

IBM SPSS Modeler 18.4 用户指南



注

在使用本资料及其支持的产品之前，请阅读第 217 页的『[注意事项](#)』中的信息。

产品信息

本版本适用于的版本 18、发行版 4、IBM® SPSS Modeler 的修订 0 以及所有后续版本和修改，除非在新版本中另有说明

© Copyright International Business Machines Corporation .

内容

第 1 章 关于 IBM SPSS Modeler.....	1
IBM SPSS Modeler 产品.....	1
IBM SPSS Modeler.....	1
IBM SPSS Modeler Server.....	1
IBM SPSS Modeler Administration Console.....	1
IBM SPSS Modeler Batch.....	2
IBM SPSS Modeler Solution Publisher.....	2
IBM SPSS Collaboration and Deployment Services 的 IBM SPSS Modeler Server 适配器.....	2
IBM SPSS Modeler 版本.....	2
文档.....	2
SPSS Modeler Professional 文档.....	3
SPSS Modeler Premium 文档.....	3
应用程序示例.....	3
Demos 文件夹.....	3
许可证跟踪.....	4
第 2 章 V18.4.0 中的新增功能.....	5
第 3 章 产品概述.....	7
入门.....	7
启动 IBM SPSS Modeler.....	7
从命令行启动.....	7
连接到 IBM SPSS Modeler Server.....	7
连接至 Analytic Server.....	9
更改 temp 目录.....	10
启动多个 IBM SPSS Modeler 会话.....	10
IBM SPSS Modeler 界面概览.....	10
IBM SPSS Modeler 流工作区.....	11
节点选用板.....	11
IBM SPSS Modeler 管理器.....	12
IBM SPSS Modeler 项目.....	13
IBM SPSS Modeler 工具栏.....	14
自定义工具栏.....	15
定制 IBM SPSS Modeler 窗口.....	15
更改流的图标尺寸.....	16
在 IBM SPSS Modeler 中使用鼠标.....	17
使用快捷键.....	17
打印.....	18
实现 IBM SPSS Modeler 的自动化.....	18
第 4 章 了解数据挖掘.....	19
数据挖掘概述.....	19
评估数据.....	20
数据挖掘策略.....	20
CRISP-DM 过程模型.....	21
模型类型.....	22
数据挖掘示例.....	25
第 5 章 构建流.....	27
流构建概述.....	27

构建数据流.....	27
使用节点.....	27
使用流.....	31
流描述.....	40
运行流.....	42
使用模型.....	42
添加注释和注解到节点和流.....	42
保存数据流.....	46
加载文件.....	48
映射数据流.....	48
提示和快捷键.....	50
第 6 章 处理数据.....	51
构建图表.....	51
图表目录.....	51
布局和术语.....	51
从图表类型库构建图表.....	52
图表类型.....	52
仪表盘.....	83
全局可视化首选项.....	83
第 7 章 处理输出.....	87
查看器.....	87
显示和隐藏结果.....	87
移动、删除和复制输出.....	88
更改初始对齐方式.....	88
更改各输出项的对齐方式.....	88
查看器大纲.....	88
编辑项并将其添加到查看器中.....	89
在查看器中查找并替换信息.....	90
复制输出到其他应用程序中.....	91
交互输出.....	92
导出输出.....	92
HTML 选项.....	93
Web 报告选项.....	93
Word 选项.....	94
Excel 选项.....	94
PowerPoint 选项.....	95
PDF 选项.....	95
文本选项.....	96
仅图像选项.....	96
图形格式选项.....	97
查看器打印.....	98
打印输出和图表.....	98
打印预览.....	98
页面属性：页眉和页脚.....	98
页面属性：选项.....	99
保存输出.....	99
保存查看器文档.....	99
透视表.....	100
透视表.....	100
操作透视表.....	100
使用层.....	103
显示和隐藏项目.....	103
表格外观.....	104
表格属性.....	105
单元格属性.....	107

脚注和文字说明.....	107
数据单元格宽度.....	109
更改列宽.....	109
显示透视表中的隐藏边框.....	109
在透视表中选择行、列和单元格.....	109
打印透视表.....	109
从透视表创建图表.....	111
遗存表.....	111
选项.....	111
选项.....	111
一般选项.....	111
查看器选项.....	111
透视表选项.....	112
输出选项.....	112
第 8 章 处理缺失值.....	113
缺失值概述.....	113
处理缺失值.....	113
处理带有缺失值的记录.....	114
处理带有缺失值的字段.....	114
处理具有缺失值的记录.....	114
归因或填充缺失值.....	116
用于缺失值的 CLEM 函数.....	116
第 9 章 构建 CLEM 表达式.....	119
关于 CLEM.....	119
CLEM 示例.....	119
值和数据类型.....	120
表达式和条件.....	121
流、会话和超节点参数.....	122
字符串的处理.....	122
处理空白值和缺失值.....	123
数字的处理.....	123
时间和日期的处理.....	123
多个字段的汇总.....	124
多个响应数据的处理.....	125
表达式构建器.....	125
访问表达式构建器.....	126
创建表达式.....	126
选择函数.....	126
选择字段、参数和全局变量.....	129
查看或选择值.....	129
检查 CLEM 表达式.....	129
查找和替换.....	130
第 10 章 CLEM 语言参考.....	133
CLEM 参考概述.....	133
CLEM 数据类型.....	133
整数.....	133
实数.....	133
字符.....	134
字符串.....	134
列表.....	134
字段.....	134
日期.....	134
时间.....	135
CLEM 运算符.....	136

函数参考.....	138
函数说明中的约定.....	138
信息函数.....	139
转换函数.....	140
比较函数.....	141
逻辑函数.....	142
数字函数.....	142
三角函数.....	144
概率函数.....	144
空间函数.....	144
位元整数运算.....	145
随机函数.....	146
字符串函数.....	147
SoundEx 函数.....	151
日期和时间函数.....	151
序列函数.....	155
全局函数.....	159
处理空白值和空值的函数.....	159
特殊字段.....	160
第 11 章 将 IBM SPSS Modeler 与存储库配合使用.....	163
关于 IBM SPSS Collaboration and Deployment Services Repository.....	163
存储和部署存储库对象.....	163
连接至存储库.....	164
输入存储库的凭证.....	164
浏览存储库凭证.....	164
浏览存储库内容.....	164
在存储库中存储对象.....	165
设置对象属性.....	165
存储流.....	166
存储工程.....	167
存储节点.....	167
存储输出对象.....	167
存储模型和模型选项板.....	167
从存储库中检索对象.....	168
选择要检索的对象.....	168
选择对象版本.....	169
在存储库中搜索对象.....	169
修改存储库对象.....	170
编辑、重命名和删除文件夹.....	170
锁定和解锁存储库对象.....	170
删除存储库对象.....	170
管理存储库对象的属性.....	171
查看文件夹属性.....	171
查看和编辑对象属性.....	171
管理对象版本标签.....	172
部署流.....	172
流部署选项.....	173
评分分支.....	174
第 12 章 将流保存到 IBM Cloud Pak for Data.....	177
第 13 章 导出到外部应用程序.....	179
关于导出到外部应用程序.....	179
打开 IBM SPSS Modeler Advantage 中的流.....	179
导入和导出 PMML 模型.....	179
支持 PMML 的模型类型.....	180

第 14 章 工程和报告	183
工程简介.....	183
CRISP-DM 视图.....	183
“类”视图.....	183
构建工程.....	184
创建新工程.....	184
添加到工程.....	184
将工程转移至 IBM SPSS Collaboration and Deployment Services Repository.....	185
设置工程属性.....	185
注解工程.....	185
对象属性.....	186
关闭工程.....	186
生成报告.....	186
保存和导出生成的报告.....	187
第 15 章 自定义 IBM SPSS Modeler	189
定制 IBM SPSS Modeler 选项.....	189
设置 IBM SPSS Modeler 选项.....	189
系统选项.....	189
设置缺省目录.....	190
设置用户选项.....	190
自定义节点选项板.....	196
自定义选项板管理器.....	197
更改选项板选项卡视图.....	199
第 16 章 流和节点的性能注意事项	201
节点的顺序.....	201
节点缓存.....	202
性能：过程节点.....	202
性能：建模节点.....	203
性能：CLEM 表达式.....	203
第 17 章 IBM SPSS Modeler 中的辅助功能	205
IBM SPSS Modeler 辅助功能概述.....	205
辅助功能支持的类型.....	205
针对视力受损用户的辅助功能.....	205
盲人用户的辅助功能选项.....	206
键盘辅助功能选项.....	206
使用屏幕阅读器.....	212
使用提示.....	213
与其他软件的冲突.....	214
JAWS 和 Java.....	214
在 IBM SPSS Modeler 中使用图形.....	214
第 18 章 Unicode 支持	215
IBM SPSS Modeler 中的 Unicode 支持.....	215
注意事项	217
商标.....	218
产品文档的条款和条件.....	218
索引	219

第 1 章 关于 IBM SPSS Modeler

IBM SPSS Modeler 是一组数据挖掘工具，通过这些工具可以采用商业技术快速建立预测性模型，并将其应用于商业活动，从而改进决策过程。IBM SPSS Modeler 参照行业标准 CRISP-DM 模型设计而成，可支持从数据到更优商业成果整个数据挖掘过程。

IBM SPSS Modeler 提供了各种借助机器学习、人工智能和统计学的建模方法。通过建模选项板中的方法，您可以根据数据生成新的信息以及开发预测模型。每种方法各有所长，同时适用于解决特定类型的问题。

SPSS Modeler 可以作为独立产品购买，也可以作为客户端与 SPSS Modeler Server 一起使用。同时提供了大量其他选项，以下各节将对这些选项进行概述。有关更多信息，请参阅 <https://www.ibm.com/analytics/us/en/technology/spss/>。

IBM SPSS Modeler 产品

IBM SPSS Modeler 系列产品及关联的软件包括以下各项。

- IBM SPSS Modeler
- IBM SPSS Modeler Server
- IBM SPSS Modeler Administration Console（包含在 IBM SPSS Deployment Manager 中）
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- IBM SPSS Collaboration and Deployment Services 的 IBM SPSS Modeler Server 适配器

IBM SPSS Modeler

SPSS Modeler 是具有完整功能的产品，它安装并运行于个人计算机上。您可以在本地方式作为独立产品运行 SPSS Modeler，也可以在分布方式下将其与 IBM SPSS Modeler Server 一起使用来提高大型数据集的性能。

借助 SPSS Modeler，您可以快速直接地构建准确的预测模型，而不进行编程。通过使用唯一可视界面，您可以轻松地查看数据挖掘过程。借助该产品随附的高级分析支持，您可以发现数据中先前隐藏的模式和趋势。您可以构建结果模型并了解影响结果的因素，从而利用业务机会并降低风险。

SPSS Modeler 推出了两个版本：SPSS Modeler Professional 和 SPSS Modeler Premium。有关更多信息，请参阅主题 [第 2 页的『IBM SPSS Modeler 版本』](#)。

IBM SPSS Modeler Server

SPSS Modeler 使用客户端/服务器体系结构将资源集约型操作的请求分发给功能强大的服务器软件，因而使大数据集的传输速度大大加快。

SPSS Modeler Server 是一个单独授权的产品，在分布分析方式下，该产品在安装了一个或多个 IBM SPSS Modeler 的服务器主机上持续运行。这种运行方式大大提高了 SPSS Modeler Server 对大型数据集的处理速度，因为在服务器上可以运行耗用内存的操作，并且无需将数据下载到客户端计算机上。IBM SPSS Modeler Server 还提供对 SQL 优化和数据库内建模功能的支持，从而在性能和自动化方面带来更多优势。

IBM SPSS Modeler Administration Console

Modeler Administration Console 是一个图形用户界面，用于管理多个 SPSS Modeler Server 配置选项，这些选项还可以通过选项文件进行配置。控制台包含在 IBM SPSS Deployment Manager，可以用于监视和配置 SPSS Modeler Server 安装，并且可供当前 SPSS Modeler Server 客户免费使用。应用程序只能安装在 Windows 计算机上；但是它可以管理安装在任何受支持平台上的服务器。

IBM SPSS Modeler Batch

数据挖掘通常是交互过程，因此，还可以从命令行运行 SPSS Modeler 而不需要图形用户界面。例如，您可能具有长时间运行或重复任务，并且希望在用户不进行干预的情况下执行这些任务。SPSS Modeler Batch 是该产品的一个特殊版本，可提供对 SPSS Modeler 完整分析性能的支持，而无需访问常规的用户界面。要使用 SPSS Modeler Batch，需要 SPSS Modeler Server。

IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher 是一个工具，它使您能够创建 SPSS Modeler 流的打包版本，该版本的流可以由外部运行时引擎运行或者可以嵌入在外部应用程序中。通过这种方式，您可以发布和部署完整的 SPSS Modeler 流以用于未安装 SPSS Modeler 的环境。SPSS Modeler Solution Publisher 作为 IBM SPSS Collaboration and Deployment Services - Scoring 服务的组成部分分发，需要单独的许可证。通过此许可证，您可以接收 SPSS Modeler Solution Publisher Runtime，它使您能够执行已发布的流。

有关 SPSS Modeler Solution Publisher 的更多信息，请参阅 IBM SPSS Collaboration and Deployment Services 文档。IBM SPSS Collaboration and Deployment Services IBM 文档包含名为“IBM SPSS Modeler Solution Publisher”和“IBM SPSS Analytics Toolkit”的部分。

IBM SPSS Collaboration and Deployment Services 的 IBM SPSS Modeler Server 适配器

IBM SPSS Collaboration and Deployment Services 的一些适配器使 SPSS Modeler 和 SPSS Modeler Server 能够与 IBM SPSS Collaboration and Deployment Services 存储库进行交互。通过这种方式，部署到存储库的 SPSS Modeler 流可以由多个用户共享，或者从瘦客户端应用程序 IBM SPSS Modeler Advantage 进行访问。请将适配器安装在托管存储库的系统上。

IBM SPSS Modeler 版本

SPSS Modeler 推出了下列版本。

SPSS Modeler Professional

SPSS Modeler Professional 提供处理大多数类型的结构化数据所需要的所有工具，例如 CRM 系统中跟踪的行为和交互、人口统计信息、采购行为和销售数据。

SPSS Modeler Premium

SPSS Modeler Premium 是一个单独授权的产品，它对 SPSS Modeler Professional 进行了扩展，以便后者能够处理专门的数据和非结构化文本数据。SPSS Modeler Premium 包含 IBM SPSS Modeler Text Analytics：

IBM SPSS Modeler Text Analytics 采用先进语言技术和自然语言处理 (NLP)，以快速处理大量非结构化文本数据，提取和组织关键概念，以及将这些概念分为各种类别。提取的概念和类别可以与现有的结构化数据（例如人口统计信息）相结合，并且可借助 IBM SPSS Modeler 的全套数据挖掘工具进行建模，以此实现更好更集中的决策。

IBM SPSS Modeler Subscription

IBM SPSS Modeler Subscription 提供与传统 IBM SPSS Modeler 客户端完全相同的预测性分析功能。通过 Subscription 版本，您可以定期下载产品更新。

文档

可从 SPSS Modeler 中的**帮助**菜单获取文档。这样会打开始始终可在产品外部访问的在线 IBM 文档。

每个产品的完整文档（包括安装指示信息）也在以下位置以 PDF 格式提供：<https://www.ibm.com/support/pages/spss-modeler-184-documentation>。

SPSS Modeler Professional 文档

SPSS Modeler Professional 文档套件（安装指示信息除外）如下。

- **IBM SPSS Modeler 用户指南。** 对于使用 SPSS Modeler 的一般简介，包括如何构建数据流、处理缺失值、构建 CLEM 表达式处理项目和报告，以及将用于部署的流打包到 IBM SPSS Collaboration and Deployment Services 或 IBM SPSS Modeler Advantage。
- **IBM SPSS Modeler Source、Process 和 Output 节点。** 描述用于以不同格式读取、处理和输出数据的所有节点。实际上这表示所有节点而非建模节点。
- **IBM SPSS Modeler Modeling 节点。** 描述所有用于创建数据挖掘模型的节点。IBM SPSS Modeler 提供了各种借助机器学习、人工智能和统计学的建模方法。
- **IBM SPSS Modeler 应用程序指南。** 本指南中的示例旨在为具体的建模方法和技术提供具有针对性的简介。还可以从“帮助”菜单获取本指南的联机版本。有关更多信息，请参阅主题 [第 3 页的『应用程序示例』](#)。
- **IBM SPSS Modeler Python 脚本编制和自动化。** 通过编写 Python 脚本实现系统自动化的相关信息，其中包括可以用于处理节点和流的属性的信息。
- **IBM SPSS Modeler 部署指南。** 有关在 IBM SPSS Deployment Manager 下以处理作业的步骤形式运行 IBM SPSS Modeler 流的信息。
- **IBM SPSS Modeler 数据库内挖掘指南。** 有关如何利用数据库的功能通过第三方算法来改进性能并增强分析功能的信息。
- **IBM SPSS Modeler Server 管理和性能指南。** 提供有关如何配置和管理 IBM SPSS Modeler Server 的信息。
- **IBM SPSS Deployment Manager 用户指南。** 有关使用 Deployment Manager 应用程序中包含的管理控制台用户界面来监视和配置 IBM SPSS Modeler Server 的信息。
- **IBM SPSS Modeler CRISP-DM 指南。** 借助 CRISP-DM 方法进行 SPSS Modeler 数据挖掘的分步指南。
- **IBM SPSS Modeler Batch 用户指南。** 提供在批处理方式下使用 IBM SPSS Modeler 的完整指导，包括批处理方式执行和命令行自变量的详细信息。本指南仅以 PDF 格式提供。

SPSS Modeler Premium 文档

SPSS Modeler Premium 文档套件（安装指示信息除外）如下。

- **SPSS Modeler Text Analytics 用户指南。** 提供有关将文本分析与 SPSS Modeler 配合使用的信息，包括文本挖掘节点、交互式工作台、模板和其他资源。

应用程序示例

SPSS Modeler 中的数据挖掘工具可以帮助解决很多业务和组织问题，应用程序示例将提供有关特定建模方法和技术的简要的针对性说明。此处使用的数据集比某些数据挖掘器管理的大量数据存储小得多，但涉及的概念和方法可扩展到实际应用程序。

要访问示例，请在 SPSS Modeler 中单击“帮助”菜单中的**应用程序示例**。

数据文件和样本流安装在产品安装目录下的 Demos 文件夹中。有关更多信息，请参阅 [第 3 页的『Demos 文件夹』](#)。

数据库建模示例。 请参阅 *IBM SPSS Modeler 数据库内挖掘指南* 中的示例。

脚本编制示例。 请参阅 *IBM SPSS Modeler 脚本编写与自动化指南* 中的示例。

Demos 文件夹

与应用程序示例配合使用的数据文件和样本流安装在产品安装目录下的 Demos 文件夹中（例如：`C:\Program Files\IBM\SPSS\Modeler\<version>\Demos`）。也可以从 Windows“开始”菜单上的 IBM SPSS Modeler 程序组访问此文件夹，或者通过单击**文件 > 打开流**对话框中最近的目录列表上的 Demos 来进行访问。

许可证跟踪

当您使用 SPSS Modeler 时，系统会定期跟踪并记录许可证使用情况。所记录的许可证度量为 *AUTHORIZED_USER* 和 *CONCURRENT_USER*，并且记录的度量类型取决于您针对 SPSS Modeler 具有的许可证类型。

产生的日志文件可由 IBM License Metric Tool 处理，通过该工具可生成许可证使用情况报告。

许可证日志文件是在记录 SPSS Modeler Client 日志文件的同一目录中创建的（缺省情况下，为 `%ALLUSERSPROFILE%/IBM/SPSS/Modeler/<version>/log`）。

第 2 章 V18.4.0 中的新增功能

IBM SPSS Modeler 在本发行版中添加了以下功能和增强功能。

- 现在，文本分析中支持 Cloud Pak for Data 模板格式 (json)。
- 现在，您可以通过 SPSS Modeler 用户界面来切换 Python 环境。
- SPSS Modeler 现在提供了增强的密码加密功能。由于向后兼容性问题，请注意以下事项：
 - 在 V18.4 中，如果使用专用密码文件运行 SPSS Modeler Server 以向专用密码数据库认证用户，那么必须使用 **pwutil** 可执行文件来重新创建该数据库。有关指示信息，请参阅配置专用密码数据库。
 - 在 V18.4 中，如果使用加密密码（通过工具菜单上的**对密码编码**选项），那么必须重新创建该密码。有关更多信息，请参阅[生成经过编码的密码](#)。
- 现在提供了一个用于管理 SPSS Modeler Server 连接的新菜单项：**工具 > 管理 Modeler Server 配置**。可用的设置与 SPSS Collaboration and Deployment Services Deployment Manager 随附的 Modeler Administration Console 的 **SPSS Modeler Server 配置** 部分中的设置类似。有关这些设置的详细信息，请参阅 [server/adminguidesource/admin_pem_config.dita](#)。
- Cognos TM1 Server 连接不再支持空密码。
- 现在支持 Windows 11。
- 现在支持 MacOS Monterey（也称为 MacOS 12）。
- 现在支持 Java 11。
- 现在支持 CPLEX 22.1。
- 现在支持 InstallAnywhere 2021 sp2。
- 现在支持 SPSS Data Access Pack 8.1.1。
- 现在支持 Netezza Performance Server 11.x。
- 现在支持 Cognos Analytics Connector 11.1.7。
- 现在支持 Amazon S3（仅用于读取数据）。
- 现在支持 ClickHouse 数据库 V22.3，但存在以下限制：
 - 数据库导出节点中的“合并”功能不支持 SQL 回送（因为 ClickHouse 对更新和删除等操作的支持较弱）
 - 时间类型不支持 SQL 回送（因为 ClickHouse 中没有时间类型）
 - 日期基线为 1970-01-01

第 3 章 产品概述

入门

作为一种数据挖掘应用程序，IBM SPSS Modeler 提供了用以寻找大数据集中有用关系的策略性方法。与更传统的统计方法相比，您在开始时不必知道您要寻找什么。您可以通过拟合不同的模型和研究不同的关系来探索您的数据，直到发现有用的信息。

启动 IBM SPSS Modeler

要启动此应用程序，请单击：

开始 > [所有] 程序 > **IBM SPSS Modeler <version>** > **IBM SPSS Modeler <version>**

主窗口将在几秒钟后显示。

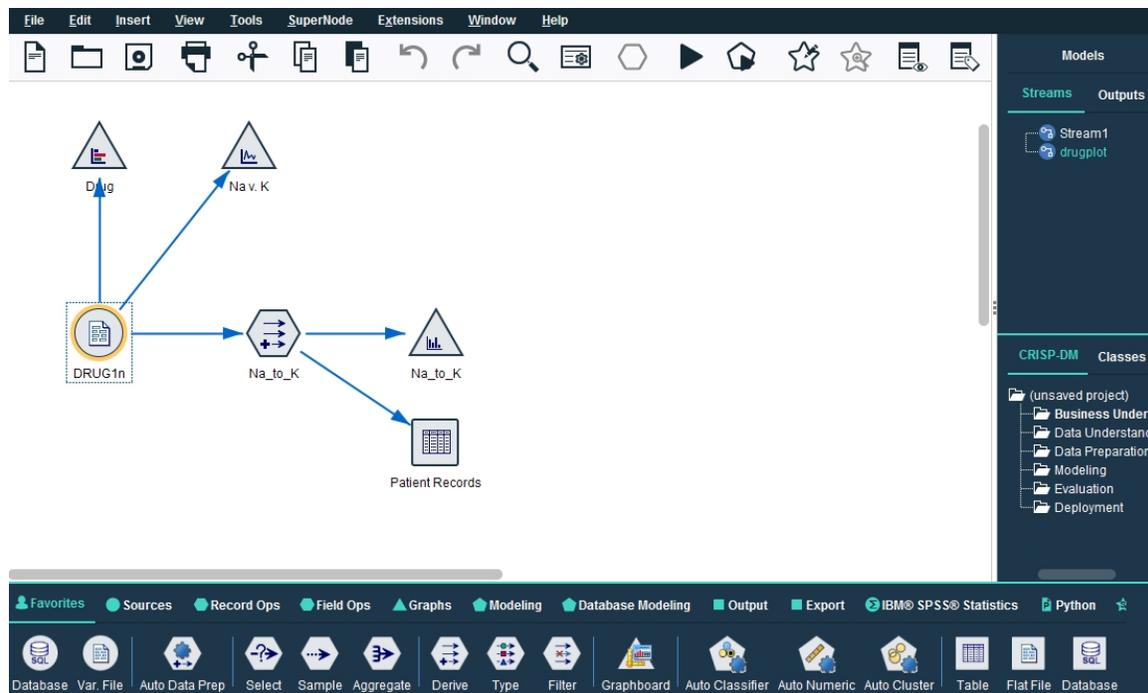


图 1: IBM SPSS Modeler 主应用程序窗口

从命令行启动

您可以使用操作系统的命令行来如下启动 IBM SPSS Modeler：

1. 在安装了 IBM SPSS Modeler 的计算机上，打开 DOS 或命令提示符窗口。
2. 要以交互方式启动 IBM SPSS Modeler 界面，请输入 modelerclient 命令，后跟必需的参数；例如：

```
modelerclient -stream report.str -execute
```

可用参数（标记）允许您连接到服务器、加载流、运行脚本或根据需要指定其他参数。

连接到 IBM SPSS Modeler Server

IBM SPSS Modeler 可作为独立的应用程序运行，或作为直接连接到 IBM SPSS Modeler Server 的客户端运行，或者作为通过进程协调器 (COP) 插件从 IBM SPSS Collaboration and Deployment Services 连接到 IBM

SPSS Modeler Server 或服务器集群的客户端运行。当前连接状态显示在 IBM SPSS Modeler 窗口的左下角。

无论何时想连接到服务器，都请手动输入想要连接的服务器的名称或选择之前已定义的名称。但是，如果您拥有 IBM SPSS Collaboration and Deployment Services，则可以从“服务器登录”对话框搜索服务器列表或服务器集群列表。可以通过进程协调器执行浏览网络上运行的 Statistics 服务的功能。

连接到服务器

1. 在“工具”菜单上，单击**服务器登录**。这将打开“服务器登录”对话框。或者，双击 IBM SPSS Modeler 窗口的连接状态区域。
2. 使用该对话框指定要连接到本地服务器计算机的选项或从表中选择连接。
 - 单击**添加**或**编辑**以添加或编辑连接。有关更多信息，请参阅主题第 8 页的『添加并编辑 IBM SPSS Modeler Server 连接』。
 - 单击**搜索**以访问进程协调器中的服务器或服务器集群。有关更多信息，请参阅主题第 9 页的『搜索 IBM SPSS Collaboration and Deployment Services 中的服务器』。

服务器表。此表包含已定义的服务器连接集。该表显示缺省连接、服务器名称、说明和端口号。您可以手动添加新的连接，以及选择或搜索现有连接。要将特定的服务器设置为缺省连接，请在表中“缺省”列中为此连接选择复选框。

缺省数据路径。指定用于服务器计算机上的数据的路径。单击省略号按钮 (...)，以浏览至所需要的位置。

设置凭证。不选中此复选框可启用**单点登录**功能，该功能尝试使您使用本地计算机用户名和密码详细信息登录服务器。如果无法使用单点登录，或您选中此复选框以禁用单点登录（例如，登录管理员帐户），则启用以下字段让您输入您的凭证。

用户标识。输入用于登录服务器的用户名。

密码。输入与指定用户名相关联的密码。

域。指定用于登录服务器的域。只有服务器计算机与客户计算机处于不同的 Windows 域时，才需要域名。

3. 单击**确定**以完成此连接。

断开与服务器的连接

1. 在“工具”菜单上，单击**服务器登录**。这将打开“服务器登录”对话框。或者，双击 IBM SPSS Modeler 窗口的连接状态区域。
2. 在此对话框中，选择“本地服务器”，然后单击**确定**。

添加并编辑 IBM SPSS Modeler Server 连接

您可以在“服务器登录”对话框中手动编辑或添加服务器连接。单击“添加”可以访问空的“添加/编辑服务器”对话框，在此对话框中可以输入服务器连接的详细信息。在“服务器登录”对话框中选择现有连接并单击“编辑”，将打开“添加/编辑服务器”对话框，其中包含所选连接的详细信息，以便可以进行任何更改。

注：不能编辑从 IBM SPSS Collaboration and Deployment Services 中添加的服务器连接，因为名称、端口及其他详细信息已在 IBM SPSS Collaboration and Deployment Services 中做过定义。最佳实践指出，应该使用相同的端口与 IBM SPSS Collaboration and Deployment Services 和 SPSS Modeler Client 进行通信。这些端口可以设置为 options.cfg 文件中的 max_server_port 和 min_server_port。

添加服务器连接

1. 在“工具”菜单上，单击**服务器登录**。这将打开“服务器登录”对话框。
2. 在此对话框中，单击**添加**。将打开“服务器登录：添加/编辑服务器”对话框。
3. 输入服务器连接的详细信息，然后单击**确定**保存此连接并返回“服务器登录”对话框。
 - **服务器。**指定可用服务器或从列表选择一个服务器。服务器计算机的名称可以使用字母数字（例如 myserver）或指派给服务器计算机的 IP 地址（例如，202.123.456.78）。
 - **端口。**指定服务器要侦听的端口号。如果缺省设置不可用，请向系统管理员索取正确的端口号。

- **描述。** 输入此服务器连接的可选描述。
- **确保安全连接（使用 SSL）。** 指定是否应该使用 SSL（安全套接字层）连接。SSL 是一个常用协议，用于确保通过网络发送的数据的安全。要使用此功能，必须在承载 IBM SPSS Modeler Server 的服务器中启用 SSL。必要时请与本地管理员联系，以了解详细信息。

编辑服务器连接

1. 在“工具”菜单上，单击**服务器登录**。这将打开“服务器登录”对话框。
2. 在此对话框中，选择希望编辑的连接，然后单击 **编辑**。将打开“服务器登录：添加/编辑服务器”对话框。
3. 更改服务器连接详细信息，然后单击**确认**保存更改内容并返回至“服务器登录”对话框。

搜索 IBM SPSS Collaboration and Deployment Services 中的服务器

在 IBM SPSS Collaboration and Deployment Services 中，可以使用进程协调器选择网络上可用的服务器或服务器集群，从而代替手动输入服务器连接。服务器集群是一组服务器，进程协调器从这组服务器中确定最适合对处理要求作出响应的服务器。

尽管可在“服务器登录”对话框中手动添加服务器，但通过搜索可用的服务器，可在无需知道正确服务器名称和端口号的情况下连接到服务器。此信息是自动提供的。但仍需输入正确的登录信息，如用户名、域和密码。

注：如果您无权访问进程协调器功能，那么仍然可以手动输入要连接的服务器名称或选择先前已定义的名称。有关更多信息，请参阅主题 [第 8 页的『添加并编辑 IBM SPSS Modeler Server 连接』](#)。

搜索服务器和服务器集群

1. 在“工具”菜单上，单击**服务器登录**。这将打开“服务器登录”对话框。
2. 在此对话框中，单击**搜索**打开“搜索服务器”对话框。如果在尝试浏览进程协调器时未登录到 IBM SPSS Collaboration and Deployment Services，则系统会提示您执行此项操作。
3. 从列表中选择服务器或服务器集群。
4. 单击**确定**以关闭对话框，然后将此连接添加到“服务器登录”对话框的表中。

连接至 Analytic Server

如果有多个 Analytic Server 可用，那么可以使用“Analytic Server 连接”对话框来定义多个要在 IBM SPSS Modeler 中使用的服务器。您的管理员可能已在 <Modeler_install_path>/config/options.cfg 文件中设置了缺省的 Analytic Server。但在定义之后，也可以使用其他可用服务器。例如，使用 Analytic Server 源节点和导出节点时，您可能希望在流的不同分支中使用不同的 Analytic Server 连接，以便在每个分支运行时，使用自己的 Analytic Server，并且不会将任何数据拉取到 IBM SPSS Modeler Server。请注意，如果某个分支包含多个 Analytic Server 连接，那么会将数据从 Analytic Server 拉取到 IBM SPSS Modeler Server。

要创建新的 Analytic Server 连接，请转至**工具 > Analytic Server 连接**，并在对话框的以下部分中提供必需的信息。

连接

URL。 以 `https://hostname:port/contextroot` 格式输入 Analytic Server 的 URL，其中 `hostname` 是 Analytic Server 的 IP 地址或主机名，`port` 是其端口号，而 `contextroot` 则是 Analytic Server 的上下文根。

租户。 输入 IBM SPSS Modeler Server 所属的租户的名称。如果不知道租户，请联系您的管理员。

认证

方式。 从以下认证方式中进行选择。

- **用户名和密码** 要求您输入用户名和密码。
- **存储的凭证** 需要您从 IBM SPSS Collaboration and Deployment Services Repository 中选择凭证。
- **Kerberos** 要求您输入服务主体名称和配置文件路径。如果不知道该信息，请联系您的管理员。

用户名。 输入 Analytic Server 用户名。

Reams。 选择要用于 Analytic Server 连接的域。

密码。 输入 Analytic Server 密码。

连接。 单击**连接**以测试新的连接。

连接

在指定上述信息并单击**连接**之后，连接将添加到该“连接”表。如果需要移除连接，请将其选择并单击**移除**。

如果管理员在 `options.cfg` 文件中定义了缺省 Analytic Server 连接，那么还可以单击**添加缺省连接**，将其添加到可用连接中。将提示您输入用户名和密码。

更改 temp 目录

IBM SPSS Modeler Server 执行的某些操作可能需要创建临时文件。缺省情况下，IBM SPSS Modeler 在系统临时目录下创建临时文件。可通过以下步骤更改临时目录的位置。

1. 创建新目录 `spss` 及其子目录 `servertemp`。
2. 编辑 `options.cfg`，该文件位于 IBM SPSS Modeler 安装目录的 `/config` 目录下。在此文件中编辑 `temp_directory` 参数，将其更改为：`temp_directory, "C:/spss/servertemp"`。
3. 完成此操作后，必须重新启动 IBM SPSS Modeler Server 服务。可通过单击 Windows 控制面板中的 **服务** 选项卡进行此服务重启操作。只需停止该服务然后将其重新启动即可激活所作的更改。重新启动机器也会重新启动该服务。

所有临时文件此时将写入该新目录。

注：必须使用正斜杠。

启动多个 IBM SPSS Modeler 会话

如果需要同时启动一个以上的 IBM SPSS Modeler 会话，则必须对 IBM SPSS Modeler 和 Windows 的设置做一些更改。例如，如果您有两个独立的服务器许可证，并且希望从同一台客户端针对两台不同的服务器运行两个流，则需要对上述设置做一些更改。

要启用多个 IBM SPSS Modeler 会话：

1. 单击：
开始 > [所有] 程序 > IBM SPSS Modeler
2. 在 IBM SPSS Modeler 快捷键（带箭头的图标）上右键单击并选择**属性**。
3. 在**目标**文本框中，将 `-noshare` 添加到该字符串的结尾。
4. 在 Windows 资源管理器中选择：
工具 > 文件夹选项...
5. 在“文件类型”选项卡上选择“IBM SPSS Modeler 流”选项，然后单击**高级**。
6. 在“编辑文件类型”对话框中，选择“使用 IBM SPSS Modeler 打开”，然后单击**编辑**。
7. 在**用于执行操作的应用程序**文本框中，在 `-stream` 参数前添加 `-noshare`。

IBM SPSS Modeler 界面概览

在数据挖掘过程中的每个时间点，您都可以通过容易使用的 IBM SPSS Modeler 界面邀请特定业务专家。建模算法（如预测、分类、细分和关联检测）可确保得到强大而准确的模型。模型结果可以轻松部署并读入到数据库、IBM SPSS Statistics 和各种其他应用程序中。

使用 IBM SPSS Modeler 即处理数据的过程，其中包括三个步骤。

- 首先，将数据读入 IBM SPSS Modeler。
- 接着，通过一系列处理来运行数据。

- 最后，将数据发送至目标。

这一操作序列称为 **数据流**，因为数据以一条条记录的形式，从数据源开始，依次经过各种操纵，最终到达目标（模型或某种数据输出）。

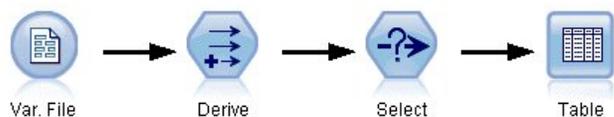


图 2: 简单流

IBM SPSS Modeler 流工作区

流工作区是 IBM SPSS Modeler 窗口中最大的区域，也是您构建和操作数据流的位置。

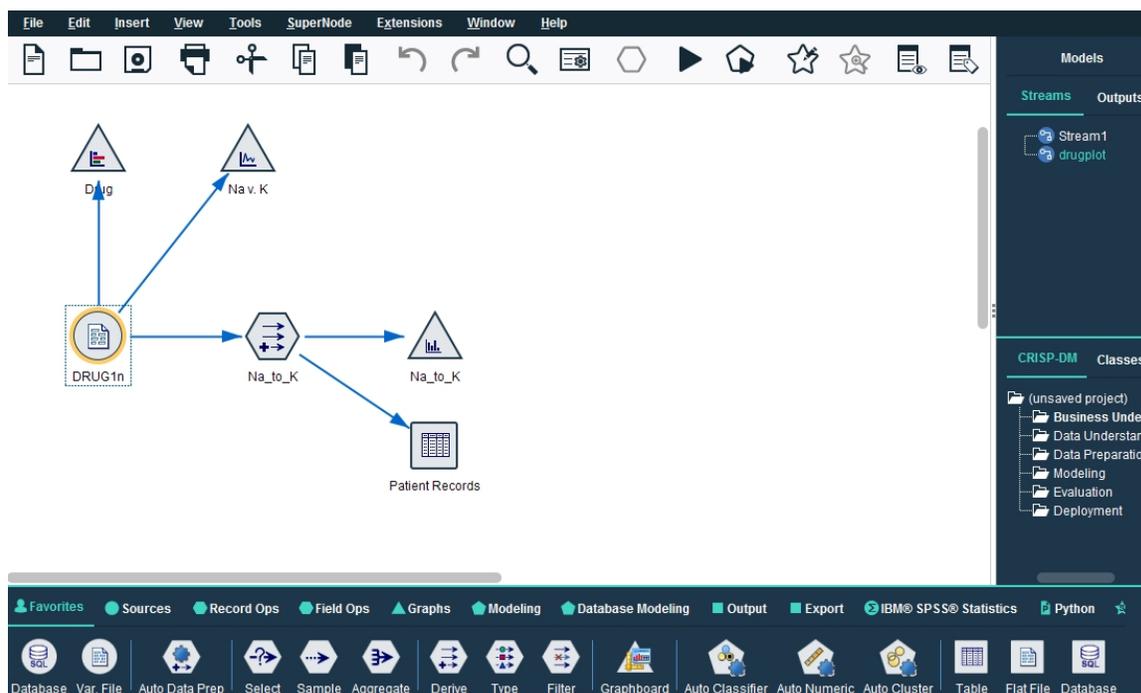


图 3: IBM SPSS Modeler 工作空间（缺省视图）

流是在界面的主画布中通过绘制与业务相关的数据操作图来创建的。每个操作都用一个图标或 **节点** 表示，这些节点通过 **流** 链接在一起，流表示数据在各个操作之间的流动。

在 IBM SPSS Modeler 中，可以在同一个流工作区或通过打开新的流工作区来同时处理多个流。会话期间，流存储在 IBM SPSS Modeler 窗口右上角的“流”管理器中。

注: 如果您使用的是 MacBook，而且已启用内置触控板的 **Force Click and haptic feedback** 设置，那么从节点选用板将节点拖放到流工作区可能会导致在工作区中添加重复的节点。为避免此问题，建议禁用 **Force Click and haptic feedback** 触控板系统首选项。

节点选用板

SPSS Modeler 中的大部分数据和建模工具都可从横跨流工作区下方窗口的底部的节点选用板中获取。

例如，**记录选项** 选用板选项卡包含可用于对数据记录执行操作（如选择、合并和附加）的节点。

要向画布添加节点，请双击节点选用板中的图标或将节点拖至画布上。随后可将各个图标连接以创建一个表示数据流动的流。



图 4: 节点选用板中的“记录选项”选项卡

每个选用板选项卡均包含一组不同的流操作阶段中使用的相关节点，如：

- **源**节点用于将数据引入到 SPSS Modeler 中。
- **记录选项**节点用于对数据记录执行操作，例如选择、合并和附加。
- **字段选项**节点用于对数据字段执行操作，例如过滤、派生新字段和确定给定字段的测量级别。
- **图形**节点用于以图形方式显示建模前后的数据。图形包括散点图、直方图、网络节点和评估图表。
- **建模**节点使用 SPSS Modeler 中提供的建模算法，例如神经网络、决策树、聚类算法和数据序列等。
- **数据建模**节点使用 Microsoft SQL Server、IBM Db2 以及 Oracle 和 Netezza 数据库中提供的建模算法。
- **输出**节点用于为数据、图表和模型结果生成可以在 SPSS Modeler 中查看的各种输出。
- **导出**节点用于生成可以在外部应用程序（例如，IBM SPSS Data Collection 或 Excel）中查看的各种输出。
- **IBM SPSS Statistics** 节点从 IBM SPSS Statistics 导入数据或向其导出数据，以及运行 IBM SPSS Statistics 过程。
- **Python** 节点可用于运行 Python 算法。
- **Spark** 节点可用于运行 Spark 算法。

随着对 SPSS Modeler 的熟悉，您也可以自定义供自己使用的选用板内容。

在“节点选用板”的左侧，您可以通过选择受监督、关联或细分对所显示的节点进行过滤。

“节点”选用板下方是一个报告窗格，此窗格提供各种操作的进度反馈，例如何时将数据读入数据流中。“节点”选用板下方还有一个状态窗格，此窗格提供有关应用程序当前正在执行的操作的信息以及何时需要用户反馈的指示信息。

注: 如果您使用的是 MacBook，而且已启用内置触控板的 **Force Click and haptic feedback** 设置，那么从节点选用板将节点拖放到流工作区可能会导致在工作区中添加重复的节点。为避免此问题，建议禁用 **Force Click and haptic feedback** 触控板系统首选项。

IBM SPSS Modeler 管理器

管理器窗格位于窗口右上角。此窗格包含用于管理流、输出和模型的三个选项卡。

可以使用“流”选项卡打开、重命名、保存和删除会话中创建的流。



图 5: “流”选项卡



图 6: “输出”选项卡

“输出”选项卡中包含由 IBM SPSS Modeler 中的流操作生成的各类文件，例如图形和表格。您可以显示、保存、重命名和关闭此选项上列出的表格、图形和报告。



图 7: 包含模型块的“模型”选项卡

在管理器选项卡中，“模型”选项卡具有最强大的功能。该选项卡中包含所有的模型块，这些模型块包含 IBM SPSS Modeler 当前会话中生成的模型。可以直接从“模型”选项卡浏览这些模型或者将它们添加到画布内的流中。

IBM SPSS Modeler 项目

窗口右侧底部是工程窗格，用于创建和管理数据挖掘工程（与数据挖掘任务相关的文件组）。您可通过两种方法来查看您在 IBM SPSS Modeler 中创建的工程：在“类”视图和 CRISP-DM 视图中查看。

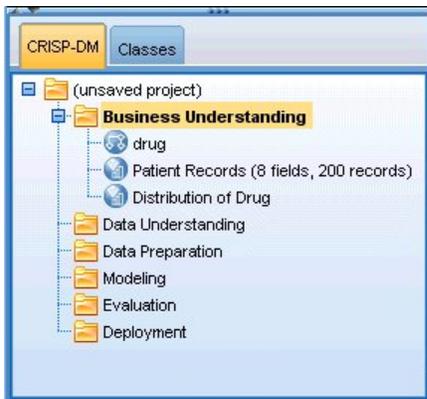


图 8: CRISP-DM 视图

依据业内认可的非专利方法“跨行业数据挖掘过程标准”，CRISP-DM 选项卡提供了一种项目组织方法。无论是有经验的数据挖掘人员还是新手，使用 CRISP-DM 工具都会使您事半功倍。

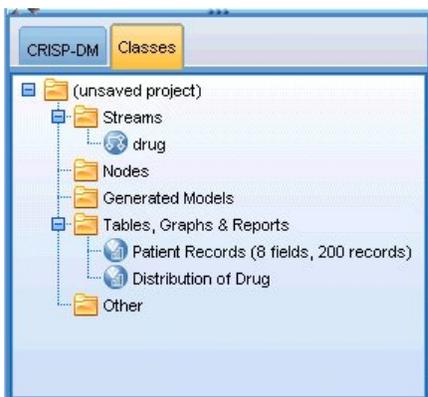


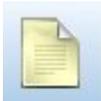
图 9: “类”视图

“类”选项卡提供一种在 IBM SPSS Modeler 中按类别（即，按您所创建的对象类型）组织工作的方式。此视图在获取数据、流、模型的详尽目录时十分有用。

IBM SPSS Modeler 工具栏

IBM SPSS Modeler 窗口顶部有一个图标工具栏，其中包含许多有用功能。下面是一些工具栏按钮及其功能。

	创建新流		打开流
	保存流		打印当前流
	剪切并移到剪贴板		复制到剪贴板
	粘贴选择		撤销上一次操作
	重做		搜索节点
	编辑流属性		预览 SQL 生成
	运行当前流		运行流选择
	停止流（仅在流处于运行状态时可用）		添加超节点

	放大（仅限于 SuperNodes）		缩小（仅限于 SuperNodes）
	流中无标记		插入注释
	隐藏流标记（如果有）		显示隐藏的流标记
	在 IBM SPSS Modeler Advantage 中打开流		

流标记由流注释、模型链接和评分分支指示组成。
在《IBM SPSS 建模节点》指南中介绍了模型链接。

自定义工具栏

您可以更改工具栏的各个方面，例如：

- 是否显示
- 图标是否有可用工具提示
- 使用大或小图标

要打开或关闭工具栏显示，请执行以下操作：

1. 在主菜单中，单击：
视图 > 工具栏 > 显示

要更改工具提示或图标大小设置，请执行以下操作：

1. 在主菜单中，单击：
视图 > 工具栏 > 定制

根据需要单击**显示工具提示**或**大按钮**。

定制 IBM SPSS Modeler 窗口

使用 SPSS Modeler 界面各部分之间的分界线，可以调整工具的大小或关闭某些工具以满足个人偏好。例如，如果要处理大型流，那么可以使用每条分界线上的小箭头来关闭节点选用板、管理器窗格和项目窗格。这样可以最大化流画布，从而为处理大型流或多个流提供足够的工作空间。

此外，从“视图”菜单上，单击**节点选用板**、**管理器**或**工程**可打开或关闭这些项目的显示。

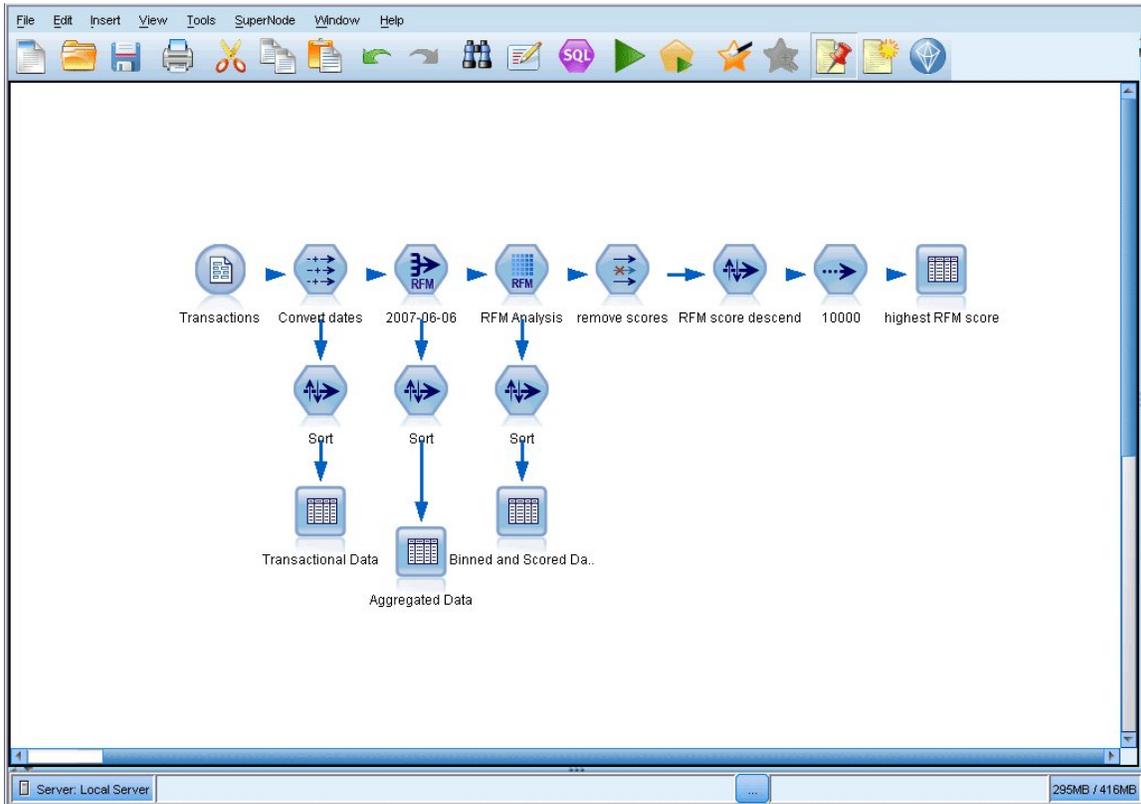


图 10: 最大化的流画布

另外一种关闭节点选用板和管理器以及工程窗格的方法是：垂直或水平移动 SPSS Modeler 窗口侧面或底部的滚动条，将流工作区当作滚动页面使用。

您也可以控制屏幕标记的显示，此标记由流注释、模型链接和评分分支指示组成。要打开或关闭此显示，请单击：

视图 > 流标记

更改流的图标尺寸

您可以通过下列方式更改流图标的大小。

- 流属性设置
- 流中的弹出菜单
- 使用键盘

您可以调整整个流视图的大小，将其调整为标准图标尺寸的 8% 至 200% 之间的某个尺寸。

要调整整个系统的大小（流属性方法）

1. 从主菜单中，选择
工具 > 流属性 > 选项 > 布局。
2. 从“图标大小”菜单中选择所需大小。
3. 单击应用以查看结果。
4. 单击确定保存更改。

要调整整个系统的大小（菜单方法）

1. 右键单击工作区上的流背景。

2. 选择图标尺寸，并选择所需的大小。

要调整整个系统的大小（键盘方法）

1. 同时按住主键盘上的 Ctrl + [-] 来缩小至下一个较小的尺寸。
2. 同时按住主键盘上的 Ctrl + Shift + [+] 来放大至下一个较大的尺寸。

请注意，这一放大方法可能不适用，具体取决于所使用的操作系统和键盘。

获取复杂流的整体视图时，此功能尤其有用。您还可以使用此功能来最大程度地减少打印流所需的页面数。

在 IBM SPSS Modeler 中使用鼠标

IBM SPSS Modeler 中最常见的鼠标用法如下所示：

- **单击。** 使用鼠标右键或左键从菜单中选择选项、打开弹出菜单，以及访问其他标准控件和选项。单击并按住按键可移动和拖动节点。
- **双击。** 双击鼠标左键可将节点放入于流画布中以及编辑现有节点。
- **中键单击。** 单击鼠标中键并拖动光标可连接流画布中的节点。双击鼠标中键可断开某个节点的连接。如果没有三键鼠标，可在单击并拖动鼠标时通过按 Alt 键来模拟此功能。

使用快捷键

IBM SPSS Modeler 中的许多可视化编程操作均有与之关联的快捷键。例如，可通过单击某个节点并按键盘上的 Delete 键将此节点删除。同样地，可在按住 Ctrl 键的同时按 S 键来快速保存某个流。控制命令（例如此命令）由 Ctrl 和其他键的组合指定，例如 Ctrl+S。

标准 Windows 操作中使用了大量快捷键，例如使用 Ctrl+X 来执行剪切操作。IBM SPSS Modeler 不仅支持这些快捷键，而且还支持下列应用程序特定的快捷键。

注：在某些情况下，IBM SPSS Modeler 中使用的旧快捷键与标准 Windows 快捷键相冲突。支持将这些旧快捷键与 Alt 键组合使用。例如，可以使用 Ctrl+Alt+C 来打开或关闭高速缓存。

快捷键	函数
Ctrl+A	全选
Ctrl+X	剪切
Ctrl+N	新建流
Ctrl+O	打开流
Ctrl+P	打印
Ctrl+C	复制
Ctrl+V	粘贴
Ctrl+Z	撤销
Ctrl+Q	选择选定节点的所有下游节点
Ctrl+W	全部不选下游节点（使用 Ctrl+Q 进行切换）
Ctrl+E	从选定节点运行
Ctrl+S	保存当前流
Alt+箭头键	向所使用的箭头方向移动流工作区上的选定节点
Shift+F10	打开选定节点的弹出菜单

表 2: 支持的旧热键快捷键

快捷键	函数
Ctrl+Alt+D	复制节点
Ctrl+Alt+L	加载节点
Ctrl+Alt+R	重命名节点
Ctrl+Alt+U	创建用户输入节点
Ctrl+Alt+C	切换缓存开关
Ctrl+Alt+F	刷新缓存
Ctrl+Alt+X	扩展超节点
Ctrl+Alt+Z	放大/缩小
删除	删除节点或连接

打印

可在 IBM SPSS Modeler 中打印下列对象：

- 流图表
- 图形
- 表
- 报告（来自报告节点和工程报告）
- 脚本（来自“流属性”、“独立脚本”或“超节点脚本”对话框）
- 模型（模型浏览器、包含当前内容的对话框选项卡、树查看器）
- 注解（使用输出的“注解”选项卡）

要打印对象：

- 要不预览就打印，请单击工具栏上的“打印”按钮。
- 要在打印前设置页面，请选择“文件”菜单中的**页面设置**。
- 要在打印前预览，请选择“文件”菜单中的**打印预览**。
- 要查看标准打印对话框中用于选择打印机以及指定外观的选项，请选择“文件”菜单中的**打印**。

实现 IBM SPSS Modeler 的自动化

由于高级数据挖掘往往是一个冗长的复杂过程，因此 IBM SPSS Modeler 包含对几种类型的编码和自动处理的支持。

- **表达式操作控制语言 (CLEM)** 是一种用于分析和操作在 IBM SPSS Modeler 流中流动的数据的语言。数据挖掘人员可在流操作中广泛使用 CLEM 语言来执行根据成本和收入数据推导利润这样的简单任务，也可以执行将 Web 日志数据转换为具有有用信息的一系列字段和记录这样的复杂任务。
- **脚本编写**是用于在用户界面上实现过程自动化的强大工具。脚本可以执行用户使用鼠标或键盘执行的同一类操作。还可以指定输出并处理生成的模型。

第 4 章 了解数据挖掘

数据挖掘概述

数据挖掘 是通过各种技术来识别隐含在数据之中的有价值的信息。通过数据挖掘提取的信息可应用于很多领域，如决策支持、预测、预报和估计。数据通常数量很多，但价值较低，并且以其原始形式显示时并没有什么直接用处。有价值的是隐藏在数据之中的信息。

在数据挖掘中，将您的（或专家的）数据知识与计算机用来识别数据底层关系和特性的高级、主动的分析技术相结合，才能获得成功。数据挖掘过程将利用历史数据生成模型，这些模型将在日后用于预测、模式识别以及更多用途。构建这些模型的技术称为 **机器学习** 或 **建模**。

建模技术

IBM SPSS Modeler 包括了一些机器学习和建模技术，这些技术大致可以根据它们要解决的问题类型进行分组。

- 预测建模方法包括决策树、神经网络和统计模型。
- 聚类模型主要用来确定相似记录的组并根据它们所属的组来为记录添加标签。聚类方法包括 Kohonen、k-means 和两步法。
- 关联规则会将某个特定的结论（如某个特定产品的购买）与一系列条件（若干其他产品的购买）相关联。
- 筛选模型可用于对数据进行筛选以找出最可能与建模相关的字段和记录，还可用于识别与已知模式不匹配的离群值。可用方法包括特征选择和异常检测。

数据操作和发现

IBM SPSS Modeler 还包括了很多可将您的专业知识应用于数据的工具：

- **数据操作**。根据现有数据项派生构建新的数据项，并将数据细分为有意义的子集。来自各种源的数据可以进行合并和过滤。
- **浏览和可视化**。使用数据审核节点显示数据的各个方面（其中包括图形和统计量），以执行初始审核。高级可视化包括交互式图形，这些图形可包含在工程报告中导出。
- **统计**。确认数据中变量之间的可疑关系。IBM SPSS Statistics 中的统计量也可在 IBM SPSS Modeler 中使用。
- **假设检验**。构建数据行为模式的模型，并验证这些模型。

通常情况下，您要使用这些工具来识别数据中的一些可能有价值的属性。然后这些属性可以提供给建模技术，用以尝试确定底层规则和关系。

典型应用

数据挖掘技术的典型应用包括下列内容：

直接邮件。确定哪些人口统计组具有最高的响应率。使用此信息可使得将来邮件的响应度达到最高。

信用评分。使用个人的信用历史记录以确定其信用。

人力资源。了解过去的招聘惯例，创建决策规则，以简化招聘过程。

医学研究。根据医疗证据，创建提出合适流程的决策规则。

市场分析。确定哪些变量（如地理位置、价格和客户特征）与销售量相关。

质量控制。分析产品制造数据，确定决定产品缺陷的变量。

政策研究。通过应用决策规则来选择最重要的变量，使用调查数据制订政策。

医疗保健。可以将用户调查和临床数据相结合，发现有利于健康的变量。

术语

属性、字段和变量 这些涉及单个数据项的术语通用于正在研究之中的所有观测。与某个特定观测相关的属性值集合称为 **记录、示例或观测值**。

评估数据

除非您要使用的数据符合特定的标准，否则数据挖掘将不会有什么成效。以下章节讲述了有关数据及其应用中值得您注意的一些问题。

确保数据可用

这可能看起来很明显，但请注意，虽然数据可能可用，但可能不会是易于使用的形式。IBM SPSS Modeler 可以从数据库（通过 ODBC）或文件导入数据。但是，数据可能在计算机上以某种其他形式保存，而无法直接访问。因此在使用之前，需要以某种合适的形式对其进行下载或转储。数据还可能散落于各种不同的数据库和源中，需要把它们放在一起。甚至这些数据可能不是在线的。如果数据只存在于纸面上，则首先需要进行数据录入，然后才能开始数据挖掘。

检查数据是否包含相关的属性

数据挖掘的目的是为了确定相关属性，因此这项检查初看起来好像很奇怪。但是，查看哪些数据可用并尝试确定可能相关但未记录的因素，这点非常有用。例如，尝试预测冰淇淋销售情况时，您可能有很多零售销路或销售历史记录的相关信息，但您可能没有关于天气和温度信息，而这些信息很可能十分重要。缺少属性不一定意味着数据挖掘无法产生有用的结果，但可能会限制所生成预测的准确性。

评估这种情况的一种快速方式是对数据执行全面的审核。开始审核之前，请将一个数据审核节点与数据源相连，并运行该节点以生成一个完整的报告。

注意噪声数据

数据通常包含错误，也可能包含主观因素，因此可能存在偏差、臆断。这些现象统称为 **噪声**。有时数据中的噪声是正常的。其中可能存在正常的底层规则，但可能不涵盖 100% 的观测值。

通常情况下，数据中的噪声越多，越难获得准确结果。但是，IBM SPSS Modeler 中的机器学习方法能够处理噪声数据，并且已成功用于包含将近 50% 噪声的数据集。

确保有充足的数据

在数据挖掘中，数据集的大小并不一定非常重要。数据集的代表性，以及它对可能结果和变量组合的覆盖范围，这些要重要得多。

通常情况下，考虑的属性越多，需要提供代表性的涉及范围的记录越多。

如果数据具有代表性并且存在通用底层规则，那么可能几千个（甚至几百个）记录组成的数据样本就将提供与一百万个记录组成的数据样本同样理想的结果，并且将更快获取结果。

找到数据专家

在很多情况下，您都是针对自己的数据进行操作，因此对其内容和意义非常熟悉。但如果您要组织对其他部门的数据进行操作，或对客户的数据进行操作，则需要您拥有了解这些数据的专家。这些专家可以指导您确定相关的属性，帮助您解释数据挖掘的结果，并从信息的“黄沙”中淘出“真金”或通过数据集的异常值得“珍宝”。

数据挖掘策略

与大多数业务方面的工作一样，如果以一种有计划的、系统的方式进行数据挖掘，其效率会高很多。即使对于最前沿的数据挖掘工具（如 IBM SPSS Modeler），数据挖掘中的大部分工作也需要一位具有深入知识的业务分析人员来跟踪该过程。为了指导您的计划，请回答下列问题：

- 您要解决的实质问题是什么？
- 哪些数据源可用，数据的哪些部分与当前问题相关？

- 开始挖掘数据之前，您需要进行哪些类型的预处理和数据清理？
- 您要使用哪些数据挖掘技术？
- 您将如何评估数据挖掘分析的结果？
- 您如何从数据挖掘获得的信息中得到最多的益处？

典型的数据挖掘过程可能很快就会变得非常复杂。有很多内容都需要进行跟踪，如复杂的业务问题、多个数据源、数据源中不同的数据质量、一组数据挖掘技术，以及多种用来度量数据挖掘成功与否的方式等等。

一直保持跟踪将有助于针对数据挖掘拥有一个明确定义的过程模型。该过程模型会帮助您回答这部分先前列出的问题，并确保解决要点。它的作用就是一个数据挖掘路线图，使您深陷在挖掘数据的复杂性时不会迷路。

推荐用于 SPSS Modeler 的数据挖掘过程是跨行业标准数据挖掘过程（Cross-Industry Standard Process for Data Mining，简称 CRISP-DM）。从它的名称可以看出，此模型是作为一个通用模型而设计的，它可以应用于多种行业，解决各种商业问题。

CRISP-DM 过程模型

通用的 CRISP-DM 过程模型包括六个用来解决数据挖掘主要问题的阶段。这六个阶段拟合在一个为将数据挖掘应用于较大业务实践而设计的循环过程中。

这六个阶段包括：

- **业务理解。** 这可能是数据挖掘的最重要阶段。商业理解包括确定业务对象、评估情况、确定数据挖掘目标以及制订工程计划。
- **数据理解。** 数据提供了数据挖掘的“原材料”。此阶段用于了解您的数据源以及这些数据的特征。此阶段包括收集初始数据、描述数据、探索数据和验证数据质量。“输出节点”选项板上提供的数据审核节点是一个用于数据理解的、不可或缺的工具。
- **数据准备。** 对数据源进行分类之后，您需要准备数据，以便进行挖掘。准备包括选择、清理、构建、集成数据以及格式化数据。
- **建模。** 此阶段毫无疑问是数据挖掘的核心部分，在此阶段将使用精巧复杂的分析方法从数据中提取信息。此阶段包括选择建模技术、生成测试设计，以及构建和评估模型。
- **评估。** 选定模型之后，就可以评估数据挖掘结果在多大程度上能够帮助您实现业务目标了。此阶段的要素包括评估结果、查看数据挖掘过程，以及确定后续步骤。
- **部署。** 既然您已经付出了上述所有努力，现在就应该有所获益了。此阶段主要是将您的新知识结合到日常的业务流程中，来解决最初的业务问题。此阶段包括计划部署、监视和维护、生成最终报告，以及复查该工程。

此过程模型中有一些关键点。首先，尽管该过程基本上都应该按顺序完成先前段落中所列出的步骤，但也有些地方，各个阶段会以一种非线性的方式互相影响。例如，数据准备通常在建模之前进行。但在建模阶段所作的决策以及所收集的信息通常可以引导您重新考虑数据准备阶段的部分，如此一来可能会出现新的建模问题。这两个阶段互相反馈，直到两个阶段都得以充分解决。与之相似，评估过程可能会引导您重新评估最初的商业理解，并且您可能认为曾经尝试回答的是一个错误的问题。此时，您可以修正商业理解，制订一个更好的目标，然后重新进行该过程的其余部分。

第二个关键点是数据挖掘的迭代特性。下面的情况很少甚至根本不会发生：您简单地规划了一个数据挖掘工程，完成该工程，然后对数据进行打包，回家。然而，数据挖掘专注于解决客户需求，因此这是一项需要不断完善的工作。从一个数据挖掘循环获得的知识几乎无所例外地会导致新的问题、新的争论以及新的机会来识别和满足客户的需求。这些新问题、新争论和新机会通常可以通过再次挖掘您的数据得以解决。这个挖掘和识别新机会的过程不仅应该成为您考虑业务的方式的组成部分，还应该成为整个业务策略的基石。

上述内容仅是针对 CRISP-DM 过程模型的概述。有关该模型的完整详细信息，请参阅下列资源：

- 《CRISP-DM 指南》，该指南与其他文档可从安装光盘上的 \Documentation 文件夹进行访问。
- CRISP-DM 帮助系统，可从“开始”菜单访问，也可通过单击 IBM SPSS Modeler“帮助”菜单中的 **CRISP-DM 帮助** 进行访问。

模型类型

IBM SPSS Modeler 提供了各种借助机器学习、人工智能和统计学的建模方法。通过建模选项板中的方法，您可以根据数据生成新的信息以及开发预测模型。每种方法各有所长，同时适用于解决特定类型的问题。

IBM SPSS Modeler 应用程序指南 为上述多种方法提供了示例以及建模过程的一般介绍。本指南作为联机教程提供，也有 PDF 格式。请参阅主题第 3 页的『应用程序示例』，以获取更多信息。

建模方法分为以下类别：

- 受监督
- 关联
- 细分

受监督模型

受监督模型使用一个或多个**输入**字段的值来预测一个或多个输出（或**目标**）字段的值。这些技术的一些示例包括：决策树（C&R 树、QUEST、CHAID 和 C5.0 算法）、回归（线性、logistic、广义线性和 Cox 回归算法）、神经网络、支持向量机和贝叶斯网络。

“受监督”模型可帮助组织预测已知的结果，例如顾客是否购买、流失或某交易是否符合某种已知的犯罪模式。其建模方法包括机器学习、规则归纳、子组标识、统计方法和多模型生成。

受监督节点



“自动分类器”节点用于创建和对比二元结果（是或否，流失或不流失等）的若干不同模型，使用户可以选择给定分析的最佳处理方法。由于支持多种建模算法，因此可以对用户希望使用的方法、每种方法的特定选项以及对比结果的标准进行选择。该节点根据指定的选项生成一组模型，并根据您指定的条件对最佳候选项进行排序。



自动数字节点使用多种不同方法估计和对比模型的连续数字范围结果。此节点和自动分类器节点的工作方式相同，因此可以选择要使用和要在单个建模传递中使用多个选项组合进行测试的算法。受支持的算法包括神经网络、C&R 树、CHAID、线性回归、广义线性回归以及支持向量机 (SVM)。可以根据相关度、相对误差或使用的变量数来比较模型。



分类和回归 (C&R) 树节点生成可用于预测或分类未来观测值的决策树。该方法通过在每个步骤最大限度降低不纯度，使用递归分区来将训练记录分割为组。如果树中某个节点中 100% 的观测值都属于目标字段的一个特定类别，那么该节点将被认定为“纯洁”。目标和输入字段可以是数字范围或分类（名义、有序或标志）；所有拆分都是二进制的（只有两个子组）。



QUEST 节点可提供用于构建决策树的二元分类法，此方法的设计目的是减少大型 C&R 树分析所需的处理时间，同时也减少在分类树方法中发现的趋势以便支持允许有多个分割的输入。输入字段可以是数字范围（连续），但目标字段必须是分类。所有分割都是二元的。



CHAID 使用卡方统计来生成决策树，以确定最佳的分割。CHAID 与 C&R 树和 QUEST 节点不同，它可以生成非二元树，这意味着有些分割将有多于两个的分支。目标和输入字段可以是数字范围（连续）或分类。穷举 CHAID 是 CHAID 的修正版，它可以更彻底地检查所有可能的拆分，但计算时间较长。



C5.0 节点构建决策树或规则集。该模型的工作原理是根据在每个级别提供最大信息收获的字段分割样本。目标字段必须为分类字段。允许进行多次多于两个子组的分割。



决策列表节点可标识子组或段，显示与总体相关的给定二元结果的似然度的高低。例如，您或许在寻找那些最不可能流失的客户或最有可能对某个商业活动作出积极响应的客户。通过定制段和并排预览备选模型来比较结果，您可以将自己的业务知识体现在模型中。决策列表模型由一组规则构成，其中每个规则具备一个条件和一个结果。规则依顺序应用，相匹配的第一个规则将决定结果。



线性回归模型根据目标与一个或多个预测变量之间的线性关系来预测连续目标。



“PCA/因子”节点提供用于降低数据复杂程度的强大数据降维技术。主成份分析（PCA）可找出输入字段的线性组合，该组合最好地捕获了整个字段集合中的方差，且组合中的各个成分相互正交（相互垂直）。因子分析则尝试识别底层因素，这些因素说明了观测的字段集合内的相关性模式。对于这两种方法，其共同的目标是找到可对原始字段集合中的信息进行有效总结的少量派生字段。



“特征选择”节点根据一组条件（例如缺失值百分比）筛选要移除的输入字段，然后，相对于指定目标对余下的输入的重要性进行排序。例如，假如某个给定数据集有上千个潜在输入，那么哪些输入最有可能用于对患者结果进行建模呢？



判别分析提出比 logistic 回归更严格的假设，但当满足这些假设时，它可以成为 logistic 回归分析的宝贵替代或补充。



Logistic 回归是一种统计方法，它可根据输入字段的值对记录进行分类。它与线性回归类似，但采用分类目标字段而不是数字范围。



广义线性模型对广义线性模型进行了扩展，这样因变量通过指定的关联函数与因子和协变量线性相关。而且，该模型还允许因变量呈非正态分布。它涵盖了大量统计模型的功能，包括计数数据的线性回归，logistic 回归，对数线性模型以及区间删失的生存模型。



广义线性混合模型 (GLMM) 扩展了线性模型，使得目标可以有非正态分布，通过指定的连接函数与因子和协变量线性相关，并且观测值可能相关。广义线性混合模型涵盖了各种模型，从简单线性回归模型到非正态纵向模型数据的复杂多级模型。



使用 Cox 回归节点，您可以在已有的检查记录中建立时间事件的生存模型。对于输入变量的给定值，该模型会生成一个生存函数，用来预测在给定时间 (t) 发生相关事件的概率。



使用支持向量机 (SVM) 节点，可以将数据分为两组，而无需过度拟合。SVM 可以与宽数据集配合使用，例如那些含有大量输入字段的数据集。



通过贝叶斯网络节点，你可以利用对真实世界认知的判断力并结合所观察和记录的证据来构建概率模型。该节点侧重于主要用于分类的树增强朴素贝叶斯 (TAN) 和马尔可夫毯网络。



