

*IBM SPSS Modeler 扩展*

**IBM**

## 注

在使用本资料及其支持的产品之前，请阅读第 49 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

<b>第 1 章 支持的语言.....</b>	<b>1</b>
R.....	1
Python for Spark.....	1
使用 Python for Spark 编写脚本.....	1
切换到其他 Python 环境.....	11
本机 Python API.....	11
“扩展”节点.....	13
“扩展导出”节点.....	13
“扩展输出”节点.....	14
“扩展模型”节点.....	17
扩展模型块.....	18
“扩展变换”节点.....	20
“扩展导入”节点.....	20
<b>第 2 章 扩展(X).....</b>	<b>23</b>
扩展中心.....	23
“浏览”选项卡.....	23
“已安装”选项卡.....	24
设置.....	25
扩展详细信息.....	25
安装本地扩展束.....	25
扩展的安装位置.....	25
必需的 R 软件包.....	26
创建并管理定制节点.....	26
定制对话框构建器布局.....	27
构建定制节点对话框.....	27
对话框属性.....	28
在对话框画布上布置控件.....	28
构建脚本模板.....	28
预览定制节点对话框.....	29
控件类型.....	30
扩展属性.....	42
管理定制节点对话框.....	44
创建定制节点对话框的本地化版本.....	46
使用 Python for Spark 导入和导出数据.....	47
使用 R 导入和导出数据.....	48
<b>注意事项.....</b>	<b>49</b>
商标.....	50
产品文档的条款和条件.....	50
<b>索引.....</b>	<b>51</b>



---

# 第 1 章 支持的语言

IBM SPSS Modeler 支持 R 和 Apache Spark（通过 Python）。请参阅以下部分以获取更多信息。

## R

---

IBM SPSS Modeler 支持 R。

### 允许的 R 语法

- 在各种“扩展”节点的语法选项卡上的语法字段中，仅允许使用 R 能够识别的语句和函数。
- 对于“扩展变换”节点和扩展模型块，通过 R 脚本（批处理）传递数据。因此，用于模型评分和过程节点的 R 脚本不应包含跨数据中行或组合行的操作，例如排序或聚集。实施此限制的目的是确保能够在 Hadoop 环境中以及数据库内挖掘期间分割数据。“扩展输出”和“扩展模型构建”节点没有此限制。
- 在“扩展变换”节点和扩展模型块中增加非批量数据传输方式，这意味着您可以在 SPSS Modeler Server 中跨越或组合数据行。
- 所有 R 节点均视为独立的全局 R 环境。因此，在两个单独的 R 节点中使用 `library` 函数需要在两个 R 脚本中装入 R 库。
- 要显示 R 脚本中定义的 R 对象的值，必须调用 `printing` 函数。例如，要显示称为 `data` 的 R 对象的值，请在 R 脚本中包含以下行：

```
print(data)
```

- 由于 IBM SPSS Modeler 将 `setwd` 函数用于控制 R 脚本输出文件的路径，因此您无法在 R 脚本中包含对该函数的调用。
- 如果在 R 脚本中使用定义用于 CLEM 表达式和脚本编制的流参数，那么无法识别这些流参数。
- IBM SPSS Modeler 不支持 R 中的交互式图
- 在流（例如，`modelerData` 和 `modelerDataModel` 数据帧）中使用扩展节点时，会自动填充某些 R 对象

## Python for Spark

---

IBM SPSS Modeler 支持 Apache Spark 的 Python 脚本。

注：

- Python 节点取决于 Spark 环境。
- Python 脚本必须使用 Spark API，因为将以 Spark DataFrame 形式表示数据。
- V17.1 中创建的旧节点仍将仅针对 IBM SPSS Analytic Server 运行（数据源自 IBM SPSS Analytic Server 源节点并且尚未抽取到 IBM SPSS Modeler 服务器）。V18.0 或更高版本中创建的新 Python 和“定制对话框构建器”可针对 IBM SPSS Modeler 服务器运行。
- 安装 Python 时，请确保所有用户都有权访问 Python 安装。
- 如果要使用“机器学习库 (MLlib)”，必须安装包含 NumPy 的 Python 版本。然后，必须配置 IBM SPSS Modeler 服务器（或 IBM SPSS Modeler 客户机中的本地服务器）以使用 Python 安装。有关详细信息，请参阅第 1 页的『使用 Python for Spark 编写脚本』。

### 使用 Python for Spark 编写脚本

IBM SPSS Modeler 可以使用 Apache Spark 框架来运行 Python 脚本，以便处理数据。此文档针对提供的接口提供 Python API 描述。

IBM SPSS Modeler 安装包含 Spark 分发版（例如，IBM SPSS Modeler 18.5 包含 Spark 3.4.0）。

## 先决条件

- 如果计划对 IBM SPSS Analytic Server 运行 Python/Spark 脚本，那么必须具有与 Analytic Server 的连接，并且 Analytic Server 必须有权访问兼容的 Apache Spark 安装。请参阅 IBM SPSS Analytic Server 文档以获取有关使用 Apache Spark 作为执行引擎的详细信息。
- 如果您计划针对 IBM SPSS Modeler Server (或 IBM SPSS Modeler Client 随附的本地服务器，需要 Windows 64 或 Mac64) 运行 Python/Spark 脚本，那么不再需要安装 Python 并编辑 `options.cfg` 以使用 Python 安装。启动 V18.1，IBM SPSS Modeler 现在将包含 Python 分发版。但是，如果需要缺省 IBM SPSS Modeler Python 分发版未包含的特定模块，请转至 `<Modeler_installation_directory>/python` 并安装额外的软件包。

即使现在随 IBM SPSS Modeler 一起提供了 Python 分发版，您仍可以通过将以下选项添加到 `options.cfg`，来指向您自己的 Python 安装（如需要）：

```
# Set to the full path to the python executable
(including the executable name) to enable use of PySpark.
eas_pyspark_python_path, ""
```

Windows 示例：

```
eas_pyspark_python_path, "C:\\Your_Python_Install\\python.exe"
```

Linux 示例：

```
eas_pyspark_python_path, "/Your_Python_Install/bin/python"
```

**注：**如果您指向自己的 Python 安装，那么它必须是 V 3.8.x。IBM SPSS Modeler 已使用 Anaconda for Python 3.8 和 Python 3.8.6 进行测试。

- 如果您计划确定 IBM SPSS Modeler 所使用的 PySpark 版本，那么可以运行以下脚本。

```
import pkg_resources
print("pandas version")
print(pkg_resources.get_distribution("pandas").version)
print("pyspark version")
print(pkg_resources.get_distribution("pyspark").version)
```

根据您的操作系统，在运行脚本以确定 PySpark 版本之前，请在相应路径中运行以下命令。

### 对于 Windows

路径: `<Modeler-InstallationDirectory>\18.5\spark\python`

命令：

```
"<Modeler-InstallationDirectory>\18.5\python\python.exe" setup.py sdist
```

### 对于 Mac

路径: `<Modeler-InstallationDirectory>/18.5/IBM SPSS Modeler.app/Contents/spark/python`

命令：

```
"<Modeler-InstallationDirectory>/18.5/IBM SPSS Modeler.app/Contents/python/bin/python3"
setup.py sdist
```

### 对于 Linux

路径: `<Modeler-InstallationDirectory>/18.5/spark/python`

命令：

```
<Modeler-InstallationDirectory>/18.5/python/bin/python3 setup.py sdist
```

**注：**`<Modeler-InstallationDirectory>` 对应于本地服务器的 IBM SPSS Modeler Thick Client 安装，而它对应于远程服务器的 IBM SPSS Modeler Server 安装。

## IBM SPSS Analytic Server 上下文对象

Analytic Server 上下文对象定义 Python/Spark 脚本的执行上下文。针对 IBM SPSS Modeler 服务器运行时，上下文对象用于 IBM SPSS Modeler 服务器安装中包含的 Analytic Server 的嵌入版本。要获取上下文对象，脚本必须包含以下命令：

```
import spss.pyspark.runtime
asContext = spss.pyspark.runtime.getContext()
```

从 Analytic Server 上下文，您可以获取 Spark 上下文和 SQL 上下文：

```
sparkContext = asc.getSparkContext()
sqlContext = asc.getSparkSQLContext()
```

请参阅 Apache Spark 文档以获取有关 Spark 上下文和 SQL 上下文的信息。

## 访问数据

在 Python/Spark 脚本和 Spark SQL DataFrame 格式的执行上下文之间传输数据。使用数据的脚本（也就是，除源节点之外的任何节点）必须从上下文检索数据帧：

```
inputData = asContext.getSparkInputData()
```

生成数据的脚本（也就是，终端节点外的任何节点）必须将数据帧返回到上下文：

```
asContext.setSparkOutputData(outputData)
```

您可以使用 SQL 上下文以从需要的 RDD 创建输出数据：

```
outputData = sqlContext.createDataFrame(rdd)
```

## 定义数据模型

生成数据的节点还必须定义数据模型，此数据模型描述节点的字段可视下游。在 Spark SQL 术语中，数据模型即模式。

Python/Spark 脚本定义 `pyspark.sql.types.StructType` 对象格式的数据数据模型。`StructType` 描述输出数据帧中的行，并且是从 `StructField` 对象构造的。每个 `StructField` 都描述输出数据模型中的一个字段。

您可以使用输入数据帧的 `:schema` 属性获取输入数据的数据模型：

```
inputSchema = inputData.schema
```

可将未经更改传递的字段从输入数据模型复制到输出数据模型。可使用 `StructField` 构造函数在输出数据模型中新建或修改字段：

```
field = StructField(name, dataType, nullable=True, metadata=None)
```

请参阅 Spark 文档以获取有关构造函数的信息。

请至少提供字段名称及其数据类型。您可以指定元数据以提供资源的测量、角色和描述（请参阅第 7 页的『数据元数据』）。

## DataModelOnly 方式

在执行节点前，IBM SPSS Modeler 需要指定节点的输出数据模型以支持下游编辑。要获取 Python/Spark 节点的输出数据模型，IBM SPSS Modeler 将在没有可用数据的特殊“仅数据模型”方式下运行脚本。脚本可在 Analytic Server 上下文对象上使用 `isComputeDataModelOnly` 方法来标识此方式。

变换节点的脚本可采用以下一般模式：

```

if asContext.isComputeDataModelOnly():
    inputSchema = asContext.getSparkInputSchema()
    outputSchema = ... # construct the output data model
    asContext.setSparkOutputSchema(outputSchema)
else:
    inputData = asContext.getSparkInputData()
    outputData = ... # construct the output data frame
    asContext.setSparkOutputData(outputData)

```

## 构建模型

构建模型的节点必须将某些内容返回给执行上下文，此内容充分描述模型，从而使应用模型的节点可在以后精确重新创建它。

按照键/值对定义模型内容，其中仅构建和评分节点了解键和值的含义，而 **Modeler** 不会以任何方式进行解释。(可选) 节点可以将 MIME 类型分配给某个值，以便 **Modeler** 可以向模型块中的用户显示具有已知类型的那些值。

此上下文中的值可以是 PMML，HTML 和图像等。要将值添加到模型内容（在构建脚本中）：

```
asContext.setModelContentFromString(key, value, mimeType=None)
```

要从模型内容检索值（在评分脚本中）：

```
value = asContext.getModelContentToString(key)
```

作为快捷方式，其中模型或模型的一部分存储到文件系统中的文件或文件夹，您可以通过一个调用绑定存储到此位置的所有内容（在构建脚本中）：

```
asContext.setModelContentFromPath(key, path)
```

请注意，在此情况下，无选项指定 MIME 类型，因为捆绑软件可能包含各种内容类型。

如果在构建模型时需要临时位置来存储内容，那么可以从上下文中获取相应的位置：

```
path = asContext.createTemporaryFolder()
```

要将现有内容检索到文件系统中的临时位置（在评分脚本中）：

```
path = asContext.getModelContentToPath(key)
```

## 错误处理

要引发错误，请从脚本抛出异常并向 IBM SPSS Modeler 用户显示。模块 `spss.pyspark.exceptions` 中预定义了一些异常。例如：

```

from spss.pyspark.exceptions import ASContextException
if ... some error condition ...:
    raise ASContextException("message to display to user")

```

## Analytic Server 上下文

“上下文”针对 Analytic Server 上下文接口提供支持以与 IBM SPSS Analytic Server 交互。

### AnalyticServerContext 对象

`AnalyticServerContext` 对象设置上下文环境，该环境提供多个用于与 IBM SPSS Analytic Server 交互的接口。想要构造此上下文实例的应用程序必须使用 `spss.pyspark.runtime.getContext()` 接口执行此操作，而不是直接实施接口。

返回 Pyspark Python SparkContext 实例：

```
cxt.getContext() : SparkContext
```



返回 Pyspark Python SQLContext 实例:

```
cxt.getSparkSQLContext() : SQLContext
```

返回 True 以描述是否只执行来计算输出数据模型。否则, 返回 False:

```
cxt.isComputeDataModelOnly() : Boolean
```

如果脚本在 Spark 环境中运行, 那么返回 True。当前, 总是返回 True:

```
cxt.isSparkExecution() : Boolean
```

从上游临时文件装入输入数据并生成 pyspark.sql.DataFrame 实例:

```
cxt.getSparkInputData() : DataFrame
```

返回从输入数据模型生成的 pyspark.sql.StructType 实例。如果输入数据模型不存在, 那么返回 None:

```
cxt.getSparkInputSchema() : StructType
```

将输出数据帧序列化为 Analytic Server 上下文并返回上下文:

```
cxt.setSparkOutputData( outDF ) : AnalyticServerContext
```

参数:

- outDF (DataFrame) : 输出数据帧值

异常:

- DataOutputNotSupported : 如果在函数 pyspark:buildmodel 中调用此接口
- ASContextException : 如果输出数据帧为 None
- InconsistentOutputDataModel : 对于两个对象公共的字段名称和存储类型信息不一致

将 outSchema StructType 实例转换为数据模型, 将其序列化为 Analytic Server 上下文, 并返回上下文:

```
cxt.setSparkOutputSchema(outSchema) : AnalyticServerContext
```

参数:

- outSchema(StructType) : StructType 对象的输出

异常:

- ASContextException : 如果输出模式实例为 None
- InconsistentOutputDataModel : 对于两个对象公共的字段名称和存储类型信息不一致

将模型构建输出的位置存储到 Analytic Server 上下文并返回上下文:

```
cxt.setModelContentFromPath(key, path, mimetype=None) : AnalyticServerContext
```

路径可以是在将目录下的所有项打包为模型内容时应使用 cxt.createTemporaryFolder() API 生成的目录路径。

参数:

- key (string) : 键字符串值
- path (string) : 模型构建输出字符串路径的位置
- mimetype (string, optional) : 内容的 MIME 类型

异常:

- ModelOutputNotSupported : 当不是从 pyspark:buildmodel 函数调用此 API 时
- KeyError : 如果密钥属性为 None 或者字符串为空

将模型构建内容、元数据或其他属性存储到 Analytic Server 上下文并返回上下文:

```
cxt.setModelContentFromString(key, value, mimetype=None) : AnalyticServerContext
```

参数:

- key (string) : 键字符串值
- value (string) : 模型元数据字符串值
- mimetype (string, optional) : 内容的 MIME 类型

异常:

- ModelOutputNotSupported : 当不是从 pyspark:buildmodel 函数调用此 API 时
- KeyError : 如果密钥属性为 None 或者字符串为空

返回 Analytic Server 管理的临时文件夹位置; 这可用于存储模型内容:

```
cxt.createTemporaryFolder() : string
```

异常:

- ModelOutputNotSupported : 当不是从 pyspark:buildmodel 函数调用此 API 时

返回匹配输入键的模型的位置:

```
cxt.getModelContentToPath(key) : string
```

参数:

- key (string) : 键字符串值

异常:

- ModelInputNotSupported : 当不是从 pyspark:applymodel 函数调用此 API 时
- KeyError : 如果密钥属性为 None 或者字符串为空
- IncompatibleModelContentType : 如果模型内容类型不是容器

返回匹配输入键的模型内容、模型的元数据或其他模型属性:

```
cxt.getModelContentToString(key) : string
```

参数:

- key (string) : 键字符串值

异常:

- ModelInputNotSupported : 当不是从 pyspark:applymodel 函数调用此 API 时
- KeyError : 如果键属性为 None、字符串为空或者键不存在
- IncompatibleModelContentType : 如果模型内容类型不一致

返回分配给输入键的 MIME 类型。如果指定的内容无 MIME 类型, 那么返回 None:

```
cxt.getModelContentMimeType(key) : string
```

参数:

- key (string) : 键字符串值

异常:

- ModelInputNotSupported : 当不是从 pyspark:applymodel 函数调用此 API 时
- KeyError : 如果键属性为 None、字符串为空或者键不存在

## 数据元数据

此部分描述如何基于 `pyspark.sql.StructField` 设置数据模型属性。

### `spss.datamodel.Role` 对象

该类枚举数据模型中每个字段的有效角色。

**BOTH**: 指示该字段可以是前提条件或结果。

**FREQWEIGHT**: 指示该字段用作频率权重；不向用户显示。

**INPUT**: 指示该字段是预测变量还是前提条件。

**NONE**: 指示在建模期间不之间使用该字段。

**TARGET**: 指示该字段是预测还是结果。

**PARTITION**: 指示该字段用于标识数据分区。

**RECORDID**: 指示该字段用于标识记录标识。

**SPLIT**: 指示该字段用于拆分数据。

### `spss.datamodel.Measure` 对象

该类枚举数据模型中字段的测量级别。

**UNKNOWN**: 指示测量类型未知。

**CONTINUOUS**: 指示测量类型为连续。

**NOMINAL**: 指示测量类型为名义。

**FLAG**: 指示字段值为两个值之一。

**DISCRETE**: 指示字段值应解释为值集合。

**ORDINAL**: 指示测量类型有序。

**TYPELESS**: 指示字段可以具有与其存储相兼容的任何值。

### `pyspark.sql.StructField` 对象

表示 `StructType` 中的字段。 `StructField` 对象有四个字段组成：

- `name` (`string`): `StructField` 名称
- `dataType` (`pyspark.sql.DataType`): 特定数据类型
- `nullable` (`bool`): 如果 `StructField` 的值可包含 `None` 值
- `metadata` (`dictionary`): 用于存储选项属性的 Python 字典

您可以使用元数据字典实例来存储指定字段的测量、角色、标签属性。这些属性的关键字为：

- `measure`: `measure` 属性的关键字
- `role`: `role` 属性的关键字
- `displayLabel`: `label` 属性的关键字

示例：

```
from spss.datamodel.Role import Role
from spss.datamodel.Measure import Measure
_metadata = {}
_metadata['measure'] = Measure.TYPELESS
_metadata['role'] = Role.NONE
_metadata['displayLabel'] = "field label description"
StructField("userName", StringType(), nullable=False,
metadata=_metadata)
```

