

对应分析

Correspondence Analysis

案例1

起名为“波澜”恰当吗



娃哈哈公司欲为其新推出的一种纯水产品起一个合适的名字，为此专门委托了当地的策划咨询公司，取了一个名字“波澜”。一个好的品牌名称至少应该满足两个条件：

- (1) 会使消费者联想到正确的产品“纯水”；
- (2) 会使消费者产生与正确产品密切相关的联想，如“纯净”、“清爽”等。

公司决定对“波澜”这一名称方案进行品牌测试。采用调查问卷的形式进行消费者调查，以便最终确定品牌名称。

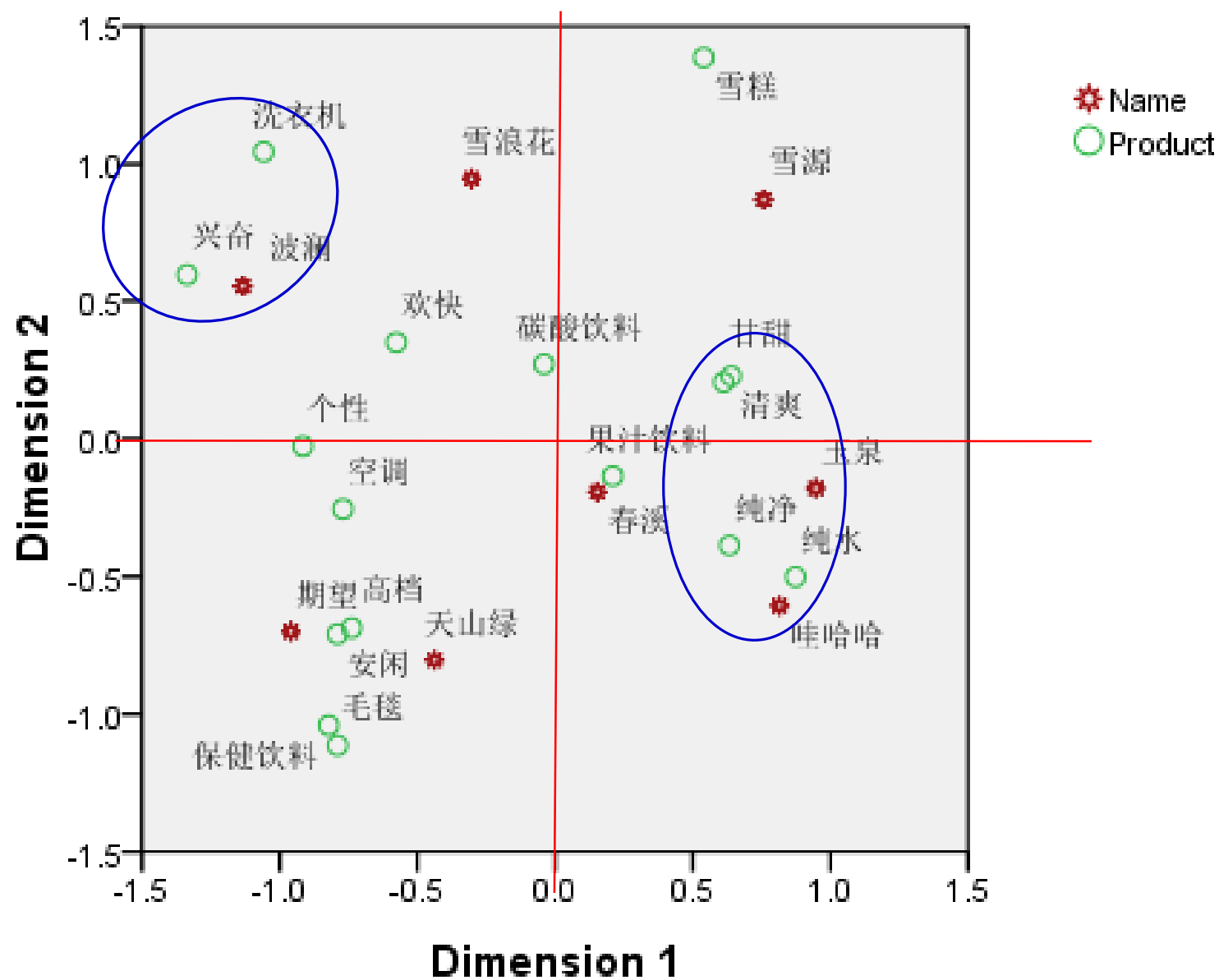
- 娃哈哈公司委托调查咨询机构，进行了一次全面的市场研究，在调查中还包括简单的名称测试。
- 调查的代码和含义如下：

调查代码和含义

代码	含义	代码	含义	代码	含义
Name1	玉泉	Product1	雪糕	Feel1	清爽
Name2	雪源	Product2	纯水	Feel2	甘甜
Name3	春溪	Product3	碳酸饮料	Feel3	欢快
Name4	期望	Product4	果汁饮料	Feel4	纯净
Name5	波澜	Product5	保健食品	Feel5	安闲
Name6	天山绿	Product6	空调	Feel6	个性
Name7	哇哈哈	Product7	洗衣机	Feel7	兴奋
Name8	雪浪花	Product8	毛毯	Feel8	高档

市场调查数据汇总

品牌名称 产品属性	玉泉	雪源	春溪	期望	波澜	天山绿	哇哈哈	雪浪花
雪糕	50	442	27	21	14	50	30	258
纯水	508	110	272	51	83	88	605	79
碳酸饮料	55	68	93	36	71	47	37	77
果汁饮料	109	95	149	41	36	125	44	65
保健食品	34	29	45	302	37	135	42	18
空调	11	28	112	146	113	39	28	31
洗衣机	30	12	54	64	365	42	8	316
毛毯	2	4	17	36	29	272	9	35
清爽	368	322	167	53	57	129	149	170
甘甜	217	237	142	41	34	95	119	116
欢快	19	25	185	105	123	44	22	193
纯净	142	140	128	47	38	123	330	68
安闲	16	16	106	166	81	164	21	36
个性	2	14	9	72	94	41	37	42
兴奋	4	11	10	78	248	35	17	81
高档	3	5	19	107	63	126	63	49



由图可以直观看出，“波澜”与“洗衣机”产品相联系，引起的感觉是“兴奋”，因此“波澜”不是合适的纯净水品牌名称。

娃哈哈公司的产品是“纯水”如果想要使该名称给人们一种“纯净”的感觉，那么“**娃哈哈**”将是最好的商品名称。如果想要使该名称给人们一种“清爽” 的感觉，那么“玉泉”将是最好的商品名称。

结论：

不如叫“玉泉”或“娃哈哈”吧！



什么是对应分析

§ 1 对应分析基本思想

- 对应分析，也称**关联分析**、**R-Q型因子分析**，是由法国人 Beozecri于1970年提出来的。
- 通过分析由**定性变量**构成的**交互汇总表**来揭示变量之间的联系。

- 在实际研究工作中，人们常常用列联表的形式来描述属性变量（定类尺度或定序尺度）的各种状态或相互关系。比如，公司的管理者为了解消费者对产品的满意情况，针对不同职业的消费者进行调查，而调查数据很自然地就以列联表的形式展示。

评价 职业	非常满意	比较满意	一般	不太满意	不满意	汇总
教师	25	42	26	10	3	106
医生	21	35	46	10	4	116
公司职员	58	72	60	17	8	215
管理者	14	22	23	9	2	70
.....						
汇总	118	171	155	46	17	

- ◆ 通过列联表，可以清楚地显示不同职业的人对该公司产品的评价，以及所有被调查者对该公司产品的整体评价、被调查者的职业构成情况等信息；
- ◆ 通过列联表，还可以看出职业分布与各种评价之间的相互关系，如“教师”与比较满意交叉单元格的数字相对较大（“相对”指应抵消不同职业在总的被调查对象中的比例的影响），则说明职业栏的“教师”这一部分与评价栏的比较满意这一部分有较强的相关性；
- ◆ 借助列联表，可以得到很多有价值的信息。

		特性 B						合计
		B_1	B_2	...	B_j	...	B_p	
特性 A	A_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1p}	$n_{1.}$
	A_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2p}	$n_{2.}$
	\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
	A_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ip}	$n_{i.}$
	\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
	A_n	n_{n1}	n_{n2}	...	n_{nj}	...	n_{np}	$n_{n.}$
合计		$n_{.1}$	$n_{.2}$...	$n_{.j}$...	$n_{.p}$	n

- 为了更方便的表示各频数之间的关系，人们往往用频率来代替频数，即将列联表中每一个元素都除以元素的总和 n ，令 $p_{ij} = \frac{n_{ij}}{n}$ ，得到如下频率意义上的列联表：

若特性A与特性B之间相互独立

		B_1	B_2	...	B_j	...	B_p	合计
特性 A	A_1	p_{11}	p_{12}	...	p_{1j}	...	p_{1p}	$p_{1\cdot}$
	A_2	p_{21}	p_{22}	...	p_{2j}	...	p_{2p}	$p_{2\cdot}$
	\vdots	\vdots					\vdots	\vdots
	A_i	p_{i1}	p_{i2}	...	p_{ij}	...	p_{ip}	$p_{i\cdot}$
	A_n	p_{n1}	p_{n2}	...	p_{nj}	...	p_{np}	$p_{n\cdot}$
合计		$p_{\cdot 1}$	$p_{\cdot 2}$...	$p_{\cdot j}$...	$p_{\cdot p}$	1

$$p_{ij} = p_{i\cdot} \times p_{\cdot j}$$

$$E(n_{ij}) = np_{ij} = n \times p_{i\cdot} \times p_{\cdot j}$$

- **A与B为属性变量**，它指的是列联表的横栏与纵栏按某种规则的分类，我们关心的是属性变量**A与B**是否独立，由此提出如下假设：

H_0 ：属性变量**A与B**相互独立；

H_1 ：属性变量**A与B**不独立。

因而构建如下统计量：

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^p \frac{[n_{ij} - E(n_{ij})]^2}{E(n_{ij})} = n \sum_{i=1}^n \sum_{j=1}^p \frac{[p_{ij} - p_{i\bullet} p_{\bullet j}]^2}{p_{i\bullet} p_{\bullet j}}$$

若 $\chi^2 > \chi_{\alpha}^2[(n-1)(p-1)]$ ，拒绝 H_0

§ 1 对应分析基本思想

- 对应分析，也称**关联分析**、**R-Q型因子分析**，是由法国人 Beozecri于1970年提出来的。
- 通过分析由**定性变量**构成的**交互汇总表**来揭示变量之间的联系。
- 可以揭示同一变量的各个类别之间的差异，以及不同变量各个类别之间的对应关系。
- 它是一种**视觉化**的数据分析方法，它能够将几组看不出任何联系的数据，通过视觉上可以接受的定位图展现出来。

§ 1 对应分析基本思想

一种非常有用的市场研究工具，可以表述一个市场的侧面（市场细分、产品定位等）。

可以在2维空间内同时表达多维的属性

可以更好地理解品牌和属性之间的关系

有助于发现市场空隙，优化产品定位

§ 1 对应分析基本思想

在因子分析中人们通常只是分析原始变量的因子结构，找出决定原始变量的公共因子，从而使问题的分析简化和清晰。这种研究对象是变量的因子分析称为**R型**因子分析。

对有些问题来说，我们还需要研究样品的结构，若对于样品进行因子分析，称为**Q型**因子分析。

当我们对数据同时进行R和Q型因子分析，并分别保留两个公共因子，则是对应分析的初步。

§ 1 对应分析基本思想

- 对应分析综合了R型和Q型分析的优点，将两者统一起来；更重要的是可以把变量和样品的载荷反映在**相同的公因子轴**上。这就把变量和样品联系起来，便于解释和推断。
- 对应分析的基本思想是将一个列联表的行和列中各元素的**比例结构以点的形式在低维**空间表示出来。它最大特点是能把众多的样品和众多的变量同时作到一张图上，直观展示。

