第五章　系统预测

第一节　系统预测概述

系统预测概述

系统预测是系统工程理论的重要组成部分，它把系统作为预测对象：

- 分析系统发展变化的规律性；
- 预测系统未来发展变化趋势；
- 为系统规划设计、经营管理和决测提供科学依据。

系统预测：根据系统发展变化的实际情况和历史数据、资料，运用现代的科学理论方法，以及对系统的各种经验、判断和知识，对系统在未来一段时期内的可能变化情况，进行推测、统计和分析，并得到有价值的系统预测结论。

预测的作用

科学的预测能为决策者提供可靠的依据。

科学的预测可以为决策者提供多种方案。

科学的预测为决策者广泛的集中了智力财富。

系统预测的分类

1.按**预测领域**分类：可分为社会预测、经济预测、科学预测、技术预测和军事预测等。

2.按**预测范围**分类：可分为宏观预测和微观预测。

3.按**预测期限**分类

近期预测　　3个月以下；　短期预测　　3个月至1年；

中期预测　　1年至5年；　　长期预测　　5年以上。

4.按**预测方法特征**分类：可分为定性预测和定量预测。

系统预测的方法

1.**定性预测方法** 依据人们对系统发展变化规律的把握、判断，用经验和直觉作出预测，如专家打分、主观评价、市场调查。常见的定性预测方法是Delphi法。

2.**因果关系预测方法** 以若干系统变量为分析对象，以样本数据为分析基础，建立系统变量之间因果数学模型。根据因果数学模型，预测某些系统变量的变化对其它系统变量的定量影响。因果关系预测方法主要有回归分析预测、状态空间预测等。

3.**时间序列分析预测方法** 主要考察系统变量随时间变化的定量关系，给出系统的演变发展规律，并对未来作出预测。

第一类为**定性方法**，第二、三类为**定量方法**。

系统预测的步骤

（1）确定预测目标

（2）选择预测方案

（3）收集、整理资料和数据

（4）建立预测模型

（5）利用模型预测

（6）预测结果分析

第二节 DELPHI定性预测方法

德尔菲（Delphi)定性预测方法

系统预测的步骤：

（1）确定预测目标；（2）选择预测方案；（3）收集、整理资料和数据 ；

（4）建立预测模型

（5）利用模型预测

选择熟悉与所预测问题相关的领域的专家10到15人，采用通信往来的方式与专家们建立联系，将预测问题的目标和任务告诉专家并提供所掌握的初步资料和数据，然后将专家关于预测分析的意见进行整理、综合、归纳，再以第一轮预测结果的形式匿名反映至各位专家进行第二轮征求意见，供他们分析判断，提出新的论证结果。如此经过多轮反复论证调查，各专家的意见逐轮趋向一致，结论的可信性也大大增加。

（6）预测结果分析

德尔菲（Delphi)定性预测方法

德尔菲法是专家会议调查法的一种发展，但其效果往往比专家会议法要好，其原因在于:

1）克服了专家会议中常见的随大流或由某些权威一言定调的缺点，专家与专家之间互不见面，各专家的意见完全独立作出，消除了心理影响；

2）为保证预测结果的可靠性，德尔菲法一般要经过多轮，并且在下一轮开始时让专家充分参考本轮甚至以前各轮的预测分析结果，不仅能保证过程和结果的客观性和公正性，而且能逐渐消弥个别固持已见的结论；

3）德尔菲法本质上是一类定性预测的方法，但其非常注重定性向定量的转化工作，使方法同时具有定性和定量的优点。

德尔菲预测法–平均数法

一种将定性预测结果转化为定量指标的方法：对预测中的每个因素给定一个分值。设有N_j个专家对可选答案j（对应分值为C_j）投了赞成票，则某轮总体预测的分值的均值和方差分别为

$$E=\frac{\sum_j C_j N_j}{\sum_j N_j},D=\frac{\sum_j (C_j-E)^2 N_j}{\sum_j N_j}$$

预测结果分析

预测结果统计分析

- 一般采用平均数或中位数统计出量化结果。
- 如果数据分布的偏态比较大，一般使用中位数以免受个别偏大偏小的判断值的影响；
- 如果数据分布的偏态比较小，一般使用平均数，以便考虑到每个判断值的影响。

预测结果分析–中位数法

常用中位数和上、下四分数来处理专家们的答案，求出预测的期望值和区间。 当有n个（包括重复）专家意见时，其由小到大的排列次序为：

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$$

求得全距 $= x_{(n)} - x_{(1)}$ ，即最大值与最小值之差。

确定中位数和上、下四分数的位置：

$$x_{中}的位置 = \frac{n+1}{2} \quad x_{下}的位置 = \frac{[x_{中}位置]+1}{2} \quad x_{上}的位置\, x_{下}位置 + [x_{中}位置]$$

其中：方括号[X]表示括号内取X的最大整数。

预测结果分析-中位数法

依据中位数及上下四分数的位置确定其相应的数值。

预测结果为：中位数表示专家组对预测的期望值；全距表示预测值的最大变动幅度（专家预测值分散程度的一种度量）；上、下四分数的区间构成预测区间，根据正态分布理论可知，有50%以上专家的预测值落在预测区间之内。

第三节 一元线性回归分析预测

一元线性回归分析预测

实际系统中，存在着大量这样的情况：

两个变量（例如x和y）彼此有一些依赖关系，由x可以部分地决定y的值（称两者具有因果关系）。然而，这种决定往往不很准确。

一般身材较高的人体重也较重，但这种因果关系是因人而异的，即仅由某一人的身高并不能准确知道他的体重。尽管如此，基于“人的身高与体重有一定联系”的认识，我们可以搜集足够多的人的身高体重数据，经过统计分析，可以发现身高和体重间服从某种统计规律。

利用统计方法研究这种因果关系的方法就是回归分析，它主要处理连续型随机变量之间的相关关系，进行某种预测分析等等。

一元线性回归预测

一元线性回归方程的模型可以表示为：

$$y_t = \alpha + \beta x_t + \varepsilon_t \qquad t = 1,2,\cdots,n$$

一元线性回归预测模型为：

$$\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$$

一元线性回归预测就是利用解释变量预期的已知值或预测值对预测期或样本以外的被解释变量作定量估计。

一元线性回归原理

设有个过程，其输入量为x，输出量为y，所观测到的数据为x_1,x_2，…,x_L,和$y_1,y_2,\ldots y_L$。假设两者是线性关系，则希望能得出下面方程组1。现在的问题是从L个方程中求出α和β的最佳估计值 $\hat{a}$、$\hat{\beta}$ 由于观测数据中混有噪声，故不能利用上述方程组求出α和β。解决的办法是假定 $\hat{a}$ 对 $\hat{\beta}$ 各组观测数据xi,yi都是有误差的，可列出方程组2：

$$\begin{cases} x_1\beta + a = y_1 \\ x_2\beta + a = y_2 \\ \qquad \cdots \\ x_L\beta + a = y_L \end{cases}$$

$$\begin{cases} x_1\hat{\beta} + \hat{a} + e_1 = y_1 \\ x_2\hat{\beta} + \hat{a} + e_2 = y_2 \\ \qquad \cdots \\ x_L\hat{\beta} + \hat{a} + e_L = y_L \end{cases}$$

一元线性回归原理-随机误差

也称为偶然误差/不定误差，是由于在测定过程中一系列有关因素微小的随机波动而形成的具有相互抵偿性的误差。其产生的原因是分析过程中种种不稳定随机因素的影响,如室温、气压等环境条件的不稳定,分析人员操作的微小差异以及仪器的不稳定等。随机误差的大小和正负都不固定,但多次测量就会发现,绝对值相同的正负随机误差出现的概率大致相等,因此它们之间常能互相抵消,所以可以通过增加平行测定的次数取平均值的办法减小随机误差。

一元线性回归原理

- 整个观测过程的误差是由各次观测误差所组成的，用怎样的方法表示整个观测过程的误差？

$$J=\sum e_i=e_1+e_2+\ldots+e_L$$

$$J=\sum|e_i|=|e_1|+|e_2|+\ldots+|e_L|$$

$$J=\sum e_i^{\ 2}=e_1^{\ 2}+e_2^{\ 2}+\ldots+e_L^{\ 2}$$

- 最后的函数J就是估计参数所用的准则函数，希望准则函数J越小越好，由于准则函数现在平方运算－最小二乘估计法。

一元线性回归原理

- **根据数学分析中寻求极值的原理，要使J达到最小，只需对α和β求偏导数，并另其为零。**

$$\begin{cases} \dfrac{\partial J}{\partial a} = -2\sum_{i=1}^{L}(y_i - \hat{a} - x_i\hat{\beta}) = 0 \\ \dfrac{\partial J}{\partial \beta} = -2\sum_{i=1}^{L}(y_i - \hat{a} - x_i\hat{\beta})x_i = 0 \end{cases}$$

$$\begin{cases} \hat{a} = \overline{y} - \hat{\beta}\overline{x} \\ \hat{\beta} = \dfrac{\sum_{i=1}^{L}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{L}(x_i - x)^2} \end{cases}$$

- **于是得到模型：**

$$y = \hat{a} + \hat{\beta}x$$

回归方程的检验

- 问题是：该方程是否符合X于Y之间的客观规律，或所拟合出来的直线在多大程度上反映出X于Y间的真实因果关系?
- 需要建立一个能反映变量X与Y之间线性关系密切程度的统计量，以便进行统计检验。

相关系数检验法

对于变量y,样本yi与均值之差的平方和可以分解为两部分：

$$L_{yy}=Q+U$$

$$L_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2$$

$$Q=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

$$U=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=\frac{L_{xy}^2}{L_{xx}}$$

$$L_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_iy_i-n\overline{xy}$$

$$L_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})=\sum_{i=1}^{n}x_i^2-n\bar{x}^2$$

Lyy反映了数据yi的总体波动情况，Q是样本值yi与回归值之差的平方和，它反映了样本偏离回归直线的程度；U是回归值与平均值之差的平方和，它反映了x与y线性关系的密切程度。

