计算机在压力加工中的应用

徐则礼 编著

中南工业大学教材科

一九九三年九月

前　言

随着计算机科学技术的不断发展，中、小型尤其是微型计算机的普及应用，几乎一切现代工程技术活动都程度不同地需要计算机的帮助，以期快速、准确地完成生产、科研、管理和设计、制造方面的任务，能否多方面有效地应用计算机，不仅是衡量一个现代工程技术人员的基本技能的重要标准，而且也在一定程度上反映出一个国家的科技发达水平。

为适应金属压力加工专业教学大纲的需要，进一步提高高年级学生的计算机应用水平，特编写了这本教材。

本教材共分三章。第一章重点介绍数学模型的建立方法和数学模型与自动控制的关系，以及用于轧制过程的典型工艺模型，模型参数的自适应校正。第二章介绍最优化技术的基本概念和常用的最优化问题算法。第三章，主要介绍计算机辅助设计与辅助制造的基本概念，必要的硬件和软件知识，交互式图形学，以及与系统开发有关的某些问题。为便于理解各部分内容，理论联系实际，还附有一定的实例和源程序清单。

在编写过程中，张晓东等同志协助编写了部分源程序的说明，并将程序调试通过，对此谨表谢意，限于编者水平和时间仓促，缺点与错误在所难免，诚恳希望批评指正。

编　者

目　录

绪 论

计算机是电子计算机的简称，从它问世以来已经历了近50年的蓬勃发展期，而且仍处在"方兴未艾"之中。计算机科学技术在目前如此兴旺发达，就是由于计算机不断更新换代，且愈来愈多地具有其他"计算机器"没有或不可能有的特点，因而具有广泛的用途所致。在国民经济各部门，以及国防、文教、卫生和科研领域，乃至人们的家庭与日常生活中，正在大量地以各种不同的类型和方式使用计算机，例如数值计算、数据处理、信息加工、通讯网络、自动控制和电子游戏等等，而与压力加工有关的，归纳起来有以下几个主要方面：

一、数值计算（数值分析）

在这方面，除一般的解题外，典型的就是用有限元或边界元法，分析金属材料或加工工具在弹性或塑性变形时的应力、变形、温度、疲劳和断裂等问题，以及用上限法或下限法计算加工载荷等等。

二、生产过程自动化

自从把计算机作为工业生产用的控制机使用以后，使生产过程自动化提高到了一个崭新的水平。在压力加工方面，由于使用计算机（主要是小型机和微型机），已经在锻压、轧制、挤压和拉拔生产中，实现了单机自动化或生产过程的全面计算机控制，例如加热炉、热处理炉的计算机控制，数控线切割机与电火花加工、轧制过程的全面计算机控制与分级控制等。

三、数字仿真

仿真是一种用模型（物理模型或数学模型）来研究真实系统或过程的技术。数字仿真则是用计算机通过数学模型来研究系统或过程的技术。数字仿真可以完成需要花费巨大投资才能完成的、或是无法完成的对真实系统或过程的实验。

在压力加工领域，对连轧过程的数字仿真技术已经逐渐完善、成熟，通过数字仿真，可以获得最佳的轧制工艺方案。在挤压和拉拔方面，苏联应用数字仿真技术，研究了挤压、拉伸变形区内的温度、速度和应力场，并得到金属连续性遭破坏的极限条件，因而也可以从中找到最好的加工工艺参数组合。此外，还有热轧工作辊温度场的仿真、塑性变形低温不稳定性的仿真，控制轧制的仿真与复合材料断裂过程的仿真等等。

四、计算机辅助设计与辅助制造

这是利用计算机协助设计者设计产品（包括工具）及从事绘图工作，并对加工、制造进行控制，管理的技术。随着这项技术的迅速发展，在压力加工工艺和工具的设计/制造方面得到了广

泛应用。例如各种加工用的模具、轧管机孔型，型材轧制孔型、工艺卡片与工艺过程的计算机辅助设计/制造等等。

五、最优化技术

最优化技术是研究和解决如何在一切可行的方案中寻求最优方案的技术。它的发展与计算机的出现和普及，以及计算技术的研究成果有密切关系。目前这项技术已经广泛用于科研、生产和管理等方面。在压力加工中，可用于工艺制度的最优化。例如板带轧制时的最优压下制度和速度制度。坏料加热时的最优加热制度，模具的最优模角，拉拔时的最优配模等等。

除上述几个主要方面外，其他方面还有象合同、工资与计划等的管理，各种资料的归档与检索，文件资料的编辑，排印等等。

本教材的第一章首先介绍数学模型及其应用，重点介绍数学模型的建立方法。然后介绍数学模型在自动控制中的作用，以便一方面了解数学模型的重要性和建立方法，一方面了解一些自动控制方面的知识。第二章介绍最优化技术的基本概念和实施的方法。由于这顶技术属于新兴的专门学科，内容十分丰富，教材内容只涉及其中极少一部分，以期达到入门的目的，第三章介绍计算机辅助设计与辅助制造技术，在介绍系统的硬件和软件组成基础上，重点介绍应用软件的开发，同样，只希望起到入门的作用。