自学响应模型 (SLRM) 节点可用于构建一个包含单个新观测值或少量新观测值的模型，通过此模型，无需使用全部数据对模型进行重新训练即可对模型进行重新评估。



时间序列节点估计时间序列数据的指数平滑模型、单变量自回归整合移动平均值 (ARIMA) 模型和多变量 ARIMA (即变换函数) 模型，并生成未来性能的预测数据。此“时间序列”节点类似于 SPSS Modeler V18 中不推荐使用的先前“时间序列”节点。但是，此较新“时间序列”节点旨在利用 IBM SPSS Analytic Server 的能力来处理大数据，并在 SPSS Modeler V17 中添加的输出查看器中显示生成的模型。



The k -最近相邻元素 (KNN) 节点将新的观测值关联到预测变量空间中与其最邻近的 k 个对象的类别或值 (其中 k 为整数)。类似观测值相互靠近，而不同观测值相互远离。



空间-时间预测 (STP) 节点使用包含位置数据、预测输入字段 (预测变量)、时间字段和目标字段的数据。每个位置在数据中都有许多行，这些行表示每个预测变量在每个测量时间的值。分析数据后，可以使用该数据来预测分析中使用的形状数据内任意位置处的目标值。

关联模型

关联模型查找您数据中的模式，其中一个或多个实体 (如事件、购买或属性) 与一个或多个其他实体相关联。这些模型构建定义这些关系的规则集。数据中的字段可以作为输入和目标。您可以手动查找这些关联，但关联规则算法可以更快速地完成，并能探索更多复杂的模式。Apriori 和 Carma 模型是使用此类算法的示例。另一种类型的关联模型是序列检测模型，后者可以在按时间建立结构的数据中查找顺序模式。

在预测多个结果时，关联模型最为有用，例如，购买了产品 X 的客户同时也购买了 Y 和 Z。关联模型可将特定结论 (例如，购买某物的决策) 与一组条件相关联。关联规则算法相对于更标准的决策树算法 (C5.0 和 C&RT) 的优势在于，它可以找到任何属性间存在的关联。决策树算法只使用单一结论来构建规则，而关联算法则试图找到更多规则，且每个规则具有不同的结论。

关联节点



“先验”节点从数据抽取一组规则，即抽取信息内容最多的规则。Apriori 节点提供五种选择规则的方法并使用复杂的索引模式来高效地处理大数据集。对于较大的问题，Apriori 训练的速度通常较快；它对可保留的规则数量没有任何限制，而且可处理最多带有 32 个前提条件的规则。“先验”要求输入和输出字段均为分类型字段，因为它专为处理此类型数据而进行优化，因而处理速度快得多。



CARMA 模型在不要求用户指定输入或目标字段的情况下从数据抽取一组规则。与 Apriori 相反，CARMA 节点提供规则支持的构建设置 (支持前项和后项)，而不仅仅是前项支持。这就意味着生成的规则可以用于更多应用程序，例如用于查找产品或服务 (前项) 的列表，这些产品或服务的后项为想在节日期间促销的商品。



序列节点可发现连续数据或与时间有关的数据中的关联规则。序列是一系列可能会以可预测顺序发生的项目集合。例如，一个购买了剃刀和须后水的顾客可能在下次购物时购买剃须膏。序列节点基于 CARMA 关联规则算法，该算法使用一个有效的两次传递方法查找序列。



“关联规则”节点与 Apriori 节点类似；但是，与 Apriori 不同，“关联规则”节点能够处理列表数据。另外，“关联规则”节点可以与 IBM SPSS Analytic Server 配合使用，以处理大型数据以及利用更快的并行处理功能。

细分模型

细分模型将数据划分为具有类似输入字段模式的记录段或聚类。细分模型只对输入字段感兴趣，没有输出或目标字段的概念。细分模型的示例为 Kohonen 网络、K-Means 聚类、二阶聚类和异常检测等。

在不知道特定结果的情况下（例如，需要识别新犯罪模式或在客户群中识别利益群体时），细分模型（也称为“聚类模型”）非常有用。聚类模型主要用来确定相似记录的组并根据它们所属的组来为记录添加标签。此方法的优点在于，不用提前了解这些组及其特征就可以使用，它使聚类模型（其中没有需要模型预测的预定义输出或目标字段）区别于其他的建模技术。对于这些模型来说，没有正确或错误的结果之分。模型的值由模型捕获数据中感兴趣的分组并提供这些分组的有用说明信息的能力来确定。聚类模型通常用于创建在后续分析中用作输入的聚类或段（例如，将潜在用户分成几个相似的子组）。

细分节点



“自动聚类”节点估算和比较识别具有类似特征记录组的聚类模型。节点工作方式与其他自动建模节点相同，使您在一次建模运行中即可试验多个选项组合。可使用基本度量对模型进行比较，尝试对聚类模型进行过滤，对其有用性进行排名，并提供基于特定字段重要性的度量。



K-Means 节点将数据集聚类到不同分组（或聚类）。此方法将定义固定的聚类数量，将记录迭代分配给聚类，以及调整聚类中心，直到进一步优化无法再改进模型。k-means 节点作为一种非监督学习机制，它并不试图预测结果，而是揭示隐含在输入字段集中的模式。



Kohonen 节点会生成一种神经网络，此神经网络可用于将数据集聚类到各个差异组。此网络训练完成后，相似的记录应在输出映射中紧密地聚集，差异大的记录则应彼此远离。您可以通过查看模型块中每个单元所捕获观测值的数量来找出规模较大的单元。这将让您对聚类的相应数量有所估计。



TwoStep 节点使用二阶聚类方法。第一步完成简单数据处理，以便将原始输入数据压缩为可管理的子聚类集合。第二步使用层级聚类方法将子聚类一步一步合并为更大的聚类。TwoStep 具有一个优点，就是能够为训练数据自动估计最佳聚类数。它可以高效处理混合的字段类型和大型的数据集。



Anomaly Detection 节点确定不符合“正常”数据格式的异常观测值（离群值）。通过此节点，即使界外值不符合任何先前已知的模式，也可以识别这些界外值，即使您不确定要查找的内容也是如此。

数据库内数据挖掘模型

IBM SPSS Modeler 支持与可从数据库供应商获取的数据挖掘和建模工具集成，包括 Oracle Data Miner 和 Microsoft Analysis Services。可以在 IBM SPSS Modeler 应用程序内的所有数据库中构建、评分和存储模型。有关完整详细信息，请参阅《IBM SPSS Modeler 数据库内挖掘指南》。

IBM SPSS Statistics 模型

如果您在计算机上拥有 IBM SPSS Statistics 安装和许可的一个副本，您可以从 IBM SPSS Modeler 访问和运行某些 IBM SPSS Statistics 例程以构建模型和给模型评分。

数据挖掘示例

从示例入手是了解数据挖掘实际应用的最佳方式。《IBM SPSS Modeler 应用程序指南》中包含很多应用程序示例，同时还提供了有关特定建模方法和技术的简要针对性说明。有关更多信息，请参阅主题 [第 3 页的『应用程序示例』](#)。

第 5 章 构建流

流构建概述

使用 IBM SPSS Modeler 进行的数据挖掘重点关注让数据流经一系列节点的过程，我们将这一过程称为**流**。这一系列节点代表要对数据执行的操作，而节点之间的链接指示数据流的方向。通常，您可以使用数据流将数据读入 IBM SPSS Modeler，对其执行一系列操作，然后将其发送至某个目标，例如表格或查看器。

例如，假定您需要打开某个数据源、添加新字段、根据新字段中的值选择记录，然后在表中显示结果。在这种情况下，您的数据流应由以下四个节点组成：



“变量文件”节点，设置此节点后，可以从数据源读取数据。



“派生”节点，用于向数据集中添加新的计算字段。



“选择”节点，用于设置选择标准，以从数据流中排除某些记录。



“表”节点，用于在屏幕上显示操作结果。

构建数据流

独具特色的 SPSS Modeler 界面使您能够使用数据流图直观地挖掘数据。从最基本的操作开始，您可采用下列步骤构建数据流：

- 将节点添加到流工作区。
- 连接节点以形成流。
- 指定任意节点或流选项。
- 运行流。

本部分包含有关使用节点创建较复杂数据流的更多详细信息。还介绍了节点和流的选项和设置。有关使用 SPSS Modeler（在程序安装的 Demos 文件夹中）随附的数据构建流的逐步示例，请参阅第 3 页的『应用程序示例』。

使用节点

在 IBM SPSS Modeler 中，节点用于帮助您研究数据。工作区中的各种节点分别代表不同的对象和操作。IBM SPSS Modeler 窗口底部的选用板包含流构建中可用的所有节点。

存在多种类型的节点。**源节点**将数据导入到流中，它位于节点选用板的“源”选项卡上。**过程节点**在单个数据记录和字段上执行操作，它位于选用板的“记录选项”和“字段选项”选项卡上。**输出节点**为数据、图表和模型结果生成各种输出，并包含在节点选用板的“图形”、“输出”和“导出”选项卡上。**建模节点**使用统计算法创建模型块，它位于节点选用板的“建模”选项卡和“数据库建模”（如激活）选项卡上。有关更多信息，请参阅主题第 11 页的『节点选用板』。

连接节点会形成流，当运行流时，可以直观地查看节点之间的关系并得出结论。流类似于脚本 - 可以保存并将其重新用于其他数据文件。

处理流数据的可运行节点称为**终端节点**。如果建模节点或输出节点位于流或流分支的末端，则为终端节点。您不能将其他节点连接到终端节点。

注: 可以自定义节点选用板。有关更多信息，请参阅主题 [第 196 页的『自定义节点选项板』](#)。

将节点添加至流

可以通过多种方法将节点从节点选用板添加至流:

- 双击选用板中的节点。注: 双击某个节点会自动将其连接到当前流。有关更多信息，请参阅主题 [第 28 页的『连接流中的节点』](#)。
- 将节点从选用板拖放到流工作区。
- 单击选用板中的节点，然后单击流工作区。
- 从 IBM SPSS Modeler 的“插入”菜单中选择合适的选项。

将节点添加至流工作区后，双击该节点可显示相应的对话框。可用选项取决于您要添加的节点的类型。有关对话框中特定控件的信息，请单击其**帮助**按钮。

移除节点

要从数据流中移除某个节点，请单击该节点并按 Delete 键，或者单击鼠标右键并从菜单中选择**删除**。

连接流中的节点

已添加到流工作区的节点在连接之前不会形成数据流。节点之间的连接指示数据从一项操作流向下一项操作的方向。连接节点以形成流的方法有以下几种: 双击、使用鼠标中键或手动连接。

通过双击添加并连接节点

形成流的最简单的方法是双击选项板中的节点。此方法会自动将新节点连接到流工作区中的选定节点。例如，如果工作区包含一个“数据库”节点，则可以选中该节点，然后在选项板中双击下一个节点，如“导出”节点。此操作会自动将“导出”节点连接到现有的“数据库”节点。您可以重复此过程，直到到达终端节点（如“直方图”或“表”节点），这时所有新节点都将连接到上游最后一个非终端节点。

使用鼠标中键连接节点

在流工作区中，可以使用鼠标中键单击某个节点并将其拖到另一个节点。（如果鼠标没有中键，可以通过按住 Alt 键的同时使用鼠标从一个节点拖到另一个节点来模拟此操作。）

手动连接节点

如果没有鼠标中键并希望手动连接节点，您可以使用节点的弹出菜单将其连接到工作区已有的另一个节点。

1. 右键单击选择连接的起始节点。这将打开节点菜单。
2. 在菜单中单击**连接**。
3. 此时，开始节点和光标处将同时显示连接图标。单击工作区中的第二个节点以连接这两个节点。

连接节点时，需要遵循以下几项规则。如果尝试进行以下任何一种连接，您将收到错误消息:

- 指向源节点的连接
- 发自终端节点的连接
- 超出节点的最大输入连接数
- 连接两个已连接的节点
- 循环（数据返回其从中流出的节点）

绕过流中的节点

如果绕过数据流中的某个节点，该节点的所有输入和输出连接都将替换为直接从其输入节点通向其输出节点的连接。如果该节点不同时具备输入连接和输出连接，则会删除该节点的所有连接，而不是更改路线。

例如，您可能有这样的一个流，它导出一个新字段、过滤字段，然后研究直方图和表中的结果。如果还希望在过滤字段之前查看相同的数据图形和数据表，您可以在流中添加新的直方图节点和表节点，也可以绕过过滤节点。绕过“过滤”节点时，将直接通过“导出”节点连接到图形和表。“过滤”节点将从流中断开连接。

绕过节点

1. 在流工作区中，使用鼠标中键双击要绕过的节点。或者，您还可使用 Alt+双击。

注：可以在“编辑”菜单中单击**撤消**或按 Ctrl+Z 撤消此操作。

禁用流中的节点

可以禁用流中具有单个输入的过程节点，这样在流运行期间将忽略此节点。这种方式使您不必删除或绕过节点，并保持其连接到其余节点。您还可以打开和编辑节点设置，不过，除非重新启用节点，否则任何更改均不会生效。

例如，您可能有一个过滤多个字段，然后使用减少后的数据集构建模型的流。如果还需要构建未过滤字段的相同模型，以查看这些字段是否改进模型结果，可以禁用“过滤”节点。禁用“过滤”节点时，到建模节点的连接将直接从“导出”节点到“类型”节点。

禁用节点

1. 在流工作区中，右键单击要禁用的节点。

2. 在弹出菜单中单击**禁用节点**。

或者，您也可以单击“编辑”菜单上的**节点 > 禁用节点**。当需要重新在流中包含节点时，可采用相同方式单击**启用节点**。

注：可以在“编辑”菜单中单击**撤消**或按 Ctrl+Z 撤消此操作。

在现有连接中添加节点

通过拖动用于连接两个节点的箭头，您可以在这两个已连接的节点之间添加新节点。

1. 使用鼠标中键单击并拖动要向其中插入节点的连接箭头。也可以在按住 Alt 键的同时单击并拖动鼠标以模拟鼠标中键。

2. 将连接拖动到您要包含的节点，然后松开鼠标按键。

注：可以通过**绕过**节点将新连接从节点中移除并复原原始节点。

删除节点之间的连接

要删除两个节点间的连接：

1. 右键单击连接箭头。
2. 在菜单中单击**删除连接**。

要删除从某个节点开始以及到该节点的所有连接，请执行以下一种操作：

- 选中节点并按 F3。
- 选择该节点，然后在主菜单中单击：

编辑 > 节点 > 断开连接

节点的设置选项

创建并连接节点后，可使用一些自定义节点的选项。右键单击节点，然后选择某个菜单选项。

- 单击**编辑**可打开选定节点的对话框。
- 单击**连接**可将一个节点手动连接到另一个节点。
- 单击**断开连接**可删除出入节点的所有链接。
- 单击**重命名并加注解**可打开“编辑”对话框的“注解”选项卡。
- 单击**新注释**可添加节点相关的注释。有关更多信息，请参阅主题 [第 42 页的『添加注释和注解到节点和流』](#)。
- 单击**禁用节点**可在处理期间“隐藏”节点。要使节点再次对处理“可见”，可以单击**启用节点**。有关更多信息，请参阅主题 [第 29 页的『禁用流中的节点』](#)。

- 单击**剪切**或**删除**可从流工作区中删除选定节点。注：单击**剪切**可以粘贴节点，单击**删除**则不能。
- 单击**复制节点**可生成不包含连接的节点副本。可以将此副本添加到新的或现有的流中。
- 单击**加载节点**可打开先前保存的节点，并将其选项加载当前选定的节点。这两个节点必须属于同一类型。
- 单击**检索节点**可从连接的 IBM SPSS Collaboration and Deployment Services Repository 中检索节点。
- 单击**保存节点**可将节点的详细信息保存在文件中。只能将节点详细信息加载其他类型相同的节点。
- 单击**存储节点**可将所选节点存储在连接的 IBM SPSS Collaboration and Deployment Services Repository 中。
- 单击**高速缓存**可展开包含对选定节点进行高速缓存选项的菜单。
- 单击**数据映射**可展开包含将数据映射到新源或指定必备字段选项的菜单。
- 单击**创建超节点**可展开包含在当前流中创建超节点选项的菜单。
- 单击**生成用户输入节点**可替换选定节点。此节点生成的示例将包含与当前节点相同的字段。
- 单击**从此处运行**可从选定节点向下运行所有终端节点。

节点的缓存选项

为优化流运行，可以在任何非终端节点上建立缓存。如果已在节点上建立了缓存，则在下一次运行数据流时流过节点的数据会充满缓存。从这时起，就会从缓存（存储在磁盘的临时目录中）而不是从数据源中读取数据了。

对于诸如排序、合并或汇总这样比较耗时的操作，缓存是最有用的。例如，假设已设置可从数据库中读取销售数据的源节点，和可按区域汇总销售数据的“汇总”节点。可以在“汇总”节点上建立缓存而不是在源节点上建立缓存，因为希望缓存存储的是已汇总的数据而不是整个数据集。

注：源节点位置的缓存仅用于存储读入 IBM SPSS Modeler 的原始数据复本，在大多数情况下，它不会带来性能方面的提升。

对于已启用缓存的节点，会在右上角显示一个小的文档图标。当节点上的数据被缓存时，文档图标将变绿。

启用缓存

1. 在流工作区中右键单击节点，然后在菜单中单击**高速缓存**。
2. 在缓存子菜单中，单击**启用**。
3. 可以通过右键单击节点，然后在缓存子菜单中单击**禁用**来关闭缓存。

数据库中的缓存节点

对于在数据库中运行的流，可于中游将数据缓存到数据库的临时表中而不是文件系统中。如果与 SQL 优化相组合，则此操作将使性能得到显著提高。例如，可以对合并多个表以创建数据挖掘视图的流的输出进行缓存并在需要时重新使用。通过为所有下游节点自动生成 SQL，性能可得到进一步的提高。

为利用数据库缓存，必须同时启用 SQL 优化和数据库缓存。注意，服务器上的优化设置将覆盖客户端上的优化设置。有关更多信息，请参阅主题 [第 34 页的『设置流的优化选项』](#)。

如果已启用数据库缓存，则只需用右键单击任意非终端节点就可以缓存该点上的数据，并且在下次运行流时，可以直接在数据库中自动创建缓存。如果未启用数据库缓存或 SQL 优化，则会转而将缓存写入文件系统中。

注：下列数据库支持用于缓存的临时表：Db2、Oracle、SQL Server 和 Teradata。其他数据库（例如 Netezza）将使用标准表进行数据库缓存。可以针对特定数据库定制 SQL 代码 - 请与服务人员联系以获取帮助。

刷新缓存

节点上的白色文档图标指示其缓存为空。当缓存已满时，该文档图标将保持绿色不变。如果要替换缓存的内容，必须先刷新缓存，然后重新运行数据流以对其进行重新填充。

1. 在流工作区中右键单击节点，然后在菜单中单击**高速缓存**。
2. 在缓存子菜单中，单击**刷新**。

保存缓存

您可以将高速缓存的内容保存为 IBM SPSS Statistics 数据文件 (*.sav)。然后，您可以将该文件重新装入为高速缓存，也可以设置一个使用该高速缓存文件作为其数据源的节点。此外，还可以加载其他工程中保存的缓存。

1. 在流工作区中右键单击节点，然后在菜单中单击**高速缓存**。
2. 在缓存子菜单中，单击**保存缓存**。
3. 在“保存缓存”对话框中，浏览至要保存缓存文件的位置。
4. 在“文件名”文本框中输入名称。
5. 确保在“文件类型”列表中选中 ***.sav**，然后单击**保存**。

加载缓存

如果在将某个缓存文件从节点中删除之前已对其进行了保存，则可以重新加载该文件。

1. 在流工作区中右键单击节点，然后在菜单中单击**高速缓存**。
2. 在缓存子菜单中，单击**加载缓存**。
3. 在“加载缓存”对话框中浏览至缓存文件所在位置，选中该文件，然后单击**加载**。

预览节点中的数据

为确保在构建流时数据以期望的方式更改，可以在每个重要步骤上通过“表”节点运行数据。为避免这样做，可以从每个节点生成会显示将要创建的数据样本的预览，从而减少构建每个节点所需的时间。

对于模型块的上流节点，预览显示输入字段；对于模型块或其下游节点（不含端节点），预览显示输入和产生的字段。

缺省显示的行数为 10，不过可以在流属性中更改此值。有关更多信息，请参阅主题 [第 32 页的『设置流的常规选项』](#)。

从**生成**菜单中，可以创建多种类型的节点。

注：预览此节点生成的数据时，将对此节点应用所有属性更改且无法取消（此行为与单击**应用**相同）。

锁定节点

为了防止其他用户修改流中一个或多个节点的设置，您可以将节点封装在称为超节点的特殊类型的节点中，然后通过使用密码保护锁定超节点。

使用流

在流工作区中连接源、过程和终端节点后，便创建了一个流。作为节点集合，可以对流进行保存、添加注解，还可以将其添加到工程。此外，还可以为流设置各种选项，如优化、日期和时间设置、参数和脚本。这些属性将在下面的主题中进行介绍。

在 IBM SPSS Modeler 中，可以在同一 IBM SPSS Modeler 会话中使用和修改多个数据流。主窗口的右侧包含管理器窗格，可帮助您在当前打开的流、输出和模型中进行导航。如果您无法看到管理器窗格，请在“视图”菜单中单击**管理器**，然后单击**流**选项卡。

通过此选项卡，可以执行下列操作：

- 访问流。
- 保存流。
- 将流保存至当前工程。
- 关闭流。
- 打开新流。

- 在 IBM SPSS Collaboration and Deployment Services 存储库中存储和检索流（如果在您的站点上可用）。有关更多信息，请参阅主题 [第 163 页的『关于 IBM SPSS Collaboration and Deployment Services Repository』](#)。

在“流”选项卡中右键单击某个流可访问这些选项。

设置流的选项

您可以指定众多选项以应用于当前流。还可将这些选项保存为缺省值以应用于所有流。这些选项如下。

- **常规。** 诸如要在流中使用的符号和文本编码等杂项。有关更多信息，请参阅[第 32 页的『设置流的常规选项』](#)。
- **日期/时间。** 有关日期和时间表达式格式的选项。有关更多信息，请参阅[第 33 页的『设置流的日期与时间选项』](#)。
- **数字格式。** 用于控制数字表达式格式的选项。有关更多信息，请参阅[第 34 页的『设置流的数字格式选项』](#)。
- **优化。** 这是用于提高流性能的选项。有关更多信息，请参阅[第 34 页的『设置流的优化选项』](#)。
- **日志记录和状态。** 用于控制 SQL 日志和记录状态的选项。有关更多信息，请参阅[第 35 页的『设置流的 SQL 日志和记录状态选项』](#)。
- **布局。** 有关流在工作区上的布局的选项。有关更多信息，请参阅[第 35 页的『设置流的布局选项』](#)。
- **Analytic Server。** 有关将 Analytic Server 与 SPSS Modeler 配合使用的选项。有关更多信息，请参阅[第 36 页的『Analytic Server 流属性』](#)。
- **地理空间。** 有关对地理空间数据进行格式编排以便在流中使用的选项。有关更多信息，请参阅[第 36 页的『设置流的地理空间选项』](#)。

设置流选项

1. 在“文件”菜单中，单击**流属性**（或从管理器窗格中的“流”选项卡上选择流，右键单击并在弹出菜单中单击**流属性**）。
2. 单击**选项**选项卡。

或者，在“工具”菜单中，单击：

流属性 > 选项

设置流的常规选项

常规选项包含一组应用于当前流的不同方面的杂项。

基本部分提供了下列基本选项：

- **十进制符号。** 选择逗号 (,) 或句号 (.) 作为小数分隔符。
- **分组符号。** 针对数字显示格式，选择用于对值进行分组的符号（例如，3,000.00 中的逗号）。选项包括“无”、“句号”、“逗号”、“空格”和“定义的语言环境”（在该情况下将采用当前语言环境的缺省设置）。
- **编码。** 指定文本编码的流缺省方法。（注：适用于变量。文件源节点和平面文件导出节点。其他节点不使用此设置；多数数据文件带有嵌入编码信息。您可以选择系统缺省值或 UTF-8。系统缺省值在 Windows 控制面板中指定，如果以分布式模式运行，则在服务器计算机上指定。有关更多信息，请参阅主题 [第 215 页的『IBM SPSS Modeler 中的 Unicode 支持』](#)。
- **规则集评估。** 确定规则集模型的评估方式。缺省情况下，规则集将采用**投票**方式综合各项规则产生的预测值，并确定最终预测值。为确保规则集缺省使用第一个匹配规则，可选择**第一个匹配项**。请注意，此选项不适用于决策列表模型，该模型始终使用由算法定义的第一个匹配项。

在数据预览中显示的最大行数。 指定当请求某个节点的数据预览时显示的行数。有关更多信息，请参阅主题 [第 31 页的『预览节点中的数据』](#)。

名义字段的最大成员数。 选择此选项可指定名义（集合）字段的最大成员数，达到此数值后字段的数据类型将变为**无类型**。当使用较大名义字段时，此选项十分有用。注：当字段的测量级别设置为**无类型**时，其角色将自动设置为**无**。这意味着这种字段无法用于建模。

限制用于 Kohonen 和 K-Means 建模的集大小。 选择此选项可指定用于 Kohonen 网络和 K-Means 建模的集合字段的最大成员数。缺省集合大小为 20，达到此数值后，将忽略该字段并发出警告，其中显示有关该字段的错误信息。

请注意，为了确保兼容，此选项还将应用于在 IBM SPSS Modeler V14 中已被替换的旧神经网络节点；一些旧流可能仍包含此节点。

执行时刷新源节点。 选择此选项将在运行当前流时自动刷新所有源节点。此操作类似于针对某个源节点点击刷新按钮，区别在于此选项会自动刷新当前流的所有源节点（用户输入节点除外）。

注：选择此选项将刷新下游节点的高速缓存，即使数据未发生更改也是如此。如果您使用工具栏中的**运行当前流**选项，但每次运行流只进行一次刷新，这意味着仍可使用下游缓存作为单次运行时的临时存储器。例如，假定您在进行某项复杂的导出操作后设置了中游高速缓存，且已为该导出节点的下游附加若干图形和报告。运行流时，将刷新导出节点的高速缓存，并仅重新填充第一个图形或报告。后续终端节点将从该“导出”节点的高速缓存中读取数据。请注意，如果您选择单独执行每一个终端节点（如果有多个终端节点），而不是使用**运行当前流**选项，那么每次执行终端节点时都会发生缓存刷新。

在输出中显示字段和值标签。 在表、图表和其他输出中显示字段和值标签。如果标签不存在，则会改为显示字段名和数据值。缺省情况下，标签处于关闭状态；但您可以在 IBM SPSS Modeler 的其他位置单独切换各个标签。此外，还可以使用工具栏中的切换按钮选择在输出窗口中显示标签。



图 11: 用于切换字段和值标签的工具栏图标

显示执行时间。 运行流之后，在“执行时间”选项卡上显示流节点的相应执行时间。有关更多信息，请参阅主题第 37 页的『查看节点执行时间』。

自动创建节点部分提供了下列用于在各个流中自动创建节点的选项。这些选项控制生成新建模块时是否将这些块插入到流工作区中。缺省情况下，这些选项仅应用于 V16 或更高版本中创建的流。在 IBM SPSS Modeler 16 或更高版本中，如果打开在 V15 或之前版本中创建的流并执行建模节点，那么不会像前发行版那样将块放入流工作区。如果您使用 IBM SPSS Modeler 16 或更高版本创建新流并执行建模节点，那么生成的块将放入流工作区。这与设计的相同，例如，**为新模型输出创建模型应用节点**选项可能会中断在 IBM SPSS Collaboration and Deployment Services 中以及不存在 IBM SPSS Modeler Server 客户机用户界面的其他环境中以批处理方式运行的 V16 版本之前的流。

- **为新模型输出创建模型应用节点。** 自动地为新模型输出创建模型应用节点。如果选中此选项，那么还可以从**创建模型更新链接**中选择是 将这些链接设置为启用、设置为禁用还是不创建这些链接。

创建新的模型应用节点或源节点时，下拉列表中的链接选项用于控制是否在构建器节点与新节点之间创建更新链接，以及创建这些链接后它们所处的模式。如果创建这些链接，那么您有机会将其启用，但这些选项使用户能够进行全面控制。

- **从源构建器创建源节点。** 自动地根据源构建器创建源节点。与上一个选项相似，如果选中此选项，那么还可以从**创建源刷新链接**下拉列表中选择是 将刷新链接设置为启用、设置为禁用还是不创建这些链接。

另存为缺省值。 指定的选项仅适用于当前流。单击此按钮以将这些选项设置为所有流的缺省值。

设置流的日期与时间选项

这些选项指定当前流中的各种日期和时间表达式要使用的格式。

将日期/时间导入为 选择是否将日期/时间用于日期/时间字段或者是否将它们导入为字符串变量。

日期格式 选择要用于日期存储字段的日期格式或 CLEM 日期函数将字符串解析为日期时使用的日期格式。

时间格式 选择要用于时间存储字段的时间格式或 CLEM 时间函数将字符串解析为时间时使用的时间格式。

回滚天数/分钟数 对于时间格式，请选择是否将负时差解释为指示前一天或前一小时。

日期基线（1 月 1 日） 选择采用单日期的 CLEM 日期函数要使用的基年（始终为 1 月 1 日）。

两位日期（起始自） 指定要为仅以两位数表示的年度添加公元位的截止年度。例如，指定 1930 为截止年度会假定 05/11/02 为 2002 年。相同的设置将针对 30 后的日期使用 20 世纪公元位，因此 05/11/73 将假定 为 1973 年。

时区 选择如何选取时区以用于 `datetime_now` CLEM 表达式。

- 如果您选择**服务器**，那么时区将取决于下列各项：
 - 如果当前流使用 Analytic Server 数据源，那么 `datetime_now` 表达式将使用 Analytic Server 中的时间；缺省情况下，此服务器将使用全球标准时间。
 - 如果当前流使用数据库源节点，那么受支持的数据库将使用 SQL 回送，并且 `datetime_now` 表达式将使用数据库时间。
 - 对于所有其他流，时区将使用 SPSS Modeler Server 中的时间。
- 如果您选择 **Modeler Client**，那么时区将反映安装了 SPSS Modeler 的机器的时区详细信息。
- 另外，您可以对时区选择任何全球标准时间值。

另存为缺省值。 所指定的选项仅适用于当前流。单击此按钮以将这些选项设置为所有流的缺省值。

设置流的数字格式选项

这些选项指定当前流中的各种数字表达式要使用的格式。

数字显示格式。 可以在“标准”(#####.###)、 “科学表示法”(#.###E+##) 和“货币”(\$###.##) 显示格式中选择。

小数位 (标准、科学表示法、货币)。 针对数字显示格式，指定显示或打印实数时要使用的小数位数。此选项将分别为每种显示格式指定一个值。

计算方式。 选择**弧度**或**角度**作为三角法 CLEM 表达式中要使用的度量单位。有关更多信息，请参阅主题 [第 144 页的『三角函数』](#)。

另存为缺省值。 指定的选项仅适用于当前流。单击此按钮以将这些选项设置为所有流的缺省值。

设置流的优化选项

您可以使用“优化”设置来优化流性能。请注意，IBM SPSS Modeler Server (如果使用的话) 上的性能和优化设置会覆盖客户端上的任何相应设置。如果在服务器中禁用了这些设置，那么客户机无法将其启用。但是，如果在服务器中启用了这些设置，那么客户机可以选择将其禁用。

注: 数据库建模和 SQL 优化需要在 IBM SPSS Modeler 计算机上启用 IBM SPSS Modeler Server 连接。通过启用此设置，您可以访问数据库算法，直接从 IBM SPSS Modeler 回送 SQL 以及访问 IBM SPSS Modeler Server。要验证当前许可证的状态，请从 IBM SPSS Modeler 菜单中选择以下项目。

帮助 > 关于 > 其他详细信息

如果启用了连接，您可以在“许可证状态”选项卡中看到选项**服务器启用**。

有关更多信息，请参阅第 7 页的『[连接到 IBM SPSS Modeler Server](#)』。

注: 是否支持 SQL 回送和优化，取决于使用的数据库类型。有关使用 IBM SPSS Modeler 支持和测试的数据库和 ODBC 驱动程序的最新信息，请参阅公司支持站点 <http://www.ibm.com/support>。

启用流重写。 选择此选项可启用 IBM SPSS Modeler 中的流重写。提供了四种重写类型，您可以选择其中的一种或多种类型。流重写会对流中的节点在后台进行重新排序，使其在不更改流语义的情况下更高效的操作。

- **优化 SQL 生成。** 利用此选项可对流中的节点进行重新排序，以便可以使用 SQL 生成回送更多的操作，使其在数据库中执行。在发现某个节点无法以 SQL 表达时，优化器将会向前检查，确定是否有任何下游节点能够以 SQL 表达并可以在不影响流语义的情况下安全地移到问题节点前面。数据库执行操作不但可以比 IBM SPSS Modeler 效率更高，而且这样的回送操作可以减小返回到 IBM SPSS Modeler 进行处理的数据集的大小。而这又可以进一步减小网络通信量，加快流操作的速度。请注意，必须选中**生成 SQL**复选框后 SQL 优化才能生效。
- **优化 CLEM 表达式。** 此选项启用优化器来搜索能够在流运行之前被预处理的 CLEM 表达式，从而提高处理速度。举个简单的示例，如果您有一个像 $\log(\text{salary})$ 这样的表达式，优化器将计算实际工资值并通过传递来对其进行处理。这可以用来提高 SQL 后推和 IBM SPSS Modeler Server 性能。
- **优化语法执行。** 这种流重写方法会提高合并多个包含 IBM SPSS Statistics 语法节点的操作的效率。通过将语法命令合并到单个操作中，而不是作为单独的操作分别运行，从而实现优化。

- **优化其他执行。** 这种流重写方法会提高无法委派给数据库的操作的效率。优化是通过尽可能早地减少流中的数据量来获得的。在保持数据完整性的同时，该流会进行重写，以使操作距离数据源更接近，因而减少了具有较高消耗的操作（如连接）的下游数据。

启用并行处理。 当在一台带有多处理器的计算机上运行时，此选项使得系统能够平衡这些处理器之间的负载，从而可能产生更佳的性能。使用多个节点或者使用下列单个节点可能会由于并行处理而受益：C5.0、合并（按键进行）、排序、分级（秩法和分位法）以及汇总（使用一个或多个键字段）。

生成 SQL。 选择此选项可启用 SQL 生成，从而可以通过使用 SQL 代码使得流操作回送到数据库以生成执行处理，因此可能能够改善性能。为了进一步改善性能，还可以选择**优化 SQL 生成**，从而实现将回送到数据库的操作数最大化。节点的操作回送到数据库之后，该节点将在流运行时以紫色突出显示。

- **数据库高速缓存。** 对于生成了要在数据库中执行的 SQL 的流，可以在中游将数据缓存到数据库中的临时表，而不是缓存到文件系统。如果与 SQL 优化相组合，则此操作将使性能得到显著提高。例如，可以对合并多个表以创建数据挖掘视图的流的输出进行缓存并在需要时重新使用。如果已启用数据库缓存，则只需单击任意非终端节点就可以缓存该点上的数据，并且在下次运行流时，可以直接在数据库中自动创建缓存。如此一来，为下游节点生成 SQL 将成为可能，从而可进一步提高性能。此外，必要时可禁用该选项，例如当策略或权限排除正被写入数据库的数据时。如果未启用数据库缓存或 SQL 优化，则会转而将缓存写入文件系统中。请参阅主题第 30 页的『节点的缓存选项』，以获取更多信息。
- **使用轻松转换。** 利用此选项可以将数据从字符串转换为数字，或从数字转换为字符串（如果以适当的格式存储）。例如，如果数据作为字符串保存在数据库中，但实际包含有意义的数字，那么可以转换数据以便在回送时使用。

注：由于 SQL 实现的微小差别，在数据库中运行的流返回的结果可能会与在 IBM SPSS Modeler 中运行时返回的结果有着轻微的不同。也是因为上述原因，这些差别还可能根据数据库供应商的不同而有所区别。

另存为缺省值。 指定的选项仅适用于当前流。单击此按钮以将这些选项设置为所有流的缺省值。

设置流的 SQL 日志和记录状态选项

这些设置包括用于控制流生成的 SQL 语句的显示，以及流处理的记录数量显示的各种选项。

在流执行期间在消息日志中显示 SQL。 指定在运行流时生成的 SQL 是否传递到消息日志。

在流准备期间在消息日志中显示 SQL 生成详细信息。 指定在流预览期间要生成的 SQL 的预览是否传递到消息日志。

显示 SQL。 指定日志中显示的任何 SQL 是否都应该包含本机 SQL 函数或格式为 `{fn FUNC(...)}` 的标准 ODBC 函数（由 SPSS Modeler 生成）。前者依赖于可能没有实现的 ODBC 驱动程序功能。

重新格式化 SQL 以提高可读性。 指定是否应将在日志中显示的 SQL 格式化以提高可读性。

显示记录的状态。 指定在记录到达终端节点后，报告记录的时间。指定一个数字，用于表示每 N 个记录更新一次状态。

另存为缺省值。 指定的选项仅适用于当前流。单击此按钮以将这些选项设置为所有流的缺省值。

设置流的布局选项

这些设置提供了有关流工作区显示和使用的众多选项。

最小流工作区宽度。 指定流工作区的最小宽度（以像素为单位）。

最小流工作区高度。 指定流工作区的最小高度（以像素为单位）。

流滚动速率。 指定流工作区的滚动速率，以控制在工作区中拖动节点时流工作区窗格的滚动速度。数值越大，滚动速率越高。

图标名称最大值。 指定流工作区中节点名称的字符限制。

图标大小。 选择适当选项，以便将整个流视图放大到标准图标尺寸的 8% 至 200% 之间的某个尺寸。

网格单元大小。 从列表中选择网格单元尺寸。此数值用于使用不可见网格对流工作区中的节点进行对齐。缺省网格单元尺寸为 0.25。

对齐网格。 选择此选项将按不可见网格模式（缺省选项）对齐图标。

生成图标的摆放位置。 选择由模型块生成的节点的图标在工作区中的位置。缺省为左上方。

另存为缺省值。 指定的选项仅适用于当前流。单击此按钮以将这些选项设置为所有流的缺省值。

Analytic Server 流属性

这些设置提供了用于与 Analytic Server 配合使用的众多选项。

Analytic Server 之外要处理的最大记录数

指定要从 Analytic Server 数据源导入 SPSS Modeler 服务器的最大记录数。

节点无法在 Analytic Server 中进行处理时的通知

此设置确定当要提交至 Analytic Server 的流包含无法在 Analytic Server 中进行处理的节点时将发生的情况。指定要发出警告并继续处理流，还是抛出错误并停止处理。

拆分模型存储设置

模型大小 (MB) 超过时按照 Analytic Server 上的引用来存储拆分模型

模型块通常存储为流的一部分。带有多个拆分的拆分模型可能产生大块，在流和 Analytic Server 之间来回移动块可能影响性能。作为一种解决方案，当拆分模型超出指定大小后，会将其存储在 Analytic Server 上，而 SPSS Modeler 中的块包含对该模型的引用。

用于在执行完成后按照 Analytic Server 上的引用来存储模型的缺省文件夹

指定要在 Analytic Server 上存储拆分模型的缺省路径。该路径应以有效的 Analytic Server 项目名称作为开始。

用于存储升级模型的文件夹

指定要存储“升级”模型的缺省路径。SPSS Modeler 会话结束后，不会清除升级模型。

设置流的地理空间选项

任何地理空间字段都具有相关联的坐标系，而无论它是形状、坐标还是单个轴值（例如 X 或 Y 或者纬度和经度）。此坐标系用于设置属性，例如原点 (0,0) 以及与值相关联的单位。

存在众多的坐标系，它们分为两类：地理坐标系和投影坐标系。SPSS Modeler 中的所有空间函数都只能与投影坐标系配合使用。

坐标系的性质决定了合并或追加来自两个不同地理空间数据源的数据时，要求这两个数据源使用同一个坐标系。因此，您必须对流中使用的任何地理空间数据指定坐标设置。

在下列情况下，将自动地重新投影数据以使用选择的流坐标系：

- 对于空间函数（例如 area、closeto 和 within），传递到函数的参数将自动进行重新投影；但是，原始行数据保持不变。
- 在空间-时间预测 (STP) 中使用构建或评分（块）节点时，将对位置字段自动进行重新投影。进行评分时，来自于块的位置是原始位置。
- 使用“地图可视化”节点时。

流坐标系。 只有在选中复选框后才可用。请单击**更改**以显示一系列可用的投影坐标系，然后选择要用于当前流的投影坐标系。

另存为缺省值。 您选择的坐标系将仅应用于当前流。要选择将此坐标系用作所有流的缺省坐标系，请单击此按钮。

选择地理空间坐标系

SPSS Modeler 中的所有空间函数都只能与投影坐标系配合使用。

“**选择流坐标系**”对话框中有一个坐标系列表，其中包含您可以为流中使用的任何地理空间数据选择的所有投影坐标系。

对于每个坐标系，列出了下列信息。

- **WKID** 每个坐标系的唯一知名标识。
- **名称** 坐标系的名称。
- **单位** 与坐标系相关联的测量单位。

除了包含所有坐标系的列表以外，此对话框还提供了**过滤**控件。如果您知道所需坐标系的完整名称或部分名称，请将其输入到对话框底部的**名称**字段中。可供您选择的一系列坐标系将自动进行过滤，以便仅显示名称包含所输入文本的坐标系。

查看流操作消息

使用“流属性”对话框中的“消息”选项卡，可以轻松查看有关运行、优化和模型构建和评估所用时间等流操作的消息。该表还会报告错误消息。

查看流消息

1. 在“文件”菜单中，单击**流属性**（或从管理器窗格中的“流”选项卡上选择流，右键单击并在弹出菜单中单击**流属性**）。
2. 单击 **消息** 选项卡。

或者，在“工具”菜单中，单击：

流属性 > 消息

除有关流操作的消息之外，此处还报告错误消息。当流运行因错误而终止时，此对话框将在“消息”选项卡中打开，并显示错误消息。此外，出错节点将在流工作区中突出显示为红色。

如果在“用户选项”对话框中启用了 SQL 优化和日志选项，则还会显示有关生成的 SQL 的信息。有关更多信息，请参阅主题 [第 34 页的『设置流的优化选项』](#)。

可以通过从保存按钮下拉列表（位于左侧“消息”选项卡的下方）中单击**保存消息**为流保存此处报告的消息。

可以通过从下拉列表中选择**清除所有消息**清除给定流的所有消息。

请注意，CPU 时间是服务器进程使用 CPU 的时间长度。耗用时间是执行开始与执行结束之间的总时间，因此还包括传输文件和呈示输出等操作所耗用的时间。当流使用了多个 CPU（并行执行）时，CPU 时间可能会大于耗用时间。如果流完全推回到用作数据源的数据库中进行处理，那么 CPU 时间将为零。

查看节点执行时间

在“消息”选项卡上可以选择显示“执行时间”，您可以在其中查看流中运行在 IBM SPSS Modeler Server 上的节点的相应执行时间。注意运行在其他区域中的流的时间可能并不准确，例如 R 或 Analytic Server。此外，无法计算某些节点的执行时间。

注：要使此功能有效，必须在**选项**选项卡的**常规**设置中选中**显示执行时间**选项。

在节点执行时间表中，各列如下所示。单击列标题可以按升序或降序对各条目进行排序（例如，以便查看哪些节点的执行时间最长）。

终端节点。 节点所属的分支的标识。此标识是分支末位处的终端节点的名称。

节点标签。 执行时间的相关节点的名称。

节点标识。 执行时间的相关节点的唯一标识。此标识在节点创建时由系统生成。

执行时间（秒）。 执行此节点所耗用的时间。请注意，执行时间可能经常不同于您在常规消息中看到的时间，因为在运行流时准备数据以及在输出中检索数据都需要时间，而这种类型的时间无法计算。

设置流和会话参数

可以定义用于 CLEM 表达式和脚本编写的参数。这些参数在效果上当同于用户定义的变量，这些变量可以在当前流、会话或超节点中保存和保留，并且可以从用户界面和脚本中访问。例如，如果保存流，则为此流设置的所有参数也将保存。（此特点使这些参数区别于仅在脚本中使用和声明的本地脚本变量。）通常将参数用于脚本编制来控制脚本的行为，可通过提供在脚本中无需进行硬编码的字段和值的相关信息。

参数的范围取决于设置参数的位置：

- 流参数可以在流脚本中或在流属性对话框中设置，且可用于流中的所有节点。这些参数显示在表达式构建器中的“参数”列表中。

- 会话参数可以在独立脚本中或在会话参数对话框中设置。这些会话参数可用于当前会话中使用的所有流（即在管理器窗格的“流”选项卡中列出的所有流）。

也可以为超节点设置参数，在这种情况下，这些参数仅对于该超节点中封装的节点可见。

通过用户界面设置流和会话参数

1. 要设置流参数，请在主菜单中单击：

工具 > 流属性 > 参数

2. 要设置会话参数，请在“工具”菜单中单击**设置会话参数**。

提示？。如果您希望在运行时提示用户为该参数输入值，请选中此复选框。

名称。 参数名在这里列出。可以通过在本字段中输入名称来创建新的参数。例如，要为最小温度创建参数，可以键入 `minvalue`。请勿包含表示 CLEM 表达式中的参数的 \$P- 前缀。该名称也用于在 CLEM 表达式构建器中显示。

长名称。 列出每个所创建参数的描述性名称。

存储器。 在列表中选择存储类型。存储类型表明数据值在参数中如何存储。例如，当所使用的值包含希望保留的先导 0 时（例如 008），应选择 **字符串** 作为存储类型。否则，先导 0 将从值中剥离。有效的存储类型为字符串、整数、实数、时间、日期及时间戳。对于日期参数，注意其值必须用下一段落所示的 ISO 标准符号指定。

值。 列出每个参数的当前值。根据需要调整参数。请注意，对于日期参数，必须以 ISO 标准表示法（即，YYYY-MM-DD）来指定值。不接受以其他格式指定的日期。

类型（可选）。 如果计划将该流部署到外部应用程序，则从列表中选择测量级别。否则，建议保留 **类型列** 的值。如果您想为参数指定值约束（如数值范围的上限或下限），请从列表中选择 **指定**。

注意只能通过用户界面为参数设置长名称、存储类型和类型选项。不能用脚本设置这些选项。

单击位于右侧的箭头可使选中的参数在可用参数列表中上下移动。使用删除按钮（标记为 X）可删除选中的参数。

指定参数值的运行时提示

如果流位于需要在不同情况下为相同参数输入不同值的位置，您可以为一个或多个流或会话参数值指定运行时提示。

参数。（可选）为参数输入值，或保留缺省值（如有）。

关闭这些提示。 如果您不希望这些提示在运行流时显示，请选择此复选框。在定义了参数的流属性或会话属性对话框上，您可以通过选择 **提示？** 复选框，使其重新显示。有关更多信息，请参阅主题 [第 37 页的『设置流和会话参数』](#)。

指定参数类型的值约束

在将流部署到读取数据建模流的外部应用程序期间，您可以启用参数的值约束。通过此对话框，可以指定运行流的外部用户可用的值。对话框中的值约束条件会呈动态变化，具体取决于数据类型。此处显示的选项即类型节点的可用选项值。

类型。 显示当前选定的测量级别。您可以根据想要在 IBM SPSS Modeler 中使用参数的方式更改此值。

存储器。 显示已知的存储类型。存储类型不受您选择在 IBM SPSS Modeler 中采用的测量级别（连续、名义或标志）影响。可以在“参数”主选项卡中更改存储类型。

根据在 **类型** 字段中选定的测量级别，对话框的下半部分会呈动态变化。

连续测量级别

下限。 指定参数值的下限。

上限。 指定参数值的上限。

标签。 可以为范围字段的任意值指定标签。单击 **标签** 按钮可打开一个新的对话框，用于指定值标签。

名义测量级别

值。 使用此选项可指定将用作名义字段的参数值。值不会在 IBM SPSS Modeler 流中强制使用，但会在外部部署应用程序的下拉列表中使用。使用箭头键和删除键可以修改现有值，以及对值进行重新排序或删除值。

标志测量级别

True。 指定满足条件时参数的标志值。

False。 指定满足条件时参数的标志值。

标签。 可以为标志字段的值指定标签。

流部署选项

通过流属性对话框的“部署”选项卡，您可以指定用于在 IBM SPSS Collaboration and Deployment Services 内部署流以进行模型刷新或自动化作业调度的选项。所有流都需要指定评分分支才能进行部署。有关更多信息，请参阅第 163 页的『[存储和部署存储库对象](#)』。

流的循环执行

使用“流属性”对话框中的“执行”选项卡可以在当前流中设置循环条件来自动执行重复性任务。

设置完这些条件后，可以将其用于介绍脚本编制，因为其使用流的基本脚本编制来填充脚本窗口。随后，您可以对该基本脚本编制进行修改 - 可能用作构建更好脚本的基础。有关更多信息，请参阅主题 [第 159 页的『全局函数』](#)。

设置流的循环

1. 在“文件”菜单中，单击**流属性**（或从管理器窗格中的“流”选项卡上选择流，右键单击并在弹出菜单中单击**流属性**）。
2. 单击**执行**选项卡。
3. 选择**循环/条件执行**执行方式。
4. 单击**循环**选项卡。

或者，在“工具”菜单中，单击：

流属性 > 执行

作为另一种替代方法，也可以右键单击节点，然后从上下文菜单中单击：

循环/条件执行 > 编辑循环设置

迭代。 您无法编辑此行数值，但是您可以使用表右侧的按钮来添加、删除、上移或下移迭代。

表头。 设置循环时，这些循环反映了您创建的迭代密钥和任何迭代变量。

查看流的全局值

使用“流属性”对话框中的“全局量”选项卡可以查看当前流的全局值集合。全局值使用设置全局量节点创建，用于确定所选字段的均值、合计或标准差等统计量。

运行“设置全局量”节点后，这些值即可用于执行各种流操作。有关更多信息，请参阅主题 [第 159 页的『全局函数』](#)。

查看流的全局值

1. 在“文件”菜单中，单击**流属性**（或从管理器窗格中的“流”选项卡上选择流，右键单击并在弹出菜单中单击**流属性**）。
2. 单击**全局**选项卡。

或者，在“工具”菜单中，单击：

流属性 > 全局

可用全局属性。 此表将列出可用的全局属性。您不能在此更改全局值，但可以使用表右侧的“清除所有值”按钮清除流的所有全局值。

搜索流中的节点

通过指定一定的搜索条件（如节点名称、类别和标识），可以搜索流中的节点。该功能对包含大量节点的复杂流特别有用。

搜索流中的节点

1. 在“文件”菜单中，单击**流属性**（或从管理器窗格中的“流”选项卡上选择流，右键单击并在弹出菜单中单击**流属性**）。
2. 单击 **搜索**选项卡。

或者，在“工具”菜单中，单击：

流属性 > 搜索

可以指定多个选项来限制搜索范围，但如果按节点标识搜索（即使用**标识等于**字段），将排除其他选项。

节点标签包含。选中此复选框，输入整个或部分节点标签以搜索特定节点。搜索不区分大小写，多个单词作为单个文本片段处理。

节点类别。选中此复选框，在列表中单击某个类别以搜索特定类型的节点。**过程节点**表示位于节点选项板的“记录选项”或“字段选项”选项卡上的节点；**应用模型节点**是指一个模型块。

关键字包含。选中此复选框，输入一个或多个完整关键字以搜索节点，这些节点在节点对话框的“注解”选项卡上的“关键字”字段中存在该文本。您输入的关键字文本必须是完全匹配。使用分号分隔多个关键字，以搜索选项（例如，输入 `proton;neutron` 将查找带有这些关键字的所有节点。有关更多信息，请参阅主题第 46 页的『注解』。

注解包含。选中此复选框，输入一个或多个单词，以搜索那些在其节点对话框“注解”选项卡上的主文本区域中包含此文本的节点。搜索不区分大小写，多个单词作为单个文本片段处理。有关更多信息，请参阅主题第 46 页的『注解』。

生成调用的字段。选中此复选框，输入某个生成字段的名称，例如 `$C-Drug`。可以使用此选项来搜索生成某个特定字段的建模节点。只输入一个必须是完全匹配的字段名称。

标识等于。选中此复选框，输入节点标识以搜索具有此标识的特定节点（选择此选项将禁用前面的所有选项）。节点标识是在创建节点时由系统分配的，可用于引用节点以编写脚本或实现自动化。只输入一个必须是完全匹配的节点标识。有关更多信息，请参阅主题第 46 页的『注解』。

在超节点中搜索。此复选框为缺省选中，表示在超节点内外的节点上执行搜索。如果只打算在流的最高级处的超节点外面的节点上执行搜索，请取消选中此框。

查找。在指定了所有的选项后，单击此按钮开始搜索。

与指定选项匹配的节点在对话框下方列出。从列表中选择节点并在流工作区中突出显示此节点。

重命名流

使用“流属性”对话框中的“注解”选项卡，可以为流添加描述性注解并创建自定义名称。在为已添加到工程窗格的流生成报告时，这些选项特别有用。有关更多信息，请参阅主题第 46 页的『注解』。

流描述

对于您创建的每个流，IBM SPSS Modeler 会为其生成包含流内容相关信息的流描述。如果您想知道流的功能但又未安装 IBM SPSS Modeler，例如通过 IBM SPSS Collaboration and Deployment Services 访问流时，这非常有用。

流描述以 HTML 文档格式显示，其中包含多个部分。

常规流信息

该部分包含流名称，以及创建和上次保存流时的详细信息。

描述与注释

这部分包含任何：

- 流注释（请参阅第 46 页的『注解』）
- 未连接到特定节点的注释
- 连接到流的建模和评分分支中的节点的注释

评分信息

该部分包含有关流的评分分支的各种标题下的信息。

- **注释。** 包含仅连接到评分分支中节点的注释。
- **输入。** 列出输入字段及其存储类型（例如，字符串、整数、实数等等）。
- **输出。** 列出输出字段及其存储类型，包括建模节点生成的其他字段。
- **参数。** 列出任何与流的评分分支相关的参数，以及每次模型评分时可以查看或编辑的参数。单击流属性对话框的**部署**选项卡上的**评分参数**按钮，可以标识这些参数。
- **模型节点。** 显示模型名称和类型（例如，神经网络以及 C&R 树等等）。这是在流属性对话框的**部署**选项卡上为**模型节点**选择的模型块。
- **模型详细信息。** 在各种标题下显示标识的模型块详细信息。如有可能，还包括模型的预测变量重要性和评估图表。

建模信息

包含流的建模分支相关信息。

- **注释。** 列出任何连接到建模分支内节点的注释或注解。
- **输入。** 列出输入字段及其在建模分支中的角色（以字段角色值形式，例如，“输入”、“目标”、“分割”等）。
- **参数。** 列出任何与流的建模分支相关的参数，以及每次模型更新时可以查看或编辑的参数。单击流属性对话框的**部署**选项卡上的**模型构建参数**按钮，可以标识这些参数。
- **建模节点。** 显示用于生成或更新模型的建模节点的名称和类型。

预览流描述

您可以在流属性对话框单击一个选项，从而在 Web 浏览器中查看流描述的内容。描述内容取决于您在对话框的“部署”选项卡上所指定的选项。有关更多信息，请参阅主题 [第 173 页的『流部署选项』](#)。

要查看流描述，请执行下列操作：

1. 在主 IBM SPSS Modeler 菜单中，单击：
 工具 > 流属性 > 部署
2. 设置部署类型、指定的评分节点以及任何评分参数。
3. 如果部署类型为“模型刷新”，则可以选择：
 - 建模节点以及任何模型构建参数
 - 在流的评分分支上的模型块
4. 单击**预览流描述**按钮。

导出流描述

可以将流描述内容导出到 HTML 文件。

要导出流描述：

1. 在主菜单中，单击：
 文件 > 导出流描述

2. 输入 HTML 文件名并单击 **保存**。

运行流

指定流的所需选项并连接所需节点后，便可以让数据流经该流中的节点，以运行该流。在 IBM SPSS Modeler 中有多种运行流的方法。您可以：

- 在“工具”菜单中，单击**运行**。
- 单击工具栏上的其中一个**运行... 按钮**。使用这些按钮，您可以运行整个流，也可以仅运行所选的终端节点。有关更多信息，请参阅主题 [第 14 页的『IBM SPSS Modeler 工具栏』](#)。
- 通过右键单击某个终端节点，并在弹出菜单中单击**运行**，运行单个数据流。
- 通过右键单击任意非终端节点并在弹出菜单中单击**从此处运行**来运行部分数据流。这将只执行那些在所选节点之后的操作。

要暂停正在运行的流，可以单击工具栏中的红色**停止按钮**，或者从“工具”菜单中选择**停止执行**。请注意，单击一下“停止”按钮向 Modeler 表明 Modeler Server 停止执行。在某些情况下，执行将立即停止，但在其他情况下，必须先完成当前步骤，然后才能停止整个执行。因此，时间将有所差异。如果双击该按钮，那么将断开服务器连接并建立新连接。在大多数情况下，这将关闭所有服务器进程（并且可能需要一些时间）。但在某些情况下，将不会停止某个服务器进程。

如果任何流的运行时间超过三秒钟，便会显示“执行反馈”对话框以指示进度。

某些节点会进一步显示有关流执行的更多信息。选择对话框中的对应行可查看这些显示。系统将自动选中第一行。

使用模型

如果流包括一个建模节点（即，来自节点选项板的“建模”或“数据库建模”选项卡的节点），则在流运行时将创建**模型块**。模型块是**模型**的容器，即规则、公式或方程式的集合，允许您针对源数据生成预测，该预测是预测分析的核心。



图 12: 模型块

当您成功运行建模节点时，相应的模型块被置于流工作区，由金色钻石形图标表示（因此名为“块”）。您可以打开块并浏览其内容以查看模型的详细信息。要查看预测，可附加并运行一个或多个终端节点，其输出将以可读取的形式显示预测。

典型的建模流由两个分支组成。**建模分支**包含建模节点以及在其之前的源节点和过程节点。**评分分支**在运行建模节点时创建，包含模型块和终端节点，或是用于查看预测的节点。

有关更多信息，请参阅《*IBM SPSS Modeler 建模节点*》指南。

添加注释和注解到节点和流

您可能需要向组织中的其他人描述流。为了帮您进行，您可以将解释注释附加到流、节点和模型块中。

其他人然后可以在屏幕上查看这些注释，或者您可以打印包括注释的流图像。

您可以列出流或超节点的所有注释，更改列表中注释的顺序，编辑注释文本和更改注释的前景色或背景色。有关更多信息，请参阅主题 [第 45 页的『列出流注释』](#)。

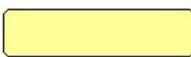
您还可以通过流属性对话框、节点对话框或模型块窗口的“注释”选项卡以文本注释的形式将注释添加到流、节点和块中。只有当“注解”选项卡打开时才能看到这些注释，例外情况是流注解也可显示为屏幕上的注释。有关更多信息，请参阅主题 [第 46 页的『注解』](#)。

注释

注释采用文本框的形式，在此可以输入任何数量的文本，还可添加任何数量的注释。注释可以是独立的（不附加到任何流对象），或者可以连接到流中的一个或多个节点或模型块。独立注释通常用于描述流的整体目的；连接的注释描述附加到的节点或块。节点和块可以附加多个注释，流可以有任意数量的独立注释。

注：虽然无法将流注解附加到节点或模型块，但可将流注解显示为屏幕上的注释。有关更多信息，请参阅主题第 45 页的『将注解转换为注释』。

文本框的外观更改为指示当前注释模式（或显示为注解的注解），如下表所示。

注释文本框	注解文本框	方式	表示	获得方式...
		编辑	注释已打开，可进行编辑。	新建注释或注解，或双击现有注释或注解。
		上次选定的	可以移动或删除注释，也可调整其大小。	在编辑后单击流背景，或单击现有注释或注解。
		视图	编辑完成。	编辑后单击另一节点、注释或注解。

当您创建新独立注释时，最初显示在流工作区的左上角。

如果您将注释附加到节点或块中，注释最初显示在附加流对象的上方。

文本框显示为白色，以显示可以输入文本。当您输入文本时，单击文本框的外部。注释背景变为黄色，以显示文本输入完成。一直选中注释，您可以对其进行移动、删除或调整大小。

再次单击时，边框变为实线以表示编辑完成。

双击注释将文本框更改为编辑模式 -- 背景变为白色，注释文本可以编辑。

您还可以将注释附加到超节点。

操作涉及注释

您可以对注释执行大量操作。您可以：

- 添加独立注释
- 将注释附加到节点或块
- 编辑注释
- 调整注释大小
- 移动注释
- 断开注释连接
- 删除注释
- 显示或隐藏流的所有注释

添加独立注释

1. 确保流上未选择任何对象。
2. 请执行下列其中一项操作：
 - 在主菜单中，单击：
插入 > 新建注释
 - 右键单击流背景并从弹出菜单中单击**新注释**。
 - 单击工具栏上的**新注释**按钮。

3. 输入注释文本（或从剪贴板粘贴文本）。
4. 在流中单击节点以保存注释。

将注释附加到节点或块

1. 在流工作区上选择一个或多个节点或块。
2. 请执行下列其中一项操作：
 - 在主菜单中，单击：
插入 > 新建注释
 - 右键单击流背景并从弹出菜单中单击**新注释**。
 - 单击工具栏上的**新注释**按钮。

3. 输入注释文本。
4. 在流中单击另一节点以保存注释。

或者，您可以：

5. 插入独立注释（请参阅上一节）。
6. 请执行下列其中一项操作：
 - 选择注释，按 F2，然后选择节点或块。
 - 选择节点或块，按 F2，然后选择注释。
 - （仅限三键鼠标）在注释上移动鼠标指针，按下中间按钮，将鼠标指针拖至节点或块上，释放鼠标按钮。

将注释附加到其他节点或块

如果注释已附加到节点或块，或者如果当前在流级别，您想将其附加到其他节点或块，进行以下其中一个操作：

- 选择注释，按 F2，然后选择节点或块。
- 选择节点或块，按 F2，然后选择注释。
- （仅限三键鼠标）在注释上移动鼠标指针，按下中间按钮，将鼠标指针拖至节点或块上，释放鼠标按钮。

编辑现有注释

1. 请执行下列其中一项操作：
 - 双击注释文本框。
 - 选中文本框，然后按下 Enter 键。
 - 右键单击文本框显示其菜单，然后单击“编辑”。
2. 编辑注释文本。您可以在编辑时使用标准 Windows 快捷键，例如使用 Ctrl+C 复制文本。在弹出菜单中列出编辑期间的其他注释选项。
3. 单击一次文本框外部，以显示调整大小控件，然后再次完成注释。

调整注释文本框大小

1. 选择注释，以显示调整大小控件。
2. 单击并拖动控件以调整文本框大小。
3. 单击文本框外部，保存更改。

移动现有注释

如果您想移动注释，而非其附加的对象（如有），进行以下其中一个操作：

- 在注释上移动鼠标指针，按下左侧鼠标按钮，将注释拖至新位置。
- 选择注释，按下 Alt 键，使用箭头键移动注释。

如果您想移动注释连同任何附加注释的节点或块：

1. 选择您想移动的所有对象。

2. 请执行下列其中一项操作：

- 在一个对象上移动鼠标指针，按下左侧鼠标按钮，将对象拖至新位置。
- 选择一个对象，按下 **Alt** 键，使用箭头键移动对象。

将注释从节点或块断开连接

1. 选择要断开连接的一个或多个注释。

2. 请执行下列其中一项操作：

- 按 **F3** 键。
- 右键单击选定注释并从其菜单中单击“断开”。

删除注释

1. 选择要删除的一个或多个注释。

2. 请执行下列其中一项操作：

- 按 **Delete** 键。
- 右键单击选定注释并从其菜单中单击“删除”。

如果注释附加到节点或块，连接线也被删除。

如果注释原来为流或超节点注解，且已转换成独立注释，注释将从工作区中删除，但其文本保留在流或超节点的“注解”选项卡中。

显示或隐藏流的注释

1. 请执行下列其中一项操作：

- 在主菜单中，单击：
查看 > 注释
- 单击工具栏中的**显示/隐藏注释**按钮。

列出流注释

您可以查看已为某个特定流或超节点添加的所有注释的列表。

在此列表上，您可以

- 更改注释的顺序
- 编辑注释文本
- 更改注释的前景色或背景色

列出注释

要列出已为流添加的注释，请执行以下一种操作：

- 在主菜单中，单击：
工具 > 流属性 > 注释
- 在管理器窗格中右键单击流，并单击**流属性**，然后单击**注释**。
- 右键单击工作区上的流背景，并单击**流属性**，然后单击**注释**。

文本。 注释文本。双击文本将字段更改为可编辑文本框。

链接。 要将注释附加到的节点的名称。如果该字段为空，则注释应用到流。

定位按钮。 这些按钮可在列表中上下移动选中的注释。

注释颜色。 要更改注释的前景色或背景色，选择注释，选择**自定义颜色**复选框，然后从**背景**或**前景**列表（或两个列表）中选择一种颜色。单击**应用**，然后单击流背景查看更改效果。单击**确定**保存更改。

将注解转换为注释

可将对流或超节点的注解转换为注释。

如果是流，则注解转换为流工作区上的独立注释（即不附加到任何节点）。

超节点注解转换为注释时，注释不附加到流工作区上的超节点，但在放大到超节点时可见。

转换流注解为注释

1. 在“工具”菜单中，单击**流属性**。（也可以右键单击管理器窗格中的流，然后单击**流属性**。）
2. 单击 **注解**选项卡。
3. 选择**注解显示为注释**复选框。
4. 单击**确定**。

转换超节点注解为注释

1. 双击工作区上的超节点图标。
2. 单击 **注解**选项卡。
3. 选择**注解显示为注释**复选框。
4. 单击**确定**。

注解

可以采用多种方式节点、流和模型添加注解。您可以添加描述性注解并指定自定义名称。在为已添加到工程窗格的流生成报告时，这些选项特别有用。对于节点和模型块，还可以添加工具提示文本，以帮助区分流工作区中的相似节点。

添加注解

编辑节点或模型块时将打开一个带有选项卡的对话框，其中包含用于设置各种注解选项的“注解”选项卡。您也可以直接打开“注解”选项卡。

1. 要为节点或模型块添加注解，请右键单击流工作区中的节点或模型块，然后单击**重命名并加注解**。此时将打开“编辑”对话框并显示“注解”选项卡。
2. 要为流添加注解，请在“工具”菜单中单击**流属性**。（也可以右键单击管理器窗格中的流，然后单击**流属性**。）单击“注解”选项卡。

名称。 选择**自定义**可调整自动生成的名称，或为流工作区中显示的节点创建唯一名称。

工具提示文本。（仅适用于节点和模型块）输入用作流工作区上工具提示的文本。使用大量相似节点时，此功能特别有用。

关键字。 指定要在工程报告中以及在搜索流中节点时或跟踪存储在存储库中的对象时使用的关键字（参阅第 163 页的『关于 IBM SPSS Collaboration and Deployment Services Repository』）。可使用分号分隔多个关键字 - 例如 `income; crop type; claim value`。将去除每个关键字开始和结尾处的空格 - 例如，`income ; crop type` 与 `income;crop type` 得到的结果相同。（但是，关键字中的空格不会去除。例如，包含一个空格的 `crop type` 与包含两个空格的 `crop type` 并不相同。）

主文本区域可用于输入有关节点操作或节点中所做决策的详细注解。例如，在共享和重用流时，对某些决策（如使用过滤节点丢弃包含许多空白值的字段）进行注解将很有用。为节点添加注解会将此类信息与节点一起存储。您也可以选择将这些注解包含在从工程窗格创建的工程报告中。有关更多信息，请参阅主题第 183 页的『工程简介』。

将注解显示为注释。（仅适用于流和超节点注解）选中此框将注解转换为在流工作区上可见的独立注释。有关更多信息，请参阅主题第 42 页的『添加注释和注解到节点和流』。

标识。 显示唯一的标识，该标识可用于引用节点以编写脚本或实现自动化。此值将在创建节点时自动生成，且不会更改。另请注意，为避免与字母“O”的混淆，节点标识中不使用数字零。使用右侧的复制按钮可以根据需要将标识复制并粘贴到脚本中或其他位置。

保存数据流

创建流之后，可以进行保存以备将来重复使用。

保存流

1. 在“文件”菜单中，单击**保存流或将流另存为**。
2. 在“保存”对话框中，浏览至要在其中保存流文件的文件夹。
3. 在“文件名”文本框中输入流的名称。
4. 如果要将保存的流添加到当前工程，请选中 **添加到工程**。

单击**保存**会以扩展名 ***.str**，将流存储在指定目录中。

自动备份文件。 每次保存流时，都会对以前保存的文件版本自动进行备份，并会在文件名末尾附加连字符（例如，**mystream.str-**）。要恢复已备份的版本，只需删除连字符并重新打开文件即可。

保存状态

除流以外，您还可以保存**状态**，其中包括当前显示的流图表和所有已创建模型块（在管理器窗格的“模型”选项卡中列出）。

保存状态

1. 在“文件”菜单中，单击：
状态 > 保存状态或将状态另存为
2. 在“保存”对话框中，浏览至要在其中保存状态文件的文件夹。

单击 **保存**会在指定目录中以扩展名 ***.cst** 存储状态。

保存节点

您也可以保存单个节点，方法是右键单击流工作区中的节点，然后在弹出菜单中单击**保存节点**。使用文件扩展名 ***.nod**。

保存多个流对象

如果在多个对象（如流、工程或模型块）未保存的情况下退出 IBM SPSS Modeler，系统将提示您在完全关闭软件之前进行保存。如果选择保存项目，则会打开一个对话框，显示保存每个对象的选项。

1. 只需选中要保存对象的复选框即可。
2. 单击**确定**在所需的位置保存每个对象。

随后系统将为您显示每个对象的标准“保存”对话框。在保存完成后，应用程序将会关闭。

保存输出

通过 IBM SPSS Modeler 输出节点生成的表、图形和报告可以采用输出对象 (***.cou**) 格式进行保存。

1. 查看要保存的输出时，请在输出窗口菜单中单击：
文件 > 保存
2. 指定输出文件的名称和位置。
3. 选择“保存”对话框中的**将文件添加到工程**可以在当前工程中包括该文件（可选操作）。有关更多信息，请参阅主题 [第 183 页的『工程简介』](#)。

或者，也可以右键单击管理器窗格中列出的任意输出对象，然后在弹出菜单中选择**保存**。

加密和解密信息

在保存流、节点、工程、输出文件或模型块时，您可以对其进行加密以防止有人在未经授权的情况下使用它。为此，应在保存时选择一项附加选项，并为要保存的项目添加密码。可以为您保存的任何项目设置密码，从而进一步增加其安全性；此加密法与在 IBM SPSS Modeler 与 IBM SPSS Modeler Server 之间传递文件时使用的 SSL 加密不同。

当您尝试打开加密项目时，系统会提示您输入密码。输入正确的密码后，项目将自动解密并正常打开。

加密项目

1. 在“保存”对话框中，针对要加密的项目单击**选项**。此时将打开“加密选项”对话框。
2. 选中 **加密此文件**。
3. 为提高安全性，可选中 **屏蔽密码**（可选操作）。此操作会将您输入的所有内容显示为一连串的点。
4. 输入密码。警告：如果忘记密码，则无法打开文件或模型。
5. 如果已选中**屏蔽密码**，请再次输入密码以确保输入无误。
6. 单击**确定**返回“保存”对话框。

注：如果保存任何加密保护项目的副本，除非您更改“加密选项”对话框中的设置，否则新项目将自动采用原有密码以加密格式保存。

加载文件

可以在 IBM SPSS Modeler 中重新加载众多已保存的对象：

- 流 (.str)
- 状态 (.cst)
- 模型 (.gm)
- 模型选用板 (.gen)
- 节点 (.nod)
- 输出 (.cou)
- 工程 (.cpj)

打开新文件

可以直接通过“文件”菜单加载流。

- 在“文件”菜单中，单击**打开流**。

使用“文件”菜单中提供的子菜单项目可以打开所有其他类型的文件。例如，要加载模型，可以在“文件”菜单中单击：

模型 > 打开模型或装入模型选用板

打开最近使用的文件

为快速加载最近使用的文件，您可使用“文件”菜单底部的选项。

选择**最近流**、**最近工程**或**最近状态**可展开最近使用的文件列表。

映射数据流

使用映射工具，可以将新数据源连接到先前存在的流。映射工具不仅会设置连接，还将帮助您指定新源中的字段如何替换现有流中的字段。您只需连接到现有的流，而不必为新数据源重新创建整个数据流。

使用数据映射工具，可以将两个流段连接在一起，并确保所有（必要）字段的名称均正确匹配。实际上，映射数据将直接导致创建新的过滤节点，该操作会通过重命名与相应字段相匹配。

以下两种方法都可以映射数据：

选择替换节点。此方法从要替换的节点入手。首先，右键单击要替换的节点；然后，使用弹出菜单中的**数据映射 > 选择替换节点**选项，选择要替换它的节点。

映射到。此方法从要引入流中的节点入手。首先，右键单击要引入的节点；然后，使用弹出菜单中的**数据映射 > 映射到**选项，选择该节点应该连接到的节点。此方法通常适用于映射到终端节点。注：不能映射到合并节点或追加节点，而只需采用常规方式将流连接到合并节点即可。

数据映射已紧密集成到流的构建过程中。如果尝试连接到已具有连接的节点，您可以选择替换连接或映射到该节点。

将数据映射到模板

要将某个模板流的数据源替换为将您自己的数据带入 IBM SPSS Modeler 的新源节点，应使用“数据映射”弹出菜单中的**选择替换节点**选项。此选项可用于所有节点，但合并节点、“汇总”节点和所有终端节点除外。使用数据映射工具执行此操作有助于确保现有流操作与新数据源之间的字段正确匹配。下列步骤简单介绍数据映射过程。

步骤 1：指定原始源节点中的基本字段。 为了正确运行流操作，需要指定必要字段。有关更多信息，请参阅主题 [第 49 页的『指定必要字段』](#)。

步骤 2：将新数据源添加到流工作区。 使用某个源节点带入新的替换数据。

步骤 3：替换模板源节点。 使用模板源节点的弹出菜单中的“数据映射”选项，单击**选择替换节点**，然后选择替换数据的源节点。

步骤 4：检查映射的字段。 在打开的对话框中，检查软件是否正从替换数据源中的字段正确映射到流中的字段。所有未映射的必要字段都显示为红色。流操作中使用的字段必须替换为新数据源中的类似字段，以正常进行下游操作。有关更多信息，请参阅主题 [第 49 页的『检查映射的字段』](#)。

使用对话框确保所有必要字段都正确映射后，原有数据源将断开连接，新数据源将使用称为映射的“过滤”节点连接到流。该过滤节点可在流中实现真正的字段映射。流工作区中还包括非映射过滤节点。非映射过滤节点可用于取消字段名映射，方法是将其添加到流中。该节点将撤消已映射的字段，但请注意，您必须编辑所有下游终端节点才能重新选择字段和重叠字段。

流之间的映射

与连接节点类似，这种数据映射方法不要求您预先设置必要字段。采用这种方法，只需使用“数据映射”弹出菜单中的**映射到**从一个流连接到另一个流即可。对于映射到终端节点和在流之间复制和粘贴的操作，这种类型的数据映射十分有用。注：使用**映射到**选项时，不能映射到合并节点、追加节点和所有类型的源节点。

在流之间映射数据

1. 右键单击要用于连接到新流的节点。
2. 在菜单中，单击：
数据映射 > 映射到
3. 使用光标选择目标流中的目标节点。
4. 在打开的对话框中确保字段完全匹配，然后单击**确定**。

指定必要字段

通常，映射到现有流时，流创建者会指定必要字段。这些必要字段可指明下游操作时是否会使用某个特定字段。例如，现有流可能会构建一个使用称为产出的字段的模型。在该流中，产出是必要字段，因为不能在没有该字段的情况下构建模型。同样，用于操控类节点（如导出节点）的字段对于导出新字段也是必要的。将此类字段明确设置为必要字段有助于确保将新源节点中的相应字段映射到这些字段。如果未映射某些必备字段，您将收到错误消息。如果确定不必要使用某些操作或输出节点，则可以将这些节点从流中删除，并将相应字段从“必要字段”列表中删除。

设置必要字段

1. 右键单击将替换的模板流的源节点。
2. 在菜单中，单击：
数据映射 > 指定基本字段
3. 使用“字段选择器”，可以在列表中添加或删除字段。要打开“字段选择器”，请单击字段列表右侧的图标。

检查映射的字段

选定要将一个数据流或数据源映射到另一个数据流或数据源的位置后，将显示一个对话框，您可以在其中选择映射字段或确保系统缺省映射正确。如果已为流或数据源设置了必要字段，但它们不匹配，这些字段将显示为红色。数据源中的所有未映射字段将以原样通过过滤节点，但请注意，您也可以映射非必要字段。

原始。 列出模板或现有流中的所有字段，即显示在较下游位置的所有字段。新数据源中的字段将映射到这些字段。

已映射。 列出已选择用于映射到模板字段的字段。这些字段的名称可能必须进行更改，以与流操作中使用的原始字段匹配。单击某个字段的表单元格可激活包含可用字段的列表。

如果不确定要映射哪些字段，在映射之前仔细检查源数据会很有用。例如，可以使用源节点中的“类型”选项卡检查源数据的摘要。

提示和快捷键

通过熟悉以下快捷键和提示，您可以快速而轻松地操作：

- **通过双击来快速构建流。** 只需双击选项板中的某个节点，即可添加该节点并将其连接到当前流。
- **使用组合键来选择下游节点。** 按 Ctrl+Q 和 Ctrl+W 可切换所有下游节点的选择。
- **使用快捷键来连接和断开节点。** 当在工作区中选中某个节点时，按 F2 可开始连接，按 Tab 可移至所需的节点，按 Shift+空格可完成连接。按 F3 可断开对所选节点的所有输入和输出连接。
- **使用您喜爱的节点来定制“节点选用板”选项卡。** 在“工具”菜单中单击**管理选用板**可打开一个对话框，用于添加、删除或移动“节点选用板”中显示的节点。
- **重命名节点并添加工具提示。** 每个节点对话框都包含一个“注解”选项卡，您可在其中为工作区中的节点指定自定义名称，以及添加工具提示以帮助您管理流。此外，还可以添加详细注解以跟踪进度、保存过程详细信息，以及表示所需或已实现的任何商业决策。
- **将值自动插入到 CLEM 表达式中。** 使用“表达式构建器”（可通过用于导出节点和过滤节点等的对话框访问此工具），可以在 CLEM 表达式中自动插入字段值。可以单击“表达式构建器”中的值按钮在现有字段值中选择。



图 13: “值”按钮

- **快速浏览文件。** 在“打开”对话框中浏览文件时，使用“文件”列表（单击对话框顶部“查找”字段旁的黄色菱形按钮）可访问先前使用的目录，以及 IBM SPSS Modeler 缺省目录。使用向前和向后按钮可滚动浏览访问的目录。
- **最大限度减少输出窗口混乱。** 使用所有输出窗口右上角的红色 X 按钮可快速关闭并删除输出。这样，您可以在管理器窗格的“输出”选项卡中仅保留所期望或感兴趣的结果。

全部键盘快捷方式都适用于此软件。有关更多信息，请参阅主题 [第 206 页的『键盘辅助功能选项』](#)。

您是否知道可以...

