

第九章 对应分析

- ❖ § 9.1 引言
- ❖ § 9.2 行轮廓和列轮廓
- ❖ § 9.3 独立性的检验和总惯量
- ❖ § 9.4 行、列轮廓的坐标
- ❖ § 9.5 对应分析图

§ 9.1 引言

- ❖ **对应分析**是用于寻找列联表的行和列之间关联的一种低维图形表示法，它同时可以揭示同一分类变量的各个类别之间的差异。
- ❖ 对应分析是由法国人Benzecri于1970年提出的，起初在法国和日本最为流行，然后引入到美国。
- ❖ 在对应分析中，列联表的每一行对应（最常二维）图中的一点，每一列也对应同一图中的一点。该图形方法特别适用于有许多类别的列联表，它能有效地用直观、简洁的图形来描述庞杂的列联表数据中所蕴含的对应关系。
- ❖ 由于列联表中行变量和列变量的地位是对称的，所以对应分析方法本身及其所得结论对于行和列也是对称的。

§ 9.2 行轮廓和列轮廓

- ❖ 一、列联表
- ❖ 二、对应矩阵
- ❖ 三、行、列轮廓

一、列联表

表9.2.1

$p \times q$ 列联表

行 \ 列					合 计
	1	2	...	q	
1	n_{11}	n_{12}	...	n_{1q}	$n_{1\cdot}$
2	n_{21}	n_{22}	...	n_{2q}	$n_{2\cdot}$
\vdots	\vdots	\vdots		\vdots	\vdots
p	n_{p1}	n_{p2}	...	n_{pq}	$n_{p\cdot}$
合 计	$n_{\cdot 1}$	$n_{\cdot 2}$...	$n_{\cdot q}$	n

二、对应矩阵

表9.2.2

对应矩阵

列 行	1	2	...	q	合 计
1	p_{11}	p_{12}	...	p_{1q}	$p_{1\cdot}$
2	p_{21}	p_{22}	...	p_{2q}	$p_{2\cdot}$
\vdots	\vdots	\vdots		\vdots	\vdots
p	p_{p1}	p_{p2}	...	p_{pq}	$p_{p\cdot}$
合 计	$p_{\cdot 1}$	$p_{\cdot 2}$...	$p_{\cdot q}$	1

这里, $p_{ij} = \frac{n_{ij}}{n}$, $p_{i\cdot} = \sum_{j=1}^q p_{ij} = \frac{n_{i\cdot}}{n}$, $p_{\cdot j} = \sum_{i=1}^p p_{ij} = \frac{n_{\cdot j}}{n}$ 。

显然有 $\sum_{i=1}^p p_{i\cdot} = \sum_{j=1}^q p_{\cdot j} = 1$ 。

❖ 称

$$\mathbf{P} = (p_{ij}) = (n_{ij}/n)$$

为对应矩阵。将对应矩阵表中的最后一列用 \mathbf{r} 表示，即

$$\mathbf{r} = \mathbf{P}\mathbf{1} = (p_{1.}, p_{2.}, \dots, p_{p.})'$$

其中 $\mathbf{1} = (1, 1, \dots, 1)' : q \times 1$ ，最后一行用 \mathbf{c}' 表示，即

$$\mathbf{c}' = \mathbf{1}'\mathbf{P} = (p_{.1}, p_{.2}, \dots, p_{.q})$$

其中 $\mathbf{1} = (1, 1, \dots, 1)' : p \times 1$ ，向量 \mathbf{r} 和 \mathbf{c} 的元素分别是行边缘频率和列边缘频率，有时称为行和列密度（或质量，mass）。

三、行、列轮廓

❖ 第*i*行轮廓:

$$\mathbf{r}'_i = \left(\frac{p_{i1}}{p_{i\cdot}}, \frac{p_{i2}}{p_{i\cdot}}, \dots, \frac{p_{iq}}{p_{i\cdot}} \right) = \left(\frac{n_{i1}}{n_{i\cdot}}, \frac{n_{i2}}{n_{i\cdot}}, \dots, \frac{n_{iq}}{n_{i\cdot}} \right)$$

其各元素之和等于1, 即 $\mathbf{r}'_i \mathbf{1} = 1, i = 1, 2, \dots, p$ 。

❖ 第*j*列轮廓:

$$\mathbf{c}_j = \left(\frac{p_{1j}}{p_{\cdot j}}, \frac{p_{2j}}{p_{\cdot j}}, \dots, \frac{p_{pj}}{p_{\cdot j}} \right)' = \left(\frac{n_{1j}}{n_{\cdot j}}, \frac{n_{2j}}{n_{\cdot j}}, \dots, \frac{n_{pj}}{n_{\cdot j}} \right)'$$

其各元素之和等于1, 即 $\mathbf{1}' \mathbf{c}_j = 1, j = 1, 2, \dots, q$ 。

行轮廓矩阵

$$\mathbf{R} = \mathbf{D}_r^{-1} \mathbf{P} = \begin{pmatrix} \mathbf{r}'_1 \\ \mathbf{r}'_2 \\ \vdots \\ \mathbf{r}'_p \end{pmatrix} = \begin{pmatrix} \frac{p_{11}}{p_{1\cdot}} & \frac{p_{12}}{p_{1\cdot}} & \dots & \frac{p_{1q}}{p_{1\cdot}} \\ \frac{p_{21}}{p_{2\cdot}} & \frac{p_{22}}{p_{2\cdot}} & \dots & \frac{p_{2q}}{p_{2\cdot}} \\ \vdots & \vdots & & \vdots \\ \frac{p_{p1}}{p_{p\cdot}} & \frac{p_{p2}}{p_{p\cdot}} & \dots & \frac{p_{pq}}{p_{p\cdot}} \end{pmatrix}$$

其中 $\mathbf{D}_r = \text{diag}(p_{1\cdot}, p_{2\cdot}, \dots, p_{p\cdot})$ 。

列轮廓矩阵

$$C = PD_c^{-1} = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_q) = \begin{pmatrix} \frac{p_{11}}{p_{\cdot 1}} & \frac{p_{12}}{p_{\cdot 2}} & \dots & \frac{p_{1q}}{p_{\cdot q}} \\ \frac{p_{21}}{p_{\cdot 1}} & \frac{p_{22}}{p_{\cdot 2}} & \dots & \frac{p_{2q}}{p_{\cdot q}} \\ \vdots & \vdots & & \vdots \\ \frac{p_{p1}}{p_{\cdot 1}} & \frac{p_{p2}}{p_{\cdot 2}} & \dots & \frac{p_{pq}}{p_{\cdot q}} \end{pmatrix}$$

