

# 时间序列分析

## 概论

陈 显

cyu@ustc.edu.cn

东区管理科研楼 1003

63602243

2022 年 8 月

- 授课方式：PPT+板书
- 考评方式：

到课+上机作业 (30%) + 期末考试 (50%) + 大作业 (20%)

- 教材：《应用时间序列分析》  
何书元 北京大学出版社
- 主要参考

- ① 《金融时间序列分析》  
Ruey S. Tsay 王远林译 人民邮电出版社
- ② 《时间序列的理论与方法》  
Peter J. Brockwell and Richard A. Davis 田铮译（第二版）  
高等教育出版社
- ③ 《Introduction to time series and forecasting》  
Peter J. Brockwell and Richard A. Davis Springer

# 课程内容、基础介绍

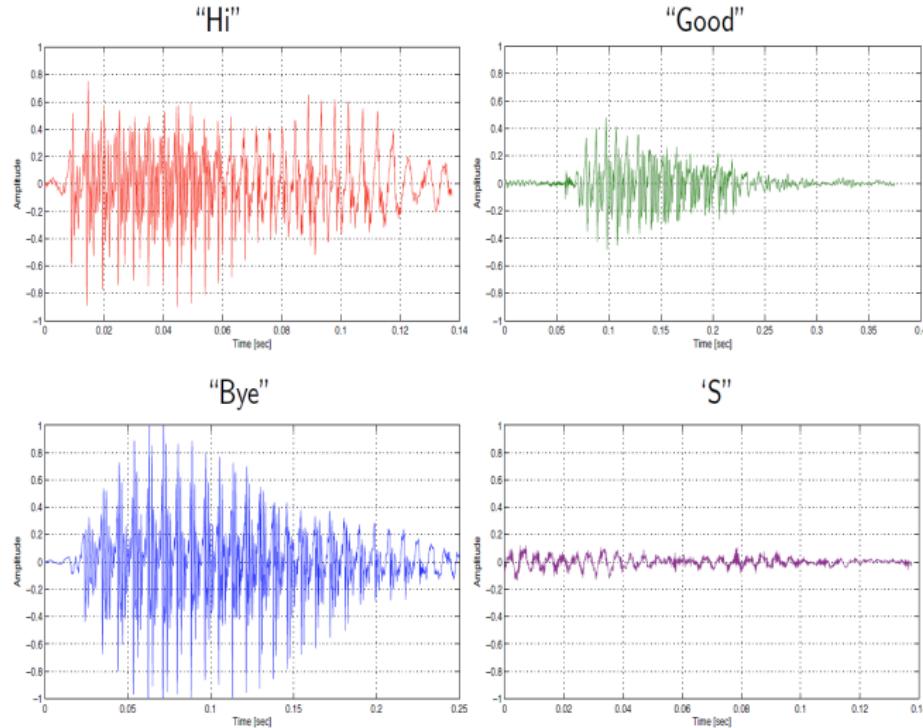
- 研究沿时间记录的数据(等间距的).
- 针对序列不同时间点的相关性, 方法是相关法和Markov方法.
- 主要使用二阶矩(自相关函数)、矩阵运算(二阶矩过程相关阵)和平稳过程.
- 也用部分的基础泛函分析、复变函数(频域分析).
- 有广泛应用但本课程主要是打基础. 应用领域:
  - ① 经济与金融, 经济数据序列, 价格序列, 指数序列。
  - ② 信号系统, 比如机械振动, 无线电信号, 音频信号。
  - ③ 社会、科学研究, 医疗等, 人体生理节律, 化学反应, 交通流量等。

# 什么是时间序列

- A variable measured sequentially at **fixed intervals** of time.
- 和统计里的样本有什么差别?  
Measurements are often **correlated** in time.

