T C

XIANDAI

TONGJI ZHISHI CONGSHU

2

责任编辑：袁　衡

现代统计知识丛书

市场统计分析

钱尚玮 编著

立信会计图书用品社出版发行

（上海中山西路2230号）

新华书店经销

立信会计常熟市印刷联营厂印刷

开本787×1092毫米 1/32 印张6.5 插页2 字数139,000

1991年1月第1版 1991年1月第1次印刷

印数：1—3,000

ISBN 7-5429-0064-1/F·0064

定价：2.90元

《现代统计知识丛书》序言

我们编写这套《现代统计知识丛书》的目的，一是为了弥补现有统计教材之不足，为统计教学增添新的内容；二是为了满足具有高中以上文化程度在职统计干部自学的需要，使他们的统计知识随着时间的推移而相应地得到更新。

在党的十一届三中全会前后，1978年12月，国家统计局在四川峨眉召开"全国统计教学、科研规划座谈会"以来，已经出版的我国学者编写的统计教材的数量，大大超过了"文化大革命"前的十七年，在一定程度上，内容也有所更新。这些教材，在满足统计教学的亟需方面，起了重要的作用。但是，四化建设和经济体制改革正在不断地向前推进，统计科学也在继续发展。这些统计教材，已经落后于形势的发展，不能完全适应四个现代化的要求。统计教材有待进行全面的充实和更新。

在职统计干部进行有计划的自学，不断提高业务能力，是我国造就统计人材的一个重要途径。我们一直在努力探索具有中国特色的统计工作道路，为实现统计现代化的目标而努力。在职统计干部现有的统计知识，有的已经适应不了统计现代化的需要；而许多现代化的统计知识，他们还没有掌握起来。广大统计干部，正面临着新的挑战，他们的统计知识也亟需得到补充和更新。

为满足上述两方面的要求，需要以马列主义、毛泽东思想为指针，从中国的实际情况出发，吸收国际上统计科学的新成果，编写一套具有中国特色的现代化的新的统计教材。但是，

经济体制的改革还在深入进行，统计工作也在不断变化，要很快编写一套在较长时期内适用的新的统计教材，条件还不够成熟。至于先就教材中的某一侧面进行比较深入的剖析与论述，编写小册子以充实统计新知识，补充统计教材之不足，为逐步更新统计教材创造有利条件，则是必要的，也是不难做到的。这就是编写这套《现代统计知识丛书》的由来。

邓小平同志提出："教育要面向现代化，面向世界，面向未来。"这是教育工作的方针，也是我们编写《现代统计知识丛书》的方针。《丛书》选题，应当包括我国三十多年来统计工作经验的总结，重点应当放在党的十一届三中全会以来经验的总结。中国统计工作的改革要立足在自己创造的经验的基础上。另一方面，我们必须向国际上先进的统计理论和实践学习，要注意在统计工作中运用数学方法和电子计算机的新方法，还要探索在统计中对信息论、控制论和系统工程论的运用问题。这也是《丛书》选题的重点。介绍外国经验，是为了根据中国的国情加以运用。当然，把外国的经验同我国的情况结合起来，需要一个过程，有时需要较长的过程。作者在坚持四项基本原则的前提下，可以阐发自己的独立见解，可以介绍和评述不同的学派，通过百家争鸣，共同探求真理。《丛书》将根据我国统计工作现代化的长期目标和中期规划的需要，有计划地进行编写。每一本书都要求在现有水平的基础上提高一步，写出新意，向深度和广度发展。

我们的这一设想，希望得到广大统计实际工作者和理论、教学工作者的支持，为《现代统计知识丛书》写稿，并提供宝贵意见，共同为促进我国统计工作现代化的实现而努力。

《现代统计知识丛书》编辑委员会

1985年12月

序　言

我国的社会主义经济，是在公有制基础上的有计划的商品经济。自从改革开放以来，随着经济体制的改革，我们在实行计划指导的同时，注意发挥市场调节的积极作用；也就是把计划经济和市场调节有机地结合起来。在经济实践中，越来越多的经济理论工作者和企业经营者意识到市场问题的重要意义。但由于生产的社会性和人民生活需求的多样性，以及市场变化的随机性，增加了市场研究的困难，于是有必要用科学的、定性和定量相结合的方法对市场进行研究与分析。对此，统计分析尤其是多元统计分析，将是十分有力的工具。

鉴于财经类高等院校，一般都开设社会经济统计学、数理统计学、市场学、市场预测与经营决策等课程，这为学习市场统计分析打下了基础。目前，从统计分析特别是从多元统计分析出发，系统地论述市场研究方面的书籍在国内尚不多。为此，我们编著了这本《市场统计分析》，它既补充了现代统计知识丛书的内容，又可满足高等财经类院校开设这类选修课的需要。因此本书不仅可作为财经、商业及管理类院校有关专业的教学参考书，亦可供广大统计干部和其他管理工作者学习、使用。

