第一章 序论 控制论 系统论 优化论

系统工程发展的几个阶段：1)第一阶段源远流长：大禹治水（“治水”与“治人”并重）；田忌赛马（系统的要素不变，但策略不同，则总体效果不同）；都江堰工程（系统的思想，整理考虑，综合寻优）；丁渭工程 2)第二阶段：突然崛起（大体上形成于上世纪 50、60 年代）：北极星导弹核潜艇计划；阿波罗登月计划（计划评审技术 PERT+图解评审技术 GERT）（错误:计算机的诞生标志着系统工程进入第二阶段）3）第三阶段：迅速扩张：经典案例的特点：问题明确，规模巨大，递归结构；基本方法：分解协调．各子系统求解过程受协调变量影响，大系统调节协调变量实现自己目标．；用大系统方法处理社会经济问题时遇到新问题：系统中包括有主观意志的人．；在中国上世纪 50、60 年代开始，80 年代掀起高潮，成功案例如载人航天工程．4)第四阶段:返璞归真(上世纪 90 年代以来)：系统工程的名称提的越来越少，系统工程的方法用的越来越多．**适合用大系统理论求解**:嫦娥计划;阿波罗登月计划;全国人口普查;十四五科技发展战略规划

计划评审技术 PERT：1）事件（Events）表示主要活动结束的那一点 2）活动（Activities）表示从一个事件到另一个事件之间的过程 3）松弛时间（Slack Time）不影响完工前提下可能被推迟完成的最大时间 4）关键路线（Critical Path）是 PERT 网络中花费时间最长的事件和活动的序列

系统的定义：系统是具特定**功能**的，相互间具有有机联系的许多要素所构成的一个整体．三个关键词：功能，要素，集合．特性：1)集合性 2）整体性 3)相关性 4)阶层性 5）目的性 6)环境适应性．系统实现环境适应性有如下 3 种方式：1）适应性自稳 2）适应性组织 3)自组织性．系统的分类：自然系统和人工系统；实体系统和概念系统；动态系统与静态系统；控制系统与行为系统．系统的功能：F，体现了系统与外部环境之间物质能量信息交换能力；系统的环境：E；系统的结构：S；系统的组成要素：C．$F=f(E,C,S)$．**系统论一个基本原理**：总体不等于部分和.**要素、结构、功能三者关系**：1)要素不同，功能不同；2）要素相同，结构不同，功能不同；3）要素，结构不同，也可能功能相同 4)同一结构，也可多种功能

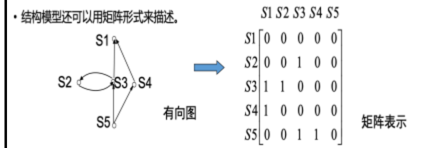
系统工程的主要特点：定量，寻优；系统工程的基本手段：建模，优化．系统工程就是用工程的方法解决系统问题．“系统”这个词最早是由古希腊哲学家提出的(赫拉克利特，德谟克里特，400BC)．由一个功能强大的元件不能构成一个系统；能够适应外界环境变换的系统才是有生命力的．错误：系统的建模只有依赖数学才能完成．不同系统之间的区别在于功能不同．错误：由一个特别强大的元件可以构造出具有特定功能的系统．不同要素，不同结构，可以实现相同功能．钱学森是中国系统工程研究的主要奠基人.**系统工程的步骤**：1)明确问题 2)设置目标,建立评价准则 3）方案的生成与未来环境预测．预测方法：定性预测，定量预测．4）把方案建模 5）对方案进行评价 6）选择一个方案 7）规划实施．系统工程的主要模块：系统分析，系统建模，系统决策,系统评价.**定量预测**：因果关系分析法(一元线性回归，多元线性回归，非线性回归)；时间序列分析(移动平均，指数平滑，ARIMA(自回归移动平均模型)).**定性分析**:专家预测法，德尔菲法.

模型的分类：结构模型，数量模型是按“模型描述的内容“分类的，数量模型按变量随时间的取值可分为连续模型和离散模型．按模型的本质，可以分类为具体模型和抽象模型；按系统输入输出情况可以分类为确定型模型，概率型模型，模糊型模型．按系统所处状态可以分为静态模型和动态模型．预测 COVID-19 确诊病例数：离散模型，动态模型．春游出发时间：建立概率型模型最合适．用电量预测：这是一个离散模型，这不是一个结构模型，这不是一个模糊型模型，这是一个时间序列分析问题，可以用 ARIMA 模型求解.铁路旅客发送量：离散模型，动态模型.系工导各章知识:结构模型,抽象模型.

Alabama 悖论：Hamilton 分配方法：先确定整数名额，再根据小数部分的大小顺序分配剩余名额，但是扩充董事会会导致丙公司代表减少．定义不公平相对差值测度$\dfrac{\frac{\text{甲州总人口}}{\text{甲州席位数}}-\frac{\text{乙州总人口}}{\text{乙州席位数}}}{\frac{\text{乙州总人口}}{\text{乙州席位数}}}$，Hauntington-Hill 分配方法：若分配一个新增席位，应该使下述比值达到最大的州 $\frac{(x_i)^2}{y_i(y_i+1)}$.两种极端的不恰当的数学模型：完全反映问题，模型无法求解；模型很好求解，严重歪曲问题．成功地应用系统工程方法基本上等价于在上述两种极端情况中找到恰当的折中.数学建模是主要的系统工程建模方法

第二章 系统建模

背景：系统由要素构成，要素之间存在逻辑关系，并可以用一定的数学模型描述．要了解系统中各要素之间的关系，需要建立系统的结构模型．结构模型：使用有向连接图来描述系统各要素间的关系，以表示一个作为要素集合体的系统的模型．**结构模型**是一种几何模型。结构模型使用由节点和有向边构成的图来描述一个系统的结构。节点：系统要素；有向边：要素之间的关系；结构模型是一种以定性分析为主的模型.**图的基本概念**：有向连接图：指由若干节点和有向边连接而成的图形．其中节点的集合是 S，有向边的集合是 E．树：没有回路的连通图就是树.关联树：在节点上带有加权值 W，而在边上有关联值 r 的树称作关联树．**邻接矩阵**：用来描述图中各节点两两之间的关系．邻接矩阵 A 的元素 a_{ij} 表示为，若 S_i 与 S_j 有关系则 $a_{ij}=1$，反之为 0．注意顺序，有向边从 S_i 到 S_j 则 $a_{ij}=1$.矩阵 A 的元素全为零的行所对应的节点称为汇点，即只有有向边进入而没有离开该节点。如 S1；矩阵 A 的元素全为零的列所对应的节点称为源点，即只有有向边离开而没有进入该节点。如 S4；对应每一节点的行中，其元素值为 1 的数量，就是离开该节点的有向边数；对应每一节点的列中，其元素值为 1 的数量，就是进入该节点的有向边数.