- 使用鼠标拖动并选择流工作区中的一组节点。
- 将节点从一个流复制并粘贴到另一个流。
- 通过每个对话框和输出窗口访问“帮助”。
- 获得有关 CRISP-DM（跨行业数据挖掘标准流程）的帮助。（在“帮助”菜单中单击 **CRISP-DM 帮助**。）

第 6 章 处理数据

您可以通过右键单击数据节点并选择**查看数据**，来检查流数据。在打开的窗口中：

- **图表**选项卡允许创建高级数据可视化以从不同的角度探索数据，并识别数据中的模式、连接和关系。
- **电子表格**选项卡以表格式显示数据的只读视图。
- **数据审计**选项卡显示每列的频率和统计信息。
- 在**仪表板**选项卡上，可以单击**开始布局设计**，并创建用于在单个页面上查看多个视图的布局。可以将布局保存为模板，并将保存的图表拖放到布局中的位置。
- **首选项**选项卡允许设置用户界面首选项，例如，语言和外观。

缺省情况下，此功能使用端口 28900。如果需要使用其他端口，请更改 `options.cfg` 文件中 `data_view_port_number` 配置设置的值。

注：此**查看数据**功能当前在 SPSS Modeler 的“预订”版本中不可用。

构建图表

图表选项卡提供一些选项，用于从预定义的图表库或从各个部分（例如，轴和条形）构建图表。要构建图表，请从提供的图表类型选项中选择库图表类型或基本元素。

构建图表时，工作区会显示该图表的预览。该预览会使用实际的变量标签，以及实际数据中有代表性的测量级别。

对于新用户，使用库是首选方法。有关使用库的信息，请参阅第 52 页的『[从图表类型库构建图表](#)』。

启动“图表构建器”

- 右键单击想要处理的数据节点，然后选择**查看数据**。

图表目录

图表目录提供了一些选项，用于通过浏览图表类别、使用关键字搜索或按图表特征过滤来查找合适的图表。可以从任何活动数据集选项卡访问图表目录。

从任何活动数据集选项卡，单击工具栏中可用的**图表构建器**控件。或者，通过**可视化 > 图表构建器...**访问目录。有多种方法可用于搜索合适的图表类型。

- **查找图表类型**字段可用于搜索特定的图表和关键字。输入图表名称或关键字时，可用图表磁贴的数量会过滤以仅为指定的图表名称或关键字提供结果。

注：您可以使用**查找图表类型**字段旁边的平铺视图和列表视图图标来控制图表类型在图表目录中的显示方式。

- 图表目录磁贴根据其相应类别进行分组。您可以在可用磁贴中滚动，直到找到所需的图表类型。
- **过滤器**窗格提供了一些选项，用于根据相应类别、用途、因变量和统计信息来过滤图表。

注释：

- 每个**过滤器**窗格标题旁边的数字标识当前为每个部分选择的过滤器数量。
- 单击**重置过滤器**以清除**过滤器**窗格设置。

布局和术语

画布

工作区是“图表构建器”对话框中用于构建图表的区域。

图表类型

列出可用的图表类型。图形元素是图表中代表数据的项。这些项包括条形、点、线条，等等。

“详细信息”窗格

“详细信息”窗格提供基本的图表构建块。

图表设置

提供选项，供您选择用来构建图表的变量、分布方法、标题和子标题字段，等等。根据选择的图表类型不同，在“详细信息”窗格中提供的选项也会有所变化。有关可用于每种图表类型的选项的详细信息，请参阅第 52 页的『图表类型』。

操作

提供选项，供您下载图表配置文件、以图像文件形式下载图表、重置图表，以及设置全局图表首选项。

从图表类型库构建图表

构建图表的最简单方法是使用图表类型库。下列是从库构建图表的一般步骤。

1. 在**图表类型**部分中，选择“图表”类别。在图表工作区上将会显示选定图表类型的预览版本。
2. 如果工作区已显示某个图表，那么新图表会替换该图表的轴系和图形元素。
 - a. 根据选择的图表类型，可用的变量会显示在“详细信息”窗格中的多个不同标题下（例如，用于条形图的**类别**，以及用于折线图的**X 轴**和**Y 轴**）。请为选择的图表类型选择相应的变量。

图表类型

库包含最常用图表的集合。这些图表包括：

三维图

三维图一般用来表示多变量函数，并包括一个 Z 轴变量，该变量同时是 X 轴变量和 Y 轴变量的因变量。

创建简式三维图

1. 在图表构建器的**图表类型**部分中，单击**三维图**图标。



工作区随即更新，以显示三维图模板。

2. 从下拉列表中选择**图表类型**。
3. 从下拉列表中选择**X 轴**变量。
4. 从下拉列表中选择**Y 轴**变量。
5. 从下拉列表中选择**Z 轴**变量。

其他特征

Type

列出可用于呈现数据的图表类型。

X 轴

列出可用于图表 X 轴的数据集变量。

Y 轴

列出可用于图表 Y 轴的数据集变量。

Z 轴

列出可用于图表 Z 轴的数据集变量。

工具提示信息

列出当鼠标指针悬停在数据点上时，可用来生成工具提示信息的变量。

颜色图

列出可用的颜色图变量。这些变量根据所指定列中值的范围，使用颜色级数在图点中表示自身。颜色图也称为等值线图。

大小图

列出可用的大小图变量。这些变量使用不同大小在图点中表示自身。

Z 比率

设置 Z 轴数据值的比例（相对于 X 轴和 Y 轴）。

旋转

启用和禁用图表旋转。

数据点工具提示

控制显示数据点工具提示的位置（数据点右侧、图表右上角或者隐藏）。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

条形图

条形图对汇总分类变量非常有用。例如，可以使用条形图显示参与调查的男女人数，也可以使用条形图显示男性和女性的平均薪水。

创建简式条形图

1. 在图表构建器的**图表类型**部分中，单击**条形图**图标。



工作区随即更新，以显示条形图模板。

2. 选择一个分类（名义或有序）变量作为**类别**变量。可以使用标度变量，但所得的结果只对几个特殊个案有用。条形图最适合显示数量有限的不同值。创建具有标度**类别**轴的条形图时，条形会十分细长，这是因为每个条形都按精确的值绘制，而且条形不能重叠在其他连续值之上。
3. 从**汇总**列表中选择统计。任何统计的结果都能确定条形图的高度。如果您所需的统计未显示在**汇总**列表中，那么表示该统计可能需要变量。请从**值**列表中选择变量，并检查该统计现在是否可用。也许还存在其他图表类型限制。例如，只能为特定统计计算误差条形图。

其他特征

类别

列出可用于图表 X 轴的数据集变量。

排序依据

选择用于变量中的类别的排序选项。

类别名称

使用类别标签对变量的类别进行排序。这些标签显示在图表中，通常作为标度标记标签或图注标签。

类别值

使用数据集中存储的值对变量的类别进行排序。类别的值是在数据集中标识类别的值。它通常与其标签不同，并且不一定是描述性的。例如，该值可能是一个数字（例如 1），而标签是类别的文本描述（例如女性）。

类别顺序

请选择变量类别的排序顺序。

如所读取

变量类别以其出现在数据集中的顺序呈现。

升序(S)

变量类别以升序排序。

降序

变量类别以降序排序。

目录

请为图形元素选择统计汇总函数。统计的结果确定图形元素在 y 轴上的位置。在二维图表中，为 x 轴上的每个值计算统计。在三维图表中，为 x 轴和 z 轴上的值的交点计算统计。

统计汇总函数分为两类。它们之间的区别至关重要，因为这种区别确定您是否需要指定值变量。

- **不需要值变量的函数。** 这些是不需要变量的函数。所有的计数和百分比统计均归入此类别。不存在已定义的值变量时，这些统计可用。
- **确实需要值变量的函数。** 这些是需要值变量的函数。例如，*Mean* 函数需要一个要计算均值的变量。存在已定义的值变量时，这些统计可用。

值

选中需要值变量的汇总函数时，就会显示此字段。请选择要用作值的变量。

分割依据

请选择一个分类变量，用于创建图表的表，分割依据变量的每个类别在其中都有一个单元格。与分组类似，在本质上，分割依据变量通过显示每个变量类别的信息，将更多维度添加到图表中。

分割类型

选择分割依据变量时，您可选择以堆积或簇状形式显示所生成的类别条形。聚类和堆积可在图表内添加维数。聚类可将一个条拆分为多个条，而堆积可在每个条中创建若干段。在为堆积选择适当的统计时应谨慎。当这些值添加到一起（堆积）时，得到的结果必须是有意义的。例如，添加和堆积平均值（平均值）通常意义不大。

条形图类型

从提供的选项中选择条形图类型。

- X 轴
- Y 轴
- 逆 X 轴
- 逆 Y 轴
- 极角度轴
- 极半径轴
- 极彩虹

标签位置

从下拉菜单中选择图表的标签位置。

- 无
- 顶
- 左
- 右
- 底
- 内
- 内左
- 内右
- 内顶
- 内底
- 内左上
- 内左下
- 内右上
- 内右下

显示参考线

该切换控件启用/禁用图表中参考线的显示。可用选项为**最小值**、**最大值**和**平均值**，分别用于在图表的最小值、最大值和平均值处显示参考线。

输入参考线值

启用**显示参考线**后，此设置允许您指定特定参考线值。单击**添加另一个列**可指定其他参考线值。

转置

启用时，图表的 X 轴与 Y 轴会转置。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

箱图

箱图显示五项统计（最小值、第一个四分位数、中位数、第三个四分位数和最大值）。箱图可用于显示标度变量的分布以及确定离群值。

创建简式箱图

1. 在图表构建器的**图表类型**部分中，单击**箱图**图标。



工作区随即更新，以显示箱图模板。

2. 选择一个或多个标度变量作为**列**变量。

注: 点图的统计是箱图。您无法进行更改。

其他特征

列

列出可用于图表 X 轴的数据集变量。

单击**添加另一个列**可添加其他列。

分割依据

请选择一个分类变量，用于创建图表的表，分割依据变量的每个类别在其中都有一个单元格。与分组类似，在本质上，分割依据变量通过显示每个变量类别的信息，将更多维度添加到图表中。

类别顺序

请选择变量类别的排序顺序。

如所读取

变量类别以其出现在数据集中的顺序呈现。

升序(S)

变量类别以升序排序。

降序

变量类别以降序排序。

数据标准化

启用此设置后，就会将数据转换为正态分布，这样就可以对来自多个数据集或多个列的数据进行比较。这项设置会为计数创建 100% 的堆积，并将统计转换为百分比。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

气泡图

气泡图将组中的类别显示为非分层的填充圆。每个圆（气泡）的大小与其值成比例。气泡图很适合用于比较数据中的关系。

创建简式气泡图

1. 在图表构建器的**图表类型**部分中，单击**气泡图**图标。

Bubble



工作区随即更新，以显示气泡图模板。

2. 从下拉列表中选择**列**变量。

注:单击**添加另一个列**可以包括其他列变量。

其他特征

组颜色

打开或关闭颜色分组。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

蜡烛图

蜡烛图是一种财务图，用于描述证券、衍生品或货币的价格变动。每个蜡烛元素通常代表一天。一个月的图表可将 20 个交易日显示为 20 个蜡烛元素。蜡烛图常常用于分析股票和货币价格模式，类似于箱图。

用来创建蜡烛图的数据集必须包含所要显示的每个时间段的开盘值、最高值、最低值和收盘值。

创建简式蜡烛图

1. 在图表构建器的**图表类型**部分中，单击**蜡烛图**图标。



工作区随即更新，以显示蜡烛图模板。

2. 选择一个变量作为 **X 轴** 变量。
3. 选择一个变量作为**最高值**变量。
4. 选择一个变量作为**最低值**变量。

其他特征

X 轴

列出可用于图表 X 轴的数据集变量。

高

列出可用于图表最高价格值的数据集变量。

最高价字段汇总

请为选定的最高价变量选择统计汇总函数。

最低价

列出可用于图表最低价格值的数据集变量。

打开

列出可用于图表开盘价格值的数据集变量。

关闭(C)

列出可用于图表收盘价格值的数据集变量。

量

列出可用于图表成交量条形的数据集变量。

类别顺序

请选择变量类别的排序顺序。

如所读取

变量类别以其出现在数据集中的顺序呈现。

升序(S)

变量类别以升序排序。

降序

变量类别以降序排序。

烛台图

切换图表数据，以显示为蜡烛或折线。

移动平均值

提供用于在图表上显示移动平均值的选项。可用选项为 5、10、20、30、60 和 200。

切换颜色

该切换控件用于切换增加和减少颜色。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

圆填充图

圆填充图将分层数据显示为一组嵌套区域，允许您可视化大量分层结构化数据。它类似于树状图，但使用的是圆而不是矩形。圆填充图使用包含（嵌套）来显示层次结构数据。

创建简式圆填充图

1. 在图表构建器的**图表类型**部分中，单击**时间散点图**图标。

Circle packing



工作区随即更新，以显示圆填充图模板。

2. 从下拉列表中选择**列**变量。

注: 单击**添加另一个列**可以包括其他列变量。

其他特征

组颜色

打开或关闭颜色分组。

目录

选择统计汇总函数。这是用于汇总每个类别的方法。

统计汇总函数分为两类。它们之间的区别至关重要，因为这种区别确定您是否需要指定值变量。

- **不需要值变量的函数。** 这些是不需要变量的函数。所有的计数和百分比统计均归入此类别。不存在已定义的值变量时，这些统计可用。
- **确实需要值变量的函数。** 这些是需要值变量的函数。例如，*Mean* 函数需要一个要计算均值的变量。存在已定义的值变量时，这些统计可用。

值

选中需要值变量的汇总函数时，就会显示此字段。请选择要用作值的变量。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

定制图表

定制图表选项可用来粘贴/编辑 JSON 代码，以创建所需的图表。

创建定制图表

1. 在图表构建器的**图表类型**部分中，单击**定制**图标。



2. 将包含图表规范的 JSON 代码粘贴到“详细信息”窗格中提供的字段中。
3. 单击**生成图表**。

谱系图图表

谱系图图表与树状图类似，通常用于说明网络结构（例如，分层结构）。谱系图图表由根节点组成，根节点通过边缘或分支连接到从属节点。层次结构中最后的节点称为叶节点。

创建简单的谱系图图表

1. 在图表构建器的**图表类型**部分中，单击**谱系图**图标。

Dendrogram



画布会更新以显示谱系图图表模板。

2. 从**字段名称**列表中至少选择两个变量。您可以选择性地选择**全选**选项以选择所有可用变量。建议最多使用 300 个变量。

其他特征

字段名称

该列表提供表示叶节点的所有可用变量。您至少应选择两个变量。

显示轴

该切换控件启用或禁用距离值轴的显示。

链接

链接条件将观测值集之间的距离确定为观测值之间的成对距离的函数。选择**平均距离**、**最小距离**或**最大距离**。选择的链接选项会影响相关长度

树形布局

从左到右

根节点显示在左侧，叶节点显示在右侧。

从右到左

根节点显示在右侧，叶节点显示在左侧。

从上到下

根节点显示在顶部，叶节点显示在底部。

从下到上

根节点显示在底部，叶节点显示在顶部。

辐射

根节点显示在中间，叶节点从根节点辐射。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

双 Y 轴图

双 Y 轴图可用来汇总或绘制两个具有不同域的 Y 轴变量。例如，可以在一个轴上绘制观测值数目，并在另一个轴上绘制平均薪酬。此图表也可以是不同图形元素的混搭，以便双 Y 轴图中包含上文所探讨的多种不同图表类型。它可以将计数显示为折线，并将每个类别的平均值显示为条形。

创建简式双 Y 轴图

1. 在图表构建器的**图表类型**部分中，单击**双 Y 轴**图标。



工作区随即更新，以显示双 Y 轴图模板。

2. 选择一个变量作为 **X 轴** 变量。
3. 选择一个变量作为第一个 **Y 轴** 变量，然后选择代表该变量的图表类型（**条形图**、**折线图**或**散点图**）。
4. 选择一个变量作为第二个 **Y 轴** 变量，然后选择代表该变量的图表类型（**条形图**、**折线图**或**散点图**）。

注: 您可使用上移和下移控件来变更 Y 轴顺序。

其他特征

X 轴

列出可用于图表 X 轴的数据集变量。

Y 轴

列出可用于图表双 Y 轴的数据集变量。

目录

启用后，就会显示选项，供您选择用于汇总每个类别的方法。

左 Y 轴汇总

设置显示在图表左侧的 Y 轴的汇总方法。选项包括**合计**、**平均值**、**最大值**和**最小值**。

右 Y 轴汇总

设置显示在图表右侧的 Y 轴的汇总方法。选项包括**合计**、**平均值**、**最大值**和**最小值**。

数据标准化

启用后，数据就会变换为正态分布。这样，您就可以轻松比较来自于多个数据集或列的数据。

第二轴线

启用后，图表的第二轴线随即显示。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

误差条形图

误差条形图表示数据的可变性，并指出所报告测量的误差（或不确定性）。误差条形可帮助您确定差异在统计上是否显著。误差条形也可以指出给定函数的拟合优度。

创建简式误差条形图

1. 在图表构建器的**图表类型**部分中，单击**误差条形图**图标。



工作区随即更新，以显示误差条形图模板。

2. 选择一个标度变量作为**类别**变量。这是要在 X 轴上表示数据的变量。
3. 选择一个变量作为**Y 轴**变量。这是要在 Y 轴上表示数据的变量。

其他特征

类别

列出可用于图表 X 轴的数据集变量。

Y 轴

列出可用于图表 Y 轴的数据集变量。

类别顺序

请选择变量类别的排序顺序。

如所读取

变量类别以其出现在数据集中的顺序呈现。

升序(S)

变量类别以升序排序。

降序

变量类别以降序排序。

分割依据

请选择一个分类变量，用于创建图表的表，分割依据变量的每个类别在其中都有一个单元格。与分组类似，在本质上，分割依据变量通过显示每个变量类别的信息，将更多维度添加到图表中。

参考线

启用时，就会在图表上显示参考线。参考线与选择的**统计方法**相关。

误差条形图

启用时，在图表上就会显示表示误差范围的线条。

测量

选择误差条形所代表的测量类型：

置信区间

设置所选变量的置信区间。缺省值为 0.95 (95%)，此值会反映在**表示值**字段中。

标准误差

测量所选变量的标准误差。

标准差

测量所选变量的标准差。

置信度级别

此值表示所选**测量**的置信区间。缺省值为 0.95 (95%)。

统计方法

选择描述集中趋势的方法：

平均值

将比率求和，然后将结果除以比率总数所得的结果。

中位数

小于此值的比率数与大于此值的比率数相同。

显示方式

选择显示**统计方法**选定项的方式（条形、折线或圆圈）。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

评估图表

评估图表类似于直方图或集合图。评估图表显示模型如何执行对特定结果的预测。评估图表的工作原理是：根据预测值及预测的置信度对记录排序，将记录分割为大小相等的组（**分位数**），并按由高到低的顺序为每个分位数绘制标准值。在散点图中，将以单独的线条显示多个模型。

通过将具体值或值的范围定义为**匹配**，处理结果。通常，匹配表示相关的某类别（如向顾客销售）或某事件（如某项医疗诊断）成功执行。

标志

输出字段是正向的，即匹配表现为 *true* 值。

名义

对于名义输出字段，集合中的第一个值确定是否匹配。

连续

对于连续输出字段，大于字段范围中点的值即为匹配。

评估图表也可以累积，因此每个点等于相应分位数的值加上所有更高分位数的值。累积图表通常能够更好的表现模型性能，而非累积图则更有利于指出模型中可能存在问题的地方。

创建简式评估图表

1. 在图表构建器的**图表类型**部分中，单击**评估**图标。

Evaluation



工作区随即更新，以显示评估图表模板。

2. 设置**目标字段**、**预测字段**和**置信度字段**变量。目标字段可以是任意实例化标志或者具有两个或多个值的名义字段。预测字段定义了用作预测值的变量。置信度字段定义了用于建立预测置信度的变量。

注：预测字段变量类型必须匹配为目标字段选择的变量类型。

3. 指定用于指示**用户定义的匹配项**的定制条件。此选项更适合于定义相关结果，而不是从目标字段类型和值的顺序中推测结果。

必须为匹配条件指定 CLEM 表达式。例如，@TARGET = "YES" 是一个有效条件，指示目标字段的值 Yes 将在评估中计为命中。指定的条件将用于所有目标字段。

其他特征

累积散点图

启用时创建累积图表。累积图表中绘制的值代表每个分位数与所有更高分位数的和。

显示方式

设置控制了哪些图表显示在预览方式和输出中。

传统方式

选中此项后，模型分类调优图、分界值图、矩阵条形图、ROC 图、增益图、投资回报图和利润图在预览方式和输出中显示。

单一方式

选中此项后，模型分类调优图是在预览方式和输出中显示的唯一图表。

完全方式

选中此项后，模型分类调优图、分界值图、矩阵条形图、ROC 图、增益图、投资回报图、利润图、GINI 图、提升图和响应图在预览方式和输出中显示。

评估图表

分界值

分界值图显示所选变量针对给定分界值的预测值与实际值。

矩阵条形图

矩阵条形图是确定多个变量之间是否存在线性相关性的好方法。

ROC

ROC（受试者工作特征）评估分类方案的性能，其中有一个变量带有要用于对主体分类的两个类别。

收益

收益的定义是相对于全部匹配，发生于每个分位数中的匹配的百分比。其计算方法为（分位数中的匹配数量/全部匹配数量）× 100%。

投资回报率

投资回报 (ROI) 也需要确定收入和成本，从这一点上来说，它与利润相同。ROI 将分位数的成本和利润进行比较。其计算方法为（分位数利润/分位数成本）× 100%。

利润

利润等于每条记录的收入减去该记录的成本。也就是说，分位数的利润就是位于该分位数内的所有记录的利润总和。这里假定收入仅应用于匹配项，但成本可应用于所有的记录。利润及成本都可以是固定的，也可以由数据中的字段决定。其计算方法为（分位数中所有记录收入的总和 - 分位数中所有记录成本的总合）。

Kolmogorov-Smirnov

比较带有指定理论分布（可以是正态、均匀、指数或泊松）的变量的观测累积分布函数。

GINI

GINI 测量统计离散，旨在表示收入或财富分布。它是最常用的不平等性测量。

增益

提升将每个分位数中匹配记录的百分比与在全部训练数据中匹配的百分比进行比较。其计算方式为（在分位数中的匹配/在分位数中的记录）/（全部匹配/全部记录）。

响应

响应即分位数中，匹配记录的比例。响应的计算方法为（分位数中的命中数/分位数中的记录数）× 100%。

评估图表设置

以下设置仅适用于利润图和投资回报图。

成本

指定与每个记录关联的固定成本。

收入

指定与表示匹配项的每个记录关联的固定收入。

权重

如果数据中的记录代表多个单元，那么可以使用频率权重来调整结果。指定与每个记录关联的固定权重。

热图

热图将矩阵中包含的各个值表示为颜色，以此方式呈现数据。

创建简式热图

1. 在图表构建器的**图表类型**部分中，单击**热图**图标。



工作区随即更新，以显示热图模板。

2. 选择一个变量作为**列**变量。每个变量类别都表示为单个的图列表。
3. 选择一个变量作为**行**变量。每个变量类别都表示为单个的图表行。

其他特征

列

列出可用于图列表的数据集变量。每个变量类别都表示为单个的图列表。

行

列出可用于图表行的数据集变量。每个变量类别都表示为单个的图表行。

类别顺序

请选择变量类别的排序顺序。

如所读取

变量类别以其出现在数据集中的顺序呈现。

升序(S)

变量类别以升序排序。

降序

变量类别以降序排序。

目录

请为图形元素选择统计汇总函数。统计的结果确定图形元素在 y 轴上的位置。在二维图表中，为 x 轴上的每个值计算统计。在三维图表中，为 x 轴和 z 轴上的值的交点计算统计。

统计汇总函数分为两类。它们之间的区别至关重要，因为这种区别确定您是否需要指定**值**变量。

- **不需要值变量的函数。** 这些是不需要变量的函数。所有的计数和百分比统计均归入此类别。不存在已定义的值变量时，这些统计可用。
- **确实需要值变量的函数。** 这些是需要值变量的函数。例如，*Mean* 函数需要一个要计算均值的变量。存在已定义的值变量时，这些统计可用。

值

选中需要值变量的**汇总**函数时，就会显示此字段。请选择要用作值的变量。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

直方图

直方图的外观类似于条形图，但是每个条形代表数据在单个类别中的分布情况，而不是对各个类别进行比较或者寻找随时间变化的趋势。每个条形都代表连续的数据范围，或者特定数据点的频数。

直方图对于显示单个标度变量的分布情况非常有用。数据使用计数或百分比统计进行分级和汇总。频数多边形是直方图的一种变体，它与典型直方图类似，但使用的是面积图形元素而不是条状图形元素。

人口金字塔是直方图的另一种变体。它的名称源自于其最常见用途：对人口数据进行汇总。在用于人口数据时，它按性别拆分，以提供两个紧挨着的有关年龄数据的水平直方图。在人口为年轻型的国家/地区中，所产生的图形呈现金字塔形状。

创建直方图

1. 在图表构建器的**图表类型**部分中，单击**直方图**图标。



工作区随即更新，以显示直方图模板。

2. 选择一个标度变量作为 **X 轴** 变量。

注: 直方图的统计是直方图或直方图百分比。这些统计量对数据进行分箱化并为每个分箱计算计数值。

其他特征

X 轴

列出可用于图表 X 轴的数据集变量。

分割依据

请选择一个分类变量，用于创建图表的表，分割依据变量的每个类别在其中都有一个单元格。与分组类似，在本质上，分割依据变量通过显示每个变量类别的信息，将更多维度添加到图表中。

显示 KDE 曲线

启用后，在图表上就会显示内核密度估算曲线。

显示分布曲线

启用时，在图表上就会显示分布拟合曲线。

分布

此下拉列表提供以下分布选项。

自动拟合分布

这是缺省设置。

Beta

返回 Beta 分布中包含指定形状参数的值。

指数

返回指数分布中的值。

伽玛

返回伽玛分布中包含指定形状和尺度参数的值。

对数正态

返回对数正态分布中包含指定参数的值。

正态

返回正态分布中包含指定均值和标准差的值。

三角

返回三角分布中包含指定参数的值。

均匀

返回均匀分布中最小值与最大值之间的值。

韦伯

返回 Weibull 分布中包含指定参数的值。

分级宽度

滑块控制用于将数据分割为组的区间大小。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

折线图

折线图在图形上绘制一系列数据点，并以折线将其相连。显示存在细微差别的趋势线，或者数据线相互交叉的趋势线时，折线图特别适合。您可使用折线图来汇总分类变量，在这种情况下，它类似于条形图（请参阅第 53 页的『条形图』）。折线图也适用于时间序列数据。

创建简式时间序列折线图

1. 在图表构建器的**图表类型**部分中，单击**折线图**图标。



工作区随即更新，以显示折线图模板。

2. 选择一个日期变量作为 **X 轴** 变量。
3. 选择一个标度变量作为 **Y 轴** 变量。这是已记录一段时间内的值的变量。

其他特征

X 轴

列出可用于图表 X 轴的数据集变量。

Y 轴

列出可用于图表 Y 轴的数据集变量。

分割依据

请选择一个分类变量，用于创建图表的表，分割依据变量的每个类别在其中都有一个单元格。与分组类似，在本质上，分割依据变量通过显示每个变量类别的信息，将更多维度添加到图表中。

区域

启用时，折线下方的区域将以不同的颜色填充。

平滑

启用后，图表会显示平滑曲线。

显示数据点

启用后，数据点会在图表中显示。

重新排序

切换控件会根据 X 轴和 Y 轴值对数据重新排序。

显示参考线

启用后，就会在图表上显示基于指定 X 轴和 Y 轴值的参考线。

在 X 轴上输入参考线值

启用**显示参考线**后，此设置允许您指定 X 轴的特定参考线值。单击**添加另一个列**可指定其他参考线值。

在 Y 轴上输入参考线值

启用**显示参考线**后，此设置允许您指定 Y 轴的特定参考线值。单击**添加另一个列**可指定其他参考线值。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

地图

地图通常用来在地理区域之间比较值和显示类别，最适合在数据包含地理信息（国家、地区、省/自治区/直辖市、县、邮政编码，等等）时使用。

创建简式地图

1. 在图表构建器的**图表类型**部分中，单击**地图**图标。



工作区随即更新，以显示地图模板。

2. 从**地图服务**下拉列表选择用于提供地图图像的服务。该列表提供了涵盖特定全球区域的选项。
3. 从下拉菜单中选择地图**图表类型**。根据所选图表类型，以下选项可用：

经度

从下拉列表中选择要用作经度值的变量。

纬度

从下拉列表中选择要用作纬度值的变量。

组

从下拉菜单中选择用于对数据点位置分组的变量。

类别

选择要可视化的列变量。

目录

选择统计汇总函数。这是用于汇总每个类别的方法。

统计汇总函数分为两类。它们之间的区别至关重要，因为这种区别确定您是否需要指定**值**变量。

- **不需要值变量的函数。** 这些是不需要变量的函数。所有的计数和百分比统计均归入此类别。不存在已定义的**值**变量时，这些统计可用。
- **确实需要值变量的函数。** 这些是需要**值**变量的函数。例如，*Mean* 函数需要一个要计算均值的变量。存在已定义的**值**变量时，这些统计可用。

值

选中需要值变量的**汇总**函数时，就会显示此字段。请选择要用作值的变量。

其他特征

地图服务

列出可用于提供地图图像的服务。请参阅第 67 页的『[地图服务选项](#)』。

Type

列出可用于呈现数据的图表类型。

经度

列出可用作经度值的变量。

纬度

列出可用作纬度值的变量。

组

列出可用于对数据点位置分组的变量。

类别

列出列变量。

目录

选择统计汇总函数。这是用于汇总每个类别的方法。

统计汇总函数分为两类。它们之间的区别至关重要，因为这种区别确定您是否需要指定**值**变量。

- **不需要值变量的函数。** 这些是不需要变量的函数。所有的计数和百分比统计均归入此类别。不存在已定义的**值**变量时，这些统计可用。
- **确实需要值变量的函数。** 这些是需要**值**变量的函数。例如，*Mean* 函数需要一个要计算均值的变量。存在已定义的**值**变量时，这些统计可用。

值

选中需要值变量的**汇总**函数时，就会显示此字段。请选择要用作值的变量。

工具提示信息

列出当鼠标指针悬停在数据点上时，可用来生成工具提示信息的变量。

大小图

列出可用的大小图变量。这些变量使用不同大小在图点中表示自身。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

地图服务选项

在“地图图表”的“详细信息”窗格上，可以选择要用于提供地图图像的地图服务。

配置地图服务选项

1. 停止 IBM SPSS Modeler Server。
2. 如果您要应用本地 geojson 格式化的地图文件，请将其放在 <Modeler_installation_directory>/dataview/conf/public/mapfiles 目录中。如果该目录不存在，请进行创建。
3. 在文本编辑器中打开 <Modeler_installation_directory>/dataview/conf/application.conf 文件，并将以下选项添加到 map.resources 部分。在此示例中，添加了两个名为 World - Local Map 和 World - Online Map 的新映射。

```
map.resources {
  "mapservices" :
  [
    {
      "type"      : "geojson",
      "name"      : "World (Bulit-in)",
      "location"  : "built_in_world",
      "id"        : "built_in_world"
    },
    {
      "type"      : "geojson",
      "name"      : "Blank (Built-in)",
      "location"  : "built_in_blank",
      "id"        : "built_in_blank"
    },
    {
      "type"      : "geojson",
      "name"      : "World - Local Map",
      "location"  : "/mapfiles/world.json",
      "useProxy"  : "false",
      "id"        : "world_local"
    },
    {
      "type"      : "geojson",
      "name"      : "World - Online Map",
      "location"  : "http://map.example.com/map/world.json",
      "useProxy"  : "true",
      "id"        : "world_online"
    }
  ]
}
```

以下参数可用于地图选项：

选项	值	Type	必需?	描述
type	geojson	String	是	当前，此选项仅支持值 geojson。

表 4: 地图服务参数 (继续)				
选项	值	Type	必需?	描述
name	地图名称	String	是	输入地图的名称。此名称显示在地图图表的详细信息窗格上的地图服务下拉列表中。
location	地图位置	String	是	输入地图的 URL 或本地文件位置。对于 URL, 它必须以 http:// 或 https:// 开头。对于本地文件, 它必须以路径 / mapfiles/ 开头。
useProxy	true 或 false	String	是	如果地图位置是 URL, 请使用 true。如果地图位置是本地文件, 请使用 false。
id	地图标识	String	是	为每个地图使用唯一标识。

注:

- 不要移除缺省的 **World (Built-in)** 和 **Blank (Built-in)** 选项。
- 请务必使用逗号分隔内容块, 如示例中所示, 并确保新的 map.resources 部分是有效的 JSON 格式化内容。
- IBM SPSS Modeler Server (或不带服务器的独立 SPSS Modeler 客户机) 必须能够访问地图位置。

4. 启动 IBM SPSS Modeler Server。

现在, 使用地图图表时, 在详细信息窗格上, 您将看到添加的新地图服务。

数学曲线图

数学曲线图基于用户输入的表达式绘制数学方程式曲线。

创建简式数学曲线图

1. 在图表构建器的图表类型部分中, 单击数学曲线图标。

Math curve



工作区随即更新, 以显示数学曲线图模板。

2. 在 **X 值开始于** 字段中输入 X 轴的开始值。
3. 在 **X 值结束于** 字段中输入 X 轴的结束值。
4. 在“方程式”字段中输入用于绘制图形曲线的方程式。单击“添加另一个列”可包括其他方程式。

每个方程式将 x 视为因变量。方程式允许以下项:

- +、-、*、/、% 和 ^
- abs(x)
- ceil(x)

- floor(x)
- log(x)
- max(a,b,c...)
- min(a,b,c...)
- random()
- round(x)
- sqrt(x)
- sin
- cos
- exp
- tan
- atan
- atan2
- asin
- acos

其他特征

X 值开始于

X 轴开始值。

X 值结束于

X 轴结束值

方程式

用户输入的方程式，用于绘制图形曲线。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

多重图

多重图提供了创建多个图表的选项。这些图表可以为相同或不同类型，并且可以包含相同数据集中的不同变量。

创建简式多重图

1. 在图表构建器的**图表类型**部分中，单击**多重图**图标。

Multi-chart



工作区随即更新，以显示多重图模板。

2. 单击**添加另一个子图表**可添加子图表。
3. 从**类型**下拉列表中选择图表类型。
4. 根据选择的图表类型，可用的选项如下所示：

条形图和饼图

从下拉列表中选择**类别**变量。

折线散点图和散点图

从下拉列表中选择**X 轴**变量。

从下拉列表中选择 **Y 轴** 变量。

5. 单击**添加另一个子图表**以包括其他图表类型。

其他特征

标题

子图表标题。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

多系列图

多系列图类似于折线图，只不过可以在 Y 轴上绘制多个变量。

创建简式多系列图

1. 在图表构建器的**图表类型**部分中，单击**多系列图**图标。



工作区随即更新，以显示多系列图模板。

2. 选择一个变量作为 **X 轴** 变量。

3. 至少选择两个标度变量作为 **Y 轴** 变量。

其他特征

X 轴

列出可用于图表 X 轴的数据集变量。

Y 轴

列出可用于图表 Y 轴的数据集变量。

从下拉列表中选择图表类型（条形图、折线图或散点图）。

单击**添加另一个列**可在图表上包括更多的列。

序列样式

提供用于定义 Y 轴方向的选项。可用选项为：

- 缺省
- 单独的 Y 轴
- 显示第二个 Y 轴
- 双 Y 轴

数据标准化

启用此设置后，就会将数据转换为正态分布，这样就可以对来自多个数据集或多个列的数据进行比较。这项设置会为计数创建 100% 的堆积，并将统计转换为百分比。

次要 Y 轴

列出可用于图表次要 Y 轴的数据集变量。

从下拉列表中选择图表类型（条形图、折线图或散点图）。

单击**添加另一个列**可在图表上包括更多的列。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

平行图

在将高维空间几何可视化，以及分析多变量数据时，平行图十分有用。平行图类似于时间序列数据的折线图，但是轴并不对应于时间点（没有自然顺序）。

创建简式平行图

1. 在图表构建器的**图表类型**部分中，单击**平行图**图标。



工作区随即更新，以显示平行图模板。

2. 请至少选择两个变量作为**列**变量。每个列都代表图表中的一个垂直并行轴。

注：列顺序对于查找特征至关重要。在典型的数据分析中，可能需要对列进行多次重新排序。

其他特征

列

列出可用于图表 Y 轴的数据集变量。

单击**添加另一个列**可添加其他列。

颜色图

列出可用的颜色图变量。这些变量根据所指定列中值的范围，使用颜色级数在图点中表示自身。颜色图也称为等值线图。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

饼图

饼图适合对比例进行比较。例如，您可使用饼图来展示特定班级中女性成员的比例较大。

创建简式饼图

1. 在图表构建器的**图表类型**部分中，单击**饼图**图标。



工作区随即更新，以显示饼图模板。

2. 从**类别**列表中选择分类（名义或有序）变量。此变量中的类别确定饼图中的分区数。
3. 请为图形元素选择统计汇总函数。对于饼图，您通常需要基于计数的统计或者总和。统计结果确定每个分区的大小。

其他特征

类别

选择一个分类（名义或有序）变量，用于确定饼图中的分区数。

目录

请为图形元素选择统计汇总函数。对于饼图，您通常需要基于计数的统计或者总和。统计结果确定每个分区的大小。

统计汇总函数分为两类。它们之间的区别至关重要，因为这种区别确定您是否需要指定**值**变量。

- **不需要值变量的函数。** 这些是不需要变量的函数。所有的计数和百分比统计均归入此类别。不存在已定义的值变量时，这些统计可用。
- **确实需要值变量的函数。** 这些是需要值变量的函数。例如，*Mean* 函数需要一个要计算均值的变量。存在已定义的值变量时，这些统计可用。

值

选中需要标度变量的**汇总**函数时，就会显示此字段。请选择要用作标度变量的变量。

饼图类型

可用的样式包括：

正态

饼图段显示为常规分区。

环

饼图段显示为环。这种样式也称为圆环图。

玫瑰

与使用公共半径的常规饼图不同，在玫瑰式饼图中，饼图的段大小随它们的值不同而异。

玫瑰区域

与使用公共半径的常规饼图不同，在玫瑰式饼图中，饼图的段大小因其区域而异。

玫瑰圆环

与使用公共半径的常规饼图不同，在玫瑰式饼图中，饼图的段大小因其值而异，并且这些段显示为圆环。

半玫瑰

与玫瑰相同，只是图表表示为饼图的一半。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

人口金字塔图

人口金字塔图（也称为“年龄-性别金字塔”）一般用来根据年龄和性别来呈现和分析人口信息。

创建简式人口金字塔图

1. 在图表构建器的**图表类型**部分中，单击**人口金字塔图**图标。



工作区随即更新，以显示人口金字塔图模板。

2. 从下拉列表中选择 **Y 轴** 变量。
3. 从下拉列表中选择**分割依据**变量。

其他特征

Y 轴

列出可用于图表 Y 轴的数据集变量。

分割依据

请选择一个分类变量，用于创建图表的表，分割依据变量的每个类别在其中都有一个单元格。与分组类似，在本质上，分割依据变量通过显示每个变量类别的信息，将更多维度添加到图表中。

分级宽度

滑块控制用于将数据分割为组的区间大小。

显示分布曲线

启用时，在图表上就会显示分布拟合曲线。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

Q-Q 图

Q-Q（分位数-分位数）图通过两两绘制概率分布的分位数，对两种概率分布进行比较。Q-Q 图用于比较分布形状，以提供图形视图，说明两种分布中的位置、标度和偏度等属性有何相似或不同之处。

创建简式 Q-Q 图

1. 在图表构建器的**图表类型**部分中，单击 **Q-Q 图** 图标。



工作区随即更新，以显示 Q-Q 图模板。

2. 选择一个变量作为 **X 轴** 变量。

其他特征

X 轴

列出可用于图表 X 轴的数据集变量。

分布

此下拉列表提供所有可用的分布方法。

自动拟合分布

这是缺省设置。

Beta

返回 Beta 分布中包含指定形状参数的值。

指数

返回指数分布中的值。

伽玛

返回伽玛分布中包含指定形状和尺度参数的值。

对数正态

返回对数正态分布中包含指定参数的值。

正态

返回正态分布中包含指定均值和标准差的值。

均匀

返回均匀分布中最小值与最大值之间的值。

Student t

返回 Student t 分布中包含指定自由度的值。

图类型

请选择 Q-Q（分位数-分位数）图或 P-P（百分比-百分比）图。

自动拟合

启用时，就会自动估计所选**分布**的数据参数。禁用后，**形状**和**标度**分布值就会显示。

Shape1

为 **Beta** 分布设置 shape1 值。仅在未启用**自动拟合**并且将 **Beta** 选作**分布**时，此设置才可用。

Shape2

为 **Beta** 分布设置 shape2 值。仅在未启用**自动拟合**并且将 **Beta** 选作**分布**时，此设置才可用。

形状

为所选分布设置 shape 值。 仅在未启用**自动拟合**并且将**伽马**或**对数正态**选作**分布**时，此设置才可用。

Scale

为所选分布设置 scale 值。 仅在未启用**自动拟合**并且将**指数**、**伽马**或**对数正态**选作**分布**时，此设置才可用。

平均值

为**正态**分布设置均值。 仅在未启用**自动拟合**并且将**正态**选作**分布**时，此设置才可用。

标准差

为**正态**分布设置标准差值。 仅在将**正态**选作**分布**时，此设置才可用。

最小值

为**均匀**分布设置最小值。 仅在未启用**自动拟合**并且将**均匀**选作**分布**时，此设置才可用。

最大值

为**均匀**分布设置最大值。 仅在未启用**自动拟合**并且将**均匀**选作**分布**时，此设置才可用。

自由度 (df)

为 **Student t** 分布设置自由度。 仅在未启用**自动拟合**并且将 **Student t** 选作**分布**时，此设置才可用。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

雷达图

雷达图比较多个定量变量，有助于可视化哪些变量有类似值，或者各个变量中是否有离群值。雷达图由一系列辐条组成，其中每个辐条表示单个变量。雷达图还有助于确定哪些变量在数据集内的评分高或低。

创建简式雷达图

1. 在图表构建器的**图表类型**部分中，单击**雷达图**图标。

Radar



工作区随即更新，以显示雷达图模板。

2. 从下拉列表中选择**列**变量。

注:单击**添加另一个列**可包含其他列。 必须定义至少三个列变量。

其他特征

类别

选择一个分类（名义或有序）变量。 如果选择**无**作为类别值，将单独显示所有值，而不应用汇总方法。

目录

选择分类变量时，可用的汇总统计如下：

计数

观测值总数。

总和

值的总和。

平均值

算术平均值；总和除以观测值的数量。

最大值

最大（最高）值。

最小值

最小（最低）值。

雷达图布局

确定雷达图的后台图像布局：

圆形

选中此项时，雷达图将在圆形布局上显示。

多边形

选中此项时，雷达图将在多边形布局上显示。

分割依据

请选择一个分类变量，用于创建图表的表，分割依据变量的每个类别在其中都有一个单元格。与分组类似，在本质上，分割依据变量通过显示每个变量类别的信息，将更多维度添加到图表中。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

关系图

关系图可用于确定变量之间的关系。

创建简式关系图

1. 在图表构建器的**图表类型**部分中，单击**关系图**图标。



工作区随即更新，以显示关系图模板。

2. 请至少选择两个变量作为**列**变量。

注：单击**添加另一个列**可以包括其他列变量。

其他特征

列

列出可用的数据集变量。

单击**添加另一个列**可添加其他列。

线型

控制相关数据点之间的线型。

曲线

选中此项表示要在相关数据点之间绘制曲线。

直线

选中此项表示要在相关数据点之间绘制直线。

标注阈值

对于值超过已定义值的数据点，显示标签。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

散点图和点图

使用点图图形元素创建了几大类图表。

- **散点图。** 散点图适合于绘制多变量数据。它们可以帮助您确定标度变量之间的潜在关系。简式散点图使用二维坐标系来绘制两个变量。三维散点图使用三维坐标系来绘制三个变量。需要绘制更多变量时，您可尝试重叠散点图和散点图矩阵 (SPLOM)。重叠散点图显示重叠的 x-y 变量配对，每一对变量之间以颜色或形状区分。SPLOM 创建二维散点图的矩阵，每个变量都与 SPLOM 中的每个其他变量两两绘图。
- **点图。** 就像直方图一样，点图适合于显示单个标度变量的分布情况。数据会进行分级，但是，每个分级中所有的点均会显示并堆积，而非每个分级只有一个值（就像计数一样）。这些图形有时称为密度图。
- **汇总点图。** 这种图与条形图很相似，只不过在本来显示条形顶端的位置显示点。因为汇总点图与条形图是如此相似，请参阅第 53 页的『条形图』，以获取有关创建方式的信息。
- **垂线图。** 垂线图是一种特殊的汇总点图。各个点会进行分组，而且会绘制线条，穿过每个类别的各个点。垂线图适合在分类变量之间比较某一项统计。

创建简式散点图

1. 在图表构建器的**图表类型**部分中，单击**散点图**图标。



工作区随即更新，以显示散点图模板。

2. 选择一个标度变量作为 **X 轴** 变量。
3. 选择一个标度变量作为 **Y 轴** 变量。不需要指定统计，因为散点图通常显示原始值。

其他特征

X 轴

列出可用于图表 X 轴的数据集变量。

Y 轴

列出可用于图表 Y 轴的数据集变量。

颜色图

列出可用的颜色图变量。这些变量根据所指定列中值的范围，使用颜色级数在图点中表示自身。颜色图也称为等值线图。

大小图

列出可用的大小图变量。这些变量使用不同大小在图点中表示自身。

形状图

列出可用的形状图变量。这些变量使用不同形状在图点中表示自身。

拟合线

在拟合线中，数据点拟合到一条线，该线条通常只会穿过部分数据点。拟合线代表数据的趋势。某些拟合线基于回归。另一些拟合线基于迭代加权最小平方。请从下拉列表中选择拟合线选项。

梯度气泡图

该切换控件启用/禁用图表气泡中颜色梯度和三维效果的显示。选择**颜色图**变量时，会禁用该设置。

最小气泡大小

设置最小气泡大小。输入介于 5 和 20 之间的值。

最大气泡大小

设置最大气泡大小。输入介于 20 和 80 之间的值。

显示参考线

启用后，就会在图表上显示基于指定 X 轴和 Y 轴值的参考线。

在 X 轴上输入参考线值

启用**显示参考线**后，此设置允许您指定 X 轴的特定参考线值。单击**添加另一个列**可指定其他参考线值。

在 Y 轴上输入参考线值

启用显示参考线后，此设置允许您指定 Y 轴的特定参考线值。单击添加另一个列可指定其他参考线值。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

散点矩阵图

散点图矩阵是确定多个变量之间是否存在线性相关性的好方法。

创建散点矩阵图

1. 在图表构建器的图表类型部分中，单击散点矩阵图图标。



工作区随即更新，以显示散点矩阵图模板。

2. 请至少选择两个刻度列变量。

注：单击添加另一个列可以包括其他列变量。

每个选定的变量都会与其他每一个变量两两绘图，以创建各个散点图的矩阵。

其他特征

列

请至少选择两个矩阵变量。这些变量必须是数字（不能是日期格式）。

单击添加另一个列可添加其他列。

颜色图

列出可用的颜色图变量。这些变量根据所指定列中值的范围，使用颜色级数在图点中表示自身。颜色图也称为等值线图。

相关性

启用时，将会显示所选变量的线性相关性信息（强、中、弱）。

显示 KDE 曲线

启用后，在图表上就会显示内核密度估算曲线。

梯度气泡图

该切换控件启用/禁用图表气泡中颜色梯度和三维效果的显示。选择颜色图变量时，会禁用该设置。

最小气泡大小

设置最小气泡大小。输入介于 5 和 20 之间的值。

最大气泡大小

设置最大气泡大小。输入介于 20 和 80 之间的值。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

序列数组图表

序列数组图表包含单独的子图表，并在图例中显示所有子图表的 Y 轴。

创建序列数组图表

1. 在图表构建器的**图表类型**部分中，单击**序列数组**图标。

Series array



画布会更新以显示序列数组图表模板。

2. 单击**添加另一个子图表**来定义第一个序列数组子图表。
3. 选择一个变量作为 **X 轴** 变量。
4. 选择一个变量作为 **Y 轴** 变量。
5. 单击**添加另一个子图表**来定义其他序列数组子图表。

其他特征

X 轴

列出可用于子图表的 x 轴的数据集变量。

Y 轴

列出可用于子图表的 y 轴的数据集变量。

分割依据

请选择一个分类变量，用于创建图表的表，分割依据变量的每个类别在其中都有一个单元格。与分组类似，在本质上，分割依据变量通过显示每个变量类别的信息，将更多维度添加到图表中。

平滑

启用后，图表会显示平滑曲线。

显示数据点

启用后，数据点会在图表中显示。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

X 轴标签

显示在图表的 X 轴上的标签。

Y 轴标签

显示在图表的 Y 轴上的标签。

旭日图

旭日图有助于可视化分层数据结构。旭日图由一个内圆以及包围在外面的更深层次结构级别的圆环组成。每个段的角度与值成比例，或在其内段下进行等分。图表段根据所属的类别或分层级别进行着色。

创建筒式旭日图

1. 在图表构建器的**图表类型**部分中，单击**旭日图**图标。

Sunburst



工作区随即更新，以显示旭日图模板。

2. 从**列列表**中选择分类（名义或有序）变量。此变量中的类别确定图表中的段数。
3. 单击**添加另一个列**并从**列列表**中选择另一个分类（名义或有序）变量。此变量中的类别确定图表的第二个圆环中的段数，并表示分层级别。
4. 为图形元素选择统计汇总函数（基于计数的统计或总和）。统计结果确定每个段的大小。选择**总和**时，请从**值列表**选择标度变量以表示数据集内要汇总的值。

5. 选择旭日图布局选项（传统或发散）。

其他特征

列

选择一个分类（名义或有序）变量，用于确定图表中的段数。

目录

为图形元素选择统计汇总函数（基于计数的统计或总和）。统计结果确定每个分区的大小。

统计汇总函数分为两类。它们之间的区别至关重要，因为这种区别确定您是否需要指定值变量。

- **不需要值变量的函数。** 这些是不需要值变量的函数。所有的计数和百分比统计均归入此类别。不存在已定义的值变量时，这些统计可用。
- **确实需要值变量的函数。** 这些是需要值变量的函数。例如，*Sum* 函数需要一个要计算汇总的变量。

值

选中需要标度变量的汇总函数时，就会显示此字段。请选择要用作标度变量的变量。

放射状布局

可用的选项包括**传统**和**发散**。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

时间散点图

时间散点图按连续时间间隔展示数据点。您绘制的时间序列包含数字值，并且假设这些序列在其中的周期一致的一个时间范围内发生。时间图提供有关进本统计和检验的时间序列数据特征的初步分析，并因此在建模前生成有关数据的有用洞察。时间散点图包含分解、扩展 Dickey-Fuller 检验 (ADF)、相关性 (ACF/PACF) 和谱分析之类的分析方法。

创建简式时间散点图

1. 在图表构建器的**图表类型**部分中，单击**时间散点图**图标。

Time plot



工作区随即更新，以显示时间散点图模板。

2. 从**值**下拉列表，选择 y 轴的值。

其他特征

日期

从下拉列表中选择日期（如果适用）。每个观测值基本上按相同的时间间隔分隔。如果选择日期变量，将显示重新采样选项。可以使用此选项来汇总值字段以匹配指定间隔。

时间散点图算法

要用于分析时间序列数据的时间散点图算法：

- **分解。** 将时间序列分解为三个成分（趋势循环、季节和不定期）。分解以加性方式执行。
- **ADF 测试。** 扩展 Dickey-Fuller (ADF) 检验以下原假设：序列中有单位根，并且序列不是稳定的。如果检验结果拒绝原假设，这意味着序列是稳定的，或者可以使用不同的模型将序列表示为稳定的。
- **ACF/PACF。** 序列的相关性。

- **谱分析。** 频率域中的分析工具。最高频率标记为钻石。

交换图表位置

反转图表在平面直角坐标系和极坐标系上的位置。

显示转折点

显示或隐藏图表中的转折点。基于时间序列分解的趋势循环成分，我们探索序列是否存在某个总体趋势，或是否存在会改变趋势模式方向的某些转折点。

显示离群值

显示或隐藏任何离群值。时间序列的离群值从时间序列分解的不定期成分进行分析。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

主题江河图

主题江河图是专门的流程图，用于显示随时间推移的变化。

创建简式主题江河图

1. 在图表构建器的**图表类型**部分中，单击**主题江河图**图标。

Theme River



工作区随即更新，以显示主题江河图模板。

2. 选择 **X 轴** 变量。
3. 选择**类别** 变量。

限制: 如果类别字段有超过 50 个不同类别，那么只有前 50 个最大类别将用作图表中的事件。

其他特征

排序依据

根据所选的 X 轴值，您可能可以指定类别顺序是基于类别名称还是类别值。

类别顺序

根据所选的 X 轴值，您可能可以指定类别顺序是升序、降序还是按照从数据集读取的顺序。

目录

选择统计汇总函数。这是用于汇总每个类别的方法。

统计汇总函数分为两类。它们之间的区别至关重要，因为这种区别确定您是否需要指定**值**变量。

- **不需要值变量的函数。** 这些是不需要变量的函数。所有的计数和百分比统计均归入此类别。不存在已定义的值变量时，这些统计可用。
- **确实需要值变量的函数。** 这些是需要值变量的函数。例如，*Mean* 函数需要一个要计算均值的变量。存在已定义的值变量时，这些统计可用。

值

选中需要值变量的**汇总**函数时，就会显示此字段。请选择要用作值的变量。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

树图表

树图表以类似于树的结构来表示层次结构。树图表的结构是由根节点（它没有父节点）、称为“分支”的线条连接（代表成员之间的关系和连接）以及叶节点（它们没有子节点）所组成。

创建简式树图表

1. 在图表构建器的**图表类型**部分中，单击**树**图标。

Tree



工作区随即更新，以显示树图表模板。

2. 从下拉列表中选择**列**变量。

注:单击**添加另一个列**可以包括其他列变量。

其他特征

列

列出可用于表示图表列的数据集变量。

目录

请为图形元素选择统计汇总函数。统计结果确定图形元素的位置。

统计汇总函数分为两类。它们之间的区别至关重要，因为这种区别确定您是否需要指定**值**变量。

- **不需要值变量的函数。** 这些是不需要值变量的函数。所有的计数和百分比统计均归入此类别。不存在已定义的**值**变量时，这些统计可用。
- **确实需要值变量的函数。** 这些是需要**值**变量的函数。例如，*Mean* 函数需要一个要计算均值的变量。存在已定义的**值**变量时，这些统计可用。

值

选中需要值变量的**汇总**函数时，就会显示此字段。请选择要用作汇总值的基础的变量。

树形布局

从左到右

根节点显示在左侧，叶节点显示在右侧。

从右到左

根节点显示在右侧，叶节点显示在左侧。

从上到下

根节点显示在顶部，叶节点显示在底部。

从下到上

根节点显示在底部，叶节点显示在顶部。

辐射

根节点显示在中间，叶节点从根节点辐射。

叶深度

为叶节点设置向下钻取级别值。

显示叶节点标签

启用后，就会显示每个叶节点的标签。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

树状图

树状图是将树形图的分层结构可视化的替代方法，在这种图中，还会显示每一个类别的数量。树状图对于您识别数据中的模式来说特别有用。树分支以矩形表示，每个子分支以更小的矩形表示。

创建简式树状图

1. 在图表构建器的**图表类型**部分中，单击**树状图**图标。



工作区随即更新，以显示树状图模板。

2. 从下拉列表中选择**列**变量。

注:单击**添加另一个列**可以包括其他列变量。

其他特征

列

列出可用于表示图表列的数据集变量。

目录

请为图形元素选择统计汇总函数。统计结果确定图形元素的位置。

统计汇总函数分为两类。它们之间的区别至关重要，因为这种区别确定您是否需要指定**值**变量。

- **不需要值变量的函数。** 这些是不需要值变量的函数。所有的计数和百分比统计均归入此类别。不存在已定义的**值**变量时，这些统计可用。
- **确实需要值变量的函数。** 这些是需要**值**变量的函数。例如，*Mean* 函数需要一个要计算均值的变量。存在已定义的**值**变量时，这些统计可用。

值

选中需要值变量的**汇总**函数时，就会显示此字段。请选择要用作汇总值的基础的变量。

叶深度

为叶节点设置向下钻取级别值。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

字云图

字云图将数据显示为字，其中，任意单个字的大小和位置均由其权重确定。

创建简式字云图

1. 在图表构建器的**图表类型**部分中，单击**字云图**图标。



工作区随即更新，以显示字云图模板。

2. 选择一个变量作为**源**变量。在图表中，每个变量类别都会根据其权重值来呈现。
3. 选择图表的**形状**值。所生成的图表数据将以选定的形状呈现。

其他特征

源

列出可用作图表源的数据集变量。在图表中，每个变量类别都会根据其权重值来呈现。

形状

列出可用的图表形状。所生成的图表数据将以选定的形状呈现。

主标题

图表标题。

子标题

图表子标题，它显示在图表标题正下方。

脚注

图表脚注，它显示在图表下方。

仪表板

可以创建图表仪表板布局以同时查看多个图表。将布局保存为模板，并将保存的图表拖放到布局中的位置。

1. 右键单击数据节点并选择**查看数据**后，单击**操作**部分中的“仪表板”控件。



图 14: “仪表板”控件

“仪表板”显示并提供下列设置。

模板

在此部分中，可以创建新的布局模板，或者从预定义模板或从保存的模板中进行选择。

选择布局模板

从可用布局模板中进行选择，或者以新模板开始。

操作

单击**编辑布局**图标以编辑所选的布局模板。您还可以导入仪表板文件。完成后，保存模板。

完成后单击**离开布局编辑方式**图标。

内容

使用此部分可将项拖放到仪表板布局。

选择保存的图表

此部分提供保存的图表的列表。将图表拖放到仪表板布局上的所需位置。

选择对象

还可以将 HTML 文本或图像拖动到仪表板布局。

全局可视化首选项

您可以覆盖标题、范围滑块、网格线和鼠标追踪行为的缺省设置。您还可以指定其他颜色方案模板。

1. 右键单击数据节点并选择**查看数据**后，单击**操作**部分中的“全局可视化首选项”控件。



图 15: “全局可视化首选项”控件

“全局可视化首选项”对话框随即显示，其中会提供下列设置。

标题

此部分提供全局图表标题设置。

全局标题

启用或禁用所有图表的全局标题。

全局主标题

启用或禁用全局主图表标题的显示。启用时，您在此处输入的顶级图表标题就会应用于所有图表，从而有效覆盖每个图表各自的主标题设置。

全局子标题

启用或禁用全局图表子标题的显示。启用时，您在此处输入的图表子标题就会应用于所有图表，从而有效覆盖每个图表各自的子标题设置。

缺省标题

针对所有图表启用或禁用缺省标题。

标题对齐方式

提供了标题对齐选项：**左对齐**、**居中对齐**（缺省设置）和**右对齐**。

工具(T)

此部分提供用于控制图表行为的选项。

范围滑块

启用或禁用每个图表的范围滑块。启用后，您可通过每个图表下提供的范围滑块来控制所显示的图表数据量。

网格线

控制 X 轴（垂直）和 Y 轴（水平）网格线的显示。

鼠标追踪程序

启用时，就会追踪与图表数据相关的鼠标光标位置，并在鼠标光标停留在图表上时显示其位置。

工具箱

启用或禁用每个图表的工具箱。根据图表类型，屏幕右侧的工具箱提供工具，例如，缩放、另存为图像、复原、选择数据和清除选择。

ARIA

启用后，身有残障的用户更容易访问 Web 内容和 Web 应用程序。

过滤掉空值

启用或禁用空图表数据的过滤。

X 轴置于原点

启用后，X 轴将位于另一个轴的原点位置。禁用后，X 轴始终从 0 开始。

Y 轴置于原点

启用后，Y 轴将位于另一个轴的原点位置。禁用后，Y 轴始终从 0 开始。

显示 X 轴标签

启用或禁用 X 轴标签。

显示 Y 轴标签

启用或禁用 Y 轴标签。

显示 X 轴线

启用或禁用 X 轴线。

显示 Y 轴线

启用或禁用 Y 轴线。

显示 X 轴名称

启用或禁用 X 轴名称。

显示 Y 轴名称

启用或禁用 Y 轴名称。

Y 轴名称位置

下拉列表提供用于指定 Y 轴名称位置的选项。选项包括**开始**、**中间**和**结束**。

截断长度

指定的值用于设置字符串长度。超过指定长度的字符串将被截断。缺省值为 10；如果指定了 0，那么将关闭截断功能。

X 轴刻度标签十进制值

设置 X 轴的刻度标签十进制值。缺省值为 3。

Y 轴刻度标签十进制值

设置 Y 轴的刻度标签十进制值。缺省值为 3。

X 轴刻度标签旋转

设置 X 轴刻度标签旋转值。缺省值为 0（无旋转）。您可以指定范围在 -90 到 90 度之间的值。

主题

选择模板，以更改具有分组或堆积变量的图表中使用的颜色。所选模板文件中定义的任何元素属性都会覆盖这些元素属性的缺省模板设置。

2. 单击**应用**以保存设置，或单击**取消**以废弃更改。

第 7 章 处理输出

运行某些流时，将通过模型块节点的**模型**选项卡或**高级**选项卡，在查看器中提供结果。在“查看器”中，您可以容易地浏览到要查看的输出。还可以操作输出并创建刚好包含您需要的输出的文档。某些图形输出还使用查看器。