其中 $D_c = \text{diag}(p_{\cdot 1}, p_{\cdot 2}, \dots, p_{\cdot q})$ 。



$$\mathbf{r} = \mathbf{P}\mathbf{1} = (\mathbf{P}\mathbf{D}_c^{-1})(\mathbf{D}_c\mathbf{1}) = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_q) \begin{pmatrix} p_{\cdot 1} \\ p_{\cdot 2} \\ \vdots \\ p_{\cdot q} \end{pmatrix} = \sum_{j=1}^q p_{\cdot j} \mathbf{c}_j$$

➤ 可见， \mathbf{r} 是各列轮廓的加权平均，可看成是 $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_q$ 的（某种）中心。

❖ 类似地，

$$\mathbf{c}' = \mathbf{1}'\mathbf{P} = (\mathbf{1}'\mathbf{D}_r)(\mathbf{D}_r^{-1}\mathbf{P}) = \sum_{i=1}^p p_{i\cdot} \mathbf{r}'_i$$

➤ 即 \mathbf{c}' 是各行轮廓的加权平均，可看成是 $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_p$ 的（某种）中心。

- ❖ 例9.2.1 将由 $n=1660$ 个人组成的样本按心理健康状况与父母社会经济地位进行交叉分类，分类结果见表9.2.3。

表9. 2. 3 心理健康状况-父母社会经济状况数据

心理健康 状况 \ 父母社会 经济地位					
	A（高）	B	C	D	E（低）
0（好）	121	57	72	36	21
1（轻微症状形成）	188	105	141	97	71
2（中等症状形成）	112	65	77	54	54
3（受损）	86	60	94	78	71

表9. 2. 4

从表9. 2. 3算得的对应矩阵

父母社会 经济状况 心理健康 状况	A（高）	B	C	D	E（低）	合 计
0（好）	0.073	0.034	0.043	0.022	0.013	0.185
1（轻微症状形成）	0.113	0.063	0.085	0.058	0.043	0.363
2（中等症状形成）	0.067	0.039	0.046	0.033	0.033	0.218
3（受损）	0.052	0.036	0.057	0.047	0.043	0.234
合 计	0.305	0.173	0.231	0.160	0.131	1.000

➤ 表9.2.4给出的行边缘频率和列边缘频率向量为

$$\mathbf{r} = \begin{pmatrix} 0.185 \\ 0.363 \\ 0.218 \\ 0.234 \end{pmatrix}, \quad \mathbf{c}' = (0.305, 0.173, 0.231, 0.160, 0.131)$$

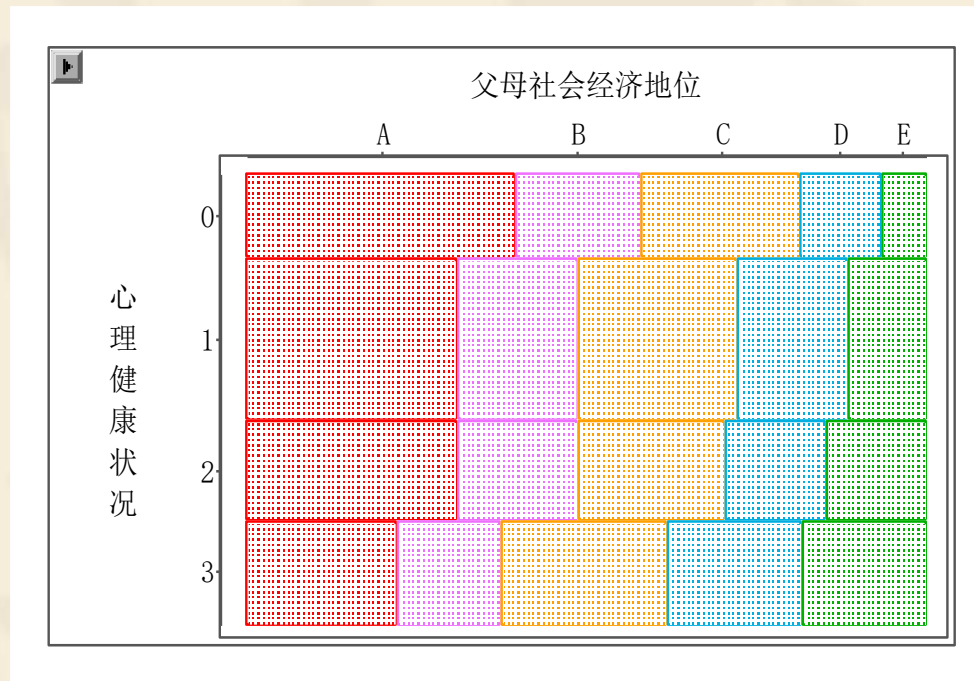
➤ 行轮廓矩阵为

$$\mathbf{R} = \mathbf{D}_r^{-1} \mathbf{P} = \begin{pmatrix} 0.394 & 0.186 & 0.235 & 0.117 & 0.068 \\ 0.312 & 0.174 & 0.234 & 0.161 & 0.118 \\ 0.309 & 0.180 & 0.213 & 0.149 & 0.149 \\ 0.221 & 0.154 & 0.242 & 0.201 & 0.183 \end{pmatrix}$$

