CDA LEVEL 1案例分析培训
人大经济论坛

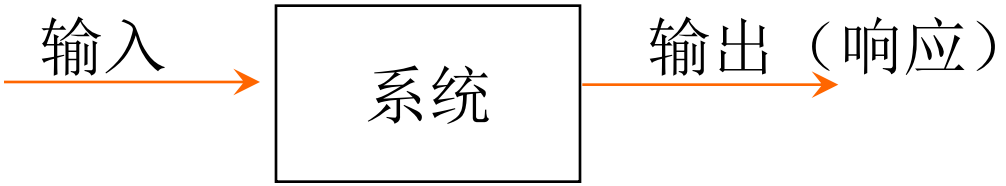
翟祥
北京林业大学统计系
zhaixbh@126.com

目录

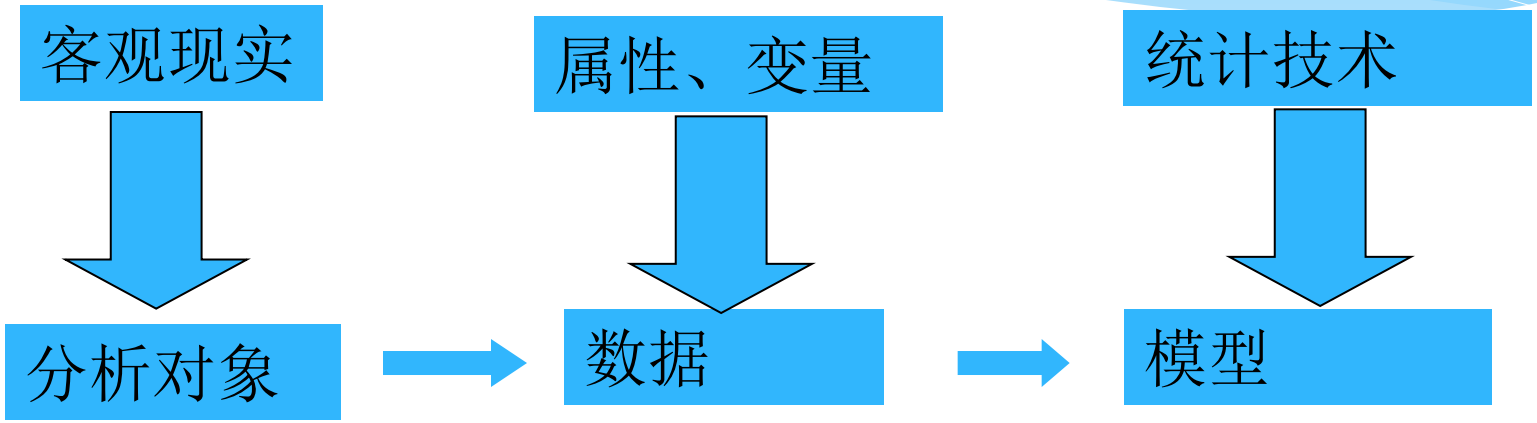
目录

数据分析（挖掘）逻辑

* 从系统的观点看：系统的记忆性，也就是某一时刻或空间进入系统的输入对系统后继行为的影响，图示如下：



数据分析（挖掘）逻辑



模型分类

* 机理模型

根据数据提供的信息体现其中的关系

* 经验模型

利用数据体现的规律进行预测（predicte和forecast）

* 模拟模型

通过模型全面的展现事情发生发展的整个过程

模型分类

验证性模型	有监督训练	预测模型
探索性模型	无监督训练	模式发现

模型分类

预测模型

回归分析

决策树

神经网络

模式发现

聚类分析

探索性因子分析

关联规则

数据分类（review）

测定层次	特征	运算功能	举例
1、定类测定	分类	计数	产业分类
2、定序测定	分类；排序	计数；排序	企业等级
3、定距测定	分类；排序；有基本测量单位	计数；排序；加减	产品质量差异
4、定比测定	分类；排序；有基本测量单位；有绝对零点	计数；排序；加减乘除	商品销售额

数据分类

* 观测数据
  * 通过调查或观测收集到的数据
* 实验数据
  * 在实验中控制实验对象而收集到的数据
* 截面数据
  * 在相同或者相近的时间点上收集的数据
* 时间序列数据
  * 在不同的时间上收集到的数据

数据和时间的关系

(1)中国1993年—1998年的GDP增长率 （%）

1993	1994	1995	1996	1997	1998
14.2	13.5	10.5	9.6	8.8	7.8

（2）横截面数据。一个或多个变量在同一时点上收集的数据。

1992年实际GDP增长

国家/地区	加拿大	智利	墨西哥	秘鲁	美国	中国	香港	日本
GDP	0.9	12.3	3.6	-1.7	2.7	14.2	6.3	1

（3）混合数据

国家和地区	实际GDP增长率						
	1992年	1993年	1994年	1995年	1996年	1997年	1998年
加拿大	0.9	2.5	3.9	2.2	1.2	4.0	3.1
智利	12.3	7.0	5.7	10.6	7.4	7.1	3.4
墨西哥	3.6	2.0	4.4	-6.2	5.2	7.0	4.8
秘鲁	-1.7	6.4	13.1	7.4	2.5	6.9	0.3
美国	2.7	2.3	3.5	2.0	2.8	3.9	3.9
中国	14.2	13.5	12.6	10.5	9.6	8.8	7.8
香港	6.3	6.1	5.4	3.9	4.6	5.3	-5.1
日本	1.0	0.3	0.6	1.5	3.9	1.4	-2.8

数据分析（挖掘）支撑点

* 模型可解释性
* 模型和技术是否对数据有严格的假定
* 能否有效的抵御维度“诅咒”
* 能否稳健的应对异常值
* 数据测量层次的难度（定性数据问题）
* 缺失值是否需要事先处理
* 计算的复杂性

目录

分析流程（SEMMA）

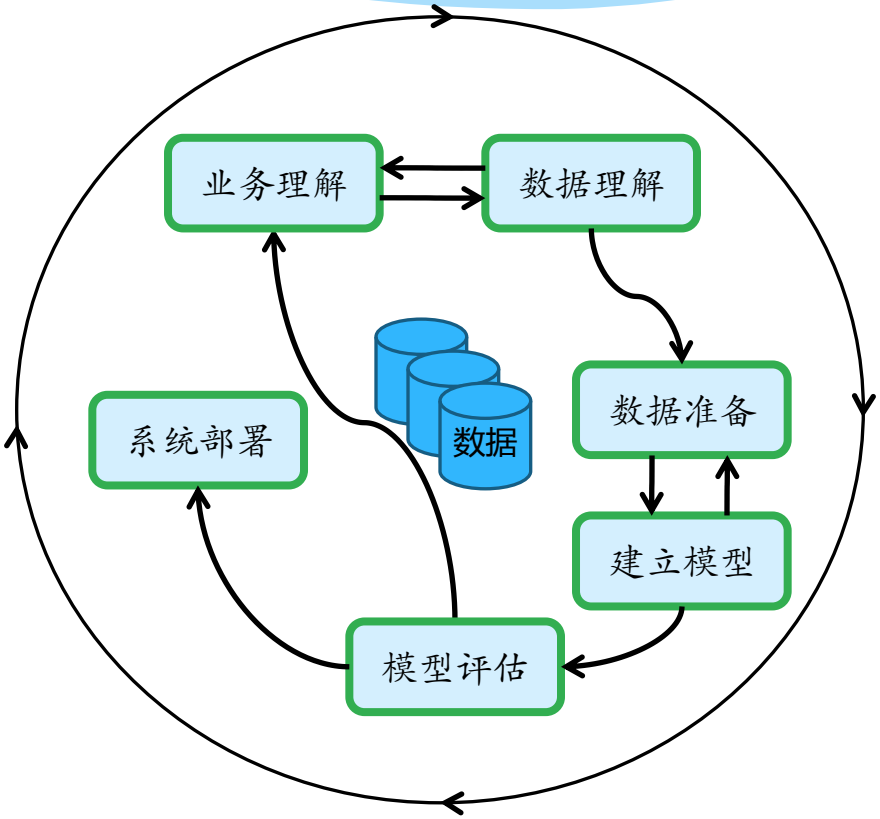
分析流程（SEMMA）

* Sample（数据选取）
* Explore（数据探索）
* Modify（数据预处理）
* Model（模型建立）
* Assessment（模型评价）
* Deploy（模型部署）
* Revise（模型修正）

模型开发步骤

- 业务理解
  - 目标定义
  - 分析窗口定义
- 数据探索
  - 基本情况
  - 单变量显著性分析
- 数据准备（预处理）
  - 数据质量检查
  - 数据预处理
  - 宽表构建
- 模型构建
  - 抽样：训练集、验证集
  - 运用相应算法训练数据
  - 模型参数解释
  - 重要变量分析
- 模型评估
- 模型应用

数据挖掘方法论



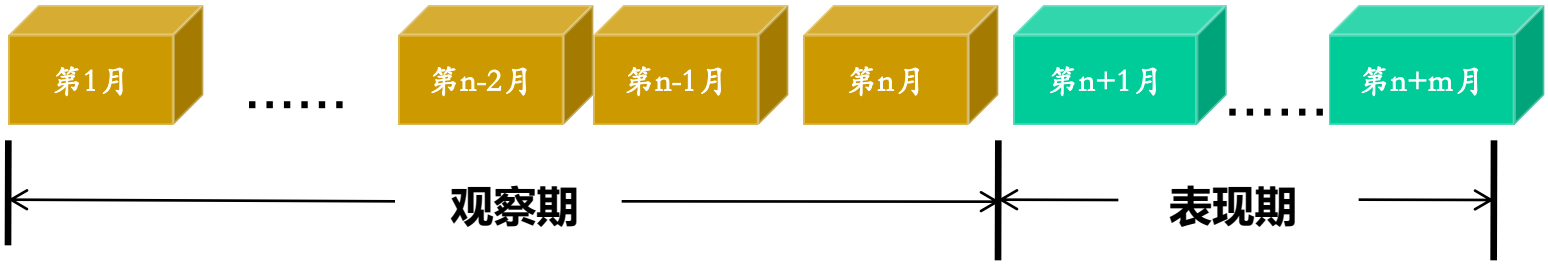
业务理解

业务问题定义

* 面对的分析对象是什么。
* 样本数据是在观察期内正常的合同，在表现期已经提前接清（流失）的合同为目标值1，否则为0。

分析时间窗口

- 基于业务特征和行业经验，我们一般把时间窗口设计为8-18个月，时间窗口设计为观察期6-12个月，表现期2-6个月。
- 模型是通过观察期的分析对象特征和持有客户的行为特征来预测分析对象在表现期的状态。



行业数据挖掘主题

分析模型介绍（基础模型—必做）

分析模型	模型应用介绍
客户细分	在客户理解和营销策划阶段，业务人员根据不同客户群自然属性和行为特征，进行差异化的客户分析和客户服务，从而指导针对性营销工作
客户响应	预测客户对新产品、套餐、营销活动的响应情况，指导新产品和套餐销售策略，提升营销活动的推出效果
流失预警	预测客户在未来一段时间内的流失可能性，采取不同的客户维系和挽留的方法，提高运营商存量保有水平，可以用于主动拆机流失、话务量流失、零次户流失、欠费停机流失等预测
客户信用	针对客户不同信用等级，在客户业务受理、业务使用、帐单催欠等环节进行不同等级的信用控制，从而提供运营商的客户信用管理水平

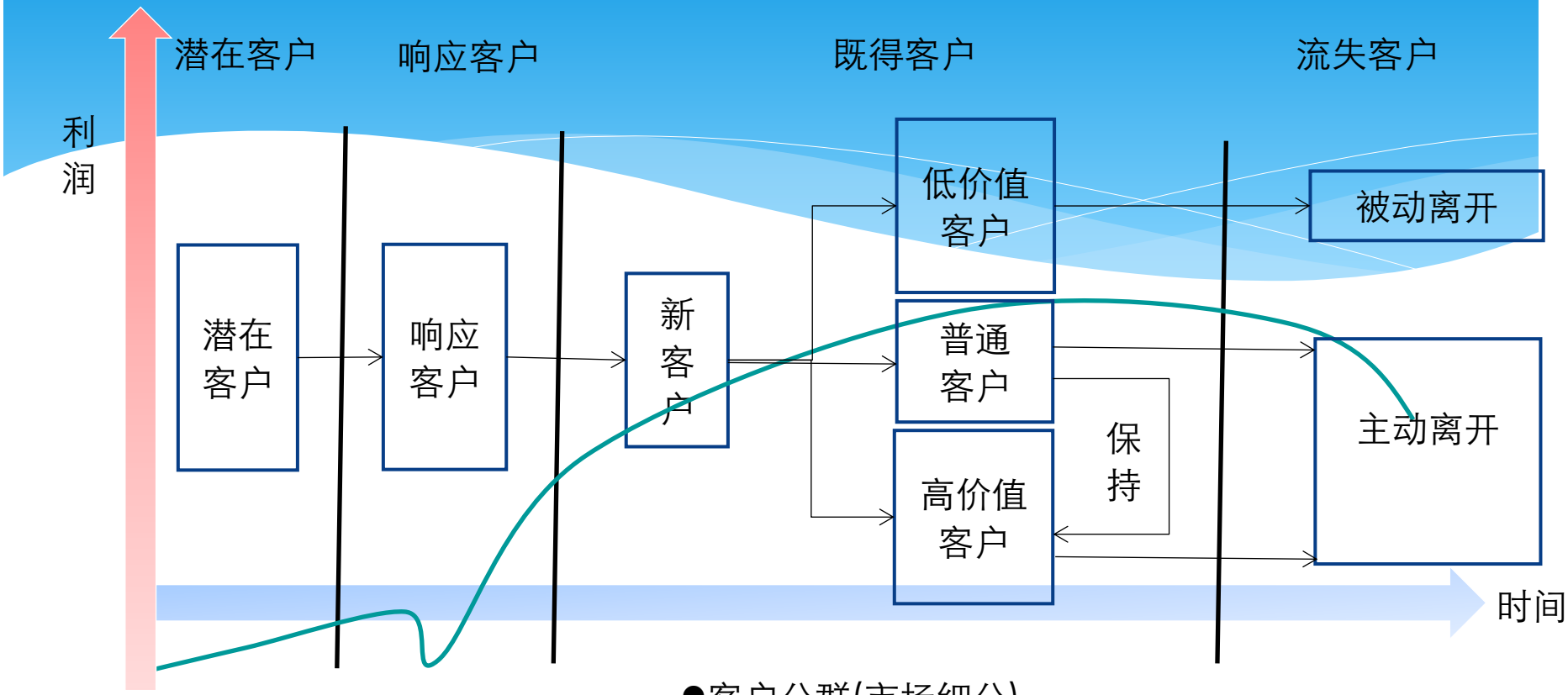
分析模型介绍（可选—业务需求驱动）

分析模型	模型应用介绍
业务量预测	在客户理解阶段，通过分客户群、分产品、分地域，**预测运营商**不同客户群体或者单个客户未来一段时间内的业务量的变化趋势，根据预测的不同情况，采取不同的营销措施
收入预测	在客户理解阶段，通过分客户群、分产品、分地域，**预测**运营商不同客户群体或者单个客户未来一段时间内的收入的变化趋势，根据预测的不同情况，采取不同的营销措施
客户价值	综合考虑客户当前价值和潜在价值,分析客户对运营商的贡献，根据客户价值高低，采取不同的营销、销售和服务策略
欠费预测	分析影响客户欠费的因素，进行欠费预测分析中，生成欠费预测结果，从而适时地对客户进行重点跟踪，并在必要时采取措施，以减少损失
渠道配置优化	对渠道的人、财、物的资源优化配置，分析客户在各个渠道的行为特征，结合不同渠道营销成本和渠道的特点，帮助营销执行人员选择合理的渠道，对客户进行针对性营销，降低营销成本
客户动因/偏好	通过分析客户产品偏好，客户偏好渠道等，可以及时、准确的洞察客户动因和偏好，跟踪和把握客户需求，全面深入理解客户心里需求，提高针对性营销有效性

分析模型介绍（可选—业务需求驱动）

分析模型	模型应用介绍
欺诈发现	通过发现客户欺诈模式，在客户业务受理、业务使用、帐单催欠等环节进行控制，提高风险控制水平。
交叉销售	通过分析产品之间关联性，找出产品组合的规则，指导组合产品的设计，提高产品组合营销的效果，降低单产品的流失率
资费定价对比	通过套餐测算方程，研究套餐对于客户的区隔，以及套餐对不同ARPU区段客户的影响，为合理选择套餐方案和套餐分档设计提供有力的支持
套餐收入测算	完成目标客户在新资费政策下的消费行为预演，预演结果与原资费政策下用户的消费情况进行对比评估，验证资费是否与预期营销目的相符，用于电信运营商对资费政策和营销计划的调整
客户满意度	分析影响客户满意度的关键因素，找出客户满意度的评价指标，改善客户服务水平
客户忠诚度	分析客户忠诚度的影响因素，能够找出影响客户忠诚度的评价指标，提高客户对自身的依赖程度

客户关系(CRM)管理方面与数据挖掘运用场景



- 发掘潜在客户

- 客户获取
- 初始信用评分
- 客户价值预测

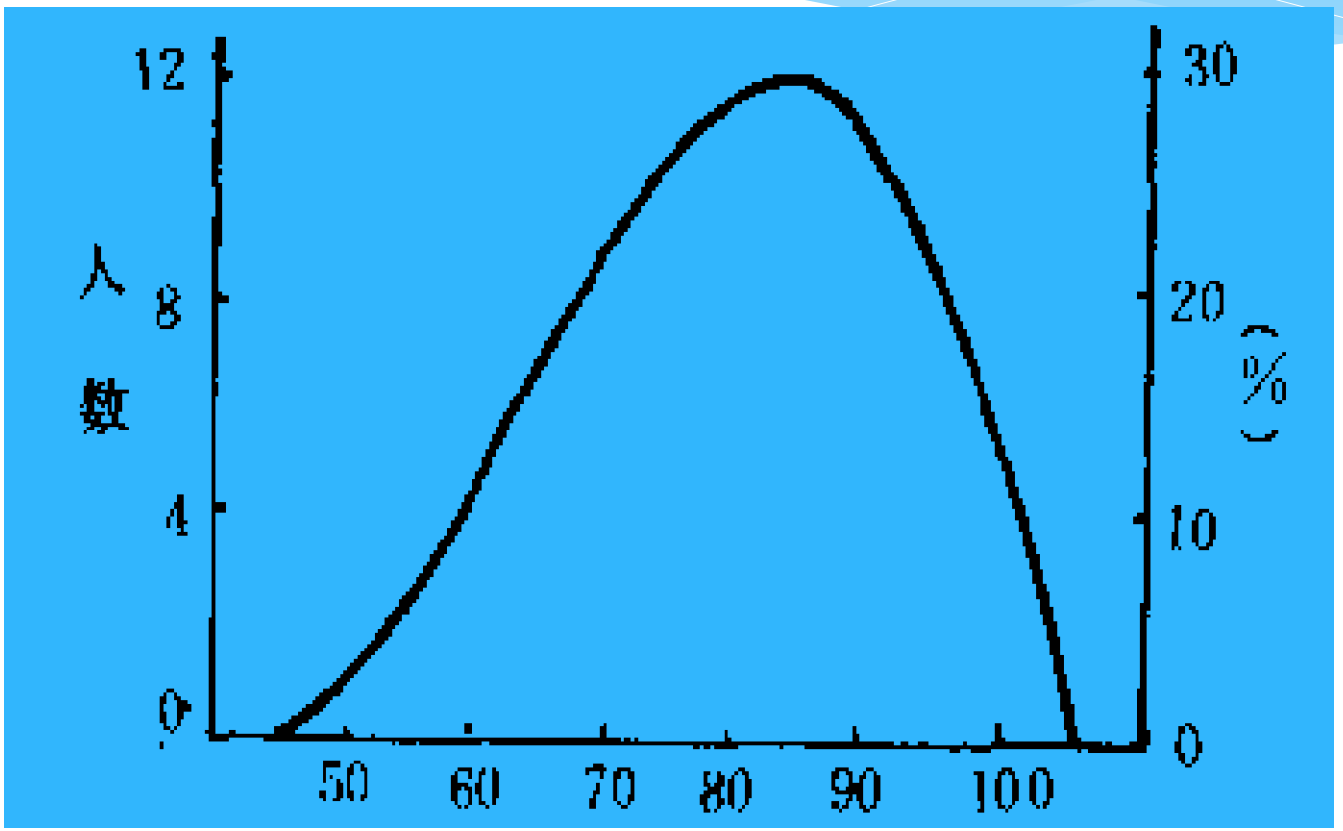
- 客户分群(市场细分)
- 交叉销售
- 产品精准营销
- 行为信用评分
- 欺诈侦测
- 客户保留
- 客户关系网

- 流失客户时间判断
- 流失客户类型判断

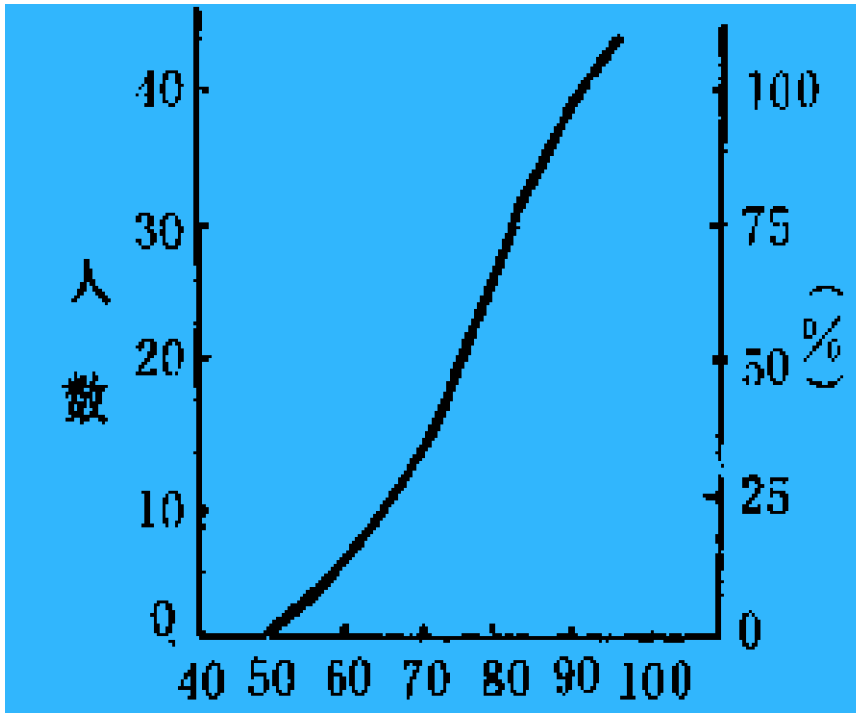
数据探索

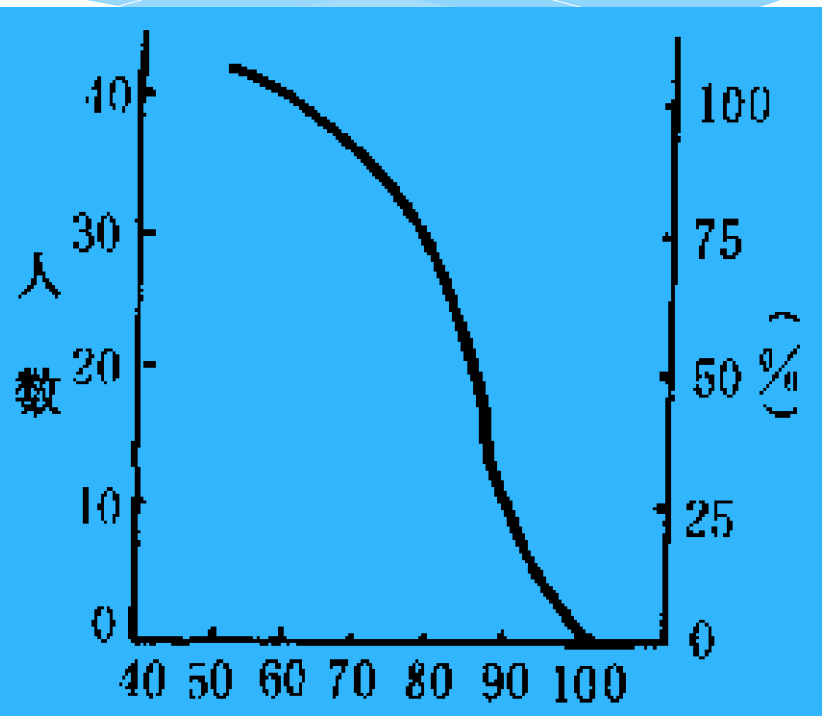
* 重要变量的分布探索
* 密度函数(P(X=x))
* 分布函数(P(X≤x))
* 生存函数(P(X>x))
* 变量间关系探索
* 相关分析
* 维度诅咒

密度函数



分布函数和生存函数





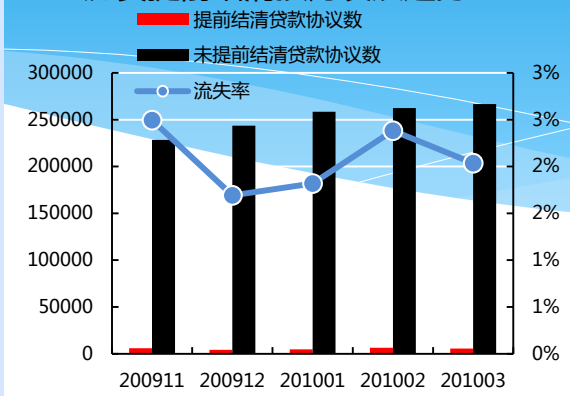
数据探索

房贷合同发展基本状况

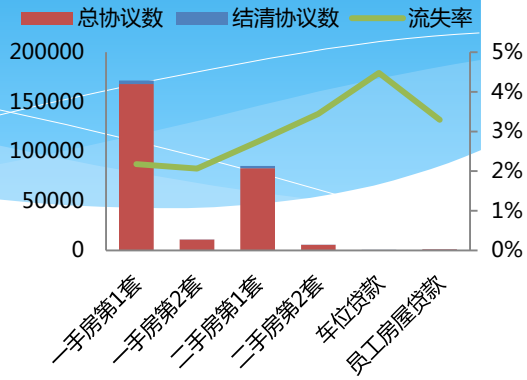
数据理解 | 业务理解

> 房贷合同结清状况趋势分析
> 房贷合同的构成分析
> 提前结清房贷合同构成分析
> ……

房贷提前结清按月发展趋势



房贷提前结清按合同类型分析



特征变量显著性分析

变量对提前结清房贷行为的影响程度

> 房贷合同基本信息

房贷合同的类型、短期期限的贷款、贷款余额较少、已还款期数较多、当前贷款执行利率较高都能显著地影响客户的提前结清还款的倾向。

> 房贷客户基本信息

客户年龄较大、高学历、已婚人士更易提前偿付贷款。

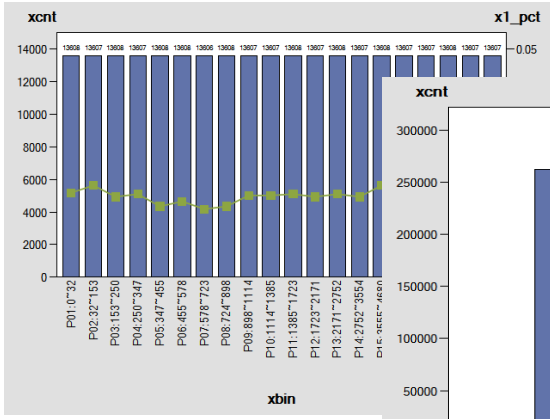
> 房贷客户交易行为

客户储蓄存款余额多寡、储蓄存款次数金额、接收转账汇款的频度和额度都能影响到客户提前还贷。

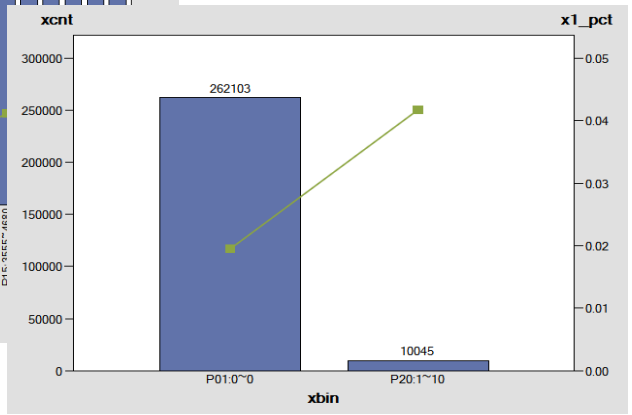
> 历史还贷情况

历史提前还贷次数、金额和单笔金额对提前结清贷款有显著影响。

储蓄账户余额



近6月个贷提前还款次数



最近半年有提前还贷行为的客户，其未来2月内提前结清贷款比例为4.17%，是平均水平2.04%的2倍多。

数据探索-相关性分析

* 散点图
* 相关系数
  * 皮尔森相关系数
  * 斯皮尔曼相关系数
  * 肯德尔相关系数
  * 霍弗丁相关系数

数据探索-相关性分析

分析目的

> 分析各特征变量间相互影响的强弱程度，剔除彼此强相关的变量，减少变量冗余程度。

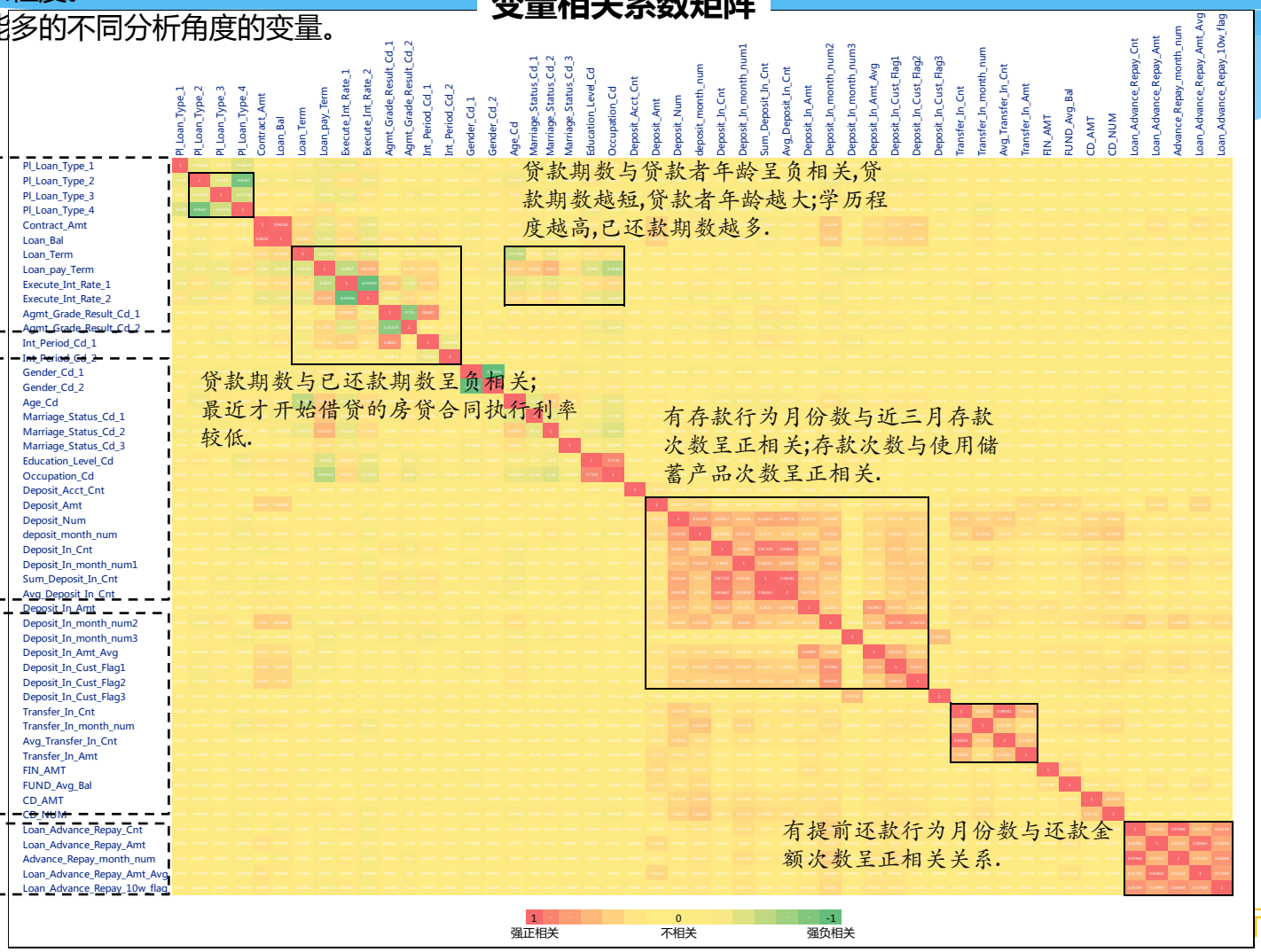
> 为模型构建选取尽可能多的不同分析角度的变量。

房贷合同基本信息

客户基本信息

客户交易信息

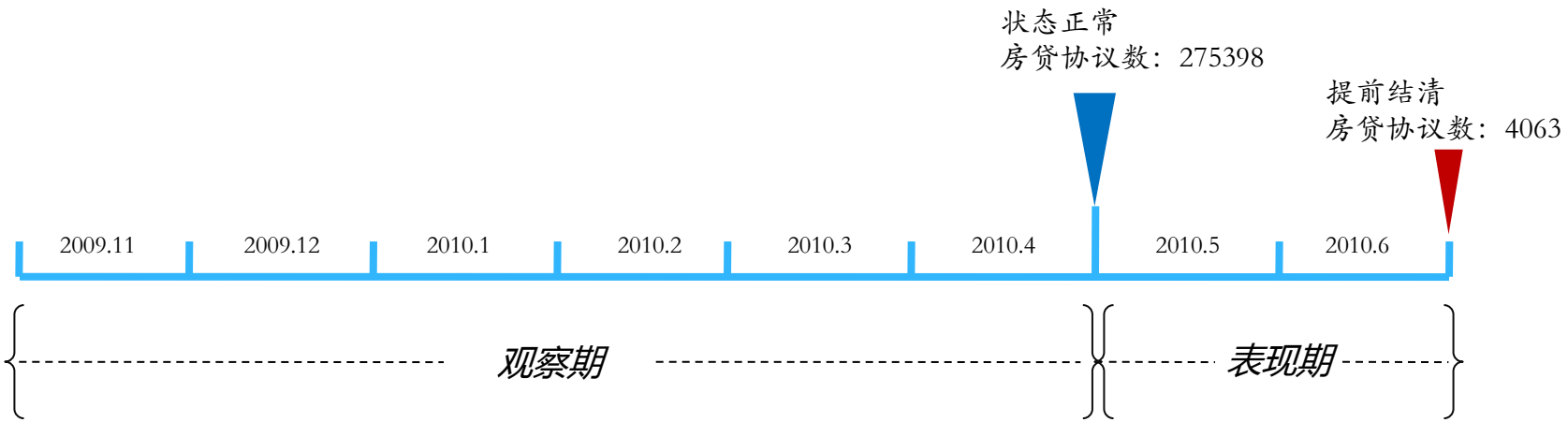
历史还贷记录



数据准备（预处理）

分析客户群选择

* 分析客户群
  > 分析客户群：2010年4月底状态正常的客户。
  > 分析样本观察窗口是2009年11月-2010年4月，表现窗口是2010年5-6月。

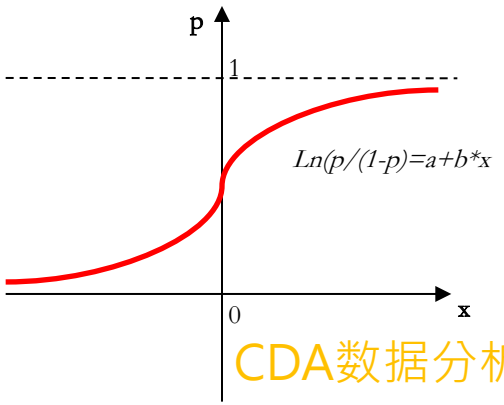


模型构建-抽样设计

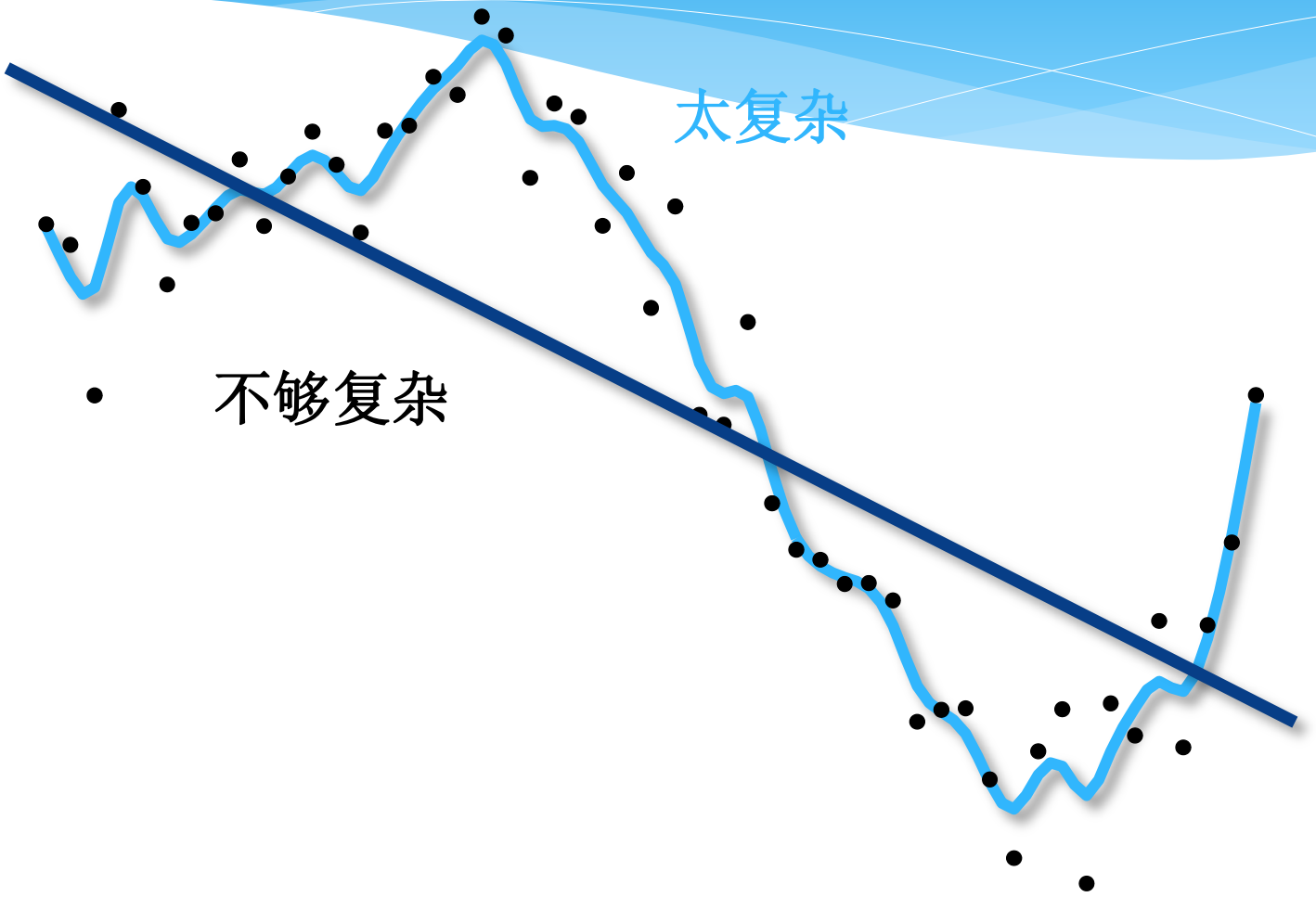
* 在建模过程中，需要使用抽样技术，从符合考察范围的用户总体中以适当方法，抽取一定量的样本，进行分析。从而保证模型的可靠性和预测模型的稳定性，同时，降低建模初期的数据准备压力。
* 针对不同的模型的实际数据的差异性，采用不同的抽样方法来进行抽取模型的样本。
  * 简单无重复随机抽样
  * 分层抽样: 分层等比例随机抽样、分层不等比例随机抽样
  * 分类抽样
* 将样本数据集进行划分为训练集、验证集，比例分别占60%和40%。

