# 数学建模竞赛方法之
# 数据处理与模拟验证

东华大学统计系 胡良剑

Ljhu@dhu.edu.cn

**https://pan.baidu.com/s/1aPSj2rlDd3Qi5TPSqy5GwQ**

**提取码：9o23**

# 内容提要

- 数据导入
- 数据预处理
- 数据可视化
- 数据建模方法
- 随机模拟

# 数据导入(Matlab)

- 非编程：剪贴板、工具栏“导入数据”
- xlsread 读取Excel文件
- textscan读取文本文件
- imread 读取图像文件
- Importdata 读取各类数据文件
- readtable读取各种表格
- 其他：百度去吧

# 例子

- 读Excel数据(2011年A题重金属污染)
  - As = xlsread('cumcm2011A附件_数据.xls','附件2',‘b4:b322');
- 读txt数据(2004年B题输电阻塞管理)
  - data=importdata('chuli.txt')
- 读jpg数据(2013年碎纸片拼接)
  - data=imread (‘000.bmp')

# 数据为什么要做预处理？

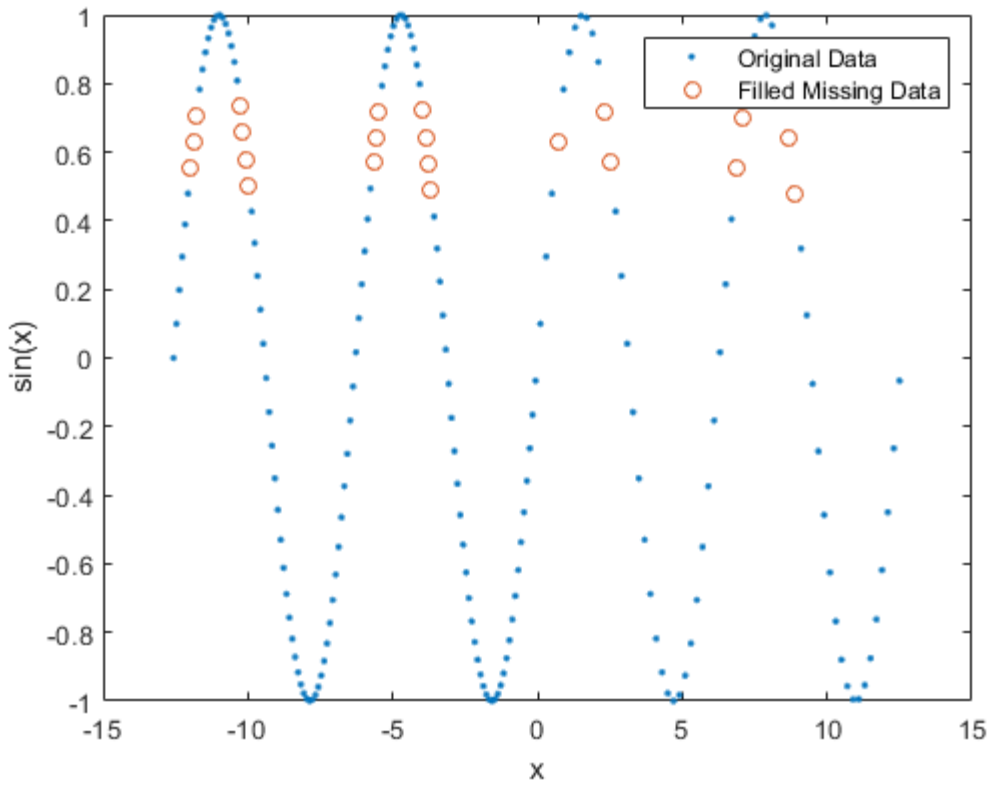
- **不完整性**
  - 缺失值
- **不一致性**
  - 数据来源不一致
  - 名称、单位的不一致性
- **噪声数据**
  - 离群或错误值
  - 信息干扰（噪声）
- **复杂性**
  - 冗余信息
  - 高维数

# 数据预处理的主要任务

- 数据清洗
  - 填入缺失数据
  - 平滑噪声数据
  - 检测和处理离群值
- 数据集成
  - 多个数据库、文件系统的集成
  - 解决不一致性
- 数据标准化
  - 规范化、聚集等
- 数据归约
  - 在可能获得相同或相似结果的前提下，对数据的属性进行有效的缩减
- 数据离散化
  - 以区间值来代替实际数据值，以减少属性值的个数.

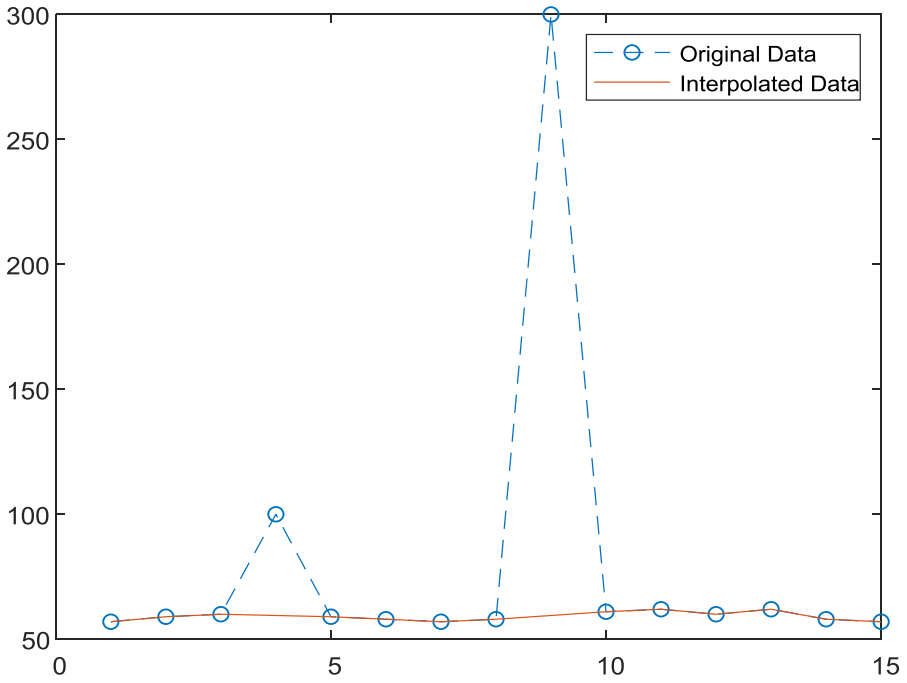
# 数据清洗

- 缺失值处理
  - 删除法
  - 高频替代法
  - 均值法
  - 插值法
  - 回归法
  - 聚类法



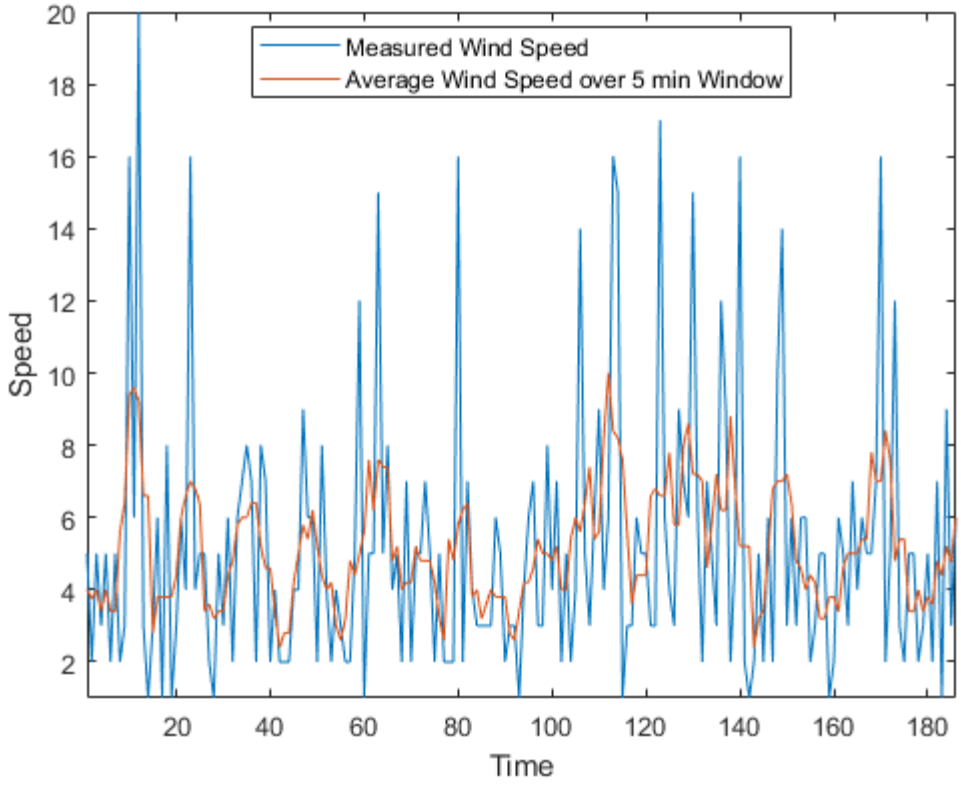
# 数据清洗

- 离群值的危害：造成统计量计算严重偏差
- 离群值检测
  - 近邻检测
  - 3σ(6σ)原则
  - 聚类检测
- 离群值处理
  - 删除
  - 插值法
  - 边界值替代
  - 不处理



