极值学院
edu.mathor.com
数学家旗下
在线教育平台

《数学建模基础课程》
课程地址

基础数学模型

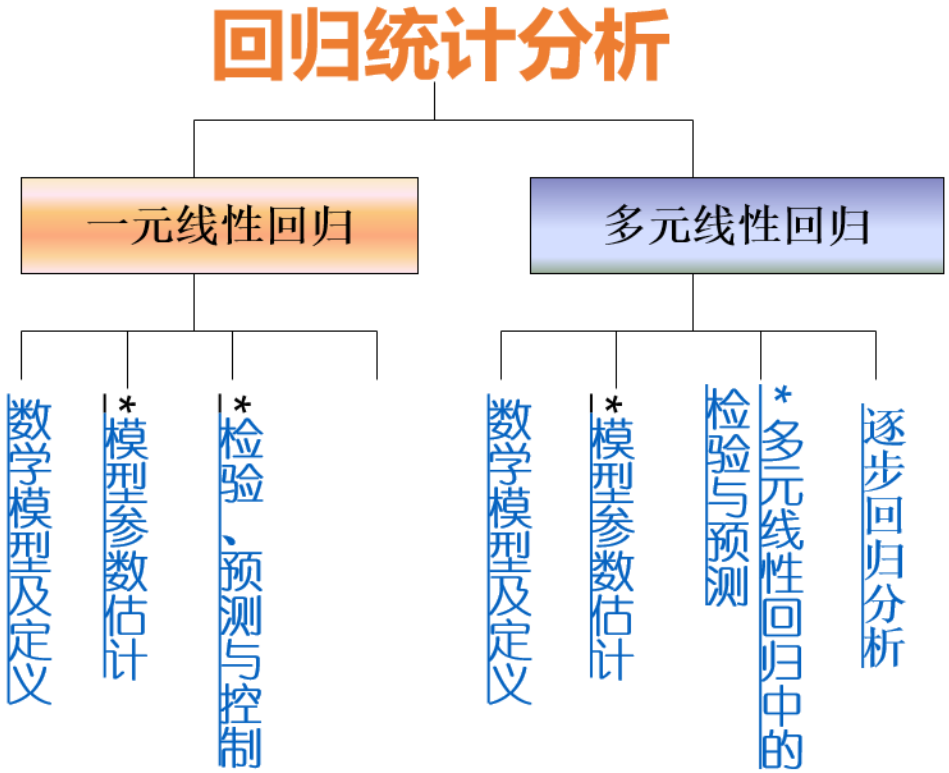
回归与统计模型

主讲人 张文斌

上海交通大学机械与动力工程学院博士生
曾获美国数学建模特等奖（Outstanding）
研究数学建模多年，掌握一定数模获奖技巧
熟悉数学建模方法、编程及论文写作

目录

1、回归和统计模型基本概念



1、回归和统计模型基本概念

➢ 基本概念

一般地，称由 $y=\beta_0+\beta_1 x+\varepsilon$ 确定的模型为**一元线性回归模型**，

记为

$$\begin{cases} y=\beta_0+\beta_1 x+\varepsilon \\ E\varepsilon=0, D\varepsilon=\sigma^2 \end{cases}$$

$Y=\beta_0+\beta_1 x$，称为 **y 对 x 的回归直线方程.**

固定的未知参数 β_0、β_1 称为回归系数，自变量 x 也称为回归变量.

一元线性回归分析的**主要任务**是：

1. 用试验值（样本值）对 β_0、β_1 和 σ 作点估计；
2. 对回归系数 β_0、β_1 作假设检验；
3. 在 $x=x_0$ 处对 y 作预测，对 y 作区间估计.

2、回归和统计模型的说明

➢ 模型参数估计

1. 回归系数的最小二乘估计

有 n 组独立观测值（x_1，y_1），（x_2，y_2），…，（x_n，y_n）

$$\text{设}\begin{cases} y_i = \beta_0 + \beta x_1 + \varepsilon_i, i = 1,2,...,n \\ E\varepsilon_i = 0, D\varepsilon_i = \sigma^2 \quad \text{且}\varepsilon_1\varepsilon_2,...,\varepsilon_n\text{相互独立} \end{cases}$$

$$\text{记} \quad Q = Q(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

最小二乘法就是选择 β_0 和 β_1 的估计 $\hat{\beta}_0$，$\hat{\beta}_1$ 使得

$$Q(\hat{\beta}_0, \hat{\beta}_1) = \min_{\beta_0, \beta_1} Q(\beta_0, \beta_1)$$

2、回归和统计模型的说明

➢ 模型参数估计

1. 回归系数的最小二乘估计

$$\text{解得}\quad \begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \dfrac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} \end{cases} \qquad \text{或}\quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

其中 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$, $\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$.