相关系数检验法

- 数据yi的总体波动Lyy由两部分引起，一方面是由x的变化所引起y的波动U，称为回归平方和；另一方面是其他偶然因素引起Y的波动Q，称为剩余平方和。
- 对于给定的一组样本数据，如果点（x,y)的分布很接近于一条直线，则Q值相对于U值较小，反之则较大。
- 定义：变量x与变量y的相关系数r(介于-1和1之间）

$$|r|=\sqrt{1-\frac{Q}{L_{xy}}}$$

相关系数检验法

- 构造统计量t,

$$t = \frac{\sqrt{N-2r}}{\sqrt{1-r^2}}$$

- 定理：设r是总体（x,y)的相关系数，但假设H0:r=0 成立时，统计量t服从自由度为n-2的t分布。
- 对于给定的显著性水平a（alpha)以及自由度（n-2），查t分布表得出相应的临界值ta，从而对H0进行假设检验，即
- 当t>ta时，否定原假设，认为x与y存在线性关系；否则，接受原假设，认为二者不存在线性关系。
- 另外还有一种检验方法:F检验法。

相关系数检验法

- 如果不存在线性关系，则 b=0,只需检验是否b=0
- 构造统计量F,

$$F=\frac{U/f_U}{Q/f_Q}=\frac{(N-2)U}{Q}$$

- 定理：假设H0:b=0 成立时，Lyy，U，Q分别是自由度为N-1,1和N-2的χ2变量(卡方），且Q和U相对独立，于是统计量F服从自由度为（fU,fQ)的F分布。
- 对于给定的显著性水平a（alpha)以及自由度（1,n-2），查F分布表得出相应的临界值Fa，从而对H0进行假设检验，即
- 当F>Fa时，否定原假设，认为x与y存在线性关系；否则，接受原假设，认为二者不存在线性关系

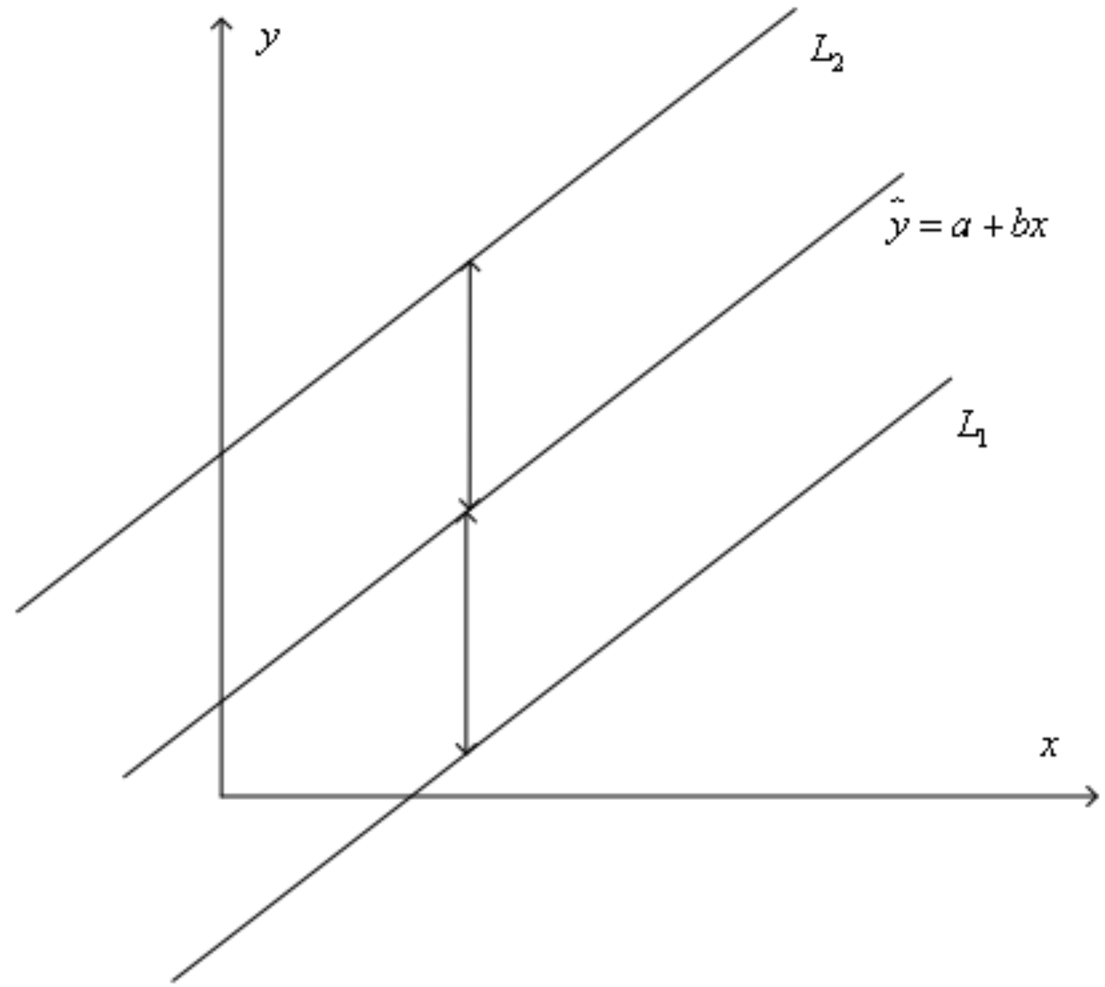
一元线性回归预测的精度分析

- 预测精度分析是衡量一元线性回归方程有效性的一个重要指标，它直接影响预测模型的取舍。--采样区间估计方法，给定置信度a，求出预测变量所取的可能取值范围。
- 根据区间，说明全部可能出现的观察值y中，大约有（1-a)*100%的概率落在该区间的范围内。

$$S_\delta = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{(1-r^2)L_{yy}}{n-2}}$$

$$(\hat{y}_0 - Z_{a/2}S_\delta,\ \hat{y}_0 + Z_{a/2}S_\delta)$$

一元线性回归预测的精度分析



L_1与L_2越靠近，预测精度越高

一元线性回归预测的步骤

① 数据线性变换（为方便计算考虑）-类似归一化；

② 计算线性回归

③ 得到回归方程

④ 进行统计检验 判断线性模型是否成立（查分布表，临界值）；

⑤ 如线性模型成立，求出S值，查出正态分布上a/2百分位点值。

⑥ 给出预测区间。

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$L_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \qquad L_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

$$L_{yy} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

$$\hat{b} = L_{xy} / L_{xx} \qquad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

$$Q = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$U = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \frac{L_{xy}^2}{L_{xx}}$$

$$S = \sqrt{Q/(n-2)}$$

应用实例

编号	成分A(x)	成分B(y)
1	0.009	4
2	0.013	3.44
3	0.006	3.6
4	0.025	1
5	0.022	2.04
6	0.007	4.74
7	0.036	0.6
8	0.014	1.7
9	0.016	2.92
10	0.014	4.8
11	0.016	3.28
12	0.012	4.16
13	0.02	3.35
14	0.018	2.2

一元非线性回归分析预测

一元非线性回归分析预测

线性关系是特殊情况，非线性是普遍存在的。

非线性回归的处理方法：

（1）将复杂的非线性通过某种数学变换转化为线性关系，然后采用线性回归分析方法；

（2）直接进行非线性回归分析的方法。

函数变换线性化

（1）根据机理或图形判断，选择一种易于线性化的非线性函数；

（2） 将自变量和因变量通过适当的数学变换得到线性关系；

（3）对变换后的自变量和因变量数据作线性回归分析；

（4）将得到的线性回归方程通过逆变换得到自变量和因变量间的非线性关系。

常用的易于线性化的非线性函数：

（1）幂指数（y=ax^b）；（2）指数函数(y=ab^x)；

（3）双曲函数(1/y=a+1/x)

相关指数

为判断得到的非线性方程是否有效，仿照检验线性回归方程的相关系数定义一个指标：相关系数

$$|R| = \sqrt{1 - \frac{Q}{L_{yy}}} = \sqrt{1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}}$$

其中，$\hat{y}_i$ 为因变量的y的回归值，R不是表示自变量与因变量间的线性相关性，也不是表示经过变换后的线性相关性，而是表示在总变差平方和中用非线性表达的份额。

多项式变换线性化

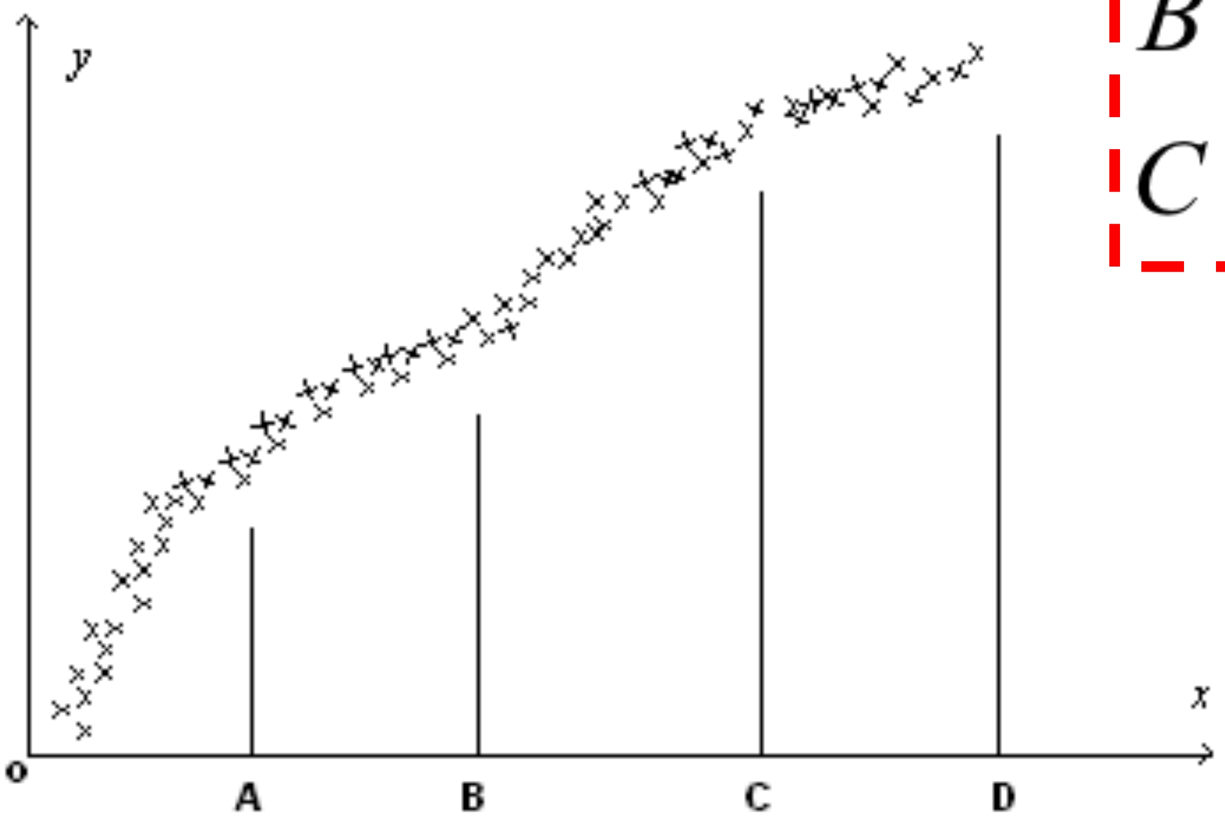
当已知变量间存在某种非线性函数关系，但这种非线性函数又无法确定时，运用非线性函数的函数逼近论原理，可以采用多项式近似模型。

