最大限度分辨的多元分析方法

王　元　琦

（湖南沅江气象站）

在只有两个因子的时候，我们可以在散布图上，用一条直线来区分两类天气，也可以用一个经验公式来表达这条直线。如果有多个因子，就只能用公式来描写区别这两类天气的界面了。我们希望求出的公式能将历史个例最大限度地或完全地区别开来。

一、方法简述

在一个以 X_1 为纵轴、X_2 为横轴的直角座标中，若求出的直线方程为：

$$AX_1+BX_2+C=0,$$

则二元分辨函数可以下式表示：

$$Y=AX_1+BX_2,$$

此时分辨指标 $Y_C=-C$。为了简便起见，使 $C=0$，则直线方程可写为：

$$AX_1+BX_2=0,$$

或：

$$X_1-kX_2=0,$$

其中，$k=-B/A$，为区分两类天气的直线的斜率。在为了提高分辨效果而需要把几个因子组合成这样一个方程时，则可先求出：

$$Y_1=X_1-k_2X_2 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (1)$$

然后引进因子 X_3，求出：$Y_2=Y_1-k_3X_3\cdots\cdots(2)$

在引进第n个因子时得出：

$$Y=Y_{n-2}-k_nX_n \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (3)$$

如果此时分辨效果已令人满意，即可将 Y_{n-2}，………………Y_2，Y_1 顺次代回，所得出的就是我们所希望的经验公式。〔式（1）—（3）中的k_2，k_3，……………kn分别是用 2 个、3 个………n 个因子进行组合时的系数。〕

显而易见，为了得出这样一个最大限度的分辨公式，关键在于求出各个k值和依次组合因子。

我们设想，当n个因子都对预报对象Y投正确票时，那么无论怎样组合，结果都会是正确的。因此，在求k值和组合因子时，应当着重分析那些投错误票多的历史个例，力图最大限度地以致完全地加以分辨。

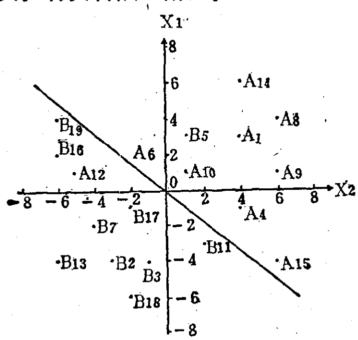
表 1

例序	y	x_1	x_2	$y=x_1+0.83x_2$	例序	y	x_1	x_2	$y=x_1+0.83x_2$
1	A	3	4	6.32	12	A	1	$-5^×$	$-3.15^×$
2	B	−4	−3	−6.49	13	B	−4	−6	−8.98
3	B	−4	−1	−4.83	14	A	6	4	9.32
4	A	$-1^×$	4	2.32	15	A	$-4^×$	6	0.98
5	B	$3^×$	$1^×$	$3.83^×$	16	B	$2^×$	−6	−2.98
6	A	2	$-2^×$	0.34	17	B	−1	−2	−2.66
7	B	−2	−4	−5.32	18	B	−6	−2	−7.66
8	A	4	6	8.96	19	B	$4^×$	−6	−0.98
9	A	1	6	5.98	单因子区分率		14/19	15/19	
10	A	1	1	1.83					
11	B	−3	$2^×$	−1.34	组合区分率				17/19

有×号的是错报因子

二、求k值的方法

现举例说明如下。表 1 中的Y是预报对象，X_1、X_2 是两个相关因子，它们的正值都对应Y的A类，负值对应B类。将资料点入附图。



附图

可以看出，这种情况，属于 $k<0$。此时第一象限（$X_1>0$，$X_2>0$）中的点子，A类是必然可辨个例，B类是必然混杂个例。落入第三象限（$X_1<0$，$X_2<0$）的点子，A类是必然混杂个例，B类是必然可辨个例。它们已无需另分析。

点子落在第二象限（$X_1>0$，$X_2<0$），如果要分辨出是A类点子，则要求$X_1-kX_2>0$，即$k>X_1/X_2$；如果要分辨出B类点子，则要求 $X_1-kX_2<0$，即$k<X_1/X_2$。

