·现代经济学与管理学文库·

LIBRARY OF MODERN ECONOMICS AND MANAGEMENT SCIENCE

方法与工具系列丛书

METHODS AND TOOLS SERIES

经济管理数据、模型与计算

——方法、实现及案例

DATA, MODEL AND COMPUTATION IN ECONOMICS AND MANAGEMENT

王文平 杨洲木 来向红／编著

東南大學出版社
SOUTHEAST UNIVERSITY PRESS

现代经济学与管理学文库·方法与工具系列丛书

经济管理数据、模型与计算

——方法、实现及案例

东南大学出版社
·南京·

内 容 提 要

本书以经济管理中的实际问题为背景，介绍经济管理领域数据处理的常用模型、方法及计算机实现。与现有的大多数统计学和数据处理教材不同，本书在注重数学逻辑和严密性的同时，强调数据处理模型的思想和原理的分析，而略去了繁琐复杂的数学推导过程，并特别重视各模型和方法的实际应用。每个章节都配有详细的经济管理案例的分析、建模、求解和编程实现。

针对经济管理领域不同层次的数据处理需求，本书根据常用数据处理方法的特征，将内容分为三个部分：经济管理中数据特征和数据处理的任务、数据处理的经典统计分析方法、经济管理中复杂数据处理的智能化方法。其中，数据处理的经典统计分析方法包括：线性回归、时间序列、数据异常点与结构变化的判断、复杂数据处理的结构分解方法、主成分分析、聚类分析、组合预测、数据包络分析以及用于实证研究的结构方程模型等；经济管理中复杂数据处理的智能化方法包括：生物智能化的不确定信息处理方法、生物体智能产生的结构仿生方法——人工神经网络以及生物智能化的过程仿生算法——模拟进化计算。

本书可以作为高等院校经管类、理工类相关专业高年级本科生以及研究生的教科书，也可以作为从事经济管理工作的专业人员的参考书。

图书在版编目(CIP)数据

经济管理数据、模型与计算——方法、实现及案例/王文平等编著. —南京：东南大学出版社，2010.6

ISBN 978-7-5641-2193-8

Ⅰ.①经… Ⅱ.①王… Ⅲ.①经济管理—高等学校—教材 Ⅳ.①F2

中国版本图书馆 CIP 数据核字(2010)第 070330 号

东南大学出版社出版发行
(南京四牌楼 2 号 邮编：210096)
出版人：江建中
江苏省新华书店经销 南京新洲印刷有限公司印刷
开本：700mm×1000mm 1/16 印张：18.5 字数：405 千字
2010 年 11 月第 1 版 2010 年 11 月第 1 次印刷
ISBN 978-7-5641-2193-8
印数：1—3000 册 定价：40.00 元

(本社图书若有印装质量问题，请直接与读者服务部联系。电话(传真)：025-83792328)

前　言

本书写作目的

科学发展至今，已经产生了诸多数据处理模型和方法，从线性回归、时间序列等经典统计模型，到以人工神经网络、模拟进化计算等为代表的生物体智能化算法，数据处理的新模型和新方法还在不断涌现，这些模型和方法在自然科学、管理科学、经济和社会等学科领域得到广泛应用。

以介绍数据处理模型和方法为主的各种课程是经济管理类专业的主干课程之一，其目的是面向现实经济管理问题，培养学生基于科学理论和方法体系的分析问题、解决问题的能力；是为学生研究本专业问题、处理相关数据、获得科学深入信息提供工具。这就要求这类课程的教材应简化数学推理，在阐明模型基本原理和思想的基础上，重点讲述如何应用这些模型解决具体经济管理问题。对于计算较为复杂的模型，还应该辅以相关的计算软件或编程代码，着重讲解这类模型如何实现，才能克服长期以来我国经济管理类专业的应用统计或数据处理类教材大多以繁琐的数学证明和手工计算为主，缺乏较为详尽的经济管理案例分析，以及与模型相匹配的计算编程实现的支持所带来的偏颇。

为此，本书从内容的选取与组织上，力求遵循“在保证数学逻辑严密性的同时，强化培养分析问题、解决问题的应用能力”的原则，让读者在掌握经济管理领域常用的数据处理模型的基本思想和原理的同时，实现规范、科学地解决经济管理问题的逻辑思维能力和应用能力的协同提高。