教材涉及到一定的数学知识和计算机方面的知识，在学习之前应对高等数学、线性代数、概率论和计算机原理、计算方法等有一定程度的了解，而在学习中则应以弄清概念，掌握解决问题的思路与具体方法为主。

为便于学习和理解教材内容，附有一定数量的实用程序并加以说明，而且可以通过上机操作，实验这些程序。

第一章　数学模型及其应用

通常，将描述某种过程或系统的一定内在规律的数学表达式(或表格)称作数学模型。它所描述的内在规律，可能是过程或系统的静态特性(状态特征)。也可能是动态特性(运动规律)。因为数学表达式具有量的内涵，所以也可以说，数学模型是描述过程或系统中表达因果关系的量与量之间一定客观关系的数学表达式(或表格)。

根据被抽象的过程或系统的性质与范围大小，以及研究问题的具体要求，数学模型有简有繁，有时只是一个简单的代数式，有时则是一组有机结合的代数或微分方程式。由此看来，数学模型虽然是一个或一组数学表达式，但与一般的数学表达式有原则上的区别。因为前者与某种过程或系统有直接关系，而后者仅仅是一种数学式子。例如：

$$S=\frac{1}{2}gt^2$$

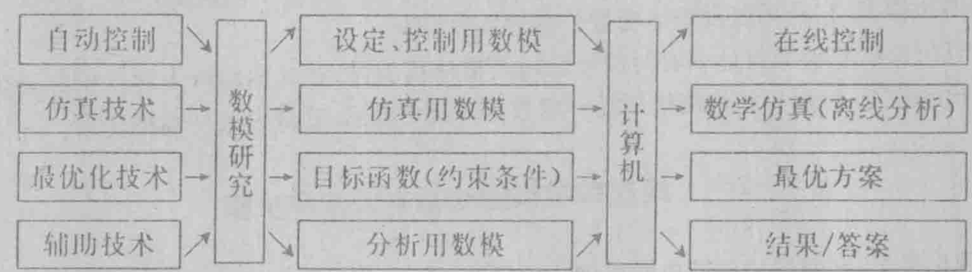
是描述不考虑空气阻力时，自由落体过程的时间与距离之间关系的数学模型，不是一般的一元二次方程或函数。又如

$$\sigma_S=\alpha_0+\alpha_1 S_i+\alpha_2 Mg+\alpha_3 Fe$$

是一个描述 Si－Mg 系铝合金的化学成分与其屈服强度关系的数学模型，也不是通常的代数方程或线性函数表达式。

数学模型是计算机控制、计算机仿真(数学仿真)、最优化技术和计算机辅助技术等不可或缺的重要内容。当用计算机控制某一生产过程或其他过程(系统)时，必须先建立描述过程中量与量之间因果关系的数学模型，编好程序并输入计算机，才能对过程或系统进行控制。计算机仿真是在计算机上用计算机模型(仿真模型)，对系统或过程进行实验的技术，而计算机模型的建立，则是在描述系统或过程内在规律的数学模型基础上，设计一种恰当的算法后得到的。在最优化技术中，首要的问题是将最优化问题表示成数学模型(又称目标函数)。然后才能运用最优化技术寻求问题的最优解。计算机辅助技术涉及许多分析功能，如辅助设计构件时的力学分析等，它们都离不开数学模型。

因此，对数学模型的研究主要包括以下如图所示的几个方面：



本章重点介绍建立数学模型的数学方法，并在简介自动控制原理基础上，说明数学模型在计算机控制中的作用。

§1－1　数学模型的分类

数学模型有多种分类方法，通常按以下几种方式予以分类：

一、按模型与时间变量 t 的关系分

1、静态模型：模型与时间变量 t 无关，即模型中不含时间变量的模型。例如前面关于合金成分与其强度关系的数学模型。

2、动态模型：模型中含有时间变量 t。例如前面关于自由落体运动的数学模型，就是一个简单的动态模型。

一般而言，静态模型都是代数方程式，动态模型则多属微分方程。

二、按模型建立的依据或方法分

1. 理论模型：根据基础理论或专业技术理论，从机理上建立的模型叫做理论模型或机理模型。例如前面关于自由落体运动的模型，就是一个理论模型。

2. 统计模型：它是根据实际抽样观测结果，用统计分析方法建立的模型。

理论模型的结构比较严谨，物理概念清晰，但结构往往比较复杂；当机理还不十分清楚时，由于作出种种假设，所以对模型精度有影响。统计模型通常具有比较简单的结构，并在一定条件下保证要求的精度，但有时无法从理论上得到解释，在不同条件下精度差别较大，甚至不能使用；这主要是因为统计模型中的量具有随机性，变量间是一种相关关系所致，因此产生了第三种数学模型。

