

IBM SPSS Modeler 15 R 建模节点

The IBM logo, consisting of the letters 'IBM' in a bold, sans-serif font, with each letter formed by eight horizontal stripes of varying lengths.

注释

在使用本资料及其支持的产品之前，请阅读第 7 页的『声明』中的信息。

产品信息

此版本适用于 IBM SPSS Modeler V15.0.0，及所有后续发行版和修订版，除非在新版本中另有说明。

目录

IBM SPSS Modeler R 节点	1	R 模型块 -“图形输出”选项卡	5
IBM SPSS Modeler R 节点 - 概述	1	R 模型块 -“文本输出”选项卡	5
“R 构建”节点	1	R 模型块 -“控制台输出”选项卡	5
R 构建节点 - 语法选项卡	1	脚本属性	5
“R 构建”节点 -“模型选项”选项卡	2	声明	7
“R 构建”节点 -“控制台输出”选项卡	2	商标	8
句法	2	索引	9
R 模型块	4		
R 模型块 -“语法”选项卡	4		
R 模型块 -“模型选项”选项卡	4		

IBM SPSS Modeler R 节点

IBM SPSS Modeler R 节点 - 概述

作为 IBM® SPSS® Modeler 及其数据挖掘功能的补充，IBM SPSS Modeler R 节点使专业的 R 用户能够输入其自己的 R 脚本以执行进一步的模型构建和模型评分。

如果安装了 R 的兼容副本，那么可以从 IBM SPSS Modeler 与它连接，并使用可在 IBM SPSS Modeler 中部署的定制 R 算法执行模型构建和模型评分。还必须安装了 IBM SPSS Modeler - Essentials for R 的副本。IBM SPSS Modeler - Essentials for R 为您提供了开发与 IBM SPSS Modeler 一起使用的自定义 R 应用程序所需的一切工具。请参阅发行说明获得关于安装指示信息以及版本兼容性的信息。

注意：建议您在**使用 IBM SPSS Modeler R 节点之前**在“类型”节点中实例化数据。

IBM SPSS Modeler R 插件包含下列节点：



“R 构建”节点使您输入定制 R 脚本以执行 IBM SPSS Modeler 中部署的模型构建和模型评分。执行“R 构建”节点会生成 R 模型块。有关更多信息，请参阅主题『“R 构建”节点』。



R 模型块与标准的 IBM SPSS Modeler 模型块相似，并且在将模型从管理器窗格的**模型**选项卡中添加到 IBM SPSS Modeler 画布时，为生成的模型定义要使用的容器。编辑 R 模型块可查看模型输出的各种格式。有关更多信息，请参阅主题第 4 页的『R 模型块』。

注意：IBM SPSS Collaboration and Deployment Services 的评分服务目前不支持这些节点。

“R 构建”节点

借助“R 构建”节点，您可以使用 IBM SPSS Modeler 内的 R 脚本编制执行模型构建和模型评分。这使其可以使用以 R 编写的算法执行模型构建和评分，并使用户开发针对特定问题定制的建模方法。执行“R 构建”节点会生成 R 模型块。

要使用此节点，您的计算机上必须同时安装了 IBM SPSS Modeler - R 的必备要素以及 R 的兼容版本。请参阅发行说明获得安装指示信息和兼容信息。

R 构建节点 - 语法选项卡

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 脚本编制语法。

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 脚本编制语法。执行该节点后，会将此字段中的 R 脚本复制到 R 模型块中。执行 R 模型块时仅运行脚本本身。

注意：有关这些字段中支持使用的语法的更多信息，请参阅第 2 页的『句法』主题。

运行。要创建 R 模型块，请单击**运行**。向“模型”选用板添加 R 模型块，也可以视需要向流画布添加 R 模型块。

“R 构建”节点 -“模型选项”选项卡

模型名称。选择**自动**后，会自动将模型名称设置为字符串“R Syntax”。选择**定制**以在毗邻的文本字段中指定定制模型名称。如果已执行过节点一次，且再次执行该节点之前为指定其他模型名称，那么先前执行中的模型将被覆盖。