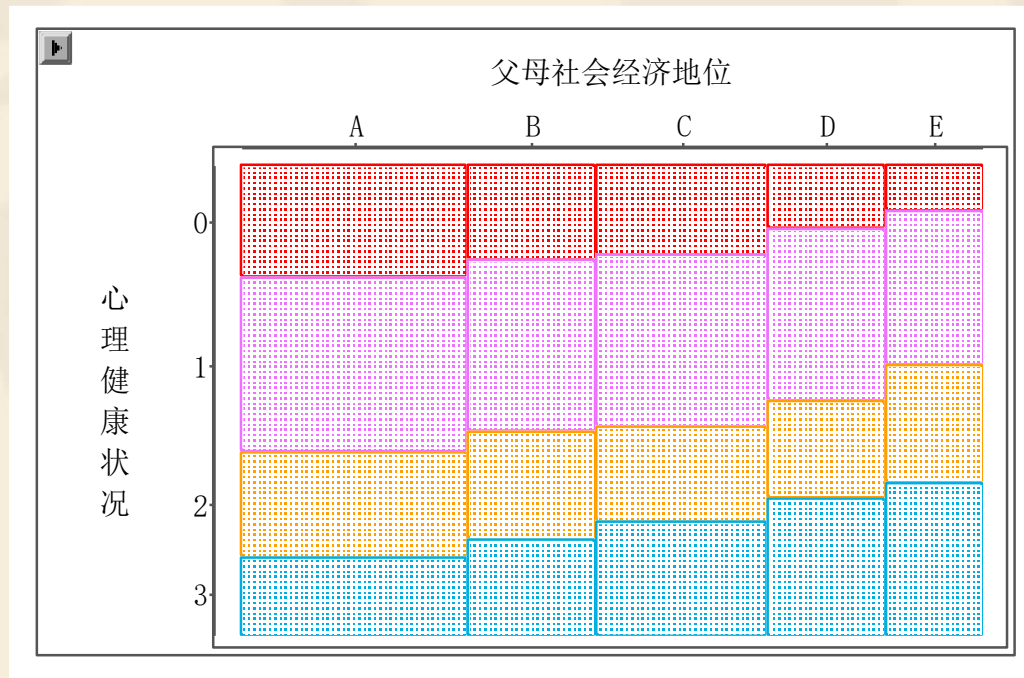
➤ 列轮廓矩阵为

$$\mathbf{C} = \mathbf{P} \mathbf{D}_c^{-1} = \begin{pmatrix} 0.239 & 0.199 & 0.188 & 0.136 & 0.097 \\ 0.371 & 0.366 & 0.367 & 0.366 & 0.327 \\ 0.221 & 0.226 & 0.201 & 0.204 & 0.249 \\ 0.170 & 0.209 & 0.245 & 0.294 & 0.327 \end{pmatrix}$$

两个马赛克图



对心理健康的每一种状况，A,B,C,D,E五个小方块的宽度显示了行轮廓，0,1,2,3四种心理健康状况的小方块高度显示了行边缘频率。



对父母社会经济的每一种地位，0,1,2,3四个小方块的高度显示了列轮廓，A,B,C,D,E五种父母社会经济地位的小方块宽度显示了列边缘频率。

§ 9.3 独立性的检验和总惯量

- ❖ 一、行、列独立的检验
- ❖ 二、总惯量

一、行、列独立的检验

- ❖ 在列联表中，检验行变量和列变量相互独立假设的统计量为

$$\chi^2 = n \sum_{i=1}^p \sum_{j=1}^q \frac{(p_{ij} - p_{i.}p_{.j})^2}{p_{i.}p_{.j}}$$

- ❖ 当独立性的原假设为真，且样本容量 n 充分大，期望频数 $np_{i.}p_{.j} \geq 5$ ， $i = 1, 2, \dots, p$ ， $j = 1, 2, \dots, q$ 时，

$$\chi^2 \sim \chi^2[(p-1)(q-1)]$$

- ❖ 拒绝规则为

若 $\chi^2 \geq \chi_{\alpha}^2[(p-1)(q-1)]$ ，则拒绝独立性的原假设

❖ χ^2 值取决于 n 和 $\sum_{i=1}^p \sum_{j=1}^q \frac{(p_{ij} - p_{i \cdot} p_{\cdot j})^2}{p_{i \cdot} p_{\cdot j}}$ 这两部分。

- $\sum_{i=1}^p \sum_{j=1}^q \frac{(p_{ij} - p_{i \cdot} p_{\cdot j})^2}{p_{i \cdot} p_{\cdot j}}$ 越大，表明实际频率 p_{ij} 与独立假设下的期望频率 $p_{i \cdot} p_{\cdot j}$ 总体上差异越大，也就认为样本数据越是偏离行、列变量相互独立的情形，从而越应拒绝独立性的原假设。
- n 越大，表明样本所含的信息越多，越易检测出对原假设的偏离。
- ❖ 如果表9.2.1的列联表中有些单元格的频数很小或为零，上述的 χ^2 近似就不会很令人满意，在这种情况下或需借助于对应分析将一些具有相近行轮廓（或列轮廓）的类别合并以增加单元格的频数。

二、总惯量

❖

$$\text{总惯量} = \frac{\chi^2}{n} = \sum_{i=1}^p \sum_{j=1}^q \frac{(p_{ij} - p_{i.}p_{.j})^2}{p_{i.}p_{.j}}$$

❖ 总惯量可作为行、列变量之间关联性的度量。

❖ 例9.3.1 例9.2.1中， $\chi^2=45.594 > 21.026 = \chi_{0.05}^2(12)$ ，

故拒绝心理健康状况与父母社会经济地位相互独立的原假设（ $p=8.15 \times 10^{-6}$ ）。

$$\text{总惯量} = \frac{\chi^2}{n} = \frac{45.594}{1660} = 0.0275$$

❖ 总惯量还可以行轮廓和列轮廓的形式表达如下：

$$\text{总惯量} = \sum_{i=1}^p p_{i\cdot} \sum_{j=1}^q \frac{(p_{ij}/p_{i\cdot} - p_{\cdot j})^2}{p_{\cdot j}} = \sum_{i=1}^p p_{i\cdot} (\mathbf{r}_i - \mathbf{c})' \mathbf{D}_c^{-1} (\mathbf{r}_i - \mathbf{c})$$

$$\text{总惯量} = \sum_{j=1}^q p_{\cdot j} \sum_{i=1}^p \frac{(p_{ij}/p_{\cdot j} - p_{i\cdot})^2}{p_{i\cdot}} = \sum_{j=1}^q p_{\cdot j} (\mathbf{c}_j - \mathbf{r})' \mathbf{D}_r^{-1} (\mathbf{c}_j - \mathbf{r})$$

其中

$$(\mathbf{r}_i - \mathbf{c})' \mathbf{D}_c^{-1} (\mathbf{r}_i - \mathbf{c}) = \sum_{j=1}^q \frac{(p_{ij}/p_{i\cdot} - p_{\cdot j})^2}{p_{\cdot j}}$$

称为第*i*行轮廓 \mathbf{r}_i 到行轮廓中心 \mathbf{c} 的卡方 (χ^2) 距离，它可看作是一个权数为 $1/p_{\cdot j}$ ($j=1, 2, \dots, q$) 的加权平方欧氏距离。

❖ 同样，

$$\left(\mathbf{c}_j - \mathbf{r}\right)' \mathbf{D}_r^{-1} \left(\mathbf{c}_j - \mathbf{r}\right) = \sum_{i=1}^p \frac{\left(p_{ij} / p_{\cdot j} - p_{i\cdot}\right)^2}{p_{i\cdot}}$$

