

时间序列分析 (TIME SERIES ANALYSIS)

主讲：吴尚

复旦大学管理学院统计与数据科学系

教材与参考书

- 教材：克莱尔 等著，潘红宇 等 译，时间序列分析及应用（R语言），机械工业出版社，2011
- 参考书：詹姆斯·D·汉密尔顿，时间序列分析(上下册)，中国人民大学出版社，2015

成绩评定

- 平时成绩40分（评教占3分，剩下的37分按照如下比例分配）
 - 考勤12.5%
 - 作业37.5%
 - 期中考试50%，闭卷
- 期末考试60分，闭卷

联系方式

■ 主讲：吴尚，统计与数据科学系讲师

- 邮箱：shangwu@fudan.edu.cn
- 办公地点：思源楼615室
- 开放办公时间：每周五下午13:00-17:00，节假日除外

■ 助教：

- 苗子：22210690188@m.fudan.edu.cn
- 程子英：22210690187@m.fudan.edu.cn

概率统计回顾

■ 概率模型：

- 给定概率模型，计算概率、期望等
- 离散模型，连续模型.....

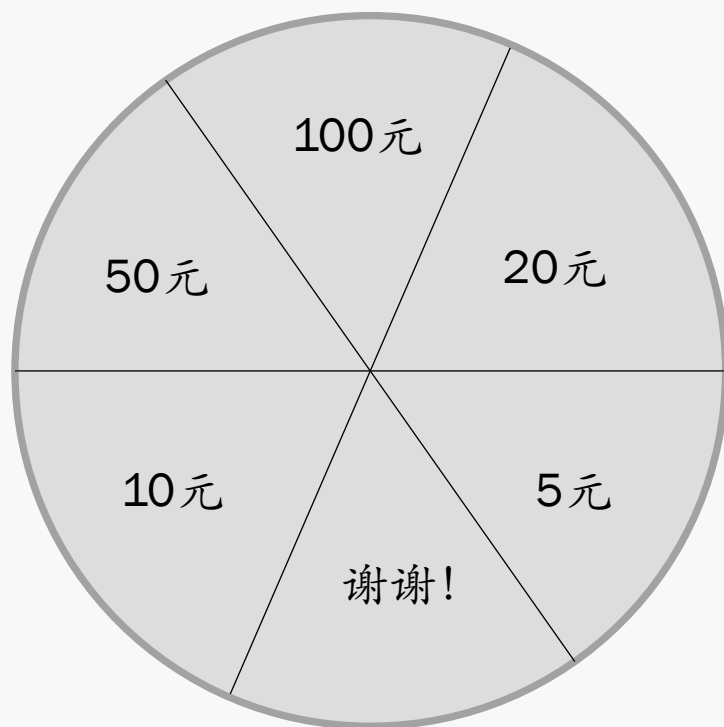
■ 统计模型：

- 给定数据，选择合适的模型
- 参数估计、模型诊断、模型优化、预测.....

■ 统计的基础是概率：

- 对于想要选择的模型，需要先熟悉其概率性质

例一：轮盘赌



概率模型：

假设

$$P(X=100)=0.01$$

$$P(X=50)=0.02$$

$$P(X=20)=0.03$$

$$P(X=10)=0.04$$

$$P(X=5)=0.05$$

$$P(X=0)=0.85$$

则玩一次期望得到的金额为：

$$0.01*100+0.02*50+0.03*20+0.04*10+0.05*5+0=3.25$$

如果玩一次需要5元，那么。。

轮盘赌

- 该模型是一个多元分布模型（离散）。
- 假设我们一共玩了200次，其中出现了2次100元，4次50元，6次20元，8次10元，10次5元，剩下170次都是“谢谢！”。
- 通过最大似然估计（对于本问题，即用频率估计概率），我们得到了上述模型参数。
- 思考题：通过该估计得到的模型的准确性如何？
- 容易发现，玩的总次数越多，所得的模型越接近真实模型。

例二：均值模型

- 对于一个零件的长度进行反复测量，每次测量结果都是独立的，满足

$$X_i = \mu + \varepsilon_i$$

- 其中，假设 $E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2$, 且 ε_i 相互独立。
- 则该模型具有如下性质：
- $E(X_i) = \mu, Var(X_i) = \sigma^2, Cov(X_i, X_j) = 0, \dots$
- 若 ε_i 的分布关于0对称（如正态分布），则 $P(X_i > \mu) = 0.5$

均值模型

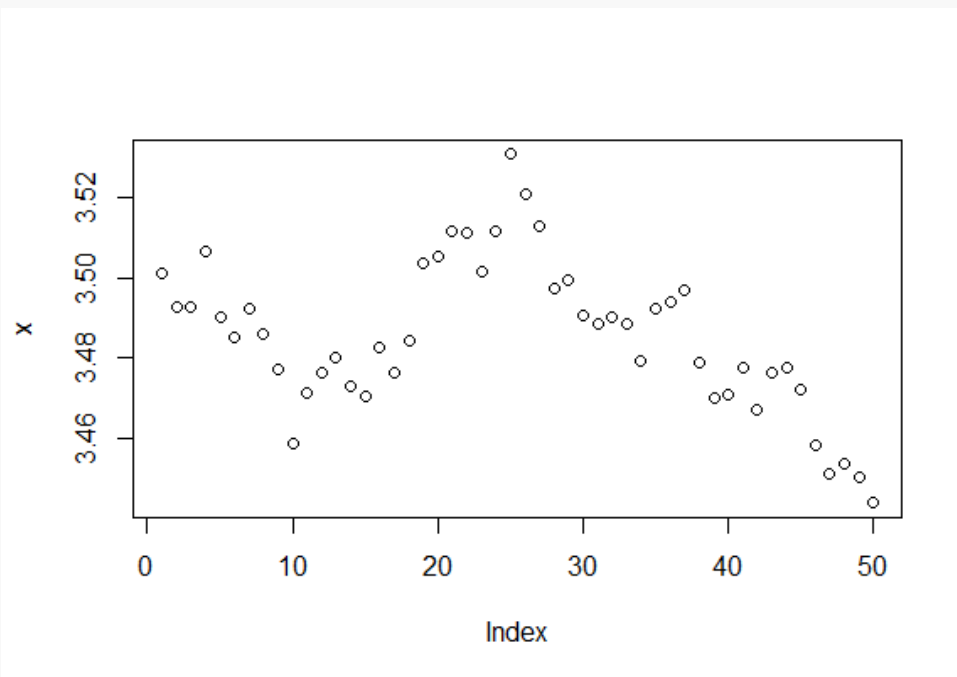
- 假设经过7次测量，得到的测量结果为
- 3.5, 3.49, 3.55, 3.51, 3.46, 3.47, 3.52
- 则样本均值为

$$(3.5+3.49+3.55+3.51+3.46+3.47+3.52)/7=3.5$$

- 问题一：样本均值和真实均值差多少？
- 问题二：真实均值是3可能吗？为什么？

均值模型

- 假设我们用了一种带有“惯性”的测量工具，多次测量结果的绘图如下：



问题三：如果我们继续用上述模型，会产生什么问题？

问题四：如何选择更合适的模型？

什么是时间序列

- 同一现象在不同时间上的相继观察值排列而成的序列。
- 用 t 表示观察的时刻, X 表示观察的数值, 则 $X_t, t = 0, 1, 2, \dots$ 为时间 t 上的观察值, 时间序列可表示为:

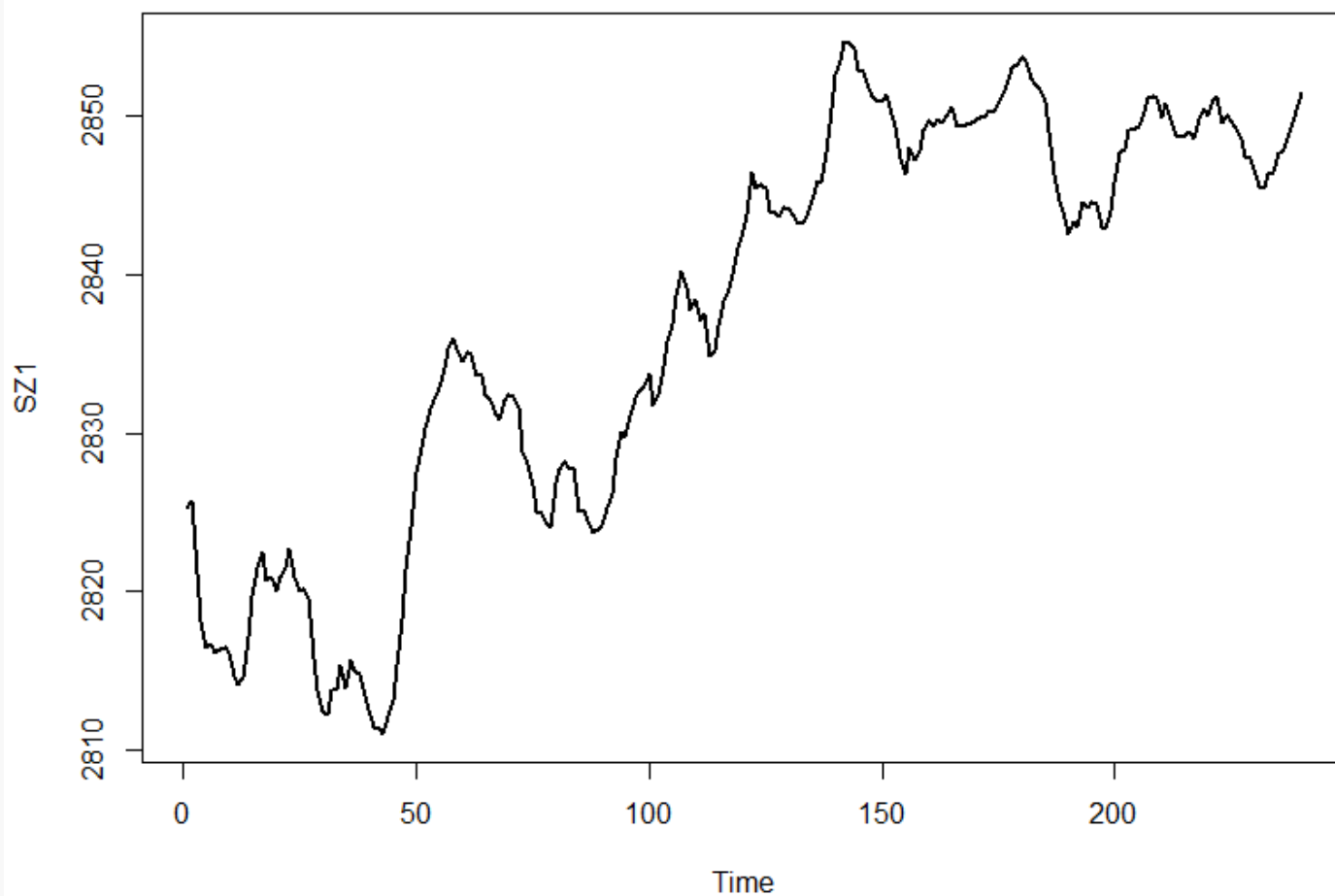
$$\{X_t, t = 0, 1, 2, \dots\}$$

时 间(t)	0	1	...	n	...
观察值(X_t)	X_0	X_1	...	X_n	...

- 时间序列亦可从 $t = 1$ 开始, 甚至追溯到 $X_{-1}, X_{-2}, X_{-3}, \dots$
- 当然, 也可用其他大写字母表示时间序列, 如 Y_t, W_t 等