阅读数据选项。通过这些选项，可以指定缺少的带有日期或日期时间格式的值和变量的处理方式。

- 将缺少的值转换为 **R“not available”值 (NA)**。如果选择此选项，那么会将所有缺少的值转换为 R NA 值。
- 将日期/时间字段转换为带有对时区的特殊控制的 **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：
 - **R POSIXct**。会将带有日期或日期时间格式的变量转换为 R POSIXct 对象。
 - **R POSIXlt (列表)**。会将带有日期或日期时间格式的变量转换为 R POSIXlt 对象。

*请注意：*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 格式不适用于带有时间格式的变量。

- **输出选项。**通过这些选项，您可以指定来自 R 的输出的显示方式。
 - 以 **HTML 格式显示 R 图形**。如果选择此选项，那么 R 图形会以 HTML 格式显示在 R 模型块的**图形输出**选项卡上。**图形输出**选项卡仅显示在执行在**语法**选项卡中的 **R 模型构建语法**字段中的 R 脚本中生成的那些图。有关更多信息，请参阅主题第 5 页的『R 模型块 -“图形输出”选项卡』。
 - **显示 R 文本输出**。如果选择此选项，那么在执行 **R 模型构建语法**字段中的 R 脚本时生成的任何文本输出都显示在 R 模型块的**文本输出**选项卡上。有关更多信息，请参阅主题第 5 页的『R 模型块 -“文本输出”选项卡』。如果希望将文本输出保存到文件，那么请在脚本中包含对 R sink 功能的调用。**文本输出**选项卡上不显示在对 sink 功能的调用之后生成的所有输出。由执行 R 模型构建脚本产生的 R 错误消息或警告始终显示在 R 构建节点的**控制台输出**选项卡。

“R 构建”节点 -“控制台输出”选项卡

控制台输出选项卡包含在执行**语法**选项卡上的 **R 模型构建语法**字段中的 R 脚本时从 R 控制台接收到的所有输出。此输出可能包括在执行 R 脚本时生成的 R 错误消息或警告以及 R 控制台的文本输出。输出可主要用于调试 R 脚本。**控制台输出**选项卡还包含 **R 模型构建语法**字段中的 R 脚本。每次执行模型构建脚本时，都会使用从 R 控制台接收到的输出覆盖**控制台输出**选项卡的内容。控制台输出不能编辑。

如果在**模型选项**选项卡上选择**显示 R 文本输出**，那么可以在 R 模型块的**文本输出**选项卡上查看 R 控制台中的文本输出。执行 R 脚本时生成的所有 R 错误消息或警告仍将显示在**控制台输出**选项卡上。有关更多信息，请参阅主题第 5 页的『R 模型块 -“文本输出”选项卡』。

句法

在“R 构建”节点对话框的**语法**选项卡上的 **R 模型构建语法**和 **R 模型评分语法**字段内，仅允许 R 识别的语句和功能。

在 R 模型评分脚本中，要使用必须通过包含对 R library 功能的调用来装入的库中的功能，即使已将该库装入 R 模型构建脚本中，也必须将该库装入 R 模型评分脚本中。