查看器将用于 IBM SPSS Modeler 中的以下输出：

- TCM 模型块
- STP 模型块
- TwoStep-AS 聚类模型块
- GSAR 模型块
- 地图可视化图形节点

查看器

结果显示在查看器中。您可以将查看器用于：

- 浏览结果
- 显示或隐藏选中的表格和图表
- 通过移动选中的项更改结果的显示顺序
- 在查看器和其他应用程序间移动各项

查看器分为两个窗格：

- 左窗格包含内容的概要视图。
- 右窗格包含统计表、图表和文本输出。

可以单击概要中的项直接转到相应的表格或图表。可以单击并拖动概要窗格的右边框来更改概要窗格的宽度。

显示和隐藏结果

在查看器中，您可以选择性地显示和隐藏整个过程中的个别表或结果。当您想减少内容窗格中的可视输出量时，此过程十分有用。

隐藏表格和图表

1. 双击查看器的概要窗格中项的书形图标。
或
2. 单击某项将其选中。
3. 从菜单中选择：
视图 > 隐藏
或
4. 单击“概要”工具栏上的合上的书形（隐藏）图标。

打开的书形（显示）图标变为活动图标，表示该项现在已隐藏。

隐藏过程结果

1. 在概要窗格中单击过程名称左边的框。

这将隐藏程序的所有结果，从而折叠概要视图。

移动、删除和复制输出

可以通过复制、移动或删除一项或一组项来重新排列结果。

在查看器中移动输出

1. 选择概要窗格或内容窗格中的某项。
2. 将选中的项拖放到不同的位置。

您在步骤 1 中剪切的项将移动到所选位置。

在查看器中删除输出

1. 选择概要窗格或内容窗格中的某项。
2. 按 **Delete** 键。

或

3. 从菜单中选择：

编辑 > 删除

更改初始对齐方式

缺省情况下，所有的结果最初都是左对齐的。要更改新输出项的初始对齐方式：

1. 从菜单中选择：

编辑 > 选项

2. 单击查看器选项卡。
3. 在“初始输出状态”组中，选择项类型（例如，透视表、图表、文本输出）。
4. 选择所需的对齐选项。

更改各输出项的对齐方式

1. 在概要或内容窗格中，单击要对齐的项。
2. 从菜单中选择：

格式 > 左对齐

或

格式 > 居中

或

格式 > 右对齐

查看器大纲

“大纲”窗格提供查看器文档的目录。可以使用概要窗格浏览结果并控制显示。概要窗格中的大多数操作对内容窗格有相应的影响。

- 在概要窗格中选择一项将在内容窗格中显示相应的项。
- 在概要窗格中移动一项将移动内容窗格中相应的项。
- 折叠概要视图将隐藏折叠级别中的所有项的结果。

控制概要显示

要控制概要显示，您可以：

- 展开和折叠概要视图

- 更改选定各项的概要级别
- 更改概要显示中的各项的大小
- 更改概要显示中使用的字体

折叠和展开概要视图

1. 单击要折叠或展开的概要项左边的框。

或

2. 单击概要中的项。
3. 从菜单中选择：

视图 > 折叠

或

视图 > 展开

更改概要级别

1. 在概要窗格中单击项。
2. 从菜单中选择：

编辑 > 概要 > 上升

或

编辑 > 概要 > 降级

更改概要项的大小

1. 从菜单中选择：
查看 > 概要大小
2. 选择概要大小（小、中或大）。

更改概要中的字体

1. 从菜单中选择：
查看 > 概要字体...
2. 选择一种字体。

编辑项并将其添加到查看器中

在查看器中，您可以编辑和添加项，例如标题、新文本、图表或来自其他应用程序的材料。

添加标题或文本

可将未连接到表格或图表的文本项添加到查看器。

1. 单击将放在标题或文本前面的表格、图表或其他对象。
2. 从菜单中选择：

插入 > 新标题

或

插入 > 新文本

3. 双击该新对象。
4. 输入文本。

添加文本文件

1. 在查看器的概要窗格或内容窗格中，单击要放在文本前面的表格、图表或其他对象。
2. 从菜单中选择：
 插入 > 文本文件...
3. 选择文本文件。

要编辑文本，请双击该文本。

将对象粘贴到查看器

可以将其他应用程序中的对象粘贴到查看器中。 您可使用**粘贴到后面**或**选择性粘贴**。这两种类型的粘贴都会将新对象放在查看器中当前选定的对象之后。 如果要选择所粘贴对象的格式，请使用**选择性粘贴**。

在查看器中查找并替换信息

1. 要在查看器中查找或替换信息，请从菜单中选择：

编辑 > 查找

 或

编辑 > 替换

您可使用“查找和替换”来：

- 搜索整个文档或仅选中项。
- 从当前位置向下或向上搜索。
- 搜索窗格或限制搜索内容或概要窗格。
- 搜索隐藏项。其中包括内容窗格中隐藏的任何项（例如，缺省情况下隐藏的“注释”表）和透视表中隐藏的行和列。
- 限制搜索标准至大小写匹配。
- 限制透视表中的搜索标准至整个单元格内容匹配。
- 限制透视表中的搜索标准至仅限脚注标记。如果查看器中的选项包含透视表以外的其他任何内容，此选项不可用。

隐藏项目和透视表层

- 多维透视表当前可视层以下的层并不视为隐藏，并且会包含在搜索范围中，即使隐藏项不包含在搜索中。
- 隐藏项目包括内容窗格中的隐藏项（概要窗格中有合上的书形图标的项，或包含在概要窗格的折叠块中的项）和透视表中缺省情况下隐藏的行和列（例如，缺省情况下隐藏的空行和空列）或手动隐藏的行和列（通过编辑表并选择隐藏特定行或列）。隐藏项目仅在您明确选择**包含隐藏项目**时包含在搜索中。
- 在两种情况下，包含搜索文本或值的隐藏或不可见元素会在找到时显示，但项目随后即返回其原始状态。

在透视表中查找值范围

要在透视表中查找属于指定范围值中的值：

1. 在查看器中激活透视表或者选择一个或多个透视表。确保只选中了透视表。如果选中了其他任何对象，“范围”选项不可用。
2. 从菜单中选择：
 编辑 > 查找
3. 单击**范围**选项卡。
4. 指定范围类型：之间、大于或等于、小于或等于。
5. 选择定义范围的一或多个值。

- 如果任一值包含非数值的字符，将两个值视为字符串。
- 如果两个值都是数字，只搜索数值。
- 您不能使用“范围”选项卡取值。

此功能对遗存表不可用。有关更多信息，请参阅主题 [第 111 页的『遗存表』](#)。

复制输出到其他应用程序中

输出对象可以复制并粘贴到其他应用程序中，如字处理程序或电子表格。可以粘贴多种格式的输出。根据目标应用程序和选定输出对象，可以使用以下部分或所有的格式：

图元文件。 WMF 和 EMF 图元文件格式。这些格式只在 Windows 操作系统上可用。

RTF (富文本格式)。 可以 RTF 格式复制和粘贴多个选定对象、文本输出、透视表。对于透视表，在大多数应用程序中，这意味着将表作为表格粘贴，然后可在其他应用程序中编辑它。根据透视表选项设置，将对文档宽度过宽的透视表进行换行、按比例缩小以适合文档刻度或不进行更改。有关更多信息，请参阅主题 [第 112 页的『透视表选项』](#)。

注: Microsoft Word 可能无法正确显示过宽的表格。

图像。 JPG 和 PNG 图像格式。

BIFF。 可以 BIFF 格式粘贴透视表和文本输出到电子表格。透视表中的数字保留数值精确度。此格式只在 Windows 操作系统上可用。

文本。 可以文本格式复制和粘贴透视表和文本输出。此过程对于电子邮件这样的应用程序很有用，因为这种应用程序只接受或传送文本。

Microsoft Office 图形对象。 支持该格式的图表可以复制到 Microsoft Office 应用程序并在这些应用程序中作为原生 Microsoft Office 图表进行编辑。由于 SPSS Statistics/SPSS Modeler 图表和 Microsoft Office 图表之间存在差异，SPSS Statistics/SPSS Modeler 图表的一些功能不会保留在复制的版本中。不支持复制多个“Microsoft Office 图形对象”格式的选中图表。

如果目标应用程序支持多种可用格式，则可能包含允许您选择格式“选择性粘贴”菜单项，或自动显示可用格式列表。

注: 必须要有 Microsoft Office V16 或更高版本才能复制和粘贴箱图和直方图。

复制和粘贴多个输出对象

当粘贴多个输出对象到其他应用程序时，以下限制适用：

- **RTF 格式。** 在大多数应用程序中，透视表作为表格粘贴，并可在该应用程序中编辑。以图像格式粘贴图表、树和模型视图。
- **图元文件和图像格式。** 在其他应用程序中，所有选定输出对象皆作为单一对象粘贴。
- **BIFF 格式。** 排除图表、树和模型视图。

选择性复制

当复制和粘贴大量输出，特别是大量透视表时，您可以使用 **编辑 > 选择性复制** 来限制复制到剪贴板的格式数，以提高操作速度。

您也可以保存选定格式作为复制到剪贴板的缺省设置格式。此项设置在会话之间保持不变。

复制为

您可以在输出查看器中右键单击选中的对象并选择 **编辑 > 复制为** 以复制为最受欢迎的格式（例如，全部、图像或 Microsoft Office 图形对象）。选择 **编辑 > 复制** 将复制全部。请注意，如果 **复制为** 呈灰色或者对于选中的对象不存在，那么此复制格式不适用于此特定对象。

交互输出

交互输出对象包含多个相关的输出对象。在一个对象中的选择可以更改在另一个对象中显示或突出显示的内容。例如，在某个表中选择某行可能会突出显示某个地图中的区域或显示另一类别的图表。

交互输出对象不支持编辑功能，例如更改文本、颜色、字体或表边界。可以将各个对象从交互对象复制到查看器。可以在透视表编辑器中编辑从交互输出中复制的表。

从交互输出中复制对象

文件复制到查看器用于将各个输出对象复制到“查看器”窗口。

- 可用选项取决于交互输出的内容。
- **图表和地图**用于创建图表对象。
- **表**用于创建可以在透视表编辑器中编辑的透视表。
- **快照**用于创建当前视图的图像。
- **模型**用于创建当前交互输出对象的副本。

编辑复制对象用于将各个输出对象复制到剪贴板。

- 将复制的对象粘贴到查看器中等同于单击**文件>复制到查看器**。
- 将对象复制到另一应用程序中会将对象作为图像进行粘贴。

缩放和平移

对于地图，您可以使用**视图>缩放**来缩放地图的视图。在已缩放的地图视图中，您可以使用**视图>平移**来移动视图。

打印设置

文件>打印设置用于控制打印交互对象的方式。

- **仅打印可视视图**。仅打印当前显示的视图。此选项为缺省设置。
- **打印所有视图**。打印交互输出中包含的所有视图。
- 所选选项还可以确定用于导出输出对象的缺省操作。

导出输出

导出输出会以 HTML、文本、Word/RTF、Excel、PowerPoint（需要 PowerPoint 97 或更高版本）和 PDF 格式保存查看器输出。图表也可以以多种不同图形格式导出。

注：导出到 PowerPoint 仅在 Windows 操作系统上可用。

导出输出

1. 使查看器成为活动对话框（单击此对话框中的任意位置）。
2. 单击工具栏上的**导出按钮**，或右键单击输出窗口并选择**导出**。
3. 输入文件名（或图表前缀）并选择导出格式。

要导出的对象。您可以导出查看器中的所有对象、所有可视对象或者仅选定的对象。

文档类型。可用选项包括：

- **Word/RTF (*.doc)**。透视表导出为 Word 表，并带有所有完整的格式属性（例如，单元格边框、字体样式和背景色）。文本输出导出为格式化的 RTF。图表、树形图和模型视图包括在 PNG 格式中。请注意，Microsoft Word 可能无法正确显示过宽的表格。
- **Excel 97-2004 (*.xls)/Excel 2007 及更高版本 (*.xlsx)**。透视表行、列和单元格导出为 Excel 行、列和单元格，并带有所有完整的格式属性（例如单元格边框、字体样式和背景色）。文本输出连同所有字体属性

一起导出。文本输出中的一行即为 Excel 文件中的一行，文本输出中行的所有内容包含在单个单元格中。图表、树形图和模型视图包括在 PNG 格式中。输出可以作为 *Excel 97-2004* 或 *Excel 2007* 及更高版本导出。

- **HTML (*.htm)**。透视表导出为 HTML 表。文本输出导出为预设置格式的 HTML。图表、树形图和模型视图都按所选图形格式嵌入文档中。需要使用与 HTML 5 兼容的浏览器才能查看以 HTML 格式导出的输出。
- **可移植文档格式 (*.pdf)**。所有输出都将按打印预览中的显示导出，所有格式设置特性都不变。
- **文本 - 纯文本/UTF8/UTF16 (*.txt)**。文本输出格式包括纯文本、UTF-8 和 UTF-16。透视表可以制表符分隔格式或空格分隔格式导出。所有文本输出以空格分隔格式导出。对于图表、树形图和模型视图，文本文件中将针对每个图片插入一个相应的行，指示图像文件名称。
- **无 (仅图形)**。可用导出格式包括：EPS、JPEG、TIFF、PNG 和 BMP。在 Windows 操作系统上还可使用 EMF (增强型图元文件) 格式。

打开包含文件夹。打开包含导出所创建文件的文件夹。

HTML 选项

HTML 导出需要与 HTML 5 兼容的浏览器。

以下选项可用于以 HTML 格式导出输出：

透视表中的层。缺省情况下，包括或排除透视表层由各个透视表的表属性控制。您可以覆盖此设置，包含所有层或排除当前可视层之外的所有层。有关更多信息，请参阅第 107 页的『表属性：打印』主题。

将分层表导出为交互式表。分层表将按其在查看器中的外观显示，而且您可以在浏览器中以交互方式更改显示的层。如果不选择此选项，则每个表层都会显示为一个独立的表。

表为 HTML。此选项控制导出透视表包括的样式信息。

- **使用样式和固定列宽导出**。所有透视表样式信息（字体样式、背景色等）同时保留列宽。
- **不带样式导出**。透视表转换为缺省 HTML 表格。不保留样式属性。自动确定列宽。

包括脚注及文字说明。控制包括或排除所有透视表脚注及文字说明。

模型的视图。缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。有关更多信息，（注：包括表在内的所有模型视图将导出为图形。）

注：对于 HTML，您还可以控制导出图表的图像文件格式。有关更多信息，请参阅主题第 97 页的『图形格式选项』。

设置 HTML 导出选项

1. 选择 **HTML** 作为导出格式。
2. 单击**更改选项**。

Web 报告选项

Web 报告是与大部分浏览器兼容的交互式文档。查看器中可用的许多透视表交互功能在 Web 报告中也可用。

报告标题。报告头中显示的标题。缺省情况下使用文件名。您可以指定定制标题以代替文件名使用。

格式。报告格式有两个选项：

- **SPSS Web 报告 (HTML 5)**。此格式需要与 HTML 5 兼容的浏览器。
- **Cognos 活动报告 (mht)**。此格式需要支持 MHT 格式文件或 Cognos 活动报告应用程序的浏览器。

排除对象。您可以从报告中排除选定对象类型：

- **文本**。非日志的文本对象。此选项包括含有关于活动数据集的信息的文本对象。
- **日志**。包含已运行的命令语法列表的文本对象。日志项还包含未生成任何查看器输出的命令遇到的警告和错误消息。

- **注释表。** 统计和图表绘制过程的输出包含“注释”表。此表包含有关使用的数据集、缺失值和用于运行过程的命令语法的信息。
- **警告和错误消息。** 来自统计和绘图过程的警告和错误消息。

重新设置表和图表的样式以使其与 Web 报告匹配。 此选项将标准 Web 报告样式应用于所有表和图表。此设置将覆盖查看器中显示的输出中的任何字体、颜色或其他样式。您无法修改标准 Web 报告样式。

Web 服务器连接。 您可以包括正在运行 IBM SPSS Statistics Web 报告应用程序服务器的一个或多个应用程序服务器的 URL 位置。Web 应用程序服务器提供了用于透视表、编辑图表以及保存已修改的 Web 报告的功能。

- 对于每个要包括在 Web 报告中的应用程序服务器，请选择**使用**。
- 如果 Web 报告包含 URL 指定项，那么 Web 报告将连接到该应用程序服务器以提供额外的编辑功能。
- 如果您指定了多个 URL，那么 Web 报告将尝试按照指定 URL 的顺序来连接到每个服务器。

可以从 <http://www.ibm.com/developerworks/spsdevcentral> 下载 IBM SPSS Statistics Web Report Application Server。

Word 选项

以下选项可用于以 Word 格式导出输出：

保留分界点。 如果您已定义分界点，那么这些设置将保留在 Word 表格中。

包括脚注和文字说明。 控制包括或排除所有透视表脚注及文字说明。

模型的视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。有关更多信息，（注：包括表在内的所有模型视图将导出为图形。）

导出的页面设置。 这将打开一个对话框，在此可以定义导出文档的页面大小和边距。用于确定换行和缩小行为的文档宽度是页面宽度减去左右边距。

设置 Word 导出选项

1. 选择 **Word/RTF** 作为导出格式。
2. 单击**更改选项**。

Excel 选项

以下选项可用于以 Excel 格式导出输出：

创建工作表或工作簿，或者修改现有工作表。 缺省情况下，将创建一个新工作簿。如果拥有指定名称的文件已存在，将被覆盖。如果您选择创建工作表的选项，且拥有指定名称的工作表已存在于指定文件中，则将被覆盖。如果您选择修改现有工作表的选项，您还必须指定工作表名称。（这对创建工作表为可选选项。）工作表名称不能超过 31 个字符，不能包含正斜杠或反斜杠、方括号、问号或星号。

导出到 Excel 97-2004 时，如果您修改一个现有工作表，图表、模型视图和树形图则不包括在导出输出中。

工作表中的位置。 控制导出输出在工作表中的位置。缺省情况下，导出输出将添加到拥有任何内容的最后一列之后，从第一行开始，而不修改任何现有内容。这对于将新列添加到现有工作表是一个不错的选择。将导出输出添加到最后一行之后对于将新行添加到现有工作表是一个不错的选择。将导出输出添加到特定单元格位置的开始将覆盖添加导出输出的区域中的任何现有内容。

透视表中的层。 缺省情况下，包括或排除透视表层由各个透视表的表属性控制。您可以覆盖此设置，包含所有层或排除当前可视层之外的所有层。有关更多信息，请参阅第 107 页的『表属性：打印』主题。

包括脚注和文字说明。 控制包括或排除所有透视表脚注及文字说明。

模型的视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。有关更多信息，（注：包括表在内的所有模型视图将导出为图形。）

设置 Excel 导出选项

1. 选择 **Excel** 作为导出格式。
2. 单击**更改选项**。

PowerPoint 选项

对于 PowerPoint，可用的选项如下所示：

透视表中的层。 缺省情况下，包括或排除透视表层由各个透视表的表属性控制。您可以覆盖此设置，包含所有层或排除当前可视层之外的所有层。有关更多信息，请参阅第 107 页的『表属性：打印』主题。

宽透视表。 控制当表相对于定义文档宽度时过宽的处理方法。缺省情况下，表将换行以适合宽度。表将划分为不同的部分，且行标签对每个表部分重复。另外，您还可以缩小宽表，或者不改动宽表，而允许其延伸到定义的文档宽度以外。

包括脚注及文字说明。 控制包含或排除所有透视表脚注及文字说明。

使用查看内略图项作为幻灯片标题。 在导出操作创建的每个幻灯片上包括标题。每个幻灯片都包含从查看器中导出的单个项。标题是根据查看器概要窗格中的项的概要条目构造而成。

模型的视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。有关更多信息，（注：包括表在内的所有模型视图将导出为图形。）

导出页面设置。 这将打开一个对话框，您可以在其中定义导出的文档的页面大小和边距。用于确定换行和缩小行为的文档宽度是页面宽度减去左右边距。

设置 PowerPoint 导出选项

1. 选择 **PowerPoint** 作为导出格式。
2. 单击**更改选项**。

注：导出到 PowerPoint 仅在 Windows 操作系统上可用。

PDF 选项

对于 PDF，可用选项有：

嵌入书签。 此选项提供 PDF 文档中对应于查看器概要条目的书签。与查看器概要窗格一样，通过书签可以更轻松地浏览具有大量输出对象的文档。

内嵌字体。 内嵌字体确保 PDF 文档在所有计算机上外观相同。否则，如果文档中使用的一些字体在用于查看（或打印）PDF 文档的计算机上不可用，则替换字体的效果可能欠佳。

透视表中的层。 缺省情况下，包括或排除透视表层由各个透视表的表属性控制。您可以覆盖此设置，包含所有层或排除当前可视层之外的所有层。有关更多信息，请参阅第 107 页的『表属性：打印』主题。

模型的视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。有关更多信息，（注：包括表在内的所有模型视图将导出为图形。）

设置 PDF 导出选项

1. 选择**可移植文档格式**作为导出格式。
2. 单击**更改选项**。

其他影响 PDF 输出的设置

页面设置/页面属性。 PDF 文档中的页面大小、方向、边距、页眉及页脚的内容和显示以及打印的图表大小均由页面设置选项和页面属性选项控制。

表格属性/表格外观。 宽表和/或长表的比例调整以及表层打印由每个表的表属性控制。这些属性也可保存在表格外观中。

缺省/当前打印机。 PDF 文档的分辨率 (DPI) 是缺省打印机或当前选定的打印机的当前分辨率设置（可以使用“页面设置”进行更改）。最高分辨率为 1200 DPI。如果打印机的设置更高，则 PDF 文档的分辨率将为 1200 DPI。

注意：在低分辨率的打印机上打印高分辨率文档时，打印的效果可能会比较差。

文本选项

设置文本输出选项

对于文本输出，可用选项有：

透视表格式。 透视表可以制表符分隔格式或空格分隔格式导出。对于空格分隔格式，您还可以控制：

- **列宽。** **自动适应**不会对任何列内容换行，每列宽度为该列中最宽标签或值的宽度。**定制**为表中的所有列设置最大列宽，超过该宽度的值将换行，从而在该列的下一行中继续显示。
- **行/列边框字符。** 控制用于创建行和列边框的字符。如果您不希望显示行边框和列边框，请输入空格作为值。

透视表中的层。 缺省情况下，包括或排除透视表层由各个透视表的表属性控制。您可以覆盖此设置，包含所有层或排除当前可视层之外的所有层。有关更多信息，请参阅第 107 页的『表属性：打印』主题。

包括脚注及文字说明。 控制包括或排除所有透视表脚注及文字说明。

模型的视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。有关更多信息，（注：包括表在内的所有模型视图将导出为图形。）

设置文本导出选项

1. 选择**文本**作为导出格式。
2. 单击**更改选项**。

仅图像选项

以下选项仅可用于导出图像：

1. 使“工作簿”选项卡成为活动选项卡（单击选项卡中的任意位置）。
2. 选中要导出的项，单击垂直省略符控件，然后从菜单选择**导出**。您还可以单击工具栏上的**导出按钮**。这样会显示**导出工作簿**对话框。

文档设置

Type

选择**仅图像**作为文档类型。

图像格式

选择已导出图像的图像格式。

PNG

选中时，将以可移植网络图形 (*.png) 格式导出图像。此设置提供了一些选项，用于选择**图像比例 (%)**和**色深**。

JPG

选中时，将以联合摄影专家组 (*.jpg) 格式导出图像。此设置提供了一些选项，用于选择**图像比例 (%)**和**转换为灰度**。

TIFF

选中时，将以标记图像文件格式 (*.tiff) 导出图像。此设置提供用于选择**图像比例 (%)**的选项。

BMP

选中时，将以位图 (*.bmp) 格式导出图像。此设置提供了一些选项，用于选择**图像比例 (%)**和**压缩图像以减小文件大小**。

打开所在文件夹

（可选）选择在导出后打开导出保存文件位置文件夹的设置。

要导出的对象

要包括的对象

选择导出**全部**或**选定**（缺省设置）输出对象。

3. 指定“仅图像”导出设置后，单击**导出**。

图形格式选项

对于 HTML 与文本文档，如果仅导出图表，您可以选择图形格式，并为每种图形格式控制不同可选设置。

要选择导出图表的图形格式和选项：

1. 选择 **HTML**、**文本**或**无（只有图形）** 作为文档类型。
2. 从下拉列表中选择图形文件格式。
3. 单击**更改选项**以更改选定图形文件格式的选项。

JPEG 图表导出选项

图像大小 (%)

原始图表大小的百分比，可达 200%。

转换为灰度

将颜色转换成灰色阴影。

BMP 图表导出选项

图像大小 (%)

原始图表大小的百分比，可达 200%。

压缩图像，以减小文件大小

创建较小的文件而不影响图像质量的无损压缩技术。

PNG 图表导出选项

图像大小 (%)

原始图表大小的百分比，可达 200%。

色深

决定导出图表中的颜色的数量。在任何深度下保存的图表最少的颜色数是实际使用的颜色数，最大则是深度允许的颜色数。例如，如果图表包含三种颜色（红色、白色和黑色），而您将其保存为 16 色，那么图表仍将保持为三种颜色。

注：如果图表中的颜色数超出了该深度的颜色数，则颜色将进行仿色处理，以复制图表中的颜色。

EMF 和 TIFF 图表导出选项

图像大小。 原始图表大小的百分比，最高可达 200%。

注意：EMF（增强型图元文件）格式只在 Windows 操作系统上可用。

EPS 图表导出选项

图像大小。 您可以将大小指定为原图像大小的百分比（可达 200%），或指定图像的像素宽度（高度取决于宽值和宽高比）。导出的图像总是与原始图像成比例。

包括 TIFF 预览图像。 将包含 EPS 图像的预览保存为 TIFF 格式，便于在不能在屏幕上显示 EPS 图像的应用程序中进行显示。

字体。 控制 EPS 图像中的字体处理。

- **使用字体参考。** 如果图表中使用的字体在输出设备上可用，那么将使用这些字体。否则，输出设备将使用备用字体。
- **将字体替换为曲线。** 将字体转换为 PostScript 曲线数据。文本本身不再像可以编辑 EPS 图形的应用程序中的文本那样可编辑。如果图表中使用的字体对输出设备不可用，则该选项十分有用。

查看器打印

有两个选项可用于打印“查看器”窗口的内容：

所有可见输出。 仅打印当前显示在内容窗格中的项。不打印隐藏的项（概要窗格中有合上的书形图标的项，或者折叠的概要层中隐藏的项）。

选择。 仅打印概要窗格和/或内容窗格中当前选中的项。

打印输出和图表

1. 使查看器成为活动窗口（单击该窗口中的任意位置）。
2. 从菜单中选择：
 文件 > 打印...
3. 选择所需的打印设置。
4. 单击**确定**进行打印。

打印预览

“打印预览”显示查看器文档的每一页上将会打印的内容。我们建议您在实际打印查看器文档前查看“打印预览”，这是因为“打印预览”将显示查看器的内容窗格中可能不显示的项，其中包括：

- 分页符
- 透视表隐藏的层
- 宽表中的分界符
- 打印在每一页上的页眉和页脚

如果在查看器中当前选中了任何输出，那么预览将仅显示选中的输出。要查看所有输出的预览，请确保查看器中没有选择任何输出。

页面属性：页眉和页脚

页眉和页脚是打印在每一页的顶部和底部的信息。您可以输入任何文本用作页眉和页脚。还可以使用对话框中间的工具栏插入：

- 日期和时间
- 页码
- 查看器文件名
- 概要标题标签
- 页面标题和子标题
- **设为缺省**使用此处指定的设置作为新查看器文档的缺省设置。（注：此选项会将“页眉/页脚”选项卡和“选项”选项卡上的当前设置变为缺省设置。）
- 概要标题标签表示每一页上的第一项的第一、第二、第三和/或第四层的概要标题。
- 页面标题和子标题打印当前的页面标题和子标题。它们是用查看器“插入”菜单上的“新建页面标题”或 TITLE 和 SUBTITLE 命令创建的。如果没有指定任何页面标题或子标题，则该设置将被忽略。

注意：新页面标题和子标题的字体特征由“选项”对话框（通过选择“编辑”菜单中的“选项”来访问）的“查看器”选项卡控制。现有页面标题和子标题的字体特征可以通过在查看器中编辑标题进行更改。

要查看页眉和页脚在打印页面上的效果，可选择“文件”菜单上的打印预览。

插入页面的页眉和页脚

1. 使查看器成为活动窗口（单击该窗口中的任意位置）。
2. 从菜单中选择：

文件 > 页眉和页脚...

3. 输入要显示在每一页上的页眉和/或页脚。

页面属性：选项

该对话框控制打印图表大小、打印输出项之间的空格和页编号。

- **打印图表尺寸。** 控制相对于定义的页面大小的打印图表的大小。图表的宽高比（宽与高的比）不受打印图表大小的影响。图表的整体打印大小受其高度和宽度的限制。当图表的外边框达到页面的左右边框时，图表大小就无法再扩展以填充更多页面高度。
- **各项之间的间距。** 控制打印的各项之间的间距。每个透视表、图表和文本对象都是一个独立的项。该设置不影响各项在查看器中的显示。
- **起始页码。** 用指定的数字开始按顺序编排页码。
- **设为缺省。** 此选项使用此处指定的设置作为新查看器文档的缺省设置。（请注意，此选项将当前页眉/页脚设置和选项设置设定为缺省值。）

更改打印图表大小、页编号和打印项之间的空格

1. 使查看器成为活动窗口（单击该窗口中的任意位置）。
2. 从菜单中选择：
文件 > 页面属性...
3. 单击**选项**选项卡。
4. 更改设置并单击**确定**。

保存输出

可以保存查看器的内容。

- **输出对象 (*.cou)。** 此格式保存整个输出容器，包括图形、制表符、注释等等。可以在 IBM SPSS Modeler 中打开和查看此格式，将其添加到项目，并使用 IBM SPSS Collaboration and Deployment Services Repository 进行发布和跟踪。此格式与 IBM SPSS Statistics 不兼容。
- **查看器文件 (*.spv)。** 这是用于在查看器窗口中显示文件的格式。在 IBM SPSS Modeler 中从模型块保存为此格式时，仅保存**模型**选项卡中的查看器内容。

要控制用于保存 Web 报告或以其他格式（例如文本、Word、Excel）保存结果的选项，请使用**文件**菜单上的**导出**。

保存查看器文档

1. 从查看器窗口菜单中选择：

文件 > 保存

2. 输入文档的名称，然后单击**保存**。

此外，您还可以进行下列动作：

锁定文件以防止在 IBM SPSS Smartreader 中进行编辑

如果查看器文档已锁定，那么您可以处理透视表（交换行和列、更改显示层等），但无法编辑任何输出，也无法将任何更改保存到 IBM SPSS Smartreader（用于处理查看器文档的单独产品）中的查看器文档。此设置对在 IBM SPSS Statistics 或 IBM SPSS Modeler 中打开的查看器文档没有影响。

使用密码加密文件

您可以通过使用密码加密文件以保护存储在查看器文档的密件信息。一旦加密，文档就只能通过提供密码来打开。要提供密码来打开文件，还需要 IBM SPSS Smartreader 用户。

要加密查看器文档：

- a. 在“将输出另存为”对话框中，选择**使用密码加密文件**。
- b. 单击**保存**。

- c. 在“加密文件”对话框中，提供密码并重新在“确认密码”文本框中输入。密码限制在 10 个字符并区分大小写。

警告：密码丢失后将无法恢复。如果密码丢失，那么将无法打开文件。

创建强密码

- 至少使用八个字符。
- 在密码中使用数字、符号甚至标点符号。
- 避免使用数字序列或字符序列（例如 "123" 和 "abc"）并避免重复，例如 "111aaa"。
- 不要创建使用个人信息（例如生日或昵称）的密码。
- 定期更改密码。

注：不支持将已加密的文件存储到 IBM SPSS Collaboration and Deployment Services Repository 中。

修改加密文件

- 如果打开加密文件，对其进行修改并选择“文件”>“保存”，修改后的文件将以相同的密码保存。
- 您可以通过打开文件、重复加密步骤并在“加密文件”对话框中指定不同的密码，在加密的文件上更改密码。
- 通过打开文件，选择“文件 > 另存为”并在“将数据“另存为”对话框中取消选择**使用密码加密文件**，可以保存加密文件的未加密版本。

注意：在 V21 之前的 IBM SPSS Statistics 版本中，无法打开经过加密的数据文件和输出文档。在 V22 之前的版本中，无法打开经过加密的语法文件。

使用输出文档存储所需的模型信息

只有在输出文档中存在需要辅助信息来启用一些交互式功能的模型查看器项时，此选项才适用。单击**更多信息**将显示这些模型查看器项和需要辅助信息的交互式功能的列表。使用输出文档存储此信息可能会极大增大文档大小。如果您选择不存储此信息，那么仍可以打开这些输出项，但指定的交互式功能将不可用。

透视表

透视表

很多结果都以可交互旋转的表形式显示。也就是说，您可以重新排列行、列和层。

操作透视表

用于操纵数据透视表的选项包括：

- 转置行和列
- 移动行和列
- 创建多维层
- 对行和列进行分组和取消分组
- 显示和隐藏行、列和其他信息
- 旋转行标签和列标签
- 查找术语的定义

激活透视表

在操作或修改透视表之前，您需要**激活表**。要激活表：

1. 双击该表。

或

2. 右键单击该表，然后从弹出菜单中选择**编辑内容**。
3. 从子菜单中选择**在查看器中或在单独窗口中**。

透视表

表具有三个维度：行、列和层。维度中可包含多个元素（或不包含元素）。可以在维度之间或内部移动元素，以更改表的组织。要移动元素，只需将其拖放到所需的位置。

更改元素在维度内的显示顺序

要更改表维度（行、列或层）中元素的显示顺序：

1. 如果透视托盘尚未打开，则从“透视表”菜单中选择：

透视 > 透视托盘

2. 将维度内的元素拖放到透视托盘中。

在维度元素中移动行和列

1. 在表（而非透视托盘）上，单击您想要移动的行或列的标签。
2. 将标签拖到新位置。

转置行和列

如果只需要变换行和列，可以通过透视托盘采用一种简单方法：

1. 从菜单中选择：

透视 > 转置行和列

这与将所有行元素拖到列维度，并将所有列元素拖到行维度中的效果相同。

对行或列分组

1. 选择您想要分组在一起的行或列的标签（单击并拖动，或者按住 Shift 并单击以选择多个标签）。
2. 从菜单中选择：

编辑 > 组

将自动插入一个组标签。双击组标签可编辑标签文本。

注意：要将行或列添加到现有的组，必须先对当前组中的项取消分组。然后创建包含附加项的新组。

对行或列取消分组

取消分组会自动删除组标签。

旋转行标签或列标签

您可以在表中最内部列标签和最外部行标签的水平和垂直显示之间旋转标签。

1. 从菜单中选择：

格式 > 旋转内部列标签

或

格式 > 旋转外部行标签

仅可旋转最内部的列标签和最外部的行标签。

对行排序

排序透视表行：

1. 激活该表。
2. 在列中选择任何用作排序依据的单元格。要仅对一组选定的行进行排序，请从列中选择两个或两个以上要用作排序依据的连续单元格。
3. 从菜单中选择：
编辑 > 对行排序
4. 从子菜单选择**升序**或**降序**。
 - 如果行维度包含组，那么排序仅影响包含所选内容的组。
 - 无法跨组边界进行排序。
 - 无法对行维度中存在多个项的表进行排序。

插入行和列

要在透视表中插入行或列：

1. 激活该表。
 2. 选择表中的任何单元格。
 3. 从菜单中选择：
在前面插入
或
在后面插入
从子菜单中选择：
行
或
列
- 加号 (+) 在新行或新列的每个单元格中插入，以防止新行或新列因为空白而自动隐藏。
 - 在具有嵌套或层维度的表中，会在每个对应的维度插入列和行。

控制变量和值标签的显示。

如果变量包含描述性的变量标签或值标签，那么您可以对透视表中变量名称、变量标签、数据值和值标签的显示进行控制。

1. 激活透视表。
2. 从菜单中选择：
视图 > 变量标签
或
视图 > 值标签
3. 从子菜单选择下列选项之一：
 - **名称或值**。仅显示变量名称（或值）。描述性标签不会显示。
 - **标签**。仅显示描述性标签。不显示变量名称（或值）。
 - **两者**。显示名称（或值）及描述性标签。

更改输出语言

要更改透视表中的输出语言：

1. 激活透视表

2. 从菜单中选择:

查看 > 语言

3. 选择下列可用语言之一:

更改语言只影响应用程序生成的文本，如表标题、行和列标签以及脚注文本。变量名称、描述性变量及值标签不受影响。

导航大型表

要使用导航窗格导航大型表:

1. 激活该表。

2. 从菜单中选择:

视图 > 导航

撤销更改

您可以撤销最近对激活的透视表进行的更改，也可以撤销对其进行的所有更改。这两种操作仅适用于自最近激活该表以来所作的更改。

要撤销最近的更改，请完成下列步骤:

1. 从菜单中选择:

编辑 > 撤销

要撤销所有更改，请完成下列步骤:

2. 从菜单中选择:

编辑 > 恢复

使用层

您可以为每个类别或类别的组合显示一个单独的二维表。可将该类表视为层的堆积，并且只显示最上层。

创建并显示层

要创建层:

1. 如果透视托盘尚未开启，请从**透视托盘**菜单中选择:

透视 > 透视托盘

2. 将一个元素从行维度或列维度拖到层维度。

将元素移至层维度中，会创建一个多维表，但只会显示单个二维分区。所显示的表为最上层的表。例如，如果层维度中有一个“是/否”分类变量，那么多维表将有两个层：一个用于“是”类别，而另一个用于“否”类别。

更改显示的层

1. 从层（在透视表本身中，而非透视表托盘中）的下列列表中选择类别。

转至层类别

“转到层类别”使您可以在透视表中更改层。当存在许多层或选定层包含许多类别时，此对话框尤其有用。

显示和隐藏项目

许多类型的单元格都可以隐藏，其中包括:

- 维度标签
- 类别，包括行或列中的标签单元格和数据单元格

- 类别标签（无需隐藏数据单元格）
- 脚注、标题和标注

隐藏表中的行和列

显示表中的隐藏行和列

1. 从菜单中选择：

视图 > 显示全部类别

这将显示表中的所有隐藏行和列。（如果在此表的“表格属性”中选择了**隐藏空行列**，则完全为空的行或列仍会隐藏。）

隐藏和显示维度标签

1. 在维数中选择维数标签或任何类别标签。
2. 从“查看”菜单或弹出菜单中，选择**隐藏维度标签**或**显示维度标签**。

隐藏和显示表标题

要隐藏标题：

1. 激活透视表。
2. 选择标题。
3. 从“视图”菜单中选择**隐藏**。

要显示隐藏的标题：

4. 从“查看”菜单中选择**全部显示**。

表格外观

表格外观是定义表的外观的一组属性。您可以选择先前定义的表格外观，或创建自己的表格外观。

- 在应用表格外观前后，可以使用单元格属性更改个别单元格或单元格组的单元格格式。即使您应用新的表格外观，编辑的单元格格式仍保持不变。
- 您还可以选择将所有单元格重置为由当前表格外观定义的单元格格式。此选项将重置已编辑的所有单元格。如果在“表格外观文件”列表中选择了**按显示**，则任何已编辑的单元格都将重置为当前的表格属性。
- 仅“表格属性”对话框中定义的表格属性将保存在表格外观中。TableLook 不包括单个单元格修改。

应用表格外观

1. 激活数据透视表。
2. 从菜单中选择：
格式 > 表格外观...
3. 从文件列表表中选择一个表格外观。要从另一个目录中选择文件，请单击**浏览**。
4. 单击**确定**将表格外观应用到选定的透视表。

编辑或创建表格外观

1. 在“表格外观”对话框中，从文件列表表中选择一个表格外观。
2. 单击**编辑外观**。
3. 为需要的属性调整表格属性，然后单击**确定**。
4. 单击**保存外观**以保存已编辑的 TableLook，或者单击**另存为**以将其保存为新的 TableLook。

- 编辑表格外观仅影响选定的透视表。已编辑的表格外观不会应用于任何其他使用该表格外观的表，除非您选择这些表并重新应用表格外观。
- 仅在“表格属性”对话框中定义的表格属性会保存在表格外观中。表格外观不包括单个单元格修改。

表格属性

“表属性”允许您设置表的常规属性，并且允许设置表的各个部分的单元格样式。您可以：

- 控制常规属性，例如隐藏空行或列，以及调整打印属性。
- 控制脚注标记的格式和位置。
- 确定数据区域中的单元格的特定格式、行和列标签的特定格式以及表的其他区域的特定格式。
- 控制形成表的每个区域的边框的线条的宽度和颜色。

更改透视表属性

1. 从菜单中选择：
 格式 > 表属性...
2. 选择一个选项卡（**常规**、**脚注**、**单元格格式**、**边框**或**打印**）。
3. 选择您想要的选项。
4. 单击**确定**或**应用**。

新的属性将应用到选定的透视表。

表属性：常规

多个属性作为一个整体应用到表。您可以执行以下操作：

- 显示或隐藏空的行与列。（空的行或列在任何数据单元格中均不存在任何内容。）
- 控制行标签的位置，它们可以位于左上角或嵌套。
- 控制最大和最小列宽（以磅表示）。

要更改常规表的属性：

1. 单击**常规**选项卡。
2. 选择您想要的选项。
3. 单击**确定**或**应用**。

设置待显示行

注：此功能仅适用于遗存表。

缺省情况下，行数较多的表显示在多个部分（每个部分 100 行）中。控制在表中显示的行数：

1. 选择**按行显示表格**。
2. 单击**设置待显示行**。
 或
3. 从已激活透视表的“视图”菜单中，选择**按行显示表格**和**设置待显示行**。

待显示行。控制一次显示的最大行数。导航控件允许您移动到表格的不同部分。最小值为 10。缺省值为 100。

短行/孤立行容差。控制表格最内层行维度的最大行数，以便在表格的显示视图之间拆分。例如，如果最内层行维度的每个组中有 6 个类别，则指定值 6 可以防止任何组在显示视图之间拆分。该设置可能导致某个显示视图中的总行数超过指定的最大待显示行数。

表属性：注释

“表属性”对话框的“注释”选项卡控制脚注格式编排和表注释文本。

脚注。 脚注标记的属性包括样式和相对于文本的位置。

- 脚注标记的样式为数字（1、2、3...）或字母（a、b、c...）。
- 脚注标记可以作为上标或下标附加在文本上。

注释文本。 您可以向每个表格添加注释文本。

- 如果您在查看器中将光标悬停在表格上，注释文本就会显示在工具提示中。
- 当焦点位于表上时，屏幕阅读器将朗读注释文本。
- 查看器中的工具提示仅显示注释的前 200 个字符，但屏幕阅读器会朗读全部文本。
- 如果将输出导出到 HTML，会使用注释文本作为替代文本。

表属性：单元格格式

表可以按不同的区域设置格式：标题、层、角标签、行标签、列标签、数据、文字说明和脚注。对于表的每个区域，您都可以修改相关联的单元格格式。单元格格式包括文本特征（如字体、大小、颜色以及样式）、水平和垂直对齐、背景色以及内部单元格边距。

将单元格格式应用到区域（信息的类别）。它们不是个别单元格的特征。当对表进行旋转时，此区别是重要的考虑因素。

例如：

- 如果指定粗体作为列标签的单元格格式，那么不管列维度中当前显示的信息是什么，列标签总是显示为粗体。如果您将一项从列维度移至其他维度，它不会保留列标签的粗体特征。
- 如果您仅是通过突出显示已激活的透视表中的单元格，并单击工具栏上的“粗体”按钮而将列标签变为粗体，则不管您将那些单元格移至哪个维度，其内容总是保留粗体，而对于其他移至列维度的项，列标签不会保留粗体特征。

要更改单元格格式：

1. 选择**单元格格式**选项卡。
2. 从下拉列表中选择`一个区域`，或单击样本的一个区域。
3. 选择该区域的特征。所选区域会显示在示例中。
4. 单击**确定**或**应用**。

交替行颜色

将另一种背景和/或文本颜色应用于表的“数据”区域中的交替行：

1. 从“区域”下拉列表中选择**数据**。
2. 选择（选中）背景颜色组中的**交替行颜色**。
3. 选择用于交替行背景和文本的颜色。

交替行颜色仅影响表格的数据区域。它们不会影响行或列标签区域。

表属性：边框

对于表中的每个边框位置，您都可以选择线样式和颜色。如果选择**无**作为线样式，则在选定位置不会出现线条。

要更改表边框：

1. 单击**边框**选项卡。
2. 选择边框位置，方法是在列表中单击边框的名称，或者在“样本”区域中单击线。
3. 选择一种线样式或选择**无**。
4. 选择一种颜色。
5. 单击**确定**或**应用**。

表属性：打印

您可以控制所打印透视表的以下属性：

- 打印表的所有层或仅打印顶层，并在单独页面上打印每一层。
- 水平或垂直地收缩表，以适合页面打印。
- 如果表对于定义的页面大小太宽和/或太长，则通过控制包含在表的任何打印部分的最小行数和列数来控制短行/孤立行。

注意：如果由于在表的上方存在其他输出而导致该表在当前页面中放不下，但是定义的页面长度能够容纳该表，那么将自动在新页面中打印该表，而不考虑短行/孤立行设置。

- 包含无法容纳在单页中的表的接续文本。您可以在每页的底部和顶部显示接续文本。如果未选择任一选项，则不会显示接续文本。

要控制透视表打印属性：

1. 单击**打印**选项卡。
2. 选择您想要的打印选项。
3. 单击**确定**或**应用**。

单元格属性

单元格属性会应用到选定的单元格。您可以更改字体、值的格式、对齐方式、边距和颜色。单元格属性会覆盖表格属性；所以，更改表格属性不会影响到任何单独应用的单元格属性。

要更改单元格属性：

1. 选择表中的单元格。
2. 从“格式”菜单或弹出菜单中选择**单元格属性**。

字体和背景

“字体及背景”选项卡控制表中选定单元格的字体样式、颜色和背景颜色。

格式值

“格式值”选项卡控制选定单元格的值格式。您可以选择数字、日期、时间或货币的格式，并且可以调整显示的小数位。

对齐和页边距

“对齐与页边距”选项卡控制选定单元格的值的水平和垂直对齐方式，以及上下左右边距。**混合水平对齐**按照单元格的类型对齐其内容。例如，日期为右对齐，而文本值为左对齐。

脚注和文字说明

您可以向表添加脚注和文字说明。还可以隐藏脚注或标注，更改脚注标记以及重新对脚注编号。

添加脚注和文字说明

要向表添加文字说明：

1. 从“插入”菜单中选择**文字说明**。

可以将脚注附加到表中的任意项。要添加脚注：

1. 在激活的透视表中，单击标题、单元格或标注。
2. 从“插入”菜单中选择**脚注**。
3. 在提供的区域中插入脚注文本。

隐藏或显示文字说明

要隐藏文字说明：

1. 选择文字说明。
2. 从“视图”菜单中选择**隐藏**。

要显示隐藏的文字说明：

1. 从“查看”菜单中选择**全部显示**。

隐藏或显示表中的脚注

要隐藏脚注：

1. 右键单击包含脚注参考的单元格，然后从弹出菜单中选择**隐藏脚注**。

或

2. 从表的脚注区域中选择脚注，并从弹出菜单中选择**隐藏**。

注意：对于旧表，选择表的脚注区域，并从弹出菜单中选择**编辑脚注**，然后取消选择（取消选中）您要隐藏的任何脚注的“可见”属性。

如果单元格包含多个脚注，则使用后一种方法选择隐藏脚注。

要隐藏表中的所有脚注，请完成下列步骤：

1. 在表的脚注区域中选择所有脚注（通过单击并拖动或者在按住 Shift 键的情况下单击来选择脚注），然后从“查看”菜单中选择**隐藏**。

要显示隐藏的脚注：

1. 从“视图”菜单中选择**显示全部脚注**。

脚注标记

“脚注标记”更改可用于对脚注进行标记的字符。缺省情况下，标准脚注标记为顺序字母或数字，具体取决于表属性设置。您也可以指定特殊标记。在您对脚注重新编号，或者切换用于标准标记的数字和字母时，不会影响特殊标记。标准标记的数字或字母显示以及脚注标记的下标或下标位置由“表属性”对话框的“脚注”选项卡控制。

要更改脚注标记：

1. 选择脚注。
2. 从**格式**菜单中选择**脚注标记**。

特殊标记限制为 2 个字符。在表的脚注区域中，带有特殊标记的脚注位于具有顺序字母或数字的字母或数字之前；因此更改特殊标记可能会对脚注列表进行重新排序。

对脚注重新编号

当通过交换行、列和层对表进行旋转之后，可能会打乱脚注的顺序。要对脚注重新编号：

1. 从“格式”菜单中选择**对脚注重新编号**。

编辑遗存表中的脚注

对于遗存表，您可以使用“编辑脚注”对话框来输入和修改脚注文本和字体设置，更改脚注标记，以及选择隐藏或删除脚注等。

当您在遗存表中插入新的脚注时，“编辑脚注”对话框将自动打开。要使用“编辑脚注”对话框来编辑现有脚注（而不创建新脚注），请完成下列步骤：

标记。缺省情况下，标准脚注标记为顺序字母或数字，具体取决于表属性设置。要指定特殊标记，只需在“标记”列中输入新的标记值。在您对脚注重新编号，或者切换用于标准标记的数字和字母时，不会影响特殊

标记。标准标记的数字或字母显示以及脚注标记的下标或下标位置由“表属性”对话框的“脚注”选项卡控制。请参阅第 105 页的『表属性：注释』主题以获取更多信息。

要将特殊标记改回标准标记，请右键单击“编辑脚注”对话框中的标记，并从弹出菜单中选择**脚注标记**，然后在“脚注标记”对话框中选择“标准”标记。

脚注。脚注内容。其显示反映当前字体和背景设置。您可以使用“格式”子对话框来更改各个脚注的字体设置。请参阅第 109 页的『脚注字体和颜色设置』主题以获取更多信息。将对所有脚注应用单一背景色，您可以在“单元格属性”对话框的“字体和背景”选项卡中更改该背景色。有关更多信息，请参阅第 107 页的『字体和背景』主题。

可视。所有脚注缺省情况下可见。取消选择（取消选中）“可见”复选框将隐藏脚注。

脚注字体和颜色设置

对于遗存表，您可以使用“格式”对话框来更改一个或多个选定脚注的字体系列、样式、大小和颜色：

1. 在“编辑脚注”对话框的“脚注”网格中，选择（单击）一个或多个脚注。
2. 单击**格式**按钮。

选定的字体系列、样式、大小和颜色将应用于所有选定的脚注。

背景颜色、对齐和边距可以在“单元格属性”对话框中设置，并应用于所有脚注。您无法对各单独脚注更改这些设置。有关更多信息，请参阅第 107 页的『字体和背景』主题。

数据单元格宽度

“设置数据单元格宽度”用于将所有数据单元格设置为相同的宽度。

要为所有数据单元格设置宽度：

1. 从菜单中选择：
格式 > 设置数据单元格宽度...
2. 输入单元格宽度的值。

更改列宽

1. 单击并拖动列边框。

显示透视表中的隐藏边框

对于不含很多可视边框的表，可以显示隐藏边框。这可以简化更改列宽之类的任务。

1. 从“视图”菜单中选择**网格线**。

在透视表中选择行、列和单元格

您可以选择整行或整列，或指定的数据和标签单元格组。

要选择多个单元格：

选择 > 数据和标签单元格

打印透视表

有多个因素会影响打印透视表外观的方式，可以通过更改透视表属性控制这些因素。

- 对于多维透视表（带有多层的表），您可以打印所有层，也可以只打印顶层（可视层）。有关更多信息，请参阅第 107 页的『表属性：打印』主题。
- 对于长或宽的透视表，可以自动调整表的大小以适合页面，也可以控制表分隔符和分页符的位置。有关更多信息，请参阅第 107 页的『表属性：打印』主题。

可以使用“文件”菜单上的“打印预览”查看打印的透视表的外观。

控制宽表和长表的表分隔符

有的透视表太宽或太长，无法在定义的页面大小中打印，程序会自动分割它们，并将它们打印在多个部分中。您可以执行以下操作：

- 控制拆分大型表的行或列的位置。
- 指定分割表时应该保留在一起的行和列。
- 重定大型表的比例以适应定义的页面大小。

为打印的透视表指定行和列分隔符

1. 激活透视表。
2. 单击要插入分隔符位置左边的列中的任何单元格，或者单击要插入分隔符的行之前的行中的任何单元格。

3. 从菜单中选择：

格式 > 分界点 > 垂直分界点

或

格式 > 分界点 > 水平分界点

1. 激活透视表。
2. 单击要插入分隔符位置左边的列中的任何单元格，或者单击要插入分隔符的行之前的行中的任何单元格。

3. 从菜单中选择：

格式 > 分界点 > 垂直分界点

或

格式 > 分界点 > 水平分界点

要指定要保留在一起的行或列：

1. 选择要保留在一起的行或列的标签。单击并拖动或按住 Shift 键的同时单击来选择多个行或列标签。
2. 从菜单中选择：

格式 > 分界点 > 保持在一起

查看分界点和“保留在一起”分组：

1. 从菜单中选择：

格式 > 分界点 > 显示分界点

分界点显示为垂直线或水平线。“保留在一起”分组显示为采用深色边框围起来的灰色矩形区域。

注意：遗存表不支持显示分界点和“保留在一起”分组。

清除分界点和“保留在一起”分组

清除分界点：

1. 单击垂直分界点左侧列中的任何单元格，或单击垂直分界点上方行中的任何单元格。
2. 从菜单中选择：

格式 > 分界点 > 清除分界点或组

清除“保留在一起”分组：

3. 选择用于指定分组的列或行标签。

4. 从菜单中选择：

格式 > 分界点 > 清除分界点或组

在您旋转或重新排序任何行或列时，所有分界点和“保留在一起”分组将自动清除。此行为不适用于遗存表。

从透视表创建图表

1. 双击透视表以将其激活。
2. 选择想要在图中显示的行、列或单元格。
3. 右键单击选定区域中的任意位置。
4. 从弹出菜单中选择**创建图形**并选择一个图表类型。

遗存表

您可以选择将表作为遗存表来呈现（称为版本 19 中的全功能表），它们与 20 之前的版本 IBM SPSS Statistics 完全兼容。遗存表呈现可能较为缓慢，仅在需要与 20 之前的版本兼容时推荐使用。有关如何创建遗存表的信息，请参阅第 112 页的『透视表选项』。

选项

选项

选项控制各种设置。

更改选项设置

1. 从菜单中选择：
 编辑 > 选项...
2. 单击要更改的设置所在的选项卡。
3. 保存设置。
4. 单击**确定**或**应用**。

一般选项

最大线程数

这是计算结果时多线程过程使用的线程数。**自动**设置基于可用处理核心数。多线程过程运行时，如果您要使更多处理资源可供其他应用程序使用，请指定较小的值。在分布式分析模式下，此选项处于禁用状态。

输出

显示小数值的前导零。对于仅包含小数部分的数字值，显示前导零。例如，显示前导零时，值 .123 将显示为 0.123。此设置不适用于具有货币或百分比格式的数字值。除了固定 ASCII 文件 (*.dat) 之外，数据保存到外部文件中时，将不包括前导零。

测量系统。用于指定属性（如透视表单元格边距、单元格宽度和打印时表之间的空白）的测量系统（点、英寸或厘米）。

查看器选项

查看器输出显示选项仅影响更改设置后生成的新输出。已经显示在查看器中的输出不受这些设置更改的影响。

初始输出状态

控制每次运行过程时自动显示或隐藏的项，以及各项的初始对齐方式。您可以控制下列项目的显示：日志、警告、注释、标题、透视表、图表、树图和文本输出。您还可以在登录或注销时显示命令。可以从日志复制命令语法，并将其保存在语法文件中。

注：在查看器中所有输出项都按照左对齐方式显示。只有打印输出的对齐方式受对齐方式设置的影响。居中和右对齐项用一个小符号表示。

标题

控制新输出标题的字体样式、大小和颜色。字体大小列表提供了一组预定义的大小，但您可以手动输入其他支持的大小值。

页面标题

控制新页面标题以及由 TITLE 和 SUBTITLE 命令语法生成的或由插入菜单上的新页面标题创建的页面标题的字体样式、大小和颜色。字体大小列表提供了一组预定义的大小，但您可以手动输入其他支持的大小值。

文本输出(O)

用于文本输出的字体。文本输出设计成与等宽（固定间距）字体一起使用。如果选择比例字体，表格输出将不会正确对齐。字体大小列表提供了一组预定义的大小，但您可以手动输入其他支持的大小值。

缺省页面设置

控制用于打印的方向和页边距缺省选项。

透视表选项

“透视表”选项可设置与透视表显示有关的各种选项。

表格外观

从文件列表选择一个表格外观，并单击**确定**或**应用**。可以使用随 IBM SPSS Statistics 一起提供的表格外观，或者在透视表编辑器（从“格式”菜单选择**表格外观**）中创建自己的表格外观。

- **浏览**。使您可以从另一个目录中选择表格外观。
- **设置表格外观目录**。使您可以更改缺省的表格外观目录。使用**浏览**导航到您要使用的目录，选择该目录中的“表格外观”，然后选择**设置表格外观目录**。

注：以 IBM SPSS Statistics 的早期版本创建的表格外观不能用于 16.0 或更高版本。

调整列宽

这些选项控制透视表中列宽的自动调整。

- **仅针对标签进行调整**。将列宽调整为列标签的宽度。这会生成结构更紧凑的表，但宽度超过标签的数据值可能会被截去。
- **针对所有表的标签和数据进行调整**。将列宽调整为列标签的宽度或最大数据值的宽度中的较大者。这会生成较宽的表，但可以保证显示所有的值。

缺省编辑模式

此选项控制透视表在查看器窗口或独立窗口中的激活。缺省情况下，双击透视表可激活查看器窗口中的所有除了特大表之外的表格。可以选择在独立窗口中激活透视表，或者选择一个大小设置，使小于这个设置的透视表在查看器中打开，而大于这个设置的透视表在独立的窗口中打开。

以丰富文本格式复制宽表格至剪贴板

当以 Word/RTF 格式粘贴透视表时，文档宽度过宽的表格会被换行、按比例缩小以适合文档宽度或不进行更改。

输出选项

输出选项控制某些输出选项的缺省设置。

屏幕朗读者辅助功能。控制屏幕朗读者如何朗读透视表行标签和列标签。可以朗读每个数据单元格的行标签和列标签，也可以只朗读在表中数据单元格之间移动时更改的标签。

第 8 章 处理缺失值

缺失值概述

在数据挖掘的数据准备阶段，通常需要替换数据中的缺失值。**缺失值**是数据集中未知、未收集或输入不正确的值。通常，这些值不可用于字段中。例如，性别字段应该包含值 *M* 和 *F*。如果您在字段中发现值 *Y* 或 *Z*，那么可以肯定此类值无效，因此应解释为空白。同样地，年龄字段出现负值也毫无意义，应将其解释为空白值。此类明显错误通常是由于问卷过程中人为输入或保留字段为空以示拒绝回答造成的。有时候，您可能会进一步检查这些空白字段，以弄清拒绝提供本人年龄等行为是否会影响具体预测结果。

某些建模技术在处理缺失值方面具有明显的优势。例如，C5.0 和 Apriori 可以很好地处理在类型节点中明确声明为“缺失”的值。其它建模技术在处理缺失值时比较麻烦，并且需要较长的培训时间，且生成的模型不够精确。

IBM SPSS Modeler 可识别的缺失值类型有以下几种：

- **Null 或系统缺失值。**这两种类型是数据库或源文件中留空、并且尚未在源节点或类型节点中专门定义为“缺失”的非字符串值。系统缺失值会显示为 **\$null\$**。请注意，空字符串在 IBM SPSS Modeler 中不会被视为 Null，但它们可能会被某些数据库视为 Null。
- **空字符串和空白。**空字符串值和空白（带有不可见字符的字符串）不被视为空值。对于大多数用途，空字符串都视为相当于空白。例如，如果您选择在源节点或类型节点中将空白视为空白值的选项，则此设置也应用于空字符串。
- **空白或用户定义的缺失值。**这些是在源节点或类型节点中被明确定义为缺失的值（例如 unknown、99 或 -1）。您还可以将空和空白视为空白值，这样将使得它们被标记为进行特殊处理并排除在大多数计算之外。例如，您可以使用 @BLANK 函数将这些值与其他类型的缺失值一起视为空白值。

读入混合数据。注意，当您读取以数字形式（整数、实数、时间、时间戳或日期）存储的字段时，所有非数值型字段都将设置为空值或系统缺失值。这是因为 Clementine 与其他应用程序不同，它不允许使用混合存储类型的字段。为避免发生这种情况，可以根据需要更改源节点或外部应用程序中的字段存储类型，以字符串的形式读取包含混合数据的字段。

从 Oracle 读取空字符串。在 Oracle 数据库中进行值的读写时，要注意，与 IBM SPSS Modeler 及大多数其他数据库不同，Oracle 将字符型空值等同于空值对待并存储。这表示同样的数据从 Oracle 数据库中提取和从文件或其他数据库中提取其表现可能有所不同，可能会返回不同的结果。

处理缺失值

您应根据自己所从事的业务或经营领域常识来确定如何处理缺失值。为了减少培训时间和提高精确度，可能需要移除数据集中的空白值。从另一方面讲，空白值的出现还可能会带来新的业务机会或其它灵感。选择最佳方法时，应考虑数据的以下几个方面：

- 数据集的大小
- 包含空白值的字段数
- 缺失信息量

通常有两种方法可供选择：

- 可以排除带有缺失值的字段或记录
- 可以使用各种方法归因、替换或强制缺失值

使用数据审核节点可以在很大程度上实现上述两种方法的自动化。例如，可以生成过滤节点（该节点将具有过多缺失值的字段排除，不在建模中使用），并生成为任何或所有保留字段归因缺失值的超节点。审核的真正强大功能就在此时得以体现，您不仅可以评估数据的当前状态，还可以基于评估结果执行操作。

处理带有缺失值的记录

如果大部分缺失值都集中在少量记录中，您只需排除这些记录。例如，银行通常会保存详细而完整的贷款客户记录。但是，如果银行在审批内部职员的贷款时管制不严，则收集到的员工贷款数据中可能会存在多个空白字段。此种情况下，有两种方法可以处理缺失值：

- 可以使用选择节点删除员工记录。
- 如果数据集较大，可以放弃所有带有空白值的记录。

处理带有缺失值的字段

如果大部分缺失值都集中在少量字段中，您可以通过字段而不是记录查找这些缺失值。此方法还允许您先检验特定字段相对于建模的重要性，然后确定如何处理缺失值。如果某个字段对于建模而言并不重要，那么可能不值得保留，而与其中包含的缺失值数目无关。

例如，某市场调查公司可能会从包含 50 个问题的普通问卷中收集数据。很多人拒绝提供年龄和政治派别信息。此种情况下，年龄和政治派别就会有大量缺失值。

字段测量级别

确定要采用的方法时，您还应考虑带有缺失值的字段的测量级别。

数字字段。对于数值字段类型（如连续），您应在构建模型前清除所有非数字值，因为如果数值字段中包含空白值，很多模型将无效。

分类字段。对于分类字段（如名义和标志），虽然不必更改缺失值，但更改后可以提高模型的精度。例如，使用性别字段的模型即使含有无意义值（如 Y 和 Z）也仍然有效，不过如果删除除 M 和 F 以外的值将提高模型的精度。

筛选或删除字段

要筛选带有大量缺失值的字段，您可以采用以下几种方法：

- 使用数据审核节点根据质量过滤字段。
- 您可以使用特征选择节点来筛选缺失值超过指定百分比的字段，并根据相对于特定目标的重要性来对字段进行排序。
- 除删除字段以外，还可以使用“类型”节点将字段角色设置为无。这会将这些字段保留在数据集中，但将它们排除在建模流程之外。

处理具有缺失值的记录

什么是系统缺失值？

系统缺失值表示未知或不适用的数据值。在数据库中，这些值通常称为 *NULL* 值。

系统缺失值不同于空白值。空白值通常在“类型”节点中定义为特定值或值的范围，可以将其视为用户定义的缺失。空白值在建模上下文中以不同方式处理。

构造系统缺失值

在从数据源读取的数据中可能存在系统缺失值（例如，数据库表可能包含 *NULL* 值）。

可以通过在表达式中使用值 **undef** 来构造系统缺失值。例如，以下 CLEM 表达式返回 Age（如果小于或等于 30）或缺失值（如果大于 30）：

```
if Age > 30 then undef else Age endif
```

在以下情况下也可创建缺失值：执行外连接时，数字除以零时，计算负数的平方根时，以及其他情况。

显示系统缺失值

系统缺失值在表和其他输出中显示为 \$null\$。

测试系统缺失值

如果自变量值是系统缺失值，请使用特殊函数 `@NULL` 返回 `true`，例如：

```
if @NULL(MyFieldName) then 'It is null' else 'It is not null' endif
```

传递到函数的系统缺失值

传递到函数的系统缺失值通常将缺失值传播到输出。例如，如果字段 `f1` 的值在特定行中是系统缺失值，那么表达式 `log(f1)` 也求值为该行的系统缺失值。`@NULL` 函数除外。

涉及算术运算符的表达式中的系统缺失值

将算术运算符应用于包含系统缺失值的值将产生系统缺失值。例如，如果特定行中字段 `f1` 的值是系统缺失值，那么表达式 `f1 + 10` 也会求值为该行的系统缺失值。

涉及逻辑运算符的表达式中的系统缺失值

处理涉及逻辑运算符的表达式中的系统缺失值时，三值逻辑规则 (`true`、`false` 和 `missing`) 适用并可在事实表中进行描述。下表中显示了常见逻辑运算符 `not`、`and` 和 `or` 的事实表。

操作数	NOT 操作数
true	false
false	true
缺少	缺少

Operand1	Operand2	操作数 1 AND 操作数 2
true	true	true
true	false	false
true	缺少	缺少
false	true	false
false	false	false
false	缺少	false
缺少	true	缺少
缺少	false	false
缺少	缺少	缺少

Operand1	Operand2	操作数 1 OR 操作数 2
true	true	true
true	false	true
true	缺少	true
false	true	true

表 7: OR 事实表 (继续)

Operand1	Operand2	操作数 1 OR 操作数 2
false	false	false
false	缺少	缺少
缺少	true	true
缺少	false	缺少
缺少	缺少	缺少

涉及比较运算符的表达式中的系统缺失值

比较系统缺失值和非系统缺失值时，结果求值为系统缺失值而非 true 或 false 结果。系统缺失值可以相互比较；两个系统缺失值视为相等。

if/then/else/endif 表达式中的系统缺失值

使用条件表达式并且条件表达式返回系统缺失值时，else 子句中的值从条件表达式返回。

“选择”节点中的系统缺失值

如果针对特定记录，选择表达式求值为缺失值，那么记录不是从“选择”节点输出（此操作适用于“包含”和“废弃”方式）。

“合并”节点中的系统缺失值

使用关键字进行合并时，不会合并在关键字段中具有系统缺失值的任何记录。

汇总中的系统缺失值

汇总列中的数据时，在计算中不包含缺失值。例如，在具有三个值 { 1、2 和 undef } 的列中，列中值的总和计算为 3；平均值计算为 1.5。

归因或填充缺失值

在仅有几个缺失值的情况下，可以用插入值替换空白值。可以在数据审核报告中实现上述操作，在此报告中您可以为特定字段指定相应选项，然后生成一个超节点采用多种方法对值进行归因。这种方法最为灵活，还可以指定在单个节点中处理大量字段。

下列方法可用于输入缺失值：

已修正。 替换为固定值（可以字段平均值、范围中间值，或者您指定的常数）。

随机。 替换为基于正态分布或均匀分布产生的随机值。

表达式。 用于指定定制表达式。例如，您可以使用设置全局量节点创建的全局变量替换值。

算法。 基于 C&RT 算法替换为模型预测的值。对于使用此方法输入的每个字段，都会有一个单独的 C&RT 模型，还有一个填充节点会使用该模型预测的值替换空白值和空值。然后使用过滤节点删除该模型生成的预测字段。

如果还要为特定字段强制赋值，则可以使用类型节点来确保字段类型仅包含合法值，然后将需要替换空白值字段的检查列设置为 **强制**。

用于缺失值的 CLEM 函数

有多个函数可用于处理缺失值。选择节点和填充节点中经常会用以下函数来放弃或填充缺失值：

- `count_nulls(LIST)`
- `@BLANK(FIELD)`
- `@NULL(FIELD)`
- `undef`

@ 函数可以与 @FIELD 函数一起使用来识别一个或多个字段中是否存在空白值或空值。当出现空白值或空值时，一般会对此类字段进行标记，也可以用替换值填充或者在各种其它操作中使用此类字段。

如下所示，您可以计算字段列表中的空值的数量：

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

在使用接受字段列表作为输入的任何函数时，可以使用特殊函数 @FIELDS_BETWEEN 和 @FIELDS_MATCHING，如以下示例中所示：

```
count_nulls(@FIELDS_MATCHING('card*'))
```

您可以通过 undef 函数用显示为 **\$null\$** 的系统缺失值来填充字段。例如，要替换任何数字值，可以使用条件语句，例如：

```
if not(Age > 17) or not(Age < 66) then undef else Age endif
```

这会将不在范围内的任何内容替换为系统缺失值（显示为 **\$null\$**）。通过使用 not() 函数，您可以捕获所有其他数字值，包括任何负数。有关更多信息，请参阅主题 [第 159 页的『处理空白值和空值的函数』](#)。

放弃记录时的注意事项

请注意，使用选择节点放弃记录时，语法将使用三值逻辑并在选择语句中自动包含空值。要在选择表达式中排除 Null 值（系统缺失值），必须在表达式中使用 and not 来显式指定此值。例如，要选择并包含处方药类型为药品 C 的所有记录，您可以使用以下选择语句：

```
Drug = 'drugC' and not(@NULL(Drug))
```

早期版本遇到上述情况时会排除空值。

第 9 章 构建 CLEM 表达式

关于 CLEM

表达式操作控制语言 (CLEM) 是一种用于分析和操纵在 IBM SPSS Modeler 流中流动的数据的功能强大的语言。数据挖掘人员可在流操作中广泛使用 CLEM 语言来执行根据成本和收入数据推导利润这样的简单任务,也可以执行将 Web 日志数据转换为具有有用信息的一系列字段和记录这样的复杂任务。

在 IBM SPSS Modeler 中, CLEM 用于:

- 对加载记录字段上的条件进行比较和评估。
- 为生成新字段导出值。
- 为现有字段导出新值。
- 合理安排记录序列。
- 将记录中的数据插入到报告中。

CLEM 表达式对于 IBM SPSS Modeler 中的数据准备是不可或缺的,此外它还广泛适用于从记录和字段操作(选择、平衡和填充)到散点图和输出(分析、报告和表)的一系列节点。例如,您可以在导出节点中使用 CLEM 来根据公式(如率)创建一个新字段。

CLEM 表达式还可用于全局搜索和替换操作。例如,可以在“填充”节点中使用表达式 `@NULL(@FIELD)` 将系统缺失值替换为整数 0。(要替换用户缺失值(也称为空白),请使用 `@BLANK` 函数。)

还可以创建更多较为复杂的 CLEM 表达式。例如,您可以根据一组条件规则来派生新字段,如通过使用以下表达式创建的新值类别: `If: CardID = @OFFSET(CardID,1), Then: @OFFSET(ValueCategory,1), Else: 'exclude'`。