§ 2 对应分析的数学原理

由于R型因子分析和Q型因子分析是反映一个整体的不同侧面，R型因子分析是从列来讨论（对变量），Q型因子分析是从行来讨论（对样品），因此他们之间存在内在的联系。

设原始数据矩阵为：

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n \times p}$$

由于因子分析都是基于协方差矩阵或相关系数矩阵完成的，所以必须从变量和样品的协方差矩阵入手来进行分析。

$$X^* = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{np} - \bar{x}_p \end{bmatrix}_{n \times p}$$

$$X^{*'} X^*$$

$$= \begin{pmatrix} x_{11} - \bar{x}_1 & x_{21} - \bar{x}_1 & \cdots & x_{n1} - \bar{x}_1 \\ x_{12} - \bar{x}_2 & x_{22} - x_2 & \cdots & x_{n2} - x_2 \\ \vdots & \vdots & & \vdots \\ x_{1p} - \bar{x}_p & x_{2p} - \bar{x}_p & \cdots & x_{np} - \bar{x}_p \end{pmatrix} \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - x_2 & \cdots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - x_2 & \cdots & x_{np} - \bar{x}_p \end{pmatrix}$$

变量的积叉矩阵

$$\Sigma_R = (X^*)' X^* \quad (p \times p)$$

样品的积叉矩阵

$$\Sigma_Q = X^* (X^*)' \quad (n \times n)$$

显然，变量和样品的积叉矩阵的阶数不同，一般来说，他们的非零特征根也不一样，那么能否将观测值做变换。

$$X \rightarrow Z$$

$Z'Z$ 和 ZZ' 具有相同的特征根。

(一) 规格化矩阵

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n \times p}$$

$x_{i.}$ 为行和, $x_{.j}$ 为列和

$x_{..}$ 为总和

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{matrix} x_{1.} \\ x_{2.} \\ \vdots \\ x_{n.} \end{matrix}$$

$$\begin{matrix} x_{.1} & x_{.2} & \cdots & x_{.p} & x_{..} \end{matrix}$$

$$p_{ij} = x_{ij} / x_{..}$$

$$X = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1p} \\ p_{21} & p_{22} & \cdots & p_{2p} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{np} \end{bmatrix}_{n \times p}$$

我们可以把 p_{ij} 解释成概率，因为所有的元素之和为1。

$$\text{行和: } p_{i.} = \sum_{j=1}^p p_{ij}$$

$$\text{列和: } p_{.j} = \sum_{i=1}^n p_{ij}$$

$$\therefore \frac{p_{ij}}{p_{i.}} = \frac{x_{ij} / x_{..}}{\sum_{j=1}^p p_{ij}} = \frac{x_{ij} / x_{..}}{\sum_{j=1}^p x_{ij} / x_{..}} = \frac{x_{ij}}{x_{i.}}$$

$$\therefore \begin{bmatrix} \frac{p_{i1}}{p_{i.}} & \frac{p_{i2}}{p_{i.}} & \dots & \frac{p_{ip}}{p_{i.}} \end{bmatrix} = \begin{bmatrix} \frac{x_{i1}}{x_{i.}} & \frac{x_{i2}}{x_{i.}} & \dots & \frac{x_{ip}}{x_{i.}} \end{bmatrix} i = 1, 2, 3, \dots, n$$

称为行形象，行轮廓(row profile)

例如：考察某一文章中各种词汇出现的次数，词汇分为如下种类：

n=名词， v=动词， a=形容词， av=副词，

l=冠词， o=其它

x_{ij} 表示在第i 篇文章中属于j 种词汇的次数。

	n	v	a	av	l	o	
i'	80	30	25	15	60	55	$x_{i'.}=265$
i''	160	60	50	30	120	110	$x_{i''.}=530$

它们的行形象：

i'	0.30	0.11	0.09	0.06	0.23	0.21
i''	0.30	0.11	0.09	0.06	0.23	0.21

这两行的形象相同，由此可以断定这两文的用词手法相同。

$$N(R) = \begin{bmatrix} \frac{p_{11}}{p_{1.}} & \frac{p_{12}}{p_{1.}} & \dots & \frac{p_{1p}}{p_{1.}} \\ \frac{p_{21}}{p_{2.}} & \frac{p_{22}}{p_{2.}} & \dots & \frac{p_{2p}}{p_{2.}} \\ \vdots & \vdots & & \vdots \\ \frac{p_{n1}}{p_{n.}} & \frac{p_{n2}}{p_{n.}} & \dots & \frac{p_{np}}{p_{n.}} \end{bmatrix}$$

行形象点集

第j个列变量的期望:

$$E\left(\frac{p_{ij}}{p_{i.}}\right) = \sum_{i=1}^n \frac{p_{ij}}{p_{i.}} \times p_{i.} = p_{.j}, j = 1, 2, \dots, p$$

因为原始变量的数量等级可能不同，所以为了尽量减少**各变量尺度差异**，将行形象中的(各列元素)都除以其期望的平方根。得矩阵D(R)

$$D(R) = \begin{bmatrix} \frac{p_{11}}{p_{1.}\sqrt{p_{.1}}} & \frac{p_{12}}{p_{1.}\sqrt{p_{.2}}} & \dots & \frac{p_{1p}}{p_{1.}\sqrt{p_{.p}}} \\ \frac{p_{21}}{p_{2.}\sqrt{p_{.1}}} & \frac{p_{22}}{p_{2.}\sqrt{p_{.2}}} & \dots & \frac{p_{2p}}{p_{2.}\sqrt{p_{.p}}} \\ \vdots & \vdots & & \vdots \\ \frac{p_{n1}}{p_{n.}\sqrt{p_{.1}}} & \frac{p_{n2}}{p_{n.}\sqrt{p_{.2}}} & \dots & \frac{p_{np}}{p_{n.}\sqrt{p_{.p}}} \end{bmatrix}$$

第j个列变量的**期望**:

$$E\left(\frac{p_{ij}}{p_{i.}\sqrt{p_{.j}}}\right) = \sum_{i=1}^n \frac{p_{ij}}{p_{i.}\sqrt{p_{.j}}} \times p_{i.} = \frac{1}{\sqrt{p_{.j}}} p_{.j} = \sqrt{p_{.j}}, j = 1, 2, \dots, p$$

第k个变量与第j个变量的协方差：

$$\begin{aligned}
 s_{kj} &= \sum_{a=1}^n \left[\frac{p_{ak}}{p_{a.} \sqrt{p_{.k}}} - \sqrt{p_{.k}} \right] \left[\frac{p_{aj}}{p_{a.} \sqrt{p_{.j}}} - \sqrt{p_{.j}} \right] \times p_{a.} \\
 &= \sum_{a=1}^n \left[\frac{p_{ak}}{\sqrt{p_{a.}} \sqrt{p_{.k}}} - \sqrt{p_{.k}} \sqrt{p_{a.}} \right] \left[\frac{p_{aj}}{\sqrt{p_{a.}} \sqrt{p_{.j}}} - \sqrt{p_{.j}} \sqrt{p_{a.}} \right] \\
 &= \sum_{a=1}^n \left[\frac{p_{ak} - p_{a.} p_{.k}}{\sqrt{p_{a.}} \sqrt{p_{.k}}} \right] \left[\frac{p_{aj} - p_{a.} p_{.j}}{\sqrt{p_{a.}} \sqrt{p_{.j}}} \right] \\
 &= \sum_{a=1}^n z_{ak} z_{aj}
 \end{aligned}$$

$$z_{ak} = \frac{p_{ak} - p_{a.}p_{.k}}{\sqrt{p_{a.}p_{.k}}} = \frac{x_{ak} - x_{a.}x_{.k}}{\sqrt{x_{a.}x_{.k}}}$$

令 \mathbf{Z} 为 z_{ij} 所组成的矩阵，则 $\mathbf{A} = \mathbf{Z}'\mathbf{Z}$

称
$$\begin{bmatrix} \frac{p_{1j}}{p_{.j}} & \frac{p_{2j}}{p_{.j}} & \dots & \frac{p_{nj}}{p_{.j}} \end{bmatrix} = \begin{bmatrix} \frac{x_{1j}}{x_{.j}} & \frac{x_{2j}}{x_{.j}} & \dots & \frac{x_{nj}}{x_{.j}} \end{bmatrix} j = 1, 2, 3, \dots, p$$

为列形象。

$$N(Q) = \begin{bmatrix} \frac{p_{11}}{p_{.1}} & \frac{p_{12}}{p_{.2}} & \dots & \frac{p_{1p}}{p_{.p}} \\ \frac{p_{21}}{p_{.1}} & \frac{p_{22}}{p_{.2}} & \dots & \frac{p_{2p}}{p_{.p}} \\ \vdots & \vdots & & \vdots \\ \frac{p_{n1}}{p_{.1}} & \frac{p_{n2}}{p_{.2}} & \dots & \frac{p_{np}}{p_{.p}} \end{bmatrix}$$

第*i*个行变量的期望：

$$E\left(\frac{p_{ij}}{p_{.j}}\right) = \sum_{j=1}^p \frac{p_{ij}}{p_{.j}} \cdot p_{.j} = p_{i.}$$

因为原始变量的数量等级可能不同，所以为了尽量减少各变量尺度差异，将列形象中的各行元素均除以其期望的平方根。得矩阵 $D(Q)$

$$D(Q) = \begin{bmatrix} \frac{p_{11}}{p_{.1}\sqrt{p_{1.}}} & \frac{p_{12}}{p_{.2}\sqrt{p_{1.}}} & \dots & \frac{p_{1p}}{p_{.p}\sqrt{p_{1.}}} \\ \frac{p_{21}}{p_{.1}\sqrt{p_{2.}}} & \frac{p_{22}}{p_{.2}\sqrt{p_{2.}}} & \dots & \frac{p_{2p}}{p_{.p}\sqrt{p_{2.}}} \\ \vdots & \vdots & & \vdots \\ \frac{p_{n1}}{p_{.1}\sqrt{p_{n.}}} & \frac{p_{n2}}{p_{.2}\sqrt{p_{n.}}} & \dots & \frac{p_{np}}{p_{.p}\sqrt{p_{n.}}} \end{bmatrix}$$

第 i 个行变量的期望：

$$E\left(\frac{p_{ij}}{p_{.j}\sqrt{p_{i.}}}\right) = \sum_{j=1}^p \frac{p_{ij}}{p_{.j}\sqrt{p_{i.}}} \cdot p_{.j} = \sqrt{p_{i.}}$$

第k个样品与第 l 个样品的协方差：

$$\begin{aligned}
 b_{kl} &= \sum_{i=1}^p \left[\frac{p_{ki}}{p_{.i} \sqrt{p_{k.}}} - \sqrt{p_{k.}} \right] \left[\frac{p_{li}}{p_{.i} \sqrt{p_{l.}}} - \sqrt{p_{l.}} \right] \times p_{.i} \\
 &= \sum_{i=1}^p \left[\frac{p_{ki}}{\sqrt{p_{.i}} \sqrt{p_{k.}}} - \sqrt{p_{.i}} \sqrt{p_{k.}} \right] \left[\frac{p_{li}}{\sqrt{p_{.i}} \sqrt{p_{l.}}} - \sqrt{p_{.i}} \sqrt{p_{l.}} \right] \\
 &= \sum_{i=1}^p \left[\frac{p_{ki} - p_{k.} p_{.i}}{\sqrt{p_{.i}} \sqrt{p_{k.}}} \right] \left[\frac{p_{li} - p_{.i} p_{l.}}{\sqrt{p_{l.}} \sqrt{p_{.i}}} \right] \\
 &= \sum_{i=1}^p z_{ki} z_{li}
 \end{aligned}$$

令 \mathbf{Z} 为 z_{ij} 所组成的矩阵，则 $\mathbf{B} = \mathbf{Z}\mathbf{Z}'$

因此将矩阵变换成矩阵 \mathbf{Z} ，则很容易求出 \mathbf{A} 和 \mathbf{B} 存在着的简单对应关系。由特征根和特征向量的性质， **\mathbf{A} 和 \mathbf{B} 有相同的非零特征根。**

设 λ_k 是 $\mathbf{A} = \mathbf{Z}'\mathbf{Z}$ 的非零特征根，则 $\mathbf{Z}'\mathbf{Z}\mathbf{u}_k = \lambda_k \mathbf{u}_k$

在上式的两边都左乘 \mathbf{Z} ，则

$$\mathbf{Z}\mathbf{Z}'(\mathbf{Z}\mathbf{u}_k) = \lambda_k (\mathbf{Z}\mathbf{u}_k)$$

