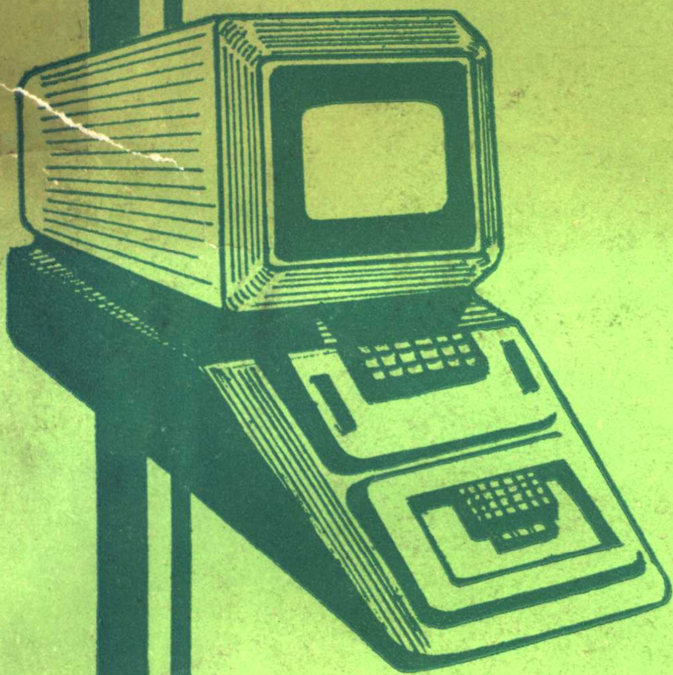

环境问题的数学解法及计算机应用

张孟威　康德梦　编著

中国环境科学出版社

环境问题的数学解法及计算机应用

张孟威　康德梦　编著

中国环境科学出版社

1989

内 容 简 介

本书较系统地介绍了环境保护问题常用的数学方法及相应的计算机程序。可用于制订环境监测方案及监测数据的分析与处理；为环境影响评价、环境污染的控制与预报、环境保护的规划提供了实用的数学方法；为环境管理、环境标准的制订、排污收费办法以及环境法规的制订提供了定量的依据。

本书侧重于基础知识与实际应用，亦编入了某些最新方法，因此，它既可作为初学者的教材，又可作为科研攻关有参考价值的资料。

本书可供从事环境保护、水文地质、气象及有关工业、农业、卫生部门的科研、设计、管理人员使用，亦可作为有关院校师生的参考书。

环境问题的数学解法及计算机应用

张孟威　康德梦　编著

责任编辑　李静华

*

中国环境科学出版社出版

北京崇文区东兴隆街69号

三河县二百户印刷厂印刷

新华书店总店科技发行所发行　　各地新华书店经售

*

1989年5月第一版　　开本　787×1092　1/16

1989年5月第一次印刷　　印张　15 3/4

印数　1—3000　　字数　370千字

ISBN 7-80010-204-1/X·151

定价：5.50元

前　言

当今环境保护问题的研究，已经从定性研究走向定性与定量相结合，从监测单一污染物发展到监测多种污染物，从微观分析进入到微观与宏观并举地进行研究的阶段。环保工作者要能够做到正确地处理环境监测数据，科学地制订环境标准，准确地提供环境污染预报，对环境进行现代化的管理，就需要很好地掌握数学解法与计算机应用。

环境污染问题从宏观上讲，是全球性的、区域性的，是具有时（间）、空（间）性的问题。污染物在时、空中进行着生物、化学、物理等作用的迁移转化，因此要求环境监测要多组分、多监测点和快速地进行测量。尽管如此，从数理统计观点认为，这些监测数据只是从实际环境中取得的一个子样品，必须用数理统计的方法，通过子样确定母体的污染状况。某些环境问题还具有模糊性，就需要采用模糊数学方法去认识它。采用不同的数学方法去处理监测数据，多数情况是需要借助于计算机。例如，研究湘江污染问题时，需要分别算出沿江800多公里长的多种结果。又如研究山西能源基地对华北地区大气污染问题，需要给出以污染源为中心，数百公里范围内的大气污染状况。这些问题的计算量之大，数据之多，只有用计算机才能胜任。