可达矩阵：用矩阵形式来描述有向连接图各节点之间，经过一定长度的通路后可以到达的程度．矩阵运算：逻辑乘取小，逻辑加(取大).A^2的元素为 1，相应变量间有二次通道；为 0 的话则无二次通道．结论：n 个变量的邻接矩阵 A，当 k 大于或等于 n 后，A^k 的非对角线上不会有首次为 1 的元素。所以，n 个变量的有向图，若两个变量间没有 1，2，…，$n-1$ 次通道，它们之间就不会有通道。所以，研究变量间有无通道，只需看 $A,A^2,\cdots,A^{n-1}$．故有向图的可达矩阵：$R=I+A+A^2+\cdots+A^{n-1}$，且由于单位矩阵运算的性质，有 $R=(I+A)^{n-1}$．而如果有 $m<n-1$ 满足 $(I+A)^m=(I+A)^{m+1}$，则：$R=(I+A)^m$.

$$\underset{A}{\begin{bmatrix}0&1&0&0\\0&0&1&1\\0&0&0&1\\1&0&0&0\end{bmatrix}}\times\underset{A}{\begin{bmatrix}0&1&0&0\\0&0&1&1\\0&0&0&1\\1&0&0&0\end{bmatrix}}=\underset{A^2}{\begin{bmatrix}0&0&1&1\\1&0&0&1\\1&0&0&0\\0&1&0&0\end{bmatrix}}$$

ISM 问题是由美国 John Warfield 教授开发的，不能解决定量化建模问题，一个系统可以由一个有向连接图表示,不是一种动态结构化技术．把复杂的系统分解为若干子系统(要素)最终将系统构造成一个多级递阶的结构模型．乒乓球，围棋不适合用解释性结构建模方法进行排序，跑步，铁饼适合

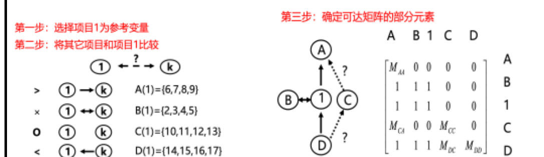
ISM 问题一般提法：给定一组变量，一组满足传递性的有向关系，要求完全表示其相互关系的骨架图．确定骨架图的步骤：1）确定邻接矩阵 2)计算可达矩阵 3）做层次划分 4）确定骨架图．层次划分：若变量是“叶子节点”，所有变量都指向它，则是顶层变量；如果一个变量没有指向它的变量，则是最底层的变量．利用以下规则就可以确定骨架图：1)同层变量或者互通或者不通（根据可达矩阵判断）2）每层变量仅指向**相邻的**上层变量（根据可达矩阵判断）3）每层变量不指向下层变量．求骨架图也就是在反复求顶层变量．顶层变量特征：1）不达到其他变量 2)如能达到某个变量，则该变量也能达到它．结论：变量 i 是顶层变量当且仅当其满足 $E(i)\subset F(i)$，其中，$E(i)$ 表示变量 i 能达到的变量的集合，$F(i)$ 表示能达到变量 i 的变量的集合．用邻接矩阵不能直接确定骨架图.可以用逐次求底层变量的方法构建骨架图．错误：在骨架图中，最优方案一定在底层.

$$\begin{array}{c|c} & \begin{matrix}1&2&3&4&5&6&7&8\end{matrix}\\ \begin{matrix}1\\2\\3\\4\\5\\6\\7\\8\end{matrix} & \begin{bmatrix}1&1&1&0&1&1&0&1\\0&1&1&0&0&0&0&1\\0&1&1&0&0&0&0&1\\1&1&1&1&1&1&0&1\\0&0&0&0&1&0&0&0\\1&1&1&0&1&1&0&1\\1&1&1&1&1&1&1&1\\0&1&1&0&0&0&0&1\end{bmatrix}\end{array}$$

否

$E(1)=\{1,2,3,5,6,8\}$，$F(1)=\{1,4,6,7\}$ 否；$E(2)=\{2,3,8\}$，$F(2)=\{1,2,3,4,6,7,8\}$ 是．确定了 2，3，5，8 是顶层变量，所以就去掉了.然后把 1，4，6，7 四个变量及其可达矩阵摘出来，发现 1，6 是顶层变量，4，7 不是．最后 2，3，5，8 是顶层，1，6 是二层，4 是三层，7 是四层.

确定骨架图：基本步骤：1）选择参考变量 2）将所有变量逐个和参考变量比较 3）考虑间接影响 4）对所有变量分类 5）以分析方法确定骨架图．任务，建立 17 个目标的结构模型，变量：17 个目标项目，关系：A 不比 B 差(也就是有向连接)，任务：确定项目相对优劣．第一步，选择项目 1 为参考变量，第二步：将其它项目和项目 1 比较．第四步：确定对角块．第五步：确定非对角块：把对角块画出来，然后分别比较，最终获得骨架图和可达矩阵.



黑箱：不清楚物理结构，或结构过于复杂；不了解机理规律，或机理过于复杂．黑箱建模：根据观测的输入输出数据，寻找规律，建立数学模型．黑箱建模（曲面拟合，回归)方法：选择由待定参数决定的一类函数 $f(x\mid\theta)$，获取样本数据 $x(t),y(t),t=1,2,\cdots$，拟合样本数据 $\min_\theta\sum_t\big(y(t)-f(x(t)\mid\theta)\big)^2$，获得经验模型 $y\approx f(x\mid\hat\theta)$．关键问题:选择什么函数类.

黑箱建模方法的基本指标：黑箱模型对原系统的逼近能力，参数是否便于估计，拟合效果如何，预测效果如何，模型是否好用．模型的好用性与模型的

预测效果是等价的，模型的逼近能力与模型的拟合性是等价的.错误：预测效果都好的模型一定是一个好的黑箱模型．正确：最小二乘法能使用的前提是残差白噪声．正确:多项式模型一定具有很好的逼近效果．正确：黑箱建模问题的难点在于找到拟合效果和预测效果的平衡．错误：一般的多项式模型能满足黑箱建模的三项指标．

多项式逼近：逼近能力：对任意阶可导函数，由泰勒定理保证．对连续函数，由魏尔斯特拉斯定理保证．结论：对任何连续函数，存在可以和其任意靠近的多项式函数序列．多项式逼近，容易确定模型参数．采用最小二乘方法（Least Square）估计参数 $\hat{\theta}=(\Phi\Phi^T)^{-1}\Phi Y^T$.观测值 $x_1(t),x_2(t),\cdots,x_n(t),y(t)$，逼近多项式$f(x\mid\theta)=a_0+a_1x_1+\cdots+a_nx_n+b_{11}x_1^2+b_{12}x_1x_2+\cdots+b_{nn}x_n^2$，待确定参数$\theta=[a_0,a_1,\cdots,a_n,b_{11},b_{12},\cdots,b_{nn},c_{111},c_{112},\cdots,c_{nnn},\cdots]^T_{n\times1}$，向量表示 $\phi(t)=[1x_1(t)\cdots x_n(t)x_1(t)^2x_1(t)x_2(t)\cdots]^T_{n\times1}$．其中向量偏导数$\frac{\partial\left(F(\theta)G^T(\theta)\right)}{\partial\theta}=\frac{\partial F(\theta)}{\partial\theta}G^T(\theta)+F(\theta)\frac{\partial G^T(\theta)}{\partial\theta}$．多项式逼近的本质：只在一点研究问题，通过不断分析该点各阶变化趋势逼近任意远处的函数值，但是某些点的偏差可能被大幅度放大．泰勒展开$f(x)=\sum_{n=0}^{\infty}a_n(x-x_0)^n$，$a_n=\frac{f^{(n)}(x_0)}{n!}$．求导时行向量与列向量相合，行行与列列不相合．错误：利用泰勒展开获得的多项式基函数是局部基函数．