在模型评分过程期间，要显示 R 脚本中定义的 R 对象的值，必须在 **R 模型评分语法**字段中包含对 R print 功能的调用。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)
```

R 对象 data 的值会显示在 R 模型块的**控制台输出**选项卡上。

由于 IBM SPSS Modeler 将 R setwd 功能用于控制 R 脚本输出文件的路径，因此您无法在 R 脚本中包含对该功能的调用。

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。

在 **R 模型构建语法** 字段中，必须将执行模型构建脚本时生成的模型对象分配到 R 对象 `modelerModel`。IBM SPSS Modeler 将此模型对象保留在 R 模型块中以便在对数据评分时传递回 R。有关详细信息，请参阅示例。如果您将多个模型对象分配到模型构建脚本中的 `modelerModel`，那么将只保留最后一个模型对象用于对数据评分。

另外，当“R 构建”节点和 R 模型块用于流中时，存在某些会自动填充的 R 对象：

- **modelerData**。这是会使用流入“R 构建”节点和 R 模型块的数据进行自动填充的 R 数据帧。
- **modelerDataModel**。这是会使用流入“R 构建”节点和 R 模型块的数据模型进行自动填充的 R 数据帧。该数据模型描述了流入节点的数据（即元数据）的类型和结构。
- **modelerModel**。分配到模型构建脚本中的 R 对象 `modelerModel` 的模型对象会自动传递到 R 模型块并且可在模型评分脚本中作为引用。

如果在 R 模型评分脚本中使用在 **R 模型构建语法** 字段中的 R 脚本中定义的任何其他 R 对象，那么将无法识别这些 R 对象。如果希望引用模型评分脚本中的这些 R 对象，那么必须在 **R 模型评分语法** 字段中的 R 脚本中重新定义这些 R 对象。

输入 **R 模型构建语法** 和 **R 模型评分语法** 字段中的 R 脚本用于处理 R 对象 `modelerData` 和 `modelerDataModel`。例如，您可能希望使用模型评分 R 脚本添加到数据模型 `modelerDataModel`。必须修改数据模型 `modelerDataModel` 以匹配对数据 `modelerData` 所作的任何更改。在成功执行“R 构建”节点后，会生成新的模型并创建 R 模型块。R 对象 `modelerData` 会自动用作 R 模型块的输出数据。R 对象 `modelerDataModel` 会自动用作 R 模型块的输出数据模型。

示例

在此示例中，通过将变量 `Age` 用作模型输入字段且将变量 `Na` 用作模型目标字段来将线性模型与示例数据集 `DRUG1n` 拟合。然后，该线性模型用于对同一数据集进行评分。

1. 将“变量文件”节点从“源”选用板添加到流画布。
2. 双击“变量文件”节点可打开节点对话框。
3. 单击文件字段右侧的省略号按钮 (...) 以选择 `DRUG1n` 数据集。在演示文件夹中可找到包含 `DRUG1n` 数据集的文件。
4. 单击确定关闭“变量文件”节点。
5. 将“R 构建”节点从“建模”选用板添加到流画布并将其连接到“变量文件”节点。
6. 双击“R 构建”节点可打开节点对话框。
7. 在语法选项卡上的 **R 模型构建语法** 字段中，输入以下 R 脚本：