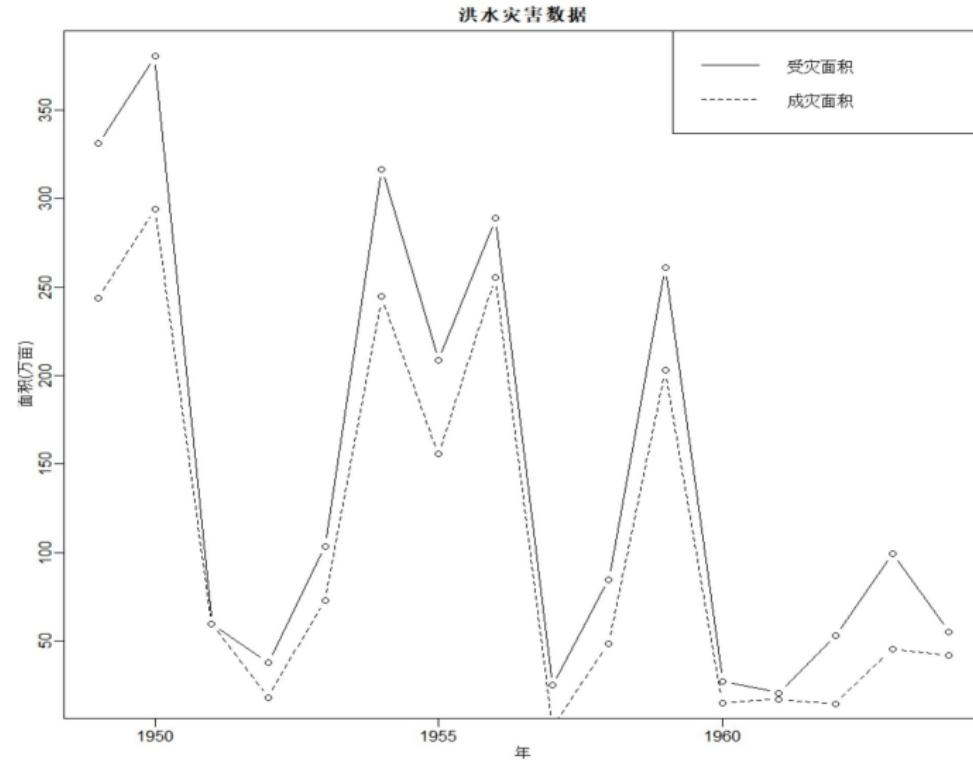
# 例子

## 语音识别系统



# 例子

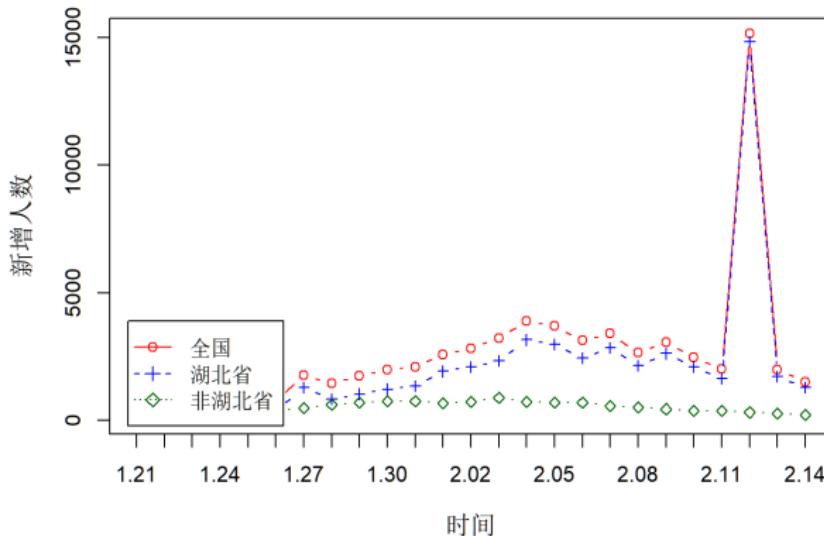
## 洪水灾害时序图



# 新冠病毒确诊人数时序图

对于新型冠状病毒，新增确诊人数

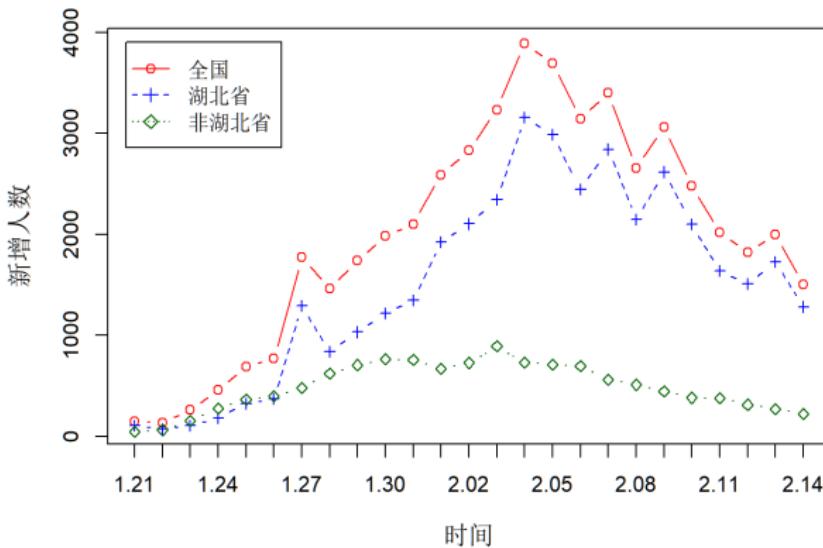
## 新型肺炎确诊病例数时序图



# 新冠病毒确诊人数时序图

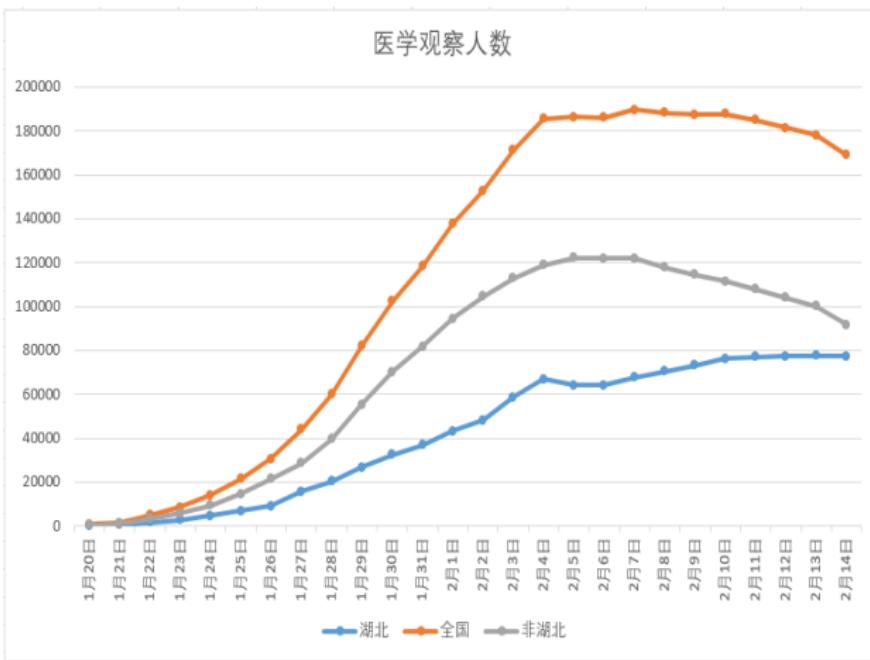