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

$$y = a_0 + a_1 t_1 + a_2 t_2 + \ldots + a_n t_n$$

分段线性化

$$O-A: \quad y = a_0 + a_1 x$$
$$A-B: \quad y = b_0 + b_1 x$$
$$B-C: \quad y = c_0 + c_1 x$$
$$C-D: \quad y = d_0 + d_1 x$$



直接非线性回归方法

设变量 *x,y* 服从非线性关系 *y=f(x)*

$$\hat{y}_i = f(x_i, \hat{\theta}), \qquad \hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k]^T$$

这样非线性方程中存在k个未知参数，预测误差：

$$\varepsilon_i = y_i - \hat{y}_i = y_i - f(x_i, \hat{\theta})$$

$$J(\hat{\theta}) = \sum_i \omega_i \varepsilon_i^2$$

w表示样本的重要程度。采用最小二乘法，得到的参数就是非线性回归参数。

第四节 多元线性回归分析预测

多元线性回归分析预测

对因变量y有影响的自变量共有p（x1,x2,...,xp)个，而且他们都对因变量y是线性关系，可以用多元线性回归分析来研究他们间的关系。

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

对n个样本带入上式，则对应的预测值为：

$$\begin{aligned}
\hat{y}_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_p x_{1p} & \varepsilon_1 &= y_1 - \hat{y}_1 \\
\hat{y}_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_p x_{2p} & \varepsilon_2 &= y_2 - \hat{y}_2 \\
&\vdots & &\vdots \\
\hat{y}_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_p x_{np} & \varepsilon_n &= y_n - \hat{y}_n
\end{aligned}$$

多元线性回归分析预测

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad X=\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ & & \ddots & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$\beta^T=[\beta_0 \quad \beta_1 \quad \beta_2 \quad \cdots \quad \beta_p]$$

$$\varepsilon^T=[\varepsilon_0 \quad \varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_p]$$

可以表示成矩阵形式： $Y=X\beta+\varepsilon$

$$\min J(\hat{\beta})=\min\sum\varepsilon^2=\min\sum(yi-yi)^2$$

多元线性回归分析预测

$$A=\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \cdots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \cdots & \sum x_{i1}x_{ip} \\ & & \ddots & & \\ \sum x_{ip} & \sum x_{i1}x_{ip} & \sum x_{i2}x_{ip} & \cdots & \sum x_{ip}^2 \end{bmatrix}$$

$$B^T=[\sum y_i \quad \sum x_{i1}y_i \quad \sum x_{i2}y_i \quad \cdots \quad \sum x_{ip}y_i]$$

$$X=\begin{bmatrix} 1 & x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \dots & x_{1p}-\bar{x}_p \\ 1 & x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \dots & x_{2p}-\bar{x}_p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \dots & x_{np}-\bar{x}_p \end{bmatrix}$$

$$A=X^TX \qquad B=X^TY$$

$$(X^TX)\hat{\beta}=(X^TY)$$

$$\hat{\beta}=(X^TX)^{-1}(X^TY)$$

多元线性回归的计算方法

① **样本数据预处理，构成预测模型：**

$$y_i = \mu_0 + \beta_1(x_{i1} - \bar{x}_1) + \beta_2(x_{i2} - \bar{x}_2) + \dots + \beta_p(x_{ip} - \bar{x}_p) + \varepsilon_i, \qquad i = 1,2,\dots,n$$

$$\mu_0 = \beta_0 + \sum_{i=1}^{p} \beta_i \bar{x}_i$$

② **计算相应矩阵**

$$B = \begin{bmatrix} \sum y_i \\ l_{1y} \\ \vdots \\ l_{py} \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \dots & x_{1p} - \bar{x}_p \\ 1 & x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \dots & x_{np} - \bar{x}_p \end{bmatrix}$$

多元线性回归的计算方法

$$A=\begin{bmatrix} n & 0 & 0 & \cdots & 0 \\ 0 & l_{11} & l_{12} & \cdots & l_{1p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & l_{p1} & l_{p2} & \cdots & l_{pp} \end{bmatrix}=\begin{bmatrix} n & 0 \\ 0 & L \end{bmatrix}$$

$$l_{ij}=\sum_{k=1}^{n}x_{ki}x_{kj}-n\bar{x}_i\bar{x}_j, \qquad i,j=12,\ldots,p$$

$$l_{iy}=\sum_{j=1}^{n}x_{ij}y_i-n\bar{x}_i\bar{y}, \qquad i=12,\ldots,p$$

$$\boxed{\begin{aligned} A\mu &= B \\ \mu &= A^{-1}B \end{aligned}}$$

$$\boxed{\hat{\mu}^T=[\hat{\mu}_0 \quad \hat{\beta}_1 \quad \hat{\beta}_2 \quad \cdots \quad \hat{\beta}_p]}$$

多元线性回归的计算方法

$$\hat{\mu}^T = [\hat{\mu}_0 \quad \hat{\beta}_1 \quad \hat{\beta}_2 \quad \cdots \quad \hat{\beta}_p] = \begin{bmatrix} 1/n & 0 \\ 0 & L^{-1} \end{bmatrix} B$$

$$L = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1p} \\ l_{21} & l_{22} & \cdots & l_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pp} \end{bmatrix}$$

$$\hat{\mu}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

$$[\hat{\beta}_1 \quad \hat{\beta}_2 \quad \ldots \quad \hat{\beta}_p]^T = L^{-1}[l_{1y} \quad l_{2y} \quad \ldots \quad l_{py}]^T$$

多元线性回归的计算方法

原始的回归系统为：

$$\hat{\mu}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

$$\beta_0 = \bar{y} - \sum_{i=1}^{n} \hat{\beta}_i \bar{x}_i$$

$$[\hat{\beta}_1 \quad \hat{\beta}_2 \quad \dots \quad \hat{\beta}_p]^T = L^{-1}[l_{1y} \quad l_{2y} \quad \dots \quad l_{py}]^T$$

多元线性回归的显著性检验

- **对于给定的n个样本数据，总可以按照最小二乘来建立一个多元线性回归。**
- **Lyy=Q+U**

$$L_{yy} = Q + U$$

$$L_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$Q = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$U = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

多元线性回归的显著性检验

假设 $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$

成立时，L_{yy}，U，Q 分别是自由度 $f_y = n-1, f_U = p, f_Q = n-p-1$ 的 χ^2 变量，并且 Q 与 U 相互独立。

在假设 H₀ 成立时，统计量

$$F = \frac{U/p}{Q/(n-p-1)} = \frac{(n-p-1)U}{pQ}$$

服从自由度为 $(p, n-p-1)$ 的 F 分布，这样按给定的显著性水平 α 查 F 分布表，得到相应的临界值 F_α：

若 $F > F\alpha$，则否定原假设 H_0，认为 y 与 $x_1, x_2, \cdots x_p$ 之间存在线性关系；

若 $F \le F_\alpha$，则接受原假设 H_0 认为 y 与 $x_1, x_2, \cdots x_p$ 之间不存在线性关系。

多元线性回归的预测精度

对于多元线性回归的预测模型

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p$$

仿照一元线性回归问题的处理方法，可以用

$$S_\delta=\sqrt{\frac{Q}{n-p-1}}$$

近似地表示 y 偏离回归平面的误差，于是，可以预测在各变量 $x_1,x_2,\cdots x_p$ 取固定的样本值时，预测值 $\hat{y}$ 将以 $1-\alpha$ 的概率落在下述区域内：

$$(\hat{y}_0-Z_{\alpha/2}S_\delta,\hat{y}_0+Z_{\alpha/2}S_\delta)$$

其中，$\hat{y}_0$ 是采用 $\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2,\cdots\hat{\beta}_p$ 作为最佳回归参数时的预测值，$Z_{\alpha/2}$ 是标准正态分布上 $\alpha/2$ 百分位点的值。

多元偏线性回归

多元线性回归分析预测

对因变量y有影响的自变量共有p（x1,x2,…,xp)个，而且他们都对因变量y是线性关系，可以用多元线性回归分析来研究他们间的关系。

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

对n个样本带入上式，则对应的预测值为：

$$\begin{aligned}
\hat{y}_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_p x_{1p} & \varepsilon_1 &= y_1 - \hat{y}_1 \\
\hat{y}_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_p x_{2p} & \varepsilon_2 &= y_2 - \hat{y}_2 \\
&\vdots & &\vdots \\
\hat{y}_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_p x_{np} & \varepsilon_n &= y_n - \hat{y}_n
\end{aligned}$$

多元线性回归分析预测

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ & & \ddots & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$\beta^T = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \cdots \quad \beta_p]$$

$$\varepsilon^T = [\varepsilon_0 \quad \varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_p]$$

可以表示成矩阵形式： $Y = X\beta + \varepsilon$

$$\min J(\hat{\beta}) = \min \sum \varepsilon^2 = \min \sum (yi - yi)^2$$

多元线性回归分析预测

$$A=\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \cdots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \cdots & \sum x_{i1}x_{ip} \\ & & \ddots & & \\ \sum x_{ip} & \sum x_{i1}x_{ip} & \sum x_{i2}x_{ip} & \cdots & \sum x_{ip}^2 \end{bmatrix}$$

$$B^T=[\sum y_i \quad \sum x_{i1}y_i \quad \sum x_{i2}y_i \quad \cdots \quad \sum x_{ip}y_i]$$

$$X=\begin{bmatrix} 1 & x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \dots & x_{1p}-\bar{x}_p \\ 1 & x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \dots & x_{2p}-\bar{x}_p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \dots & x_{np}-\bar{x}_p \end{bmatrix}$$

$$\boxed{\begin{aligned} &A=X^TX \qquad B=X^TY \\ &(X^TX)\hat{\beta}=(X^TY) \\ &\hat{\beta}=(X^TX)^{-1}(X^TY) \end{aligned}}$$

多元线性偏回归分析

- **关于复共线问题**

$$A = X^T X \qquad B = X^T Y$$
$$(X^T X)\hat{\beta} = (X^T Y)$$
$$\hat{\beta} = (X^T X)^{-1}(X^T Y)$$