基函数方法：限制基函数起作用的区域，用局部基函数代替全局基函数．辐射基函数 RBF$\rho(x\mid\beta(k))=\eta\left(\frac{x-c_k}{\delta_k}\right)$，高斯 RBF $\rho(x\mid\beta(k))=\exp\left(\frac{-\|x-c_k\|^2}{\delta_k^2}\right)$．岭函数$\rho(x\mid\beta(k))=\sigma(w_k^Tx+d_k)$，Sigmoid 函数$\sigma(u)=\frac{1}{1+\exp(-u)}$．两类常用人工神经网络：高斯 Radial Basis 神经网络 $\sum_k\alpha_k\exp\left(\frac{-\|x-c_k\|^2}{\delta_k^2}\right)$；Sigmoid 神经网络 $\sum_k\alpha_k\frac{1}{1+\exp(w_k^Tx+d_k)}$．逼近能力：都能逼近任意连续函数；模型参数：可用基于导数的非线性规划算法，部分参数可用最小二乘公式．有效克服多项式函数的缺陷．1D：可选用 hstep (x) 作基函数，它只在 I 区间内起作用，在其他区间不起作用，是局部基函数。辐射基函数类神经网络：用基函数显著大于 0 的部分 $\sum_k\alpha_k\eta\left(\frac{x-c_k}{\delta_k}\right)\approx\sum_k g(z_k)hstep_{I_k}(x)$，岭函数类神经网络：用基函数接近 1 的部分，所需要的节点数目 $M=m^n$，$\sum_k\alpha_k\sigma(w_kx+d_k)\approx g(z_1)+\sum_k\left(g(z_k)-g(z_{k-1})\right)\sigma(w_kx+d_k)$．错误：step 函数和 hstep 函数是局部基函数．正确：证明 RBFNN 逼近效果是利用了其基函数与超钟型函数的相似性．

回归分析方法：从一组数据出发，确定因变量和自变量之间的关系式；对关系式中的参数进行估计，并进行假设检验；筛选自变量，找出影响显著的，剔除不显著的；用求得的回归模型做预测；对预测结果进行分析，评价．线性回归问题对优化问题 $\min_{\theta\in R^n}\sum_{t=1}^N\left(y(t)-\theta^Tx(t)\right)^2$，如果存在 $(XX^T)^{-1}$，最优解为 $\hat{\theta}=(XX^T)^{-1}XY^T$.

一元线性回归：$y=a+bx+\epsilon$，则有$\hat{a}=\bar{y}-\hat{b}\bar{x}$，$\hat{b}=\frac{\sum X_iY_i}{\sum X_i^2}=\frac{L_{xy}}{L_{xx}}$，正则方程：$\frac{\partial(\sum e_i^2)}{\partial a}=-2\sum(y_i-a-bx_i)=0$，$\frac{\partial(\sum e_i^2)}{\partial b}=-2\sum x_i(y_i-a-bx_i)=0$．回归方程检验：相关系数分解法：因为 $L_{yy}=\sum_{i=1}^N(y_i-\bar{y})^2=\sum_{i=1}^Ny_i^2-N\bar{y}^2=\sum_{i=1}^N(\hat{y}_i-\bar{y})^2+\sum_{i=1}^N(y_i-\hat{y}_i)^2$，记作总平方和 TSS = 解释平方和 ESS + 剩余平方和 RSS，定义 $r^2=\frac{ESS}{TSS}=\frac{\sum_{i=1}^N(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^N(y_i-\bar{y})^2}$，表示总平方和中由回归解释了的部分，最小二乘法将使这个部分达到最大.，$r=\pm\sqrt{r^2}$ 称为相关系数，r 符号与 b 相同．相关系数为 0，说明建立的一元线性回归模型无效．

也可以运用**假设检验方法**刻画回归方程的线性因果关系，构造统计量 $t=\frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$，设 r 是总体 (x,y) 的相关系数，当假设 $H_0:r=0$ 成立时，统计量 t 服从自由度（degree of freedom）为 $N-2$ 的 t 分布．当 $t>t_\alpha$ 时，否定原假设（null hypothesis），认为 x 与 y 存在线性关系．F 检验法：在假设 $H_0:b=0$ 成立时，TSS，ESS，RSS 分别是自由度为 $f_T=N-1,f_E=1,f_r=N-2$ 的 χ^2 变量，并且 RSS 与 ESS 相互独立，于是统计量 $F=\frac{ESS/f_E}{RSS/f_R}=\frac{(N-2)ESS}{RSS}$ 服从自由度为 $(1,N-2)$ 的 F 分布．当 $F>F_\alpha$ 时，否定原假设，认为 x 与 y 存在线性关系．

精度分析：设 S_δ 为 y 的剩余均方差，它表示变量 y 偏离回归直线的误差 $S_\sigma=\sqrt{\frac{\sum_{i=1}^N(y_i-\hat{y})^2}{N-2}}=\sqrt{\frac{(1-r^2)L_{yy}}{N-2}}$．给定显著性水平 α，对某一 x_0，相应的 y_0 将以 $(1-\alpha)$ 的概率落在下述区间（称为置信区间）．式中，$\hat{y}_0$ 是对应于 x_0 的 y_0 的预测值，$Z_{\alpha/2}$ 是标准正态分布上 $\alpha/2$ 百分位点的值．

一元线性回归步骤：数据平移和归一化压缩变换，得到变量 $x'=\frac{x-c_1}{d_1},y'=\frac{y-c_2}{d_2}$，计算新变量的系数 $\hat{b}'=L_{x'y'}/L_{x'x'}$，$\hat{a}'=\bar{y}'-\hat{b}'\bar{x}'$，带回原变量 $\frac{y-c_2}{d_2}=\hat{a}'+\hat{b}'\frac{x-c_1}{d_1}$，进行假设检验$ESS'=\hat{b}'^2L_{x'x'}=\hat{b}'L_{x'y'}$，$RSS'=L_{y'y'}-ESS'$，$F'=\frac{(N-2)ESS}{RSS}$，求置信区间，对回归直线进行预测$S_\delta'=\sqrt{\frac{RSS'}{N-2}}$，$S_\delta=d_2S_\delta'$.