对于新型冠状病毒，新增（临床诊断去除）确诊人数

## 新型肺炎确诊病例数时序图



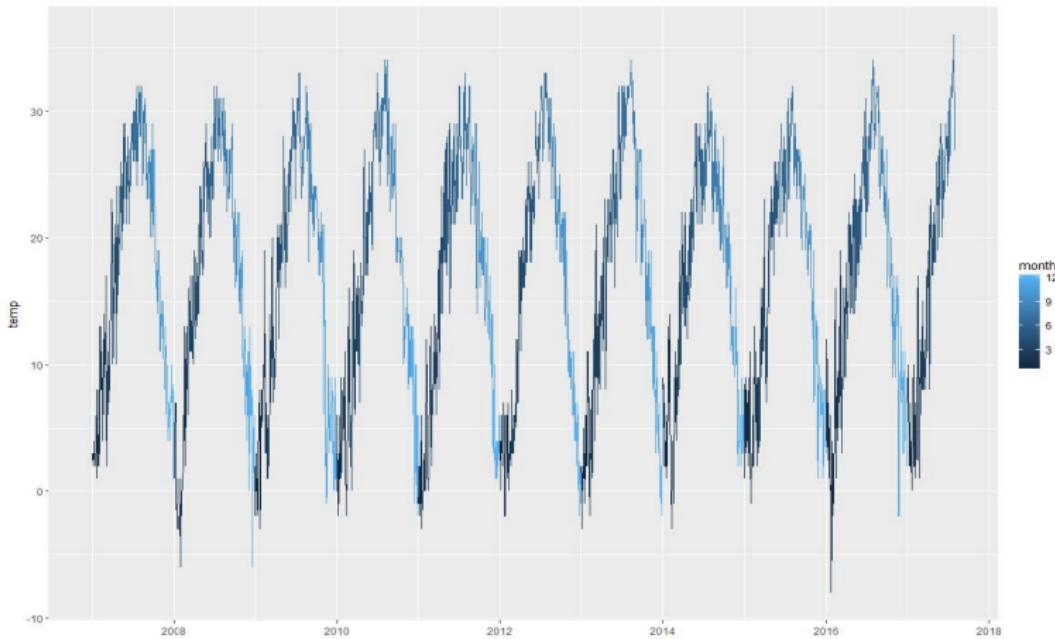
# 时间序列实际例子

对于新型冠状病毒，我们特别关注全国正在接受医学观察的人数



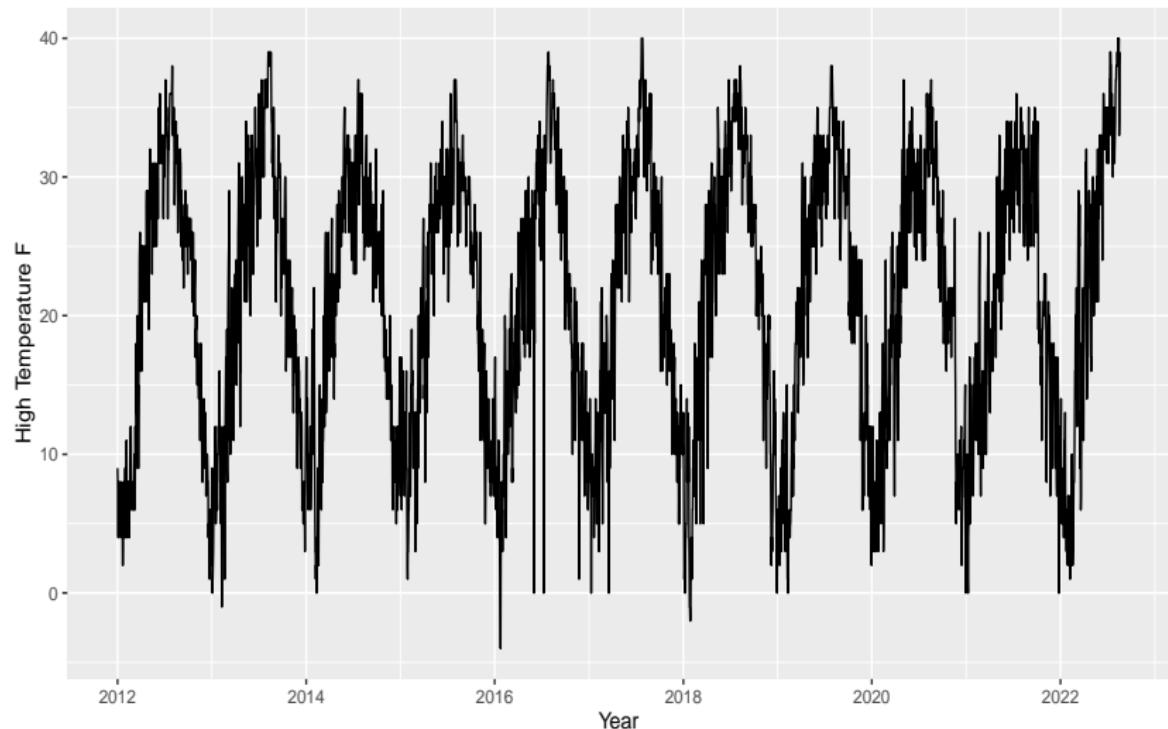
# 合肥的平均气温序列

下图能看出合肥天气是变冷了还是变暖了？



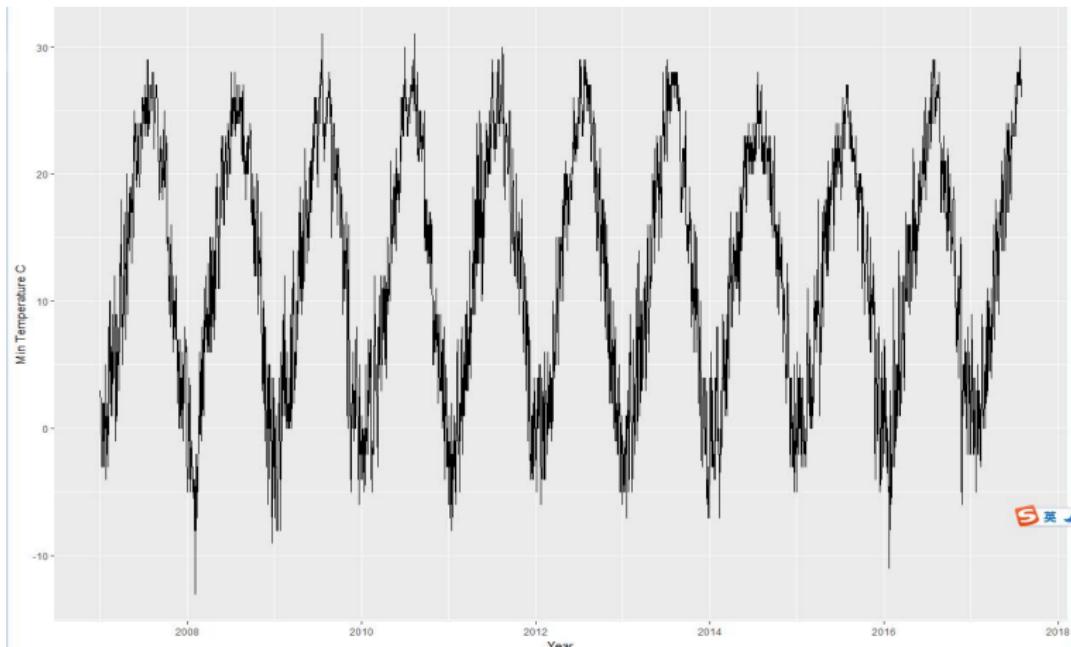
# 合肥的最高气温序列

下图能看出合肥天气是夏天越来越热?

