN(x_2) + 1.257E03x_3 \\
 & - 3.376EXP(x_3/100) + 6.322E04x_4 + 1.247E05x_5 \\
 & + 4.658E02x_6 + 4.932E03x_7
 \end{aligned}$$

其中:  $x_1$ ——建筑总面积,  $m^2$ ;

$x_2$ ——人员总数, 人;

$x_3$ ——客房总数, 间;

$x_4$ ——冷库总数, 个;

$x_5$ ——是否准备商务餐, 是/否;

$x_6$ ——HDD18;

$x_7$ ——CDD26。

### 5.3.2 四星级酒店建筑能耗基准模型

$$\begin{aligned}
 y = & -1.826E06 + 32.760x_1 + 2.680E05LN(x_2) - 5.974E02EXP\left(\frac{x_3}{100}\right) \\
 & + 1.385E02x_4 + 1.443E03x_5
 \end{aligned}$$

其中:  $x_1$ ——建筑总面积,  $m^2$ ;

$x_2$ ——人员总数, 人;

$x_3$ ——客房总数, 间;

$x_4$ ——HDD18;

$x_5$ ——CDD26。

### 5.3.3 三星级酒店建筑能耗基准模型

$$y = -1.50E06 + 14.880x_1 + 1.829E03x_2 + 2.593E02x_3 + 3.214E03x_4$$

其中:  $x_1$ ——建筑总面积,  $m^2$ ;

$x_2$ ——人员总数, 人;

$x_3$ ——HDD18;

$x_4$ ——CDD26。

### 5.3.4 汇总

以上就是拟合出的全部公式, 具体的 R 值和 F 检验值见表 36

---

---

表36 模型的 R 值和 F 值

酒店类型	总能耗模型	
	Adjusted R <sup>2</sup>	F
五星	0.759 (106 栋)	42.394
四星	0.602 (95 栋)	29.384
三星	0.780 (25 栋)	22.296

从拟合的结果来看，所有的公式全部通过 F 检验，自变量的显著性较好，而且各模型的 R 值的大小同 Energy Star 相比，几乎接近。

在进行三星、四星酒店数据拟合时发现，三星、四星酒店的统计数据相比五星级酒店要差，尤其是三星级酒店的数据，很多都出现了缺失的情况，这样就使参与进入拟合的酒店数目减少，对拟合的准确性有一定的影响，同时发现三星、四星的数据比五星级酒店的数据有较强的波动性，说明在统计时出现了一些错误数据。虽然三星模型的拟合优度较高，但酒店数量较少，F 值较低，并不能证明三星模型的性能更高。这也证明了，模型的 R<sup>2</sup> 不是体现模型性能的唯一指标，必须与其他指标联合使用。

## 第六章 评价方法的建立

### 6.1 建筑能耗评价工具的建立

评价一个建筑是否节能,可以采用很多方法。指标法是最简单、最易于理解,也是应用最广泛的方法。根据指标数量的不同,将指标法分为单一指标法和复合指标法。一般单一指标法所采用的指标来源于大量的建筑能耗数据的统计,或建筑能耗模拟数据。由于建筑能耗差异巨大,采用单一指标对大量建筑进行建筑是否节能的判别,存在一定不合理性。建筑业主或使用者很容易提出对单一指标公平性的疑问。一般的修正方法是对建筑类型进行更加细致的划分,例如德国 VDI 3807 标准的技术思路。但对于我国,建筑数量巨大,建筑功能复杂,如何建立合理的建筑类型划分标准,又成了一个新的技术难题。

