



时间序列分析

联络方式:

办公室: 诚信楼506

电 话: 64494530

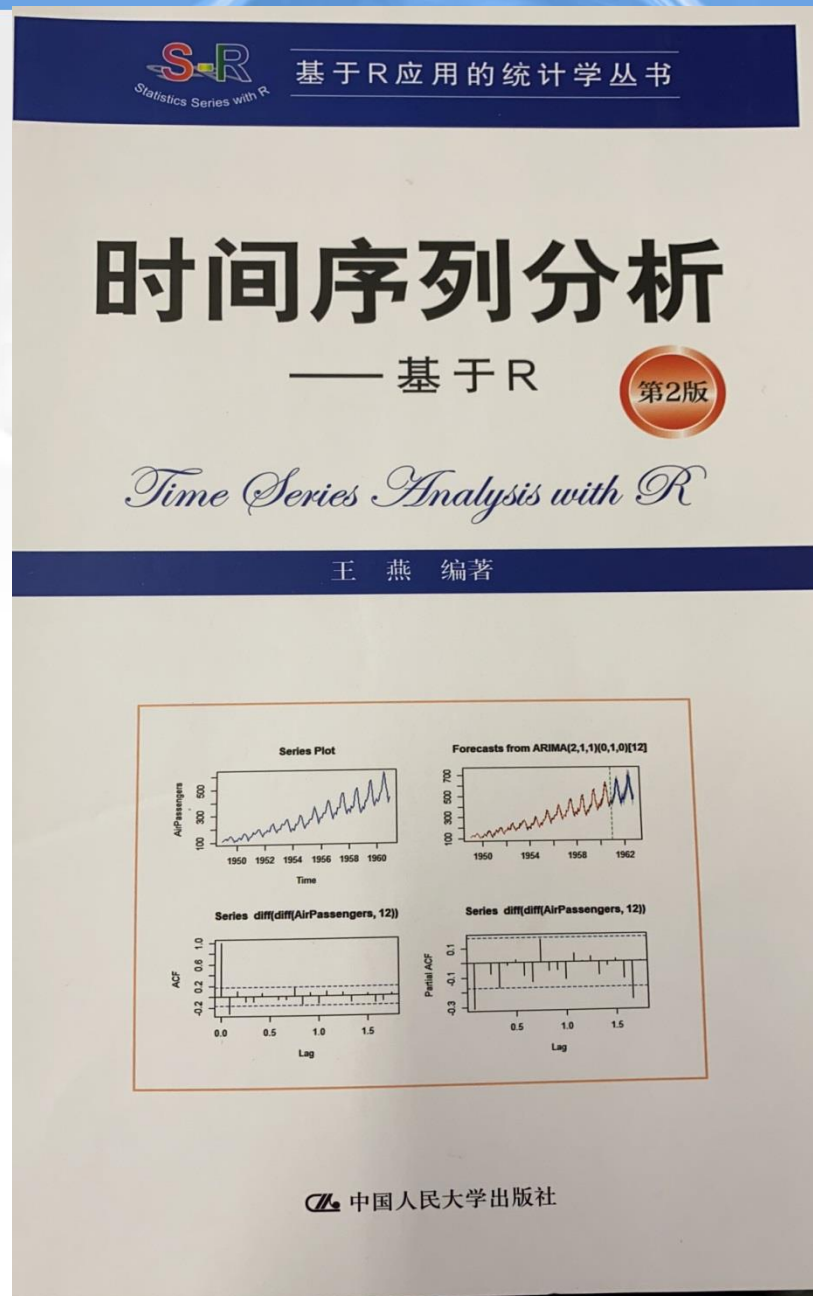
邮 箱: wangjingle@uibe.edu.cn

考核要求:

1. 期中测验, 闭卷, 随堂考。
2. 课上随机点名, 抽查迟到, 早退, 旷课情况, 病、事假请于课前告知教师。
3. 总评成绩包括: 平时、作业、期中、期末。

教材：《时间序列分析-基于R》。

参考书：《时间序列分析及应用：R语言》，Jonathan D. Cryer, Kung-sik Chan著，潘红宇译，机械工业出版社。
应用软件：R软件



第一章、时间序列分析简介

1. 引言

2. 时间序列的定义

3. 时间序列分析方法简介

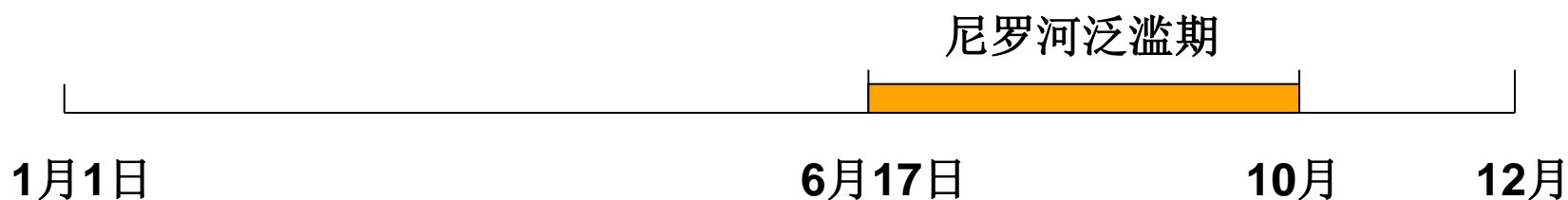
4. 时间序列分析软件

一、引言

按照时间的顺序把随机事件变化发展的过程记录下来就构成了一个时间序列。对时间序列进行观察、研究、找寻它变化发展的规律，预测它将来的走势就是时间序列分析。

7000年前，古埃及人把尼罗河涨落的情况逐天记录下来，就构成所谓的时间序列。

对这个时间序列长期的观察使他们发现尼罗河的涨落非常有规律。当天狼星第一次和太阳同时升起的那一天之后，再过200天左右，尼罗河就开始泛滥，泛滥期将持续70~80天，洪水过后，土地肥沃，随意播种就会有丰厚的收成。



由于掌握了尼罗河泛滥的规律，古埃及农业迅速发展，从而创建了埃及灿烂的史前文明。

二、时间序列定义

在统计研究中，有大量的数据是按照时间顺序排列的，使用数学方法表述即用一组随机序列

$$\cdots, X_1, X_2, \cdots, X_t, \cdots$$

表示随机事件的时间序列.

随机序列的有序观察值，记为：

$$\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$$

观察值序列是随机序列的一个实现。

与随机样本的区别：

在数理统计中，容量为n的随机样本可表示为：

$$X_1, X_2, \dots, X_n$$

而时间序列也可以写成如上形式。

上述都表示一组随机变量序列，有何不同？

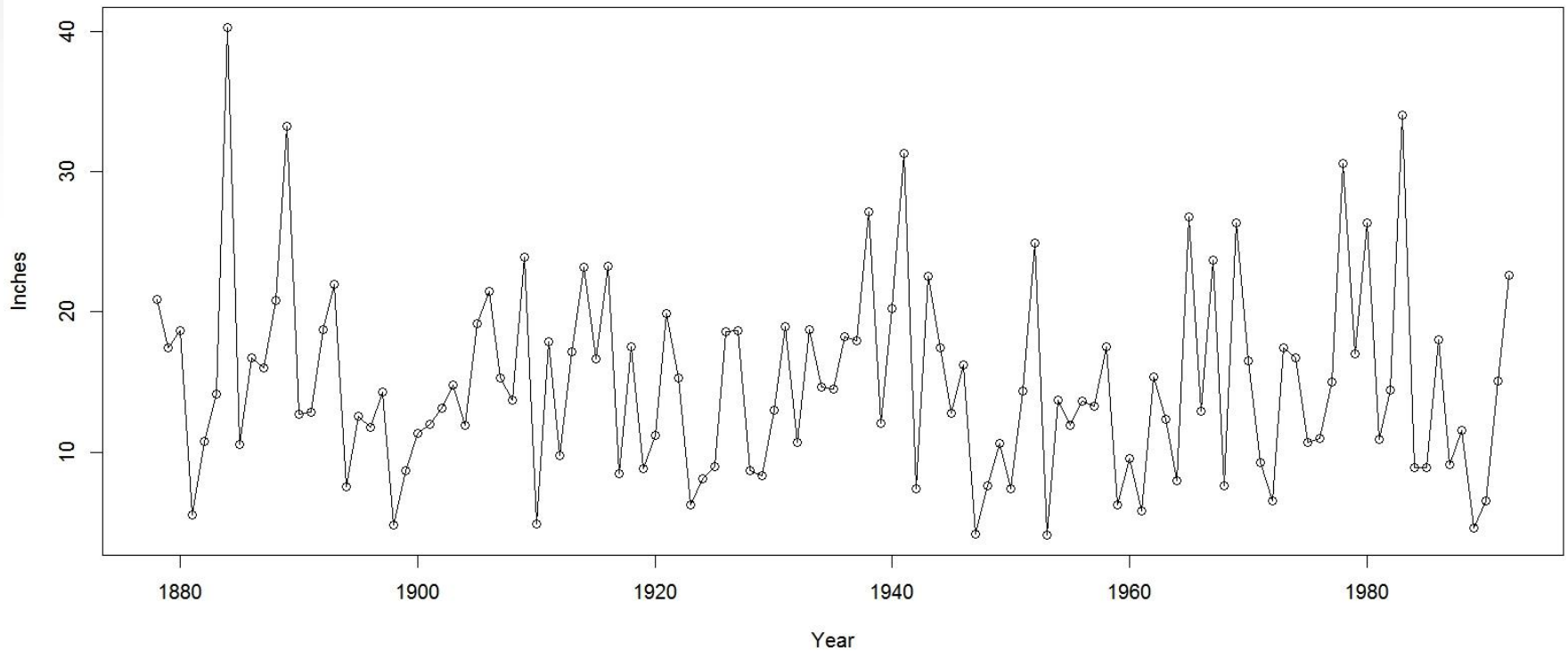
1、随机样本中各变量是同分布的，从而每个样本观察值可以看作来自同一个总体。而时间序列没有**同分布**这个要求。

