合理进行多元分析——度量型多维尺度分析

胡纯严[1]，胡良平[1,2*]

（1. 军事科学院研究生院，北京 100850；
2. 世界中医药学会联合会临床科研统计学专业委员会，北京 100029
*通信作者：胡良平，E-mail：lphu927@163.com）

【摘要】 本文目的是介绍与度量型多维尺度分析有关的基本概念、计算方法、两个实例以及SAS实现。基本概念包括度量型与多维尺度、应用场合、距离矩阵与相似系数矩阵、研究对象、拟合构图与构图、多维尺度分析的基本思想；计算方法涉及经典多维尺度分析和最小平方多维尺度分析；两个实例分别为“甘肃省12个主要城市之间的距离”和“8个测试项目之间的相关系数矩阵”；借助SAS软件，对两个实例分别进行了多维尺度分析，并对SAS输出结果做出了解释。

【关键词】 度量型；多维尺度；距离矩阵；相似系数矩阵；拟合构图

中图分类号：R195.1 **文献标识码**：A doi：10.11886/scjsws20230726003

Reasonably carry out multivariate analysis: metric multidimensional scaling analysis

Hu Chunyan[1], *Hu Liangping*[1,2*]

（1. *Graduate School, Academy of Military Sciences PLA China, Beijing* 100850, *China*;
2. *Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing* 100029, *China*
*Corresponding author: *Hu Liangping*, *E-mail*: *lphu927@163.com*）

【**Abstract**】 The purpose of the paper was to introduce the basic concepts, calculation methods, two examples and SAS implementation related to the metric multidimensional scaling analysis. Basic concepts included metric and multidimensional scaling, application occasions, distance matrix and similarity coefficient matrix, research objects, configuration and composition, basic ideas of multidimensional scaling analysis. The calculation method involved classical multidimensional scaling analysis and least square multidimensional scaling analysis. The data in the two examples were "distance between 12 major cities in Gansu province" and "correlation coefficient matrix between 8 test items". With the help of SAS software, the multidimensional scale analysis was carried out on the data in the two examples, and an explanation was given for the output results of SAS.

【**Keywords**】 Metric; Multidimensional scale; Distance matrix; Similarity coefficient matrix; Configuration

多维尺度分析又称为多维标度分析，是多元统计分析方法中的一个重要内容。该方法最早由Torgerson在20世纪50年代提出，适用于可获得研究对象之间精确的相似性或相异性数据的情形。多维尺度分析可以对数据进行探索性分析，可用来进行相似性评价，也是检验观察数据是否符合研究者提出的结构关系的一种理想方法。多维尺度分析可以分为度量型多维尺度分析和非度量型多维尺度分析。本文将结合实际问题介绍度量型多维尺度分析的基本概念、计算方法以及SAS实现，并对输出结果进行解释。

1 基本概念

1.1 度量型与多维尺度

所谓度量型，就是所提供的反映任何两个受试对象之间相似或相异程度的数值具有较高的精确度。所谓多维尺度，就是假定任何一个观测对象都由多个维度来决定，例如，观测了很多个体的身高、体重、胸围、心像面积，即每个个体对应着一个四维空间中的一个点。一般来说，每个个体对应着一个m维欧氏空间中的一个点。如果存在这样的两个个体，他们在m个维度上的取值都非常接近，即这两个个体在m维欧氏空间中所对应的两个点的位置十分接近，可以说这两点间的欧氏距离接近0。

1.2 应用场合

假设：有n个个体，每个个体由m个指标（变量）反映其在m维欧氏空间的位置。但反映个体的指标个数m的具体取值不清楚，甚至指标本身是什么也是模糊的，更难以对它们进行观察或直接测量，唯一知道的是这n个个体中任何两者之间的某种距离（不一定是通常的欧氏距离）或者某种相似性或

