

第三节 回归分析与相关分析实验

同一现象中的若干变量,往往存在着某种相互依存的关系.一般来说,这种关系并不能由一些变量完全决定另一些变量,它们之间存在的是一种统计关系,一般可以用回归分析与相关分析来研究.

一、一元直线回归与简单相关

仅有一个自变量的线性回归称为一元直线回归.设 x 为回归变量,是可以控制的变量; y 为响应变量,是服从正态分布的随机变量.一元直线回归的目的是确定一个回归方程 $y = \beta_0 + \beta x + \epsilon$,其中 β_0 为回归截距, β 为回归系数, ϵ 为随机误差.

例 1 某水稻品种在 5 月 5 日至 8 月 5 日播种(每隔 10 天播一期),测定播种至齐穗的天数 X 和播种至齐穗的总积温 Y (单位:日·度)的数据如下表,试确定 Y 与 X 的关系式.

X	70	67	55	52	51	52	51	60	64
Y	1616.3	1610.9	1440.0	1440.7	1423.3	1471.3	1421.8	1547.1	1533.0

(1) 在 SAS 中可用一般回归分析过程 REG 进行一元直线回归. REG 过程是一个通用回归过程,它利用最小二乘法来估计线性回归模型. REG 过程的一般调用格式如下

```
PROC REG 选项;
label:MODEL 因变量表 = 回归变量表/选项;
OUTPUT OUT = 数据集 关键字 = 名称表;
BY 变量表;
```

其中“PROC REG 选项;”中常用的选项有

```
DATA = 数据集:指定要分析的数据集,默认为最新建立的数据集;
ALL:要求输出各种输出项;
SIMPLE:为每个变量打印简单统计量;
NOPRINT:抑制正常的打印输出;
CORR:打印模型中所有变量的相关阵;
USSCP:打印所有变量的平方和及交叉积阵.
```

“Label:MODEL 因变量 = 回归变量/选项”;中 Label 是模型标号,可以省略. REG 过程可以有多个 MODEL(模型)语句,如果使用多个 MODEL 语句,可给予模型标号名称以便于区别. 该语句常用选项有

```
NOPRINT:抑制回归分析结果的打印输出;
NOINT:抑制模型中常数项的出现,缺省时模型中包括常数项;
I:打印 X'X 的逆矩阵;
XPX:打印 X'X 阵;
ALL:要求输出所有项;
P:打印观测值号、实测值、预测值及残差;
R:要求进行残差分析,包括预测值及残差的标准误、学生氏残差以及 Cook's 统计量 D;
CLM:打印每个观测值的因变量期望值的 95% 置信区间的上下限,给出参数估计的变异范围,而不是预测区间;
CLI:要求为每一个观测值打印 95% 置信度的上下限;
DW:要求计算 Durbin-Waston 统计量,检验误差是否有一阶自相关;
PARTIAL:要求打印每个回归变量的偏回归影响图;
```

“OUTPUT OUT = 数据集 参数表;”语句指定一个输出数据集,其中包括每

个观测值计算的统计量,统计量名称由 OUTPUT 语句之后的参数表指定,每个参数后的名称顺序与 MODEL 语句中的因变量顺序一致.常用参数项如下

- PREDICTED(或 P) = 观测值;
- RESIDUAL(或 R) = 残差,即为实测值与预测值之差;
- L95 = 每个预测值的 95% 置信区间下限;
- U95 = 每个预测值的 95% 置信区间上限;
- STDR = 残差的标准误;
- STDI = 单个观测值的标准误;
- STUDENT = 学生氏残差,其值为残差除以标准误.

“BY 变量表;”语句对指定的变量分组分别进行分析.

REG 过程输入的数据可以是原始样本数据,也可以是相关阵,还可以打印模型中的参数估计、置信区间估计、预测值及残差分析等,并可作线性假设检验.本例的分析程序如下

```
DATA LEX01;
INPUT x y@@@;
CARDS;
70 1616.3 67 1610.9 55 1440.0 52 1440.7 51 1423.3
52 1471.3 51 1421.8 60 1547.1 64 1533.0
PROC REG CORR;
MODEL y = x/p CLI;
RUN;
```

程序输出的主要分析结果有相关系数矩阵、方差分析表、参数估计、回归系数、t 检验值及置信度、实测值、预测值、残差和 95% 置信度的上下限.主要分析结果如下

Correlation					
	CORR	X	Y		
X		1.0000	0.9594		
Y		0.9594	1.0000		
Analysis of Variance					
Source	DF	Sumof Squares	Mean Square	F Value	Prob>F
Model	1	45167.28658	45167.28658	81.053	0.0001
Error	7	3900.78231	557.25462		
C Total	8	49068.06889			

Root MSE	23.60624	R-square	0.9205	C. V.	1.57324
Dep Mean	1500.48889	Adj R-sq	0.9091		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter = 0	Prob > T
INTERCEP	1	915.498799	65.45226849	13.987	0.0001
X	1	10.086036	1.12030262	9.003	0.0001

(2) 在 SPSS 中可用 Linear 过程进行线性回归分析,进行线性回归分析的一般步骤为:

- 在数据窗口中建立数据文件,定义变量并输入数据.
- 在“Statistics”菜单中,选择“Regression”中的“Linear”过程,打开线性回归主对话框.
- 在对话框的源变量框中选择一数值变量作为因变量进入“Dependent”框中;再选择一个或多个变量作为自变量进入“Independent”框中.
- 在“Method”框中选择一种变量分析方式.可供选择的方式有
 - ① Enter:是默认方式,所选择的自变量将全部进入回归方程中.
 - ② Remove:建立的回归方程将会根据设定的条件剔除部分自变量.
 - ③ Stepwise:根据“Options”对话框所设定的标准,在计算过程中逐步加入或剔除单个变量,直到所建立的方程中不能再加入变量或含有可剔除的变量为止.
 - ④ Backward:在计算过程中每次加入一个变量,然后再剔除某个变量,直到所建立的方程中不再含有可剔除的变量为止.
 - ⑤ Forward:根据“Options”对话框中设定的标准,在计算过程中一次加入单个变量,直至加入所有符合条件的变量为止.
- 根据需要,可单击“Statistics”按钮进行下列选择
 - ① Estimates:是默认选项,提供各变量之间的回归系数、相关统计,包括回归系数 B、标准误、标准化回归系数 beta、t 的双侧显著性水平检验、beta 的 t 值、t 的概率偏相关系数、最小容忍度等统计量.
 - ② Confidence intervals:每一个非标准化回归系数 95% 的置信区间.
 - ③ Covarice matrix:非标准化回归系数的方差 - 协方差矩阵.
 - ④ Descriptive:变量的均值、标准差、相关系数、单侧显著性水平矩阵.
 - ⑤ Model fit:是默认选项,提供相关系数 R、决定系数 R²、调整系数、估计标准误、ANOVA 的图表.

⑥ Block summary: 提供回归过程中每一步的统计量.

⑦ Durbin Watson: 德宾—沃森检验

⑧ Collinearity: 共线性诊断.

· 根据需要, 可单击“Options”按钮进行下列选择, 改变用于进行逐步回归时的内部数值的设定, 以及对缺值的处理方式

① Use probability of F: 如果一个变量的 F 显著性水平值小于所设定的进入值(Entry Value), 那么这个变量将会被选入方程式中; 如果它的 F 显著性水平值大于所设定的剔除值(Removal value), 那么这个变量将会被剔除. 所设定的进入值必须小于剔除值.

② Use F Value: 如果一个变量的 F 值大于所设定的进入值, 那么这个变量将会被选入回归方程中; 如果它的 F 值小于剔除值, 那么这个变量将会被剔除.