由于复合指标法使用各种类型评价指标,更多的用于建筑综合性能的评价,例如绿色建筑评价、LEED 评价等。但对于多指标的评价系统,又增加了一个技术难题:各指标之间的权重如何确定。国内一些学者提出用建筑用能系统中的多个性能参数进行综合评价的方法,例如周权<sup>(61)</sup>对建筑用能系统中各大主要设备制定了指标,包括水管路输送能效比、风管路单位风量耗功率、冷机 COP 等指标。对于各一个指标,给出优、良、中、差四个等级和对应的数值范围。复合指标法一般不会给出建筑总体能耗水平。通过评价各分项指标,给出各设备的能耗利用效率情况。复合指标法更多的是为建筑节能诊断提供依据。

本文第五章中已经建立了建筑能耗基准模型。输入建筑的运行参数和建筑基本信息,根据该模型可以计算出建筑的预测能耗即该建筑对应的能耗基准。很显然,建筑的实际能耗与能耗基准比值的高低,在一定程度上能够表征该建筑的实际能源利用的水平高低。比值越大,说明实际耗能比能耗基准要多,代表该建筑能源利用水平低;反之,比值越小,说明该建筑能源利用水平高。美国 EnergyStarBenchmarking 工具给出了建筑实际能耗与能耗基准的比值分布情况。以下也给出由前文所建立的建筑能耗基准模型计算出来的比值分布情况。

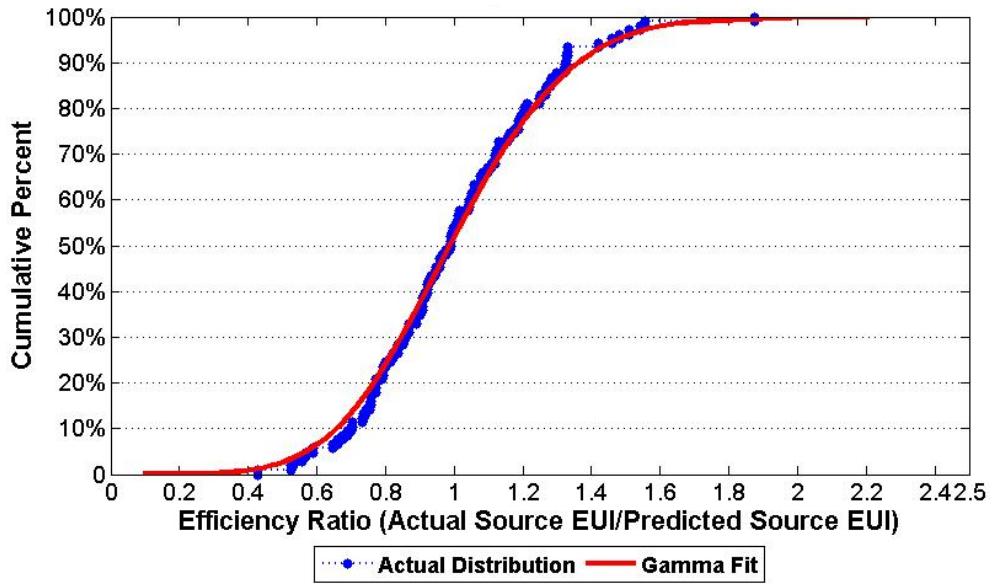


图34 能源利用率的累积分布曲线

图 34 为能源利用率的累积分布曲线。横坐标为能源利用率，等于建筑实际能耗与能耗基准的比值。纵坐标为累积百分比，表示低于某能源利用率的建筑数量，占总建筑数量的百分比。例如，能源利用率低于 0.2 的建筑累积百分比为 0%，代表所有建筑的能源利用率都不会低于 0.2。能源利用率低于 2.2 的建筑累积百分比为 100%，代表所有建筑的能源利用率都低于 2.2。

为了便于多个建筑之间进行能源利用水平高低的比较，建立 1~100 分的评分体系，每一分代表建筑总数的 1%。为了获得分值与建筑能效水平之间的关系，对能源利用率（能源利用率=建筑实际能耗/基准能耗）的累积概率分布曲线进行分析。发现，该分布曲线符合伽马概率分布的特点，因此采用伽马分布曲线进行拟合回归，如图 34 中所示的红线表示通过最小二乘法拟合回归得到的伽马分布曲线。得到伽马分布曲线的概率分布公式：

$$f(x) = \frac{(x/\beta)^{\alpha-1} \exp(-x/\beta)}{\beta \Gamma(\alpha)}, \quad x, \alpha, \beta > 0$$

