

轻松快速地识别群体及预测结果

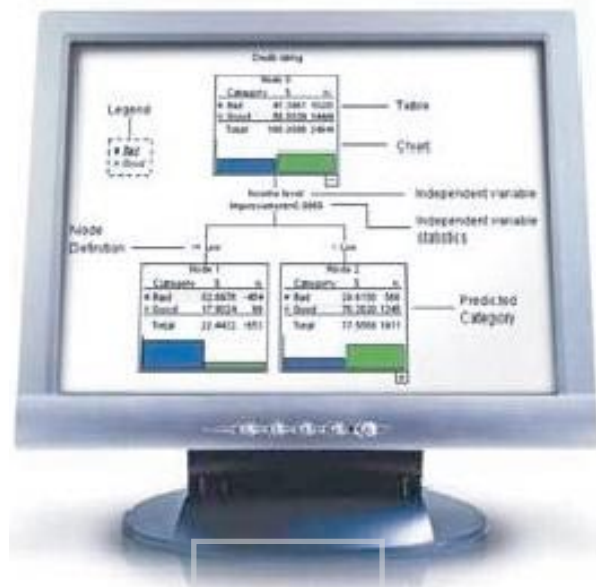
SPSS Decision Trees附加模块，能够在SPSS环境下直接创建分类决策树，帮助您快速并准确地识别群体，发现群体之间的关系并预测未来事件。您可应用分类决策树于分段、分层、预测、数据降维、变量筛选、类别合并、以及连续变量离散化。

高度形象化的图解使您能够以非常直观的方式展现分类结果 – 这样您可以更加清楚地把分类结果解释给业务人员。这些树方便您探索结果，并直观地确定模型是如何展开的。直观地结果能够帮助您找出具体的子群以及通过传统的统计方法难以发现的关系。因为分类树可以把数据分到不同的枝和节点，您可以轻松直观地看到某个群体是在何处被分开的及在何处终结的。

SPSS Decision Trees广泛应用于多个领域：

■ 数据库营销

- 选择一个响应变量对客户细分（回复/未回复邮件的；高-、中-、和低利润客户；营销目标囊括那些延长服务期的客户）
- 基于其它属性概括客户群，如人口统计学或客户活动
- 针对某具体客户群进行新的个性化营销，以减少成本，提高投资回报率(ROI)



■ 市场研究

- 进行客户、雇员，或招募满意度的调查
- 选择一个变量作为满意度的度量（例如，取值“1-5”等级变量）
- 根据响应者对其它问题的回答描述满意度水平
- 改善影响满意度的因素，例如工作环境或产品质量

■ 信用风险评分

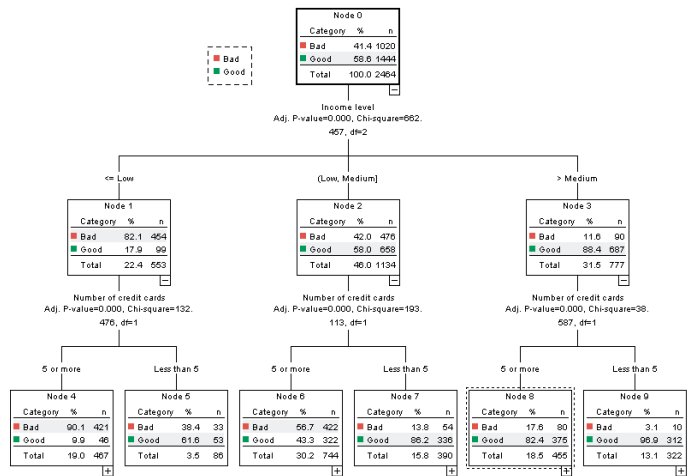
- 确定风险组（高、中或低）
- 基于客户信息描述风险群体特征，例如帐户活动
- 基于风险群为贷款申请人提供合理的贷款额度

■ 项目目标

- 选择变量用以表示是否发生预期结果(例如，一福利项目的成功实施)
- 根据申请人信息，揭示决定成功的因素
- 开展新的项目，以满足更多人的需要

■ 公共部门营销

- 选择一个响应变量划分您的客户群（例如，高校潜在申请人中实际申请者与未申请者）
- 基于其它属性概括客户群，如人口统计学或客户活动
- 针对某具体客户群进行新的个性化营销，以减少成本，提高投资回报率(ROI)



SPSS Data Editor window showing a table of model results for 13 nodes.