此示例使用 `@OFFSET` 函数来表达如下的意思:如果某一给定记录的 `CardID` 字段值和先前记录的该字段值相同,则返回先前记录的名为 `ValueCategory` 字段的值。否则,将为其赋予字符串“exclude”。换言之,如果邻近记录的 `CardID` 值相同,就应该赋予它们相同的 `ValueCategory`。(带有字符串“exclude”的记录可在稍后通过选择节点调用。)

CLEM 示例

为阐释使用 CLEM 的正确语法和可能的表达式类型,我们给出了表达式的示例。

简单表达式

下面这个公式可根据字段 `After` 和 `Before` 的值导出一个新字段,您的公式也可以像这个公式一样简单。

```
(After - Before) / Before * 100.0
```

注意:在引用字段值的时候不需要给字段名加引号。

类似的,如下的表达式只是简单地返回字段 `salary` 的对数值。

```
log(salary)
```

复杂表达式

表达式也可以很长而且非常复杂。如果两个字段 (`$KX-Kohonen` 和 `$KY-Kohonen`) 的值落在了指定范围之外,如下的表达式将返回 `true`。注意:由于此处的字段名包含有特殊字符,因此给它们加上了单引号。

```
(' $KX-Kohonen ' >= -0.2635771036148072 and ' $KX-Kohonen ' <= 0.3146203637123107 and ' $KY-Kohonen ' >= -0.18975617885589602 and ' $KY-Kohonen ' <= 0.17674794197082522) -> T
```

少数几个函数（如字符串函数）需要使用正确的语法键入几个参数值。下面例子中使用的函数 `substrs` 将返回 `produce_ID` 字段的第一个字符，用以指明该物品究竟是属于有机、转基因还是常规。表达式的结果将由 `-> `result`` 来描述。

```
substrs(1,produce_ID) -> `c`
```

类似的表达式有：

```
stripchar(`3`,`123`) -> `12`
```

需要注意的是字符要用反单引号括起来。

表达式中的组合函数

通常，CLEM 表达式是由组合函数构成的。下列函数将 `substr` 和 `lowertoupper` 结合使用，以返回 `produce_ID` 的第一个字符并将其转换为大写。

```
lowertoupper(substr(1,produce_ID)) -> `C`
```

同样的表达式还可以简写为：

```
lowertoupper(produce_ID(1)) -> `C`
```

另一个经常使用的函数组合是：

```
locchar_back(`n`, (length(web_page)), web_page)
```

该表达式用以确定从 `web_page` 字段值的最后一个字符往前数起字符 ``n`` 所在的位置。通过引入 `length` 函数，表达式可动态计算当前值的长度从而避免了使用可能无效的静态数字（以 7 为例，如果值的长度小于 7 个字符，将导致无效）。

特殊函数

有许多可用的特殊函数（函数名前加有符号 `@`）。常用的函数包括：

```
@BLANK('referrer ID') -> T
```

通常，特殊函数通过组合的方式加以使用，这种方式常用于一次性标记多个字段中的空白值。

```
@BLANK(@FIELD) -> T
```

CLEM 文档中还将讨论更多的示例。有关更多信息，请参阅主题 [第 133 页的『CLEM 参考概述』](#)。

值和数据类型

CLEM 表达式类似于由值、字段名、运算符和函数构成的公式。最简单的有效 CLEM 表达式是一个值或一个字段名。有效值的示例如下：

```
3  
1.79  
'banana'
```

字段名的示例如下：

```
Product_ID  
'$P-NextField'
```

其中，`Product` 是市场购物篮数据集中的字段名，`'$P-NextField'` 是一个参数名，而表达式的值则是一个指定字段的值。通常，字段名称以字母开头，还可能包含数字和下划线 (`_`)。如果将名称括在引号内，那么可以使用不遵循这些规则的名称。CLEM 值可以是如下所列的任何一种类型：

- 字符串 - 例如, "c1"、"Type 2"、"a piece of free text"
- 整数 - 例如 12、0 和 -189
- 实数 - 例如 12.34、0.0 和 -0.0045
- 日期/时间字段 - 例如, 05/12/2002、12/05/2002、12/05/02

还有可能使用到下列元素:

- 字符代码 - 例如 `a` or 3
- 项的列表 - 例如, [1 2 3]、['Type 1' 'Type 2']

字符代码和列表一般不作为字段值出现。通常, 它们用作 CLEM 函数的参数。

引用规则

虽然软件在确定 CLEM 表达式中某部分是字段、值、参数还是字符串时非常灵活, 但在创建表达式时遵循以下的一般规则将十分有利于您的实践操作:

- **字符串** - 在编写字符串时始终使用双引号 ("Type 2" 或 "value")。也可以使用单引号, 但容易与带引号字段混淆。
- **字符** - 始终使用反单引号, 例如 `。例如, 请注意函数 `stripchar(`d`,`drugA`)` 中的字符 d。此类情形中的唯一例外是使用整数来引用字符串中的特定字符。例如, 请注意函数 `lowertoupper("drugA"(5))` -> "A" 中的字符 5。注: 在标准的 U.K. 和 U.S. 键盘上, 反引号字符 (重音符, Unicode 0060) 键就在 Esc 键的正下方。
- **字段** - 在 CLEM 表达式 (`subscr(2,arrayID)`) -> CHAR) 中使用, 字段通常未加引号。必要时可以使用单引号将空格或其他特殊字符括起来 ('Order Number')。数据集中加引号但未定义的字段将会被误读为字符串。
- **参数** - 始终使用单引号 ('\$P-threshold')。

表达式和条件

CLEM 表达式可以返回一个结果 (在导出新值时使用) - 例如:

```
Weight * 2.2
Age + 1
sqrt(Signal-Echo)
```

或者, 它们也可判断 *true* 或 *false* (根据某一条件进行选择时使用) - 例如:

```
Drug = "drugA"
Age < 16
not(PowerFlux) and Power > 2000
```

您可以在 CLEM 表达式中任意组合使用运算符和函数 - 例如:

```
sqrt(abs(Signal)) * max(T1, T2) + Baseline
```

括号和运算符优先级将确定表达式求值的顺序。本例中, 求值的顺序为:

- 首先计算 `abs(Signal)`, 随后再将 `sqrt` 应用于上一步计算结果。
- 计算 `max(T1, T2)`。
- 上述两个结果相乘: **x** 的优先级高于 +。
- 最后, 将 `Baseline` 添加到结果中。

优先级按降序排列 (即, 从第一个执行的运算到最后一个运算) 如下:

- 函数参数
- 函数调用
- **xx**

- `x / mod div rem`
- `+ -`
- `> < >= <= /== == = /=`

如果您希望覆盖优先级，或者您对求值顺序不确定，利用圆括号可使求值顺序固定下来 - 例如：

```
sqrt(abs(Signal)) * (max(T1, T2) + Baseline)
```

流、会话和超节点参数

可以定义用于 CLEM 表达式和脚本编写的参数。这些参数在效果上当同于用户定义的变量，这些变量可以在当前流、会话或超节点中保存和保留，并且可以从用户界面和脚本中访问。例如，如果保存流，则为此流设置的所有参数也将保存。（此特点使这些参数区别于仅在脚本中使用和声明的本地脚本变量。）通常将参数用于脚本编制来控制脚本的行为，可通过提供在脚本中无需进行硬编码的字段和值的相关信息。

参数的范围取决于设置参数的位置：

- 流参数可以在流脚本中或在流属性对话框中设置，且可用于流中的所有节点。这些参数显示在表达式构建器中的“参数”列表中。
- 会话参数可以在独立脚本中或在会话参数对话框中设置。这些会话参数可用于当前会话中使用的所有流（即在管理器窗格的“流”选项卡中列出的所有流）。

也可以为超节点设置参数，在这种情况下，这些参数仅对于该超节点中封装的节点可见。

在 CLEM 表达式中使用参数

CLEM 表达式中的参数由 `$P-pname` 表示，其中 `pname` 是参数名。在 CLEM 表达式中，使用的参数必须用单引号括起来 - 例如 `'$P-scale'`。

可用的参数可直接通过表达式构建器来查看。查看当前的参数：

1. 在任何可使用 CLEM 表达式的对话框中，单击“表达式构建器”按钮。
2. 从“字段”列表中，选择**参数**。

从列表中选择要插入到 CLEM 表达式的参数。有关更多信息，请参阅主题 [第 129 页的『选择字段、参数和全局变量』](#)。

字符串的处理

有很多可用于字符串的运算，包括：

- 将字符串转换为大写或小写 - `uppertolower(CHAR)`。
- 消除某个字符串变量中指定的字符（例如 ``ID_`` 或 ``$``） - `stripchar(CHAR, STRING)`。
- 确定字符串变量的长度（字符数） - `length(STRING)`。
- 检查字符串值的字母顺序 - `alphabefore(STRING1, STRING2)`。
- 移除值中的前端或尾部空格 - `trim(STRING)`、`trim_start(STRING)` 或 `trimend(STRING)`。
- 从字符串中抽取开头或末尾的 n 个字符 - `startstring(LENGTH, STRING)` 或 `endstring(LENGTH, STRING)`。例如，假设您有一个名为项的字段，该字段将产品名称与四位数的标识代码（ACME CAMERA-D109）组合在一起。要创建仅包含四位数代码的新字段，请在“派生”节点中指定以下公式：

```
endstring(4, item)
```

- 与特定模式匹配 - `STRING matches PATTERN`。例如，要选择职位中任何位置包含“市场”二字的人员，您可以在“选择”节点中指定以下内容：

```
job_title matches "*market*"
```

- 替换字符串中某个子串的所有实例 - `replace(SUBSTRING, NEWSUBSTRING, STRING)`。例如，要在文本挖掘之前将不受支持字符的所有实例（例如，竖线 (|)）替换为分号，请在“填充”节点中使用

replace 函数。在**填入字段**：下，选中可能出现该字符的所有字段。对于**替换**：条件，请选择**始终**，并在**替换为**：中指定以下条件：

```
replace('|', ';', @FIELD)
```

- 根据特定子串的存在与否，导出标志字段。例如，可以在“派生”节点中使用字符串函数，针对每个响应生成单独的标志字段，使用的表达式如下所示：

```
hassubstring(museums, "museum_of_design")
```

有关更多信息，请参阅主题 [第 147 页的『字符串函数』](#)。

处理空白值和缺失值

对数据挖掘人员而言，替换空白值和缺失值是常见的数据准备任务。CLEM 提供了多种自动处理空白值的工具。最常用于处理空白值的是“填充”节点；除此以外，在任何可以使用 CLEM 表达式的节点中，还可使用下列函数：

- @BLANK(FIELD) 可用来确定特定字段（例如，Age）的值为空的记录。
- @NULL(FIELD) 可用于确定指定字段中值为系统缺失值的记录。在 IBM SPSS Modeler 中，系统缺失值显示为 \$null\$ 值。

有关更多信息，请参阅主题 [第 159 页的『处理空白值和空值的函数』](#)。

数字的处理

在 IBM SPSS Modeler 中，存在众多可用于数字值的标准运算，例如：

- 计算指定角度的正弦 - sin(NUM)
- 计算数字字段的自然对数 - log(NUM)
- 计算两个数字之和 - NUM1 + NUM2

有关更多信息，请参阅主题 [第 142 页的『数字函数』](#)。

时间和日期的处理

时间和数据格式可能会因为您数据源和语言环境的不同而有所不同。可在“流属性”对话框中设置具体到每个流的日期和时间格式。下面的示例是处理日期/时间字段时常用的函数。

计算已流逝的时间

利用与下列函数相似的一族函数，您可以轻松地计算出从基线日期开始的时间流逝。此函数以实数形式返回从基线日期到日期字符串 DATE 所表示日期的时间（以月为单位）。这是基于每月 30.0 天的近似数字。

```
date_in_months(Date)
```

日期/时间值的比较

利用与下列函数类似的函数，可以在记录之间比较日期/时间字段的值。如果日期字符串 DATE1 表示的日期早于日期字符串 DATE2 表示的日期，函数将返回 true。否则，此函数将返回值 0。

```
date_before(Date1, Date2)
```

计算差值

您可以使用函数来计算两个时间或两个日期之间的差值，如：

```
date_weeks_difference(Date1, Date2)
```

此函数以实数形式返回从日期字符串 DATE1 所表示日期到日期字符串 DATE2 所表示日期的时间（以周为单位）。这基于每周 7.0 天。如果 DATE2 早于 DATE1，那么此函数将返回一个负数。

当前日期

可使用函数 @TODAY 将当前日期添加至数据集。当前日期可作为字符串，以“流属性”对话框中选定的日期格式添加到指定字段或新字段中。有关更多信息，请参阅主题 [第 151 页的『日期和时间函数』](#)。

多个字段的汇总

CLEM 语言包含多个针对多个字段返回汇总统计的函数。这些函数在分析调查数据时特别有用，因为调查数据中某个问题的多个答案会存储在多个字段中。有关更多信息，请参阅主题 [第 125 页的『多个响应数据的处理』](#)。

比较函数

您可以使用 min_n 和 max_n 函数在多个字段值之间进行比较，例如：

```
max_n(['card1fee' 'card2fee' 'card3fee' 'card4fee'])
```

您还可以使用一些计数函数来获得满足特定标准的值的计数，即便这些值存储在多个字段中也无妨。例如，计算持有时间在五年以上的卡的数量：

```
count_greater_than(5, ['cardtenure' 'card2tenure' 'card3tenure'])
```

计数某字段集内的空值：

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

注意：本例中的计数是持有的卡的数量，并非持有卡的人数。有关更多信息，请参阅主题 [第 141 页的『比较函数』](#)。

要计算指定值在多个字段中出现的次数，您可以使用 count_equal 函数。下例计算在列表中包含值 Y 的字段数。

```
count_equal("Y", [Answer1, Answer2, Answer3])
```

假设列表中的字段具有以下值，则该函数对于值 Y 返回所示结果。

表 8: 函数值

Answer1	Answer2	Answer3	计数
Y	N	Y	2
Y	N	N	1

数字函数

您可以使用 sum_n、mean_n 和 sdev_n 函数来获取涉及多个字段的统计，例如：

```
sum_n(['card1bal' 'card2bal' 'card3bal'])
```

```
mean_n(['card1bal' 'card2bal' 'card3bal'])
```

有关更多信息，请参阅主题 [第 142 页的『数字函数』](#)。

生成字段列表

使用任何接受一系列字段作为输入的函数时，可以使用特殊函数 @FIELDS_BETWEEN(start, end) 和 @FIELDS_MATCHING(pattern) 作为输入。例如，假定字段顺序如以上 sum_n 示例所示，则下面的表达式等效：

```
sum_n(@FIELDS_BETWEEN(card1bal, card3bal))
```

或者，要计算所有以“card”开头的字段中的空值数目：

```
count_nulls(@FIELDS_MATCHING('card*'))
```

有关更多信息，请参阅主题 [第 160 页的『特殊字段』](#)。

多个响应数据的处理

有许多比较函数可用于分析多个响应数据，其中包括：

- `value_at`
- `first_index` / `last_index`
- `first_non_null` / `last_non_null`
- `first_non_null_index` / `last_non_null_index`
- `min_index` / `max_index`

例如，假设有一个多响应问题，内容是在个人购买过程中最看重的三个因素（如价格、他人推荐、评论、本地供应商等）。在这种情况下，您可以通过导出字段中首次包括的索引来确定价格的重要性：

```
first_index("price", [Reason1 Reason2 Reason3])
```

同样地，假定您已让客户按照购买的可能性对三辆汽车进行了排序，并且已经在三个不同的字段中对这些答案进行了编码，如下所示：

表 9: 汽车排序示例			
客户标识	car1	car2	car3
101	1	3	2
102	3	2	1
103	2	3	1

在本例中，您可以使用 `min_index` 函数，确定他们最喜欢（排名第一位，或排名最低）的汽车字段索引：

```
min_index(['car1' 'car2' 'car3'])
```

有关更多信息，请参阅主题 [第 141 页的『比较函数』](#)。

引用多重响应集

`@MULTI_RESPONSE_SET` 特殊函数可用于引用多重响应集中的所有字段。例如，如果前例中的三个 `car` 字段都包括在名为 `car_rankings` 的多重响应集中，那么以下函数将返回相同的结果：

```
max_index(@MULTI_RESPONSE_SET("car_rankings"))
```

表达式构建器

您可手动键入 CLEM 表达式，也可使用表达式构建器，后者将显示一个 CLEM 函数和运算符的完整列表以及当前流中的数据字段，因此，您无需记住字段或函数的准确名称即可快速构建表达式。此外，“构建器”控件自动为字段和值添加适当的引号，使得创建语法正确的表达式变得更加容易。

注：“表达式构建器”在脚本编译或参数设置中并不受支持。

注：如果要更改数据源，那么在更改数据源之前，您应检查表达式构建器是否仍支持选择的函数。由于并非所有数据库都支持所有函数，因此对新数据源运行时，可能会发生错误。

访问表达式构建器

所有使用 CLEM 表达式的节点中都有“表达式构建器”，这些节点包括选择、平衡、导出、填充、分析、报告和表。单击公式字段右侧的“计算器”按钮即可将其打开。

创建表达式

“表达式构建器”不仅提供了字段、函数和运算符的完整列表，如果您的数据已经实例化，它还可以用来访问数据值。

使用“表达式构建器”创建表达式

1. 参考函数和字段列表，在表达式字段输入公式。
- 或者
2. 从滚动的列表中选择所需的字段和函数。
3. 双击或单击黄色方向按钮可将字段或函数添加到表达式字段中。
4. 使用位于对话框中心的运算符按钮可将运算插入到表达式中。

选择函数

函数列表显示了所有可用的 CLEM 函数和运算符。滚动列表以从中选择函数，为了方便搜索，也可使用下拉列表来显示函数或运算符的子集。可用的函数按类别分组，以便更方便的搜索。

在 CLEM 语言说明的参考部分中描述了其中的多数类别。有关更多信息，请参阅 [第 138 页的『函数参考』](#)。

其他类别如下所示：

- **一般函数**包含精选的最常用函数。
- **最近使用**包含当前会话中使用的 CLEM 函数列表。
- **@ 函数**包含所有特殊函数的列表，其名称前面带有 "@" 符号。
注：@DIFF1(FIELD1, FIELD2) 和 @DIFF2(FIELD1, FIELD2) 函数要求两个字段的类型相同（例如，均为 Integer、Long 或 Real）。
- **数据库函数**。如果流包含数据库连接（通过“数据库源”节点），该选项列出在此数据库中可用的函数，包括用户定义的函数 (UDF)。有关更多信息，请参阅 [第 126 页的『数据库函数』](#)。
- **数据库汇总**。如果流包含数据库连接（通过“数据库源”节点），该选项列出在此数据库中可用的汇总选项。这些选项在“汇总”节点的表达式构建器中提供。
- **数据库窗口汇总**。如果流包含数据库连接（通过“数据库源”节点），该选项列出在此数据库中可以使用的窗口汇总选项。这些选项仅在**字段操作**选用板中的节点内的表达式构建器中提供。

注：由于 SPSS Modeler 从数据库系统视图中获取 **Window 汇总函数**，因此可用的选项取决于数据库行为。

尽管称这些选项为“汇总”，但其并不是设计用于“汇总”节点；这些选项更适用于“派生”或“选择”这样的节点。原因是这些选项的输出是标量而不是真正的汇总，也就是说，它们与“汇总”节点一样不会减少输出中显示的数据的量。例如，您可以使用此种汇总在数据行中提供诸如“当前行和所有先前的平均值”这样的移动平均值。

- **内置汇总**。包含可使用的聚集的可能方式列表。
- **运算符**列出在构建表达式时可使用的所有运算符。另外，也可以通过对话框中央的按钮来使用运算符。
- **所有函数**包含了可用 CLEM 函数的完整列表。

选定了一组函数之后，双击即可将函数插入到表达式字段中光标指示的位置。

数据库函数

数据库函数可以在许多不同位置列出；下表显示 SPSS Modeler 查找函数详细信息时搜索的位置。数据库管理员可以使用此表来确保用户有权访问所需区域以便能够使用不同的函数。

另外，此表列出了一些条件，这些条件用于在某个函数可供使用时根据数据库和函数类型进行过滤。

注：如果使用 Amazon Redshift 中的数据库函数，您的数据库管理员可能需要授予您对以下六个数据库对象的许可权。前四个是系统目录表，最后两个是模式。

- pg_type
- pg_proc
- pg_namespace
- pg_aggregate
- information_schema
- pg_catalog

表 10: 表达式构建器中的数据库函数

数据库	函数类型	函数的查找位置	用于过滤函数的条件
Db2 LUW	UDF	SYSCAT.ROUTINES SYSCAT.ROUTINEPARMS	ROUTINETYPE 为 F，而 FUNCTIONTYPE 为 S
Db2 LUW	UDA	SYSCAT.ROUTINES SYSCAT.ROUTINEPARMS	ROUTINETYPE 为 F，而 FUNCTIONTYPE 为 C
Db2 iSeries	UDF	QSYS2.SYSROUTINES QSYS2.SYSPARMS	ROUTINE_TYPE 为 F，而 FUNCTION_TYPE 为 S
Db2 iSeries	UDA	QSYS2.SYSROUTINES QSYS2.SYSPARMS	ROUTINE_TYPE 为 F，而 FUNCTION_TYPE 为 C
Db2 z/OS	UDF	SYSIBM.SYSROUTINES SYSIBM.SYSPARMS	ROUTINETYPE 为 F，而 FUNCTIONTYPE 为 S
Db2 z/OS	UDA	SYSIBM.SYSROUTINES SYSIBM.SYSPARMS	ROUTINETYPE 为 F，而 FUNCTIONTYPE 为 C
SQL Server	UDF	SYS.ALL_OBJECTS SYS.ALL_PARAMETERS SYS.TYPES	TYPE 为 FN 或 FS
SQL Server	UDA	SYS.ALL_OBJECTS SYS.ALL_PARAMETERS SYS.TYPES	TYPE 为 AF
Oracle	UDF	ALL_ARGUMENTS ALL_PROCEDURES	满足以下所有条件： <ul style="list-style-type: none"> • OBJECT_TYPE 为 FUNCTION • AGGREGATE 为 NO • PLS_TYPE 不为 NULL
Oracle	UDA	ALL_ARGUMENTS ALL_PROCEDURES	满足以下所有条件： <ul style="list-style-type: none"> • ARGUMENT_NAME 为 NULL • AGGREGATE 为 YES • PLS_TYPE 不为 NULL

表 10: 表达式构建器中的数据库函数 (继续)

数据库	函数类型	函数的查找位置	用于过滤函数的条件
Teradata	UDF	DBC.FUNCTIONS DBC.ALLRIGHTS	满足以下所有条件: <ul style="list-style-type: none"> • FUNCTIONTYPE 为 F • COLUMNNAME 为 RETURN0 • SPPARAMETERTYPE 为 0 • ACCESSRIGHT 为 EF
Teradata	UDA	DBC.FUNCTIONS DBC.ALLRIGHTS	满足以下所有条件: <ul style="list-style-type: none"> • FUNCTIONTYPE 为 A • COLUMNNAME 为 RETURN0 • SPPARAMETERTYPE 为 0 • ACCESSRIGHT 为 EF
Netezza	UDF	#####._V_FUNCIO N NZA._V_FUNCTION INZA._V_FUNCIO N	对于 #####._V_FUNCTION, 符合下列条件: <ul style="list-style-type: none"> • RESULT 不包含具有值的字符串, 例如 TABLE% • FUNCTION 不包含具有值的字符串, 例如 '/_%' escape '/' • VARARGS 为 FALSE 对于 NZA._V_FUNCTION 和 INZA._V_FUNCTION, 以下条件适用: <ul style="list-style-type: none"> • RESULT 不包含具有值的字符串, 例如 TABLE% • FUNCTION 不包含具有值的字符串, 例如 '/_%' escape '/' • BUILTIN 为 f • VARARGS 为 FALSE
Netezza	UDA	#####._V_AGGREGA TE NZA._V_FUNCTION INZA._V_FUNCIO N	同时满足以下两个条件: <ul style="list-style-type: none"> • AGGTYPE 为 ANY 或 GROUPED • VARARGS 为 FALSE
Netezza	WUDA	#####._V_AGGREGA TE NZA._V_FUNCTION INZA._V_FUNCIO N	对于 #####._V_AGGREGATE, 符合下列条件: <ul style="list-style-type: none"> • AGGTYPE 为 ANY 或 ANALYTIC • AGGREGATE 不为 MAX_LABEL • VARARGS 为 FALSE 对于 NZA._V_FUNCTION 和 INZA._V_FUNCTION, 以下条件适用: <ul style="list-style-type: none"> • AGGTYPE 为 ANY 或 ANALYTIC • BUILTIN 为 f • VARARGS 为 FALSE

表中使用的术语关键字

- UDF 用户定义的函数

- UDA 用户定义的聚集函数
- WUDA 用户定义的窗口聚集函数
- ##### 当前连接到的数据库。

选择字段、参数和全局变量

字段列表显示了数据流的这个位置上的所有可用字段。滚动列表以从中选择一个字段。双击或单击黄色箭头按钮可将字段添加到表达式中。

有关更多信息，请参阅主题 [第 122 页的『流、会话和超节点参数』](#)。

除了这些字段，您还可以从以下项目中进行选择：

多重响应集。 有关更多信息，请参阅 *IBM SPSS Modeler Source, Process, and Output Nodes* 指南。

最近使用的包含了当前对话中所使用的字段、多重响应集、参数和全局值的列表。

参数。 有关更多信息，请参阅主题 [第 122 页的『流、会话和超节点参数』](#)。

全局值。 有关更多信息，请参阅 *IBM SPSS Modeler Source, Process, and Output Nodes* 指南。

查看或选择值

系统的很多位置都可以用来查看字段值，包括表达式构建器、数据审核报告以及在时间间隔节点中编辑未来值时。注意：要使用该功能，数据必须在源或类型节点中完全实例化，以便掌握数据的存储、类型和值。

要从“表达式构建器”或时间间隔节点查看字段值，请选择要查看的字段并单击“值选取器”按钮，这将打开一个列有所选字段字段值的对话框。随后可选择一个值并单击 **插入** 已将值粘贴到当前表达式或列表中。



图 16: “值选取器”按钮

对于标志和名义字段，将列出所有定义的值。对于连续（数值范围）字段，显示最小和最大值。

检查 CLEM 表达式

单击“表达式构建器”中的**检查**（右下角）以验证表达式。尚未检查的表达式显示为空格。如果遇到错误，将显示一条指示起因的消息。

将检查以下的事项：

- 值和字段名的正确引用
- 参数和全局变量的正确使用
- 运算符的有效使用
- 引用的字段存在
- 引用的全局量存在及定义正确

如果您遇到语法错误，请尝试使用列表和运算符按钮来创建表达式，而不要手动键入表达式。这种方法可自动为字段和值添加适当的引号。

在 IBM Analytical Decision Management 中构建表达式时，请注意下列限制。表达式不得包含下列任意项：

- 对 IBM SPSS Modeler 流参数的引用
- 对 IBM SPSS Modeler 流全局量的引用
- 对数据库函数的引用
- 对下列其中一个特殊字段或字段值 @ 函数的引用：
 - @TARGET
 - @PREDICTED

- @FIELD
- @PARTITION_FIELD
- @TRAINING_PARTITION
- @TESTING_PARTITION
- @VALIDATION_PARTITION

注: 包含分隔符的字段名称必须用引号括起。要自动添加引号, 可以使用列表和运算符按钮创建表达式, 而不是手动输入表达式。字段名称中出现以下字符可能会导致错误: • ! "# \$% & ' () = ~ | - ^ ¥ @ " + * " <?> . , / ; ; → (箭头标记)、□ △ (图形标记等)

查找和替换

可在编辑脚本或表达式文本的位置 (包括脚本编辑器和 CLEM 表达式构建器) 或定义“报告”节点中的模板时使用“查找/替换”对话框。在上述任一区域编辑文本时, 按 **Ctrl+F** 可访问此对话框, 并确保光标的焦点位于文本区域中。例如, 处理填充节点时, 可以通过“设置”选项卡的任一文本区域或表达式构建器中的文本字段访问此对话框。

1. 使用文本区域中的光标, 按 **Ctrl+F** 可访问“查找/替换”对话框。
2. 输入要搜索的文本, 或从最近搜索项下拉列表中选择。
3. 输入替换文本 (如果有的话)。
4. 单击**查找下一个**开始搜索。
5. 单击**替换**替换当前选定的内容, 或单击**全部替换**更新所有项或选定的实例。
6. 每次操作完成后, 此对话框将关闭。从任一文本区域中按 **F3** 键, 可重复上一次查找操作, 或按 **Ctrl+F**, 可再次访问该对话框。

搜索选项

匹配大小写。 指定查找操作是否区分大小写; 例如 *myvar* 是否与 *myVar* 匹配。无论怎样设置, 替换文本始终完全按照输入插入。

全字匹配。 指定查找操作是否匹配单词中内嵌的文本。如果选中, *spider* 的搜索结果将不会包括 *spiderman* 或 *spider-man*。

正则表达式。 指定是否使用正则表达式语法 (请参阅下一节)。如果选中, **仅限于整个单词**选项将禁用并且会忽略其值。

仅所选文本。 控制使用**全部替换**选项时的搜索范围。

正则表达式语法

使用正则表达式, 您可以搜索特殊字符 (如选项卡或换行字符)、字符的类或范围 (如 *a* 到 *d*)、任何数字或非数字以及边界 (如行首或行尾)。支持的表达式类型如下。

字符	匹配
x	字符 x
\\	反斜杠字符
\\0n	含八进制值的字符 0n (0 <= n <= 7)
\\0nn	含八进制值的字符 0nn (0 <= n <= 7)
\\0mnn	含八进制值的字符 0mnn (0 <= m <= 3, 0 <= n <= 7)
\\xhh	含十六进制值的字符 0xhh
\\uhhhh	含十六进制值的字符 0xhhhh
\\t	制表符 ('\\u0009')

表 11: 字符匹配 (继续)

字符	匹配
\n	换行符 ('\u000A')
\r	回车符 ('\u000D')
\f	换页符 ('\u000C')
\a	警报 (蜂鸣) 符 ('\u0007')
\e	转义符 ('\u001B')
\cx	x 对应的控制字符

表 12: 匹配字符类

字符类	匹配
[abc]	a、b、或 c (简单类)
[^abc]	除 a、b、或 c 之外的所有字符 (相减)
[a-zA-Z]	a 到 z 或 A 到 Z, 包含 (范围)
[a-d[m-p]]	a 到 d 或 m 到 p (合并)。也可指定为 [a-dm-p]
[a-z&&[def]]	a 到 z 和 d、e、或 f (交集)
[a-z&&[^bc]]	a 到 z, 除 b 和 c 外 (相减)。也可指定为 [ad-z]
[a-z&&[^m-p]]	a 到 z, 而非 m 到 p (相减)。也可指定为 [a-lq-z]

表 13: 预定义字符类

预定义字符类	匹配
.	任意字符 (可能或不可能与行终止符匹配)
\d	任意数字: [0-9]
\D	非数字: [^0-9]
\s	空格字符: [\t\n\x0B\f\r]
\S	非空格字符: [^\s]
\w	单词字符: [a-zA-Z_0-9]
\W	非单词字符: [^\w]

表 14: 边界匹配

边界匹配符	匹配
^	行首
\$	行尾
\b	单词边界
\B	非单词边界
\A	输入的头
\Z	除最后终止符外 (如果有), 输入的结尾
\z	输入的结尾

第 10 章 CLEM 语言参考

CLEM 参考概述

本节介绍表达式操作控制语言 (CLEM)，这是一个强大的工具，可用于分析和操控 IBM SPSS Modeler 流中使用的数据。可以在节点内使用 CLEM，以执行从评估条件或推导值到将数据插入报告等一系列任务。

CLEM 表达式由值、字段名、运行符和函数组成。使用正确的语法，可以创建各种功能强大的数据操作。

CLEM 数据类型

CLEM 数据类型可以由下列任意一项构成：

- 整数
- 实数
- 字符
- 字符串
- 列表
- 字段
- 日期/时间

引用规则

虽然 IBM SPSS Modeler 在确定 CLEM 表达式中使用的字段、值、参数和字符串时非常灵活，但在创建表达式时遵循以下一般规则将有利于您的实践操作：

- 字符串 - 在编写字符串时始终使用双引号，例如 "Type 2"。也可以使用单引号，但容易与带引号字段混淆。
- 字段 - 仅在必要时使用单引号将空格或其他特殊字符括起来，例如 'Order Number'。加引号但数据集中未定义的字段会误读为字符串。
- 参数 - 使用参数时始终使用单引号，例如 '\$P-threshold'。
- 字符 - 始终使用单反引号 (')，例如 stripchar('d', "drugA")。

以下主题详细介绍了这些规则。

整数

整数表示为十进制数字序列。可以选择在整数前加负号 (-) 来表示负数，例如 1234、999 和 -77。

CLEM 语言可以处理任意精度的整数。最大整数大小取决于您的平台。如果值太大而无法显示在整数字段中，那么将字段类型更改为 Real 通常会恢复该值。

实数

实数是指浮点数。实数由一个或多个数字来表示，后跟小数点，小数点后再跟一个或多个数字。CLEM 实数以双精度保存。

或者，您可以在实数之前放置一个减号 (-) 来表示负数，例如，1.234、0.999、-77.001。使用 `<number> e <exponent>` 形式以指数表示法来表示实数，例如，1234.0e5、1.7e-2。当 IBM SPSS Modeler 应用程序从文件中读取数字字符串并将它们自动转换为数字时，接受小数点前没有前导数字或小数点后没有数字的那些数字，例如，999. 或 .11。不过，这些形式在 CLEM 表达式中是非法的。

注：在 CLEM 表达式中引用实数时，无论当前流或语言环境的设置如何，都必须使用句点作为小数分隔符。例如，指定

```
Na > 0.6
```

而不是

```
Na > 0,6
```

即使在流属性对话框中选择了逗号作为小数符号，此规则也适用，它与代码语法应独立于任何特定语言环境或惯例的一般准则一致。

字符

在 CLEM 表达式中通常使用字符（通常显示为 CHAR）来对字符串执行测试。例如，您可以使用函数 `isuppercode` 来确定字符串的第一个字符是否为大写。下面的 CLEM 表达式使用一个字符来表示应对字符串的第一个字符进行这样的检验：

```
isuppercode(subscrs(1, "MyString"))
```

要在 CLEM 表达式中表达特定字符的代码（与位置相反），请使用 `<character>` 格式的单反引号 - 例如，`'A'`、`'Z'`。

注：对于字段而言，没有 CHAR 存储类型，如果使用会产生 CHAR 存储类型的表达式派生或填充字段，则结果将转换为字符串。

字符串

通常，应该将字符串括在双引号内。字符串示例为 `"c35product2"` 和 `"referrerID"`。要指示字符串中的特殊字符，请使用反斜杠，例如 `"\$65443"`。（要表示反斜杠字符，应使用两个反斜杠 `\\`。）您可以使用单引号将字符串括起来，但是结果会与带引号的字段（`'referrerID'`）难以区分。请参阅主题第 147 页的『字符串函数』，以获取更多信息。

列表

列表是元素的有序排列，其中的元素可以是混合类型。列表括在方括号 ([]) 中。列表的示例为 `[1 2 4 16]` 和 `["abc" "def"]`。列表并不用作 IBM SPSS Modeler 字段的值。它们用于向函数提供参数，例如 `member` 和 `oneof`。

注：列表可以仅包含静态对象（例如字符串、数字或字段名称），而不包含函数调用。

字段

CLEM 表达式中的名称，如若不是函数名称，则视其为字段名称。您可以将这些内容简单地写为 `Power`、`val27`、`state_flag` 等，但如果名称以数字开头或包含非字母字符，例如，空格（下划线除外），请将名称括在单引号内，例如 `'Power Increase'`、`'2nd answer'`、`'#101'`、`'$P-NextField'`。

注：加引号但数据集中未定义的字段会误读为字符串。

日期

日期将以在“流属性”对话框中指定的“基线”日期为依据进行计算。缺省基线日期为 1900 年 1 月 1 日。

CLEM 语言支持以下日期格式。

表 15: CLEM 语言日期格式	
格式	示例
DDMMYY	150163
MMDDYY	011563

表 15: CLEM 语言日期格式 (继续)	
格式	示例
YYMMDD	630115
YYYYMMDD	19630115
YYYYDDD	前四位数表示年, 后三位数表示一年中的某一天 - 例如, 2000032 表示 2000 年的第 32 天, 即 2000 年 2 月 1 日。
DAY	当前语言环境下一周中的某一天 例如, 英语环境下的 Monday 和 Tuesday 等等。
MONTH	当前语言环境下的月份, 例如 January 和 February 等等。
DD/MM/YY	15/01/63
DD/MM/YYYY	15/01/1963
MM/DD/YY	01/15/63
MM/DD/YYYY	01/15/1963
DD-MM-YY	15-01-63
DD-MM-YYYY	15-01-1963
MM-DD-YY	01-15-63
MM-DD-YYYY	01-15-1963
DD.MM.YY	15.01.63
DD.MM.YYYY	15.01.1963
MM.DD.YY	01.15.63
MM.DD.YYYY	01.15.1963
DD-MON-YY	15-JAN-63, 15-jan-63, 15-Jan-63
DD/MON/YY	15/JAN/63, 15/jan/63, 15/Jan/63
DD.MON.YY	15.JAN.63, 15.jan.63, 15.Jan.63
DD-MON-YYYY	15-JAN-1963, 15-jan-1963, 15-Jan-1963
DD/MON/YYYY	15/JAN/1963, 15/jan/1963, 15/Jan/1963
DD.MON.YYYY	15.JAN.1963, 15.jan.1963, 15.Jan.1963
MON YYYY	Jan 2004
q Q YYYY	日期表示为代表季度的数字 (1-4), 后跟字母 Q 和四位数字的年份 - 例如, 2004 年 12 月 25 日将表示为 4 Q 2004。
ww WK YYYY	前两位数字代表一年中的某周, 后跟字母 WK 以及代表年的四位数字。计算一年中的某周时, 假设这周的第一天是星期一且第一周至少有一天。

时间

CLEM 语言支持以下时间格式。

表 16: CLEM 语言时间格式

格式	示例
HHMMSS	120112, 010101, 221212
HHMM	1223, 0745, 2207
MMSS	5558, 0100
HH:MM:SS	12:01:12, 01:01:01, 22:12:12
HH:MM	12:23, 07:45, 22:07
MM:SS	55:58, 01:00
(H)H:(M)M:(S)S	12:1:12, 1:1:1, 22:12:12
(H)H:(M)M	12:23, 7:45, 22:7
(M)M:(S)S	55:58, 1:0
HH.MM.SS	12.01.12, 01.01.01, 22.12.12
HH.MM	12.23, 07.45, 22.07
MM.SS	55.58, 01.00
(H)H.(M)M.(S)S	12.1.12, 1.1.1, 22.12.12
(H)H.(M)M	12.23, 7.45, 22.7
(M)M.(S)S	55.58, 1.0

CLEM 运算符

CLEM 语言提供了下列运算符。

表 17: CLEM 语言运算符

运算	注释	优先级 (参阅下一节)
or	用于两个 CLEM 表达式之间。如果其中一个表达式为真或两个表达式同时为真, 则返回真值。	10
and	用于两个 CLEM 表达式之间。如果两个表达式同时为真, 则返回真值。	9
=	用于任意两个可比项之间。若 ITEM1 等于项目 ITEM2, 则返回值为 true。	7
==	与 = 相同。	7
/=	用于任意两个可比项之间。若 ITEM1 不等于 ITEM2, 则返回值为真。	7
/==	与 /= 相同。	7
>	用于任意两个可比项之间。若 ITEM1 完全大于 ITEM2, 则返回值为 true。	6
>=	用于任意两个可比项之间。若 ITEM1 大于或等于 ITEM2, 则返回值为 true。	6
<	用于任意两个可比项之间。如果 ITEM1 严格小于 ITEM2, 则返回真值。	6

表 17: CLEM 语言运算符 (继续)

运算	注释	优先级 (参阅下一节)
<=	用于任意两个可比项之间。若 ITEM1 小于或等于 ITEM2, 则返回值为 true。	6
&&=_0	用于两个整数之间。相当于布尔表达式 INT1 && INT2 = 0。	6
&&/=_0	用于两个整数之间。相当于布尔表达式 INT1 && INT2 /= 0。	6
+	将两个数字相加: NUM1 + NUM2。	5
><	连接两个字符串, 例如 STRING1 >< STRING2.	5
-	将两个数字相减: NUM1 - NUM2。也可以用在数字之前: - NUM。	5
*	用于将两个数字相乘: NUM1 * NUM2。	4
&&	用于两个整数之间。结果为 INT1 和 INT2 的位元“与”。	4
&&~~	用于两个整数之间。结果为 INT1 的位元“与”和 INT2 的位元补码。	4
	用于两个整数之间。结果为 INT1 和 INT2 的位元“或”。	4
~~	用于整数前。生成整数的位元补码。	4
/&	用于两个整数之间。结果为 INT1 和 INT2 的位元“异或”。	4
INT1 << N	用于两个整数之间。产生整数向左偏移 N 个位置的位元模式。	4
INT1 >> N	用于两个整数之间。产生整数向右偏移 N 个位置的位元模式。	4
/	用于将两个数相除: NUM1 / NUM2。	4
**	用于两个数字之间: BASE ** POWER。返回 BASE 的 POWER 次幂。	3
rem	用于两个整数之间: INT1 rem INT2。返回余数, INT1 - (INT1 div INT2) * INT2。	2
div	用于两个整数之间: INT1 div INT2。执行整数除。	2

运算符优先级

优先级决定复杂表达式 (尤其是含有不止一个中缀运算符且未加括号的表达式) 的解析顺序。例如,

```
3 + 4 * 5
```

解析为 $3 + (4 * 5)$ 而不是 $(3 + 4) * 5$, 因为相对优先顺序指示在 + 之前解析 *。CLEM 语言中的每个运算符都具有与之关联的优先顺序值; 此值越低, 它在解析列表中就越重要, 这意味着将比其他具有较高优先顺序值的运算符更快进行处理。

函数参考

下列 CLEM 函数可用于在 IBM SPSS Modeler 中处理数据。您可以在各种对话框（例如“派生”节点和“设为标志”节点）中输入这些函数作为代码，也可以使用表达式构建器来创建有效的 CLEM 表达式，而后一种方式无需记忆函数列表或字段名。

函数类型	描述
信息	用于深入了解字段值。例如，函数 <code>is_string</code> 针对类型为字符串的所有记录返回真值。
转换	用于构建新字段或转换存储类型。例如，函数 <code>to_timestamp</code> 会将选定字段转换为时间戳。
比较	用于字段值的相互比较或与指定字符串进行比较。例如， <code><=</code> 用于比较两个字段的值是否存在小于或等于关系。
逻辑	用来进行逻辑运算，例如， <code>if</code> 、 <code>then</code> 、 <code>else</code> 运算。
数字	用来进行数值计算，例如对字段值取自然对数。
三角函数	用来进行三角计算，例如指定角度的反余弦。
概率 (Probability)	返回基于各种分布的概率，例如学生氏 t 分布中某个值小于特定值的概率。
空间	用于对地理数据执行空间计算。
按位	用于以位元模式操作整数。
随机	用于随机选择项或生成数值。
字符串	用于对字符串执行各种操作，例如 <code>stripchar</code> 用来移除指定的字符。
SoundEx	用于在不知道字符串准确拼写的情况下根据某些字母的假设发音查找字符串。
日期和时间	用于对日期、时间和时间戳记字段执行各种操作。
序列	用于深入了解数据集的记录序列，或者根据该序列执行操作。
全局	用于访问由“设置全局量”节点创建的全局值。例如， <code>@MEAN</code> 用于引用整个数据集内某个字段的所有值的平均值。
空白和 Null	用于访问、标记或填充用户指定的空白值或系统缺失值。例如， <code>@BLANK(FIELD)</code> 用于为存在空白值的记录添加一个真值标志。
特殊字段	用于标示检查中的特定字段。例如，在派生多个字段时使用 <code>@FIELD</code> 。

函数说明中的约定

涉及函数中的项时，以下惯例通用于本指南。

惯例	描述
<code>BOOL</code>	布尔值或标志，例如真或假。
<code>NUM</code> 、 <code>NUM1</code> 和 <code>NUM2</code>	任意数字。
<code>REAL</code> 、 <code>REAL1</code> 和 <code>REAL2</code>	任何实数，例如，1.234 或 -77.01。

表 19: 函数说明中的约定 (继续)

惯例	描述
INT、INT1 和 INT2	任何整数，例如 1 或 -77。
Char	字符代码，例如 `A`。
STRING	字符串，例如 "referrerID"。
列表	项目列表，例如 ["abc" "def"]。
ITEM	字段，例如 Customer 或 extract_concept。
date	其值具有 DD-MON-YYYY 这样格式的日期字段，例如 start_date。
时间	其值具有 HHMMSS 这样格式的时间字段，例如 power_flux。

本指南中的函数以函数、结果类型（整数、字符串等）和说明（如果有）各占一列的形式一一列举说明。例如，函数 rem 的描述如下。

表 20: rem 函数描述

函数	结果	描述
INT1 rem INT2	数字	返回 INT1 除以 INT2 所得的余数。例如，INT1 - (INT1 div INT2) * INT2。

有关用法约定的详情（例如，如何在函数中列出项目或指定字符）将在其他文档中介绍。有关更多信息，请参阅主题 第 133 页的『CLEM 数据类型』。

信息函数

信息函数用于深入了解特定字段的值。它们通常用于派生标志字段。例如，您可以使用 @BLANK 函数来创建一个标志字段，指示所选字段的值为空的记录。同样，您可以使用任何存储类型函数（例如，is_string）来检查字段的存储类型。

表 21: CLEM 信息函数

函数	结果	描述
@BLANK(FIELD)	布尔值	根据上游类型节点或源节点（“类型”选项卡）中所设置的空白值处理规则，对所有空白值的记录返回真值。
@NULL(ITEM)	布尔值	为未定义值的所有记录返回真值。未定义的值是系统 Null 值，在 IBM SPSS Modeler 中显示为 \$null\$。
is_date(ITEM)	布尔值	为所有日期类型的记录返回真值。
is_datetime(ITEM)	布尔值	为所有日期、时间或时间戳类型的记录返回真值。
is_integer(ITEM)	布尔值	为所有整数类型的记录返回真值。
is_number(ITEM)	布尔值	为所有数值类型的记录返回真值。
is_real(ITEM)	布尔值	为所有实数类型的记录返回真值。
is_string(ITEM)	布尔值	为所有字符串类型的记录返回真值。
is_time(ITEM)	布尔值	为所有时间类型的记录返回真值。
is_timestamp(ITEM)	布尔值	为所有时间戳类型的记录返回真值。

转换函数

转换函数可用来构建新字段和转换现有文件的存储类型。例如，可通过将字符串连接在一起或分拆字符串来形成新字符串。要连接两个字符串，请使用运算符 $><$ 。例如，如果字段 *Site* 的值为 "BRAMLEY"，那么 "xx" $><$ *Site* 将返回 "xxBRAMLEY"。即使参数不是字符串， $><$ 的结果也始终是字符串。因此，如果字段 *V1* 为 3，字段 *V2* 为 5，那么 *V1* $><$ *V2* 将返回 "35"（字符串而不是数字）。

转换函数（及其他要求特定类型输入（如日期或时间值）的函数）取决于“流选项”对话框中指定的当前格式。例如，如果想转换值为 *Jan 2003*、*Feb 2003* 等的字符串字段，请选择匹配的日期格式 **MON YYYY** 作为流的缺省时间格式。

函数	结果	描述
ITEM1 $><$ ITEM2	<i>String</i>	将两个字段的值并置，并返回结果字符串 <i>ITEM1ITEM2</i> 。
to_integer(ITEM)	整数	将指定字段的存储值转换为整数。
to_real(ITEM)	实数	将指定字段的存储值转换为实数。
to_number(ITEM)	数字	将指定字段的存储值转换为数值。
to_string(ITEM)	<i>String</i>	将指定字段的存储值转换为字符串。使用此函数将实数转换为字符串时，将返回在基点后存在 6 位数的值。
to_time(ITEM)	时间	将指定字段的存储值转换为时间。
to_date(ITEM)	<i>date</i>	将指定字段的存储值转换为日期。
to_timestamp(ITEM)	时间戳记	将指定字段的存储值转换为时间戳记。
to_datetime(ITEM)	日期/时间	将指定字段的存储值转换为日期、时间或时间戳记值。
datetime_date(ITEM)	<i>date</i>	返回数字、字符串或时间戳记的日期值。请注意，这是唯一允许您将数字（以秒数记）转换回日期的函数。如果 <i>ITEM</i> 为字符串，将以当前日期格式解析字符串来创建日期。流属性对话框中指定的日期格式必须正确，这样此函数才能成功执行。如果 <i>ITEM</i> 是一个数字，那么会将其解释为自基本日期（或戳记）以来的秒数。不足一天的部分将截断。如果 <i>ITEM</i> 为时间戳，将返回时间戳的日期部分。如果 <i>ITEM</i> 为日期，将返回未更改的日期。
stb_centroid_latitude(ITEM)	整数	返回与地理散列自变量的质心相对应的纬度的整数值。
stb_centroid_longitude(ITEM)	整数	返回与地理散列自变量的质心相对应的经度的整数值。
to_geohash(ITEM)	<i>String</i>	<p>返回对应于纬度和经度并使用密度的指定位数的地理散列字符串。</p> <p>地理散列是一个代码，用于根据纬度和经度详细信息标识一组地理坐标。to_geohash 的三个参数为：</p> <ul style="list-style-type: none"> • 纬度：范围 (-180, 180)，并且在 WGS84 坐标系中单位为度。 • 经度：范围 (-90, 90)，并且在 WGS84 坐标系中单位为度。 • 位：要用于存储散列的位数。范围为 [1,75]。这将同时影响所返回的字符串的长度（对每 5 个位使用 1 个字符）以及散列的准确性。例如，5 个位（1 个字符）表示大约 2500 千米或 45 个位（9 个字符）表示大约 2.3 米。

比较函数

比较函数用于字段值的相互比较或与指定字符串进行比较。例如，您可以使用 = 检查字符串是否相等。字符串相等验证示例：Class = "class 1"。

对数值比较来说，大于 表示离正无穷更近，小于 表示离负无穷更近。即，所有负数均小于任意正数。

函数	结果	描述
count_equal(ITEM1, LIST)	整数	返回字段列表中等于 ITEM1 的值的个数，如果 ITEM1 为空，则返回空。
count_greater_than(ITEM1, LIST)	整数	返回字段列表中大于 ITEM1 的值的个数，如果 ITEM1 为空，则返回空。
count_less_than(ITEM1, LIST)	整数	返回字段列表中小于 ITEM1 的值的个数，如果 ITEM1 为空，则返回空。
count_not_equal(ITEM1, LIST)	整数	返回字段列表中不等于 ITEM1 的值的个数，如果 ITEM1 为空，则返回空。
count_nulls(LIST)	整数	返回字段列表中空值的个数。
count_non_nulls(LIST)	整数	返回字段列表中非空值的个数。
date_before(DATE1, DATE2)	布尔值	用于检查日期值的排序。DATE1 早于 DATE2，则返回真值。
first_index(ITEM, LIST)	整数	返回字段列表中包含 ITEM 的第一个字段的索引，如果找不到该值，则返回 0。仅支持字符串、整数和实数类型。
first_non_null(LIST)	任意	返回所提供字段列表中的第一个非空值。支持所有存储类型。
first_non_null_index(LIST)	整数	返回字段列表中包含非空值的第一个字段的索引，如果所有值都为空值，则返回 0。支持所有存储类型。
ITEM1 = ITEM2	布尔值	为 ITEM1 等于 ITEM2 的记录返回真值。
ITEM1 /= ITEM2	布尔值	两个字符串不完全相同时返回真值；完全相同时返回 0。
ITEM1 < ITEM2	布尔值	为 ITEM1 小于 ITEM2 的记录返回真值。
ITEM1 <= ITEM2	布尔值	为 ITEM1 小于等于 ITEM2 的记录返回真值。
ITEM1 > ITEM2	布尔值	为 ITEM1 大于 ITEM2 的记录返回真值。
ITEM1 >= ITEM2	布尔值	为 ITEM1 大于等于 ITEM2 的记录返回真值。
last_index(ITEM, LIST)	整数	返回字段列表中包含 ITEM 的最后一个字段的索引，如果找不到该值，则返回 0。仅支持字符串、整数和实数类型。
last_non_null(LIST)	任意	返回所提供字段列表中的最后一个非空值。支持所有存储类型。
last_non_null_index(LIST)	整数	返回字段列表中包含非空值的最后一个字段的索引，如果所有值都为空值，则返回 0。支持所有存储类型。
max(ITEM1, ITEM2)	任意	返回 ITEM1 和 ITEM2 两项中较大的一个。
max_index(LIST)	整数	返回数字字段列表中包含最大值的字段索引，如果所有值为空值，则返回 0。例如，如果所列的第三个字段包含最大值，将返回索引值 3。如果有多个字段包含最大值，则将返回第一个列出的字段（位于最左侧）。

表 23: CLEM 比较函数 (继续)

函数	结果	描述
max_n(LIST)	数字	返回数字字段列表中的最大值, 如果所有字段值为空, 则返回空。
member(ITEM, LIST)	布尔值	如果 ITEM 是指定 LIST 的成员, 则返回真值。否则返回假值。另外, 还可以指定字段名称列表。
min(ITEM1, ITEM2)	任意	返回 ITEM1 和 ITEM2 两项中较小的一个。
min_index(LIST)	整数	返回数字字段列表中包含最小值的字段索引, 如果所有值为空值, 则返回 0。例如, 如果所列的第三个字段包含最小值, 将返回索引值 3。如果有多个字段包含最小值, 则将返回第一个列出的字段 (位于最左侧)。
min_n(LIST)	数字	返回数字字段列表中的最小值, 如果所有字段值为空, 则返回空。
time_before(TIME1, TIME2)	布尔值	用于检查时间值的排序。TIME1 早于 TIME2, 则返回真值。
value_at(INT, LIST)		在偏移 INT 处返回每个列出字段的值; 如果偏移超出了有效值的范围 (即小于 1 或大于所列字段的个数), 则返回空值。支持所有存储类型。

逻辑函数

CLEM 表达式可用来执行逻辑运算。

表 24: CLEM 逻辑函数

函数	结果	描述
COND1 and COND2	布尔值	此运算为逻辑合取运算, 在 COND1 和 COND2 同时为真时返回真值。如果 COND1 为假, 则不求 COND2 的值; 此时可以构造这样的合取, 让 COND1 首先检验 COND2 中运算是合法的。例如, length(Label) >=6 和 Label(6) = 'x'。
COND1 or COND2	布尔值	此运算为逻辑 (包含性) 析取运算, 在 COND1 或 COND2 为真或者这两者同时为真时返回真值。如果 COND1 为真, 则不求 COND2 的值。
not(COND)	布尔值	此运算为逻辑否运算, 如果 COND 为假, 则返回真。否则, 此运算将返回值 0。
if COND then EXPR1 else EXPR2 endif	任意	此运算为条件求值。如果 COND 为真, 此运算将返回 EXPR1 的结果。否则, 将返回 EXPR2 的求值结果。
if COND1 then EXPR1 elseif COND2 then EXPR2 else EXPR_N endif	任意	此运算为多分支条件求值。如果 COND1 为真, 那么此运算将返回 EXPR1 的结果。否则, 如果 COND2 为真, 此运算将返回 EXPR2 的求值结果。否则, 将返回 EXPR_N 的求值结果。

数字函数

CLEM 包含许多常用数值函数。

表 25: CLEM 数字函数

函数	结果	描述
-NUM	数字	用于求 NUM 的相反数。返回具有相反符号的对应数字。
NUM1 + NUM2	数字	返回 NUM1 与 NUM2 相加所得的和。
NUM1 - NUM2	数字	返回 NUM2 减去 NUM1 所得的差。
NUM1 * NUM2	数字	返回 NUM1 乘以 NUM2 所得的值。
NUM1 / NUM2	数字	返回 NUM1 除以 NUM2 所得的商。
INT1 div INT2	数字	用于进行整数除法。返回 INT1 除以 INT2 所得的商。
INT1 rem INT2	数字	返回 INT1 除以 INT2 所得的余数。例如, INT1 - (INT1 div INT2) * INT2。
INT1 mod INT2	数字	此函数已不推荐使用。请改为使用 rem 函数。
BASE ** POWER	数字	返回 BASE 的 POWER 次方, 底和幂可以是任意数字 (一种情况除外: 如果 POWER 为除整数 0 之外其他类型的 0, 则 BASE 不能为 0)。如果 POWER 为整数, 则可通过将 BASE 连续乘 POWER 次计算出结果。因此, 如果 BASE 为整数, 则结果为整数。如果 POWER 为整数 0, 则结果将始终是与 BASE 类型相同的 1。如果 POWER 不是整数, 则用公式 $\exp(\text{POWER} * \log(\text{BASE}))$ 计算结果。
abs(NUM)	数字	返回 NUM 的绝对值 (始终为同一类型的数值)。
exp(NUM)	实数	返回 e 的 NUM 次幂, 其中 e 是自然对数的底数。
fracof(NUM)	实数	返回定义为 NUM-intof(NUM) 的 NUM 的小数部分。
intof(NUM)	整数	将其参数截为整数。它返回与 NUM 相同符号的整数, 并且具有最大的量级, 例如 abs(INT) <= abs(NUM)。
log(NUM)	实数	返回 NUM 的自然对数 (以 e 为底), NUM 不能为零 (无论何种类型)。
log10(NUM)	实数	返回 NUM 的对数 (以 10 为底), 不能为零。此函数定义为 $\log(\text{NUM}) / \log(10)$ 。
negate(NUM)	数字	用于求 NUM 的相反数。返回具有相反符号的对应数字。
round(NUM)	整数	用于将 NUM 舍入为整数, 如果 NUM 为正, 那么取 intof(NUM+0.5); 如果 NUM 为负, 那么取 intof(NUM-0.5)。
sign(NUM)	数字	用于确定 NUM 的符号。如果 NUM 为整数, 此运算将返回 -1、0 或 1。如果 NUM 为实数, 将返回 -1.0、0.0 或 1.0, 这取决于 NUM 是负数、零还是正数。
sqrt(NUM)	实数	返回 NUM 的平方根。NUM 必须是正数。
sum_n(LIST)	数字	返回数字字段列表中字段值的和, 如果所有字段均为空, 则返回空值。
mean_n(LIST)	数字	返回数字字段列表中字段值的均值, 如果所有字段值均为空值, 则返回空值。
sdev_n(LIST)	数字	返回数字字段列表中字段值的标准差, 如果所有字段值均为空值, 则返回空值。

三角函数

本部分介绍的函数以角度为参数或返回结果为角度。无论哪种情况，角度单位（弧度或度数）均由相关流选项的设置控制。

函数	结果	描述
arccos(NUM)	实数	计算指定角度的反余弦值。
arccosh(NUM)	实数	计算指定角度的反双曲余弦值。
arcsin(NUM)	实数	计算指定角度的反正弦值。
arcsinh(NUM)	实数	计算指定角度的反双曲正弦值。
arctan(NUM)	实数	计算指定角度的反正切值。
arctan2(NUM_Y, NUM_X)	实数	计算 NUM_Y / NUM_X 的反正切值，然后使用两个数的符号推导出象限信息。结果是范围 $-\pi < \text{ANGLE} \leq \pi$ (radians) - $180 < \text{ANGLE} \leq 180$ (degrees) 内的一个实数
arctanh(NUM)	实数	计算指定角度的反双曲正切值。
cos(NUM)	实数	计算指定角度的余弦值。
cosh(NUM)	实数	计算指定角度的双曲余弦值。
pi	实数	此常数是最接近 pi 的实数。
sin(NUM)	实数	计算指定角度的正弦值。
sinh(NUM)	实数	计算指定角度的双曲正弦值。
tan(NUM)	实数	计算指定角度的正切值。
tanh(NUM)	实数	计算指定角度的双曲正切值。

概率函数

概率函数返回基于各种分布的概率，例如，学生氏 t 分布中某个值将会小于特定值的概率。

函数	结果	描述
cdf_chisq(NUM, DF)	实数	返回具有指定自由度的卡方分布中某个值将会小于指定数值的概率。
cdf_f(NUM, DF1, DF2)	实数	返回自由度为 $DF1$ 和 $DF2$ 的 F 分布中某个值将会小于指定数值的概率。
cdf_normal(NUM, MEAN, STDDEV)	实数	返回具有指定均值和标准差的正态分布中某个值将会小于指定数值的概率。
cdf_t(NUM, DF)	实数	返回具有指定自由度的学生氏 t 分布中某个值将会小于指定数值的概率。

空间函数

空间函数可以与地理空间数据配合使用。例如，它们使您能够计算两个点之间的距离以及多边形的面积等信息。并且，还存在要求对多个基于空间预测（处于和接近于，等等）的地理空间数据集进行合并的情况，此任务可以通过合并条件完成。

这些空间函数与工具 > 流属性 > 选项 > 地理空间中指定的坐标系配合使用。

注: 这些空间函数不适用于三维数据。如果将三维数据导入到流, 那么这些函数仅使用前两个维度。将忽略 z 轴值。

函数	结果	描述
close_to(SHAPE, SHAPE, NUM)	布尔值	检验两个形状之间的相互距离是否不超过特定的 DISTANCE。如果使用了投影坐标系, 那么 DISTANCE 以米为单位。如果未使用坐标系, 那么单位任意。
crosses(SHAPE, SHAPE)	布尔值	检验两个形状是否相交。此函数适用于两个线串形状或者一个线串和一个多边形。
overlap(SHAPE, SHAPE)	布尔值	检验两个多边形之间是否存在交集并且该交集是否在这两个形状之内。
within(SHAPE, SHAPE)	布尔值	检验是否整个 SHAPE1 包含在 POLYGON 之内。
area(SHAPE)	实数	返回所指定 POLYGON 的面积。如果使用了投影系统, 那么此函数将返回以平方米为单位的结果。如果未使用坐标系, 那么单位任意。形状必须是 POLYGON 或 MULTIPOLYGON。
num_points(SHAPE, LIST)	整数	返回点字段 (MULTIPOINT) 中包含在 POLYGON 的边界内的点数。SHAPE1 必须是 POLYGON 或 MULTIPOLYGON。
distance(SHAPE, SHAPE)	实数	返回 SHAPE1 与 SHAPE2 之间的距离。如果使用了投影坐标系, 那么此函数将返回以米为单位的结果。如果未使用坐标系, 那么单位任意。SHAPE1 和 SHAPE2 可以是任何地理测量类型。

位元整数运算

这些函数使整数能够被处理为表示两个补码值的位模式, 其中位的位置 N 具有权重 2^{*N} 。位从 0 开始往上数。这些运算就好像是把整数的符号位向左无限延伸。因此, 最高有效位之前的所有位, 正整数均为 0, 负整数均为 1。

函数	结果	描述
~~ INT1	整数	产生整数 INT1 的位元补码。即, 对于 INT1 具有 0 的每个位的位置, 结果中都有 1。~~ INT = -(INT + 1) 始终是正确的。
INT1 INT2	整数	此运算的结果是 INT1 与 INT2 的按位“或”。即, INT1 和 INT2 其中一个的对应位上是 1 或两者的对应位上同时是 1 时“或”后的结果为 1。
INT1 /& INT2	整数	此运算的结果是 INT1 与 INT2 的按位“异或”。即, INT1 和 INT2 其中一个的对应位是 1 但非同时为 1 时“异或”后的结果为 1。
INT1 && INT2	整数	产生整数 INT1 和 INT2 的位元“与”。即, INT1 和 INT2 的对应位同时为 1 时进行这一运算后的结果才为 1。

表 29: CLEM 位元整数运算 (继续)

函数	结果	描述
$INT1 \ \&\&\sim\sim \ INT2$	整数	将对 $INT1$ 和 $INT2$ 位元求补后的结果进行位元“与”运算。即, $INT1$ 的对应位是 1 且 $INT2$ 的对应位是 0 时进行这一运算后的结果才为 1。此运算与 $INT1 \ \&\&(\sim\sim INT2)$ 运算的结果相同, 可用于清除在 $INT2$ 中设置的 $INT1$ 的位。
$INT \ll N$	整数	产生 $INT1$ 向左偏移 N 个位置的位元模式。如果 N 为负值, 则向右移。
$INT \gg N$	整数	产生 $INT1$ 向右偏移 N 个位置的位元模式。如果 N 为负值, 则向左移。
$INT1 \ \&\&=_0 \ INT2$	布尔值	等价于布尔表达式 $INT1 \ \&\& \ INT2 \ /== \ 0$, 但更有效。
$INT1 \ \&\&/=_0 \ INT2$	布尔值	等价于布尔表达式 $INT1 \ \&\& \ INT2 \ == \ 0$, 但更有效。
<code>integer_bitcount(INT)</code>	整数	计算 INT 的二补数表示中 1 位或 0 位的数目。如果 INT 是非负数, 那么 N 为 1 的个数。如果 INT 为负数, 那么 N 为 0 的个数。由于符号扩展的原因, 非负整数中将有无数个 0 位, 而负整数中将有无数个 1 位。对于此运算, 始终有 $integer_bitcount(INT) = integer_bitcount(-(INT+1))$ 成立。
<code>integer_leastbit(INT)</code>	整数	返回整数 INT 中最低有效位组的数位位置 N 。 N 是保证可以对 INT 进行整除的条件下, 2 的最高幂。
<code>integer_length(INT)</code>	整数	以二补整数形式返回 INT 的长度 (以位计)。即, N 是最小整数, 因此 $INT < (1 \ll N)$ if $INT \geq 0$ $INT \geq (-1 \ll N)$ if $INT < 0$ 。如果 INT 为非负数, 则表示为无符号整数的 INT 需要一个至少 N 位的字段。另外, 最少需要用 $N+1$ 个数位来表示有符号整数 INT (不论符号是正是负)。
<code>testbit(INT, N)</code>	布尔值	测试整数 INT 中位置 N 处的位, 并以布尔值形式返回第 N 位的状态 (true 表示 1, false 表示 0)。

随机函数

以下函数用于随机选择项目或随机生成数值。

表 30: CLEM 随机函数

函数	结果	描述
<code>oneof(LIST)</code>	任意	返回一个从 $LIST$ 中随机选取的元素。应以 $[ITEM1, ITEM2, \dots, ITEM_N]$ 的形式输入列表项。注意, 还可以指定字段名列表。
<code>random(NUM)</code>	数字	返回类型相同 (INT 或 $REAL$) 的均匀分布随机数, 其范围从 1 至 NUM 。如果使用的是整数, 则仅返回整数。如果使用 (十进制) 实数, 则返回实数 (由流选项确定的十进制精度)。此函数所返回的最大随机数可以等于 NUM 。
<code>random0(NUM)</code>	数字	这具有与 <code>random(NUM)</code> 相同的属性, 但从 0 开始。该函数返回的最大随机数将永远不等于 NUM 。

字符串函数

在 CLEM 中您可以对字符串执行以下操作：

- 比较字符串
- 创建字符串
- 访问字符

在 CLEM 中，字符串是匹配的双引号 ("string quotes") 之间的任何字符序列。字符 (CHAR) 可以是任何单个字母数字字符。它们在 CLEM 表达式中使用 `` 格式的单反引号进行声明，例如 `z`、`A` 或 `2`。如果字符超出边界或为字符串中的负指数，则会导致未定义的行为。

注：在使用与不使用 SQL 回送的字符串之间的对比可能生成存在尾部空格的不同结果。

函数	结果	描述
<code>allbutfirst(N, STRING)</code>	String	返回字符串删除前 N 个字符的字符值。
<code>allbutlast(N, STRING)</code>	String	返回字符串 字符串 移除后 N 个字符后所生成的新字符串。
<code>alphabefore(STRING1, STRING2)</code>	布尔值	用于检查字符串的数字字母顺序。如果 <i>STRING1</i> 在 <i>STRING2</i> 之前，则返回真值。
<code>endstring(LENGTH, STRING)</code>	String	从指定字符串中抽取后 N 个字符。如果字符串长度小于或等于指定长度，则字符串不发生改变。
<code>hasendstring(STRING, SUBSTRING)</code>	整数	此函数与 <code>isendstring(SUBSTRING, STRING)</code> 相同。
<code>hasmidstring(STRING, SUBSTRING)</code>	整数	此函数的功能等同于 <code>ismidstring(SUBSTRING, STRING)</code> (嵌入的子串)。
<code>hasstartstring(STRING, SUBSTRING)</code>	整数	此函数的功能等同于 <code>isstartstring(SUBSTRING, STRING)</code> 。
<code>hassubstring(STRING, N, SUBSTRING)</code>	整数	此函数的功能等同于 <code>issubstring(SUBSTRING, N, STRING)</code> ，其中 N 的缺省值为 1。
<code>count_substring(STRING, SUBSTRING)</code>	整数	返回字符串中指定字符串出现的次数。例如， <code>count_substring("foooo.txt", "oo")</code> 返回结果为 3。
<code>hassubstring(STRING, SUBSTRING)</code>	整数	此函数的功能等同于 <code>issubstring(SUBSTRING, 1, STRING)</code> ，其中 N 的缺省值为 1。
<code>isalphacode(CHAR)</code>	布尔值	如果 <i>CHAR</i> 是字符代码为字母的指定字符串 (常为字段名) 中的某个字符，则返回真值。否则，此函数将返回值 0。例如， <code>isalphacode(produce_num(1))</code> 。

表 31: CLEM 字符串函数 (继续)

