对应分析新旧方法差异的探讨

刘照德，叶舒婷，林海明，Miraj Ahmed Bhuiyan
广东财经大学经济学院，广东 广州 510320

摘　　要： 对应分析新方法显著提升传统方法的有效性，为了说明新旧方法及其结果的差异，这里找出了十项比较内容和差异度量方法，进行逐项比较和实证。结果表明新方法结果与原始数据差异小，新方法能优良地保留原始数据的特征，能更好的达到对应分析目的，能更深入地分析和解释数据。

关 键 词： 对应分析；新旧方法；差异；比较

Discussion on the Differences between New and Old Methods of Corresponding Analysis

Liu Zhaode,Ye Shuting, Lin Haiming, Miraj Ahmed Bhuiyan
School of Economics, Guangdong University of Finance and Economics, Guangzhou, Guangdong 510320

Abstract: Corresponding analysis has recently produced new methods to illustrate the differences between old and new methods and their results. Here, we find out ten comparative items and methods of difference measurement, and compare and demonstrate them item by item. The difference between the new method results and the original data is small. The new method can preserve the characteristics of the original data well enough, better achieve the purpose of corresponding analysis, and analyze and interpret the data more deeply.

Keywords: corresponding analysis; new and old methods; differences, comparison

引言

对应分析于1933年由 Richardson和 Kuder提出[1]，迄今已有90年历史。对应分析的实际背景涉及自然科学和社会科学的许多领域，已得到了广泛的重视。对应分析的目的是将数据阵中的列变量与行变量图表示在低维坐标系中，能直观地找出列变量之间、行变量之间、列变量与行变量之间的关系。

从20世纪的30年代到70年代的40年间，许多著名的统计学家如 Fisher(1940)[2]、Maung(1941)[3]、Guttman(1941)、Williams (1952)[4]、Lancaster(1953)[5]等参与研究对应分析模型和计算准则，各自声称独立建立了一种新的统计方法，并冠以不同的名字，但这些方法的优化准则基本等价，计算结果基本一致，这在学科发展史上是比较罕见的[6]。迄今国内外流行的对应分析法，是1970年 Benz é cri[7]给出的，其是对数据阵的行、列变量进行对等变换 +R 型因子载荷阵 +Q 型因子载荷阵的图，下称 B 氏方法或旧方法。

但有国内外的专家就对等变换提出了质疑：对等变换对列、行变量中的每个分量进行了非线性变换，如所周知，变量的非线性变换会扭曲变量之间的关系，即列、行变量的非线性变换会扭曲列变量之间、行变量之间、行变量与列变量之间的关系，这样做能达到对应分析目的吗？又如杜子芳 (2016) 认为[8]：当列变量量纲不同时，对等变换中对同一行变量的值相加的做法是不适应的 (如某教室10张凳子 +10张桌子不具有可加性，加的结果20是错误的)，从而 B 氏方法一开始就是不适应的。

因此，2018年刘照德等用具有优良性的因子分析模型 L，给出了改进的对应分析法 (见附1)[9]，其是对数据阵的列变量进行标准化变换 +R 型因子载荷阵 + 因子值矩阵的图，下称 L 氏方法或新方法。

毫无疑问，B 氏方法、L 氏方法及其结果是有差异的，人们在用数据阵作对应分析时，为了解决问题，自然希望找到与原始数据差

项目信息：本文获国家社科基金西部项目“因子分析最小误差模型及其检验的建立与应用”（23XTJ008）资助。

作者简介：
刘照德（1970–），湖南武冈人，教授，博士，硕士生导师，研究方向：经济统计、多元统计、计量分析等；
叶舒婷（2000–），硕士研究生，研究方向：经济统计；
林海明（1959–），教授，研究方向：多元统计等；
Miraj Ahmed Bhuiyan，（1986–），副教授，研究方向：计量模型等。

此文得到方开泰教授的支持和指导，特此鸣谢。

异小、能达到对应分析目的的更好方法与结果，于是有问题：

（1）B氏方法、L氏方法及其结果有哪些差异?

（2）哪个方法得到的结果与原始数据差异小、能更好地达到对应分析目的?