NodeID	PredictedValue	PredictedProbability 1	PredictedProbability 2
1	14	1.00	.43
2	12	.00	.81
3	10	.00	.94
4	10	.00	.94
5	14	1.00	.43
6	14	1.00	.43
7	13	.00	.56
8	10	.00	.94
9	10	.00	.94
10	12	.00	.81
11	10	.00	.94
12	11	.00	.78
13	13	.00	.56

模型结果在SPSS中直接对新的案例评分



可选择四种决策树算法

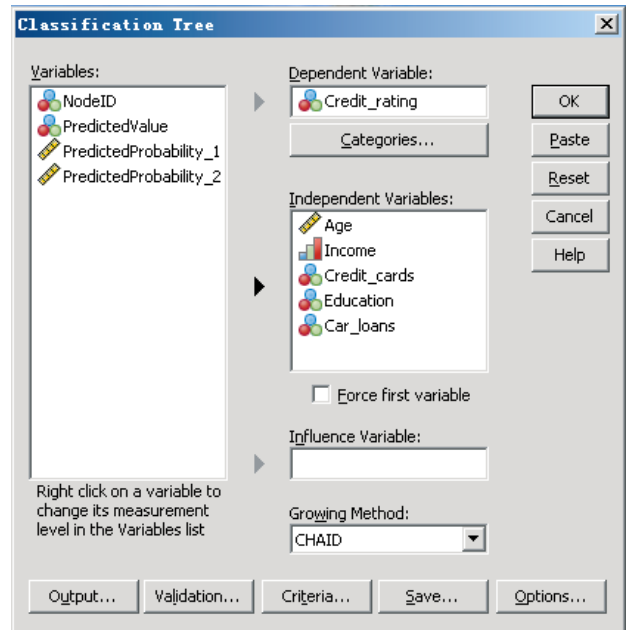
SPSS Decision Trees包括四种成熟的树生成算法：

- CHAID – 快速、多分枝的统计树算法，使您能够迅速有效地探索数据，可根据所希望的分类结果建立分段及资料概括说明
- Exhaustive CHAID – 改进的CHAID算法，会检查预测因子的每种可能分割
- Classification & regression trees (C&RT) – 一个完全的二叉树算法，能将分割数据为精确、类似同质的子集合
- QUEST—可以无偏差地选择变量，迅速有效地建立二叉树的算法