模型构建-模型选择

- 逻辑回归算法Logistic Regression
  算法基本公式：
  $P = \exp(S) / (1 + \exp(S))$
  $S = a + b1 * X1 + \sim + bn * Xn$
  $Odds = P / (1-P)$
- LOGISTIC模型主要用于处理因变量为分类变量的问题，如：客户信用的好、坏；在分析时，我们感兴趣的是因变量取各个值的概率P。
  - P表示因变量取某个值(Y=1)的概率：
    $1+\exp(a + b1 * X1 + \sim + bn * Xn)$
  - 经变换后生成对数似然比（Odds）：
    $\ln(p/(1-p))=a + b1 * X1 + \sim + bn * Xn$
    揭示了n个自变量X与响应变量Y之间的非线性关系；
  - 因变量Y是分类型变量（如二分变量0/1）；用p表示Y=1发生的概率；预测p与n个自变量X的关系。
- 逻辑斯蒂模型的优点：
  - 变量的可解释性强
  - 对变量分布正态性和方差齐性不做要求
  - 对自变量类型不做要求
  - 有较强的预测稳定性和准确性



模型复杂度



一点困扰

* 数据分区

* 业务支持

模型结果解释

评分公式

计算相应的概率：

$P = \exp(S)/(1 + \exp(S))$

其中，$S= a + b1* X1+ \cdots + bn* Xn$，重要变量Xi及其参数bi如前所述。

变量分析

变量的对数似然率odds rate 表示该变量取值增加1个单位,对目标p与非目标值1-p间的差异贡献的程度，odds rate越大,说明该变量对目标变量的区分能力强。

。

重要影响变量

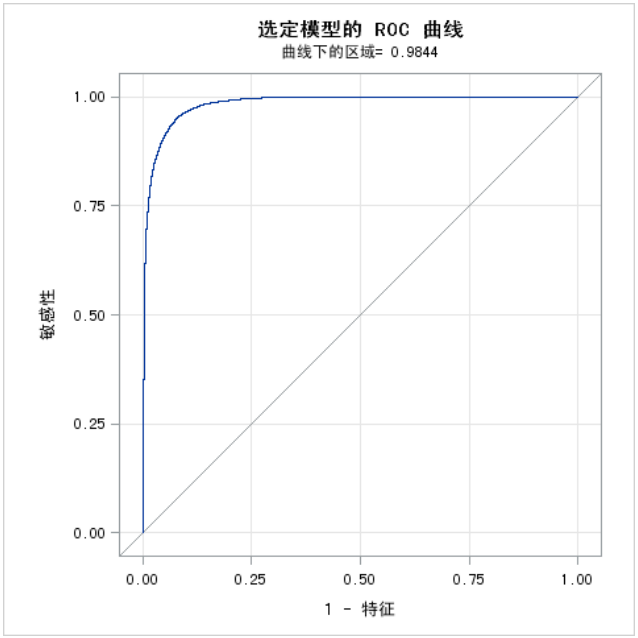
依据各变量的odds rates估计值和变量间相关关系，可知影响客户的重要变量有：

■。

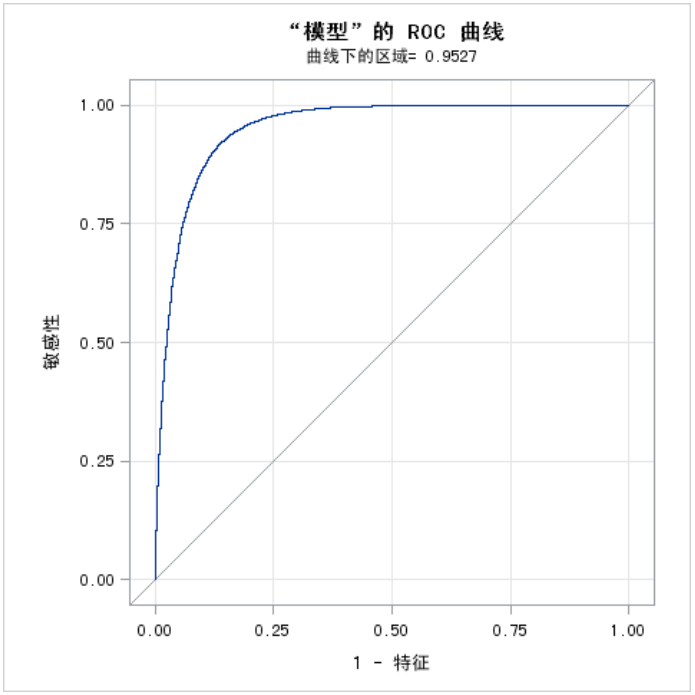
模型评估

* 验证方法
  * 样本内验证（In sample Validation）
    * 把建模样本分成两部分，如60%：40%，一部分用来建模，另一部分用来验证。
  * 样本外验证（Out of Sample Validation）
    * 把应用到建模样本以外的客户群进行验证，如：不同时间段的相似客户群。
  * 两种验证方法结合保证了模型的稳定性。

模型评估-ROC图



V.S.



逐步法:16个解释变量

因子分析法:8个解释变量

其他验证指标

* K-S曲线（图1）

国际上用K-S指标来衡量评估结果是否优于期望值，具体标准是，如果K-S大于40%，模型具有较好的预测功能，发展的模型具有成功的应用价值。

* 准确率和查全率（图2）
* Lorentz曲线（图3）

“捕获”的目标人数占目标总人数的百分比。

* 提升率、覆盖率（图4）

图1

K-S统计量代表了预测响应/非响应对象能达到的最大提升度（与随机判断比较）。

$$KS = \max_{s} |B(s) - G(s)|$$

rank	%ile for Combined Populatio	Total Account	Good	Bad	Score Range	Bad%	Cum Pct Bad	Cum Bad Ratio
0	1	2495	2263	232	474—594	9.30%	17.9%	9.3%
1	2	2451	2336	115	595—606	4.69%	26.8%	7.0%
2	3	2562	2475	87	607—614	3.40%	33.5%	5.8%
3	4	2642	2574	68	615—620	2.57%	38.7%	4.9%
4	5	2139	2084	55	621—624	2.57%	43.0%	4.5%
5	6	2443	2395	48	625—628	1.96%	46.7%	4.1%
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	17	2978	2955	23	654—655	0.77%	73.5%	2.2%
17	18	1567	1557	10	656—656	0.64%	74.3%	2.2%
18	19	3226	3210	16	657—658	0.50%	75.5%	2.1%
19	20	1596	1588	8	659—659	0.50%	76.2%	2.0%

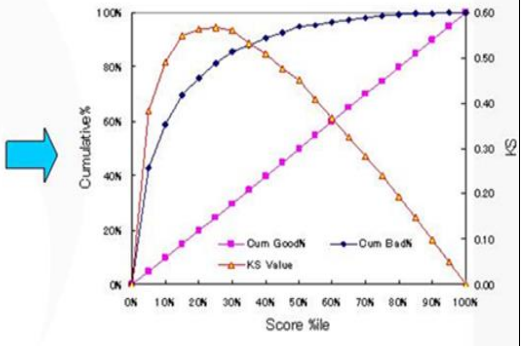


图2

		预测值 0	预测值 1
实际值	0	True Negative	False Positive
	1	True Negative	False Positive

准确率=TP/(FP+TP)

查全率（覆盖率）=TP/(FN+TP)

图3

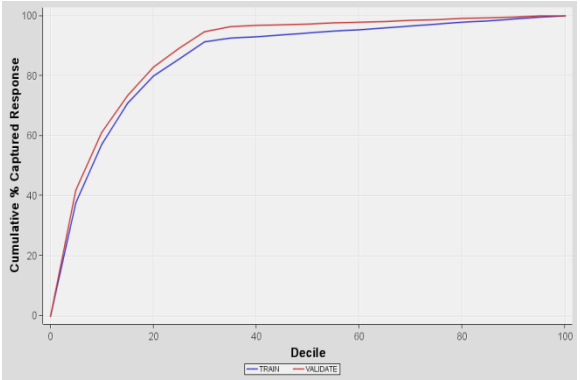
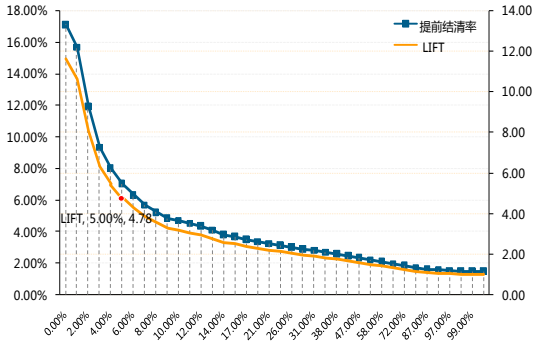


图4

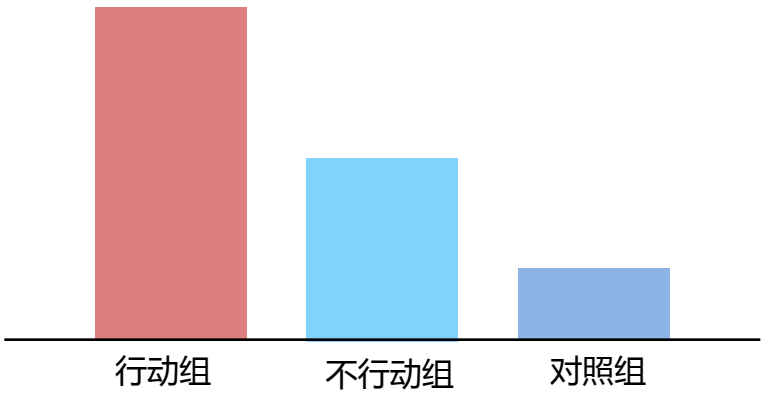


模型应用

- 业务应用
  - 制定客户维系营销策略
    - 对提前接清倾向评分较高的客户，根据其年龄、性别、职业特点进行相应的理财产品推荐。
    - 捆绑销售其它个人产品
    - 对提前接清倾向评分较高、价值较低的客户提供优惠。
  - 对其它相关产品开发提供支持

- 营销活动效果评估
  - 模型预测效果评估
  - 营销效果评估

不同组别的营销活动成功率



各组客户描述:

- 行动组:模型预测评分较高、参与营销活动。
- 不行动组:模型预测评分较高、不参与营销活动。
- 对照组:随机抽取、参与营销活动。

行动组和对照组营销效果对比体现了模型预测效果的优劣。
行动组营销效果与和不行动组办理业务情况对比体现了营销活动效果。

数据挖掘模型业务流程的应用


私人理财顾问


客户经理


低柜理财经理

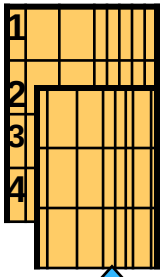

高柜柜员


外呼系统


网上银行

交易时提示销售信息

OCRM系统




市场管理人员
技术分析专家

营销
策略

销售结果反馈

销售过程监控

销售活动执行

销售渠道选择

销售任务创建

原型开发
模型训练

挖掘模型

模型评分

评分清单


数据分析人员

ECIS数据集市

数据转换
数据加载

EDW

销售目标设定


销售管理人员

目录

案例简介

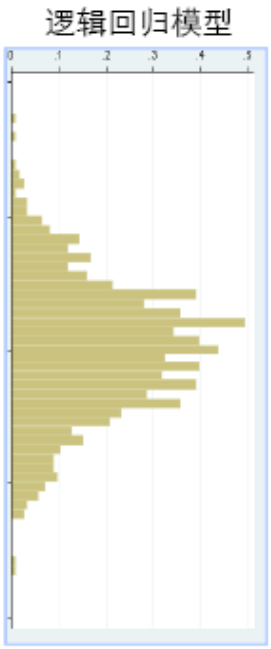
很多人都听说过美国的西南航空等公司的大名。这些公司成功的秘诀就是擅于提高上座率。在基于CRM数据库的营销方面，恰当的流失客户挽留工作是提高上座率的有效途径。

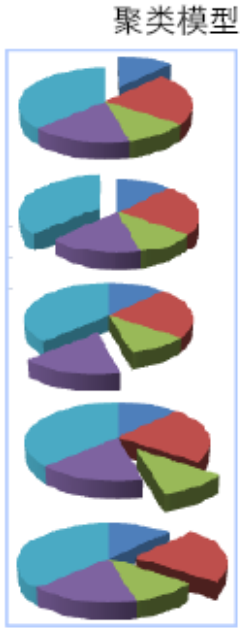
附件数据集来自国内某航空公司的会员数据，共有62988个样本，每个样本有63个属性，各属性说明见“变量含义”Sheet页。除了每个客户的基本资料外，该数据集还包含了一个观测窗（2年）内8个季度的用户飞行数据，包括乘机次数、里程、积分等。

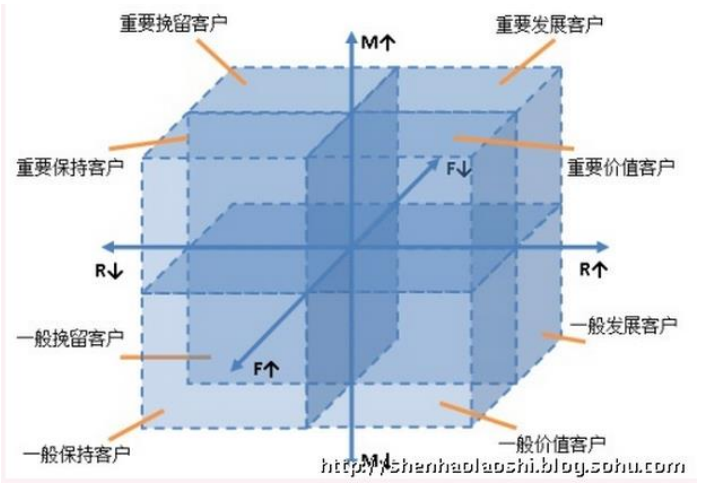
客户部门希望通过这些数据构建客户流失预警模型，。而且由于营销资源有限，希望结合客户特征进行有针对性地、高效率的开展客户挽留。

分析思路

这个任务分三个部分：
1、根据客户的属性、业务信息等预测客户流失概率；
2、根据类RFM模型对客户进行画像，区分客户的价值类型；3、讲两者结合指定有针对性的维护策略。







客户流失概率　　类RFM客户聚类模型　　有针对性地营销策略

流失倾向打分

* 背景和目标
* 数据探索
* 数据准备
* 模型构建
  * 业务分析
  * 变量选择（变量聚类）
  * 工具变量
* 模型评估
* 模型应用

客户分群

1、聚类分析的数据准备

变量和观测选择

分布分析

量纲剔除

2、聚类分析过程

聚类分析

聚类分析根据一批样品的许多观测指标，按照一定的数学公式具体地计算一些样品或一些参数(指标)的相似程度，把相似的样品或指标归为一类，把不相似的归为一类。

例如对上市公司的经营业绩进行分类；据经济信息和市场行情，客观地对不同商品、不同用户及时地进行分类。又例如当我们对企业的经济效益进行评价时，建立了一个由多个指标组成的指标体系，由于信息的重叠，一些指标之间存在很强的相关性，所以需要将相似的指标聚为一类，从而达到简化指标体系的目的。

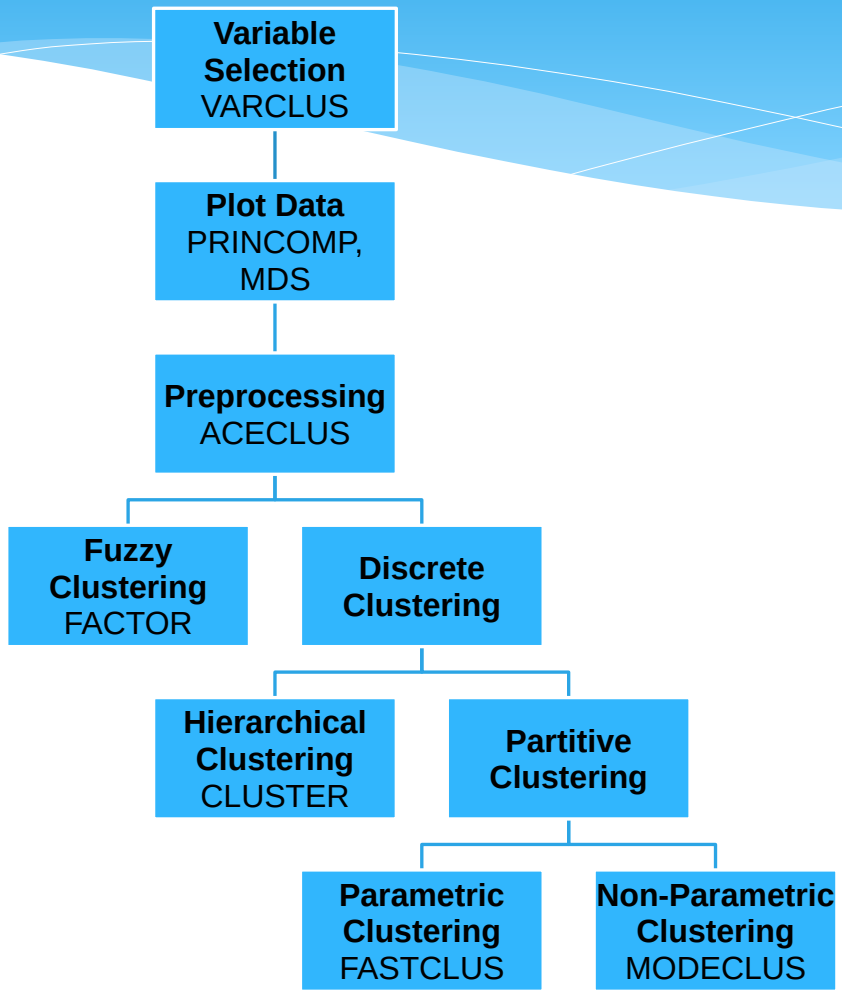
聚类分析数据准备

1. 数据和样本选择(*Who am I clustering?*)
2. 变量选择(*What characteristics matter?*)
3. 数据探索(*What shape/how many clusters?*)
4. 标准化(*Are variable scales comparable?*)
5. 数据变换(*Are variables correlated? Are clusters elongated?*)

示例：客户聚类分析

* 在众多客户当中，如果不进行分类，通常无法进行行为和习惯的分析.
  * Bargain hunter
  * Man/woman on a mission
  * Impulse shopper
  * Weary parent
  * DINK (dual income, no kids)

聚类分析概况



客户分群

根据航空公司独有的特点，对传统的RFM 指标做了适当的调整，确定了L、R、F、M、C 五个指标作为航空公司客户细分的参数。L 代表客户关系长度（FFP_days=LOAD_TIME-FFP_DATE），R 代表客户最近一次消费距今时间长度(DAYS_FROM_LAST_TO_END)，F 代表客户在一定时间内的消费频率(L1Y_Flight_Count)，M 代表客户在一定时间内的升级里程（L1Y_Flight_Count），C 代表客户在一定时间内所乘航班的平均舱位折扣系数（avg_discount)

聚类结果分析

根据每个类中变量的取值情况，为每个类取名称

聚类均值						
聚类	入会之日算	最近一次消费距今时间长度	代表客户在一定时间内的消费频率	在一定时间内的升级里程	在一定时间内所乘航班的平均舱位折扣系	客户类型
1	-0.60	-0.92	1.31	1.33	0.34	重要价值
2	-0.45	-0.12	-0.05	-0.35	-1.19	重要发展
3	1.19	-1.34	1.74	1.75	0.63	重要价值
4	-1.23	0.36	-0.37	-0.21	0.49	一般发展
5	0.19	1.27	-1.29	-1.18	1.05	一般挽留
6	1.26	-0.12	0.26	0.25	-0.18	重要保持
7	-0.22	1.34	-1.37	-1.35	-0.88	一般保持
8	0.05	-0.49	0.25	0.33	0.39	一般价值

分类流失预测结果

输入数据 | 代码 | 日志 | 输出数据

修改任务(Y) | 过滤和排序(L) | 查询生成器(Q) | 数据

	MEMBER_NO	CLUSTER	IP_1
1	'00024465	重要价值	0.9828248195
2	'00051394	重要价值	0.967715571
3	'00051864	重要价值	0.9656044766
4	'00061117	重要价值	0.9632355981
5	'00043193	重要价值	0.9489573518
6	'00022396	重要价值	0.9367048326
7	'00025327	重要价值	0.9297742172
8	'00023777	重要价值	0.9098395947
9	'00023289	重要价值	0.9047079114
10	'00019962	重要价值	0.8999916189
11	'00042627	重要价值	0.8799112879
12	'00051523	重要价值	0.8654337002
13	'00050910	重要价值	0.8650911579
14	'00022629	重要价值	0.8520828411
15	'00035299	重要价值	0.8488855213
16	'00005872	重要价值	0.8137244076
17	'00024935	重要价值	0.8072396517

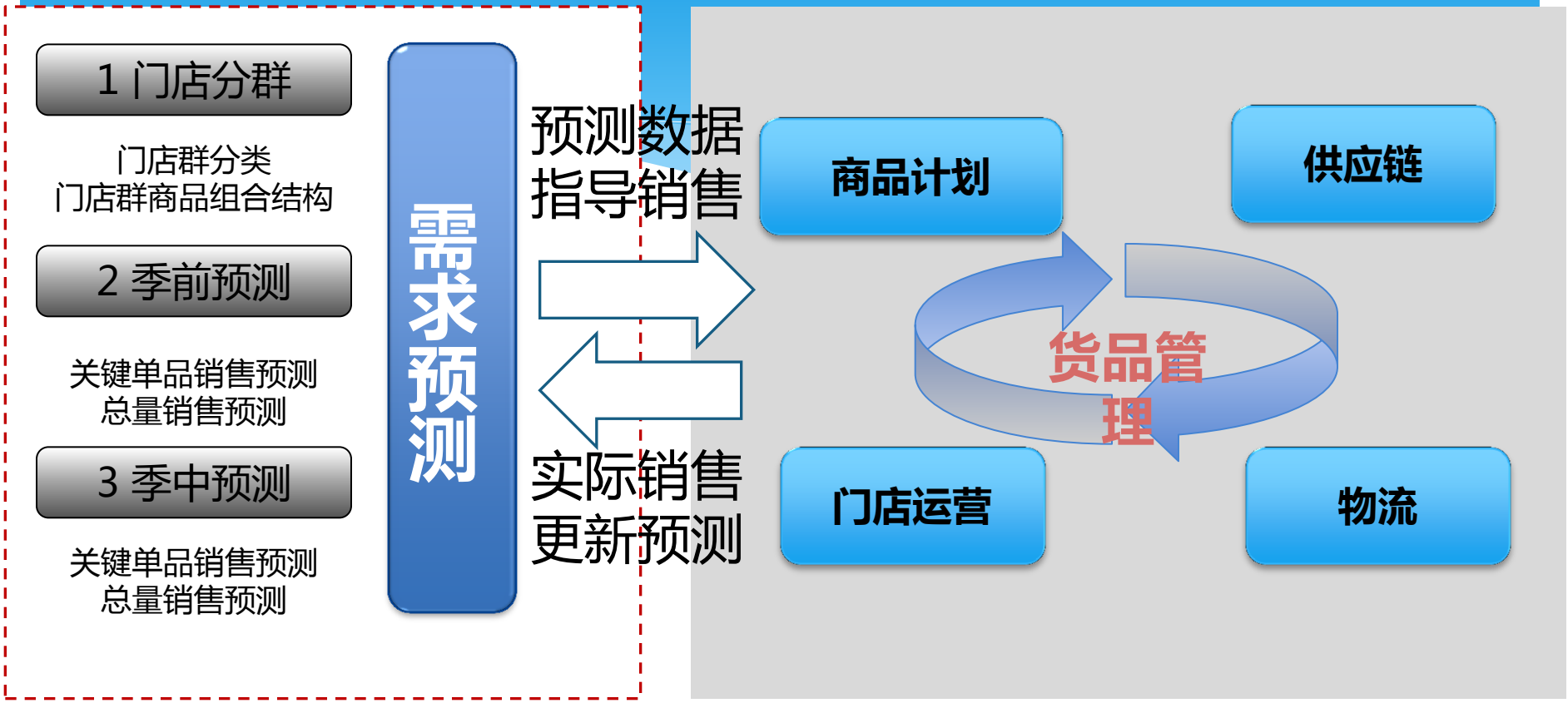
目录

案例简介

随着国家运动总局大力推行全民健身计划，到2020将会有40%的国民积极参加各类运动活动，运动消费需求大幅上涨，运动零售市场的发展潜力巨大。

自2010年以来，我国运动品行业遭遇了史上最为严重的业绩下滑和高库存积累，相关企业纷纷通过密集关店的方式谋求生存。仅2012年一年，李宁、安踏、特步等6家行业上市公司累计关闭门店数达5000家以上。这种现象产生的原因是相关公司未能判断运动产品的生命周期，未能对销售需求及销售量进行合理的预测，造成了库存增加、利润下滑、积累高库存、出现亏损等一系列连锁反应，这也是我国体育用品企业目前普遍面对的问题。

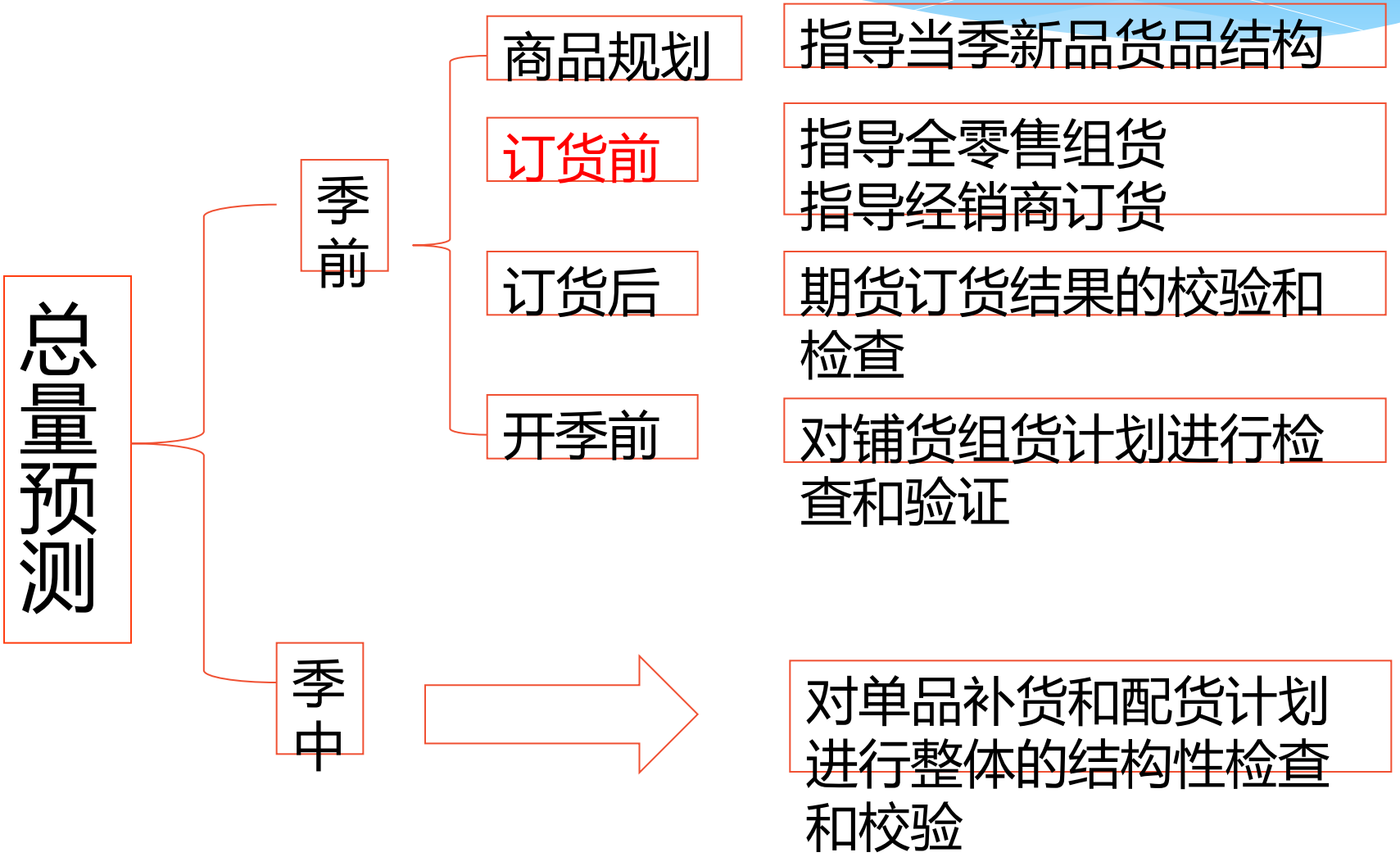
分析思路



需求预测模型包含以下三部分功能：

1. 门店分群
2. 季前预测（总量、关键单品）
3. 季中预测（总量、关键单品）

业务分析框架


总量预测
季前
季中
商品规划
订货前
订货后
开季前
指导当季新品货品结构
指导全零售组货
指导经销商订货
期货订货结果的校验和检查
对铺货组货计划进行检查和验证
对单品补货和配货计划进行整体的结构性检查和校验

模型构建的方法

针对总量预测模型，主要采用时间序列的方法的方法进行到产品类别的销售总量预测；对于新品SKU销售量，我们会采用符合公司销售特点的扩散模型（Bass模型，Gompertz 模型或者Logistic模型）。

模型构建的方法

Bass模型

$$N_t = N_{t-1} + p\left(m - N_{t-1}\right) + q\,\frac{N_{t-1}}{m}\left(m - N_{t-1}\right)$$