## 日期、时间和时间戳记

对于使用日期、时间或时间戳记类型数据的操作，将基于值 1970-01-01:00:00:00（使用全球标准时间）将值转换为实数值。

对于日期，该值基于值 1970-01-01（使用全球标准时间）表示天数。

对于时间，该值表示 24 小时制的秒数。

对于时间戳记，该值基于值 1970-01-01:00:00:00（使用全球标准时间）来表示秒数。

## 异常

此部分描述可能的异常实例。

### MetadataException 对象

Python 异常的子类。

如果操作元数据对象期间发生错误，那么将抛出此异常。

### UnsupportedOperationException 对象

Python 异常的子类。

如果特定操作不允许执行，那么将抛出此异常。

### InconsistentOutputDataModel 对象

Python 异常的子类。

如果调用 `setSparkOutputSchema` 和 `setSparkOutputData` 但是两个对象的公共字段名称和存储类型不一致性，那么将抛出此异常。

### IncompatibleModelContentType 对象

Python 异常的子类。

在以下场景中，将抛出此异常：

- 使用 `setModelContentFormString` 来设置模型但是使用 `getModelContentToPath` 以获取值
- 使用 `setModelContentFormPath` 来设置模型但是使用 `getModelContentToString` 以获取值

### DataOutputNotSupported 对象

Python 异常的子类。

在函数 `pyspark:buildmodel` 处理的执行的 `setSparkOutputData` 中发出此异常。

### ModelInputNotSupported 对象

Python 异常的子类。

仅当脚本未在 `pyspark:applymodel` 函数中调用 `getModelContentPathByKey` 和 `getModelContentToString` API 时发生此异常。

### ModelOutputNotSupported 对象

Python 异常的子类。

仅当脚本未在 `pyspark:buildmodel` 函数中调用 `setModelContentFromPath` 和 `setModelContentFromString` API 时发生此异常。

## ASContextException 对象

Python 异常的子类。

如果发生意外的运行时异常，那么抛出此异常。

## 示例

该部分包含 Python for Spark 脚本编制示例。

### 用于处理数据的基本脚本编制示例

```
import spss.pyspark.runtime
from pyspark.sql.types import *

cxt = spss.pyspark.runtime.getContext()

if cxt.isComputeDataModelOnly():
    _schema = cxt.getSparkInputSchema()
    cxt.setSparkOutputSchema(_schema)
else:
    _structType = cxt.getSparkInputSchema()
    df = cxt.getSparkInputData()
    _newDF = df.sample(False, 0.01, 1)
    cxt.setSparkOutputData(_newDF)
```

### 使用 LinearRegressionWithSGD 算法的示例模型构建脚本

```
from pyspark.context import SparkContext
from pyspark.sql.context import SQLContext
from pyspark.sql import Row
from pyspark.mllib.regression import
LabeledPoint, LinearRegressionWithSGD, LinearRegressionModel
from pyspark.mllib.linalg import DenseVector
import numpy
import json

import spss.pyspark.runtime
from spss.pyspark.exceptions import ASContextException

ascontext = spss.pyspark.runtime.getContext()
sc = ascontext.getSparkContext()
df = ascontext.getSparkInputData()

# field settings amd algorithm parameters
target = '%%target_field%%'
predictors = [%%predictor_fields%%]
num_iterations=%%num_iterations%%
prediction_field = "$LR-" + target

# save linear regression model to a filesystem path
def save(model, sc, path):
    data =
sc.parallelize([json.dumps({"intercept":model.intercept,"weights":model.weights.tolist()})])
    data.saveAsTextFile(path)

# print model details to stdout
def dump(model,predictors):
    print(prediction_field+" = " + str(model.intercept))
    weights = model.weights.tolist()
    for i in range(0,len(predictors)):
        print("\t+ "+predictors[i]+"*"+ str(weights[i]))

# check that required fields exist in the input data
input_field_names = [ty[0] for ty in df.dtypes[:]]
if target not in input_field_names:
    raise ASContextException("target field "+target+" not found") for predictor in
predictors:
    if predictor not in input_field_names:
        raise ASContextException("predictor field "+predictor+" not found")
```

```

# define map function to convert from dataframe Row objects to mllib LabeledPoint
def row2LabeledPoint(target,predictors,row):
    pvals = []
    for predictor in predictors:
        pval = getattr(row,predictor)
        pvals.append(float(pval))
    tval = getattr(row,target)
    return LabeledPoint(float(tval),DenseVector(pvals))

# convert dataframe to an RDD containing LabeledPoint

training_points = df.rdd.map(lambda row:
row2LabeledPoint(target,predictors,row))

# build the model

model = LinearRegressionWithSGD.train(training_points,num_iterations,intercept=True)

# write a text description of the model to stdout

dump(model,predictors)

# save the model to the filesystem and store into the output model content

modelpath = ascontext.createTemporaryFolder()
save(model,sc,modelpath)
ascontext.setModelContentFromPath("model",modelpath)

```

## 使用 LinearRegressionWithSGD 算法的示例模型评分脚本

```

import json
import spss.pyspark.runtime
from pyspark.sql import Row
from pyspark.mllib.regression import
LabeledPoint,LinearRegressionWithSGD, LinearRegressionModel
from pyspark.mllib.linalg import DenseVector
from pyspark.sql.context import SQLContext
import numpy
from pyspark.sql.types import DoubleType, StructField

ascontext = spss.pyspark.runtime.getContext()
sc = ascontext.getSparkContext()

prediction_field = "$LR-" + '%%target_field%%'
predictors = [%%predictor_fields%%]

# compute the output schema by adding the prediction field
outputSchema = ascontext.getSparkInputSchema()
outputSchema.fields.append(StructField(prediction_field,
DoubleType(), nullable=True))

# make a prediction based on a regression model and Dataframe Row object
# return a list containing the input row values and the predicted value
def predict(row,model,predictors,infields,prediction_field_name):
    pvals = []
    rdict = row.asDict()
    for predictor in predictors:
        pvals.append(float(rdict[predictor]))
    estimate = float(model.predict(pvals))
    result = []
    for field in infields:
        result.append(rdict[field])
    result.append(estimate)
    return result

# load a serialized model from the filesystem

def load(sc, path):
    js = sc.textFile(path).take(1)[0]
    obj = json.loads(js)
    weights = numpy.array(obj["weights"])
    intercept = obj["intercept"]
    return LinearRegressionModel(weights,intercept)

ascontext.setSparkOutputSchema(outputSchema)

if not ascontext.isComputeDataModelOnly():
    # score the data in the input data frame

```

```

indf = ascontext.getSparkInputData()

model_path = ascontext.getModelContentToPath("model")
model = load(sc,model_path)

# compute the scores
infield_names = [ty[0] for ty in indf.dtypes[:]]
scores_rdd = indf.rdd.map(lambda
row:predict(row,model,predictors,infield_names,prediction_field))

# create an output DataFrame containing the scores
sqlCtx = SQLContext(sc)
outdf = sqlCtx.createDataFrame(scores_rdd,schema=outputSchema)

# return the output DataFrame as the result
ascontext.setSparkOutputData(outdf)

```

## 切换到其他 Python 环境

您可以在 IBM SPSS Modeler 中更改 "扩展" 节点和 "自定义对话框构建程序" 所使用的 Python 环境:

1. 授予 `options.cfg` 的编辑许可权:
  - a. 从 IBM SPSS Modeler Server 安装目录中, 在文本编辑器中打开 `config/options.cfg`。
  - b. 修改以下行以授予用户编辑许可权:

```
administrators, "*"
```

- c. 保存该文件, 然后重新启动 IBM SPSS Modeler Server。
2. 打开 IBM SPSS Modeler 并连接到 IBM SPSS Modeler Server。
3. 转至 **工具 > Manage Modeler Server Configuration**。
4. 在 **数据文件访问** 下, 修改 Python 可执行文件路径, 然后单击 **确定**。
5. 重新启动 IBM SPSS Modeler 服务器, 然后在 IBM SPSS Modeler 中重新连接到该服务器。
6. 转至 **工具 > 管理 Modeler Server 配置**, 并验证是否已保存新的 Python 路径。

## 本机 Python API

您可以从脚本中调用本机 Python API 以与 IBM SPSS Modeler 进行交互。

支持以下 API。

要查看示例, 您可以下载 [样本流](#) 并将其导入到 IBM SPSS Modeler 中。

### 用于数据模型的 API

- **modelerpy.isComputeDataModelOnly()**

您可以使用此 API 来检查当前运行是计算输出数据还是仅计算输出数据模型。当它返回 `true` 时, 脚本不得执行依赖于输入或输出数据的任何任务, 否则运行将失败。

- **modelerpy.getDataModel()**

此 API 联系 IBM SPSS Modeler 以获取输入数据集的数据模型。返回值是 `class DataModel` 的实例, 用于描述输入数据集的元数据, 包括字段计数, 字段名称, 字段存储类型等。

- **modelerpy.setOutputDataModel(dataModel)**

此 API 将类 `DataModel` 的实例发送回 IBM SPSS Modeler, 并且必须在脚本将数据集传递到 IBM SPSS Modeler 之前调用此 API。IBM SPSS Modeler 将使用此 `DataModel` 实例中描述的元数据来处理 IBM SPSS Modeler 端的数据。

### 用于建模的 API

- **modelerpy.saveModel(model, name='model')**

此 API 将 Python 模型转换为 IBM SPSS Modeler 模型，然后由 IBM SPSS Modeler 保存。在构建 Python 模型时，应该从建模节点调用此 API。调用此 API 后，会将保存的模型复制到生成的模型块。

- **modelerpy.loadModel(name='model')**

此 API 装入 IBM SPSS Modeler 保存的模型，并为保存的模型创建 Python 对象。从模型块调用此 API 以装入保存的模型进行进一步处理，例如评分。

## 输入/输出数据集的 API

- **modelerpy.readPandasDataframe()**

此 API 将数据集从 IBM SPSS Modeler 读取到 Python。返回值为 Python Pandas DataFrame (二维数据结构，例如二维数组或包含行和列的表)。

- **modelerpy.writePandasDataframe(df)**

此 API 将 Python Pandas DataFrame 从 Python 写入 IBM SPSS Modeler。

## 软件包的 API

- **modelerpy.installPackage(package)**

此 API 从 pypi.org 中拉取软件包并将其安装。

- **modelerpy.uninstallPackage(package)**

此 API 将卸载已安装的软件包。

## 用于元数据的 API

下列与元数据相关的类应与 `modelerpy.getDataModel` 和 `modelerpy.setOutputDataModel` 配合使用。

- **modelerpy.DataModel**

此 API 是元数据的主条目类。它包含 class `Field` 的实例数组，并包含以下方法

- **modelerpy.DataModel.getFields**

此方法返回 class `Field` 实例的数组。

- **modelerpy.DataModel.addField**

此方法将 `Field` 的实例添加到元数据数组。

- **modelerpy.Field**

`Field` 类是存储实际元数据信息的位置，包括字段名称，存储和度量。

- **modelerpy.Field.getName**

此方法返回字段的名称。

- **modelerpy.Field.getStorage**

此方法返回字段的存储。有效存储器包括: `integer`, `real`, `string`, `date`, `time` 和 `timestamp`。

- **modelerpy.Field.getMeasure**

此方法返回字段的测量。有效度量包括: `discrete`, `flag`, `nominal`, `ordinal` 和 `continuous`。

以下示例代码通过使用 `modelerpy.Field` 数组调用 `modelerpy.DataModel` 构造函数来构造 `DataModel` 对象。 `modelerpy.Field` 构造函数接受字段名，字段存储和字段测量作为其输入参数 (字段存储和字段测量是必需的; 字段测量是可选的)。

```
dataModel = modelerpy.DataModel([
#                               %FieldName%, %StorageType%, %MeasurementType%
    modelerpy.Field('StringField', 'string', 'nominal'),
    modelerpy.Field('FloatField', 'real', 'continuous'),
    modelerpy.Field('IntegerField', 'integer', 'ordinal'),
```



```

        modelerpy.Field('BooleanField', 'integer', 'flag'),
        modelerpy.Field('DatetimeField', 'timestamp', 'continuous'),
        modelerpy.Field('TimeField', 'time', 'continuous'),
        modelerpy.Field('DateField', 'date', 'continuous'),
    ])
# StorageType could be: integer, real, string, date, time, timestamp
# MeasurementType could be: discrete, flag, nominal, ordinal, continuous

outputDataModel = modelerDataModel
outputDataModel.addField(modelerpy.Field(field_outlier, "real", measure="flag"))
outputDataModel.addField(modelerpy.Field(field_dist_hp, "real", measure="continuous"))

```

## “扩展”节点

作为对 IBM SPSS Modeler 及其数据挖掘功能的补充，“扩展”节点使专业用户能够输入自己的 R 脚本或 Python for Spark 脚本以执行数据处理、模型构建和模型评分。

### “扩展导出”节点

通过“扩展导出”节点，您可以运行 R 或 Python for Spark 脚本来导出数据。

#### “扩展导出”节点 -“语法”选项卡

选择语法类型 - **R** 或 **Python for Spark**。请参阅以下部分以获取更多信息。语法就绪时，您可以单击运行来执行“扩展导出”节点。

#### R 语法

**R 语法。** 您可以在此字段中输入或粘贴用于数据分析的定制 R 脚本语法。

**转换标志字段。** 指定标志字段的处理方式。共有两个选项：**将字符串转换为因子**，**将整数和实数转换为双精度数和逻辑值（True 和 False）**。如果选择**逻辑值（True 和 False）**，那么标志字段的原始值将丢失。例如，如果某个字段的值为 Male 和 Female，那么这些值将更改为 True 和 False。

**将缺失值转换为 R “不可用”值 (NA)。** 选中时，任何缺失值都将转换为 R NA 值。R 使用值 NA 来标识缺失值。您使用的某些 R 函数可能有一个参数，可用于控制当数据包含 NA 时函数的行为方式。例如，该函数可能会允许您选择自动排除包含 NA 的记录。如果未选择此选项，那么所有缺失值都将按原样传递到 R，并可能导致执行 R 脚本时发生错误。

**将日期/时间字段转换为特殊时区控制的 R 类。** 如果选择此选项，那么会将带有日期或日期时间格式的变量转换为 R 日期/时间对象。必须选择下列选项之一：

- 将具有日期或日期时间格式的 **R POSIXct**。变量将转换为 R POSIXct 对象。
- **R POSIXlt**（列表）。将具有日期或日期时间格式的变量转换为 R POSIXlt 对象。

**注：**POSIX 格式是高级选项。仅当您的 R 脚本指定以需要这些格式的方式处理日期时间字段时才使用这些选项。POSIX 格式不适用于具有时间格式的变量。

#### Python 语法

**Python 语法。** 您可以针对数据分析向此字段中输入或粘贴定制的 Python 脚本语法。有关 Python for Spark 的更多信息，请参阅 [Python for Spark](#) 和 [Python for Spark 的脚本编制](#)。

#### “扩展导出”节点 -“控制台输出”选项卡

**控制台输出选项卡**包含当“语法”选项卡上的 R 脚本或 Python for Spark 脚本运行时接收到的任何输出（例如，如果使用 R 脚本，当执行语法选项卡上的 **R 语法** 字段中的 R 脚本时，它显示从 R 控制台接收到的输出）。此输出可能包括执行 R 或 Python 脚本时生成的 R 或 Python 错误消息或警告。输出可主要用于调试脚本。**控制台输出选项卡**还包含 **R 语法**或 **Python 语法** 字段中的脚本。

每次执行“扩展导出”脚本时，都会使用从 R 控制台或 Python for Spark 接收到的输出来覆盖**控制台输出**选项卡的内容。输出不能编辑。

## 发布流

系统使用以下任何一个标准导出节点直接从 IBM SPSS Modeler 发布流：数据库、平面文件、统计信息 导出、扩展导出、数据收集导出、SAS 导出、Excel 和 XML 导出节点。导出节点的类型决定每次使用 IBM SPSS Modeler Solution Publisher Runtime 或外部应用程序执行发布的流时要写入的结果格式。例如，如果您想每次运行发布的流时将结果写入数据库，则请使用数据库导出节点。

## 发布流

1. 以普通方式打开或构建一个流，并在最后附加一个导出节点。
2. 在导出节点中的“发布”选项卡上，指定已发布文件的根名（即，.pim、.par 和 .xml 扩展将附加到的文件名）。
3. 单击**发布**以发布该流，或选择**发布流**以便每次执行该节点时自动发布流。

**已发布的名称。** 指定已发布图像和参数文件的根名。

- **图像文件 (\*.pim)** 提供了 Runtime 执行发布的流时所需的所有信息，这些信息与导出时完全相同。如果您确信不需要更改流的任何设置（如输入数据源或输出数据文件），则可以只部署该图像文件。
- **参数文件 (\*.par)** 包含有关数据源、输出文件和执行选项的可配置信息。如果您希望能够在不重新发布流的情况下控制流的输入或输出，则同时需要参数文件和图像文件。
- **元数据文件 (\*.xml)** 用于描述图像及其数据模型的输入和输出。它旨在供内嵌 runtime 库并需要了解输入数据和输出数据结构的应用程序使用。

**注：** 仅当您选择了**发布元数据**选项时，才会生成此文件。

**发布参数。** 如果需要，可以在 \*.par 文件中包含流参数。在通过编辑 \*.par 文件或通过运行时 API 执行图像时，可以更改这些流参数值。

此选项用于启用**参数**按钮。单击该按钮时，将显示“发布参数”对话框。

通过在**发布**列中选择相关选项来选择您要包含在已发布图像中的参数。

**在流执行时。** 指定执行节点时是否自动发布流。

- **导出数据。** 以标准方式执行导出节点，但不发布流。（基本上，节点在 IBM SPSS Modeler 中的执行方式与 IBM SPSS Modeler Solution Publisher 不可用时的执行方式相同。）如果您选择此选项，那么只有通过单击导出节点对话框上的**发布**进行明确发布时，才会发布该。另外，您还可以通过使用工具栏上的发布工具，或通过使用脚本来发布当前流。
- **发布流。** 使用 IBM SPSS Modeler Solution Publisher 发布流以用于部署。如果您希望每次执行节点时都自动发布流，则请选择此选项。

**注：**

- 如果您计划使用新数据或更新后的数据运行发布的流，则要注意的重要一点是，输入文件中字段的顺序必须与发布的流中指定的源节点输入文件中的字段顺序相同。
- 发布到外部应用程序时，请考虑对无关的字段进行过滤，或者对字段进行重命名以符合输入要求。通过在导出节点之前使用过滤节点，可以完成上述两个操作。