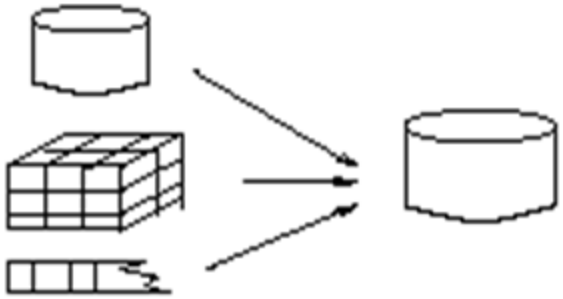
# 数据清洗

- 噪声数据的平滑
  - 移动平均法
  - 中值滤波
  - 盒子法(bin)
  - 高斯窗法
  - 指数法
  - 聚类法
  - 函数拟合法



# 数据集成

- 多个数据源的匹配
- 统一数据格式
- 消除冗余和冲突数据
- 特征编码



# 数据集成

- 数据类型冲突
  - 性别：string(Male、Female)、Char（M、F）、Interger（0、1）
  - 日期：Date、DateTime、String
- 数据标签冲突：解决同名异义、异名同义
  - 学生成绩、分数
- 度量单位冲突
  - 学生成绩
    - 百分制：100 ~ 0
    - 五分制： A 、B、C、D、E
    - 字符表示：优、良、及格、不及格
- 概念不清
  - 最近交易额：前一个小时、昨天、本周、本月？

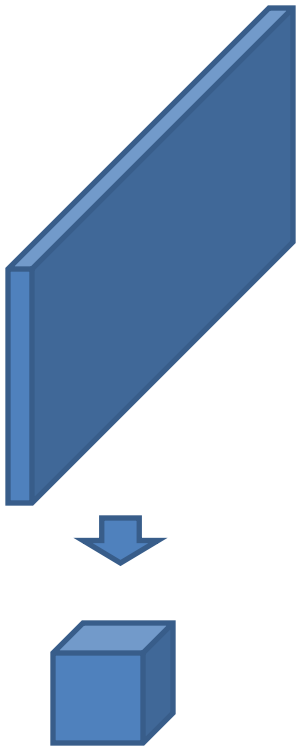
# 数据集成

- 特征编码: 名义数据的数值化
  - {是，否}→{0, 1}
  - {低，中，高} →{0, 1, 2}
  - {东北，东南，西北，西南 } →{1, 2, 3, 4}
  - {东北，东南，西北，西南 } →

| | | |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 1 |

# 数据的标准化

- 数据的标准化变换是消除属性间数量级差异的影响。
  - 均值-方差(Z-score)标准化
  - 极差(Min-Max)归一化
  - 分位数标准化
  - 小数定标标准化
  - 数学变换：取对数, Logistic函数等

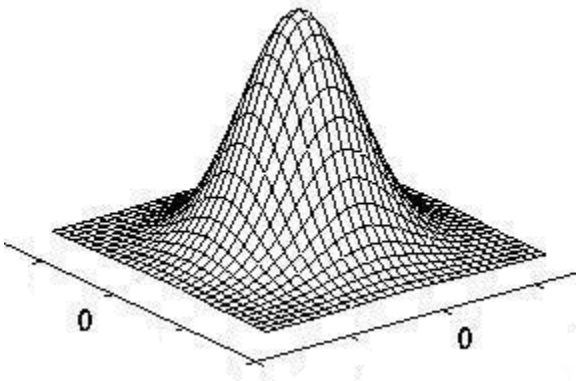


# 均值-方差(Z-score)标准化

$$\mathbf{X}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \mathbf{X}^{*}=\begin{pmatrix} x_{11}^{*} & x_{12}^{*} & \cdots & x_{1p}^{*} \\ x_{21}^{*} & x_{22}^{*} & \cdots & x_{2p}^{*} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1}^{*} & x_{n2}^{*} & \cdots & x_{np}^{*} \end{pmatrix},$$

$$x_{ij}^{*}=\frac{x_{ij}-\bar{x}_j}{s_j}, \quad i=1,2,\cdots,n, \quad j=1,2,\cdots,p,$$

$$\bar{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij}, \quad s_j=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}, \quad j=1,2,\cdots,p\ .$$

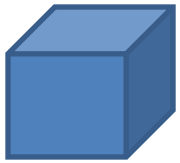



**Z-score标准化后各列数据的均值为0，方差为1.**

# 极差(Min-Max)归一化

$$\mathbf{X}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \qquad \mathbf{X}^R=\begin{pmatrix} x_{11}^R & x_{12}^R & \cdots & x_{1p}^R \\ x_{21}^R & x_{22}^R & \cdots & x_{2p}^R \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1}^R & x_{n2}^R & \cdots & x_{np}^R \end{pmatrix},$$

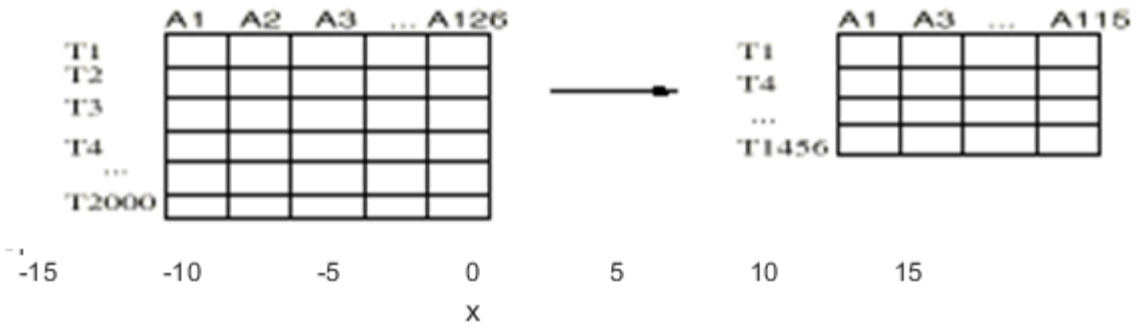
$$x_{ij}^R=\frac{x_{ij}-\min\limits_{1\le k\le n} x_{kj}}{\max\limits_{1\le k\le n} x_{kj}-\min\limits_{1\le k\le n} x_{kj}}, \quad i=1,2,\cdots,n, \quad j=1,2,\cdots,p,$$



**Min-Max归一化后各列数据的最小值为0，最大值为1.**

# 数据归约

- 数据立方体聚集：聚集操作作用于立方体中的数据
- 减少数据维度（维归约）：可以检测并删除不相关、弱相关或者冗余的属性或维。
- 数据压缩：使用编码机制压缩数据集
- 数值压缩：用替代的、较小的数据表示替换或估计数据



# 维归约(特征提取)

- 维归约：通过删除不相关的属性（或维）减少数据量
- 主成分分析、因子分析。
- 特征选取 (属性子集的选取):
  - 选取最小的特征属性集合，得到的数据挖掘结果与所有特征参加的数据挖掘结果相近或完全一致
  - 特征提取，对于$d$ 个属性来说，具有$2^d$ 个可能的子集

# 数据压缩

- 数据压缩：应用数据编码或变换，以便得到数据的归约或压缩表示
  - 无损压缩：原数据可以由压缩数据重新构造而不丢失任何信息
    - 字符串压缩是典型的无损压缩
    - 现在已经有许多很好的方法但是它们只允许有限的数据操作
  - 有损压缩：只能重新构造原数据的近似表示
    - 影像文件的压缩是典型的有损压缩
    - 典型的方法：小波变换、主成分分析

# 数值压缩

- 数值归约：通过选择替代的、“较小”的数据表示形式来减少数据量
  - 有参的方法
    - 假设数据符合某些模型，通过评估模型参数，仅需要存储参数，不需要存储实际数据（孤立点也可能被存放）
    - 典型方法：对数线性模型，它估计离散的多维概率分布
  - 无参的方法
    - 不存在假想的模型
    - 典型方法:直方图、聚类和抽样