环境污染多是由于工业生产、人民生活、旅游等因素造成的。另外，由于水、气、土壤污染相互影响形成二次污染。因此，解决环境污染就必须由单一学科、单一行业发展到多学科、多行业协同作战。在环境治理和规划方面，则由单一治理污染源，发展到兼顾环境容量，即挖掘环境资源（环境自身消化污染物的能力）以及充分考虑到可能的治理投资等因素。对以上各点必须综合研究、系统分析，要采用近代数学的重要分支——系统工程学方法，其中包括有线性规划、非线性规划、动态规划以及图论、博奕论、决策论、排队论和存贮论等。环境科学的发展，促使环保工作者掌握新的数学方法去解决环境问题，环境科学与数学的密切结合就形成了环境问题的定量方法，这就是本书介绍的环境问题的数学解法。很显然，环境问题的数学解法只是一种简便的提法，是指对环境问题的一种分析方法，由这些数学方法得到的结果必须同定性分析相结合，才能得到解决环境问题切实可行的方案。

计算机已经发展到研制第五代计算机，它更新换代如此之快，原因在于人类社会发展需要它，当然也包括环境科学发展需要它。然而，目前在我国，计算机在环境保护中的应用尚未普及，众多环保工作者对计算机比较陌生。这是因为，环保问题的定量方法还没有被人们掌握，要求使用计算机的迫切性就不那么明显；其次是对计算机的基本知识尚不熟悉，尤其是对计算机程序设计不够熟悉。而计算机程序是将环境、数学、计算机三者融为一体，有了一批具有实际应用价值的，简便易行的计算机程序，会有助于环保工作者进一步认识环境问题，避免繁杂的数学推导，减轻程序调试的工作量，工作效率会显著提高。本书的目的就是为了促进计算机在环境保护中的应用。

作　者

1988.5

目　录

第二篇　环境污染的控制与预报

第三篇　环境保护问题的系统分析与规划

第一篇　环境监测数据的分析与处理

对于得到的大量监测数据，只有采用正确的数学方法去分析和处理，才能探索到科学规律，得到正确的结论。本篇将系统地介绍有关的数学方法及其计算机程序。究竟选用哪种方法来处理某一批监测数据，主要是根据监测目的及希望得到怎样的结论；另一方面需要提醒读者的是，为了使数据处理得到满意的结果，应在监测前，预先选定书中介绍的某种数学方法，参照该法制定监测方案是必要的。

第一章　监测数据的公式化

第一节　监测数据公式化的两种方法

将一组监测数据拟合成一个数学表达式，称为监测数据公式化，由于对所求公式的要求不同，可分为两种不同的情况，将分别采用两种不同的解法。

第一类情况　要求所得公式的计算值必须等于监测数据中的一个或数个数值，其几何意义是，监测点（实测值）通过曲线（公式），如图1-1。该类问题应采用第二节介绍的数据公式化插值解法。

