

IBM SPSS Categories 23

IBM

注释

使用本信息及其支持的产品之前，请阅读 第 43 页的『声明』中的信息。

产品信息

此版本适用于 IBM SPSS Statistics V23R0M0 及所有后续发行版和修改版，除非在新版本中另有说明。

目录

第 1 章 分类数据的最优刻度过程介绍 . . .	1
什么是最优刻度? . . .	1
为什么要使用最优刻度? . . .	1
最佳度量水平和测量级别 . . .	1
选择最佳度量水平. . .	2
转换图 . . .	2
类别代码. . .	2
哪种过程最适合于您的应用? . . .	4
分类回归. . .	4
分类主成份分析 . . .	5
非线性正态协变量分析 . . .	5
对应分析. . .	6
多重对应分析 . . .	6
多维刻度. . .	7
多维展开. . .	7
最优刻度图表中的宽高比 . . .	7
第 2 章 分类回归 (CATREG) . . .	9
在“分类回归”中定义刻度 . . .	10
分类回归: 离散 . . .	10
分类回归: 缺失值 . . .	10
分类回归: 选项 . . .	11
分类回归: 规则化 . . .	11
分类回归: 输出 . . .	12
分类回归: 保存 . . .	12
分类回归: 转换图 . . .	12
CATREG 命令附加功能 . . .	13
第 3 章 分类主成份分析 (CATPCA) . . .	15
在 CATPCA 中定义度量和权重 . . .	16
分类主成份分析: 分箱化 . . .	16
分类主成份分析: 缺失值 . . .	16
分类主成份分析: 选项 . . .	17
分类主成份分析: 输出 . . .	18
分类主成份分析: 保存 . . .	18
分类主成份分析: 对象图 . . .	19
分类主成份分析: 类别图 . . .	19
分类主成份分析: 载荷图 . . .	19
分类主成份分析自助抽样 . . .	19
CATPCA 命令附加功能 . . .	20
第 4 章 非线性正态协变量分析 (OVERALS) . . .	21
定义范围和刻度 . . .	22
定义范围 . . .	22
非线性正态协变量分析: 选项. . .	22
OVERALS 命令附加功能 . . .	23

第 5 章 对应分析 . . .	25
在“对应分析”中定义行范围 . . .	26
在“对应分析”中定义列范围 . . .	26
对应分析: 模型 . . .	26
对应分析: 统计 . . .	27
对应分析: 图 . . .	27
CORRESPONDENCE 命令附加功能 . . .	28
第 6 章 多重对应分析 . . .	29
在“多重对应分析”中定义变量权重 . . .	29
多重对应分析: 分箱化 . . .	30
多重对应分析: 缺失值 . . .	30
多重对应分析: 选项. . .	30
多重对应分析: 输出. . .	31
多重对应分析: 保存. . .	31
多重对应分析: 对象图 . . .	32
多重对应分析: 变量图 . . .	32
MULTIPLE CORRESPONDENCE 命令附加功能 . . .	32
第 7 章 多维刻度 (PROXSCAL) . . .	33
矩阵中的跨列近似值 . . .	34
多列中的近似值 . . .	34
单列中的近似值 . . .	34
从数据中创建近似值 . . .	34
从数据中创建测量 . . .	35
定义多维刻度模型 . . .	35
多维刻度: 约束 . . .	36
多维刻度: 选项 . . .	36
多维刻度: 图, 版本 1 . . .	37
多维刻度: 图, 版本 2 . . .	37
多维刻度: 输出 . . .	37
PROXSCAL 命令附加功能 . . .	38
第 8 章 多维展开 (PREFSCAL) . . .	39
定义多维展开模型 . . .	39
多维展开: 约束 . . .	40
多维展开: 选项 . . .	40
多维展开: 图 . . .	41
多维展开: 输出 . . .	42
PREFSCAL 命令附加功能. . .	42
声明 . . .	43
商标 . . .	44
索引 . . .	47

第 1 章 分类数据的最优刻度过程介绍

类别过程使用最优刻度来分析标准统计过程难以分析或无法分析的数据。本章描述每种过程所执行的工作、每种过程最适合的情形、过程之间的关系，以及这些过程与其对应的标准统计过程的关系。

注：这些过程及其 IBM® SPSS® Statistics 实现是由 Data Theory Scaling System Group (DTSS) 开发的，该小组由莱顿大学 (Leiden University) 社会和行为科学分院的教育和心理学系的成员组成。

什么是最优刻度？

最优刻度的基本理念是将数值定量分配给每个变量的类别，从而允许将标准过程用于获取已量化变量的解。

最优刻度值根据所用过程的优化标准分配给每个变量的类别。与分析中的名义变量或有序变量的原始标签不同，这些刻度值具有度量属性。

在大多数类别过程中，每个刻度化变量的优化定量都是通过称为**交替最小平方**的迭代方法获取的，在这种方法中，在将当前定量用于找到一个解之后，该定量会使用该解进行更新。更新后的定量接着用于查找新解，新解再用于更新定量，依此类推，直至达到通知过程停止的某个标准。

为什么要使用最优刻度？

分类数据通常可以在市场研究、民意调查研究以及社会和行为科学研究中找到。事实上，很多研究人员几乎专门处理分类数据。

虽然大多数标准模型的改编版本都专用于分析分类数据，但是通常，对于具有以下特点的数据集，它们的效果并不好：

- 观察值过少
- 变量过多
- 每个变量的值过多

通过量化类别，最优刻度方法避免了上述情况下的问题。此外，即使在专用方法适用的情况下，它们同样很有用。

对最优刻度输出的解释通常基于图形显示，而不基于解释参数估计值。最优刻度方法提供了极佳的探索性分析，它很好地补充了其他 IBM SPSS Statistics 模型。通过缩小调查重点的范围，利用最优刻度使数据可视化可构成以解释模型参数为中心的分析的基础。

最佳度量水平和测量级别

在首次使用类别过程时，这可能是非常令人困惑的概念。在指定级别时，您指定的不是变量被度量的级别，而是它们被**调整**的级别。其理念是，无论变量是如何度量的，要量化的变量都可能具有非线性关系。

对于类别而言，存在三种基本测量级别：

- **标定级别**表示变量的值代表无序类别。可能为名义变量的示例有地区、邮政编码区域、宗教信仰以及多选类别。
- **有序级别**表示变量的值代表有序类别。示例包括代表满意度或信心以及首选评级得分的态度刻度。

- **数值级别**表示变量的值代表具有有意义度规的有序类别，这样类别之间适合比较距离。示例包括以年为单位的年龄和以千美元为单位的收入。

例如，假定像下表这样对地区、工作和年龄这几个变量进行编码。

表 1. 地区、作业和年龄的编码方案

区域代码	区域值	作业代码	作业值	年龄
1	北部	1	实习生	20
2	南部	2	销售代表	22
3	东部	3	经理人	25
4	西部			27

所示值代表了每个变量的类别。地区为名义变量。地区有四个类别，没有内在顺序。值 1 到 4 仅表示四种类别；编码方案完全是任意的。而工作可视为有序变量。原始类别构成了从实习生到经理人的晋级过程。较大的代码代表公司体制中职位较高的工作。但是，只有顺序信息是已知的，这意味着相邻类别之间谈不上有距离。相反，年龄可视为数值变量。对于年龄，值之间的距离是有内在意义的。20 与 22 之间的距离与 25 与 27 之间的距离相同，而 22 和 25 之间的距离大于前面的两个距离。

选择最佳度量水平

变量中没有任何内在属性自动预定义了您应为变量指定哪种最佳度量水平，理解这一点非常重要。您可以通过有意义的并且可使解释更简单的任何方法研究数据。例如，通过在有序级别分析数值级变量，使用非线性转换可能在更小的维数中得出解。

下面的两个示例演示了为什么“显而易见”的测量级别可能不是最佳的最佳度量水平。假设某个变量将对象分类为年龄组。虽然年龄可调整为数值变量，但是实际情况可能是对于 25 岁以下的人，安全性与年龄成正向关系，而对于 60 岁以上的人，安全性与年龄成负向关系。在此情况下，最好将年龄视为名义变量。

另一个示例是，根据政治偏好对人进行分类的变量似乎本质上是标定的。但是，如果按从政治左派到政治右派的顺序对政党进行排序，则可能要使用有序分析级别对政党进行量化以符合此顺序。

尽管变量没有预定义的属性可使其专属哪个级别，还是存在一些可以帮助新用户的通用准则。对于单名义量化，您通常不知道类别的顺序，但是您希望分析过程施加一个顺序。如果类别的顺序已知，则应该尝试有序量化。如果类别不可排序，则可以尝试多名义量化。

转换图

每个变量可调整到的不同级别对量化施加了不同的约束。转换图说明了从选定的最佳度量水平得出的量化和原始类别之间的关系。例如，当将变量视为数值时，将产生线性转换图。视为有序的变量将产生非递减转换图。视为标定的 U 形（或反转的 U 形）变量的转换图显示二次关系。通过完全改变类别的顺序，名义变量还可以生成没有明显趋势的转换图。下图显示了样本转换图。

转换图特别适合于确定选定的最佳度量水平的效果。如果多个类别得到了相似的量化，则确保可以将这些类别合并为一个类别。或者，如果视为标定的变量得到了显示递增趋势的量化，则有序转换可能生成相似的拟合。如果该趋势是线性的，则可能数值处理是合适的。但是，如果能确保合并类别或更改刻度级别，则分析不会有显著变化。

类别代码

当为分类变量进行编码时，应多加小心，因为一些编码方案可能生成不理想的输出或不完整的分析。下表中显示了作业的可行编码方案。

表 2. 作业的替代编码方案

类别(Y)	A	B	C	D
实习生	1	1	5	1
销售代表	2	2	6	5
经理人	3	7	7	3

某些类别过程要求定义每个所用变量的范围。此范围之外的任何值都视为缺失值。最小类别值总是 1。最大类别值由用户提供。此值不是变量类别的数目—它是最大类别值。例如，在表中，方案 A 的最大类别值为 3，方案 B 的最大类别值为 7，两种方案都对相同的三个类别进行编码。

变量范围确定了分析中将忽略哪些类别。分析中将忽略代码在定义范围之外的任何类别。这是忽略类别的简单方法，但是可能会产生不理想的分析结果。未正确定义的最大类别可能从分析中忽略有效类别。例如，对于方案 B，将最大类别值定义为 3 表示工作有编码为 1 到 3 的类别；经理人类别被视为缺失。因为实际上没有类别编码为 3，所以分析中的第三类别不包含个案。如果您本来就想忽略所有经理人类别，则此分析可能是合适的。但是，如果应该将经理人包括在内，则最大类别必须定义为 7，并且缺失值必须以大于 7 或小于 1 的值进行编码。

对于视为标定或名义的变量，类别范围不影响结果。对于名义变量，只有标签是重要的，而与该标签关联的值不重要。对于有序变量，通过量化保持类别的顺序；类别值本身不重要。生成相同类别排序的所有编码方案将得到相同的结果。例如，如果以有序级别分析工作，则表中前三个方案的作用是等效的。这三个方案的类别顺序是相同的。而方案 D 颠倒了第二类别和第三类别，将产生与其他方案不同的结果。

虽然一个变量的很多编码方案的作用相同，但是代码之间差值较小的方案为首选方案，因为代码影响过程产生的输出量。用 1 和用户定义的最大值之间的值编码的所有类别都是有效的。如果这些类别中有任何类别为空，则根据过程的不同，对应的定量将为系统缺失值或 0。虽然这两个赋值都不会影响分析，但是会为这些类别生成输出。因此，对于方案 B，工作有四个得到系统缺失值的类别。对于方案 C，也有四个得到系统缺失指示符的类别。相比之下，方案 A 没有系统缺失值定量。对视为标定或名义的变量使用连续整数作为代码将产生更少的输出，同时又不影响结果。

对视为数值的变量的编码方案比有序变量的编码方案更受限制。对于这些变量，连续类别之间的差值非常重要。下表显示了年龄的三种编码方案。

表 3. 年龄的替代编码方案

类别(Y)	A	B	C
20	20	1	1
22	22	3	2
25	25	6	3
27	27	8	4

对数值变量的任何重新编码都必须保持类别之间的差值。使用原始值是确保保留差值的一种方法。但是，这可能产生很多有系统缺失指示符的类别。例如，方案 A 采用原始观察值。对于除“对应分析”之外的所有类别过程，最大类别值为 27，最小类别值设置为 1。前 19 个类别为空，并得到系统缺失指示符。如果最大类别远大于 1，则输出可能会很快变得相当繁琐，并且在 1 和最大值之间会有很多空类别。

为了减少输出的量，可以重新编码。但是，对于数值个案，不应使用“自动重新编码”工具。编码为连续整数将生成所有相邻类别之间的差值（值为 1），因此，所有定量都成为等间距的。在将变量视为数值时被视作重要的度规特征将因为重新编码为连续整数而受到破坏。例如，表中的方案 C 对应于自动重新编码年龄。类别 22 和 25 之间的差值从三更改为一，这种定量将影响后面的差值。

保持类别之间的差值的另一种重新编码方案是从每个类别减去最小类别值，然后在每个差上加 1。方案 B 出自于这种转换。从每个类别中减去最小类别值 20，然后对每个结果加 1。转换后的代码最小为 1，所有差值与原始数据的差值相同。最大类别值现在为 8，第一个非零定量之前的零定量已全部排除。但是，从方案 B 得出的对应于每个类别的非零定量与从方案 A 得出的定量相同。

哪种过程最适合于您的应用？

这些过程中的四个过程（对应分析、多重对应分析、分类主成份分析和非线性正态协变量分析）所使用的方法属于称为降维的多变量数据分析的一般性领域。也就是说，变量间的关系尽可能以少量维数（如，二或三）表示。这使您能够描述其原始丰富性和复杂性中难以理解的关系中的结构或模式。在市场研究应用中，这些方法可以是知觉图的形式。这些过程的主要优点是它们容纳具有不同最佳度量水平的数据。

分类回归描述分类响应变量和分类预测变量组合之间的关系。每个预测变量对响应变量的影响由相应的回归权重描述。与其他过程一样，可使用不同级别的最优刻度分析数据。

多维刻度和多维展开使用对象间的近似性描述低维空间中对象间的关系。

以下是每个过程的简要准则：

- 使用“分类回归”从分类自变量的组合中预测分类因变量的值。
- 使用“分类主成份分析”解释混合最佳度量水平的单个变量集中的变动模式。
- 使用“非线性正态协变量分析”评估两个或多个混合最佳度量水平的变量集的相关程度。
- 使用“对应分析”分析双向列联表或可表示为双向表的数据，例如品牌偏好或社会测量选择数据。
- 当您不想更强烈地假设以标定级别分析所有变量时，应使用“多重对应分析”分析分类多变量数据矩阵。
- 使用“多维刻度”分析近似值数据以查找低维空间中单个对象集的最小平方表示形式。
- 使用“多维展开”分析近似值数据以查找低维空间中两个对象集的最小平方表示形式。

分类回归

在分析目标为从一组自变量（预测变量）预测一个因变量（响应）的情况下，最适合使用分类回归。与所有最优刻度过程一样，刻度值分配给每个变量的每个类别，以使这些值对于回归是最优的。分类回归的解使已转换的响应与加权的已转换预测变量组合之间的平方相关性最大化。