本书的特点

本书是作者在多年数据处理课程教学经验基础上编写而成，具有以下特点：在内容安排上，提取经济管理中常用数据处理模型和方法的共同特征，将其分类整合，形成了包括：经济管理中数据处理的综合方法（包括：输入—输出综合、指标综合等）、经济管理中数据处理的智能化方法（包括：人工神经网络、模拟进化计算等）等的模块化结构，更加有利于读者学习和掌握；在具体模型的讲解上，着重模型、方法的原理和思想阐释，简化繁琐的数学证明和推导，强化问题建模和计算机实现；在整体结构上，强调问题导向，每个章节都配有典型的经济管理案例分析，详尽讲解如何针对具体问题进行分析、建模、求解及采用 MATLAB 或 SPSS

编程实现。

本书的适用范围

本书可以作为高等院校经管类、理工类相关专业高年级本科生、研究生的教材。对于从事经济管理工作或其他人文社会科学领域的实际工作者和研究人员，本书也是一本很好的参考书。

本书各章节的内容简介

本书分三个部分，共计 10 章。

第一部分介绍了经济管理中数据特征、数据处理的任务和步骤。着重归纳了数据处理的四个任务，即预测、控制、统计诊断和系统评价，以及现代数据处理的六个步骤。

第 2 章至第 7 章构成第二部分，主要讲述数据处理的经典统计分析方法，主要包括：线性回归、时间序列、数据异常点与结构变化的判断、复杂数据处理的结构分解方法、复杂数据建模的综合方法（包括：主成分分析、聚类分析、组合预测、数据包络分析），以及用于实证研究的结构方程模型等。每个章节在重点讲述模型和方法的基本思想和原理的基础上，详细给出了模型和方法在经济管理领域中的实际应用案例及编程实现。

第三部分为经济管理中复杂数据处理的智能化方法，包括第 8 至第 10 章，包括生物智能化的不确定信息处理方法，具体又包括模糊聚类分析、模糊模式识别和模糊综合评判等方法；另外还包括生物体智能产生的结构仿生方法——人工神经网络，以及生物智能化的过程仿生算法——模拟进化计算。由于这部分的建模和计算相对复杂，书中给出了应用实例和详尽的 MATLAB 程序供参考。

本书是在教学过程中不断总结和修改而成的，先后有作者的多名研究生参加了资料收集、文字录入及修改校对等工作，他们是：王卫东、花磊、章慧、林秋月、何署子等，最后由王文平、杨洲木、来向红修改定稿。全文由王文平统稿。在此，向所有为本书付印做出贡献的朋友和同仁表示挚诚的感谢。由于作者水平有限，书中难免出现错误，敬请读者批评指正。

王文平

2010 年 11 月

目　录

第一部分 绪论

第1章 经济管理中数据特征和数据处理的任务

数据是企业、政府等机构的一种重要资源，是进行科学决策的基础。经营管理者管理过程中几乎所有的行为都需要数据，面对大量数据资源，如何进行科学地处理，使之服务于管理决策，无论是现实中还是理论上都具有重要意义。为此，人们做了大量努力，有关的数据处理方法众多，但这些方法在实际应用中，特别是在我国经济管理者中并没有得到广泛应用，人们在决策时凭主观感觉、“拍脑袋”的现象仍普遍存在。随着管理科学化的深入，这些数据处理方法的应用和普及将不再是遥远的事情。本书写作的目的，即是为当代经济管理者提供各种数据处理的方法，服务于经济管理。本书特点是在保证数学严密性的前提下，避免艰涩的数学论证，以应用为目的，主要向读者介绍实际中常用的数据处理方法。

1.1 数据特征与分类

经济管理中人们所面临的系统行为数据由于受随机因素的影响，大多具有随机性特征。

例如：“记录某公共汽车站某日上午某时刻的等车人数”、“从一批产品中，依次任选三件，记录出现正品与次品的件数”、“考察某地区10月份的平均气温”、“从一批灯泡中任取一只，测试其寿命”等。我们根据这些具有随机性的数据，可以从中总结出其统计规律性。

与随机性数据相对的是确定性数据。所谓确定性数据，一般是指能够据此找出确定性因果关系的数据，这种确定性因果关系，也就是一一对应关系。例如：圆的半径与圆的面积(R,S)，若忽略测量系统误差和随机误差，则可视为确定性数据，因为R与S的关系是一种确定性关系。

1965年，美国自动化控制论教授 L. A. Zadeh 的开创性论文“Fuzzy Sets, Information and Control”，不仅拓宽了经典数学的理论基础，也使人们对现实中的数据特征和处理方法有了更加合理的认识，可谓是关于模糊数学与模糊性数据处理的开山之作。