是第 j 列轮廓 \mathbf{c}_j 到列轮廓中心 \mathbf{r} 的卡方距离。

- ❖ 故总惯量可看成是行轮廓到其中心的卡方距离的加权平均，也可看成是列轮廓到其中心的卡方距离的加权平均。它既度量了行轮廓之间的总变差，也度量了列轮廓之间的总变差。
- ❖ 由此可见，行和列之间的关联性越强，行（列）轮廓之间的差异性就越大；反之亦然。

总惯量为零的等价情形

- ❖ 总惯量为零与以下三种情形的任一种等价：
 - (1) $p_{ij} = p_{i.}p_{.j}$, $i = 1, 2, \dots, p$, $j = 1, 2, \dots, q$, 或表示为 $\mathbf{P} = \mathbf{r}\mathbf{c}'$;
 - (2) 所有的行轮廓相等, 即 $\mathbf{r}'_1 = \mathbf{r}'_2 = \dots = \mathbf{r}'_p = \mathbf{c}'$;
 - (3) 所有的列轮廓相等, 即 $\mathbf{c}_1 = \mathbf{c}_2 = \dots = \mathbf{c}_q = \mathbf{r}$ 。
- ❖ 所以, 如果行变量与列变量相互独立, 则我们可以期望 (由样本数据构成的) 列联表中所有的行有相近的轮廓, 所有的列亦有相近的轮廓。

- ❖ 例9.3.2 在表9.3.1中，显然所有的行轮廓都相同，从而总惯量必为零，故这也意味着所有的列轮廓也都相同，且每一行、列类别组合的实际频数等于相应的期望频数。

表9.3.1 行（或列）轮廓都相同的数据

列变量 行变量					合计
	A	B	C	D	
1	10	34	7	4	55
2	100	340	70	40	550
3	30	102	21	12	165
合计	140	476	98	56	770

总惯量的分解

❖ 对 $\mathbf{P} - \mathbf{r}\mathbf{c}'$ 构造标准化矩阵

$$\mathbf{Z} = \mathbf{D}_r^{-1/2} (\mathbf{P} - \mathbf{r}\mathbf{c}') \mathbf{D}_c^{-1/2}$$

其元素为

$$z_{ij} = \frac{p_{ij} - p_{i\cdot}p_{\cdot j}}{\sqrt{p_{i\cdot}p_{\cdot j}}}$$

记 $k = \text{rank}(\mathbf{Z})$, 有 $k \leq \min(p-1, q-1)$, 这是因为

$$(\mathbf{P} - \mathbf{r}\mathbf{c}')\mathbf{1} = \mathbf{P}\mathbf{1} - \mathbf{r}\mathbf{c}'\mathbf{1} = \mathbf{r} - \mathbf{r} = \mathbf{0}$$

$$\mathbf{1}'(\mathbf{P} - \mathbf{r}\mathbf{c}') = \mathbf{c}' - \mathbf{c}' = \mathbf{0}'$$

❖ 对 \mathbf{Z} 进行奇异值分解，得

$$\mathbf{Z} = \mathbf{U}\mathbf{A}\mathbf{V}' = \sum_{i=1}^k \lambda_i \mathbf{u}_i \mathbf{v}_i'$$

其中 $\mathbf{U}=(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k)$, $\mathbf{V}=(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$, $\mathbf{A}=\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$, 这里 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ 是一组 p 维正交单位向量, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ 是一组 q 维正交单位向量, 即有 $\mathbf{U}'\mathbf{U}=\mathbf{V}'\mathbf{V}=\mathbf{I}_k$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$ 是 \mathbf{Z} 的 k 个奇异值。于是, $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_k^2 > 0$ 是 $\mathbf{Z}\mathbf{Z}'$ 的正特征值。

❖

$$\begin{aligned} \text{总惯量} &= \sum_{i=1}^p \sum_{j=1}^q \frac{(p_{ij} - p_{i.} p_{.j})^2}{p_{i.} p_{.j}} = \sum_{i=1}^p \sum_{j=1}^q z_{ij}^2 \\ &= \text{tr}(\mathbf{Z}\mathbf{Z}') = \sum_{i=1}^k \lambda_i^2 \end{aligned}$$

§ 9.4 行、列轮廓的坐标

❖ 令

$$\underset{p \times k}{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k) = \underset{r}{D}^{1/2} \mathbf{U}, \quad \underset{q \times k}{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k) = \underset{c}{D}^{1/2} \mathbf{V}$$

❖ $\text{rank}(A) = \text{rank}(\mathbf{U}) = k$, $\text{rank}(B) = \text{rank}(\mathbf{V}) = k$

$\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$ ——一组线性无关的 p 维向量

$\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$ ——一组线性无关的 q 维向量

❖ $\mathbf{r}'_i - \mathbf{c}' = x_{i1} \mathbf{b}'_1 + x_{i2} \mathbf{b}'_2 + \dots + x_{ik} \mathbf{b}'_k, \quad i = 1, 2, \dots, p$

即它在由 $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$ 构成的坐标系中的坐标为

$(x_{i1}, x_{i2}, \dots, x_{ik}), i = 1, 2, \dots, p$ 。

❖ 中心化的第 j 列轮廓为

$$\mathbf{c}_j - \mathbf{r} = y_{j1}\mathbf{a}_1 + y_{j2}\mathbf{a}_2 + \cdots + y_{jk}\mathbf{a}_k, \quad j = 1, 2, \cdots, q$$

即它在由 $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k$ 构成的坐标系中的坐标为
 $(y_{j1}, y_{j2}, \cdots, y_{jk}), j=1, 2, \cdots, q$ 。

❖

$$\begin{aligned} 0 &= \sum_{i=1}^p p_i (\mathbf{r}'_i - \mathbf{c}') = \sum_{i=1}^p p_i (x_{i1}\mathbf{b}'_1 + x_{i2}\mathbf{b}'_2 + \cdots + x_{ik}\mathbf{b}'_k) \\ &= \left(\sum_{i=1}^p p_i x_{i1} \right) \mathbf{b}'_1 + \left(\sum_{i=1}^p p_i x_{i2} \right) \mathbf{b}'_2 + \cdots + \left(\sum_{i=1}^p p_i x_{ik} \right) \mathbf{b}'_k \end{aligned}$$

从而

$$\sum_{j=1}^p p_j x_{ji} = 0, \quad i = 1, 2, \cdots, k$$

即各行点在坐标轴 \mathbf{b}_i 上坐标的加权平均值为0, $i=1,2,\dots,k$ 。

❖ 同理可得

$$\sum_{j=1}^q p_{\cdot j} y_{ji} = 0, \quad i=1,2,\dots,k$$

即各列点在坐标轴 \mathbf{a}_i 上坐标的加权平均值也为0, $i=1,2,\dots,k$ 。

❖
$$\sum_{j=1}^p p_{\cdot j} x_{ji}^2 = \sum_{j=1}^q p_{\cdot j} y_{ji}^2 = \lambda_i^2, \quad i=1,2,\dots,k$$