模型引入三个参量来预测N（消费者在第n期购买该产品的数量）：

m=市场潜力， 即潜在使用者总数。

p=创新系数(外部影响)， 即尚未使用该产品的人，受到大众传媒或其他外部因素的影响，开始使用该产品的可能性。

q=模仿系数(内部影响)， 即尚未使用该产品的人，受到使用者的口碑影响，开始使用该产品的可能性。

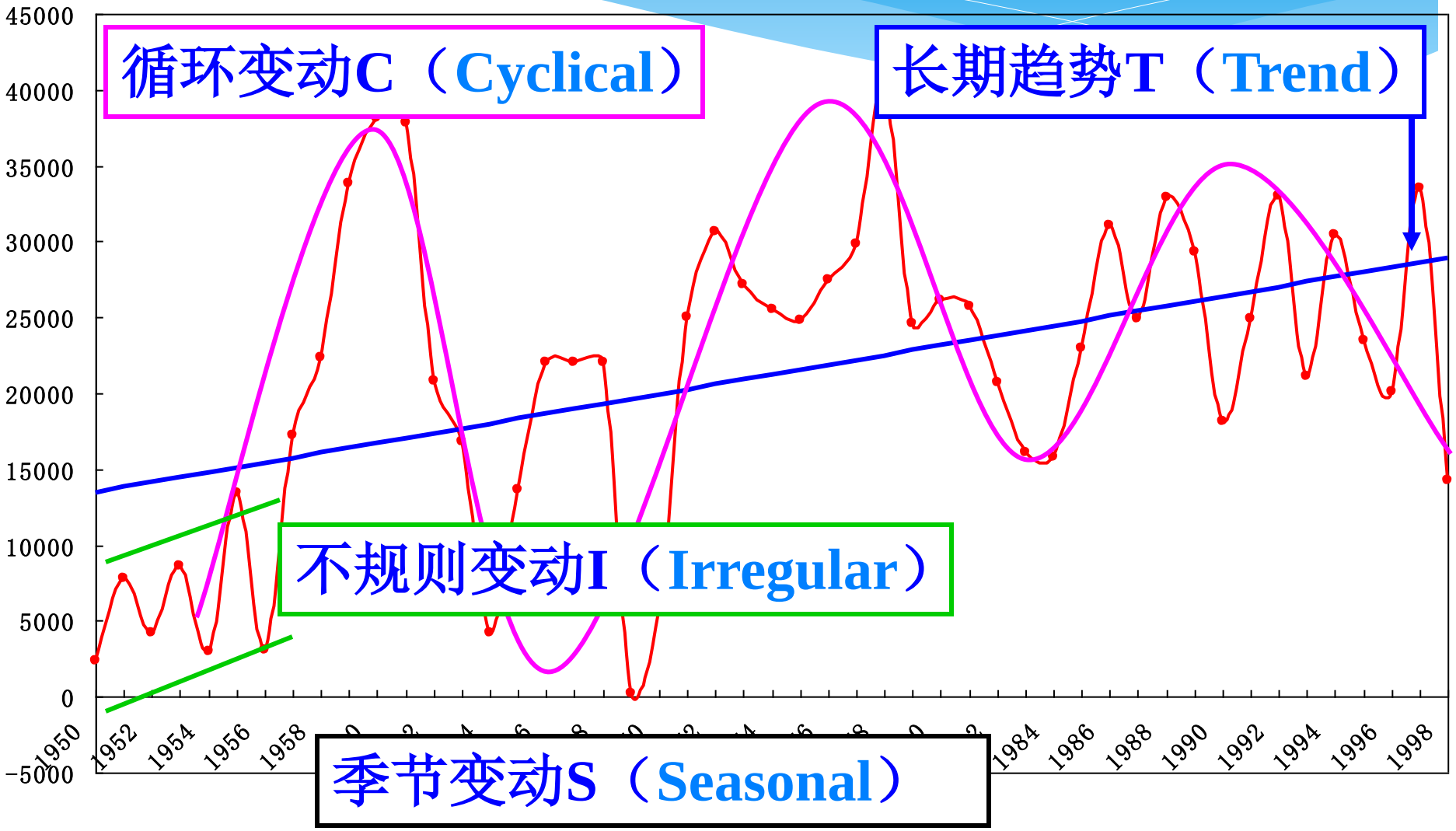
模型构建的方法

时间序列模型

ARIMA模型全称为求和自回归移动平均模型(Autoregressive Integrated Moving Average Model,简记ARIMA)，是由博克思(Box)和詹金斯(Jenkins)于70年代初提出的一著名时间序列预测方法，所以又称为box-jenkins模型、博克思-詹金斯法。其中ARIMA（p，d，q）称为差分自回归移动平均模型，AR是自回归, p为自回归项; MA为移动平均，q为移动平均项数，d为时间序列成为平稳时所做的差分次数。

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

1950-1998年中国水灾受灾面积(单位：千公顷)



随机时间序列分析

* 时域模型
  积分自回归滑动平均模型（ARIMA）

* 频域模型
  傅立叶（Flourier）分析，小波（Wavelets）分析

随机过程的概率结构分类

- 独立随机过程
- 独立增量随机过程
- 马尔可夫过程
- 平稳随机过程

（1）独立随机过程

设$\{X(t)，\ t\in T\}$对任意 n 个不同的t_1，t_2，…，$t_n\in T$

$X(t_1)$，$X(t_2)$，…，$X(t_n)$ 是相互独立的

则称 $X(t)$ 为具有独立随机变量的随机过程，

简称独立随机过程。

（2）独立增量随机过程

设$\{X(t),\ t\in T\}$对任意 n 个不同的$t_1,\ t_2,\ \cdots,\ t_n\in T$

且$t_1<t_2<\cdots<t_{n-1}<t_n$

$X(t_2)-X(t_1),\ X(t_3)-X(t_2),\ \cdots,\ X(t_n)-X(t_{n-1})$

是相互独立的，

则称$X(t)$为具有独立增量的随机过程。

例 甲、乙两路公共汽车都通过某一车站，两路汽车的到达分别服从10分钟1辆（甲），15分钟1辆（乙）的泊松分布。假定车总不会满员，试问可乘坐甲或乙两路公共汽车的乘客在此车站所需等待时间的概率分布及其期望。

解 反映甲、乙两路公共汽车到达情况的泊松分布

$X_1(t)$ 和 $X_2(t)$ 的生起率分别为 $\lambda_1 = 1/10$，$\lambda_2 = 1/15$

下面证明两路车混合到达过程 $X(t)$ 服从生起率为

$\lambda = \lambda_1 + \lambda_2$ 的泊松分布

事实上 $X(t)=X_1(t)+X_2(t)$ 是独立增量

且 $X(t+s)-X(t)$ 是相互独立地服从泊松分布的随机变量

$X_1(t+s)-X_1(t)$ 及 $X_2(t+s)-X_2(t)$ 的和，

所以由泊松过程的定义可知

$X(t)$ 服从均值为 λs 的泊松分布。

因此 $X(t)$ 服从生起率为 $\lambda=1/10+1/15=1/6$ 的泊松过程。

由定理1知公共汽车的到达时间间隔服从均值为6分钟的指数分布。 再由指数分布的无记忆性，

这位乘客的等待时间也服从均值为6分钟的指数分布。

（3）马尔可夫过程

设$\{X(t)，\ t \in T\}$对任意 n 个不同的t_1 ，t_2 ，…，$t_n \in T$

且$t_1 < t_2 < \cdots < t_{n-1} < t_n$

$$P(X(t_n) \le x_n \mid X(t_{n-1}) = x_{n-1}，\ \cdots，\ X(t_1) = x_1)$$

$$= P(X(t_n) \le x_n \mid X(t_{n-1}) = x_{n-1})，$$

则称$X(t)$为马尔可夫过程

简称马氏过程。

马氏过程的特点

当随机过程在时刻t_{n-1}的状态已知的条件下，

它在时刻t_n（$t_n > t_{n-1}$）所处的状态

仅与时刻t_{n-1}的状态有关，

而与过程在时刻t_{n-1}以前的状态无关

称这个特性为马尔可夫性，简称马氏性。

马氏性实质上是无后效性，所以也称马氏过程为无后效过程。

例 市场占有率预测

设某地有1600户居民，某产品只有甲、乙、丙3厂家在该地销售。经调查，8月份买甲、乙、丙三厂的户数分别为480，320，800。9月份里，原买甲的有48户转买乙产品，有96户转买丙产品；原买乙的有32户转买甲产品，有64户转买丙产品；原买丙的有64户转买甲产品，有32户转买乙产品。用状态1、2、3分别表示甲、乙、丙三厂，试求

（1）转移概率矩阵；
（2）9月份市场占有率的分布；
（3）12月份市场占有率的分布；
（4）当顾客流如此长期稳定下去市场占有率的分布。

解 （1） 由题意得频数转移矩阵为

$$N=\begin{bmatrix}336 & 48 & 96\\ 32 & 224 & 64\\ 64 & 32 & 704\end{bmatrix}$$

再用频数估计概率，得转移概率矩阵为

$$P_1=\begin{bmatrix}0.7 & 0.1 & 0.2\\ 0.1 & 0.7 & 0.2\\ 0.08 & 0.04 & 0.88\end{bmatrix}$$

（2）以1600除以N中各行元素之和，得初始概率分布（即初始市场占有率）

$$P(0)=(p_1,p_2,p_3)=(0.3 \quad 0.2 \quad 0.5)$$

所以9月份市场占有率分布为

$$P(1)=P(0)P_1=(0.3\quad 0.2\quad 0.5)\begin{bmatrix}0.7 & 0.1 & 0.2\\ 0.1 & 0.7 & 0.2\\ 0.08 & 0.04 & 0.88\end{bmatrix}$$

$$=(0.27\quad 0.19\quad 0.54)$$

（3）12月份市场占有率分布为

$$P(4)=P(0)P_1^4$$

$$=(0.3\quad 0.2\quad 0.5)\begin{bmatrix}0.7 & 0.1 & 0.2\\ 0.1 & 0.7 & 0.2\\ 0.08 & 0.04 & 0.88\end{bmatrix}^4$$

$$=(0.2319\quad 0.1698\quad 0.5983)$$

（4）由于该链不可约、非周期、状态有限正常返的，所以是遍历的。

解方程组

$$\begin{cases} \pi_1 = 0.7\pi_1 + 0.1\pi_2 + 0.08\pi_3 \\ \pi_2 = 0.1\pi_1 + 0.7\pi_2 + 0.04\pi_3 \\ \pi_3 = 0.2\pi_1 + 0.2\pi_2 + 0.88\pi_3 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases}$$

即得当顾客流如此长期稳定下去是市场占有率的分布为

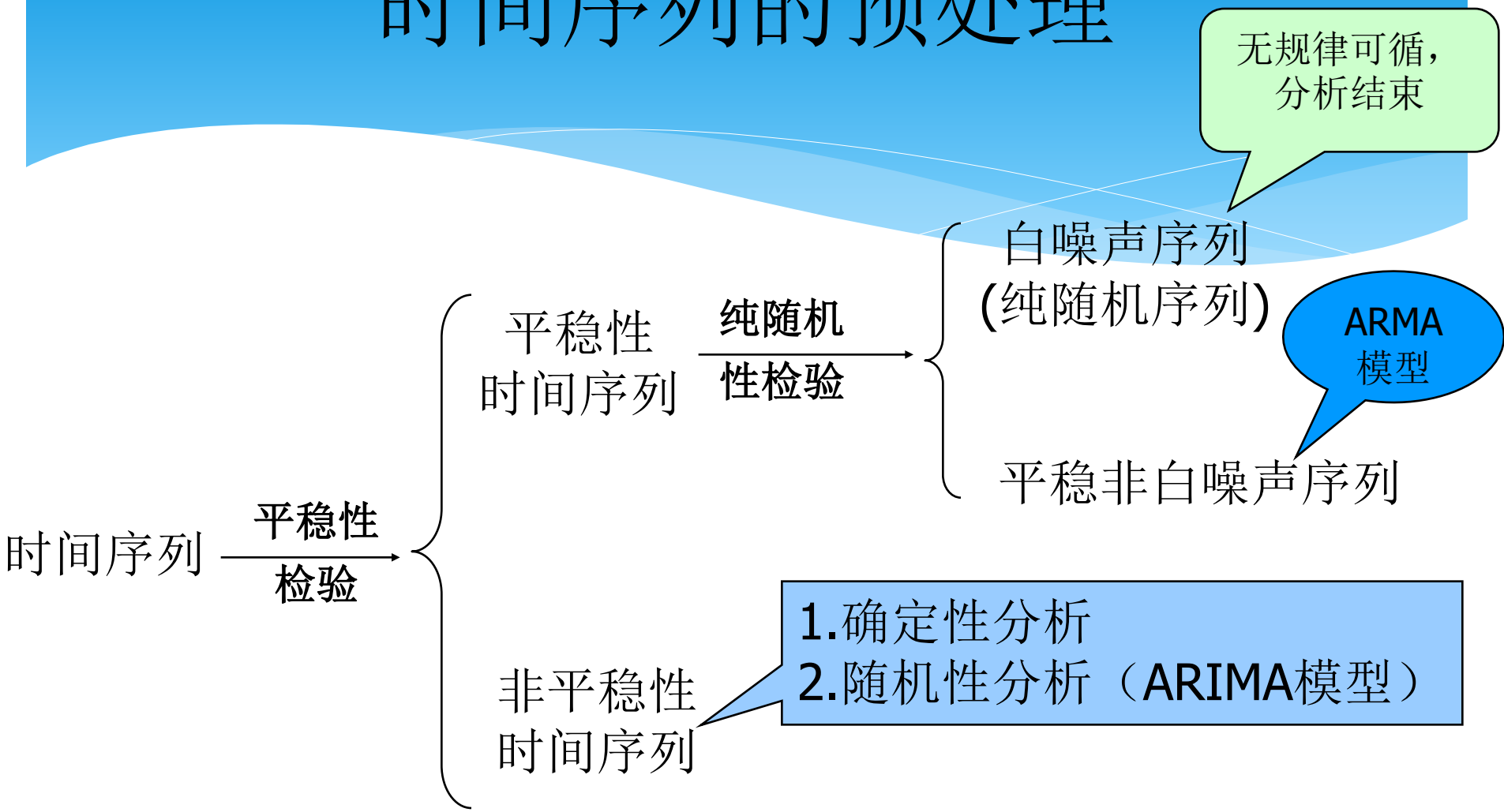
$$(\pi_1, \pi_2, \pi_3) = (0.219 \quad 0.156 \quad 0.625)$$

（4）平稳随机过程

平稳过程的统计特性与马氏过程不同，它不随参数的推移而变化，过程的“过去”可以对“未来”有不可忽视的影响。

在实际中，有相当多的随机过程，不仅它现在的状态，而且它过去的状态，都对未来状态的发生有着很强的影响.

时间序列的预处理



平稳时间序列的意义

* 时间序列数据结构的特殊性
  * 可列的多个随机变量，而每个变量只有一个样本观察值
* 平稳性的重大意义
  * 极大地减少了随机变量的个数，并增加了待估变量的样本容量
  * 极大地简化了时序分析的难度，减少了待估参数的个数

模型

* AR模型（Auto Regression Model）
* MA模型（Moving Average Model）
* ARMA模型（Auto Regression Moving Average model）

AR模型的定义

* 具有如下结构的模型称为 p阶自回归模型，简记为 $AR(p)$

$$\begin{cases} x_t = \varphi_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t \\ \varphi_p \neq 0 \\ E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E x_s \varepsilon_t = 0, \forall s < t \end{cases}$$

自回归模型(Auto regressive model,AR)

(一).一阶自回归模型，AR(1)

1.设$\{x_t\}$为零均值的平稳过程，如果关于x_t的合适模型为:

$$x_t = \varphi_1 x_{t-1} + \varepsilon_t$$

其中：(1) ε_t是白噪声序列（$E\,\varepsilon_t=0, Var(\varepsilon_t)=\sigma^2$, $cov(\varepsilon_t, \varepsilon_{t+k})=0$,$k\neq 0$）,(2)假定：$E(x_t, \varepsilon_s)=0$ $(t<s)$，那么我们就说x_t遵循一个一阶自回归或AR(1)随机过程。

（二）二阶自回归模型，AR(2)

* 1.设{x_t}为零均值的平稳过程，如果关于x_t的合适模型为

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \varepsilon_t$$

其中：(1) ε_t是白噪声序列,(2)假定：$E(x_t, \varepsilon_s)=0$ $(t<s)$，那么我们就说x_t遵循一个二阶自回归或AR(2)随机过程。上述模型就是AR(2)模型。

MA模型的定义

* 具有如下结构的模型称为 q阶自回归模型，简记为

$$MA(q)$$

$$\begin{cases} x_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \\ \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s) = 0, s \neq t \end{cases}$$

移动平均模型（moving average model,MA）

* （一）一阶移动平均模型，MA(1)
* 如果关于x_t（假设同前）的合适的模型如下：

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

其中：ε_t为白噪声序列，那么就称x_t满足一阶移动平均模型，记作MA(1)

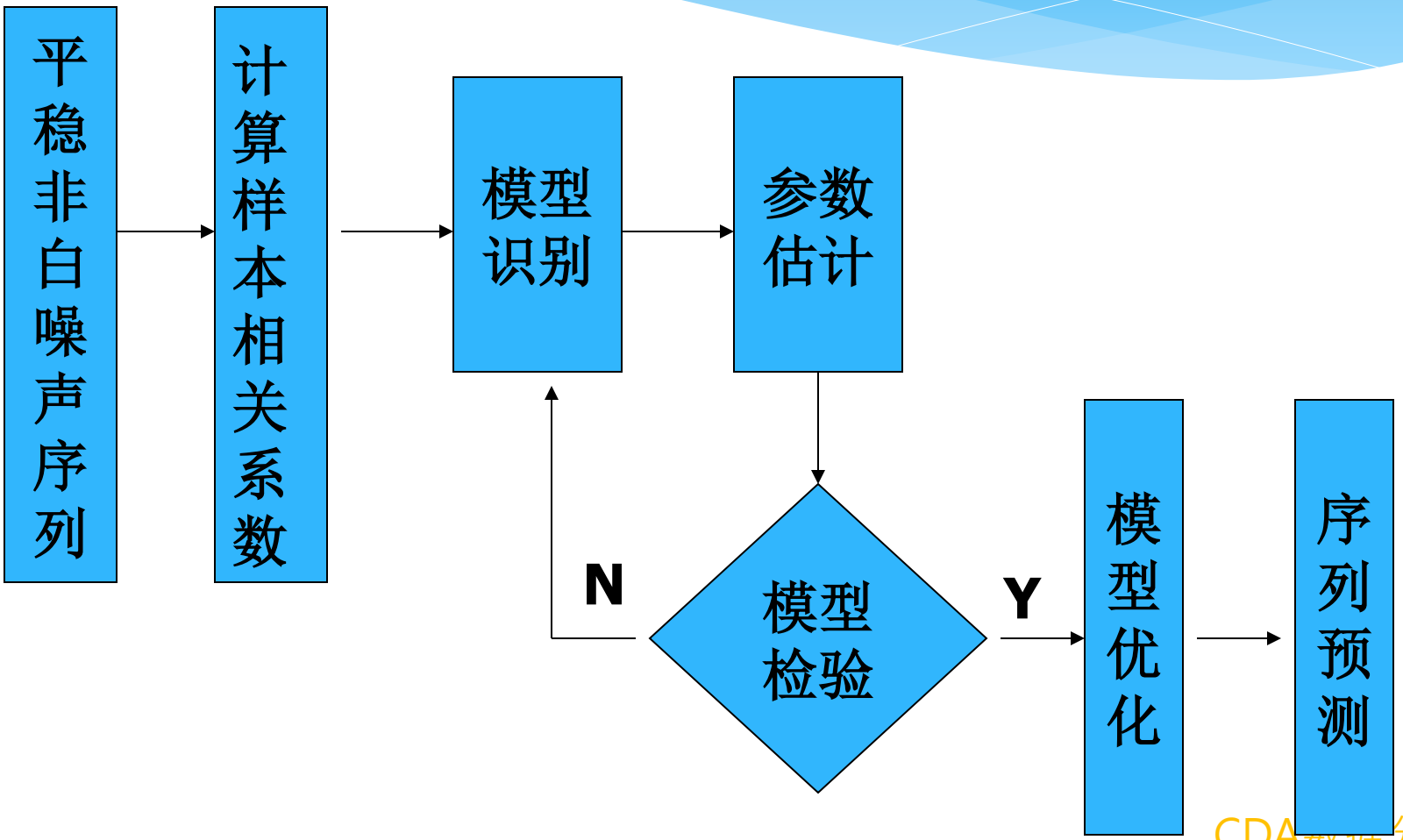
ARMA模型的定义

* 具有如下结构的模型称为自回归移动平均模型，简记为

$$ARMA(p,q)$$

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \quad \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ Ex_s \varepsilon_t = 0, \forall s < t \end{cases}$$

建模步骤



模型的平稳性

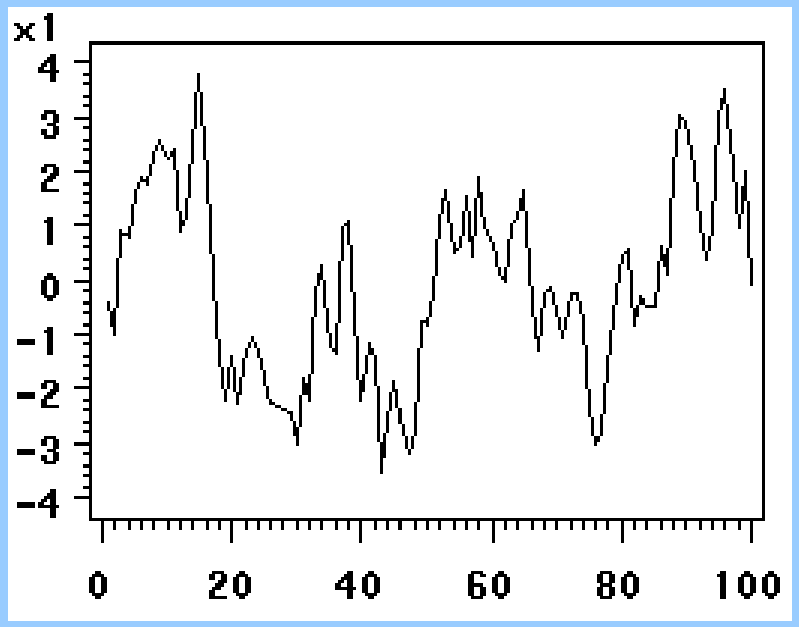
$$(1)x_t = 0.8x_{t-1} + \varepsilon_t$$

$$(2)x_t = -1.1x_{t-1} + \varepsilon_t$$

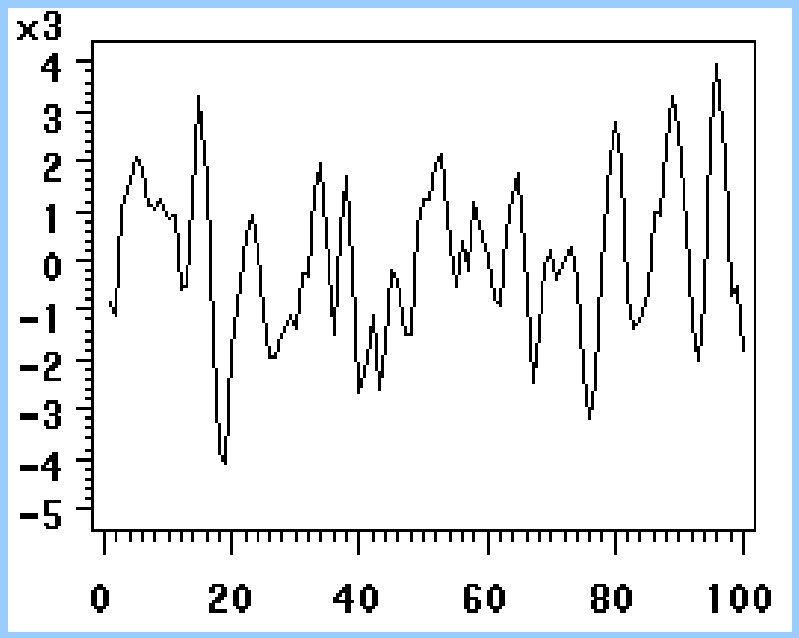
$$(3)x_t = x_{t-1} - 0.5x_{t-2} + \varepsilon_t$$

$$(4)x_t = x_{t-1} + 0.5x_{t-1} + \varepsilon_t$$

平稳序列时序图

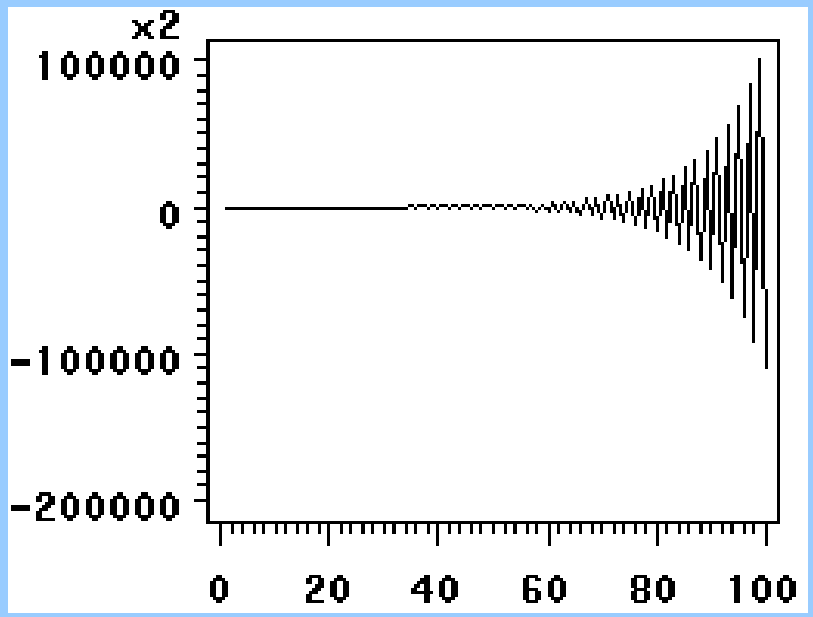


$$(1)x_t = 0.8x_{t-1} + \varepsilon_t$$

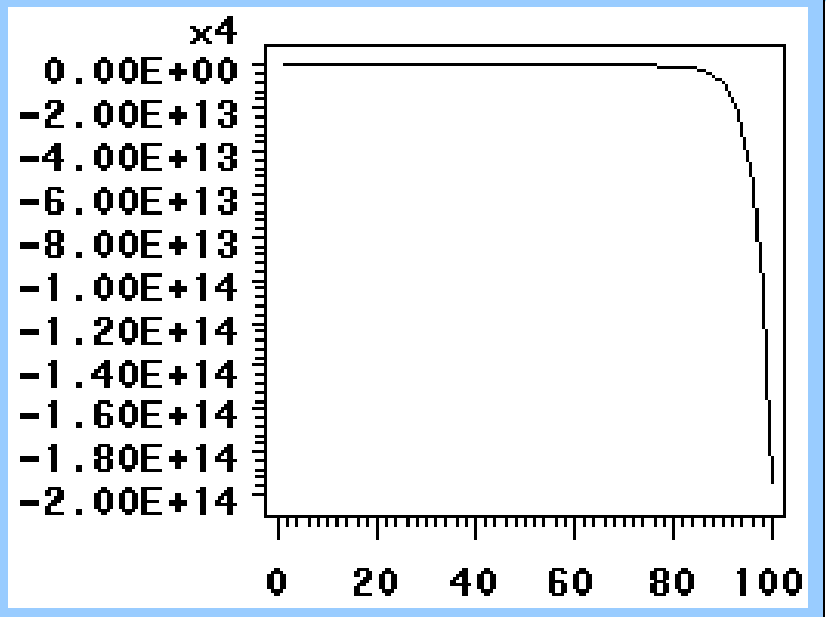


$$(3)x_t = x_{t-1} - 0.5x_{t-2} + \varepsilon_t$$

非平稳序列时序图



$$(2)x_t = -1.1x_{t-1} + \varepsilon_t$$



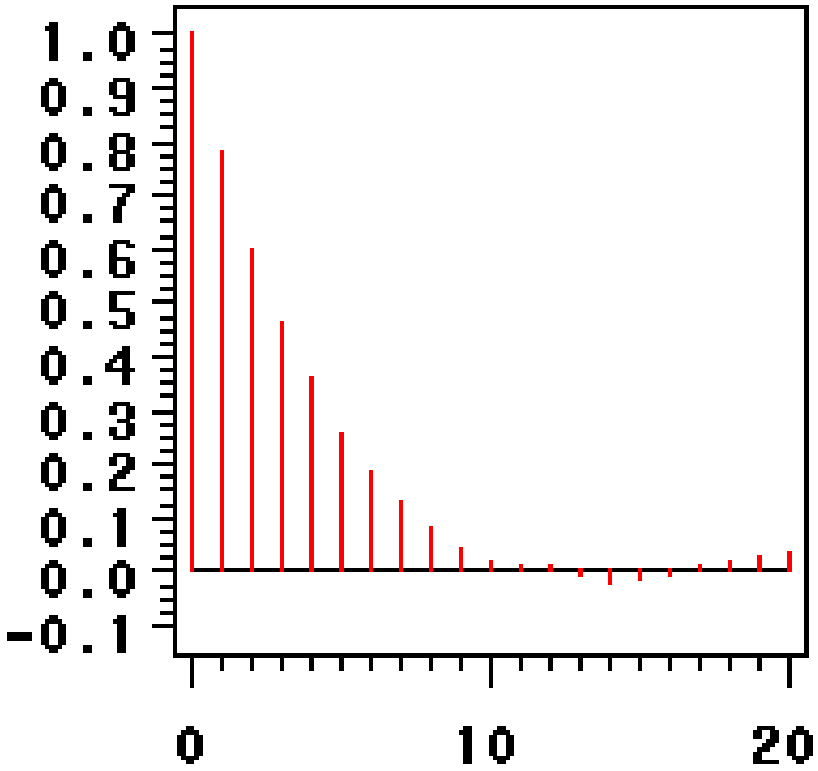
$$(4)x_t = x_{t-1} + 0.5x_{t-1} + \varepsilon_t$$

平稳模型的统计性质

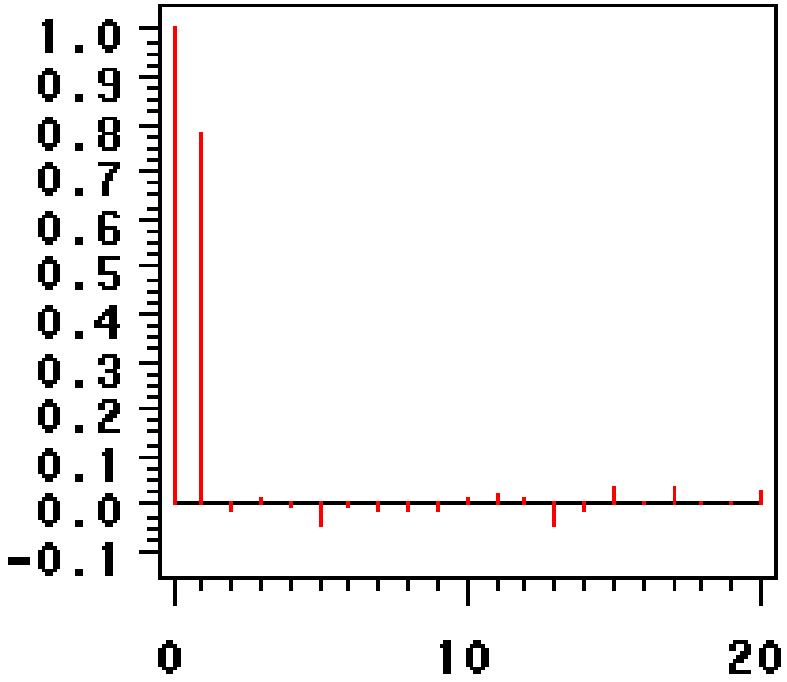
* 均值
* 方差
* 协方差
* 自相关系数
* 偏自相关系数

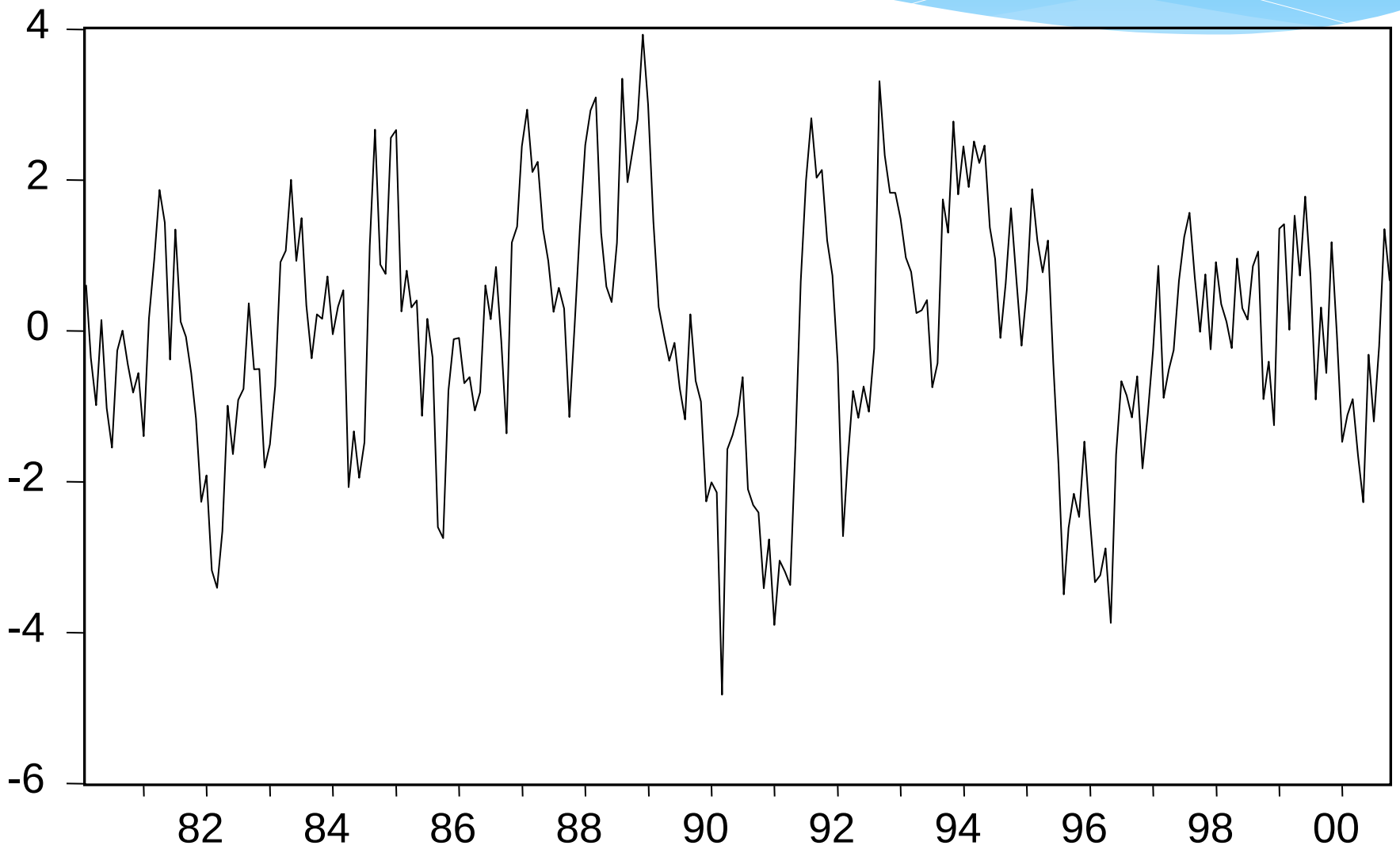
$$(1) x_t = 0.8 x_{t-1} + \varepsilon_t$$

