管理科学研究方法

Methods of Management Science

统计与运筹优化应用

朱顺泉 编著

清华大学出版社

管理科学研究方法

——统计与运筹优化应用

朱顺泉　编著

清华大学出版社
北　京

内 容 简 介

本书讲述管理科学研究的方法，介绍了数理统计、计量经济、多元统计与运筹优化模型及其应用。本书分为两篇：数理统计、计量经济与多元统计篇包括一些常用的随机变量分布、参数估计、假设检验、线性回归等一些常用内容和计量经济模型的检验，以及主元分析、因子分析、聚类分析、判别分析等多元统计分析及其应用等内容；运筹优化篇向读者介绍常用的优化模型及其应用，主要包括线性规划模型、整数线性规划模型、非线性规划模型、多目标决策模型、神经网络模型以及模拟决策模型及其应用等内容。

本书内容充实，通俗易懂，涉及面广。可作为大、中专院校各类学生学习数据、模型与决策、商务决策数量方法、管理科学、运筹学、数理统计学、计量经济学、多元统计学等课程的教材或参考书。也可供从事数量经济分析方法的企业管理人员参考，对从事金融、财务、会计、企业管理等实证研究的读者，本书也是一部良好的教材和参考书。

图书在版编目(CIP)数据

管理科学研究方法：统计与运筹优化应用/朱顺泉编著. —北京：清华大学出版社，2007. 2
ISBN 978-7-302-14202-7

Ⅰ. 管…　Ⅱ. 朱…　Ⅲ. ①数理统计—应用—管理学 ②运筹学—应用—管理学　Ⅳ. C931.1

中国版本图书馆 CIP 数据核字(2006)第 143794 号

责任编辑：王　青　陆浥晨
责任校对：宋玉莲
责任印制：王秀菊
出版发行：清华大学出版社　　**地　　址**：北京清华大学学研大厦 A 座
http://www.tup.com.cn　　**邮　　编**：100084
c - service@ tup.tsinghua.edu.cn
社 总 机：010-62770175　**邮购热线**：010-62786544
投稿咨询：010-62772015　**客户服务**：010-62776969
印 刷 者：北京市清华园胶印厂
装 订 者：北京国马印刷厂
经　　销：全国新华书店
开　　本：185×230　**印　张**：17　**字　数**：324 千字
版　　次：2007 年 2 月第 1 版　　**印　　次**：2007 年 2 月第1 次印刷
印　　数：1 ~ 4000
定　　价：29.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题，请与清华大学出版社出版部联系调换。联系电话：(010)62770177 转 3103　　产品编号：021947 - 01

前　言

《管理科学研究方法——统计与运筹优化应用》一书分为两篇，即上篇：数理统计、计量经济与多元统计；下篇：运筹优化。在数理统计、计量经济与多元统计篇中，向读者介绍一些常用的随机变量分布、参数估计、假设检验、回归分析模型等一些常用的数理统计知识和计量经济模型检验等内容，以及主元分析、因子分析、聚类分析、判别分析等多元统计分析及其在经济管理中的应用等内容。在运筹优化篇中，主要包括一些常用的线性规划模型、整数线性规划模型、非线性规划模型、多目标决策模型、神经网络模型以及模拟决策模型及其应用等内容。编写本书的目的是为准备从事经济、管理、金融、财务、会计实证研究的人员提供一些数量分析的工具和数量经济分析的方法。

本书融数理统计、计量经济、多元统计、运筹优化等内容于一体，内容丰富，通俗易懂，涉及的范围广。可作为大、中专院校商学院、管理学院、经济学院、财经学院、信息学院等各专业，尤其是工商管理、市场营销、统计学、数量经济、财务管理、会计学、金融学、经济学、信息管理与信息系统、物流管理、管理科学与工程、技术经济及管理、工业工程、系统工程等各专业的本科高年级学生与研究生、MBA 学员学习"商务数量方法决策"、"数据、模型与决策"、"运筹学"、"数理统计学"、"计量经济学"、"多元统计学"等课程的教材或参考书。对从事管理科学与工程、统计学、数量经济学、财务管理、会计学、金融学、金融工程、经济学、信息管理与信息系统、工业工程、物流管理、系统工程等广大在职的企业管理人员也是一本难得的参考书。