一元非线性回归：函数变换线性化方法；多项式变换线性化方法；分段线性化方法；直接非线形回归分析方法(最小二乘准则，非线性规划)．

多元线性回归：$y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n$，定义以下向量和矩阵：$\mathbf{Y}=[y_1\quad y_2\quad\cdots\quad y_N]$，$\mathbf{X}=\begin{bmatrix}1&1&\cdots&1\\x_{11}&x_{21}&\cdots&x_{N1}\\x_{12}&x_{22}&\cdots&x_{N2}\\\vdots&\vdots&\ddots&\vdots\\x_{1n}&x_{2n}&\cdots&x_{Nn}\end{bmatrix}$，$\boldsymbol{\beta}=\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\\\vdots\\\beta_n\end{bmatrix}$，$\boldsymbol{\varepsilon}=[\boldsymbol{\varepsilon}_1\quad\boldsymbol{\varepsilon}_2\quad\cdots\quad\boldsymbol{\varepsilon}_N]$，则回归方程及回归预测误差可表示为 $\mathbf{Y}=\boldsymbol{\beta}^T\mathbf{X}+\boldsymbol{\varepsilon}$，其中 $\hat{\boldsymbol{\beta}}=(\mathbf{XX}^T)^{-1}(\mathbf{XY}^T)$．又有 $\mathbf{A}=\begin{bmatrix}N&\sum_{i=1}^Nx_{i1}&\sum_{i=1}^Nx_{i2}&\cdots&\sum_{i=1}^Nx_{in}\\\sum_{i=1}^Nx_{i1}&\sum_{i=1}^Nx_{i1}^2&\sum_{i=1}^Nx_{i1}x_{i2}&\cdots&\sum_{i=1}^Nx_{i1}x_{in}\\\vdots&\vdots&\vdots&\ddots&\vdots\\\sum_{i=1}^Nx_{in}&\sum_{i=1}^Nx_{i1}x_{in}&\sum_{i=1}^Nx_{i2}x_{in}&\cdots&\sum_{i=1}^Nx_{in}^2\end{bmatrix}$，$\mathbf{B}=\begin{bmatrix}\sum_{i=1}^Ny_i\\\sum_{i=1}^Nx_{i1}y_i\\\vdots\\\sum_{i=1}^Nx_{in}y_i\end{bmatrix}$，有$\mathbf{A}=\mathbf{XX}^T$，$\mathbf{B}=\mathbf{XY}^T$.

有数据预处理的方法：$y_i=\mu_0+\beta_1(x_{i1}-\bar{x}_1)+\beta_2(x_{i2}-\bar{x}_2)+\beta_3(x_{i3}-\bar{x}_3)+\varepsilon_i$，$i=1,2,\cdots,49$，则方程变化为 $l_{11}=\sum_{i=1}^{49}x_{i1}^2-\frac{1}{49}\left(\sum_{i=1}^{49}x_{i1}\right)^2$，$l_{21}=l_{12}=\sum_{i=1}^{49}x_{i1}x_{i2}-\frac{1}{49}\left(\sum_{i=1}^{49}x_{i1}\right)\left(\sum_{i=1}^{49}x_{i2}\right)$，$l_{1y}=\sum_{i=1}^{49}x_{i1}y_i-\frac{1}{49}\left(\sum_{i=1}^{49}x_{i1}\right)\left(\sum_{i=1}^{49}y_i\right)$，则$\mathbf{A}=\begin{bmatrix}N&0\\0&\mathbf{L}\end{bmatrix}$可得.

显著性检验：同样有：TSS=ESS+RSS．在假设 $H_0:\beta_1=\beta_2=\cdots=\beta_n=0$ 成立时，统计量 $F=\frac{ESS/f_E}{RSS/f_R}=\frac{(N-n-1)\cdot ESS}{n\cdot RSS}$ 服从自由度为 $(n,N-n-1)$ 的 F 分布.当 $F>F_\alpha$ 时，否定原假设，认为 x 与 y 存在线性关系．$TSS=\sum_{i=1}^n(y_i-\bar{y})^2$，$ESS=\sum_{i=1}^n(\hat{y}_i-\bar{y})^2$，$RSS=\sum_{i=1}^n(y_i-\hat{y}_i)^2$．可以用 $S_\delta=\sqrt{\frac{RSS}{N-n-1}}$，预测值 $\hat{y}$ 将以 $(1-\alpha)$ 的概率落在下述区域内，即 $(\hat{y}_0-Z_{\alpha/2}S_\delta,\hat{y}_0+Z_{\alpha/2}S_\delta)$.

病态线性回归问题．产生原因：样本数据中回归变量间严格线性相关！XX^T 的秩不会大于 B_m 的秩，没有逆矩阵！先求最大无关向量，再变换得到结果.确定满足$B_m^TB_m=I_m$，$x(t)=B_mz(t),1\le t\le N$，的 $B_m,z(t),1\le t\le N$ 和最小的 m,．由于$y(t)\approx c^Tx(t)=c^TB_mz(t)=d^Tz(t)$，先估计$\hat{d}=(ZZ^T)^{-1}ZY^T$，再利用$z(t)=B_m^TB_mz(t)=B_m^Tx(t)$.

方法：从 n 递减寻找符合要求的 m．严格病态线性回归问题的处理方法：从 $\hat{m}=n-1$ 开始逐渐减少 $\hat{m}$，依次求解，$\min\sum^N\left(x(t)-Lv(t)\right)^T\left(x(t)-Lv(t)\right)$，s. t. $L^TL=I_{\hat{m}},v(t)\in R^{\hat{m}}$，找到使最优目标值等于零的最小 m，其对应的最优解就可用作所需要的B_m 和 $z(t),1\le t\le N$．对于严格病态回归问题，可以通过多次求解优化问题来确定独立变量个数 m．错误：上述优化问题目标函数一直远大于零，说明对应的 m 还不是我们要找的 m．因为一旦 $\hat{m}<m$，下述优化问题的最优目标值一定大于零.正确:只要优化问题目标函数一直等于 0,说明对应的 m 还不是我们要找的 m.正确：严格病态回归问题，估计参数的公式就不能使用了.求解严格病态回归问题的思路是先建立 Y 与 Z(由线性无关的变量构成)之间的回归方程,再利用 Z 与 X 的关系,建立 Y 与 X 的方程．X=BZ,其中 Z 是 X 在 B 组成的空间上的投影.

实际病态线性回归：$XX^T\approx BZZ^TB^T$接近奇异，其逆矩阵即使存在，参数估计值$\hat{c}=(XX^T)^{-1}XY^T$也很不可靠．将实对称矩阵XX^T正交对角化，XX^T接近奇异，本质上就是某些特征根远比其它特征根小.类似严格病态线性回归的处理方法，优化问题的最优目标值为$\sum_{i=m+1}^n\lambda_i$，则处理病态线性回归问题的基本方法是找出使逼近误差$\sum_{i=m+1}^n\lambda_i$可以接受的最小正整数m,确定$Q_m=[q(1)q(2)\cdots q(m)]$，$Z=Q_m^TX$，$\hat{d}=(ZZ^T)^{-1}ZY^T$，$\hat{c}=Q_m\hat{d}$，得到$y\approx\hat{d}^Tz=\hat{d}^TQ_m^Tx=\hat{c}^Tx$．实际病态问题理论分析：参数估计误差为$-\Lambda_m^{-1}Z\mu^T$，特征根趋近于 0 的话倒数趋近无穷大，线性回归误差被严重放大．注意$(ZZ^T)^{-1}=\Lambda_m^{-1}$，$\hat{d}=(ZZ^T)^{-1}ZY^T=\Lambda_m^{-1}Q_m^TXY^T$．理论：若$A$为 n 阶实对称矩阵，则 A 的特征根皆为实数，R^n中属于 A 的不同特征值的特征向量必正交，存在正交矩阵 C，使得 $C^{-1}AC$ 为对角阵．矩阵特征值之和等于矩阵的迹．行列式等于所有特征值的乘积.如果 n 阶矩阵 A 的秩小于 n，则 A 的行列式等于 0.