函数	结果	描述
isendstring(SUBSTRING, STRING)	整数	如果字符串 字符串 以子串 SUBSTRING 结尾, 则该函数将返回字符串 字符串 中子串 SUBSTRING 的整数下标。否则, 此函数将返回值 0。
islowercode(CHAR)	布尔值	如果 CHAR 是指定字符串 (通常为字段名) 中的某个小写字母字符, 则将返回真值。否则, 此函数将返回值 0。例如, islowercode(``) 和 islowercode(country_name(2)) 都是有效的表达式。
ismidstring(SUBSTRING, STRING)	整数	如果 SUBSTRING 是字符串的子串, 但首字符或尾端字符与字符串不同, 则该函数将返回此子串的起始下标。否则, 此函数将返回值 0。
isnumbercode(CHAR)	布尔值	如果指定字符串 (通常为字段名) 的 CHAR 为数字字符, 则返回真值。否则, 此函数将返回值 0。例如, isnumbercode(product_id(2))。
isstartstring(SUBSTRING, STRING)	整数	如果字符串 STRING 以子串 SUBSTRING 开头, 那么此函数将返回下标 1。否则, 此函数将返回值 0。
issubstring(SUBSTRING, N, STRING)	整数	从字符串 字符串 中搜索 (从第 N 个字符开始) 等于 SUBSTRING 的子串。如果找到子串, 则该函数返回匹配子串的起始整数下标。否则, 此函数将返回值 0。如果未指定 N, 那么此函数将缺省为 1。
issubstring(SUBSTRING, STRING)	整数	从字符串 字符串 中搜索 (从第 N 个字符开始) 等于 SUBSTRING 的子串。如果找到子串, 则该函数返回匹配子串的起始整数下标。否则, 此函数将返回值 0。如果未指定 N, 那么此函数将缺省为 1。
issubstring_count(SUBSTRING, N, STRING):	整数	返回字符串中第 N 次出现的 SUBSTRING 的索引。如果 SUBSTRING 的出现次数少于 N 次, 则返回结果为 0。
issubstring_lim(SUBSTRING, N, STARTLIM, ENDLIM, STRING)	整数	此函数的功能与 issubstring 相同, 但匹配必须从下标 STARTLIM 或其之前开始, 并且必须从下标 ENDLIM 或其之前结束。通过为任一参数提供值 false (例如, issubstring_lim(SUBSTRING, N, false, false, STRING) 与 issubstring 相同), 可以禁用 STARTLIM 或 ENDLIM 约束。
isuppercode(CHAR)	布尔值	如果 CHAR 是大写字母字符, 则返回真值。否则, 此函数将返回值 0。例如, isuppercode(``) 和 isuppercode(country_name(2)) 都是有效的表达式。

表 31: CLEM 字符串函数 (继续)

函数	结果	描述
last(CHAR)	String	返回字符串（长度至少为一个字符）的最后一个字符 CHAR。
length(STRING)	整数	返回字符串 字符串 的长度，即字符串内的字符数。
locchar(CHAR, N, STRING)	整数	<p>用于标识号字段中的字符位置。此函数将在字符串 字符串 中搜索字符 CHAR（从字符串的第 N 个字符开始搜索）。此函数的返回值表示（从 N 开始）找到字符的位置。如果找不到该字符，那么此函数将返回值 0。如果该函数具有无效偏移量 (N)（例如，超出字符串长度的偏移量），那么此函数将返回 \$null\$。</p> <p>例如，locchar(`n`, 2, web_page) 会在名为 web_page 的字段中搜索以字段值中的第二个字符开头的 `n` 字符。</p> <p>注：请用反单引号将指定字符引起。</p>
locchar_back(CHAR, N, STRING)	整数	<p>与 locchar 函数类似，所不同的是搜索反向进行并且从第 N 个字符开始。例如，locchar_back(`n`, 9, web_page) 将从 web_page 的第 9 个字符开始搜索并且是从后向前移动搜索。如果此函数存在无效偏移（例如，某个偏移超出了字符串长度），则函数返回结果为 \$null\$。理想情况下，您应该将 locchar_back 与函数 length(<field>) 结合使用，以动态使用该字段的当前值的长度。例如 locchar_back(`n`, (length(web_page)), web_page)。</p>
lowertoupper(CHAR) lowertoupper (STRING)	CHAR 或字符串	<p>可以输入字符串或字符，此函数将返回一个类型相同项，小写字符将转换为相应的大写字符。例如，lowertoupper(`a`)、lowertoupper("My string") 和 lowertoupper(field_name(2)) 都是有效的表达式。</p>
matches	布尔值	<p>如果字符串与指定模式相匹配，则将返回真值。模式必须是字符串字符，而不能是包含模式的字段名。问号 (?) 可以包含在模式中以便匹配某一字符，星号 (*) 可以匹配零个或多个字符。要匹配文字问号或星号（而不是将其用作通配符），可以使用反斜杠 (\) 作为转义符号。</p>
replace(SUBSTRING, NEWSUBSTRING, STRING)	String	<p>在指定字符串中，用 NEWSUBSTRING 替换 SUBSTRING 中的所有示例。</p>
replicate(COUNT, STRING)	String	<p>返回一个由原始字符串指定个数的复本所构成的字符串。</p>

表 31: CLEM 字符串函数 (继续)

函数	结果	描述
stripchar(CHAR, STRING)	String	借助此函数, 您可从某个字符串 或字段中移除指定字符。例如, 您可以使用此函数从数据中移除多余符号 (如货币符号) 以获取一个简单的数值或名称。例如, 通过语法 stripchar('\$', 'Cost') 可以移除所有值中的美元符号并返回一个新字段。 注: 请用反单引号将指定字符引起。
skipchar(CHAR, N, STRING)	整数	在字符串 字符串 中搜索除 CHAR 以外的字符 (从第 N 个字符开始搜索)。此函数返回一个表示发现子串位置的整数子串, 如果从第 N 个位置开始的每个字符是 CHAR, 则返回 0。如果此函数存在无效偏移 (例如, 某个偏移超出了字符串长度), 则函数返回结果为 \$null\$。 locchar 函数经常与 skipchar 结合使用以确定 N (开始搜索字符串的位置) 的值。例如 skipchar('s', (locchar('s', 1, "MyString")), "MyString")。
skipchar_back(CHAR, N, STRING)	整数	与 skipchar 函数类似, 所不同的是搜索 反向 进行并且从第 N 个字符开始。
startstring(LENGTH, STRING)	String	从指定字符串中抽取前 N 个字符。如果字符串长度小于或等于指定长度, 则字符串不发生改变。
strmember(CHAR, STRING)	整数	等同于 locchar(CHAR, 1, STRING) 函数。它返回一个整数子串, 指示 CHAR 首次出现的点或 0。如果此函数存在无效偏移 (例如, 某个偏移超出了字符串长度), 则函数返回结果为 \$null\$。
subscrs(N, STRING)	Char	返回输入字符串 字符串 的第 N 个字符 CHAR。此函数也可采用简写形式 STRING(N)。例如, lowertoupper("name"(1)) 是有效的表达式。
substring(N, LEN, STRING)	String	返回从位于下标 N 的字符开始的字符串 SUBSTRING, 此子串由字符串 字符串 的 LEN 个字符组成。
substring_between(N1, N2, STRING)	String	返回 字符串 的子串, 该字符串从下标 N1 开始并结束于下标 N2。
trim(STRING)	String	移除指定字符串的前导和尾部空格。
trim_start(STRING)	String	移除指定字符串的前导空格。
trimend(STRING)	String	移除指定字符串的尾部空格。
unicode_char(NUM)	Char	输入必须是十进制值, 而不能是十六进制值。返回 Unicode 值为 NUM 的字符。
unicode_value(CHAR)	NUM	返回 CHAR 的 Unicode 值。

表 31: CLEM 字符串函数 (继续)

函数	结果	描述
uppertolower(CHAR) uppertolower (STRING)	CHAR 或字符串	可以输入字符串或字符，此函数将返回一个类型相同项，大写字符将转换为相应的小写字符。 注：指定字符串时一定要用英文双引号，而指定字符时一定要用反向单引号。简单字段名不必使用引号。

SoundEx 函数

SoundEx 是在知道字符串的发音却不知其准确拼写的情况下用来查找字符串的一种方法。此种方法在 1918 年提出，它在对特定字母发音进行语音假定的基础上搜索出类似发音的字。此种方法可用于在数据库中搜索名称，例如相似名称的拼写和发音可能有所不同。基本 SoundEx 算法记录在多个源中，尽管存在已知限制（例如，虽然 ph 和 f 之类的前导字母组合听起来相同，但它们并不匹配），但大多数数据库都以某种形式支持此算法。

表 32: CLEM SoundEx 函数

函数	结果	描述
soundex(STRING)	整数	返回指定字符串的四字符 SoundEx 代码。
soundex_difference(STRING1, STRING2)	整数	返回 0 和 4 之间的某个整数，此整数表示两个字符串中 SoundEx 编码相同的字符的个数，0 表示不相似，4 表示两个字符串极相似或完全相同。

日期和时间函数

CLEM 包含一系列处理含有字符串日期时间存储变量的字段的函数，这些字符串变量代表日期和时间。可在“流属性”对话框中指定具体到每个流的日期和时间格式。日期和时间函数根据当前选定的格式来解析日期和时间字符串。

如果用两位数指定日期中的年（即未指定世纪），则 IBM SPSS Modeler 将采用“流属性”对话框中所指定的缺省世纪。

注：如果将数据函数后推到 SQL 或 IBM SPSS Analytic Server，那么在 Analytic Server 数据源后面的分支中，该数据内的任何数据格式字符串 (to_date) 都必须与 SPSS Modeler 流中指定的数据格式相匹配。

表 33: CLEM 日期及时间函数

函数	结果	描述
@TODAY	String	如果在流属性对话框中选中了 翻转日/分钟 ，那么此函数将采用当前日期格式的当前日期字符串。如果使用两位数的日期格式，并且未选择 回滚天数/分钟数 ，那么此函数将在当前服务器上返回 \$null\$。
to_time(ITEM)	时间	将指定字段的存储值转换为时间。
to_date(ITEM)	date	将指定字段的存储值转换为日期。
to_timestamp(ITEM)	时间戳记	将指定字段的存储值转换为时间戳记。
to_datetime(ITEM)	日期/时间	将指定字段的存储值转换为日期、时间或时间戳记值。

表 33: CLEM 日期及时间函数 (继续)

函数	结果	描述
datetime_date (ITEM)	date	返回数字、字符串或时间戳记的日期值。请注意，这是唯一允许您将数字（以秒数记）转换回日期的函数。如果 ITEM 为字符串，将以当前日期格式解析字符串来创建日期。流属性对话框中指定的日期格式必须正确，这样此函数才能成功执行。如果 ITEM 是一个数字，那么会将其解释为自基本日期（或戳记）以来的秒数。不足一天的部分将截断。如果 ITEM 为时间戳记，将返回时间戳记的日期部分。如果 ITEM 为日期，将返回未更改的日期。
date_before (DATE1, DATE2)	布尔值	如果 DATE1 所表示的日期或时间戳记在 DATE2 之前，那么返回真值。否则，此函数将返回值 0。
date_days_difference (DATE1, DATE2)	整数	以整数的形式返回从日期或时间戳记 DATE1 到日期或时间戳记 DATE2 的天数。如果 DATE2 在 DATE1 之前，那么该函数返回负值。
date_in_days (DATE)	整数	以整数形式返回从基线日期到日期或时间戳记 DATE 的天数。如果 DATE 在基线日期之前，那么该函数返回负数。为了正确进行计算，必须包括有效日期。例如，不应将日期指定为 2001 年 2 月 29 日。因为 2001 年并非闰年，所以上述日期不存在。
date_in_months (DATE)	实数	以实数形式返回从基线日期到日期或时间戳记 DATE 的月数。这是基于每月 30.4375 天的近似数字。如果 DATE 在基线日期之前，那么该函数返回负数。为了正确进行计算，必须包括有效日期。例如，不应将日期指定为 2001 年 2 月 29 日。因为 2001 年并非闰年，所以上述日期不存在。
date_in_weeks (DATE)	实数	以实数形式返回从基线日期到日期或时间戳记 DATE 的周数。这基于每周 7.0 天。如果 DATE 在基线日期之前，那么该函数返回负数。为了正确进行计算，必须包括有效日期。例如，不应将日期指定为 2001 年 2 月 29 日。因为 2001 年并非闰年，所以上述日期不存在。
date_in_years (DATE)	实数	以实数形式返回从基线日期到日期或时间戳记 DATE 的年数。这是基于每年 365.25 天的近似数字。如果 DATE 在基线日期之前，那么该函数返回负数。为了正确进行计算，必须包括有效日期。例如，不应将日期指定为 2001 年 2 月 29 日。因为 2001 年并非闰年，所以上述日期不存在。
date_months_difference (DATE1, DATE2)	实数	以实数形式返回从日期或时间戳记 DATE1 到日期或时间戳记 DATE2 的月数。这是基于每月 30.4375 天的近似数字。如果 DATE2 在 DATE1 之前，那么该函数返回负值。
datetime_date (YEAR, MONTH, DAY)	date	创建给定 YEAR、MONTH、DAY 的日期值。参数必须为整数。
datetime_day (DATE)	整数	对给定的 DATE 或时间戳记，返回其所在月份的天数。结果为范围 1 到 31 内的一个整数。
datetime_day_name (DAY)	String	返回给定 DAY 的全称。参数必须是 1（周日）至 7

表 33: CLEM 日期及时间函数 (继续)

函数	结果	描述
datetime_hour(TIME)	整数	返回某个 <i>TIME</i> 或时间戳记的小时部分。结果为 0 到 23 范围内的整数。
datetime_in_seconds(TIME)	实数	返回 <i>TIME</i> 存储的秒钟部分。
datetime_in_seconds(DATE), datetime_in_seconds(DATE TIME)	实数	返回从当前 <i>DATE</i> 或 <i>DATETIME</i> 到基线日期 (1900-01-01) 之间的差值的累积秒数。
datetime_minute(TIME)	整数	返回某个 <i>TIME</i> 或时间戳记的分钟部分。结果为 0 到 59 范围内的整数。
datetime_month(DATE)	整数	返回某个 <i>DATE</i> 或时间戳记的月份部分。结果为 1 到 12 范围内的整数。
datetime_month_name (MONTH)	<i>String</i>	返回给定 <i>MONTH</i> 的全称。参数必须为范围 1 到 12 内的整数。
datetime_now	时间戳记	以时间戳记形式返回当前时间。
datetime_second(TIME)	整数	返回某个 <i>TIME</i> 或时间戳记的秒钟部分。结果为 0 到 59 范围内的整数。
datetime_day_short_name (DAY)	<i>String</i>	返回给定 <i>DAY</i> 的简称。参数必须是 1 (周日) 至 7
datetime_month_short_name (MONTH)	<i>String</i>	返回给定 <i>MONTH</i> 的简称。参数必须为范围 1 到 12 内的整数。
datetime_time(HOUR, MINUTE, SECOND)	时间	返回给定 <i>HOUR</i> 、 <i>MINUTE</i> 和 <i>SECOND</i> 的时间值。参数必须为整数。
datetime_time(ITEM)	时间	返回给定 <i>ITEM</i> 的时间值。
datetime_timestamp(YEAR, MONTH, DAY, HOUR, MINUTE, SECOND)	时间戳记	返回指定 <i>YEAR</i> 、 <i>MONTH</i> 、 <i>DAY</i> 、 <i>HOUR</i> 、 <i>MINUTE</i> 以及 <i>SECOND</i> 的时间戳记值。
datetime_timestamp(DATE , TIME)	时间戳记	返回给定 <i>DATE</i> 和 <i>TIME</i> 的时间戳记值。
datetime_timestamp (NUMBER)	时间戳记	返回给定秒数的时间戳记值。
datetime_weekday(DATE)	整数	对给定 <i>DATE</i> 或时间戳记, 返回其所在周的天数。
datetime_year(DATE)	整数	返回某个 <i>DATE</i> 或时间戳记的年份部分。结果为一个整数, 如 2002。
date_weeks_difference (DATE1, DATE2)	实数	以实数形式返回从日期或时间戳记 <i>DATE1</i> 到日期或时间戳记 <i>DATE2</i> 的周数。这基于每周 7.0 天。如果 <i>DATE2</i> 在 <i>DATE1</i> 之前, 那么该函数返回负值。
date_years_difference (DATE1, DATE2)	实数	以实数形式返回从日期或时间戳记 <i>DATE1</i> 到日期或时间戳记 <i>DATE2</i> 的年数。这是基于每年 365.25 天的近似数字。如果 <i>DATE2</i> 在 <i>DATE1</i> 之前, 那么该函数返回负值。
date_from_ywd(YEAR, WEEK, DAY)	整数	将年份、年中的第几个星期以及星期几转换为使用 ISO 8601 标准的日期。

表 33: CLEM 日期及时间函数 (继续)

函数	结果	描述
date_iso_day(DATE)	整数	返回使用 ISO 8601 标准的日期中的星期几。
date_iso_week(DATE)	整数	返回使用 ISO 8601 标准的日期中的年中第几个星期。
date_iso_year(DATE)	整数	返回使用 ISO 8601 标准的日期中的年份。
time_before(TIME1, TIME2)	布尔值	如果 <i>TIME1</i> 所表示的时间或时间戳记在 <i>TIME2</i> 之前, 则返回真值。否则, 此函数将返回值 0。
time_hours_difference(TIME1, TIME2)	实数	以实数形式返回时间或时间戳记 <i>TIME1</i> 与 <i>TIME2</i> 之间的小时差。如果在“流属性”对话框中选择 翻转日/分钟 , 则用 <i>TIME1</i> 的较大值表示前一天。如果不选择翻转日选项, 则较大的 <i>TIME1</i> 值将造成返回值为负数。
time_in_hours(TIME)	实数	以实数的形式返回 <i>TIME</i> 所表示的小时数。例如, 采用时间格式 HHMM 时, 表达式 <code>time_in_hours('0130')</code> 的返回值为 1.5。 <i>TIME</i> 可表示时间或时间戳。
time_in_mins(TIME)	实数	以实数的形式返回 <i>TIME</i> 所表示的分钟数。 <i>TIME</i> 可表示时间或时间戳。
time_in_secs(TIME)	整数	以实数的形式返回 <i>TIME</i> 所表示的秒数。 <i>TIME</i> 可表示时间或时间戳。
time_mins_difference(TIME1, TIME2)	实数	以实数形式返回时间或时间戳记 <i>TIME1</i> 与 <i>TIME2</i> 之间的分钟差。如果在“流选项”对话框中选择 翻转日/分钟 , 则将采用 <i>TIME1</i> 的较大值来指代前一天 (仅当在当前格式中指定分和秒时, 指代前一小时)。如果不选择翻转日选项, <i>TIME1</i> 的较大值会造成返回值为负数。
time_secs_difference(TIME1, TIME2)	整数	返回 <i>TIME1</i> 和 <i>TIME2</i> 所表示的时间或时间戳记之间的时差秒数 (整数)。如果在“流选项”对话框中选择 翻转日/分钟 , 则将采用 <i>TIME1</i> 的较大值来指代前一天 (仅当在当前格式中指定分和秒时, 指代前一小时)。如果不选择翻转日选项, 则较大的 <i>TIME1</i> 值将造成返回值为负数。

转换日期和时间值

请注意, 转换函数 (及其他要求特定类型输入 (如日期或时间值) 的函数) 取决于“流选项”对话框中指定的当前格式。例如, 如果某个名为 *DATE* 的字段是以 *Jan 2003* 和 *Feb 2003* 等值的字符串进行存储的, 则可以使用以下函数将其存储为日期格式:

```
to_date( DATE )
```

为了使此转换生效, 请选择匹配的日期格式 **MON YYYY** 作为流的缺省日期格式。

例如, 若要使用填充节点将字符串值转换为日期, 请参阅流 *broadband_create_models.str*, 它安装在 *streams* 子文件夹下的 *|Demos* 文件夹中。

存储为数字的日期。 请注意, 上一个示例中的 *DATE* 是字段的名称, 而 *to_date* 则是 CLEM 函数。如果您将日期存储为数字, 那么可以使用 *datetime_date* 函数对其进行转换, 其中数字被解释为自基本日期 (或戳记) 以来的秒数。

```
datetime_date( DATE )
```

通过将日期转换为秒数（再反向转换），您可以进行用当前日期加减固定天数的计算，例如：

```
datetime_date((date_in_days(DATE)-7)*60*60*24)
```

序列函数

对于某些操作而言，事件序列很重要。通过该应用程序，您可以处理以下记录序列：

- 序列和时间序列
- 序列函数
- 记录索引
- 求值的平均数、总和以及对值进行比较
- 监视变更--区别
- @SINCE
- 偏移值
- 其它序列工具

对于许多应用程序，每个通过流传递的记录可以看成是独立于所有其它记录的单观测值例。此种情况下，记录的顺序通常并不重要。

然而对于某些类别的问题，记录序列非常重要。这些一般是发生在时间序列中的情况，其中记录序列代表事件的有序排列。每条记录代表着时间中的某个特定时刻的快照，大部分最重要的信息可能并未包含在瞬时值中，而是包含在随时间的流逝不断变更发展的方式中。

当然，相关参数可能并不是时间。例如，记录可以代表沿着某条线的不同位置对事物所采取的分析，但也将应用相同的原则。

通过以下特征可以立即识别出序列和特殊函数：

- 它们都以@为前缀。
- 函数名称采用大写。

序列函数可以引用节点当前处理的记录、已通过节点的记录、甚至是（某种情况下）尚未通过节点的记录。序列函数可与其他 CLEM 表达式部分自由组合，虽然某些函数对参数有所限制。

示例

您会发现了解自某一特定事件发生或条件为真以来的时间长度非常有用。使用函数 @SINCE 来执行此操作 - 例如：

```
@SINCE(Income > Outgoings)
```

该函数返回最后一条满足此条件的记录的偏移量，即在此记录之前满足条件的记录数。如果条件从未为 true，那么 @SINCE 将返回 @INDEX + 1。

有时，您可能希望在 @SINCE 所使用的表达式中引用当前记录的值。您可以使用函数 @THIS 来执行此操作，此函数指定字段名称始终应用于当前记录。要查找 Concentration 字段值为当前记录两倍以上的一条记录的偏移量，您可以使用：

```
@SINCE(Concentration > 2 * @THIS(Concentration))
```

在某些情况下，按定义为 @SINCE 指定的当前记录的条件为 true - 例如：

```
@SINCE(ID == @THIS(ID))
```

基于这种原因，@SINCE 函数将不对当前记录的条件求值。如果要对当前记录以及前面记录的条件求值，请使用类似函数 @SINCE0；如果当前记录的条件为真，则 @SINCE0 返回值为 0。

表 34: CLEM 序列函数

函数	结果	描述
MEAN(FIELD)	实数	返回指定字段 <i>FIELD</i> 或 <i>FIELDS</i> 的值的均值。
@MEAN(FIELD, EXPR)	实数	返回当前节点所接收的最后 <i>EXPR</i> 条记录 (包括当前记录) 中的 <i>FIELD</i> 值的均值。 <i>FIELD</i> 必须是数值型字段的名称。 <i>EXPR</i> 可以是值为大于 0 的整数的任意表达式。 如果省略了 <i>EXPR</i> , 或者超过了到目前为止接收的记录数, 那么将返回到目前为止接收的所有记录的平均值。
@MEAN(FIELD, EXPR, INT)	实数	返回当前节点所接收的最后 <i>EXPR</i> 条记录 (包括当前记录) 中的 <i>FIELD</i> 值的均值。 <i>FIELD</i> 必须是数值型字段的名称。 <i>EXPR</i> 可以是值为大于 0 的整数的任意表达式。 如果省略了 <i>EXPR</i> , 或者超过了到目前为止接收的记录数, 那么将返回到目前为止接收的所有记录的平均值。 <i>INT</i> 指定要向后查看的值的最大数量。 这比使用两个参数更为有效。
@DIFF1(FIELD)	实数	返回 <i>FIELD</i> 的一阶微分差。 因此单自变量形式将返回该字段当前值和前一个值的差。 如果先前的相关记录不存在, 那么返回 \$null\$ 。
@DIFF1(FIELD1, FIELD2)	实数	两参数形式给出了 <i>FIELD1</i> 对 <i>FIELD2</i> 的一阶微分差。 如果先前的相关记录不存在, 那么返回 \$null\$ 。 此值计算为 @DIFF1(FIELD1)/@DIFF1(FIELD2)。
@DIFF2(FIELD)	实数	返回 <i>FIELD</i> 的二阶微分差。 因此单自变量形式将返回该字段当前值和前一个值的差。 如果先前的相关记录不存在, 那么返回 \$null\$ 。 @DIFF2 计算为 @DIFF(@DIFF(FIELD))。
@DIFF2(FIELD1, FIELD2)	实数	两参数形式给出了 <i>FIELD1</i> 对 <i>FIELD2</i> 的二阶微分差。 如果先前的相关记录不存在, 那么返回 \$null\$ 。 这是一个复杂的计算 -- @DIFF1(FIELD1)/@DIFF1(FIELD2) - @OFFSET(@DIFF1(FIELD1),1)/@OFFSET(@DIFF1(FIELD2)) / @DIFF1(FIELD2)。
@INDEX	整数	返回当前记录的索引。 记录到达当前节点时, 即将索引分配给相应的记录。 为第一条记录分配索引 1, 每条后续记录的索引以 1 递增。
@LAST_NON_BLANK(FIELD)	任意	根据上游源或类型节点中的定义, 返回不是空白值的 <i>FIELD</i> 的最后一个值。 如果目前所读取的记录中的 <i>FIELD</i> 没有非空白值, 则返回 \$null\$ 。 注意, 可以为每个字段单独定义空白值 (又称用户缺失值)。
@MAX(FIELD)	数字	返回指定 <i>FIELD</i> 的最大值。
@MAX(FIELD, EXPR)	数字	返回目前为止所接收的最后 <i>EXPR</i> 条记录 (包括当前记录) 中的 <i>FIELD</i> 字段的最大值。 <i>FIELD</i> 必须是数值型字段的名称。 <i>EXPR</i> 可以是值为大于 0 的整数的任意表达式。

表 34: CLEM 序列函数 (继续)

函数	结果	描述
@MAX(FIELD, EXPR, INT)	数字	返回目前为止所接收的最后 EXPR 条记录 (包括当前记录) 中 FIELD 的最大值。FIELD 必须是数值型字段的名称。EXPR 可以是值为大于 0 的整数的任意表达式。如果省略了 EXPR, 或者超过了到目前为接收的记录数, 那么将返回到目前为接收的所有记录中的最大值。INT 指定要向后查看的值的最大数量。这比使用两个参数更为有效。
@MIN(FIELD)	数字	返回指定 FIELD 的最小值。
@MIN(FIELD, EXPR)	数字	返回目前为止所接收的最后 EXPR 条记录 (包括当前记录) 中的 FIELD 的最小值。FIELD 必须是数值型字段的名称。EXPR 可以是值为大于 0 的整数的任意表达式。
@MIN(FIELD, EXPR, INT)	数字	返回目前为止所接收的最后 EXPR 条记录 (包括当前记录) 中的 FIELD 的最小值。FIELD 必须是数值型字段的名称。EXPR 可以是值为大于 0 的整数的任意表达式。如果省略了 EXPR, 或者超过了到目前为接收的记录数, 那么将返回到目前为接收的所有记录中的最小值。INT 指定要向后查看的值的最大数量。这比使用两个参数更为有效。
@OFFSET(FIELD, EXPR)	任意	<p>返回当前记录中值为 EXPR 的记录偏移量中的 FIELD 值。正偏移引用已经经过的记录 (即“lookback”), 而负偏移指定尚未到达的记录 (即“lookahead”)。例如, @OFFSET(Status, 1) 将返回先前记录中 Status 字段的值, 而 @OFFSET(Status, -4) 则“向前查看”序列中的四条记录 (即, 尚未通过此节点的记录) 以获取该值。请注意, 必须将负 (向前检查) 偏移指定为常数。 (仅限于正偏移) EXPR 还可以是任意的 CLEM 表达式, 将通过此表达式给出当前记录的偏移值。在此情况下, 采用该函数的三参数版本可以提高性能 (参阅下个函数)。如果表达式返回的不是非负整数, 将出现错误 - 即, 计算的期待偏移量不合法。</p> <p>注: 自引用的 @OFFSET 函数无法使用文字前瞻。例如, 在填充节点中, 不能使用 @OFFSET(field1, -2) 之类的表达式来替换 field1 的值。</p> <p>注: 在“填充”节点中填充字段时, 实际上存在该字段的两个不同值, 即填充前的值和填充后的值。当 @OFFSET 引用自身时, 将引用填充后的值。仅对于过去的行, 才存在填充后的值, 因此自引用 @OFFSET 只能引用过去的行。由于自引用 @OFFSET 不能引用未来的行, 因此将对偏移量执行下列检查:</p> <ul style="list-style-type: none"> • 如果偏移量为文字值, 并且指向未来, 那么将在执行开始前报告错误。 • 如果偏移量为表达式并且在运行时求值为未来, 那么 @OFFSET 将返回 \$null\$。 <p>注: 不支持在一个节点中同时使用“lookahead”和“lookback”。</p>

表 34: CLEM 序列函数 (继续)

函数	结果	描述
@OFFSET(FIELD, EXPR, INT)	任意	与 @OFFSET 函数执行相同的操作，但附带第三个参数 <i>INT</i> （用于指定向后查看值的最大数量）。当从表达式计算偏移时，采用第三个参数可以提高性能。 例如，在诸如 @OFFSET(Foo, Month, 12) 之类的表达式中，系统知道仅保留 Foo 的最后 12 个值；否则，它必须存储每个值以防万一。如果偏移值为常数包括必须为常数的负“期待”偏移量第三个参数将毫无意义，此时应当使用该函数的双参数版本。另请参阅在前述双参数版本中自引用函数的说明。 注：不支持在一个节点中同时使用“lookahead”和“lookback”。
@SDEV(FIELD)	实数	返回指定的 <i>FIELD</i> 或 <i>FIELDS</i> 值的标准差。
@SDEV(FIELD, EXPR)	实数	返回当前节点所接收的最后 <i>EXPR</i> 条记录（包括当前记录）中的 <i>FIELD</i> 值的标准偏差。 <i>FIELD</i> 必须是数值型字段的名称。 <i>EXPR</i> 可以是值为大于 0 的整数的任意表达式。如果省略了 <i>EXPR</i> ，或者超过了到目前为止接收的记录数，那么将返回到目前为止接收的所有记录的标准偏差。
@SDEV(FIELD, EXPR, INT)	实数	返回当前节点所接收的最后 <i>EXPR</i> 条记录（包括当前记录）中的 <i>FIELD</i> 值的标准偏差。 <i>FIELD</i> 必须是数值型字段的名称。 <i>EXPR</i> 可以是值为大于 0 的整数的任意表达式。如果省略了 <i>EXPR</i> ，或者超过了到目前为止接收的记录数，那么将返回到目前为止接收的所有记录的标准偏差。 <i>INT</i> 指定要向后查看的值的最大数量。这比使用两个参数更为有效。
@SINCE(EXPR)	任意	如果 <i>EXPR</i> （可以是任何的 CLEM 表达式）的结果为真，则返回已传递的记录数。
@SINCE(EXPR, INT)	任意	添加第二个参数（ <i>INT</i> ）时，用于指定向后查看的最大记录数。如果 <i>EXPR</i> 结果不为真，则 <i>INT</i> 为 @INDEX+1。
@SINCE0(EXPR)	任意	对于当前记录，@SINCE 不返回任何值；而如果当前记录的 <i>EXPR</i> 结果为真，则 @SINCE0 返回 0。
@SINCE0(EXPR, INT)	任意	添加第二个参数（ <i>INT</i> ）时，用于指定向后查看的最大记录数。
@SUM(FIELD)	数字	返回指定字段 <i>FIELD</i> 或 <i>FIELDS</i> 的值的和。
@SUM(FIELD, EXPR)	数字	返回当前节点所接收的最后 <i>EXPR</i> 条记录（包括当前记录）中的 <i>FIELD</i> 值的和。 <i>FIELD</i> 必须是数值型字段的名称。 <i>EXPR</i> 可以是值为大于 0 的整数的任意表达式。如果省略了 <i>EXPR</i> ，或者超过了到目前为止接收的记录数，那么将返回到目前为止接收的所有记录的总和。

表 34: CLEM 序列函数 (继续)

函数	结果	描述
@SUM(FIELD, EXPR, INT)	数字	返回当前节点所接收的最后 EXPR 条记录（包括当前记录）中的 FIELD 值的和。FIELD 必须是数值型字段的名称。EXPR 可以是值为大于 0 的整数的任意表达式。如果省略了 EXPR，或者超过了到目前为止接收的记录数，那么将返回到目前为止接收的所有记录的总和。INT 指定要向后查看的值的最大数量。这比使用两个参数更为有效。
@THIS(FIELD)	任意	返回当前记录中名为 FIELD 的字段的价值。仅用于 @SINCE 表达式。

全局函数

函数 @MEAN、@SUM、@MIN、@MAX 和 @SDEV 最多可以处理读取的直至当前记录（包括当前记录）的所有记录。但在某些时候，它可用于检测如何将当前记录中的值与整个数据集中的值进行比较。如果使用设置全局量节点来生成整个数据集中的值，则您可在使用全局函数的 CLEM 表达式中访问这些值。

例如，

```
@GLOBAL_MAX(Age)
```

在数据集中返回 Age 的最高值，而表达式则

```
(Value - @GLOBAL_MEAN(Value)) / @GLOBAL_SDEV(Value)
```

将给出该记录 Value 和作为标准偏差的全局平均数之间的差。仅当设置全局量节点计算出全局值后，您方可使用它们。通过单击“流属性”对话框“全局量”选项卡中的 **Clear Global Values** 按钮，您可以撤消当前所有全局值。

表 35: CLEM 全局函数

函数	结果	描述
@GLOBAL_MAX(FIELD)	数字	返回超出整个数据集(先前由“设置全局量”节点生成)的 FIELD 的最大值。FIELD 必须是数字字段、日期/时间/日期时间字段或者字符串字段的名称。如果未设置相应的全局值，则将出现错误。
@GLOBAL_MIN(FIELD)	数字	返回超出整个数据集(先前由“设置全局量”节点生成)的 FIELD 的最小值。FIELD 必须是数字字段、日期/时间/日期时间字段或者字符串字段的名称。如果未设置相应的全局值，则将出现错误。
@GLOBAL_SDEV(FIELD)	数字	返回超出整个数据集(先前由“设置全局量”节点生成)的 FIELD 的标准差。FIELD 必须是数值型字段的名称。如果未设置相应的全局值，则将出现错误。
@GLOBAL_MEAN(FIELD)	数字	返回超出整个数据集(先前由“设置全局量”节点生成)的 FIELD 的值的均值。FIELD 必须是数值型字段的名称。如果未设置相应的全局值，则将出现错误。
@GLOBAL_SUM(FIELD)	数字	返回超出整个数据集(由“设置全局量”节点生成)的 FIELD 值的和。FIELD 必须是数值型字段的名称。如果未设置相应的全局值，则将出现错误。

处理空白值和空值的函数

借助 CLEM，您可以指定在某个字段中用作“空白值”或缺失值的特定值。下列函数用于处理空白值。

表 36: CLEM 空白值函数和 Null 值函数

函数	结果	描述
@BLANK(FIELD)	布尔值	根据上游类型节点或源节点 (“类型”选项卡) 中所设置的空白值处理规则, 对所有空白值的记录返回真值。
@LAST_NON_BLANK(FIELD)	任意	根据上游源或类型节点中的定义, 返回不是空白值的 <i>FIELD</i> 的最后一个值。如果目前所读取的记录中的 <i>FIELD</i> 没有非空白值, 则返回 \$null\$。注意, 可以为每个字段单独定义空白值 (又称用户缺失值)。
@NULL(FIELD)	布尔值	如果 <i>FIELD</i> 的值是系统缺失值, 则返回 true \$null\$。对于所有其他值 (包括用户定义的空白值), 则返回 false。如果要同时检查这两者, 请使用 @BLANK(FIELD) 和 @NULL(FIELD)。
undef	任意	通常用于在 CLEM 中输入 \$null\$ 值, 在过滤节点中用空值替代空白值。

在过滤节点中可以“填写”空白值字段。在“填充”节点和“派生”节点 (仅限多模式) 中, 特殊 CLEM 函数 @FIELD 引用当前正在检查的字段。

特殊字段

特殊函数用于指示所检查的具体字段, 或用于生成输入字段列表。例如, 当一次导出多个字段时, 应使用 @FIELD 函数来指示“对选定字段执行导出操作”。使用表达式 log(@FIELD) 可为每个选定字段派生一个新的日志字段。

表 37: CLEM 特殊字段

函数	结果	描述
@FIELD	任意	对表达式中指定的所有字段执行操作。
@TARGET	任意	在用户定义的分析函数中使用 CLEM 表达式时, @TARGET 表示正在分析的目标/预测对的目标字段或“正确的值”。此函数一般用于分析节点。
@PREDICTED	任意	当 CLEM 表达式用于用户定义的分析函数时, @PREDICTED 代表正在分析的目标/预测对的预测值。此函数一般用于分析节点。
@PARTITION_FIELD	任意	替换当前分区字段的名称。
@TRAINING_PARTITION	任意	返回当前训练分区的值。例如, 要用选择节点选择训练记录, 请使用 CLEM 表达式: @PARTITION_FIELD = @TRAINING_PARTITION 此表达式将确保无论数据中的每个分区使用哪个值表示, 选择节点都将照常工作。
@TESTING_PARTITION	任意	返回当前测试分区的值。
@VALIDATION_PARTITION	任意	返回当前验证分区的值。

表 37: CLEM 特殊字段 (继续)

函数	结果	描述
@FIELDS_BETWEEN(start, end)	任意	根据数据中字段的自然（即插入）顺序，返回指定开始和结束字段（包含）之间的字段名列表。
@FIELDS_MATCHING(pattern)	任意	<p>返回与指定模式相匹配的字段名的列表。问号 (?) 可以包含在模式中以便匹配某一字符，星号 (*) 可以匹配零个或多个字符。要匹配文字问号或星号（而不是将其用作通配符），可以使用反斜杠 (\) 作为转义符号。</p> <p>注: 这需要以字符串字面值为自变量；它不能使用嵌套表达式来生成该自变量。</p>
@MULTI_RESPONSE_SET	任意	返回已命名的多重响应集的字段列表。

第 11 章 将 IBM SPSS Modeler 与存储库配合使用

关于 IBM SPSS Collaboration and Deployment Services Repository

SPSS Modeler 可以与 IBM SPSS Collaboration and Deployment Services 存储库结合使用，使您能够管理数据挖掘模型和相关预测对象的生命周期，并使这些对象可供企业应用程序、工具和解决方案使用。可以通过此方式共享的 IBM SPSS Modeler 对象包括流、节点、流输出、项目和模型。这些对象存储在中央存储库中，可以和其他应用程序共享，也可以使用扩展的版本、元数据和搜索功能进行跟踪。

在您通过存储库使用 SPSS Modeler 之前，需要在存储库主机上安装一个适配器。没有该适配器，您在尝试从某些 SPSS Modeler 节点或模型访问存储库对象时会看到以下消息：

```
The repository may need updating to support new node, model and output types.
```

有关安装适配器的指示信息，请参阅产品下载过程中以 PDF 文件形式提供的《SPSS Modeler 部署安装》指南。在 SPSS Modeler 部署指南中提供了如何从 IBM SPSS Deployment Manager 访问 IBM SPSS Modeler 存储库对象的详细信息。

以下部分提供从 SPSS Modeler 中访问存储库的信息。

可扩展的版本和搜索支持

存储库提供了全面的对象版本和搜索功能。例如，假定创建了一个流，并将其存储到可与其他部门的研究人员共享的存储库中。如果稍后在 SPSS Modeler 中更新了此流，则可以将此更新后的版本添加到存储库中而不用覆盖以前的版本。可以始终访问所有的版本，并可以按照名称、标签、使用的字段或其他属性搜索这些版本。例如，可以搜索所有将净收入用作输入的模型版本，或搜索所有由特定程序设计员创建的模型。

（要在传统的文件系统中执行此操作，必须以不同的文件名保存每个版本，且各个版本间的关系对软件而言是未知的。）

单点登录

单点登录功能使用户能够连接到而不必每次都输入用户名和密码详细信息。用户现有的本地网络登录详细信息提供了 IBM SPSS Collaboration and Deployment Services 所需的必要认证信息。此功能取决于以下条件：

- IBM SPSS Collaboration and Deployment Services 必须配置以使用单点登录提供者。
- 用户必须登录到与提供者兼容的主机中。

有关更多信息，请参阅第 164 页的『[连接至存储库](#)』。

存储和部署存储库对象

在 IBM SPSS Modeler 中创建的流可以**存储**在存储库中，就像存储为具有扩展名 `.str` 的文件一样。这样，单个流便可由企业中的多个用户访问。有关更多信息，请参阅主题第 165 页的『[在存储库中存储对象](#)』。

另外，还可以在存储库中部署流。部署的流将存储为具有附加元数据的文件。部署的流可以充分利用 IBM SPSS Collaboration and Deployment Services 的企业级功能，例如自动化评分和模型刷新。例如，可以按定期计划间隔在新数据可用时对模型进行自动更新。另外，也可以部署一组流用于 Champion Challenger 分析，该分析将比较各个流以确定哪个流包含最有效的预测模型。

注：关联规则、STP 和 TCM 建模节点不支持 IBM SPSS Collaboration and Deployment Services 中的 Model Evaluation 或 Champion Challenger 步骤。

您可以部署流（扩展名为 `.str`）。部署为流将使该流能够由瘦客户端应用程序 IBM SPSS Modeler Advantage 使用。有关更多信息，请参阅主题第 179 页的『[打开 IBM SPSS Modeler Advantage 中的流](#)』。

有关更多信息，请参阅第 173 页的『[流部署选项](#)』。

其他部署选项

IBM SPSS Collaboration and Deployment Services 不仅为管理企业内容提供了极其广泛的功能，还提供部署或导出流的其他工具，包括：

- 使用 IBM SPSS Modeler Solution Publisher Runtime 导出流和模型，供以后使用。
- 导出一个或多个 PMML（用于对模型信息编码的基于 XML 的格式）中的模型。有关更多信息，请参阅主题 第 179 页的『导入和导出 PMML 模型』。

连接至存储库

1. 要连接到存储库，请在 IBM SPSS Modeler 主菜单中单击：

工具 > 存储库 > 选项...

2. 在 **存储库 URL** 字段中，输入或选择您要访问的存储库安装的目录路径或 URL。一次只能连接一个存储库。

每个站点或安装的设置都是特定的。有关特定的登录详细信息，请与本地系统管理员联系。

设置凭证。 不选中此复选框可启用单点登录功能，该功能尝试使您使用本地计算机用户名和密码详细信息登录。如果无法使用单点登录，或者选中此选项以禁用单点登录（例如，登录管理员帐户），那么会显示一个窗口让您输入凭证。

输入存储库的凭证

根据设置不同，在“存储库：凭证”对话框中，下列字段可能是必填项：

用户标识和密码。 指定用于登录的有效用户名和密码。如果需要，请联系本地管理员以获取更多信息。

提供程序。 选择一个安全提供者用于认证。存储库可配置使用不同的安全提供者；如有必要，请与本地管理员联系以了解详细信息。

记住存储库和用户标识。 将当前设置保存为缺省设置，以便在每次想要连接时不必再次加入这些设置。

浏览存储库凭证

从 Analytic Server、Cognos、ODBC 或 TM1 源节点连接到存储库时，可以选择先前记录的凭证来连接到存储库。这些凭证列在“**选择存储库凭证**”对话框中。要选择此对话框，请单击**凭证**字段旁边的“浏览”。

在“**选择存储库凭证**”对话框中，突出显示提供的列表中的凭证，然后单击“确定”。如果列表过大，请使用**过滤器**字段输入名称或名称的一部分，以查找所需的凭证。

浏览存储库内容

在存储库中，可以按照与 Windows Explorer 类似的方式浏览其中存储的内容；也可以浏览每个已存储对象的版本。

1. 要打开 IBM SPSS Collaboration and Deployment Services Repository 窗口，在 SPSS Modeler 菜单中，单击：

工具 > 存储库 > 浏览...

1. 如果需要，指定存储库的连接设置。有关更多信息，请参阅主题 第 164 页的『连接至存储库』。有关特定端口、密码和其他连接的详细信息，请与本地系统管理员联系。

浏览器窗口初始情况下显示文件夹层次结构的树形视图。单击文件夹名称可显示其内容。

此时，符合当前选择或搜索标准的对象将列在右侧窗格中，已选择版本的详细信息将显示在右下窗格中。显示的属性应用于最新的版本。

在存储库中存储对象

您可以将流、节点、模型、模型选项板、工程和输出对象保存在存储库中，其他用户和应用程序可从这里对这些内容进行访问。

您也可以其他用户通过使用 IBM SPSS Collaboration and Deployment Services Deployment Portal 在因特网上能够查看的格式将流输出发布到存储库。

设置对象属性

存储对象时，将显示“存储库：存储”对话框，以使您能够设置该对象的众多属性的值。您可以：

- 选择用于存储对象的名称和存储库文件夹
- 添加对象的相关信息，例如版本标签及其他可搜索属性
- 指定一个或多个分类主题到对象
- 设置对象的安全选项

以下几节将详细介绍您可设置的属性。

选择存储对象的位置

在“存储库：存储”对话框中，输入以下内容。

保存位置。 显示当前文件夹，即对象将保存到的位置。双击列表中的文件夹名称，将此文件夹设为当前文件夹。使用“上一层文件夹”按钮导航到父文件夹。使用“新建文件夹”按钮在当前层级中创建一个文件夹。

文件名。 存储对象时使用的名称。

存储。 在当前位置存储对象。

添加存储对象的相关信息

“存储库：存储”对话框的“信息”选项卡上的所有字段都是可选的。

作者。 在存储库中创建对象的用户名。缺省情况下，将显示用于存储库连接的用户名，但可以在此更改该名称。

版本标签。 从列表中选择标签以指示对象版本，或单击**添加**以创建新标签。避免在标签中使用 "[" 字符。如果不打算为此对象版本指定标签，请确保未选中任何复选框。有关更多信息，请参阅主题 [第 171 页的『查看和编辑对象属性』](#)。

描述。 对象的描述。用户可以按描述搜索对象（见注释）。

关键字。 与对象相关的一个或多个关键字，可用于搜索目的（见注释）。

到期日期。 对象在此日期之后不再对一般用户可见，但其所有者和存储库管理员仍然可以查看。要设置到期日期，选择**日期**选项并输入日期，或使用**日历**按钮选择日期。

存储。 在当前位置存储对象。

注：**描述**和**关键字**字段中的信息被当做和对象的“注解”选项卡上的 SPSS Modeler 中输入的任何信息不同。按描述或关键字的存储库搜索没有从“注解”选项卡返回信息。有关更多信息，请参阅主题 [第 169 页的『在存储库中搜索对象』](#)。

为存储对象指定主题

主题是面向存储库中所存储内容的层次分类系统。您可以在存储对象时从可用主题中进行选择，用户还可按主题搜索对象。可用主题列表由具有相应特权的存储库用户设置（有关更多信息，请参阅《*Deployment Manager* 用户指南》）。

要对该对象指定主题，请在“存储库：存储”对话框的“主题”选项卡上执行以下操作：

1. 单击**添加**按钮。
2. 从可用主题列表中单击主题名称。

3. 单击**确定**。

要删除主题指定：

4. 在已指定的主题列表中选择主题。

5. 单击**删除**。

设置存储对象的安全选项。

可以在“存储库：存储”对话框的“安全”选项卡上针对已存储的对象设置或更改众多安全选项。对于一个或多个**主体**（即用户或用户组），您可以：

- 分配对象的访问权限
- 修改对象的访问权限
- 删除对象的访问权限

Principal。 对此对象具有访问权限的用户或组的存储库用户名。

许可权。 该用户或组对对象的访问权限。

添加。 允许您添加一个或多个用户或组到对该对象具有访问权限的用户/组列表。有关更多信息，请参阅主题 [第 166 页的『添加用户到权限列表』](#)。

修改。 允许您修改选定用户或组对该对象的访问权限。缺省情况下授予读取访问权限。使用此选项可以授予其他访问权限，包括所有者、写入、删除和修改权限。

删除。 从该对象的权限列表中删除选定用户或组。

添加用户到权限列表

在“存储库：存储”对话框的“安全”选项卡上选择**添加**时，下列字段可用：

选择提供程序。 选择一个安全提供者用于认证。存储库可配置使用不同的安全提供者；如有必要，请与本地管理员联系以了解详细信息。

查找。 输入您要添加的用户或组的存储库用户名，并单击**搜索**以在用户列表中显示该名称。要一次添加多个用户名，请保留该字段为空，并单击**搜索**即可显示所有存储库用户名的列表。

用户列表。 从列表中选择一个或多个用户名，单击“**确定**”以将其添加到权限列表中。

修改对象的访问权限

在“存储库：存储”对话框的“安全”选项卡上选择**修改**时，下列字段可用：

所有者。 选择此选项为该用户或组授予对此对象的所有者访问权限。所有者对对象具有完全控制权限，包括“删除”和“修改”访问权限。

读取。 缺省情况下，非对象所有者的用户或用户组仅具有对象的读取访问权限。选中适当的复选框，以便为该用户或组添加“写入”、“删除”和“修改”权限等访问权限。

存储流

您可以在存储库中将流存储为 `.str` 文件，其他用户可以从这里访问它。

注：有关部署流以充分利用存储库附加功能的详细信息，请参阅 [第 172 页的『部署流』](#)。

要存储当前流：

1. 在主菜单中，单击：

文件 > 存储 > 存储为流...

2. 如果需要，指定存储库的连接设置。有关更多信息，请参阅主题 [第 164 页的『连接至存储库』](#)。有关特定端口、密码和其他连接的详细信息，请与本地系统管理员联系。

3. 在“存储库：存储”对话框中，选择要存储对象的文件夹，指定要记录的任何其他信息，然后单击**存储**按钮。有关更多信息，请参阅主题 [第 165 页的『设置对象属性』](#)。

存储工程

您可以在存储库中将整个 IBM SPSS Modeler 工程存储为 .cpj 文件，以便其他用户从这里访问它。

因为工程文件是其他 IBM SPSS Modeler 对象的容器，因此需要告诉 IBM SPSS Modeler 将工程的对象存储到存储库中。通过使用“工程属性”对话框中的设置可以执行此操作。有关更多信息，请参阅主题 [第 185 页的『设置工程属性』](#)。

一旦将工程配置为在存储库中存储对象，则不论何时在工程中添加新对象，IBM SPSS Modeler 都可自动提示存储该对象。

结束 IBM SPSS Modeler 会话后，必须存储工程文件的新版本，以便此工程文件可记录所添加的内容。工程文件可自动包含（及检索）其对象的最新版本。如果在 IBM SPSS Modeler 会话过程中没有向工程添加任何对象，则不必重新存储工程文件。但是，必须存储已更改的工程对象（流、输出等等）的新版本。

要存储工程

1. 在 IBM SPSS Modeler 的管理器窗格的 CRISP-DM 或“类别”选项卡上选择工程，然后在主菜单中单击：
文件 > 项目 > 存储项目...
2. 如果需要，指定存储库的连接设置。有关更多信息，请参阅主题 [第 164 页的『连接至存储库』](#)。有关特定端口、密码和其他连接的详细信息，请与本地系统管理员联系。
3. 在“存储库：存储”对话框中，选择要存储对象的文件夹，指定要记录的任何其他信息，然后单击**存储**按钮。有关更多信息，请参阅主题 [第 165 页的『设置对象属性』](#)。

存储节点

您可以在存储库中将当前流的单独节点定义存储为 .nod 文件，其他用户可以从这里访问它。

要存储节点：

1. 右键单击流工作区中的节点，然后单击**存储节点**。
2. 如果需要，指定存储库的连接设置。有关更多信息，请参阅主题 [第 164 页的『连接至存储库』](#)。有关特定端口、密码和其他连接的详细信息，请与本地系统管理员联系。
3. 在“存储库：存储”对话框中，选择要存储对象的文件夹，指定要记录的任何其他信息，然后单击**存储**按钮。有关更多信息，请参阅主题 [第 165 页的『设置对象属性』](#)。

存储输出对象

您可以在存储库中将当前流的输出对象存储为 .cou 文件，其他用户可以从这里访问它。

要存储输出对象：

1. 在 SPSS Modeler 的管理器窗格的“输出”选项卡上单击对象，然后在主菜单中单击：
文件 > 输出 > 存储输出...
2. 此外，可右键单击“输出”选项卡中的对象，然后单击**存储**。
3. 如果需要，指定存储库的连接设置。有关更多信息，请参阅主题 [第 164 页的『连接至存储库』](#)。有关特定端口、密码和其他连接的详细信息，请与本地系统管理员联系。
4. 在“存储库：存储”对话框中，选择要存储对象的文件夹，指定要记录的任何其他信息，然后单击**存储**按钮。有关更多信息，请参阅主题 [第 165 页的『设置对象属性』](#)。

存储模型和模型选项板

您可以在存储库中将单独模型存储为 .gm 文件，其他用户可以从这里访问它。您还可在存储库中将“模型”选项板的全部内容存储为 .gen 文件。

存储模型：

1. 在 SPSS Modeler 的“模型”选用板上单击对象，然后在主菜单中单击：
文件 > 模型 > 存储模型...

2. 此外，可右键单击“模型”选用板中的对象，然后单击**存储模型**。
3. 从“完成存储过程”继续。

存储模型选用板：

1. 右键单击模型选用板的背景。
2. 在弹出菜单中，单击**存储选用板**。
3. 从“完成存储过程”继续。

完成存储过程：

1. 如果需要，指定存储库的连接设置。有关更多信息，请参阅主题 [第 164 页的『连接至存储库』](#)。有关特定端口、密码和其他连接的详细信息，请与本地系统管理员联系。
2. 在“存储库：存储”对话框中，选择要存储对象的文件夹，指定要记录的任何其他信息，然后单击**存储按钮**。有关更多信息，请参阅主题 [第 165 页的『设置对象属性』](#)。

从存储库中检索对象

可以检索已存储在存储库中的流、模型、模型选项板、节点、工程和输出对象。

注：除使用此处介绍的菜单选项外，还可以在 SPSS Modeler 窗口右上角的管理器窗格的适当选项卡上单击右键以检索流、输出对象、模型和模型选用板。

1. 要检索流，在 IBM SPSS Modeler 主菜单中，单击：

文件 > 检索流...

2. 要检索模型、模型选用板、工程或输出对象，在 IBM SPSS Modeler 主菜单中单击：

文件 > 模型 > 检索模型...

或者

文件 > 模型 > 检索模型选用板...

或者

文件 > 项目 > 检索项目...

或者

文件 > 输出 > 检索输出...

3. 此外，可右键单击管理器或工程窗格，并在弹出菜单中单击**检索**。
4. 要检索节点，在 IBM SPSS Modeler 主菜单中，单击：

插入 > 来自存储库的节点（或超节点）...

- a. 如果需要，指定存储库的连接设置。有关更多信息，请参阅主题 [第 164 页的『连接至存储库』](#)。有关特定端口、密码和其他连接的详细信息，请与本地系统管理员联系。

5. 在“存储库：检索”对话框中，浏览到对象，将其选中并单击**检索按钮**。请参阅主题，以获取更多信息。

选择要检索的对象

在“存储库：检索/搜索”对话框中，存在下列字段：

查找范围。 显示当前文件夹的文件夹层次结构。要导航到其他文件夹，请从列表选中并直接导航到相应位置，或使用该字段下面的对象列表进行导航。

“上层文件夹”按钮。 导航到层次机构中当前文件夹的上一级。

“新建文件夹”按钮。 在层次结构中的当前层级上创建新文件夹。

文件名。 选定对象的存储库文件名。要检索该对象，单击**检索**。

文件类型。 您选择要检索的对象类型。在对象列表中只显示该类型的对象以及文件夹。要显示不同类型的对象以进行检索，请从列表中选择对应的对象类型。

以锁定状态打开。缺省情况下，当检索对象时，它将在存储库中被锁定，以便不会被其他用户更新。如果不希望在检索对象时将其锁定，请取消选中此复选框。

描述、关键字。如果在存储对象时定义了对对象的其他详细信息，都会显示在这里。有关更多信息，请参阅主题第 165 页的『添加存储对象的相关信息』。

版本。要检索最新版本以外的对象版本，请单击此按钮。此时显示出所有版本的详细信息，因此有助于选择所需的版本。

选择对象版本

要选择存储库对象的特定版本，请在“存储库：选择版本”对话框中执行以下操作：

1. (可选) 按版本、标签、大小、创建日期或创建用户排序列表，请单击相应列的标题。
2. 选择您要使用的对象版本。
3. 单击“继续”。

在存储库中搜索对象

可以按名称、文件夹、类型、标签、日期或其他标准搜索对象。

按名称搜索对象

1. 在 IBM SPSS Modeler 主菜单中，单击：

工具 > 存储库 > 探索...

- a. 如果需要，指定存储库的连接设置。有关更多信息，请参阅主题第 164 页的『连接至存储库』。有关特定端口、密码和其他连接的详细信息，请与本地系统管理员联系。
2. 单击 **搜索** 选项卡。
 3. 在 **搜索命名的对象** 字段中，指定要查找的对象名称。

当按名称搜索对象时，可以用星号(*)作为通配符匹配任何字符串，用问号(?)匹配任何单个字符。例如，*cluster* 将匹配在名称中任意位置包含字符串 cluster 的所有对象。搜索字符串 m0?_* 匹配 M01_cluster.str 和 M02_cluster.str 但不匹配 M01a_cluster.str。搜索不区分大小写 (cluster 与 Cluster 和 CLUSTER 都匹配)。

注: 如果对象数量巨大，那么搜索可能需要较长时间。

按其他条件搜索

可以根据标题、标签、日期、作者、关键字、索引内容或描述执行搜索。只能找到与所有指定的搜索标准匹配的对象。例如，可以找到所有包含一个或多个聚类模型的流，这些流还应用了特定标签，并且修改于特定日期后。

对象类型。您可以将搜索限于模型、流、输出、超节点、项目、模型选用板或其他类型的对象。

- **模型。**您可以按类别 (分类、近似值、聚类等) 或者通过特定的建模算法 (如 Kohonen) 来搜索模型。
还可以按所使用的字段进行搜索例如，搜索将名称为 *income* 的字段用作输入或输出 (目标) 的所有模型。
- **流。**对于流，可以按流中使用的字段或包含的模型类型 (类别或算法) 限定搜索。

主题。您可以从具有相应特权的存储库用户设置的列表中搜索与特定主题相关联的模型 (有关更多信息，请参阅《Deployment Manager 用户指南》)。要获得此列表，请选中此复选框，然后单击出现的“添加主题”按钮，从列表选择一个或多个主题，并单击“确定”。

标签。用特定对象版本标签限定搜索。

日期。可以指定创建或修改日期，然后搜索指定日期范围之前、之后或者之间的对象。

作者。限定搜索特定用户创建的对象。

关键字。 搜索特定关键字。在 IBM SPSS Modeler 中，可在“注解”选项卡中为流、模型或输出对象指定关键字。

描述。 在说明字段中搜索特定项。在 IBM SPSS Modeler 中，可在“注解”选项卡中为流、模型或输出对象指定说明。多个搜索词组可用分号隔开，例如，income; crop type; claim value。（注意在搜索词组中，空格很重要。例如，包含一个空格的 crop type 与包含两个空格的 crop type 并不相同。）

修改存储库对象

您可以直接从 SPSS Modeler 修改存储库中的现有对象。您可以：

- 创建、重命名或删除文件夹
- 锁定或解锁对象
- 删除对象

编辑、重命名和删除文件夹

1. 要对存储库中的文件夹执行操作，在 SPSS Modeler 主菜单中单击：

工具 > 存储库 > 浏览...

- a. 如果需要，指定存储库的连接设置。有关更多信息，请参阅主题第 164 页的『[连接至存储库](#)』。有关特定端口、密码和其他连接的详细信息，请与本地系统管理员联系。
2. 确保文件夹选项卡处于活动状态。
 3. 要创建新文件夹，可右键单击父文件夹并单击**新建文件夹**。
 4. 要重命名文件夹，可右键单击该文件夹并单击**重命名文件夹**。
 5. 要删除文件夹，可右键单击该文件夹并单击**删除文件夹**。

锁定和解锁存储库对象

您可以锁定一个对象，以防止其他用户更新任一现有版本或新建版本。对象图标上的挂锁符号指示对象被锁定。



图 17: 锁定的对象

要锁定对象

1. 在存储库资源管理器窗口中，右键单击所需的对象。
2. 单击**锁定**。

要解锁对象

1. 在存储库资源管理器窗口中，右键单击所需的对象。
2. 单击**解锁**。

删除存储库对象

在删除存储库中的对象之前，必须确定是要删除对象的所有版本还是仅删除其特定版本。

删除对象的所有版本

1. 在存储库资源管理器窗口中，右键单击所需的对象。
2. 单击**删除对象**。

删除对象的最新版本

1. 在存储库资源管理器窗口中，右键单击所需的对象。
2. 单击**删除**。

删除对象的以前版本

1. 在存储库资源管理器窗口中，右键单击所需的对象。
2. 单击**删除版本**。
3. 选择要删除的版本并单击 **确定**。

管理存储库对象的属性

可以从 SPSS Modeler 中控制各类对象属性。您可以：

- 查看文件夹的属性
- 查看和编辑对象的属性
- 创建、应用和删除对象的版本标签

查看文件夹属性

要在存储库窗口中查看任何文件夹的属性，可右键单击要查看的文件夹。单击**文件夹属性**。

“常规”选项卡

此选项卡将显示文件夹的名称、创建和修改日期。

“权限”选项卡

在此选项卡中，您可以指定文件夹的读许可权和写许可权。所有具有父文件夹访问权的用户和组的将列出。许可权具有层次结构。例如，如果没有读权限，就不能有写权限。如果没有写权限，就不能有删除权限。

用户和组。 列出对此文件夹至少具有“读取”访问权限的存储库用户和组。选择“写入”和“删除”复选框，将此文件夹的这些访问权限添加到特定用户或组。单击“权限”选项卡右侧的**添加用户/组**图标可将访问权限分配给其他用户和组。可用用户和组的列表由管理员控制。

建立权限级联。 选择一个选项以控制当前文件夹的更改如何应用到其子文件夹（如果有）。

- **为所有权限建立级联。** 将当前文件夹的权限设置级联到所有下级文件夹。这是一次为多个文件夹快速设置权限的方法。根据需要为父文件夹设置权限，然后根据需要进行级联。
- **仅为更改建立级联。** 仅级联上次应用更改之后所做的更改。例如，添加了一个新组，并希望授予该组对销售分支下所有文件夹的访问权限，则可授予该组对根销售文件夹的访问权限并将更改级联到所有子文件夹。现有子文件夹的所有其他权限保存不变。
- **不要建立级联。** 所有更改仅应用于当前文件夹，而不级联到子文件夹。

查看和编辑对象属性

在“对象属性”对话框中，您可以查看和编辑属性。虽然某些属性无法更改，但通常可以通过添加新版本来更新某个对象。

1. 在存储库窗口中，右键单击所需的对象。
2. 单击**对象属性**。

一般选项卡

名称。 存储库中显示的对象名称。

创建日期。 对象（而非版本）的创建日期。

上次修改时间。 最新版本的修改日期。

作者。 用户的登录名。

描述。 缺省情况下，此处包含在 SPSS Modeler 的此对象的“注解”选项卡上指定的说明。

链接的主题。 通过存储库可在需要时根据主题组织模型和相关对象。可用主题的列表由具有相应特权的存储库用户设置（有关更多信息，请参阅《*Deployment Manager* 用户指南》）。

关键字。可在“注解”选项卡上为流、模型或输出对象指定关键字。多个关键字之间应由空格进行分隔，且最多可包含 255 个字符。（如果关键字中包含空格，则使用引号对其进行分隔。）

版本选项卡

存储在存储库中的对象可能有多个版本。“版本”选项卡中可显示有关每个版本的信息。

可为已存储对象的特定版本指定或修改下列属性：

版本。根据存储版本的时间生成该版本的唯一标识。

标签。版本的当前标签（如果有的话）。与版本标识不同，标签可从对象的一个版本移动到另一个版本。

还为每个版本显示了文件大小、创建日期和创建者。

编辑标签。单击“版本”选项卡右上角的**编辑标签**图标以定义、应用或删除存储对象的标签。有关更多信息，请参阅主题 [第 172 页的『管理对象版本标签』](#)。

权限选项卡

在“权限”选项卡上，可以为对象设置读写权限。所有具有当前对象访问权的用户和组的将列出。许可权具有层次结构。例如，如果没有读权限，就不能有写权限。如果没有写权限，就不能有删除权限。

用户和组。列出对此对象至少具有“读取”访问权限的存储库用户和组。选择“写入”和“删除”复选框，将此对象的这些访问权限添加到特定用户或组。单击“权限”选项卡右侧的**添加用户/组**图标可将访问权限分配给其他用户和组。可用用户和组的列表由管理员控制。

管理对象版本标签

使用“编辑版本标签”对话框可以：

- 应用标签到所选对象
- 从所选对象删除标签
- 定义新标签并将其应用到对象

应用标签到对象

1. 在 **可用标签** 列表中选择一个或多个标签。
2. 单击向右方向按钮移动所选标签到 **已应用标签** 列表。
3. 单击**确定**。

从对象删除标签

1. 在 **已应用用标签** 列表中选择一个或多个标签。
2. 单击向左方向按钮移动所选标签到 **可用标签** 列表。
3. 单击**确定**。

定义新标签并将其应用到对象

1. 在 **新建标签** 字段中键入标签名称。
2. 单击向右方向按钮移动新标签到 **已应用标签** 列表。
3. 单击**确定**。

部署流

要使流能够与瘦客户机应用程序 IBM SPSS Modeler Advantage 一起使用，必须将其部署为存储库中的流（.str 文件）。

注：无法部署评分分支中存在多个源节点的流。

流可以充分利用 IBM SPSS Collaboration and Deployment Services 的企业级功能。有关更多信息，请参阅主题 [第 163 页的『存储和部署存储库对象』](#)。

要部署当前流（文件菜单方法）

1. 在主菜单中，单击：
文件 > 存储 > 部署
2. 选择部署类型并根据需要填写剩余对话框内容。
3. 单击**部署为流**以部署要与 IBM SPSS Modeler Advantage 或 IBM SPSS Collaboration and Deployment Services 配合使用的流。
4. 单击**存储**。有关更多信息，请单击**帮助**。
5. 从“完成部署过程”继续。

要部署当前流（工具菜单方法）

1. 在主菜单中，单击：
工具 > 流属性 > 部署
2. 选择部署类型，根据需要完成“部署”选项卡的其余部分，然后单击**存储**。有关更多信息，请参阅主题 [第 173 页的『流部署选项』](#)。

完成部署过程

1. 如果需要，指定存储库的连接设置。有关更多信息，请参阅主题 [第 164 页的『连接至存储库』](#)。有关特定端口、密码和其他连接的详细信息，请与本地系统管理员联系。
2. 在“存储库：存储”对话框中，选择要存储对象的文件夹，指定要记录的任何其他信息，然后单击**存储**。有关更多信息，请参阅主题 [第 165 页的『设置对象属性』](#)。

流部署选项

“流选项”对话框中的“部署”选项卡可指定用于部署流的选项。

部署流时，将其作为扩展名为 `.str` 的文件存储在存储库中。

通过部署流，可以充分利用 IBM SPSS Collaboration and Deployment Services 附带的其他功能，如多用户访问、自动评分、模型刷新和 Champion Challenger 分析。

注：关联规则、STP 和 TCM 建模节点不支持 IBM SPSS Collaboration and Deployment Services 中的 Model Evaluation 或 Champion Challenger 步骤。

从“部署”选项卡中，还可以预览 IBM SPSS Modeler 为流创建的流描述。有关更多信息，请参阅主题 [第 40 页的『流描述』](#)。

部署类型。 选择流的部署方式。所有流都需要指定评分分支才能进行部署；其他要求和选项取决于部署类型。

注：关联规则、STP 和 TCM 建模节点不支持 IBM SPSS Collaboration and Deployment Services 中的 Model Evaluation 或 Champion Challenger 步骤。

- **<none>**。该流将不会部署到存储库。除流描述预览外的所有选项均被禁用。
- **仅限于评分**。当您单击**存储**按钮时，流被部署到存储库。可使用在**评分节点**字段中指定的节点对数据进行评分。
- **模型刷新**。同“仅限于评分”，除此之外，还可使用在**建模节点**和**模型块**字段中指定的对象于存储库中更新模型。请注意，缺省情况下在 IBM SPSS Collaboration and Deployment Services 中不支持自动模型刷新，因此如果要在从存储库运行流时使用此功能，那么必须选择此部署类型。有关更多信息，请参阅[第 175 页的『模型刷新』](#)。

评分节点。 选择图形、输出或导出节点，以识别用于对数据评分的流分支。虽然流实际上可以包含任何数目的有效分支、模型和终端节点，但必须指定一个且唯一一个评分分支以进行部署。这是部署任何流时最基本的要求。

评分参数。 通过此选项可指定参数，在运行评分分支时可对这些参数进行修改。有关更多信息，请参阅[第 174 页的『评分和建模参数』](#)。

建模节点。 对于模型刷新，指定用于在存储库中重新生成或更新模型的建模节点。此建模节点的类型必须与为**模型块**指定的类型相同。

模型构建参数。 通过此选项可指定参数，在运行建模节点时可对这些参数进行修改。有关更多信息，请参阅第 174 页的『[评分和建模参数](#)』。

模型块。 对于模型刷新，指定将在存储库中每次流更新（通常作为计划作业的一部分）时更新或重新生成的模型块。模型必须位于评分分支上。尽管评分分支可能存在多个模型，但只能指定一个模型。请注意，最初创建流时，这可能实际上是一个占位模型，此模型将在新数据可用时进行更新或重新生成。

检查。 单击此按钮以查看此流是否为可部署的有效流。所有流都需要指定评分分支才能进行部署。如果未满足这些条件，那么将显示错误消息。

存储。 如果流有效，则将其部署。否则将显示错误消息。单击**修订**按钮，更正错误并重试。

预览流描述。 允许您查看 IBM SPSS Modeler 为流创建的流描述内容。有关更多信息，请参阅第 40 页的『[流描述](#)』。

注：关联规则、STP 和 TCM 建模节点不支持 IBM SPSS Collaboration and Deployment Services 中的 Model Evaluation 或 Champion Challenger 步骤。

评分和建模参数

将流部署到 IBM SPSS Collaboration and Deployment Services 时，可以选择每次对模型进行更新或评分时，可查看或编辑哪些参数。例如，您可以指定每次运行作业时的最大和最小值，或可更改的其他值。

1. 要使某个参数可见以便在部署流后对其进行查看或编辑，请在“评分参数”对话框的列表中将其选中。

可用参数的列表在流属性对话框的“参数”选项卡中进行定义。有关更多信息，请参阅主题第 37 页的『[设置流和会话参数](#)』。

评分分支

如果您正部署流，必须将流的一个分支指定为**评分分支**（即包含评分节点的分支）。在您将某个分支指定为评分分支后，该分支将在流工作区上突出显示，就像评分分支上指向模型块的模型链接那样。这一直观表示在存在多个分支的复杂流情况下特别有用，此时评分分支可能并不明显可见。

注：只能指定单个流分支作为评分分支。

如果流已定义有评分分支，则新指定的分支将替换它成为评分分支。通过“自定义颜色”选项，可以设置评分分支的指示颜色。有关更多信息，请参阅主题第 191 页的『[设置显示选项](#)』。

通过“显示/隐藏流标记”工具栏按钮，可以显示或隐藏评分分支指示。



图 18: “显示/隐藏流标记”工具栏按钮

标识评分分支以供部署

您可以从终端节点的弹出菜单或“工具”菜单中指定评分分支。如果使用弹出菜单，则会在流属性的“部署”选项卡中自动设置评分节点。

要将分支指定为评分分支（弹出菜单）

1. 将模型块连接到终端节点（位于模型块下游的处理或输出节点）。
2. 右键单击终端节点。
3. 在菜单中，单击**用作评分分支**。

要将分支指定为评分分支（工具菜单）

1. 将模型块连接到终端节点（位于模型块下游的处理或输出节点）。
2. 在主菜单中，单击：

工具 > 流属性 > 部署

3. 在**部署类型**列表中，根据需要单击**仅评分**或**模型刷新**。有关更多信息，请参阅主题 [第 173 页的『流部署选项』](#)。
4. 单击**评分节点**字段，并从列表中选择终端节点。
5. 单击**确定**。

模型刷新

模型刷新是使用较新数据重新构建流中现有模型的过程。在存储库中的流本身不会变化。例如，算法类型和流相关设置保持不变，但模型会采用新数据；如果新版本的模型工作性能优于旧的模型，则会进行更新。

在一个流中只能将单个模型块设置为刷新，这称为**刷新模型**。如果在流属性的“部署”选项卡上单击**模型刷新**选项（参阅第 173 页的『流部署选项』），则您所指定的模型块将成为刷新模型。您还可以从模型块的弹出菜单中指定模型为刷新模型。要完成此操作，模型块必须已位于评分分支上。

如果关闭模型块的“刷新模型”状态，这等效于将流的部署类型设置为“仅评分”，同时“流属性”对话框的“部署”选项卡将相应更新。通过当前评分分支上的模型块的弹出菜单中的**用作刷新模型**选项，可以打开和关闭此状态。

从评分分支中删除模型块的模型链接时，还将删除该模型块的“刷新模型”状态。您可通过“编辑”菜单或工具栏取消删除模型链接；这样还会恢复模型块的“刷新模型”状态。

如何选择刷新模型

与评分分支相同，在流中也将突出显示到刷新模型的链接。选为刷新模型的模型块，以及由此突出显示的链接，取决于流中存在多少个模型块。

流中仅单个模型

如果在标识评分分支时其上面只有单个链接模型块，则该模型块将成为流的刷新模型。

流中有多个模型

如果流中存在多个链接模型块，则按以下方式选择刷新模型。

如果在“流属性”对话框的“部署”选项卡中定义了模型块，且该模型块也存在于流中，则该模型块成为刷新模型。

如果未在“部署”选项卡上定义模型块，或定义的模型块不在评分分支上，则最接近终端节点的模型块成为刷新模型。

如果您后面取消选择这两个模型链接作为刷新链接，则只会突出显示评分分支，而不包括链接。部署类型被设置为“仅评分”。

注：您可以选择设置一个链接为“替换”状态，而不是其他链接。在这种情况下，选为刷新模型的模型块具有刷新链接，并且在指定评分分支时最接近终端节点。

流中无模型

如果流中不存在模型，或所有模型均未带有模型链接，部署类型将被设置为“仅评分”。

检查评分分支是否存在错误

在指定评分分支时，将检查其是否存在错误。

如果发现了错误，评分分支将以评分分支错误颜色突出显示，同时显示错误消息。通过“自定义颜色”选项，可以设置评分分支的错误颜色。有关更多信息，请参阅主题 [第 191 页的『设置显示选项』](#)。

如果发现了错误，请执行以下操作：

1. 根据错误消息的内容纠正错误。
2. 在主菜单中，单击：

工具 > 流属性 > 部署

并单击**检查**。

3. 如有必要，重复此过程直到不再发现错误。

第 12 章 将流保存到 IBM Cloud Pak for Data

您可以将流保存到 IBM Cloud Pak for Data 服务器，在其中您可以将这些流作为流程打开，并利用 Cloud Pak for Data 中提供的许多功能。

Cloud Pak for Data 简化了您在整个企业中收集、组织和分析数据以及注入 AI 的方式。无缝集成广泛的数据和 AI 软件、领先的治理和安全性，以及统一的协作体验。有关更多信息，请参阅 <https://www.ibm.com/products/cloud-pak-for-data>。如果您对 Cloud Pak for Data 感兴趣，请联系 IBM 代表。

配置服务器连接

这些指示信息基于 Firefox。步骤可能会因您使用的 Web 浏览器而异。

1. 打开 Firefox，在地址栏中粘贴 Cloud Pak for Data URL，然后按 Enter 键。

您可以通过打开 Cloud Pak for Data Web 客户端并查看浏览器地址栏中的 URL 来获取 URL（即截至但不包括 /zen 的所有内容）。缺省情况下，URL 采用格式 `https://<namespace>-cpd-<namespace>.apps.<cluster-subdomain>`

其中：

- <namespace> 是您的名称空间（例如，ab123）。
- <cluster-subdomain> 是集群的子域名（例如，ocp454x86-aqt-mycompany.com）

在此示例中，URL 为 `https://ab123-cpd-ab123.apps.ocp454x86-aqt-mycompany.com`

有关 Cloud Pak for Data URL 的更多信息，请参阅 [Cloud Pak for Data 文档](#)。

2. 单击位于地址栏左侧的黄色挂锁图标，然后单击**连接不安全**旁边的箭头。
3. 单击**更多信息**。此时将打开“页面信息”对话框。
4. 在“页面信息”对话框中，转至**安全性**，然后单击**查看证书**。
5. 向下滚动到**其他**并下载 PEM 证书文件。
6. 从命令行，将认证文件导入到 IBM SPSS Modeler 客户机目录，如下所示：

在 Windows 上：

```
"%JAVA_HOME%\bin\keytool.exe -importcert -trustcacerts -file <Certification file absolute path> -alias <alias> -keystore "%Modeler client installation path%\jre\lib\security\cacerts
```

在 MacOS 上：

```
"%JAVA_HOME%/bin/keytool -importcert -trustcacerts -file <Certification file absolute path> -alias <alias> -keystore "<Modeler client installation path>/jre/lib/security/cacerts
```

提示时，输入 `changeit` 以获取密钥库密码，并输入 `yes` 以信任证书。

7. 在 IBM SPSS Modeler 中的主菜单上，单击：

工具 > IBM Cloud Pak for Data 服务器 > IBM Cloud Pak for Data 服务器连接...