可见 λ_k 也是 $\mathbf{Z}\mathbf{Z}'$ 的特征根，**相应的特征向量是 $\mathbf{Z}\mathbf{u}_k$**

因此将原始数据矩阵 X 变换成矩阵 Z ，则变量和样品的协差阵分别可表示为 $A = Z'Z$ 和 $B = ZZ'$ ， A 和 B 具有相同的非零特征值，相应的特征向量有很密切的关系。

这样就可以用相同的因子轴去同时表示变量和样品，把变量和样品同时反映在具有相同坐标轴的因子平面上。

§ 3 对应分析图

设 $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_i (0 < i < \min(n, p))$ 为矩阵 \mathbf{A} 和 \mathbf{B} 的非零特征根,其相应的特征向量为

$$\mathbf{u}_1 = [u_{11} \quad u_{21} \quad \cdots \quad u_{p1}]'$$

$$\mathbf{u}_2 = [u_{12} \quad u_{22} \quad \cdots \quad u_{p2}]'$$

$$\mathbf{v}_1 = [v_{11} \quad v_{21} \quad \cdots \quad v_{n1}]'$$

$$\mathbf{v}_2 = [v_{12} \quad v_{22} \quad \cdots \quad v_{n2}]'$$

我们知道因子载荷矩阵的含义是原始变量与公共因子之间的相关系数，所以如果我们构造一个平面直角坐标系，将第一公共因子的载荷与第二个公共因子的载荷看成平面上的点，在坐标系中绘制散点图，则构成对应图。

对应分析的步骤

1、获取对应分析数据

首先要规定研究的目的，然后选择对应分析中所需数据，应该包括的背景资料。

2、建立列联表

3、对应分析

4、对应图并解释结果的意义。

例 交叉列联表（表1）总结了260个消费者对于四种不同软件的性能评价，假设B软件是公司自己的产品，我们想了解消费者对B软件的评价如何？与其他竞争对手的产品形象有何不同？

表1 消费者对四种软件的性能评价

<div>软件性能 软件名称</div>	易学	操作简单	运行速度 快	可视化	算法丰富	界面友好	扩展能力 强
A软件	140	120	130	100	140	110	130
B软件	160	150	180	180	160	160	160
C软件	170	180	110	115	120	140	100
D软件	100	150	200	150	180	120	170

Contingency Table

	易学	操作简单	运行速度快	可视化	算法丰富	界面友好	扩展能力强	Sum
A软件	140	120	130	100	140	110	130	870
B软件	160	150	180	180	160	160	160	1150
C软件	170	180	110	115	120	140	100	935
D软件	100	150	200	150	180	120	170	1070
Sum	570	600	620	545	600	530	560	4025

将表1中数据除以总和4025，得到对应矩阵，见表2。

表2 消费者对四种软件性能评价的对应矩阵

软件性能 软件名称	易学	操作简单	运行速度快	可视化	算法丰富	界面友好	扩展能力强	$p_{i\bullet}$
A软件	0.0348	0.0298	0.0323	0.0248	0.0348	0.0273	0.0323	0.2161
B软件	0.0398	0.0373	0.0447	0.0447	0.0398	0.0398	0.0398	0.2857
C软件	0.0422	0.0447	0.0273	0.0286	0.0298	0.0348	0.0248	0.2323
D软件	0.0248	0.0373	0.0497	0.0373	0.0447	0.0298	0.0422	0.2658
$p_{\bullet j}$	0.1416	0.1491	0.1540	0.1354	0.1491	0.1317	0.1391	1

$$E(n_{ij}) = np_{ij} = n \times p_{i\bullet} \times p_{\bullet j}$$

Contingency Table

	易学	操作简单	运行速度快	可视化	算法丰富	界面友好	扩展能力强	Sum
A软件	140	120	130	100	140	110	130	870
B软件	160	150	180	180	160	160	160	1150
C软件	170	180	110	115	120	140	100	935
D软件	100	150	200	150	180	120	170	1070
Sum	570	600	620	545	600	530	560	4025

Chi-Square Statistic Expected Values

	易学	操作简单	运行速度快	可视化	算法丰富	界面友好	扩展能力强
A软件	123.205	129.689	134.012	117.801	129.689	114.559	121.043
B软件	162.857	171.429	177.143	155.714	171.429	151.429	160.000
C软件	132.410	139.379	144.025	126.602	139.379	123.118	130.087
D软件	151.528	159.503	164.820	144.882	159.503	140.894	148.870

Observed Minus Expected Values							
	易学	操作简单	运行速度快	可视化	算法丰富	界面友好	扩展能力强
A软件	16.7950	-9.6894	-4.0124	-17.8012	10.3106	-4.5590	8.9565
B软件	-2.8571	-21.4286	2.8571	24.2857	-11.4286	8.5714	0.0000
C软件	37.5901	40.6211	-34.0248	-11.6025	-19.3789	16.8820	-30.0870
D软件	-51.5280	-9.5031	35.1801	5.1180	20.4969	-20.8944	21.1304

Contributions to the Total Chi-Square Statistic								
	易学	操作简单	运行速度快	可视化	算法丰富	界面友好	扩展能力强	Sum
A软件	2.2895	0.7239	0.1201	2.6900	0.8197	0.1814	0.6627	7.4874
B软件	0.0501	2.6786	0.0461	3.7877	0.7619	0.4852	0.0000	7.8095
C软件	10.6715	11.8388	8.0381	1.0633	2.6944	2.3149	6.9586	43.5796
D软件	17.5224	0.5662	7.5091	0.1808	2.6339	3.0986	2.9992	34.5102
Sum	30.5335	15.8075	15.7134	7.7218	6.9100	6.0801	10.6206	93.3867

将表1中数据除以总和4025，得到对应矩阵，见表2。

表2 消费者对四种软件性能评价的对应矩阵

软件性能 软件名称	易学	操作简单	运行速度快	可视化	算法丰富	界面友好	扩展能力强
A软件	0.0348	0.0298	0.0323	0.0248	0.0348	0.0273	0.0323
B软件	0.0398	0.0373	0.0447	0.0447	0.0398	0.0398	0.0398
C软件	0.0422	0.0447	0.0273	0.0286	0.0298	0.0348	0.0248
D软件	0.0248	0.0373	0.0497	0.0373	0.0447	0.0298	0.0422

计算行轮廓矩阵：

$$N(R) = \begin{bmatrix} 0.1609 & 0.1379 & 0.1494 & 0.1149 & 0.1609 & 0.1264 & 0.1494 \\ 0.1391 & 0.1304 & 0.1565 & 0.1565 & 0.1391 & 0.1391 & 0.1391 \\ 0.1818 & 0.1925 & 0.1176 & 0.1230 & 0.1283 & 0.1497 & 0.1070 \\ 0.0935 & 0.1402 & 0.1869 & 0.1402 & 0.1682 & 0.1121 & 0.1589 \end{bmatrix}$$

将表1中数据除以总和4025，得到对应矩阵，见表2。

表2 消费者对四种软件性能评价的对应矩阵

软件性能 软件名称	易学	操作简单	运行速度快	可视化	算法丰富	界面友好	扩展能力强
A软件	0.0348	0.0298	0.0323	0.0248	0.0348	0.0273	0.0323
B软件	0.0398	0.0373	0.0447	0.0447	0.0398	0.0398	0.0398
C软件	0.0422	0.0447	0.0273	0.0286	0.0298	0.0348	0.0248
D软件	0.0248	0.0373	0.0497	0.0373	0.0447	0.0298	0.0422

计算列轮廓矩阵：

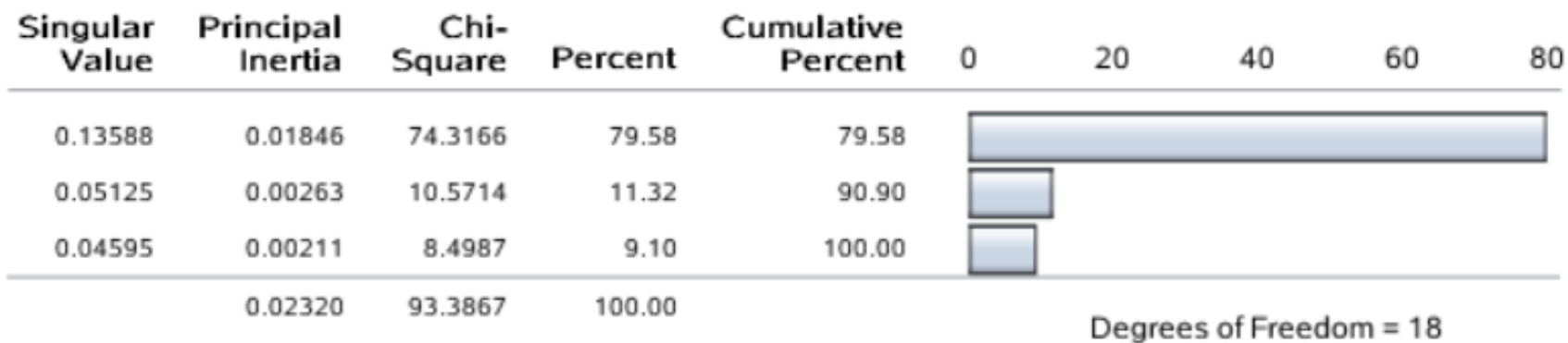
$$N(Q)=\begin{bmatrix} 0.2456 & 0.2000 & 0.2097 & 0.1835 & 0.2333 & 0.2075 & 0.2321 \\ 0.2807 & 0.2500 & 0.2903 & 0.3303 & 0.2667 & 0.3019 & 0.2857 \\ 0.2982 & 0.3000 & 0.1774 & 0.2110 & 0.2000 & 0.2642 & 0.1786 \\ 0.1754 & 0.2500 & 0.3226 & 0.2752 & 0.3000 & 0.2264 & 0.3036 \end{bmatrix}$$

计算奇异值、主惯量以及贡献率

表3 奇异值、主惯量以及贡献率

	奇异值	主惯量	卡方	贡献率	累计贡献率
1	0.13588	0.01846	74.3166	79.58	79.58
2	0.05125	0.00263	10.5714	11.32	90.90
3	0.04595	0.00211	8.4987	9.10	100.00
		0.02320	93.3867		

Inertia and Chi-Square Decomposition



计算行变量和列变量前两维的坐标矩阵

表4 行坐标（Row Coordinates）

软件名称	Dim1	Dim2
A软件	0.0132	0.0637
B软件	-0.0215	-0.075
C软件	0.2115	0.0076
D软件	-0.1724	0.0222

表5 列坐标（Column Coordinates）

软件性能	Dim1	Dim2
易学	0.2210	0.0146
操作简单	0.1296	0.0354
运行速度快	-0.1588	0.0016
可视化	-0.0553	-0.1049
算法丰富	-0.0889	0.0592
界面友好	0.0962	-0.0467

