

线性回归模型与软件计算

1. 引言

2004年全国数模竞赛B题“电力市场的输电阻塞管理”的第1个问题是这样：

某电网有8台发电机组，6条主要线路，表1和表2中的方案0给出了各机组的当前出力和各线路上对应的有功潮流值，方案1~32给出了围绕方案0的一些实验数据，试用这些数据确定各线路上有功潮流关于各发电机组出力的近似表达式.

表 1 各机组出力方案 (单位: 兆瓦, 记作 MW)

方案\机组	1	2	3	4	5	6	7	8
0	120	73	180	80	125	125	81.1	90
1	133.02	73	180	80	125	125	81.1	90
2	129.63	73	180	80	125	125	81.1	90
3	158.77	73	180	80	125	125	81.1	90
4	145.32	73	180	80	125	125	81.1	90
5	120	78.596	180	80	125	125	81.1	90
6	120	75.45	180	80	125	125	81.1	90
7	120	90.487	180	80	125	125	81.1	90
8	120	83.848	180	80	125	125	81.1	90
9	120	73	231.39	80	125	125	81.1	90
10	120	73	198.48	80	125	125	81.1	90
25	120	73	180	80	125	125	60.582	90
26	120	73	180	80	125	125	70.962	90
27	120	73	180	80	125	125	64.854	90
28	120	73	180	80	125	125	75.529	90
29	120	73	180	80	125	125	81.1	104.84
30	120	73	180	80	125	125	81.1	111.22
31	120	73	180	80	125	125	81.1	98.092
32	120	73	180	80	125	125	81.1	120.44

表 2 各线路的潮流值（各方案与表 1 相对应，单位：MW）

方案\线路	1	2	3	4	5	6
0	164.78	140.87	-144.25	119.09	135.44	157.69
1	165.81	140.13	-145.14	118.63	135.37	160.76
2	165.51	140.25	-144.92	118.7	135.33	159.98
3	167.93	138.71	-146.91	117.72	135.41	166.81
4	166.79	139.45	-145.92	118.13	135.41	163.64
5	164.94	141.5	-143.84	118.43	136.72	157.22
6	164.8	141.13	-144.07	118.82	136.02	157.5
7	165.59	143.03	-143.16	117.24	139.66	156.59
8	165.21	142.28	-143.49	117.96	137.98	156.96
9	167.43	140.82	-152.26	129.58	132.04	153.6
10	165.71	140.82	-147.08	122.85	134.21	156.23
25	162.21	141.21	-144.13	116.03	135.5	154.26
26	163.54	141	-144.16	117.56	135.44	155.93
27	162.7	141.14	-144.21	116.74	135.4	154.88
28	164.06	140.94	-144.18	118.24	135.4	156.68
29	164.66	142.27	-147.2	120.21	135.28	157.65
30	164.7	142.94	-148.45	120.68	135.16	157.63
31	164.67	141.56	-145.88	119.68	135.29	157.61
32	164.69	143.84	-150.34	121.34	135.12	157.64

设 6 条线路上有功潮流为 $y_j (j = 1, 2, \dots, 6)$ ，8 台发电机组出力为 $x_i (i = 1, 2, \dots, 8)$ ，

该问题变为寻找函数关系表达式：

$$y_j = f_j(x_1, x_2, \dots, x_8) \quad (j = 1, 2, \dots, 6) \quad (1)$$

对本问题，我们采用多元线性回归分析，效果很好。

2. 回归分析方法

回归分析，就是对平面上一些散布的点，采用一条最好的直线去表达。

如图1是12组儿子身高 y 和父亲身高 x 数据关系的散布点，采用直线拟合。

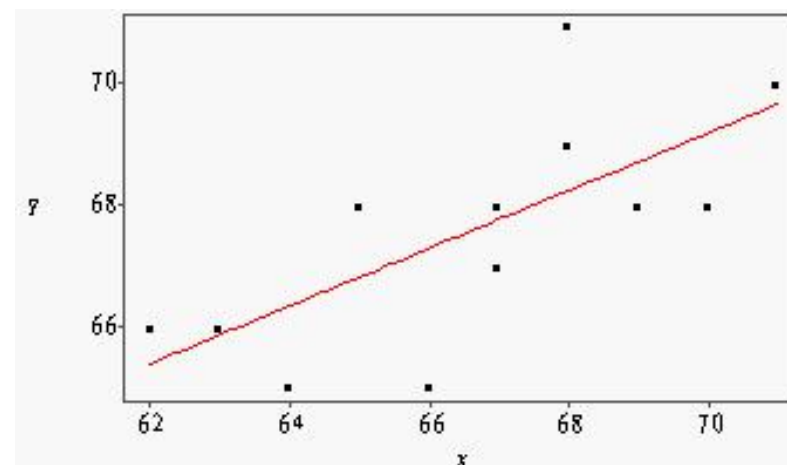


图1

2.1 一元线性回归

模型: $y = \alpha + \beta x + \varepsilon$ (2)

其中 $\varepsilon \sim N(0, \sigma^2)$

对一组观测值 $(x_i, y_i) (i=1, 2, \dots, n)$, 满足:

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (3)$$

其中各 ε_i 相互独立且 $\varepsilon_i \sim N(0, \sigma^2)$ ($i=1, 2, \dots, n$)