* 自相关系数



偏自相关系数

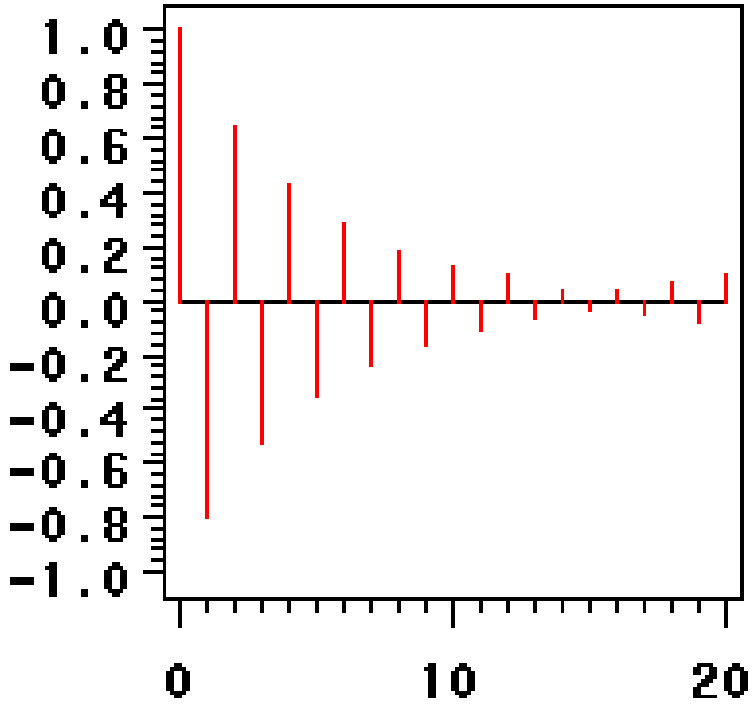




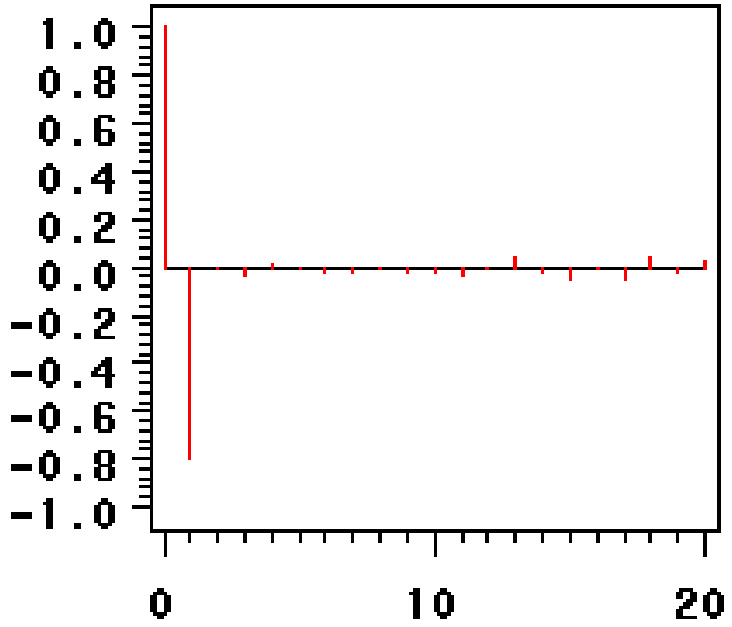
模拟生成的AR(1)过程趋势图

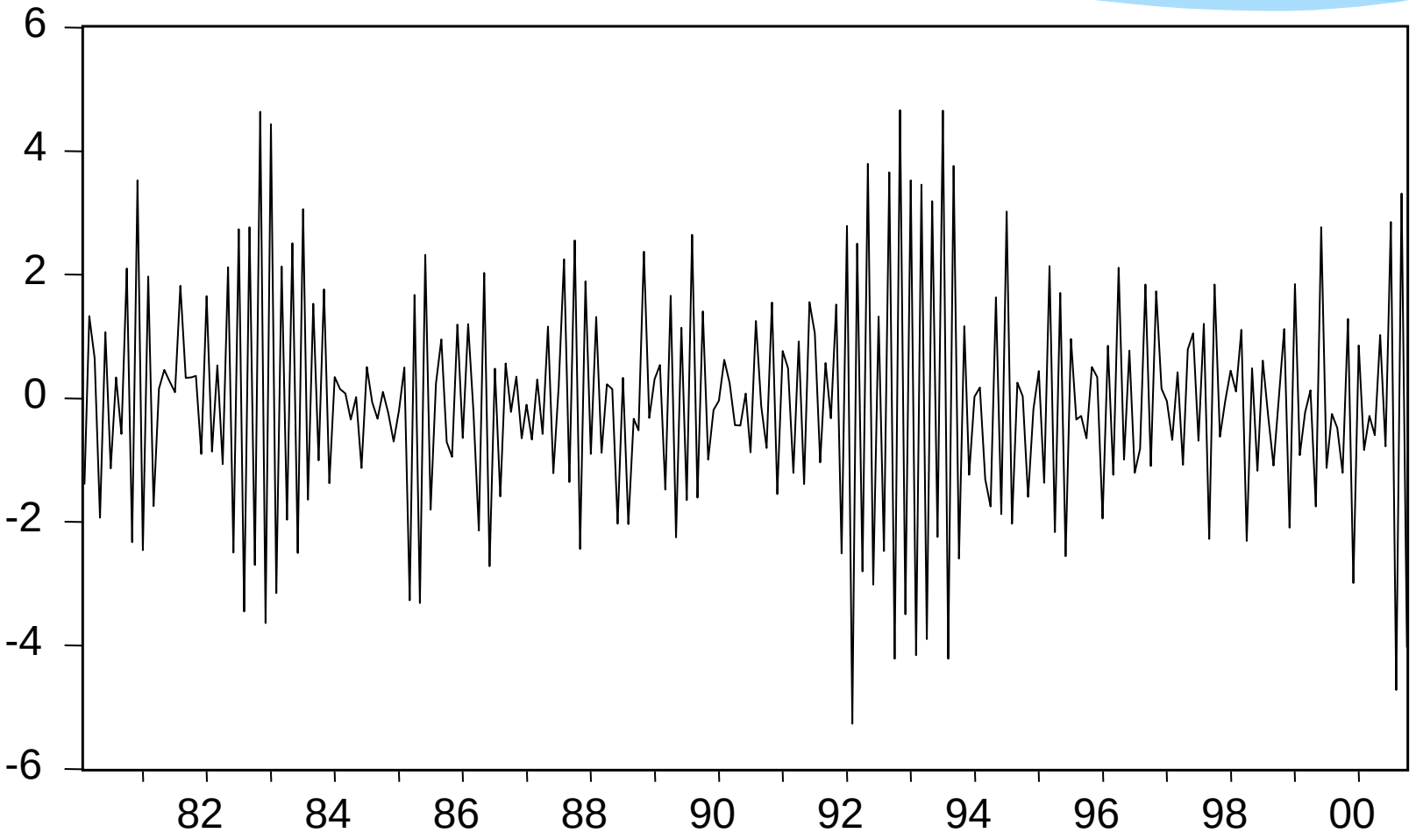
$$(2)x_t = -0.8x_{t-1} + \varepsilon_t$$

自相关系数



偏自相关系数

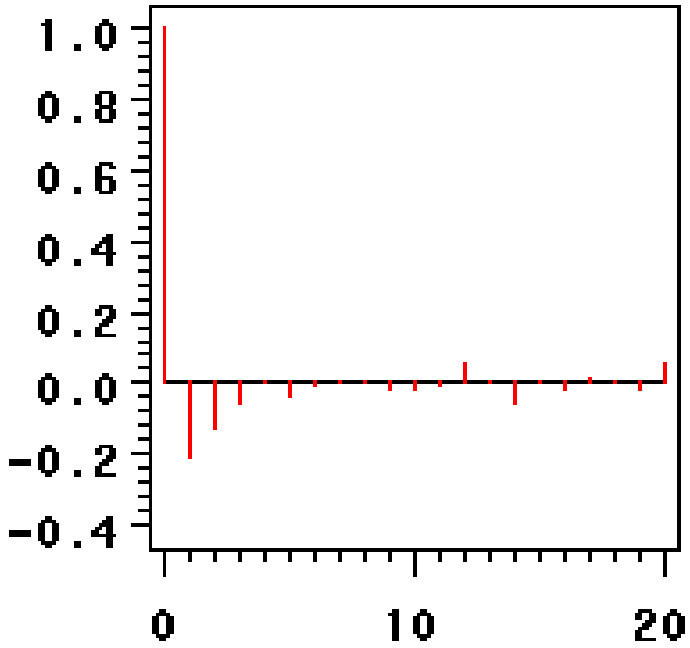




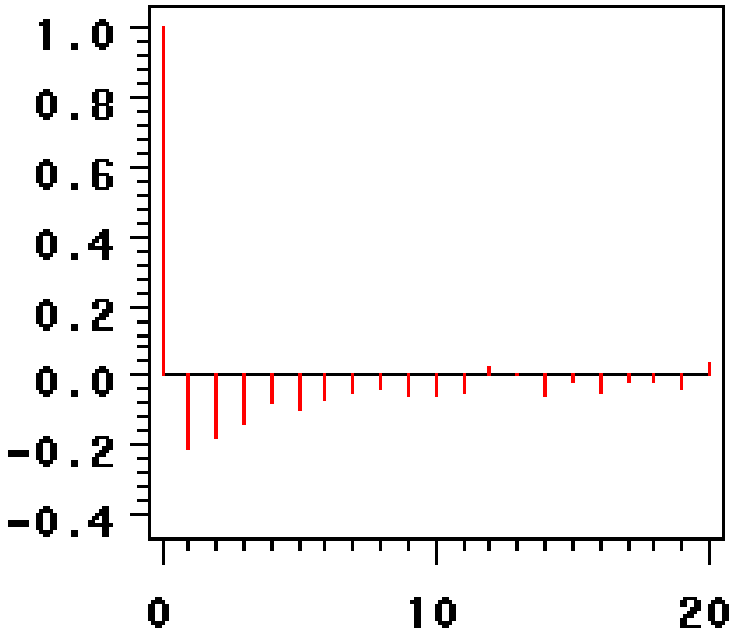
模拟生成的AR(1)过程趋势图

ARMA模型自相关系数和偏自相关系数拖尾性

- 样本自相关图



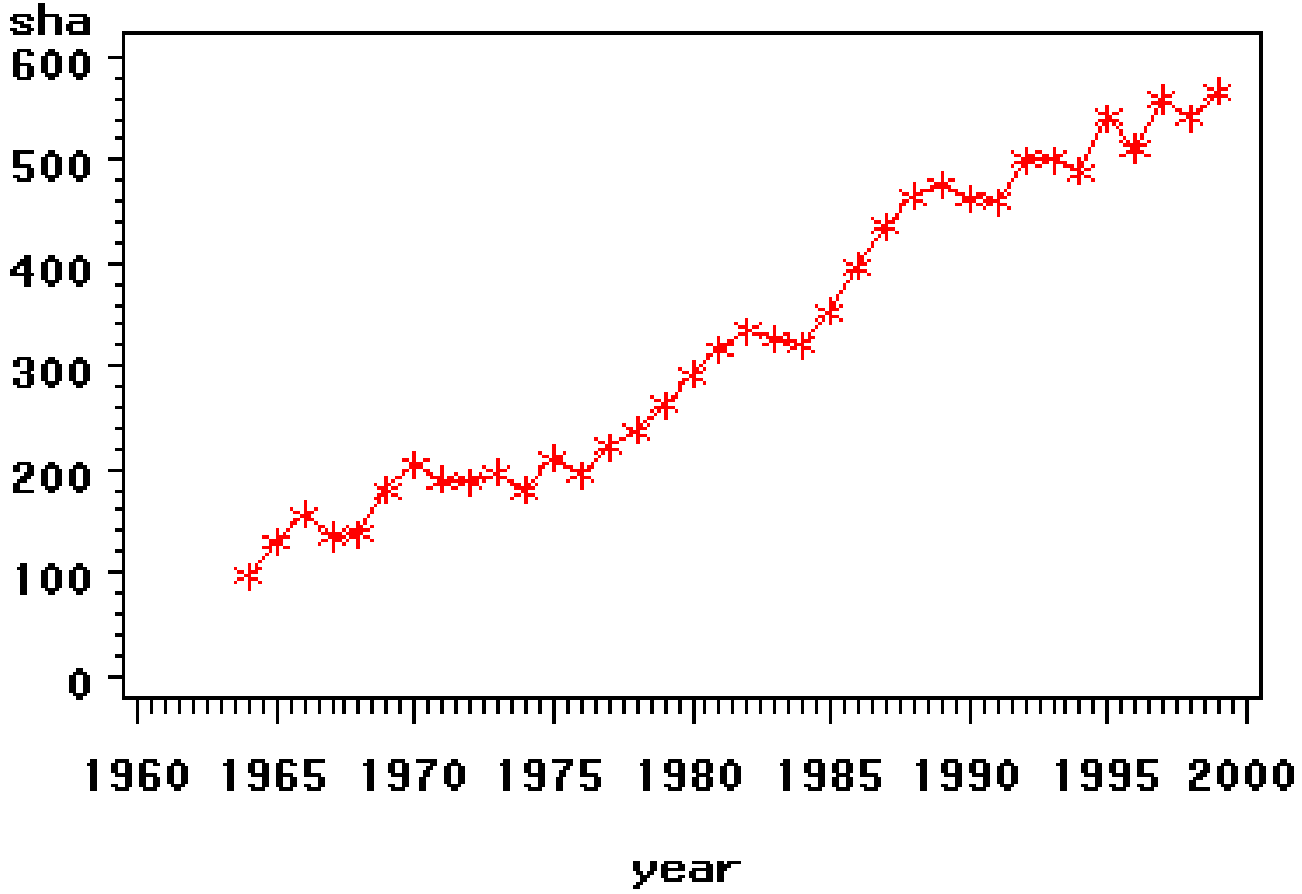
- 样本偏自相关图



ARMA模型相关性特征

模型	自相关系数	偏自相关系数
AR(P)	拖尾	P阶截尾
MA(q)	q阶截尾	拖尾
ARMA(p,q)	拖尾	拖尾

非平稳序列时序图



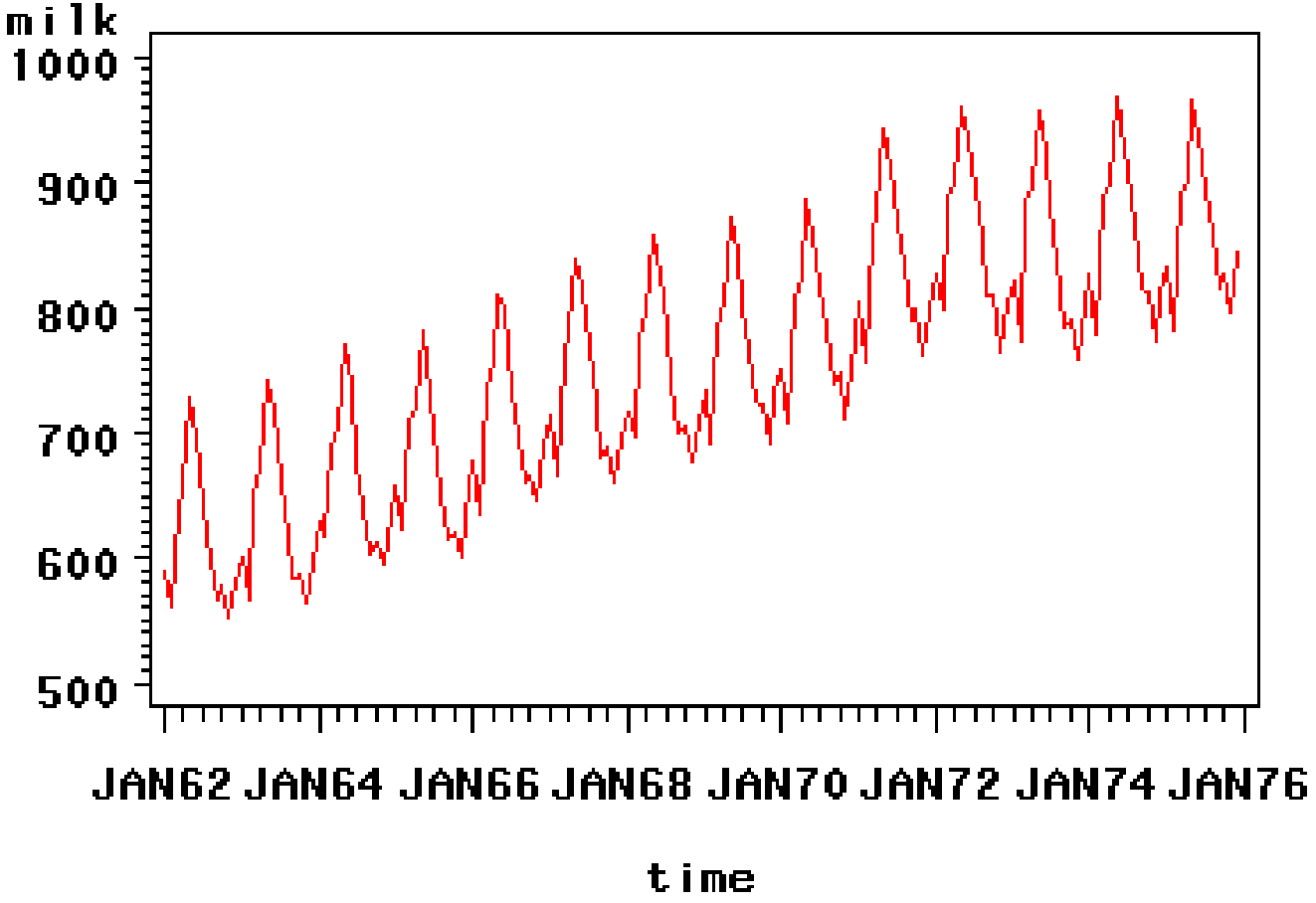
自相关图

Autocorrelations

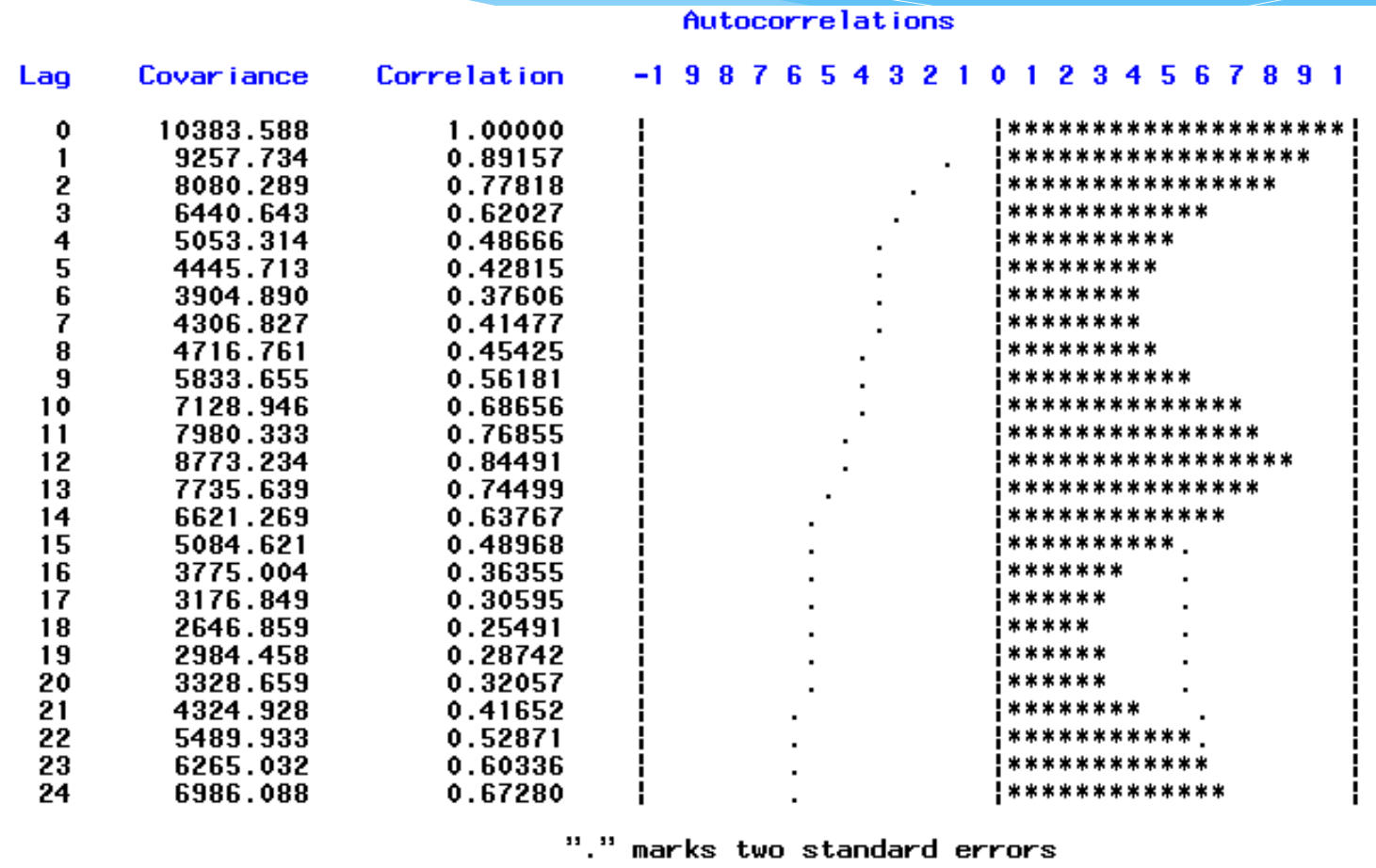
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	21741.103	1.00000	|**********|
1	19869.670	0.91392	|********** . |
2	18336.945	0.84342	|********** . |
3	16679.644	0.76719	|********** . |
4	15119.827	0.69545	|********** . |
5	13234.768	0.60874	|********** . |
6	11822.365	0.54378	|********** . |
7	10355.425	0.47631	|********** . |
8	8597.171	0.39543	|******** . |
9	6977.227	0.32092	|****** . |
10	5262.589	0.24206	|***** . |
11	3185.458	0.14652	|*** . |
12	1257.065	0.05782	|* . |
13	-717.129	-.03298	| * . |
14	-2356.762	-.10840	| ** . |
15	-3657.864	-.16825	| *** . |
16	-4675.021	-.21503	| **** . |
17	-5645.938	-.25969	| ***** . |
18	-6662.959	-.30647	| ****** . |
19	-7523.279	-.34604	| ******* . |
20	-8300.856	-.38180	| ******** . |
21	-9068.912	-.41713	| ******** |
22	-9409.375	-.43279	| ********* |

"." marks two standard errors

非平稳时序图



自相关图

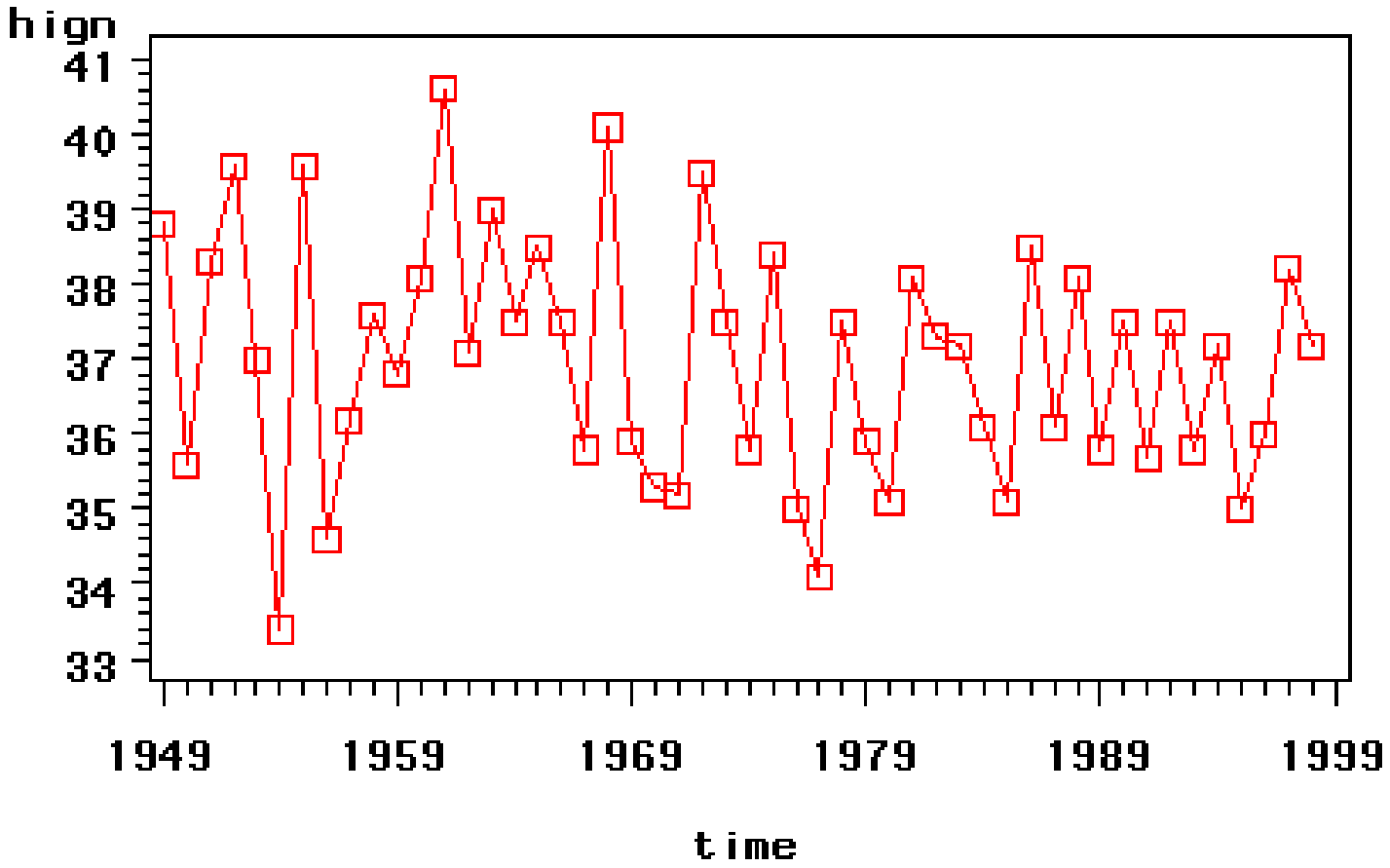


Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	10383.588	1.00000	|**********|
1	9257.734	0.89157	|********** . |
2	8080.289	0.77818	|********** . |
3	6440.643	0.62027	|********** . |
4	5053.314	0.48666	|********** . |
5	4445.713	0.42815	|********** . |
6	3904.890	0.37606	|********** . |
7	4306.827	0.41477	|********** . |
8	4716.761	0.45425	|********** . |
9	5833.655	0.56181	|********** . |
10	7128.946	0.68656	|********** . |
11	7980.333	0.76855	|********** . |
12	8773.234	0.84491	|********** . |
13	7735.639	0.74499	|********** . |
14	6621.269	0.63767	|********** . |
15	5084.621	0.48968	|********** . |
16	3775.004	0.36355	|******* . |
17	3176.849	0.30595	|****** . |
18	2646.859	0.25491	|***** . |
19	2984.458	0.28742	|****** . |
20	3328.659	0.32057	|****** . |
21	4324.928	0.41652	|******** . |
22	5489.933	0.52871	|********** . |
23	6265.032	0.60336	|********** . |
24	6986.088	0.67280	|********** . |

"." marks two standard errors

平稳序列时序图



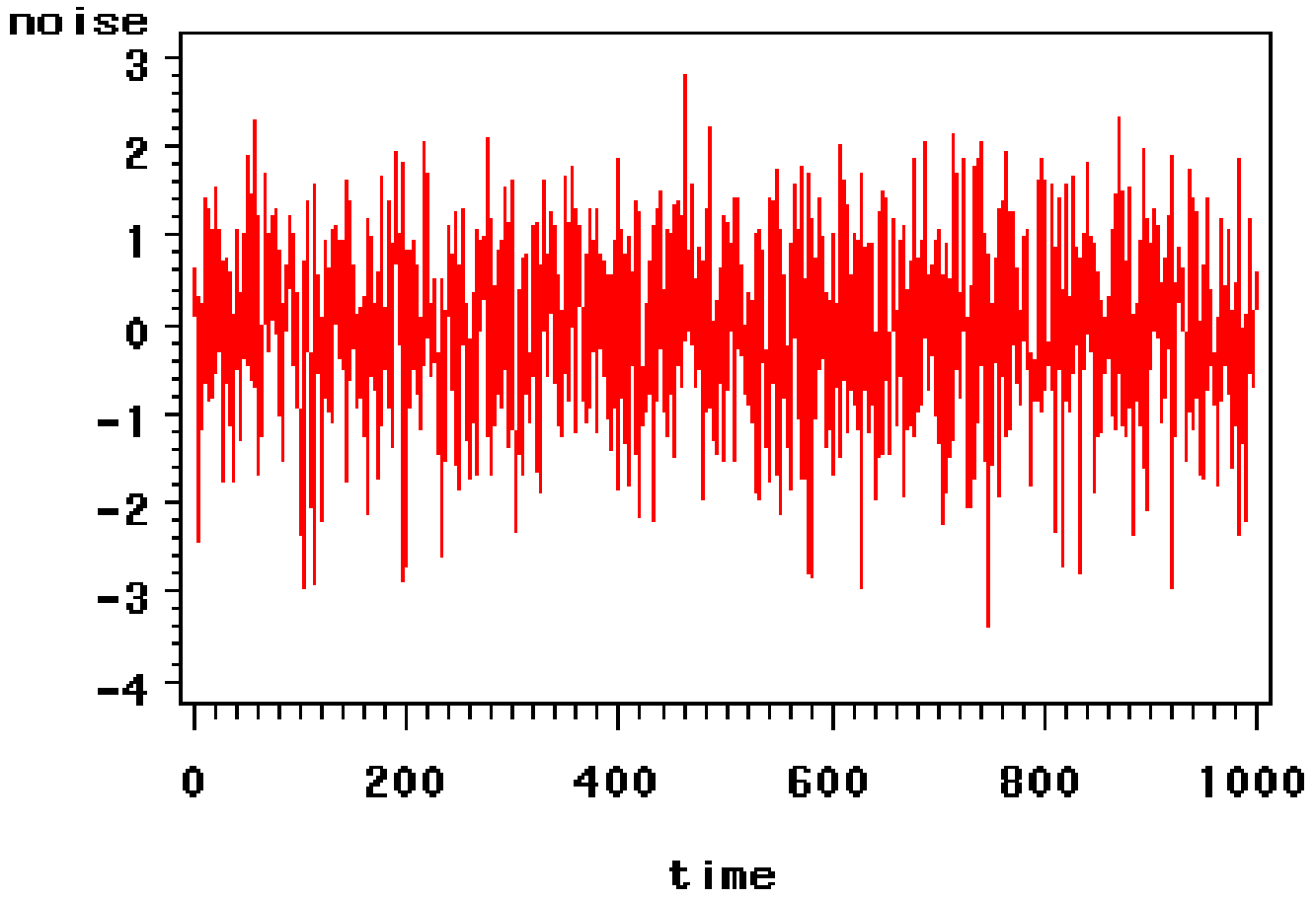
自相关图

Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	2.569604	1.00000	|********************|
1	-0.449960	-.17511	|. ****| . |
2	-0.0091078	-.00354	|. | . |
3	0.463204	0.18026	|. |**** |
4	0.059232	0.02305	|. | . |
5	-0.421428	-.16400	|. *** | . |
6	0.253512	0.09866	|. |** |
7	-0.067559	-.02629	|. * | . |
8	-0.0083274	-.00324	|. | . |
9	-0.057247	-.02228	|. | . |
10	0.148917	0.05795	|. |* |
11	0.095461	0.03715	|. |* |
12	-0.267799	-.10422	|. ** | . |
13	0.260969	0.10156	|. |** |
14	0.011069	0.00431	|. | . |
15	-0.069243	-.02695	|. * | . |

"." marks two standard errors

标准正态白噪声序列时序图



标准正态白噪声序列纯随机性检验

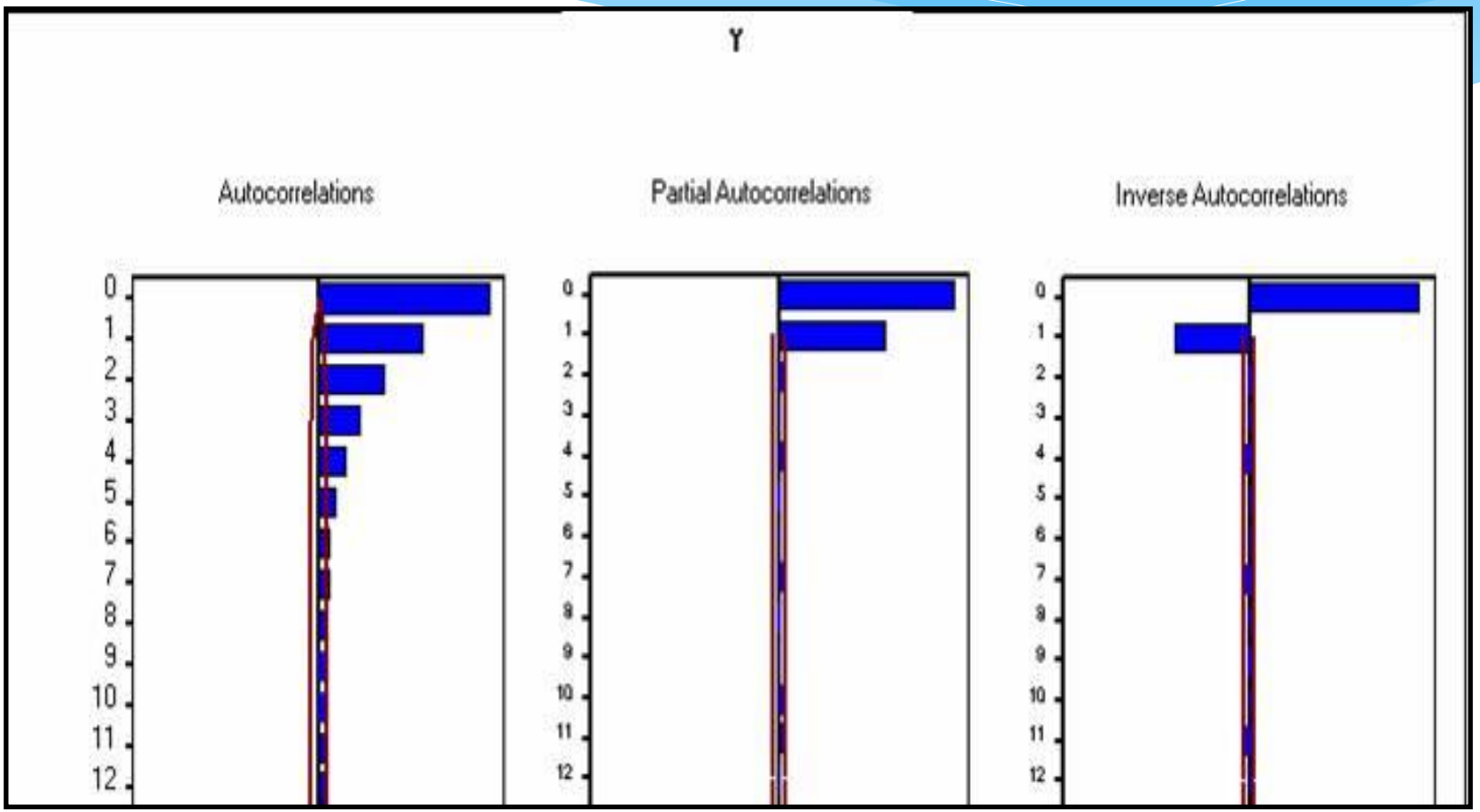
样本自相关图

Autocorrelations

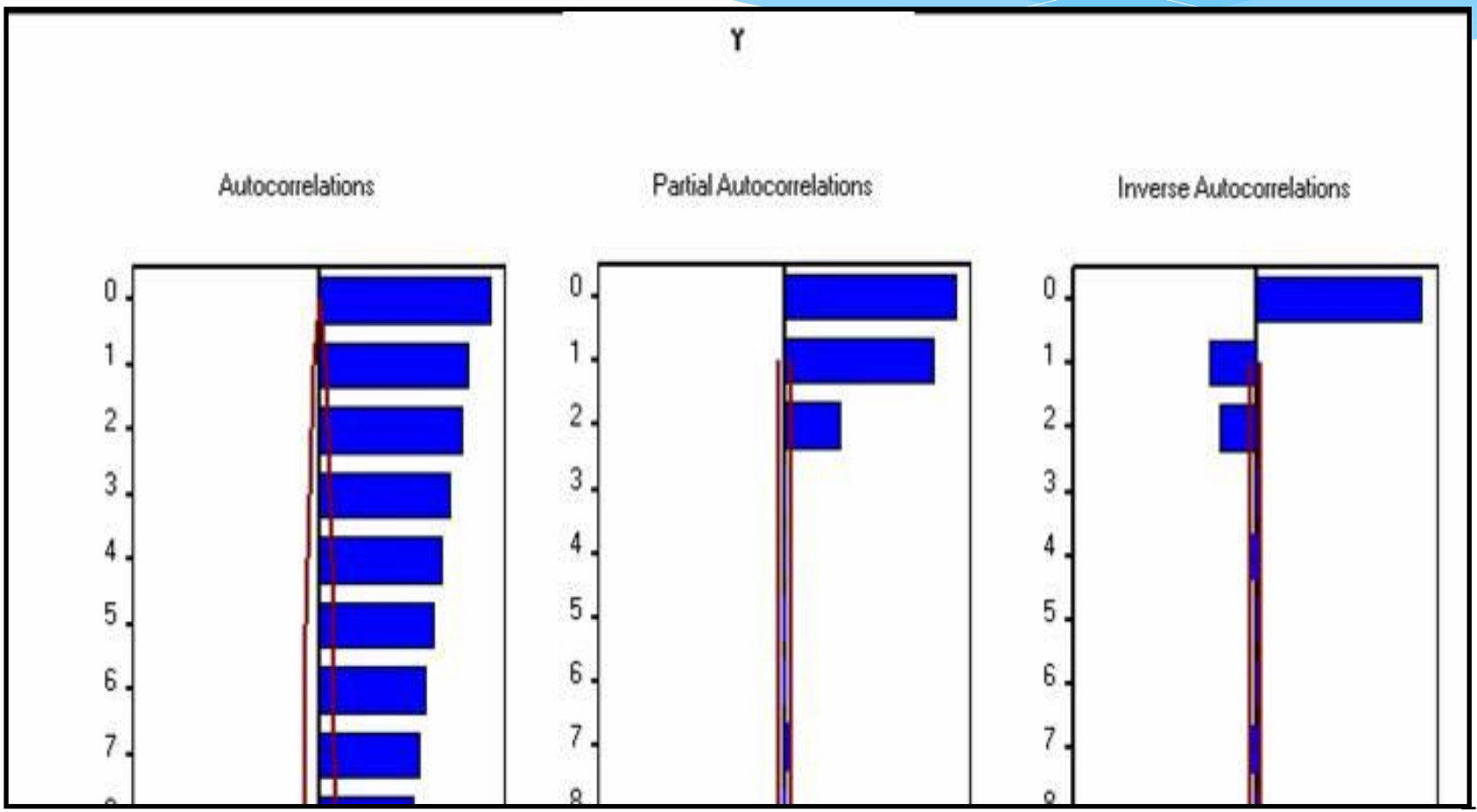
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	1.005528	1.00000	|**********|
1	-0.0010647	-.00106	.| .
2	-0.036747	-.03655	*| .
3	-0.0062567	-.00622	.| .
4	0.011938	0.01187	.| .
5	-0.025139	-.02500	*| .
6	-0.014428	-.01435	.| .
7	0.0088565	0.00881	.| .
8	-0.010179	-.01012	.| .
9	-0.026893	-.02674	*| .
10	-0.024882	-.02475	.| .
11	-0.014021	-.01394	.| .
12	0.035527	0.03533	.|*