时间序列每个变量只有一个样本观察值。

2、随机样本中各变量相互独立，从而样本是独立同分布的随机变量序列。而时间序列则研究**非独立性的结构**。

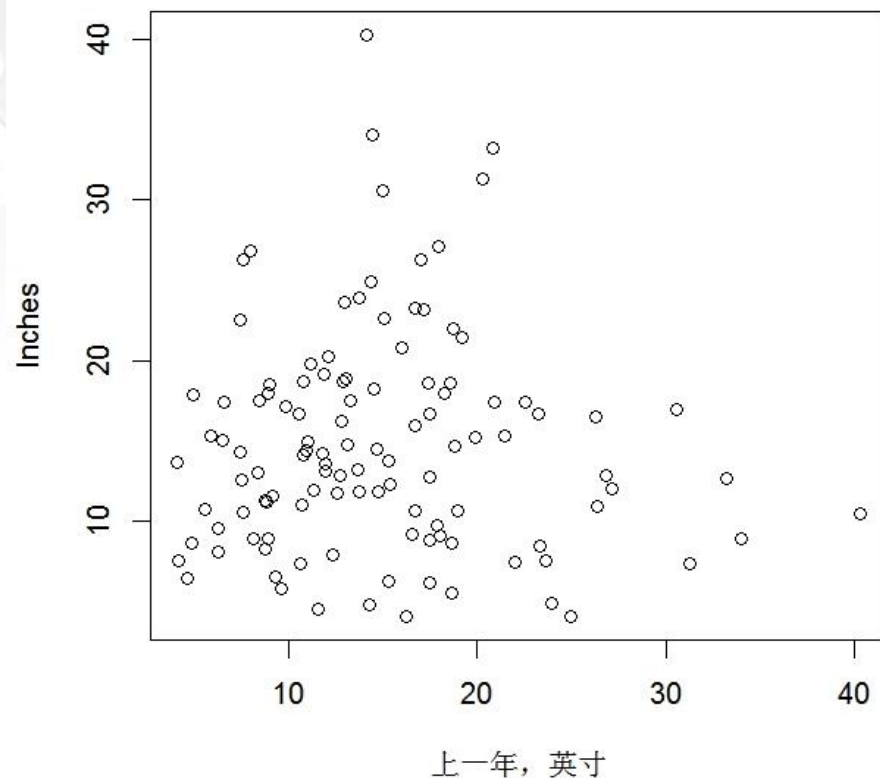
一些时间序列实例：

1、洛杉矶100多年的降水量时间序列图；

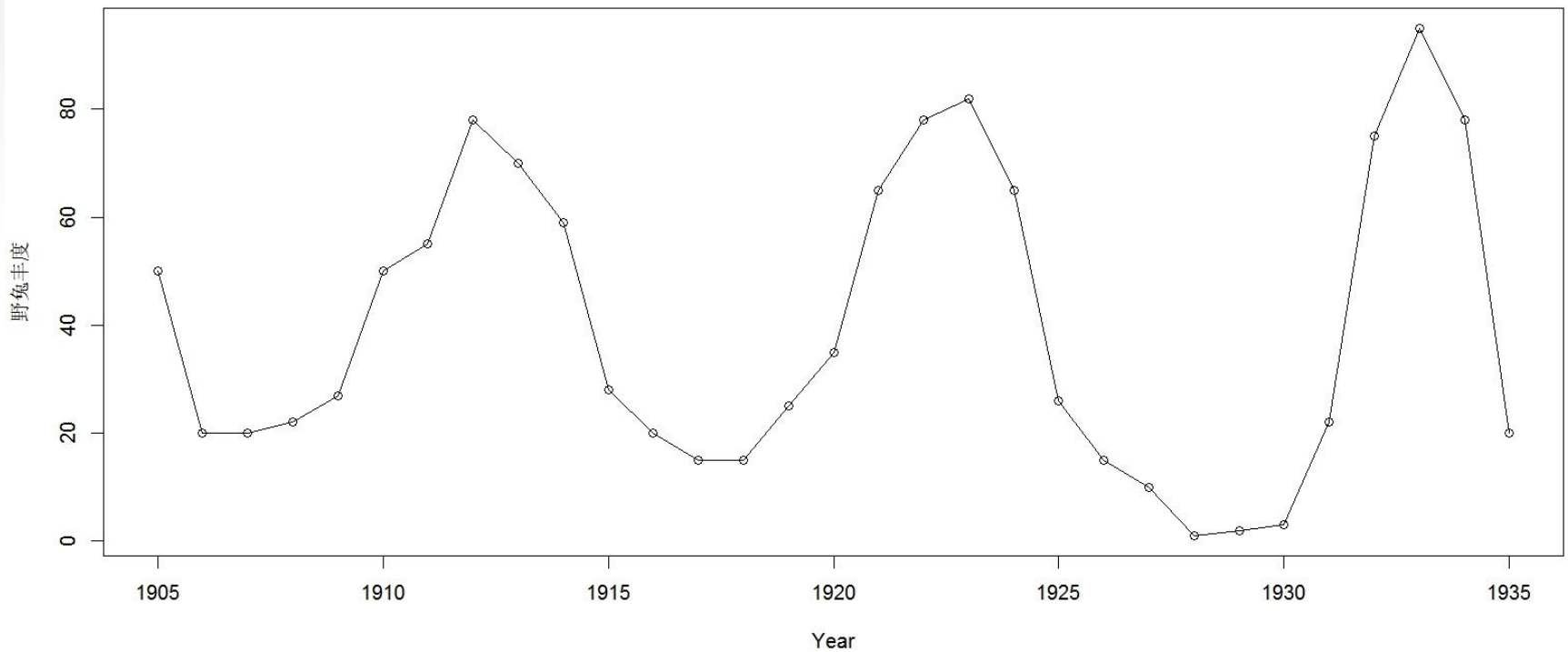


此图是当年降水量与去年降水量散点图。

从第二幅图看出，前一年的降水量与下一年并没有什么特殊关系。