不相似性，希望仅依据这样的信息就能比较准确地确定这n个个体在m维空间中的相对位置。这就是多维尺度分析要解决的实际问题。

多维尺度分析就是在仅知道全部n个个体（或泛称样品）中任何一对样品之间的“相对距离”（真实距离或相似性度量或不相似性度量）的前提下，通过数学处理，计算出n个样品中每一个样品在m维空间中的坐标（通常m取2或3），再用图形将n个样品在m维空间中呈现出来。从直观角度出发，通常仅在二维空间（即二维平面）中呈现全部n个样品。相当于把m维空间中的n个点投影到二维空间中，以便直观感受到它们之间的相对位置，是以“图示法”呈现样品聚类分析结果的一种方法。

1.3 距离矩阵与相似系数矩阵

一个$n\times n$阶矩阵$D=(d_{ij})_{n\times n}$，如果满足$D=D'$，$d_{ii}=0$，$d_{ij}\geqslant 0(i,j=1,2,\cdots,n)$，则称$D$为广义距离矩阵，$d_{ij}$为第$i$点与第$j$点之间的距离。

一个$n\times n$阶矩阵$C=(d_{ij})_{n\times n}$，如果满足$C=C'$，$c_{ii}\leqslant c_{ij}$ $(i,j=1,2,\cdots,n)$，则称C为相似系数矩阵，c_{ij}为第i点与第j点之间的相似系数。

1.4 研究对象

多维尺度分析利用的是研究对象之间的近似数据，这些近似数据可以分为相异性数据和相似性数据。相异性数据是用较大的数值表示不相似，较小的数值表示相似，也称为距离数据；相似性数据则相反，数值较大表示相似，数值较小表示不相似。与这两者相对应的分别是距离矩阵和相似系数矩阵。

1.5 拟合构图与构图

设D为基于观测的n个样品计算所得的距离矩阵，又设$\hat{D}$为基于推算的n个样品计算所得的距离矩阵。再设与$\hat{D}$对应的n个样品的坐标为X，用矩阵形式将它呈现出来，见式(1)。

$$X=\left(x_1,x_2,\cdots,x_n\right)' \tag{1}$$

则称X为D的一个近似解，或称多维尺度经典解。统计学家形象地称X为距离矩阵D的一个拟合构图（也叫感知图，即采用图形将n个样品在一个事先确定的k维空间中的相对位置呈现出来，通常取k=2）。如果$\hat{D}=D$，则称X为D的一个构图或精确解[1-2]。

1.6 多维尺度分析的基本思想

当已知n个研究个体或样品之间的相似度（或距离）时，多维尺度分析方法可以在低维空间中找出与之相对应的n个点，使得这些点之间的距离与原来的相似度基本匹配。统计学家将这些低维空间中的点以图形的形式呈现出来，它们之间的距离就直观地反映了n个研究个体或样品之间相似程度。在分析过程中，要解决的核心问题是确定这些点的坐标，具体方法被称为“多维尺度分析法的经典解”[3-4]。

2 计算方法

2.1 概述

设有n个研究对象，用δ_{ij}表示第i个与第j个对象之间的相似性或相异性，该数据是已知的。$X_1,\cdots,X_n$代表m维空间中的n个点，第i个点X_i的坐标为$(x_{i1},x_{i2},\cdots,x_{im})$，它是未知的。任意两点$X_i$和$X_j$之间的距离为$\hat{d}_{ij}$，一般采用欧氏距离。多维尺度分析就是要用$X_1,\cdots,X_n$分别表示待研究的$n$个对象，进一步将第$i$个和第$j$个对象之间的相似性或相异性$\delta_{ij}$映射成为$m$维空间中$X_i$与$X_j$之间的距离$\hat{d}_{ij}$，其映射函数见式(2)。

$$f:\delta_{ij}\rightarrow\hat{d}_{ij} \tag{2}$$

对上述映射函数进行相应变换，得到的结果见式(3)。

$$f(\delta_{ij})=\hat{d}_{ij} \tag{3}$$

在实际应用中，由于多维尺度分析采用了降维的技术，n个观测对象的所有信息不可能被全部保留，不同对象之间的差异也不会被完全保留。在降维的过程中，总是有少量信息损失，故一般只能得到尽可能接近的结果。