③ Include constant equation: 选择此项表示在回归方程中包括常数项.

④ Missing Values: 缺失值处理方式, 其中“Exclude cases Listwise”在使用所有带有合法值的变量的同时, 排除数据中的剩余观测量; “Exclude cases Pairwise”只使用那些数据完整并且具有相关关系, 用来计算相关系数的观测量; “Replace with mean”利用变量的平均数代替缺失值.

· 根据需要, 可单击“Plots”按钮, 用于绘制残差散点图、直方图、奇异值图或正态概率图.

· 根据需要, 可单击“Save”按钮, 每项选择都会增加一个或更多的新变量进入现有数据文件, 或建立新文件, 包括预测值, 残差等统计量.

· 单击主对话框中的“OK”按钮进行统计分析; “RESET”按钮可重新设置上述各种选择项.

本例中, 定义变量 X(天数)和 Y(积温), 并输入数据; 单击“Statistics”→“Regression”→“Linear”过程, 打开线性回归主对话框, 选 Y 为因变量, X 为自变量; 单击“Statistics”按钮, 选择“Estimates”、“Confidence Interval”、“Covariance Matrics”、“Model Fit”、“R Squared Change”、“Collinearity Diagnostic”和“Durbin-Watson”; 其他使用默认选择, 得如下结果

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.959	.921 ^a	.909	23.6062	.921	81.053	1	7	.000	2.888

a Predictors: (Constant), 天数

b Dependent Variable: 积温

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	45167.287	1	45167.287	81.053	.000 ^a
	Residual	3900.782	7	557.255		
	Total	49068.069	8			

a Predictors: (Constant), 天数 b Dependent Variable: 积温

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions ^a	
				(Constant)	天数
1	1	1.993	1.000	.00	.00
	2	7.253E-03	16.576	1.00	1.00

a Dependent Variable: 积温

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics		
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF	
1	常数	915.499	65.452		13.987	.000	760.729	1070.268		
	天数	10.086	1.120	.959	9.003	.000	7.437	12.735	1.000	1.000

a Dependent Variable: 积温

Coefficient Correlations^a

Model	天数	Model	天数
1	Correlations	1	Covariances
	天数	1.000	1.255

a Dependent Variable: 积温

Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1429.8866	1621.5214	1500.4889	75.1393	9
Residual	-30.2308	31.3273	2.526E-14	22.0816	9
Std. Predicted Value	-.940	1.611	.000	1.000	9
Std. Residual	-1.281	1.327	.000	.935	9

a Dependent Variable: 积温

二、相关分析与偏相关分析

相关分析是两个变量之间相似性程度的一种度量.简单相关是在不排除其他变量影响的情况下,两个变量所具有的线性关系,相关系数越大,这两个变量的线性依赖程度越强;偏相关是在排除了其他所有变量影响的前提下,两个变量之间的净相关,它反映了两个变量之间的真实相依程度.

例 2 九个品种的六个性状资料见下表,试作相关分析.

	冬季分蘖	株高	每穗粒数	千粒重	抽穗期	成熟期
7014-R ₀	11.5	95.3	26.4	39.2	4/9	6/2
7576/3矮7	9.0	97.7	30.8	46.8	4/17	6/6
68G1287	7.5	110.7	39.7	39.1	4/17	6/3
70190-1	9.1	89.0	35.4	35.3	4/18	6/2
9615-11	11.6	88.0	29.3	37.0	4/20	6/7
9615-13	13.0	87.7	24.6	44.8	4/19	6/7
73136	11.6	79.7	25.6	43.7	4/19	6/5
丰3	10.7	119.3	29.9	38.8	4/19	6/5
矮3	11.1	87.7	32.2	35.6	4/18	6/3

(1) 在 SAS 中可用 CORR 过程进行相关分析. CORR 过程的用法在实验六已详细介绍,这里不再赘述.本例中,若计算品种间的相关、性状间的相关与协方差,以及冬季分蘖与每穗粒数之间去除其他性状影响后的净相关,用下述程序来实现

```

data lex02;
  do i = 1 to 6;
    input x1 - x9@; output;
  end;
cards;
11.50 9.00 7.50 9.10 11.60 13.00 11.60 10.70 11.10
95.30 97.70 110.70 89.00 88.00 87.70 79.70 119.30 87.70
26.40 30.80 39.70 35.40 29.30 24.60 25.60 29.90 32.20
39.20 46.80 39.10 35.30 37.00 44.80 43.70 38.80 35.60
9.00 17.00 17.00 18.00 20.00 19.00 19.00 19.00 18.00
2.00 6.00 3.00 2.00 7.00 7.00 5.00 5.00 3.00
proc corr;

```

```

var x1 - x9;
run;
proc transpose out = xex20 prefix = y;
var x1 - x9;
proc corr cov;
var y1 - y6;
run;
proc corr;
var y1 y3;
partial y2 y4 - y6;
run;

```

主要执行结果如下

Correlation Analysis

9 'VAR 工' Variables: X1 X2 X3 X4 X5 X6 X7 X8 X9

Covariance Matrix DF = 5

	X1	X2	X3	X4	X5
X1	185.402667	1178.104000	1348.558667	1059.936667	1008.282000
X2	1178.104000	1185.591000	1349.036000	1062.904000	1010.669000
X3	1348.558667	1349.036000	1573.094667	1237.978667	1164.592000
X4	1059.936667	1062.904000	1237.978667	978.030667	916.992000
X5	1008.282000	1010.669000	1164.592000	916.992000	870.663000
X6	1020.757333	1025.871000	1157.181333	912.253333	874.233000
X7	931.216667	939.960000	1059.282667	838.192667	799.050000
X8	1435.292667	1429.477000	1654.910667	1297.508667	1237.877000
X9	1032.950667	1033.690000	1198.414667	945.944667	891.336000

	X6	X7	X8	X9
X1	1020.757333	931.216667	1435.292667	1032.950667
X2	1025.871000	939.960000	1429.477000	1033.690000
X3	1157.181333	1059.282667	1654.910667	1198.414667
X4	912.253333	838.192667	1297.508667	945.944667
X5	874.233000	799.050000	1237.877000	891.336000
X6	894.657667	818.133333	1238.582333	890.839333
X7	818.133333	751.634667	1125.518667	816.732667

X8	1238.582333	1125.518667	1773.309667	1263.870667
X9	890.839333	816.732667	1263.870667	917.214667

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
X1	6	30.56667	34.42968	183.40000	2.00000	95.30000
X2	6	34.55000	34.43241	207.30000	6.00000	97.70000
X3	6	36.16667	39.66226	217.00000	3.00000	110.70000
X4	6	31.46667	31.27348	188.80000	2.00000	89.00000
X5	6	32.15000	29.50700	192.90000	7.00000	88.00000
X6	6	32.68333	29.91083	196.10000	7.00000	87.70000
X7	6	30.76667	27.41596	184.60000	5.00000	79.70000
X8	6	37.11667	42.11068	222.70000	5.00000	119.30000
X9	6	31.26667	30.28555	187.60000	3.00000	87.70000

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Y1	9	10.56667	1.70441	95.10000	7.50000	13.00000
Y2	9	95.01111	12.59975	855.10000	79.70000	119.30000
Y3	9	30.43333	4.86081	273.90000	24.60000	39.70000
Y4	9	40.03333	4.12765	360.30000	35.30000	46.80000
Y5	9	17.33333	3.27872	156.00000	9.00000	20.00000
Y6	9	4.44444	2.00693	40.00000	2.00000	7.00000