规范化措施：样本数据规范化$\bar{x}_i(t)=\frac{x_i(t)-e(x_i)}{\sqrt{\delta^2(x_i)}}$，其中$e(x_i)=\frac{1}{N}\sum_{t=1}^Nx_i(t)$，$\delta^2(x_i)=\frac{1}{N-1}\sum_{t=1}^N\left(x_i(t)-e(x_i)\right)^2$，归一化的具体作用是统一样本的统计分布特性．一般来说,线性回归不一定要做归一化，但此时比较各回归系数就不那么方便了．规范化：相对误差$\frac{\sum_{t=1}^N\left(x(t)-Q_mQ_m^Tx(t)\right)^T\left(x(t)-Q_mQ_m^Tx(t)\right)}{\sum_{t=1}^Nx^T(t)x(t)}=\frac{\sum_{i=m+1}^n\lambda_i}{\sum_{i=1}^n\lambda_i}$

交通流量预测．对时序的研究主要基于两种方法，一种是用随机过程的理论建立线性关系模型，如回归模型，ARIMA 模型；另外一种方法是利用非线性动力学方法，研究低自由度的混沌系统．预测方法举例：时空自回归滑动平均求和模型

STARIMA(Pfeifer 在 1980 年提出)，多变量自适应回归样条模型 MARS(Jerome Freidman 1991). 移动平均法：$M_t^{(1)} = \frac{1}{N}(y_t + y_{t-1} + \cdots + y_{t-N+1})$，增量形式：$M_t^{(1)} = M_{t-1}^{(1)} + \frac{1}{N}(y_t - y_{t-N})$，指数平滑法 $\hat{x}_h(1) = \alpha \sum_{j=0}^{\infty}(1-\alpha)^j x_{h-j}$，增量形式$\hat{x}_h(1) = \alpha x_h + (1-\alpha)\hat{x}_{h-1}(1)$.

第三章 系统分析

主成分分析：将坐标做平移和旋转变换，使得新坐标的原点与样本数据群的重心重合，第一主轴与数据变异最大的方向对应. 数据分类时数据分散更利于划分. 该问题的最优解$\hat{y}_1(t) = \hat{l}_1(1)\tilde{x}_1(t) + \hat{l}_2(1)\tilde{x}_2(t) + \hat{l}_3(1)\tilde{x}_3(t)$就是这组样本数据的第一主成分. 一般情况下，给定一组样本数据，首先求出规格化的数据，确定 m 个主成分的优化模型为 $\max\sum_{t=1}^{N}\sum_{k=1}^{m}\left(y_k(t)\right)^2$，s. t. $y_k(t) = \sum_{i=1}^{n} l_i(k)\tilde{x}_i(t)$, $k = 1,2,\cdots m \le n$，$\sum_{i=1}^{n}\left(l_i(k)\right)^2 = 1$, $k = 1,2,\cdots m$，$\sum_{i=1}^{n} l_i(k) l_i(j) = 0$, $\forall k \ne j$. 问题转化为$\max\sum_{k=1}^{m} l^T(k)\tilde{X}\tilde{X}^T l(k)$，st$L^T L = I_m$. 用$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$表示$\tilde{X}\tilde{X}^T$ 的顺序递减的特征根，$q(1), q(2), \cdots, q(n)$ 是它们对应的规范化的特征向量，则所求主成分为$\hat{y}_k(t) = q^T(k)\tilde{x}(t), k = 1,2,\cdots m$. **主成分样本均值**$e(\hat{y}_k) = \sum_{i=1}^{n} q_i(k) e(\tilde{x}_i) = 0, \forall k$，主成分样本方差$\delta^2(\hat{y}_k) = \lambda_k/(N-1), \forall k$，主成分样本方差之和$\sum_{k=1}^{n}\delta^2(\hat{y}_k) = n$，因为 $(N-1)\sum_{k=1}^{n}\delta^2(\hat{y}_k) = \sum_{k=1}^{n}\sum_{t=1}^{N}\tilde{x}^T(t) q(k) q^T(k)\tilde{x}(t)$.

分类变量个数选择准则，选择最小的 $\hat{m}$ 作为 m，满足$\frac{\sum_{k=1}^{\hat{m}}\delta^2(\hat{y}_k)}{\sum_{k=1}^{n}\delta^2(\hat{y}_k)} = \frac{1}{n}\sum_{k=1}^{\hat{m}}\delta^2(\hat{y}_k) \ge \gamma$. **数据压缩问题**：给定一组样本数据，首先求出其规格化的数据，求解优化问题$\min\sum_{t=1}^{N}\left(\tilde{x}(t) - Ly(t)\right)^T\left(\tilde{x}(t) - Ly(t)\right)$，假定 L 列满秩且$L^T L = I_m$，则等价于求解 $\max\sum_{k=1}^{m} l^T(k)\tilde{X}\tilde{X}^T l(k)$, st$L^T L = I_m$，注意下标是 m，最优压缩变量是$\hat{y}_k(t) = q^T(k)\tilde{x}(t), k = 1,2,\cdots m$，它就是前 m 个主成分. 相对逼近误差：$\frac{\sum_{t=1}^{N}\tilde{x}^T(t)\tilde{x}(t) - \sum_{t=1}^{N}\tilde{x}^T(t)\hat{L}\hat{L}^T\tilde{x}(t)}{\sum_{t=1}^{N}\tilde{x}^T(t)\tilde{x}(t)} = \frac{\sum_{i=m+1}^{n}\lambda_i}{\sum_{i=1}^{n}\lambda_i} = \frac{\sum_{k=m+1}^{n}\delta^2(\hat{y}_k)}{n}$. 主成分压缩数据所需要存储的数据个数：假设样本数据是 n 维向量，共有 N 个，我们提取出了 m 个主成分，则需要存储数据$m \times N + m \times n + n + n$. 第一部分是在 m 个主成分方向上的投影，第二部分是 m 个主成分向量，n 是样本方差，n 是样本均值. **主成分数据压缩计算过程**：对样本数据进行归一化；计算归一化样本数据协方差矩阵XX^T，计算XX^T的特征值和特征向量；对特征值进行排序，选取前 m 个特征值所对应的特征向量作为主成分方向；计算各样本数据在主成分方向上的投影 $y(t) = [q(1), q(2), \cdots, q(m)]^T x(t)$，计算压缩率$\eta = \frac{m\times N + m\times n + n + n}{n \times N}$.

PCA 和病态线性回归的联系：二者原理相同，病态回归消除线性相关，PCA 找到最为关键的少数综合变量，与原系统数据保持了较高的一致性.

聚类分析：聚类问题实际上是将包含若干元素的集合，按照某种测度，划分成若干子类。测度是指定义在每个类上的函数，我们的目标就是使其达到最大或最小. 聚类问题的本质是划分问题：属于 NP 难问题. **向量聚类**：测度的定义是类内所有点与其中心点距离之和. **变量聚类**：r 是两个变量的协方差，希望类内协方差的值越大越好，即类内两个变量的夹角越小越好，类间：希望整体分类结果相关性最强最好，角度之和最小.

