



信阳师范学院  
数学与统计学院  
SCHOOL OF MATHEMATICS AND STATISTICS

## 第3章 数据拟合与数据插值



讲授人：牛言涛



日期：2020年2月20日

# 目录

## CONTENTS



多项式的表示与运算



数据拟合



一维数据插值



高维数据插值



## 3.2 数据拟合

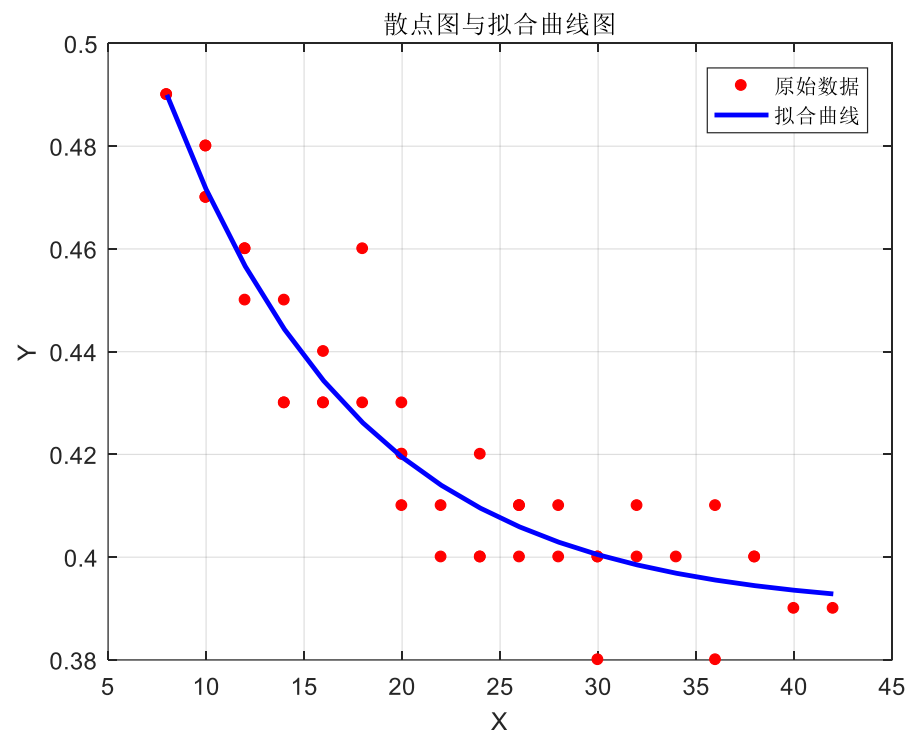


- 对于可控变量 $x$ 和随机变量 $y$ 的 $m$  ( $m > n$ )次独立的观测 $(x_i, y_i), i = 1, 2, \dots, m$ ,  $y$ (响应变量)和 $x$  (自变量)之间的多项式回归模型 (曲线拟合curve fitting) 为:

$$\begin{cases} y_i = p_1 x_i^n + p_2 x_i^{n-1} + \dots + p_n x_i + p_{n+1} + \varepsilon_i, \\ \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2), i = 1, 2, \dots, m. \end{cases}$$

- 曲线拟合是工程中经常要用到的技术之一。函数 `polyfit` 给出在最小二乘意义下最佳拟合系数。该函数的调用格式为:  $[P, S, mu] = \text{polyfit}(x, y, n)$