第二类情况　只要求监测数据均最优地接近所求公式的计算值。几何意义是监测点均接近曲线（包括直线），如图1-2。这类问题需采用第二章介绍的回归分析法。

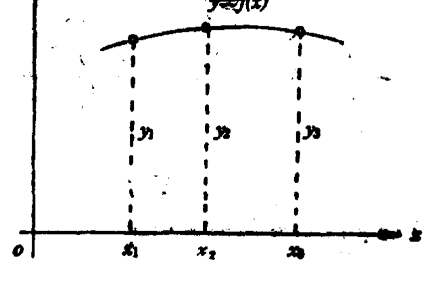

图 1-1　有限个监测点通过曲线（插值问题）

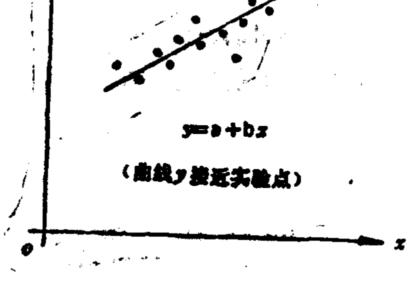

图 1-2　监测点接近曲线（回归问题）

第二节 监测数据公式化的插值法

一、要求通过两个实验点建立一个公式

已知 两个实验点（x_0, y_0）及（x_1, y_1）。

求 $y=f(x)$

根据已知条件，可得下列公式

$$y=y_0\cdot\left(\frac{x-x_1}{x_0-x_1}\right)+y_1\cdot\left(\frac{x-x_0}{x_1-x_0}\right) \tag{1-1}$$

式中 $\frac{x-x_1}{x_0-x_1}=y_0(x)$ 及 $\frac{x-x_0}{x_1-x_0}=y_1(x)$ 称为基函数

二、实 例

已知 河流水质环境评价中需要计算污染物的实测值C_i与标准值 L_i 的比值，即C_i/L_i，当河水中溶解氧为零时，$C_i/L_i=A_{max}$；溶解氧的实测值恰等于标准值时，即$C_i=L$时，$C_i/L_i=1$

求解 当溶解氧的实测值在零与标准值之间时，即$0\leqslant C_i\leqslant L_i$时，相对污染值的表达式。

解：按已知条件，要求两个特定点（0，A_{max}）及（L_i，1）必须通过所求的公式。将已知点代入式（1-1）得

$$(C_i/L_i)=A_{max}\cdot\left(\frac{C_i-L_i}{0-L_i}\right)+1\cdot\left(\frac{C_i-0}{L_i-0}\right)$$

整理后得

$$(C_i/L_i)=A_{max}\cdot(1-C_i/L_i)+C_i/L_i$$

图1-3表示了本例题在参考文献[1]中的应用

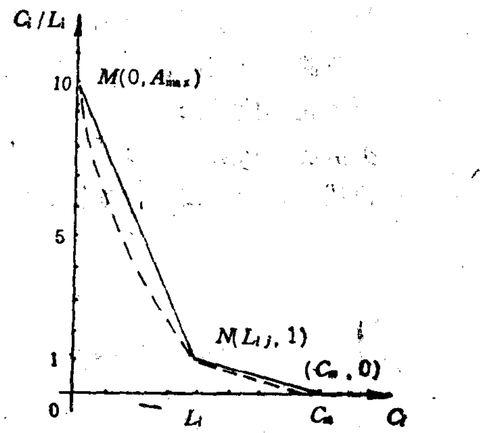

图 1-3 河流评价溶解氧污染值的插值式

第三节 通过三个实验点的插值公式

已知 三个实验点的数值分别为（x_0, y_0），（x_1, y_1），（x_2, y_2）。

求 建立一曲线方程式$y=f(x)$通过此三点。

插值公式

$$y = y_0 \cdot A_0(x) + y_1 \cdot A_1(x) + y_2 \cdot A_2(x) \tag{1-2}$$

其中 $A_0(x) = \dfrac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}$， $A_1(x) = \dfrac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}$，

$$A_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$$

第四节　通过$n+1$个实验点建立一个公式

已知　$n+1$个实验点的监测值（x_0，y_0），（x_1，y_1），…，（x_n，y_n）。

求　建立一个曲线方程（n次多项式）

$$y = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

使得上式通过以上$n+1$个实验点

确定上式的系数a_0，a_1，a_2，…，a_n由下列表达式给出：

$$y = \sum_{i=0}^{n} y_i A_i(x) \tag{1-3}$$

式中 $$A_i(x) = \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - x_k}{x_i - x_k}, \qquad i = 0,\ 1,\ \cdots,\ n$$

