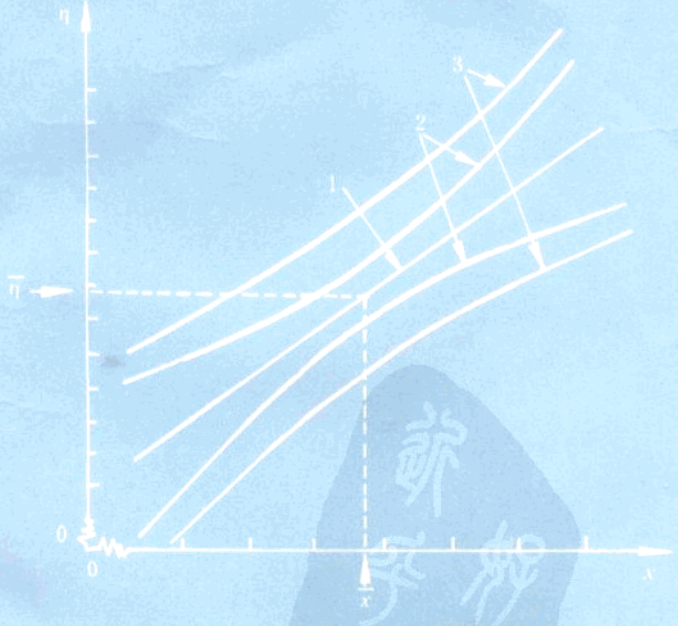
多元回归分析方法
及
在船舶科研设计中的应用



国防工业出版社

多元回归分析方法及在船舶科研设计中的应用

周连第 编

国防工业出版社

内容简介

本书所叙述的回归分析方法是数理统计的一个分支，它在船舶科研设计的许多领域中得到广泛的应用。书中介绍了多元回归分析的基本原理、数值计算方法，以及在船舶科研设计某些领域中的实际应用。书末列出了一部分参考资料，以及计算框图和程序，供参考。

本书可供从事船舶科研设计人员及有关大专院校师生参考。

多元回归分析方法及在船舶科研设计中的应用

周连第 编

*

国防工业出版社 出版

北京市书刊出版业营业许可证出字第074号

国防工业出版社印刷厂印装　内部发行

*

850×1168 1/32　印张 6　149千字

1979年4月第一版　1979年4月第一次印刷　印数：0,001— 1,000册

统一书号：N15034·1754　定价：0.78元

前　言

迄今，设计一艘船舶（包括螺旋桨）完全基于理论尚不切实际。试验和经验辅助两者都是必要的。传统的经验设计方法，一般都是统计以往大量的已设计使用船舶的线型和性能，得出简单的经验公式，或是进行变化参数的系列试验，把试验结果绘制成专供设计使用的图谱或按选定间距给出表列数值，然后根据这些简单的经验公式、图谱和表列数值，用手工计算的方法来进行设计。随着电子计算机的迅速发展，国内外都在探索利用计算机进行设计的技巧来代替传统的经验设计方法。利用计算机进行设计，不但由于其计算速度快，可以减少设计人员手工计算的工作量，缩短设计周期，更为主要的是，它能使船舶和螺旋桨设计者在初步设计阶段研究许多变量，使设计最优化，从而可以提高设计质量，所以它是一个先进的设计方法。但是利用计算机进行设计，必须迅速、准确地将传统经验设计所采用的图谱、图解或按选定间距给出的表列数值等设计数据以及大量的已设计船舶的统计数据转变成由数学方程式所定义的信息，这样才能使计算机充分推广用于解决设计问题。其正确的途径是用数理统计的方法来处理大量的试验或统计数据。其中回归分析方法是最有效的工具之一。应用这个方法，可以在电子计算机上对大量的数据进行运算、处理，从而得出项数较少，又有可能反映参数变化规律的回归方程。计算机用这个方程来进行设计及预测性能就方便、可靠。由此可见，回归分析方法在由经验设计转变到利用电子计算机进行设计的过程中扮演着何等重要的角色。

