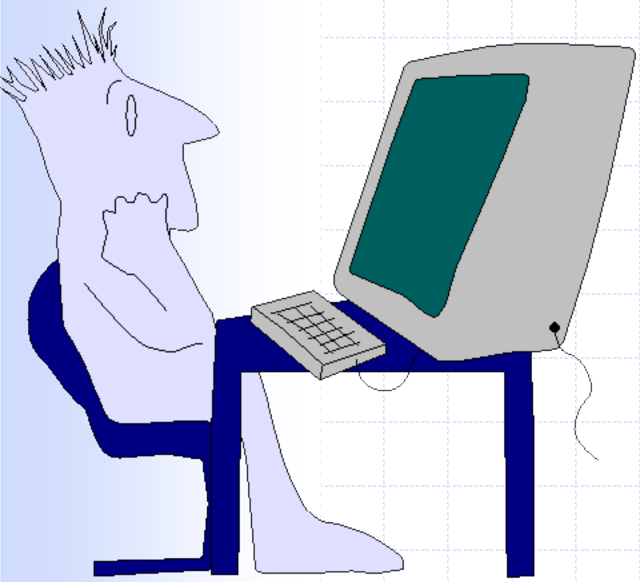
# 数学试验

# Experiments in Mathematics

赵小艳

**办公地址：理科楼214**



# 试验13 人口预测与数据拟合

实验目

1、学会用MATLAB软件进行数据拟合。

2、了解利用最小二乘法进行数据拟合基本思想，掌握用数据拟正当寻找最正确拟合曲线方法。

3、了解多元函数极值在数据拟正当中应用。

4、经过对实际问题进行分析研究，初步掌握建立数据拟合数学模型方法。

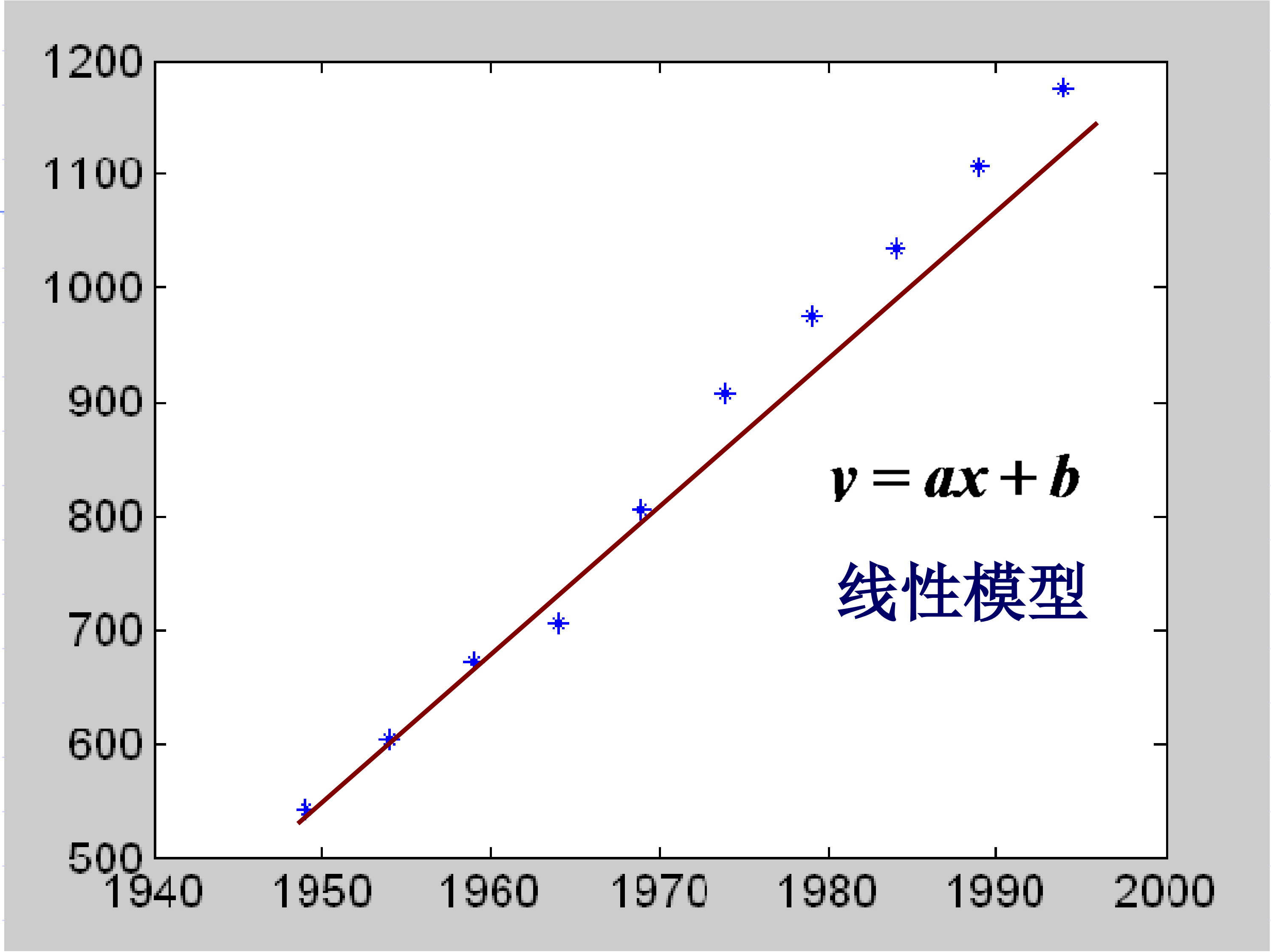
# 试验问题

据人口统计年鉴，知我国从1949年至1994年人口数据资料以下：（人口数单位为：百万）

| 年份 | 1949 | 1954 | 1959 | 1964 | 1969 |
|---|---|---|---|---|---|
| 人口数 | 541.67 | 602.66 | 672.09 | 704.99 | 806.71 |