与其他类别过程的关系。使用最优刻度的分类回归类似于具有两个集（其中一个集仅包含因变量）的最优刻度正态协变量分析。在后一种方法中，集的相似性是通过将每个集与位于所有集之间的某个未知变量进行比较而得出的。在分类回归中，已转换的响应与已转换的预测变量的线形组合的相似性是直接评定的。

与标准方法的关系。在标准线性回归中，分类变量或者可以重新编码为指示符变量，或者以同样的方式视为区间级变量。在第一种方法中，该模型对于分类变量级别的每个组合都包含一个独立的截距和斜率。这产生了大量要解释的参数。在第二种方法中，对于每个变量只估计一个参数。但是，由于类别编码的任意性，不可能进行广义化。

如果某些变量不是连续的，则替代分析可用。如果响应是连续的并且预测变量是分类的，则通常采用方差分析。如果响应是分类的，而预测变量是连续的，则可能适用 Logistic 回归或判别分析。如果响应和预测变量都是分类的，则通常使用对数线形模型。

使用最优刻度的回归为每个变量提供三个刻度级别。这些级别的组合可解释任何单个“标准”方法都不适用的大量非线性关系。因此，最优刻度比具有最小附加复杂性的标准方法提供的灵活性更大。

此外，预测变量的非线性转换通常降低了预测变量之间的依赖性。如果将预测变量的相关性矩阵的特征值与最优刻度预测变量的相关性矩阵的特征值进行比较，则后者集合的可变性通常比前者更低。换言之，在分类回归中，最优刻度使预测变量相关性矩阵的较大的特征值较小，而较小的特征值较大。

分类主成份分析

当您想解释混合最佳度量水平的单个变量集中的变化模式时，最适宜使用分类主成份分析。此项方法尝试在尽可能多地考虑到变化的同时减少变量集的维数。刻度值分配给每个变量的每个类别，以使这些值对于主成分解是最优的。分析中的对象根据量化的数据得到成分得分。成分得分图揭示了分析中对象间的模式，并且可以揭示数据中的异常对象。分类主成份分析的解使对象得分与指定的成分的数量（维数）的每个量化变量的相关性最大化。

分类主成份的一个重要应用是检查偏好数据，在此应用中，响应者根据其偏好对一定数量的项进行排序或评级。在常规 IBM SPSS Statistics 数据配置中，行为个人，列为项的测量值，行中的得分是偏好得分（例如，0 到 10 的范围），使数据依行而定。对于偏好倾向数据，可以将个人视为变量。使用变换过程可变换数据。评分者成为变量，并且所有变量声明为有序。允许使用比 CATPCA 中的对象更多的变量。

与其他类别过程的关系。如果所有变量都声明为多名义，则分类主成份分析将生成等效于对相同变量运行的多重对应分析的分析。因此，分类主成份分析可视为多重对应分析的一种类型，在此类型中，某些变量声明为有序或数值。

与标准方法的关系。如果所有变量按数值级别调整，则分类主成份分析等效于标准主成分分析。

更广义地说，分类主成份分析是计算非数值刻度之间的相关性，并使用标准主成分或因子分析方法分析这些相关性的一种替代方法。仅将常规 Pearson 相关性系数用作有序数据的相关性测量可能导致估计相关性时出现严重偏差。

非线性正态协变量分析

非线性正态协变量分析是具有多种不同应用的常规过程。非线性正态协变量分析的目标是分析两个或多个变量集之间的关系，而不似主成分分析那样，分析变量本身之间的关系。例如，您可能有两个变量集，其中一个变量集可能是一组响应者的人口统计学背景项，另一组变量集可能是对一组态度项的响应。分析中的刻度级别可以是名义、有序和数值的任何组合。最优刻度正态协变量分析通过同时将每个集中的典型变量与分配给对象的一组折算得分进行比较来确定集之间的相似性。

与其他类别过程的关系。如果有两个或多个变量集，并且每个集只有一个变量，则最优刻度正态协变量分析等效于最优刻度主成分分析。如果“每集单变量”分析中的所有变量为多名义，则最优刻度正态协变量分析等效于多重对应分析。如果有两个变量集，其中一个变量集只包含一个变量，则最优刻度正态协变量分析等效于使用最优刻度的分类回归。

与标准方法的关系。标准正态协变量分析是一种统计方法，它查找相关性最大的一个变量集的线性组合和另一个变量集的线性组合。对于这组线性组合，正态协变量分析可查找后面的独立线性组合集（称为典型变量），直到最大数量等于较小集中的变量数。

如果分析中有两组变量集并且所有变量都定义为数值，则最优刻度正态协变量分析等效于标准正态协变量分析。虽然 IBM SPSS Statistics 没有正态协变量分析过程，但是可以从多变量方差分析中获得许多相关统计。

最优刻度正态协变量分析有多种其他应用。如果有两个变量集，并且其中一个集包含声明为单名义的名义变量，则可以用类似于解释回归分析结果的方式解释最优刻度正态协变量分析结果。如果将变量视为多名义，则最优刻度分析是判别分析的替代方法。将变量分组为两个以上的集合将提供多种分析数据的方法。

对应分析

对应分析的目标是生成对应表的双标图。在对应表中，行和列变量假设为表示无序类别；因此，将始终使用标定最佳度量水平。检查这两种变量仅仅是为了获得其标定信息。也就是说，唯一需要考虑的是某些对象属于同一个类别而其他对象不属于。对于同一变量的类别之间的距离或顺序不作任何假设。

对应分析的一种特定用途是分析二阶列联表。如果表有 r 个活动行和 c 个活动列，则对应分析解中的维数最小为 r 减 1 或 c 减 1，以较小者为准。换言之，可在维度空间中完好地表示列联表的行类别或列类别。但是实际上，您希望在低维空间（如二维）中表示双向表的行和列类别，因为二维图比多维空间表示形式更容易理解。

当使用的维数小于可能的最大维数时，分析中生成的统计将描述行和列类别在低维表示形式中的表示效果。如果二维解表示形式的质量良好，则可以查看行点和列点图，以了解哪些行变量的类别相似，哪些列变量的类别相似，以及哪些行类别与哪些列类别相似。

与其他类别过程的关系。简单对应分析限于双向表。如果有两个以上需要处理的变量，则可以组合变量以创建交互变量。例如，对于变量地区、工作和年龄，可组合地区和工作以创建具有 12 个类别的新变量地区工作，如下表所示。此新变量与年龄一起构成了一个双向表（12 行，4 列），可在对应分析中分析此表。

表 4. 地区和作业的组合

类别代码	类别定义	类别代码	类别定义
1	北部, 实习生	7	东部, 实习生
2	北部, 销售代表	8	东部, 销售代表
3	北部, 经理人	9	东部, 经理人
4	南部, 实习生	10	西部, 实习生
5	南部, 销售代表	11	西部, 销售代表
6	南部, 经理人	12	西部, 经理人

此方法的一个缺点是任意一对变量都可以组合。我们可组合工作和年龄，产生另一个 12 个类别的变量。或者，可以组合地区和年龄，生成一个新的 16 个类别的变量。这些交互变量的每一个都与其余的变量构成一个双向表。对这三个表的对应分析不会生成相同的结果，但每次分析都是有效的。此外，如果有四个或更多变量，则可以构建将一个交互变量与另一个交互变量进行比较的双向表。要分析的可能存在的表的数量可以变得很大，即使只有几个表。可以选择其中一个表进行分析，或者可分析所有表。或者，可以使用多重对应分析过程同时检查所有变量，无需构建交互变量。

与标准方法的关系。在独立性作为分析中共同焦点的情况下，交叉表格过程也可用于分析列联表。但是，即使在小型表中，检测背离独立性的原因可能也很困难。对应分析的作用在于显示任何规模的双向表的此类模式。如果行和列变量之间存在关联（即如果卡方值很大），则对应分析可能有助于揭示关系的本质。

多重对应分析

多重对应分析尝试生成一个解，该解中同类别的对象绘制为紧靠在一起，而不同类别的对象绘制为远远分开。每个对象都尽可能地接近应用于该对象的类别的类别点。这样，类别可将对象分成数个均一子组。当变量将相同类别中的对象分类到同一子组中时，这些变量被视为是均一的。

对于单维解，多重对应分析将最优刻度值（类别定量）分配给每个变量的每个类别，以使总体上，平均而言，类别有最大分布。对于二维解决方案，多重对应分析查找与每个变量的类别的第一组定量无关的第二组定量，并再次尝试最大化分布，如此一直继续。因为变量的类别得到与维数一样多的得分，所以分析中的变量被假设为最佳度量水平的多名义。

多重对应分析还为分析中的对象分配得分，使类别量化为该类别中的对象的对象得分平均值，即质心。

与其他类别过程的关系。多重对应分析也称为同质性分析或双重刻度。当只有两个变量时，它提供了与对应分析类似但不相同的结果。对应分析生成了总结解的拟合与表示形式的质量的唯一输出，包括稳定性信息。因此，对应分析通常更适合于双变量情况下的多重对应分析。两种过程的另一区别是，多重对应分析的输入是数据矩阵，其中行为对象，列为变量；而对应分析的输入可以是相同的数据矩阵、常规近似值矩阵，或者联合列联表，后者是行和列都表示变量的类别的汇总矩阵。还可以将多重对应分析看作是在多名义级别调整的数据的主成分分析。

与标准方法的关系。可将多重对应分析视为多阶列联表的分析。多阶列联表也可使用交叉表格过程进行分析，但是交叉表格提供了每个控制变量的每个类别的单独的汇总统计。使用多重对应分析时，通常可通过单个二维图总结所有变量之间的关系。多重对应分析的一种高级用法是用第一维的最优刻度值替换原始类别值，并执行二次多变量分析。由于多重对应分析用数值刻度值替换了类别标签，因此需要数值数据的很多不同过程可在多重对应分析之后应用。例如，因子分析过程生成等效于多重对应分析的第一维的主成分。第一维的成分得分等于对象得分，并且平方成份载荷等于区分测量。但是，第二个多重对应分析维度不等于因子分析的第二维。

多维刻度

当分析目标是查找单组对象或个案之间的一组距离测量中的结构时，最适合使用多维刻度。这是通过将观察值指定给概念低维空间中的特定位置实现的，这样使空间中的点之间的距离尽可能与给定的（不）相似性相匹配。结果是一个该低维空间中的对象的最小平方表示形式，这在许多个案中将帮助您进一步理解您的数据。

与其他类别过程的关系。当您有用来创建距离的多变量数据，并且随后用多维刻度分析这些数据时，其结果类似于使用具有主要对象标准化的“分类主成份分析”分析数据。这类 PCA 也称为主坐标分析。

与标准方法的关系。类别多维刻度过程 (PROXSCAL) 提供了对 Statistics Base 选项中可用的刻度过程的多种改进 (ALSCAL)。PROXSCAL 提供了对某些模型的加速算法，可以对公共空间施加约束。此外，PROXSCAL 尝试使标准化初始强度而不是 S-stress（也称为张力）达到最小。标准化初始强度通常是首选的，因为它是基于距离测量的，而 S-stress 是基于平方距离。

多维展开

当分析目标是查找两组对象（指行对象和列对象）之间的一组距离测量值中的结构时，最适合使用多维展开。这是通过将观察值指定给概念低维空间中的特定位置实现的，这样使空间中的点之间的距离尽可能与给定的（不）相似性相匹配。结果是一个该低维空间中的行和列对象的最小平方表示形式，这在许多情况下可帮助您进一步理解您的数据。

与其他类别过程的关系。如果数据由单组对象（正方形对称矩阵）之间的距离构成，则使用多维刻度。

与标准方法的关系。类别多维展开过程 (PREFSCAL) 提供了对 Statistics Base 选项中可用的展开功能的多种改进（通过 ALSCAL）。使用 PREFSCAL 可以对公共空间施加约束；此外，PREFSCAL 尝试使惩罚强度度量最小化，这有助于避免退化解（容易对其使用较老的算法）。

最优刻度图表中的宽高比

最优刻度图中的宽高比是等方性的。在二维图中，表示维度 1 中的一个单位的距离等于表示维度 2 中的一个单位的距离。如果更改二维图中某个维的范围，则系统将更改另一维的大小，以保持物理距离相等。对于优化刻度过程，不可覆盖等方宽高比。

第 2 章 分类回归 (CATREG)

分类回归通过为类别指定数值来量化分类数据，从而生成转换后变量的最优线性回归方程。分类回归也用缩写词 CATREG 来表示（代表 *categorical regression*）。

标准线性回归分析涉及使响应变量（因变量）和预测变量（自变量）的加权组合之间的平方差之和达到最小。变量通常是定量的，（名义）分类数据重新编码为二元变量或对比变量。因此，分类变量用于分离个案组，并且该技术估计每个组的独立的参数集。估计的系数反映了预测变量的变化对响应的影响程度。对于预测变量值的任何组合都可以预测响应。

另一种方法需要对分类预测变量值本身进行响应回归。这样，将为每个变量分别估计一个系数。但是，对于分类变量，类别值是任意的。以不同的方式编码类别将产生不同的系数，这样，在对同样的几个变量的分析进行比较时，难度就增大了。

CATREG 通过同时调整名义、序数和数值变量扩展了标准方法。该过程量化分类变量以使量化反映初始类别的特征。该过程以与处理数值变量相同的方式处理量化的分类变量。使用非线性转换允许在各种级别分析变量以查找最佳拟合模型。

示例。 分类回归可用于描述工作满意度对工作类别、地理区域和旅行量的依赖程度。您可能会发现高满意度对应于经理和低旅行量。生成的回归方程可用于针对三个自变量的任何组合预测工作满意度。

统计和图。 频率、回归系数、ANOVA 表、迭代历史记录、类别量化、未转换的预测变量之间的相关性、转换后的预测变量之间的相关性、残差图和转换图。

分类回归数据注意事项

数据。 CATREG 在类别指示符变量上运行。类别指示符应为正整数。可使用“分箱化”对话框将小数值变量和字符串变量转换为正整数。

假设。 只允许一个响应变量，但是预测变量的最大数目为 200。该数据必须至少包含三个有效个案，并且有效个案数必须大于预测变量数加一。

相关过程。 CATREG 等效于使用最优刻度的分类正态协变量分析 (OVERALS)，该分析有两个变量集，其中一个只包含一个变量。将所有变量调整为数值级别对应于标准多重回归分析。

获取分类回归

1. 从菜单中选择:

 分析 > 回归 > 最优刻度 (CATREG)...