在确定性、随机性和模糊性数据之外，还存在一种情况，即样本数据远远达不到显示系统统计规律或确定性规律的数量，即数据信息不完备，这样的数据即可称之为具有灰色性。严格地说，现实中经营管理者所面对的数据都不同程度地具有灰色性。1982年华中理工大学（现为华中科技大学）自控系邓聚龙教授开创的灰色系统理论，提出了处理灰色性数据的理论和方法体系。

1.2 经济管理中的数据处理方法

现有的数据处理方法繁多，这些方法从数据特征来看，对应着不同特征的数据，有着相应的处理方法。当人们关注所掌握数据的随机性时，有统计处理方法。现有统计处理方法可分为两大类，即静态统计分析和动态统计分析。所谓静态统计分析有两层含义：一是指对取自同一总体的具有随机性与相互独立性的样本的统计分析，这些数据取自同一总体，次序可以任意排列，数据之间相互独立，毫无关联；另一含义是表示对某些静止状态下有关数据的分析方法，如主成分分析法、聚类分析等。动态统计分析则是针对动态数据的统计分析方法，所谓动态数据，其每一数据都取自不同的总体，但这些不同时刻的总体存在着某种统计相依关系，使得序列中数据间不仅相互关联，而且次序也不可随意变化，有着其特定含义。目前最主要的动态统计分析方法即时间序列分析，在许多领域得到广泛应用。对确定性数据一般有机理分析、曲线拟合等处理方法。对模糊性数据，有模糊统计分析方法、模糊聚类分析等处理方法。对灰色数据，有灰统计分析、灰关联度分析等处理方法。

上述这些数据处理方法，按方法分类，可分为结构分解方法，该类方法主要将非平稳时间序列分解为确定性部分和随机性部分之和进行分析；综合方法，包括指标综合如聚类分析、主成分分析、关联度分析等；输入—输出综合，如 DEA 方法、神经网络方法等。

在动态数据统计分析中，针对不同模式的数据，也有着各种不同的处理方法。对平稳数据序列、非平稳时间序列及季节性或非季节性时间序列，分别有不同处理方法。某些非平稳时间序列，还可经过差分分离趋势等方法进行平稳化处理。

1.3 经济管理中数据处理的任务

数据处理在经济管理中的任务大致可分为如下几个方面：

1.3.1 建模

所谓数据建模，即根据一定数量的客观数据，参照系统的特征，概括或近似地表达系统的数学结构。按时间对模型的影响，可分为时变与时不变模型、静态与动态模型；按变量情况可分为离散模型与连续模型、确定性模型和随机性模型等；按研究对象所处的实际领域可分为经济模型、社会模型、交通模型等。

模型不仅可帮助决策者分解所面临的系统，而且也可为决策者进行决策、预测、评价提供模型基础。

1.3.2 预测

模型的建立是研究预测问题的关键，预测是经营管理中的重要环节，是决策的基础，数据处理的重要任务是为经济管理中的预测提供方法和手段。现有的预测方法较多，通常采用的有：回归分析法、德尔菲法、马尔柯夫预测法、模型法、指数平滑法、残差辨识法、灰色预测法等。

普遍认为经济管理系统所采用的数据都带有某种随机的性质，统计的时间分布是一个时间序列，对时间序列的分析已经形成一个学科分支，预测是其中重要的组成部分，根据时间序列在时刻 t 得到的观测值，可以预报时刻 $t+1$ 的未来值，可以为经济管理提供优化基础。在实际中与预测相联系的三个重要问题是：预测时间序列、估计适应的系统模型和设计控制系统。

1.3.3 控制

经济管理中的控制，多属于预测控制，即通过系统行为数据序列，按照已掌握的系统规律，通过预测，得到系统未来行为预测值，根据该预测值，采取控制策略，达到预期目的。控制问题是预测问题的反向问题，两者之间的关系非常密切，但是控制问题往往比预测问题更为困难。

1.3.4 统计诊断

对实际工作中逐步积累的历史数据或围绕某一特定目标收集起来的数据，经过加工处理，之后通常的做法是将其纳入一个方便有效的统计模型中进行研究，但任何统计模型都只能是客观复杂过程的一种近似描述，不可避免地包含某些假设，甚至模型本身就是一种假定。统计推断则是寻找一种诊断方法，判别实际数据是否与既定模型有较大偏离，并采取相应对策，通过统计诊断，可以找出严重偏离既定模型的数据点，即异常值；也可以区分对统计推断影响特别大的点，即强影响点；还可以找出那些远离数据主体的点，即高杠杆点。对数据进行这些初步诊断后，还