其中：  $\alpha=15.194$ ;

$\beta=0.0659$ 。

因此得出的伽马曲线公式为：

$$f(x) = \frac{\left(\frac{x}{0.0659}\right)^{14.194} \exp\left(-\frac{x}{0.0659}\right)}{0.0659\Gamma(15.194)}$$

通过伽马曲线公式可以计算出每一分值所对应的建筑能源利用率的数值,即建筑实际能耗与基准能耗的比值。根据该方法建立如表所示的查询表格。通过查询表格给出的比值范围,可以很容易获得该比值对应的分值,也能说明该建筑所处的建筑能效水平。例如,伽马曲线上的1%对应的分值为99分,代表仅有1%的建筑能效水平比该建筑高。那么25%对应的分值应为75分,代表仅有25%的建筑能效水平比该建筑高。美国EnergyStar定义75分作为标准值,代表所有建筑的“前25%”的高水平,高于75分的建筑可以申请获得EnergyStar认证<sup>(62)</sup>。

通过前文所建立的建筑能耗基准模型,可得到我国酒店建筑能耗评价工具。

表37 能源利用率查询表

分值	累积百分比	能源利用率		分值	累积百分比	能源利用率	
		>=	<			>=	<
100	0%	0.0000	0.5017	50	50%	0.9794	0.9858
99	1%	0.5017	0.5467	49	51%	0.9858	0.9922
98	2%	0.5467	0.5767	48	52%	0.9922	0.9986
97	3%	0.5767	0.5999	47	53%	0.9986	1.0051
96	4%	0.5999	0.6194	46	54%	1.0051	1.0116
95	5%	0.6194	0.6363	45	55%	1.0116	1.0182
94	6%	0.6363	0.6513	44	56%	1.0182	1.0248
93	7%	0.6513	0.6650	43	57%	1.0248	1.0315
92	8%	0.6650	0.6776	42	58%	1.0315	1.0382
91	9%	0.6776	0.6894	41	59%	1.0382	1.0450
90	10%	0.6894	0.7004	40	60%	1.0450	1.0519
89	11%	0.7004	0.7109	39	61%	1.0519	1.0589
88	12%	0.7109	0.7209	38	62%	1.0589	1.0660
87	13%	0.7209	0.7304	37	63%	1.0660	1.0731

分值	累积百分比	能源利用率		分值	累积百分比	能源利用率	
		>=	<			>=	<
86	14%	0.7304	0.7396	36	64%	1.0731	1.0804
85	15%	0.7396	0.7484	35	65%	1.0804	1.0877
84	16%	0.7484	0.7570	34	66%	1.0877	1.0952
83	17%	0.7570	0.7653	33	67%	1.0952	1.1028
82	18%	0.7653	0.7733	32	68%	1.1028	1.1106
81	19%	0.7733	0.7812	31	69%	1.1106	1.1185
80	20%	0.7812	0.7889	30	70%	1.1185	1.1265
79	21%	0.7889	0.7964	29	71%	1.1265	1.1347
78	22%	0.7964	0.8038	28	72%	1.1347	1.1431
77	23%	0.8038	0.8111	27	73%	1.1431	1.1517
76	24%	0.8111	0.8182	26	74%	1.1517	1.1605
75	25%	0.8182	0.8252	25	75%	1.1605	1.1696
74	26%	0.8252	0.8321	24	76%	1.1696	1.1789
73	27%	0.8321	0.8389	23	77%	1.1789	1.1885
72	28%	0.8389	0.8457	22	78%	1.1885	1.1984
71	29%	0.8457	0.8524	21	79%	1.1984	1.2086
70	30%	0.8524	0.8590	20	80%	1.2086	1.2192
69	31%	0.8590	0.8655	19	81%	1.2192	1.2302
68	32%	0.8655	0.8720	18	82%	1.2302	1.2417
67	33%	0.8720	0.8785	17	83%	1.2417	1.2537
66	34%	0.8785	0.8849	16	84%	1.2537	1.2663
65	35%	0.8849	0.8913	15	85%	1.2663	1.2795
64	36%	0.8913	0.8976	14	86%	1.2795	1.2935
63	37%	0.8976	0.9040	13	87%	1.2935	1.3083
62	38%	0.9040	0.9103	12	88%	1.3083	1.3242