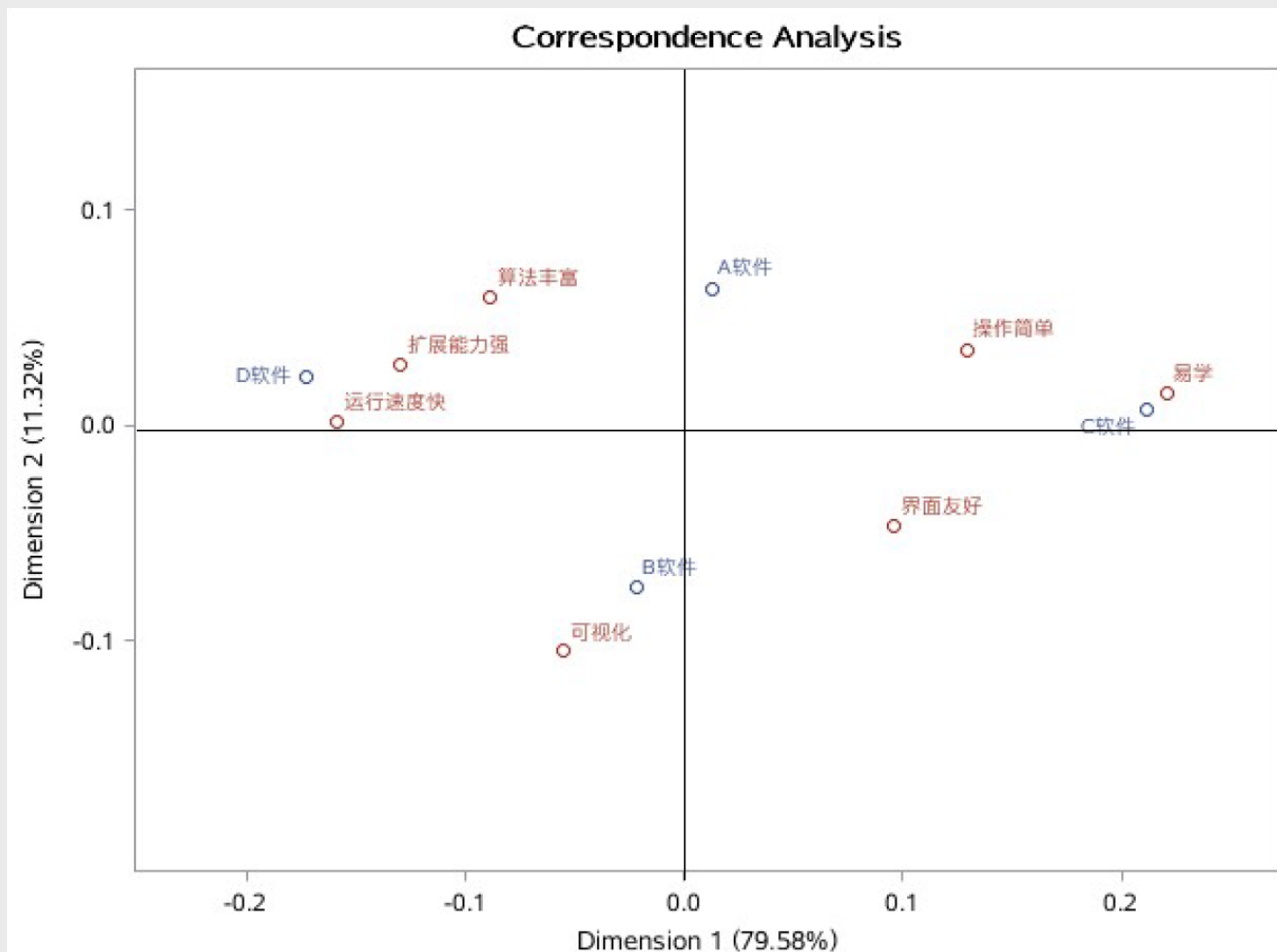


图1 对应分析图

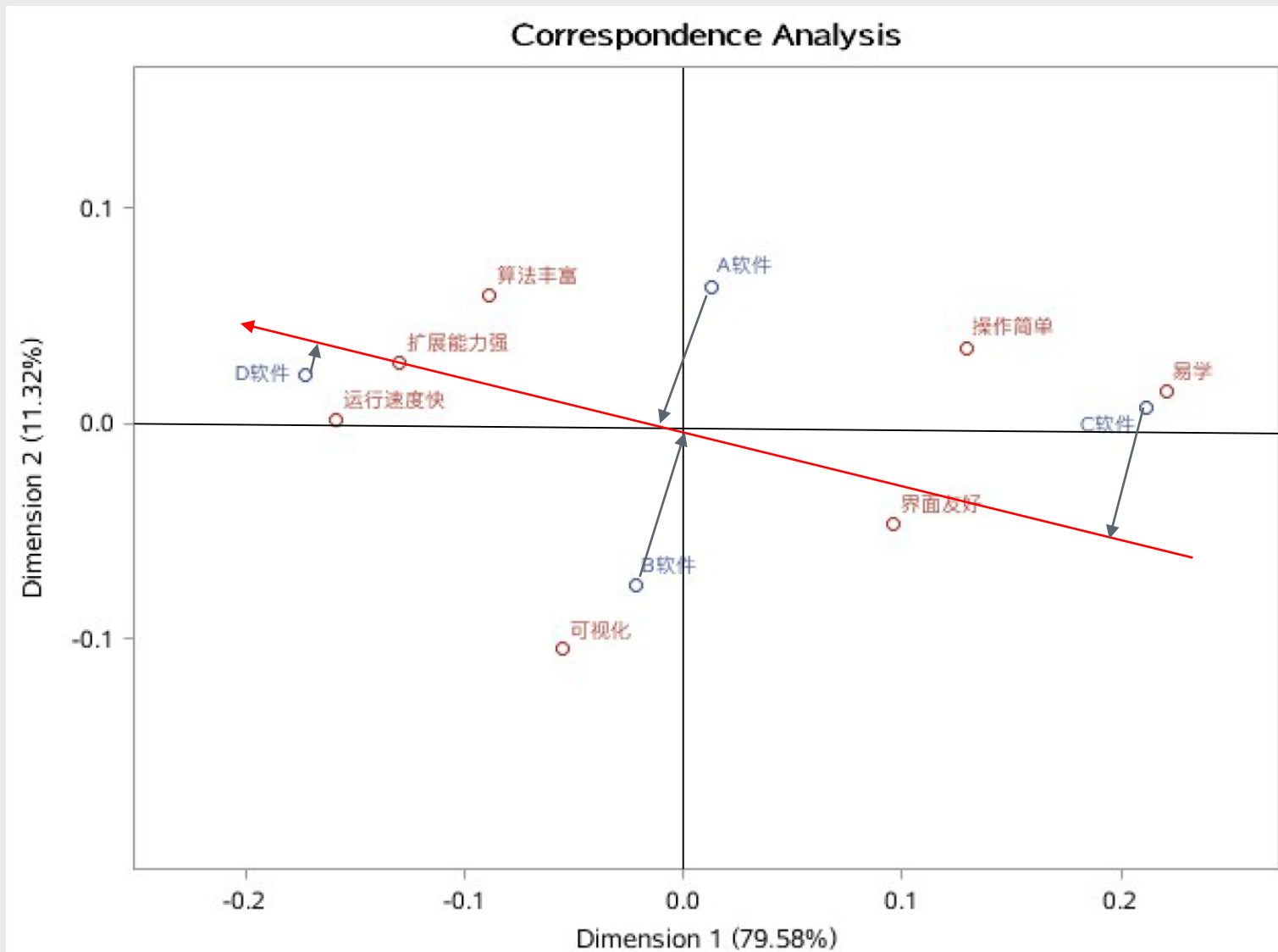


图1 对应分析图

对应分析图解读

1.总体观察

2.观察邻近区域

3.向量分析——偏好排序

从中心向任意点连线--向量，例如从中心向“扩展能力强”作向量，然后让所有“软件”往这条向量及延长线上作垂线，垂点越靠近向量正向的表示扩展能力越强。

4.向量的夹角——余弦定理

从向量夹角的角度看不同“软件”或不同“性能”间的相似情况，从余弦定理的角度看相似性！

5.从距离中的位置看：越靠近中心，越没有特征，越远离中心，说明特征越明显。各个类别之间的距离表示相对密切关系

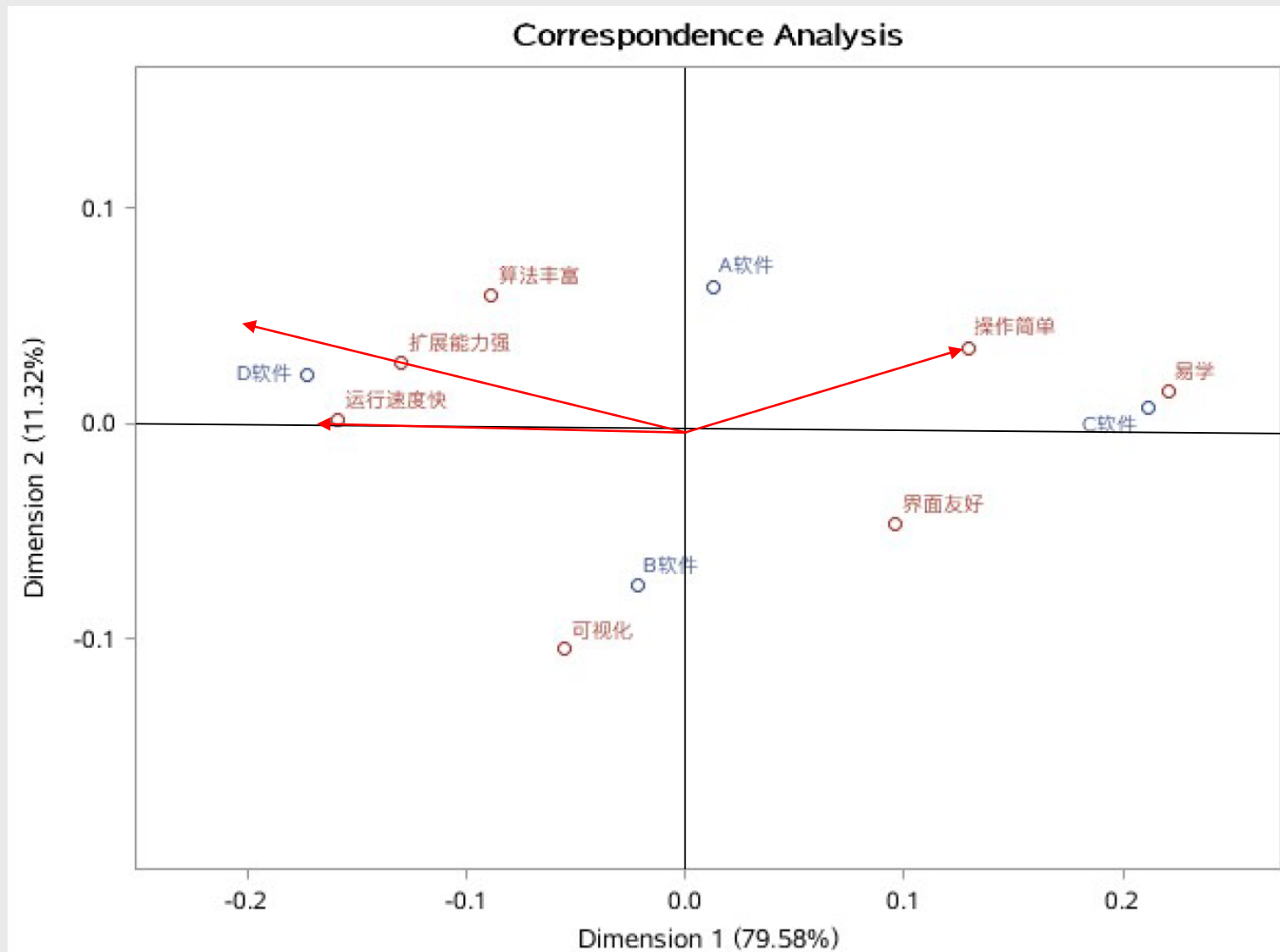


图1 对应分析图

案例分析：车主特征与车辆信息的对应分析

聚焦用户特征，开展精准营销能将营销的力度发挥到最大程度。不同的车型可能受到不同类型车主的青睐。一项著名的调查提供了车主特征（如收入、婚姻状况、性别等）和车辆基本信息（产地、型号、用途等）。合理分析这些信息，挖掘数据背后的联系，将为市场研究部门确定广告的宣传对象及营销策略提供重要依据。

表1 指标数据（文件名：car）

	American	European	Japanese	大型	中型	小型	家用	跑车	代步
低收入	58	18	74	20	57	73	69	55	26
高收入	67	26	91	22	84	78	105	51	28
已婚无孩	37	13	51	9	42	50	50	35	16
已婚有孩	50	15	44	21	51	37	79	12	18
单身	32	15	62	11	40	58	35	57	17
单身有孩	6	1	8	1	8	6	10	2	3
女	58	21	70	17	70	62	83	40	26
男	67	23	95	25	71	89	91	56	28

SAS 程序

```
proc corresp data=car out=ao rp cp all;
```

```
var x1-x9;
```

```
id usercha;
```

```
run;
```

SAS 输出

输出1： 计算期望频数

卡方统计量期望值 Chi-Square Statistic Expected Values									
	American	European	Japanese	大型	中型	小型	家用	跑车	代步
低收入	56.3251	19.8264	74.3491	18.9252	63.5347	68.0407	78.4045	46.2617	24.3324
高收入	69.0921	24.3204	91.2016	23.2150	77.9359	83.4633	96.1762	56.7477	29.8478
已婚无孩	37.9256	13.3498	50.0617	12.7430	42.7800	45.8141	52.7924	31.1495	16.3838
已婚有孩	40.9296	14.4072	54.0270	13.7523	46.1686	49.4429	56.9740	33.6168	17.6816
单身	40.9296	14.4072	54.0270	13.7523	46.1686	49.4429	56.9740	33.6168	17.6816
单身有孩	5.6325	1.9826	7.4349	1.8925	6.3535	6.8041	7.8405	4.6262	2.4332
女	55.9496	19.6943	73.8535	18.7991	63.1111	67.5871	77.8818	45.9533	24.1702
男	68.2160	24.0120	90.0451	22.9206	76.9476	82.4049	94.9566	56.0280	29.4693