## “扩展输出”节点

如果在“扩展输出”节点对话框的**输出**选项卡上选择了**输出到屏幕**，那么将在输出浏览器窗口中显示屏幕上的输出。这还会将输出添加到输出管理器。输出浏览器窗口具有它自己的菜单集，使用这些菜单，可以打印或保存输出，也可以将输出导出为其他格式。**编辑**菜单仅包含**复制**选项。“扩展输出”节点的输出浏览器有两个选项卡：用于显示文本输出的**文本输出**选项卡，以及用于显示图形和图表的**图形输出**选项卡。

如果在“扩展输出”节点对话框的**输出**选项卡上选择了**输出到文件**，那么在成功执行“扩展输出”节点后，不会显示输出浏览器窗口。

## “扩展输出”节点 -“语法”选项卡

选择语法类型 - **R** 或 **Python for Spark**。请参阅以下部分以获取更多信息。语法就绪时，您可以单击**运行**来执行“扩展输出”节点。输出对象将添加到输出管理器，或者选择性地添加到**输出选项卡**上的**文件名**字段中指定的文件中。

### R 语法

**R 语法**。您可以在此字段中输入或粘贴用于数据分析的定制 R 脚本语法。

**转换标志字段**。指定标志字段的处理方式。共有两个选项：**将字符串转换为因子**，**将整数和实数转换为双精度数和逻辑值 (True 和 False)**。如果选择**逻辑值 (True 和 False)**，那么标志字段的原始值将丢失。例如，如果某个字段的值为 Male 和 Female，那么这些值将更改为 True 和 False。

**将缺失值转换为 R "不可用" 值 (NA)**。选中时，任何缺失值都将转换为 R NA 值。R 使用值 NA 来标识缺失值。您使用的某些 R 函数可能有一个参数，可用于控制当数据包含 NA 时函数的行为方式。例如，该函数可能会允许您选择自动排除包含 NA 的记录。如果未选择此选项，那么所有缺失值都将按原样传递到 R，并可能导致执行 R 脚本时发生错误。

**将日期/时间字段转换为特殊时区控制的 R 类**。如果选择此选项，那么会将带有日期或日期时间格式的变量转换为 R 日期/时间对象。必须选择下列选项之一：

- 将具有日期或日期时间格式的 **R POSIXct**。变量将转换为 R POSIXct 对象。
- **R POSIXlt (列表)**。将具有日期或日期时间格式的变量转换为 R POSIXlt 对象。

**注**: POSIX 格式是高级选项。仅当您的 R 脚本指定以需要这些格式的方式处理日期时间字段时才使用这些选项。POSIX 格式不适用于具有时间格式的变量。

### Python 语法

**Python 语法**。您可以针对数据分析向此字段中输入或粘贴定制的 Python 脚本语法。有关 Python for Spark 的更多信息，请参阅 [Python for Spark](#) 和 [Python for Spark 的脚本编制](#)。

## “扩展输出”节点 -“控制台输出”选项卡

**控制台输出选项卡**包含当“语法”选项卡上的 R 脚本或 Python for Spark 脚本运行时接收到的任何输出（例如，如果使用 R 脚本，当执行**语法**选项卡上的 **R 语法**字段中的 R 脚本时，它显示从 R 控制台接收到的输出）。此输出可能包括执行 R 或 Python 脚本时生成的 R 或 Python 错误消息或警告。输出可主要用于调试脚本。**控制台输出选项卡**还包含 **R 语法**或 **Python 语法**字段中的脚本。

每次执行“扩展输出”脚本时，都会使用从 R 控制台或 Python for Spark 接收到的输出来覆盖**控制台输出选项卡**的内容。输出不能编辑。

## “扩展输出”节点 -“输出”选项卡

**输出名称**。指定执行此节点时生成的输出的名称。选择**自动**时，输出名称将自动设置为“R 输出”或“Python 输出”，具体取决于脚本类型。（可选）可以选择**定制**以指定其他名称。

**输出到屏幕**。选择此选项以在新窗口中生成并显示输出。这还会将输出添加到输出管理器。

**输出到文件**。选择此选项可将输出保存到文件。执行此操作将启用**输出图形**和**输出文件**单选按钮。

**输出图形**。只有在选择了**输出到文件**的情况下才会启用此选项。选择此选项可以将执行“扩展输出”节点所产生的任何图形保存到文件中。请在**文件名**字段中指定要用于生成的输出的文件名。单击省略符按钮 (...) 以选择特定文件和位置。请在**文件类型**下拉列表中指定文件类型。可用的文件类型如下所示：

- 输出对象 (.cou)
- HTML (.html)

**输出文本**。只有在选择了**输出到文件**的情况下才会启用此选项。选择此选项可以将执行“扩展输出”节点所产生的任何文本输出保存到文件中。请在**文件名**字段中指定要用于生成的输出的文件名。单击省略符按钮 (...) 以指定特定文件和位置。请在**文件类型**下拉列表中指定文件类型。可用的文件类型如下所示：

- HTML (.html)

- 输出对象 (.cou)
- 文本文档 (.txt)

## 扩展输出浏览器

如果在“扩展输出”节点对话框的**输出选项卡**上选择了**输出到屏幕**，那么将在输出浏览器窗口中显示屏幕上的输出。这还会将输出添加到输出管理器。输出浏览器窗口具有它自己的菜单集，使用这些菜单，可以打印或保存输出，也可以将输出导出为其他格式。**编辑菜单**仅包含**复制**选项。“扩展输出”节点的输出浏览器有两个选项卡：

- **文本输出**选项卡显示文本输出
- **图形输出**选项卡显示图形和图表

如果在“扩展输出”节点对话框的**输出选项卡**上选择了**输出到文件**，而不是**输出到屏幕**，那么在成功执行“扩展输出”节点后，不会显示输出浏览器窗口。

### 扩展输出浏览器 -“文本输出”选项卡

**文本输出**选项卡显示执行“扩展输出”节点的**语法**选项卡上的 R 脚本或 Python for Spark 脚本时生成的任何文本输出。

**注:** 由于执行扩展输出脚本而产生的 R 或 Python for Spark 错误消息或警告始终显示在“扩展输出”节点的**控制台输出**选项卡上。

### 扩展输出浏览器 -“图形输出”选项卡

Python Spark 的“扩展输出”节点现在包含一个类似于 R 的**图形输出**选项卡。此选项卡显示执行“扩展输出”节点的**语法**选项卡上的 R 脚本或 Python for Spark 脚本时生成的任何图形或图表。例如，如果您的 R 脚本包含对 R plot 函数的调用，那么产生的图形会显示在此选项卡上。现在，您可以使用如下脚本来生成图形：

```
import spss.pyspark.runtime

import numpy
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

ascontext = spss.pyspark.runtime.getContext()
indf = ascontext.getSparkInputData()
sns.pairplot(indf.toPandas(), hue='K')
sns.pairplot(indf.toPandas(), hue='Age')
plt.show()
```

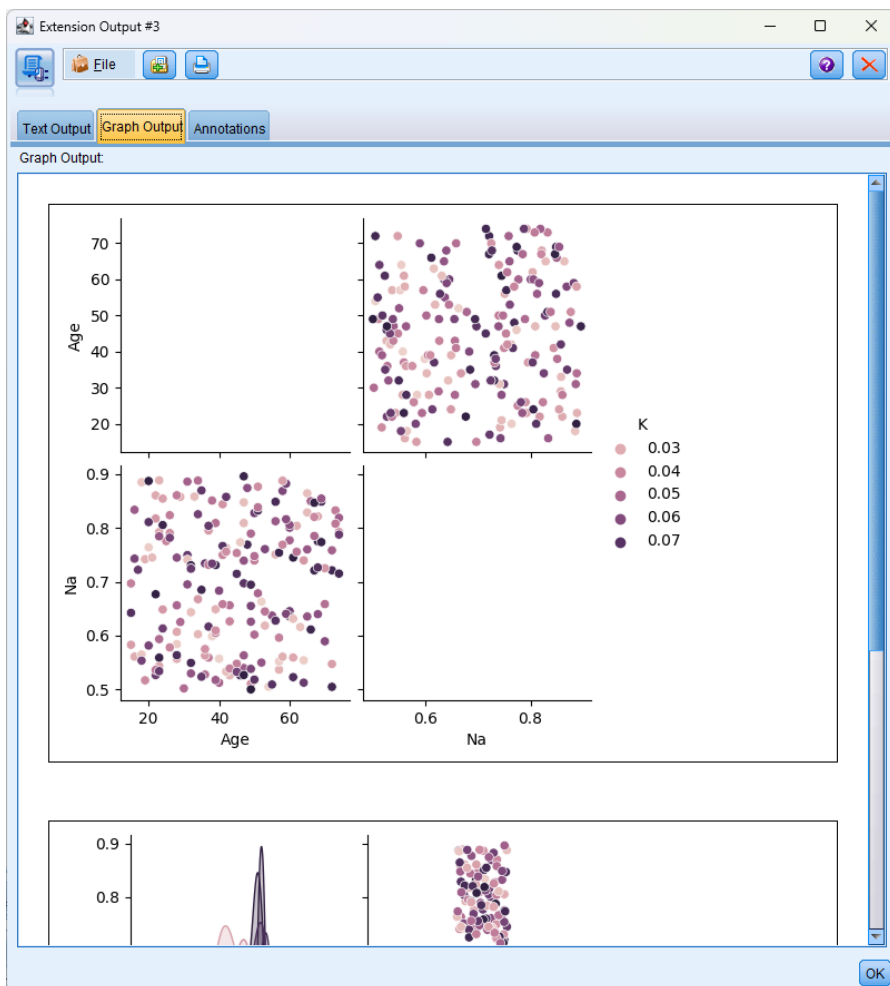


图 1: “图形输出”选项卡

## “扩展模型”节点

使用“扩展模型”节点，您可以运行 R 或 Python for Spark 脚本以构建模型并对其进行评分。

### “扩展模型”节点 -“语法”选项卡

选择语法类型 - R 或 Python for Spark。然后，向以下某个字段输入或粘贴定制的脚本语法。语法就绪时，您可以单击运行来执行“扩展模型”节点。

#### R 语法

**R 模型构建语法。** 您可以针对模型构建向此字段输入或粘贴定制的 R 脚本语法。

**R 模型评分语法。** 您可以针对模型评分向此字段输入或粘贴定制的 R 脚本语法。

#### Python for Spark 语法

**Python 模型构建语法。** 您可以针对模型构建向此字段输入或粘贴定制的 Python 脚本语法。

**Python 模型评分语法。** 您可以针对模型评分向此字段输入或粘贴定制的 Python 脚本语法。

有关 Python for Spark 的更多信息，请参阅 [第 1 页的『Python for Spark』](#) 和 [第 1 页的『使用 Python for Spark 编写脚本』](#)。

## “扩展模型”节点 -“模型选项”选项卡

**模型名称。** 您可以根据目标或标识字段（或未指定此类字段时的模型类型）自动生成模型名称，或者指定定制名称。

## “扩展模型”节点 -“控制台输出”选项卡

**控制台输出**选项卡包含当“语法”选项卡上的 R 脚本或 Python for Spark 脚本运行时接收到的任何输出（例如，如果使用 R 脚本，当执行**语法**选项卡上的 **R 语法**字段中的 R 脚本时，它显示从 R 控制台接收到的输出）。此输出可能包括执行 R 或 Python 脚本时生成的 R 或 Python 错误消息或警告。输出可主要用于调试脚本。**控制台输出**选项卡还包含 **R 语法**或 **Python 语法**字段中的脚本。

每次执行“扩展模型”脚本时，都会使用从 R 控制台或 Python for Spark 接收到的输出来覆盖**控制台输出**选项卡的内容。输出不能编辑。

## “扩展模型”节点 -“文本输出”选项卡

如果通过在“扩展模型”节点对话框的**模型选项**选项卡上选中**显示 R 文本输出**复选框进行请求，那么**文本输出**选项卡存在于“扩展模型”节点中。此选项卡仅能显示文本输出。此选项卡显示由执行 R 模型构建脚本生成的所有文本输出。如果您再次执行模型构建脚本，而未先对模型指定其他名称，那么来自先前执行的**文本输出**选项卡的内容将被覆盖。文本输出不能编辑。

如果您在脚本中包含对 R `sink` 函数的调用，那么在此函数之后生成的所有输出会保存到指定文件并且不会显示在**文本输出**选项卡。

**注：**由于执行模型构建脚本而产生的 R 或 Python for Spark 错误消息或警告始终显示在“扩展模型”节点的**控制台输出**选项卡上。

## 扩展模型块

执行“扩展模型”节点后会生成扩展模型块并将其置于“模型”选用板上，该模型块包含定义模型构建和模型评分的 R 或 Python for Spark 脚本。缺省情况下，扩展模型块包含用于模型评分的脚本、用于读取数据的选项以及 R 控制台或 Python for Spark 的任何输出。（可选）扩展模型块还可以包含各种其他格式的模型输出，例如图形和文本输出。在生成扩展模型块且将其添加到流画布后，可将输出接点与其进行连接。此时，该输出接点通常会在 IBM SPSS Modeler 流中用于获取有关数据和模型的信息以及导出各种格式的数据。

要将此节点与 R 配合使用，您必须安装 IBM SPSS Modeler - Essentials for R。请参阅 *IBM SPSS Modeler - Essentials for R: Installation Instructions* 以获取安装指示信息和兼容性信息。另外，还必须在计算机上安装 R 的兼容版本。

## 扩展模型块 -“语法”选项卡

**语法**选项卡始终存在于扩展模型块中。

**R 模型评分语法。** 如果使用 R，那么用于模型评分的 R 脚本显示在此字段中。缺省情况下，该字段为启用状态，但不可编辑。要编辑 R 模型评分脚本，请单击**编辑**。

**Python 模型评分语法。** 如果使用 Python for Spark，那么用于模型评分的 Python 脚本显示在此字段中。缺省情况下，该字段为启用状态，但不可编辑。要编辑 Python 模型评分脚本，请单击**编辑**。

**编辑。** 单击**编辑**以使评分语法字段可编辑。然后，您可以通过在评分语法字段中进行输入来编辑模型评分脚本。例如，如果您在执行扩展模型块后在模型评分脚本中识别了错误，那么您可能希望编辑您的模型评分脚本。如果您通过执行“扩展模型”节点重新生成模型，那么您在扩展模型块中对模型评分脚本所做的任何更改都将丢失。

## 扩展模型块 -“模型选项”选项卡

**模型选项**选项卡始终存在于扩展模型块中。

**读取数据选项。** 这些选项仅适用于 R，不适用于 Python for Spark。借助这些选项，您可以指定格式为日期或日期时间的缺失值、标志字段和变量的处理方式。

- **以批处理方式读取数据。** 如果要处理大量数据（例如，数据太大无法适合 R 引擎的内存），请使用此选项将数据划分为可单独发送和处理的批次。指定要包含在每个批次中的数据记录的最大数量。
- 对于“扩展变换”节点和扩展评分块，通过 R 脚本（批处理）传递数据。因此，用于在 Hadoop 或“数据库”环境中运行的模型评分节点和流程节点的脚本不应包括跨数据行或组合数据行的操作，例如排序或汇总。实施此限制的目的是确保能够在 Hadoop 环境中以及数据库内挖掘期间分割数据。如果在 SPSS Modeler Server 中运行模型评分脚本，那么此限制不适用。“扩展输出”和“扩展模型”节点没有此限制。
- **转换标志字段。** 指定标志字段的处理方式。共有两个选项：**将字符串转换为因子，将整数和实数转换为双精度数和逻辑值（True 和 False）**。如果选择**逻辑值（True 和 False）**，那么标志字段的原始值将丢失。例如，如果某个字段的值为 Male 和 Female，那么这些值将更改为 True 和 False。
  - **将缺失值转换为 R “不可用”值 (NA)。** 选中时，任何缺失值都将转换为 R NA 值。R 使用值 NA 来标识缺失值。您使用的某些 R 函数可能有一个参数，可用于控制当数据包含 NA 时函数的行为方式。例如，该函数可能会允许您选择自动排除包含 NA 的记录。如果未选择此选项，那么所有缺失值都将按原样传递到 R，并可能导致执行 R 脚本时发生错误。
  - **将日期/时间字段转换为具有特殊时区控件的 R 类** 如果选择此选项，那么会将带有日期或日期时间格式的变量转换为 R 日期/时间对象。必须选择下列选项之一：

- 将具有日期或日期时间格式的 **R POSIXct** 变量将转换为 R POSIXct 对象。
- **R POSIXlt**（列表）。将具有日期或日期时间格式的变量转换为 R POSIXlt 对象。

**注：**POSIX 格式是高级选项。仅当您的 R 脚本指定以需要这些格式的方式处理日期时间字段时才使用这些选项。POSIX 格式不适用于格式为时间的变量。

针对数据库运行扩展模型块时，未识别您为**转换标志字段、将缺失值转换为 R “不可用”值 (NA)** 以及**将日期/时间字段转换为具有时区特殊控件的 R 类**控件选择的选项。针对数据库运行节点时，将改为使用这些控件的缺省值：

- 转换标志字段设置为**将字符串转换为因子，将整数和实数转换为双精度数**。
- 选择了**将缺失值转换为 R 的“不可用”值 (NA)**。
- 未选择将日期/时间字段转换为具有时区特殊控件的 R 类。

## 扩展模型块 -“图形输出”选项卡

如果通过在“扩展模型”节点对话框的**模型选项**选项卡上选中**以 HTML 格式显示 R 图形**复选框进行请求，那么**图形输出**选项卡存在于扩展模型块中。由执行模型构建 R 脚本产生的图形可显示在此选项卡上。例如，如果您的 R 脚本包含对 R plot 函数的调用，那么产生的图形会显示在此选项卡上。如果您再次执行模型构建脚本，而未先对模型指定其他名称，那么来自先前执行的**图形输出**选项卡的内容将被覆盖。

## 扩展模型块 -“文本输出”选项卡

如果通过在“扩展模型”节点对话框的**模型选项**选项卡上选中**显示 R 文本输出**复选框进行请求，那么**文本输出**选项卡存在于扩展模型块中。此选项卡仅能显示文本输出。此选项卡显示由执行“扩展模型”脚本生成的任何文本输出。如果您再次执行“扩展模型”脚本，而未先对模型指定其他名称，那么来自先前执行的**文本输出**选项卡的内容将被覆盖。文本输出不能编辑。

**注：**

- 如果您在脚本中包含对 R sink 函数的调用，那么在此函数之后生成的所有输出会保存到指定文件并且不会显示在**文本输出**选项卡。
- 由于执行“扩展模型”脚本而产生的错误消息或警告始终显示在“扩展模型”节点**控制台输出**选项卡上。