Correlation Analysis

4 ' PARTIAL ' Variables: Y2 Y4 Y5 Y6

2 ' VAR ' Variables: Y1 Y3

Simple Statistics

Variable	N	Mean	Std Dev	Sum
Y2	9	95.011111	12.599746	855.100000
Y4	9	40.033333	4.127651	360.300000

Y5	9	17.333333	3.278719	156.000000
Y6	9	4.444444	2.006932	40.000000
Y1	9	10.566667	1.704406	95.100000
Y3	9	30.433333	4.860813	273.900000

Simple Statistics

Variable	Minimum	Maximum	Variance	Partial Std Dev
Y2	79.700000	119.300000	.	.
Y4	35.300000	46.800000	.	.
Y5	9.000000	20.000000	.	.
Y6	2.000000	7.000000	.	.
Y1	7.500000	13.000000	2.600548	1.612621
Y3	24.600000	39.700000	17.546053	4.188801

(2) 在 SPSS 中可用 Bivariate Correlations 过程进行变量间的简单相关分析,用 Partial Correlations 过程进行两个变量在去除其他变量影响后的偏相关分析.本例中,先在数据输入窗口定义变量 x1~x6,均为数值型,分别表示冬季分蘖、株高、每穗粒数、千粒重、抽穗期和成熟期,将品种作为观测值输入数据.在“Statistics”菜单中,选择“Correlate”中的“Bivariate”过程,打开简单相关主对话框,选择所有变量进行相关分析;打开“Options”选择框,选择统计框中的两个选项,返回并单击“OK”,得主要结果如下

Descriptive Statistics

	Mean	Std. Deviation	N
X1	10.5667	1.7044	9
X2	95.0111	12.5997	9
X3	30.4333	4.8608	9
X4	40.0333	4.1277	9
X5	17.3333	3.2787	9
X6	4.4444	2.0069	9

		Correlations					
		X1	X2	X3	X4	X5	X6
X1	Pearson Correlation	1.000	-.493	-.898**	.145	.042	.440
	Sig. (2-tailed)	.	.177	.001	.710	.914	.236
	Sum of Squares and Cross-products	23.240	-84.757	-59.530	8.150	1.900	12.033
	Covariance	2.905	-10.595	-7.441	1.019	.237	1.504
	N	9	9	9	9	9	9
X2	Pearson Correlation	-.493	1.000	.437	-.085	-.101	-.126
	Sig. (2-tailed)	.177	.	.240	.827	.796	.746
	Sum of Squares and Cross-products	-84.757	1270.029	214.077	-35.503	-33.333	-25.544
	Covariance	-10.595	158.754	26.760	-4.438	-4.167	-3.193
	N	9	9	9	9	9	9
X3	Pearson Correlation	-.898**	.437	1.000	-.471	.104	-.492
	Sig. (2-tailed)	.001	.240	.	.201	.791	.178
	Sum of Squares and Cross-products	-59.530	214.077	189.020	-75.580	13.200	-38.433
	Covariance	-7.441	26.760	23.628	-9.448	1.650	-4.804
	N	9	9	9	9	9	9
X4	Pearson Correlation	.145	-.085	-.471	1.000	.026	.568
	Sig. (2-tailed)	.710	.827	.201	.	.947	.110
	Sum of Squares and Cross-products	8.150	-35.503	-75.580	136.300	2.800	37.667
	Covariance	1.019	-4.438	-9.448	17.037	.350	4.708
	N	9	9	9	9	9	9

续表

		X1	X2	X3	X4	X5	X6
X5	Pearson Correlation	.042	-.101	.104	.026	1.000	.583
	Sig. (2-tailed)	.914	.796	.791	.947	.	.100
	Sum of Squares and Cross-products	1.900	-33.333	13.200	2.800	86.000	30.667
	Covariance	.237	-4.167	1.650	.350	10.750	3.833
	N	9	9	9	9	9	9
X6	Pearson Correlation	.440	-.126	-.492	.568	.583	1.000
	Sig. (2-tailed)	.236	.746	.178	.110	.100	.
	Sum of Squares and Cross-products	12.033	-25.544	-38.433	37.667	30.667	32.222
	Covariance	1.504	-3.193	-4.804	4.708	3.833	4.028
	N	9	9	9	9	9	9

* * Correlation is significant at the 0.01 level (2-tailed).

然后在“Statistics”菜单中,选择“Correlate”中的“Partial”过程,打开偏相关系数主对话框,相关系数变量选择 x_1 和 x_3 ,控制变量为 x_2, x_4, x_5 和 x_6 进行偏相关分析,得如下结果

```

- - - PARTIAL CORRELATION COEFFICIENTS - - -
Controlling for.. X6          X4          X5          X2
                  X1          X3
X1                1.0000        -.9267
                  ( 0)          ( 3)
                  P= .          P= .024
X3                -.9267        1.0000
                  ( 3)          ( 0)
                  P= .024        P= .
    
```

(Coefficient / (D.F.) / 2-tailed Significance)

"," is printed if a coefficient cannot be computed

在数据区再定义一变量 y , 字符型宽度为 2, 输入 $y_1 \sim y_9$. 单击“Data”菜单中的“Transpose”过程, 变量框输入 $x_1 \sim x_6$, 名称变量输入 y . 执行后将观测变为变量, 然后在“Statistics”菜单中, 选择“Correlate”中的“Bivariate”过程, 打开简单相关主对话框, 选择变量 $y_1 \sim y_9$ 进行相关分析, 其他使用默认选项. 返回并单击“OK”, 得结果如下

续表

		Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9
Y8	Pearson Correlation	.990	.986	.991	.985	.996	.983	.975	1.000	.991
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.001	.	.000
	N	6	6	6	6	6	6	6	6	6
Y9	Pearson Correlation	.991	.991	.998	.999	.997	.983	.984	.991	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.
	N	6	6	6	6	6	6	6	6	6

三、多元线性回归与复相关

在回归分析中如果涉及到多个回归变量时,就必须用多元线性回归方法寻求因变量与多个自变量之间的回归关系,并检验这种线性关系是否合宜,以及检验各个自变量对因变量是否是重要的,必要时可对自变量作适当的取舍.

1. 一般多元线性回归

一般多元线性回归用来检测一个因变量与一组独立变量之间的关系,并对这种关系的显著性进行检验.

例3 研究同一地区土壤内所含植物可给态磷情况得到如下数据表,其中 x_1 表示土壤内含无机磷的浓度, x_2 表示土壤内溶于 K_2CO_3 溶液并溶于溴化物水解的有机磷, x_3 表示土壤内溶于 K_2CO_3 溶液但不溶于溴化物的有机磷, y 表示栽在 $20^\circ C$ 土壤内的玉米中的可给态磷,单位均为百万分之一.已知 y 对 x_1, x_2, x_3 存在着线性相关,试求它们的回归方程,并进行回归方程显著性、回归系数显著性的检验.