最小二乘估计为方差最小的线性无偏估计，在正态分布下是方差最小无偏的。

多元线性偏回归分析

- 实际情况下，该估计结果不理想，矩阵求逆存在问题。
- 如果X^TX的逆不存在，表明X的列向量是线性相关的，无法求解。
- 即使存在，如果其行列式很小，表明X列向量接近线性相关。该估计的性能也会变差。

$$A = X^T X \qquad B = X^T Y$$
$$(X^T X)\hat{\beta} = (X^T Y)$$
$$\hat{\beta} = (X^T X)^{-1}(X^T Y)$$

多元线性偏回归分析

假设已知x_1，x_2与y的关系服从线性回归模型

$$y=10+2x_1+3x_2+\varepsilon$$

	序号	1	2	3	4	5	6	7	8	9	10
(1)	x_1	1.1	1.4	1.7	1.7	1.8	1.8	1.9	2.0	2.3	2.4
(2)	x_2	1.1	1.5	1.8	1.7	1.9	1.8	1.8	2.1	2.4	2.5
(3)	ε_i	0.8	-0.5	0.4	-0.5	0.2	1.9	1.9	0.6	-1.5	-1.5
(4)	y_i	16.3	16.8	19.2	18.0	19.5	20.9	21.1	20.9	20.3	22.0

$\hat{\beta}_0=11.292$，$\hat{\beta}_1=11.307$，$\hat{\beta}_2=-6.591$

- **$rho_{12}=0.972$**

多元线性偏回归分析

- 一般来说，如果变量之间存在完全的线性关系，他们之间的的相关系数为1。

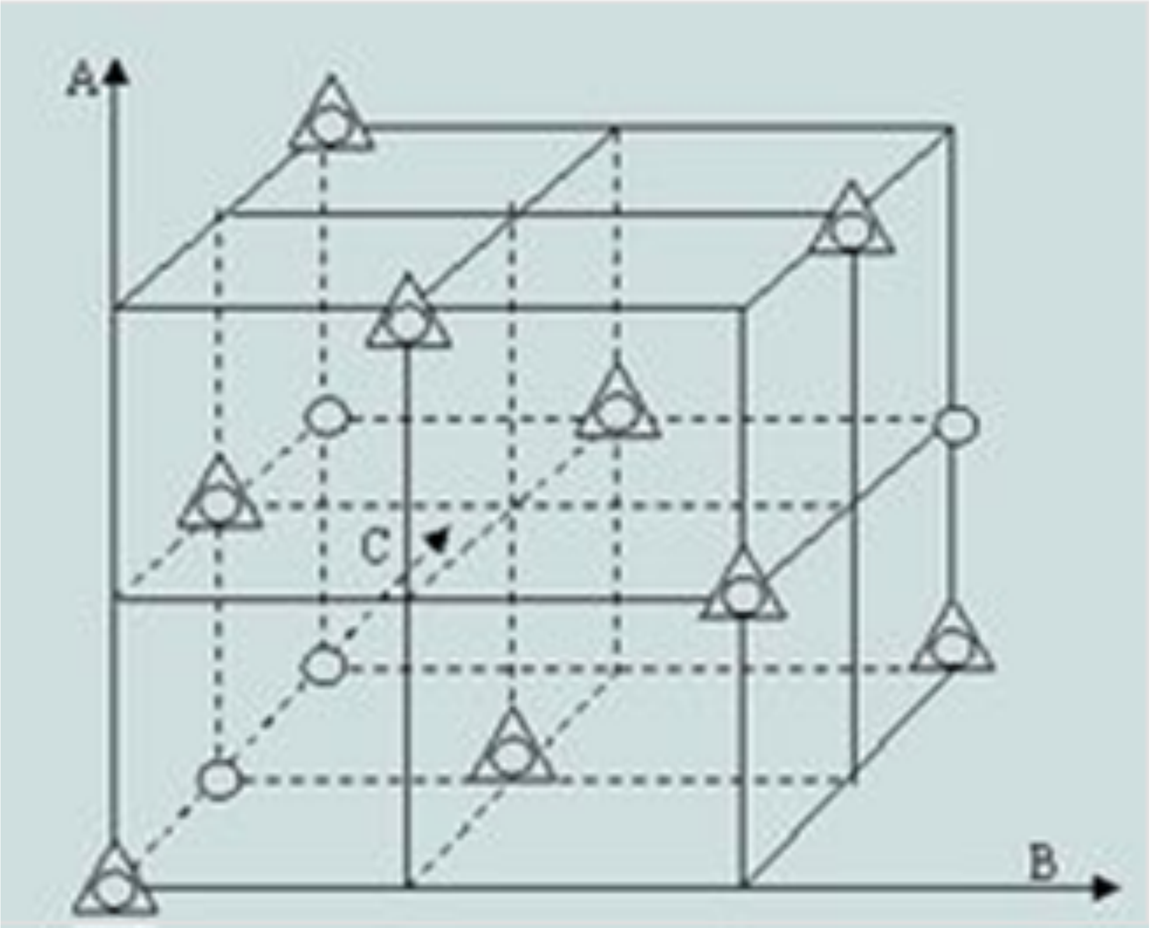
$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

多元线性偏回归分析

- 变量间存在复共线时，最小二乘估计性能下降，主要表现在以下几个方面：

1. 自变量完全相关时，回归系统无法确定；
2. 存在不完全的复共线时，估计的回归系数方差将随着自变量的相关程度迅速扩大；
3. 自变量高度相关的条件下，回归系数的估计值对样本数据的微小变化将变得非常敏感，估计值的稳定性将变得很差；
4. 回归模型的统计检验存在困难；
5. 模型的物理解释存在问题。

正交试验法



岭回归分析

- 岭回归分析是一种专用于共线性数据分析的有偏估计回归方法，实质上是一种改良的最小二乘估计法，通过放弃最小二乘法的无偏性，以损失部分信息、降低精度为代价获得回归系数更为符合实际、更可靠的回归方法，对病态数据的耐受性远远强于最小二乘法。

岭回归分析

- 岭回归的原理较为复杂。根据高斯马尔科夫定理，多重相关性并不影响最小二乘估计量的无偏性和最小方差性。
- 虽然最小二乘估计量在所有线性无偏估计量中是方差最小的，但是这个方差却不一定小。
- 实际上可以找一个有偏估计量，这个估计量虽然有微小的偏差，但它的精度却能够大大高于无偏的估计量。
- 岭回归分析就是依据这个原理，通过在正规方程中引入有偏常数而求得回归估计量的。

岭回归分析

- 当自变量间存在复共线性时，｜X′X｜≈0，设想给X′X加上一个正常数矩阵kI，（k＞0)，那么X′X+kI接近奇异的程度就会比X′X接近奇异的程度小得多。
- 考虑到变量的量纲问题，我们先对数据做标准化，为了记号方便，标准化后的设计阵仍然用X表示

$$\boxed{\hat{\beta}(k) = (X^T X + kI)^{-1} X^T Y}$$

- 因为岭参数k不是唯一确定的，所以我们得到的岭回归估计实际是回归参数β的一个估计族。

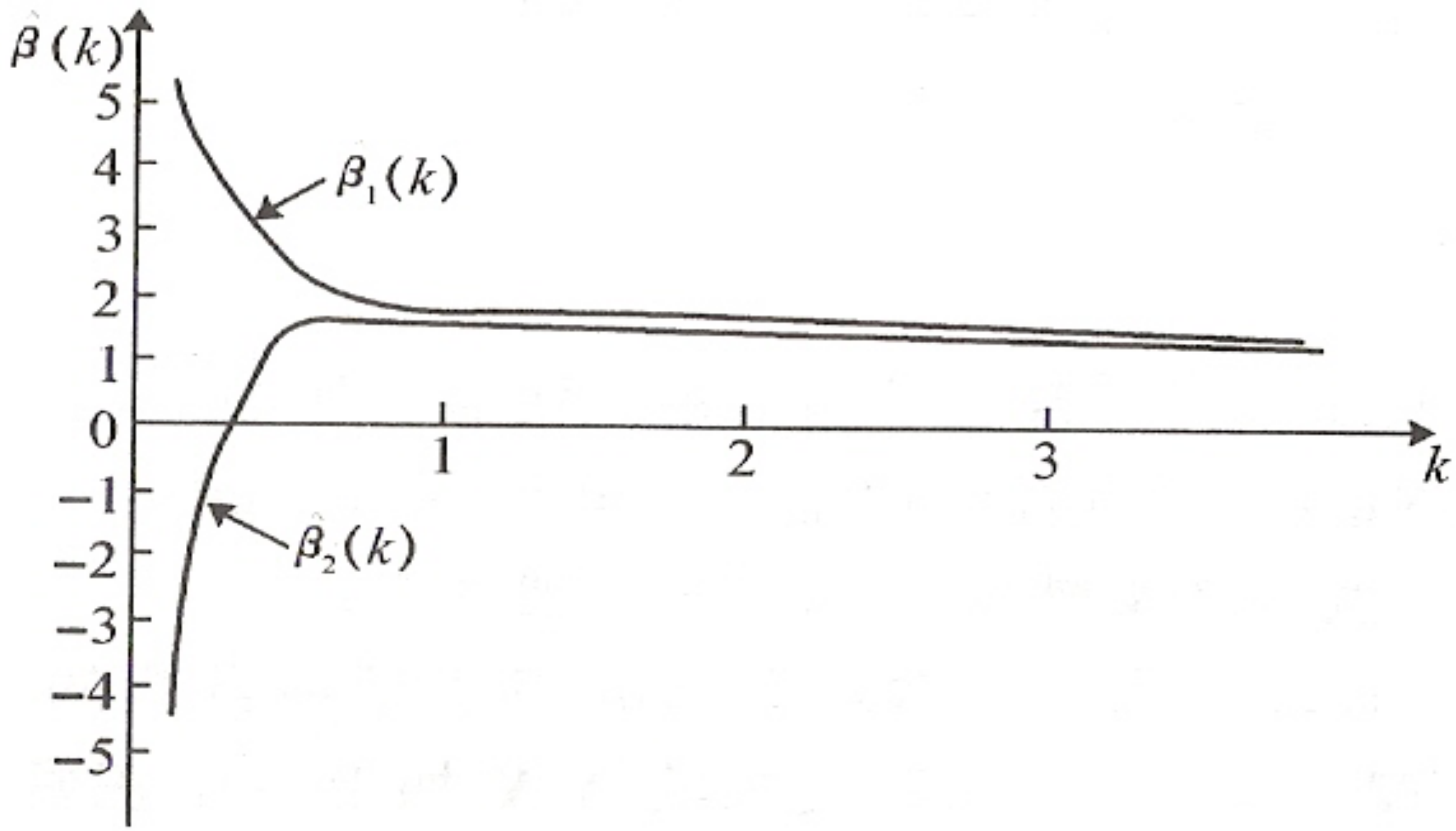
岭回归分析

```
x =

    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000
    1.1000    1.4000    1.7000    1.7000    1.8000    1.8000    1.9000    2.0000
    1.1000    1.5000    1.8000    1.7000    1.9000    1.8000    1.8000    2.1000

>> y

y =

   16.3000   16.8000   19.2000   18.0000   19.5000   20.9000   21.1000   20.9000

>> c=inv(x*x')*(x*y')

c =

   11.2924
   11.3073
   -6.5907
```

```
>> i=[1 0 0;0 1 0; 0 0 1];
>> c=inv(x*x'+i)*(x*y')

c =

    4.8349
    4.0748
    3.6592

>> c=inv(x*x'+2*i)*(x*y')

c =

    3.7599
    4.1664
    4.0037

>> c=inv(x*x'+3*i)*(x*y')

c =

    3.3092
    4.1778
    4.1055
```