据国内外有关文献检索，如近期文献[10]-[15]等，没有研究上述问题。这里从两种方法的基本原理入手，进行方法、结果和原始数据的比较，解决问题。

一、B氏方法与L氏方法的原理

设数据阵$X=(x_{ij})_{n\times p}$，X对应的p维列变量$x=(x_1,x_2,\cdots,x_p)'$、n个行变量为$x_{(1)},x_{(2)},\cdots,x_{(n)}$，对应分析的目的可描述为：将$p$维列变量x与$n$个行变量$x_{(1)},x_{(2)},\cdots,x_{(n)}$图表示在低维坐标系中，能直观地找出列变量x之间、行变量$x_{(1)},x_{(2)},\cdots,x_{(n)}$之间、行变量$x_{(1)},x_{(2)},\cdots,x_{(n)}$与列变量x之间的关系。

（一）B氏方法的原理

R型因子分析的因子载荷阵可表示列变量（估计），Q型因子分析的因子载荷阵可表示行变量（估计）。但由于列变量的协差阵与行变量的协差阵的特征值不等，不能建立R型因子载荷阵与Q型因子载荷阵之间的关系[16]。B氏方法受到列联表两因素独立性卡方检验统计量的启发，对数据阵X的行、列变量进行对等变换，

$$z_{ij}=(x_{ij}-x_{i.}x_{.j}/x_{..})/(x_{i.}x_{.j})^{1/2},$$

得数据阵$Z=(z_{ij})_{n\times p}$，这里$x_{i.}=\sum_{j=1}^{p}x_{ij}$，$x_{.j}=\sum_{i=1}^{n}x_{ij}$，$x_{..}=\sum_{i=1}^{n}\sum_{j=1}^{p}x_{ij}$，此时Z的行变量与列变量的协差阵非零特征值相等，主成分法下，能建立Z的R型因子载荷阵A_R（列变量）与Q型因子载荷阵A_Q（行变量）之间的密切关系，从而得出B氏方法结果：数据阵X行、列变量对等变换+R型因子载荷阵A_R+Q型因子载荷阵A_Q的图，即B氏方法结果是数据阵Z行、列变量的降维图。B氏方法能依据列变量、行变量的接近程度，揭示列变量之间、行变量之间、列变量与行变量之间的关系，使得问题的分析带来许多方便。[16]

（二）L氏方法的原理[9]

因子分析模型L有因子载荷阵及其因子解：主成分法下，R型因子载荷阵L及其回归的因子f（不是估计）。L氏方法受到因子载荷阵L（列变量）与因子值矩阵F（行变量）可表示在同一因子坐标系图中的启发，对数据阵X的列变量进行标准化变换，

$$s_{ij}=(x_{ij}-\bar{x}_j)/\sigma_j,$$

得数据阵$S=(s_{ij})_{n\times p}$，这里$\bar{x}_j=\sum_{i=1}^{n}x_{ij}/n$，$\sigma_j=[\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)/(n-1)]^{1/2}$，此时主成分法下，R型因子载荷阵L（列变量）与因子值矩阵F（行变量）有直接关系，从而得出L氏方法结果：数据阵X列变量标准化变换+R型因子载荷阵L+因子值矩阵F的图，即L氏方法结果是数据阵S行、列变量的降维图。所以，L氏方法是常规R型因子分析中因子载荷阵+因子值矩阵的图结果。如所周知，其能依据列变量、行变量的接近程度，揭示列变量之间、行变量之间、列变量与行变量之间的关系，能深入地分析和解释数据。

二、B氏方法与L氏方法及其结果的差异

（一）方法差异

表1 方法差异比较

	方法	B氏方法	L氏方法
1	行变量表示	Q型因子载荷阵A_Q（是估计）。	因子值矩阵F（是解，不是估计）。
2	行、列变量关系	数据阵X的R型因子载荷阵与Q型因子载荷阵不能直接建立联系。	数据阵X的R型因子载荷阵L（列变量）与因子值矩阵F（行变量）是直接关系。
3	变换意图	对等变换意图是建立变换后数据阵Z的R型、Q型因子载荷阵的联系。	标准化变换意图是消除列变量量纲不同的影响，得出行变量之间的相对可比性。
4	适应性	X列变量量纲不同时，行数值无可加性，对等变换中$x_{i.}=\sum_{j=1}^{p}x_{ij}$无适应性。	没有X行数值的加法运算，有更广泛的适应性。
5	变换特性	对等变换是非线性变换，会扭曲列变量之间、行变量之间、行变量与列变量之间的关系。	标准化变换是线性变换，完整保留了列变量之间、行变量之间、行变量与列变量之间的关系。