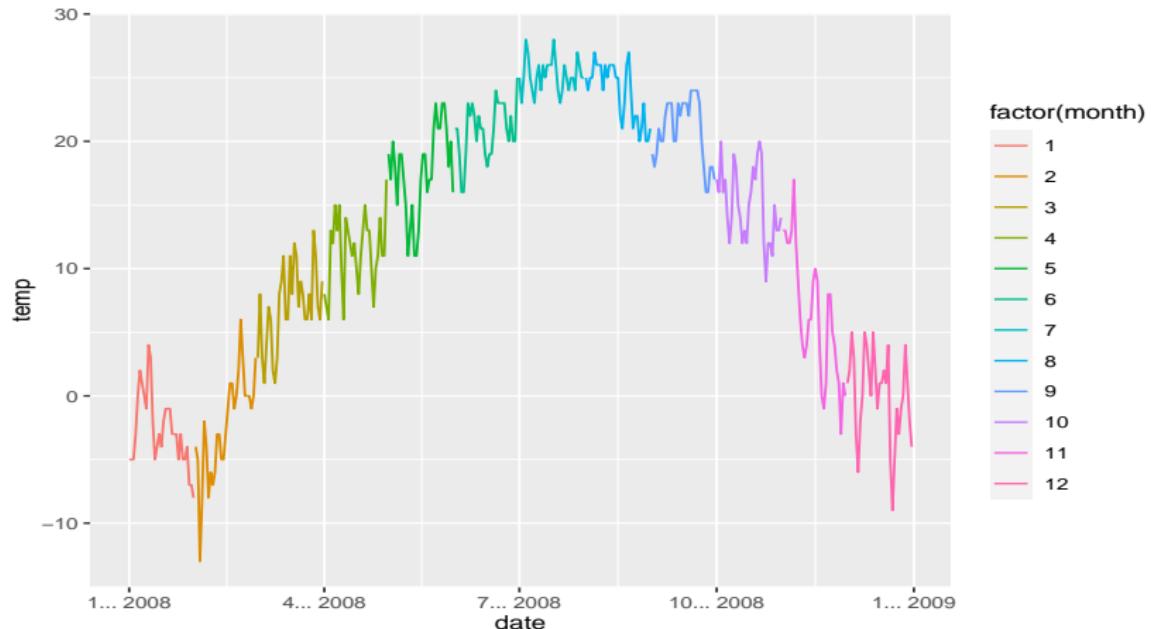


# 合肥的最低气温序列

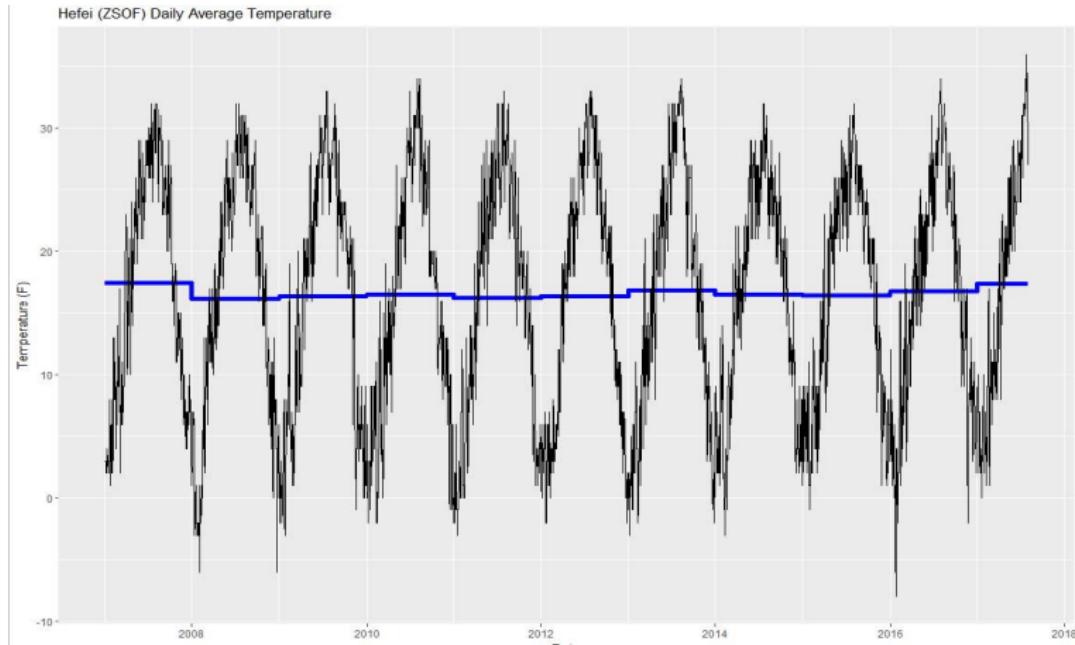
下图能看出合肥天气是冬天变暖了？



# 2008年合肥的最低气温序列

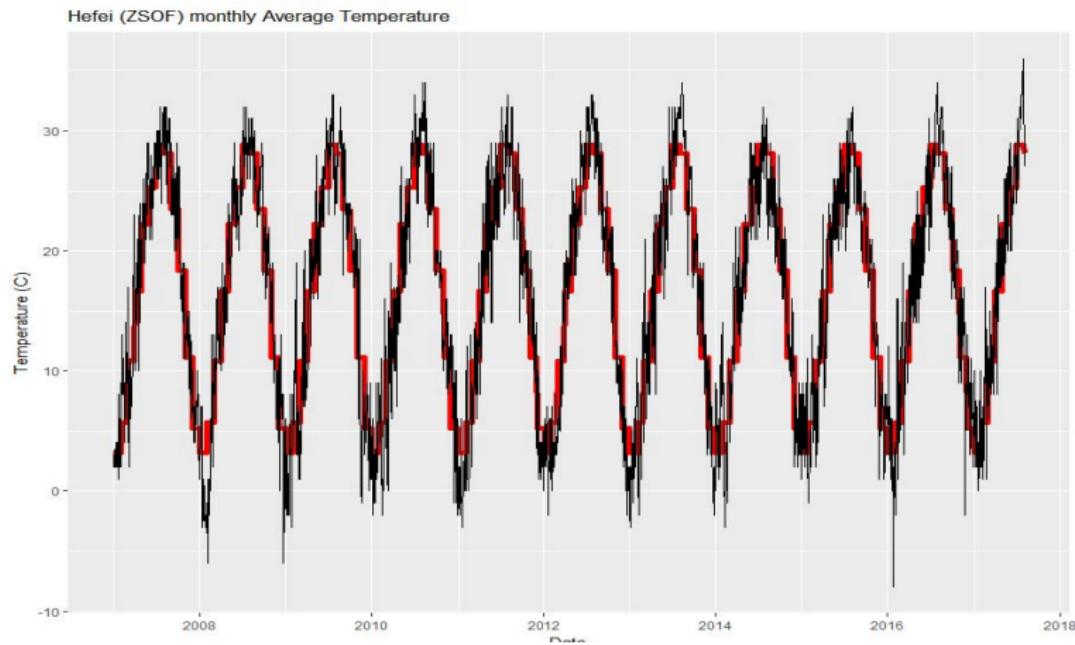


# 合肥的年平均气温序列



# 合肥的月平均气温序列

红色线为月平均序列.