土壤样本	x_1	x_2	x_3	y	土壤样本	x_1	x_2	x_3	y
1	0.4	53	158	64	10	12.6	58	112	51
2	0.4	23	163	60	11	10.9	37	111	76
3	3.1	19	37	71	12	23.1	46	114	96
4	0.6	34	157	61	13	23.1	50	134	77
5	4.7	24	59	54	14	21.6	44	73	93
6	1.7	65	123	77	15	23.1	56	168	95
7	9.4	44	46	81	16	1.9	36	143	94
8	10.1	31	117	93	17	26.8	58	202	168
9	11.6	29	173	93	18	26.9	51	124	99

(1) 在 SAS 中用 REG 过程进行多元线性回归分析.下述程序将 REG 过程

及 MODEL 语句的所有可能输出选项全部输出,程序如下

```
DATA LEX03;
INPUT x1 - x3 y@@@;
CARDS;
0.4 53 158 64 0.4 23 163 60 3.1 19 37 71
0.6 34 157 61 4.7 24 59 54 1.7 65 123 77
9.4 44 46 81 10.1 31 117 93 11.6 29 173 93
12.6 58 112 51 10.9 37 111 76 23.1 46 114 96
23.1 50 134 77 21.6 44 73 93 23.1 56 168 95
1.9 36 143 94 26.8 58 202 168 26.9 51 124 99
;
PROC REG ALL;
MODEL y = x1 - x3/ALL;
RUN;
```

程序执行主要结果如下:

Descriptive Statistics

Variables	Sum	Mean	Uncorrected SS	Variance	Std Deviation
INTERCEP	18	1	18	0	0
X1	212	11.7778	4150.62	97.278300654	9.8629762574
X2	758	42.1111	35076	185.63398693	13.624756399
X3	2214	123	307894	2092.4705882	45.743530562
Y	1503	83.5	137219	689.32352941	26.254971518

Uncorrected Sums of squares and Crossproducts

USSCP	INTERCEP	X1	X2	X3	Y
INTERCEP	18	212	758	2214	1503
X1	212	4150.62	9986.5	27273	20485.2
X2	758	9986.5	35076	96598	65265
X3	2214	27273	96598	307894	193262
Y	1503	20485.2	65265	193262	137219

Correlation

CORR	X1	X2	X3	Y
X1	1.0000	0.4635	0.1561	0.6322
X2	0.4635	1.0000	0.3175	0.3243

X3	0.1561	0.3175	1.0000	0.4111
Y	0.6322	0.3243	0.4111	1.0000

Model: MODEL1

Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	5892.58216	1964.19405	4.720	0.0177
Error	14	5825.91784	416.13699		
C Total	17	11718.50000			
Root MSE		20.39944	R - square	0.5028	
Dep Mean		83.50000	Adj R - sq	0.3963	
C. V.		24.43046			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter = 0	Prob > T
INTERCEP	1	45.913937	18.39490260	2.496	0.0257
X1	1	1.623780	0.56615994	2.868	0.0124
X2	1	-0.125949	0.42691151	-0.295	0.7723

Covariance of Estimates

COVB	INTERCEP	X1	X2	X3
INTERCEP	338.3724418	0.8075688402	-4.738873754	-1.017933718
X1	0.8075688402	0.3205370746	-0.106830195	-0.000683293
X2	-4.738873754	-0.106830195	0.1822534351	-0.013640639
X3	-1.017933718	-0.000683293	-0.013640639	0.0130114135

Correlation of Estimates

CORRB	INTERCEP	X1	X2	X3
INTERCEP	1.0000	0.0775	-0.6034	-0.4851
X1	0.0775	1.0000	-0.4420	-0.0106
X2	-0.6034	-0.4420	1.0000	-0.2801

X3 - 0.4851 - 0.0106 - 0.2801 1.0000

Test of First and Second Moment Specification

DF: 9 Chisq Value: 5.9514 Prob>Chisq: 0.7448

Dep Var Predict Std Err Lower95 % Upper95 % Lower95 % Upper95 %

Obs	Y	Value	Predict	Mean	Mean	Predict	Predict
1	64.0000	70.4161	10.899	47.0409	93.7912	20.8109	120.0
2	60.0000	75.1606	11.270	50.9886	99.3326	25.1750	125.1
3	71.0000	55.7036	11.947	30.0788	81.3283	4.9995	106.4
4	61.0000	72.9406	8.881	53.8929	91.9883	25.2218	120.7
5	54.0000	61.9226	9.624	41.2810	82.5642	13.5454	110.3
6	77.0000	64.2531	14.158	33.8878	94.6183	10.9959	117.5
7	81.0000	64.5236	10.364	42.2960	86.7512	15.4487	113.6
8	93.0000	81.0158	6.417	67.2533	94.7783	35.1499	126.9
9	93.0000	94.5234	10.217	72.6103	116.4	45.5902	143.5
10	51.0000	80.7086	8.539	62.3948	99.0223	33.2779	128.1
11	76.0000	80.3998	5.230	69.1823	91.6173	35.2323	125.6
12	96.0000	99.6561	7.727	83.0834	116.2	52.8701	146.4
13	77.0000	103.0	7.450	87.0376	119.0	56.4376	149.6
14	93.0000	89.5505	9.301	69.6022	109.5	41.4651	137.6
15	95.0000	108.8	8.624	90.3341	127.3	61.3288	156.3
16	94.0000	72.0946	7.557	55.8860	88.3033	25.4364	118.8
17	168.0	121.2	11.654	96.1612	146.1	70.7671	171.5
18	99.0000	107.1	9.047	87.7248	126.5	59.2666	155.0

Std Err Student Cook's

Obs	Residual	Residual	Residual	- 2	- 1	- 0	1	2	D
1	- 6.4161	17.244	- 0.372						0.014
2	- 15.1606	17.004	- 0.892		*				0.087
3	15.2964	16.535	0.925			*			0.112
4	- 11.9406	18.365	- 0.650		*				0.025
5	- 7.9226	17.987	- 0.440						0.014

6	12.7469	14.687	0.868		*		0.175
7	16.4764	17.571	0.938		*		0.076
8	11.9842	19.364	0.619		*		0.011
9	-1.5234	17.656	-0.086				0.001
10	-29.7086	18.526	-1.604		***		0.137
11	-4.3998	19.718	-0.223				0.001
12	-3.6561	18.879	-0.194				0.002
13	-26.0166	18.990	-1.370		**		0.072
14	3.4495	18.156	0.190				0.002
15	-13.8302	18.487	-0.748		*		0.030
16	21.9054	18.948	1.156		**		0.053
17	46.8445	16.743	2.798		*****		0.948
18	-8.1288	18.284	-0.445				0.012
Sum of Residuals			0				
Sum of Squared Residuals			5825.9178				
Predicted Resid SS (Press)			10822.9418				

(2) 在 SPSS 中,可用 Linear 过程进行多元线性回归分析.在数据编辑窗口中建立数据文件,定义数值型变量 x1, x2, x3 和 y,然后输入数据.在“Statistics”菜单中的“Regression”中选择线性回归过程“Linear”,打开主对话框,在主对话框中选择变量 y 为因变量,选择 x1, x2, x3 为自变量.回归方式“Method”选择默认方式“Enter”;参数统计“Statistics”选用全部估计选项;残差分析选择“Durbin - Watson”;“Options”使用默认值,最后单击“OK”按钮,得主要结果如下

Descriptive Statistics

	Mean	Std. Deviation	N
Y	83.5000	26.2550	18
X1	11.7778	9.8630	18
X2	42.1111	13.6248	18
X3	123.0000	45.7435	18

Correlations					
		Y	X1	X2	X3
Pearson Correlation	Y	1.000	.632	.324	.411
	X1	.632	1.000	.464	.156
	X2	.324	.464	1.000	.318
	X3	.411	.156	.318	1.000
Sig. (1 - tailed)	Y	.	.002	.095	.045
	X1	.002	.	.026	.268
	X2	.095	.026	.	.100
	X3	.045	.268	.100	.
N	Y	18	18	18	18
	X1	18	18	18	18
	X2	18	18	18	18
	X3	18	18	18	18