第二章　环境监测数据的线性方程式

第一节　环境监测数据线性方程式的计算方法

一、问题的提出

对于只有两个参数的监测数据，当需要研究两个参数之间的相关性时，可根据监测数据建立两者之间的线性方程式。如在大气环境污染问题中，风速往往是一个重要因素，若研究城区街道内风速（y）与气象站在距地面10m的平均风速（X）间的相关性，可建立其间的线性方程式$Y=a+bX$。又如在环境噪声问题中，噪声Y与道路上车流量X间的相关性，可表示为$Y=a+bX$。采用最小二乘法由监测数据确定方程中的系数a、b，该法的几何意义是在平面直角坐标系中，全部监测数据对应一批点，使得回归方程直线与这批点的距离之和为最小。其解析表达式将在计算方法中给出。在建立线性方程式的过程中，需研究两者之间是否具有线性关系，若具有线性关系，才能建立起两者间的线性方程式，否则只能建立两者间的曲线方程式。检验是否有线性关系是通过计算相关系数R，并作显著性检验来完成的。第三是要计算剩余标准差S，用S来说明回归方程计算y时，其精度如何。以上三点是线性方程式的基本计算。

二、线性方程式的计算方法

当给定n对监测数据：（x_1，y_1），（x_2，y_2），…（x_n，y_n），要求计算线性方程式

$$Y=a+bX \tag{2-1}$$

其计算方法如下：

（一）求回归系数a及b

$$b=\frac{L_{XY}}{L_{XX}}\text{及 } a=\overline{Y}-b\cdot\overline{X} \tag{2-2}$$

其中　均值 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}x_i$，$\overline{Y}=\frac{1}{n}\sum_{i=1}^{n}y_i$

X 的离差平方和 $L_{XX}=\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2\Big/n$

Y 的离差平方和 $L_{YY}=\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2\Big/n$

X，Y 的乘积和 $L_{XY}=\sum_{i=1}^{n}x_i\cdot y_i-\left(\sum_{i=1}^{n}x_i\right)\cdot\left(\sum_{i=1}^{n}y_i\right)\Big/n$

回归系数a及b的计算，根据最小二乘法，取误差函数J

$J=\sum_{i=1}^{n}[y_i-(a+bx_i)]^2$，使$J$达到极小值。按数学分析极值原理，解下列方程组：

$$\left.\begin{aligned}\frac{\partial J}{\partial a}&=-2\sum_{i=1}^{n}(y_i-a-bx_i)=0\\ \frac{\partial J}{\partial b}&=-2\sum_{i=1}^{n}(y_i-a-bx_i)x_i=0\end{aligned}\right\}$$

则可得回归系数a及b。

（二）求相关系数并作相关性检验

$$\text{相关系数 } R=\frac{L_{XY}}{\sqrt{L_{XX}\cdot L_{YY}}} \tag{2-3}$$

相关系数R表示了用已求得的回归方程来描述Y与X间线性关系密切程度。相关系数的取值范围是$0\leqslant|R|\leqslant1$而R值所表示的密切程度必须达到某个精度要求，才可以断定Y与X间已具有线性关系。这需要对R作显著性检验。检验的方法如下：

$$|R|\geqslant R^{a}_{N-2} \tag{2-4}$$

式中R^{a}_{N-2}由查相关系数检验表（附录Ⅰ）得到。其中上角标a称置信度，如$a=0.05$表示我们的判断有5％的可能性是错误的。下角标是监测数据组数N减2。当计算值R与查表值R^{a}_{N-2}满足式（2-4），则认为Y与X之间在a水平上，线性相关性是显著的，否则认为是不显著的。当不显著时，需要拟合成非线性式（见第三章）。