即相比之下，L氏方法是对列变量进行等价的标准化变换，有更广泛的适应性，行变量用因子值矩阵表示，直接有行、列变量的关系；放弃了对等变换、Q型因子分析。

（二）结果差异

表2 结果差异比较

	结果	B氏方法	L氏方法
1	变换等价性	两因素独立时，对等变换将数据阵X变为数据阵$Z=0$[9]，X与Z不等价。	标准化变换将数据阵X变为数据阵S，X与S等价。
2	图	数据阵Z的降维图。	数据阵S的降维图。
3	因子表示式	无。	有。
4	因子意义和方向	无。	有，能旋转，因子意义较清晰。
5	行、列变量关系	不能说明列变量之间[11]、行变量之间、行变量与列变量之间的关系。	能优良地保留列变量之间、行变量之间、行变量与列变量之间的关系。

即相比之下，L氏方法的结果，能作出贴近实际意义的数据分析和解释，能优良地保留列变量之间、行变量之间、行变量与列变量之间的关系，能更好的达到对应分析目的。

（三）结果差异的度量

（1）列变量差异度量，原始数据阵X的列变量关系用相关阵

R= $(r_{ij})_{p\times p}$度量，B(L)氏方法列变量关系用R型因子载荷阵A_R(L)计算的相关阵$R_B=(r_{Bij})_{p\times p}$[$R_L=(r_{Lij})_{p\times p}$]度量，如果max｜$r_{ij}-r_{Bij}$｜(max｜$r_{ij}-r_{Lij}$｜)大，则B(L)氏方法结果差异大，否则B(L)氏方法结果差异小。

（2）行变量差异度量，原始数据阵X的行变量$x_{(j)}$是列变量x的取值，按列变量x有大小顺序，故X的行变量关系用$x_{(j)}$在列变量x_i的排序k度量。设$x_{(j)}$在列变量x_i的排序为k，如果B(L)氏方法的行变量$x_{(j)}$在x_i的排序是k的附近，则B(L)氏方法结果差异小，否则B(L)氏方法结果差异大。

注：这里用排序差度量行变量差异，而不用距离差异度量行变量差异，原因为列变量变换后距离差异不易测量，而排序差异容易测量，其也是人们常用的行变量差异度量。

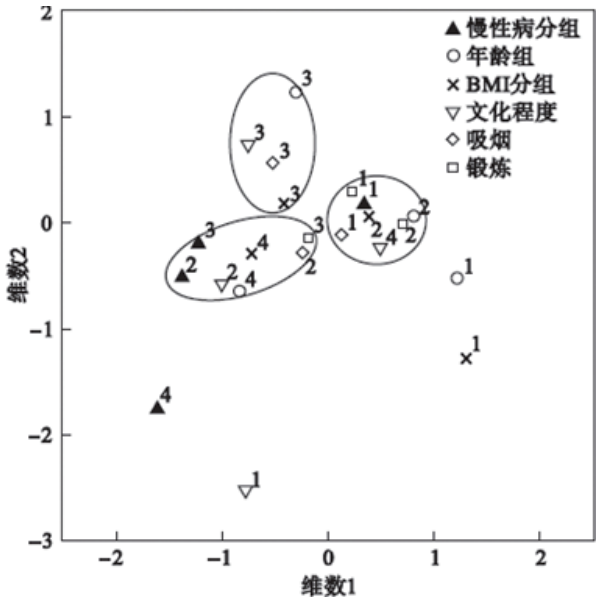
（3）行变量受列变量的影响差异度量，X中$x_{(j)}$的x_i值（排序）越大，$x_{(j)}$受x_i的影响越大，如果B(L)氏方法中$x_{(j)}$、x_i在一起，则B(L)氏方法结果差异小，否则B(L)氏方法结果差异大。