2. 选择因变量和自变量。

3. 单击**确定**。

或者，更改每个变量的刻度级别。

在“分类回归”中定义刻度

可设置因变量和自变量的最佳度量水平。缺省情况下，它们调整为带两个内部结点的二度单调性样条（有序）。此外，还可以设置分析变量的权重。

最佳度量水平。 还可以选择用于量化每个变量的刻度级别。

- **有序样条。** 已观察到的变量的类别顺序保存在最优刻度化变量中。类别点将位于一条通过原点的直线上（向量）。生成的转换是一个选定次数的平滑单调的分段多项式。这些段是由内部结点的用户指定的数量和过程决定的位置来指定的。
- **名义样条。** 已观察到的变量中保留在最优刻度化变量中的唯一信息是类别中对象的分组。已观察到的变量的类别的顺序不会保留。类别点将位于一条通过原点的直线上（向量）。生成的转换是一个选定次数的平滑的、可能非单调的分段多项式。这些段是由内部结点的用户指定的数量和过程决定的位置来指定的。
- **有序。** 已观察到的变量的类别顺序保存在最优刻度化变量中。类别点将位于一条通过原点的直线上（向量）。生成的转换比有序样条转换拟合得好，但是平滑度较低。
- **名义。** 已观察到的变量中保留在最优刻度化变量中的唯一信息是类别中对象的分组。已观察到的变量的类别的顺序不会保留。类别点将位于一条通过原点的直线上（向量）。生成的转换比样条名义转换拟合得更好，但是平滑度较低。
- **数值。** 类别被视为是有序的并且是等间距的（区间度）。已观察到的变量的类别顺序以及类别号之间的等间距保留在最优刻度化变量中。类别点将位于一条通过原点的直线上（向量）。当所有变量都为数值级别时，分析与标准主成分分析类似。

分类回归：离散

使用“分箱化”对话框可以选择对变量重新编码的方法。除非另有指定，否则小数值变量将分组成具有近似正态分布的七个类别（如果变量的不同值的数目小于七，则将按此数目划分类别）。通过按照升序字母数值顺序分配类别指示符，字符串变量总是转换为正整数。字符串变量的分箱化适用于这些整数。缺省情况下，其他变量保留原样。随后，分析中将使用离散化变量。

方法。 在分组、等级和乘之间选择。

- **分组。** 重新编码为指定数量的类别或者按区间重新编码。
- **等级。** 通过对个案等级排序来对变量进行离散化。
- **乘。** 变量的当前值是乘以 10 且经过四舍五入的离散化值，并且加上了一个常数以使最低离散值为 1。

分组。 在按分组分箱化变量时，可以使用以下选项：

- **类别数。** 指定一定数量的类别，并指定变量的值是否应该遵循在这些类别间的近似正态分布或均匀分布。
- **相等区间。** 变量重新编码到由这些等大小的间隔所定义类别中。必须指定间隔的长度。

分类回归：缺失值

使用“缺失值”对话框可以选择处理分析变量和补充变量中的缺失值的方案。

方案。 选择排除带有缺失值的对象（成列删除）或插补缺失值（积极处理）。

- **排除此变量具有缺失值的对象。** 选定的变量有缺失值的对象将从分析中排除。此方案不适用于补充变量。
- **插补缺失值。** 选定的变量有缺失值的对象将对这些值插补。您可以选择插补方法。选择**众数**将缺失值替换为最频繁的类别。当有多个众数时，将使用具有最小类别指示符的众数。选择**附加类别**将缺失值替换为相同的附加类别定量。这意味着此变量有缺失值的对象被视为属于同一（附加）类别。

分类回归: 选项

“选项”对话框允许您选择初始配置样式，指定迭代和收敛性条件，选择补充对象和设置绘图标记。

补充对象。使用此项可以指定您要视作补充对象的对象。只需键入补充对象的编号（或指定个案范围），然后单击**添加**。不可对补充变量加权（指定的权重将被忽略）。

初始配置。如果没有变量被视作名义变量，则选择**数值配置**。如果至少有一个变量被视为名义，则选择**随机配置**。

或者，如果至少一个变量具有有序或有序样条刻度级别，则通常的模拟拟合算法可能导致欠佳解。选择具有所有可能的待检验符号模式的**多系统性起点**，可以始终寻找最优解，但由于数据集中的有序和有序样条变量数量增加，因此所需的处理时间也快速增加。可以指定方差损失阈值百分比以减少检验模式数量，即阈值越高，排除的符号模式越多。采用此选项，尽管无法保证获得最优解，但也消除了得到欠佳解的可能性。此外，如果找不到最优解，欠佳解与最优解也不会差别太大。在要求多系统性起点时，每个起点的回归系数符号被写入外部 IBM SPSS Statistics 数据文件或当前会话中的数据集中。请参阅第 12 页的『分类回归: 保存』主题以获取更多信息。

先前采用多系统性起点的运行结果允许您为**回归系数使用固定符号**。符号（采用 1 和 -1 指示）需要在指定数据集或文件的某行中。整数值的起始数字为包含待使用符号的文件对应行的个案号。

标准。可以指定回归可在其计算中执行的最大迭代次数。还可以选择收敛性标准值。如果上两次迭代之间的总拟合之差小于收敛性值，或者达到了最大迭代次数，则回归停止迭代。

标注图。可用于指定在图中将使用变量和值标签还是变量名称和值。还可指定标签的最大长度。

分类回归规则化

方法。规则化方法可以向 0 方向缩小回归系数估计，以降低其可变性，从而改善模型的预测误差。在要求规则化方法时，每个惩罚系数的规则化模型和系数被写入外部 IBM SPSS Statistics 数据文件或当前会话中的数据集中。请参阅第 12 页的『分类回归: 保存』主题以获取更多信息。

- **Ridge 回归。**Ridge 回归引入惩罚项以缩小系数，惩罚项等于系数平方乘以**惩罚系数**的总和。该系数可从 0（无惩罚）到 1 变化；如果指定了范围与增量，过程将搜索“最佳”的惩罚值。
- **套索。**套索的惩罚项是基于绝对系数的总和，惩罚系数的指定与 Ridge 回归类似，但套索涉及更密集的计算。
- **弹性网络。**“弹性网络”简单地组合套索和 Ridge 回归惩罚，在指定的值网格中搜索以发现“最佳”的套索和 Ridge 回归惩罚系数。对于给定的套索与 Ridge 回归惩罚，“弹性网络”的计算量并不比套索多很多。

显示规则化图。这些是回归系数与规则化惩罚图。在搜索某个值范围以寻找“最佳”惩罚系数时，它提供了有关回归系数在该范围上如何变化的视图。

弹性网络图。对于“弹性网络”方法，由 Ridge 回归惩罚值产生单独的规则化图。**所有可能图**使用指定的最小和最大 Ridge 回归惩罚值所确定范围中的每个值。**为部分 Ridge 惩罚**允许您指定由最小和最大 Ridge 回归惩罚值所确定范围的值子集。只需键入惩罚值的编号（或指定值范围），然后单击**添加**。

分类回归：输出

“输出”对话框允许您选择显示在输出中的统计。

表。为以下项目生成表：

- **复 R**。包括 R^2 、调整后的 R^2 以及将最优刻度考虑在内的调整后的 R^2 。
- **ANOVA**。此选项包括回归、残差平方和、均方和 F 。将显示两个 ANOVA 表：一张的回归自由度等于预测变量的数量，一张的回归自由度将最优刻度考虑在内。
- **系数**。此选项提供三张表：包括 beta、beta 的标准误差、 t 值和显著性的“系数”表；含有将最优刻度自由度考虑在内的 beta 标准误差的“系数最优刻度”表；以及含有零阶、部分和偏相关性、转换后预测变量的 Pratt 的相对重要测度量，以及转换前后容差的表。
- **迭代历史记录**。对于每次迭代，包括算法的起始值在内，将显示复 R 和回归误差。复 R 中的增量从第一次迭代开始列出。
- **初始变量的相关性**。显示未转换的变量之间的相关性的矩阵将被显示。
- **转换变量的相关性**。显示已转换的变量之间的相关性的矩阵将被显示。
- **规则化模型和系数** 显示每个规则化模型的惩罚值、 R 方和回归系数。如果指定了重新抽样方法，或指定了补充对象（检验个案），它还可显示预测误差或检验 MSE。

重新抽样。重新抽样方法为您提供有关模型预测误差的估计。

- **交叉验证**。交叉验证将样本分割为许多子样本，或群。然后，生成分类回归模型，并依次排除每个子样本中的数据。第一个模型基于第一个样本折的个案之外的所有个案，第二个模型基于第二个样本折的个案之外的所有个案，依此类推。对于每个模型，估计其预测误差的方法是将模型应用于生成它所排除的子样本。
- **.632 Bootstrap**。采用 Bootstrap 方法，可以通过放回方式从数据中随机抽取观察值，多次重复该过程以获得大量 Bootstrap 样本。为所有 Bootstrap 样本拟合模型。将应用拟合模型估计出来的每个模型的预测误差应用到不在 Bootstrap 样本中的个案。

类别量化。显示选定的变量转换后的值的表将显示。

描述统计。显示选定的变量的频率、缺失值和众数的表将显示。

分类回归：保存

使用“保存”对话框可以将预测值、残差和转换后的值保存到活动数据集和/或将离散化数据、转换后的值、规则化模型和系数以及回归系数符号保存到当前会话中的外部 IBM SPSS Statistics 数据文件或数据集中。

- 数据集在当前会话期间可用，但在后续会话期间不可用，除非显式将其保存为数据文件。数据集名称必须符合变量命名规则。
- 文件名或数据集名称对于保存的每种数据类型都必须各不相同。

只要在规则化对话框中选择规则化方式，即会保存规则化模型和系数。缺省情况下，该过程以唯一名称创建新数据集，当然您可以自行指定名称，或将其写入外部文件。

只要在选项对话框上使用多系统性起点作为初始配置，即会保存回归系数符号。缺省情况下，该过程以唯一名称创建新数据集，当然您可以自行指定名称，或将其写入外部文件。

分类回归：转换图

使用“图”对话框可以指定将生成转换图和残差图的变量。

转换图。 对于这些变量中的每个变量，将根据初始类别值绘制类别量化。空类别出现在水平轴上，但是不影响计算。这些类别由连接定量的行中的中断符标识。

残差图。 对于这些变量中的每个变量，将根据类别指示符绘制残差（针对从除了有问题的预测变量之外的所有预测变量预测出的因变量进行计算），并且最优类别量化根据类别指示符乘以 `beta`。

CATREG 命令附加功能

如果将选定的内容粘贴到语法窗口中，并编辑生成的 CATREG 命令语法，则可以定制分类回归。使用命令语法语言还可以：

- 将已转换的变量保存到活动数据集时为其指定根名称（使用 `SAVE` 子命令）。

请参阅 [命令语法参考](#) 以获取完整的语法信息。

第 3 章 分类主成份分析 (CATPCA)

此过程在减少数据维数的同时量化分类变量。分类主成份分析也表示为缩写词 CATPCA（代表 *categorical principal components analysis*）。

主成份分析的目标是将初始变量集缩减为表示初始变量中发现的大部分信息的较小不相关主成份集合。当大量变量妨碍有效解释对象（主体和单元格）间关系时，该方法最为有用。通过减少维数，您只需解释少量主成份，而不是大量变量。

标准主成份分析假设数值变量间为线性关系。另一方面，通过最优刻度方法，可以将变量调整为不同级别。分类变量在指定维数内得到最优化。因此，可以为变量间的非线性关系建模。

示例。 分类主成份分析可用于以图形方式显示工作类别、工作部门、地区、旅行量（高、中、低）和工作满意度之间的关系。您可能会发现两个维占据了大量方差。第一维可能将工作类别与地区分开，而第二维可能将工作部门与旅行量分开。您可能还会发现较高的工作满意度与中等旅行量相关。

统计和图。 频率；缺失值；最佳度量水平；众数；按质心坐标、向量坐标、每变量和每维总计解释的方差；向量化变量的成份载荷；类别量化和坐标；迭代历史记录；转换后变量和相关性矩阵特征值的相关性；初始变量与相关性矩阵特征值的相关性；对象得分；类别图；联合类别图；转换图；残差图；投影质心图；对象图；双标图；三标图和成份载荷图。

分类主成份分析数据注意事项

数据。 字符串变量值总是按升序字母数值顺序转换为正整数。用户定义的缺失值、系统缺失值以及小于 1 的值都视为缺失值；可重新编码值小于 1 的变量，或者给值小于 1 的变量加上一个常数，以使其成为非缺失值。

假设。 数据必须至少包含三个有效个案。该分析基于正整数数据。分箱化选项通过将其值分组成具有接近正态分布的类别，将自动分类小数值变量，并且将自动把字符串变量的值转换为正整数。可指定其他分箱化方案。

相关过程。 将所有变量调整为数值级别对应于标准主成份分析。在标准线性主成份分析中使用转换后的变量可获得交替绘图功能。如果所有变量都有多名义刻度级别，则分类主成份分析等同于多重对应分析。如果需要处理的是变量集，则应使用分类（非线性）正态协变量分析。

获取分类主成份分析

1. 从菜单中选择：
 分析 > 降维 > 最优刻度...
2. 选择有些变量不是无序多分类。
3. 选择一个集合。
4. 单击定义。
5. 选择至少两个分析变量并指定解的维数。
6. 单击确定。

也可以指定补充变量，它们适合于找到的解，或者为图标记变量。

在 CATPCA 中定义度量和权重

可以为分析变量和补充变量设置最佳度量水平。缺省情况下，它们调整为带两个内部结点的二度单调性样条（有序）。此外，还可以设置分析变量的权重。

变量权重。 可以选择为每个变量定义权重。指定的值必须为正整数。缺省值为 1。

最佳度量水平。 还可选择用于量化每个变量的刻度级别。

- **有序样条。** 已观察到的变量的类别顺序保存在最优刻度化变量中。类别点将位于一条通过原点的直线上（向量）。生成的转换是一个选定次数的平滑单调的分段多项式。这些段是由内部结点的用户指定的数量和过程决定的位置来指定的。
- **名义样条。** 已观察到的变量中保留在最优刻度化变量中的唯一信息是类别中对象的分组。已观察到的变量的类别的顺序不会保留。类别点将位于一条通过原点的直线上（向量）。生成的转换是一个选定次数的平滑的、可能非单调的分段多项式。这些段是由内部结点的用户指定的数量和过程决定的位置来指定的。
- **多标定。** 已观察到的变量中保留在最优刻度化变量中的唯一信息是类别中对象的分组。已观察到的变量的类别的顺序不会保留。类别点将在特定类别中的对象的质心中。多表示为每一维获取不同的量化集。
- **有序。** 已观察到的变量的类别顺序保存在最优刻度化变量中。类别点将位于一条通过原点的直线上（向量）。生成的转换比有序样条转换拟合得好，但是平滑度较低。
- **名义。** 已观察到的变量中保留在最优刻度化变量中的唯一信息是类别中对象的分组。已观察到的变量的类别的顺序不会保留。类别点将位于一条通过原点的直线上（向量）。生成的转换比样条名义转换拟合得更好，但是平滑度较低。
- **数值。** 类别被视为是有序的并且是等间距的（区间度）。已观察到的变量的类别顺序以及类别号之间的等间距保留在最优刻度化变量中。类别点将位于一条通过原点的直线上（向量）。当所有变量都为数值级别时，分析与标准主成分分析类似。

分类主成份分析：分箱化

使用“分箱化”对话框可以选择对变量重新编码的方法。除非另有指定，否则小数值变量将分组为具有近似正态分布的七个类别（如果变量的不同值的数目小于七，则将按此数目划分类别）。通过按照升序字母数值顺序分配类别指示符，字符串变量总是转换为正整数。字符串变量的分箱化适用于这些整数。缺省情况下，其他变量保留原样。随后，分析中将使用离散化变量。

方法。 在分组、等级和乘之间选择。

- **分组。** 重新编码为指定数量的类别或者按区间重新编码。
- **等级。** 通过对个案等级排序来对变量进行离散化。
- **乘。** 变量的当前值是乘以 10 且经过四舍五入的离散化值，并且加上了一个常数以使最低离散值为 1。