分值	累积百分比	能源利用率		分值	累积百分比	能源利用率	
		>=	<			>=	<
61	39%	0.9103	0.9165	11	89%	1.3242	1.3413
60	40%	0.9165	0.9228		90%	1.3413	1.3599
59	41%	0.9228	0.9291		91%	1.3599	1.3802
58	42%	0.9291	0.9353		92%	1.3802	1.4028
57	43%	0.9353	0.9416		93%	1.4028	1.4283
56	44%	0.9416	0.9479		94%	1.4283	1.4578
55	45%	0.9479	0.9541		95%	1.4578	1.4929
54	46%	0.9541	0.9604		96%	1.4929	1.5368
53	47%	0.9604	0.9667		97%	1.5368	1.5965
52	48%	0.9667	0.9730		98%	1.5965	1.6935
51	49%	0.9730	0.9794		99%	1.6935	>1.6935

## 6.2 应用举例

使用本方法对建筑能耗水平进行评分，应遵循以下五个步骤：

### (1) 用户输入建筑基本信息

以 THE 酒店为例。该建筑的基本信息如下：

#### ■ 能耗数据：

- 全年电力消耗：4849357kWh
- 全年天然气消耗：116905m<sup>3</sup>

#### ■ 运行数据：

- 建筑面积：58000m<sup>2</sup>
- 人员总数：409 人
- 客房总数：264 间
- 冷库总数：6 个
- 是否准备商务餐：不提供
- HDD18：1982

---

---

## ○ CDD26: 306

### (2) 计算标准化能耗

$$\begin{aligned}\text{建筑总能耗 (标煤)} &= \text{电力消耗 (标煤)} + \text{天然气消耗 (标煤)} \\ &= 1568384 + 141957 \\ &= 1710342 (\text{kg 标煤})\end{aligned}$$

### (3) 计算基准能耗

- THE 酒店属于五星级酒店，代入对应的基准模型，计算基准能耗为：  
2319787kg 标煤。

### (4) 计算能源利用率

- 能源利用率 =  $1710342 / 2319787 = 0.7373$

### (5) 查询分数

- 查询表格发现，0.7373 介于 0.7304 和 0.7396 之间，因此，该建筑的得分为 86 分。

## 6.3 建筑能耗评价工具的应用

该评价体系建立在公平的基础之上。根据用户输入的建筑信息，将影响建筑能耗的主要因素消除，得到标准化后的能耗，保证建筑之间进行能耗比较的公平性。通过统一的基准模型，为每一栋建筑计算各自的基准能耗。模型同时通过评分系统，将评价结果以 0 分~100 分的形式表示，便于理解和比对。得出的结论即能让人信服又能使高能耗建筑的业主意识到楼宇的高能耗来源于自己在管理水平或技术水平上的差距。

本文在节能监管方面有重要的作用。由于本评价系统是以得分制来评价建筑用能的，并且对分数进行了相应的分级，因此就给节能的监管工作带来的方便，同时也可成为节能监管的一种重要手段。将各个级别分别作为能源基准，这样可以通过评分和分级容易地了解每幢建筑的能耗现状以及该建筑的能源利用是否合理等多方面的信息，对能源的监管和节能的潜力的挖掘都有重要的作用。