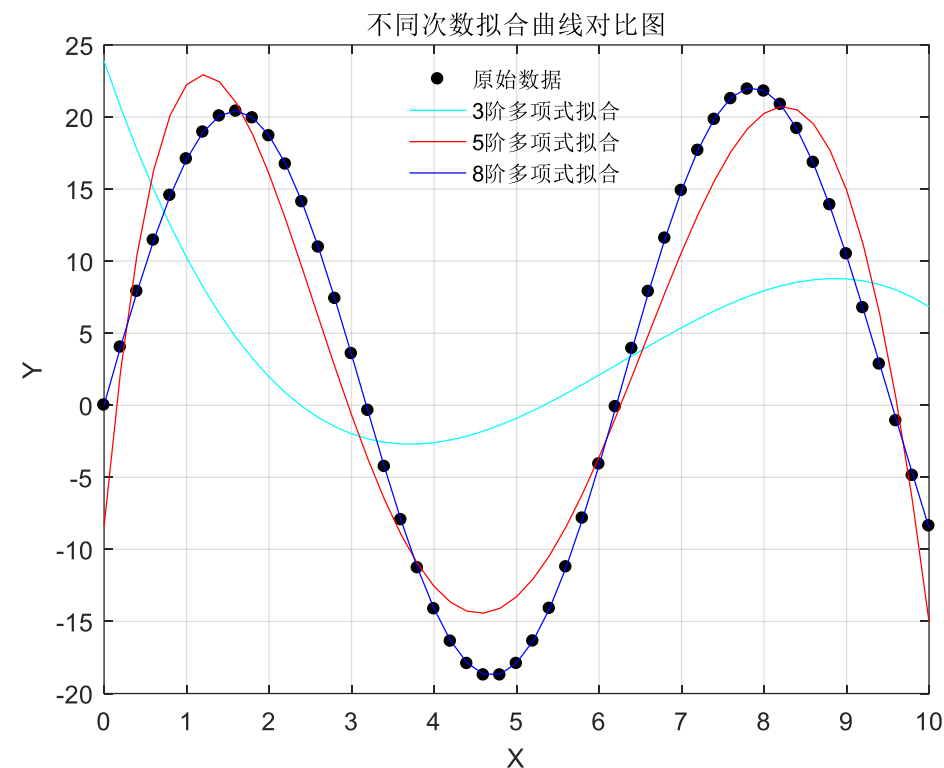
说明: 函数根据采样点 $x$ 和采样点 $y$ , 产生一个 $n$ 次多项式 $P$ 及其在采样点的误差向量 $S$ 。其中,  $x$ 、 $y$ 是两个等长的向量,  $P$ 是一个长度为 $m + 1$ 的向量。 $mu(1)$  是  $\text{mean}(x)$ ,  $mu(2)$  是  $\text{std}(x)$ 。



# 1. 拟合函数示例

- 例：用3、5和8阶多项式拟合函数  $f(x) = 0.25x + 20\sin(x)$

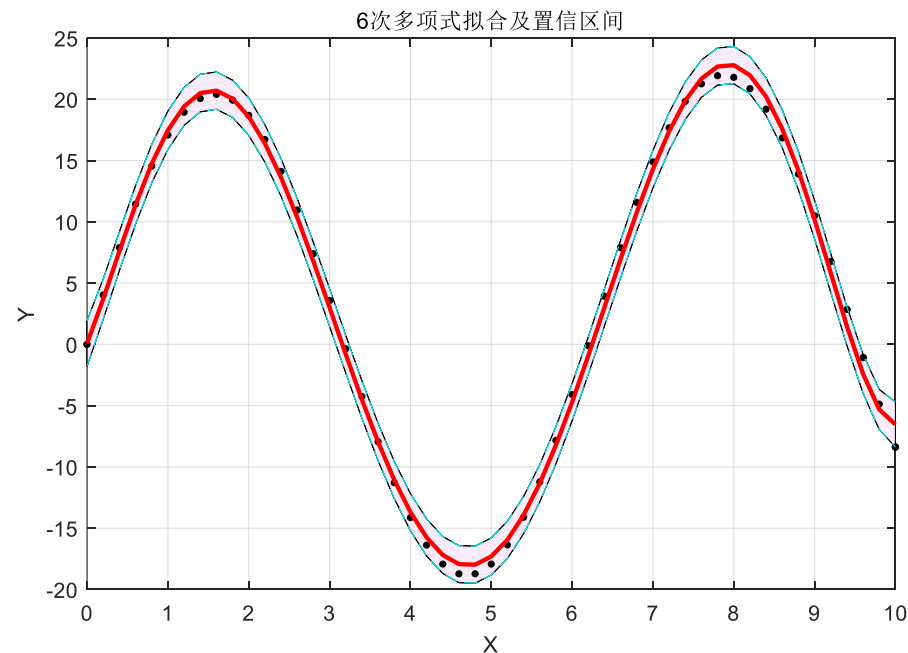
```
x=0:0.2:10; y=0.25*x+20*sin(x);  
plot(x,y,'k.','MarkerSize',15)  
grid on; hold on  
[p1,s1,mu1]=polyfit(x,y,3); %3阶多项式拟合  
y1=polyval(p1,x,s1,mu1);  
[p2,s2,mu2]=polyfit(x,y,5); %5阶多项式拟合  
y2=polyval(p2,x,s2,mu2);  
[p3,s3,mu3]=polyfit(x,y,8); %8阶多项式拟合  
y3=polyval(p3,x,s3,mu3);  
plot(x,y1,'c-', x,y2, 'r-', x,y3, 'b-');  
xlabel('X'); ylabel('Y');  
legend('原始数据','3阶多项式拟合','5阶多项式拟合','8阶多项式拟合','Location','best');  
legend('boxoff')  
title('不同次数拟合曲线对比图')
```