（经验）回归方程为:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = \bar{y} + \hat{\beta}_1 (x - \bar{x})$$

2、回归和统计模型的说明

➢ 模型参数估计

2．σ^2的无偏估计

$$记\quad Q_e = Q(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

称 Q_e 为残差平方和或剩余平方和.

σ^2 的无偏估计为 $\hat{\sigma}_e^2 = Q_e/(n-2)$

称 $\hat{\sigma}_e^2$ 为剩余方差（残差的方差），$\hat{\sigma}_e^2$ 分别与 $\hat{\beta}_0$、$\hat{\beta}_1$ 独立.

$\hat{\sigma}_e$ 称为剩余标准差.

2、回归和统计模型的说明

➢ 检验、预测与控制

1. 回归方程的显著性检验

对回归方程 $Y=\beta_0+\beta_1 x$ 的显著性检验，归结为对假设

$$H_0:\beta_1=0;H_1:\beta_1\neq 0$$

进行检验.

假设 $H_0:\beta_1=0$ 被拒绝，则回归显著，认为 y 与 x 存在线性关系，所求的线性回归方程有意义；否则回归不显著，y 与 x 的关系不能用一元线性回归模型来描述，所得的回归方程也无意义.

2、回归和统计模型的说明

➢ 检验、预测与控制

1. 回归方程的显著性检验

（I）F检验法

当 H_0 成立时，$F=\dfrac{U}{Q_e/(n-2)}\sim F(1,n\text{-}2)$

其中 $U=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$（回归平方和）

故 $F>F_{1-\alpha}(1,n-2)$，拒绝 H_0，否则就接受 H_0.

（II）t 检验法

当 H_0 成立时，$T=\dfrac{\sqrt{L_{xx}}\hat{\beta}_1}{\hat{\sigma}_e}\sim t(n\text{-}2)$

故 $|T|>t_{1-\frac{\alpha}{2}}(n-2)$，拒绝 H_0，否则就接受 H_0.

其中 $L_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2$

2、回归和统计模型的说明

➢ 检验、预测与控制

1．回归方程的显著性检验

（III）*r*检验法

记 $$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

当$|r|>\mathrm{r}_{1-\alpha}$时，拒绝 H_0；否则就接受 H_0.

其中 $r_{1-\alpha}=\sqrt{\frac{1}{1+(n-2)/F_{1-\alpha}(1,n-2)}}$

2、回归和统计模型的说明

➢ 检验、预测与控制　　2. 回归系数的置信区间

β_0 和 β_1 置信水平为 1-α 的置信区间分别为

$$\left[\hat{\beta}_0 - t_{1-\frac{\alpha}{2}}(n-2)\hat{\sigma}_e\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{L_{xx}}}, \hat{\beta}_0 + t_{1-\frac{\alpha}{2}}(n-2)\hat{\sigma}_e\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{L_{xx}}}\right]$$

和 $$\left[\hat{\beta}_1 - t_{1-\frac{\alpha}{2}}(n-2)\hat{\sigma}_e/\sqrt{L_{xx}}, \hat{\beta}_1 + t_{1-\frac{\alpha}{2}}(n-2)\hat{\sigma}_e/\sqrt{L_{xx}}\right]$$

σ^2 的置信水平为 1-α 的置信区间为

$$\left[\frac{Q_e}{\chi^2_{1-\frac{\alpha}{2}}(n-2)}, \frac{Q_e}{\chi^2_{\frac{\alpha}{2}}(n-2)}\right]$$