在编写本书时，我们力求注意到：(1)系统地、课入浅出地论述市场研究中广泛应用的近代实用的多元统计分析方法；(2)联系市场研究中的各类问题，把统计原理、方法与实际应用密切结合起来；(3)叙述既注意科学性，又考虑通俗性，将

一些重要的理论说明另列单节，使有余力的读者在阅读本书时更有兴趣。

考虑到丛书的篇幅，对书中所提及的一般熟悉的统计分析原理与方法及有关的线性代数知识不作专门阐述。

在编写本书的过程中，曾得到中国统计学会和该会的教学研究组的大力支持，特别是佟哲晖教授、莫曰达研究员、林青副研究员及袁槐副教授给予作者的热情的帮助与指导；上海财经大学黄树颜教授、张福宝副教授审阅了本书初稿，并提出不少宝贵意见与建议，在此一一表示诚挚的谢意。

由于作者水平所限，纰误可能不少，恳请读者批评指正。

作　者

1988年5月于杭州

目　　录

绪　论

近年来，我国实行了对外开放、对内搞活经济的方针，实行了计划经济与市场调节相结合的经济政策。商品经济的迅速发展，改变了长期以来传统的经营方式，企业由生产型向生产经营型转变，企业经营者逐步认识到只有在了解消费者和潜在消费者的欲望、需要、动向及其爱好的基础上制定正确的产品决策和经营决策，使产品适销对路，满足消费者的需要，企业才能得到生存与发展，产品才能长期占有市场。但是由于生产的社会性和人民生活需求的多样性，以及市场变化的随机性，要真正实现这种经营思想的转变，还需要有一套科学的市场研究的方法。市场研究是二十世纪新兴的一门科学，开始美国用于商业研究，直到第二次世界大战以后，由于统计学特别是多变量统计的迅速发展，电脑技术的兴起与普及，市场研究在经济发达国家越来越引起重视，已经成为指导生产、开拓市场的重要环节。

市场研究有广义和狭义之分。狭义的市场研究可以说是市场调查，广义的市场研究却应包括一切有关市场营销活动的分析和研究。由于市场研究是一门新兴的科学，所以其定义尚未有统一完整的解释。根据美国市场协会的解释，市场研究包括各种不同的研究，诸如：(1)市场分析是关于市场之大小、位置、性质和特征的研究；(2)销售分析是关于销售资料的分析；(3)消费者研究，即研究、探测、分析消费者态度、反应及嗜好；(4)广告研究，主要是对广告管理、作用的分析。

从这意义上说，市场研究应指广义的市场研究。本书由于是从统计分析出发来研究与分析一系列市场问题，故以广义的市场研究为其范畴。

市场研究的内容较广，包括：市场销售潜量研究；销售趋势研究；产品包装、试销、评估、比较等研究；销售网点研究；价格研究；广告研究；消费者行为研究；市场竞争研究；市场占有率研究；市场销售成本与利润分析。这些研究有的属于探测性研究，有的属于描述性研究，有的属于因果关系研究，有的属于预测性研究。不论那一类研究都离不开统计方法。若从统计分析角度考察，近代实用的多元统计分析方法，结合市场研究的各类问题，在国外市场研究公司中的应用已经十分普遍。表 1 给出的是美国与西欧公司常用的市场研究的统计方法和应用各方法的百分数，虽然仅仅是 1977 年的一份国外的调查资料，但还是可以从中窥见一斑。

表1

分 析 方 法	美国公司(%)	西欧公司(%)
统计假设检验	63	48
回归与相关分析	25	82
非参数统计	缺	18
变异数分析	35	22
判别分析	24	15
主成份与因子分析	31	20
聚类分析	22	8
时间数列分析	65	20

在我国，这几年来不论是高等院校，还是理论界或企业界，开始重视与学习现代西方市场学和市场研究技术，并且

结合我国实际，在教学研究和企业经营管理上得到了较好的应用。可以预料,随着我国经济体制改革的不断深入,用统计分析方法研究与解决市场问题将会得到更大的重视和发展。

第一章　产品需求分析(I)
——多元线性回归

§1.1　提出问题

在市场研究中，很多问题往往归结为了解一些有关变量间的联系，例如某产品的销售量与产品的价格、消费者的收入，该销售地区的人口以及所花费的广告费用等因素有关联。这种联系，实际上要求市场研究者寻求一个因变量Y与它的自变量$X_1,X_2,\cdots,X_k$之间的具体表达式。由于因变量Y所取的值与自变量$X_1,X_2,\cdots,X_k$所取的值的关系没有密切到可以唯一确定的程度。如上面所提及的各因素的变动虽然会直接影响到销售量Y的大小，但它们不能唯一地决定Y，所以市场研究者所要研究的变量间的关系是非确定性的，也称为相关关系。在统计中研究这种相关关系，就归结为弄清Y在给定$X_1=x_1$，$X_2=x_2$，…，$X_k=x_k$时的概率分布随$x_1,x_2,\cdots,x_k$而变化的规律。