本文的评价体系还对能耗定额的制定有重大的意义。由于影响建筑能耗的因素较多，过去在建立建筑能耗定额时，需要考虑建筑能耗可比性的问题。需要

考虑不同类型建筑、不同运营特点建筑，如何放在同一基础上进行建筑能耗的比较。本文的建筑能耗基准模型为能耗定额工作提供了强大的数据支持，使不同形式建筑之间进行能耗比较变成可能。

在制定节能激励政策方面，本文的评价系统同样有着重要的意义。节能激励政策存在公平性的问题。激励措施的实施与节能措施的效果有直接关系。那么节能效果评定的公平性，将直接影响激励措施实施的公平性，也关系到该措施能否推广，能否得到切实可靠的效果。因此，在制定了能耗定额之后，为了鼓励业主和用户节能，以及加强节能意识，同时做出相应的节能改造，以简单易行的分数形式来反映节能效果是较好的一种方法。通过变量输入后得到的能耗与实际能耗的比较最终得到一个分数，这个分数能反映出建筑用能的现状以及节能的潜力。为了激励建筑节能，业主可以通过各种激励政策来鼓励节能，其评判标准就是分数的提升，分数较以前提高了就说明建筑较以前节能了，节能的程度反映在分数的提高幅度，这样就能使业主的奖励制度有章可循，建筑能耗基准模型也保证了节能效果评价的公平性。对于节能监管和政府部门来说，节能量是他们最关心的方面，有了这个评价体系之后，节能的多少直接就能从分数的多少来判别，这样就能制定详细的奖惩措施和更加细致的管理办法来督促和监督建筑节能。对于节能工作做得好的单位可以根据建筑得分来给予相应的奖励，而对于节能工作做得不利的企业或业主可以给予相应的敦促和处罚。

---

---

## 第七章结论与展望

### 7.1 结论

本文以全国 470 栋酒店建筑 2009 年全年能耗数据和运行参数为基础数据，对酒店建筑能耗基准模型进行了研究。采用多元拟合回归分析方法，分别建立了五星级、四星级和三星级酒店建筑能耗基准模型。主要结论如下：

1、本文分析了国内建筑能耗基准研究现状，系统全面地分析了国外主流的几种建筑能耗基准模型的原理和特点，结合我国公共建筑能耗分布的特点，提出了建筑能耗基准模型宜使用数据统计的方法进行建立的思想，为建筑能耗基准模型的建立明确了方向。

2、根据建筑能耗基准模型的建立要求，对多种统计工具进行了比对研究，提出了适合建筑能耗基准模型特点的多元拟合回归分析法；采用 F 值检验的一元非线性回归分析法，对国外常用的多元线性拟合回归模型进行了改进，大幅度提高了模型整体的准确性并降低了部分自变量的误差，为建立能耗基准模型提供了方法及工具。

3、使用 SPSS 软件，采用多元线性拟合回归分析法，对 470 栋酒店的建筑能耗数据进行分析，并通过样本筛选、自变量筛选、中心化等关键技术，建立了酒店建筑的能耗基准模型，为其他类型建筑能耗基准模型的建立奠定了基础。

4、根据建筑能效比率的累积概率曲线的分布特点，建立建筑能耗评价工具。实现了酒店类建筑能耗基准模型在实际中的应用。

5、建筑能耗基准模型使不同建筑之间的能耗具有可比性，建立在实测数据基础上的评价模型，在保证建筑能耗评价公平性的同时，便于理解、易于推广，对节能监管、能耗定额及制定节能激励政策等工作的开展，提供了量化工具，具有重要指导意义。