作　者

2006 年 9 月于广州

目　　录

下篇 运筹优化

引　言

20 世纪以来，管理科学的发展大致经历了以下几个阶段。

第一阶段是 20 世纪 20 年代，以泰勒（Taylor）为代表的科学管理学派，其主要观点是通过提高效率来提高生产率，并通过科学方法的应用来增加工人的工资，其原理强调应用科学，创造集体的协调与合作，达到最大的产出量，培养工人的能力。

第二阶段是 20 世纪 30 年代，以梅奥（Mayo）为代表的行为科学学派，该理论主张以人为中心，激励人的积极性。

第三阶段是 20 世纪 40 年代，出现了数学管理学派，该学派主张用定量化的手段、数学模型的方法进行管理。20 世纪 80 年代以前，数学管理主要以运筹学模型的应用为主，而 20 世纪 90 年代以后，在国外，数理统计模型的应用要比运筹学模型的运用广泛得多；近年来，我国管理科学方面的各种实证研究论文，也可看到这一趋势，即统计学模型的应用越来越广泛。

第四阶段是 20 世纪 50 年代，此时出现了计算机管理学派。这是一股势力，他们把计算机广泛用于管理，继 1954 年用于工资管理和人口统计以后，在 50 年代末至 60 年代初形成了计算机管理的第一次热潮。

第五阶段是 20 世纪 70 年代，此时出现了系统工程学派。该理论提出用系统的观点、工程的观点来考虑管理问题。

20 世纪 80 年代，管理科学中出现了权变学派、比较管理学派等。

进入 20 世纪 90 年代以后，出现了学习型组织、虚拟组织等新的管理组织，企业流程再造（*BPR*）、知识管理等成了管理界人们研究的热点。

从上可见，管理科学中大量地使用数学、统计、运筹等方法和计算机工具等。

因此，这里我们将管理科学理解为“运用数学模型，对人、财、物等进行系统和定量的分析，以作出科学决策的管理理论与方法”，该理论的主要特点：

① 管理领域的各项活动都以经济效果的优劣作为评价标准；

② 使衡量各项活动效果的标准定量化，借助数学模型描述事物的现状及发展规律，并找出最优的实施方案；

③ 强调使用先进的科学理论和管理方法，如数理统计、计量经济、多元统计、运筹优化、神经网络等数学方法与模型；

④ 强调利用计算机进行各项管理。

在发达国家，大部分成功企业均将定量化方法应用于企业生产和管理，并取得了很大的成功。为促进管理科学理论与方法在实际工作中的应用，美国大部分大学都强调计算机的应用。

自20世纪80年代以来，管理科学在我国得到了迅速的发展，许多大学建立了管理科学系，培养了大批管理人才。然而，管理科学在实际工作中的应用还远未普及。究其原因，除了我国企业的管理水平有待提高外，另一个重要问题是我国管理科学教学中存在理论与实际分离的状况。由于教学中常常过于强调数学问题，如数学公式的推导等，而对管理科学的思想、从实际问题中建立模型的技术，以及定量化方法在实际管理中的应用有所忽略。其结果，一方面使得不少管理人员望而却步，将管理科学看成深奥的、难以掌握的、抽象的数学问题；另一方面，管理科学难以在实践中普及。

本书将应用数理统计、计量经济、多元统计、运筹优化等内容，面向实际问题，建立模型，并应用计算机软件工具（SPSS，Excel 等）进行求解，着重讨论管理科学中模型的建立与应用。其目的是使读者能理解并掌握管理科学的一般理论与方法，并将其应用于管理工作的实践，以促进管理科学与实践的有机融合。

上篇

数理统计、计量经济与多元统计

第 1 章　数理统计模型及其应用

利用样本对总体进行统计推断，主要有两类问题：一类是估计问题，另一类是统计检验问题。

估计问题是多种多样的，如总体参数的估计，分布函数、密度函数的估计等。估计的方法也很多，如最小二乘法、最大似然法、矩估计法等。

统计推断的另一类问题与估计问题的提法不同，它是从样本值出发去判断关于总体分布的某种“看法”是否成立。例如从 6 000 名学生中随机地抽取 100 名学生测量体重 X，利用 100 名学生的体重数据来判断看法“平均体重 $E(X)=62\text{kg}$”是否成立。“看法”又叫做“假设”，这就是所谓的统计检验问题。

