

数理统计天气预报讲义

下 册

湖南省气象局统计预报训练班编印

一九七五年元月



目 录

第五章 判别分析

5.1 判别分析的基本原理	5-1
5.2 二组判别	5-4
5.3 多组判别	5-13
5.4 逐多二组分析	5-22

第六章 天气预报中的马尔科夫链方法

6.1 马尔科夫链的定义与组成	6-11
6.2 用马尔科夫链预报天气现象	6-11
6.3 用多维马尔科夫链作晴雨预报	6-22

第七章 平稳时间序列预报

7.1 随机过程概论	7-1
7.2 一维平稳时间序列预报	7-8
7.3 平稳时间序列回归系数的估计方法	7-21
7.4 多维平稳时间序列预报	7-35
7.5 非平稳时间序列分析	7-40

第八章 统计决策在概率天气预报中的应用

8.1 统计决策研究的问题	8-1
8.2 策略论与预报服务	8-2
8.3 统计决策函数理论的基本概念	8-10
8.4 在两类天气预报中的应用	8-15

第九章 统计动力预报

- §1 动力预报的误差及统计动力预报的一种思路 - - - - 9-1
- §2 经验正交函数展开随机场的统计动力预报 - - - - 9-4
- §3 动力预报模式的随机化及参数化 - - - - 9-10
- §4 预报模式解的逐步校正——长内滤波 - - - - 9-12
- §5 相空间中的统计动力模式 - - - - 9-21

第十章 用电子计算机程序例题

- §1 电子计算机简介 - - - - 10-1
- §2 X-2 机的数据格式 - - - - 10-3
- §3 统计预报方法和若干程序 - - - - 10-6
- §4 X-2 机上的操作介绍 - - - - 10-19

第五章 判别分析

§1. 判别分析的基本原理

一. 什么是判别分析

在天气预报工作中，我们经常遇到预报对象^时属于分类（分组）的性质。晴、雨、小雨、中雨、大雨、暴雨，有旱还是无旱，有水雹还是无水雹，冰冻严重还是不严重等。对于这类天气现象出现与否的预报，常用的工具之一就是点聚图。例如我们可以用两个预报因子 x_1 和 x_2 去预报预报对象（ y ）出现与否。这时分别以 x_1 和 x_2 组成一个直角坐标，将记录分别标在图上，于 x_1 与 x_2 相交处绘制一点，并用不同符号标明 y 出现还是不出现。然后再在不同符号之间划一条分界线。预报时，看预报因子落在该分界线的那一侧以此来预报对象出现与否。这种方法就是一种图解判别的方法，它简单但较粗糙，而且在同样资料的情况下，不同的人所得预报结果不会完全相同。特别是当因子多于二个的时候，图解法就很难以映出来。因此人们就设法采用事物的数量特征对这类事物进行分辨。例如，为了预报次日是晴或雨，就可以把影响降水的天气因子综合为一个标准，以此来判别是晴或是雨。这种根据观测到的一些数量特征，按一定的原则将客观事物进行分类，分辨事物的种类和属性的方法，在近代统计学中称之为判别分析，我们简称为分辨法。它是一种采用事物数量特征来区分不同性质事物的方法。这种方法在各个学科中有广泛的应用。例如，在考古学中根据发掘出的人的头盖骨高度等特征数量，可以分辨为男女性别；对于两类相近植物，根据其叶茎长度等特点可以分辨其所属类别等。这里介绍用来制作天气预报，在天气预报中，根据气象要素及其变化的数量可以分辨天气，就可以进行定性的分类，作一些定性的预报，也可以直接进行分组预报。

二. 分辨法的概念

现在我们用一个简单的例子来说明分辨法的概念。假设用两个因子 x_1 和 x_2 预报未来 24 小时的晴雨。 x_1 — 14 时的 24 小时变压 (ΔP_{24}), x_2 — 14 时的 “T-e”, 晴天用 A 表示, 雨天用 B 表示。

如果选择的 x_1 和 x_2 是预报晴雨较好的因子。那么晴天之前的 14 时的 x_1 和 x_2 多数出现在某些数值范围内, 雨天前 14 时的 x_1 和 x_2 多数出现在另一些数值范围内。比如晴天之前 ΔP_{24} 常为正值, T-e 常较大, 雨天之前, ΔP_{24} 常为负值, T-e 常较小, 甚至是负值。我们的目的就是要由这些不同的数值范围来区分晴、雨这两类事物。为了根据 x_1 和 x_2 的这个特征来区分晴、雨, 我们采用一种最简单的线性组合形式把 x_1 和 x_2 组合起来:

$$y = C_1 x_1 + C_2 x_2 \quad (1-1)$$

式中的 C_1 和 C_2 是常数, 可根据历史资料按一定原则确定之。将历史资料中晴天前的 x_1 和 x_2 代入 (1-1) 式, 可得许多个 y_i 值, 记为 y_i^A , 雨天前 14 时的 x_1 和 x_2 代入 (1-1) 式, 亦可得许多个 y_i 值, 记为 y_i^B 。由于雨天前和晴天前的 x_1 和 x_2 数值不同, 自然 y_i^A 和 y_i^B 亦不相同, 这样可以设法找到一个 y_c 值将 y_i^A 和 y_i^B 区分开来。就用 y_c 来区分晴、雨, 我们称它为分辨指标。日常预报中将 x_1 和 x_2 代入 (1-1) 式就得到一个 y 值, 若 $y > y_c$ 报未来 24 小时晴, 若 $y < y_c$ 报未来 24 小时雨。因此, 在判别分析中把 y 称之为判别函数, C_1 、 C_2 称为判别系数, 整个 (1-1) 式称为判别方程, 或叫判别函数。

这个例子可以用几何图形来进行说明 (见 1-1)。

图例说明

- ○ 表示A类
- ○ 表示B类
- • 表示在 x_1, x_2 平面上
的点

- ○ 表示在空间的点
- 表示平面在空间的
交线
- == 表示空间交线在 x_1, x_2
平面上的投影
- - - 表示投影线

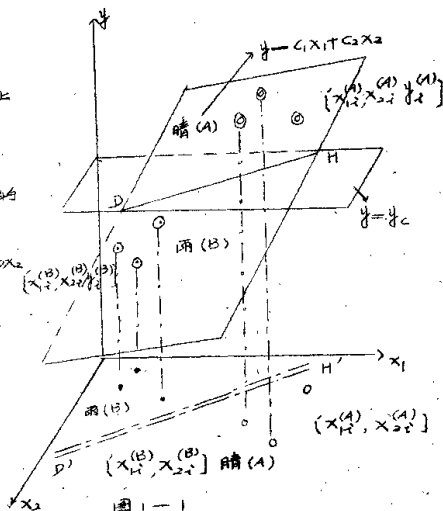


图 1-1

在 x_1, x_2 平面上，若较大的 x_1, x_2 对应未来24小时是晴天（A类表示），较小的 x_1, x_2 对应未来24小时是雨天（B类表示），那么，我们用（1-1）式把 x_1, x_2 组合起来，那么，每一个晴天点 $(x_1^{(A)}, x_2^{(A)})$ 便对应空间一个点 $(x_1^{(A)}, x_2^{(A)}, y_1^{(A)})$ ，每一个雨天点 $(x_1^{(B)}, x_2^{(B)})$ 也对应一个空间点 $(x_1^{(B)}, x_2^{(B)}, y_1^{(B)})$ 。由图可知，晴天点较高，即 $y_1^{(A)}$ 较大，雨天点较低，即 $y_1^{(B)}$ 较小。根据历史资料拟合方程 $y = C_1x_1 + C_2x_2$ ，它是通过坐标原点的一个空间斜线。我们所需要的判别指标 y_c ，它的方程为 $y = y_c$ ，是一个与 x_1, x_2 平面相平行的平面，与 y 轴相交于 y_c ，和 $y = C_1x_1 + C_2x_2$ 平面相交于 DH。DH 将 $y = C_1x_1 + C_2x_2$ 平面分为两部份：晴区和雨

区, DH 就成了晴雨的分界线. 若 $y > y_c$ 表示处于落在晴区, 应报晴; $y < y_c$ 表示处于落在雨区, 应报雨. 而 X_1, X_2 平面上的 DH 是空间分界线 DH 在 X_1, X_2 平面上的投影, 实际上就是决策图方法中的晴雨分界线.

我们的目的是要用 y 来预报未来的晴雨, 那么 y 值多大时报晴, 多大时报雨? 虽然 y 是与 C_1 和 C_2 有密切关系的. 因此, 利用历史资料按照一定的原则来决定 C_1, C_2 时, 要能决定未来的 C_1 和 C_2 起到以下作用: (1) 要使得 y_c^A 与 y_c^B 能够区别, 而且两者相差越大越好. 如 y_c^A 的平均值记为 $\bar{y}^A$, y_c^B 的平均值记为 $\bar{y}^B$, 选择 C_1, C_2 时, 要使得 $(\bar{y}^A - \bar{y}^B)$ 大, 越大越能判列分明. 即要求两类差别要大; (2) 要能使晴天的观测值其平均值 $\bar{y}^A$ 不要太过, 对雨天的观测也一样. 亦即同类的差别要小. 偏差的程度用方差平方和表示, 即:

$$\sum_{i=1}^{N_A} (y_i^A - \bar{y}^A)^2 \text{ 和 } \sum_{i=1}^{N_B} (y_i^B - \bar{y}^B)^2$$

要小. N_A 和 N_B 分别是晴天和雨天的次数; (3) 要反映出哪一个因子对未来晴雨的代表性好, 如果 X_1 比 X_2 的代表性好, 那么 C_1 要比 C_2 大. X_1 和 X_2 对晴雨的代表性通过 C_1 和 C_2 加以放大而影响到 y .