落在第四象限（$X_1<0$，$X_2>0$）的点子，如要能够区别开A、B类，对k值的要求与第二象限的情况恰恰相反。

由此可见，在 $k<0$ 的情况下，只要分析第二、四象限中的个例，加以最大限度的分辨即可。（$k>0$ 的情况，可依此类推。）

仍以表 1 和图 1 的资料为例，可以看出，第一、三象限中，只有例 5（图中 B_5）是必然混杂个例，其他 11 个点子都是必然可辨个例。第二、四象限中有 7 个点子，要把A类点子分辨出，需要满足以下 4 个不等式：例 4 要求 $k<-1/4$，例 6 要求 $k>2/-2$，例 12 要求 $k>1/-5$，例 15 要求 $k<-4/6$；要把B类点子分辨出，则例 11 要求 $k>-3/2$，例 16 要求 $k<$

2/－6，例 19 要求 k＜4/－6。分析这七个不等式，可以得出当－1＜k＜－0.67 时，除例 12 外的六个不等式均可成立。取区间的中值－0.83 为k，则得

$$Y = X_1 + 0.83X_2,$$

$Y_C = 0$。当Y＞0 时为A类，Y＜0 时为B类，区分率 17/19，亦即除例 5 和例 12（图中的 B_5 和 A_{12}）以外均可分辨。

三、引进因子组合公式的步骤

在有 n 个因子的情况下，先统计出每一例的错报因子的数目。然后抽出错报因子最多的个例，用前述不等式分析的方法找出使这些特例能全部区分开的两因子组合，定出 k′ 区间。再把用上述的两因子组合对全样本进行不等式分析，求出组合后的区分率，定出k区间。

选择对区分特例有贡献，即 k 与 k′ 有共同解域的组合，以共解的中值为一步过渡函数的常系数，组成一步过渡的经验公式：

$$Y_1 = X_1 - k_2X_2$$

再以 Y_1 为一组合因子，与其他因子结合进行不等式分析，挑选组合后区分率有提高的因子，组成二步过渡的经验公式：

$$Y_2 = Y_1 - k_3X_3$$

然后用同样方法与其他因子组合，直到区分率不再提高或达到完全分辨为止。最后将 Y_{n-2}……Y_2，Y_1 等顺次代回，即得我们所希望的经验公式。现举一实例加以说明。

预报对象Y，其A类表示 4 月上中旬有倒春寒，B类表示无倒春寒。选出的 7 个预报因子如下：

X_1——头年 7 月的日照时数

X_2——头年小暑雨量

X_3——头年 5—9 月总雨量

X_4——头年 8 月雷暴日数

X_5——头年白露平均气温

表 2

年	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	错报因子数		y_{11} $x_1+1.1x_3$	y_{12} $x_1+0.15x_4$	y_{21} $y_{11}+1.2x_4$	y_{22} $y_{12}+0.93x_7$	y_{31} $y_{21}+1.5x_7$
1956	A	$-1^{\times}$	3	1	8	3	3	2	1		0.1	0.20	9.7	2.06	12.7
1957	B	$4^{\times}$	－1	0	－2	－2	－6	－8	1		$4.0^{\times}$	$3.70^{\times}$	$1.6^{\times}$	－3.74	－10.4
1958	B	－2	－2	－1	－2	－3	－4	－1	0		－3.1	－2.30	－5.5	－3.23	－7.0
1959	B	$1^{\times}$	－3	－1	－3	－2	－2	－4	1		－0.1	$0.55^{\times}$	－3.7	－3.17	－9.7
1960	A	7	3	5	2	2	2	7	0		12.5	7.30	14.9	13.81	25.4
1961	B	－1	0	$4^{\times}$	－5	－1	－2	－1	1		$3.4^{\times}$	－1.75	－2.6	－2.68	－4.1
1962	A	6	3	3	3	$-3^{\times}$	0	9	1		9.3	6.45	12.9	14.82	26.4
1963	A	2	2	1	6	1	8	$-3^{\times}$	1		3.1	2.90	10.3	0.11	5.8
1964	A	$-1^{\times}$	3	3	7	1	$-3^{\times}$	$-6^{\times}$	3		2.3	0.05	10.7	$-5.53^{\times}$	1.7
1965	A	3	3	3	2	$-1^{\times}$	3	$-2^{\times}$	2		6.3	3.30	8.7	1.44	5.7
1966	A	5	3	$-4^{\times}$	1	$-3^{\times}$	$-2^{\times}$	9	3		0,6	5.15	1.8	13.52	15.3
1967	A	1	$-6^{\times}$	2	$-6^{\times}$	$-1^{\times}$	3	3	3		3.2	0.10	$-4.0^{\times}$	2.89	0.5
1968	B	－1	$3^{\times}$	－5	$5^{\times}$	－6	－3	－5	2		－6.5	－0.25	－0.5	－4.90	－8.0
1969	A	$-3^{\times}$	0	4	2	1	$-3^{\times}$	3	2		1.4	$-2.70^{\times}$	3.8	0.09	8.3
1970	B	－6	－10	－8	$8^{\times}$	$4^{\times}$	－1	0	2		－14.8	－4.80	－5.2	－4.80	－5.2
1971	B	－7	－10	－7	－1	－3	$3^{\times}$	－1	1		－14.7	－7.15	－15.9	－8.08	－17.4
1972	A	9	3	2	5	$-7^{\times}$	$-4^{\times}$	$-7^{\times}$	3		12.3	9.75	18.3	3.24	7.8
1973	B	－1	$2^{\times}$	$6^{\times}$	－5	－2	－4	－1	2		$5.6^{\times}$	－1.75	－0.4	－2.68	－1.9
1974	B	－3	－6	－6	$6^{\times}$	－4	－2	－5	1		－9.6	－2.10	－2.4	－6.75	－9.9
1975	B	－7	－10	$4^{\times}$	－6	－3	－5	－3	1		－2.6	－7.90	－9.8	－10.69	－14.3
单因子区分率		15/20	15/18	15/19	16/20	14/20	14/19	15/19		组合区分率	17/20	17/20	18/20	19/20	20/20