- 即各行点和列点在第 i 坐标轴上的坐标平方的加权平均都等于 λ_i^2 , 称之为第 i 主惯量或第 i 惯量, $i=1,2,\dots,k$ 。
- 主惯量度量了在每一坐标轴上的变差, 类似于主成分的方差。
- ❖ 总惯量可以分解为各主惯量之和。
- ❖ 从视觉上可以接受作图时将行点和列点置于同一个坐标系中, 并使用同一坐标刻度。

§ 9.5 对应分析图

- ❖ 一、对应分析图的构建
- ❖ 二、行（列）点之间的距离
- ❖ 三、行点和列点相近的意涵

❖ 当 $\sum_{i=1}^m \lambda_i^2 / \sum_{i=1}^k \lambda_i^2$ 足够大时，有

$$\mathbf{r}'_i - \mathbf{c}' \approx x_{i1} \mathbf{b}'_1 + \cdots + x_{im} \mathbf{b}'_m, \quad i = 1, 2, \dots, p$$

$$\mathbf{c}_j - \mathbf{r} \approx y_{j1} \mathbf{a}_1 + \cdots + y_{jm} \mathbf{a}_m, \quad j = 1, 2, \dots, q$$

- ❖ 为了作图的目的，我们通常取维数 $m = 1, 2, 3$ ，可将各（中心化的）行轮廓和列轮廓在 m 维坐标系上用点标出，并同时作到同一张图上。
- ❖ 在以下的讨论中，为表述方便，我们只涉及最常用、最典型的 $m=2$ 时的情形， $m=1$ 或 3 时情形的讨论是完全类似的。

一、对应分析图的构建

- ❖ $\mathbf{r}'_i - \mathbf{c}'$ 在由 \mathbf{b}_1 和 \mathbf{b}_2 构成的平面坐标系中的坐标是 (x_{i1}, x_{i2}) , $\mathbf{c}_j - \mathbf{r}$ 在由 \mathbf{a}_1 和 \mathbf{a}_2 构成的平面坐标系中的坐标是 (y_{j1}, y_{j2}) 。
- ❖ 将上述两个平面坐标系重叠在一个坐标系中, \mathbf{b}_1 和 \mathbf{a}_1 重叠在第一维坐标轴上, 具有同一主惯量 λ_1^2 , 其对总惯量的贡献率为 $\lambda_1^2 / \sum_{i=1}^k \lambda_i^2$ 。 \mathbf{b}_2 和 \mathbf{a}_2 重叠在第二维坐标轴上, 具有同一主惯量 λ_2^2 , 其对总惯量的贡献率为 $\lambda_2^2 / \sum_{i=1}^k \lambda_i^2$ 。

- ❖ 前二维对总惯量的累计贡献率为 $(\lambda_1^2 + \lambda_2^2) / \sum_{i=1}^k \lambda_i^2$, 该值如很大, 则说明所作的对应分析图几乎解释了列联表数据的所有变差。
- ❖ 虽然 \mathbf{b}_1 与 \mathbf{b}_2 及 \mathbf{a}_1 与 \mathbf{a}_2 一般都不正交, 但作图时我们仍将第一维和第二维坐标轴画成是垂直的。如果还要取第三维坐标轴, 则这三个维的坐标轴应作成彼此垂直。

二、行（列）点之间的距离

- ❖ 在累计贡献率 $(\lambda_1^2 + \lambda_2^2) / \sum_{i=1}^k \lambda_i^2$ 足够大的对应分析图中，第*i*个行点 (x_{i1}, x_{i2}) 与第*j*个行点 (x_{j1}, x_{j2}) 之间的平方欧氏距离

$$d_{ij}^2(r) = (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2$$
$$\approx (\mathbf{r}_i - \mathbf{r}_j)' \mathbf{D}_c^{-1} (\mathbf{r}_i - \mathbf{r}_j) = \tilde{d}_{ij}^2(r)$$

- ❖ 类似地，第*i*个列点 (y_{i1}, y_{i2}) 与第*j*个列点 (y_{j1}, y_{j2}) 之间的平方欧氏距离

$$d_{ij}^2(c) = (y_{i1} - y_{j1})^2 + (y_{i2} - y_{j2})^2$$
$$\approx (\mathbf{c}_i - \mathbf{c}_j)' \mathbf{D}_r^{-1} (\mathbf{c}_i - \mathbf{c}_j) = \tilde{d}_{ij}^2(c)$$

- ❖ 可见，如果两个行（列）点越接近（远离），则表明相应的两个行（列）轮廓越相似（不相似）。
- ❖ 此外，对应分析图中行（列）点的方位是富有意义的，而行点与列点之间的距离并没有意义。

三、行点和列点相近的意涵

- ❖ 如果对应分析图上第 i 个行点和第 j 个列点相近，即有

$$(x_{i1}, x_{i2}) \approx (y_{j1}, y_{j2})$$

- ❖ 则在 $(\lambda_1^2 + \lambda_2^2) / \sum_{i=1}^k \lambda_i^2$ 足够大的条件下，有

$$\frac{n_{ij} - n_{i.}n_{.j}/n}{n_{i.}n_{.j}/n} \approx \frac{x_{i1}^2}{\lambda_1} + \frac{x_{i2}^2}{\lambda_2} \geq 0$$

$$\frac{n_{ij} - n_{i.}n_{.j}/n}{n_{i.}n_{.j}/n} \approx \frac{y_{j1}^2}{\lambda_1} + \frac{y_{j2}^2}{\lambda_2} \geq 0$$

- ❖ 可见，如果一个行点和一个列点相近，则表明行、列两个变量的相应类别组合发生的实际频数 (n_{ij}) 一般会高于这两个变量相互独立情形下的期望频数 ($n_{i.}n_{.j}/n$)，也就意味着该行类别与该列类别相关联。
- ❖ 一般地，对于相近的行点和列点，它们离原点越远，其关联性就越强，也就是其类别组合的实际频数越是明显高于两变量独立情形下的期望频数。如果它们都在原点附近，则其关联性一般较弱、甚至可能几乎无关联性。

❖ 例9.5.1 在例9.2.1中，经计算，奇异值、主惯性以及贡献率等的计算结果列于表9.5.1中。总惯量的94.75%可由第一维来解释，前二维解释了高达99.76%的总惯量，几乎解释了列联表数据的所有方差。