§1.2　多元线性回归模型

设　Y对$X_1,X_2,\cdots,X_k$的依赖关系可分解为两部分

$$Y=f(X_1,X_2,\cdots,X_k)+u,\ 且\ E(u)=0 \tag{1.1}$$

其中：u为一随机变量，

$E(u)$为u的数学期望。

(1.1)式被称为理论的回归模型，模型(1.1)式中主要部分由回归函数f表示，它所反映的是一种"平均趋势"，余下部分u可视作某种随机干扰，通常称为模型误差或随机误差。显然，Y与$X_1,X_2,\cdots,X_k$的关系的不确定性程度主要取决于该u的影响的大小。

对(1.1)中的回归函数一种情况是毫无所知，另一种情况是总的数学形式已知，称前者所反映的回归模型为非参数性的，后者所反映的已知数学形式若包含若干个未知参数β_0，$\beta_1,\cdots,\beta_k$，且f为$X_1,X_2,\cdots,X_k$的线性函数：

$$f(X_1,X_2,\cdots,X_k)=\beta_0+\beta_1X_1+\cdots+\beta_kX_k+u \qquad (1.2)$$

称该回归模型为线性的。本章所讨论的就是这一类回归模型。

为了估计模型(1.2)中各参数$\beta_0,\beta_1,\cdots,\beta_k$的值，以及利用此模型进一步研究其它各项统计问题。就需要通过观察或试验取得样本，假定我们进行了n次观察或调查，在第i次中，$X_1,X_2,\cdots,X_k$和Y分别取值$X_{i1},X_{i2},\cdots,X_{ik}$和$Y_i(i=1,2,\cdots,n)$，相应的设该次观察中随机误差$u$取值$u_i$，$(i=1,2,\cdots,n$，注意，$u_1,u_2,\cdots,u_n$的值是无法观察到的)。这样得到了$n$个方程：

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\cdots+\beta_kX_{ik}+u_i \quad i=1,2,\cdots,n \qquad (1.3)$$

模型(1.3)被称为样本的线性回归模型。

由于各次观察的随机误差$u_1,u_2,\cdots,u_n$本身是随机变量，且满足如下假定：

(1) $E(u_1)=E(u_2)=\cdots=E(u_n)=0$ （零均值性）

(2) $D(u_1)=D(u_2)=\cdots=D(u_n)=\sigma_u^2$ （等方差性）

(3) $\mathrm{cov}(u_i,u_j)=0$，当$i\neq j$，$i,j=1,2,\cdots,n$ （不相关性）

(4) $u_1,u_2,\cdots,u_n$相互独立，$u_i\sim N(0,\sigma_u^2)$，$i=1,2,\cdots,n$ （正态性）

注意，上述的误差方差 σ_u^2 均为未知的。(2)、(3)两点可理解为各次观测不论其中自变量取什么值，所受到的随机影响程度相同，且任意两次观察的误差大小无关联。

§1.3 回归模型的矩阵表示及其统计性质

一、模型的矩阵表示

$$Y=\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix},\quad X=\begin{pmatrix} X_{10} & X_{11} & \cdots & X_{1k} \\ X_{20} & X_{21} & \cdots & X_{2k} \\ \cdots & \cdots & & \cdots \\ X_{n0} & X_{n1} & \cdots & X_{nk} \end{pmatrix},\ \text{其中}\ X_{i0}=1,\quad i=1,2,\cdots,n$$

$$\beta=\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix},\quad u=\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

(1.3)式可改写成

$$Y=X\beta+u \tag{1.4}$$

又若对一随机向量 $\xi=(\xi_1,\xi_2,\cdots,\xi_n)'$ 的均值向量定义为 $E(\xi)=[E(\xi_1),\cdots,E(\xi_n)]'$，方差阵记为 $Var(\xi)$。此时，模型的几条假定可改写如下：

(1) $E(u)=0$ （此 0 表示 $n\times 1$ 的列向量）

(2) $Var(u)=\sigma_u^2 I_n$，$0<\sigma_u^2<\infty$

(3) $u\sim N(0,\ \sigma_u^2 I_n)$ （其中 I_n 为 $n\times n$ 单位阵）

二、β 之最小二乘估计

为求 β 之最小二乘估计量 $\hat{\beta}$，考虑误差平方和

$$\sum_{i=1}^{n} u_i^2=u'u=(Y-X\beta)'(Y-X\beta)=Y'Y-2\beta'X'Y+\beta'X'X\beta$$

使上式为最小的必要条件是

$$\frac{\partial}{\partial\beta}(u'u)=-2X'Y+2X'X\beta=0$$

即 $(X'X)\beta=X'Y$

故 $\hat{\beta}=(X'X)^{-1}X'Y$ (1.5)