```
modelerModel<-lm(Na~Age,data=modelerData)
plot(x=modelerData$Na,y=modelerData$Age,xlab="Na",ylab="Age")cor(modelerData$Na,modelerData$Age)
```

R 对象 `modelerData` 会使用 `DRUG1n` 数据集自动填充。

执行该节点后，R 对象 `modelerModel` 包含线性模型分析的结果。

8. 在模型选项选项卡上，选择以 **HTML 格式显示 R 图形**。执行该节点后，R 模型块的图形输出选项卡上会显示针对输入字段 `Age` 的目标字段 `Na` 的图。

- 在**模型选项**选项卡上，选择**显示 R 文本图形**。执行该节点后，目标字段 *Na* 与输入字段 *Age* 之间的关联将写入 R 模型块的**文本输出**选项卡。
- 在**语法**选项卡上的 **R 模型评分语法**字段中，输入以下 R 脚本：

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

执行 R 模型块后，会创建以下 R 对象：

- R 对象 *result* 包含目标字段 *Na* 的预测值，该值是从模型 *modelerModel* 中获取。
 - R 对象 *modelerData* 是包含带有附加字段的原始数据的数据帧，该附加字段包含目标字段的预测值。
 - R 对象 *var1* 为描述了目标字段预测值的类型和结构的数据模型设置新的字段。
 - R 对象 *modelerDataModel* 包含带有目标字段预测值的附加字段的原始数据的数据模型。
- 单击**运行**以执行“R 构建”节点。将 R 模型块添加到“模型”选用板。
 - 将 R 模型块添加到流画布。
 - 将“表”节点从“输出”选用板添加到流画布。
 - 要查看目标字段的预测值，请将“表”节点连接到 R 模型块，双击“表”节点，然后单击**运行**。
 - 该表在名为 *NaPrediction* 的字段中包含预测值；该字段由模型评分 R 脚本创建。

R 模型块

执行“R 构建”节点后会生成 R 模型块并将其置于“模型”选用板上，该选用板包含定义模型构建和模型评分的 R 脚本。缺省情况下，R 模型块包含用于模型评分的 R 脚本、用于读取数据的选项以及 R 控制台的所有输出。

（可选）R 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 R 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。

要使用此节点，您的计算机上必须同时安装了 IBM SPSS Modeler - R 的必备要素以及 R 的兼容版本。请参阅发行说明获得安装指示信息和兼容信息。

R 模型块 -“语法”选项卡

语法选项卡始终存在与 R 模型块中。

R 模型评分语法。该字段中显示用于模型评分的 R 脚本。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**。

编辑。单击**编辑**以使 **R 模型评分语法**字段可编辑。此时，您可以通过在 **R 模型评分语法**字段中进行输入来编辑 R 模型评分脚本。例如，如果您在执行 R 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 R 模型评分脚本。如果您通过执行“R 构建”节点重新生成模型，那么您在 R 模型块中对 R 模型评分脚本所作的全部更改都将丢失。

R 模型块 -“模型选项”选项卡

模型选项选项卡始终存在与 R 模型块中。

阅读数据选项。通过这些选项，可以指定缺少的带有日期或日期时间格式的值和变量的处理方式。

- 将缺少的值转换为 **R“not available”值 (NA)**。如果选择此选项，那么会将所有缺少的值转换为 R NA 值。

- 将日期/时间字段转换为带有对时区的特殊控制的 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：
 - R POSIXct. 会将带有日期或日期时间格式的变量转换为 R POSIXct 对象。
 - R POSIXlt (列表)。 会将带有日期或日期时间格式的变量转换为 R POSIXlt 对象。

请注意: 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。

R 模型块 -“图形输出”选项卡

如果通过在“R 构建”节点对话框的模型选项选项卡上选择以 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 R 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 功能的调用，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。

R 模型块 -“文本输出”选项卡

如果通过在“R 构建”节点对话框的模型选项选项卡上选择显示 R 文本输出复选框进行请求，那么文本输出选项卡存在于 R 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 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 功能的调用，那么在此功能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。

注意: 由执行 R 模型构建脚本产生的 R 错误消息或警告始终显示在 R 构建节点的控制台输出选项卡。

R 模型块 -“控制台输出”选项卡

控制台输出选项卡始终存在与 R 模型块中。 它包含在执行 R 模型块的语法选项卡上的 R 模型评分语法字段中的 R 脚本时从 R 控制台接收到的所有输出。 此输出包括在执行 R 脚本时生成的所有 R 错误消息或警告以及 R 控制台的所有文本输出。 输出可主要用于调试 R 脚本。 每次执行模型评分脚本时，都会使用从 R 控制台接收到的输出覆盖控制台输出选项卡的内容。 控制台输出不能编辑。

脚本属性

本节列示了特定于“R 构建”节点和 R 模型块的脚本编制属性。

“R 构建”节点

“R 构建”节点的脚本编制类型为 buildr。

示例