## 扩展模型块 -“控制台输出”选项卡

**控制台输出**选项卡始终存在于扩展模型块中。它包含当“语法”选项卡上的 R 脚本或 Python for Spark 脚本运行时接收到的任何输出（例如，如果使用 R 脚本，当执行扩展模型块**语法**选项卡上的**R 模型评分语法**字段中的 R 脚本时，它显示从 R 控制台接收到的输出）。此输出包括在执行 R 或 Python 脚本时生成的任何 R 或 Python 错误消息或警告，以及 R 控制台的任何文本输出。输出可主要用于调试脚本。

每次执行模型评分脚本时，都会使用从 R 控制台或 Python for Spark 接收到的输出来覆盖**控制台输出**选项卡的内容。控制台输出不能编辑。



## “扩展变换”节点

借助“扩展变换”节点，您可以使用 R 脚本编写或 Python for Spark 脚本编写从 IBM SPSS Modeler 流中获取数据并对该数据应用变换。完成修改后，数据将返回到流中以进行进一步处理、模型构建和模型评分。“扩展变换”节点支持使用以 R 或 Python for Spark 编写的算法来变换数据，并使用户能够开发针对特定问题进行定制的数据变换方法。

要将此节点与 R 配合使用，您必须安装 IBM SPSS Modeler - Essentials for R。请参阅 *IBM SPSS Modeler - Essentials for R: Installation Instructions* 以获取安装指示信息和兼容性信息。另外，还必须在计算机上安装 R 的兼容版本。

### “扩展变换”节点 -“语法”选项卡

选择语法类型 - **R** 或 **Python for Spark**。请参阅以下部分以获取更多信息。语法就绪时，您可以单击运行来执行“扩展变换”节点。

#### R 语法

**R 语法。** 您可以在此字段中输入或粘贴用于数据分析的定制 R 脚本语法。

**转换标志字段。** 指定标志字段的处理方式。共有两个选项：**将字符串转换为因子**，**将整数和实数转换为双精度数和逻辑值 (True 和 False)**。如果选择**逻辑值 (True 和 False)**，那么标志字段的原始值将丢失。例如，如果某个字段的值为 Male 和 Female，那么这些值将更改为 True 和 False。

**将缺失值转换为 R “不可用”值 (NA)。** 选中时，任何缺失值都将转换为 R NA 值。R 使用值 NA 来标识缺失值。您使用的某些 R 函数可能有一个参数，可用于控制当数据包含 NA 时函数的行为方式。例如，该函数可能会允许您选择自动排除包含 NA 的记录。如果未选择此选项，那么所有缺失值都将按原样传递到 R，并可能导致执行 R 脚本时发生错误。

**将日期/时间字段转换为特殊时区控制的 R 类。** 如果选择此选项，那么会将带有日期或日期时间格式的变量转换为 R 日期/时间对象。必须选择下列选项之一：

- 将具有日期或日期时间格式的 **R POSIXct**。变量将转换为 R POSIXct 对象。
- **R POSIXlt (列表)**。将具有日期或日期时间格式的变量转换为 R POSIXlt 对象。

注：POSIX 格式是高级选项。仅当您的 R 脚本指定以需要这些格式的方式处理日期时间字段时才使用这些选项。POSIX 格式不适用于具有时间格式的变量。

#### Python 语法

**Python 语法。** 您可以针对数据分析向此字段中输入或粘贴定制的 Python 脚本语法。有关 Python for Spark 的更多信息，请参阅 [Python for Spark](#) 和 [Python for Spark 的脚本编制](#)。

### “扩展变换”节点 -“扩展输出”选项卡

**控制台输出选项卡**包含当“语法”选项卡上的 R 脚本或 Python for Spark 脚本运行时接收到的任何输出（例如，如果使用 R 脚本，当执行语法选项卡上的 **R 语法** 字段中的 R 脚本时，它显示从 R 控制台接收到的输出）。此输出可能包括执行 R 或 Python 脚本时生成的 R 或 Python 错误消息或警告。输出可主要用于调试脚本。**控制台输出选项卡**还包含 **R 语法**或 **Python 语法** 字段中的脚本。

每次执行“扩展变换”脚本时，都会使用从 R 控制台或 Python for Spark 接收到的输出来覆盖**控制台输出选项卡**的内容。输出不能编辑。

## “扩展导入”节点

通过“扩展导入”节点，您可以运行 R 或 Python for Spark 脚本来导入数据。

### “扩展导入”节点 -“语法”选项卡

选择语法类型 - **R** 或 **Python for Spark**。然后，为导入数据输入或粘贴定制脚本。语法就绪时，您可以单击运行来执行“扩展导入”节点。



## R 示例

```
# import R demo data cars to modeler
modelerData <- cars

# write the data model that matches the data
var1<-c(fieldName="speed",fieldLabel="",fieldStorage="integer",fieldMeasure="",fieldFormat="",
fieldRole="")
var2<-c(fieldName="dist",fieldLabel="",fieldStorage="integer",fieldMeasure="",fieldFormat="",
fieldRole="")
modelerDataModel<-data.frame(var1, var2)
```

## Python for Spark 示例

```
import spss.pyspark.runtime
from pyspark.sql import SQLContext
from pyspark.sql.types import *

cxt = spss.pyspark.runtime.getContext()
if cxt.isComputeDataModelOnly():
    _schema = StructType([StructField("Age", LongType(), nullable=True), \
                             StructField("Sex", StringType(), nullable=True), \
                             StructField("BP", StringType(), nullable=True), \
                             StructField("Cholesterol", StringType(), nullable=True), \
                             StructField("Na", DoubleType(), nullable=True), \
                             StructField("K", DoubleType(), nullable=True), \
                             StructField("Drug", StringType(), nullable=True)])
    cxt.setSparkOutputSchema(_schema)
else:
    sqlContext = cxt.getSparkSQLContext()
    # the demo data is in modeler installation path
    df = sqlContext.read.option("inferSchema", "true").option("header", "true").csv("/opt/IBM/
SPSS/ModelerServer/Cloud/demos/DRUG1n")
    cxt.setSparkOutputData(df)
    df.show()
    # print (df.dtypes[:])
```

## “扩展导入”节点 -“控制台输出”选项卡

**控制台输出**选项卡包含当“语法”选项卡上的 R 脚本或 Python for Spark 脚本运行时接收到的任何输出（例如，如果使用 R 脚本，当执行语法选项卡上的 R 语法字段中的 R 脚本时，它显示从 R 控制台接收到的输出）。此输出可能包括执行 R 或 Python 脚本时生成的 R 或 Python 错误消息或警告。输出可主要用于调试脚本。**控制台输出**选项卡还包含 **R 语法**或 **Python 语法**字段中的脚本。

每次执行“扩展导入”脚本时，都会使用从 R 控制台或 Python for Spark 接收到的输出来覆盖**控制台输出**选项卡的内容。输出不能编辑。

## 过滤或重命名字段

您可以在流中的任意时间点上重命名或排除字段。例如，作为医学研究人员，您可能不关心患者（记录级别数据）的钾水平（字段级别数据）；因此，您可以过滤掉 K（钾）字段。也可以使用单独的“过滤”节点，或者源节点或输出节点上的“过滤”选项卡实现此操作。无论使用哪种节点，结果都是一样的。

- 可以在将数据从源节点（如变量文件、固定文件、统计信息文件、XML 或扩展导入）读入 IBM SPSS Modeler 时对字段进行重命名或过滤。
- 使用过滤节点，可以在流的任何位置对字段进行重命名或过滤。
- 通过统计信息导出、统计信息变换、统计信息模型和统计信息输出节点，可以对字段进行过滤或重命名，使之符合 IBM SPSS Statistics 命名标准。
- 您可以使用以上任何节点中的“过滤器”选项卡来定义或编辑多响应集。
- 最后，可以使用“过滤”节点将一个源节点中的字段映射至另一个源节点。



## 第 2 章 扩展(X)

扩展是可扩展 IBM SPSS Modeler 的功能的定制组件。扩展打包在扩展束（.mpe 文件）中，并安装到 IBM SPSS Modeler。任何用户都可创建扩展，并可通过共享关联扩展束与其他用户共享。

提供了以下实用程序来处理扩展：

- 第 23 页的『扩展中心』（可从**扩展 > 扩展中心**访问）是用于从 GitHub 上的 IBM SPSS 预测性分析集合搜索、下载和安装扩展的接口。从“扩展中心”对话框，还可查看计算机上安装的扩展的详细信息，获取安装的扩展的更新以及移除扩展。
- 您可以从**扩展 > 安装本地扩展束**安装本地计算机上存储的扩展束。
- 您可以使用针对扩展的定制对话框构建器创建包含用户界面的扩展，此用户界面称为定制节点对话框。定制节点对话框会生成 R 脚本或 Python for Spark 脚本，用于执行与此扩展关联的任务。在设计定制对话框时设计生成的脚本。

### 扩展中心

从“扩展中心”对话框，可执行以下任务：

- 探索 GitHub 上的 IBM SPSS 预测性分析集合中提供的扩展。您可以选择扩展以立即安装，也可以下载选定的扩展并在以后安装。
- 获取计算机上已经安装的扩展的更新版本。
- 查看有关计算机上已经安装的扩展的详细信息。
- 移除计算机上已经安装的扩展。

要下载或移除扩展：

1. 从菜单选择：**扩展 > 扩展中心**
2. 选择要下载或移除的扩展并单击**确定**。单击**确定**时将处理在“浏览”和“已安装”选项卡上进行的所有选择。

缺省情况下，选择下载的扩展将下载并安装在计算机上。从“设置”选项卡，可选择将选定扩展下载到指定位置而不进行安装。您可在以后选择**扩展 > 安装本地扩展束**安装它们。对于 Windows，可以通过双击扩展束文件来安装扩展。

**要点：**

- 对于 Windows 7 以及更高版本，安装现有扩展束的更新后版本可能需要使用管理员特权运行 IBM SPSS Modeler。您可以通过右键单击 IBM SPSS Modeler 的图标并选择**以管理员身份运行**来使用管理员权限启动 IBM SPSS Modeler。具体而言，如果您接收到表明未能安装一个或多个扩展束的错误消息，请尝试使用管理员权限运行。
- 如果是通过代理连接到因特网，在尝试通过**扩展 > 扩展中心**菜单选项打开“扩展中心”时，可能会收到如下错误：未检测到因特网连接，因此某些功能不可用。要解决此问题，请将以下参数添加到 jvm.cfg 文件（位于 SPSS Modeler 安装的 config 目录）中的 # JVM 选项中。保存文件并重新启动 SPSS Modeler。

```
options, "-DproxyHost=proxyIP"  
options, "-DproxyPort=proxyPort"
```

**注：**可通过单击“已安装”选项卡上的扩展的**更多信息 ...**，来查看您在安装扩展时同意的许可。

### “浏览”选项卡

“浏览”选项卡显示在 GitHub(<https://ibmpredictiveanalytics.github.io/>) 上的 IBM SPSS 预测性分析 集合中提供的所有扩展。在“探索”选项卡中，可以选择要下载和安装的新扩展，并且可以为计算机上已安装的扩展选择更新。“浏览”选项卡需要因特网连接。

- 对于每个扩展，会显示最新版本号和该版本的关联日期。还会提供扩展的简要摘要。针对计算机上已安装的扩展，还会显示安装的版本号。
- 您可以通过单击**更多信息**来查看有关扩展的详细信息。在更新可用时，**更多信息**显示有关更新的信息。
- 您可以通过单击**先决条件**，查看运行扩展的先决条件，例如，是否需要 IBM SPSS Modeler - Integration Plug-in for R。更新可用时，**先决条件**会显示有关更新的信息。

## 优化依据

您可以优化显示的扩展集。您可以按以下项进行优化：扩展的常规类别、实现扩展所采用的语言、提供扩展的组织类型以及扩展的状态。针对每个组（例如，类别），可以选择按其优化显示的扩展列表的多个项。还可以按搜索项进行优化。搜索不区分大小写，星号 (\*) 的处理与其他任何字符相同，不指示通配符搜索。

- 要优化显示的扩展列表，请单击**应用**。光标处于**搜索框**中时按 Enter 键与单击**应用**的作用一样。
- 要重置列表以显示所有可用扩展，请删除**搜索框**中的任何文本，取消选择所有项，并单击**应用**。

## 如何获取集成插件

要获取 **IBM SPSS Modeler - Integration Plug-in for R**，请执行以下操作：

安装 IBM SPSS Modeler - Essentials for R，可以从 [https://github.com/IBMPredictiveAnalytics/R\\_Essentials\\_Modeler/releases/](https://github.com/IBMPredictiveAnalytics/R_Essentials_Modeler/releases/) 或 IBM SPSS Statistics 社区（网址为 <https://www.ibm.com/products/spss-statistics/support>）获取。IBM SPSS Modeler - Essentials for R 包含 IBM SPSS Modeler - Integration Plug-in for R。R 必备知识 但它不包括 R 编程语言。在安装 IBM SPSS Modeler - Essentials for R 之前，如果尚未安装 R 版本 4.0，则需要安装它。可以从 <https://cran.r-project.org/> 获取该版本。建议下载和安装 R 4.0.x。

注：如果要在无法访问互联网的计算机上安装 R 必备知识，并且计划使用 R 必备知识 随附的 R 脚本，那么必须获取这些脚本所需的任何 R 软件包，并手动将其安装在 R 中。要确定特定 R 脚本需要哪些 R 软件包，请打开扩展中心对话框（“扩展” > “扩展中心”），单击“**已安装**”选项卡，然后单击“**更多信息**”以获取期望的扩展。所需 R 程序包将在“扩展详细信息”对话框中列出。可以从任何 R CRAN 镜像站点（可从 <http://www.r-project.org/> 访问）获取 R 程序包。确保获取与 R 版本匹配的程序包的版本。特定于版本的程序包在 CRAN 镜像站点的“Contributed Packages”页面上的链接上提供。

## “已安装”选项卡

“已安装”选项卡会显示计算机上已安装的所有扩展。从“已安装”选项卡中可以选择 GitHub 上的 IBM SPSS 预测性分析集合提供的已安装扩展的更新和移除扩展。要获取安装的扩展的更新，必须具有因特网连接。

- 对于每个扩展，将显示已安装的版本号。提供了因特网连接时，将显示最新版本数以及此版本的关联日期。还提供扩展的简述。
- 您可以通过单击**更多信息**来查看有关扩展的详细信息。在更新可用时，**更多信息**显示有关更新的信息。
- 您可以通过单击**先决条件**，查看运行扩展的先决条件，例如，是否需要 IBM SPSS Modeler - Integration Plug-in for R。更新可用时，**先决条件**会显示有关更新的信息。

## 优化依据

您可以优化显示的扩展集。您可以按以下项进行优化：扩展的常规类别、实现扩展所采用的语言、提供扩展的组织类型以及扩展的状态。针对每个组（例如，类别），可以选择按其优化显示的扩展列表的多个项。还可以按搜索项进行优化。搜索不区分大小写，星号 (\*) 的处理与其他任何字符相同，不指示通配符搜索。

- 要优化显示的扩展列表，请单击**应用**。光标处于**搜索框**中时按 Enter 键与单击**应用**的作用一样。
- 要重置列表以显示所有可用扩展，请删除**搜索框**中的任何文本，取消选择所有项，并单击**应用**。

## 专用扩展

专用扩展是计算机上已经安装但是 GitHub 上的 IBM SPSS 预测性分析集合尚未提供的扩展。用于优化显示的扩展集和查看运行扩展的先决条件的功能不适用于专用扩展。

注：在没有因特网连接的情况下使用“扩展中心”时，“已安装”选项卡的部分功能可能不可用。

## 设置

“设置”选项卡指定是下载并安装针对下载选中的安装，还是下载但不安装。此设置适用于新扩展和现有扩展的更新。如果您正下载扩展以分发给组织内的其他用户，那么可以选择下载扩展而不进行安装。如果不具有用于运行扩展的必备软件但计划获取必备软件，那么还可以选择下载但不安装扩展。

如果选择下载扩展而不进行安装，那么稍后可通过选择**扩展 > 安装本地扩展束**进行安装。对于 Windows，可以通过双击扩展束文件来安装扩展。

## 扩展详细信息

“扩展详细信息”对话框中显示扩展作者提供的信息。除必需信息（例如“摘要”和“版本”）外，作者还可提供指向相关位置（例如作者的主页）的 URL。如果扩展是从“扩展中心”下载的，那么它包含许可证，可以通过单击**查看许可证**来查看。

**定制节点。**“定制节点”表列出了扩展包含的定制节点对话框。

**注：**安装包包含定制节点对话框的扩展可能需要重新启动 IBM SPSS Modeler，以便在定制节点表中查看节点对话框的输入。

**相关性。**“相关性”组列出运行扩展中包含的组件所需的附加组件。

- **R 的集成插件** 扩展的组件可能需要 Integration Plug-in for R。
- **R 程序包。** 列出扩展所需的任何 R 程序包。请参阅主题第 26 页的『必需的 R 软件包』，了解更多信息。

## 安装本地扩展束

要安装存储在本地计算机上的扩展束：

1. 从菜单中选择：

**扩展 > 安装本地扩展束...**

2. 选择扩展束。扩展束的文件类型为 mpe。

**要点：**对于 Windows 7 以及 Windows 更高版本的用户，安装现有扩展束的更新后版本可能需要使用管理员特权运行 IBM SPSS Modeler。您可以通过右键单击 IBM SPSS Modeler 的图标并选择**以管理员身份运行**来使用管理员权限启动 IBM SPSS Modeler。具体而言，如果您接收到表明未能安装一个或多个扩展束的错误消息，请尝试使用管理员权限运行。

## 扩展的安装位置

缺省情况下，扩展会安装到操作系统的常规用户可写位置。

您可以通过使用 `IBM_SPSS_MODELER_EXTENSION_PATH` 环境变量定义路径来覆盖该缺省位置。指定的位置在目标计算机上必须存在。在设置 `IBM_SPSS_MODELER_EXTENSION_PATH` 后，必须重新启动 IBM SPSS Modeler 才能使更改生效。

要在 Windows 上创建环境变量，请从控制面板：

### Windows 7

1. 选择“用户帐户”。
2. 选择**更改我的环境变量**。
3. 单击 **新建**，在 **变量名称** 字段中输入环境变量的名称（例如，`IBM_SPSS_MODELER_EXTENSION_PATH`），并在“变量值”字段中输入一个或多个路径。

### Windows 8 或更高版本

1. 选择“系统”。
2. 选择“高级”选项卡并单击**环境变量**。“高级”选项卡是从高级系统设置访问的。

3. 在“用户变量”部分中，单击 **新建**，在 **变量名称** 字段中输入环境变量的名称（例如，IBM\_SPSS\_MODELER\_EXTENSION\_PATH），并在“变量值”字段中输入一个或多个路径。

**要点:** 对于 Windows 7 以及 Windows 更高版本的用户，安装现有扩展束的更新后版本可能需要使用管理员特权运行 IBM SPSS Modeler。您可以通过右键单击 IBM SPSS Modeler 的图标并选择**以管理员身份运行**来使用管理员权限启动 IBM SPSS Modeler。具体而言，如果您接收到表明未能安装一个或多个扩展束的错误消息，请尝试使用管理员权限运行。