岭回归分析

k	0	0.1	0.15	0.2	0.3	0.4	0.5	1.0	1.5	2	3
B1	11.31	3.48	2.99	2.71	2.39	2.20	2.06	1.66	1.43	1.27	1.03
B2	-6.59	0.63	1.02	1.21	1.39	1.46	1.49	1.41	1.28	1.17	0.98

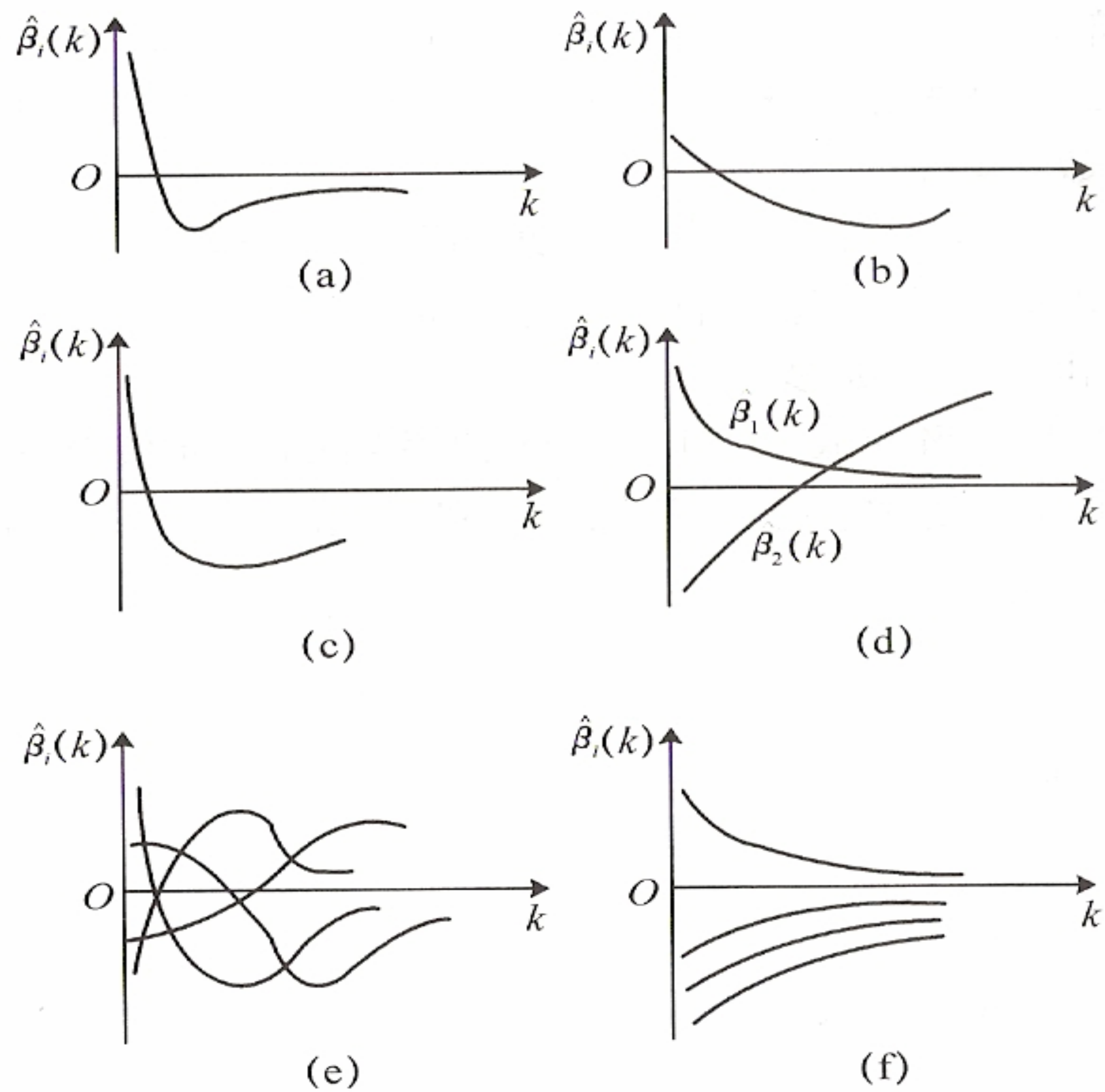


岭回归分析

当岭参数k在（0，∞）内变化时，$\hat{\beta}_j(k)$ 是k的函数，在平面坐标系上把函数 $\hat{\beta}_j(k)$ 描画出来。画出的曲线称为岭迹。在实际应用中，可以根据岭迹曲线的变化形状来确定适当的M值和进行自变量的选择。

在岭回归中，岭迹分析可用来了解各自变量的作用及自变量间的相互关系。

岭回归分析



岭回归分析

岭迹法选择k值的一般原则：

（1）各回归系数的岭估计基本稳定；

（2）用最小二乘估计时符号不合理的回归系数，其岭估计的符号变得合理；

（3）回归系数没有不合乎经济意义的绝对值；

（4）残差平方和增大不太多。

偏最小二乘回归分析

偏最小二乘回归法（PLSR：partial least squares regression)：是一种新型的多元统计数据分析方法，它主要研究的是多因变量对多自变量的回归建模，是对自变量数据矩阵X和因变量数据矩阵Y进行多元统计投影变换，建立X和Y的线性正交投影关系。当各变量内部高度线性相关时，用偏最小二乘回归法更有效。另外，偏最小二乘回归较好地解决了样本个数少于变量个数等问题。

偏最小二乘回归分析

设有p个自变量{ $x_1, \dots, x_p$ }和q因变量{ $y_1, \dots, y_q$ }。为了研究因变量和自变量的统计关系,我们观测了n个样本点,由此构成了自变量与因变量的数据表

X={ $x_1, \dots, x_p$ }$_{n\times p}$和Y={ $y_1, \dots, y_q$ }$_{n\times q}$。

如果X的变量之间以及Y的变量之间的相关性很弱，则可以应用经典的多元最小二乘回归方法进行求解。如果X^TX的逆不存在或是临界非奇异的，最小二乘无法求解或得到不可靠的参数。

偏最小二乘回归的思想

设有p个自变量{ $x_1,\ldots,x_p$ }和q因变量{ $y_1,\ldots,y_q$ }。为了研究因变量和自变量的统计关系,我们观测了n个样本点,由此构成了自变量与因变量的数据表

X={ $x_1,\ldots,x_p$ }$_{n\times p}$和Y={ $y_1,\ldots,y_q$ }$_{n\times q}$。

偏最小二乘回归分别在X与Y中提取出成分t_1和u_1(也就是说,t_1是 $x_1,\ldots,x_p$的线性组合,u_1是 $y_1,\ldots,y_q$的线性组合)。

偏最小二乘回归的思想

在提取这两个成分时,为了回归分析的需要,有两个要求:

(1)t_1和u_1应尽可能大地携带它们各自数据表中的变异信息;

(2)t_1和u_1的相关程度能够达到最大。

这两个要求表明：t_1和u_1应尽可能好地代表数据表X和Y,同时,自变量的成分t_1对因变量的成分u_1又有很强的解释能力。

偏最小二乘回归的思想

在第一个成分t_1和u_1被提取后,偏最小二乘回归分别实施X对t_1的回归以及Y对t_1的回归。如果回归方程已经达到满意的精度,则算法终止;否则,将利用 X被t_1解释后的残余信息以及Y被t_1解释后的残余信息进行第二轮的成分提取。如此往复,直到能达到一个较满意的精度为止。

偏最小二乘回归的思想

若最终对 X共提取了m个成分t_1、t_2、…、t_m，偏最小二乘回归将通过实施y_k(k=1、2、…、q)对t_1、t_2、…、t_m的回归,然后表达成y_k关于原变量x_1、x_2、…、x_p的回归方程。

偏最小二乘回归的特点

1.PLS是一种可以处理多个因变量对多个自变量的回归建模方法。特别当各变量集合内部存在较高程度的相关性时,用PLS进行回归建模分析,比对逐个因变量做多元回归更加有效,其结论更加可靠,整体性更强。

2.PLS可以较好地解决许多以往用普通多元回归分析方法无法解决的重要问题。例如自变量之间的多重相关性问题和样本点容量不宜太少等问题。

偏最小二乘回归的特点

3.PLS可以实现多种数据分析方法的综合应用。它可以集多元线性回归方法、主成分分析法和典型相关分析的基本功能为一体。在一次PLS计算后,不但可以得到多因变量对多自变量的回归模型,而且可以分析2组变量之间的相关关系,以及观察样本点间的相似性结构。这使得数据系统的分析内容更加丰富,同时还可以对所建立的回归模型给予许多更详细深入的实际解释。

偏最小二乘回归的特点

4.PLS允许在最终模型中包含原来全部自变量,最大限度地利用数据信息,使得PLS在相同的数据信息情况下比普通多元二乘回归模型具有更高的有效性。

5.在建模的同时实现了数据结构的简化,可以在二维平面上对多维数据的特性进行观察,图形功能强大。

因此,许多统计分析专家称PLS为第二代回归分析方法。

时间序列分析模型

时间序列分析模型

- 前面介绍的都是与时间无关的建模方法。满足独立采样的样本，无论如何打乱样本顺序，经回归得到的模型是不变的—基于稳态数据的建模方法；
- 实际生产中存在其数据与系统的演变过程本身有关，时间是样本数据的一个重要特征，预测模型必须包含时间因素在内—动态数据的建模方法。