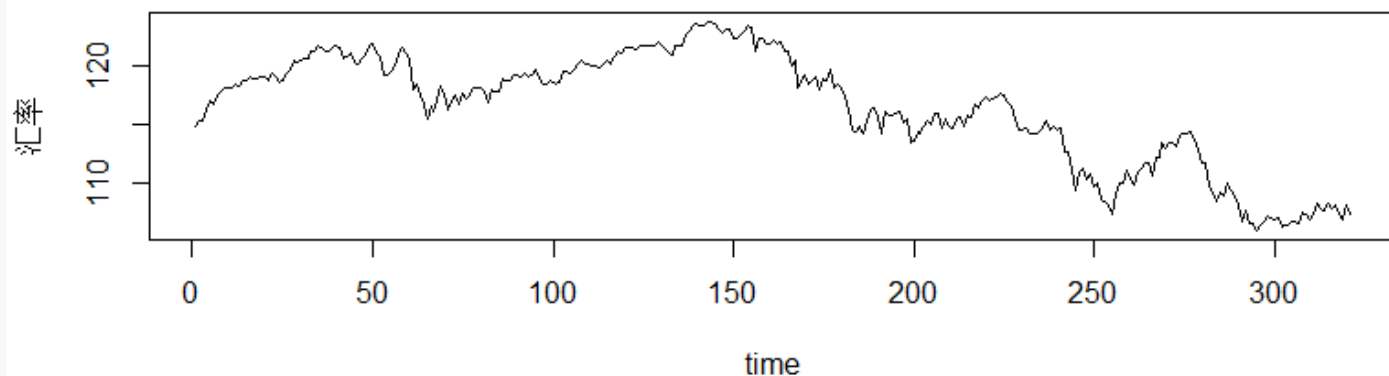
时间序列的例子

上证综合指数2011/1/4 1分钟高频数据

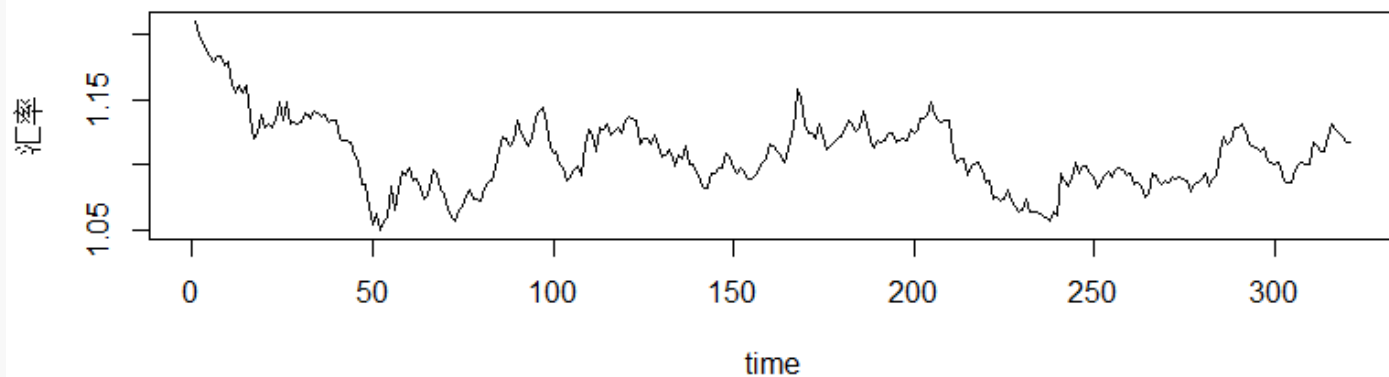


时间序列的例子

2015/1/1-2016/3/24 美元兑日元汇率

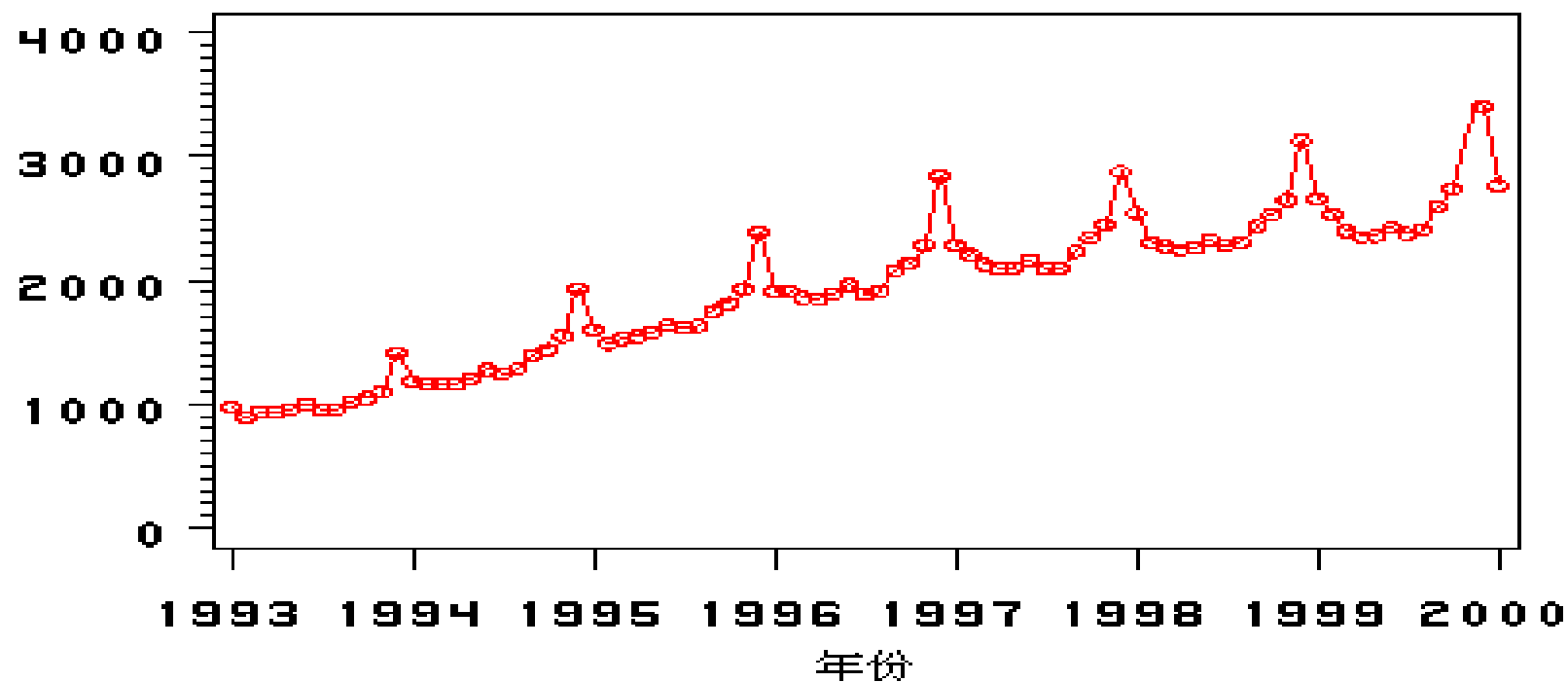


2015/1/1-2016/3/24 美元兑欧元汇率



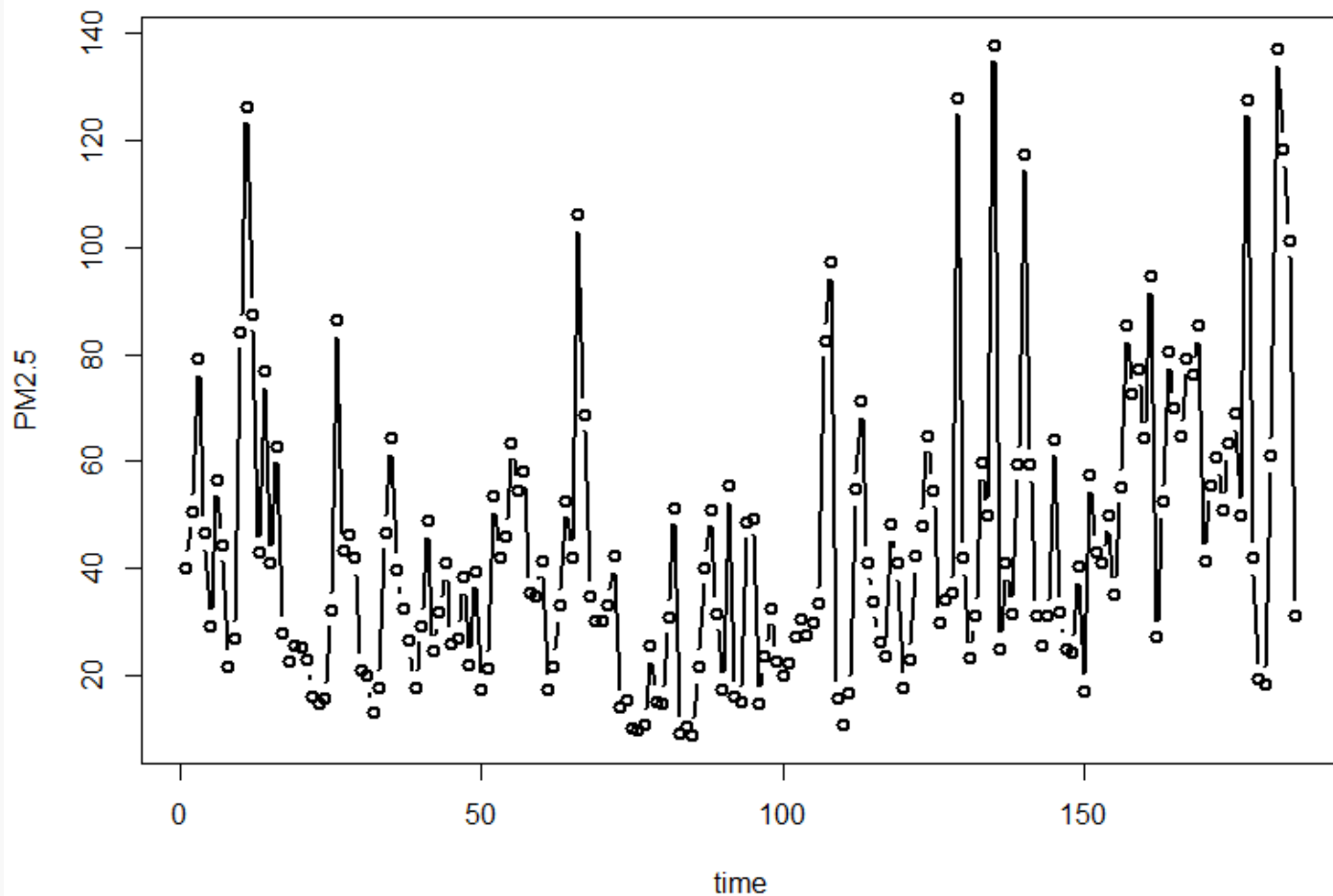