# 数据离散化

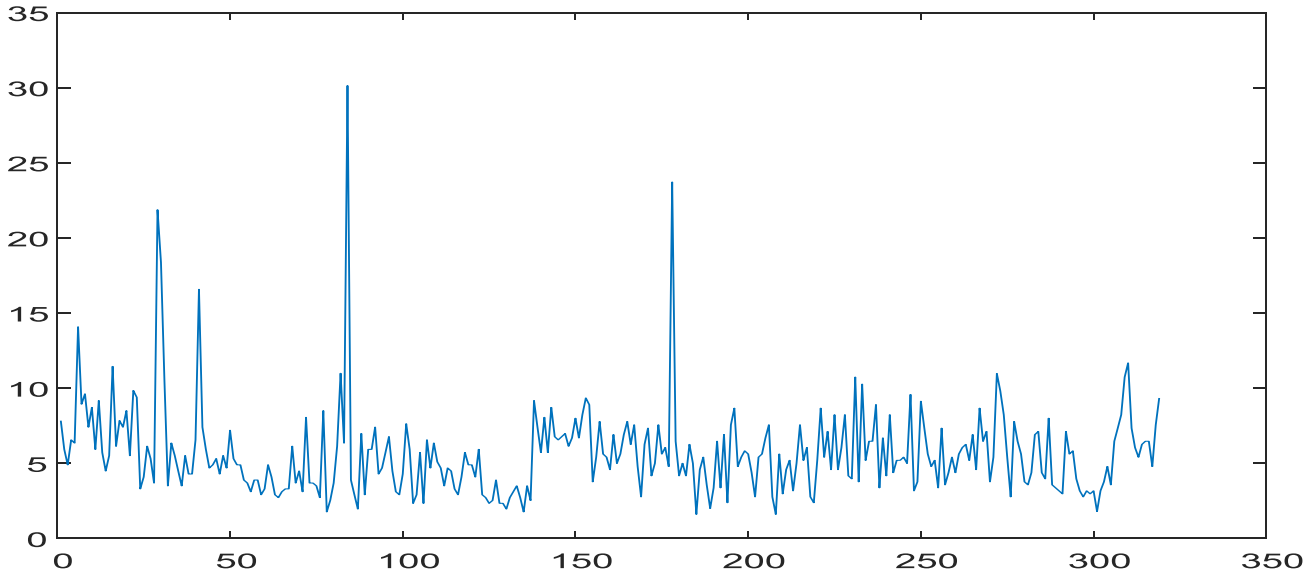
- 属性值分类
  - 枚举型
    - 有序的
    - 无序的
  - 连续型：如 Real类型
- 数据离散化
  - 对于一个特定的连续属性，可以把属性值划分成若干区间，以区间值来代替实际数据值，以减少属性值的个数。
- 概念层次
  - 利用高层的概念（如儿童、青年、中年、老年等）来代替低层的实际数据值（实际年龄），以减少属性值的个数。

# Matlab数据预处理

- 缺失值处理ismissing, fillmissing, fixunknowns
- 离群值处理isoutlier, filloutliers, movmean
- 去噪声smooth, smoothts, smoothdata
- 标准化zscore, mapstd,mapminmax
- 降维(规约)pca, processpca

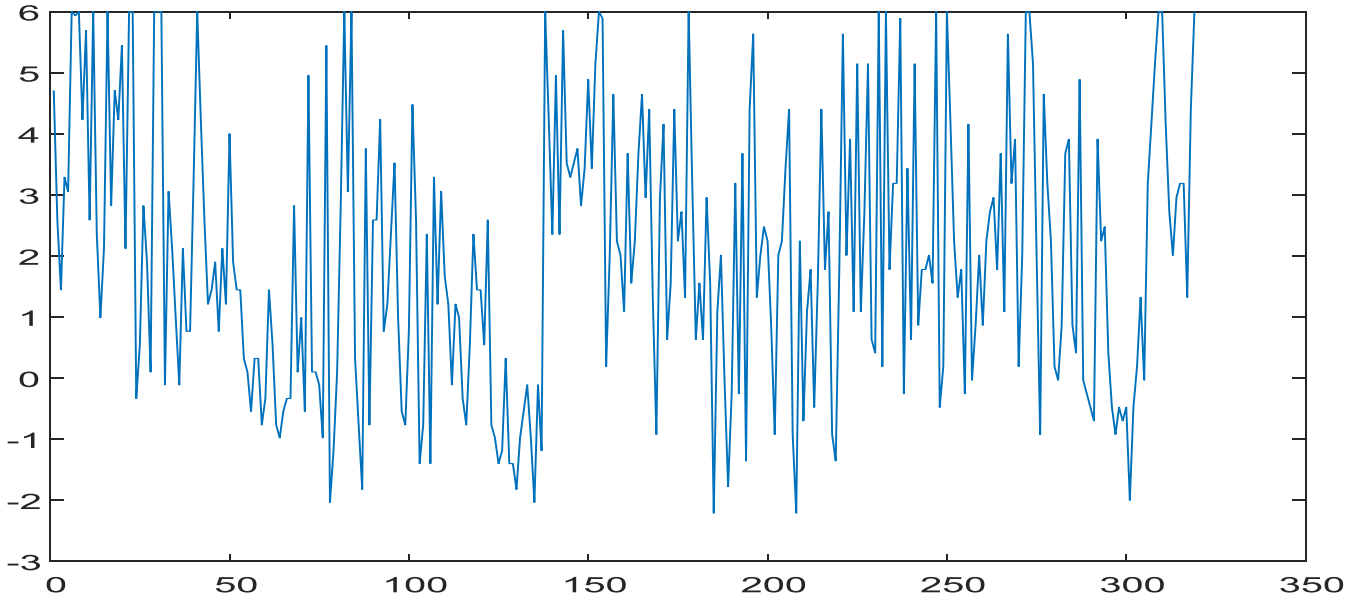
# 例1 重金属污染

- isoutlier(As) %找离群值
- filloutliers(As,‘nearest’,‘mean’)%用临近均值修正离群值(本题这样处理是错误的)