### 7.2 创新点

1、本文首次使用较大规模建筑能耗实测数据，建立了我国酒店类建筑的能耗基准模型和评价方法。

2、首次提出了样本筛选、自变量筛选的标准、数据中心化处理及模型修正

---

---

方法。建筑能耗基准模型采用了耗能强度作为因变量，可以更加客观反映各自变量的影响程度。

3、在国内首次使用 F 值检验的一元非线性拟合回归方法，对多元线性拟合回归模型进行优化，优化后模型的  $R^2$  明显优于美国 EnergyStar 的模型（详见表 34）。

### 7.3 展望

1、按照美国 EnergyStar 的模式，建立建筑能耗基准模型，证明在我国是可行的。通过对在模型建立过程中遇到问题的分析，发现我国在建筑能耗基准建立工作上与国外的差距。最大的差距是没有完善的建筑能耗及相关运行参数等基础数据的采集制度。所有人都能认识到数据是开展任何工作的基础。但目前国内建筑能耗数据的采集工作还处于相对混乱的状态。采集方法不统一，时间间隔不同，采集的数据信息不统一，数据共享难度大。于是国内建筑能耗数据呈现出种种乱象。建议未来采取统一、科学、完善的建筑能耗数据采集制度和方法，有利于科研，有利于设计优化，有利于运行管理，更有利于全社会建筑节能工作的开展。

2、本文基本建立了酒店建筑的能耗基准模型和评价方法。对于其他类型建筑，虽然建立基准模型和评价方法的技术思路差不多，但自变量和数据的选择必然存在较大差异。未来应逐步建立各类型建筑的能耗基准模型。

3、由于建筑能耗基准模型采用建筑各类能源的总和，那么如何将各类能源进行统一折算，是一个涉及到模型精确度的重要问题。普通化石能源可以简单的按照当量热值进行折算，但如何将城市热力进行折算，是个需要未来考虑的问题。

4、建筑能耗基准模型和评价方法的建立属于建筑节能的基础性工作，不能带来直接的节能收益，但对建筑节能工作的意义重大。有了这个可靠公平的基准，才有各建筑之间、各地区之间、各国之间建筑能耗利用水平的比较，才有可能评价建筑节能的潜力空间和开展工作的方向。而且建筑基准模型也可以用于对设计方案的评价，对节能改造后的效果进行评价。有了这个基准工具，为节能工作指明大方向，希望能够得到“抛砖引玉”的效果。

---

---

## 参考文献

- [1]. CCTV 国际频道. 纪录片《与气候一起变化》. 北京, 2009 年.
- [2]. USGBC. LEED 介绍. USGBC 官方网站. [ 联机 ]  
<http://www.usgbc.org/DisplayPage.aspx?CMSPageID=2492>.
- [3]. 百度. 百度百科-LEED. [联机] <http://baike.baidu.com/view/482625.htm>.
- [4]. 中国建筑科学研究院. 《公共建筑节能设计标准》GB 50189-2005. 北京: 中国标准出版社, 2005. GB 50189-2005.
- [5]. 张旭. 几个节能措施对夏热冬冷地区居住建筑能耗的敏感性分析. 暖通空调. 2008, 38.
- [6]. 文精卫. 公共建筑能效基准及能效评价. 煤气与热力. 2008, 11.
- [7]. 于涛, 方修睦. 多层建筑能耗分析软件的开发与应用. 暖通空调. 2003, 33.
- [8]. 李骥. 建筑能耗模拟软件的特点及应用中存在的问题. 建筑科学. 2010, 2.
- [9]. DeruM. Energy Savings Modeling and Inspection Guidelines for Commercial Building Federal Tax Deductions. 美国: National Renewable Energy Laboratory, 2007.
- [10]. USGBC. LEED for Existing Buildings: Operation & Maintenance. 美国: USGBC, 2008 年 9 月.
- [11]. 中国建筑科学研究院. 民用建筑能效测评标识技术导则(试行). 北京市: 中国城乡与住房建设部, 2008 年 6 月.
- [12]. 谢艳群. 长沙市住宅能耗调查实测及其影响因素的统计分析研究. 湖南大学硕士论文. 2007 年.
- [13]. 严智勇. 福州地区大型办公建筑能耗的多元线性回归分析. 能源与环境. 2009, 1.
- [14]. 郑晓卫. 国内外建筑能耗基准评价工具的研究与应用. 上海节能. 2006, 6.
- [15]. 韩连华. 基于回归分析和数据挖掘的建筑能耗基准评价模型研究. 北京工业大学硕士论文. 2009 年.
- [16]. 明雷. 基于标杆建筑的公共建筑能耗技术定额编制方法. 第七届绿色建筑

---

---

大会会议论文. 2011 年.