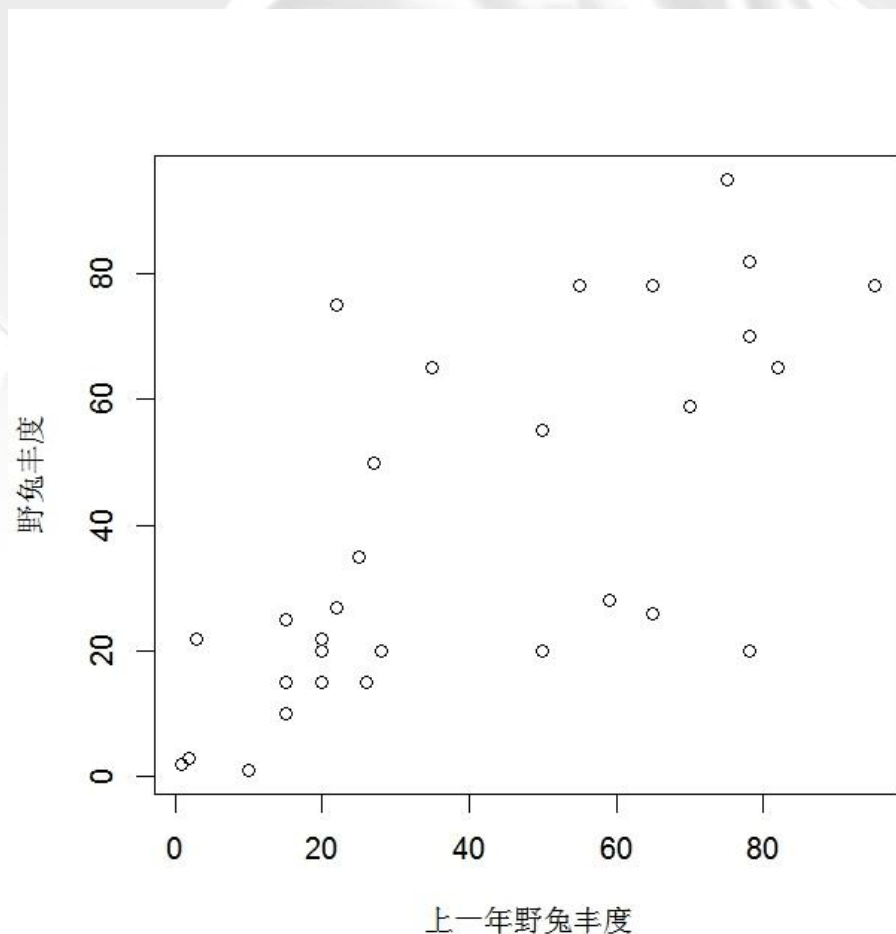


2、加拿大30年的野兔丰度时间序列图；

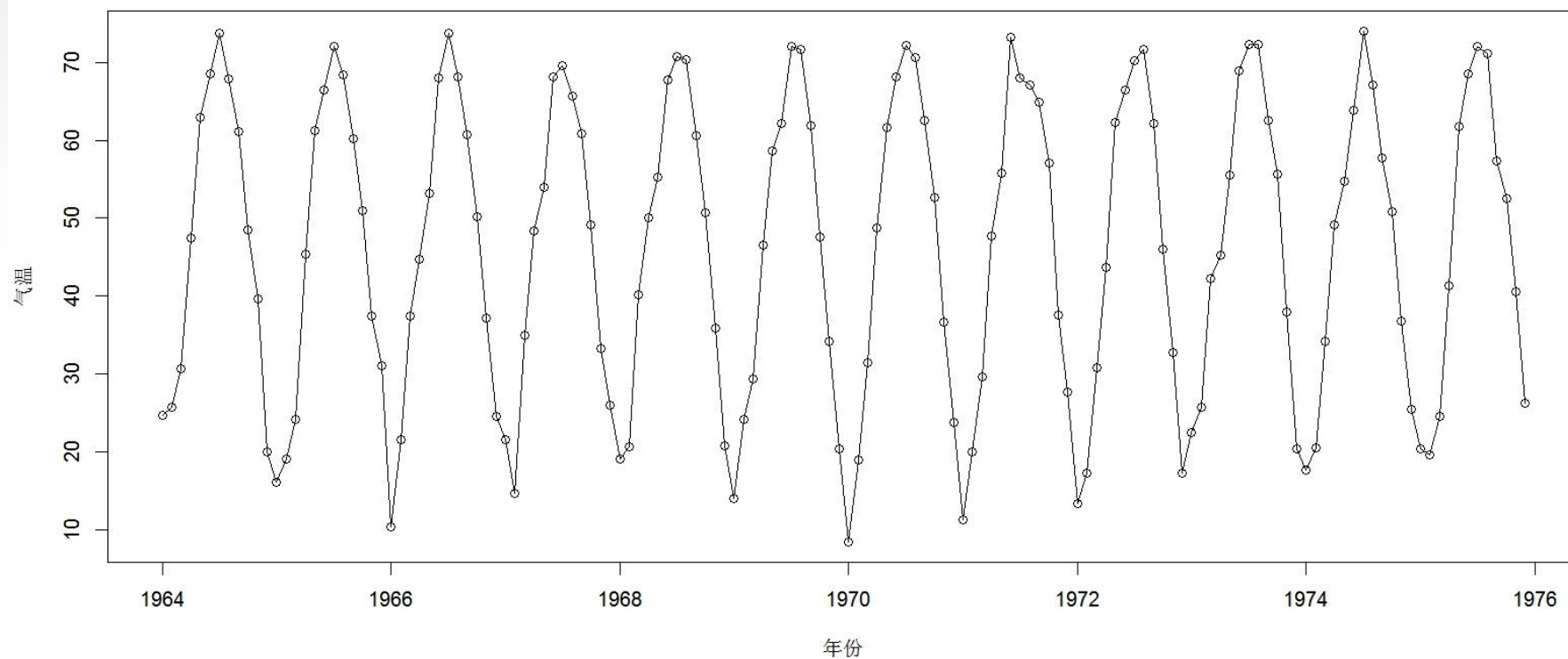


第二幅图是上一年与当年的丰度散点图。

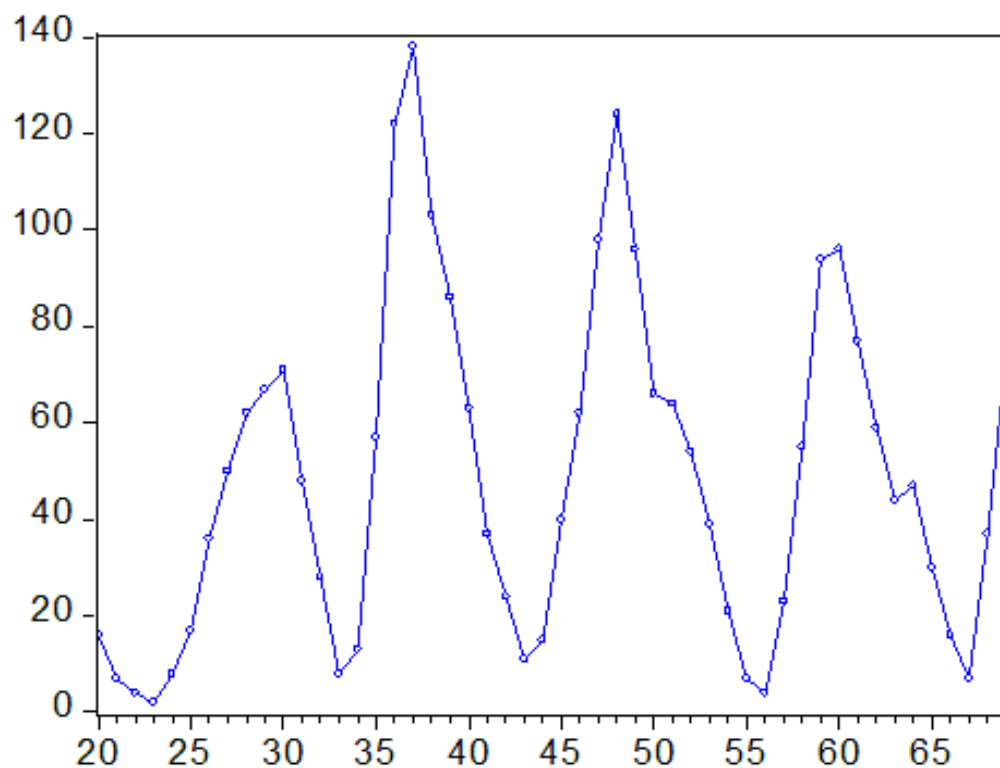
从第二幅图看出，前一年与下一年有一定的线性趋势。



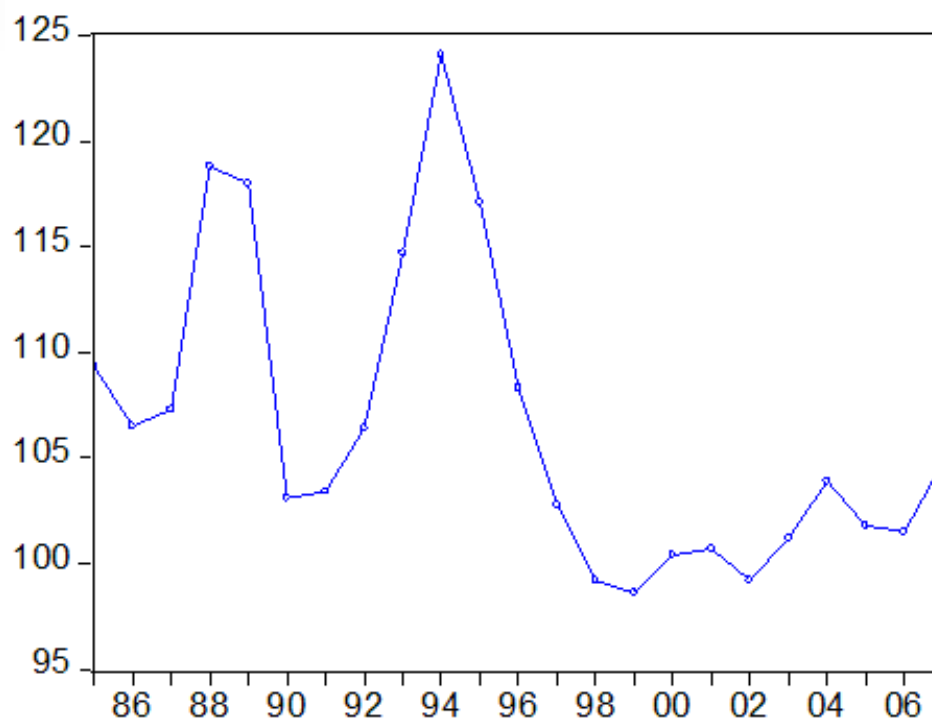
3、艾奥瓦州迪比克市月平均气温



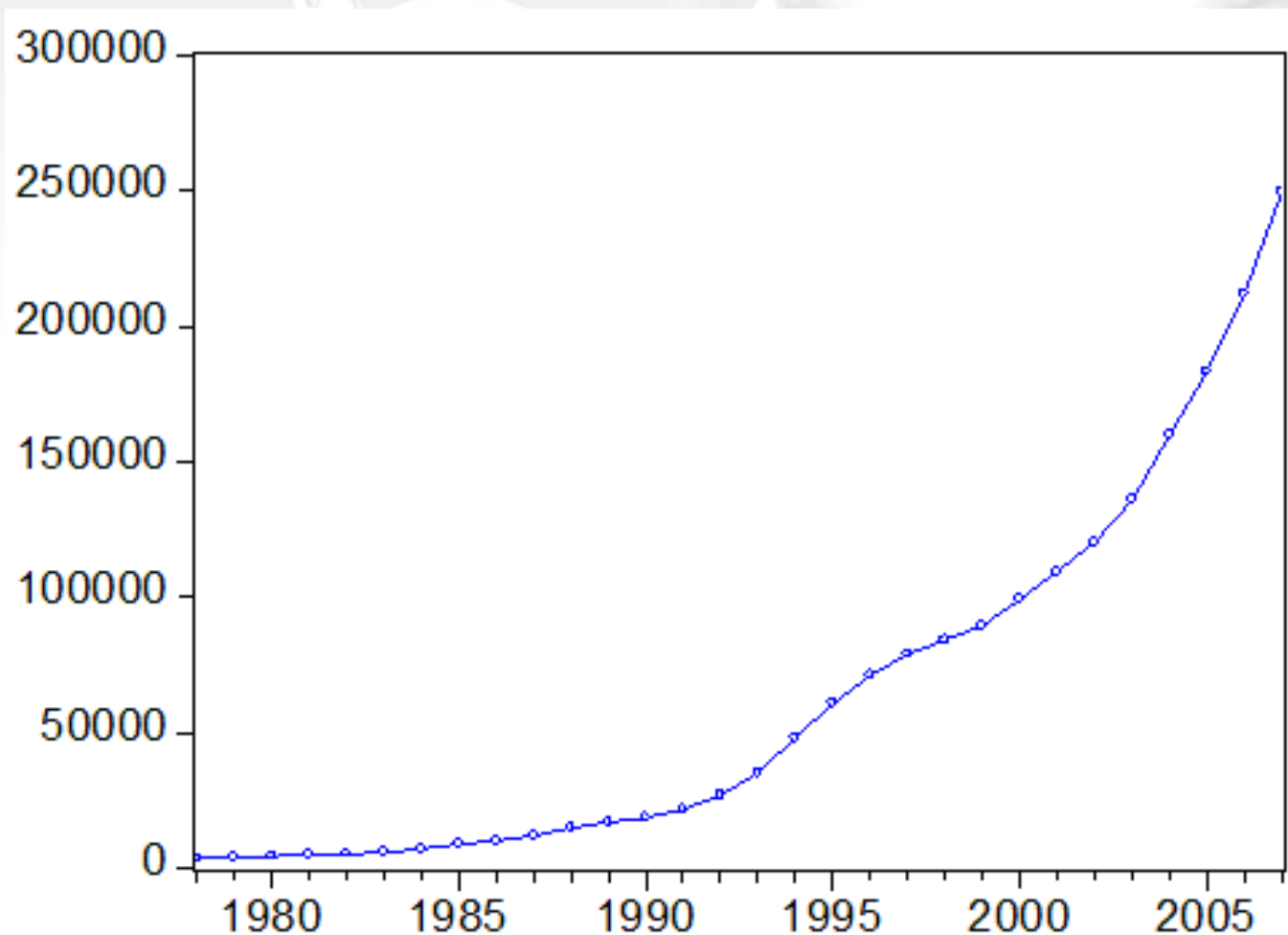
