

# 单因素方差分析的BASIC程序

裴 兰 生

(冶金部花园职工中专)

本文介绍一个因素的方差分析BASIC程序,并举例演示了该程序的使用方法。该程序通用性强,在生产中有一定的实用价值。

关键词: 方差分析; BASIC程序



工作方法

方差分析是一种很有效的统计分析方法。在地质工作中,常用方差分析研究各种地质条件对某种金属含量的影响、地层的划分、对比和估计储量计算的精度等。它是用数理统计方法分析因素对指标的作用。对单因素方差分析来说,就是研究改变实验条件,对试验结果有无显著的影响。

设有 $A_1, A_2, \dots, A_b$ 个实验条件,在每一个试验条件下,进行 $a$ 次实验,试验结果是 $x_{ij}$ 。 $x_{ij}$ 表示在 $A_i$ 条件下所进行的第 $j$ 个实验,见下列原始数据(表1):

方差分析原始数据表 表 1

	$A_1$	$A_2$	...	$A_i$	...	$A_b$
1	$x_{11}$	$x_{21}$	...	$x_{i1}$	...	$x_{b1}$
2	$x_{12}$	$x_{22}$	...	$x_{i2}$	...	$x_{b2}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$j$	$x_{1j}$	$x_{2j}$	...	$x_{ij}$	...	$x_{bj}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$a$	$x_{1a}$	$x_{2a}$	...	$x_{ia}$	...	$x_{ba}$

$$\text{令 } P = \frac{1}{ab} \left( \sum_{i=1}^b \sum_{j=1}^a x_{ij} \right)^2 \quad (1-1)$$

$$Q = \frac{1}{a} \sum_{i=1}^b \left( \sum_{j=1}^a x_{ij} \right)^2 \quad (1-2)$$

$$R = \sum_{i=1}^b \sum_{j=1}^a x_{ij}^2 \quad (1-3)$$

由此有:

组间平方和 $S_1 = Q - P$ , 自由度 $Z_1 = b - 1$ ; 组内平方和 $S_2 = R - Q$ , 自由度 $Z_2 = b(a - 1)$ ; 总平方和 $S = S_1 + S_2 = R - P$ , 自由度 $Z = ab - 1$ 。

上述公式只适用于每组样品个数相等的情况。但由于种种原因,获得的数据往往不甚整齐,各组的样品个数可能不等,此时,公式需作适当的修改。根据一般情况,设 $A_1, A_2, \dots, A_b$ 相应的实验次数为 $a_1, a_2, \dots, a_b$ ,则有:

$$P = \frac{1}{\sum_{i=1}^b a_i} \left( \sum_{i=1}^b \sum_{j=1}^{a_i} x_{ij} \right)^2 \quad (2-1)$$

$$Q = \sum_{i=1}^b \frac{1}{a_i} \left( \sum_{j=1}^{a_i} x_{ij} \right)^2 \quad (2-2)$$

$$R = \sum_{i=1}^b \sum_{j=1}^{a_i} x_{ij}^2 \quad (2-3)$$

组间自由度仍为 $b - 1$ , 组内自由度改为 $\sum_{i=1}^b (a_i - 1)$ , 总自由度为 $\left( \sum_{i=1}^b a_i - 1 \right)$ , 其余仍同前。显然, (2-1)、(2-2) 和 (2-3) 式包含了 (1-1)、(1-2) 和 (1-3) 式。本文提出的程序是根据 (2-1)、(2-2) 和 (2-3) 式编写的,适用于 PC-1500 机的 BASIC 程序。

在程序中,  $P$  单元执行  $\sum_{i=1}^b \sum_{j=1}^{a_i} x_{ij}$ ,  $R$  单元执行  $\sum_{i=1}^b \sum_{j=1}^{a_i} x_{ij}^2$ ,  $L$  单元在内层循环中执行  $\sum_{j=1}^{a_i} x_{ij}$ 。由于47语句的存在,  $K$  单元在内层循环中, 累加的是第  $i$  组的实际样品个数  $a_i$ ;  $Q$  单元设置在外层循环之中, 它与87语句相配合, 执行  $\sum_{i=1}^b \frac{1}{a_i} \left( \sum_{j=1}^{a_i} x_{ij} \right)$ ;  $T$  单元累加的是样品总数  $\sum_{i=1}^b a_i$ ; 110语句计算  $P$  值,

至此  $P$ 、 $Q$ 、 $R$  已全部计算完毕。最后, 则依据公式计算并打印  $S_1$ 、 $Z_1$ 、 $S_2$ 、 $Z_2$ 、 $S$ 、 $Z$ 、 $\bar{S}_1$ 、 $\bar{S}_2$  及实测  $F$  值, 打印结束后停机。

#### 两个实例:

##### 1. 每组样品个数相等的情况

例1 在某矿体的4个不同层位上, 分别在5个剖面处取了铅的组合样(为单个样品铅含量的几何平均数, 见表2)。试通过方差分析, 来确定层位(因素)对铅含量的(指标)影响。

层位 ( $i$ )	1	2	3	4
样品 ( $j$ )				
1	2.45	4.65	3.76	4.80
2	2.32	3.84	5.57	3.30
3	2.12	3.39	7.85	5.18
4	2.44	4.73	6.94	5.75
5	1.89	2.40	5.05	5.63