时间序列的例子

社会消费品零售总额（亿）

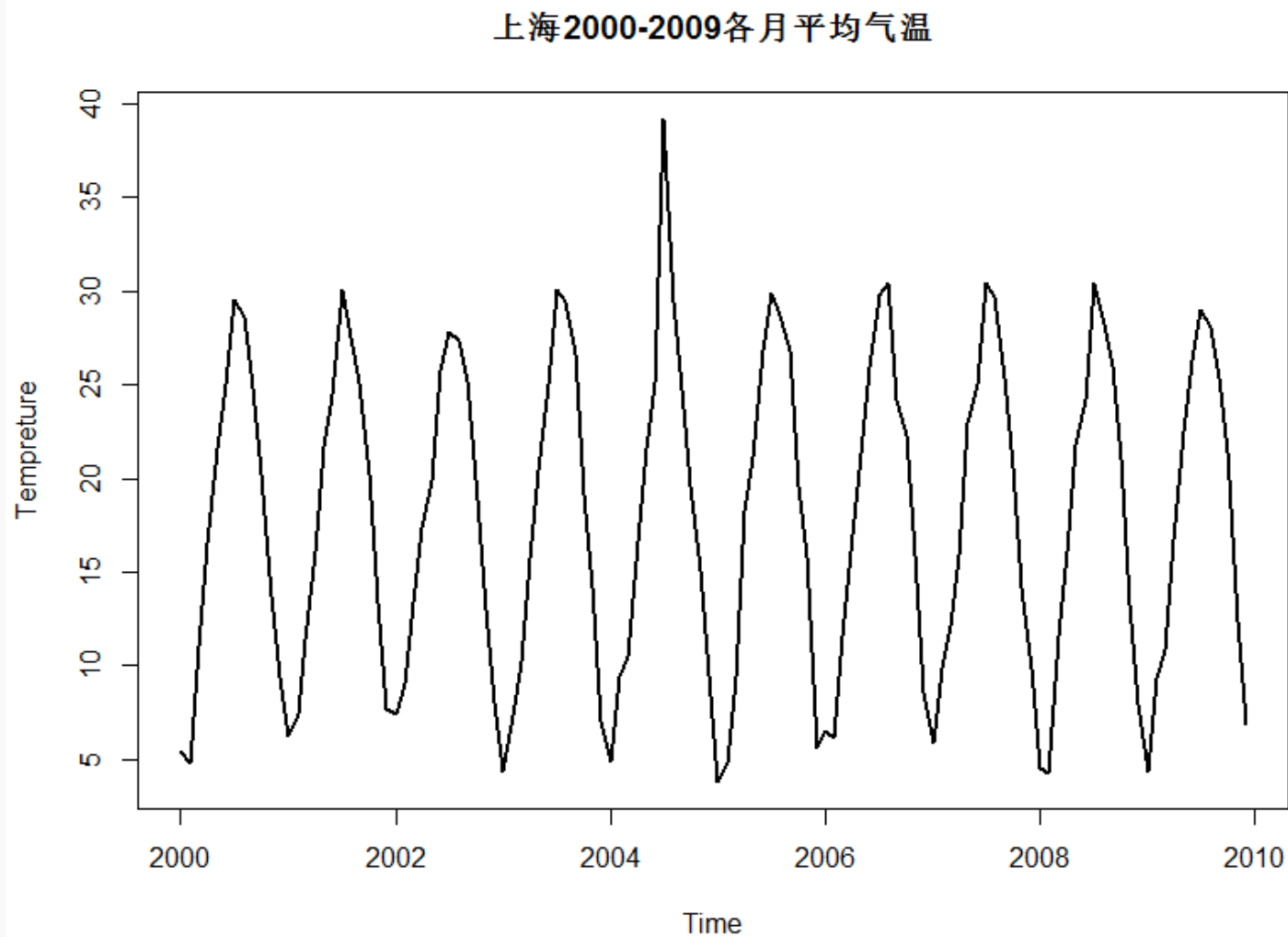


时间序列的例子

2014/7/1---2014/12/31 上海浦东PM2.5



时间序列的例子



课程安排

- 时间序列的初步分解
- 时间序列的平稳性、ARMA模型的性质
- 非平稳时间序列ARIMA模型
- 模型识别、参数估计、模型诊断与优化
- 条件数学期望与时间序列的预测
- 带季节的ARIMA模型
- 时间序列回归模型
- 异方差模型：ARCH模型、GARCH模型