4、1820年—1869年的太阳黑子数时间序列图。该图中，横轴是时间指标 t （在这里的 t 以年为单位），纵轴表示在时间 t 内太阳黑子个数的观测值。



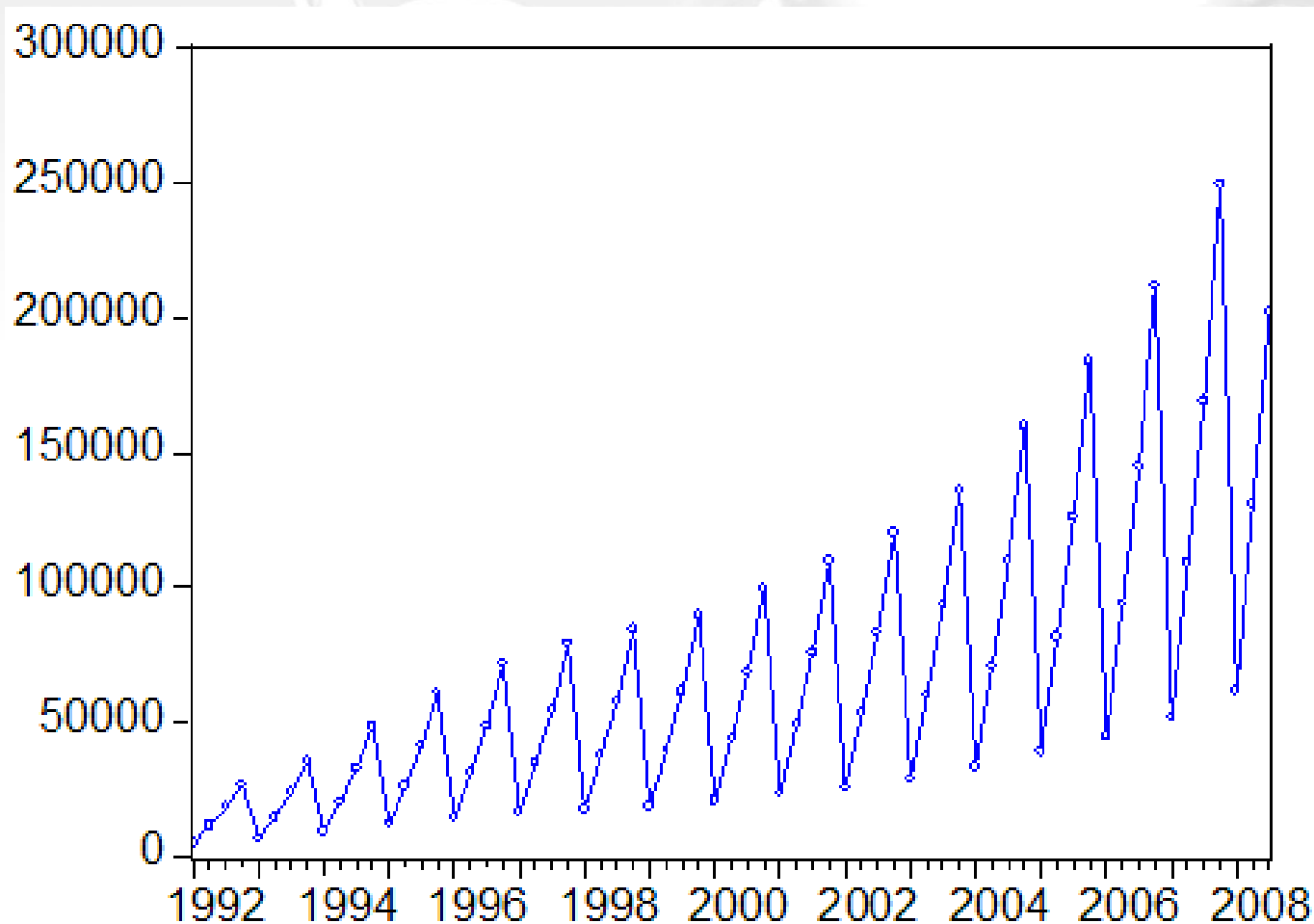
5、居民消费价格指数(Consumer Price Index), 英文缩写CPI, 是反映与居民生活有关的产品及劳务价格统计出来的物价变动指标, 通常作为观察通货膨胀水平的重要指标。本例给出了我国1985年—2007年的CPI年度数据时间序列图。