"." marks two standard errors

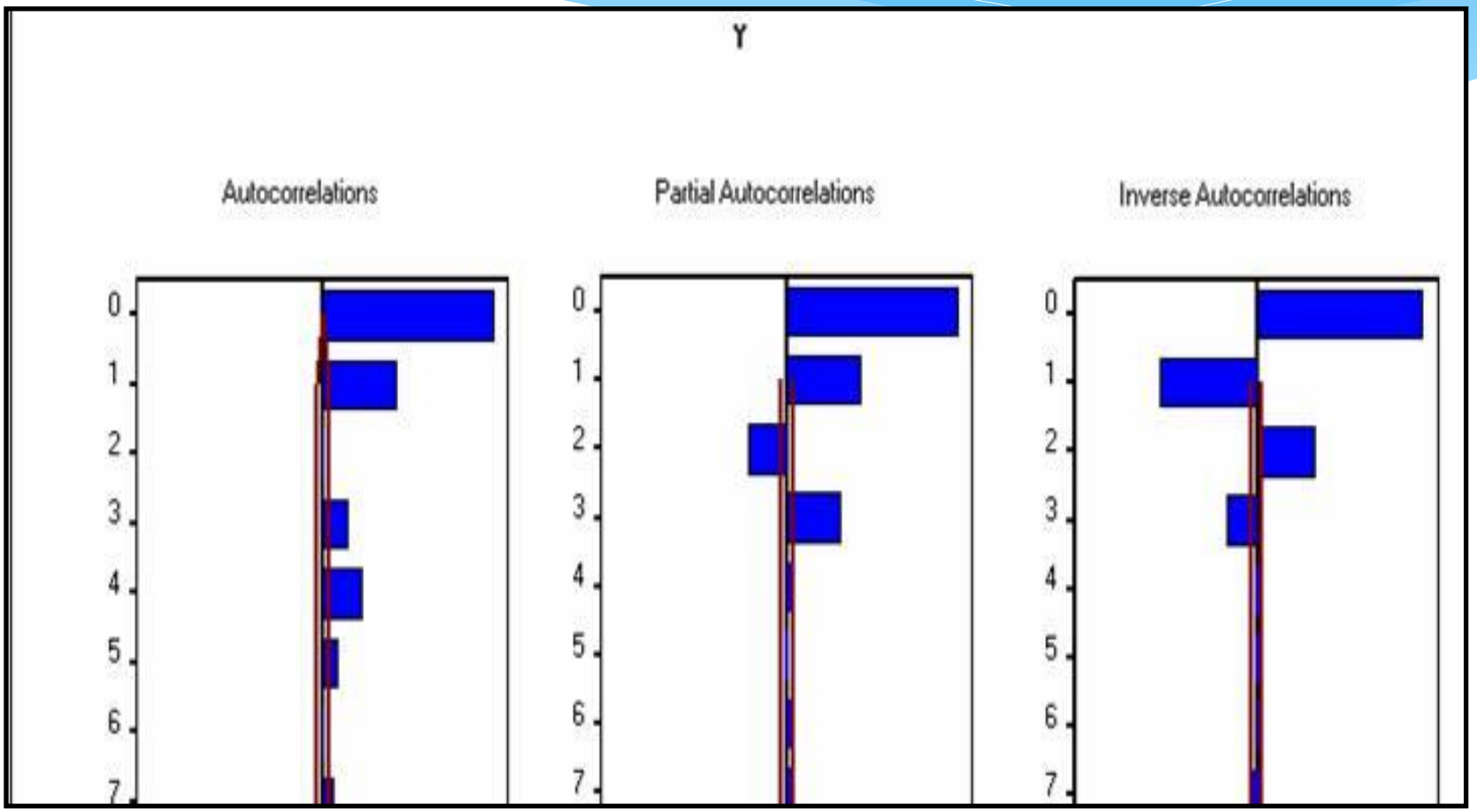
Autocorrelation Plots for an AR(1) Time Series



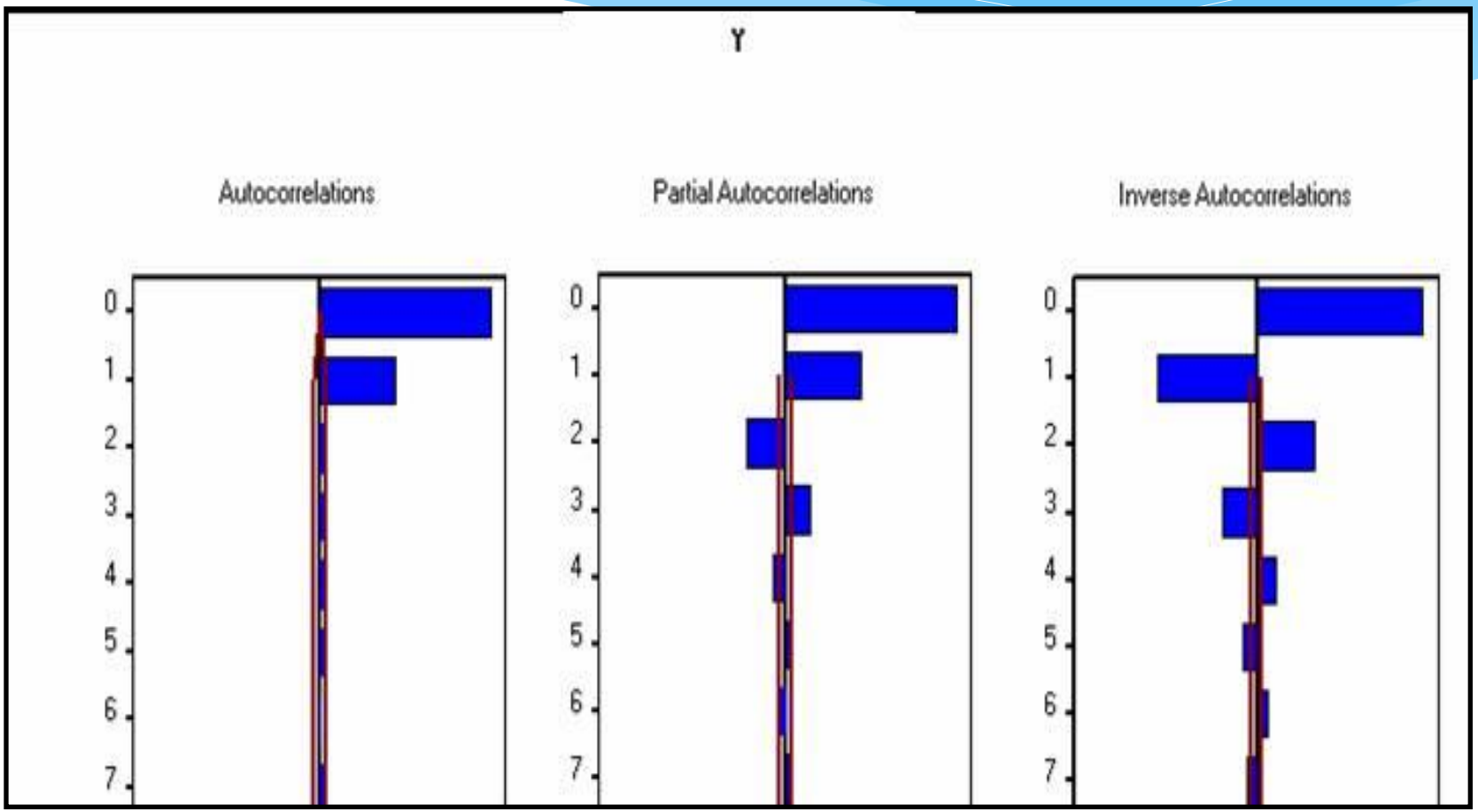
Autocorrelation Plots for an AR(2) Time Series



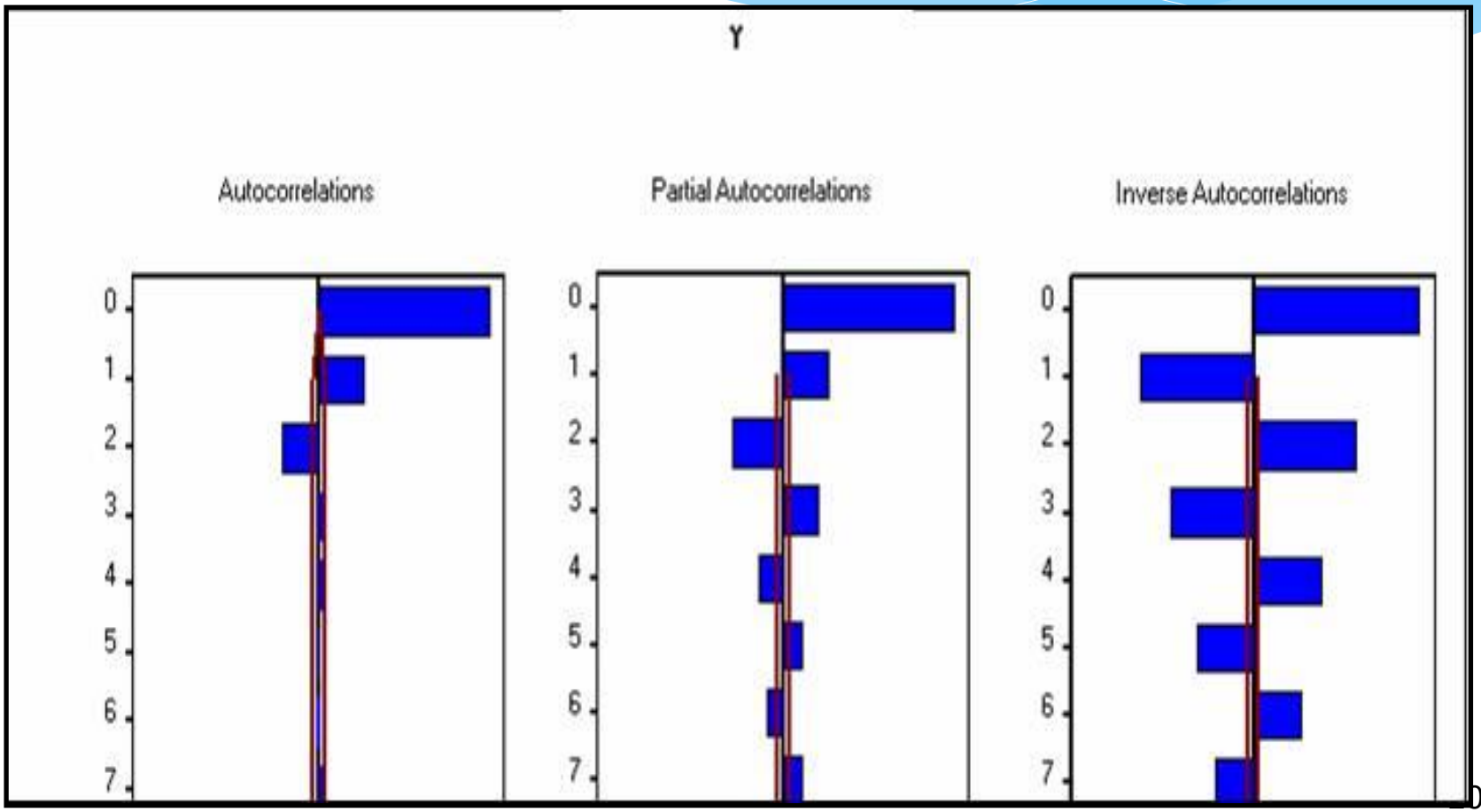
Autocorrelation Plots for an AR(3)



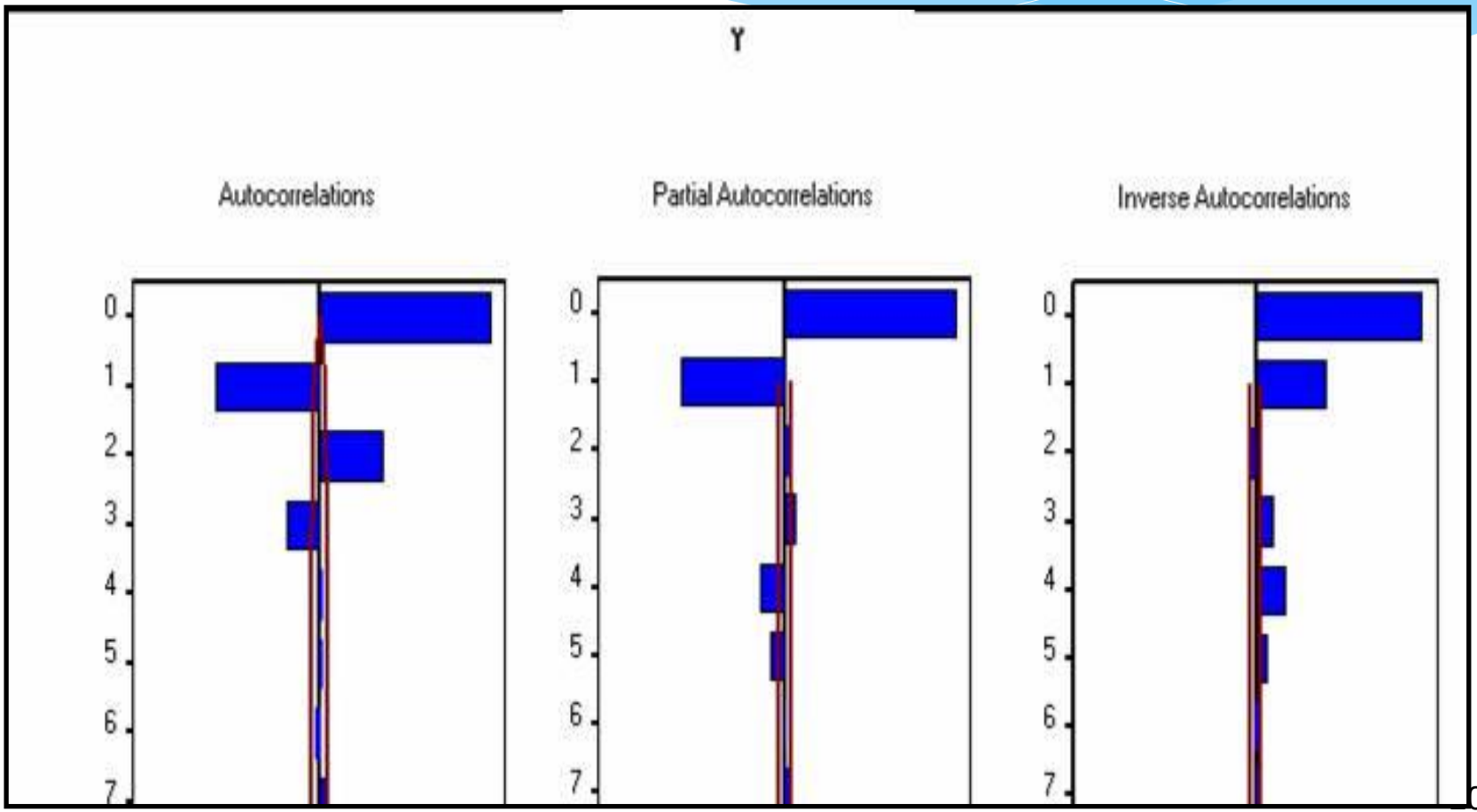
Autocorrelation Plots for an MA(1)



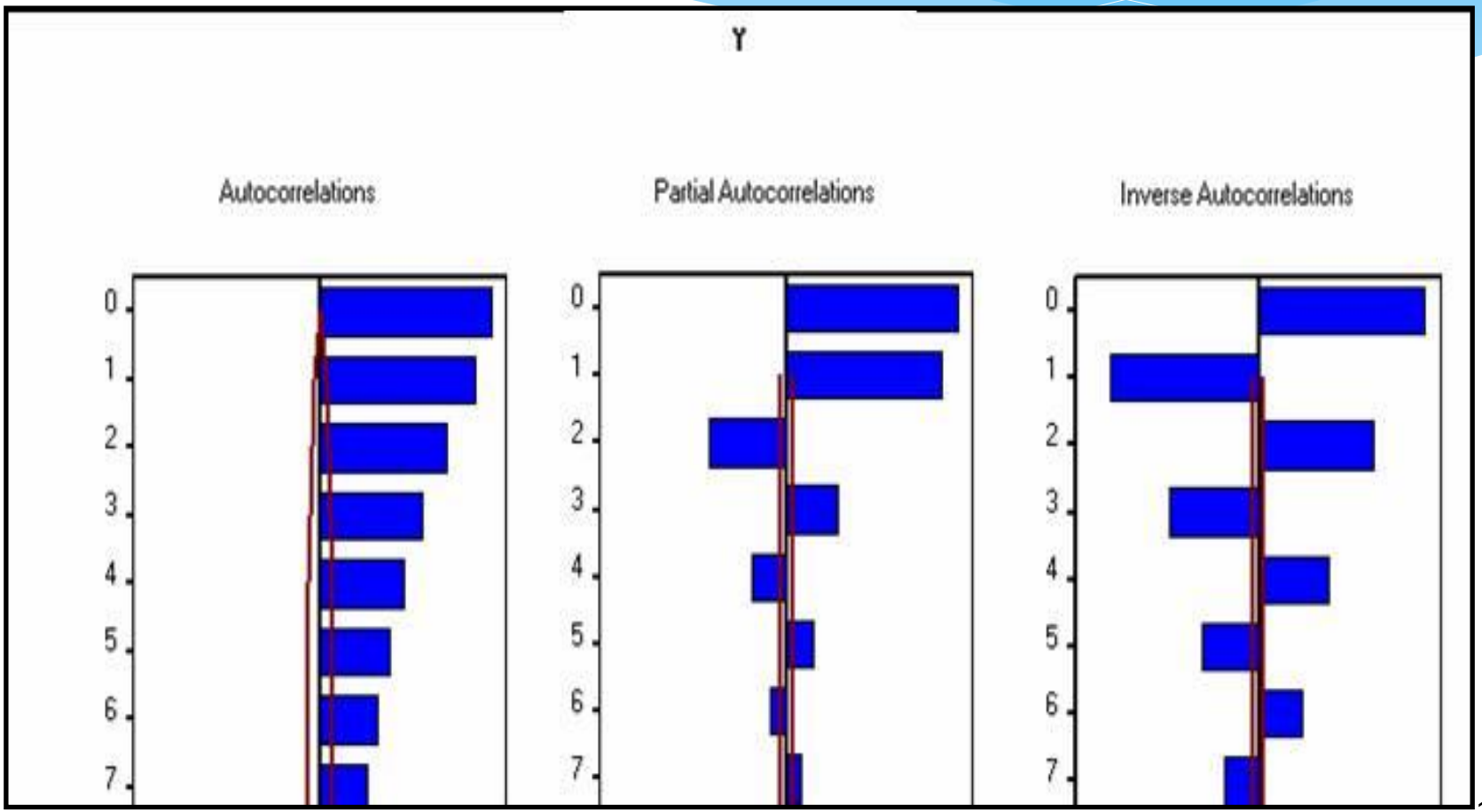
Autocorrelation Plots for an MA(2)



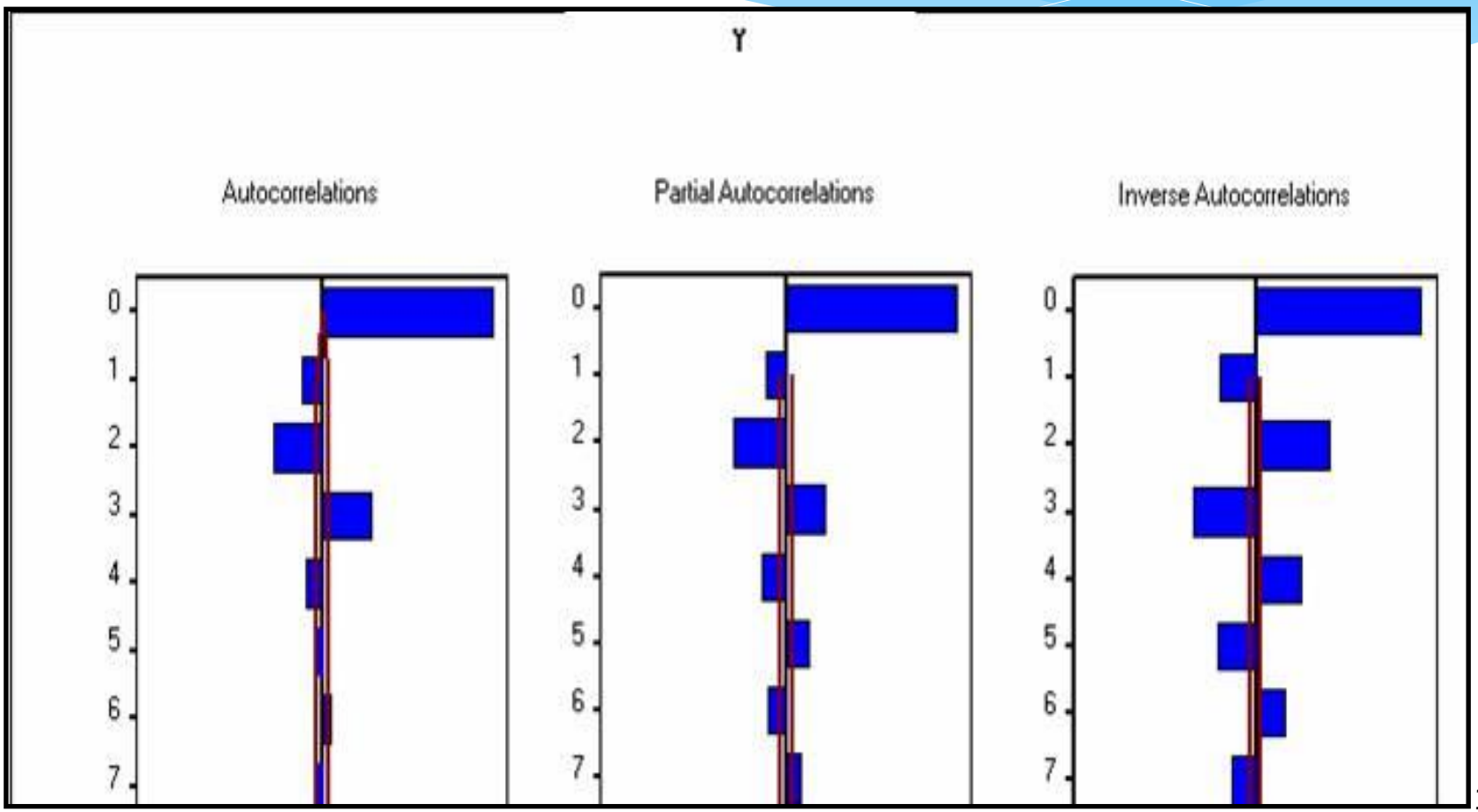
Autocorrelation Plots for an MA(3)



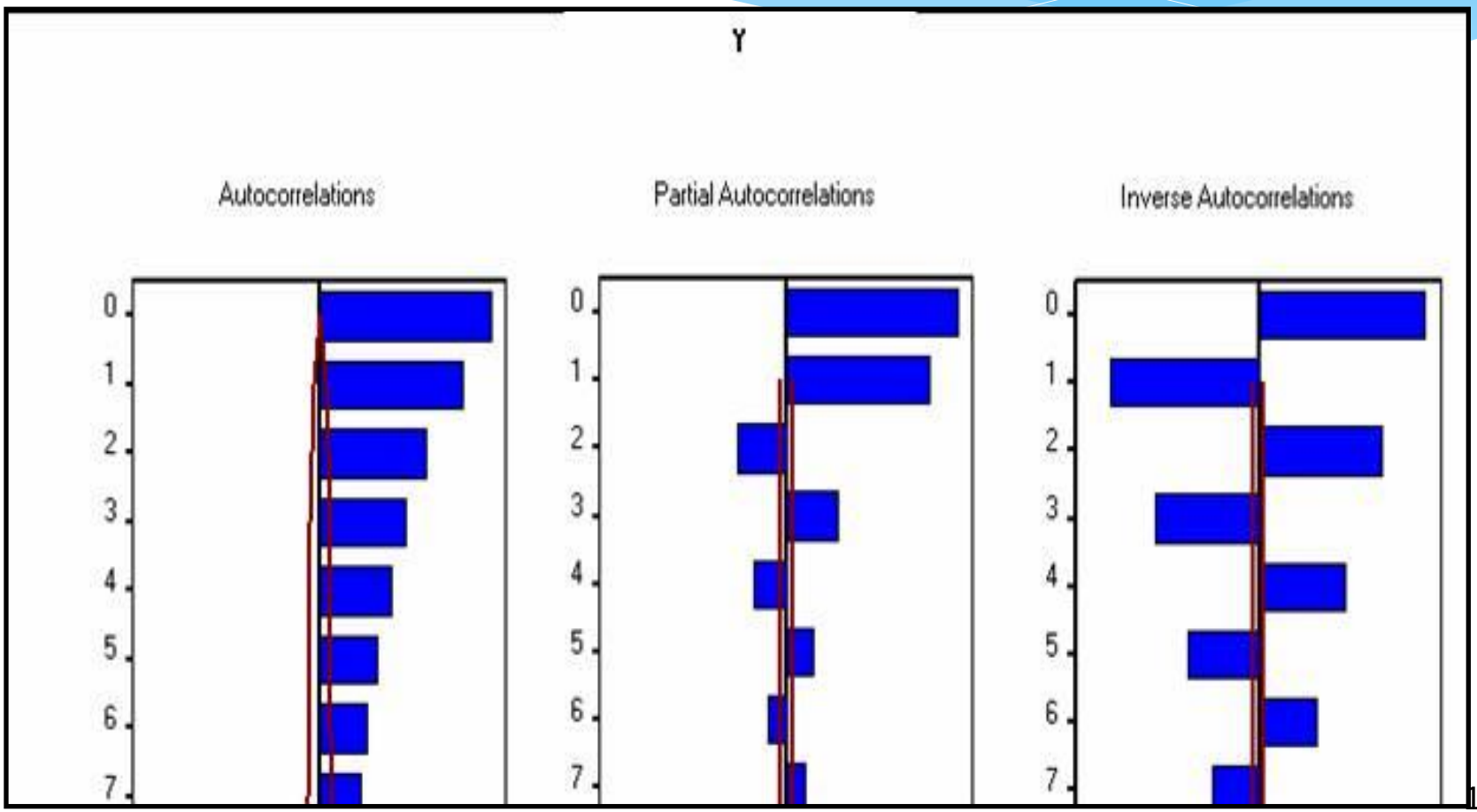
Autocorrelation Plots for an ARMA(1,1)



Autocorrelation Plots for an ARMA(1,2)



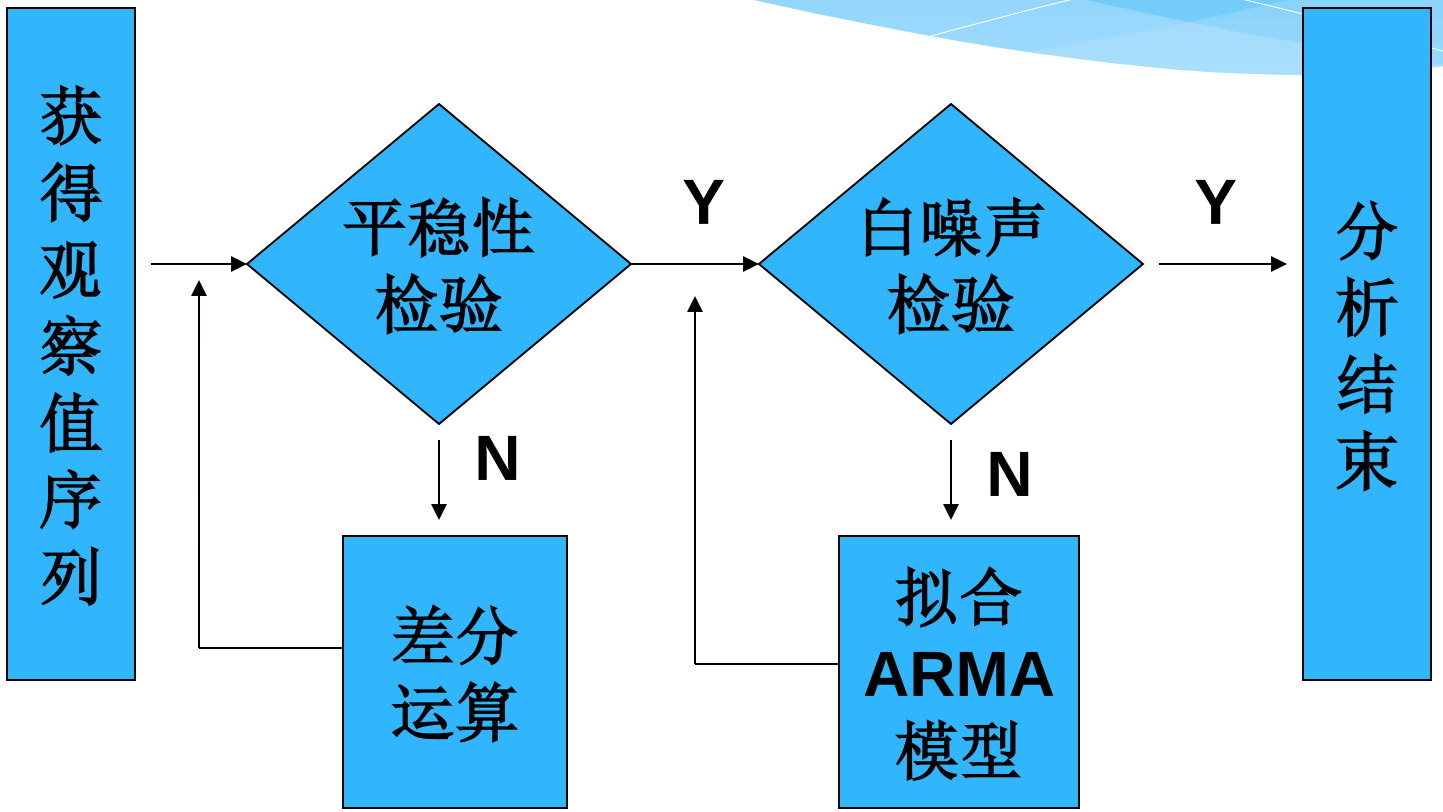
Autocorrelation Plots for an ARMA(2,1)



非平稳序列差分方式的选择

* 序列蕴含着显著的线性趋势，一阶差分就可以实现趋势平稳
* 序列蕴含着曲线趋势，通常低阶（二阶或三阶）差分就可以提取出曲线趋势的影响
* 对于蕴含着固定周期的序列进行步长为周期长度的差分运算，通常可以较好地提取周期信息