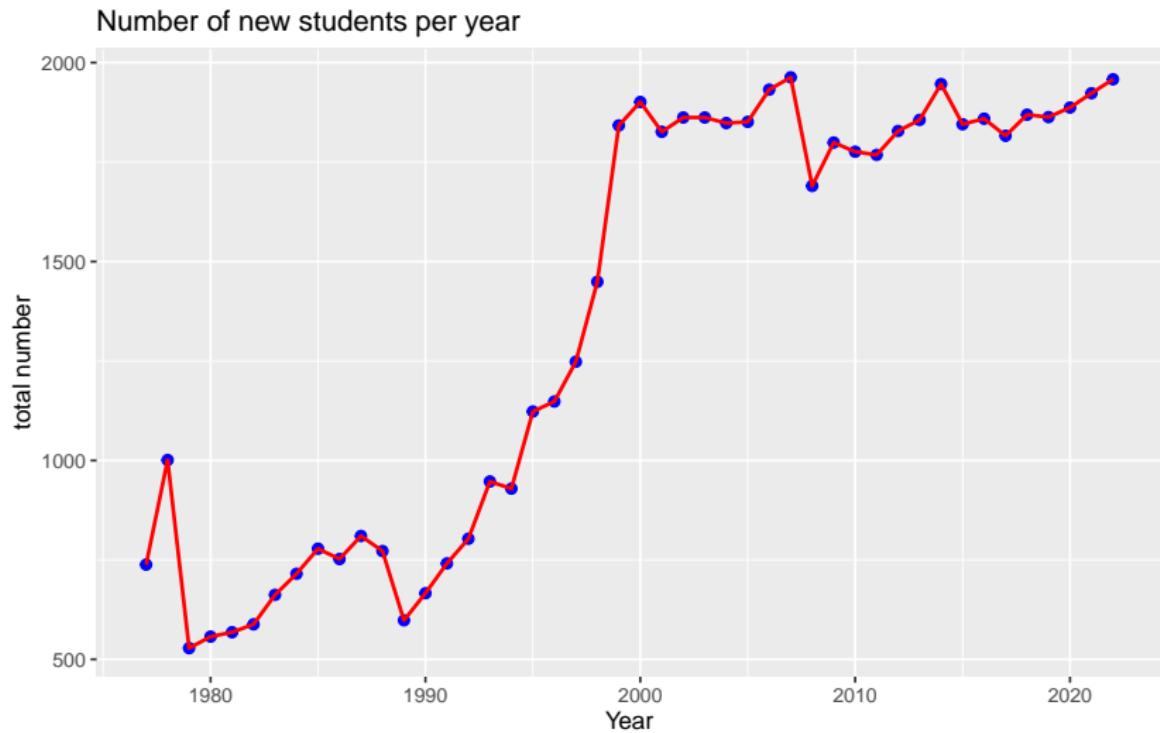


# 第一章 时间序列概论

1. 引言
2. 时间序列的定义
3. 时间序列的分析方法
  - 3.1. 时域分析方法
  - 3.2. 统计分析方法
4. 时间序列分析软件
5. 课程主要内容

1. 最早的时间序列分析可以追溯到7000年前的古埃及.
  - 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. 由于掌握了尼罗河泛滥的规律, 使得古埃及的农业迅速发展, 从而创建了埃及灿烂的史前文明.
2. 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, 找寻它长期变化发展的统计规律, 预测它将来的走势 就是时间序列分析.

# 科大每年的录取新生数



## 2. 时间序列的定义

- **随机序列**: 按时间顺序排列的一组随机变量 $\{X_t, t \in T\}$ ,

$$\dots, X_1, X_2, \dots, X_t, \dots$$

- **观察值序列**: 随机序列的 $n$ 个有序观察值, 称之为序列长度为 $n$ 的观察值序列

$$x_1, x_2, \dots, x_n.$$

- **随机序列和观察值序列之间的关系**

- ◆ 观察值序列是随机序列的一个实现.
- ◆ 我们研究的目的是想揭示随机时序的性质.
- ◆ 实现的手段都是通过观察值序列的性质进行推断.

### 3 时间序列的分析方法

- 描述性时序分析

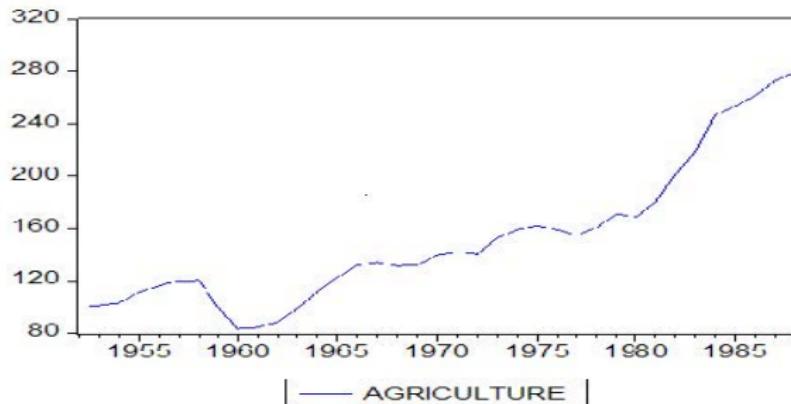
- 通过直观的数据比较或绘图观测，寻找序列中蕴含的**发展规律**，这种分析方法就称为描述性时序分析。
- 描述性时序分析方法具有**操作简单、直观有效** 的特点，它通常是人们进行统计时序分析的第一步。