3、理论－统计模型：利用理论模型的结构形式，按实际抽样观测数据确定模型参数建立的模型，就是理论一统计模型。显然，这种模型兼有前两种模型的优点，并可有效地克服它们的缺点，因此在工程上得到了广泛应用。

三、按变量关系的表达方式分

1、公式法模型：用数学公式表达变量间内在关系的模型叫做公式法模型。

2、表格法模型：它是用数表表达变量关系的数学模型。

除计算机控制和辅助设计中有时需要使用表格法模型外，在其他场合大多使用公式法模型。而且在必要时，可以将表格法模型经一定数学方法处理后，变成公式法模型使用。

§1－2　建立数学模型的基本方法和步骤

对于理论模型，在问题比较简单或机理比较清楚时，一般可以直接引用。但在工程技术问题中，需要建立许多统计模型或理论一统计模型。以弥补理论模型的缺乏或不足。为了建立（研制）实用、可靠的统计模型或理论一统计模型，通常必须采取下列基本方法和步骤：

一、确定最佳试验方案和方法

由于工程技术问题具有很强的工艺性，因此除少数试验必须在试验室进行外，应尽可能从生产性试验或生产记录中获取数据，同时为了用尽可能少的试验取得足够多的有效试验数据，扩大模型的使用范围和加强模型的稳定性，在变量较多的情况下，有必要采用回归设计的方法制订最佳试验方案，试验时必须保证试验条件稳定，仪表精度足够。对试验数据，尤其是生产数据要正确判断、筛选和分析，必要时整理出图表。

二、确定合理的模型结构

因为模型结构反映过程或系统的内在规律，对试验数据的拟合精度有本质的影响。所以必须根据基础理论或专业技术理论，切实分析、判断，以确定正确的模型结构；在必要与可能时，应作出试验数据的散点图。

三、确定模型中的最佳参数

在确定模型参数时，目前广泛采用基于小二乘法的回归分析方法，以及相应的算法。

四、试验验证

模型是否与实际过程或系统吻合，需要反复试验验证，直到确认以后才能投入使用。在使用过程中，还要根据实际使用效果不断进行修正，使其逐步完善。

§1－3　模型识别与参数估计

所谓模型识别，是指认定（确定）模型的结构，参数估计则是指确定模型参数的具体数值，例如，认定模型为：

$$\hat{y}=a_0+a_1x_1+a_2x_2+\cdots+a_mx_m$$

线性结构后，具体确定 a_0、a_1，$\cdots a_m$ 共计 $m+1$ 个系数（参数）的值就是参数估计。

在模型识别与参数估计时，将碰到三种情况：

一、对过程或系统的内在规律比较清楚，能够从理论上确定模型的结构，并用数学表达式将其描述出来，但模型参数还不能具体确定，这时只需具体确定模型的参数就可以了。

二、对过程或系统的内在规律一无所知，或不十分清楚，只能知道对其输入的信息和相应的输出信息，因此这种情况就象一个看不清里面的黑箱子，只看见往里面装了什么，以及在里面经过一定作用后拿出来的是什么，而不知道里面的结构。这类问题在工程上叫做“黑箱”问题。

三、介于以上两种情况之间，即对过程或系统的内在规律只在一定程度上掌握其定性的结构，究竟如何，又不清楚，这样的问题叫做“灰箱”问题。可见，黑箱与灰箱问题都存在寻求反映过程或系统内在规律的模型结构识别与参数估计问题。对于第一类问题，可以通过试验或借助

生产数据，经一定的数学处理后推断出确定的参数值。黑箱与灰箱之类的问题，一般要用统计分析等方法才能建立它们的数学模型。

§1－4　建立线性模型的回归分析方法

在生产实践与科学实验，乃至日常生活当中，人们常碰到存在于变量之间的两种关系，一种是函数关系，又叫做确定性关系或决定性关系。如自由落体运动方程式，它表达了变量S(自由落体的距离)与t(时间)之间的关系，只要知道t值，就能准确地求出S值，这种关系就是函数关系。另一种是相关关系，即变量之间彼此相互联系，相互制约，但不能由一个或几个变量的数值精确地求出另一变量的值。例如，铝合金材料的屈服强度与其所含化学成分及加工条件等紧密相关，但又不能从其中诸化学成分的具体含量及加工具体条件等，精确确定材料的屈服强度；又如人的血压与年龄有关，但多大年龄应有多高血压才算正常，也无法精确确定。上述材料屈服强度与化学成分、加工条件之间，人的血压与年龄之间存在的关系，就是一种相关关系。