## 在包含扩展的批处理模式下运行流

要将定制对话框构建器扩展与 Modeler Batch 结合使用，请执行以下操作：

1. 在 Modeler Client 中，使用定制对话框构建器来创建和安装定制节点。在缺省情况下，扩展节点文件将位于以下目录：
  - **Windows:** C:\ProgramData\IBM\SPSS\Modeler\\CDB\
  - **Mac:** /Users/yourname/Library/Application Support/IBM/SPSS/Modeler/<version>/CDB/
2. 将 CDB 目录复制到 Modeler Batch 机器。
3. 将环境变量 IBM\_SPSS\_MODELER\_EXTENSION\_PATH 导出到新目录。现在，您可以以批处理模式运行流。

## 必需的 R 软件包

如果您的计算机无法访问 Internet，则需要向其他拥有特定扩展的 R 程序包的人员索取。在扩展安装后，您可以从“扩展详细信息”对话框中查看所需 R 程序包的列表。有关更多信息，请参阅主题 [第 25 页的『扩展详细信息』](#)。软件包可以从 <http://www.r-project.org/> 下载，然后从 R 内安装。有关详细信息，请参阅随 R 分发的 *R 安装和管理指南*。

**注:** 对于 UNIX（包括 Linux）用户，程序包以源代码形式下载，需要进行编译。因此必须在您的计算机上安装相应的工具。请参阅 *R 安装与管理指南* 获得详细信息。具体而言，Debian 用户应当通过 `apt-get install r-base-dev` 安装 r-base-dev 程序包。

## 创建并管理定制节点

针对扩展的定制对话框构建器会创建节点，以便在 SPSS Modeler 流中使用。

使用针对扩展的定制对话框构建器，您可以：

- 创建定制节点对话框，以便执行以 R 或 Apache Spark（通过 Python）实现的节点。有关更多信息，请参阅 [第 28 页的『构建脚本模板』](#)。
- 打开包含定制节点对话框（此对话框可能由其他用户创建）规范的文件，并将此对话框添加到您的 IBM SPSS Modeler 安装中，从而选择性地进行您自己的修改。
- 保存定制节点对话框的规范，以便其他用户可以将其添加到他们的 IBM SPSS Modeler 安装中。
- 创建定制节点和编写 Python for Spark 脚本，以从任何数据源读取数据，并以 Apache Spark 支持的任何数据格式写出数据。有关更多信息，请参阅 [第 47 页的『使用 Python for Spark 导入和导出数据』](#)。
- 创建定制节点和编写 R 脚本，以从数据源所在位置读取数据，将数据写出到 R 支持的任何数据格式。有关更多信息，请参阅 [第 48 页的『使用 R 导入和导出数据』](#)。

在“针对扩展的定制对话框构建器”中，可以在扩展中创建或修改定制节点对话框。打开“针对扩展的定制对话框构建器”时，会创建包含空的定制节点对话框的新扩展。从“针对扩展的定制对话框构建器”保存或安装定制节点对话框时，会将其作为扩展的一部分进行保存或安装。

**注:**

- 无法针对标准 IBM SPSS Modeler 节点创建您自己的节点对话框版本。
- 使用定制对话框构建器创建的节点不支持脚本编制，包括定制对话框构建器 R 节点和定制对话框构建器 Python 节点。



## 如何启动“针对扩展的定制对话框构建器”

从菜单中选择 **扩展 > 定制节点对话框构建器**

注:

- Python 节点取决于 Spark 环境。
- Python 脚本必须使用 Spark API，因为将以 Spark DataFrame 形式表示数据。
- V17.1 中创建的旧节点仍将仅针对 IBM SPSS Analytic Server 运行（数据源自 IBM SPSS Analytic Server 源节点并且尚未抽取到 IBM SPSS Modeler 服务器）。V18.0 或更高版本中创建的新 Python 和“定制对话框构建器”可针对 IBM SPSS Modeler 服务器运行。
- 安装 Python 时，请确保所有用户都有权访问 Python 安装。
- 如果要使用“机器学习库 (MLlib)”，必须安装包含 NumPy 的 Python 版本。然后，必须配置 IBM SPSS Modeler 服务器（或 IBM SPSS Modeler 客户机中的本地服务器）以使用 Python 安装。有关详细信息，请参阅 [第 1 页的『使用 Python for Spark 编写脚本』](#)。

## 定制对话框构建器布局

### 对话框画布

对话框画布是定制对话框构建器的一个区域，您可以在其中设计节点对话框的布局。

### “属性”窗格

属性窗格是定制对话框构建器的一个区域，您可以在其中指定节点对话框中包含的控件的属性以及对话框自身的属性，例如节点类型。

### 工具选用板

工具选用板提供了一组可以包括在定制节点对话框中的控件。您可以从 **视图** 菜单中选择“工具选用板”，以显示或隐藏工具选用板。

### 脚本模板

脚本模板可指定由定制节点对话框生成的 R 脚本或 Python for Spark 脚本。您可以通过单击 **移至新窗口**，将脚本模板窗格移至单独的窗口。要将单独的“脚本模板”窗口移动回“定制对话框构建器”，请单击 **复原到主窗口**。

## 构建定制节点对话框

构建定制节点对话框时涉及的基本步骤如下所示：

1. 指定节点对话框自身的属性，例如启动此节点对话框时显示的标题，以及新节点在 IBM SPSS Modeler 选用板中的位置。有关更多信息，请参阅主题 [第 28 页的『对话框属性』](#)。
2. 指定构成此节点对话框以及任何子对话框的控件，例如字段选择器和复选框。有关更多信息，请参阅主题 [第 30 页的『控件类型』](#)。
3. 创建脚本模板，用于指定节点对话框所生成的 R 代码或 Python for Spark 代码。有关更多信息，请参阅主题 [第 28 页的『构建脚本模板』](#)。
4. 指定包含节点对话框的扩展的属性。有关更多信息，请参阅主题 [第 42 页的『扩展属性』](#)。
5. 将包含节点对话框的扩展安装到 IBM SPSS Modeler，并且/或将相应扩展保存到扩展束 (.mpe) 文件中。有关更多信息，请参阅主题 [第 44 页的『管理定制节点对话框』](#)。

在构建节点对话框时，可以进行预览。有关更多信息，请参阅主题 [第 29 页的『预览定制节点对话框』](#)。

## 对话框属性

“定制对话框构建程序”窗口显示了节点对话框的属性以及所选用户界面控件的属性。要查看和设置对话框属性，请单击画布上任何控件外的区域。如果画布上没有控件，“对话框属性”将始终可见。

**对话框名称。**“对话框名称”属性是必填项，它指定要与节点对话框相关联的唯一名称。为了尽量降低名称冲突可能性，您可以使用贵公司的标识（例如 URL）作为名称的前缀。

**标题。**“标题”属性指定要显示在节点对话框的标题栏中的文本。

**帮助文件。**“帮助文件”属性是可选项，用于指定节点对话框的帮助文件的路径。这是用户单击对话框中的帮助按钮时将要启动的文件。帮助文件必须为 HTML 格式。安装或保存节点对话框后，节点对话框的规范随附指定帮助文件的副本。如果没有关联的帮助文件，运行时间对话框的帮助按钮将被隐藏。

- 当您添加帮助文件时，和帮助文件位于相同目录中的帮助文件本地化版本将会自动添加到节点对话框。帮助文件本地化版本的名称为 <Help File>\_<language identifier>.htm。有关更多信息，请参见主题第 46 页的『创建定制节点对话框的本地化版本』。
- 可以将图像文件和样式表等支持文件添加到节点对话框，方法是先保存节点对话框，然后，您可以将支持文件手动添加到节点对话框文件 (.cfe)。有关访问和手动修改定制节点对话框文件的信息，请参见主题第 46 页的『创建定制节点对话框的本地化版本』中题为“本地化对话框字符串”的部分。

**脚本类型。**指定可用于构建脚本模板的脚本的类型。在 IBM SPSS Modeler 中，可使用 R 脚本编制或 Python for Spark 脚本编制。

**使用模型进行评分。**指定使用模型构建脚本构建的模型是否用于评分。

**节点类型。**指定安装节点对话框时将创建的节点的类型。

**选用板。**指定安装节点对话框时要将新创建的节点添加到的选用板。

**节点图标。**单击省略符 (...) 按钮可以选择要用作新创建的节点的节点图标的图像。选择的图像必须是 .gif 文件。

## 在对话框画布上布置控件

通过将控件从工具选用板拖放到对话框画布上，可以将这些控件添加到定制节点对话框。为了确保与内置的节点对话框保持一致，对话框画布划分为四个可以在其中放置控件的功能列。

- 第一列（最左侧的列）主要用于放置字段选择器控件。
- “子对话框”按钮必须位于左右侧列中（例如，如果仅使用三个列，那么位于第三列），其他控件可位于与“子对话框”按钮相同的列中。在此方面，第四列仅包含子对话框按钮。

虽然未在对话框画布上显示，但将节点对话框安装到 IBM SPSS Modeler 时，相应的按钮将添加到对话框中，例如**确定**、**取消**、**应用**、**重置**以及（适当时）**帮助**和**运行**。这些按钮的存在及其位置均为自动确定。但是，如果没有帮助文件关联到节点对话框（由“对话框属性”中的“帮助文件”属性指定），那么**帮助**按钮将被隐藏。

通过向上或向下拖动控件，可以更改这些控件在一列中的垂直顺序，但是各个控件的确切位置将自动确定。在运行时，当对话框自身调整大小时，各个控件将相应地调整大小。字段选择器之类的控件将自动展开，以填充其下方的可用空间。

## 构建脚本模板

脚本模板指定定制节点对话框将生成的 R 脚本或 Python for Spark 脚本。可以使用单个定制节点对话框来指定一项或多项操作，这些操作将按顺序运行。

脚本模板可以包含静态文本。静态文本与静态文本控件不同；这是节点运行时始终生成的 R 或 Python for Spark 代码。例如，与用户输入无关的命令名和子命令规范是静态文本。另外，脚本模板还可以包含控件标识，在运行时，这些标识将替换为相关联定制节点对话框控件的值。例如，在字段选择器中指定的字段集将通过该字段选择器控件的控件标识来表示。



## 要构建脚本模板

1. 对于与用户指定的值无关的静态文本，请像在“R 构建”节点的 **R 模型构建语法** 字段之类的位置输入 R 脚本或 Python for Spark 脚本一样输入该脚本。
2. 在要插入由控件生成的 R 脚本或 Python for Spark 脚本的位置添加格式为 `%%Identifier%%` 的控件标识，其中 `Identifier` 是控件的“标识”属性值。
  - 您可以通过选中标识表中的某个行、右键单击并选择**添加到脚本模板**来插入控件标识。您还可以通过右键单击画布上的控件并选择**添加到脚本模板**来插入控件标识。
  - 还可通过按 `Ctrl+空格` 键，从可用控件标识列表中进行选择。此列表依次包含控件标识以及通过脚本自动补全功能提供的项。

如果您手动输入标识，请保留任何空格，因为标识中的任何空格均有意义。

在运行时，对于除复选框、复选框组和静态文本控件以外的所有控件，每个标识都将替换为相关联控件的脚本属性的当前值。如果控件在运行时为空，那么不会生成任何脚本。对于复选框和复选框组，该标识将替换为相关联控件的“选中 R 脚本”或“未选中 R 脚本”属性的当前值，具体取决于该控件的当前状态 - 选中或未选中。有关更多信息，请参阅主题 [第 30 页的『控件类型』](#)。

## 示例：在 R 脚本模板中包含运行时值

在此示例中，定制节点对话框将生成并运行 R 脚本，并借助此处显示的特征符对 `R lm` 函数进行调用，从而构建线性回归模型并对该模型进行评分。

```
lm(formula, data)
```

- `formula` 用于指定 `Na~Age` 之类的表达式，其中 `Na` 是模型的目标字段，`Age` 是模型的输入字段。
- `data` 是一个数据帧，其中包含公式中指定的字段的值。

假定有一个定制节点对话框，其中包含单个字段选择器控件，该控件允许用户选择线性模型的输入字段。您在**脚本**选项卡上输入脚本模板，此脚本模板生成并运行用于构建模型的 R 脚本，并且可能如下所示：

```
modelerModel <- lm(Na~%%input%%, data=modelerData)
```

- `%%input%%` 是字段选择器控件的“标识”属性的值。在运行时，它将替换为该控件的脚本属性的当前值。
- 将字段选择器控件的脚本属性定义为 `%%ThisValue%%` 表示，在运行时，此属性的当前值将是该控件的值，也就是从字段选择器中选择的字段。

假设此定制节点对话框的用户选择 `Age` 字段作为模型的输入字段。节点对话框将生成以下 R 脚本：

```
modelerModel <- lm(Na~Age, data=modelerData)
```

您在**评分脚本**选项卡上输入脚本模板，此脚本模板生成并运行用于对模型进行评分的 R 脚本，并且可能如下所示：

```
result <- predict(modelerModel, newdata=modelerData)
var1 <- c(fieldName="predicted",
fieldLabel="", fieldStorage="real", fieldMeasure="", fieldFormat="",
fieldRole="")
modelerDataModel <- data.frame(modelerDataModel, var1)
```

这个 R 脚本不依赖于任何由用户指定的值，而仅依赖于使用模型构建 R 脚本构建的模型。因此，就像在“R 构建”节点的 **R 模型评分语法** 字段中输入模型评分 R 脚本一样输入该脚本。

## 预览定制节点对话框

您可以预览当前在定制对话框构建器中打开的节点对话框。从 IBM SPSS Modeler 中节点的运行时，按原样显示对话框。

- 字段选择器以虚字段进行填充。
- 单击**确定**可关闭预览。

- 如果指定了帮助文件，则会启用**帮助**按钮，并可用它来打开指定文件。如果未指定帮助文件，则预览时会禁用帮助按钮；当实际对话框运行时，此按钮也将被隐藏。

要预览定制节点对话框，请从定制对话框构建器的菜单中选择**文件 > 预览对话框**。

## 控件类型

工具选用板提供了定制节点对话框中可能需要的所有标准控件。

- **字段选择器**：活动数据集中所有字段的列表。有关更多信息，请参阅主题 [第 30 页的『字段选择器』](#)。
- **复选框**：单个复选框。有关更多信息，请参阅主题 [第 31 页的『复选框』](#)。
- **组合框**：用于创建下拉列表的组合框。有关更多信息，请参阅主题 [第 32 页的『组合框』](#)。
- **列表框**：用于创建单选或多选列表的列表框。有关更多信息，请参阅主题 [第 32 页的『组合框』](#)。
- **文本控件**：接受任意文本作为输入内容的文本框。有关更多信息，请参阅主题 [第 33 页的『“文本”控件』](#)。
- **数字控件**：仅接受数字值作为输入内容的文本框。有关更多信息，请参阅主题 [第 34 页的『数字控件』](#)。
- **日期控件**：用于指定日期/时间值（包括日期、时间和日期时间）的微调器 (spinner) 控件。有关更多信息，请参阅主题 [第 35 页的『日期控件』](#)。
- **安全文本**：使用星号掩饰用户输入的文本框。有关更多信息，请参阅主题 [第 35 页的『安全文本』](#)。
- **静态文本控件**：用于显示静态文本的控件。有关更多信息，请参阅主题 [第 36 页的『静态文本控件』](#)。
- **颜色选取器**：用于指定颜色和生成关联 RGB 值的控件。有关更多信息，请参阅主题 [第 36 页的『颜色选取器』](#)。
- **表控件**：用于创建具有固定数目列以及在运行时添加的可变数目行的表。有关更多信息，请参阅主题 [第 37 页的『表控件』](#)。
- **项目组**：用于对一组控件（如一组复选框）进行分组的容器。有关更多信息，请参阅主题 [第 38 页的『项目组』](#)。
- **单选组**：一组单选按钮。有关更多信息，请参阅主题 [第 39 页的『单选组』](#)。
- **复选框组**：此容器包含一组通过单个复选框作为组进行启用或禁用的控件。有关更多信息，请参阅主题 [第 39 页的『复选框组』](#)。
- **文件浏览器**：用于浏览文件系统以打开或保存文件的控件。有关更多信息，请参阅主题 [第 40 页的『文件浏览器』](#)。
- **选项卡**：单个选项卡。有关更多信息，请参阅主题 [第 41 页的『选项卡』](#)。
- **子对话框按钮**：用于打开子对话框的按钮。有关更多信息，请参阅主题 [第 41 页的『子对话框按钮』](#)。

## 字段选择器

" 字段选择器 " 控件显示可用于节点对话框的最终用户的字段列表。您可以显示活动数据集中的所有字段（默认），也可以根据类型和测量级别过滤列表，例如，具有刻度测量级别的数字字段。还可以将任何其他“字段选择器”指定为当前“字段选择器”的字段源。“字段选择器”控件具有下列属性：

**标识。** 控件的唯一标识。

**标题。** 控件上方显示的可选标题。对于多行标题，请使用 \n 来指定换行符。

**标题位置。** 指定标题相对于控件的位置。值为“顶部”和“左侧”，其中“顶部”是缺省值。仅当选择器类型设置为选择单个字段时，此属性才适用。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。仅当鼠标指针悬停在此控件的标题区域上时，才会显示指定的文本。将鼠标悬停在某个列示字段上将显示字段名称和标签。

**助记键。** 这是标题中的可选字符，用作控件的键盘快捷键。这些字符在标题中显示时带有下划线。按 Alt + [助记键] 可激活快捷方式。

**选择器类型。** 指定可使用定制节点对话框中的字段选择器从字段列表选择单个字段还是多个字段。

**分隔符类型。** 指定生成的脚本中的所选字段之间的定界符。允许的分隔符是空格、逗号和加号 (+)。您还可以输入任意单个字符作为分隔符。

**最小字段数。** 必须为控件指定的最小字段数 (如果具有)。

**最大字段数。** 可以为控件指定的最大字段数 (如果具有)。

**执行时必需。** 指定该控件中是否需要值, 以便继续执行操作。如果指定了 **True**, 那么节点对话框的用户必须为此控件指定值, 否则单击**确定**按钮将出错。如果指定了 **False**, 那么此控件中缺少值并不会对**确定**按钮的状态造成任何影响。

**变量过滤。** 允许过滤在控件中显示的字段集。可以按字段类型与测量级别过滤, 并可指定在字段列表中包括多个响应集。单击省略符 (...) 按钮将打开“过滤器”对话框。您也可以通过双击画布上的“字段选择器”控件来打开“过滤器”对话框。请参阅主题第 31 页的『对字段列表进行过滤』, 了解更多信息。