下面我们介绍时间序列的初步分解技巧

时间序列的分解

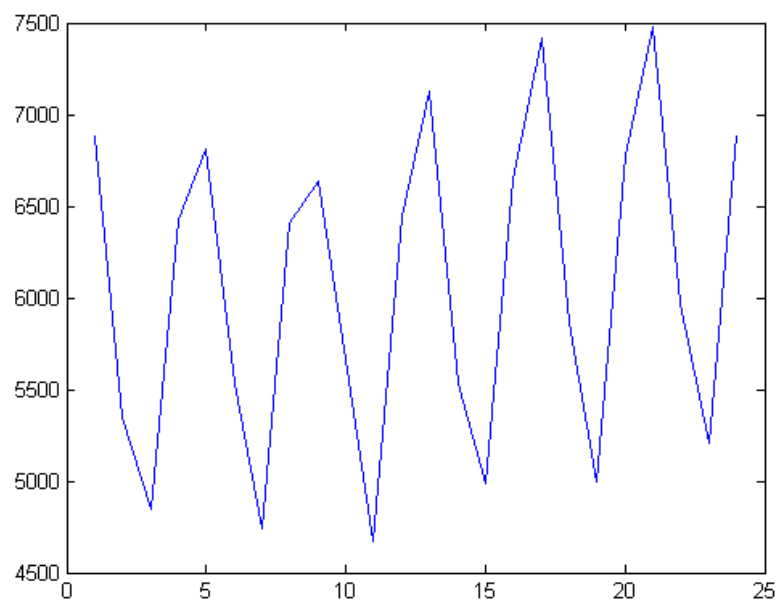
- 趋势项 (Trend) 表示时间序列的某种长期的演化趋势;
- 季节项 (Seasonal) 表示时间序列中随着四季、月、周等的周期变化规律;
- 不规则项 (Irregular) 表示时间序列中的不规则变化, 通常是由随机因素造成的, 它是在原时间序列中将趋势项、季节项去掉后的误差项。
- 加法模型: $Y_t = T_t + S_t + I_t$

分解的步骤（加法模型）

- 趋势项估计 $\{\hat{T}_t\}$
 - 分段趋势
 - 趋势拟合
 - 移动平均
- 去掉趋势项后,所得数据 $\{y_t - \hat{T}_t\}$
 - 由季节项和不规则项组成
 - 季节项估计 $\{\hat{S}_t\}$ 为该数据的季节平均
- 不规则项估计即为 $\{y_t - \hat{T}_t - \hat{S}_t\}$

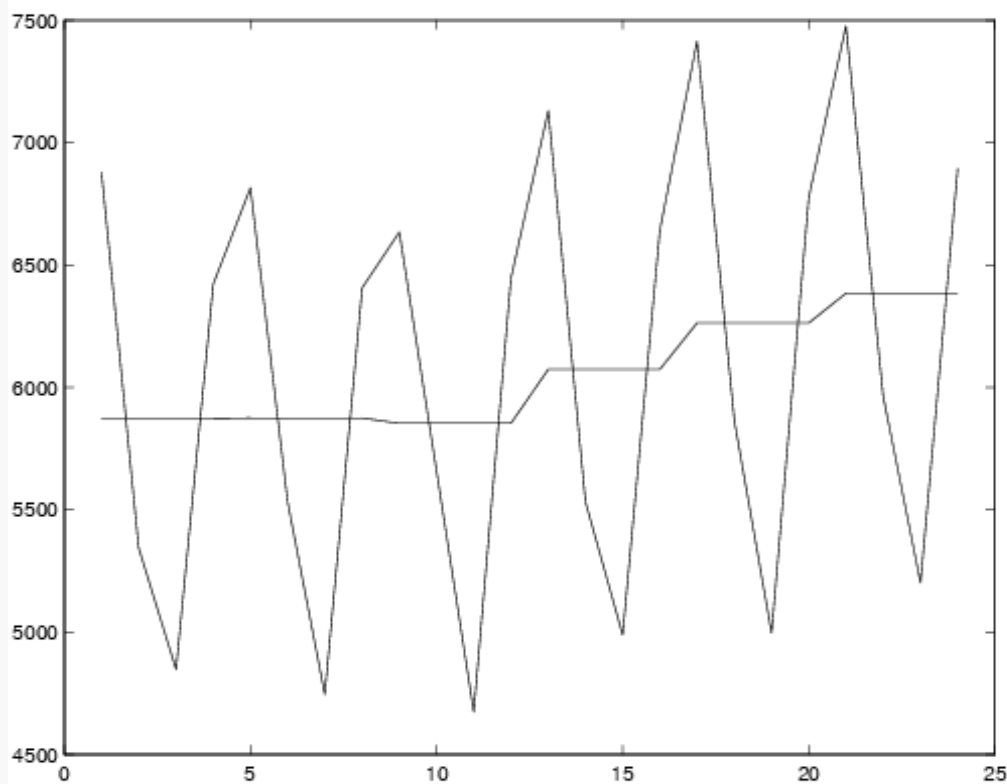
例：某城市居民季度用煤消耗量

年	1季度	2 季度	3 季度	4 季度	年平均
1991	6878.4	5343.7	4847.9	6421.9	5873.0
1992	6815.4	5532.6	4745.6	6406.2	5875.0
1993	6634.4	5658.5	4674.8	6445.5	5853.3
1994	7130.2	5532.6	4989.6	6642.3	6073.7
1995	7413.5	5863.1	4997.4	6776.1	6262.6
1996	7476.5	5965.5	5202.1	6894.1	6384.5
季平均	7058.1	5649.3	4909.6	6597.7	