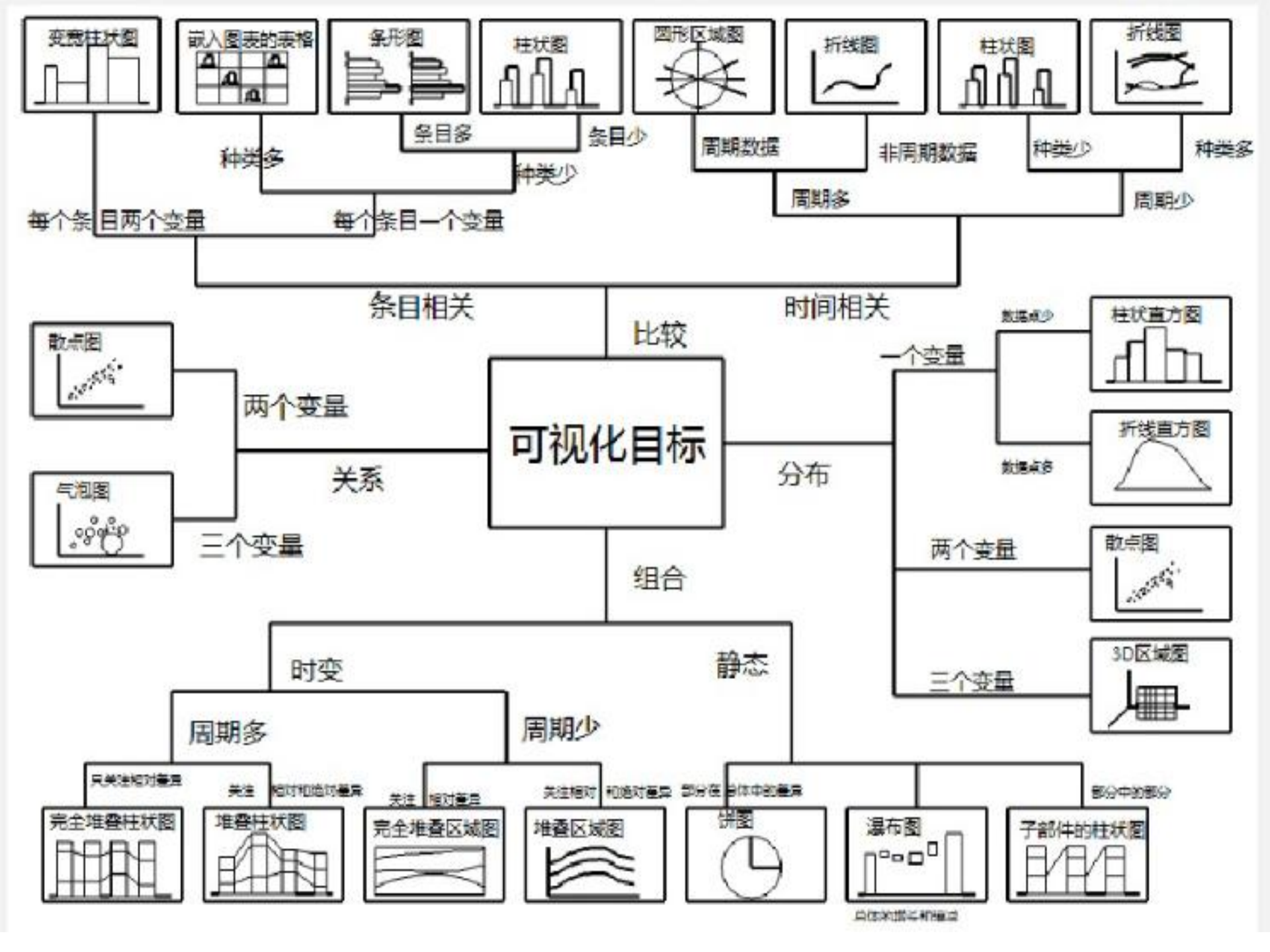
# 例1 重金属污染

- 本题利用背景值处理离群值或噪声值
- As2=max(-6, min(6, (As-3.6)/0.9))



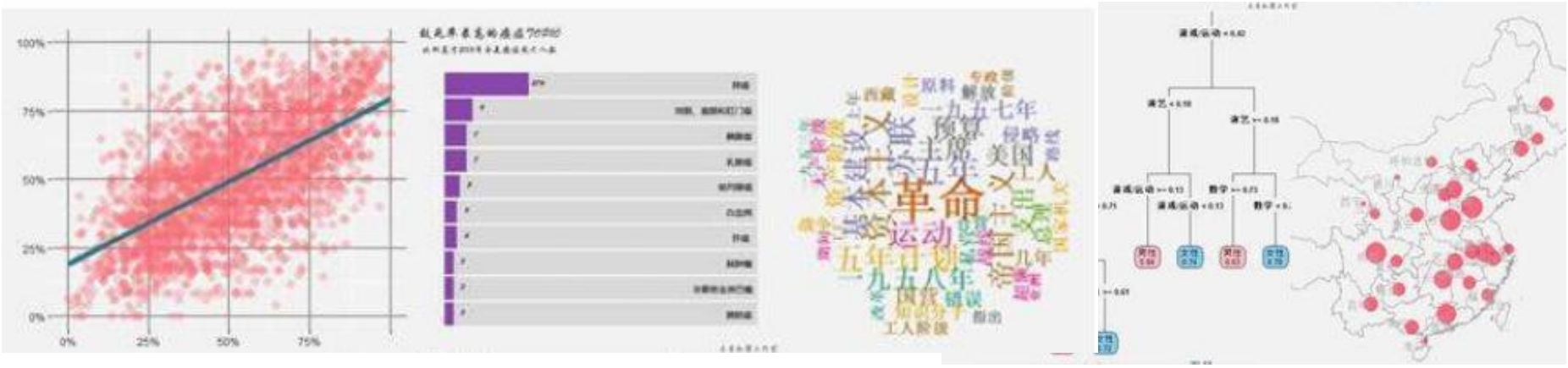
# 数据可视化

- 数据可视化: 旨在借助于图形化手段，清晰有效地传达与沟通信息。
  - 以下面两种方式观察数据：
    - 在不同的粒度或抽象层面观察
    - 属性或维度的不同结合
  - 数据可以被表示成不同的格式，柱状图、饼状图、散点图、三维立方体、曲线、数据分布图表、气泡图、热力图、词云等


可视化目标
比较
条目相关
时间相关
变宽柱状图
嵌入图表的表格
条形图
柱状图
圆形区域图
折线图
柱状图
折线图
条目多
条目少
种类多
种类少
每个条目两个变量
每个条目一个变量
周期数据
非周期数据
种类少
种类多
周期多
周期少
散点图
两个变量
气泡图
三个变量
关系
组合
分布
一个变量
数据点少
数据点多
柱状直方图
折线直方图
两个变量
散点图
三个变量
3D区域图
时变
静态
周期多
周期少
完全堆叠柱状图
堆叠柱状图
完全堆叠区域图
堆叠区域图
饼图
瀑布图
子部件的柱状图

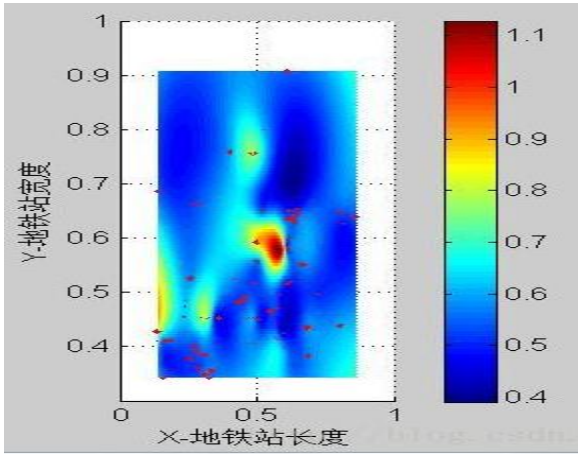
# 数据可视化

- 通用软件： R, Python, Matlab， Excel
- 可视化软件： Echarts, Tabulae, DataV, PowerBI
- 数据地图： Power map for excel，地图慧