- 统计时序分析

- 频域分析方法。
- 时域分析方法。

# 描述性时序分析案例

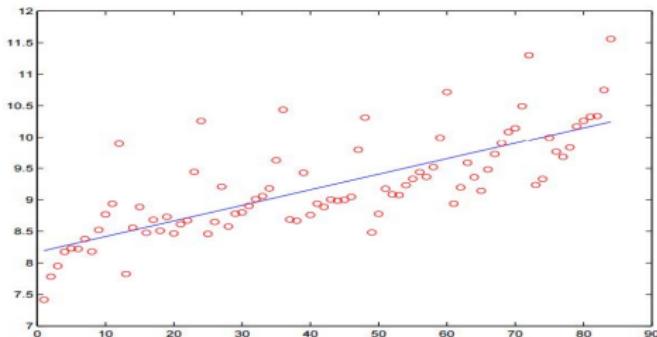
趋势性Trend: 时间序列存在向上或者向下的运动趋势



# 描述性时序分析案例

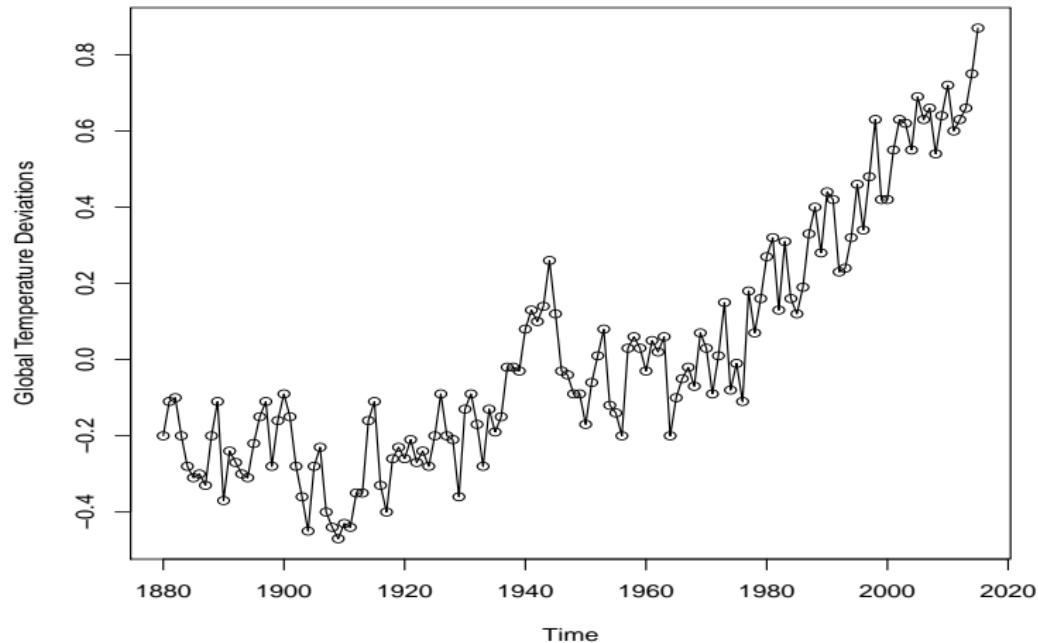
趋势性Trend：时间序列存在向上或者向下的运动趋势

Trend



# 描述性时序分析案例-全球变暖

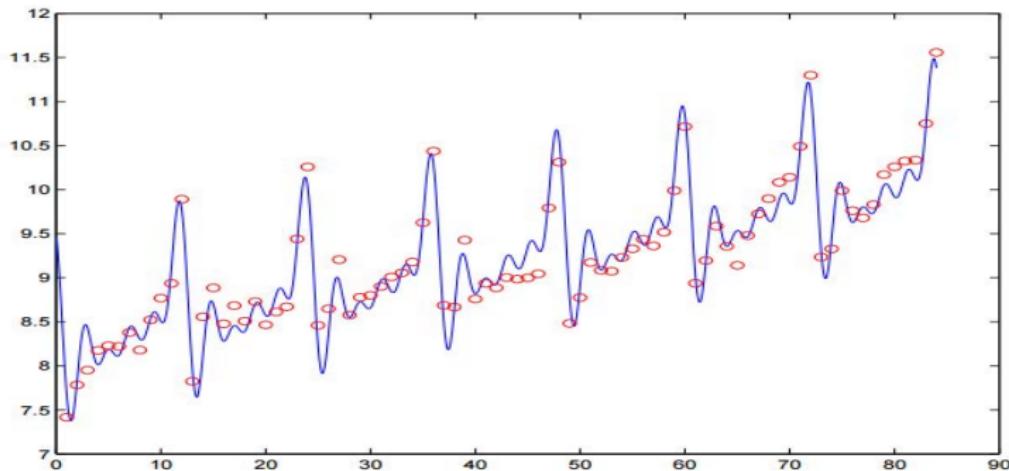
## 全球气温序列



# 描述性时序分析案例

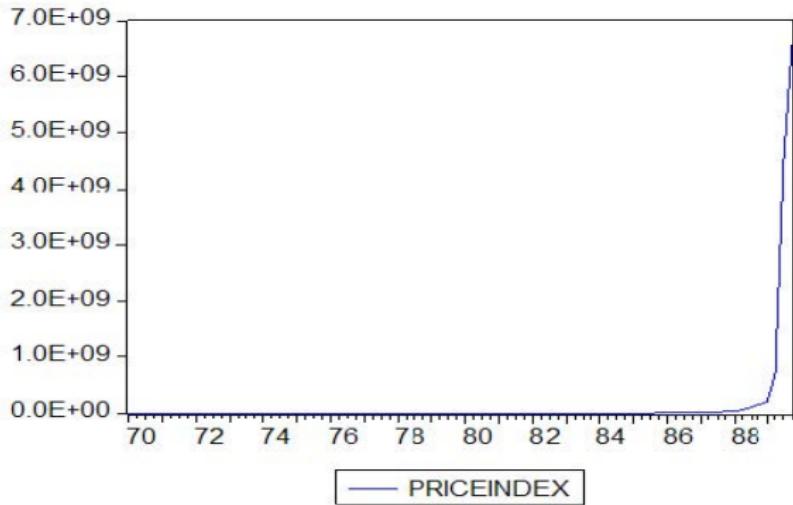
季节性seasonal：某个季节的观测值与其他季节的观测值显著不同

## Trend and seasonal variation

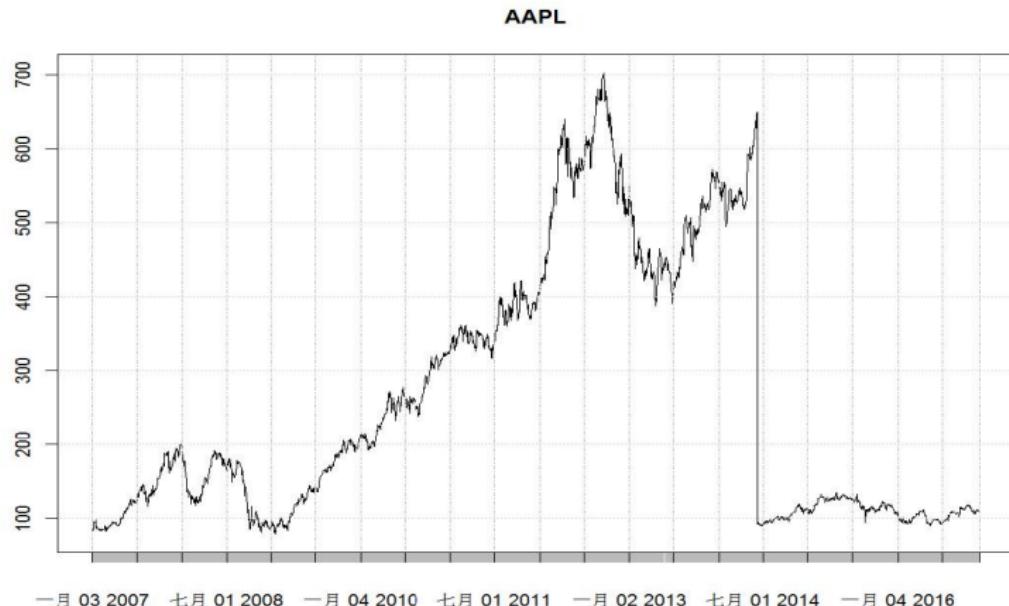


# 描述性时序分析案例

异常观察值：奇异观察值



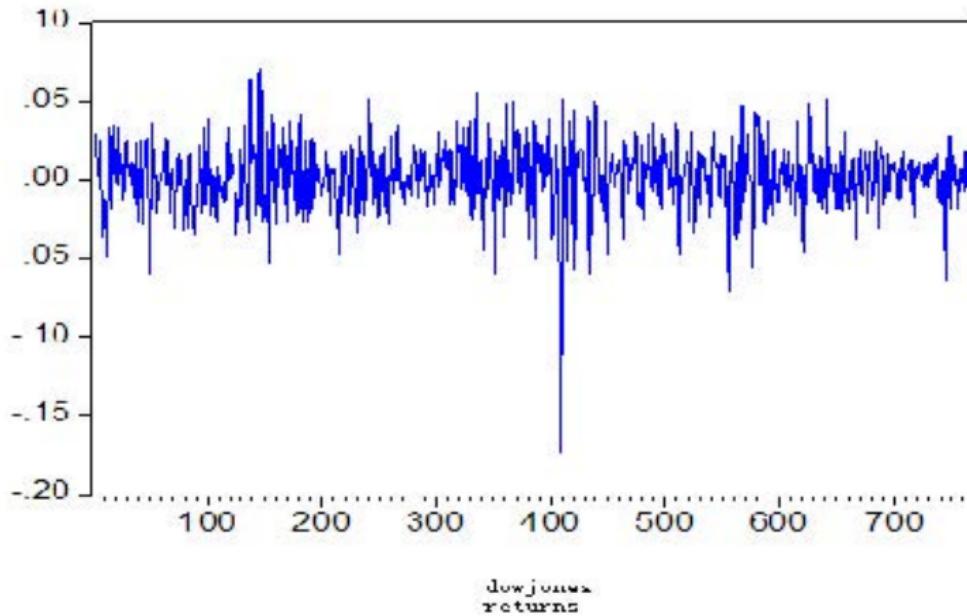
# 描述性时序分析案例-苹果股价



美国时间6月9日(星期一), 苹果公司(AAPL)实施7:1比例拆股后, 股价从上周五收盘价645.57美元除权为92.22美元

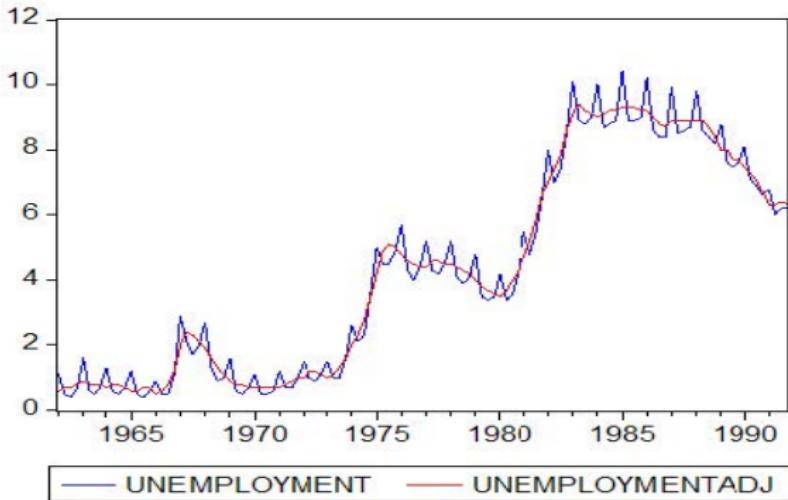
# 描述性时序分析案例

条件异方差性：波动积聚性(某一特征的值重复出现)



# 描述性时序分析案例

非线性：线性以外的一切情况