回归分析方法是一种处理变量之间的相关关系的数理统计方法。在建立数学模型中主要用到这种方法的原因也在这里。

用回归分析方法建立起来的回归方程(即数学模型)，根据变量间的关系，分为线性的与非线性的两类。我们在加工工艺中所要研究的数学模型主要是线性模型，而对许多非线性模型，又可将其线性化以后按线性回归方法处理。因此着重介绍线性回归方法。

回归分析的两大基本任务是：

(1)从一定数量的观测数据组(包括选定的因变量与所对应的若干自变量的实测值)出发，确定变量之间的关系式——回归方程式；

(2)对回归方程的显著性(或可信度)作统计检验，并同时作精度估计。

一、观测数据的散点图

以一个自变量为例，假设因变量y与自变量x之间，经5次观测后取得的观测数据如表1－1．其中下标i表示观测次数(i＝1，2，…，5)。

表 1－1

i	x_i	y_i
1	15.0	39.4
2	25.8	42.9
3	30.0	41.0
4	36.6	43.1
5	44.4	49.2

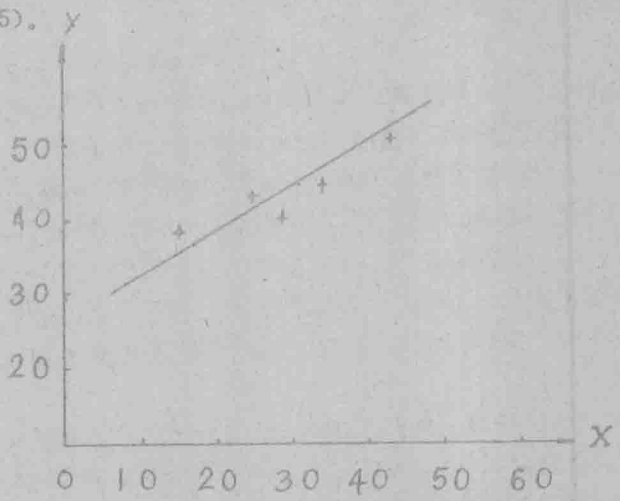


图 1－1　散点图

当将 x,y 之数值一一对应标于 x—y 平面上时，得到如图 1—1 之五点分布图形。我们称此图为观测数据的散点图。

由散点图可以看出，五个观测点大体分布在一条直线上，因此可以将 y 和 x 的关系大体确定为如下线性方程：

$$y=b_0+b_1x \tag{1—1}$$

式中 b_0、b_1 为方程的待定系数，从数学模型建立的角度来看。选定(1—1)式就是确定了模型的结构(式)，确定系数 b_0、b_1 的具体数值就是进行参数估计。

如果 y 与 x 的关系不呈线性关系，则散点图上观测点分布的轨迹将呈曲线形式，这就要用非线性方程(模型)来描述了。

因此散点图对判断变量间的关系特征很有帮助，当我们根据理论或实践经验能够判断变量间的关系特征时，可以不作散点图而直接认定模型结构，并进行参数估计。

二、曲线拟合及最小二乘解

如上所述，散点图上的观测点轨迹不是大体呈直线，便是大体呈曲线形式(当然，如变量间根本不存在任何关系时，既不呈直线也不呈曲线的情况也是可能的)，描述它们的方程不是线性方程便是非线性方程。现在的问题是，当用直线或曲线描述因变量 y 与自变量 x 的关系特征时，是否必须做到观测点全部落在表征变量关系的曲线上？回答是：原则上并不特别要求曲线真正通过某一个(或一些)观测点，而只要求它尽可能地在观测点附近通过。这样处理还可以部分抵消数据中含有的随机误差。我们把确定具有如上特点的曲线的方程问题叫做曲线拟合。

显然，曲线拟合的好坏，取决于确定方程中待定系数(如 1—1 式中的 b_0、b_1)的好坏。

假如(1—1)式中的系数 b_0、b_1 已被确定，并用 $\hat{b}_0$、$\hat{b}_1$ 表示，则每给出一个 x 值，就可以算得一个 y 值。为区别起见，将此算得的 y 值记为 $\hat{y}$，即：

$$\hat{y}_i=\hat{b}_0+\hat{b}_1x_i, \qquad i=1,2,\cdots,5 \tag{1—2}$$

但 $\hat{y}_i$ 与观测值 y_i 并不是一回事，二者可能相等，也可能不相等，而实际的情况是极少可能相等。于是在 $\hat{y}_i$ 与 y_i 之间存在差别，一般称这种差值为“残差”，用 e 表示，即：

$$e_i=y_i-\hat{y}_i=y_i-(\hat{b}_0+\hat{b}_1x_i) \qquad i=1,2,\cdots,5 \tag{1—2$'$}$$