在度量型多维尺度分析中，函数f可以为恒等函数、线性函数、对数函数、指数函数等，一般情况下采用单调函数。根据求解方法的不同，度量型多维尺度分析主要包括两类：经典多维尺度分析和最小平方多维尺度分析，以下分别进行讨论。

2.2 经典多维尺度分析

对于距离矩阵$D=(d_{ij})_{n\times n}$，如果存在某个正整数m及m维空间中的n个点$X_1,\cdots,X_n$，使得式(4)成立，则称D为欧氏距离矩阵。

$$d_{ij}^2=(X_i-X_j)'(X_i-X_j),\ i,j=1,2,\cdots,n \tag{4}$$

由距离矩阵D,可构造出矩阵$B=(b_{ij})_{n\times n}$,其中的元素由式(5)定义。

$$b_{ij}=\frac{1}{2}\left(-d_{ij}^2+\frac{1}{n}\sum_{i=1}^{n}d_{ij}^2+\frac{1}{n}\sum_{j=1}^{n}d_{ij}^2-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}d_{ij}^2\right) \quad (5)$$

一个$n\times n$距离矩阵D是欧氏距离矩阵的充要条件为$B\geqslant 0$。

用$X=(X_1,X_2,\cdots,X_n)'$表示求得的m维空间中的n个点,这些点之间的距离矩阵$\hat{D}$与D尽可能接近,称X为距离矩阵D的拟合构造点。如果$\hat{D}=D$,则称X为D的构造点。当D是欧氏距离矩阵时,可以得到构造点X;当D不是欧氏距离阵时,不存在D的构造点,只能寻求D的拟合构造点。

在实际应用中,即使D是欧氏距离矩阵,它的构造点X也是$n\times m$阵,当m较大时,往往没有实用的价值,这时一般用低维的拟合构造点$\hat{X}$代替构造点X。

矩阵B称为X的中心化内积阵,求B的特征根$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_m$,对应的单位特征向量为$e_1,e_2,\cdots,e_m$。$\Gamma=(e_1,e_2,\cdots,e_m)$是以单位特征向量为列组成的矩阵,则有$X=(\sqrt{\lambda_1}\,e_1,\sqrt{\lambda_2}\,e_2,\cdots,\sqrt{\lambda_m}\,e_m)$,矩阵$X$中的每一行对应$m$维空间中的一个点,第$i$行即为$X_i$。令$\Lambda=diag(\lambda_1,\lambda_2,\cdots,\lambda_m)$,则有如下结果,见式(6)和式(7)。

$$B=XX'=\Gamma\Lambda\Gamma' \quad (6)$$

$$X=\Gamma\Lambda^{1/2} \quad (7)$$

当B的所有特征根非负,表明$B\geqslant 0$,D是欧氏距离矩阵,X是D的构造点;若有负特征根,D一定不是欧氏距离矩阵,X是D的拟合构造点。

基于上述思想得到的构造点或拟合构造点称为多维尺度分析方法的经典解[5-6]。由以上讨论,可以给出求经典解的基本步骤:①由距离阵D,根据式(5)计算b_{ij},构造出中心化内积阵B;②计算矩阵B的特征根和k个最大特征根对应的单位特征向量,k是低维空间的维数,它的确定有两种方法,一是事先确定k=1、2或3,二是通过计算前k个大于零的特征根占全体特征根的比例,该比例相当于主成分分析中的累积贡献率,见式(8),让式(8)的计算结果大于预先给定的一个贡献率;③根据式(7)计算拟合构造点$\hat{X}$,根据$\hat{X}$可以在k维空间中绘出每个研究对象对应的点,进而判断研究对象之间的关系。

$$\frac{\lambda_1+\lambda_2+\cdots+\lambda_k}{|\lambda_1|+|\lambda_2|+\cdots+|\lambda_n|} \quad (8)$$

2.3 最小平方多维尺度分析

在度量型最小平方多维尺度分析中,通过最小化损失函数S得到低维空间中各个点的坐标[7]。损失函数由研究对象之间的相异性δ_{ij}与低维空间中第i与第j两点之间的距离$\hat{d}_{ij}$定义,也可以由δ_{ij}的函数$f(\delta_{ij})$与距离$\hat{d}_{ij}$定义。