目标函数:

$$\sum_{i=1}^n (y_i - \hat{a} - \hat{\beta} x_i)^2 = \min_{\alpha, \beta} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \quad (4)$$

记

$$S(a, \beta) = \sum_{i=1}^n (y_i - a - \beta x_i)^2 \quad (5)$$

则

$$\frac{\partial S}{\partial \alpha} = 2 \sum_{i=1}^n (y_i - \alpha - \beta x_i) = 0$$

$$\frac{\partial S}{\partial \beta} = 2 \sum_{i=1}^n (y_i - \alpha - \beta x_i) x_i = 0$$

有：

$$\left. \begin{aligned} n\hat{\alpha} + n\bar{x}\hat{\beta} &= n\bar{y} \\ n\bar{x}\hat{\alpha} + \sum_{i=1}^n x_i^2 \hat{\beta} &= \sum_{i=1}^n x_i y_i \end{aligned} \right\} \quad (6)$$

这里, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

解得

$$\begin{cases} \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \\ \hat{\beta} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{cases} \quad (7)$$

$$\sigma^2 \text{ 的无偏估计为: } \hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2}{n-2} \quad (8)$$

2.2 多元线性回归模型

$$\text{模型: } y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \varepsilon \quad (9)$$

$\varepsilon \sim N(0, \sigma^2), \beta_0, \beta_1, \cdots, \beta_m, \sigma^2$ 是未知参数.

设 $(x_{i1}, x_{i2}, \dots, x_{im}, y_i)$ ($i=1, 2, \dots, n$) 是 $(x_1, x_2, \dots, x_m, y)$ 的 n 个观测值, 则:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im} + \varepsilon_i \quad (i=1, 2, \dots, n) \quad (10)$$

其中各 ε_i 相互独立, 且 $\varepsilon_i \sim N(0, \sigma^2)$.

$$\text{令 } \beta = (\beta_0, \beta_1, \dots, \beta_m)^T, \varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T, \quad Y = (y_1, y_2, \dots, y_n)^T$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad (12)$$

则方程组用矩阵表达为

$$Y = X\beta + \varepsilon \quad (13)$$

假定矩阵 X 的秩等于 $m+1$. 即列满秩.

$$\text{则} \quad X^T Y = (X^T X) \hat{\beta}$$

$$\text{解得} \quad \hat{\beta} = (X^T X)^{-1} X^T Y \quad (14)$$

$$\sigma^2 \text{ 的无偏估计} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^m x_{ij} \hat{\beta}_j)^2}{n - m - 1} \quad (15)$$

当 $m=1$ 时, 就变成一元回归分析

2.3 回归模型的假设检验

2.3.1 回归方程的显著性检验

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_m = 0 \quad H_1: \text{至少有一个 } \beta_j \neq 0 (j = 1, 2, \cdots, m)$$

当原假设 H_0 成立时, 说明回归方程不显著, 采用线性进行回归是不适合的.

当备选假设 H_1 成立时, 说明回归方程显著, 采用线性回归有意义.

令 $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$, 考虑总离差平方和

$$S_T = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \quad (16)$$

$$= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = S_e + S_R$$

$$S_e = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \text{称为剩余残差平方和.}$$

$$S_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \text{称为回归平方和.}$$

在 H_0 成立的条件下，可以证明

$$S_e / \sigma^2 \sim \chi^2(n-m-1), S_R / \sigma^2 \sim \chi^2(m) \quad (17)$$

且 S_e 与 S_R 相互独立，

则

$$F = \frac{S_R / m}{S_e / (n-m-1)} \sim F(m, n-m-1) \quad (18)$$

对给定显著水平 α ，可查表得 $F_\alpha(m, n-m-1)$ ，计算统计量 F 的数值 f 。

若 $f \geq F_\alpha(m, n-m-1)$ ，则拒绝 H_0 ，即认为各系数不为零，线性回归方程是显著的。

否则接受 H_0 ，即认为线性回归方程不显著。

2.3.2 回归系数的显著性检验

检验假设

$$H_0: \beta_j = 0 \leftrightarrow H_1: \beta_j \neq 0 (j = 1, 2, \dots, m)$$

当原假设 H_0 成立时, 说明自变量 x_j 对 y 不起作用, 在回归模型中可以去掉.

当备选假设 H_1 成立时, 说明自变量 x_j 对 y 有作用, 在回归模型中不能去掉.