三、新旧方法结果差异的实证比较

文[17]给出了2013年北京市青壮年患慢性病数据(见表5)，列变量为x_1-不患慢性病、x_2-仅患高血压、x_3-仅患糖尿病、x_4-患糖尿病和高血压。

（一）B氏方法实证结果

文[17]用表5数据给出了B氏方法结果图1及其四个结论：



> 图1 B氏方法对应分析图

第一：不患病(▲[1])与30岁年龄段(○[2])、BMI正常(×[2])、高中及以上文化程度(▽[4])、不吸烟(◇[1])、近半年不锻炼和每周锻炼次数小于1次(□[1]、□[2])距离较近。

第二：仅患高血压、仅患糖尿病(▲[2]、▲[3])两点距离较近，与这两点临近的有50 ~ 59岁(○[4])、BMI肥胖(×[4])、小学文化程度(▽[2])、非每天吸烟(◇[2])、近半年每周锻炼次数1次及以上(□[3])。

第三：40岁年龄段(○[3])与每天吸烟(◇[3])、初中文化程度组(▽[3])、BMI超重(×[3])距离较近。

第四：患高血压和糖尿病(▲[4])、没上过学(▽[1])、BMI偏瘦(×[1])、20岁年龄段(○[1])这四点与其余各点距离均较远。

现在用表5原始数据、图1结果，指出B氏方法结论与原始数据差异大的情况：

文[17]第一结论的差异：从表5及其排序结果有，不患慢性病受年龄组20~、30~，BMI组0~18.5、18.5~24，高中及以上，从不锻炼影响较大，排序分别为第1、2、3、4、5、6，而文[17]B氏方法结果出现了不患慢性病与年龄组20~、BMI组0~18.5不在一起的结果。按行变量受列变量的影响差异度量有，B氏方法结果差异大。

文[17]第二结论的差异：从表5及其排序结果有，仅患高血压受年龄组50 ~ 59，BMI组24-28、肥胖，没上过学、小学、初中影响排序分别为第1、6、2、4、3、5，而文[17]出现了仅患高血压病与BMI组24-28、没上过学、初中不在一起的结果；同理，文[17]B氏方法结果出现了仅患糖尿病与没上过学、初中影响不在一起的结果。按行变量受列变量的影响差异度量有，B氏方法结果差异大。

文[17]第三结论的差异：从表5及其排序结果有，40岁年龄段(○[3])与每天吸烟(◇[3])、初中文化程度组(▽[3])、BMI超重(×[3])在x_1、x_2、x_3、x_4的排序差距分别大至6、6、7、8，排序不在附近。而文[17]B氏方法结果出现了距离较近的结果。按行变量影响差异度量有，B氏方法结果差异大。

文[17]第四结论的差异：从表5及其排序结果有，同时患高血压和糖尿病受年龄组50 ~ 59，BMI组肥胖，没上过学、小学、初中、非每天吸烟影响较大，排序分别为第2、1、3、4、6、5，而文[17]B氏方法结果同时患高血压和糖尿病与年龄组50 ~ 59，BMI组肥胖，没上过学、小学、初中、非每天吸烟全部不在一起的结果。按行变量受列变量的影响差异度量有，B氏方法结果差异大。综上所述，B氏方法的结果差异大。

（二）L氏方法实证结果

按照L氏方法应用步骤[9]用表5原始数据有结果：因子个数为1，因子

$f_1=-0.261x_1+0.258x_2+0.248x_3+0.255x_4$（列变量是标准化变换后的变量），

f_1解释了95.66%的信息，f_1与4个变量显著相关，故f_1命名为患慢性病影响水平因子，f_1是越大越不好的逆向变量，行变量、列变量在因子轴上的数值和图见表3、图2。

表1 行变量、列变量在因子轴上的数值表示

行变量	因子值矩阵	序	列变量	初始因子载荷阵
20~（年龄组）	-1.815	18	x_1	-0.997
30~	-1.434	17	x_2	0.987
40~	-0.271	11	x_3	0.951
50~59	1.714	1	x_4	0.976
0-18.5(BMI组)	-1.212	16		
18.5-24	-0.672	15		