分段趋势法

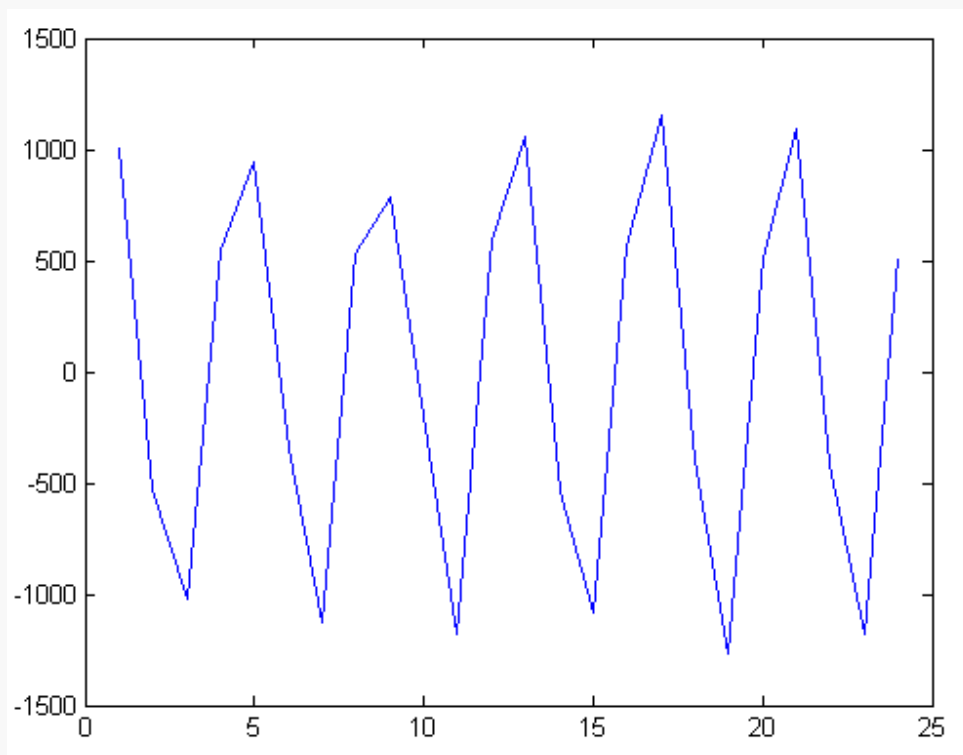
分段趋势图(年平均)



$$\begin{aligned}\hat{T}_1 &= \hat{T}_2 = \hat{T}_3 = \hat{T}_4 = 5873.0 \\ \hat{T}_5 &= \hat{T}_6 = \hat{T}_7 = \hat{T}_8 = 5875.0 \\ \hat{T}_9 &= \hat{T}_{10} = \hat{T}_{11} = \hat{T}_{12} = 5853.0 \\ \hat{T}_{13} &= \hat{T}_{14} = \hat{T}_{15} = \hat{T}_{16} = 6073.7 \\ \hat{T}_{17} &= \hat{T}_{18} = \hat{T}_{19} = \hat{T}_{20} = 6262.6 \\ \hat{T}_{21} &= \hat{T}_{22} = \hat{T}_{23} = \hat{T}_{24} = 6384.5\end{aligned}$$

去掉趋势项后的时间序列图

$$SI_t = y_t - \hat{T}_t$$



季节项估计

季节项估计为去掉趋势项之后的季节平均

如 $\hat{S}_1 = (SI_1 + SI_5 + SI_9 + SI_{13} + SI_{17} + SI_{21})/6$

$$\begin{aligned}\hat{S}_1 &= \hat{S}_5 = \hat{S}_9 = \hat{S}_{13} = \hat{S}_{17} = \hat{S}_{21} = 1004.4 \\ \hat{S}_2 &= \hat{S}_6 = \hat{S}_{10} = \hat{S}_{14} = \hat{S}_{18} = \hat{S}_{22} = -404.3 \\ \hat{S}_3 &= \hat{S}_7 = \hat{S}_{11} = \hat{S}_{15} = \hat{S}_{19} = \hat{S}_{23} = -1144.1 \\ \hat{S}_4 &= \hat{S}_8 = \hat{S}_{12} = \hat{S}_{16} = \hat{S}_{20} = \hat{S}_{24} = 544.0\end{aligned}$$

最终不规则项估计为

$$\hat{I}_t = y_t - \hat{T}_t - \hat{S}_t, t = 1, 2, \dots, 24.$$

趋势拟合法

- 有些时间序列具有非常显著的趋势，我们分析的目的就是要找到序列中的这种趋势，并利用这种趋势对序列的发展作出合理的预测。
- 趋势拟合法就是把时间作为自变量，相应的序列观察值作为因变量，建立序列值随时间变化的回归模型的方法。
 - 线性拟合
 - 非线性拟合

线性拟合

- 例：无季节项简单线性趋势模型

$$\begin{cases} Y_t = a + bt + I_t, t = 1, 2, \dots \\ E(I_t) = 0, \text{Var}(I_t) = \sigma^2 \end{cases}$$

- 参数估计方法：最小二乘估计

非线性拟合

■ 使用场合

- 长期趋势呈现出非线性特征

■ 参数估计

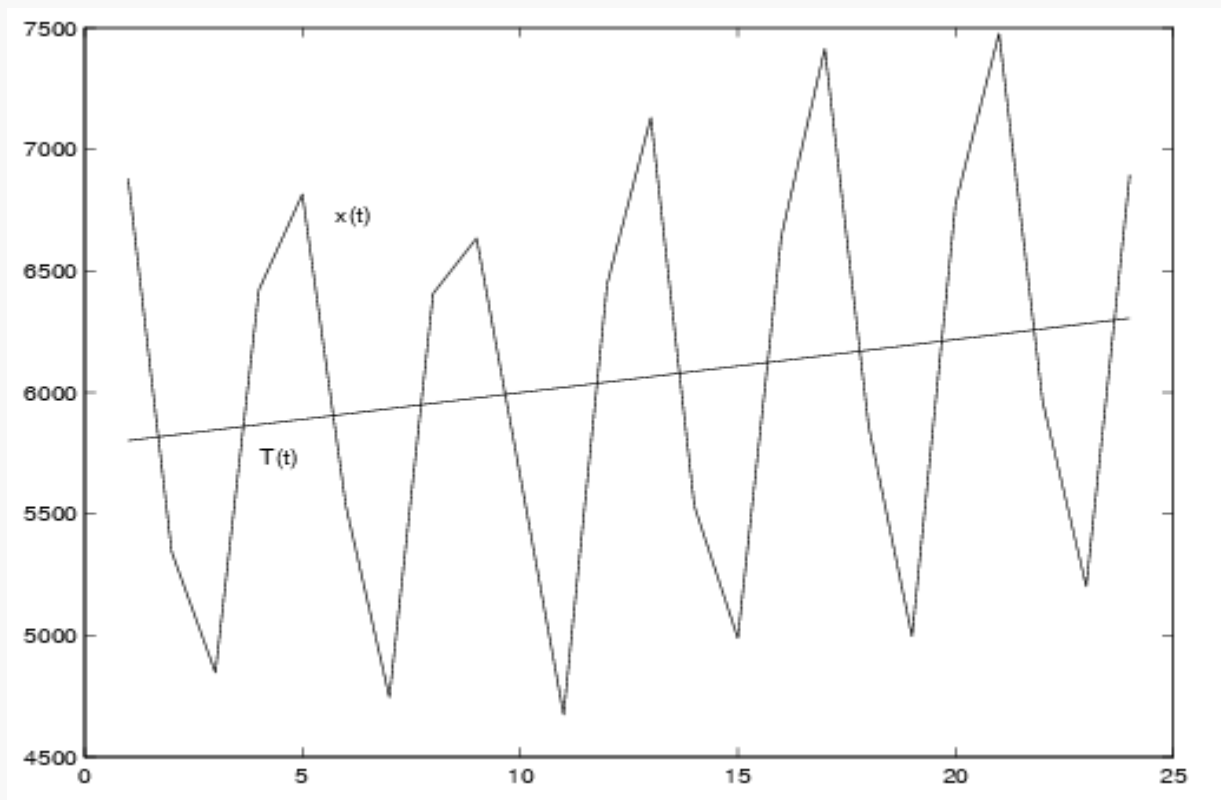
- 能转换成线性模型的都转换成线性模型，用最小二乘法进行参数估计
- 不能转换成线性的，就用迭代法进行参数估计