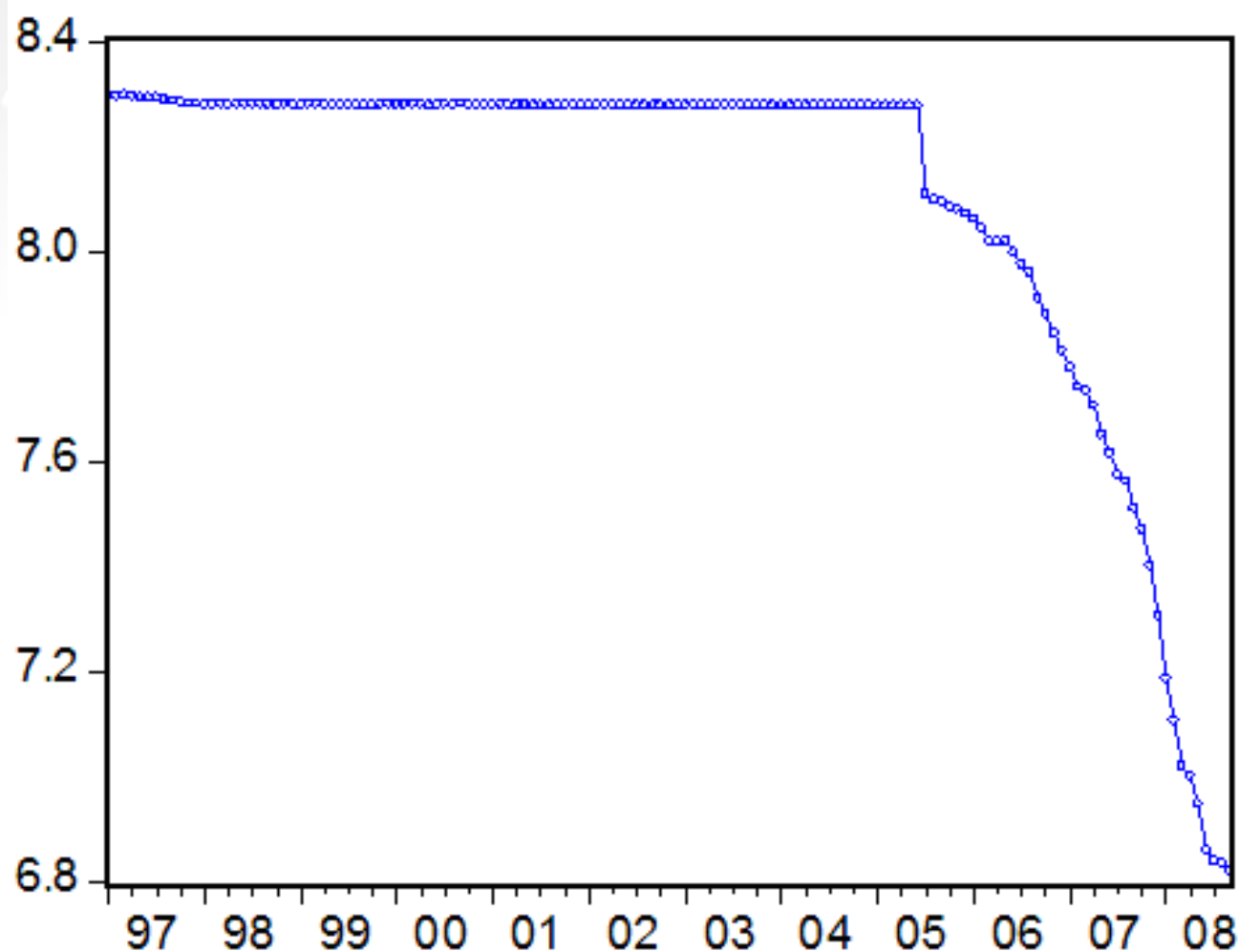
6、本例给出我国1978年—2007年GDP数据（单位：亿元）的时间序列图。



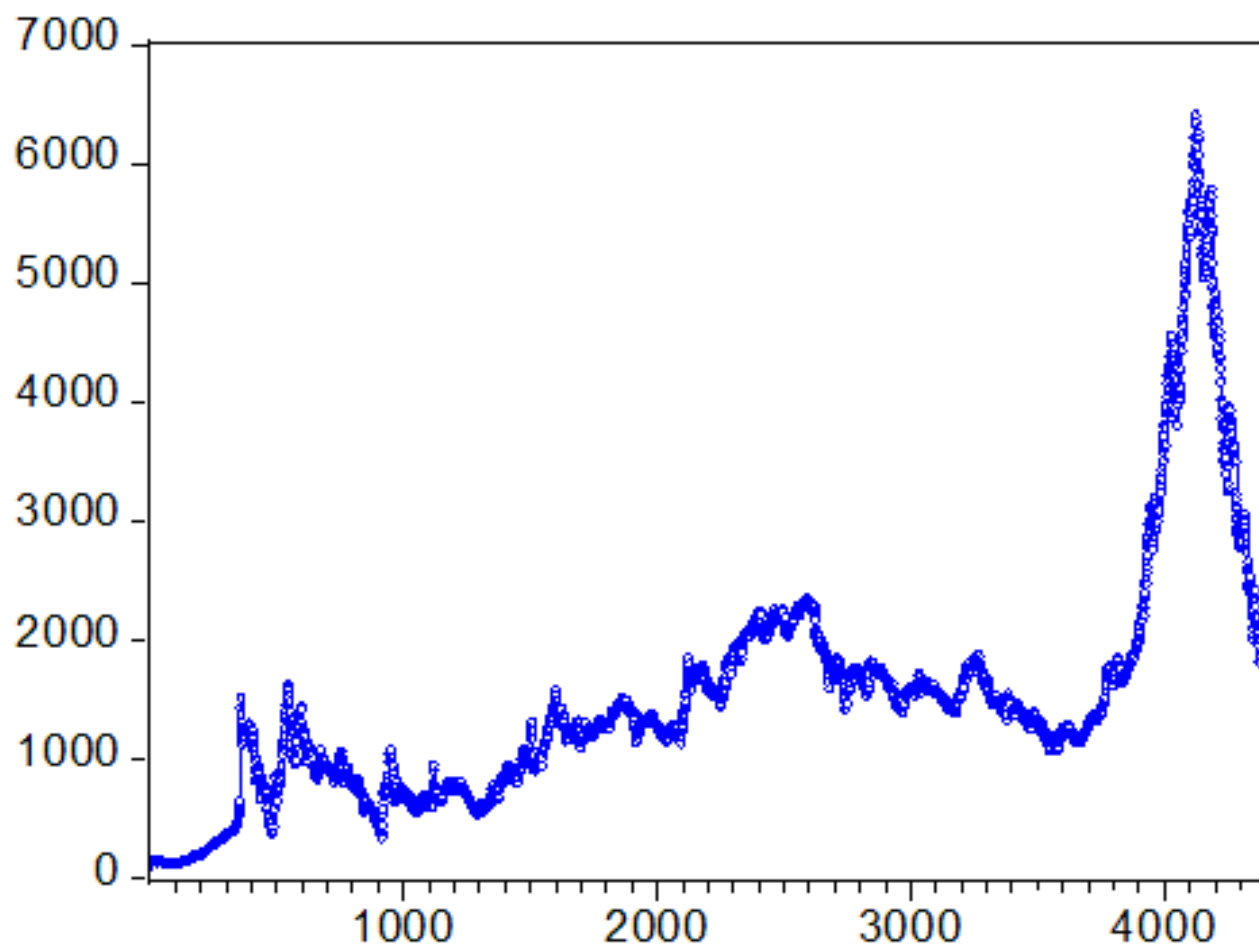
7、本例给出了1992年第一季度至2008年第三季度我国GDP季度数据（单位：亿元）。



8、本例给出了1997年1月—2008年9月美元对人民币汇率的月度数据（单位：元）时序图。



9、本例给出了1990年12月19日—2008年11月6日上证A股指数日数据（除去节假日，共4386个数据）时序图。





时间序列分析依赖于不同地应用背景，有着不同的目的。

分析的基本任务是揭示支配观测到的时间序列的随机规律，通过所了解的这个随机规律，我们可以理解所要考虑的动态系统，预报未来的事件，并且通过干预来控制将来事件。上述即为时间序列分析的三个目的。

三、时间序列分析方法

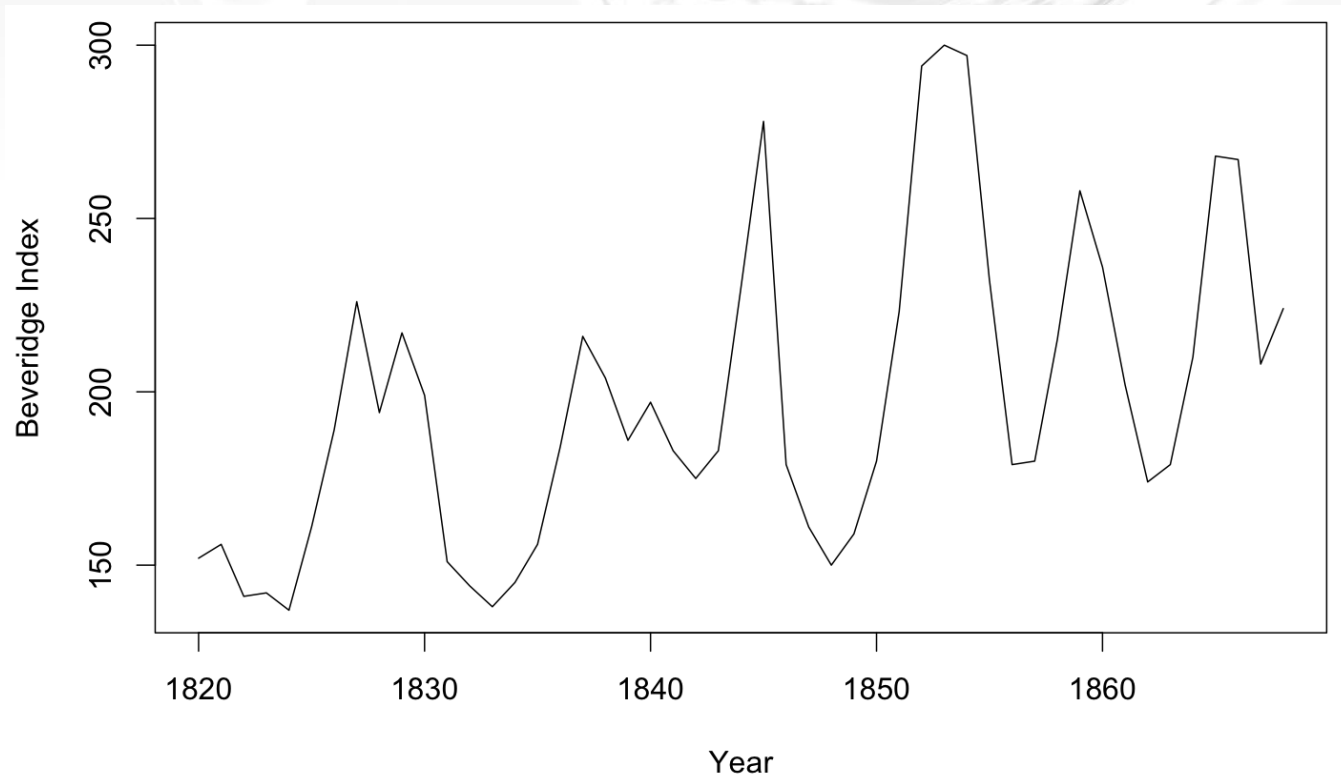


描述性时间序列分析

- 早期的时序分析通常都是通过直观的数据比较或绘图观测，寻找序列中蕴含的发展规律，这种分析方法就称为描述性时序分析。
- 古埃及人发现尼罗河泛滥的规律就是依靠这种分析方法。
- 在天文、物理、海洋学等自然科学领域，这种简单的描述性时序分析方法也常常能使人们发现意想不到的规律。