8. 指定服务器的连接设置。使用步骤 1 中使用的相同 URL。如果您不确定要使用哪个用户名或密码，请与本地系统管理员联系。如果选择 API 密钥登录方法，请使用以下方法来获取 URL：

- 登录到 Cloud Pak for Data 集群，并运行命令 `oc get routes -n ibm-common-services`，然后复制 `cp-console` 的 `HOST/PORT` 值。在此示例中，URL 为 `https://cp-console.apps.wml46x04.cp.myserver.com`：

```
[root@wml46x04-inf ocp-scripts]# oc get routes -n ibm-common-services
NAME          HOST/PORT          PATH          SERVICES
PORT         TERMINATION        WILDCARD
cp-console    cp-console.apps.wml46x04.cp.myserver.com    icp-management-ingress
https        reencrypt/Redirect None
cp-proxy      cp-proxy.apps.wml46x04.cp.myserver.com      nginx-ingress-controller
https        passthrough/Redirect None
```

9. 单击**连接**以连接到服务器。

将流保存到 Cloud Pak for Data

1. 在主菜单中，单击：

文件 > 存储为流...。流在迁移到 Cloud Pak for Data 之后称为流程。

或者，也可以在流管理器窗格中右键单击并选择**存储为流程...**

2. 选择先前配置的 Cloud Pak for Data 服务器。然后，选择要将流程保存到的项目。

提示: 您还可以选择要同时保存的多个流。在主菜单上，单击**工具 > IBM Cloud Pak for Data 服务器... > 将流批量上传到 IBM Cloud Pak for Data 服务器**。系统将提示您从本地文件系统中选择流，选择服务器，然后选择项目。

第 13 章 导出到外部应用程序

关于导出到外部应用程序

IBM SPSS Modeler 提供了若干机制，用以将整个数据挖掘过程导出到外部应用程序，以便您所作的准备数据和构建模型工作还可在 IBM SPSS Modeler 之外得以利用。

在上一部分中介绍了如何将流部署到 IBM SPSS Collaboration and Deployment Services 存储库以充分利用其多用户访问、作业计划和其他功能。类似地，IBM SPSS Modeler 流也可以与以下应用程序结合使用：

- IBM SPSS Modeler Advantage
- 能够以 PMML 格式导入和导出文件的应用程序

有关在 IBM SPSS Modeler Advantage 中使用流的更多信息，请参见第 179 页的『[打开 IBM SPSS Modeler Advantage 中的流](#)』。

有关将模型以 PMML 文件形式导出和导入，以便与其他支持此格式的应用程序共享模型的信息，请参阅第 179 页的『[导入和导出 PMML 模型](#)』。

打开 IBM SPSS Modeler Advantage 中的流

IBM SPSS Modeler 流可与精简客户端应用程序 IBM SPSS Modeler Advantage 一起使用。您可以完全在 IBM SPSS Modeler Advantage 中创建自定义应用程序，还可以使用在 IBM SPSS Modeler 中创建的流作为应用工作流的基础。

要打开 IBM SPSS Modeler Advantage 中的流：

1. 在 IBM SPSS Collaboration and Deployment Services 存储库中部署流，确保单击**部署为流**选项。有关更多信息，请参阅主题第 172 页的『[部署流](#)』。
2. 单击“在 IBM SPSS Modeler Advantage 中打开”工具栏按钮，或者在主菜单中单击：

文件 > 在 IBM SPSS Modeler Advantage 中打开

1. 如果需要，指定存储库的连接设置。有关更多信息，请参阅主题第 164 页的『[连接至存储库](#)』。有关特定端口、密码和其他连接的详细信息，请与本地系统管理员联系。

注：存储库服务器还必须安装有 IBM SPSS Modeler Advantage 软件。

1. 在“存储库：存储”对话框中，选择要存储对象的文件夹，指定要记录的任何其他信息，然后单击**存储按钮**。有关更多信息，请参阅主题第 165 页的『[设置对象属性](#)』。

进行此操作可启动 IBM SPSS Modeler Advantage，此时流已经打开。流在 IBM SPSS Modeler 中关闭。

导入和导出 PMML 模型

PMML（也称为预测模型标记语言）是一种 XML 格式，用于描述数据挖掘和统计模型，包括模型的输入、用于为数据挖掘准备数据的变换，以及定义模型自身的参数。IBM SPSS Modeler 可导入和导出 PMML，这使得其能够与其他支持此格式的应用程序（如 IBM SPSS Statistics）共享模型。

有关 PMML 的详细信息，请参阅数据挖掘组网站 (<http://www.dmg.org>)。

导出模型

PMML 导出支持大多数模型类型，这些模型类型生成在 IBM SPSS Modeler 中。有关更多信息，请参阅主题第 180 页的『[支持 PMML 的模型类型](#)』。

1. 右键单击模型选项板上的模型块。（或者，双击工作区上的模型块并选择“文件”菜单。）
2. 在菜单上，单击**导出 PMML**。
3. 在“导出”（或“保存”）对话框中，指定此模型的目标目录及唯一名称。

注:

您可在“用户选项”对话框中更改 PMML 导出选项。在主菜单中, 单击:

工具 > 选项 > 用户选项

然后单击 PMML 选项卡。

有关更多信息, 请参阅主题 [第 192 页的『设置 PMML 导出选项』](#)。

导入以 PMML 格式保存的模型

以 PMML 格式从 IBM SPSS Modeler 或其他应用程序中导出的模型可以导入到模型选用板中。有关更多信息, 请参阅主题 [第 180 页的『支持 PMML 的模型类型』](#)。

1. 在模型选用板上, 右键单击选用板并从菜单中选择**导入 PMML**。
2. 选择要导入的文件并根据需要为变量标签指定选项。
3. 单击**打开**。

如果模型中存在变量标签, 请使用变量标签。 PMML 可为数据字典中的变量同时指定变量名和变量标签 (例如, Referrer 标识, 简称 *RefID*)。如果在最初导出的 PMML 中存在变量标签, 则选中此选项可以使用这些变量标签。

如果已选中变量标签选项但在 PMML 中没有变量标签, 则按常规使用变量名。

支持 PMML 的模型类型

PMML 导出

IBM SPSS Modeler 模型。 在 IBM SPSS Modeler 中创建的下列模型都可导出为 PMML 4.0 格式:

- C&R 树
- QUEST
- CHAID
- 神经网络
- C5.0
- Logistic 回归
- Genlin
- SVM
- Apriori
- Carma
- K-Means
- Kohonen
- 二阶
- 二阶 AS
- GLMM (对于所有 GLMM 模型都会导出 PMML, 但是 PMML 只有固定效应)
- 决策列表
- Cox
- 序列 (不支持序列 PMML 模型评分)
- 随机树
- 树 AS
- 线性
- 线性 AS

- 回归
- Logistic
- GLE
- LSVM
- 异常检测
- KNN
- 关联规则

数据库本机模型。对于使用数据库本机算法生成的模型，PMML 导出不可用。无法导出使用 Microsoft 的 Analysis Services 或 Oracle Data Miner 创建的模型。

PMML 导入

IBM SPSS Modeler 可以导入并评分由所有 IBM SPSS Statistics 产品的当前版本生成的 PMML 模型，包括从 IBM SPSS Modeler 导出的模型和由 IBM SPSS Statistics 17.0 或以后版本生成的模型或转换 PMML。这实质上意味着评分引擎可评分的任何 PMML，以下除外：

- 无法导入 Apriori、CARMA、异常检测、序列和关联规则模型。
- 将 PMML 模型导入到 IBM SPSS Modeler 中后，虽然可以对其进行评分，但不能进行浏览。（注意，其中包括最初从 IBM SPSS Modeler 中导出的模型。为避免此限制，可将模型按生成的模型文件 (*.gm) 导出而不是按 PMML 导出。）
- 在导入时会执行有限的验证，但在试图对模型评分时会执行全面验证。因此有可能导入成功，但评分却失败或产生不正确的结果。

注：对于导入到 IBM SPSS Modeler 中的第三方 PMML，IBM SPSS Modeler 将尝试对可以识别并进行评分的有效 PMML 进行评分。但是，无法保证将对所有 PMML 进行评分，也无法保证以应用程序生成 PMML 的方式对 PMML 进行评分。

第 14 章 工程和报告

工程简介

工程 是一组与数据挖掘任务相关的文件。工程包括数据流、图形、生成的模型、报告以及在 IBM SPSS Modeler 中创建的任何其他内容。乍一看，IBM SPSS Modeler 工程似乎只是一种组织输出的方式，但实际上其作用远不止此。利用工程，可以：

- 注解工程文件中的每个对象。
- 使用 CRISP-DM 方法指导数据挖掘工作。工程还包含 CRISP-DM 帮助系统，该系统针对使用 CRISP-DM 进行的数据挖掘提供详细信息和现实示例。
- 将非 IBM SPSS Modeler 对象添加到工程中，如用于表示数据挖掘目标的 PowerPoint 幻灯片放映或与计划使用的算法的白皮书。
- 根据注解生成全面更新报告和简单更新报告。为了便于在组织的内联网上发布，可以用 HTML 的形式生成这些报告。

注：如果 IBM SPSS Modeler 窗口中未显示工程窗格，请在“视图”菜单中单击**工程**。

可通过两种方式查看添加到工程中的对象：**类视图**和 **CRISP-DM 视图**。添加到工程的任何内容都会添加到这两个视图中，并且您可以在这两个视图之间切换以创建效果最佳的组织。

CRISP-DM 视图

通过支持跨行业数据挖掘标准过程 (CRISP-DM)，IBM SPSS Modeler 工程提供业已证明的非专有方式来组织数据挖掘工作的各个方面。CRISP-DM 使用六个阶段来描述从开始（收集业务需求）到结束（部署结果）的整个过程。虽然某些阶段通常并未涉及 IBM SPSS Modeler 中的工作，但工程窗格仍旧包括全部六个阶段，以便您有一个中心位置来存储和跟踪与工程关联的所有材料。例如，业务了解阶段通常包括收集需求并与同事会谈以确定目标，而不是在 IBM SPSS Modeler 中处理数据。使用工程窗格，可以将会谈记录存储在 *Business Understanding* 文件夹中，以便将来参考及纳入到报告中。

工程窗格中的 CRISP-DM 视图还配备了自己的帮助系统，该系统可全程指导您完成整个数据挖掘过程。在 IBM SPSS Modeler 中，可通过在“帮助”菜单中单击 **CRISP-DM 帮助** 来访问此帮助。

注：如果窗口中未显示工程窗格，请在“视图”菜单中单击**工程**。

设置缺省工程阶段

添加到工程的对象会被添加到 CRISP-DM 的缺省阶段。这表示，您需要根据使用对象的数据挖掘阶段来手动组织对象。最好将缺省文件夹设置为您当前正在使用的阶段。

选择要用作缺省值的阶段：

1. 在 CRISP-DM 视图中，右键单击要设置为缺省文件夹的阶段的文件夹。
2. 在菜单中，单击**设为缺省值**。

缺省文件夹将以粗体显示。

“类”视图

工程窗格中的类视图根据所创建对象的类型分门别类地组织您在 IBM SPSS Modeler 中所做的工作。可将保存后的对象添加到以下任何类别中：

- 流
- 节点数
- 模型
- 表、图形、报告

- 其他（非 IBM SPSS Modeler 文件，如与数据挖掘工作相关的幻灯片放映或白皮书）

将对象添加到类视图也会将它们添加到 CRISP-DM 视图的缺省阶段文件夹中。

注：如果窗口中未显示工程窗格，请在“视图”菜单中单击**工程**。

构建工程

工程本质上是一个文件，包含对与该工程关联的所有文件的引用。这表示工程项既单独保存又作为工程文件（.cpj）中的引用保存。因为这种引用结构，请注意以下几点：

- 必须先单独保存工程项，然后才能将其添加到工程中。如果未保存某项，系统将提示您先保存它，然后再将它添加到当前工程中。
- 单独更新的对象（如流）也会在工程文件中更新。
- 手动从文件系统中移动或删除对象（如流、节点和输出对象）会导致工程文件中的链接无效。

创建新工程

在 IBM SPSS Modeler 窗口中新建工程非常简单。您可以从构建新工程开始（如果没有工程打开），也可以关闭现有工程并从头开始。

在主菜单中，单击：

文件 > 项目 > 新建项目...

添加到工程

创建或打开工程后，便可以使用多种方法来添加对象，如数据流、节点和报告。

从管理器添加对象

使用 IBM SPSS Modeler 窗口右上角的管理器，可以添加流或输出。

1. 从其中一个管理器选项卡中选择一个对象（如表或流）。
2. 右键单击，然后单击**添加到工程**。

如果以前保存过该对象，则会自动将它添加到适当的对象文件夹（在类视图中）或缺省阶段文件夹（在 CRISP-DM 视图中）。

3. 或者，您可以将对象从管理器拖放到工程窗格中。

注：系统可能会让您先保存对象。保存时，一定要在“保存”对话框中选中**将文件添加到工程**。这样，在您保存对象后，将会自动将它添加到工程中。

从工作区添加节点

您可以使用“保存”对话框从流工作区添加单个节点。

1. 在工作区中选择一个节点。
2. 右键单击，然后单击**保存节点**。或者，在主菜单中，单击：

编辑 > 节点 > 保存节点...

3. 在“保存”对话框中，选中**将文件添加到工程**。
4. 为节点创建名称，然后单击 **保存**。

这样会保存文件并将它添加到工程中。节点会被添加到类视图中的 *Nodes* 文件夹以及 CRISP-DM 视图中的缺省阶段文件夹。

添加外部文件

您可以将各种非 IBM SPSS Modeler 对象添加到工程中。这在 IBM SPSS Modeler 中管理整个数据挖掘过程时十分有用。例如，可以将指向数据、注释、演示文稿和图片的链接存储在工程中。在 CRISP-DM 视图中，可将外部文件添加到您选择的文件夹。在类视图中，只能将外部文件保存到 *Other* 文件夹。

要将外部文件添加到工程：

1. 将文件从桌面拖到工程中。
或者
 2. 右键单击 CRISP-DM 视图或类视图中的目标文件夹。
 3. 在菜单中，单击**添加到文件夹**。
 4. 在对话框中选择一个文件，然后单击 **打开**。
- 这会在 IBM SPSS Modeler 工程内添加对所选对象的引用。

将工程转移至 IBM SPSS Collaboration and Deployment Services Repository

可以一步将整个工程（包括所有组件文件）转移到 IBM SPSS Collaboration and Deployment Services Repository。所有已经在目标位置的对象均不移动。此功能也可以反过来使用：可以将整个工程从 IBM SPSS Collaboration and Deployment Services Repository 转移到本地文件系统。

转移工程

确保在工程窗格中打开要转移的工程。

要转移工程，请完成下列步骤：

1. 右键单击根工程文件夹，然后单击**转移工程**。
2. 如果系统提示，请登录 IBM SPSS Collaboration and Deployment Services Repository。
3. 为工程指定新位置，然后单击 **确定**。

设置工程属性

可以使用工程属性对话框来自定义工程的内容和文档。要访问工程属性，请执行以下操作：

1. 右键单击工程窗格中的对象或文件夹，然后单击**工程属性**。
2. 单击 **工程**选项卡以指定基本工程信息。

已创建。 显示工程的创建日期（不可编辑）。

摘要。 您可以输入将在工程报告中显示的数据挖掘工程的摘要。

内容。 列出工程文件引用的组件的类型和数量（不可编辑）。

将未保存的对象另存为。 指定是应将未保存的对象保存到本地文件系统还是存储在存储库中。有关更多信息，请参阅主题 [第 163 页的『关于 IBM SPSS Collaboration and Deployment Services Repository』](#)。

装入项目时更新对象引用。 选中此选可更新工程对其组件的引用。注：添加到工程中的文件不会保存在工程文件本身中。更确切地说，工程中存储的是对这些文件的引用。这意味着移动或删除文件会将该对象从工程中删除。

注解工程

工程窗格提供了多种方式来注解数据挖掘工作。工程级注解常用于跟踪“宏观”目标和决策，而文件夹或节点注解提供其他详细信息。“注解”选项卡提供了足够空间供您记录工程级详细信息，比如具有不可恢复丢失数据的数据排除、数据探索过程中形成的可能假设等。

要注解工程：

1. 在 CRISP-DM 视图或类视图中选择工程文件夹。
2. 右键单击该文件夹，然后单击**工程属性**。
3. 单击 **注解**选项卡。
4. 输入用于描述工程的关键字和文本。

文件夹属性和注解

可以对（CRISP-DM 视图和类视图中的）单个工程文件夹进行注解。在 CRISP-DM 视图中，这可能是记录组织在每个数据挖掘阶段的目标的一种极为有效的方式。例如，对 *Business Understanding* 文件夹使用注解

工具，就可以进行这样的记录：“此研究的业务目标是减少高价值客户的流失”。然后，可通过选中 **纳入报告** 选项自动将此文本包括在工程报告中。

要注解文件夹：

1. 在工程窗格中选定一个文件夹。
2. 右键单击该文件夹，然后单击 **文件夹属性**。

在 CRISP-DM 视图中，用每个阶段的目的摘要以及完成相关数据挖掘任务的指导注解文件夹。您可以删除或编辑这些注解中的任何一个。

名称。 此区域显示所选字段的名称。

工具提示文本。 创建将鼠标指针悬停在工程文件夹上将显示的自定义工具提示。这在 CRISP-DM 视图中非常有用，例如，快速概览每个阶段的目标或标记某个阶段的状态（如“进行中”或“完成”）。

注解字段。 此字段用于进行可在工程报告中详细核对的更长注解。CRISP-DM 视图在注解中包括每个数据挖掘阶段的描述，但您可自由地为自己的工程自定义此描述。

包含在报告中。 要在报告中纳入注解，请选中 **纳入报告**。

对象属性

您可以查看对象属性并选择是否在工程报告中纳入单个对象。要访问对象属性，请执行以下操作：

1. 右键单击工程窗格中的对象。
2. 在菜单中，单击 **对象属性**。

名称。 此区域列示已保存的对象的名称。

路径。 此区域列示已保存的对象的位置。

包含在报告中。 选中此选项将在生成的报告中纳入对象详细信息。

关闭工程

当您退出 IBM SPSS Modeler 或打开新工程时，现有工程文件 (.cpj) 将被关闭。

一些与工程（如流、节点或图形）关联的文件仍可能打开。如果要使这些文件保持打开状态，请针对以下消息回复否：**是否要保存并关闭这些文件？**

如果在关闭工程之后修改并保存任何关联文件，则下次打开工程时，会把这些更新过的文件包括进来。如果不希望将更新过的文件包括进来，请将该文件从工程中删除或用其他文件名保存它。

生成报告

工程最有用的一项功能是能够根据工程项和注解生成报告。正如 CRISP-DM 方法通篇所讨论的，这是有效数据挖掘的关键要素。您可以直接生成若干种文件类型的报告，也可以直接输出到屏幕窗口以便立即查看。从输出窗口中，可以打印、保存或查看 Web 浏览器中的报告。还可以将保存的报告分发给组织中的其他人。

在数据挖掘过程中，经常会根据工程文件多次生成报告以分发给工程参与人员。报告从工程文件以及所创建的全部注解中精选有关所引用对象的信息。您可以基于类视图或 CRISP-DM 视图创建报告。

要生成报告：

1. 在 CRISP-DM 视图或类视图中选择工程文件夹。
2. 右键单击该文件夹，然后单击 **工程报告**。
3. 指定报告选项并单击 **生成报告**。

报告对话框中的选项提供多种方式来生成所需类型的报告：

输出名称。 如果选择将报告输出发送到屏幕，请指定输出窗口的名称。您可以指定自定义名称，或让 IBM SPSS Modeler 自动为您命名窗口。

输出到屏幕。 选择此选项可生成报告并在输出窗口中进行显示。请注意，您可以选择将报告从输出窗口导出为各种类型的文件。

输出到文件。 选择此选项可生成报告并将其保存为在“文件类型”列表中指定类型的文件。

文件名。 为生成的报告指定文件名。缺省情况下，这些文件将保存到 IBM SPSS Modeler `\bin` 目录中。使用省略号按钮 (...) 可指定其他位置。

文件类型。 可用文件类型包括：

- **HTML 文档。** 报告保存为单个 HTML 文件。如果报告包含图形，则会将它们保存为 PNG 文件并被 HTML 文件引用。在因特网上发布报告时，一定要上传 HTML 文件以及其引用的所有图像。
- **文本文档。** 报告保存为单个文本文件。如果报告包含图形，则报告中只包括文件名和路径引用。
- **Microsoft Word 文档。** 报告保存为单个文档，且所有图形均直接嵌入在该文档中。
- **Microsoft Excel 文档。** 报告保存为单个电子表格，且所有图形均直接嵌入在该电子表格中。
- **Microsoft PowerPoint 文档。** 每个阶段均显示在一张新的幻灯片上。所有图形均直接嵌入在 PowerPoint 幻灯片中。
- **输出对象。** 在 IBM SPSS Modeler 中打开此文件 (.cou) 时，它与**报告格式**组中的**输出到屏幕**选项相同。

注：要导出为 Microsoft Office 文件，必须安装相应的应用程序。

标题。 指定报告的标题。

报告结构。 选择 **CRISP-DM** 或类。CRISP-DM 视图提供状态报告，其中包含“宏观”概要以及有关每个数据挖掘阶段的详细信息。类视图是基于对象的视图，它更适合对数据和流进行内部跟踪。

作者。 显示缺省用户名，但您可以更改它。

报告包含。 选择在报告中包括对象的方法。选择**所有文件夹和对象**将包括添加到工程文件中的所有项。您也可以根据是否在对象属性中选中了**纳入报告**来包括项。或者，要检查未报告项，可以选择只包括标记为排除的项（此时未选中**纳入报告**）。

选择。 使用此选项，可通过只选择报告中的**最近项目**来提供工程更新。或者，可通过为**旧项目**设置参数来跟踪较旧并且可能尚未解决的问题。选择**所有项目**即是在生成报告时不考虑时间因素。

排序依据。 您可以选择下列对象特征的组合以便在文件夹中为它们排序：

- **类型。** 按类型将对象分组。
- **名称。** 按字母顺序组织对象。
- **添加日期。** 根据对象添加到工程中的日期为对象排序。

保存和导出生成的报告

在屏幕上生成的报告显示在新输出窗口中。此报告中包括的所有图形都会显示为内嵌图像。

报告术语

报告中列出了每一个流的节点总数。该数显示在下列标题下，这些标题使用的是 IBM SPSS Modeler 术语，而不是 CRISP-DM 术语：

- **数据阅读器。** 源节点。
- **数据编写器。** 导出节点。
- **模型构建器。** 构建或建模节点。
- **模型应用器。** 生成的模型（也称为块）。
- **输出构建器。** 图形节点或输出节点。
- **其他。** 任何其他与工程相关的节点。例如，节点选项板上“字段选项”选项卡或“记录选项”选项卡上的可用节点。

要保存报告：

1. 在“文件”菜单中，单击**保存**。
2. 指定文件名。

报告即保存为输出对象。

要导出报告：

3. 在“文件”菜单中，单击**导出**以及要导出为的文件类型。
4. 指定文件名。

报告即以您选择的格式保存。

您可以将报告导出为如下类型的文件：

- HTML
- Text
- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint

注：要导出为 Microsoft Office 文件，必须安装相应的应用程序。

使用窗口顶部的按钮可以：

- 打印报告。
- 在外部 Web 浏览器中以 HTML 形式查看报告。

第 15 章 自定义 IBM SPSS Modeler

定制 IBM SPSS Modeler 选项

您可以通过多种操作来根据需要自定义 IBM SPSS Modeler。本自定义的内容主要包括设置特定的用户选项，如内存分配、缺省目录以及声音和颜色的用法。您也可以自定义位于 IBM SPSS Modeler 窗口底部的节点选用板。

设置 IBM SPSS Modeler 选项

定制和设置 IBM SPSS Modeler 选项有以下几种方式：

- 通过单击工具> 选项菜单上的**系统选项**来设置系统选项，例如内存使用情况和语言环境。
- 通过单击工具> 选项菜单上的**用户选项**来设置用户选项，例如显示字体和颜色。
- 通过单击工具> 选项菜单上的**帮助应用程序**来指定使用 IBM SPSS Modeler 的应用程序的位置。
- 通过在“文件”菜单中单击**设置目录**或**设置服务器目录**，指定在 IBM SPSS Modeler 中使用的缺省目录。

另外，还可以设置应用于某些或全部流的选项。有关更多信息，请参阅主题 [第 32 页的『设置流的选项』](#)。

系统选项

您可以通过单击工具> 选项菜单上的**系统选项**来指定 IBM SPSS Modeler 的首选语言或语言环境。还可以在此设置 SPSS Modeler 的最大内存使用量，并指定自动保存流的频率。请注意，在此对话框中所做的更改只有在重新启动 SPSS Modeler 后才会生效。

最大内存。 选择此选项可限制 IBM SPSS Modeler 的内存使用量 (MB)。在某些平台上，如果计算机资源有限或负载过重，则 SPSS Modeler 将限制其进程大小以减少计算机的资源耗费。如果所处理的数据数量很大，这有可能导致“内存不足”错误。可以通过指定新阈值来减轻内存负载。

例如，尝试显示非常大的决策树可能会引起内存错误。在这种情况下，我们建议您将内存量增大到最大值 4096Mb。在有可能处理非常大量的数据的此类情况下，增大内存限额之后，请关闭 SPSS Modeler，然后从命令行将其启动，以确保使用最大限度的内存量来处理数据。

要从命令行启动（假定 SPSS Modeler 安装在缺省位置中），请在命令提示符窗口中输入以下命令：

```
C:\Program Files\IBM\SPSS\Modeler\18.4.0\bin\modelerclient.exe" -J-Xss4096M
```

注：由于 Java 约束，您可以指定的最大内存阈值为 58368Mb。

使用系统语言环境。 缺省情况下，此选项处于选中状态并被设置为“英语(美国)”。取消选中此选项可从可用语言和语言环境的下拉列表中选择其他语言。

流自动保存时间间隔（分钟）。 指定希望 SPSS Modeler 自动保存流的频率。最大值为 60 分钟，最小值为 1 分钟，缺省值为 5 分钟。

管理内存

除了在“系统选项”对话框中指定**最大内存**设置外，还可以通过其他方式优化内存使用：

- 在流属性对话框中调整**最大名义字段数量**选项。此选项指定字段的测量级别变为无类型后名义字段的最大成员数。有关更多信息，请参阅主题 [第 32 页的『设置流的常规选项』](#)。
- 单击窗口右下角（此处显示 IBM SPSS Modeler 正使用的内存和已分配的内存 (xx MB/ xx MB)），强迫 IBM SPSS Modeler 释放内存。单击此区域使其变暗，随后内存分配数将下降。一旦此区域恢复正常颜色，则表示 IBM SPSS Modeler 已释放了所能释放的全部内存。

设置缺省目录

通过选择“文件”菜单中的**设置目录**或**设置服务器目录**，可以指定用于文件浏览器和输出的缺省目录。

- **设置目录**。可以使用此选项来设置工作目录。缺省的工作目录是根据您 IBM SPSS Modeler 版本的安装路径进行确定的，或者是用于启动 IBM SPSS Modeler 的命令行路径。在本地模式下，工作目录是用于所有客户端操作和输出文件（如果是通过相对路径引用文件）的路径。
- **设置服务器目录**。只要有远程服务器连接，“文件”菜单中的“设置服务器目录”选项都处于启用状态。使用此选项可以为所有服务器文件和用于输入或输出的数据文件指定缺省目录。缺省服务器目录为 *\$CLEO/data*，其中 *\$CLEO* 是服务器版本 IBM SPSS Modeler 的安装目录。在命令行模式下，还可以通过使用带有 `modelerclient` 命令行参数的 `-server_directory` 标志来覆盖此缺省目录。

设置用户选项

您可以通过从**工具> 选项**菜单中选择**用户选项**来设置 IBM SPSS Modeler 的常规选项。这些选项适用于 IBM SPSS Modeler 中使用的所有流。

通过单击相应选项卡可以设置以下类型的选项：

- 通知选项，如模型覆盖和错误消息。
- 显示选项，如图形和背景颜色。
- 语法颜色显示选项。
- PMML 导出选项，将模型导出为预测模型标记语言 (PMML) 时使用。
- 用户或作者信息，例如姓名、姓名首字母和电子邮件地址。此信息可能会显示在节点以及您创建的其他对象的“注解”选项卡上。
- 在传统方式和 Analytic Server 方式之间切换。

要设置特定于流的选项，例如十进制分隔符、时间格式及数据格式、优化、流布局以及流脚本，请使用可以从“文件”和“工具”菜单中打开的“流属性”对话框。

设置通知选项

使用“用户选项”对话框的“通知”选项卡，可以设置有关 IBM SPSS Modeler 中警告和确认窗口的出现及类型的各种选项。您还可以指定生成新的输出和模型时管理器窗格中“输出”和“模型”选项卡的行为。

显示流执行反馈对话框 选中此项可在流运行三秒时显示包含进度指示器的对话框。此对话框还包括有关流所创建的输出对象的详细信息。

- **完成之后关闭对话框** 缺省情况下，当流完成运行时关闭该对话框。如果您希望此对话框在流完成后保持可见，请取消选中此复选框。

当节点覆盖文件时发出警告 选中此项将在节点操作覆盖现有文件时发出警告并显示错误消息。

当节点覆盖数据库表时发出警告 选中此项将在节点操作覆盖现有数据库表时发出警告并显示错误消息。

声音提示

使用该列表指定是否在发生事件或错误时用声音进行通知。这里提供了多种声音。使用“播放”（扬声器）按钮可以播放选择的聲音。使用省略号按钮 (...) 可浏览和选择一种声音。

注：IBM SPSS Modeler 中用于创建声音的 .wav 文件存储在安装位置的 `/media/sounds` 目录下。

- **全部静音** 选中此项将关闭所有事件的声音通知。

可视提示

此组中的选项用于指定生成新项时显示的内容中右上角的管理器窗格中“输出”和“模型”选项卡的行为。从列表中选择**新建模型**或**新建输出**可指定相应选项卡的行为。

下列选项适用于**新建模型**：

替换原有模型 如果选中（缺省），将覆盖“模型”选项卡和流工作区中该流的现有模型。如果取消选中此框，那么该模型将添加到选项卡和画布上的现有模型中。请注意，此设置将被模型链接上的模型替换设置覆盖。

下列选项适用于**新建输出**：

输出超出 [n] 时覆盖 选择当“输出”选项卡上的项数超过预先指定的数量时是否显示警告。缺省数量为 20；但是，可以根据需要更改此缺省值。

下列选项在所有情况下都可用：

选择选项卡 选择在流运行过程中有相应对象生成时是否切换到“输出”或“模型”选项卡。

- 选择**始终**将切换到管理器窗格中的相应选项卡。
- 选择**若由当前流生成**则只为由工作区中当前可见流生成的对象切换到相应选项卡。
- 选择**从不**将限制软件以切换到相应选项卡的方式来通知您生成了输出或模型。

闪烁选项卡 选择当新的输出或模型已生成时是否闪烁管理器窗格中的“输出”或“模型”选项卡。

- 选择**若未选择**将在管理器窗格中有新对象生成时闪烁相应选项卡（如果尚未选定）。
- 选择**从不**将限制软件通过闪烁相应选项卡来通知您有对象生成。

滚动选项板至可见（仅限**新建模型**）。选择是否自动滚动管理器窗格中的“模型”选项卡以使最新生成的模型可见。

- 选择**始终**将启用滚动。
- 选择**若由当前流生成**则仅对由工作区中当前可见流生成的对象滚动。
- 选择**从不**将限制软件自动滚动“模型”选项卡。

打开窗口（仅限**新建输出**）。选择是否在生成后自动打开输出窗口。

- 选择**始终**将总是打开新的输出窗口。
- 选择**若由当前流生成**则为由工作区中当前可见流生成的输出打开一个新窗口。
- 选择**从不**将限制软件自动为生成的输出打开新窗口。

要恢复为此选项卡的系统缺省设置，请单击**缺省值**。

设置显示选项

使用“用户选项”对话框的“显示”选项卡，可以设置 IBM SPSS Modeler 中字体和颜色的显示选项。

启动时显示“欢迎”对话框。选择此选项可在启动时显示“欢迎”对话框。欢迎对话框提供了选项，这些选项可以启动应用程序示例教程、打开演示流或者现有流或工程以及创建新流。

显示流和超节点标记。如果选中，可默认显示流和超节点中的标记（如果有）。标记包括流注释、模型链接以及评分分支突出显示。

标准字体和颜色（重新启动时生效）。此控件框中的选项用于指定 IBM SPSS Modeler 屏幕设计、颜色方案以及显示字体的大小。此处选择的选项直到您关闭并重新启动 IBM SPSS Modeler 后才会生效。

- **外观**。请选择标准颜色方案和屏幕设计。您可以从下列各项中进行选择：
 - **SPSS 标准**，缺省设计。
 - **SPSS Classic**，这是 SPSS Modeler 较早版本的用户熟悉的设计。
 - **Windows**，此 Windows 设计有助于加强流工作区和选用板的对比度。
 - **Analytics Carbon**，具有光滑图标和颜色的现代化设计。
- **节点的缺省字体大小**。指定要用于节点选用板以及用于流工作区中显示的节点的字体大小。
- **指定固定宽度字体**。要选择固定宽度字体以及相关联的字体大小以用于脚本和 CLEM 表达式控件，请选中此复选框。缺省字体为 Monospace plain；单击**更改...**以显示您可以选择的其他字体的列表。

注：您可以在流属性对话框的“选项”选项卡的“布局”窗格中设置流的节点图标大小。从主菜单中，选择**工具 > 流属性 > 选项 > 布局**。

定制颜色。 此表列出用于各种显示项的当前所选颜色。对于表中所列的每个项，您可以双击**颜色**列中的相应行并从列表中选择一种颜色，以更改当前颜色。要指定定制颜色，请滚动到列表底部，然后单击**颜色...** 条目。

图表类别颜色顺序。 此表列出用于新创建的图形中的显示的当前所选颜色。这些颜色的顺序反映了它们用于图表的顺序。例如，如果使用了名义字段作为包含四个唯一值的颜色覆盖，那么将只使用此处列示的前四种颜色。对于表中所列的每个项，您可以双击**颜色**列中的相应行并从列表中选择一种颜色，以更改当前颜色。要指定定制颜色，请滚动到列表底部，然后单击**颜色...** 条目。此处所作的更改不会影响先前创建的图形。

要恢复为此选项卡的系统缺省设置，请单击**缺省值**。

设置语法显示选项

使用“用户选项”对话框的“语法”选项卡，可以设置在 IBM SPSS Modeler 中创建的脚本中的字体属性和显示颜色的选项。

语法突出显示。 此表列出了当前所选的用于各种语法项目（包括字体和可在其中显示字体的窗口）的颜色。对于表中所列的每个项目，可通过单击行中对应下拉列表，并从列表中选择一种颜色，以更改颜色。此外，对于字体项目，可以选择添加粗体和斜体强调。

预览。 此表显示了一个使用在**语法突出显示**表中选择的颜色和字体属性的语法显示示例。此预览将在您更改任何选择后更新。

单击 **缺省值** 可将此选项卡还原为系统缺省设置。

设置 PMML 导出选项

在“PMML”选项卡上，可以控制 IBM SPSS Modeler 如何将模型导出为预测模型标记语言 (PMML)。有关更多信息，请参阅主题 [第 179 页的『导入和导出 PMML 模型』](#)。

导出 PMML。 可以在此配置最适合目标应用程序的 PMML 变体。

- 选择**带有扩展**允许将 PMML 扩展名用于没有等效标准 PMML 的特例。注意，多数情况下，这会得到与标准 PMML 相同的结果。
- 选择**作为标准 PMML...** 以导出与 PMML 标准尽可能接近的 PMML。

标准 PMML 选项。 当选中**作为标准 PMML...** 选项时，您可以选择以下两种有效方式之一来导出线性回归模型和逻辑回归模型：

- 作为 **PMML <GeneralRegression>模型**
- 作为 **PMML <Regression>模型**

有关 PMML 的详细信息，请参阅数据挖掘组网站 (<http://www.dmg.org>)。

设置用户信息

用户/作者信息。 您在此处输入的信息，可以显示在您创建的节点和其他对象的“注解”选项卡上。

设置方式

Modeler 方式设置。 在方式选项卡上，可以从以下方式中进行选择：

- **传统 SPSS Modeler 模式** 在用户界面中显示所有可用节点和表达式。
- **Analytic Server 模式** 仅显示 Analytic Server 支持的节点和表达式。但请注意，某些节点和某些 CLEM 表达式即使不受 Analytic Server 的完全支持也仍然会显示。下表提供了哪些节点受到 Analytic Server 支持、部分支持以及不受支持的一般信息。

请参阅 [Analytic Server 文档](#) 以了解有关 Analytic Server 的更多信息。

请注意，如果配置 SPSS Modeler 以在“数据库建模”选用板上显示数据库节点，那么它们在切换模式时不受影响。将总是显示数据库节点。如果使用 IBM Db2 for z/OS 或 IBM Netezza 集成，那么在某些情况下，在切换到 Analytic Server 模式后这些节点可能在“数据库建模”选用板上消失。如果发生此情况，请转至 **工具 > 选项 > 帮助应用程序**，并重置复选框。

表 38: 节点支持

节点类型 (选用板名称)	受 Analytic Server 支持	受 Analytic Server 部分支持	不受 Analytic Server 支持
源	<ul style="list-style-type: none"> • Analytic Server 源节点 		<ul style="list-style-type: none"> • 数据库 • 变量 文件 • 固定文件 • 统计文件 • 数据收集 • IBM Cognos • TWC 导入 • TM1 导入 • SAS 文件 • Excel • XML 用户输入 • Sim Gen • 数据视图 • 地理空间 • Object Storage • 扩展导入 • R 导入 • SNA (扩散分析和组分析)
记录选项	<ul style="list-style-type: none"> • 选择 • 排序 • 平衡 • 区分 • RFM 汇总 • 追加 • 流式 TS • 扩展变换 • 流式 TCM 	<ul style="list-style-type: none"> • 样本 (仅对简单方法支持随机百分比。复杂不受支持。) • 合并 (仅支持按键和条件合并) • 汇总 (第一个分位数、第三个分位数和中位数不受 Analytic Server 支持) 	<ul style="list-style-type: none"> • 空间时间限制 • CPLEX Optimization

表 38: 节点支持 (继续)

节点类型 (选用板名称)	受 Analytic Server 支持	受 Analytic Server 部分支持	不受 Analytic Server 支持
字段选项	<ul style="list-style-type: none"> • Type • 过滤 • 导出 • 填充 • 重新分类 • 整体 • 设为标志 • 重建 • 字段重排 • 重新投影 • 时间间隔 	<ul style="list-style-type: none"> • 自动数据准备 (仅支持变换) • 分级 (如果选中“结”属性保留在当前分级中, 那么不支持分级方法 equalFreq) • RFM 分析 (如果选中“结”属性保留在当前分级中, 那么不支持分级方法 Tiles) • 分区 (不受 Analytic Server 支持, 除非使用唯一字段重复地向分区分配行) 	<ul style="list-style-type: none"> • 匿名化 • 历史记录 • 转置
图形	<ul style="list-style-type: none"> • 统计图 • 多重散点图 • 时间散点图 • 分布 • 直方图 • 收集 • Web • 评估 • 地图可视化 • E-Plot (Beta) • t-SNE 	<ul style="list-style-type: none"> • 图形板 (仅对测量级别为离散、名义、有序或标志的字段支持汇总方式函数) 	

表 38: 节点支持 (继续)

节点类型 (选用板名称)	受 Analytic Server 支持	受 Analytic Server 部分支持	不受 Analytic Server 支持
建模	<ul style="list-style-type: none"> • 时间序列 • TCM • Isotonic-AS • 随机树 • 树 AS • 线性 AS • GLE • LSVM • STP • 二阶 AS • 关联规则 • XGBoost-AS • K-Means-AS 	<ul style="list-style-type: none"> • 自动分类器 • 自动数值 (这两个节点仅支持分割, 当使用“自动分类器”选项启用分割的情况下在 Analytic Server 上运行时必须提供具有分割角色的字段) • 扩展 (R 语法构建模型不受 Analytic Server 支持) • 以下节点仅支持分割和 PSM: <ul style="list-style-type: none"> - C&R 树 - 线性 - 神经网络 - CHAID - Quest 	<ul style="list-style-type: none"> • 自动聚类 • 决策列表 • C5.0 • 回归 • PCA/因子 • 特征选择 • 判别 • Logistic • GenLin • GLMM • 贝叶斯网络 • Apriori • Carma • 序列 • K-Means • Kohonen • 二阶 • 异常 • KNN • R • 随机森林 • 以下节点具有 asl 但只是读/写 asl: <ul style="list-style-type: none"> - Cox - SVM - SLRM
输出	<ul style="list-style-type: none"> • 表 • 矩阵 • 分析 • 数据审核 • 变换 • 统计信息 • 平均值 • 报告 • 设置全局量 	<ul style="list-style-type: none"> • 扩展输出 • R 语法输出 	<ul style="list-style-type: none"> • 模拟评估 • 模拟拟合 • R 输出

表 38: 节点支持 (继续)

节点类型 (选用板名称)	受 Analytic Server 支持	受 Analytic Server 部分支持	不受 Analytic Server 支持
导出	<ul style="list-style-type: none"> Analytic Server 导出节点 	<ul style="list-style-type: none"> 扩展导出 R 语法导出 	<ul style="list-style-type: none"> 数据库 平面文件 统计导出 数据收集 Excel IBM Cognos 导出 TM1 导出 SAS XML 导出 Object Storage R 导出
IBM SPSS Statistics			<ul style="list-style-type: none"> 统计文件 统计变换 统计统计信息 Statistics 输出 统计导出
IBM SPSS Text Analytics	<ul style="list-style-type: none"> 文本链接分析 文本挖掘 语言节点 		<ul style="list-style-type: none"> 文件列表 Web 订阅源 文本链接分析 翻译 文本挖掘 文件查看器
Python			<ul style="list-style-type: none"> SMOTE 单类 SVM XGBoost Tree XGBoost Linear t-SNE 随机森林 HDBSCAN
Spark	全部受支持		

自定义节点选项板

流是通过使用节点构建的。IBM SPSS Modeler 窗口底部的节点选用板包含构建流时可能需要使用的所有节点。有关更多信息，请参阅主题 [第 11 页的『节点选用板』](#)。

可以通过以下两种方式重新组织节点选项板：

- 自定义选项板管理器。有关更多信息，请参阅主题 [第 197 页的『自定义选项板管理器』](#)。
- 更改包含子选用板的选用板选项卡在节点选用板上的显示方式。有关更多信息，请参阅主题 [第 198 页的『创建子选项板』](#)。

自定义选项板管理器

根据 IBM SPSS Modeler 对您的用途，您可以自定义选用板管理器。例如，如果您经常分析数据库中的时间序列数据，则可能需要确保在单个选项板选项卡中可同时使用“数据库源”节点、时间间隔节点、时间序列节点，以及时间散点图节点。通过选项板管理器，只要在节点选项板中创建自定义选项板选项卡，您就可以轻松进行这些调整。

通过选项板管理器，您可以执行以下多种任务：

- 控制流工作区中节点选用板上所显示的选用板选项卡。
- 更改选用板选项卡在节点选用板上的显示顺序。
- 创建并编辑您自己的选用板选项卡以及任何相关联的子选用板。
- 编辑选项卡上的缺省节点选择。

要访问选用板管理器，请在“工具”菜单中单击**管理选项板**。

选用板名称。 列出了所有可用的选项板选项卡，无论其是否显示在节点选项板上。这包含已创建的任何选项板选项卡。有关更多信息，请参阅主题 [第 197 页的『创建选项板选项卡』](#)。

节点数。 每个选项板选项卡上所显示的节点的数量。此处的数目较大意味着创建子选用板以便对选项卡上的节点进行划分会更为方便。有关更多信息，请参阅主题 [第 198 页的『创建子选项板』](#)。

是否已显示？ 选择此字段可显示节点选项板上的选项板选项卡。有关更多信息，请参阅主题 [第 197 页的『显示节点选项板上的选项板选项卡』](#)。

子选用板。 要选择选项板选项卡上显示子选项板，请突出显示所需的**选项板名称**，然后单击此按钮以显示“子选项板”对话框。有关更多信息，请参阅主题 [第 198 页的『创建子选项板』](#)。

恢复缺省值。 要完全删除您对选项板和子选项板所做的所有更改和添加，并恢复缺省的选项板设置，请单击此按钮。

创建选项板选项卡

要创建自定义选项板选项卡：

1. 在“工具”菜单中，打开“选用板管理器”。
2. 在已显示? 列的右侧，单击“添加选用板”按钮；将显示“创建/编辑选用板”对话框。
3. 键入唯一的 **选用板名称**。
4. 在 **可用节点** 区域，选择要添加至选用板选项卡的节点。
5. 单击“添加节点”右箭头按钮，将突出显示的节点移动到**选定节点**区域。重复该操作，直到添加完所需的所有节点。

在添加所有必需节点之后，可以更改这些节点在选用板选项卡上的显示顺序：

6. 使用简单的方向按钮可将节点上移或下移一行。
7. 使用行方向按钮可将节点移至列表底部或顶部。
8. 要从选项板中删除节点，请突出显示该节点，然后单击**所选节点**区域右侧的“删除”按钮。

显示节点选项板上的选项板选项卡

在 IBM SPSS Modeler 中有一些可用的选项，您可能从来都不需要使用它们；在此情况下，可以使用“选用板管理器”隐藏包含这些节点的选项卡。

要选择将在节点选项板上显示的选项卡：

1. 在“工具”菜单中，打开“选用板管理器”。
2. 使用已显示? 列中的相应复选框，选择是包含还是隐藏每个选用板选项卡。

要从节点选用板中永久除去选用板选项卡，请突出显示该节点，并单击已显示? 列右侧的“删除”按钮。一旦删除，选项板选项卡将无法恢复。

注：除“收藏夹”选项卡以外，您无法删除 IBM SPSS Modeler 提供的任何缺省选用板选项卡。

更改节点选项板上的显示顺序

在选择要显示的选用板选项卡之后，可以更改它们在“节点选用板”上的显示顺序：

1. 使用简单的箭头按钮可将选项板选项卡上移或下移一行。将这些选用板选项卡上移可将它们移至节点选用板的左侧，反之亦然。
2. 使用行箭头按钮可将选项板选项卡移至列表底部或顶部。位于列表顶部的选项板选项卡将显示在节点选项板左侧。

在选项板选项卡上显示子选项板

以控制选用板选项卡在“节点选用板”上的显示顺序的方式，可以控制哪些子选用板显示在它们的父选用板选项卡中。

要选择子选用板以便将其显示在选用板选项卡上，请执行以下操作：

1. 在“工具”菜单中，打开“选用板管理器”。
2. 选择所需的选用板。
3. 单击“子选用板”按钮，将显示“子选用板”对话框。
4. 使用已显示?列中的相应复选框，选择是否在选用板选项卡上包含每个子选用板。**所有**子选用板会始终显示，且无法删除。
5. 要从选用板选项卡中永久除去子选用板，请突出显示子选用板，并单击已显示?列右侧的“删除”按钮。

注：无法删除“建模”选用板选项卡随附的缺省子选用板。

更改选项板选项卡上的显示顺序

在选择要显示子选用板之后，可以更改它们在父选用板选项卡上的显示顺序：

1. 使用简单的方向按钮将子选用板向上或向下移动一行。
2. 使用“线箭头”按钮可以将子选用板移至列表底部或顶部。

您创建的子选用板将在您选择其父选用板选项卡时显示在“节点选用板”中。有关更多信息，请参阅主题 [第 199 页的『更改选项板选项卡视图』](#)。

创建子选项板

由于可以将任何现有的节点添加到您创建的自定义选项板选项卡中，因此您可以选择更多的节点，这样无需滚动屏幕就可以在屏幕上查看这些节点。为了避免必须进行滚动，您可以创建子选用板，并将您为选用板选项卡选择的节点放入其中。例如，如果您创建了一个选用板选项卡，其中包含通常用于创建流的节点，那么可以创建 4 个子选用板以便按源节点、字段操作、建模和输出来划分选择。

注：您只能从那些已添加到父选用板选项卡的子选用板节点中进行选择。

要创建子选用板，请执行以下操作：

1. 在“工具”菜单中，打开“选用板管理器”。
2. 选择要将子选用板添加到的选用板。
3. 单击“子选用板”按钮，将显示“子选用板”对话框。
4. 在已显示?列的右侧，单击“添加子选用板”按钮；将显示“创建/编辑子选用板”对话框。
5. 键入唯一的 **子选用板名称**。
6. 在**可用节点**区域，选择要添加至子选项板的节点。
7. 单击“添加节点”右方向按钮，将所选节点移动到**选定节点**区域。
8. 添加完所需的节点后，单击**确认**以返回至“子选项板”对话框。

您创建的子选用板将在您选择其父选用板选项卡时显示在“节点选用板”中。有关更多信息，请参阅主题 [第 199 页的『更改选项板选项卡视图』](#)。

更改选项板选项卡视图

由于 IBM SPSS Modeler 中存在很多可用的节点，因此如果不滚动至节点选用板左侧或右侧，则可能无法在较小的屏幕上显示所有的节点；这种情况在建模选用板选项卡上尤其明显。要减少滚动操作，您可以选择仅显示子选项板（如果存在）中包含的节点。有关更多信息，请参阅主题 [第 198 页的『创建子选项板』](#)。

要更改选用板选项卡上显示的节点，请选择选用板选项卡，然后从左侧的菜单中选择显示所有节点或者仅显示特定子选用板中的节点。

第 16 章 流和节点的性能注意事项

可对流进行设计以获得最佳性能，主要方法有：使用最有效的配置排列节点；必要时启用节点缓存；以及注意本节中详细列出的其他注意事项。

除了此处讨论的注意事项外，您通过有效地使用数据库（尤其是通过对 SQL 进行优化处理），通常可以获得额外的更加实质性的性能提高。

节点的顺序

即使未在使用 SQL 优化，流中节点的顺序也会影响性能。一般目标是尽量减少下游处理；因此，当您具有的节点减少了数据量时，请将它们放置在流的开头附近。IBM SPSS Modeler Server 可以在编译期间自动应用一些重新排序规则，以便在证实可以安全操作时将某些节点提前。（此功能缺省处于启用状态。请向系统管理员核实，以确保它在您的安装中处于启用状态。）

使用 SQL 优化时，您会希望最大程度地提高它的可用性和效率。由于当流包含无法在数据库中执行的操作时优化会停止，因此，最好在流的起点将 SQL 优化的操作组成一组。此策略可将大多数处理保持在数据库中进行，因此进入 IBM SPSS Modeler 中的数据较少。

在大多数数据库中，均可执行下列操作。尝试在流的起点将它们组为一组：

- 按关键字合并（连接）
- 选择
- 汇总
- 排序
- 样本
- 追加
- 在包括模式下执行的可区分操作，在该模式下将选择所有字段
- 填充操作
- 使用标准算术或字符串操纵的基本派生操作（具体取决于数据库支持哪些操作）
- 设为标志

在大多数数据库中，均无法执行下列操作。应在完成先前列表中的操作之后将下列操作放置在流中：

- 针对任何非数据库数据（如平面文件）的操作
- 按顺序合并
- 平衡
- 在丢弃模式下或只选择了部分字段作为可区分字段的情况下执行的可区分操作
- 需要访问来自记录（正在处理之中的记录除外）的数据的任何操作
- 状态和计数字段导出
- 历史记录节点操作
- 涉及“@”（时间序列）函数的操作
- 类型检查模式“警告”和“中止”
- 模型构造、应用和分析

注：决策树、结果集、线性回归和因素生成的模型可以生成 SQL，因此可以将其推送回到数据库中。

- 在除正在处理数据的数据库以外的任何位置进行的数据输出

节点缓存

为优化流运行，可以在任何非终端节点上建立缓存。如果已在节点上建立了缓存，则在下一次运行数据流时流过节点的数据会充满缓存。从这时起，就会从缓存（存储在磁盘的临时目录中）而不是从数据源中读取数据了。

对于诸如排序、合并或汇总这样比较耗时的操作，缓存是最有用的。例如，假设已设置可从数据库中读取销售数据的源节点，和可按区域汇总销售数据的“汇总”节点。可以在“汇总”节点上建立缓存而不是在源节点上建立缓存，因为希望缓存存储的是已汇总的数据而不是整个数据集。

注：源节点位置的缓存仅用于存储读入 IBM SPSS Modeler 的原始数据复本，在大多数情况下，它不会带来性能方面的提升。

对于已启用缓存的节点，会在右上角显示一个小的文档图标。当节点上的数据被缓存时，文档图标将变绿。

启用缓存

1. 在流工作区中右键单击节点，然后在菜单中单击**高速缓存**。
2. 在缓存子菜单中，单击**启用**。
3. 可以通过右键单击节点，然后在缓存子菜单中单击**禁用**来关闭缓存。

数据库中的缓存节点

对于在数据库中运行的流，可于中流将数据缓存到数据库的临时表中而不是文件系统中。如果与 SQL 优化相组合，则此操作将使性能得到显著提高。例如，可以对合并多个表以创建数据挖掘视图的流的输出进行缓存并在需要时重新使用。通过为所有下游节点自动生成 SQL，性能可得到进一步的提高。

为利用数据库缓存，必须同时启用 SQL 优化和数据库缓存。注意，服务器上的优化设置将覆盖客户端上的优化设置。有关更多信息，请参阅主题第 34 页的『设置流的优化选项』。

如果已启用数据库缓存，则只需右键单击任意非终端节点就可以缓存该点上的数据，并且在下次运行流时，可以直接在数据库中自动创建缓存。如果未启用数据库缓存或 SQL 优化，则会转而将缓存写入文件系统中。

注：下列数据库支持用于缓存的临时表：Db2、Oracle、SQL Server 和 Teradata。其他数据库（例如 Netezza）将使用标准表进行数据库缓存。可以针对特定数据库定制 SQL 代码 - 请与服务人员联系以获取帮助。

性能：过程节点

排序。排序节点必须读取整个输入数据集才能为该集进行排序。数据存储在内存在中，直到达到某个限制，并且超出限制的数据会溢出到磁盘上。排序算法是一种组合算法：将数据读取到内存中（直到达到某个限制）并使用混合式快速排序算法为其排序。如果所有数据均在内存中，则排序完成。否则，会应用合并排序算法。将排序后的数据写至文件中，并将下一个数据组块读取到内存中，进行排序，然后写到磁盘上。此处理会一直重复，直到读取了所有数据；然后，合并排序后的组块。合并可能需要多次转移存储在磁盘上的数据。在使用高峰期，排序节点将在磁盘上拥有两个完整的数据集副本：已排序和未排序。

算法的总体运行时间按 $N \times \log(N)$ 的顺序排列，其中 N 是记录数。在内存中排序比从磁盘进行合并速度快，因此通过分配更多内存来进行排序可减少实际运行时间。该算法为自己分配一部分物理 RAM（由 IBM SPSS Modeler Server 配置选项内存使用倍增器控制）。要增加用于排序的内存，请提供更多物理 RAM 或增加此值。请注意，当使用的内存比率超过过程的工作集从而导致部分内存分页到磁盘上时，性能会下降，原因是：内存内排序算法的内存访问模式是随机的，可导致过度分页。排序算法由排序节点以外的多个节点使用，但是应用的性能规则相同。

分箱。分箱节点读取整个输入数据集来计算分箱边界，然后为记录分箱。计算边界时会缓存数据集；然后，重新扫描数据集以进行分配。当分级方法是定宽或平均值+标准差时，直接将数据集缓存到磁盘上。这些方法具有线性运行时间，并且需要足够的磁盘空间来存储整个数据集。当分级方法为秩或分位数时，使用之前介绍的排序算法为数据集排序，并且将排序后的数据集作为缓存使用。排序为这些方法提供了 $M \times N \times \log(N)$ 的运行时间，其中 M 是分箱字段数， N 是记录数；它需要等于数据集大小两倍的磁盘空间。

根据已生成的箱生成导出节点将提高后续传递的性能。派生操作比分箱速度快很多。

按键合并（连接）。当合并方法为关键字时（等效于数据库连接），合并节点按键字段为它的每个输入数据集排序。此部分过程的运行时间为 $M \times N \times \log(N)$ ，其中 M 是输入数， N 是最大输入中的记录数；它需要足够的磁盘空间来存储其所有输入数据集，以及最大数据集的第二个副本。合并本身的运行时间与输出数据集的大小成比例，输出数据集的大小取决于匹配关键字的频率。在“输出是输入的笛卡尔乘积”这一最糟糕的情况下，运行时间可能会接近于 NM 。这种情况很少见，大多数连接的匹配关键字很少。如果一个数据集相对大于其他数据集，或者，如果已经按键字段为传入数据排序，则可以使用“优化”选项卡提高此节点的性能。

分类汇总。未设置键是连续的选项时，此节点先读取（但不存储）它的整个输入数据集，然后生成任意聚合的输出。在其他极端的情况下（聚合数据的大小达到某个限制，该限制由 IBM SPSS Modeler Server 配置选项内存使用倍增器确定），将会为数据集的其余部分进行排序并对其进行处理，就像设置了键是连续的选项一样。如果设置了此选项，则不会存储任何数据，原因是：聚合输出记录在读取输入数据时生成。

非重复值“区分”节点将所有唯一键字段存储在输入数据集中；如果所有字段都是键字段且所有记录都是唯一的，则其存储整个数据集。缺省情况下，“区分”节点排序键字段上的数据，然后从每个组选择（或丢弃）第一个区分记录。对于包含少量区分关键字的较小型数据集，或者对于预先进行了排序的数据集，可以选择选项以提高处理速度和效率。

类型。在一些情况下，类型节点在读取值时缓存输入数据；缓存用于下游的处理。缓存需要足够的磁盘空间来存储整个数据集，但会加速处理。

评估。评估节点必须为输入数据排序才能计算分位数。因为在每种情况下分值和所产生的记录顺序是不同的，所以会对所评估的每个模型重复进行排序。运行时间为 $M \times N \times \log(N)$ ，其中 M 是模型数量， N 是记录数量。

性能：建模节点

神经网络和 Kohonen。神经网络训练算法（包括 Kohonen 算法）会多次传递训练数据。数据存储在内存中，直到达到某个限制，并且超出限制的数据会溢出到磁盘上。从磁盘访问训练数据的成本很高，原因是：访问方法是随机的，这可能会导致磁盘活动过多。您可以针对这些算法禁用磁盘存储，强制将所有数据存储在内存中，方法是：在节点对话框的“模型”选项卡上选择**优化速度**选项。请注意，如果存储数据所需的内存量大于服务器过程的工作集，则会将数据的一部分分页到磁盘上，并且性能将相应地降低。

当**优化内存**启用时，会根据 IBM SPSS Modeler Server 配置选项**建模内存限制百分比**的值将一定百分比的物理 RAM 分配给算法。要使用更多内存来训练神经网络，请提供更多 RAM 或增加此选项的值，但是请注意，将该值设置得太大将导致分页。

神经网络算法的运行时间取决于所需的准确性水平。您可以通过在节点的对话框中设置停止条件来控制运行时间。

K-Means。K-Means 聚类算法具有与神经网络算法相同的内存使用控制选项。但是，因为对数据的访问是按顺序进行的，所以存储在磁盘上的数据的性能会较好。

性能：CLEM 表达式

对数据流进行回顾的 CLEM 序列函数（@ 函数）必须存储足够的数据才能满足最长的回顾。对于其回顾度未作限制的操作，必须存储字段的所有值。无界运算是指偏移值不是字面值整数的运算；例如，`@OFFSET(Sales, Month)`。偏移值是字段名称 *Month*，在执行之前其值是未知的。服务器必须保存 *Sales* 字段的所有值以确保结果准确。在已知上限的情况下，您应该将其作为附加参数提供；例如，`@OFFSET(Sales, Month, 12)`。此操作指示服务器最多存储 *Sales* 的 12 个最新值。序列函数（限制的或未限制的）几乎始终禁止 SQL 生成。

第 17 章 IBM SPSS Modeler 中的辅助功能

IBM SPSS Modeler 辅助功能概述

IBM SPSS Modeler 为所有用户提供辅助功能支持，还为视力和有其他功能性损伤的用户提供针对性支持。本节描述了使用辅助功能选项增强功能（例如，屏幕朗读器和键盘快捷键）的操作特点和方法。

辅助功能支持的类型

无论您视力受损还是依赖于键盘进行操作，都可以通过多种替代方法来使用此数据挖掘工具包。例如，您无需使用鼠标即可构建流，指定选项以及读取输出。后续主题列出了可用的键盘快捷键。另外，IBM SPSS Modeler 还广泛支持屏幕朗读器（例如 JAWS for Windows）。您还可以优化颜色方案，以提供更高的对比度。以下主题讨论了这些支持类型。

针对视力受损用户的辅助功能

您可以在 IBM SPSS Modeler 中指定多项属性，它们将增强使用此软件的能力。

显示选项

可以为图形显示选择颜色。也可以选择将特定的 Windows 设置用于此软件。这样可能有助于增强视觉对比度。

1. 要设置显示选项，请在“工具”菜单上单击**用户选项**。
2. 单击**显示选项卡**。此选项卡上的选项包括软件颜色方案、图表颜色和节点字体大小。

注：屏幕朗读器无法读取图形，因此有视力障碍的用户无法访问图形。

使用声音进行通知

通过开启或关闭声音，您可以控制接收针对软件中特定操作的警报的方式。例如，可以为创建和删除节点或者生成新的输出或模型之类的事件激活声音。

1. 要设置通知选项，请在“工具”菜单上单击**用户选项**。
2. 单击**通知选项卡**。

控制新窗口的自动启动

“用户选项”对话框上的“通知”选项卡还用于控制是否在另一个窗口中启动新生成的输出，例如表和图表。对您而言，禁用此选项并且仅在需要时打开输出窗口可能更为简单。

1. 要设置这些选项，请在“工具”菜单上单击**用户选项**。
2. 单击**通知选项卡**。
3. 在该对话框中，从**可视化通知组**的列表中选择**新建输出**。
4. 在**打开窗口**下，选择**从不**。

节点大小

可以使用标准大小或较小大小显示节点。您可能希望调整这些大小以满足您的需求。

1. 要设置节点大小选项，请在“文件”菜单上单击**流属性**。
2. 单击**布局选项卡**。
3. 从**图标大小**列表中，选择**标准**。

盲人用户的辅助功能选项

对盲人用户的支持主要通过屏幕阅读器实现，例如 JAWS for Windows。为了优化屏幕阅读器与 IBM SPSS Modeler 的配合使用，您可以指定多项设置。

显示选项

屏幕的视觉对比度越高，屏幕阅读器的运行效果越好。如果已具有高对比度的 Windows 设置，那么可以选择将这些 Windows 设置用于软件自身。

1. 要设置显示选项，请在“工具”菜单上单击**用户选项**。
2. 单击**显示选项卡**。

注：屏幕阅读器无法读取图形，因此有视力障碍的用户无法访问图形。

使用声音进行通知

通过开启或关闭声音，您可以控制接收针对软件中特定操作的警报的方式。例如，可以为创建和删除节点或者生成新的输出或模型之类的事件激活声音。

1. 要设置通知选项，请在“工具”菜单上单击**用户选项**。
2. 单击**通知选项卡**。

控制新窗口的自动启动

“用户选项”对话框上的“通知”选项卡还用于控制是否在另一个窗口中启动新生成的输出。对您而言，禁用此选项并且在需要时打开输出窗口可能更为简单。

1. 要设置这些选项，请在“工具”菜单上单击**用户选项**。
2. 单击**通知选项卡**。
3. 在该对话框中，从**可视化通知组**的列表中选择**新建输出**。
4. 在**打开窗口**下，选择**从不**。

键盘辅助功能选项

可以通过键盘访问产品功能。最基本的功能包括：按住 Alt 键加相应键激活窗口菜单（例如，按 Alt+F 可以访问“文件”菜单）或者按 Tab 键在对话框控件间进行滚动。但是，存在与产品的每个主窗口相关的特殊问题，并且浏览对话框时需要一些有用的提示信息。

本节将介绍了键盘辅助功能的亮点，包括打开流、使用节点对话框以及对输出执行操作等。另外，还提供了一个键盘快捷键列表，以提高浏览效率。

用于浏览主窗口的快捷方式

大部分数据挖掘工作都在 IBM SPSS Modeler 的主窗口中完成。该主要区域称为**流工作区**，用于构建和运行数据流。该窗口的底部包含**节点选用板**，其中包含了所有可用节点。选用板组织在选项卡上，这些选项卡与每个节点组的数据挖掘操作类型相对应。例如，用于将数据引入 IBM SPSS Modeler 的节点在“源”选项卡上进行分组，而用于派生、过滤字段或对字段设置类型的节点在“字段操作”选项卡上进行分组。