回归分析方法是数理统计的一个分支，已有几十年的发展历史，并且已有严格的理论体系。但是在国外，在船舶科研设计领

域内广泛应用回归分析方法，也是近十余年来随着电子计算机的迅速发展，并提出适合于电子计算机的数值计算方法之后才开始的。近几年来，我国对回归分析方法及其在船舶科研设计中的应用方面也开展了工作，提出了自己的数值计算方法，并在某些应用方面取得了一定的成果。本书旨在向我国从事船舶科研设计工作的同志介绍回归分析方法及其应用，主要侧重于应用方面。本书是在假定读者已具备概率论和数理统计的基本知识的前提下编写而成的。若需要系统了解这方面基础知识的读者，可参阅资料〔1〕。

本书内容分为两个部分。其中第一部分介绍多元回归分析的基本原理和数值计算方法，计有三章。关于多元回归分析的数值计算方法已有几种，目前国内外采用较多的是逐步回归法，这方面已有专门书籍和资料介绍，本书不再多述。本书只着重介绍我们在多元回归分析的应用实践过程中自己提出的，特别适用于多元多项式回归的模长界限控制的格拉姆-施米特（Gram-Schmidt）正交化方法。它的计算框图和程序载于本书的附录（一）和（二）中。第二部分介绍回归分析方法在船舶科研设计中的应用，计有六章。其中前三章是介绍我们自己工作的结果，后三章则主要介绍国外在某些方面的应用情况。

本书在编写中得到六机部七〇二所各级领导和同志们的关怀及支持，在第三、四、五、六章和§9.3编写时曾得到杨昌培、叶元培、沈贻德、韩瑞德、郑永敏、叶永兴等同志的合作和协助，以及柳忠杰、王玲娣同志帮助描绘了本书的插图，在此一并表示谢意。

由于笔者水平所限，书中难免有错误之处，请广大读者批评指正。

周连第

目　录

第一章　什么是回归分析 …… 1

§1.1　两种不同类型的变量关系——函数与相关 …… 1

§1.2　什么叫回归 …… 2

§1.3　什么是回归分析 …… 5

第二章　多元正态线性回归 …… 6

§2.1　多元正态线性回归的参数估计和分布 …… 7

§2.2　多元线性回归方程效果的检验 …… 14

§2.3　如何检验每个自变量对因变量的影响程度 …… 18

§2.4　剔除一个自变量时回归系数及回归平方和的调整 …… 22

§2.5　多元线性回归的置信区间和预测区间 …… 25

§2.6　一个完整的多元回归分析所应包含内容的建议 …… 29

第三章　多元多项式回归的一个数值计算方法 …… 30

§3.1　多元多项式回归问题化为多元线性回归问题 …… 31

§3.2　"最优"回归方程的选择 …… 33

§3.3　模长界限控制的正交化方法 …… 38

第四章　螺旋桨图谱设计电子计算机化 …… 50

§4.1　螺旋桨敞水系列试验数据的回归分析 …… 50

§4.2　$\sqrt{B_P}-\delta$ 设计图谱的多项式表达 …… 55

§4.3　螺旋桨图谱设计的计算机程序 …… 63

第五章　导管螺旋桨空泡筒系列试验数据回归分析处理方法 …… 74

§5.1　大气情况时试验数据的回归分析 …… 75

§5.2　第一、二阶段空泡现象起始曲线的回归分析 …… 76

§5.3　空泡影响因子及其回归分析 …… 84

§5.4　筒壁效应修正及其结果的回归分析 …… 91

第六章　船模推进试验数据回归分析处理方法 …… 96

§6.1　目前各种自航试验方法剖析 …… 96

§6.2　船模推进试验数据回归分析处理方法 …… 99

第七章　系列或随机形式的模型阻力和推进数据的回归分析 …… 112
§7.1 国外系列或随机形式的模型阻力和推进数据回归分析方法应用情况概述 …… 112
§7.2 回归分析方法处理系列船模阻力和推进数据的实例 …… 116
§7.3 某些静水力系数和形状参数的回归分析 …… 123
第八章　船模-实船相关因子和航行数据的统计分析 …… 125
§8.1 船模-实船相关因子的统计分析 …… 125
§8.2 船舶航行数据的统计分析 …… 133
第九章　回归分析在其它方面的应用 …… 144
§9.1 单桨船模在迎浪中运动和推进试验数据的回归分析 …… 144
§9.2 设计满载吃水百分数时阻力和推进因子数据的统计分析 …… 150
§9.3 螺旋桨升力面修正因子的回归分析 …… 154
附　录 …… 157
附录一　RA3-程序框图 …… 157
附录二　RA3-程序 …… 166
附表Ⅰ　F分布表 …… 167
附表Ⅱ　t分布表 …… 170
附表Ⅲ　B系列回归多项式的幂次和系数 …… 171
附表Ⅳ　无侧斜螺旋桨升力面修正因子回归多项式的幂次和系数 …… 175
参考资料 …… 179