分组。 在按分组分箱化变量时，可以使用以下选项：

- **类别数。** 指定一定数量的类别，并指定变量的值是否应该遵循在这些类别间的近似正态分布或均匀分布。
- **相等区间。** 变量重新编码到由这些等大小的间隔所定义的类别中。必须指定间隔的长度。

分类主成份分析：缺失值

使用“缺失值”对话框可以选择用于处理分析变量和补充变量中缺失值的方案。

策略。 选择排除缺失值（消极处理）、插补缺失值（积极处理）或者排除带缺失值的对象（成列删除）。

- **排除缺失值；量化后为相关性插补。** 选定的变量有缺失值的对象对于此变量的分析不起作用。如果消极处理所有变量，则所有变量都有缺失值的对象将视为补充对象。如果在“输出”对话框中指定了相关性，则（分析后）缺失值通过初始变量相关性变量的最频繁类别（即众数）进行插补。对于最优刻度化变量的相关性，可选择插补方法。
 - **众数。** 将缺失值替换为最优刻度化变量的众数。
 - **附加类别。** 将缺失值替换为附加类别的量化。此设置表示该变量中具有缺失值的对象被视为属于同一（附加）类别。
 - **随机类别。** 根据变量的类别的边际频率，在具有另一随机类别数的量化值的变量中插补每个缺失值。
- **插补缺失值。** 选定的变量有缺失值的对象将对这些值插补。您可以选择插补方法。
 - **众数。** 将缺失值替换为最频繁的类别。当有多个众数时，将使用具有最小类别指示符的众数。
 - **附加类别。** 将缺失值替换为附加类别的同一量化。此设置表示该变量中具有缺失值的对象被视为属于同一（附加）类别。
 - **随机类别。** 根据类别的边际频率，在具有另一随机类别数的变量中插补每个缺失值。
- **排除此变量具有缺失值的对象。** 选定的变量有缺失值的对象将从分析中排除。此方案不适用于补充变量。

分类主成份分析：选项

使用“选项”对话框可以选择初始配置、指定迭代和收敛条件、选择正态化方法、选择标记图的方法以及指定补充对象。

补充对象。 指定要其成为附加对象的对象的个案编号，或者对象范围的第一个和最后一个个案编号，然后单击添加。如果将某个对象指定为附加对象，则对于该对象将忽略个案权重。

标准化方法。 可以指定用于标准化对象得分和变量的五个选项之一。在每个分析中只能使用一种正态化方法。

- **主要变量。** 此选项可优化变量之间的关联。对象空间中的变量坐标是成份加载（与主成分的相关性，如维和对象得分）。如果您主要对变量之间的相关性感兴趣，那么此方法很有用。
- **主要对象。** 此选项可优化对象间的距离。如果您主要对对象之间的区别或相似性感兴趣，那么此方法很有用。
- **对称。** 如果主要对对象和变量之间的关系感兴趣，则使用此标准化选项。
- **独立。** 如果您想单独检查对象之间的距离和变量之间的相关性，请使用此标准化选项。
- **定制。** 可指定封闭区间 $[-1, 1]$ 中的任何实数值。值 1 表示“对象主成分”方法。值 0 表示“对称”方法。值 -1 表示“变量主成分”方法。通过指定大于 -1 小于 1 的值，可在对象和变量上分布特征值。此方法对于制作合适的双标图或三标图很有用。

标准。 可以指定该过程可在其计算中执行的最大迭代次数。还可以选择收敛性标准值。如果上两次迭代之间的总拟合之差小于收敛性值，或者达到了最大迭代次数，则算法停止迭代。

标注图。 您可以指定在图中使用变量和值标签还是变量名称和值。还可指定标签的最大长度。

图维数。 您可以控制在输出中显示的维数。

- **显示解中的所有维数。** 解中的所有维数都显示在散点图矩阵中。
- **限制维数。** 显示的维数限制为绘制的对。如果限制维数，则必须选择要绘制的最低和最高维数。最低维数的范围可从 1 到解中的维数减 1，并且针对较高维数绘制。最高维数值的范围可从 2 到解中的维数，表示要在绘制维对时使用的最高维数。此指定项适用于所有请求的多维图。

旋转。 您可以选择用于获取旋转结果的旋转方法。

注：如果您在**自助抽样**对话框中选择了**执行自助抽样**，那么这些旋转方法不可用。

- **最大方差法**。一种正交旋转方法，它使得在每个成分上具有高负载的变量的数目达到最小。它简化了成分的解释。
- **四次幂极大法**。一种旋转方法，它可使得解释每个变量所需的成分最少。它简化了实测变量的解释。
- **等量最大法**。一种旋转方法，它是简化成分的最大方差法与简化变量的最大四次方方法的组合。它可以使得高度依赖成分的变量的个数以及解释变量所需的成分的个数最少。
- **斜交法**。斜交（非正交）旋转法。当 Δ 等于 0 时，成分的斜交度最大。 Δ 为负的程度越高，成分的斜交度越低。正值允许额外的成分相关性。**Delta** 的值必须小于或等于 0.8。
- **最优斜交法**。斜交旋转，可使成分相关联。此旋转的计算速度比直接最小斜交旋转更快，因此它可用于大型数据集。所允许的相关性量由 κ 参数控制。**Kappa** 的值必须大于或等于 1 并且小于 10,000。

配置。您可以从包含配置的坐标的文件中读取数据。此文件中的第一个变量包含第一维的坐标。第二个变量包含第二维的坐标，依次类推。

- **初始**。指定的文件中的配置用作分析起点。
- **固定**。指定的文件中的配置用于拟合变量。拟合的变量必须选择作为分析变量，但是因为配置是固定的，所以它们视为补充变量（因此不需要选择它们作为补充变量）。

分类主成份分析：输出

“输出”对话框用于控制结果的显示。

对象得分。显示对象得分并且有以下选项：

- **包括类别**。显示选定的分析变量的类别指示符。
- **标注对象得分**。从指定为标注变量的变量列表中，您可以选择一个用于标注对象的变量。

成份加载。显示未获得多标定刻度级别的所有变量的成份加载。您可以按大小对成分载荷进行排序。

迭代历史记录。对于每次迭代，显示偏差考虑情况以及偏差考虑情况中的损失和增加项。

初始变量的相关性。显示初始变量的相关性矩阵以及该矩阵的特征值。

转换变量的相关性。显示已转换的（最优刻度）变量的相关性矩阵以及该矩阵的特征值。

方差所占百分比。显示按质心坐标、向量坐标以及每个变量和每个维总计（合并的质心坐标和向量坐标）解释的方差量。

类别量化。给出所选变量每个维的类别量化和坐标。

描述统计。显示所选变量的频率、缺失值数目和众数。

分类主成份分析：保存

“保存”对话框中的控件用于将离散化数据、对象分数、变换后的值以及其他结果保存到活动数据集、当前会话中的新数据集或外部文件中。

- 数据集在当前会话期间可用，但在后续会话期间不可用，除非显式将其保存为数据文件。数据集名称必须符合变量命名规则。
- 文件名或数据集名称对于保存的每种数据类型都必须各不相同。
- 如果将对对象得分或已转换的值保存到活动数据集中，则可以指定多名义维数的数量。

- 仅当您在“自助抽样”对话框中选择**执行自助抽样**时，**自助抽样置信椭圆**组中的选项才可用。

分类主成份分析：对象图

使用“对象和变量图”对话框可以指定所要的图类型以及将为其生成图的变量。

对象点。 显示对象点的图。

对象和变量（双标图）。 对象点是用您所选择的变量坐标（成份加载或变量质心）绘制的。

对象、加载和质心（三标图）。 对象点是用多名义刻度级别变量的质心以及其他变量的成份加载绘制的。

双标图和三标图变量。 可选择将所有变量用于双标图和三标图或选择一个子集。

标签对象。 可选择用选定的变量的类别（可在“选项”对话框中选择类别指示符值或值标签）或其个案号标注对象。如果选择**变量**，则为每个变量生成一个图。

分类主成份分析：类别图

使用“类别图”对话框可以指定所要的图类型以及将为其生成图的变量。

类别图。 对于选定的每个变量绘制质心和向量坐标图。对于多名义刻度级别的变量，类别在该特定类别中的对象的质心中。对于所有其他刻度级别，类别在通过原点的向量上。

联合类别图。 这是每个选定的变量的质心和向量坐标的单个图。

转换图。 显示最优类别量化与类别指示符的比较图。可以为拥有多名义刻度级别的变量指定维数；将为每一维分别生成一个图。还可以选择显示每个选定的变量的残差图。

质心投影。 可选择一个变量并将其质心投影到选定的变量上。不能选择拥有多名义刻度级别的变量作为投影的对象。当请求此图时，含有投影的质心的坐标的表也会同时显示。

分类主成份分析：载荷图

“载荷图”对话框用于控制图中包含的变量、在载荷图中显示质心以及显示方差所占百分比的图。

方差所占百分比。 对于每个维，将显示方差所占百分比的图。

显示成份加载。 如果选中，将显示成份载荷图。

载荷变量。 可选择将所有变量用于成份载荷图或选择一个子集。

包含质心。 具有多名义刻度级别的变量没有成分加载，但是可以选择在图中包含这些变量的质心。可选择使用所有多名义变量或选择一个子集。

分类主成份分析自助抽样

“自助抽样”对话框用于指定自助抽样分析的参数。

执行自助抽样。 执行自助重新抽样。

- 如果对载荷、类别或组件分数图发出了请求，那么将显示额外的图。这些图显示数据样本的点和自助抽样估算。转换图包括置信区域。还将显示特征值的图。

- 如果指定了二维解决方案，那么将显示特征值的置信椭圆图、成分载荷、类别点和对象点。
- 如果您在“选项”对话框中指定了旋转方法，那么自助重新抽样不可用。

您可以选择**已平衡**或**未平衡**自助抽样。

样本数。这是用于计算自助抽样估算的自助抽样数。此值必须为正整数。

置信度级别。这是自助抽样估算的置信度级别，以百分比表示。此值必须为小于 100 的正数。

匹配方法。两种替代方法为**普氏**和**反射**。

置信椭圆。用于控制图中置信椭圆的阈值区域。对于面积大于 (> 运算符) 指定值的区域，椭圆的数目将随阈值的增大而减少。

- 仅当主对话框中指定的维数为 2 时，此组中的设置才可用。
- 仅在“载荷图”对话框中选择了**显示成分载荷**时，载荷图的置信椭圆才可用。
- 仅在“对象和变量图”对话框中选择了**对象点**时，对象图的置信椭圆才可用。
- 仅当在“类别图”对话框的**类别图**列表中指定了一个或多个变量时，类别图的置信椭圆才可用。

椭圆轮廓点数。图点数用于绘制每个置信椭圆。值越大，生成的椭圆越平滑。此值必须为小于或等于 100 的正整数。

CATPCA 命令附加功能

如果将选中的内容粘贴到语法窗口中，并编辑得到的 CATPCA 命令语法，则可以定制分类主成份分析。使用命令语法语言还可以：

- 将已转换的变量、对象得分和近似值保存到活动数据集时（使用 SAVE 子命令），为其指定根名称。
- 分别为每个图指定最大标签长度（使用 PLOT 子命令）。
- 为残差图指定单独的变量列表（使用 PLOT 子命令）。

请参阅**命令语法参考**以获取完整的语法信息。

第 4 章 非线性正态协变量分析 (OVERALS)

非线性正态协变量分析对应于使用最优刻度的分类正态协变量分析。此过程的目的是确定分类变量集相互之间的相似程度。非线性正态协变量分析也用缩写词 OVERALS 来表示。

标准正态协变量分析是多重回归的扩展，其中第二个集不包含单响应变量，而是包含多响应变量。其目标是尽可能解释低维空间中两个数值变量集之间的关系中的方差。最初，每个集内的变量进行线性组合以使线性组合有最大的相关性。有了这些组合，就可以确定后续线性组合与前面的组合无关，并可确定其具有可能的最大相关性。

最优刻度方法在三个重要方面扩展了标准分析。首先，OVERALS 允许两个以上的变量集。其次，变量或者可调整为名义、有序，或者调整为数值。因此，可以分析变量间的非线性关系。最后，变量集与一个由对象得分定义的未知折中集进行比较，而不是使变量集之间的相关性最大化。

示例。 使用最优刻度的分类正态协变量分析可用于以图形方式显示包含工作类别和教育年限的一个变量集与包含居住地区和性别的另一个变量集之间的关系。您可能会发现教育年限与居住地区的区别程度比其余变量高。您还可能发现教育年限在第一维上区别最大。

统计和图。 频率、质心、迭代历史记录、对象得分、类别量化、权重、成份加载、单拟合和多拟合、对象得分图、类别坐标图、成份载荷图、类别质心图、转换图。

非线性正态协变量分析数据注意事项

数据。 使用整数来编码分类变量（名义或有序刻度级别）。要最小化输出，请使用从 1 开始的连续整数来编码每个变量。调整为数值级别的变量不应重新编码为连续整数。要为每个调整为数值级别的变量最大程度地减少输出，请从每个值中减去最小实测值然后加上 1。小数后的分数值将被截断。

假设。 变量可分成两个或更多变量集。分析中的变量调整为多名义、单名义、有序或数值。过程中使用的最大维数取决于变量的最佳度量水平。如果所有变量都指定为有序、单名义或数值，则最大维数是以下两个值中的较小者：观察值数量减 1 或变量总数。但是，如果只定义了两个变量集，则最大维数为较小集中的变量数。如果某些变量为多名义，则最大维数为多名义类别的总数加上非多名义变量的数目减去多名义变量的数目。例如，如果分析涉及五个变量，其中一个变量是带有四种类别的多名义变量，则最大维数为 $(4 + 4 - 1)$ ，即 7。如果指定了大于最大值的数，则会使用最大值。

相关过程。 如果每个集只包含一个变量，则非线性正态协变量分析等效于使用最优刻度的主成分分析。如果所有这些变量都是多名义，则分析对应于多重对应分析。如果涉及两个变量集，并且其中一个仅包含一个变量，则分析等同于使用最优刻度的分类回归。

获取非线性正态协变量分析

1. 从菜单中选择:

分析 > 降维 > 最优刻度...