案例、Beveridge小麦价格指数序列

贝弗里奇（Beveridge）小麦价格指数序列，它由1500-1869年逐年估计的小麦价格构成，可以清晰地看到该序列有一个13年左右的周期



- 德国业余天文学家施瓦贝（S.H.Schwabe）发现太阳黑子的活动具有11-12年左右的周期



粮食价格波动的天文学解释：

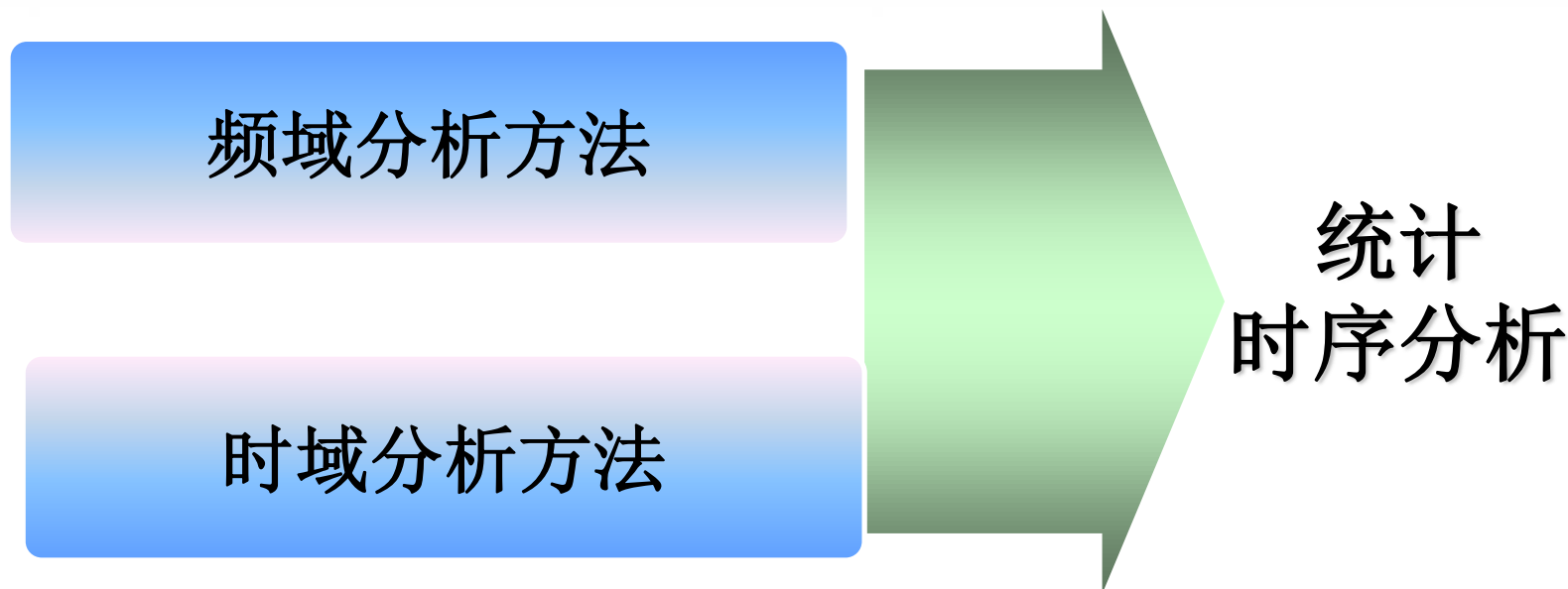
- 太阳黑子的运动周期和农业生产的周期长度非常接近，这引起了英国天文学家、天王星的发现者威廉·赫歇尔（F. W. Herschel）的关注。最后他发现当太阳黑子变少时，地球上的雨量也会减少。所以在没有良好人工灌溉技术的时代，农业生产会呈现出和太阳黑子近似的变化周期。
- 我们没有采用任何复杂的模型或分析方法，仅仅是沿着时间顺序收集数据，描述和呈现序列的波动，就了解到小麦产量的周期波动特征，以及产生该周期波动的气候成因及该周期对价格的影响。

描述性时序分析

- 通过直观的数据比较或绘图观测，寻找序列中蕴含的发展规律，这种分析方法就称为描述性时序分析
- 描述性时序分析方法具有操作简单、直观有效的特点，它通常是人们进行统计时序分析的第一步。
- 局限性：它只能展示非常明显的规律性。
- 而在金融、保险、法律、人口、心理学等社会科学研究领域，随机变量的发展通常会呈现出非常强的随机性，想通过对序列简单的观察和描述，总结出随机变量发展变化的规律，并准确预测出它们将来的走势通常是非常困难的。

统计时序分析方法

19世纪末-20世纪初，是现代时间序列分析方法萌芽期。我们课本上学的主要方法，都是从这个时期开始产生的。人们运用数理统计原理来分析时间序列。这个时期产生了两种不同的时序分析方法。



频域分析方法

原理：假设任何一种无趋势的时间序列都可以分解成若干不同频率的周期波动

发展过程

- 早期借助傅里叶分析从频率的角度揭示时间序列的规律
- 后来借助了傅里叶变换，来逼近某个函数
- 20世纪60年代，引入最大熵谱估计理论，进入现代谱分析阶段

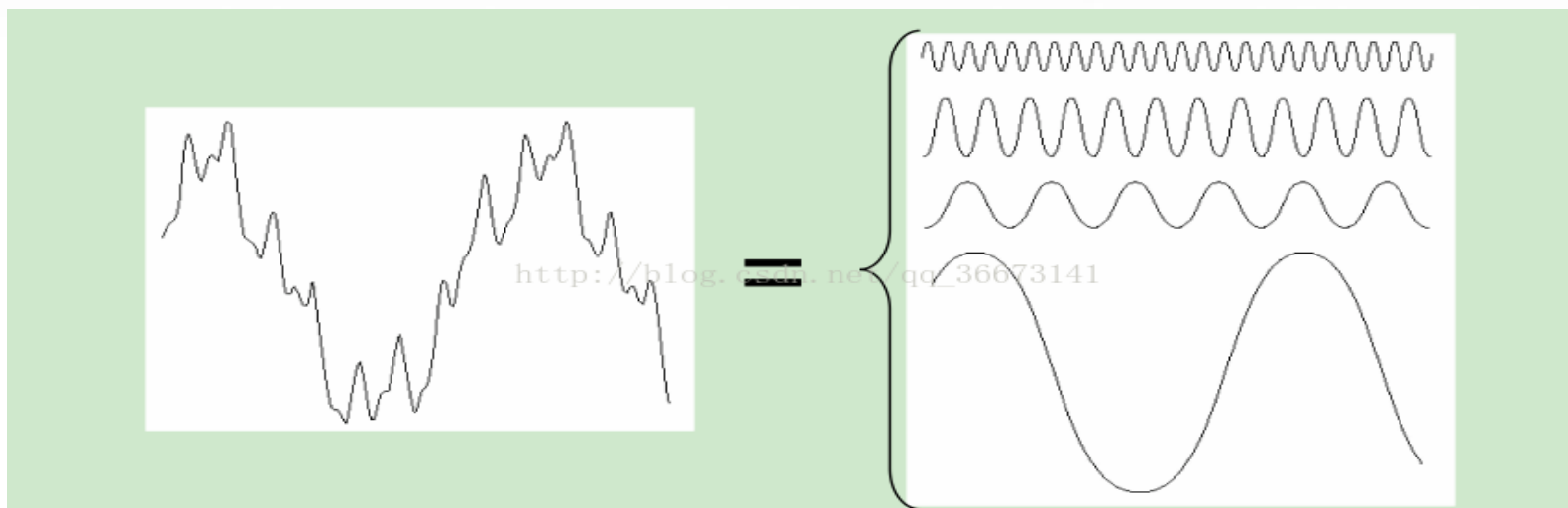
特点：非常有用的动态数据分析方法，但是由于分析方法复杂，结果抽象，有一定的使用局限性

频域分析举例：

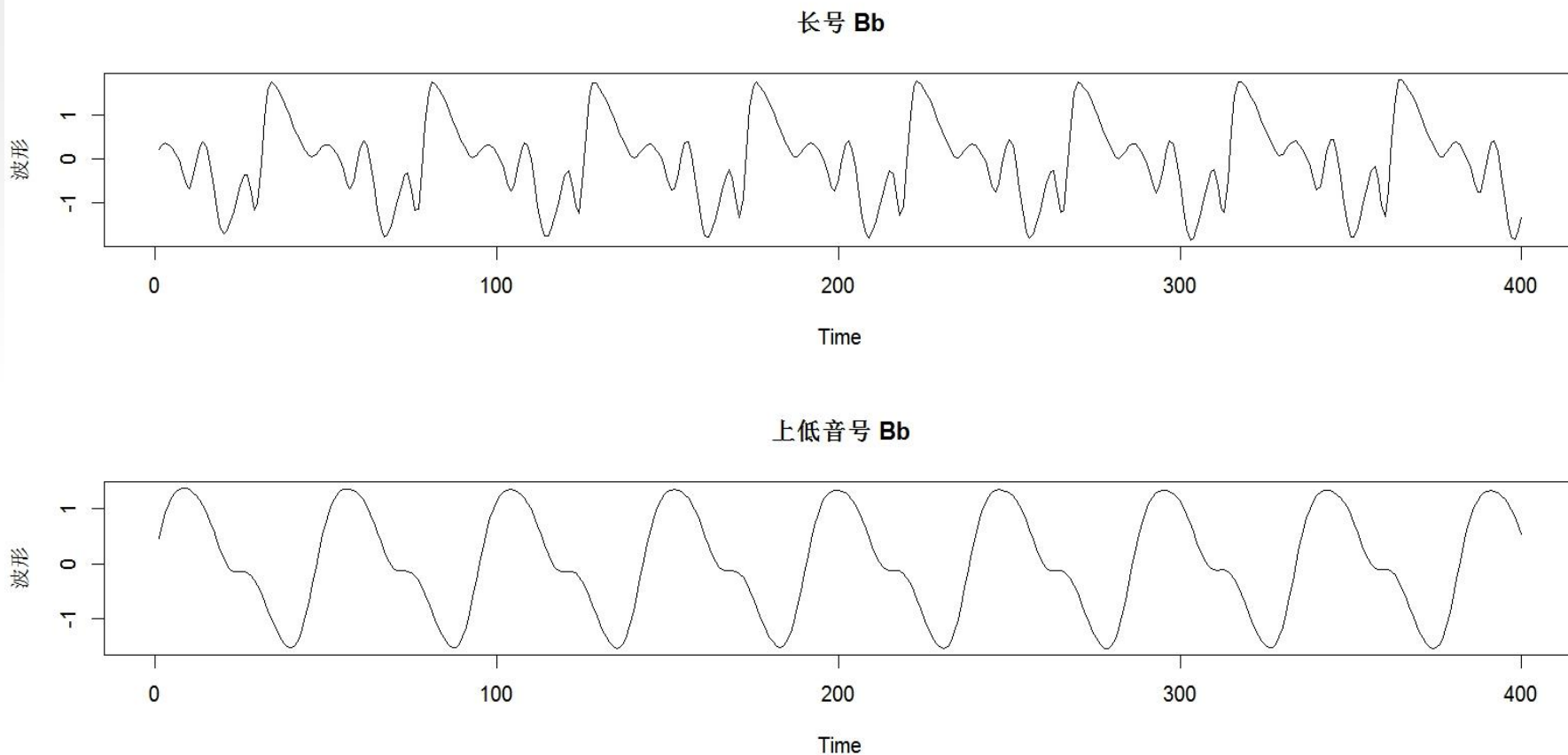
回顾：任何一个函数都可以展开为如下的傅里叶级数

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(n\omega t) + b_n \sin(n\omega t)]$$