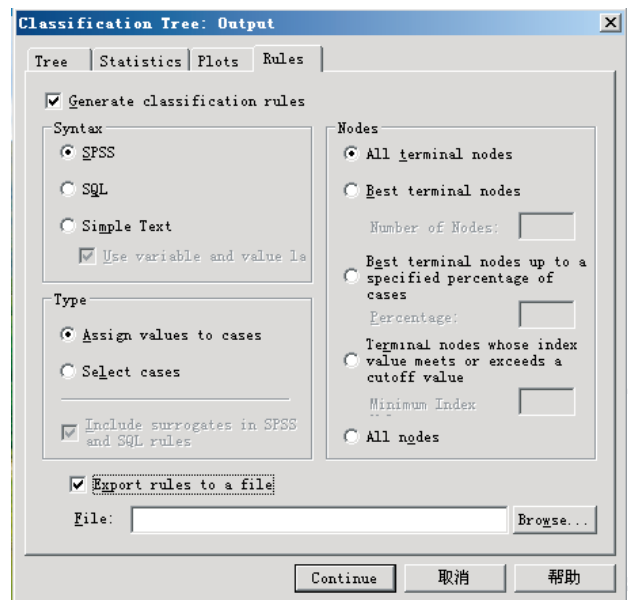
四种算法，使您能够尝试不同类型的树生成方法，并找到最佳拟合数据的模型

在SPSS中扩展您的结果，并应用于将来的分析之中

由于SPSS Decision Trees是SPSS的附加模块，所以您在SPSS环境下，就可以直接创建分类树，并可方便地将结果应用于数据集。您并不需要在SPSS和其它软件之间反复地操作。而且，您可以生成SPSS Syntax、SQL语句、以及文本格式的选择或分类/预测规则。您可以在输出窗口显示结果，并把结果保存为外部文件以备后用，如对新的个体或观测进行预测。如果您希望利用结果对其它数据文件评分，则可以把相关的模型信息直接添加到数据集，或将模型以XML格式导出以供SPSS Server调用，实现数据评分。



利用SPSS中的CHAID， Exhaustive CHAID， C&RT， 或者QUEST创建树模型



根据模型结果直接在SPSS中生成用于选择个案/预测的规则，或者将规则导出到外部文件，供后用

功能

树

- 显示树状图解、树映射图、节点内显示条形图、数据表格
- 功能全面易于理解的交互界面，使您可以很轻松地建立分类决策树模型，您可以设定：
 - 测度水平（名义、序次、连续）
 - 子变量
 - 因变量
 - 影响变量
 - 生长方法
 - 输出设置，包括树、统计量、图和规则
 - 分割样本校验或交叉检验
 - 终止条件
 - 保存变量，包括预测值、概率、和XML模型
- 从四种树生成方法中选择
- 通过几种方法查看树节点：节点中显示目标变量的条形图/表，或同时显示二者
- 合并或扩展分枝，并更改其它修饰树的属性，如字体、颜色
- 审视和打印树
- 设置在输出视窗显示的树模型大小的百分比
- 利用production mode自动化树模型的建立
 - 在交互界面下自动生成语法
- 强迫某个预测因子进入模型
- 指定先验概率、误分类成本、收入、费用、利润

树生成方法

- 利用四种功能强大的树生成算法实现
 - Kass (1980) 的CHAID方法
 - Biggs, de Ville, 和Suen (1991)的Exhaustive CHAID方法
 - Breiman, Friedman, Olshen, 和Stone (1984)的Classification & regression trees (C&RT) 方法
 - Loh和Shih (1997)的QUEST方法
- 两种方法处理缺失值：归为新类别或利用替代值填充
- 根据指定的类别数目，离散化连续型预测变量
- C&RT 和QUEST具有修剪能力
- 从原数据中随机抽取样本用于分割样本校验或利用某一变量分割样本

模型评估

- 生成风险和分类表
- 根据评估图表选择最佳的分割：
 - 增益
 - 索引（提升）
 - 响应
 - 均值
 - 平均利润
 - 投资收益（ROI）
- 数据分为训练和测试两部分，检验模型精度
- 节点摘要窗口显示选定节点的摘要图或分类规则

部署

导出

- 树形图和表：导出格式包括：
 - WMF、BMP、JPEG、TIFF、PNG、Macintosh[®] PICT
 - HTML、text、Word/RTF、及Excel格式的摘要表格
- 通过创建新变量来保存模型信息于工作数据集中
- 以SQL格式导出选定分割的决策规则，用以评分数据库、以SPSS Syantx格式导出以评分SPSS文件，或导出简单的文本（通过语法）
- 将树模型以XML格式导出，以供在SPSS Server及SmartScore[®] 下用于对新的个案或数据文件评分
- 将树模型以图像和静态/交互表的形式发布到SmartViewer[®] Web Server[™]
- 要进行深入的分析，可以通过树节点直接在工作数据集中选择感兴趣的分割，然后进行更多的分析

系统要求

- 软件：SPSS Base 17.0
- 其它的系统需求根据平台的不同而