行变量	因子值矩阵	序	列变量	初始因子载荷阵
24-28	0.316	7		
28及以上	1.383	3		
没上过学（文化程度）	1.456	2		
小学	1.011	4		
初中	0.649	5		
高中及以上	−0.591	14		
不吸（吸烟情况）	−0.153	10		
非每天吸	0.358	6		
每天吸	0.101	9		
从不锻炼（近半年每周锻炼次数）	−0.574	13		
不到1次	−0.496	12		
1次及以上	0.230	8		

> 图2 因子双重信息图

由图2，在因子f_1轴上行变量和列变量分为两类：

第一类“正向”类：行变量有年龄组50~59、BMI组24-28、28及以上、没上过学、小学、初中、锻炼次数1次及以上、非每天吸烟、每天吸烟，变量有x_2、x_3、x_4。

第二类“负向”类：行变量有年龄组20~、30~、40~、BMI组0-18.5、18.5-24、高中及以上、锻炼次数从不锻炼、不到1次、不吸烟，变量有x_1。

表4 列变量相关系数

相关系数	x_1,x_2	x_1,x_3	x_1,x_4	x_2,x_3	x_2,x_4	x_3,x_4
r_{ij}−原始数据	−0.996	−0.930	−0.973	0.905	0.959	0.886
r_{Lij}−L氏方法	−1.000	−1.000	−1.000	1.000	1.000	1.000

列变量方面 由表4，通过数据阵X列变量相关系数r_{ij}、L氏方法列变量相关系数r_{Lij}[1]进行比较有，差异最大值 max｜$r_{ij}-r_{Lij}$｜=0.114，保留了列变量的意义、方向和相关性。按列变量差异度量有，L氏方法结果差异小，优良地保留了列变量之间的关系。

行变量方面 由表5，原始数据列变量值与因子值的同行排序值（f_1中的x_1是负号，故x_1的排序要颠倒排序），只有一个差异值为3，其余差异值为0~2。按行变量差异度量有，L氏方法结果差异小，优良地保留了行变量之间的关系。

表5 2013年北京市青壮年患慢性病原始数据（%）、因子值与排序

分组	x_1	序	x_2	序	x_3	序	x_4	序	f_1	序
20~（年龄组）	99.21	1	0.42	18	0.32	18	0.05	18	−1.815	18
30~	94.69	2	4.01	17	1.01	17	0.28	17	−1.434	17
40~	82.40	8	12.62	11	3.22	10	1.79	14	−0.271	11
50~59	61.78	18	26.11	1	5.77	2	6.34	2	1.714	1
0~18.5（偏瘦）(BMI组)	93.32	3	4.06	16	1.43	16	1.19	16	−1.212	16
18.5~24（正常）	87.73	4	7.98	15	2.79	13	1.51	15	−0.672	15
24~28（超重）	75.58	13	17.59	6	3.56	9	3.27	8	0.316	7
28及以上（肥胖）	63.89	17	25.45	2	3.92	5	6.74	1	1.383	3
没上过学（文化程度）	66.12	16	21.86	4	6.01	1	6.01	3	1.456	2
小学	67.70	15	23.31	3	4.21	4	4.78	4	1.011	4
初中	73.16	14	18.40	5	4.54	3	3.91	6	0.649	5
高中及以上	85.93	5	9.84	14	2.28	14	1.95	12	−0.591	14
不吸（吸烟情况）	81.18	9	13.19	10	2.91	11	2.72	10	−0.153	10
非每天吸	76.95	12	14.99	8	3.75	8	4.32	5	0.358	6
每天吸	79.02	10	14.35	9	3.79	7	2.84	9	0.101	9
从不锻炼（近半年每周锻炼次数）	84.94	7	11.15	12	1.94	15	1.97	11	−0.574	13
不到1次	85.31	6	10.06	13	2.82	12	1.81	13	−0.496	12
1次及以上	77.75	11	15.08	7	3.85	6	3.32	7	0.230	8