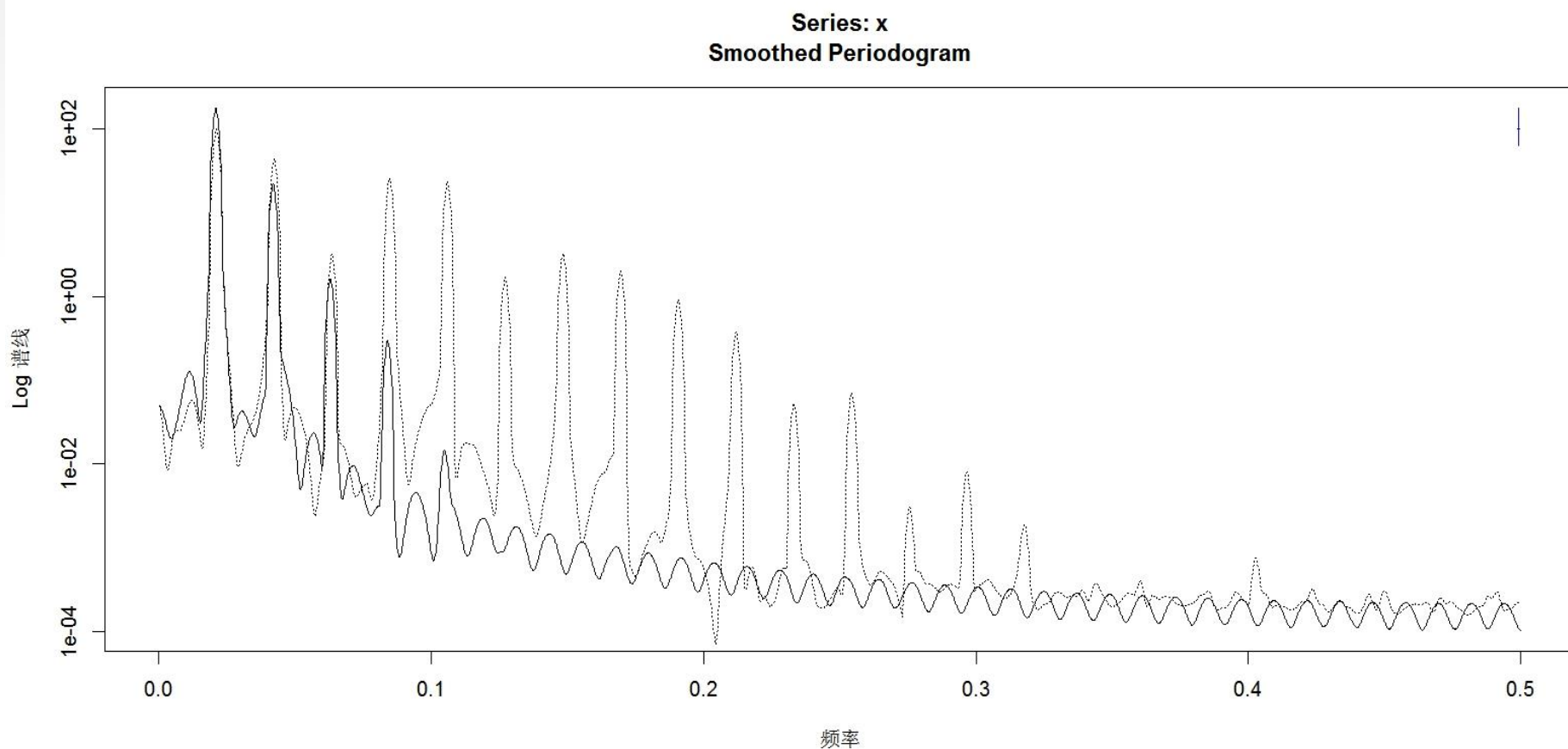
⑥



长号和上低音号演奏的降B音的时间序列图



长号 and 上低音号波形对应的谱估计。虚线为长号，实线为上低音号。（样本谱密度即样本协方差函数的傅里叶变换）



■ 原理

- 事件的发展通常都具有一定的惯性，这种惯性用统计的语言来描述就是序列值之间存在着一定的相关关系，这种相关关系通常具有某种统计规律。

■ 目的

- 寻找出序列值之间相关关系的统计规律，并拟合出适当的数学模型来描述这种规律，进而利用这个拟合模型预测序列未来的走势

■ 特点

- 理论基础扎实，操作步骤规范，分析结果易于解释，是时间序列分析的主流方法

时域分析方法的分析步骤

- 考察观察值序列的特征
- 根据序列的特征选择适当的拟合模型
- 根据序列的观察数据确定模型的口径
- 检验模型，优化模型
- 利用拟合好的模型来推断序列其它的统计性质或预测序列将来的发展

时域分析方法的发展过程



■ G.U.Yule

- 英国数学家。1927年，Yule提出用线性回归方程来模拟一个时间序列，这是最早的AR模型。

■ G.T.Walker

- 英国数学家，天文学家。
- 1931年，Walker利用Yule的分析方法研究了衰减正弦时间序列，得出Yule-Walker方程



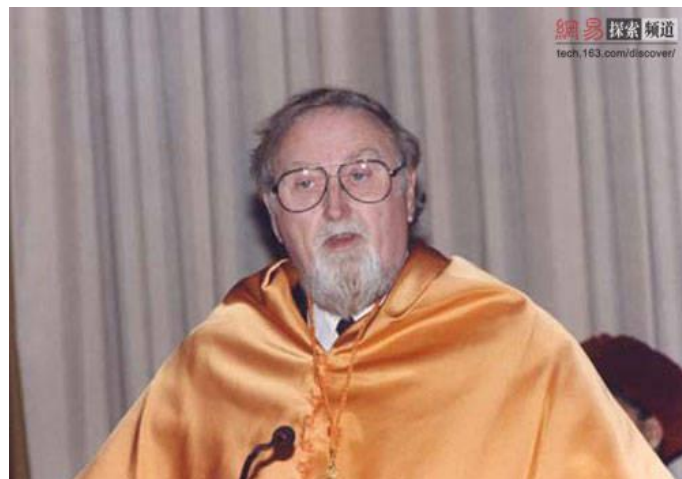
- G.E.P.Box和 G.M.Jenkins
 - 1970年，他们出版了《Time Series Analysis Forecasting and Control》一书
 - 书中，他们系统地阐述了ARIMA模型的识别、估计、检验及预测的原理及方法。这些知识现在被称为经典时间序列分析方法，是时域分析方法的核心内容。
 - 为了纪念Box和Jenkins对时间序列发展的特殊贡献，现在人们也常把ARIMA模型称为Box-Jenkins模型。
- ARIMA模型的实质
 - 单变量、同方差场合的线性模型

完善阶段

- 异方差场合
 - Robert F.Engle, 1982年, ARCH模型
 - Bollerslov, 1985年, GARCH模型
 - Nelson等人提出了GARCH模型的多种衍生模型
- 多变量场合
 - C.Granger, 1987年提出了协整 (co-integration) 理论
- 非线性场合
 - 汤家豪等, 1980年, 门限自回归模型
 - C.Granger, 1978年, 双线性模型



Robert F.Engle



C.Granger

四、时间序列分析软件

■ 常用软件

- S-plus, Matlab, Gauss, TSP, Eviews、SAS 和 R

■ 推荐软件——R

- R从开发伊始就定位为开源软件，所以它所有的代码和帮助文件免费向全球用户开放。
- 目前，R语言已经有若干个专门用于时间序列分析的程序包。借助这些程序包，我们能够完成序列读入，绘图，识别，建模，预测等一系列分析工作。R语言编程简洁，绘图功能强大，分析结果准确，是进行时间序列分析与预测的常用软件之一。

- R是完全免费的
- R可以在运行于UNIX, Windows和Macintosh的操作系统上.
- R具有很强的作图能力.
- 通过R语言的许多内嵌统计函数, 容易掌握R语言的语法.
- 可以编制自己的函数来扩展现有的R语言(这就是为什么它在不断升级完善!!)
-

R 的网站资源

- R主页: <http://www.r-project.org>
- CRAN (Comprehensive R Archive Network),
<http://cran.r-project.org>
- CRAN的镜像站点
<http://cran.r-project.org/mirrors.html>

统计分析软件包

– CRAN提供了许多便于统计分析的宏包：

[http://cran.r-](http://cran.r-project.org/src/contrib/PACKAGES.html)

[project.org/src/contrib/PACKAGES.html](http://cran.r-project.org/src/contrib/PACKAGES.html)

- stable -- 稳妥(分布)广义回归分析
- tseries – 时间序列分析
- VaR – 风险值分析
- matrix – 矩阵运算
- cinterface – C与R的接口

还有很多.....