| 年份 | 1974 | 1979 | 1984 | 1989 | 1994 |
| 人口数 | 908.59 | 975.42 | 1034.75 | 1106.76 | 1176.74 |

（1）在直角坐标系上作出人口数图象。

（2）建立人口数与年份函数关系，并估算1999年人口数。



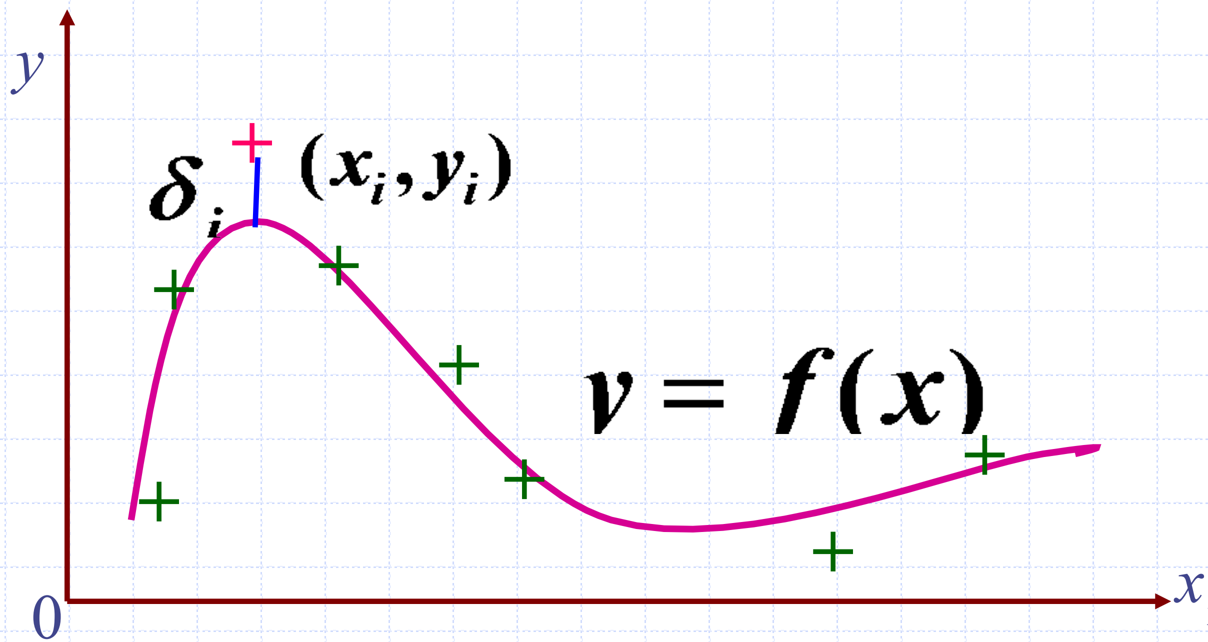
1200
1100
1000
900
800
700
600
500
1940
1950
1960
1970
1980
1990
2000
$y = ax + b$
线性模型


# 怎样确定*a,b*?

# 一、曲线拟合

## 1 曲线拟合问题提法:

已知一组（二维）数据，即平面上 n 个点$(x_i, y_i)$，$i = 1, 2, \cdots, n$，$x_i$ 互不相同，寻求一个函数（曲线）$y = f(x)$，使 $f(x)$ 在观察点x处所取得值f(x)分别与观察值y在某种准则下最为靠近，即曲线拟合得最好，如图