运用本程序, 将计算结果填入方差分析表3中, 然后, 再查  $F$  分布表, 找到  $F_{0.05}$  及  $F_{0.01}$  的值, 当  $F < F_{0.05}$  则无显著性;  $F_{0.05} < F < F_{0.01}$  时, 有显著性, 用 “\*” 表示;  $F > F_{0.01}$  时, 有特别显著性, 用 “\*\*” 表示。

例1的方差分析表

表 3

来源	平方和	自由度	均方	$F$	$F_{0.05}$	$F_{0.01}$	显著性
组间	35.9504	3	11.9835	10.5835	3.24	5.29	**
组内	18.1164	16	1.1323				
总计	54.0668	19					

从表3可得出, 矿体的不同层位对铅含量的高低有特别显著的影响。

##### 2. 各组样品个数不等的情况

例2 在某多金属矿山中, 矿脉产于7种不同类型的围岩内。对每种围岩各取若干样品, 测出其中矿石体重数据列入表4, 试求不同类型的围岩对矿石体重有无显著的影响?

矿石体重原始数据表

表 4

围岩 ( $i$ )	1	2	3	4	5	6	7
样品 ( $j$ )							
1	2.95	2.60	2.65	2.55	2.75	2.80	2.60
2	2.50	2.95	2.80	2.85	2.45	2.50	2.55
3	2.55	2.70	2.75	2.60	3.00	2.85	2.70
4	2.80	2.90	2.60	2.65	2.50	2.95	2.95
5	2.60	2.65	3.40	3.10	3.00	2.95	2.80
6	2.75	3.25		2.70	3.60	2.90	3.15
7		2.50		3.10		3.40	

运用程序可得出例2的方差分析(见表5)。

例2的方差分析表

表 5

来源	平方和	自由度	均方	$F$	$F_{0.05}$	$F_{0.01}$	显著性
组间	0.1920	6	0.0320	0.4145	2.36	3.34	无
组内	2.8560	37	0.0772				
总计	3.0480	43					

因此, 可得出不同类型的围岩对矿石体重影响不明显的结论。

#### 程序及具体用法:

```

10: CLEAR: INPUT B, A
20: DIM X(B, A)
25: P=0:L=0:R=0:K=0:T=0
30: FOR I=1 TO B
40: FOR J=1 TO A
45: READ X(I, J)
47: IF X(I, J)=0 GOTO 55
50: P=P+X(I, J)

```

```

51: L=L+X(I, J)
52: R=R+X(I, J)*X(I, J)
53: K=K+1: T=T+1
55: NEXT J
80: Q=Q+L*L/K
87: L=0: K=0
100: NEXT I
110: P=P*P/T.
120: S1=Q-P: S2=R-Q: S=S2-P
130: LPRINT "S1="; S1
135: LPRINT "Z1="; B-1
140: LPRINT "S2="; S2
145: LPRINT "Z2="; T-B
150: LPRINT "S="; S
155: LPRINT "Z="; T-1
160: S1=S1/(B-1): S2=S2/(T-B)
170: LPRINT "SS1="; S1
180: LPRINT "SS2="; S2
190: LPRINT "F="; S1/S2
200: END

```

对例1，在“200:END”之后，把表2的数据写入DATA语句：

```

210: DATA 2.45, 2.32, 2.12, 2.44, 1.89, 4.65,
3.84, 3.39, 4.73, 2.40
220: DATA 3.76, 5.57, 7.85, 6.94, 5.05, 4.80,
3.30, 5.18, 5.75, 5.63

```

“RUN”后，输入“4”、“5”，印出：

```

例1结果
S1=35.95042
Z1=3
S2=18.1164
Z2=16
S=54.06682
Z=19
SS1=11.98347333
SS2=1.132275
F=10.58353609

```

```

例2结果
S1=0.1919691
Z1=6
S2=2.8559286
Z2=37
S=3.0478977
Z=43
SS1=0.03199485
SS2=7.718725946E-02
F=4.145094699E-01

```

例2与例1类同，差异仅在于先将表4空格填“0”，空几格，填几个“0”写进DATA语句（数据多，则多写几个DATA语句），“RUN”后，输入“7”，“7”即印出结果。

空格填“0”的目的有两个：①“0”为47语句充当转移标志，以使工作单元P、L、R、K、T正确计数；②使原始数据输入机内，仍保持表4中原有的顺序。为避免READ—DATA语句的读数混乱，在易混淆的场合，可用“-1”来作为转移标志，此时47句应作相应的修改。

本程序通用性强，无论每组样品个数是否相同都可进行处理，故在生产中有实用价值。

上述两例均选自武汉地质学院1984年编写的《地质工作中常用的数理统计方法》方差分析部分。

#### 参 考 文 献

中国科学院数学研究所统计组，《常用数理统计方法》，北京，科学出版社，1979，第56~61页

### A Computer Program for Single Factor Variance Analysis

Pei Lansheng

A BASIC computer Program for single factor variance analysis is introduced with an example to illustrate the operation method. This program is of universal applicability and adaptable to a computer PC-1500.