第一章　什么是回归分析

§1.1　两种不同类型的变量关系——函数与相关

在数学分析中，我们已熟悉两个变量间的函数关系，即对于一个变量，在某个范围内的每一个数值，都有另一个变量的一个（单值函数）或几个（多值函数）完全确定的数值与之对应。例如，通过具有一个电阻R的电路中的电流I与加在这电路两端的电压V之间关系是遵循欧姆定律的，即

$$I=\frac{U}{R}$$

这就是说，对一定的电压值，电流强度就可由上式完全确定。再如，从物理学知道，一定质量的理想气体，当温度不变时，在容器中的空气压强p与体积V的反比关系：

$$pV=cT$$

这又是一种函数关系，其中c是一常数，T是绝对温度。

但是在实际问题中，绝大多数情形下，变量之间的关系就没有这么简单。例如，在船模拖曳试验时，对于同一条船模，其拖曳速度V和阻力R之间就不存在这种确定性的函数关系。这就是说，在不同次船模拖曳试验时，对同一拖曳速度V值（当然，对不同次试验要做到拖曳速度相同也是不容易的，这里姑且假定能做到这一点），所测量到的阻力值R都是各不相同的。造成这种情况，是由于在试验时各种试验因素的影响的复杂性。如受到试验操作人员心理和生理上的因素以及外界的温度、湿度等的影响。其中有些是属于人们一时还没有认识或掌握的，有些是已认识但暂时还无法控制或测量的，再加上在测定一些变量的量值时或多

或少都有些误差，所有这些偶然因素的综合作用造成了变量之间关系的不确定性。变量之间的这种不确定性也许是由于它们之间根本不存在什么关系的结果。但是，在相当多的情况下，这种不存在确定性关系的变量之间都有一定的规律性可寻找。这是因为大量的偶然性中蕴含着必然性的规律，如果我们经过大量的试验，就会发现许多变量之间确实存在着某种客观规律。例如，上述对于同一条船模的同一拖曳速度值V，虽然不同次试验测量到的阻力值R都各不相同，而且当试验次数不多时试验值的分布是完全没有规则的，杂乱无章的，没有什么显著的规律性。但是，当试验次数增加时，分布就开始呈现一些规律性，试验次数越多，规律性就越清楚。可以发现试验值都分布在某一典型数值附近，也就是说有一定的分布规律。我们称这种变量之间的关系为统计相关或简称相关。其确切的数学定义是：

设ξ和η是两个随机变量，可以把ξ看作为自变量，η为因变量。若对于ξ的每一个固定数值$\xi=x$，就对应着因变量η的一个概率分布（条件概率分布）即$P\{\eta<y|x\}$，则称ξ和η有相关关系，其概率密度记作$f(y|x)$●。

§1.2 什么叫回归

一、回归的定义

当两个随机变量ξ和η之间不存在函数关系而存在着相关关系时，对于自变量ξ的每一个数值$\xi=x$，因变量η并没有一个确定的数值$\eta=y$与之对应。虽然如此，为了便于描述这两个变量之间的数量变化关系，我们不妨选定一个足以代表因变量η的典型数值与自变量$\xi=x$相对应，这个典型数值我们选定为当$\xi=x$时η的条件概率分布的数学期望，简称条件数学期望，

● 本书中用到的有关概率论和数理统计的术语和符号都和〔1〕一致。

记作：

$$M(\eta|\xi=x)=\int_{-\infty}^{+\infty}yf(y|x)dy \qquad (1-1)$$

事实上，典型值的这种选法在我们的试验数据处理中已经习惯采用。例如，我们在船模拖曳试验中，对某一个拖曳速度V重复做多次试验，测量到各种不同的阻力数值R，习惯上都取这些阻力数值的算术平均值作为对应这个拖曳速度的船模阻力。数学期望就是通常的算术平均值的数学抽象（见资料〔1〕的第三章）。