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin - Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.709 ^a	.503	.396	20.3994	.503	4.720	3	14	.018	1.790

a Predictors: (Constant), X3, X1, X2 b Dependent Variable: Y

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5892.582	3	1964.194	4.720	.018 ^a
	Residual	5825.918	14	416.137		
	Total	11718.500	17			

a Predictors: (Constant), X3, X1, X2 b Dependent Variable: Y

Coefficients ^a													
Model	Unstan- dardized Coefficients		Stand- ardized Coeffi- cients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics		
	B	Std. Error				Beta	Lower Bound	Upper Bound	Zero- order	Partial	Part	Toler- ance	VIF
1	常数	45.914	18.395		2.496	.026	6.461	85.367					
	X1	1.624	.566	.610	2.868	.012	.409	2.838	.632	.608	.540	.785	1.27
	X2	-.126	.427	-.065	-.295	.772	-1.042	.790	.324	-.079	-.056	.724	1.38
	X3	.193	.114	.337	1.694	.112	-.051	.438	.411	.412	.319	.899	1.112

a. Dependent Variable: Y

2. 多元逐步线性回归

在进行多元线性回归时,往往需要从全体自变量中找出作用最显著的部分自变量来建立回归方程,这一过程一般是通过不断引入和剔除自变量逐步建立回归方程的,因而称为多元逐步线性回归。

例 4 某地区二化螟的第一代成虫发生量 y 与四个因素有关,它们是

x_1 : 秋季积雪期限(单位为周); x_2 : 每年化雪日期(以 2 月 1 日为 1);

x_3 : 二月份平均气温(℃); x_4 : 三月份平均气温(℃);

y 为二化螟发生总量(头)。试建立二化螟发生总量的回归方程。

(1) 在 SAS 中可用逐步回归分析过程 STEPWISE 进行多元逐步线性回归。

STEPWISE 过程的一般调用格式如下

PROC STEPWISE 选项;

MODEL 因变量表 = 回归变量表/选项;

BY 变量表;

STEPWISE 过程中常用选项只有“DATA = 数据集”一个,它指定要分析的数据集,默认为最新建立的数据集。

“MODEL 因变量 = 回归变量/选项”,中常用选项有

NOINT: 抑制模型中常数项的出现,缺省时模型中包括常数项;

FORWARD 或 F: 从没有变量开始,每次引入一个最显著的变量,直到
剩余变量均不显著为止;

BACKWARD 或 B: 从含所有变量开始,每次剔除一个最不显著的变
量,直到剩余变量均显著为止;

STEPWISE:为默认方法,每次引入一个最显著的变量,然后剔除所有不显著的变量,直到不再有能引入或剔除的变量为止;

MAXR:从没有变量开始,每次引入一个 R^2 最大的变量,然后将引入的所有变量与未引入的所有变量比较,决定是否用能产生最大 R^2 的变量代替模型中的变量,直到没有可替换的变量为止,此法要求回归变量不超过 20 个;

MINR:与 MAXR 法类似,不同的是每次寻找最小 R^2 的变量.

“BY 变量表”;语句对指定的变量分组分别进行分析.

本例分别用 STEPWISE 法和 FORWARD 法进行分析,程序如下:

```
DATA LEX04;
  INPUT Y X1 - X4 @@;
  CARDS;
  9 10 26 0.3 3.6 7 12 26 -1.4 4.4
  34 14 40 -0.8 1.7 42 16 32 0.2 1.4
  40 19 51 -1.4 0.9 27 16 33 0.2 2.1
  4 7 26 2.7 2.7 27 7 25 1.0 4.0
  13 12 17 2.2 3.7 56 11 24 -0.8 3.0
  15 12 16 -0.5 4.9 8 7 16 2.0 4.1
  20 11 15 1.1 4.7
  ;
  PROC STEPWISE;
  MODEL Y = X1 - X4;
  MODEL Y = X1 - X4/F;
  RUN;
```

分析结果如下

Stepwise Procedure for Dependent Variable Y

Step 1 Variable X4 Entered R-square = 0.34747962 C(p) = 6.41323180

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	1	1077.98871014	1077.98871014	5.86	0.0340
Error	11	2024.31898217	184.02899838		
Total	12	3102.30769231			

VaXriable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	46.09702823	10.16942415	3781.28486500	20.55	0.0009
X4	-7.21508172	2.98110505	1077.98871014	5.86	0.0340

Bounds on condition number: 1, 1

Step 2 Variable X3 Entered R-square = 0.48725987 C(p) = 5.11147222

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	2	1511.63005719	755.81502860	4.75	0.0354
Error	10	1590.67763512	159.06776351		
Total	12	3102.30769231			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	43.76202872	9.55980728	3333.33370941	20.96	0.0010
X3	-4.58832279	2.77894344	433.64134705	2.73	0.1297
X4	-5.94374815	2.87653673	679.14596771	4.27	0.0657

Bounds on condition number: 1.077183, 4.308731

Step 3 Variable X2 Entered R-square = 0.64233432 C(p) = 3.44844723

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	3	1992.71871602	664.23957201	5.39	0.0213
Error	9	1109.58897629	123.28766403		
Total	12	3102.30769231			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	124.55477034	41.75660949	1096.95774436	8.90	0.0154
X2	-1.70329571	0.86225815	481.08865883	3.90	0.0796
X3	-8.99497009	3.31086133	909.99098087	7.38	0.0237
X4	-16.57748897	5.94904358	957.33253359	7.77	0.0212

Bounds on condition number: 7.880502, 47.39292

All variables left in the model are significant at the 0.1500 level.

No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y

Step	Entered	Removed	Number In	Partial R* * 2	Model R* * 2	C(p)	F	Prob>F
1	X4		1	0.3475	0.3475	6.4132	5.8577	0.0340
2	X3		2	0.1398	0.4873	5.1115	2.7261	0.1297
3	X2		3	0.1551	0.6423	3.4484	3.9022	0.0796

Forward Selection Procedure for Dependent Variable Y

Step 1 Variable X4 Entered R-square = 0.34747962 C(p) = 6.41323180

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	1	1077.98871014	1077.98871014	5.86	0.0340
Error	11	2024.31898217	184.02899838		
Total	12	3102.30769231			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	46.09702823	10.16942415	3781.28486500	20.55	0.0009
X4	-7.21508172	2.98110505	1077.98871014	5.86	0.0340

Bounds on condition number: 1, 1

Step 2 Variable X3 Entered R-square = 0.48725987 C(p) = 5.11147222

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	2	1511.63005719	755.81502860	4.75	0.0354
Error	10	1590.67763512	159.06776351		
Total	12	3102.30769231			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	43.76202872	9.55980728	3333.33370941	20.96	0.0010
X3	-4.58832279	2.77894344	433.64134705	2.73	0.1297
X4	-5.94374815	2.87653673	679.14596771	4.27	0.0657

Bounds on condition number: 1.077183, 4.308731

Step 3 Variable X2 Entered R-square = 0.64233432 C(p) = 3.44844723

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	3	1992.71871602	664.23957201	5.39	0.0213
Error	9	1109.58897629	123.28766403		
Total	12	3102.30769231			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
----------	--------------------	----------------	------------------------	---	--------

INTERCEP	124.55477034	41.75660949	1096.95774436	8.90	0.0154
X2	-1.70329571	0.86225815	481.08865883	3.90	0.0796
X3	-8.99497009	3.31086133	909.99098087	7.38	0.0237
X4	-16.57748897	5.94904358	957.33253359	7.77	0.0212
Bounds on condition number:		7.880502,	47.39292		

No other variable met the 0.5000 significance level for entry into the model.