## 2. 多项式评价和置信区间估计

- $[Y, \text{DELTA}] = \text{polyconf}(p, x, S, \alpha, \text{MU})$ 
  - 求polyfit所得的回归多项式在 $x$ 处的预测值 $Y$ 及预测值的显著性为 $1-\alpha$ 的置信区间DELTA;  $\alpha$ 缺省时为0.05。其中,  $p$ 是polyfit函数的返回值;  $x$ 和polyfit函数的 $x$ 值相同;  $S$ 和polyfit函数的 $S$ 值相同;  $\text{MU}$ 和polyfit函数的 $\text{MU}$ 值相同。
  - 它假设polyfit函数数据输入的误差是独立正态的, 并且方差为常数。

```
>> x = 0:0.2:10;  
>> y = 0.25*x + 20*sin(x); %数值运算  
>> [P, S, MU] = polyfit(x, y, 6); %6阶多项式拟合  
>> [Y, DELTA] = polyconf(P, x, S, 0.05, MU); %预测值与置信区间  
>> fill([x, flip1r(x)], [Y-DELTA, flip1r(Y+DELTA)], [0.9706 0.9216 0.9804])  
>> hold on  
>> plot(x, y, 'k.', 'MarkerSize', 10)  
>> plot(x, Y, 'r-', 'LineWidth', 2)  
>> plot(x, Y-DELTA, 'c--')  
>> plot(x, Y+DELTA, 'c--')  
>> grid on  
>> title('6次多项式拟合及置信区间')  
>> xlabel('X')  
>> ylabel('Y')
```



### 3. 曲线拟合案例分析

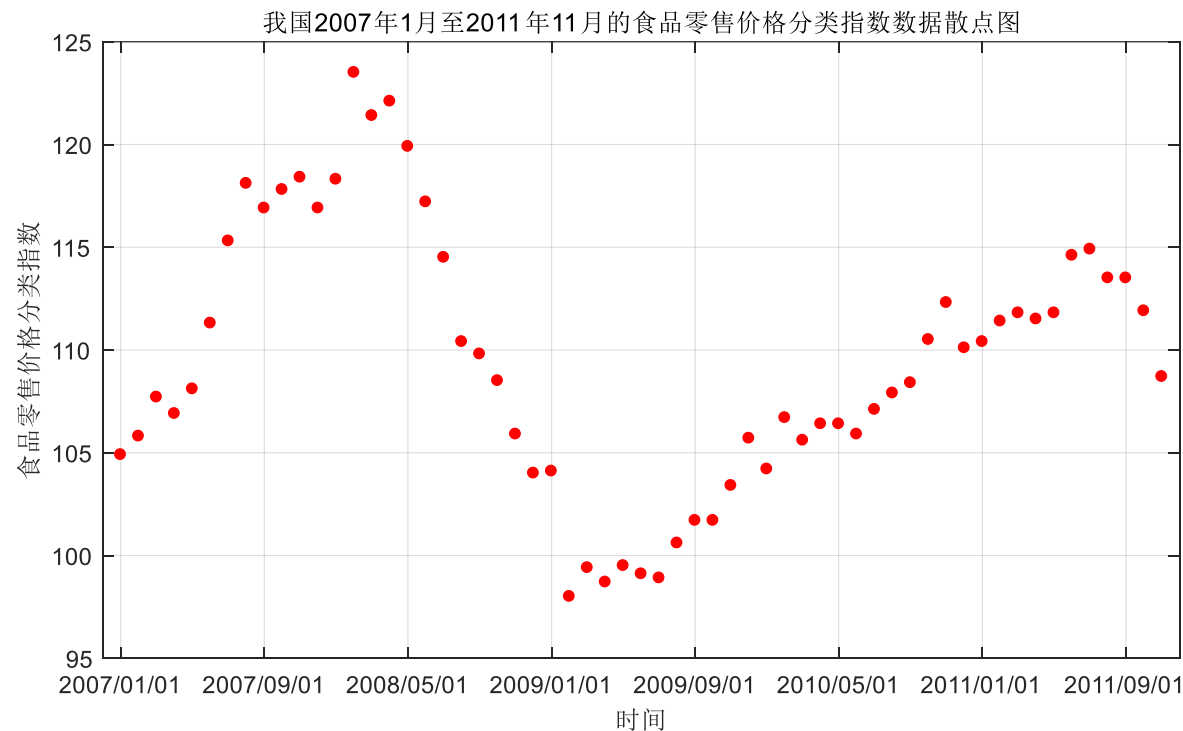
例：现有我国2007年1月至2011年11月的食品零售价格分类指数数据，如下表所示。数据来源：中华人民共和国国家统计局网站月度统计数据。试根据这59组统计数据研究全国食品零售价格分类指数  $y$ （上年同月 = 100）和时间  $x$  之间的关系。