除了统计检验外，还有计量检验问题，我们在第 2 章将作介绍。

在介绍估计与统计检验之前，下面首先对常用的随机变量分布作一个简要的回顾。

1.1　常用的随机变量分布

1.1.1　正态分布

在实际问题中，我们考虑的总体特征量（即随机变量 X）的规律有很多都可以用正态分布来描述。正态分布是总体的一种理论分布，有严格的数学定义。

如果随机变量 X 的概率分布密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\frac{(x-\mu)^2}{2\sigma^2},\quad -\infty<x<+\infty$$

则称随机变量 X 服从均值为 $\mu(-\infty<\mu<+\infty)$，方差为 $\sigma^2(\sigma>0)$ 的正态分布，记为 $X\sim N(\mu,\sigma^2)$。

正态分布的性质和特点如下。

① 正态分布完全由其均值 μ 和方差 σ^2 决定；

② 正态分布的密度函数曲线是中间高两边低、对称、光滑的“钟形”曲线；

③ 正态分布位置特征量均值、众数和中位数相等。

④ 正态分布的偏度和峰度均为 0。偏度和峰度是用于描述分布形状的特征量，偏度

用于刻画总体分布是否对称；峰度用于衡量“分布的尾重”。

正态分布的分布密度函数为

$$F(x)=P(X\leqslant x)=\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{x}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)\mathrm{d}y,\quad -\infty<x<+\infty。$$

特别地，当 $\mu=0,\sigma^2=1$ 时的正态分布称为标准正态分布，记为 $X\sim N(0,1)$。常用 $\varphi(x),\Phi(x)$ 表示其概率密度和分布函数，即

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right),\quad -\infty<x<+\infty$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{y^2}{2}\right)\mathrm{d}y,\quad -\infty<y<+\infty$$

对于任意的 x，查 $\Phi(x)$ 的函数值表即可求标准正态分布的函数值。例如，$\Phi(1)=0.841\,3$，$\Phi(1.64)=0.949\,5$。

由于标准正态分布曲线同横轴所包围的面积是常数 1，故

$$\Phi(-x)=1-\Phi(x)。$$

若 $X\sim N(0,1)$，还可得到以下结论：

$$P(a\leqslant X\leqslant b)=\Phi(b)-\Phi(a),$$

$$P(|X|\leqslant a)=P(-a\leqslant X\leqslant a)=\Phi(a)-[1-\Phi(a)]=2\Phi(a)-1。$$

例如

$$\Phi(-1)=1-\Phi(1)=1-0.841\,35=0.158\,65,$$

$$P(-1\leqslant X\leqslant 1)=2\Phi(1)-1=2\times 0.841\,35-1=0.682\,7=68.27\%,$$

$$P(-2\leqslant X\leqslant 2)=2\Phi(2)-1=2\times 0.977\,25-1=0.954\,5=95.45\%,$$

$$P(-3\leqslant X\leqslant 3)=2\Phi(3)-1=2\times 0.998\,65-1=0.997\,3=99.73\%。$$

对于一般的正态分布 $N(\mu,\sigma^2)$，只需对 X 进行线性变换，即设 $Z=\frac{X-\mu}{\sigma}$，则随机变量 Z 服从标准正态分布，即 $Z\sim N(0,1)$。因而求一般正态分布在某区间上的概率，就转化为求标准正态分布在相应区间上的概率。

若 $X\sim N(\mu,\sigma^2)$，有以下结论：

$$P(X\leqslant x)=P\left(\frac{X-\mu}{\sigma}\leqslant\frac{x-\mu}{\sigma}\right)=P\left(Z\leqslant\frac{x-\mu}{\sigma}\right)=\Phi\left(\frac{x-\mu}{\sigma}\right),$$

$$P(X>x)=1-P(X\leqslant x)=1-\Phi\left(\frac{x-\mu}{\sigma}\right),$$

$$P(a\leqslant X\leqslant b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)。$$

易得，$X\sim N(0,1)$ 时，

$$P(|X-\mu|\leqslant\sigma)=68.27\%,$$
$$P(|X-\mu|\leqslant 2\sigma)=95.45\%,$$
$$P(|X-\mu|\leqslant 3\sigma)=99.73\%。$$

显然 $|X-\mu|>3\sigma$ 的概率很小，因而可以认为 X 的值几乎一定落在 $(\mu-3\sigma,\mu+3\sigma)$ 内，这就是重要的 3σ 准则（或小概率原则），以后会经常用到。