常用非线性模型

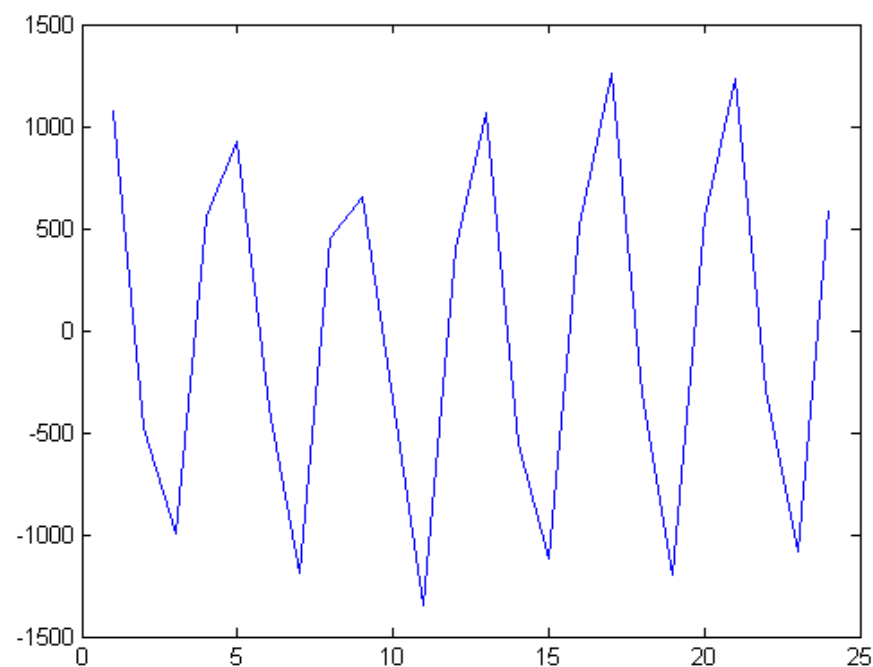
模型	变换	变换后模型	参数估计方法
$T_t = a + bt + ct^2$	$t_2 = t^2$	$T_t = a + bt + ct_2$	线性最小二乘估计
$T_t = ab^t$	$T_t' = \ln T_t$ $a' = \ln a$ $b' = \ln b$	$T_t' = a' + b't$	
$T_t = a + bc^t,$ $e^{a+bc^t}, \frac{1}{a+bc^t},$	—	—	迭代法