显然，因变量η的条件数学期望是依赖于自变量x的取值的，因此是x的一个函数：

$$M(\eta|\xi=x)=\mu(x) \qquad (1-2)$$

当x变动，点$(x,\ \mu(x))$的轨迹将描述出一条曲线。由此曲线的形态，我们可得到有关η条件数学期望$M(\eta|\xi=x)$在x的不同位置上的信息。这样的曲线称为η倚ξ的回归曲线，函数$\mu(x)$称之为η倚ξ的回归，或简称为回归，而方程

$$y=\mu(x) \qquad (1-3)$$

叫做η倚ξ的回归方程，或简称为回归方程。

引进了回归的概念以后我们就可以指出，函数与相关虽然是两种不同类型的变量关系，但是它们之间并无严格的界限。一方面，如上所述，相关的变量之间尽管没有确定的关系，但在一定的条件下，从一定的统计意义上来看，它们之间又存在着某种确定的函数关系，即回归。另一方面，尽管从理论上说一定质量气体的体积，压强和绝对温度之间存在着函数关系，但是如果我们作了多次反复的实测，则每次得到的比值pV/T并不见得都是个常数。这就是说，实际测量得到的总是非确定性关系，这是由于实际测定的数据中总是存在着误差的缘故。实验科学（包括物理学）中的许多确定性定律正是通过大量实验数据的分析和处理，经过总结和提高，从感性到理性，最后才得到更能深刻地反映变量之间关系的客观规律。在这过程中就必须运用数理统计的方法。

二、回归函数的“最小”性质

式（1-2）所定义的回归函数有一个重要的“最小”性质，这就是当 $g(x)=M(\eta|\xi=x)$ 时，$M(\eta-g(x))^2$ 为最小●。这就告诉我们，当试验时测到自变量 ξ 的某个数值 $\xi=x$ 时，要根据它来对因变量 η 作一估计，那末 $M(\eta|\xi=x)$ 是一切对 η 估计值中均方误差最小的一个。它的几何意义是随机点 (x,η) 到回归曲线 $y=\mu(x)$ 的垂直距离的数学期望为最小，见图 1-1。

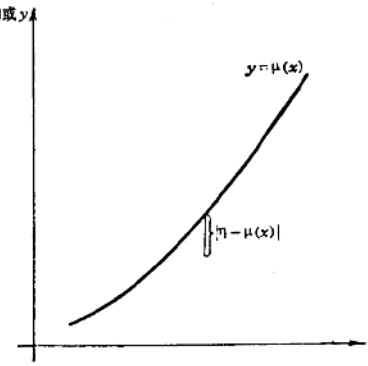


图1-1　η倚ξ的回归

回归函数的这个最小性质在实际应用中具有更进一步的指导意义。事实上，尽管我们假定了对应于每个 $\xi=x$，η 有一个确定的条件概率分布，但要确定它的条件数学期望常常是困难的。然而，我们可以假定回归函数 $\mu(x)$ 属于某个类型的函数，比方说 $\mu(x)$ 是一个 n 次多项式。这样一来，利用回归函数 $\mu(x)$ 的这个最小性质，就可以定出这个多项式的未知常数。至于怎样判定函数 $\mu(x)$ 的型式，一方面可以从理论上分析，另一方面可以由试验点的形态靠经验来判断。

上述定义的因变量 η 倚 ξ 的回归函数 $\mu(x)$ 可类似地推广到多个自变量 $\xi_1,\xi_2,\cdots\xi_n$ 的情况，此时多元回归函数 $\mu(x_1,x_2,\cdots x_n)$ 的定义是：

$$M(\eta|\xi_1=x_1,\ \xi_2=x_2,\ \cdots\xi_n=x_n)=\mu(x_1,\ x_2\cdots x_n)\quad(1\text{-}4)$$