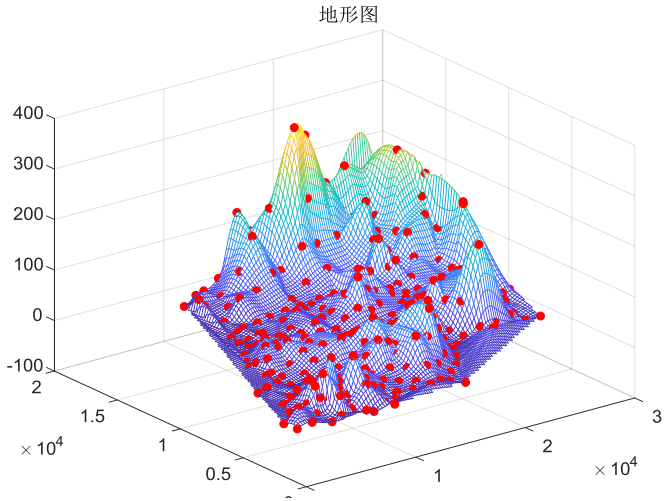


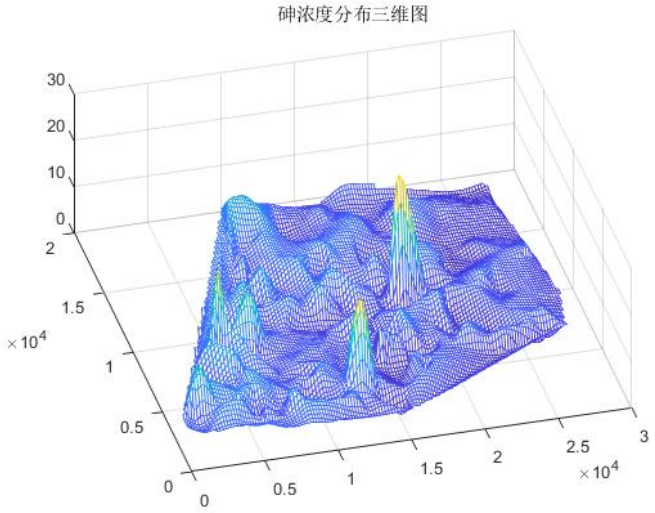
# Matlab可视化

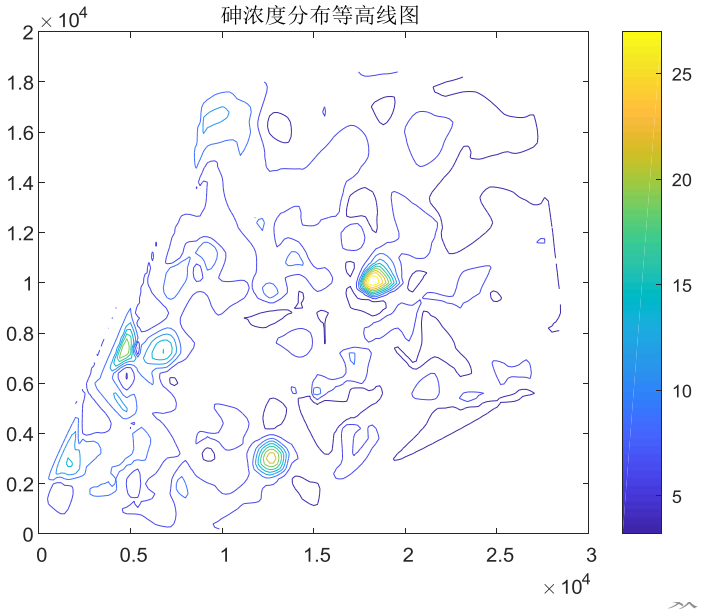
- 曲线图plot, fplot, plot3
- 曲面图meshgrid, ndgrid, mesh, surf
- 等高线图contour, contourf, contour3
- 柱图bar, bar3
- 饼图pie, pie3
- 散点图plot, scatter, scatter3, stem, stem3
- 直方图histogram
- 雷达图polarplot, polarhistogram
- 热力图heatmap
- 地理气泡图geobubble
- 动画comet, comet3
- 词云Wordcloud

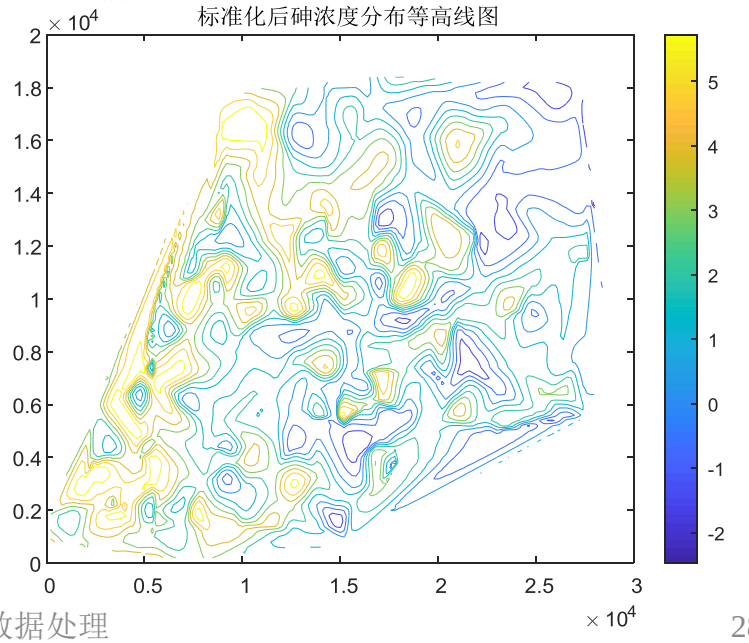



# 例1 重金属污染


地形图
400
300
200
100
0
-100
2
1.5
1
0.5
× 10^4
1
2
3
× 10^4


砷浓度分布三维图
30
20
10
0
2
1.5
1
0.5
0
× 10^4
0
0.5
1
1.5
2
2.5
3
× 10^4


砷浓度分布等高线图
× 10^4
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
2
2.5
3
× 10^4
25
20
15
10
5


标准化后砷浓度分布等高线图
× 10^4
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
2
2.5
3
× 10^4
5
4
3
2
1
0
-1
-2

# 数学建模方法的分类

- 机理分析方法
- 数据拟合方法
- 计算机模拟方法

数学建模区别于数据挖掘之处：比起结果的精度和算法的速度，建立揭示规律性的模型显得更重要。

# 数据建模方法

- 综合评价问题
  - 简单加权
  - 熵权法: Shannon熵
  - Topsis法: 接近理想解的偏好排序
  - 层次分析法: 成对比较阵
  - 模糊综合评判：模糊关系矩阵
  - 主成分分析: 多元统计

# 数据建模方法

- 预测问题
  - 差分方程或微分方程：机理模型
  - 插值（外推）或拟合：计算数学
  - 回归分析：线性回归，Lasso回归，岭回归
  - 时间序列：Hot-Winters, ARIMA, GARCH
  - 灰色模型：GM(1,1)，适用少数据且噪音小
  - 马氏链: 多值概率分布
  - 神经网络：非线性模型

# 数据建模方法

- 分类问题
  - 判别分析(有监督学习)
    - KNN：k个近邻
    - Logistic回归：因变量为0-1
    - 朴素Bayes，Fisher判别
    - 神经网络：非线性
    - 决策树：离散化, C4.5, CART
    - 支持向量机: 超平面上的间隔
  - 聚类分析（无监督学习）
    - 层次聚类BIRCH
    - 划分聚类K-means
    - 密度聚类DBSCAN
    - 模糊聚类FCM

# 数据建模方法

- 最优化问题
  - 数学规划：线性规划、非线性规划、整数规划、0-1整数规划、目标规划等
  - 启发式算法：贪心算法、随机搜索、禁忌搜索、遗传算法、蚁群算法、模拟退火等

# 数据建模常见的误区

- 滥用层次分析法
- 滥用灰色预测
- 滥用高次函数拟合
- 各种过拟合
- 用建模数据做检验
- 过度依赖现成软件包

# Matlab 数据拟合

- interp1 - 一维数据插值
- interp2 - 二维数据插值
- interp3 - 三维数据插值
- interpn - n维数据插值
- spline - 样条插值
- caspe - 样条插值
- griddata - 散乱数据插值（2-3维）

- polyfit - 多项式拟合
- lsqnonlin - 最小二乘法
- lsqlin - 约束线性拟合
- lsqcurvefit - 曲线（面）拟合
- casps - 样条拟合
- 应用工具Curve Fiting -二元拟合
- 应用工具Neural Net Fiting 神经网络拟合

# Matlab统计分析

- 极大似然估计mle
- 正态性检验kstest
- 分布拟合度检验chi2gof
- 相关分析corrcoef, canoncorr
- 主成分分析pca
- 因子分析factoran
- 判别分析classify, fitcdiscr
- 线性回归regress
- 非线性回归nlinfit
- 逐步回归stepwisefit
- 时间序列分析模型arima (估计estimate，预测forcast)
- GARCH模型garch

