

微型计算机应用丛书

统 计 程 序 包

马 速 成	编
高 文 培	审
范 兴 业	校

长 春 地 质 学 院

前 言

从六十年代初开始,地质学开创了一个新的时代,这就是地质学与数学及电子计算机的结合。在这一过程中,越来越多的地质方法要求应用概率统计的方法以解决地质学中的许多新课题。与此同时产生了一些新的学科,如数学地质科学。因此,使用计算机进行概率统计分析的计算也越来越普及。

自七十年代起,由于廉价的微型机的大量涌现,这就给地质数据处理带来了极大的方便。因此,本书的宗旨是考虑到微型机的应用已深入到地质科学的各个领域这一特点,利用统计分析的方法,在微型机上对地质数据进行多元统计分析和数据处理。在本书中,给出各种方法的一般原理、模型、计算公式、算法、计算流程和通用的FORTRAN IV源程序。这样便于用户在各种型号的计算机上使用本程序包。

本书虽然偏重微型机上地质统计程序包,但是,从数学角度看,各种统计分析方法和程序对其它科学有着广泛的应用。因此,也可以称为统计程序包。当然它不是统计学的全部,但是它包括了统计学的主要内容和方法。

各章内容的格式

第一章给出统计程序包的概述。其它各章按如下内容的格式描述:

职能: 给出所描述的分析方法和程序的职能。**算法:** 给出求解问题的数学描述和具体实施的各种方法。即给出数学公式和计算机实现这一问题的算法。有了编程的算法,阅读程序就比较方便了。这一部分的描述是较为详细的,因为它是进行程序设计的基础。对于误差分析和解线性代数方程组仍然启用

传统的算法，如用户感到精度不够，可以选用更好的方法。

程序说明、计算流程和程序

这一部分采用叙述的方式给出程序说明和计算流程图，便于读者了解程序中每一块的功能。最后给出完整的程序。

关于统计学程序包在国外的计算机上已是很成熟的成果，国内也有不少机器配备了统计学程序包。我们的目的是为开发微型计算机的应用，所以在微型机上建立统计学程序包。本书的内容主要参考〈1〉数字计算机上的数字方法第三卷（统计分析）；〈2〉Davis, J. C的《Statistics and data analysis in Gaology》；〈3〉P. M. Mather的《Computational Methods of Multivariate Analytis in Physical Geography》。大部分程序是从这些书上移植过来的。

值得提出的是：在我们搞这项任务之前，由徐中信、吴锡生付教授领导的化探数据处理科研组，曾在地质部北京计算中心的M—160机上做了很多工作，为我们搞这项工作打下了一定的基础；在初稿写完后，得到孙成栋、陈天与付教授的审阅，为本书提出了宝贵的意见，我们表示衷心的感谢。

为开发微型机的应用，我们编写这本程序包，对工程技术人员查阅公式、算法、流程和程序是很方便的。可供工科大专院校，尤其是地质院校师生做教学参考用书。

书中所提供的程序已在地质科学的某些分支得到了应用，但还未经过大量的考核，加上我们的水平所限，缺点和错误一定存在，诚恳地希望广大读者和用户提出批评指导。

编者

一九八三年初。于长春地质学院

目 录

第一章 统计程序概述	1
第二章 基本统计分析程序	3
职能	4
算法	4
一、数字统计计算及其统计检验	4
1 位置特征计算	4
2 散度特征计算	5
3 分布特征计算	6
4 相关特征计算	6
5 统计检验	7
二、经验分布计算及其拟合检验	8
1 划分分组区间	8
2 计算经验频率	9
3 拟合检验	9
三、实验数据的变换与校正	10
1 数据变换	10
2 过失误差校正	10
四、常用分布函数的数值计算	11
1 标准正态分布函数的计算	11
2 Γ 分布函数的计算	12
3 χ^2 分布函数的计算	13
4 T 分布函数的计算	15
5 F 分布函数的计算	18
程序说明, 计算流程和程序	21