序号	统计月度	上年同月 = 100			上年同期 = 100		
		全国	城市	农村	全国	城市	农村
1	2007年1月	104.9	104.4	105.9	104.9	104.4	105.9
2	2007年2月	105.8	105.2	106.9	105.3	104.8	106.4
3	2007年3月	107.7	107.4	108.3	106.1	105.7	107
...	...	...	...	...	...	...	...
57	2011年9月	113.5	113.4	114	112.6	112.4	113.1
58	2011年10月	111.9	111.8	112.2	112.5	112.3	113
59	2011年11月	108.7	108.8	108.6	112.2	112	112.6

### 3. 曲线拟合案例分析

#### 绘制散点图

```
[Data,Textdata] = xlsread('examp41.xls');  
x = Data(:,1);  
y = Data(:,3);  
timestr = Textdata(3:end,2);  
plot(x,y,'r.','Markersize',15);  
% xtick坐标刻度  
% numel数组元素的个数（或用length）  
% xticklabel坐标显示的字符串  
set(gca,'xtick',1:8:numel(x),'xticklabel',timestr(1:8:end));  
xlabel('时间');  
ylabel('食品零售价格分类指数');  
title('我国2007年1月至2011年11月的食品零售价格分类指数数据散点图');  
grid on
```



### 3. 曲线拟合案例分析

```
[p4,s4,mu4] = polyfit(x,y,4); %4次拟合
```

```
pf4 = vpa(poly2sym(p4),5); %显示4次拟合多项式
```

```
[p6,s6,mu6] = polyfit(x,y,6); %6次拟合
```

```
[p9,s9,mu9] = polyfit(x,y,9); %9次拟合
```

```
[p11,s11,mu11] = polyfit(x,y,11); %11次拟合
```

```
yd4 = polyval(p4,x,s4,mu4); %多项式求值
```

```
yd6 = polyval(p6,x,s6,mu6); %多项式求值
```