Summary of Forward Selection Procedure for Dependent Variable Y

Step	Entered	In	Partial R ²	Model R ²	C(p)	F	Prob>F
1	X4	1	0.3475	0.3475	6.4132	5.8577	0.0340
2	X3	2	0.1398	0.4873	5.1115	2.7261	0.1297
3	X2	3	0.1551	0.6423	3.4484	3.9022	0.0796

(2) 在 SPSS 中,可用 Linear 过程进行多元逐步线性回归分析.在数据编辑窗口中建立数据文件,定义数值型变量 x1,x2,x3,x4 和 y,然后输入数据.在“Statistics”菜单中的“Regression”中选择线性回归过程“Linear”,打开主对话框,在主对话框中选择变量 y 为因变量,选择 x1,x2,x3,x4 为自变量.回归方式“Method”选择默认方式“Stepwise”;参数统计“Statistics”选用除描述统计与偏相关之外的全部估计选项,残差分析选择 Durbin - Watson;“Options”中引入变量概率选 0.13,剔除变量概率选 0.50,其他使用默认值.单击“OK”按钮,得主要结果如下

Model Summary^d

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin - Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.589 ^a	.347	.288	13.5657	.347	5.858	1	11	.034	
2	.698 ^b	.487	.385	12.6122	.140	2.726	1	10	.130	
3	.801 ^c	.642	.523	11.1035	.155	3.902	1	9	.080	2.141

a Predictors: (Constant), X4

b Predictors: (Constant), X4, X3

c Predictors: (Constant), X4, X3, X2

d Dependent Variable: Y

ANOVA^d

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1077.989	1	1077.989	5.858	.034 ^a
	Residual	2024.319	11	184.029		
	Total	3102.308	12			
2	Regression	1511.630	2	755.815	4.752	.035 ^b
	Residual	1590.678	10	159.068		
	Total	3102.308	12			
3	Regression	1992.719	3	664.240	5.388	.021 ^c
	Residual	1109.589	9	123.288		
	Total	3102.308	12			

a Predictors: (Constant), X4

b Predictors: (Constant), X4, X3

c Predictors: (Constant), X4, X3, X2

d Dependent Variable: Y

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics		
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF	
1	常数	46.097	10.169	4.533	.001	23.714	68.480			
	X4	-7.215	2.981	-.589	-2.420	.034	-13.776	-.654	1.000	1.000
2	常数	43.762	9.560	4.578	.001	22.461	65.063			
	X4	-5.944	2.877	-.486	-2.066	.066	-12.353	.466	.928	1.077
	X3	-4.588	2.779	-.388	-1.651	.130	-10.780	1.604	.928	1.077
3	常数	124.555	41.757	2.983	.015	30.095	219.015			
	X4	-16.577	5.949	-1.354	-2.787	.021	-30.035	-3.120	.168	5.944
	X3	-8.995	3.311	-.761	-2.717	.024	-16.485	-1.505	.507	1.973
	X2	-1.703	.862	-1.105	-1.975	.080	-3.654	.247	.127	7.881

a Dependent Variable: Y

若在选项“Options”中引入变量概率选 0.05, 剔除变量概率选 0.10, 其他使用默认值, 则得到的结果有所不同, 请同学们自己练习。

四、非线性回归分析

在实际问题中,变量与变量之间的关系并非都是线性的,因而,常常需要解决变量间的非线性回归关系问题.由于任一连续可导函数都可以分段用多项式来逼近它,故多项式回归是非线性回归中最常见的回归问题.

1. 多项式回归分析

一般的多项式回归可以是 n 次多项式,但最常见的多项式回归还是二次多项式回归.多项式回归一般包含常数项、一次项、二次项等,还可以有交叉项,各种项都包含的 m 元 n 次多项式是比较复杂的.

例 5 现有不同月龄西门答尔去势公犊肌肉占胴体比例资料如下表,试用 X 的二次式来拟合 Y ,并作显著性检验.

去势公犊月龄(X)	1	7	13	19	25
肌肉占胴体比例(Y)	55	66	68	65	55

(1) 在 SAS 中可用多项式回归分析过程 RSREG 进行一元多项式回归分析.RSREG 过程的回归模型中可包含自变量的平方项及交叉乘积项,因此又称二次响应曲面拟合模型.SAS 中的其他过程,如 GLM 过程也可解决响应曲面问题,但 RSREG 过程更有效.RSREG 过程的调用格式如下

```
PROC RSREG 选项;
MODEL 因变量 = 自变量/选项
BY 变量表;
```

其中“PROC RSREG 选项;”语句中的选项只有 DATA = 数据集,指定输入数据集,缺省时为最新建立的数据集.

“MODEL 因变量 = 自变量/选项;”语句中常用的选项有

LACKFIT:要求进行失拟检验.若选此项,则须先将自变量排序.

COVAR = n :指定模型中前 n 个自变量不含二次项.

NOPRINT:抑制打印方差分析及典型分析.

“BY 自变量表;”语句对由变量所定义的观测值分组分别进行分析.

本例的分析程序如下

```
DATA LEX05;
INPUT X Y@@;
CARDS;
1 55 7 66 13 68 19 65 25 55
;
PROC RSREG;
```

MODEL Y = X;

RUN;

程序运行结果如下

Coding Coefficients for the Independent Variables

Factor	Subtracted off	Divided by
X	13.000000	12.000000

Response Surface for Variable Y

Response Mean	61.800000
Root MSE	0.676123
R - Square	0.9942
Coef. of Variation	1.0941

	Degrees of Freedom	Type I Sum of Squares	R - Square	F - Ratio	Prob > F
Regression					
Linear	1	0.100000	0.0006	0.219	0.6860
Quadratic	1	157.785714	0.9936	345.2	0.0029
Crossproduct	0	0	0.0000	.	.
Total Regress	2	157.885714	0.9942	172.7	0.0058

	Degrees of Freedom	Sum of Squares	Mean Square
Residual			
Total Error	2	0.914286	0.457143

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0:	Prob > T
INTERCEPT	1	52.971032	0.736949	71.879	0.0002
X	1	2.407937	0.135284	17.799	0.0031
X * X	1	-0.093254	0.005019	-18.578	0.0029

Parameter Estimate from Coded Data

INTERCEPT	68.514286
X	-0.200000
X * X	-13.428571

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F - Ratio	Prob > F
--------	--------------------	----------------	-------------	-----------	----------

X	2	157.885714	78.942857	172.7	0.0058
---	---	------------	-----------	-------	--------

Canonical Analysis of Response Surface

(based on coded data)

Critical Value

Factor	Coded	Uncoded
X	-0.007447	12.910638

Predicted value at stationary point 68.515030

Eigenvectors

Eigenvalues	X
-13.428571	1.000000

Stationary point is a maximum.