损失函数有很多种形式,在多维尺度分析中,应用最广的是应力函数,此处给出Sammon建议的损失函数,见式(9)。

$$S=\sum_{i<j}\delta_{ij}^{-1}(\hat{d}_{ij}-\delta_{ij})^2\Big/\sum_{i<j}\delta_{ij} \quad (9)$$

通过距离$\hat{d}_{ij}$求损失函数关于各点坐标的偏导数并使之为0,就可得到一个以各点坐标为未知数的方程组。解出该方程组,便可获得各个点的坐标值,进而在低维空间中绘出各点。该方程组的求解需要采用数值解法,例如梯度法。

3 实例与SAS实现

3.1 问题与数据结构

3.1.1 2个实际问题及数据

【例1】表1列出了甘肃省12个主要城市之间的距离[7]。这些距离是公路里程,不是城市之间的最短距离,可视为这些城市之间的近似距离。希望利用这些距离数据绘制一张平面地图,标出这12个城市的位置,使之尽量接近表1给出的距离数据,从而反映它们真实的地理位置。

【例2】一项智力测验共有8个测试项目,这些项目的相关系数矩阵见表2[7]。试通过相关系数矩阵考查这8个测试项目之间的结构关系。

3.1.2 对数据结构的分析

例1中,每个城市到自身的距离为0,所以主对角线上的元素都是0。表1中的数字越大,说明两个城市之间的距离越远。

例2为测试项目之间的相关系数矩阵,该矩阵是对称矩阵。所有的相关系数均大于0,相关系数越大,说明项目之间的关系越紧密,相似程度越高。

3.1.3 创建SAS数据集

分析例1资料,设所需要的SAS数据步程序如下:

表 1 甘肃省 12 个主要城市之间的距离(km)

Table 1 Distance between 12 major cities in Gansu province

城 市	距 离											
	兰州	敦煌	酒泉	张掖	武威	白银	定西	合作	平凉	庆阳	天水	陇南
兰州	0											
敦煌	1 140	0										
酒泉	735	405	0									
张掖	514	626	221	0								
武威	276	864	459	238	0							
白银	90	1 230	825	604	366	0						
定西	104	1 244	839	618	380	182	0					
合作	259	1 399	994	773	535	349	308	0				
平凉	334	1 474	1 069	848	610	364	284	542	0			
庆阳	496	1 636	1 231	1 010	772	521	446	704	162	0		
天水	327	1 467	1 062	841	603	385	223	442	254	367	0	
陇南	450	1 590	1 185	964	726	540	385	362	536	649	282	0

表 2 8 个测试项目之间的相关系数矩阵

Table 2 Correlation coefficient matrix between 8 test items

项 目	1	2	3	4	5	6	7	8
1	1.00	0.40	0.25	0.12	0.67	0.39	0.26	0.19
2	0.40	1.00	0.31	0.39	0.50	0.24	0.18	0.52
3	0.25	0.31	1.00	0.46	0.28	0.38	0.42	0.49
4	0.12	0.39	0.46	1.00	0.20	0.14	0.29	0.55
5	0.67	0.50	0.28	0.20	1.00	0.38	0.26	0.26
6	0.39	0.24	0.38	0.14	0.38	1.00	0.40	0.22
7	0.26	0.18	0.42	0.29	0.26	0.40	1.00	0.25
8	0.19	0.52	0.49	0.55	0.26	0.22	0.25	1.00

```
data a1;
input city$ x1-x12;
cards;
兰州 0 . . . . . .
敦煌 1140 0 . . . . . .
酒泉 735 405 0 . . . . . .
张掖 514 626 221 0 . . . . . .
武威 276 864 459 238 0 . . . . . .
(此处省略输入内容,见前文表1)
天水 327 1467 1062 841 603 385 223 442 254 367 0 .
陇南 450 1590 1185 964 726 540 385 362 536 649 282 0
;
run;
```