§2 二般判列

一. 公式的推导

(一) 符号

第一个因子 X_1 , 各次晴天时的 X_1 是, $x_{11}^A, x_{12}^A, x_{13}^A, \dots, x_{1N_A}^A$

记为 $x_{1i}^A, i = 1, 2, 3, \dots, N_A$.

各次雨天时的 X_1 是, $x_{11}^B, x_{12}^B, x_{13}^B, \dots, x_{1N_B}^B$

记为 $x_{1i}^B, i = 1, 2, 3, \dots, N_B$.

第二个因子 X_2 同样分别记为 X_{12}^A 和 X_{12}^B 。
其平均值：晴天以前的记为 $\bar{X}_1^A = \frac{1}{N_A} \sum_{i=1}^{N_A} X_{1i}^A$, $\bar{X}_2^A = \frac{1}{N_A} \sum_{i=1}^{N_A} X_{2i}^A$

雨天以前的记为 $\bar{X}_1^B = \frac{1}{N_B} \sum_{i=1}^{N_B} X_{1i}^B$, $\bar{X}_2^B = \frac{1}{N_B} \sum_{i=1}^{N_B} X_{2i}^B$

晴天的 y 记为 y_i^A , ($i=1, 2, \dots, N_A$) 平均值为

$$\bar{y}^A = \frac{1}{N_A} \sum_{i=1}^{N_A} y_i^A = C_1 \bar{X}_1^A + C_2 \bar{X}_2^A$$

雨天的 y 记为 y_i^B ($i=1, 2, 3, \dots, N_B$) 平均值记为

$$\bar{y}^B = \frac{1}{N_B} \sum_{i=1}^{N_B} y_i^B = C_1 \bar{X}_1^B + C_2 \bar{X}_2^B$$

(一) 决定判别系数的原则:

根据前节的讨论, 很显然, 要想使 y_c 成为良好的分解指标, 决定判别系数时必须遵循下述原则:

1. 要同类之间的差别尽量地大, 即 $(\bar{y}^A - \bar{y}^B)^2 = E$ 为最大。
2. 要使得两类现象本身的离散平方和尽量地小, 即:

$$\sum_{i=1}^{N_A} (y_i^A - \bar{y}^A)^2 + \sum_{i=1}^{N_B} (y_i^B - \bar{y}^B)^2 = F \text{ 为最小。}$$

3. 结合起来考虑即可求:

$$I = \frac{(\bar{y}^A - \bar{y}^B)^2}{\sum_{i=1}^{N_A} (y_i^A - \bar{y}^A)^2 + \sum_{i=1}^{N_B} (y_i^B - \bar{y}^B)^2} = \frac{E}{F} \text{ 为最大 (1-2)}$$

分子为组内离散平方和, 分子为组间平均值之差的平方。分子越大, 说明两类中心点之间的绝对距离越远, 分母越小, 说明各类现象的离散程度越小, 这样就越好判别。I 大则说明判别好, I 小则说明不好判别。因此把 I 叫做判别的标佳。

(三) 公式推导

I 是 y 的函数, 而 y 又是 C_1 和 C_2 的函数, 所以 I 是 C_1 、 C_2 的函数。根据求极值原理, 使 I 达到最大的 C_1 、 C_2 由下列联立方程式解出:

$$\begin{cases} \frac{\partial I}{\partial C_1} = 0 \\ \frac{\partial I}{\partial C_2} = 0 \end{cases}$$

$$\text{而 } \frac{\partial I}{\partial C_1} = \frac{\partial}{\partial C_1} \left(\frac{E}{F} \right) = \frac{F \frac{\partial E}{\partial C_1} - E \frac{\partial F}{\partial C_1}}{F^2} = 0$$

$$\frac{\partial I}{\partial C_2} = \frac{\partial}{\partial C_2} \left(\frac{E}{F} \right) = \frac{F \frac{\partial E}{\partial C_2} - E \frac{\partial F}{\partial C_2}}{F^2} = 0$$

整理后即为:

$$\begin{cases} I \frac{\partial F}{\partial C_1} = \frac{\partial E}{\partial C_1} \\ I \frac{\partial F}{\partial C_2} = \frac{\partial E}{\partial C_2} \end{cases} \quad (1-3)$$

$$\text{其中的 } E = (\bar{y}^A - \bar{y}^B) = (C_1 \bar{x}_1^A + C_2 \bar{x}_2^A) - (C_1 \bar{x}_1^B + C_2 \bar{x}_2^B)^2$$

$$F = \sum_{i=1}^{N_1} (C_1 x_{1i}^A + C_2 x_{2i}^A) - (C_1 \bar{x}_1^A + C_2 \bar{x}_2^A)^2 + \sum_{i=1}^{N_2} (C_1 x_{1i}^B + C_2 x_{2i}^B) - (C_1 \bar{x}_1^B + C_2 \bar{x}_2^B)^2$$