(2) 在 SPSS 中,可用 Curve Estimation 过程进行多项式回归分析.在数据编辑窗口中建立数据文件,定义数值型变量 x 和 y,然后输入数据.在“Statistics”菜单中的“Regression”中选择曲线估计过程“Curve Estimation”,打开主对话框,在主对话框中选择变量 y 为因变量,选择 x 为自变量.模型选择“Quadratic”、方程包含常数项、作图,并显示方差分析表.单击“OK”按钮,得结果如下

MODEL: MOD __ 1. Dependent variable.. Y Method.. QUADRATI

Listwise Deletion of Missing Data

Multiple R	.99712
R Square	.99424
Adjusted R Square	.98849
Standard Error	.67612

Analysis of Variance:

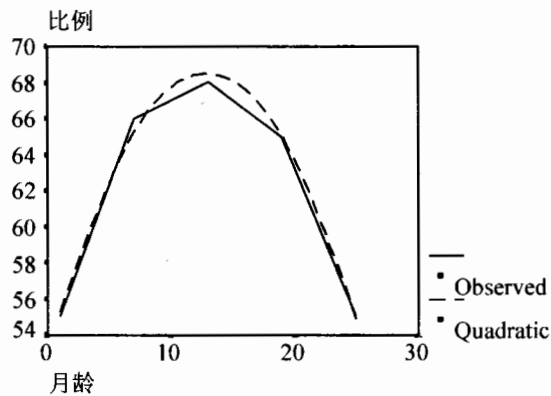
	DF	Sum of Squares	Mean Square
Regression	2	157.88571	78.942857
Residuals	2	.91429	.457143

F = 172.68750 Signif F = .0058

----- Variables in the Equation -----

Variable	B	SE B	Beta	T Sig T
X	2.407937	.135284	3.625526	17.799 .0031

X * * 2	-.093254	.005019	-3.784262	-18.578	.0029
(Constant)	52.971032	.736949		71.879	.0002



例6 有一大麦氮磷肥用量配比试验,施氮量 X_1 为五个水平,取值分别为 0.0, 2.5, 5.0, 7.5, 10.0; 施磷量 X_2 为四个水平,取值分别为 0, 2, 4, 6. 测得氮磷肥配比试验平均产量 Y , 试建立 Y 与 X_1 及 X_2 的二元二次回归方程.

(1) 在 SAS 中的分析程序如下

```
DATA LEX06;
INPUT X1 X2 Y@@@;
CARDS;
0.0 0 84.5 0.0 2 105.5 0.0 4 156.0 0.0 6 154.0
2.5 0 100.0 2.5 2 131.5 2.5 4 177.0 2.5 6 188.0
5.0 0 142.0 5.0 2 165.5 5.0 4 211.0 5.0 6 217.0
7.5 0 175.5 7.5 2 193.0 7.5 4 245.0 7.5 6 255.0
10.0 0 161.0 10.0 2 172.0 10.0 4 233.5 10.0 6 235.5
;
PROC RSREG;
MODEL Y = X1 X2;
RUN;
```

程序运行结果如下

Coding Coefficients for the Independent Variables		
Factor	Subtracted off	Divided by
X1	5.000000	5.000000
X2	3.000000	3.000000

Response Surface for Variable Y

Response Mean					175.125000	
Root MSE					15.111235	
R-Square					0.9284	
Coef. of Variation					8.6288	
Degrees of Type I Sum						
Regression	Freedom	of Squares	R-Square	F-Ratio	Prob > F	
Linear	2	39193	0.8775	85.817	0.0000	
Quadratic	2	2268.598214	0.0508	4.967	0.0234	
Crossproduct	1	5.445000	0.0001	0.0238	0.8795	
Total Regress	5	41467	0.9284	36.319	0.0000	
		Degrees of	Sum of			
Residual		Freedom	Squares	Mean Square		
Total Error		14	3196.891786	228.349413		
					Parameter	
	Degrees of	Parameter	Standard	T for H0:	Estimate from	
Parameter	Freedom	Estimate	Error	Parameter = 0	Prob > T	Coded Data
INTERCEPT	1	74.147857	11.119123	6.668	0.0000	191.825893
X1	1	18.047714	3.605044	5.006	0.0002	43.800000
X2	1	19.627500	5.704366	3.441	0.0040	42.435000
X1 * X1	1	-0.948571	0.323092	-2.936	0.0108	-23.714286
X2 * X1	1	0.066000	0.427410	0.154	0.8795	0.990000
X2 * X2	1	-0.968750	0.844744	-1.147	0.2707	-8.718750
	Degrees of	Sum of				
Factor	Freedom	Squares	Mean Square	F-Ratio	Prob > F	
X1	3	21158	7052.710238	30.886	0.0000	
X2	3	20314	6771.286667	29.653	0.0000	

Canonical Analysis of Response Surface

(based on coded data)

Critical Value

Factor	Coded	Uncoded
X1	0.975447	9.877233
X2	2.488929	10.466786
Predicted value at stationary point		265.997016

Canonical Analysis of Response Surface
(based on coded data)

Eigenvectors

Eigenvalues	X1	X2
- 8.702428	0.032956	0.999457
- 23.730608	0.999457	- 0.032956

Stationary point is a maximum.

(2) 在 SPSS 中,用 Nonlinear Regression 过程进行多项式回归分析.在数据编辑窗口中建立数据文件,定义数值型变量 x1, x2 和 y, 然后输入数据.在“Statistics”菜单中的“Regression”中选择非线性回归过程“Nonlinear”, 打开主对话框,在主对话框中选择变量 y 为因变量,在模型表达式中输入

$$p1 + p2 * x1 + p3 * x2 + p4 * x1 * x1 + p5 * x1 * x2 + p6 * x2 * x2$$

并给 p1~p6 输入初始值,可全部输 0,其他使用默认选项,最后单击“OK”按钮,得结果如下

All the derivatives will be calculated numerically.

Iteration	Residual SS	P1	P2	P3	P4
		P5	P6		
1	58038.7500	.000000000	.000000000	.000000000	.000000000
		.000000000	.000000000		
1.1	364146.0301	10.7539471	1.55727370	2.51674538	.158974750
		.359162932	.424924803		
2	364146.0301	10.7539471	1.55727370	2.51674538	.158974750
		.359162932	.424924803		
2.1	52547.21158	39.4397451	4.76797851	7.58702207	.429203488
		.843374271	1.16208763		
3	52547.21158	39.4397451	4.76797851	7.58702207	.429203488
		.843374271	1.16208763		

3.1	3682.973597	83.6159103	13.5013220	15.5285388	-.54750095
		.221568060	-.40947493		
4	3682.973597	83.6159103	13.5013220	15.5285388	-.54750095
		.221568060	-.40947493		
4.1	3196.891786	74.1478609	18.0477135	19.6274966	-.94857137
		.066000079	-.96874953		
5	3196.891786	74.1478609	18.0477135	19.6274966	-.94857137
		.066000079	-.96874953		
5.1	3196.891786	74.1478575	18.0477143	19.6275003	-.94857144
		.066000039	-.96875009		

Run stopped after 10 model evaluations and 5 derivative evaluations.

Iterations have been stopped because the relative reduction between successive residual sums of squares is at most SSSCON = 1.000E-08

Nonlinear Regression Summary Statistics Dependent Variable Y

Source	DF	Sum of Squares	Mean Square
Regression	6	654841.85821	109140.30970
Residual	14	3196.89179	228.34941
Uncorrected Total	20	658038.75000	
(Corrected Total)	19	44663.43750	

R squared = 1 - Residual SS / Corrected SS = .92842

Asymptotic 95 %

Parameter	Estimate	Std. Error	Confidence Interval	
			Lower	Upper
P1	74.147857513	11.119124694	50.299706887	97.996008139
P2	18.047714280	3.605044901	10.315661968	25.779766593
P3	19.627500321	5.704365984	7.392852095	31.862148548
P4	-.948571441	.323091791	-1.641534414	-.255608469
P5	.066000039	.427410398	-.850704092	.982704170
P6	-.968750095	.844743599	-2.780544919	.843044730

Asymptotic Correlation Matrix of the Parameter Estimates

	P1	P2	P3	P4	P5	P6
P1	1.0000	-.6445	-.5940	.3632	.5766	.3039
P2	-.6445	1.0000	.1332	-.8962	-.3557	.0000
P3	-.5940	.1332	1.0000	.0000	-.3746	-.8885
P4	.3632	-.8962	.0000	1.0000	.0000	.0000
P5	.5766	-.3557	-.3746	.0000	1.0000	.0000
P6	.3039	.0000	-.8885	.0000	.0000	1.0000

2. 一般非线性回归分析

一般常见的非线性回归分析大多是一个因变量,一个自变量,若干个参数,这种回归大多不能用线性回归的办法估计参数,即使有些可以化为线性处理,但得到的参数已经不是最佳回归方程的参数,只有用非线性最小二乘法才能得到较为理想的结果.