系统中的某一变量或指标的数值或统计观测值按时间顺序排成一个数值序列就称为时间序列。

时间序列分析模型

- 时间序列中的每一个Xi都被看作为随机变量，所以时间序列实际上是随机变量序列，或称为随机序列。随机序列中不同时刻的xi,xj有一定的相互依赖关系。
- 从统计角度，时间序列中的每一个xi作为随机变量都应该有均值E{xi}和方差E{(xi-E{xi})^2}。若将之看做时间变量，则E{xi}是i的函数。

确定性与随机性时间序列分析

时间序列依据其特征，有以下几种表现形式，并产生与之相适应的分析方法：

（1）长期趋势变化

受某种基本因素的影响，数据依时间变化时表现为一种确定倾向，它按某种规则稳步地增长或下降。

使用的分析方法有：移动平均法、指数平滑法、模型拟和法等；

确定性与随机性时间序列分析

（2）季节性周期变化

受季节更替等因素影响，序列依一固定周期规则性的变化，又称商业循环。采用的方法：季节指数；

（3）循环变化

周期不固定的波动变化。

（4）随机性变化

由许多不确定因素引起的序列变化。它所使用的分析方法就是我们要讲的时间序列分析。

确定性与随机性时间序列分析

- 时间序列分析
  - 确定性变化分析
    - 趋势变化分析
    - 周期变化分析
    - 循环变化分析
  - 随机性变化分析 AR、MA、ARMA模型

时间序列分析模型

- **ARMA模型是一类常用的随机时间序列模型，是一种精度较高的时间序列短期预测方法，其基本思想是：某些时间序列是依赖于时间 t的一族随机变量，构成该时间序列的单个序列值虽然具有不确定性，但整个序列的变化却有一定的规律性，可以用相应的数学模型近似描述。**

时间序列分析模型

- **ARMA模型有三种基本类型：**
  - **自回归（AR：Auto-regressive）模型**
  - **移动平均（MA：Moving Average）模型**
  - **自回归移动平均（ARMA：Auto-regressive Moving Average）模型**

时间序列分析模型

如果时间序列 X_t 是它的前期值和随机项的线性函数，即可表示为

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + u_t$$

称为 p 阶自回归模型，记为AR（p）

- 实参数 $\varphi_1, \varphi_2, \cdots, \varphi_p$ 称为自回归系数，是待估参数。随机项 u_t 是相互独立的白噪声序列，且服从均值为0、方差为 σ^2 的正态分布，随机项与滞后变量不相关。

时间序列分析模型

如果时间序列 X_t 是它的**当期和前期的随机误差项的**的线性函数，即可表示为

$$X_t = u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \cdots - \theta_q u_{t-q}$$

称为 q 阶移动平均模型，记为MA（q）

- 实参数 $\theta_1, \theta_2, \cdots, \theta_q$ 称为移动回归系数，是待估参数。

时间序列分析模型

如果时间序列 X_t 是它的**当期和前期的随机误差项以及前期值的线性函数，即可表示为**

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \cdots - \theta_q u_{t-q}$$

称为（p,q）阶自回归移动平均模型，记为ARMA（p,q）

- 实参数 $\varphi_1, \varphi_2, \cdots, \varphi_p$, $\theta_1, \theta_2, \cdots, \theta_q$ 称为待估参数。

样本自协方差函数

- 构成时间序列的每个序列值 $X_t, X_{t-1}, X_{t-2}, \cdots, X_{t-k}$ 之间的简单相关关系称为自相关。自相关程度由自相关系数 γ_k 度量，表示时间序列中相隔k期的观测值之间的相关程度

$$\gamma_k = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$$

模型的识别与建立

- MA(q)的自协方差函数与样本自相关函数

$$\gamma_k=\begin{cases}\left(1+\theta_1^2+\cdots+\theta_q^2\right)\sigma^2, & k=0\\ \left(-\theta_k+\theta_1\theta_{k+1}+\cdots\theta_{q-k}\theta_q\right)\sigma^2, & 1\le k\le q\\ 0, & k>q\end{cases}$$

$$\rho_k=\frac{\gamma_k}{\gamma_0}=\begin{cases}1, & k=0\\ \dfrac{-\theta_k+\theta_1\theta_{k+1}+\cdots\theta_{q-k}\theta_q}{1+\theta_1^2+\cdots+\theta_q^2}, & 1\le k\le q\\ 0, & k>q\end{cases}$$

截尾性

模型的识别与建立

- **AR(p)的自相关函数满足差分方程**

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \ldots + \varphi_p \rho_{k-p} \quad k>0$$

- **该式不是截尾序列，其受负指数函数控制，具有“拖尾”性。**

ARMA(p,q）序列的自相关函数是拖尾的.

AIC赤池准则确定模型的阶数

S是模型的未知参数的总数

$\hat{\sigma}^2$是用某种方法得到的方差 σ^2 的估计值

N为样本大小，则定义AIC准则函数 $AIC(S) = \ln \hat{\sigma}^2 + \frac{2S}{N}$

用AIC准则定阶是指在(p,q)的一定变化范围内，寻求使得AIC最小的点 $(\hat{p}, \hat{q})$ 作为 (p,q)的估计值。

$$AIC = \ln \hat{\sigma}^2 + \frac{2p}{N} \qquad AIC = \ln \hat{\sigma}^2 + \frac{2(p+q)}{N}$$

BIC贝叶斯准则确定模型的阶数

$$\boxed{BIC = \ln(L) \ * N + S * \ln(N)}$$

$$\boxed{AIC = \ln(L) \ * N + S * 2}$$

模型的参数估计

在阶数给定的情形下，模型参数的估计的首要任务是估计模型的相关参数：

（1）均值估计

$$\hat{\mu} = \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

（2）自协方差函数估计

$$\hat{r}_k = \frac{1}{N}\sum_{i=1}^{N-k} x_i x_{i+k}; \qquad \hat{r}_0 = \hat{\sigma}^2$$

根据N的大小可分为有偏和无偏估计

（3）自相关函数估计

$$\hat{\rho}_k = \frac{\hat{r}_k}{\hat{r}_0}$$

模型的参数估计

在阶数给定的情形下模型参数的估计有三种基本方法：矩估计法、逆函数估计法和最小二乘估计法.

AR模型：

$$\begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{p-2} \\ \vdots & \vdots & & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{pmatrix}$$

白噪声序列的方差的矩估计为

$$\hat{\sigma}^2 = \gamma_0 - \sum_{j=1}^{p} \hat{\varphi}_j \hat{\gamma}_j$$

模型的参数估计

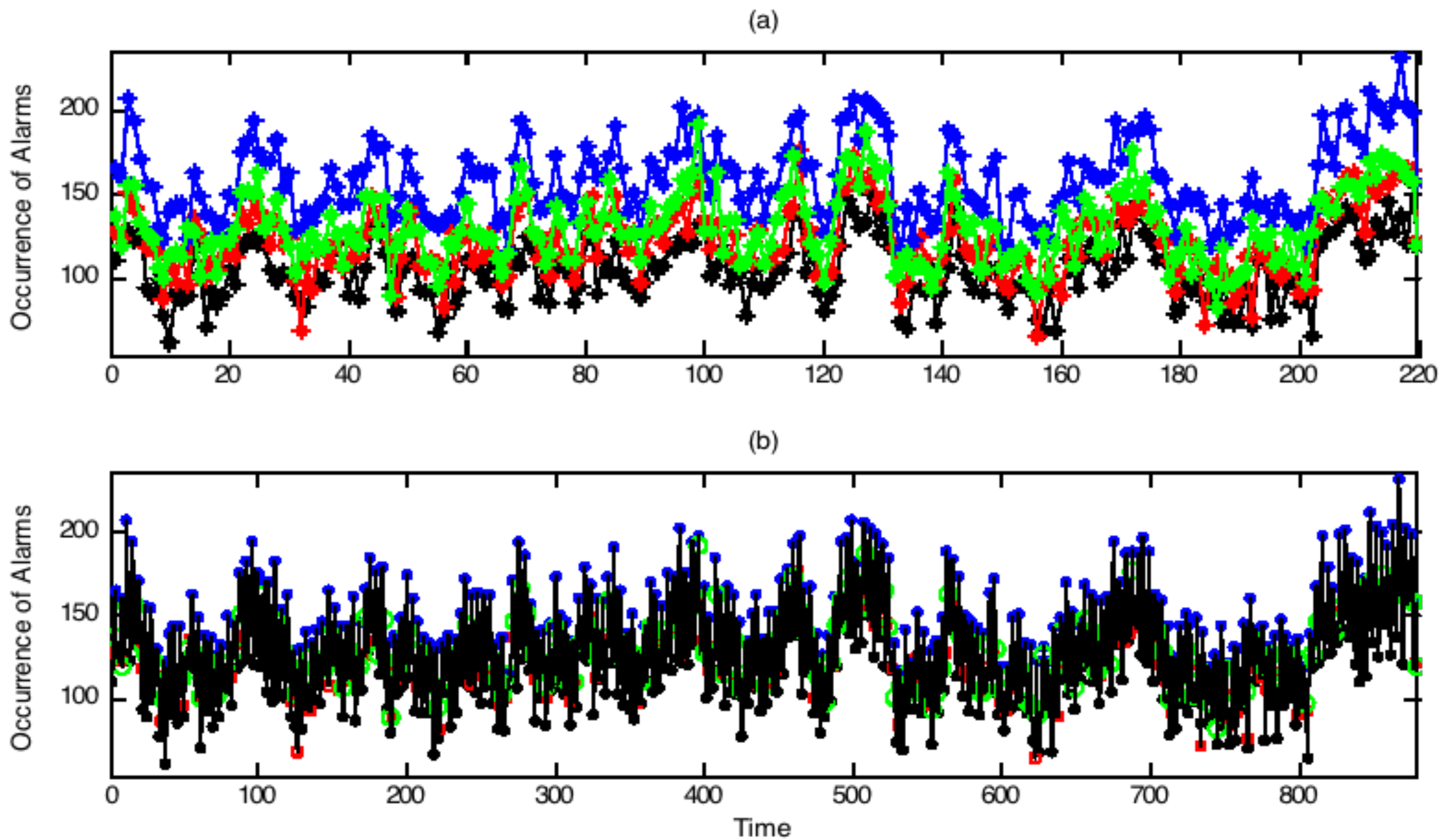
MA模型：由自协方差函数

$$
\begin{cases}
\left(1+\hat{\theta}_1^2+\cdots+\hat{\theta}_q^2\right)\hat{\sigma}^2=\hat{\gamma}_0 \\
\left(-\hat{\theta}_k+\hat{\theta}_1\hat{\theta}_{k+1}+\cdots+\hat{\theta}_{q-k}\hat{\theta}_q\right)\hat{\sigma}^2=\hat{\gamma}_k, k=1,\cdots,q
\end{cases}
$$