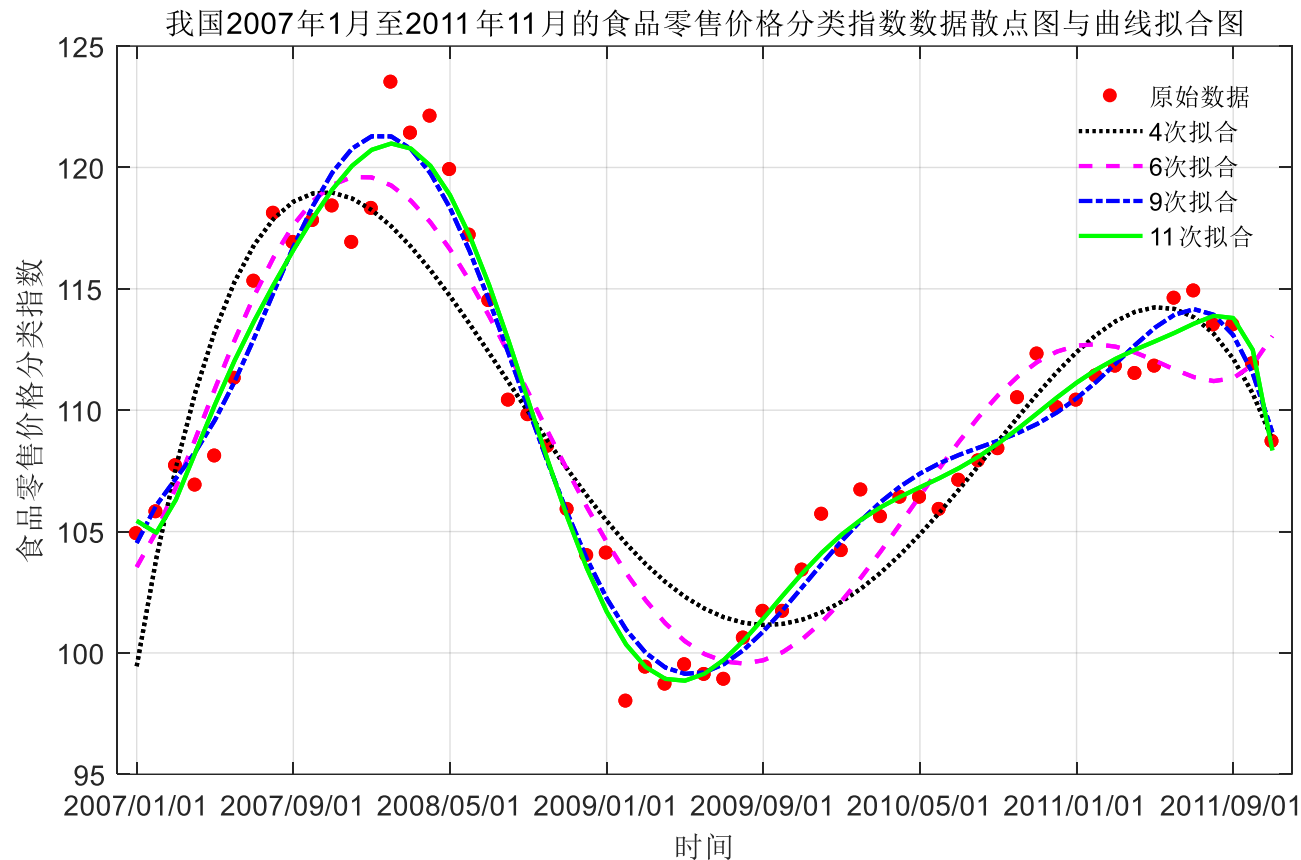
```
yd9 = polyval(p9,x,s9,mu9); %多项式求值
```

```
yd11 = polyval(p11,x,s11,mu11); %多项式求值
```

```
hold on %保持当前图像
```

```
plot(x,yd4,'k:',x,yd6,'m--',x,yd9,'b-.',x,yd11,'g-', 'LineWidth',1.5);
```

```
legend('原始数据','4次拟合','6次拟合','9次拟合','11次拟合'); legend('boxoff')
```





## 4. 自定义函数拟合

- 对于给定的数据，根据经验拟合为带有待定常数的自定义函数：**nlinfit()**。
- 调用格式：**[beta, r, J]=nlinfit(x, y, 'fun', beta0)**：beta返回函数fun中的待定常数；r表示残差；J表示雅可比矩阵。x, y为数据；fun自定义函数；beta0待定常数初值。
- 例：在化工生产中获得的氯气的级分y随生产时间x下降，假定在 $x \geq 8$ 时，y与x之间有如下形式的非线性模型：

$$y = a + (0.49 - a)e^{-b(x-8)}$$

现收集了44组数据，利用该数据通过拟合确定非线性模型中的待定常数。

## 4. 自定义函数拟合

% 2、命名窗口输入:

```
>> x = [8.00 8.00 10.00 10.00 10.00 10.00 12.00 12.00 12.00 14.00 14.00 14.00 16.00 16.00 16.00 18.00 18.00  
20.00 20.00 20.00 20.00 22.00 22.00 24.00 24.00 24.00 26.00 26.00 26.00 28.00 28.00 30.00 30.00 30.00  
32.00 32.00 34.00 36.00 36.00 38.00 38.00 40.00 42.00]';
```

```
>> y = [0.49 0.49 0.48 0.47 0.48 0.47 0.46 0.46 0.45 0.43 0.45 0.43 0.43 0.44 0.43 0.43 0.46 0.42 0.42  
0.43 0.41 0.41 0.40 0.42 0.40 0.40 0.41 0.40 0.41 0.41 0.40 0.40 0.40 0.38 0.41 0.40 0.40 0.41 0.38 0.40  
0.40 0.39 0.39]';
```

```
>> beta0 = [0.30 0.02];
```

```
>> betafit = nlinfit(x,y,'mfun',beta0)
```