- 启动R, 它由三部分组成: 主菜单 工具条

R console (R的运行窗口)

- 你可以得到在线帮助

`help()` 得到相应函数的帮助,例如 `help(plot)`

`help.start()` HTML格式的关于R的帮助文件

`demo()` 得到R提供的几个示例; `demo(fm)`

`q()` 退出R

程序包的安装和调用

程序包的安装两种方式：

- 1、通过菜单栏选项下载安装。（见教材）
- 2、输入指令安装。例如安装tseries程序包：

```
install.packages("tseries")
```

程序包的调用：

```
library(tseries)
```

我们用一些例子来看R软件的特点。

输入以下语句：

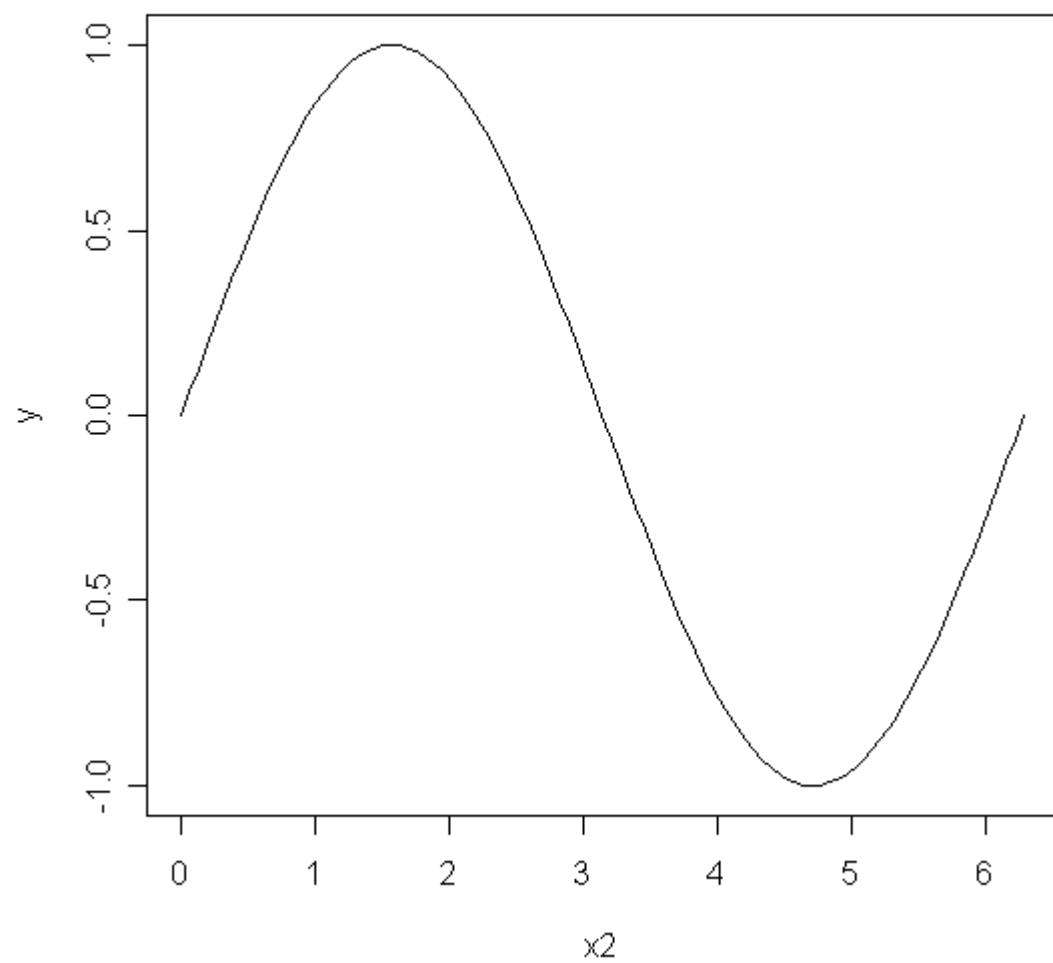
```
> x1 <- 0:100
```

```
> x2 <- x1*2*pi/100
```

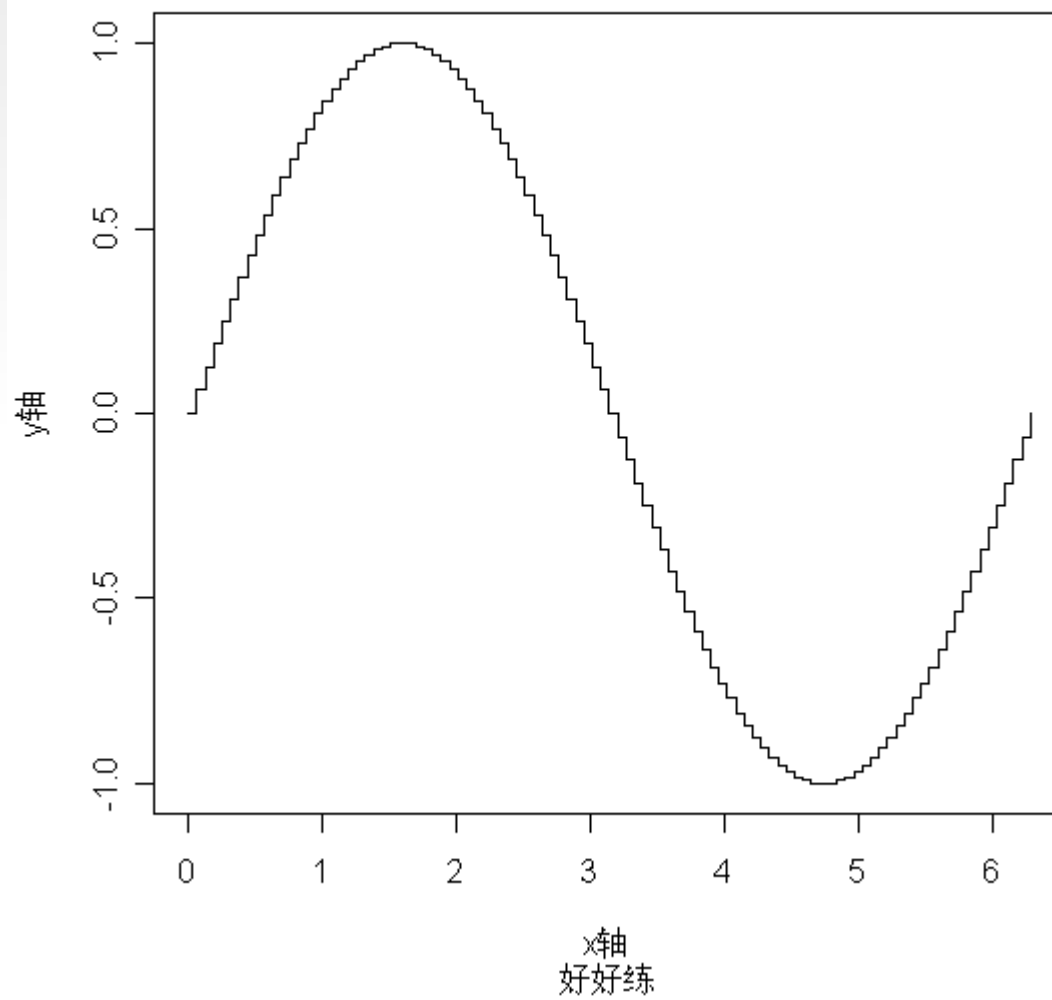
```
> y <- sin(x2)
```

```
> plot(x2,y, type='l')
```

```
> plot(x2,y,main="画图练习", type="s",  
sub="好好练", xlab="x轴",ylab='y轴')
```



画图练习



常用统计量和数学符号（见教材）

sum, mean, var, sd, min, max, range,
median, IQR（四分位间距）等统计量,
sort, order, rank与排序有关,
其它还有ave, fivenum, mad, quantile, stem等。

abs, sqrt: 绝对值, 平方根

log, exp, log10, log2: 对数与指数函数

sin, cos, tan, asin, acos, atan: 三角函数

sinh, cosh, tanh, asinh, acosh: 双曲函数

下面我们看一看S的简单统计功能:

```
> marks <- c(10, 6, 4, 7, 8)
```

```
> mean(marks)
```

```
> sd(marks)
```

```
> median(marks)
```

```
> min(marks)
```

```
> max(marks)
```

```
> boxplot(marks)
```

在R中生成时间序列数据

事件序列数据是一种包含观测值、起始时间你、终止时间以及周期的结构。只有将数据转成时间序列数据后，我们才能用各种时间序列方法对其进行分析、建模和绘图。

```
> sales <- c(18, 33, 41, 7, 34, 35, 24, 25,  
             24, 21, 25, 20, 22, 31, 40, 29, 25, 21,  
             22, 54, 31, 25, 26, 35)
```

```
> tsales<-ts(sales, start=c(2003, 1),  
             frequency=12)
```

```
tsales
```

```
plot(tsales)
```

```
start(tsales)
```

```
end(tsales)
```

```
frequency(tsales)
```

```
tsales.subset<- window(tsales,
```

```
start=c(2003, 5), end=c(2004, 6))
```

从其他文件中读取数据

R可以从其他文件格式中读取数据。

导入csv数据:

```
x<-read.table('H:/file1',sep="," ,header=T)
```

```
x$yield
```

导入txt数据:

```
data1 <- read.table('E:/5.2.txt',sep = '\t',header =  
TRUE)
```