

预测结果和揭示分类数据中的关系

通过预测分析、统计学习、启发性的概念映射、最优尺度、偏好尺度和数据降维技术，揭示数据中全部潜在的关系。SPSS Categories为您提供深入分析复杂的分类数据和高维数据的全部工具。

例如，消费者会认为哪种特征与您的产品或者品牌联系最紧密，或者比起您或您的竞争对手提供的其他产品，确定消费者对您的某种产品的理解。

利用SPSS Categories，当结果和预测变量都是数值的、有序的或者名义的时候，您可以做回归处理，还可以可视化地解释数据，在得分、计数、分级、秩或者相似表中查看行和列是如何相关的。这使您可以：

- 通过类似传统的回归、主成分分析、和典型相关分析，处理和了解序次及名义数据
- 处理数值数据中的非正态残差，或者预测变量和结果变量之间的非现行关系。为数值和类别数据使用Ridge回归、Lasso、Elastic Net、变量选择、模型选择这些选项
- 在相关分析中使用二维图或者三维图表示对象（示例）、类别、变量集合之间的关系
- 在概念映射中把一个或者两个对象结合之间的相似性表示为距离

把您的定性数据转化为定量数据

SPSS Categories中的高级程序使您能够实现更多形式的分类数据的统计分析。

利用SPSS Categories的最优尺度您可以指定测量单位以及分类数据中的零点。通过允许您分析具有混合测量水平的变量（例如名义、序次、和数值变量），为您提供一套新的统计分析功能。

SPSS Categories的对应分析和多重对应分析程序帮助您数字化地评估两个或多于两个名义变量的相似性。并且，利用主成分分析程序，您可以根据重要成分来描述数据。或者把不同测量水平的变量合并为一组，然后进行非线性典型相关分析。

图形化展示数据中的潜在关系

无论您研究的分类数据是什么类型，例如市场细分、亚文化研究、政党研究、或者生物物种 – SPSS Categories的最优尺度程序使您不再受限于二维的表格，帮您在更大的框架中展现数据变量之间的关系。您不仅仅能够看到统计报告，您也可以看到数据映射。

SPSS Categories数据降维技术使您不拘泥于复杂的表格分析形式。相反，您可以利用概念映射和双标图阐明数据之间的关系。

- 高解析度的概念映射图能够图形化地展现相似的变量或相近的类别。它们为您提供独特的洞察力来了解两个或多于两个类别的分类变量之间的关系
- 双标图帮助您了解观测、变量和类别之间的关系。例如，您可以了解产品、顾客和人口统计特征之间的关系。现在，利用新增的偏好尺度分析，您能够进一步可视化地展现对象之间的关系。这一程序所基于的突破性算法使您能够实现序次变量的非度量性多维标度分析，并得到有意义的结果。

如何使用SPSS Categories

以下程序可为您的数据分析增添有意义的结果：

分类回归（CATREG）



利用最优尺度的技术从一组分类预测因子中预测名义、序次或数值变量的取值。您可以利用最优尺度回归描述诸如工作满意度如何与工作种类、地域和工作相关的差旅行程相关联等问题。

对应分析 (CORRESPONDENCE)

使您能够分析行列包含对应量度的二维表格。最常见的对应表格就是单元格包含频数计数的列联表。

SPSS Categories通过概念映射图描述名义变量之间的关系，也能够直观展现变量不同类别之间的关系。

多重对应分析 (MULTIPLE CORRESPONDENCE)

用来分析多类别（大于等于2）的分类数据。程序类似对应分析，但不局限于两个变量。所有变量在多重对应分析中都是作为名义变量进行分析的。

例如，您可以利用多重对应分析来探察最受欢迎的电视节目秀、年龄组和性别之间的相关性。通过SPSS Categories生成的低维映射图，您可以了解每个节目秀被哪些群体喜欢，同时了解哪些节目秀是最相似的。

分类的主成分分析 (CATPCA)

运用最优尺度法来一般化主成分分析，以涵盖具然而，每个评判人员或许用单独的标准进行评判。这种情况下，每个人的评判构成一组，OVERALS在移除相关性后基于不同标准进行平均，然后合并不同评判人员的打分。

多维尺度 (PROXSCAL)

可进行一个和多个含有相同和相异矩阵的多维尺度分析。另外，您也可以输入PROXSCAL的多维数据观测之间的距离。

PROXSCAL可在低维图上显示距离，您可以获得对象关系的“空间”理解。在多重相似矩阵的情形，PROXSCAL将分析它们的共同度，并以图形展现它们之间的差异。

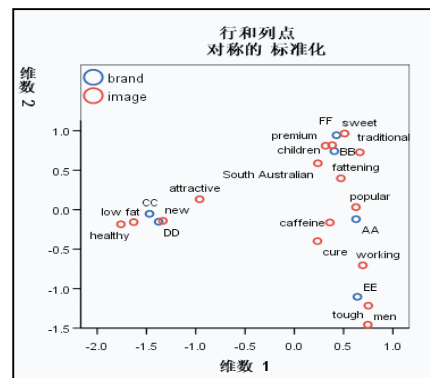
例如，PROXSCAL可展现不同可乐口味在不同年龄层之消费者的偏好。您也许会发现青年人更注重传统和新口味之间的区别，而成人则注重是否是低热量的可乐。