需研究解决方案，若实际数据仅有个别点与既定模型偏离较大，这时往往肯定模型，而对这些个别点再作进一步考察，若实际数据中许多点都与既定模型偏差较大，在多数情况下，仍希望保留方便有效的既定模型，为此，可对数据集进行合适的数据变换，使变换后的数据能符合既定模型，从而进行必要的统计分析。

1.3.5 系统评价

在经济管理中，对目标系统的评价是决策和管理的重要依据。其基本过程主要包括：

(1) 首先确定指标体系，收集有关定量指标相关数据。经济管理系统的评价指标体系的选择，是为实现对目标系统的综合评价提供数据支持，一般应遵循以下原则：

① 指标体系的科学性和先进性原则；

② 系统性原则，即指标体系层次结构应合理，协调统一；

③ 定性与定量相结合原则；

④ 可行性和可操作性原则，即设计指标应具有可量化、数据可采集性等特点，各项指标能有效地测度和统计。

(2) 选择评价方法

现有评价方法较多，据不完全统计，有数百种之多，常用的评价方法有：

① 专家评价法：该方法是以评价者的主观判断为基础的一种评价方法，常用“分数”或“指数”等作为评价尺度；

② 经济分析法：是以经济指数为尺度，定量表示收益等经济指数；

③ 运筹学评价法：用运筹学方法评价的基本原则是利用有关数据，建立数学模型，对系统进行定量动态评价，基本步骤为：建立模型、决定目标函数，然后求解；

④ 综合评价法：由专家评价法、经济分析法、运筹评价法进行不同方式的组合形成的评价方法。

1.4 数据处理的步骤

1.4.1 确定数据处理的目标

数据处理是制定决策的辅助工具，数据的采集以及数据处理方法的选择，均取决于决策问题的决策目标。

1.4.2 搜集和整理数据

数据是数据处理和最终决策的依据，应根据决策问题本身和拟采用的数据处理手段，采集尽可能系统全面的数据。

1.4.3 对数据进行背景分析

对数据进行背景分析包括数据的可靠性分析、数据的可用性分析及数据自身结构特征分析，具体包括：数据的来源是否可靠、数据中是否含有异常值、数据中隐含的分析对象的结构变化特征等。这些背景分析，可以帮助确定采用何种模型来分析解决问题。

1.4.4 数据的预处理

当数据中含有异常值时，则需要对数据进行稳健处理，也称为异常值剔除处理。有时，即使数据中没有异常值，若数据离散程度较大，为了减小数据处理误差，也要对数据进行预处理。

1.4.5 选择适当的数据处理方法

根据数据处理的目的和数据特征，选择适当的方法群或模型群。

1.4.6 分析结果并进行动态调整

针对各种不同方法和模型，比较其有效性，从而确定一个或多个具体模型和方法，然后根据计算结果对模型和方法进行调整和控制，并结合定性分析，最终辅助决策，使方案在实践中实施。

第二部分　数据处理的经典统计分析方法

第2章　线性回归分析

2.1　引　　言

二战初期，德国对法国发动攻势后，英国首相丘吉尔应法国的请求，动用了十几个防空中队对德作战。由于防空中队的飞机需要在欧洲大陆的机场进行维护，使得空战中英国飞机损失惨重。这时，法国总理请求英国继续增派十几个中队的飞机，丘吉尔决定同意这一要求。英国内阁知道此事后，请求统计学家利用线性回归模型对出动飞机与战损飞机的数据进行统计分析，发现如果飞机的补充率与损失率不变，飞机数量的下降是非常快的："以现在的损失率损失两周，英国在法国的飓风式战斗机就一架也不存在了。"内阁希望丘吉尔收回他的决定，最终，丘吉尔同意了内阁的要求，并在几天内撤回了在法国的飓风式战机，只留下了三个中队，为以后英国本土的保卫战保留了实力。

线性回归分析是统计学中的常用方法，是处理变量之间线性关系的重要方法之一。

2.1.1　确定性关系与非确定性关系

客观世界中变量之间的关系一般可分为两种类型。一种类型是，变量之间存在着确定性关系，或称为函数关系。例如，圆的面积 S 与半径 r 之间有关系 $S=\pi r^2$；牛顿第二定律中的力 F、质量 m 和加速度 a 之间有关系 $F=ma$；欧姆定律中的电压 U、电流 I 和电阻 R 之间有关系 $U=IR$ 等。这类关系的本质特点是：一个变