SAS 输出

输出2：实际频数减期望频数

观测值减期望值 Observed Minus Expected Values									
	American	European	Japanese	大型	中型	小型	家用	跑车	代步
低收入	1.6749	-1.8264	-0.3491	1.0748	-6.5347	4.9593	-9.4045	8.7383	1.6676
高收入	-2.0921	1.6796	-0.2016	-1.2150	6.0641	-5.4633	8.8238	-5.7477	-1.8478
已婚无孩	-0.9256	-0.3498	0.9383	-3.7430	-0.7800	4.1859	-2.7924	3.8505	-0.3838
已婚有孩	9.0704	0.5928	-10.0270	7.2477	4.8314	-12.4429	22.0260	-21.6168	0.3184
单身	-8.9296	0.5928	7.9730	-2.7523	-6.1686	8.5571	-21.9740	23.3832	-0.6816
单身有孩	0.3675	-0.9826	0.5651	-0.8925	1.6465	-0.8041	2.1595	-2.6262	0.5668
女	2.0504	1.3057	-3.8535	-1.7991	6.8889	-5.5871	5.1182	-5.9533	1.8298
男	-1.2160	-1.0120	4.9549	2.0794	-5.9476	6.5951	-3.9566	-0.0280	-1.4693

SAS 输出

输出3： 计算卡方统计量

对卡方统计量的贡献 Contributions to the Total Chi-Square Statistic										
	American	European	Japanese	大型	中型	小型	家用	跑车	代步	Sum
低收入	0.0498	0.1683	0.0016	0.0610	0.6721	0.3615	1.1281	1.6506	0.1143	4.2072
高收入	0.0633	0.1160	0.0004	0.0636	0.4718	0.3576	0.8095	0.5821	0.1144	2.5789
已婚无孩	0.0226	0.0092	0.0176	1.0994	0.0142	0.3825	0.1477	0.4760	0.0090	2.1781
已婚有孩	2.0101	0.0244	1.8609	3.8196	0.5056	3.1314	8.5152	13.9004	0.0057	33.7734
单身	1.9482	0.0244	1.1766	0.5508	0.8242	1.4810	8.4750	16.2649	0.0263	30.7713
单身有孩	0.0240	0.4870	0.0429	0.4209	0.4267	0.0950	0.5948	1.4908	0.1320	3.7142
女	0.0751	0.0866	0.2011	0.1722	0.7519	0.4619	0.3363	0.7712	0.1385	2.9949
男	0.0217	0.0427	0.2727	0.1887	0.4597	0.5278	0.1649	0.0000	0.0733	1.7513
Sum	4.2148	0.9584	3.5739	6.3762	4.1263	6.7986	20.1716	35.1360	0.6135	81.9694

SAS 输出

输出4： 计算行轮廓系数

行轮廓Row Profiles									
	American	European	Japanese	大型	中型	小型	家用	跑车	代步
低收入	0.128889	0.040000	0.164444	0.044444	0.126667	0.162222	0.153333	0.122222	0.057778
高收入	0.121377	0.047101	0.164855	0.039855	0.152174	0.141304	0.190217	0.092391	0.050725
已婚无孩	0.122112	0.042904	0.168317	0.029703	0.138614	0.165017	0.165017	0.115512	0.052805
已婚有孩	0.152905	0.045872	0.134557	0.064220	0.155963	0.113150	0.241590	0.036697	0.055046
单身	0.097859	0.045872	0.189602	0.033639	0.122324	0.177370	0.107034	0.174312	0.051988
单身有孩	0.133333	0.022222	0.177778	0.022222	0.177778	0.133333	0.222222	0.044444	0.066667
女	0.129754	0.046980	0.156600	0.038031	0.156600	0.138702	0.185682	0.089485	0.058166
男	0.122936	0.042202	0.174312	0.045872	0.130275	0.163303	0.166972	0.102752	0.051376

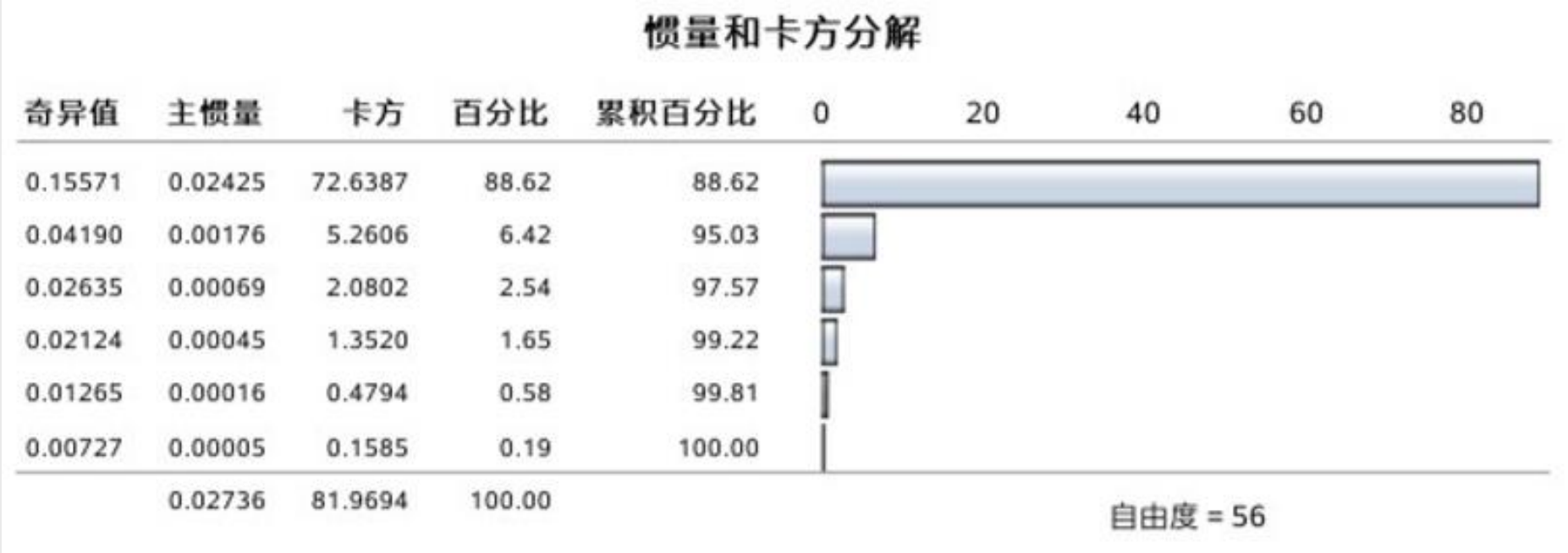
SAS 输出

输出5： 计算列轮廓系数

列轮廓 Column Profiles									
	American	European	Japanese	大型	中型	小型	家用	跑车	代步
低收入	0.154667	0.136364	0.149495	0.158730	0.134752	0.161148	0.132184	0.178571	0.160494
高收入	0.178667	0.196970	0.183838	0.174603	0.198582	0.172185	0.201149	0.165584	0.172840
已婚无孩	0.098667	0.098485	0.103030	0.071429	0.099291	0.110375	0.095785	0.113636	0.098765
已婚有孩	0.133333	0.113636	0.088889	0.166667	0.120567	0.081678	0.151341	0.038961	0.111111
单身	0.085333	0.113636	0.125253	0.087302	0.094563	0.128035	0.067050	0.185065	0.104938
单身有孩	0.016000	0.007576	0.016162	0.007937	0.018913	0.013245	0.019157	0.006494	0.018519
女	0.154667	0.159091	0.141414	0.134921	0.165485	0.136865	0.159004	0.129870	0.160494
男	0.178667	0.174242	0.191919	0.198413	0.167849	0.196468	0.174330	0.181818	0.172840

SAS 输出

输出6：奇异值、主惯量以及贡献率



SAS 输出

输出7： R型因子分析
中公因子载荷

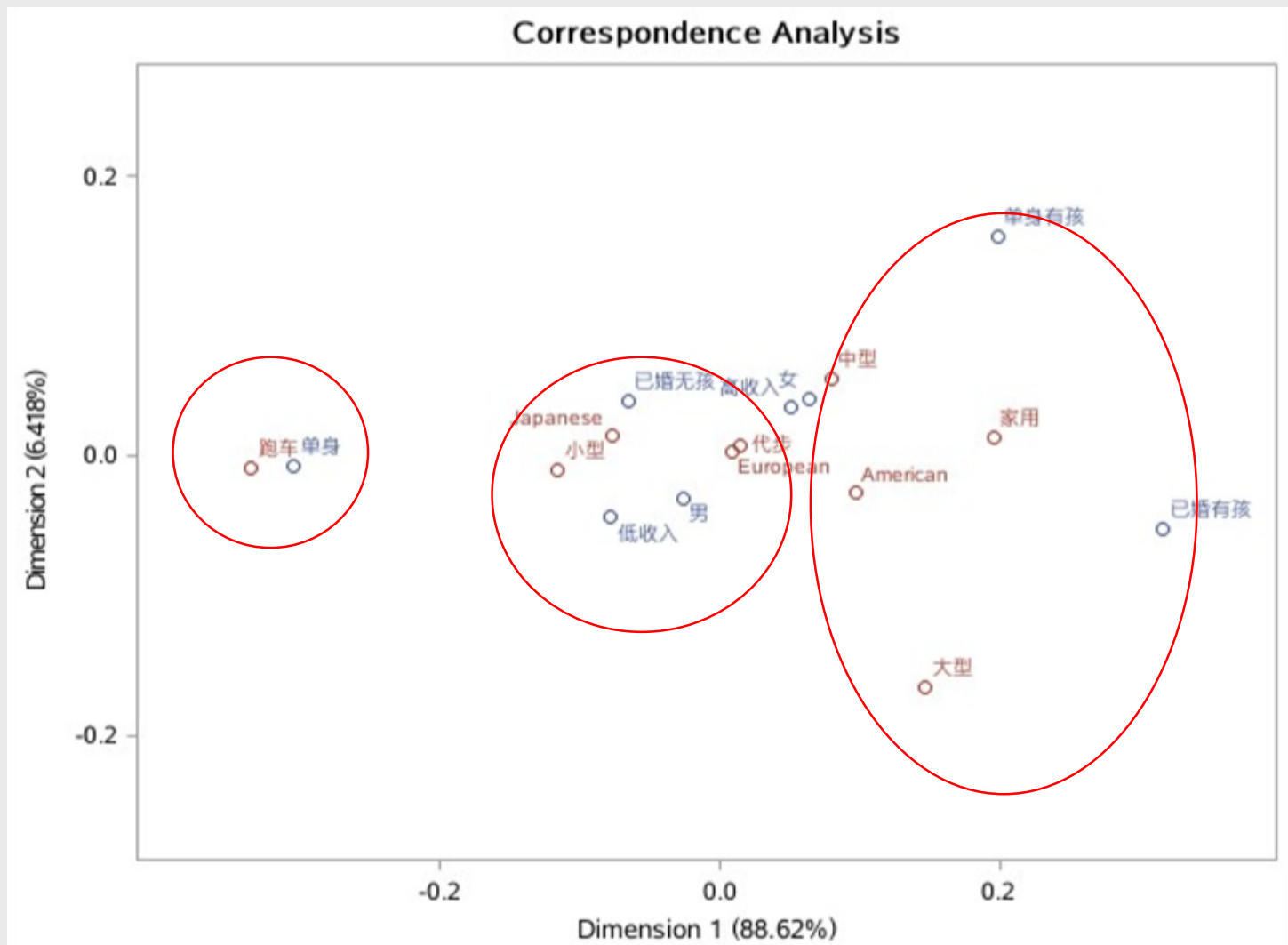
行坐标 Row Coordinates		
	Dim1	Dim2
低收入	-0.0786	-0.0438
高收入	0.0511	0.0345
已婚无孩	-0.0659	0.0393
已婚有孩	0.3165	-0.0527
单身	-0.3047	-0.0073
单身有孩	0.1982	0.1564
女	0.0641	0.0404
男	-0.0262	-0.0306