第三章 方差分析程序	41
职能	41
算法	41
一、一种方式分组的方差分析	41
二、二种方式分组的方差分析	44
三、三重套合分析	46
程序说明, 计算流程和程序	49
第四章 回归分析程序	61
职能	61
算法	61
一、一元回归的几何意义和最小性质	62
二、用最小二乘法求一元线性回归方程	63
三、一元线性回归方程相关系数及其显著性检验	64
四、一元线性回归的方差分析	66
五、多元线性回归及方差分析	68
六、线性回归方程的某种推广	71
七、逐步回归及其检验	72
程序说明, 计算流程和程序	79
第五章 判别分析与聚类分析程序	97
职能	97
算法	98
程序说明, 计算流程和程序	104
第六章 其它常用的统计方法	112
一、主成份分析	112
二、因子分析	121
三、趋势分析	139

附录 1: 正交趋势分析程序.....	155
附录 2: 网格化程序.....	163
附录 3: 移动平均程序.....	166
第七章 宽行打图子程序.....	170
一、一元线性回归散点图.....	170
二、点群分析谱系图.....	175
三、统计分析直方图.....	183
四、主成份分析椭圆图.....	187
五、判别分析图.....	194
六、正交趋势图.....	205
七、正交趋势散点图.....	211
八、移动平均和网格化图.....	217
第八章: 各种专用子程序.....	223

程 序 清 单

- 1、正态分布函数计算: SUBROUTINE ZT (A,B,N)
- 2、T分布函数计算: SUBROUTINE GAMMA
 (X,Z)
- 3、 χ^2 分布函数计算: SUBROUTINE CF (KA,N,C
 H,CHS,D,L)
- 4、T分布函数计算: SUBROUTINE TF(IA,N,TL,
 CH,D,L)
- 5、F分布函数计算: SUBROUTINE FFB (IX,IY,
 N1,N2,F,N,P,D)
- 6、均值计算: FUNCTION AMEAN (C,K)
- 均方差计算: FUNCTION VR (B,N)

- 7、统计分析: PROGRAM TEST
- 8、一种方式分组方差分析: PROGRAM FORONEWAY
- 9、二种方式分组方差分析: PROGRAM FQRTWOWAY
- 10、三重套合方差分析: PROGRAM FQRTTHREEWAY
- 11、一元线性回归: PROGRAM ONEREG
- 12、一元线性回归并多项 PROGRAM REGRES
 式回归:
- 13、多元线性回归: PROGRAM MULR
- 14、多元线性并逐步回归: PROGRAM STEPRA
- 15、判别分析: PROGRAM DECIDE
- 16、聚类分析: PROGRAM CLUS
- 17、主成份分析: PROGRAM PCA
- 18、因子分析: PROGRAM FACT
- 19、趋势分析: PROGRAM TRENDM
- 20、正交趋势分析: PROGRAM TSADOP
- 21、网格化分析: PROGRAM NET
- 22、移动平均: PROGRAM MOVED
- 23、一元线性回归散点图: SUBROUTINE PLOTB (X1
 N1,X2,N2)
- 24、点群分析谱系图: SUBROUTINE DENDRO (I
 AIR,XLEV,M,M1,ISIM)
- 25、统计分析椭圆图: SUBROUTINE PLOTA(W,N,
 X,XMID,IS,NB,PEAN,F
 ENG)
- 26、主成份分析椭圆图 SUBROUTINE PLOTP (X,N,
 A,B,C,G,)