同样，多元回归函数 $\mu(x_1,x_2,\cdots x_n)$ 也有最小性质，在此不

● 它的证明可参阅〔1〕的第三章。

再赘述。

§1.3 什么是回归分析

回归分析是处理变量之间相关关系的一种数理统计方法。上面已经提到，由于相关的变量之间不存在确定性的函数关系，但若进行了大量的试验观察，就能看出这些相关关系所呈现的规律性。但是，在客观上又只允许我们进行次数有限的试验。从表面看来这是矛盾的，然而只要我们充分利用试验测量到的数据，以及局部和整体之间的内在联系来进行分析与推断，仍然是能够认识这些规律性的。回归分析就是应用数学方法，对试验测量到的数据加以去粗取精、去伪存真、由此及彼、由表及里的处理，从而得出反映事物内部规律性的东西。它的主要任务是从有限次试验观察数据对回归函数估计、分析与推断，以及对变量进行预测和控制。它包括如下三个方面的内容：

1）确定几个特定的变量之间是否存在相关关系，如果存在的话，找出它们之间回归函数的合适的数学表达式。在实际上往往先假定变量之间存在相关关系，而回归函数的类型也是假定已知的，余下的问题是确定回归函数中的某些未知参数，这在形式上和熟知的最小二乘数据拟合是完全类似的，以下两个问题才是回归分析所独有的。

2）进行因素分析。例如对于共同影响一个变量的许多变量（因素）之间，找出哪些变量是重要因素，哪些是次要的，甚至是可以忽略的因素，这些因素之间又有什么关系等等。

3）根据一个或几个变量的值，预测或控制另一个变量的取值，并且要知道这种预测或控制可以达到什么样的精确度。

回归分析方法有很广泛的应用，在生产和科学研究工作中的许多问题都可以用这种方法得到帮助和解决。近十几年来，随着电子计算机的迅速发展，它在船舶科研设计工作中也得到广泛的应用。为用电子计算机进行船舶和螺旋桨设计创造了有利的条件。

第二章　多元正态线性回归

本章讨论的是多个自变量的问题。因为在船舶科研设计中，在绝大多数的情况下，影响因变量的因素不是一个，而是多个。例如船模拖曳试验，如果把船模的因素也考虑进去，那末影响船模阻力的因素就不只是船模的拖曳速度，而应把船模的各种几何参数也包括进去。我们称这种多个自变量的回归问题为多元回归分析。

一般的多元回归问题很复杂，我们着重讨论简单而又最一般的多元正态线性回归问题。这是因为许多非线性情形都可以化为多元线性回归来做，而正态分布则是一种最常见的概率分布[1]，在一般误差理论中都认为误差是服从正态分布的。

所谓多元是指自变量是多个，假定为 n 个 $(x_1, x_2, \cdots x_n)$；所谓正态是指因变量 η 的条件概率分布已知是正态分布，即它的概率密度为：

$$\frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(y-\mu)^2}{2\sigma^2}}$$

所谓线性是指它的条件数学期望 μ（即回归函数）是一个多元线性函数：

$$y = \mu(x_1, x_2, \cdots x_n) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \qquad (2\text{-}1)$$

其中 β_0 称之为常数项，β_i 称之为 η 对 x_i（$i = 1, 2, \cdots, n$）的偏回归系数。它的意义是，除去其它所有自变量的影响（即保持其它变量不变）后，自变量 x_i 对因变量 η 的影响。现在的问题是根据有限次试验数据，即

$$(x_{1,r}, x_{2,r}, \cdots\cdots, x_{n,r}, \eta_r) \qquad r = 1, \cdots\cdots, m$$

在正态、线性的假定下解决 §1.3 所提出的三个问题。

§2.1 多元正态线性回归的参数估计和分布

一、估计问题

多元正态线性回归的基本数学模型可以归结为：

$$\left.\begin{aligned}&\eta_r=\beta_0+\beta_1x_{1,r}+\beta_2x_{2,r}+\cdots+\beta_nx_{n,r}+\varepsilon_r\\&r=1,\ 2,\cdots,m\end{aligned}\right\}\tag{2-2}$$

其中

m 是试验次数；

η_r 是因变量 η 的第 r 次试验值；

$x_{k,r}$ 是自变量 x_k 的第 r 次试验值；

ε_r 是误差项。

由 η 的条件概率分布是正态的假定以及抽样的要求，误差 ε_r 满足如下条件：

1）ε_r 是正态分布的随机变量；

2）无偏性，它们的数学期望都是 0，即

$$M(\varepsilon_r)=0,\quad r=1,\cdots,m\tag{2-3}$$

3）等方差性，它们的方差都等于 σ，即

$$D(\varepsilon_r)=\sigma^2,\quad r=1,\cdots,m\tag{2-4}$$

4）独立性，它们是相互独立的，即

$$M(\varepsilon_i\varepsilon_j)=0\qquad i\neq j\tag{2-5}$$

概括地说，误差 $\varepsilon_1,\varepsilon_2,\cdots\cdots,\varepsilon_m$ 相互独立地遵循正态分布 $N(0,\sigma^2)$。