输出8： Q型因子分
析中公因子载荷

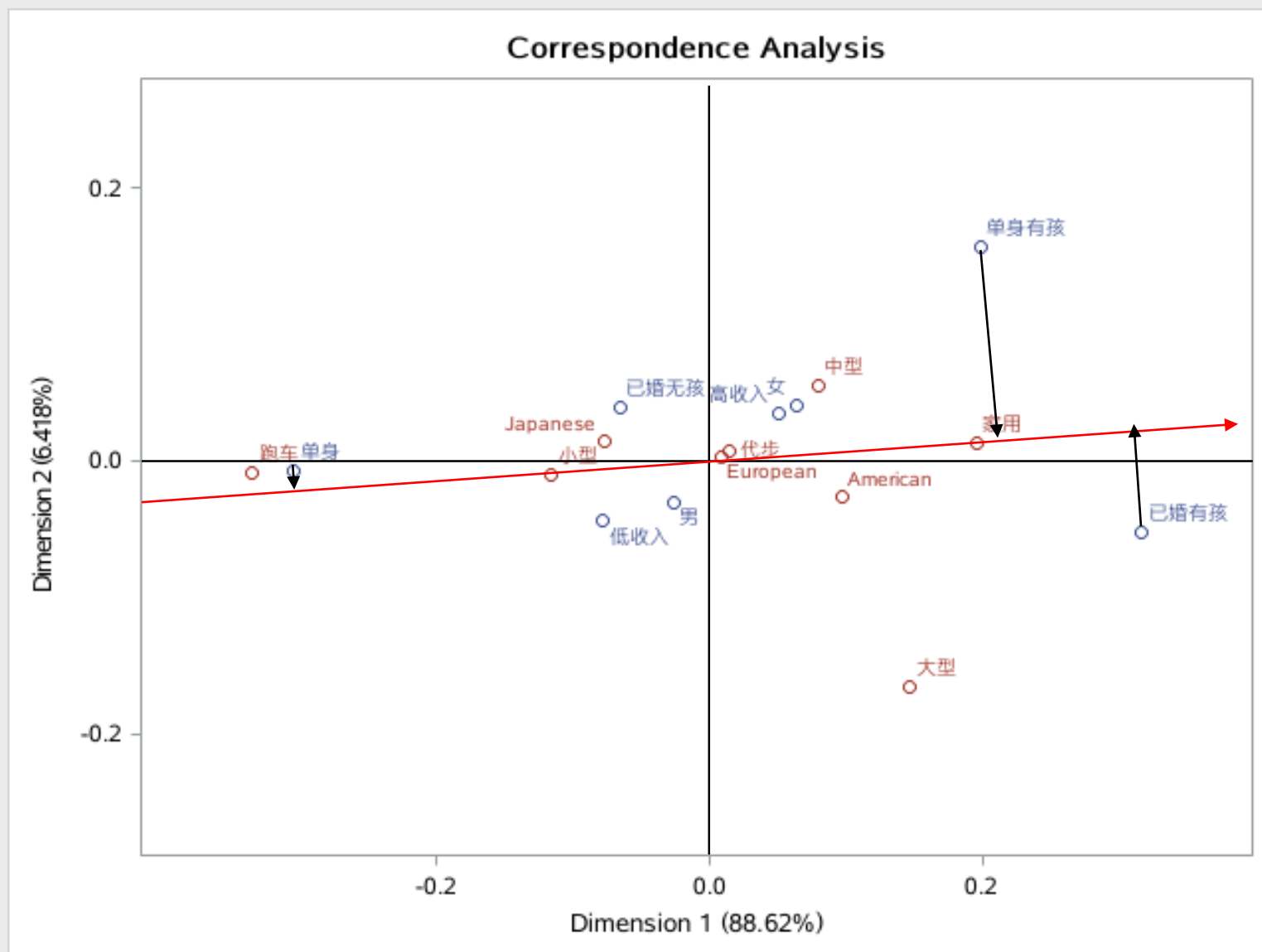
列坐标Column Coordinates		
	Dim1	Dim2
American	0.0968	-0.0266
European	0.0085	0.0034
Japanese	-0.0767	0.0143
大型	0.1471	-0.1656
中型	0.0791	0.0549
小型	-0.1160	-0.0106
家用	0.1956	0.0124
跑车	-0.3358	-0.0086
代步	0.0149	0.0065

SAS 输出

输出9: 对应分析图



结果分析



案例分析：居民收入来源的对应分析

随着我国经济发展，各地区居民的收入水平日益提高，人民生活质量逐渐提高。但我国各个省份的经济发展特点、产业结构和水平差异较大，居民的收入来源和结构有较大的差别。为探究各个地区的收入来源差异，对居民收入来源进行对应分析。

目前，我国民居收入的来源有四种，包括：工资性收入（X1）、经营净收入(X2)、财产净收入(X3)、转移净收入(X4)，因此选取这四种收入作为分析指标。

表1 指标数据（文件名：income）

地区	工资性收入	经营净收入	财产净收入	转移净收入
北京	35216.6	1408.3	9305.9	11299.0
天津	23165.0	3262.2	3504.9	7090.3
河北	13003.5	3210.8	1467.3	3802.6
山西	11957.1	2624.1	1227.9	4610.9
内蒙古	13899.7	6363.8	1287.6	4661.2
辽宁	14596.2	4881.9	1342.7	7014.6
吉林	10631.3	4712.7	898.7	5125.6
黑龙江	10318.8	4499.3	993.0	5394.8
上海	34365.4	1532.6	9030.1	14059.9
江苏	20399.2	4994.2	3238.6	6392.1
浙江	24137.3	7123.4	4741.6	6043.4
安徽	11920.9	4878.9	1227.7	3835.8
福建	17380.1	5600.1	2885.1	4182.4
江西	12553.1	3760.9	1397.0	4320.4
山东	15532.3	5892.6	1831.3	3673.8
河南	10108.1	4574.5	1237.3	4250.1
湖北	11830.6	5157.3	1501.9	5267.4
湖南	11836.6	4483.5	1626.8	5155.7
广东	23052.9	4420.9	3602.0	1927.5
广西	9819.3	5014.1	1171.5	3899.8
海南	13371.2	4285.7	1365.4	3530.9
重庆	12603.8	4016.7	1525.5	6007.0
四川	10013.6	4263.7	1362.9	4939.6
贵州	8642.8	3842.1	903.3	3315.5
云南	8468.1	4771.0	1850.9	3258.3
西藏	7839.7	4482.3	753.4	2381.9
陕西	11254.5	2629.5	1179.8	5571.4
甘肃	8798.4	2982.2	1043.7	3186.7
青海	11351.0	2861.2	949.0	3839.9
宁夏	12270.3	3628.2	819.8	3843.3
新疆	10907.2	4743.7	739.3	3584.9

SAS 程序

```
proc corresp data=income out=incmoe_out rp cp short;
```

```
var x1 x2 x3 x4;
```

```
id region;
```

```
run;
```

SAS 输出

输出1：行轮廓系数

Row Profiles				
	工资性收入	经营净收入	财产净收入	转移净收入
北京	0.615354	0.024608	0.162606	0.197432
天津	0.625702	0.088114	0.094670	0.191514
河北	0.605259	0.149449	0.068297	0.176995
山西	0.585558	0.128506	0.060132	0.225803
内蒙古	0.530274	0.242779	0.049122	0.177825
辽宁	0.524375	0.175385	0.048237	0.252003
吉林	0.497527	0.220546	0.042058	0.239869
黑龙江	0.486600	0.212172	0.046827	0.254401
上海	0.582583	0.025982	0.153084	0.238352
江苏	0.582433	0.142593	0.092468	0.182506
浙江	0.574073	0.169420	0.112773	0.143734
安徽	0.545247	0.223155	0.056153	0.175445
福建	0.578417	0.186374	0.096017	0.139192
江西	0.569782	0.170706	0.063409	0.196102
山东	0.576766	0.218812	0.068002	0.136420

Row Profiles				
	工资性收入	经营净收入	财产净收入	转移净收入
河南	0.501145	0.226797	0.061344	0.210714
湖北	0.497980	0.217084	0.063219	0.221718
湖南	0.512349	0.194069	0.070416	0.223165
广东	0.698503	0.133953	0.109141	0.058403
广西	0.493316	0.251905	0.058855	0.195924
海南	0.592874	0.190026	0.060541	0.156559
重庆	0.521832	0.166302	0.063160	0.248706
四川	0.486574	0.207179	0.066225	0.240022
贵州	0.517418	0.230015	0.054078	0.198489
云南	0.461520	0.260024	0.100876	0.177580
西藏	0.507184	0.289979	0.048741	0.154095
陕西	0.545403	0.127428	0.057174	0.269995
甘肃	0.549522	0.186259	0.065186	0.199032
青海	0.597386	0.150581	0.049944	0.202088
宁夏	0.596758	0.176455	0.039870	0.186916
新疆	0.546040	0.237481	0.037011	0.179468

SAS 输出

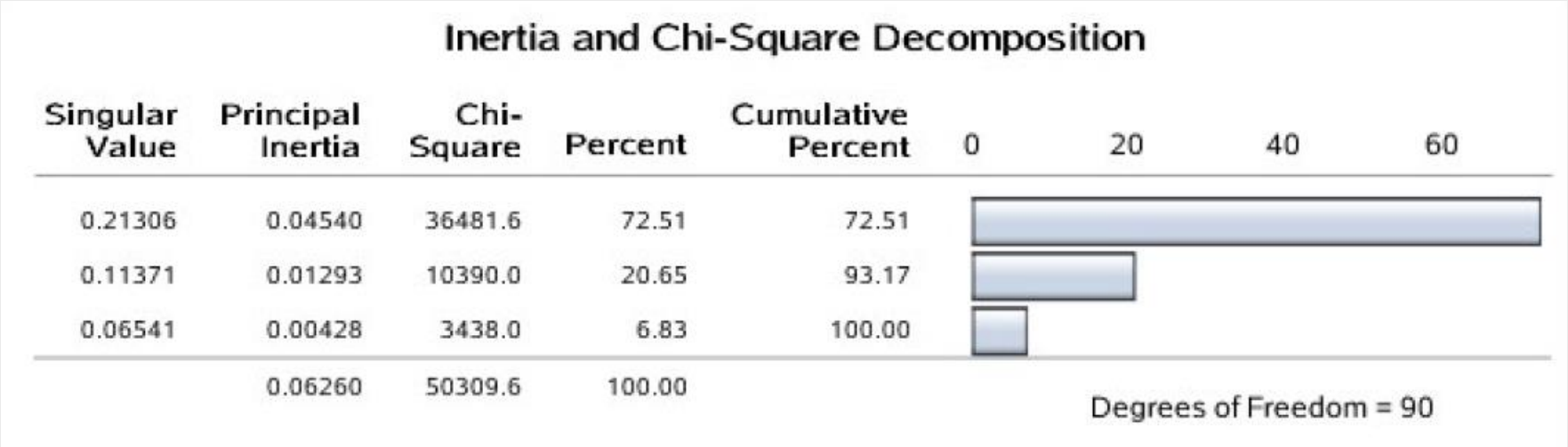
输出2：列轮廓系数

Column Profiles				
	工资性收入	经营净收入	财产净收入	转移净收入
北京	0.078043	0.010758	0.140973	0.072678
天津	0.051336	0.024921	0.053095	0.045607
河北	0.028817	0.024528	0.022228	0.024459
山西	0.026498	0.020046	0.018601	0.029658
内蒙古	0.030803	0.048615	0.019506	0.029982
辽宁	0.032347	0.037294	0.020340	0.045120
吉林	0.023560	0.036002	0.013614	0.032969
黑龙江	0.022867	0.034371	0.015043	0.034701
上海	0.076157	0.011708	0.136795	0.090437
江苏	0.045207	0.038152	0.049061	0.041116
浙江	0.053491	0.054418	0.071829	0.038873
安徽	0.026418	0.037271	0.018598	0.024673
福建	0.038516	0.042781	0.043706	0.026902
江西	0.027819	0.028731	0.021163	0.027790
山东	0.034421	0.045015	0.027742	0.023631