（三）求剩余标准差S

$$S=\sqrt{\frac{(1-R^2)\cdot L_{YY}}{N-2}} \tag{2-5}$$

S值表示用回归方程计算Y值的精度，S值的单位（量纲）与Y值相同。S值的作用是用回归方程计算Y，其精度由S来表示。

三、对监测数据的要求及数据的进一步处理

（一）对监测数据的要求

对监测数据进行处理其目的是从中找到规律，若使得找到的规律更加真实。且精度高，首先要求提高监测数据本身的精度（包括选用合理的监测方法，高灵敏度的仪器，以及减少观测误差等）；其次要求增加观测次数，即增加样品数；第三是尽量扩大监测数据中自变量的监测区间（即线性方程式中X的起始点的距离），监测区间越大，还可以增大回归方程式的应用区间（一般只允许回归方程式在监测区间中使用，而不能任意外推）。

（二）对监测数据的进一步处理

1．计算回归系数b的标准差S_b

$$S_b=\frac{S}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}}=\frac{S}{\sqrt{L_{XX}}} \tag{2-6}$$

式中符号含义如前。S_b的大小，可以说明监测数据的代表性如何。S_b越小，代表性越强。S_b越大，代表性越差。所谓代表性，是指在取样时，环境条件及监测方法不变的情况下，在监测过程中只包括随机因素的正常情况下，该批数据所反映的监测对象的客观规律。由上式可知，当监测区间比较大（越大）可使S_b值变小。否则，当监测区间较小时（即自变量x_1与x_n较近时），回归方程中的系数b的值就不会精确。

2．计算回归系数a的标准差S_a

$$S_a = S \cdot \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{L_{XX}}} \tag{2-7}$$

式中符号含义如前。S_a表示回归方程$Y = a + bX$中系数a的波动情况。求得的S_a值越小，说明a值越精确。由上式可知，S_a值与S值、L_{XX}有关，还与数据对数（组数）n有关。即监测区间越宽、数据组数n越大，所求的回归方程系数a越精确。

3．计算回归方程Y的标准差S_Y

$$S_Y = S \cdot \sqrt{1 + \frac{1}{n} + \frac{(X_i - \overline{X})^2}{L_{XX}}} \tag{2-8}$$

式中符号含义同前，S_Y表示根据回归方程$Y = a + bX$来预报Y值时的精度，S_Y值越小，精度越好。可知，当n相当大且X值离$\overline{X}$越近时，可使S_Y值变小，则预报精度越好，此时S_Y值可近似地以S值表示。

四、根据回归方程预报和控制的问题

根据监测数据可以计算出它们的回归方程式$Y = a + bX$，这个方程式主要是用于预报和控制Y的取值。

（一）用回归方程作预报

对于任意给定的X_0，相应的Y值由于随机因素，有可能出现不同的值，但它的平均值是由回归方程得到的，即$Y_0 = a + bX_0$，而Y的其他取值是以Y_0为中心对称分布（服从正态分布），愈靠近Y_0的地方出现的机率愈大，离Y_0较远的地方出现的机率就愈小，且与剩余标准差S之间有下列关系：

落在$Y_0 \pm 0.5S$的区间内约占38%

落在$Y_0 \pm S$的区间内约占68%

落在$Y_0 \pm 2S$的区间内约占95%

落在$Y_0 \pm 3S$的区间内约占99.7%

由上可知，当已知车流量(X)与环境噪声(Y)之间的回归方程为$Y = 59.45 + 0.018X$。若问当车流量为269辆/h，试预报环境噪声值：

$$Y = 59.45 + 0.018 \times 269 = 64.29\text{dB}$$

当已知剩余标准差为$S = 1.693$，则可预报知将可能有：约38%的Y_0值是在64.29 ± 0.85范围内

约68%的Y_0值是在64.29 ± 1.693范围内

约95%的Y_0值是在64.29 ± 3.386范围内

约99.7%的Y_0值是在64.29 ± 5.079范围内

（二）用回归方程进行控制的问题

由回归方程$Y = a + bX$控制Y的取值的问题，是指当给定Y值，来计算相应的X取值。

由回归方程可得 $$X_0 = \frac{1}{b}(Y_0 - a)$$

这类问题在环境污染控制中，有着重要意义。如根据车流量与环境噪声的回归方程式$Y = 59.45 + 0.018X$，若要求环境等效噪声控制在65dB，左右不超过3dB，那么每小时的车流量应控制在多少辆？