$\hat{\beta}_j \sim N(\beta_j, c_{jj}\sigma^2)$, c_{jj} 是 $C = (X^T X)^{-1}$ 的主对角线上第 $j+1$ 个元素

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{C_{jj}\sigma^2}} \sim N(0, 1) \quad (19)$$

而 $\frac{S_e}{\sigma^2} \sim \chi^2(n-m-1)$, 且 S_e 与 $\hat{\beta}_j$ 独立, 则在 H_0 成立的条件下, 有

$$T_j = \frac{\hat{\beta}_j}{\sqrt{C_{jj}S_e/(n-m-1)}} = \frac{\hat{\beta}_j}{\sqrt{C_{jj}}\hat{\sigma}} \sim t(n-m-1) \quad (20)$$

对给定的显著水平 α ，查表得 $t_{\alpha/2}(n-m-1)$ ，计算统计量 T_j 的数值 t_j 。

若 $|t_j| \geq t_{\alpha/2}(n-m-1)$ 则拒绝 H_0 ，即认为 β_j 显著不为零。

若 $|t_j| < t_{\alpha/2}(n-m-1)$ 则接受 H_0 ，即认为 β_j 等于零。

2.3.3 复相关系数

对一个回归方程来说,对回归好坏程度的度量采用复相关系数度量。

$$\text{复相关系数定义} \quad R^2 = \frac{S_R}{S_T} = 1 - \frac{S_e}{S_T} \quad (21)$$

当残差平方和 S_e 越小，则复相关系数越大。

该指标反映了采用一组自变量 x_1, x_2, \dots, x_m 解释因变量 y 的程度。

$0 < R^2 \leq 1$ ，当 R^2 越接近 1，表示因变量 y 与各自变量 x_i 之间线性相关程度越强。

调整的复相关系数(Adjust R^2).其定义如下:

$$aR^2 = 1 - \frac{S_e / (n - m - 1)}{S_T / (n - 1)} \quad (22)$$

当 R^2 和 aR^2 越接近 1,表示因变量 y 与各自变量 x_i 之间线性相关程度越强.

3. 软件实现

3.1 SAS8软件求解过程

- 1). 启动SAS软件,鼠标点击Solutions->Analysis->Analyst,启动分析员.
- 2). 在弹出的表中输入数据,结果如图2.其中1~32行为32组试验数据

	X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2	Y3	Y4	Y5	Y6
1	133.02	73	180	80	125	125	81.1	90	165.81	140.13	-145.14	118.63	135.37	160.76
2	129.63	73	180	80	125	125	81.1	90	165.51	140.25	-144.92	118.7	135.33	159.98
3	158.77	73	180	80	125	125	81.1	90	167.93	138.71	-146.91	117.72	135.41	166.81
4	145.32	73	180	80	125	125	81.1	90	166.79	139.45	-145.92	118.13	135.41	163.64
5	120	78.596	180	80	125	125	81.1	90	164.94	141.5	-143.84	118.43	136.72	157.22
6	120	75.45	180	80	125	125	81.1	90	164.8	141.13	-144.07	118.82	136.02	157.5
7	120	90.487	180	80	125	125	81.1	90	165.59	143.03	-143.16	117.24	139.66	156.59
8	120	83.848	180	80	125	125	81.1	90	165.21	142.28	-143.49	117.96	137.98	156.96
9	120	73	231.39	80	125	125	81.1	90	167.43	140.82	-152.26	129.58	132.04	153.6
10	120	73	198.48	80	125	125	81.1	90	165.71	140.82	-147.08	122.85	134.21	156.23
11	120	73	212.64	80	125	125	81.1	90	166.45	140.82	-149.33	125.75	133.28	155.09
12	120	73	190.55	80	125	125	81.1	90	165.23	140.85	-145.82	121.16	134.75	156.77
13	120	73	180	75.857	125	125	81.1	90	164.23	140.73	-144.18	119.12	135.57	157.2
14	120	73	180	65.958	125	125	81.1	90	163.04	140.34	-144.03	119.31	135.97	156.31
15	120	73	180	87.258	125	125	81.1	90	165.54	141.1	-144.32	118.84	135.06	158.26
16	120	73	180	97.824	125	125	81.1	90	166.88	141.4	-144.34	118.67	134.67	159.28
17	120	73	180	80	150.71	125	81.1	90	164.07	143.03	-140.97	118.75	133.75	158.83
18	120	73	180	80	141.58	125	81.1	90	164.27	142.29	-142.15	118.85	134.27	158.37
19	120	73	180	80	132.37	125	81.1	90	164.57	141.44	-143.3	119	134.88	158.01
20	120	73	180	80	156.93	125	81.1	90	163.89	143.61	-140.25	118.64	133.28	159.12
21	120	73	180	80	125	138.88	81.1	90	166.35	139.29	-144.2	119.1	136.33	157.59
22	120	73	180	80	125	131.21	81.1	90	165.54	140.14	-144.19	119.09	135.81	157.67
23	120	73	180	80	125	141.71	81.1	90	166.75	138.95	-144.17	119.15	136.55	157.59
24	120	73	180	80	125	149.29	81.1	90	167.69	138.07	-144.14	119.19	137.11	157.65
25	120	73	180	80	125	125	60.582	90	162.21	141.21	-144.13	116.03	135.5	154.26
26	120	73	180	80	125	125	70.962	90	163.54	141	-144.16	117.56	135.44	155.93
27	120	73	180	80	125	125	64.854	90	162.7	141.14	-144.21	116.74	135.4	154.88
28	120	73	180	80	125	125	75.529	90	164.06	140.94	-144.18	118.24	135.4	156.68
29	120	73	180	80	125	125	81.1	104.84	164.66	142.27	-147.2	120.21	135.28	157.65
30	120	73	180	80	125	125	81.1	111.22	164.7	142.94	-148.45	120.68	135.16	157.63
31	120	73	180	80	125	125	81.1	98.092	164.67	141.56	-145.88	119.68	135.29	157.61
32	120	73	180	80	125	125	81.1	120.44	164.69	143.84	-150.34	121.34	135.12	157.64

图2 SAS数据输入图

3) 鼠标点击 Statistics->Regression->Linear...在弹出对话框中(见图 3),将左边文本框中将 8 个自变量 x_1, x_2, \dots, x_8 选入 Explanatory 框中,将因变量 y_1, y_2, \dots, y_6 选入 Dependent 框中.然后点击 OK 即可执行回归分析.