- 27、判别分析图: SUBROUTINE PLOTS (X,Y,
N1,N2,N3,N,A,B,C,S1,
S2)
- 28、正交趋势图: SUBROUTINE MAP(NL,NC,
ALPHA,VEC,REFCON,CO
NT,XAXIS,YAXIS,NO,
IORD,VMAX,UMIN)
- 29、正交趋势散点图: SUBROUTINE DAPLOT (X,
N,XAXIS,YAXIS,VMA
X,VMAX,UMIN,VMIN,
NL,NC)
- 30、移动平均和网格化图: SUBROUTINE PLOTCC(Y,NR
其下为各种专用子程序 MC,ISD)
- 31、输入子程序: SUBROUTINE READM (A,
N,M) (附有几种变形形式)
- 32、输出子程序: SUBROUTINE PRINTM (A,
N,M) (附有变形形式)
- 33、标准化子程序: SUBROUTINE STAND (X,
N,M)
- 34、矩阵相减子程序: SUBROUTINE SUBM (A,B,
C,N,M)
- 35、矩阵相乘子程序: SUBROUTINE MMULT (A,B,
C,L,N,M) (附自身左右乘)
- 36、求因子负载子程序: SUBROUTINE LOADING (A1,
A2,M,L)

37. 求矩阵转置子程序: SUBROUTINE TRANS(X,Y,
N,M)
38. 矩阵传送程序: SUBROUTINE SAVE
(A1,A3,M)
39. 求相关系数子程序: SUBROUTINE RCOEF (X,N,
M,A)
40. 矩阵求逆子程序: SUBROUTINE MINV (A,B,
N,DET)
41. 求COSQ子程序: SUBROUTINE CTHETA
(X,N,M,A)
42. 方差极大旋转子程序: SUBROUTINE VARMAX
(F,M,L)
43. 求实对称阵特征值及: SUBROUTINE EIGENJ
特征向量的Jacobi法 (A,B,N)
44. 求协方差子程序: SUBROUTINE COV (X,N,
M,A)
45. 求椭圆长、短轴及方向: SUBROUTINE MAPP (X,A
MPP,N,M,A1,A2,A,B,
C,D)
46. 求予测值: SUBROUTINE PREDIC (A
D,N,IORD1,C)
47. 求正规方程系数: SUBROUTINE COEFFI (A,
B,C,N,IORD1,XP)
48. 求离差数: SUBROUTINE SUMMY (D,
N,SY,SY2,SYC,SYC2)

- 49、取两列子程序: SUBROUTINE ~~EXH~~ ~~PH~~, A,
N, M, L1, ~~2P~~
- 50、求沿回归线成等间距值: SUBROUTINE SPAGH(A,
N, C, E, IORD, ~~N1~~)
- 51、解方程: SUBROUTINE SLE(A, B, N,
ZERO)
- 52、移动平均: SUBROUTINE MOVE(A, N,
B, IZ, JZ, DX, DY, W, YM
AX, XMIN)
- 53、网格化: SUBROUTINE GRID(DAT
A, AMAP, N, IH, IW, X2M
AX, X1MIN, DX1, DX2,
SMALL)
- 54、调一列: SUBROUTINE AINTO(A, N,
M, J, L)
- 55、剔除异常点: SUBROUTINE REMOV(D, C,
N1, N, K, DM, SD)
- 56、谱系分组配对: SUBROUTINE WPGA(X, M,
M1, IPAIR, XLEV, ISIM)
- 57、求距离系数: SUBROUTINE DIST(X, N,
M, N1, M1, A, M2)
- 58、异常下限: SUBROUTINE TRVA(DAT
A, N, M, N1, M1)
- 59、求判别得分: SUBROUTINE SCORE(A, N,
M, B)