行变量与列变量方面 由表5逐一对比，L氏方法行变量与列变量之间的结果与原始数据的特征基本相同：①表5原始数据排序结果有，年龄组20~、30~、BMI组0~18.5（偏瘦）、18.5~24（正常）、高中及以上、近半年每周锻炼次数不到1次受不患慢性病（x_1）的影响较大（排序分别为第1、2、3、4、5、6）；图2结果有，年龄组20~、30~、BMI组0~18.5（偏瘦）、18.5~24（正常）、高中及以上、近半年每周锻炼次数不到1次与不患慢性病（x_1）在一起。②表5原始数据排序结果有，年龄组50 ~ 59，BMI组28以上、没上过学、小学、初中、非每天吸烟受仅患高血压（x_2）慢性病的影响较大（排序分别为第1、2、4、3、5、8），受仅患糖尿病（x_3）慢性病的影响较大（排序分别为第2、5、1、4、3、8），受患糖尿病和高血压（x_4）慢性病影响较大（排序分别为第2、1、3、4、6、5）；图2结果有，年龄组50~59、BMI组28以上、没上过学、小学、初中、非每天吸烟与仅患高血压（x_2），仅患糖尿病（x_3），患糖尿病和高血压（x_4）慢性病在一起。按行变量受列变量的影响差异度量有，L氏方法结果差异小，优良地保留了行变量与列变量之间的关系。

四、结论

B氏（旧）方法通过对等变换，建立了变换后数据阵列变量与行变量之间的关系，但变换后数据阵与原始数据差异大，以至旧方法结果与原始数据阵差异大，不能达到对应分析目的。L氏（新）方法用标准化变换替代对等变换，用因子列向量值矩阵作为

1 图相关系数的计算，设变量x_i在坐标系$(f_1,\cdots,f_m)$中的坐标为$(l_{i1},\cdots,l_{im})$ $(i=1,\cdots,p)$，变量x_i是中心化时，即$E(x_i)=0$，x_i与x_j的相关系数为：$r_{ij}=\sum_{s=1}^{m}l_{is}l_{js}/\sqrt{\sum_{s=1}^{m}l_{is}^2\sum_{s=1}^{m}l_{js}^2}$。

行变量替代Q型因子分析，R型因子载荷阵（列变量）与因子值矩阵（行变量）有直接关系，标准化变化是数据阵的等价变换，新方法结果与原始数据阵差异小，能更好的达到对应分析目的，能更深入地分析和解释数据。

附1：对应分析新模型（方法）如下：设数据阵 $X=(x_{ij})_{n\times p}$ 的列变量 $x=(x_1,\cdots,x_p)'$ 是标准化变换后的列变量，降维因子坐标系为 $f=(f_1,\cdots,f_m)'$，$m\leqslant p$。

x_i 在f中的表示为：$x_i=l_{i1}f_1+\cdots+l_{im}f_m+\varepsilon_i$，$l_{ij}$ 是 x_i 在 f_j 上的投影，$i=1,\cdots,p$，即：

$$x=Lf+\varepsilon \tag{1}$$

这里 $L=(l_{ij})_{p\times m}$ 称为x在f上的投影矩阵，$\varepsilon=(\varepsilon_1,\cdots,\varepsilon_p)'$ 是误差向量，

$$E(f)=0,\ E(\varepsilon)=0,\ Cov(f)=I_m,\ Cov(f,\varepsilon)=0 \tag{2}$$

$$tr(LL')\text{达到最大} \tag{3}$$

X的行变量 $x_{(j)}$ 在f中的近似表示为：f对 $x_{(j)}$ 的投影 $(f_{1j},\cdots,f_{mj})=f'|x=x_{(j)}$，$j=1,\cdots,n$，即

$$F=(f_{ij})_{p\times m} \tag{4}$$

F是f对行变量 $x_{(1)},x_{(2)},\cdots,x_{(n)}$ 的投影，称为因子值矩阵。

在坐标系 $f=(f_1,\cdots,f_m)'$ 中，以L的行 $(l_{i1},\cdots,l_{im})$ 为列变量 $x_i(i=1,\cdots,p)$ 的坐标，以F的行 $(f_{1j},\cdots,f_{mj})$ 为行变量 $x_{(j)}(j=1,\cdots,n)$ 的坐标点图，即得对应分析优化模型。