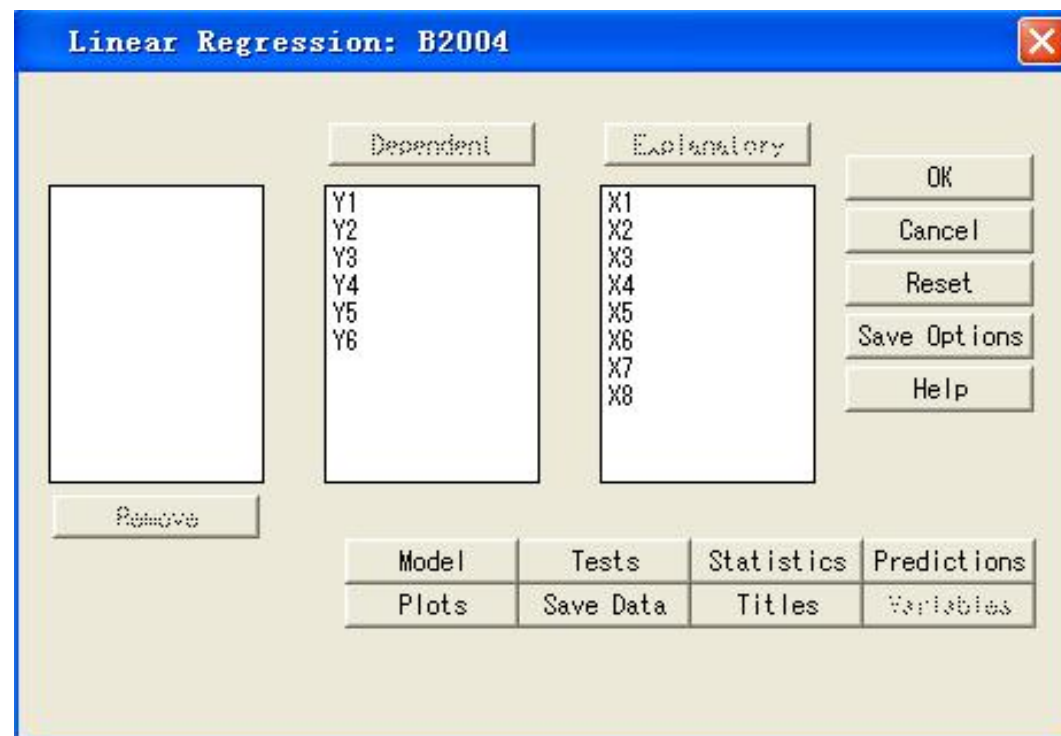


图3 SAS线性回归对话框

4) SAS进行回归分析结果见下面表3

表3

The REG Procedure
Dependent Variable: Y1
Analysis of Variance

	Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
	Model	8	60.73531	7.59191	5861.52	<.0001
	Error	23	0.02979	0.00130		
	Corrected Total	31	60.76510			
	Root MSE		0.03599	R-Square	0.9995	
	Dependent Mean		165.17031	Adj R-Sq	0.9993	
	Coeff Var		0.02179			
	Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	
Intercept	1	110.29651	0.44512	247.79	<.0001	
X1	1	0.08284	0.00084653	97.86	<.0001	
X2	1	0.04828	0.00191	25.21	<.0001	
X3	1	0.05297	0.00064256	82.44	<.0001	
X4	1	0.11993	0.00149	80.24	<.0001	
X5	1	-0.02544	0.00093315	-27.26	<.0001	
X6	1	0.12201	0.00126	96.45	<.0001	
X7	1	0.12158	0.00146	82.99	<.0001	
X8	1	-0.00123	0.00103	-1.19	0.2450	

从表中可以得到,总离差平方和 $S_T=60.76510$, 回归平方和 $S_R=60.73531$,残差平方和 $S_e=0.02979$; $F=5861.52$,而概率 $P\{F > 5861.52\} < 0.0001$,故不管取检验水平 $\alpha = 0.05$ 或 $\alpha = 0.1$ 都说明回归显著.

均方误差 $\hat{\sigma}^2 = 0.03599$,复相关系数 $R^2=0.9995$,调整的复相关系数 $aR^2=0.9993$.

该回归方程为:

$$y_1 = 110.29651 + 0.08284x_1 + 0.04828x_2 + 0.05297x_3 + 0.11993x_4 - 0.02544x_5 \\ + 0.12201x_6 + 0.12158x_7 - 0.00123x_8$$

从表可以看到常数项及 x_1, x_2, \dots, x_7 都通过了 T 检验, x_8 未通过 T 检验.但考虑该到实际问题, 8 台机器都为各线路的潮流值有贡献, 因此回归模型中考虑所有机组的出力.