60. 计算判别函数: SUBROUTINE DISCRI(X,Y,
Z,N1,N2,N3,M,A,C,P,Q,
R,S1,S2,IK)
61. 矩阵两列对调 SUBROUTINE EACH (A,N,
M,IE)
62. 求解求逆: SUBROUTINE SOLUT (X,Y,
M,K)
63. 逐步回归子程序 SUBROUTINE STEP(A,N,M,
AM,SGM,S,R,PY,PW,Q,
AP,SM,F,B,C,X,XMID,
IS)
64. 计算误差: SUBROUTINE WCHB (A,B,
S,N,M,BO)
65. 求正交趋势: FUNCTION ZED (ALPHA,V
EC,NO,N,U,V)
66. 求趋势增置: SUBROUTINE ADDCOL (X,
N,M,K,MEANS)
67. 修正GRAM—SCH MIT正交化: SUBROUTINE GSORTH (X,
U,C,D,N,M,MM,NUL L,
K1,K2,I3,EPS)
68. 把正交系数转成最小 = 乘数: SUBROUTINE RECOV (C,D,
LD,M,DD,K1,K2,COEFP,
S,WORK,IORD,RSS)
69. 求线性对称方程组解的分解法 (1): SUBROUTINE DJN(A,N,M)

- 70、求线性对称方程组解的 SUBROUTINE DJN1(A,L,B,
分解法 (2) : N,M)
- 71、正定对称方程组的平方 SUBROUTINE CHOL (A,N,
根法 (1) : M,FLAG)
- 72、同71 (2) : SUBROUTINE CHOL2 (A,L,
B,N,M,FLAG)
- 73、共轭斜量法解线代数 SUBROUTINE GOEAX(A,B,
方程组 X,N,IL,IMAX,FLAG)
- 74、赛德尔迭代法: SUBROUTINE SADEL(A,B,
X,N,IXF,FLAG,IMA X,
EP)
- 75、正交变换化实对称阵 SUBROUTINE TRED2(NM,
为三对角阵: N,A,D,E,Z)
- 76、求三角对阵特征根、 SUBROUTINE TQL2(NM,N,
特征向量: D,E,Z,IERR)
- 77、同76: SUBROUTINE IMTQL2(NM,
N,D,E,Z,IERR)

附录: 几点说明

第一章 统计程序包

随着微型计算机的深入发展和广泛应用，给地质数据处理带来了广阔的前景。地质体是在慢长的历史进程中形成的，并受到各种复杂因素的影响和控制。由于对地质体的深入研究，人们已获得了大量的地质数据，为了揭示其内在的规律，并预测其发展，必须对大量的地质数据进行去粗取精，去伪存真的处理，分析和检验，并由表及里地深入研究，找到解决问题的线索和途径。对大量地质数据进行处理，其重要的数学方法就是统计分析方法。使用统计分析方法处理地质数据时，如采用人工去处理，不但繁杂，而且也是难以实现的。因此，人们必须借助于用计算机处理，才能快速地完成数据处理任务。

目前，有许多地质队都配置了微型计算机，各队都有大量的数据等待使用多种方法进行处理。因此，本程序包主要是为了满足用户对各种统计分析程序的要求：节省不必要的重复性劳动；以廉价的微型机为基础，方便微型机用户和地质队的使用。

近些年来，多元统计分析方法已广泛地应用于地质科学的许多领域，而且在一些高度复杂的统计分析计算中，一些新的分析技术也正在不断地出现。在这种形势下，不用说对那些无能力编制程序的用户，就是对那些有独立编制程序能力的用户来说，要由个人的力量去承担全部编程的工作，无疑是极为困难的。由此可见，不论对于什么样的用户，都要求有一套比较完整，使用方便灵活，操作简单的统计学程序包。这样做即可以得到经济收益，又可以获得事半功倍的成果。因此，在微型机上建立统计程序包势在必行。

统计程序包结构

地质统计程序包并没有一个严格的定义。一般说来，地质统计程序包应具有程序结构上的特点，也就是：

- 1、把具有共同输入数据形式的分析方法的程序集中起来；
- 2、把各种处理功能和分析方法的程序构成一个统一的集合体；
- 3、通用、简单、方便用户、便于移植。

统计程序包的内容

本书收集了对观测数据进行正态， χ^2 ，T和F分布检验的一些基本统计分析程序，以及方差分析，回归分析，判别分析，因子分析，点群分析和趋势分析构成地质统计程序包。该程序包是各种专用统计分析程序的集合体。它们都以文件的形式存贮在磁盘上的而且独立于程序所使用的数据。