- [17]. 周智勇. 基于统计数据编制的公共建筑能耗定额. 煤气与热力. 2009, 29.
- [18]. 徐强. 公共建筑用能定额研究现状及思考 . 建设科技. 2010.
- [19]. 庄智. 上海市大型公共建筑用能定额研究. 第七届中国绿色建筑大会. 2011 年.
- [20]. 刘刚. 深圳市公共建筑能耗定额标准编制思路与编制要点. 第七届中国绿色建筑大会. 2011 年.
- [21]. 劳伦斯伯克利国家实验室. [联机] [http://poet.lbl.gov/cal-arch/lin\\_ks.html](http://poet.lbl.gov/cal-arch/lin_ks.html).
- [22]. 澳大利亚新南威尔士州政府. NABERS (the National Australian Built Environment Rating System). [联机] 2010 年年. <http://www.nabers.com.au/>.
- [23]. 美国能源之星. [联机] 美国环保署. <http://www.energystar.gov>.
- [24]. OMB. Commercial Buildings Energy Consumption Survey for 2003. 美国 , 2003.
- [25]. EnergySmartTool. [联机] 新加坡. <http://www.esu.com.sg/>.
- [26]. 香港机电工程署. [联机] 香港. <http://www.emsd.gov.hk>.
- [27]. 美国太平洋煤电公司. [联机] <http://www.pge.com/chinese/myhome/>.
- [28]. Labs21 工具 . 劳伦斯伯克利国家实验室 . [ 联机 ] <http://www.labs21century.gov/toolkit/benchmarking.htm>.
- [29]. 美国加州 CalArch 工具. [联机] <http://poet.lbl.gov/cal-arch/links.html>.
- [30]. BRECSU. Energy consumption in hospitals. 英国 : CROWN COPYRIGHT FIRST, 1996.
- [31]. BRECSU. Energy performance in the government's civil estate. 英国 : CROWN COPYRIGHT FIRST, 2000.
- [32]. LohseRüdiger. 非居住建筑的节能: 德国建筑节能法律框架和单位电耗的统计方法.
- [33]. 王甫友, 田稳苓. 天津市河北区建筑能耗统计分析. 建筑节能. 2010,38.
- [34]. 李志生,刘旭红,朱雪梅,王晓霞,李冬梅. 广州市 20 栋大型公共建筑能耗特征分析. 建筑科学. 2009,25.