分析例 2 资料,设所需要的 SAS 数据步程序如下:

```
data a2;
input item $ x1-x8;
cards;
1 1.00 0.40 0.25 0.12 0.67 0.39 0.26 0.19
2 0.40 1.00 0.31 0.39 0.50 0.24 0.18 0.52
3 0.25 0.31 1.00 0.46 0.28 0.38 0.42 0.49
4 0.12 0.39 0.46 1.00 0.20 0.14 0.29 0.55
5 0.67 0.50 0.28 0.20 1.00 0.38 0.26 0.26
6 0.39 0.24 0.38 0.14 0.38 1.00 0.40 0.22
7 0.26 0.18 0.42 0.29 0.26 0.40 1.00 0.25
8 0.19 0.52 0.49 0.55 0.26 0.22 0.25 1.00
;
run;
```

3.2 用 SAS 实现统计分析

3.2.1 分析例 1 的资料

设所需 SAS 过程步程序如下[8]:

```
proc mds data=a1 level=absolute coef=identity fit=1 formula=1 pfinal;
id city;
run;
```

【SAS 程序说明】proc mds 语句表示调用 MDS 过程,该语句中的选项 level 决定了多维尺度分析是度量型还是非度量型。当 level 取值为 absolute、ratio、interval 或 loginterval 时,采用度量型多维尺度分析,不同的取值表示对距离进行不同的变换。本程序中,level=absolute 表示未对距离进行任何变换。选项 coef 决定采用加权多维尺度模型还是未加权模型,本资料中只有一个数据矩阵,采用未加权模型,此时 coef=identity。选项 fit 表示使用距离、距离的平

方、距离的 n 次幂还是距离的对数，本程序中，fit=1 表示使用距离本身。选项 formula 用来指定拟合劣度标准，也就是损失函数，本程序中，formula=1 规定使用未校正的平方和。选项 pfinal 规定输出模型中各项参数的最终估计值。

【SAS输出结果及解释】例 1 资料迭代计算的过程和结果见表 3。经过 4 次迭代，收敛标准得到满足，模型最终收敛。拟合劣度标准为 0.042<0.050，说明模型对资料的拟合效果较好。

表 3　基于模型拟合例 1 资料的迭代计算过程和结果

Table 3　Iterative calculation process and results based on the model fitting example 1 data

迭　代	类　型	拟合劣度准则	准则中的更改	收敛测度
0	Initial	0.054	-	0.608
1	Lev-Mar	0.042	0.012	0.098
2	Gau-New	0.042	0.000	0.028
3	Gau-New	0.042	0.000	0.011
4	Gau-New	0.042	0.000	0.005

基于模型算出的二维拟合构图中与各地对应的坐标见表 4。表 4 是各城市在二维空间中所对应的点的坐标，也就是拟合构图或感知图中各点的横坐标与纵坐标。这部分结果是由 pfinal 选项产生的，默认状态下它们不会被输出。

表 4　基于模型算出的二维拟合构图中 2 个坐标轴上的坐标

Table 4　Coordinates on two coordinate axes in the two-dimensional configuration calculated based on the model

地　区	dim1	dim2	地　区	dim1	dim2
兰州	-93.28	-28.56	定西	-193.50	-22.36
敦煌	1079.04	29.31	合作	-217.18	-320.26
酒泉	670.44	14.76	平凉	-371.80	224.35
张掖	446.02	4.41	庆阳	-501.52	337.04
武威	201.65	-11.43	天水	-418.87	-3.15
白银	-106.14	57.86	陇南	-494.86	-281.96

各城市在二维坐标平面内相对位置的拟合构图见图 1。图 1 中，甘肃省 12 个主要城市可分为四类：第一类包括庆阳、平凉；第二类包括天水、定西、白银、兰州、武威；第三类包括陇南、合作；第四类包括张掖、酒泉、敦煌。从数学角度来考量，图中各城市的位置是相对的，不一定是真实的地理位置。

反映模型对资料拟合效果的散布图见图 2。横坐标代表二维空间中两点（即两个城市）之间的距离，纵坐标代表原始数据。如果模型拟合效果较好，则所有的散点应该在一条直线上。本资料中绝大部分散点基本位于同一直线上，可认为模型的拟合效果较好。