表9.5.1

奇异值、主惯量以及贡献率

维数	1	2	3	
奇异值	0.1613	0.0371	0.0082	总值
主惯量	0.0260	0.0014	0.0001	0.0275
贡献率	0.9475	0.0501	0.0024	1.0000
累计贡献率	0.9475	0.9976	1.0000	

➤ 行点和列点的前二维坐标矩阵为

$$\mathbf{X}_1 = \begin{pmatrix} 0.260 & -0.013 \\ 0.030 & -0.023 \\ -0.014 & 0.070 \\ -0.237 & -0.020 \end{pmatrix}, \quad \mathbf{Y}_1 = \begin{pmatrix} 0.183 & 0.016 \\ 0.059 & 0.022 \\ -0.009 & -0.042 \\ -0.165 & -0.043 \\ -0.288 & 0.062 \end{pmatrix}$$

将各行点和列点置于同一坐标系中，构成对应分析图，如下图所示。

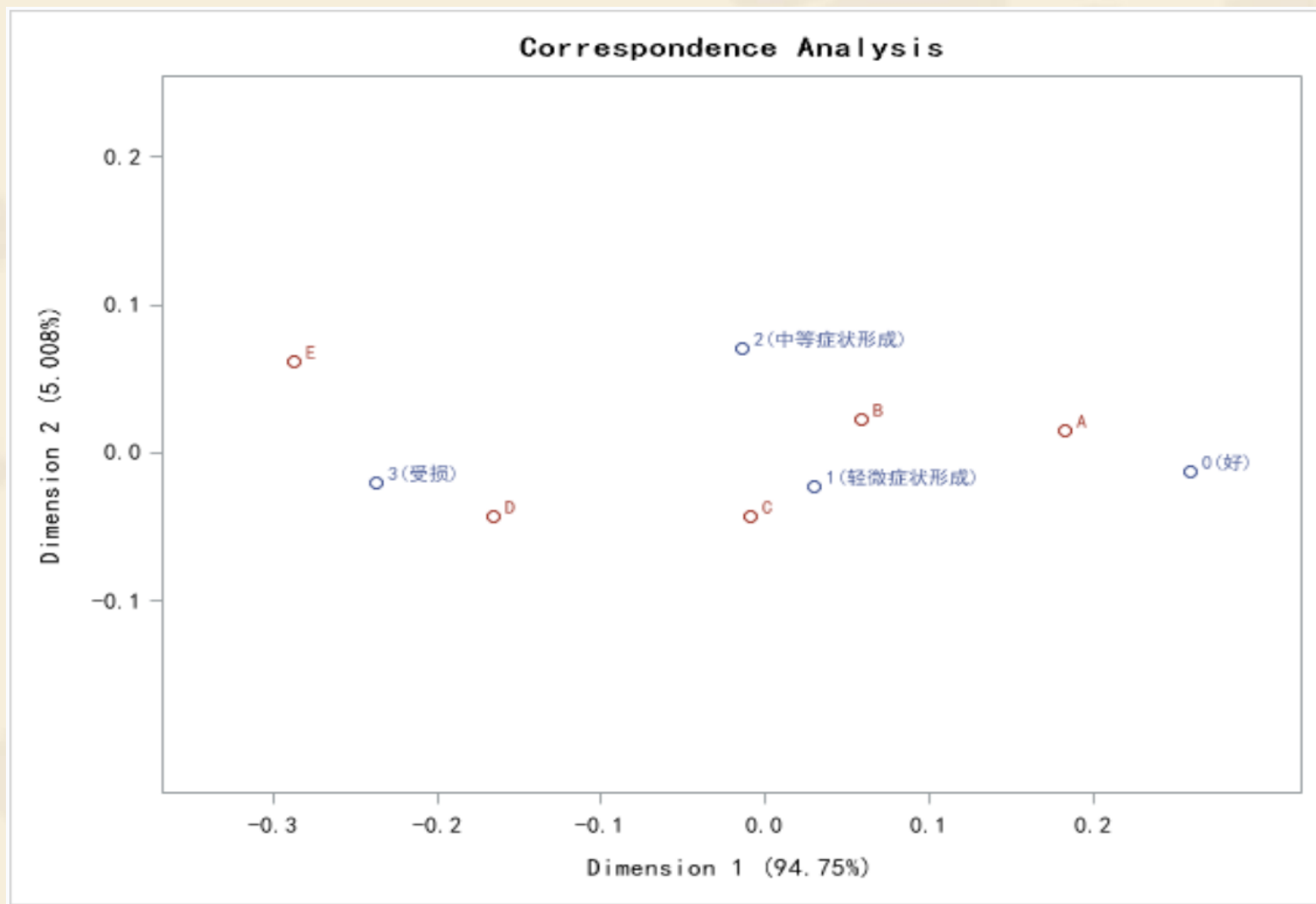


图9.5.1 心理健康状况-父母社会经济地位数据的对应分析图

表9.5.2 行点和列点靠近的分类组合频数及行、列独立情形下的频数期望值

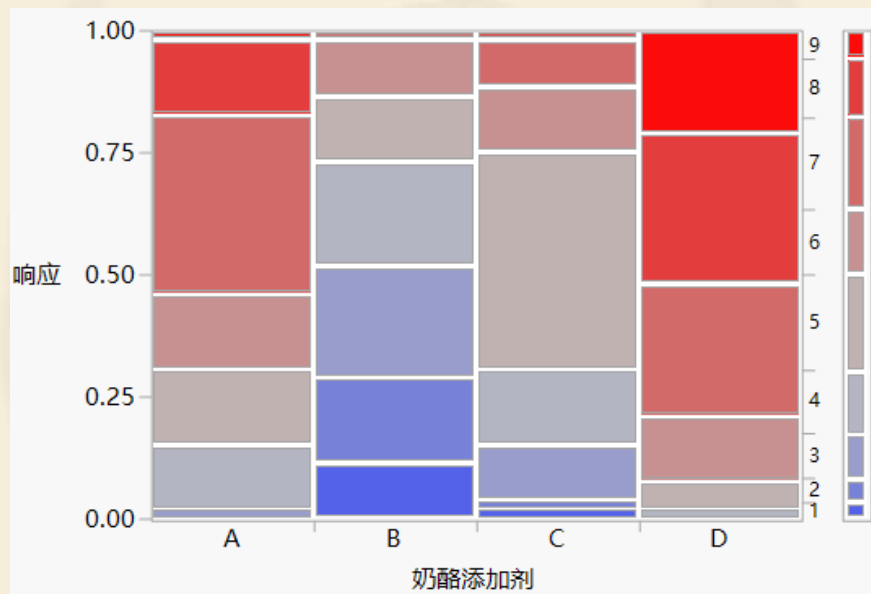
父母社会 经济地位 心理健康状况					
	A（高）	B	C	D	E（低）
0（好）	121（93.8）	57	72	36	21
1（轻微症状形成）	188	105（104.1）	141（139.3）	97	71
2（中等症状形成）	112	65（62.6）	77	54	54
3（受损）	86	60	94	78（62.1）	71（50.9）