2、回归和统计模型的说明

➢ 检验、预测与控制

3. 预测与控制

（1）预测

用 y_0 的回归值 $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ 作为 **y_0 的预测值.**

y_0 的置信水平为 $1-\alpha$ 的**预测区间**为

$$\left[\hat{y}_0 - \delta(x_0), \hat{y}_0 + \delta(x_0)\right]$$

其中 $\delta(x_0) = \hat{\sigma}_e t_{1-\frac{\alpha}{2}}(n-2)\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{L_{xx}}}$

特别，当 n 很大且 x_0 在 $\bar{x}$ 附近取值时，y 的置信水平为 $1-\alpha$ 的**预测区间近似为**

$$\left[\hat{y} - \hat{\sigma}_e u_{1-\frac{\alpha}{2}}, \hat{y} + \hat{\sigma}_e u_{1-\frac{\alpha}{2}}\right]$$

2、回归和统计模型的说明

➢ 检验、预测与控制　　3．预测与控制　　（2）控制

要求：$y=\beta_0+\beta_1 x+\varepsilon$ 的值以 $1-\alpha$ 的概率落在指定区间 (y',y'')

只要控制 x 满足以下两个不等式

$$\hat{y}-\delta(x)\geq y',\hat{y}+\delta(x)\leq y''$$

要求 $y''-y'\geq 2\delta(x)$.若 $\hat{y}-\delta(x)=y',\hat{y}-\delta(x)=y''$ 分别有解 x' 和 x''，即 $\hat{y}-\delta(x')=y',\hat{y}+\delta(x'')=y''$.

则 (x',x'') 就是所求的 x 的控制区间.

2、回归和统计模型的说明

➢ 数学模型及定义

通常选择的六类曲线如下：

（1）双曲线 $\frac{1}{y}=a+\frac{b}{x}$

（2）幂函数曲线 $y=a\,x^{b}$，其中 $x>0,a>0$

（3）指数曲线 $y=a\,\mathrm{e}^{bx}$ 其中参数 $a>0.$

（4）倒指数曲线 $y=a\,\mathrm{e}^{b/x}$ 其中 $a>0$，

（5）对数曲线 $y=a+b\log x,x>0$

（6）S 型曲线 $y=\frac{1}{a+b\mathrm{e}^{-x}}$

2、回归和统计模型的说明

➢ 数学模型及定义

一般称 $\begin{cases} Y = X\beta + \varepsilon \\ E(\varepsilon) = 0, \mathrm{COV}(\varepsilon, \varepsilon) = \sigma^2 I_n \end{cases}$

为高斯-马尔可夫线性模型（k 元线性回归模型），并简记为 $(Y, X\beta, \sigma^2 I_n)$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$ 称为回归平面方程.

线性模型 $(Y, X\beta, \sigma^2 I_n)$ 考虑的主要问题是：

（1） 用试验值（样本值）对未知参数 β 和 σ^2 作点估计和假设检验，从而建立 y 与 $x_1, x_2, \ldots, x_k$ 之间的数量关系；

（2） 在 $x_1 = x_{01}, x_2 = x_{02}, \ldots, x_k = x_{0k}$，处对 y 的值作预测与控制，即对 y 作区间估计.

2、回归和统计模型的说明

➢ 模型参数估计

1. 对 β_i 和 σ^2 作估计

用最小二乘法求 $\beta_0,\ldots,\beta_k$ 的估计量：作离差平方和

$$Q = \sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_{i1} - \ldots - \beta_k x_{ik}\right)^2$$

选择 $\beta_0,\ldots,\beta_k$ 使 Q 达到最小.

解得估计值 $\hat{\beta} = (X^T X)^{-1}(X^T Y)$

得到的 $\hat{\beta}_i$ 代入回归平面方程得：

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$

称为经验回归平面方程. $\hat{\beta}_i$ 称为经验回归系数.

注意： $\hat{\beta}$ 服从 $p+1$ 维正态分布，且为 β 的无偏估计，协方差阵为 $\sigma^2 C$.$C=L^{-1}=(c_{ij})$, $L=X^T X$

2、回归和统计模型的说明

➢ 模型参数估计

2. 多项式回归

设变量 x、Y 的回归模型为

$$Y=\beta_0+\beta_1 x+\beta_2 x^2+\ldots+\beta_p x^p+\varepsilon$$

其中 p 是已知的，$\beta_i(i=1,2,\cdots,p)$ 是未知参数，ε 服从正态分布 $N(0,\sigma^2)$.

$$Y=\beta_0+\beta_1 x+\beta_2 x^2+\ldots+\beta_k x^k$$

称为回归多项式.上面的回归模型称为多项式回归.