设 $L^*=(l_{ij}^*)_{p\times m}$ 是主成分法的因子载荷阵，$f^*=(f_{i1}^*,\cdots,f_{im}^*)'$ 是 L^* 回归的因子（得分），$(f_{i1}^*,\cdots,f_{im}^*)'=f^{*\prime}|x=x_{(j)}$，因子值矩阵 $F^*=(f_{ij}^*)_{n\times m}$。

因子载荷阵 L^*、因子 f^*、因子值矩阵 F^*，都能用SAS、SPSS等软件计算。

定义1　以因子 f^* 为坐标系，因子载荷阵 L^* 的第 i 行 $(l_{i1}^*,\cdots,l_{im}^*)$ 为列变量 $x_i(i=1,2,\cdots,p)$ 点图，$f^{*\prime}|x=x_{(j)}\backsimeq(f_{j1}^*,\cdots,f_{jm}^*)$ 为行变量 $x_{(j)}(j=1,\cdots,n)$ 点图，称为因子双重信息图。

定理1　对应分析优化模型的解（不是估计）是因子双重信息图。

性质1　因子双重信息图，优良地反映了列变量之间、行变量之间、行变量与列变量之间的关系，能达到对应分析目的。

性质2　因子双重信息图有旋转功能，因子有较清晰的意义和方向。

对应分析优化模型的分类标准　因子双重信息图中，方向靠近的列变量是一类，它们之间的相关性较高；位置靠近的行变量是一类，它们之间的优劣性特征较相近；行变量受靠近的列变量影响较大。显然，$m\leqslant 3$ 时，能作图；$m>3$ 时，不能作图，但可用坐标系中行变量、列变量的坐标值，确定行变量、列变量、行变量与列变量的远近，进行分类；也可在f的两两因子轴平面上作图。

参考文献

[1]Richardson, M., and Kuder, G.F. ***Making a rating scale that measures***. Personnel Journal, 12, 1933, 36-40.

[2]Fisher R.A. ***The precision of discriminant functions***. Annals of Eugenics, 10, 1940, 422-429.

[3]Guttman, L. ***The quantification of a class of attributes: A theory and Method of scale construction***. ***In The Committee on Social Adjustment***(ed.), The Prediction of Personal Adjustment. New York: Social Science Research Council, 1941.

[4]Lancaster, H.O. ***A reconciliation of x2, considered from metrical and enumerative aspects***. Sankhya, 13, 1953, 1-10.

[5]Williams, E.J. ***Use of scores for the analysis of association in contingency tables***. Biometrika, 39, 1952, 274-89

[6]李卫东编著. 应用多元统计分析. 北京：北京大学出版社，2015。

[7]Benz é cri, J.P., et al. L′ ***Analyse des données: II. L'Analyse des correspondences***. Paris: Dunod, 1973.

[8]杜子芳著. 多元统计分析. 北京：清华大学出版社，2016。

[9]刘照德，等. 对应分析法的改进与应用. 数理统计与管理，2018，2：243-254.

[10]张文婷，等. 体检人群慢性病指标与中医体质的多重对应分析［J］. 护理研究，2024，38 (19)：3429-3434.

[11]罗盛，等. 地区恶性肿瘤死亡率的对应分析. 数理统计与管理，2009.28(3)：566-570.

[12]李苍舒. 我国金融业效率的测度及对应分析［J］. 统计研究，2014，1：91-97.

[13]Audigier V., Husson F., Josse J., ***MIMCA: multiple imputation for categorical variables with multiple correspondence analysis***, Stat Comput, 2017, 27:501 - 518.

[14]Blasius J., Greenacre M., Groenen P.J.F., Velden M., ***Special issue on correspondence analysis and related methods***, Computational Statistics and Data Analysis, 2009, 53: 3103-3106.

[15]Beh E. J., ***Elliptical confidence regions for simple correspondence analysis***, Journal of Statistical Planning and Inference, 2010, 140: 2582 - 2588.

[16]方开泰编著. 实用多元统计分析. 上海：华东师范大学出版社，1989.

[17]陈卓然，等. 北京市青壮年人群慢性病患病相关因素的多重对应分析，中国卫生统计，2017.34 (1)：40-46.