ARIMA模型建模步骤



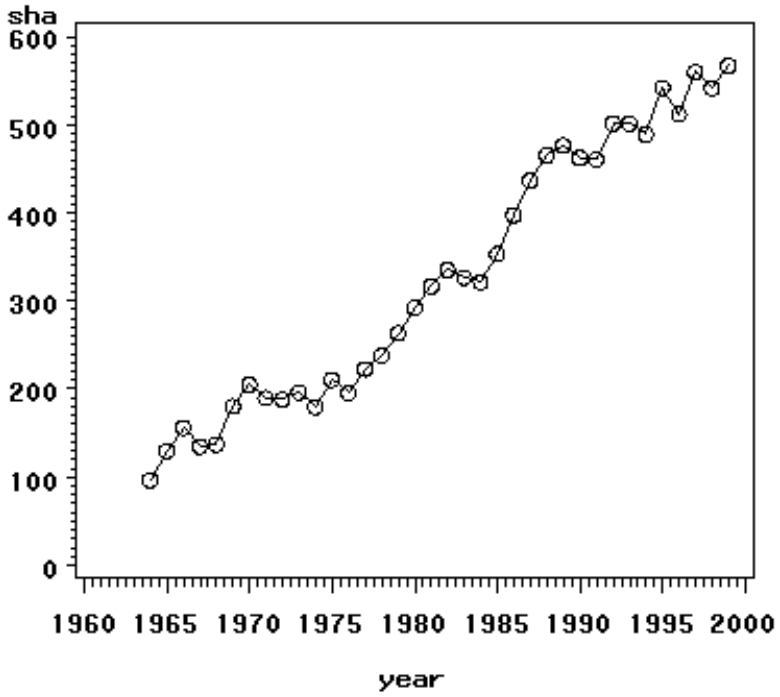
1964年——1999年中国纱年产量序列蕴含着一个近似线性的递增趋势。对该序列进行一阶差分运算

$$\nabla x_t = x_t - x_{t-1}$$

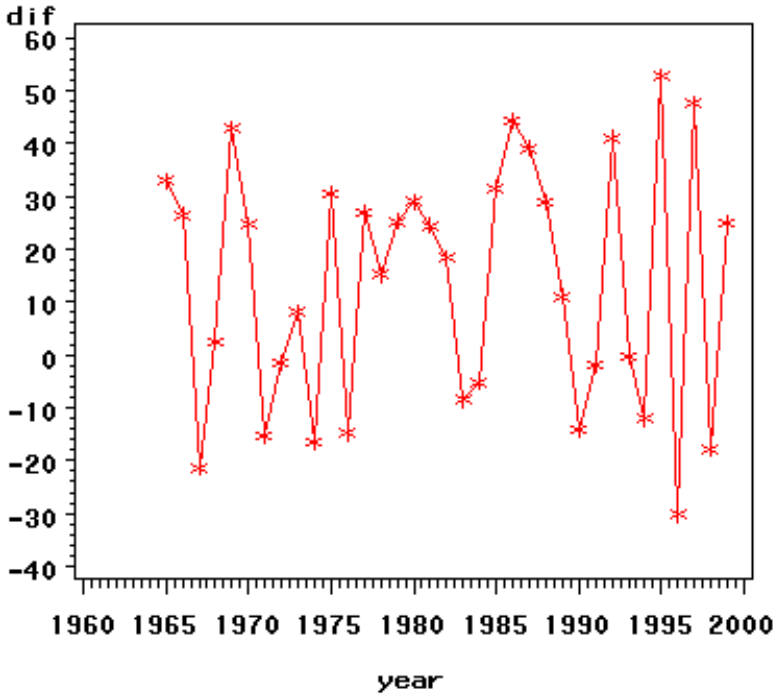
考察差分运算对该序列线性趋势信息的提取作用

差分前后时序图

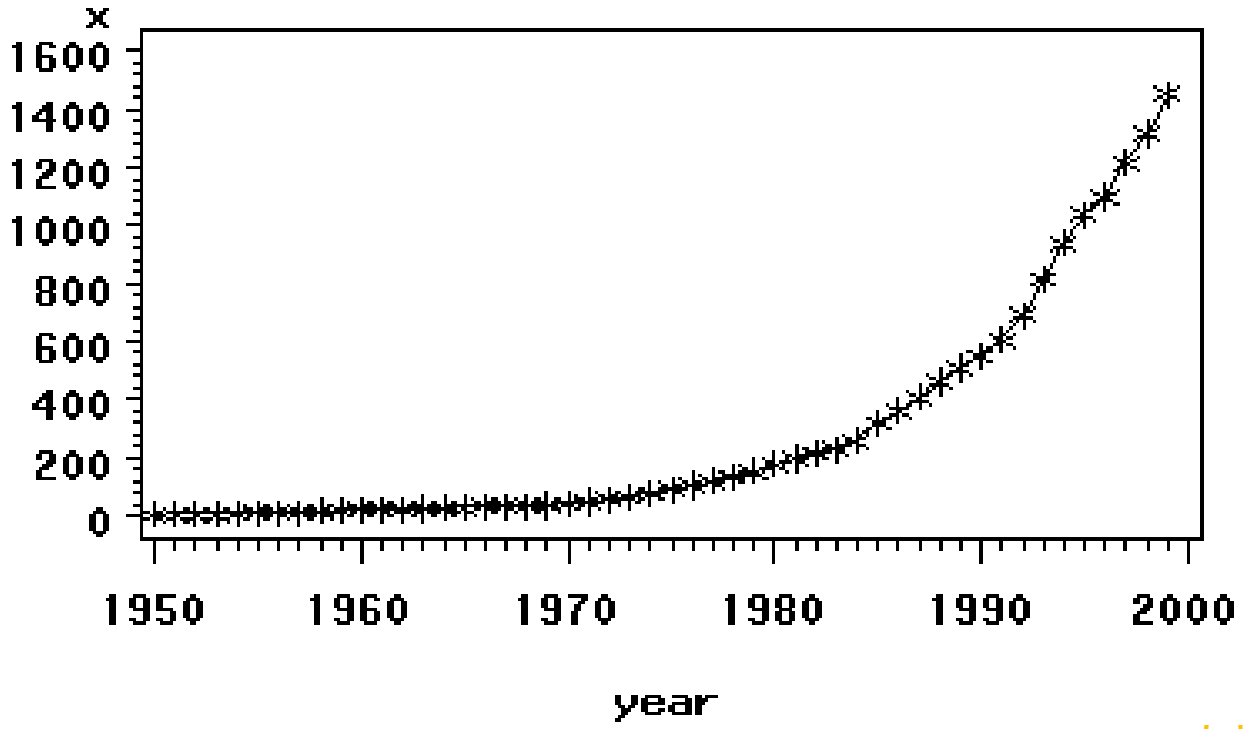
原序列时序图



差分后序列时序图

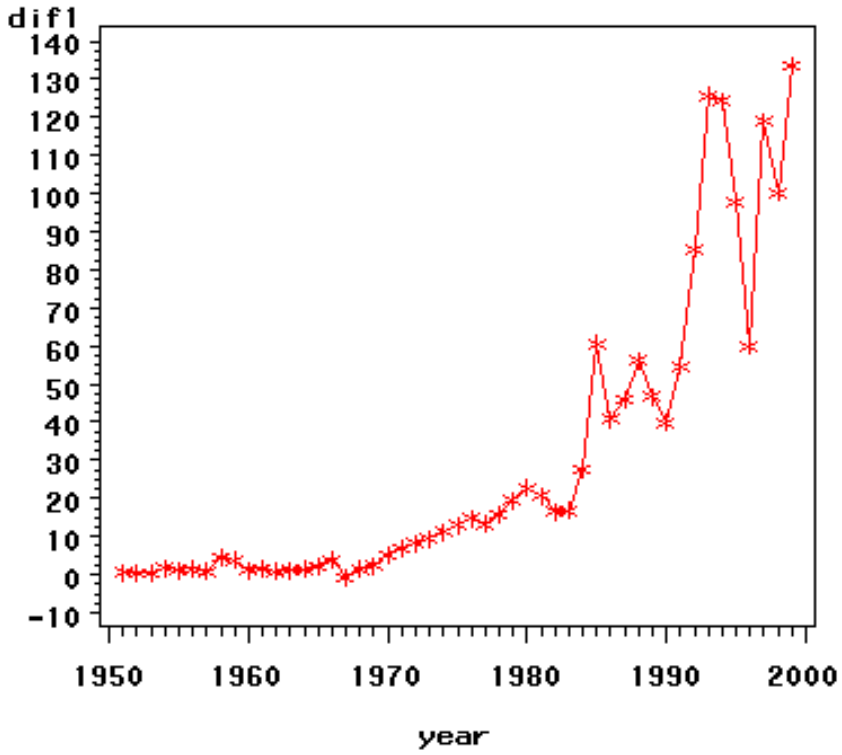


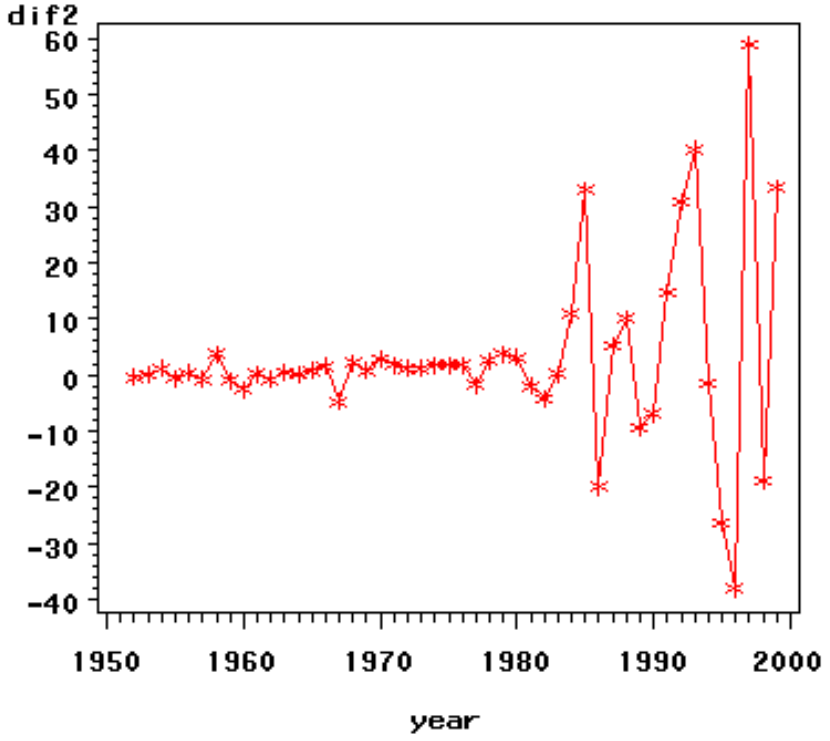
* 尝试提取1950年——1999年北京市民用车辆拥有量序



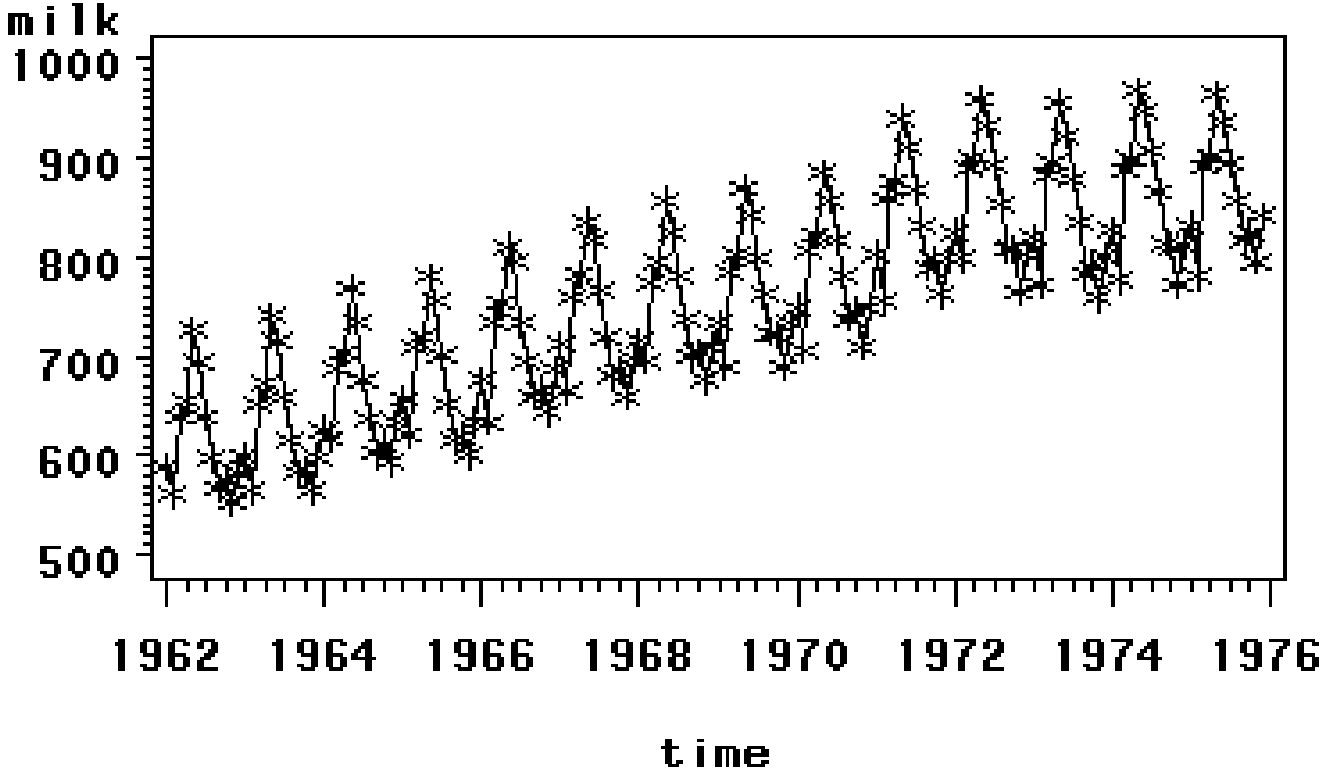
差分后序列时序图

- 一阶差分
- 二阶差分



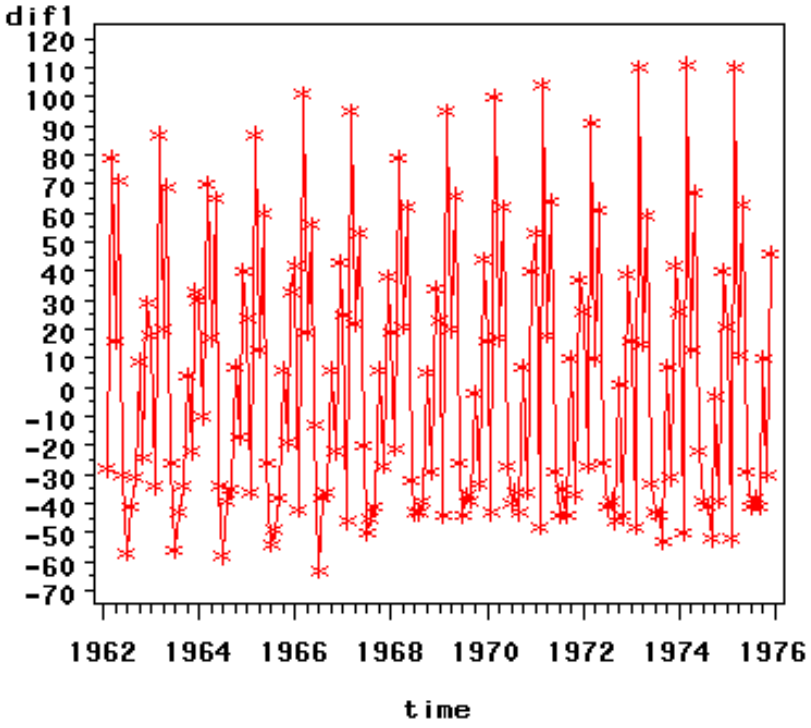


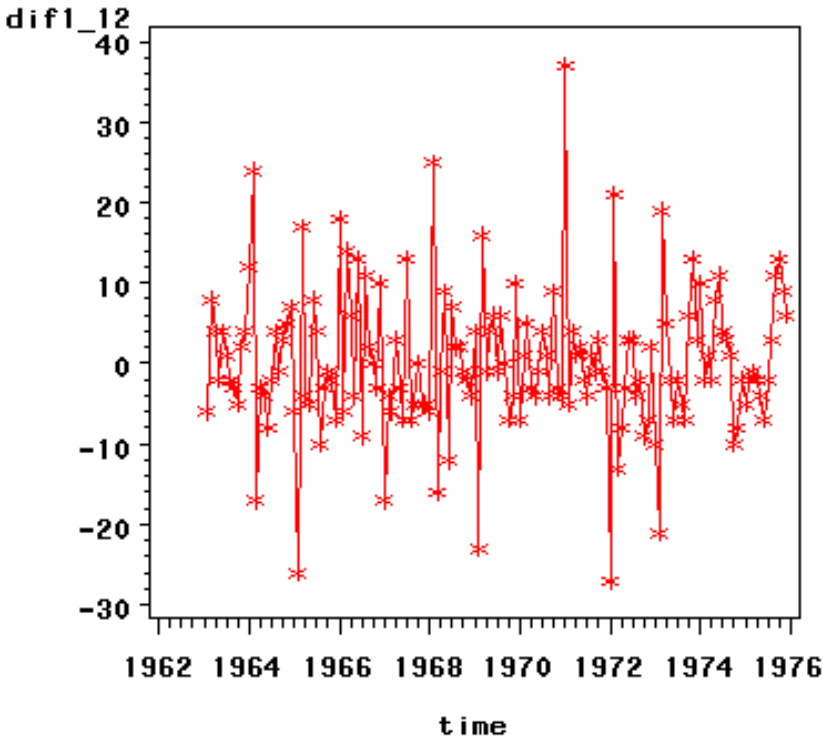
* 差分运算提取1962年1月——1975年12月平均每头奶牛的月产奶量序列中的确定性信息



差分后序列时序图

- 一阶差分
- 1阶一12步差分





模型参数估计

* 一、模型参数的矩方法估计
* 二、最小二乘估计
* 三、极大似然估计

模型检验

* 模型的显著性检验
  * 整个模型对信息的提取是否充分
* 参数的显著性检验
  * 模型结构是否最简

模型优化

* 问题提出
  * 当一个拟合模型通过了检验，说明在一定的置信水平下，该模型能有效地拟合观察值序列的波动，但这种有效模型并不是唯一的。
* 优化的目的
  * 选择相对最优模型

模型选择统计量

* R-Square:

$$R^2 = 1 - \sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2 \Big/ \sum_{t=1}^{n}(Y_t - \bar{Y}_t)^2$$

* Root Mean Squared Error:

$$\mathrm{MSE} = \frac{1}{n-k}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$$

模型选择统计量

* Mean Absolute Percent Error:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} | Y_t - \hat{Y}_t | / Y_t$$

* Mean Absolute Error:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} | Y_t - \hat{Y}_t |$$

continued...

模型选择统计量

* 准则函数
* 一般公式:

* IC = 拟合程度 + 模型复杂性惩罚

* Akaike's A Information Criteria（AIC）:

$$\mathrm{AIC} = -2\log(L) + 2k$$

* Schwarz's Bayesian Information Criteria（SBC）:

$$\mathrm{SBC} = -2\log(L) + k\log(n)$$

模型选择统计量

* Akaike's A Information Criteria:

$$\mathrm{AIC} = n\log(\mathrm{SSE}/n) + 2k$$

* Schwarz's Bayesian Information Criteria:

$$\mathrm{SBC} = n\log(\mathrm{SSE}/n) + k\log(n)$$

时间序列预测

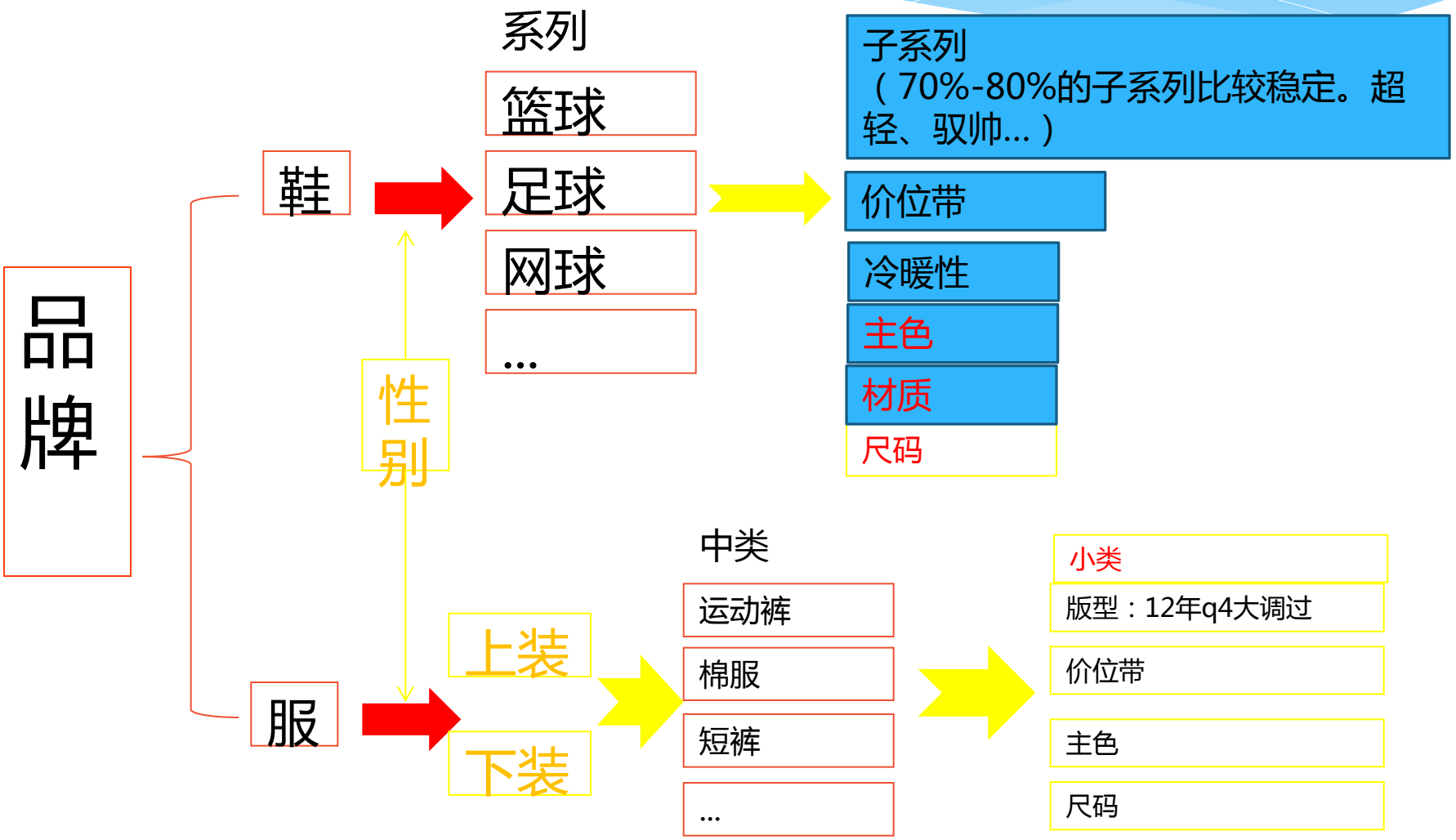
设当前时刻为t,且零均值平稳序列$\{x_t\}$在时刻t及以前的观察值为$x_t, x_{t-1}, \cdots\cdots$,用序列$\{x_t\}$对时刻$t$以后的观察值$x_{t+l}(l>0)$进行预测,这种预测称为以$t$为原点的,向前期步长为$l$的预测,预测值记为$\hat{x}_t(l)$.

选择变量

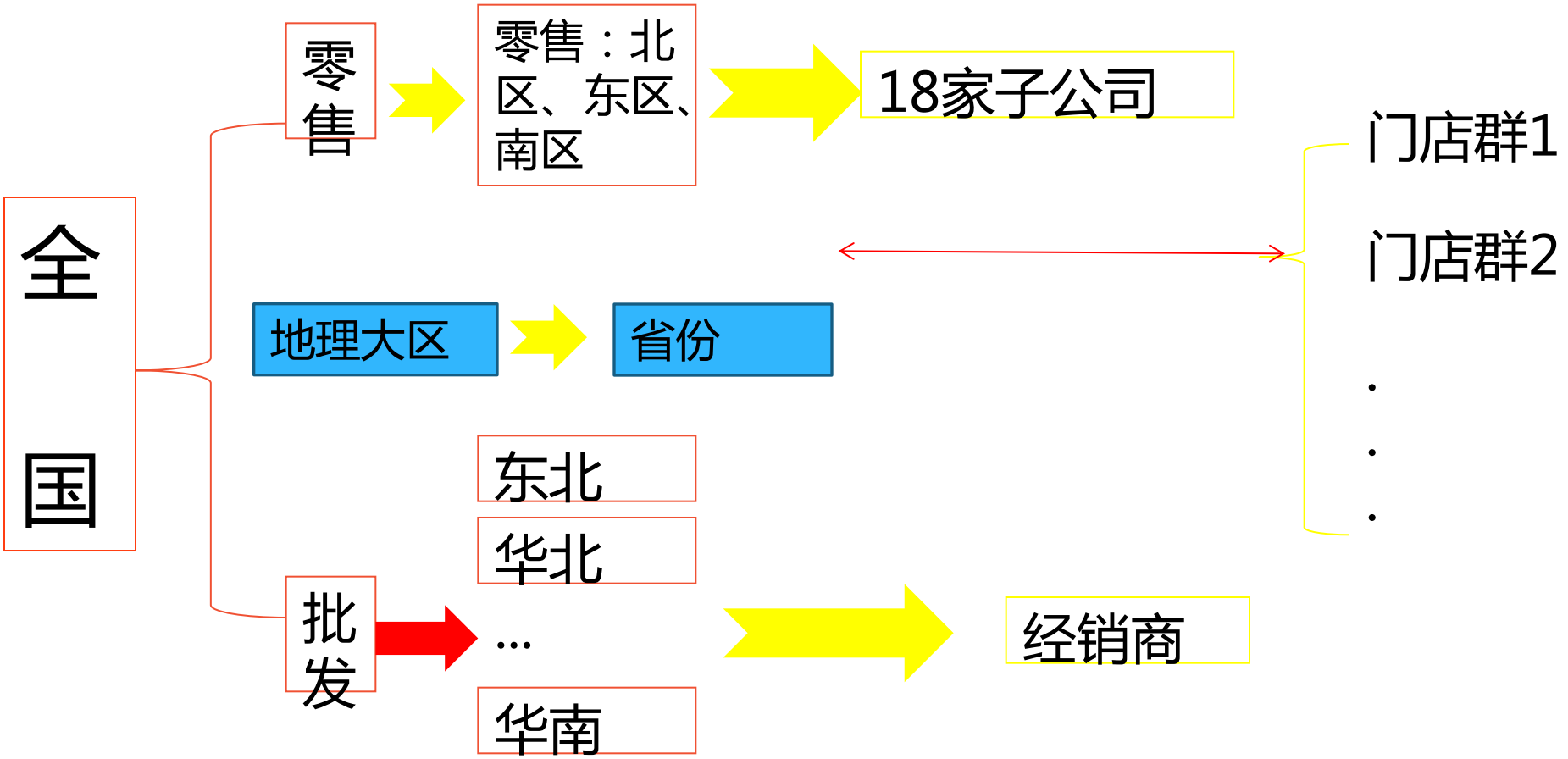
由销售记录形成整体的数据字典；

使用产品代码记录和门店代码记录，进行数据数据清洗，整理出可以进行分析和预测的产品和渠道维度。

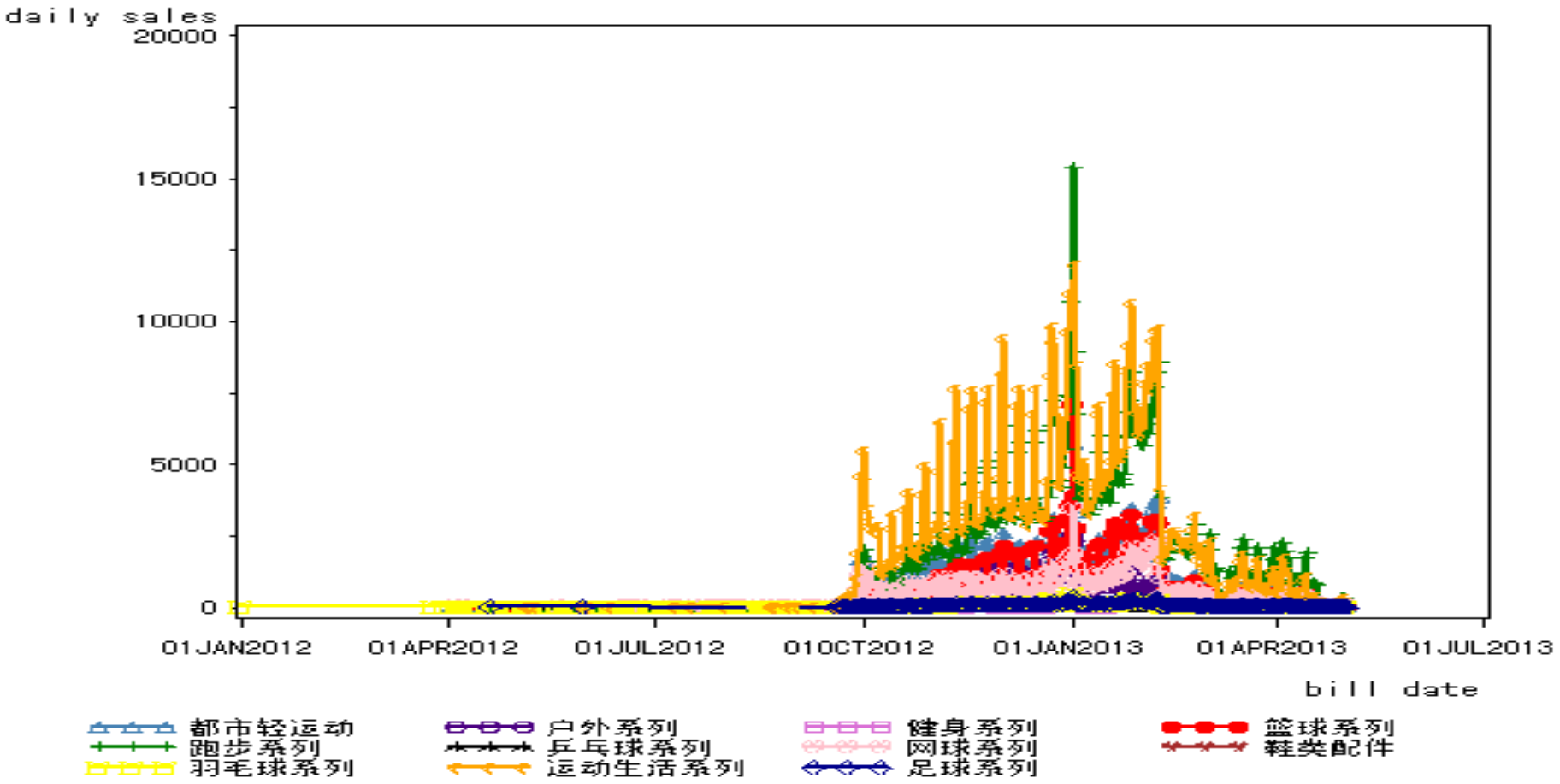
业务分析内容（产品）


品牌
鞋
服
性别
系列
篮球
足球
网球
...
子系列
（70%-80%的子系列比较稳定。超轻、驭帅...）
价位带
冷暖性
主色
材质
尺码
上装
下装
中类
运动裤
棉服
短裤
...
小类
版型：12年q4大调过
价位带
主色
尺码