令 $x_i=x^i$，i=1，2，…，k 多项式回归模型变为多元线性回归模型.

2、回归和统计模型的说明

➢ 多元线性回归中的检验与预测

1. 线性模型和回归系数的检验

假设 $H_0 : \beta_0 = \beta_1 = \cdots = \beta_k = 0$

其中 $U = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ （回归平方和）

$Q_e = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

（I）*F* 检验法

当 H_0 成立时，$F = \dfrac{U/k}{Q_e/(n-k-1)} \sim F(k, n-k-1)$

如果 $F > F_{1-\alpha}(k, n\text{-}k\text{-}1)$，则拒绝 H_0，认为 y 与 $x_1, \cdots, x_k$ 之间显著地有线性关系；否则就接受 H_0，认为 y 与 $x_1, \cdots, x_k$ 之间线性关系不显著.

（II）*r* 检验法

定义 $R = \sqrt{\dfrac{U}{L_{yy}}} = \sqrt{\dfrac{U}{U+Q_e}}$ 为 y 与 $x_1, x_2, \ldots, x_k$ 的多元相关系数或复相关系数.

由于 $F = \dfrac{n-k-1}{k} \dfrac{R^2}{1-R^2}$，故用 F 和用 R 检验是等效的.

2、回归和统计模型的说明

➢ 多元线性回归中的检验与预测

2. 预测

（1）点预测

求出回归方程 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\cdots+\hat{\beta}_kx_k$，对于给定自变量的值 $x_1^*,\cdots,x_k^*$，用 $\hat{y}^*=\hat{\beta}_0+\hat{\beta}_1x_1^*+\cdots+\hat{\beta}_kx_k^*$ 来预测 $y^*=\beta_0+\beta_1x_1^*+\cdots+\beta_kx_k^*+\varepsilon$.称 $\hat{y}^*$ 为 y^* 的点预测.

（2）区间预测

y 的 $1-\alpha$ 的预测（置信）区间为 $(\hat{y}_1,\hat{y}_2)$,其中

$$\begin{cases}\hat{y}_1=\hat{y}-\hat{\sigma}_e\sqrt{1+\sum_{i=0}^{k}\sum_{j=0}^{k}c_{ij}x_ix_j}\,t_{1-\alpha/2}(n-k-1)\\ \hat{y}_2=\hat{y}+\hat{\sigma}_e\sqrt{1+\sum_{i=0}^{k}\sum_{j=0}^{k}c_{ij}x_ix_j}\,t_{1-\alpha/2}(n-k-1)\end{cases}$$

$C=L^{-1}=(c_{ij})$, $L=X^{\mathrm{T}}X$

2、回归和统计模型的说明

➢ 逐步回归分析

“最优”的回归方程就是包含所有对Y有影响的变量，而不包含对Y影响不显著的变量回归方程。

选择“最优”的回归方程有以下几种方法：

（1）从所有可能的因子（变量）组合的回归方程中选择最优者；

（2）从包含全部变量的回归方程中逐次剔除不显著因子；

（3）从一个变量开始，把变量逐个引入方程；

（4）“有进有出”的逐步回归分析.

以第四种方法，即逐步回归分析法在筛选变量方面较为理想.

2、回归和统计模型的说明

➢ 逐步回归分析

逐步回归分析法的思想：

- 从一个自变量开始，视自变量Y对作用的显著程度，从大到小地依次逐个引入回归方程.
- 当引入的自变量由于后面变量的引入而变得不显著时，要将其剔除掉.
- 引入一个自变量或从回归方程中剔除一个自变量，为逐步回归的一步.
- 对于每一步都要进行Y值检验，以确保每次引入新的显著性变量前回归方程中只包含对Y作用显著的变量.
- 这个过程反复进行，直至既无不显著的变量从回归方程中剔除，又无显著变量可引入回归方程时为止.