由 (1-3) 式可知, 必须求出 $\frac{\partial F}{\partial C_1}$, $\frac{\partial E}{\partial C_1}$, $\frac{\partial F}{\partial C_2}$, $\frac{\partial E}{\partial C_2}$ 。

1. 求 $\frac{\partial E}{\partial C_1}$ 和 $\frac{\partial F}{\partial C_1}$

$$\begin{aligned} \frac{\partial E}{\partial C_1} &= \frac{\partial}{\partial C_1} (C_1 \bar{x}_1^A + C_2 \bar{x}_2^A) - (C_1 \bar{x}_1^A + C_2 \bar{x}_2^A)^2 \\ &= 2[(C_1 \bar{x}_1^A + C_2 \bar{x}_2^A) - (C_1 \bar{x}_1^B + C_2 \bar{x}_2^B)](\bar{x}_1^A - \bar{x}_1^B) \\ &= 2[C_1(\bar{x}_1^A - \bar{x}_1^B) + C_2(\bar{x}_2^A - \bar{x}_2^B)](\bar{x}_1^A - \bar{x}_1^B) \\ &= 2[C_1 \alpha_1 + C_2 \alpha_2] \alpha_1 \end{aligned}$$

$$\text{其中的 } \alpha_1 = \bar{x}_1^A - \bar{x}_1^B, \alpha_2 = \bar{x}_2^A - \bar{x}_2^B$$

$$\begin{aligned}
\frac{\partial F}{\partial C_1} &= \frac{\partial}{\partial C_1} \left\{ \sum_{i=1}^{N_A} [(C_1 x_{1i}^A + C_2 x_{2i}^A) - (C_1 \bar{x}_1^A + C_2 \bar{x}_2^A)]^2 \right. \\
&\quad \left. + \sum_{i=1}^{N_B} [(C_1 x_{1i}^B + C_2 x_{2i}^B) - (C_1 \bar{x}_1^B + C_2 \bar{x}_2^B)]^2 \right\} \\
&= 2 \sum_{i=1}^{N_A} [(C_1 x_{1i}^A + C_2 x_{2i}^A) - (C_1 \bar{x}_1^A + C_2 \bar{x}_2^A)] (x_{1i}^A - \bar{x}_1^A) \\
&\quad + 2 \sum_{i=1}^{N_B} [(C_1 x_{1i}^B + C_2 x_{2i}^B) - (C_1 \bar{x}_1^B + C_2 \bar{x}_2^B)] (x_{1i}^B - \bar{x}_1^B) \\
&= 2 \sum_{i=1}^{N_A} [C_1 (x_{1i}^A - \bar{x}_1^A) + C_2 (x_{2i}^A - \bar{x}_2^A)] (x_{1i}^A - \bar{x}_1^A) \\
&\quad + 2 \sum_{i=1}^{N_B} [C_1 (x_{1i}^B - \bar{x}_1^B) + C_2 (x_{2i}^B - \bar{x}_2^B)] (x_{1i}^B - \bar{x}_1^B) \\
&= 2C_1 \left[\sum_{i=1}^{N_A} (x_{1i}^A - \bar{x}_1^A)^2 + \sum_{i=1}^{N_B} (x_{1i}^B - \bar{x}_1^B)^2 \right] \\
&\quad + 2C_2 \left[\sum_{i=1}^{N_A} (x_{1i}^A - \bar{x}_1^A) (x_{2i}^A - \bar{x}_2^A) + \sum_{i=1}^{N_B} (x_{1i}^B - \bar{x}_1^B) (x_{2i}^B - \bar{x}_2^B) \right] \\
&= 2C_1 S_{11} + 2C_2 S_{12}
\end{aligned}$$

其中 $S_{11} = \sum_{i=1}^{N_A} (x_{1i}^A - \bar{x}_1^A)^2 + \sum_{i=1}^{N_B} (x_{1i}^B - \bar{x}_1^B)^2$

$$S_{12} = \sum_{i=1}^{N_A} (x_{1i}^A - \bar{x}_1^A) (x_{2i}^A - \bar{x}_2^A) + \sum_{i=1}^{N_B} (x_{1i}^B - \bar{x}_1^B) (x_{2i}^B - \bar{x}_2^B)$$

S_{11} 表示第一个因子在两类不同条件下的距平方和，即为 x_1 的组内相关。 S_{12} 为第一个因子 x_1 和第二个因子 x_2 的距平方相互乘积和，即为 x_1, x_2 间的相关。