由上式知$X_0 = \frac{1}{0.018}(65 - 59.45) = 308$，即车流量控制在每小时308辆时，实际发生的环境噪声，将有95%的可能性为65 ± 3dB。这样的推算是否准确，还要取决于所求回归方程是否足够精确。为了使得回归方程足够精确，应当对有关参数进行研究（这些参数包括S、S_a、S_b、S_y），以便进一步研究监测区域是否足够大，以及监测数据取样个数n是否足够多。

第二节 环境监测数据线性方程实例

一、环境气象数据的回归计算

（一）环境气象数据实例

在大气环境污染问题中，风速往往是一个重要因素，在城镇气象数据中，常可从当地气象站了解到日平均风速（距地面10m高空处），而城镇街道距地面2m高空处的风速，常可由气象站测得的风速推算得到。推算方法是，预先测定n对风速数据，每对风速由日平均风速与街道风速组成，而且必须是同一天的（同步数据），然后求得这两种风速间的回归方程式。由回归方程式便可推算得街道风速。

例如某城镇实测的风速数据如表2-1，试计算两种风速间的线性方程式并作线性相关性检验，以及计算剩余标准差。

表 2-1 风速监测数据 单位（m/s）

编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
平均(X)	4.3	2.7	3.3	4.7	4.3	5.7	6.0	6.0	5.3	6.0	5.0	3.7	2.7	1.0	1.0	2.7	1.0	0.7	0.7	0.7
街道(Y)	3.0	3.5	3.5	4.0	4.5	2.5	4.0	4.0	4.5	4.5	3.5	2.2	1.8	1.2	1.0	2.0	1.2	1.0	1.0	0.4

（二）计算过程及计算结果

计算过程：将表2-1数据分别按公式（2-2）、（2-3）及（2-5）计算。本例由计算

机程序R01计算。步骤是：

1. 将数据输入到语句600至630
2. 查表（附表Ⅰ）得$R_{20-2}^{\alpha=0.01}=0.561$
3. 计算机运行时输入N＝20，RA＝0.561，ALPHA＝0.01

计算结果：

$$Y=0.604896+0.610401\cdot X$$

$$R=0.875496 \qquad S=0.691175$$

计算结果表示如图2-1。

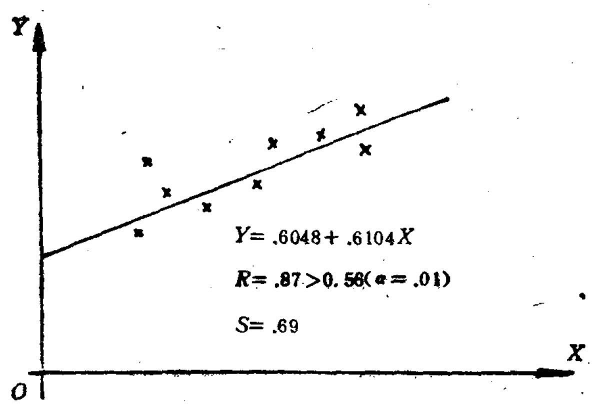

图 2-1 环境气象数据的线性方程

二、环境噪声数据的回归计算

（一）环境噪声数据实例

实测某城市街道机动车流量与噪声数据见表2-2。该数据为全天每小时的平均值。试求车流量与噪声之间的回归方程式及其相应的参数。

（二）计算过程及计算结果

1. 计算过程

（1）在计算机上调出已有的一元线性回归通用程序——R01，运行该程序，打印结果无误时，证明该程序可用。

（2）将表2-2数据输入计算机中，方法是修改程序中的DATA语句。修改顺序为

600 DATA 77.8，61.0，58.4，59.9…

（3）运行程序。由键盘输入以下数据N＝24，ALPHA＝0.05，RA＝0.424

2. 计算结果

$Y=59.59+0.018X$，$R=0.9494$，

$S=1.8219$，$S_a=4.738$，$S_b=1.28$

表 2-2　车流量与噪声实测值

监测项目 \ 取样时间	时间（第×小时）											
	1	2	3	4	5	6	7	8	9	10	11	12
车流量X_i（$\frac{辆}{h}$）	77.8	58.4	31.4	37.1	52.6	269	650.4	924.7	831.8	831.9	724.2	607.1
等效声级Y_i（dB）	61.0	59.9	55.8	58.3	60.2	66.7	71.5	73.0	73.6	74.0	72.5	72.0