3、回归和统计模型的求解

多元线性回归

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

1. 确定回归系数的点估计值：

b=regress(Y，X)

2. 求回归系数的点估计和区间估计、并检验回归模型：

[b, bint,r,rint,stats]=regress(Y,X,alpha)

3. 画出残差及其置信区间：

rcoplot（r，rint）

相关系数 r^2 越接近 1，说明回归方程越显著；

$F > F_{1-\alpha}(k, n-k-1)$ 时拒绝 H_0，F 越大，说明回归方程越显著；

与 F 对应的概率 $p < \alpha$ 时拒绝 H_0，回归模型成立.

3、回归和统计模型的实例分析

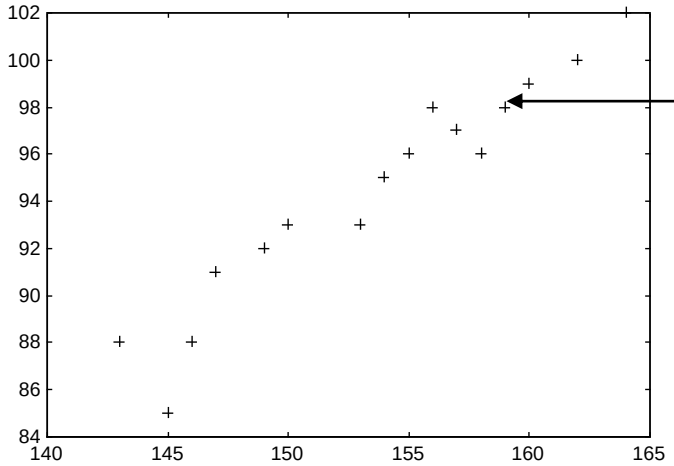
例1-1 测16名成年女子的身高与腿长所得数据如下：

身高(cm)	143	145	146	147	149	150	153	154	155	156	157	158	159	160	162	164
腿长(cm)	88	85	88	91	92	93	93	95	96	98	97	96	98	99	100	102

以身高x为横坐标，以腿长y为纵坐标将这些数据点（x_i，y_i）在平面直角坐标系上标出.

3、回归和统计模型的实例分析

➢ 例1-1



散点图

Example_1_1

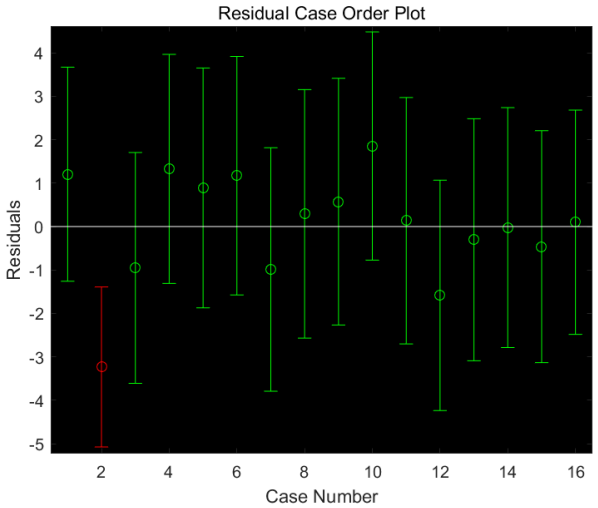
$$y = \beta_0 + \beta_1 x + \varepsilon$$

```
b =
  -16.0730
    0.7194
bint =
  -33.7071    1.5612
    0.6047    0.8340
stats =  0.9282  180.9531    0.0000    1.7437
```

3、回归和统计模型的实例分析

残差分析，作残差图：

rcoplot(r,rint)



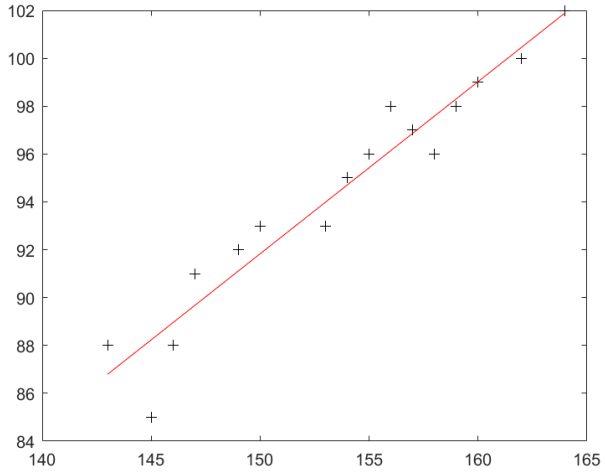
从残差图可以看出，除第二个数据外，其余数据的残差离零点均较近，且残差的置信区间均包含零点，这说明回归模型 y=-16.073+0.7194x能较好的符合原始数据，而第二个数据可视为异常点.