例 7 越冬红棉铃虫的化蛹进度见下表,试进行 Logistic 曲线回归.

<i>t</i>	5	10	15	20	25	30	35	40	45	50
<i>y</i>	3.5	6.4	14.6	31.4	45.6	60.4	75.2	90.2	95.4	97.5

(1) 在 SAS 中可用非线性回归分析过程 NLIN 进行非线性模型参数的最小二乘估计. NLIN 过程的一般调用格式如下

PROC NLIN 选项;

PARAMETERS 或 PARMS 参数 = 初值…;

MODEL 因变量表 = 表达式;

DER 参数 = 表达式;

OUTPUT OUT = 数据集 关键字 = 变量;

其中前三条语句是必须的;

“PROC NLIN 选项;”中常用的选项有

DATA = 数据集:指定要分析的数据集,默认为最新建立的数据集;

PLDT:当只有两个参数时,打印残差平方和的等值散点图;

METHOD = MARQUARDT | NEWTON | GAUSS | GRADIENT |

DUD:选用这五种迭代法之一,默认为有 DER 语句时用 GAUSS 法,否则用 DUD 法;

EFORMAT:以科学记数法打印所有数值;

MAXITER = n:指定迭代次数,默认为 50 次.

“PARAMETERS 参数 = 初值…;”指定模型中各个参数的初始值,初始值

可以是一个单值,也可以是多个用逗号隔开的指定值,还可以是由 TO 和 BY 连接的起止值,如 B0=1 TO 4,B1=0 TO 0.6 BY 0.2,B2=0,0.5,1 TO 4.

“MODEL 因变量 = 表达式;”指定非线性模型表达式;

“DER” 指定非线性模型参数的一、二阶偏导数表达式;

“DER. 参数 = 表达式;”指定参数的一阶偏导数表达式;

“DER. 参数. 参数 = 表达式;”指定参数的二阶偏导数表达式;

“OUTPUT OUT = 数据集 关键字 = 变量;”指定一个输出数据集,其中含有关键字表示的统计量,这些统计量的名称由等号右面的变量给出,其关键字可以是

PREDICTED 或 P:因变量的观测值;

RESIDUAL:残差值,即实际值 - 预测值;

L95:预测值的 95% 置信区间的下限;

U95:预测值的 95% 置信区间的上限;

STDI:观测值的标准误.

本例使用 Marquardt 迭代法的分析程序如下

```
DATA LEX07;
INPUT T Y@@@;
CARDS;
5 3.5 10 6.4 15 14.6 20 31.4 25 45.6
30 60.4 35 75.2 40 90.2 45 95.4 50 97.5
;
PROC NLIN METHOD = MARQUARDT;
PARMS K = 100 A = 60 B = 0.1;
MODEL Y = K/(1 + A * EXP(- B * T));
DER. K = 1/(1 + A * EXP(- B * T));
DER. A = - K * EXP(- B * T)/(1 + A * EXP(- B * T)) * * 2;
DER. B = K * A * T * EXP(- B * T)/(1 + A * EXP(- B * T)) * * 2;
RUN;
```

程序执行结果为:

Non - Linear Least Squares Iterative Phase
Dependent Variable Y Method: Marquardt

Iter	K	A	B	Sum of Squares
0	100.000000	60.000000	0.100000	7879.607726
1	114.786732	31.186891	0.110023	266.754259
2	109.827503	27.632540	0.115641	91.203516

3	102.615756	35.125968	0.132155	43.815059
4	101.857435	43.166880	0.141118	35.129898
5	101.965897	44.196099	0.141090	33.969996
6	101.993150	44.139058	0.141000	33.968964
7	101.992876	44.144105	0.141004	33.968962
8	101.992955	44.143670	0.141003	33.968962

NOTE: Convergence criterion met.

Non-Linear Least Squares Summary Statistics			Dependent Variable Y
Source	DF	Sum of Squares	Mean Square
Regression	3	39344.371038	13114.790346
Residual	7	33.968962	4.852709
Uncorrected Total	10	39378.340000	
(Corrected Total)	9	12317.536000	

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval	
			Lower	Upper
			K	101.9929553
A	44.1436702	8.3383277090	24.426499947	63.86084036
B	0.1410031	0.0084796534	0.120951735	0.16105444

Asymptotic Correlation Matrix

Corr	K	A	B
K	1	-0.537767562	-0.741637939
A	-0.537767562	1	0.9418241958
B	-0.741637939	0.9418241958	1

(2) 在 SPSS 中,用 Nonlinear Regression 过程进行非线性回归分析.在数据编辑窗口中建立数据文件,定义数值型变量 t 和 y,然后输入数据.在“Statistics”菜单中的“Regression”中选择非线性回归过程“Nonlinear”,打开主对话框,在主对话框中选择变量 y 为因变量,在模型表达式中输入:

$$K / (1 + A * \text{EXP}(- B * T))$$

并分别给 K,A,B 输入初始值 100,60,0.1,其他使用默认选项,最后单击“OK”按钮,得结果如下

All the derivatives will be calculated numerically.

Iteration	Residual SS	K	A	B
1	7879.607726	100.000000	60.0000000	.100000000
1.1	1560473.213	57.3205089	- 54.742872	.115637455
1.2	4675.530599	105.479329	51.8272552	.103265513
2	4675.530599	105.479329	51.8272552	.103265513
2.1	742.6670482	114.139364	37.5835790	.109681601
3	742.6670482	114.139364	37.5835790	.109681601
3.1	210.8804805	96.6909077	30.7867912	.128164273
4	210.8804805	96.6909077	30.7867912	.128164273
4.1	42.40389536	101.650656	42.7268100	.142346714
5	42.40389536	101.650656	42.7268100	.142346714
5.1	33.97501226	101.922941	44.1381858	.141091730
6	33.97501226	101.922941	44.1381858	.141091730
6.1	33.96896468	101.991612	44.1439293	.141005410
7	33.96896468	101.991612	44.1439293	.141005410
7.1	33.96896215	101.992926	44.1436822	.141003147
8	33.96896215	101.992926	44.1436822	.141003147
8.1	33.96896214	101.992949	44.1437072	.141003126

Run stopped after 17 model evaluations and 8 derivative evaluations.

Iterations have been stopped because the relative reduction between successive residual sums of squares is at most SSSCON = 1.000E - 08

Nonlinear Regression Summary Statistics		Dependent Variable Y	
Source	DF	Sum of Squares	Mean Square
Regression	3	39344.37104	13114.79035
Residual	7	33.96896	4.85271
Uncorrected Total	10	39378.34000	
(Corrected Total)	9	12317.53600	

R squared = 1 - Residual SS / Corrected SS = .99724

Parameter	Asymptotic 95 % Confidence Interval			
	Estimate	Std. Error	Lower	Upper
K	101.99294913	2.407850206	96.299288143	107.68661012
A	44.143707151	8.338333145	24.426682378	63.860731925
B	.141003126	.008479657	.120951923	.161054328

Asymptotic Correlation Matrix of the Parameter Estimates

	K	A	B
K	1.0000	-.5378	-.7416
A	-.5378	1.0000	.9418
B	-.7416	.9418	1.0000