趋势项的估计—回归直线法

■ 直线趋势项

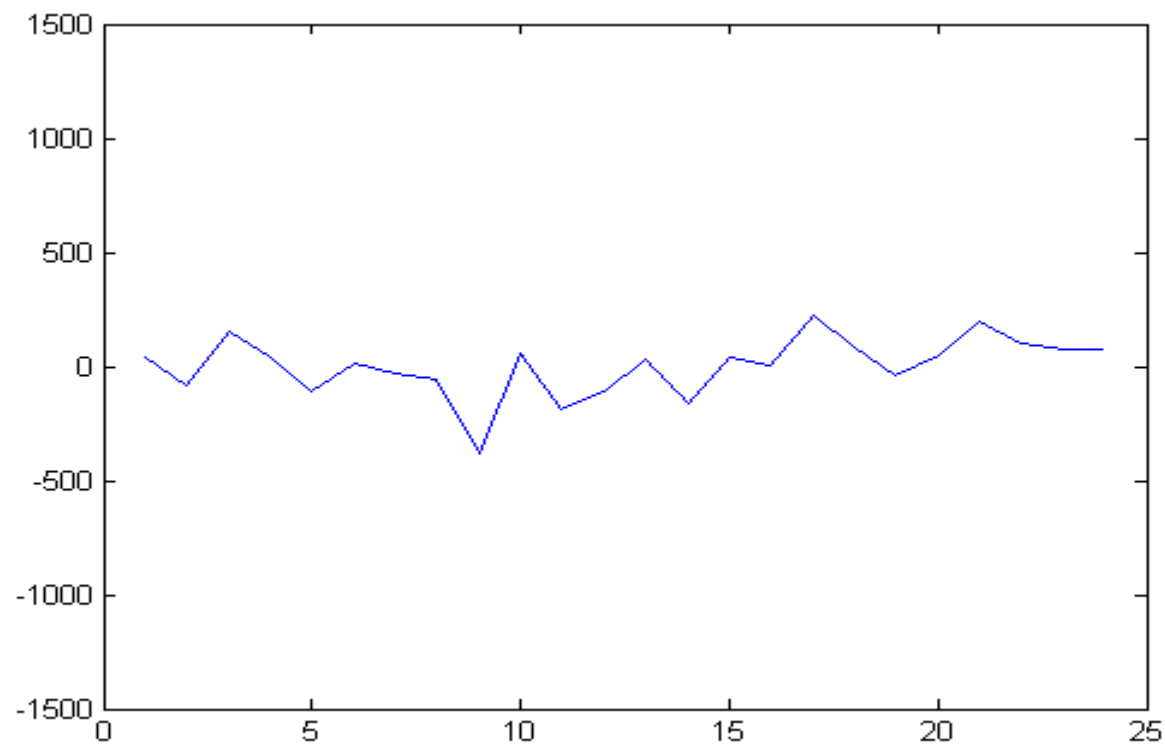


去掉直线趋势项后的时间序列图



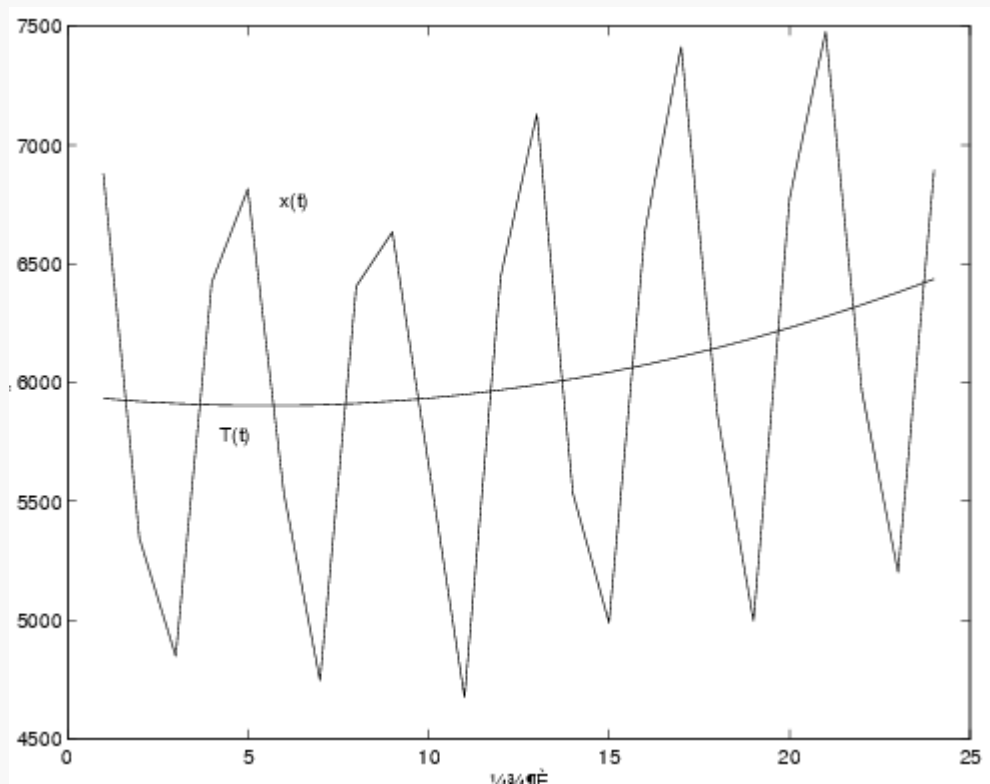
不规则项

$$\hat{I}_t = y_t - \hat{T}_t - \hat{S}_t, t = 1, 2, \dots, 24.$$



趋势项的估计—二次曲线法

■ 二次趋势项估计



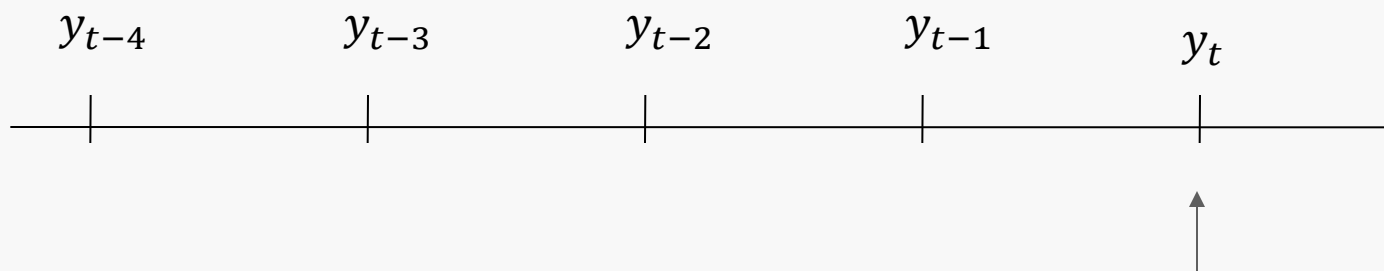
移动平均法

- 作用：削弱短期随机波动对序列的影响，使序列平滑化，显示出趋势变化的规律，有时可消除周期效应。
- 分类：
 - n 期移动平均
 - n 期中心移动平均

n期移动平均

$$\tilde{y}_t = \frac{1}{n} (y_t + y_{t-1} + \cdots + y_{t-n+1})$$

5
期
移
动
平
均

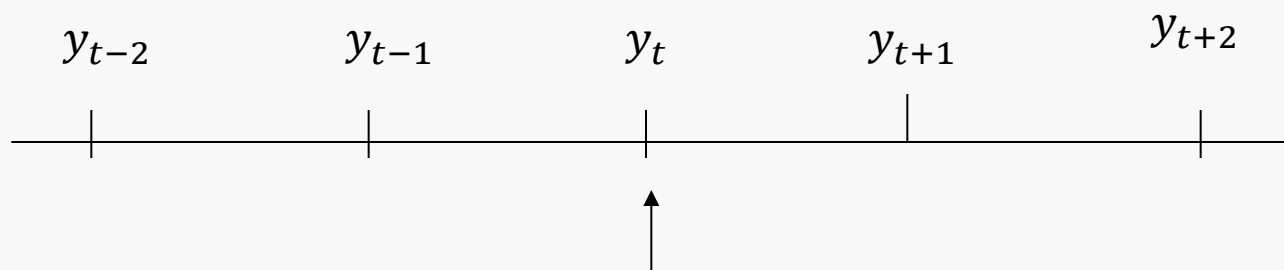


$$\tilde{y}_t = \frac{y_{t-4} + y_{t-3} + y_{t-2} + y_{t-1} + y_t}{5}$$

n期中心移动平均

$$\tilde{y}_t = \begin{cases} \frac{1}{n} (y_{t-\frac{n-1}{2}} + y_{t-\frac{n-1}{2}+1} + \cdots + y_t + \cdots + y_{t+\frac{n-1}{2}-1} + y_{t+\frac{n-1}{2}}), & n \text{ 为奇数} \\ \frac{1}{n} (\frac{1}{2} y_{t-\frac{n}{2}} + y_{t-\frac{n}{2}+1} + \cdots + y_t + \cdots + y_{t+\frac{n}{2}-1} + \frac{1}{2} y_{t+\frac{n}{2}}), & n \text{ 为偶数} \end{cases}$$

5
期
中
心
移
动
平
均



The diagram shows a horizontal line with five tick marks. Above the first tick mark is the label y_{t-2} , above the second is y_{t-1} , above the third is y_t , above the fourth is y_{t+1} , and above the fifth is y_{t+2} . An upward-pointing arrow is positioned below the third tick mark, pointing to the y_t term in the formula below.

$$\tilde{y}_t = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}}{5}$$

移动平均期数确定的原则

■ 有无周期性

- 以周期长度作为移动平均的间隔长度，以消除周期效应的影响（为什么？）
- 对比分段趋势？

■ 对趋势平滑的要求

- 移动平均的期数越多，拟合趋势越平滑（为什么？）

■ 对趋势反映近期变化敏感程度的要求

- 移动平均的期数越少，拟合趋势越敏感