## 1. 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**.

## 2. 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**.

## 3. 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.

- **数据的简洁描述**  $X_t = T_t + S_t + Y_t$
- **解释** 平稳过程建模、季节调整等
- **预测** 预测趋势
- **控制** 例如根据失业率调整货币政策
- **假设检验** 例如全球变暖
- **模拟** 灾难性事件发生的概率等

## 1. 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

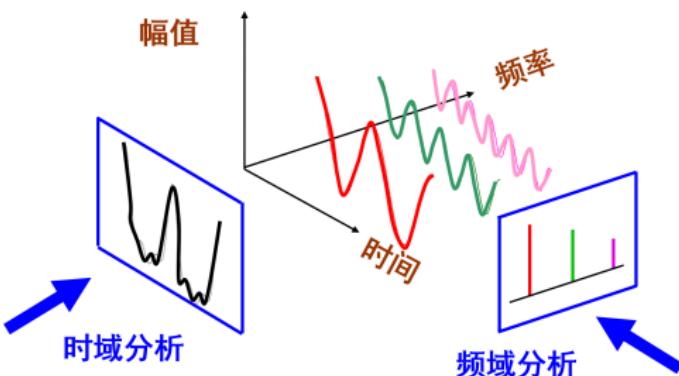
## 2. 发展过程

- 早期的频域分析方法借助富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，用正弦、余弦项之和来逼近某个函数.
- 20世纪60年代，Burg引入最大熵谱估计理论，进入现代谱分析阶段. 用在工程方面，海洋学，天文学等领域.

## 3. 特点

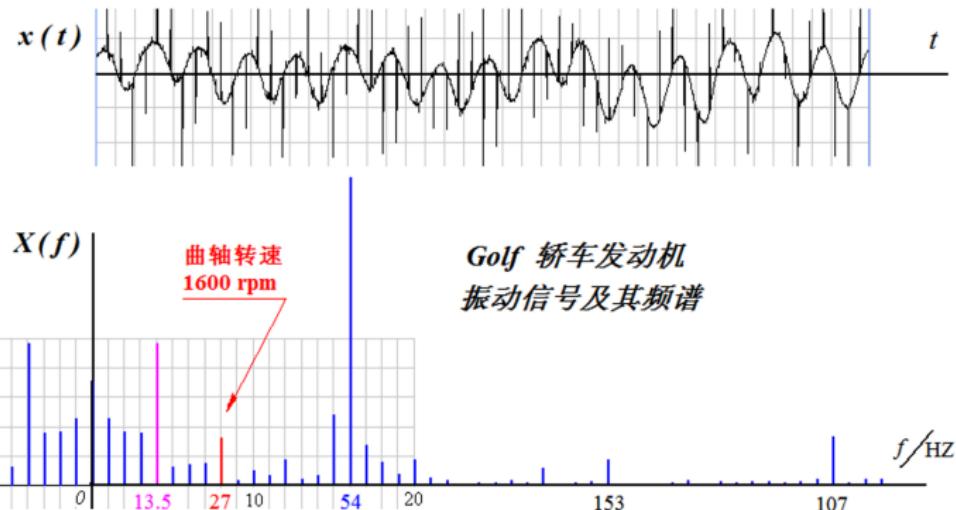
- 非常有用的动态数据分析方法，但是由于分析方法复杂，结果抽象，有一定的使用局限性.
- 将平稳过程看成余弦和正弦形式的周期函数的加权和，目标确定不同频率中的重要周期如何影响过程的行为.

# 时域与频域分析方法比较



不同的分析方法不会改变时间序列的性质，只是分析问题的角度不同。根据需要选取不同的分析方法，经济和金融分析以时域为主，信号与系统等工科以频域为主。

# 时域与频域分析方法比较



寻找发动机故障使用频域会更方便一些.