量的取值随着其他变量取值的确定而确定，即因变量的取值随着自变量取值的确定而确定。另一种类型是，变量之间存在非确定性关系。例如，根据遗传规律，父辈的身高与子辈的身高间有一定的关系，一般而言，父辈的身高越高，其子辈的身高也越高，然而，当父辈的身高已知时，却未必能精确地确定出其子辈的身高；人的血压与年龄之间也存在着关系，但同年龄的人的血压往往并不相同；气象领域中的温度与湿度之间的关系也是这样。这是因为我们涉及的变量（如身高、血压、湿度）是随机变量，上面所说的变量之间的关系是非确定性的，称为相关关系。回归分析是研究相关关系的一种数学工具，它能帮助我们分析一个被解释变量（或称因变量）与一个或多个解释变量（或称自变量）之间的统计关系。

2.1.2 回归的含义

回归（regression）一词是英国著名人类学家和气象学家 Francis Galton（1822—1911）于 1885 年引入的。在“身高遗传中的平庸回归”的论文中，Galton 阐述了他的重要发现：虽然高个子的父辈会有高个子子辈，但子辈的身高趋向于比他们的父辈更加平均，就是说如果父辈身材高大，则子辈的身高要比父辈矮小一些；如果父辈身材矮小，则子辈的身材要比父辈高大一些。换言之，子辈的身高有向平均值靠拢的趋向。因此，他用回归一词来描述子辈身高与父辈身高的这种关系。而后，他的朋友，英国著名统计学家 K. Pearson 等人收集了上千个家庭成员的身高数据，分析出儿子的身高 y 与父亲的身高 x 大致可归结为以下关系：$y=0.516x+33.73$（单位为英寸），从而进一步证实了 Galton 的“回归定律”。这就是回归一词最初在遗传学上的含义。回归的现代意义远比其原始意义要广泛得多，具体地说，回归的内容包括如何确定因变量与自变量之间的回归模型；如何根据样本观测数据估计并检验回归模型及其未知参数；从众多的自变量中，判断哪些变量对因变量的影响是显著的，哪些变量的影响是不显著的；根据自变量的已知值或给定值来估计和预测因变量的平均值并给出预测精度。而回归分析就是研究随机因变量与可控自变量之间相关关系的统计方法，分为线性回归分析与非线性回归分析，本章着重介绍线性回归分析，它是两类回归分析中较简单的一类，也是应用较多的一类。

2.2 一元线性回归

2.2.1 一元线性回归模型

一元线性回归描述两个变量间的线性相关关系，先看一个例子。

例 2.1 某快餐连锁店分布在全国多个城市，连锁店最佳位置在大学附近。

管理人员确信这些连锁店的季销售收入与学生人数是正相关的。统计数据如下：

表 2.1　10 家快餐连锁店的学生人数和季度销售收入数据

连锁店 i	学生人数(x_i)(单位：千人)	季销售收入(y_i)(单位：千元)
1	2	58
2	6	105
3	8	88
4	8	118
5	12	117
6	16	137

由上述数据可见，学生人数 x 与季销售收入 y 之间近似存在线性关系，可表示为：

$$y = a + bx + \varepsilon \tag{2.1}$$

其中 a、b 是未知常数，称为回归系数，ε 为随机干扰或随机误差，表示其他随机因素对季销售收入的影响，称式(2.1)为一元线性回归模型。一般的，一元线性回归模型有以下假定：

① 两变量 x 与 y 之间的关系为线性关系；

② x 为非随机变量，它的值是可观测的；

③ 随机误差项的数学期望为 0，即 $E(\varepsilon)=0$；

④ 对于所有的观测值，误差项具有相同的方差，即 $\mathrm{Var}(\varepsilon)=\sigma^2$；

⑤ 各随机误差 ε_i 之间相互独立，即对任意的 $i\neq j$，$E(\varepsilon_i\varepsilon_j)=0$；

⑥ 随机误差项服从正态分布，即 $\varepsilon\sim N(0,\sigma^2)$。

对模型(2.1)，主要考虑如下问题：① 用 n 对试验观测数据(x_i,y_i)，$i=1,2,\cdots,n$，对 a，b 和 σ^2 作估计；② 对回归系数 b 作假设检验；③ 对 y 作预测。