- ❖ 例9.5.2 表9.5.3中的数据来源于奶酪品尝的实验，实验记录了九种不同响应和四种不同奶酪添加剂的交叉频数。九种不同的响应是从最不喜欢到最喜欢，品尝者依次打分为1,2,...,9，四种不同的奶酪添加剂分别为A,B,C,D。

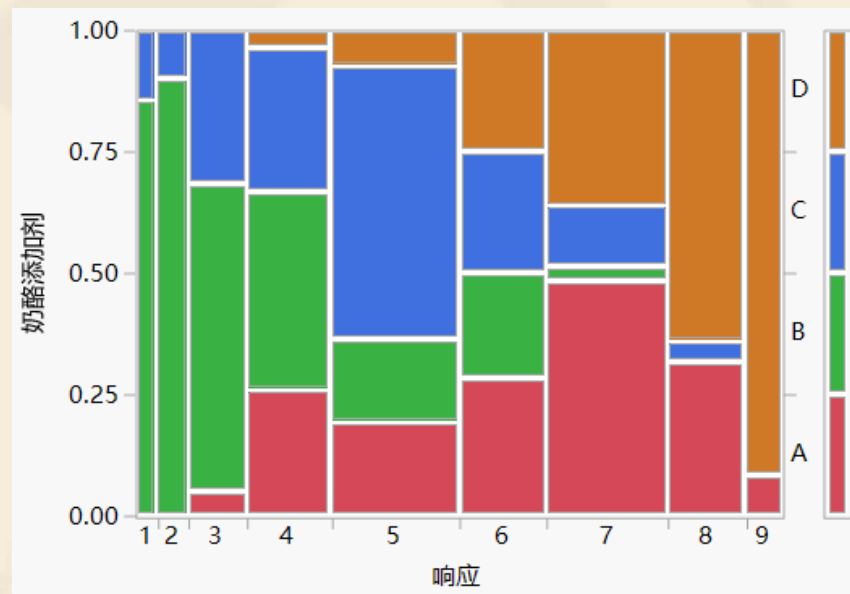
表9.5.3

奶酪品尝的实验数据

编号	奶酪添加剂	响应	频数	编号	奶酪添加剂	响应	频数
1	A	1	0	19	C	1	1
2	A	2	0	20	C	2	1
3	A	3	1	21	C	3	6
4	A	4	7	22	C	4	8
5	A	5	8	23	C	5	23
6	A	6	8	24	C	6	7
7	A	7	19	25	C	7	5
8	A	8	8	26	C	8	1
9	A	9	1	27	C	9	0
10	B	1	6	28	D	1	0
11	B	2	9	29	D	2	0
12	B	3	12	30	D	3	0
13	B	4	11	31	D	4	1
14	B	5	7	32	D	5	3
15	B	6	6	33	D	6	7
16	B	7	1	34	D	7	14
17	B	8	0	35	D	8	16
18	B	9	0	36	D	9	11



(1) 奶酪添加剂轮廓及边缘频率



(2) 响应轮廓及边缘频率

图9.5.2 奶酪添加剂-响应数据的轮廓及边缘频率

详细信息			
奇异值	惯量	对应部分	累积
0.73609	0.54183	0.6936	0.6936
0.42010	0.17649	0.2259	0.9195
0.25070	0.06285	0.0805	1.0000
奶酪添加剂	c1	c2	c3
A	-0.3763	-0.2528	-0.3865
B	0.9553	0.4728	-0.0554
C	0.3981	-0.5540	0.2467
D	-0.9771	0.3340	0.1952
响应	c1	c2	c3
1	1.190	0.7764	-0.0490
2	1.222	0.8811	-0.1006
3	0.964	0.2628	0.0900
4	0.507	-0.0588	-0.1693
5	0.328	-0.6068	0.2705
6	-0.065	-0.0617	-0.0472
7	-0.623	-0.1480	-0.3510
8	-0.991	0.2634	0.0443
9	-1.259	0.6786	0.5852