Column Profiles				
	工资性收入	经营净收入	财产净收入	转移净收入
河南	0.022400	0.034946	0.018744	0.027338
湖北	0.026218	0.039398	0.022752	0.033881
湖南	0.026231	0.034251	0.024644	0.033163
广东	0.051087	0.033772	0.054566	0.012398
广西	0.021760	0.038304	0.017747	0.025084
海南	0.029632	0.032740	0.020684	0.022712
重庆	0.027931	0.030685	0.023109	0.038638
四川	0.022191	0.032572	0.020646	0.031773
贵州	0.019153	0.029351	0.013684	0.021326
云南	0.018766	0.036447	0.028039	0.020958
西藏	0.017374	0.034242	0.011413	0.015321
陕西	0.024941	0.020087	0.017873	0.035837
甘肃	0.019498	0.022782	0.015811	0.020498
青海	0.025155	0.021858	0.014376	0.024699
宁夏	0.027192	0.027717	0.012419	0.024721
新疆	0.024171	0.036238	0.011199	0.023059

SAS 输出

输出3：主惯量和卡方分解



SAS 输出

输出4: 行坐标

Row Coordinates		
	Dim1	Dim2
北京	-0.4467	0.0021
天津	-0.1914	-0.0026
河北	-0.0120	-0.0373
山西	-0.0335	0.0875
内蒙古	0.2324	-0.0373
辽宁	0.0997	0.1522
吉林	0.2096	0.1203
黑龙江	0.1864	0.1561
上海	-0.4172	0.1061
江苏	-0.0661	-0.0281
浙江	-0.0487	-0.1308
安徽	0.1754	-0.0434
福建	0.0164	-0.1397
江西	0.0494	0.0100
山东	0.1358	-0.1429

Row Coordinates		
	Dim1	Dim2
河南	0.1844	0.0432
湖北	0.1620	0.0709
湖南	0.0981	0.0745
广东	-0.1457	-0.3404
广西	0.2418	0.0052
海南	0.0878	-0.0892
重庆	0.0535	0.1415
四川	0.1393	0.1165
贵州	0.2005	0.0140
云南	0.1861	-0.0499
西藏	0.3343	-0.0995
陕西	-0.0195	0.1983
甘肃	0.0827	0.0159
青海	0.0274	0.0291
宁夏	0.0990	-0.0083
新疆	0.2407	-0.0303

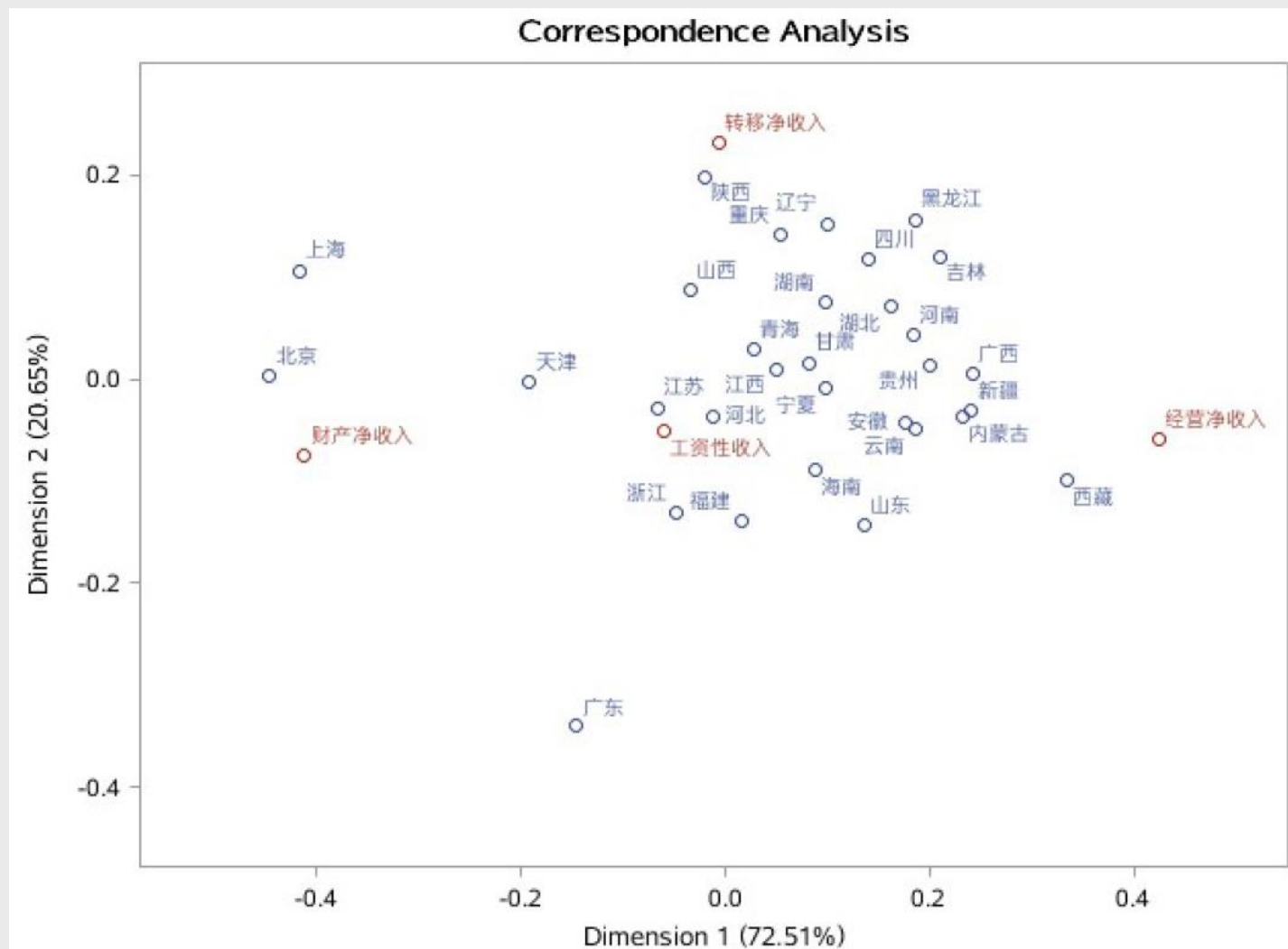
SAS 输出

输出5：列坐标

Column Coordinates		
	Dim1	Dim2
工资性收入	-0.0609	-0.0517
经营净收入	0.4242	-0.0591
财产净收入	-0.4127	-0.0755
转移净收入	-0.0053	0.2318