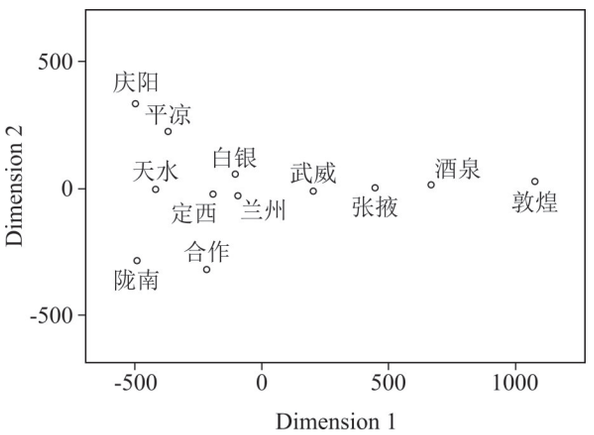

图 1　甘肃省 12 个主要城市的拟合构图

Figure 1　Configuration of 12 major cities in Gansu province

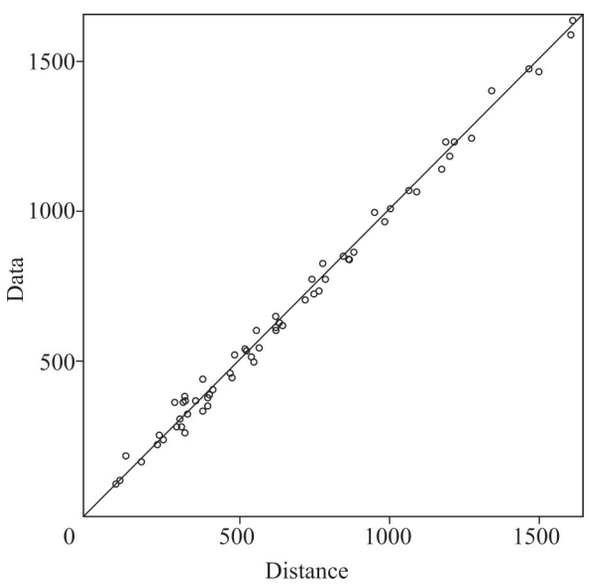

图 2　基于多维尺度分析模型对甘肃省 12 个主要城市拟合的散布图

Figure 2　Scatter diagram of fitting 12 major cities in Gansu province based on a multidimensional scale analysis model

3.2.2　分析例 2 的资料

设所需要的SAS过程步程序如下[8]：

```
proc mds data=a2 level=ratio pfinal similar;
id item;
run;
```

【SAS 程序说明】在 proc mds 过程步中，选项 level=ratio 表示对距离进行线性变换，也就是给距离乘以一个常数。选项 similar 指定本资料中的数据为相似性数据，此时 SAS 系统会将原始数据进行转换，用数据矩阵中的最大值减去每一个数据，将相似性数据转化为相异性数据。

【SAS 输出结果及解释】例 2 资料迭代计算的过程和结果见表 5。结果显示，经过 13 次迭代，收敛标准得到了满足，模型最终收敛。拟合劣度标

准为 0.082<0.100，说明模型对资料的拟合效果较好。

表 5 基于模型拟合例 2 资料的迭代计算过程和结果

Table 5 Iterative calculation process and results based on the model fitting example 2 data

迭 代	类 型	拟合劣度准则	准则中的更改	收敛测度
0	Initial	0.106	-	0.646
1	Lev-Mar	0.083	0.024	0.111
2	Gau-New	0.083	0.000	0.069
…	…	…	…	…
12	Gau-New	0.082	0.000	0.011
13	Gau-New	0.082	0.000	0.009

基于模型算出的二维拟合构图中与各地区对应的 2 个坐标轴上的坐标见表 6。

表 6 基于模型算出的二维拟合构图中 2 个坐标轴上的坐标

Table 6 Coordinates on two coordinate axes in the two-dimensional configuration calculated based on the model