K 均值：考虑向量聚类问题$\min_{\Omega}\sum_{i=1}^{k}\sum_{t\in\varpi_i}\left(x(t) - e_{\varpi_i}(x)\right)^T\left(x(t) - e_{\varpi_i}(x)\right)$，其中$e_{\varpi}(x) = \frac{1}{|\varpi|}\sum_{t\in\varpi} x(t)$. 步骤：首先确定分类数目 k，然后在所有样本中挑选 k 个作为初始中心点，逐个利用每个样本修改中心点. 对所有样本：顺序进行下述计算：将其归入与其最近的中心点所在的类，重新计算该类的重心，并用新的重心替换中心点. **优点**：算法简单，快速，易于实现；聚类结果容易解释，适用于高维数据的聚类；实验发现，当各个类的分布近似为高斯分布时，效果较好. **不足**：K 均值为贪婪策略，可能陷入局部最优，大规模数据集上求解效率低；对离群点和噪声点敏感；不同初始点选取可能会导致不同的聚类结果；K 值选择较为困难；不适于发现非凸形状或者大小差别很大的聚类. **改进**：Kmeans++：使初始的聚类中心点相互之间的距离尽可能远！一定收敛，不能找到全局最优解.

基于密度的聚类方法：DBSCAN，OPTICS，DENCLUE. DBSCAN 向量聚类：密度：领域内样本点个数；核心点：密度不小于 MinPots；边界点：不为核心点，但领域内有核心点；噪音点：不为核心点也不为边界点. **步骤**：初始化参数，确定核心点集合，寻找核心点队列集合，噪音点确定. 优点：可以对任意形状的稠密数据集进行聚类；可以在聚类的同时发现噪音点；不需要事先指定类别数目；聚类结果不依赖节点的遍历顺序. 不足：数据集过大时收敛时间长；聚类质量依赖于距离公式的选取；密度差异较大时，超参 ϵ，MinPots 选取较为困难.

系统聚类方法：变量聚类问题. 是一种贪婪算法. 基本步骤：1）首先将每个变量视为一类，得到 n 类变量 2）每次选择最相关的两个类合并，顺序得到 $n-1, n-2, n-3, \cdots$ 直至一类变量 3）记录合并过程生成聚类谱系图 4）设定阈值，根据聚类谱系图决定最终分类. $\rho = \min R\left(x_i, x_j\right)$.

动态聚类方法：考虑密度最大的几个点作为初始中心点，把距离某个中心点最近的点划归该中心点对应的类，用每类的重心替换中心点. 如果当前的中心点和重心点都很接近，停止分类.

SOM 网络定义了一种非线性映射，通常 SOM 是一个有监督的两层神经网络，其中第二层可以是二维矩形点阵或六边形.

因子分析：N 个同学考试，考试科目为 n，成绩记为$x(t) = [x_1(t)x_2(t)\cdots x_n(t)]^T 1 \le t \le N$，若有$m < n$种公共能力，则考试成绩可分解为$x_i(t) = \sum_{k=1}^{m} a_i(k) f_k(t) + s_i(t)$，其中$A_m = [a(1)a(2)\cdots a(m)]_{n\times m}$为载荷矩阵，$x(t) = A_m f(t) + s(t)$. 公共因子和特殊因子正交. 样本数据规格化，公共因子规格化，$e(f_k) = \frac{1}{N}\sum_{t=1}^{N} f_k(t) = 0$，$\delta^2(f_k) = \frac{1}{N-1}\sum_{t=1}^{N}\left(f_k(t)\right)^2 = 1$，因子之间线性无关. 满足条件：$1_N F^T = 0, 1_N S^T = 0, SF^T = 0, \frac{1}{N-1}FF^T = I_m$，$\frac{1}{N-1}SS^T = \mathrm{diag}\{\phi_1, \phi_2, \cdots \phi_n\}$. 由样本相关矩阵$R_X = A_m A_m^T + \Phi$. 可以用最小二乘估计，但是是假设白噪声. 因子得分 $\hat{f}(t) = (A_m \Phi^{-1} A_m^T)^{-1}\Phi^{-1} A_m x(t)^T$. Φ 是对角矩阵. **基于主成分的因子分析**：利用前 m 个主成分建立近视的 m-因子模型.

第四章 系统决策

20 世纪 70 年代提出了“决策分析系统（DSS）”概念. **决策的过程**：1）情报：收集信息；2)设计：形成候选方案 3)抉择：根据不同标准排序，寻优 4）实施：观察决策效果，收集反馈信息. **西蒙**：管理就是决策. 若存在公认的最好方案，此时不需要决策. 但如果面临多目标，不确定因素的情况，方案好坏因人而异，此时才需要决策. 决策分析就是对特定的备选项进行的系统评估. **决策分析**是一种标准的方法，而不是一种描述方法. 它展示了为最大限度的达成目标，决策者如何运用一系列制定决策的逻辑规则. 决策是分配资源的过程，结果是决策的效果. 如果决策在制定时是最佳选择，这个决策就是一个好决策；如果决策者达到了预期目标，我们说他得到了理想的结果. **决策分析的基本依据**：决策者对不同决策后果的主观偏好：决策者对所有可能的决策后果存在，合理的主观偏好. **决策分析的重点和难点**：如何有效地获取决策者的主观偏好. **决策环境**：以方案选择为主要内容的决策过程也随环境不同而有很大差别，决策环境处于完全可以预测和极难预测两种情况之间. **决策环境**可以归纳为三种类型：确定型：未来环境完全可以预测(然而包含多个目标：多目标决策问题)；风险型：未来环境有几种可能的状态和相应的后果，可以观测每种状态和后果出现的概率(概率已知：风险型决策问题)；不确定型：未来环境完全不可预测(概率未知：不确定型决策问题)

期望值法：若采用决策目标（准则）是期望收益最大，则选择收益期望值最大的行动方案为最优方案. **决策树法**：绘制决策树，计算各行动方案的益损期望值，将计算所得的各种行动方案的益损期望值加以比较，选择其中最大的期望值对应的方案为最优方案. **灵敏度分析**：对状态概率的估计往往不一定十分准确. 当状态概率发生变化时，会对所选最优方案产生怎样的影响？或者当状态概率在何范围内变化时，所选最优方案不变？

决策所需情报的种类：完全情报，即完全可以肯定某一状态发生的情报；非完全情报（或称抽样情报），即不能完全肯定某一状态发生的情报. 在决策分析过程中，如果得不到完全情报，或者采集完全情报所花代价太大，则可以采用非完全情报作为补充信息对原来的状态概率进行修正. 原来的状态概率称为：先验概率. 修正后的状态概率称为：后验概率. 贝叶斯决策就是用来估计由于获得了非完全情报而提高决策的效果（即情报的价值）的方法. $p(\theta_i \mid B) = \frac{p(\theta_i)p(B|\theta_i)}{\sum_{j=1}^{n} p(\theta_j)p(B|\theta_j)}$.