由(1—2)′式可知，残差愈小，表明曲线拟合得愈好。而残差的大小仅取于系数 $\hat{b}_0$、$\hat{b}_1$。因此 $\hat{b}_0$、$\hat{b}_1$ 的确定应以使得残差最小为准则。

使残差最小可以有各种各样的原则，例如：

(1)使残差的最大绝对值达到最小，即令

$$T=\max_i|e_i| \quad 最小$$

(2)使残差的绝对值之和达到最小，即令

$$A=\sum_i|e_i| \quad 最小$$

(3)使残差的平方和达到最小，即令

$$J=\sum_i e_i^2 \quad 最小$$

显然，如用残差和达到最小(即 $\sum_i e_i$ 最小)作为残差最小的准则是不充分的，因为 e_i 有正有负，所以就有可能使 $\sum_i e_i=0$，这也许是最大的残差被符号相反的残差之和抵消之后得出的

结果，于是将实际存在的最大的残差掩盖掉了。(1)、(2)两种原则比较好，能真实反映预验情况，但计算繁难。原则(3)既能反映残差真实情况（将残差一律变成正值以后再相加），计算起来又较容易，因此用得较多。用残差平方和最小这一原则确定系数（如 $\hat{b}_0$、$\hat{b}_1$）的方法叫做最小二乘法，确定的系数就叫最小二乘解。

三、线性回归分析计算

线性回归分析计算的中心问题，就是根据观测数据，对线性回归方程（直线方程）进行最佳拟合。换句话说，就是要使一般的线性回归方程

$$y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_mX_m+\varepsilon$$

或
$$y=\beta_0+\sum_{i=1}^{m}\beta_iX_i+\varepsilon \qquad (1-3)$$

在确定回归系数以后，方程代表的直线（回归直线）与全部观测点的距离为最小（即式中 ε 最小）。

式中：y——因变量；

X_1、X_2、$\cdots X_m$——m 个自变量；

β_0、β_1、$\cdots$、β_m——m+1 个回归系数；

ε——残差，因具有随机性，又叫随机误差，是除 X 以外的许多随机因素对 Y 影响的总和，一般被认为是服从正态分布 $N(O,\sigma^2)$ 的随机变量。

确定回归系数以后的回归直线方程，因满足残差 ε 最小的原则，故可写为：

$$\hat{y}=b_0+b_1x_1+b_2x_2+\cdots\cdots+b_mx_m \qquad (1-4)$$

式中：$\hat{y}$——因变量的回归值或预报值；

b_0、b_1，$\cdots$，b_m——m+1 个经回归计算后确定的回归系数，即回归系数 β_0、β_1、$\cdots\beta_m$ 的估计值。

我们用最小二乘法求出系数 b_0、b_1、$\cdots$，b_m 的具体数值，即求 β_0、β_1、$\cdots\beta_m$ 的最小二乘解。

设从 N 次观测中获得 N 组相互独立的观测数据为：

$$\begin{gathered}y_1, x_{11}, x_{12}, \cdots\cdots\cdots, x_{1m};\\ y_2, x_{21}, x_{22}, \cdots\cdots\cdots, x_{2m};\\ \vdots\\ y_N, x_{N1}, x_{N2}, \cdots\cdots\cdots, x_{Nm}\end{gathered}$$

其中 M 为自变量的个数，且一般为 N>m。如已知 y 与 x_1、x_2，$\cdots$，x_m 的内在联系是线性的，则可用线性回归方程(1-3)式来表达这种关系：

$$\begin{aligned}y_1&=\beta_0+\beta_1X_{11}+\beta_2X_{12}+\cdots+\beta_mX_{1m}+\varepsilon_1\\ y_2&=\beta_0+\beta_1X_{21}+\beta_2X_{22}+\cdots+\beta_mX_{2m}+\varepsilon_2\\ &\vdots\\ y_N&=\beta_0+\beta_1X_{N1}+\beta_2X_{N2}+\cdots+\beta_mX_{Nm}+\varepsilon_N\end{aligned} \qquad (1-5)$$

或一般地写为：

$$y_n=\beta_0+\beta_1X_{n1}+\beta_2X_{n2}+\cdots+\beta_mX_{nm}+\varepsilon_n \qquad (1-5)'$$

如系数 β_i 的最小二乘解 b_i 已经求得，下标 $i=0,1,2,\cdots m$，则回归方程可写为：

$$\begin{aligned}\hat{y}_1&=b_o+b_1x_{11}+b_2x_{12}+\cdots+b_mx_{1m}\\ \hat{y}_2&=b_o+b_1x_{21}+b_2x_{22}+\cdots+b_mx_{2m}\\ &\vdots\\ \hat{y}_N&=b_o+b_1x_{N1}+b_2x_{N2}+\cdots+b_mx_{Nm}\end{aligned}\tag{1—6}$$