SAS8 可以同时完成了 6 个回归模型参数及各指标的计算.上面只列出了 y_1 的回归计算.

3.2. SPSS软件求解过程

- 1). 启动SPSS软件,鼠标点击File->New->Data,启动数据编辑器。
- 2). 将表1中的后32组数据直接拷贝到数据编辑器的表格中。
- 3) 此时14列数据的变量名分别被系统自动命名为var00001到var00014, 鼠标点击数据表格下端的Variable View, 将前8个变量名修改为x1,x2,...,x8, 后6个变量名修改为y1,y2,...,y6。并将小数点显示为3位。再点Data View, 就可以看到图4所示数据表

	x1	x2	x3	x4	x5	x6	x7	x8	y1	y2	y3	y4	y5	y6
1	133.020	73.000	180.000	80.000	125.0	125.00	81.100	90.00	165.81	140.13	-145.140	118.6	135	160.760
2	129.630	73.000	180.000	80.000	125.0	125.00	81.100	90.00	165.51	140.25	-144.920	118.7	135	159.980
3	158.770	73.000	180.000	80.000	125.0	125.00	81.100	90.00	167.93	138.71	-146.910	117.7	135	166.810
4	145.320	73.000	180.000	80.000	125.0	125.00	81.100	90.00	166.79	139.45	-145.920	118.1	135	163.640
5	120.000	78.596	180.000	80.000	125.0	125.00	81.100	90.00	164.94	141.50	-143.840	118.4	137	157.220
6	120.000	75.450	180.000	80.000	125.0	125.00	81.100	90.00	164.80	141.13	-144.070	118.8	136	157.500
7	120.000	90.487	180.000	80.000	125.0	125.00	81.100	90.00	165.59	143.03	-143.160	117.2	140	156.590
8	120.000	83.848	180.000	80.000	125.0	125.00	81.100	90.00	165.21	142.28	-143.490	118.0	138	156.960
9	120.000	73.000	231.390	80.000	125.0	125.00	81.100	90.00	167.43	140.82	-152.260	129.6	132	153.600
10	120.000	73.000	198.480	80.000	125.0	125.00	81.100	90.00	165.71	140.82	-147.080	122.9	134	156.230
11	120.000	73.000	212.640	80.000	125.0	125.00	81.100	90.00	166.45	140.82	-149.330	125.8	133	155.090
12	120.000	73.000	190.550	80.000	125.0	125.00	81.100	90.00	165.23	140.85	-145.820	121.2	135	156.770
13	120.000	73.000	180.000	75.857	125.0	125.00	81.100	90.00	164.23	140.73	-144.180	119.1	136	157.200
14	120.000	73.000	180.000	65.958	125.0	125.00	81.100	90.00	163.04	140.34	-144.030	119.3	136	156.310
15	120.000	73.000	180.000	87.258	125.0	125.00	81.100	90.00	165.54	141.10	-144.320	118.8	135	158.260
16	120.000	73.000	180.000	97.824	125.0	125.00	81.100	90.00	166.88	141.40	-144.340	118.7	135	159.280
17	120.000	73.000	180.000	80.000	150.7	125.00	81.100	90.00	164.07	143.03	-140.970	118.8	134	158.830
18	120.000	73.000	180.000	80.000	141.6	125.00	81.100	90.00	164.27	142.29	-142.150	118.9	134	158.370
19	120.000	73.000	180.000	80.000	132.4	125.00	81.100	90.00	164.57	141.44	-143.300	119.0	135	158.010
20	120.000	73.000	180.000	80.000	156.9	125.00	81.100	90.00	163.89	143.61	-140.250	118.6	133	159.120
21	120.000	73.000	180.000	80.000	125.0	138.88	81.100	90.00	166.35	139.29	-144.200	119.1	136	157.590
22	120.000	73.000	180.000	80.000	125.0	131.21	81.100	90.00	165.54	140.14	-144.190	119.1	136	157.670
23	120.000	73.000	180.000	80.000	125.0	141.71	81.100	90.00	166.75	138.95	-144.170	119.2	137	157.590
24	120.000	73.000	180.000	80.000	125.0	149.29	81.100	90.00	167.69	138.07	-144.140	119.2	137	157.650
25	120.000	73.000	180.000	80.000	125.0	125.00	60.582	90.00	162.21	141.21	-144.130	116.0	136	154.260
26	120.000	73.000	180.000	80.000	125.0	125.00	70.962	90.00	163.54	141.00	-144.160	117.6	135	155.930
27	120.000	73.000	180.000	80.000	125.0	125.00	64.854	90.00	162.70	141.14	-144.210	116.7	135	154.880
28	120.000	73.000	180.000	80.000	125.0	125.00	75.529	90.00	164.06	140.94	-144.180	118.2	135	156.680
29	120.000	73.000	180.000	80.000	125.0	125.00	81.100	104.8	164.66	142.27	-147.200	120.2	135	157.650
30	120.000	73.000	180.000	80.000	125.0	125.00	81.100	111.2	164.70	142.94	-148.450	120.7	135	157.630
31	120.000	73.000	180.000	80.000	125.0	125.00	81.100	98.09	164.67	141.56	-145.880	119.7	135	157.610
32	120.000	73.000	180.000	80.000	125.0	125.00	81.100	120.4	164.69	143.84	-150.340	121.3	135	157.640