# Matlab统计学习

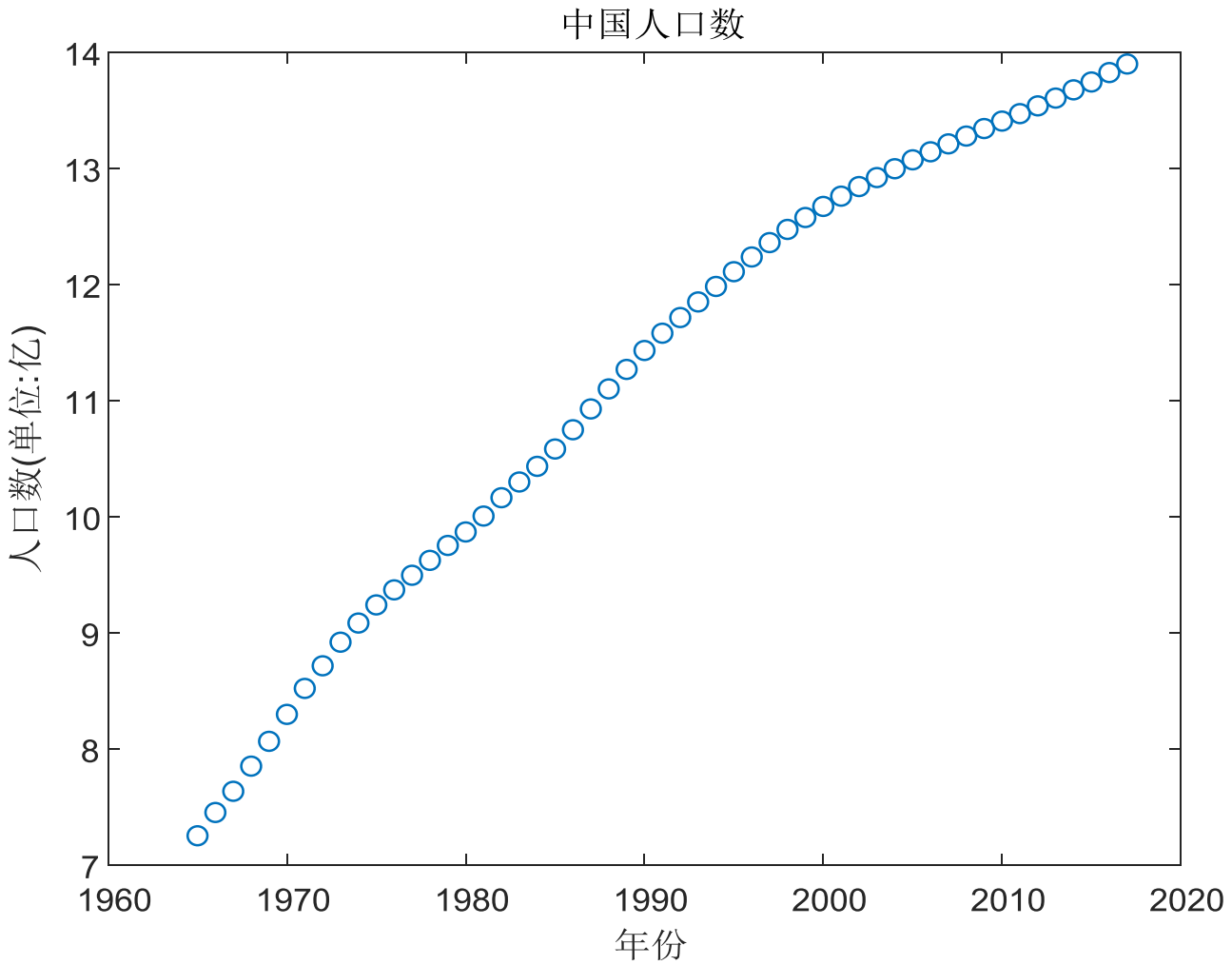
- 多变量输出的线性回归mvregress
- Lasso回归lasso
- 岭回归ridge
- 支持向量机分类svmclassify
- K-means聚类kmeans
- 系统聚类clusterdata
- 决策树fitctree
- 回归树fitrtree
- 朴素贝叶斯fitcnb
- k-近邻fitcknn
- 隐马尔可夫hmmtrain
- 集成学习fitensemble
- 应用工具Neural Net clusting神经网络聚类
- 应用工具Neural Net Time Seires神经网络时间序列

# 例2：中国人口预测

问题：下表中国人口数据(单位：千人，来自国家统计局网站)，试建立数学模型预测中国人口增长。

| 年份 | 人数 | 年份 | 人数 | 年份 | 人数 | 年份 | 人数 |
|---|---|---|---|---|---|---|---|
| 1965年 | 72538 | 1980年 | 98705 | 1995年 | 121121 | 2010年 | 134091 |
| 1966年 | 74542 | 1981年 | 100072 | 1996年 | 122389 | 2011年 | 134735 |
| 1967年 | 76368 | 1982年 | 101654 | 1997年 | 123626 | 2012年 | 135404 |
| 1968年 | 78534 | 1983年 | 103008 | 1998年 | 124761 | 2013年 | 136072 |
| 1969年 | 80671 | 1984年 | 104357 | 1999年 | 125786 | 2014年 | 136782 |
| 1970年 | 82992 | 1985年 | 105851 | 2000年 | 126743 | 2015年 | 137462 |
| 1971年 | 85229 | 1986年 | 107507 | 2001年 | 127627 | 2016年 | 138271 |
| 1972年 | 87177 | 1987年 | 109300 | 2002年 | 128453 | 2017年 | 139008 |
| 1973年 | 89211 | 1988年 | 111026 | 2003年 | 129227 | | |
| 1974年 | 90859 | 1989年 | 112704 | 2004年 | 129988 | | |
| 1975年 | 92420 | 1990年 | 114333 | 2005年 | 130756 | | |
| 1976年 | 93717 | 1991年 | 115823 | 2006年 | 131448 | | |
| 1977年 | 94974 | 1992年 | 117171 | 2007年 | 132129 | | |
| 1978年 | 96259 | 1993年 | 118517 | 2008年 | 132802 | | |
| 1979年 | 97542 | 1994年 | 119850 | 2009年 | 133450 | | |

# 数据图示


中国人口数
人口数(单位:亿)
年份
14
13
12
11
10
9
8
7
1960
1970
1980
1990
2000
2010
2020

# 建模方法

- 拟合效果挺好，但讲不出道理的方法
  - 多项式拟合
  - 灰色预测
  - 神经网络
  - 。。。。
- 讲得出道理的方法（机理分析）
  - 差分方程
  - 微分方程

# 模型1：Malthus指数增长模型

考虑人口增长是一个随时间连续变化的过程，
假设每年增长率 $r$是常数，

$x(t)$ ~时刻$t$的人口, $[t, t+\Delta t]$ 人口的增量

$$x(t+\Delta t)-x(t)=rx(t)\Delta t,$$

即 $$\frac{x(t+\Delta t)-x(t)}{\Delta t}=rx(t),$$

令$\Delta t\to 0$，得微分方程模型 $$\frac{dx}{dt}=rx$$

设起始人口 数$x_0$， 求解得到指数增长模型

$$x(t)=x_0e^{rt}$$

# 最小二乘拟合

- 数据($t_0,x_0$),($t_1,x_1$), …($t_n,x_n$)拟合函数 $x=f(t;c_0, \ldots,c_m)$, 其中$c_0, \ldots,c_m$为未知参数, $m<n$. 求$c_0, \ldots,c_m$使得误差平方和取得最小值。

$$\sum_{i=0}^{n}[x_i - f(t_i; c_0, \cdots, c_m)]^2$$

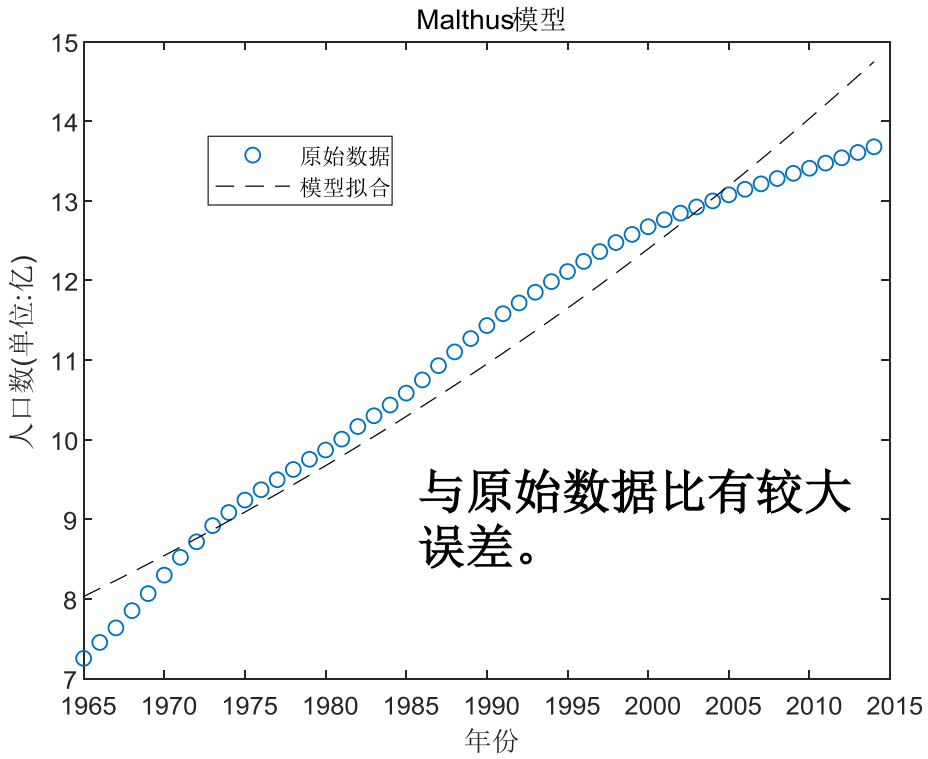
# 参数$x_0$, $r$的估计

- 指数增长模型：$x(t) = x_0 e^{rt}$
- 线性化拟合：变换为$\ln x = \ln x_0 + rt$，这样$\ln x$与$t$是线性函数关系，用线性拟合求得$\ln x_0$和$r$。