DSS 是在计算机用于管理的过程中产生的. DSS 是一种能够帮助决策者利用数据和模型，解决半结构化的以计算机为基础的交互作用系统. 西蒙（1960)年提出：决策问题分为结构化和半结构化两大类. 结构化问题是指在决策过程开始前能够准确识别，可用计算机实现全部自动化求解的问题. 对于结构化问题，管理者只关心决策的效率. 半结构化问题至今没有统一定义. Stabell 于 1979 年提出了半结构化问题的特征：1）目标不明确且为非操作的，或目标可操作，但目标多且相互矛盾；2）事后难于确定决策效益变化的原因，事前也难预测决策者采取措施对于决策效益的影响；3）决策者采取什么措施会影响决策效益是不确定的. 对于半结构化问题，管理者关心决策的效能(首先保证决策合理，然后寻求提高效率). DSS 问题管理的信息结构包括：决策任务、决策问题、求解方案、求解结果、决策结果和实施结果等. DSS 数据管理一般采用关系型数据模型. DSS 模型管理包括结构模型、关系模型和基于知识表示模型的管理. DSS 知识管理包括知识获取、修改、删除、求精和一致性检验等.

冲突分析：冲突分析是研究冲突现象的数学理论和方法，运用数学模型来描述冲突现象. 冲突分析是决策论的一个分支. 冲突模型的基本要素：1)决策人（局中人）：具有独立决策权的参与者 2)行动：可供局中人选择的动作 3)策略：一组可行的完整行动方案；策略空间：所有策略的集合. 4)结局：当所有局中人都选择了某一个策略后，冲突所形成的结果.

风险决策问题：决策后果 $g(s \mid a), s \in S, a \in A$，给定决策方案后不同状态发生的概率 $\hat{p}(s \mid a), s \in S, a \in A$，决策后果集$C = \{g(s \mid a), s \in S, a \in A\}$. 不同决策方案所产生的差异仅在于所有后果发生的概率不一样. 对于风险决策问题，决策者的偏好本质上

是对不同的概率向量 $p=[p_1\quad p_2\cdots\quad p_k]^T$ 的偏好，其中 p_i 是后果 i 出现的概率．我们将这种向量称为展望（Prospect），将所有展望组成的集合称为展望集．风险决策分析的关键就是如何确定决策者对展望集中不同元素的偏好．解决风险决策分析问题的基本思路:设法在展望集上定义一个实函数（效用函数）$u(p),\forall p\in P$．合理的偏好应该满足：1)连通性：展望集中的任意两个元素之间存在明确的优劣关系 2)传递性 3)单调性（复合传递性）4)连续性（有限优越性）．**关于有限优越性的争论**．**效用函数**：结论：如果决策者对风险决策问题的偏好满足以上四条假定，则一定存在具有以下性质的效用函数 $\hat{u}(p),\forall p\in P$，一致性，线性，正线性变换下的唯一性．规范化的效用函数：$u(e(1))=0$，$u(e(k))=1$．一般情况下不会根据期望效益最大决策．：圣彼得堡悖论．**决策者对风险的态度**:冒险型决策者．保守型决策者，中立型决策者．$\lambda_i=\frac{x_i-x_1}{x_k-x_1}$．风险态度本质上反映了在不同基础上对增加单位收入的感受！保守：对增加单位收入的满意程度递减；冒险：对增加单位收入的满意程度递增．$\mathbf{v(x_i)=u(e(i))}$，其中$e(i)$代表以概率 1 获得 x_i 元．

引入 e(i)的作用是利用效用函数的线性性质，确定效用函数的任务可以简化为确定以概率 1 发生每个结果的展望的效用值的问题，具体含义是表示后果 ci 以概率 1 发生的展望．

有限理性原则：Simon，1947．**前景理论**：人在面临获利时不愿冒风险；而在面临损失时，人人都成了冒险家．1)大多数人在面临获利的时候是风险规避的（确定效应）2）大多数人在面临损失的时候是风险喜好的（反射效应）3）大多数人对得失的判断往往根据参考点决定（参照依赖）4)大多数人对损失比对收益更敏感（损失效应）

典型决策准则：平均准则：$\max_{a\in A}\sum_{s\in S}v(g(s\mid a))$ 有利于小概率结果；悲观(保守)准则:$\max_{a\in A}\left(\min_{s\in S}v(g(s\mid a))\right)$，乐观冒险准则：$\max_{a\in A}\left(\max_{s\in S}v(g(s\mid a))\right)$，折衷准则：$\max_{a\in A}\left(\alpha\min_{s\in S}v(g(s\mid a))+(1-\alpha)\max_{s\in S}v(g(s\mid a))\right)$；极小化最大后悔值准则：$r(s\mid a)=\max_{\hat{a}\in A}v(g(s\mid\hat{a}))-v(g(s\mid a))$，$R(a)=\max_{s\in S}r(s\mid a)$，决策准则$\min_{a\in A}R(a)$．

钱的非线性效用：买卖点的工作点不一样．不会成交．一个保守的人做买方时，是一个保守型决策者，在做卖方时，是一个冒险型决策者．

群决策分析：社会选择问题：群由 m 个成员组成，决策问题有 n 个方案可供选择，每个成员对这组方案有自己的偏好顺序，如何确定群的偏好顺序？**几种常用的选择规则**：1）简单多数（Plurality）：每个成员选一个方案，得票最多的获胜 2）绝对多数(Majority)：每个成员选一个方案，得票选过 50%的获胜，如果没有获胜方案，选择得票多的重新投票 3）加权投票：每个成员对不同方案给出不同的分值，总得分最多的方案获胜．例子：Borda 规则：每个成员对其最骗好的方案给 $n-1$ 分，第二偏好的 $n-2$ 分，如此类推，总分最高的获胜．4）批准投票：每个成员列出其认可的方案（不限数目），得到最多成员认可的方案获胜．批准投票可以看成是一种特殊的加权投票方法，每个成员对其认可的方案给 1 分，不认可的给 0 分．结论：在群的每个成员偏好不变的情况下群的选择结果强烈地依赖于选择规则！**合理的选择规则**：1)公理 1：连通性．简单多数规则满足公理 1；2）公理 2：传递性．简单多数规则可能违背公理 2 3）条件 1：完全域．简单多数规则满足条件 1 4）条件 2：无关方案的独立性．简单多数规则满足条件 2 5）条件 3：群偏好和成员偏好的正的联系．即：群中成员对某些方案偏好顺序一样，这群（整体)对这些方案偏好顺序一样．简单多数规则满足条件 3．Borda 规则可能违背条件 3 6）条件 4：Pareto 原则．简单多数规则满足条件 4．7)条件 5：非独裁性．简单多数规则满足条件 5.**Arrow 的不可能定理**：没有一个群的选择规则能够同时满足前面的两个公理和五个条件．简单多数票规则满足公理 1（ 连通性）；条件 1（ 完全域）;条件 2（无关方案的独立性）；条件 3（群偏好和成员偏好的正的联系）；条件 4（ Pareto 原则）；条件 5（ 非独裁性）．简单多数票规则不满足公理 2（ 传递性）．一定可以设计一种偏好断面，使得一个人成决定性子群，出现独裁?不可能定理的根本原因:序数效用的局限性．