2. 求 $\frac{\partial F}{\partial C_2}$ 和 $\frac{\partial F}{\partial C_3}$

$$\begin{aligned}
\frac{\partial F}{\partial C_2} &= \frac{\partial}{\partial C_2} [(C_1 \bar{x}_1^A + C_2 \bar{x}_2^A) - (C_1 \bar{x}_1^B + C_2 \bar{x}_2^B)]^2 \\
&= 2[(C_1 \bar{x}_1^A + C_2 \bar{x}_2^A) - (C_1 \bar{x}_1^B + C_2 \bar{x}_2^B)] (\bar{x}_2^A - \bar{x}_2^B) \\
&= 2(C_1 d_1 + C_2 d_2) d_2
\end{aligned}$$

$$\begin{aligned}\frac{\partial F}{\partial C_2} &= \frac{\partial}{\partial C_2} \left\{ \sum_{i=1}^{N_A} [(C_1 x_{i1}^A + C_2 x_{i2}^A) - (C_1 \bar{x}_1^A + C_2 \bar{x}_2^A)]^2 \right. \\ &\quad \left. + \sum_{i=1}^{N_B} [(C_1 x_{i1}^B + C_2 x_{i2}^B) - (C_1 \bar{x}_1^B + C_2 \bar{x}_2^B)]^2 \right\} \\ &= 2(C_1 S_{21} + C_2 S_{22})\end{aligned}$$

$$\text{其中 } S_{22} = \sum_{i=1}^{N_A} (x_{i2}^A - \bar{x}_2^A)^2 + \sum_{i=1}^{N_B} (x_{i2}^B - \bar{x}_2^B)^2$$

$$S_{21} = \sum_{i=1}^{N_A} (x_{i2}^A - \bar{x}_2^A)(x_{i1}^A - \bar{x}_1^A) + \sum_{i=1}^{N_B} (x_{i2}^B - \bar{x}_2^B)(x_{i1}^B - \bar{x}_1^B)$$

S_{22} 表示 x_2 在 A 类和 B 类条件下的距平方和，即为 x_2 的自相关。
 S_{21} 表示 x_2, x_1 在 A 类和 B 类条件下的距平方乘积和，即为 x_2, x_1 的互相关。

3. 将求得的 $\frac{\partial F}{\partial C_1}, \frac{\partial F}{\partial C_2}, \frac{\partial F}{\partial d_1}, \frac{\partial F}{\partial d_2}$ 代入 (1-3) 式得：

$$I d(C_1 S_{11} + C_2 S_{12}) = 2(C_1 d_1 + C_2 d_2) d_1$$

$$I d(C_1 S_{21} + C_2 S_{22}) = 2(C_1 d_1 + C_2 d_2) d_2$$

$$C_1 S_{11} + C_2 S_{12} = (C_1 d_1 + C_2 d_2) d_1 / I$$

$$C_1 S_{21} + C_2 S_{22} = (C_1 d_1 + C_2 d_2) d_2 / I$$

右边有一个共同的系数 $(C_1 d_1 + C_2 d_2) / I$ ，它对线性方程的解 C_1, C_2 起着扩大和缩小的作用，并不影响 C_1, C_2 的相对大小，对判别分析来说没有影响。因此令 $(C_1 d_1 + C_2 d_2) / I = 1$ ，上式变为：

$$\begin{cases} C_1 S_{11} + C_2 S_{12} = d_1 \\ C_1 S_{21} + C_2 S_{22} = d_2 \end{cases} \quad (1-4)$$

方程中的 $S_{11}, S_{12}, S_{22}, d_1, d_2$ 都可以从历史资料中求出，代入 (1-4) 式便可求出 C_1 和 C_2 。

4. 多个因子的情况

上已是一个简例，仅只二因子且只分或二级的情况，下已再推导在因子 m 时的二级分阶公式。

判别函数为：

$$y = C_1 x_1 + C_2 x_2 + \dots + C_m x_m = \sum_{k=1}^m C_k x_k \quad (1-5)$$

其中 m 为因子个数， C_k 为判别系数， $k=1, 2, 3, \dots, m$ 。

原则和上已一样，即要求：

$$\frac{\partial}{\partial C_k} \left(\frac{(\bar{y}^A - \bar{y}^B)}{\sum_{i=1}^{N_A} (y_i^A - \bar{y}^A)^2 + \sum_{i=1}^{N_B} (y_i^B - \bar{y}^B)^2} \right) = 0$$

$$\text{令：(1) } (\bar{y}^A - \bar{y}^B) = E$$

$$(2) \sum_{i=1}^{N_A} (y_i^A - \bar{y}^A)^2 + \sum_{i=1}^{N_B} (y_i^B - \bar{y}^B)^2 = F$$

$$(3) E/F = I$$

$$\text{即由 } \frac{\partial I}{\partial C_k} = \frac{\partial}{\partial C_k} \left(\frac{E}{F} \right) = 0 \quad \text{有：} \frac{1}{I} \frac{\partial E}{\partial C_k} = \frac{\partial E}{\partial C_k} \quad (1-6)$$