项 目	dim1	dim2
1	1.41	0.64
2	-0.15	1.18
3	-0.75	-0.62
4	-1.68	0.13
5	1.13	0.71
6	1.19	-0.94
7	0.02	-1.63
8	-1.17	0.54

呈现各项目在二维坐标平面内相对位置的拟合构图见图 3。图 3 中，项目 1 与项目 5 最接近，项目 4 与项目 8 比较接近，其他 4 个项目彼此相差很大。

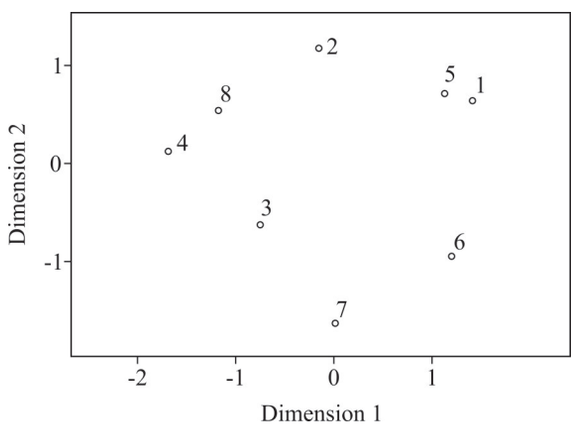

图 3 一项智力测验中 8 个项目的拟合构图

Figure 3 Configuration of 8 items in an intelligence test

反映模型对资料拟合效果的散布图见图 4。横坐标代表二维空间中两点（即两个项目）之间的距离，纵坐标代表原始数据。本资料中绝大部分散点基本位于同一直线上，可认为模型对资料的拟合效果较好。

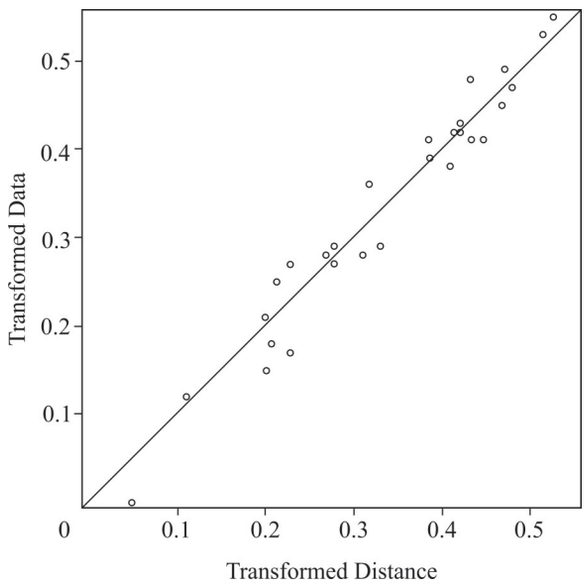

图 4 基于多维尺度分析模型对 8 个项目拟合的散布图

Figure 4 Scatter plot of fitting 8 items based on multidimensional scale analysis model

4 讨论与小结

4.1 讨论

4.1.1 相似性与相异性数据的转化

在进行多维尺度分析时，通常需要将相似性数据转化为相异性数据。用 d_{ij} 表示第 i 个对象与第 j 个对象之间的相异性，c_{ij} 表示两者之间的相似性，可通过以下几种形式完成从相似性到相异性的转化，见式（10）、式（11）、式（12）、式（13）。

$$d_{ij} = 1 - c_{ij} \quad (10)$$

$$d_{ij} = c - c_{ij}，\text{其中} c \text{为给定的常数} \quad (11)$$

$$d_{ij} = \sqrt{2(1 - c_{ij})} \quad (12)$$

$$d_{ij} = \sqrt{c_{ii} + c_{jj} - 2c_{ij}} \quad (13)$$

4.1.2 低维空间维数的确定

多维尺度分析的目标是以较少的维数空间较好地拟合获得的数据。通常维数越高，空间图对资料的拟合度越高，随着维数的减少，维度的实际意义将更容易归纳，但这是以损失部分原始数据信息为代价的。所以，选择合适的维数至关重要[6]。在古典多维尺度分析中，可根据特征根的累积贡献率来确定空间的维数。除此之外，还可以通过专业知识、文献资料或预试验的结果确定维数。一般来说，要想解释三维以上的空间图是很困难的，所以实际应用中空间的维数通常不会超过三维，其中使用最普遍的是二维空间。