2. 或者选择所有变量都是无序多分类，或者选择有些变量不是无序多分类。

3. 选择多个集合。

4. 单击定义。

5. 定义至少两个变量集。选择要包含在第一个集内的变量。若要移到下一个集合，单击下一个，然后选择要包含在第二个集合中的变量。可以添加更多集合。单击上一个返回到先前定义的变量集。

6. 为每个选定的变量定义值范围和测量刻度（最佳度量水平）。
7. 单击**确定**。
8. 或者：
 - 选择一个或多个要为对象得分图提供点标签的变量。各个变量分别产生一个图，由该变量的值标注各个点。必须为各个图标签变量分别定义一个范围。在使用该对话框时，单个变量不能既在分析中使用，又作为标签变量使用。如果想要用在分析中使用的变量标注对象得分图，请使用“计算”工具（可从“转换”菜单上获取）创建该变量的副本。使用新变量标注图。或者，可使用命令语法。
 - 指定解中您所要的维数。通常，根据需要选择尽量少的维数来解释大多数变异。如果分析涉及两个以上的维，则将生成前三维的三维图。通过编辑图表可显示其他维。

定义范围和刻度

必须为每个变量定义范围。指定的最大值必须是整数。小数数据值会在分析中被截断。指定范围之外的类别值在分析中会被忽略。为了使输出最小，请使用“自动重新编码”工具（可从“转换”菜单上获取）为视作名义或有序的变量创建从 1 开始的连续类别。对于调整为数值级别的变量，建议不要重新编码为连续整数。要最小化视为数值的变量的输出，对于每个变量，从每个值中减去最小值然后加上 1。

还必须选择将用于量化每个变量的刻度。

- **有序**。将观察变量的类别顺序保存在量化变量中。
- **单标定**。在已量化变量中，相同类别中的对象得到相同得分。
- **多标定**。每个维的定量可以不同。
- **离散数**。类别被视为有序且等间距。观察到的变量的类别号和类别顺序之间的差别保留在已量化的变量中。

定义范围

必须为每个变量定义范围。指定的最大值必须是整数。小数数据值会在分析中被截断。指定范围之外的类别值在分析中会被忽略。要最小化输出，请使用“自动重新编码”工具（可从“转换”菜单上获取）创建从 1 开始的连续类别。

还必须为用于标注对象得分图的每个变量定义范围。但是，数据值在为变量定义的范围之外的类别的标签确实显示在图中。

非线性正态协变量分析：选项

使用“选项”对话框可以选择可选统计和图、将对象得分保存为活动数据集中的新变量、指定迭代和收敛性标准以及为分析指定初始配置。

输出。可用统计包括边际频率（计数）、质心、迭代历史记录、权重和成份加载、类别量化、对象得分以及单拟合和多拟合统计。

- **质心 (Centroids)**。这是指对象（个案）的类别量化以及对象得分的投影平均值和实际平均值，这些对象包括在那些属于同一变量类别的个案的每个集合中。
- **权重和成分加载 (Weights and component loadings)**。集合中每个已量化的变量的每个维度的回归系数（其中，在已量化的变量上对对象得分进行回归）以及已量化的变量在对象空间中的投影。它指示每个变量对每个集合中的维度的贡献。
- **单拟合和多拟合 (Single and multiple fit)**。对于对象，是对单和多类别坐标 / 类别量化的拟合优度的测量。
- **类别量化 (Category quantifications)**。分配给变量类别的最优刻度值。

- **对象得分 (Object scores)**. 分配给特定维度中某个对象 (个案) 的最优得分。

图。可生成类别坐标图、对象得分图、成份载荷图、类别质心图以及转换图。

保存对象得分。 可将对象得分保存为活动数据集中的新变量。对象得分针对在主对话框中指定的维数保存。

使用随机初始配置。 如果部分或全部变量为单名义, 则应使用随机初始配置。如果未选择此选项, 则使用嵌套初始配置。

标准。 可以指定非线性正态协变量分析可在其计算中执行的最大迭代次数。还可以选择收敛性标准值。如果上两次迭代之间的总拟合之差小于收敛性值, 或者达到了最大迭代次数, 则分析停止迭代。

OVERALS 命令附加功能

如果将选中的内容粘贴到语法窗口中, 并编辑生成的 OVERALS 命令语法, 则可以定制非线性正态协变量分析。使用命令语法语言还可以:

- 指定要绘制的维对, 而不是绘制所有抽取的维 (在 PLOT 子命令上使用 NDIM 关键字)。
- 指定用于在图上标注点的值标签字符的数目 (使用 PLOT 子命令)。
- 指定五个以上的变量作为对象得分图的标记变量 (使用 PLOT 子命令)。
- 选择在分析中使用的变量作为对象得分图的标记变量 (使用 PLOT 子命令)。
- 选择变量以便为定量得分图提供点标签 (使用 PLOT 子命令)。
- 如果不想使用活动数据集中所有个案, 请指定要包含在分析中的个案数 (使用 NOBSERVATIONS 子命令)。
- 为通过保存对象得分创建的变量指定根名称 (使用 SAVE 子命令)。
- 指定要保存的维数, 而不是保存所有抽取的维 (使用 SAVE 子命令)。
- 将类别量化写入矩阵文件 (使用 MATRIX 子命令)。
- 生成可能比常规高分辨率图更易于读取的低分辨率图 (使用 SET 子命令)。
- 仅对指定变量生成质心和转换图 (使用 PLOT 子命令)。

请参阅命令语法参考以获取完整的语法信息。

第 5 章 对应分析

对应分析的目标之一是描述低维空间中对应表中的两个名义变量之间的关系，同时描述每个变量的类别之间的关系。对于每个变量，类别点在图中的距离反映了相似的类别绘制为相互靠近的类别间的关系。从原点到另一个变量的类别点的向量上的一个变量的投影点描述了变量之间的关系。

列联表的分析通常包括检查行和列概要文件以及通过卡方统计检验自变量。但是，概要文件的数量可能非常大，并且卡方检验不会揭示因变量结构。交叉表格过程提供了多个相关性测量和关联性检验，但是不能以图形方式表示变量间的任何关系。

因子分析是描述低维空间中的变量之间的关系的标准技术。但是，因子分析需要定距数据，并且观察数应为变量数的五倍。而对应分析采用名义变量，并且可以描述每个变量的类别之间的关系和变量之间的关系。此外，对应分析可用于分析任何正对应度量的表。

示例。 对应分析可用于以图形的方式显示员工类别和吸烟习惯之间的关系。您可能会发现，就吸烟而言，初级经理人有别于秘书，但是秘书与高级经理人没差别。您可能还会发现重度烟瘾与初级经理人相关，而轻度烟瘾与秘书相关。

统计和图。 对应度量、行和列概要文件、奇异值、行和列得分、惯量、质量、行和列得分置信统计、奇异值置信统计、转换图、行点图、列点图和双标图。

对应分析数据注意事项

数据。 要分析的分类变量调整为名义。对于汇总数据或除频率之外的对应度量，应使用具有正相似性值的加权变量。或者，对于表数据，使用语法读取表。

假设。 该过程中使用的最大维数取决于活动行和列类别的数目以及等同性约束的数目。如果未使用等同性约束并且所有类别都是活动的，则最大维数比具有最少类别的变量的类别数少一。例如，如果一个变量有五个类别，另一个变量有四个类别，则最大维数为三。补充类别不是活动类别。例如，如果一个变量有五个类别，其中两个是补充类别，而另一个变量有四个类别，则最大维数为二。将所有约束为相等的类别集视为一个类别。例如，如果一个变量有五个类别，其中三个约束为相等，则在确定最大维数时，应将该变量视为有三个类别。其中两个类别未受约束，第三个类别对应于三个受约束的类别。如果指定大于最大值的维数，则会使用最大值。

相关过程。 如果涉及两个以上的变量，则使用多重对应分析。如果变量应调整为有序，则使用分类主成份分析。

获取对应分析

1. 从菜单中选择:

 分析 > 降维 > 对应分析...

2. 选择行变量。

3. 选择列变量。

4. 定义变量的范围。

5. 单击**确定**。

在“对应分析”中定义行范围

必须为行变量定义范围。指定的最小值和最大值必须为整数。小数数据值会在分析中被截断。指定范围之外的类别值在分析中会被忽略。

初始状态下，所有类别都不受约束且是活动的。可将行类别约束为等于其他行类别，或者可将行类别定义为补充类别。

- **类别必须相等。** 类别必须具有相等的得分。如果所获得的类别顺序不理想或不直观，请使用等同性约束。可约束为相等的行类别的最大数量等于活动行类别总数减 1。若要对类别集施加不同的等同性约束，请使用语法。例如，使用语法，将类别 1 和 2 约束为相等，将类别 3 和 4 约束为相等。
- **类别为补充型。** 补充类别不影响分析，但会出现在由活动类别定义的空间中。补充类别对定义维不起作用。补充行类别的最大数目为行类别总数减 2。

在“对应分析”中定义列范围

必须为列变量定义范围。指定的最小值和最大值必须为整数。小数数据值会在分析中被截断。指定范围之外的类别值在分析中会被忽略。

初始状态下，所有类别都不受约束且是活动的。可将列类别约束为等于其他列类别，或者可将列类别定义为补充类别。

- **类别必须相等。** 类别必须具有相等的得分。如果所获得的类别顺序不理想或不直观，请使用等同性约束。可约束为相等的列类别的最大数目等于活动列类别总数减 1。要对多组类别设定不同的等同性约束，请使用语法。例如，使用语法，将类别 1 和 2 约束为相等，将类别 3 和 4 约束为相等。
- **类别为补充型。** 补充类别不影响分析，但会出现在由活动类别定义的空间中。补充类别对定义维不起作用。补充列类别的最大数目为列类别总数减 2。

对应分析：模型

使用“模型”对话框可以指定维数、距离测量、标准化方法以及标准化方法。

解的维数。 指定维数。通常，根据需要进行选择尽量少的维数来解释大多数变异。最大维数取决于分析中使用的活动类别数以及等同性约束。最大维数是以下两项中的较小者：

- 活动行类别数减去约束为相等的行类别数，加上受约束的行类别集的数目。
- 活动列类别数减去约束为相等的列类别数，加上受约束的列类别集的数目。

距离测量。 可选择对应表的行和列之间距离的测量。选择下列选项之一：

- **卡方。** 使用加权概要文件距离，其中权重是行或列的质量。此度量是标准对应分析所必需的。
- **欧几里德距离。** 使用行对和列对之间平方差之和的平方根。

标准化方法。 选择下列选项之一：

- **行和列平均值已删除。** 行和列都居中。此方法是标准对应分析所必需的。
- **行平均值已删除。** 只有行居中。
- **列平均值已删除。** 只有列居中。
- **使行总和相等，移去删除。** 使行居中之前，使行边距相等。
- **使列总和相等，移去删除。** 使列居中之前，使列边距相等。

标准化方法。 选择下列选项之一：

- **对称。** 对于每个维，行得分为列得分的加权平均值除以对应的奇异值，列得分为行得分的加权平均值除以对应的奇异值。如果想要检查两个变量的类别之间的差异或相似性，则使用此方法。
- **主要。** 行点和列点之间的距离是对应表中对应于所选距离测量的距离的近似值。如果想要检查一个或两个变量的类别之间的差别，而非两个变量之间的差别，则使用此方法。
- **主要行。** 行点之间的距离是对应表中对应于所选距离测量的距离近似值。行得分是列得分的加权平均值。如果想要检查行变量的类别之间的差分或相似性，则使用此方法。
- **主要列。** 列点之间的距离是对应表中对应于所选距离测量的距离近似值。列得分是行得分的加权平均值。如果想要检查列变量的类别之间的差异或相似性，则使用此方法。
- **定制。** 必须指定介于 -1 和 1 之间的值。值为 -1 对应于主要列。值为 1 对应于主要行。值为 0 对应于对称。所有其他值不同程度地将惯量分布于行得分和列得分上。此方法对于制作合适的双标图很有用。

对应分析：统计

使用“统计”对话框可以指定生成的数值输出。

对应表。 带有行和列边际总计的输入变量的交叉表格。

行点概览。 对于每个行类别，有得分、质量、惯量、对维惯量的贡献，以及维对点惯量的贡献。

列点概览。 对于每个列类别，有得分、质量、惯量、对维惯量的贡献，以及维对点惯量的贡献。

行概要文件。 对于每个行类别，为跨列变量类别的分布。

列概要文件。 对于每个列类别，为跨行变量类别的分布。

对应表的排列。 对应表重新组织为：行和列根据第一维上的得分按递增顺序排列。或者，可指定将为其生成置换表的最大维数。将为从 1 到指定数字的每一维分别生成一个置换表。

行点的置信统计。 包括所有非补充行点的标准差和相关性。

列点的置信统计。 包括所有非补充列点的标准差和相关性。

对应分析：图

使用“图”对话框可以指定生成哪些图。

散点图。 生成维的所有成对图矩阵。可用散点图包括：

- **双标图。** 生成行点和列点的联合图矩阵。如果选择了主标准化，则双标图不可用。
- **行点。** 生成行点图的矩阵。
- **列点。** 生成列点图的矩阵。

或者，可指定在标注点时要使用多少个值标签字符。该值必须为小于或等于 20 的非负整数。

线图。 为所选变量的每一维生成一个图。可用的线图包括：

- **已转换的行类别。** 根据初始行类别值的对应行得分生成这些值的图。
- **已转换的列类别。** 根据初始列类别值的对应列得分生成这些值的图。

或者，可指定在标注类别轴时要使用多少个值标签字符。该值必须为小于或等于 20 的非负整数。

图维数。 可用以控制在输出中显示的维数。

- **显示解中的所有维数。**解中的所有维数都显示在散点图矩阵中。
- **限制维数。**显示的维数限制为绘制的对。如果限制维数，则必须选择要绘制的最低和最高维数。最低维数的范围可从 1 到解中的维数减 1，并且针对较高维数绘制。最高维数值的范围可从 2 到解中的维数，表示要在绘制维对时使用的最高维数。此指定项适用于所有请求的多维图。

CORRESPONDENCE 命令附加功能

如果将选定的内容粘贴到语法窗口中，并编辑生成的 CORRESPONDENCE 命令语法，则可以定制对应分析。使用命令语法语言还可以：

- 指定表数据作为输入，而不是使用个案情况数据（使用 TABLE = ALL 子命令）。
- 指定用于标注每种类型的散点图矩阵或双标图矩阵的点的值标签字符数（使用 PLOT 子命令）。
- 指定用于标注每种类型的线图的点的值标签字符数（使用 PLOT 子命令）。
- 将行和列得分的矩阵写入矩阵数据文件（使用 OUTFILE 子命令）。
- 将奇异值和得分的置信统计（方差和协方差）的矩阵写入矩阵数据文件（使用 OUTFILE 子命令）。
- 指定多个类别集相等（使用 EQUAL 子命令）。

请参阅命令语法参考以获取完整的语法信息。

第 6 章 多重对应分析

多重对应分析通过为个案（对象）和类别分配数值量化标定（分类）数据，以使同类别的对象紧靠在一起，不同类别的对象远远分开。每个对象都尽可能地接近应用于该对象的类别的类别点。这样，类别可将对象分成数个均一子组。当变量将相同类别中的对象分类到同一子组中时，这些变量被视为是均一的。

示例。 多重对应分析可用以图形方式显示工作类别、少数民族分类和性别之间的关系。您可能发现在人群之间少数民族分类和性别有区别，但是工作类别没有区别。您可能还会发现拉丁裔美国人和非洲裔美国人类别相似。

统计和图。 对象得分、区分测量、迭代历史记录、原始变量和转换后变量的相关性、类别量化、描述统计、对象点图、双标图、类别图、联合类别图、转换图和区分测量图。

多重对应分析数据注意事项

数据。 字符串变量值总是按升序字母数值顺序转换为正整数。用户定义的缺失值、系统缺失值以及小于 1 的值都视为缺失值；可重新编码值小于 1 的变量，或者给值小于 1 的变量加上一个常数，以使其成为非缺失值。

假设。 所有变量都有多名义刻度级别。数据必须至少包含三个有效个案。该分析基于正整数数据。分箱化选项通过将其值分组成具有接近正态分布的类别，将自动分类小数值变量，并且将自动把字符串变量的值转换为正整数。可指定其他分箱化方案。

相关过程。 对于两个变量，多重对应分析类似于对应分析。如果相信变量具有有序或数值属性，则应使用分类主成份分析。如果需要处理变量集，则应使用非线性正态协变量分析。

获取多重对应分析

1. 从菜单中选择：

 分析 > 降维 > 最优刻度...