```
set :buildr.score_syntax = ""
result <- predict(modelerModel, newdata= modelerData)
modelerData <- cbind(modelerData, result)
var1<-c(fieldName="test", fieldLabel="", fieldStorage="real", fieldFormat="", fieldMeasure="",
fieldRole="")
modelerDataModel<-data.frame(modelerDataModel, var1) ""
```

除了通用节点属性外，还有下列属性可用。

表 1. *buildr* 属性.

buildr 属性	值	属性说明
build_syntax	<i>string</i>	用于模型构建的 R 脚本编制语法。
score_syntax	<i>string</i>	用于模型评分的 R 脚本编制语法。
convert_datetime	<i>flag</i>	将带有日期或日期时间格式的变量转换为 R 日期/时间格式的选项。
convert_datetime_class	POSIXct POSIXlt	指定将带有日期或日期时间格式的变量转换为哪种格式的选项。
convert_missing	<i>flag</i>	将缺少的值转换为 R NA 值的选项。
output_html	<i>flag</i>	在 R 模型块中的选项卡上显示图形的选项。
output_text	<i>flag</i>	将 R 控制台文本输出写入 R 模型块中的选项卡的选项。

R 模型块

R 模型块的脚本编制类型是 `applyr`。

除了通用节点属性外，还有下列属性可用。

表 2. *applyr* 属性

applyr 属性	值	属性说明
convert_datetime	<i>flag</i>	将带有日期或日期时间格式的变量转换为 R 日期/时间格式的选项。
convert_datetime_class	POSIXct POSIXlt	指定将带有日期或日期时间格式的变量转换为哪种格式的选项。
convert_missing	<i>flag</i>	将缺少的值转换为 R NA 值的选项。

声明

这些信息开发用于在全球提供的产品和服务。

IBM 可能在其他国家或地区不提供本文中讨论的产品、服务或功能特性。有关您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、程序或服务的操作，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。您可以用书面形式将许可查询寄往：

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

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

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

以下段落对于英国和与当地法律有不同规定的其他国家或地区均不适用：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、适销和适用于某特定用途的保证。某些国家或地区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。

本信息可能包含技术方面不够准确的地方或印刷错误。本信息将定期更改；这些更改将编入本信息的新版本中。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。

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

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

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

IBM 软件部ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

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。

此处包含的任何性能数据都是在受控环境中测得的。因此，在其他操作环境中获得的结果可能会有明显的不同。某些测量可能是在开发级的系统上进行的，因此不保证与一般可用系统上进行的测量结果相同。此外，有些测量可能是通过推算估计出来的。实际结果可能会有差异。本文档的用户应当验证其特定环境的适用数据。

涉及非 IBM 产品的信息可从这些产品的供应商、其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。

所有关于 IBM 未来方向或意向的声明都可随时变更或收回，而不另行通知，它们仅仅表示了目标和意愿而已。

本信息包含了在日常业务运作中所使用到的数据和报告的示例。为了尽可能全面地说明这些数据和报告，该示例包含了个人、公司、商标和产品的名称。所有这些名字都是虚构的，若现实生活中实际业务企业使用的名字和地址与此相似，纯属巧合。

如果您正在查看本信息的软拷贝，图片和彩色图例可能无法显示。

商标

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

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

Linux 是 Linus Torvalds 在美国和@3B72其他国家或地区的注册商标。

Microsoft、Windows、Windows NT 以及 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/或其他国家或地区的注册商标。

Java 和所有基于 Java 的商标及徽标皆为 Oracle 和/或其附属公司的商标或注册商标。

其他产品和服务名称可能是 IBM 或其他公司的商标。

索引

[J]

脚本属性 5

I

IBM SPSS Modeler R 节点 1, 5

R

R 构建节点

“语法”选项卡 1

R 模型块 4

关于 4

“控制台输出”选项卡 5

“模型选项”选项卡 4

“图形输出”选项卡 5

“文本输出”选项卡 5

“语法”选项卡 4

[特别字符]

“控制台输出”选项卡

R 模型块 5

“图形输出”选项卡

R 模型块 5

“文本输出”选项卡

R 模型块 5

“语法”选项卡

R 模型块 4

“R 构建”节点 1

允许的语法 2

“控制台输出”选项卡 2

“模型选项”选项卡 2



Printed in China