根据 (1-5) 式有：

$$\begin{aligned} E &= (C_1 \bar{x}_1^A + C_2 \bar{x}_2^A + C_3 \bar{x}_3^A + \dots + C_m \bar{x}_m^A \\ &\quad - C_1 \bar{x}_1^B - C_2 \bar{x}_2^B - C_3 \bar{x}_3^B - \dots - C_m \bar{x}_m^B)^2 \\ &= [C_1 (\bar{x}_1^A - \bar{x}_1^B) + C_2 (\bar{x}_2^A - \bar{x}_2^B) + \dots + C_m (\bar{x}_m^A - \bar{x}_m^B)]^2 \end{aligned}$$

$$\begin{aligned} F &= \sum_{i=1}^{N_A} [C_1 x_{1i}^A + C_2 x_{2i}^A + \dots + C_m x_{mi}^A - C_1 \bar{x}_1^B - C_2 \bar{x}_2^B - \dots - C_m \bar{x}_m^B]^2 \\ &\quad + \sum_{i=1}^{N_B} [C_1 x_{1i}^B + C_2 x_{2i}^B + \dots + C_m x_{mi}^B - C_1 \bar{x}_1^A - C_2 \bar{x}_2^A - \dots - C_m \bar{x}_m^A]^2 \\ &= \sum_{i=1}^{N_A} [C_1 (x_{1i}^A - \bar{x}_1^A) + C_2 (x_{2i}^A - \bar{x}_2^A) + \dots + C_m (x_{mi}^A - \bar{x}_m^A)]^2 \\ &\quad + \sum_{i=1}^{N_B} [C_1 (x_{1i}^B - \bar{x}_1^B) + C_2 (x_{2i}^B - \bar{x}_2^B) + \dots + C_m (x_{mi}^B - \bar{x}_m^B)]^2 \end{aligned}$$

$$\begin{aligned}\frac{\partial E}{\partial c_k} &= \frac{\partial}{\partial c_k} \left\{ c_1 (\bar{x}_1^A - \bar{x}_1^B) + c_2 (\bar{x}_2^A - \bar{x}_2^B) + \dots + c_m (\bar{x}_m^A - \bar{x}_m^B) \right\}^2 \\ &= 2 \left\{ c_1 (\bar{x}_1^A - \bar{x}_1^B) + c_2 (\bar{x}_2^A - \bar{x}_2^B) + \dots + c_m (\bar{x}_m^A - \bar{x}_m^B) \right\} (\bar{x}_k^A - \bar{x}_k^B) \\ &= 2 \left[c_1 \alpha_1 + c_2 \alpha_2 + \dots + c_m \alpha_m \right] \alpha_k\end{aligned}$$

$$\begin{aligned}\frac{\partial F}{\partial c_k} &= \frac{\partial}{\partial c_k} \left\{ \sum_{i=1}^{N_A} \left[c_1 (x_{1i}^A - \bar{x}_1^A) + c_2 (x_{2i}^A - \bar{x}_2^A) + \dots + c_m (x_{mi}^A - \bar{x}_m^A) \right]^2 \right. \\ &\quad \left. + \sum_{i=1}^{N_B} \left[c_1 (x_{1i}^B - \bar{x}_1^B) + c_2 (x_{2i}^B - \bar{x}_2^B) + \dots + c_m (x_{mi}^B - \bar{x}_m^B) \right]^2 \right\} \\ &= 2 \sum_{i=1}^{N_A} \left[c_1 (x_{1i}^A - \bar{x}_1^A) + c_2 (x_{2i}^A - \bar{x}_2^A) + \dots + c_m (x_{mi}^A - \bar{x}_m^A) \right] (\bar{x}_k^A - \bar{x}_k^A) \\ &\quad + 2 \sum_{i=1}^{N_B} \left[c_1 (x_{1i}^B - \bar{x}_1^B) + c_2 (x_{2i}^B - \bar{x}_2^B) + \dots + c_m (x_{mi}^B - \bar{x}_m^B) \right] (\bar{x}_k^B - \bar{x}_k^B)\end{aligned}$$

$$= 2 \left[c_1 S_{k1} + c_2 S_{k2} + \dots + c_m S_{km} \right]$$

其中: $\alpha_k = \bar{x}_k^A - \bar{x}_k^B$ $\bar{x}_k$ 在两种情况下的平均值之差。

$$S_{kk} = \sum_{i=1}^{N_A} (x_{ki}^A - \bar{x}_k^A)^2 + \sum_{i=1}^{N_B} (x_{ki}^B - \bar{x}_k^B)^2$$

$$S_{kp} = \sum_{i=1}^{N_A} (x_{ki}^A - \bar{x}_k^A) (x_{pi}^A - \bar{x}_p^A) + \sum_{i=1}^{N_B} (x_{ki}^B - \bar{x}_k^B) (x_{pi}^B - \bar{x}_p^B)$$

($k, p = 1, 2, 3, \dots, m$, $k \neq p$)