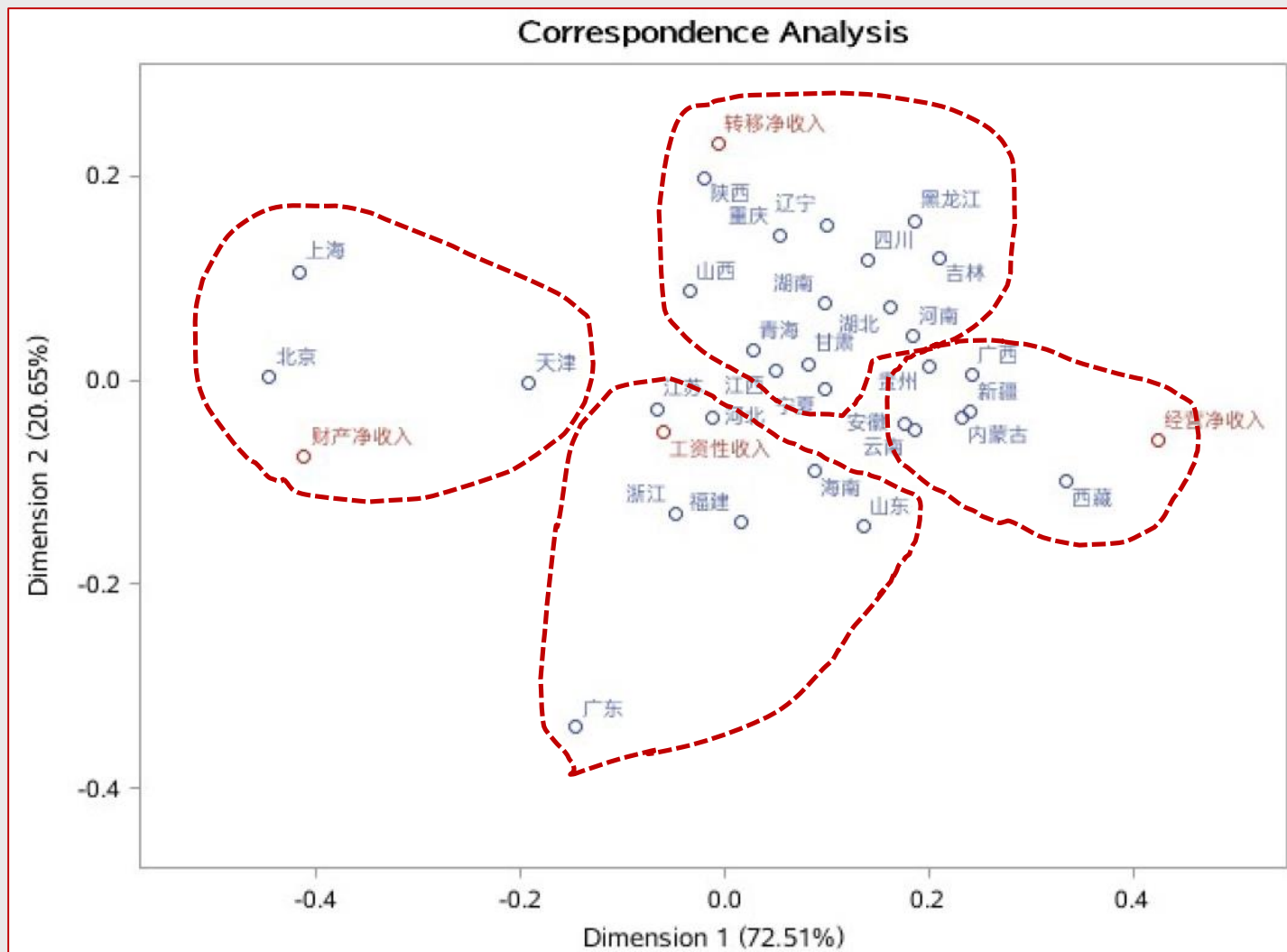
SAS 输出

输出6: 对应分析图



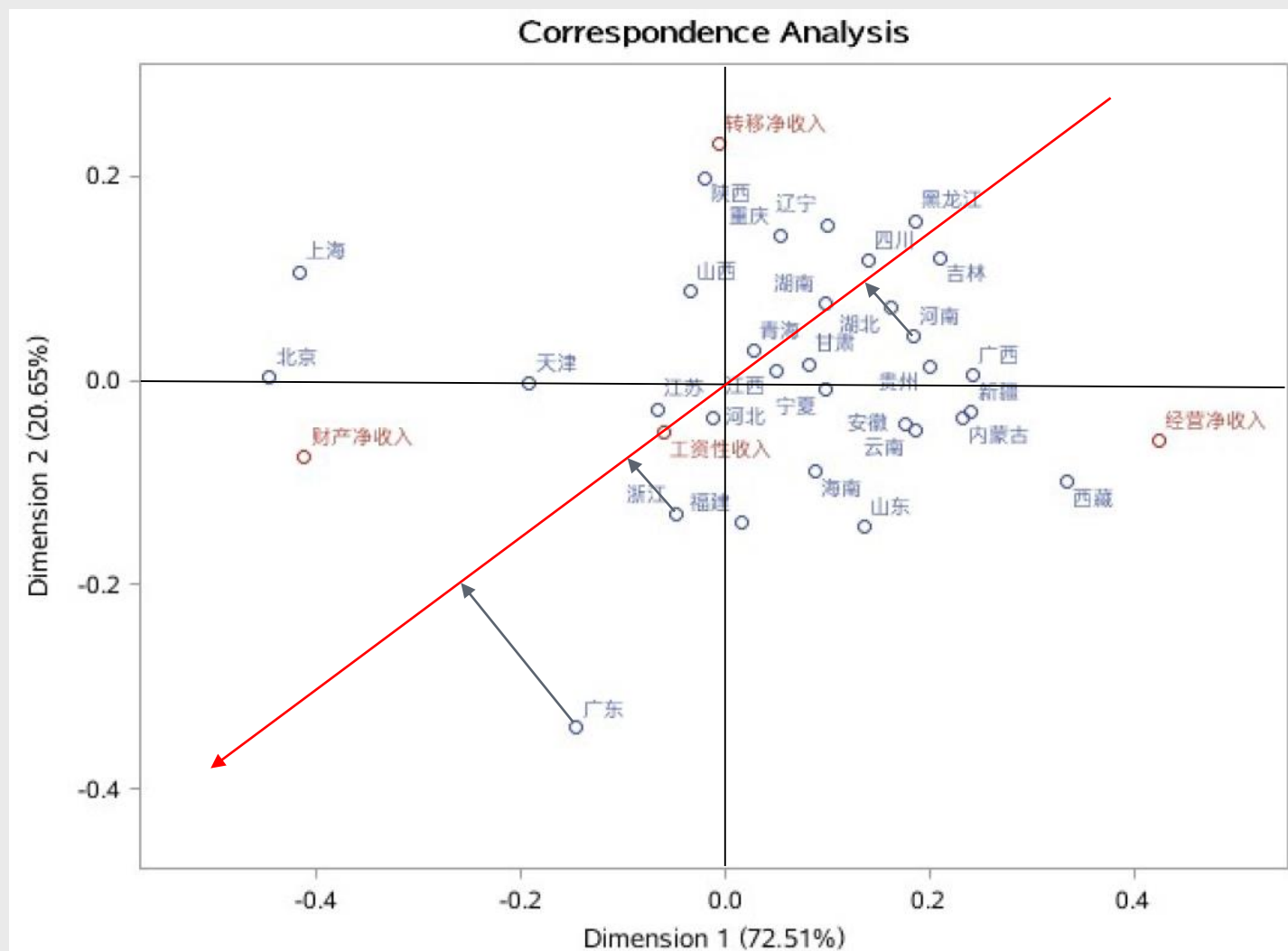
结果分析

(1) 关联性分析



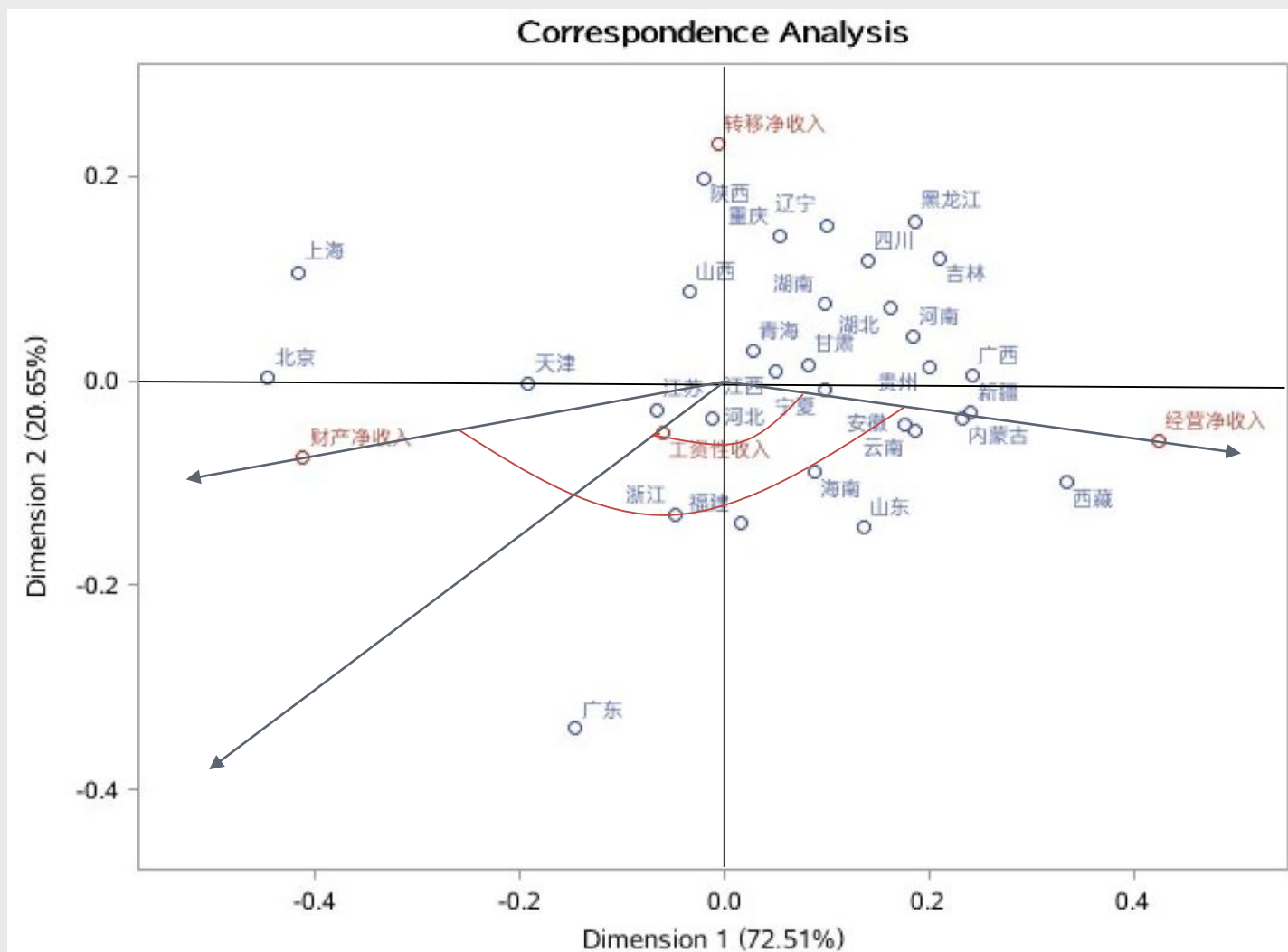
结果分析

(2) 向量分析



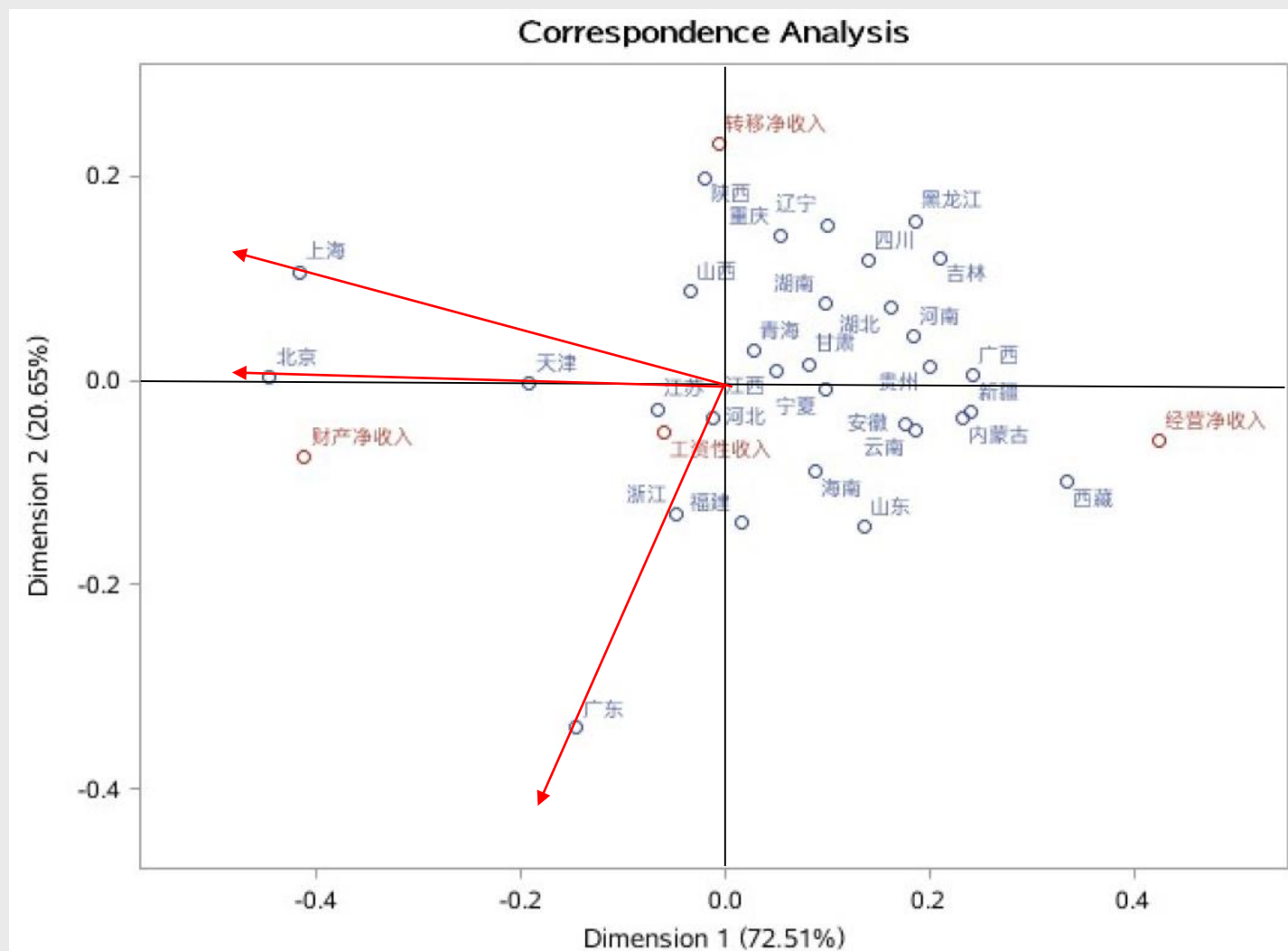
结果分析

(3) 向量夹角分析



结果分析

(3) 向量夹角分析



四、对应分析方法的优缺点

优点

- (1) 定性变量划分的类别越多，这种方法的优越性越明显
- (2) 揭示行变量类间与列变量类间的联系
- (3) 直观地表变量所含类别间的关系

缺点：

- (1) 不能用于相关关系的假设检验
- (2) 受极端值的影响

主要应用领域

- 概念发展 (Concept Development)
- 新产品开发 (New Product Development)
- 市场细分 (Market Segmentation)
- 竞争分析 (Competitive Analysis)
- 广告研究 (Advertisement Research)

主要回答以下问题

- 谁是我的用户？
- 还有谁是我的用户？
- 谁是我竞争对手的用户？
- 相对于我的竞争对手的产品，我的产品的定位如何？
- 与竞争对手有何差异？
- 我还应该开发哪些新产品？
- 对于我的新产品，我应该将目标指向哪些消费者？

- 对应分析是一种数据分析技术，它能够帮助我们研究由定性变量构成的交互汇总表来揭示变量间的联系。
- 交互表的信息以图形的方式展示。
- 主要适用于有多个类别的定类变量
Category Data，可以揭示同一个变量的各个类别之间的差异，以及不同变量各个类别之间的对应关系。适用于两个或多个定类变量。
- 也可以，揭示样品和变量间的内在联系

原始数据

```
proc corresp data=tmp1.examp911 out=results rp cp all;  
    tables 行变量名, 列变量名;  
    weight 变量名;  
run;  
proc plot data=results;  
    plot dim1*dim2=_NAME_ /box vspace=6 hspace=10  
    haxis=-.30 to .30 by .15 vaxis=-.30 to .30 by .15;  
run;
```

作图字符由变量=_NAME_ 给出, box 指要求画出的边框围住整个图形, vspace=6, hspace=10规定图中纵坐标、横坐标单位格在图中的实际长度.

数据是列联表形式

```
data examp911;  
  input mental $ A B C D E;  
  cards;
```

```
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
```

```
;
```

```
proc corresp data=examp911 out=results rp cp short;
```

```
  var A B C D E;  
  id mental;  
run;
```

VAR 语句，指示列联表中的列是A ,B,C,D,E，
id 语句，指定用mental的值作为输出列联表中行的
名称。

```
proc plot data=results;
```

```
  plot dim1*dim2="*$ mental /box vspace=6 hspace=10 haxis=-.30  
to .30 by .15 vaxis=-.30 to .30 by .15;  
run;
```

```
ods graphics on;  
proc corresp data=tmp1.examp911 out=results rp cp all;  
  tables 行变量名, 列变量名;  
  weight 变量名;  
run;  
ods graphics off;
```

Output Delivery System (ODS) 输出传递系统,
应用ods graphics得到SAS过程步的统计图



浙江工商大學
ZHEJIANG GONGSHANG UNIVERSITY

Thank You