一般，把$(X'X)\beta=X'Y$称为正规方程，而把$\hat{\beta}=(X'X)^{-1}X'Y$称为正规方程的解，正规方程有解的充要条件是$(X'X)^{-1}$存在。

三、$\hat{\beta}$ 的性质及 σ_u^2 的估计

1．$\hat{\beta}$ 是 β 的线性无偏估计

先定义 $\hat{\beta}$ 的数学期望$E(\hat{\beta})$为：

$$E(\hat{\beta})=E\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \vdots\\ \hat{\beta}_k\end{pmatrix}=\begin{pmatrix}E\hat{\beta}_0\\ E\hat{\beta}_1\\ \vdots\\ E\hat{\beta}_k\end{pmatrix}$$

由于 $\hat{\beta}=(X'X)^{-1}X'Y$，故 $\hat{\beta}$ 是 $Y_1, Y_2,\cdots,Y_k$ 的线性函数，即 $\hat{\beta}$ 是一个线性估计。又：

$$\begin{aligned}E(\hat{\beta})&=E[(X'X)^{-1}X'Y]=(X'X)^{-1}X'E(X\beta+u)\\&=(X'X)^{-1}X'(X\beta)+E(u)=\beta \quad (\because E(u)=0)\end{aligned}$$

（注意，在上面的推导中用到了这样的事实，即X与 β 均不是随机变量，只有 $\hat{\beta}$ 和 u 才是随机变量。）

由此可知 $\hat{\beta}$ 是 β 的无偏估计，即用 $\hat{\beta}$ 估计 β 时可能有偏差，但多次的平均结果接近于真值 β。

2．$Var(\hat{\beta})$是 $\hat{\beta}$ 的方差阵

$$Var(\hat{\beta})=\sigma_u^2(X'X)^{-1} \tag{1.6}$$

事实上，

$$\begin{aligned}Var(\hat{\beta})&=E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)']\\&=E[(X'X)^{-1}X'(uu')X(X'X)^{-1}]\\&=(X'X)^{-1}X'E(uu')X(X'X)^{-1}\\&=(X'X)^{-1}X'Var(u)X(X'X)^{-1}\end{aligned}$$

$$= (X'X)^{-1}X'\sigma_u^2 I_n X(X'X)^{-1}$$
$$= \sigma_u^2 (X'X)^{-1}$$

此结果表明，$\hat{\beta}_i$ 的方差是矩阵 $(X'X)^{-1}$ 的第 i 个对角元素乘以随机误差 u 的方差。显见，$Var(\hat{\beta}_i)$的大小可作为估计量 $\hat{\beta}_i$ 好坏的标准，即 $Var(\hat{\beta}_i)$愈小，$\hat{\beta}_i$ 愈好，又$Var(\hat{\beta}_i)$依赖于随机误差的方差 σ_u^2，故 σ_u^2愈大，则模型误差的影响愈大，而 β_i 愈不易准确估计。

3．σ_u^2 的无偏估计为

$$\hat{\sigma}_u^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n-k-1} \tag{1.7}$$

事实上，只需说明 $E(\hat{\sigma}_u^2) = \sigma_u^2$ 即可，由于

$$E(Y - X\hat{\beta})'(Y - X\hat{\beta}) = E[u - X(X'X)^{-1}X'u]'$$
$$[u - X(X'X)^{-1}X'u] = E[u'u - u'X(X'X)^{-1}X'u]$$
$$= E(u'u) - E(u'X(X'X)^{-1}X'u)$$
$$= n\sigma_u^2 - E(u'Au) \quad [A = X(X'X)^{-1}X']$$
$$= n\sigma_u^2 - \sigma_u^2 trA = (n-k-1)\sigma_u^2$$

4．若 u 的正态性假定 $u \sim N(0, \sigma_u^2 I_n)$ 成立，则可以证明：

(1) $(n-k-1)\hat{\sigma}_u^2/\sigma_u^2 \sim \chi_u^2(n-k-1)$，

(2) $\hat{\sigma}_u$ 与 $\hat{\beta}$ 独立。

§1.4 高斯—马尔柯夫 (Gauss-Markov)定理

定理 对线性回归模型 $Y = X\beta + u$，在 u 满足 $E(u) = 0$，$Var(u) = \sigma_u^2 I_n$ 的假定下，最小二乘估计量 $\hat{\beta} = (X'X)^{-1}X'Y$ 是 β 的最小方差无偏估计。

在证明定理之前，先说明定理的意义。在§3中已证明