预测及作图：

z=b(1)+b(2)*

plot(x,Y,'k+',x,z,'r')

Example_1_2



3、回归和统计模型的求解

多项式回归

（一）一元多项式回归 $y=a_1x^m+a_2x^{m-1}+\ldots+a_mx+a_{m+1}$

1．回归：

（1）确定多项式系数的命令：[p，S]=polyfit（x，y，m）

其中 $x=(x_1, x_2, \cdots, x_n)$，$y=(y_1, y_2, \cdots, y_n)$；

$p=(a_1, a_2, \cdots, a_{m+1})$ 是多项式 $y=a_1x^m+a_2x^{m-1}+\cdots+a_mx+a_{m+1}$ 的系数；S 是一个矩阵，用来估计预测误差.

（2）一元多项式回归命令：polytool（x，y，m）

2．预测和预测误差估计：

（1）Y=polyval（p，x）求polyfit所得的回归多项式在x处 的预测值Y；

（2）[Y，DELTA]=polyconf（p，x，S，alpha）

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Y DELTA；alpha缺省时为0.05.

3、回归和统计模型的求解

（二）多元二项式回归

命令：**rstool（x，y，’model’, alpha）**

由下列 4 个模型中选择 1 个（用字符串输入，缺省时为线性模型）：

linear（线性）：$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$

purequadratic（纯二次）：$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{j=1}^{n} \beta_{jj} x_j^2$

interaction（交叉）：$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{1 \le j \ne k \le m} \beta_{jk} x_j x_k$

quadratic（完全二次）：$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{1 \le j,k \le m} \beta_{jk} x_j x_k$

3、回归和统计模型的实例分析

例1-2 设某商品的需求量与消费者的平均收入、商品价格的统计数据如下，建立回归模型，预测平均收入为1000、价格为6时的商品需求量.

需求量	100	75	80	70	50	65	90	100	110	60
收入	1000	600	1200	500	300	400	1300	1100	1300	300
价格	5	7	6	6	8	7	5	4	3	9

选择纯二次模型，即 $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_{11}x_1^2+\beta_{22}x_2^2$

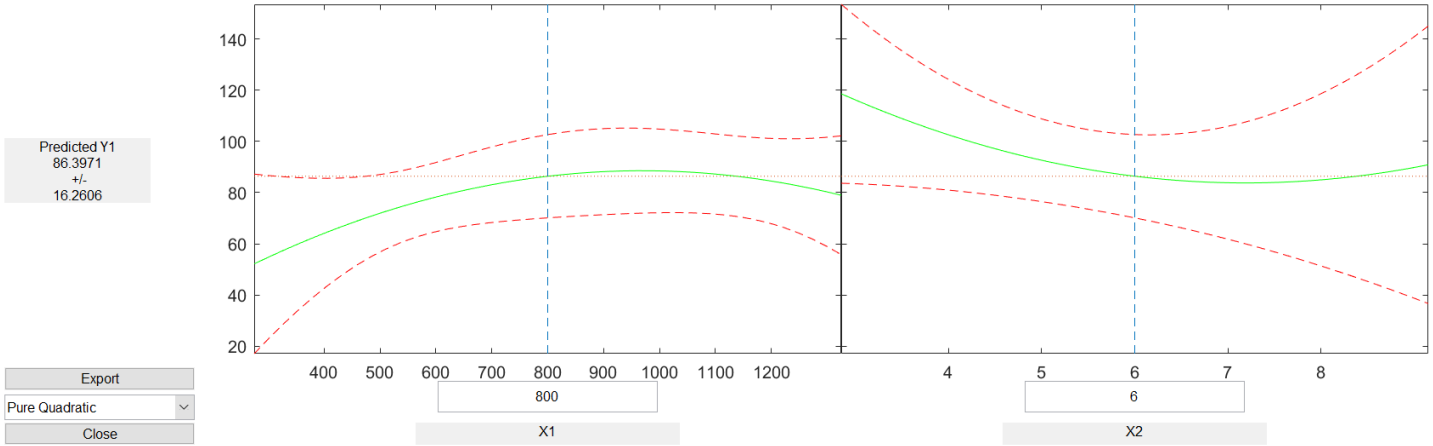
3、回归和统计模型的实例分析

直接用多元二项式回归：

Example_1_2

在画面左下方的下拉式菜单中选”all”, 则beta．rmse和residuals都传送到MATLAB工作区中.