2.2.2　一元线性回归模型的参数估计

(1) a，b 的最小二乘估计

对于模型(2.1)，确定 a、b 的估计值 $\hat{a}$、$\hat{b}$ 的方法主要有最小二乘估计法和极大似然估计法，我们只介绍前者。

假设有 x、y 的 n 组独立观测数据$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$，则由(2.1)可得：

$$\begin{cases} y_i=a+bx_i+\varepsilon_i, i=1,2,\cdots,n \\ E(\varepsilon_i)=0, \mathrm{Var}(\varepsilon_i)=\sigma^2 \end{cases} \tag{2.2}$$

记

$$Q = Q(a,b) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \tag{2.3}$$

称 $Q(a,b)$ 为偏离真实直线的偏差平方和，或为离差平方和。

最小二乘法就是确定 a、b 的估计 $\hat{a}$、$\hat{b}$

使得

$$Q(\hat{a},\hat{b}) = \min_{a,b} Q(a,b) \tag{2.4}$$

为此将(2.3)分别对 a、b 求偏导数得：

$$\begin{cases} \dfrac{\partial Q}{\partial a} = -2\sum\limits_{i=1}^{n} (y_i - a - bx_i) \\ \dfrac{\partial Q}{\partial b} = -2\sum\limits_{i=1}^{n} x_i (y_i - a - bx_i) \end{cases} \tag{2.5}$$

令 $\begin{cases} \dfrac{\partial Q}{\partial a} = -2\sum\limits_{i=1}^{n} (y_i - a - bx_i) = 0 \\ \dfrac{\partial Q}{\partial b} = -2\sum\limits_{i=1}^{n} x_i (y_i - a - bx_i) = 0 \end{cases}$，并用 $\hat{a}$、$\hat{b}$ 代替 a、b 得

$$\begin{cases} \sum\limits_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i) = 0 \\ \sum\limits_{i=1}^{n} x_i (y_i - \hat{a} - \hat{b}x_i) = 0 \end{cases} \tag{2.6}$$

于是有

$$\begin{cases} n\hat{a} + \hat{b}\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=0}^{n} y_i \\ \hat{a}\sum\limits_{i=1}^{n} x_i + \hat{b}\sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i y_i \end{cases} \tag{2.7}$$

(2.7)称为正规方程组。

由正规方程组解得

$$\begin{cases} \hat{b} = \dfrac{\sum\limits_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} \\ \hat{a} = \bar{y} - \hat{b}\bar{x} \end{cases} \tag{2.8}$$

其中 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$，$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$，分别称为 x 和 y 的样本均值。

为了便于计算，引入下列记号：

$$L_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2,\ L_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}),\ L_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2$$

则(2.8)可简记为

$$\begin{cases}\hat{b}=\dfrac{L_{xy}}{L_{xx}}\\ \hat{a}=\bar{y}-\hat{b}\bar{x}\end{cases}\tag{2.9}$$

注： ① 对模型(2.1)，若用极大似然估计法求 a、b 的估计值，则结果与最小二乘估计值相同(证明留给读者作为练习)；

② 在求得 a、b 的最小二乘估计值 $\hat{a}$、$\hat{b}$ 后，称 $\hat{y}=\hat{a}+\hat{b}x$ 为 y 关于 x 的经验回归(直线)方程，称 $\hat{a}$、$\hat{b}$ 为经验回归系数，将 $\hat{a}=\bar{y}-\hat{b}\bar{x}$ 代入 $\hat{y}=\hat{a}+\hat{b}x$，经验回归(直线)方程可改写为

$$\hat{y}=\bar{y}+\hat{b}(x-\bar{x})。\tag{2.10}$$

例 2.1 中，代入数据可得：$\hat{y}=60+5x$。现在可用回归模型来进行预测了，如果有一家连锁店位于有 16 000 名学生的校园附近，则可根据所得回归方程预测其季销售收入为

$$\hat{y}=\hat{a}+\hat{b}x=60+5\times16=140(千元)$$

例 2.2 某饮料公司发现，饮料的销售量与气温之间存在相关关系，即气温越高，人们对饮料的需求量越大。表 2.2 列出了该饮料公司通过实际记录所得到的饮料销售量和气温的观察值，求销售量关于气温的线性回归方程，并给出当气温为 35℃时销售量的预测值。

表 2.2 某饮料公司饮料销售量和气温的观察值

时期	销售量(箱)	气温(℃)	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$\hat{y}$	$(y-\hat{y})^2$
1	430	30	150	9	2 500	409	441
2	335	21	270	36	2 025	322	169
3	520	35	1 120	64	19 600	458	3 844
4	490	42	1 650	225	12 100	526	1 296