- 
- 
- [35]. 段小平,郎莺. 杭州市公共建筑能耗现状及节能分析. 建筑科学. 2010, 36.
  - [36]. 张吉礼,陈永攀, 赵天怡, 卢振, 李琳琳,马良栋,. 大连市首批典型公共建筑能耗统计分析. 建筑热能通风空调. 2009, 28.
  - [37]. 刘涛,刘大龙,杨柳. 西安市大型办公建筑能耗调研分析. 制冷与空调. 2010, 24.
  - [38]. 朱炜,王竹,陈海啸,龚敏. 杭州政府机关办公建筑能耗分析. 建筑学报. 2009.
  - [39]. 缪晨光,尹峰,胡丽君,魏小清,杨勇. 长沙市政府办公建筑能耗调研及节能途径. 煤气与热力. 2009, 29.
  - [40]. 苏华,郑澍奎,王守亮,赵天蓉,向勇. 成都市民用建筑能耗调查分析. 四川建筑科学. 2010, 36.
  - [41]. 陈诤岚,常艳新. 浙江省医院建筑能耗调查及其影响因素研究. 建筑科学. 2010, 26.
  - [42]. 张炜伟. 福州市区机关办公建筑能耗调查分析. 能源与环境. 2009.
  - [43]. 毕业亮,邱烷南,张华玲. 深圳地区大型酒店能耗现状及节能分析. 制冷与空调. 2009, 23.
  - [44]. 王树健. 重庆市既有大中型商场建筑特征及能耗调研分析. 建筑节能. 2008, 36.
  - [45]. 潘毅群,殷荣欣,楼振飞. 上海 10 幢大型公共建筑节能状况调研. 暖通空调. 2010, 40.
  - [46]. 谭颖. 重庆市大型商场类公共建筑能源使用与管理的现状与问题分析. 建筑节能. 2009,1.
  - [47]. 龙惟定. 上海公共建筑能耗现状及节能潜力分析. 暖通空调. 1998, 28.
  - [48]. 曹印革. 统计学原理与实务. 北京：电子工业出版社, 2009 年. ISBN 978-7-121-09300-5.
  - [49]. yanfengshen. 品味统计. 百度博客. [联机] 2009 年 8 月 21 日.  
<http://hi.baidu.com/yanfengshen/blog/item/f4f051fc390b171f09244dc0.html>.
  - [50]. 中华人民共和国国家质量监督检验检疫总局. 空气调节系统经济运行. 北京：中国标准出版社, 2007. GB/T 17981-2007.

- 
- 
- [51]. 维基百科. 词条“多变量分析”的解释. 维基百科. [联机] 2010年5月11日. <http://zh.wikipedia.org/wiki/%E5%A4%9A%E8%AE%8A%E9%87%8F%E5%88%86%E6%9E%90>.
- [52]. 王保进. 多变量分析-统计软件与数据分析. 北京: 北京大学出版社, 2007年8月第1版. ISBN 978-7-301-12473-4/C.0446.
- [53]. James D. LeeperPh.D. Choosing the Correct Statistic. 无出版地: College of Community Health Sciences University of Alabama, 2007年.
- [54]. 高桥信. 漫画统计学. 北京: 科学出版社, 2009年8月第1版. ISBN 978-7-03-024963-0.
- [55]. 百度. 百度百科. [联机] <http://baike.baidu.com/view/130328.htm#5>.
- [56]. 现代统计分析方法与应用.
- [57]. 何晓群. 应用回归分析. 北京市: 中国人民大学出版社, 2007. ISBN: 9787300082356.
- [58]. EPA.US. ENERGY STAR® Performance Ratings Technical Methodology for Hotel Portfolio Manager. [ 联机 ] [http://www.energystar.gov/ia/business/evaluate\\_performance/hotel\\_tech\\_desc.pdf](http://www.energystar.gov/ia/business/evaluate_performance/hotel_tech_desc.pdf).
- [59]. SPSS Inc. Help Documents of SPSS 15.0. 美国: SPSS Inc., 2007年.
- [60]. 项静恬. 多元回归模型在实际应用中的几种推广. 数理统计与管理. 第3期, 1994, 13.
- [61]. 周权. 公共建筑空调系统节能评价指标及体系. 哈尔滨工业大学硕士学位论文. 2007年.
- [62]. 美国 EPA. 美国 EnergyStar 建筑认证方法. [ 联机 ] [http://www.energystar.gov/index.cfm?c=business.bus\\_bldgs](http://www.energystar.gov/index.cfm?c=business.bus_bldgs).
- [63]. 清华大学建筑节能研究中心. 中国建筑节能2008年度发展研究报告. 北京: 中国建筑工业出版社, 2008.
- [64]. 国家质量监督检验检疫总局. GB/T 17981-2007 空气调节系统经济运行. 北京: 中国标准出版社, 2007.