betafit =

0.3896 0.1011

$$y = 0.3896 + (0.49 - 0.3896)e^{-0.1011(x-8)}$$

% 1、定义函数文件

```
function y = mfun(beta,x)
```

```
    a = beta(1);
```

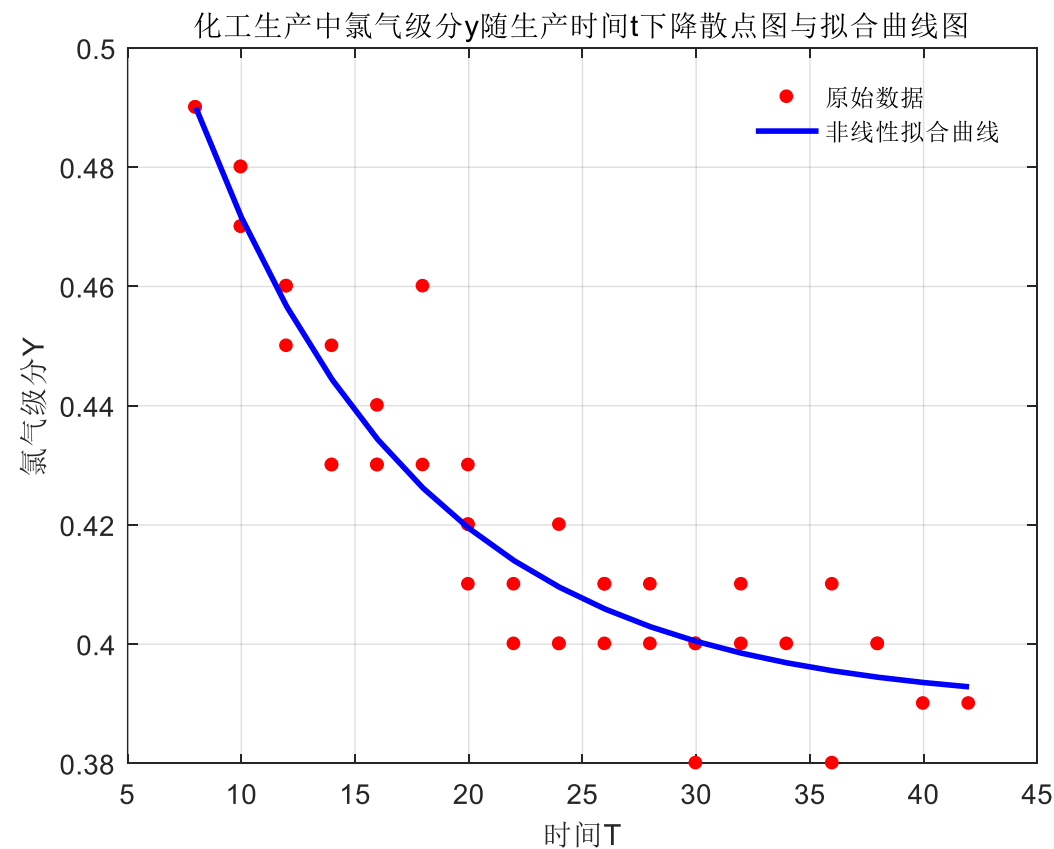
```
    b = beta(2);
```

```
    y = a + (0.49 - a) * exp(-b * (x - 8));
```

```
end
```

## 4. 自定义函数拟合

```
>> plot(x,y,'r','MarkerSize',15)
>> hold on
>> yi = mfun(betafit,x);
>> plot(x,yi,'b-','LineWidth',2)
>> grid on
>> xlabel('时间T')
>> ylabel('氯气级分Y')
>> legend('原始数据','非线性拟合曲线')
>> legend('boxoff')
>> title('化工生产中氯气级分y随生产时间t下降散点图与拟合曲线图')
```





---

# 感谢聆听

---