2. 选择所有变量都是无序多分类。

3. 选择一个集合。

4. 单击**定义**。

5. 选择至少两个分析变量并指定解的维数。

6. 单击**确定**。

也可以指定补充变量，它们适合于找到的解，或者为图标记变量。

在“多重对应分析”中定义变量权重

可以设置分析变量的权重。

变量权重。 可以选择为每个变量定义权重。指定的值必须为正整数。缺省值为 1。

多重对应分析：分箱化

使用“分箱化”对话框可以选择对变量重新编码的方法。除非另有指定，否则小数值变量分组成具有近似正态分布的七个类别（如果变量的相异值的数目小于七，则按此数目划分类别）。通过按照升序字母数值顺序分配类别指示符，字符串变量总是转换为正整数。字符串变量的分箱化适用于这些整数。缺省情况下，其他变量保留原样。随后，分析中将使用离散化变量。

方法。 在分组、等级和乘之间选择。

- **分组。** 重新编码为指定数量的类别或者按区间重新编码。
- **等级。** 通过对个案等级排序来对变量进行离散化。
- **乘。** 变量的当前值是乘以 10 且经过四舍五入的离散化值，并且加上了一个常数以使最低离散值为 1。

分组。 在按分组分箱化变量时，可以使用以下选项：

- **类别数。** 指定一定数量的类别，并指定变量的值是否应该遵循在这些类别间的近似正态分布或均匀分布。
- **相等区间。** 变量重新编码到由这些等大小的间隔所定义的类别中。必须指定间隔的长度。

多重对应分析：缺失值

使用“缺失值”对话框可以选择处理分析变量和补充变量中的缺失值的方案。

缺失值方案。 选择排除缺失值（消极处理）、插补缺失值（积极处理）或者排除带缺失值的对象（成列删除）。

- **排除缺失值；量化后为相关性插补。** 选定的变量有缺失值的对象对于此变量的分析不起作用。如果消极处理所有变量，则所有变量都有缺失值的对象将视为补充对象。如果在“输出”对话框中指定了相关性，则（分析后）缺失值通过初始变量相关性变量的最频繁类别（即众数）进行插补。对于最优刻度化变量的相关性，可选择插补方法。选择**众数**将缺失值替换为最优刻度化变量的众数。选择**附加类别**将缺失值替换为附加类别的定量。这意味着此变量有缺失值的对象被视为属于同一（附加）类别。
- **插补缺失值。** 选定的变量有缺失值的对象将对这些值插补。您可以选择插补方法。选择**众数**将缺失值替换为最频繁的类别。当有多个众数时，将使用具有最小类别指示符的众数。选择**附加类别**将缺失值替换为相同的附加类别定量。这意味着此变量有缺失值的对象被视为属于同一（附加）类别。
- **排除此变量具有缺失值的对象。** 选定的变量有缺失值的对象将从分析中排除。此方案不适用于补充变量。

多重对应分析：选项

使用“选项”对话框可以选择初始配置、指定迭代和收敛性标准、选择标准化方法、选择标记图的方法以及指定附加对象。

补充对象。 指定要其成为附加对象的对象的个案编号，或者对象范围的第一个和最后一个个案编号，然后单击**添加**。继续操作，直到指定完所有附加对象。如果将某个对象指定为附加对象，则对于该对象将忽略个案权重。

标准化方法。 可以指定用于标准化对象得分和变量的五个选项之一。给定分析中只能使用一种标准化方法。

- **主要变量。** 此选项可优化变量之间的关联。对象空间中的变量坐标是成份加载（与主成分的相关性，如维和对象得分）。如果您主要对变量之间的相关性感兴趣，这将非常有用。
- **主要对象。** 此选项可优化对象间的距离。如果您主要对对象之间的区别或相似性感兴趣，这将非常有用。
- **对称。** 如果主要对对象和变量之间的关系感兴趣，则使用此标准化选项。
- **独立。** 如果您想单独检查对象之间的距离和变量之间的相关性，请使用此标准化选项。

- **定制。** 可指定封闭区间 $[-1, 1]$ 中的任何实数值。值为 1 等同于“主要对象”方法；值为 0 等同于“对称”方法；值为 -1 等同于“主要变量”方法。通过指定大于 -1 小于 1 的值，可在对象和变量上分布特征值。此方法对于制作合适的双标图或三标图很有用。

标准。 可以指定该过程可在其计算中执行的最大迭代次数。还可以选择收敛性标准值。如果上两次迭代之间的总拟合之差小于收敛性值，或者达到了最大迭代次数，则算法停止迭代。

标注图。 可用于指定在图中将使用变量和值标签还是变量名称和值。还可指定标签的最大长度。

图维数。 可用于控制在输出中显示的维数。

- **显示解中的所有维数。** 解中的所有维数都显示在散点图矩阵中。
- **限制维数。** 显示的维数限制为绘制的对。如果限制维数，则必须选择要绘制的最低和最高维数。最低维数的范围可从 1 到解中的维数减 1，并且针对较高维数绘制。最高维数值的范围可从 2 到解中的维数，表示要在绘制维对时使用的最高维数。此指定项适用于所有请求的多维图。

配置。 可从包含配置的坐标的文件中读取数据。文件中的第一个变量应包含第一维的坐标，第二个变量应包含第二维的坐标，依此类推。

- **初始。** 指定的文件中的配置将用作分析起点。
- **固定。** 指定的文件中的配置将用于拟合变量。拟合的变量必须选择作为分析变量，但是，因为配置是固定的，所以它们被视为补充变量（因此不需要选择它们作为补充变量）。

多重对应分析：输出

使用“输出”对话框可以为对象得分、区分测量、迭代历史、原始变量和转换后变量的相关性、选定的变量的类别量化和选定的变量的描述统计生成表。

对象得分。 显示对象得分，包括质量、惯量和贡献，并且有以下选项：

- **包括类别。** 显示选定的分析变量的类别指示符。
- **标注对象得分。** 从指定为标签变量的变量列表中，可选择一个用以标注对象的变量。

区分测量。 显示每个变量和每一维的区分测量。

迭代历史记录。 对于每次迭代，显示偏差考虑情况以及偏差考虑情况中的损失和增加项。

初始变量的相关性。 显示初始变量的相关性矩阵以及该矩阵的特征值。

转换变量的相关性。 显示已转换的（最优刻度）变量的相关性矩阵以及该矩阵的特征值。

类别量化和贡献。 针对选定的变量的每一维给出类别量化（坐标），包括质量、惯量和贡献。

注意： 坐标和贡献（包括质量和惯量）显示在透视表输出的不同层上，缺省显示坐标。要显示贡献，在表上激活（双击）“层”下拉列表并从中选择“贡献”。

描述统计。 显示选定的变量的频率、缺失值的数量以及众数。

多重对应分析：保存

使用“保存”对话框可以将离散化数据、对象得分和转换后的值保存到当前会话中的外部 IBM SPSS Statistics 数据文件或数据集中。还可以将转换后的值和对象得分保存到活动数据集中。

- 数据集在当前会话期间可用，但在后续会话期间不可用，除非显式将其保存为数据文件。数据集名称必须符合变量命名规则。
- 文件名或数据集名称对于保存的每种数据类型都必须各不相同。
- 如果将对象得分或已转换的值保存到活动数据集中，则可以指定多名义维数的数量。

多重对应分析：对象图

使用“对象图”对话框可以指定所要的图类型和要绘制的变量。

对象点。 显示对象点的图。

对象和质心（双标图）。 对象点用变量质心绘制。

双标图变量。 可选择将所有变量用于双标图或选择一个子集。

标签对象。 可选择用选定的变量的类别（可在“选项”对话框中选择类别指示符值或值标签）或其个案号标注对象。如果选择**变量**，则为每个变量生成一个图。

多重对应分析：变量图

使用“变量图”对话框可以指定所要的图类型和要绘制的变量。

类别图。 对于选定的每个变量绘制质心坐标图。类别在特定类别中的对象的质心中。

联合类别图。 这是每个选定的变量的质心坐标的单个图。

转换图。 显示最优类别量化与类别指示符的比较图。可指定维数；将为每一维分别生成一个图。还可以选择显示每个选定的变量的残差图。

区分测量。 为选定的变量生成区分测量的单个图。

MULTIPLE CORRESPONDENCE 命令附加功能

如果将选中的内容粘贴到语法窗口中，并编辑生成的 MULTIPLE CORRESPONDENCE 命令语法，则可以定制多重对应分析。使用命令语法语言还可以：

- 将已转换的变量、对象得分和近似值保存到活动数据集时（使用 SAVE 子命令），为其指定根名称。
- 分别为每个图指定最大标签长度（使用 PLOT 子命令）。
- 为残差图指定单独的变量列表（使用 PLOT 子命令）。

请参阅命令语法参考以获取完整的语法信息。

第 7 章 多维刻度 (PROXSCAL)

多维刻度尝试查找对象间的一组近似值度量的结构。该过程是通过将观察值分配到概念低维度空间中的特定位置实现的，这样使空间中的点之间的距离尽可能与给定的（不）相似性相匹配。结果是一个该低维空间中的对象的最小平方表示形式，这在许多个案中将帮助您进一步理解您的数据。

示例。 多维刻度对于确定感知关系很有用。例如，在考虑产品形象时，可进行调查以获取描述您的产品与竞争对手产品的感知相似性（或近似性）的数据集。通过使用这些近似性和自变量（如价格），您可尝试确定哪些变量对于人们如何看待这些产品至关重要，以对产品形象做出相应的调整。

统计和图。 迭代历史记录、应力测量、应力分解、公共空间坐标、最终配置中的对象距离、私有空间权重、私有空间、转换后的近似值、转换后的自变量、应力图、公共空间散点图、私有空间权重散点图、私有空间散点图、转换图、Shepard 残差图，以及自变量转换图。

多维刻度数据注意事项

数据。 数据可以以近似值矩阵或转换为近似值矩阵的变量的形式提供。矩阵的格式可以为单列，也可以为多列。近似值可以在比率、区间、有序或样条刻度级别上进行处理。

假设。 至少必须指定三个变量。维数不能超过对象数减一。如果与多随机起点组合，则省略维数减少。如果仅指定一个源，则所有模型都等同于恒等模型；因此，该分析缺省设置为恒等模型。

相关过程。 调整所有数值级别变量对应于标准多维刻度分析。

获取多维刻度

1. 从菜单中选择：

分析 > 刻度 > 多维刻度 (PROXSCAL)...

这将打开“数据格式”对话框。

2. 指定数据的格式：

数据格式。 指定您的数据是由近似值度量组成，还是您要从数据中创建近似值。

源的数目。 如果数据是近似值，则指定近似值度量是单源还是多源。

一个源。 如果有一个近似值源，则指定数据集是格式化为跨列矩阵中的近似值，还是单列中的近似值（单列中有两个单独的变量用于标识每个近似值的行和列）。

- **近似值在跨列矩阵中** (*The proximities are in a matrix across columns*). 近似值矩阵跨与对象同数量的列分布。这生成“跨列矩阵”对话框中的近似值。
- **近似值位于单列中** (*The proximities are in a single column*). 近似值矩阵拼并到单个列或变量中。需要两个附加变量，以识别每个单元格的行和列。这生成“单列”对话框中的近似值。

多个源。 如果有多个近似值源，则指定数据集是格式化为跨列堆积矩阵中的近似值（在多个列中，每列一个源）还是单列中的近似值。

- **近似值位于跨列的堆积矩阵中** (*The proximities are in stacked matrices across columns*). 近似值矩阵跨与对象同数量的列分布，并跨行相互堆积（行数等于对象的数量乘以源的数量）。这生成“跨列矩阵”对话框中的近似值。

- 近似值位于多列中，每列一个源 (*The proximities are in columns, one source per column*). 近似值矩阵拼并到多个列或变量中。需要两个附加变量，以识别每个单元格的行和列。这生成“多列”对话框中的近似值。
- 近似值累积在单列中 (*The proximities are stacked in a single column*). 近似值矩阵拼并到单个列或变量中。需要三个附加变量，以识别行、列和每个单元格的源。这生成“单列”对话框中的近似值。

3. 单击定义。

矩阵中的跨列近似值

如果在“数据格式”对话框中为单个源或多个源选择了矩阵数据模型中的近似值，那么执行以下操作：

1. 选择三个或更多近似值变量。（确保列表中的变量顺序与近似值的列顺序一致。）
2. 或者，选择与近似值变量数相等的权重变量数。（确保权重的顺序与它们所加权的近似值的顺序一致。）
3. 或者，如果存在多个源，则选择源变量。（每个近似值变量中的个案数应等于近似值变量的数目乘以源的数目。）

此外，还可以定义多维刻度模型、对公共空间施加约束、设置收敛性标准、指定要使用的初始配置以及选择图和输出。

多列中的近似值

如果在“数据格式”对话框中为多个源选择了多列模型，那么执行以下操作：

1. 选择两个或更多近似值变量。（每个变量被假设为来自单独源的近似值矩阵。）
2. 选择一个行变量以定义每个近似值变量中近似值的行位置。
3. 选择一个列变量以定义每个近似值变量中近似值的列位置。（近似值矩阵中没有得到行/列指定项的单元格被视为缺失。）
4. 或者，选择与近似值变量数相等的权重变量数。

此外，还可以定义多维刻度模型、对公共空间施加约束、设置收敛性标准、指定要使用的初始配置以及选择图和输出。

单列中的近似值

如果在“数据格式”对话框中为单个源或多个源选择了单列模型，那么执行以下操作：

1. 选择一个近似值变量。（它被视作一个或多个近似值矩阵。）
2. 选择一个行变量以定义近似值在近似值变量中的行位置。
3. 选择一个列变量以定义近似值在近似值变量中的列位置。
4. 如果存在多个源，请选择源变量。（对于每个源，近似值矩阵中没有得到行/列指定项的单元格被视为缺失。）
5. 或者，选择一个权重变量。

此外，还可以定义多维刻度模型、对公共空间施加约束、设置收敛性标准、指定要使用的初始配置以及选择图和输出。

从数据中创建近似值

如果在“数据格式”对话框中选择从数据中创建近似值，那么执行以下操作：

1. 如果创建两个变量之间的距离（请参见“从数据中创建测量”对话框），则选择至少三个变量。这些变量将用于创建近似值矩阵（如果有多个源，则创建多个矩阵）。如果创建个案之间的距离，则只需要一个变量。
2. 如果存在多个源，请选择源变量。
3. 或者，选择一个用于创建近似值的测量。

此外，还可以定义多维刻度模型、对公共空间施加约束、设置收敛性标准、指定要使用的初始配置以及选择图和输出。

从数据中创建测量

多维刻度使用非相似性数据创建刻度分析解。如果您的数据为多变量数据（度量到的变量的值），就必须创建非相似性数据才能计算多维刻度解。可以指定从数据创建非相似性测量的详细信息。