或一般地写为：

$$\hat{y}_n=b_o+b_1x_{n1}+b_2x_{n2}+\cdots+b_mx_{nm}\tag{1—6}'$$

由最小二乘法知道，估计值 $b_o, b_1, \cdots b_m$ 应使回归值 $\hat{y}_n$ 与实际观测值 y_n 的残差平方和为最小，即使：

$$J=\sum_{n=1}^{N}e_n^2=\sum_{n=1}^{N}(Y_n-\hat{y}_n)^2$$

$$=\sum_{n=1}^{N}(y_n-b_o-b_1x_{n1}-b_2x_{n2}-\cdots-b_mX_{nm})^2$$

为最小。

由于 $y_n, x_{n1}, x_{n2}, \cdots, x_{nm}$ 是已知的观测值，因此 J 仅是 $b_o, b_1, \cdots, b_m$ 的非负二次式，其最小值一定存在。根据数学分析中求极值的原理，$b_o, b_1, \cdots, b_m$ 应是下列方程组的解：

$$\begin{aligned}&\frac{\partial J}{\partial b_0}=0\\ &\frac{\partial J}{\partial b_i}=0 \qquad (i=1,2,\cdots\cdots,m)\end{aligned}\tag{1—7}$$

或 $\dfrac{\partial J}{\partial b_i}=\dfrac{\partial}{\partial b_i}\sum\limits_{n}(y_n-b_o-b_1x_{n1}-b_2x_{n2}-\cdots\cdots-b_mx_{nm})^2=0$

$$i=0,1,2,\cdots,m$$

式中 $\sum\limits_{n}$ 表示对 n 从 1 到 N 求和(下同)。

按(1—7)式对 J 分别求偏导数并令其等于零后得到：

$$-2\sum_{n}(y_n-b_o-b_1x_{n1}-b_2x_{n2}-\cdots-b_mx_{nm})\cdot 1=0$$

$$-2\sum_{n}(y_n-b_o-b_1x_{n1}-b_2x_{n2}-\cdots-b_mx_{nm})x_{ni}=0$$

即

$$\begin{aligned}&\sum_{n}(y_n-b_o-b_1x_{n1}-b_2x_{n2}-\cdots b_mx_{nm})=0\\ &\sum_{n}(y_n-b_o-b_1x_{n1}-b_2x_{n2}-\cdots b_mx_{nm})x_{ni}=0\\ &i=1,2,\cdots,m\end{aligned}\tag{1—8}$$

方程式(1—8)叫做正规方程组或法方程。

我们将(1—8)式按 i 的顺序展开，进一步具体化为：

$$
\begin{aligned}
&Nb_0+(\sum_n X_{n1})b_1+(\sum_n X_{n2})b_2+\cdots+(\sum_n X_{nm})B_m=\sum_n X_{n1}y_n\\
&(\sum_n X_{n1})b_0+(\sum_n X_{n1}^2)b_1+(\sum_n X_{n1}X_{n2})b_2+\cdots+(\sum_n X_{n1}X_{nm})b_m=\sum_n X_{n1}y_n\\
&(\sum_n X_{n2})b_0+(\sum_n X_{n2}X_{n1})b_1+(\sum_n X_{n2}^2)b_2+\cdots+(\sum_n X_{n2}X_{nm})b_m=\sum_n X_{n2}y_n\\
&\vdots\\
&(\sum_n X_{nm})b_0+(\sum_n X_{nm}X_{n1})b_1+(\sum_n X_{nm}X_{n2})b_2+\cdots+(\sum_n X_{nm}^2)b_m=\sum_n X_{nm}y_n
\end{aligned}
\tag{1-9}
$$

由(1－9)式可以看出，正规方程组的系数构成的矩阵是对称的。下面将正规方程组用矩阵格式表示为：

$$AB=C \tag{1-10}$$

式中

$$
A=\begin{bmatrix}
N & \sum_n X_{n1} & \sum_n X_{n2} & \cdots\cdots & \sum_n X_{nm}\sum_n X_{n1}\\
\sum_n X_{n1} & \sum_n X_{n1}{}^2 & \sum_n X_{n1}X_{n2} & \cdots\cdots & \sum_n X_{n1}X_{nm}\\
\sum_n X_{n2} & \sum_n X_{n2}X_{n1} & \sum_n X_{n2}^2 & \cdots\cdots & \sum_n X_{n2}X_{nm}\\
\vdots & & & & \\
\sum_n X_{nm} & \sum_n X_{nm}X_{n1} & \sum_n X_{nm}X_{n2} & \cdots & \sum_n X_{nm}^2
\end{bmatrix}
$$