**资源源。** 指定另一个“字段选择器”是当前“字段选择器”的字段源。未设置“字段源”属性时, 字段的源为活动数据集。单击省略符 (...) 按钮以打开对话框并指定字段源。

**脚本。** 指定由此控件生成并运行脚本, 并且可插入到脚本模板中。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本, 请单击省略符 (...) 按钮, 并在脚本属性对话框中输入脚本。
- 值 `%%ThisValue%%` 指定控件的运行时值, 它是字段列表。这是缺省选项。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时, 才会显示“启用规则”属性。

## 为“字段选择器”指定“字段源”

“字段源”对话框可指定“字段选择器”中显示的资源的源。源可以是任何其他“字段选择器”。您可以从活动数据集选择显示包含在选定控件中的字段或不包含在选定控件中的的字段。

## 对字段列表进行过滤

通过使用与字段选择器控件相关联的“过滤器”对话框, 可以对活动数据集中能够在列表中显示的字段类型进行过滤。还可以指定是否包含与活动数据集关联的多响应集。数字字段包括除日期格式和时间格式以外的所有数字格式。

## 复选框

“复选框”控件是一个简单的复选框, 它可以针对选中和未选中状态生成并运行不同的 R 脚本或 Python for Spark 脚本。复选框控件具有下列属性:

**标识。** 控件的唯一标识。

**标题。** 控件上方显示的可选标题。对于多行标题, 请使用 `\n` 来指定换行符。

**工具提示。** 当用户悬停在控件上时, 显示的可选工具提示文本。

**助记键。** 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。按 `Alt+[记忆键]` 可激活快捷方式。

**缺省值。** 复选框的缺省状态为选中或未选中。

**选中/未选中脚本。** 指定在选中以及未选中此控件的情况下, 生成并运行的 R 脚本或 Python for Spark 脚本。要将脚本包括在脚本模板中, 请使用“标识”属性的值。生成的脚本将插入到该标识的指定位置, 而它是根据“选中脚本”属性还是“未选中脚本”属性生成无关。例如, 如果标识为 `checkbox1`, 那么在运行时, 如果选中该复选框, 则脚本模板中 `%%checkbox1%%` 的实例将替换为选中的脚本属性的值; 如果未选中该复选框, 则替换为未选中的脚本属性的值。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本, 请单击省略符 (...) 按钮, 并在脚本属性对话框中输入脚本。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时, 才会显示“启用规则”属性。

## 组合框

使用“组合框”控件可以创建下拉列表，此下拉列表可以生成并运行特定于所选列表项的 R 脚本和 Python for Spark 脚本。它限制为单选。组合框控件具有下列属性：

**标识。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 控件上方显示的可选标题。对于多行标题，请使用 \n 来指定换行符。

**标题位置。** 指定标题相对于控件的位置。值为“顶部”和“左侧”，其中“顶部”是缺省值。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**列表项。** 单击省略号 (...) 按钮以打开列表项属性对话框，该对话框允许您指定控件的列表项。您还可以通过双击画布上的“组合框”控件来打开“列表项属性”对话框。

**助记键。** 这是标题中的可选字符，用作控件的键盘快捷键。这些字符在标题中显示时带有下划线。快捷键通过按下 Alt+[助记键] 激活。

**可编辑。** 指定组合框是否可编辑。在控件可编辑时，可在运行时输入定制值。

**脚本。** 指定由此控件在运行时生成且可以插入到脚本模板中的 R 脚本或 Python for Spark 脚本。

- 值 `%%ThisValue%%` 指定控件的运行时值，也是缺省值。如果手动定义列表项，那么运行时值是所选列表项的“脚本”属性的值。如果列表项基于目标列表控件，那么运行时值为选定列表项的值。对于多选列表框控件，运行时值为以空格分隔的选定项列表。请参阅主题第 32 页的『为组合框和列表框指定列表项』，了解更多信息。
- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。

**报价处理。** 当脚本属性包含 `%%ThisValue%%` 作为带引号字符串的一部分时，指定处理运行时值 `%ThisValue%` 中的引号。在此上下文中，加引号的字符串是用单引号或双引号引起的字符串。引号处理仅适用于类型与用于引起 `%%ThisValue%%` 的引号相同的引号。提供了以下类型的引号处理。

### Python

运行时值 `%%ThisValue%%` 中与所含引号匹配的引号将使用反斜杠 () 字符进行转义，例如，如果脚本属性为 `'%%ThisValue%%'`，并且组合框的运行时值为**组合框的值**，则生成的脚本为 `'Combo box\s value'`。请注意，`%%ThisValue%%` 使用三重引号引起时，不会执行引号处理。

### R

运行时值 `%%ThisValue%%` 中与所含引号匹配的引号将使用反斜杠 () 字符进行转义，例如，如果脚本属性为 `'%%ThisValue%%'`，并且组合框的运行时值为**组合框的值**，则生成的脚本为 `'Combo box\s value'`。

### 无

与括起的引号匹配的 `%%ThisValue%%` 运行时值中引号将保留，不进行任何修改。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

## 为组合框和列表框指定列表项

使用“列表项属性”对话框可以指定组合框或列表框控件的列表项。

**手动定义的值。** 允许您显式指定各个列表项。

- **标识。** 列表项的唯一标识。
- **名称。** 在列表中出现的该项的名称。名称为必需的字段。
- **缺省值。** 对于组合框，指定列表项是否为组合框中显示的缺省项。对于列表框，指定列表项是否缺省选中。
- **脚本。** 指定选择此列表项时生成的 R 脚本或 Python for Spark 脚本。
- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。



**注:** 可以在现有列表底部的空白行中添加新的列表项。输入标识以外的任何属性会生成唯一的标识, 您可以保留或修改此标识。要删除列表项, 请单击其标识单元格, 并按“删除”。

## 列表框

使用“列表框”控件可以显示项目列表, 这些项目支持单选或多选, 并将生成特定于所选项目的 R 脚本和 Python for Spark 脚本。列表框控件具有下列属性:

**标识。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 控件上方显示的可选标题。对于多行标题, 请使用 \n 来指定换行符。

**工具提示。** 当用户悬停在控件上时, 显示的可选工具提示文本。

**列表项。** 单击省略号 (...) 按钮以打开列表项属性对话框, 您可以从中指定控件的列表项。还可通过双击画布上的列表框控件以打开“列表项属性”对话框。

**助记键。** 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。快捷键通过按下 Alt+[助记键] 激活。

**列表框类型。** 指定列表框是仅支持单选, 还是支持多选。还可指定列表项显示为复选框列表。

**分隔符类型。** 在生成的脚本中的所选列表项之间指定定界符。允许的分隔符是空格、逗号和加号 (+)。还可以输入任意单个字符作为分隔符。

**最小所选项数。** 必须在控件中选择的最小项数 (如果具有)。

**最大所选项数。** 可以在控件中选择的最大项数 (如果具有)。

**脚本。** 指定由此控件在运行时生成且可以插入到脚本模板中的 R 脚本或 Python for Spark 脚本。

- 值 `%%ThisValue%%` 指定控件的运行时值, 它是缺省值。如果手动定义列表项, 那么运行时值是所选列表项的“脚本”属性的值。如果列表项基于目标列表控件, 那么运行时值为选定列表项的值。对于多选列表框控件, 运行时值为以指定分隔符类型分隔 (缺省为空格分隔) 的选定项列表。有关更多信息, 请参阅主题第 32 页的『为组合框和列表框指定列表项』。
- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本, 请单击省略符 (...) 按钮, 并在脚本属性对话框中输入脚本。

**引号处理。** 当脚本属性包含 `%%ThisValue%%` 作为带引号字符串的一部分时, 指定处理运行时值 `%%ThisValue%%` 中的引号。在此上下文中, 加引号的字符串是用单引号或双引号引起来的字符串。引号处理仅适用于类型与用于引起 `%%ThisValue%%` 的引号相同的引号。提供了以下类型的引号处理。

### Python

运行时值 `%%ThisValue%%` 中与所含引号匹配的引号将使用反斜杠 (\) 进行转义。例如, 脚本属性为 `%%ThisValue%%` 且选定的列表项为 `List item's value`, 那么将生成的脚本为 `'List item\'s value'`。请注意, `%%ThisValue%%` 使用三重引号引起时, 不会执行引号处理。

### R

运行时值 `%%ThisValue%%` 中与所含引号匹配的引号将使用反斜杠 (\) 进行转义。例如, 脚本属性为 `%%ThisValue%%` 且选定的列表项为 `List item's value`, 那么将生成的脚本为 `'List item\'s value'`。

### 无

与括起的引号匹配的 `%%ThisValue%%` 运行时值中引号将保留, 不进行任何修改。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时, 才会显示“启用规则”属性。

## “文本”控件

“文本”控件是一个简单的文本框, 它可以接受任意输入并具有下列属性:

**标识。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 控件上方显示的可选标题。对于多行标题, 请使用 \n 来指定换行符。

**标题位置。** 指定标题相对于控件的位置。值为“顶部”和“左侧”, 其中“顶部”是缺省值。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**助记键。** 这是标题中的可选字符，用作控件的键盘快捷键。这些字符在标题中显示时带有下划线。按 Alt + [助记键] 可激活快捷方式。

**文本内容。** 指定内容是任意的，还是文本框必须包含遵守 IBM SPSS Modeler 字段名称规则的字符串。

**缺省值。** 文本框的缺省内容。

**宽度。** 指定控件的文本区域的宽度（字符）。允许的正整数。空值表示将自动确定宽度。

**执行需要。** 指定该控件中是否需要值，以便继续执行操作。如果指定了 **True**，那么节点对话框的用户必须为此控件指定值，否则单击**确定**按钮将出错。如果指定了 **False**，那么此控件中缺少值并不会对**确定**按钮的状态造成任何影响。缺省为 **False**。

**脚本。** 指定 R 脚本或 Python for Spark 脚本，由此控件在运行时生成和运行，可插入到脚本模板中。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。
- 值 `%%ThisValue%%` 指定控件的运行时值，它是文本框的内容。这是缺省值。
- 如果“脚本”属性包含 `%%ThisValue%%`，并且文本框的运行时值为空，那么文本框控件不会生成任何脚本。

**引号处理。** 当脚本属性包含 `%%ThisValue%%` 作为带引号字符串的一部分时，指定处理运行时值 `%%ThisValue%%` 中的引号。在此上下文中，加引号的字符串是用单引号或双引号引起来的字符串。引号处理仅适用于类型与用于引起 `%%ThisValue%%` 的引号相同的引号。提供了以下类型的引号处理。

#### **python**

使用反斜杠字符 (\) 对运行时值 `%%ThisValue%%` 中与括起的引号匹配的引号进行转义。例如，如果脚本属性为 `'%%ThisValue%%'`，文本控件的运行时值为 `Text box's value`，那么生成的脚本将为 `'Text box\'s value'`。`%%ThisValue%%` 使用三重引号引起时，不会执行引号处理。

#### **R**

使用反斜杠字符 (\) 对运行时值 `%%ThisValue%%` 中与括起的引号匹配的引号进行转义。例如，如果脚本属性为 `'%%ThisValue%%'`，文本控件的运行时值为 `Text box's value`，那么生成的脚本将为 `'Text box\'s value'`。

#### **无**

与括起的引号匹配的 `%%ThisValue%%` 运行时值中引号将保留，不进行任何修改。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

## **数字控件**

“数字”控件为文本框，它用于输入数值，并具有下列属性：

**标识符。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 控件上方显示的可选标题。对于多行标题，请使用 \n 来指定换行符。

**标题位置。** 指定标题相对于控件的位置。值为“顶部”和“左侧”，其中“顶部”是缺省值。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**助记键。** 这是标题中的可选字符，用作控件的键盘快捷键。这些字符在标题中显示时带有下划线。按 Alt + [助记键] 可激活快捷方式。

**数值类型。** 指定对输入内容的任何限制。值“Real”指定对输入值除了必须为数值外，没有其它限制。如果值为“Integer”，表示输入值必须为整数。

**微调输入。** 指定控件是否显示为微调器 (spinner)。缺省值为 **False**。

**增量。** 控件显示为微调器 (spinner) 时的增量。

**缺省值。** 缺省值，如果有的话。

**最小值。** 最小允许值，如果有的话。

**最大值。** 最大允许值，如果有的话。

**宽度。** 指定控件的文本区域的宽度（字符）。允许的正整数。空值表示将自动确定宽度。

**执行需要。** 指定该控件中是否需要值，以便继续执行操作。如果指定了 **True**，那么节点对话框的用户必须为此控件指定值，否则单击**确定**按钮将出错。如果指定了 **False**，那么此控件中缺少值并不会对**确定**按钮的状态造成任何影响。缺省为 **False**。

**脚本。** 指定 R 脚本或 Python for Spark 脚本，由此控件在运行时生成和运行，可插入到脚本模板中。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。
- 值 `%%ThisValue%%` 指定控件的运行时值，它是数值。这是缺省值。
- 如果“脚本”属性包含 `%%ThisValue%%`，并且数字控件的运行时值为空，那么数字控件不会生成任何脚本。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

## 日期控件

日期控件是用于指定日期/时间值（包括日期、时间和日期时间）的微调器 (spinner) 控件。日期控件具有下列属性：

**标识符。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 控件上方显示的可选标题。对于多行标题，请使用 `\n` 来指定换行符。

**标题位置。** 指定标题相对于控件的位置。值为“顶部”和“左侧”，其中“顶部”是缺省值。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**助记键。** 这是标题中的可选字符，用作控件的键盘快捷键。这些字符在标题中显示时带有下划线。按 `Alt + [助记键]` 可激活快捷方式。

**类型。** 指定控件是用于日期、时间还是日期时间值。

### 日期

控件指定格式为 `yyyy-mm-dd` 的日历日期。缺省运行时值通过“缺省值”属性指定。

### 时间

控件指定格式为 `hh:mm:ss` 的时间。缺省运行时值为当前的时间。

### 日期时间

控件指定格式为 `yyyy-mm-dd hh:mm:ss` 的日期和时间。缺省运行时值为当前日期和时间。

**缺省值。** 在类型为“日期”时控件的缺省运行时值。您可以指定以显示当前日期或特定日期。

**脚本。** 指定 R 脚本或 Python for Spark 脚本，由此控件在运行时生成和运行，可插入到脚本模板中。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。
- 值 `%%ThisValue%%` 指定控件的运行时值。这是缺省值。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

**注：** 在 R18 之前的 IBM SPSS Modeler 发行版中都不支持日期控件。

## 安全文本

安全文本控件是使用星号掩饰用户输入的文本框。

**标识符。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 控件上方显示的可选标题。对于多行标题，请使用 `\n` 来指定换行符。

**标题位置。** 指定标题相对于控件的位置。值为“顶部”和“左侧”，其中“顶部”是缺省值。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**助记键。** 这是标题中的可选字符，用作控件的键盘快捷键。这些字符在标题中显示时带有下划线。按 Alt + [助记键] 可激活快捷方式。

**宽度。** 指定控件的文本区域的宽度（字符）。允许的正整数。空值表示将自动确定宽度。

**执行所必需。** 指定该控件中是否需要值，以便继续执行操作。如果指定了 **True**，那么节点对话框的用户必须为此控件指定值，否则单击**确定**按钮将出错。如果指定了 **False**，那么此控件中缺少值并不会对**确定**按钮的状态造成任何影响。缺省值为 **False**。

**脚本。** 指定 R 脚本或 Python for Spark 脚本，由此控件在运行时生成和运行，可插入到脚本模板中。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。
- 值 `%%ThisValue%%` 指定控件的运行时值，它是文本框的内容。这是缺省值。
- 如果脚本属性包含 `%%ThisValue%%` 并且受保护的文本控件的运行时值为空，那么受保护的文本控件不会生成任何 R 脚本或 Python for Spark 脚本。

**引号处理。** 指定在以下情况下处理运行时值 `%%ThisValue%%` 中的引号：当脚本属性在引起来的字符串中包含 `%%ThisValue%%` 时。在此上下文中，加引号的字符串为使用单引号或双引号引起的字符串。引号处理仅适用于类型与用于引起 `%%ThisValue%%` 的引号相同的引号，且仅当 `Encrypt passed value=False` 时适用。提供了以下类型的引号处理。

#### Python

使用反斜杠字符 (\) 对运行时值 `%%ThisValue%%` 中与括起的引号匹配的引号进行转义。例如，如果脚本属性为 `'%%ThisValue%%'`，控件的运行时值为 `Secured Text's value`，那么生成的脚本为 `'Secured Text\'s value'`。在三重引号括起 `%%ThisValue%%` 时，不执行引号处理。

#### R

使用反斜杠字符 (\) 对运行时值 `%%ThisValue%%` 中与括起的引号匹配的引号进行转义。例如，如果脚本属性为 `'%%ThisValue%%'`，控件的运行时值为 `Secured Text's value`，那么生成的脚本为 `'Secured Text\'s value'`。

#### 无

与括起的引号匹配的 `%%ThisValue%%` 运行时值中引号将保留，不进行任何修改。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

注：在 R18 之前的 IBM SPSS Modeler 发行版中都不支持安全文本控件。

## 静态文本控件

使用“静态文本”控件可以向节点对话框添加文本块，此控件具有下列属性：

**标识符。** 控件的唯一标识。

**标题。** 文本块的内容。针对多行内容，使用 \n 指定换行符。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

## 颜色选取器

颜色选取器控件为用于指定颜色和生成关联 RGB 值的用户界面。颜色选取器控件具有下列属性：

**标识符。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 控件上方显示的可选标题。对于多行标题，请使用 \n 来指定换行符。

**标题位置。** 指定标题相对于控件的位置。值为“顶部”和“左侧”，其中“顶部”是缺省值。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**助记键。** 这是标题中的可选字符，用作控件的键盘快捷键。这些字符在标题中显示时带有下划线。按 Alt + [助记键] 可激活快捷方式。



**脚本。** 指定 R 脚本或 Python for Spark 脚本，由此控件在运行时生成和运行，可插入到脚本模板中。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。
- 值 `%%ThisValue%%` 指定控件的运行时值，即选中的颜色的 RGB 值。RGB 值表示为采用以下顺序的整数的空格分隔列表：R 值、G 值和 B 值。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

注：“颜色选择器”控件在 IBM SPSS Modeler R18 之前的发行版中不受支持。

## 表控件

表控件用于创建具有固定数目列以及在运行时添加的可变数目行的表。表控件具有下列属性：

**标识符。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 控件上方显示的可选标题。对于多行标题，请使用 \n 来指定换行符。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**助记键。** 这是标题中的可选字符，用作控件的键盘快捷键。这些字符在标题中显示时带有下划线。快捷键通过按下 Alt+[助记键] 激活。