## 1. 画出时序图

寻找趋势, 季节因素, 阶梯变换, 异常点

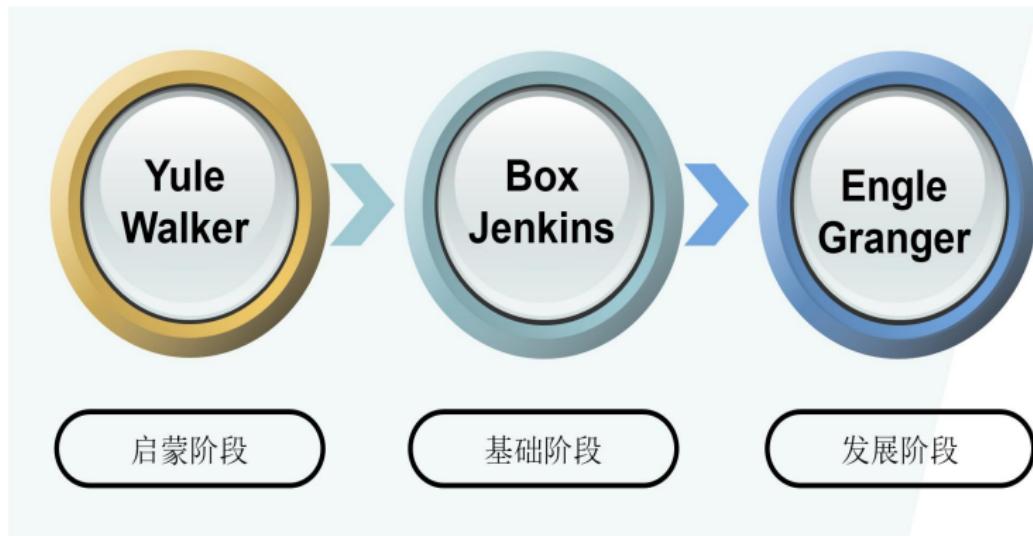
## 2. 变化数据(建立模型)使得残差是平稳过程

- 估计和减去趋势和季节性  $T_t, S_t$ .
- 差分
- 非线性变化( $\log, \sqrt{\cdot}$ )

## 3. 对残差进行建模.

1. **模型建立**: 通过观察时间序列数据的特征, 如自相关性等来建立模型
2. **方法选择**: 频域分析和时域分析, 如ARMA模型主要利用了自相关函数。
3. **数据选择**: 变量选择和时间长度的选择.
4. **预测**: 样本外预测前提—样本内数据与样本外数据具有相似性 (平稳性)

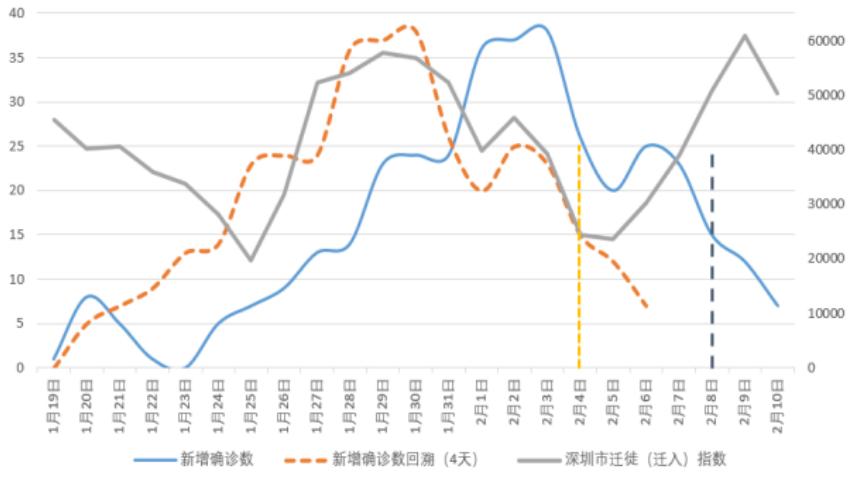
# 时域分析方法的发展过程

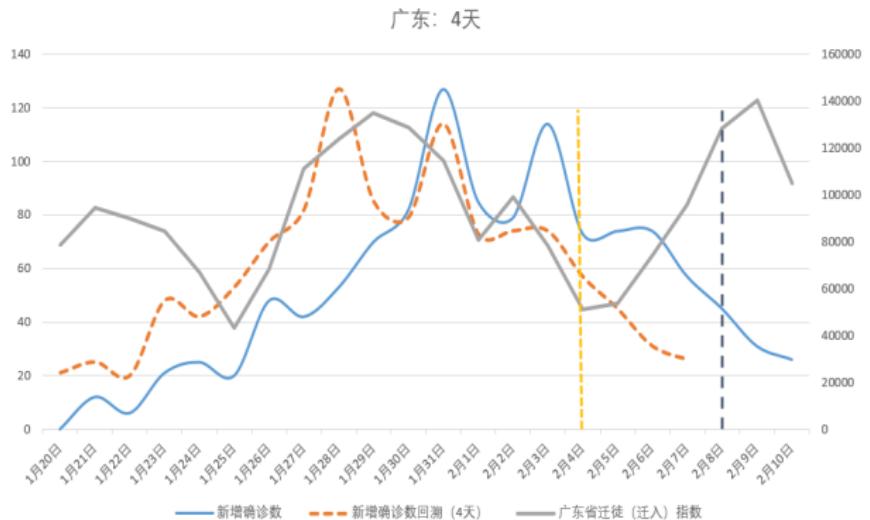


# 课程主要内容

- 平稳时间序列
- 平稳ARMA过程
- 平稳过程预报及ARMA模型的估计
- 平稳过程谱表示
- 非平稳时间序列ARIMA,FARIMA
- ARCH和GARCH
- 向量自回归VAR

## 深圳市：按确诊日滞后期4天





- 常用软件
  - Python, R, S-plus, Matlab, Gauss, TSP, Eviews 和SAS
- 推荐软件——Python, R, SAS 和Eviews.
  - 在SAS系统中有一个专门进行计量经济与时间序列分析的模块: SAS/ETS. SAS/ETS编程语言简洁, 输出功能强大, 分析结果精确, 是进行时间序列分析与预测的理想软件.
  - 由于SAS系统具有全球一流的数据仓库功能, 因此在进行海量数据的时间序列分析时它具有其它统计软件无可比拟的优势
- <https://www.machinelearningplus.com/time-series/time-series-analysis-python/>

- 著名S语言的GNU应用，该软件在R小组和世界各地统计学家的贡献下发展的
- <http://www.r-project.org> 可访问该网站免费下载此软件
- 使用R程序中的TSA程序包， leaps, locfit, MASS, mgcv, tseries, uroot, fgarch等

1. 非常有用的R包，可以直接下载一些金融数据，比如Yahoo Finance, Google Finance and the Federal Reserve Economic Data (FRED) of Federal Reserve Bank of St. Louis.
2. 得到数据命令getSymbols(可以选择感兴趣的区间:  
`from="2007-01-03",to="2016-12-02"`)
3. 时序图命令chartSeries. (`theme="white"` of chartSeries, 选择时序图的背景色)

- library(quantmod)
- setSymbolLookup(ICBC=list(name="601398.ss",src="yahoo"))
- getSymbols("ICBC",from="2015-06-02",to="2017-07-31")
- a1=100\*dailyReturn(ICBC)
- chartSeries(ICBC,theme="white")

- 本课程主要介绍时间序列分析的基本理论和方法,时间序列分析的一些新发展及其在现代经济、金融学中的应用. 要求大家通过学习,熟悉社会、信号、经济领域特别是金融领域中常用的时间序列分析手段,掌握常用的时间序列模型,能对社会、经济、金融 时间序列数据进行处理、建立模型,以预测、解释社会现象、经济理论或检验经济假说.
  - 能够掌握时间序列分析的基本方法.
  - 能够应用时间序列方法解决实际问题.