$$
=\begin{bmatrix}
1 & 1 & \cdots\cdots & 1\\
X_{11} & X_{21} & \cdots\cdots & X_{N1}\\
X_{12} & X_{22} & \cdots\cdots & X_{N2}\\
\vdots & & & \\
X_{1m} & X_{2m} & \cdots\cdots & X_{Nm}
\end{bmatrix}
\begin{bmatrix}
1 & X_{11} & X_{12} & \cdots\cdots & X_{1m}\\
1 & x_{21} & X_{22} & \cdots\cdots & X_{2m}\\
\vdots & & & & \\
\vdots & & & & \\
1 & X_{N1} & X_{N2} & \cdots\cdots & X_{Nm}
\end{bmatrix}
=X'X;
$$

$$
B=\begin{bmatrix} b_0\\ b_1\\ b_2\\ \vdots\\ b_m \end{bmatrix};
$$

$$
C=\begin{bmatrix}
\sum_n y_n\\
\sum_n x_{n1}y_n\\
\sum_n x_{n2}y_n\\
\vdots\\
\sum_n x_{nm}y_n
\end{bmatrix}
=\begin{bmatrix}
1 & 1 & \cdots\cdots & 1\\
X_{11} & X_{21} & \cdots\cdots & X_{N1}\\
X_{12} & X_{22} & \cdots\cdots & X_{N2}\\
\vdots & & & \\
X_{1m} & X_{2m} & \cdots\cdots & X_{Nm}
\end{bmatrix}
\begin{bmatrix} y_1\\ y_2\\ \vdots\\ \vdots\\ y_n \end{bmatrix}
=X'Y
$$

因此又可将(1－10)式改写成：

$$(X'\ X)B=X'\ Y \tag{1-10$'$}$$

分别称 A 为(m+1)×(m+1)维系数矩阵,B 为(m+1)维(未知数)列向量,C 为(m+1)维(常数)列向量,X 为[N×(m+1)]维结构矩阵,X' 为 X 的转置矩阵。

当系数矩阵 A 满秩(或非奇异)时,A 的逆阵 A^{-1}存在,因此

$$B=A^{-1}C=(X'X)^{-1}(X'Y) \tag{1-11}$$

(1-11)式就是(1-5)式中系数的最小二乘解。

为了简化矩阵运算,可以对正规方程组的系数矩阵等作降低一阶的处理(推导过程此处从略,可以参阅有关回归分析理论的专著)。经降阶后的系数矩阵 A 及常数项矩阵 C(列向量)变为:

$$A=\begin{bmatrix} S_{11} & S_{12} & \cdots\cdots & S_{1m} \\ S_{21} & S_{22} & \cdots\cdots & S_{2m} \\ \vdots & & & \\ S_{m1} & S_{m2} & \cdots\cdots & S_{mm} \end{bmatrix},$$

$$C=\begin{bmatrix} S_{1y} \\ S_{2y} \\ \vdots \\ S_{my} \end{bmatrix}$$

而以上矩阵中诸元素的计算公式分别是:

$$S_{ij}=S_{ji}=\sum_n (X_{ni}-\overline{X}_i)(X_{nj}-\overline{X}_j)$$

$$=\sum_n X_{ni}X_{nj}\sum_n X_{ni})(\sum_n X_{nj})/N \tag{1-12}$$

$$S_{iy}=\sum_n (X_{ni}-\overline{X}i)(y_n-\overline{y})$$

$$=\sum_n X_{ni}y_n(\sum_n X_{ni})(\sum_n y_n)/N \tag{1-13}$$

$$i,j=1,2,\cdots,m$$

式中

$$\overline{X}_i=\frac{1}{N}\sum_n x_{ni} \qquad \overline{y}=\frac{1}{N}\sum_n y_n$$

当 i=j 时,有

$$S_{ii}=\sum_n (X_{ni}-\overline{X}_i)^2 \qquad i=1,2,\cdots,m \tag{1-14}$$

这时的列向量 B 是:

$$B=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}=A^{-1}C \tag{1-15}$$

回归方程的常数项

$$b_o=\overline{y}-\sum_{i=1}^{m} b_i\overline{x}_i \tag{1-16}$$

例题 1-1:

设从生产中获得如表 1-2 所列的 11 组观测数据,若从经验知道数据之间大致呈线性关系,试建立它们的回归方程。

解:为了计算方便,首先列出如表 1-3 所示的计算格式表,然后用降阶后的矩阵格式计

算。

表 1—2

n	1	2	3	4	5	6	7	8	9	10	11
y_n	1.80	1.53	1.80	1.84	1.74	1.81	2.00	1.70	1.84	1.84	1.60
x_{n1}	1.20	0.90	1.50	1.30	1.20	1.00	1.60	0.80	1.40	1.10	1.00
X_{n2}	2.00	2.00	2.10	1.80	2.20	2.00	2.40	1.80	2.30	2.20	1.90