度量。 允许您指定进行分析的非相似性测量。从与您的数据类型相对应的“度量”组选择一个选项，然后从与该类度量相对应的下拉列表选择一种度量。可以使用的选项有：

- **区间。** 欧氏距离、平方 Euclidean 距离、Chebychev、区组、Minkowski 或定制。
- **计数。** 卡方统计测量或 Phi 平方统计测量。
- **二元。** 欧氏距离、平方 Euclidean 距离、刻度差分、模式差分、方差或 Lance 和 Williams。

创建距离矩阵。 可以选择要分析的单位。选项有“变量之间”或“个案之间”。

转换值。 在某些情况下（例如，当用相差很大的刻度测量变量时），您会希望先将值标准化，然后再计算近似值（对二分类数据不可用）。从“标准化”下拉列表选择一种标准化方法（如果不需要标准化，请选择**无**）。

定义多维刻度模型

使用“模型”对话框可以指定刻度模型、其最大和最小维数、近似值矩阵的结构、要用于近似值的转换，以及近似值是在每个源中单独转换还是在源上无条件转换。

刻度模型。 选择下列选项之一：

- **恒等。** 所有源都具有相同的配置。
- **加权欧几里得。** 该模型是个体差异模型。每个源都具有一个私有空间，在该空间中公共空间的每个维都有不同的加权。
- **广义欧几里得。** 该模型是个体差异模型。每个源都具有私有空间（等于公共空间的一个旋转度，后跟各个维数的不同加权）。
- **减少的等级。** 此模型是广义欧几里得模型，您可以为其指定私有空间的等级。您必须指定大于等于 1 且小于最大维数的等级。

形状。 指定近似值应从近似值矩阵的下三角部分获取还是从上三角部分获取。可指定使用满矩阵，在此情况下，将分析上三角部分和下三角部分的权重总和。虽然将只用到指定的部分，但是任何情况下，都应该指定完整的矩阵，包括对角线。

近似值。 指定近似值矩阵是包含相似性测量还是非相似性测量。

近似值转换。 选择下列选项之一：

- **比率。** 转换后的近似值与原始近似值成比例。这只对于正值近似值是允许的。
- **区间。** 转换后的近似值与原始近似值成比例，外加一个截距项。该截距确保所有转换后的近似值都是值。

- **有序。** 转换后的近似值与初始近似值具有相同的顺序。指定结近似值应保持打结还是允许解开。
- **样条。** 转换后的近似值是初始近似值的平滑非递减的分段多项式转换。指定多项式的次数以及内部结点数。

应用转换。 指定是只有每个源中的近似值相互比较还是在源上进行无条件比较。

维。 缺省情况下，在两个维中（最小值 = 2，最大值 = 2）计算解。可选择最小为 1 最大为对象数减一的整数（只要最小值小于等于最大值）。该过程计算最大维数中的解，然后逐步减去维数，直至达到最低值。

多维刻度：约束

使用“约束”对话框可以对公共空间施加约束。

公共空间的约束。 指定约束类型。

- **无约束。** 对公共空间不施加约束。
- **某些固定坐标。** 选定的第一个变量包含对象在第一维上的坐标，第二个变量对应于在第二维上的坐标，依此类推。缺失值表示维上的坐标是自由的。选定的变量的数量必须等于所请求的最大维数。
- **自变量的线性组合。** 公共空间被限制为选定的变量的线性组合。

约束变量。 选择定义公共空间的约束的变量。如果指定线性组合，则为约束变量指定区间、名义、有序或样条转换。无论哪种情况，每个变量的个案数必须等于对象数。

多维刻度：选项

使用“选项”对话框可以选择初始配置样式、指定迭代和收敛性准则，以及选择标准更新或不严格的更新。

初始配置。 选择下列选项之一：

- **单形体。** 在最大维上，对象相互之间等距离放置。执行一次迭代以改进这种高维配置，随后执行降维操作以获取具有您在“模型”对话框中指定的最大维数的初始配置。
- **Torgerson。** 一个经典刻度解用作初始配置。
- **单随机起点。** 随机选择配置。
- **多随机起点。** 随机选择多个配置，并且标准化初始强度最低的配置用作初始配置。
- **定制。** 选择包含您自己的初始配置的坐标的变量。选定的变量的数量应等于指定的最大维数，并且第一个变量与维 1 上的坐标对应，第二个变量与维 2 上的坐标对应，依此类推。每个变量中的个案数应等于对象数。

迭代标准。 指定迭代标准值。

- **应力收敛性。** 当连续标准化初始强度值之间的差小于此处指定的数字（该数字必须介于 0.0 到 1.0 之间）时，算法将停止迭代。
- **最小应力。** 当标准化初始强度小于此处指定的数字（该数字必须介于 0.0 到 1.0 之间）时，算法将停止。
- **最大迭代次数。** 算法将执行指定的迭代次数（除非已先满足了上述某个条件）。
- **使用不严格的更新。** 不严格的更新将加速算法；这些更新不能与除恒等模型之外的其他模型一起使用，也不能与约束一起使用。

多维刻度：图，版本 1

使用“图”对话框可以指定将生成哪些图。如果您具有“多列中的近似值”数据格式，那么本主题将显示“图”对话框。对于私有空间权重、初始近似值对比转换后的近似值和转换后的近似值对比距离图，指定应为其生成图的源。可用源的列表为主对话框中的近似值变量列表。

应力。 生成标准化初始强度与维数的关系图。仅当维数的最大值大于维数的最小值时，才能生成此图。

公共空间。 显示公共空间坐标的散点图矩阵。

私有空间。 对于每个源，私有空间的坐标显示在散点图矩阵中。仅当在“模型”对话框中指定其中某个个体差异模型时才可能具有此行为。

私有空间权重。 生成私有空间权重的散点图。仅当在“模型”对话框中指定其中某个个体差异模型时才可能具有此行为。对于加权欧几里得模型，权重显示在图中，每个轴上有一维。对于广义欧几里得模型，每一维生成一个图，指示该维的旋转和权重。减少的等级模型生成与广义欧几里得模型相同的图，但是减少私有空间的维数。

初始近似值与转换后的近似值。 生成初始近似值与转换后的近似值的关系图。

转换后的近似值对比距离。 绘制转换后的近似值与距离。

转换后的自变量。 为自变量生成转换图。

变量和维数相关性。 显示自变量和公共空间维数之间的相关性的图。

多维刻度：图，版本 2

使用“图”对话框可以指定将生成哪些图。如果您的数据格式是除“多列中的近似值”之外的任何其他格式，那么本主题会显示“图”对话框。对于私有空间权重、初始近似值对比转换后的近似值和转换后的近似值对比距离图，指定应为其生成图的源。输入的源号必须是在主对话框中指定的源变量的值，并且范围必须是从 1 到源的数目。

多维刻度：输出

使用“输出”对话框可以控制显示的输出的数量并将其中的一部分保存到单独的文件中。

输出。 选择以下一项或多项进行显示：

- **公共空间坐标。** 显示公共空间坐标。
- **私有空间坐标。** 只有在模型不是恒等模型的情况下，才会显示私有空间的坐标。
- **私有空间权重。** 只有在指定了个体差异模型之一的情况下，才会显示私有空间权重。根据模型的不同，空间权重被分解为旋转权重和维权重，这两种权重也会显示。
- **距离。** 显示配置中对象之间的距离。
- **转换后的近似值。** 显示配置中对象之间的转换后的近似值。
- **输入数据。** 包括原始近似值，如果存在数据权重、初始配置和自变量的固定坐标，则还包括这些项。
- **随机起点的应力。** 显示每个随机起点的随机数种子以及标准化初始强度值。
- **迭代历史记录。** 显示主要算法的迭代历史记录。
- **多应力测量。** 显示不同的应力值。该表包含标准化初始强度值、Stress-I 值、Stress-II 值、S-Stress 值、离散考虑情况 (DAF) 值和同余 Tucker's 系数值。

- **应力分解。** 显示对象和源的最终标准化初始强度的分解，包括每个对象的平均值和每个源的平均值。
- **转换后的自变量。** 如果选择了线性组合约束，则会显示转换后的自变量和对应的回归权重。
- **变量和维数相关性。** 如果选择了线性组合约束，则会显示自变量和公共空间维数之间的相关性。

保存到新文件。 可将公共空间坐标、私有空间权重、距离、转换后的近似值以及转换后的自变量保存到单独的 IBM SPSS Statistics 数据文件中。

PROXSCAL 命令附加功能

如果将选中的内容粘贴到语法窗口中，并编辑生成的 PROXSCAL 命令语法，则可以定制近似值分析的多维刻度。使用命令语法语言还可以：

- 为转换和残差图指定单独的变量列表（使用 PLOT 子命令）。
- 为私有空间权重、转换和残差图指定单独的源列表（使用 PLOT 子命令）。
- 指定要显示的自变量转换图的子集（使用 PLOT 子命令）。

请参阅 [命令语法参考](#) 以获取完整的语法信息。

第 8 章 多维展开 (PREFSCAL)

“多维展开”过程尝试查找允许您直观地检查两组对象之间关系的公共定量刻度。

示例。 您已经让 21 个人按照各自的偏好对 15 种早餐食品进行排序，顺序为 1 到 15。使用多维展开，您可以确定他们从以下两个主要方面区分早餐食品：使用多维展开，您可以确定他们从以下两个主要方面区分早餐食品：软面包和硬面包之间，以及增肥食品和不增肥食品之间。

或者，您让一组驾驶员就 10 种属性对 26 款车型进行评分，分数为 6 分制，从 1=“完全不真实”到 6=“非常真实”。对个体取平均值后，得到的值被视为相似性。通过使用多维展开，您将找到相似模型的聚类以及与这些模型最密切相关的属性。

统计和图。 多维展开过程可生成迭代历史记录、应力测量、应力分解、公共空间坐标、最终配置中的对象距离、私有空间权重、私有空间、转换后的近似值、应力图、公共空间散点图、私有空间权重散点图、私有空间散点图、转换图和 Shepard 残差图。

多维展开数据注意事项

数据。 数据以矩形近似值矩阵的形式提供。每个列被视为一个单独的列对象。近似值矩阵的每一行被视为独立的行对象。当存在多个近似值源时，矩阵堆积。

假设。 至少必须指定两个变量。解中的维数不能超过对象数减一。如果仅指定一个源，则所有模型都等同于恒等模型；因此，该分析缺省设置为恒等模型。

获取多维展开

1. 从菜单中选择：

分析 > 刻度 > 多维展开 (PREFSCAL)...

2. 选择标识矩形近似值矩阵中的列的两个或更多变量。每个变量表示独立的列对象。
3. 或者，选择与列对象变量的数目相等的权重变量数。权重变量的顺序应与它们加权的列对象的顺序一致。
4. 或者，选择一个行变量。此变量的值（或值标签）用于在输出中标记行对象。
5. 如果存在多个源，也可以选择源变量。数据文件中的个案数应等于行对象的数目乘以源的数目。

此外，还可以定义多维展开模型、对公共空间施加约束、设置收敛性标准、指定要使用的初始配置以及选择图和输出。

定义多维展开模型

使用“模型”对话框可以指定刻度模型、其最大和最小维数、近似值矩阵的结构、要用于近似值的转换，以及近似值是在行上进行有条件转换、在源上进行有条件转换，还是在源上进行无条件转换。

刻度模型。 选择下列选项之一：

- **恒等。** 所有源都具有相同的配置。
- **加权欧几里得。** 该模型是个体差异模型。每个源都具有一个私有空间，在该空间中公共空间的每个维都有不同的加权。

- **广义欧几里得。** 该模型是个体差异模型。每个源都具有私有空间（等于公共空间的一个旋转度，后跟各个维数的不同加权）。

近似值。 指定近似值矩阵是包含相似性测量还是非相似性测量。

维。 缺省情况下，在两个维中（最小值 = 2，最大值 = 2）计算解。只要最小值小于等于最大值，就可以选择最小为 1 最大为对象数减 1 的整数。该过程计算最大维数中的解，然后逐步减去维数，直至达到最低值。

近似值转换。 选择下列选项之一：

- **无。** 不转换近似值。或者选择**包括截距**，在这种情况下，近似值可通过一个常数项移位。
- **线性。** 转换后的近似值与原始近似值成比例；即转换函数估算一个斜率，并且截距修正为 0。这又称为比率转换。或者选择**包括截距**，在这种情况下，近似值可通过一个常数项移位。这又称为区间转换。
- **样条。** 转换后的近似值是初始近似值的平滑非递减的分段多项式转换。可指定多项式的次数和内部结点数。或者选择**包括截距**，在这种情况下，近似值可通过一个常数项移位。
- **平滑。** 转换后的近似值与原始近似值有着相同的顺序，包括将后继值之间的差值考虑在内的约束。结果为“平滑有序”转换。可以指定结近似值应保持打结还是允许解开。
- **有序。** 转换后的近似值与初始近似值具有相同的顺序。可以指定结近似值应保持打结还是允许解开。

应用转换。 指定是只有每行中的近似值相互比较，还是只有每个源中的近似值相互比较，或者是在行或源上无条件进行比较；即，转换是逐行执行、逐源执行，还是一次性对所有近似值执行。

多维展开：约束

使用“约束”对话框可以对公共空间施加约束。

公共空间的约束。 可选择固定行和/或列对象在公共空间中的坐标。

行/列约束变量。 选择包含约束的文件，然后选择定义对公共空间的约束的变量。选定的第一个变量包含对象在第一维上的坐标，第二个变量对应于在第二维上的坐标，依此类推。缺失值表示维上的坐标是自由的。选定的变量的数量必须等于所请求的最大维数。每个变量的个案数必须等于对象数。

多维展开：选项

使用“选项”对话框可以选择初始配置样式，指定迭代和收敛性标准以及为应力设置惩罚项。

初始配置。 选择下列选项之一：

- **经典。** 矩形近似值矩阵用于补充完整对称 MDS 矩阵的内部块（行之间和列之间的值）。完整矩阵一旦形成，经典刻度解将用作初始配置。可使用三角不等性或 Spearman 距离通过插补法填充内部块。
- **Ross-Cliff。** Ross-Cliff 起点将双中心和正方形近似值矩阵的奇异值分解的结果用作行和列对象的初始值。
- **对应。** 对应起点使用对反向数据（相似性而非非相似性）的对应分析的结果，且对行和列的得分进行对称的标准化。
- **质心。** 该过程是通过使用特征值分解将行对象定位在配置中来启动的。然后列对象定位在指定选择项的质心。对于选择项的数目，请指定介于 1 和近似值变量的数目之间的正整数。
- **多随机起点。** 为多个随机选择的初始配置计算解，并且惩罚强度最低的解作为最佳解显示。
- **定制。** 可选择包含您自己的初始配置的坐标的变量。选定的变量的数量应等于指定的最大维数，并且第一个变量与维 1 上的坐标对应，第二个变量与维 2 上的坐标对应，依此类推。每个变量中的个案数应等于行和列对象合并的数目。行和列坐标应该是堆积的，列坐标在行坐标后面。

迭代标准。 指定迭代标准值。

- **应力收敛性。** 当连续惩罚强度值中的相对差值小于此处指定的数字（必须为非负）时，算法将停止迭代。
- **最小应力。** 当惩罚强度小于此处指定的数字（必须为非负）时，算法将停止。
- **最大迭代次数。** 算法将执行此处指定的迭代次数（除非已先满足了上述某一条件）。