可解出：

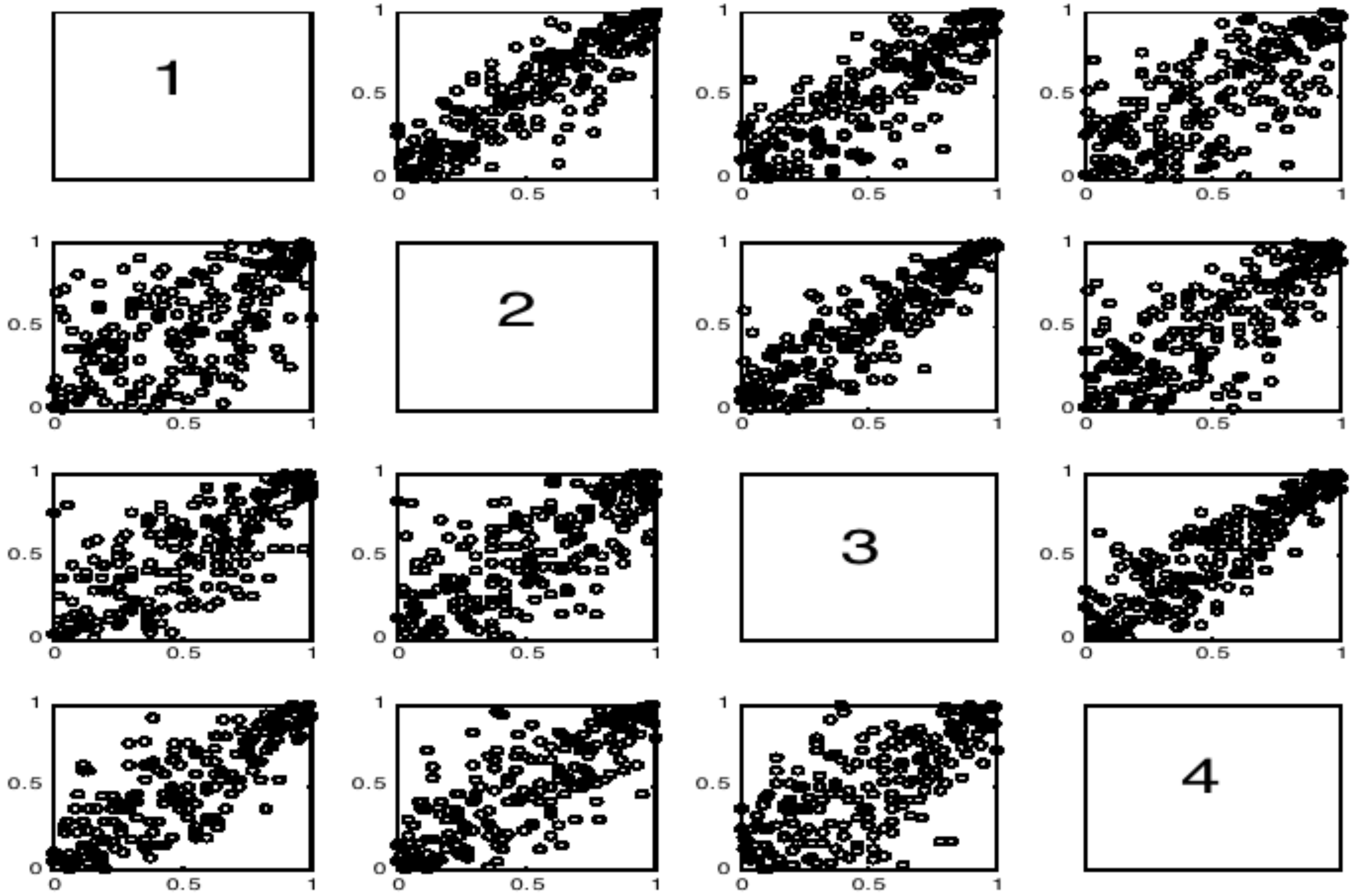
$$
\boxed{
\begin{aligned}
&\hat{\sigma}^2=\hat{r}_0\left(\sum_{i=1}^{q}\hat{\theta}_i^2\right)^{-1}, \theta_0=1 \\
&\hat{\theta}_k=-\frac{\hat{r}_k}{\hat{\sigma}^2}+(\hat{\theta}_1\hat{\theta}_{k+1}+\ldots+\hat{\theta}_{q-k}\hat{\theta}_q), 1\le k\le q
\end{aligned}
}
$$

时间序列模型举例

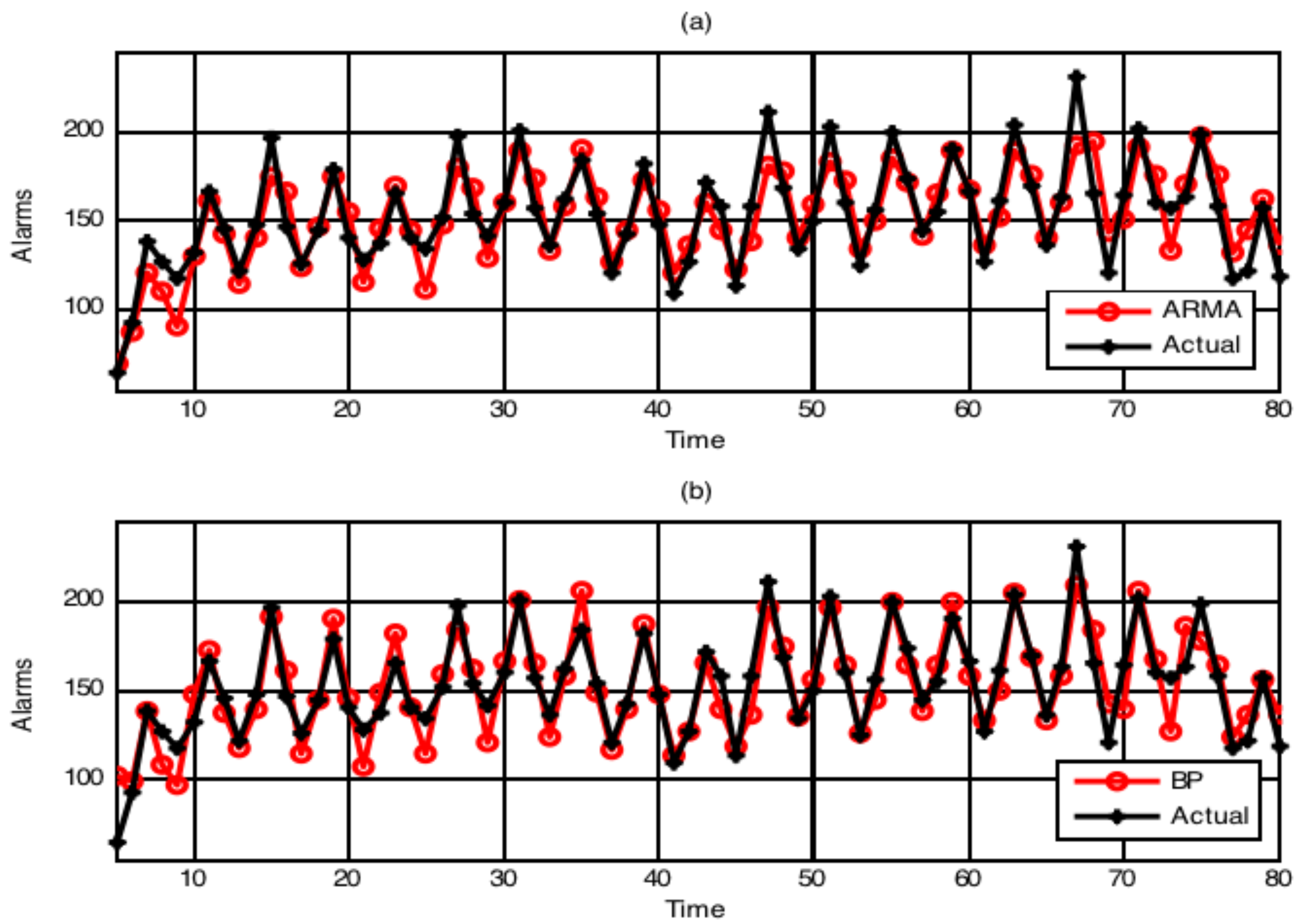


(a)
Occurrence of Alarms
200
150
100
0
20
40
60
80
100
120
140
160
180
200
220
(b)
Occurrence of Alarms
200
150
100
0
100
200
300
400
500
600
700
800
Time