X_6——1 月的日照时数

X_7——2 月上旬极端最高气温

第一步：将资料列入表 2，可见这 7 个因子都是正值对应A类，负值对应B类。

第二步：统计每一个例的错报因子数（见表 2），将错报因子最多的四个个例抽出，列于表 3。分析这四例的资料，可以得出：如果从 7 个因子中任取两个因子加以组合，能够把这四例区分开的只有两种组合。把这两种组合扩大到全部资料来分析，对区分都有贡献。于是就用它求出一步过渡函数，如表 4。

表 3　特例资料

年	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1964	A	$-1^{\times}$	3	3	7	1	$-3^{\times}$	$-6^{\times}$
1966	A	5	3	$-4^{\times}$	1	$-3^{\times}$	$-2^{\times}$	9
1967	A	1	$-6^{\times}$	2	$-6^{\times}$	$-1^{\times}$	3	3
1972	A	9	3	3	5	$-7^{\times}$	$-4^{\times}$	$-7^{\times}$

表 4

一特例分析	用全资料分析	鉴别	一步过渡函数
能区分特例的因子组合	最大分辨率		
x_1x_3 $(-5/4<k'<-1/3)$ x_1x_4 $(-1/6<k'<-1/7)$	17/20 $(-5/4<k<1)$ 17/20 $(-1/6<k<1/7)$	有共解，分辨率提高	$y_{11}=x_1+1.1x_3$ $y_{12}=x_1+0.15x_4$

第三步：分别以 Y_{11} 和 Y_{12} 为组合因子，与剩下的其他因子结合作不等式分析，如下：

$Y_{11}(17/20)$
- X_2：出现必然混杂个例，剔除。
- X_4：18/20$(-6.5/5<k<-5.6/5)$，挑上，组成 $Y_{21}=Y_{11}+1.2X_4$；
- X_5：17/20，区分率未提高，剔除；
- X_6：17/20，区分率未提高，剔除；
- X_7：17/20，区分率未提高，剔除。

把 Y_{12} 与其他因子分别结合作同样的分析，得出：Y_{12} 与 X_7 组合后区分率为 19/20 $(-2.9/3<k<-2.7/3)$，组成 $Y_{22}=Y_{12}+0.93X_7$，其他因子剔除。

第四步：以 Y_{21} 和 Y_{22} 为组合因子（其资料见表2），再分别与各自的剩余因子结合进行分析，得出：

$Y_{21}(18/20)$
- X_2：出现必然混杂个例，剔除；
- X_5：出现必然混杂个例，剔除；
- X_6：18/20，区分率未提高，剔除；
- X_7：20/20，$(10.7/-6<k<-4.0/3)$，挑上，组成 $Y_{31}=Y_{21}+1.5X_7$，完全分辨。

把 Y_{22} 与剩余的四个因子组合，区分率都未提高，因子均剔除。

第五步：由于上面得出的 $Y_{31}=Y_{21}+1.5X_7$ 式是唯一的可以完全分辨的公式，所以把 Y_{21} 和 Y_{11} 代回，得出：

$$Y=X_1+1.1X_3+1.2X_4+1.5X_7$$

$Y_C=0$，当 $Y>0$ 时为A类，$Y<0$ 时为B类，区分率20/20。

1976—1978年使用上述公式预报有无倒春寒，结果均与实况相符。