从几何上讲，并不要求曲线严格经过已知点，但要求曲线在各数据点和已知数据点之间总体误差最小，通常称为数据拟合。

而我们经常是确定f(x)使得偏差平方和

$$\sum_{i=1}^{n}\delta_i^2=\sum_{i=1}^{n}[f(x_i)-y_i]^2$$ 到达最小。

$\longrightarrow$ 最小二乘准则

# 数据插值

已知一组（二维）数据，即平面上 n 个点$(x_i, y_i)$，$i=1,2,\cdots,n$，$x_i$互不相同，寻求一个函数（曲线）$y=f(x)$

使$f(x)$在观察点$x_1,\cdots,x_n$处满足$f(x_i)=y_i, i=1,\cdots,n$,

$\longrightarrow$ 数据插值

# 2．用什么样曲线拟合已知数据?

1）画图观察

2）理论分析

$$f(x)=a_1r_1(x)+a_2r_2(x)+\cdots+a_mr_m(x)$$

**惯用曲线函数系ri(x)类型：**

直线： $y=a_1x+a_0$

多项式： $y=a_mx^m+\cdots+a_1x+a_0$

双曲线（一支）： $y=\dfrac{a_1}{x}+a_2$

指数曲线： $y=a_1e^{a_2x}$

# 比如

$$f(x) = a_1 x^m + a_2 e^{\lambda x} + a_3 \sin(x)$$

$1, x, x^2, \mathrm{L}\ \ x^m$ 多项式拟合

$e^{\lambda_1 x}, \mathrm{L}\ \ e^{\lambda_m x}$ 指数函数拟合

$\sin(x), \cos(x), \sin(2x),$
$\cos(2x), \mathrm{L}\ \ \sin(mx), \cos(mx)$ 三角函数拟合

## 3 拟合函数组中系数确定

以$f(x)=a_1x+a_2$为例，即确定$a_1$，$a_2$使得

$$J(a_1,a_2)=\sum_{i=1}^{n}\delta_i^2=\sum_{i=1}^{n}[a_1x_i+a_2)-y_i]^2$$达到最小。

为此，只需利用极值的必要条件$\dfrac{\partial J}{\partial a_k}=0(k=1,2)$

$a_1,a_2$

$$\begin{cases}\sum_{i=1}^{n}2[a_1x_i+a_2)-y_i]x_i=0\\ \sum_{i=1}^{n}2[a_1x_i+a_2)-y_i]=0\end{cases}\Rightarrow a_1,a_2$$

# 4 用matlab软件进行数据拟合

## （1）lsqcurvefit命令-----最小二乘拟合

**a= lsqcurvefit(fun,x0,xdata,ydata)**

**[a,resnorm]=lsqcurvefit(fun,x0,xdata,ydata)**

**是依据给定数据xdata,ydata，按照函数文件fun给定函数，以x0为初值做最小二乘拟合，返回函数中系数向量a和残差平方和resnorm。**

# 例

首先编写函数文件

```
function y=f(a,x)
f=a(1)*exp(x)+a(2)*x.^2+a(3)*x.^3
```

保留为f.m，其次调用该函数

```
x=0:0.1:1;
y=[3.1,3.27,3.81,4.5,5.18,6,7.05,8.56,9.69,11.25,13.17];
a0=[0 0 0];
[x,resnorm]=lsqcurvefit(@f,a0,x,y)
```

# 也能够用inline命令定义函数

```
x=0:0.1:1;
y=[3.1,3.27,3.81,4.5,5.18,6,7.05,8.56,9.69,11.25,13.17];
f=inline('a(1)*exp(x)+a(2)*x.^2+a(3)*x.^3','a','x');
a0=[0 0 0];
[a,resnorm]=lsqcurvefit(f,a0,x,y)
plot(x,y,'*')
hold on
g=a(1)*exp(x)+a(2)*x.^2+a(3)*x.^3;
plot(x,g,'r-')
```