业务分析内容（渠道）

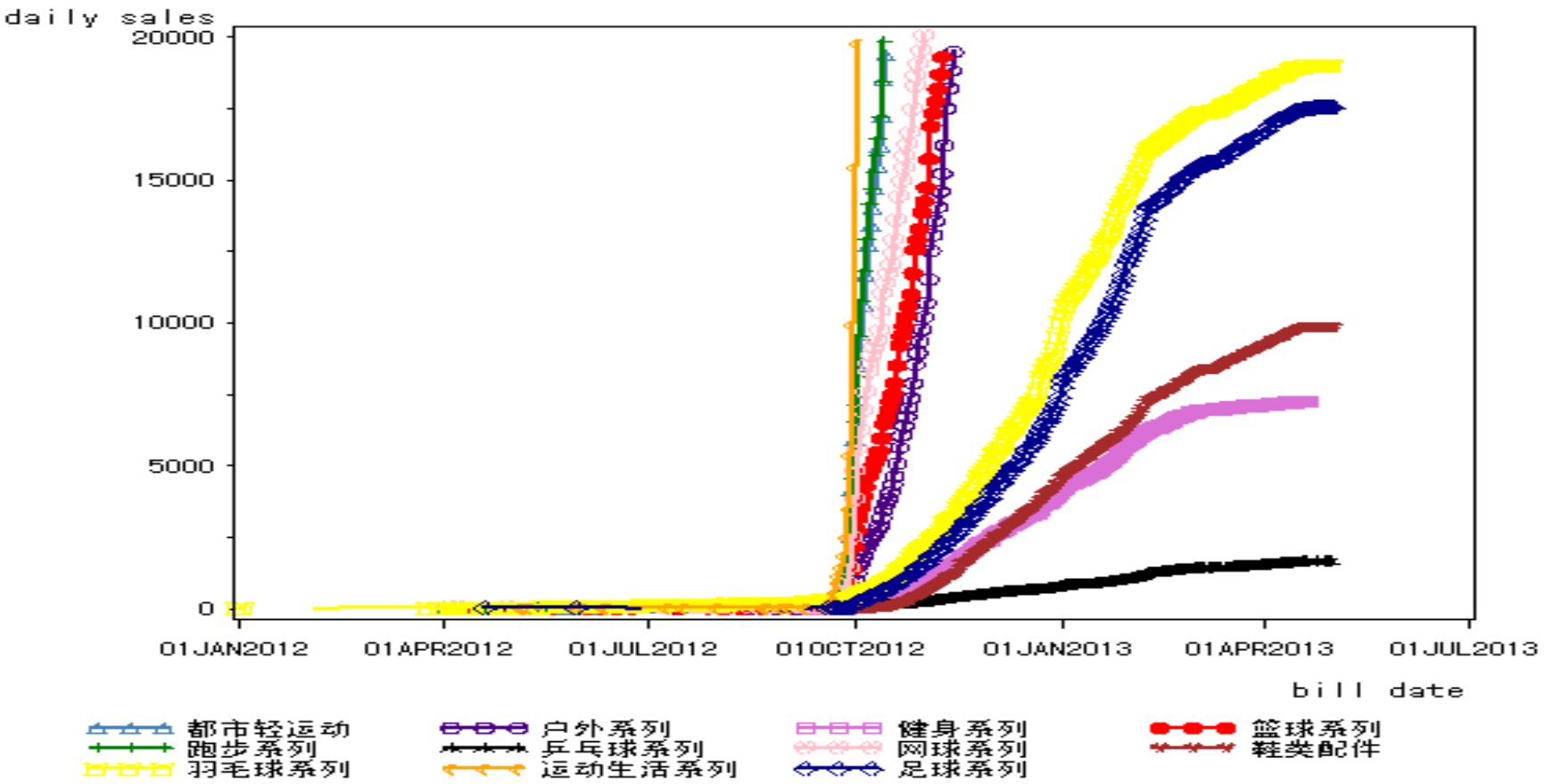


Shoes Daily Sale



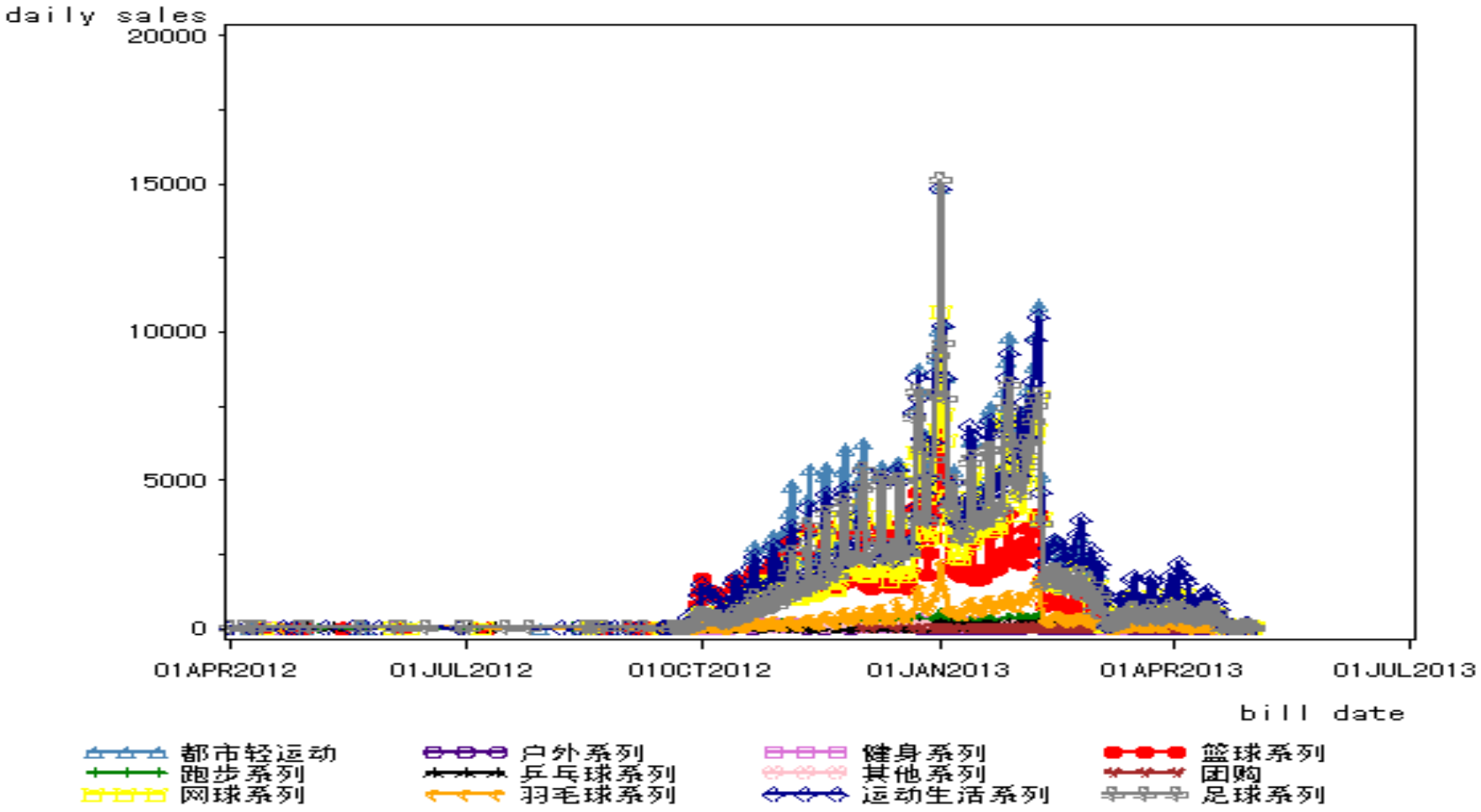
Source: 2012-2013 Trade
trade details

Shoes Cumulative Sales



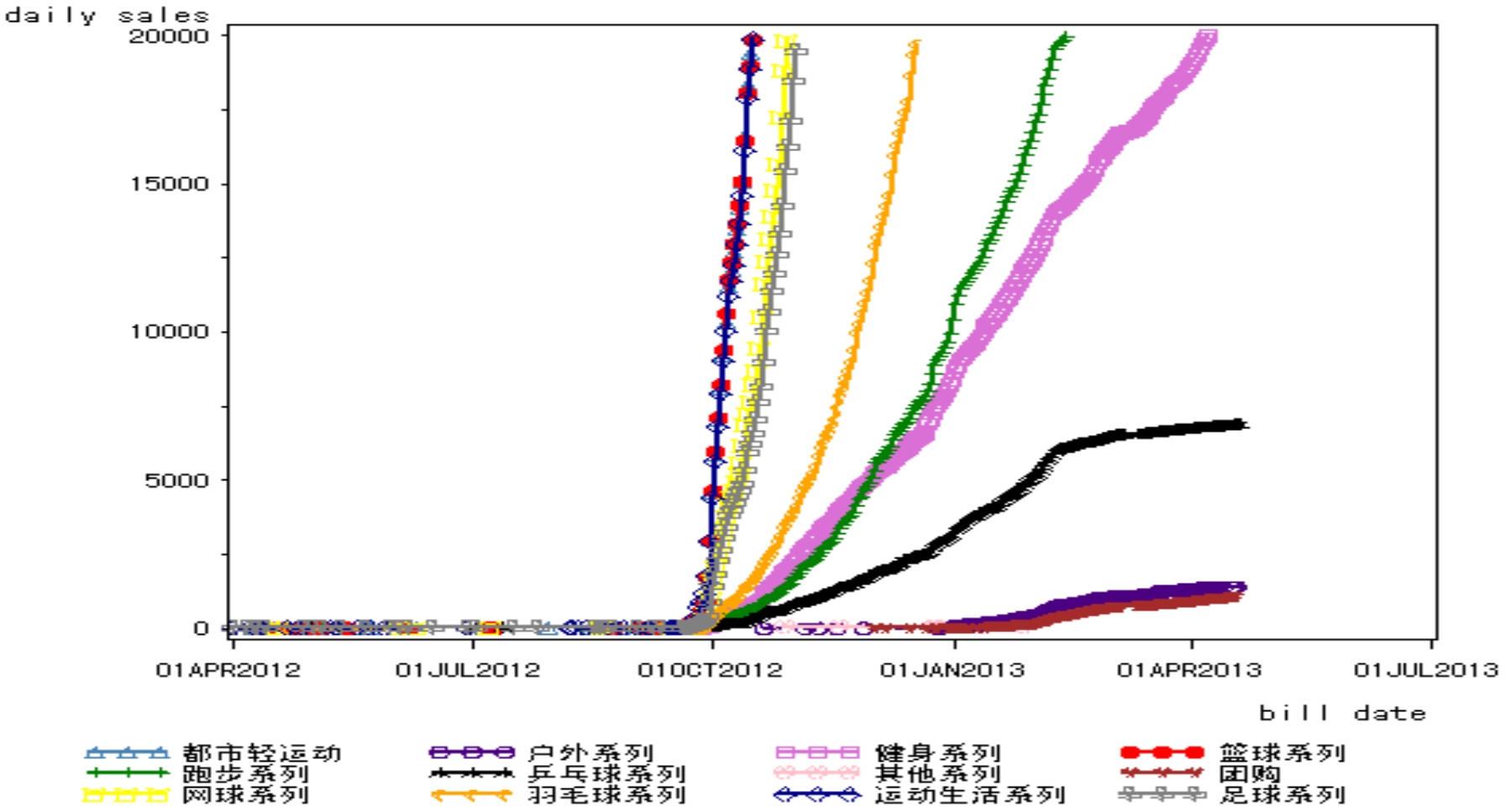
Source: 2012-2013 Trade trade details

Clothe Daily Sale



Source: 2012-2013 Trade
trade details

Clothe Cumulative Sales



Source: 2012-2013 Trade trade details

附录：指数平滑模型

* 一次指数平滑预测法
* 二次指数平滑预测法
* 三次指数平滑预测法

一次指数平滑预测法

* 原理
  * 一次指数平滑预测法是把一次指数平滑值直接作为预测值的方法。而一次指数平滑又称为简单指数平滑。
* 公式：是根据移动平均公式推导而来的。

$$F_{t+1} = F_t + \frac{1}{n}(y_t - y_{t-n})$$

式中 n 表示移动的步长，t 表示任意时刻。我们用 F_t 代替 y_{t-n}，则有

$$F_{t+1} = F_t + \frac{1}{n}(y_t - y_t)$$

$$F_{t+1} = \frac{1}{n}y_t + \left(1 - \frac{1}{n}\right)F_t$$

令 $\alpha = \frac{1}{n}$，显然 $0 < \alpha < 1$，有

$$F_{t+1} = \alpha y_t + (1 - \alpha)F_t$$

上式就是一次指数平滑公式，式中 α 称为平滑常数或平滑系数。 F_t 表示 t 时刻上的平滑值。 F_{t+1} 即表示 $t+1$ 时刻上的平滑值，又是 $t+1$ 时的预测值。

根据习惯一次指数平滑公式还可以写成：

$$S_{t+1} = \alpha y_t + (1-\alpha) S_t$$

上式一般是用于把平滑值直接用于预测的情形。但是当平滑值不是直接用于预测，而主要是对数列进行平滑时，公式还可改写为：

$$S_t = \alpha y_t + (1-\alpha) S_{t-1}$$

一次指数平滑的特点。它实际是对过去的历史数据进行呈指数递减的加权平均。因为，

$$S_t = \alpha y_{t-1} + (1-\alpha) S_{t-1}$$

$$S_{t-1} = \alpha y_{t-1} + (1-\alpha) S_{t-2}$$

……

所以有

$$S_t = \alpha y_t + (1-\alpha) S_{t-1}$$

$$= \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots + \alpha(1-\alpha)^n y_{t-n+1} + (1-\alpha)^n S_{t-n+1}$$

显然，上式的系数呈指数衰减，这也是指数平滑名称的由来。

例　对某产品的月度销售量进行预测，α值分别取 0.2，0.5，0.7，初始值取为数列的第一个值。结果如下表：

表　　某产品的月度销售量一次指数平滑预测表

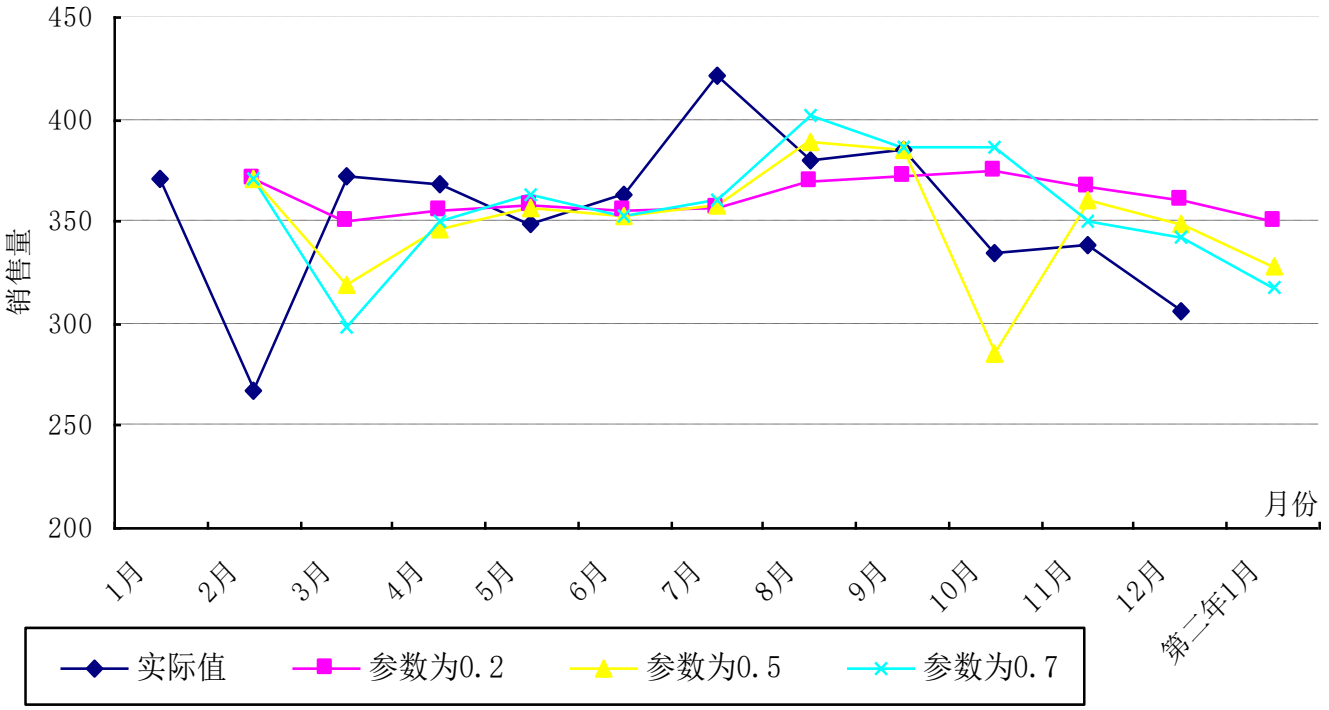
月份	某产品的月度销售量（见万箱）y_t	一次指数平滑预测		
		$\alpha=0.2$	$\alpha=0.5$	$\alpha=0.7$
1 月	371.5			
2 月	267.4	371.5	371.5	371.5
3 月	372.4	350.68	319.45	298.63
4 月	368.2	355.02	345.93	350.27
5 月	349.4	357.66	357.06	362.82
6 月	362.8	356.01	353.23	353.43
7 月	420.9	357.37	358.02	359.99
8 月	380.4	370.07	389.46	402.63
9 月	385.6	372.14	384.93	387.07
10 月	335.0	374.83	285.27	386.04
11 月	338.5	366.86	360.13	350.31
12 月	306.6	361.19	349.32	342.04
第二年 1 月		350.27	327.96	317.23

其中计算过程如下如α=0.2时

$$S_1 = 0.2 \times 371.5 + 0.8 \times 371.5 = 371.5$$

$$S_2 = 0.2 \times 267.4 + 0.8 \times 371.5 = 350.68$$

依次类推。根据表中数据作图如下：



由例题图可以看出 α =0.2时平滑值的波动幅度最小，即修匀效果好，α =0.7时平滑值的波动幅度最大。因此，α 的取值对平滑效果影响较大，α 越小平滑效果越显著。

一次指数平滑的特点。它实际是对过去的历史数据进行呈指数递减的加权平均。因为，

$$S_t = \alpha y_{t-1} + (1-\alpha) S_{t-1} \qquad S_{t-1} = \alpha y_{t-2} + (1-\alpha) S_{t-2}$$

$$S_t = \alpha y_t + (1-\alpha) S_{t-1} =$$

$$\alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots + \alpha(1-\alpha)^n y_{t-n+1} + (1-\alpha)^n S_{t-n+1}$$

显然，上式的系数呈指数衰减，这也是指数平滑名称的由来。同时也体现了各期的数据离 t 期越远，它的系数越小，因此对预测值的影响也越小。下表列出不同的 α 取值时各期观察值的权数。

表

期数	权数值		
	α =0.3	α =0.2	α =0.1
1	0.012106	0.026844	0.038742
2	0.01729	0.033554	0.043047
3	0.02471	0.041943	0.047830
4	0.03529	0.052429	0.053144
5	0.05042	0.065536	0.059049
6	0.07203	0.08192	0.06561
7	0.1029	0.1024	0.0729
8	0.147	0.128	0.081
9	0.21	0.16	0.09
10	0.3	0.2	0.1

表明，α 取值的大小决定了在平滑值中起作用的观察值项数的多少，当 α取值较大时，各观察值权数的递减速度很快，因此在平滑值中起作用的观察值的项数就少；而当 α取值较小时，各观察值权数的递减速度很慢，因此在平滑值中起作用的观察值的项数就多。

应用一次指数平滑进行预测应注意以下三点：

①平滑常数的选择。

一般来说，α取值的大小应当视所预测现象的特点及预测期的长短而定。在早期的研究中，一般认为简单指数平滑的平滑系数α一般选择在 0.10~0.30 之间，如果α超过 0.30，则应当配合更为复杂的模型。然而，后来的研究发现无论理论研究和经验数据都证明应当考虑比这一范围更宽的系数范围。同时，有证据表明，低估最佳参数的后果往往比高估更为严重。因此，参数值α的确定应当根据具体的数据而定，一般可以采用多种α值进行试算比较，选择其中的最佳参数，选择的原则是使得一期预测误差平方和 SSE 或平均平方误差 MSE 或平均绝对误差 MAE 为最小。

②初始值的确定。

初始值的选择正确与否直接关系着预测效果及平滑系数的选择。根据反映指数平滑由来的公式可以看出，当递推到n单位的时间后，初始值的权数为$(1-\alpha)^n$，因此α取值的大小与初始值在平滑值中所起作用的大小有直接的关系。α取值越小，在递推期数一定时，初始值权数越大；α取值越大，在递推期数一定时，初始值权数越小；同时，在α取值一定时，递推期数越少，初始值的权数越大。

初始值的确定一般有如下的方法：

第一、将资料分为两个部分，第一部分用于估计初始值，第二部分用于估计平滑系数，一般要有 8 个观察值或 3 个季节周期的数据用于估计初始值。

第二、用倒推法，即将数据倒置，从最近期的数据进行平滑，一直推到最远期，用最远期的平滑值作为初始值按照正常的顺序进行预测。

第三、用前几个（如十个）数的平均数作为初始值。

第四、当只有有限数据存在，不能用以上方法时，可用第一期数据作为初始值。

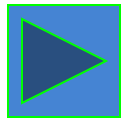
③误差分析。

一次指数平滑预测实际是根据本期预测误差对本期预测值作出一定的调整后得到下一个预测值，即：

$$S_{t+1} = S_t + \alpha(y_t - S_t)$$

$$S_{t+1} = S_t + \alpha(e_t)$$

新的预测值=预测值+ α × 本期预测值误差



二次指数平滑预测法

* 原理：二次指数平滑预测法是根据一次指数平滑值、二次指数平滑值的结果建立模型进行预测方法。
  * 二次指数平滑又称双重指数平滑，它是对一次指数平滑值再进行一次平滑。
  * 当预测对象存在线性发展趋势时，一次指数平滑值滞后于实际值（如表）
  * 这样就需要进行一定的修正，因此，预测的一般模型应为：

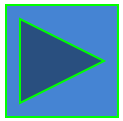
一次指数平滑和二次指数平滑公式为：

$$S_t^{(1)} = \alpha Y_t + (1-\alpha) S_{t-1}^{(1)}$$

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)}$$

表　数列存在线性趋势时一次、二次指数平滑值滞后实例

时期	观察值	一次指数平滑值	二次指数平滑值
1	2	1.00	0.500
2	4	2.50	0.375
3	6	4.25	2.313
4	8	6.13	4.219
5	10	8.06	6.141
6	12	10.03	8.086
7	14	12.02	10.053
8	16	14.01	12.031
9	18	16.00	14.016



$$F_{t+m} = a_t + b_t m$$

a_t第t期数列水平的平滑值

b_t第t期趋势值的平滑值

m预测的超前期数

显然：

一次指数平滑是直接利用平滑值作为预测值。

二次指数平滑是用平滑值对时序存在的线性趋势进行修正。

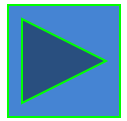
由于 a_t b_t 的估计方法不同，产生两种不同的公式。

1、布朗（Brown）单一参数线性指数平滑预测法。

2、霍特（HOLT）双参数指数平滑预测法

1、布郎（Brown）单一参数线性指数平滑预测法

* 首先：计算两次平滑值
  * 进行一次指数平滑和二次指数平滑，即
* 其次：建立预测模型，估计相应的参数
  * 预测模型为：
* 应用实例
* 特点以及适用场合
  * 适用于具有线性变化趋势的时序进行短期预测 。



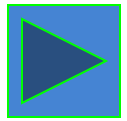
首先：计算两次平滑值

进行一次指数平滑和二次指数平滑，即

$$S_t^{(1)} = \alpha Y_t + (1-\alpha) S_{t-1}^{(1)}$$

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)}$$

当预测对象存在线性趋势时，简单指数平滑值滞后于实际值，布郎证明，这种滞后量在初始调整阶段后，为 $\frac{1-\alpha}{\alpha} b_t$，其中 b_t 为第 t 期趋势值的平滑值。即当趋势存在时，一次和二次平滑值都滞后于实际值，将一次平滑值和二次平滑值之差加在一次平滑值上，则可对趋势进行修正。



预测模型为：

$$F_{t+m} = a_t + b_t m$$

a_t第t期数列水平的平滑值

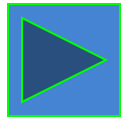
b_t第t期趋势值的平滑值

m预测的超前期数

由此得 a_t b_t 的估计公式为：

$$a_t = S_t^{(1)} + \left[S_t^{(1)} - S_t^{(2)}\right] = 2S_t^{(1)} - S_t^{(2)}$$

$$b_t = \frac{\alpha}{1-\alpha}\left[S_t^{(1)} - S_t^{(2)}\right]$$

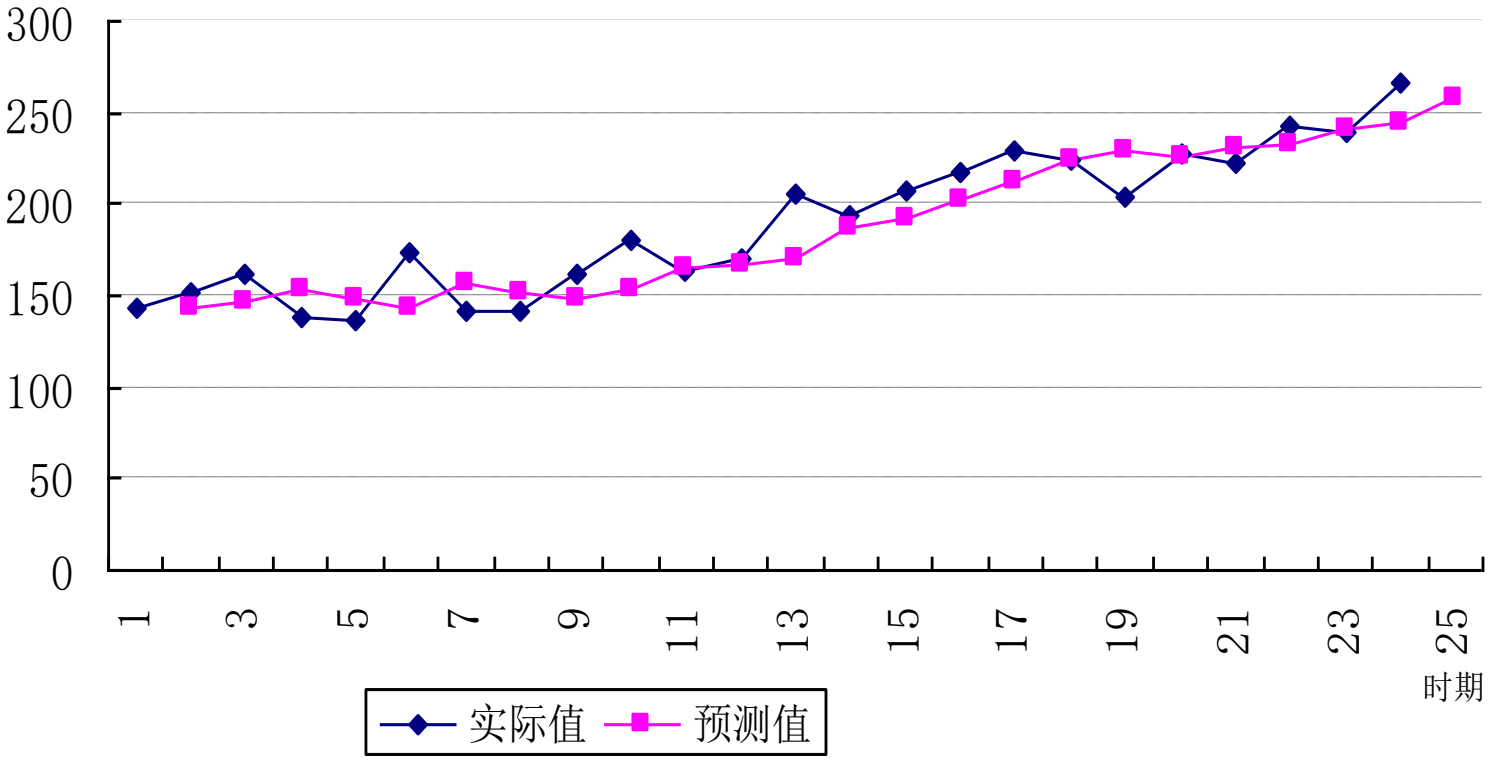


某银行月度居民储蓄存款余额如下表：

月份	期数	存款余额（万元）
1	1	143
2	2	152
3	3	161
4	4	139
5	5	137
6	6	174
7	7	142
8	8	141
9	9	162
10	10	180
11	11	164
12	12	171
1	13	206
2	14	193
3	15	207
4	16	218
5	17	229
6	18	225
7	19	204
8	20	227
9	21	223
10	22	242
11	23	239
12	24	266

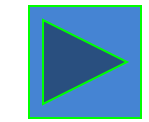
	A	B	C	D	E	F	G	H
1	月份	期数	存款余额	一次指数平滑	二次指数平滑	a值	b值	预测值
2		0		143	143			
3	1	1	143	143	143	143	0	
4	2	2	152	144.8	143.36	146.24	0.36	143
5	3	3	161	148.04	144.296	151.784	0.936	146.6
6	4	4	139	146.232	144.6832	147.7808	0.3872	152.72
7	5	5	137	144.3856	144.6237	144.1475	-0.05952	148.168
8	6	6	174	150.3085	145.7606	154.8563	1.13696	144.088
9	7	7	142	148.6468	146.3379	150.9557	0.577229	155.9933
10	8	8	141	147.1174	146.4938	147.7411	0.155912	151.5329
11	9	9	162	150.0939	147.2138	152.9741	0.720032	147.897
12	10	10	180	156.0752	148.9861	163.1642	1.772268	153.6941
13	11	11	164	157.6601	150.7209	164.5994	1.734808	164.9365
14	12	12	171	160.3281	152.6423	168.0139	1.921442	166.3342
15	1	13	206	169.4625	156.0064	182.9186	3.36403	169.9353
16	2	14	193	174.17	159.6391	188.7009	3.632724	186.2826
17	3	15	207	180.736	163.8585	197.6135	4.21938	192.3336
18	4	16	218	188.1888	168.7245	207.653	4.866065	201.8329
19	5	17	229	196.351	174.2498	218.4522	5.5253	212.5191
20	6	18	225	202.0808	179.816	224.3456	5.566199	223.9775
21	7	19	204	202.4647	184.3458	220.5836	4.529726	229.9118
22	8	20	227	207.3717	188.9509	225.7925	4.605195	225.1133
23	9	21	223	210.4974	193.2602	227.7345	4.309287	230.3977
24	10	22	242	216.7979	197.9678	235.628	4.707534	232.0438
25	11	23	239	221.2383	202.6219	239.8548	4.654111	240.3356
26	12	24	266	230.1907	208.1356	252.2457	5.513756	244.5089
27		25						257.7594

预测效果见下图：



2、霍特（HOLT）双参数指数平滑预测法

* 原理：
  * 首先：计算两类平滑值：
    * 一次指数平滑值　趋势值平滑值
  * 其次：建立预测模型，进行参数估计
    * 预测模型
    * 这种方法是对数列的水平和趋势值分别进行平滑，然后利用这两次平滑的结果进行线性外推预测。
* 应用实例
* 特点及适用场合：
  * 适用于具有线性变化趋势的时序进行短期预测 。
  * 与布郎法的比较



首先：计算两类平滑值。一是，一次指数平滑值（对原数列进行平滑）；

$$S_t = \alpha y_t + (1-\alpha)(S_{t-1} + b_{t-1})$$

这里是把上一期的趋势值b_{t-1}加到S_{t-1}上，以便消除一个滞后，修正S_t，使其与实际的观察值y_t尽可能地接近。

二是，计算趋势值平滑值（对序列的趋势进行平滑）。

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}$$

上式类似于一次指数平滑公式，所不同的是用一个差值$S_t - S_{t-1}$来表示序列的趋势值，因此是对序列的趋势进行平滑

$$F_{t+m} = S_t + b_t m$$

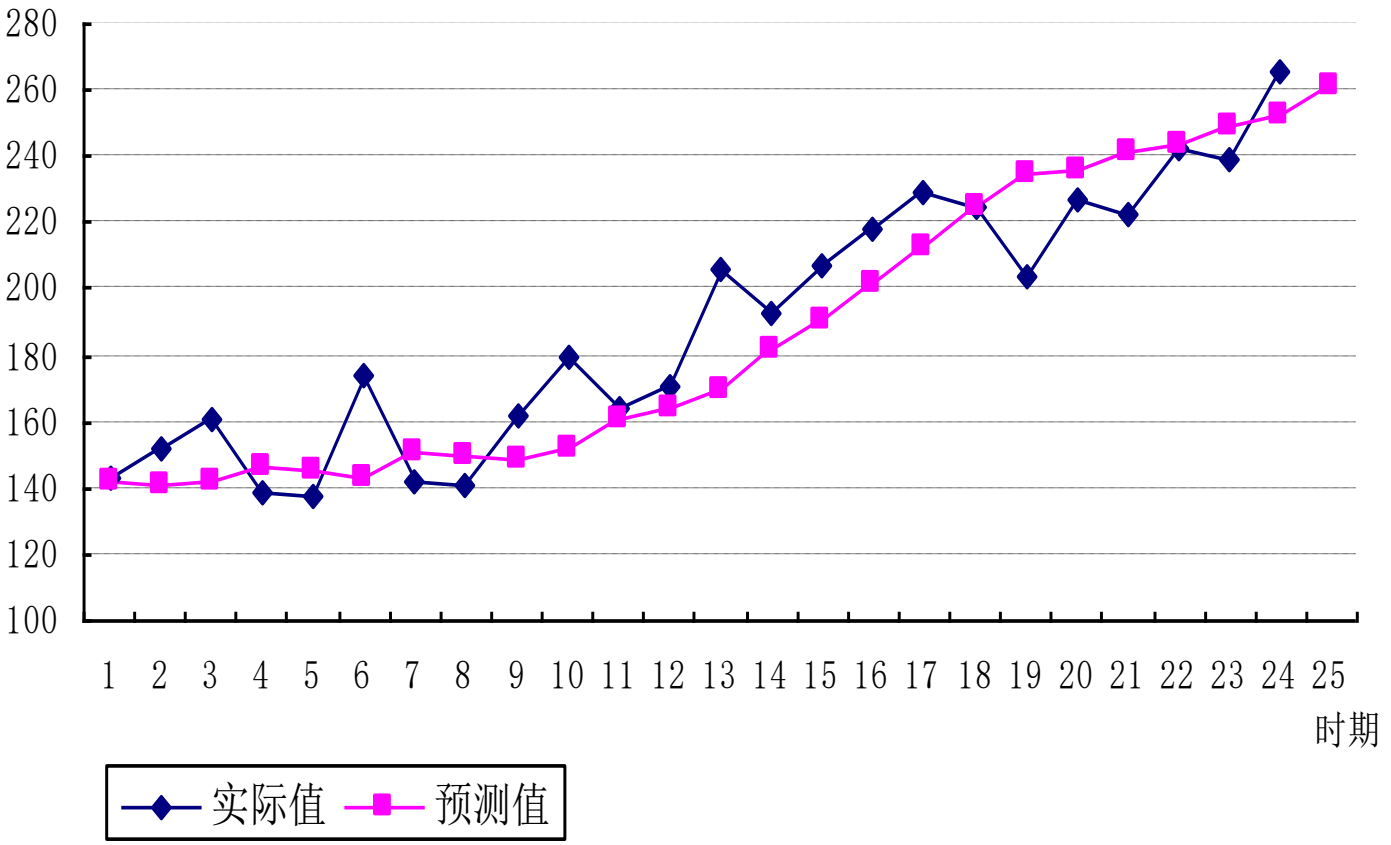
S_t 表示第 t 期数列水平的平滑值

b_t 表示第 t 期趋势值的平滑值

m 表示预测的超前期数

	A	B	C	D	E	F
1	月份	期数	存款余额	数据的平滑	趋势的平滑	预测值
2		0		143	-1.33	
3	1	1	143	141.936	-1.2502	141.67
4	2	2	152	142.9486	-0.57135	140.6858
5	3	3	161	146.1018	0.546014	142.3773
6	4	4	139	145.1183	0.087144	146.6478
7	5	5	137	143.5643	-0.40518	145.2054
8	6	6	174	149.3273	1.445269	143.1592
9	7	7	142	149.0181	0.918913	150.7726
10	8	8	141	148.1496	0.382694	149.937
11	9	9	162	151.2258	1.190757	148.5323
12	10	10	180	157.9333	2.845762	152.4166
13	11	11	164	161.4232	3.03902	160.779
14	12	12	171	165.7698	3.431285	164.4622
15	1	13	206	176.5609	5.639221	169.2011
16	2	14	193	184.3601	6.287216	182.2001
17	3	15	207	193.9178	7.268379	190.6473
18	4	16	218	204.549	8.277206	201.1862
19	5	17	229	216.0609	9.247636	212.8262
20	6	18	225	225.2469	9.229122	225.3086
21	7	19	204	228.3808	7.400563	234.476
22	8	20	227	234.0251	6.873682	235.7813
23	9	21	223	237.319	5.799757	240.8988
24	10	22	242	242.895	5.732631	243.1188
25	11	23	239	246.7021	5.154972	248.6276
26	12	24	266	254.6857	6.003547	251.8571
27		25				260.6892

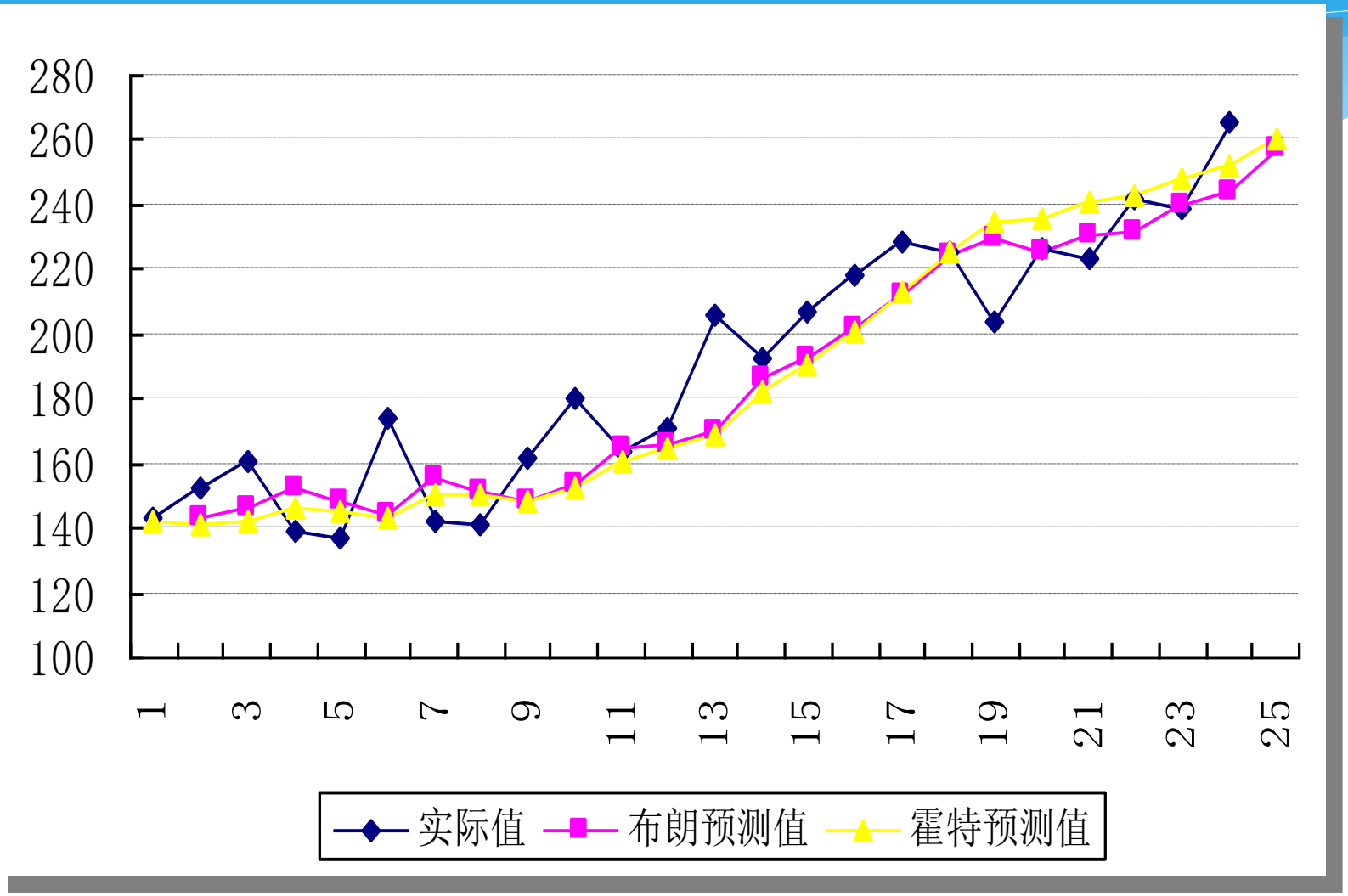
预测效果见下图：


280
260
240
220
200
180
160
140
120
100
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
时期
实际值
预测值