将界面下方方框中的“800”改成1000，右边图形下方的方框中仍输入6.则画面左边的“Predicted Y”下方的数据由原来“86.3971”变为88.4791，即预测出平均收入为1000，价格为6时的商品需求量为88.4791.



3、回归和统计模型的求解

非线性回归

1．回归：

（1）确定回归系数的命令：

[beta，r，J]=nlinfit（x,y,'model',beta0）

（2）非线性回归命令：nlintool（x，y，'model', beta0，alpha）

2．预测和预测误差估计：

[Y，DELTA]=nlpredci（'model', x，beta，r，J）

求nlinfit 或lintool所得的回归函数在x处的预测值Y及预测值的显著性水平为1-alpha的置信区间Y ±DELTA.

3、回归和统计模型的求解

逐步回归

逐步回归的命令是：

stepwise（x，y，inmodel，alpha）

运行**stepwise**命令时产生三个图形窗口：**Stepwise Plot**，**Stepwise Table**，**Stepwise History**.

在**Stepwise Plot**窗口，显示出各项的回归系数及其置信区间.

Stepwise Table 窗口中列出了一个统计表，包括回归系数及其置信区间，以及模型的统计量剩余标准差（**RMSE**）、相关系数（**R-square**）、**F**值、与**F**对应的概率**P**.

3、回归和统计模型的实例分析

例1-3 水泥凝固时放出的热量y与水泥中4种化学成分x_1、x_2、x_3、x_4有关，今测得一组数据如下，试用逐步回归法确定一个 线性模型.

序号	1	2	3	4	5	6	7	8	9	10	11	12	13
x_1	7	1	11	11	7	11	3	1	2	21	1	11	10
x_2	26	29	56	31	52	55	71	31	54	47	40	66	68
x_3	6	15	8	8	6	9	17	22	18	4	23	9	8
x_4	60	52	20	47	33	22	6	44	22	26	34	12	12
y	78.5	74.3	104.3	87.6	95.9	109.2	102.7	72.5	93.1	115.9	83.8	113.3	109.4

3、回归和统计模型的实例分析

1．数据输入：

```
x1=[7 1 11 11 7 11 3 1 2 21 1 11 10]';
x2=[26 29 56 31 52 55 71 31 54 47 40 66 68]';
x3=[6 15 8 8 6 9 17 22 18 4 23 9 8]';
x4=[60 52 20 47 33 22 6 44 22 26 34 12 12]';
y=[78.5 74.3 104.3 87.6 95.9 109.2 102.7 72.5 93.1 115.9 83.8 113.3 109.4]';
x=[x1 x2 x3 x4];
```

2．逐步回归：

（1）先在初始模型中取全部自变量：

stepwise(x,y)