图4 SPSS数据输入图

4) 鼠标点击菜单Analyze->Regression->Linear...。弹出图5所示的线性回归对话框。将左边编辑框中的x1,x2,...,x8选入右边的Independent(s)编辑框中作回归分析的自变量。将y1选入右边的Dependent编辑框作因变量。

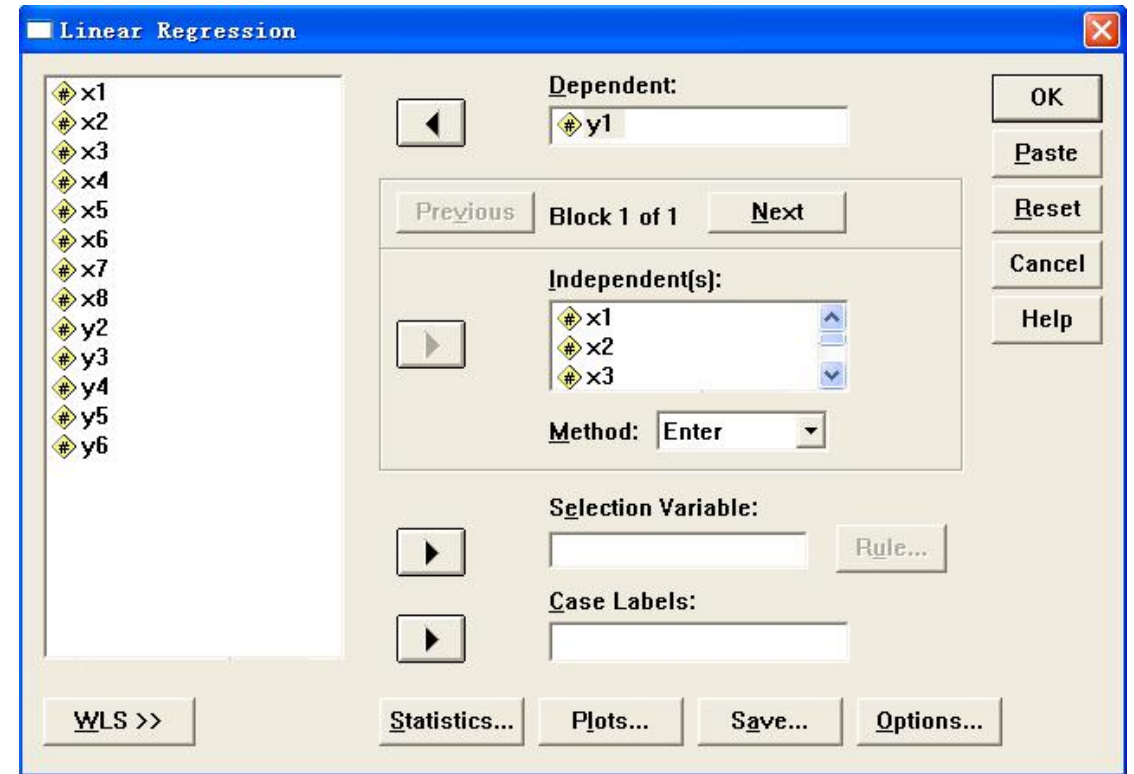


图5 SPSS线性回归对话框

5) 在线性回归对话框
中点按钮OK。
得到表4所示的回归
分析结果。

$$R^2 = 1, a.R^2 = 0.999, \text{RMSE} = 0.035989$$

$$F = 5861.519, P\{F > 5861.519\} \approx 0.000$$

$$y_1 = 110.297 + 0.08284x_1 + 0.04828x_2 + 0.05297x_3 + 0.120x_4 - 0.0254x_5 + 0.122x_6 + 0.122x_7 - 0.00123x_8$$

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	1.000 ^a	1.000	.999	3.5989E-02

a. Predictors: (Constant), X8, X4, X2, X3, X1, X5, X7, X6

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	60.735	8	7.592	5861.519	.000 ^a
	Residual	2.979E-02	23	1.295E-03		
	Total	60.765	31			

a. Predictors: (Constant), X8, X4, X2, X3, X1, X5, X7, X6

b. Dependent Variable: Y1

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	110.297	.445		247.791	.000
	X1	8.284E-02	.001	.495	97.860	.000
	X2	4.828E-02	.002	.127	25.213	.000
	X3	5.297E-02	.001	.417	82.438	.000
	X4	.120	.001	.372	80.238	.000
	X5	-2.54E-02	.001	-.139	-27.263	.000
	X6	.122	.001	.491	96.454	.000
	X7	.122	.001	.422	82.992	.000
	X8	-1.23E-03	.001	-.006	-1.193	.245

a. Dependent Variable: Y1

表4

3.2 线性回归的Matlab实现

Matlab中regress函数:

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

Y为列向量，X为矩阵，alpha为置信水平，缺省时取0.05.

b-----参数 β 的取值，列向量.

r-----残差向量，取值为Y-X.b.

rint-----残差的置信度为(1-alpha)的置信区间

stats----stats(1)为复相关系数,stats(2)为F值,stats(3)为F值对应的概率值.