偏好尺度 (PREFSCAL)

以可视化的方式来检视变量之间的关系。偏好尺度通过多维展现的方法，找出两个对象集间共同的数量度量尺度。可帮助您找出变量之群集。例如，一些驾驶员根据十种特性和六个等级来评价二十六种车子，您可以找出种类相似的车子集合，以及它们彼此的相关特性。

更好地理解顾客感知

南澳大利亚的市场研究人员想更好地理解顾客对六种冰咖啡的品牌认知。他们请顾客根据十六种不同的分类属性对每个品牌进行评分（下图中品牌以AA到FF标识）。



研究人员研究消费者南澳大利亚销售的六个品牌冰咖啡的顾客认知。品牌以AA到FF表示，不同特点以分类属性表示，如“healthy”。此图为SPSS Categories中对应分析生成的对应图

96个单元格的表格对于分析人员来说，很难清楚的了解品牌和认知的属性之间的关系。市场研究人员利用SPSS中的对应分析程序，发现品牌和属性之间隐含着两种非常强的关系。通过在每个维度内给每个品牌和属性赋予具体数值，相关信息就通过易于理解的图表（通常称为概念影视图）展示出来。例如，您可以清楚地看到品牌AA被认知为“最受市场欢迎”的。类似地，研究人员轻易地发现CC和DD品牌多数被认知为健康及低脂肪的。而FF则被认知为营养丰富、甜的、并且是质优价高的品牌。

*数据和例子来源： Kennedy, R., C. Riquier, and Byron Sharp. 1996. “Practical Applications of Correspondence Analysis to Categorical Data in Market Research,” Journal of Targeting, Measurement and Analysis for Marketing, Vol. 5, No. 1, pp. 56-70

功能

统计

CATREG

- 通过最优尺度进行分类回归
 - 指定您要分析的每个变量最优尺度水平
 - 可选择：有序样条（单调的）、名义样条（非单调的）、有序、名义、多重名义、或者数值
 - 连续变量离散化或者把字符串转换成整数值变量，可以通过乘法、秩或分组实现。您可根据选择的分布（如正态、均匀），按照预先设定的类别数/预先选定的区间，将数值分组。秩和分组选项也可以用来对数据重新分类
 - 指定缺失值处理方式：利用众数/额外类别填充或列删
 - 指定补充对象
 - 指定初始值的计算方法
 - 控制迭代次数
 - 指定收敛标准
 - 图形输出结果，可以输出：
- 转换图(与原始分类数值对应的优化量表)
- 残差图
 - 存储到工作文件：变换后的变量、预测值、残差
 - 打印结果，包括：
- 多重R、 R^2 、调整 R^2
- 标准化回归系数、标准误、零阶顺序相关、部分相关、偏相关、转换后预测值的Pratt相对重要性测度、转换前后的容限、F统计量

- 描述统计量表格，包括边际数、转换类型、缺失值数量、众数
- 迭代历史
- 显示拟合值及模型参数的表格：根据最优尺度水平得到的带有自由度信息的ANOVA表、最优尺度的模型概述表（包括调整的 R^2 ，T值和显著性水平）、包括零阶顺序相关，部分和偏相关以及变换前后的重要性的容限的表格
- 转换后预测值之间及特征值之间的相关性
- 原始的预测值之间及特征值之间的相关性
- 分类量化
 - 把离散化及转换后的数据写入外部文件
- 三种新的规则化方法：Ridge回归、Lasso、Elastic Net
 - 通过稳定参数估计改进预测精度
 - 分析高维数据（变量比对象多）
 - 从预测集中得到自动变量选择
 - 把规则化模型和参数写到一个新的数据集中以备后用
- 两种新的模型选择和预测精度评估方法： $.632$ bootstrap和交叉验证（CV）
 - 用 $.632$ bootstrap和交叉验证（CV）发现最优的预测模型
 - 用bootstrap获取系数的标准误差的非参数估计

- Systematic multiple start
 - 当包含单调变换的时候发现全局最优解
 - 把回归系数的符号写到一个新的数据集中以备重用。

CORRESPONDENCE

- 对应分析
 - 输入数据可以为个案文件或表格文件
 - 指定解决方案的维数
 - 两种距离度量方法：对应分析的 χ^2 距离、或者双标图的欧式距离
 - 五种标准化方法：行和列均值已删除、行均值已删除、列均值已删除、使行总和相等，删除均值、使列总和相等，删除均值
 - 五种规范化方法：对称、主要行、主要列、主要、自定义
 - 打印结果，包括：
- 对应表
- 摘要表格：奇异值、惯量、由维度计算的惯量比例及其累积、最大维度的置信统计量、行点概览、列点概览
- 行列点概览：质量、得分、惯量、点对维惯量的贡献、维对点惯量的贡献
- 行列置信度统计量：当前行列点的标准偏差和相关性
- 行列置信度统计量：当前行列点的标准偏差和相关性
- 置换表格：给定维度的按行列得分排序的行列表格