**Matlab线性化拟合**

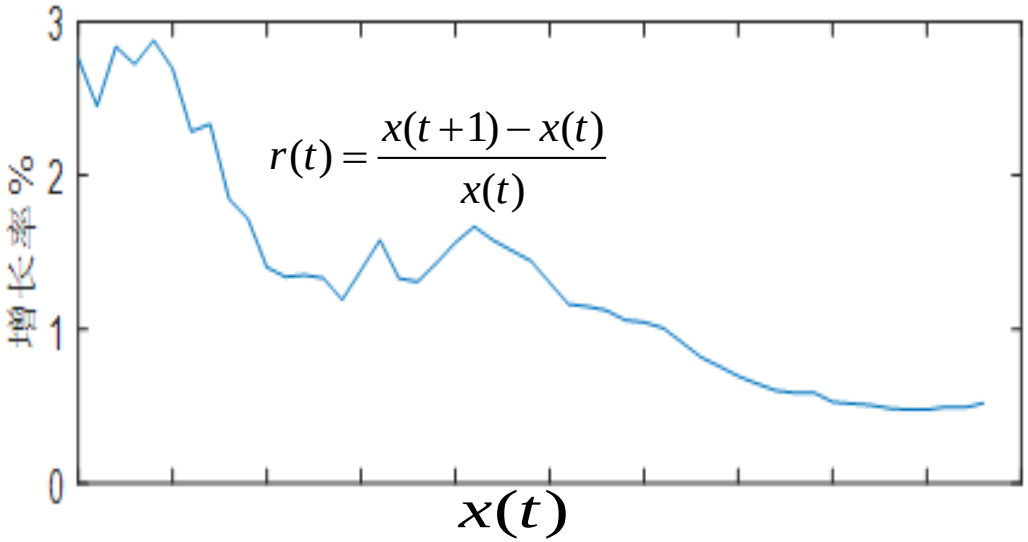
**$x_0$=8.0310,**

**$r$= 0.0124.**



# 模型2：阻滞增长模型(Logistic)

•仔细分析发现：人口增长率*r*不是常数(逐渐下降)




$$r(x)=r(1-\frac{x}{x_m})$$

$$\frac{dx}{dt}=r(x)x$$

$$\frac{dx}{dt}=rx(1-\frac{x}{x_m})$$

$$x(t)=\frac{x_m}{1+(\frac{x_m}{x_0}-1)e^{-rt}}$$

# 最小二乘拟合

- 数据($t_0$,$x_0$),($t_1$,$x_1$), …($t_n$,$x_n$)拟合函数 $x=f(t;c_0, \ldots,c_m)$, 其中$c_0, \ldots,c_m$为未知参数, $m<n$. 求$c_0, \ldots,c_m$使得误差平方和取得最小值。

$$\Sigma_{i=0}^{n}[x_i - f(t_i; c_0, \cdots, c_m)]^2$$

平均拟合误差

$$e = \sqrt{\frac{1}{n}\Sigma_{i=0}^{n}[x_i - f(t_i; c_0, \cdots, c_m)]^2}$$

# 参数$x_0$, $r$, $x_m$的估计

- Logistic增长模型：

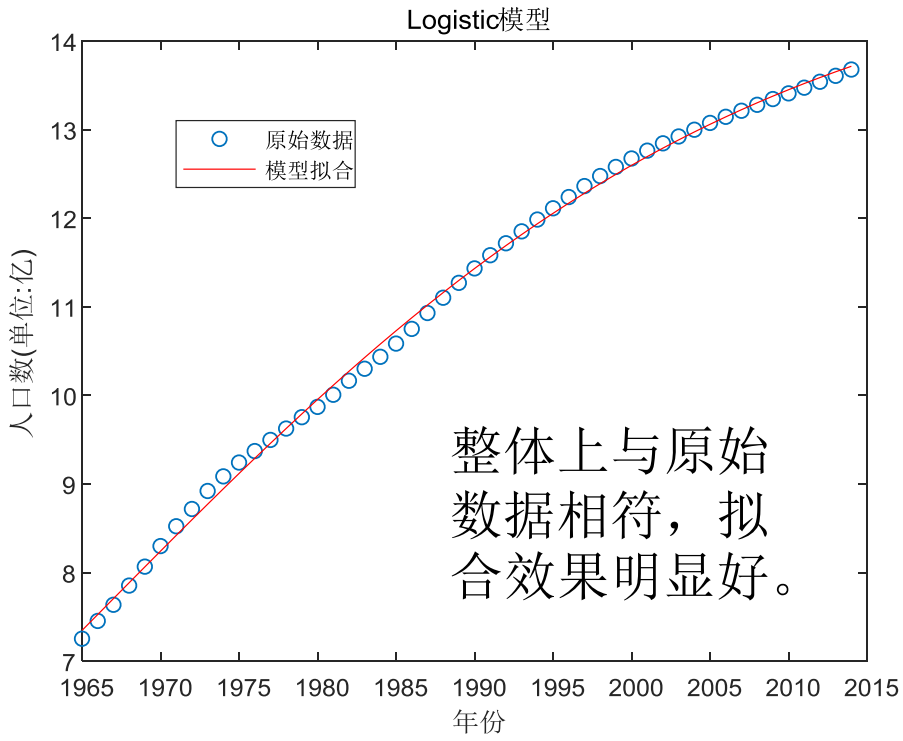
$$x(t)=\frac{x_m}{1+(\frac{x_m}{x_0}-1)e^{-rt}}$$

**Matlab非线性拟合**

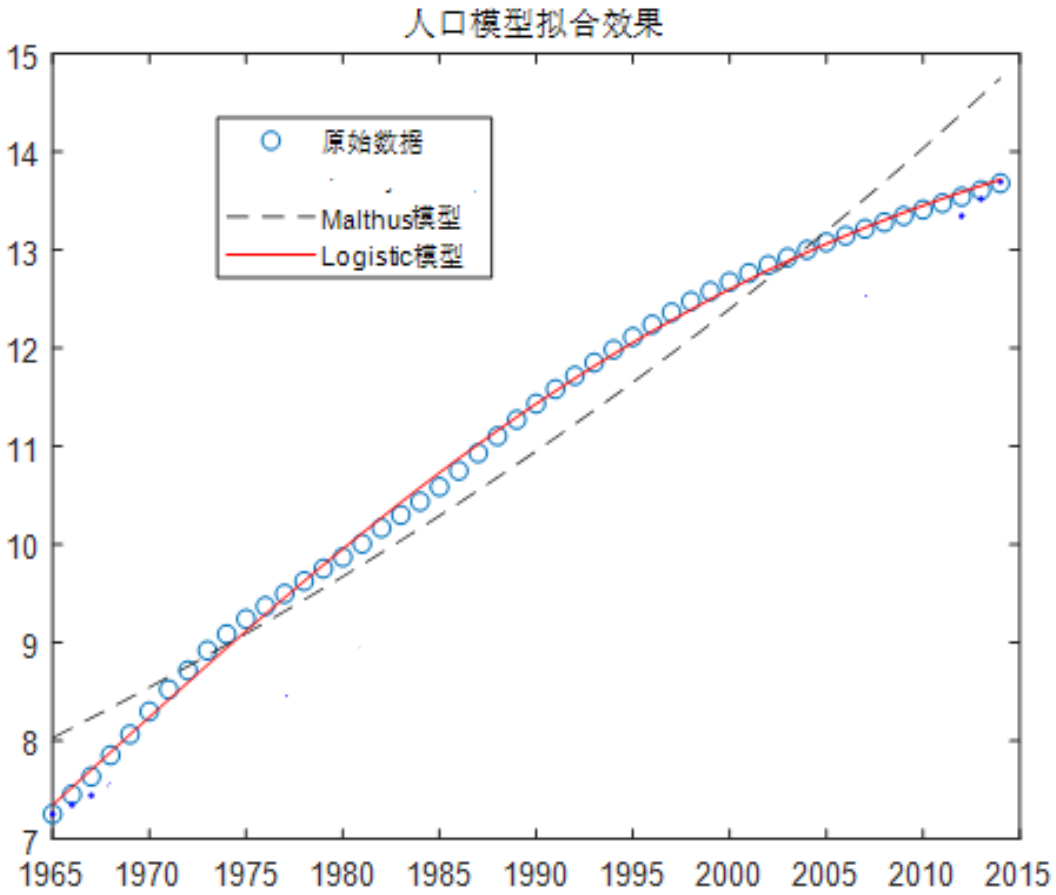
**$x_0$ =7.3475，**

**r = 0.0474，**

**$x_m$=15.1436。**



# 拟合效果比较




| 模型 | 平均拟合误差（亿人） |
|---|---|
| Malthus | 0.4362 |
| Logistic | 0.0763 |

Logistic更好。

# 模型检验

表 2015-2017年模型预测值(单位：亿人)

| 模型 | 2015年 | 2016年 | 2017年 | 误差 |
|---|---|---|---|---|
| 真实值 | 13.7462 | 13.8271 | 13.9008 | |
| Malthus模型 | 14.9319 | 15.1183 | 15.3069 | 1.2975 |
| Logistic模型 | 13.7752 | 13.8331 | 13.8887 | 0.0185 |