附录1 问题1回归分析的MatLab程序

%围绕方案0的32组实验数据(8台机组出力)

x=[133.02 73	180	80	125	125	81.1	90
129.63 73	180	80	125	125	81.1	90
158.77 73	180	80	125	125	81.1	90
145.32 73	180	80	125	125	81.1	90
120 78.596	180	80	125	125	81.1	90
120 75.45	180	80	125	125	81.1	90
120 90.487	180	80	125	125	81.1	90
120 83.848	180	80	125	125	81.1	90
120 73	231.39	80	125	125	81.1	90
120 73	198.48	80	125	125	81.1	90
120 73	212.64	80	125	125	81.1	90
120 73	190.55	80	125	125	81.1	90
120 73	180	75.857	125	125	81.1	90
120 73	180	65.958	125	125	81.1	90
120 73	180	87.258	125	125	81.1	90
120 73	180	97.824	125	125	81.1	90

120	73	180	80	150.71	125	81.1	90
120	73	180	80	141.58	125	81.1	90
120	73	180	80	132.37	125	81.1	90
120	73	180	80	156.93	125	81.1	90
120	73	180	80	125	138.88	81.1	90
120	73	180	80	125	131.21	81.1	90
120	73	180	80	125	141.71	81.1	90
120	73	180	80	125	149.29	81.1	90
120	73	180	80	125	125	60.582	90
120	73	180	80	125	125	70.962	90
120	73	180	80	125	125	64.854	90
120	73	180	80	125	125	75.529	90
120	73	180	80	125	125	81.1	104.84
120	73	180	80	125	125	81.1	111.22
120	73	180	80	125	125	81.1	98.092
120	73	180	80	125	125	81.1	120.44];

%围绕方案0的32组实验数据(6条线路的潮流值)

y=[165.81	140.13	-145.14	118.63	135.37	160.76
165.51	140.25	-144.92	118.7	135.33	159.98
167.93	138.71	-146.91	117.72	135.41	166.81
166.79	139.45	-145.92	118.13	135.41	163.64
164.94	141.5	-143.84	118.43	136.72	157.22
164.8	141.13	-144.07	118.82	136.02	157.5
165.59	143.03	-143.16	117.24	139.66	156.59
165.21	142.28	-143.49	117.96	137.98	156.96
167.43	140.82	-152.26	129.58	132.04	153.6
165.71	140.82	-147.08	122.85	134.21	156.23
166.45	140.82	-149.33	125.75	133.28	155.09
165.23	140.85	-145.82	121.16	134.75	156.77
164.23	140.73	-144.18	119.12	135.57	157.2
163.04	140.34	-144.03	119.31	135.97	156.31
165.54	141.1	-144.32	118.84	135.06	158.26
166.88	141.4	-144.34	118.67	134.67	159.28
164.07	143.03	-140.97	118.75	133.75	158.83
164.27	142.29	-142.15	118.85	134.27	158.37

164.57	141.44	-143.3	119	134.88	158.01
163.89	143.61	-140.25	118.64	133.28	159.12
166.35	139.29	-144.2	119.1	136.33	157.59
165.54	140.14	-144.19	119.09	135.81	157.67
166.75	138.95	-144.17	119.15	136.55	157.59
167.69	138.07	-144.14	119.19	137.11	157.65
162.21	141.21	-144.13	116.03	135.5	154.26
163.54	141	-144.16	117.56	135.44	155.93
162.7	141.14	-144.21	116.74	135.4	154.88
164.06	140.94	-144.18	118.24	135.4	156.68
164.66	142.27	-147.2	120.21	135.28	157.65
164.7	142.94	-148.45	120.68	135.16	157.63
164.67	141.56	-145.88	119.68	135.29	157.61
164.69	143.84	-150.34	121.34	135.12	157.64];

x0=[120 73,180,80,125,125,81.1,90]';%方案0的8台机组出力

y0=[164.78,140.87,-144.25,119.09,135.44,157.69]';

%方案0的6条线路的潮流值

```

yp=zeros(6,1); err=zeros(6,1);
X=[ones(32,1),x]; alpha=0.05;
for i=1:6 %考虑6条线路分别进行回归分析
    Y=y(:,i); %获得第i条线路潮流值
    [b,bint,r,rint,stats]=regress(Y,X,alpha);%回归函数
    fprintf('第%2d条线路回归方程参数:\n',i);
    fprintf('系数:');
    for k=1:9 fprintf('%8.5f ',b(k)); end ; fprintf('\n');
    fprintf('统计量值R^2=%8.4f,F=%8.4f,p=%8.5f\n',stats(1),stats(2),stats(3));
    temp=b(2:9);
    yp(i)=b(1)+sum(temp.*x0);%计算方案0中对第i条线路潮流预测值
    err(i)=abs(yp(i)-y0(i))/abs(y0(i))*100; %计算预测相对误差的百分比
end
fprintf('方案0的原始值,预测值,相对误差百分比:\n');
for i=1:6
    fprintf('%8.4f %8.4f %8.4f\n',y0(i),yp(i),err(i));
end