式(2-2)可写成：

$$\eta_r-\bar{\eta}=\beta_1(x_{1,r}-\bar{x}_1)+\beta_2(x_{2,r}-\bar{x}_2)+\cdots+\beta_n(x_{n,r}-\bar{x}_n)+\varepsilon_r,\quad r=1,\ 2,\ \cdots,\ m\tag{2-6}$$

而常数项 β_0 则有：

$$\beta_0=\bar{\eta}-\beta_1\bar{x}_1-\beta_2\bar{x}_2-\cdots-\beta_n\bar{x}_n \tag{2-7}$$

其中

$$\bar{\eta}=\frac{1}{m}\sum_{r=1}^{m}\eta_r \tag{2-8}$$

$$\bar{x}_k=\frac{1}{m}\sum_{r=1}^{m}x_{k,r}\quad k=1,\ 2,\ \cdots,\ n \tag{2-9}$$

现在欲从这m次试验观察值来定出回归系数β_k的估计值$\hat{\beta}_k$，而β_0的估计值则由式(2-7)可知为：

$$\hat{\beta}_0=\bar{\eta}-\hat{\beta}_1\bar{x}_1-\hat{\beta}_2\bar{x}_2-\cdots-\hat{\beta}_n\bar{x}_n \tag{2-10}$$

按照回归函数的最小性质，要求的估计值$\hat{\beta}_k$是使得试验点到回归函数之间的距离的平方和为最小，即有

$$Q_2=\sum_{r=1}^{m}\{(\eta_r-\bar{\eta})-[\hat{\beta}_1(x_{1,r}-\bar{x}_1)+\cdots+\hat{\beta}_n(x_{n,r}-\bar{x}_n)]\}^2=\min \tag{2-11}$$

这就是通常熟知的最小二乘法形式。式(2-11)中花括弧内的部分

$$e_r=\eta_r-\bar{\eta}-[\hat{\beta}_1(x_{1,r}-\bar{x}_1)+\cdots+\hat{\beta}_n(x_{n,r}-\bar{x}_n)]$$
$$r=1,\ \cdots,\ m \tag{2-12}$$

称之为剩余或残差。

为书写简便起见，我们采用矩阵、向量的书写形式。

令η、β、$\hat{\beta}$、ε、e分别为如下的列向量：

$$\eta=\begin{pmatrix}\eta_1-\bar{\eta}\\ \eta_2-\bar{\eta}\\ \vdots\\ \eta_m-\bar{\eta}\end{pmatrix},\quad \beta=\begin{pmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_n\end{pmatrix},\quad \hat{\beta}=\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_n\end{pmatrix}$$

$$\varepsilon=\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_m\end{pmatrix}\qquad e=\begin{pmatrix}e_1\\ e_2\\ \vdots\\ e_m\end{pmatrix} \tag{2-13}$$

$\mathbf{X}$ 为矩阵：

$$\mathbf{X}=\begin{pmatrix} x_{1,1}-\bar{x}_1 & x_{2,1}-\bar{x}_2 & \cdots & x_{n,1}-\bar{x}_n \\ x_{1,2}-\bar{x}_1 & x_{2,2}-\bar{x}_2 & \cdots & x_{n,2}-\bar{x}_n \\ \vdots & \vdots & & \vdots \\ x_{1,m}-\bar{x}_1 & x_{2,m}-\bar{x}_2 & \cdots & x_{n,m}-\bar{x}_m \end{pmatrix} \tag{2-14}$$

则式(2-6)可以简写为：

$$\boldsymbol{\eta}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \tag{2-15}$$

条件式(2-3)简写为：

$$M(\boldsymbol{\varepsilon})=0 \tag{2-16}$$

条件式(2-4)和(2-5)则可简写为：

$$M(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')=\sigma^2\mathbf{I} \tag{2-17}$$