窗口的右侧包含一些用于管理流、输出和项目的工具。右侧的上半部分包含**管理器**以及分别用于管理流、输出和已生成模型的三个选项卡。通过选择选项卡，然后从列表中选择对象，您可以访问这些对象。右侧的下半部分包含**项目窗格**，您可以通过它将工作分组为项目。此区域中有两个选项卡，它们反映了项目的两个不同视图。**类视图**会按类型对项目对象进行排序，而**CRISP-DM 视图**按相关的数据挖掘阶段（例如，数据准备或建模）对项目对象进行排序。帮助系统和《用户指南》中介绍了该 IBM SPSS Modeler 窗口的这些不同部分。

以下是用于在 IBM SPSS Modeler 主窗口中移动以及构建流的快捷键表。后续主题中列出了用于对话框和输出的快捷键。请注意，只有在主窗口中才可以使用这些快捷键。

快捷键	函数
Ctrl+F5	将焦点移至节点选用板。
Ctrl+F6	将焦点移至流工作区。
Ctrl+F7	将焦点移至管理器窗格。
Ctrl+F8	将焦点移至项目窗格。

快捷键	函数
Ctrl+N	创建新的空白流工作区。
Ctrl+O	显示“打开”对话框，您可以从中选择并打开现有流。
Ctrl+数字键	将焦点移至窗口或窗格上的相应选项卡。例如，在选项卡式窗格或窗口中，Ctrl+1 用于将焦点移至左起第一个选项卡，Ctrl+2 用于将焦点移至左起第二个选项卡，依此类推。
Ctrl+向下方向键	在节点选用板中，用于将焦点从选用板选项卡移至该选项卡下的第一个节点。
Ctrl+向上方向键	在节点选用板中，用于将焦点从某个节点移至其选用板选项卡。
进入法	如果在节点选用板中选择了某个节点（包括已生成的模型选用板中的优化模型），那么按下此键会将该节点添加至流工作区。在工作区中选择了某个节点后，按 Enter 将打开该节点的对话框。
Ctrl+Enter	在选用板中选择某个节点后，此快捷键会将该节点添加至流工作区而无需将其选中，并且会将焦点移至选用板中的第一个节点。
Alt+Enter	在选用板中选择某个节点后，此快捷键会将该节点添加至流工作区并选中它，同时会将焦点移至选用板中的第一个节点。
Shift+空格键	焦点位于选用板中的节点或注释上时，此快捷键用于在选中与取消选中该节点或注释之间进行切换。 如果还选中了任何其他节点或注释，那么这会将它们取消选中。
Ctrl+Shift+空格键	焦点位于流中的节点或注释上，或者位于选用板中的节点或注释上时，此快捷键用于在选中与取消选中该节点或注释之间进行切换。 这不会影响任何其他所选节点或注释。
向左/向右方向键	如果焦点位于流工作区，那么此快捷键用于在屏幕上水平移动整个流。如果焦点位于选用板选项卡上，那么此快捷键用于在选项卡之间循环移动。如果焦点位于某个选用板节点上，那么此快捷键用于在选用板的节点之间进行移动。
向上/向下方向键	如果焦点位于流工作区中，那么此快捷键用于在屏幕上垂直移动整个流。如果焦点位于某个选用板节点上，那么此快捷键用于在选用板的节点之间进行移动。如果焦点位于某个子选用板上，那么此快捷键用于在此选用板选项卡的其他子选用板之间进行移动。
Alt+向左/向右方向键	按方向键的方向在流工作区中水平移动所选节点和注释。
Alt+向上/向下方向键	按方向键的方向在流工作区中垂直移动所选节点和注释。
Ctrl+A	选择流中的所有节点。

表 40: 节点和流快捷键 (继续)	
快捷键	函数
Ctrl+Q	焦点位于某个节点上时, 此快捷键用于选择该节点及下游的所有节点, 并用于取消选择上游的所有节点。
Ctrl+W	焦点位于所选节点上时, 此快捷键用于取消选择该节点以及下游的所有所选节点。
Ctrl+Alt+D	复制所选节点。
Ctrl+Alt+L	如果选择了流中的模型块, 那么此快捷键用于打开“插入”对话框, 以便您将已保存的模型从 .nod 文件装入流中。
Ctrl+Alt+R	针对所选节点显示“注释”选项卡, 从而使您能够将该节点重命名。
Ctrl+Alt+U	创建“用户输入”源节点。
Ctrl+Alt+C	开启或关闭节点的缓存。
Ctrl+Alt+F	清空节点的缓存。
制表符	在流工作区中当前流的所有源节点和注释之间循环移动。在节点选用板上的各个节点之间移动。在所选子选用板上, 移至该子选用板中的第一个节点。
Shift+Tab	按逆序执行 Tab 键所执行的操作。
Ctrl+Tab	在焦点位于管理器窗格或工程窗格的情况下, 将焦点移至流工作区。在焦点位于节点选用板上时, 此快捷键用于在节点与其选用板选项卡之间移动焦点。
任何字母键	当焦点位于当前流中的某个节点上时, 此快捷键使焦点循环移至其名称以所按键开头的下一个节点。
F1	打开帮助系统并转到与焦点相关的主题。
F2	对工作区中选择的节点启动连接过程。使用 Tab 键可以移至工作区上的所需节点, 然后按 Shift+空格键完成连接。
F3	删除工作区中所选节点的所有连接。
F6	在管理器窗格、项目窗格和节点选用板之间移动焦点。
F10	打开“文件”菜单。
Shift+F10	打开节点或流的弹出菜单。
删除	从工作区中删除所选节点。
Esc	关闭弹出菜单或对话框。
Ctrl+Alt+X	展开超节点。
Ctrl+Alt+Z	放大超节点。
Ctrl+Alt+Shift+Z	缩小超节点。
Ctrl+E	当焦点位于流工作区中时, 此快捷键用于运行当前流。

IBM SPSS Modeler 中还使用了一些标准快捷键, 例如使用 Ctrl+C 进行复制。有关更多信息, 请参阅主题 [第 17 页的『使用快捷键』](#)。

对话框和表的快捷键

对对话框、表和对话框中的表进行操作时, 一些快捷键和屏幕朗读者键非常有用。特殊键盘快捷键和屏幕朗读者快捷键的完整列表如下所示。

表 41: 对话框和表达式构建器快捷键

快捷键	函数
Alt+4	用于关闭所有打开的对话框或输出窗口。可以从管理器窗格中的“输出”选项卡中检索输出。
Ctrl+End	不论焦点集中于表达式构建器中的哪个控件，此快捷键都可将插入点移动到表达式的末尾处。
Ctrl+1	在表达式构建器中，可将焦点移动到表达式编辑控件。
Ctrl+2	在表达式构建器中，可将焦点移动到函数列表。
Ctrl+3	在表达式构建器中，可将焦点移动到字段列表。

表快捷键

表快捷键可用于节点（例如类型、过滤和合并节点）的输出表和对话框表控件。通常，您将使用 Tab 键在表单元格之间移动，并使用 Ctrl+Tab 来离开表控件。注：有时候，屏幕阅读器可能不会立即开始朗读单元格内容。按一次或两次方向键将重置软件并启动发音。

表 42: 表快捷键

快捷键	函数
Ctrl+W	对于表，阅读所选行的简短描述。例如，“所选的第 2 行的值为性别、标志、m/f 等。”
Ctrl+Alt+W	对于表，阅读所选行的详细描述。例如，“所选的第 2 行的值为字段 = 性别、类型 = 标志、性别 = m/f 等。”
Ctrl+D	对于表，阅读所选区域的简短描述。例如，“所选区域为一行六列。”
Ctrl+Alt+D	对于表，提供所选区域的详细描述。例如，“所选区域为一行六列。所选列为字段、类型和缺失值。所选行为 1。”
Ctrl+T	对于表，提供所选列的简短描述。例如，“字段、类型和缺失值。”
Ctrl+Alt+T	对于表，提供所选列的详细描述。例如，“所选列为字段、类型和缺失值。”
Ctrl+R	对于表，提供表中记录的编号。
Ctrl+Alt+R	对于表，提供表中记录的编号及列名。
Ctrl+I	对于表，阅读具有焦点的单元格的信息或内容。
Ctrl+Alt+I	对于表，阅读具有焦点的单元格的信息（列名及单元格内容）的详细描述。
Ctrl+G	对于表，提供简短的基本选择信息。
Ctrl+Alt+G	对于表，提供详细的基本选择信息。
Ctrl+Q	对于表，提供表单元格之间的快速切换。当您使用方向键在表中移动时，按 Ctrl+Q 将阅读详细描述，例如“性别 = 女性”。再次按 Ctrl+Q 将切换到简短描述（单元格内容）。
F8	对于表，当焦点在表中时，将焦点设置到列标题。
空格键	对于表，当焦点在列标题上时，启用列排序。

注释的快捷键

处理屏幕上的注释时，您可以使用下列快捷键。

快捷键	函数
Alt+C	切换显示/隐藏注释功能。
Alt+M	插入新注释（如果当前显示了注释）；显示注释（如果注释当前处于隐藏状态）。
制表符	在流工作区中当前流的所有源节点和注释之间循环移动。
进入法	注释具有焦点时，指示编辑开始位置。
Alt+Enter 或 Ctrl+Tab	结束编辑并保存编辑更改。
Esc	取消编辑。在编辑期间所作的更改将丢失。
Alt+Shift+向上方向键	将文本区域的高度缩小一个网格单元格（或一个像素），前提是对齐网格功能处于开启（或关闭）状态。
Alt+Shift+向下方向键	将文本区域的高度增大一个网格单元格（或一个像素），前提是对齐网格功能处于开启（或关闭）状态。
Alt+Shift+向左方向键	将文本区域的宽度缩小一个网格单元格（或一个像素），前提是对齐网格功能处于开启（或关闭）状态。
Alt+Shift+向右方向键	将文本区域的宽度增大一个网格单元格（或一个像素），前提是对齐网格功能处于开启（或关闭）状态。

聚类浏览器和模型查看器的快捷键

可以使用快捷键在聚类浏览器和模型查看器窗口中进行浏览。

快捷键	函数
制表符	将焦点移到下一个屏幕控件。
Shift+Tab	将焦点移到上一个屏幕控件。
向下方向键	如果下拉列表具有焦点，那么将打开该列表或者移至该列表中的下一项。 如果菜单具有焦点，那么将移至该菜单中的下一项。 如果缩略图具有焦点，那么将移至集合中的下一项（如果最后一个缩略图具有焦点，那么将移至第一项）。
向上方向键	如果已打开下拉列表，那么此键用于移至该列表中的上一项。 如果菜单具有焦点，那么此键用于移至该菜单中的上一项。 如果缩略图具有焦点，那么此键用于移至集合中的下一项（如果第一个缩略图具有焦点，那么此键用于移至最后一项）。
进入法	关闭已打开的下拉列表，或者在打开的菜单中进行选择。
F6	在窗口的左右窗格之间切换焦点。

快捷键	函数
左右方向键	如果选项卡具有焦点, 那么此键用于移至上一个或下一个选项卡。 如果菜单具有焦点, 那么此键用于移至上一个或下一个菜单。
Alt+letter	选择名称中包含这个带下划线字母的按钮或菜单。
Esc	关闭已打开的菜单或下拉列表。

仅限“聚类查看器”

聚类查看器包含一个“聚类”视图, 该视图包含聚类-特征网格。

要选择该“聚类”视图而不是“模型摘要”视图:

1. 重复按 Tab 键直到选中**视图**按钮。
2. 按向下方向键两次以选中**聚类**。

您可以在此处选择网格中的单个单元格:

3. 重复按 Tab 键, 直到您到达可视化工具栏中的最后一个图标为止。



图 19: 显示“可视化树”图标

4. 再按一次 Tab 键, 然后按空格键, 再按方向键。

现在, 下列键盘快捷键可用:

快捷键	函数
方向键	在网格中的各个单元格之间移动焦点。右侧窗格中的单元格分布显示将在焦点移动时发生变化。
Ctrl+, (逗号)	选择或取消选择网格中单元格具有焦点的整列。 要将某列添加到选择中, 请使用方向键浏览到该列中的某个单元格, 然后再次按 Ctrl+,。
制表符	使焦点移出网格并移至下一个屏幕控件。
Shift+Tab	使焦点移出网格并移回到上一个屏幕控件。
F2	进入编辑方式 (仅限于标签和描述单元格)。
进入法	保存编辑更改并退出编辑方式 (仅限于标签和描述单元格)。
Esc	退出编辑方式而不保存更改 (仅限于标签和描述单元格)。

快捷键示例: 构建流

为了让依赖键盘或屏幕朗读器的用户更清晰地了解流构建过程, 下面提供了一个不使用鼠标构建流的示例。在此示例中, 您将通过下列步骤, 构建一个包含变量文件节点、派生节点和直方图节点的流:

1. 启动 **IBM SPSS Modeler**。第一次启动 IBM SPSS Modeler 时, 焦点位于节点选用板的“收藏夹”选项卡上。
2. **Ctrl + 向下箭头**。将焦点从该选项卡移至选项卡主体。
3. **向右箭头**。将焦点移至“变量文件”节点。

4. **空格键**。选择“变量文件”节点。
5. **Ctrl + Enter**。将“变量文件”节点添加至流工作区。此组合键还会保持选中“变量文件”节点，以便添加的下一个节点将连接到该节点。
6. **Tab**。将焦点移回节点选用板。
7. **按向右箭头 4 次**。移至“派生”节点。
8. **空格键**。选择“派生”节点。
9. **Alt + Enter**。将该“派生”节点添加至工作区，并将所选内容移至该“派生”节点。现在，此节点可以连接到下一个所添加的节点。
10. **Tab**。将焦点移回节点选用板。
11. **按向右箭头 5 次**。将焦点移至选用板中的“直方图”节点。
12. **空格键**。选择“直方图”节点。
13. **进入法**。将该节点添加到流，并将焦点移至流工作区。

继续完成下一个示例，或者如果您希望稍后再尝试下一个示例，请保存流。

快捷键示例：编辑节点

在此示例中，您将使用上一示例中构建的流。该流包含一个“变量文件”节点、一个“导出”节点和一个“直方图”节点。在本指示信息的开头部分，焦点位于流中第三个节点（即，“直方图”节点）上。

1. **按 Ctrl + 向左箭头 2 次**。将焦点移回“变量文件”节点。
2. **进入法**。打开“变量文件”对话框。一直按 Tab 键以转到“文件”字段，然后输入文本文件路径和名称以选择该文件。按 Ctrl+Tab 以浏览到对话框的下半部分，并一直按 Tab 键转到“确定”按钮，然后按 Enter 键关闭对话框。
3. **Ctrl + 向右箭头**。将焦点移至第二个节点（即，“派生”节点）。
4. **进入法**。打开“派生节点”对话框。一直按 Tab 键以选择字段，并指定派生条件。按 Ctrl+Tab 以浏览到“确定”按钮，然后按 Enter 键关闭对话框。
5. **Ctrl + 向右箭头**。将焦点移动到第三个节点（即，“直方图”节点）。
6. **进入法**。打开“直方图节点”对话框。一直按住 Tab 键以选择字段，并指定图形选项。对于下拉列表，请按向下方向键以打开该列表并突出显示某个列表项，然后按 Enter 键以选择该列表项。一直按 Tab 键以浏览到“确定”按钮，然后按 Enter 键关闭对话框。

此时，您可以添加更多节点或运行当前流。构建流时，请记住下列提示：

- 手动连接节点时，使用 F2 创建连接起点，并按 Tab 键移至终点，然后使用 Shift+空格键完成连接。
- 使用 F3 键可断开工作区中所选节点的所有连接。
- 创建流之后，使用 Ctrl+E 可以运行当前流。

提供了一个完整的快捷键列表。有关更多信息，请参阅主题 [第 206 页的『用于浏览主窗口的快捷方式』](#)。

使用屏幕朗读者

市场上有各种各样的屏幕朗读者。IBM SPSS Modeler 配置为支持 JAWS for Windows，后者使用随 IBM SPSS Modeler 一起安装的 Java Access Bridge。如果已安装 JAWS，那么只需在启动 IBM SPSS Modeler 之前启动 JAWS 即可使用此产品。

注：我们建议您至少留出 6GB 的空间，以便将 JAWS 与 SPSS Modeler 一起运行。

由于 IBM SPSS Modeler 具备独特的以图形方式表示数据挖掘过程的特征，因此可通过最佳可视化方式使用图表和图形。但是，您可能需要使用屏幕朗读者根据以文本方式查看的输出和模型来了解并作出决策。

注：对于 64 位客户端机器，某些辅助技术功能无效。这是因为 Java Access Bridge 并非针对 64 位操作设计。

使用 IBM SPSS Modeler 字典文件

随 JAWS 提供了可用的 IBM SPSS Modeler 字典文件 (*Awt.JDF*)。要使用此文件，请完成下列步骤：

1. 浏览至 IBM SPSS Modeler 安装版本的 */accessibility* 子目录，然后复制该字典文件 (*Awt.JDF*)。
2. 将其复制到包含 JAWS 脚本的目录。

如果正在运行其他 Java 应用程序，那么您的计算机可能已包含名为 *Awt.JDF* 的文件。在这种情况下，您可能无法使用此字典文件，而必须手动编辑该字典文件。

将屏幕阅读器用于 HTML 输出

在 IBM SPSS Modeler 中使用屏幕阅读器查看显示为 HTML 的输出时，可能会遇到一些困难。受影响的一些输出类型包括：

- 在回归、Logistic 回归和因子/PCA 节点的“高级”选项卡中看到的输出
- 报告节点输出

在上述每个窗口或对话框中，工具栏上提供了可用于在缺省浏览器中启动输出的工具，这将提供标准的屏幕阅读器支持。然后，您可以使用屏幕阅读器传达输出信息。

交互树窗口中的辅助功能选项

“交互式树”窗口中决策树的标准显示可能会导致屏幕阅读器的问题。要访问易访问版本，请在“交互式树”菜单中单击：

视图 > 可访问窗口

此选项将显示一个类似于标准树映射的视图，但该视图可由 Jaws 正确阅读。您可以使用标准的箭头键向上、向下、向左、向右移动。当您浏览可访问窗口时，“交互式树”窗口中的焦点将相应地移动。使用空格键可以更改选择，使用 Alt+/ 可以扩展当前选择。

使用提示

一些提示可帮助您更方便地访问 IBM SPSS Modeler 环境。以下是在 IBM SPSS Modeler 中工作时的一般提示。

- **退出展开的文本框。** 使用 Ctrl+Tab 可退出展开的文本框。请注意 Ctrl+Tab 也用于退出表控件。
- **使用 Tab 键而不是箭头键。** 选择对话框的选项时，请使用 Tab 键在选项按钮之间进行移动。方向键在该环境中不起作用。
- **下拉列表。** 在对话框的下拉列表中，可以使用 Esc 键或空格键选择某项，然后关闭下拉列表。也可以使用 Esc 键关闭在跳转到另一个控件时未关闭的下拉列表。
- **执行状态。** 对大型数据库运行流时，JAWS 可能会在阅读流状态时出现滞后现象。定时按 Ctrl 键可更新状态报告。
- **使用节点选用板。** 第一次进入节点选用板的选项卡时，JAWS 有时会阅读“组框”，而不是该节点的名称。在这种情况下，可以按 Ctrl+向右方向键，然后按 Ctrl+向左方向键以重置屏幕阅读器并收听节点名。
- **读取菜单。** 有时，在您第一次打开某个菜单时，JAWS 可能不会阅读第一个菜单项。如果您怀疑发生了这种情况，请使用向下方向键，然后使用向上方向键收听菜单中的第一个菜单项。
- **级联菜单。** JAWS 不会阅读级联菜单的第一级菜单。如果您在菜单中移动时听到发音中断，请按向右方向键收听子菜单项。

另外，如果已安装 IBM SPSS Modeler Text Analytics，那么以下提示可使交互式工作台界面更易于使用。

- **进入对话框。** 进入对话框时，您可能需要按 Tab 键以便将焦点置于第一个控件上。
- **退出展开的文本框。** 使用 Ctrl+Tab 可退出展开的文本框并移至下一个控件。请注意 Ctrl+Tab 也用于退出表控件。
- **输入第一个字母以查找树列表中的元素。** 在“类别”窗格、“抽取的结果”窗格或库树中查找某个元素时，您可以在焦点位于此窗格上的情况下键入该元素的首字母。此操作将选中下一个出现的以您所输入字母开头的元素。
- **下拉列表。** 在对话框的下拉列表中，可以使用空格键选择某项，然后关闭列表。

与其他软件的冲突

使用屏幕阅读器（例如，JAWS）测试 IBM SPSS Modeler 时，我们的开发小组发现在贵组织内使用 Systems Management Server(SMS) 可能与 JAWS 的功能发生冲突，而无法阅读基于 Java 的应用程序，例如 IBM SPSS Modeler。禁用 SMS 将改善此状况。请访问 Microsoft 的 Web 站点以获取有关 SMS 的更多信息。

JAWS 和 Java

不同版本的 JAWS 将为基于 Java 的软件应用程序提供不同级别的支持。虽然 IBM SPSS Modeler 可以使用所有 JAWS 的最新版本，但某些版本在应用到基于 Java 的系统时还是会出现一些小问题。请访问 JAWS for Windows 的网站 <http://www.FreedomScientific.com>。

在 IBM SPSS Modeler 中使用图形

信息的可视化显示（例如，直方图、评估图表、多重散点图和散点图）很难使用屏幕阅读器进行说明。但是请注意，可以使用输出窗口中提供的文本摘要查看 Web 图形和分布。

第 18 章 Unicode 支持

IBM SPSS Modeler 中的 Unicode 支持

IBM SPSS Modeler 10.0 IBM SPSS Modeler 和 IBM SPSS Modeler Server 均完全支持 Unicode。这样就可以与其他支持 Unicode 的应用程序（包括多语言数据库）交换数据，同时不会由于针对特定语言环境的编码转换而导致信息丢失。

- IBM SPSS Modeler 将 Unicode 数据存储在内部，并且可以读写数据库中以 Unicode 存储的多语言数据而不会发生数据丢失。
- IBM SPSS Modeler 可以读写用 UTF-8 编码的文本文件。文本文件的导入和导出将缺省使用语言环境编码，但也支持 UTF-8 作为替代编码。此设置可以在文件导入和导出节点中指定，也可以在流属性对话框中更改缺省编码设置。有关更多信息，请参阅主题 第 32 页的『设置流的常规选项』。
- 以语言环境编码保存的 Statistics、SAS 和文本数据文件在导入时将转换为 UTF-8 格式，导出时又转换为原来的格式。写入任何文件时，如果存在语言环境字符集中不存在的 Unicode 字符，则这些字符将被替代并显示警告信息。仅当从支持 Unicode 并且包含来自另一语言环境或来自多个语言环境或字符集的字符的数据源（数据库或 UTF-8 文本文件）导入数据时，才会发生此种情况。
- IBM SPSS Modeler Solution Publisher 图像使用 UTF-8 编码，且真正实现了平台间和语言环境间的可移植。

关于 Unicode

Unicode 标准的目标是提供一种多语言文本编码的统一方法，以便可以轻松地跨地界、语言环境和应用程序共享文本。Unicode 标准现已发展到版本 4.0.1，它定义了一个作为当今世界上所有常用字符集的超级集合的字符集，并为每个字符分配一个唯一名称和码点。这些字符及其代码点与 ISO-10646 定义的通用字符集 (UCS) 的字符及其代码点相同。有关更多信息，请参阅 [Unicode 主页](#)。

注意事项

本信息是为在美国提供的产品和服务编写的。IBM 可能会提供其他语言形式的本资料。但是，您可能需要拥有该语言的产品副本或产品版本，才能对其进行访问。

IBM 可能在其他国家或地区不提供本文档中讨论的产品、服务或功能。有关您所在区域当前可获得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。但是，评估和验证任何非 IBM 产品、程序或服务，则由用户自行负责。

IBM 可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并不意味着授予用户使用这些专利的任何许可。您可以以书面形式将许可查询寄往：

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US

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，或以书面形式将查询寄往：

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

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。某些管辖区域在某些交易中不允许免除明示或暗含的保证。因此本条款可能不适用于您。

本信息中可能包含技术方面不够准确的地方或印刷错误。此处的信息将定期更改；这些更改将编入本资料的新版本中。IBM 可以随时对本出版物中描述的产品和/或程序进行改进和/或更改，而不另行通知。

本信息中对非 IBM Web 站点的任何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 站点中的资料不是本 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 Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US

只要遵守适当的条件和条款，包括某些情形下的一定数量的付费，都可获得这方面的信息。

本文档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 国际程序许可协议或任何同等协议中的条款提供。

所引用的性能数据和客户示例仅作参考用途。实际的性能结果可能会因特定的配置和运营条件而异。

涉及非 IBM 产品的信息是从这些产品的供应商、已出版说明或其他可公开获得的资料中获取。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。

关于 IBM 未来方向或意向的声明都可随时更改或收回，而不另行通知，它们仅仅表示了目标和意愿而已。

本信息包含在日常业务操作中使用的数据和报告的示例。为了尽可能完整地说明这些示例，示例中可能会包括个人、公司、品牌和产品的名称。所有这些名字都是虚构的，若与实际个人或业务企业相似，纯属巧合。

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。IBM 商标的最新列表可从 Web 上的“Copyright and trademark information”处获得，网址为：www.ibm.com/legal/copytrade.shtml。

Adobe、Adobe 徽标、PostScript 以及 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 或其子公司在美国或其他国家或地区的商标或注册商标。

Linux 是 Linus Torvalds 在美国和/或其他国家或地区的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国和/或其他国家或地区的商标。

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 和/或其子公司的商标或注册商标。

产品文档的条款和条件

根据以下条款和条件授予这些出版物的使用许可权。

适用性

这些条款和条件是对 IBM Web 站点的任何使用条款的补充。

个人使用

您可以复制这些出版物供个人非商业性使用，但前提是保留所有专有权声明。未经 IBM 明确同意，您不可以分发、展示或制作这些出版物或其中任何部分的演绎作品。

商业性使用

您仅可在贵公司内部复制、分发和显示这些出版物，但前提是保留所有专有权声明。未经 IBM 明确同意，您不可以制作这些出版物的演绎作品，或者在您的企业外部复制、分发或展示这些出版物或其中的任何部分。

权利

除非本许可权中明确授予，否则不得授予对这些出版物或其中包含的任何信息、数据、软件或其他知识产权的任何许可权、许可证或权利，无论明示的还是暗含的。

只要 IBM 认为这些出版物的使用会损害其利益或者 IBM 判定未正确遵守上述指示信息，IBM 将有权撤销本文授予的许可权。

只有您完全遵循所有适用的法律和法规，包括所有的美国出口法律和法规，您才可以下载、出口或再出口该信息。

IBM 对这些出版物的内容不作任何保证。这些出版物“按现状”提供，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适销性、非侵权和适用于某种特定用途的保证。

索引

Special Characters

"工作簿"选项卡
 移动输出 [88](#)
"合并"节点
 性能 [202](#)
"汇总"节点
 性能 [202](#)
"区分"节点
 性能 [202](#)
"数据审核"节点
 用于数据挖掘 [20](#)
 在探索中使用 [19](#)
@BLANK 函数 [116, 139, 159](#)
@DIFF 函数 [155](#)
@FIELD 函数 [116, 160](#)
@FIELDS_BETWEEN 函数 [116, 124, 160](#)
@FIELDS_MATCHING 函数 [116, 124, 160](#)
@INDEX 函数 [155](#)
@LAST_NON_BLANK 函数 [155, 159](#)
@MAX 函数 [155](#)
@MEAN 函数 [155](#)
@MIN 函数 [155](#)
@MULTI_RESPONSE_SET 函数 [125, 160](#)
@NULL 函数 [116, 139, 159](#)
@OFFSET 函数
 性能注意事项 [203](#)
@PARTITION_FIELD 函数 [160](#)
@PREDICTED 函数 [160](#)
@SDEV 函数 [155](#)
@SINCE 函数 [155](#)
@SUM 函数 [155](#)
@TARGET 函数 [160](#)
@TESTING_PARTITION 函数 [160](#)
@THIS 函数 [155](#)
@TODAY 函数 [151](#)
@TRAINING_PARTITION 函数 [160](#)
@VALIDATION_PARTITION 函数 [160](#)
案例 [19](#)
版本标签, IBM SPSS Collaboration and Deployment Services Repository 对象 [172](#)
保存
 多个对象 [47](#)
 节点 [46](#)
 流 [46](#)
 输出对象 [47](#)
 状态 [47](#)
保存工作簿
 文本格式 [92](#)
 Excel 格式 [92](#)
 HTML [92](#)
 HTML 格式 [92](#)
 PDF 格式 [92](#)
 PowerPoint 格式 [92](#)
 Word 格式 [92](#)
保存输出
 文本格式 [92, 96](#)

保存输出 (继续)
 Excel 格式 [92, 94](#)
 HTML [92, 93](#)
 HTML 格式 [92](#)
 PDF 格式 [92, 95](#)
 PowerPoint 格式 [92, 95](#)
 Web 报告 [93](#)
 Word 格式 [92, 94](#)
保存图表
 元文件 [92](#)
 BMP 文件 [92, 97](#)
 EMF 文件 [92](#)
 EPS 文件 [92, 97](#)
 JPEG 文件 [92, 97](#)
 PICT 文件 [92](#)
 PNG 文件 [97](#)
 PostScript 文件 [97](#)
 TIFF 文件 [97](#)
报告
 保存输出 [47](#)
 设置属性 [186](#)
 生成 [186](#)
 添加到工程 [183](#)
备份流文件
 恢复 [46](#)
背景颜色 [107](#)
比较函数 [141](#)
必备字段 [49](#)
必要字段 [48, 49](#)
边框
 显示隐藏的边框 [109](#)
编码 [32, 215](#)
变量
 对话框中的显示顺序 [111](#)
变量标签
 在对话框中 [111](#)
 在概要窗格中 [112](#)
 在枢轴表中 [112](#)
变量名
 在对话框中 [111](#)
标签
 插入组标签 [101](#)
 删除 [101](#)
 显示 [32](#)
 value [179](#)
 variable [179](#)
标签, IBM SPSS Collaboration and Deployment Services Repository 对象 [172](#)
标题
 添加到查看器 [89](#)
表
 保存输出 [47](#)
 背景颜色 [107](#)
 边距 [107](#)
 单元格属性 [107](#)
 对齐 [107](#)
 控制表分隔符 [110](#)

- 表 (继续)
 - 添加到工程 [183](#)
 - 字体 [107](#)
- 表达式 [133](#)
- 表达式构建器
 - 访问 [126](#)
 - 概述 [125](#)
 - 使用 [126](#)
- 表格外观
 - 创建 [104](#)
 - 应用 [104](#)
- 表图 [111](#)
- 饼图 [71](#)
- 不等运算符 [141](#)
- 部署 [163](#)
- 部署类型 [173](#)
- 部署选项 [173](#)
- 参数
 - 会话 [38](#)
 - 流 [38](#)
 - 模型构建 [174](#)
 - 评分 [174](#)
 - 运行时提示 [38](#)
 - 在 CLEM 表达式中 [129](#)
 - type [38](#)
- 测量系统 [111](#)
- 层
 - 创建 [103](#)
 - 打印 [98, 105, 107](#)
 - 显示 [103](#)
 - 在透视表中 [103](#)
- 插入组标签 [101](#)
- 查看器
 - 保存文档 [99](#)
 - 查找和替换信息 [90](#)
 - 概要 [88](#)
 - 概要窗格 [87](#)
 - 更改概要大小 [89](#)
 - 更改概要级别 [89](#)
 - 更改概要字体 [89](#)
 - 结果窗格 [87](#)
 - 删除输出 [88](#)
 - 输出项之间的空格 [99](#)
 - 搜索和替换信息 [90](#)
 - 显示变量标签 [112](#)
 - 显示变量名 [112](#)
 - 显示数据值 [112](#)
 - 显示选项 [111](#)
 - 显示值标签 [112](#)
 - 隐藏结果 [87](#)
 - 展开概要 [89](#)
 - 折叠概要 [89](#)
- 查找和替换
 - 查看器文档 [90](#)
- 常用对数函数 [142](#)
- 撤销 [14](#)
- 成员函数 [141](#)
- 呈示视图时出错
 - 内存不足 [189](#)
- 垂线图 [76, 77](#)
- 垂直标签文本 [101](#)
- 从 IBM SPSS Collaboration and Deployment Services Repository 中检索对象 [168](#)
- 簇状条形图 [53](#)
- 错误消息 [37](#)
- 打开
 - 节点 [48](#)
 - 流 [48](#)
 - 模型 [48](#)
 - 输出 [48](#)
 - 项目 [184](#)
 - 状态 [48](#)
- 打印
 - 层 [98, 105, 107](#)
 - 打印预览 [98](#)
 - 调整表的大小 [105, 107](#)
 - 控制表分隔符 [110](#)
 - 流 [16, 29](#)
 - 输出项之间的空格 [99](#)
 - 透视表 [98](#)
 - 图表 [98](#)
 - 图表大小 [99](#)
 - 文本输出 [98](#)
 - 页码 [99](#)
 - 页眉和页脚 [98](#)
- 大小
 - 在概要中 [89](#)
- 大于运算符 [141](#)
- 单点登录 [7](#)
- 单点登录, IBM SPSS Collaboration and Deployment Services Repository [163, 164](#)
- 单元格属性 [107](#)
- 导出
 - 流描述 [41](#)
 - PMML [179, 180](#)
- 导出工作簿
 - Excel 格式 [92](#)
 - HTML 格式 [92](#)
 - PDF 格式 [92](#)
 - PowerPoint 格式 [92](#)
 - Word 格式 [92](#)
- 导出输出
 - Excel 格式 [92, 94](#)
 - HTML [93](#)
 - HTML 格式 [92](#)
 - PDF 格式 [92, 95](#)
 - PowerPoint 格式 [92](#)
 - Web 报告 [93](#)
 - Word 格式 [92, 94](#)
- 导出图表 [92, 97](#)
- 导航
 - 键盘快捷键 [206](#)
- 导入
 - PMML [179, 180](#)
- 地理空间坐标
 - 显示格式 [36](#)
 - 选择坐标系 [36](#)
- 地理空间坐标格式 [36](#)
- 地理空间坐标系选择 [36](#)
- 地图 [65](#)
- 地图服务 [67](#)
- 登录 IBM SPSS Modeler Server [7](#)
- 典型应用 [19](#)
- 调整
 - 透视表 [105](#)
- 调整比例
 - 透视表 [107](#)
- 调整大小 [15](#)

- 定制图表 [58](#)
- 逗号 [32](#)
- 端口号
 - IBM SPSS Modeler Server [7, 8](#)
- 堆积条形图 [53](#)
- 堆栈溢出错误 [189](#)
- 对话框
 - 变量显示顺序 [111](#)
 - 显示变量标签 [111](#)
 - 显示变量名 [111](#)
- 对齐
 - output [88, 111](#)
- 对齐方式
 - output [88, 111](#)
- 对象
 - 属性 [186](#)
 - 对象属性, IBM SPSS Collaboration and Deployment Services Repository [171](#)
- 对行或列分组 [101](#)
- 多二分集
 - 在 CLEM 表达式中 [125](#)
- 多个 IBM SPSS Modeler 会话 [10](#)
- 多个打开的数据文件 [111](#)
- 多类别集
 - 在 CLEM 表达式中 [125](#)
- 多系列图 [70](#)
- 多折线图 [65](#)
- 多重图 [69](#)
- 多重响应集
 - 在 CLEM 表达式中 [125, 129](#)
- 多轴图 [59](#)
- 发球方
 - 登录 [7](#)
 - 缺省目录 [190](#)
 - 添加连接 [8](#)
 - 通过 COP 搜索服务器 [9](#)
- 翻转日 [33](#)
- 非运算符 [142](#)
- 分表符 [110](#)
- 分布函数 [144](#)
- 分割表
 - 控制表分隔符 [110](#)
- 分级节点
 - 性能 [202](#)
- 分支, 建模和评分 [42, 174, 175](#)
- 分组符号
 - 数字显示格式 [32](#)
- 分组散点图 [76](#)
- 分组折线图 [65](#)
- 符号函数 [142](#)
- 符合 508 要求 [205](#)
- 辅助功能
 - IBM SPSS Modeler 中的提示 [213](#)
- 辅助功能选项
 - 示例 [211, 212](#)
 - IBM SPSS Modeler 中的功能 [205](#)
- 复制 [14](#)
- 复制和粘贴输出到其他应用程序 [91](#)
- 复制特别 [91](#)
- 概率函数 [144](#)
- 概要
 - 更改级别 [89](#)
 - 在查看器中 [88](#)
 - 展开 [89](#)

- 概要 (继续)
 - 折叠 [89](#)
- 高速缓存文件节点
 - 正在装入 [48](#)
- 工具栏 [14](#)
- 工具提示
 - 为节点添加注解 [46](#)
- 工作簿
 - 导出 [92](#)
- 工作区 [11](#)
- 构建规则节点
 - 正在装入 [48](#)
- 关键字
 - 为节点添加注解 [46](#)
- 关系图 [75](#)
- 管理器 [12](#)
- 规则集
 - 评估 [32](#)
- 滚动
 - 设置选项 [35](#)
- 过程节点
 - 性能 [202](#)
- 函数
 - @BLANK [116](#)
 - @FIELD [125, 160](#)
 - @GLOBAL_MAX [159](#)
 - @GLOBAL_MEAN [159](#)
 - @GLOBAL_MIN [159](#)
 - @GLOBAL_SDEV [159](#)
 - @GLOBAL_SUM [159](#)
 - @PARTITION [160](#)
 - @PREDICTED [125, 160](#)
 - @TARGET [125, 160](#)
 - 处理缺失值 [116](#)
 - 示例 [119](#)
 - 数据库 [126](#)
 - 用户定义的函数 (UDF) [126](#)
 - 在 CLEM 表达式中 [126](#)
- 弧度
 - 度量单位 [34](#)
- 欢迎对话框 [191](#)
- 缓存
 - 保存 [30](#)
 - 启用 [189](#)
 - 设置高速缓存 [29](#)
 - 刷新 [30, 32](#)
- 换行
 - 控制换行文本的列宽 [105](#)
- 回归 [213](#)
- 汇总点图 [76](#)
- 会话参数 [38](#)
- 或运算符 [142](#)
- 货币显示格式 [34](#)
- 机器学习 [19](#)
- 基于树的分析
 - 典型应用 [19](#)
- 集合 [32](#)
- 记录
 - 矩阵表中的 [114](#)
 - 缺少值 [114](#)
- 剪切 [14](#)
- 检查 CLEM 表达式 [129](#)
- 简介
 - IBM SPSS Modeler [7, 189](#)

- 建模
 - 分支 [42](#)
- 建模节点
 - 建模选项板选项卡自定义 [199](#)
 - 性能 [203](#)
- 键盘快捷键 [206, 208, 210](#)
- 将输出居中 [88, 111](#)
- 交互输出 [92](#)
- 交互树窗口
 - 辅助功能选项 [213](#)
- 交换行和列 [101](#)
- 交替行颜色
 - 透视表 [106](#)
- 角度
 - 度量单位 [34](#)
- 脚本编制 [18, 119](#)
- 脚本颜色
 - 设置 [192](#)
- 脚注
 - 标记 [105](#)
 - 重新编号 [108](#)
- 接续文本
 - 透视表的 [107](#)
- 节点
 - 保存 [46](#)
 - 编辑 [29](#)
 - 从选项板中删除 [197](#)
 - 的顺序 [201](#)
 - 复制 [29](#)
 - 简介 [27](#)
 - 将注释添加到 [42](#)
 - 禁用 [29](#)
 - 删除 [28](#)
 - 删除连接 [29](#)
 - 设置选项 [29](#)
 - 数据预览 [31](#)
 - 搜索 [40](#)
 - 锁定 [31](#)
 - 添加 [28, 29](#)
 - 添加到工程 [184](#)
 - 性能 [202, 203](#)
 - 选项板选项卡自定义 [199](#)
 - 预览数据 [31](#)
 - 在 IBM SPSS Collaboration and Deployment Services Repository 中存储 [167](#)
 - 在流中禁用 [29](#)
 - 在流中连接 [28](#)
 - 在选项板上显示 [197](#)
 - 正在装入 [48](#)
 - 执行时间 [37](#)
 - 注解 [42, 46](#)
 - 自定义选项板创建 [197](#)
 - 自定义子选项板创建 [198](#)
- 节点名称 [46](#)
- 节点数
 - 启用 [29](#)
 - 添加到工程 [183](#)
 - 在流中绕过 [28](#)
- 节点选项板选择 [197](#)
- 解锁 IBM SPSS Collaboration and Deployment Services Repository 对象 [170](#)
- 进程协调器 [9](#)
- 禁用节点 [29](#)
- 警告
 - 警告 (继续)
 - 设置选项 [190](#)
 - 矩阵表中的
 - 记录中的 [114](#)
 - 矩阵散点图 [76, 77](#)
 - 句号 [32](#)
 - 距离函数 [144](#)
 - 决策树
 - 辅助功能选项 [213](#)
 - 卡方分布
 - 概率函数 [144](#)
 - 科学表示法
 - 输出中取消 [111](#)
 - 显示格式 [34](#)
 - 可视化编程 [10](#)
 - 客户机 (client)
 - 缺省目录 [190](#)
 - 空白值 [113, 123](#)
 - 空白值处理
 - CLEM 函数 [159](#)
 - 空格
 - 从字符串中删除 [122, 147](#)
 - 空间函数 [144](#)
 - 空值 [123](#)
 - 控制显示行数 [105](#)
 - 块
 - 已定义 [12](#)
 - 快捷键
 - 键盘 [17, 206, 208, 210](#)
 - 一般用法 [50](#)
 - 快速透视表 [112](#)
 - 宽表
 - 粘贴到 Microsoft Word [91](#)
 - 蜡烛图 [56](#)
 - 雷达图 [74](#)
 - 类 [13, 183](#)
 - 类型节点
 - 缺少值 [116](#)
 - 性能 [202](#)
 - 连接
 - 到 IBM SPSS Analytic Server [9](#)
 - 到 IBM SPSS Collaboration and Deployment Services Repository [164](#)
 - 到 IBM SPSS Modeler Server [7-9](#)
 - 服务器集群 [9](#)
 - 连接字符串 [140](#)
 - 列
 - 更改枢轴表中的宽度 [109](#)
 - 在枢轴表中选择 [109](#)
 - 列表 [133, 134](#)
 - 列出流的所有注释 [45](#)
 - 列宽
 - 控制换行文本的宽度 [105](#)
 - 控制缺省宽度 [112](#)
 - 控制最大宽度 [105](#)
 - 透视表 [109](#)
- 流
 - 保存 [46](#)
 - 保存到 IBM Cloud Pak for Data [177](#)
 - 备份文件 [46](#)
 - 部署选项 [173](#)
 - 查看执行时间 [37](#)
 - 地理空间坐标系 [36](#)
 - 构建 [27](#)

流 (继续)

- 禁用节点 [29](#)
- 连接节点 [28](#)
- 绕过节点 [28](#)
- 缩放以查看 [16](#)
- 添加到工程 [183, 184](#)
- 添加节点 [28, 29](#)
- 添加注释 [42](#)
- 选项 [32-36](#)
- 运行 [42](#)
- 在 IBM SPSS Collaboration and Deployment Services Repository 中存储 [166](#)
- 正在装入 [48](#)
- 重命名 [40, 46](#)
- 注解 [42, 46](#)

流参数 [38](#)

流工作区

- 设置 [35](#)

流描述 [40, 41](#)

流名称 [46](#)

流缺省编码 [32](#)

流属性

- Analytic Server [36](#)

逻辑函数 [142](#)

密码

- IBM SPSS Analytic Server [9](#)

- IBM SPSS Modeler Server [7](#)

幂 (指数) 函数 [142](#)

面积图

- 空间函数 [144](#)

命令行

- 启动 IBM SPSS Modeler [7](#)

模板

- 图表 [83](#)

模板字段 [49](#)

模型

- 导出 [192](#)
- 刷新 [175](#)
- 替换 [190](#)
- 添加到工程 [183](#)
- 在 IBM SPSS Collaboration and Deployment Services Repository 中存储 [167](#)

模型块 [42](#)

模型刷新 [173](#)

模型选项板 [167](#)

目录

- 缺省 [190](#)

内存

- 堆栈溢出错误 [189](#)
- 管理 [189](#)

内存不足 [189](#)

排序

- 透视表行 [101](#)

排序节点

- 性能 [202](#)

频数多边形图 [63](#)

平方根函数 [142](#)

平行图 [71](#)

评分

- 分支 [42, 174, 175](#)

评估节点

- 性能 [202](#)

评估图表 [61](#)

凭证

凭证 (继续)

- 用于 IBM SPSS Collaboration and Deployment Services Repository [164](#)

屏幕阅读器

- 示例 [211, 212](#)

谱系图图表 [58](#)

启动对话框 [191](#)

启用节点 [29](#)

气泡图 [56](#)

求反函数 [142](#)

取消字段映射 [48](#)

取整函数 [142](#)

全局函数 [159](#)

全局值

- 在 CLEM 表达式中 [129](#)

缺少值

- 记录中的 [114](#)

- 系统 [114](#)

- CLEM 表达式 [116](#)

缺省

- 工程阶段 [183](#)

热键 [17](#)

热图 [62](#)

人口金字塔 [63](#)

人口金字塔图 [72](#)

日期

- 操作 [154](#)

- 转换 [154](#)

日期/时间值 [123](#)

日期格式 [33, 134, 135](#)

日期时间函数

- datetime_date [151](#)

- datetime_day [151](#)

- datetime_day_name [151](#)

- datetime_day_short_name [151](#)

- datetime_hour [151](#)

- datetime_in_seconds [151](#)

- datetime_minute [151](#)

- datetime_month [151](#)

- datetime_month_name [151](#)

- datetime_month_short_name [151](#)

- datetime_now datetime_second [151](#)

- datetime_time [151](#)

- datetime_timestamp [151](#)

- datetime_weekday [151](#)

- datetime_year [151](#)

日志文件

- 显示生成的 SQL [35](#)

三角函数 [144](#)

三维散点图 [76](#)

三维条形图 [53](#)

三维图 [52](#)

散点图

- 点图 [76](#)

- 分组 [76](#)

- 简单 [76](#)

- 矩阵 [76, 77](#)

- 三维 [76](#)

- 一维 [76](#)

- 重叠 [76](#)

删除输出 [88](#)

神经网络节点

- 大型集合 [32](#)

- 性能 [203](#)

- 生产工具
 - 使用日志文件的命令语法 [111](#)
- 生成的模型选用板 [12](#)
- 时间格式 [33](#), [134](#), [135](#)
- 时间函数
 - [time_before](#) [141](#), [151](#)
 - [time_hours_difference](#) [151](#)
 - [time_in_hours](#) [151](#)
 - [time_in_mins](#) [151](#)
 - [time_in_secs](#) [151](#)
 - [time_mins_difference](#) [151](#)
 - [time_secs_difference](#) [151](#)
- 时间和日期函数 [134](#), [135](#)
- 时间散点图 [79](#)
- 时间字段
 - 转换 [154](#)
- 实数 [133](#)
- 示例
 - 概述 [3](#)
 - 应用程序指南 [2](#)
- 枢轴表中的单元格
 - 宽度 [109](#)
 - 选择 [109](#)
- 输出
 - 保存 [99](#)
 - 查看器 [87](#)
 - 导出 [92](#)
 - 复制 [88](#)
 - 加密 [99](#)
 - 交互式 [92](#)
 - 删除 [88](#)
 - 显示 [87](#)
 - 移动 [88](#)
 - 隐藏 [87](#)
 - 粘贴到其他应用程序中 [91](#)
- 输出对象
 - 在 IBM SPSS Collaboration and Deployment Services Repository 中存储 [167](#)
- 输出节点 [27](#)
- 输出文件
 - 保存 [47](#)
- 属性
 - 报告阶段 [186](#)
 - 表 [105](#)
 - 工程文件夹 [185](#)
 - 数据流 [32](#)
 - 透视表 [105](#)
- 鼠标
 - 在 IBM SPSS Modeler 中使用 [17](#), [28](#)
- 鼠标中键
 - 模拟 [17](#), [28](#)
- 树图表 [81](#)
- 树状图 [82](#)
- 数据
 - 预览 [31](#)
- 数据编辑器
 - 多个打开的数据文件 [111](#)
 - 描述统计选项 [112](#)
 - 图表目录 [51](#)
- 数据库
 - 函数 [126](#)
- 数据库函数
 - 用户定义的函数 (UDF) [126](#)
 - 在 CLEM 表达式中 [126](#)
- 数据类型
 - 在参数中 [38](#)
- 数据流
 - 构建 [27](#)
- 数据挖掘
 - 策略 [20](#)
 - 应用程序示例 [25](#)
- 数据挖掘的应用 [20](#)
- 数据文件
 - 多个打开的数据文件 [111](#)
- 数据映射工具 [48](#), [49](#)
- 数学曲线图 [68](#)
- 数值函数 [142](#)
- 数字 [123](#), [133](#)
- 数字显示格式 [34](#)
- 刷新
 - 源节点 [32](#)
- 刷新模型 [175](#)
- 双 Y 轴图 [59](#)
- 双轴图 [59](#)
- 搜索
 - 流中的节点 [40](#)
- 搜索和替换
 - 查看器文档 [90](#)
- 缩放 [14](#)
- 缩放流以查看 [16](#)
- 锁定 IBM SPSS Collaboration and Deployment Services Repository 对象 [170](#)
- 锁定节点 [31](#)
- 特殊函数 [160](#)
- 特殊字符
 - 从字符串中删除 [122](#)
- 特征选择节点
 - 缺少值 [114](#)
- 提示
 - 辅助功能 [213](#)
 - 一般用法 [50](#)
- 提示, 运行时 [38](#)
- 题注 [107](#)
- 替换模型 [190](#)
- 添加
 - 到工程 [184](#)
 - 添加 IBM SPSS Modeler Server 连接 [8](#), [9](#)
 - 添加组标签 [101](#)
- 填充节点
 - 缺少值 [116](#)
- 条形图 [53](#)
- 停止执行 [14](#)
- 通过 COP 搜索连接 [9](#)
- 通知
 - 设置选项 [190](#)
- 统计文件
 - 编码 [215](#)
- 透视表
 - 背景颜色 [107](#)
 - 边距 [107](#)
 - 边框 [106](#)
 - 编辑 [100](#)
 - 变量标签 [102](#)
 - 操作 [100](#)
 - 层 [103](#)
 - 插入行和列 [102](#)
 - 插入组标签 [101](#)
 - 常规属性 [105](#)

透视表 (继续)

- 撤销更改 [103](#)
 - 从表创建图表 [111](#)
 - 打印层 [98](#)
 - 打印大型表 [110](#)
 - 单元格格式 [106](#)
 - 单元格宽度 [109](#)
 - 单元格属性 [107](#)
 - 导出为 HTML [92](#)
 - 调整比例以适合页面 [105](#), [107](#)
 - 对齐 [107](#)
 - 对行或列分组 [101](#)
 - 对行或列取消分组 [101](#)
 - 对行排序 [101](#)
 - 更改外观 [104](#)
 - 更改显示顺序 [101](#)
 - 交换行和列 [101](#)
 - 交替行颜色 [106](#)
 - 脚注 [107-109](#)
 - 脚注属性 [105](#)
 - 接续文本 [107](#)
 - 控制表分隔符 [110](#)
 - 控制显示行数 [105](#)
 - 快速呈现表格 [112](#)
 - 快速透视表 [112](#)
 - 缺省列宽的调整 [112](#)
 - 删除组标签 [101](#)
 - 使用图标 [101](#)
 - 属性 [105](#)
 - 题注 [107](#)
 - 透视 [101](#)
 - 网格线 [109](#)
 - 文字说明 [108](#)
 - 显示和隐藏单元格 [103](#)
 - 显示隐藏的边框 [109](#)
 - 新表的缺省外观 [112](#)
 - 旋转 [100](#)
 - 旋转标签 [101](#)
 - 选择行和列 [109](#)
 - 移动行和列 [101](#)
 - 遗存表 [111](#)
 - 隐藏 [87](#)
 - 语言 [102](#)
 - 粘贴到其他应用程序中 [91](#)
 - 粘贴为表 [91](#)
 - 值标签 [102](#)
 - 字体 [107](#)
- ## 透视表中的单元格
- 格式 [106](#)
 - 显示 [103](#)
 - 隐藏 [103](#)
- ## 透视表中的颜色
- 边框 [106](#)
- ## 图标
- 设置选项 [16](#), [35](#)
- ## 图表
- 3D [52](#)
 - 保存输出 [47](#)
 - 饼图 [71](#)
 - 垂线图 [76](#)
 - 从透视表创建 [111](#)
 - 导出 [92](#)
 - 点图 [76](#)
 - 定制 [58](#)

图表 (继续)

- 多系列 [70](#)
 - 多重 [69](#)
 - 多轴 [59](#)
 - 关系 [75](#)
 - 汇总点图 [76](#)
 - 蜡烛 [56](#)
 - 雷达 [74](#)
 - 列 [53](#)
 - 模板 [83](#)
 - 频数多边形图 [63](#)
 - 平行 [71](#)
 - 评估 [61](#)
 - 谱系图 [58](#)
 - 气泡 [56](#)
 - 热图 [62](#)
 - 人口金字塔 [63](#), [72](#)
 - 散点图 [76](#)
 - 时间散点图 [79](#)
 - 树 [81](#)
 - 树状图 [82](#)
 - 数学曲线 [68](#)
 - 双 Y 轴 [59](#)
 - 条形 [53](#)
 - 误差条形 [53](#), [60](#)
 - 箱图 [55](#)
 - 旭日 [78](#)
 - 序列数组 [77](#)
 - 仪表板 [83](#)
 - 隐藏 [87](#)
 - 圆填充 [57](#)
 - 折线图 [65](#)
 - 直方图 [63](#)
 - 主题江河 [80](#)
 - 字云 [82](#)
 - map [65](#), [67](#)
 - Q-Q 图 [73](#)
- ## 图表构建器
- 布局 [51](#)
 - 库 [52](#)
 - 术语 [51](#)
- ## 图表目录
- 数据编辑器 [51](#)
- ## 图形
- 保存输出 [47](#)
 - 添加到工程 [183](#)
- ## 网格线
- 透视表 [109](#)
- ## 为节点和流命名
- [46](#)
- ## 位元函数
- [145](#)
- ## 文本编码
- [32](#)
- ## 文本数据文件
- 编码 [215](#)
- ## 文档
- [2](#)
- ## 文件
- 将文本文件添加到查看器 [90](#)
- ## 文件夹, IBM SPSS Collaboration and Deployment Services Repository
- [170](#), [171](#)
- ## 文字说明
- [108](#)
- ## 误差条形图
- [53](#), [60](#)
- ## 系统
- 选项 [189](#)
- ## 显示
- 标题 [104](#)

- 显示 (继续)
 - 脚注 [108](#)
 - 结果 [87](#)
 - 维度标签 [104](#)
 - 文字说明 [108](#)
 - 行或列 [104](#)
- 显示格式
 - 地理空间坐标 [36](#)
 - 分组符号 [34](#)
 - 货币 [34](#)
 - 科学表示法 [34](#)
 - 数字 [34](#)
 - 小数位 [34](#)
- 显示顺序 [101](#)
- 线性回归
 - 导出为 PMML [192](#)
- 相等运算符 [141](#)
- 箱图 [55](#)
- 项目
 - “类”视图 [183](#)
 - 对象属性 [186](#)
 - 构建 [184](#)
 - 关闭 [186](#)
 - 设置缺省文件夹 [183](#)
 - 设置属性 [185](#)
 - 生成报告 [186](#)
 - 添加对象 [184](#)
 - 文件夹属性 [185](#)
 - 新建 [184](#)
 - 在 IBM SPSS Collaboration and Deployment Services Repository 中 [185](#)
 - 在 IBM SPSS Collaboration and Deployment Services Repository 中存储 [167](#)
 - 注解 [185](#)
 - CRISP-DM 视图 [183](#)
- 消息
 - 显示生成的 SQL [35](#)
- 小数符号
 - 数字显示格式 [32](#)
- 小数位
 - 显示格式 [34](#)
- 小于运算符 [141](#)
- 新增功能 [5](#)
- 信息函数 [139](#)
- 行
 - 在枢轴表中选择 [109](#)
- 性能
 - 过程节点的 [202](#)
 - 建模节点的 [203](#)
 - CLEM 表达式 [203](#)
- 旭日图 [78](#)
- 序列函数 [155](#)
- 序列数组图表 [77](#)
- 旋转标签 [101](#)
- 选项
 - 查看器 [111](#)
 - 常规 [111](#)
 - 流属性 [32-37](#)
 - 输出标签 [112](#)
 - 数据编辑器里的描述统计 [112](#)
 - 透视表外观 [112](#)
 - 显示 [191](#)
 - 用户 [190](#)
 - 语法 [192](#)
- 选项 (继续)
 - IBM SPSS Modeler 的 [189](#)
 - PMML [192](#)
- 选项板选项卡自定义 [199](#)
- 选用板
 - 自定义 [197](#)
- 选择方法
 - 在透视表中选择行和列 [109](#)
- 颜色
 - 设置 [191](#)
- 页编号 [99](#)
- 页脚 [98](#)
- 页眉 [98](#)
- 页面设置
 - 图表大小 [99](#)
 - 页眉和页脚 [98](#)
- 仪表板
 - 图表 [83](#)
- 移动行和列 [101](#)
- 移去组标签 [101](#)
- 遗存表 [111](#)
- 因子 [213](#)
- 隐藏
 - 标题 [104](#)
 - 过程结果 [87](#)
 - 脚注 [108](#)
 - 维度标签 [104](#)
 - 文字说明 [108](#)
 - 行和列 [104](#)
- 应用 [19](#)
- 应用程序示例 [2](#)
- 映射数据 [49](#)
- 映射字段 [48](#)
- 用户标识
 - IBM SPSS Modeler Server [7](#)
- 用户定义的函数 (UDF) [126](#)
- 用户选项 [190](#)
- 优先级 [136](#)
- 与运算符 [142](#)
- 语言
 - 更改输出语言 [102](#)
 - 选项 [189](#)
- 预览
 - 节点数据 [31](#)
- 域名 (Windows)
 - IBM SPSS Modeler Server [7](#)
- 元文件
 - 导出图表 [92](#)
- 圆填充图 [57](#)
- 源节点
 - 数据映射 [49](#)
 - 刷新 [32](#)
- 约定 [138](#)
- 运算符
 - 连接字符串 [140](#)
 - 在 CLEM 表达式中 [126](#)
- 运算符优先级 [136](#)
- 运行流 [42](#)
- 在 CLEM 表达式中的反斜杠字符 [134](#)
- 在 IBM SPSS Collaboration and Deployment Services Repository 中存储对象 [165](#)
- 在 IBM SPSS Collaboration and Deployment Services Repository 中搜索对象 [169](#)
- 噪声数据 [20](#)

- 粘贴 [14](#)
- 粘贴输出到其他应用程序中 [91](#)
- 折线图
 - 垂线图 [76, 77](#)
 - 时间散点图 [79](#)
- 整数 [133](#)
- 正态分布
 - 概率函数 [144](#)
- 正在装入
 - 节点 [48](#)
 - 状态 [48](#)
- 知识发现 [19](#)
- 执行时间, 查看 [37](#)
- 直方图 [63](#)
- 值
 - 添加到 CLEM 表达式中 [129](#)
 - 通过数据审核查看 [129](#)
- 值标签
 - 在概要窗格中 [112](#)
 - 在枢轴表中 [112](#)
- 指数函数 [142](#)
- 质量节点
 - 缺少值 [114](#)
- 终端节点 [27](#)
- 重叠散点图 [76](#)
- 重命名
 - 节点 [46](#)
 - 流 [40](#)
- 重新排序行和列 [101](#)
- 主窗口 [11](#)
- 主机名
 - IBM SPSS Modeler Server [7, 8](#)
- 主题江河图 [80](#)
- 注解
 - 节点 [42, 46](#)
 - 流 [42, 46](#)
- 注释
 - 的注释 [42](#)
 - 键盘快捷键 [210](#)
 - 列出流中的所有注释 [45](#)
 - 文件夹 [185](#)
 - 转换为注释 [45](#)
 - project [185](#)
- 柱状图 [53](#)
- 转换函数 [140](#)
- 状态
 - 保存 [47](#)
 - 正在装入 [48](#)
- 字典文件 [212](#)
- 字段
 - 查看值 [129](#)
 - 在 CLEM 表达式中 [129](#)
- 字符 [133, 134](#)
- 字符串
 - 匹配 [122](#)
 - 替换 [122](#)
 - 在 CLEM 表达式中操作 [122](#)
- 字符串函数 [147](#)
- 字体
 - 在概要窗格中 [89](#)
- 字云图 [82](#)
- 自定义选项板创建
 - 子选项板创建 [198](#)
- 自动化 [119](#)

- 自然对数函数 [142](#)
- 子选项板
 - 创建 [198](#)
 - 从选项板选项卡中删除 [198](#)
 - 在选项板选项卡上显示 [198](#)
- 租户
 - IBM SPSS Analytic Server [9](#)
- 组标签 [101](#)
- 最小化 [15](#)

A

- abs 函数 [142](#)
- allbutfirst 函数 [147](#)
- allbutlast 函数 [147](#)
- alphabefore 函数 [147](#)
- arccos 函数 [144](#)
- arccosh 函数 [144](#)
- arcsin 函数 [144](#)
- arcsinh 函数 [144](#)
- arctan 函数 [144](#)
- arctan2 函数 [144](#)
- arctanh 函数 [144](#)
- area 函数 [144](#)

B

- BMP 文件
 - 导出图表 [92, 97](#)

C

- cdf_chisq 函数 [144](#)
- cdf_f 函数 [144](#)
- cdf_normal 函数 [144](#)
- cdf_t 函数 [144](#)
- Champion Challenger 分析 [163, 173](#)
- CLEM
 - 表达式 [121, 133](#)
 - 构建表达式 [126](#)
 - 函数 [126](#)
 - 检查表达式 [129](#)
 - 简介 [18, 119](#)
 - 示例 [119](#)
 - 数据类型 [133, 134](#)
 - 语言 [133](#)
- CLEM 表达式
 - 性能 [203](#)
- CLEM 函数
 - 比较 [141](#)
 - 概率 [144](#)
 - 可用列表 [138](#)
 - 空白值和空置 [159](#)
 - 空间 [144](#)
 - 逻辑 [142](#)
 - 全局 [159](#)
 - 缺少值 [116](#)
 - 日期时间 [151](#)
 - 三角法 [144](#)
 - 数值 [142](#)
 - 随机 [146](#)
 - 特殊函数 [160](#)
 - 信息 [139](#)

CLEM 函数 (继续)
 序列 [155](#)
 逐位 [145](#)
 转换 [140](#)
 字符串 [147](#)
close_to
 空间函数 [144](#)
close_to 函数 [144](#)
Cognos 活动报告 [93](#)
conditions [121](#)
COP [9](#)
cos 函数 [144](#)
cosh 函数 [144](#)
count_equal 函数 [124, 141](#)
count_greater_than 函数 [124, 141](#)
count_less_than 函数 [124, 141](#)
count_non_nulls 函数 [141](#)
count_not_equal 函数 [124, 141](#)
count_nulls 函数 [116, 124, 141](#)
count_substring 函数 [147](#)
CRISP-DM
 工程视图 [183](#)
CRISP-DM 过程模型 [20, 21](#)
crosses
 空间函数 [144](#)
crosses 函数 [144](#)

D

date 函数
 @TODAY 函数 [151](#)
 date_before [141, 151](#)
 date_days_difference [151](#)
 date_in_days [151](#)
 date_in_months [151](#)
 date_in_weeks [151](#)
 date_in_years [151](#)
 date_months_difference [151](#)
 date_weeks_difference [151](#)
 date_years_difference [151](#)
date_before 函数 [141](#)
datetime_date 函数 [140](#)
DIFF 函数 [155](#)
distance
 空间函数 [144](#)
div 函数 [142](#)
DTD [179](#)

E

endstring 函数 [147](#)
EPS 文件
 导出图表 [92, 97](#)
Excel 格式
 导出工作簿 [92](#)
 导出输出 [92, 94](#)

F

f 分布
 概率函数 [144](#)
first_index 函数 [125, 141](#)
first_non_null 函数 [125, 141](#)

first_non_null_index 函数 [125, 141](#)
fracof 函数 [142](#)

H

hasendstring 函数 [147](#)
hasmidstring 函数 [147](#)
hasstartstring 函数 [147](#)
hassubstring 函数 [147](#)
HTML
 导出工作簿 [92](#)
 导出输出 [92, 93](#)
HTML 输出
 屏幕阅读器 [213](#)

I

IBM SPSS Analytic Server
 多个连接 [9](#)
 连接 [9](#)
IBM SPSS Collaboration and Deployment Services [163](#)
IBM SPSS Collaboration and Deployment Services
 Repository
 存储对象 [165](#)
 单点登录 [164](#)
 对象属性 [171](#)
 检索对象 [168](#)
 将工程转移至 [185](#)
 连接到 [164](#)
 浏览 [164](#)
 凭证 [164](#)
 删除对象和版本 [170](#)
 搜索范围 [169](#)
 锁定和解锁对象 [170](#)
 文件夹 [170, 171](#)
IBM SPSS Modeler
 从命令行运行 [7](#)
 辅助功能 [205](#)
 概述 [7, 189](#)
 提示和快捷键 [50](#)
 文档 [2](#)
 新手入门 [7](#)
 选项 [189](#)
IBM SPSS Modeler Advantage [163](#)
IBM SPSS Modeler Server
 端口号 [7, 8](#)
 密码 [7](#)
 用户标识 [7](#)
 域名 (Windows) [7](#)
 主机名 [7, 8](#)
if、then、else 函数 [142](#)
INDEX 函数 [155](#)
integer_bitcount 函数 [145](#)
integer_leastbit 函数 [145](#)
integer_length 函数 [145](#)
intof 函数 [142](#)
is_date 函数 [139](#)
is_datetime 函数 [139](#)
is_integer 函数 [139](#)
is_number 函数 [139](#)
is_real 函数 [139](#)
is_string 函数 [139](#)
is_time 函数 [139](#)

is_timestamp 函数 [139](#)
isalphacode 函数 [147](#)
isendstring 函数 [147](#)
islowercode 函数 [147](#)
ismidstring 函数 [147](#)
isnumbercode 函数 [147](#)
isstartstring 函数 [147](#)
issubstring 函数 [147](#)
issubstring_count 函数 [147](#)
issubstring_lim 函数 [147](#)
isuppercode 函数 [147](#)

J

Java [214](#)
JAWS [205](#), [212](#), [214](#)
JPEG 文件
 导出图表 [92](#), [97](#)

K

K-Means 节点
 大型集合 [32](#)
 性能 [203](#)
Kohonen 节点
 大型集合 [32](#)
 性能 [203](#)

L

last_index 函数 [125](#), [141](#)
LAST_NON_BLANK 函数 [155](#)
last_non_null 函数 [125](#), [141](#)
last_non_null_index 函数 [125](#), [141](#)
length 函数 [147](#)
LOCALE
 选项 [189](#)
locchar 函数 [147](#)
locchar_back 函数 [147](#)
Logistic 回归
 导出为 PMML [192](#)
lowertoupper 函数 [147](#)

M

matches 函数 [147](#)
max 函数 [141](#)
MAX 函数 [155](#)
max_index 函数 [125](#), [141](#)
max_n 函数 [124](#), [141](#)
MEAN 函数 [155](#)
mean_n 函数 [124](#), [142](#)
min 函数 [141](#)
MIN 函数 [155](#)
min_index 函数 [125](#), [141](#)
min_n 函数 [124](#), [141](#)
mod 函数 [142](#)

N

num_points
 空间函数 [144](#)
num_points 函数 [144](#)

O

OFFSET 函数 [155](#)
oneof 函数 [146](#)
output
 对齐 [88](#), [111](#)
 更改输出语言 [102](#)
 集中 [88](#), [111](#)
overlap
 空间函数 [144](#)
overlap 函数 [144](#)

P

PDF
 导出工作簿 [92](#)
 导出输出 [92](#), [95](#)
pi 函数 [144](#)
PMML
 导出模型 [179](#), [180](#)
 导出选项 [192](#)
 导入模型 [179](#), [180](#)
PMML 模型
 线性回归 [192](#)
 Logistic 回归 [192](#)
PNG 文件
 导出图表 [92](#), [97](#)
PostScript 文件 (压缩)
 导出图表 [92](#), [97](#)
PowerPoint
 将输出导出为 PowerPoint [95](#)
PowerPoint 格式
 导出工作簿 [92](#)
 导出输出 [92](#)
PowerPoint 文件 [183](#)

Q

Q-Q 图 [73](#)

R

random 函数 [146](#)
random0 函数 [146](#)
rem 函数 [142](#)
replace 函数 [147](#)
replicate 函数 [147](#)

S

SAS 文件
 编码 [215](#)
SDEV 函数 [155](#)
sdev_n 函数 [124](#), [142](#)
sin 函数 [144](#)
SINCE 函数 [155](#)
sinh 函数 [144](#)
skipchar 函数 [147](#)
skipchar_back 函数 [147](#)
SoundEx 函数 [151](#)
soundex_difference 函数 [151](#)
SPLOM [76](#)
SQL 生成

SQL 生成 (继续)
 [日志记录 35](#)
 [预览 35](#)
[startstring 函数 147](#)
[stripchar 函数 147](#)
[strmember 函数 147](#)
[subscrs 函数 147](#)
[substring 函数 147](#)
[substring_between 函数 147](#)
[SUM 函数 155](#)
[sum_n 函数 124, 142](#)

T

t 分布
 [概率函数 144](#)
[tan 函数 144](#)
[tanh 函数 144](#)
[temp 目录 10](#)
[testbit 函数 145](#)
text
 [将工作簿导出为文本 92](#)
 [将输出导出为文本 92, 96](#)
 [将文本文件添加到查看器 90](#)
 [添加到查看器 89](#)
[THIS 函数 155](#)
TIFF 文件
 [导出图表 92, 97](#)
[time_before 函数 141](#)
[to_date 函数 140, 151](#)
[to_dateline 函数 151](#)
[to_datetime 函数 140](#)
[to_integer 函数 140](#)
[to_number 函数 140](#)
[to_real 函数 140](#)
[to_string 函数 140](#)
[to_time 函数 140, 151](#)
[to_timestamp 函数 140, 151](#)
[trim 函数 147](#)
[trim_start 函数 147](#)
[trimend 函数 147](#)

U

[undef 函数 159](#)
[Unicode 支持 215](#)
[unicode_char 函数 147](#)
[unicode_value 函数 147](#)
[uppertolower 函数 147](#)
URL
 [IBM SPSS Analytic Server 9](#)
[UTF-8 编码 32, 215](#)

V

[value_at 函数 125, 141](#)

W

Web 报告
 [导出输出 93](#)
within
 [空间函数 144](#)

[within 函数 144](#)
Word 格式
 [导出工作簿 92](#)
 [导出输出 92, 94](#)
 [宽表 92](#)