图9.5.3 奇异值、主惯量、贡献率以及行、列点的坐标

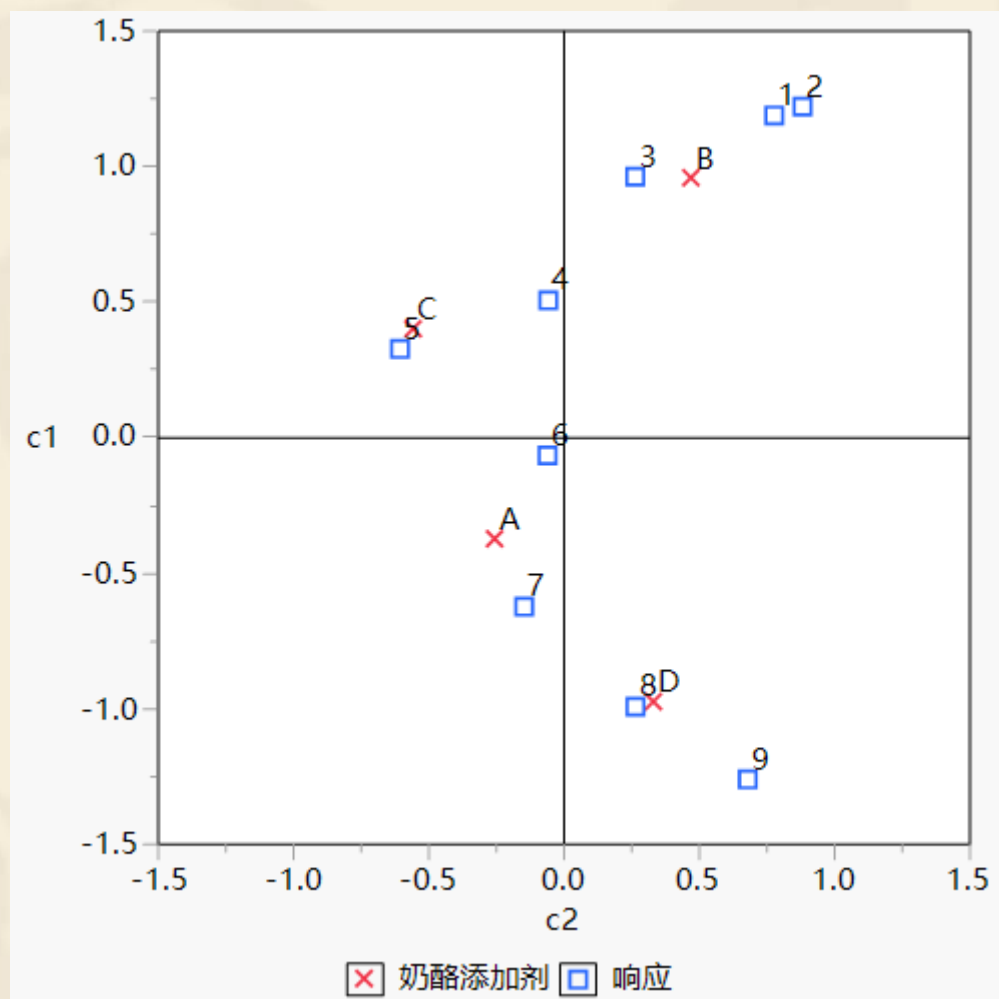


图9.5.4 奶酪添加剂-响应数据的对应分析图

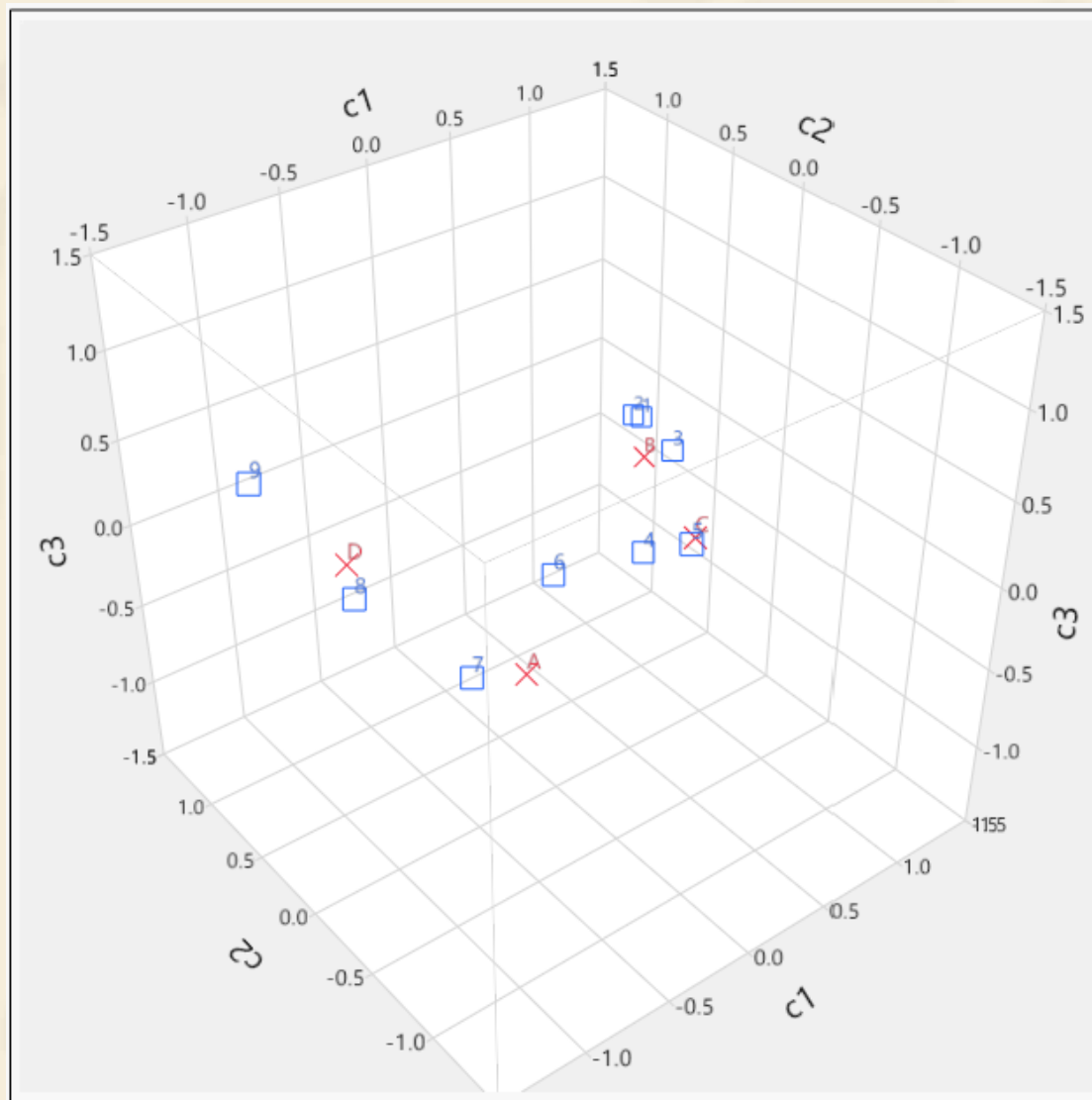


图9.5.5 奶酪添加剂-响应数据的三维对应分析图

		响应									
计数	1	2	3	4	5	6	7	8	9	合计	
期望值											
奶酪添加剂	A	0	0	1	7	8	8	19	8	1	52
		1.75	2.5	4.75	6.75	10.25	7	9.75	6.25	3	
	B	6	9	12	11	7	6	1	0	0	52
		1.75	2.5	4.75	6.75	10.25	7	9.75	6.25	3	
	C	1	1	6	8	23	7	5	1	0	52
		1.75	2.5	4.75	6.75	10.25	7	9.75	6.25	3	
	D	0	0	0	1	3	7	14	16	11	52
		1.75	2.5	4.75	6.75	10.25	7	9.75	6.25	3	
	合计	7	10	19	27	41	28	39	25	12	208

图9.5.6 类别组合的实际频数及两变量独立情形下的期望频数

		响应								
奶酪添加剂	行百分比	1	2	3	4	5	6	7	8	9
	A	0.00	0.00	1.92	13.46	15.38	15.38	36.54	15.38	1.92
	B	11.54	17.31	23.08	21.15	13.46	11.54	1.92	0.00	0.00
	C	1.92	1.92	11.54	15.38	44.23	13.46	9.62	1.92	0.00
	D	0.00	0.00	0.00	1.92	5.77	13.46	26.92	30.77	21.15

图9.5.7 奶酪添加剂的轮廓（%）