**重新排序按钮。** 指定是否将上移和下移按钮添加到表中。在运行时会使用这些按钮来对表的行进行重新排序。

**表列。** 单击省略号 (...) 按钮以打开 表列对话框，您可以在其中指定表的列。

**最小行数。** 必须位于表中的最小行数。

**最大行数。** 可位于表中的最大行数。

**运行时必需。** 指定该控件中是否需要值，以便继续执行操作。如果指定了 **True**，那么节点对话框的用户必须为此控件指定值，否则单击**确定**按钮将出错。如果指定了 **False**，那么此控件中缺少值并不会对**确定**按钮的状态造成任何影响。

**脚本。** 指定由此控件在运行时生成且可以插入到脚本模板中的 R 脚本或 Python for Spark 脚本。

- 值 `%%ThisValue%%` 指定控件的运行时值，也是缺省值。运行时值是表中每个列（从最左边列开始）生成的脚本的空格分隔列表。如果脚本属性包含 `%%ThisValue%%`，且列都未生成脚本，那么整个表不会生成任何脚本。
- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

注：“表”控件在 IBM SPSS Modeler R18 之前的发行版中不受支持。

## 指定表控件的列

“表列”对话框指定表控件的列的属性。

**标识。** 列的唯一标识。

**列名。** 表中显示的列名。

**内容。** 指定列的数据类型。值**实数**指定输入的值必须是数字，此外没有任何限制。值 **Integer** 指定输入值必须为整数。值**任何**指定对输入值没有任何限制。值为 **Variable Name** 指定此值必须满足 IBM SPSS Statistics 中有效变量名称的需求。

**缺省值。** 在运行时将新行添加到表时此列的缺省值（如果具有）。

**分隔符类型。** 在生成的脚本中列的值之间指定定界符。允许的分隔符是空格、逗号和加号 (+)。还可以输入任意单个字符作为分隔符。

**加上引号。** 指定在生成的脚本中是否使用双引号将列中每个值引起。

**引号处理。** 指定“加引号”属性为 true 时处理列的单元格输入的引号。引号处理仅适用于单元格值中双引号。提供了以下类型的引号处理。

**python**

单元格值中的双引号用反斜杠 (\) 转义。例如，如果单元格值是“This "quoted" value”，则生成的脚本为“This \"quoted\" value”。

**R**

单元格值中的双引号用反斜杠 (\) 转义。例如，如果单元格值是“This "quoted" value”，则生成的脚本为“This \"quoted\" value”。

**无**

单元格值中的双引号将保留不做修改。

**宽度（字符）。** 指定列的宽度（字符数）。允许的正整数。

**脚本。** 指定此列在运行时生成的 R 脚本或 Python for Spark 脚本。针对表整体生成的脚本是一个表中每个列（从最左侧的列开始）生成的脚本的以空格分隔的列表。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。
- 值 %%ThisValue%% 指定列的运行值，它是列中值的列表，以指定分隔符分隔。
- 如果列的“脚本”属性包含 %%ThisValue%%，并且列的运行值为空，那么列不会生成任何脚本。

**注：**您可以在“表列”对话框中现有列表底部的空行中，针对一个新表列添加一行。输入标识以外的任何属性会生成唯一的标识，您可以保留或修改此标识。您可以通过单击“表”列的标识单元并按 Delete 键来删除“表”列。

## 链接到控件

您可以将“表”控件链接至“字段选择器”控件。将表控件链接到字段选择器时，针对字段选择器中每个字段，表中会具有一行。通过向“字段选择器”添加字段将行添加到表。通过从字段选择器移除字段会从表删除行。例如，已链接的“表”控件可用于指定“字段选择器”中选中的字段的属性。

要启用链接，表必须具有包含“内容的变量名称”属性的列，并且画布上至少有一个字段选择器控件。

要将表控件链接到字段选择器，请从“表列”对话框的“链接到控件”组中的“可用控件”列表中指定字段选择器。然后，选择称为**链接的列表**列，该列定义链接。在呈现表时，链接的列显示字段选择器中的当前字段。仅可以链接至多字段选择器。

## 项目组

项目组控件是包含其他控件的容器，使用此控件可以对多个控件生成的脚本进行分组和控制。例如，您有一组用于为子命令指定可选设置的复选框，但希望只有在至少选中了一个复选框的情况下，才生成该子命令的脚本。这通过使用“项目组”控件作为复选框控件的容器来实现。可以在项目组中包含以下类型的控件：字段选择器、复选框、组合框、列表框、文本控件、数字控件、静态文本、单选组和文件浏览器。“项目组”控件具有下列属性：

**标识符。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 这个组的可选标题。对于分布在多行的标题，使用 \n 来指定换行符。

**脚本。** 指定 R 脚本或 Python for Spark 脚本，由此控件在运行时生成和运行，可插入到脚本模板中。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。
- 可以将标识包含在“项目组”中的任何控件中。在运行时，这些标识将替换为控件所生成的 R 脚本或 Python 脚本。
- 值 %%ThisValue%% 将生成项目组中各个控件所生成的 R 脚本或 Python 脚本的列表（以空格分隔），并按其相应控件在此组中的显示顺序排列（从上到下）。这是缺省值。如果“脚本”属性包含 %%ThisValue%%，并且此项目组中的所有控件均未生成脚本，那么项目组整体不生成任何脚本。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

## 单选组

“单选按钮组”控件是一组单选按钮的容器，每个单选按钮又可包含一组嵌套控件。单选按钮组控件具有下列属性：

**标识符。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 适用于组的可选标题。对于分布在多行的标题，使用 \n 来指定换行符。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**单选按钮。** 单击省略号 (...) 按钮可打开 单选按钮组属性对话框，该对话框允许您指定单选按钮的属性，以及添加或移除按钮组中的按钮。在给定的单选按钮下嵌套控件的功能是单选按钮的属性，在“单选按钮组属性”对话框中进行设置。请注意，还可通过双击画布上的“单选组”控件以打开“单选组属性”对话框。

**脚本。** 指定由此控件在运行时生成且可以插入到脚本模板中的 R 脚本或 Python for Spark 脚本。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。
- 值 `%%ThisValue%%` 指定单选按钮组的运行时值，它是所选单选按钮的脚本属性的值。这是缺省值。如果“脚本”属性包括 `%%ThisValue%%`，且所选单选按钮未生成脚本，那么单选按钮组也不会生成任何脚本。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

## 定义单选按钮

使用“单选按钮组属性”对话框中可以指定一组单选按钮。

**标识。** 单选按钮的唯一标识。

**列名。** 在单选按钮旁边显示的名称。名称为必需的字段。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**助记键。** 在名称中用作记忆键的可选字符。指定字符必须存在于名称中。

**嵌套组。** 指定是否可以将其他控件嵌套在此单选按钮之下。缺省为 `False`。如果嵌套组属性设置为 `True`，那么会在相关单选按钮下显示矩形放置区，并且可以嵌套和缩进。可以在单选按钮下嵌套以下控件：字段选择器、复选框、文本控件、静态文本、数字控件、组合框、列表框和文件浏览器。

**缺省。** 指定此单选按钮是否为缺省选择。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

**脚本。** 指定选择单选按钮时生成的 R 脚本或 Python for Spark 脚本。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。
- 对于包含嵌套控件的单选按钮，值 `%%ThisValue%%` 将生成各个嵌套控件所生成的 R 脚本或 Python for Spark 脚本的列表（以空格分隔），并按其相应控件在此单选按钮下的显示顺序排列（从上到下）。

可以在现有列表底部的空白行中添加新的单选按钮。输入标识以外的任何属性会生成唯一的标识，您可以保留或修改此标识。通过单击单选按钮的标识单元格并按 `Delete` 键，可以删除该按钮。

## 复选框组

“复选框组”控件是包含一组通过单个复选框进行启用或禁用的控件的容器。在“复选框组”中可以包含下列控件类型：字段选择器、复选框、组合框、列表框、文本控件、数字控件、静态文本、单选组和文件浏览器。。复选框组控件具有下列属性：

**标识。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 组的可选标题。对于分布在多行的标题，使用 \n 来指定换行符。

**复选框标题。** 与控制复选框一起显示的可选标签。对于多行标题，可使用 \n 指定换行。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**助记键。** 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。按 Alt+[记忆键] 可激活快捷方式。

**缺省值。** 控制复选框的缺省状态为选中或未选中。

**选中/未选中 R 脚本。** 指定在选中以及未选中此控件的情况下，生成的 R 脚本。要将 R 脚本包括在脚本模板中，请使用“标识”属性的值。生成的 R 脚本将插入到该标识的指定位置，而与它是根据“选中 R 脚本”属性还是“未选中 R 脚本”属性生成无关。例如，如果标识为 `checkboxgroup1`，那么在运行时，如果选中该复选框，则脚本模板中 `%%checkboxgroup1%%` 的实例将替换为选中的 R 脚本 属性的值；如果未选中该复选框，则替换为未选中的 R 脚本 属性的值。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。
- 可以将标识包含在“复选框组”中的任何控件中。在运行时，这些标识将替换为控件所生成的 R 脚本。
- 值 `%%ThisValue%%` 可在选中的 R 脚本 或未选中的 R 脚本 属性中使用。它将生成复选框组中各个控件所生成的 R 脚本的列表（以空格分隔），并按其相应控件在此组中的显示顺序排列（从上到下）。
- 缺省情况下，选中的 R 脚本 属性的值为 `%%ThisValue%%`，未选中的 R 脚本 属性为空白。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

## 文件浏览器

“文件浏览器”控件包含一个文件路径文本框和一个浏览器按钮，它可打开一个标准 IBM SPSS Modeler 对话框以打开或保存文件。“文件浏览器”控件具有下列属性：

**标识符。** 控件的唯一标识。这是在脚本模板中引用控件时要使用的标识。

**标题。** 控件上方显示的可选标题。对于多行标题，请使用 \n 来指定换行符。

**标题位置。** 指定标题相对于控件的位置。值为“顶部”和“左侧”，其中“顶部”是缺省值。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**助记键。** 这是标题中的可选字符，用作控件的键盘快捷键。这些字符在标题中显示时带有下划线。按 Alt + [助记键] 可激活快捷方式。

**文件系统操作。** 指定由浏览器按钮打开的对话框是用来打开文件还是保存文件。值“打开”表示浏览对话框将验证指定的文件是否存在。如果值为“Save”，表示浏览器对话框不会验证指定文件是否存在。

**浏览器类型。** 指定浏览对话框是用于选择文件（查找文件）还是选择文件夹（查找文件夹）。

**文件过滤器。** 单击省略号 (...) 按钮以打开文件过滤对话框，您可以从中指定打开或保存对话框的可用文件类型。缺省情况下允许所有文件类型。请注意，您也可以通过双击画布上的“文件浏览器”控件来打开“文件过滤器”对话框。

**文件系统类型。** 在分布式分析方式下，此属性指定打开或保存对话框是浏览运行 IBM SPSS Modeler Server 的文件系统还是浏览本地计算机的文件系统。选择**服务器**浏览服务器的文件系统，或选择**客户端**浏览本地计算机的文件系统。该属性对本地分析模式没有影响。

**执行需要。** 指定该控件中是否需要值，以便继续执行操作。如果指定了 **True**，那么节点对话框的用户必须为此控件指定值，否则单击**确定**按钮将出错。如果指定了 **False**，那么此控件中缺少值并不会对**确定**按钮的状态造成任何影响。缺省值为 **False**。

**缺省值。** 控件的缺省值。

**脚本。** 指定由此控件在运行时生成且可以插入到脚本模板中的 R 脚本或 Python for Spark 脚本。

- 您可以指定过任何有效 R 脚本或 Python for Spark 脚本。针对多行脚本或长脚本，请单击省略符 (...) 按钮，并在脚本属性对话框中输入脚本。
- 值 `%%ThisValue%%` 指定文本框的运行时值，此值是括在双引号内的文件路径，由用户手动指定或者由浏览对话框填写。这是缺省值。
- 如果“脚本”属性包含 `%%ThisValue%%`，并且文本框的运行时值为空，那么文件浏览器控件不会生成任何脚本。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

## 文件类型过滤器

使用“文件过滤器”对话框可以指定一些文件类型，这些文件类型将显示在通过“文件系统浏览器”控件访问的打开及保存对话框的“文件类型”和“另存为类型”下拉列表中。缺省情况下允许所有文件类型。

要指定未在对话框中显式列出的文件类型，请完成下列步骤：

1. 选择“其他”。
2. 输入文件类型名称。
3. 以 `*.后缀名` 形式输入文件类型，例如 `*.xls`。您可以指定多种文件类型，各种文件类型之间以分号分隔。

## 选项卡

选项卡控件用于向节点对话框添加选项卡。可以向新选项卡添加任何其他控件。选项卡控件具有下列属性：

**标识。** 控件的唯一标识。

**标题。** 选项卡的标题。

**位置。** 指定此选项卡在节点对话框上的位置，此位置与节点对话框上的其他选项卡位置相对。

**脚本。** 指定此控件在运行时生成和运行并可插入到脚本模板中的 R 脚本或 Python for Spark 脚本。

- 您可以指定任何有效的 R 脚本或 Python for Spark 脚本，并可以使用 `\n` 作为换行符。
- 值 `%%ThisValue%%` 生成选项卡中每个控件生成的 R 脚本或 Python for Spark 脚本的空格分隔的列表，并按照选项卡中出现的顺序进行排序（从上到下，从左到右）。这是缺省选项。
- 如果“脚本”属性包含 `%%ThisValue%%`，并且此选项卡中的所有控件均未生成 R 脚本或 Python for Spark 脚本，那么选项卡整体不生成任何脚本。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

## 子对话框按钮

“子对话框按钮”指定用于打开子对话框的按钮，以便在对话框生成器中访问子对话框。子对话框按钮控件具有下列属性：

**标识。** 控件的唯一标识。

**标题。** 在按钮中显示的文本。

**工具提示。** 当用户悬停在控件上时，显示的可选工具提示文本。

**子对话框。** 单击省略号 (...) 按钮可打开子对话框的“自定义对话框构建程序”。还可双击子对话框按钮，以打开生成器。

**助记键。** 这是标题中的可选字符，用作控件的键盘快捷键。这些字符在标题中显示时带有下划线。按 `Alt + [助记键]` 可激活快捷方式。

**启用规则。** 指定用于确定何时启用当前控件的规则。单击省略符 (...) 按钮以打开对话框并指定规则。仅当画布上存在可用于指定启用规则的控件时，才会显示“启用规则”属性。

注：“子对话框按钮”控件不能添加到子对话框。

## 子对话框的对话框属性

查看和设置子对话框属性：

1. 在主对话框中双击子对话框按钮以打开子对话框，或单击子对话框按钮并单击省略号 (...) 按钮以显示子对话框属性。
2. 在子对话框中，单击任何控件外部区域中的画布。如果画布上没有控件，子对话框属性将始终可见。

**子对话框名称。** 子对话框的唯一标识。“子对话框名称”为必需字段。

注：如果您在脚本模板（如在 %%My Sub-dialog Name%%）中将子对话框名称指定为标识，那么它将在运行时被替换为子对话框中每个控件生成的脚本的空格分隔列表，按顺序显示（从上到小，从左到右）。

**标题。** 指定在子对话框的标题栏中显示的文本。“标题”属性为可选，但建议填入值。

**帮助文件。** 为子对话框指定可选帮助文件的路径。这是用户单击子对话框中的**帮助**按钮时启动的文件，并可以就是对主对话框指定的帮助文件。帮助文件必须为 HTML 格式。请参阅[对话框属性](#)中“帮助文件”属性的描述以获取更多信息。

## 为控件指定“启用规则”

您可以指定用于确定何时启用控件的规则。例如，您可以指定填充字段选择器时启用单选按钮组。用于指定启用规则的可用选项取决于用于定义规则的控件的类型。

### 字段选择器

您可以指定在使用至少一个字段（非空字段）填充字段选择器时启用当前控件。还可以指定在未填充字段选择器时启用当前控件。

### 复选框或复选框组

您可以指定在选中复选框或复选框组时启用当前控件。还可以指定在未选中复选框或复选框组时启用当前控件。

### 组合框或单选项列表框

您可以指定在组合框或单选项列表框中选择特定值时启用当前控件。还可以指定在组合框或单选项列表框中未选择特定值时启用当前控件。

### 多选列表框

您可以指定“多选列表框”中的选定值中包含某个特定值时启用当前控件。还可以指定特定值不属于多选列表框中选定值时启用当前控件。

### 单选按钮组

您可以指定在选择特定单选按钮时启用当前控件。或者，您可以指定未选中某个特定单选按钮时启用当前控件。

可以为指定启用规则的控件具有关联“启用规则”属性。

注：

- 不管是否启用了用于定义规则的控件，启用规则都适用。例如，请考虑指定填充字段选择器时启用单选按钮组的规则。只要填充字段选择器，就会启用单选按钮组，无论是否启用字段选择器都是如此。
- 禁用选项卡控件时，会禁用选项卡上的所有控件，不管这些控件中任何控件是否具有达到需求的启用规则。
- 禁用复选框组时，就会禁用该组中的所有控件，不管是否已选中控制复选框。

## 扩展属性

“扩展属性”对话框指定用于扩展的自定义对话框构建程序中的当前扩展的相关信息，例如扩展的名称和扩展中的文件。

- 在用于扩展的自定义对话框构建程序中创建的所有定制节点对话框都是扩展的一部分。
- 必须先指定“扩展属性”对话框的“必需”选项卡上的字段，才能安装扩展和其中包含的定制节点对话框。

要指定扩展的属性，请从用于扩展的自定义对话框构建程序中的菜单选择：

## 扩展 > 属性

### 扩展的必需属性

#### 名称

与扩展关联的唯一名称。它可以包含最多三个单词，且不区分大小写。仅限使用 7 位 ASCII 字符。为降低名称冲突可能性，您可能需要使用多个单词的名称，其中首个单词包含您组织的标识（例如 URL）。请注意，在保存扩展时，缺省情况下此名称也将用于扩展束 (.mpe) 文件名。建议在保存时使用缺省名称。如果使用其他名称进行保存，那么以后将无法卸载扩展。

#### 目录

关于扩展的简短描述，用于在单行中显示。

#### 版本

x.x.x 格式的版本标识，其中，标识的每个组件必须是一个整数，例如，1.0.0。如果未提供，那么将隐含零。例如，版本标识 3.1 会表示为 3.1.0。版本标识与 IBM SPSS Modeler 版本无关。

#### 最低 SPSS Modeler 版本

运行扩展所需的最低 SPSS Modeler 版本

#### 文件

“文件”列表显示了扩展中当前包含的文件。单击**添加**以将文件添加到扩展中。还可以从扩展中除去文件，并将文件抽取到指定文件夹中。

- 定制节点对话框的文件类型为 .cfe。
- 从“可选”选项卡上的“本地化”设置中添加扩展组件的翻译文件。
- 您可以在扩展中添加自述文件。将文件名指定为 `ReadMe.txt`。用户可以从显示扩展详细信息的对话框中访问此自述文件。您可以包含自述文件的本地化版本（指定为 `ReadMe_<language identifier>.txt`），如法语版本的 `ReadMe_fr.txt` 所示。