```

例2 某种水泥在凝固时放出的热量Y (Cal)与水泥中下列4种化学成份有关:

x_1 :3CaO·Al₂O₃ x_2 :3CaO·SiO₂

x_3 :4CaO·Al₂O₃·Fe₂O₃ x_4 :2CaO·SiO₂

通过试验得到数据列于表5中
求Y对 x_1, x_2, x_3, x_4 回归方程

表 5 水泥放热数据表

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

利用SAS8.0求解过程如下:

- 1. 启动SAS软件,鼠标点击Solutions->Analysis->Analyst,启动分析员.
- 2. 在弹出的表中输入数据,结果如图6

	x1	x2	x3	x4	Y
1	7	26	6	60	78.5
2	1	29	15	52	74.3
3	11	56	8	20	104.3
4	11	31	8	47	87.6
5	7	52	6	33	95.9
6	11	55	9	22	109.2
7	3	71	17	6	102.7
8	1	31	22	44	72.5
9	2	54	18	22	93.1
10	21	47	4	26	115.9
11	1	40	23	34	83.8
12	11	66	9	12	113.3
13	10	68	8	12	109.4

图6 SAS数据输入图

1. 鼠标点击Statistics->Regression->Linear...在弹出对话框中(见图7), 将左边文本框中将四个自变量x1,x2,x3,x4选入Explanatory框中, 将因变量Y选入Dependent框中.然后点击OK即可执行回归分析.

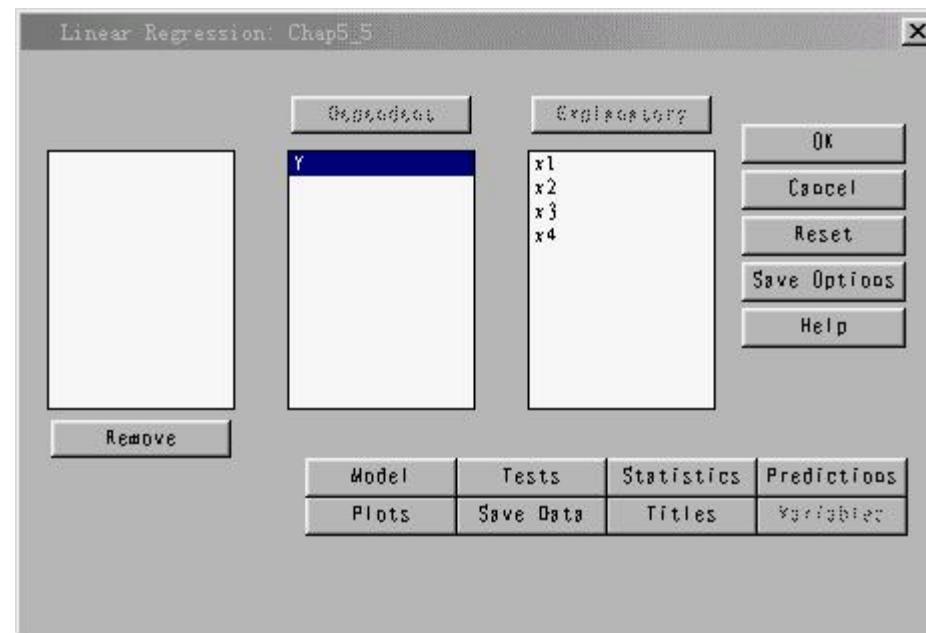


图7 SAS线性回归对话框

4. SAS回归分析结果见表6

$$\hat{\sigma}^* = 2.44601$$

$$R^2 = 0.9824$$

$$aR^2 = 0.9736$$

在参数估计中,大多数不能通过显著性检验,因此回归方程有问题。

表6 SAS回归分析结果表

The REG Procedure					
Dependent Variable: Y					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2667.89944	666.97486	111.48	<.0001
Error	8	47.86364	5.98295		
Corrected Total	12	2715.76308			
Root MSE		2.44601	R-Square	0.9824	
Dependent Mean		95.42308	Adj R-Sq	0.9736	
Coeff Var		2.56333			
Parameter Estimates					
Variable	DF	Estimate	Error	t Value	Pr > t
Intercept	1	62.40537	70.07096	0.89	0.3991
x1	1	1.55110	0.74477	2.08	0.0708
x2	1	0.51017	0.72379	0.70	0.5009
x3	1	0.10191	0.75471	0.14	0.8959
x4	1	-0.14406	0.70905	-0.20	0.8441

通过尝试或采用逐步回归的方法,采用去掉常数项最好,最后得到的结果见表7。

$$\hat{\sigma}^* = 2.41774$$

$$R^2 = 0.9996$$

$$aR^2 = 0.9994$$

$$y = 2.19305x_1 + 1.15333x_2 + 0.75851x_3 + 0.48632x_4$$

通过显著性检验

表7 SAS回归分析计算结果表

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	121035	30259	5176.47	<.0001
Error	9	52.60916	5.84546		
Uncorrected Total	13	121088			
Root MSE		2.41774	R-Square	0.9996	
Dependent Mean		95.42308	Adj R-Sq	0.9994	
Coeff Var		2.53370			
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
X1	1	2.19305	0.18527	11.84	<.0001
X2	1	1.15333	0.04794	24.06	<.0001
X3	1	0.75851	0.15951	4.76	0.0010
X4	1	0.48632	0.04141	11.74	<.0001

谢 谢！