- 结果图表：行得分、列得分、双标图（行列得分的联合图）
- 把行得分、列得分、置信统计量（方差和协方差）写入到外部数据文件

MULTIPLE CORRESPONDENCE

- 多重对应分析（该过程取代了SPSS 13.0版本之前的HOMALS）
- 指定权重变量
- 连续变量离散化或者把字符串转换成整数变量，可以通过乘法、秩或分组实现。您可根据选择的分布（如正态、均匀），按照预先设定的类别数/预先选定的区间，将数值分组。秩和分组选项也可以用来对数据重新分类
- 指定缺失值的处理方式。仅排除数据矩阵中含有无效值的单元格、利用众数/额外类别填充或列删
- 指定对象或标量作为补充值（类别中的补充对象也会在结果中输出）
- 指定解决方案的维数
- 生成包含某配置下的坐标分量以及在此约束配置下得出的拟合变量文件
- 五种正则化方法：以变量为主（最优化变量之间的关联性），以对象为主（最优化对象之间的距离）、对称化（最优化对象和变量之间的关联性）、独立性、或者自定义（用户指定可应用于以变量或对象为主的数值）

- 控制迭代次数
- 指定收敛准则
- 打印结果，包括
- 模型概述
- 迭代统计量和历史
- 描述统计量（频数、缺失值、众数）
- 根据变量和维度的区分测量
- 类别量化（中心坐标）、质量、类别的惯量、类别对维惯量的贡献、维对类别惯量的贡献
- 转换后预测值之间及特征值之间的相关性
- 原始的预测值之间及特征值之间的相关性
- 对象得分
- 对象贡献量：质量、惯量、对象对维惯量的贡献、维对对象惯量的贡献
- 打印结果，包括：
- 类别图：类别点、转换图(与原始分类数值对应的优化量化表)、选定变量的残差图、选定变量的联合类别图
- 对象得分
- 区分测量
- 选定变量的对象和质心双标图
- 把转换的变量和对象得分添加到外部文件
- 把离散化的数据、转换的数据、和得分写入到外部文件

CATPCA

- 利用最优尺度方法实现分类变量的主成分分析
- 指定要分析的每个变量的最优尺度水平可选择：有序样条（单调的）、名义样条（非单调的）、序、名义、多重名义、或者数值
- 指定权重变量
- 连续变量离散化或者把字符串转换成整数变量，可以通过乘法、秩或分组实现。您可根据选择的分布（如正态、均匀），按照预先设定的类别数/预先选定的区间，将数值分组。秩和分组选项也可以用来对数据重新分类
- 指定缺失值处理方式：仅排除数据矩阵中含有无效值的单元格、利用众数/额外类别填充或列删
- 指定缺失值的处理方式。仅排除数据矩阵中含有无效值的单元格、利用众数/额外类别填充或列删
- 指定对象或标量作为补充值（类别中的补充对象也会在结果中输出）
- 指定解决方案的维数
- 生成包含某配置下的坐标分量以及在此约束配置下得出的拟合变量文件
- 五种正则化方法：以变量为主（最优化变量之间的关联性），以对象为主（最优化对象之间的

距离)、对称化(最优化对象和变量之间的关联性)、独立性、或者自定义(用户指定可以变量或对象为主的数值)

- 控制迭代次数

- 指定收敛准则

- 打印结果, 包括:

■ 模型概述

■ 迭代统计量和历史

■ 描述统计量(频数、缺失值、众数)

■ 被变量和维度解释的方差

■ 类别量化以及每一维度类别坐标(向量或者质心坐标)

■ 转换后变量间相关系数、特征值的相关系数矩阵原始变量间相关系数、特征值的相关系数矩阵

■ 对象(主成分)得分

- 打印结果, 包括:

■ 类别图: 类别点、转换图(与原始分类数值对应的优化量化表)、选定变量的残差图、选定变量的联合类别图

■ 对象(主成分)得分图

■ 因子载荷图

PREFSCAL

■ 直观地展现两组对象集内变量之间的关系, 以便找到共同的量化尺度

- 读入一个或更多长方形相似矩阵

- 读入权重、初始配置、坐标约束

- 选择多种方法转换近似值: 线性、有序、平滑、或样条函数

- 指定多维展现的方法: 标识、加权欧几里德、或者广义欧几里德

- 指定行坐标的约束、列坐标的约束

- 指定初始配置(经典: 归因依据包括三角形、Spearman, Ross-Cliff, 对应、质心、多随机起点、定制)、迭代标准、惩罚项

- 绘制多起点图、初始公共空间、每维的应力、最终的公共空间、空间权重、私有权重、拟合散点图、转换图、Shepard图

- 指定输出, 可包括输入数据、多起点、初始公共空间、迭代历史、拟合度量、应力分解、最终公共空间、空间权重、私有空间、拟合距离、转换近似值

- 可以把公共空间坐标、私有空间权重、距离、转换的近似值保存到新文件

系统要求

■ 软件: SPSS Statistics Base 17.0

■ 其它的系统需求根据平台的不同而异