### 扩展的可选属性

#### 常规属性

##### 描述

关于扩展的描述说明，比**摘要**字段中的描述更加详细。例如，您可以列出扩展提供的主要功能。

##### 日期

扩展的当前版本的可选日期。未提供格式化。

##### 作者

扩展的作者。您可能希望包括一个电子邮件地址。

##### 链接

与扩展关联的一组 URL，例如作者的主页。此字段的格式不受限制，因此请确保使用空格、逗号或其他有效的分隔符来分隔多个 URL。

##### 关键字

与扩展关联的一组关键字。

##### 平台

有关在特定操作系统平台上使用扩展时所适用的任何限制的信息。

### 相关性

#### 最高 SPSS Modeler 版本

可在其中运行扩展的 IBM SPSS Modeler 的最高版本。

#### Integration Plug-in for R 必需

指定是否需要 Integration Plug-in for R。

如果扩展需要来自 CRAN 程序包库的任何 R 程序包，那么请在“所需 R 程序包”控件中输入这些程序包的名称。名称区分大小写。要添加第一个程序包，在“所需 R 程序包”控件中单击任意位置以突出显示输入字段。保持光标在给定行中，按 **Enter** 键将创建新行。选中某行并按 **Delete** 键可删除此行。

## 本地化

### 定制节点

您可以在扩展中为定制节点对话框添加属性文件的已翻译版本（指定节点对话框中显示的所有字符串）。要为特定节点对话框添加翻译，请选中此对话框，单击**添加翻译**，并选择包含已翻译的版本的文件夹。特定节点对话框的所有已翻译的文件必须位于相同文件夹中。有关创建翻译文件的指示信息，请参阅第 46 页的『创建定制节点对话框的本地化版本』主题。

### 转换目录文件夹

可以为最终用户从“扩展中心”查看扩展详细信息时显示的扩展的**摘要**和**描述**字段提供本地化版本。扩展的所有本地化文件都必须放在名为 lang 的文件夹中。请浏览至包含本地化文件的 lang 文件夹，并选择此文件夹。

要提供 **摘要** 和 **描述** 字段的本地化版本，请为正在提供翻译的每种语言创建名为 <extension name>\_<language-identifier>.properties 的文件。在运行时，如果找不到当前用户界面语言的 .properties 文件，会使用在“必需”选项卡和“可选”选项卡上指定的**摘要**和**描述**字段的值。

- <extension name> 是扩展的**名称**字段的值，所有空格都替换为下划线字符。
- <language-identifier> 是特定语言的标识。IBM SPSS Modeler 支持的语言的标识如下所示。

例如，名为 MYORG MYSTAT 的扩展的法语翻译将存储在文件 MYORG\_MYSTAT\_fr.properties 中。

.properties 文件必须包含以下两行，它们指定两个字段的本地化文本：

```
Summary=<localized text for Summary field>  
Description=<localized text for Description field>
```

- 关键字 Summary 和 Description 必须是英语，本地化文本必须与关键字在同一行上，不带换行符。
- 文件必须使用 ISO 8859-1 编码。无法在此编码中直接表示的字符必须带 Unicode 转义符 (“\u”)。

包含本地化文件的 lang 文件夹必须具有名为 <language-identifier> 的子文件夹，其中包含特定语言的本地化 .properties 文件。例如，法语 .properties 文件必须在 lang/fr 文件夹中。

### 语言标识

**de**。德语

**en**。英语

**es**。西班牙语

**fr**。法语

**it**。意大利语

**ja**。日语

**ko**。韩国语

**pl**。波兰语

**pt\_BR**。巴西葡萄牙语

**ru**。俄语

**zh\_CN**。简体中文

**zh\_TW**。繁体中文

## 管理定制节点对话框

通过用于扩展的自定义对话框构建程序可以管理由您或其他用户在扩展中创建的定制节点对话框。定制节点对话框必须安装到需要它们的 SPSS Modeler 客户机或 SPSS Modeler Batch 的所有实例，才可使用。请注意，以服务器方式使用定制对话框节点，不需要在 SPSS Modeler Server 上进行任何安装。

**注：**您只能对 IBM SPSS Modeler 中创建的定制节点对话框进行修改。



## 打开包含定制节点对话框的扩展

您可以打开扩展束文件 (.mpe)，这些文件包含一个或多个定制节点对话框的规范，或者可以打开已安装的扩展。可以修改扩展中的任何节点对话框，然后保存或安装扩展。安装扩展也会安装扩展中包含的节点对话框。保存扩展也会保存已对扩展中任何节点对话框作出的更改。

要打开扩展束文件，请从用于扩展的自定义对话框构建程序中的菜单选择：

文件 > 打开

要打开已安装的扩展，请从用于扩展的自定义对话框构建程序中的菜单选择：

文件 > 打开已安装的

注：如果打开已安装的扩展以对其进行修改，选择**文件 > 安装**将重新安装该扩展，并替换现有版本。对于通过定制对话框构建器创建的节点，使用它的上下文菜单中的**编辑**并不会在定制对话框构建器中打开该节点对话框。

## 保存至扩展束文件

保存在用于扩展的自定义对话框构建程序中打开的扩展也会保存扩展中包含的定制节点对话框。扩展将保存到扩展束文件 (.mpe)。建议保留与您在“扩展属性”对话框 **名称** 字段中指定的名称匹配的缺省文件名。

从用于扩展的自定义对话框构建程序中的菜单选择：

文件 > 保存

## 安装扩展

安装在用于扩展的自定义对话框构建程序中打开的扩展也会安装扩展中包含的定制节点对话框。安装现有扩展将会替换现有版本，包括替换已安装的扩展中的所有定制节点对话框。

要安装当前打开的扩展，请从用于扩展的自定义对话框构建程序中的菜单选择：

文件 > 安装

缺省情况下，扩展会安装到操作系统的常规用户可写位置。有关更多信息，请参阅主题 [第 25 页的『扩展的安装位置』](#)。

注：在打开的流中，不会替换扩展中包含的节点对话框的现有版本。您打开包含已重新安装的“定制对话框构建器”节点的流时，将接收到一条警告消息。

## 卸载扩展

从用于扩展的自定义对话框构建程序中的菜单选择：

文件 > 卸载

卸载扩展也会卸载扩展中包含的所有定制节点对话框。您也可以从“扩展中心”卸载扩展。

注：要成功卸载扩展，扩展束 .mpe 文件名必须与“扩展属性”对话框中指定的名称匹配。这是缺省文件名。如果修改了文件名，那么请进行重命名以匹配“名称”字段，然后重试卸载。

## 导入定制对话框包文件

可以将定制对话框包 (.cfd) 文件导入用于扩展的自定义对话框构建程序中。系统会将 .cfd 文件转换为 .cfe 文件并添加到新扩展。

从用于扩展的自定义对话框构建程序中的菜单选择：

文件 > 导入

您也可以从“扩展属性”对话框将 .cfe 文件添加到扩展，该对话框可从用于扩展的自定义对话框构建程序中的 **扩展 > 属性** 访问。

## 将定制节点对话框添加到扩展

可以将新的定制节点对话框添加到扩展。

从用于扩展的自定义对话框构建程序中的菜单选择：

扩展 > 新建对话框

## 在扩展中的多个定制节点对话框之间切换

如果当前扩展包含多个定制节点对话框，您可以在这些对话框之间切换。

从用于扩展的自定义对话框构建程序中的菜单选择：

扩展 > 编辑对话框并选择要使用的定制节点对话框。

## 创建新扩展

在用于扩展的自定义对话框构建程序中创建新扩展时，系统会将新的空定制节点对话框添加到扩展。

要创建新扩展，请从用于扩展的自定义对话框构建程序中的菜单选择：

文件 > 新建

## SPSS Modeler Batch 或 IBM SPSS 协作和部署服务 中的扩展

要在 SPSS Modeler Batch 或 IBM SPSS 协作和部署服务 安装中使用扩展，请确保在目标环境上已定义环境变量 `IBM_SPSS_MODELER_EXTENSION_PATH`，并且此环境变量指向包含扩展的位置。如果在定义 `IBM_SPSS_MODELER_EXTENSION_PATH` 环境变量之前，包含定制节点的流存储到 IBM SPSS Collaboration and Deployment Services Repository，那么必须将流重新存储到存储库，这些流才会成功运行。

注：请确保 SPSS Modeler 的 SPSS Modeler Batch 或 IBM SPSS 协作和部署服务 适配器的版本与创建扩展的 SPSS Modeler 客户机的版本相匹配。

## 创建定制节点对话框的本地化版本

您可以为 IBM SPSS Modeler 支持的任何语言，创建定制节点对话框的本地化版本。可以将定制节点对话框中显示的任何字符串本地化，并可以将可选的帮助文件本地化。

### 要本地化对话框字符串

您必须为计划部署的每种语言创建与定制节点对话框相关联的属性文件的副本。此属性文件包含所有与节点对话框相关联的可本地化字符串。

从扩展中抽取定制节点对话框文件 (.cfe)，方法是在（用于扩展的自定义对话框构建程序中的）“扩展属性”对话框中选择该文件，然后单击**抽取**。然后，抽取 .cfe 文件的内容。 .cfe 文件就是 .zip 文件。抽取的 .cfe 文件内容包括每种受支持语言的属性文件，其中特定语言的文件是按照 `<Dialog Name>_<language identifier>.properties` 的方式来命名（请参阅下表中遵循的语言标识符）。

1. 使用支持 UTF-8 的文本编辑器（例如 Windows 上的记事本打开您计划翻译的每个属性文件。修改需要本地化的任何属性的相关值，但不要更改属性名称。与特定控件关联的属性是以控件的标识为前缀。例如，某个标识为 `options_button` 的控件的“工具提示”属性为 `options_button_tooltip_LABEL`。标题属性是按照 `<identifier>_LABEL` 的规则来简单地命名，例如 `options_button_LABEL`。
2. 从“扩展属性”对话框的“可选”选项卡上的“本地化”设置将属性文件的本地化版本重新添加到定制节点对话框文件 (.cfe)。有关更多信息，请参阅主题 第 43 页的『扩展的可选属性』。

启动节点对话框时，IBM SPSS Modeler 将搜索语言标识与当前语言相匹配的属性文件，其中，当前语言由“选项”对话框中“常规”选项卡上的“语言”下拉菜单指定。如果找不到此类属性文件，则使用缺省文件 `<Dialog Name>.properties`。

## 对帮助文件进行本地化

1. 创建与定制节点对话框相关联的帮助文件的副本，然后针对所需语言将文本本地化。
2. 使用下表中的语言标识符，将副本重命名为 <Help File>\_<language identifier>.htm。例如，如果帮助文件为 myhelp.htm，且您打算创建德语版本的文件，则本地化后的帮助文件应命名为 myhelp\_de.htm。

将帮助文件的所有本地化版本存储在与非本地化版本相同的目录中。当您从对话框属性的“帮助文件”属性添加非本地化的帮助文件时，本地化版本会自动添加到节点对话框。

如果存在其它需要本地化的辅助文件，例如图像文件，则必须手动修改主帮助文件中的相应路径，以指向本地化后的版本。必须将包括本地化版本在内的辅助文件手动添加到定制节点对话框 (.cfe) 文件。请参见前面题为“本地化对话框字符串”的部分，了解关于访问和手动修改定制节点对话框文件的信息。

启动节点对话框时，IBM SPSS Modeler 将搜索语言标识与当前语言相匹配的帮助文件，其中，当前语言由“选项”对话框中“常规”选项卡上的“语言”下拉菜单指定。如果找不到这样的帮助文件，那么将使用对节点对话框指定的帮助文件（对话框属性的“帮助文件”属性中指定的文件）。

语言标识

**de。** 德语

**en。** 英语

**es。** 西班牙语

**fr。** 法语

**it。** 意大利语

**ja。** 日语

**ko。** 韩国语

**pl。** 波兰语

**pt\_BR。** 巴西葡萄牙语

**ru。** 俄语

**zh\_CN。** 简体中文

**zh\_TW。** 繁体中文

注意：定制节点对话框及相关帮助文件中的文本不限于 IBM SPSS Modeler 所支持的语言。您可以采用任何语言任意编写节点对话框和帮助文本，而不必创建特定于语言的属性文件和帮助文件。随后，所有节点对话框用户都将以该语言查看文本。

## 使用 Python for Spark 导入和导出数据

通过使用针对扩展的定制对话框构建器，您可以创建定制节点和编写 Python for Spark 脚本，以从数据源所在位置读取数据，将数据写出到 Apache Spark 支持的任何数据格式。

例如，用户希望将他的数据写入数据库。他使用针对扩展的定制对话框构建器和 Python for Spark 来创建定制导出 JDBC 节点，然后运行模型以将数据写入数据库。要从数据库读取数据，还可以创建定制导入 JDBC 节点。例如，还可以使用此方法将数据从 JSON 文件读入 SPSS Modeler。然后，在将其数据读入 SPSS Modeler 后，他可以使用所有可用的 SPSS Modeler 节点来处理业务问题。

注：如果要将 JDBC 与 Python 配合使用以实现 Spark 导入和导出功能，那么必须将 JDBC 驱动程序文件复制到 IBM SPSS Modeler 安装目录内的 as/lib 目录。

### 使用 Python for Spark 导入/导出数据

1. 转至扩展 > 定制节点对话框构建器。
2. 在对话框属性下，为脚本类型选择 **Python for Spark**，为节点类型选择**导入**或**导出**。
3. 根据需要输入其他属性，如对话框名称。

4. 在“脚本”部分中，为导入或导出数据输入或粘贴 Python for Spark 脚本。
5. 单击**安装**以安装 Python for Spark 脚本。新的定制导入节点将添加到“源”选用板，新的定制导出节点将添加到“导出”选用板。

## 使用 R 导入和导出数据

通过使用针对扩展的定制对话框构建器，您可以创建定制节点和编写 R 脚本，以从数据源所在位置读取数据，将数据写出到 R 支持的任何数据格式。

例如，用户希望将她的数据写入数据库。她使用针对扩展的定制对话框构建器和 R 脚本编制来创建定制导出 JDBC 节点，然后运行模型以将数据写入数据库。要从数据库读取数据，还可以创建定制导入 JDBC 节点。例如，还可以使用此方法将数据从 JSON 文件读入 SPSS Modeler。然后，在将其数据读入 SPSS Modeler 后，他可以使用所有可用的 SPSS Modeler 节点来处理业务问题。

### 使用 R 导入/导出数据

1. 转至**扩展 > 定制节点对话框构建器**。
2. 在对话框属性下，为脚本类型选择 **R**，为节点类型选择**导入或导出**。
3. 根据需要输入其他属性，如对话框名称。
4. 在“脚本”部分中，为导入或导出数据输入或粘贴 R 脚本。
5. 单击**安装**以安装 R 脚本。新的定制导入节点将添加到“源”选用板，新的定制导出节点将添加到“导出”选用板。

## 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](http://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](http://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

- .par 文件 [14](#)
- .pim 文件 [14](#)
- “控制台输出”选项卡
  - 扩展模型块 [19](#)
- “扩展”节点 [13](#)
- “扩展变换”节点
  - “控制台输出”选项卡 [20](#)
- “扩展导出”节点
  - “控制台输出”选项卡 [13](#)
- “扩展导入”节点
  - “控制台输出”选项卡 [21](#)
- “扩展模型”节点
  - “控制台输出”选项卡 [18](#)
  - “模型选项”选项卡 [18](#)
  - “文本输出”选项卡 [18](#)
- “扩展输出”节点
  - “控制台输出”选项卡 [15](#)
  - “输出”选项卡 [15](#)
  - “语法”选项卡 [15](#)
- “图形输出”选项卡
  - 扩展模型块 [19](#)
- “文本输出”选项卡
  - 扩展模型块 [19](#)
- “语法”选项卡
  - 扩展模型块 [18](#)
- “IBM SPSS Modeler 扩展”节点 [13](#)
- 定制对话框构建器
  - 安全文本 [35](#)
  - 帮助文件 [28](#)
  - 本地化对话框和帮助文件 [46](#)
  - 表控件 [37](#)
  - 表控件列 [37](#)
  - 布局规则 [28](#)
  - 单选按钮组 [39](#)
  - 对话框属性 [28](#)
  - 对字段列表进行过滤 [31](#)
  - 复选框 [31](#)
  - 复选框组 [39](#)
  - 脚本模板 [28](#)
  - 静态文本控件 [36](#)
  - 列表框 [33](#)
  - 列表框列表项 [32](#)
  - 启用规则 [42](#)
  - 日期控件 [35](#)
  - 数字控件 [34](#)
  - 文本控件 [33](#)
  - 文件类型过滤器 [41](#)
  - 文件浏览器 [40](#)
  - 项目组控件 [38](#)
  - 选项卡 [41](#)
  - 颜色选取器 [36](#)
  - 预览 [29](#)
  - 字段选择器 [30](#)
  - 字段源 [31](#)
  - 子对话框按钮 [41](#)

- 定制对话框构建器 (继续)
  - 子对话框属性 [42](#)
  - 组合框 [32](#)
  - 组合框列表项 [32](#)
  - Python for Spark 数据导出 [47](#)
  - Python for Spark 数据导入 [47](#)
  - R 数据导出 [48](#)
  - R 数据导入 [48](#)
- 发布流
  - IBM SPSS Modeler Solution Publisher [14](#)
- 扩展
  - 安装扩展的更新 [23](#)
  - 查看安装的扩展 [24](#)
  - 查找和安装新扩展 [23](#)
  - 除去扩展 [24](#)
  - 扩展详细信息 [25](#)
- 扩展模型块
  - “控制台输出”选项卡 [19](#)
  - “模型选项”选项卡 [18](#)
  - “图形输出”选项卡 [19](#)
  - “文本输出”选项卡 [19](#)
  - “语法”选项卡 [18](#)
  - 关于 [18](#)
- 扩展输出浏览器 [16](#)
- 扩展束
  - 安装扩展束 [25](#)
- 用于扩展的自定义对话框构建程序
  - 安装包含节点对话框的扩展 [44](#)
  - 保存包含节点对话框的扩展 [44](#)
  - 打开包含节点对话框的扩展 [44](#)
  - 扩展束文件 [44](#)
  - 修改已安装的扩展中的节点对话框 [44](#)
- 允许的 R 语法 [1](#)

## A

- Apache Spark [1](#)

## I

- IBM SPSS Modeler 扩展 [1](#)
- IBM SPSS Modeler Solution Publisher [14](#)

## P

- Python for Spark
  - 切换环境 [11](#)
  - API [1](#), [4](#), [7-9](#)

## R

- R
  - 允许的语法 [1](#)