先计算 A、C 矩阵诸元素：

$$S_{11}=\sum_{n}(X_{n1}-\overline{X}_1)^2=0.636$$

$$S_{22}=\sum_{n}(X_{n2}-\overline{X}_2)^2=0.385$$

$$S_{12}=S_{21}=\sum_{n}(X_{n1}-\overline{X}_1)(X_{n2}-\overline{X}_2)=0.323$$

$$S_{1y}=\sum_{n}X_{n1}y_n(\sum_{n}X_{n1})(\sum_{n}y_n)/N=23.33-13\times19.55/11=0.227$$

$$S_{2y}=\sum_{n}X_{n2}y_n(\sum_{n}X_{n2})(\sum_{n}y_n)/N=40.48-22.7\times19.55/11=0.136$$

于是有：

$$A=\begin{bmatrix}S_{11} & S_{12}\\ S_{21} & S_{22}\end{bmatrix}=\begin{bmatrix}0.636 & 0.323\\ 0.323 & 0.385\end{bmatrix}$$

$$C=\begin{bmatrix}S_{1y}\\ S_{2y}\end{bmatrix}=\begin{bmatrix}0.227\\ 0.136\end{bmatrix}$$

因此得：

$$B=A^{-1}C=\begin{bmatrix}0.636 & 0.323\\ 0.323 & 0.385\end{bmatrix}^{-1}\begin{bmatrix}0.227\\ 0.136\end{bmatrix}=\begin{bmatrix}0.310\\ 0.093\end{bmatrix}$$

即 $b_1=0.310, b_2=0.093$

$$b_0=\overline{y}-b_1\overline{x}_1-b_2\overline{x}_2=1.78-0.31\times1.18-0.093\times2.06=1.223$$

最后得所需建立的回归方程：

$$\hat{y}=1.223+0.31X_1+0.093X_2$$

四、回归方程的显著性水平及精度检验

线性回归方程建立起来之后，该方程是否真实描述了因变量 Y 与自变量 X_1、X_2、……X_m 之间的相关关系？或者说，所建立的回归方程是否有意义？如果有意义，它的精度如何？必须对此作统计检验并估计精度。

为了便于检验，需有一个刻画 y 与诸 X 间线性相关的紧密程度的数量性指标，这个指标就是复相关系数 R。现在从多元回归的方差分析导出复相关系数的计算公式，然后说明它的意义。

变量 y 是个随机变量，其取值必然存在波动，我们称此波动现象为变差，对每次 y 的观测值来说，其变差的大小可以用实际取值（观测值）偏离 Y 的平均数 $\overline{y}$的大小（叫做离差）"$y-\overline{y}$"来表示，而对 n 次观测值的总变差（即总的波动大小），则用各次离差的平方和表示。即：

表 1—3

n	y_n	x_{n1}	x_{n2}	$x_{n1}\cdot y_n$	$x_{n2}\cdot y_n$	$x_{n1}-\bar{x}_1$	$x_{n2}-\bar{x}_2$	$(x_{n1}-\bar{x}_1)^2$	$(x_{n2}-\bar{x}_2)^2$	$(x_{n1}-\bar{x}_1)(x_{n2}-\bar{x}_2)$
1	1.80	1.20	2.00	2.16	3.60	0.02	-0.06	0.0004	0.0036	-0.0012
2	1.58	0.90	2.00	1.42	3.16	-0.28	-0.06	0.0784	0.036	0.0168
3	1.80	1.50	2.10	2.70	3.78	0.32	0.04	0.1024	0.0016	0.0128
4	1.84	1.30	1.80	2.39	3.31	0.12	-0.26	0.0144	0.0676	-0.0312
5	1.74	1.20	2.20	2.09	3.83	0.02	0.14	0.0004	0.0196	0.0028
6	1.81	1.00	2.00	1.81	3.62	-0.18	-0.06	0.0324	0.0036	0.0108
7	2.00	1.60	2.40	3.20	4.80	0.42	0.34	0.1764	0.1156	0.1428
8	1.70	0.80	1.80	1.36	3.06	-0.31	-0.26	0.0961	0.0676	0.0806
9	1.84	1.40	2.30	2.58	4.23	0.22	0.24	0.0484	0.0576	0.0528
10	1.84	1.10	2.20	2.02	4.05	-0.08	0.14	0.0064	0.0196	-0.0112
11	1.60	1.00	1.90	1.60	3.04	-0.18	-0.16	0.0324	0.0256	0.0288
$\sum_n$	19.55	13.00	22.70	23.33	40.48	0.02	0.04	0.5881	0.3856	0.3046

$$\bar{y}=\frac{1}{N}\sum_n y_n=1.78,\quad \bar{x}_1=\frac{1}{N}\sum_n x_{n1}=1.18,\quad \bar{x}_2=\frac{1}{N}\sum_n x_{n2}=2.06$$