可见，Logistic预测效果更好。

注意：预测模型检验时，不能用建模中已使用的数据。

# 模型应用

- 加入2015-2017三年数据，重新估计Logistic模型参数(有微小变化，说明模型稳健性好)

  $x_0$ =7.3468 ， r = 0.0474，$x_m$= 15.1370 。

- 2018年至2025年的预测值(单位：亿人)依次为：

  13.9396，13.9908， 14.0401，14.0874，
  14.1327， 14.1763， 14.2181，14.2582。

# 例3 销售量预测

- 例：2009~2012四年销售量的季度数据。

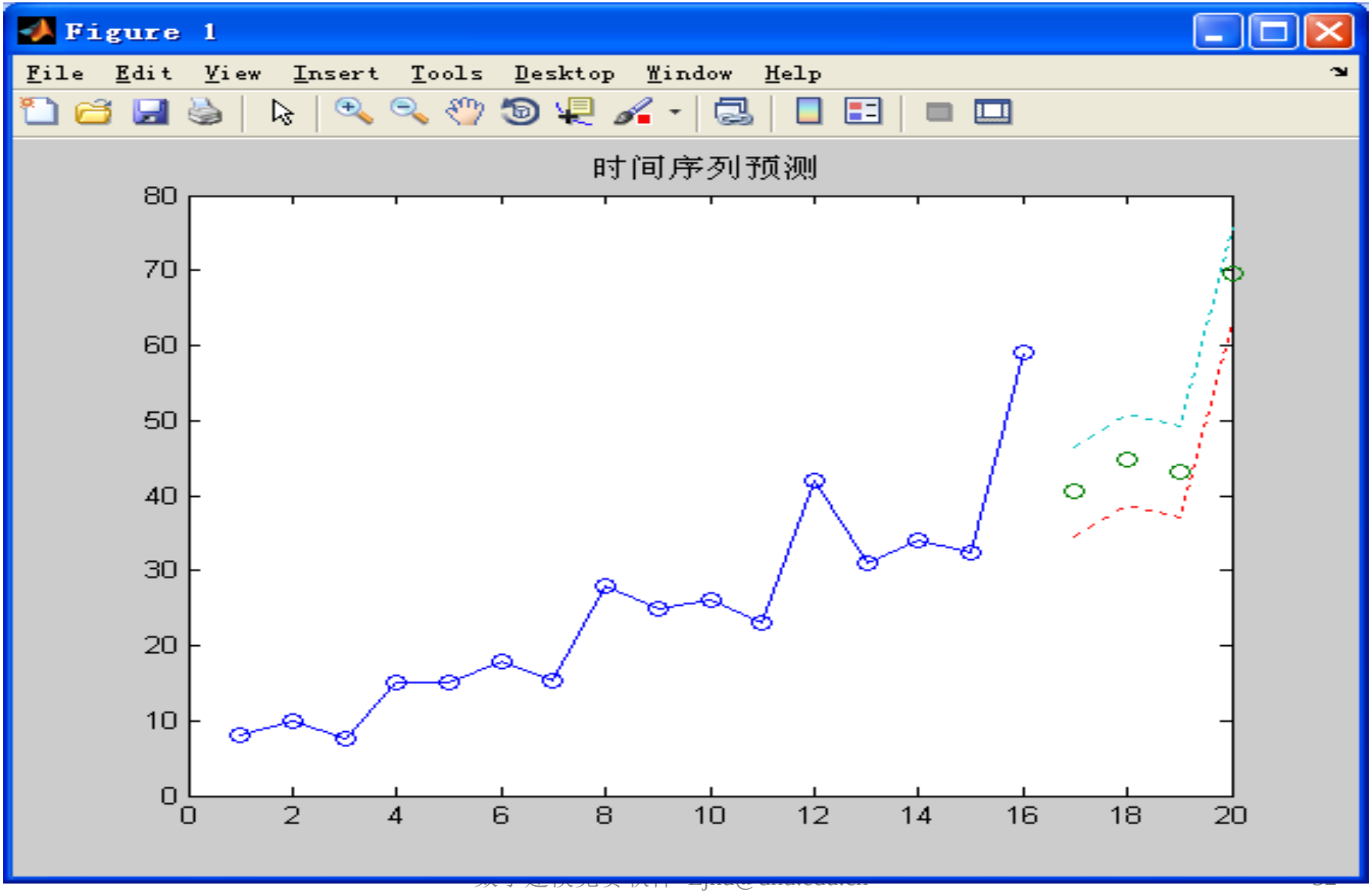
| | | | |
|---|---|---|---|
| 8 | 10 | 7.7 | 15 |
| 15 | 18 | 15.3 | 28 |
| 25 | 26 | 23 | 42 |
| 31 | 34 | 32.5 | 59 |

- 据此预测2013年度各季度销售量。

- 本题无法建立机理模型，故采用数据建模。如时间序列、神经网络、灰色预测等。

```
%时间序列预测
clear;close all;
y=[8,10,7.7,15,15,18,15.3,28,25,26,23,42,31,34,32.5,59]';
model =
arima('Constant',0,'D',1,'ARLags',1,'MALags',1,'Seasonality',4)
fit1 = estimate(model,y)
[yf,ymse] = forecast(fit1,4,'Y0',y)
upper = yf + 1.96*sqrt(ymse)
lower = yf - 1.96*sqrt(ymse)
plot(1:16,y,'-o',17:20,yf,'o',17:20,lower,':',17:20,upper,':');
title('时间序列预测')
```

# 预测结果：41，45，43，70



# 随机模拟

**Monte Carlo原理**：设$\xi$是一个分布已知的随机变量，为了求取$\eta = f(\xi)$的概率分布或数字特征，生成$N$个($N$足够大)服从$\xi$的分布的随机数$x_1$，$x_2$，…，$x_N$，令$y_i = f(x_i)$, $i=1,2,\ldots,N$, 那么

$$P\{\eta \in A\} \approx \frac{N_A}{N} = \frac{1}{N}\sum_{i=1}^{N} 1_{y_i \in A}$$

$$E[\eta] \approx \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$D[\eta] \approx S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2$$

# 随机模拟(Matlab)

- rand - [0，1]区间均匀分布随机数
- randn - 标准正态分布随机数
- randperm - 1...n 随机排列
- random -各种分布随机数
- normrnd – 一般正态分布随机数
- normpdf –正态分布概率密度函数
- normcdf –正态分布分布函数
- norminv –正态分布逆分布函数(分位数)
- .....均匀分布，二项分布， 泊松分布等

# 例4 零件参数设计

- 1997年建模竞赛
- $x_i$正态分布已知，i=1,2,...7

$$y = 174.42\left(\frac{x_1}{x_5}\right)\left(\frac{x_3}{x_2 - x_1}\right)^{0.85} \times \sqrt{\frac{1 - 2.62\left[1 - 0.36\left(\frac{x_4}{x_2}\right)^{-0.56}\right]^{\frac{3}{2}}\left(\frac{x_4}{x_2}\right)^{1.16}}{x_6 x_7}}$$

- 损失函数

$$z = \begin{cases} 9000 & |y - 1.5| > 0.3 \\ 1000 & 0.3 \geq |y - 1.5| > 0.1 \\ 0 & |y - 1.5| \leq 0.1 \end{cases}$$

- 求平均损失（程序jm1997B.m）

# 例5 自动化车床管理

大意：自动化车床连续加工某种零件, 工作人员通过检查零件来确定工序是否出现故障。现积累有100次刀具故障记录，故障出现时该刀具完成的零件数459, 362, 624,….（略）。现计划在刀具加工一定件数后定期更换新刀具。

故障时产出的零件损失费用 f=200元/件；进行检查的费用 t=10元/次； 发现故障进行调节使恢复正常的平均费用 d=3000元/次(包括刀具费)；未发现故障时更换一把新刀具的费用 k=1000元/次。

假定工序故障时产出的零件均为不合格品，正常时产出的零件均为合格品, 试对该工序设计效益最好的检查间隔（生产多少零件检查一次）和刀具更换策略。

参考论文：自动化车床

主模型：闭式解+数值积分+最优化

模拟验证：随机模拟+枚举

# 资料下载

https://pan.baidu.com/s/1aPSj2rlDd3Qi5TPSqy5GwQ

提取码：9o23