多目标决策问题：决策变量 $x=[x_1\quad x_2\quad\cdots\quad x_n]$，可行集 $S\subseteq R^n$，目标函数 $f(x)=[f_1(x)f_2(x)\cdots f_m(x)]^T$，寻找 $\hat{x}\in S$，满足不存在 $x\in S$ 使得 $f(x)>f(\hat{x})$，其中 $>$ 表示决策者的优先关系．假设每个目标都是成本型目标，即越小越好，如果存在 $x\in S$ 满足 $f_i(x)\le f_i(\hat{x})$，$\forall 1\le i\le m$，并且至少有一个目标，比如 $f_k(x)$ 满足 $f_k(x)<f_k(\hat{x})$，那么 $\hat{x}$ 肯定不是所求决策．上述 $\hat{x}$ 被称为劣解，不是劣解的就叫非劣解，有效解，或 Pareto 解．多目标决策实际上是在有效解集中进行决策．**有效解**：负直角锥和可行目标集只交于一点；**弱有效解**：负直角锥内部和可行目标集不交；**劣解**：负直角锥内部和可行目标集相交．**权重是主观的**

可根据如何获取偏好信息对不同方法分类：1）在优化之前：先获取所需要的偏好信息，再进行优化；较易实现，但适用范围小 2）在优化之中：获取偏好信息和优化过程交互进行（DSS）；较难实现，但适用范围大 3）在优化之后：先优化产生足够的有效解，再进行决策；适用范围很小

确定情况下多目标加权方法：1，同等重要，3，稍重要，5，重要，7，重要得多，9，极端重要．判断矩阵．一致性条件 $a_{ij}=\frac{w_i}{w_j}=\frac{1}{\frac{w_j}{w_i}}=\frac{1}{a_{ji}}$ 容易保证，$a_{ik}a_{kj}=\frac{w_i}{w_k}\times\frac{w_k}{w_j}=a_{ij}$ 难以保证．总结：如果判断矩阵能表示为一组权系数之比，任意求得判断矩阵的最大特征根所对应的一个特征向量，再将其规范化，就可以唯一确定权向量．总结：如果判断矩阵的最大特征根和 m 的差比较小，就可以用其规范化的特征向量做权向量，否则不行．最后要解决的问题：$\lambda_{\max}-m$ 大到什么程度不接受．$C.R.(m)=\frac{1}{N}\sum_{t=1}^{N}\frac{\lambda_{\max}(t)-m}{m-1}$，具体的判断矩阵的 C.I.（$A\mid m)=\frac{\lambda_{\max}-m}{m-1}$，若认为 C.I.（$A\mid m)<0.1C.R.(m)$，可认为一致性满足要求．总结前面讨论的确定权系数的方法：1）决策者对目标两两比较给出判断矩阵 2）计算判断矩阵的最大特征根 $\lambda_{\max}$ 3）如果 $\frac{\lambda_{\max}-m}{m-1}<0.1C.R.(m)$，对最大特征根的特征向量进行规范化得到权向量，否则决策者应改进判断矩阵．目的规划法：基本思想：用目标期望值反映决策偏好，通过逼近目标期望值获得决策者满意的方案．逐步进行法．第一步求理想点，第二步确定各目标权系数，第三步按照加权悲观原则求非劣解，第四步停止或修改偏好．

第五章 系统评价

System Analysis：由美国 RAND 公司最早于 20 世纪 40 年代提出，早期用于武器系统的成本和效益分析，采用定量分析．70 年代左右，推广到更广泛的领域，常常与制定政策相关 80 年代后，特别针对信息系统建设的中系统分析方法应用广泛：结构法，原型法，面向对象，构件法．**定义**：广义：等同于系统工程;狭义：通过一系列步骤，帮助领导者选择最优方案的一种系统方法;是实现科学决策的重要工具．**需要考虑的要素**:目标,可行方案,费用,模型,效果,准则,结论．**系统评价过程需遵循的原则**:内部因素与外部因素相结合;近期与远期利益相结合；局部效益与总体效益相结合;定性分析与定量分析相结合．

层次分析法 AHP：起源：20 世纪 70 年代由美国 Saaty 教授提出;特点：定性与定量分析相结合；适用：不能完全用数学模型表示的多目标，多准则，群决策问题；方法：问题分层，因素权重分析，方案排序，一致性检验等整套方法．;应用：80 年代初期介绍到中国，在工程技术，社会科学领域应用较广泛．**步骤**：Step1：将问题按照决策要求进行层次分解，得到决策层次 Decision Hierarchy；Step2：采用两两比较 pairwise comparison 方法得到各决策元素值；Step3：构造判断矩阵 judgments matrix 对决策元素值进行一致性检验；若判断不一致，返回 Step2，重新进行两两比较；若满足一致性，进入 Step4；Step4：计算决策表的相对权重 weights；Step5：归一化处理相对权重值，并得到各方案的分数值及排序情况 scores and hence rankings

判断矩阵：若矩阵 $A=(a_{ij})_{n\times n}$ 满足 $\forall i,j=1,2,\cdots,n,a_{ij}>0$，则称其为正的，如果 $\forall i,j=1,2,\cdots,n,a_{ji}=\frac{1}{a_{ij}}$，则称其为互反的．如果矩阵 $A=(a_{ij})_{n\times n}$ 是正的，互反的，且元素以 scale [1,9] 取值，则称 A 为判断矩阵．**如何由判断矩阵计算出权重？**Saaty 提出特征向量方法 Eigenvector Method（EM）．设 $\lambda_i(i=1,2,\cdots,n)$ 是判断矩阵 A 的特征值，即$Aw=\lambda_i w,w\ne 0$，设 $\lambda_{\max}=\max_i(\lambda_i)$，那么，如下向量 w 就是我们所希望的权重向量，$Aw=\lambda_{\max}w$，$w\in D$，权重向量就是最大特征值对应的规范特征向量．．**“However, the validity of EM has never been fully proved.”** —— Sekitani，Yamaki(1999)．如何得到各项分数值？对每一个因素,或再分解后的下一级因素,对不同方案进行两两比较,得到各个判断矩阵．采用 EM 方法,对每一项因素分别求解最大特征值,特征向量，归一化处理，得到权重向量．

一致性检验：满足以下条件的矩阵 $A=(a_{ij})_{n\times n}$ 是一致的$a_{ij}=a_{ik}a_{kj}$．定义判断矩阵 A 的一致性指标 consistency index（C.I.)如下：C.I. ＝ $(\lambda_{\max}-n)/(n-1)$．A 的一致性程度 consistency rate（C.R.）定义为：C.R. ＝ C.I./ R.I.．若 C.R.小于 0.1，则认为一致．可证明结论$\lambda_{\max}\ge n$，因为$\lambda_{\max}=\sum_{j=1}^{n}a_{ij}\left(\frac{x_j}{x_i}\right)$．一致的判断矩阵秩为 1．

步骤：Step1 建立层次结构模型．Step2：构造判断矩阵．Step3：层次单排序．Step4 层次总排序 Step5：一致性检验．

矩阵特征值求解：$\det(\lambda I-A)=0$求解特征值，再代入$(\lambda I-A)v=0$．秩 1 矩阵的特征值：$\lambda_1=\lambda_2=\cdots\lambda_{n-1}=0$，也可得$\lambda_n=\sum_{i=1}^{n}a_{ii}$．秩一矩阵$A=ab'$，列向量，所有和 b 正交的向量都是 A 的特征值为 0 的特征向量．另外$Aa=ab'a=a(b'a)=(b'a)a$，所以 a 是特征值为 b‘a 的特征向量．

钱学森是工程控制论创始人，汪应洛严广乐编写教材，Daniel Kacheman 和 Amos Tversky 提出前景理论，Herbert Simon 提出前景理论．