其中 $\boldsymbol{\varepsilon}'$ 为向量 $\boldsymbol{\varepsilon}$ 的转置，即是一个行向量，$\mathbf{I}$ 为单位矩阵。

同样式(2-11)可简写为：

$$Q_2=\mathbf{e}'\mathbf{e}=(\boldsymbol{\eta}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{\eta}-\mathbf{X}\hat{\boldsymbol{\beta}})=\min \tag{2-18}$$

其中右上角带“′”的记号表示矩阵或向量的转置，以下不另说明。欲求确定最小值的 $\hat{\boldsymbol{\beta}}$，可对

$$\begin{aligned} Q_2&=\mathbf{e}'\mathbf{e}=(\boldsymbol{\eta}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{\eta}-\mathbf{X}\hat{\boldsymbol{\beta}}) \\ &=\boldsymbol{\eta}'\boldsymbol{\eta}-\hat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{\eta}-\boldsymbol{\eta}'\mathbf{X}\hat{\boldsymbol{\beta}}+\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\ &=\boldsymbol{\eta}'\boldsymbol{\eta}-2\hat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{\eta}+\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \end{aligned} \tag{2-19}$$

进行微分并使之等于 0 求得，即由

$$\frac{\partial Q_2}{\partial \hat{\boldsymbol{\beta}}'}=-\mathbf{X}'\boldsymbol{\eta}+2(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}=0 \tag{2-20}$$

推得

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}=\mathbf{X}'\boldsymbol{\eta} \tag{2-21}$$

由式(2-12)，方程(2-20)也可以写成：

$$\mathbf{X}'(\boldsymbol{\eta}-\mathbf{X}\hat{\boldsymbol{\beta}})=\mathbf{X}'\mathbf{e}=0 \tag{2-22}$$

方程(2-21)称之为多元线性回归的正规方程。矩阵：

$$\mathbf{A}=\mathbf{X}'\mathbf{X}=\begin{pmatrix} a_{11}a_{12} & \cdots & a_{1n} \\ a_{21}a_{22} & \cdots & a_{2n} \\ \vdots & & \\ a_{n1}a_{n2} & & a_{nn} \end{pmatrix} \tag{2-23}$$

称为正规方程的系数矩阵，其元素是：

$$a_{kj}=a_{jk}=\sum_{r=1}^{m}(x_{k,r}-\bar{x}_k)(x_{j,r}-\bar{x}_j) \qquad k,\ j=1,\cdots,n \qquad (2\text{-}24)$$

方程(2-21)的右端项则记作：

$$g=X'\eta=\begin{pmatrix} a_{1\eta} \\ a_{2\eta} \\ \vdots \\ a_{n\eta} \end{pmatrix} \qquad (2\text{-}25)$$

其中

$$a_{k\eta}=\sum_{r=1}^{m}(x_{k,r}-\bar{x}_k)(\eta_r-\bar{\eta}) \qquad (2\text{-}26)$$

由上面讨论可知，为求一般的 n 元线性回归方程，最后归结为解一个具有 n 个未知数的线代数方程组，即正规方程(2-21)。关于线代数方程组的具体解法很多，有行列式法、逆矩阵法、消元法、迭代法……。但是，由于多元线性回归有它自己固有的特点，即要寻求所谓“最优”回归方程，因此就提出了求解正规方程（2-21）所特有的数值解法(见第三章)。本章为了定性讨论方便起见，不妨假定已经求得了系数矩阵 A 的逆矩阵，记作 $(X'X)^{-1}$。这样，回归系数的估计值 $\hat{\beta}$ 可以求得为：

$$\hat{\beta}=(X'X)^{-1}X'\eta \qquad (2\text{-}27)$$

二、分布问题

1. $\hat{\beta}$ 的分布

把式(2-15)代入式(2-27)有：

$$\begin{aligned}\hat{\beta}&=(X'X)^{-1}X'(X\beta+\varepsilon)\\&=(X'X)^{-1}X'X\beta+(X'X)^{-1}X'\varepsilon\\&=\beta+(X'X)^{-1}X'\varepsilon\end{aligned} \qquad (2\text{-}28)$$

由上式看出 $\hat{\beta}$ 是 ε 的线性函数，所以由 ε 的正态分布可以知道 $\hat{\beta}$ 的分布也是正态分布，它的数学期望是：