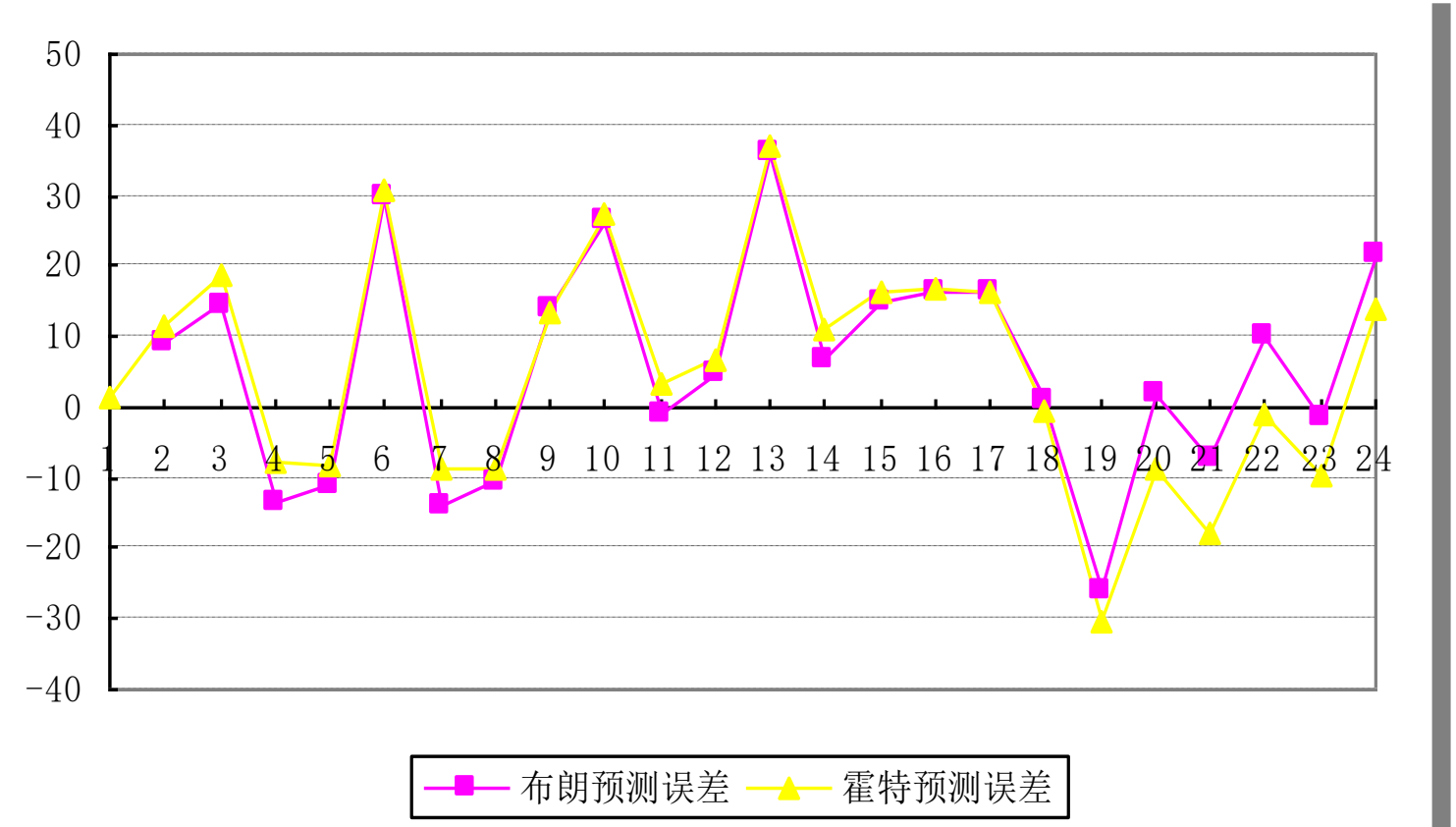
与布郎法比较，霍特方法由于多使用了一个参数，因此预测时更具灵活性，尤其是在自动地用于不包含长期趋势的大量时间数列的预测时，霍特法可以通过参数 γ 的选择使 b_t 接近 0 的数，从而使模型接近于简单的指数平滑预测模型，而布郎法则无法进行类似的调节，对所有的数列都只能按照线性趋势进行预测。因此实际应用中霍特法的预测精度一般要高于布郎法。但是霍特法选择平滑系数的工作量要比布郎法大。

事实上，布郎法实际是霍特法的一个特例。因为，如果用 α_k 表示霍特法中的参数 α，用 α_b 表示布郎法中的参数 α。当 $\alpha_k = 1-(1-\alpha_b)^2$，$\gamma = 1-\frac{2(1-\alpha_b)}{2-\alpha_b}$ 时，两种方法是相同的。只要将 α_k，γ 代入霍特的公式中，就可得到验证。

根据例题，布郎单参数指数平滑预测法与霍特（HOLT）双参数指数平滑预测法其预测结果比较可如下图：


280
260
240
220
200
180
160
140
120
100
1
3
5
7
9
11
13
15
17
19
21
23
25
实际值
布朗预测值
霍特预测值

他们的误差图如下：

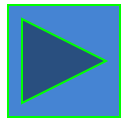

50
40
30
20
10
0
-10
-20
-30
-40
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
布朗预测误差
霍特预测误差

三、三次指数平滑预测法

* 布朗三次指数平滑预测法
* 温特线性和季节性指数平滑预测方法（乘法模型）
* 温特线性和季节性指数平滑预测方法（加法模型）

1、布朗三次指数平滑预测法

* 原理：
  * 首先：计算三次指数平滑值
  * 其次：建立预测模型，估计相应的参数
    * 显然，这里是一个非线性平滑模型，类似于一个二次多项式，能够表现时序的一种曲线变化趋势。
* 应用实例
* 特点及适用场合
  * 具有非线性趋势的预测



首先计算三次指数平滑值，

$$S_t^{(1)} = \alpha Y_t + (1-\alpha) S_{t-1}^{(1)}$$

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)}$$

$$S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha) S_{t-1}^{(3)}$$

其次：建立预测模型，估计相应的参数，

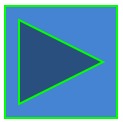
$$F_{t+m} = a_t + b_t m + \frac{1}{2} c_t m^2$$

其中：

$$a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)}$$

$$b_t = \frac{\alpha}{2(1-\alpha)}\left[(6-5\alpha)S_t^{(1)} - (10-8\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}\right]$$

$$c_t = \frac{\alpha^2}{(1-\alpha)^2}\left(S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}\right)$$



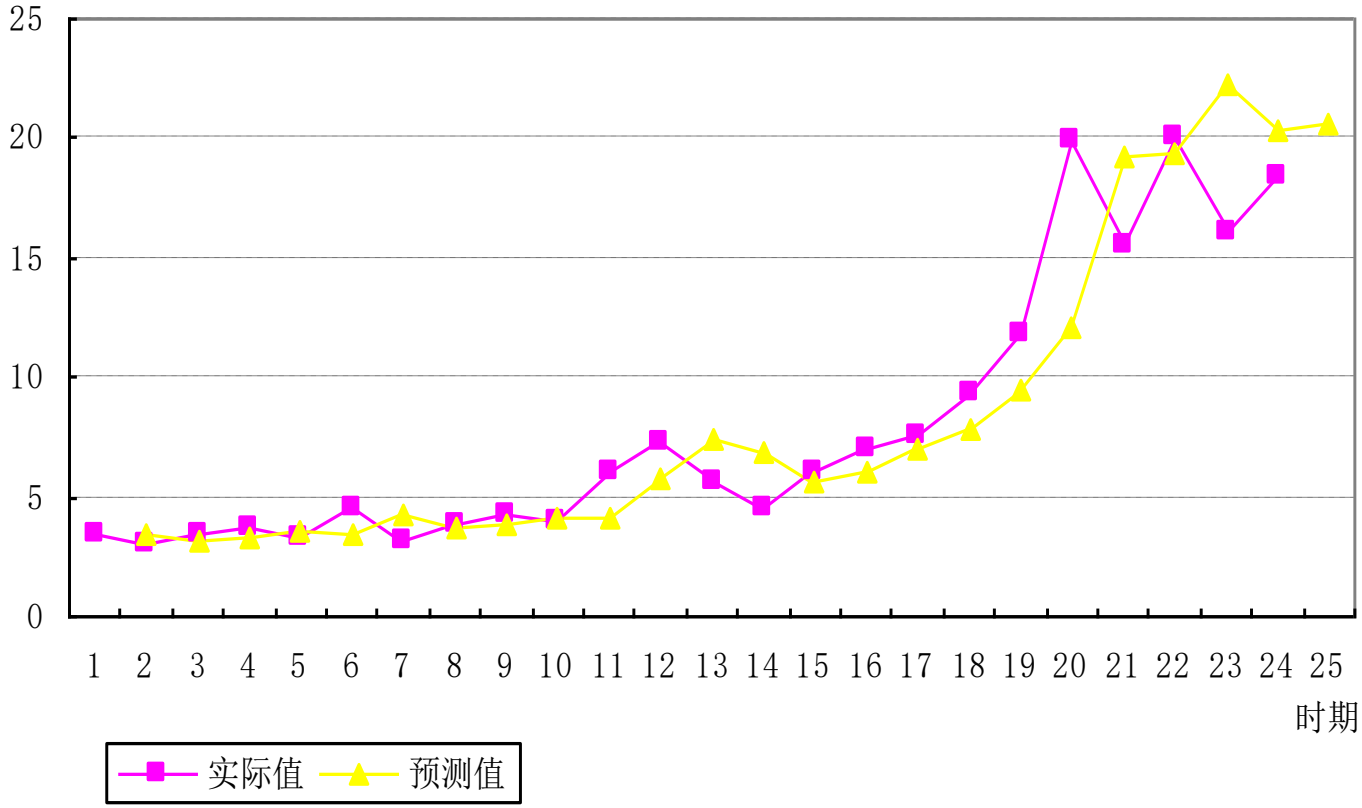
（1）应用实例。

某电视机销售量资料如下表：

月份	期数	电视机销售量（万台）
1	1	3.4
2	2	3
3	3	3.4
4	4	3.7
5	5	3.3
6	6	4.6
7	7	3.2
8	8	3.8
9	9	4.2
10	10	4
11	11	6.1
12	12	7.3
1	13	5.7
2	14	4.5
3	15	6
4	16	7
5	17	7.6
6	18	9.3
7	19	11.8
8	20	19.9
9	21	15.5
10	22	20.1
11	23	16.1
12	24	18.4

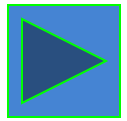
	A	B	C	D	E	F	G	H	I	J
1	月份	期数	电视机销售量	一次平滑值	二次平滑值	三次平滑值	a值	b值	c值	预测值
2		0		3.4	3.4	3.4				
3	1	1	3.4	3.4	3.4	3.4	3.4	-3.9E-16	0	
4	2	2	3	3.3	3.375	3.39375	3.16875	-0.06563	-0.00313	3.4
5	3	3	3.4	3.325	3.3625	3.385938	3.273438	-0.02266	-0.00078	3.1
6	4	4	3.7	3.41875	3.376563	3.383594	3.510156	0.049609	0.002734	3.25
7	5	5	3.3	3.389063	3.379688	3.382617	3.410742	0.012012	0.000684	3.5625
8	6	6	4.6	3.691797	3.457715	3.401392	4.103638	0.206409	0.009875	3.423438
9	7	7	3.2	3.568848	3.485498	3.422418	3.672467	0.042422	0.001126	4.319922
10	8	8	3.8	3.626636	3.520782	3.447009	3.764569	0.058453	0.001782	3.716016
11	9	9	4.2	3.769977	3.583081	3.481027	4.041714	0.123573	0.004713	3.824805
12	10	10	4	3.827483	3.644181	3.521816	4.071719	0.105109	0.003385	4.170001
13	11	11	6.1	4.395612	3.832039	3.599372	5.29009	0.426845	0.018384	4.180214
14	12	12	7.3	5.121709	4.154457	3.738143	6.6399	0.720318	0.030608	5.735319
15	1	13	5.7	5.266282	4.432413	3.91171	6.413317	0.504132	0.017398	7.390825
16	2	14	4.5	5.074711	4.592987	4.08203	5.527201	0.139461	-0.00162	6.934847
17	3	15	6	5.306033	4.771249	4.254334	5.858688	0.191168	0.000993	5.665038
18	4	16	7	5.729525	5.010818	4.443455	6.599577	0.348873	0.008408	6.050848
19	5	17	7.6	6.197144	5.307399	4.659441	7.328674	0.471205	0.013433	6.956858
20	6	18	9.3	6.972858	5.723764	4.925522	8.672803	0.74198	0.025047	7.813312
21	7	19	11.8	8.179643	6.337734	5.278575	10.8043	1.17929	0.043486	9.439831
22	8	20	19.9	11.10973	7.530734	5.841615	16.57861	2.557913	0.104993	12.02708
23	9	21	15.5	12.2073	8.699875	6.55618	17.07845	2.154057	0.075763	19.24152
24	10	22	20.1	14.18047	10.07002	7.434641	19.76599	2.435475	0.081948	19.30827
25	11	23	16.1	14.66036	11.21761	8.380383	18.70863	1.584905	0.03364	22.28341
26	12	24	18.4	15.59527	12.31202	9.363293	19.21303	1.336009	0.018584	20.32717
27		25								20.56762

预测效果见下图：


25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
时期
实际值
预测值

2、温特线性和季节性指数平滑预测方法（乘法模型）

* 原理：
  * 首先，计算三类平滑值
    * 水平平滑值
    * 趋势平滑值
    * 季节平滑值
  * 其次，建立预测模型，估计相应的参数
* 应用实例
* 特点以及适用场合
  * 既有线性趋势又有季节变动而且季节变动的幅度随着趋势的增长而增加的时间序列的短期预测



首先，计算三类平滑值。

一是，总平滑值

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1})$$

二是，趋势平滑值

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}$$

季节平滑值

$$I_t = \beta \frac{y_t}{S_t} + (1-\beta)I_{t-L}$$

L 为季节周期长度， I 为季节调整因子。

$$F_{t+m} = (S_t + b_t m) I_{t-L+m}$$

模型中包括时序的三种成分：平稳性（S_t），趋势性（b_t）和季节性（I_t）。

（2）应用实例

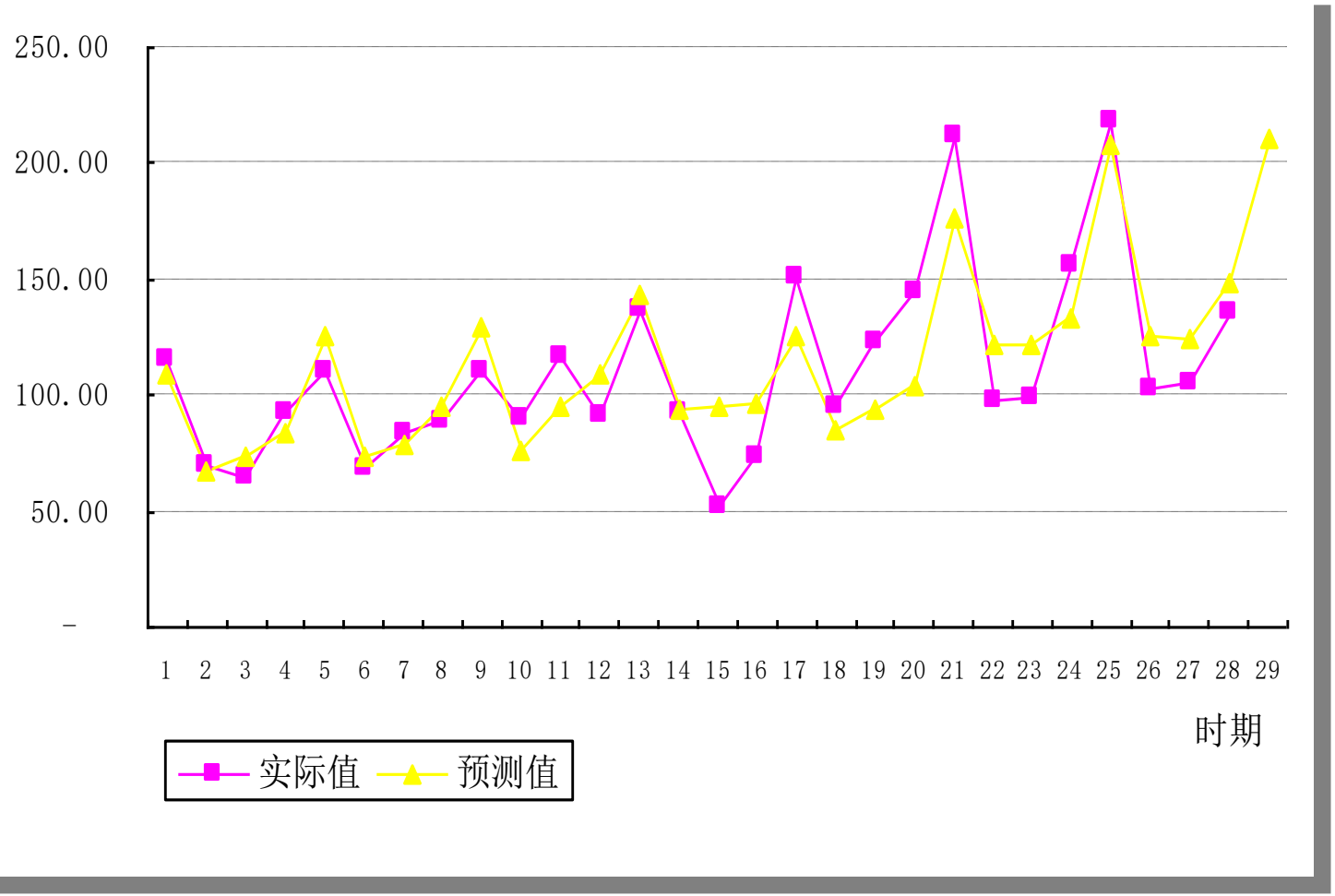
某地区电视机季度销售资料如下

某地区电视机季度销售量（万台）

年份 季度	1996	1997	1998	1999	2000	2001	2002
一	115	109.9	110.1	136.6	151.6	212.2	218
二	69.9	68.2	90.3	93	94.6	97.6	103
三	64.5	84.2	117.3	51.7	123.6	99.6	105.6
四	92.7	89.4	90.9	74	144.8	156.6	135.2

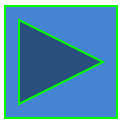
作图如（　　），可见存在明显的季节变动，即：一季度高，二季度、三季度低，四季度回升，同时还存在增长趋势，随着销售量的增长，季节变动的幅度也随之增大，因此属于相乘型季节变动，可以用温特线性和季节性指数平滑预测。

	A	B	C	D	E	F	G
1	季度	时期	销售量	数据平滑	趋势平滑	季节平滑	预测值
2		-3				1.37	
3		-2				0.81	
4		-1				0.851	
5		0		77.254	2.335	0.964	
6	一	1	115.00	80.4595	2.5091	1.39	109.0369
7	二	2	69.90	83.6342	2.64221	0.82	67.2046
8	三	3	64.50	84.1797	2.22288	0.83	73.5286
9	四	4	92.70	88.3545	2.61325	0.99	83.2921
10	一	5	109.90	88.6123	2.14218	1.34	126.2438
11	二	6	68.20	89.2838	1.84804	0.80	74.2131
12	三	7	84.20	93.3036	2.28239	0.85	78.9125
13	四	8	89.40	94.5376	2.0727	0.98	94.5876
14	一	9	110.10	93.678	1.48624	1.29	129.7979
15	二	10	90.30	98.6621	2.18582	0.84	76.2809
16	三	11	117.30	108.323	3.68085	0.92	95.0493
17	四	12	90.90	108.223	2.92463	0.94	109.359
18	一	13	136.60	110.046	2.70438	1.28	143.7199
19	二	14	93.00	112.458	2.64586	0.83	94.2227
20	三	15	51.70	103.336	0.29223	0.79	95.2236
21	四	16	74.00	98.7236	-0.6886	0.88	96.9387
22	一	17	151.60	102.161	0.13663	1.34	125.2424
23	二	18	94.60	104.55	0.58698	0.85	85.2207
24	三	19	123.60	115.269	2.61353	0.88	93.5192
25	四	20	144.80	127.227	4.48239	0.96	103.7
26	一	21	212.20	137.052	5.55094	1.40	176.4182
27	二	22	97.60	136.924	4.41506	0.81	121.868
28	三	23	99.60	135.785	3.30422	0.83	121.982
29	四	24	156.60	143.991	4.28464	1.00	133.138
30	一	25	218.00	149.717	4.57282	1.42	207.8985
31	二	26	103.00	148.799	3.47477	0.78	125.292
32	三	27	105.60	147.144	2.44881	0.80	124.754
33	四	28	135.20	146.814	1.89306	0.97	149.043
34		29					210.9116


250.00
200.00
150.00
100.00
50.00
-
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
时期
实际值
预测值

3、温特线性和季节性指数平滑预测方法（加法模型）

* 原理：
  * 首先，计算三类平滑值
    * 水平平滑值
    * 趋势平滑值
    * 季节平滑值
  * 其次，建立预测模型，估计相应的参数
* 应用实例
* 特点以及适用场合
  * 既有线性趋势又有季节变动而且季节变动的幅度随着趋势的增长而增加的时间序列的短期预测



（1）原理

首先，计算三类平滑值。

一是，总平滑值

$$S_t = \alpha(y_t - I_{t-L}) + (1-\alpha)(S_{t-1} + b_{t-1})$$

二是，趋势平滑值

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}$$

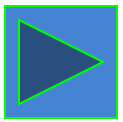
季节平滑值

$$I_t = \beta(y_t - S_t) + (1-\beta)I_{t-L}$$

L 为季节周期长度， I 为季节增减量。

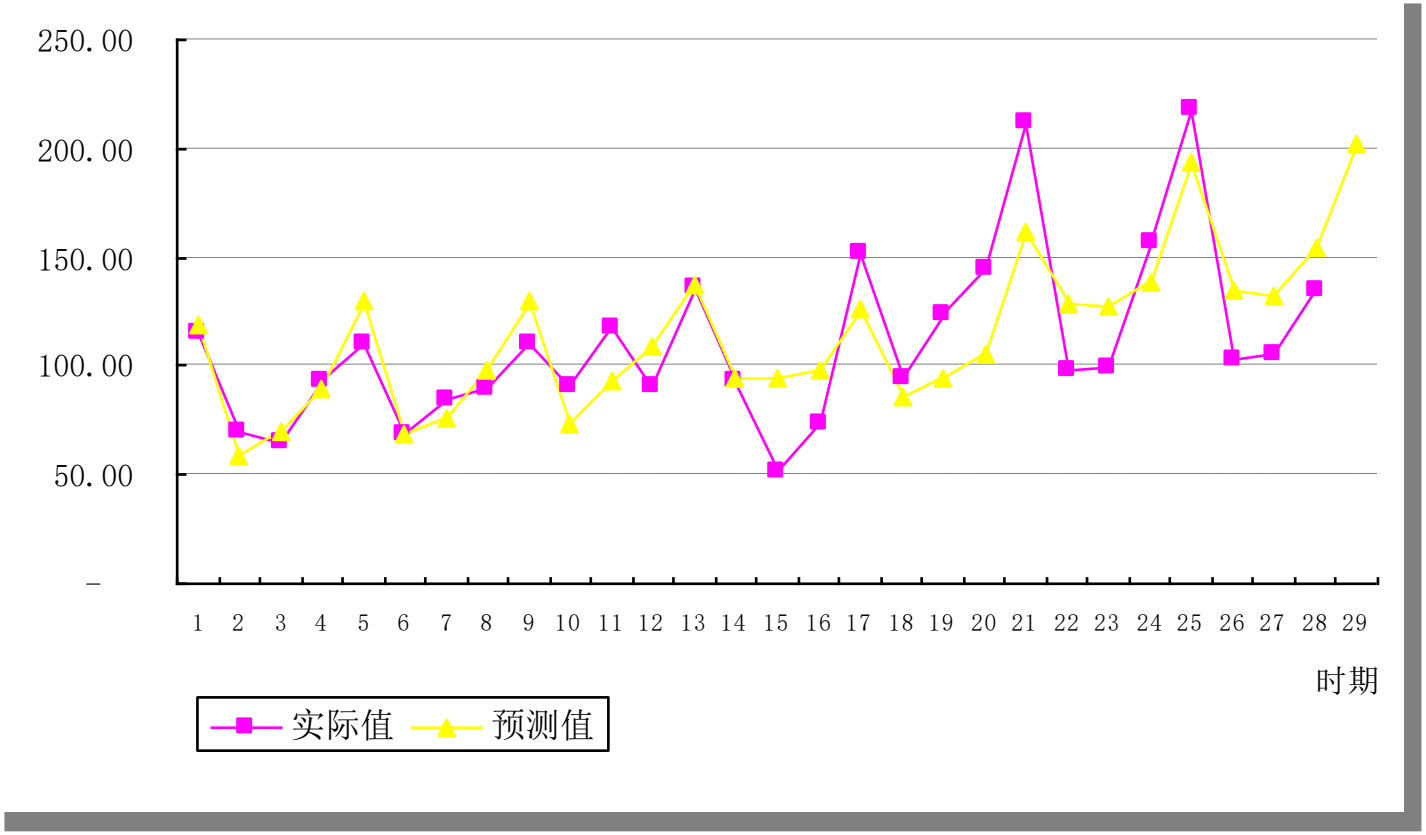
其次，建立预测模型，估计相应的参数

$$F_{t+m} = (S_t + b_t m) + I_{t-L+m}$$

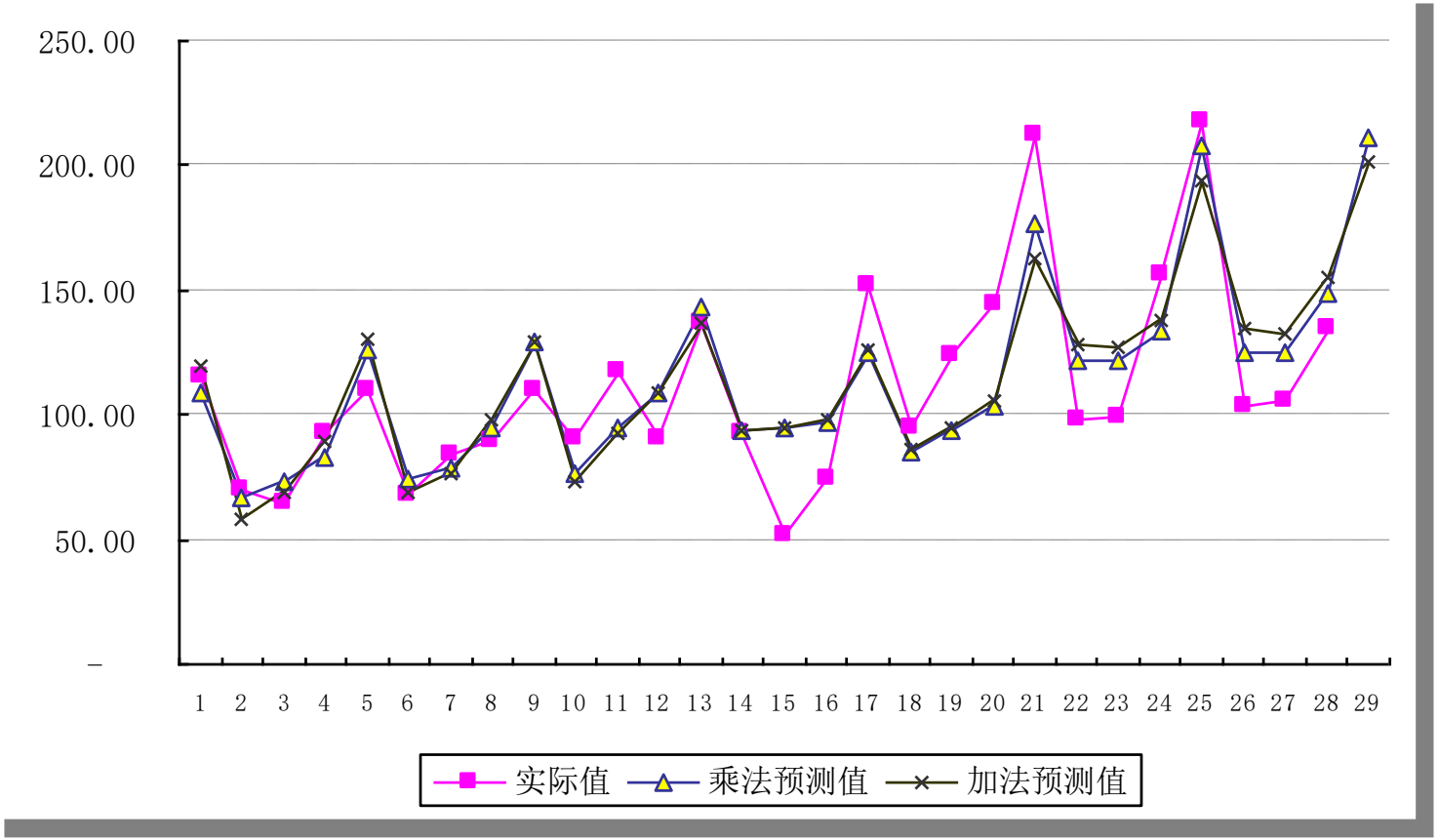


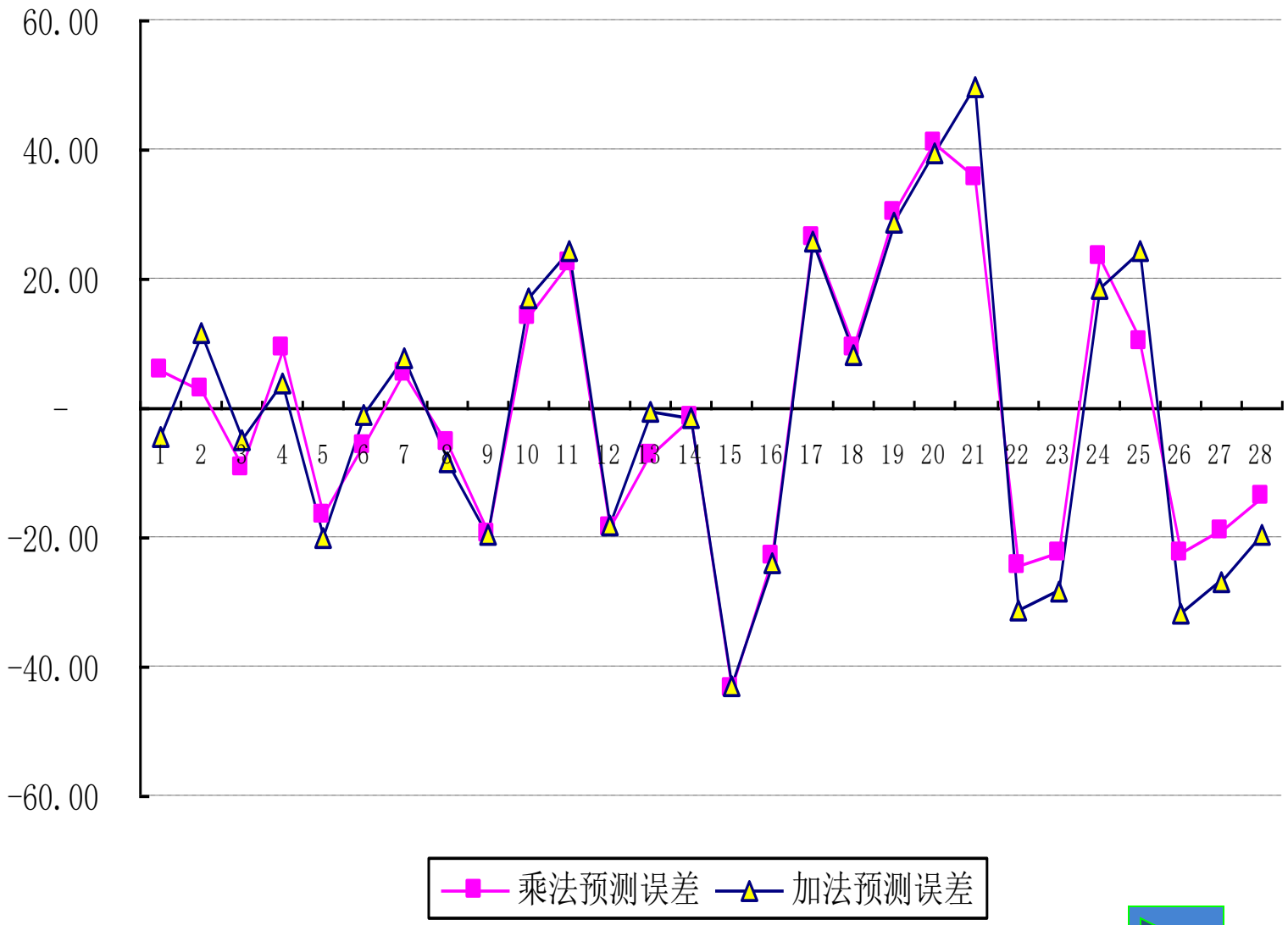
	A	B	C	D	E	F	G
1	季度	时期	销售量	数据平滑	趋势平滑	季节平滑	预测值
2		-3				39.76786	
3		-2				-22.6321	
4		-1				-18.3607	
5		0		77.254	2.335	1.225	
6	一	1	115.00	78.7176	2.16073	38.72	119.3569
7	二	2	69.90	83.2091	2.62688	-19.84	58.2463
8	三	3	64.50	85.2409	2.50786	-19.07	69.3881
9	四	4	92.70	88.494	2.65691	2.12	88.9738
10	一	5	109.90	87.1563	1.85799	33.93	129.8732
11	二	6	68.20	88.8185	1.81882	-20.07	69.1791
12	三	7	84.20	93.1648	2.32432	-16.04	76.4143
13	四	8	89.40	93.8474	1.99599	0.15	97.6084
14	一	9	110.10	91.909	1.2091	29.21	129.7721
15	二	10	90.30	96.5685	1.89919	-15.93	73.0479
16	三	11	117.30	105.443	3.29415	-7.67	92.6949
17	四	12	90.90	105.139	2.57471	-4.17	108.886
18	一	13	136.60	107.65	2.56185	29.13	136.9215
19	二	14	93.00	109.955	2.51057	-16.24	94.282
20	三	15	51.70	101.847	0.38681	-20.41	94.5619
21	四	16	74.00	97.4206	-0.5759	-9.94	98.0666
22	一	17	151.60	101.97	0.44915	35.28	125.9749
23	二	18	94.60	104.103	0.78589	-14.22	86.1815
24	三	19	123.60	112.714	2.35093	-11.02	94.6501
25	四	20	144.80	123	3.93808	-0.42	105.121
26	一	21	212.20	136.935	5.93733	47.28	162.2187
27	二	22	97.60	136.661	4.69513	-21.67	128.655
28	三	23	99.60	135.21	3.46585	-18.40	127.651
29	四	24	156.60	142.345	4.19964	3.98	138.255
30	一	25	218.00	151.38	5.16685	53.08	193.82
31	二	26	103.00	150.172	3.89177	-29.32	134.877
32	三	27	105.60	148.051	2.68923	-25.62	132.34
33	四	28	135.20	146.836	1.90833	-0.70	154.722
34		29					201.8228

预测效果见下图：


250.00
200.00
150.00
100.00
50.00
-
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
时期
实际值
预测值

乘法模型与加法模型的预测结果比较如下：




60.00
40.00
20.00
-
-20.00
-40.00
-60.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
乘法预测误差
加法预测误差