为了便于数据的存贮保留，数据独立于程序，且以文件形式存贮在磁盘上，这样就给数据处理带来极大的方便。为了便于用户阅读和使用，程序包中的程序在以下各章分别列出。考虑到用户的修改和从一个机器移植到另一个机器上，我们将一些公用子程序放到最后两章。因此，读者要注意各章中的程序所调用的子程序序列。对于某些专用程序，我们给出几种程序，用户使用时，可以根据情况进行选择。

第二章 基本统计分析程序

大量的观测数据或实验数据都是随机的数据，它们是否具有一定的统计规律？这就需要对数据进行有效的整理，检验和分析，利用概率统计分析方法，计算这些数据的统计特征，然后推断数据的理论统计参数和统计分布规律，为深入研究这些数据提供比较合理的数学模型和公式。

首先我们看一下观测数据或实验数据的基本特点：

1、数据是离散的且具有波动性，即在同样的条件下所获得的观测数据（或实验数据）是不完全相同的。

2、数据存在一定的规律性。也就是说数据虽然波动，但并非杂乱无章的，而是具有一定的规律性。

数据的波动也就是存在着实验误差。误差按其性质而言可分为三类：

①随机误差。是由偶然因素引起的，但它随着实验观测次数的增加其算术平均值趋于零。

②系统误差。即在观测过程中，服从确定性规律的误差。大多数情况下，它是一个常量，可以通过一定的方法去识别和消除它。

③过失误差。即明显歪曲了实验结果的误差，包含过失误差的实验观测数据称为异常点，它是由于测错，传送错或记录错等不正常的因素而造成的。

由此可见。在对实验数据进行统计分析之前，必须对数据进行整理和检验。

职 能

在获得大量数据的基础上,进行数字特征计算和统计检验;进行统计分布计算及其拟合检验;进行数据的变换和校正;以及对常用的统计分布函数进行数值计算。

算 法

一、数字特征计算及其统计检验

对实验观测数据

$$X_1, X_2, \dots, X_N$$

的数字特征计算就是给出描述数据的取值的大致范围、离散程度、分布特征及相关特征。

1、位置特征计算

描述位置特征的量有均值,极大值,极小值。

均 值

常用的是算术平均值,其算法有:

(1) 直接法

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

(2) 递推法

对 $\bar{X}_0 = 0, n = 1, 2, \dots, N$ 计算

$$\bar{X}_n = \frac{n-1}{n} \bar{X}_{n-1} + \frac{1}{n} X_n = \bar{X}_{n-1} + \frac{1}{n} (X_n - \bar{X}_{n-1})$$

$$\bar{X} = \bar{X}_N$$

(3) 二次均值法

$$\bar{X}_1 = \bar{X} + \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})$$

极小值

$$a = \min\{X_1, X_2, \dots, X_N\}$$

极大值

$$b = \max\{X_1, X_2, \dots, X_N\}$$

2、散度特征计算

散度特征常用极差、均方差和标准差来描述。

极差

$$L = b - a$$

方差和均方差

$$S^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - (\bar{x})^2$$

$$S = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

方差S的常用算法是：

(1) 直接法

$$S^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - (\bar{x})^2$$

(2) 递推法

令 $\bar{x}_0 = 0$, $S_0^2 = 0$, 对 $n = 1, 2, \dots, N$ 计算

$$\bar{x}_n = \bar{x}_{n-1} + \frac{1}{n} (x_n - \bar{x}_{n-1})$$

$$S_n^2 = \frac{n-1}{n} \left[S_{n-1}^2 + \frac{1}{n} (x_n - \bar{x}_{n-1})^2 \right]$$

$$S^2 = S_N^2$$