将上述结果代入 (1-6) 式

$$2 (c_1 S_{k1} + c_2 S_{k2} + \dots + c_m S_{km}) = 2 \frac{(c_1 \alpha_1 + c_2 \alpha_2 + \dots + c_m \alpha_m) \alpha_k}{I}$$

同样可以令 $(c_1 \alpha_1 + c_2 \alpha_2 + c_3 \alpha_3 + \dots + c_m \alpha_m) / I = 1$, 则

$$c_1 S_{k1} + c_2 S_{k2} + c_3 S_{k3} + \dots + c_m S_{km} = \alpha_k$$

展开就是:

$$\begin{cases} C_1 S_{11} + C_2 S_{12} + C_3 S_{13} + \dots + C_m S_{1m} = \alpha_1 \\ C_1 S_{21} + C_2 S_{22} + C_3 S_{23} + \dots + C_m S_{2m} = \alpha_2 \\ \dots \dots \dots \\ C_1 S_{m1} + C_2 S_{m2} + C_3 S_{m3} + \dots + C_m S_{mm} = \alpha_m \end{cases} \quad (5-7)$$

二. 分辨指标的确定

求得判别系数 C_k 以后, 则有:

$$y = \sum_{k=1}^m C_k x_k$$

若通过这个方法可分别求出 y^A 的平均 $\bar{y}^A$ 和 y^B 的平均 $\bar{y}^B$ 为:

$$\bar{y}^A = \sum_{k=1}^m C_k \bar{x}_k^A \quad \bar{y}^B = \sum_{k=1}^m C_k \bar{x}_k^B$$

于是可得到预报判别为:

$$y_c = \frac{N_A \bar{y}^A + N_B \bar{y}^B}{N_A + N_B}$$

当 $\bar{y}^A > \bar{y}^B$ 时, 若 $y > y_c$, 预报出现 A 类天气.

$y < y_c$, 预报出现 B 类天气.

当 $\bar{y}^A < \bar{y}^B$ 时, 若 $y > y_c$, 预报出现 B 类天气.

$y < y_c$, 预报出现 A 类天气.

三. 举例说明统计步骤.

为了便于掌握统计步骤, 现举一个简单例子, 说明当预报因子选好后如何进行统计. 至于因子的选择方法和建立判别函数的原理将在后文介绍.

某站用 x_1 (上一年 T_{12}) 和 x_2 (上一年 R_{11}) 来判定 4 月份的降水 (R_4) 是否大于等于 50mm. 资料由表 1-1 给出

表 1-1

y	129.8	4.6	28.0	14.4	70.5	104.3	22.5	41.6	87.2	14.7
x ₁	4.1	5.3	4.3	1.2	3.4	4.8	3.9	2.1	2.3	-1.7
x ₂	38.7	58.7	76.6	87.5	39.5	59.5	23.3	79.6	28.6	166.4

根据数据要求分成两组:

表 1-2

x ^A	x ₁₁ ^A	4.1	5.3	4.8	3.9	2.3	$\bar{x}_1^A = 3.70$
	x ₂₁ ^A	38.7	58.7	59.5	23.3	28.6	$\bar{x}_2^A = 38.00$
x ^B	x ₁₁ ^B	5.3	4.3	1.2	2.1	-1.7	$\bar{x}_1^B = 2.40$
	x ₂₁ ^B	58.7	76.6	87.6	79.6	166.4	$\bar{x}_2^B = 81.76$

1. 计算各因子的平均值

$$\text{计算公式: } \bar{x}_k^A = \frac{1}{N_A} \sum_{i=1}^{N_A} x_{ki}^A \quad \bar{x}_k^B = \frac{1}{N_B} \sum_{i=1}^{N_B} x_{ki}^B$$

本例中 $N_A = N_B = 5$ $k=2$ 结果见上表 1-2。2. 计算自相关系数: (S_{11} 和 S_{22})

$$\text{计算公式: } S_{kk} = \sum_{i=1}^{N_A} (x_{ki}^A - \bar{x}_k^A)^2 + \sum_{i=1}^{N_B} (x_{ki}^B - \bar{x}_k^B)^2$$

$$S_{11} = \sum_{i=1}^5 (x_{1i}^A - \bar{x}_1^A)^2 + \sum_{i=1}^5 (x_{1i}^B - \bar{x}_1^B)^2$$

$$S_{22} = \sum_{i=1}^5 (x_{2i}^A - \bar{x}_2^A)^2 + \sum_{i=1}^5 (x_{2i}^B - \bar{x}_2^B)^2$$

结果见表 1-3。

3. 计算因子间的相关系数: (S_{12})

$$\text{计算公式: } S_{kp} = \sum_{i=1}^{N_A} (x_{ki}^A - \bar{x}_k^A)(x_{pi}^A - \bar{x}_p^A) + \sum_{i=1}^{N_B} (x_{ki}^B - \bar{x}_k^B)(x_{pi}^B - \bar{x}_p^B)$$