惩罚项。 该算法尝试最小化惩罚强度，该惩罚强度是一个拟合优度测量，等于 Kruskal 的 Stress-I 和基于转换后的近似值的变异系数的惩罚项的乘积。使用这些控件可以设置惩罚项的强度和范围。

- **强度。** 强度参数的值越小，惩罚越强。请指定介于 0.0 和 1.0 之间的值。
- **范围。** 此参数设置惩罚有效的起始时刻。如果设置为 0.0，则惩罚无效。增加该值将使算法搜索转换后的近似值之间变差更大的解。请指定一个非负值。

多维展开：图

使用“图”对话框可以指定将生成哪些图。

图。 以下图是可用的：

- **多起点。** 显示惩罚强度的堆积直方图，该直方图同时显示应力和惩罚。
- **初始公共空间。** 显示初始公共空间的坐标的散点图矩阵。
- **每维的应力。** 生成惩罚强度对比维数的线图。仅当维数的最大值大于维数的最小值时，才能生成此图。
- **最终公共空间。** 显示公共空间坐标的散点图矩阵。
- **空间权重。** 生成私有空间权重的散点图。仅当在“模型”对话框中指定其中某个个体差异模型时才可能具有此行为。对于加权欧几里得模型，所有源的权重显示在一张图中，并且每个轴上一维。对于广义欧几里得模型，每一维生成一个图，指示该维对每个源的旋转和权重。
- **私有空间。** 显示每个源的私有空间坐标的散点图矩阵。仅当在“模型”对话框中指定其中某个个体差异模型时才可能具有此行为。
- **转换图。** 生成原始近似值对比转换后的近似值的散点图。根据应用转换的方式，将为每行或每个源分配不同的颜色。无条件转换将生成单色。
- **Shepard 图。** 原始近似值对比转换后的近似值和距离。距离由点表示，转换后的近似值由线表示。根据应用转换的方式，将为每行或每个源生成不同的线。无条件转换将生成一条线。
- **拟合散点图。** 显示转换后的近似值对比距离的散点图。如果指定多个源，则会为每个源指定单独的颜色。
- **残差图。** 显示转换后的近似值对比残差（转换后的近似值减去距离）的散点图。如果指定多个源，则会为每个源指定单独的颜色。

行对象样式。 这些样式使您可以进一步控制行对象在图中的显示方式。可选颜色变量的值用于循环选择所有颜色。可选标记变量的值用于循环选择所有可能的标记。

源图。 对于私有空间、拟合散点图和残差图（如果转换由源应用，则还包括转换图和 Shepard 图），可指定应为其生成图的源。输入的源号必须是在主对话框中指定的源变量的值，范围为从 1 到源的数目。

行图。 如果转换由行应用，则对于转换图和 Shepard 图，可指定应为其生成图的行。输入的行号必须在从 1 到行数的范围之内。

多维展开: 输出

使用“输出”对话框可以控制显示的输出的数量并将其中的一部分保存到单独的文件中。

输出。 选择以下一项或多项进行显示:

- **输入数据。** 包括原始近似值, 如果存在数据权重、初始配置和固定坐标, 则还包括这些项。
- **多起点。** 显示每个随机起点的随机数种子以及惩罚强度值。
- **初始数据。** 显示初始公共空间的坐标。
- **迭代历史记录。** 显示主要算法的迭代历史记录。
- **拟合测量。** 显示不同的测量。该表包含多个拟合优度、拟合劣度、相关性、变差和非退化度量。
- **应力分解。** 显示惩罚强度的对象、行和源分解, 包括行、列和源平均值以及标准差。
- **转换后的近似值。** 显示转换后的近似值。
- **最终公共空间。** 显示公共空间坐标。
- **空间权重。** 显示私有空间权重。仅当指定其中某个个体差异模型时, 该选项才可用。根据模型的不同, 空间权重被分解为旋转权重和维权重, 这两种权重也会显示。
- **私有空间。** 显示私有空间的坐标。仅当指定其中某个个体差异模型时, 该选项才可用。
- **拟合距离。** 显示配置中对象之间的距离。

保存到新文件。 可将公共空间坐标、私有空间权重、距离以及转换后的近似值保存到不同的 IBM SPSS Statistics 数据文件中。

PREFSCAL 命令附加功能

如果将选中的内容粘贴到语法窗口中, 并编辑生成的 PREFSCAL 命令语法, 则可以定制近似值分析的多维展开。使用命令语法语言还可以:

- 在有多个源可用的情况下, 为私有空间、拟合散点图和残差图指定多源列表, 并且在矩阵有条件转换的情况下, 还可以为转换图和 Shepard 图指定多源列表 (使用 PLOT 子命令)。
- 在有条件转换的情况下, 为转换图和 Shepard 图指定多行列表 (使用 PLOT 子命令)。
- 指定多行而不是行 ID 变量 (使用 INPUT 子命令)。
- 指定很多源而不是源 ID 变量 (使用 INPUT 子命令)。

请参阅命令语法参考以获取完整的语法信息。

声明

本信息是为在美国提供的产品和服务编写的。

IBM 可能在其他国家或地区不提供本文中讨论的产品、服务或功能特性。有关您当前所在区域的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。不过，用户应自行负责评估和验证任何非 IBM 产品、程序或服务的工作情况。

IBM 公司可能已拥有或正在申请与本文档内容有关的各项专利。您获得该文档并不意味着授予您任何这些专利许可。您可以将书面许可查询函件发送至：

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

有关双字节 (DBCS) 信息的许可查询，请与您所在国家或地区的 IBM 知识产权部门联系，或用书面方式将查询寄往：

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

本条款不适用英国或任何这样的条款与当地法律不一致的国家或地区：International Business Machines Corporation“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某种特定用途的保证。有些州/省不允许针对某些交易的明示或暗示免责条款，因此本声明可能不适用于您。

本信息可能含有技术误差或排版错误。此处的信息会定期进行更改；这些更改会体现在本出版物的新版本中。IBM 可以随时对本资料中描述的产品和/或程序进行改进和/或更改，而不另行通知。

本信息中对任何非 IBM Web 站点的引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。

本程序的被许可方如果要了解有关程序的信息以达到如下目的：（i）允许在独立创建的程序和其他程序（包括本程序）之间进行信息交换，以及（ii）允许对已经交换的信息进行相互使用，请与下列地址联系：

IBM Software Group
ATTN: Licensing

200 W. Madison St.
Chicago, IL; 60606
U.S.A.

此类信息的提供应遵照相关条款和条件，其中包括在某些情况下支付适当费用。

本资料中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际软件许可协议或任何同等协议中的条款提供。

此处所含的性能数据均在受控环境下决定。因此，在其他操作环境中获得的结果可能差异较大。有些测量可能在开发级的系统中进行，不保证这些测量结果与常用系统上的测量结果相同。此外，有些测量结果可能通过推断来估计得出。实际结果可能有所差异。此文档的用户应针对其具体环境验证适用的数据。

非 IBM 产品的相关信息来自这些产品的供应商，及其发布的公告或其他公开来源。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

有关 IBM 未来方向或意向的所有声明均可能未经通知即变更或撤销，并且仅代表目标和目的。

本信息包含日常业务运营中使用的数据和报告的示例。为了尽可能详尽地对其进行说明，示例中包含了人员的姓名、公司、品牌和产品的名称。所有这些名称均为虚构，与真实商业企业使用的名称和地址的任何雷同纯属巧合。

版权许可：

本信息包括源语言形式的样本应用程序，这些样本说明不同操作平台上的编程方法。如果是为按照在编写样本程序的操作平台上的应用程序编程接口 (API) 进行应用程序的开发、使用、经销或分发为目的，您可以任何形式对这些样本程序进行复制、修改、分发，而无须向 IBM 付费。这些示例并未在所有条件下作全面测试。因此，IBM 不能担保或暗示这些程序的可靠性、可维护性或功能。本样本程序仍然是“按现状”提供的，不附有任何种类的保证。对于因使用样本程序所引起的任何损害，IBM 概不负责。

凡这些实例程序的每份拷贝或其任何部分或任何衍生产品，都必须包括如下版权声明：

©（贵公司的名称）（年）。此部分代码是根据 IBM Corp. 公司的样本程序衍生出来的。

© Copyright IBM Corp. _（输入年份）_. All rights reserved.

商标

IBM、IBM 徽标和 ibm.com 是 International Business Machines Corp.，在全球许多管辖区域注册的商标或注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。当前的 IBM 商标列表，可从 Web 站点 www.ibm.com/legal/copytrade.shtml 上“版权和商标信息”部分获取。

Adobe、Adobe 徽标、PostScript 和 PostScript 徽标是 Adobe Systems Incorporated 在美国和/或其他国家或地区的注册商标或商标。

Intel、Intel 徽标、Intel Inside、Intel Inside 徽标、Intel Centrino、Intel Centrino 徽标、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 是 Intel Corporation 或其子公司在美国和其他国家或地区的商标或注册商标。

Linux 是 Linus Torvalds 在美国、其他国家或地区或两者的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国、其他国家或地区或两者的商标。

UNIX 是 The Open Group 在美国和其他国家或地区的注册商标。

Java 和所有基于 Java 的商标和徽标是 Oracle 及/或其分支机构的商标和注册商标。

索引

[B]

- 变量权重
 - 在“多重对应分析”中 29
 - 在“分类主成份分析”中 16
- 标准化
 - 在“对应分析”中 26
- 不严格的更新
 - 在多维刻度中 36
- 补充对象
 - 在“分类回归”中 11

[C]

- 残差图
 - 在“多维展开”中 41
- 惩罚项
 - 在“多维展开”中 40
- 成份载荷
 - 在“非线性正态协变量分析”中 22
 - 在“分类主成份分析”中 18
- 成份载荷图
 - 在“分类主成份分析”中 19
- 初始公共空间图
 - 在“多维展开”中 41
- 初始配置
 - 在多维刻度中 36
 - 在“多维展开”中 40
 - 在“非线性正态协变量分析”中 22
 - 在“分类回归”中 11

[D]

- 迭代标准
 - 在多维刻度中 36
 - 在“多维展开”中 40
- 迭代历史记录
 - 在多维刻度中 37
 - 在“多重对应分析”中 31
 - 在“多维展开”中 42
 - 在“分类主成份分析”中 18
- 对象得分
 - 在“多重对应分析”中 31
 - 在“非线性正态协变量分析”中 22
 - 在“分类主成份分析”中 18
- 对象点图
 - 在“多重对应分析”中 32
 - 在“分类主成份分析”中 19
- 对应分析 25, 26, 27
 - 命令附加功能 28
 - 统计 25

- 对应分析 (续)
 - 图 25
- 多重对应分析 29, 30
 - 保存变量 31
 - 命令附加功能 32
 - 最优刻度级别 29
- 多起点图
 - 在“多维展开”中 41
- 多维刻度 33, 34, 35
 - 命令附加功能 38
 - 模型 35
 - 输出 37
 - 统计 33
 - 图 33, 37
 - 选项 36
 - 约束 36
- 多维展开 39
 - 公共空间的约束 40
 - 命令附加功能 42
 - 模型 39
 - 输出 42
 - 统计 39
 - 图 39, 41
 - 选项 40

[F]

- 方差解释情况
 - 在“分类主成份分析”中 18
- 非线性正态协变量分析 21, 22
 - 命令附加功能 23
 - 统计 21
 - 图 21
- 分类回归 9
 - 保存 12
 - 规则化 11
 - 命令附加功能 13
 - 统计 9
 - 图 9
 - 最优刻度级别 10
- 分类主成份分析 15, 17
 - 保存变量 18
 - 命令附加功能 20
 - 自动抽样 19
 - 最优刻度级别 16
- 分箱化
 - 在“多重对应分析”中 30
 - 在“分类回归”中 10
 - 在“分类主成份分析”中 16
- 复相关系数
 - 在“分类回归”中 12

[G]

- 公共空间的约束
 - 在“多维展开”中 40
- 公共空间图
 - 在多维刻度中 37
 - 在“多维展开”中 41
- 公共空间坐标
 - 在多维刻度中 37
 - 在“多维展开”中 42
- 惯量
 - 在“对应分析”中 27
- 广义欧几里得模型
 - 在“多维展开”中 39

[H]

- 恒等模型
 - 在“多维展开”中 39
- 回归系数
 - 在“分类回归”中 12

[J]

- 加权欧几里得模型
 - 在“多维展开”中 39
- 近似值转换
 - 在“多维展开”中 39
- 距离
 - 在多维刻度中 37
 - 在“多维展开”中 42
- 距离测量
 - 在“对应分析”中 26

[K]

- 刻度模型
 - 在“多维展开”中 39
- 空间权重图
 - 在“多维展开”中 41

[L]

- 类别量化
 - 在“多重对应分析”中 31
 - 在“非线性正态协变量分析”中 22
 - 在“分类回归”中 12
 - 在“分类主成份分析”中 18
- 类别图
 - 在“多重对应分析”中 32

类别图 (续)
在“分类主成份分析”中 19
联合类别图
在“多重对应分析”中 32
在“分类主成份分析”中 19

[M]

描述统计
在“分类回归”中 12

[N]

拟合
在“非线性正态协变量分析”中 22
拟合散点图
在“多维展开”中 41

[Q]

区分测量
在“多重对应分析”中 31
区分测量图
在“多重对应分析”中 32
权重
在“非线性正态协变量分析”中 22
缺失值
在“多重对应分析”中 30
在“分类回归”中 10
在“分类主成份分析”中 16

[S]

三标图
在“分类主成份分析”中 19
双标图
在“对应分析”中 27
在“多重对应分析”中 32
在“分类主成份分析”中 19
私有空间权重
在多维刻度中 37
在“多维展开”中 42
私有空间权重图
在多维刻度中 37
在“多维展开”中 41
私有空间图
在多维刻度中 37
在“多维展开”中 41
私有空间坐标
在“多维展开”中 42

[T]

弹性网络
在“分类回归”中 11
套索
在“分类回归”中 11
投影质心图
在“分类主成份分析”中 19
图
在多维刻度中 37
在“对应分析”中 27
在“非线性正态协变量分析”中 22
在“分类回归”中 12

[W]

维度
在“对应分析”中 26

[X]

相关性
在多维刻度中 37
相关性矩阵
在“多重对应分析”中 31
在“分类主成份分析”中 18
相关性图
在多维刻度中 37

[Y]

应力测量
在多维刻度中 37
在“多维展开”中 42
应力图
在多维刻度中 37
在“多维展开”中 41
约束
在多维刻度中 36

[Z]

质心
在“非线性正态协变量分析”中 22
置信统计
在“对应分析”中 27
转换后的近似值
在多维刻度中 37
在“多维展开”中 42
转换后的自变量
在多维刻度中 37
转换图
在多维刻度中 37
在“多重对应分析”中 32
在“多维展开”中 41

转换图 (续)
在“分类主成份分析”中 19
自助抽样
分类主成份分析 19
最优刻度级别
在“多重对应分析”中 29
在“分类主成份分析”中 16
最终公共空间图
在“多维展开”中 41

A

ANOVA
在“分类回归”中 12

P

PREFSCAL 39

R

Ridge 回归
在“分类回归”中 11

S

Shepard 图
在“多维展开”中 41



Printed in China