## （2） polyfit命令——多项式曲线拟合

**a = polyfit(xdata,ydata,n)**

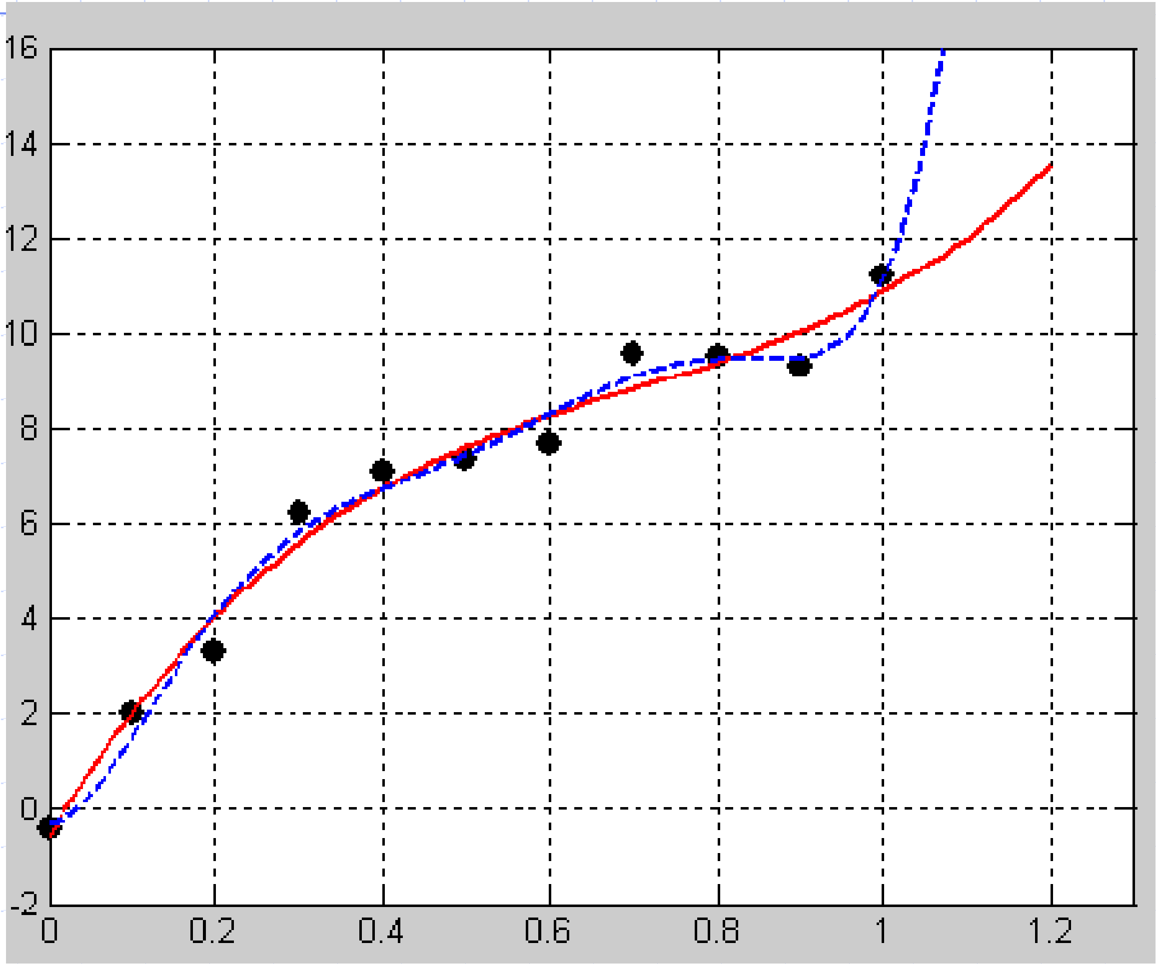
- 其中n表示多项式最高阶数
- **xdata**，**ydata** 为要拟合数据，它是用向量方式输入。
- 输出参数$a$为拟合多项式 $y = a_n x^n + \ldots + a_1 x + a_0$系数$a$ = [$a_n$, …, $a_1$, $a_0$]。
- 多项式在x处值y可用下面程序计算。

**y = polyval (*a*, x)**

因为高次多项式曲线改变不稳定，所以多项式次数选取不宜过高。

# 比如

```
clear;clc;
x=0:0.1:1;
y=[-0.447 1.978 3.28 6.16 7.08 7.34 7.66,9.56,9.48,9.3,11.2];
plot(x,y,'k.','markersize',25);
axis([0 1.3 -2 16]);
p3=polyfit(x,y,3)
p6=polyfit(x,y,6)
t=0:0.01:1.2;
s=polyval(p3,t);
s1=polyval(p6,t);
hold on
plot(t,s,'r-','linewidth',2);
plot(t,s1,'b--','linewidth',2);
grid
```



16
14
12
10
8
6
4
2
0
-2
0
0.2
0.4
0.6
0.8
1
1.2

# 二、人口预测线性模型

## 编写程序调用matlab命令

```
x=1949:5:1994;
y=[541.67,602.66,672.09,704.99,806.71,908.59,975.42,1034.75,1106.76,1176.74];
plot(x,y,'r*','linewidth',2)
grid
f=inline('a(1)+a(2)*x','a','x');
a0=[0 5];
[a,resnorm]=lsqcurvefit(f,a0,x,y)
hold on
g=a(1)+a(2)*x;
plot(x,g,'b-','linewidth',2)
```