时间序列模型举例

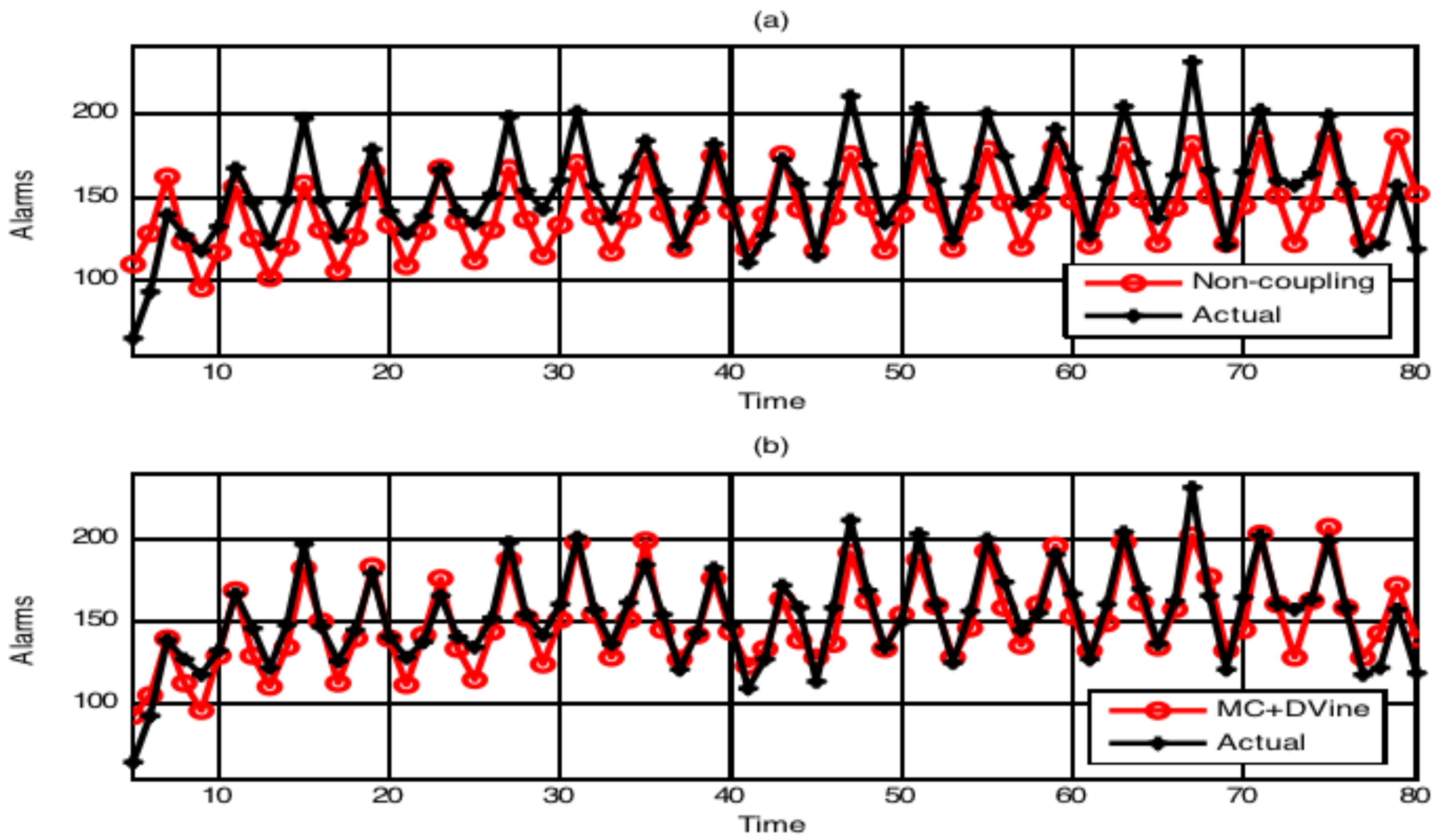


时间序列模型举例



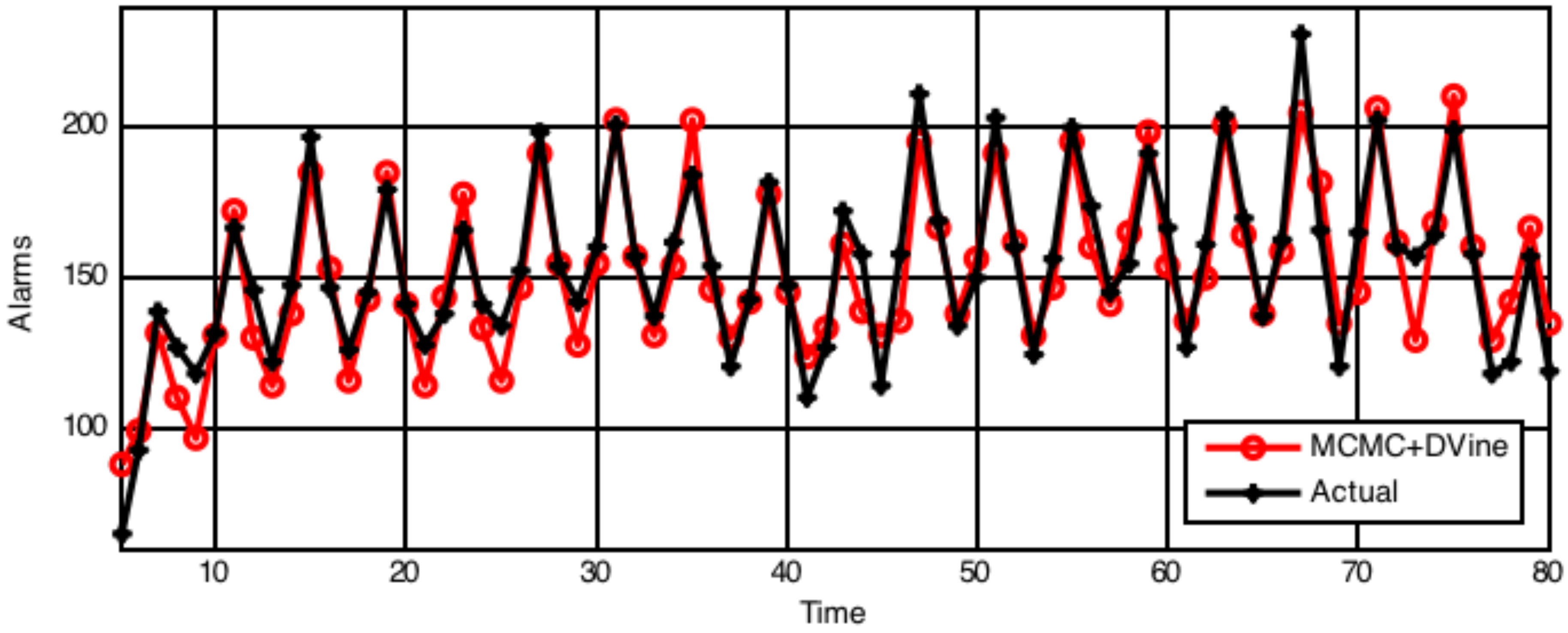
(a)
Alarms
200
150
100
10
20
30
40
50
60
70
80
Time
ARMA
Actual
(b)
Alarms
200
150
100
10
20
30
40
50
60
70
80
Time
BP
Actual

时间序列模型举例



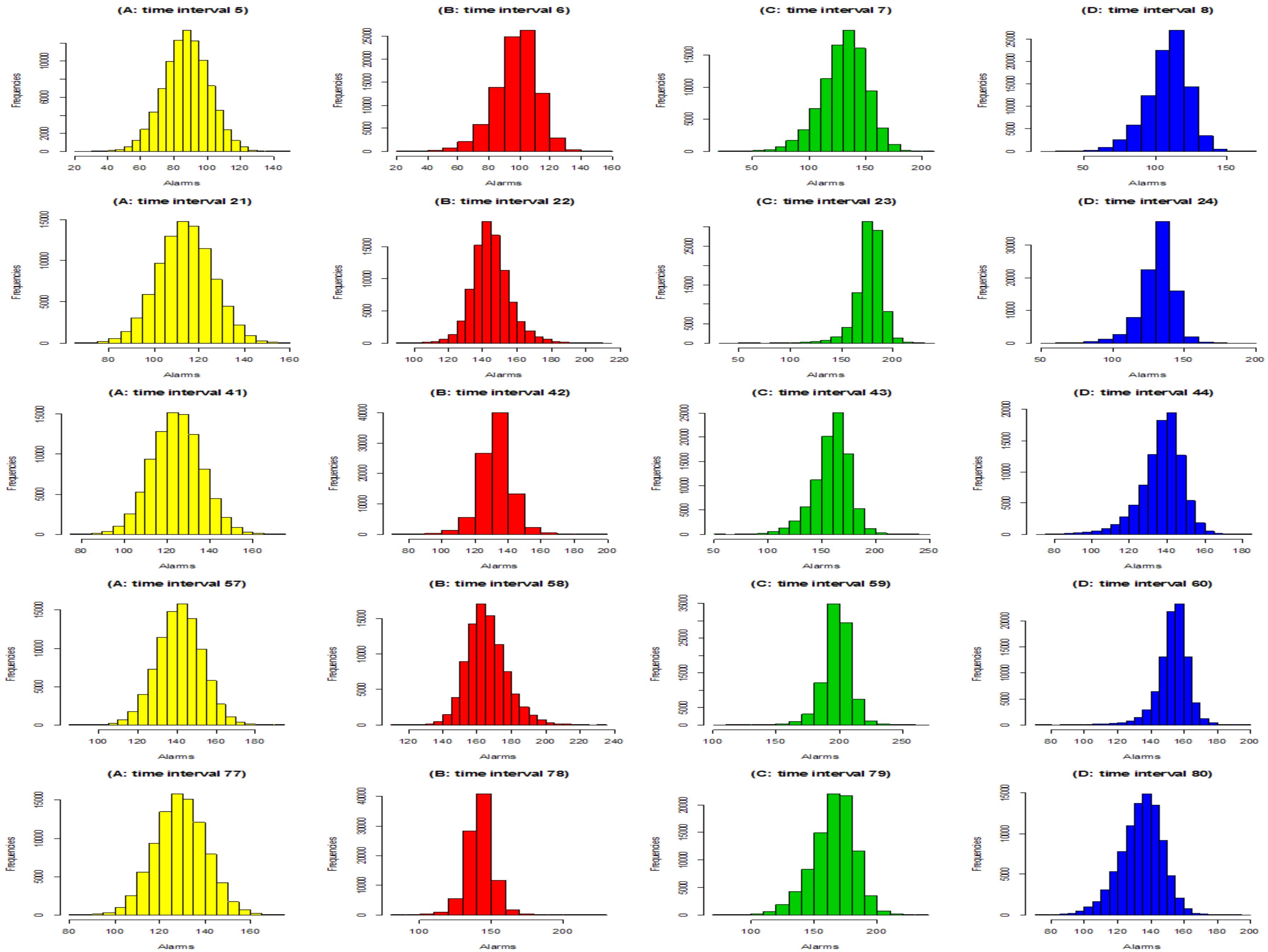
(a)
200
150
100
Alarms
10
20
30
40
50
60
70
80
Time
Non-coupling
Actual
(b)
200
150
100
Alarms
10
20
30
40
50
60
70
80
Time
MC+DVine
Actual

时间序列模型举例



200
150
100
Alarms
10
20
30
40
50
60
70
80
Time
MCMC+DVine
Actual

时间序列模型举例


(A: time interval 5)
(B: time interval 6)
(C: time interval 7)
(D: time interval 8)
(A: time interval 21)
(B: time interval 22)
(C: time interval 23)
(D: time interval 24)
(A: time interval 41)
(B: time interval 42)
(C: time interval 43)
(D: time interval 44)
(A: time interval 57)
(B: time interval 58)
(C: time interval 59)
(D: time interval 60)
(A: time interval 77)
(B: time interval 78)
(C: time interval 79)
(D: time interval 80)
Frequencies
Alarms

作业

- **1. P163页第6题；**