1.1.2 χ^2 分布

如果随机变量 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}\mathrm{e}^{-x/2}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

则称 X 服从自由度为 n 的 χ^2（卡方）分布，记作 $X\sim\chi^2(n)$；这里 $\Gamma(s)=\int_0^{\infty}x^{s-1}\mathrm{e}^{-x}\mathrm{d}x$，$s>0$，

$$\Gamma(s+1)=s\Gamma(s),\quad \Gamma(n+1)=n!。$$

χ^2 分布的数学期望、方差分别为

$$E(X)=n,\quad D(X)=2n。$$

χ^2 分布适用于对总体方差的统计推断、拟合优度检验、独立性检验等。它是一个以自由度 n 为参数的分布族，自由度 n 决定了分布的形状，对于不同的 n，χ^2 分布的密度曲线也不同，它是一个不对称分布，随着 n 的增大，χ^2 分布逐渐趋于正态分布。

χ^2 分布有如下重要结论：

(1) 若 X,Y 相互独立，且分别服从自由度为 n_1,n_2 的 χ^2 分布，则 $X+Y$ 服从自由度为 n_1+n_2 的 χ^2 分布，即 $X+Y\sim\chi^2(n_1+n_2)$。

(2) 若随机变量 $X_1,X_2,\cdots,X_n$ 相互独立，且都服从 $N(0,1)$，它们的平方和 $\sum X_i^2$ 服从自由度为 n 的 χ^2 分布，记作 $\sum X_i^2\sim\chi^2(n)$。

(3) 若随机变量 $X_1,X_2,\cdots,X_n$ 相互独立，且都服从 $N(\mu,\sigma^2)$，则

$$\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)。$$

其中，$\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$ 与 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 相互独立。

最后简要说明自由度的含义。

由线性代数知，对于变量 $X_1, X_2, \cdots, X_n$，如果存在一组不全为0的常数 $c_1, c_2, \cdots, c_n$，使得

$$c_1 X_1 + c_2 X_2 + \cdots + c_n X_n = 0$$

成立，则称变量 $X_1, X_2, \cdots, X_n$ 之间存在一个线性约束条件。如果存在 k 个独立的线性约束条件，则 $X_i(i=1,2,\cdots,n)$ 中独立变量的个数为 $n-k$，称它为自由度。自由度也可粗略解释为：可以自由选择数值的变量个数。

例如，$\sum_{i=1}^{n} X_i^2$ 由 n 个独立的随机变量 X_i 构成，由于它们之间没有线性约束条件（即 $k=0$），所以它的自由度为 n。

$\sum_{i=1}^{n} (X_i - \mu)^2$ 的自由度为 n。

$\sum_{i=1}^{n} (X_i - \overline{X})^2 = (n-1)S^2$ 的自由度为 $n-1$，这是因为计算 S^2 时要用 $\overline{X}$，它有一个限制条件 $\sum_{i=1}^{n} (X_i - \overline{X}) = 0$，即相对于 $\overline{X}$ 的 n 个离差 $X_1 - \overline{X}, X_2 - \overline{X}, \cdots, X_n - \overline{X}$，只有 $n-1$ 个可以任意确定，第 n 个失去了“自由”，所以其自由度为 $n-1$。

1.1.3 t 分布

如果随机变量 X 的概率密度为

$$f(x) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad -\infty < x < \infty,$$

则称 X 服从自由度为 n 的 t 分布，记为 $X \sim t(n)$。

t 分布密度曲线的形状取决于自由度 n。t 分布与标准正态分布的密度曲线类似，都为对称分布，且取值范围都是从 $-\infty$ 到 $+\infty$。但 t 分布曲线的顶部低于标准正态分布，而两尾部又高于标准正态分布，自由度越小这种区别就越明显，随着自由度的增大，t 分布趋于标准正态分布 $N(0,1)$。

t 分布的均值、方差分别为

$$E(X) = 0, \quad D(X) = \frac{n}{(n-2)}, \quad n > 2。$$

t 分布有如下重要结论：

(1) 若随机变量 $X \sim N(0,1)$，$Y \sim \chi^2(n)$，且 X, Y 相互独立，则 $t = \frac{X}{\sqrt{Y/n}} \sim t(n)$。