得图**Stepwise Plot** 和表**Stepwise Table**

3、回归和统计模型的实例分析

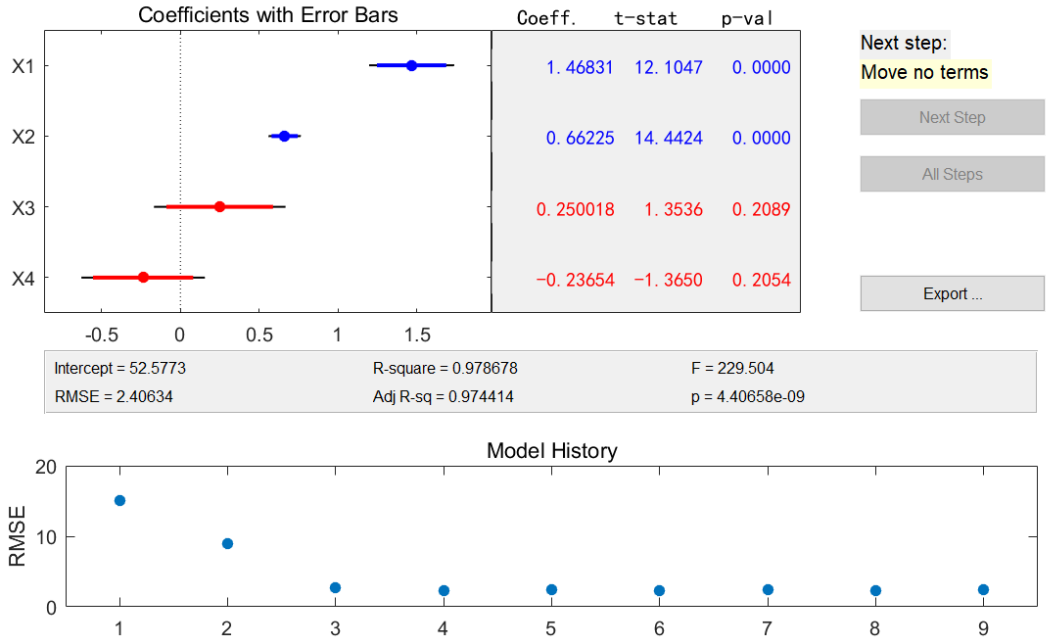
Example_1_3_1

图Stepwise Plot中四条直线都是实线，说明模型的显著性不好

从表Stepwise Table中看出变量x_3和x_4的显著性最差.

3、回归和统计模型的实例分析

（2）在图Stepwise Plot中点击直线3和直线4，移去变量x_3和x_4



Example_1_3_1

移去变量x_3和x_4后模型具有显著性.

虽然剩余标准差（RMSE）没有太大的变化，但是统计量F的值明显增大，因此新的回归模型更好.

3、回归和统计模型的实例分析

（3）对变量y和x_1、x_2作线性回归：

```
X=[ones(13,1) x1 x2];
b=regress(y,X)
```

Example_1_3_2

得结果：b =

52.5773
1.4683
0.6623

故最终模型为：$y=52.5773+1.4683x_1+0.6623x_2$

练习题

财政收入预测问题：财政收入与国民收入、工业总产值、农业总产值、总人口、就业人口、固定资产投资等因素有关.表中列出了1952–1981年的原始数据（具体数据可查看附件：财政收入预测问题.pdf），试构造预测模型.

年份	国民收入（亿元）	工业总产值(亿元)	农业总产值（亿元）	总人口（万人）	就业人口（万人）	固定资产投资（亿元）	财政收入(亿元)
1952	598	349	461	57482	20729	44	184
1953	586	455	475	58796	21364	89	216
1954	707	520	491	60266	21832	97	248
1955	737	558	529	61465	22328	98	254
1956	825	715	556	62828	23018	150	268
1957	837	798	575	64653	23711	139	286
1958	1028	1235	598	65994	26600	256	357
1959	1114	1681	509	67207	26173	338	444
1960	1079	1870	444	66207	25880	380	506
1961	757	1156	434	65859	25590	138	271
1962	677	964	461	67295	25110	66	230
1963	779	1046	514	69172	26640	85	266
1964	943	1250	584	70499	27736	129	323
1965	1152	1581	632	72538	28670	175	393
1966	1322	1911	687	74542	29805	212	466
1967	1249	1647	697	76368	30814	156	352
1968	1187	1565	680	78534	31915	127	303
1969	1372	2101	688	80671	33225	207	447
1970	1638	2747	767	82992	34432	312	564
1971	1780	3156	790	85229	35620	355	638
1972	1833	3365	789	87177	35854	354	658
1973	1978	3684	855	89211	36652	374	691
1974	1993	3696	891	90859	37369	393	655
1975	2121	4254	932	92421	38168	462	692
1976	2052	4309	955	93717	38834	443	657
1977	2189	4925	971	94974	39377	454	723
1978	2475	5590	1058	96259	39856	550	922
1979	2702	6065	1150	97542	40581	564	890
1980	2791	6592	1194	98705	41896	568	826
1981	2927	6862	1273	100072	43280	496	810

感谢各位聆听!

Thanks for Listening

Q&A