本例中:

$$S_{12} = \sum_{i=1}^5 (x_{1i}^A - \bar{x}_1^A)(x_{2i}^A - \bar{x}_2^A) + \sum_{i=1}^5 (x_{1i}^B - \bar{x}_1^B)(x_{2i}^B - \bar{x}_2^B)$$

5~12

计算结果见表 1-3

表 1-3

S_{11}	S_{22}	S_{12}
33.694	1973.934	-150.364

4. 解线性方程组:

$$\begin{cases} 33.694C_1 - 150.364C_2 = 3.70 - 2.24 \\ -150.364C_1 + 1973.934C_2 = 38.00 - 81.76 \end{cases}$$

由方程组求得判别系数为: $C_1 = -0.084$ $C_2 = -0.029$

5. 求出 $\bar{y}_A$, $\bar{y}_B$ 和 y_c

$$\bar{y}_A = -0.084 \times 3.7 - 0.029 \times 38.00 = -1.413$$

$$\bar{y}_B = -0.084 \times 2.24 - 0.029 \times 81.76 = -2.559$$

$$\bar{y}_c = \frac{N_A \bar{y}_A + N_B \bar{y}_B}{N_A + N_B} = \frac{5 \times (-1.413) + 5 \times (-2.559)}{5 + 5} = -1.986$$

本例中因为 $\bar{y}_A > \bar{y}_B$, 所以当 $y > y_c$ 时, 报 A 类, 即预报 4 月份的降水量 $\geq 50 \text{mm}$, 反之当 $y < y_c$ 时, 预报 B 类, 即预报 4 月份的降水量 $< 50 \text{mm}$ 。

四、因子的选择

对某种预报对象的类型进行分辨, 要判别到分明必须选择好的预报因子。这里仅介绍两种根据分辨法的原理而进行因子筛选的方法。

1. 计算单相关系数。

计算单相关系数不必采用一般的方法, 可根据分辨法的特殊点将计算公式进行某种简化。设要分辨的 A、B 类事物中, A 类有 n_1 个样本, B 类有 n_2 个样本, 总样本数 $n = n_1 + n_2$ 个

$$\bar{y}_A = \frac{n_1}{n} \quad \bar{y}_B = \frac{-n_2}{n}$$

计算相关系数的一般公式为

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2}}$$

式中

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{那么 } \bar{x}^A = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^A, \quad \bar{x}^B = \frac{1}{n_2} \sum_{i=1}^{n_2} x_i^B$$

$$\begin{aligned} \text{因为 } \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \left(\sum_{i=1}^{n_1} y_i^A + \sum_{i=1}^{n_2} y_i^B \right) \\ &= \frac{1}{n} \left(n_1 \frac{n_2}{n} - n_2 \frac{n_1}{n} \right) = \frac{1}{n} \left(\frac{n_1 n_2}{n} - \frac{n_1 n_2}{n} \right) = 0 \end{aligned}$$

由此分子变为

$$\begin{aligned} \sum_{i=1}^n y_i (x_i - \bar{x}) &= \sum_{i=1}^{n_1} y_i^A (x_i^A - \bar{x}^A) + \sum_{i=1}^{n_2} y_i^B (x_i^B - \bar{x}^B) \\ &= \sum_{i=1}^{n_1} \frac{n_2}{n} (x_i^A - \bar{x}^A) + \sum_{i=1}^{n_2} \left(-\frac{n_1}{n} \right) (x_i^B - \bar{x}^B) \\ &= \frac{n_2}{n_1} \left(\sum_{i=1}^{n_1} x_i^A - \sum_{i=1}^{n_1} \bar{x}^A \right) - \frac{n_1}{n} \sum_{i=1}^{n_2} x_i^B + \frac{n_1}{n} \sum_{i=1}^{n_2} \bar{x}^B \\ &= \frac{n_2}{n} n_1 \bar{x}^A - \frac{n_1 n_2}{n} \bar{x} - \frac{n_1}{n} n_2 \bar{x}^B + \frac{n_1 n_2}{n} \bar{x} \\ &= \frac{n_1 n_2}{n} (\bar{x}^A - \bar{x}^B) \end{aligned}$$

又因为

$$\begin{aligned} \sum_{i=1}^n y_i^2 &= \sum_{i=1}^{n_1} (y_i^A)^2 + \sum_{i=1}^{n_2} (y_i^B)^2 = n_1 \left(\frac{n_2}{n} \right)^2 + n_2 \left(-\frac{n_1}{n} \right)^2 \\ &= \frac{n_1 n_2^2 + n_2 n_1^2}{n^2} = \frac{n_1 n_2 (n_2 + n_1)}{n^2} = \frac{n_1 n_2}{n} \end{aligned}$$

将上述结果代入计算相关系数的一般公式，于是便得到：