监测项目 \ 取样时间	时间（第×小时）											
	13	14	15	16	17	18	19	20	21	22	23	24
车流量X_i（$\frac{辆}{h}$）	507.2	530.1	696.6	714.8	707.2	762.1	533.9	592.1	367.2	323.6	186.7	105.6
等效声级Y_i（dB）	68.0	71.8	71.6	71.8	73.6	72.6	71.2	71.8	69.7	65.4	64.0	62.1

第三节　环境监测数据线性方程程序
——一元线性回归通用程序

一、程序功能

已知一组数据（x_1，y_1），（x_2，y_2），…，（x_n，y_n）

求　线性方程式$Y=a+bX$；求相关系数R并作显著性检验；求剩余标准差S并打印出回归方程的绝对误差$Y-Y_i$，相对误差（$Y-Y_i$）/Y_i

二、计算步骤

计算步骤详见程序框图（图2-2）

三、程序中的变量标志符

输入变量　N——数据对数；RA——相关系数检验值；ALPHA——置信水平（以上两数查附表Ⅰ）；X(N，2）——监测数据（本程序调试数据为表2-1。）

输出变量　SX（N）-Σx及Σy；X-AVARGE-$\overline{X}$及$\overline{Y}$；ANOTHER——线性相关性不显著；OK-OK-OK——线性相关性显著；$Y=a+bX$——回归方程；R——相关系数；S——剩余标准差；$X-X_i$；$Y-Y_i$；Y^*-Y的计算值，$Y-Y^*$——实测值与计算值之差（即绝对误差）；$Y-Y^*/Y$——相对误差；SA——表示S_a，即回归系数a的标

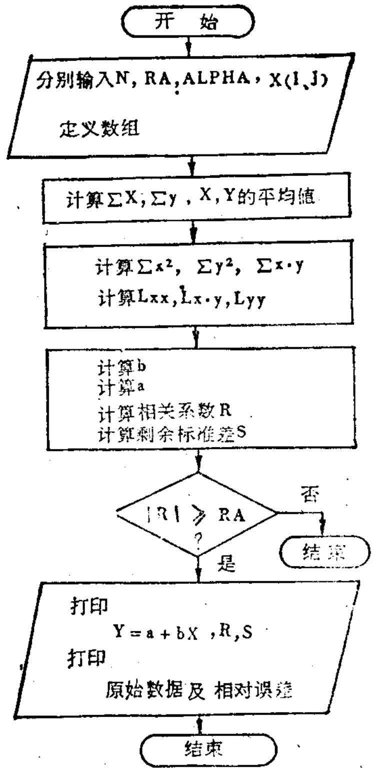

图 2-2 一元线性回归方程通用程序框图

准差；*SB*——表示S_b，即回归系数*b*的标准差；此外将显示S_y的值。

四、程序的使用方法

(一) 将程序输入到计算机中。运行后再输入N＝20；RA＝0.561；ALPHA＝0.01。运行结果若与本书结果相同，证明程序输入无误。否则应消除输入时的错误。

(二) 若有另一批新的数据，首先修改程序中的DATA语句，然后运行该程序，并依次输入相应的N，RA及ALPHA。

本程序调试数据详见第二节环境气象数据的回归计算。

本程序验证数据详见第二节二、环境噪声数据的回归计算。

五、环境监测数据线性方程式BASIC程序——一元线性回归通用程序——R01