4.1.3 多维尺度分析的解并不唯一

由于欧氏距离在正交变换和平移变换下具有不变性，多维尺度分析的解并不唯一。多维尺度分析的解在经过上述两种变换之后，仍然是它的解。并且，用多维尺度分析法构造出的图形结构，只能确定各对象或个体之间的相对位置，其绝对位置不能完全确定。多维尺度分析法的结果是试探性的，而不是结论性的[1]。

4.1.4 权重多维尺度分析

在实际应用中，有时需确定多个距离矩阵的拟合构图。例如，请20位啤酒品尝师分别对10种不同品牌的啤酒进行品尝，每位啤酒品尝师都能给出一个对10种啤酒的评价结果（10×10的距离矩阵），共有20个类似的距离矩阵。将它们整合在一起进行多维尺度分析的方法，被称为权重多维尺度分析法[3]。

4.1.5 多维尺度分析与其他分析方法的关系

多维尺度分析可以看成是因子分析的一种替代，目的是识别潜在的有意义的维度使得研究者能够解释被调查对象之间的相似性或相异性。在因子分析中，变量之间的相似性是用相关系数矩阵表示的，但在多维尺度分析里，研究者可以分析任何形式的相似矩阵或距离矩阵，包括相关系数矩阵[7]。

4.2 小结

本文介绍了与多维尺度分析有关的基本概念、计算方法、两个实例以及使用SAS实现计算的方法。基本概念包括度量型与多维尺度、应用场合、距离矩阵与相似系数矩阵、研究对象、拟合构图与构图、多维尺度分析的基本思想；计算方法涉及经典多维尺度分析和最小平方多维尺度分析；两个实例分别为“甘肃省12个主要城市间的距离”和“8个测试项目之间的相关系数矩阵”；借助SAS软件，对两个实例中的数据分别进行了多维尺度分析，并对SAS输出结果做出了解释。

参考文献

[1] 余锦华，杨维权. 多元统计分析与应用[M]. 广州：中山大学出版社，2005：232-250.
Yu JH，Yang WQ. Multivariate statistical analysis and application [M]. Guangzhou：Sun Yat-sen University Press，2005：232-250.

[2] 张润楚. 多元统计分析[M]. 北京：科学出版社，2006：288-311.
Zhang RC. Multivariate statistical analysis[M]. Beijing：Science Press，2006：288-311.

[3] 李卫东. 应用多元统计分析[M]. 北京：北京大学出版社，2008：239-258.
Li WD. Applied multivariate statistical analysis [M]. Beijing：Peking University Press，2008：239-258.

[4] Johnson RA，Wichern DW. 实用多元统计分析[M]. 6版. 北京：清华大学出版社，2008：671-756.
Johnson RA，Wichern DW. Applied multivariate statistical analysis [M]. 6th edition. Beijing：Tsinghua University Press，2008：671-756.

[5] 何晓群. 多元统计分析[M]. 2版. 北京：中国人民大学出版社，2008：227-254.
He XQ. Multivariate statistical analysis[M]. 2nd edition. Beijing：China Renmin University Press，2008：227-254.

[6] Armitage P，Colton T. Encyclopedia of biostatistics [M]. 2nd edition. New York：John Wiley & Sons，Inc，2005：3635-3643.

[7] 胡良平. 面向问题的统计学：(3)试验设计与多元统计分析[M]. 北京：人民卫生出版社，2012：303-317.
Hu LP. Problem-oriented statistics：(3) experimental design and multivariate statistical analysis [M]. Beijing：People′s Mental Publishing House，2012：303-317.

[8] SAS Institute Inc. SAS/STAT®15.1 user′s guide[M]. Cary，NC：SAS Institute Inc，2018：2514-2578，2997-3216.

（收稿日期：2023-07-26）
（本文编辑：陈 霞）